دانشکده مهندسی کامپیوتر

بسم الله الرحمن الرحیم

مجموعه سمینارهای کارشناسی ارشد

مهندسی کامپیوتر – معماری کامپیوتر

مهندسی فناوری اطلاعات – شبکه های کامپیوتری

ورودیهای ۱۳۹۰

فهرست مطالب

امنیت در رایانش ابر

علی چاوشی [۱]، مرتضی آنالویی [۲]

[۱] دانشجوی ارشد گرایش معماری کامپیوتر، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
chavoshi@iust.ac.ir

[۲] دانشیار و عضو هیات علمی، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
analoui@iust.ac.ir

چکیده

پیش‌بینی‌ها بر آن است که آینده پردازش و فناوری اطلاعات مبتنی بر رایانش ابر خواهد بود. اما این فناوری با تمام قابلیت‌هایی که دارد، با مشکلی جدی در بحث امنیت مواجه است که تبدیل به مانعی برای رشد سریع آن شده است. در این گزارش مشکلات مطرح در امنیت رایانش ابر معرفی شده و رویکردهای مواجه شدن با این مشکلات بیان شده است. در پایان، امنیت داده به عنوان مهمترین بخش امنیت رایانش ابر معرفی شده است که اغلب مباحث امنیت در رایانش ابر منتج به آن است. از این رو بیشترین تحقیقات مربوط به امنیت رایانش ابر در این حوزه انجام می‌شود.

کلمات کلیدی

رایانش ابر، امنیت.

۱- مقدمه

رایانش ابر یکی از فناوری‌های مطرح در سالهای اخیر بوده است که پیش‌بینی می‌شود در سالهای پیش رو گسترش این فناوری در حوزه‌های اقتصادی، تجاری و تحقیقاتی بیش از پیش گسترش یابد. از ویژگیهای رایانش ابر که منجر به استقبال گسترده از آن شده است میتوان به انعطاف‌پذیری، مقیاس‌پذیری بر حسب تقاضا، پرداخت هزینه مبتنی بر استفاده و دسترس-پذیری آن نام برد. ولی مشکلی که رایانش ابر همچنان به طور جدی با آن روبرو است و به گلوگاهی برای رشد آن مبدل شده است بحث امنیت در آن است. سازمان‌ها، شرکتها و مشتریان شخصی رایانش ابر نیاز دارند معیارها و ضوابط کارآمدی برای برآورده شدن امنیت در رایانش ابر از طرف تامین کنندگان آن ارائه گردد. بحث محرمانگی داده یکی از جدی‌ترین مباحث مطرح در این حوزه است. امنیت داده برای شرکتها وسازمانهایی که صحت و پوشیدگی داده برای آنها، لازمه موفقیت تجاری و اقتصادی است از اهمیت فوق‌العاده‌ای برخوردار است. در این گزارش در ابتدا به معرفی اجمالی ومختصر فناوری رایانش ابر اشاره شده است تا خواننده با تعاریف اولیه آشنایی داشته باشد. سپس در بخش دوم به مسئله امنیت در رایانش ابر پرداخته شده است. نگرانی‌ها و ریسک‌های موجود بررسی و راهکارهای عملی که کاربر باید به آنها توجه کند تا این ریسک‌ها کاهش یابد ذکر شده است. در ادامه بحث، امنیت در بخش‌های مختلف رایانش ابر مورد بررسی قرار گرفته است و آسیب‌پذیری ابر در هر بخش مورد ارزیابی واقع شده است. در هر بخش مزایا و معایب

راهکارهای استفاده شده تاکنون برای ارتقاء امنیت، بیان شده است.

۲- معرفی رایانش ابر

یکی از مسائل مطرح در فناوری‌های سالهای گذشته بهبود و تکامل سیستمهای محاسباتی بوده است. این پیشرفت و تکامل در ابعاد مختلفی همچون افزایش سرعت پردازش اطلاعات، افزایش ظرفیت ذخیره سازی اطلاعات، بهبود در دسترس پذیری آنها و ... انجام پذیرفته است. رایانش ابر ساختاری را معرفی کرده است که بتوان، توان سیستمهای پردازشی و محاسباتی را در تمام این ابعاد راحتتر و کم هزینه تر توسعه داد. در ساختار رایانش ابر، محاسبات، پردازش و هر آنچه که می‌توان در یک سیستم پردازشی مورد توجه قرار داد را به صورت سرویس بر روی اینترنت که در هرجایی قابل دسترس است به کاربر ارائه می‌شود. از این رو توسعه نرم افزارهایی که به جای اجرا بر روی رایانه‌های شخصی، بر روی ابر اجرا می‌شوند و در هرجایی در دسترس هستند، در حال افزایش است. در واقع یکی از رویکردهای اصلی مطرح شده در رایانش ابر، این است که خدمات اینترنتی به مانند یک رایانه واحد در اختیار تمام کاربرانی که به آن متصل هستند قرار گیرد[۱].

.

۲-۱- معماری رایانش ابر

رایانش ابر به ابر یا کلاسترهایی از رایانه‌های توزیع شده گفته میشود که بر حسب تقاضا منابع و خدمات را بر روی شبکه(اینترنت) به شکل سرویس ارائه میدهد[۱]. هر کاربر از یک سرور مجازی در سمت ابر سرویس می‌گیرد و در صورت تغییر نیاز کاربر می‌توان این سرورها را به شکل پویا تغییر مقیاس داد.

۲-۲- انواع ابر[۱]

ابرها را می‌توان از دو نگاهه، دسته بندی با توجه به مقیاس-پذیری ابرها و دسته بندی با توجه به دسترس‌پذیری ابرها تقسیم بندی کرد.

دسته‌بندی بر اساس مقیاس پذیری شامل دو نوع ساختار زیر است.

۱. توده‌های ابری که نمونه‌های محاسباتی را فراهم می-کنند و به گونه‌ای طراحی شده‌اند که با افزایش تعداد نمونه‌های محاسباتی، قابلیت مقیاس پذیری دارند.
۲. توده‌های ابری که ظرفیت محاسباتی را فراهم می‌آورند و با ایجاد مقیاس‌پذیری در ظرفیت، برای پشتیبانی از برنامه‌های کاربردی استفاده می‌شوند.

نوع دسته‌بندی دوم که با نگاه دسترس پذیری ابرها انجام می-شود شامل سه دسته زیراست.

۱. **ابر خصوصی:** برای استفاده انحصاری یک مشتری ایجاد می‌شوند بطوریکه بتواند بیشترین حد کنترل بر روی داده، امنیت و کیفیت سرویس را داشته باشد. در واقع استفاده کننده از ابر خصوصی صاحب زیرساخت است و روی چونگی ارائه برنامه‌های کاربردی کنترل دارد.

۲. **ابر عمومی:** شکلی از ابر است که سرویس‌هایش به طور عمومی در دسترس همگان است. موقعیت قرار گیری ابر اصولا خارج از محل قرارگیری مشتری خواهد بود. در ابرهای عمومی درخواستهای مشتریان مختلف با رعایت امنیت، در یک سیستم ابر ذخیره سازی و پردازشی قرار میگیرد. ابرهای عمومی باعث میشوند هزینه و ریسک ایجاد زیر ساخت و توسعه آن برای مشتریان آن حذف شود.

۳. **ابر ترکیبی:** هر دو مدل خصوصی و عمومی در کنار هم قرار می‌گیرند. توانایی ترکیب یک ابر خصوصی با منابع یک ابر عمومی میتواند برای تامین سطح سرویس مورد نیاز در زمان مواجه با نوسانات حجم کار استفاده برد. از یک ابر ترکیبی می‌توان برای مدیریت بهتر محاسبات ناگهانی[۱] که سیستم با آن مواجه میشود استفاده کرد. بدین شکل که محاسبات متعارف در ابر خصوصی انجام می‌پذیرد و در صورت بروز محاسبات ناگهانی خارج از توان سیستم، از ابر عمومی برای انجام محاسبات استفاده میشود.

۲-۳- سرویس‌های رایانش ابر

در رایانش ابر، می‌توان از لایه سخت افزار تا به لایه نرم افـزار بـه عنوان سرویس، بهره برد. از این رو سرویس‌های مختلف رایـانش ابر را در سه گروه زیر، دسته بندی میکنند:

۲-۳-۱- نرم‌افزار به عنوان سرویس[۲]

در این شکل سرویس، یک برنامه یا همان نرم افزار بصورت یـک سرویس برحسب تقاضا ارائه میشود. برنامه مورد نظر بر روی ابـر اجرا شده و به مشتریان وکاربران نهایی سرویس میدهد.

۲-۳-۲- سکو به عنوان سرویس[۳]

در این شکل سرویس، یک لایه از نرم‌افزار به عنوان سرویس ارائه میشود تا بتـوان از آن بـرای ایجـاد سـرویس‌هـای سـطح بـالاتر استفاده کرد. با ارائه این سرویس، تامین کننده ابـر یـک محـیط توسعه را به‌عنوان سرویس برای مشتری فراهم میکند. و از طرفی مشتری با استفاده از این سرویس، برنامه کاربردی که نیاز دارد را پیاده‌سازی می‌کند.

۲-۳-۳- زیرساخت به‌عنوان سرویس[۴]

این شکل سرویس، قابلیت‌های محاسباتی و ذخیره‌سازی اولیه را به عنوان سرویس‌های استاندارد در شبکه ارائه می‌دهد. سـرورها، سیستم‌های ذخیره‌سازی، سوئیچ‌ها، روترها از منابعی هستند کـه به عنوان سرویس بر حسب تقاضا در اختیار مشتری قرار بگیرند. در این ساختار از مجازی سـازی اسـتفاده میشـود تـا زیرسـاخت مورد نیاز مشتری شکل بگیرد[۱].

۲-۴- اهمیت فناوری رایانش ابر

بدون شک، مهم‌ترین فناوری که دنیای IT و به طور خاص دنیای شبکه را در چند سال آینده متحول خواهد کرد، فناوری رایـنش ابر است. شرکت اینتل در آخرین کنفرانس رایانش ابر در مـارس سال ۲۰۱۱ با بیان این‌که در سال ۲۰۱۵ نزدیک به ۱۵ میلیارد اتصال اینترنت شامل انواع دستگاه های همراه، خودرو، تلویزیـون، دستگاه‌های خاص منظوره و... در کنار صدها هگزابایت اطلاعـات روی اینترنت خواهیم داشت، پیش بینی کرده کـه تنهـا راه بـاقی مانده استفاده از فناوری ابر است. اینتـل در ایـن کنفـرانس سـه مشکل کنونی ابـر را امنیـت ایـن فنـاوری، چـالش‌هـای قـانونی دولت‌ها و استفاده مؤثر و بهینه از آن، می‌داند و می‌گوید که تمام تلاش‌ها در چند سال آینده باید معطوف به برطرف‌سازی این سه مشکل اصلی باشد. در این گزارش، اینتل سال ۲۰۱۵ را نقطه اوج فناوری‌های ابر پیش بینی کرده است. مؤسسه تحقیقـاتی IDC[۵] نیز طی یک گزارش در اکتبر امسال، عنوان کرده از سال ۲۰۱۱ تا سال ۲۰۱۵ سرویس هـای عمـومی و ذخیـره‌سـازی اطلاعـات مبتنی بر ابـر، ۲۳/۶ درصـد افـزایش خواهنـد یافـت. همچنـین سرویس‌های خصوصی ابر که توسـط شـرکت‌هـای بـزرگ مـورد استفاده قرار می‌گیرند، ۲۸/۹ درصد رشـدخواهنـد کـرد. در ایـن گزارش تراکنش مالی روی فناوری‌های ابر در سال ۲۰۱۵ نزدیک به ۲۲/۶ میلیارد دلار تخمین زده شده است[۲].

۲-۵- مزایا و معایب[۵]

مزایا:

- کاهش ریسک تهیه زیرساخت:
سازمان‌های IT می‌توانند از ابـر بـرای کـاهش ریسـک موجود در خرید سرورهای فیزیکی اسـتفاده کننـد. در شروع یک پـروژه شـاید تخمـین مناسـب و دقیقـی از موفقیت، بازده‌ای و سودآوری آن پروژه وجـود نداشـته باشد. اگر پروژه با موفقیت و استقبال پیش برود نیاز به افزایش زیر ساخت خواهـد بـود و در صـورت شکسـت پروژه هزینه انجام شده برای زیر ساخت نیز یک ضـرر اضافی را ایجاد میکند. با بهره گیری از رایانش ابر برای فراهم کردن زیرسـاخت مشـکلات اینچنینـی برطـرف میشود. از طرفی اگر ارائه یک سرویس بر روی یک ابـر خصوصی قرار داشته باشد، امکان محاسبات ناگهانی را که ناشی از افـزایش لحظـه‌ای تقاضـاها بـوده اسـت را میتوان به یک ابر عمومی انتقال داد. یعنی میتـوان بـا ریسک کمتری شرایط افـزایش محاسـبات را مـدیریت کرد.

- موانع کمتر برا ی ورود به بازار
با کمک رایانش ابر که می‌توان زیر سخت را کرایه کرد، هزینه سرمایه‌گـذاری زیرسـاخت و در نتیجـه سـرعت ورود به بازار افزایش می‌یابد.

- مقیاس پذیری راحت‌تر بر حسب تقاضا
در رایانش ابر متناسب با تقاضای کاربر، منابع پردازشی و ذخیره‌سازی اختصاص می‌یابد. در صورت تغییر نیـاز کاربر، می‌توان در کمترین زمان ممکن بـه طـور پویـا

منابع را بازتخصیص داد. این مسئله منجر به پرداخت هزینه متناسب با نیاز کاربر می‌شود.

- مدیریت بهتر حجم بالای داده

در رایانش ابر می‌توان حجم بالایی از فضای ذخیره-سازی را کرایه کرد. پردازش داده بر روی حجم عظیم داده تنها به کمک مجموعه‌ای از ابرها امکان‌پذیر خواهد بود. در این شرایط دسترس‌پذیری به داده نسبت به حالت سنتی بسیار آسان‌تر خواهد بود.

مشکلات ومعایب:

- دسترسی راه دور

یکی از نگرانی‌های مطرح در فناوری رایانش ابر، متفاوت بودن محل ذخیره داده با موقعیت کاربر است. در حالت سنتی که داده بر روی ماشین کاربر ذخیره می‌شد کنترل فیزیکی و منطقی بیشتری بر داده وجود داشت. ولی در رایانش ابر به خاطر ماهیت ذاتی آن، این کنترل و قابلیت اعتماد ناشی از آن از بین می‌رود وکاربر نظارت مستقیم خود بر داده را از دست می‌دهد.

- مشکلات ناشی از محدودیت پهنای باند

تمام سرویس‌های رایانش ابر مبتنی بر اینترنت است. در صورت پائین بودن پهنای باند ارتباطی بین کاربر و فراهم کننده ابر، ارائه سرویس‌های رایانش ابر با محدودیت‌های زیادی مواجه خواهد شد. بنابراین یکی از نیازمندی‌های اولیه استفاده از رایانش ابر دسترسی به پهنای باند مناسب است.

- محدودیتها ونگرانی‌های امنیتی

مهمترین مشکل رایانش ابر، فراهم کردن امنیت داده و سرویس قابل اعتماد برای کاربران آن است. به علت اهمیت این موضوع، مفصلا در ادامه به آن پرداخته خواهد شد.

۳- امنیت در رایانش ابر

سازمان‌ها برای به کار بردن سرویس‌های رایانش ابر، نیاز به تضمین کافی امنیت داده‌های خصوصی و سازمانی خود را دارند. پیش از این برای حفظ امنیت داده‌های خصوصی در شبکه از روش‌هایی همچون دیواره‌های آتش و VPN استفاده میشد. با ظهور رایانش ابر ،محل قرار گیری داده‌های خصوصی دیگر در اختیار کاربر نیست. علاوه بر آن، وضعیت برنامه‌های کاربردی که داده‌ها با آن‌ها سروکار دارند مشخص نیست. یعنی برای استفاده کنندگان از سرویسهای ابر مشخص نیست چه کدهایی بر روی داده‌های آنها اجرا می‌شود. برنامه‌هایی که بر روی ابر اجرا می-شوند از طرف یک شرکت ثالث پیاده سازی شده است و برای کاربر ممکن است، امکان هیچگونه کنترلی بر روی داده‌های شخصی خود مهیا نباشد. از طرفی ماهیت سرویس‌های ابر به شکلی است که از طریق مرورگرها ارائه میشوند. بنابراین تهدیداتی که همیشه در مورد مرورگرها وجود داشته است سرویسهای رایانش ابر را نیز در معرض خطر قرار میدهد.

۳-۱- مشکلات، ریسک‌ها و تهدیدات در رایانش ابر و مدیریت آنها

به خاطر ماهیت رایانش ابر نیاز است که ارزیابی ریسک در زمینه‌هایی مانند صحت ودرستی ، محرمانه‌گی و قابلیت بازیابی داده انجام شود. در اینجا هفت موضوع امنیتی که گارتنر[۶] بیان میکند مشتریان قبل از انتخاب سرویس‌دهنده ابر، باید این موضوعات را با آن مطرح کنند، ارائه می‌شود.[۴].

۳-۱-۱- مدیران و مسئولان سازمان ابر

پردازش داده‌های کاربر خارج از موقعیت فیزیکی آن، به طور ذاتی دارای ریسک است. چرا که در چنین شرایطی کنترل‌های فیزیکی، منطقی و شخصی که بر روی برنامه‌های درون-خانه‌ای[۷] امکان پذیر است، بر روی داده منبع خارجی[۸] از دست می‌رود. در حد امکان در مورد افرادی که برای مدیریت سیستم ابر دسترسی به داده‌های شما دارند اطلاعات کسب کنید. از ارائه دهنده سرویس بخواهید در مورد مدیران وکارمندانش که به داده‌های شما دسترسی دارند، اطلاعات ویژه‌ای را برایتان فراهم کند و سطح دسترسی آنها را مشخص کند.

۳-۱-۲- برآوردهای منظم[۹]

ارائه‌دهندگان قدیمی سرویس ابر در معرض نفوذ ممیزی‌های خارجی و گواهی‌های امنیتی هستند. اگر یک ارائه دهنده سرویس ابری به این ممیزی امنیتی وفادار نباشند، منجر به کاهش آشکار در اعتماد مشتریان می‌شود.

۳-۱-۳- موقعیت داده

زمانیکه یک کاربر از ابر استفاده میکند، به احتمال زیاد نمی‌داند داده‌هایش دقیقا در کجا ذخیره شده‌اند. در واقع شاید حتی نداند در کدام کشور قرار دارند. کاربر باید از فراهم کنند ابر بپرسد که آیا موقعیت قرارگیری داده‌ها و پردازش آنها از قوانین قضایی خاصی تبعیت میکند. آیا فراهم کننده ابر، قردادی الزام‌آور برای پذیرفتن امنیت شخصی داده‌ها می‌پذیرند.

۳-۱-۴- تفکیک داده[۱۰]

داده‌های کاربران مختلف در ابر، نوعا در کنار یکدیگر از یک محیط مشترک استفاده میکنند. استفاده از رمزگذاری موثر است ولی چاره تمام مشکلات نیست. گارتنر می‌گوید: باید بررسی کرد که در تفکیک[۱۱] داده چه اتفاقی می‌افتد. فراهم کننده ابر باید مدارکی ارائه کند که نشان دهد روش‌های رمزگذاری استفاده شده بر روی داده، توسط گروه مجرب و آگاهی، تست و ارزیابی شده است. مشکل در رمزگذاری ممکن است داده را غیر قابل استفاده کند و حتی رمزگذاری بدون اشکال ممکن است دسترس‌پذیری داده را سخت کند.

۳-۱-۵- بازیابی[۱۲]

حتی اگر کاربر نداند که داده‌هایش در کجا قرار دارد، باید ارائه کننده سرویس ابر به مشتری خود بیان کند در صورت بروز هرگونه حادثه مخرب در سیستم ابر چه اتفاقی بر روی داده‌های او می‌افتد. هر گونه راه‌حلی که ذخیره چند نسخه از داده‌ها را بر روی سایت‌های مختلف در خود نداشته باشد به هر حال ممکن است با مشکل مواجه شود. کاربر باید از فراهم کننده سرویس ابر خود بپرسد که آیا توانایی بازیابی تمام داده‌ها را دارد. این رویه چه مدت زمان به طول خواهد انجامید.

۳-۱-۶- پشتیبانی بازرسی[۱۳]

بازرسی بی‌رویه در ابر به هر حال امکان پذیر نیست. گارتنر هشدار میدهد" سرویس‌های ابر را مشخصا بسیار سخت می‌توان بازرسی کرد. چرا که گزارشات و داده‌های کاربران مختلف با هم ذخیره شده باشد و یا حتی ممکن است بر روی مراکز داده مختلفی پخش شده باشد. اگر کاربر نتواند قرارداد مشخصی را برای پشتیبانی از یک شکل بازرسی از ارائه‌کننده سرویس درخواست کند، آنگاه در صورت بروز مشکل، امکان بازرسی از فعالیت‌های تامین کننده ابر، برای بازیابی داده‌ها امکان پذیر نخواهد بود.

۳-۱-۷- پایداری[۱۴] طولانی مدت

حالت ایدال این است که فراهم کننده ابر هیچگاه متوقف نشود و یا توسط شرکت‌های بزرگتر حذف نشود. ولی کاربر جهت اطمینان از فراهم کننده باید بپرسد در صورت بروز چنین مشکلی چه بر سر داده‌هایش میفتد. آیا امکان بازپس گیری داده‌ها وجود دارد. بازپس گیری داده‌ها به چه فورمتی امکان‌پذیر خواهد بود.

۳-۲- اهمیت امنیت در رایانش ابر و چگونگی ارزیابی آن

امنیت با اولویت ترین نگرانی مشتریان ابر می‌باشد. یکی از موارد تاثیرگذار در خرید سرویس ابر، شهرت سرویس دهنده در زمینه‌های امنیت، محرمانگی داده‌ها و توانایی بازگشت از خرابی‌ها می‌باشد. از این رو یکی از موارد رقابتی در بازار سرویس‌های ابر موضوعات امنیتی است.

غالبا ارائه دهندگان سرویس‌های ابر ادعا می‌کنند، با توجه به موافقت سطح سرویس[۱۵] میتوان تضمین کافی برای رفع مشکلات بیان شده ارائه کرد. ولی مشکل از آنجا ناشی میشود که به خاطر ماهیت رایانش ابر که نمی‌توان نظارت مستقیم بر آن داشت، رعایت موافقت سطح سرویس دارای ضمانت کافی نیست.

برای رفع این مشکل و با توجه به ریسک‌هایی مطرح شده در بخش قبل، روشهای زیر توصیه شده است.

- استفاده از SLA دقیق و وجود ضمانت قابل قبول برای رعایت آن
- وجود تضمین مناسب برای تداوم فعالیت تجاری
- وجود تضمین مناسب برای بازیابی از سوانح[۱۶]
- بیان جزئیات مربوط به سیاست‌های امنیتی و پیاده‌سازی‌های انجام شده
- ارائه اطلاعات کامل زیرساختی در زمان انجام مذاکرات و نیز در هر لحظه از طول دوره ارائه سرویس
- نشان دادن رویه‌ها وسیاست‌ها توسعه نرم‌افزار، سیاست‌های تست امنیتی و سیاست‌های اعلام آسیب‌پذیری
- انجام عملیات تست توسط سرویس دهندگان ثالث
- انجام تست‌های نفوذ[۱۷] به صورت منظم و دوره‌ای و در دسترس بودن نتایج تست در صورت تقاضا

البته امکان اینکه یک سرویس‌دهنده بتواند در این سطح شفافیت عملکرد داشته باشد خود بحث دیگری است.

با توجه به تمام این موارد مشخص است که امنیت با اولویت ترین نگرانی مشتریان ابر خواهد بود. یکی از موارد تاثیرگذار در خرید سرویس ابر، شهرت سرویس دهنده در زمینه امنیت و محرمانگی داده‌ها، توانایی آن در بازگشت از خرابی می‌باشد. از این رو یکی از موارد رقابتی در بازار سرویس‌های ابر موضوعات امنیتی است. از طرفی برقراری امنیت هزینه بر است. ولی اگر معیارهای امنیتی در مقیاس بزرگ پیاده‌سازی شوند بسیار ارزان - تر تمام خواهند شد. بنابراین در رایانش ابر به خاطر گسترده بودن سیستم و کاربران، سرمایه گذاری برای امنیت سودمندتر از حالت شخصی خواهد بود. از این رو شرکت‌های بزرگتر و با پشتوانه قویتر بهتر می‌توانند، برای برقراری امنیت سرمایه‌گذاری کنند. . برخی از این زمینه‌های سرمایه‌گذاری در زیر توضیح داده شده است.

- چندین محل[۱۸]: اغلب ارئه کنندگان سرویس ابر امکان نسخه برداری از داده در چندین محل را فراهم میکنند. این افزونگی سیستمی باعث میشود در صورت خرابی در یک محل، امکان بازیابی داده از مابقی محل‌ها وجود داشته باشد.
- پاسخ‌های بجا به اتفاقات[۱۹]: سیستم‌های با مقیاس بزرگتر توانایی و قابلیت بالاتری را برای شناسایی و پاسخ به عوامل مشکل ساز مانند نرم‌افزارهای بدخواه را دارند.
- مدیریت تهدید[۲۰]: ارائه دهندگان سرویس ابر می‌توانند از عهده استخدام متخصصین برای مقابله با تهدیدهای امنیتی برآیند ولی شرکت‌های کوچک توانایی این کار را براحتی نخواهند داشت.

شرکت‌ها برای آنکه هزینه بالای فراهم کردن ساختار امنیت را پرداخت نکنند به رایانش ابر رو می‌آورند. ولی تمام ارائه کنندگان رایانش ابر نیز قادر به تامین امنیت کامل نیستند. از آنجا که یک محیط ابر از اجزاء مختلفی تشکیل شده است، امنیت رایانش ابر از امنیت ضعیف‌ترین جزء آن بیشتر نخواهد بود. یک حمله‌کننده سایبری با تشخیص نقاط ضعف امنیتی یک ابر میتواند آنرا مورد حمله قرار دهد. ذات معماری ابر قابلیت حملات همزمان را افزایش میدهد. در نتیجه بدون امنیت کافی در ابر، با یک حمله سایبری، یک مجموعه از عملکردهای مختلف ابر میتواند با مشکل مواجه شود[۳].

۳-۳- امنیت داده

برای انتقال از محیط سنتی پردازش به محیط ابر، برای یک مشتری ابر دو موضوع قابل توجه خواهد بود. اول اینکه محل ذخیره داده مکانی غیر از مکان ماشین کاربر خواهد بود. دوم اینکه محل ذخیره داده از یک محیط تک کاربره به یک محیط چند کاربره تغییر پیدا می‌کند. این تغییرات باعث بوجود آمدن نگرانی مهم از بابت افشای داده می‌شود. از دید سازمانها، مشکل افشای داده یک ریسک بزرگ می‌باشد [۸].

از این کنترل دسترسی به داده یک از مواردی است که بیشترین تحقیقات را در بحث امنیت داده ابر به خود اختصاص داده است. در صورت نبود امنیت کافی داده، این امکان وجود دارد که یک نفوذگر، به داده‌های ذخیره شده در ابر دسترسی پیدا کند. بنابراین یکی از موارد مهم و قابل توجه در عرصه رقابت ارائه دهندگان ابر ارائه سیاست‌های مطمئن و کارآمد برای برقرای امنیت داده میباشد[۵]. در این بخش توضیحی در مورد مدل امنیت داده و روشهای مطرح در این بخش ارائه خواهد شد.

۳-۳-۱- اصول امنیت داده

تمام روش‌های امنیت داده بر اساس سه اصل محرمانه‌گی، درستی و دسترس پذیریی داده بنا نهاده شده‌اند. محرمانه‌گی داده به مخفی ماندن اصل داده و اطلاعات به خصوص در محیط - های نظامی و حساس گفته میشود. البته محرمانگی داده فقط اشاره به اجتناب از دسترسی‌های غیر مجاز دیگر کاربران و یا یک مهاجم به داده‌های یک کاربر ندارد. بلکه موضوع قابل توجه دیگر در این بخش محرمانگی داده کاربر از سازمان ارائه دهنده ابر هم می‌باشد. در موضوع صحت و درستی[۲۱] داده، نیاز به تضمینی برای صحت داده در موارد حذف و اصلاح‌های غیرمجاز می‌باشد. دسترس‌پذیری داده به معنی این است که کاربر در هر جا که از سرویس‌های ابر بهره می‌برد بتواند به داده‌هایش نیز دسترسی داشته باشد.

۳-۳-۲- روشهای مطرح برای برقراری امنیت داده

- ذخیره‌سازی داده بصورت رمز شده باشد. تا در صورت دسترسی غیر مجاز به داده‌ها،امکان تفسیر آنها وجود نداشته باشد.

- انتقال داده‌ها نیز به صورت رمز شده باشد. این رویه به‌خصوص در ابرهای عمومی که از یک زیرساخت عمومی برای انتقال داده‌ها استفاده میکنند قابل توجه است.
- احراز هویت قوی بین اجزای برنامه بطوریکه داده تنها برای بخش‌های شناخته شده ارسال شود.
- استفاده از الگوریتم‌های رمز نگاری قویتر
- مدیریت دسترسی اشخاص به برنامه ها و نحوه این دسترسی
- استفاده از احراز هویت قوی ومبتنی بر شناسه
- بررسی امن بودن سرور احراز هویت برای ورود کاربران

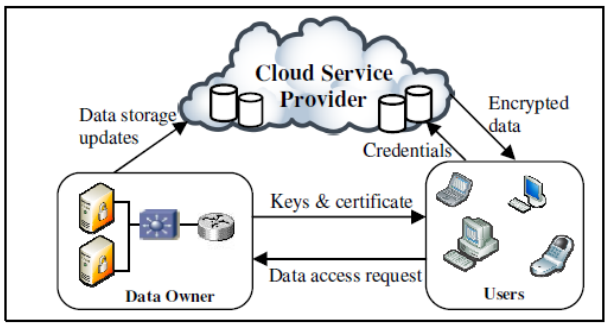
- البته از آنجا که خود رویه رمزنگاری داده ممکن است منجر به بروز مشکلاتی شود، تامین کننده ابر باید شرایط زیر را در آن برآورده کرده باشد[۳].
- استفاده از رمزنگاری آزمایش شده مطمئن برای داده‌ها در فضاهای ذخیره‌سازی مشترک
- ایجاد ذخیره پشتیبان[۲۲] داده، زمانبندی شده[۲۳]

۳-۴- کارهای انجام شده در امنیت داده

در این بخش به دو طرح ارائه شده در امنیت پرداخته میشود. طرح اول در مورد دسترس‌پذیری داده رمز شده است و طرح دوم مربوط به چگونگی تعیین سطح کنترل دسترسی کاربران به داده‌های ابر می‌باشد.

۳-۴-۱- سناریوی دسترس‌پذیری امن به داده

همانطور که اشاره شد برای برقراری امنیت داده آنرا به صورت رمز شده بر روی ابر قرار میدهند. در [۵] سناریویی از دسترسی به داده ارائه شده است و روش رمز نگاری داده در آن تحلیل شده است. در این طرح، سیستم ابر از سه جزء مالک داده[۲۴] ، فراهم کننده سرویس ابر[۲۵] و کاربر تشکیل شده است. مالک داده، داده مورد نیاز کاربر را بر روی ابر قرار میدهد. اگر فراهم کننده سرویس ابر از نظر مالک داده قابل اعتماد نباشد داده‌ی رمز شده بر روی ابر قرار خواهد گرفت. با دریافت درخواست داده از طرف کاربر، مالک داده کلیدها[۲۶]و گواهی[۲۷] مورد نیاز را برای کاربر ارسال میکند. کاربر با ارائه گواهی به فراهم کننده سرویس ابر و پس از اعتبارسنجی موفق از طرف سرویس دهنده ابر داده‌ی رمز شده را دریافت میکند(شکل ۱).



شکل (۱) مدل دسترسی امن داده [۵]

مزیت این سناریو در آن است که حتی سازمان سرویس دهنده ابر به اصل داده‌ها دسترسی ندارند ولی مشکل آن از آنجا ناشی می‌شود که نیاز است مالک داده، در زمان درخواست کاربر برخط[۲۸] باشد.

در ادامه مقاله، نویسنده طرحی را ارائه کرده است که این مشکل را بهبود داده است که در اینجا از ذکر آن اجتناب می‌شود.

۳-۴-۲- مدل ریاضی توزیع داده برای فراهم کردن امنیت داده

در [۶] روشی برای مدل کردن داده ذخیره شده در ابر ارائه شده‌است. داده در ابر به صورت رمز شده و توزیع شده ذخیره می‌شود. البته می‌توان این طرح را یک مدل کلی برای قرارگیری داده در ابر شناخت. در این مدل فایل f به بخش‌های کوچکتری تقسیم میشود. هر بخش به صورت رمز شده بر روی یک سرور ذخیره‌سازی داده قرار می‌گیرد. اطلاعات توزیع فایل در ماتریس Df نگهداری می‌شود. مدیریت توزیع فایل و ایجاد ماتری Df با سرور NameNode میباشد. این سرور ارزیابی هویت کاربر و کنترل سطح دسترسی آنرا نیز انجام می‌دهد. روابط زیر نشان دهنده ارتباط بین این اجزاء است.

$$D_f = C(Namenode) \quad (1)$$

$$K_f = f * D_f \quad (2)$$

- C(x): ویزیت گره x می‌باشد.
- D_f: ماتریس توزیع فایل
- K_f: وضعیت توزیع داده در سرورهای ذخیره داده
- f: نماد یک فایل است، هر فایل با تقسیم بندی زیر مشخص میشود.

$$f = \{F(1), F(2), \dots, F(n)\} \quad (3)$$

رابطه بالا بیانگر تقسیم بندی فایل f به n بلاک است که بین هر بلاک شرایط زیر برقرار است.

$$F(i) \cap F(j) = \emptyset, \quad i \neq j; \qquad i,j \in 1,2,3,\dots n \quad (4)$$

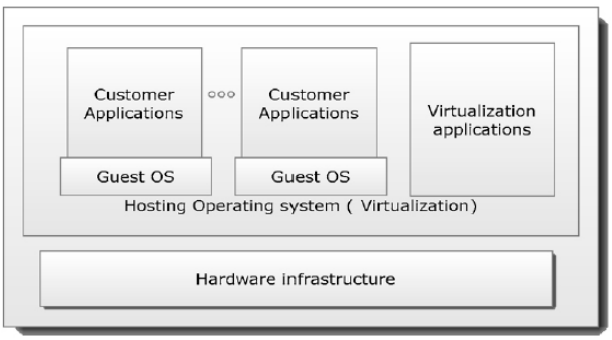
D_f یک ماتریس صفر و یک، با ابعاد L x L میباشد. که پارامتر L تعداد سرورهای ذخیره داده را مشخص می‌کند.

۳-۵- امنیت مجازی‌سازی

مجازی سازی از فناوری‌های مهم و پایه‌ای در ساختار رایانش ابر میباشد. مجازی سازی این امکان را فراهم میسازد تا بتوان نرم افزار را مستقل از سخت افزارهای مختلف موجود در ابر اجرا کرد. همچنین با ایجاد انعطاف پذیری بالا، میتوان مطابق نیاز کاربر منابعی همچون میزان پردازنده و یا حافظه را به شکل پویا برای محاسبات اختصاص داد. [۱].

۳-۵-۱- اجزاء مجازی سازی[۲۹]

مجازی‌سازی به سازمانهای فناوری اطلاعات کمک میکند تا بازده‌ای برنامه‌های کاربردیشان را با توجه به هزینه‌ها بهینه کنند. اینکار با ایجاد سرورهای مجازی صورت می‌گیرد. عبارت ماشین مجازی[۳۰] به یک نرم‌افزار کامپیوتری اشاره دارد که بتواند به مانند یک سخت‌افزار کامپیوتر، یک سیستم عامل و برنامه‌های کاربردی مورد نیاز را اجراء کند. به سیستم‌عاملی که بر روی ماشین مجازی اجراء می‌شود سیستم عامل مهمان[۳۱] گفته میشود. به لایه‌ای که وظیفه ایجاد، مدیریت و کنترل زیربخشهای مجازی ماشین مجازی را دارد، پایش ماشین مجازی[۳۲] یا مدیر ماشین مجازی[۳۳] گفته می‌شود[۱۰]. (شکل ۲ را مشاهده کنید)



شکل (۲) مجازی سازی مبتنی بر سیستم عامل[۱۰]

۳-۵-۲- روشهای مجازی سازی [۱۰]

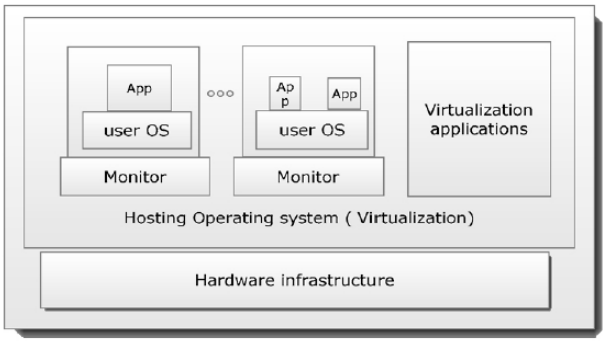
در یک محیط سنتی، سرورهای فیزیکی با یک سوئیچ[۳۴] فیزیکی به هم متصل می‌شدند. در چنین ساختاری سازمان‌های فناوری اطلاعات(IT) ، اطلاعات مدیریتی ترافیک داده بین سرورها را از این سوئیچ‌های فیزیکی دریافت می‌کردند. متاسفانه این سطح اطلاعات مدیریتی توسط یک سوئیچ مجازی فراهم نمی‌شود. در نتیجه در یک محیط مجازی، پایش[۳۵] ترافیک داده بین ماشین‌های مجازی از بین می‌رود. چندین رویکرد معمول مجازی‌سازی که در کنترل ماشین‌های مجازی با هم تفاوت دارند وجود دارد.

۱. مجازی‌سازی مبتنی بر سیستم[۳۶]

در این رویکرد(شکل ۲)، مجازی‌سازی با یک سیستم‌عامل میزبان[۳۷] فعال می‌شود که می‌تواند چندین سیستم‌عامل مجازی مهمان را بر روی یک سخت‌افزار مشترک پشتیبانی کند. در این حالت، تمام سیستم‌عامل‌های مهمان بر روی یک هسته سیستم‌عامل مشترک که دسترسی به سخت افزار را مقدور می‌سازد اجراء میشوند. در نتیجه سیستم عامل میزبان به ماشین‌های مجازی دید وکنترل دارد. این رویکرد ساده ولی آسیب‌پذیر است. برای مثال یک مهاجم با حمله به هسته سیستم‌عامل میزبان میتواند اجرای تمام سیستم‌عامل‌های مهمان را با مشکل مواجه سازد و حتی کنترل ماشین‌های مجازی آنها را به عهده بگیرد.

۲. مجازی‌سازی مبتنی بر برنامه

در این رویکرد نیز، مجازی‌سازی بر روی سیستم‌عامل میزبان می‌نشیند. در این روش هر ماشین مجازی سیستم‌عامل مهمان خود را با برنامه کاربردی آن اجرا میکند. این روش در محیط‌های تجاری خیلی معمول نیست. شکل ۳ این ساختار را نشان میدهد.

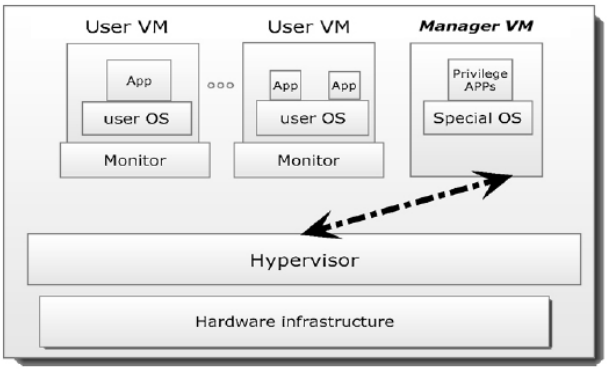


شکل (۳) مجازی سازی مبتنی بر برنامه[۱۰]

۳. مجازی‌سازی مبتنی بر هایپرویزور[۳۸]

یک هایپرویزور بخشی از زیرساخت سخت‌افزار یا هسته سیستم‌عامل میزبان می‌باشد(شکل ۴). یکی از روش‌های مطرح مجازی‌سازی استفاده از هایپرویزور است که اجازه میدهد چندین سیستم عامل(مهمان) به طور همزمان بر روی یک کامپیوتر اجرا شوند. گاهی به این روش، مجازی‌سازی سخت‌افزار نیز گفته می‌

شود. این قابلیت به سیستم عامل‌های مهمان یک سکوی مجازی برای اجراء ارائه میدهد. هایپرویزور منابع به اشتراک گذاشته شده بین ماشین‌های مجازی را کنترل کرده و آنها را پایش میکند. تعدادی ازاین ماشین‌های مجازی دارای اولویت هستند تا مدیریت سکو مجازی‌سازی را انجام دهند. در این معماری، بخشهای اولویت دار به دیگر ماشین‌های مجازی دید وکنترل دارند. این رویکرد محیط با قابلیت کنترل بالاتری ایجاد می‌کند و می‌تواند ابزارهای امنیتی مانند سیستم تشخیص نفوذ، را اجراء کند. اما این رویکرد هم آسیب‌پذیر است از آن جهت که هایپرویزور نقطه ضعف[۳۹] آن به حساب می‌آید. اگر هایپرویزور دچار مشکل شود و یا مهاجم کنترل آن را به دست بگیرد، آنگاه تمام ماشین‌های مجازی در کنترل مهاجم خواهند بود. ولی به هر حال به دست گرفتن کنترل هایپرویزور از ماشین مجازی بسیار مشکل‌تر است.



شکل (۴) مجازی‌سازی مبتنی بر های‌پرویزور [۱۰]

۳-۵-۳- امنیت ماشین مجازی و تهدیدات آن

مجازی سازی خود به عنوان یک فناوری مطرح پیش از رایانش ابر، دارای تهدیدات امنیتی شناخته شده‌ای بود که این مشکلات را به محیط ابر نیز وارد کرده است.

در هایپرویزو، تمام کاربران، سیستم خود را یک کامپیوتر مجزا از بقیه کاربران مشاهده میکند حتی اگر تمام کاربران با یک ماشین سرویس داده شوند. در این ساختار، یک ماشین مجازی یک سیستم‌عامل است که با یک برنامه کنترلی از لایه زیر مدیریت می‌شود. در این سطح، تهدیدات و حملات متنوعی وجود دارد که در ادامه ذکر می‌شود.

- حملات در سطح ماشین مجازی[۴۰]
- آسیب‌پذیری فراهم کننده ابر
- گسترش سطح حمله در زیر ساخت شبکه
- شناسایی و کسب اجازه[۴۱] برای کاربر ابر
- کنترل داده در ابر
- ارتباطات در سطح مجازی‌سازی

۳-۵-۴- امنیت هایپرویزور

در یک محیط مجازی ماشین‌های مجازی که دارای پوشش امنیتی مشخصی هستند نمی‌توانند با ماشین‌های مجازی دیگری که پوشش امنیتی متفاوتی دارند دسترسی پیدا کنند. در محیط مجازی یک هایپرویزور دارای محدوده امنیتی منحصر به خود می‌باشد. هایپرویزورها می‌توانند بر تمام فعالیتهای ماشین‌های مجازی اجراء شده بر روی میزبان مجازی سازی نظارت و کنترل داشته باشند[۷]. از این رو اگر یک مهاجم بتواند کنترل یک هایپرویزور را به دست بگیرد، می‌تواند به تمام محیط نظارت هایپرویزور کنترل پیدا کند.

با توجه به اشارات قبلی، هایپرویزورها ابزارهای مدیریتی هستند که با محیط امنیتی مشخص سعی در ایجاد منطقه امن در مجازی سازی دارند. در مشخصه‌های امنیتی، سه سطح اصلی در مدیریت امنیت هایپرویزور وجود دارد که در ادامه ذکر می‌شود.

- شناسایی[۴۲]: باید کاربر با یک سری مکانیزم‌های مطمئن، استاندار و مناسب شناسایی شود.
- کسب اجازه[۴۳] : سطح دسترسی کاربران و عملیاتی که می‌خواهند انجام دهند باید مشخص شود
- شبکه: باید ارتباط بین شبکه کاربرکه دارای سطح امنیتی متفاوت از شبکه هایپرویزور است با مکانیزم-هایی مطمئن شود.

با توجه به بحث مطرح شده برقراری امنیت در مجازی سازی به هیچ وجه نباید فقط مبتنی بر امنیت بالای شبکه ارتباطی آن باشد. چرا که ارتباط داخلی ماشین‌های مجازی و هایپرویزورها خود میتواند منجر به مشکلات متعددی شود.

۳-۶- امنیت شبکه و سرویس‌های ابر

ماهیت رایانش ابر مبتنی بر ارائه سرویس بر روی زیرساخت اینترنت (شبکه) میباشد. این ویژگی خود یک مسئله چالش برانگیز است. چرا که امنیت در این بخش از جنبه‌های مختلف طراحی سیستم، پیاده سازی، استقرار آن و ارائه سرویس‌ها قابل بررسی است.در سرویس‌های رایانش ابر به مانند ما بقی سرویس-

های ارائه شده بر روی اینترنت احتمال بروز حملاتی همچون Sniffing، Spoofing، Man in the middle و مانندآنها می‌تواند زیاد باشد. در جدول زیر برخی از تهدیدات مطرح در سرویس‌های رایانش ابر به اختصار آمده است.

جدول ۱- برخی از تهدیدات شناخته شده در خصوص استفاده از رایانش ابر

تهدیدات	شرح
قطع شدن سرویس	قطع شدن اتصال با اینترنت یا ارتباط با سرویس دهنده و سایر اختلالات مشابه
حملات DOS	مختل کردن یا از کار انداختن سرویسها
XSS[۴۴]	قرار دادن کدها و اسکریپتهای مخرب در داخل صفحات وب
CSRF[۴۵]	ترکیبی از حمله XSS و استفاده از URLهای تغییر شکل داده شده
سرویس‌های نامعتبر	ارائه سرویس توسط سرویس دهندگان نامعتبر
DNS poisoning exploit	ارجاع کاربر به یک سایت تقلبی با ظاهری مشابه سایت اصلی

با انجام اقدامات امنیتی لازم در این بخش این ریسک‌ها را می‌توان کاهش داد. به عنوان نمونه در پیاده سازی، توسعه دهندگان باید مراقب کدهایی که منجر به آسیب‌رسانی ابر در مقابل تکنیک‌هایی نظیر Buffer overflow و یا SQL injection میشود باشند. سیستم‌عامل و نرم افزارهای ارائه شده بر روی ابر باید با آخرین وصله‌های امنیتی به روز شده باشند تا امن باشند. بعضی از رویکردهای امنیت شبکه به شرح زیر است:

- استفاده از دامنه‌های امن برای گروه‌بندی ماشین‌های مجازی با همدیگر و سپس کنترل دسترسی با دامنه از طریق قابلیت‌های port filtering سرویس دهنده ابری.
- کنترل ترافیک با استفاده از فیلترینگ مبتنی بر پورت، یا بکارگیری فیلترینگ بسته یا دیواره آتش در مکان‌های مناسب امکان‌پذیر است.

ولی ممکن است تامین کننده ابر تمامی این اقدامات را به طور کامل و درست انجام ندهد. برای برقراری امنیت در زیر ساخت شبکه، استانداردها و ضوابط مشخصی از پیش تعریف و مشخص شده است. این بخش از امنیت فناوری رایانش ابر به طور ویژه در محدوده محققان فناوری اطلاعات و شبکه مطرح میشود. از این رو بیش از این در این نگارش به این حوزه پرداخته نشده است.

۴- نتیجه

در این گزارش، امنیت ابر در مباحثی چون ریسک‌ها و مشکلات مطرح در آن، امنیت داده، امنیت مجازی‌سازی و امنیت سرویس بررسی شد. با بررسی دقیق تمام مباحث مطرح شده، به دو موضوع اصلی امنیت داده و امنیت سرویس در ابر می‌رسیم. بحث ریسک‌ها و مشکلات امنیت ابر تماما به نگرانی‌های این دو حوزه می‌پردازد. امنیت در مجازی‌سازی تضمین کننده امنیت داده و امنیت سرویس خواهد بود. همینطور امنیت زیرساخت شبکه ابر هم منتهی به امنیت سرویس میشود. با توجه به این نگاه، مسائل اصلی در امنیت رایانش ابر نهایتا به امنیت داده و امنیت سرویس میرسد. مباحث مربوط به امنیت سرویس تماما در حوزه فناوری اطلاعات می‌باشد ولی امنیت داده از گسترده‌گی بالایی برخوردار است و معماری‌های مختلفی برای ایجاد محرمانگی[۴۶]، پوشیدگی[۴۷] و صحت داده ارائه شده است. از طرفی مهمترین نگرانی امنیت رایانش ابر در حوزه امنیت داده دیده میشود و به آن پرداخته میشود.

مراجع

[۱] محمد کاظم اکبری، مرتضی سرگلزایی جوان **"محاسبات ابری"** انتشارات دانشگاه صنعتی امیرکبیر، بهار ۱۳۸۹.

[۲] مجله شبکه، **"ده فناوري تأثيرگذار بر شبکه‌هاي کامپيوتري"**، ترجمهٔ صدیقی مشکنانی، ویرایش سوم، اصفهان، نشر شیخ بهایی، بهار ۱۳۹۱.

[3] J.Harauz, Lori M. Kaufman, B.Potter, *"Data Security in the World of Cloud Computing"* Published by the IEEE Computer and Reliability Societies, JULY/AUGUST 2009

[4] Gartner, *"Seven cloud-computing security risks,"* Network World , July 2008

[5] S.Sanka, C.Hota, M.Rajarajan, *"Secure Data Access in Cloud Computing"*. 2010 IEEE

[6] D.Yuefa, W.B.G.Yaqiang, Z.Quan, T.Chaojing "*Data Security Model for Cloud Computing*" International Workshop on Information Security and Application, 2009 ACADEMY PUBLISHER.

[7] Texiwill. (2009). "*Is Network Security the Major Component of Virtualization Security?"* Available: http://www.virtualizationpractice.com/blog/?p=350

[8] C. Almond, "*A Practical Guide to Cloud Computing Security*," ,27 August 2009

[9] K.Mukherjee - G.Sahoo "*A Secure Cloud Computing*", *International Conference on Recent Trends in Information*, Telecommunication and Computing 2010
[10] Farzad Sabahi, "*Virtualization-Level Security in Cloud Computing*", 2011 IEEE.
[11] D.Borthakur, "Apache Hadoop FileSystem and its Usage in Facebook", Presented at UC Berkeley, April 2011

زیرنویس‌ها

---

۱ Surge Computing
۲ Software as a Service
۳ Platform as a Service
٤ Infrastructure as a Service
٥ International Data Corporation
٦ Gartner
۷ In-house
۸ Outsourced
۹ Regulatory compliance
۱۰ Data segregation
۱۱ Segregate
۱۲ Recovery
۱۳ Investigative Support
۱٤ Viability
۱٥ Service Level Agreement
۱٦ Disaster Recovery
۱۷ Intrusion
۱۸ Replicate
۱۹ Improve timeliness of response to incidents
۲۰ Threat Management
۲۱ Integrity
۲۲ Backup
۲۳ Scheduled
۲٤ Data Owner
۲٥ Cloud Service Provider
۲٦ Key
۲۷ Certificate
۲۸ Online
۲۹ Virtualization Components
۳۰ Virtual Machine
۳۱ Guest OS
۳۲ VM Monitor
۳۳ VM Manager
۳٤ Switch
۳٥ Monitor
۳٦ Operating system-based virtualization
۳۷ Host
۳۸ Hypervisor-based virtualization
۳۹ Single point of failure
٤۰ VM level attacks
٤۱ Authorization
٤۲ Authentication
٤۳ Authorization
٤٤ Cross-Side-Scripting
٤٥ Cross-site request forgery
٤٦ confidentiality
٤۷ Privacy

بررسی آسیب‌پذیری‌ها و مخاطرات لایه‌ی MAC در شبکه‌های Wi-Fi و WiMax و راه‌های مقابله

داور احمدپور[1]، پیمان کبیری[2]

[1] دانشجوی کارشناسی ارشد، دانشکده‌ی مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
تهران، ایران
davar_ahmadpour@comp.iust.ac.ir
[2] استاد راهنما، دانشکده‌ی مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
تهران، ایران
peyman.kabiri@iust.ac.ir

چکیده

گسترش روزافزون شبکه‌های Wi-Fi و WiMax سبب پیدایش طیف نسبتا وسیعی از مخاطرات با هدف به چالش کشاندن امنیت این شبکه‌ها شده است. بسیاری از این مخاطرات نشأت گرفته از آسیب‌پذیری‌های موجود در فریم‌های مدیریتی و کنترلی لایه‌ی MAC هستند که هیچ‌گونه عملیات رمزنگاری و احراز هویت برای آن‌ها صورت نمی‌گیرد. هدف از این نوشتار، معرفی آسیب‌پذیری‌های موجود در لایه‌ی MAC شبکه‌های Wi-Fi و WiMax با تمرکز بر فریم‌های مدیریتی و کنترلی، و همچنین بررسی راه‌کارهای ارائه شده برای تشخیص و پیش‌گیری نوع خاصی از حملات Wi-Fi با عنوان جعل آدرس MAC است.

کلمات کلیدی

استاندارد IEEE 802.11 و IEEE 802.16، MAC، حمله، جعل MAC، تشخیص و پیش‌گیری از حمله.

۱- مقدمه

سهولت استفاده از شبکه‌های بی‌سیم محلی مبتنی بر استاندارد IEEE 802.11 (Wi-Fi) سبب شده تا استفاده از این شبکه‌ها به طور چشمگیری در خانه‌ها، ادارات و اماکن عمومی گسترش یابد. این رشد فزاینده به گونه‌ای بوده است که شرکت تحقیقاتی Gartner پیش‌بینی می‌کند که در سال ۲۰۱۵ تعداد دستگاه‌های قابل اتصال به Wi-Fi به بیش از سه میلیارد خواهد رسید (این در حالی است که در سال ۲۰۱۰ کمتر از یک میلیارد دستگاه با قابلیت اتصال Wi-Fi وجود داشتند)[1]. علاوه بر این، استفاده از این شبکه‌ها در محیط‌های حساس مانند بیمارستان‌ها و کاربردهای نظامی که دسترس‌پذیری و قابلیت اتکا بر شبکه نقشی کلیدی ایفا می‌کنند نیز در حال افزایش است [2].

استاندارد IEEE 802.16 نیز که با نام WiMax هم شناخته می‌شود، به دلیل پهنای باند مناسب، تأخیر اندک و پشتیبانی از کیفیت سرویس، یک جایگزین مناسب برای شبکه‌های سیمی و بخشی از شبکه‌های نسل آینده[1] (NGN) جهت ارتباطات بی‌سیم درون شهری[2] (WMAN) است.

علی‌رغم تلاش صورت گرفته برای حفظ محرمانگی، جامعیت داده‌ها و احراز هویت، *دسترس‌پذیری* شبکه‌های Wi-Fi و WiMax مسئله‌ای است که کماکان با چالش‌های جدی روبرو است. منظور از دسترس‌پذیری امکان استفاده‌ی عادلانه‌ی کاربران از شبکه است. دسترس‌پذیری مسئله‌ای است که نه تنها با مسدودسازی[3] کانال فرکانسی شبکه مخدوش می‌شود، بلکه با تغییر پارامترهای لایه‌ی MAC[4] و ارسال فریم‌های کنترلی یا مدیریتی مخدوش نیز تحت شعاع قرار می‌گیرد.

علاوه بر این، جعل آدرس MAC در فریم‌های مدیریتی دیگر چالش جدی شبکه‌های بی‌سیم مبتنی بر 802.11 است که علی‌رغم تلاش‌های صورت گرفته برای تدوین استاندارد 802.11w جهت تأمین امنیت فریم‌های مدیریتی، جعل آدرس MAC کماکان مرتفع نشده است [33].

مباحثی که در ادامه مورد بررسی قرار خواهند گرفت بدین شرح خواهند بود: در بخش ۲ شرحی از استانداردهای IEEE 802.11 و IEEE 802.16 ارائه خواهد شد. در بخش ۳ و ۴ مخاطرات و کاستی‌های امنیتی موجود در لایه‌ی MAC این دو استاندارد مورد بررسی قرار خواهند گرفت. در کنار آن، سایر آسیب‌پذیری‌های موجود نیز به طور اجمالی معرفی خواهند شد. در بخش ۵ راه‌کارهای ارائه شده برای مقابله با نوع خاصی از حملات Wi-Fi تحت عنوان جعل آدرس MAC مورد بررسی قرار خواهند گرفت و بخش ۶ شامل نتیجه‌گیری خواهد بود.

۲- IEEE 802.11 و IEEE 802.16

این بخش به معرفی کلی استانداردهای 802.11 و 802.16 اختصاص دارد.

۲-۱- IEEE 802.11

در این بخش در رابطه ساختار کلی IEEE 802.11 صحبت خواهد شد.

۲-۱-۱- معماری و سرویس‌ها

استاندارد IEEE 802.11 مجموعه‌ای از استانداردهای لایه‌ی فیزیکی و لایه‌ی MAC است که برای ارتباطات شبکه‌های بی‌سیم محلی در فرکانس‌های ۲.۴ و ۵ GHz و در دو حالت زیرساخت (BSS) و اقتضایی (IBSS) طراحی شده است. لایه‌ی فیزیکی در این استاندارد می‌تواند از هر یک از فنآوری‌های FHSS، DSSS، OFDM و Infrared استفاده کند و در برگیرنده‌ی دو زیر لایه‌ی PLCP[۵] و PMD[۶] است که وظیفه‌ی تبدیل فریم‌های داده‌ی لایه‌ی MAC (MPDU) به یک فرمت مناسب جهت انتقال روی رسانه‌ی بی‌سیم و تعیین ویژگی‌های این انتقال را بر عهده دارند. زیر لایه‌ی MAC نیز که در نیمه‌ی تحتانی لایه‌ی پیوند داده‌ها قرار گرفته است وظیفه‌ی مدیریت دسترسی ایستگاه‌ها به کانال انتقال، تضمین دریافت فریم‌های داده توسط گیرنده و امنیت فریم‌های داده را بر عهده دارد.

۲-۱-۲- لایه‌ی MAC

همانطور که در بخش ۲-۱-۱ نیز گفته شد، لایه‌ی MAC در 802.11 وظیفه‌ی تضمین دریافت داده‌ها، کنترل دسترسی رسانه وامنیت را بر عهده دارد. در این بخش هر یک از این وظایف مورد بررسی قرار خواهند گرفت.

تضمین دریافت داده‌ها

مانند تمامی شبکه‌های بی‌سیم، شبکه‌های Wi-Fi نیز تحت تأثیر عواملی نظیر نویز، سایه‌افکنی، محوشدگی و تصادم قرار می‌گیرند که این عوامل منجر به از دست رفتن بخشی از فریم‌های در حال انتقال می‌گردند. برای این منظور، IEEE 802.11 شامل یک مکانیزم تبادل فریم دو مرحله‌ای است، بدین مفهوم که هر زمان که یک ایستگاه یک فریم داده دریافت کند، در پاسخ یک فریم کنترلی ACK برای فرستنده ارسال می‌کند. اگر فرستنده‌ی فریم داده طی زمان مشخصی فریم ACK را دریافت نکند، فریم داده را مجددا ارسال می‌کند.

علاوه بر این، برای اطمینان بیشتر از عدم بروز تصادم، یک مکانیزم چهار مرحله‌ای برای ارسال فریم‌ها نیز در نظر گرفته شده است. در این روش، ایستگاه فرستنده‌ی فریم داده ابتدا یک فریم کنترلی RTS برای گیرنده ارسال می‌کند. گیرنده با دریافت فریم RTS، یک فریم CTS برای فرستنده ارسال می‌کند تا آمادگی خود برای دریافت فریم داده را اعلام نماید. تمامی ایستگاه‌هایی که یکی از فریم‌های RTS و CTS را دریافت می‌کنند، به اندازه‌ی مدت زمانی که در این فریم‌ها مشخص شده (موسوم به NAV) سکوت می‌کنند تا از بروز تصادم جلوگیری شود. در نهایت، فرستنده داده‌ی خود را ارسال و در پاسخ یک فریم ACK دریافت می‌کند.

کنترل دسترسی رسانه

هدف از کنترل دسترسی رسانه، ارائه‌ی روشی برای به اشتراک گذاشتن کانال انتقال میان ایستگاه‌های شبکه است. برای این منظور، گروه کاری 802.11 دو روش برای پیاده‌سازی در لایه‌ی MAC ارائه کرده است: DCF[۷] که مانند Ethernet، تصمیم برای انتقال داده از طریق یک مکانیزم شنود حامل میان گره‌های شبکه توزیع می‌گردد؛ و PCF[۸] که نقطه‌ی دسترسی شبکه (AP) تعیین می‌کند که کدام گره اجازه‌ی انتقال دارد.

در روش DCF، از الگوریتم CSMA/CA[۹] استفاده شده است. روش کار این الگوریتم بدین شرح است:

۱. ایستگاهی که قصد ارسال فریم دارد، ابتدا کانال انتقال را شنود می‌کند. اگر کانال آزاد باشد، آن‌گاه ایستگاه به اندازه‌ی یک زمان از پیش تعیین شده به نام DIFS منتظر می‌ماند. اگر در طی این زمان نیز کماکان کانال آزاد باشد، آن‌گاه ایستگاه فریم خود را ارسال می‌کند.
۲. اگر کانال مشغول باشد (در لحظه‌ای که ایستگاه شروع به شنود کانال می‌کند یا در حین سپری شده زمان DIFS)، ایستگاه عقب کشیده و تا زمان آزاد شدن فعلی کانال منتظر می‌ماند.
۳. زمانی که کانال آزاد شود، ایستگاه کانال را دوباره شنود کرده و به اندازه‌ی یک زمان DIFS منتظر می‌ماند. اگر پس از سپری شدن این زمان کانال کماکان آزاد باشد، آن‌گاه ایستگاه به اندازه‌ی یک زمان تصادفی عقب کشیده و انتقال خود را به تعویق می‌اندازد. طی این زمان تعویق، که به آن زمان عقب‌گرد گفته می‌شود، ایستگاه کماکان کانال را شنود می‌کند و در صورت مشغول بودن کانال، زمان‌سنج عقب‌گرد

متوقف، و با آزاد شدن مجدد کانال زمان‌سنج از همان نقطه‌ی توقف به حرکت خود ادامه می‌دهد. در نهایت پس از پایان زمان عقب‌گرد، ایستگاه فریم خود را ارسال می‌کند.

۴. اگر پس از ارسال، ایستگاه فریم ACK دریافت نکند، آن‌گاه ارسال موفقیت‌آمیز نبوده و تصادم رخ داده است.

برای اطمینان از ثبات روش عقب‌گرد، الگوریتمی موسوم به BEB[۱۰] برای این کار مورد استفاده قرار می‌گیرد. در این الگوریتم، ایستگاه زمان عقب‌گرد خود را از بازه‌ی $[0 \ldots CW_{max}]$[۱۱] به طور تصادفی انتخاب می‌کند و در صورت بروز تصادم، مقدار CW_{max} که توانی از ۲ است دو برابر می‌شود. این کار موجب می‌شود که احتمال بروز مجدد تصادم کمتر شود.

روش دوم برای به اشتراک‌گذاری کانال، PCF است. در این روش، AP در نقش کنترل‌کننده عمل کرده و به هر ایستگاه که فریمی برای ارسال دارد یک روزنه‌ی زمانی جهت انتقال فریم اختصاص می‌دهد. به طور پیش‌فرض، از روش DCF در لایه‌ی MAC شبکه‌های Wi-Fi می‌شود.

امنیت

پروتکل WEP اولین پروتکل رمزنگاری 802.11 بود که با هدف تأمین محرمانگی فریم‌های داده معرفی شد. در این پروتکل، از یک کلید WEP ۴۰ یا ۱۰۴ بیتی برای رمزنگاری داده‌ها توسط الگوریتم RC4 استفاده می‌شود. برای احراز هویت نیز دو مد سیستم باز[۱۲] و کلید اشتراکی[۱۳] در نظر گرفته شده است. در مد اول هیچ عملیات احراز هویتی میان گره‌ی کاربر و AP صورت نمی‌گیرد و تنها از کلید WEP برای رمزنگاری فریم‌های داده استفاده می‌شود. در مد دوم، از کلید WEP برای انجام یک دست‌دهی ۴ مرحله‌ای استفاده می‌شود که طی آن، گره‌ی کاربر یک پیام تصادفی که از سوی AP برای آن ارسال شده است را با کلید WEP رمز، و برای AP ارسال می‌کند. در صورت صحیح بودن کلید، پیام توسط AP رمزگشایی و هویت کاربر تصدیق می‌شود.

پس از شناسایی ضعف‌های امنیتی فراوان در WEP، استاندارد IEEE 802.11i و پروتکل‌های WPA و WPA2 برای فراهم کردن امنیت بیشتر معرفی شدند. برخلاف WEP که در آن از یک کلید ثابت برای رمزنگاری فریم‌ها استفاده می‌شود، WPA از روش TKIP استفاده می‌کند که در آن از یک کلید رمزنگاری ۱۲۸ بیتی مجزا برای رمز کردن هر فریم استفاده می‌شود. هم‌چنین، به جای CRC، که دارای ضعف امنیتی است، از MIC[۱۴] استفاده می‌شود.

پس از WPA، پروتکل WPA2 معرفی شد. پروتکل WPA2 از پروتکل CCMP که مبتنی بر روش رمزنگاری AES است استفاده می‌کند و امنیت بیشتری نسبت به WPA فراهم می‌کند.

علاوه بر این، دو حالت شخصی[۱۵] و تشکیلاتی[۱۶] برای WPA و WPA2 در نظر گرفته شده اند. حالت شخصی برای کاربردهای عمومی نه چندان حساس (مانند خانه، اداره‌های کوچک و...) طراحی شده است که در آن از روش احراز هویت کلید پیش‌اشتراکی[۱۷] که مانند روش کلید اشتراکی در WEP عمل می‌کند استفاده می‌شود. در حالت تشکیلاتی، که با نام 802.1X هم شناخته می‌شود، از یک سرویس‌دهنده‌ی مجزا موسوم به RADIUS[۱۸] برای عملیات احراز هویت استفاده می‌شود.

در کل، اگرچه WPA2 دارای امنیت نسبتا مناسبی است، اما این امنیت صرفا برای تأمین محرمانگی فریم‌های داده طراحی شده است و هیچ عملیات احراز هویت و رمزنگاری بر روی فریم‌های غیر داده‌ای (مانند فریم‌های کنترلی و مدیریتی) صورت نمی‌گیرد.

فریم MAC

به طور کلی، علاوه بر فریم‌های داده، دو نوع فریم در استاندارد 802.11 وجود دارد: فریم‌های کنترلی و فریم‌های مدیریتی، که به ترتیب به منظور تضمین انتقال فریم‌های داده (مقابله با تصادم و از دست رفتگی) و مدیریت کانال ارتباطی مورد استفاده قرار می‌گیرند. جدول (۱) انواع فریم‌های موجود در استاندارد IEEE 802.11 را نشان می‌دهد.

جدول (۱): فریم‌های کنترلی و مدیریتی 802.11

شرح زیرنوع	ارزش فیلد زیرنوع	نوع فریم	ارزش فیلد نوع
درخواست پیوست[۱۹]	۰۰۰۰	مدیریتی	۰۰
پاسخ پیوست[۲۰]	۰۰۰۱	مدیریتی	۰۰
درخواست بازپیوست[۲۱]	۰۰۱۰	مدیریتی	۰۰
پاسخ بازپیوست[۲۲]	۰۰۱۱	مدیریتی	۰۰
درخواست جست-	۰۱۰۰	مدیریتی	۰۰
پاسخ جست‌وجو[۲۴]	۰۱۰۱	مدیریتی	۰۰
راهنما[۲۵]	۱۰۰۰	مدیریتی	۰۰
TIM[۲۶]	۱۰۰۱	مدیریتی	۰۰
قطع پیوست[۲۷]	۱۰۱۰	مدیریتی	۰۰
احراز هویت	۱۰۱۱	مدیریتی	۰۰
قطع احراز هویت[۲۸]	۱۱۰۰	مدیریتی	۰۰
PS-Poll	۱۰۱۰	کنترلی	۰۱
RTS	۱۰۱۱	کنترلی	۰۱
CTS	۱۱۰۰	کنترلی	۰۱
Ack	۱۱۰۱	کنترلی	۰۱
CF-End	۱۱۱۰	کنترلی	۰۱
CF-End+CF-Ack	۱۱۱۱	کنترلی	۰۱

IEEE 802.16 – ۲-۲

در این بخش، معماری و ساختار کلی 802.16 بررسی خواهد شد.

۲-۲-۱- معماری و سرویس‌ها

ابتدا 802.16 برای ارتباط دید مستقیم (LOS) میان ایستگاه پایه[۲۹] (BS) و کاربران در فرکانس ۱۰ تا ۶۶ GHz طراحی شده بود. اما بعدها ارتباط دید غیرمستقیم(NLOS) در فرکانس ۲ تا ۱۱ GHz به همراه تکنیک‌های OFDM، OFDMA و MIMO نیز به آن اضافه شد. از نسخه‌ی 802.16e-2005 امکان تحرک کاربران را نیز فراهم شد. آخرین نسخه از استاندارد (تحت عنوان 802.16m-2011) دارای نرخ داده‌ی ۱۰۰ Mbps برای کاربران متحرک[۳۰] (MS) و ۱ Gbps برای کاربران ایستا است.

مانند 802.11، 802.16 نیز در دو لایه‌ی فیزیکی و MAC تعریف می‌شود. لایه‌ی MAC خود شامل سه زیرلایه‌ی CS[۳۱]، SS[۳۳] است که به ترتیب وظیفه‌ی ارتباط با لایه‌های بالاتر و تبدیل داده‌های آن‌ها به قالب لایه‌ی MAC، برقراری و مدیریت ارتباط، و تضمین امنیت را بر عهده دارند. در لایه‌ی فیزیکی، کانال ارسال[۳۴] (UL) از تکنیک TDMA[۳۵] و کانال دریافت[۳۶] (DL) از TDM استفاده می‌کنند.

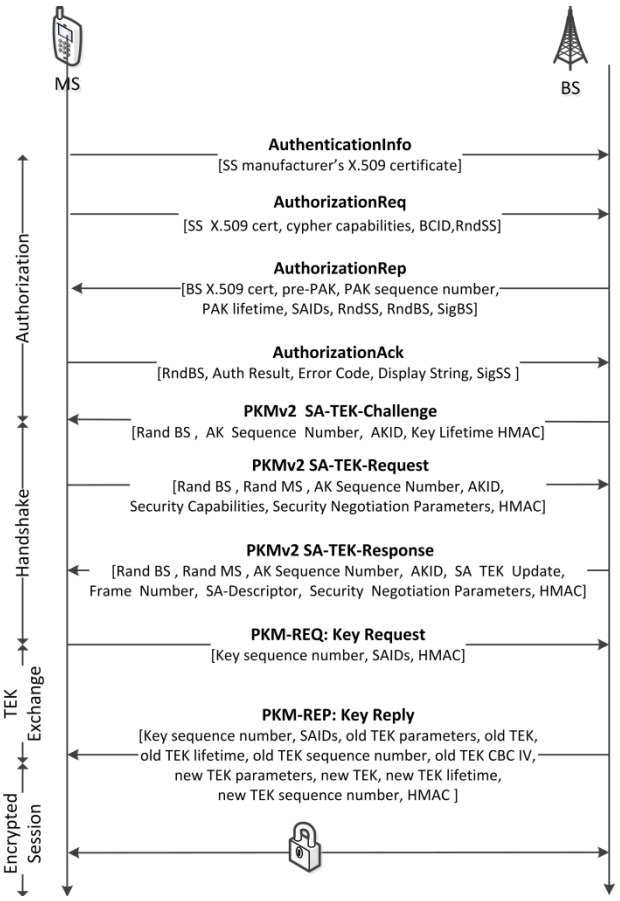
۲-۲-۲- ورود به شبکه

برای ورود به شبکه، MS فرکانس کانال‌های UL و DL را برای پیدا کردن پیام‌های DL-MAP[۳۷] و UL-MAP که توسط BS همه‌پخشی می‌شوند جست‌وجو می‌کند. بعد از به دست آوردن پارامترهای لازم از داخل این پیام‌ها، MS تلاش می‌کند تا به اطلاعات زمان‌بندی و تنظیم قدرت مناسب طی فرآیندی که به آراستن[۳۸] موسوم است برسد. این فاز با تبادل دو پیام RNG-REQ[۳۹] و RNG-RSP میان MS و BS انجام می‌شود. پس از این فاز، MS، BS را از قابلیت‌های خود آگاه و کلید-های امنیتی لازم میان MS و BS مبادله می‌شوند. در نهایت، فاز ثبت-نام صورت می‌گیرد که به معنای اجازه‌ی ورود MS به شبکه است.

لازم به ذکر است که پیام‌هایی طی مراحل ورود به شبکه میان MS و BS مبادله می‌شوند به صورت رمزنشده انتقال می‌یابند.

۲-۲-۳- امنیت

امنیت 802.16 بر عهده زیرلایه‌ی SS است که وظیفه‌ی تأمین محرمانگی، جامعیت، احراز هویت و محافظت در برابر سرقت داده‌ها توسط سرویس‌دهنده را بر عهده دارد. هسته‌ی اصلی امنیت 802.16 پروتکل PKM[۴۰] است که اجازه‌ی دسترسی کاربران به شبکه، توزیع کلید و رمزکردن ترافیک را بر عهده دارد. شکل (۱) به طورخلاصه عملکرد پروتکل PKMv2 را نشان می‌دهد.



شکل (۱): PKMv2 [34]

۳- حملات Wi-Fi

در حالت کلی، می‌توان حملات و مخاطرات امنیتی 802.11 را در قالب چندین گروه دسته‌بندی کرد. این گروه‌ها به همراه حملات مربوط به هر کدام در ادامه بررسی خواهند شد.

۳-۱- مسدودسازی فیزیکی

هدف از این حملات مسدود کردن کانال رادیویی شبکه از طریق ارسال سیگنال‌های پرقدرت است. در کار ارائه شده‌ی Konings و همکاران [2] سه نوع از حملات مسدود سازی معرفی شده‌اند؛ در *مسدودسازی با نرخ ثابت* حمله‌کننده به طور پیوسته و با نرخ ثابت اقدام به ارسال سیگنال یا نویز می‌کند. در *مسدودسازی تصادفی* حمله‌کننده با هدف صرف‌جویی در انرژی و نیز عدم شناسایی شدن، اقدام به ارسال سیگنال یا نویز به طور تصادفی می‌کند. اما پیچیده‌ترین نوع از حملات مسدودسازی، نوع *واکنشی* است که طی آن حمله‌کننده زمانی که یک پیام خاص را مشاهده کند اقدام به ارسال سیگنال یا نویز می‌کند تا از دریافت پیام توسط گیرنده جلوگیری کند. این حمله توسط Bayraktaroglu و همکاران [3] بررسی و پیاده‌سازی شده است.

۳-۲- حملات لایه‌ی MAC

در این بخش، حملات و آسیب‌پذیری‌های مربوط به لایه‌ی MAC در Wi-Fi بررسی خواهند شد.

۳-۲-۱- حملات 802.11i

این حملات مربوط به ضعف‌ها و آسیب‌پذیری‌های امنیتی پروتکل‌های WPA و WPA2 هستند. در کار گزارش شده‌ی Glass و همکاران [4] امکان انجام حمله‌ی ممانعت از سرویس[۴۱] (DoS) در برابر پروتکل TKIP که در WPA استفاده می‌شود بررسی شده است. در کار گزارش شده‌ی Xing و همکاران [5] کاستی‌های امنیتی مربوط به احراز هویت و محرمانگی داده‌ها در 802.11i مورد بررسی قرار گرفته‌اند و بیان شده است که حملات درون شبکه‌ای و نیز حملات برون‌خطی و حدس زدن کلمات عبور در 802.11i امکان‌پذیر هستند.

۳-۲-۲- جعل آدرس MAC[۴۲]

یکی از تهدیدهای جدی برای شبکه‌های بی‌سیم، مخصوصا شبکه‌های Wi-Fi، جعل آدرس MAC است. تمامی واسط‌های رادیویی دارای یک آدرس منحصر به فرد MAC هستند. اگرچه این آدرس توسط کارخانه‌ی سازنده‌ی واسط تعیین می‌شود، اما تغییر دادن آن کار ساده‌ای است (با اجرای دستور ifconfig در لینوکس یا با ایجاد تغییر در رجیستری ویندوز). هم‌چنین، هیچ‌گونه عملیات رمزنگاری یا احراز هویت برای آدرس‌های MAC انجام نمی‌شود. علاوه بر این، جدی بودن این آسیب‌پذیری از آنجا نشأت می‌گیرد که علاوه بر قابلیت اختلال در سرویس‌گیری سایر گره‌های شبکه، بستری برای انجام حملات دیگر است. در کارهای [6-10] انواع حملاتی که توسط جعل آدرس MAC قابل انجام هستند معرفی شده‌اند. این حملات عبارت‌اند از:

حمله‌ی مرد میانی[۴۳] (MITM). در این حمله، حمله‌کننده آدرس MAC و نام AP شبکه (BSSID) را جعل و با ارسال فریم‌های قطع احراز هویت برای یکی از گره‌های شبکه، ارتباط گره با AP قانونی شبکه را قطع می‌کند. گره‌ی بیرون انداخته‌شده مجددا برای ارتباط با شبکه اقدام می‌کند. در این حالت، با توجه به تشابه آدرس MAC و BSSID حمله‌کننده و AP، ممکن است گره به AP حمله‌کننده وصل شود. از طرفی، حمله‌کننده می‌تواند با استفاده از یک واسط رادیویی دیگر و جعل آدرس MAC گره‌ی قربانی، به AP اصلی (در نقش گره قربانی) نیز وصل و مانند یک پل ارتباطی میان AP و قربانی عمل کند. بسته‌ی نرم‌افزاری AirJack [11] حاوی برنامه‌ای با نام monkey_jack است که قادر است این حمله را به طور خودکار انجام دهد.

حمله‌ی قطع احراز هویت/ قطع پیوست. در IEEE 802.11، هر گره باید برای برقراری ارتباط با ارسال فریم‌های درخواست پیوست/احراز هویت خود را به AP معرفی کند. هم‌چنین برای قطع ارتباط از فریم‌های قطع احراز هویت/ قطع پیوست استفاده می‌شود. در این حالت، حمله‌کننده قادر است با تغییر آدرس MAC خود به آدرس AP، به طور پیوسته فریم‌های جعلی قطع احراز هویت/ قطع پیوست برای گره‌های شبکه ارسال و مانع از برقراری ارتباط آن‌ها با AP شود. نتیجه‌ی این عمل یک حمله‌ی DoS خواهد بود. برای ایجاد فریم‌های قطع احراز هویت/ قطع پیوست می‌توان از نرم‌افزارهای AirJack یا void11 [12] استفاده کرد.

حمله در حالت ذخیره‌ی انرژی. در 802.11 یک حالت ذخیره‌ی انرژی در نظر گرفته شده است که طی آن، گره برای ذخیره‌ی انرژی باتری خود به حالت خواب می‌رود. فریم‌هایی که در طول این زمان برای گره ارسال می‌شوند توسط AP بافر می‌شوند، تا زمانی که گره با ارسال فریم مدیریتی PS-Poll فریم‌های بافرشده‌ی خود را از AP درخواست کند. با جعل آدرس MAC گره و قرار دادن آن در یک فریم PS-Poll، حمله‌کننده می‌تواند فریم‌های بافر شده‌ی گره را از AP درخواست و در نتیجه باعث از دست رفتن آن‌ها شود. این حمله توسط Gu و همکاران [13] بر روی دستگاه‌های مختلف که دارای پیاده‌سازی‌های متفاوتی از مکانیزم مدیریت انرژی هستند اجرا و تبعات آن بررسی شده است.

جعل فریم راهنما[۴۴]. در استاندارد 802.11، فریمی موسوم به فریم راهنما به طور متناوب (به طور پیش فرض هر ۱۰۰ ms) توسط AP همه‌پخشی می‌شود. در این فریم اطلاعات لازم برای شناسایی شبکه، همگام‌سازی ساعت گره‌ها با ساعت AP و نام گره‌هایی که فریم بافر شده برای آن‌ها در AP وجود دارد قرار دارند. با توجه به اینکه هیچ‌گونه احراز هویت برای این فریم صورت نمی‌گیرد، حمله‌کننده قادر است با جعل آدرس AP شبکه در یک فریم راهنما و قرار دادن اطلاعات نادرست (مانند ساعت اشتباه) در آن سبب ایجاد اختلال در عملکرد شبکه شود. حملات مربوط به فریم راهنما توسط Martinez و همکاران [9] بررسی شده‌اند.

۳-۲-۳- حملات فریم‌های کنترلی و CSMA/CA

پیشرفت نرم‌افزارها و توسعه‌ی ابزارها و درایورهای متن‌باز نظیر MadWifi [14] این امکان را فراهم کرده تا بتوان پارامترهای مربوط به پروتکل‌های کنترل دسترسی شبکه را نیز تغییر داد. حملات فریم‌های کنترلی و پروتکل CSMA/CA مربوط به آسیب‌پذیری‌های فریم‌ها و الگوریتم‌های مورد استفاده در مکانیزم DCF می‌شوند. این حملات به دو دسته تقسیم می‌شوند: حملات DoS یا مسدودسازی[۴۵] مجازی با هدف ممانعت از سرویس شبکه، و رفتارهای حریصانه با هدف استفاده‌ی بیشتر از پهنای باند شبکه. در کارهای گزارش شده‌ی [14-19] به بررسی این حملات که ناشی از عدم اعمال هر گونه مکانیزم امنیتی بر روی فریم‌های کنترلی است پرداخته شده است. از جمله‌ی این حملات می‌توان به افزایش مقدار فیلد مدت زمان (مقدار NAV) در فریم‌های RTS/CTS به منظور سکوت گره‌های شبکه، کم کردن مقدار پارامتر DIFS به منظور استفاده‌ی بیشتر حمله‌کننده از شبکه و تغییر در پارامترهای الگوریتم BEB و جعل فریم ACK اشاره کرد. در کارهای [19, 20] این حملات بر اساس آگاهی حمله‌کننده از نحوه‌ی عملکرد سیستم تشخیص نفوذ شبکه[۴۶] (IDS) به دو دسته‌ی

ساده و هوشمندانه تقسیم شده‌اند و انواع پیچیده‌تری از آن‌ها که قادر به دور زدن روش‌های تشخیص رایج هستند معرفی شده‌اند.

۳-۳- آسیب‌پذیری‌های مربوط به پیاده‌سازی

این آسیب‌پذیری‌ها به نحوه‌ی پیاده‌سازی محصولات توسط شرکت‌ها مرتبط می‌شود. در کار Ferrari و همکاران [21] با ارسال حجم وسیعی از فریم‌های مدیریتی جست‌وجو، احراز هویت و پیوست به APهای مختلف، تأثیر این حملات بر APها مورد بررسی قرار گرفته‌اند. هم‌چنین در کار گزارش شده توسط Gu و همکاران [13]، آسیب‌پذیری‌های APها و واسط‌های رادیویی مختلف در برابر حملات مربوط به مکانیزم مدیریت انرژی (بخش ۳-۲-۲) مورد بررسی قرار گرفته‌اند.

در جدول (۲) خلاصه‌ای از انواع حملات و آسیب‌پذیری‌های موجود در استاندارد IEEE 802.11 آمده‌اند.

جدول (۲): حملات IEEE802.11

نام حمله	قابل اجرا
مسدودسازی فیزیکی کانال	
یکنواخت	IBSS, BSS
انفجاری	IBSS, BSS
تصادفی	IBSS, BSS
واکنشی	IBSS, BSS
حملات 802.11i	
TKIP	IBSS, BSS
EAP	IBSS, BSS
RSN IE	IBSS, BSS
دست‌دهی ۴ طرفه	IBSS, BSS
جعل آدرس MAC	
MITM	IBSS, BSS
حمله در مد ذخیره‌ی انرژی	BSS
تغییر پارامترهای فریم راهنما	BSS
Deauth/Disass DoS	BSS
حمله‌ی تعویض کانال (802.11h) [2]	IBSS, BSS
حمله‌ی سکوت (802.11h) [2]	IBSS, BSS
حملات DCF	
جعل RTS/CTS	IBSS, BSS
جعل ACK	IBSS, BSS
تغییر اندازه‌ی (D/S)IFS	IBSS, BSS
تغییر پارامترهای BEB	IBSS, BSS
حملات Block ACK (802.11n) [2]	
جعل BlockAck	BSS
جعل ADDBA	BSS
جعل DELBA	BSS
حملات مربوط به پیاده‌سازی	
سرریز بافر	IBSS, BSS
DoS فریم‌های مدیریتی [۲۱]	IBSS, BSS

۴- حملات WiMax

در این بخش، حملات مربوط به WiMax مورد بررسی قرار خواهند گرفت.

۴-۱- حملات فاز آراستن

یکی از فازهای اولیه برای ورود به شبکه، فاز آراستن است. طی این فاز، BS و MS اطلاعات لازم برای زمان‌بندی و تنظیم قدرت ارسال را مبادله می‌کنند. طی این فاز، دو پیام RNG-REQ و RNG-RSP میان MS و BS مبادله می‌شوند. این دو پیام حاوی اطلاعات لازم برای فاز آراستن هستند. به طورمشخص، RNG-RSP که از سوی BS برای MS ارسال می‌شود، حاوی ID این MS، اطلاعات مربوط به سطح قدرت RF و اطلاعات زمان‌بندی و تنظیم فرکانس است. هم‌چنین، در پیام RNG-RSP یک فیلد وضعیت وجود دارد که می‌تواند یکی از سه مقدار موفقیت‌آمیز، ادامه و شکست را داشته باشد. در حالت ادامه، MS و BS به ادامه‌ی تبادل RNG-REQ و RNG-RSP برای برقراری ارتباط بهتر می‌پردازند و در حالت شکست، MS سیکل ورود به شبکه را (با جست‌جو برای فرکانس DL) از نو آغاز می‌کند.

با توجه به عدم اعمال رمزنگاری و تصدیق جامعیت برای این دو پیام، اطلاعات درونی آن‌ها قابل شنود و پیام‌ها قابل جعل هستند. در نتیجه حمله‌کننده می‌تواند با شنود ID مربوط به MS و ارسال پیام RNG-RSP با پارامترهای مخدوش، عملکرد MS را تحت تأثیر قرار دهد. برای مثال، حمله‌کننده می‌تواند پارامتر فرکانس در پیام RNG-RSP را تغییر و MS را مجبور به تغییر کانال فرکانسی کند، یا می‌تواند با تغییر مقدار فیلد وضعیت، MS را مجبور به ارسال مجدد پیام RNG-REQ یا شروع مجدد سیکل ورود به شبکه کند. در نوع خاصی از این حمله که با نام حمله‌ی شکنجه‌ی‌آبی[۴۷] شناخته می‌شود [34]، حمله‌کننده پارامتر قدرت سیگنال را به بیشترین مقدار ممکن تغییر و سبب کاهش زودهنگام توان باتری MS می‌شود. انواع مختلفی از حملات آراستن در [36-34] مورد بررسی قرار گرفته اند که در جدول (۳) در بخش ضمایم معرفی شده‌اند.

۴-۲- حملات در حالت ذخیره‌ی انرژی

پشتیبانی WiMax از گره‌های متحرک، مسئله‌ی توان محدود باتری‌های MSها و نیاز به ذخیره‌ی انرژی را به همراه دارد. در حالت ذخیره‌ی انرژی، MS به اندازه‌ی یک مدت زمان مذاکره‌شده با BS، برخی از کارکردهای[۴۸] خود را خاموش می‌کند و به حالت خواب می‌رود. حالت خواب با دو بازه‌ی *دسترسی* و *عدم‌دسترسی* در کانال DL یا UL تعریف می‌شود. در بازه‌ی دسترسی، MS می‌تواند به ترافیک‌های روی کانال DL یا UL دسترسی داشته باشد. در بازه‌ی عدم دسترسی، MS نمی‌تواند به/از BS ارسال/دریافت داشته باشد، بنابراین، اگر طی زمان خواب، BS ترافیک برای ارسال به MS داشته باشد، باید آن را بافر کند یا با ارسال پیام MOB_TRF-IND[۴۹] MS را از حالت خواب

بیدار کند. حالت خواب در نهایت با ارسال یک پیام مدیریتی خاص از BS برای MS خاتمه می‌یابد.

حملات این دسته، مربوط به مکانیزم ذخیره‌ی انرژی و پیام‌هایی است که برای این منظور مبادله می‌شوند. در کار Kolias و همکاران [34] این حملات مورد بررسی قرار گرفته‌اند. در نوعی از این حملات، حمله‌کننده با ارسال ترافیک بیهوده (مثلا بسته‌های تهی TCP/IP) که به یک MS خاص آدرسی‌دهی شده اند، سبب خروج MS از حالت خواب می‌شود. اگرچه موانع زیادی برای انجام این حمله توسط نویسندگان معرفی شده (از جمله اینکه ممکن است BS ترافیک را برای MS بافر کند)، اما انجام این حمله در شرایط خاص امکان پذیر است. ارسال پیام‌های جعلی MOB_TRF-IND جهت بیدار کردن MSهای در حالت خواب، ارسال پیام‌های جعلی MOB-SLP-REQ[۵۰] با آدرس جعل شده‌ی یک MS خاص برای BS جهت قرار دادن وضعیت خواب برای MS، انواع دیگری از این حملات هستند.

۴-۳- حملات دگرسپاری

دگرسپاری در WiMax فرآیند قطع اتصال MS از یک BS و اتصال آن به BS همسایه است. این فرآیند شامل چندین مرحله است:

- BSهای همسایه توسط MS شناسایی می‌شوند. برای این منظور، هر BS به صورت متناوب یک پیام MOB_NBR_ADV[۵۱] ارسال می‌کند که شامل اطلاعات لازم است.
- MS اقدام به آغاز دگرسپاری می‌کند. این مرحله می‌تواند از سوی BS نیز آغاز شود.
- MS با استخراج پارامترهای DL و UL با BS جدید همگام می‌شود.
- فاز آراستن میان MS و BS جدید آغاز می‌شود.
- ارتباط MS با BS قدیمی قطع می‌شود.

طی این مراحل، جامعیت پیام MOB_NBR_ADV محافظت نمی‌شود. در نتیجه حمله‌کننده قادر است با جعل این پیام MS را از BS جدا کند [34-36].

۴-۴- حملات پیام‌های کنترلی

این حملات مربوط به پیام‌های کنترلی هستند که بین MS و BS تبادل می‌شوند. با توجه به عدم تصدیق جامعیت و رمزنگاری این پیام-ها، حمله‌کننده قادر است با تغییر پارامترهای آن‌ها عملکرد شبکه را مخدوش کند. یکی از انواع این حملات که در کارهای [34-36] به آن اشاره شده است، مربوط به پیام کنترل سریع قدرت[۵۲] (FPC) است. این پیام از سوی BS برای یک یا چندین MS جهت کنترل قدرت ارسال آن‌هاارسال می‌شود. با تغییر پارامترهای این پیام، حمله‌کننده می‌تواند سبب افزایش یا کاهش قدرت ارسال MSها شود. حالت اول منجر به کاهش زودهنگام توان باتری MSها می‌شود. در حالت دوم، MS برای افزایش قدرت خود پیام‌های متعددی در کانال UL برای BS ارسال می‌کند تا به قدرت لازم برسد که اتلاف پهنای باند UL را به همراه دارد.

یکی دیگر از آسیب‌پذیری‌های پیام‌های کنترلی مربوط به پیام SBC_REQ[۵۳] است. این پیام در فاز اولیه‌ی ورود به شبکه از سوی MS برای BS جهت آگاهی BS از قابلیت‌های MS ارسال می‌شود. با توجه به اینکه این پیام قبل از آغاز هرگونه ارتباط رمزشده میان BS و MS رد و بدل می‌شود، هیچ‌گونه عملیات تصدیق جامعیت برای این پیام اعمال نمی‌شود. در نتیجه حمله‌کننده می‌تواند پارامترهای آن را تغییر دهد. برای مثال، حمله‌کننده می‌تواند پارامترهای این پیام را به گونه‌ای تغییر دهد که ارتباط میان BS و MS یک ارتباط رمزنشده باشد.

سایر پیام‌های کنترلی که مستعد انجام حمله هستند، پیام‌های DBPC-REQ[۵۴] و RES-CMD[۵۵] هستند که توسط Kolias و همکاران [34] مورد بررسی قرار گرفته‌اند.

۴-۵- حملات مربوط به مکانیزم‌های امنیتی

این حملات مربوط به آسیب‌پذیری‌های موجود در روش‌های احراز هویت و رمزنگاری استفاده شده در WiMax هستند. در کار Kolias و همکاران [34] هفت نوع از این حملات مطرح شده‌اند که در جدول(۳) آمده‌اند.

۴-۶- حملات مربوط به مکانیزم همه/چندپخشی

در WiMax امکان همه‌پخشی یا چندپخشی یک پیام از سوی BS برای گروه خاصی از MSها وجود دارد. برای این کار، پیام با یک کلید GTEK[۵۶] که از قبل میان MSهای عضو گروه توسط BS به اشتراک گذاشته شده است رمزنگاری می‌شود. اما مشکل آن‌جاست که هر یک از اعضاء گروه می‌تواند از این کلید برای رمزکردن هر پیام دلخواه استفاده و آن را برای سایر اعضاء ارسال کند، بدون آنکه دریافت کننده بتواند تشخیص دهد که این پیام از جانب BS ارسال شده یا یکی از اعضاء گروه.

نوع دیگری از این حملات که تحت عنوان سرقت اطلاعات از آن یاد شده [34]، از آنجا ناشی می‌شود که هر MS جدید که به گروه اضافه می‌شود قادر است ترافیک‌هایی که قبل از ورودش به گروه ردوبدل شده‌اند را نیز رمزگشایی کند (در صورتی که این ترافیک در جایی ذخیره شده باشد). اگر چه کلید GTEK به طور متناوب به روز می‌شود اما ترافیک‌هایی که در طول مدت معتبر بودن GTEK فعلی انتقال یافتند قابل رمزگشایی هستند.

نوع دیگری از این حملات مربوط به پیام MCA-REQ[۵۷] است. این پیام از سوی BS برای MS جهت پیوستن یا ترک یک گروه ارسال می‌شود. از آنجا که این پیام رمز نمی‌شود، لذا حمله‌کننده با جعل آن می‌تواند سبب خروج یک MS از گروه همه‌پخشی شود.

۵- راه‌کارهای امنیتی مربوط به جعل آدرس MAC در Wi-Fi

در حالت کلی، راه‌کارهای مقابله با جعل آدرس MAC در دو دسته قرار می‌گیرند: پیش‌گیری و تشخیص. در راه‌کارهای پیش‌گیری هدف جلوگیری از انجام حمله است. این‌گونه روش‌ها معمولا با ایجاد تغییر در ساختار پروتکل و استفاده از روش‌های رمزنگاری عمل می‌کنند. از این‌رو، هزینه‌ی پیاده‌سازی و استفاده از این روش‌ها نسبتا بالا است.

از طرف دیگر، راه‌کارهای تشخیص با هدف تشخیص نفوذ از طریق استفاده از یک IDS مستقل در شبکه عمل می‌کنند. بنابراین، استفاده از این روش‌ها نیازمند ایجاد تغییر در ساختار پروتکل نیست و دارای هزینه‌ی کمتری هستند. همچنین، مقیاس‌پذیری آن‌ها و قابلیت به روزرسانی جهت مقابله با مخاطرات جدید نیز بالاتر است. معمولا IDSها به همراه یک سیستم پاسخ به نفوذ[۵۸] (IRS) به کار می‌روند که وظیفه‌ی پاسخ به نفوذ را برعهده دارد.

در این بخش، راه‌کارهای ارائه شده برای پیش‌گیری و تشخیص حملات مربوط به جعل آدرس MAC در Wi-Fi بررسی می‌شوند.

۵-۱- روش‌های پیش‌گیری

در کار گزارش شده‌ای توسط Nagarajan و همکاران [22] یک روش پیش‌گیری مبتنی بر قدرت سیگنال دریافتی[۵۹] (RSS) ارائه شده است. بر این اساس، ابتدا یک دست‌دهی میان AP و گره‌ی کاربر صورت می‌گیرد. طی این فاز، حداقل قدرت لازم برای ارسال فریم‌ها از گره به AP (Pmin) بر اساس فاصله‌ی گره از AP مشخص، و یک دنباله‌ی تصادفی از اعداد بزرگتر از Pmin میان آن‌ها مبادله می‌شود. پس از این فاز، قدرت ارسال تمامی فریم‌هایی که میان AP و گره مبادله می‌شوند می‌بایست از این دنباله اعداد تصادفی پیروی کند. روش ارائه شده دارای هزینه‌ی پیاده‌سازی فراوانی است، چرا که تغییر قدرت ارسال مستلزم سخت‌افزارهای ویژه است و همواره امکان‌پذیر نیست (محدودیت‌های محیطی و تداخل با دستگاه‌های دیگر). علاوه‌بر این، این روش مستلزم برقراری یک دست‌دهی رمز شده میان AP و گره است که با هر بار تغییر مکان گره می‌بایست انجام شود.

در کار Laishun و همکاران [23] یک روش برای پیش‌گیری از حمله‌ی DoS احراز هویت ارائه شده است. در این حمله، حمله‌کننده حجم وسیعی از فریم‌های درخواست احراز هویت با آدرس‌های MAC مختلف برای AP ارسال می‌کند. به ازاء هر درخواست احراز هویت، AP یک فریم پاسخ احراز هویت برای حمله‌کننده ارسال و منتظر فریم ACK باقی می‌ماند. حمله‌کننده از ارسال ACK خودداری و در نتیجه AP ملزم به ارسال مجدد فریم پاسخ احراز هویت می‌شود. با ادامه‌ی این کار، منابع محاسباتی AP اشغال و سرویس سایر گره‌های شبکه مختل یا از کیفیت آن‌ها کاسته می‌شود. اگرچه حمله‌ی معرفی شده توسط آن‌ها نوعی DoS و با هدف اشغال منابع محاسباتی AP انجام می‌شود، اما راه‌حل ارائه شده توسط آن‌ها می‌تواند برای مقابله با جعل آدرس MAC در مورد فریم‌های احراز هویت (یا حتی سایر فریم‌های مدیریتی) اعمال شود. روش ارائه شده‌ی آن‌ها AP را ملزم به ارسال یک پازل برای گره‌های درخواست کننده‌ی احراز هویت می‌کند و در صورتی که گره بتواند پازل را حل کند، فریم‌های بعدی وی توسط AP پذیرفته می‌شوند. ایراد عمده‌ی این روش نیازمندی آن به ایجاد تغییرات فراوان در سخت‌افزار و پروتکل است. هم‌چنین لازمه‌ی این روش یافتن یک پازل مناسب است.

برای پیش‌گیری از حمله‌ی DoS قطع احراز هویت/ قطع پیوست (بخش ۳-۲-۲)، پیشنهاد شده است که پس از دریافت هر فریم قطع احراز هویت/ قطع پیوست توسط AP، یک تأخیر ۵ تا ۱۰ ثانیه‌ای اعمال شود [24]. اگر در طی این زمان فریم‌های دیگری از همان آدرس MAC برای AP ارسال شود به معنای حمله است. اما این روش چندان کارا نیست، چرا که حمله‌کننده می‌تواند ۱۰ ثانیه صبر کرده و سپس اقدام به ارسال فریم‌های بعدی نماید [7].

برای پیش‌گیری از DoS قطع احراز هویت/ قطع پیوست، استفاده از تئوری اعداد اول نیز پیشنهاد شده است [10]. بر این اساس، تمامی گره‌های شبکه ملزم به انتخاب دو عدد اول بزرگ p و q (N=p*q) هستند. به طور مشابه، AP نیز دو عدد اول بزرگ s و t (M=s*t) انتخاب می‌کند. طی یک فاز احراز هویت، مقادیر N و M میان AP و گره‌ی خواهان اتصال به شبکه ردوبدل می‌شوند. هر زمان که گره خواهان قطع ارتباط با AP باشد، باید مقدار p (یا q) را همراه با فریم قطع احراز هویت (یا قطع پیوست) خود برای AP ارسال نماید. در صورتی که مقدار ارسال شده صحیح باشد، آن‌گاه N بر p (یا q) بخش‌پذیر خواهد بود و درخواست گره برای جدا شدن از شبکه توسط AP انجام خواهد شد ایراد اساسی این روش امکان انجام حمله‌ی DoS بر روی AP با ارسال حجم وسیعی از مقادیر N تصادفی است.

برای پیش‌گیری از حمله‌ی DoS قطع احراز هویت/ قطع پیوست ، پیشنهاد شده است که فریم‌های قطع احراز هویت/ قطع پیوست به کلی نادیده گرفته شوند [7]. اما این روش مشکلاتی در شبکه‌های متشکل از چندین AP که نیازمند دگرسپاری[۶۰] گره‌ها میان APها هستند به وجود خواهد آورد.

استفاده از پروتکل 802.1X و رمزنگاری و اعمال احراز هویت روی فریم‌های مدیریتی نیز روش دیگر برای محافظت از فریم‌های مدیریتی در برابر جعل MAC است [7]. با این‌حال، این امر مستلزم مدیریت کلیدها است. هم‌چنین استفاده از رمزنگاری سبب افزایش زمان پردازش و پهنای باند می‌گردد.

۵-۲- روش‌های تشخیص

استفاده از روش‌های تشخیص (به جای پیش‌گیری) برای مقابله با جعل MAC عملی‌تر است، چرا که نیازمند ایجاد تغییر در پروتکل و سخت‌افزارهای گره‌های شبکه نیست. روش‌های تشخیص بر این اساس عمل

می‌کنند که با بررسی یک یا مجموعه‌ای از ویژگی‌های موجود در ترافیک، فریم‌های اصلی از جعلی متمایز و وجود حمله تشخیص داده می‌شوند. بر همین مبنا، می‌توان کارهای صورت گرفته را بر اساس ویژگی‌هایی که برای تشخیص حمله مورد استفاده می‌دهند به چند دسته تقسیم کرد که در ادامه مورد بررسی قرار خواهند گرفت.

۵-۲-۱- فیلد کنترل دنباله[۶۱]

استفاده از فیلد کنترل دنباله در سرآیند فریم‌ها می‌تواند برای تشخیص جعل MAC مورد استفاده قرار گیرد. کاربرد این فیلد در حفظ ترتیب فریم‌های داده و مدیریتی است (فریم‌های کنترلی فاقد این فیلد هستند). ایده‌ی اصلی آن است که اگر دو فریم دارای آدرس‌های MAC یکسان باشند اما اختلاف کنترل دنباله‌های آن‌ها از یک حد آستانه تجاوز کند، آن‌گاه یکی از دو فریم جعلی است. این روش توسط Li و همکاران [27] ارائه و به تفضیل مورد بررسی قرار گرفته است. همچنین در کار گزارش شده توسط Bansal و همکاران [6] این روش پیاده‌سازی و نرخ تشخیص ۹۴.۴۸ درصد برای آن به دست آمده است. با اینکه این روش یکی از مطرح‌ترین روش‌های موجود است، اما چند ایراد اساسی به آن وارد است. اول نرخ FP[۶۲] بالای این روش. در کار Bansal و همکاران [6]، FP به دست آمده ۸۱.۹۱ درصد است. دوم آنکه این روش بر روی فریم‌های کنترلی قابل اجرا نیست، و سوم آنکه فیلد کنترل دنباله خود قابل شنود و جعل است. در کار گزارش شده‌ی Chandrasekaran و همکاران [28] استفاده از این فیلد در کنار فیلدهای نوع فریم[۶۳] و اولویت QoS[۶۴] برای تشخیص این حمله در شبکه‌های 802.11e پیشنهاد شده است.

در کار گزارش شده توسط Madory [29] نیز از فیلد کنترل دنباله برای تشخیص جعل MAC استفاده شده است، با این تفاوت که نرخ ارسال فریم‌ها نیز در نظر گرفته شده است. بنابراین، اثر از دست - رفتن فریم‌ها به دلیل خطای کانال کم‌تر می‌شود. در این روش که SNRA نام دارد، اگر S(i) و S(i-۱) فریم‌های iام و (i-۱)ام باشند و T(i) و T(i-۱) زمان‌های ورود آن‌ها، تابعی تحت عنوان Gap بر اساس رابطه‌ی (۱) محاسبه می‌شود.

$$Gap=\{S(i) - S(i-1)\}/\{T(i) - T(i-1)\} \qquad (1)$$

اگر مقدار این Gap از حداکثر تعداد فریم‌های قابل ارسال در ۱ ثانیه بیشتر باشد، آنگاه پیام اخطار حمله صادر می‌شود. این روش توسط Bansal و همکاران [6] پیاده‌سازی و نرخ تشخیص ۳۵.۰۲ درصد برای آن به دست آمده است.

۵-۲-۲- RSS

مقدار RSS کمیت دیگری است که برای تشخیص جعل MAC می‌توان از آن استفاده کرد. این کمیت تابعی است از قدرت ارسال فرستنده، فاصله‌ی میان فرستنده و گیرنده، و محیط انتقال. روش‌های مبتنی بر RSS بر این فرضیه استوار هستند که حمله‌کننده‌ای که از نظر فیزیکی در مجاورت قربانی قرار ندارد مقادیر RSSاش با مقادیر RSS قربانی متفاوت است. بر این اساس، در [30, 31] مقادیر RSS برای تمامی آدرس‌های MAC استخراج و الگوریتم خوشه‌بندی k-means بر روی آن‌ها اعمال می‌شود؛ اگر اختلاف مرکز ثقل خوشه dB ۶ یا بیشتر باشد بیانگر حمله است.

در کار ارائه شده‌ی Sheng و همکاران [8] یک مدل مختلط گوسی (GMM) برای هر یک از گره‌های شبکه بر اساس RSS دریافتی از آن‌ها در هر یک از گیرنده‌های IDS ساخته می‌شود. در نتیجه، اگر حمله‌ای صورت گیرد، توزیع RSS مشاهده شده در گیرنده‌ها با GMM موجود مطابقت نخواهد داشت.

ایراد عمده‌ی روش‌های مبتنی بر RSS آن است که تحرک گره‌ها استفاده از این روش‌ها را با مشکل مواجه می‌کند [30]. هم‌چنین، اگر حمله‌کننده از نظر فیزیکی در مجاورت قربانی قرار داشته باشد، مقادیر RSS آن‌ها نزدیک به هم خواهد بود.

۵-۲-۳- ویژگی‌سنجی[۶۵] یا انگشت‌نگاری[۶۶]

ویژگی‌سنجی یا انگشت‌نگاری دستگاه‌های شبکه روش دیگر برای تشخیص جعل آدرس MAC است. ایده‌ی اصلی این روش آن است که برای هر یک از گره‌های شبکه یک مجموعه ویژگی یا اثرانگشت منحصر به فرد ایجاد گردد تا بتوان از آن برای تمییز دادن فریم‌های اصلی از جعلی استفاده کرد. در کار ارائه شده توسط Franklin و همکاران [32] فراوانی فریم‌های درخواست جست‌وجو گره‌های شبکه محاسبه و در صورت مشاهده‌ی تغییر ناگهانی در آن وجود حمله‌ی جعل MAC تشخیص داده می‌شود. با این‌وجود، اگرچه این روش متحرک بودن گره‌ها را پشتیبانی می‌کند، اما حمله کننده می‌تواند با ابزارهایی نظیر MadWifi [14] ارسال فریم‌های درخواست جست‌وجو را غیرفعال کند و مانع از تشخیص حمله شود [30].

روش دیگر برای انگشت‌نگاری استفاده از ویژگی‌های کارت‌های رادیویی شبکه است [8]. این روش بر این اساس استوار است که هر کارت رادیویی دارای ویژگی‌های منحصر به فردی است که سبب می‌شود بتوان سیگنال‌های ارسال شده از آن را از سیگنال‌های سایر کارت‌ها متمایز کرد. برای این منظور، Hall و همکاران [37] پیشنهاد کرده‌اند تا ویژگی‌های سیگنال‌های رادیویی در حوزه‌ی فرکانس استخراج و بر اساس آن یک اثرانگشت منحصر به فرد برای هر یک از گره‌های شبکه ساخته شود. سپس با استفاده از شبکه‌های عصبی سیگنال‌های دریافتی از ترافیک با اثرانگشت‌ها مقایسه و وجود حمله تشخیص داده شود. اگرچه این روش قابل دور زدن در سطح نرم‌افزار نیست، اما مستلزم نمونه برداری و تحلیل ترافیک با نرخی معادل با فرکانس شبکه (۲.۴ GHz یا ۵ GHz) است، لذا استفاده از آن چندان کم‌هزینه نیست [8].

۵-۲-۴- سایر روش‌ها

یک روش دیگر برای تشخیص جعل آدرس MAC، استفاده از سرآیند PLCP است [25]. در سرآیند پروتکل PLCP پارامترهای لایه‌ی فیزیکی مانند نرخ انتقال و ویژگی‌های مدولاسیون قرار می‌گیرند. با توجه به اینکه تغییر این پارامترها کار چندان راحتی نیست، می‌توان دو فریم با آدرس MAC یکسان اما از سوی دو فرستنده‌ی مختلف را از روی مقادیر PLCP تمییز داد. با این وجود، یک حمله‌کننده‌ی ماهر قادر به تغییر این پارامتر توسط درایورهای متن‌باز خواهد بود. هم‌چنین احتمال یکسان بودن این پارامترها برای حمله‌کننده و گره‌ی قربانی نیز وجود دارد.

جعل آدرس MAC با استفاده از پروتکل RARP نیز می‌تواند بررسی شود [26]. با توجه به اینکه به ازاء هر آدرس MAC تنها یک آدرس IP وجود دارد، اگر یک آدرس MAC جعل شود آن‌گاه دو آدرس IP با یک MAC وجود خواهند داشت و با استفاده از پروتکل RARP می‌توان آن را تشخیص داد. واضح است که این روش چندان قابل اتکا نیست، چرا که حمله‌کننده می‌تواند آدرس IP را نیز جعل نماید [7].

۶- نتیجه

در این نوشتار، حملات و مخاطرات Wi-Fi و WiMax با تمرکز بر آسیب‌پذیری‌های لایه‌ی MAC، به ویژه جعل آدرس MAC در شبکه‌های Wi-Fi، بررسی شدند. این حملات در نتیجه‌ی عدم وجود یک مکانیزم امنیتی قوی برای فریم‌های کنترلی و مدیریتی به وجود آمده‌اند. با توجه به اینکه استفاده از روش‌های رمزنگاری دارای هزینه و سربار بوده و بحث مدیریت کلید را به همراه دارد، راه‌کارهایی برای پیش‌گیری و تشخیص حملات مربوط به جعل آدرس MAC در Wi-Fi ارائه شده‌اند. روش‌های مبتنی بر تشخیص به دلیل هزینه و سربار کمتر روش‌های مناسب‌تری هستند، چرا که نیازمند ایجاد تغییر در ساختار پروتکل نیستند. از میان روش‌های تشخیص ارائه شده، استفاده از RSS، انگشت‌نگاری و فیلد کنترل دنباله دارای ثبات و دقت بیشتری نسبت به سایر روش‌ها هستند. با این وجود، هر یک دارای معایب و محدودیت‌هایی هستند. لذا، ارائه‌ی یک روش تشخیص مناسب برای حملات ناشی از جعل آدرس MAC که دارای دقت، ثبات و جامعیت بیشتر باشد کماکان مسئله‌ای باز است.

ضمایم

جدول (۳): حملات WiMax [34]

پروتکل	نتیجه	اندازه	مدت زمان	ریسک	دشواری	هزینه	خطر	حمله	گروه
802.16-'09	DoS	بزرگ	بلند	بالا	آسان	گران	زیاد	RNG-RSP DoS	آراستن
802.16-'09	اذیت	متوسط	بلند	بالا	قابل‌حل	گران	کم	RNG-RSP مخدوش	
802.16-'09	اذیت	متوسط	بلند	متوسط	آسان	گران	کم	RNG-RSP شکنجه آبی	
802.16-'09	اذیت	متوسط	بلند	متوسط	سخت	گران	کم	RNG-RSP DoS	
802.16-'09	DoS	بزرگ	بلند	پایین	آسان	ارزان	زیاد	RNG-RSP مخدوش	
802.16-'09	اذیت	متوسط	بلند	متوسط	قابل‌حل	گران	کم	MOB_ASC-REP DoS	
802.16-'09	اذیت	متوسط	کوتاه	بالا	آسان	گران	کم	MOB_TRF-IND شکنجه آبی	ذخیره‌ی انرژی
802.16-'09	اذیت	بزرگ	کوتاه	پایین	سخت	ارزان	کم	UL DoS در خواب	
802.16-'09	DoS	بزرگ	بلند	پایین	آسان	ارزان	زیاد	LU DDoS امن	
802.16-'09	اذیت	کوچک	بلند	بالا	سخت	گران	کم	MOB_NBR-ADV مخدوش	دگرسپاری
802.16-'09	اذیت	کوچک	بلند	بالا	سخت	گران	کم	MOB_NBR-ADV DoS	
802.16-'04	شنود	کوچک	بلند	کم	سخت	گران	کم	SBC-REQ مخدوش	کنترلی
802.16-'09	اذیت	بزرگ	بلند	متوسط	قابل‌حل	گران	متوسط	FPC مخدوش	
802.16-'09	اذیت	متوسط	بلند	متوسط	آسان	گران	کم	FPC شکنجه آبی	
802.16-'09	اذیت	کوچک	کوتاه	متوسط	سخت	گران	کم	RES-CMD DoS	
802.16-'09	اذیت	متوسط	بلند	متوسط	سخت	گران	کم	DBPC-REQ DoS	
–	–	–	–	–	–	–	–	برگ‌برگ‌سازی	مکانیزم امنیتی
–	–	–	–	–	–	–	–	بازاجرای AUTH-REQ	
802.16-'09	DoS	متوسط	بلند	پایین	قابل‌حل	ق.م.*	متوسط	AUTH-REQ DoS	
802.16-'09	DoS	بزرگ	بلند	متوسط	آسان	گران	زیاد	PKM-RSP DoS	
–	–	–	–	–	–	–	–	بازاستفاده TEK	
802.16-'09	شنود	کوچک	کوتاه	پایین	سخت	ق.م.	کم	حمله‌ی DES CBC IV	
802.16-'09	شنود	کوچک	کوتاه	پایین	سخت	ق.م.	کم	حمله‌ی DES CBC	
802.16j	DoS	متوسط	بلند	پایین	آسان	ارزان	متوسط	DoS نوسازی GTEK	همه/چند پخشی
802.16-'09	سرقت	متوسط	بلند	پایین	آسان	ق.م.	زیاد	سرقت اطلاعات	
802.16-'09	DoS	بزرگ	بلند	متوسط	آسان	گران	زیاد	MCA-REQ DoS	

*ق.م.: قابل مدیریت

مراجع

[1] Tom Taulli, *Recent IPO Filings Show the Power of Mobile*, Jan. 2012, http://www.investorplace.com/ipo-playbook/ipo-filings-show-power-of-mobile/, as visited on 11/1/2012.

[2] Konings, B., Schaub, F., Kargl, F., Dietzel, S., *"Channel Switch and Quiet Attack: New DoS Attacks Exploiting the 802.11 Standard"*, IEEE 34th Conf. on Local Computer Networks (LCN), pp. 14-21, 2009.

[3] Bayraktaroglu, E., King, C., Liu, X., Noubir, G., Rajaraman, R., Thapa, B., *"On the Performance of IEEE 802.11 under Jamming"*, IEEE 27th Conf. on Computer Communications (INFOCOM), pp. 1265-1273, 2008.

[4] Glass, S., Muthukkumarasamy, V., *"A Study of the TKIP Cryptographic DoS Attack"*, IEEE 15th Int. Conf. on Networks (ICON), pp. 59-65, 2007.

[5] Xinyu Xing, Shakshuki, E., Benoit, D., Sheltami, T., *"Security Analysis and Authentication Improvement for IEEE 802.11i Specification"*, IEEE Global Telecommunications Conf. (GLOBECOM), pp. 1-5, 2008.

[6] Bansal, R., Tiwari, S., Bansal, D., *"Non-cryptographic methods of MAC spoof detection in wireless LAN"*, IEEE 16th Int. Conf. on Networks (ICON), pp. 1-6, 2008.

[7] Aslam, B., Islam, M.H., Khan, S.A., *"802.11 Disassociation DoS Attack and Its Solutions: A Survey"*, Proc. of the First Mobile Computing and Wireless Communication Int. Conf., pp. 221-226, 2006.

[8] Yong Sheng, Tan, K., Guanling Chen, Kotz, D., Campbell, A., *"Detecting 802.11 MAC Layer Spoofing Using Received Signal Strength"*, IEEE 27th Conf. on Computer Communications (INFOCOM), pp.1768-1776, 2008.

[9] Martinez, A., Zurutuza, U., Uribeetxeberria, R., Fernandez, M., Lizarraga, J., Serna, A., Velez, I., *"Beacon Frame Spoofing Attack Detection in IEEE 802.11 Networks"*, Third Int. Conf. on Availability, Reliability and Security (ARES), pp. 520-525, 2008.

[10] Nguyen, T.D., Nguyen, D., Tran, B.N., Vu, H., Mittal, N., *"A Lightweight Solution for Defending Against Deauthentication/Disassociation Attacks on 802.11 Networks"*, Proc. of 17th Int. Conf. on Computer Communications and Networks (ICCCN), pp. 1-6, 2008.

[11] AirJack: http://www.sourceforge.net/projects/airjack/., as visited on 11/2/2012.

[12] void11: http://www.wirelessdefence.org/Contents/Void11Main, as visited on 11/13/2012.

[13] Wenjun Gu, Zhimin Yang, Dong Xuan, Weijia Jia, Can Que, *"Null Data Frame: A Double-Edged Sword in IEEE 802.11 WLANs", IEEE Transactions on Parallel and Distributed Systems, Vol. 21, No. 7, pp. 897-910, 2010.*

[14] MADWiFi: Multiband Atheros Driver for WiFi. http://www.madwifi-project.org, as visited on 11/13/2012.

[15] Rachedi, A., Benslimane, A., *"Smart Attacks based on Control Packets Vulnerabilities with IEEE 802.11 MAC"*, Int. Wireless Communications and Mobile Computing Conf. (IWCMC), pp. 588-593, 2008.

[16] Raya, M., Aad, I., Hubaux, J.-P., El Fawal, A., *"DOMINO: Detecting MAC Layer Greedy Behavior in IEEE 802.11 Hotspots"*, IEEE Transactions on Mobile Computing, Vol. 5, No.12, pp.1691-1705, 2006.

[17] Radosavac, S., Baras, J.S., *"Application of Sequential Detection Schemes for Obtaining Performance Bounds of Greedy Users in the IEEE 802.11 MAC"*, IEEE Communications Magazine, Vol. 46, No. 2, pp.148-154, 2008.

[18] Venkatarama, A., Corbett, C., Beyah, R., *"A Wired-side Approach to MAC Misbehavior Detection"*, IEEE Int. Conf. on Communications (ICC), pp. 1-6, 2010.

[19] Lei Guang, Assi, C., Benslimane, A., *"MAC Layer Misbehavior in Wireless Networks: Challenges and Solutions"*, IEEE Wireless Communications, Vol. 15, No. 4, pp. 6-14, 2008.

[20] Guang, L., Assi, C., *"Mitigating Smart Selfish MAC Layer Misbehavior in Ad Hoc Networks"*, IEEE Int. Conf. on Wireless and Mobile Computing, Networking and Communications (WiMob), pp. 116-123, 2006.

[21] Ferreri, F., Bernaschi, M., Valcamonici, L., *"Access points Vulnerabilities to DoS Attacks in 802.11 Networks"*, IEEE Wireless Communications and Networking Conf. (WCNC), Vol. 1, pp. 634-638, 2004.

[22] Nagarajan, V., Arasan, V., Dijiang Huang, *"Using Power Hopping to Counter MAC Spoof Attacks in WLAN"*, IEEE 7th Consumer Communications and Networking Conf. (CCNC), pp. 1-5, 2010.

[23] Zhang Laishun, Zhang Minglei, Guo Yuanbo, *"A Client Puzzle Based Defense Mechanism to Resist DoS Attacks in WLAN"*, Int. Forum on Information Technology and Applications (IFITA), Vol. 3, pp. 424-427, 2010.

[24] J. Bellardo and S. Savage, *"802.11 Denial-of-Service Attacks. Real Vulnerabilities and Practical Solutions"*, Proc. of the 12th USENIX Security Symposium, pp. 15-28, 2003.

[25] Chumchu, P., Saelim, T., Sriklauy, C., *"A New MAC Address Spoofing Detection Algorithm Using PLCP Header"*, Int. Conf. on Information Networking (ICOIN), pp. 48-53, 2011.

[26] E. D Cardenas, *MAC Spoofing: An Introduction*, http://www.giac.org/practical/GSEC/, as visited on 11/2/2012.

[27] Qing Li, Trappe, W., *"Detecting Spoofing and Anomalous Traffic in Wireless Networks via Forge-Resistant Relationships"*, IEEE Transactions on Information Forensics and Security, Vol. 2, No. 4, pp. 793-808, 2007.

[28] Chandrasekaran, G., Francisco, J.-A., Ganapathy, V., Gruteser, M., Trappe, W., *"Detecting Identity Spoofs in IEEE 802.11e Wireless Networks"*, IEEE Global Telecommunications Conf. (GLOBECOM), pp. 1-6, 2009.

[29] Douglas Madory, *New Methods of Spoof Detection in 802.11b Wireless Networking*, Master of Science Thesis, Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire, 2006.

[30] Yingying Chen, Trappe, W., Martin, R.P., *"Detecting and Localizing Wireless Spoofing Attacks"*, 4th Annual IEEE Communications Society Conf. on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), pp. 193-202, 2007.

[31] Yingying Chen, Jie Yang, Trappe, W., Martin, R.P., *"Detecting and Localizing Identity-Based Attacks in Wireless and Sensor Networks"*, IEEE Transactions on Vehicular Technology, Vol. 59, No. 5, pp. 2418-2434, 2010.

[32] J. Franklin, D. McCoy, P. Tabriz, V. Neagoe, J. V. Randwyk, D. Sicker., *"Passive Data Link Layer 802.11Wireless Device Driver Fingerprinting"*, Proc. of the 15th USENIX Security Symposium, 2006.

[33] Gopinath K N, Hemant Chaskar, white paper, AirTight Networks, 2009, http://www.blog.airtightnetworks.com/802-11w-tutorial/, As visited on 11/5/2012.
[34] Kolias, C., Kambourakis, G., Gritzalis, S., "*Attacks and Countermeasures on 802.16: Analysis and Assessment* ", IEEE Communications Surverys & Tutorials, Vol. pp, No. 99, pp. 1-28, 2012.
[35] Naseer, S., Younus, M., Ahmed, A., "*Vulnerabilities Exposing IEEE 802.16e Networks To DoS Attacks: A Survey*", Ninth ACIS Int. Conf. on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD), pp. 344-349, 2008.
[36] Han, J., Alias, M.Y., Goi Bok Min, "*Potential Denial of Service Attacks in IEEE802.16e- 2005 Networks*", Ninth Int. Symp. on Communications and Information Technology (ISCIT), pp. 1207-1212, 2009.
[37] J. Hall, M. Bareau, E. Karankis, '' *Using Transceiverprints for Anomaly Based Intrusion Detection''*, Proc. of 3rd IASTED, CIIT, pp. 22-24, 2004.

زیرنویس‌ها

[1] Next Generation Network
[2] Wireless Metropolitan Area Network
[3] Jamming
[4] Medium Access Control
[5] Physical Layer Convergence Procedure
[6] Physical Medium Dependent
[7] Distributed Coordination Function
[8] Point Coordination Function
[9] Carrier Sense Multiple Access Collision Avoidance
[10] Binary Exponential Back off
[11] Contention Window
[12] Open system
[13] Shared key
[14] Message Integrity Check
[15] Personal
[16] Enterprise
[17] Pre-Shared Key
[18] Remote Authentication Dial In User Service
[19] Association request
[20] Association response
[21] Reassociation request
[22] Reassociation response
[23] Probe request
[24] Probe response
[25] Beacon
[26] Time Indication Map
[27] Disassociation
[28] Deauthentication
[29] Base Station
[30] Mobile Subscriber
[31] Convergence Sub-layer
[32] Common Part Sub-layer
[33] Security Sub-layer
[34] Uplink
[35] Time Division Multiple Access
[36] Downlink
[37] Downlink Medium Access Protocol
[38] Ranging
[39] Ranging Request
[40] Primary Key Management
[41] Denial of Service
[42] MAC spoofing
[43] Man in the Middle
[44] Beacon spoofing
[45] Virtual Jamming
[46] Intrusion Detection System
[47] Water torture
[48] Function
[49] Mobile Traffic Indicator
[50] Mobile Sleep Request
[51] Mobile Neighbor Advertisement
[52] Fast Power Control
[53] SS Basic Capability Request
[54] Downlink Burst Profile Change Request
[55] Reset Command
[56] Group Temporary Encryption Key
[57] Multicast Assignment Request
[58] Intrusion Response System
[59] Received Signal Strength
[60] Handoff
[61] Sequence Control
[62] False Positive
[63] Frame type
[64] QoS Priority
[65] Profiling
[66] Fingerprinting

برآورد امنیت سیستم‌های مبتنی بر RF TAG با کاربرد در حوزه مدیریت ترافیک شهری

سید ابراهیم امام‌جمعه[۱]، سید وحید ازهری[۲]

[۱] دانشجوی مقطع کارشناسی ارشد (گرایش شبکه‌های کامپیوتری)
Emamjomeh@Comp.iust.ac.ir

[۲] استاد راهنما
Azharivs@iust.ac.ir

چکیده

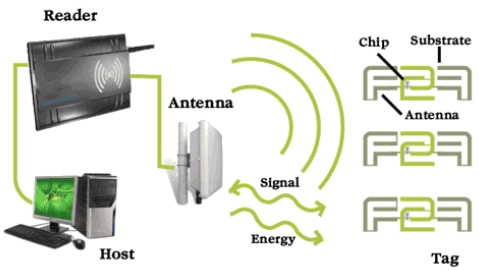
تکنولوژی شناسایی رادیویی یا به اختصار RFID[۱] نیز همانند سایر تکنولوژی‌های پرکاربرد امروزی در صنعت، نیازمند تأمین و تضمین سطوح امنیتی مشخصی خواهد بود. ازآنجائیکه هریک از انواع برچسب‌های رادیویی دارای ویژگی‌های مشخصی هستند، بسته به اینکه برای پیاده‌سازی سیستم مدیریت ترافیک شهری با تمرکز بر روی جرایم رانندگی از کدامیک از انواع برچسب‌های رادیویی استفاده می‌شود، سطح امنیتی ارائه شده نیز متفاوت خواهد بود. بنابراین لازمه ارائه راه‌کارهای امنیتی برای حل چنین چالش‌هایی، شناخت دقیق نقاط ضعف و قوت سیستم است تا بتوانیم درنهایت سیستمی با هدف ثبت شناسه وسایل نقلیه عبوری در شرایط مختلف سرعت حرکت و ازدحام ترافیکی مبتنی‌بر تکنولوژی RFID، با سطح امنیتی مناسب به منظور مدیریت ترافیک شهری پیاده‌سازی کنیم. هدف اصلی این مقاله نیز بررسی میزان آسیب‌پذیری تکنولوژی RFID در کاربرد مدیریت ترافیک شهری نسبت به برخی از مهم‌ترین تهدیدات امنیتی، و همچنین ارائه راه‌کارهایی به منظور افزایش سطح امنیتی چنین سیستمی است.

کلمات کلیدی

تکنولوژی RFID، برچسب، برچسب‌خوان، تهدید امنیتی، احرازهویت، مدیریت ترافیک شهری، راه‌کارهای امنیتی.

۱- مقدمه

تکنولوژی RFID به اشیاء یک شناسه رادیویی اختصاص می‌دهد. این شناسه درون یک برچسب رادیویی یا اصطلاحاً RF Tag گنجانده شده است. دستگاه برچسب‌خوان[۲] با استفاده از امواج رادیویی امکان شناسایی این اشیاء را فراهم می‌نماید. شکل ۱ نمای کلی یک سیستم RFID را نمایش می‌دهد.



شکل ۱ : نمای کلی یک سیستم RFID [۳۹]

تکنولوژی RFID کاربردهای فراوانی در حوزه‌هایی چون پزشکی و بهداشت، کنترل دسترسی، پرداخت حمل‌ونقل، سیستم‌های اموال اداری، گذرنامه، مدیریت ترافیک و حمل‌ونقل داراست. یکی از مهم‌ترین چالش‌های استفاده از برچسب‌های رادیویی، تضمین امنیت و اختفای اطلاعات است. ازآنجائیکه هریک از انواع برچسب‌های رادیویی دارای ویژگی‌های خاص خودشان هستند، تهدیدات امنیتی نیز به مختصات هریک از انواع برچسب‌های رادیویی مربوط می‌شود.

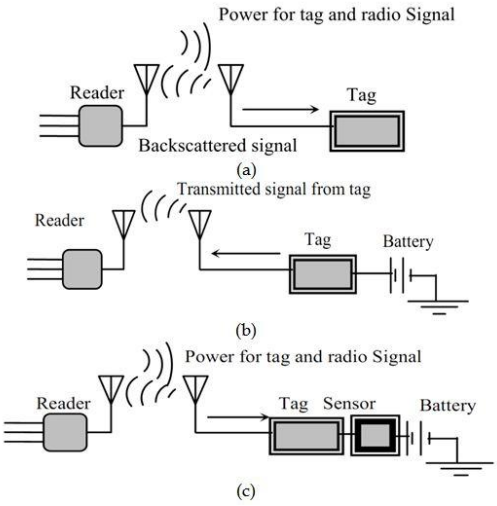
در این سمینار قصد داریم با انواع تهدیدات امنیتی در حوزه تکنولوژی RFID آشنا شده و در نهایت به بررسی این تهدیدات در کاربرد مدیریت ترافیک شهری بپردازیم. سپس برخی از راه‌کارهای ارتقای سطح امنیتی چنین سیستمی را بررسی خواهیم نمود.

این سمینار از ۷ بخش اصلی تشکیل شده است. در بخش ۲ با ویژگی‌های اصلی انواع برچسب‌های رادیویی آشنا می‌شویم. در بخش ۳ دسته‌بندی از تهدیدات امنیتی تکنولوژی RFID ارایه می‌شود. در

بخش ۴ بطور مختصر با تکنیک‌های احرازهویت در سیستم‌های RFID آشنا می‌شویم. در بخش ۵ بطور اختصاصی تهدیدات امنیتی در حوزه مدیریت ترافیک شهری بررسی شده، برای هریک از آنها راه-کارهای مقابله بیان خواهد شد. در بخش ۶ نیز محدودیت‌های امنیتی مختص هریک از انواع برچسب‌های رادیویی به همراه راه‌کارهای امنیتی مربوطه بیان می‌شود. درنهایت در بخش ۷ نتیجه‌گیری کلی از چالش-های امنیتی مرتبط با سیستم مدیریت ترافیک شهری مبتنی‌بر انواع برچسب‌های رادیویی ارائه می‌شود.

۲- معرفی انواع برچسب‌های رادیویی[۳]

بطورکلی می‌توان برچسب‌های رادیویی را به ۳ دسته کنش‌پذیر[۴]، نیمه-کنش‌پذیر[۵] و کنش‌گر[۶] تقسیم کرد [۲۳،۲۶]. البته نوع دیگری از برچسب‌های رادیویی که در بخش RF از تکنولوژی UWB استفاده می‌کنند نیز وجود دارد که در گروه برچسب‌های فوق باند وسیع قرار می‌گیرند [۱۹،۲۰،۳۷]. ازآنجائیکه برچسب‌های فوق باند وسیع از نظر امنیت لایه فیزیکی نسبت به سایر برچسب‌ها متفاوت هستند، آنها را در دسته جداگانه‌ای بررسی می‌کنیم. در ادامه با مهم‌ترین ویژگی هریک از آنها آشنا می‌شویم. شکل ۲ نحوه ارتباط بین برچسب‌های رادیویی مختلف و برچسب‌خوان را نشان می‌دهد.



شکل ۲ : نحوه ارتباط بین برچسب‌های رادیویی و برچسب‌خوان [۴۰]

۲-۱- برچسب‌های کنش‌پذیر

این نوع برچسب نسبت به سایرین اندازه کوچکتری داشته و وزن آن نیز کمتر است. همچنین به دلیل وجود سادگی در ساختار برچسب و عدم استفاده از منبع انرژی (باتری)، قیمت آن نیز بسیار پایین است. دو مورد از مهم‌ترین محدودیت‌های این نوع از برچسب‌ها عبارتنداز: کوچکی برد مفید خواندن برچسب (بطور میانگین حداکثر ۱۰ متر) و عدم امکان انجام محاسبات سنگین همچون تکنیک‌های احراز هویت و رمزنگاری‌های پیچیده. قسمت (a) از شکل ۲ نحوه ارتباط یک برچسب کنش‌پذیر با برچسب‌خوان را نشان می‌دهد.

۲-۲- برچسب‌های نیمه-کنش‌پذیر

این نوع برچسب‌ها به دلیل استفاده از منبع انرژی، نسبت به برچسب-های کنش‌پذیر اندازه بزرگتری داشته و وزن آنها نیز بیشتر است. اما از طرفی امکان انجام محاسبات سنگین نیز در آنها وجود دارد. برچسب-های نیمه-کنش‌پذیر همانند برچسب‌های کنش‌پذیر برای ارسال داده به سمت برچسب‌خوان از همان انرژی سیگنال ارسالی توسط برچسب-خوان استفاده می‌کنند، به همین دلیل از نظر برد خواندن تقریباً شبیه به برچسب‌های کنش‌پذیر بوده و تنها چند متر از آنها بیشتر است. درواقع این نوع برچسب‌ها از منبع انرژی تنها برای انجام محاسبات داخلی خودشان استفاده می‌کنند. قسمت (c) از شکل ۲ نحوه ارتباط یک برچسب نیمه-کنش‌پذیر با برچسب‌خوان را نشان می‌دهد.

۲-۳- برچسب‌های کنش‌گر

این نوع برچسب‌ها همانند برچسب‌های نیمه-کنش‌پذیر بوده اما نسبت به آنها دو مزیت اصلی دارند: استفاده از منبع انرژی قوی‌تر و توانایی آنها در استفاده از منبع انرژی برای ارسال داده‌ها به سمت برچسب‌خوان. به همین دلیل می‌توان محدوده خواندن را حتی به بیشتر از ۱۰۰ متر نیز رساند و بدین ترتیب قابلیت اطمینان ارتباط بیسیم بین برچسب و برچسب‌خوان را افزایش داد. برچسب‌های کنش‌گر نسبت به برچسب‌های کنش‌پذیر اندازه بزرگتری داشته و وزن آنها نیز بیشتر است؛ ولی از نظر اندازه و وزن معمولاً شبیه به برچسب-های نیمه-کنش‌پذیر هستند. قسمت (b) از شکل ۲ نحوه ارتباط یک برچسب کنش‌گر با برچسب‌خوان را نشان می‌دهد.

۲-۴- برچسب‌های فوق باند وسیع[۷]

درحالت معمول در لایه فیزیکی هر سه نوع برچسب معرفی شده، از باند فرکانسی باریک در یکی از فرکانس‌های LF, HF, VHF, UHF, Microwave استفاده می‌شود که همین مسئله ممکن است منجربه ایجاد حساسیت نسبت به انواع تداخلات رادیویی شود [۶،۱۹،۲۲،۳۷،۳۸]. درحالیکه با استفاده از تکنیک UWB می‌توان داده‌ها را با توان ارسالی کمتر و در بازه فرکانسی بزرگتری ارسال کرد. بنابراین ارتباط بیسیم بین برچسب و برچسب‌خوان نسبت به انواع تداخلات و برخی تهدیدات امنیتی ایمن می‌شود. نکته مهم اینجاست که تکنولوژی UWB تنها به لایه فیزیکی برچسب‌های رادیویی اشاره می‌کند (امن کردن لایه فیزیکی برچسب‌های رادیویی) [۶،۱۹،۲۰].

شکل ۳ : فرکانس مورداستفاده توسط برچسب‌های رادیویی [۴۱]

۳- تهدیدات امنیتی در حوزه تکنولوژی RFID

از یک دیدگاه کلی ارتباطات موجود در یک سیستم مبتنی بر تکنولوژی RFID را می‌توان به دو بخش تقسیم نمود. بخش اول مربوط به ارتباط بیسیم مابین برچسب و برچسب‌خوان بوده و چنین ارتباطی مبتنی بر فرکانس رادیویی و بصورت بیسیم می‌باشد[۸]. بخش دوم مربوط به ارتباط بین برچسب‌خوان و سیستم مرکزی بوده و چنین ارتباطی مبتنی بر IP و بصورت سیمی می‌باشد[۹] [۲۶،۳۰،۳۵]. هریک از این دو بخش ارتباطی در سیستم دارای مسائل امنیتی خاص خودشان هستند. هدف این سمینار بررسی تهدیدات امنیتی مرتبط با ارتباط بیسیم بین برچسب‌ها و برچسب‌خوان است. بطورکلی می‌توان انواع تهدیدات امنیتی مرتبط با کانال بیسیم بین برچسب و برچسب‌خوان را در ۸ گروه اصلی قرار داد. در ادامه هریک از این تهدیدات امنیتی و حملات موجود در هر گروه را بطور مختصر شرح می‌دهیم.

۳-۱- حملات فیزیکی[۱۰]

تمرکز اصلی حملاتی که در این گروه قرار می‌گیرند، تغییر داده‌های موجود بر روی برچسب است [۱۰،۱۳]. گروه حملات فیزیکی را نباید با حملات لایه فیزیکی اشتباه گرفت؛ چراکه اکثر حملات لایه فیزیکی در گروه حملات اختلال در سرویس قرار می‌گیرند که در بخش بعدی راجع‌به آنها نیز صحبت خواهیم کرد. در ادامه به یکی از مهم‌ترین حملات فیزیکی اشاره می‌کنیم [۱۰].

۳-۱-۱- اصلاح برچسب[۱۱]

هدف اصلی در این حمله تغییر (اصلاح) داده‌های نوشته شده روی برچسب است. بدین منظور می‌توان برچسب را مجبور به اعمال تغییراتی در حافظه‌اش کرده و داده‌های آن را تغییر داد [۲،۱۰،۱۴،۲۶،۳۲،۳۷]. چنانچه فرد متخاصم پس از دسترسی فیزیکی به برچسب، با علم به ساختار داخلی برچسب آن را مجدداً و مطابق با اهداف خود برنامه‌ریزی کند، به چنین حمله‌ای اصطلاحاً برنامه‌ریزی مجدد برچسب[۱۲] گفته می‌شود [۲،۳۲].

۳-۲- حملات اختلال در سرویس[۱۳]

حملات اختلال در سرویس به حملاتی اشاره می‌کنند که به هرنحوی منجربه ایجاد اختلال و یا قطع سرویس‌دهی سیستم موردنظر شوند. معمولاً انجام این نوع حملات آسان بوده و مقابله با آن بسیار دشوار است [۴،۱۲،۱۳،۲۶،۳۰]. انواع مختلفی از حملات در این گروه قرار می‌گیرند که در ادامه آنها را بطور مختصر شرح می‌دهیم.

۳-۲-۱- برداشتن برچسب[۱۴]

در این حملات فرد متخاصم برچسب را از دستگاه یا شیء متصل به آن جدا کرده و مانع از حضور برچسب در سیستم شود [۱۰،۲۶].

۳-۲-۲- دستکاری فیزیکی[۱۵]

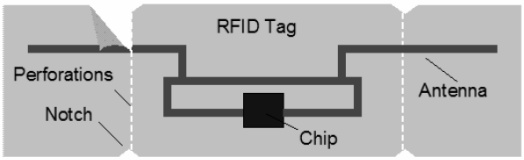
در این حملات فرد متخاصم سعی می‌کند با تحت تأثیر قرار دادن برچسب یا برچسب‌خوان در شرایط دمای بالا/پایین، رطوبت بیش ازحد، ضربه و ... باعث ایجاد نقص در تجهیزات مذکور شود [۲،۱۰،۲۶].

۳-۲-۳- سوزاندن[۱۶]

در این حملات فرد متخاصم سعی می‌کند با نزدیک شدن به برچسب و ارسال یک پالس کوتاه ولتاژ-بالا، منجربه سوزاندن و از کار انداختن برچسب شود. ذکر این نکته ضروری است که اکثراً برچسب‌های کنش-پذیر نسبت به این نوع حمله آسیب‌پذیر هستند [۴،۷،۲۲].

۳-۲-۴- برش (چیدن) برچسب[۱۷]

در این حملات فرد متخاصم سعی می‌کند با جدا کردن آنتن برچسب از چیپ متصل به آن مانع کارکرد صحیح برچسب شود. البته اکثر برچسب‌ها قادرند حتی بدون آنتن نیز در فواصل کمتر از ۱۰ سانتی‌متر کار کنند [۱،۱۵،۲۲،۲۹].



شکل (۴) : یک برچسب برش خورده [۱۵]

۳-۲-۵- دستور غیرفعال سازی[۱۸]

کلیه برچسب‌های تحت استاندارد EPC از دستوری تحت عنوان Kill پشتیبانی می‌کنند که برچسب در صورت دریافت چنین دستوری از طرف برچسب‌خوان، خودش را برای همیشه غیرفعال می‌کند. بنابراین با سوء استفاده از این دستور می‌توان برچسب‌ها را برای همیشه غیرفعال کرد [۱،۴،۲۲،۲۷،۲۸].

۳-۲-۶- پارازیت (مسدود کردن)[۱۹]

در این حملات فرد متخاصم سعی می‌کند با ارسال سیگنال پارازیت در کانال بیسیم، منجربه ایجاد اختلال در عملکرد صحیح برچسب یا برچسب‌خوان شود [۲،۶،۷،۱۴،۲۲،۲۶،۲۸،۳۰].

۳-۲-۷- تصادم[۲۰]

در این حملات فرد متخاصم می‌تواند به هرنحوی (مثلاً ارسال بیش از یک پاسخ به درخواست ارسالی از برچسب‌خوان)، منجربه ایجاد تصادم در کانال ارتباطی بیسیم بین برچسب و برچسب‌خوان شده و ارتباط صحیح آنها را مختل کند [۲۶،۲۸،۳۰،۳۸].

۳-۲-۸- سوءاستفاده از منابع پروتکل[۲۱]

در این حملات فرد متخاصم سعی می‌کند با سوءاستفاده از برخی محدودیت‌های موجود در پروتکل‌های مورد استفاده در تجهیزات، منجربه ایجاد شرایطی خاص شده و با ایجاد تغییر در مقادیر برخی پارامترها باعث ایجاد اختلال درکارکرد صحیح سیستم شود [۲,۳۲]. در اینجا منظور از منابع پروتکل، متدها و پارامترهای مورداستفاده در هریک از آنها است.

۳-۲-۹- آسیب رساندن به برچسب‌خوان[۲۲]

این نوع حمله مختص برچسب‌خوان بوده و به مواردی مثل جدا کردن، جابجایی و یا آسیب رساندن به برچسب‌خوان اشاره می‌کند [۳۲].

۳-۲-۱۰- تخلیه باتری[۲۳]

این نوع حمله، خاص دو نوع برچسب نیمه-کنش‌پذیر و کنش‌گر است. در این نوع حملات فرد متخاصم سعی می‌کند با تحریک برچسب، منجربه انجام محاسبات داخلی ناخواسته در برچسب شده تا بخشی از باتری برچسب مصرف شود تا در نهایت با تکرار مکرر چنین حمله‌ای، باتری برچسب به مرور زمان به اتمام رسیده و به دلیل وابستگی کارکرد برچسب به باتری، برچسب از کار می‌افتد [۱۷].

۳-۳- حملات جعل هویت[۲۴]

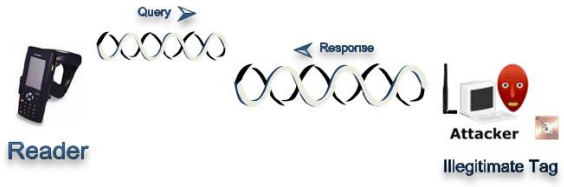
همانطورکه از نام چنین حملاتی مشخص است، در این نوع حملات فرد متخاصم سعی می‌کند خودش را جای یک موجودیت مجاز دیگر در سیستم معرفی کرده و با هویتی غیر از هویت واقعی خودش در سیستم حضور می‌یابد. به هرحال پیش شرط وقوع چنین حملاتی، اطلاع داشتن از یک شناسه مجاز در سیستم است تا بتوان به واسطه جعل چنین شناسه‌ای سیستم را دچار اشتباه کرد [۲, ۱۰,۱۲,۱۴,۲۵,۳۰,۳۲,۳۵,۳۸] . انواع مختلفی از حملات در این گروه قرار می‌گیرند که در ادامه آنها را بطور مختصر شرح می‌دهیم.

۳-۳-۱- تعویض برچسب[۲۵]

در این حملات فرد متخاصم سعی می‌کند با جابجایی برچسب‌ها بین موجودیت‌های سیستم، هویت واقعی خودش را پشت هویت جعل شده و البته مجاز در سیستم پنهان کند [۲,۲۶,۳۰,۳۲].

۳-۳-۲- سوءاستفاده از فاصله[۲۶]

در این نوع حملات فرد متخاصم سعی می‌کند اندکی زودتر و قبل از دریافت درخواست (چالش) از طرف برچسب‌خوان، پاسخ خود را ارسال کند. بدین ترتیب برچسب‌خوان اینطور تصور می‌کند که برچسب مذکور در فاصله‌ای نزدیک و مجاز قرار دارد و نسبت به چنین فاصله اشتباهی متقاعد می‌شود [۲,۸,۳۲].



شکل (۵) : نحوه وقوع حمله سوءاستفاده از فاصله

۳-۳-۳- جعل شناسه برچسب[۲۷]

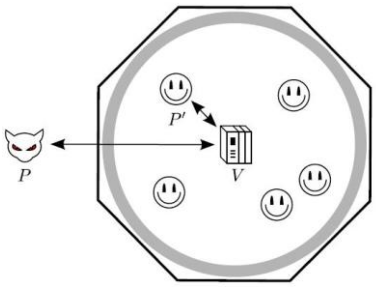
در این حملات فرد متخاصم سعی می‌کند شناسه و یا هرگونه اطلاعات مربوط به یک برچسب مجاز در سیستم را جعل کرده و به نوعی اطلاعات مربوط به یک موجودیت مجاز در سیستم را تحویل برچسب-خوان بدهد [۲,۲۱,۲۲,۲۸,۳۰,۳۱,۳۲].

۳-۳-۴- کپی‌برداری برچسب[۲۸]

در این نوع حملات فرد متخاصم سعی می‌کند به کمک تکنیک‌هایی مثل مهندسی معکوس، یک کپی فیزیکی کاملاً مشابه از یک برچسب مجاز در سیستم ایجاد نموده و کلیه اطلاعات جعل شده را نیز بر روی این برچسب فیزیکی جعلی کپی می‌کند [۲,۲۶,۲۸,۳۰,۳۱,۳۲].

۳-۳-۵- سرقت فاصله[۲۹]

در این نوع حملات فرد متخاصم سعی می‌کند با سوءاستفاده از فاز اندازه‌گیری فاصله موجود در پروتکل‌های حد فاصله[۳۰]، برچسب‌خوان را نسبت به فاصله اشتباه بین یک موجودیت غیرمجاز تا برچسب‌خوان متقاعد کند [۸,۲۱]. حمله دیگری تحت عنوان سرقت مکان[۳۱] نیز وجود دارد که حالت پیچیده‌تری داشته و در آن فرد متخاصم سعی می‌کند از اطلاعات چندین برچسب مجاز در محدوده برچسب‌خوان سوءاستفاده کرده و برچسب‌خوان را نسبت به مکان اشتباه برچسب غیرمجاز متقاعد کند [۸].



شکل (۶) : سناریوی واقعی حمله سرقت فاصله؛ P برچسب غیرمجاز، V برچسب‌خوان، 'P یکی از برچسب‌های مجاز [۸]

۳-۴- حملات بازپخش[۳۲]

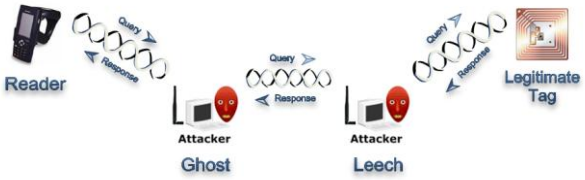
حملات گروه بازپخش حملاتی هستند که عنصر اصلی در آنها وجود یک شخص ثالث (فرد متخاصم) است. در این نوع حملات فرد حمله-کننده سعی می‌کند با قرار گرفتن میان برچسب و برچسب‌خوان، اطلاعات ردوبدل شده بین آنها را دریافت نموده و آنها را مجدداً و یا حتی پس از اعمال برخی تغییرات ارسال نماید [۲،۸،۱۰،۲۲،۳۰،۳۱،۳۸]. انواع مختلفی از حملات در این گروه قرار می‌گیرند که در ادامه آنها را بطور مختصر شرح می‌دهیم.

۳-۴-۱- مرد میانی[۳۳]

در این نوع حملات فرد متخاصم مابین برچسب‌ها و برچسب‌خوان قرار گرفته و اطلاعات ردوبدل شده بین آنها را دریافت می‌کند. اگر فرد متخاصم تنها داده‌های دریافتی را بلافاصله برای موجودیت دیگری ارسال کند، این حمله حالت غیرفعال[۳۴] خواهد داشت. ولی اگر داده‌های دریافتی دچار برخی تغییرات شده و سپس ارسال شوند، حمله از نوع فعال[۳۵] است [۵،۲۱،۲۲،۳۱،۳۲].

۳-۴-۲- فریب مافیا[۳۶]

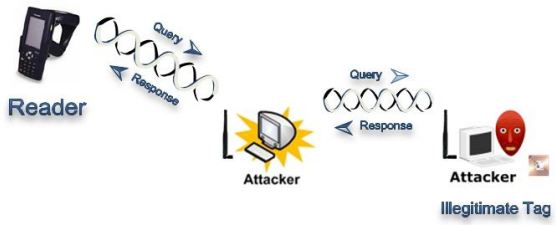
در این نوع از حملات فرد متخاصم با قرار دادن تجهیزات خود مابین برچسب مجاز و برچسب‌خوان، سعی می‌کند پیام‌های ردوبدل شده بین آنها را بدون اعمال هیچگونه تغییری رِله (بازپخش) کند. تفاوت اصلی این حمله با نوع غیرفعال حمله مرد میانی این است که در حمله فریب مافیا محدوده حمله افزایش می‌یابد؛ چراکه فرد متخاصم از ۲ دستگاه برای رِله کردن پیام‌ها استفاده می‌کند (یکی در نزدیکی برچسب و دیگری در نزدیکی برچسب‌خوان) [۲،۸،۲۱،۲۲،۲۹،۳۲].



شکل (۷) : نحوه وقوع حمله فریب مافیا

۳-۴-۳- فریب تروریست[۳۷]

حمله فریب تروریست تاحدودی شبیه به حمله فریب مافیا می‌باشد، ولی این دو حمله دارای ۲ تفاوت اصلی نیز هستند. تفاوت نخست آنکه در حمله فریب مافیا تنها با یک فرد حمله‌کننده سروکار داریم، درحالیکه حمله فریب تروریست حاصل همکاری ۲ فرد حمله‌کننده است. درواقع در حمله فریب تروریست یکی از افراد متخاصم در نزدیکی برچسب‌خوان قرار گرفته و دیگری در مکان دیگری قرار دارد. تفاوت دوم این دو حمله آن است که در حمله فریب مافیا اطلاعات یک برچسب مجاز رِله می‌شود، درحالیکه در حمله فریب تروریست این اطلاعات متعلق به یک برچسب غیرمجاز است [۲،۸،۲۱،۳۲].



شکل (۸) : نحوه وقوع حمله فریب تروریست

۳-۴-۴- ارسال مجدد[۳۸]

این حمله تاحدود زیادی شبیه به حمله مرد میانی است، با این تفاوت که در حمله ارسال مجدد پس از دریافت اطلاعات ردوبدل شده بین برچسب و برچسب‌خوان، این اطلاعات بصورت دست‌نخورده و یا حتی تغییر یافته، در زمان دیگری برای طرف دیگر ارتباط ارسال می‌شود تا بتوان بدین وسیله موجودیت‌های سیستم را با اینکه عملاً در لحظه وقوع حمله هیچگونه اطلاعاتی از طرف مقابل در ارتباط ارسال نشده است، فریب داد [۲،۵،۲۲،۲۸،۳۲،۳۵،۳۸].

۳-۵- حملات کانال مجاور[۳۹]

بطورکلی به هر نوع حمله‌ای که مبتنی‌بر اطلاعات بدست آمده از پیاده‌سازی فیزیکی یک سیستم باشد، حملات کانال مجاور گفته می-شود. درواقع برای وقوع موفقیت‌آمیز چنین حملاتی نیازمند داشتن دانش فنی کافی از عملیات درون سیستم موردنظر و همچنین تجهیزات خاصی هستیم [۲،۱۰،۱۸،۲۱،۲۵،۲۸]. در ادامه دو نوع از مهم‌ترین حملات کانال مجاور را بطور مختصر شرح می‌دهیم.

۳-۵-۱- زمان‌سنجی[۴۰]

این نوع از حملات به تحلیل مدت زمان لازم برای انجام محاسبات داخلی برچسب‌ها (شناسایی، احرازهویت، رمزنگاری، رمزگشایی و...) و یا کل یک فاز درخواست/پاسخ اشاره می‌کنند [۱۰،۲۸].

۳-۵-۲- تحلیل مصرف توان[۴۱]

این نوع از حملات به تحلیل میزان توان مصرفی برچسب‌های دارای منبع انرژی برای انجام محاسبات داخلی خود (شناسایی، احرازهویت، رمزنگاری، رمزگشایی و...) اشاره می‌کنند [۱۰،۱۸،۲۸].

۳-۶- حملات استراق‌سمع[۴۲]

بطورکلی منظور از حملات استراق سمع، شنود غیرمجاز اطلاعات در یک سیستم است. تنها حمله‌ای که می‌توان در این گروه خاص قرار داد، حمله بو کشیدن (تحلیل ترافیک) است که در ادامه به آن اشاره می‌کنیم [۲،۵،۶،۱۴،۲۸،۳۰،۳۱].

۳-۶-۱- بو کشیدن[۴۳]

عموماً اصطلاح «بو کشیدن» به نسخه الکترونیکی «استراق‌سمع» اشاره می‌کند. در این حمله فرد متخاصم سعی می‌کند با قرار گرفتن در محدوده کاری سیستم و یا حتی دورتر از محدوده کاری سیستم (استفاده از تجهیزات گیرنده قوی‌تر)، اطلاعات ردوبدل شده بین موجودیت‌های مختلف سیستم را شنود کند. در اکثر موارد چنین حمله‌ای فاز نخست برای سایر حملات محسوب می‌شود [۵،۱۰،۱۳،۲۲،۲۴،۳۱،۳۷].

۳-۷- حملات تجسس[۴۴]

هدف اصلی این گروه از حملات کسب اطلاعات با استفاده از تجهیزاتی خاص و به شکلی فعالانه و غیرمجاز در یک سیستم است. بنابراین پیش‌نیاز چنین حمله‌ای، حملات گروه جعل هویت هستند [۲،۵،۱۳،۲۲،۲۸،۳۱]. در ادامه مهم‌ترین حمله از گروه حملات تجسس را بطور خلاصه شرح می‌دهیم.

۳-۷-۱- خواندن سطحی[۴۵]

در این نوع حملات فرد متخاصم سعی می‌کند تا بصورت غیرمجاز خودش را برای برچسب‌ها به عنوان یک برچسب‌خوان مجاز در سیستم نشان داده و برای آنها درخواست ارسال می‌کند تا بتواند با دریافت پاسخ از برچسب‌ها به اطلاعات خاصی دست پیدا کند. درحقیقت می‌توان این حمله را بطور خلاصه اینطور تعریف نمود: «دسترسی غیرمجاز به داده‌های ذخیره شده روی برچسب» [۱۴،۲۲،۲۶،۲۸،۲۹،۳۷].

۳-۸- تهدیدات اختفاء[۴۶]

به دلیل ماهیت تکنولوژی RFID و رسالت اصلی آن یعنی شناسایی خودکار، قطعاً این تکنولوژی در معرض برخی تهدیدات در حوزه حریم شخصی مالک شناسه (برچسب) می‌باشد [۲،۱۲،۱۴،۲۴،۳۲]. در این بخش قصد داریم ۲ مورد از مهم‌ترین تهدیدات مرتبط با اختفاء اطلاعات را بطور مختصر معرفی کنیم.

۳-۸-۱- ردگیری غیرقانونی[۴۷]

این نوع از تهدیدات به این موضوع اشاره دارند که به دلیل بیسیم بودن کانال ارتباطی بین برچسب و برچسب‌خوان، درصورت امن نبودن چنین کانال بیسیمی به هرحال امکان شنود و ردگیری غیرقانونی افراد براساس شناسه برچسب آنها وجود دارد [۲،۱۲،۲۲،۲۸،۲۹،۳۰،۳۲].

۳-۸-۲- بمب RFID[۴۸]

تهدید امنیتی بمب RFID به این موضوع اشاره دارد که می‌توان مدار فعال‌سازی یک بمب را با اندکی تغییر به یک سیستم مبتنی‌بر RFID متصل نمود و آن را طوری برنامه‌ریزی کرد که به محض قرارگیری یک برچسب و یا حتی وجود برچسبی با شناسه خاص در محیط، فعال شده و بمب منفجر شود. حتی می‌توان این مدار را با یک برچسب مرتبط نمود تا بمب در صورت دریافت درخواست از طرف برچسب‌خوان منفجر شود [۲،۱۰،۱۴،۳۲،۳۷].

۴- تکنیک‌های احرازهویت

۴-۱- آشنایی با تکنیک‌های احرازهویت

یکی از بهترین روش‌های پیشگیری نسبت به تهدیدات امنیتی در اکثر سیستم‌ها، استفاده از تکنیک‌های احرازهویت است تا طرفین ارتباط با امنیت بیشتری با یکدیگر تعامل داشته باشند. بطورکلی متداول‌ترین تکنیک‌های احرازهویت در سیستم‌های IT عبارتند از: تکنیک‌های احرازهویت مبتنی‌بر کلید متقارن[۴۹] و تکنیک‌های احرازهویت مبتنی‌بر کلید نامتقارن[۵۰] [۲۲،۲۶،۳۳،۳۷،۳۸]. همانطورکه از نام این تکنیک‌ها مشخص است، مهم‌ترین تفاوت آنها در نوع کلیدی است که در توابع موردنظر استفاده می‌شود. بدین معناکه اگر طرفین ارتباط از کلیدی مشابه در توابع احرازهویت استفاده کنند، روش مذکور در گروه تکنیک‌های احراز هویت مبتنی‌بر کلید متقارن قرار گرفته و اگر طرفین ارتباط از کلیدهای نامتقارن (کلید عمومی/خصوصی) استفاده کنند، تکنیک مورداستفاده در گروه تکنیک‌های احرازهویت نامتقارن قرار می‌گیرد که به آن اصطلاحاً روش مبتنی‌بر کلید عمومی[۵۱] هم گفته می‌شود [۱،۲۲،۲۴،۲۷،۳۳،۳۴].

۴-۱-۱- شناسایی[۵۲] و احراز هویت

توجه به نکته‌ای ظریف در رابطه با مفهوم احرازهویت مهم است. قبل از فاز احرازهویت لازم است تا موجودیت موردنظر از نظر هویت واقعی‌اش شناسایی شود که به این فاز اصطلاحاً فاز شناسایی هویت گفته می‌شود. درحقیقت فاز شناسایی هویت بیانگر مرحله‌ای است که به دنبال کلید موردنظر برای شناسه خاصی بگردیم تا بتوانیم از این کلید منحصربه‌فرد در فاز بعدی یعنی فاز احرازهویت استفاده کنیم. درصورت شناسایی موجودیتی خاص و پیدا شدن کلید متناظر با آن، می‌توانیم به سراغ فاز احرازهویت برویم. لذا هریک از این ۲ فاز از نظر مصرف منابع و زمان، سربار خاص خودشان را خواهند داشت [۳۳،۳۴،۳۸].

۴-۱-۲- کلاس‌های مختلف احراز هویت

متأسفانه امروزه به دلیل وجود برخی محدودیت‌ها در برچسب‌های رادیویی، اکثر چنین سیستم‌هایی از تکنیک‌های احرازهویت امنی استفاده نمی‌کنند. برمبنای هزینه محاسباتی و حجم عملیاتی که برچسب‌های رادیویی پشتیبانی می‌کنند، می‌توان تکنیک‌های

احرازهویت در تکنولوژی RFID را به ۴ کلاس مختلف تقسیم نمود [۳۵]؛ در ادامه این ۴ کلاس را معرفی می‌کنیم.

کلاس کاملاً تکامل‌یافته[۵۳]

این کلاس بیانگر تکنیک‌هایی است که نیازمند پشتیبانی از توابع رمزنگاری مرسوم، توابع درهم‌ریزی[۵۴] و یا حتی الگوریتم‌های رمزنگاری کلید عمومی هستند. درصورت استفاده از الگوریتم‌های مبتنی‌بر کلید عمومی، می‌توان از مفهوم امضای دیجیتال[۵۵] نیز برای هرچه امن‌تر شدن عملیات احرازهویت استفاده نمود [۲۶،۳۵].

کلاس ساده[۵۶]

این کلاس بیانگر تکنیک‌هایی است که بر روی برچسب‌ها تنها از یک تابع تولیدکننده اعداد تصادفی ساختگی[۵۷] و یا توابع درهم‌ریزی یکطرفه[۵۸] استفاده می‌کنند [۳۵].

کلاس سبک[۵۹]

این کلاس بیانگر تکنیک‌هایی است که نیازمند پشتیبانی از یک تابع تولیدکننده اعداد تصادفی ساختگی و یا توابع ساده‌ای مثل CRC Checksum هستند [۳۵].

کلاس فوق سبک[۶۰]

این کلاس بیانگر تکنیک‌هایی است که تنها نیازمند انجام عملیات بیتی ساده هستند (عملیاتی مثل XOR،AND،OR،Rotate و غیره) [۳۵].

ازآنجائیکه هریک از برچسب‌های کنش‌پذیر، نیمه-کنش‌پذیر و کنش‌گر دارای ویژگی‌های خاص خودشان هستند، لذا می‌توان اینطور بیان کرد که معمولاً بر روی برچسب‌های کنش‌پذیر از کلاس سبک و یا فوق سبک احرازهویت استفاده می‌شود، درحالیکه بر روی برچسب-های نیمه-کنش‌پذیر و کنش‌گر به دلیل وجود منبع انرژی مستقل و همچنین وجود حافظه و تعداد گیت‌های بیشتر، امکان استفاده از کلاس کاملاً تکامل‌یافته و یا کلاس ساده احرازهویت وجود دارد [۳۵]. به هرحال یکی از مهم‌ترین چالش‌ها در رابطه با تکنیک‌های احراز هویت و همچنین رمزنگاری کانال انتقال بین برچسب و برچسب‌خوان، پیاده‌سازی تکنیک‌های امنیتی سبک به منظور کاهش تعداد سیکل-های موردنیاز برای انجام عملیات مربوطه و همچنین کاهش توان مصرفی برچسب و برچسب‌خوان است. قطعاً استفاده از تکنیک‌های احراز هویت و رمزنگاری سبک‌تر منجربه تحمیل سربار کمتری بر روی سیستم خواهد شد. بدین ترتیب می‌توان علاوه‌بر افزایش سطح امنیتی سیستم، کارایی آن را نیز در حد معمول حفظ نمود.

۵- مسائل مرتبط با امنیت تکنولوژی RFID در کاربرد مدیریت ترافیک شهری

همانطورکه می‌دانیم تکنولوژی RFID در حوزه‌های مختلف کاربردهای زیادی دارد و به دلیل ویژگی‌های چشمگیر این تکنولوژی شاهد استفاده روزافزون از آن هستیم. یکی از کاربردهای نوظهور این تکنولوژی، مدیریت ترافیک شهری است. هدف اصلی در این کاربرد آن است که بتوانیم با استفاده از این تکنولوژی بطور خودکار وقوع برخی از جرائم رانندگی وسائل نقلیه (با تمرکز بر روی موتورسیکلت‌ها) در حوزه شهری را مدیریت کنیم. برای مثال می‌توان مواردی مثل جمع‌آوری عوارض جاده‌ای، تشخیص سرعت غیرمجاز، عبور از چراغ قرمز، ردگیری وسایل نقلیه دزدیده‌شده، طرح‌های ترافیکی و مناطق ورود ممنوع، ورود و خروج پارکینگ‌ها و مواردی از این قبیل را به کمک چنین سیستمی کنترل کرد. قطعاً پیاده‌سازی چنین سیستمی دارای مسائل غیرامنیتی و امنیتی خاص خودش می‌باشد، لذا لازم است تا با ساختار و نقاط ضعف و قوت چنین سیستمی به خوبی آشنا شویم. بدین منظور انواع تهدیدات امنیتی تکنولوژی RFID در حوزه مدیریت ترافیک شهری به همراه مهم‌ترین راه‌کارهای امنیتی مربوطه در جدول ۱ بررسی شده است. شکل ۹ بیانگر یک سیستم مبتنی‌بر تکنولوژی RFID به منظور شناسایی تردد وسایل نقلیه عبوری برای کاربرد ثبت تخلفات رانندگی است.

شکل (۹): سیستم کنترل تردد وسایل نقلیه مبتنی‌بر تکنولوژی RFID

جدول ۱ : مقایسه تهدیدات امنیتی تکنولوژی RFID و راه‌کارهای مقابله با آن در کاربرد مدیریت ترافیک شهری

گروه حملات	توضیحات	راه‌کارهای امنیتی مرتبط
فیزیکی	درصورت پیاده‌سازی سیستم با برچسب‌های کنش‌پذیر، سیستم بیشتر در معرض وقوع این نوع حملات خواهد بود [۱۰].	• حفظ بهتر امنیت فیزیکی برچسب‌ها [۱،۱۰،۲۷]. • افزایش پیچیدگی ساختار داخلی برچسب [۱]. • استفاده از تکنیک‌های احرازهویت قدرتمند [۱،۱۰،۱۲].
اختلال در سرویس	یکی از مهم‌ترین چالش‌ها در پیاده‌سازی سیستم، محل نصب برچسب‌ها بر روی وسائل نقلیه و برچسب‌خوان‌ها در معابر شهری است که باید به خوبی انجام شود. به هرحال احتمال وقوع این نوع حملات نسبتاً زیاد است [۲].	• حفظ بهتر امنیت فیزیکی برچسب‌ها [۱،۱۰،۲۷] و همچنین استفاده از حسگرها به منظور فعال‌سازی سیستم هشداردهی دستکاری برچسب و یا برچسب‌خوان [۱،۹،۱۰،۲۶]. • استفاده از برچسب‌های مقاوم در برابر دستکاری و آسیب[۶۱] [۱،۹،۱۰]. • استفاده از تکنیک UWB به عنوان لایه فیزیکی برچسب‌ها [۶،۱۹،۲۰]. • ایزوله‌سازی محدوده خواندن برچسب‌ها (مثلاً استفاده از آنتن‌های جهت‌دار) [۱۰]. • برطرف نمودن نقاط ضعف موجود در استانداردها و پروتکل‌های مورداستفاده در سیستم.
جعل هویت	انجام موفقیت‌آمیز چنین حملاتی نیازمند استفاده از تکنیک‌های خاص مثل مهندسی معکوس و همچنین آشنایی با ساختار پروتکل-های مورداستفاده در سیستم است [۸]. درواقع پیچیدگی وقوع چنین حملاتی نسبتاً بالاست.	• حفظ بهتر امنیت فیزیکی برچسب‌ها [۱،۱۰،۲۷] و همچنین استفاده از حسگرها به منظور فعال‌سازی سیستم هشداردهی دستکاری برچسب و یا برچسب‌خوان [۱،۹،۱۰،۲۶]. • استفاده از برچسب‌های مقاوم در برابر دستکاری و آسیب [۱،۹،۱۰]. • استفاده از کلمه عبور و یا اطلاعات زیست‌سنجی (افزایش سطح امنیتی احرازهویت) [۹،۱۰،۲۵،۳۰]. • کاهش محدوده خواندن برچسب‌ها (مثلاً کاهش توان ارسالی و یا استفاده از آنتن‌های جهت‌دار) [۱۰]. • رمزنگاری و ذخیره نمودن شناسه‌ها برروی برچسب‌ها و تغییر آنها در بازه‌های زمانی مناسب [۱،۲۵،۳۴]. • برقراری ارتباط و تحیل وابستگی بین اطلاعات موجود در سیستم به کمک تجهیزاتی مثل سیستم تشخیص نفوذ RFID[۶۲] [۳۵]، نگهبان RFID[۶۳] [۲،۳،۱۰،۱۱،۲۹]، پروکسی تقویت‌کننده RFID[۶۴] [۱۶] و دستگاه Prover Honeypot [۳۶] به منظور تشخیص موجودیت‌های غیرمجاز [۲،۸،۳۲]. • استفاده از تکنیک‌های قدرتمند احرازهویت (مثلاً تکنیک‌های احرازهویت دوطرفه) [۱،۱۲،۲۶،۲۷]. • امن کردن کانال ارتباط به کمک پروتکل‌های SSL/TLS [۸].
بازپخش	به دلیل بیسیم بودن کانال ارتباطی بین برچسب و برچسب‌خوان، احتمال وقوع چنین حملاتی (با فرض استفاده از تجهیزات خاص) نسبتاً زیاد است. بنابراین استفاده از برخی راه‌کارهای امنیتی تنها احتمال وقوع آنها را کاهش خواهد داد.	• استفاده از کلمه عبور و یا اطلاعات زیست‌سنجی (افزایش سطح امنیتی احرازهویت) [۹،۱۰،۲۵،۳۰]. • کاهش محدوده خواندن برچسب‌ها (مثلاً کاهش توان ارسالی و یا استفاده از آنتن‌های جهت‌دار) [۱۰]. • استفاده از تکنیک‌های قدرتمند احرازهویت (مثلاً تکنیک‌های احرازهویت دوطرفه) [۹،۲۷]. • امن کردن کانال ارتباط به کمک پروتکل‌های SSL/TLS [۸]. • منحصربه‌فرد بودن چالش‌ها (استفاده از تمبرهای زمانی[۶۵]، کلمه‌های عبور یکبارمصرف، رشته تصادفی Nonce و تطابق زمان‌سنج[۶۶]) [۱،۲،۹،۳۷]. • برآورد فاصله مجاز برچسب و برچسب‌خوان (مثلاً از روی قدرت سیگنال ارسالی برچسب‌خوان) [۸،۱۰].
کانال مجاور	احتمال وقوع چنین حملاتی در کاربرد مدیریت ترافیک شهری بسیار اندک بوده و می‌توان از چنین تهدیدی صرف‌نظر کرد.	• استفاده از تکنیک‌های مبهم برای فاز چالش- پاسخ [۲۸]. • اعمال تأخیر عمدی و تصادفی در انجام فرآیند محاسبات [۹]. • استفاده از عاملی به منظور مصرف مقدار تصادفی از توان [۹].
استراق سمع	جلوگیری از این نوع حملات می‌تواند تاحد زیادی به افزایش امنیت سیستم و پیشگیری از برخی سایر حملات کمک کند.	• رمزنگاری کانال انتقال [۹،۲۷] و استفاده از تکنیک UWB به عنوان لایه فیزیکی برچسب‌ها [۶،۱۹،۲۰] • عدم ذخیره‌سازی داده‌های مهم و غیرضروری روی برچسب [۹،۲۶] و همچنین رمزنگاری شناسه‌ها و ذخیره نمودن شناسه‌های رمزشده بر روی برچسب‌ها و تغییر آنها در بازه‌های زمانی مناسب [۱،۳۴]. • کاهش محدوده خواندن برچسب‌ها (مثلاً کاهش توان ارسالی و یا استفاده از آنتن‌های جهت‌دار) [۱۰].
تجسس	از آنجائیکه برای وقوع چنین حملاتی نیازمند استفاده از تجهیزات خاصی برای جعل هویت یک برچسب‌خوان مجاز و سپس برقراری ارتباط با برچسب‌ها هستیم، احتمال وقوع چنین حملاتی نیز نسبتاً پایین است.	• رمزنگاری و ذخیره نمودن شناسه‌ها برروی برچسب‌ها و تغییر آنها در بازه‌های زمانی مناسب [۱،۲۵،۳۴]. • کنترل بهتر دسترسی به داده‌های موجود بر روی برچسب‌ها (استفاده از تکنیک‌های رمزنگاری و احرازهویت قدرتمند) [۲۷،۳۷،۳۸]. • استفاده از دستگاه‌هایی مثل Blocker Tag [۱۴،۲۱]، نگهبان RFID [۲،۳،۱۰،۱۱،۲۹]، پروکسی تقویت-کننده RFID [۱۶]. • استفاده از دستور غیرفعال‌سازی Kill در شرایط خاص(به منظور غیرفعال کردن دائمی برچسب‌ها) [۳].
تهدیدات اختفاء	از آنجائیکه برچسب در پاسخ به درخواست برچسب‌خوان، شناسه خودش را ارسال می‌کند، در ساده‌ترین حالت هویت برچسب (وسیله نقلیه) در معرض ردگیری غیرقانونی خواهد بود.	• استفاده از تکنیک‌های قدرتمند احراز هویت و رمزنگاری کانال انتقال [۲۶،۲۷]. • رمزنگاری و ذخیره نمودن شناسه‌ها برروی برچسب‌ها و تغییر آنها در بازه‌های زمانی مناسب [۱،۲۵،۳۴]. • استفاده از دستگاه‌هایی مثل Blocker Tag [۱۴،۲۱]، نگهبان RFID [۲،۳،۱۰،۱۱،۲۹]، پروکسی تقویت-کننده RFID [۱۶]. • استفاده از راه‌کارهای فیزیکی مثل به خواب بردن برچسب[۶۷]، مسدود کردن برچسب[۶۸]، غیرفعال کردن دائمی برچسب، برش برچسب و برچسب‌زنی مجدد[۶۹] [۱،۹،۲۲،۲۴،۲۶].

۶- بررسی امنیت برچسب‌های رادیویی

در بخش ۵ به انواع تهدیدات امنیتی و راه‌کارهای مقابله با آنها در کاربرد مدیریت ترافیک شهری، بدون درنظر گرفتن نوع برچسب رادیویی مورداستفاده در سیستم اشاره شد. در این بخش قصد داریم برخی از مهم‌ترین محدودیت‌های امنیتی برچسب‌های رادیویی را معرفی کنیم، تا با چالش‌های امنیتی مرتبط با هریک از انواع آنها آشنا شویم؛ همچنین برای هریک از آنها به راه‌کارهای امنیتی خاصی اشاره می‌کنیم. درضمن آسیب‌پذیری برچسب‌های مختلف نسبت به انواع تهدیدات امنیتی در جدول ۲ (بخش ضمایم) مطرح شده است.

۶-۱- برچسب‌های کنش‌پذیر

۶-۱-۱- محدودیت‌های امنیتی و غیرامنیتی

باتوجه به کاربردهای مطرح شده در حوزه مدیریت ترافیک شهری، ممکن است استفاده از برچسب‌های کنش‌پذیر به دلیل وجود محدودیت‌های زیر انتخاب مناسبی نباشد؛ بطورخلاصه مهم‌ترین محدودیت‌ها عبارتنداز :

- قابل استفاده نبودن در برخی کاربردها به دلیل وجود محدودیت در فاصله بین برچسب و برچسب‌خوان.
- سادگی ساختار داخلی برچسب‌ها (محدودیت حافظه و مشکلات مربوط به مدیریت حافظه برچسب‌ها، عدم قابلیت افزودن حسگرها به منظور تشخیص دستکاری و آسیب رساندن به برچسب‌ها) [۲۳].
- عدم توانایی انجام محاسبات سنگین (عدم توانایی پیاده‌سازی برخی از خصیصه‌های امنیتی).
- عدم توانایی ثبت و ارسال خودکار رخدادها.
- عدم امکان استفاده از تکنیک‌های قدرتمند احراز هویت، توابع درهم‌ریزی[۷۰]، کد احرازهویت پیام[۷۱]، عملیات رمزنگاری کلید متقارن[۷۲]، رمزنگاری کلید نامتقارن[۷۳] [۲۱،۳۳،۳۸].
- عدم امکان استفاده از تکنیک‌های پیچیده برای فاز چالش-پاسخ[۷۴].
- عدم امکان استفاده از پروتکل‌هایی مثل SSL/TLS [۸].

۶-۱-۲- راه‌کارهای امنیتی

- حفظ بهتر امنیت فیزیکی برچسب‌ها (عدم امکان دسترسی فیزیکی به برچسب‌ها و برقراری اتصال فیزیکی هرچه بهتر آنها) [۱،۱۰،۲۷].
- کاهش محدوده عملیات سیستم (مثلاً ایزوله کردن محدوده عملیات، دستکاری توان ارسالی برچسب‌خوان و همچنین استفاده از آنتن‌های جهت‌دار) [۱۰].
- برطرف نمودن نقاط ضعف موجود در استانداردهای معرفی شده در برچسب‌های کنش‌پذیر.
- استفاده از تجهیزاتی مثل سیستم تشخیص نفوذ RFID، نگهبان RFID [۲،۳،۱۰،۱۱،۲۹،۳۵] ، پروکسی تقویت‌کننده RFID [۱۶]، Blocker Tag [۱۴].
- رمزنگاری شناسه‌ها و ذخیره نمودن شناسه‌های رمزشده بر روی برچسب‌ها و تغییر آنها در بازه‌های زمانی مناسب [۱،۳٤].
- استفاده از راه‌کارهای فیزیکی مانند به خواب بردن، برش و یا حتی غیرفعال کردن دائمی برچسب‌ها در شرایط خاص [۱].

باتوجه به راه‌کارهای امنیتی ارائه شده در رابطه با برچسب‌های کنش‌پذیر و همچنین وجود محدودیت‌های امنیتی بیان شده، متأسفانه امکان افزایش سطح امنیت سیستم بیشتر از حد مشخصی وجود ندارد؛ چنانچه برای پیاده‌سازی سیستم مدیریت ترافیک شهری از برچسب‌های کنش‌پذیر استفاده شود، چنین سیستمی در معرض انواع مختلفی از تهدیدات امنیتی قرار خواهد گرفت.

۶-۲- برچسب‌های نیمه-کنش‌پذیر و کنش‌گر

باتوجه به کاربردهای مطرح شده در حوزه مدیریت ترافیک شهری، به دلیل استفاده از باتری در برچسب‌های نیمه-کنش‌پذیر و کنش-گر، سیستم نسبت به سیستم‌های مبتنی بر برچسب‌های کنش‌پذیر هم از لحاظ امنیتی مقاوم شده است و هم از لحاظ محدوده خواندن شاهد افزایش فاصله هستیم. درضمن با استفاده از برچسب‌های نیمه-کنش‌پذیر و کنش‌گر، امکان استفاده از انواع مختلف حسگرها به منظور مانیتور کردن وضعیت وسیله نقلیه فراهم می‌شود. اما در کنار تمام ویژگی‌های ذکرشده، همواره مشکلات مرتبط با باتری در چنین سیستمی دیده می‌شود. به عنوان مثال عمر برچسب وابسته به عمر باتری متصل به آن است و در برخی موارد امکان شارژ مجدد و یا تعویض باتری وجود نداشته و نیازمند تعویض کامل برچسب هستیم [۲۳].

ازآنجائیکه دو نوع برچسب نیمه-کنش‌پذیر و کنش‌گر از نظر ساختار داخلی و منبع انرژی مورداستفاده در اکثر موارد شبیه به یکدیگر هستند، لذا راه‌کارهای امنیتی آنها نیز تقریباً یکسان است. تنها تفاوت این دو نوع برچسب از نظر محدودیت‌های امنیتی این است که در برچسب‌های نیمه-کنش‌پذیر به دلیل وجود محدودیت در حداکثر فاصله بین برچسب و برچسب‌خوان (معمولاً کمتر از ۱۰ یا ۱۵ متر)، انجام محاسبات با سربار زیاد در بازه زمانی کوتاه مشکل‌ساز است؛ چراکه ممکن است در این مدت برچسب از برد دستگاه برچسب‌خوان خارج شود.

۶-۲-۱- راه‌کارهای امنیتی

- حفظ بهتر امنیت فیزیکی برچسب‌ها (عدم امکان دسترسی فیزیکی به برچسب‌ها و برقراری اتصال فیزیکی هرچه بهتر آنها) [۱،۱۰،۲۷].
- کاهش محدوده عملیات سیستم (مثلاً ایزوله کردن محدوده عملیات، دستکاری توان ارسالی برچسب‌خوان و همچنین استفاده از آنتن‌های جهت‌دار) [۱۰].
- برطرف نمودن نقاط ضعف موجود در استانداردهای معرفی شده در برچسب‌های نیمه-کنش‌پذیر و کنش‌گر.
- استفاده از تجهیزاتی مثل سیستم تشخیص نفوذ و پروکسی تقویت‌کننده RFID، نگهبان RFID [۲،۳،۱۰،۱۱،۲۹]، Blocker Tag [۱۴،۲۱].
- رمزنگاری شناسه‌ها و ذخیره نمودن شناسه‌های رمزشده بر روی برچسب‌ها و تغییر آنها در بازه‌های زمانی مناسب [۱،۳٤].
- استفاده از تکنیک‌های قدرتمند احرازهویت، رمزنگاری کلید متقارن/نامتقارن و تکنیک‌های پیچیده فاز چالش-پاسخ [۱،۲۸].
- استفاده از راه‌کارهای فیزیکی مانند به خواب بردن، برش و یا حتی غیرفعال کردن دائمی برچسب‌ها در شرایط خاص.

۶-۳- برچسب‌های کنش‌پذیر و نیمه-کنش‌پذیر UWB

در برچسب‌های کنش‌پذیر و نیمه-کنش‌پذیر از نوع UWB، کانال ارتباطی از برچسب‌خوان به برچسب (کانال روبه‌جلو[۷۵]) در یکی از فرکانس‌های LF,HF,VHF,UHF,Microwave بوده و بصورت UWB نمی‌باشد، ولی کانال ارتباطی از برچسب به برچسب‌خوان (کانال روبه‌عقب[۷۶]) بصورت UWB است. چنین ارتباطی را اصطلاحاً نیمه- فوق باند وسیع[۷۷] می‌نامند؛ در چنین ارتباطی هنوز هم کانال ارتباطی از برچسب‌خوان به برچسب از نظر برخی خطرات امنیتی تهدید می‌شود [۶،۱۹،۲۰].

۶-۴- برچسب‌های کنش‌گر UWB

در برچسب‌های کنش‌گر از نوع UWB هر دو کانال روبه‌جلو و روبه-عقب بصورت UWB هستند، لذا امکان وقوع حملاتی مثل شنود غیرمجاز، پارازیت، تصادم و تخلیه باتری نیز بسیار دشوار خواهد شد؛ بنابراین امنیت نسبت به ۲ نوع برچسب قبلی در سطوح بالاتری فراهم می‌شود [۶،۱۹].

۷- نتیجه‌گیری

باتوجه به مسائل مطرح شده راجع به محدودیت‌های امنیتی و غیرامنیتی هریک از انواع برچسب‌های رادیویی، کاملاً مشخص است که به‌هرحال استفاده از برچسب‌های کنش‌پذیر در کاربرد مدیریت ترافیک شهری، بیشترین مشکلات امنیتی را به همراه خواهد داشت. این درحالیست که با استفاده از برچسب‌های نیمه-کنش‌پذیر و یا کنش‌گر، به دلیل وجود منبع انرژی مستقل برای انجام محاسبات خاص (مثل عملیات احرازهویت، رمزنگاری و رمزگشایی) سطح امنیتی سیستم افزایش چشمگیری خواهد داشت، ولی درعوض هزینه پیاده‌سازی سیستم نیز اندکی افزایش می‌یابد. درنهایت چنانچه به دنبال بالاترین سطح امنیتی (کمترین احتمال موفقیت در وقوع حملات) هستیم، بهتر است از برچسب‌های نوع UWB استفاده کنیم. به عنوان یک جمع‌بندی کلی می‌توان اینطور بیان کرد که چالش‌های اصلی استفاده از برچسب‌های کنش‌پذیر عبارتنداز: کاهش محدوده خواندن برچسب و کاهش سطح امنیتی سیستم (عدم امکان انجام محاسبات). چالش‌های اصلی استفاده از برچسب‌های نیمه-کنش‌پذیر عبارتنداز: وجود محدودیت در طول عمر برچسب (وابستگی به باتری)، هزینه نسبتاً بالای پیاده‌سازی سیستم، اندازه و وزن نسبتاً بالای برچسب (وجود دشواری در نصب برچسب‌ها، نسبت به برچسب‌های کنش‌پذیر) و محدوده خواندن نسبتاً پایین (تنها چند متر بیشتر از برچسب‌های کنش‌پذیر). چالش‌های اصلی استفاده از برچسب‌های کنش‌گر عبارتنداز: وجود محدودیت در طول عمر برچسب (وابستگی به باتری)، هزینه بالای پیاده‌سازی سیستم، اندازه و وزن بالای برچسب (وجود دشواری در نصب برچسب‌ها، نسبت به برچسب‌های کنش‌پذیر). بطورکلی از نقطه نظر امنیت، برچسب‌های کنش‌پذیر ساده‌ترین و ضعیف‌ترین نوع برچسب‌های رادیویی، و برچسب‌های کنش‌گر UWB پیچیده-ترین و قوی‌ترین نوع برچسب‌های رادیویی محسوب می‌شوند.

ضمایم

جدول ۲ : مقایسه آسیب‌پذیری برچسب‌های مختلف نسبت به حملات در تکنولوژی RFID

نوع برچسب / تهدید امنیتی	Passive	Semi-Passive	Active	UWB (Passive)	UWB (Semi-Passive)	UWB (Active)
برداشتن برچسب	!	!	!	!	!	!
تعویض برچسب	!	!	!	!	!	!
سوزاندن برچسب	!	~	~	!	~	~
برش (چیدن) برچسب	!	!	!	!	!	!
دستکاری فیزیکی	!	~	~	!	~	~
دستور غیرفعال‌سازی دائمی برچسب	!	~	~	~	~	~
آسیب رساندن به برچسب‌خوان	!	!	!	!	!	!
پارازیت (مسدود کردن)	!	!	!	~	~	-
تصادم	!	!	!	~	~	-
سوءاستفاده از منابع پروتکل	-	!	!	-	!	!
تخلیه باتری	-	!	!	-	-	-
کپی‌برداری فیزیکی برچسب	!	!	!	!	!	!
برنامه‌ریزی مجدد برچسب	!	-	-	!	-	-
استراق‌سمع (شنود غیرمجاز)	!	!	!	~	-	-
ردگیری غیرقانونی	~	-	-	-	-	-
بمب RFID	!	~	~	~	-	-
جعل برچسب	!	-	-	~	-	-
زمان‌سنجی	-	-	-	-	-	-
تحلیل مصرف توان	-	-	-	-	-	-
خواندن سطحی و اصلاح برچسب	!	-	-	!	-	-
مرد میانی	~	-	-	~	-	-
ارسال مجدد	!	-	-	!	-	-
فریب مافیا	!	-	-	~	-	-
فریب تروریست	!	-	-	~	-	-
سوءاستفاده از فاصله	!	-	-	~	-	-
سرقت فاصله	!	-	-	~	-	-
سرقت مکان	!	-	-	~	-	-

راهنمایی : ! : مستعد حمله می‌باشد. - : مستعد حمله نمی‌باشد. ~ : احتمال وقوع حمله خیلی کم است.

مراجع

[۱] Syed Ahson, Mohammad Ilyas. "*RFID Handbook : Applications, Technology, Security, and Privacy*", CRC Press (۲۰۰۸).

[۲] Ranasinghe, Damith C., Quan Z. Sheng, and Sherali Zeadally, eds. "*Unique Radio Innovation for the ۲۱st Century: Building Scalable and Global RFID Networks.*" Springer, ۲۰۱۰.

[۳] Pateriya, R. K., and Sangeeta Sharma. "*The evolution of RFID security and privacy: A research survey.*" In Communication Systems and Network Technologies (CSNT), ۲۰۱۱ International Conference on, pp. ۱۱۵-۱۱۹. IEEE, ۲۰۱۱.

[۴] D'Arco, Paolo, Alessandra Scafuro, and Ivan Visconti. "*Revisiting DoS Attacks and Privacy in RFID-Enabled Networks.*" Algorithmic Aspects of Wireless Sensor Networks (۲۰۰۹): ۷۶-۸۷.

[۵] Van Deursen, Ton, and Sasa Radomirovic. "*Attacks on RFID protocols.*" IACR eprint Archive ۳۱۰ (۲۰۰۸).

[۶] Yu, Pengyuan, Patrick Schaumont, and Dong Ha. "*Securing RFID with ultra-wideband modulation.*" In Proceedings of Workshop on RFID Security (RFIDSec), pp. ۲۷-۳۹. ۲۰۰۶.

[٧] Oren, Yossef, Dvir Schirman, and Avishai Wool. "*RFID Jamming and Attacks on Israeli e-Voting*." ITG-Fachbericht-Smart SysTech ٢٠١٢ (٢٠١٢).

[٨] Cremers, Cas, Kasper Bonne Rasmussen, Benedikt Schmidt, and Srdjan Capkun. "*Distance hijacking attacks on distance bounding protocols*." In Security and Privacy (SP), ٢٠١٢ IEEE Symposium on, pp. ١١٣-١٢٧. IEEE, ٢٠١٢.

[٩] Xiao, Qinghan, Thomas Gibbons, and Hervé Lebrun. "*RFID Technology, Security Vulnerabilities, and Countermeasures*." Supply Chain the Way to Flat Organization, Publisher-Intech (٢٠٠٩): ٣٥٧-٣٨٢.

[١٠] Mitrokotsa, Aikaterini, Melanie R. Rieback, and Andrew S. Tanenbaum. "*Classifying RFID attacks and defenses*." Information Systems Frontiers ١٢, no. ٥ (٢٠١٠): ٤٩١-٥٠٥.

[١١] Rieback, Melanie R., Georgi Gaydadjiev, Bruno Crispo, Rutger FH Hofman, and Andrew S. Tanenbaum. "*A platform for RFID security and privacy administration*." In USENIX LISA, pp. ٨٩-١٠٢. ٢٠٠٦.

[١٢] Thompson, Dale R. "*RFID technical tutorial*." Journal of Computing Sciences in Colleges ٢١, no. ٥ (٢٠٠٦): ٨-٩.

[١٣] Shih, Dong-Her, Chin-Yi Lin, and Binshan Lin. "*RFID tags: privacy and security aspects*." International Journal of Mobile Communications ٣, no. ٣ (٢٠٠٥): ٢١٤-٢٣٠.

[١٤] Thompson, Dale R., Neeraj Chaudhry, and Craig W. Thompson. "*RFID security threat model*." In Conf. on Applied Research in Information Technology. ٢٠٠٦.

[١٥] Moskowitz, Paul A., Andris Lauris, and Stephen S. Morris. "*A privacy-enhancing radio frequency identification tag: Implementation of the clipped tag*." In Pervasive Computing and Communications Workshops, ٢٠٠٧. PerCom Workshops' ٠٧. Fifth Annual IEEE International Conference on, pp. ٣٤٨-٣٥١. IEEE, ٢٠٠٧.

[١٦] Juels, Ari, Paul Syverson, and Dan Bailey. "*High-power proxies for enhancing RFID privacy and utility*." In Privacy Enhancing Technologies, pp. ٢١٠-٢٢٦. Springer Berlin/Heidelberg, ٢٠٠٦.

[١٧] Seetharam, Deva, and Richard Fletcher. "*Battery-Powered RFID*." In SenseID ٢٠٠٧ ١st ACM Workshop on Convergence of RFID and Wireless Sensor Networks and their Applications. ٢٠٠٧.

[١٨] Park, JeaHoon, HoonJae Lee, and ManKi Ahn. "*Side-channel attacks against aria on active rfid device*." In Convergence Information Technology, ٢٠٠٧. International Conference on, pp. ٢١٦٣-٢١٦٨. IEEE, ٢٠٠٧.

[١٩] Ha, Dong Sam, and Patrick R. Schaumont. "*Replacing cryptography with ultra wideband (UWB) modulation in secure RFID*." In RFID, ٢٠٠٧. IEEE International Conference on, pp. ٢٣-٢٩. IEEE, ٢٠٠٧.

[٢٠] Yu, Pengyuan, Patrick Schaumont, and Dong Ha. "*Securing RFID with ultra-wideband modulation*." In Proceedings of Workshop on RFID Security (RFIDSec), pp. ٢٧-٣٩. ٢٠٠٦.

[٢١] Duc, Dang Nguyen, Hyunrok Lee, Divyan M. Konidala, and Kwangjo Kim. "*Open issues in RFID security*." In Internet Technology and Secured Transactions, ٢٠٠٩. ICITST ٢٠٠٩. International Conference for, pp. ١-٥. IEEE, ٢٠٠٩.

[٢٢] Peris-Lopez, Pedro, Julio Cesar Hernandez-Castro, Juan M. Estevez-Tapiador, and Arturo Ribagorda. "*Attacking RFID Systems*." Information Security Management Handbook ٥ (٢٠١١): ٣١٣.

[٢٣] Sarma, Sanjay, and Daniel W. Engels. "*On the future of RFID tags and protocols*." White paper, Auto-ID Center, Massachusetts Institute of Technology (٢٠٠٣).

[٢٤] Juels, Ari. "*RFID security and privacy: A research survey*." Selected Areas in Communications, IEEE Journal on ٢٤, no. ٢ (٢٠٠٦): ٣٨١-٣٩٤.

[٢٥] Saparkhojayev, Nurbek, and Dale R. Thompson. "*Matching electronic fingerprints of RFID tags using the hotelling's algorithm*." In Sensors Applications Symposium, ٢٠٠٩. SAS ٢٠٠٩. IEEE, pp. ١٩-٢٤. IEEE, ٢٠٠٩.

[٢٦] Karygiannis, Tom, Bernard Eydt, Greg Barber, Lynn Bunn, and Ted Phillips. "*Guidelines for securing radio frequency identification (RFID) systems*." *NIST Special publication* ٨٠ (٢٠٠٧): ١-١٥٤.

[٢٧] Manfred Aigner , Trevor Burbridge, Alexander llic, David Lyon, Andrea Soppera, Mikko Lehtonen, "*White Paper RFID Tag Security*.", Technical Report , BRIDGE Project (Supported by the European Commision).

[٢٨] Burmester, Mike, and Breno De Medeiros. "*RFID security: attacks, countermeasures and challenges*." Computer Science Department, Florida State University (٢٠٠٧).

[٢٩] Tan, Chiu C., and Jie Wu. "*١٠ Security in RFID Networks and Communications*." Temple University, USA.

[٣٠] Huang, Chia-hung. "*An Overview of RFID Technology, Application, and Security/Privacy Threats and Solutions*." George Mason University, Scholarly paper (٢٠٠٩).

[٣١] Ranasinghe, Damith C., and Peter H. Cole. "*Security in low cost RFID*." Auto-ID Labs University of Adelaide, White Paper (٢٠٠٦).

[٣٢] Rieback, Melanie Rose. "*Security and privacy of radio frequency identification*." Vrije Universiteit, ٢٠٠٨.

[٣٣] Cheng, Shu, "*Security and authentication schemes in RFID*", Master of Computer Science - Research thesis, School of Computer Science and Software Engineering, University of Wollongong, ٢٠١١.

[٣٤] Muhammed Ali, "*SECURITY ANALYSIS OF RFID AUTHENTICATION PROTOCOLS BASED ON SYMMETRIC CRYPTOGRAPHY AND IMPLEMENTATION OF A FORWARD PRIVATE SCHEME*", M.Sc. Thesis, Department of Electronics and Communications Engineering, ISTANBUL

TECHNICAL UNIVERSITY ⋆ INSTITUTE OF SCIENCE AND TECHNOLOGY, ۲۰۱۲

[۳۵] Wang, Hongliang, Bruno Crispo, and Melanie Reiback. "*RFID Guardian Back-end Security Protocol.*" Department of Computer Science, Vrije Universiteit (۲۰۰۸).

[۳۶] White, Jonathan. "*RFID Honeytokens as a Mitigation Tool against Illicit Inventorying.*" Department of Computer Science and Computer Engineering, University of Arkansas.

[۳۷] Nemade, Nikhil, and Krishna C. Konda. "*Security Issues in RFID systems.*" , Department of Electrical & Computer Engineering, Texas A&M University, ۲۰۰۶.

[۳۸] I.Stojmenovic. "*RFID Systems*.", University of Ottawa, ۲۰۱۱.

[۳۹] Http://www.rfid-f۲f.eu/hardware.asp , Last Accessed, November, ۲۰۱۲

[۴۰] Http://www.intechopen.com/books/current-trends-and-challenges-in-rfid/rfid-technology-perspectives-and-technical-considerations-of-microstrip-antennas-for-multi-band-rfid , Last Accessed, November, ۲۰۱۲

[۴۱] Http://www.ubisense.net/, Last Accessed, November, ۲۰۱۲

زیرنویس‌ها

۱ Radio Frequency IDentification
۲ Tag Reader
۳ RFID
٤ Passive
٥ Semi-Passive
٦ Active
۷ UWB (Ultra-Wideband)
۸ Front-end Communication
۹ Back-end Communication
۱۰ Physical Attacks
۱۱ Tag Modification
۱۲ Tag Reprogramming
۱۳ DoS (Denial of Service) Attacks
۱٤ Tag Removal
۱٥ Physical Tampering
۱٦ Zapping
۱۷ Tag Clipping
۱۸ Kill Command
۱۹ Jamming
۲۰ Collision
۲۱ Protocol Resource Exhaustion
۲۲ Reader Vandalization
۲۳ Battery-Depletion
۲٤ Spoofing Attacks
۲٥ Tag Swapping
۲٦ Distance Fraud
۲۷ Tag Counterfeiting (Duplication)
۲۸ Tag Cloning
۲۹ Distance Hijacking
۳۰ Distance Bounding Protocols
۳۱ Location Hijacking
۳۲ Relay Attacks
۳۳ MitM (Man-in-the-Middle)
۳٤ Passive
۳٥ Active
۳٦ Mafia Fraud
۳۷ Terrorist Fraud
۳۸ Replay
۳۹ Side-Channel Attacks
٤۰ Timing
٤۱ Power Consumption
٤۲ Eavesdropping
٤۳ Sniffing
٤٤ Snooping
٤٥ Skimming
٤٦ Privacy Threats
٤۷ Illicit Tracking
٤۸ RFID Bomb
٤۹ Symmetric-key-Based Authentication Scheme
٥۰ Asymmetric-key-Based Authentication Scheme
٥۱ Public-Key
٥۲ Identification
٥۳ Full-fledged Class
٥٤ Hash Functions
٥٥ Digital Signature
٥٦ Simple Class
٥۷ PRNG (Pseudo-Random Number Generator)
٥۸ One-way Hash Functions
٥۹ Lightweight Class
٦۰ Ultra-Lightweight Class

[٦١] Tamper-Resistant (Tamper-evident) Tags
[٦٢] RFID IDS (Intrusion Detection System)
[٦٣] RFID Guardian
[٦٤] REP (RFID Enhancer Proxy)
[٦٥] Timestamps
[٦٦] Clock Synchronization
[٦٧] Tag Sleeping
[٦٨] Tag Blocking
[٦٩] Relabeling
[٧٠] Hash Functions
[٧١] MAC (Message Authentication Code)
[٧٢] Symmetric-Key Encryption
[٧٣] Public-Key Encryption
[٧٤] Challenge-Response
[٧٥] Forward Channel
[٧٦] Backward Channel
[٧٧] Semi-UWB

بررسی قابلیت سیستم‌های مبتنی بر RFID برای مدیریت ترافیک شهری

سعید میرزایی[1]، سید وحید ازهری[2]

[1]دانشجوی کارشناسی ارشد
saeid_mirzaee@comp.iust.ac.ir

[2]استاد راهنما
azharivs@iust.ac.ir

چکیده

با رشد چشم‌گیر وسایل نقلیه و تردد آنها در جاده‌های مختلف شهری، شناسایی خودکار آنها برای مدیریت ترافیک شهری، امری ضروری است. در این مقاله، با معرفی تکنولوژی RFID و اجزای آن، قابلیت آن در حوزه‌ی مدیریت ترافیک شهری مورد بررسی قرار می‌گیرد. با نگاهی به سیستم‌های غیر RFID در مدیریت ترافیک شهری و مسائل و مشکلات آنها، نقاط قوت تکنولوژی RFID مطرح شده و کاربردهایی را که حاصل استفاده از این تکنولوژی در حوزه مدیریت ترافیک شهری است، تشریح می‌شود. از آنجا که هر سیستمی برای امکان‌پذیر شدن با مسائل و چالش‌هایی روبروست، چالش‌های مرتبط با تکنولوژی RFID برای شناسایی وسایل نقلیه نیز بررسی شده و برخی از راهکارهای ارائه شده، مطرح می‌شود.

کلمات کلیدی

RFID، شناسه، برچسب رادیویی، برچسب‌خوان، مدیریت ترافیک شهری.

۱- مقدمه

چگونگی مدیریت ترافیک شهری و مسائل مرتبط با حمل و نقل یکی از معضلات اغلب جوامع امروزی است. حجم وسایل نقلیه به صورت نمایی در حال گسترش است و این مسئله کنترل و مدیریت آنها را با مشکلاتی روبرو می‌کند [3]. با افزایش تخلفات رانندگی که زیان‌های جانی و مالی بیشماری را به بار می‌آورد، نیازمند سیستمی هستیم که به صورت خودکار (بدون دخالت انسان) و مؤثر تمامی وسایل نقلیه را شناسایی کرده و تخلفات انجام گرفته توسط آنها را ثبت نماید. RFID به عنوان یک تکنولوژی فراگیر، به علت ساده‌گی، کارایی و هزینه نسبتاً پایین می تواند در این زمینه مورد استفاده قرار گیرد.

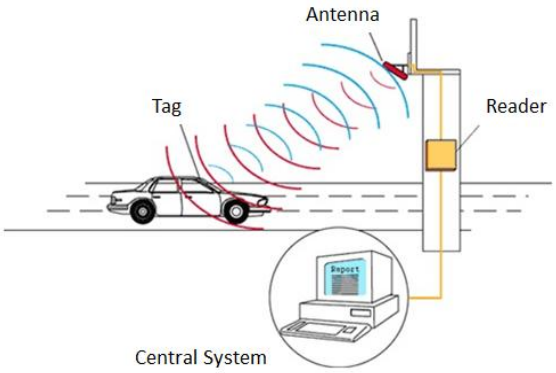
اصطلاح RFID، سرنام کلمات Radio Frequency IDentification و به معنای شناسایی از طریق امواج رادیویی است [1,7]. تکنولوژی RFID نوظهور نیست و تاریخچه آن به جنگ جهانی دوم برمی‌گردد که در قالب سیستمی به نام IFF[1] توسط انگلیسی‌ها برای شناسایی هواپیماهای خودی از هواپیماهای دشمن پیاده شده بود [2,22]. با آغاز قرن ۲۱ و با پیشرفت‌های به عمل آمده در زمینه ساخت تراشه‌های بسیار ریز، RFID بیش از پیش مورد توجه قرار گرفت و تقریباً در هر زمینه‌ای که نیاز به شناسایی اجسام بود، به کار گرفته شد. از مهمترین کاربردهایی که هم‌اکنون از RFID بهره می‌گیرند، می‌توان به شناسایی کالاها در فروشگاه‌ها، شناسایی بیماران و تجهیزات پزشکی در بیمارستان‌ها، شناسایی حیوانات، انبارداری، زنجیره تأمین، کنترل دسترسی، کارت‌های هوشمند و غیره نام برد [1,4].

RFID در زمره‌ی تکنولوژی‌های «شناسایی خودکار[2]» قرار می‌گیرد. در واقع به هر نوع سیستمی که در آن عملیات شناسایی یک فرد و یا شیء به صورت خودکار انجام گیرد، شناسایی خودکار گویند که هدف اینگونه سیستم‌ها، افزایش کارایی، سهولت، کاهش خطا در ورود اطلاعات و عدم نیاز به عامل انسانی در فرایند شناسایی است. وقتی RFID در حوزه مدیریت ترافیک شهری مورد استفاده قرار گیرد، در دسته «سیستم‌های حمل و نقل هوشمند[3]» نیز قرار می‌گیرد [1,23].

تمامی سیستم‌های RFID شامل سه جزء: برچسب رادیویی[4]، برچسب‌خوان[5] و سیستم‌مرکزی (پایگاه داده) می‌باشند [1,2]. **برچسب رادیویی** شامل یک تراشه الکترونیکی است. این برچسب بر روی هر شیئی که خواهان شناسایی خودکار آن هستیم، نصب (چسبانده) می‌شود. تراشه موجود در برچسب حاوی یک شناسه (ID) است که هویت شیئی که بر روی آن نصب شده است را مشخص می‌کند. **برچسب‌خوان** دستگاهی ثابت و یا دستی است که با قرار گرفتن برچسب‌ها در محدوده آن، شناسه‌های آنها را از طریق امواج

رادیویی می‌خواند [10]. شناسه‌های خوانده شده توسط برچسب‌خوان برای انجام پردازش‌های لازم به **سیستم مرکزی** ارسال می‌شوند.

در کاربرد مدیریت ترافیک شهری، برچسب‌های رادیویی بر روی وسایل نقلیه (ماشین، موتور و غیره) قرار می‌گیرند و برچسب‌خوان‌ها نیز در مکان‌هایی در کنار جاده‌ها نصب می‌شوند. با عبور هر وسیله نقلیه از کنار برچسب‌خوان، یک ارتباط بیسیم بین برچسب و برچسب‌خوان برقرار شده و شناسه هر وسیله نقلیه خوانده می‌شود.



شکل (۱): اجزای سیستم RFID [5]

این مقاله به صورت زیر طبقه‌بندی شده است. در بخش ۲ جزئیات مرتبط با تکنولوژی RFID بررسی می‌شود. در بخش ۳ استانداردهای RFID تشریح می‌شود. سیستم‌های غیر RFID برای مدیریت ترافیک شهری در بخش ۴ مورد بحث قرار می‌گیرد. بخش ۵ کاربردهای RFID در مدیریت ترافیک شهری را بیان می‌کند. چالش‌های مرتبط با پیاده‌سازی سیستم RFID برای مدیریت ترافیک شهری در بخش ۶ و در نهایت یک نتیجه‌گیری در بخش ۷ ارائه می‌شود.

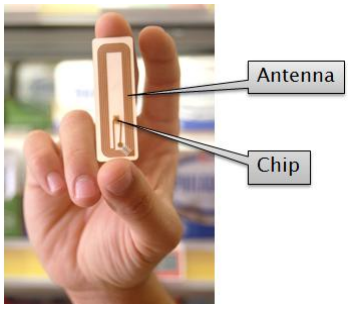
۲- تکنولوژی RFID و جزئیات آن

فرآیند شناسایی در تکنولوژی RFID، بدون اینکه تماسی (برخوردی) بین برچسب و برچسب‌خوان برقرار شود، انجام می‌گیرد [1,6,23]. همچنین برخلاف روش‌هایی مثل بارکد نیازی به خط دید مستقیم بین برچسب و برچسب‌خوان نیست [6]. از دیگر ویژگی‌های بارز RFID می‌توان به شناسایی منحصر بفرد هر شیء، توانایی خواندن صدها برچسب توسط برچسب‌خوان در هر لحظه و عدم کاهش کارایی در شرایط بد آب و هوایی [8] اشاره کرد.

از آنجائیکه RFID یک تکنولوژی بیسیم است، لذا در لایه «فیزیکی» و «پیوند داده» مدل OSI کار می‌کند. در ادامه پس از شرح انواع برچسب‌های RFID، مسائلی مانند فرکانس‌های کاری RFID در لایه فیزیکی و پروتکل‌های anti-collision بکار رفته در لایه MAC را بررسی می‌کنیم.

۲-۱- برچسب‌های RFID

برچسب‌های RFID در ساده‌ترین حالت شامل یک تراشه الکترونیکی، آنتن و در برخی از انواع آن شامل باتری نیز می‌باشند [10].



شکل (۲): یک نمونه برچسب رادیویی [9]

برچسب‌های RFID از نظر منبع انرژی به سه دسته کلی تقسیم‌بندی می‌شوند [1,10]:

- **برچسب‌های غیرفعال**[۶]

برچسب‌های غیرفعال منبع انرژی (باتری) نداشته و برای فعال شدن IC و ارسال شناسه خود نیازمند دریافت انرژی از سیگنال‌های ارسالی از برچسب‌خوان هستند [1,10,26]؛ به عبارت دیگر از تکنیک backscatter برای ارسال‌های خود استفاده می‌کنند. میزان انرژی دریافتی از امواج برچسب‌خوان بسیار ناچیز است؛ لذا همین امر موجب بروز محدودیت‌هایی در این نوع از برچسب‌ها می‌شود. برخی از این محدودیت‌ها عبارتنداز: کاهش محدوده خواندن برچسب‌ها، عدم توانایی انجام محاسبات سنگین، عدم توانایی ثبت و ارسال خودکار رخدادها و لینک ارتباطی ضعیف بین برچسب و برچسب‌خوان.

علیرغم محدودیت‌های ذکر شده، سه ویژگی عمده‌ی اندازه و وزن پایین، هزینه پایین و طول عمر بالا منجر به گستردگی استفاده از این نوع برچسب‌ها شده است.

- **برچسب‌های فعال**[۷]

برچسب‌های فعال دارای باتری بوده و تمامی پردازش‌های خود و همچنین ارسال شناسه به برچسب‌خوان را با استفاده از این منبع انرژی درونی انجام می‌دهند. به دلیل وجود باتری، این برچسب‌ها قابلیت «آغازگر ارتباط بودن» را دارا خواهند بود. برخی از مهمترین مزیت‌های برچسب‌های فعال به شرح است:

افزایش محدوده خواندن تا ۱۰۰ متر و بالاتر، قابلیت اطمینان[۸] بیشتر لینک ارتباطی بین برچسب و برچسب‌خوان، قابلیت خواندن تعداد برچسب‌های بیشتر در مدت زمان کوتاه حتی در سرعت‌های بالا، امکان پیاده‌سازی برخی از خصیصه‌های امنیتی، قابلیت استفاده از انواع مختلفی از حسگرها و توانایی ثبت رخدادها، قابلیت ارتباط برچسب‌ها با یکدیگر در صورت لزوم (ad-hoc capability)، دارای حافظه داخلی بیشتر نسبت به سایر برچسب‌ها، داشتن نرخ انتقال بالاتر به علت داشتن توان و پهنای باند بالاتر [10,34] (برچسب‌های فعال معمولاً در فرکانس‌های بالا (Microwave و UWB) کار می‌کنند؛ در فرکانس‌های بالاتر به علت خواص فرکانسی و قابلیت داشتن پهنای باند بالا، می‌توان به نرخ‌های بالاتری دست یافت [45]).

اما برچسب‌های فعال معایبی نیز دارند که باعث ناکارآمد شدن آنها در بسیاری از کاربردها می‌شود: اندازه و وزن بالا به علت استفاده از باتری، هزینه بالای هر برچسب، هزینه‌بر بودن نگهداری بلند مدت کل سیستم و طول عمر پایین برچسب (وابسته به باتری) [10].

- **برچسب‌های نیمه-فعال**[9]

برچسب‌های نیمه-فعال برخلاف برچسب‌های غیرفعال دارای منبع انرژی (باتری) بوده و به برچسب‌های battery-assisted نیز مشهورند [1]. این برچسب‌ها از این حیث که دارای باتری هستند، شبیه به برچسب‌های فعال می‌باشند. اما از طرف دیگر همانند برچسب‌های غیرفعال برای ارسال پاسخ به برچسب‌خوان، از انرژی دریافتی از سیگنال ارسال شده توسط برچسب‌خوان استفاده می‌کنند (backscatter). در حقیقت برچسب‌های نیمه-فعال از لحاظ مزایا و معایب، مابین دو مورد قبلی قرار می‌گیرند. به عنوان مثال هزینه بیشتر نسبت به برچسب‌های غیرفعال و کمتر نسبت به برچسب‌های فعال.

مهمترین دلایل استفاده از برچسب‌های نیمه-فعال عبارتند از: افزایش محدوده‌ی خواندن در برخی از کاربردها و مجهز کردن برچسب‌ها به حسگرهای خاص که نیازمند منبع انرژی برای ثبت رخدادها هستند. مقایسه کلی این سه نوع برچسب در جدول ۱ آمده است.

خصیصه / نوع برچسب	غیرفعال (Passive)	نیمه-فعال (Semi-Passive)	فعال (Active)
اندازه/وزن	کوچک/پایین	نسبتاً بزرگ/بالا	بزرگ/بالا
قیمت	پایین	نسبتاً بالا	بالا
منبع انرژی	ندارد	باتری	باتری
طول عمر	بدون محدودیت	محدود (وابسته به باتری)	محدود (وابسته به باتری)
نگهداری بلند مدت سیستم	هزینه پایین	هزینه بالا	هزینه بالا
دمای کاری	-40°C to +80°C	-20°C to +70°C	-20°C to +70°C
فرکانس	LF, HF, UHF	UHF, Microwave	UHF, Microwave
توانایی پردازش	بسیار محدود	بالا	بالا
قابلیت اطمینان لینک	پایین	پایین	بالا
محدوده خواندن	حداکثر ۱۰ متر	بطور متوسط حداکثر ۱۵ متر	۱۰۰ متر و بیشتر
نرخ خواندن	پایین	نسبتاً پایین	بالا
تأثیر سرعت بالا در خواندن برچسب	ممکن است مشکل ساز شود	نسبت به Passive بهتر است	نسبت به بقیه بهتر است.

جدول (۱): مقایسه انواع برچسب‌های رادیویی

۲-۲- فرکانس‌های استفاده شده در RFID

اغلب سیستم‌های RFID پیاده شده، در یکی از فرکانس‌های زیر کار می‌کنند [10,35]:

- فرکانس LF[10] (۱۲۵ تا ۱۳۴ کیلوهرتز)
- فرکانس HF[11] (۱۳.۵۶ مگاهرتز)
- فرکانس UHF[12] (۸۶۰ تا ۹۶۰ مگاهرتز)
- فرکانس Microwave (۲.۴ گیگاهرتز)
- فرکانس UWB[13] (۳.۱ تا ۱۰.۶ گیگاهرتز)

فرکانس‌های LF و HF محدوده‌ی خواندن بسیار پایینی دارند (در حد چندین اینچ تا چندین فوت). همچنین نرخ بیت در این فرکانس‌ها نیز پایین است. بیشتر برای کاربردهای کنترل دسترسی، کتابخانه‌ها و immobilizer استفاده می‌شوند. برچسب‌هایی که از این نوع فرکانس استفاده می‌کنند، برچسب‌های غیرفعال هستند.

فرکانس UHF پرکاربردترین فرکانس RFID می‌باشد که تقریباً در هر نوع کاربردی قابل استفاده است. دارای نرخ ارسال بالاتری نسبت به دو مورد قبلی بوده و محدوده‌ی خواندن آن نیز بیشتر است.

فرکانس Microwave که معمولاً برچسب‌های نوع فعال از آن استفاده می‌کنند، نرخ انتقال بالاتری نسبت به موارد قبلی دارد. در کاربردهایی که در آنها شناسایی اجسام در حال حرکت مد نظر است، در صورت استفاده از فرکانس‌های بالا براحتی می‌توان با پدیده شیفت داپلر مقابله کرد. لذا برای کاربردهای مدیریت ترافیک، فرکانس‌های UHF و Microwave مناسب هستند.

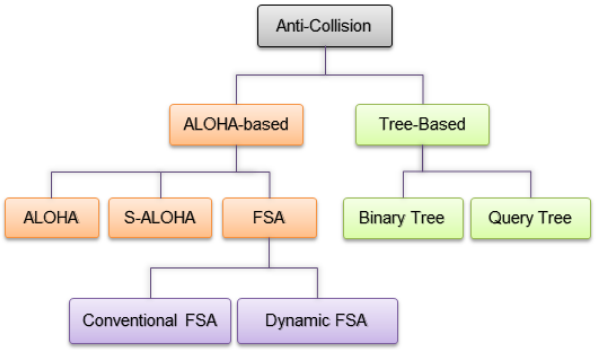
سیگنال‌هایی که از فرکانس‌های بیان شده‌ی فوق استفاده می‌کنند، همگی به صورت باند باریک بوده و تمام مشکلات مرتبط با سیگنال‌های باند باریک مانند قرار گرفتن در فید عمیق و غیره را دارند. در نقطه مقابل این سیستم‌ها، سیستم‌های UWB قرار دارند که از پهنای باند به مراتب بزرگتری نسبت به سیستم‌های باند باریک استفاده می‌کنند.

بر طبق تعریف ارائه شده توسط FCC[14]، به هر نوع سیستمی که دارای ۵۰۰ مگاهرتز پهنای باند و یا ۲۰ درصد فرکانس مرکزی پهنای باند در بازه ۳.۱ تا ۱۰.۶ گیگاهرتز باشد، UWB گفته می‌شود [36]. با استفاده از UWB می‌توان انتقال اطلاعات را با نرخ بالا در یک فاصله کوتاه و با توان پایین انجام داد. طبق قانون شانون، نرخ بیت رابطه خطی با پهنای باند و رابطه لگاریتمیک با توان مصرفی دارد، پس می‌توان با پهنای باند بسیار بالا، نرخ بیت بالایی را در قبال استفاده از توان بسیار ناچیز بدست آورد [36]. همچنین استفاده از پهنای باند زیاد با توان کم باعث شبیه به نویز بودن سیگنال‌های ارسالی شده و تشخیص اطلاعات ارسالی را برای افراد غیرمجاز دشوار می‌کند [11]. برچسب‌های RFID ای که از فرکانس UWB استفاده می‌کنند، معمولاً فعال هستند؛ هرچند که می‌توان مکانیزمی را اتخاذ کرد که بتوان برچسب‌های غیرفعال را نیز قادر به استفاده از این محدوده‌ی فرکانس کرد.

یکی از مهمترین دلایل استفاده از برچسب‌هایی که در بازه فرکانسی UWB کار می‌کنند، مقابله با یکسری از تهدیدات امنیتی است. علاوه بر این به دلیل عوض شدن ماهیت انتقال، این تکنیک مزایای دیگری به شرح زیر دارد: کاهش توان مصرف، duty cycle پایین، مقابله با پدیده فیدینگ چند مسیره و ساده شدن مدارات داخلی برچسب و برچسب‌خوان به دلیل عدم نیاز به مدوله کردن سیگنال بر روی یک فرکانس حامل برای انتقال اطلاعات. یکی از مهمترین معایب تکنیک UWB نیز دشوار بودن همگام سازی بین فرستنده و گیرنده است؛ چرا که انتقال اطلاعات در این تکنیک در پالس‌های بسیار کوتاه (در حد نانو و پیکو ثانیه) انجام می‌گیرد.

۲-۳- پروتکل‌های Anti-Collision

اگر بیش از یک برچسب در محدوده یک برچسب‌خوان قرار بگیرد، همگی آنها همزمان اقدام به ارسال شناسه خود کرده و در نتیجه تصادم بوجود می‌آید چرا که برچسب‌خوان قادر به شناسایی هیچ یک از آنها نخواهد بود. برای مقابله با این مشکل باید از پروتکل‌های anti-collision استفاده کنیم. پروتکل‌های anti-collision استفاده شده در تکنولوژی RFID، به دو دسته کلی **مبتنی بر درخت** و **مبتنی بر ALOHA** تقسیم می‌شوند [1,13-16].



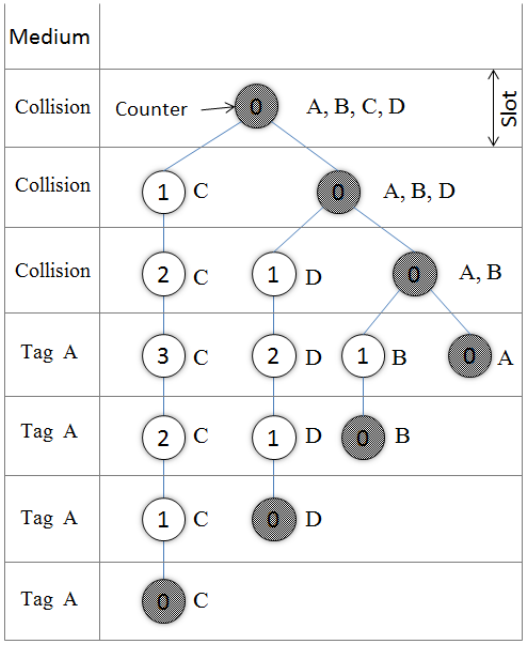
شکل (۳): الگوریتم‌های anti-collision

هر دوی این پروتکل‌ها بر اساس تقسیم زمانی[15] عمل می‌کنند.

- **پروتکل‌های مبتنی بر درخت**

این پروتکل‌ها در فرآیند شناسایی برچسب‌ها، از لحاظ مفهومی تشکیل یک درخت را می‌دهند. روش کار آنها بدین صورت است که تمامی برچسب‌های موجود در در محدوده برچسب‌خوان را به دو مجموعه تقسیم می‌کنند. تقسیم‌بندی تا زمانی که هر مجموعه تنها یک برچسب را شامل شود، ادامه پیدا می‌کند که در این هنگام این برچسب شناسایی می‌شود؛ سپس به صورت بازگشتی به سراغ بقیه مجموعه‌ها رفته و همین عمل ادامه پیدا می‌کند [1,13]. دو پروتکل مهمی که در این دسته قرار می‌گیرند پروتکل‌های binary tree و query tree هستند [1,14] که از لحاظ روش تقسیم‌بندی برچسب‌ها با یکدیگر متفاوتند.

در پروتکل binary tree، تقسیم‌بندی بر اساس شماره‌ای که در برچسب‌ها وجود دارد انجام می‌گیرد [1]. مقدار اولیه تمامی شماره‌ها صفر است. وقتی که برچسب‌خوان درخواست خواندن را ارسال می‌کند، هر برچسبی که شماره صفر داشته باشد، شناسه خود را ارسال می‌کند. بنابراین در اولین سیکل تمامی برچسب‌ها شناسه خود را ارسال کرده و تصادم می‌کنند. در سیکل بعد، تمامی برچسب‌ها یک عدد باینری تصادفی (۰ یا ۱) را انتخاب کرده و به شماره خود می‌افزایند. در نتیجه تمامی برچسب‌ها به دو مجموعه تقسیم می‌شوند. دوباره کار از سر گرفته می‌شود و برچسب‌های با شماره صفر شروع به ارسال می‌کنند. در هر سیکلی که تصادم رخ می‌دهد، برچسب‌هایی که در آن تصادم شرکت نکرده‌اند، یک واحد به شماره خود می‌افزایند. در هر سیکلی هم که تصادم رخ نمی‌دهد، تمامی برچسب‌ها یک واحد از شماره خود می‌کاهند. این تقسیم‌بندی تا جایی پیش می‌رود که هر سیکل تنها شامل یک برچسب باشد که ارسال آن موفقیت آمیز خواهد بود. عملیات تا زمانی که تمامی برچسب‌ها شناسایی شوند، ادامه پیدا می‌کند [1]. با توجه به شکل (۴) اگر در هر سیکل بیش از یک برچسب دارای شماره صفر باشند، رخدادن تصادم قطعی است.



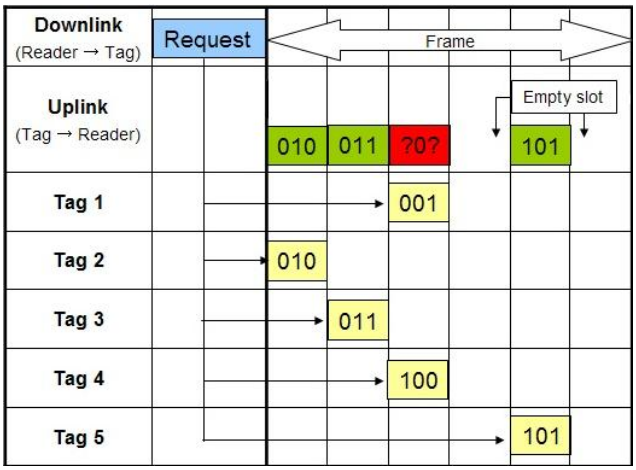
شکل (۴): مثالی از الگوریتم binary tree با چهار برچسب

ایده پروتکل query tree، نیز به همین صورت است، با این تفاوت که عملیات تقسیم‌بندی بر اساس شناسه خود برچسب انجام می‌گیرد [1]. به این صورت که هر برچسب‌خوان به همراه درخواست خواندن یک رشته بیت هم ارسال می‌کند. هر برچسبی که شناسه‌اش با این رشته بیت آغاز شود، شناسه خود را ارسال می‌کند در صورت رخ دادن تصادم، یک بیت به رشته ارسالی اضافه شده و فرآیند از سرگرفته می‌شود. در این روش اگر شناسه‌ها خیلی شبیه به هم باشند. به علت تصادم‌های زیاد، فرآیند خواندن طولانی می‌شود.

مهمترین عیب پروتکل‌های مبتنی بر درخت، تصادم‌های زیاد و در پی آن طولانی شدن پروسه خواندن برچسب‌ها می‌باشد. در [1] روش‌های بهبود یافته این پروتکل‌ها شرح داده شده است.

- **پروتکل‌های مبتنی بر ALOHA**

در این روش‌ها، مهمترین پروتکلی که استفاده می‌شود، FSA[۱۶] و نسخه‌های بهبود یافته آن است [14]. روش کار بدین صورت است که برچسب‌خوان یک فریم که شامل تعدادی اسلات زمانی است برای برچسب‌ها ارسال می‌کند. برچسب‌ها یک اسلات از این فریم را به صورت تصادفی انتخاب می‌کنند تا شناسه خود را در آن زمان ارسال کنند. اگر بیش از یک برچسب، شناسه خود را در یک اسلات ارسال کنند. تصادم رخ می‌دهد. برچسب‌های تصادم کرده بار دیگر در فریم بعد در فرآیند شناسایی شرکت می‌کنند. در این روش اگر تعداد اسلات‌ها تقریباً برابر با تعداد برچسب‌های موجود در محیط باشد، بیشترین کارایی را خواهیم داشت [14,15]. لذا الگوریتمی به نام Dynamic FSA مطرح شده است که در هر دور، اندازه فریم را به صورت داینامیک تغییر می‌دهد [1,14]. این تغییر، نیازمند الگوریتمی برای تخمین زدن تعداد برچسب‌ها است که در [14] مورد بررسی قرار گرفته است.



شکل (۵): مثالی از الگوریتم FSA [44]

همانطور که در شکل (۵) می‌بینید، برچسب‌های ۲، ۳ و ۵ به ترتیب اسلات‌های زمانی ۱، ۲ و ۵ را انتخاب کرده و با موفقیت ارسال خود را انجام می‌دهند. اما برچسب‌های ۱ و ۴ که هر دو اسلات سوم را انتخاب کرده‌اند، دچار تصادم شده و باید در فریم بعد شانس خود را برای ارسال مجدد امتحان کنند.

۳- استانداردهای RFID

بطور کلی دو موسسه اصلی کار استانداردسازی تکنولوژی RFID را انجام می‌دهند. یکی از آنها سازمان ISO[۱۷] است که استانداردهای مختص تکنولوژی RFID را با شماره ISO-18000 نامگذاری کرده است؛ استاندارد ISO-18000 دارای ۷ بخش می‌باشد که هر یک از این بخش‌ها بیانگر نوع برچسب، الگوریتم MAC و فرکانس مورد استفاده می باشند [10].

موسسه دیگر، انجمن EPC Global است که استانداردهای این انجمن کلاس‌بندی‌های مختلفی را برحسب نوع برچسب، نوع حافظه، نوع پروتکل مورداستفاده و غیره ارائه می‌دهد.

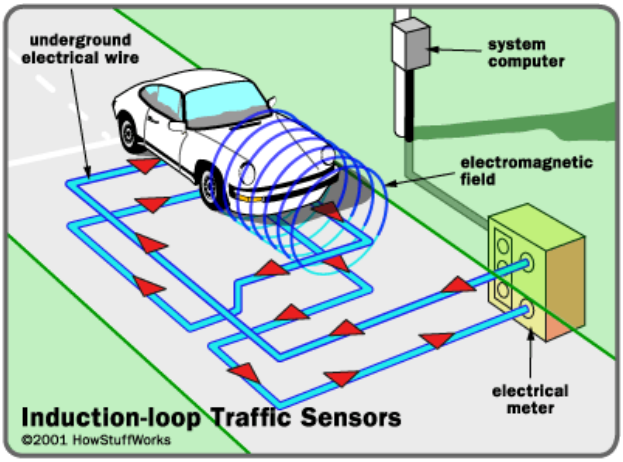
نام کلاس	نوع برچسب	نوع حافظه
Class 0	غیرفعال	فقط خواندنی
Class 1 Gen 1	غیرفعال	WORM[۱۸]
Class 1 Gen 2	غیرفعال	خواندنی/نوشتنی
Class 2	غیرفعال (قابلیت بیشتر)	خواندنی/نوشتنی
Class 3	نیمه-فعال	خواندنی/نوشتنی
Class 4	فعال	خواندنی/نوشتنی
Class 5	فعال (برچسب‌خوان)	خواندنی/نوشتنی

۴- بررسی سیستم‌های غیر RFID برای مدیریت ترافیک شهری

قبل از پرداختن به کاربردهای RFID، در این بخش برخی از سیستم‌های غیر RFID را برای مدیریت ترافیک شهری بررسی کرده و مزایا و معایب هر یک را شرح می‌دهیم و جایگاه تکنولوژی RFID را در کنار این سیستم‌ها بیان می‌کنیم.

۴-۱- حلقه‌های القایی[۱۹]

یکی از روش‌های تشخیص وجود یا عدم وجود وسایل نقلیه در قسمت‌های مختلف جاده، استفاده از حلقه‌های القایی است. در این روش خطوطی از جاده را که خواهان تشخیص وسایل نقلیه در آنها هستیم، به شکل مربع و یا مستطیل شکاف داده (به عمق ۵ الی ۱۰ سانتی‌متر) و حلقه‌هایی از سیم را در این شکاف‌ها قرار می‌دهیم. دو سر این حلقه را به دستگاه شناساگر متصل می‌کنیم. شناساگر جریانی را در این حلقه‌ها به‌وجود می‌آورد که این جریان منجر به ایجاد یک میدان مغناطیسی در اطراف حلقه می‌شود. با عبور وسایل نقلیه (اجسام فلزی) از روی این حلقه، تغییراتی در خواص الکتریکی حلقه ایجاد می‌شود که شناساگر متوجه این تغییرات شده و در نتیجه وجود وسیله نقلیه تشخیص داده می‌شود [37]. از جمله کاربردهای این روش در مدیریت ترافیک شهری می‌توان به «شمارش تعداد خودروها در یک مسیر و تشخیص ازدحام» اشاره کرد.



شکل (۶): تشخیص وسایل نقلیه از طریق حلقه القایی [38]

اما اگر از دشواری‌های نصب و نگهداری و سایر مشکلات موجود در تشخیص وسایل نقلیه در اینگونه سیستم‌ها هم چشم‌پوشی کنیم، باز هم یک عیب بزرگی که این سیستم‌ها دارند، عدم شناسایی منحصر به فرد هر وسیله نقلیه است. در حالیکه که همانطور که قبلاً بیان کردیم یکی از نقاط قوت سیستم‌های مبتنی بر RFID، شناسایی منحصر به فرد هر وسیله نقلیه است. بسیاری از کاربردهای مدیریت ترافیک و حمل و نقل بر اساس همین شناسایی منحصر به فرد میسر می‌شود. این کاربردها در بخش ۵ شرح داده شده‌اند.

۴-۲- VANETs[۲۰]

یکی دیگر از مهمترین و مشهورترین شبکه‌هایی که در زمینه مدیریت ترافیک مطرح است، VANET می‌باشد. بر طبق [17]، دو هدف اصلی شبکه‌های VANET، فراهم کردن ایمنی و راحتی رانندگان و مسافران است که مورد اول معمولاً با ارتباطات V2V[۲۱] و مورد دوم با ارتباطات V2I[۲۲] میسر می‌شود. در واقع در شبکه‌های VANET هر وسیله نقلیه مجهز به یک رابط رادیویی فعال (شبیه برچسب‌خوان در RFID) است که همین امر هزینه بسیار زیادی را به سیستم تحمیل می‌کند. در حالیکه در شبکه‌های معمول RFID، برچسب‌خوان‌ها در کنار جاده‌ها نصب شده و هر وسیله نقلیه مجهز به یک برچسب است که در مقابل برچسب‌خوان هزینه بسیار اندکی دارد [19].

تکنولوژی RFID در نقطه مقابل VANET قرار نگرفته است که بخواهیم تنها یک مورد از این دو را انتخاب کنیم، بلکه هر یک اهدافی متفاوت را دنبال می‌کنند؛ هر چند ممکن است که در برخی موارد مشترکاتی نیز با هم داشته باشند. لذا RFID می‌تواند در کنار VANET نیز مورد استفاده قرار گیرد. طبق [18,19]، RFID در دسته ارتباطات V2I قرار می‌گیرد و از این نظر می‌توان برای کاهش هزینه‌های VANET، این بخش از VANET را در برخی از کاربردها، با استفاده از RFID پیاده کرد.

۴-۳- تحلیل تصاویر ضبط شده توسط دوربین‌ها

در بین سیستم‌های مدیریت ترافیک، دوربین‌های ترافیکی از لحاظ اهدافی که دنبال می‌کنند، بیشترین شباهت را با سیستم‌های RFID دارند؛ بگونه‌ای که می‌توان اغلب کاربردهای RFID را با استفاده از این دوربین‌ها پیاده کرد (مثلاً سرعت غیر مجاز، ورود غیر مجاز به طرح‌های ترافیکی و غیره). اکثر کشورهای جهان از جمله ایران در مدیریت ترافیک شهری از دوربین‌های ترافیکی بهره گرفته‌اند.

در کاربردهایی که شناسایی منحصر به فرد هر وسیله نقلیه مد نظر است، نیازمند این هستیم که دوربین‌ها از پلاک وسیله نقلیه عکس بگیرند که در این صورت هیچ مانعی نباید در فاصله بین دوربین و پلاک وجود داشته باشد. کثیف شدن پلاک به علت بارش، مخفی شدن پلاک پشت برخی از خودروها در مکان‌های شلوغ، عدم وجود نور کافی در شب برای دیده شدن پلاک و غیره از مهمترین مسائلی هستند که باعث ناکارآمد شدن دوربین‌ها در این زمینه می‌شوند [39]. همچنین تحلیل تصاویر ویدئویی نیازمند الگوریتم‌های پیچیده و زمانبری است که دقت آنها نیز بر کیفیت سیستم اثرگذار است [6].

۵- کاربردهای RFID در مدیریت ترافیک شهری

مهمترین کاربردهایی را که می‌توان از طریق سیستم RFID در مدیریت ترافیک شهری پیاده کرد، عبارتند از: تشخیص داشتن بیمه نامه و معاینه فنی خودرو، عبور از چراغ قرمز، سرعت غیر مجاز، طرح ترافیک و مناطق ورود ممنوع (طرح زوج و فرد، ورود ممنوع، خیابان-

های یکطرفه)، ردگیری وسایل نقلیه دزدیده شده، پارک ممنوع (توقف ممنوع)، دور زدن ممنوع، گردش به چپ و گردش به راست ممنوع، جمع‌آوری عوارض جاده‌ای، کنترل ورود و خروج پارکینگ‌ها، چراغ ترافیکی هوشمند، تشخیص ازدحام در مسیرهای مختلف و پیشنهاد مسیر و ردیابی وسایل نقلیه عمومی (اتوبوس، تاکسی). در ادامه برخی از این کاربردها را با جزئیات بیشتر بررسی کرده و نحوه کار هر یک را بیان می‌کنیم.

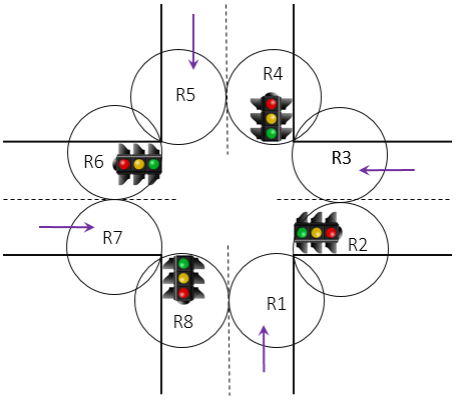
۵-۱- بیمه‌نامه و معاینه فنی خودرو

داشتن بیمه‌نامه و معاینه فنی خودرو برای تمامی وسایل نقلیه الزامی است. روش تشخیص وسایل نقلیه فاقد بیمه‌نامه و معاینه فنی بدین صورت است که با عبور وسایل نقلیه از محدوده برچسب‌خوان نصب شده در کنار جاده، شناسه (ID) آنها خوانده شده و به سیستم مرکزی ارسال می‌شود؛ سپس بررسی می‌شود که آیا برای شناسه‌های خوانده شده بیمه‌نامه و معاینه فنی معتبر ثبت شده است یا خیر؛ که در صورت نداشتن هر یک از موارد فوق جریمه مورد نظر صادر شده و به پلیس‌های مستقر در راه، مسیر حرکت متخلفان اعلام شده تا در جهت متوقف کردن آنها، اقدامات لازم انجام گیرد.

۵-۲- عبور از چراغ قرمز

چراغ‌های راهنمایی می‌توانند در مکان‌ها و با حالت‌های متفاوتی نصب گردند، در اینجا ما حالت عمومی آن را بررسی می کنیم. در حالت عمومی چهارراهی وجود دارد (با ۴ چراغ) که هر راه آن شامل دو مسیر می‌باشد. هر مسیر را به یک برچسب‌خوان مجهز می‌کنیم، یعنی در کل ۸ برچسب‌خوان در یک چهارراه نصب می‌کنیم. برچسب هر وسیله نقلیه برای عبور از چهارراه توسط دو برچسب‌خوان خوانده می‌شود و از آنجا که برچسب‌خوان‌ها زمان سبز یا قرمز بودن چراغ‌ها را می‌دانند، متخلفان قابل تشخیص خواهند بود[21].

در واقع برچسب‌خوان اول مکان ورود به تقاطع را مشخص می‌کند و برچسب‌خوان دوم تعیین می‌کند که آیا در آن مقطع زمانی (زمان سبز یا قرمز بودن چراغ‌ها) آن وسیله نقلیه حق عبور از برچسب‌خوان دوم (مسیر خروجی) را داشته است یا خیر.



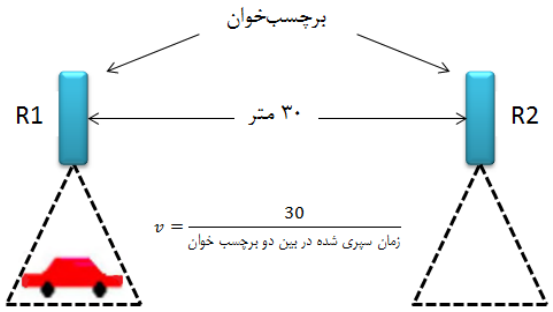
شکل (۷): چیدمان برچسب‌خوان‌ها در چهارراه

۵-۳- سرعت غیر مجاز

اغلب روش‌هایی که امروزه برای تشخیص سرعت استفاده می‌شود، مبتنی بر پدیده شیفت داپلر می‌باشند. با استفاده از تکنولوژی RFID می‌توان دو برچسب‌خوان به فاصله معینی از یکدیگر (D) نصب نمود؛ با عبور وسیله نقلیه از برچسب‌خوان اول زمان T1 برای آن وسیله نقلیه ثبت شده و با عبور از برچسب‌خوان دوم نیز زمان T2 ثبت می‌شود؛ سپس با استفاده از رابطه ساده زیر سرعت وسیله نقلیه محاسبه می‌شود [20]:

$$V = \frac{D}{T2 - T1}$$

سرعت بدست آمده، سرعت متوسط در بین دو برچسب‌خوان است.



شکل (۸): نحوه تشخیص سرعت

۵-۴- طرح ترافیک و مناطق ورود ممنوع

در سطح شهر، مکان‌ها و خیابان‌هایی هستند که یکسری از وسایل نقلیه برحسب نوع آن وسیله، شماره پلاک و غیره اجازه تردد در آن خیابان‌ها را ندارند. در یک دسته بندی می‌توان موارد زیر را درنظر گرفت:

طرح زوج و فرد: در این حالت با ورود هر وسیله نقلیه به محدوده طرح، شناسه آن توسط برچسب‌خوان‌های نصب شده در آن محدوده‌ها خوانده شده و مجاز بودن یا نبودن آن بررسی می‌شود و در صورت مجاز نبودن به تردد در آن محدوده، جریمه وی صادر می‌شود.

خیابان‌های یکطرفه: در این حالت نیاز به نصب دو برچسب‌خوان داریم، یکی در کنار تابلویی که علامت ورود ممنوع را نشان می‌دهد و یکی به فاصله چند متر داخل آن خیابان. بدین ترتیب با وارد شدن وسیله نقلیه به خیابان مذکور، برچسب آن خوانده شده و ذخیره می‌شود؛ حال اگر وسیله نقلیه با وارد شدن به خیابان در جهت خلاف شروع به حرکت نماید، برچسب خوان دوم دوباره برچسب آن را خوانده و تخلف آن وسیله نقلیه ثبت می شود.

خطوط ویژه و تونل‌ها: خطوط ویژه و تونل‌ها از جمله معابری هستند که برای برخی از وسایل نقلیه (موتورها و غیره) ممنوع هستند. در هنگام ورود وسایل نقلیه به تونل، برچسب آنها خوانده شده و نوع وسیله نقلیه مشخص می‌شود و در صورت مجاز نبودن وسیله نقلیه برای ورود به تونل، تخلف آن ثبت می‌شود.

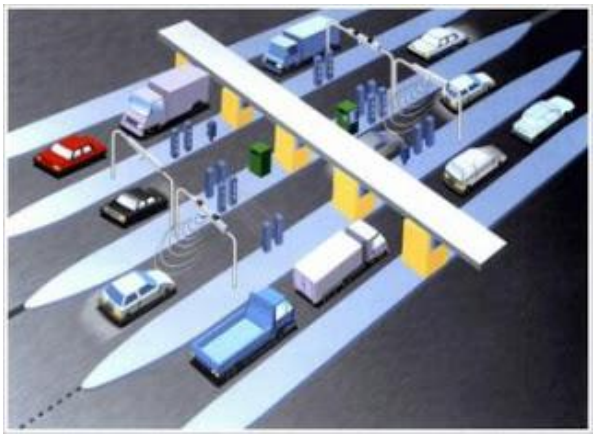
به طور کلی تمامی مکان‌های ورود ممنوع برای وسایل نقلیه مختلف را می‌توان با استفاده از تکنولوژی RFID کنترل کرد.

۵-۵- ردگیری وسایل نقلیه دزدیده شده

در این حالت، مالک وسیله نقلیه به سرقت رفتن خودرو خود را گزارش می‌دهد. حال اگر سارق با آن وسیله نقلیه تردد کند، برچسب‌خوان‌های نصب شده در سطح شهر، شناسه موجود در برچسب آن وسیله نقلیه را خوانده و مسیر حرکت وی مشخص می‌شود و اقدامات لازم برای متوقف کردن آن وسیله نقلیه انجام می‌گیرد (مکان تمامی برچسب‌خوان‌ها مشخص است). در واقع به اینگونه سیستم‌ها، سیستم‌های RTSV گفته می‌شود که در مقاله [5] مورد بررسی قرار گرفته‌اند.

۵-۶- جمع‌آوری عوارض جاده‌ای

از بارزترین کاربردهای RFID در عرصه حمل و نقل، جمع آوری عوارض جاده‌ای است که در دهه ۹۰ در برخی از کشورها پیاده شده و مورد بهره‌برداری قرار گرفته است.



شکل (۹): جمع‌آوری عوارض جاده‌ای [42]

از جمله این سیستم‌ها می‌توان سیستم E-ZPass [40] را نام برد که در ایالات متحده استفاده می‌شود. این سیستم‌ها اصطلاحاً به سیستم‌های ETC[۲۳] معروفند. شرح این سیستم‌ها، فواید و نحوه نصب و راه‌اندازی آنها در مقالات [25-23] آمده است.

به طور خلاصه، روش انجام کار بدین صورت است که برچسب‌خوان‌هایی در محل‌های مورد نظر نصب می‌شود و با عبور هر وسیله نقلیه از کنار آنها برچسب موجود در وسایل نقلیه خوانده شده و میزان عوارض برحسب نوع وسیله نقلیه از حساب مالک آن کسر شده و یا ماهانه قبض صادر شده و برای مالک خودرو ارسال می‌شود. خودروهای فاقد برچسب نیز توسط روش‌هایی مانند استفاده از دوربین شناسایی می‌شوند. مهمترین مزیت این سیستم‌ها، عدم نیاز به توقف وسایل نقلیه برای پرداخت عوارض است.

۵-۷- کنترل ورود و خروج پارکینگ‌ها

طبق [27] مهمترین مشکلات پارکینگ‌ها (اعم از گاراژ، پارکینگ کنارجاده‌ای و غیره)، هزینه پرداختی توسط هر وسیله نقلیه است. انجام این کار به شیوه دستی با مشکلاتی همراه است که در [27] بیان شده است. پیاده‌سازی سیستم‌های پارکینگ با استفاده از تکنولوژی RFID، باعث افزایش سرعت، امنیت و سهولت رانندگان می‌شود [26].

برای کنترل کردن ورود/خروج وسایل نقلیه به/از پارکینگ‌ها، در نقاط ورود/خروج پارکینگ‌ها، برچسب‌خوان هایی نصب می‌شود. در هنگام ورود خودرو به پارکینگ، شناسه آن خوانده شده و با ثبت زمان ورود، خودرو وارد پارکینگ می‌شود. در هنگام خروج نیز، شناسه خودرو بار دیگر توسط برچسب‌خوان خروجی خوانده شده و برحسب مدت زمانی که در پارکینگ بوده، هزینه مورد نظر از حساب مالک کسر می‌شود.

شکل (۱۰): ورود/ خروج به/از پارکینگ [43]

۵-۸- تشخیص ازدحام در مسیرهای مختلف و پیشنهاد مسیر به رانندگان

این نوع کاربرد در مقالات [6] و [20] بررسی شده است. در [20] وضعیت ترافیکی مسیرهای مختلف (شلوغ یا خلوت بودن آنها) بر اساس متوسط سرعت خودروهای آن مسیرها مشخص می‌شود (برای بدست آوردن سرعت متوسط مسیرهای مختلف از همان رابطه بخش ۵-۳ استفاده شده است)؛ و سپس با این فرض که دستگاهی بر روی خودروها نصب شده است که شامل نقشه شهر بوده و امکان برقراری ارتباط با سیستم مرکزی را از هر طریقی (مثلاً سیستم GPRS، WiMax و غیره) دارد، راننده می تواند مقصد خود را به سیستم مرکزی اعلام کند و سیستم مرکزی نیز با اعمال یک الگوریتم مسیریابی (مثلاً دایجسترا) سریعترین مسیر را به وی پیشنهاد می کند

۵-۹- ردیابی وسایل نقلیه عمومی (اتوبوس و تاکسی)

برخی از وسایل نقلیه مانند تاکسی‌ها ملزم به ارائه خدمت به مسافرین در ساعت‌های مشخصی از روز می‌باشند (به علت سهمیه بنزین بیشتر و غیره). با بکار بستن RFID در جاده‌های مختلف شهری، می‌توان تمام تاکسی‌ها را ردیابی کرده و از محل آنها باخبر شد؛ در نتیجه می‌توان تاکسی‌هایی را که در مسیر خاص خود حضور ندارند، شناسایی کرد.

همچنین می‌توان در ایستگاهای اتوبوس (مترو) موجود در شهر تابلوهایی را نصب کرد که محل اتوبوس‌ها (قطارها) ی مختلف را در مسیر مورد نظر نمایش دهد؛ بدین ترتیب مسافرین می‌توانند از زمان ورود اتوبوس (قطار) ها به ایستگاه‌ها مطلع شوند. این سیستم می‌تواند به اندازه‌ای کارا و قابل اطمینان باشد که حتی دارندگان وسایل نقلیه شخصی مایل باشند که از وسایل نقلیه عمومی برای رفت‌وآمد خود استفاده کنند [22].

۵-۱۰- سایر کاربردها

تا بدین‌جا با برخی از مهمترین کاربردهای RFID در مدیریت ترافیک شهری آشنا شدیم. کاربردهایی مانند گردش به چپ ممنوع، گردش به راست ممنوع، پارک ممنوع، چراغ ترافیکی هوشمند و غیره را نیز می‌توان با نصب یک، دو و یا چند برچسب‌خوان پیاده کرد.

برخی از کاربردها نیز نیازمند تجهیزات بیشتری برای محقق شدن می‌باشند. مانند تشخیص سبقت ممنوع که در [29] بررسی شده است. در این مقاله برچسب‌خوان‌ها در وسایل نقلیه قرار گرفته و تمامی وسایل نقلیه نیز به حسگرهایی برای تشخیص خط، مجهز هستند. برچسب‌ها نیز بر روی تابلوهای سبقت ممنوع نصب شده‌اند. هر وسیله نقلیه با خواندن برچسب نصب شده بر روی تابلوی «سبقت ممنوع»، متوجه می‌شود که در منطقه سبقت ممنوع قرار گرفته است؛ حال با انجام سبقت غیرمجاز، حسگرهای نصب شده در خودرو، آن را تشخیص داده و با استفاده از روشهایی مانند GPRS[۲۴] و غیره تخلف در سیستم مرکزی ثبت می‌شود. با خواندن برچسب‌هایی که در تابلوهای «پایان منطقه سبقت ممنوع» نصب شده است، حسگرها غیرفعال می‌شوند.

تکنولوژی RFID می‌تواند با هدف کاهش خطا در ورود اطلاعات نیز به کار برده شود. شاید همه ما شنیده باشیم که برخی از رانندگان خودروهای سواری به دلیل نداشتن کلاه ایمنی جریمه شده‌اند. این موارد در واقع از خطاهایی که در ورود دستی اطلاعات به سیستم رخ می‌دهد، ناشی می‌شوند. برای حذف اینگونه خطاها می‌توان یک برچسب‌خوان دستی به هر یک از مأموران پلیس داد؛ حال اگر راننده‌ای تخلفی انجام دهد. مأموران راهنمایی و رانندگی می‌توانند با استفاده از آن برچسب‌خوان دستی، شناسه موجود در برچسب وسایل نقلیه را خوانده و جریمه مورد نظر را اعمال کنند و سپس جریمه ثبت شده با استفاده از تجهیزات بیسیم و یا سیمی به سیستم مرکزی منتقل شود. [30]

۶- چالش‌های مرتبط با پیاده‌سازی سیستم

در پیاده‌سازی هر سیستم RFID، یکی از مهمترین سوالاتی که باید پاسخ داده شود، این است که: «**چه نوع برچسبی را برای کاربرد خود انتخاب کنیم؟**». انتخاب نوع برچسب در کاربرد مدیریت ترافیک، وابسته به مسائلی چون محدوده خواندن مورد نیاز، سرعت وسایل نقلیه، پروتکل‌های anti-collision، مسائل مرتبط با کانال بیسیم در محیط شهری، نرخ خواندن و غیره است. در ادامه به بررسی این مسائل پرداخته و در انتها یک جمع‌بندی ارائه می‌کنیم.

۶-۱- محدوده خواندن

مهمترین معیاری که در بررسی کارایی سیستم‌های RFID مطرح است، محدوده‌ی خواندن برچسب‌ها است [46]. در کاربردهای مدیریت ترافیک شهری، فاصله بین برچسب و برچسب‌خوان نسبتاً زیاد است. طبق [8]، برچسب‌های فعال به دلیل فراهم کردن محدوده‌ی خواندن بیشتر، گزینه مناسبی برای این نوع کاربردها هستند. اما در کاربردهایی مانند تشخیص توقف ممنوع، پارکینگ و موارد مشابه دیگر به دلیل کم بودن فاصله بین برچسب و برچسب‌خوان و همچنین کم بودن تعداد برچسب‌هایی که باید همزمان خوانده شود، برچسب غیرفعال نیز می‌تواند مورد استفاده قرار گیرد. و اما محدوده خواندن از یکسری عوامل تأثیر می‌پذیرد که در ادامه آنها را مورد بررسی قرار می‌دهیم.

۶-۱-۱- محدودیت‌های برچسب

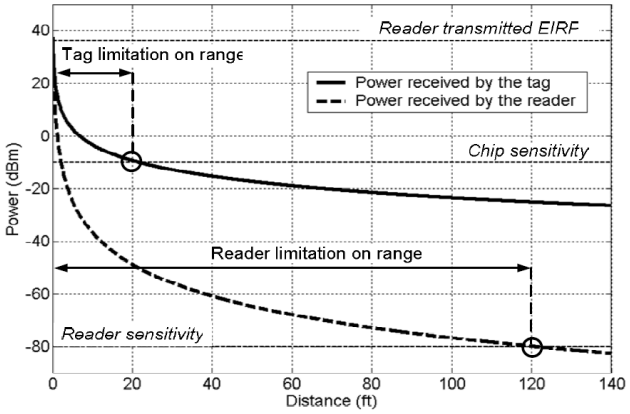
مهمترین عامل محدود کننده برچسب، **آستانه حساسیت**[۲۵] گیرنده آن است که حداقل میزان توان سیگنال دریافتی را مشخص می‌کند که برچسب برای کدگشایی سیگنال دریافتی نیاز دارد. هر چقدر این آستانه کمتر باشد حساسیت بالا بوده و لذا محدوده خواندن بیشتری خواهیم داشت. یکی دیگر از عوامل مهم، **بهرهٔ آنتن**[۲۶] برچسب است که در صورت افزایش آن، محدوده خواندن افزایش می‌یابد. عامل دیگر **پلاریزاسیون**[۲۷] آنتن برچسب است که باید با آنتن برچسب‌خوان مطابقت داشته باشند. در صورت استفاده از پلاریزاسیون دایره‌ای در برچسب‌خوان، می‌توان این حساسیت را نادیده گرفت ولی نیاز به افزایش دو برابری توان ارسال در برچسب‌خوان خواهیم داشت. همچنین **تطبیق امپدانس**[۲۸] بین آنتن برچسب و تراشه داخلی آن نیز از عواملی است که مستقیماً بر محدوده خواندن تأثیر می‌گذارد؛ چرا که این عامل مشخص می‌کند چه مقدار از توان دریافتی قابل جذب بوده و چه مقدار انعکاس می‌یابد [46,47].

۶-۱-۲- محیط انتشار

افت مسیر[۲۹] مهمترین عاملی است که در هر محیط انتشاری وجود دارد و بسته به نوع محیط، ضریب آن از ۱ تا ۴ متغیر است (در فضای آزاد ضریب افت مسیر، ۲ است). این عامل در واقع بیانگر میزان افت توان سیگنال به دلیل طی مسافت است [46]. علاوه بر این تداخل امواج، پدیده محو شدگی، انتشار چند مسیری و به طور کلی هر عاملی که در انتقال بیسیم تأثیرگذار است در محدوده‌ی خواندن برچسب‌های RFID نیز تأثیر می‌گذارد.

۶-۱-۳- محدودیت‌های برچسب‌خوان

توان ارسالی و **بهرهٔ آنتن** برچسب‌خوان از دیگر عوامل مهم تأثیرگذار بر محدوده خواندن است که با افزایش آنها می‌توان محدوده خواندن را افزایش داد. اما استانداردهای جهانی مانع از افزایش بیش از حد توان ارسالی می‌شوند [46,47]. **آستانه حساسیت برچسب‌خوان** نیز در محدده خواندن اثر گذار است ولی معمولاً این آستانه بسیار پایین است (حساسیت بالا است) و همانظور که از شکل (11) پیداست، معمولاً آستانه حساسیت برچسب است که محدودیت اساسی ایجاد می‌کند و عامل تعیین کننده در محدوده خواندن می‌شود [46]. در شکل (۱۱)، محدوده خواندن ۲۰ فوت خواهد بود که براساس آستانه حساسیت برچسب تعیین می‌شود.



شکل (۱۱): توان دریافتی برچسب و برچسب‌خوان در مسافت‌های مختلف [46].
(Reader Sensitivity: -80 dBm, Tag Sensitivity: -10 dBm)

و در نهایت اینکه **فرکانس** استفاده شده در انتقال اطلاعات بین برچسب و برچسب‌خوان نیز بر محدوده خواندن برچسب اثر می‌گذارد.

۶-۲- سرعت وسایل نقلیه

برای خوانده شدن برچسب، حداکثر سرعتی که یک وسیله نقلیه باید داشته باشد از رابطه زیر محاسبه می‌شود [8]:

$$\text{سرعت خودرو} = \frac{\text{محدوده خواندن (متر)}}{\text{کل زمان مورد نیاز برای انتقال شناسه (ثانیه)}}$$

هر چه سرعت وسایل نقلیه بالاتر باشد، زمان کمتری برای خواندن برچسب وجود خواهد داشت. طبق رابطه فوق برای جبران این محدودیت زمان، می‌توان محدوده خواندن برچسب را افزایش داد تا برچسب مدت زمان بیشتری در محدوده برچسب‌خوان باقی بماند.
یک راهکار برای افزایش محدوده خواندن، استفاده از بر چسب‌های فعال است [32]. همچنین، استفاده از پروتکل anti-collision مناسب، داشتن نرخ بیت بالا، استفاده از برچسب‌ها و برچسب‌خوان‌هایی که مدارات سریعتری دارند [32]، باعث کاهش زمان انتقال شناسه شده و خوانده شدن برچسب در سرعت‌های بالاتر را امکان‌پذیر می‌کند.
البته متحرک بودن برچسب‌ها، باعث می‌شود که شیفت داپلر نیز اثر گذار شود [32] که خود شیفت داپلر هم از عواملی چون سرعت وسایل نقلیه و فرکانس کاری تأثیر می‌پذیرد. هر چه سرعت وسیله نقلیه بالاتر باشد، شیفت داپلر نیز بیشتر می‌شود. از طرفی استفاده از فرکانس‌های بالا، می‌تواند باعث کاهش شیفت داپلر شود. پس در کاربردهای مدیریت ترافیک شهری استفاده از فرکانس‌های بالا توصیه می‌شود.

۶-۳- پروتکل Anti-Collision مورد استفاده

یکی دیگر از مسائل مهم، پروتکل anti-collision استفاده شده در لایه MAC می‌باشد. پروتکل FSA و نسخه‌های بهبود یافته آن به علت تأخیر و مصرف انرژی کمتر برای کاربرد مدیریت ترافیک شهری بهتر از پروتکل‌های مبتنی بر درخت می‌باشند [15]. در [16] یک پروتکل مبتنی بر ALOHAی بهبود یافته برای شناسایی وسایل نقلیه پیشنهاد شده است. ایده پیشهنادی این مقاله، گروه‌بندی برچسب‌ها بر اساس فاصله تا برچسب‌خوان و سپس خواندن به ترتیب هر گروه است.

۶-۴- مسائل مرتبط با کانال بیسیم

به مشکلات موجود در کانال بیسیم، در بخش ۶-۱-۲ مختصراً اشاره شد. در کاربرد مدیریت ترافیک شهری، جاده‌هایی با موانع مختلف و اشیای فلزی (خودروها)، محیط انتشار را تشکیل می‌دهند. فراهم کردن لینک ارتباطی قابل اطمینان با استفاده از برچسب‌های غیرفعالی که با توان بسیار ناچیزی شناسه خود را ارسال می‌کنند، در چنین محیط‌هایی دشوار است. لذا با استفاده از برچسب‌های فعال که توان بیشتری در اختیار دارند، می‌توان تا حدی با این مشکلات مقابله کرد. دقت کنید که وجود اکثر این مشکلات به دلیل باند باریک بودن سیگنال است و همانطور که قبلاً بیان شد، راه‌حل اساسی مقابله با این مشکلات، استفاده از برچسب‌های UWB است.

۶-۵- نرخ خواندن

به تعداد برچسب‌های خوانده شده در واحد زمان، نرخ خواندن گفته می‌شود که این عامل اساساً از تمامی مسائل مطرح شده‌ی قبلی تأثیر می‌پذیرد. به عبارت دیگر ارائه یک راهکار مناسب برای مشکلات مطرح شده قبلی، باعث افزایش نرخ خواندن می‌شود. البته نرخ بیت ارسالی نیز بر روی این عامل تأثیر گذار است. همانطور که قبلاً گفته شد، برچسب‌هایی که در فرکانس‌های بالا (UHF و Microwave) کار می‌کنند. به علت داشتن پهنای باند بیشتر نرخ بیت بالاتری را نیز می‌توانند فراهم کنند. لذا تعداد برچسب‌های بیشتری نیز قابل خواندن می‌شود.

۶-۶- سایر چالش‌ها

۶-۶-۱- مکان نصب برچسب در وسایل نقلیه

برچسب باید در مکانی نصب شود که مانعی برای ارتباط بین برچسب و برچسب‌خوان وجود نداشته باشد. همچنین تا حد ممکن نباید برچسب بر روی سطوح فلزی نصب شود؛ چرا که دریافت سیگنال را با مشکلاتی مواجه می‌کند. طبق [48] نصب برچسب بر روی یک سطح فلزی می‌تواند به میزان قابل توجهی باعث کاهش بهره آنتن برچسب گردد. اندازه و وزن برچسب نیز از دیگر عواملی است که کار نصب را تحت‌الشعاع قرار می‌دهد. برچسب‌های فعال دارای اندازه بزرگتری هستند، لذا نسبت به برچسب‌های غیرفعال، مشکلات بیشتری برای نصب دارند.

۶-۶-۲- نصب برچسب‌خوان‌ها در جاده‌ها

در سطح شهر معابر مختلفی با شرایط متفاوت وجود دارد که نصب برچسب‌خوان‌ها در هریک از آنها مسائل خاص خود را خواهد داشت. به عنوان مثال در هنگام نصب برچسب‌خوان در جاده‌های عریض باید دقت شود که نقطه کوری باقی نماند؛ و یا در خطوط ویژه نیز باید فقط خط ویژه تحت پوشش قرار گیرد تا به اشتباه برچسب وسیله نقلیه دیگری که در مسیر صحیح خود در حال رانندگی است، خوانده نشود.

۶-۶-۳- نگهداری بلند مدت سیستم

ازآنجائیکه برچسب‌های غیرفعال دارای باتری نمی‌باشند، به دلیل عدم نیاز به مسائل مرتبط با بررسی وضعیت باتری‌ها، هزینه نگهداری سیستم‌های مبتنی بر برچسب‌های غیرفعال نسبت به سایر برچسب‌ها پایین‌تر است.

همچنین با فرض رعایت امنیت فیزیکی برچسب‌ها (نسبت به دما، رطوبت و کلیه آسیب‌های احتمالی)، طول عمر برچسب‌های غیرفعال هم نسبت به سایر برچسب‌ها بیشتر است.

۶-۷- جمع‌بندی

به طور کلی می‌توان نتیجه گرفت که استفاده از برچسب فعال باعث ارتباط بیسیم بهتر بین برچسب و برچسب‌خوان می‌شود و لینک ارتباطی را قابل اطمینان‌تر می‌کند. ولی هزینه‌ی زیادی را هم به سیستم تحمیل می‌کنند. همین امر باعث می‌شود که به سراغ سایر برچسب‌ها نیز برویم و آنها را نیز مورد بررسی قرار دهیم. جدول (۲) برخی از مهمترین کاربردها را به همراه نوع برچسب قابل استفاده نمایش می‌دهد.

کاربرد	نوع برچسب قابل استفاده			توضیحات
	Passive	Semi-Passive	Active	
جمع آوری عوارض	؟	؟	✓	سیستم‌های EZPass [40] و FastPass [41] از برچسب نوع Active استفاده کرده‌اند.
طرح‌های ترافیکی	؟	؟	✓	به نکته ۱ مراجعه شود.
ردیابی وسایل نقلیه عمومی	✓	✓	✓	سیستم پیشنهاد شده در [22] از نوع Passive استفاده کرده است.
عبور از چراغ قرمز	؟	؟	✓	به نکته ۱ مراجعه شود.
سرعت غیر مجاز	؟	؟	✓	به نکته ۱ مراجعه شود.
سبقت غیر مجاز	-	-	-	نیاز به تجهیزات بیشتری دارد (قراردادن برچسب‌خوان و حسگر بر روی وسیله نقلیه [29])
توقف ممنوع	✓	✓	✓	وسیله نقلیه متوقف است و فاصله بین برچسب و برچسب‌خوان کم است. لذا برچسب نوع Passive نیز می‌تواند استفاده شود. در [30] که یک سیستم جریمه الکترونیکی پیشنهاد شده است، از برچسب نوع Passive استفاده شده است.
پارکینگ	✓	✓	✓	
سیستم جریمه الکترونیکی	✓	✓	✓	

نکته ۱: علامت ؟ به این معنی است که تعداد وسایل نقلیه موجود در محدوده، سرعت آنها، پروتکل‌های استفاده شده و در کل تمامی چالش‌های مطرح شده، در انتخاب این نوع برچسب تأثیرگذار است و بسته به شرایط می‌تواند قابل استفاده باشد و یا نباشد.

نکته ۲: برچسب‌های نوع Active تقریباً برای تمامی کاربردهای مدیریت ترافیک قابل استفاده است ولی هزینه بالایی دارد.

نکته ۳: در کاربردهایی که هر سه نوع برچسب قابل استفاده است، انتخاب برچسب Passive به علت مقرون به صرفه بودن توصیه می‌شود.

نکته مهم: اکثر کارهای انجام شده در حوزه مدیریت ترافیک شهری، برچسب‌های Active را به دلیل محدوده خواندن بیشتر و داشتن لینک ارتباطی مطمئن‌تر، گزینه مناسبی برای این کابرد دانسته‌اند.

جدول (۲): کاربردهای مختلف مدیریت ترافیک شهری و نوع برچسب مناسب

۷ – نتیجه

مشکلاتی که افزایش بیش از حد وسایل نقلیه در کلان‌شهرهایی چون تهران به بار می‌آورد بر هیچ‌کس پوشیده نیست. بررسی‌های انجام شده در این مقاله حاکی از آن است که تکنولوژی RFID، پتانسیل بالایی برای وارد شدن به حیطه مدیریت ترافیک شهری و حل این مشکلات دارد؛ و این مهم با مجهز کردن وسایل نقلیه به یک برچسب RFID ممکن می‌شود. سهولت، هزینه نسبتاً پایین، کاهش نیاز بـه نیـروی انسانی، تنوع کاربردهایی که RFID در مدیریت حمل و نقل به ارمغان می‌آورد و همچنین کنترل منظم و یکپارچه بر رعایت قـوانین، ایـن تکنولوژی را بر سایر تکنولوژی‌ها ارجح می‌کند.

در این مقاله بررسی مختصری نیز از روش‌های غیر RFID برای مدیریت ترافیک بعمل آمد؛ با توضیحاتی که در این مقاله ارائه شـد، دیدیم که می‌توان تکنولوژی RFID را در کنار سایر سیستم‌ها نیز، بکار بست. بدین ترتیب نیازی به انجـام تغییـرات کلـی در سیسـتم مدیریت حمل و نقل فعلی نخواهد بود.

و اما استفاده از RFID برای مدیریت ترافیک شهری نیز همانند هر سیستم دیگری با چالش‌هایی روبروست. در انتهـای مقالـه ایـن چالش‌ها با محوریت «نوع برچسب» مورد بررسی قرار گرفته شـد و راه‌کارهایی نیز ارائه شد.

مراجع

[1] Syed Ahson, Mohammad Ilyas, "*RFID Handbook Applications, Technology, Security and Privacy*", CRC press, 2008.

[2] Albert Lozano-Nieto, "*RFID Design Fundamentals and Applications*", CRC Press, 2011.

[3] Khan, A.A.; Yakzan, A.I.E.; Ali, M.; , "*Radio Frequency Identification (RFID) Based Toll Collection System*," Computational Intelligence, Third International Conference on Communication Systems and Networks (CICSyN), 2011, vol., no., pp.103-107, 26-28 July 2011.

[4] Yunus A. kathawala; Benjamin Tueck, "*The use of RFID for Traffic Management*", International Journal of Technology, Policy and Management 2008 - Vol. 8, No.2 pp. 111 – 125.

[5] Pandit, Anala Aniruddha, Jyot Talreja, and A. K. Mundra. "*RFID Tracking System for Vehicles (RTSV)*." In Computational Intelligence, Communication Systems and Networks, 2009. CICSYN'09. First International Conference on, pp. 160-165. IEEE, 2009.

[6] Hongjian, Wang, and Tang Yuelin. "*RFID Technology Applied to Monitor Vehicle in Highway*." In Digital Manufacturing and Automation (ICDMA), 2012 Third International Conference on, pp. 736-739. IEEE, 2012.

[7] Prof Somprakash Bandyopadhyay, "*Traffic Congestion Management Using RFID & Wireless Technologies*", Technical Report, Indian Institute of Management Calcutta, May 2010.

[8] Yu, Minghe, Dapeng Zhang, Yurong Cheng, and Mingshun Wang. "*An RFID electronic tag based automatic vehicle identification system for traffic iot applications*." In Control and Decision Conference (CCDC), 2011 Chinese, pp. 4192-4197. IEEE, 2011.

[9] Lee, Eun-Kyu, Young Min Yoo, Chan Gook Park, Minsoo Kim, and Mario Gerla. "*Installation and evaluation of rfid readers on moving vehicles*." In Proceedings of the sixth ACM international workshop on VehiculAr InterNETworking, pp. 99-108. ACM, 2009.

[10] M. Ward, R. van Kranenburg and G. Backhouse. "*RFID: Frequency, standards, adoption and innovation*", JISC Technology and Standards Watch, 2006.

[11] Lee, Kin Keung, Hakon A. Hjortland, and Tor Sverre Lande. "*IR-UWB technology on next generation RFID systems*." In NORCHIP, 2011, pp. 1-4. IEEE, 2011.

[12] Jain, Shweta, and Samir R. Das. "*Collision avoidance in a dense RFID network*." In Proceedings of the 1st international workshop on Wireless network testbeds, experimental evaluation & characterization, pp. 49-56. ACM, 2006.

[13] Cheng, Tao, and Li Jin. "*Analysis and Simulation of RFID Anti-collision Algorithms*." International Conference on Advanced Communication Technology (ICACT), 2007.

[14] Park, Jongho, Min Young Chung, and Tae-Jin Lee. "*Identification of rfid tags in framed-slotted ALOHA with tag estimation and binary splitting*." In Communications and Electronics, 2006. ICCE'06. First International Conference on, pp. 368-372. IEEE, 2006.

[15] Yoon, Won-Ju, Sang-Hwa Chung, and Seong-Joon Lee. "*Implementation and performance evaluation of an active RFID system for fast tag collection*." Computer Communications 31, no. 17 (2008): 4107-4116.

[16] Kai, Zhang, and Tang Niuming. "*Study on improved slotted ALOHA algorithm to vehicle identification*." In Information Science and Technology (ICIST), 2011 International Conference on, pp. 870-875. IEEE, 2011.

[17] Yousefi, Saleh, Mahmoud Siadat Mousavi, and Mahmood Fathy. "*Vehicular ad hoc networks (VANETs): challenges and perspectives*." In ITS Telecommunications Proceedings, 2006 6th International Conference on, pp. 761-766. IEEE, 2006.

[18] Cheng, Wei, Xiuzhen Cheng, Min Song, Biao Chen, and Wendy W. Zhao. "*On the Design and Deployment of RFID Assisted Navigation Systems for VANETs*." Parallel and Distributed Systems, IEEE Transactions on 23, no. 7 (2012): 1267-1274.

[19] Munilla, J., A. Ortiz, and A. Peinado. "*What can RFID do for VANETs? A cryptographic point of view*." In Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on, pp. 1-4. IEEE, 2010.

[20] Manikonda, Prithvinath, Anil Kumar Yerrapragada, and Sai Sasank Annasamudram. "*Intelligent traffic management system*." In Sustainable Utilization and Development in Engineering and Technology (STUDENT), 2011 IEEE Conference on, pp. 119-122. IEEE, 2011.

[21] Singh, Harpal, Satinder Jeet Singh, and Ravinder Pal Singh. "*Red Light Violation Detection Using RFID*", 2012.

[22] Assaf, M.H.; Williams, K.M.; , "*RFID for optimisation of public transportation system*," Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2011 Seventh International Conference on , vol., no., pp.407-412, 6-9 Dec. 2011

[23] Feng, Zhihui, Yanjie Zhu, Pengtao Xue, and Mingjie Li. "*Design and realization of expressway vehicle path recognition and ETC system based on RFID*." In Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on, vol. 7, pp. 86-90. IEEE, 2010.

[24] Zhengang, Ren, and Gao Yingbo. "*Design of electronic toll collection system in expressway based on RFID*." In Environmental Science and Information Application Technology, 2009. ESIAT 2009. International Conference on, vol. 3, pp. 779-782. IEEE, 2009.
[25] Kamarulazizi, Khadijah, Dr Widad Ismail, M. EL MARRAKI, G. AL HAGRI, A. KARTIT, M. EL MARRAKI, A. RADI et al. "*Electronic Toll Collection System Using Passive Rfid Technology*." Journal of Theoretical and Applied Information Technology (JATIT) 22, no. 2 (2011).
[26] Pala, Zeydin, and Nihat Inanc. "*Smart parking applications using RFID technology*." In RFID Eurasia, 2007 1st Annual, pp. 1-3. IEEE, 2007.
[27] Ostojic, G., S. Stankovski, M. Lazarevic, and Vukica Jovanovic. "*Implementation of RFID technology in parking lot access control system*." In RFID Eurasia, 2007 1st Annual, pp. 1-5. IEEE, 2007.
[28] Dong, Yuejun, Yan Cui, Fangfang Shan, and Zhiyong Dong. "*The research and application in intelligent parking management system of RFID anti-collision algorithm*." In Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on, pp. 468-471. IEEE, 2011.
[29] Nejati, Omid. "*Smart Recording of Traffic Violations via M-RFID*." In Wireless Communications, Networking and Mobile Computing (WiCOM), 2011 7th International Conference on, pp. 1-4. IEEE, 2011.
[30] Vali Derhami, Mohammad Ghasemzadeh, Mina AmirSadeghi, "*An Innovation in Using RFID Technology in Automation of Traffic Fine Issue and Management*", Studies in Informatics and Control, ISSN 1220-1766, vol. 19 (4), pp. 403-410, 2010.
[31] Kai, Zhang, and Tang Niuming. "*Study on improved slotted ALOHA algorithm to vehicle identification*." In Information Science and Technology (ICIST), 2011 International Conference on, pp. 870-875. IEEE, 2011.
[32] Zhang, Xiaoqiang; Lakafosis, Vasileios; Traille, Anya; Tentzeris, Manos M.; , "*Performance analysis of fast-moving RFID tags in state-of-the-art high-speed railway systems*," RFID-Technology and Applications (RFID-TA), 2010 IEEE International Conference on , vol., no., pp.281-285, 17-19 June 2010.
[33] Jo, Minho, Hee Yong Youn, Si-Ho Cha, and Hyunseung Choo. "*Mobile RFID tag detection influence factors and prediction of tag detectability*." Sensors Journal, IEEE 9, no. 2 (2009): 112-119.
[34] www.rfidjournal.com, last accessed, november 2012.
[35] en.wikipedia.org/wiki/RFID
[36] Intel Corporation, "*Ultra-Wideband (UWB Technology)*" White Paper, 2005
[37] www.marshproducts.com, last accessed, november 2012.
[38] www.iwatchsystems.com/technical/2010/06/26/loop-detector/, last accessed, november 2012.
[39] Chintalacheruvu, Naveen, and Venkatesan Muthukumar. "*Video Based Vehicle Detection and its Application in Intelligent Transportation Systems*." Journal of Transportation Technologies 2, no. 4 (2012): 305-314.
[40] http://en.wikipedia.org/wiki/E-ZPass, last accessed, november 2012.
[41] http://en.wikipedia.org/wiki/SunPass, last accessed, november 2012.
[42] http://hanuriworld.blogspot.com, last accessed, november 2012.
[43] www.ercls.com, last accessed, november 2012.
[44] Cornel Turcu, "*Current Trends and Challenges in RFID*", InTech, 2012.
[45] Chawla, Vipul, and Dong Sam Ha. "*An overview of passive RFID*."Communications Magazine, *IEEE* 45, no. 9 (2007): 11-17.
[46] Nikitin, Pavel V., and K. V. S. Rao. "*Performance limitations of passive UHF RFID systems*." In Proceedings of the IEEE Antennas and Propagation Symposium, pp. 1011-1014. 2006.
[47] Keskilammi, M., L. Sydänheimo, and M. Kivikoski. "*Radio frequency technology for automated manufacturing and logistics control. Part 1: passive RFID systems and the effects of antenna parameters on operational distance*." The International Journal of Advanced Manufacturing Technology 21, no. 10 (2003): 769-774.
[48] Griffin, Joshua D., Gregory Durgin, Andreas Haldi, and Bernard Kippelen. "*Radio link budgets for 915 MHz RFID antennas placed on various objects*." In Proc. 2005 Texas Wireless Symposium, Austin, TX, pp. 22-26. 2005.

زیرنویس‌ها

[1] Identify: Friend or Foe
[2] Auto ID
[3] Intelligent Transportation Systems
[4] RF Tag (Radio Ferequency Tag)
[5] Reader
[6] Passive Tags
[7] Active Tags
[8] Reliability
[9] Semi-Passive Tags
[10] Low Ferequency
[11] High Ferequency
[12] Ultra High Ferequency
[13] Ultra Wideband
[14] Federal Communications Commission
[15] Time Division
[16] Framed-Slotted ALOHA
[17] International Standads Organizations
[18] Write Once Read Many
[19] Inductive Loops
[20] Vehicular Ad-hoc NETworks
[21] Vehicle to Vehicle
[22] Vehicle to Infrastructure
[23] Electronic Toll Collection
[24] General Packet Radio Service
[25] Sensitivity Threshold
[26] Antenna Gain
[27] Polarization
[28] Impedance Match
[29] Path Loss

مطالعه‌ی محدوده‌ی بار قابل تحمیل به شبکه‌ی SIP با حفظ پایداری آن و قابلیت فیلتر کردن بار

وحید قاسم‌خانی[۱]، سید وحید ازهری[۲]

[۱]دانشجوی کارشناسی ارشد
vahidgk@gmail.com
[۲]استاد راهنما
azharivs@iust.ac.ir

چکیده

هدف اصلی ما در این سمینار، بررسی روش‌های مختلف کنترل اضافه‌بار[۱] در شبکه‌های SIP[۲] و ارایه‌ی یک دسته‌بندی مناسب از آن‌ها در جهت مطالعه‌ی بار قابل تحمیل به شبکه‌ی تحت کنترل هر دسته از این روش‌ها می‌باشد. در حقیقت، هدف ما بررسی مسایل مربوط به محدوده‌ی پایداری[۳] شبکه می‌باشد. پایداری را به طور عرفی می‌توان عدم رفتن شبکه به حالتی که نتواند از اثرات منفی (گذردهی نزدیک به صفر و تاخیر بالا) اضافه‌بار رهایی یابد و یا عدم نوسانات شدید در گذردهی معنی کرد.

برای کنترل اضافه‌بار شبکه‌های SIP روش‌های مختلفی ارایه شده است؛ اغلب آن‌ها تا حد زیادی کارایی سیستم را در شرایط اضافه‌بار نسبتا سنگین بالا نگه می‌دارند. اما برای هریک از این روش‌ها یک محدوده‌ی اضافه‌باری وجود دارد که اگر بار ورودی بیش از آن شود، روش با شکست مواجه خواهد شد. این شکست منجر به ناپایداری گذردهی در حالت باقی ماندن اضافه‌بار و یا حتی بعد از رهایی از اضافه‌بار خواهد شد.

کلمات کلیدی

اضافه‌بار، SIP، کنترل اضافه‌بار، پایداری، فیلتر کردن بار

۱- مقدمه

SIP یک پروتکل لایه‌ی کاربرد برای ایجاد، نگه‌داری و اتمام جلسه‌های چندرسانه‌ای می‌باشد که توسط IETF[۴] استاندارد شده است [۱]. این پروتکل همچنین توسط موسساتی مانند ITU[۵]، ۳GPP[۶] و ETSI[۷] به عنوان هسته‌ی اصلی پروتکل سیگنالینگ شبکه‌های نسل آینده برای ارایه‌ی خدماتی چون VoIP[۸]، کنفرانس‌های صوتی و تصویری، ویدئو، VoD[۹] و پیام‌های فوری[۱۰] به رسمیت شناخته شده است. برای پذیرش تجاری SIP در چنین سطح فراگیری، لازمست زیرساخت یک شبکه‌ی SIP از قابلیت اطمینان بسیار بالایی برخوردار باشد. وقتی که شبکه به دلایلی مانند طراحی نامناسب شبکه، شروع به کار یکباره تعداد زیادی از کاربران، ازدحام آنی، خرابی عناصر شبکه و حملات که در [۲] بررسی شده‌اند تحت اضافه‌بار قرار می‌گیرد، کارایی‌اش به شدت افت می‌کند، لذا لازمست که به مساله‌ی اضافه‌بار نگاهی دقیق و ویژه شود.

سرورهای SIP نیز همانند سایر سرورها از اضافه بار رنج می‌برند. مساله‌ی اضافه‌بار در SIP به دلایلی مانند معماری چندگامی[۱۱] SIP با مسیریابی سطح کاربرد و هزینه‌ی بالای رد کردن درخواست‌ها، متفاوت‌تر و مهم‌تر از اضافه‌بار در سایر سرورها می‌باشد [۱۴]. در [۵] یک دسته از متریک‌ها برای ارزیابی و محک[۱۲] کارایی انواع سرورهای SIP مطرح شده است. گذردهی‌مفید[۱۳] و تاخیر برقراری تماس مهم‌ترین معیارهایی هستند که به طور گسترده مورد مطالعه قرار گرفته‌اند. وقتی سرور تحت اضافه‌بار قرار می‌گیرد گذردهی آن به طور چشم‌گیری کاهش می‌یابد و حتی به صفر نیز می‌رسد. علاوه بر این، تاخیر برقراری تماس نیز غیر قابل تحمل می‌شود [۳،۸،۲۷،۲۸].

سرورهای SIP اغلب به اندازه‌ی نیازهای کاربران طراحی و مهندسی می‌شوند و اقتصادی و یا امکان‌پذیر نمی‌باشد که آن‌ها را برای یک ترافیک با قله‌ی[۱۴] خیلی بالا طراحی کرد. از آنجایی که به طور کامل نمی‌توان جلوی اضافه‌بار را گرفت، مهم است که SIP را با یک مکانیزم که بتواند به طور موثر اضافه‌بار را کنترل کند مجهز کرد [۲]. یک روش ساده این است که سرور درخواست‌های اضافی را نادیده بگیرد [۱]. متاسفانه نادیده گرفته شدن درخواست‌ها باعث ارسال مجدد آن‌ها می‌شود و در نتیجه بار سرور تحت اضافه‌بار کم نمی‌شود که هیچ، زیادتر نیز می‌شود. شبیه‌سازی‌ها نشان می‌دهد که این روش ساده می‌تواند باعث ناپایداری سیستم شود؛ در این شرایط سرور حتی

پس از اتمام اضافه‌بار ناپایدار خواهد ماند حتی اگر مقدار بار کمتر از ظرفیتش نیز بشود [۳].

جهت رفع مشکل فوق، استاندارد SIP یک مکانیزم کنترل اضافه‌بار از طریق پاسخ کد ۵۰۳ فراهم آورده است [۱]. سرور از بالادستی[۱۵]اش می‌خواهد برای مدتی که در سراسر[۱۶] Retry-After مشخص شده است از ارسال دست بردارد و دوباره بعد از زمان مشخص شده به ارسال بپردازد. این الگوی فرستادن/نفرستادن، به دلیل ایجاد نوسان در ترافیک، کارایی شبکه‌ی SIP را خراب می‌کند [۲،۳]. اگر از Retry-After استفاده نشود، هر درخواست به طور مجزا رد می‌شود و درخواست‌های بعدی می‌توانند ارسال شوند. بدین ترتیب جلوی الگوی فرستادن/نفرستادن گرفته می‌شود و می‌تواند کارایی بهتری را ارایه دهد [۳]. اما در اضافه‌بارهای سنگین، در نهایت به دلیل صرف اغلب منابع پردازشی برای رد کردن درخواست‌ها منجر به گذردهی صفر خواهد شد. جهت مدیریت موثر اضافه‌بار در شبکه‌های SIP، روش‌های زیادی که به سه دسته‌ی کلی محلی[۱۷]، گام به گام[۱۸] و انتها به انتها[۱۹] تقسیم می‌شوند [۴]، وجود دارد که در بخش ۴ بررسی می‌شوند.

کنترل اضافه‌بار می‌تواند در هر یک از لایه‌های پیونده داده‌ها، شبکه، انتقال و کاربرد پیاده‌سازی شود [۶]. از آنجایی SIP یک پروتکل لایه‌ی کاربرد می‌باشد، اغلب روش‌های کنترل اضافه‌بار در لایه‌ی کاربرد می‌باشند. ولی روش‌هایی نیز برای کنترل اضافه‌بار در لایه‌ی انتقال به هنگام استفاده از TCP، جهت بهبود رویه‌ی کنترل جریان TCP برای کنترل اضافه‌بار SIP مورد بررسی قرار گرفته است [۷].

درست است که اغلب روش‌ها تا حد زیادی جلوی اثرات منفی اضافه‌بار را می‌گیرند، اما همانطور که در بخش‌های بعدی نشان خواهیم داد، عواملی که همگی منجر به واکنش با تاخیر به اضافه‌بار خواهند شد، موجب ناپایداری خواهند شد.

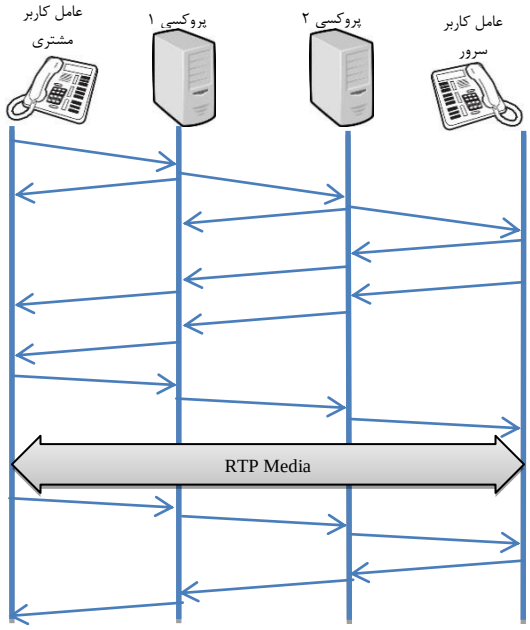
هدف ما در این سمینار مطالعه‌ی بحث پایداری و مسایل تاثیرگذار روی محدوده‌ی پایداری می‌باشد. مواردی مانند نوع بازخورد، الگوریتم کنترل اضافه‌بار، معیار تشخیص اضافه‌بار، میزان همکاری عناصر شبکه با یکدیگر و فاصله‌ی بین کنترل‌ها، روی تاخیر واکنش به اضافه‌بار و در نتیجه روی پایداری تاثیرگذار می‌باشند. اینکه با چه دقتی و با چه سرعتی اضافه‌بار و یا خروج از اضافه‌بار را تشخیص داد و نسبت به هر کدام یک عمل مناسب انجام داد، مواردی هستند که می‌توانند باعث هرچه بیشتر شدن محدوده‌ی پایداری بشوند. در نهایت، ما می‌خواهیم با پیشگیری از وقوع یک اضافه‌بار سنگین، محدوده‌ی پایداری را بهبود ببخشیم. ما دریافتیم که اگر یک کنترل کننده بالاتر از روش‌های کنترل موجود قرار داد که وظیفه‌اش یافتن یک مقدار حداکثر بار با توجه به ظرفیت اسمی سرور و تاخیر واکنش روش کنترل محلی است؛ می‌توان از وقوع یک اضافه‌بار سنگین که در تمامی روش‌های موجود به دلیل تاخیر واکنش به اضافه‌بار اتفاق می‌افتد پیشگیری کرد. بدین صورت که هر پایین‌دستی[۲۰] فیلترهایی را تعریف کند که حاوی مقداری هستند که پایین‌دستی در حالت حداکثر می‌تواند آن مقدار بار بپذیرد. بار ارسالی توسط بالادستی، تحت هر شرایطی نباید از آن مقدار تجاوز کند. مهم‌ترین بحث در اینجا، یافتن مناسب این مقادیر می‌باشد.

در ادامه، بخش دوم به معرفی کوتاهی از SIP می‌پردازد. بخش سوم مساله‌ی اضافه‌بار در SIP را بیان می‌کند. بخش چهارم یک دسته‌بندی کامل از روش‌های کنترل اضافه‌بار ارایه می‌کند. بخش پنجم به توضیح مساله پایداری می‌پردازد. بخش ششم مسایل تاثیرگذار روی پایداری را با جزییات کامل بررسی می‌کند. بخش هفتم به معرفی چندین الگوریتم از روش‌های مختلف می‌پردازد. در نهایت، بخش هفتم شامل طرح پیشنهادی برای پیشگیری از ناپایداری می‌باشد.

۲- آشنایی با SIP

۲-۱- سیگنالینگ برقراری ارتباط

عناصر اصلی شبکه‌ی SIP عبارتند از: ۱) عامل‌های کاربر که با نقش عامل کاربر مشتری[۲۱] (UAC) و عامل کاربر سرور[۲۲] (UAS) اقدام به برقراری ارتباط می‌کنند. ۲) سرورهای‌ثبت‌کننده[۲۳] که وظیفه ثبت و نگه‌داری اطلاعات کاربران را دارند. ۳) سرورهای پروکسی که وظیفه‌ی مسیریابی پیام‌های سیگنالینگ بین عامل‌های کاربر را برعهده دارند.



شکل (۱): سیگنالینگ برقراری جلسه در SIP

SIP می‌تواند روی لایه‌های انتقال UDP، TCP، SCTP[۲۴]کار کند.

شکل ۱ مراحل مختلف سیگنالینگ لازم برای ایجاد و اتمام یک نشست ساده را نشان می‌دهد. عامل کاربر مشتری یک پیام INVITE از طریق پروکسی‌های درون مسیر به عامل کاربر سرور ارسال می‌کند. هر پروکسی، پذیرش درخواست را با ارسال پاسخ Trying ۱۰۰، به گام قبلی اطلاع می‌دهد تا زمانبند ارسال مجددش را متوقف سازد. این عمل با فرض پیکربندی حالتمند[۲۵] پروکسی‌ها صورت می‌گیرد. وقتی که عامل کاربر سرور درخواست INVITE را دریافت کرد، یک پاسخ

Ringing ۱۸۰ به سمت کاربر مشتری از همان مسیری که درخواست آمده بود ارسال می‌کند. همچنین، بعد از پذیرش تماس، یک پاسخ OK ۲۰۰ به سمت کاربر مشتری ارسال می‌کند. در نهایت، کاربر مشتری یک پاسخ ACK به سمت مشتری سرور ارسال می‌کند. بعد از رسیدن ACK به مشتری سرور، دست‌دهی‌سه‌طرفه[۲۶] به پایان می‌رسد و جلسه برقرار می‌شود. حال، داده‌های چندرسانه‌ای بدون دخالت پروکسی‌ها،توسط پروتکل‌های انتقال بلادرنگ مانند RTP[۲۷] بین کاربران ردوبدل می‌شود. در نهایت، یکی از کاربران با ارسال درخواست BYE خواستار قطع ارتباط می‌شود. طرف دیگر با ارسال پیام ۲۰۰ OK جلسه را خاتمه می‌دهد.

تراکنش تماس موفق، به ارسال موفق تمام پیام‌های مربوط به دو تراکنش INVITE و BYE گفته می‌شود. تاخیر برقراری تماس، فاصله بین ارسال درخواست INVITE توسط کاربر مشتری و دریافت پیام ۲۰۰ OK مربوط به INVITE توسط کاربر مشتری می‌باشد [۵].

۲-۲- زمان‌بندها

SIP برای مقابله با فقدان بسته‌ها[۲۸] از مکانیزم ارسال مجدد در لایه‌ی کاربرد استفاده می‌کند. این کار را با نگه‌داشتن زمانبندهای مختلف و فعال کردن آن‌ها به هنگام ارسال هر درخواست انجام می‌دهد. در صورت عدم دریافت پاسخ مربوطه در مدت زمان مشخص، درخواست، ارسال مجدد می‌شود. در کل، درخواست‌ها به دو دسته‌ی INVITE و non-INVITE تقسیم می‌شوند. زمانبند A برای INVITE و زمانبند E برای non-INVITEها (BYE) می‌باشد. زمانبند دیگری برای ۲۰۰ OK مربوط به درخواست INVITE وجود دارد که از لحاظ خواص شبیه به تایمر E می‌باشد و جهت کنترل دریافت ACK می‌باشد. [۱] زمانبند T_1 را برای ارسال مجدد تعریف می‌کند، یک درخواست INVITE برای بار اول بعد از گذشت T_1 ثانیه ارسال مجدد می‌شود، بعد از هر ارسال مجدد، این بازه‌ی زمانی ۲ برابر می‌شود. فرستنده وقتی که یک پاسخ Trying ۱۰۰ دریافت کند و یا اینکه از اولین ارسال مجدد $T1\times64$ ثانیه گذشته باشد، ارسال مجدد را متوقف خواهد ساخت. مقدار پیش‌فرض T_1 برابر با ۵۰۰ میلی‌ثانیه می‌باشد.

برای non-INVITEها زمانبند T_2 نیز با مقدار پیش‌فرض ۴ ثانیه تعریف می‌شود. درخواست‌ها در ابتدا بعد از T_1 ثانیه ارسال مجدد می‌شوند و بعد از هر ارسال مجدد، مقدار بازه‌ی زمانی دو برابر می‌شود. با رسیدن این مقدار به T_2، فاصله بین ارسال مجددها T_2 خواهد ماند. با دریافت پاسخ یا گذشت $T_1\times64$ ثانیه ارسال مجدد قطع خواهد شد.

این زمانبندها بجز زمانبند مربوط به OK ۲۰۰، اغلب در زمان استفاده از پروتکل‌های انتقال اطمینان‌پذیر غیرفعال می‌باشد.

گرچه ارسال مجددها برای اطمینان ارتباط در حالاتی که لایه‌ی انتقال غیرقابل اطمینان است مفید است، اما در شرایطی باعث افزایش بار و کاهش سنگین کارایی شبکه SIP می‌شود [۳،۶]. با این حال، برای برقراری سریع تماس، برنامه‌ها اغلب از پروتکل غیرقابل اطمینان UDP استفاده می‌کنند و محبوبیت بیشتری نسبت به TCP دارد.

۳- مساله‌ی اضافه بار در SIP

اضافه‌بار زمانی رخ می‌دهد که نرخ ورود درخواست‌ها به یکی از سرورهای SIP بیشتر از ظرفیت منابع آن باشد. این منابع می‌توانند شامل ظرفیت پردازشی پردازشگر، حافظه، پهنای باند شبکه، ورودی/خروجی یا منابع دیسک باشد. وقتی سرور تحت اضافه‌بار قرار می‌گیرد کارایی آن به شدت کاهش می‌یابد، تا حدی که گذردهی مفید آن به صفر می‌رسد [۳،۶،۸]. این وضعیت زمانی که از UDP به عنوان پروتکل لایه‌ی انتقال استفاده شود بدتر است [۳]، زیرا SIP برای اطمینان ارتباط از چندین زمانبند ارسال مجدد استفاده می‌کند و در صورتیکه پاسخ پیام‌ها به موقع نرسد درخواست‌ها ارسال مجدد خواهند شد. در نتیجه، بار سرور تشدید خواهد شد.

هنگام اضافه‌بار، پیام‌ها یا دچار تاخیر زیاد می‌شوند و یا به دلیل پرشدن بافر سرریز خواهند شد. پیام‌هایی که دچار تاخیر زیاد شده‌اند در نهایت منجر به تماس ناموفق خواهند شد، زیرا در صورت عدم برقراری تماس طی ۳۲ ثانیه، تماس نادیده گرفته خواهد شد [۱]. حتی اگر فرض کنیم که هیچ ارسال مجددی نباشد، سرریز شدن بافر باعث از دست رفتن پیام‌های مربوط به یک تماس در حال برقراری و در نهایت منجر به تماس ناموفق خواهد شد [۶،۹،۱۰،۱۱]. از طرف دیگر، حتی در صورت عدم سرریز بافر و با فرض بزرگ بودن آن، صف پر از ارسال مجدد و جلساتِ منقضی شده خواهد شد که هیچ ارزشی ندارند.

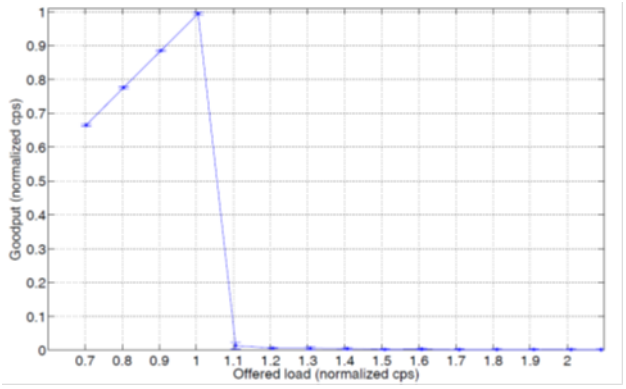
شکل ۲ کارایی سرور تحت اضافه‌بار را به نمایش می‌گذارد. محور افقی نشانگر نرخ بار ورودی و محور عمودی نشانگر گذردهی سرور می‌باشد. هر دو محور بر روی ظرفیت سرور نرمال شده‌اند. همانطور که مشاهده می‌شود، بعد از اینکه بار ورودی بیش از مقدار بار قابل پردازش سرور می‌شود، با یک افت ناگهانی، گذردهی به صفر می‌رسد.

این نمودار با فرض ادامه‌دار بودن اضافه‌بار ترسیم شده است، اما اگر اضافه‌بار موقتی باشد، گذردهی نیز بطور موقتی شدیدا افول خواهد کرد. حتی در شرایطی امکان دارد با برطرف شدن اضافه‌بار، سیستم به گذردهی قبلی خود بازنگردد. این پدیده بدلیل تجمع بیش از حد پیام‌ها در بافر سرور تحت اضافه‌بار است که حتی تحت بار عادی باعث افزایش تاخیر تماس و در نتیجه ارسال مجددهای بیهوده می‌شود [۳].

هنگام وقوع اضافه‌بار در یک سرور، تعداد زیادی از درخواست‌ها توسط بالادستی‌هایش ارسال مجدد می‌شوند. در چنین شرایطی نه تنها بار روی سرور تحت اضافه‌بار تشدید می‌شود، بلکه منجر به تحت اضافه‌بار قرار گرفتن بالادستی‌هایش نیز می‌شود. در این حالت اضافه‌بار می‌تواند به کل شبکه گسترش یابد و کل شبکه را از کار بیندازد [۳].

عواملی مانند پیکربندی‌های مختلف سرورها، اندازه‌ی حافظه‌ی گیرنده و پروتکل انتقال مورد استفاده، روی کارایی سرورها در شرایط بار عادی و اضافه‌بار موثر می‌باشند [۳،۲۹،۳۰،۳۸،۳۹،۴۰،۴۱]. درست است که برخی موارد نتایج بهتری را ارایه می‌دهند، ولی هیچ یک از آن‌ها، از رفتن به یک ناپایداری عمیق جلوگیری نخواهند کرد. لذا نیاز به روشی موثر برای مقابله با اضافه‌بار و پیشگیری از ناپایداری وجود

دارد. بدون یک روش کنترل اضافه‌بار، کارایی به شدت کاهش پیدا خواهد کرد [۳،۱۲،۱۳،۱۴].



شکل (۲): رفتار سرور تحت بارهای مختلف [۲۷]

۴- دسته‌بندی روش‌های کنترل اضافه‌بار

۴-۱- از دید ساختار

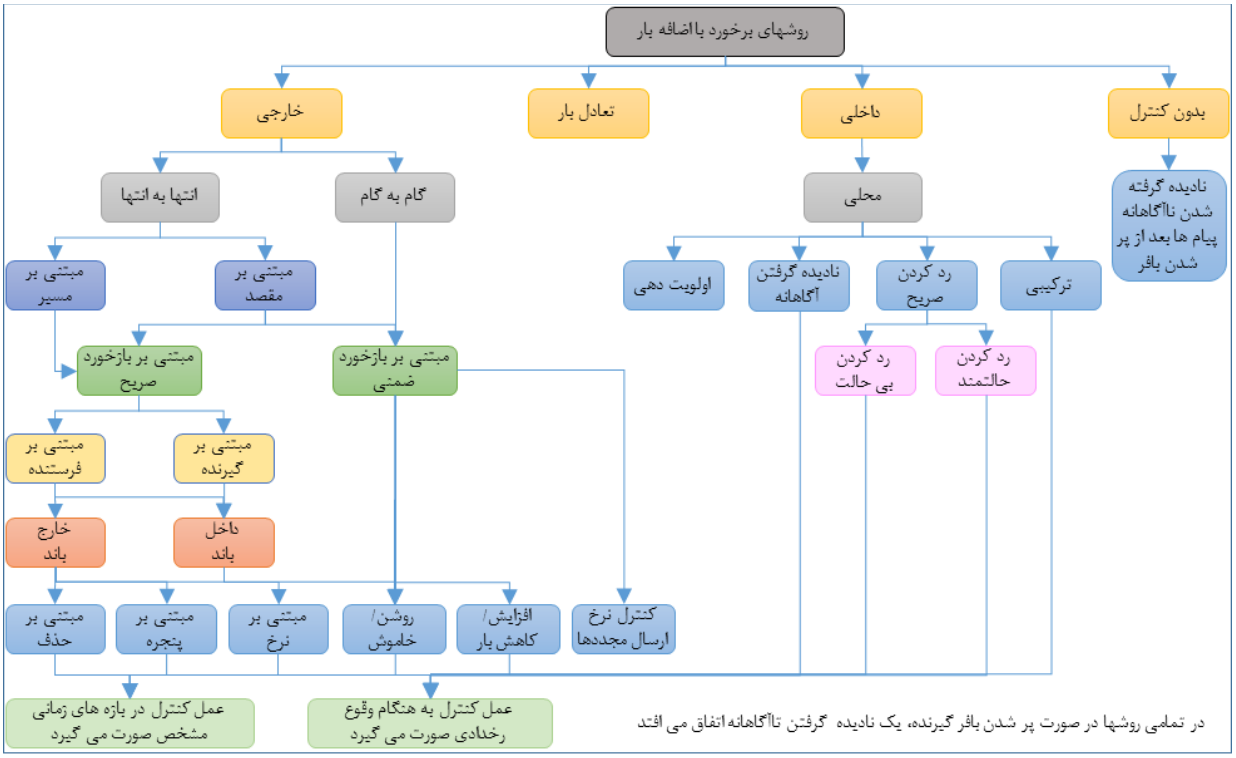
روش‌های کنترل اضافه‌بار از جنبه‌های مختلفی از یکدیگر متفاوت هستند. برای کاهش پیچیدگی بیان آن‌ها، ابتدا آن‌ها را در شکل ۳ به صورت درختی نمایش می‌دهیم و سپس به توضیح مختصر هریک می‌پردازیم. اگر در شکل ۳ از سمت ریشه به سمت یک برگ حرکت کنیم، موارد مشاهده شده در مسیر، یک روش را ایجاد می‌کنند.

۴-۱-۱- بدون کنترل

سرور تحت اضافه‌بار هیچ‌گونه تمهیدی برای مقابله با اضافه‌بار نمی‌اندیشد. در این حالت، زمانی که بافر گیرنده به دلیل اضافه‌بار پر شد، پیام‌های اضافی به سادگی نادیده گرفته خواهند شد که منجر به مشکلات مطرح شده در بخش ۳ می‌شود.

۴-۱-۲- کنترل داخلی

ایده‌ی پایه‌ی کنترل داخلی این است که سرور در صورت تشخیص اضافه‌بار، به جای نادیده گرفتن درخواست‌ها، به صورت محلی شروع به رد کردن درخواست‌ها با استفاده از پاسخ کد ۵۰۳ بدون سرایند Retry-After نماید. منطق پشت کنترل اضافه‌بار محلی این است که، رد کردن یک درخواست، از لحاظ مصرف منابع به صرفه‌تر از سرویس‌دهی به آن است، زیرا یک درخواست به دنبال خود پردازش شش پیام دیگر را نیز در بر دارد. همچنین به دلیل رد کردن آشکار درخواست‌های اضافی، درخواست‌ها ارسال مجدد نخواهند شد و در نتیجه بار سرور تقویت نخواهد شد. بنابراین در روش کنترل محلی، نیاز به همکاری بین سرورها وجود ندارد. اما این روش تنها در اضافه‌بارهای سبک خوب و کارا می‌باشد. در بارهای سنگین، یک سرور تحت اضافه‌بارِ مجهز به این روش، در نهایت بیشتر منابع پردازشی‌اش را صرف رد کردن درخواست‌ها خواهد نمود؛ که منجر به گذردهی پایین و تاخیر غیرقابل قبول خواهد شد [۳،۱۲،۱۴]. برای تسریع بخشیدن به پروسه‌ی رد کردن، می‌توان از رد کردن بدون حالت که هیچ حالتی برای درخواست رد شده نگه‌داری نمی‌شود استفاده کرد [۳،۱۲]. این روش باید به همراه دیگر مکانیزم‌ها و به عنوان آخرین نقطه‌ی دفاعی استفاده شود. اما تنها راه درمان اضافه‌بار، زمانی که اضافه‌بار به طور مستقیم توسط کاربران ایجاد می‌شود است [۱۳،۱۴]. روش‌های دیگری نیز برای کنترل محلی وجود دارد. مانند اولویت‌دهی



شکل (۳): روش‌های مختلف برخورد با اضافه‌بار از دید ساختار

به پیام‌ها به صورت‌های مختلف [۱۱]، رد کردن آگاهانه بدون تولید پاسخ صریح برای فرستنده [۱۰] و ترکیبی از این سه [۹].

۴-۱-۳- کنترل خارجی

این نوع کنترل خود بر دو نوع گام به گام و انتها به انتها تقسیم می‌شود، که مشخص کننده‌ی درجه همکاری عناصر درگیر در کنترل اضافه‌بار می‌باشند [۳،۱۳]. در نوع گام به گام، یک حلقه‌ی کنترل مجزا بین سرور و هر سرور بالادستی‌اش می‌باشد. در نوع انتها به انتها، حلقه‌ی کنترل در حالت ایده‌آل بین هر سرور ورودی (منبع ترافیک) و هر سرور مقصد می‌باشد. یعنی در سرور مبدا تصمیم گرفته می‌شود که آیا پیام به مقصدی مشخص فرستاده شود یا خیر. این دو روش را به صورت‌های مختلفی می‌توان پیاده‌سازی کرد که در شکل ۲ به طور کامل آمده‌اند و در این جا توضیح مختصری از هریک ارایه می‌دهیم:

مبتنی بر بازخورد صریح: پایین‌دستی اطلاعاتی را که از وضعیت خود به دست آورده است، برای بالادستی‌هایش ارسال می‌کند. بالا دستی نیز بر اساس این اطلاعات، تصمیمات لازم را اتخاذ می‌کند.

مبتنی بر بازخورد ضمنی[۲۹]: بالادستی(ها) به صورت ضمنی از وجود یا عدم وجود اضافه‌بار در سرور(های) پایین‌دستی مطلع می‌شود. برای مثال، از تاخیر برقراری تماس به عنوان یک معیار استفاده کند [۱۲].

مبتنی بر گیرنده: این حالت زمانی که از روش مبتنی بر بازخورد استفاده می‌شود امکان‌پذیر است. در این حالت، عنصر نظارتگر و کنترل کننده در سمت گیرنده می‌باشند. گیرنده باید با توجه به وضعیتش که از نظارتگر به دست می‌آورد، یک مقدار برای بالادستی ارسال کند، این مقدار نوع برخورد با اضافه‌بار را مشخص می‌کند.

مبتنی بر فرستنده: این حالت نیز هنگام استفاده از بازخورد صریح معنا دارد. عنصر نظارتگر در سمت گیرنده و کنترل کننده در فرستنده می‌باشد. فرستنده با استفاده از اطلاعات صریح دریافتی از نظارتگر، مقداری را برای تنظیم بار ارسالی‌اش به دست می‌آورد.

داخل باند: این اصطلاح مربوط به روش ارسال بازخوردهای صریح می‌باشد، منظور از داخل باند، ارسال بازخوردها از طریق خود پیام‌های SIP می‌باشد. پیش‌نویس‌های [۱۵،۱۶،۱۷] یک مکانیزم جهت قرار دادن بازخوردها درون پاسخ‌های SIP ارایه کرده‌اند.

خارج باند: در این حالت، اطلاعات بازخورد از طریق پیام‌های دیگری همانند مکانیزم معرفی شده در [۱۸] ارسال می‌شوند.

مبتنی بر مسیر: این حالت که در روش انتها به انتهای مبتنی بر بازخورد صریح امکان پذیر است، به معنای شناسایی مسیرهای مختلف به یک سرور مقصد می‌باشد. یعنی سرور ورودی بداند که چه مسیری از مسیرهای رسیدن به مقصد در حال تجربه‌ی اضافه‌بار می‌باشد. این روش به دلیل آگاهی‌اش از تمامی مسیرها و تصمیم‌گیری بهتر و دقیق می‌تواند از منابع شبکه به خوبی استفاده نماید، ولی دارای پیچیدگی و سربار بسیار زیادی می‌باشد که پیاده‌سازی آن را دشوار می‌کند [۲۸].

مبتنی بر مقصد: این حالت در روش انتها به انتها و در هر دو حالت بازخورد صریح و ضمنی امکان‌پذیر است. فرستنده فقط کافی است که بداند مسیری دچار اضافه‌بار است، یعنی لازم به شناسایی تمام مسیرها نمی‌باشد. این روش دارای سربار و پیچیدگی کمی است [۲۸].

اطلاعاتی که الگوریتم‌های کنترل اضافه‌بار در هر روشی خروجی می‌دهند، می‌تواند بر اساس یکی از موارد زیر باشد.

مبتنی بر حذف[۳۰]: در این روش، کنترل کننده یک درصد را به عنوان خروجی می‌دهد. بالادستی هر مقدار بار که به مقصد پایین‌دستی به دستش می‌رسد را فقط به مقدار درصد مشخص شده به پایین‌دستی هدایت کند [۱۳،۱۴].

مبتنی بر نرخ[۳۱]: کنترل کننده مقداری را که نشانگر ظرفیت پایین‌دستی در آن برهه از زمان کنترل می‌باشد، به دست می‌آورد. این مقدار در روش مبتنی بر گیرنده، می‌تواند به صورت مساوی و یا نامساوی بین بالادستی‌ها تقسیم شود [۱۳،۱۴،۱۹،۲۳،۲۷].

مبتنی بر پنجره[۳۲]: خروجی الگوریتم کنترل یک عدد به عنوان اندازه پنجره می‌باشد. وقتی تعداد درخواست‌های بی‌پاسخ به اندازه‌ی پنجره رسید، ارسال متوقف می‌شود. با هر پاسخ، شمارنده کاهش و با هر ارسال درخواست، شمارنده یک واحد افزایش می‌یابد [۱۳،۱۴،۸،۱۹].

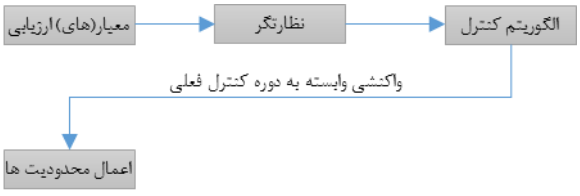
روشن/خاموش[۳۳]: خروجی الگوریتم کنترل یک مدت زمان می‌باشد. در این مدت زمان نباید هیچ درخواستی به گیرنده ارسال شود. بعد از اتمام مدت زمان، دوباره تمامی درخواست‌ها ارسال می‌شود. مثالی از این روش، ارسال پاسخ کد ۵۰۳ با سرایند Retry-After می‌باشد که در [۳،۱۳،۱۹] مورد بررسی قرار گرفته است.

مبتنی بر زمان[۳۴]: عمل کنترل در بازه‌های زمانی مرتبی صورت می‌گیرد. اندازه‌ی بازه‌ی زمانی باید بزرگتر یا مساوی فاصله‌ی زمانی عمل نظارت باشد [۱۴].

مبتنی بر رخداد[۳۵]: در این حالت عمل کنترل به مجرد وقوع یک رخداد مانند تکمیل سرویس یک جلسه انجام می‌شود [۱۴].

۴-۲- نوع برخورد با اضافه‌بار

۴-۲-۱- روش‌های تشخیص‌دهنده[۳۶]

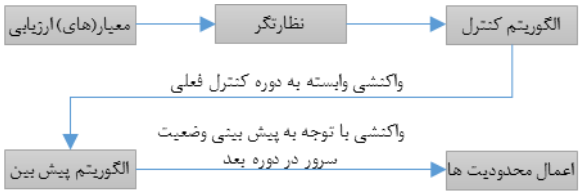


شکل (۴): بلوک دیاگرام روش‌های تشخیص‌دهنده

این گونه روش‌ها مجهز به الگوریتمی هستند که بطور متناوب یا به مجرد اتفاق افتادن یک رخداد، اطلاعات جمع‌آوری شده توسط نظارتگر را مورد تحلیل قرار می‌دهند. خروجی الگوریتم، یک واکنش پویا نسبت به وضعیت فعلی سرور تحت کنترل می‌باشد. واکنش می‌تواند اضافه یا کم کردن بار، اولویت‌دهی و غیره باشد [۸،۹،۱۲،۱۴،۱۹،۲۱،۲۲،۲۳،۲۶،۳۱].

۴-۲-۲- روش‌های پیش‌بینانه[۳۷]

این‌گونه روش‌ها علاوه بر الگوریتم تشخیص اضافه‌بار، مجهز به الگوریتمی هستند که سعی بر پیش‌بینی مقادیر خروجی الگوریتم کنترل برای دوره‌های بعدی می‌کند. در واقع با توجه به خروجی فعلی الگوریتم کنترل، مقداری را برای دوره‌ی بعد پیش‌بینی می‌کند و یک قدم جلوتر از روش‌های تشخیص دهنده‌ی محض می‌باشد و می‌تواند واکنش سریع‌تر و بهتری نشان دهد [۲۷]. میزان قابل پیش‌بینی بودن رفتار ترافیک، دقت و درستی خروجی آن عامل‌هایی هستند که می‌توانند به هر چه بیشتر شدن موفقیت آن کمک کنند. شکل ۵ بلوک دیاگرام این‌گونه روش‌ها را نشان می‌دهد.



شکل (۵): بلوک دیاگرام روش‌های پیشبینانه. با الهام از [۲۷]

۴-۲-۳- روش‌های کاهنده‌ی[۳۸] اضافه‌بار

این‌گونه روش‌ها هیچ‌گونه الگوریتمی جهت تشخیص و پیش‌بینی اضافه‌بار ندارند. تنها کاری که در این روش‌ها صورت می‌گیرد، اعمال سیاست‌های از پیش تعیین شده مانند اولویت دادن به پیام‌ها [۹,۱۱] یا تخصیص مناسب حافظه [۳۰] است که در صورت وقوع اضافه‌بار، تاثیر منفی آن نسبت به حالتی که هیچ کنترل اضافه‌باری وجود ندارد کم‌تر شود. در نتیجه، هیچ واکنشی در زمان وقوع اضافه‌بار وجود ندارد.

۴-۲-۴- روش‌های پیشگیرانه[۳۹]

اگر بتوان رفتار ترافیک شبکه را دانست، می‌توان از وقوع اضافه‌بار تا حدی پیشگیری کرد. برای مثال اگر بدانیم در زمان مشخص تعداد تماس‌ها از یک دامنه مشخص قرار است باعث اضافه‌بار در شبکه شود، می‌توان تعداد تماس‌ها از مبدا مشخص را محدود ساخت. تعادل بار را نیز می‌توان به دلیل سعی آن‌ها جهت پیشگیری و کاهش احتمال وقوع اضافه‌بار، در این دسته قرار داد [۳۵].

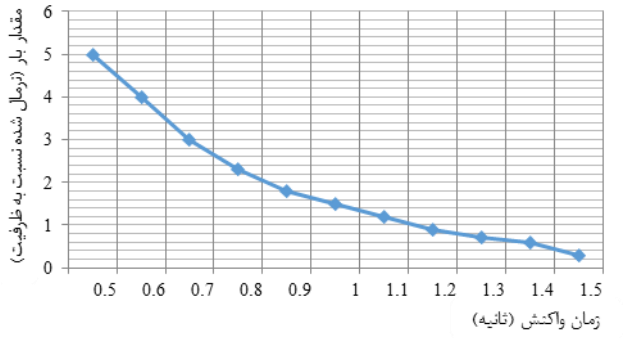
۵- مساله پایداری شبکه‌های SIP

ارائه‌ی تعریف دقیقی از پایداری یک سیستم نیازمند مقدمات ریاضی فراوان است. لذا در اینجا یک تعریف عرفی از پایداری شبکه‌ی SIP ارائه خواهیم داد که با مفاهیم پایداری نیز انطباق داشته باشد. در این نوشتار یک سرور SIP پایدار فرض می‌شود اگر بتواند طول صف درخواستهای خود را محدود نگه دارد. لازم به ذکر است که این لزوما به مفهوم ارائه‌ی حداکثر گذردهی نیست.

پایداری یعنی مقاومت روش کنترل اضافه‌بار در برابر بار اضافی ورودی. عدم مقاومت در برابر این اضافه‌بار باعث ناموفق‌شدن الگوریتم خواهد شد که نتیجه‌ی آن میل کردن به سوی یک گذردهی غیرقابل قبول و نزدیک به صفر و برنگشتن به حالت طبیعی و یا برگشتن در یک مدت زمان طولانی می‌باشد. از سوی دیگر پایداری را می‌توان عدم نوسان حول گذردهی قابل قبول مطرح کرد. یعنی، گذردهی به طور متناوب بین گذردهی خوب و بد نوسان کند. این حالت باعث افت کیفیت و عدم توانایی در ارایه یک سرویس تضمین شده می‌شود.

پایداری به شدت وابسته به سرعت واکنش الگوریتم کنترل اضافه‌بار در برابر افزایش بار ورودی می‌باشد. در [۲۰] نشان داده شده است که زمان واکنش به اضافه‌بار، چه تاثیری روی میزان اضافه‌بار قابل تحمل در شبکه و همچنین زمان همگرایی به حالت پایدار دارد.

اهمیت پایداری تنها برای سروری که از اضافه‌بار محافظت می‌شود نیست. وقتی الگوریتم کنترل اضافه‌بار با لختی به اضافه‌بار واکنش نشان می‌دهد، نه تنها باعث ناپایدار شدن سرور تحت کنترل می‌شود، بلکه به دلیل وجود ارسال مجددها، خود سرور کنترل کننده نیز دچار کاهش ظرفیت می‌شود و ممکن است تحت اضافه‌بار قرار گیرد. در [۲۰] نشان داده شده است که سطوح ارسال مجدد چه تاثیری روی قابلیت تحمل بار سرور خواهد گذاشت.



شکل (۶): رابطه‌ی تاخیر واکنش نسبت به اضافه‌بار و حداکثر بار قابل تحمل

هر الگوریتم با توجه به زمان واکنش‌اش یک محدوده پایداری دارد، به عبارت دیگر، اگر بار ورودی بیش از میزان قابل تحمل‌اش باشد الگوریتم برای مدتی و یا برای همیشه در مواجهه با اضافه‌بار، با شکست مواجه خواهد شد. حتی ممکن است پس از برگشتن بار ورودی به کمتر از ظرفیت سرور نیز، الگوریتم قادر به پایدار نمودن سیستم نباشد [۳،۲۰]. می‌توان شکل ۶ را به عنوان یک نمودار مفهومی برای رابطه‌ی حداکثر زمان واکنش به اضافه‌بار و حداکثر بار قابل تحمل ارایه داد [۲۰]. با توجه به شکل ۶ می‌بینیم که هرچقدر میزان تاخیر واکنش روشی کم‌تر باشد، سرور توانایی اداره کردن اضافه‌بار بیشتری را دارد.

عوامل زیادی روی تاخیر واکنش موثر هستند که در بخش بعدی به بررسی آن‌ها می‌پردازیم.

۶- عوامل موثر روی پایداری

۶-۱- نوع بازخورد

۶-۱-۱- بازخورد ضمنی در مقابل بازخورد صریح

در روش‌های مبتنی بر بازخورد ضمنی، فرستنده بدون اینکه از گیرنده

بازخورد صریحی دریافت کند سعی بر حس اضافه‌بار در پایین‌دستی می‌کند و در صورت لازم بار ارسالی را کاهش می‌دهد [۴]. در نتیجه، در صورت تحت اضافه‌بار قرار گرفتن گیرنده، فرستنده از این پدیده به طور ضمنی مطلع می‌شود و از ارسال درخواست‌ها و بدتر کردن شرایط اضافه‌بار پیشگیری می‌کند.

مهم‌ترین مزیت روش مبتنی بر بازخورد ضمنی این است که حتی پیچیده‌ترین اضافه‌بارها را نیز تشخیص خواهد داد چون نیاز به تولید بازخورد خاصی ندارد و از روی تاخیر مطلب را می‌فهمد. در شرایطی ممکن است که منابع پردازشی و حافظه‌ای کافی در اختیار باشد ولی اضافه‌بار به دلایل دیگری رخ دهد. در این شرایط، از طریق بازخوردهای صریح که اغلب بر اساس اندازه‌گیری روی پردازشگر و حافظه به دست می‌آیند، نمی‌توان اضافه‌بار را تشخیص داد ولی بازخورد ضمنی این اضافه‌بار را می‌تواند تشخیص دهد. بعلاوه، در توپولوژی‌ها و معماری‌های پیچیده‌ی شبکه‌ی SIP که عناصر غیر SIP نیز وجود دارد، عناصر غیر SIP امکان تولید بازخورد را ندارند.

مزیت دیگر روش‌های ضمنی، عدم نیاز به ارسال داخل یا خارج باند که به ترتیب موجب تغییر در پروتکل و ارسال ترافیک اضافه به شبکه می‌شوند، می‌باشد. لذا مشکل فقدان بازخوردها نیز وجود ندارد.

مزیت مهم دیگر بازخورد ضمنی، عدم تحمیل یک سربار پردازشی سنگین به سرور تحت اضافه‌بار برای عمل نظارت و ایجاد اطلاعات کنترلی توسط الگوریتم کنترل می‌باشد [۲۶]. لذا این امر نیز یک مساله مهم در افزایش محدوده‌ی پایداری سرور تحت اضافه‌بار می‌باشد.

علیرغم مزایای فوق، بازخورد ضمنی تاخیر واکنشش زیاد است و برای معماری‌هایی که تشخیص اضافه بار در آن‌ها بسادگی با بازخورد صریح قابل انجام است، باعث کوچک شدن محدوده‌ی بار قابل تحمل و پایداری سیستم خواهد شد.

مزیت روش‌های مبتنی بر بازخورد صریح اینست که سرور تحت اضافه‌بار خود به طور مستقیم و با توجه به وضعیت فعلی‌اش محدودیت‌ها را محاسبه و با تاخیر کمتری به بالادستی‌ها اعلام می‌کند. ولی بزرگترین مشکل این روش‌ها، سربار پردازشی روی سرور تحت اضافه‌بار، عدم ارسال موفقیت‌آمیز محدودیت‌ها در اضافه‌بارهای سنگین و یا دریافت آن‌ها با تاخیر زیاد در فرستنده می‌باشد [۲۶]. این عوامل موجب افزایش تاخیر واکنش به اضافه‌بار، بدتر کردن شرایط اضافه‌بار و کاهش محدوده‌ی بار قابل تحمل و پایداری سیستم خواهد شد. در روش‌های مبتنی بر فرستنده، تنها سربار نظارت، روی سرور تحت اضافه‌بار می‌باشد، ولی سربار کنترل بر روی فرستنده می‌باشد [۲۲].

۶-۲- نوع الگوریتم کنترل مورد استفاده

۶-۲-۱- مبتنی بر حذف

در این روش‌ها هرگونه تغییر در بار ورودی به سرعت روی بار ارسالی تاثیر خواهد گذاشت. زیرا برای بار ارسالی یک سقف نمی‌توان تعیین کرد [۱۳،۱۷]. در نتیجه، افزایش و یا کاهش‌های ناگهانی باعث نوسان در گذردهی خواهد شد. اگر بار ورودی به شدت افزایش یابد سرور برای مدتی تحت اضافه‌بار سنگین قرار می‌گیرد. بنابراین، بعد از مدتی بالادستی از اضافه‌بار مطلع خواهد شد و درصد بار ارسالی را کاهش خواهد داد. بعد از مدتی که شرایط عادی گشت، باز درصد ارسالی افزایش خواهد یافت. این الگو می‌تواند تکرار شود و موجب نوسان شدید در گذردهی شود. نمونه‌ای از ناپایداری به صورت نوسان در [۱۲] جایی که یک الگوریتم OCC بررسی شده است، آمده است. این روش به شدت به تاخیر واکنش وابسته می‌باشد و در صورت ادامه‌دار بودن اضافه‌بار سنگین می‌تواند منجر به ناپایداری عمیق شود.

۶-۲-۲- مبتنی بر نرخ

در این روش تضمین می‌شود که یک سقف برای مقدار بار ارسالی تعیین می‌شود [۱۳،۱۷،۱۴،۱۹،۲۳،۲۷]. نکته‌ی قابل توجه در این روش این است که این روش نسبت به افزایش سریع و یا معمول بار مقاوم‌تر می‌باشد. لذا انتظار می‌رود که کمتر باعث نوسانات و عدم ناپایداری شود. ولی احتمال آن هنوز زیاد می‌باشد. زیرا همانطور که در بخش ۵ گفته شد، عامل اصلی در بحث میزان پایداری هر روشی تاخیر واکنش به اضافه‌بار می‌باشد. یعنی ممکن است نیاز به کاهش بار باشد ولی به دلیل تاخیر اعمال محدودیت جدید، نرخ غیر قابل قبولی ارسال شود و شرایط را بدتر نماید. در نتیجه، این روش نیز مشکلات روش قبل در حالات مشابه را با شدت کم‌تری دارا می‌باشد.

۶-۲-۳- مبتنی بر پنجره

این روش دارای یک مکانیزم خود بازدارندگی[۴۰] می‌باشد. یعنی بعد از مدتی حتی در صورت هرگونه تاخیر و عدم دریافت بازخورد، از ارسال خودداری خواهد کرد. عیب عمده‌ی دو روش قبلی این است که در هر دوره‌ی کنترل مقداری که برای نرخ و درصد به دست می‌آید تا دوره‌ی بعدی اعمال می‌شود. یعنی اگر بین فاصله‌ی کنترل فعلی و کنترل بعدی، حتی شبکه در حال تجربه‌ی اضافه‌بار سنگین باشد، این مقادیر اعمال خواهند شد و شرایط را بدتر خواهند کرد. اما در این روش به جای نرخ و درصد، تعداد کنترل می‌شود [۸،۱۲،۱۴،۱۹]. اگر این تعداد در هر زمانی به حد آستانه برسد، ارسال را متوقف می‌کند و منتظر اعمال کنترل دوره‌ی بعد نمی‌ماند. لذا نسبت به تاخیر واکنش و فاصله‌ی بین کنترل‌ها کمتر آسیب‌پذیر می‌باشد. بعلاوه، در برابر بارهای انفجاری[۴۱] مقاوم‌تر از روش مبتنی بر نرخ می‌باشد. با تعیین یک اندازه پنجره‌ی مناسب، سیستم دچار ناپایداری مطلق نخواهد شد و تنها باید جلوی نوسانات گذردهی گرفته شود. بطور کلی، محدوده‌ی پایداری بیشتری حتی از نقطه نظر نوسانات، نسبت به سایر روش‌ها دارد.

۶-۲-۴- روشن/خاموش

این روش به دلیل ایجاد الگوی فرستادن و نفرستادن همه‌ی ترافیک، باعث نوسان شدید در گذردهی و ایجاد مشکل در بازگشت از ناپایداری سیستم می‌شود [۲،۳،۱۹]. [۱۹] با تنظیم بهتر زمان روشن/خاموش،

کمی این روش را بهبود بخشیده است. به هرحال، این روش نسبت به روش‌های بالا از کارایی پایین‌تری برخوردار است.

۶-۳- معیار تشخیص یا پیش‌بینی اضافه‌بار

۶-۳-۱- تاخیر پاسخ‌دهی

همانطور که از ذات این معیار بر می‌آید، یک معیار انتها به انتهاست که روش‌های مبتنی بر بازخورد ضمنی از آن استفاده می‌کنند. تاخیر پاسخ‌دهی شامل زمان پردازش پیام‌ها و میزان انتظار آن‌ها در صف هر یک از عناصر میانی می‌باشد. در شرایط ایده‌آل، یعنی زمانی که سیستم اغلب بیکار است، اغلب پیام‌ها هیچ‌گونه زمان انتظاری در صف‌های مختلف تجربه نمی‌کنند و تنها، زمان پردازش آن‌ها به عنوان تاخیر پاسخ‌دهی در نظر گرفته می‌شود. اما، در صورتی که سیستم در حال تجربه‌ی بار سنگین و یا اضافه‌بار باشد، زمان پاسخ‌دهی برای اغلب درخواست بیشتر شامل انتظار آن‌ها در صف‌ها خواهد بود. در نتیجه، می‌توان از سطوح تغییرات در تاخیر پاسخ‌ها، برای تشخیص اضافه‌بار و میزان اضافه‌بار استفاده کرد [۳۴]. خوبی این معیار، لحاظ شدن تاخیر حاصل از تمامی عناصر میانی می‌باشد. برای مثال، حتی اگر سرورهای DNS، پایگاه‌های داده یا سرورهای کاربرد نیز گلوگاه باشند، تاثیرشان در تاخیر برقراری تماس پدیدار خواهد شد.

تاخیر پاسخ‌دهی، فاصله‌ی زمانی بین ارسال درخواست و دریافت پاسخ مورد نظر می‌باشد. مقاله‌های [۱۲،۳۳] از پاسخ تکمیل کننده‌ی تراکنش OK ۲۰۰ استفاده می‌کنند. ایرادی که می‌توان به این روش‌ها گرفت این است که، تاخیری که کاربر مقصد در برقراری تماس ایجاد می‌کند نیز وارد تاخیر کل می‌شود. مقاله‌ی [۸] این مشکل را با استفاده از پاسخ موقت Ringing ۱۸۰ تا حدی حل می‌کند. اما، این پاسخ نمی‌تواند نشانگر درستی از تکمیل تراکنش باشد و ممکن است که OK ۲۰۰ با تاخیر زیاد برسد و یا اصلا نرسد.

با استفاده از این معیار احتمال دارد افزایش تاخیرهای طبیعی به اشتباه به عنوان اضافه‌بار و افزایش‌های غیر طبیعی به عنوان شرایط عادی تلقی گردند که این باعث کاهش محدوده پایداری سیستم بدلیل لختی در تشخیص اضافه بار خواهد شد.

ایراد دیگر این است که بازخوردی که می‌گیریم با تاخیر می‌باشد، زیرا باید منتظر ماند تا تراکنش تکمیل شود و تاخیر آن را اندازه گرفت. این مساله نیز در تشخیص بلادرنگ اضافه‌بار مشکل ایجاد خواهد کرد. در [۲۶] روشی آمده است که این مشکل را با نظارت کردن تعداد تراکنش‌های فعال تا حدی حل کرده است.

۶-۳-۲- دریافت پاسخ کد ۵۰۳

این معیار نیز طبیعتی انتها به انتها دارد. لازمه‌ی استفاده از آن این است که تمامی سرورهای میانی دارای یک مکانیزم کنترل اضافه‌بار محلی باشند. مقاله‌های [۲۲،۲۸] با استفاده از الگوریتم‌هایی، بسته به تعداد پاسخ ۵۰۳ دریافتی و یا عدم دریافت آن‌ها برای مدتی، بار ارسالی را کاهش یا افزایش می‌دهند. ایراد بزرگ استفاده از این معیار، وابستگی روش کنترل اضافه‌بار به سرعت و دقت روش‌های کنترل اضافه‌بار محلی موجود در سرورهای درون مسیر می‌باشد. بعلاوه، مشکلاتی مانند فقدان پاسخ‌ها و تحمیل سربار به سرور تحت اضافه‌بار جهت انجام نظارت و کنترل را دارد. در اضافه‌بارهای سبک، الگوریتم کنترل می‌تواند سریع‌تر از روش قبلی اضافه‌بار را تشخیص دهد، زیرا در صورت وقوع اضافه‌بار پاسخ ۵۰۳ به اولین درخواست تراکنش داده می‌شود. اما در معیار مذکور قبلی، بازخورد بعد از اتمام تراکنش به دست می‌آید. اما به نظر می‌رسد که در اضافه‌بارهای سنگین، بدلیل متکی بودن به یک روش محلی، شکست خواهد خورد.

۶-۳-۳- منقضی شدن درخواست‌ها

ایراد این معیار زیاد بودن تاخیر واکنش آن است. تاخیر تشخیص اضافه‌بار وابسته به زمانبندهای تراکنش SIP جهت تشخیص فقدان پیام‌ها می‌باشد. این معیار به تنهایی دارای لختی زیادی می‌باشد و اصلا نمی‌تواند واکنش سریع به هنگام وقوع اضافه‌بار فراهم آورد. این معیار یک معیار کمکی برای سایر معیارها می‌تواند باشد، آنگونه که در [۲۲] به همراه معیار پاسخ ۵۰۳ استفاده شده است.

اینگونه به نظر می‌رسد که اگر عمل نظارت روی هر سه معیار مذکور انجام شود، بتوان یک روش قدرتمند ارایه نمود.

۶-۳-۴- میزان اشغال صف در برابر میزان اشغال پردازنده

هنگام استفاده از معیار اندازه یا تاخیر صف، نظارتگر به طور مرتب بر میزان اشغال صف در بافر نگاه می‌کند، این نگاه کردن می‌تواند به طور ساده یک اندازه برگرداند و یا اینکه بر اساس نوع پیام‌ها، تعداد آن‌ها درون صف را به دست آورد و با توجه به نرخ سرویس، تاخیر صف را محاسبه کند و به کنترل کننده بدهد.

برای به دست آوردن میزان اشغال پردازشگر، نظارتگر در فاصله‌های زمانی تعیین شده، میزان اشغال پردازنده را محاسبه می‌کند و به عنصر کنترلی می‌دهد تا آن را با یک مقدار که حداکثر بهره‌وری قابل قبول است مقایسه کند. مقاله‌ی [۱۴] نشان داده است که روش‌های مبتنی بر تاخیر می‌توانند عملکرد بهتری نسبت به روش‌های مبتنی بر میزان اشغال پردازشگر ارایه دهند. این ادعا همیشه صدق نمی‌کند و به شدت وابسته به الگوریتم کنترل هر روش می‌باشد. اما از آنجایی که در روش‌های مبتنی بر میزان اشغال پردازشگر، به پردازشگر اجازه داده نمی‌شود که به حداکثر بهره‌وری خود برسد، نمی‌توان به حداکثر گذردهی نیز دست یافت. لذا با فرض ایده‌آل بودن الگوریتم کنترل هر دو روش، روش مبتنی بر بافر نتیجه‌ی بهتری خواهد داد. یک راه حل برای بهبود کارایی روش مبتنی بر میزان اشغال پردازشگر در اضافه‌بارهای سنگین، استفاده از فاصله‌ی زمانی بسیار کوچک بین کنترل‌ها می‌باشد. زیرا در این حالت می‌توان مقدار حداکثر بهره‌وری را افزایش داد. اما دو مشکل اساسی دارد: اولا، به روزرسانی مقدار اشغال پردازشگر در فاصله‌های زمانی بسیار کوچک در هر سیستمی به

سادگی نمی‌باشد. ثانیا، نشان داده شده است که استفاده از فاصله زمانی‌های خیلی کوچک، در اضافه‌بارهای سبک باعث افت کارایی می‌شود [۱۴].

۶-۴- درجه‌ی همکاری

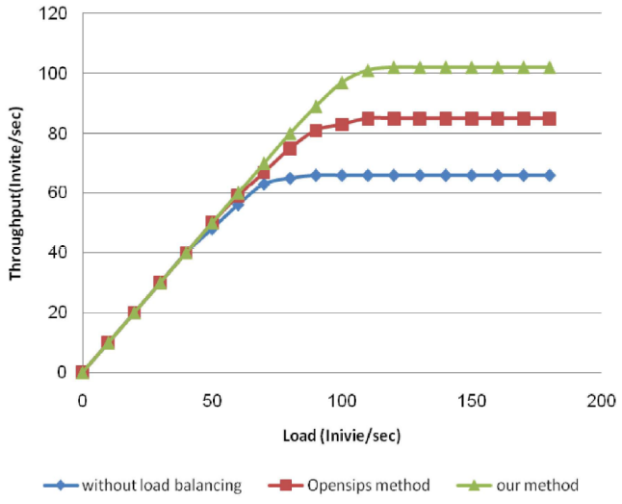
درجه‌ی همکاری یکی از عوامل مهم تاثیرگذار روی میزان باری که شبکه می‌تواند تحمل کند می‌باشد. در حالت ایده‌آل، درخواست‌هایی که قرار است در نهایت رد بشوند، باید در منبع ترافیک رد شوند. در غیر این صورت درخواست‌ها بعد از مصرف میزان زیادی از منابع شبکه، در نهایت رد خواهند شد. این قابلیت را روش‌های انتها به انتها فراهم می‌آورند. اگر بتوان یک روش انتها به انتهای مناسب پیاده‌سازی نمود، قطعا محدوده‌ی پایداری آن از هر روش محلی و گام به‌گام بیشتر خواهد شد. برای مثال، [۲۸] روش انتها به انتهای خود را با یک روش محلی و گام به گام در دو توپولوژی رایج و محک مقایسه کرده است و نشان داده است که روش انتها به انتها هم در گذردهی و هم در تاخیر نتایج بهتری ارایه می‌دهد. از این دیدگاه، با فرض ایده‌آل بودن تمامی الگوریتم‌های موجود، روش‌های انتها به انتها، گام به گام و محلی به ترتیب دارای کارایی و محدوده‌ی پایداری بالاتری خواهند بود.

۶-۵- فاصله‌ی بین کنترل‌ها و نظارت‌ها

اگر فاصله‌ی زمانی بین کنترل‌ها زیاد شود، تاخیر واکنش نیز زیاد خواهد شد. از طرف دیگر، اگر فاصله‌ی بین کنترل‌ها خیلی کوچک باشد، باعث سربار پردازشی زیاد و کاهش کارایی خواهد شد. لذا یک موازنه‌ی خوب باید بین تاخیر و سربار یافت. همچنین، این کار را باید برای عمل نظارت نیز انجام داد. مقاله‌ی [۱۴] تاثیر فاصله‌ی زمانی بین کنترل‌ها روی محدوده‌ی پایداری چند الگوریتم مختلف را مورد بررسی قرار داده و نشان داده است که در اضافه‌بارهای سنگین هرچقدر این مقدار کم‌تر باشد بهتر است. فاصله‌ی بین نظارت‌ها باید کمتر یا مساوی فاصله‌ی بین کنترل‌ها باشد و بین دو عمل کنترل بهتر است که چندین عمل نظارت انجام داد تا دقت اطلاعات بالاتر باشد.

۶-۶- استفاده یا عدم استفاده از تعادل‌بار[۴۲]

اینکه تمامی منابع شبکه به یک میزان مورد بهره‌وری قرار گیرند در میزان پایداری خیلی با اهمیت می‌باشد. در صورت استفاده از یک متعادل کننده‌ی بار خوب، تمامی اجزای درونی شبکه دارای یک میزان بهره‌وری خواهند بود، در نتیجه، گلوگاه شبکه، یک عنصر منفرد با کمترین ظرفیت باقیمانده نخواهد بود. بلکه کل عناصر میانی شبکه گویی دارای ظرفیت باقیمانده‌ی یکسانی هستند. بنابراین، حداقل بار قابل اعمال به سیستم برای ایجاد وضعیت اضافه بار و ناپایداری، افزایش می‌یابد. در نتیجه، تمامی منابع به صورت حداکثر استفاده خواهد شد و محدوده‌ی پایداری شبکه افزایش خواهد یافت. شکل ۷ این میزان افزایش در محدوده‌ی پایداری را به وضوح نشان می‌دهد. در واقع، استراتژی تعادل بار، ترافیک ورودی جدید را به هر سرور بر اساس ظرفیت در دسترس‌اش توزیع می‌کند. بنابراین، احتمال اینکه اضافه‌بار در سرور مشخصی اتفاق بیفتد را کاهش می‌دهد [۳۷].



شکل (۷): تاثیر استفاده از تعادل بار بر روی محدوده‌ی پایداری [۳۵]

۷- بررسی محدوده‌ی پایداری روش‌های مختلف

۷-۱- روش‌های محلی

همانطور که قبلا گفته شد، روش‌های محلی ضعیف‌ترین روش مقابله با اضافه‌بار می‌باشند، اما روش‌های مختلف کنترل محلی دارای محدوده بار قابل کنترل مختلفی هستند.

در [۶] روشی مبتنی بر اندازه بافر مطرح شده است که سرور تنها دو حالت دارد، حالت کم‌بار، که تمامی درخواست‌ها پذیرفته می‌شوند و حالت اضافه‌بار که هیچ درخواستی پذیرفته نمی‌شود. روش [۲۱] کمی این روش را با اضافه کردن یک حالت مابین و رد کردن احتمالی درخواست‌ها با توجه به طول صف، بهبود بخشیده است. نشان داده شده است که روش [۶] وقتی بار ورودی به حدی می‌رسد دچار ناپایداری می‌شود. ناپایداری آن به دلیل رد کردن تمامی درخواست‌ها هنگام اضافه‌بار و پذیرفتن تمام درخواست‌ها هنگام خروج از اضافه‌بار می‌باشد. این روش، به دلیل رد کردن احتمالی پایداری را با کم کردن نوسانات گذردهی مقداری بهبود بخشیده است. اما به طور کلی، این دو روش از نقطه نظر دچار گذردهی صفر شدن و عدم خارج شدن از آن به مدت طولانی دارای محدوده بار تقریبا یکسانی می‌باشند.

[۱۱] یک روش کاهنده ارایه می‌کند که دارای دو صف اولویت FIFO می‌باشد، که اولویت بالا به پیام‌های غیر INVITE اختصاص داده می‌شود و آن را با یک روش تک صفی مقایسه کرده است و نشان داده است که روش مطرح شده پایداری بهتری ارایه می‌دهد.

[۱۰] یک روش مبتنی بر تشخیص ارایه می‌دهد که نتایج [۱۱] را تایید می‌کند. هنگام اضافه‌بار، به پیام‌ها به صورت اولویت‌دار سرویس‌دهی می‌شود. اولویت بالا به غیر INVITEها اختصاص می‌یابد. وقتیکه اضافه‌بار رفع شد، دوباره به حالت تک اولویتی باز می‌گردد. نشان داده شده است که بدون استفاده از ۵۰۳ نتایج بهتری را می‌دهد.

[۹] ترکیب دو روش [۶،۲۱] را مورد استفاده قرار داده است، اما از یک روش کاهنده نیز استفاده کرده است. بدین صورت که پیام‌ها به کلاس‌های مختلفی تقسیم می‌شوند که هر کلاس اولویت جدایی برای پردازش دارد. بالاترین اولویت به پیام‌های non-INVITE داده شده است و اولویت‌های بعدی به ترتیب برای INVITEهای بار اول تا بار هفتم ارسال شده می‌باشد. این روش به دلیل دادن اولویت به پیام‌های جلسه‌ی پذیرفته شده و تکمیل آن‌ها در ابتدا، گذردهی و محدوده‌ی پایداری را نسبت به دو روش قبلی کمی بهبود بخشیده است.

[۳] یک الگوریتم مبتنی بر میزان اشغال پردازشگر را با الگوریتم مبتنی بر بافر [۶] مقایسه کرده است و نشان داده است که روش مبتنی بر میزان اشغال پردازشگر دارای محدوده‌ی پایداری بیشتری می‌باشد. زیرا در مبتنی بر بافر، هنگام اضافه‌بار، سرور شروع به رد کردن تمامی درخواست‌ها می‌کند که نتیجه‌ی آن یک افت ناگهانی در گذردهی می‌باشد. بعلاوه، اجازه داده می‌شود که پردازشگر به بهره‌وری بالایی برسد. در کنترل اضافه‌بار محلی، پردازشگر نیاز به ظرفیت کافی برای رد کردن درخواست‌ها دارد، در این روش‌ها ممکن است زمانیکه اندازه بافر به حد بالا رسیده است، این ظرفیت در دسترس نباشد [۳]. با این حال، نمی‌توان گفت که روش‌های مبتنی میزان اشغال پردازنده بهتر از روش‌های مبتنی بر اندازه‌ی صف می‌باشند، زیرا کارایی هر روش به شدت وابسته به الگوریتم کنترل می‌باشد. برای مثال، مقاله‌ی [۳۲] دو الگوریتم مبتنی بر صف و میزان اشغال پردازشگر را مقایسه کرده و نشان داده است که روش مبتنی بر میزان اشغال پردازشگر به دلیل تخمین کم هزینه‌ی پردازش تماس کارایی پایین‌تری دارد.

می‌توان گفت که تمامی روش‌های کنترل محلی بعد از اینکه بار از نرخ رد کردن سرور عبور می‌کند گذردهی‌شان نزدیک به صفر می‌شود. تفاوت آن‌ها، اغلب در میزان گذردهی‌شان قبل از این محدوده می‌باشد. بعد از اینکه بار از نرخ رد کردن سرور عبور می‌کند، دیگر نمی‌توانند به راحتی به شرایط عادی برسند. در نهایت می‌توان گفت که روش‌های تقلیل مبتنی بر اولویت‌دهی به پیام‌ها دارای محدوده‌ی پایداری بیشتری نسبت به سایر روش‌های محلی می‌باشد.

۷-۲- روش‌های گام به گام

[۱۴] پنج الگوریتم مختلف را مقایسه می‌کند؛ برخی مبتنی بر تاخیر صف و برخی مبتنی بر میزان اشغال پردازشگر. نتایج نشان می‌دهد که روش‌های مبتنی بر صف نتایج بهتری را ارایه می‌دهند.

در [۱۹] چهار الگوریتم مختلف آزمایش شده است، الگوریتم اول یک نسخه‌ی بهبود یافته از مکانیزم Retry-After موجود در SIP می‌باشد، دومی و سومی دو الگوریتم مبتنی بر نرخ می‌باشند که به ترتیب مبتنی بر میزان اشغال پردازشگر و تاخیر صف هستند. آزمایش نشان می‌دهد که سومی بهتر از دومی عمل می‌کند. الگوریتم چهارم یک روش مبتنی بر پنجره‌ی فرستنده-محور می‌باشد که دریافت ۱۰۰ Trying را برای افزایش و عدم دریافت پاسخ و گرفتن کد ۵۰۳ را برای کاهش پنجره استفاده می‌کند. آزمایش‌ها نشان می‌دهند که روش مبتنی بر پنجره بهتر از سه الگوریتم دیگر می‌باشد.

[۲۳] روشی را معرفی می‌کند که از چهار معیار زمان سرویس، زمان پاسخ‌دهی پایگاه داده، تاخیر صف و تاخیر دریافت پاسخ ۱۰۰ Trying استفاده می‌کند و جهت تشخیص اضافه‌بار، آن‌ها را به یک طبقه‌بند[۴۳] تحت عنوان SVM[۴۴] می‌دهد. شبیه‌سازی نشان داده است که این روش عملکردی بهتر از روش ARC[۴۵] موجود در [۱۴] عمل می‌کند. استدلال نویسنده برای کارایی بهتر، استفاده از چند معیار به جای یک معیار برای تشخیص اضافه‌بار می‌باشد.

[۲۷] یک روش ترکیبی ارایه می‌کند که روش CFP [۱۰] را به عنوان الگوریتم محلی و از روش مبتنی بر نرخ مبتنی بر صف [۱۴] برای الگوریتم توزیعی استفاده می‌کند. اما یک روش متفاوت از آنچه که در [۱۴] آمده است استفاده می‌کند که تخمین خیلی دقیق‌تری نسبت به آن ارایه می‌دهد. در نتیجه واکنش به اضافه‌بار خیلی دقیق‌تر و سریع‌تر انجام می‌شود. در نتیجه شبیه‌سازی‌ها نشان داده است که این روش محدوده‌ی پایداری را افزایش می‌دهد. همچنین این روش از یک پیش‌بین NLMS[۴۶] برای پیش‌بینی مقادیر بازخورد استفاده می‌کند که کارایی روش و محدوده پایداری را افزایش می‌دهد.

[۳۶] با کنترل نرخ ارسال مجددها توسط بالادستی‌ها اثرات اضافه‌بار را در اضافه‌بارهای سبک و موقت کاهش می‌دهد. ایده‌ی آن این است که به هنگام اضافه‌بار موقت، ممکن است نیازی به رد کردن تماس‌ها نباشد و می‌توان بدون رد کردن تماس‌ها، نرخ بلوکه شدن را کاهش داد و کارایی را بهبود بخشید. کار اصلی آن تشخیص دلیل ارسال مجددها می‌باشد. در صورتی که دلیل ارسال مجددها اضافه‌بار در پایین‌دستی باشد، نرخ ارسال مجددها را کنترل می‌کند.

۷-۳- روش‌های انتها به انتها

همانطور که قبلا اشاره شد، روش‌های انتها به انتها به دلیل پیچیدگی بالای روش‌های مبتنی بر مسیر، اغلب به صورت مبتنی بر مقصد پیاده‌سازی می‌شوند. روش‌های مبتنی بر مقصد اغلب از تاخیر برقراری تماس، منقضی شدن درخواست‌ها، دریافت پاسخ کد ۵۰۳ و یا ترکیبی از این سه به عنوان معیاری برای تشخیص اضافه‌بار استفاده می‌کنند.

[۱۲] یک روش مبتنی بر پنجره‌ی ضمنی را ارایه می‌کند که از تاخیر برقراری تماس به عنوان معیاری برای تشخیص اضافه‌بار استفاده می‌کند. نشان داده شده است که این روش در مقایسه با یک روش انتها به انتهای OCC [۳] بهتر عمل می‌کند، زیرا به دلیل طبیعت نوع بازخورد مورد دوم، الگوریتم دایما در حال بالا و پایین بردن بهره‌وری پردازشگر و در نتیجه‌ی آن بالا و پایین بردن گذردهی آن می‌باشد.

[۲۲] از نرخ دریافت ۵۰۳ها و نرخ منقضی شدن درخواست‌ها استفاده می‌کند، یعنی ترکیبی از روش صریح و ضمنی. این روش از سه بخش کاهش ضربتی، افزایش ضربتی و افزایش خطی تشکیل شده است. تفاوت اساسی آن با روش AIMD [۲۴] در نحوه‌ی افزایش و کاهش می‌باشد، در این روش میزان کاهش بر اساس انحراف معیار محاسبات انجام شده با یک مقدار از قبل تعیین شده می‌باشد. به

همین دلیل نسبت به افزایش و کاهش‌های زیاد و کم به طور مناسب عکس‌العمل نشان می‌دهد. نشان داده شده است که گذردهی یکسانی را با روش‌های کنترل اضافه‌بار مبتنی بر گیرنده‌ی [۲۵] ارایه می‌دهد.

[۲۶] با استفاده از تاخیر برقراری تماس و تعداد تراکنش‌های فعال، هدفش این است که تعداد تراکنش‌های فعال بیش از یک حدی نباشد. ادعا شده است که این روش دارای تاخیر کم‌تری برای تشخیص اضافه‌بار می‌باشد. زیرا در هر دوره‌ی کنترل، از روی نرخ بار ورودی به سمت سرور مقصد مورد نظر و تاخیر پاسخ‌های قبلی، تعداد تراکنش‌هایی را که فعال خواهند شد را به دست می‌آورد و آن را با تعداد تراکنش‌های فعال فعلی مقایسه می‌کند. اگر بیشتر بود درخواست را ارسال و در غیر این صورت درخواست را رد می‌کند.

[۲۸] یک روش مبتنی بر فرستنده با پاسخ صریح با استفاده از پیام‌های ۵۰۳ دریافتی می‌باشد. مقاله نتایج را با یک روش محلی و یک روش گام به گام که هر دو مبتنی بر میزان اشغال پردازشگر هستند مقایسه کرده است و نشان داده است که در دو توپولوژی مختلف، الگوریتم پیشنهادی نتایج بهتری را ارایه می‌کند.

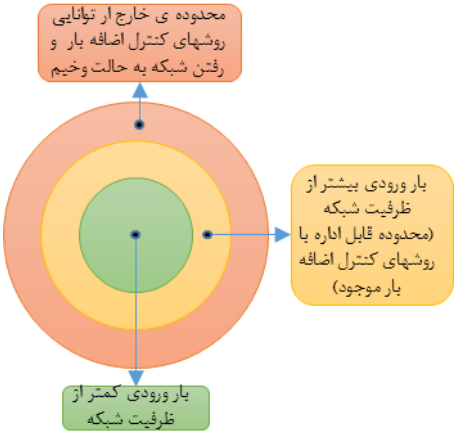
[۳۱] ترکیب روش تعادل بار و روش انتها به انتهای مبتنی بر مقصد را ارایه کرده است. کنترل فقط در سرورهای لبه‌ی ورودی انجام می‌شود. پیام‌های هر مقصد وارد صف جداگانه‌ای می‌شود و با توجه به اندازه‌ی صف مربوط به یک مقصد، برای آن مقصد یک مقدار که نشان دهنده‌ی زمان بین ارسال هر درخواست است تعیین می‌شود. نتایج نشان می‌دهد که کارایی بهتری از یک روش محلی و گام به گام دارد.

[۳۳] یک روش مبتنی بر بازخورد ضمنی که از معیار تاخیر تکمیل تراکنش‌ها استفاده می‌کند ارایه کرده است. در هر ثانیه تاخیرهای محاسبه شده در ۵ ثانیه‌ی قبل را بررسی میکند. اگر ۹۵ درصد از تراکنش‌ها تاخیر کمتر از ۰٫۵ ثانیه داشته باشند، شرایط عادی تلقی شده و به تعداد حداکثر تراکنش‌های فعال می‌افزاید. در غیر این صورت این تعداد را کاهش می‌دهد. ادعا شده است که مقادیر ذکر شده برای پارامترها، باعث هرچه بیشتر شدن پایداری و عدم نوسان می‌شود.

۸- طرح پیشنهادی

هدف از این سمینار و بررسی مساله‌ی پایداری، ارایه‌ی روشی برای پیشگیری از وقوع یک اضافه‌بار سنگین و رفتن شبکه به یک وضعیت وخیم می‌باشد. روش‌های مذکور در بخش ۴ به فراوانی مورد مطالعه و بررسی قرار گرفته‌اند. تفاوت اغلب آن‌ها در ایجاد یک بهبود نسبی نسبت به روش‌های قبلی می‌باشد و هر کدام مشکلات ناپایداری مربوط به خود را آنچنان که مورد بررسی قرار گرفت دارا می‌باشند. هدف ما بهبود محدوده‌ی پایداری اما با استفاده از روش‌های موجود می‌باشد.

با توجه به شکل مفهومی ۸ که نشانگر وضعیت شبکه تحت بارهای مختلف است و بررسی‌های ارایه شده در بخش‌های قبل، اگر میزان بار ورودی به شبکه از میزان بار قابل کنترل (ناحیه‌ی زرد رنگ) توسط روش‌های کنترل مختلف بیشتر شود شبکه وارد ناحیه‌ی بحرانی خواهد شد. ما متوجه شدیم که اگر بتوان توسط روشی، با توجه به تاخیر واکنش روش کنترل محلی و ظرفیت اسمی هر سرور، یک حداکثر بار برای سرور تعریف نمود، می‌توان از این پدیده پیشگیری کرد. روش پیشنهادی، بالای روش‌های موجود کار می‌کند و وظیفه‌ی اصلی‌اش تعیین این مقدار حداکثر در یک مقیاس زمانی بزرگتر نسبت به روش‌های موجود می‌باشد. هر سرور مقداری را محاسبه می‌کند و آن را درون فیلترهایی قرار می‌دهد و برای فیلتر کردن بار به بالادستی(های) خود ارسال می‌کند. فیلتر کردن بار بدین معناست که بالادستی‌ها تحت هر شرایطی اجازه‌ی ارسال مقداری بیش از مقدار درون فیلتر ندارند. این مقدار به گونه‌ای انتخاب می‌شود که سرور را می‌تواند دچار اضافه‌بار کند. با وقوع اضافه‌بار، هر الگوریتم کنترل اضافه‌باری که در سطح زیرین قرار دارد وظیفه‌ی اداره‌ی آن را دارد. اما این مقدار به اندازه‌ای است که شبکه از محدوده‌ی دایره دوم در شکل ۸ خارج نشود.

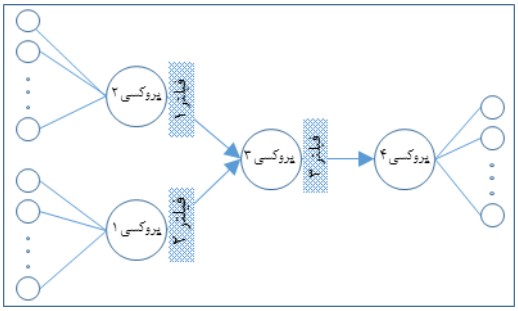


شکل (۸): نمودار مفهومی از ناحیه‌های مختلف کاری شبکه

اینکه این مقدار بیشتر از ظرفیت اسمی سرور می‌باشد بدین دلیل است که هدف پیشگیری از اضافه‌بار نمی‌باشد، بلکه پیشگیری از یک اضافه‌بار سنگین و غیرقابل کنترل می‌باشد. زیرا با فرض مشخص بودن ظرفیت لحظه‌ای سرور، تقسیم آن بین بالادستی‌ها به طور مساوی یا هر طور دیگر که جمعشان برابر با ظرفیت سرور شود وظیفه‌ی همان روش‌های کنترل اضافه‌بار زیرین می‌باشد که به طور دایم این مقدار را به نحوی برای بالادستی‌ها به روز می‌کنند. ما قصد ارایه‌ی چنین روشی را نداریم. در روش پیشنهادی، برای مثال، مقداری که پروکسی ۳ در شکل ۹ به دست می‌آورد بیانگر مقداری است که پروکسی می‌تواند در بدترین حالت همه را رد کند تا بتواند از مساله سطوح مختلف ارسال مجدد که در [۲۰] آمده است و منجر به سقوط کارایی می‌شود پیشگیری کند و مانع از خروج از ناحیه‌ی دوم شکل ۸ شود. مهم‌ترین مساله در اینجا، محاسبه‌ی مناسب این مقدار در هر دوره‌ی کنترل، فاصله‌ی زمانی تنظیم مجدد مقادیر و توزیع آن‌ها می‌باشد. ایده‌ای که پشت این طرح می‌باشد این است که زمان لازم برای این محاسبه در مقیاس بزرگتری می‌باشد و می‌توان این مقدار را با دقت مناسبی محاسبه نمود و با سیاست‌های مختلفی بین بالادستی‌ها تقسیم کرد.

برای نصب، ارسال و به روزرسانی این فیلترها می‌توان از مکانیزم یا بستر ارایه شده در [۱۸] استفاده نمود که مکانیزم اشتراک-اعلان[۴۷] و

قالب XML را برای فیلترها پیشنهاد می‌کند. فیلترها می‌توانند حاوی اطلاعاتی به جز مقدار بار حداکثر باشند. مانند زمان اعمال فیلتر و دامنه‌ی تحت تاثیر توسط این فیلتر. یعنی می‌توان مشخص کرد که این فیلتر روی چه دامنه‌های مبدا یا مقصدی جهت برقراری عدالت و افزایش کیفیت سرویس اعمال شود.



شکل (۹): نحوه‌ی نصب فیلترها در روش پیشنهادی

۹- نتیجه

در این سمینار مسایل مربوط به محدوده‌ی پایداری شبکه‌های SIP را مورد بررسی قرار دادیم و متوجه شدیم که مهم‌ترین عامل تاثیر گذار روی محدوده‌ی پایداری سیستم، تاخیر واکنش به اضافه‌بار می‌باشد. از آنجاییکه وقوع اضافه‌بار در این شبکه‌ها پدیده‌ای اجتناب‌ناپذیر می‌باشد باید سیستم را مجهز به یک روش کنترل اضافه‌بار موثر نمود تا هرچه بیشتر کارایی سیستم را بالا نگه دارد. روش‌های متنوعی برای کنترل اضافه‌بار در شبکه‌های SIP وجود دارد. اغلب این روش‌ها تا حد زیادی کارایی شبکه را نسبت به حالت بدون کنترل در شرایط اضافه‌بار بهبود می‌بخشند. اما با بررسی‌های انجام شده در این سمینار، به طور کیفی نشان دادیم که هیچ یک از این روش‌ها قادر به جلوگیری کامل ناپایداری سیستم نمی‌باشند. تمامی روش‌های مذکور بعد از اینکه بار ورودی به سیستم بیش از حدی می‌شود، به دلیل تاخیر در واکنش به اضافه‌بار موجب ناپایداری شبکه خواهند شد. مواردی مانند نوع بازخورد، الگوریتم کنترل اضافه‌بار، معیار تشخیص اضافه‌بار، میزان همکاری عناصر شبکه بایکدیگر و فاصله‌ی بین کنترل‌ها، روی تاخیر واکنش به اضافه‌بار و در نتیجه روی محدوده‌ی پایداری تاثیرگذار می‌باشند. اینکه با چه دقتی و با چه سرعتی اضافه‌بار و یا خروج از اضافه‌بار را تشخیص داد و نسبت به هر کدام یک عمل مناسب انجام داد، مسایلی هستند که می‌توانند باعث هرچه بیشتر شدن محدوده‌ی پایداری بشنود.

همچنین متوجه شدیم که استفاده از روش‌هایی مانند تعادل بار و روش‌های کاهنده می‌توانند باعث افزایش محدوده‌ی پایداری بشوند.

مراجع

[۱] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, "*SIP: Session Initiation Protocol*", IETF RFC ۳۲۶۱, June ۲۰۰۲.

[۲] J. Rosenberg, "*Requirements for Management of Overload in the Session Initiation Protocol*", IETF RFC ۵۳۹۰, December ۲۰۰۸.

[۳] V. Hilt, I. Widjaja, "*Controlling overload in networks of SIP servers*", in: Proceedings of IEEE ICNP, October ۲۰۰۸, pp. ۸۳–۹۳.

[۴] V. Hilt, E. Noel, C. Shen, A. Abdelal, "*Design considerations for session initiation protocol (SIP) overload control*", IETF RFC ۶۳۵۷, August ۲۰۱۱.

[۵] SCHULZRINNE, H., NARAYANAN, S., LENNOX, J., AND DOYLE, M., "*SIPstone – Benchmarking SIP server performance*", Tech. rep., Columbia University, Apr. ۲۰۰۲.

[۶] M. Ohta: "*Overload Control in a SIP Signaling Network*", Proceedings of World Academy of Science, Engineering and Technology, vol.۱۲, Mar. (۲۰۰۶).

[۷] C. Shen and H. Schulzrinne, "*On TCP-based SIP server overload control*", in IPTComm, ۲۰۱۰, pp. ۷۱-۸۳.

[۸] Azhari, Seyed Vahid, Maryam Homayouni, Hani Nemati, Javad Enayatizadeh, and Ahmad Akbari. "*Overload control in SIP networks using no explicit feedback: A window based approach*." Computer Communications (۲۰۱۲).

[۹] Inwhee Joe, and Janghyun Lee. "*An Overload Control Algorithm based on Priority Scheduling for SIP Proxy Server*." International Conference on Internet Computing, ۲۰۱۲

[۱۰] R. G. Garroppo, S. Giordano, S. Niccolini, and S. Spagna, "*Queueing strategies for local overload control in SIP server*", in Proc. Globecom, Nov. ۲۰۰۹.

[۱۱] M. Ohta, "*Overload protection in a SIP signaling network*", Internet Surveillance Protection, International Conference on, p. ۱۱, Aug. ۲۰۰۶.

[۱۲] Homayouni, Maryam, Mojtaba Jahanbakhsh, Vahid Azhari, and Ahmad Akbari. "*Overload control in sip servers: Evaluation and improvement*." In Telecommunications (ICT), ۲۰۱۰ IEEE ۱۷th International Conference on, pp. ۶۶۶-۶۷۲. IEEE, ۲۰۱۰.

[۱۳] V. Hilt, E. Noel, C. Shen, A. Abdelal, "*Design considerations for session initiation protocol (SIP) overload control*", IETF RFC ۶۳۵۷, August ۲۰۱۱.

[۱۴] C. Shen, H. Schulzrinne, and E. Nahum, "*Session Initiation Protocol (SIP) Server Overload Control: Design and Evaluation*", Principles, Systems and Applications of IP Telecommunica-tions (IPTComm), July ۲۰۰۸.

[۱۵] V. Gurbani, Ed., V. Hilt, H. Schulzrinne. "*Session Initiation Protocol (SIP) Overload Control draft-gurbani-soc-overload-control-۰۳*", August ۲۳, ۲۰۱۰

[۱۶] V. Hilt, H. Schulzrinne. "*Session Initiation Protocol (SIP) Overload Control draft-ietf-sipping-overload-۰۸*", April ۲۶, ۲۰۱۰

[۱۷] V. Gurbani, Ed., V. Hilt, H. Schulzrinne. "*Session Initiation Protocol (SIP) Overload Control draft-ietf-soc-overload-control-۰۹*", July ۰۶, ۲۰۱۲

[۱۸] Shen, C., Schulzrinne, H., and A. Koike, "*A Session Initiation Protocol (SIP) Load Control Event Package*",

draft-ietf-soc-load-control-event-package-۰۴' (work in progress), July ۲۰۱۲.

[۱۹] E. Noel and C. Johnson, "*Novel Overload Controls for SIP Networks*", ۲۱st International Teletraffic Congress, ۲۰۰۹, pp. ۱ – ۸.

[۲۰] هانی نعمتی، "*ارایه‌ی روش کنترل بار مبتنی بر نرخ و پنجره در سرورهای SIP*"، پایان نامه‌ی کارشناسی ارشد، دانشگاه علم و صنعت ایران-دانشکده کامپیوتر، اردیبهشت ۱۳۹۰

[۲۱] J. Yang, F. Huang, and S. Gou: "*An Optimized Algorithm for Overload Control of SIP signaling Network*", ۵th International Conference on Wireless Communications, Networking and Mobile Computing (WiCom), pp.۱-۴, ۲۰۰۹.

[۲۲] Abdelal, A.; Matragi, W., "*Signal-Based Overload Control for SIP Servers*", Consumer Communications and Networking Conference (CCNC), ۲۰۱۰ ۷th IEEE , vol., no., pp.۱-۷, ۹-۱۲ Jan. ۲۰۱۰

[۲۳] Chentouf, Zohair. "*SIP overload control using automatic classification*." In Electronics, Communications and Photonics Conference (SIECPC), ۲۰۱۱ Saudi International, pp. ۱-۶. IEEE, ۲۰۱۱.

[۲۴] D-M. Chiu and R. Jain, "*Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks*", Computer Networks and ISDN Systems, vol. ۱۷, pp. ۱–۱۴, ۱۹۸۹.

[۲۵] E. Noel, C. Johnson, "*Initial Simulation Results That Analyze SIP Based VoIP Networks Under Overload*", International Teletraffic Congress (ITC'۰۷), Ottawa, Canada, June ۲۰۰۷.

[۲۶] Happenhofer, Marco, and Christoph Egger. "*Implicit SIP proxy overload detection mechanism based on response behavior*." Software, Telecommunications and Computer Networks (SoftCOM), ۲۰۱۱ ۱۹th International Conference on. IEEE, ۲۰۱۱.

[۲۷] Garroppo, Rosario G., Stefano Giordano, Saverio Niccolini, and Stella Spagna. "*A prediction-based overload control algorithm for SIP servers*." Network and Service Management, IEEE Transactions on ۸, no. ۱ (۲۰۱۱): ۳۹-۵۱.

[۲۸] Liao, Jianxin, Jinzhu Wang, Tonghong Li, Jing Wang, Jingyu Wang, and Xiaomin Zhu. "*A distributed end-to-end overload control mechanism for networks of SIP servers*." Computer Networks (۲۰۱۲).

[۲۹] Homayouni, Maryam, Sayed Vahid Azhari, Mojataba Jahanbakhsh, Ahmad Akbari, Alireza Mansoori, and Nahid Amani. "*Configuration of a sip signaling network: An experimental analysis*." In INC, IMS and IDC, ۲۰۰۹. NCM'۰۹. Fifth International Joint Conference on, pp. ۷۶-۸۱. IEEE, ۲۰۰۹.

[۳۰] مجتبی جهانبخش، سید وحید ازهری، احمد اکبری، "*بهبود کارایی پروکسی پروتکل SIP با تعیین مناسب حافظه آن*"، دانشگاه علم و صنعت-دانشکده کامپیوتر

[۳۱] Wang, Yaogong. "*SIP overload control: a backpressure-based approach*." ACM SIGCOMM Computer Communication Review ۴۰, no. ۴ (۲۰۱۰): ۳۹۹-۴۰۰.

[۳۲] Montagna, S., & Pignolo, M. (۲۰۰۸, April). "*Performance evaluation of load control techniques in sip signaling servers*." In Systems, ۲۰۰۸. ICONS ۰۸. Third International Conference on (pp. ۵۱-۵۶). IEEE.

[۳۳] Egger, Christoph, Marco Happenhofer, and Peter Reichl. "*SIP proxy high-load detection by continuous analysis of response delay values*." In Software, Telecommunications and Computer Networks (SoftCOM), ۲۰۱۱ ۱۹th International Conference on, pp. ۱-۵. IEEE, ۲۰۱۱.

[۳۴] Egger, Christoph, Marco Happenhofer, and Michael Hirschbichler. "*A study of SIP proxy load patterns*." In Measurements and Networking Proceedings (M&N), ۲۰۱۱ IEEE International Workshop on, pp. ۱۴۰-۱۴۵. IEEE, ۲۰۱۱.

[۳۵] Karimi, Alireza, MehdiAgha Sarram, and Mohammad Ghasemzadeh. "*Two Stage Architecture for Load Balancing and Failover in SIP Networks*." Middle-East Journal of Scientific Research ۶, no. ۱ (۲۰۱۰): ۸۸-۹۲.

[۳۶] Hong, Y., Huang, C., & Yan, J. (۲۰۱۱b). "*Controlling Retransmission Rate for Mitigating SIP Overload*." In Proceedings of IEEE ICC, Kyoto, Japan.

[۳۷] Hong, Yang, Changcheng Huang, and James Yan. "*A Comparative Study of SIP Overload Control Algorithms*." Internet and Distributed Computing Advancements: Theoretical Frameworks and Practical Applications (۲۰۱۲).

[۳۸] Scharf, M., & Kiesel, S. "*Head-of-line blocking in TCP and SCTP: analysis and measurements*." In Global Telecommunications Conference, ۲۰۰۶. GLOBECOM'۰۶. IEEE (pp. ۱-۵). IEEE.

[۳۹] G Camarillo, R Kantola, H Schulzrinne. "*Evaluation of transport protocols for the session initiation protocol*." Network, IEEE ۱۷, no. ۵ (۲۰۰۳): ۴۰-۴۶.

[۴۰] Gurbani, Vijay K., and Rajnish Jain. "*Transport protocol considerations for session initiation protocol networks*." Bell Labs Technical Journal ۹, no. ۱ (۲۰۰۴): ۸۳-۹۷.

[۴۱] Masataka Ohta, "*Performance Comparisons of Transport Protocols for Session Initiation Protocol Signaling*", ۲۰۰۸

[۴۲] Dorgham Sisalem," *SIP Overload Control: Where are We Today?*" , Trust Worthy Internet, Springer Milan, ۲۰۱۱

زیرنویس‌ها

۱ Overload
۲ Session Initiation Protocol
۳ Stability
۴ Internet Engineering Task Force
۵ International Telecommunication Union
۶ ۳rd Generation Partnership Project
۷ European Telecommunications Standards Institute
۸ Voice over Internet Protocol
۹ Video over Demand
۱۰ Instant Messaging
۱۱ Multi hop
۱۲ Benchmark

[١٣] Goodput
[١٤] Peak
[١٥] Upstream
[١٦] header
[١٧] Local
[١٨] Hop-to-Hop
[١٩] End-to-End
[٢٠] Downstream
[٢١] User Agent Client
[٢٢] User Agent Server
[٢٣] Registration Server
[٢٤] Stream Control Transmission Protocol
[٢٥] Stateful
[٢٦] ٣-Way Handshaking
[٢٧] Real-time Transfer Protocol
[٢٨] Packet Loss
[٢٩] Implicit
[٣٠] Loss-based
[٣١] Rate-based
[٣٢] Window-based
[٣٣] ON/OFF
[٣٤] Time-Driven
[٣٥] Event-Driven
[٣٦] Detection
[٣٧] Prediction
[٣٨] Mitigation
[٣٩] Prevention
[٤٠] Self Limiting
[٤١] Burst
[٤٢] Load Balancing
[٤٣] Classifier
[٤٤] Support Vector Machine
[٤٥] Addaptive Rtae Control
[٤٦] Normalized Last Mean Square
[٤٧] Subscribe-Notify

مطالعه و بررسی روشهای مدلسازی سرور پراکسی SIP

محمد نعمتی[۱] ، احمد اکبری[۲]

[۱]دانشجوی مقطع کارشناسی ارشد (گرایش شبکه های کامپیوتری)

Nemati.moh@gmail.com

[۲] استاد راهنما

akbari@iust.ac.ir

چکیده

پروتکل SIP پروتکل لایه کاربرد است که به عنوان پروتکل سیگنالینگ در سیستم های تلفنی مبتنی براینترنت [۱] استفاده می شود و وظیفه اصلی آن برقراری و مدیریت تماس است. مهمترین عنصر در یک شبکه SIP، پراکسی سرور است . پراکسی سرور یک نرم افزار کاربردی است که بسته های ورودی SIP را در یک شبکه VoIPمسیریابی و هدایت می کند. ارائه مدل قوی و تحلیل دقیق معیارهای هدف در شبکه می تواند به در اختیار داشتن یک شبکه با قابلیت اطمینان و مقیاس پذیری بالاتر منجر شود. هدف در این سمینار مطالعه و بررسی مدل های ارزیابی کارایی سرور پراکسی به منظور شناخت بهتر رفتار یک پراکسی سرور تحت شرایط مختلف ترافیکی شبکه VoIP است.

کلمات کلیدی

IP تلفنی ، VoIP، پروتکل SIP، پراکسی سرور، مدل های ارزیابی کارایی، تئوری صف

[۱] Internet telephony

۱-مقدمه

پروتکل SIP به عنوان پروتکل منتخب برای سیگنالینگ در شبکه های VOIP استفاده می شود و در واقع وظیفه برقراری، نگهداری، انتقال و مدیریت جلسه در یک نشست را بر عهده دارد. مقایسه شبکه های VOIP با شبکه های تلفن انتقال صوت (PSTN) در شناخت بهتر و مناسبتر این پروتکل و در نهایت بررسی دقیقتر مدلهای ارائه شده به منظور ارزیابی کارایی پراکسی سرور ها کمک می کند. در این سمینار هدف ما بررسی و مطالعه مدل های مورد استفاده در ارزیابی کارایی سرور های پراکسی به عنوان قلب شبکه های VoIP می باشد. در بخش ۲ به معرفی عناصر موجود در شبکه VoIP و مقایسه نحوه برقراری تماس در SIP و PSTN می پردازیم. در بخش ۳مدل های ارائه شده برای ارزیابی کارایی و همچنین شرایط مختلف و ملاحضاتی که در برخی مدل ها لحاظ شده است به همراه شبیه سازی های انجام شده برای اعتبار سنجی مدل های استفاده شده را مورد بحث قرار می دهیم. در بخش ۴ نیز نتیجه گیری کلی از مدلهای ارائه شده و نحوه اثر بخشی آنها در پیاده سازی های بسترهای SIP ارائه می شود.

۲- فرآیند برقراری تماس و معرفی عناصر اصلی

برقراری تماس[۲] درشبکه های سنتی PSTN توسط پروتکل SS۷ صورت می گیرد. ابتدا به معرفی عناصر اصلی در دو حالت PSTN و SIP می پردازیم.

۲-۱ عناصر اصلی در PSTN و SIP

عنصر اصلی در شبکه های PSTN ، اداره مرکزی [۳] است. خطوط مربوط به انتقال داده های اصلی صوت و کنترلی در PSTN جدا می باشند. یعنی اطلاعات سیگنالینگ و کنترلی در PSTN در یک کانال مجزا ارسال می شوند و نوع سوئیچینگ از نوع مداری است. اما در یک شبکه VoIP مبتنی بر پروتکل SIP ، UAC [۴]دستگاه شروع کننده نشست و UAS [۵]مقصد یا همان مشترک مورد نظر است. سرور رجیستر SIP پیام های مربوط به ثبت کاربرها را در شبکه مدیریت می کند. سرور پراکسی درخواست های SIP که توسط عامل های کاربری تولید می شوند را دریافت کرده و با گرفتن اطلاعات از سرور رجیستر به سمت مقصد هدایت می کند. و در واقع وظیفه مسیر یابی را بر عهده دارد. سرور مسیردهی هم نقش درگاه و روتر را در شبکه های SIP بازی می کند. پراکسی سرور ها نقش های دیگری همچون احراز هویت، شناسایی کاربر، کنترل دسترسی به شبکه، ارسال مجدد درخواست ها در صورت عدم ارسال و همچنین تامین امنیت را ایفا می کند. نکته مهمی که باید بدان اشاره کرد این است که سرور رجیستر در واقع آدرس اینترنتی عامل های کاربری را به یک آدرس SIP نگاشت میدهد. آدرس SIP مشابه با آدرس های ایمیل و به شکل sip:userid@gateway.com است.

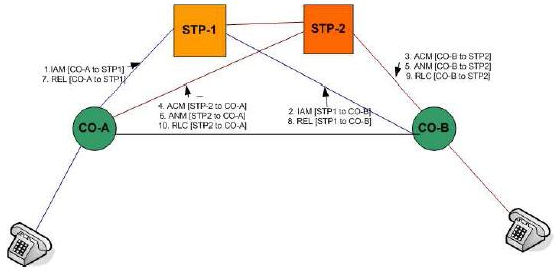
۲-۲ برقرای تماس در PSTN و SIP

مشترک آغاز کننده نشست که به اداره مرکزی مبدا متصل است با شماره گیری مشترک مورد نظر و ارسال آن به مرکز، مرکز یک پیام IAM تشکیل داده و به مرکزی که مشترک موردنظر بدان متصل است ارسال می کند. البته در صورتی که هر دو طرف نشست در یک حوزه [۶]قرار داشته باشند آنگاه این پیغام مستقیما به مقصد ارسال می شود. مرکز مقصد در دسترس بودن مشترک مورد نظر را بررسی کرده و در صورت در دسترس بودن مشترک یک پیغام تکمیل آدرس [۷] را برای مرکز مبدا ارسال می کند و در پی آن با جواب دادن مشترک مورد نظر توسط مرکز B که همان مرکز مقصد است، یک پیغام پاسخ [۸]به اداره مرکز A که همان مرکز مبدا است، ارسال می شود. در ادامه هر یک از طرفین می توانند با ارسال پیغام خاتمه [۹]به نشست خاتمه دهند. این فرآیند در شکل ۱ نمایش داده شده است. حال سناریوی مربوط به برقراری تماس در SIP[۱۰] را بررسی می کنیم.

---

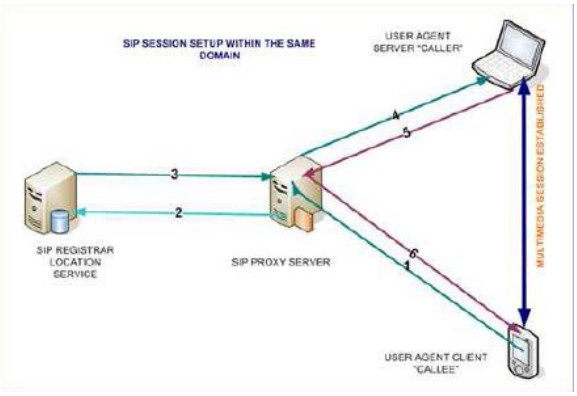
۲ Call setup
۳ Central office(CO)
۴ User agent client
۵ User agent server

۶ Domain
۷ Address complete message
۸ Answer Message
۹ Release message
۱۰ SIP call setup

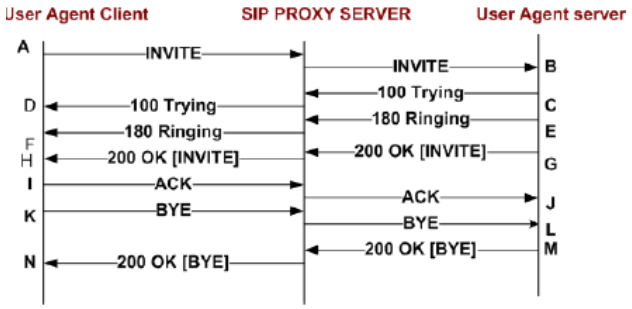


شکل ۱: برقراری تماس در PSTN [۳]

ابتدا عامل کاربری تماس مبدا (UAC) و مقصد (UAS) آدرس ها و در دسترس بودن خود را توسط سرور رجیستر ثبت می کنند. UAC پیغام INVITE را برای پراکسی سرور ارسال و سپس پراکسی سرور اطلاعات آدرس و مسیریابی مقصد را از سرور رجیستر دریافت می کند. این پیغام INVITE توسط پراکسی سرور به سمت UAS ارسال شده و در صورت تایید توسط مقصد، پیغام تایید برای UAC ارسال می شود و UAC یک پیغام ACK را ارسال می کند. در این مرحله فاز برقراری تماس پایان می پذیرد. شکل ۲، روند برقراری تماس بین UAC و UAS در یک دامین و شکل ۳ دیاگرام پیغام های مبادله شده بین عناصر مربوطه را نمایش می دهد.



شکل۲: برقراری ارتباط در SIP [۳]



شکل۳: دیاگرام پیغام های مبادله شده بین UACو UAS [۳]

۳- ارزیابی کارایی

در این بخش به مطالعه مدل های ارزیابی کارایی سرورهای پراکسی می پردازیم. به طور خاص منظور از ارزیابی کارایی در گام اول، تعیین دسته ای از معیار های مورد نظر برای سنجش بوده و در گام بعد مقایسه این معیار ها تحت شرایط مختلف شبکه است. بیشتر کارهایی که در حوزه VoIP مرتبط با SIP صورت گرفته است مربوط به قواعد مهندسی و تعریفات و الحاقات پروتکل است. کار اندکی در حوزه مدلسازی کارایی سرورهای پراکسی انجام شده است [۲]. در [۹] یک شبکه IMS مبتنی بر SIP ارائه شده است و کارایی شبکه با انتخاب معیارها و معیار های مورد نظر ارزیابی شده است. مدل های پیشنهادی مبتنی بر مدل های صف هستند. در [۹] کارایی SIP با حضور عنصر SIP-T(SIP telephone) مورد آنالیز قرار می گیرد. در حقیقت عملیات سیگنالینگ تلفن را در قالب پیغام های SIP کپسوله سازی و در مقصد ترجمه می کند. پیغام های مرتبط با برقرای تماس که بین سوئیچ های PSTN در جریان هستند کپسوله سازی شده و به عنوان داده در شبکه SIP انتقال داده می شوند. قالبی که توضیح داده شد در واقع مجتمع سازی دو نوع شبکه در یک قالب است که توسط SIP-T صورت گرفته است. SIP-T علاوه بر این روند نرمال اشاره شده، گاهی نیز برخی داده های مرتبط با سرآیند برقراری تماس در PSTN را به نزدیکترین حالت معادل ممکن موجود در SIP تبدیل می کند که به مسیریابی در پراکسی سرورهای SIP کمک می کند. در [۱۰] همچنین آنالیزی که صورت گرفته با مدل صف تاخیری و با استفاده از زنجیره مارکوف و با مدل صف M/G/۱ است.در مقابل در[۲] آنالیز کارایی تحت شرایط نرخ ورود [۱۱]متغیر، نرخ سرویس متغیر [۱۲]و تاخیرات ابتدا انتهای [۱۳]متغیر صورت گرفته است. همچنین یکی از کارهای مهمی که در [۲] صورت گرفته، ارائه

[۱۱] Arrival Rate
[۱۲] Service rate
[۱۳] End to end delay

یک مدل برای ارزیابی دو پارامتر مهم در دسترس بودن و نرخ از بین رفتن تماس در SIP است که به لحاظ مقیاس پذیری شبکه های VoIP تحت SIP بسیار مهم بوده و همچنین آنالیزی در مورد استفاده از SIP در شبکه IMS مرتبط با پروژه ۳GPP[۱۴]صورت گرفته است[۱۱]. این تحلیل در ارتباط با استفاده از SIP در قالب یک معماری متمرکز کنترل[۱۵]شده است. در [۱۲] یک رهیافت برای استفاده از مدل شبکه های صف به منظور تحلیل یک شبکه ساده SIP ارائه شده است و تمرکز آن بیشتر بر روی ویژگی های گذرا مربوط به تعداد تماس های پردازش شده قبل از یک حالت خرابی سرور[۱۶] است. همانطور که بعدا اشاره می شود در [۲] فرآیند ارزیابی کارایی در قالب یک مدل سلسله مراتبی انجام می شود که در سطح بالا یک مدل مارکوفی و در سطح پایین یک مدل شبکه صف را شاهد هستیم. همچنین یک رهیافت در مورد بررسی اعتبار و ارزش یک مدل ارائه شده، استفاده از معادلات فرم بسته است که ویژگی های اصلی یک مدل کارایی را با هدف اعتارسنجی مدل مورد بررسی قرار می دهد. در ادامه کارهای انجام شده در این زمینه به طور مبسوط ارائه می شود.

۳-۱ آنالیز پراکسی سرور با استفاده از مدل شبکه صف[۱۷]

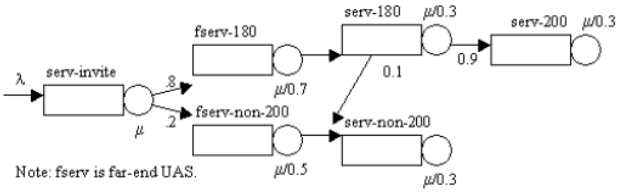
۳-۱-۱ معرفی مدل

همانطور که در بخش ۲-۲ اشاره شد، در شبکه PSTN، سیگنالینگ توسط پروتکل SS۷ صورت می گیرد که روی یک کانال مجزا انجام می شود و از لحاظ قابلیت اطمینان بسیار بالاست.

اما در سیستم VoIP که در بستر اینترنت و بر اساس سیگنالینگ SIP صورت می گیرد، دیگر یک شبکه مجزای سیگنالینگ مانند PSTN وجود ندارد و ارسال بسته های داده به صورت Best effort است و حساسیت قابل توجهی نسبت به پارامترهای شبکه دارد که در این صورت ارائه یک مدل ارزیابی بسیار حائز اهمیت است.

معیار های مورد علاقه برای اندازه گیری در [۲] یکی میانگین زمان پاسخ[۱۸] و دیگری میانگین تعداد کارهای[۱۹]سیستم در حالت پایدار هستند. میانگین زمان پاسخ معادل با زمان t۱ از زمانی که پیغام درخواست INVITE ارسالی به دست پراکسی سرور مبدا می رسد تا t۲ زمانی که پراکسی سرور یک پاسخ به سمت مبدا ارسال می کند در نظر گرفته می شود. پارامتر میانگین تعداد کار در سیستم نیز به عنوان تعداد متوسط تماس هایی است که در حال حاضر در حال پردازش می باشند یا اینکه در صف برقراری تماس برای آینده قرار گرفته اند.

در ادامه به بررسی رفتار معیار های مورد علاقه، به عنوان تابعی از پارامترهای متغیر شبکه و سپس مقایسه نتایج شبیه سازی و نتایج به دست آمده از تحلیل پرداخته می شود. در [۲] رفتار دو معیار مذکور را تحت شرایط زمان سرویس متغیر[۲۰]پیغام های INVITE و نرخ ورودی متغیر[۲۱]پیغام INVITE و همچنین تاخیر انتشار بین پراکسی سرورهای مبدا و مقصد بررسی می شود. مدلی که در [۲] تشریح می شود در شکل ۴ نمایش داده شده است.



شکل۴: مدل ارزیابی بدون تاخیرشبکه [۲]

همانطور که مشاهده می شود پراکسی سرور به صورت یک شبکه صف باز جلو رونده[۲۲]مدل می شود و کارهای ورودی همان درخواست های INVITE هستند که توسط پراکسی از سمت UAC دریافت می شوند ۶ ایستگاه در شکل ۴ مشاهده می شوند و هر ایستگاه معادل با یک مرحله در توالی فرآیند برقراری تماس است. دو ایستگاه fserv-۱۸۰ و fserv-non۲۰۰ مربوط به UAS مقصد و باقی ۴ ایستگاه دیگر مربوط به سرور پراکسی هستند. ملاحضاتی در این مطالعه اعمال شده است که در شبکه های عملیاتی واقعی این ملاحضات ما را با مشکل دچار می کند. مثلا در مورد این مدل تولید پاسخ ringing-۱۸۰ و پاسخ serv-۲۰۰ که به نشانه پاسخ مثبت UAS به درخواست INVITE است بدون لحاظ تاخیر بین آنها صورت می گیرد و این در شبکه عملیاتی به صورت حتمی نیست. در صورت اعمال تاخیر در بین

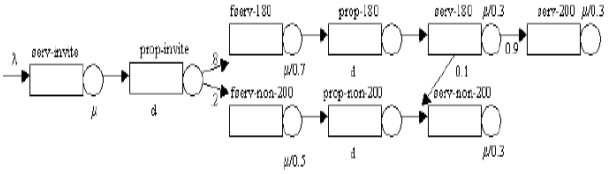
[۱۴] Third Generation partnership protocol
[۱۵] Centrally controlled
[۱۶] Server down state
[۱۷] Queue network
[۱۸] Mean response time
[۱۹] Mean number of jobs
[۲۰] Variable service time
[۲۱] Variable arrival rate
[۲۲] Open Feed-forward Queue network

تولید این دو پیغام، محاسبات در مورد معیار های مورد نظر تغییر خواهد یافت.

همچنین مدل شکل ۴ بدون تاخیر انتشار در شبکه صف در نظر گرفته شده است. احتمالاتی نیز در مورد گذر به حالت های مختلف مشاهده می شود که این احتمالات بر اساس داده های شبیه سازی جمع آوری شده از بسترهای مختلف عملیاتی اعمال شده است و خود مساله ای مهم در صحت این مدل است.

یکی دیگر از مسائلی که در مورد این مدل جای بحث و مطالعه دارد نرخ سرویس هایی است که در مورد ایستگاه های مختلف در نظر گرفته ایم. این نرخ ها به عنوان فرضیات مدل آورده شده اند و اینجا نیز در مورد صحت مدل که محاسبات اندازه گیری کارایی آن بر اساس این فرض در مورد نرخ سرویس ها صورت گرفته، جای بحث دارد. حالتی که برای مدل پیشنهادی در شکل ۴ در نظر گرفته شده، حالتی است که هر دو عامل کاربری UAC و UAS در یک دامین یکسان قرار دارند و مدل جریان درخواست ها و پاسخ ها به شکل UAC=> proxy => UAS است و حالتی دیگر نیز بررسی می شود که UAC و UAS در دامین های مختلف قرار دارند. هر کدام از ایستگاه ها یک صف M/M/۱ هستند و داده های ورودی از یک منبع خارجی وارد می شوند و هیچ فیدبکی بین ایستگاه ها وجود ندارد. با روش های استانداردی که وجود دارد (معادلات مربوط به صف M/M/۱) مقادیر پارامترهای مورد نظر که همان زمان پاسخ میانگین و تعداد کارهای میانگین هستند محاسبه شده اند.

در یک حالت دیگر که به حالت شبکه های عملیاتی نزدیکتر است تاخیرات انتشار بین سرور پراکسی های مجاور و UAC و UAS نیز لحاظ می شود. تاخیرات انتشار در مسیر به صورت یک سرور تاخیر یا با اسمی دیگر می توان گفت یک ایستگاه صف M/M/∞ مدل می شود که زمان سرویس میانگین آن همان میانگین تاخیر انتشار است. منظور از تعداد سرور ∞ در واقع معادل گرفتن این تاخیر با تعدادی نامشخص از سرورهای پراکسی بین UAC و UAS است. این مدل بسط داده شده در شکل ۵ آورده شده است.



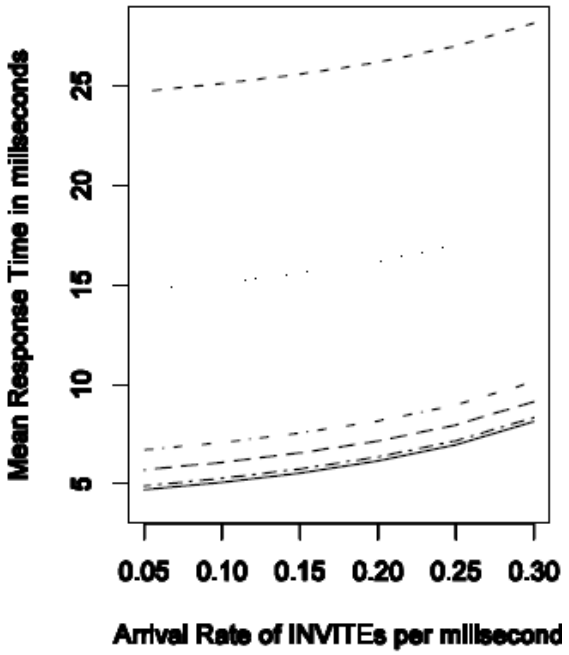
شکل ۵: مدل ارزیابی با تاخیر شبکه[۲]

ایستگاه هایی که با prop در ابتدای آنها نامگذاری شده اند مربوط به تاخیر انتشار از پراکسی به عناصر بعدی SIP هستند. این مدل با تاخیر صفر، مانند مدل شکل ۴ است.

محاسبات مربوط به معیار های مورد علاقه نیز مانند قبل صورت می گیرد با توجه به این نکته که زمان سرویس میانگین برای شبکه صف در این مدل همان زمان میانگین تاخیر است.

۳-۱-۲ نتایج آنالیز کارایی برای سرورهای پراکسی

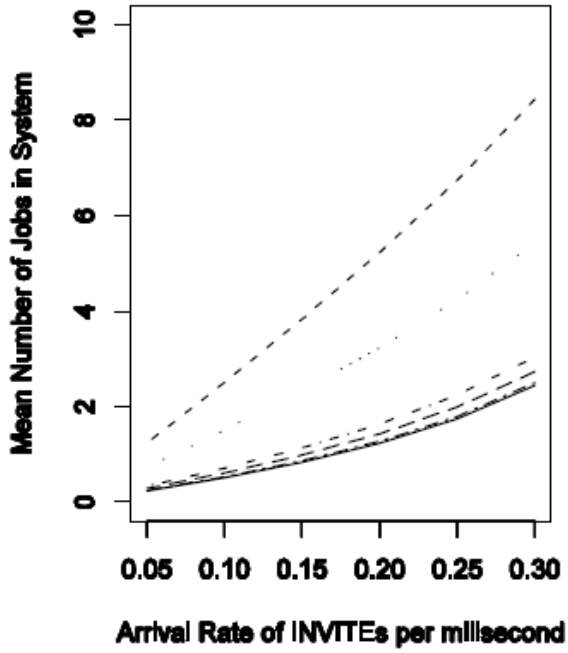
۳-۱-۲-۱ نتایج مدل معیار با روش پیشنهادی، دو معیار مذکور در مورد پراکسی سرور محاسبه و نتایج مربوطه در شکل ۶ نمایش داده شده اند.



شکل ۶: زمان پاسخدهی میانگین تحت نرخ ورودی متغیر[۲]

همانطور که مشاهده می شود با افزایش تاخیر انتشار بین پراکسی سرور ها، میانگین زمان پاسخ (R) افزایش یافته و تقریبا این مقدار R به صورت خطی با نرخ ورود پیغام های INVITE تغییر می یابد. البته فرض ثابت بودن نرخ سرویس در نظر گرفته شده است. در دامنه مقادیر نرخ ورودی و تاخیرات انتشار، مقادیر به دست آمده برای R برای سیستم های عملیاتی قابل قبول هستند.

شکل ۷ مقادیر به دست آمده برای متوسط تعداد کارهای سیستم به ازای تغییرات نرخ ورودی را نمایش می دهد.

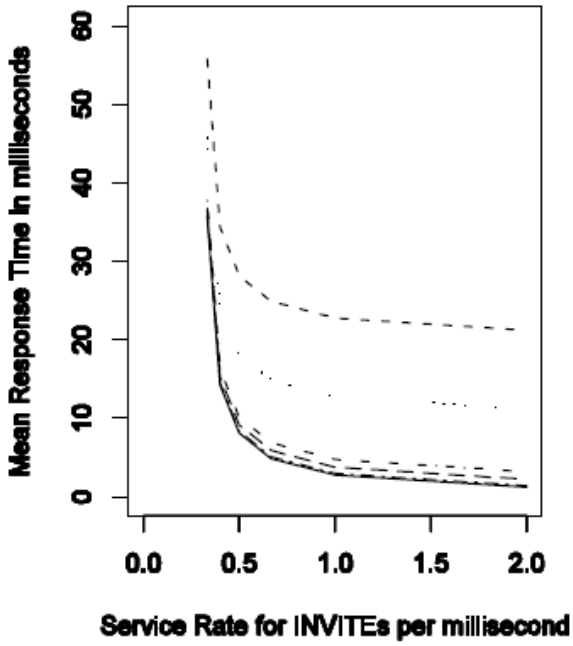


شکل ۷: تعداد میانگین کار تحت نرخ ورودی متغیر[۲]

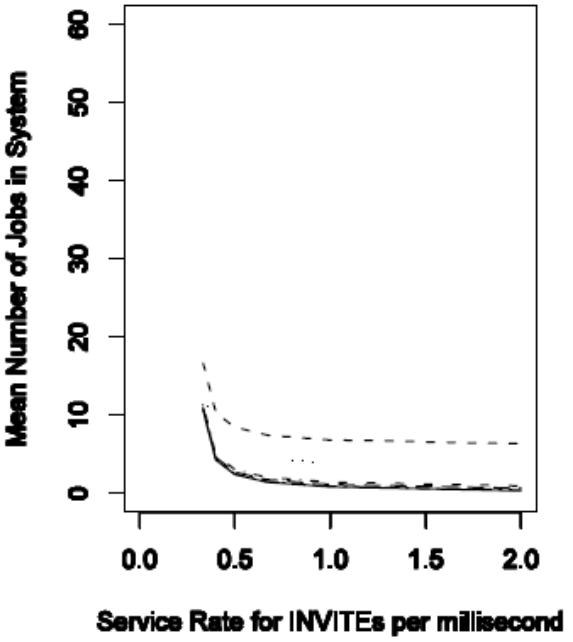
تعداد کارهای موجود در سیستم بسیار کم می باشد حتی در مورد تاخیرات انتشار بالا نیز اینطور مشاهده می شود. کاری که باید اینجا صورت گیرد بررسی رفتار این دو معیار در حضور تغییرات نرخ سرویس برای پردازش درخواست های INVITE است که شکل های ۸ و ۹ این رفتار را نشان می دهند. تغییرات میانگین زمان پاسخ بعد از یک مقدار نرخ سرویس به صورت خطی تغییر می کند که به لحاظ تحلیل حساسیتی می توان نتیجه گرفت که برای نرخ سرویس های کمتر از ۰٫۶، بر اساس داده های این مدل شبکه دارای پایداری نیست و کماکان متوسط تعداد کارهای موجود در سیستم حتی در تاخیرهای انتشار بالا بسیار کم می باشد. در بخش های بعدی صحت این داده ها با مقادیر شبیه سازی بررسی می شود.

۳-۱-۲-۲ بسط مدل با سرور های چندتایی

در مدل توسعه داده شده در [۲،۳] حالتی در نظر گرفته شده است که به منظور بالا بردن مقیاس پذیری از پراکسی سرور چندتایی [۲۳]استفاده شده است. دقیقا معادل با مدل شکل ۴ است اما به جای صف های M/M/۱ از صف M/M/m که m تعداد سرورها را بیان می کند استفاده می کنیم. نتایج آنالیز این مدل توسعه داده شده در شکل ۱۰ آورده شده است.

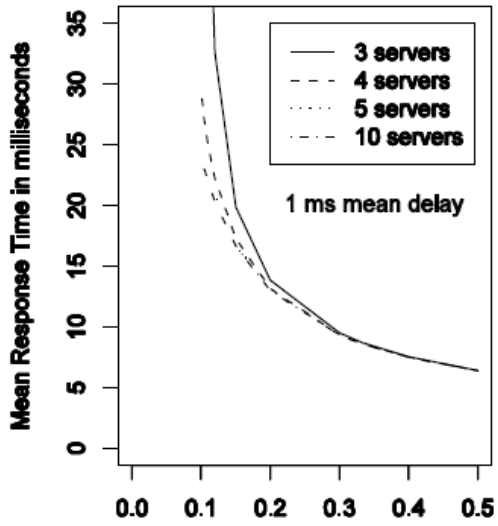


شکل ۸: زمان پاسخ میانگین تحت نرخ سرویس متغیر[۲]



شکل ۹: متوسط تعداد کارها در ازای نرخ سرویس متغیر[۲]

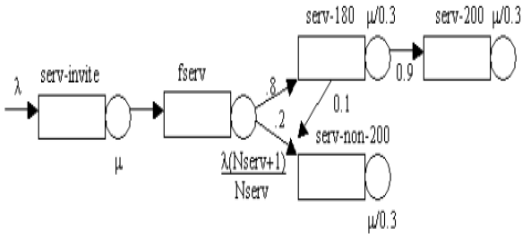
[۲۳] Multi-server proxy host



شکل ۱۰: متوسط زمان پاسخ سرور پراکسی چندتایی در ازای نرخ سرویس متغیر و تاخیر شبکه[۲]

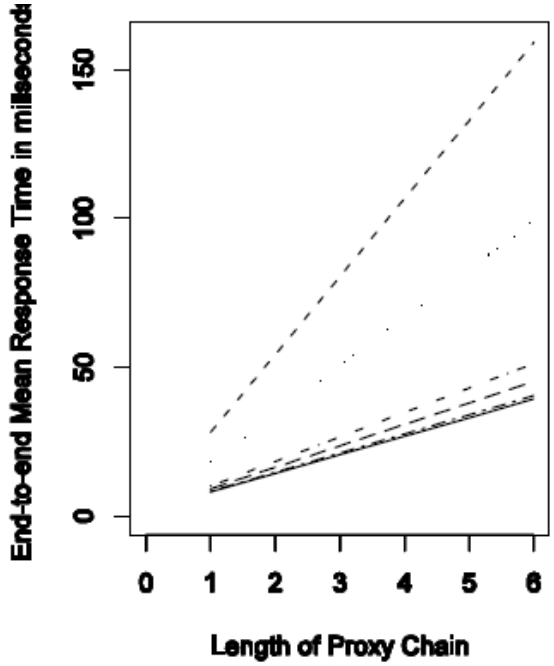
در شکل ۱۰ به ازای مقادیر کمتر از یک آستانه معین نرخ سرویس، زمان میانگین پاسخ بسیار به تغییرات نرخ سرویس حساسیت نشان می دهد. مشاهده می شود که حد آستانه برای پارامتر زمان سرویس در مورد سناریوی پراکسی سرورهای چندتایی که مقادیر کمتر از این حد، پایداری شبکه را تضمین نمی کند و حتی همان سناریوی بخش **۳=۱=۲=۱** عملکرد بهتری از خود نشان می دهد. همچنین به ازای مقادیر بیشتر از این حد آستانه، علاوه بر حساسیت کمتر R نسبت به تغییرات نرخ سرویس، مقادیر R نسبت به تعداد سرورها تغییری نداشته و مستقل از این مساله است. می توان نتیجه گرفت تعداد کمی سرور چندتایی با یک نرخ سرویس آستانه برای تامین R مناسب، کافی بوده و نیازی به افزایش تعداد سرورهای پراکسی چندتایی نیست.

می توان سناریویی در نظر گرفت که در آن یک خوشه ای[۲۴] از پراکسی سرور ها داشته باشیم. در [۲] ، این شبکه خوشه ای پراکسی ها را همانند مدل شکل ۴ اما با تغییراتی در آن مدل کرده است که در شکل ۱۱ این مدل بسط داده شده، نمایش داده شده است.



شکل ۱۱: مدل بسط داده شده برای شبکه خوشه ای پراکسی ها[۲]

نتایج آنالیز بر اساس این مدل در شکل ۱۲ آورده شده است. شکل ۱۲ نشان می دهد که با افزایش زمان تاخیر میانگین، شیب تغییرات زیاد شده و حساسیت معیار های مورد نظر به ازای طول زنجیره پروکسی ها افزایش می یابد.



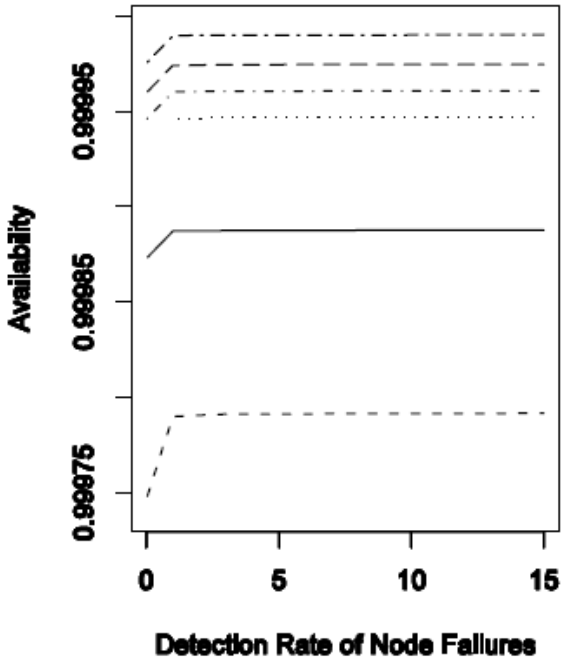
شکل ۱۲:متوسط زمان پاسخ انتها به انتها در ازای تغییرات طول زنجیره پراکسی[۲]

نتیجه ای که از این تحلیل می توان گرفت به دست آوردن یک حد برای طول زنجیر در سناریوی شبکه پراکسی سرورها است که در طول های بالا، R به دلیل بالا رفتن زیاد تاخیر انتشار بین عناصر این زنجیر بسیار بالا می رود .

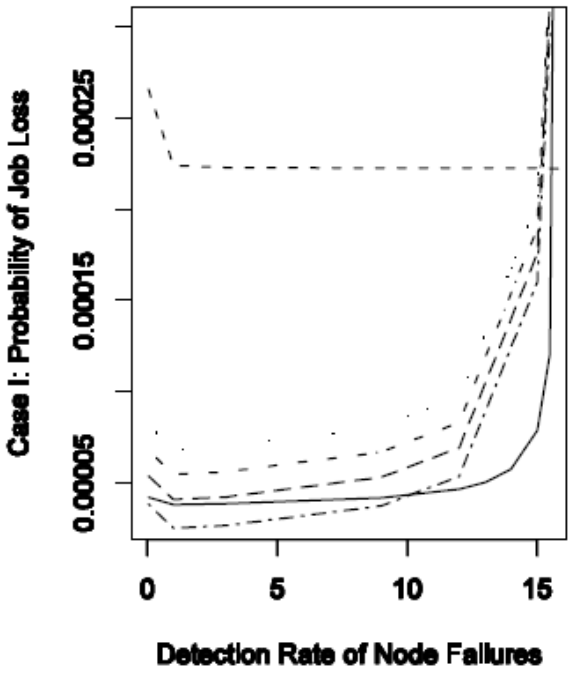
[۲۴] Cluster

۳-۱-۳ تحلیل قابلیت اطمینان[۲۵]

آنالیز قابلیت اطمینان پراکسی سرور، دو معیار مدنظر است که یکی فراهم بودن [۲۶]سرور و دیگری احتمال از دست دادن کار، در حالت پایدار است.در [۱۳] یک مدل قابلیت اطمینان سلسله مراتبی در دو سطح بالا و پایین ارائه شده است که به دلیل استفاده از عبارات فرم بسته برای محاسبه دو معیار فراهم بودن و احتمال از دست رفتن تماس، مدلی نسبتا قوی و کارآمد محسوب می شود و در کارهای بعدی نیز به عنوان مدل مرجع در نظر گرفته شده است. مساله ای که در اینجا مطرح است جلوگیری از دست رفتن تماس ها یا همان کارهای ورودی سیستم به هر روش ممکن است که این کار در مدل ارائه شده در[۲] بر اساس ایده تکرار [۲۷]صورت می گیرد. پراکسی سرور های تکرار کنار پراکسی سرورهای اصلی وجود دارند که فراهم پذیری سرور پراکسی را بالا برده و تعداد تماس های از دست رفته را کم می کند. نتایج به دست آمده از این مدل در سناریو های مختلف که از پراکسی سرور تکراراستفاده شود یا نشود در [۲،۱۳] در شکل ۱۳ و ۱۴ نمایش داده شده اند. پارامتر "نرخ کشف خطاهای گره" مربوط به نرخ کشف خطاهایی است که در سطح یک گره منجر به از دست رفتن یک تماس یا مسدود کردن یک جریان پیغام های درخواست / پاسخ می شود، که با بالا رفتن آن چون هزینه سربار این الگوریتم بسیار بالا می رود ترافیک شبکه نیز رشد قابل توجهی پیدا کرده و احتمال از دست رفتن تماس افزایش می یابد. حالت های قبول برای شبکه های عملیاتی در حالت سرورهای پراکسی تکرار صورت می گیرد و عدم حضور این سرور پراکسی های تکرار، عملا شبکه را از لحاظ قابلیت اطمینان نامطمئن می کند.



شکل ۱۳: فراهم بودن سرور[۲]



شکل ۱۴: احتمال از دست رفتن تماس[۲]

مساله دیگری که در تحلیل قابلیت اطمینان شبکه های SIP وجود دارد، سیاست اتخاذ شده در طول زمان خراب بودن

[۲۵] Reliability analysis
[۲۶] Availability
[۲۷] Replication

[۲۸]سرور پراکسی است که یک راه حل ساده، در نظر گرفتن بافر مناسب برای تماس هایی است که در زمان خراب بودن سرور وارد سیستم شده اند. نتایج آنالیز با استفاده از مدل قابلیت اطمینان در [۲،۱۳] نشان می دهد که زمان مورد نیاز برای پردازش تماس های بافر شده پس از شروع مجدد پراکسی سرور، در شرایط عدم استفاده از پراکسی سرورها در شبکه های عملیاتی قابل قبول نیست [۲].

۳-۱-۴ مقایسه داده های آنالیزی با داده های شبیه سازی شده

نتایج به دست آمده از آنالیز های مدل ارائه شده در بخش ۳ برای صف M/M/۱ با نتایج شبیه سازی شده، مقایسه و در شکل های ۱۵ و ۱۶ ارائه شده اند[۲].

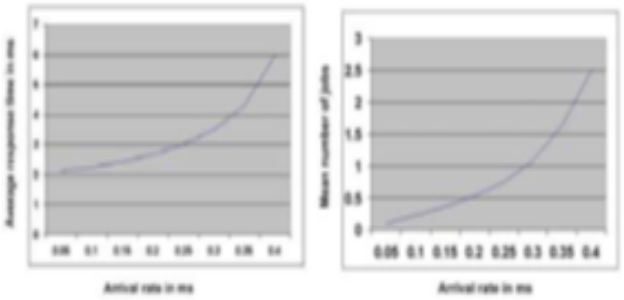
λ	0.05	0.10	0.15	0.20	.25	0.30	0.35	0.40
L	0.22	0.49	0.82	1.2	1.7	2.4	3.4	5.4
W in ms	5.9	6.3	6.9	7.5	8.6	10.2	12.5	15.5

شکل ۱۵: نتایج تحلیلی مدل ارزیابی کارایی[۳]

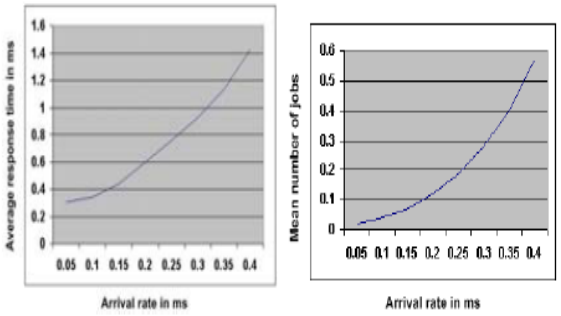
λ	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
L	0.02	0.04	0.07	0.12	0.19	0.28	0.40	0.57
W in ms	0.31	0.35	0.44	0.65	0.76	0.93	1.13	1.42

شکل ۱۶: نتایج شبیه سازی شده مدل ارزیابی کارایی[۳]

در [۳] یک تغییر در نوع صف پراکسی سرور اعمال شده است و صف های سرور M/D/۱ در نظر گرفته شده اند که در آن نرخ سرویس درخواست ها و پاسخ ها ثابت و مشخص فرض شده است. این تغییر، نتایجی که در شکل های ۱۷ و ۱۸ نمایش داده شده اند به همراه داشته است.



شکل ۱۷: نتایج تحلیلی صف M/D/۱ [۳]



شکل ۱۸: نتایج شبیه سازی شده صف M/D/۱ [۳]

جالب است که در مورد مدل ارزیابی با استفاده از صف M/M/۱ که نرخ سرویس را فقط در مورد پیغام های درخواست INVITE ثابت در نظر گرفته شده است، نتایج تحلیلی و شبیه سازی شده تفاوت فاحشی دارند. اما در مورد حالت M/D/۱ که نرخ سرویس در مورد همه پیغام های درخواست و پاسخ SIP، ثابت فرض شده، عملکرد مدل بهبود قابل توجهی پیدا کرده و نتایج شبیه سازی و آنالیز بسیار به هم نزدیک هستند[۳].

در مدل شبکه صف ارائه شده، همانطور که اشاره شد تعداد میانگین کارهای موجود در سیستم حتی در مورد سناریوی شبکه ای از پراکسی سرورها نیز بسیار کم هستند که خود این مساله می تواند ماهیت پروتکل SIP را زیر سوال ببرد. البته یک توجیه این مساله می تواند مرتبط با لحاظ نکردن زمان بین ارسال پیغام پاسخ ۱۸۰-RINGING و ۲۰۰-ok باشد که به زمان برداشتن گوشی توسط کاربر مقصد بستگی داشته، متغیر است و با در نظر گرفتن این پارامتر تعداد بیشتری کار در حالت زنگ خوردن [۲۹]به سیستم اضافه می شوند.

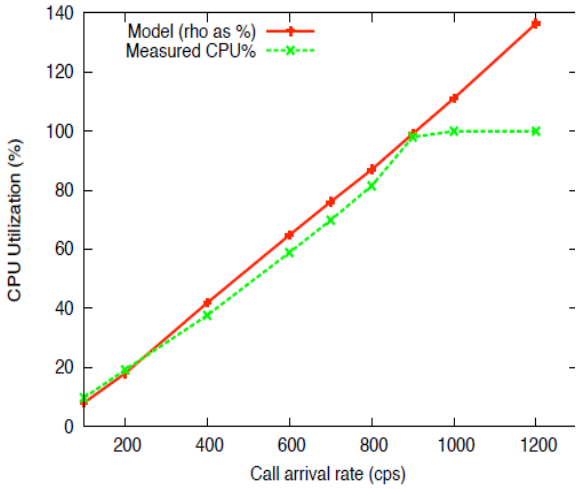
۳-۲ ارزیابی با استفاده از مدل صف M/G/۱

در[۴،۶] ازOpen SIPS SPS که یک نرم افزار پراکسی سرور SIPS متن باز است به منظور بررسی رفتار دو معیار زمان انتظار

[۲۸] Server down-time

[۲۹] Ringing State

و بهره وری پردازنده به ازای تغییرات نرخ ورودی درخواست های INVITE استفاده شده است. یک مدل صف M/G/۱ با استفاده از SPS ارائه شده که در آن دو پارامتر مهم که در بسیاری از کارهای مشابه نادیده گرفته شده اند را اثر داده و محاسبات با حضور این دو پارامتر انجام شده اند. یکی از این دو پارامتر مربوط به تاخیر صف است که با افزایش نرخ ورود تماس ها به صف، تماس های متاخر بافر شده و زمان انتظار افزایش می یابد. پارامتر دوم مربوط به سربار وقفه [۳۰]است. با افزایش نرخ ورود پاکت، تعداد وقفه ها به طور مستقیم زیاد می شود. این وقفه ها در قالب زمان مورد نیاز برای مدیریت وقفه، زمان تعویض زمینه و همچنین زمان صرف شده به دلیل عدم وجود روتین های فراخوانی شده بعدی توسط پشته شبکه در حافظه کش پردازنده، هستند [۵]. نتایج شبیه سازی شده و تحلیلی این مدل در شکل های ۱۹ و ۲۰ نمایش داده شده اند.



شکل ۱۹: نتایج شبیه سازی شده و تحلیلی بهره وری پردازنده پراکسی سرور SPS [۶]

شباهت بسیار زیاد نتایج تحلیلی و شبیه سازی شده در مورد دو معیار مورد علاقه در مدل صف M/G/۱ با لحاظ کردن دو پارامتر مذکور در ۳-۲ ، مدل مذکور را به عنوان یک رهیافت مناسبی برای ارزیابی کارایی سرور پروکسی SIP ارائه می کند [۶].

در [۸] دو معیار جدید مورد بررسی قرار گرفتند. نقش شناسایی کاربر و پروتکل لایه کاربردی بر روی کارایی پروکسی سرور در بستر OpenSER که یک نرم افزار متن باز پراکسی سرور SIP است آنالیز شده اند و نتایج به دست آمده از این آنالیزها تغییرات بسیار شدید معیار گذردهی [۳۱]را به صورت نمایی از درجه ۲ و ۴ بیان می کنند. فرآیند شناسایی کاربر به دلیل دسترسی زیاد به پایگاه داده، بیشترین تاثیر بر روی کارایی پراکسی سرور را دارد. بار سیستم نیز تاثیر زیادی بر روی توزیع مقادیر معیار تاخیر در سیستم به خصوص درشرایط اضافه بار [۳۲]ایجاد می کند.

۴- نتیجه گیری

همانطور که مشاهده شد با استفاده از مدل های مختلف و توسعه آنها تحت شرایط گوناگون، رفتارمعیار های کارایی مورد علاقه به ازای تغییرات برخی پارامترها همانند نرخ ورود بسته های INVITE، زمان سرویس پردازش بسته ها، تاخیر انتشار بین گره های مختلف و ... بررسی شدند.

تحلیل بااستفاده از مدل های بسط داده شده که در آنها شبکه ای از سرور های پراکسی در نظر گرفته شده اند صورت گرفت. نتیجه گرفته شد که تعداد سرورهای اضافه شده به منظور بالا بردن مقیاس پذیری و بهتر کردن مقادیر پارامترهای هدف، یک حد بالا داشته و با گذر از این حد، کارایی به طور چشم گیری کاهش می یابد.

در یک روند کلی، برای نزدیکتر کردن هرچه بیشتر نتایج شبیه سازی شده و نتایج مدل، یک مدل مرجع بسط داده می شود. بسط مدل با لحاظ کردن عواملی هستند که در قدیم در نظر گرفته نشده اند. مثلا می توان عناصری مانند سرور احراز هویت که برای فاز شناسایی کاربر استفاده می شود را از نقطه نظر بار ترافیکی و به تبع آن تاخیری که در زمان پاسخ نهایی ایجاد می کند، را مهندسی مجدد نمائیم یا تاثیر آن در کارایی را بسنجیم.

در حالتی دیگر می توان تاثیر تغییرات در لایه های پایین تر مانند لایه شبکه یا انتقال را در کارایی پارامترهای هدف سنجش کرد. مثلا در لایه انتقال اگر انتقال امن TLS استفاده شود چه تاثیری بر روی کارایی پرارمترهای هدف در سرور پراکسی SIP می گذارد.

[۳۰] Interrupt overhead

[۳۱] Throughput
[۳۲] Overload

Model Parameters	Call Arrival Rate								
	100cps	200cps	400cps	600cps	700cps	800cps	900cps	1000cps	1200cps
λ (packets/sec)	600	1200	2400	3600	4200	4800	5400	6000	7200
$\alpha(\lambda)$	3.0	5.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
$E[X] = K^s_{sockq}(\lambda) + K_{snd} + K_{copy} + T_{sip}$	133.4	149.88	174.53	179.98	181.09	181.185	183.49	185.23	189.5
$\rho = \lambda E[X]$	0.080	0.1798	0.4188	0.6479	0.7606	0.8698	0.9908	1.111	1.3644
W (model)	6.0	16.89	64.21	169.19	293.98	617.77	10171.9	N/A	N/A
W (measured) $= K^w_{sockq}$	5.485	13.59	53.27	190.92	333.52	664.57	9049.07	27997	36194

شکل ۲۰: نتایج شبیه سازی شده و تحلیلی زمان انتظار [۶]

مراجع

[۱] J.L. Wang, " Traffic Routing and Performance Analysis of Common Channel Signaling System No. ۷ Network", *Global Telecommunications Conf.(GLOBECOM), pp.* ۳۰۱-۳۰۵*, vol.* ۱, Texas, USA, December ۱۹۹۱.

[۲] V.K.Gurbani, L. Jagadeesan, V.B. Mendiritta, "Charecterizing the Session Initiation Protocol (SIP) Network Peformance and Reliability", *ISAS* ۲۰۰۵*: LNCS* ۳۶۹٤*, pp.* ۱۹٦-۲۱۱, ۲۰۰۵

[۳] Sureshkumar V. Subramanian, Rudra Dutta "Comparative Study of M/M/۱ and M/D/۱ Models of a SIP Proxy Server", IP Communications Business Unit, CISCO, Research Triangle Park, North Carolina ۲۷۷۰۹, USA

[٤] OpenSIPS: Open source implementation of a SIP server. http://www. opensips.org/.

[٥] F. Liu *et al. Characterizing and Modeling the Behavior of Context Switch Misses. In* Proceedings of ACM PACT, October ۲۰۰۸.

[٦] Ramesh Krishnamurthy, George N. Rouskas "Evaluation of SIP Proxy Server Performance: Packet-Level Measurements and Queuing Model", Department of Computer Science, North Carolina State University, Raleigh, NC ۲۷٦۹٥-۸۲۰٦ USA

[۷] M. Cortes, J. R. Ensor, and J. O. Esteban. On SIP performance. Bell Labs Technical Journal, ۹(۳):۱٥٥{۱۷۲, Nov ۲۰۰٤.

[۸] Erich M. Nahum, John Tracey, and Charles P. Wright "Evaluating SIP Proxy Server Performance**",** IBM T.J. Watson Research Center Hawthorne, NY, ۱۰٥۳۲ , fnahum,traceyj,cpwrightg@us.ibm.com

[۹] Vemuri and J. Peterson, "Session Initiation Protocol for Telephones (SIP-T): Context and Architectures", IETF RFC ۳۳۷۲, September ۲۰۰۲, <http://www.ietf.org/rfc/rfc۳۳۷۲.txt>

[۱۰] J-S. Wu and P-Y Wang, "The performance analysis of SIP-T signaling system in carrier class VoIP network", Proceedings of the ۱۷th IEEE International Conference on Advanced Information Networking and Applications (AINA), ۲۰۰۳.

[۱۱] Zhu, "Analysis of SIP in UMTS IP Multimedia Subsystem", MSc. Thesis, Computer Engineering, North Carolina State University, ۲۰۰۳.

[۱۲] F. Lipson, "Verification of Service Level Agreements with Markov Reward Models," South African Telecommunications Networks and Applications Conference, September ۲۰۰۳

[۱۳] S. Garg, et al., "Performance and Reliability Evaluation of Passive Replication Schemes in Application Level Fault Tolerance," Proceedings of the ۲۹th Annual International Symposium on Fault-Tolerant Computing, Madison, WI, June ۱۹۹۹

سرویس توزیع محتوی بر روی شبکه‌های نسل سوم تلفن همراه

سیما راست خدیو[1]، مرتضی آنالویی[2]

[1] دانشجوی کارشناسی ارشد رشته‌ی مهندسی فنآوری اطلاعات
rastkhadiv@comp.iust.ac.ir

[2] دانشیار گروه فنآوری اطلاعات دانشکده‌ی کامپیوتر دانشگاه علم و صنعت ایران
analoui@iust.ac.ir

چکیده

امروزه دسترسی به اطلاعات در کمترین زمان ممکن بسیار مورد توجه کاربران شبکه‌های اینترنت و موبایل قرار گرفته است. در گذشته مدل‌های سرور/مشتری برای ارائه‌ی سرویس و محتوی به کاربران مورد استفاده قرار گرفته می‌شد، اما این روش‌ها با افزایش تعداد کاربران و نیز افزایش محتوی و سرویس‌های ارائه شده سازگار نبوده و قابل گسترش نیستند. در نتیجه مدل‌هایی تحت عنوان شبکه های توزیع محتوی معرفی گشتند که با قرار دادن تعدادی سرور، به عنوان جانشین سرور اصلی، در کشورها و مناطق مختلف بار روی سرور اصلی را کم کرده و وظیفه‌ی ارسال محتوی و ارائه‌ی سرویس به کاربران را به نزدیکترین سرور جانشین به آنها محول می‌کند. این روش که بسیار قابل گسترش است بازدهی بسیار خوبی نسبت به راه‌حل‌های مشابه از خود نشان داده است.

با ظهور بسترهای موبایل و فراگیر شدن شبکه‌های نسل سوم تلفن همراه این ایده به وجود آمد که با استفاده از مزایای منحصر به فرد این شبکه‌ها، همانند سرعت دسترسی به اینترنت بسیار بالا برای کاربران در حال حرکت، سرویس‌های شبکه‌های توزیع محتوی بر این بستر ارائه شوند. در این تحقیق به مطالعات اولیه و بررسی چالش‌های پیش رو جهت دستیابی به این هدف پرداخته شده است.

کلمات کلیدی

شبکه های توزیع محتوی، شبکه های توزیع محتوای موبایل،شبکه‌های نسل سوم تلفن همراه، سرورهای جانشین، مسیریابی

۱- مقدمه

امروزه دسترسی سریع کاربران به اطلاعات به ویژه داده‌های چند رسانه‌ای[1] به شدت مورد توجه قرار گرفته است و با پیشرفت فناوری‌های دسترسی به اینترنت پرسرعت بستر لازم برای این کار فراهم گشته است. با این وجود ارائه‌ی این سرویس به تعداد زیادی از کاربران که در مناطق متعددی پراکنده شده‌اند با چالش‌های زیادی رو به رو است. ارائه‌ی کیفیت سرویس[2] حداقلی و نیازمندی به مدل‌های سرویس چند پخشی[3] از چالش‌هایی است که به دلیل عدم وجود امکان برآورده‌سازی آنها در شبکه‌ی اینترنت موجود انگیزه‌های ارائه‌ی راه‌حل‌های جدید شده‌اند. در این راستا شبکه‌های توزیع محتوی با هدف توزیع محتوی و در عین حال بهینه کردن مصرف پهنای باند، بهبود دسترس‌پذیری[4] و صحت داده‌ها به عنوان راه‌حلی کارا معرفی گشتند. [۱]

در ادامه‌ی این مقاله، در بخش ۲ به معرفی شبکه‌های توزیع محتوی و معماری آنها پرداخته می‌شود. در بخش ۳ جزئیات ساختار و دسته‌بندی این شبکه‌ها عنوان می‌گردد. در بخش ۴ نکاتی در باب شبکه‌های توزیع محتوای موبایل ذکر می‌گردد و در قسمت ۵ به معرفی شبکه‌های نسل سوم تلفن همراه پرداخته می‌شود. بخش ۶ به توضیحی در باب جایابی محتوی در شبکه‌های توزیع محتوای موبایل اختصاص می‌یابد و در بخش ۷ به نتیجه‌گیری و جمع‌بندی مطالب مطرح شده پرداخته می‌شود. در انتها نیز تمامی مراجعی که در تدوین این مقاله مورد استفاده قرار گرفته‌اند ارائه خواهند شد.

۲- شبکه‌های توزیع محتوی

در مراجع مختلف تعاریف متعددی برای شبکه‌های توزیع محتوی ارائه شده است، اما به طور کلی می‌توان گفت شبکه‌های توزیع محتوی مجموعه‌ای از چندین PoP[5] هستند که در تلاشند با بیشترین کارایی و بازدهی و در کمترین زمان ممکن محتوی را به کاربران برسانند. در حقیقت شبکه‌های توزیع محتوی به ارائه‌ی سرویس‌هایی می‌پردازند که

با بیشینه کردن پهنای باند و بهبود دسترس‌پذیری و پشتیبانی صحیح باعث افزایش کارایی کل شبکه می‌شوند. [۲، ۴، ۵، ۶]

منظور از محتوی در این مقاله داده‌های دیجیتالی است که شامل داده‌های رمز شده[۶] و ابرداده‌ها[۷] می‌باشند. داده‌های رمز شده به محتوایی گفته می‌شود که مورد درخواست کاربران است و به سه دسته‌ی داده-های ایستا[۸] (همانند صفحات وب، قطعات کد قابل اجرا و...)، داده‌های پویا[۹] (همانند اسکریپت‌ها و انیمیشن‌ها) و داده‌های رشته‌ای[۱۰] (صوت و ویدئو) تقسیم می‌گردند. ابرداده‌ها نیز محتوایی هستند که امکان شناسایی، کشف، مدیریت داده‌های چند رسانه‌ای و تعبیر آنها را می-دهند. محتوی می‌تواند یک فایل ضبط شده باشد و یا به صورت زنده از منابع خود گرفته شوند (مانند پخش یک برنامه‌ی زنده). در عین حال این محتوی می‌تواند پایدار و یا گذرا باشد.

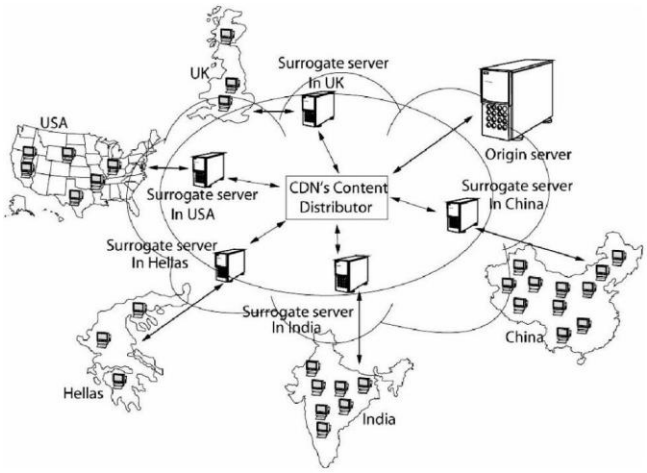
شبکه‌های توزیع محتوی شامل یک سرور اصلی[۱۱] هستند که طرف قرارداد با مشتریان خود، شامل سازمانها، سایت‌های اطلاع رسانی بزرگ مانند BBC و غیره می‌باشد. این سرور محتوای مربوطه را از مشتریان خود دریافت کرده و با کمک یک سیستم توزیع‌کننده[۱۲] این اطلاعات را به صورت تکراری بر روی سرورهای جانشین[۱۳] خود، که در سطح منطقه بر اساس مکانیزم‌های مشخصی توزیع شده‌اند، قرار می‌دهد. این سرورهای جانشین وظیفه‌ی انتقال اطلاعات به مشتریان در کمترین زمان ممکن را بر عهده دارند. به عبارت دیگر در شبکه‌های توزیع محتوی یک ارتباط سرور/مشتری به دو ارتباط سرور جانشین/مشتری و سرور اصلی/سرور جانشین تبدیل شده است و همین مسئله به کاهش ازدحام و دسترس‌پذیری بالای داده‌ها منجر می‌شود. این مدل در (شکل ۱) نشان داده شده است.

از مزایای استفاده از شبکه‌های توزیع محتوی می‌توان به کاهش بار سرور اصلی، کاهش تاخیر در ارسال اطلاعات به کاربران، افزایش گذردهی، مقیاس‌پذیری بالا و مواجهه سریع با مشکلاتی همچون Flash Crowd[۱۴] (مانند واقعه‌ی یازده سپتامبر) [۸] اشاره نمود. در این شبکه-ها مشکلاتی که در شبکه‌های سنتی سرور/مشتری وجود داشت، از جمله خرابی و از کارافتادن سرور اصلی و در نتیجه قطع کامل سرویس و ازدحام و افزایش بار بر روی سرور اصلی با افزایش درخواست‌های کاربران، در حد قابل قبولی حل شده است.

تفاوت اصلی شبکه های توزیع محتوی با سیستم‌های کش[۱۵] در این است که اولا در کش تنها داده‌های پربازدید ذخیره می‌گردند، درحالیکه در شبکه‌های توزیع محتوی انتخاب اینکه چه داده‌ای بر روی سرورهای جانشین قرار گیرد بر عهده‌ی مدیر شبکه است، و ثانیا کش برای استفاده‌های محلی طراحی شده است و شبکه‌های توزیع محتوی به صورت توزیع شده در منطقه‌ی بزرگتری مورد استفاده قرار می گیرند. [۶]

هزینه استفاده از شبکه‌های توزیع محتوی تقریبا بالاست و پرداخت آن از سوی شرکت‌ها و سازمان‌های کوچک عموما امکان‌پذیر نیست. همچنین عموما برای گرفتن سرویس، کاربرانی که دورتر از سرور اصلی هستند هزینه‌ی بیشتری نسبت به کاربران نزدیکتر پرداخت می‌کنند. عواملی که در هزینه‌ی ارائه سرویس توسط شبکه‌های توزیع محتوی نقش دارند عبارتند از [۵]:

- عرض باند مصرفی
- حجم محتوای نگهداری شده در سرورهای جانشین
- تعداد سرورهای جانشین
- پایداری و امنیت شبکه

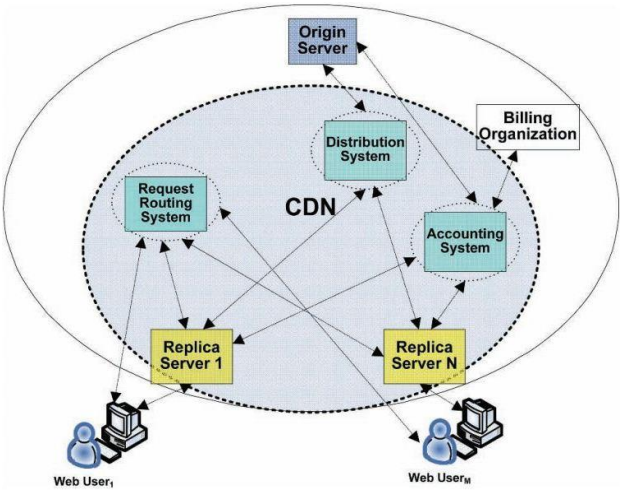


شکل ۱: مدل شبکه های توزیع محتوی [۱]

۲-۱- معماری شبکه های توزیع محتوی

به صورت کلی شبکه‌های توزیع محتوی از چهار زیر ساخت اصلی (شکل ۲) تشکیل شده‌اند که عبارتند از [۱]:

- تحویل محتوی[۱۶]: همان سرورهای جانشین یا سرورهای لبه-ای[۱۷] هستند. این سرورها در تلاشند تا بار را از روی سرور اصلی کم کنند و به جای آن به ارسال اطلاعات بپردازند.
- مسیریابی درخواست‌ها[۱۸]: این بخش مسئول هدایت کردن درخواست‌های کاربران به مناسب‌ترین سرور جانشین موجود است. معیارهای مناسب بودن یک سرور می‌تواند شامل نزدیکی سرور، قدرت پردازشی سرور و یا میزان بار در حال پردازش بر روی سرور باشد.
- توزیع[۱۹]: این بخش مسئول انتقال داده‌ها از سرور اصلی به سرورهای جانشین و نیز اطمینان از پایداری و به روز بودن داده ها با کمک بخش مسیریابی درخواست‌ها است.
- حسابرسی[۲۰]: هدف این بخش گزارش‌گیری، صدور صورت حساب و مدیریت شبکه‌ی توزیع محتوی است.



شکل ۲: زیرساخت های شبکه های توزیع محتوی [۱]

تمرکز اصلی در طراحی و ساخت شبکه‌های توزیع محتوی مبتنی بر نگهداری و مدیریت محتوی، توزیع محتوی در میان سرورهای جانشین، مدیریت کش، ارسال انواع داده‌ها به کاربران، تهیه پشتیبان و بازیابی اطلاعات پس از اتفاقات و مشکلات از پیش تعیین نشده، ارزیابی کارایی و کیفیت و در نهایت گزارش‌گیری و مدیریت ارسال داده‌ها به کاربران است.

معماری شبکه‌های توزیع محتوی را می‌توان از نقطه‌نظرهای گوناگونی طبقه بندی نمود. اجزای این شبکه می‌توانند به صورت همگن یا ناهمگن با یکدیگر همکاری و تعامل داشته باشند و ساختار آنها می‌تواند به صورت متمرکز[۲۱]، هرمی[۲۲] و یا کاملا غیر متمرکز[۲۳] باشد.

ساختار لایه‌ای شبکه‌های توزیع محتوی را می‌توان به صورت زیر عنوان کرد:

- لایه‌ی بنیادی[۲۴]: لایه‌ی زیرین شبکه که شامل تمامی منابع محاسباتی و سخت‌افزارهای شبکه مانند Cluster، File Server، Index Server و... می‌باشد. این منابع توسط لینک‌هایی با پهنای باند بالا به یکدیگر متصل شده‌اند.
- لایه‌ی ارتباطات و اتصالات[۲۵]: این لایه شامل تمامی پروتکل‌های اینترنت مورد استفاده در شبکه شامل FTP، TCP/UDP و... می‌باشد.
- لایه‌ی توزیع داده[۲۶]: این لایه شامل تمامی کاربردها و ویژگی‌های اصلی شبکه‌های توزیع محتوی همانند انتخاب سرور جانشین، مسیریابی درخواست‌ها و... می‌باشد.
- کاربران[۲۷]: لایه‌ی نهایی در معماری شبکه‌های توزیع محتوی.

۲-۲- راه حل‌های پیشین

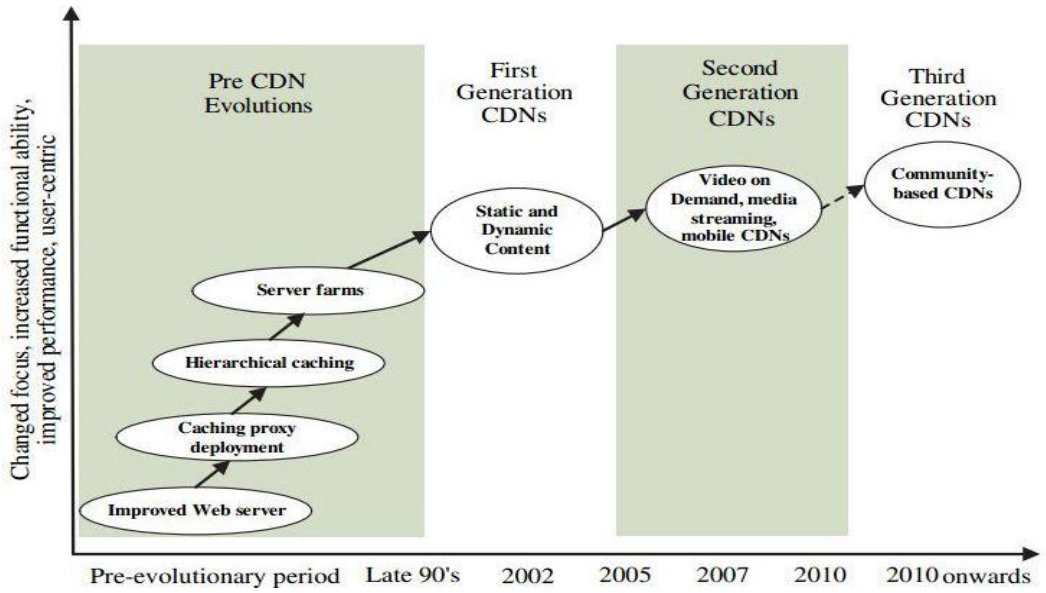
با بالا رفتن بار ترافیکی در شبکه‌ی اینترنت، سازمان‌ها و شرکت‌های بزرگ بر آن شدند تا راه‌حلی برای فائق آمدن بر این مشکلات ارائه دهند [۱] (شکل ۳ **Error! Reference source not found.**). اولین ایده این بود که با بهبود ظرفیت لینک‌ها و قدرت پردازشی سرورها این حجم بالای ترافیک را کنترل کنند ولی این ایده در عمل امکان پذیر و قابل گسترش نبود.

راه‌حل‌های بعدی استفاده از متدهای کش کردن و نیز مزرعه سرورها[۲۸] بود که هر کدام در عین حال که تا حدی مشکل را حل می‌نمایند معایب خاص خود را نیز به همراه دارند. همانگونه که گفته شد متدهای کش تنها داده‌هایی را نگهداری می‌کنند که پربازدید هستند و با افزایش تعداد کاربران و تنوع در درخواست‌های کاربران کارایی پایینی از خود نشان می‌دهند. مزرعه‌ی سرورها نیز با وجود این که بسیار کارا و قابل گسترش هستند به دلیل این که تمامی سرورها در کنار هم و نزدیک یکدیگر قرار دارند بهبود اندکی در کارایی کل شبکه، به خصوص در شرایط ازدحام، ایجاد می‌کند.

بنابراین شبکه‌های توزیع محتوی با بهره‌گیری از مزایای راه‌حل‌های گذشته به عنوان راه‌حلی بهینه معرفی شد. در مدل‌های اولیه‌ی شبکه‌های توزیع محتوی تمرکز اصلی بر روی انتقال داده‌های ایستا و پویا بود، در حالی که در مدل‌های بعدی به انتقال داده‌های رشته‌ای همچون صدا و تصویر بیشتر پرداخته شده است.

به طور کلی اهداف شبکه‌های توزیع محتوی را می‌توان در موارد زیر خلاصه کرد:

- مقیاس‌پذیری[۲۹]: شبکه در برابر تغییراتی همچون افزایش داده‌ها، کاربران و/یا تراکنشها بدون اعمال تغییرات زیاد در ساختار شبکه و کاهش کارایی. مقیاس‌پذیری باعث می‌شود که از افراط در تامین منابع و سرمایه‌گذاری غیرضروری برای پاسخگویی به نیازهای کاربران پرهیز شود.
- امنیت در نگهداری و انتقال اطلاعات. در غیر این صورت شبکه‌های توزیع محتوی هدف حملات و سرقت‌های متعددی قرار می‌گیرد که به علت گسترش جغرافیایی و معماری توزیع شده‌ی این سیستم، تشخیص این حملات بسیار دشوار است [۹].
- قابلیت اعتماد[۳۰]: این ویژگی به بازه‌ای از زمان که انتظار می‌رود در طول آن سرویس به درستی ارائه شود اطلاق می‌گردد. شبکه‌های توزیع محتوی با توزیع محتوی در مناطق جغرافیایی گوناگون و به کارگیری سازوکارهای قابل اعتماد برای توزیع بار بین سرورهای جانشین قابلیت اعتماد در کل سیستم را تامین می‌کنند.
- بهبود ارائه، سرعت و کارایی سرویس‌ها. کارایی در شبکه‌ها عموما بر اساس زمان پاسخگویی به کاربران و یا تاخیری که در سرویس‌دهی به کاربران مشاهده می‌شود سنجیده می‌شود. قابلیت اعتماد و کارایی شبکه‌های توزیع محتوی تحت تاثیر مکانیزم‌های مسیریابی، مکان توزیع محتوی، تکثیر داده‌ها و نیز استراتژی‌های ذخیره‌سازی محتوی می‌باشند.



شکل ۳: سیر تکامل شبکه های توزیع محتوی [۱]

۳- دسته بندی[۳۱] شبکه‌های توزیع محتوی

از چندین نقطه نظر می‌توان به بررسی شبکه‌های توزیع محتوی پرداخت که در ادامه به معرفی آنها پرداخته خواهد شد.

۳-۱- ساختار شبکه‌های توزیع محتوی[۳۲]

ساختار شبکه‌های توزیع محتوی بر اساس محتوی و سرویسی که ارائه می‌دهند با یکدیگر متفاوت است. در این ساختار چندین سرور جانشین وجود دارند که زیرساخت توزیع محتوی را بر عهده دارند، درخواست-های کاربران توسط مکانیزم‌ها و الگوریتم‌هایی به دست این سرورهای جانشین می‌رسند. تمامی اجزای شبکه نیز با استفاده از پروتکل‌های تعاملی متعددی با یکدیگر در ارتباطند مسائل گوناگونی در شکل‌گیری ساختار این شبکه‌ها دخیلند که این عوامل در ادامه ذکر می‌گردند.

۳-۱-۱- سازمان دهی شبکه‌های توزیع محتوی[۳۳]

دو رویکرد در سازمان‌دهی شبکه‌های توزیع محتوی وجود دارد [۱۰]. اولین رویکرد Overlay Approach است که در این رویکرد سرویس‌ها بر روی ساختار اینترنت موجود به کاربران ارائه می‌گردد. Overlay Network در حقیقت یک شبکه منطقی و مجازی است که برای ارائه سرویس‌هایی که در شبکه‌ی اینترنت کنونی وجود ندارند طراحی و پیاده‌سازی شده‌اند. در این گونه از شبکه‌ها المان‌های شبکه همانند سوئیچ‌ها و مسیریاب‌ها جز برقراری ارتباط و تضمین کیفیت سرویس هیچ نقش دیگری ندارند و سرویس‌ها توسط سرورهایی تحت عنوان سرورهای Application Specific ارائه می‌شوند. ارائه‌دهندگان بزرگ شبکه‌های توزیع محتوی همچون Akamai از این شیوه استفاده می-کنند [۱۷].

رویکرد دیگر Network Approach است. بدین صورت که المان-های شبکه برای سرویس‌دهی به کاربران دستکاری شده و با استفاده از یک سری قواعد و سیاست‌های از پیش تعیین شده کلیه‌ی عملیات توزیع محتوی را انجام می‌دهند. در حقیقت المان‌های شبکه خود شبکه‌ی توزیع محتوی را شامل می‌شوند و نقشی اساسی در انتقال محتوی به کاربران ایفا می‌کنند.

۳-۱-۲- سرورها

شبکه‌های توزیع محتوی شامل دو نوع سرور هستند. سرورهای اصلی که مسئولیت نگهداری نسخه‌ی اصلی داده‌ها را بر عهده دارند. این داده-ها توسط سازمان‌ها و تولیدکنندگان محتوی مرتبا به روز می‌شوند. نوع دیگر سرورها، سرورهای جانشین هستند که مسئولیت رساندن داده‌ها به دست کاربران را برعهده دارند. سرورهای جانشین شامل Media Server ها برای نگهداری داده‌های رشته‌ای و چندرسانه‌ای همانند صوت و تصویر، Web Server ها برای نگهداری داده‌های ایستا و پویا همچون داده‌های مبتنی بر وب و Cache Server ها برای نگهداری یک نسخه از پر بازدیدترین داده‌ها می‌باشند.

۳-۱-۳- ارتباطات

روابط متعددی بین المان‌های شبکه‌های توزیع محتوی برقرار است. نمونه‌هایی از این روابط می‌تواند شامل ارتباط میان کاربران، سرورهای جانشین و سرور اصلی (در Overlay Approach) و یا ارتباط بین کاربران، المان‌های شبکه و سرورهای کش (در Network Approach) باشد. تمامی این ارتباطات تحت پوشش پروتکل‌هایی استاندارد شده‌اند.

۳-۱-۴- پروتکل‌های تعاملی

همانگونه که گفته شد پروتکل‌ها در راستای استانداردسازی ارتباطات در شبکه‌های توزیع محتوی تعریف و ارائه شده‌اند. در ارتباطات میان المان-های شبکه و نیز سرورهای کش پروتکل‌های متعددی همچون

NECP[۳۴]، ICP[۳۵]، CARP[۳۶] و HTCP[۳۷] طراحی شده و مورد استفاده قرار گرفته‌اند.

۳-۱-۵- نوع محتوی و سرویس ارائه شده

همانگونه که در مقدمه‌ی این مقاله عنوان شد در شبکه‌های توزیع محتوی داده‌ها انواع متفاوتی دارند:

- داده‌های ایستا که نرخ تغییر پایینی دارند،
- داده‌های پویا که به ازای هر کاربر در حال تغییرند و
- داده‌های رشته‌ای که شامل صوت و تصویر می‌باشند و دو نوع زنده و On-Demand دارند.

شبکه‌های توزیع محتوی سرویس‌هایی همچون انتقال فایل، دایرکتوری و... نیز ارائه می‌دهند. در حقیقت شبکه‌های توزیع محتوی به مشتریان خود اجازه می‌دهند تا از منابعشان در راستای ارائه سرویس‌های ارزش افزوده[۳۸] به کاربران خود استفاده نمایند.

۳-۲- مدیریت و توزیع محتوی

در این نقطه نظر به بررسی ارسال بهینه داده‌ها به کاربران و افزایش کارایی کل سیستم پرداخته می‌شود.

۳-۲-۱- جایابی سرورهای جانشین[۳۹]

با جایابی درست و بهینه‌ی سرورهای جانشین در مکان‌هایی نزدیک به کاربران و به تعداد مناسب با در نظر گرفتن میزان ترافیک منطقه می‌توان زمان ارسال اطلاعات به کاربران را کمینه نمود و در مصرف و هزینه‌ی عرض باند صرفه جویی کرد [۱۱].

استراتژی‌های متعددی در جایابی سرورهای جانشین وجود دارد که از جمله‌ی آنها می‌توان به Greedy Methods، Hot Spot، Tree-Based Replica Placement و Center Placement Problem اشاره کرد.

به طور کلی دو رویکرد در جایابی سرورهای جانشین وجود دارد: رویکرد Single ISP و Multi ISP [۶]. در رویکرد اول یک ISP[۴۰] و تعدادی سرور جانشین در شبکه (در کشورهای مختلف) توزیع شده‌اند (تعداد این سرورها معمولا به ۴۰ عدد می‌رسد) که مسئولیت انتقال اطلاعات به کاربران زیر نظر ISP مرکزی را بر عهده دارند [۱۲]. یکی از معایب این روش این است که سرورهای جانشین از کاربران فاصله دارند، زیرا در هر کشور و منطقه یک و یا حداکثر دو سرور جانشین وجود دارد. از مزایای این رویکرد می‌توان به Hit Rate بالای هر کدام از سرورهای جانشین اشاره نمود.

در رویکرد دوم چندین ISP که هر کدام چندین سرور جانشین را مدیریت می‌کنند در شبکه پراکنده شده‌اند. در این روش ISP ها با یکدیگر در ارتباطند. همچنین سرورهای جانشین به کاربران نزدیکتر شده‌اند، در نتیجه ارسال اطلاعات سریعتر انجام می‌گیرد، ولی میزان Hit Rate در هر سرور جانشین کم شده و به طور کلی کارایی سیستم پایین می‌آید [۵]. رویکرد اول برای سایت‌هایی با بازدید کم تا متوسط و رویکرد دوم برای سایت‌های با ترافیک متوسط به بالا و پربازدید مناسب است [۱۲].

۳-۲-۲- انتخاب و ارسال محتوی[۴۱]

دو رویکرد اصلی در انتخاب و ارسال داده‌ها به کاربران وجود دارد. در یک رویکرد که با نام Full Site شناخته می‌شود کل محتوی بر روی سرورهای جانشین قرار می‌گیرد. تنها مزیت این راه‌حل سادگی آن است، در حالی که از چالش‌های آن می‌توان به کمبود فضای ذخیره‌سازی با توجه به افزایش بی رویه داده‌ها و نیز دشواری به روز کردن تمامی این داده‌ها اشاره کرد.

رویکرد دوم که Partial Site گفته می‌شود بدین صورت است که تنها المان‌ها و اشیا محاط شده[۴۲] در داده‌ها در سرورهای جانشین نگهداری می‌شوند و اطلاعات پایه و HTML اصلی از سرور اصلی گرفته می‌شود. از آنجایی که اشیا محاط شده به ندرت تغییر می‌کنند نرخ به روز رسانی داده‌ها در سرورهای جانشین کاهش یافته و این رویکرد می‌تواند کارایی بیشتر نسبت به رویکرد قبلی داشته باشد.

۳-۲-۳- واگذاری محتوی[۴۳]

سه راه وجود دارد تا اطلاعات مورد درخواست کاربران از سرور اصلی به سرورهای جانشین منتقل گردد [۵]:

- Cooperative Push Based: در این راهکار اطلاعات از قبل بر روی سرورهای جانشین منتقل شده‌اند و در صورت درخواست آن، کاربر با متدهای مختلفی همچون DNS Redirection و یا URL Rewriting به نزدیک‌ترین سرور جانشین متصل می‌گردد. در این روش بین سرورهای جانشین همکاری و تعامل وجود دارد. این روش هنوز به صورت عمومی عملی نشده و تنها به صورت تئوری ارائه شده است [۱۳].
- Non Cooperative Pull Based: در این راهکار داده‌ها در سرور اصلی وجود دارند و با درخواست کاربران بر روی سرورهای جانشین بارگذاری می‌شوند. بنابراین در این روش درخواست کاربر ابتدا به نزدیکترین سرور جانشین فرستاده می‌شود و در صورتیکه داده در آن سرور وجود نداشت اطلاعات از سرور اصلی گرفته می‌شود.
- Cooperative Pull Based: این روش همانند روش قبل است، با این تفاوت که سرورهای جانشین با یکدیگر در تعامل و همکاری هستند و در صورتیکه داده در اولین سرور جانشین تعیین شده وجود نداشت درخواست به سرورهای جانشین مجاور منتقل می‌شود تا محتوای مورد نیاز یافت شود. در این روش محتوی و سرورها اندیس‌گذاری[۴۴] می‌شوند

تا بتوان سروری را که داده مورد نظر بر روی آن قرار دارد به راحتی و در کمترین زمان ممکن یافت.

۳-۳- مسیریابی درخواست‌ها

انتخاب بهترین سرور جانشین و انتقال درخواست کاربران به آن یکی از مهمترین مسائل مطرح شده در شبکه‌های توزیع محتوی می‌باشد. معیارهای متعددی در تعیین بهترین سرور جانشین وجود دارد، عواملی همچون فاصله، سرعت پردازش و انتقال اطلاعات، قابلیت اعتماد و هزینه ارسال از این دست معیارها هستند [۱۴]. روش‌های متعددی برای مسیریابی درخواست‌ها وجود دارد که دو روش متداول در این حوزه عبارتند از DNS Redirection و URL Rewriting.

۳-۳-۱- DNS Redirection

در این روش از DNS برای ترجمه‌ی بین اسم و آدرس IP سرورها استفاده می‌شود و بدین ترتیب کاربران به سرورهای جانشین مناسب متصل می‌گردند. از معایب این روش می‌توان به تاخیر بالا به دلیل استفاده از DNS اشاره نمود. در رویکرد Full site از این روش استفاده می‌شود [۶].

۳-۳-۲- URL Rewriting

در این روش سرور اصلی خود درخواستهای کاربران را به سرورهای جانشین منتقل می‌کند. در حقیقت پس از دریافت درخواست از سوی کاربر URL آن دسته از اطلاعاتی که در سرورهای جانشین ذخیره شده است دوباره‌نویسی شده و درخواست به جانب آن سرورها فرستاده می‌شود. پس از دریافت آن داده ها و ترکیب آنها با اطلاعات ذخیره شده در خود سرور اصلی، کل داده‌ی مورد درخواست به کاربر مربوطه ارسال می‌گردد. از این روش در رویکرد Partial Site استفاده می‌شود [۶].

۳-۴- ارزیابی کارایی

از آنجایی که شبکه‌های توزیع محتوی دو دسته مشتری دارند، معیارهای کارایی و بهره‌وری از دیدگاه هر دسته کاملا متفاوت است [۱۲]. دسته‌ی اول مشتریان[۴۵] شبکه‌های توزیع محتوی هستند که اطلاعات خود را بر روی سرورهای اصلی و جانشین بارگذاری می کنند. معیارهای ارزیابی کارایی شبکه‌های توزیع محتوی از دیدگاه مشتریان عبارتند از:

- Hit Ratio: این نسبت در حقیقت رابطه‌ی بین میزان داده‌هایی که با موفقیت در سرورهای جانشین یافت می‌شوند نسبت به کل داده‌های درخواستی از سوی کاربران را نشان می‌دهد. هر چه این نسبت بزرگتر باشد کارایی شبکه بهتر و مورد قبول‌تر خواهد بود.
- عرض باند: هر چه عرض باند از سرورهای جانشین و منتهی به سرور اصلی کاهش یابد آن شبکه مناسبتر خواهد بود، زیرا هزینه‌ی کمتری بابت پهنای باند پرداخته می‌شود.
- تاخیر: هر چه تاخیر در ارسال اطلاعات کاهش یابد عرض باند مصرفی نیز کاهش می‌یابد و شبکه هزینه‌ی کمتر و بازدهی بیشتری خواهد داشت.
- بهره‌وری سرورهای جانشین: میزان زمانی که سرورهای جانشین در حال کار می باشند برابر با بهره‌وری آنهاست. این پارامتر نباید از یک میزانی کمتر (که منجر به بیکاری سرور و اتلاف منابع آن می‌شود) و از یک میزانی بیشتر (که منجر به سربار بالا و کند شدن سرویس دهی می‌شود) باشد.
- قابلیت اعتماد: هر چه میزان از بین رفتن بسته‌ها در حین ارسال کمتر باشد دسترس پذیری محتوی افزایش می‌یابد.

دسته‌ی دوم از کسانی که با شبکه‌های توزیع محتوی سر و کار دارند کاربران آن هستند، کسانی که لایه‌ی آخر از این شبکه‌ها را شامل می‌شوند و برای دستیابی به اطلاعات مورد نیاز خود به این شبکه‌ها درخواست می‌دهند. معیارهای ارزیابی کارایی یک شبکه توزیع محتوی از دید کاربرانش شامل موارد زیر می‌گردد:

- هزینه: همانگونه که بیان شد هزینه‌ی استفاده از شبکه‌های توزیع محتوی زیاد است، ولی با توجه به کاهش عرض باند مصرفی، این هزینه‌ها در کل کاهش می‌یابند. به طور معمول هزینه بر اساس میزان عرض باند مصرفی توسط کاربر (به ازای هر بایت مصرفی) از او گرفته می‌شود.
- کارایی: اندازه‌گیری کارایی معیارهای متعددی دارد، از جمله: میزان فضای ذخیره‌سازی، هزینه‌ی عرض باند، گذردهی، تاخیر ارسال، تاخیر در پیدا کردن مناسبترین سرور جانشین و... .
- دسترس‌پذیری: این معیار با کارایی ارتباط نزدیک دارد. شبکه‌های توزیع محتوی میزان خطاهایی که کاربران به دلایل متعددی همچون DNS Lookup Fail و یا Connection Time Out با آنها مواجه می‌شوند را کاهش داده و دسترس‌پذیری محتوی را افزایش می‌دهند.

۴- شبکه‌های توزیع محتوای موبایل

امروزه با ظهور شبکه های موبایل نیاز به ارائه‌ی سرویس‌های شبکه‌های توزیع محتوی بر روی این زیرساخت نیز احساس می‌گردد. در حقیقت از آنجایی که امروزه نیاز به دسترسی به محتوی در کمترین زمان و بدون محدود شدن به مکان رو به افزایش است راه‌حل‌هایی که این نیاز را برای کاربران در حال جابه‌جایی فراهم می‌کنند بسیار مورد استقبال قرار گرفته‌اند [۱۵، ۱۶].

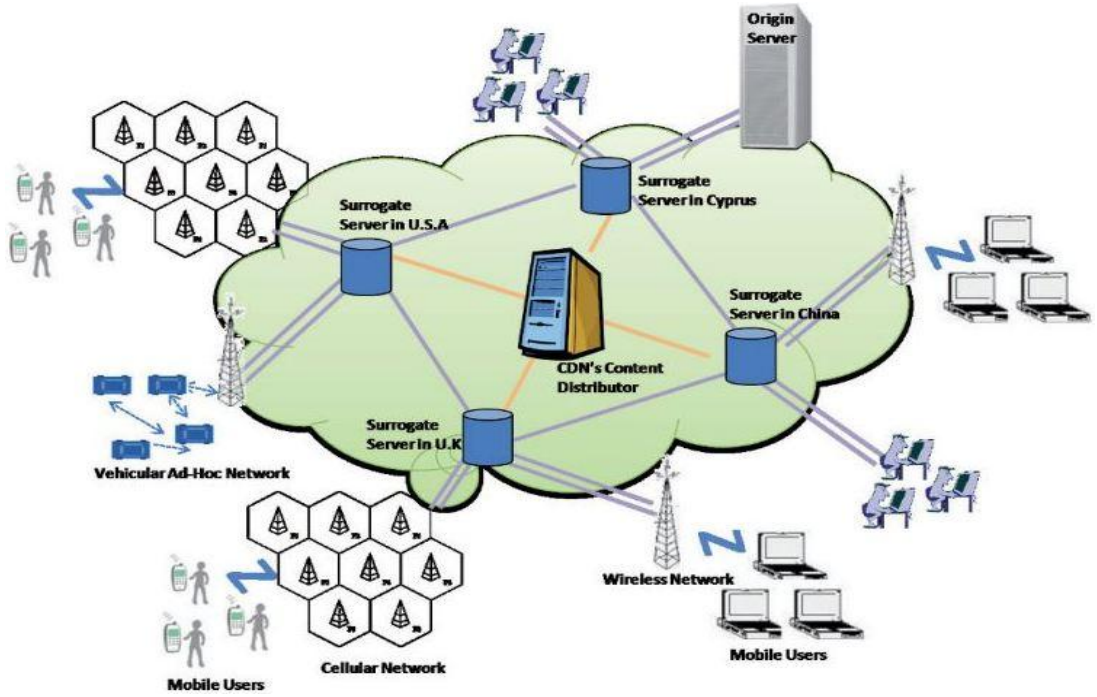
مفهوم شبکه‌های موبایل در حقیقت دسترسی بیسیم به داده‌های دیجیتال از طریق دستگاه‌های موبایل است. این دستگاه‌ها می‌تواند شامل گوشی‌های تلفن همراه، کامپیوترهای جیبی و یا تبلت‌ها باشند. در عین حال که معماری کنونی شبکه‌های توزیع محتوی تا حد قابل قبولی مناسب و قابل پیاده‌سازی بر روی شبکه‌های موبایل است، با این

وجود این معماری به دلیل عدم ارائه راه‌حلی برای پوشش برخی از ویژگی‌های این شبکه‌ها، همچون جابه‌جایی کاربران در شبکه، کافی نمی‌باشد و بایستی تغییراتی در آن اعمال گردد. علاوه بر این چالش‌های متعددی در این راه وجود دارد که به طور کلی می‌تواند آنها را به صورت زیر دسته‌بندی نمود [۱]:

- قطعی مداوم اتصال به شبکه: این قطعی به دلایل متعددی همچون (۱) مدت زمان کوتاه اتصال به شبکه، که از ویژگی‌های ذاتی ارتباطات در شبکه‌های موبایل است، (۲) وجود موانع متعدد بر سر راه ارتباط همچون ساختمان‌های بلند، درخت‌ها و غیره که باعث قطعی ارتباط می‌شوند و همچنین (۳) گاها خارج شدن کاربر از محدوده تحت پوشش شبکه می‌تواند اتفاق بیافتد.
- چندپارگی[۴۶] اطلاعات: به دلیل قطعی مداوم ارتباطات، احتمال چند تکه شدن اطلاعات در حال انتقال بسیار زیاد است که برای بازسازی آنها بایستی تدابیر متعددی اندیشید.
- محدودیت‌های دستگاه‌های موبایل: دستگاه‌های موبایل به دلیل ابعاد کوچک و باتری‌های کم توانی که دارند محدودیت های زیادی در میزان توان، حافظه و قدرت پردازش دارند.

به دلیل محدودیت‌هایی که اشاره شد برای طراحی شبکه‌های توزیع محتوی بر روی زیرساخت موبایل لازم است که وضعیت هر کدام از کاربران در شبکه مشخص باشد تا از میزان اتلاف ترافیک و عرض باند جلوگیری شود. به عنوان مثال در صورتیکه کاربری در حین بارگیری اطلاعاتی به دلیل کمبود باتری خاموش گردد و یا ارتباطش با شبکه قطع شود، شبکه باید از وضعیت کنونی این کاربر در اسرع وقت آگاهی یافته و با هدف صرفه‌جویی در عرض باند مصرفی از ادامه ارسال اطلاعات به آن کاربر جلوگیری کند.

اجزای سازنده‌ی شبکه‌های توزیع محتوای موبایل همانگونه که در (شکل **Error! Reference source not found.**۴) نشان داده شده‌اند عبارتند از: سرور اصلی، مجموعه‌ای از سرورهای جانشین که در منطقه پراکنده شده‌اند، توزیع کننده‌ی محتوی و المان‌های دسترسی موبایل به شبکه همچون ایستگاه‌های اصلی[۴۷] و نقاط دسترسی[۴۸].



شکل ۴: شبکه توزیع محتوای موبایل [۱]

به طور کلی شبکه‌های موبایل شامل دو بخش هستند: (۱) بخش سیمی شبکه که مسئول فراهم ساختن زیرساخت لازم برای ارتباطات میان سرور اصلی و سرورهای جانشین و میان سرورهای جانشین و المان‌های شبکه (سوئیچ‌ها، مسیریاب‌ها، ایستگاه‌های اصلی و نقاط دسترسی) است و (۲) بخش بیسیم که مسئول ارتباطات بیسیم بین کاربران و المان‌های شبکه است. بدین ترتیب ارتباط مشتری/سرور با سه ارتباط زیر جایگزین خواهد شد:

- ارتباط میان کاربران و لبه‌های شبکه (نقاط دسترسی و ایستگاه‌های اصلی)
- ارتباط میان لبه‌های شبکه و سرورهای جانشین
- ارتباط میان سرورهای جانشین و سرور اصلی

از آنجایی که سرورهای جانشین باید بیشترین پوشش را در سطح شبکه ایجاد کنند بهتر است که در کنار ایستگاه‌های اصلی یا نقاط دسترسی بنا شوند. بنابراین تعداد سرورهای جانشین در شبکه‌های توزیع محتوای موبایل بیشتر از تعداد سرورهای جانشین در شبکه‌های توزیع محتوای معمولی است.

علاوه بر این در صورت ارائه‌ی سرویس بارگذاری برای کاربران بر روی شبکه بایستی معماری سرورهای جانشین تغییر کند، بدین صورت

که فضای حافظه آنها به دو بخش تقسیم گردد که یک بخش اطلاعاتی که از سرور اصلی گرفته می‌شود را در خود جای دهد و بخش دیگر به محتوایی که کاربران بر روی شبکه بارگذاری می‌کنند اختصاص یابد. همچنین به دلیل آگاهی نسبی از موقعیت جغرافیایی کاربران امکان ارائه سرویس‌های مبتنی بر مکان[۴۹] نیز توسط شبکه‌های توزیع محتوای موبایل وجود دارد.

راهکار انتخاب شده برای واگذاری محتوی در شبکه‌های توزیع محتوای موبایل راهکار Cooperative Push Based است که بر اساس تحقیقات انجام شده در مرجع [۱۳] روی هم رفته بازدهی بیشتری نسبت به دیگر راهکارها از خود نشان می‌دهد.

با این تفاصیل می‌توان المان‌های معماری شبکه توزیع محتوای موبایل را بدین صورت تعریف نمود:

- مجموعه‌ای از سرورهای جانشین که در سطح دنیا و در نزدیکی ایستگاه های اصلی پراکنده شده‌اند
- زیرساخت شبکه (سیمی و بیسیم) جهت مسیریابی و هدایت درخواست‌ها به مناسب‌ترین سرور جانشین
- مکانیزم‌های مدیریتی جهت کنترل پارامترهای شبکه مانند عرض باند قابل استفاده، تاخیر و...
- مکانیزم‌های مدیریتی جهت انتخاب مناسب‌ترین داده‌ها برای نگهداری بر روی سرورهای جانشین
- مکانیزم‌های مدیریت مکانی محتوی برای انتخاب بهترین مکان برای نگهداری هر داده
- مکانیزم‌های حسابرسی جهت جمع‌آوری گزارشات و اطلاعات حساب کاربران و ارسال آنها به سرور اصلی و شرکت‌های تولید کننده محتوی

تمامی ویژگی‌ها و پارامترهای عنوان شده تاکنون در (جدول ۱) خلاصه شده اند.

۵- شبکه‌های نسل سوم[۵۰] تلفن همراه

شبکه‌های نسل سوم تلفن همراه با ساختار مبتنی بر بسته[۵۱] پس از نسلهای اول و دوم تلفن همراه پا به عرصه‌ی ظهور گذاشتند. این شبکه‌ها در مقایسه با تکنولوژی‌های قبلی از سرعت انتقال اطلاعات بیشتری برخوردار هستند (در حدود ۱۰۰ مگابیت در ثانیه). از دیگر مزیت‌های این شبکه‌ها می‌توان به امنیت بسیار بالای ارتباطات و انتقال اطلاعات اشاره نمود. ویژگی‌ای که این شبکه‌ها را از شبکه‌های نسل قبلی خود متمایز می‌سازد امکان انتقال اطلاعات چندرسانه‌ای است. این امکان در شبکه‌های قبلی که تنها برای انتقال صوت و داده طراحی شده بودند وجود نداشت، از این رو سرویس‌هایی همچون Video on Demand، Mobile TV، Video Conferencing، Tele-Medicine و Location Based Services بر بستر شبکه‌های نسل سوم ارائه شدند [۳].

با توجه به ویژگی‌های بالقوه‌ای که این شبکه‌ها دارند بستر بسیار مناسبی جهت ارائه‌ی سرویس‌های توزیع محتوی به کاربران می‌باشند.

۶- جایابی محتوی در شبکه‌های توزیع محتوای موبایل

همانگونه که ذکر شد به دلیل تحرک کاربران در شبکه‌های موبایل مساله‌ی جایابی محتوای ذخیره شده در سرورهای جانشین از بزرگ‌ترین چالش در ارائه‌ی راهکاری برای پیاده‌سازی سرویس‌های شبکه‌های توزیع محتوی بر بستر شبکه‌های موبایل به شمار می‌رود. بایستی بتوان محتوای مورد نیاز کاربران را همواره در نزدیکی آنها نگهداری نمود تا در اسرع وقت به درخواست‌های کاربران رسیدگی شده و محتوای مذکور به آنها ارسال گردد، که در ادامه‌ی این تحقیق به ارائه‌ی راهکاری بهینه در جهت برآوردن این نیاز پرداخته می‌شود.

شبکه‌ی مورد بررسی ترکیبی از نسل سوم شبکه‌های تلفن همراه و شبکه‌ی توزیع محتوی خواهد بود که به دو دسته کاربران ثابت و متحرک سرویس دهی می‌کند. فرض بر این است که سرورهای جانشین می‌توانند در هر مکانی، در نزدیکی سوئیچ های مرکزی و یا در کنار BTS[۵۲]ها، بنا شوند. محیط مورد بررسی فضای شهری و بدون محدودیت در پهنای باند سرورها، و در نتیجه بدون ازدحام می‌باشد، بنابراین تنها معیار در انتخاب بهترین سرور جانشین نزدیکی آن به کاربر می‌باشد. علاوه بر این یک سیستم هدایت درخواست وجود دارد که به هر کاربر متقاضی نزدیکترین سرور جانشینی که محتوای مورد درخواست را دارا می‌باشد معرفی می‌کند.

تعدادی از کاربران در این شبکه در حال جابجایی هستند. هدف از انجام این پروژه آن است که با کمترین هزینه برای کاربری که جابجا شده است یک نسخه‌ی المثنی از محتوای مورد نیازش تولید شود. بدین صورت با تحرک کاربران نسخه‌های المثنی محتوی نیز جابجا شده و در نزدیکی کاربران نگه داشته می‌شوند.

در محاسبه‌ی هزینه‌ی کل مدیریت کاربران در حال تحرک در شبکه‌های توزیع محتوای موبایل دو نوع هزینه وجود دارد. یکی هزینه‌ی تولید نسخه‌ی المثنی برای کاربران و دیگری هزینه‌ی پوشش و سرویس‌دهی کاربران متحرک توسط سرور جانشین است. در هزینه‌ی تولید نسخه‌ی المثنی هزینه‌ی نگهداری نسخه‌ها و نیز هزینه‌ی انتقال اطلاعات از سرور اصلی به سرور جانشین وجود دارد. در هزینه‌ی پوشش و سرویس‌دهی کاربران نیز هزینه‌ی انتخاب نزدیکترین سرور جانشین که محتوای مورد نیاز را دارا می‌باشد و نیز هزینه‌ی ارتباط کاربران با این سرور جانشین انتخاب شده نهفته است. با کمینه کردن هزینه‌ی کل مدیریت کاربران در حال تحرک می‌توان راهکار کارا و مورد قبولی برای ارائه‌ی سرویس توزیع محتوای موبایل ارائه داد.

۷ - نتیجه‌گیری

در این تحقیق ابتدا به بررسی شبکه‌های توزیع محتوی بر بستر اینترنت و معماری و ساختار آنها پرداخته شد. منظور از ساختار جزئیاتی در باب اجزای سازنده، پروتکل‌ها و استراتژی‌های مورد استفاده در این شبکه‌ها می‌باشد. در ادامه به معرفی شبکه‌های توزیع محتوی بر بستر موبایل پرداخته شد و چالش‌هایی که در این زمینه وجود دارند مشخص گردید.

ویژگی ها	شبکه های توزیع محتوای معمولی	شبکه های توزیع محتوای موبایل
نوع محتوی	ایستا، پویا، رشته ای	ایستا، پویا، رشته ای
مکان کاربران	ثابت	متحرک
مکان سرورهای جانشین	ثابت	ثابت
توپولوژی سرورهای جانشین	در نزدیکی ارائه دهندگان سرویس اینترنت	در نزدیکی ایستگاه های اصلی
هزینه‌ی سرورهای جانشین	متوسط	زیاد
سرویس ها	سرویسهای کاربردی و چندرسانه‌ای	سرویس های کاربردی مبتنی بر مکان و چند رسانه‌ای
مکانیزم واگذاری محتوی	Cooperative/Uncooperative Pull-Based Scheme	Cooperative Push-Based Scheme

جدول ۱: مقایسه شبکه های توزیع مجتوای معمولی با شبکه های توزیع محتوای موبایل

در انتها نیز پس از معرفی اجمالی شبکه‌های نسل سوم تلفن همراه و ذکر ویژگی‌های آنها، به معرفی ایده‌ای برای جایابی محتوی در شبکه-های توزیع محتوای موبایل پرداخته شد.

همانگونه که عنوان شد می‌توان با کمینه نمودن هزینه‌ی تولید نسخه‌های المثنی از محتوی کارایی ارائه‌ی سرویس توزیع محتوی بر روی شبکه‌های موبایل را افزایش داد.

مراجع

[1] Buyaa, R., Pathan, M., *Content Delivery Networks*, Vakali, A. (Eds.), Vol. 9, Springer-Verlag Berlin Heidelberg, 2008.

[2] Held, G., *A Practical Guide to Content Delivery Networks*, Taylor and Francis Group, New York, USA, 2006.

[3] Bannister, J., Mather, P., Coope, S., *Convergence Technologies for 3G Networks*, Chichester: John Wiley & Sons, 2004.

[4] Pathan, M., Buyya, R., *"A Taxonomy and Survey of Content Delivery Networks"*

[5] Pallis, G., Vakali, A., *"Insight and Perspectives for Content Delivery Networks"*, Communications of the ACM, Vol. 49, No. 1, January 2006.

[6] Vakali, A., Pallis, G., *"Content Delivery Networks: Status and Trends"*, IEEE Internet Computing, November/December 2003.

[7] Rabinovich, M., Spatscheck, O., *"Web Caching and Replication"*, Addison Wesley, USA, 2002.

[8] Arlitt, M., Jin, T., *"A Workload Characterization Study of 1998 World Cup Website"*, IEEE Network, 2000.

[9] Brussee, R., et al., *"Content Distribution Network State of the Art"*, Telematica Institut, 2001.

[10] Lazar, I., Terrill, W., *"Exploring Content Delivery Networking"*, IT Pro, August 2001.

[11] Sivasubramanian, S., et al., *"Analysis of Caching and Replication Strategies for Web Applications"*, IEEE Internet Computing, 2007.

[12] Douglis, F., Kaashoek, M., *"Scalable Internet Services"*, IEEE Internet Computing, 2001.

[13] Chen, Y. et al., *"Efficient and Adaptive Web Replication using Content Clustering"*, IEEE Journal, Vol. 21, No. Y, 2003.

[14] Dilley, J. et al., *"Globally Distributed Content Delivery"*, IEEE Internet Computing, September/October 2002.

[15] Aioffi, W.M., et al., *"Dynamic Content Distribution for Mobile Enterprise Networks"*, IEEE Journal, 2005.

[16] Wu, T., Dixit, S., *"The Content Driven Mobile Internet"*, Wireless Personal Communications: An International Journal, 2003.

[17] Akamai Technologies, Inc., www.akamai.com, 2012.

[1] Multimedia
[2] Quality of Service (QoS)
[3] Multicast
[4] Availability
[5] Point of Presence
[6] Encoded Data
[7] Meta Data
[8] Static
[9] Dynamic
[10] Stream
[11] Origin Server

[12] CDN's Content Distributor
[13] Surrogate Server
[14] Flash Crowd به افزایش ۸۰٪ در تعداد مخاطبین ۱۰٪ از محتوی گفته می‌شود.
[15] Cache
[16] Content Delivery
[17] Edge Server
[18] Request Routing
[19] Distribution
[20] Accounting
[21] Centralized
[22] Hierarchical
[23] Decentralized
[24] Base Layer

[25] Communication Layer
[26] Distribution Layer
[27] Users
[28] Server Farm
[29] Scalability
[30] Reliability
[31] Taxonomy
[32] CDN Composition
[33] CDN Organization
[34] Network Element Control Protocol
[35] Internet Cache Protocol
[36] Cache Array Routing Protocol
[37] HyperText Element Control Protocol
[38] Value Added Service (VAS)
[39] Surrogate Placement
[40] Internet Service Provider (ISP)
[41] Content Selection and Delivery
[42] Embedded Objects
[43] Content Outsourcing
[44] Indexing
[45] Client
[46] Fragmentation
[47] Base Station
[48] Access Point
[49] Geo-Location Service
[50] Third Generation (3G)
[51] Packet-Based
[52] Base Transciever Station

بررسی شبکه‌های بدنی بی‌سیم[1] با تمرکز بر کاربرد آن در پزشکی[2]

منصور حسینی[1]، محمود فتحی[2]

[1]دانشجوی کارشناسی ارشد
Mansoor_hoseini@comp.iust.ac.ir
[2]استاد راهنما
mahfathy@iust.ac.ir

چکیده

در سال‌های اخیر شبکه‌های بدنی بی‌سیم (WBAN) به عنوان یک تکنولوژی جدید برای کاربردهای بهداشت و درمان[3] مطرح شده است.انتظار می‌رود WBAN انقلابی در نظارت‌های سلامتی انسان‌ها بوجود آورد.اما این تکنولوژی در مراحل اولیهٔ توسعهٔ خود قرار دارد.WBAN به منظور برقراری ارتباط بی‌سیمِ دستگاه‌های شبکه (حسگرها[4]) روی بدن[5]،درون بدن[6] یا نزدیک بدن انسان استفاده می‌شود(اما به انسان محدود نمی‌شود).این مقاله خلاصه‌ای از شبکه های بدنی بی‌سیم را ارائه می‌دهدکه شامل کاربردها،معماری،امنیت و چالش‌های WBAN می‌باشد.علاوه بر آن استاندارد 802.15.6 که برای شبکه‌های WBAN به کار می‌رود ،مرور و دو قسمت اصلی آن شامل لایه فیزیکی،لایه کنترل دسترسی رسانه[7](MAC) مورد بحث قرار می‌گیرد.در پایان موضوعاتی که باید برروی آن مطالعات و تحقیقات بیشتری صورت پذیرد،مطرح شده است.

کلمات کلیدی

WBAN،شبکه‌های بدنی بی‌سیم، سیستم‌های بهداشت و درمان، حسگر، معماری، امنیت، 802.15.6 ، توپولوژی

۱ – مقدمه

پیشرفت‌های جدید در زمینه‌های مدارهای مجتمع،ارتباطات بی-سیم،تکنولوژی‌های نیمه هادی و علم کوچک سازی[8] باعث رشد استفادهٔ شبکه حسگر در کاربردهای وسیعی از جمله پزشکی و سازمان بهداشت و درمان شده است [1,2].از طرفی دیگر افزایش بیماری‌ها و هزینه‌های درمان ناشی از آن سبب پیدایش تکنیک‌هایی برای حل این مشکلات شده است.یکی از این تکنیک‌ها به کار گیری شبکه‌های بی-سیم بدنی (WBAN) می‌باشد.WBAN شامل چندین حسگر کوچک،قابل حمل و اتوماتیک می‌باشد که روی بدن یا درون بدن نصب می‌شود.این حسگرها علائم حیاتی بدن انسان مثل ضربان قلب و فشار خون را پایش[9] کرده و به یک هماهنگ کننده[10] می‌فرستند.این علائم ثبت شده در هماهنگ کننده می‌تواند برای اهداف پزشکی مورد استفاده قرار بگیرد [3].ابداع WBAN برای پایش علائم حیاتی و دیگر کاربردها،انعطاف پذیری و صرفه جویی در هزینه را هم برای بیماران و هم برای سازمان بهداشت و درمان فراهم کرده است.WBAN که به منظور اهداف پزشکی مورد استفاده قرار می‌گیرند دو مزیت مهم در مقایسه با سیستم‌های پایش قدیمی دارند .اولین مزیت به دلیل ماهیت بی‌سیم بودن شبکه‌های WBAN می‌باشد که باعث راحتی هرچه بیشتر بیمار می‌شود.به طوری که بیمار می‌تواند تمام فعالیت‌های روزانهٔ خود را (با وجود حسگرها) بدون هیچ مزاحمتی انجام دهد.دومین مزیت، سهولت پایش کردن بیمار (مشاهدهٔ علائم فیزیولوژی بدن بیمار) به دلیل مستقل بودن از موقعیت بیمار است.به این معنی که یک فرد بیمار چه در محیط خانه چه در محل کار می تواند به طور دائم مورد پایش قرار گیرد و نتیجهٔ این عمل، پایش بیمار برای دراز مدت است.اطلاعات جمع‌آوری شده از بیمار طی زمان طولانی نقش بسزایی در تشخیص و درمان بسیاری از بیماری‌ها دارد[4].

یکی از مشخصات اصلی WBAN که آن را از دیگر تکنولوژی‌های موجود همچون شبکه‌های حسگر بی‌سیم[11](WSN) متمایز کرده است، محدودهٔ کمی است که توسط این شبکه پوشش داده می‌شود.این محدوده در حدود ۲ متر است و در بعضی کاربردها به ۵ متر هم می‌رسد.دیگر مشخصات WBAN در جدول (۱) نشان داده شده است [5].

جدول (۱): مشخصات WBAN

مشخصه	مقدار
فاصله	۲ متر ،در بعضی حالات ۵ متر
زمان راه اندازی	< ۱۰۰ نانو ثانیه
زمان برپایی شبکه	< ۱ ثانیه به ازای هر دستگاه
مصرف انرژی[12]	تقریبا ۱ میلی وات به ازای هر ۱ مگا بیت بر ثانیه

تراکم شبکه	۲ الی ۴ شبکه در هر متر مربع
اندازه هر شبکه (تراکم حسگرها)	حداکثر ۱۰۰ حسگر در هر شبکه
تاخیر انتها به انتها	۱۰ میلی ثانیه

ادامه سمینار به شرح زیر تقسیم می‌شود.در بخش ۲ زیرساخت[۱۳] شبکه‌های WBAN را توصیف می‌کنیم.بخش ۳ کاربردهای WBAN را مطرح می‌کند.معماری شبکه WBAN در بخش ۴ پوشش داده می‌شود.در بخش ۵ نگاهی به به استاندارد 802.15.6 می‌اندازیم.در بخش ۶ امنیت شبکه‌های بدنی بی‌سیم بحث می‌شود. بخش‌های ۷ و۸ به ترتیب به چالش‌های موجود و مباحثی که هنور جای مطالعه و بررسی دارد،اختصاص دارد.در بخش ۹ طرح پیشنهادی ارائه می‌شود و سرانجام نتیجه گیری از مقاله را در بخش ۱۰ می‌آوریم.

۲- زیرساختِ WBAN

همان طور که در شکل (۱) مشاهده می‌کنید، یک WBAN متشکل از تعدادی گره حسگر و یک هماهنگ کننده می‌باشد.هر گره از باتری، حسگر، عملگر[۱۴]، پردازشگر، حافظه و فرستنده- گیرنده تشکیل شده است [6].وظیفهٔ هر گره حسگر دریافت علائم حیاتی بیمار و ارسال آن برای هماهنگ کننده می‌باشد.دو نوع مختلف از حسگر ها از لحاظ محل قرار گیری آن مطرح است:

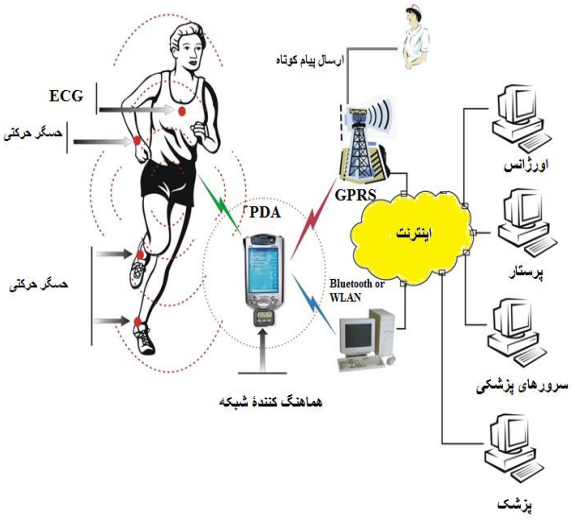
- **حسگرهای کاشتنی**[۱۵] که در زیر بدن انسان کار گذاشته می‌شوند.مثل کپسول های آندسکوپی
- **حسگرهای پوشیدنی**[۱۶] که اساسا برروی بدن نصب می‌شوند. مثل حسگرهای اندازه گیری درجه حرارت بدن.

از جمله حسگرهای موجود می‌توان به حسگرهای ثبت امواج الکتریکی مغز[۱۷](EEG)، ثبت ضربان قلب[۱۸] (ECG)، آنالیز خون، تعیین درجه حرارت بدن، اندازه گیری قند خون و... اشاره کرد.هماهنگ کننده که به آن دستیار دیجیتال شخصی[۱۹] (PDA) هم گفته می‌شود اطلاعات دریافت شده از حسگرها را ذخیره و آن را در صورت لزوم برای متخصصان، مراکز درمانی، اورژانس، پرستاران و سرورهای پزشکی ارسال می‌کند.PDA از راه‌های مختلفی مثل استانداردهای ۸۰۲.۱۱ ، بلوتوث یا شبکه های سلولی اقدام به ارسال اطلاعات می‌کند.مراکز و افراد مربوطه پس از دریافت اطلاعات، بر حسب نوع اطلاعات دریافتی (اورژانسی بودن یا نبودن) تصمیماتی را اتخاذ می‌نمایند.از نقطه نظر اینکه WBAN چگونه تصمیمی را عملی می‌کند سه نوع زیرساخت مطرح شده که در ادامه به آن می‌پردازیم [6].

۱. شبکه‌های بدنیِ بی‌سیمِ مدیریت شده[۲۰] (MWBAN): PDA پس از دریافت اطلاعات، آنالیزهایی بر روی آن انجام می‌دهد و در صورت تشخیص مشکل، پیام هشداری[۲۱] به نزدیکترین بیمارستان یا متخصص ارسال می‌کند.متخصص یا سازمان مربوطه تصمیمی را اتخاذ کرده و به PDA برمی‌گرداند.PDA براساس تصمیم گرفته شده عملی را برروی بدن انسان انجام می‌دهد.برای این کار، شبکه لازم است تا از طریق ارتباطات در محدودهٔ وسیعی مثل شبکه های سلولی یا استاندارد های ۸۰۲.۱۱ به اینترنت وصل شود.مزیت این نوع ساختار، نظارت مستقیم متخصص مربوطه بر روال انجام کار است.علت نام گذاری این نوع زیرساخت به این دلیل است که شخص سومی مدیریت عملکرد شبکهٔ WBAN را بر عهده دارد.

۲. شبکه‌های بدنیِ بی‌سیمِ خود مختار[۲۲] (AWBAN): در این حالت PDA به شبکه‌هایی با محدودهٔ وسیع ارتباط برقرارنمی‌کند.PDA در واقع دستگاهی با هوشِ بالاتری است و طوری برنامه ریزی شده است که خودش می‌تواند تصمیماتی را بگیرد.بنابراین اگر مشکلی رخ دهد، PDA قادر خواهد بود ورودی‌های مختلف را آنالیز کند، تشخیصی را انجام دهد وبه عمل کننده‌ها دستور دهد به منظور حل کردن مشکل عملی را برروی بدن انسان دهند.به عنوان مثال PDA قند خون را اندازه گیری می‌کند و اگر از حد خاصی بالاتر بود به عمل کننده‌ها دستور می‌دهد تا مقداری انسولین در بدن فرد تزریق کنند.از جمله مزایای این نوع شبکه‌ها ، استقلال شبکه و عدم نیاز به شبکه‌هایی با محدودهٔ وسیع است.به هرحال تشخیص و درمان یک بیماری توسط کامپیوتر به خوبی این کار توسط پزشک نخواهد بود.

۳. شبکه‌های بدنیِ بی‌سیمِ هوشمند[۲۳] (IWBAN): در این شبکه‌ها هر دو روش مدیریت شده و خود مختار با یکدیگر ترکیب می‌شوند.در حالت‌های ساده خود WBAN می‌تواند عملی را بر روی بدن انجام دهد(AWBAN) اما زمانی که با وضعیت های پیچیده تری مواجه می‌شویم که نیاز به مهارت‌های انسانی است،WBAN هشداری را برای پزشک ارسال می‌کند و وی بهترین تصمیم را می‌گیرد.



شکل (۱): زیرساخت WBAN برای کاربردهای پزشکی

۳- کاربردهای WBAN

کاربردهای WBAN به دو دستهٔ کلی پزشکی و غیرپزشکی تقسیم می‌شوند که در شکل (۲) نشان داده شده است [7]. کاربردهای پزشکی شامل جمع آوری اطلاعات حیاتی بیمار به طور پیوسته و ارسال به آن ایستگاه‌های راه دور برای تحلیل بیشتر است.این حجم زیاد از اطلاعاتِ بیمار می‌تواند در جلوگیری از رخ دادن حملات قبلی و همچنین مراقبت در برابر بیماری‌های خطرناکی مثل سرطان، آسم،اختلالات اعصاب و... موثر باشد.موارد متعددی از به کار گیری WBAN برای تشخیص و درمان بیماری وجود دارد.بسیاری از محققان به تحقیق در این رابطه پرداخته‌اندکه خلاصه‌ای از آن در جدول (۲) بیان شده است [5].WBAN همچنین برای افرادی که ناتوانی جسمی دارند، مفید است.برای مثال چیپ‌های مصنوعیِ شبکیه چشم می‌تواند درون چشم انسان قرار بگیرد وسطحی از دیدن را فراهم کند.کاربردی در [8] به منظور کمک به افراد معلول مطرح شده است.کاربردهای غیرپزشکی شامل کاربردهای نظامی[۲۴]، بازی ها[۲۵] و شبکه‌های اجتماعی[۲۶] می‌باشد.در کاربردهای نظامی می‌توان به استفاده از یک WBAN در میدان جنگ اشاره کرد.در این کاربرد، WBAN به منظور مرتبط کردن سربازان و گزارش فعالیت‌های آنها به فرماندهان استفاده شود[2] .مثلا میزان هوشیاری سربازان توسط حسگرهایی اندازه گیری و به فرمانده ارسال می‌شود.در بازی‌ها، حسگرهای WBAN می‌توانند جابجایی مختصاتِ قسمت‌های مختلف بدن را جمع‌آوری کند و متعاقبا این حرکت را به شخصیت مورد نظر در بازی انتقال دهد. به عنوان مثال بازی تنیس.با استفاده از WBAN، شبکه‌های اجتماعی به کاربرانشان اجازه می‌دهند تا پروفایل دیجیتالی خود را با تکان دادن دست‌ها عوض کنند.

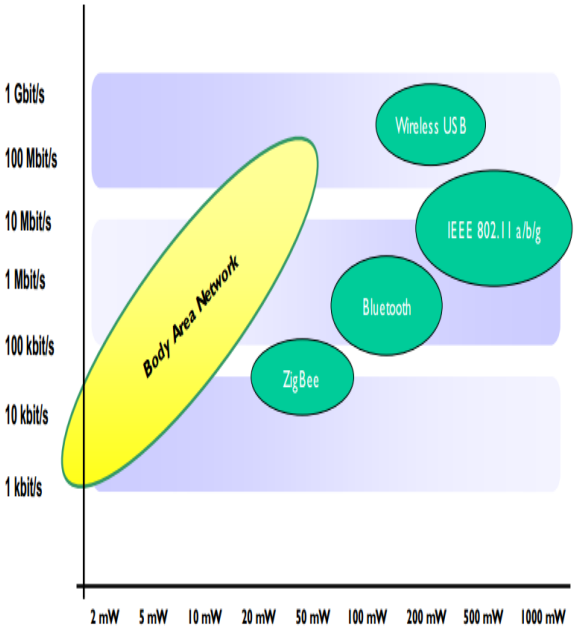
شکل (۲): کاربردهای WBAN

جدول (۲): نقش WBAN در سیستم‌های بهداشت ودرمان

زمینهٔ کاربرد/ بیماری	حسگرها	نقش WBAN
بیماری های قلبی ۳۰٪ کل مرگ و میر ۱۷/۵ میلیون در سال و در سال ۲۰۱۵ به ۲۰ میلیون خواهد رسید.	حسگر ضربان قلب حسگر ECG	کادر پزشکی اگر اطلاعات حیاتی مثل ضربان قلب یا بی نظمی قلب، بیمار را داشته باشند می‌توانند درمان مناسبی برای وی در نظر بگیرند.
سرطان ۱۲/۷ میلیون سرطانی در جهان وجود دارد و تخمین زده می‌شود که ۷/۶ میلیون نفر آنها به مرگ منجر خواهد شد.	حسگر اکسید نیتریک	حسگر می‌تواند در محل مشکوک قرار بگیرد و پزشک می‌تواند درمان را با تشخیص سلول‌های سرطانی هرچه سریعتر آغاز کند.
آلزایمر، افسردگی ، فشارخون ۳۵۷ میلیون نفر در سال ۱۹۹۰ و پیش بینی می‌شود در سال ۲۰۲۵ به ۷۶۱ میلیون نفر برسد.		WBAN می تواند در صورت مشاهدهٔ هرگونه وضعیت غیر معمولی افراد تنها و مسن، هشداری را به خانواده، همسایه یا به نزدیکترین بیمارستان بفرستد.
بیماری قند (دیابت) ۲۴۶ میلیون نفر در جهان به این بیماری مبتلا می‌باشند.	یک حسگر برای اندازه گیری قند خون/ یک عمل کننده به منظور تزریق انسولین	اگر حسگر افت ناگهانی قند خون را مشاهده کرد یک سیگنال به عمل کننده جهت تزریق انسولین می‌فرستد.
آسم ۳۰۰ میلیون در جهان به این بیماری مبتلا می‌باشند.	حسگرهای حساسیت	حسگرهایی که می‌توانند عوامل الرژی زا را در هوا حس کنند و مرتبا بیمار را از این موضوع آگاه کنند.
نظارت های بعد از عمل	حسگرهای درجه حرارت حسگرهای فشار خون حسگرهای ضربان قلب	بیمار دیگر نیازی نیست برای مدت طولانی در تخت بیمارستان بماند.
فشار خون بالا باعث بروز سکته در بیش از ۷/۱۲ میلیون نفر در جهان می‌شود.	حسگر فشار خون همراه با یک عمل کننده	اگر حسگر تغییری در در فشار خون مشاهده کند و این تغییر بیش از حد مجاز باشد، به عمل کننده سیگنالی ارسال و تزریق دارو در بدن صورت می‌گیرد در نتیجه شانس کمتری برای بروز سکته وجود دارد.

همان طور که بیان شد WBAN طیف گسترده‌ای از کاربردها را شامل می‌شود.برخی از کاربردها به نرخ داده[۲۷] و توان ارسالی کمی نیاز دارند همچون حسگرهای پزشکی که علائم منفردی از بدن را مشاهده می‌کنند.در مقابل کاربردهایی وجود دارد که به نرخ داده و توان بالایی نیاز دارند.از جمله این کاربردها می‌توان به عکس

برداری و فیلم برداری‌هایی اشاره کرد که توسط حسگر ها از داخل بدن صورت می‌گیرد.شکل (۳) طیفی از نرخ داده و توان مصرفی استفاده شده در کاربردهای WBAN و دیگر استاندارد های شبکه‌های بی‌سیم را نشان می‌دهد.



شکل(۳): نرخ داده در مقابل توان ارسالی

۴ – معماری WBAN

در این بخش به معماری شبکه‌های WBAN از دو منظر نگاه می‌کنیم.ابتدا چگونگی ارتباط بین حسگرها (گره) و PDA (سینک[۲۸]) را تحت عنوان توپولوژی ستاره[۲۹] و چند پرشی بررسی[۳۰] می‌کنیم و سپس در یک دید کلی‌تر به بررسی چگونگی ارتباط WBAN با مراکز مربوطه، تحت عنوان معماری سه لایه‌ای می‌پردازیم.

۴–۱ – بررسی توپولوژی ستاره و چند پرشی

معماری شبکه، سازماندهی منطقیِ دستگاه‌های ارتباطی در یک سیستم می‌باشد [9].معماری شبکه بر اساس ویژگی های سیستم در شرایط خاصی انتخاب می‌شود و این انتخاب می‌تواند بروری کارایی[۳۱] سیستم از طرق مختلفی همچون مصرف انرژی، توانایی در رفتار کردن با انواع بارهای ترافیکی، پایداری شبکه[۳۲]، انتخاب پروتکل MAC ، تاخیر[۳۳] انتقال و تداخل[۳۴] کاربران تاثیر بگذارد.

۴–۱–۱ – مصرف انرژی

انتخاب معماری باید به گونه‌ای صورت پذیرد که در شرایط ایده‌آل مصرف انرژی را برروی کل شبکه به طور یکنواخت توزیع کند.

۴–۱–۲ – سازگاری[۳۵] با انواع بارهای ترافیکی

شبکه باید انواع بارهای ترافیکی را پشتیبانی کند که این مساله تا حدودی به معماری شبکه بستگی دارد.تعدادی از معماری‌ها صرفا جهت پشتیبانی از ترافیک‌ها با نرخ داده‌ای کم طراحی شده‌اند.

۴–۱–۳ – پایداری شبکه

در برخی از معماری‌ها عملکرد کل شبکه ممکن است به یک یل چند گرهٔ خاص بستگی داشته باشد.در این شرایط خرابی گره باعث از کار افتادن شبکه می‌شود.

۴–۱–۴ – انتخاب پروتکل MAC

انتخاب پروتکل MAC و معماری شبکه به شدت به یکدیگر وابسته می‌باشند به طوری که معمولا انتخاب یک معماری خاص،پروتکل MAC مخصوصی را می‌پذیرد یا بالعکس.

۴–۱–۵ – تاخیر انتقال

جدای شرایط ترافیکی، برخی از معماری‌ها تاخیر زیادی را نتیجه می‌دهند و این به دلیل تعداد پرش‌هایی است که داده برای رسیدن به سینک طی می‌کند.

۴–۱–۶ – تداخل کاربران

حضور کاربران در یک مکان باعث تداخل سیگنال‌های انتقالی آنها می‌شود.برخی از معماری‌ها که زمینه را برای انتقال با توان بالا فراهم می‌کنند، باعث افزایش تداخل سیگنال‌های انتقالی کاربران می‌شود.

در مقایسه با شبکه‌های حسگر بی‌سیم که به منظور مشاهدات محیطی به کار می‌روند، WBAN دارای محدودیت‌های مضاعفی در برقراری ارتباطات و دریافت داده از بدن انسان می‌باشد به عنوان مثال به دلیل محدودیت‌هایِ بدنِ انسان تعدادی کمی حسگر می‌تواند روی بدن پخش شود، با توجه به این محدودیت‌ها و به دلیل مقیاس کوچک WBAN و فاصلهٔ کوتاه بین لینک‌ها (ارتباط بین گره و سینک) توپولوژی ستاره به عنوان یک انتخاب طبیعی و پیش فرض برای WBAN محسوب می‌شود [10].از سوی دیگر به دلیل حرکت کاربر و ماهیت بدن انسان، کانال‌های ارتباطی بی‌سیم روی بدن دائما تغییر می‌کند [11].در نتیجه نیاز به برقراری ارتباط بی‌سیم مطمئن که انرژی را به طور موثر بگیرد احساس می‌شود.به همین دلیل معماری دیگری تحت عنوان چند پرشی مطرح می‌شود که در آن حسگرها ممکن است به عنوان گره‌های رله[۳۶] استفاده شوند تا کارایی انتقال را افزایش دهند.در [10] آزمایشی به منظور تحلیل کارایی معماری ستاره و چند پرشی صورت گرفته است.نتایج بدست آمده در این مقاله تایید می‌کند که معماری چند پرشی در مقایسه با معماری ستاره کاراتر می‌باشد و در کاربردهای پزشکی، معیارهای قابلیت اطمینان[۳۷] و مصرف

انرژی به طور موثر[38] را فراهم می‌کند.همچنین در[12] معماری‌های ستاره و چند پرشی مورد بررسی قرار گرفته است.در این مقاله دو نمونه از شبکه‌های چند پرشی مطرح شده است.شبکه‌ای با معماری چند پرشی به منظور بهینه کردن حداکثر نرخ تحویل دادن بسته[39] (PDR) و شبکه‌ای با معماری چند پرشی به منظور بهینه کردن حداقل میانگین تعداد انتقال دوباره[40] (ANR).نتایج بدست آمده از این مقاله در پایین لیست شده است.

- معماری ستاره برای محیط‌های پویا[41] نا مناسب است.معماری چند پرشی چون از گره‌های رله استفاده می‌کند می‌تواند خود را با تغییرات وفق دهد.
- اگر یک شبکه به PDR بالایی نیاز داشته باشد و در محیط پویایی قرار داشته باشد آنگاه باید معماری چند پرشی ـ با حداکثر PDR به کار گرفته شود.
- اگر طراحِ WBAN بخواهد شبکه‌ای طراحی کند که حداقل انرژی را مصرف کند و کمترین تاخیر را هم داشته باشد، معماری چند پرشی ـ با حداقل ANR بهترین انتخاب است.
- از نقطه نظر تداخل کاربران، معماری ستاره نباید در محیط‌های باز استفاده شود.چون در این محیط‌ها تنها راه افزایش قابلیت اطمینان، افزایش توان است و این افزایش توان به شدت بر تداخل کاربران تاثیر می‌گذارد.

در [13] نگاهی به معماری‌های شبکهٔ WBAN می‌اندازد و نتیجه می‌گیرد از نقطه نظر مصرف انرژی، معماری چند پرشی سودمندتر است.همچنین مولفان نشان می‌دهند که در بعضی شرایط تنها گزینهٔ موجود استفاده از معماری چند پرشی می‌باشد.

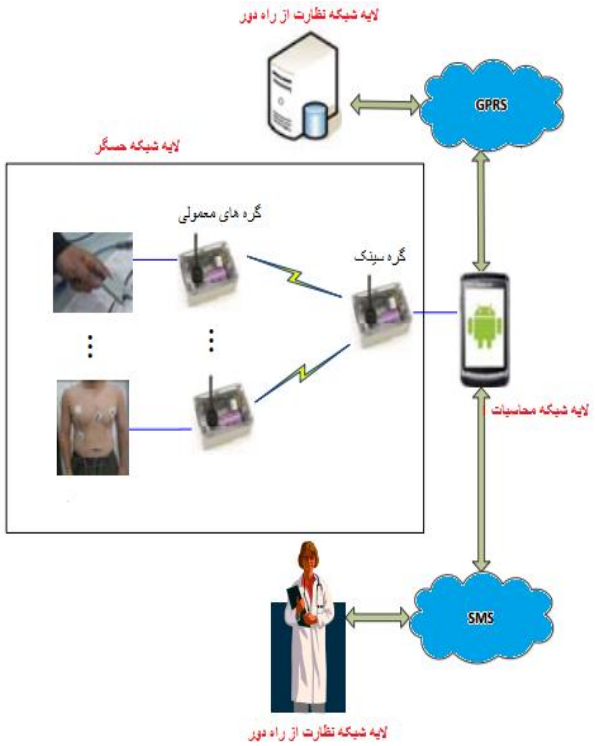
۴-۲- معماری سه لایه‌ای سیستم نظارت بر سلامتی بیمار

در این بخش سیستم نظارت بر سلامتی بیمار که متشکل از سه لایه است، بررسی می‌شود.لایه اول:لایه شبکه حسگر، لایه دوم:لایه شبکه محاسبات، لایه سوم:لایه شبکه نظارت از راه دور.شکل (۴) جزئیات و ارتباط بین لایه ها را نشان می‌دهد [14].

لایه اول اساسا شامل حسگرهای پزشکی بی‌سیم می‌باشد.این لایه در تماس با بدن انسان قرار دارد.این حسگرها بر اساس توپولوژی‌های مطرح شده در قسمت قبل به یکدیگر متصل می‌شوند.یکی از این حسگرها نقش سینک را ایفا می‌کند و بقیهٔ حسگرها به عنوان گره‌های معمولی شناخته می‌شوند.گرهٔ سینک به عنوان یک پل ارتباطی بین لایه اول و لایه دوم است.

لایه دوم شامل یک PDA یا یک تلفن هوشمند است.این تلفن هوشمند به عنوان یک ترمینال برای کاربران در نظر گرفته می‌شود.کاربران از طریق این ترمینال می‌توانند اطلاعات را مشاهده کنند.اطلاعات در این تلفن هوشمند جمع‌آوری و پس انجام محاسباتی برروی آن (جهت تشخیص رخدادی) از طریق شبکه‌های سلولی یا وای-فای به لایه سوم ارسال می‌شود.

لایه سوم که بالاترین لایه در این معماری است شامل تلفن هوشمند پزشکان و مراکز مشاهدهٔ از راه دورِ مستقر در بیمارستان یا اورژانس‌ها می‌باشد.این مراکز داده‌های بدست آمده را ذخیره و دسته بندی می‌کنند و از این داده‌ها جهت انجام عملیات خاصی استفاده می‌شود.در این لایه همچنین امکان به کار گیری GPS و GSM جهت پیدا کردن موقعیت بیمار در نظر گرفته شده است.این امکان برای بیماری‌های ناگهانی مثل سکته‌های مغزی و قلبی بسیار مناسب است [15].



شکل (۴): معماری سه لایه‌ای سیستم نظارت [14]

۵- مروری بر استاندارد 802.15.6

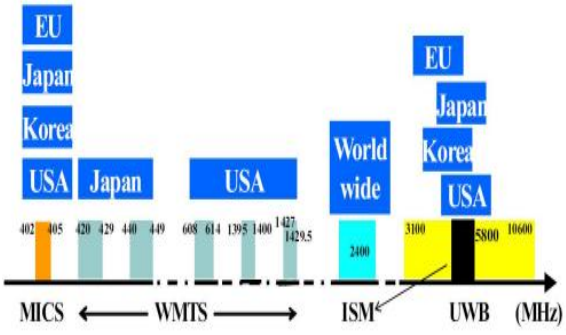
به منظور پیاده سازی موفق WBAN نیاز به یک استاندارد جامع به شدت احساس می‌شد.به همین منظور IEEE 802 یک گروه کاری به نام IEEE 802.15.6 را برای استاندارد کردن WBAN در سال ۲۰۰۷ مشخص کرد.هدف این گروه فراهم کردن یک استاندارد بین-المللی به منظور برقراری یک ارتباط بی‌سیم برای یک محدودهٔ کوتاه با توان کم و قابلیت اطمینان بالا بود که در نزدیکی یا داخل بدن انسان بتواند کار کند [7].از جمله ویژگی دیگری که در این استاندارد باید لحاظ می‌شد،نرخ دادهٔ بالایی بود که در برخی کاربردهای سرگرمی و سیستم نظارت بر سلامت بیمار نیاز است (تا ۱۰ مگابیت بر

ثانیه).استاندارد شبکه‌های شخصی موجود[۴۲] (PAN) پارامتر های پزشکی همچون خطر نداشتن برای انسان را در نظر نمی‌گیرند همچنین PAN ترکیبی از قابلیت اطمینان بالا، کیفیت سرویس[۴۳] (QoS)، توان پایین، انواع نرخ داده و عدم تداخل مورد نیاز را پشتیبانی نمی‌کند [16].

هدف اصلی 802.15.6 تعریف لایه فیزیکی و MAC جدید برای شبکه‌های بدنی بی‌سیم است.انتخاب باند فرکانسی مناسب برای این استاندارد یکی از مهمترین مسائل بود که باید به آن توجه می‌شد.شکل (۵) به طور مختصر برخی از باندهای فرکانسی قابل دسترس برای WBAN در کشورهای مختلف را نشان می‌دهد.باند خدمات ارتباطی ایمپلنت پزشکی[۴۴] (MICS) یک باند دارای مجوز[۴۵] است که برای ارتباطات ایمپلنت پزشکی استفاده می‌شود و در بیشتر کشورها دارای محدوده فرکانسی (۴۲۰-۴۰۵ مگا هرتز) یکسانی است.باند خدمات تله متری پزشکی بی‌سیم[۴۶] (WMTS) نیز یک باند دارای مجوز است که برای سیستم‌های تله متری پزشکی به کار می‌رود.پهنای باند کمِ WMTS و MICS برای کاربردهایی که نرخ داده‌ای بالایی نیاز دارند مناسب نیست.باند پزشکی،علمی،صنعتی[۴۷] (ISM) نیاز به مجوز ندارد و در کل جهان در دسترس است.برای کاربردهایی با نرخ داده‌ای بالا نیز مناسب است ولی به دلیل وجود دستگاه‌های بی‌سیم زیادی همچون بلوتوث احتمال تداخل در آن وجود دارد.

برای ارتباطات امن 802.15.6 سه سطح را تعریف کرده است: ۱) سطح ۰ - ارتباطات نا امن، ۲) سطح ۱ - فقط اهراز هویت[۴۸]، ۳) سطح ۲ – هم اهراز هویت و هم رمز نگاری[۴۹].هر کاربر با توجه به شرایطش هر کدام از این سطوح را که برایش بهترین باشد انتخاب می‌کند.زمانی که یک گره به شبکه ملحق می‌شود، سطح امنیتی خود را انتخاب می‌کند.

دو نوع توپولوژی در این استاندارد در نظر گرفته شده است: توپولوژی ستاره – تک پرشی و توپولوژی ستاره – دو پرشی.در حالت دوم ارتباط بین گره‌ها و گره سینک از طریق گرهٔ رله صورت می‌پذیرد.در ادامه به دلیل اهمیت لایه فیزیکی و MAC بیشتر به آن می‌پردازیم.



شکل (۵): باند فرکانسی برای WBAN [7]

۵-۱- مشخصات لایه فیزیکی

استاندارد 802.15.6 سه لایه فیزیکی مختلف را پشتیبانی می‌کند. NB[۵۰]، UWB[۵۱]، HBC[۵۲] [17] .

۵-۱-۱- NB

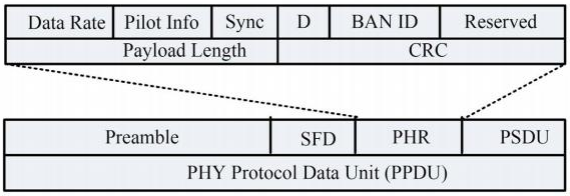
شکل (۶) قالب واحد داده‌ای لایه فیزیکی[۵۳] (PPDU) در NB را نشان می‌دهد که از سه مولفهٔ اصلی تشکیل شده است.مقدمهٔ[۵۴] PLCP[۵۵]، هدر PLCP و PSDU[۵۶].مقدمه به گیرنده در سنکرون شدن و بازیابی افستِ فرکانس حامل کمک می‌کند. هدر PLCP شامل داده‌های ضروری برای دیکد کردن بسته در گیرنده است.PSDU هم در واقع قاب داده ای است که از لایه MAC دریافت کرده است.حداکثر نرخ داده در NB ۴۸۵ کیلو بیت بر ثانیه است.در نتیجه NB برای کاربردهایی که نیاز به نرخ داده‌ای بالاتری دارند مناسب نیست.

۵-۱-۲- UWB

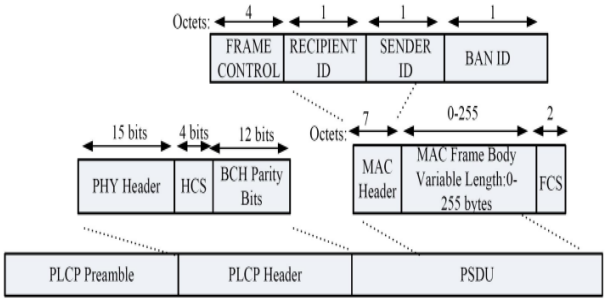
UWB به دلیل پهنای باند زیادی که به کار می‌گیرد،دامنهٔ وسیعی از پیاده سازی‌ها با کارایی بالا، استحکام، پیچیدگی کم و توان مصرفی فوق‌العاده پایین را فراهم کرده است.علاوه بر این، مزیتِ UWB در این واقعیت نهفته است که مقدار توانی که UWB به کار می‌گیرد منطبق بر باند MICS است.بنابراین توان ارسالی به قدری است که نه برای بدن انسان ضرری دارد و نه با دیگر دستگاه‌ها تداخل پیدا می‌کند.ساختار قاب داده لایه فیزیکی در شکل (۷) نشان داده شده است.PPDU از SHR[۵۷]، PHR[۵۸] و PSDU تشکیل شده است.SHR به دو قسمت تقسیم می‌شود.قسمت اول مقدمه به منظور سنکرون کردن زمانبندی،تشخیص بسته و بازیابی آفستِ فرکانس حامل در نظر گرفته می‌شود.قسمت دوم SFD[۵۹] برای سنکرون کردن قاب استفاده می‌شود.PHR اطلاعاتی در مورد نرخ دادهٔ PSDU، طول قاب لایه MAC و ... را ارسال می‌کند.PSDU که هم شامل داده اصلی است که از لایه MAC می‌آید.UWB در دو باند فرکانسی کار می‌کند.باند کوتاه[۶۰] و باند بلند[۶۱].جدول (۳) باندهای فرکانسی که در UWB به کار گرفته می‌شود را نشان می‌دهد.

۵-۱-۳- HBC

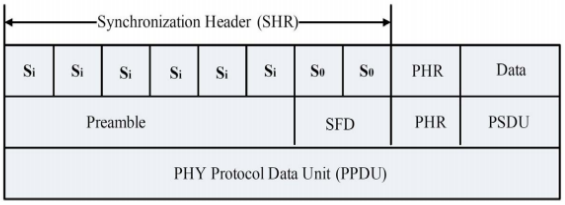
HBC در دو باند فرکانسی با فرکانس مرکزی ۱۶ مگا هرتز و ۲۷ مگا هرتز کار می‌کند.پهنای باند هر دو ۴ مگا هرتز است.هر دو باند برای آمریکا، ژاپن و کره معتبر است و باندی که در ۲۷ مگا هرنز کار می‌کند فقط برای اروپا معتبر است.شکل (۸) ساختار PPDU لایه فیزیکیِ HBC را توصیف می‌کند.این ساختار متشکل از مقدمه، SFD، PHR، و PSDU است.وظیفهٔ مولفه‌های بیان شده در بخش های قبلی (NB و UWB) توضیح داده شده است. نرخ داده‌ای که HBC می‌تواند فراهم کند از ۱۶۴ کیلو بیت بر ثانیه تا ۳۱۲۵/۱ مگا بیت بر ثانیه است.



شکل (۸): ساختار PPDU در HBC



شکل (۶): ساختار PPDU در NB



شکل (۷): ساختار PPDU در UWB

جدول (۳): باندهای فرکانسی UWB

نوع باند	شماره کانال	فرکانس مرکزی	پهنای باند (مگا هرتز)	خصوصیت کانال
باندپایین	۰	۴/۳۴۹۴	۲/۴۹۹	اختیاری
	۱	۶/۳۹۹۳	۲/۴۹۹	اجباری
	۲	۸/۴۴۹۲	۲/۴۹۹	اختیاری
باند بالا	۳	۶/۶۴۸۹	۲/۴۹۹	اختیاری
	۴	۸/۶۹۸۸	۲/۴۹۹	اختیاری
	۵	۰/۷۴۸۸	۲/۴۹۹	اختیاری
	۶	۲/۷۹۸۷	۲/۴۹۹	اجباری
	۷	۴/۸۴۸۶	۲/۴۹۹	اختیاری
	۸	۶/۸۹۸۵	۲/۴۹۹	اختیاری
	۹	۸/۹۴۸۴	۲/۴۹۹	اختیاری
	۱۰	۰/۹۹۸۴	۲/۴۹۹	اختیاری

۵-۲- زیر لایه کنترل دسترسی رسانه(MAC)

در این بخش سعی می‌شود پروتکل MAC استاندارد 802.15.6 ارائه شود.این استاندارد حالت‌های مختلف دسترسی،فازها و مکانیسم‌های دسترسی به رسانه را مشخص می‌کند.

در WBAN کانال به صورت سوپرفریم‌هایی از هم جدا می‌شود.طول هر سوپرفریم توسط بیکن[۶۲] هایی که به صورت دوره‌ای ارسال می‌شود مشخص می‌شود.این سوپرفریم‌ها دارای طول مساوی می‌باشند.استاندارد 802.15.6 در یکی از حالت‌های زیر کار می‌کند.

۵-۲-۱- Beacon mod with super frame

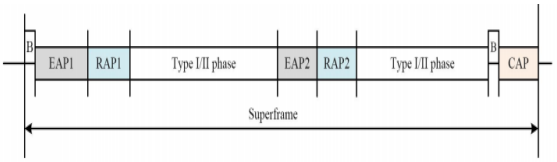
در این حالت، بیکن‌ها توسط هاب در هر سوپرفریم به جز سوپرفریم‌های غیر فعال انتقال داده می‌شود.شکل (۹) ساختار سوپرفریم 802.15.6 را نشان می‌دهد که به قسمت‌های بیکن، فاز دسترسی انحصاری ۱[۶۳] (EAP1)، فاز دسترسی تصادفی ۱[۶۴] (RAP1)، فاز نوع ۲/۱، EAP2، RAP2، فاز نوع ۲/۱ و فاز دسترسی رقابتی[۶۵] (CAP) تقسیم می‌شود.در دوره های EAP، RAP و CAP گره‌ها برای تخصیص منابع با استفاده از رویه‌های دسترسی CSMA/CA یا الوهای برهه‌ای[۶۶] با یکدیگر رقابت می‌کنند.EAP1 و EAP2 برای ترافیک‌ها با اولویت بالا مثل گزارش رخدادهای اورژانسی استفاده می‌شوند.در حالی که RAP1، RAP2 و CAP فقط برای ترافیک معمولی استفاده می‌شوند.فاز نوع ۲/۱ برای تخصیص فواصل بالاسو[۶۷]، پایین‌سو[۶۸] استفاده می‌شود.در این فاز ار روش سرکشی[۶۹] برای تخصیص منابع استفاده می‌شود.با توجه به کاربرد مورد نظر هماهنگ کننده می‌تواند هریک از این فازها را با تخصیص زمان صفر برای آنها غیر فعال کند.

۵-۲-۲- non-Beacon mode with super frame

در این حالت تمام طول سوپر فریم توسط فاز نوع ۱ یا ۲ پوشش داده می‌شود.ولی همزمان توسط هر دوی آنها پوشش داده نمی‌شود.

۵-۲-۳- non-Beacon mode without super frame

در این حالت هماهنگ کننده فقط تخصیص را از نوع ۲ فراهم می‌کند



شکل (۹): ساختار سوپرفریم IEEE 802.15.6

۶- امنیت در شبکه‌های بدنی بی‌سیم

از آنجایی که گره‌های WBANاطلاعات حساسی (علائم حیاتی بدن) را جمع آوری می‌کنند و ممکن است که در محیط‌های خصمانه قرار بگیرند، به کارگیری مکانیسم‌های قوی امنیتی امری ضروری است.در کاربردهای پزشکی تهدیدات امنیتی ممکن است شرایط خطرناکی را برای بیمار ایجاد کند و در گاهی اوقات باعث مرگ وی شود.فرض کنید برروی یک سیگنالی که حاوی اطلاعاتی مبنی بر بالا بودن قند خون فرد بیمار است، حمله ای صورت پذیرد.عدم رسیدن این سیگنال به پزشک و یا مراکز درمانی باعث بروز تشنج و در صورت عدم تزریق انسولینِ به موقع باعث مرگ وی می‌شود.امنیت یکی از مهمترین چالش‌های شبکه‌های بدنی بی‌سیم محسوب می‌شود[18]. امنیت در شبکه های WBAN در سه مرحله ارائه شده است:

- **رمزنگاری**: اطلاعات حیاتی انسان نباید توسط افراد مختف مورد استفاده قرار گرفته و خوانده شود. پس لزوم رمز نگاری‌هایی منطبق بر شبکهٔ WBAN امری ضروری است.استفاده از رمزنگاری‌های متقارن [۷۰] کاراتر می‌باشد.زیرا رمزنگاری کلید عمومی [۷۱] به دلیل محاسبت سنگینش و بالطبع مصرف انرژی زیاد برای گره هایی که انرژی محدودی دارند،مناسب نیست [19].
- **احراز هویت**: برخی از حسگرها با توجه به عامل‌هایی که برای آنها در نظر گرفته شده است قادرند تا عملی را برروی بدن انسان انجام دهند.دسترسی هر کس جهت اقدام برروی بدن انسان بدون هیچ مجوزی، خطرات بسیاری را پدید می‌آورد.پس احراز هویت باید انجام شود.در [20] با استفاده از رمزنگاری کلید متقارن یک روش احراز هویت ترکیبی مطرح و بررسی شده است.
- **تشخیص و تصحیح خطای انتقال**: محافظت در برابر خطای انتقال باید حتما در نظر گرفته شود. فرض کنید خطایی در انتقال داده‌های اورژانسی صورت پذیرد انگاه ممکن است جان انسانی بدلیل نرسیدن اطلاعات به پزشک یا مرکز مربوطه به خطر بیافتد

در [21] نیازمندی‌های امنیتی که باید برای WBAN در نظر گرفته شود به صورت دیگری بیان شده است:

- قابلیت اعتماد [۷۲] داده‌ها: فاش نشدن اطلاعات حیاتی برای دیگران
- یکپارچگی [۷۳] داده‌ها: جلوگیری از کم و زیاد کردن داده‌ها توسط مخاصم
- احراز هویت: تایید اینکه داده‌های دریافتی از منبع صحیح صورت گرفته است.
- موقعیت یابی امن: جلوگیری از اقدامات مخاصم که مانع از موقعیت یابی بیمار شود. در کاربردهای پزشکیِ WBAN می‌توان پس از تشخیص اورژانسی بودن وضعیت بیمار، وی را از طریق تکنولوژی‌های مختلفی همچون GPS،GSM موقعیت یابی کرد و اقدامات لازم مثل اعزام آمبولانس به محل مربوطه را انجام داد.
- قابلیت دسترس پذیری [۷۴]: جلوگیری از فعالیت‌های مخاصم جهت ناتوان ساختن دسترس پذیری اطلاعات توسط پزشکان
- مدیریت امن: این مدیریت باید برای PDA لحاظ شود.به طوری که بتواند به منظور عملیات رمزنگاری،کلید ها را به روش امنی توزیع کند. PDA وظیفه دارد به روش امنی گره‌ای را حذف یا اظافه کند. مثلا زمانی که گره‌ای را حذف کرد،کلید اختصاص یافته به آن را هم حذف کند. این کار مانع دسترسی گرهٔ حذف شده به شبکه می‌شود یا اگر از یک کلید یکسان برای تمام گره‌ها استفاده می‌کند پس از حذف گره، کلید را تغییر دهد.

یک WBAN در برابر تعدادی از حملات آسیب پذیر است[38]. این حملات به روش‌های مختلفی انجام می‌شود.حملات جلوگیری از سرویس (DoS)، حملات نقض محرمانگی و حملات فیزیکی.به دلیل محدودیت‌های توان مصرفی حسگرها، محافظت در برابر این نوع حملات امری چالش برانگیز است.یک حسگر قدرتمند (خارج از شبکه) به راحتی می‌تواند با ارسال یک سیگنال قوی برروی گره حسگر مانع از جمع‌آوری داده های بیمار شود.(Jamming)

حملات بر روی WBAN به سه دسته‌ی اصلی تقسیم می‌شوند : (الف) حملات برروی محرمانه بودن و احراز هویت، مخاصم از طریق استراق سمع و جعل و ارسال دوباره بسته این نوع حملات را انجام می‌دهد، (ب) حملات بر روی یکپارچگی سرویس، شبکه مجبور به دریافت اطلاعات غلطی می‌شود. (پ) حملات جلوگیری از سرویس، قابلیت دسترسی به شبکه را مختل می‌کند.

۷- چالش‌های WBAN

توسعه WBAN چالش‌های تحقیقاتی همچون نیاز به طراحی بهتر حسگرها، مجتمع شدن با سیستم‌های درمانی، امنیت و قابلیت اطمینان اطلاعات[1]، کیفیت سرویس و مصرف انرژی به طور موثر را به وجود آورده است [3].در ادامه این چالش‌ها با جزئیات بیشتری بررسی می‌شود.

۷-۱- بهبود طراحی حسگرها

یکی از مولفه‌های اصلی که در طراحی حسگرها باید در نظر گرفت کوچک بودن آنها است.از طرفی دیگر حسگرهایی که درون بدن انسان قرار می‌گیرند باید سازگار با بدن انسان باشند به طوری که هیچ گونه عوارضی برای بیمار نداشته باشند [23].پیشرفت‌های جدید در

تکنولوژی های زیست شناختی[۷۵]، شیمی، الکترونیک و مکانیک باعث عرضه شدن حسگرهای جدیدی سازگار با بدن انسان شده است.همچنین ارتقا در ساختِ حسگرها و تکنیک‌های مهندسی نانو پتانسیل ساخت حسگرهای کوچکتر را فراهم می‌کند.یکی از این حسگرها که از مهندسی نانو بهره می‌گیرد و تحت پیاده سازی هم می‌باشد در [22] بحث شده است.

۷-۲- مجتمع شدن با سیستم‌های درمانی

مجتمع شدن حسگرها و سیستم‌های درمانی و گرفتن بازخورد[۷۶] از این سیستم مجتمع شده، نقش مهمی از شبکه‌های بدنی را در عملکردهای بالینی[۷۷] و پزشکی تعریف می‌کند [24].این موضوع در کاربردهایی که حسگرها عمل تزریق دارو را نیز انجام می‌دهند ملموس‌تر است.در این شبکه‌ها تزیق دارو براساس شرایط فعلی بیمار، علائم حیاتی بیمار، سوابق موجود بیمار در مراکز درمانی و در نهایت به دستور پزشک صورت می‌پذیرد.همان‌طور که مشخص است یک تعامل کامل بین شبکه‌های بدنی بی‌سیم و سیتم‌های درمانی (پزشکان، مرکز بهداشت و درمان، اورژانس و...) نیاز است که باید در طراحی WBAN لحاظ شود.

۷-۳- امنیت و قابلیت اطمینانِ اطلاعات

از لحاظ امنیتی، داده‌های WBAN باید با استفاده از رمزنگاری‌های قوی و همچنین دیگر روش‌های امنیتی از دسترسی افراد غیر مجاز مصون بمانند.این رمز نگاری‌های قوی به منابع و محاسبات وسیعی احتیاج دارد.با در نظر گرفتن محدودیت منابع (انرژی مورد نیاز) که گره‌های WBAN با آن مواجه هستند باید یک مصالحه‌ای بین حداکثر کردن امنیت و حداقل استفاده از منابع صورت پذیرد.علاوه بر این به دلیل طبیعت بسیار پویای WBAN روش های احراز هویت که برای محیط‌های ایستا در نظر گرفته می‌شود، قابل اجرا نخواهد بود.حتی روش‌های رمزنگاری متقارن که برای شبکه‌های موردی به کار می‌رود محاسبات سنگینی را بر روی کاربردهای WBAN تحمیل می‌کند.

از طرفی دیگر قابلیت اطمینان شبکه به طور مستقیم بر روی کیفیت نظارت بیمار تاثیر می‌گذارد و در بدترین حالت ممکن است به دلیل عدم تشخیص رخدادی که زندگی بیمار را تهدید می‌کند، باعث مرگ وی شود.به هر حال به دلیل محدودیت‌های که روی پهنای باند و انرژی وجود دارد، استفاده از تکنیک‌های که برای ایجاد قابلیت اطمینان در شبکه‌های قدیمی به کار می‌رفت مثل مکانیسم انتقال دوبارهٔ بسته‌ها در پروتکل TCP،چندان برای شبکه‌های WBAN عملی نیست.محققان روش‌های متعددی برای بهبود قابلیت اطمینان مطرح کرده‌اند.در ساده‌ترین روش بسته چند بار ارسال می‌شود تا اعلام وصول آن دریافت شود.این ارسال‌های چند باره سربار[۷۸] زیاده بر روی شبکه تحمیل می‌کنند.روش دیگر برای افزایش قابلیت اطمینان استفاده از کدهای تصحیح کنندهٔ خطا می‌باشد.در [25] با استفاده از کدهای تبدیل لوبی[۷۹] که برای تصحیح خطا به کار می‌رود روشی برای افزایش قابلیت اطمینان مطرح شده است.کدهای تبدیل لوبی براساس XOR کردن بیت‌های اطلاعات محاسبه می‌شود.چون ساختار این کد ساده است انرژی کمی مصرف می‌شود.در [26] به منظور افزایش قابلیت اطمینان توپولوژی شبکه را مورد بحث قرار داده است.در این مقاله نشان داده شده است که توپولوژی ستاره که به منظور سادگی در بیشتر WBAN ها به کار گرفته می‌شود قابلیت اطمینان بالایی را تضمین نمی‌کند. در عوض توپولوژی درختی محدود شده[۸۰] (RTT) را معرفی می‌کند.RTT یک شبکه دو پرشی است.در این نوع توپولوژی گره‌هایی به عنوان گره‌های رله برای برقراری ارتباط بین نودهای انتهایی و هماهنگ کننده معرفی می‌شود.

۷-۴- کیفیتِ سرویس (QoS)

به دلیل وجود کاربردهای بلادرنگ[۸۱] و غیر بلادرنگ در شبکه‌های بدنی، نیاز به سطوح مختلف سرویس دهی و تجریه و تحلیل و برآورد دقیقی از سطوح کیفیت سرویس احساس می‌شود.بسته‌های حاوی اطلاعات اورژانسی بیمار باید با استفاده از سطوح مختلف QoS طوری سرویس بگیرند که بتوانند در موقعیت‌های مختلف که بیمار در شرایط خطرناکی قرار دارد به وی کمکی قابل اعتماد و فراگیری انجام دهند.این موضوع وقتی ظرفیت شبکه محدود است،بحرانی‌تر خواهد بود [27].در میان پارامترهای مختلف QoS دسترس پذیری، یکپارچگی، تاخیر تحویل داده، قابلیت اطمینان و تحرک پذیری[۸۲] به عنوان نیازهای اصلی سیستم‌های بهداشت و روان در نظر گرفته می‌شود [28].چالش‌های عمده‌ای برای پشتیبانی از کیفیت سرویس در [29] مطرح شده است.این چالش‌ها عبارتنداز: محدودیت منابعی همچون باتری و پهنای باند، الگوهای ترافیک غیر قابل پیش بینی، بی‌ثباتی شبکه، توزیع یکنواخت انرژی بین گره‌ها، بحرانی بودن برخی بسته‌ها، پشتیبانی از ترافیک‌های نا متعادل و... .

۷-۵- مصرف انرژی

مدیریت توان در شبکه‌های بدنی یک موضوع کاربردی بسیار مهم است.مصرف توان با بهینه کردن فرآیندهای لایه فیزیکی و MAC به حداقل می‌رسد.یک لایه فیزیکی با انتخاب کدینگ و مودلاسیون مناسب می‌تواند احتمال ارسال بسته به طور موفقیت آمیز را افزایش دهد.هرچه این احتمال بشتر باشد به همان نسبت هم در مصرف انرژی صرفه جویی می‌شود.توان مصرفی یک گره با افزایش احتمال ارسال موفق بسته می‌تواند بهینه شود.از طرف دیگر لایه MAC با بکار گیری تکنیک‌هایی از جمله تکنیک‌های دسترسی به کانال، تکنیک‌های سیگنالینگ هوشمند و همچنین استفاده از ساختار بهینه بسته می‌تواند در کاهش مصرف توان نقش بسیار مهمی داشته باشد.

MAC های مختلفی به منظور اینکه انرژی را به طور موثری بکار گیرند طراحی و بررسی شده است.در [30] یک پروتکل MAC به نام

body mac مطرح شده است.body mac با استفاده از تخصیص پهنای باندِ انعطاف پذیر باعث کاهش مصرف انرژی می‌شود.دلیلش این است که به دلیل پهنای باند انعطاف پذیر احتمال تداخل بسته و در نتیجه ارسال چند بارهٔ بسته کاهش می‌یابد.همچنین در این MAC ملاحظاتی در ارتباط با زمان خوابِ[83] گره در نظر گرفته شده است.پروتکل MAC دیگری در [31] به نام medmac عنوان شده است که برای کاهش مصرف انرژی، تغییراتی را در نحوهٔ همگام سازی گره‌ها لحاظ کرده است.پایه و اساس هر دو پروتکل بیان شده، TDMA می‌باشد.

۸- بحث و گفتگو

در این بخش مروری بر مسائلی که هنوز باید تحقیقات و مطالعات بیشتری در ارتباط با آنها صورت پذیرد خواهیم داشت و چندین راه حل پیشنهاد می‌شود.

۸-۱- بهبود طول عمر باتری

سلول‌های سوختی کوچک[84] انرژی زیاد و حجیمی را فراهم می‌کنند و برای شارژ کردن آنها فقط کافی است کارتریج آن عوض شود.این ویژگی، سلول‌های سوختی را برای کاربردهای قابل حمل مثل WBAN جذابتر می‌سازد.به کار بردن انرژی خورشیدی و همچنین استفاده از گرما ولرزش بدن نیز روشی برای بهبود عمر باتری است.در [32] یک حسگر بی‌سیم که خودش به تنهایی انرژی لازم را از گرمای بدن انسان بدست می‌آورد، طراحی شده است.همچنین امکان به کارگیری تکنیک‌هایی مثل شارژ باتری از راه دور وجود دارد.اخیرا محققان در دانشگاه MIT گزارش داده‌اند امکان انتقال انرژی بی‌سیم در یک محدودهٔ کوتاه (چند متر) با استفاده از امواج ناپایدار امکان پذیر است.

۸-۲- پایین آوردن مصرف انرژی

بیشتر انرژی مصرفی به ارتباطات بی‌سیم تخصیص داده می‌شود.یک راه حل برای کاهش مصرف انرژی استفاده باند UWB است زیرا هم نرخ داده‌ای بالایی دارد و هم مصرف انرژی کمی دارد.استفاده از الگوریتم‌های فشرده سازی داده‌ها نیز مفید است زیرا با استفاده از این الگوریتم‌ها تعداد بیت‌هایی که باید ارسال شود کاهش می‌یابد.

۸-۳- مزاحم نبودن سنسورها

مولفهٔ اصلی در بزرگی و وزن یک حسگر، باتری آن می‌باشد بنابراین روش‌هایی که که اندازهٔ باتری را کاهش می‌دهند مثل پذیرش تکنولوژی سلول سوختی، می تواند پتانسیل زیادی برای ساختن حسگرهایی که مزاحمت کمتری دارند، ایجاد کند.همچنین استفاده از تکنولوژی‌هایی که سطح بالاتری از مجتمع شدن را فراهم می‌کند می‌تواند در کاهش اندازه حسگرها کمک کند.

۸-۴- سیستم‌های بهداشت و درمان پیشگیرانه

تکنولوژی‌های شبکه‌های بدنی موجود اغلب بر مبنای تقاضا[85] توسعه داده شده است.این WBAN ها در برابر اتفاقات فیزیلوژی بدن پاسخی را می‌دهند.با استفاده از برنامه ریزی پویا محیط و شناخت رابطه‌ها، ما قادر خواهیم بود تا با اندازه گیری چندین پارامتر در بدن انسان و استفاده از این داده ها در پیشگیری از بیماری و تشخیص بیماری کمک کنیم.

۸-۵- مسیریابی[86]

همان طور که توضیح داده شد به کار گیری توپولوژی‌های چند پرشی در مقابل توپولوژی ستاره قابلیت اطمینان بیشتری فراهم می‌کند وهمچنین انرژی کمتری مصرف می‌کند.در صورت بکارگیری توپولوژی چند پرشی مفهوم مسیریابی مطرح می‌شود.هدف، طراحی پروتکل‌هایی است که انرژی را به طور موثر به کار گیرد و یک ارتباط مطمئن بین دو دستگاه انتها به انتها[87] را فراهم کند.

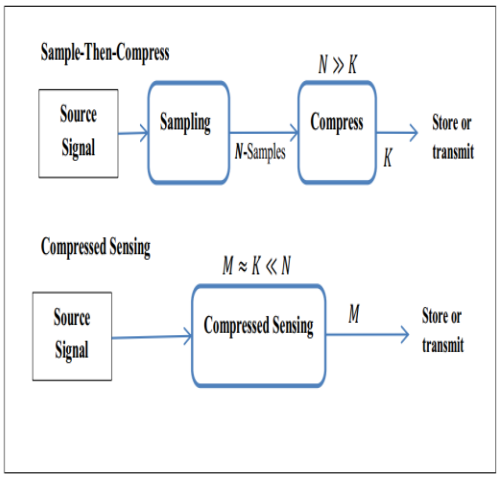
۹- طرح پیشنهادی

هدف از انجام پروژه پایانی به کارگیری روش‌های امنیتی دقیق به منظور محافظت در برابر انواع حملات می‌باشد.اهمیت و راهکارهای امنیتی مناسب در قسمت‌های قبلی مورد بحث قرار گرفت.یکی از اصلی ترین چالش‌ها در بکارگیری روش‌های امنیتی میزان مصرف انرژی است که گره‌ها باید مد نظر داشته باشند.روش‌های مختلفی برای کاهش مصرف انرژی در WBAN در نظر گرفته شده است.یک دسته از این روش‌ها بر پروتکل کنترل دسترسی رسانه (MAC) تمرکز دارد و تلاش شده است رویه هایی را در MAC به کار گیرند که انرژی را به طور موثر استفاده کند.برخی دیگر از تحقیقات به منظور کاهش مصرف انرژی به بررسی توپولوژی شبکه پرداخته‌اند.در این تحقیقات نشان داده شده که بجای توپولوژی ستاره که یک انتخاب طبیعی و پیش فرض برای شبکه‌های بی‌سیم بدنی است می‌توان از انواع توپولوژی چند پرشی به منظور کاهش مصرف انرژی استفاده کرد. یکی دیگر از روش‌های که در سال‌های اخیر به آن توجه ویژه‌ای شده است،به کار گیری نظریه نمونه برداری فشرده[88] (CS) است. این نظریه باعث کاهش حجم داده شده و بالطبع مقدار انرژی مصرفی نیز تا حد قابل توجهی کاهش می‌یابد.این نظریه می‌گوید به تعداد کمی از اندازه گیری‌های خطی تصادفی، برای جمع آوری، پردازش، انتقال و بازیابی سیگنال نیاز است[33,34] .در [35] نشان داده شده است با بکار گیری نظریه CS می‌توان تا ۳۵ درصد در مصرف انرژی صرفه جویی کرد بدون اینکه میزان قابلیت اطمینان یا دسترس پذیری دچار افت شود. این آزمایش بر روی سیگنال‌های ECG صورت پذیرفته است. در واقع CS یک روش جدید فشرده سازی و کدینگ محسوب می‌شود.در روش های کلاسیک فشرده سازی، سیگنال به طور کامل

اخذ شده (n- نمونه) و سپس ضرایب[۸۹] مربوط به تبدیل مورد نظر بدست می‌آید.بعد از اعمال تبدیل مورد نظر، k ضریب بزرگتر حفظ شده و بقیه ضرایب (n-k) دور ریخته می‌شوند.سپس این k مقدار به همراه مکان‌های مربوطه در سیگنال اصلی کد می‌شوند.بکار گیری این روش (نمونه برداری - سپس - فشرده سازی[۹۰]) سه مشکل ذاتی دارد:

- ما تعداد n نمونه را ذخیره می‌کنیم (ممکن است n مقدار بزرگی باشد) در حالی که ممکن است k مقدار خیلی کوچکی باشد.
- کدگذار باید ضرایب مربوط به دامنه تبدیل[۹۱] را برای هر n- نمونه حساب کند با وجودی که فقط به ضرایب k- مقدار نیاز دارد
- کد کردن مکانِ مقادیرِ قابل توجه، سربار را روی کد گذار ایجاد می‌کند.

نمونه برداری فشرده با حذف مرحله میانی اخذ n- نمونه، به صورت مستقیم سیگنال را به نمایش فشرده آن در دامنه تنک[۹۲] ساز مربوطه تبدیل می‌کند.در شکل (۱۰) تفاوت بین روش‌های نمونه برداری کلاسیک و روش نمونه برداری فشرده به تصویر کشیده شده است.



شکل (۱۰) : مقایسه روش‌های نمونه برداری کلاسیک و نمونه برداری فشرده

از طرف دیگر روشهای کلاسیک نمونه برداری سیگنال‌ها، از نرخ نمونه برداری شانون استفاده می‌کنند ، که بیان می‌کند: برای بازسازی مناسب سیگنال، نرخ نمونه برداری باید حداقل دو برابر حداکثر فرکانس (f_{max}) موجود در سیگنال باشد (نرخ نایکویست)[36]. تئوری نمونه برداری فشرده اثبات می‌کند که، سیگنال‌های معینی را می‌توان با تعداد نمونه‌های کمتری نسبت به انچه در تئوری شانون بیان شده است، باز سازی کرد[37].

ما در این پروژه قصد داریم از راه کارهای امنیتی قوی برای محافظت در برابر حملات استفاده کنیم و از آنجایی که این راهکارهای امنیتی قوی(مثلا رمز نگاری کلید عمومی) مصرف انرژی در گره‌ها را افزایش می‌دهد، با به کارگیری تئوری نمونه برداری فشرده در سیگنال‌های پزشکی میزان مصرف انرژی را جبران می‌کنیم.

۱۰- نتیجه

در این سمینار ابتدا تکنولوژی شبکه‌های بدنی بی‌سیم را معرفی و برخی از ویژگی‌های آن را بیان کردیم.سپس بستری برای پیاده سازی این تکنولوژی جهت کاربردهای پزشکی توصیف شد که شامل سه نوع زیر ساخت برای WBAN است که عبارتنداز: شبکه‌های بدنی مدیریت شده، خود مختار و هوشمند.انواع کاربردهای WBAN نیز به طور خلاصه ذکر گردید.در ادامه معماری‌های رایج برای WBAN بحث شد و با وجود اینکه توپولوژی ستاره به طور طبیعی برای این نوع شبکه‌ها به کار می‌رود نشان داده شد که این توپولوژی از لحاظ مصرف انرژی و برقراری ارتباط مطمئن چندان کارآمد نیست و در عوض توپولوژی چند پرشی عملکرد بهتری دارد.IEEE 802 یک گروه کاری به نام IEEE 802.15.6 را برای استاندارد کردن WBAN مشخص کرد.هدف این گروه ایجاد یک استاندارد به منظور برقراری یک ارتباط بی‌سیم برای محدودهٔ کوتاه با توان کم و قابلیت اطمینان بالا بود که در نزدیکی یا داخل بدن انسان بتواند کار کند.که در این سمینار به مرور این استاندارد پرداختیم.به دلیل ماهیتِ اطلاعات رد و بدل شده در WBAN، امنیت یکی از اساسی ترین موضوعات این شبکه‌ها است.امنیت شامل سه فرآیند رمزنگاری، احراز هویت و تصحیح خطای انتقال است که در این سمینار بررسی شد.در آخر هم چالش‌های مهم در این نوع تکنولوژی که از مهمترین آن می‌توان به مصرف انرژی اشاره کرد، بیان گردید.و البته با توجه با جدید بودن این نوع تکنولوژی مسائل بسیاری وجود دارد که باید بر روی آن مطالعات و تحقیقات بیشتری صورت پذیرد.در آخر سمینار برخی از این مسائل به طور گذرا بحث شد.

مراجع

[1] Guang Zhong Yang (Ed.)., *" Body sensor network ©"* Springer-Verlag London Limited 2006

[2] Sana ollah, Pervez khan, shahnaz saleem and kyung sup kwak., *"A Review of Wireless Body Area Network for Medical Application "* , Int'l J. of Communications, Network and System Sciences, Vol. 2 No. 8, 2009, pp. 797-803. doi: 10.4236/ijcns.2009.28093.

[3] Chin, C.A.; Crosby, G.V.; Ghosh, T.; Murimi, R. *"Advances and challenges of wireless body area networks for healthcare applications "* , Computing, Networking and Communications (ICNC), 2012

[4] Jamil. Y. Khan and Mehmet R. Yuce., *"Wireless Body Area Network (WBAN) for Medical Applications "*,

[17] " *IEEE Standard for Local and metropolitan area networks Part 15.6: Wireless Body Area Networks*" 29 February 2012

[18] Dr. Shinyoung Lim , Dr. Tae Hwan Oh , Dr. Young B. Choi, Dr. Tamil Lakshman., "*Security Issues on Wireless Body Area Network for Remote Healthcare Monitoring* " ,2010 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing

[19] M. Somasundaram and R. Sivakumar, " *Security in Wireless Body Area Networks: A survey* " , 2011 International Conference on Advancements in Information Technology With workshop of ICBMG 2011

[20] Wassim Drira, Eric Renault and Djamal Zeghlache., " *A Hybrid Authentication and Key Establishment Scheme for WBAN "* 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications.

[21] S. Saleem, S. Ullah, and K.S. Kwak, " *A Study of IEEE 802.15.4 Security Framework for Wireless Body Area Networks"*, Sensors, vol.11, No.2, pp. 1383-1395, 2011.

[22] Pandolfino JE, Ritcher JE, Ours T, Jason RN, Guardino M, Chapman J, et al. " *Ambulatory esophageal pH monitoring using a wireless system*" . American Journal of Gastroenterology 2003; 98(4):740-749.

[23] Prem Chand Jain, " *wireless body area network for health care*" , November 11, 2012

[24] Scanlon WG, Evans NE. " *Radiowave propagation from a tissue-implanted source at 418MHz and 916.5MHz* ". IEEE Transactions on Biomedical Engi-neering 2000; 47(4):527-534.

[25] Yusuke Hamada,Kenichi Takizawa, and Tetsushi Ikegami., *" Highly reliable wireless body area network using error correcting codes* " , in 2011

[26] Hind Chebbo, Saied Abedi , haraka A. Lamahewa, David B. Smith, Dino Miniutti and Leif Hanlen., *"Reliable Body Area Networks Using Relays: Restricted Tree Topology* " IEEE ٢٠١٢

[27] Mrinmoy Barua, M.S.Alam, Xiaohui liang Student Member IEEE, and Xuemin shen., " *Secure and Quality of Service Assurance Scheduling Scheme for WBAN with Application to eHealth"* IEEE WCNC 2011-Network

[28] D. Vergados, D. Vergados, and I. Maglogiannis, "*Ngl03-6: Ap-plying wireless diffserv for qos provisioning in mobile emer-gency telemedicine,*" in Global Telecommunications Conference, 2006. GLOBECOM '06. IEEE, nov. 2006, pp. 1 –5.

[29] M.A. Ameen, Ahsanun Nessa, Kyung Sup Kwak, " *QoS issues with focus on Wireless Body Area Networks* ", Third 2008 International Conference on Convergence and Hybrid Information Technology

[30] Gengfa, Fang; Dutkiewicz, E., " *BodyMAC: Energy efficient TDMA-based MAC protocol for Wireless Body Area Networks*", IEEE, Communications and Information Technology, 2009. ISCIT 2009. 9th International Symposium.

[31] Timmons, N.F.; Scanlon, W.G., "*An adaptive energy efficient MAC protocol for the medical body area network*". Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on Topic(s): Aerospace ; Communication, Networking & Broadcasting ; Computing & Processing (Hardware/Software) ; Signal Processing & Analysis ; Transportation.

[32] V.Leonov , et al. " *Thermoelectric Converters of human Warmth for Self-Powered Wireless Sensor Node*, " IEEE Sensors Journal, vol.7,pp.650-657, 2007

January 1, 2010 under CC BY-NC-SA 3.0 license, in subject Biomedical Engineering

[5] Barakah, D.M.; Ammad-uddin M., "*A Survey of Challenges and Applications of Wireless Body Area Network (WBAN) and Role of a Virtual Doctor Server in Existing Architecture*" , Intelligent Systems, Modelling and Simulation (ISMS), 2012 Third International Conference
On topic(s): communication, networking & broadcasting;computing & processing

[6] Arif Onder ISIKMAN, Loris Cazalon,feiquan Chen, peng li., "*body area network "*, 2010

[7] Kwak, K.S., *"An overview of IEEE 802.15.6 standard* " , Applied Sciences in Biomedical and Communication Technologies (ISABEL), 2010 3rd International Symposium on

[8] J. A. Ruiz and S. Shimamoto, "Novel Communication Services Based on Human Body and Environment Inter-action: Applications inside Trains and Applications for Handicapped People," Proc. IEEE WCNC 2006, Las Vegas, NV, 2006.

[9] Anirudh Natarajan, Mehul Motani, Buddhika de Silva, Kok-Kiong Yap & K. C. Chua,. " *Investigating network architectures for body sensor networks"* , Published in Proceeding HealthNet '07 Proceedings of the 1st ACM SIGMOBILE international workshop on Systems and networking support for healthcare and assisted living environments in 2007

[10] Yu Ge; Liang Liang; Wei Ni; Aung Aung Phyo Wai; Gang Feng., *" A measurement study and implication for architecture design in wireless body area networks* ", Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference .

[11] Reusens, E.; Joseph, W.; Latre, B.; Braem, B.; Vermeeren, G.; Tanghe, E.; Martens, L.; Moerman, I.; Blondia.*" Characterization of On-Body Communication Channel and Energy Efficient Topology Design for Wireless Body Area Networks* ",in 2009

[12] Natarajan, A.; de Silva, B.; Kok-Kiong Yap; Motani, M., ." *To Hop or Not to Hop: Network Architecture for Body Sensor Networks* ", Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON '09. 6th Annual IEEE Communications Society Conference

[13] Shankar, V.; Natarajan, A.; Gupta, S.K.S.; Schwiebert, L.," *Energy-efficient protocols for wireless communication in biosensor networks"*, Personal, Indoor and Mobile Radio Communications, 2001 12th IEEE International Symposium on

[14] Changhong Wang; Qiang Wang; Shunzhong Shi. , " *A distributed wireless body area network for medical supervision "* , Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE *International* .

[15] Md.Asdaque Hussain and Kyung Sup Kwak., "*Positioning in Wireless Body Area Network using GSM "*, International Journal of Digital Content Technology and its Applications Volume 3, Number 3, September 2009

[16] Saeed Rashwand, Jelena Mi˘ si´ c, and Hamzeh Khazaei., *" IEEE 802.15.6 under Saturation : some problems to Be Expected "*, Journal of communications and Networks, VOL. 13, NO. 2, APRIL 2011

[33] Donoho, D.L. " *Compressed Sensing* " , Information Theory, IEEE Transactions on , Publication Year: 2006 , Page(s): 1289 - 1306
[34] Justin Romberg, Michael Wakin., "*Compressed Sensing: A Tutorial* ", IEEE Statistical Signal Processing Workshop Madison, Wisconsin August 26, 2007
[35] Mohammadreza Balouchestani, Kaamran Raahemifar, and Sridhar Krishnan., "*Low Power Wireless Body Area Networks with Compressed Sensing Theory* " ,Circuits and Systems (MWSCAS), 2012 IEEE 55th International Page(s): 916 – 919,
[36] ANDREW S. TANENBAUM., *" Computer Network* " , © 2003 Pearson Education, Inc. Publishing as Prentice Hall PTR Upper Saddle River, New Jersey 0745
[37] Yihong Wu, " Shannon Theory for Compressed Sensing " September 2011
[38] Ramli, S.N.; Ahmad, R.; , "*Surveying the Wireless Body Area Network in the realm of wireless communication*," Information Assurance and Security (IAS), 2011 7th International Conference on , vol., no., pp.58-61, 5-8 Dec. 2011

زیرنویس‌ها

[1] Wireless Body Area Network
[2] Medical
[3] Healthcare
[4] sensor
[5] On-body
[6] In-body
[7] Medium Access Control
[8] Miniaturization
[9] Monitor
[10] Coordinator
[11] Wireless Sensor Network
[12] Energy consumption
[13] Infrastructure
[14] Actuator
[15] Implant
[16] Wearable
[17] Electroencephalography
[18] Electrocardiography
[19] Personal digital assistant
[20] Management WBAN
[21] Alert
[22] Autonomous WBAN
[23] Intelligent WBAN
[24] Military
[25] Games
[26] Social network
[27] Data rate
[28] Sink
[29] Star
[30] Multi hop
[31] Performance
[32] Robustness
[33] Delay
[34] Interference
[35] Compatibility
[36] Relay
[37] Reliability
[38] Efficiency consumption
[39] Packet Delivery Ratio
[40] Average Number Retransmission
[41] Dynamic
[42] Personal Area Network
[43] Quality of Service
[44] Medical Implant Communication Service
[45] License
[46] Wireless Medical Telemetry Service
[47] Industrial Science Medical
[48] Authentication
[49] Encryption
[50] Narrow Band
[51] Ultra Wide Band
[52] Human Body Communication
[53] Physical Protocol Data Unit
[54] Preamble
[55] Physical Layer Convergence Procedure
[56] Physical Service Data Unit
[57] Synchronization Header
[58] Physical Header
[59] Start Frame Delimiter
[60] Low band
[61] High band
[62] Beacon
[63] Exclusive Access Phase
[64] Random Access Phase
[65] Contention Access Phase
[66] Slotted Aloha
[67] Uplink
[68] Downlink
[69] Polling
[70] Symmetric
[71] Public key
[72] Confidentiality
[73] Integrity
[74] Availability
[75] Biological
[76] Feedback
[77] Clinical
[78] Overhead
[79] LOBY Transform
[80] Restrict Tree Topology
[81] Real time
[82] Mobility
[83] Sleep time
[84] Micro-fuel
[85] On demand
[86] Routing
[87] End to End
[88] Compressed Sensing
[89] Coefficients
[90] Sample-Then-Compress
[91] Transform Domain
[92] spars

معماری شبکه های رادیوی شناختی سلولی

پروین عباسی[۱]، رضا برنگی[۲]

[۱] گروه سخت افزار دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت
تهران، ایران
pabbasi@iust.ac.ir

[۲] استادیار دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت
تهران، ایران
rberangi@iust.ac.ir

چکیده

افزایش کاربران و تنوع سرویس های بی سیم سبب افزایش تقاضابرای طیف فرکانسی شده است. از طرفی طیف فرکانسی موجود قادر به پاسخگویی به نیاز های جدید نیستند. بنابراین مفهوم رادیو شناختی برای حل مشکل کمبود طیف مطرح شده است. شبکه های رادیو شناختی دارای معماری های متنوعی هستند و اخیرا از معماری سلولی برای گسترش پوشش و توسعه سرویس ها در این نوع شبکه ها استفاده می شود. از آنجاییکه طراحی الگوریتم ها و پروتکل های لایه های مختلف شبکه بدون در نظر گرفتن معماری شبکه امکان پذیر نیست، لذا در این سمینار ضمن بررسی معماری های شبکه های رادیو شناختی به مرور کارهای انجام شده در این زمینه و معماری مورد استفاده آنها پرداخته و زمینه های باز تحقیق در این حوزه شناسایی و معرفی شده است.

کلمات کلیدی

رادیو شناختی، معماری سلولی، شبکه های دارای زیر ساخت، دسترسی پویا به طیف

۱- مقدمه

افزایش کاربران و تنوع سرویس های بی سیم سبب افزایش تقاضا برای طیف فرکانسی شده است. بیشتر طیف فرکانسی در دسترس، به شبکه های بی سیم موجود اختصاص یافته است. لذا تنها بخش کوچکی از طیف فرکانسی می تواند به کاربردهای بی سیم جدید تخصیص یابد و باندهای فرکانسی موجود قادر به پاسخگویی به نیازهای جدید نیستند. بنابراین در حال حاضر کمبود طیف فرکانسی به مسأله ای مهم تبدیل شده است . باندهای فرکانسی در همه ی کشور ها توسط دولت تنظیم می شود و تخصیص فرکانس یا تخصیص طیف نامیده می شود . براین اساس طیف فرکانسی به دو بخش تقسیم می شود:

- باند های فرکانسی مجوزدار: این باند های فرکانسی توسط اپراتور ها خریداری می شود و تنها کاربران این شبکه ها حق ارسال در آن را دارند. شبکه های سلولی نظیر GSM* و WIMAX† نمونه هایی از این نوع شبکه ها هستند.
- باند های فرکانسی آزاد یا ISM‡ : باند های فرکانسی از ۲,۴ تا ۵ گیگاهرتز از این نوع می باشند. امکان ارسال در این باندهای فرکانسی برای همه ی کاربران بدون نیاز به مجوز، به شرطی که محدودیت های مشخصی مانند توان را رعایت نمایند، مهیا است. شبکه های بی سیم محلی(Bluetoothو WLAN [۱])از جمله شبکه های بی سیمی هستند که از این باندهای فرکانسی استفاده می نمایند.

بررسی های FCC [۲] نشان داده است طیف فرکانسی مجوزدار به طور کامل مورد استفاده قرار نمی گیرد . بر اساس گزارشی که FCC در سال ۲۰۰۲ منتشر کرده است، میزان بهره برداری از طیف بسته به زمان و مکان بین ۱۵ الی ۸۵ درصد می باشد[۱]. محدودیت طیف دردسترس و راندمان پایین استفاده از طیف، ضرورت معرفی الگویی

* Global System for Mobile communication
† Worldwide Interoperability for Microwave Access
‡ Industrial,Scientific and Medical

جدید برای بهره بردن از طیف موجود به صورت فرصت طلبانه را پیشنهاد داده است برای افزایش ظرفیت و در نتیجه بیش تر شدن سودمندی طیف به کاربران دیگر اجازه داده شود هنگامی که کاربران مجوز دار از طیف استفاده نمی کنند از این فرصت های طیفی بهره ببرند. فرصت های طیفی یا حفره های طیفی به باندهای فرکانسی گفته می شود که به کاربران مجوزدار اختصاص یافته است،اما در بعضی زمان ها یا مکان ها مورد استفاده قرار نگرفته اند و بنابراین می توانند توسط کاربران دیگر مورد بهره برداری قرار گیرند.

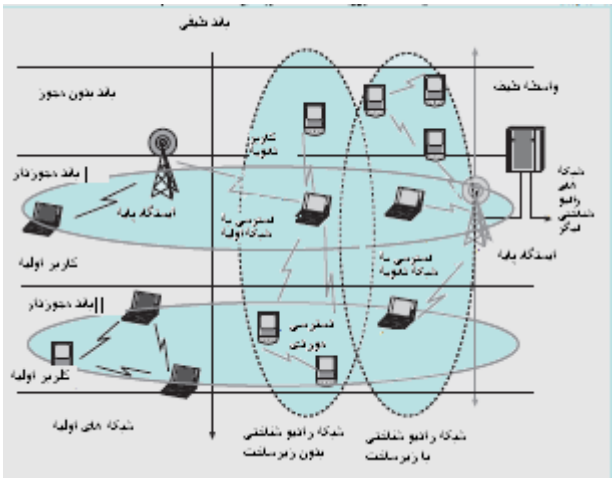
با توجه به این که حفره های طیفی در طول زمان متغیر هستند، سیستم هایی که می خواهند از این فرصت های طیفی استفاده نمایند باید دارای انعطاف پذیری برای دسترسی پویا به طیف باشند. سیستم های مخابراتی بی سیم معمول، برای فعالیت در باند فرکانسی مشخصی طراحی شده اند. لذا مفهوم رادیو شناختی[۳] معرفی شده است. رادیو شناختی با به اشتراک گذاری طیف فرکانسی بین کاربران اولیه که مجوز استفاده از آنرا دارند و کاربران ثانویه که بدون مجوز هستند، به عنوان فناوری کلیدی در [۳],[۲] برای حل مشکل کمبود طیف مطرح شده است. در واقع، یک رادیو شناختی به عنوان رادیویی که پارامترهای ارسال خود را بر اساس تعامل با محیط اطرافش تغییر می دهد، تعریف می شود. رادیو شناختی یک واحد خود مختار در محیط های ارتباطی است که معمولا با شبکه هایی که به رادیو شناختی های دیگر دسترسی دارند،اطلاعات مبادله می کند. به عبارت دیگر، یک رادیو شناختی، یک رادیو مبتنی بر نرم افزار[۴] است در حالی که یک رادیو دیجیتال هم هست [۴]. به مجموعه ای از رادیو شناختی ها که با یکدیگر ارتباط دارند، شبکه رادیو شناختی گفته می شود.

۲- معماری شبکه رادیو شناختی

برای توسعه و ایجاد پروتکل های ارتباطی رادیو شناختی که چالش های دسترسی پویا به طیف را کنترل می کنند، توصیف دقیق و جامع معماری شبکه رادیو شناختی ضروری است. در این بخش از مقاله معماری شبکه رادیو شناختی ارائه می شود.

۲-۱- مولفه شبکه

مولفه های معماری شبکه رادیو شناختی نشان داده شده در شکل (۱) به دو دسته تقسیم می شوند: شبکه اولیه (یا شبکه مجوزدار) و شبکه رادیو شناختی (بعلاوه به آن شبکه ثانویه، شبکه با دسترسی پویا به طیف، شبکه بدون مجوز هم گفته می شود). شبکه اولیه به یک شبکه موجود اشاره دارد که در آن کاربران اولیه مجاز به استفاده از باند طیفی مشخصی هستند. اگر شبکه های اولیه دارای زیر ساخت[۵] باشند، فعالیت های کاربران اولیه از طریق ایستگاه های پایه[۶] کنترل می شوند و به علت اولویت آنها در دسترسی به طیف، عملیات کاربران اولیه نباید تحت تاثیر کاربران ثانویه(یا بدون مجوز) قرار گیرد. شبکه رادیو شناختی مجوز استفاده از باند طیفی مورد نظر را ندارد. از اینرو، کاربران رادیو شناختی بایستی قابلیت به اشتراک گذاری طیف مجوزدار را داشته باشند. شبکه های رادیو شناختی نیز می توانند به ایستگاه های پایه مجهز باشند تا اتصال تک پرشی[۷] برای کاربران رادیو شناختی فراهم کنند. نهایتا، شبکه های رادیو شناختی می توانند شامل واسطه طیف[۸] باشند که نقش توزیع کننده منابع طیفی بین شبکه های رادیو شناختی مختلف را بازی می کند[۵] .



شکل (۱): معماری شبکه های رادیو شناختی

۲-۲- ناهمگنی طیف

کاربران رادیو شناختی قابلیت دسترسی به هر دو بخش دارای مجوز(که توسط کاربران اولیه استفاده می شود) و بدون مجوز طیف را دارند (دسترسی به این بخش طیف از طریق فناوری دسترسی باند وسیع امکان پذیر است). پس، عملکرد شبکه های رادیو شناختی به دو دسته تقسیم می شود: عملکرد در باند دارای مجوز و عملکرد در باند بدون مجوز.

- عملکرد در باند دارای مجوز: باند دارای مجوز اساسا توسط شبکه اولیه استفاده می شود. بنابراین، شبکه های رادیو شناختی در این مورد عمدتا روی شناسایی کاربران اولیه تمرکز دارند. ظرفیت کانال به تداخل با کاربران اولیه مجاور بستگی دارد. همچنین، اگر کاربران اولیه در باند طیفی که توسط کاربران رادیو شناختی اشغال شده است، پدیدار شوند، کاربران رادیو شناختی بایستی باند طیفی را رها کرده و فورا به باند طیفی دیگری منتقل شوند.
- عملکرد در باند بدون مجوز: در غیاب کاربران اولیه، کاربران رادیو شناختی حق دسترسی به طیف دارند. لذا، روش های کارآمد اشتراک گذاری طیف برای کاربران رادیو شناختی به جهت رقابت در باند بدون مجوز لازم است.

۲-۳- ناهمگنی شبکه

چنانچه در شکل (۱) نشان داده شده، کاربران رادیو شناختی ۳ نوع دسترسی فرصت طلبانه متفاوت اجرا می کنند:

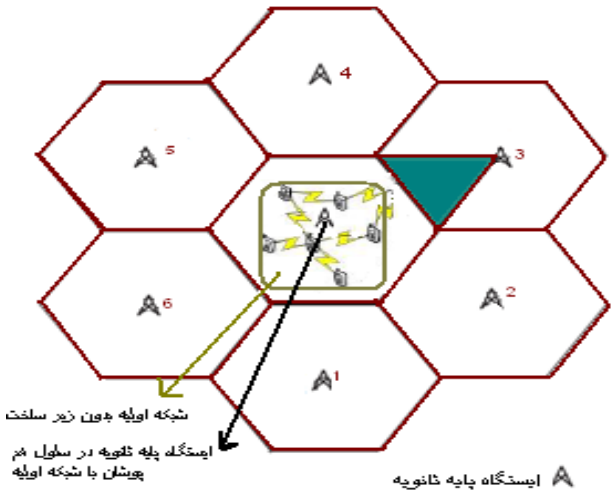
- دسترسی به شبکه رادیو شناختی: کاربران رادیو شناختی به ایستگاه پایه خودشان در هر دو باندهای طیفی دارای مجوز و بدون مجوز دسترسی دارند، چون همه ی تعاملات ، داخل شبکه رادیو شناختی اتفاق می افتد، پس سیاست اشتراک گذاری طیفی آنها مستقل از سیاست اشتراک گذاری طیف در شبکه اولیه است.
- دسترسی موردی[۹] رادیو شناختی: کاربران رادیو شناختی با دیگر کاربران رادیو شناختی از طریق اتصال موردی در هر دو باندهای طیفی دارای مجوز و بدون مجوز ارتباط دارند.
- دسترسی به شبکه اولیه: کاربران رادیو شناختی همچنین می توانند به ایستگاه پایه اولیه از طریق باند دارای مجوز دسترسی داشته باشند. بر خلاف انواع دسترسی های دیگر، کاربران رادیو شناختی نیازمند پروتکل کنترل دسترسی به رسانه[۱۰] سازگار هستند تا قادر به گردش بین چندین شبکه اولیه با فناوری های دسترسی متفاوت باشند.

شکل (۱) ، قابلیت های مختلف مورد نیاز برای پشتیبانی مدیریت طیف در شبکه های رادیو شناختی را نشان می دهد.

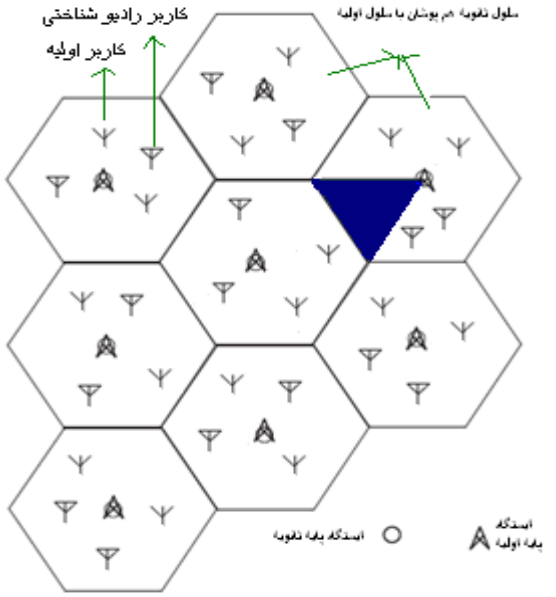
۳- معماری شبکه های رادیو شناختی سلولی

در شبکه های رادیو شناختی برای حصول اهدافی همچون پوشش بیشتر و افزایش ظرفیت و گذردهی نیازمند استفاده از معماری سلولی هستیم. در معماری سلولی ارتباط بین سلول های مختلف از طریق ایستگاه های پایه برقرار می شوداز اینرو شبکه های رادیو شناختی سلولی در هر سلول خود مجهز به یک ایستگاه پایه یا نقطه دسترسی بوده و دارای زیر ساخت می باشند. فایده ساختار سلولی در اینست که اگر در یک سلول، ایستگاه پایه فرصت های طیفی شبکه اولیه مجاور خود را شناسایی کند، این فرصت های طیفی بین همه ایستگاه های پایه ثانویه (ایستگاه های پایه در شبکه رادیو شناختی)به اشتراک گذاشته می شود و با استفاده از مفهوم بازیافت فرکانسی در حوزه زمان و مکان می توان همان فرصت های طیفی محدود را بین کاربران رادیو شناختی در سلول های مختلف و با تکنیک تقسیم زمانی به اشتراک گذاشت. در نتیجه، ظرفیت بیشتر و پوشش گسترده تر نسبت به حالت غیر سلولی بدست می آید. اما از طرفی، در معماری رادیو شناختی سلولی با چالش های بیشتری در مقایسه با معماری غیر سلولی مواجه هستیم. زیرا به مکانیزم های تخصیص منبع و کنترل تداخل سازگار با معماری سلولی نیاز داریم تا تداخل بین کاربران ثانویه در سلول های هم فرکانس(سلول هایی که از همان باند های طیفی به طور همزمان استفاده می کنند و در رنج تداخل یکدیگر قرار دارند) و همچنین تداخل کاربران ثانویه با کاربران اولیه را کنترل نمائیم و اجازه ندهیم میزان تداخل از آستانه قابل تحمل فراتر رود(این آستانه قابل تحمل در موقعیت های جغرافیایی متفاوت و بر حسب فاصله سلول های رادیو شناختی هم فرکانس از یکدیگر و تاثیر فاکتور افت مسیر[۱۱]، تاثیر محوشدگی[۱۲] و نیز تاثیر سایه اندازی[۱۳] محیط متفاوت خواهد بود).

در بخش ۲-۱- گفتیم معماری شبکه های رادیو شناختی دارای دو مولفه است: شبکه اولیه و شبکه رادیو شناختی. پس در معماری شبکه های رادیو شناختی سلولی، ساختار یا معماری شبکه اولیه نیز مهم است. شبکه اولیه می تواند دارای زیر ساخت(یعنی فعالیت کاربران اولیه از طریق ایستگاه های پایه کنترل شود)یا بدون زیر ساخت باشد. پس اگر شبکه اولیه بدون زیر ساخت باشد، معماری شبکه های رادیو شناختی سلولی به صورت شکل (۲) خواهد بود و اگر شبکه اولیه دارای زیر ساخت باشد، معماری شبکه های رادیو شناختی سلولی به صورت شکل (۳) خواهد بود(یعنی خود شبکه اولیه نیز دارای معماری سلولی می باشد). در این معماری هر سلول اولیه با یک سلول رادیو شناختی همپوشنای دارد و با بکارگیری مدولاسیون OFDM در شبکه رادیو شناختی، زیر حامل هایی که در سلول اولیه توسط کاربران اولیه استفاده نمی شوند در سلول رادیو شناختی همپوشان با آن، بین کاربران رادیو شناختی به اشتراک گذاشته می شود. بنابراین بین یک سلول اولیه با یک سلول رادیو شناختی همپوشان با آن در حالت ایده آل تداخلی وجود ندارد(یعنی با پدیدار شدن کاربر اولیه در زیر حامل اشغال شده توسط کاربر رادیو شناختی، بلافاصله کاربر رادیو شناختی این زیر حامل را رها کرده و به زیر حامل دیگری منتقل می شود و بنابراین، تداخلی با کاربر اولیه ایجاد نخواهد کرد.این حالت مطلوب با حس کردن دقیق و قابل اطمینان طیف توسط کاربر رادیو شناختی امکان پذیر خواهد بود).اما بین یک سلول اولیه با سلول های اولیه و سلول های رادیو شناختی هم فرکانس موجود در حلقه(تایر) اول تداخل وجود دارد و همین حالت برای سلول های رادیو شناختی نیز وجود دارد و هر سلول رادیو شناختی دو نوع تداخل تحمل می کند: یکی با سلول های اولیه هم فرکانس مجاور و دیگری با سلول های رادیو شناختی هم فرکانس مجاور. از اینرو به مکانیزم ها و الگوریتم های تخصیص منبع جدید منعطف با این معماری نیازمندیم تا ضمن اشتراک گذاری طیف مجوزدار بین کاربران اولیه و کاربران ثانویه، نیازهای کیفیت سرویس شبکه اولیه و شبکه رادیو شناختی را تامین نمائیم.

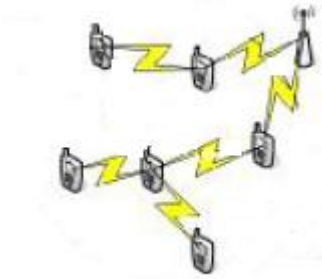


شکل (۲): معماری شبکه رادیو شناختی سلولی با شبکه اولیــه بــدون زیر ساخت

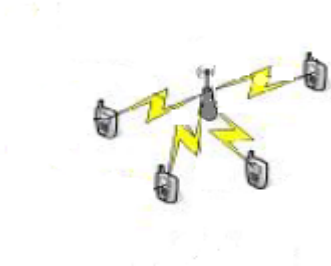


شکل (۳): معماری شبکه رادیو شناختی سلولی با شبکه اولیــه دارای زیرساخت

از آنجایی که، رادیو شناختی از دو روش برای اشتراک گذاری طیف بین کاربران اولیه و ثانویه استفاده می کند تا بهره وری طیف شبکه اولیه و ظرفیت و پهنای باند شبکه رادیو شناختی را افزایش دهد. بنابراین، کاربران در سلول رادیو شناختی با ایستگاه پایه به صورت تک پرشی یا چند پرشی[۱۴]می توانند ارتباط برقرار کنند. چنانچه در شبکه رادیو شناختی از دسترسی زیر گستر برای اشتراک گذاری طیف استفاده شود، ارتباط کاربران در سلول رادیو شناختی با ایستگاه پایه به صورت چند پرشی خواهد بود. شکل (۴) زیرا، در این روش توان ارسالی کاربر ثانویه محدود می شود تا امکان همزیستی کاربران اولیه و ثانویه در باند فرکانسی یکسان فراهم گردد. در واقع، در این روش با کم کردن توان ارسال کاربر ثانویه سعی می شود از میزان تداخل کاسته و ارسال کاربر ثانویه برای کاربر اولیه مانند نویز تلقی شود و به علت کاهش توان ارسال، امکان اتصال تک پرشی بین ایستگاه پایه و کاربر ثانویه وجود ندارد. یکی از روش های ممکن برای دسترسی زیر گستر، ارسال سیگنال در باند فرکانسی بسیار وسیع است که نرخ داده ی بالایی را با توان ارسالی پایین به دست می دهد. در عمل این روش بر اساس فرض بدترین حالت[۱۵] است که کاربران اولیه در همه ی زمان ها در حال ارسال هستند. بنابراین، در روش دسترسی زیر گستر از حفره های طیفی استفاده نمی شود. همچنین اگر در شبکه رادیو شناختی، از روش دسترسی رو گستر(یا دسترسی فرصت طلبانه) برای اشتراک گذاری طیف استفاده شود، نیازی به اعمال محدودیت روی توان ارسالی کاربران ثانویه نخواهد بود.این روش به کاربران اجازه می دهد حفره های طیفی در زمان، مکان و فرکانس را شناسایی و از آنها استفاده کنند. بنابراین، اتصال کاربران در سلول رادیو شناختی با ایستگاه پایه به صورت تک پرشی امکان پذیر است، شکل (۵).



شکل (۴):ارتباط چند پرشی کاربران رادیو شناختی با ایستگاه پایه در یک سلول



شکل (۵): ارتباط تک پرشی کاربران رادیو شناختی با ایستگاه پایه در یک سلول

پس از معرفی انواع معماری های سلولی برای شبکه های رادیو شناختی، در این بخش می خواهیم به مرور کارهایی که تا کنون در زمینه تخصیص منبع(مثل توان، بیت و زیر حامل)، مسیریابی و دگرسپاری[۱۶] در شبکه های رادیو شناختی با معماری های سلولی انجام گرفته است، پرداخته و زمینه های باز تحقیق برای علاقمندان را شناسایی نمائیم. بخش بعدی این مقاله به مرور مطالعات گذشته می پردازد.

۴- کارهای مربوط به معماری سلولی

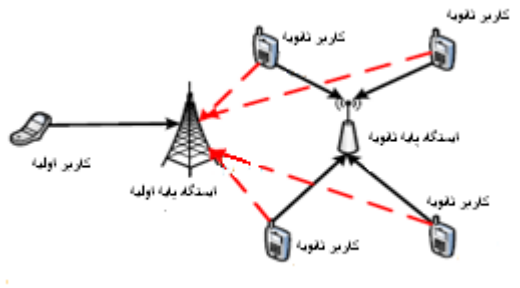
طراحی الگوریتم ها و پروتکل های تخصیص منبع، مسیریابی و دگرسپاری برای شبکه های رادیو شناختی بر اساس یک معماری معین انجام می گیرد. از آنجایی که در این مقاله، ما روی معماری های سلولی این شبکه ها تمرکز داریم، در زیر به مرور مقالات محدودی که الگوریتم های رادیو شناختی برای معماری سلولی ارائه کرده اند، می پردازیم.

۴-۱- کارهای مربوط به تخصیص منبع با معماری سلولی

در شبکه های رادیو شناختی به دلیل اینکه منابع از پیش تخصیص یافته وجود ندارد، به الگوریتم های تخصیص منبع احتیاج داریم تا ضمن حفظ نیازهای کیفیت سرویس شبکه اولیه، کارایی شبکه ثانویه را(از قبیل افزایش گذردهی و پوشش شبکه، کاهش تاخیر مسیر یابی)افزایش دهند. به همین دلیل، در مقاله [۷] مسئله تخصیص منبع

(توان، بیت و زیر حامل)در سیستم های رادیو شناختی مبنی بر OFDM که در آنها یک یا بیش از یک حفره طیفی بین باند های فرکانسی کاربران اولیه وجود دارد، بررسی شده و کاربران ثانویه می توانند هر بخشی از باند فرکانسی را تا زمانیکه تداخل اضافی با کاربران اولیه نداشته باشند، استفاده کنند. از آنجاییکه سودمندی هر کاربر در شبکه های رادیو شناختی سلولی به توان ارسال خود و توان ارسالی کاربران دیگر وابسته است بنابراین، مسئله بهینه سازی توان در انها یک مساله غیر خطی و غیر محدب است از اینرو در این مقاله مسئله تخصیص منبع به صورت مسئله کوله پشتی چند بعدی فرموله شده است که چون یک مسئله NP-hard است، نویسنده از روش شبه حریصانه استفاده کرده است اما راه حل، بهینگی و حتی جواب نزدیک به بهینه را در همه ی شرایط تضمین نمی کند. در مقاله [۶] با الهام از مقاله [۷] الگوریتم ماکسیمم-مینیمم به جهت فراهم کردن راه حل های نزدیک به بهینه ارائه شده است. طرز کار این الگوریتم بدین نحو است که برای افزایش نرخ ارسال کاربران ثانویه و افزایش کارایی شبکه رادیو شناختی، بیت بعدی را به زوج زیر کانال/ایستگاه پایه ای اضافه می کند که بیشترین سودمندی را دارد. در حقیقت منظور از بیشترین سودمندی، ایجاد تداخل کمتر است چون ارسال بیت اضافی احتیاج به افزایش توان ارسال دارد و افزایش توان سبب افزایش تداخل هم فرکانس می گردد، پس بیت بعدی به زوجی افزوده می شود که تداخل کمتری ایجاد کند. معماری بررسی شده در این مقاله یک شبکه رادیو شناختی سلولی با ۴ سلول است و در هر سلول کاربران ثانویه با کاربران اولیه همزیست هستند. پیچیدگی این الگوریتم$O(RK^3M)$ است؛ R تعداد کل بیت های بارگذاری شده و K تعداد ایستگاه های پایه و M تعداد زیر کانال ها است.از نقاط ضعف این مقاله آنست که مسئله تخصیص منبع فقط در انتقال فروسو مطالعه شده و انتقال فراسو را نادیده گرفته و شبکه اولیه را تنها با ۲ کاربر اولیه در نظر گرفته و حالتی که تراکم جمعیت کاربران اولیه زیاد باشد را به دلیل پیچیدگی بررسی نکرده است(معماری بررسی شده در این مقاله متناظر با شکل (۲) است).

در مقاله[۸] محدودیت های تخصیص توان و مسئله بهینه سازی کنترل توان غیر خطی در شبکه های رادیو شناختی سلولی مبتنی بر CDMA بررسی شده و با استفاده از برنامه نویسی هندسی[۱۷] آنرا به مسئله بهینه سازی محدب تبدیل کرده است ولی از آنجایی که برنامه نویسی هندسی معمولا به مجموعه ای از اطلاعات کلی سیستم و محاسبات متمرکز نیاز دارد باعث افزایش شدید سربار ارتباطی و تحمیل بار محاسباتی زیاد به ایستگاه های پایه می شود از اینرو از تکنیک تجزیه دوگانه برای تبدیل مسئله برنامه نویسی هندسی به چندین زیر مسئله استفاده نموده تا قابل پیاده سازی در شبکه های سلولی باشد. نقطه ضعف اصلی این مقاله اینست که با وجود ادعای معماری سلولی، اما شبکه CDMA را با تنها یک ایستگاه پایه و N کاربر بدون مجوز به عنوان شبکه ثانویه بررسی کرده و این شبکه ها را در حالتی که تراکم کاربران (یعنی N)زیاد باشد مورد مطالعه قرار نداده است، همین طور به نحوه ارتباط ایستگاه های پایه در ساختار خوشه ای شبکه های سلولی و تخصیص توان به کاربران ثانویه در باند های فرکانسی فراسو کاربران اولیه پرداخته نشده است. در مقاله[۹] مسئله بهینه سازی تخصیص توان کاربران ثانویه در باندهای فرکانسی فراسو کاربران اولیه و با حفظ کیفیت سرویس این کاربران بررسی و دو الگوریتم زمانبندی ارائه شده است. اولی به صورت حریصانه همواره کاربری را جهت ارسال در شبکه ثانویه انتخاب می کند که امکان ارسال توان بیشتری داشته باشد اما در الگوریتم دوم با تکنیک پنجره لغزان، عدالت بین کاربران ثانویه فراهم شده و ایستگاه پایه، آن کاربری که متوسط نرخ ارسال کمتری دارد برای دسترسی به کانال انتخاب می کند. به این ترتیب حداکثر گذردهی سیستم قربانی منصفانه عمل کردن می شود. نویسنده ادعا کرده معماری سلولی بررسی شده در این مقاله متناظر با شکل (۳) است، با این حال، تنها یک سلول از شبکه اولیه با یک کاربر اولیه و تعدادی کاربر ثانویه که در باند فراسو کاربر اولیه برای ارسال با یکدیگر رقابت می کنند مدلسازی شده شکل (۶). و حالت پیچیده تر وجود چندین کاربر اولیه در نظر گرفته نشده است.



شکل (۶): معماری تک سلولی [۹]

در [۱۰] نویسنده روشی جهت تعیین نقاب توان طیف[۱۸] برای هر دو ارسال فراسو و فروسو کاربران ثانویه و نقاط دسترسی[۱۹] ثانویه در هر زیر کانال بر اساس احتمال SINR[۲۰] خروجی کاربر اولیه و در باند طیفی فروسو آن بر مبنای معماری سلولی پیشنهاد داده و بیشینه کردن مجموع نرخ وزنی سیستم را به صورت مسئله بهینه سازی، مدل نموده است و معماری تحت بررسی آن، همان معماری نشان داده شده در شکل (۳) است.

چالش های تخصیص منبع در شبکه های رادیو شناختی سلولی:

برای تخصیص منبع به کاربران در شبکه های رادیو شناختی سلولی که در باند طیفی فروسو کاربران اولیه رقابت می کنند و همچنین شبکه اولیه دارای زیر ساخت و معماری سلولی و با یک یا چند مورد از شرایط زیر باشد:

- **تراکم کاربران اولیه زیاد باشد.**
- **تخصیص منبع به صورت پویا باشد.**
- **کاربران اولیه تحرک و جابجایی داشته باشند.**

تا کنون راه حلی ارائه نشده است. بعلاوه، در تمامی مطالعات انجام گرفته در این زمینه، همواره معماری سلولی برای شبکه رادیو شناختی شامل دو شبکه اولیه و ثانویه جداگانه است. معماری ای که در آن شبکه اولیه و ثانویه یکی باشند در نظر گرفته نشده است. به عبارت دیگر بجای اینکه حفره های طیفی شبکه اولیه را به یک شبکه دیگر(همان شبکه ثانویه)تخصیص دهیم، می توانیم این حفره ها را با تکنیک رادیو شناختی به خود کاربران اولیه در نقاط پر تراکم[۲۱] برای سرویس دهی بهتر تخصیص دهیم. هنوز در زمینه این ایده جدید هیچ راه حلی پیشنهاد نشده است.

۴-۲- کارهای مربوط به مسیریابی با معماری سلولی

در شبکه های بی سیم سنتی، همه گره های شبکه باند طیفی ثابت و معینی برای استفاده دارند. برای مثال، WLAN از باندهای ۲,۴ تا ۵ GHz و GSM از باند های ۹۰۰ و ۱۸۰۰ MHz استفاده می کند. در شبکه های رادیو شناختی، ممکن است چنین طیف از پیش تخصیص یافته ای وجود نداشته باشد تا توسط هر گره در هر زمان استفاده شود و طیف فرکانسی که برای ارتباط از هر گره به گره دیگر بکار می رود ممکن است بسیار متفاوت باشد.این ویژگی جدید شبکه های دسترسی پویا به طیف(یا رادیو شناختی)، چالش های زیادی به شبکه های بی سیم، بخصوص در زمینه مسیریابی تحمیل می کند. اگر دو گره همسایه کانال مشترک نداشته باشند، یا آنها کانال های مشترک داشته باشند اما به یک فرکانس کوک نشوند، اتصال چند پرشی امکان پذیر نخواهد بود. بنابراین، الگوریتم های مسیریابی جدید به جهت تطبیق با پویایی طیف نیاز است تا کارایی شبکه مثل ظرفیت و گذردهی زیاد، تاخیر کم و نرخ پایین از دست رفتن بسته ها[۲۲]را تضمین کند.از اینرو، در [۱۱] الگوریتم های مسیریابی بر اساس تخصیص حداقل توان و حداقل تداخل برای شبکه های رادیو شناختی CDMA یا IEEE ۸۰۲,۱۱ b/g (متناظر با شکل(۳)) پیشنهاد شده است. در [۱۲] الگوریتمی برای زمانبندی کنترل توان بر اساس حداقل سازی پهنای باند شبکه های رادیو شناختی سلولی ارائه شده است،اما این الگوریتم برای شبکه های سلولی چند پرشی متناظر با شکل (۴) مناسب نیست چون پیچیدگی پیاده سازی آن خیلی زیاد است. در [۱۳] الگوریتم مسیر یابی تداخل کنترل شده با استفاده از احتمال SINR خروجی کاربر اولیه پیشنهاد شده و نویسنده سعی کرده است بر مبنای معماری سلولی شکل (۳) بین ارسال تک پرشی و چند پرشی، بسته به شرایط حالت بهینه تر را انتخاب کند.از آنجایی که تاخیر در یافتن مسیر و نگهداری مسیرهای چند پرشی در شبکه های رادیو شناختی مهم است، بنابراین در [۱۷-۱۴] معیار های مسیریابی و مولفه های تاخیر بیان شده اند.این مولفه ها شامل تاخیر راه گزینی، تاخیر دسترسی به رسانه بر اساس پروتکل دسترسی به رسانه مورد استفاده و تاخیر صف است. در [۱۸] الگوریتمی برای زمانبندی دگرسپاری و مسیریابی در شبکه های رادیو شناختی چند پرشی پیشنهاد شده است. نویسنده ثابت کرده است، حداقل سازی تاخیر دگرسپاری طیفی در شبکه های رادیو شناختی چند پرشی یک مسئله NP-hard است و دو الگوریتم حریصانه متمرکز و توزیع شده پیشنهاد نموده است که راه حل نزدیک به بهینه را بدست می آورند.الگوریتم متمرکز بر اساس محاسبه حداکثر مجموعه پیوند[۲۳]هایی که با هم تضاد ندارند عمل می کند و مناسب شبکه هایی با معماری چند پرشی سلولی(دارای زیر ساخت)نشان داده شده در شکل (۴) می باشد. الگوریتم توزیع شده از معیار هزینه پیوند که با زمان نگهداری پیوند و کیفیت پیوند رابطه عکس دارد، برای مسیریابی استفاده می کند. این الگوریتم با از دست رفتن مسیر بین گره های مبدا و مقصد، مسیر جدیدی با حداقل هزینه پیدا می کند. در [۱۹] پروتکل مسیریابی SEARCH برای شبکه های رادیو شناختی چند پرشی با معماری سلولی پیشنهاد شده است.ایده اصلی پشت SEARCH شناسایی چندین مسیراز مبدا به مقصد است که در گره مقصد این مسیرها با هم ترکیب شده و مسیری با کمترین تعداد پرش ممکن انتخاب می گردد. در [۲۰] الگوریتم حریصانه برای تخصیص توان و مسیر یابی متناظر با معماری سلولی شکل (۴) با هدف بهینه سازی پوشش کاربران ثانویه ارائه شده است.این الگوریتم از فاکتور چند لگاریتمی برای بهینه سازی استفاده و راه حل نزدیک به بهینه را پیدا می کند.

چالش های بدون راه حل در این زمینه:

در شبکه های چند پرشی سنتی، گره ها با ارسال پیام های کنترلی به صورت دوره ای روی همه کانال ها با یکدیگر تبادل اطلاعات دارند ولی در شبکه های رادیو شناختی سلولی چند پرشی که رنج فرکانس ها از یک گره به گره دیگر بسیار متغیر است و مجموعه کانال از پیش تخصیص یافته ای وجود ندارد، هماهنگی بین گره ها برای یافتن توپولوژی شبکه و یافتن مسیر برای رد و بدل کردن بسته های بین مبدا به ایستگاه پایه و یا از ایستگاه پایه به گره مقصد هنوز یک چالش اساسی است.

۴-۳- کارهای مربوط به دگرسپاری با معماری سلولی

در روش دسترسی زیر گستر، به محض پدیدار شدن کاربر اولیه در باند فرکانسی اشغال شده توسط کاربر ثانویه، کاربر ثانویه مجبور به تغییر باند(یا باندهای) فرکانس عملیاتی[۲۴] خود می شود که به آن تحرک طیفی[۲۵] گفته می شود. تحرک طیفی سبب نوع جدیدی از دگرسپاری در شبکه های رادیو شناختی می شود(یعنی دگرسپاری طیفی). پروتکل های لایه های مختلف از پشته پروتکل[۲۶] شبکه باید با پارامترهای کانال فرکانس عملیاتی سازگار شوند و برنامه های کاربردی[۲۷] نباید متوجه دگرسپاری طیفی و تاخیر ناشی از آن شوند. هر زمان که یک کاربر رادیو شناختی فرکانس عملیاتی خود را تغییر می دهد، لازم است پروتکل های شبکه با پارامترهای عملیاتی تغییر کنند. پس در شبکه های رادیو شناختی هدف اطمینان از سرعت و

نرمی[۲۸] در دگرسپاری طیفی است تا حداقل کاهش کارایی را در حین دگرسپاری داشته باشیم. نیاز اساسی پروتکل های مدیریت تحرک[۲۹] در شبکه های رادیو شناختی، آگاهی از مدت زمان لازم برای دگرسپاری طیفی است. این آگاهی توسط الگوریتم های حس کردن طیف[۳۰] فراهم می شود. بعد از کسب اطلاعات تاخیر ناشی از دگرسپاری، ارتباط پیوسته با اندک کاهش کارایی امکان پذیر خواهد بود. بنابر این، به علت خصوصیات ذاتی شبکه های رادیو شناختی، با دو مفهوم جدید مواجه هستیم:تحرک طیفی و دگرسپاری طیفی. تا کنون هیچ تلاشی برای رفع مشکلات ناشی از دگرسپاری طیفی انجام نشده است. اگر چه مکانیزم های دگرسپاری مبتنی بر تحرک در شبکه های رادیو شناختی با معماری سلولی بررسی شده است،مثل [۲۱]، اما هنوز زمینه باز تحقیق وجود دارد.

زمینه های باز تحقیق برای تحرک طیفی موثر و کارا در شبکه های رادیو شناختی سلولی به شرح زیر است:

- **تحرک طیفی در حوزه زمان:**شبکه های رادیو شناختی با دسترسی فرصت طلبانه به حفره های طیفی شبکه اولیه، پهنای باند مورد نیاز خود را تامین می کنند و چون این حفره های طیفی در زمان متغیر هستند، بنابر این تامین کیفیت سرویس در این شبکه ها چالش برانگیز است.
- **تحرک طیفی در مکان:** باندهای طیفی موجود همچنین با حرکت کاربر از یک نقطه به نقطه دیگر تغییر می کنند. بنابراین، تخصیص طیف پیوسته چالش برانگیز است.

۵- نتیجه

در این سمینار کارهای انجام شده بر روی شبکه های رادیوی شناختی که دارای ساختار سلولی می باشند مورد بررسی قرار گرفت. در واقع، در شبکه های رادیو شناختی برای رسیدن به اهدافی همچون پوشش بهتر و گذردهی بیشتر نیازمند استفاده از ساختار سلولی هستیم، اما بررسی های انجام گرفته در این سمینار نشان می دهد تا کنون مطالعات بسیار کمی در زمینه شبکه های رادیو شناختی سلولی انجام شده است و با وجود اینکه در این مراجع محدود موجود، معماری سلولی برای شبکه های رادیو شناختی مبنای کار قرار گرفته است اما اکثرا تنها به بررسی یک سلول محدود اکتفا شده است. و هنوز چالش های زیادی در شبکه های رادیو شناختی سلولی بدون راه حل باقی مانده اند که می توان به آنها پرداخت. با معرفی این چالش ها، زمینه های باز برای تحقیقات آینده را در این حوزه معرفی نموده ایم.

مراجع

[۱] Federal Commiunications Commission, "Spectrum Policy Task Force Report",*FCC ۰۲-۱۳۵،۲۰۰۲*

[۲] S. Haykin, "Cognitive radio: brain-empowered wireless comminications," *IEEE Journal on Selected Areas in Communications*, vol. ۲۳,no. ۲,pp.۲۰۱-۲۲۰،۲۰۰۵

[۳] J. Mitola, G. Maguire, "Cognitive radio: Making software radios more personal",*IEEE Personal Communications* ,vol. ۶,no. ۴,pp. ۱۳-۱۸,Aug.۱۹۹۹

[۴] Friedrich K.Jondral, "Software-Defined Radio—Basics and Evolution to Cognitive Radio", *EURASIP Journal on Wireless Commiunications and Networking* ۲۰۰۵:۳, ۲۷۵-۲۸۳

[۵] O. Ileri, D. Samardzija, and N. B. Mandayam, "Demand Responsive Pricing and Competitive Spectrum Allocation via Spectrum Server," *Proc. IEEE DySPAN ۲۰۰۵,nov. ۲۰۰۵, pp. ۱۹۴-۲۰۲.*

[۶] Y. Zhang and C. Leung, "Subcarrier, Bit and Power Allocation for Multiuser OFDM-based Multi-Cell Cognitive Radio Systems" *IEEE ۹۷۸-۱-۴۲۴۴-۱۷۲۲.*

[۷] Y. Akcay, H. Li, "Greedy algorithm for the general multidimensional Knapsack problem," *Annals of Operations Research,vol. ۱۵۰ ,no. ۱,pp. ۱۷-۲۹،۲۰۰۷.*

[۸] Q. Jin, D. Yuan, "Distributed Geometric-Programming-Based Power Control in Cellular Cognitive Radio Networks," *IEEE, ۹۷۸-۱-۴۲۴۴-۲۵۱۷*

[۹] J. Zhang, Z. Zjang, "Uplink Scheduling for Cognitive Radio Cellular Network with Primary User,s QoS Protection" *IEEE ۹۷۸-۱-۴۲۴۴-۶۳۹۸.*

[۱۰] Y.Ma, D. I. Kim, "Weighted Sum Rate Optimization of Multicell cognitive radio networks," *Global Telecommunications Conference, ۲۰۰۸,Nov. ۲۰۰۸.*

[۱۱] CW. Pyo, M. Hasegawa, "Minimum weighted routing based on a common link control radio for cognitive wireless ad hoc networks," *IWCMC, ۰۷,۲۰۰۷,pp. ۳۹۹-۴۰۴.*

[۱۲] CW. Pyo, M. Hasegawa, "Minimum weighted routing based on a common link control radio for cognitive wireless ad hoc networks," *IWCMC, ۰۷,۲۰۰۷,pp. ۳۹۹-۴۰۴.*

[۱۳] Y.Shi , Y. Hou, "A distributed optimization algorithm for multi-hop cognitive radio networks," *IEEE, ۲۰۰۸.*

[۱۴] M. Xie, W. Zhang, "Ageometric approach to improve spectrum efficiency for cognitive relay networks," *IEEE ,۲۰۱۰.*

[۱۵] H. Ma, L.Zheng, "Spectrum aware routing for multi-hop cognitive radio networks with a single transceiver," *CrownCom ۲۰۰۸.*

[۱۶] G. change, W. Liu, "Spectrum aware on-demand routing in cognitive radio networks," *IEEE, DySPAN ۲۰۰۷.*

[۱۷] G. Cheng, W. Liu, "Joint on-demand routing and spectrum assignment in cognitive radio networks," *IEEE, ICC, ۰۷, ۲۰۰۷.*

[۱۸] G. Zhu, I. Akyildiz, "a spectrum-tree based on-demand routing protocol for multi-hop cognitive radio networks," *IEEE, GLOCOM. ۲۰۰۸.*

[۱۹] W. Feng ,J. Cao, "Joint optimization of spectrum handoff scheduling and routing in multi-hop multi-radio cognitive networks," *ICDCS , ۲۰۰۹.*

[۲۰] Q. U. Catholique, L. La-Neuve, "Joint Admission Control,Channel Assignment and QoS Routing for Coverage Optimization in Multi-Hop Cognitive Radio Cellular Networks," IEEE, Oct, ۲۰۱۱.

[۲۱] *IEEE ۸۰۲,۲۲ WRAN*, "A PHY/MAC proposal for IEEE ۸۰۲,۲۲ WRAN system Part ۲: The Cognitive MAC," *Mar. ۲۰۰۶.*

زیرنویس‌ها

۱ Wireless Local Area Network
۲ Federal Communications Commission
۳ Cognitive Radio
٤ Software Defined Radio
٥ Infrastructure
٦ Base Stations
۷ Single Hop
۸ Spectrum Broker
۹ Ad hoc
۱۰ Medium Access Control
۱۱ Path Loss
۱۲ Fading
۱۳ Shadowing
۱٤ Multiple Hops
۱٥ Worst Case
۱٦ Handoff
۱۷ Geometric Programming
۱۸ Power Spectrum Mask
۱۹ Access Points
۲۰ Signal to Interference Noise Ratio
۲۱ Hot Spot
۲۲ Packet Loss Rate
۲۳ Link
۲٤ Operational Frequency Band
۲٥ Spectrum Mobility
۲٦ Protocol Stack
۲۷ Applications
۲۸ Smooth
۲۹ Mobility Management
۳۰ Spectrum Sensing

دگرسپاری رادیوشناختی در شبکه‌های بیسیم سلولی

رضا برنگی[1]
صدف تفضلی[2]

[1] استادیار رشته‌ی فناوری اطلاعات، دانشگاه علم و صنعت ایران
sadaftafazoli@comp.iust.ac.ir
[2] دانشجوی کارشناسی ارشد رشته‌ی فناوری اطلاعات، دانشگاه علم و صنعت ایران
rberangi@iust.ac.ir

چکیده

شبکه‌های رادیوشناختی به دلیل بهره‌وری بهتر از پهنای باند نقش مهمی در فناوری ارتباطات و مخابرات نسل آینده خواهند داشت. در شبکه‌های سلولی پیوستگی ارتباط در خدمات بلادرنگ از اهمیت خاصی برخوردار است، از اینرو دگرسپاری یعنی انتقال ارتباط به یک سرویس دهنده دیگر برای حفظ پیوستگی ارتباط مورد توجه می‌باشد.

در شرایطی که پهنای باند برای دگرسپاری از یک سلول به سلول دیگر موجود نباشد و یا تاخیر موجود در عملیات دگرسپاری از یک حدی بیشتر باشد، مکالمه قطع می‌شود. این امکان وجود دارد که با توجه به اینکه منابع شبکه کاملا شناخته شده می‌باشد، با روش‌های به کار گرفته شده در شبکه‌های رادیوی‌شناختی از فرکانس‌های همان شبکه که در فاصله‌های دورتر مورد استفاده قرار می‌گیرد برای تامین منابع مورد نیاز برای دگرسپاری استفاده شود. در این سمینار ضمن معرفی شبکه‌های سلولی و رادیوی‌شناختی، به بررسی کارهای انجام شده در زمینه‌ی دگرسپاری در شبکه‌های سلولی با روش رادیوشناختی می‌پردازد. در حال حاضر کارهای کمی در زمینه‌ی دگرسپاری در شبکه‌های سلولی با روش‌های رادیوشناختی انجام شده است که این می‌تواند زمینه خوبی را برای انجام تحقیقات فراهم کند.

کلمات کلیدی

شبکه‌هایی رادیوشناختی، شبکه‌های سلولی، دگرسپاری

۱- مقدمه

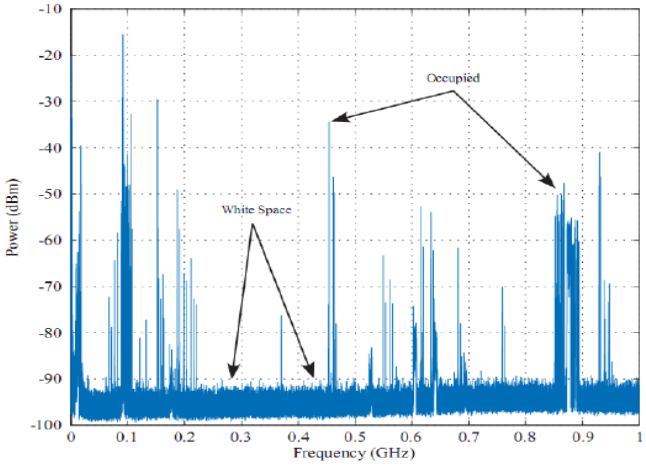
پیوستگی ارتباط در خدمات بلادرنگ از اهمیت خاصی در سیستم‌های مخابراتی برخوردار است. با توجه به محدودیت پوشش سیستم مخابراتی و تحرک کاربران، قطع ارتباط کاربر با یک سرویس دهنده بسیار محتمل بوده و از این رو دگرسپاری[1] یعنی انتقال ارتباط به سرویس دهنده دیگر برای حفظ پیوستگی ارتباط مورد توجه قرار گرفته است. اولین دگرسپاری‌ها در شبکه‌های بیسیم سلولی برای حفظ ارتباط صوتی مورد توجه قرار گرفته و تکامل یافتند. امروزه در شبکه‌های نامتجانس که هر دو نوع ارتباط بیسیم و با سیم را پشتیبانی می‌نمایند و همچنین دارای تنوع سیستم‌های مختلف بیسیم می‌باشد نه تنها برای حفظ پیوستگی سرویس بلکه برای بهبود کیفیت سرویس و کاهش انواع هزینه‌ها مورد توجه قرار گرفته است. دگرسپاری دارای چهار مرحله متمایز است که شامل تشخیص دگرسپاری، رزرو منابع (تامین منابع) ، انجام دگرسپاری و آزاد سازی منابع می‌باشد. هرکدام از این مراحل موضوع تحقیقات مفصلی بوده است. در این سمینار تمرکز ما بر روی تامین منابع دگرسپاری در سیستم‌های بیسیم سلولی می‌باشد. هدف از منبع در اینجا پهنای باند مورد نیاز برای برقراری سرویس است که باید در سیستم جدیدی که کاربر به آن نقل مکان می‌نماید در اختیار کاربر قرار گیرد.

اختصاص طیف در شبکه‌های بیسیم فعلی به صورت ایستا می‌باشد، به عبارت دیگر تخصیص طیف فرکانسی توسط آژانس‌های دولتی صورت می‌گیرد، به این صورت که این سازمان‌ها باند فرکانسی خاصی را به استفاده‌کنندگان حقوقی تخصیص می‌دهند و آن کاربران،

تنها مجاز به بهره گیری از باند تخصیص داده شده می‌باشند. قسمت اعظمی از این پهنای باند تخصیص داده شده عموماً گاه‌به‌گاه و به ندرت مورد استفاده قرار می گیرد [1]. استفاده از طیف در زمـان و مکـان مختلف عموماً بین ۱۵ تا ۸۵ درصد می‌باشد (شکل ۱). اگر چه ایـن سیاست تخصیص در گذشته مناسب بوده، اما با توجه به نیازمندی‌های نسل چهارم موبایل، محدودیت طیف دردسترس و استفاده غیر بهینه از طیف موجود مسیر تحقیقات را به سمت استفاده از روش‌هایی جهت بالا بردن بهره‌وری طیفی هدایت کرده است[2].

رادیوی‌شناختی کلید استفاده پویا و دسترسی فرصت طلبانه به طیف فرکانسی است. توانایی حس کردن محیط و انطباق با شرایط موجود محیطی از جمله ویژگی‌های شبکه‌های رادیوی‌شناختی می‌باشد[3]. استفاده از رادیوی‌شناختی برای اولین بار با توسط J.Mitolla مطرح گردید[4] ، و افرادی مانند Akyildiz جایگاه ویژه آن را در شبکه‌های نسل آینده مورد بررسی قرار دادند[5].

بخش‌های مختلف این سمینار به این ترتیب می‌باشد، در بخش دوم مروری بر شبکه‌های سلولی و کارهایی که در رابطه با فرایند دگرسپاری در شبکه‌های سلولی انجام شده است آمده است. در بخش سوم به معرفی رادیوی‌شناختی پرداخته شده است و روش‌های مختلف دسترسی طیف توسط رادیوی‌شناختی آمده است. چهارمین بخش به کارگیری رادیوی‌شناختی در شبکه‌های بیسیم را بررسی کرده و ملاحظات لازم برای دگرسپاری در شبکه‌های رادیوشناختی مطرح گردیده است. بخش پنجم به بررسی کارهای انجام شده در زمینه‌ی دگرسپاری رادیوشناختی در شبکه‌های سلولی پرداخته است و در نهایت در بخش ششم نتیجه‌گیری می‌باشد.



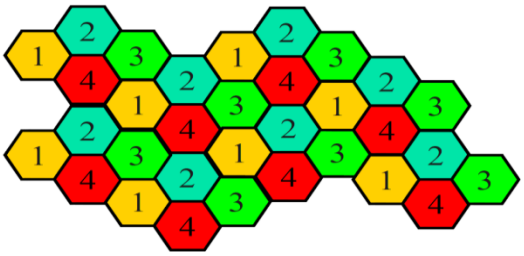
شکل ۱: بخش‌های استفاده شده و بدون استفاده‌ی طیف [6]

۲- شبکه‌های سلولی

محدوده‌ای را که یک آنتن مرکزی می‌تواند پوشش بدهد را سلول می‌گویند. اگر در یک سیستم منفرد دو آنتن مختلف موجود باشد، هرکدام یک سیستمی را برای خود تشکیل خواهند داد که با دیگری هیچ ارتباطی نخواهد داشت. در شبکه‌های رادیویی سلول‌های منفرد از طریق کابل زمینی توسط یک مرکز سوئیچ با هم ارتباط برقرار می‌کنند به این طریق می‌توان بین یک مشترک که در داخل یک سلول هست با مشترک دیگری که در داخل سلول دیگریست ارتباط برقرار کرد پوشش رادیویی متناسب است با ارتفاع آنتن، قدرت فرستنده و حساسیت گیرنده می‌باشد که در این بین ارتفاع آنتن از اهمیت بیشتری برخوردار است هرچه ارتفاع آنتن بیشتر باشد وسعتی که سلول پوشش می‌دهد گسترده‌تر است. در شبکه‌های سلولی برای افزایش ظرفیت سیستم از بازیابی فرکانسی استفاده می‌کنند، به همین دلیل برای پوشش وسیع، از تعداد زیاد ایستگاه‌های پایه (آنتن‌ها) که هر یک پوشش محدودی دارند، استفاده می‌کنند.

منابع اصلی شبکه موبایل پهنای باند فرکانسی اختصاص یافته به آن است. پهنای باند را به دو دسته تقسیم می‌کنند قسمت بالای باند که به آن لینک فروسو[۲] می‌گویند و قسمت پایین باند که به آن لینک فراسو[۳] میگویند، فاصله‌ای را هم بین آنها در نظر می‌گیرند که به آن محافظ می‌گویند.

برای استفاده‌ی بهینه از فرکانس، مفهومی به اسم بازیابی فرکانسی توسعه پیدا کرد. بازیابی فرکانسی در دو حوزه زمان (باند فرکانسی که به یک موبایل اختصاص داده شده است را در صورتی که او از آن استفاده نکند بتوان در اختیار دیگران قرار داد) و مکان (یک باند فرکانسی را که در یک محدوده توسط یک موبایل استفاده می‌گردد، در ناحیه جغرافیایی دورتر نیز مورد استفاده قرار گیرد) مطرح است. برای اینکه بتوان از این بازیابی فرکانسی در حوزه مکان بهره گیری کنند باند فرکانسی که به یک شبکه موبایل اختصاص داده شده است را به تعدادی زیر باند تقسیم بندی می‌کنند آنگاه آن زیر باندها را خوشه بندی می‌کنند. یک الگوی خوشه K تایی از باندها را می‌شود چنان تصور کرد که هر سلول آن خوشه بتواند از یک دسته باند استفاده کند این خوشه‌ها را می‌توان کنار هم چید و فضای جغرافیایی را پر کرد، بدین ترتیب یک توزیع متقارن از سلول‌های با فرکانس‌های مشابه در پهنای باند جغرافیایی ایجاد می‌شود (شکل ۲).



شکل ۲: یک الگوی خوشه‌ی چهار تایی [7]

۲-۱- تحرک در شبکه‌های سلولی و دگرسپاری

وقتی که موبایل در درون یک سلول از یک کانال مخابراتی استفاده می‌کند و ارتباطش برقرار است، ممکن است حرکت کند و از حوزه یک سلول خارج بشود، چون خروج از یک سلول موجب تضـعیف مـوج مخابراتی می‌شود این باعث می‌شود که ارتباط تضعیف و نهایتاً قطع

گردد، برای اینکه این ارتباط قطع نگردد و مکالمه ادامه پیدا کند، باید ارتباط جدیدی بین موبایل و ایستگاه پایه آن سلول جدید برقرار شود، که به این فرایند دگرسپاری می‌گویند[7]. روش‌های دگرسپاری در شبکه‌های سلولی به دو دسته اصلی دگرسپاری سخت و دگرسپاری نرم تقسیم می‌شوند. دگرسپاری سخت از روش قطع قبل از وصل[4] استفاده می‌کند. یعنی ارتباط با ایستگاه پایه قدیمی قطع می‌گردد قبل از اینکه ارتباط جدیدی با ایستگاه پایه جدیدی شکل بگیرد. دگرسپاری نرم از روش وصل قبل از قطع[5] استفاده می‌کند. یعنی ارتباط با ایستگاه پایه جدید برقرار می‌گردد قبل از اینکه ارتباط موبایل با ایستگاه پایه قدیمی قطع گردد.

به منظور پشتیبانی بیشتر از تحرک کاربر در شبکه‌های سلولی از سرویس رومینگ استفاده می‌گردد. در واقع رومینگ قراردادهایی است که بین دو شبکه‌ی مخابراتی تصویب می‌گردد که براساس آن به کاربران شبکه‌ی طرف قرارداد خود اجازه استفاده از سرویس‌های شبکه خود را می‌دهند، فرض کنید فردی یک سیم کارت از شرکت ایرانسل گرفته است، بنابراین او با شرکت ایرانسل قرارداد می‌بندد و بر اساس آن قرارداد مکالماتش محاسبه می‌شود و هزینه‌ی آن از او دریافت می‌گردد. این فرد با شرکت مخابراتی دیگری به طور مستقیم قرارداد نبسته است ولی اگر وارد سلولی شد که مربوط به شرکت تالیا است اگر قرارداد رومینگ بین دو شرکت مخابرات ایرانسل و تالیا موجود باشد، تالیا این اجازه را به کاربر می‌دهد که از سیم کارتش در سلول‌های مربوط به شبکه تالیا استفاده کند. تالیا هزینه‌ی این کاربر را حساب کرده و به شرکت ایرانسل اطلاع می‌دهد، آنگاه شرکت ایرانسل آن هزینه را از کاربر می‌گیرد و بخشی از آن را به تالیا طبق قرار بینشان می‌دهد.

قرار دادهایی که به صورت کاغذی بین دوشرکت امضا می‌شوند، باید به صورت الکترونیکی قابل پیاده سازی باشد که این پیاده سازی را از طریق HLR[6] و VLR[7] انجام می‌دهند. HLR پایگاه داده‌ای می‌باشد که زمانی که موبایلی در یک شرکت ثبت می‌شود اطلاعات آن مشترک در آن شرکت در این پایگاه داده نگه داری می‌گردد و VLR پایگاه داده‌ای می‌باشد که وقتی موبایل وارد شبکه جدیدی غیر از شبکه اصلی می‌شود یک رکورد موقت از آن در این پایگاه داده نگهداری می‌شود که موبایل به طور دوره‌ای با اطلاع دادن موقعیت خود به VLR و HLR اطلاعات خود را تازه سازی می‌کند.

۲-۲- تخصیص پویای طیف در شبکه‌های سلولی

همان طور که گفته شد در شبکه‌های سلولی تخصیص طیف به صورت ایستا فرض شده است، یعنی به هر یک از سلول‌های یک خوشه در یک شبکه تعداد ثابتی فرکانس (کانال) اختصاص می‌یابد که این کانال‌ها قابل استفاده توسط کاربر در سلول‌های دیگر آن خوشه نمی‌باشد. اگر کاربر در طول یک تماس از سلولی به سلول مجاور خود حرکت کند دگرسپاری می‌کند منتها اگر سلول جدید که به آن وارد شده است تمام کانال‌هایش پر باشد ارتباط قطع می‌گردد، این درحالیست که امکان دارد در سلول‌های دیگر آن خوشه کانال خالی وجود داشته باشد بنابراین، اگر امکان اختصاص آن کانال‌ها به سلولی که ازدحام آن زیاد است ممکن بود تماس قطع نمی‌شد. در همین راستا تحقیقات بسیاری به منظور رفع این مشکل صورت گرفته است، در واقع موضوع این تحقیقات اختصاص پویای طیف می‌باشد[8]. با توجه به مفهوم شبکه‌ی سلولی امکان استفاده مجدد از یک کانال در یک مکان وجود ندارد مگر اینکه تداخل هم‌فرکانس در آن مکان به قدری ناچیز باشد که بتوان آن را به عنوان نویز در نظر گرفت[7].

الگوریتم‌های بسیاری برای اختصاص پویای طیف ارائه گردیده است در تمامی این الگوریتم‌ها زمانی که درخواستی برای تخصیص کانال تولید می‌گردد کنترل کننده‌ی شبکه اقدام به اختصاص طیف به صورت پویا می‌کند، حال چه این کنترل کننده یک MSC باشد یا یک BSC، این بستگی به متمرکز یا غیر متمرکز بودن الگوریتم اختصاص پویای طیف دارد. کنترل کننده کانالی را انتخاب می‌کند که منجر به کمترین تداخل هم‌فرکانس گردد به این منظور یک جستجوی سراسری را در شبکه انجام می‌دهد و همچنین به منظور برآوردن کیفیت بالای سرویس نیاز می‌باشد تا به محض آزاد شدن یک کانال اطلاعات مورد نیاز الگوریتم اختصاص پویای طیف به روز گردد[9].

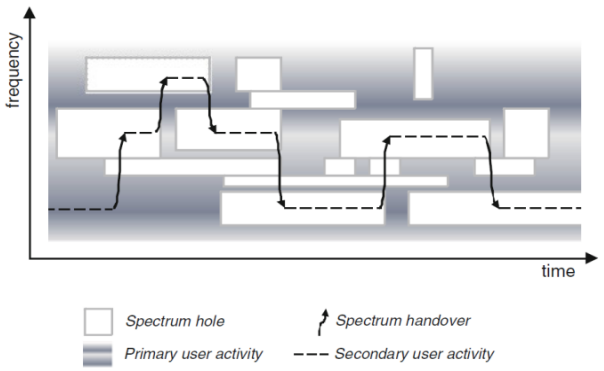
از جمله مزیت‌های اختصاص پویای طیف کمتر بودن تعداد دگرسپاری‌های ناموفق و احتمال قطع یک تماس در حال اجرا می‌باشد. منتها روش‌های اختصاص پویای طیف به دلیل استفاده از الگوریتم‌هایی که نیاز به جستجو و به روزرسانی سراسری اطلاعات دارند به اندازه‌ی کافی سریع نمی‌باشند و لذا منجر به کاهش کیفیت سرویس[8]های ارائه شده می‌گردد[10].

۳- رادیوی‌شناختی

رادیوی‌شناختی را میتوان به صورت رادیویی که شناختی می‌باشد تعریف کرد و یا به عبارت دیگر میتوان گفت آنچه فکر می‌کند، رادیوی‌شناختی است[11]. البته به دلیل عدم توافق و ناسازگاری‌هایی که روی میزان شناخت مورد نیاز برای رادیوشناختی در میان محققان وجود دارد، اختلافاتی روی تعریف رادیوشناختی ایجاد شده است. بر اساس تعریف FCC[9] رادیوی‌شناختی به این صورت تعریف میشود: رادیوی‌شناختی یک رادیویی می‌باشد که می‌تواند پارامترهای ارسال خود را بر اساس تعاملی که با محیط کاری خود دارد تغییر دهد[12] در حالیکه بر اساس تعریف ITU[10] رادیوی‌شناختی، رادیو یا سیستمی است که محیط کاری خود را حس کرده و از آن آگاه می‌باشد و به صورت خودکار و پویا می‌تواند پارامترهای کاری خود را بر طبق آن تنظیم نماید[13]. در واقع میتوان گفت که اختلافات موجود در تعریف رادیوی‌شناختی مربوط به اختلاف درانتظاراتی می‌باشد که از عملکرد رادیوی‌شناختی می‌توان داشت.

رادیوی‌شناختی بهره برداری موثر از طیف را توسط تعریف دو نوع از کاربران اولیه و ثانویه امکان پذیر نموده است. رادیوی‌شناختی در فرکانس‌هایی که در اصل و در ابتدا توسط آژانس‌های دولتی به سرویس‌های رادیویی مجوز داده شده است کار می‌کند، علاوه بر آن در باندهای فرکانسی بدون مجوز نیز می‌تواند کار کند. کاربران اولیه، کاربران مجوزداری هستند که به آنها کانال خاصی منتصب شده است و کاربران ثانویه، کاربران بدون مجوزی هستند که اجازه دارند از کانال‌های تخصیص یافته به کاربران اولیه فقط در شرایطی که هیچگونه تداخل مخربی برای کاربران اصلی ایجاد نکنند، استفاده نمایند. برای مثال در IEEE 802.22 WRAN، برج فرستنده تلویزیونی به عنوان کاربر اصلی بوده و دستگاه‌های رادیویی که از کانال‌های تلویزیونی برای برقراری ارتباط استفاده می‌کنند، کاربران ثانویه می‌باشند.

کاربران ثانویه قصد دارند توسط روش‌ها فرصت‌طلبانه به طیف موجود دسترسی پیدا کنند به نحوی که به روی کار کاربران اولیه تداخل ایجاد نکنند. یک کاربر ثانویه با حس کردن همه کانال‌های موجود بهترین کانالی را که فعلاً مورد استفاده کاربر دارای مجوز نیست را انتخاب می‌کند و برای مدتی از آن استفاده می‌کند. البته این استفاده تا زمانی باید ادامه داشته باشد که کاربر ثانویه بازنگشته باشد[14].

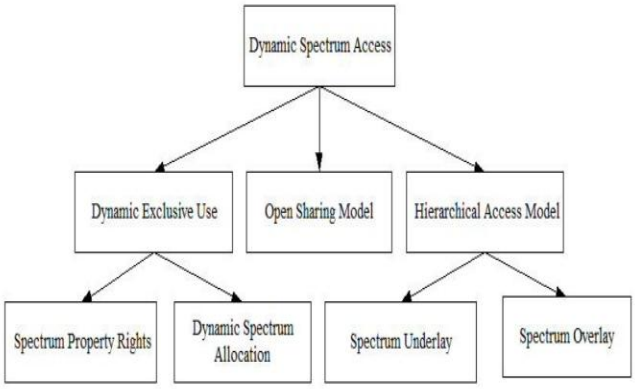


شکل ۳: دسترسی فرصت طلبانه به حفره‌های طیفی [15]

در رادیوی‌شناختی هدف استفاده بهینه از حفره‌های طیفی است. حفره‌ی طیفی همان زمان‌های عدم استفاده کاربران دارای مجوز از فرکانس تخصیص داده شده به آن‌ها می‌باشد. (شکل ۳) بیان کننده مفهوم حفره طیفی و پرش فرکانسی است.

۳-۱- روش‌های دسترسی پویای طیف[11]

روش‌های دسترسی پویای طیف به سه گروه مشابه (شکل ۴) تقسیم‌بندی می‌شود.



شکل ۴: روش‌های دسترسی پویای طیف [16]

۳-۱-۱- بهره برداری انحصاری پویا[12]

این مدل از دسترسی به طیف، مانند مدل رایج دسترسی به طیف می‌باشد که در آن باندهای طیف به فراهم کنندگان سرویس برای استفاده‌های انحصاری مجوز داده می‌شود. این مدل دارای دو روش حق مالکیت طیف[13] و تخصیص پویای طیف[14] می‌باشد. در روش حق مالکیت طیف، کاربران مجوزدار دارای حق انحصاری برای انتخاب آزادانه تکنولوژی بوده و طیف خودشان را می‌توانند بفروشند یا مبادله نمایند. در روش تخصیص پویای طیف، طیف به صورت انحصاری در زمان و مکان مشخص به فراهم کنندگان سرویس اختصاص داده میشود[16].

۳-۱-۲- اشتراک طیف باز[15]

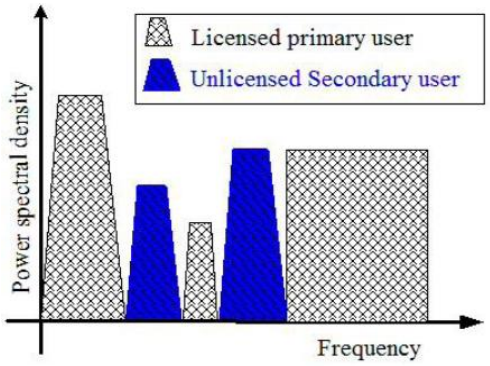
این مدل شامل سرویس‌های بیسیمی می‌باشد که در روش بدون مجوز کار می‌کنند و تمام کاربران دارای فرصت برابر برای دسترسی به طیف هستند. به هر حال کاربران رادیوشناختی می‌توانند کانال‌هایی که دارای ترافیک کمتر هستند را نسبت به کانال‌های با ترافیک زیاد انتخاب نمایند.

۳-۱-۳- دسترسی سلسله مراتبی[16]

این مدل شامل یک سلسله مراتبی بین کاربران مجوزدار و کاربران شناختی بدون مجوز(کاربران ثانویه) می‌باشد. در این مدل از اشتراک طیف، کاربران بدون مجوز می‌توانند به طور پویا به طیف (که روی آن دارای مجوز دسترسی نمی‌باشند) دسترسی یابند اما آنها باید اطمینان یابند که تداخل ایجاد شده توسط آنها در کاربران اصلی مجوزدار در ناحیه قابل تحمل باشد و یا از طیف به صورت فرصت طلبانه و بدون تداخل با کاربران اصلی استفاده نمایند. این مدل دارای دو روش پایه روگستر[17] و زیرگستر[18] می‌باشد.

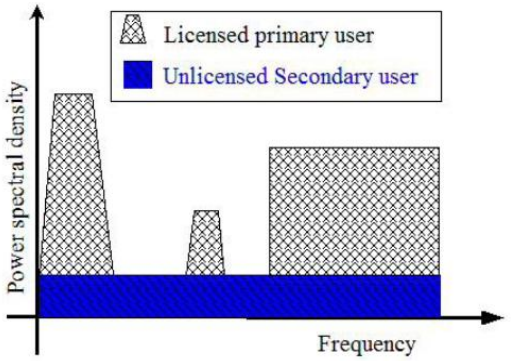
- روش روگستر: در این روش، رادیوشناختی باید باندهای بیکار طیف را که توسط سیستم مجوزدار در یک زمان و مکان خاص استفاده نشده‌اند را شناسایی کرده و از این

باندهای خالی به صورت پویا استفاده نماید. این روش در (نمودار ۱) نمایش داده شده است.



نمودار ۱: دسترسی به طیف، روش روگستر [16]

- روش زیرگستر: در این روش، کاربران ثانویه شناختی اجازه ارسال با توان ارسال ضعیف را به طور همزمان با ارسال کاربران اصلی دارند اما باید ارسال آنها به گونه‌ای باشد که تداخل ایجاد شده از ارسال کاربران شناختی در کاربران اصلی زیر یک حد آستانه قرار گرفته و باعث ایجاد تداخل مضر با کاربران اصلی نشود. این روش در (نمودار ۲) نشان داده شده است.



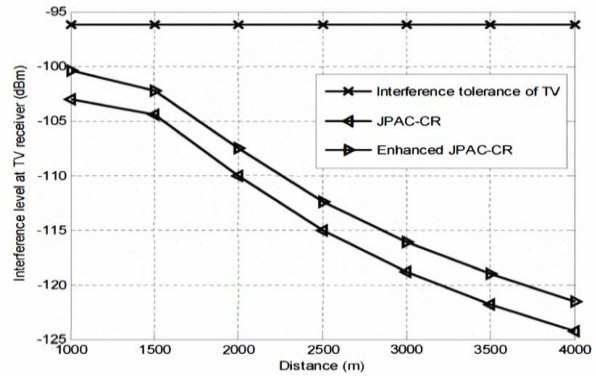
نمودار ۲: دسترسی به طیف، روش زیرگستر[16]

نکته مهمی که در رابطه با روش دسترسی زیرگستر مطرح می‌باشد تعداد کاربران ثانویه‌ی پذیرفته شده در سیستم است که بر روی کیفیت سرویس شبکه تاثیر می‌گذارد این که چه تعداد کاربر ثانویه همزمان در یک کانال حضور داشته باشند و از طرف دیگر کیفیت سرویس برای کاربران ثانویه و اولیه رضایت‌بخش باشد. یک مسئله‌ی NP-Complete می‌باشد. راه‌حل بهینه استفاده از الگوریتم BFA[۱۹] می‌باشد، که نیاز به یک جستجوی جامع دارد که دارای یک پردازش طولانی می‌باشد. در [17] الگوریتمی را برای حل این مسئله پیشنهاد می‌دهد که علاوه بر یک کنترل‌کننده‌ی پذیرش که کاربران ثانویه‌ی متقاضی را یک به یک حذف می‌کند، تا QOS[۲۰] به میزان رضایت‌بخش کاربران (اولیه و ثانویه) برسد همچنین از یک کنترل‌کننده‌ی قدرت نیز استفاده می‌کنند که باعث می‌گردد عمل جستجو به منظور حذف کاربران ثانویه سریع‌تر گردد.

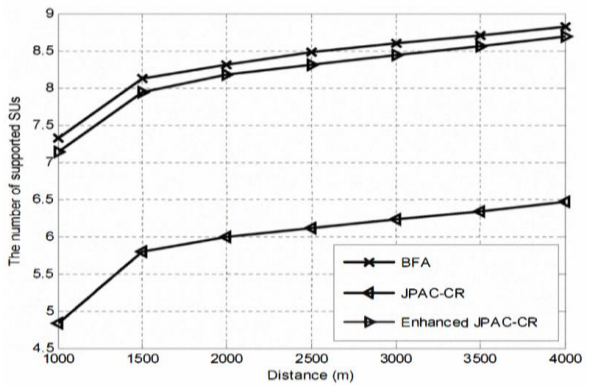
با توجه به (نمودار ۳) با افزایش فاصله از آنتن گیرنده‌ی کاربراولیه میزان تداخل بر کاربراولیه کاهش میابد، بنابراین این امکان وجود دارد که بتوان در فاصله‌ی دورتر از کاربراولیه تعداد کاربران ثانویه‌ی بیشتری را پشتیبانی کرد.

همچنین این موضوع در (نمودار ۴) نشان داده شده است، با افزایش فاصله از کاربراولیه تعداد بیشتری کاربرثانویه می‌توانند از یک کانال استفاده کنند.

در (نمودار ۳) و (نمودار ۴) معیار Enhanced JPAC-CR مربوط به حالتی است که پارامترهای الگوریتم، تغییر داده شده‌اند، که به ازای تغییر پارامترها سرعت جستجو کندتر می‌گردد، در حالی که دقت آن افزایش می‌یابد.



نمودار ۳: میزان تداخل در گیرنده‌ی تلویزیون[17]



نمودار ۴: مقایسه‌ی کاربران ثانویه‌ی قابل پشتیبانی[17]

هر یک از این دو روش دسترسی دارای مزایا و معایب خودشان هستند. برای مثال در روش روگستر، کاربران ثانویه می‌توانند با یک توان ارسال بالا ارسال خودشان را انجام دهند اما به هر حال باید باندهای خالی طیف را برای ارسال روی آنها شناسایی کنند به طور مشابه در روش دسترسی به طیف زیر گستر، کاربران ثانویه نیازمند شناسایی و تعیین فرصت‌های طیف نمی‌باشد و می‌توانند به طور همزمان با حضور کاربران اصلی ارسال خودشان را انجام دهند اما به هر حال آنها اجازه ارسال با توان ارسال بالا را حتی وقتی کل باند فرکانس رادیویی خالی باشد را ندارند[16]. بنابراین این مساله مطرح است که استفاده از کدامیک از روش‌های دسترسی کارایی بیشتری خواهد

داشت. به منظور پاسخ گویی به این مساله تحقیقات بسیاری صورت گرفته است [18, 19].

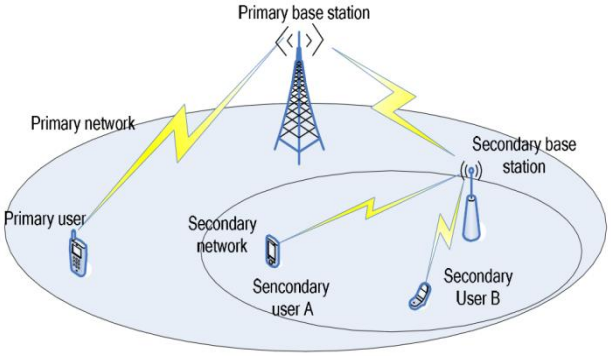
۳-۱-۴- مقایسه‌ی روش‌های دسترسی روگستر و زیرگستر بر اساس تداخل کاربر ثانویه بر کاربر اولیه و احتمال قطع ارتباط

در [18] سه روش دسترسی طیف را مورد توجه قرار داده است: روش روگستر(جلوگیری از تداخلات)، روش زیر گستر و همچنین یک روش پیوندی[۲۱] (روش زیرگستر با استفاده از جلوگیری از تداخلات) که در این روش کاربر مخابره اش را بر روی یک طیف کامل توزیع می کند و همچنین فرکانس‌هایی که کاربر اولیه ارسال می کند را مورد بررسی قرار می دهد. در این مقاله از معیار احتمال قطع ارتباط برای مقایسه‌ی این سه روش استفاده کرده و به شبیه سازی هر یک از روش‌های دسترسی با فرض نداشتن هیچ دانش سیستمی، دانش سیستم بی نقص و دانش سیستم محدود شده پرداخته است.

درست مشابه سایر کارهای موجود، وقتی که یک دانش سیستمی بی نقص فرض گرفته می شود، روش روگستر عملکرد فراتر از انتظار را در مقایسه از روش زیر گستر خواهد داشت و زمانی که هیچ دانش سیستمی وجود ندارد، هر سه روش عملکرد بسیار ضعیفی دارند منتها در مقایسه روش زیرگستر عملکرد بهتری از خود نشان می‌دهد. زمانی که جلوگیری از تداخلات با به اشتراک گذشتن طیف‌ها ترکیب می شود، روش زیرگستر همراه با پیشگیری از تداخلات احتمال قطع کمتری را ضمانت می کند نسبت به وقتی که جلوگیری از تداخلات خالص مد نظر است. زمانی که میزان دانش سیستمی، به واقعیت نزدیک است یعنی زمانی که دانش سیستمی محدود شده فرض گرفته شده است، اهمیت روش‌های پیوندی خود را بروز می دهند. روش روگستر با داشتن اطلاعات محدود، در شناسایی حفره‌های طیفی نقص دارد. به صورت دقیق تر، یک کاربرثانویه می تواند از یک کانال استفاده کند درحالیکه که یک مصرف کننده اولیه در حال استفاده از کانال باشد منتها زمانی که از روش زیرگستر همراه با جلوگیری از تداخلات استفاده می شود، تداخلاتی که در مصرف کننده اولیه ایجاد می‌شود به حداقل می رسد.

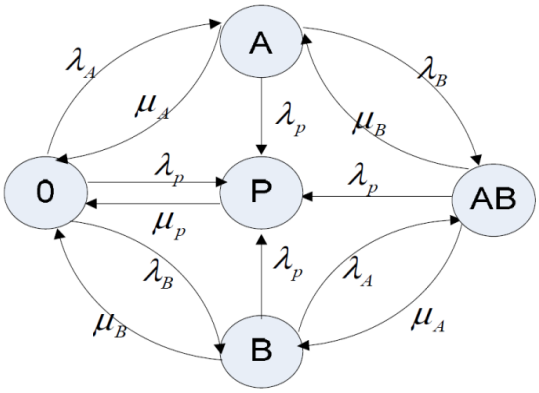
۳-۱-۵- مقایسه‌ی روش‌های دسترسی روگستر و زیرگستر با استفاده از مدل مارکف

در [19] به مدل کردن دسترسی پویای طیف زیرگستر، با در نظر گرفتن SINR[۲۲] در کاربر اولیه با مدل مارکف، پرداخته است. اگر تداخل ناشی از کاربران ثانویه و نویز محیط در کاربر اولیه کمتر از یک حد آستانه باشد طوری که مزاحمتی برای کاربر اولیه ایجاد نکند، در این صورت کاربر اولیه و کاربران ثانویه، می‌توانند با یکدیگر از یک کانال، استفاده کنند. همانطور که در (شکل ۵) مشاهده می‌شود، اشتراک طیف بین یک کاربر اولیه و دو کاربر ثانویه صورت گرفته است

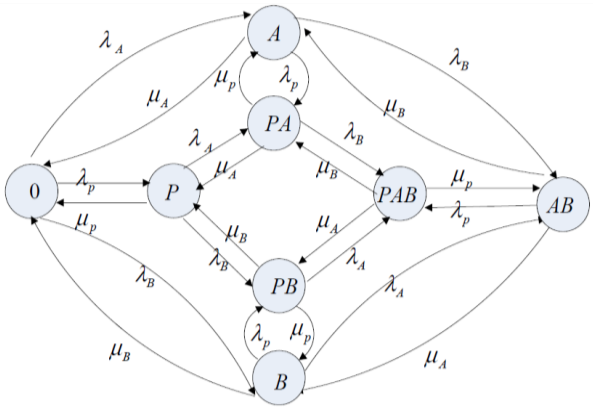


شکل ۵: اشتراک طیف بین کاربران اولیه و ثانویه[19]

مدل مارکف مربوط به این اشتراک طیف، در دو حالت دسترسی پویای روگستر و زیرگستر در زیر آورده شده است.



شکل ۶: زنجیره‌ی مارکف برای مدل کردن دسترسی روگستر[19]



شکل ۷: زنجیره‌ی مارکف برای مدل کردن دسترسی زیرگستر[19]

در این مدل‌ها تفسیر هر از حالات به صورت زیر میباشد: حالت '0' بیانگر حضور هیچ کاربری در کانال می‌باشد، حالت 'P'، بیانگر حضور کاربر اولیه در کانال، حالت 'A' بیانگر حضور کاربر ثانویه‌ی A، حالت 'B' بیانگر حضور کاربر ثانویه‌ی B، حالت 'AB' بیانگر حضور کاربر ثانویه‌ی A و کاربر ثانویه‌ی B، حالت 'PA' بیانگر حضور کاربر اولیه‌ی P و کاربر ثانویه‌ی A، حالت 'PB' بیانگر حضور کاربر اولیه‌ی P و کاربر ثانویه‌ی B و نهایتا حالت 'PAB' بیانگر حضور کاربر اولیه‌ی P، کاربر ثانویه‌ی A و کاربر ثانویه‌ی B می‌باشد.

در این مدل‌ها فرض شده است که ترافیک ورود کاربران اولیه و ثانویه پواسن است که در این (شکل ۶) و (شکل ۷) λ_P و λ_Y به ترتیب معرف این دو پارامتر هستند. و زمان سرویس‌دهی برای هر دو کاربر اولیه و ثانویه نمایی در نظر گرفته شده که پارمترهای μ_P و μ_Y نیز به ترتیب بیانگر این دو پارامتر می‌باشند با کمک حل معادلات مدل مارکف برای هر دو حالت می‌توان نرخ داده را برای کاربران ثانویه به دست آورد. با شبیه سازی‌هایی که در نرم‌افزار MATLAB انجام شده‌است، نتیجه‌گیری شده که گذردهی در حالت زیرگستر نسبت به حالت روگستر بیشتر می‌باشد.

۴ - شبکه‌های رادیوشناختی

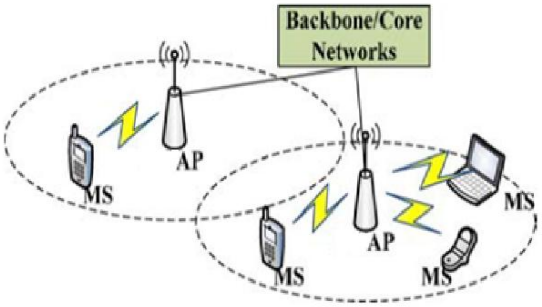
دو قابلیت اصلی رادیوی شناختی عبارتند از قابلیت شناخت و قابلیت پیکربندی مجدد.

- قابلیت شناخت: قابلیت شناخت به قابلیتی از تکنولوژی رادیویی گفته می‌شود که به کمک آن می‌توان اطلاعات محیط رادیویی را حس کرد و یا بدست آورد. بدست آوردن اطلاعات محیط رادیویی به آسانی و با نظارت قدرت سیگنال رادیویی بدست نمی‌آید بلکه نیاز به تکنیکهای هوشمندتری برای بدست آوردن رفتار پویای طیف می‌باشد. به کمک این قابلیت می‌توان با بررسی رفتار پویای طیف، قسمتهای بدون استفاده طیف را برای استفاده بقیه کاربران شناسایی کرد و ویژگیهای آنها را مشخص نمود.
- قابلیت پیکربندی مجدد: قابلیت شناخت، اطلاعات لازم برای شناخت رفتار پویای طیف را دراختیار می‌گذارد درحالی‌که قابلیت پیکربندی مجدد امکان استفاده پویا از طیف به کمک اطلاعات بدست آمده از قابلیت شناخت را فراهم می‌سازد.

یکی از مهمترین اهداف رادیوی‌شناختی فراهم نمودن بهترین طیف موجود به کمک قابلیت‌های ذکر شده می‌باشد. شبکه‌ای که از رادیوی‌شناختی یعنی از دو قابلیت شناخت و پیکربندی مجدد استفاده می‌کند، شبکه‌ی رادیوشناختی نامیده می‌شود. اجزاء پایه در شبکه‌های رادیوشناختی ایستگاه‌های متحرک(MS[۲۳])، نقاط دسترسی(AP[۲۴]) و ستون فقرات[۲۵] شبکه می‌باشند. با این سه جزء پایه، سه نوع معماری برای شبکه‌های رادیوشناختی ممکن است: معماری متمرکز(زیرساختی)، معماری توزیع شده(ادهاک) و معماری ترکیبی(مش).

۴-۱-۱- معماری متمرکز

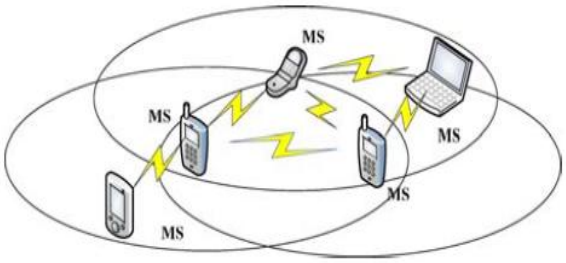
در این معماری یک MS می‌تواند فقط به یک AP و در یک روش یک پرشی دسترسی یابد. در این روش MSهای تحت پوشش ارسال یک AP می‌توانند با یکدیگر از طریق AP ارتباط برقرار نمایند. ارتباط بین سلول‌های مختلف از طریق بخش مرکزی شبکه مسیردهی می‌شود AP می‌تواند یک یا چند پروتکل ارتباطی را برای برآورده کردن نیازمندی‌های مختلف MSها را اجرا نمایند. یک ترمینال رادیوشناختی نیز می‌تواند به انواع مختلفی از سیستم‌های ارتباطی از طریق APهای آنها دسترسی یابد[20]. این معماری در (شکل ۸) نشان داده شده است.



شکل ۸: معماری متمرکز برای شبکه‌ی رادیوشناختی[20]

۴-۱-۲- معماری توزیع شده

در معماری توزیع شده از شبکه‌های رادیوشناختی، هیچ پشتیبانی زیرساختی وجود ندارد. در این معماری، شبکه روی هوا تاسیس می‌شود. اگر یک MS تشخیص دهد که یک MS دیگر در آن نزدیکی وجود دارد و از طریق پروتکل‌های ارتباطی مشخص قابل دستیابی می‌باشد، آنها می‌توانند یک ارتباط را برقرار کرده و بنابراین یک شبکه ادهاک را تشکیل دهند. باید توجه شود که این ارتباطات بین گره‌ها، می‌توانند توسط تکنولوژی‌های ارتباطی مختلفی برپا شوند. به علاوه، دو ترمینال رادیوشناختی می‌توانند با همدیگر توسط استفاده از پروتکل‌های ارتباطی موجود (Wi-Fi، Bluetooth و ...) و یا استفاده پویا از حفره‌های طیف، ارتباط برقرار نمایند[20]. (شکل ۹) معماری توزیع شده برای شبکه‌های رادیوشناختی را نشان می‌دهد.

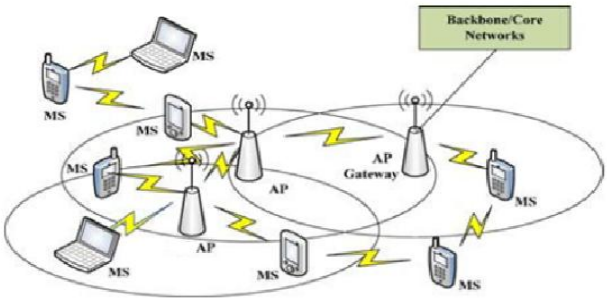


شکل ۹: معماری توزیع شده برای شبکه‌ی رادیوشناختی[20]

۴-۱-۳- معماری ترکیبی

این معماری ترکیبی از معماری‌های متمرکز و توزیع شده بوده، به علاوه ارتباطات بیسیم بین APها را امکان پذیر می‌سازد این معماری شبکه، مشابه با شبکه‌های بیسیم مش پیوندی[۲۶] می‌باشد. در این معماری، APها به صورت مسیریاب‌های بیسیم کار کرده و ستون فقرات بیسیم را تشکیل می‌دهند. MSها نیز می‌توانند به APها به صورت مستقیم و یا با استفاده از دیگر MSها به عنوان گره‌های رله

چند پرشی دسترسی یابند. بعضی از AP ها می‌توانند به شبکه ستون فقرات سیمی اتصال یافته و به عنوان دروازه عمل نمایند. چون AP ها می‌توانند بدون اینکه الزاما به شبکه‌های ستون فقرات سیمی اتصال یابند، توسعه داده شوند، بنابراین انعطاف پذیری آنها بیشتر بوده و همچنین هزینه کمتری برای طراحی مکان APها مورد نیاز می‌باشد. این معماری در (شکل ۱۰) نشان داده شده است.

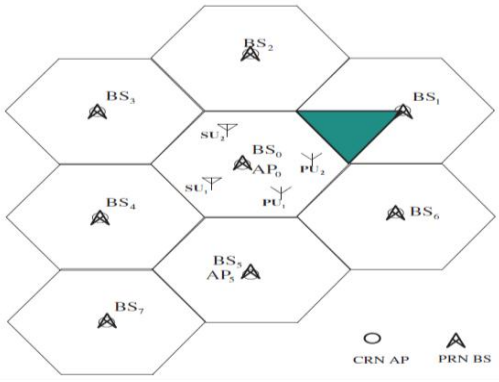


شکل ۱۰: معماری ترکیبی برای شبکه‌ی رادیوی‌شناختی[20]

اگر APها دارای توانایی‌های رادیوی‌شناختی باشند می‌توانند از حفره‌های طیف برای برقراری ارتباط با یکدیگر استفاده کنند. زیرا به دلیل بهره برداری فعلی ناکارآمد از طیف، حفره‌های زیادی در طیف قابل شناسایی می‌باشد. بنابراین ظرفیت اتصالات ارتباطی بیسیم بین APهای رادیوشناختی می‌تواند زیاد باشد و این مسئله، ستون فقرات سیمی را برای خدمت رسانی به ترافیک بیشتر، توانا می‌نماید[20].

نکته‌ای که باید مورد توجه قرار بگیرد منابع فرکانسی مورد استفاده‌ی کاربران شبکه‌های رادیوشناختی(کاربران ثانویه) می‌باشد، که در واقع فرکانس‌های مختص کاربران اولیه می‌باشد به عبارت دیگر به شبکه‌های رادیوشناختی هیچ فرکانسی اختصاص داده نمی‌شود و کاربران ثانویه با دسترسی پویا و به صورت فرصت طلبانه از منابع فرکانسی سایر کاربردهای بیسیم مانند پخش کننده‌های رادیویی و تلویزیونی، اپراتورهای رادیویی موبایل، شرکت‌ها، ارتش، آژانس‌های امنیت عمومی و ... استفاده می‌کنند.

در [21] از معماری متمرکز برای طراحی شبکه‌ی رادیوشناختی استفاده شده است، نکته‌ی مهمی که در این نوع معماری در نظر گرفته شده است محل قرار دادن نقاط دسترسی شبکه رادیوشناختی و منابع فرکانسی مورد استفاده‌ی کاربران آن می‌باشد. در این شبکه‌ی رادیوشناختی AP بر روی دکل‌های BS شبکه سلولی قرار داده شده و برای تامین منابع کاربران خود از کانال‌های فرکانسی مختص اپراتورهای رادیویی موبایل استفاده کرده است. در (شکل ۱۱) معماری این شبکه نشان داده شده است، CRN AP[۲۷] معرف نقطه دسترسی درشبکه‌ی رادیوشناختی و PRN BS[۲۸] معرف ایستگاه پایه در شبکه اولیه یا همان شبکه‌ی موبایل می‌باشد. از آنجا که در شبکه‌های سلولی به منظور افزایش ظرفیت از مفهوم بازیابی فرکانسی در زمان و مکان استفاده می‌گردد باید برای جلوگیری از تداخل هم‌فرکانس کاربران ثانویه علاوه بر اینکه باید در زمانی که کاربر اولیه در کانال حضور ندارد از کانال استفاده شود، همچنین از ایجاد تداخل هم‌فرکانس جلوگیری کند.



شکل ۱۱: زیرساخت سلولی برای شبکه‌ی رادیوشناختی[21]

۴-۲- دگرسپاری در شبکه‌های رادیوی‌شناختی

در یک شبکه‌ی رادیوشناختی درخواست دگرسپاری توسط کاربر اولیه و کاربر ثانویه صورت می‌گیرد. درخواست دگرسپاری توسط کاربر اولیه به دلیل کاهش کیفیت کانال ارتباطی و یا وجود کانالی با کیفیت بهتر می‌باشد، درخواست دگرسپاری توسط کاربر ثانویه به دلیل بازگشت کاربر اولیه، کاهش کیفیت کانال ارتباطی، وجود کانالی با کیفیت بهتر می‌باشد. اینکه در یک شبکه رادیوشناختی از چه روشی برای اختصاص دادن کانال به کاربران استفاده شود در کیفیت سرویس دهی و استفاده بهینه از طیف و همچنین تعداد دگرسپاری‌های موفق تاثیر می‌گذارد در همین راستا الگوریتم‌های بسیاری برای اختصاص کانال در شبکه‌های رادیوشناختی ارائه شده است. برای مثال در [22] برای اختصاص دادن کانال در شبکه‌های رادیوشناختی از الگوریتمی استفاده کرده است که در آن به یک درخواست دگرسپاری نسبت به درخواست تماس جدید، اولویت بیشتری داده است. با توجه به این که در شبکه‌ی رادیوشناختی دو نوع کاربر اولیه و ثانویه وجود دارد، اگر درخواست کانال برای دگرسپاری توسط کاربر اولیه را با PH، درخواست کانال برای تماس توسط کاربر اولیه را با PI، درخواست کانال برای دگرسپاری توسط کاربر ثانویه را با SH و درخواست کانال برای تماس توسط کاربر ثانویه را با SI نشان داده شود، اولویت‌بندی انجام شده برای پاسخ به درخواست کانال کاربران به صورت زیر می‌باشد:

PH > PI > SH > SI

که این نوع اولویت بندی منجر به استفاده‌ی بهینه از کانال و کاهش دگرسپاری‌های ناموفق می‌گردد. منتها میزان ترافیک موجود در شبکه را کاهش می‌دهد. که این به دلیل دادن اولویت کمتر به درخواست‌های تماس جدید نسبت به درخواست دگرسپاری می‌باشد. به منظور جلوگیری از این امر کانال‌ها را به دو دسته تقسیم کرده است: کانال‌های مختص دگرسپاری و کانال‌های مختص تماس‌های جدید. که اگر درشبکه درخواستی برای دگرسپاری باشد، باید ازکانال‌های مختص دگرسپاری استفاده کند. منتها اگر درخواستی برای تماس باشد، به جز اینکه می‌تواند از کانال‌هایی که مختص درخواست

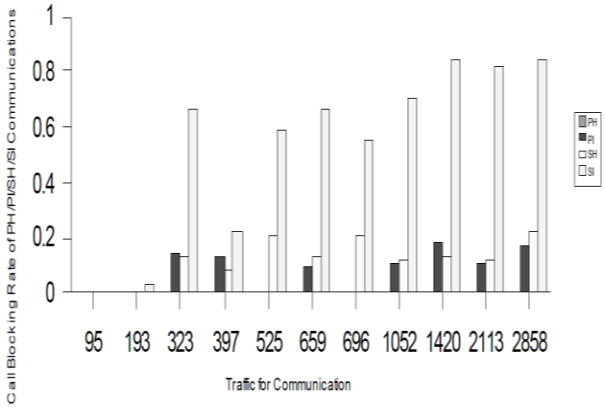
تماس می‌باشد، استفاده کند همچنین می‌تواند با درنظر گرفتن شرایط زیر، از کانال‌های مختص دگرسپاری نیز استفاده بکند:

۱. اگر زمان ملبین دو دگرسپاری موفق از یک حد آستانه‌ای بیشتر باشد.
۲. اگر فراوانی درخواست دگرسپاری کم باشد.
۳. اگر ترافیک از یک حد آستانه‌ای کمتر باشد، که به طبع آن میزان نیاز به دگرسپاری کم می‌شود.

(نمودار ۵) به مقایسه‌ی نرخ موفقیت این چهار نوع درخواست کانال(PH، IN، SH، SI) پرداخته است، که همان طور که دیده می‌شود بیشترین نرخ مربوط به درخواست دگرسپاری کاربر اولیه می‌باشد که این به دلیل دادن بیشترین الویت به این درخواست کانال می‌باشد.

نمودار ۵: مقایسه‌ی‌نرخ موفقیت‌درخواست‌های‌کانال‌توسط PH، IN، SH، SI [22]

(نمودار ۶) نشان می‌دهد که با الگوریتم پیشنهادی در این مقاله نرخ رد درخواست کانال برای دگرسپاری برای کاربر اولیه صفر می‌گردد و میزان رد درخواست کانال برای تماس توسط کلبر اولیه و نرخ درخواست کانال برای دگرسپاری توسط کاربر ثانویه با افزایش ترافیک شبکه تقریبا ثابت می‌ماند منتها نرخ درخواست دگرسپاری توسط کاربر ثانویه افزایش می‌یابد.



نمودار ۶: مقایسه‌ی نرخ رد درخواست‌های‌کانال‌توسط PH، IN، SH، SI [22]

۵- استفاده از رادیوی‌شناختی برای دگرسپاری بین شبکه‌ای در شبکه‌ی سلولی GSM[۲۹]

سیستم GSM یک شبکه موبایل زمینی است و شرکت‌های مختلف می‌توانند چنین شبکه‌ای را روی زمین نصب کنند و چندین شرکت می‌توانند همزمان با هم همکاری کنند. این شبکه از یک ساختار سلولی استفاده میکند. GSM سه بخش مشخص دارد:

- زیر سیستم رادیویی (RSS[۳۰]): مربوط به کلیه مسایل رادیویی.
- زیر سیستم شبکه و سوئیچینگ (NSS[۳۱]): کارهای انتقال مکالمه، دگرسپاری و سوئیچینگ را انجام میدهد.
- زیر سیستم عملیات: مدیریت شبکه را انجام میدهد.

در سیستم GSM برای برقراری ارتباطات اپراتورهای شبکه با منابع مختلف و تجهیزات زیرساختار سلولی، نه تنها رابط هوایی بلکه چندین رابط اصلی دیگر برای مرتبط کردن قسمت‌های مختلف این سیستم تعریف شده است، سه رابط مهم در GSM در زیر آمده است:

- رابط A که میان MSC[۳۲] و BSC[۳۳] قرار دارد.
- رابط Abis که میان BSC و BTS[۳۴] قرار دارد.
- رابط UM که میان BTS و MS قرار دارد.

زیر سیستم رادیویی GSM شامل تجهیزات و توابع مرتبط با مدیریت اتصالات مسیر رادیویی، مانند دگرسپاری‌ها می‌باشد. این زیر سیستم شامل BTS، BSC و MSC است. MS به طور قراردادی در زیرساخت رادیویی قرار گرفته است و همیشه آخرین مسیر یک مکالمه است. MS دارای قابلیت پایانه شبکه و همچنین پایانه کاربر است. هرسلول در سیستم GSM یک BTS با چندین گیرنده و فرستنده دارد. یک گروه از BTSها توسط یک BSC کنترل می‌شوند. BSC و BTS با هم به عنوان BSS شناخته می‌شوند. BSS از دید MSC به صورت یک رابط که ارتباطات لازم را با MS ها در حوزه‌ای مشخص برقرار می‌کند، به نظر می‌رسد. BSS به طور دائم با یک مدیریت کانال رادیویی، وظایف انتقال، کنترل لینک رادیویی، کیفیت و مهیاسازی سیستم برای دگرسپاری‌ها مرتبط است. هر MSC دارای یک پایگاه داده‌ی VLR و حداقل یک پایگاه دادهی HLR که از آنها برای دگرسپاری استفاده می‌کند[23].

در سیستم GSM اگر کاربری در حین مکالمه از مرزهای سلول عبور نماید بیشتر مایل خواهد بود که از منابع رادیویی در سلول جدید استفاده نماید. چون توان سیگنال فراهم شده توسط سلول قدیمی به دلیل دور شدن کاربر از ناحیه‌ی پوشش سلول قدیمی ضعیف شده است. کل فرایند قطع اتصال موجود با BTS سلول جاری و برقراری یک اتصال جدید با BTS مناسب، دگرسپاری نامیده می‌شود. براساس موقعیت و استفاده، چهار نوع مختلف دگرسپاری در سیستم‌های GSM استفاده می‌شوند (شکل ۱۲):

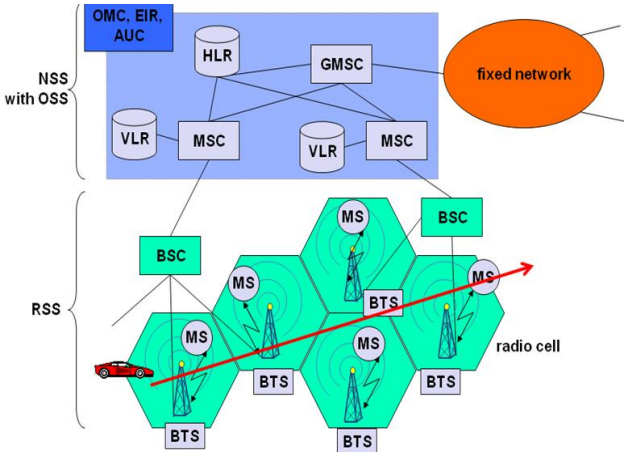
۱. دگرسپاری بین کانال‌ها (شکاف زمانی[۳۵]) در سلول یکسان یا دگرسپاری بین حاملی. در این مورد کاربر بین کانال‌های

ترافیکی مختلف داخل سلول یکسان انتقال می‌یابد. این کانال با فرکانس یا شکاف زمانی مختلف ایجاد می‌گردد. تصمیم درمورد دگرسپاری توسط BSC گرفته می‌شود که سلول را کنترل می‌کند.

۲. دگرسپاری بین سلول‌ها (BTSها) تحت کنترل BSC یکسان یا دگرسپاری داخلBSC.

۳. دگرسپاری بین سلول‌ها (BTSها) تحت کنترل BSC مختلف یا دگرسپاری بینBSC.

۴. دگرسپاری بین سلول‌های تحت کنترل MSCهای مختلف یا دگرسپاری بینMSC (در دگرسپاری بین MSC، اگر MSCها مختص شبکه‌های شرکت‌های مختلف می‌باشد دگرسپاری انجام شده از نوع دگرسپاری بین شبکه‌ای می‌باشد).



شکل ۱۲: دگرسپاری در شبکه‌ی GSM

در [23] یک روش دگرسپاری بین شبکه‌ای به منظور افزایش ظرفیت و کاهش قطع مکالمات در حال اجرا معرفی گشته است که در واقع نوعی دگرسپاری بین شبکه‌ای می‌باشد.فرض کنید در یک منطقه جغرافیایی(در سطح یک استان) سه شرکت مختلف مخابراتی A و B و C وجود دارند و با یکدیگر قراردادهای مربوط به رومینگ را هم تنظیم کرده اند، اگر درخواست دگرسپاری به دلیل جابه جایی یک کاربر که سیم کارتش را از شرکت مخابراتی A خریداری کرده صورت بگیرد اگر تمام کانال‌های سلول مقصد در شبکه‌ی A پر باشد ارتباط قطع می‌گردد، ولی اگر کاربر می‌توانست از کانال‌های مختص شبکه‌های دیگر که قابل استفاده مجدد هستند (یعنی استفاده از آنها منجر به ایجاد تداخل هم‌فرکانس در کاربران آن شبکه‌ها (کاربران اولیه) نمی‌گردد) استفاده کند تماس قطع نمی‌شد. برای حل این مشکل در این مقاله پیشنهاد شده است که به جای رادیوی نرم افزاری(SDR[۴۶]) که در موبایل‌های سیستم GSM از آنها استفاده می‌شود، از رادیوی‌شناختی استفاده شود. SDR یک رادیوی قابل برنامه‌ریزی می‌باشد که می‌تواند پارامترهای ارسال و دریافتش را به صورت پویا در طول یک مکالمه تغییر دهد و رادیوی‌شناختی یک رادیوی SDR هوشمند است که می‌تواند فرکانس‌های موجود در اطرافش را حس کند و به هر فرکانس آزاد موجود در اطرافش دسترسی پیدا کند.

رادیوی‌شناختی به منظور تامین هوش و دسترسی پویا به طیف چهار عمل را انجام میدهد: حس کردن طیف، اشتراک طیف، انخاب طیف، تحرک طیف. بنابراین زمانی که درخواستی برای تماس شکل می‌گیرد چه این تماس از طرف کاربر اولیه (کاربری که اولویت اول را برای استفاده از منابع فرکانسی شبکه اش دارد) باشد و یا کاربر ثانویه (کاربری که به صورت فرصت طلبانه از منابع فرکانسی شبکه‌های دیگر استفاده می‌کند)، رادیوی‌شناختی ابتدا بررسی می‌کند که آیا فرکانس خالی در شبکه‌ای که آن موبایل به آن تعلق دارد موجو د هست یا نه اگر نبود آنگاه عملیات زیر را به ترتیب انجام میدهد:

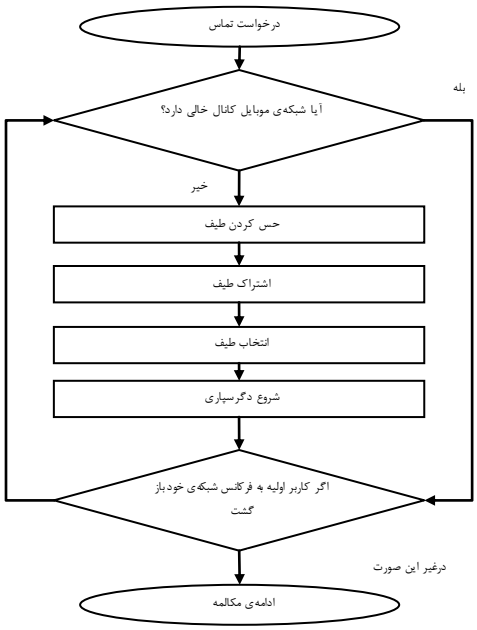
۱. حس کردن طیف: محیط اطرافش را حس می‌کند و اطلاعات مرتبط با فرکانس‌های مختلف را جمع آوری میکند و لیستی از فرکانس‌های خالی و شبکه‌های مرتبط با آنها را تهیه می‌نماید.

۲. اشتراک طیف: از آنجا که ممکن است چندین کاربر ثانویه همزمان متقاضی استفاده از فرکانس خالی مربوط به یک شبکه باشند در این مرحله بررسی‌های لازم به منظور جلوگیری از تداخل به اشتراک گذاری طیف صورت می‌گیرد.

۳. انتخاب طیف: با توجه به لیست فرکانس‌های قابل استفاده بهترین را انتخاب میکند.

۴. تحرک موبایل: در صورتی که موبایل حرکت کرد و یا اگر کاربراولیه به فرکانس شبکه‌ای خود را مورد استفاده قرار داد تمام مراحل را از ابتدا انجام می‌دهد.

(فلوچارت ۱) مراحل این الگوریتم را نشان می‌دهد. همان طور که در بالا گفته شد این نوع دگرسپاری، دگرسپاری بین شبکه‌ای می‌باشد بنابراین برای مسائل قانونی و هزینه‌ها لازم می‌باشد تا از رومینگ در این دگرسپاری استفاده شود.



فلوچارت ۱: الگوریتم دگرسپاری پیشنهادی[23]

۶- نتیجه گیری

در این سمینار مشکل کمبود پهنای باند در شبکه‌های سلولی بررسی گردید، از طرف دیگر مقالاتی در زمینه‌ی رادیوی‌شناختی و مسائل مطرح در رابطه با این تکنولوژی بررسی شد که از جمله، روش‌های دسترسی پویای طیف، ویژگی‌های روش‌های دسترسی روگستر و زیرگستر، شبکه‌ی رادیوشناختی، معماری‌های ارائه شده برای شبکه‌های رادیوشناختی و دگرسپاری در شبکه‌های رادیوشناختی مطرح گردید.

تاکنون در زمینه‌ی کاربرد تکنولوژی رادیوی‌شناختی برای دگرسپاری در شبکه‌های سلولی کارهای کمی انجام شده است که البته این مساله خود زمینه‌ی لازم را برای تحقیق بیشتر در رابطه با این موضوع فراهم می‌کند. تنها روش پیشنهاد شده در رابطه با استفاده از رادیوی‌شناختی برای دگرسپاری در شبکه‌های سلولی روش پیشنهادی در [26] می‌باشد که برای دگرسپاری بین شبکه‌ای از دگرسپاری رادیوشناختی در شبکه‌های سلولی GSM استفاده کرده است. با توجه اینکه نسل سوم موبایل تنها از یک طیف فرکانسی در سراسر شبکه استفاده می کند قابلیت استفاده از تکنولوژی رادیوی‌شناختی را ندارد اما نسل دوم و نسل چهارم موبایل این عیب را نداشته و امکان به کار گیری این تکنولوژی را دارند.

با توجه به توانایی‌های رادیوی‌شناختی این امکان وجود دارد که بتوان از روش‌های رادیوی‌شناختی برای تامین منابع مورد نیاز برای دگرسپاری در شبکه‌های سلولی نسل آینده استفاده کرد که میتوان این نوع دگرسپاری را دگرسپاری رادیوشناختی نام گذاری کرد.

مراجع

[1] Qing Zhao, Sadler, B.M., "*A Survey of Dynamic Spectrum Access*" Signal Processing Magazine, IEEE , vol.24, no.3, pp.79-89, May 2007.

[2] Mitola, J., Maguire, G.Q., Jr., "*Cognitive radio: making software radios more personal*" Personal Communications, IEEE , vol.6, no.4, pp.13-18, Aug 1999.

[3] FCC Spectrum Policy Task Force, *"Report of the Spectrum Efficiency Working Group"*, Tech. rep. 02-135, Nov2002.

[4] Haykin, S., *"Cognitive Radio: Brain-Empowered Wireless Communications"* IEEE JSAC, vol. 23, pp. 201–20, no. 2, Feb. 2005.

[5] Akyildiz, I.F., Lee, W.Y., Vuran, , M.C., Mohanty, S., *"A survey on spectrum management in cognitive radio networks"* IEEE Communications Magazine, vol 46, pp. 40-48, April 2008.

[6] Ahmed, Rehan, Arfat Ghous, Yasir, *"detection of vacant frequency bands in cognitive radio"*, Blekinge Institute of Technology Ph.D. thesis, May 2010.

[7] McDonald, V.H., *"The cellular concept "*, Bell Syst Tech. J., vol.58, pp.15-41, Jan 1997.

[8] Everitt, D., Manfield, D., *"Performance analysis of cellular mobile communication system with dynamic channel assignment"*, IEEE J. Selected Area Commun., vol. 7, no. 8, pp. 1172-1180, Oct.1989.

[9] Del Re, E., Goodman, D. J., *"dynamic resource acquisition: Distributed carrier allocation for TDMA cellular systems"*, Glob Com Journal, pp.1698-7102, 1993.

[10] Bauer, Carolin, Rees, S. John , *"classification of Handover schemes within a cellular Enviroment "*, IEEE Journal, Feb. 2002.

[11] Neel, J.d., *"Analysis And Design of Cognitive radio Networks and Distributed Radio Resource Management Algorithms"*, Virginia Polytechnic Institute and state university Ph.D. Thesis, September 2006.

[12] Fette, B., *"Introduction to Cognitive Radio"*, SDR Forum Technical Conference 2005, pp. 14-17, Nov 2005.

[13] Nekovee, M., *"Dynamic Spectrum Access Concept and Future Architectures"*, BT Technology Journal, April 2006.

[14] Akyildiz, I.F., Lee, W.Y., Vuran, , M.C., Mohanty, S., *"Next Generation Dynamic Spectrum Access Cognitive Radio Networks: A survey"*, Computer Networks Journal (Elsevier), Issue 13,50,pp. 2127-2159,Ep2006.

[15] Marinho, Jose, Monterio, Edmondo, *"Cognitive Radio: survey on communication protocols, spectrum decision issues, and future research directions"*, Wireless Netw Journal, pp. 147-164, 2012.

[16] Danda, B. Rawt, Gongjun, Yan *"Introduction to cognitive radio"*, SDR Forum Technical Conference, Nov.2005.

[17] Im, Sooyeol, Kim, Wonsop, Kang Yunsuk, Lee, Hyuckjae, *"Joint Power and Addmition Control for UnderlaySpectrum Sharing in Cognitive Radio Networks"*, IEEE conference on Advance Technology for Communications, pp.56-61, 2010.

[18] Menon, Rekha, Buehrer, R. Michael, *"Outage Probability based Comparison of Underlay and Overlay Spectrum Sharin Techniques "*, IEEE Journal, pp.101-109, 2005.

[19] Hu, Han, Zhu, Qi, *"Dynamic Spectrum Access inn Underlay Cognitive Radio System with SINR constraints"*, IEEE Journal, 2009.

[20] Chen, K., Peng, Y.J., Prasad, N., Liang, Y.C., Sun, S., *"Cognitive Radio Network Architecture: Trusted Network Layer Structure"*, National Science Council Journal, 2008.

[21] Ma, Yao, In Kim, Dong, Leith, Alex, *"Weighted Sum Rate Optimization of Multi cell Cognitive Radio Networks"*, Glob Com Journal, 2008.

[22] Viz, Chhavi, Udgata, Siba K., *"Spectrum hand-off schemes and optimum utilization of spectrum holes in Cognitive Radio Networks"*, IEEE Journal, pp.181-186, 2008.

[23] Meghana, Talasila, Praneeth, L., Sindhu Sravya, p., Apuroopa, K., *"Inter Network Handover Using Cognitive radio"*, International Journal of engineering science & Advanced Technology, Vol. 2, pp.128-132, Feb. 20012.

---

[1] Hand Over

[2] Down Link

[3] Up Link

[4] Break Before Make

[5] Make Before Break

[6] Home location Register

[7] Visitor Location Register

[8] Quality Of Service

[9] Federal Communications Commission

[10] International Telecommunication Union

[11] Dynamic spectrum access
[12] Dynamic Exclusive Use
[13] Spectrum Priority Right
[14] Dynamic spectrum Allocation
[15] Open Sharing Model
[16] Hierarchical Access Model
[17] Overlay
[18] Underlay
[19] Brute Force Algorithm
[20] Quality Of Service
[21] Hybrid
[22] Signal to Interference Plus Noise Ratio
[23] Mobile Station
[24] Access Point
[25] Back Bone
[26] Mesh Hybrid
[27] Cognitive Radio Network Access point
[28] Primary Network Access Point
[29] Global System for Mobile communication
[30] Radio Sub System
[31] Network and Switching Subsystem
[32] Mobile Switching Center
[33] Base Station Controller
[34] Base Transceiver Station
[35] Time Slot
[36] Software Defined Radio

بررسی پیشرفت‌های اخیر در طراحی مدارات اتوماتای سلولی کوانتومی(QCA)

معصومه هادیان امیری
کارشناسی ارشد معماری سیستم‌های کامپیوتری ،دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت
تهران ، ایران
masoumehadiyanamiri@comp.iust.ac.ir
دکتر محسن سریانی
استادیار ،دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت
تهران ، ایران
soryani@iust.ac.ir

چکیده

محدودیت‌های فیزیکی فناوری CMOS و سایر فناوری های متداول، در سطح نانو و هزینه بالا لیتوگرافی در این سطح، زمینه را برای ساخت عناصر سخت افزاری مبتنی بر اتوماتای سلولی کوانتومی فراهم آورده است. این فناوری، برخلاف CMOS که مبتنی بر جریان الکتریکی است، از اصل برهم‌نهی در ذرات کوانتومی کم انرژی استفاده می‌کند بنابراین به‌طور ذاتی کم مصرف اســتبا بکــارگیری قوانین موجود در طراحی مقاوم در QCA و ساده‌سازی این مدارات و کم کردن تعداد لایه‌های بکارگرفته شده در طراحی اتوماتای سلولی کوانتومی امکان دستیابی به اهداف ساخت مدارات VLSI از قبیل:توان مصرفی پایین، سرعت بالا، چگــالی زیــاد و آســانی ساخت و آزمون‌پذیری، آسان‌تر خواهد شد.

اتوماتای سلولی کوانتومی درحوزه بهبود مدارات به طراحی‌های مختلفی دست یافته که بهینه بودن این مدارات از نظر تاخیر، ناحیه مورد نیاز برای پیاده‌سازی و تعداد سلول‌های به کار برده شده در سطح مدار و تعداد لایه‌ها اهمیت فراوانی دارند. همچنین در حوزه حافظه و مدارات ترتیبی، بوجود آمدن مدارات DET(Dual Edge Triggered) امکان گذردهی داده بیشتر و بهبود ســرعت و بهبود کارآیی را در سطح مدارت اتوماتای سلولی کوانتومی افزایش داده است. در حوزه تحمل پذیری اشکال، قوانین بوجود آمده در افزایش روز افزون مقاومت و کم شدن نرخ خطا در مدارت QCA تاثیر فراوان دارد. یکی از اهداف مهم در ساخت مــدارات حافظــه اتوماتای سلولی کوانتومی که باعث بکارگیری آسان آن در مدارات بزرگتر و پیچیده می‌شود، ساخت آن‌ها به شکلی است که ورودی-ها در یک سمت و خروجی‌ها در سمت دیگرمدار قرار گیرد. با دنبال کردن این اهداف امید است روزی QCA جایگزین CMOS باشد.

کلمات کلیدی

اتوماتای سلولی کوانتومی، نانوتکنولوژی، حافظه ترکیبی، تقاطع همسطح،تمام جمع کننده QCA، تفریق کننده QCA.

۱ – مقدمه

اتوماتای سلولی کوانتومی ابزاری برای ارائه داده‌ها و اجرای محاسبات بر پایه خاصیت کوانتومی ذرات باردار می‌باشد و از اصل برهم نهی تبعیت می‌کند. ایده اتوماتای سلولی کوانتومی امیدبخش‌ترین جایگزین برای فناوری CMOS است. افزایش چگالی مدارات با فناوری اتوماتای سلولی کوانتومی چندین برابر مدارات متداول CMOS خواهد بود و سرعت سویچینگ بالا و مصرف توان پایین باعث شده است اتوماتای سلولی کوانتومی در صدر تحقیقات پژوهشی قرار بگیرد. تا کنون مدارات زیادی بر پایه اتوماتای سلولی کوانتومی طراحی شده است و پیاده سازی‌های مختلفی در ساخت سلول اتوماتای سلولی کوانتومی پیشنهاد شده و از نظر کارآیی و توان و سرعت آن مورد بررسی قرار گرفته است [1].

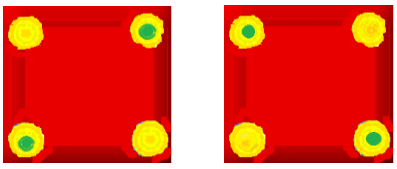
با گذشت زمان، مقیاس ساخت مدارات CMOS رو به کاهش بوده است و هم اکنون به مرحله‌ی نانومتری رسیده است. به عبارت بهتر روند پیشرفت مدارات به منحنی قانون مور نزدیک است. اما با ورود به حوزه‌ی نانومتری، مشکلات خاصی مانند جریان نشتی، مصرف بالای توان، نوسان در تزریق اکسید گیت و نویزپذیری بالا بروز می‌کنند [2,3]. از آن جا که این مشکلات ناشی از ماهیت فیزیکی این نوع مدارات و همچنین محدودیت‌های این فناوری می‌باشد، شاید بتوان گفت کاراترین راه حل برای این مشکل، استفاده از بستری جدید به جای CMOS می‌باشد. هدف از انجام این سمینار آشنایی با انواع سلول های اتوماتای سلولی کوانتومی و بررسی مدارات طراحی شده مختلف و بهبود آنها می‌باشد [2,1]. در سال‌های اخیر پیشرفت‌های زیادی در زمینه کم کردن تعداد سلول‌ها و افزایش سرعت مدارات QCA صورت گرفته است که بررسی این پیشرفت‌ها به ارائه ایده‌ایی برای بالا بردن لبه‌های علم در حوزه اتوماتای سلولی منجر می‌شود [4,5,6,7,8]. در بخش۲ با ساختار این سلول‌ها و نقاط کوانتومی درون آن‌ها آشنا خواهیم شد. در این بخش ساختار سلول، سیم، گیت اکثریت و معکوس‌کننده توضیح داده شده است. در بخش ۳ مدارات ترکیبی که بیشترین توجه را در تحقیقات QCA به خود اختصاص داده را معرفی کردیم. در بخش ۴ توضیح مختصری بر نحوه تولید ساعت مدار داده شده و به روشی جدید در تولید ساعت برای تقاطع همسطح با یک سلول اشاره کردیم. در بخش ۵ به بررسی حافظه‌ها پرداختیم که در QCA زمینه فراوانی برای تحقیقات جامع وگسترده برای آن وجود دارد. در بخش ۶ به انواع خطا و معرفی گیت اکثریت تحمل‌پذیر اشکال پرداختیم. در بخش ۷ ابزارهای موجود برای QCA معرفی شده است و در آخر کارهای آتی و تحقیقات ممکن بیان شده است.

۲ – معرفی پایه ای اتوماتای سلولی کوانتومی

عناصر پایه که در این بخش معرفی می‌شود برای ساخت مدارات ترکیبی و ترتیبی به کار گرفته می‌شود تا کنون مداراتی چون جمع-کننده، مالتی‌پلکسر، رمز گشا، انکدر، و ALU و FPGA در QCA طراحی شده اما در حوزه مدارات ترتیبی تحقیقات زیادی صورت نگرفته است. تا کنون مدارات D-FF وJK-FF نیز طراحی شده است. این مدارات به صورت SET(Single edge triggered) و (Dual DET edge triggered) پیاده سازی شده است و از نظر تاخیر و تعداد سلول‌ها و میزان گذردهی داده در آن بهبود حاصل شده است [9]. در این بخش عناصر پایه در QCA که برای ساخت مدارات ذکرشده لازم است، معرفی شده است.

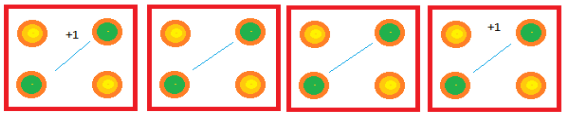
۲-۱ – معرفی عناصر ابتدای QCA

ساده‌ترین عنصر در QCA یک سلول به شکل مربع است که چهار نقطه کوانتومی با کمینه انرژی در چهار گوش این مربع قرار دارد و دارای بار معادل ۰/۵+ هستند. دو الکترون با بار ۱- در این سلول طوری در این نقاط کوانتومی قرار می‌گیرند که، همواره دارای کمینه انرژی باشند. در نتیجه دو حالت معتبر برای یک سلول وجود خواهد داشت که معرف منطق ۰ و ۱ خواهد بود [7]. شکل۱ منطق ۰ و ۱ را در این نوع سلول ساده نشان می‌دهد.



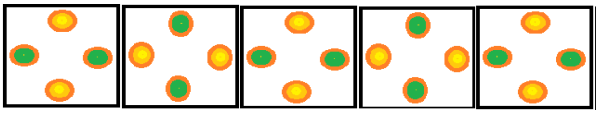
شکل(۱):نمای سمت چپ مربوط به منطق ۱ و پلاریته ۱+ و نمای سمت چپ مربوط به منطق ۰ و پلاریته ۱- است.

بدلیل خاصیت رانش و نیروهای کوانتومی، کنار هم قرار دادن این سلول‌ها به صورت خط، پلاریته یک سلول به سلول انتهایی انتقال می-یابد و یک سیم QCA ساخته می‌شود (شکل ۲). نوعی دیگر از سیم در QCA وجود دارد که سلول‌های آن یکی در میان به صورت چرخیده است که از نظر سختی در فرآیند ساخت برای جانشانی سلول چرخیده و افزایش میزان تخریب و خطای عدم همترازی[۱] استفاده از آن در مداراتی که با هدف مقاوم بودن و تحمل‌پذیری بالا در مقابل اشکال هستند توصیه نمی‌شود (شکل ۳) [10].



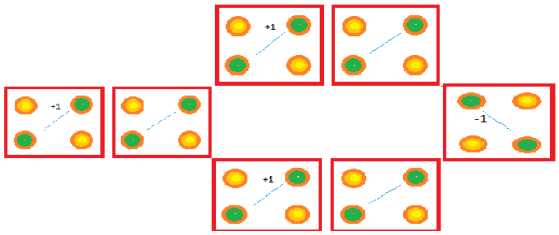
شکل (۲): قطعه سیم QCA با سلول ۹۰ درجه

[۱] Misalignment

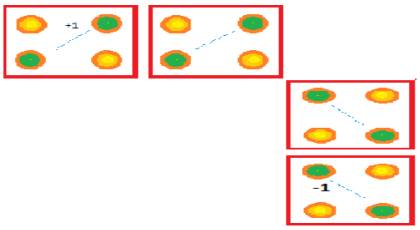


شکل (۳): قطعه سیم QCA با سلول ۴۵ درجه

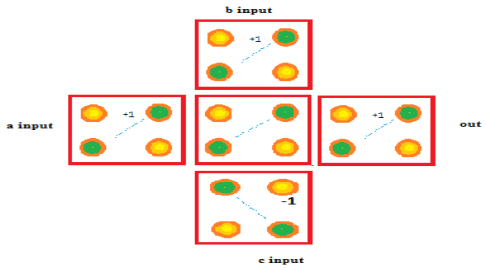
با استفاده از جابجایی سلول‌ها و ایجاد چینش جدید به مدارات پایه دیگری دست می‌یابیم، از جمله معکوس‌کننده و گیت اکثریت که در شکل ۴ دیده می‌شوند.



۴ الف)گیت معکوس کننده مقاوم



۴ب)گیت معکوس کننده ساده



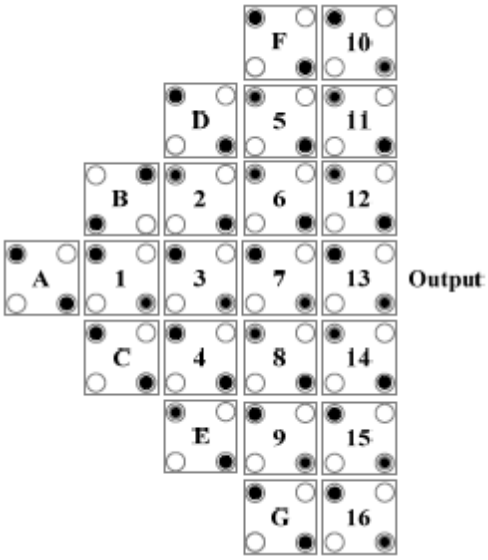
۴چ)گیت اکثریت سه ورودی

شکل(۴): گیت‌های معکوس‌کننده QCA

در گیت سه ورودی اکثریت (شکل۴)، سلول A، B و C به عنوان ورودی می‌باشند و سلول F به عنوان خروجی، بر اساس قطبش اکثریت سلول‌های ورودی، تعیین حالت می‌شود. در حقیقت گیت اکثریت توسط پنج سلول QCA به صورت یک + پیاده سازی می‌شود که معادل با رابطه‌ی F(A,B,C) = AB + AC + BC می‌باشد. همچنین می‌توان یکی از ورودی‌ها را به عنوان پایه‌ی کنترلی در نظر گرفت تا از گیت اکثریت یک گیت AND یا OR ساخته شود. با ثابت نگاه داشتن قطبش یک سلول ورودی به صورت ۰ یا ۱ منطقی، می‌توانیم گیت OR یا AND را داشته باشیم. لذا، مدارات منطقی پیچیده‌تر می‌توانند از گیت‌های AND و OR ساخته شوند [11].

۲-۲- گیت‌های اکثریت با ورودی‌های بیشتر

در این مقاله یک گیت اکثریت هفت ورودی طراحی شده است. در کارهای قبلی گیت های اکثریت ۵ ورودی پیشنهاد شده بود [11]. پیاده سازی QCA این گیت را در شکل ۵ نشان داده ایم.



شکل(۵): گیت اکثریت ۷ ورودی [11]

با استفاده از گیت اکثریت هفت ورودی می‌توان گیت AND و OR چهار ورودی را پیاده سازی نمود. برای پیاده سازی AND چهار ورودی کافیست تا سه ورودی را به صفر ست کنیم و برای OR چهار ورودی سه ورودی را به یک ست کنیم. رابطه ۱ و ۲ را در رابطه با AND و OR مشاهده کنید.

$$M(A,B,C,D,0,0,0)=ABCD \quad (۱)$$

$$M(A,B,C,D,1,1,1)=A+B+C+D \quad (۲)$$

با استفاده از این گیت‌های جدید می‌توان مدارات پیچیده را ساده سازی نمود و تاخیر و فضای فیزیکی[۲] را کاهش داد.

برای نشان دادن کاربرد این گیت‌ها باید برخی از توابع مانند XOR و MUX را ساده سازی نمود [11].

۲-۳- تقاطع سیم در QCA

در QCA، دو گزینه برای تقاطع سیم‌ها وجود دارد. یکی تقاطع همسطح و دیگری تقاطع چند لایه می‌باشد. در تقاطع همسطح، از هر دو نوع سلول معمولی و چرخیده نیاز است که برای تحمل پذیری اشکال مفید نیست. وقتی که هر دو نوع سلول به درستی جایگذاری شده باشند، با هم تعاملی نخواهند داشت و لذا تقاطع نیز به طور صحیح عمل می‌کند. گزارش‌های منتشر شده بیان می‌کنند که تقاطع های همسطح ممکن است به عدم جایگذاری صحیح خیلی حساس باشند [8]. در یک تقاطع همسطح، احتمال *انقیاد ضعیف*[۳] سیگنال و همچنین بازخورد (پسروی) از سمت دورتر وجود دارد.

---

[۲] Area

[۳] Loose Binding

ممانعت از این امر مستلزم ناحیه‌های بیشتری از پالس ساعت بین سلول‌های معمولی و چرخیده می‌باشد. یک نکته‌ی مهم این است که تعداد تقاطع‌ها در مدارات QCA دارای محدودیت است [8].

تقاطع چند لایه اگر چه از نظر مفهومی ساده است ولی در اصل این مقوله سوالاتی مطرح می‌شود. چرا که نیازمند آن است که دو لایه ی فعال با استفاده از اتصالات عمودی همپوشانی داشته باشند. کارهای پیشین احتمال وجود مدارات QCA چند لایه را بررسی کرده ولی هنوز هیچ گزارشی از پیاده سازی آن به دست نرسیده است.

در [8] روشی در تقاطع هم‌سطح به کار گرفته که تنها از سلول ۹۰ درجه استفاده شود و با استفاده از ابزاری که تهیه شده است ساخت مدارات را با این روش ممکن ساخته که از ابزار QCA Designer بهتر است و در برخورد با سه یا بیشترتقاطع هم‌سطح درست عمل خواهد کرد.

۳- مدارات ترکیبی در QCA

در این بخش، نتایج تحقیقات اخیر در رابطه با ساخت و بهبود جمع کننده و تفریق کننده و مالتی پلکسر و رمز گشا در QCA بررسی می-شود.

۳-۱- تمام جمع کننده و سلول‌های جدید QCA

در [2] سلول ۳ بعدی QCA پیشنهاد شده و با آن یک گیت اکثریت ۵ ورودی طراحی شده است. در [11] یک گیت جدید ۵ ورودی با سلول‌های ابتدایی پیشنهاد شده است که یکی از ورودی‌ها از یک سمت و ورودی‌های دیگر از دو سو بر خروجی تاثیر می‌گذارند، خروجی این گیت با سلول‌های دیگر محاط نشده‌است، بنابراین براحتی در دسترس است و برای استفاده از خروجی نیاز به سیم‌کشی اضافه و تقاطعی نخواهد بود. با استفاده از این رأی‌گیراکثریت در [12] یک تمام جمع کننده معرفی شده است. سپس دو تمام جمع کننده با استفاده از این گیت معرفی شده است که یکی از آنها دارای ساختار قابل اطمینان[۴] است [12].

البته در ساخت این سلول سه بعدی چالش وجود دارد که باتوجه به تحقیقات در این زمینه،ایده اضافه کردن میدان توسط مدار از نظر من عملی خواهد بود. در شکل ۷ نمایش میدان‌های سلول سه بعدی را می‌بینیم. این بحث بیشتر مبنای فیزیکی دارد و در اینجا بیان نشده است.

شکل ۶a طراحی گیت اکثریت را نشان داده است و شکل۶b همان گیت را با قوانین طراحی [10] پیاده سازی کرده است. بنابراین طراحی شکل ۶b برای ساخت جمع کننده‌ای که از گیت‌های اکثریت ۳ و ۵ ورودی و گیت NOT استفاده می‌کند مناسب است، در این جمع کننده معکوس نقلی به عنوان ۲ ورودی گیت اکثریت استفاده شده است [2]. در شکل ۶a مدار دارای دو ناحیه ساعت است و برای اطمینان پذیر ساختن مدار طبق قوانین [6,10] شکل ۶b با سه ناحیه ساعت طراحی می‌شود. بین هر ناحیه ساعت با ساعت بعدی ۹۰ درجه اختلاف فاز وجود دارد. اولین تمام جمع کننده در سال ۱۹۹۴ پیشنهاد شد [12] که ترکیب از ۵ تا گیت اکثریت ۳ ورودی و ۲ معکوس کننده بود (شکل ۸).

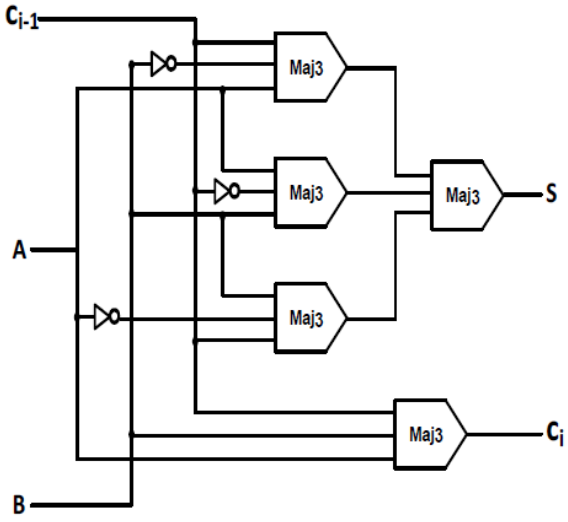
این تمام جمع کننده از تقاطع همسطح استفاده می‌کند و در یک لایه با ۱۹۲ سلول QCA پیاده سازی شده است. در این طراحی نحوه تولید ساعت در نظر گرفته شده است.

در [3] تمام جمع کننده دیگری با همین ساختار با در نظر گرفتن نحوه تولید ساعت پیشنهاد شده و با ۱۴ فاز (3.5 Cycle) خروجی را تولید می‌کند.

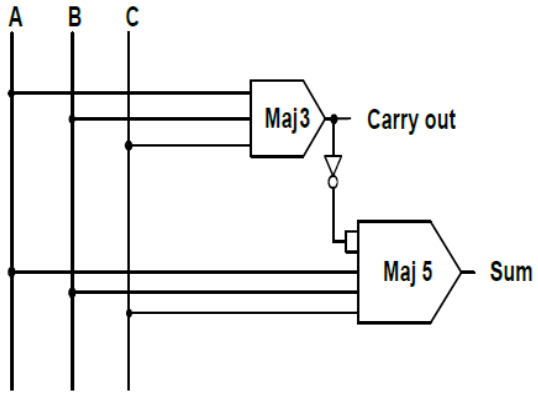
یک تمام جمع کننده ساده‌تردر [13] پیشنهاد شده که در آن از ۳ گیت اکثریت سه ورودی و ۲ معکوس کننده ساخته شده است، و در آن از تقاطع همسطح استفاده شده است و با ۵ فاز (1.5 cycle) خروجی را تولید می‌کند.

در [2] تمام جمع کننده دیگری که از دو گیت اکثریت و یک معکوس کننده تشکیل شده، معرفی شده است. این تمام جمع کننده از سلول‌های متداول QCA استفاده نمی‌کند. در شکل ۹ شماتیک این تمام جمع کننده نشان داده شده است [2]. جمع کننده دیگری که در شکل ۹ نشان داده‌ایم در [7] معرفی شده است. در مقایسه با [2] از سلول‌های معمولی QCA استفاده کرده است. در این مقاله از تقاطع چند لایه استفاده می‌شود. این طراحی با وجود اینکه در مقایسه با [2] از سلول ساده QCA استفاده می‌کند و با طراحی‌های قبلی از نظر ناحیه فیزیکی و تاخیر برابری می‌کرد، اما در طراحی‌های قبلی نامناسب بود و جمع از سمت راست خارج می‌شد و نقلی از سمت چپ، بنابراین طراحی مداری بر پایه این جمع کننده نیاز به تقاطع و سیم اضافی زیادی داشت. این امر پیچیدگی مدار را افزایش می‌دهد و طراحی کارآمد را غیر ممکن می‌ساخت. ولی در طراحی [12] ورودی‌ها در وسط قرار داشت و براحتی در دسترس نبود و در [12] خروجی‌ها با سلول‌های دیگر محاط شده و قابل دسترسی نیست. بنابراین [12] چندان برای پیاده‌سازی مدارات بزرگتر کارا نیستد. این تمام جمع کننده تمام طراحی‌های قبلی را پوشش می‌دهد و در این طراحی برخلاف طراحی‌های قبلی از تقاطع همسطح استفاده شده است. این جمع کننده تعداد سلول، ناحیه و تاخیر تمام جمع کننده‌های دیگر قبلی را پوشش می‌دهد اما در ۳ لایه پیاده سازی شده است که برای پیاده سازی عملی مفید نمی‌باشد.

به دنبال آن ۲ تمام جمع کننده دیگر با گیت اکثریت ۵ ورودی جدید معرفی شده است. در این دو طراحی ورودی و خروجی با سلولهای دیگر محاط نشده و خروجی ها از یک سمت بیرون می آید.



شکل (۸): تمام جمع کننده که در سال ۱۹۹۴ پیشنهاد شد ترکیبی از ۵ تا گیت اکثریت ۳ ورودی و ۲ معکوس کننده است [12]

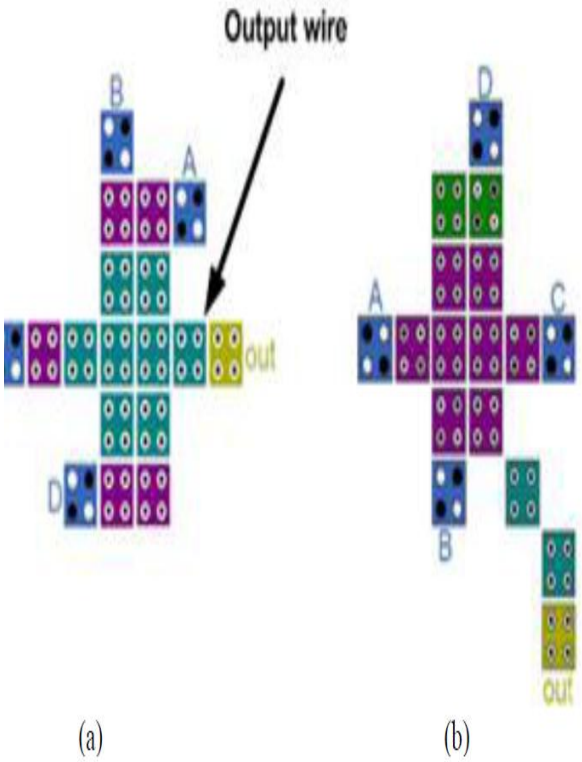


شکل (۹): شماتیک مداری تمام جمع کننده با سلول ۳ بعدی [2]

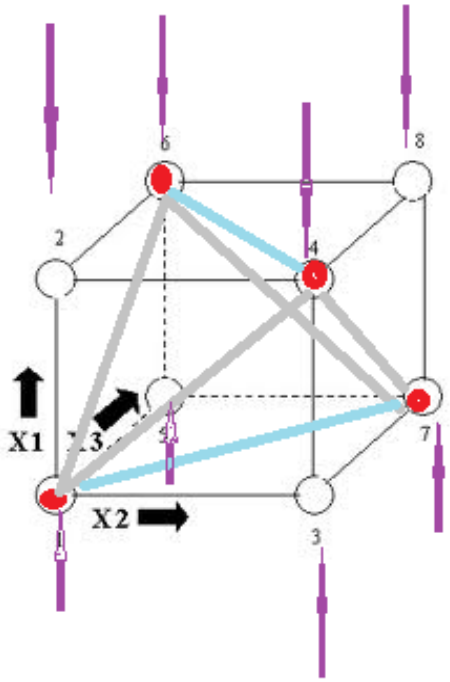
۳-۱-۱- تمام جمع کننده جدید مبتنی اتوماتای سلولی کوانتومی

با استفاده از گیت اکثریت ۵ ورودی جدید اتوماتای سلولی کوانتومی که در شکل ۶ a ,b نشان داده شده است، ساخت مدار تمام جمع کننده کارا امکان پذیر است. طرح سلول پیشنهادی با طرح‌های قبلی که معرفی شده، متفاوت است. در این طرح از طراحی چند لایه استفاده شده که در شکل ۱۰ دیده می‌شود. این تمام جمع کننده ساختاری ساده دارد و از گیت اکثریت شکل ۶ استفاده می‌کند.

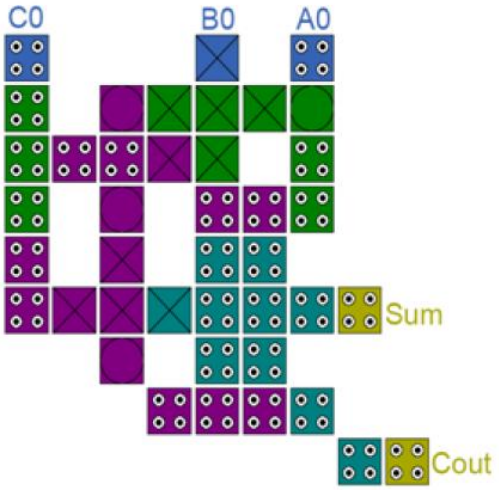
در ابتدا معکوس نقلی محاسبه می‌شود و به عنوان دو ورودی گیت اکثریت به کار گرفته می‌شود [12].



شکل (۶): شکل a یک گیت اکثریت است و با استفاده از قوانین [10,6] به صورت شکل b در آمده است [12]



شکل(۵): نمایش سلول سه بعدی و میدان از سوی مدار نحوه تولید ساعت برای پیاده سازی عملی

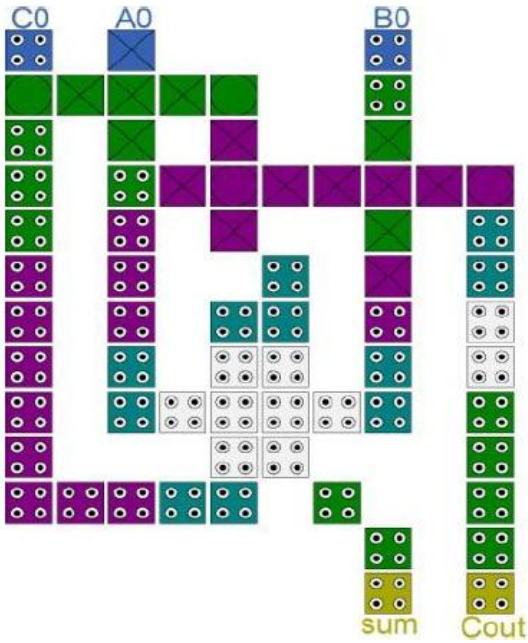


شکل (۱۰): طرح تمام جمع کننده جدید [12]

در [12] سه لایه این مدار درج است. در لایه اول یک گیت اکثریت ۵ ورودی و ۲ گیت معکوس کننده به صورت قطری قرار دارند. یکی از این معکوس کننده‌ها، معکوس نقلی را تولید می‌کند و دیگری خروجی را تولید می‌کند. معکوس نقلی به عنوان دو ورودی از گیت اکثریت به آن اعمال می‌شود [12].

لایه دوم تنها شامل ۴ سلول دایره‌ای شکل برای انتقال نقلی و A0 است. در لایه سوم دو سیم متقاطع قرار دارد (سلولهای X شکل). در شکل ۱۱ این طرح را به صورت تحمل‌پذیر طراحی کرده‌ایم [7]. در این طرح از گیت اکثریت شکل 6b استفاده شده است. در این طرح خروجی نقلی از Maj(A0, B0,C0) مستقیم بدست می‌آید [12].

اولین سلول QCA به صورت قطری قرار گرفته تا بیشترین فاصله را بین ورودی و خروجی ایجاد کند. نتایج شبیه سازی نشان می‌دهد که این کار نویز را کمتر می‌کند. در ۳ لایه مدار آن، یک گیت اکثریت ۵ ورودی، ۱ گیت اکثریت ۳ ورودی و ۳ سلول قطری که کار معکوس کننده را انجام می‌دهد، قرار دارد. یک سلول کار معکوس کردن نقلی و دو سلول دیگر برای ایجاد جمع به کار گرفته می‌شود [12].



شکل (۱۱): طرح تمام جمع کننده جدید تحمل‌پذیر [12]

دو تمام جمع کننده پیشنهاد شده [12] نتایج دقیقی را تولید می‌کنند. در یک آرایه شبیه سازی همروند طراحی تحمل پذیر اشکال از شکل ۱۲ قابل اطمینان‌تر است [7]. با استفاده از این تمام جمع کننده طراحی جمع کننده نقلی موجی با سایر کلمه‌های متفاوت امکان پذیر است.

ساختار یک جمع کننده نقلی موجی چهار بیتی با استفاده از طراحی تحمل پذیر تمام جمع کننده پیشنهادی طراحی شده است. جمع کننده نقلی موجی ۸ و ۱۶ بیتی نیز طراحی شده و با ساختارهای مقایسه شده است. خروجی این مدار تمام جمع کننده پس از دومین لبه پایین رونده ساعت ۳ ممکن خواهد بود.

مقایسه کاملی از طرح پیشنهادی با تمام جمع کننده های قبلی داده شده است. در [12] تمام جمع کننده نقلی موجی تحمل پذیر [12] را با بهترین تمام جمع کننده نقلی که از طراحی تحمل پذیر [14] استفاده می‌کند مقایسه شده است.

۳-۲- طراحی تفریق کننده با اتوماتای سلولی کوانتومی

در [4] نیم تفریق کننده و تمام تفریق کننده طراحی و شبیه سازی شده است. تفریق کننده یکی از اصلی ترین بلاک‌های ساختاری معماری است. در این تفریق کننده برای کم شدن تعداد سلول‌های طراحی از گیت اکثریت استفاده کرده است [4].

۳-۲-۱- نیم تفریق کننده

یک نیم تفریق کننده یک مدار ترتیبی است که تفاضل دو بیت را تولید می‌کند. یک خروجی نیز برای مشخص کردن قرضی دارد.

عبارت گیت اکثریت تابع در طراحی مورد نیاز است.یک تمام جمع کننده از قرضی مرحله قبل برای انجام محاسبات استفاده می-کند.

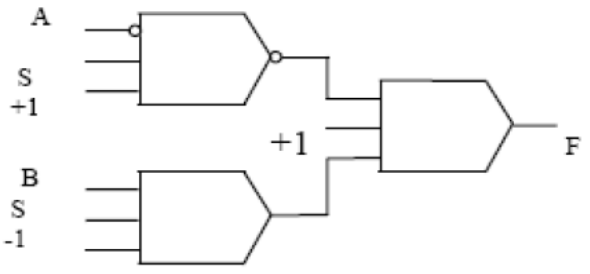
تمام جمع کننده‌ای با استفاده از الگوریتم تفریقی که در [4] بیان شده است، پیشنهاد شد [4].

۳-۳- مالتی‌پلکسر در QCA

روش‌های اخیر پیشرفت‌های زیادی را برای کم کردن تاخیر، فضا و تعداد سلول‌ها، در مالتی‌پلکسر اتوماتای سلولی کوانتومی ایجاد کردند. اما مالتی‌پلکسر پیشنهادی [5] در مقایسه با طرح های پیشنهادی قبلی پیچیدگی کمتری از نظر فضا وتاخیر دارد.

به طور کلی یک مالتی‌پلکسر دو به یک دارای دو ورودی و یک خروجی و یک خط انتخاب است که به وسیله آن یکی از دو ورودی به خروجی هدایت خواهد شد.

طرح مالتی‌پلکسر پیشنهادی [5] از دو گیت رای گیر اکثریت و یک گیت رای گیر اقلیت تشکیل شده است که طرح آن در شکل ۱۲ نشان داده شده است.



شکل (۱۲): طرح مالتی‌پلکسر

ورودی A قبل از رسیدن به گیت اقلیت معکوس می‌شود. برای OR بین A' و s که ورودی انتخاب ماست،باید سیگنال ۱ به همراه A' و S به گیت اقلیت اعمال شود. گیت OR می‌تواند با استفاده از یک گیت اکثریت نیز پیاده سازی شود.

در نهایت خروجی هر دو گیت اکثریت و اقلیت به گیت اکثریت دیگری اعمال می‌شود که ورودی دیگر آن یک سیگنال یک است و ولتاژ خروجی F را تولید می‌کند.خروجی F که خروجی یک گیت مالتی پلکسر ۲ به ۱ است.

در [5]، از ۲۳ سلول استفاده شده است و تاخیر آن ۲۷۰ است، و به اندازه لازم چگال است. تمام اطلاعات مربوط به شبیه سازی این ملتی‌پلکسر در [5] ذکر شده است. در [5] تعداد سلول و ناحیه و تاخیر انتشار سیگنال از ورودی به خروجی را کاهش داده شده است و ساده و قابل پیاده سازی است، می‌توان با استفاده از این روش، طراحی‌های بهتر مالتی پلکسر را بدست آورد.

۳-۴- رمز گشا در QCA

در مرجع [15] برای اولین بار یک رمز گشا ۲ به ۴ در اتوماتای سلولی کوانتومی معرفی شده است که در یک لایه است و هیچ تقاطع هم سطحی ندارد.

در پیاده سازی با ترانزیستور حد اقل ۴ تقاطع وجود داشت که با استفاده از ۲ رای گیر اکثریت و یک رای گیر اکثریت اقلیت و گیت‌های معکوس کننده منطق اتوماتای کوانتومی سلولی یک پیاده سازی بدون تقاطع را ممکن ساخت. در اتوماتای کوانتومی سلولی وجود حداقل تعداد تقاطع برای پیاده سازی اهمیت زیادی دارد.

در [15] نتایج پیاده سازی با QCA Designer قرار داده شده است.

۴- نحوه تولید ساعت

یک چرخه‌ی پالس ساعت در منطق QCA می‌تواند به چهار فاز *"تغییر"*[۵]، *"حفظ"*[۶]، *"آزادسازی"*[۷] و *"استراحت"*[۸] تقسیم شود. در فاز *"تغییر"*، *موانع بین نقطه‌ای*[۹] به تدریج بالا می‌رود و سلول QCA به یکی از حالات زمینه‌ی همسایگانش می‌رود. در فاز "حفظ"، مانع بین نقطه‌ای بالا می‌ماند و مانع از تونل زنی الکترون‌ها شده و سلول حالت زمینه‌ی خود را حفظ می‌کند. در فازهای "آزادسازی" و "استراحت"، مانع بین نقطه‌ای پایین می‌آید و الکترون‌های اضافی به تحرک می‌رسند. در این دو فاز، یک سلول QCA، بدون قطبش باقی می‌ماند.

در مجموع، قطبش یک سلول QCA هنگامی که در فاز "تغییر" است توسط قطبش همسایگان خود که در فاز "تغییر" و یا "حفظ" هستند، مشخص می‌شود. همسایگان بدون قطبش در فازهای "آزادسازی" و "استراحت" هیچ تأثیری در تعیین حالت سلول QCA ندارند [8,9].

در کل یک طرح QCA دارای چهار پالس ساعت مانند شکل می‌باشد. هر پالس ساعت ۹۰ درجه با پالس ساعت قبلی خود اختلاف فاز دارد.

مدل‌های نحوه تولید ساعت متفاوتی برای مدارات QCA پیشنهاد شده است که برای نوع خاصی از طراحی‌ها مناسب است. برای برخی از مدارات چند نوع ساعت معرفی شده تا به درستی و با تاخیر مناسب کار کند.

برخی از این نوع نحوه تولید ساعت‌ها برای پیاده سازی تقاطع همسطح با یک نوع سلول توصیه شده و ابزارهایی نیز برای این کار معرفی و ساخته شده است [8].

---

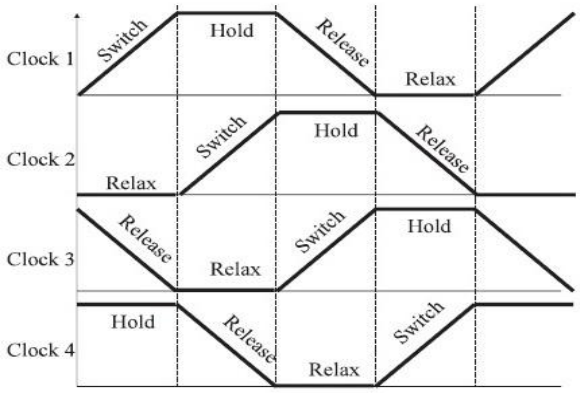
[۵] Switch

[۶] Hold

[۷] Release

[۸] Relax

[۹] Inter-dot Barriers



شکل(۱۳): چهار پالس ساعت QCA با فواصل ۹۰ درجه‌ای[6]

۴-۱- نحوه تولید ساعت Lander و Bennett

یک E-field بر روی لایه‌ای از سیم فلزی تولید می‌شود تا بر رو یا زیر سلول‌های QCA قرار گیرد تا بدین وسیله تونل‌زنی را در سلول‌ها کنترل کند. سلول مستقیما به مدار کلاک متصل نمی‌باشد. این یک خاصیت ذاتی مقیاس مولکولی است. انتقال پیوسته E-field روی لبه جلویی موج خطا و انرژی کینک را کاهش می‌دهد. ارتعاشات E-field به وسیله هر سیم ایجاد می‌شود که اختلاف فاز آن با همسایه‌اش برابر $\pi/2$ می‌باشد. لایه زمین این ارتعاش را به دیگر لایه‌های QCA انتقال می‌دهد. یک E-field که با این مدار تولید تولید می‌شود می‌تواند به شکل سینوسی فرض شود که برای نحوه تولید کلاک لندر مجاز است.

محاسبه و تغییر سلول تنها در لبه پایین‌روند موج صورت می‌گیرد. این روش نحوه تولید کلاک سفر موج و موج محاسباتی و نحوه تولید کلاک لندر نامیده می‌شود.در این نحوه تولید ساعت که به هدف تولید مدار قابل اطمینان از تغییردهی بی دررو[۱۰] استفاده می‌کند، از چهار فاز ساعت مانند شکل ۱۳ استفاده می‌نماید [16].

حداقل سه فاز برای کنترل جریان داده‌ها در لایه QCA لازم است.امکان پیاده‌سازی تولید چهار فاز کلاک QCA با سیم‌هایی که دقیقا بالای لایه QCA قرار داده شده‌اند، امکان دارد. زمینه روی طرف دیگر لایه QCA تا خط های میدان را جداسازی کند. چهار سیم ($\phi1,\phi2,\phi3,\phi4$) چهار فاز شیفت داده شده سیگنال را تولید می‌کند. توزیع اصلی روی لایه QCA از سیم کوچکتری که انشعاب به سمت حامل اصلی دارد بدست می‌آید.

روش نحوه تولید کلاک بنت مقدار میانی و محاسبه شده را به همگام کلاک سلول QCA درجا ذخیره می‌کند. این روش توان مصرفی بسیار پایین دارد و فضای کمی را اشغال می‌کند زیرا هیچ اصلاحی در طرح مدار صورت نمی‌گیرد. اما سربار زمانی در مقایسه با روش نحوه تولید کلاک لندر ایجاد می‌کند.

یک مدار با کلاکینک بنت زمان زیادی برای تولید خروجی دارد. بنابراین یک راه‌حل ترکیبی برای کاهش توان مصرفی و ممانعت از کاهش بازدهی نیاز است. این روش ترکیبی می‌تواند ترکیب خط لوله و معکوس‌پذیری باشد. روش نحوه تولید کلاک بنت را با امکانات خط لوله QCA ترکیب می‌نماید. یک تحلیل مصرف توان کامل باید در لایه کلاک انجام شود. در [17] مدار QCA به دو بخش محاسباتی و حافظه ذخیره‌سازی تقسیم شده است. مدار QCA در مرحله محاسبات از نحوه تولید کلاک بنت استفاده می‌کند و توانی مصرف نمی‌کند. مدار QCA به m مرحله بخش‌بندی می‌شود که هر بخش شامل i_j ورودی و o_j خروجی می‌باشد. مقدار $i_j=o_{j-1}$ زیرا ورودی هر مرحله خروجی مرحله قبل است [17].

۴-۲- نحوه تولید ساعت تسهیم زمانی

در [8] نحوه تولید کلاک تازه ای در تقاطع بهره گرفته می‌شد تا یک نوع سلول در تقاطع همسطح وجود داشته باشد. وجود این یک نوع سلول باعث سادگی در ساخت خواهد شد و اطمینان پذیری مدار را افزایش می‌دهد [8].

۵- عناصر حافظه درQCA

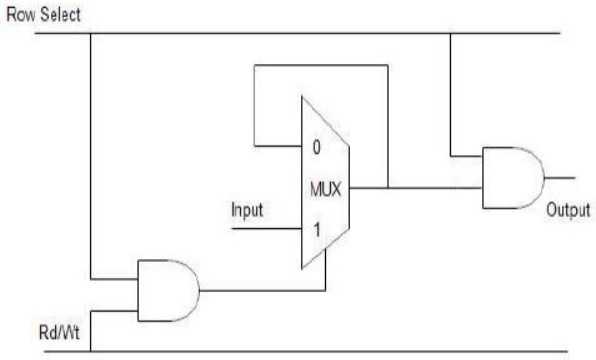
در تحقیقات اتوماتای سلولی کوانتومی دو نوع معماری حافظه و سری معرفی شده است. حافظه موازی، چرخه چند تایی حافظه تک بیتی را دارد، بنابراین تمام بیت‌های یک کلمه همزمان مورد دسترسی قرار می‌گیرد و این امر موجب تاخیر کمی خواهد شد. اما ایراد این نوع حافظه در جریان مداوم و بالایی است که برای هر بیت از کلمه مصرف می‌شود، بنابراین حافظه موازی در جایی که ناحیه مهم‌تر از تاخیر است مناسب نیست.از طرفی حافظه سری تاخیر زیاد و فضای کمی دارد و چون بیت‌ها یکی پس از دیگر در دسترس قرار می‌گیرند نیاز به جریان انرژی داده جدا برای هر بیت نیست.

۵-۱- حافظه موازی[18]

شکل ۱۴ چرخه یک بیتی حافظه موازی را نشان می‌دهد. در این حافظه یک مالتی پلکسر ۲ به ۱ وجود دارد که یکی از خروجی ها به یکی از ورودی‌های آن وصل است و در طول عملیات خواندن سیگنال RD/WR پایین است و مالتی پلکسر درحالت بازخورد قرار دارد و مانند یک چرخه حافظه عمل می‌نماید.

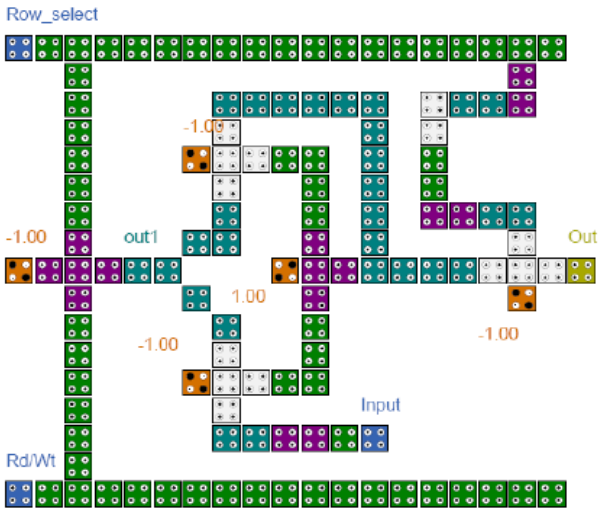
---

[۱۰] Adiabatic



شکل (۱۴): سلول حافظه موازی

وقتی سیگنال RD/WR بالا باشد، ورودی جدید به خروجی مالتی‌پلکسر می‌رود و مقدار جدید در حافظه بارگذاری می‌شود. در شکل ۱۵ شماتیک مداری و طرح QCA این سلول حافظه موازی نشان داده شده است.



شکل(۱۵): طرح QCA سلول حافظه موازی

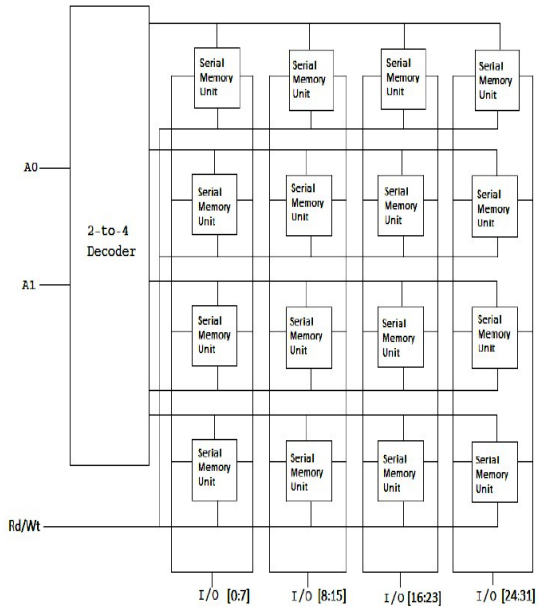
در طراحی قبلی این فرض که تمامی تاخیرهای ورودی یکسان باشند، وجود نداشت، تا به درستی کار کند.

۵-۲- حافظه سری

حافظه سری برتری خوبی نسبت به حافظه موازی دارد، زیرا در آن نیاز کمتری به چند برابر کردن اندازه مدار جریان رسانی است. برای پیاده-سازی یک حافظه سری معماری‌های متفاوتی در مقالات پیشنهاد شده است. اما در بیشتر این طرح ها از نحوه تولید ساعت رایج استفاده نشده و از نحوه تولید ساعت پیچیده استفاده می‌کنند. برای نوشتن سیگنال RD/WR باید به تعداد بیت‌های یک کلمه در پریود ساعت، بالا باشد و در خواندن حافظه سری نیز باید این سیگنال به تعداد بیت‌های کلمه، پایین باشد [19].

۵-۳- حافظه ترکیبی[11]

در بخش قبل با فواید و مشکل هر دو نوع معماری موازی و سری آشنا شدیم. بنابراین با توجه به میزان تاخیر مورد نظر و فضای متناسب ترکیبی از دو حافظه را می‌توان به کار برد. به طور مثال برای یک حافظه با کلمات ۳۲ بیتی می‌توان از واحدهای حافظه ۸ بیتی استفاده کرد تا یک کلمه کامل بدست آورد. در شکل ۱۶ این روش نشان داده شده است [1].



شکل (۱۶): طرح حافظه ترکیبی

حافظه ترکیبی از ترکیب های متفاوت حافظه سری و موازی بدست می‌آید. جدول ۲ ترکیب‌های مختلف ممکن در اندازه حافظه سری برای ساخت یک واحد حافظه ۳۲ بیتی را نشان می‌دهد.

۵-۴- قرارداد برای استفاده از حافظه سری

در واقع حافظه سری یک نوع خافظه متحرک است یعنی بیت در یک چرخه قرار دارد و در دستیابی ممکن است ترتیب بیت‌ها تغییر کند که برای حل این مشکل قراردادی می‌گذاریم و از شمارنده استفاده می-کنیم و چون داده‌ها برای پردازش باید به صورت موازی پردازش شود این داده سری را به موازی تبدیل می‌کنیم.که فضای زیادی برای این تبدیلات به کار گرفته می‌شود [1].

در جدول ۲، X نشان دهنده تعداد سلول‌های به کار گرفته شده در ارتباط بین واحدهای حافظه سری است. از [7] نشان داده شده .که این تعداد حدود ۳۰ سلول است و C عدد ثابتی برای پیاده سازی رمز گشا است برای رسیدن به بهترین ترکیب برای ساخت حافظه با کلمه ۳۲ بیت و یا هر اندازه دیگر، تابع هدف زیر را تعریف می‌کنیم.

F=(Latency)*(Total number of cell required) (۳)

[11] Hybrid memory

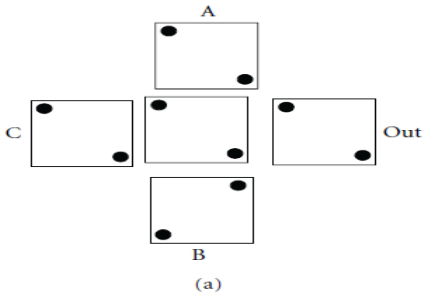
جدول (۲): ترکیب‌های مختلف در اندازه حافظه سری برای ساخت یک واحد حافظه ۳۲ بیت

Serial Memory unit size	Total Latency	Number if cells in one unit	Number of parallel units required	Total Number of cells
32	33	366+x	1	366+x +C
16	17	238+x	2	476+2x +C
8	9	174+x	4	696+4x+C
4	5	142+x	8	1136+8x+C
2	3	126+x	16	2016+16x+C
1	2	126+x	32	4032+32x+C

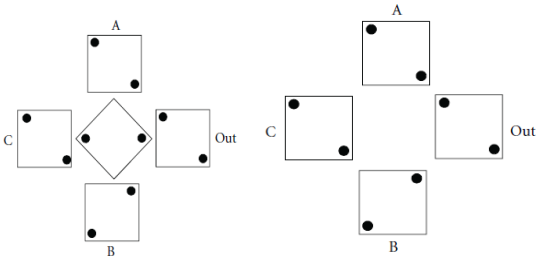
با استفاده از کمینه سازی این تابع هدف بهترین حالت را برای حافظه ترکیبی بدست می‌آوریم. البته با توجه به کاربرد تابع هدف بالا می‌تواند به هر کدام از پارامترهای تاخیر یا فضا وزن دهد [1].

۶- خطا در QCA و اطمینان پذیری در حافظه‌ها

در این تحقیقات سه نوع اشکال در سیستم معرفی شده است. خطای ناهم‌گونی [۱۲] و خطای از بین رفتن سلول [۱۳] و جابجایی سلول [۱۴] به ترتیب در شکل ۱۷ نشان داده شده است [7].



(a)



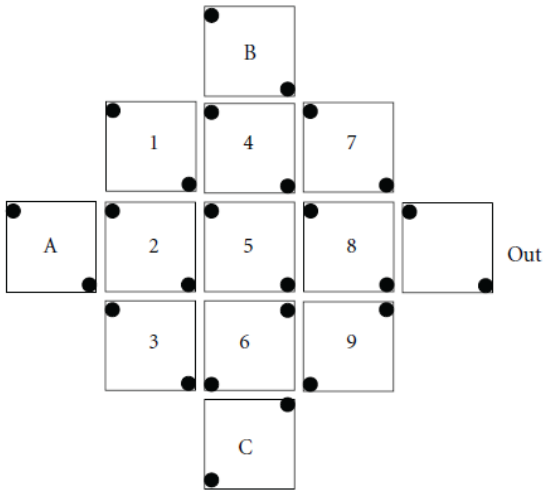
شکل(۱۷): شکل (a) خطای ناهمگونی و (b) خطای از بین رفتن سلول و (c) خطای جابجایی سلول را نشان می‌دهد.

در یک مدار اتوماتای سلولی ممکن است ترکیبی از این خطاها نیز رخ دهد.

مدل‌های مختلفی برای خطا وابسته به کاربرد،می‌تواند تعریف شود.

۶-۱- مدل جدید برای طراحی گیت اکثریت سه ورودی تحمل پذیر اشکال

در مدل پیشنهادی [7] سه ورودی با A, B, C وخروجی با Out نام-گذاری شده است و ۹ سلول میانی با شماره ۱ تا ۹ نامگذاری شده است. پلاریته سلول ورودی را ثابت نگه‌داریم وبقیه سلول ها دارای پلاریته آزاد هستند. با استفاده از ۱۳ سلول این گیت طراحی شده است(شکل۱۸).



شکل (۱۸): طراحی گیت اکثریت سه ورودی تحمل پذیر اشکال [7]

برای اثبات درستی طرح پیشنهادی، هم با استفاده از شبیه ساز و هم اثبات فیزیکی استفاده شده است برای اثبات فیزیکی، گیت به صورت یک سیستم در نظر گرفته می‌شود و آنگاه انرژی کینگ در حالات مختلف برای این سیستم بررسی می‌شود، آنگاه آن حالتی که در آن سیستم کمترین انرژی را دارا خواهد بود به عنوان حالت پایدار انتخاب می‌شود [7,10,1,11].

در [7] تحمل [۱۵] گیت را در حضور سه خطای نا هم گونی [۱۶] و خطای از بین رفتن سلول [۱۷] و جابجایی سلول [۱۸] بررسی شده و از اثبات

[۱۲] Misalignment
[۱۳] Missing cell
[۱۴] Dislocation
[۱۵] Robostness

اثبات فیزیکی و شبیه ساز QCA Designer برای اثبات درستی استفاده شده است.

با توجه به قوانین طراحی که در [10,6] و مدارات مقاوم ساخته شده در [7,12] دیدیم می‌توان مدارات حافظه را نیز بهبود داد.

۷- کارهای آتی

با استفاده از قانون تسهیم زمانی که برای تک نوع سازی سلول‌های QCA در تقاطع همسطح مطرح شد، می‌توان سایر مدارات حافظه که از تقاطع همسطح استفاده می کند را بهبود داد و در تحمل پذیری و ساده سازی و همچنین کم کردن تعداد سلول‌ها و کم کردن پیچیدگی مدارات حافظه از آن استفاده نمود .

مدارات حافظه‌ای که معرفی شد، با به کارگیری عناصر بهبود یافته جدید مالتی پلکسر قابل بهبود است. همچنین از عناصر جدید DET در طراحی آن‌ها می‌توان بهره برد تا گذردهی داده را افزایش داد.

با استفاده از عناصر تحمل‌پذیر قابل اطمینال معرفی شده می‌توان در بهبود اطمینان پذیری مدارات حافظه کارهای قابل قبولی پیشنهاد داد [10,6].

۸- نتیجه

اتوماتای سلولی کوانتومی امید بخش‌ترین جایگزین برای CMOS است که امید است در ساخت مدارات کم توان به کار گرفته شود و با توجه پیشرفت مدارات در QCA، کامپیوترهایی با چگالی بسیار بالا و مصرف توان بسیار پایین و سرعت بالا قابل دستیابی خواهد بود.

با پیشرفت مشاهده شده در حوزه تقاطع همسطح و استفاده از یک نوع سلول در آن و قابلیت ترکیب چند تقاطع فرآیند ساخت ساده-تر خواهد شد و امید برای دستیابی به مدارات دیگر افزایش می‌یابد.

با بهبود هر چه بیشتر مدارات برای ساخت حافظه‌ها و سایر مدارات ترتیبی، امید دستیابی به ساخت Cache و حافظه اصلی با سرعت و ظرفیتی صدها برابر وجود خواهد داشت.

۹- پیشنهاد پروژه

در QCA به علت استفاده از ذرات کوانتومی به جای ترانزیستورها، سلسله مراتب حافظه که در ترانزیستور وجود داشت از بین می‌رود.

در مدارات ترانزیستوری به طور کلی RAM دو نوع است. مخفف کلمات STATIC RAM به معنای حافظه ایستا و DRAM مخفف کلمات Dynamic RAM به معنای حافظه پویا. همان طور که می دانید اطلاعات در حافظه به صورت صفر و یک ذخیره می‌شود.

ساختار DRAM از خازن‌هایی تشکیل شده که شارژ بودن آن‌ها به منزله یک و خالی بودن آن‌ها به منزله صفر بودن سلول حافظه است. اما خاصیت خازن این است که با گذشت زمان تخلیه می‌شود .بنابراین باید به طور منظم و در بازه‌های زمانی منظم دوباره شارژ شود تا اطلاعات از بین نرود. زمانی که صرف شارژ مجدد حافظه می‌شود، باعث کند شدن نسبی حافظه می‌گردد. بنابراین DRAM کند است.

حافظه SRAM از یک سری مدارات منطقی تشکیل شده است که با گذشت زمان، مقدار و تعداد خود را از دست نمی‌دهد و تا جریان برق برقرار باشد صفر و یک را در خود نگاه می‌دارند.به همین دلیل نیاز به شارژ دوباره ندارد و سریع‌تر از DRAM است.

اما همین ساختاری که باعث افزایش سرعت SRAM می‌شود، باعث می‌شود از نظر اندازه نیز فضای بیشتری را نسبت به DRAM اشغال کند و در ضمن از قیمت بالاتری نیز برخوردار باشد.

به همین دلیل RAM معمولی رایانه با وجود سرعت بالاتر SRAM، از نوع DRAM است اما حافظه نهان پردازنده (cache) که حجم زیادی ندارد و در عوض به سرعت بالایی نیاز دارد، از نوع SRAM است. بنابر آنچه ذکر شد سلسله مراتب حافظه در مدارات CMOS وجود دارد، اما در QCA با دستیابی به حافظه سریع و با توان مصرفی پایین دیگر نیازی به سلسله مراتب خافظه نخواهد بود. بنابراین با توجه به اهمیت حافظه‌ها در معماری سیستم‌های کامپیوتری و اهمیت زمان دستیابی به حافظه، ساخت مدارات حافظه مبتنی بر تکنولوژی QCA که هدف آنها کاهش زمان دسترسی حافظه و کاهش تراکم مدار است، به عنوان یک چالش مطرح می‌باشد.

اگر بتوان حافظه‌های QCA را با سرعت بالا و تاخیر دسترسی پایین طراحی نمود، امکان ترکیب حافظه نهان و اصلی وجود خواهد داشت و سلسله مراتب حافظه با ظرفیت و سرعت دستیابی بالا و همچنین توان مصرفی ناچیز امکان‌پذیر خواهد بود و بدین وسیله طراحی پردازنده و سیستم‌های QCA با کارآیی و هزینه مناسب‌تر امکان‌پذیر می‌شود. با بکارگیری قوانین طراحی و ساده‌سازی مدارات QCA و کم کردن تعداد لایه‌های بکارگرفته شده درطراحی این‌گونه مدارات می‌توان بهبودی برای زمان دسترسی به حافظه و افزایش تراکم مدرات حافظه‌ای، ارائه داد.

یکی از راهکارها برای بهبود حافظه‌ها استفاده از تقاطع همسطح با یک نوع سلول است که برای ساخت حافظه اتوماتای سلولی کوانتومی برای دستیابی به تراکم بالا و زمان دسترسی کمتر، است[6]. راهکار دیگر استفاده از فلیپ فلاپ‌های دولبه QCA است که با بهره‌گیری از ویژگی گذردهی داده بالا می‌تواند برای ساخت عناصر حافظه به کار گرفته شود [3].

---

[۱۶] Misalignment

[۱۷] Missing cell

[۱۸] Dislocation

برای اجرا این تحقیقات از ابزار شبیه‌سازی QCA Designer و موتور جدیدی در MATLAB که توانایی شبیه‌سازی تقاطع‌های بیشتری در سطح مدارات را دارست، بهره گرفته خواهد شد [1].

در سال‌های اخیر تحقیقات زیادی برای طراحی مدارات QCA انجام شده است که منجر افزایش کارآیی مدارهای مبتنی بر QCA از لحاظ تاخیر، سرعت، توان، تعداد سلول‌های به کاربرده شده، حجم مدارت، آسان شدن فرآیند تولید، شده است و زمینه را برای ورود مدارات به حوضه نانومتری فراهم آورده است. اتوماتای سلولی کوانتومی در حوزه توسعه مدارات جمع کننده به طراحی‌های مختلفی رسیده که بهینه بودن این مدارات از نظر تاخیر، ناحیه مورد نیاز برای پیاده‌سازی وتعداد سلول‌های به کار برده شده در سطح مدار و تعداد لایه‌ها، مقایسه آنها و تکنیک‌های به کار گرفته شده در آن. همچنین در حوزه حافظه و مدارات ترتیبی بوجود آمدن مدارات DET(Dual Edge Triggered) امکان گذردهی داده بیشتر و بهبود سرعت و افزایش کارآیی را ایجاد کرده است[1,2,3,7].
درحوزه تحمل پذیری اشکال قوانین بوجود آمده باعث ایجاد مدارات قوی‌تر شده است. با بکارگیری قوانین موجود در طراحی مقاوم در QCA و ساده‌سازی این مدارات و کم کردن تعداد لایه‌های بکارگرفته شده در طراحی اتوماتای سلولی کوانتومی امکان دستیابی به اهدافی در ساخت مدارات حافظه از قبیل:توان پایین، سرعت بالا، چگالی زیاد آسان‌تر خواد شد. یکی از اهدافی که در بهبود طراحی مدارات حافظه در QCA مورد توجه است، ساخت مدارات به شکلی است که ورودی‌ها در یک سمت و خروجی‌ها در سمت دیگر باشد تا بکارگیری آن در مدارات بزرگتر و پیچیده حافظه آسان باشد. با دنبال کردن این اهداف بهبود مدارات حافظه در QCA روندی سریع خواهد داشت.
در سال ۲۰۱۲ مدار ترتیبی دو لبه فلیپ فلاپ JK و D طراحی شده است که نسبت به مدارات فلیپ فلاپ قبلی بهبود چشم‌گیری داشته است [3]. حافظه ترکیبی در سال ۲۰۱۲ معرفی شده است که از موازنه بین حافظه موازی و سریال بهره گرفته و بین تاخیر و حجم حافظه موازنه برقرار می‌کند [1]. با استفاده از تسهیم زمانی در سال ۲۰۱۲راهی برای پیاده سازی در تقاطع همسطح پیشنهاد شده که منجر به استفاده از یک نوع سلول در تقاطع شده و مقاومت مدار را اقزایش داده است و با ترکیب چند تقاطع حجم مدار و تعداد سلول‌ها کاهش می‌یابد[6]. در مرجع [15] برای اولین بار یک دیکدر ۲ به ۴ در اتوماتای سلولی کوانتومی معرفی شده است که در یک لایه است و هیچ تقاطع هم سطحی ندارد.

در پیاده سازی با ترانزیستور حد اقل ۴ تقاطع وجود داشت که با استفاده از ۲ رای گیر اکثریت و یک رای گیر اکثریت اقلیت و گیت‌های معکوس کننده منطق اتوماتای کوانتومی سلولی یک پیاده سازی بدون تقاطع را ممکن ساخت. در اتوماتای کوانتومی سلولی وجود حداقل تعداد تقاطع برای پیاده سازی اهمیت زیادی دارد. در [15] نتایج پیاده سازی با QCA Designer قرار داده شده است.

سپاسگزاری

با تشکر از جناب آقای مهندس مقدم ارجمند بخاطر حمایت علمی در این سمینار .

مراجع

[1] D. Agrawal and B. Ghosh, "Quantum Dot Cellular Automata Memories," *Computer Applications*, vol. 46, no. 5, pp. 27-30, May 2012.

[2] M. R. Azghadi and K. N. Omid Kavehei, "A novel design for Quantum-dot cellular automata and fulladder," *Applied Science*, vol. 22, no. 7, pp. 3460-3468, 2007.

[3] V. A, W. K, J. GA, and D. V, "Quantum Dot Cellular Automata Carry-Look-Ahead Adder and Barrel Shifter," in *IEEE Conf. Emerging Telecommunications Technologies*, Dallas, 2002, pp. 23-24.

[4] S. K. lakshmil, G.Athishi, M.Karthikeyan, and C.Ganesh, "Design of Subtractor using Nanotechnology Based QCA," in *Communication Control and Computing Technology (ICCCCT)*, 2010, pp. 384-388.

[5] D. Mukhopadhyay and P. Dutta, "Quantum Cellular Automata based Novel Unit 2:1 Multiplexer," *Computer Application*, vol. 43, no. 2, pp. 22-25, Apr. 2012.

[6] K. K, W. K, and K. R, "Towards Designing Robust QCA Architectures in the Presence of Sneak Noise Paths," in *Design, Automation and Test*, Europ, 2005, pp. 1214-1219.

[7] R. Farazkish, S. Sayedsalehi, and a. K. Navi, "Novel Design for Quantum Dots Cellular Automata to Obtain Fault-TolerantMajority Gate," *Nanotechnology*, Jan. 2012.

[8] K. P. Rajeswari D and M. Balakrishnan, "Clocking-based CoplanarWire Crossing

Scheme for QCA," in *International Conference on VLSI Design*, Bangalore, India, 2010, pp. 339-344.

[9] L.-r. XIAO, X.-x. CHEN1, and S.-y. YING, "Design of dual-edge triggered flip-flops based on quantum-dot cellular automata," *Zhejiang University-SCIENCE C*, vol. 13, no. 5, pp. 385-392, 2012.

[10] K. K, W. K, and K. R, "The Robust QCA Adder Designs Using Composable QCA Building Blocks," *IEEE Transaction CAD Integrated Circuits System*, vol. 26, pp. 176-183, 2007.

[11] K. Navi, A. M. Chabi, and S. Sayedsalehi, "A Novel Seven Input Majority Gate in Quantum-dot Cellular Automata," *Computer Science*, vol. 9, no. 1, pp. 84-89, Jan. 2012.

[12] S. Hashemi, M. Tehrani, and K. Navi, "An efficient quantum-dot cellular automata full-adder," *Scientific Research and Essays*, vol. 7, no. 2, pp. 177-189, Jan. 2012.

[13] W. Wang, K. Walus, and G. Jullien, "Quantum-dot cellular automata adders," in *IEEE Conf. Nanotechnology (NANO)*, San Francisco, 2003, pp. 461-464.

[14] H. Ismo and T. Jarmo, "Binary multipliers on quantum-dot cellular automata," vol. 58, no. 1, pp. 87-103, Jan. 2010.

[15] R. ZHOU, X. XIA, F. WANG, Y. SHI, and H. LIAO, "A Logic Circuit Design of 2-4 Decoder Using Quantum Cellular Automata," *Computational Information Systems*, vol. 8, no. 8, pp. 3463-3469, 2012.

[16] C. Lent ; B. Isaksen, "Clocked molecular quantum-dot," *Electron Devices, IEEE Transactions*, vol. 50, no. 9, p. 1890–1896, Sep. 2003.

[17] Marco Ottavi; Salvatore Pontarelli; Erik DeBenedictis, "High Throughput and Low Power Dissipation in QCA Pipelines using Bennett Clocking," *IEEE ACM International Symposium on Nanoscale Architectures*, pp. 17-22, 2010.

[18] M. O. Vamsi Vankamamidi, "A Line-Based Parallel Memory for QCA Implementation," *IEEE TRANSACTIONS ON NANOTECHNOLOGY*, vol. 4 , no. 6, 2005.

[19] A. S. M. b. Q.-D. C. A. (QCA), *IEEE TRANSACTIONS ON COMPUTERS*, vol. 57, no. 5, pp. 606-618, May 2008.

بررسی و مقایسه‌ی معماری‌های کامپیوتر، بر مبنای فناوری‌های سلول‌های کوانتومی[1] و ترانزیستورهای نانولوله‌ی کربن[2]

محمد طاهری فرد* ، محمود فتحی **
* دانشجوی کارشناسی ارشد، دانشکده‌ی مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
Taherifard@Comp.Iust.ac.ir
** دانشیار، دانشکده‌ی مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
MahFathy@Iust.ac.ir

چکیده

کوچک کردن اندازه‌ی ترانزیستورها در ابعاد نانو، صنعت ترانزیستورهای اثر میدانی نیمه هادی[3] را با مشکل‌ها و چالش‌های جدیدی روبرو کرده است. از اینرو سلول‌های کوانتومی و ترانزیستورهای نانولوله‌ی کربن، به عنوان برخی از تکنولوژهای جایگزین جهت رفع این مشکل‌ها، در نظر گرفته شده‌اند. در این سمینار به بررسی، معرفی و مقایسه‌ی مشخصه‌ها و ویژگی‌های این دو فناوری از جمله، توان مصرفی، فضای اشغالی، تأثیر دما و مقایسه‌ی آنها با فناوری ترانزیستورهای اثر میدانی، پرداخته شده است.

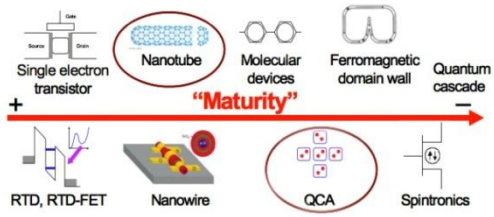
بررسی‌های انجام شده حاکی از آن است که هر یک از این دو فناوری، در معیارهای مختلف، دارای رفتارها و برتری‌های منحصر بفرد هستند و بطور قطع نمی‌توان یکی را برتر از دیگری دانست. به عنوان مثال، با هدف کاهش فراوان توان مصرفی، کاهش تأخیر، کاهش فضای اشغالی و افزایش چشمگیر فرکانس کاری، سلول‌های کوانتومی دارای برتری هستند. همچنین بلوغ پایین‌تر سلول‌های کوانتومی از یک طرف، و کاهش مناسب توان مصرفی ترانزیستورهای نانولوله‌ی کربن، حساسیت پایین آنها در مقابل دما، قابلیت اطمینان بیشتر نسبت به سلول‌های کوانتومی و امکان تولید ترانزیستورهای نانولوله‌ی کربن با استفاده از روش‌های فعلی تولید ترانزیستورهای اثر میدانی، از طرف دیگر، برای آینده‌ی نزدیک، فناوری ترانزیستورهای نانولوله‌ی کربن را به برتری در مقابل سلول-های کوانتومی می‌رساند.

کلمه‌های کلیدی

سلول‌های کوانتومی، ترانزیستورهای نانولوله‌ی کربن، توان مصرفی، تأخیر، فضای اشغالی

۱- مقدمه

گوردن مور[4] یکی از دانشمندان برجسته‌ی شرکت اینتل در سال ۱۹۶۵، دو برابر شدن تعداد ترانزیستورها در سطح تراشه را در هر ۱۸ ماه، پیش‌بینی نمود. با توجه به توسعه رو به افزایش مدارها در ابعاد نانومتر، در برخی موارد افزایش تعداد ترانزیستورها از مرز دو برابر نیز گذشته است. تولید تراشه‌ها در ابعاد نانو، صنعت نیمه‌هادی اثر میدانی را به چالش‌های جدیدی کشانده است. به عنوان مثال کوچک کردن ابعاد ترانزیستورها در اندازه‌های ۳۲ نانومتر و کوچکتر، منجر به افزایش تأثیر اتصال کوتاه کانال[۱] ، کاهش کنترل گیت[۲] ، افزایش نمایی جریان نشتی[۳,۴] ، تغییر بزرگ پارامترهای ساخت و غیره می‌گردد [۵]. فناوری‌های مختلف از جمله ترانزیستورهای تک الکترونی[5] [۶] ، کلیدهای مولکولی[6] [۷]، اسپین ترونیک[7] [۸]، سلول‌های کوانتومی و ترانزیستورهای نانولوله‌ی کربن از جمله مواردی است که به عنوان تکنولوژهای جایگزین در نظر گرفته شده است. این فناوری‌ها به ترتیب بلوغ در شکل(۱) نمایش داده شده‌اند[۹] .هر یک از فناوری‌های فوق، دارای مزیت‌ها و معایبی است که پرداختن به جزئیات آن از این مقوله خارج است. به عنوان مثال فناوری کلیدهای مولکولی، دارای فرکانسی تا مقیاس تراهرتز هستند, ولی مراحل ساخت آنها با چالش‌های بزرگی همراه است. فناوری اسپین ترونیکس، بر اساس خواص مغناطیسی، منطق صفر و یک را پیاده‌سازی می‌نماید و دارای سرعت و فشردگی بالایی است.



شکل(۱): فناوری‌های جدید در ابعاد نانو [۹]

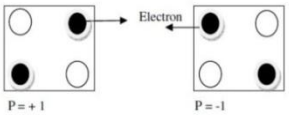
سلول‌های کوانتومی و ترانزیستورهای نانولوله‌ی کربن از جمله تکنولوژهای کاندید برای جایگزینی هستند، که در این سمینار به بررسی، معرفی و مقایسه‌ی مشخصات و ویژگی‌های آنها از جمله توان مصرفی، فضای اشغالی، تأثیر دما و مقایسه‌ی آنها با فناوری ترانزیستورهای اثر میدانی پرداخته خواهد شد. اگر چه تکنولوژی ترانزیستورهای نانولوله‌ی کربن، با بهبود مشخصات و معیارهای ترانزیستورهای اثر میدانی و بکارگیری از روش‌های تولید آنها، انتخاب

بهتری برای جایگزینی هستند، ولی سلول‌های کوانتومی نیز دارای برتری‌هایی نظیر توان مصرفی بسیار پایین, سرعت و فرکانس کاری بسیار بالا و فضای اشغالی کم هستند، که چشم‌پوشی از این فناوری نوپا را غیرممکن می‌سازد.

در ادامه‌ی گزارش و در بخش‌های ۲ و ۳ ، به معرفی فناوری سلول‌های کوانتومی و ترانزیستورهای نانولوله‌ی کربن، و بررسی خواص و مشخصات آنها خواهیم پرداخت. در بخش۴، مقایسه‌ی معیارهای مختلف سلول‌های کوانتومی و نانولوله‌ی کربن انجام خواهد شد. در بخش۵، طرح پیشنهاد پروژه و نهایتاً در بخش۶، جمع‌بندی و نتیجه-گیری نهایی ارایه خواهد شد.

۲- سلول‌های کوانتومی

سلول‌های کوانتومی یکی از عناصر تونلی و فناوری بدون ترانزیستور هستند که براساس تراکنش و تعامل میان نقطه‌های کوانتومی عمل می‌نمایند. این سلول‌ها به صورت بلوک‌های مربعی شکلی هستند که در آنها دو الکترون وجود دارد. در این ساختار برخلاف ساختارهای متداول، حالت‌ها و یا مقادیر منطقی با استفاده از سطوح ولتاژ نشان داده نمی‌شوند، بلکه براساس مکان و نوع قرار گرفتن الکترون‌ها نسبت به یکدیگر تعیین می‌شوند[۱۰,۱۱] . این مفهوم اولین بار توسط لنت و همکارانش در سال ۱۹۹۳ ارایه شد. استفاده از فناوری اتوماتای سلول کوانتومی، مساحت اشغالی و توان مصرفی تراشه و یا مدارهای طراحی شده را، به طور قابل ملاحظه‌ای کاهش، و فرکانس کاری آنها را به صورت مشهود و تا مقیاس تراهرتز[۸] افزایش می‌دهد[۱۲] . شکل(۲) نمای دو سلول کانتومی را نشان می-دهد که طرف چپ، نماد یک و طرف راست، نماد صفر منطقی است.



شکل(۲): نمایش دو سلول کوانتومی[۱۳]

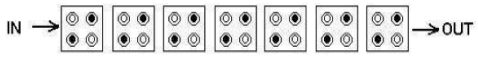
همانگونه که در شکل(۲) مشاهده می‌شود، هر سلول کوانتومی از یک فضای مربعی و دو الکترون آزاد درون آن تشکیل می‌شود. در هر گوشه چهارنقطه مشاهده می‌گردد که به آنها نقاط کوانتومی گفته می-شود. دو الکترون موجود درسلول به دلیل دافعه‌ی الکترواستاتیک میان آنها، همیشه در قطر مربع قرار می‌گیرند و هیچ‌گاه در کنار یکدیگر و در یک ضلع مستقر نمی‌گردند. نحوه‌ی قرار گرفتن الکترونها، پلاریته آنها را مشخص می‌سازد. این پلاریته دارای دو مقدار ۱+ و ۱- بوده، که در منطق دودویی به ترتیب به یک و صفر منطقی تعبیر می‌شوند [۱۴]. با این توضیحات، بدیهی است که چنانچه یک سلول که دارای پلاریته ۱+ است، به یک سلول دیگر نزدیک شود، به دلیل نیروی دافعه‌ی میان الکترون‌های دو سلول مجاور، پلاریته‌ی سلول دوم نیز به ۱+ تغییر می‌یابد.

۲-۱- اجزای پایه‌ی فناوری سلول‌های کوانتومی

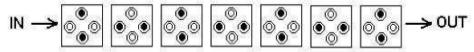
برای درک نحوه‌ی عملکرد فناوری سلول‌های کوانتومی، لازم است اجزای اولیه، نظیر سیم‌ها، گیت معکوس‌کننده، گیت اکثریت و سایر گیت‌های منطقی را معرفی نماییم.

۲-۱-۱- سیم‌های ارتباطی

همانند هر مدار الکتریکی، در فناوری سلول‌های کوانتومی سیم‌ها نقش ارتباط و انتقال داده‌ها را میان اجزا، ایفا می‌نمایند. در این فناوری، سیم‌ها به نوعی از همان سلول‌های کوانتومی تشکیل می-گردند. به اینصورت که با در کنار هم قرار گرفتن سلول‌ها و با تغییر پلاریته یک سلول، به دلیل دافعه‌ی کولومبی، این تغییر پلاریته درسایر سلول‌ها اعمال می‌گردد. در شکل(۳)، انواع سیم‌های مورد استفاده در این فناوری نمایش داده شده است.

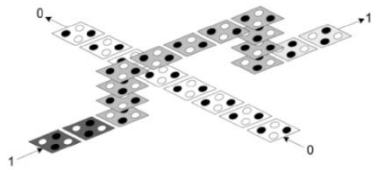


شکل(۳- الف): سیم باینری ساده[۱۵]

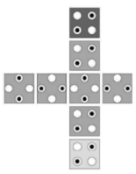


شکل(۳- ب): سیم باینری باینری با چرخش ۴۵ درجه [۱۵]

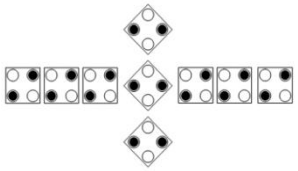
سیم‌های باینری در این فناوری دارای خواص متفاوتی نسبت به سیم‌های معمولی هستند که مهمترین آنها تلاقی و عبور سیم‌های باینری از روی یکدیگر است. شکل(۴) این ساختار را نمایش می‌دهد.



شکل(۴- الف): عبور چند لایه‌ی دو سیم [۱۶]



شکل(۴- ب): عبور دوسیم هم لایه [۱۶]



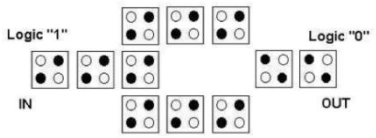
شکل(۴- ج): عبور دو سیم هم لایه [۱۷]

شکل(۴- الف) عبور چند لایه‌ی سیم‌ها را نمایش می‌دهد. بدین-صورت که هر دو سیم از سلول‌های هم نوع تشکیل شده و در هنگام عبور از یکدیگر, در دو صفحه و با دو سطح متفاوت قرار می‌گیرند. همچنین شکل (۴ – ب و ج)، عبور دو سیم در یک لایه را نمایش می‌دهد, با این تفاوت که یک سیم از سلول‌های عادی کوانتومی و سیم

دیگر از سلول چرخیده با زاویه ۴۵درجه تشکیل شده است و علی‌رغم تلاقی مستقیم آنها با همدیگر، اثری بر روی داده‌ی یکدیگر ندارند [۱۶].

۲-۱-۲- گیت معکوس‌کننده

در فناوری سلول‌های کوانتومی، گیت معکوس‌کننده به روش‌های مختلف قابل پیاده‌سازی است, که عمده‌ی تفاوت آنها, مساحت اشغالی و قابلیت اطمینان است. در [۱۵,۱۸] انواع معکوس‌کننده‌ها مورد بحث قرار گرفته است.

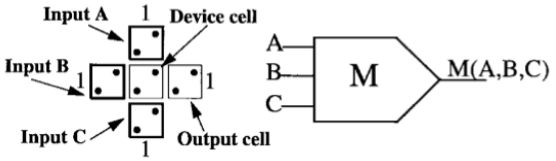


شکل(۵) معکوس‌کننده با قابلیت اطمینان بیشتر [۱۵]

لازم به ذکر است, علاوه بر پیاده‌سازی گیت معکوس‌کننده در شکل(۵)، در صورت استفاده از سیم باینری با زاویه‌ی چرخش ۴۵درجه (شکل۳ – ب)، به صورت یک در میان, هر سلول کوانتومی به صورت خودکار معکوس داده را تولید می‌نماید.

۲-۱-۳- گیت اکثریت[۹]

یکی از مهم‌ترین و اساسی‌ترین گیت‌ها در فناوری سلول‌های کوانتومی، گیت اکثریت است, که بسیاری از گیت‌های منطقی توسط این گیت ساخته می‌شود. این گیت دارای سه ورودی و یک خروجی است. دلیل این نام‌گذاری این است که خروجی، با رای‌گیری از ورودی‌ها, مقدار خود را تعیین می‌نماید. برای درک بهتر به شکل(۶) و جدول (۱) توجه فرمایید.

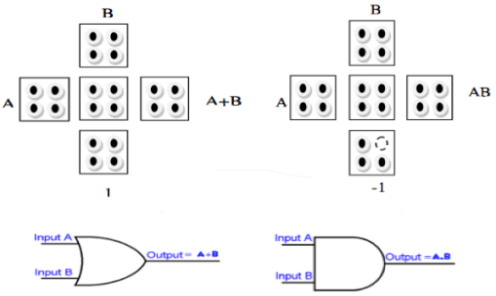


شکل(۶): گیت اکثریت

جدول(۱): جدول ارزش گیت اکثریت

A	B	C	Out
۰	۰	۰	۰
۰	۰	۱	۰
۰	۱	۰	۰
۰	۱	۱	۱
۱	۰	۰	۰
۱	۰	۱	۱
۱	۱	۰	۱
۱	۱	۱	۱

همانگونه که قبلاً هم اشاره شد، بسیاری از گیت‌های منطقی در فناوری سلول‌های کوانتومی با استفاده از گیت اکثریت پیاده‌سازی می‌شوند. به عنوان مثال گیت AND، با صفر نمودن (پلاریته ۱-) و گیت OR، با یک نمودن (پلاریته ۱+) یکی از پایه‌های گیت اکثریت بدست می‌آید. شکل (۷) این پیاده‌سازی را نمایش می‌دهد [۱۳].

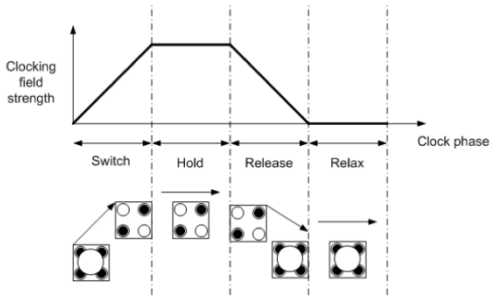


شکل(۷): پیاده‌سازی گیت AND و OR [۱۳]

لازم به ذکر است گیت‌های مختلف، به روش‌های مختلفی طراحی و پیشنهاد شده است و هر کدام سعی در بهبود کارایی، سرعت و مساحت اشغالی داشته است.

۲-۱-۴- مفهوم کلاک

کلاک در فناوری سلول‌های کوانتومی با کلاک در مدارهای کنونی تفاوت‌هایی دارد، چرا که در فناوری سلول‌های کوانتومی، کلاک نه فقط نقش هماهنگ‌کننده، بلکه به عنوان یک سیگنال خارجی تنها تامین کننده‌ی توان و عهده‌دار جهت حرکت اطلاعات در اینگونه مدارها است [۱۸]. عمده تفاوت کلاک در فناوری کوانتومی، چهار فازه بودن آن است، به این مفهوم که کلاک به چهار وضعیت سوئیچ[۱۰]، نگهداری[۱۱]، آزادسازی[۱۲] و آرامش[۱۳] تقسیم می‌گردد. این چهار بخش در شکل(۸) نمایش داده شده است.



شکل(۸): نواحی چهارگانه کلاک [۱۷]

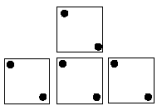
با بررسی شکل(۸) مشاهده می‌شود، در ابتدا یک سلول بدون پلاریته است. سپس در فاز اول کلاک از این حالت بدون پلاریته خارج و یک پلاریته مشخص را بخود می‌گیرد، (مثلا پلاریته ۱+). در فاز دوم که همان فاز نگهداری است، سلول یاد شده، پلاریته خود را نگهداری کرده و ثابت می‌ماند. در فاز سوم سلول کوانتومی آماده‌ی از دست دادن پلاریته‌ی خود شده و به سلول بدون پلاریته تبدیل می‌گردد و نهایتاً در فاز چهارم، در همان وضعیت بدون پلاریته باقی می‌ماند [۱۷,۱۸].

۲-۲- نقص‌های موجود درفناوری سلول‌های کوانتومی

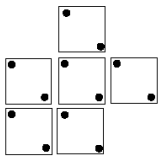
وجود نقص در محصول، مساله‌ای انکار ناپذیر است که در فناوری نانو، بدلیل کوچک شدن ابعاد مساله از اهمیت بیشتری برخوردار است. در فناوری سلول‌های کوانتومی، همانند هر فناوری دیگر، نقص‌های مختلفی وجود دارد و در [۱۹]، احتمال ایجاد نقص در سلول‌های کوانتومی در مقایسه با ترانزیستورهای اثر میدانی، حدود ده تا صد برابر گزارش شده است. مهم‌ترین و رایج‌ترین نقصها به صورت زیر خلاصه می‌شود.

۱. نقص کمبود سلول[۱۴]
۲. نقص اضافه بودن سلول[۱۵]
۳. نقص جابجایی سلول[۱۶]
۴. نقص سلول چرخیده[۱۷]
۵. نقص سلول ثابت[۱۸]

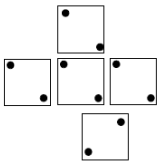
در نقص کمبود سلول و نقص اضافه بودن سلول که هردوی آنها ناشی ازعدم برداشت مناسب ماده رزیست درلیتوگرافی است، در یک طرح، کمبود و یا افزایش سلول کوانتومی، عملکرد مورد نظر را دچار خطا می‌نماید. همچنین در نقص جابجایی سلول، همانگونه که از نامش پیداست، یک سلول درست در جای خودش قرار نمی‌گیرد و بدین‌ترتیب مدار کارایی خود را از دست می‌دهد. در نقص سلول چرخیده نیز سلول دارای یک چرخش ناخواسته است و یا در نقص سلول ثابت، سلول توانایی تغییر پلاریته را ندارد. در شکل(۹) نمونه‌ای از این نقص‌ها را در گیت اکثریت نمایش می‌دهد.



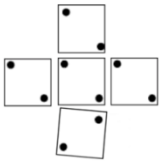
شکل(۹– الف): نقص کمبود سلول



شکل(۹– ب): نقص اضافه بودن سلول



شکل(۹– ج): نقص جابجایی سلول



شکل(۹– د): نقص سلول چرخیده

۲-۳- بررسی سایر مشخصات سلول‌های کوانتومی

در این بخش به بررسی برخی مشخصات، نظیر توان مصرفی، فضای اشغالی، تاخیر و تأثیر دما خواهیم پرداخت.

۲-۳-۱- توان مصرفی

توان مصرفی در سلول‌های کوانتومی در [۲۰] به ازای هر ورودی، حدود $۱۰^{-۱۰}$ وات، محاسبه شده است. این مقدار محاسبه شده، صرفاً جهت تغییر پلاریته‌ی اولین سلول در ورودی در نظر گرفته شده است. بدیهی است که این توان مصرفی در مقایسه با ترانزیستورهای اثر میدانی، قابل چشم‌پوشی است. همچنین برای سلول‌های کوانتومی مفهومی به عنوان توان مصرفی نشتی وجود نخواهد داشت. توان مصرفی سویچینگ نیز در مرز بین نواحی کلاک مطرح می‌گردد[۲۱] و البته با در نظر گرفتن توان مصرفی $۱۰^{-۱۰}$ وات جهت تغییر پلاریته‌ی نواحی کلاک، باز هم توان مصرفی کل، قابل چشم‌پوشی است. به عنوان مثال اگر یک مدار با دو ورودی، دارای تاخیر دو کلاک (معادل هشت ناحیه کلاک) باشد، توان مصرفی آن حدود ۱۰×۱۰^{-۱۰} خواهد بود. در [۲۱] برای تمام جمع‌کننده و گیت اکثریت، متوسط توان مصرفی مورد بررسی واقع شده و در جدول(۲) قابل مشاهده است.

جدول(۲): توان مصرفی در برخی مدارهای سلول‌های کوانتومی

نوع مدار	متوسط انرژی مصرفی(میلی الکترون ولت)	متوسط توان مصرفی (وات)
گیت اکثریت	۳۹,۳۳	62.92×10^{-22}
تمام جمع‌کننده	۵۰۷,۶۶	812.25×10^{-22}

۲-۳-۲- فضای اشغالی

در فناوری سلول‌های کوانتومی، فضای اشغالی تحت تأثیر اندازه هر سلول و فضاهای خالی میان اجزا و سیم‌ها خواهد بود. از آنجا که این سلول‌ها تحت تأثیر بسیار شدید سلول‌های مجاور قرار خواهند گرفت، مطابق روش‌ها و الگوهای طراحی شده در [۲۲] می‌باید فواصل مناسب میان سیم‌ها و سایر اجزا لحاظ شود. اندازه‌های در نظر گرفته شده برای هر سلول کوانتومی ۱۸ * ۱۸ نانومتر، قطر هر نقطه کوانتومی ۵ نانومتر و فاصله دو سلول مجاور ۲ نانومتر در نظر گرفته شده است. بدیهی است که تغییر هر کدام از پارامترهای فوق، در عملکرد مدار تأثیرگذار است. با در نظر گرفتن همه‌ی این موارد، بازهم در [۱۶] چگالی اجزای سلول‌های کوانتومی، حدود $۱۰^{۱۲}$ سلول در سانتیمترمربع گزارش شده است و این مقدار حدود ۱۰۰ برابر فشردگی بیشتر در مقایسه با تکنولوژی ترانزیستورهای اثر میدانی، است[۲۳] .

۲-۳-۳- اثر دما بر سلول‌های کوانتومی

سلول‌های کوانتومی همانند هر فناوری دیگر تحت تأثیر دمای محیط، دارای عملکردی متفاوت هستند. در[۲۴] قابلیت اطمینان

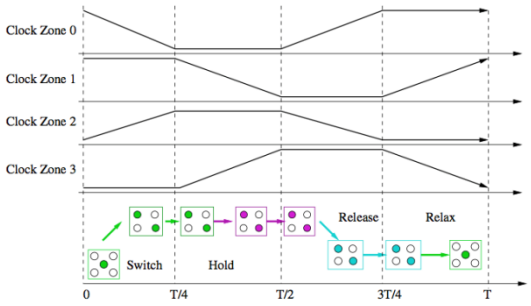
سلول‌های کوانتومی مورد بحث قرار گرفته است. نتایج بدست آمده در جدول(۳) نمایش داده شده است. با بررسی این جدول اینگونه استنباط می‌گردد که در اجزایی نظیر سیم باینری، معکوس‌کننده و گیت اکثریت، با افزایش دما احتمال خطا کاهش می‌یابد، بجز در سیم-های متقاطع که برعکس، با افزیش دما، احتمال وقوع خطا افزایش می‌یابد.

جدول(۳): تأثیردما و احتمال نقص در عناصر مختلف سلول‌های کوانتومی [۲۴]

Temperature	0 K	1 K	2 K	3 K	4 K	5 K
Binary wire	0.265	0.260	0.250	0.210	0.145	0.075
Inverter chain	0.200	0.205	0.215	0.205	0.170	0.120
Inverter	0.200	0.185	0.155	0.110	0.065	0.000
AND 110	0.280	0.280	0.270	0.255	0.235	0.130
OR 101	0.305	0.305	0.295	0.280	0.260	0.145
Crossover	0.015	0.015	0.015	0.010	0.010	0.010
XOR 00	0.232	0.216	0.192	0.144	⋯	⋯
XOR 01	0.216	0.216	0.192	0.136	⋯	⋯
XOR 10	0.018	0.02	0.014	0.000	⋯	⋯
XOR 11	0.004	0.000	0.000	0.000	⋯	⋯

۲-۳-۴- نقش کلاک در تاخیر مدار

همانگونه که پیشتر اشاره شد، هر کلاک دارای چهار فاز است و هر فاز به اندازه ۹۰ درجه جابجا[۱۹] شده است. این جابجایی در شکل(۱۰) نمایش داده شده است.



شکل(۱۰): جابجایی ۹۰ درجه‌ی کلاک[۲۵]

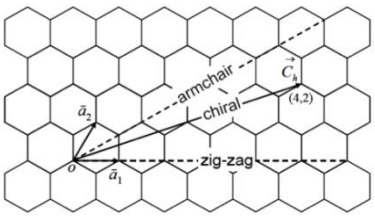
به عنوان مثال، چهار سلول کوانتومی متوالی، در هر لحظه در یکی از ستون‌های شکل(۱۰) قرار خواهند گرفت و یکی پس از دیگری مراحل چهارگانه را طی می‌نمایند. بدیهی است که تاخیر کلی یک مدار، بر اساس تعداد کلاک در مسیربحرانی[۲۰] است. لازم به ذکر است که فرکانس کاری سلول‌های کوانتومی حدود یک الی ده تراهرتز بوده، که این مقدار حدود دو برابر فرکانس کاری ترانزیستورهای اثر میدانی است [۱۷].

۲-۳-۴-۱- مشکل کلاک در سلول‌های کوانتومی

یکی از موارد بسیار مهم در طراحی کلاک، تنظیم نواحی چهارگانه است. به اینصورت که هرگونه تقسیم‌بندی نادرست مناطق کلاک، باعث بروز خطا در عملکرد مدار خواهد شد. همزمانی اعمال ورودی، طول سیم و تعداد سلول‌ها در هر فاز، از جمله مواردی است که باید مورد توجه قرار گیرد [۲۲]. به عنوان مثال در گیت اکثریت، تنظیم این نواحی در بستر مدار باید به گونه‌ای باشد که هر سه سیگنال ورودی گیت اکثریت، به صورت هم‌زمان به گیت اعمال شوند. هرگونه تأخیر در هر یک از ورودی‌ها، شکل موج غلط خروجی را به دنبال خواهد داشت. همچنین طول سیم‌ها در هر ناحیه کلاک نباید طولانی باشد، زیرا با افزایش این طول، احتمال آنکه سلول در انتهای سیم، به صورت موفقیت‌آمیزی تحت تأثیر سلول راه‌انداز در سر دیگر قرار گیرد و تغییر وضعیت دهد، کاهش می‌یابد. از طرفی طول سیم در یک ناحیه کلاک، فرکانس کلاک را تعیین می‌کند و با افزایش این طول، فرکانس کاهش می‌یابد. تعداد سلول‌ها در هر فاز کلاک نباید از حد معینی بیشتر شود، زیرا با افزایش تعداد سلول‌ها در هر فاز نه تنها فرکانس کلاک کاهش می‌یابد، بلکه به علت محدود بودن انرژی لازم جهت پلاریته شدن سلول‌ها، برخی از آنها ممکن است وضعیت نامشخصی به خود بگیرند[۲۲] .

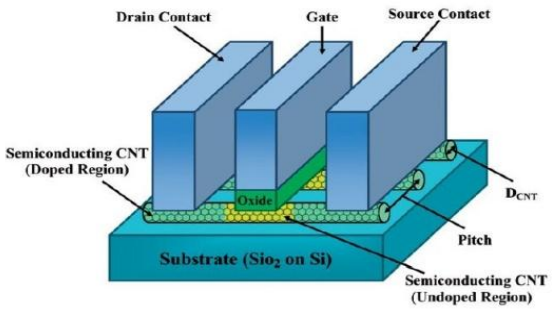
۳- ترانزیستورهای نانو لوله‌ی کربن

فناوری نانولوله‌ی کربنی در سال ۱۹۹۱ توسط دانشمندی به نام سومیو ایجیما[۲۱] کشف و تولید گردید. در یک سلول نانولوله‌ی کربنی نیز، اتم‌های کربن در ساختاری استوانه‌ای آرایش یافته که دقیقاً مشابه آرایش کربن در صفحه‌های گرافیت است. در گرافیت، شش ضلعی‌های منظم کربنی در کنار یکدیگر صفحه‌ی گرافیت را می‌سازند. در شکل(۱۱) ساختار منظم کربنی نمایش داده شده است .



شکل(۱۱): نمایش صفحه‌ی گرافیت [۹]

نانولوله‌ها به میزان قابل توجهی سخت و قوی بوده و هادی خوب جریان الکتریسیته و گرما هستند. این خواص سبب استفاده از این مواد در صنعت الکترونیک شده است. نانولوله‌های کربنی، سیم‌های مولکولی بزرگی هستند که الکترون می‌تواند آزادانه در آنها حرکت کند. از این سلول‌ها در ساختن نوع خاصی از ترانزیستورها استفاده می‌شود که به آنها ترانزیستورهای نانو لوله‌ی کربنی (CNTFET) گفته می‌شود[۲۶,۲۷] . در شکل(۱۲) ساختمان و اجزای درونی یک ترانزیستور نانو لوله‌ی کربنی نمایش داده شده است.



شکل(۱۲):ساختمان و اجزای درونی ترانزیستور نانو لوله‌ی کربن[۲۸]

این ترانزیستورها به لحاظ عملکرد و نحوه‌ی پیاده‌سازی منطق صفر و یک، شباهت زیادی به ترانزیستورهای اثر میدانی دارند، ولی از نظر تاخیر، توان مصرفی و گرمای تولید شده، در سطح بسیار مطلوبی هستند [۲۶,۲۷]. همچنین شباهت زیاد روند ساخت ترانزیستورهای نانو لوله‌ی کربنی در مقایسه با مراحل ساخت ترانزیستورهای اثر میدانی, و نیز قدرت بالای انتقال جریان آنها ، در [۹] ، مورد بحث قرار گرفته است. بدیهی است که با انجام تغییرهای مختلف در طراحی اولیه این سلول‌ها، می‌توان هر یک از پارامترهای فوق را کاهش و یا افزایش داد [۲۹,۳۰].

۳-۱- انواع نانولوله‌ی کربن

نانو لوله‌های کربنی به صورت‌های مختلف تک دیواره[۲۲] و چند دیواره[۲۳] تولید می‌شوند. قطر نانولوله‌های چند دیواره چند ده برابر قطر نانولوله‌های تک دیواره است و این در حالی است که قطر نانولوله‌های تک دیواره معمولاً از یک و یا دو نانومتر تجاوز نمی‌کند. از آنجایی که قطر نانولوله‌ها تأثیر مستقیم در کارایی اینگونه ترانزیستورها دارند، بنابراین در عمل، بیشتر، از نانولوله‌های کربنی تک دیواره استفاده شده است. همانگونه که در شکل(۱۱) قابل مشاهده است، نانولوله‌ی کربنی تک جداره به صورت یک صفحه‌ی گرافیت در نظر گرفته می‌شود که مطابق شکل(۱۳) و بر اساس بردار $\overrightarrow{C_h} = n_1.\overrightarrow{a_1} + n_2.\overrightarrow{a_2}$ ، به صورت لوله درآمده است. $[\overrightarrow{a_1}, \overrightarrow{a_2}]$ بردارهای واحد صفحه‌ی گرافیت هستند و شاخص‌های (n۲،n۱) اعداد مثبت و صحیحی هستند که نحوه‌ی پیچش لوله‌ی کربن را تعیین می‌کنند.

شکل(۱۳): صفحه‌ی گرافیت لوله شده [۹]

۳-۲- انواع نانو لوله‌های تک‌جداره

نانو لوله‌های تک جداره، بسته به عدد (n۲،n۱) به سه دسته تقسیم می‌شوند [۳۱].

نوع اول - آرم‌چیر[۲۴]

اگر n۲،n۱ با هم مساوی باشند، در این صورت بردار مجموع، نیم‌ساز دو بردار شبکه گرافیت می‌شود. اگر صفحه‌ی گرافیت را بر طبق این بردار لوله کنیم، آنگاه نانو لوله‌ی حاصل از نوع آرم‌چیر یا شکل دسته صندلی خواهد بود.

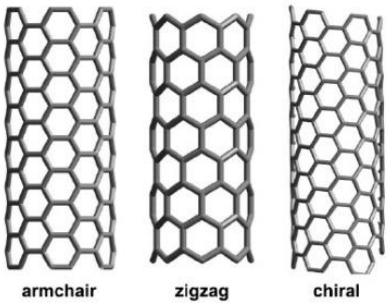
نوع دوم – زیگ زاگ[۲۵]

اگر یکی از اعداد n۱ و n۲ ، صفر باشد، آنگاه با چرخش حول محور باقیمانده، نانولوله‌ی زیگ زاگ ایجاد می‌گردد.

نوع سوم - چیرال[۲۶]

درصورتیکه هیچکدام از دوحالت قبل رخ ندهد، یعنی n۱ و n۲ غیر مساوی و غیرصفر باشند، آنگاه لوله‌ی ایجاد شده از نوع چیرال است.

این سه نمونه نانولوله‌ی توضیح داده شده، در شکل(۱۴) نمایش داده شده است.



شکل(۱۴): سه نمونه شکل نانولوله

۳-۳- پارامترهای نانولوله‌ی کربن

تغییر پارامترهای نانو لوله‌ی کربن، تأثیر مستقیم در کارایی این نوع ترانزیستورها دارد. به عنوان مثال قطر نانولوله، تعیین کننده میزان گذردهی جریان و تعیین کننده ولتاژ آستانه است. همچنین مقدار n۱ و n۲ ، نقش تاثیرگذار در خاصیت رسانایی و یا نیمه‌رسانایی نانو لوله‌ی کربن ایفا می‌کند.

۱) محیط نانولوله

محیط نانولوله‌ی کربن از طریق رابطه‌ی(۱) قابل محاسبه است [۳۱].

$$C_h = a\sqrt{n_1^2 + n_2^2 + n_1 n_2} \qquad \text{رابطه(۱)}$$

۲) قطر نانو لوله

قطر نانولوله‌ی کربن از طریق رابطه(۲) محاسبه می‌گردد [۳۱].

$$D_{CNT} = C_h / \pi \qquad \text{رابطه(۲)}$$

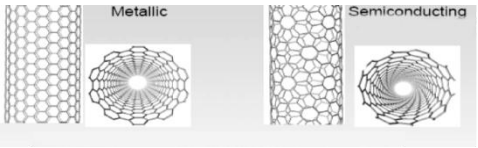
۳) ولتاژ آستانه

با فرض فاصله‌ی اتم‌های کربن، در حدود ۱,۴۴ آنگستروم، ولتاژ آستانه از رابطه‌ی(۳) محاسبه می‌گردد[۳۲] .

$$V_{th} = \frac{0.42}{D_{CNT(nm)}} V \qquad \text{رابطه(۳)}$$

۳-۴- خواص الکتریکی و رسانایی نانولوله‌ی کربن

همانگونه که قبلاً اشاره شد، با حالت‌های مختلف پارامتر (n۲، n۱) نانولوله‌های مختلفی ایجاد می‌گردد. این پارامترها در میزان رسانایی نانولوله‌ی کربن نیز اثر دارند. بدین‌صورت که چنانچه n۱ و n۲ مساوی باشند (نوع آرمچیر نانو لوله) و یا حاصل تفریق n۱-n۲ ، مضربی از عدد سه باشد، آنگاه نانولوله‌ی تولید شده دارای خاصیت رسانایی فلز خواهد بود [۳۳]. در غیر اینصورت، نانولوله‌ی تولید شده دارای خاصیت نیمه‌رسانا است. در شکل(۱۵)، دوحالت رسانایی و نیمه رسانایی نانولوله‌ی کربنی نمایش داده شده است.



شکل(۱۵): نمایش حالت رسانایی و نیمه رسانایی نانولوله‌ی کربن

۳-۵- انواع ترانزیستورهای نانو لوله‌ی کربن

خانواده‌ی ترانزیستورهای نانولوله‌ی کربن (CNTFET) از نظر عملکرد، بسیار شبیه به ترانزیستورهای اثر میدانی هستند. این شباهت عملکرد را می‌توان به صورت زیر خلاصه نمود [۹]:

۱) استفاده از خاصیت نیمه رسانایی نانو لوله‌ی کربن

۲) برقراری کانال ارتباط بوسیله نانو لوله

۳) پل زدن میان پایه‌های سورس[۲۷] و درین[۲۸]

۴) روشن و خاموش شدن ترانزیستور، بوسیله پایه گیت[۲۹]

برهمین اساس ترانزیستورهای نانولوله‌ی کربن به دوگروه کلی تقسیم می‌شوند [۹] :

الف) ترانزیستورهای نانولوله‌ی کربن شاتکی باریر[۳۰]

در ترانزیستورهای شاتکی باریر، میزان رسانایی نانولوله براساس اکثریت حامل‌هایی است که در انتها، تونل زنی را انجام می‌دهند و در نتیجه‌ی آن، مقدار جریان گذردهنده، و میزان کارایی آن بر اساس مقاومت اتصال‌ها با توجه به تونل‌های زده شده در یکی از پایانه‌های سورس و یا درین و یا هردو مشخص می‌گردد. همچنین ارتفاع و عرض این ترانزیستورها و میزان هدایت آنها، بر اساس تنظیم‌های الکترواستاتیکی گیت تعیین می‌شود [۹]. این ترانزیستورها دارای دو ایراد نسبتا مهم هستند :

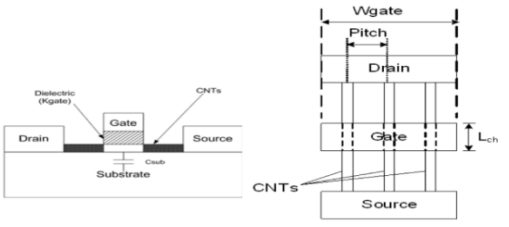
۱) کاهش جریان خروجی در حالت روشن شدن ترانزیستور

۲) داشتن رفتار و حالت چند قطبی در خروجی

لازم به تاکید است که این ایرادهای فوق موجب محدودیت در بکار-گیری آنها شده است.

ب) ترانزیستورهای نانولوله‌ی کربن شبه ترانزیستورهای اثر میدانی[۳۱]

همانگونه که از نام آنها مشخص است رفتار اینگونه ترانزیستورها شبیه ترانزیستورهای اثر میدانی سلیکونی است. بر خلاف ترانزیستورهای نانولوله‌ی کربن شاتکی باریر، دارای رفتار تک قطبی، مقیاس پذیری بالا و جریان گذردهی بالا در حالت روشن هستند و در آنها انتقال به صورت پرتابی[۳۲] ، است. به همین دلیل، این گروه از ترانزیستورهای نانولوله‌ی کربنی ، بیشتر مورد توجه قرار گرفته و ساختار آن در شکل(۱۶) قابل مشاهده است.



شکل(۱۶): ساختار ترانزیستورهای نانولوله‌ی کربنی شبه ترانزیستورهای اثر میدانی [۳۱]

در این نوع ترانزیستورها، نانولوله‌ها به صورت تزریق نشده[۳۳] بکار گرفته می‌شوند، در حالیکه سایر نقاط به صورت خیلی عمیقی تزریق شده‌اند و به عنوان نواحی درین، سورس و یا اتصال‌های درونی دو قطعه‌ی مجاور هستند. ترانزیستور نانولوله‌ی کربنی تک جداره‌ی معمول، دارای لوله‌های کربنی با قطر ۱,۴ نانومتر و پهنای باند[۳۴] حدود ۰,۶ الکترون‌ولت است. همچنین دارای گیت فوقانی[۳۵] همراه عایق و اتصال فلزکاربید[۳۶] در درین و سورس است [۳۱]. مقاومت اتصال‌ها و انحراف زیرآستانه‌ی[۳۷] اینگونه ترانزیستورها، با گونه سلیکونی موجود قابل مقایسه است. همانگونه که در ترانزیستورهای سلیکونی جریان نسبت به واحد عرض قطعه سنجیده می‌شود(مثلا میکروآمپر بر میکرومتر)، در ترانزیستورهای نانولوله‌ی کربن، جریان نسبت به تعداد لوله‌های کربن مورد سنجش واقع می‌شود [۳۱].

۳-۶- نقص‌های موجود در ترانزیستور نانولوله‌ی کربن

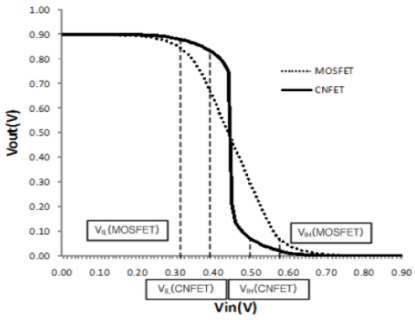
از آنجایی که روش ساخت و تولید ترانزیستورهای نانولوله‌ی کربن مشابه ترانزیستورهای اثر میدانی است، لذا درصد بسیاری از خطاها و نقص‌های این دو فناوری مشابه یکدیگر هستند. مضافاً اینکه در ترانزیستورهای نانولوله‌ی کربن احتمال وقوع خطا در تولید نانولوله‌ی کربن نیز وجود دارد. از جمله مهم‌ترین نقص‌ها می‌توان به نقص‌های زمان ساخت[۳۸]، نقص ثابت ماندن در صفر یا یک، نقص جدا شدن اتصال‌ها اشاره نمود[۳۴] .

۳-۷- مقایسه‌ی کارایی ترانزیستورهای نانولوله‌ی کربن و ترانزیستورهای اثر میدانی

با مقایسه و بررسی شبیه سازی‌ها، اینگونه استنتاج می‌شود که کارایی ترانزیستورهای نانو لوله‌ی کربنی، سیزده بار بهتر از ترانزیستورهای اثر میدانی است و دلیل آن این است که خازن تاثیرگذار گیت در ترازیستورهای کربنی به ازای هر لوله‌ی کربن در گیت، حدود چهار درصد خازن بدنه‌ی ترانزیستورهای اثر میدانی است. علاوه بر این، جریان نشتی حالت خاموش[۳۹] ترانزیستورهای نانولوله‌ی کربنی، به مراتب کمتر از انواع نمونه‌ی ترانزیستورهای اثر میدانی است [۳۱]. البته لازم به ذکر است این بهبودی در حالت ایده آل در نظر گرفته شده است و این در حالی است که نوع مقاومت تزریق شده در درین و سورس، سد شاتکی باریر در رابط‌های فلزی، خازن‌های خارجی گیت و خازن‌های سیم‌بندی میانی، می‌تواند پارامترهای مورد محاسبه را از حالت ایده‌آل دور سازد [۳۳].

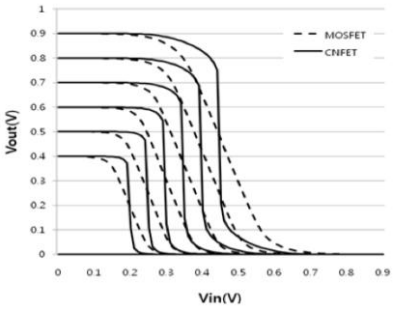
۳-۷-۱- مقایسه‌ی گیت معکوس کننده

برای مقایسه‌ی کارایی، ابعاد ۳۲ نانومتر در نظر گرفته شده است. ازآنجایی که در ترانزیستورهای نانو لوله‌ی کربنی، نرخ گذردهی نوع p به n ، برابر است[۳۵] ، نسبت ۱ به ۱ در نظر گرفته شده است. همچنین جهت تقارن بیشتر نمودار خصوصیت ولتاژ خروجی[۴۰] ، نرخ گذردهی میان ماسفت نوع p به n به صورت ۳ به ۱ در نظر گرفته شده است. این نمودار در شکل(۱۷) نمایش داده شده است.



شکل(۱۷): نمودار VTC در گیت معکوس‌کننده [۳۵]

همانگونه که به وضوح در شکل(۱۷) مشاهده می‌شود، نمودار ترانزیستورهای اثر میدانی، دارای شیب تندتر بوده و مطلوب ما است و محدوده اختلال[۴۱] را حدود ۲۲,۵ درصد بهبود می‌دهد [۳۵]. نکته‌ی جالب اینکه، این وضعیت در ولتاژهای پایین‌تر نیز وجود دارد و در شکل(۱۸) نمایش داده شده است.



شکل(۱۸): نمودار VTC گیت معکوس کننده در ولتاژ پایین [۳۵]

لازم به توضیح است که در ترانزیستورهای اثر میدانی، نرخ گذردهی میان ترانزیستور p و n، با عرض گیت قابل تغییر بوده و این مسأله در ترانزیستورهای نانو لوله‌ی کربن، با تغییر تعداد لوله‌های کربنی و یا تغییر در قطر آنها انجام می‌شود.

۳-۷-۲- توان مصرفی حاصل‌ضرب انرژی در تاخیر (PDP)[۴۲]

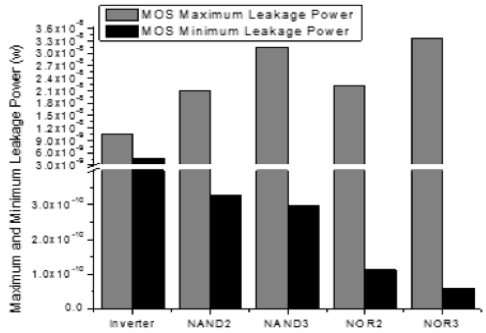
با بررسی نتایج شبیه‌سازی و در ابعاد ۳۲ نانومتر، توان مصرفی PDP در ترانزیستورهای اثر میدانی، حدود ۱۰۰ برابر ترانزیستورهای نانولوله‌ی کربنی است. این نتایج درجدول(۴) قابل مشاهده است [۳۳].

جدول(۴): مقایسه‌ی حاصل‌ضرب انرژی در تاخیر(PDP)

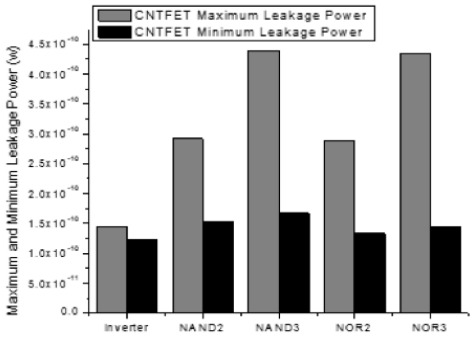
		Delay (sec)	Power (watt)	PDP (joule)
MOSFET	Inverter	1.77E-11	1.39E-06	2.46E-17
	NAND2	2.26E-11	1.96E-06	4.41E-17
	NAND3	2.99E-11	2.77E-06	8.29E-17
	NOR2	3.97E-11	2.58E-06	1.02E-16
	NOR3	6.97E-11	4.04E-06	2.82E-16
CNTFET	Inverter	2.42E-12	1.11E-07	2.69E-19
	NAND2	3.49E-12	1.89E-07	7.41E-19
	NAND3	5.06E-12	2.90E-07	1.47E-18
	NOR2	3.50E-12	1.85E-07	6.48E-19
	NOR3	5.08E-12	2.73E-07	1.39E-18

۳-۷-۳- بررسی جریان نشتی[۴۳]

با ادامه‌ی روند کاهش اندازه در ابعاد نانو، روش‌های معمول کاهش توان مصرفی دینامیک[۴۴]، تأثیر پذیری کمتری داشته است، زیرا این کاهش ابعاد، از طرف دیگر موجب افزایش توان مصرفی استاتیک[۴۵] شده است[۳۶] . به طورعمومی، توان مصرفی نشتی به ورودی اعمال شده از مدار وابسته نیست. نتایج شبیه‌سازی‌ها برای ابعاد ۳۲ نانومتر در [۳۱] ، حاکی از آن است که حداکثر توان مصرفی نشتی در ترانزیستورهای اثر میدانی، حدود هفتادوپنج برابر، بیشتر از خانواده‌ی ترانزیستورهای نانو لوله‌ی‌کربن است. همچنین حداقل توان مصرفی نشتی در ترانزیستورهای اثر میدانی، حدود ۳ برابر بیشتر از انواع ترانزیستورهای نانو لوله‌ی‌کربن است. نمودارهای مربوطه در شکل(۱۹) و شکل(۲۰)، قابل مشاهده است.



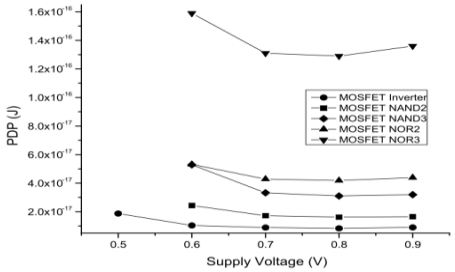
شکل(۱۹): نمودار توان مصرفی نشتی برای‌خانواده‌ی ترانزیستورهای اثر میدانی [۳۱]



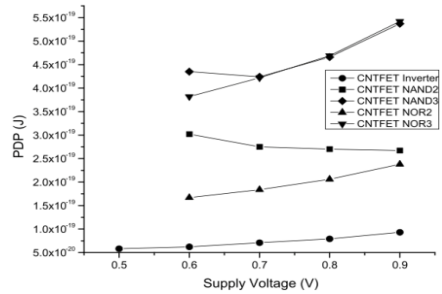
شکل(۲۰): نمودار توان مصرفی نشتی در خانواده‌ی ترانزیستورهای نانولوله‌ی‌کربن [۳۱]

۳-۷-۴- تغییر ولتاژ

کاهش توان مصرفی با توجه به کاهش ولتاژ منبع تغذیه، به افزایش حساسیت نسبت به تغییر ولتاژ منجر می‌گردد و این مساله، یک مفهوم پایه برای تشخیص میزان کارایی فناوری‌های مختلف است. به شکل(۲۱) و شکل(۲۲) توجه فرمایید. با مقایسه‌ی این دو شکل نتیجه به اینصورت است که با کاهش ولتاژ منبع تغذیه، خانواده‌ی ترانزیستورهای نانولوله‌ی کربن تغییر PDP بیشتری نسبت به خانواده‌ی ترانزیستورهای اثر میدانی دارند.لازم به تاکید است که با این شرایط در مجموع، ترانزیستورهای نانولوله‌ی کربن PDP کمتری را دارا هستند.

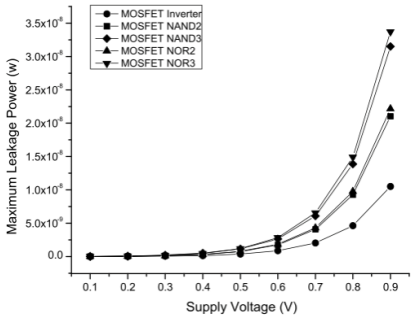


شکل(۲۱): تغییر ولتاژ منبع تغذیه در ترانزیستورهای اثر میدانی و PDP [۳۳]

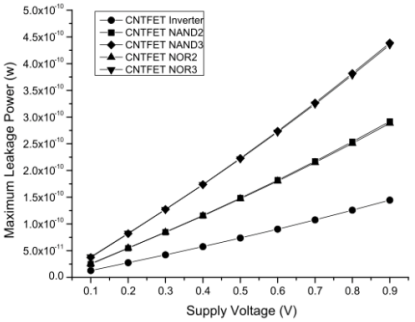


شکل(۲۲): تغییر ولتاژ منبع تغذیه در ترانزیستورهای نانولوله‌ی کربن و PDP [۳۳]

همچنن حداکثر توان مصرفی نشتی در ترانزیستورهای اثر میدانی، با افزایش ولتاژ منبع تغذیه به صورت نمایی، و در ترانزیستورهای نانولوله‌ی کربن، به صورت خطی افزایش می‌یابد. نمودار شکل(۲۳) و شکل(۲۴) این موضوع را بخوبی نشان داده است. نهایتاً ترانزیستورهای اثر میدانی، توان مصرفی نشتی بیشتری نسبت به ترانزیستورهای نانو لوله‌ی کربن دارند [۳۳].



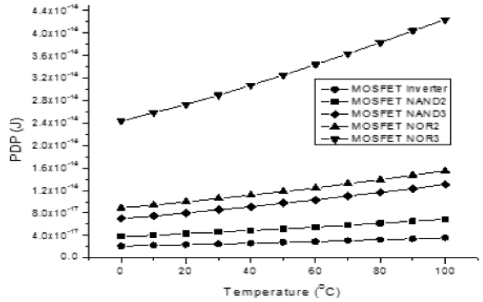
شکل(۲۳): تغییر ولتاژ منبع تغذیه در ترانزیستورهای اثر میدانی و توان مصرفی نشتی [۳۳]



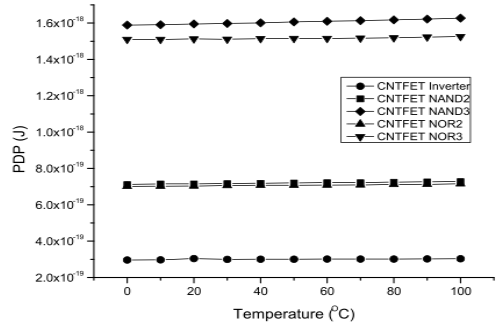
شکل(۲۴): تغییر ولتاژ منبع تغذیه در ترانزیستورهای نانو لوله‌ی کربن و توان مصرفی نشتی [۳۳]

۳-۷-۵- تغییر دما

افزایش سرعت کاری مدارهای دیجیتال و توان مصرفی بالا، موجب افزایش دما در سطح تراشه[۴۶] می‌گردد. همچنین افزایش بی-رویه‌ی توان مصرفی و دما، موجب افزایش خرابی‌های زمان اجرا و کاهش جدی قابلیت اطمینان[۴۷] سیستم می‌شود [۳۶]. در شکل(۲۵) و شکل(۲۶) نمودار تغییر دما و توان مصرفی PDP نمایش داده شده است.

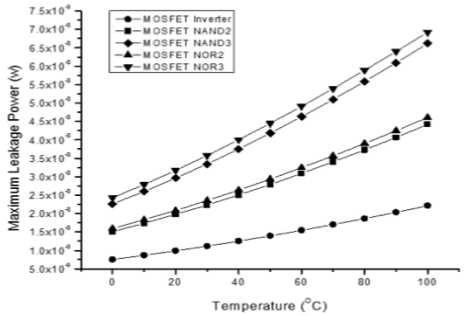


شکل(۲۵): تغییر دما در ترانزیستورهای اثر میدانی و PDP [۳۳]

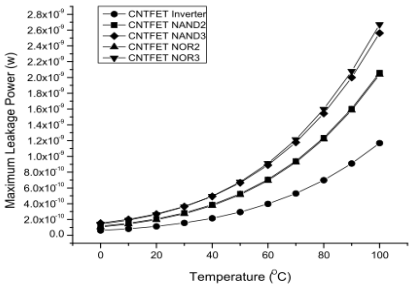


شکل(۲۶): تغییر دما در ترانزیستورهای نانو لوله‌ی‌کربن و PDP [۳۳]

در خانواده‌ی ترانزیستورهای نانولوله‌ی کربن، با افزایش دما، توان مصرفی PDP ثابت مانده است و این در حالی است که در خانواده‌ی ترانزیستورهای اثر میدانی، با روند افزایش دما، افزایش توان مصرفی PDP را به همراه داشته است. همچنین با افزایش دما، حداکثر توان مصرفی نشتی در ترانزیستورهای اثر میدانی، به صورت تقریبا خطی و در ترانزیستورهای نانولوله‌ی کربن به صورت نمایی افزایش می‌یابد. این نتیجه نیز در شکل(۲۷) و شکل(۲۸) قابل مشاهده است [۳۱].



شکل(۲۷): تغییر دما در ترانزیستورهای اثر میدانی و حداکثر توان مصرفی نشتی [۳۱]



شکل(۲۸): تغییر دما در ترانزیستورهای نانولوله‌ی کربن و حداکثر توان مصرفی نشتی [۳۱]

۳-۷-۶- فضای اشغالی

از آنجایی که ترانزیستورهای نانو لوله‌ی کربن از نظر مراحل ساخت و عملکرد، شباهت زیادی به ترانزیستورهای اثر میدانی دارند، فضای اشغال شده توسط اینگونه ترانزیستورها، تفاوت چشمگیری ندارد. در [۳۷] میزان فضای اشغال شده و میزان بهبودی مدارها، در ترانزیستورهای اثر میدانی و ترانزیستورهای نانولوله‌ی‌کربن مورد بررسی و ارزیابی قرار گرفته است. میانگین کاهش فضای اشغالی در ترانزیستورهای نانولوله‌کربن ، در جدول(۵) نمایش داده شده است.

جدول(۵): بهبود فضای اشغالی ترانزیستورهای نانو لوله‌ی کربن [۳۷]

ابعاد	۱۶ نانومتر	۲۲ نانومتر	۴۵ نانومتر
درصد کاهش	٪ ۲۴,۷۲	٪ ۲۹,۲۹	٪ ۶,۹۳

۴- مقایسه‌ی معیارهای مختلف فناوری سلول‌های کوانتومی و ترانزیستورهای نانولوله‌ی کربن

در این بخش، خلاصه نتایج و بهبود فناوری سلول‌های کوانتومی و نانولوله‌های کربن، در مقایسه با ترانزیستورهای اثر میدانی، در جدول(۶) ارایه می‌گردد.

جدول(۶): نتایج بهبود فناوری سلول‌های کوانتومی و نانولوله‌های کربن نسبت به ترانزیستورهای اثر میدانی

معیار / فناوری	سلول‌های کوانتومی	نانو لوله‌های کربن
PDP	قابل چشم پوشی	۱۰۰ برابر کمتر
توان مصرفی‌نشتی	صفر	۷۵ برابر کمتر
فضای اشغالی	۱۰۰ برابر کاهش	حداکثر ۳۰ ٪ کاهش
احتمال وقوع نقص	ده تا صد برابر بیشتر	تقریبا برابر
اثر تغییر دما	دارای رفتار نامناسب	دارای حساسیت کمتر
تغییر ولتاژ منبع تغذیه	بدون اثر	دارای رفتارخطی و نمایی بهتر

۵- طرح پیشنهادی

سنتز توابع منطقی، یکی از مراحلی است که در تولید مدارهای دیجیتال مورد توجه فراوان قرار گرفته است. فناوری سلول‌های کوانتومی نیز از این قاعده مستثنا نیست. چیدمان فیزیکی اجزای پایه‌ی سلول‌های کوانتومی، از نظر طول سیم، فواصل اجزا، نواحی کلاک وغیره، دارای محدودیت‌هایی است [۲۲]، که جانمایی تولید شده را از نظر فضا و تخصیص منابع، از حالت بهینه دور می‌سازد. همچنین در موارد و شرایط خاص، نحوه‌ی پیاده‌سازی توابع منطقی، می‌تواند باعث تولید پاسخ نادرست گردد [۲۴]. در [۳۸,۳۹,۴۰] روش‌هایی جهت ساده‌سازی توابع منطقی مطرح و در [۴۱] نرم افزار تولید خودکار جانمایی، تولید شده است که بنا به اظهار نویسنده، پاسخ بدست آمده بهینه نیست. بدیهی است که بدست آوردن مدار بهینه، از همان مراحل ابتدایی باید مورد توجه واقع شود. این مراحل شامل ساده‌سازی توابع منطقی، طراحی ماژولار جهت اتصال صحیح اجزا و نهایتاً چیدمان و تخصیص منابع است. بنابراین، موضوع پروژه، ارایه‌ی راهکار جدید در بهبود سنتز توابع و گیت‌های منطقی، در فناوری سلول‌های کوانتومی، پیشنهاد می‌گردد.

۶- نتیجه‌گیری

در این سمینار، دو فناوری سلول‌های کوانتومی و ترانزیستورهای نانولوله‌ی کربن مورد بررسی و ارزیابی قرار گرفتند. با توجه به جدول(۶) و با در نظر داشتن سه معیار توان مصرفی، فضای اشغالی و تاخیر مدار، اینگونه می‌توان نتیجه گرفت که سلول‌های کوانتومی کاندیدای مناسب‌تری جهت جایگزینی ترانزیستورهای اثر میدانی در ابعاد نانو هستند. این در حالیست که سلول‌های کوانتومی دارای احتمال وقوع نقص بالاتری نسبت به ترانزیستورهای نانولوله‌ی کربن هستند. همچنین با توجه به حساسیت پایین ترانزیستورهای نانولوله‌ی کربن در مقابل تغییر دما از یک طرف، و شباهت زیاد آنها به ترانزیستورهای اثر میدانی ، با امکان بهره‌مند شدن از روش‌های ساخت و تولید آنها از طرف دیگر، بنظر میرسد در آینده‌ی نزدیک، ترانزیستورهای نانولوله‌ی کربن، جایگزین مناسب‌تری برای ترانزیستورهای اثر میدانی بشمار بروند. بدیهی است که اگر مشکل‌ها و چالش‌های سلول‌های کوانتومی مورد بررسی و آزمایش بیشتری قرار گیرند، و پژوهش‌ها و تحقیق‌های لازم جهت افزایش قابلیت اطمینان این فناوری صورت پذیرد، سلول‌های کوانتومی به مراتب جایگزین مناسب‌تری برای ترانزیستورهای اثر میدانی محسوب می‌شوند.

۷- مراجع

[۱] Kenichi.Higashi, Yasumori.Fukushima, Alberto.O.Adan, "Analytical Threshold Voltage Model for Ultrathin SOI MOSFET's Including Short-Channel and Floating-Body Effects", in IEEE Transactions On Electron Devices, pp. ۷۲۹-۷۳۷, ۱۹۹۹.

[۲] Keunwoo.Kim, Ching.Te.Chuang, Christophe Tretz Meng.Hsueh.Chiang, "High-Density Reduced-Stack Logic Circuit Techniques Using Independent-Gate Controlled Double-Gate Devices", in IEEE Transactions On Electron Devices, pp. ۲۳۷۰-۲۳۷۷, ۲۰۰۶.

[۳] I.N.Hajj S.Bobba, "Maximum Leakage Power Estimation for CMOS Circuits", in Low-Power Design Proceedings,IEEE Alessandro Volta

Journal of Computer and Electrical Engineering, vol. ۲, pp. ۱۷۹۳-۸۱۶۳, February ۲۰۱۰.

[۱۶] Karim Mohamadi, Mohamad Javid, "Characterization and Tolerance of QCA Full Adder Under Missing Cells Defects", in International Conference on MEMS, NANO, and Smart Systems (ICMENS), pp. ۸۵-۸۸, ۲۰۱۰.

[۱۷] Subhra Mazumdar, Sudip Roy, Rajib Mall, Mayur Bubna, "Designing Cellular Automata Structures using Quantum-dot Cellular Automata", in IEEE International Conference on High Performance Computing, pp. ۳۰۱-۳۰۶, ۲۰۰۷.

[۱۸] G.Athisha, S.Karthigai lakshmi, "Efficient Design of Logical Structures and Functions using Nanotechnology Based Quantum Dot Cellular Automata Design", International Journal of Computer Applications, vol. ۳, pp. ۳۵-۴۲, June ۲۰۱۰.

[۱۹] Peter.M.Kogge, Timothy.J.Dysart, "System Reliabilities when Using Triple Modular Redundancy in Quantum-Dot Cellular Automata", in IEEE International Symposium on Defect and Fault Tolerance of VLSI Systems, pp. ۷۲-۸۰, ۲۰۰۸.

[۲۰] P.Douglas.Tougaw, Porod.Wolfgang, Gary.H.Bernstein, Craig.S.Lent, "Quantum Cellular Automata", in IEEE Internation Symposium, pp. ۴۹-۵۷, ۱۹۹۳.

[۲۱] Sudeep.Sarkar, Sanjukta.Bhanja, Saket.Srivastava, "Estimation of Upper Bound of Power Dissipation in QCA Circuits", in IEEE Transactions On NanoTechnology, pp. ۱-۱۲, ۲۰۰۸.

[۲۲] Kaijie.WU, Ramesh.KARRI, Kyosun.KIM, "Quantum-Dot Cellular Automata Design Guideline", in IEICE Transactions Fundamentals, pp. ۱۶۰۷-۱۶۱۴, ۲۰۰۶.

[۲۳] D.Z.Chen, Dysart.T.J, Hu.X.S, Kahng.A.B, Kogge.P.M, Murphy.R.C., Niernier.M.T, D.A.Antonelli, "Quantum-Dot Cellular Automata (QCA) Circuit Partitioning: Problem Modeling and Solutions", in Design Automation Conference, pp. ۳۶۳-۳۶۸, ۲۰۰۴.

[۲۴] B.D.Padgett, M.Kuntzman, M.K.Hendrichsen, I.Sturzu, G.A.Anduwan, "Fault-Tolerance and Thermal Characteristics of Quantum-Dot Cellular Automata Devices", Journal Of Applied Physics, vol. ۱۱, pp. ۱۱۴۳۰۶-۱۱۴۳۰۶-۹, ۲۰۱۰.

[۲۵] Anik Sengupta, Mamata Dalui, Biplab K Sikdar, Bibhash Sen, "Design of Testable Universal Logic Gate Targeting Minimum Wire-Crossings in QCA Logic Circuit", in ۱۳th Euromicro Conference on Digital System Design: Architectures, Methods and Tools, pp. ۶۱۳-۶۲۰, ۲۰۱۰.

[۲۶] L. Anghel, R. Leveugle, T. Dang, "CNTFET Basics and Simulation", in Design and Test of Integrated Memorial Workshop, pp. ۱۱۶-۱۲۴, ۱۹۹۹.

[۴] Niraj.K.Jha Kamal.S.Khouri, "Leakage Power Analysis and Reduction During Behavioral Synthesis", in IEEE Transaction On Very Large Scale Integration (VLSI) Systems, pp. ۸۷۶-۸۸۵, ۲۰۰۲.

[۵] A.Doostaregan, K.Navi, M.H.Moaiyeri, "Design of Energy-Efficient and Robust Ternary Circuits for Nanotechnology", in International Conference IET Circuits Devices Systems, pp. ۲۸۵–۲۹۶, ۲۰۱۱.

[۶] AK.Abu, M.El.Banna, M.A.Hakim, El.Seoud, "On Modelling and Characterization of Single Electron Transistor", International Journal of Electronics ۹۴, vol. ۶, pp. ۵۷۳-۵۸۵, ۲۰۰۷.

[۷] Pedro.A.Derosa, Luis.E.Cordova, Brian.H.Bozard Jorge.M.Seminario, "A Molecular Device Operating at Terahertz Frequencies: Theoretical Simulations", in IEEE Transactions On Nanotechnology, pp. ۲۱۵-۲۱۸, ۲۰۰۴.

[۸] Jianguo.Wang, Jian-Ping.Wang Hao.Meng, "A Spintronics Full Adder For Magnetic CPU", in IEEE Electron Device Letters, pp. ۳۶۰-۲۶۲, ۲۰۰۵.

[۹] Jie.Deng, "Device Modeing and Circuit Performance Evaluation for Nanoscale Devices: Silicon Technology beyond ۴۵nm Node and Carbon Nanotube Field Effect Transistors", PHD Thesis , Stanford University ۲۰۰۷.

[۱۰] Craig.S.Lent, P.Douglas.Tougaw, "Logical Devices Implemented Using Quantum Cellular Automata", AIP Journals & Magazines, vol. ۷۵, pp. ۱۸۱۸-۱۸۲۵, ۱۹۹۴.

[۱۱] D.Z.Chen, Dysart.T.J, Hu.X.S, Kahng.A.B, Kogge.P.M, Murphy.R.C., Niernier.M.T, D.A.Antonelli, "Quantum-Dot Cellular Automata", in Solid-State and Integrated Circuits Technology, pp. ۸۷۵-۸۸۰, ۲۰۰۴.

[۱۲] Farzaneh Ahmadi Kakhki, Ehsan Rahimi, Shahram Mohammad Nejad, "A Simple Mathematical Model for Clocked QCACells", in International Conference on Communication Systems Networks and Digital Signal Processing, pp. ۳۵۱-۳۵۴, ۲۰۱۰.

[۱۳] Abhineet.S.Tomar, Sanjay.Modi, "Logic Gate implementations for Quantum Dot Cellular Automata", in International Conference on Computational Intelligence and Communication Network, pp. ۵۶۵-۵۶۷, ۲۰۱۰.

[۱۴] P. DOUGLAS TOUGAW, CRAIG S. LENT, "A Device Architecture for Computing with Quantum Dots", IEEE Journals & Magazines, vol. ۸۵, no. ۴, pp. ۵۴۱-۵۵۷, ۱۹۹۷.

[۱۵] Pijush Kanti Bhattacharjee, "Digital Combinational Circuits Design By QCA Gates", International

Micro, vol. ۳, pp. ۱۷۸-۱۸۸, ۲۰۱۱.

[۳۸] Pallav.Gupta, Niraj.K.Jha, Rui.Zhang, "Majority and Minority Network Synthesis With Application to QCA-, SET-, and TPL-Based Nanotechnologies", in IEEE Transactions On Computer-Aided Design Of Integrated Circuits and System, pp. ۱۲۳۳-۱۲۴۵, ۲۰۰۷.

[۳۹] Mohammed Niamat, Srinivasa.Vemuru, Peng.Wang, "Minimal Majority Gate Mapping of ۴-variable Functions for Quantum Cellular Automata", in IEEE International Conference on Nanotechnology, pp. ۱۳۰۷-۱۳۱۲, ۲۰۱۱.

[۴۰] Yun.Shang, Ruqian.Lu, Kun.Kong, "An Optimized Majority Logic Synthesis Methodology for Quantum-Dot Cellular Automata", in IEEE Transactions On NanoTechnology, pp. ۱۷۰-۱۸۳, ۲۰۱۰.

[۴۱] Leonel.Sousa, Tiago.Teodosio, "QCA-LG: A tool for the automatic layout generation of QCA combinational circuits", in Norchip, pp. ۱-۵, ۲۰۰۷.

[۴۲] Yong-Bin Kim, Fabrizio Lombardi, Geunho Cho, "Assessment of CNTFET Based Circuit Performance and Robustness to PVT Variations", in IEEE International Midwest Symposium, pp. ۱۱۰۶-۱۱۰۹, ۲۰۰۹.

Systems in Nanoscale Technology, pp. ۲۸-۳۳, ۲۰۰۶.

[۲۷] Kartik Mohanram, Giovanni De Micheli, M.Haykel Ben-Jamaa, "An Efficient Gate Library for Ambipolar CNTFET Logic", in IEEE Transactions On Coputer-Aided Design Of Integrated Circuits and Systems, vol. ۳۰, pp. ۲۴۲-۲۵۵, ۲۰۱۱.

[۲۸] Reza Faghih Mirzaee, Keivan Navi, Amir Momeni, Mohammad Hossein Moaiyeri, "Design and Analysis of a High Performance CNFET-based Full Adder", International Journal of Electronics, vol. ۱, pp. ۱۱۳-۱۳۰, ۲۰۱۱.

[۲۹] K. Kalyan Babu, B.Rambabu, Y. Srinivasa Rao, P. Swapna, "Ambipolar CNTFET: Basic Characterization And Effect Of High Dielectric Material", in Transactions on Electrical and Electronic Materials International Conference, pp. ۱-۴, ۲۰۱۱.

[۳۰] K.Subbulakshmi, R.V.Jananee, T.Ravi, Navin.Greeni, "Study of Materials and Low Power Techniques for CNTFET", in Emerging trends in robotics and communication technologies international conference, pp. ۲۰۶-۲۰۸, ۲۰۱۱.

[۳۱] Young Bok Kim, "Design Methodology Based on Carbon Nanotube Field Effect Transistor(CNFET)", PHD Thesis, Northeastern University ۲۰۱۱.

[۳۲] P.Keshavarzian, M.Salari.Sardoueyeh, A.Shojaei, B.Naji.Givi, S.A.Ebrahimi, "Ultra-Low Power and High Speed Full Adder Based-on CNTFET", European Journal of Scientific Research, vol. ۸۰, pp. ۳۵۸-۳۶۵, ۲۰۱۲.

[۳۳] Yong-Bin.Kim, "Integrated Circuit Design Based on Carbon Nanotube Field Effect Transistor", in Transactions On Electrical and Electronic Materials, vol. ۱۲, pp. ۱۷۵-۱۸۸, ۲۰۱۱.

[۳۴] David de Andres, Juan-Carlos Ruiz, Pedro Gil, Daniel.Gil, "Impact of Manufacturing Defects on Carbon Nanotube Logic Circuits", in Dependable and Secure Nanocomputing Conference, pp. ۳۰۱-۳۰۵, ۲۰۰۹.

[۳۵] H.-S. Philip Wong, Jie Deng, "A Compact SPICE Model for Carbon-Nanotube Field-Effect Transistors Including Nonidealities and Its Application—Part II: Full Device Model and Circuit Performance Benchmarking", in IEEE Transactions on Electron Devices, pp. ۳۱۹۵-۳۲۰۵, ۲۰۰۷.

[۳۶] Yong-Bin.Kim, Kyung.Ki.Kim, "Optimal Body Biasing for Minimum Leakage Power in Standby Mode", in IEEE International Symposium on Circuits, pp. ۱۱۶۱-۱۱۶۴, ۲۰۰۷.

[۳۷] Ali Jahaniana, Keivan Navi, Mohammad Hossein Moaiyeri, "Comparative Performance Evaluation of Large FPGAs with CNFET- and CMOS-based Switches in Nanoscale", European Journal of Nano-

---

۱ Quantum Cellular Automata (QCA)

۲ Carbon Nano Tube Field-Effect Transistor (CNTFET)

۳Metal-Oxide Semiconductor Field-Effect Transistor (MOSFET)

۴ Gordon Moore

۵ Sinlge Electron Transistor

۶ Molecular Device

۷ Spintronic

۸ Tera Hertz

۹ Majority gate

۱۰ Switch

۱۱ Hold

۱۲ Release

۱۳ Relax

۱۴ Missing Cell Defect

۱۵ Extra Cell Defect

۱۶ Displacement Cell Defect

۱۷ Rotated Cell Defect

۱۸ Fixed Cell Defect

۱۹ Shift

۲۰ Critical Path

۲۱ Sumio Iijima

۲۲ Single Walled

[۲۳] Multi Walled
[۲۴] Arm Chair
[۲۵] Zig Zag
[۲۶] Chiral
[۲۷] Source
[۲۸] Drain
[۲۹] Gate
[۳۰] Schottky Barrier CNTFET
[۳۱] MOSFET-Like CNTFET
[۳۲] Ballestic
[۳۳] Undope
[۳۴] Bandgap
[۳۵] Top Gate
[۳۶] Metal Carbide
[۳۷] Subthreshold Slope
[۳۸] Manufacturing Fault
[۳۹] OFF-State Leakage Current
[۴۰] VTC(Voltage Transfer Characteristic)
[۴۱] Noise Margine
[۴۲] PDP(Power Delay Product)
[۴۳] Leakage Current
[۴۴] Dynamic Power Consumption
[۴۵] Static Power Consumption
[۴۶] Chip
[۴۷] Reliability

بررسی انواع حملات مبتنی بر تحلیل توان در سامانه‌های تعبیه‌شده و راه‌های مقابله با آن

بهنام رحمانی[۱] ، دکتر احمد پاطوقی[۲]، دکتر محمود فتحی[۳]

[۱] دانشجوی کارشناسی ارشد، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
تهران، ایران
Rahmani_b@comp.iust.ac.ir

[۲] استادیار گروه سخت‌افزار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
تهران، ایران
patooghy@iust.ac.ir

[۳] دانشیار گروه سخت‌افزار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
تهران، ایران
mahfathy@iust.ac.ir

چکیده

امروزه تراشه‌هایی که توانایی ذخیره و پردازش حجم بالایی از اطلاعات را دارند به طور وسیعی در سامانه‌های تعبیه‌شده مورد استفاده قرار می‌گیرند. وسایلی مانند تلفن‌های همراه، کارت‌های اعتباری بانکی و کارت‌های شناسایی هوشمند نمونه‌هایی از این سامانه‌ها هستند. اکثر این کاربردها حاوی اطلاعات با ارزشی هستند که امنیت آن‌ها باید از هر دو دیدگاه نرم‌افزاری و سخت‌افزاری تأمین گردد. یکی از جدی‌ترین چالش‌های امنیت در این سامانه‌ها حملات تحلیل توانی است. اساس کارکرد این دسته از حملات بر شنود مشخصه‌های سخت‌افزاری جهت کشف رمز بنا نهاده شده است. به‌عبارت دیگر، در حملات تحلیل توانی مهاجم خارجی با استخراج الگوی مصرف توان در سامانه و تحلیل این الگو و مطابقت دادن آن با توان مصرفی اندازه‌گیری شده از سیستم مورد نظر می‌تواند به اطلاعات محرمانه دسترسی پیدا کند. در این گزارش قصد داریم پس از معرفی انواع حملات مبتنی بر تحلیل توان، به تشریح روشهای مختلف مقابله با این نوع از حملات بپردازیم و سپس بهبود امنیت سامانه در حضور این روشها را ارزیابی کنیم.

کلمات کلیدی

کانال-جانبی، حملات مبتنی بر تحلیل توان، آنالیز توان تفاضلی، منطق پیش‌شارژ دوخطی.

۱- مقدمه

حملات کانال-جانبی[1] با دور زدن پارامترهای امنیتی سیستم‌های رمزنگار[2] موفق به استخراج داده‌های باارزش سیستم می‌شوند. این کار با استفاده از اطلاعاتی انجام می‌گیرد که اصطلاحاً کانال-جانبی نامیده می‌شوند. کانال-جانبی هر سیستم به دو دسته ورودی و خروجی تقسیم می‌گردد[۲]. توان مصرفی، تشعشعات الکترومغناطیسی، نور، زمان و صوت نمونه‌هایی از کانال-جانبی خروجی یک سیستم هستند[۹]. دسته‌ای دیگر از کانال‌های جانبی مانند منبع تغذیه، دما، نور و هر سیگنال ورودی به‌غیر از ورودی‌های سیستم رمزنگار کانال‌های جانبی ورودی سیستم به حساب می‌آیند[۳,۲]. با توجه به این دسته‌بندی، حملاتی که با استفاده از یک کانال-جانبی خروجی به سیستم اعمال می‌شوند حملات کانال-جانبی غیرفعال[3] و در مقابل حملاتی که با استفاده از کانال‌های جانبی ورودی به سیستم اعمال می‌گردند حملات کانال-جانبی فعال[4] و یا حملات تزریق اشکال[5] نامیده می‌شوند[۹].

از دید طراح سیستم حملات کانال-جانبی غیرفعال به دو دلیل خطرناک‌تر هستند. اول اینکه پیاده‌سازی و اجرای یک حمله کانال-

[1] Side-Channel
[2] Crypto-System
[3] Passive
[4] Active
[5] Fault Injection

جانبی مانند حملات توانی در مقایسه با روش‌های دیگر بسیار کم هزینه است و نیز با اندک تغییر می‌توان روش‌های قبلی را بر روی سیستم‌های جدیدتر بکار برد. دوم اینکه به دلیل نحوه عملکرد این نوع از حملات، مالک سیستم عموماً متوجه سرقت اطلاعات نخواهد شد[۲]. بنابراین سارق بدون اطلاع مالک سیستم می‌تواند اقدام به استخراج و سوءاستفاده از اطلاعات حساس سیستم نماید.

در مقابل، نحوه عمل حملات تزریق اشکال متفاوت است. در این نوع از حملات مهاجم با اعمال یک ورودی غیرمعتبر از طریق یک کانال-جانبی ورودی باعث می‌شود تا سیستم از کارکرد عادی خود خارج گردد و در نتیجه مهاجم می‌تواند با دور زدن پارامترهای امنیتی اقدام به استخراج رمز نماید[۱].

از بین حملات کانال-جانبی غیرفعال، حملات تحلیل توانی که برای اولین بار توسط کوچر [۲۸] معرفی گردیدند، به دلیل شانس موفقیت بالایی که در تشخیص اطلاعات حساس سیستم دارند از توجه ویژه‌ای هم برای طراحان مدار و هم برای مهاجمین برخوردار هستند[۶-۱۲]. علت این امر این می‌باشد که در مدارات مبتنی بر منطق CMOS میزان توان مصرفی مدار متناسب با دو پارامتر ۱) داده در حال پردازش و ۲) عملیات در حال انجام است، بنابراین با تعیین مقدار داده در حال پردازش و یا عملیات در حال اجرا می‌توان اقدام به تخمین میزان مصرف توان در سیستم نمود.

حملاتی مانند تحلیل توان ساده[۱] (SPA) ، حملات تحلیل توان تفاضلی[۲] (DPA) ، حملات تحلیل توان تطابقی[۳] (CPA) انواع مختلفی از حملات توانی هستند[9]. به منظور مقابله با این دسته از حملات روشهای مختلفی در سطوح انتزاع متفاوت ارائه گردیده است که بصورت مشروح درباره آنها صحبت خواهد شد.

این گزارش بصورت زیر ساختاربندی شده است : در بخش دوم به معرفی انواع حملات مبتنی بر تحلیل توان پرداخته خواهد شد. در بخش سوم انواع روشهای مقابله با حملات توانی به صورت دسته‌بندی شده معرفی خواهد شد و سپس به صورت خاص به تشریح مداراتی که اصطلاحاً منطق پیش‌شارژ دوخطی[۴](DRP) نامیده می‌شوند پرداخته می‌شود. بخش چهارم شامل طرح پیشنهادی برای پروژه بوده و در بخش پنجم نیز مطالب مربوط به جمع‌بندی و نتیجه‌گیری آورده شده است.

۲- حملات تحلیل توانی

حملات تحلیل توانی یکی از کارآمدترین حملات کانال-جانبی هستند[۹,۲]. به‌گونه‌ای که مطالعات بسیاری به منظور تحلیل نحوه عملکرد و مقابله با این حملات انجام گرفته است[۲۸,۹,۳,۲,۱]. در این روش مهاجم خارجی با تحلیل توان مصرفی سامانه تعبیه‌شده اقدام به استخراج اطلاعات حساس شامل کلید استفاده‌شده در سیستم‌های رمزنگار می‌نماید[۹,۲].

به منظور بررسی عملکرد این دسته از حملات از آنجا که اکثر تراشه‌های امروزی با استفاده از تکنولوژی CMOS ساخته می‌شوند در ابتدا لازم است اجزاء توان مصرفی در تراشه‌ها مورد بررسی قرار گیرند. در تکنولوژی CMOS میزان توان مصرفی متناسب با تعداد انتقال‌های $1 \rightarrow 0$ و بالعکس رخ داده در خروجی دروازه‌های منطقی است[۲۸,۵]. در حملات توانی با در نظر گرفتن یک نتیجه میانی و انتخاب مدل توانی متناسب با آن و مقایسه آن با توان مصرفی اندازه‌گیری شده، می‌توان وقوع یا عدم وقوع یک نتیجه میانی را تعیین نمود[۱۰,۹]. بنابراین برای انجام یک حمله بر روی یک سیستم رمزنگار علاوه بر دسترسی به ورودی و یا خروجی سیستم، نیازمند یک مدل توانی به منظور تخمین میزان مصرف توان در سیستم هستیم. هر چه دانش مهاجم در مورد سیستم تحت حمله بیشتر باشد می‌تواند از مدل‌های کامل‌تری به منظور تخمین میزان توان مصرفی سیستم استفاده نماید. دو مورد از پرکاربردترین مدل‌های توان عبارتند از :

مدل فاصله همینگ[۵]: چنانکه بیان شد در مدارات دیجیتال مبتنی بر تکنولوژی CMOS میزان توان مصرفی متناسب با تعداد انتقال‌های $1 \rightarrow 0$ و $0 \rightarrow 1$ رخ داده در سیستم است. ایده اولیه مدل توانی فاصله همینگ نیز شمارش تعداد این انتقال‌ها درون یک سیستم دیجیتال در یک بازه زمانی مشخص به منظور تخمین میزان توان مصرف شده در آن بازه زمانی است[۹]. با برش زمان شبیه‌سازی یک سیستم به بازه‌های زمانی کوتاه و استخراج فاصله همینگ هر کدام از این بازه‌های زمانی، یک مدل از مصرف توان سیستم مورد نظر به دست می‌آید. برای اعمال این مدل بر روی داده میانی مورد نظر باید اطلاعات مربوط به داده قبل و یا بعد از آن در دسترس باشد. این مدل با مقایسه دو دنباله پشت سر هم از داده‌های میانی و شمارش تعدا انتقال رخ داده بین آنها عددی را بعنوان خروجی مدل توانی ارائه می‌نماید[۹]. برای استفاده از این مدل اطلاعات دقیقی از نحوه پیاده‌سازی سیستم مورد نظر مورد نیاز است.

مدل وزن همینگ[۶]: در این مدل فرض بر این است که میزان مصرف توان در سیستم متناسب با تعداد بیت‌های با مقدار یک در داده میانی است. هر چند که این مدل توصیف دقیقی از مصرف توان در سیستم را ارائه نمی‌دهد ولی به دلیل اینکه این مدل بسیار ساده‌تر از مدل فاصله همینگ است و نیازی به اطلاع از نحوه پیاده‌سازی سیستم ندارد، جهت تخمین میزان مصرف توان در سیستم مقصد بسیار مورد استفاده قرار می‌گیرد[۹].

حملات مبتنی بر تحلیل توان متناسب با میزان دسترسی به سیستم تحت حمله و تعداد نمونه‌های توان قابل اندازه‌گیری دارای انواع مختلفی است که در ادامه به تشریح آنها خواهیم پرداخت.

۲-۱- حملات تحلیل توان ساده

در [۲۸] در تعریف این دسته از حملات چنین بیان شده است : "حملات تحلیل توان ساده روشی است که مستقیماً به تحلیل توان

اندازه‌گیری شده در طول انجام عمل رمزنگاری می‌پردازد". هدف از این حمله استخراج اطلاعات حساس در صورت دسترسی به تعداد محدودی نمونه توان مصرفی اندازه‌گیری شده است. در بیشتر مواقع این بدین معنی است که مهاجم تنها یک نمونه از توان مصرفی اندازه‌گیری شده را در دسترس دارد[۹]. در این روش مهاجم به دنبال استخراج الگوهایی از روی توان مصرفی است که وابسته به عملیاتی باشد که بر روی کلید سیستم انجام گرفته است. عملیات استخراج الگو در این روش بیشتر به صورت دیداری انجام می‌گیرد. به همین دلیل نیل به نتیجه درست در این روش بسیار چالش برانگیز است و مهاجم به منظور افزایش قدرت استخراج اطلاعات از یک‌سری پایگاه داده الگو بهره می‌برد. این روش تحلیل توان ساده مبتنی بر الگو نام دارد[۹].

۲-۲- حملات تحلیل توان تفاضلی

هدف از روش حمله تحلیل توان تفاضلی استخراج داده‌های حساس سیستم بر پایه تعداد زیادی نمونه توان مصرفی اندازه‌گیری شده از سیستم حین انجام عملیات رمزنگاری داده‌های مختلف است[۲۸,۹]. در تمایز با روش قبل، این روش نیازمند تعداد زیادی نمونه‌های اندازه‌گیری شده توان مصرفی است. همچنین در روش قبل ما به دنبال استخراج الگوهای مورد نظر در بازه‌ای از زمان بودیم درحالی‌که در این روش مقادیر لحظه‌ای توان اندازه‌گیری شده مورد بررسی و مقایسه قرار می‌گیرند[۹].

به منظور درک دقیق‌تر، نحوه عمل حملات مبتنی بر تحلیل توان تفاضلی در قالب مثالی بیان می‌گردد[۱۰]. در این مثال ما می‌خواهیم نحوه استخراج بایت اول از کلید رمزنگاری الگوریتم رمزنگاری AES که بر روی میکروکنترلر ۸۰۵۱ پیاده‌سازی شده است را بیان نماییم. توان مصرفی سیستم رمزنگار از طریق یک طیف‌نگار دیجیتالی نمونه‌گیری می‌شود. داده‌های خام از طریق یک واسط RS-232 از کامپیوتر به میکروکنترلر انتقال داده می‌شود و پس از اعمال عملیات رمزنگاری بر روی آن دوباره به کامپیوتر بازگردانده می‌شود. در اولین مرحله از الگوریتم AES اولین بایت از کلید با اولین بایت از داده ورودی در واحد SubBytes ترکیب می‌شود. بنابراین با در نظر گرفتن این داده میانی، به ازای ۱۰۰۰ داده ورودی عملیات نمونه‌گیری از توان مصرفی سیستم انجام می‌گیرد. ما این نمونه‌های اندازه‌گیری شده را درون یک ماتریس بنام T قرار می‌دهیم که هر سطر ماتریس بیانگر یک نمونه اندازه‌گیری شده است. در گام بعد به ازای هر کدام از ورودی‌ها مقادیر داده‌های میانی به ازای مقادیر فرضی کلید محاسبه می‌گردد. یعنی به ازای هر ورودی i عبارت $v_{i,j} = SBox(d_i \oplus k_j)$ به ازای مقادیر مختلف کلید فرضی که از صفر تا ۲۵۵ است، محاسبه می‌گردد. بنابراین ماتریس V که شامل نتایج فوق است اندازه‌ای برابر با ۱۰۰۰*۲۵۶ خواهد داشت.

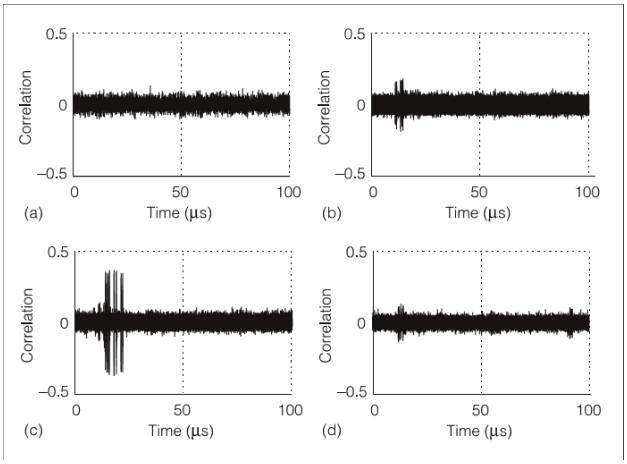
گام چهارم از حملات مبتنی بر تحلیل توان تفاضلی نگاشت مقادیر فرضی به دست آمده ماتریس V به توان مصرفی ناشی از این مقادیر میانی است. بدین منظور ما با انتخاب یک مدل توانی مناسب اقدام به نگاشت این نتایج میانی به داده‌های توان مصرفی متناظر می‌نماییم. به عنوان مثال با انتخاب مدل وزن همینگ به عنوان مدل توانی و با داشتن داده میانی (۰۱۱۰۱۰۰۱) مقدار ۴ به عنوان مقدار متناظر در موقعیت داده میانی مورد نظر، در یک ماتریس جدید با نام H نوشته می‌شود. ماتریس H ماتریس توان مصرفی متناظر با ماتریس V است.

آخرین مرحله از یک حمله مبتنی بر تحلیل توان تفاضلی استخراج داده‌های مربوط به نسبت تشابه دو ماتریس H و T است. به منظور محاسبه داده‌های مورد نظر بین دو بردار از روش تفاضل میانگین[۷] استفاده می‌شود. روش دیگر ضریب همبستگی[۸] نام دارد که رابطه آن در (۱) آمده است. از این رابطه در روش حمله تحلیل توان تطابقی استفاده می‌شود. در واقع تنها تفاوت بین این دو روش حمله در نحوه محاسبه نسبت تشابه بین دو ماتریس H و T است.

$$r_{i,j} = \frac{\sum_{d=1}^{D}(h_{d,i}-\overline{h}_i)*(t_{d,j}-\overline{t}_j)}{\sqrt{\sum_{d=1}^{D}(h_{d,i}-\overline{h}_i)^2 * \sum_{d=1}^{D}(t_{d,j}-\overline{t}_j)^2}} \qquad (۱)$$

مقادیر محاسبه شده از رابطه (۱) درون ماتریس R قرار داده می‌شود. خروجی این مرحله یک $r_{i,j}$ است که به عنوان بزرگ‌ترین عدد ماتریس R شناخته می‌شود. مقدار $r_{i,j}$ بیانگر ضریب همبستگی بین ستون i از ماتریس H با ستون j از ماتریس T است و مقدار i بیانگر مقدار صحیح از کلیدهای فرضی در نظر گرفته شده در مراحل قبل است.

شکل (۱) یک نمونه از خروجی مرحله آخر از این روش است که در آن به ازای کلیدهای فرضی ۱۱۷، ۱۱۸، ۱۱۹ و ۱۲۰ مقادیر ضرایب همبستگی به نمایش درآمده است.



شکل (۱) : مقادیر ضرایب همبستگی به ازای کلیدهای فرضی ۱۱۷،۱۱۸،۱۱۹ و ۱۲۰[۱۰]

۳- مقابله با حملات تحلیل توانی

همانطور که در بخش‌های پیشین تشریح شد، حملات مبتنی بر تحلیل توان با تحلیل توان مصرفی سیستم و توجه به این نکته که میزان مصرف توان در سیستم‌های رمزنگار متناسب با داده‌ی در حال پردازش در سیستم رمزنگاری است، موفق به استخراج اطلاعات حساس سیستم می‌گردند. روشن است که به منظور مقابله با این حملات، باید مصرف توان سیستم مستقل از داده در حال پردازش و یا عملیات در حال انجام باشد[۱۰,۹]. تمامی تحقیقات انجام گرفته در این زمینه بر پایه حذف همبستگی بین مصرف توان در سیستم و داده در حال پردازش در درون هستند[۳,۲]. این روش‌ها را می‌توان به دو دسته کلی اختفا[۹] و پوشش[۱۰] تقسیم‌بندی نمود.

۳-۱- روش‌های مبتنی بر اختفا

در روش‌های مقابله با حملات تحلیل توان مبتنی بر اختفا سعی می‌شود مصرف توان در سیستم رمزنگاری مستقل از داده در حال پردازش باشد. بدین منظور یکی از دو روش زیر به کار گرفته می‌شود.
۱) مصرف یکنواخت توان[۹] : برای یکنواخت کردن مصرف توان سخت‌افزار اضافه‌ای در سیستم در نظر گرفته می‌شود که رفتار توانی آن عکس رفتار توانی سیستم رمزنگاری اصلی است. بدین معنی که هر جا مصرف توان سیستم اصلی بالاست این سخت‌افزار توان کمی مصرف می‌کند و بالعکس. ۲) مصرف تصادفی توان : مصرف توان در سیستم رمزنگاری حالت تصادفی به خود می‌گیرد تا نتوان داده در حال پردازش را از تحلیل توان مصرفی استخراج کرد[۹].

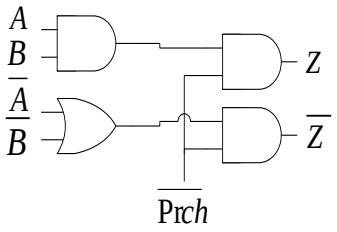
هر دو روش یکنواخت‌سازی و تصادفی کردن مصرف توان را می‌توان در سطوح مختلف پیاده‌سازی یک سیستم رمزنگار به کار برد که در ادامه یک دسته‌بندی از روش‌های مبتنی بر اختفا ارائه می‌شود.

- سطح معماری
  - بعد زمانی
    - عملیات زائد و جابجایی دستورات[۹]، درج تصادفی سیکل‌های زائد[۹]، تغییر تصادفی فرکانس مدار[۱۸]
  - بعد دامنه
    - استفاده از منابع تولید نویز به منظور کاهش پارامتر SNR[۲۶]، ایجاد افزونه برای واحدهای عملیاتی[۲۵]، فیلتر جریان منبع تغذیه با استفاده از تجهیزات فعال[۲۲]
- سطح دروازه
  - منطق پیش‌شارژ دوخطی
    - منطق تفاضلی پویای موج[۱۱] (WDDL)
- سطح ترانزیستور
  - منطق پیش‌شارژ دوخطی
    - منطق مبتنی بر تقویت‌کننده‌های حسی[۱۲] (SABL)
    - منطق پیش‌شارژ تک‌خطی سه‌فاز[۱۳] (TSPL)
    - منطق پیش‌شارژ دوخطی سه‌فاز[۱۴] (TDPL)
    - منطق بی‌دررو[۱۵]
    - ساختار مبتنی بر ترانزیستورهای گذر[۱۶] (SDMLp)

در ادامه این بخش برخی از روش‌های فوق که بیشتر توسط محققیق مورد توجه قرار گرفته‌اند و کارایی بیشتری نیز دارند به طور مبسوط مورد بحث قرار می‌گیرند.

۳-۱-۱- منطق تفاضلی پویای موج

این منطق در سطح دروازه پیاده‌سازی شده است[۲۰] و عناصر پایه آن همان عناصر استاندارد منطق CMOS هستند. در این منطق هر خروجی به صورت عادی و مکمل‌شده ارائه می‌شود. مدارات مبتنی بر این منطق دارای دو فاز پیش‌شارژ و ارزیابی هستند که در فاز پیش‌شارژ هر دو خروجی مکمل و عادی مقدار صفر به خود می‌گیرند و در فاز ارزیابی بسته به مقدار ورودی‌ها، یکی از خروجی‌ها مقدار یک منطقی را به خود می‌گیرد. به این ترتیب در این منطق در هر سیکل مدار دقیقا یک انتقال از صفر به یک رخ می‌دهد و مصرف توان در این منطق تقریباً یکنواخت است[۹]. شکل (۲) نمایانگر دروازه AND در این منطق است.



شکل (۲) : یک دروازه AND در منطق WDDL[۲۰]

در مقایسه با منطق SCMOS، مساحت مورد نیاز در این منطق بیش از دو برابر بوده و از دید سرعت نیز میزان تأخیر در مدارات مبتنی بر منطق تفاضلی پویای موج بیشتر از منطق استاندارد است. علاوه بر موارد بالا به منظور ایجاد عدم وابستگی بین مصرف توان و داده ورودی در این منطق لازم است تا خطوط مکمل متوازن گردند[۱۲]. مورد دیگر مفهوم انتشار اولیه[۱۷] است که بدلیل عدم همزمانی در ورود داده‌های ورودی بوجود می‌آید و مداراتی که فاقد مکانیزمی جهت اعمال همزمان ورودی‌ها به سیستم می‌باشند در مقابل حملات توانی آسیب‌پذیر خواهند بود[۱۴].

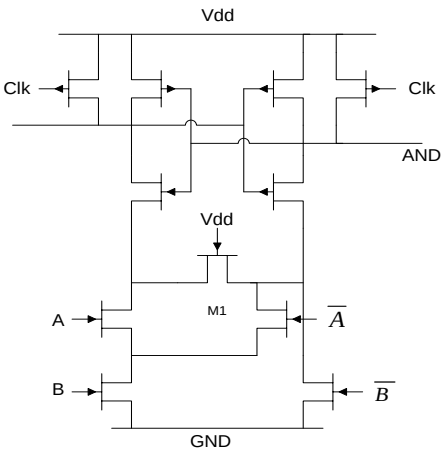
جدول (۱) مقایسه‌ای را بین WDDL و منطق استاندارد از دید میزان فضای مورد نیاز و تاخیر مدار انجام داده است.

جدول (۱) : مقایسه بین WDDL و منطق استاندارد[۲۰]

	Area(Kgates)		Delay(ns)	
	SCMOS	WDDL	SCMOS	WDDL
Kasumi	7304	26549	33.3	47.3
DES	1493	4810	7.9	8.4
AES	13241	44827	13.6	14.6

۳-۱-۲- منطق مبتنی بر تقویت‌کننده‌های حسی

یک دروازه مبتنی بر منطق SABL بر پایه عنصر حافظه StrongArm110 ساخته شده است. این منطق که در [۲۳] معرفی گردیده منطقی پویا است که از مفهوم خروجی دوخطی در آن استفاده شده است. شکل (۳) یک دروازه AND/NAND مبتنی بر منطق SABL را به نمایش گزارده است[۲۳,۲۷].



شکل (۳) : یک دروازه AND/NAND در منطق SABL[۲۳]

این منطق دارای دو فاز است. سطح منطقی صفر سیگنال ساعت مدار به عنوان فاز پیش‌شارژ و سطح یک منطقی آن به عنوان فاز ارزیابی عمل می‌نماید. دو معکوس کننده متصل به خروجی‌ها وظیفه تنظیم سطح سیگنال خروجی را بر عهده دارند و ترانزیستور M1 هم که همیشه روشن است به منظور تخلیه بارهای پارازیت تعبیه گردیده است. در این منطق در فاز پیش‌شارژ، خروجی‌های دوگان تا سطح منطقی یک شارژ می گردند و در مرحله ارزیابی بسته به ورودی‌های اعمال شده یکی از خروجی‌ها صفر می گردد.

این منطق در مقایسه با منطق استاندارد دارای میزان تغییرات توان مصرفی بسیار پایینی است و در مقابل میزان انرژی مصرفی و همچنین فضای مورد نیاز آن در مقایسه با منطق استاندارد بسیار زیاد است. جدول (۲) مقایسه‌ای بین این منطق و منطق استاندارد از دید میزان تغییرات توان مصرفی، میزان مصرف انرژی در هر سیکل و فضای مورد نیاز را نشان داده است.

جدول (۲) : مقایسه بین منطق SABL و منطق استاندارد[۲۳]

	INV	NAND	XOR	FF	S9-box			
	NED	NED	NED	NED	NED	NSD	E/Cycle	Area
SCMOS	1.000	1.000	1.000	0.821	1.000	0.293	5.92	21449
SABL	0.009	0.032	0.016	0.020	0.032	0.006	11.32	38541

در این جدول پارامتر NED عبارتست از میزان تغییرات انرژی نرمال که از رابطه زیر به دست می‌آید :

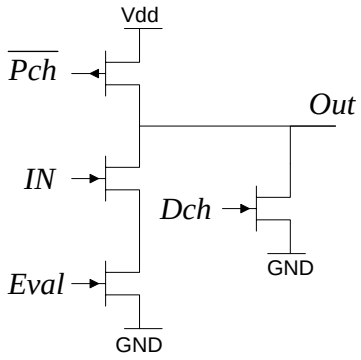
$$NED = \frac{Max\left(\frac{Energy}{Cycle}\right) - Min\left(\frac{Energy}{Cycle}\right)}{Max\left(\frac{Energy}{Cycle}\right)} \qquad (۲)$$

همچنین پارامتر NSD عبارت است از انحراف معیار نرمال میانگین مصرف توان.

۳-۱-۳- منطق پیش‌شارژ تک‌خطی سه‌فاز

منطق‌هایی با خروجی دوگان هم میزان مصرف توان را افزایش داده و هم باعث بزرگ‌تر شدن سطح تراشه می‌گردند. از طرف دیگر متوازن نمودن خروجی‌های مکمل جهت جلوگیری از نشت اطلاعات کار بسیار مشکلی است. بنابراین ارائه یک منطق تک‌خطی که در مقابل حملات توانی مقاوم باشد، می‌تواند این مشکلات را برطرف نماید[۸].

این منطق یک منطق پویا با خروجی سه‌فاز است که در آن در هر سیکل یک عملیات شارژ و یک عملیات تخلیه صورت می‌پذیرد. بنابراین مصرف توان مستقل از داده ورودی خواهد بود[۸]. شکل (۴) تصویر یک معکوس‌کننده در منطق TSPL است.



شکل (۴) : یک معکوس‌کننده در منطق TSPL[۸]

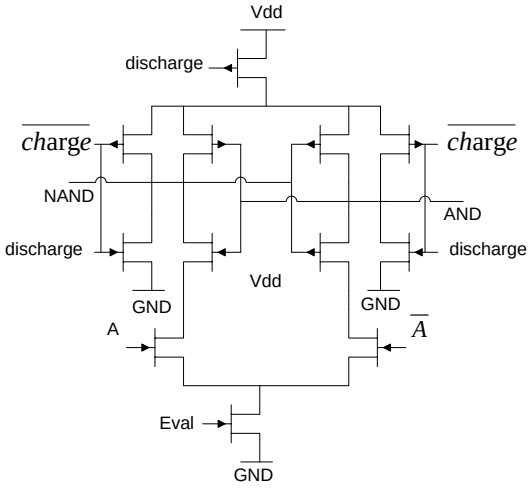
از نقاط ضعف این منطق این می‌باشد که در صورتی که حمله‌کننده بتواند فازهای مختلف را از هم تشخیص دهد قادر به انجام حمله خواهد بود[۸]. جدول (۳) مقایسه این منطق با منطق‌های دیگر را به نمایش گزارده است. با توجه به داده‌های جدول میزان افزونه توان مصرفی و فضای مورد نیاز در این منطق در مقایسه با منطق‌های معرفی شده قبلی بسیار پایین است. همچنین پارامتر NED در مقایسه با منطق‌های دیگر کاهش چشم‌گیری داشته است.

جدول (۳) : مقایسه بین منطق TSPL با منطق‌های دیگر[۸]

Logic Style	Static	TDPL	WDDL	TSPL
Delay(ps)	47	66	74	39
Active Area(λ^2)	64	300	160+320	96
Layout Area(λ^2)	1824	6840	3840+7680	2376
Max. Energy(fJ)	6.79	19.2	25.0	7.70
NED	100%	5.97%	16.6%	0.037%
Logical Effort	1.2	0.57	1.2	0.49

۳-۱-۴- منطق پیش‌شارژ دوخطی سه‌فاز

این منطق که در [۱۱] ارائه گردیده است در واقع بهبود یافته منطق SABL است که برخلاف آن نسبت به عدم توازن بار خروجی حساس نیست. در این منطق نیز به ازای هر سیکل یک مرحله شارژ و یک مرحله تخلیه در هر کدام از خروجی‌ها مشاهده می‌گردد. بنابراین میزان مصرف توان در این منطق مستقل از داده ورودی تقریبا ثابت است. شکل زیر یک معکوس‌کننده در منطق TDPL است که در مقایسه با مدار مشابه در منطق SABL یک ترانزیستور PMOS و دو ترانزیستور NMOS به مدار اضافه گردیده است.



شکل (۵) : یک معکوس‌کننده در منطق TDPL[۱۱]

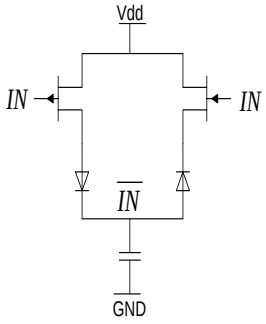
با توجه به اینکه پارامتر NED یک پارامتر مهم در مقایسه میزان تغییرات توان مصرفی مدار است، با توجه به اندازه‌گیری‌های انجام گرفته (جدول ۴) این منطق دارای میزان تغییرات توان مصرفی بسیار پایین‌تری نسبت به منطق SABL است.

جدول (۴) : مقایسه بین منطق TDPL و SABL[۱۱]

	INV		NAND/AND		XOR/XNOR	
	SABL	TDPL	SABL	TDPL	SABL	TDPL
Max(E)[fJ]	52.3	65.6	56.3	68.3	58.4	69.5
Min(E)[fJ]	31.1	65.3	35.2	66.4	39.4	68.0
NED	40.4%	0.4%	37.5%	2.7%	32.6%	2.1%
$\overline{E}[fJ]$	41.7	65.5	50.5	67.3	48.9	68.7
$\sigma_E[fJ]$	10.9	0.1	8.0	0.6	8.5	0.4
NSD	26.1%	0.2%	15.9%	0.9%	17.4%	0.6%

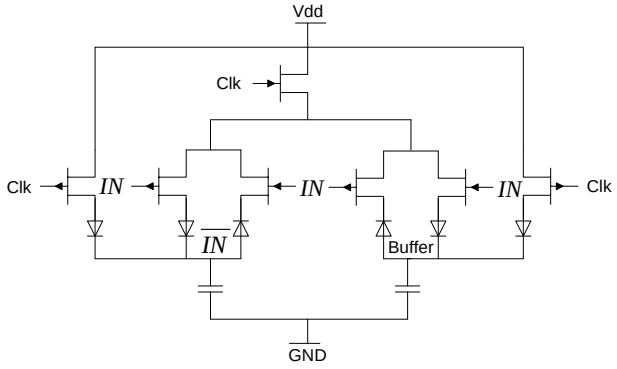
۳-۱-۵- منطق مبتنی بر فرایندهای بی‌دررو

به منظور طراحی مدارهایی با توان مصرفی پایین روشهای متفاوتی پیشنهاد گردیده است. یکی از این روشها استفاده از منبع تغذیه متناوب است که در آن انرژی مصرف شده دوباره بازیافت می‌گردد[۶]. این روش اصطلاحاً ساختار مبتنی بر فرایندهای بی‌دررو نامیده می‌شود[۷]. شکل (۶) تصویر یک معکوس کننده در این ساختار است.



شکل (۶) : یک معکوس کننده در ساختار بی‌دررو[۷]

در صورتی که بتوان ساختار دوخطی را به این منطق اضافه نمود خواهیم توانست مصرف توان را به صورت یکنواخت دربیاوریم. با افزودن دو فاز پیش‌شارژ و ارزیابی به ساختار مورد نظر نیز خواهیم توانست تعداد انتقالات سیگنال خروجی را از ورودی مستقل نماییم. شکل زیر تصویر یک معکوس‌کننده مبتنی بر مفهوم بی‌دررو با ساختار پیش‌شارژ دوخطی می‌باشد.



شکل (۷) : یک معکوس‌کننده مبتنی بر ساختار پیش‌شارژ دوخطی در منطق بی‌دررو[۷]

چنانکه از جدول (۵) نیز مشخص است، میزان مصرف توان در ساختار معرفی شده از منطق استاندارد نیز پایین تر است هر چند که میزان فضای مورد نیاز بسیار بیشتر از منطق استاندارد است.

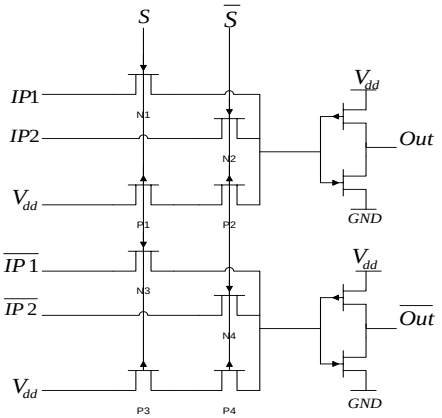
جدول (۵) : میزان مصرف توان منطق معرفی شده در مقایسه با منطق استاندارد و TDPL[۷]

Logic Style	Av. Power(watts)
Purely adiabatic	6.9068E-10 W
CMOS	1.1243E-05 W
TDPL	0.9558E-06 W
Current Style	4.0813E-09 W
Proposed Logic Style	1.6704E-09 W

۳-۱-۶- منطق مبتنی بر ترانزیستورهای گذر

سعی در طراحی مدارات مقاوم در برابر حملات تـوانی بـدلیل اینکـه میزان توان مصرفی و فضای مورد نیاز برای پیاده‌سازی آنهـا بهینـه نیست همچنان ادامه خواهد داشت. منطق مبتنی بر ترانزیستورهای گذر سعی نموده است تا اولا با استفاده از مفاهیم پیش‌شارژ و خروجی دوگان مداری مقاوم در برابر حملات توانی ارائه دهد ثانیا با استفاده از ساختار مدارات مبتنی بر ترانزیستورهای گذر میزان توان مصرفی و فضای مورد نیاز را کمینه نماید[۴].

شکل (۸) عنصر پایه در این منطق است کـه در آن بـا اعمـال ورودی‌های مختلف طبق جدول (۶) می‌توان خروجی‌های مورد نظر را تولید نمود.



شکل (۸) : ساختار عنصر پایه در منطق مبتنی بر ترانزیستورهای گذر(SDMLp)[۴]

جدول (۶) : تولید توابع مختلف توسط عنصر پایه[۴]

$IP1$	$IP2$	S	Out	$\overline{Out}$
$\overline{A}$	$\overline{B}$	B	$A.B$	$\overline{A.B}$
$\overline{B}$	$\overline{A}$	B	$A+B$	$\overline{A+B}$
A	$\overline{A}$	B	$A.\overline{B}+\overline{A}.B$	$\overline{A.\overline{B}+\overline{A}.B}$
$\overline{B}$	$\overline{A}$	S	$\overline{S}.A+S.B$	$\overline{\overline{S}.A+S.B}$
A	$\overline{A}$	B	$\bar{A}+B$	$\overline{\overline{A}+B}$
$\overline{A}$	A	B	$A+B$	$\overline{A+B}$
A	$\overline{A}$	A	A	$\overline{A}$

با توجه به شکل (۸) اگر S و $\overline{S}$ هر دو مقدار صفر بگیرند بیانگر فاز پیش‌شارژ در این منطق است و هر دو خروجی مقدار صفر به خود خواهند گرفت. در فاز ارزیابی نیز بسته به مقدار سیگنال‌های S، IP1 و IP2 و مکمل‌های آنها مقدار خروجی مشخص خواهد شد.

جدول (۷) : مقایسه میزان توان مصرفی و فضای مورد نیاز بین منطق SDMLp ، WDDL و منطق استاندارد[۴]

DES Full Chip	SCMOS	SDMLp	WDDL
Area(λ^2)	13247.21	18715.26	32714.64
Total Power(mW)	1.37	1.47	2.96
Max. Op. Freq.(MHz)	100	66.67	83.33
Inst. Current Var.($10^{-6}Amps^2$)	1892.3	5.95	229.1

در مورد پارامترهای ارزیابی کارایی مدارات مقاوم در برابر حملات، متاسفانه اظهار نظری در مورد پارامتر NED بیان نشده است ولی در مورد پارامترهایی چون NSD، میزان توان مصرفی و میزان فضای مورد نیاز با توجه به جدول (۸) بهبود خوبی حاصل گردیده است.

جدول (۸) : مقایسه میزان انحراف معیار استاندارد توان مصرفی بین منطق SDMLp، WDDL و منطق استاندارد[۴]

Gates(2x1)	SCMOS	SDMLp	WDDL
AND	5.29	7.41	13.95
OR	5.62	7.43	14.47
XOR	6.69	7.02	22.12
MUX	7.02	6.93	21.98
Avg.	6.15	7.20	18.13
Std. Dev.	0.83	0.26	4.53

۳-۲- روش‌های مبتنی بر پوشش

ایده اصلی در این دسته از روش‌های مقابله با حملات تحلیل توانی این است که با تصادفی‌سازی مصرف توان در سیستم‌های رمزنگار، رابطه بین مصرف توان و داده در حال پردازش قابل ردیابی نباشد که این کار با ایجاد عدم قطعیت در مقدار داده میانی تولید شده به دست می‌آید. بنابراین در روش‌های مبتنی بر پوشش، داده میانی تولیـد شـده بـه واسطه یک داده تصادفی پوشش داده می‌شود[۹].

پوشش داده میانی توسط یک عملگر ریاضی صورت می‌گیرد. این عملگر ممکن است جمع، ضرب و یا عمل جمـع انحصـاری[۱۸] باشـد. عملگر انتخاب شده باید بگونه‌ای باشد تا از روی داده پوشش داده‌شده نتوان به داده اصلی دست یافت. این حالت را اصطلاحاً استقلال بـین داده پوشش داده‌شده و داده اصلی می‌نامیم[۱۲]. بنابراین اگر v داده اصلی و m عدد تصادفی باشد آنگاه v_m داده پوشش داده شده است و رابطه زیر بین آنها برقرار است :

$$v_m = v * m \qquad (۳)$$

روشهایی که تاکنون از این الگو جهت مقاوم‌سازی مدارات رمزنگار استفاده نموده‌اند را می‌توان بصورت زیر دسته‌بندی نمود.

- سطح دروازه
  - پوشش تصادفی[۱۹]
  - پیش‌شارژ تصادفی[۲۰]
- سطح ترانزیستور
  - منطق پیش‌شارژ دوخطی پوشش داده شده[۲۱] (MDPL)
  - منطق راه‌گزینی تصادفی دوخطی[۲۲] (DRSL)

۳-۲-۱- پوشش تصادفی

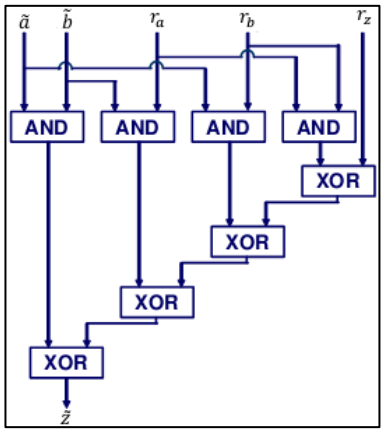
در رویکرد مبتنی بر پوشش داده میانی جهت مقابله با حملات توانی بیان شده است که اگر هم سیگنال‌های ورودی به دروازه و هم سیگنال‌های خروجی از دروازه پوشش داده شود، در آنصورت دروازه مورد نظر ضمن انجام عمل منطقی مورد نظر خواهد توانست در مقابل حملات توانی نیز مقاوم عمل نماید. بنابراین در هر دروازه منطقی به ازای تعداد ورودی‌ها به علاوه یک خروجی سیگنال پوشش وجود دارد که باعث افزایش فضای مورد نیاز و در نتیجه کاهش کارایی سیستم خواهد شد[۱۵].

یک روش پوشش دروازه منطقی در [۱۹] چنین بیان شده است : سیگنال $\tilde{a}$ بیانگر سیگنال پوشش داده شده a است. بنابراین داریم : $\tilde{a} = a \oplus r_a$. برای سیگنال $\tilde{b}$ نیز داریم : $\tilde{b} = b \oplus r_b$. برای یک دروازه منطقی AND خروجی عبارتست از سیگنال $\tilde{z}$ که پوشش یافته سیگنال z است. بین سیگنال $\tilde{z}$ و ورودی‌های پوشش یافته رابطه زیر برقرار است : $\tilde{z} = (\tilde{a} \wedge \tilde{b}) \oplus r_z$.

با گسترش ورودی‌ها در رابطه بالا رابطه (۴) به دست خواهد آمد :

$$\tilde{z} = z \oplus r_z = (\tilde{a} \wedge \tilde{b}) \oplus r_z = (\big((r_z \oplus (r_a \wedge r_b)) \oplus (r_a \wedge \tilde{b})\big) \oplus (r_b \wedge \tilde{a})) \oplus (\tilde{a} \wedge \tilde{b}) \qquad (۴)$$

شکل (۹) پیاده‌سازی رابطه (۴) را نشان می‌دهد. برای دروازه OR نیز رابطه مشابهی بیان گردیده است[۱۹].



شکل (۹) : یک دروازه AND پوشش داده شده[۱۹]

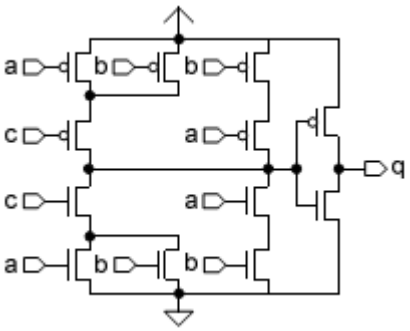
علاوه بر این روش ، تکنیک‌های دیگری چون روش پوشش مالتی‌پلکسری[۱۹] و نیز پوشش مدارات ضرب کننده[۱۵] در سطح دروازه‌های منطقی به منظور طراحی مدارات مقاوم در برابر حملات تحلیل توان پیشنهاد گردیده است.

در زمینه ارزیابی پارامترهای کارایی این روش مانند تصادفی‌سازی توان مصرفی، میزان توان مصرفی و میزان فضای مورد نیاز اطلاعاتی ارائه نگردیده است. ولی آنچه که مسلم است در زمینه میزان توان مصرفی و همچنین فضای مورد نیاز، مقدار بالاسری[۲۳] در مقایسه با منطق استاندارد بسیار زیاد است.

۳-۲-۲- منطق پیش‌شارژ دوخطی پوشش داده شده

استفاده از تکنیک پوشش داده میانی درکنار مفاهیم پیش‌شارژ دوخطی باعث شده است تا این منطق بتواند ضمن بهره از مزایای این روشها از معایب هر کدام از آنها نیز اجتناب نماید[۱۲,۵].

در روشهای قبل بیان شد که در تکنیک دوخطی، عدم توفیق در متوازن نمودن خطوط دوگان جزء نقاط ضعف روش مورد نظر به حساب می‌آید. ولی در منطق MDPL پوشش داده‌های میانی باعث شده است تا وابستگی توان مصرفی به داده‌های مورد پردازش حذف گردیده و نیاز به متوازن نمودن خطوط دوگان برطرف گردد. از طرفی بدلیل استفاده از مفهوم پیش‌شارژ، مدار تحت تاثیر تغییرات لحظه‌ای[۲۴] سیگنال قرار نخواهد گرفت[۱۶].

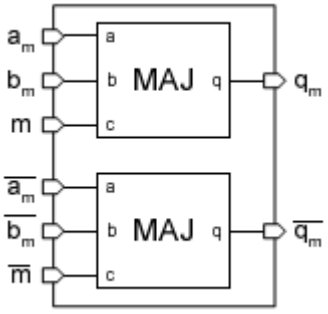


شکل (۱۰) : تابع اکثریت(MAJ) مورد استفاده در دروازه AND منطق MDPL[۱۶]

در این منطق جهت پیاده‌سازی دروازه AND از رابطه زیر استفاده می‌گردد.

$$q = ((a_m \oplus m) \wedge (b_m \oplus m) \oplus m \qquad (۵)$$

شکل (۱۰) پیاده‌سازی در سطح ترانزیستور این رابطه را نشان می‌دهد. از قرار دادن یک جفت از این تابع مطابق شکل (۱۱) دروازه AND دو ورودی در منطق MDPL ساخته می‌شود.



شکل (۱۱) : یک دروازه AND/NAND در منطق MDPL[۱۶]

در این منطق در فاز پیش‌شارژ هر دو خروجی دوگان مقدار صفر منطقی می‌گیرند. هر چند که بیان نشده است که سیگنال پیش‌شارژ چگونه و توسط کدام واحد تولید می‌گردد.

در جدول (۹) منطق مورد نظر با منطق استاندارد از دید فضای مورد نیاز، سرعت و میزان توان مصرفی مورد مقایسه قرار گرفته است.

جدول (۹) : مقایسه بین منطق MDPL و منطق استاندارد[۱۶]

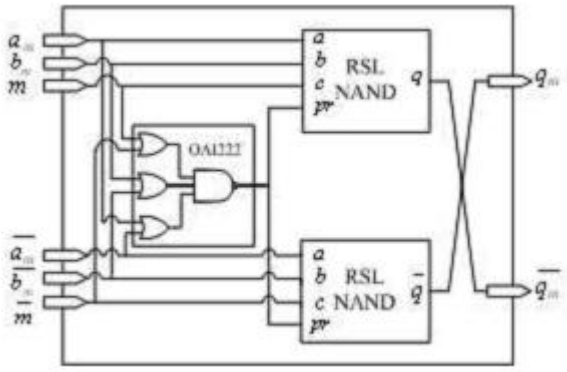
	Area(GE)	Speed(MHz; worst-case corner)	Avg. Power(uA*100kHz)
CMOS	3452	17.02	7.02
MDPL	16421	10.04	122.34
Ratio $\frac{MDPL}{CMOS}$	4.76	0.59	17.43

۳-۲-۳- منطق راه‌گزینی تصادفی دوخطی

در [۲۱] بیان شده است که در روش‌های مبتنی بر پوشش داده‌های میانی که در آن از یک بیت داده تصادفی استفاده می‌شود امکان نشت اطلاعات در برابر حملات توانی وجود دارد. بدین منظور منطقی مبتنی بر روش پیشنهاد شده در [۱۳] جهت غلبه بر این نقیصه پیشنهاد گردیده است.

در روش‌های مبتنی بر پیش‌شارژ دوخطی در صورتی که ورودی‌ها به صورت همزمان اعمال نگردند حتی در شرایطی که هیچ پرش ناخواسته‌ای در سیگنالهای ورودی مشاهده نگردد نشتی اطلاعات خواهیم داشت[۱۴]. در منطق راه‌گزینی تصادفی دوخطی ، به منظور غلبه بر پرش‌های ناخواسته سیگنالهای ورودی، از پیش‌شارژ و برای غلبه بر عدم توازن در خطوط دوگان از پوشش تصادفی و نیز برای همگام‌سازی سیگنالهای ورودی از واحد تولید سیگنال پیش‌شارژ درونی استفاده شده است[۱۳]. در این ساختار استفاده از مفهوم خطوط دوگان به منظور سهولت در پیاده‌سازی واحد تولید سیگنال پیش‌شارژ انجام گرفته است.

تصویر یک دروازه AND در منطق DRSL در شکل (۱۲) به نمایش درآمده است.



شکل (۱۲) : یک دروازه AND در منطق DRSL[۱۳]

این منطق دارای دو فاز عملیاتی است. در فاز پیش‌شارژ تمامی سیگنال‌ها اعم از سیگنال‌های تولید مقادیر تصادفی دارای سطح صفر منطقی خواهند بود و در فاز ارزیابی مقدار سیگنال پیش‌شارژ صفر گردیده و خروجی مدار به ازای ورودیهای آن مشخص می‌گردد.

جدول (۱۰) مقایسه بین این منطق و منطق استاندارد است که در آن از دید میزان فضای مورد نیاز مورد بررسی قرار گرفته‌اند.

جدول (۱۰) : مقایسه بین منطق DRSL و منطق استاندارد[۱۳]

DRSL Cell	Implementation	Area(gate equivalents)	
		DRSL	Standard
Inverter	Wire swapping	0	0.67
Buffer	2*Buffer	2.66	1.33
AND, OR(2-in)	2*RSL NAND, OAI	7.21	1.33
NAND, NOR(2-in)	2*RSL NAND, OAI	7.21	1
XOR, XNOR	2*RSL XOR, OAI	8.22	2.67

۴- طرح پیشنهادی

با بررسی مطالعات و تحقیقات انجام گرفته در زمینه روش‌های مختلف مقابله با حملات تحلیل توانی، تاکنون بررسی جامعی در زمینه ارزیابی انواع روش‌های مقابله با این نوع از حملات صورت نگرفته است. با توجه به داده‌های ارائه شده در مقالات علمی، روش‌های پیشنهاد شده عموماً جهت بررسی کارایی‌شان صرفاً به مقایسه آن با تعدادی از روش‌های پیشین و آن هم با در نظر گرفتن بخشی از پارامترهای کارایی اقدام نموده‌اند. وجود یک ارزیابی جامع در زمینه انواع روش‌های مقابله با حملات تحلیل توان باعث خواهد شد تا طراح یک سیستم با کارکرد امنیتی، دقیق‌تر و کاراتر بتواند با توجه به خواسته‌ها و نیازهای مورد نظرش روش مناسب جهت مقاوم سازی سیستم را انتخاب نماید.

در این تحقیق ما قصد داریم روش‌های مختلف پیشنهاد شده را مورد بررسی قرار داده و به ارزیابی هر کدام از این روش‌ها از دید میزان مقاومت در برابر حملات تحلیل توان بپردازیم. پارامترهای مورد نظر در این ارزیابی عبارتند از : میزان توان مصرفی، میزان فضای مورد نیاز جهت پیاده‌سازی و میزان انحراف معیار توان مصرفی.

۵- جمع‌بندی

هدف از این گزارش این بود تا ضمن بیان انواع مختلف حملات مبتنی بر تحلیل توان، روش‌های موجود جهت مقابله با این حملات مورد بررسی قرار گیرند. بنابراین با بیان انواع روش‌های حمله، نحوه عمل هر کدام بصورت خلاصه بیان گردید. سپس روش‌های مختلف مقابله با این نوع از حملات با توجه به نحوه عمل و سطوح انتزاع پیاده‌سازی، دسته‌بندی گردیده و در مورد سطوح انتزاع دروازه و ترانزیستور به صورت مبسوط صحبت گردید و هر کدام از روش‌ها از دید میزان یکنواخت‌سازی و یا تصادفی‌سازی مصرف توان، میزان توان مصرفی و میزان فضای مورد نیاز جهت پیاده‌سازی با توجه به نتایج ارائه گردیده با منطق استاندارد مورد مقایسه قرار گرفت.

مراجع

[1] S. Skorobogatov, "Physical Attacks and Tamper Resistance," in Introduction to Hardware-Oriented Security and Trust (HOST), M. Tehranipoor and C. Wang, Eds. USA: Springer, 2012, ch. 7, pp. 143-174.

[2] K. Mai, "Side Channel Attacks and Countermeasures," in Introduction to Hardware-Oriented Security and Trust (HOST), M. Tehranipoor and C. Wang, Eds. USA: Springer, 2012, ch. 8, pp. 175-194.

[3] P. Schaumont; Z. Chen, "Side-Channel Attacks and Countermeasures for Embedded Microcontrollers," in Introduction to Hardware-Oriented Security and Trust (HOST), M. Tehranipoor and C. Wang, Eds. USA: Springer, 2012, ch. 11, pp. 263-282.

[4] Ramakrishnan, L.N.; Chakkaravarthy, M.; Manchanda, A.S.; Borowczak, M.; Vemuri, R., "SDMLp: On the Use of Complementary Pass Transistor Logic for Design of DPA Resistant Circuits, "IEEE International Symposium on Hardware-Oriented Security and Trust (HOST), pp. 31-36, 2012.

[5] Moradi, A.; Kirschbaum, M.; Eisenbarth, T.; Paar, C., "Masked Dual-Rail Precharge Logic Encounters State-of-the-Art Power Analysis Methods", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 20, pp. 1578-1589, 2012.

[6] Sana, P.K.; Satyam, M., "A Low Power Secure Logic Style to Counteract Differential Power Analysis Attacks," International Symposium on VLSI Design, Automation and Test (VLSI-DAT), pp. 1-4, 2011.

[7] Sana, P.K.; Satyam, M., "An Energy Efficient Secure Logic to Provide Resistance against Differential Power Analysis Attacks," International Symposium on Electronic System Design (ISED), pp. 61-65, 2010.

[8] Menendez, E.; Mai, K., "Extended Abstract: A High-Performance, Low-Overhead, Power-Analysis-Resistant, Single-Rail Logic Style," Hardware-Oriented Security and Trust (HOST), pp. 33-36, Jun. 2008.

[9] Popp, T.; Oswald, E.; Mangard, S., "Power Analysis Attacks: Revealing the Secrets of Smart Cards," 1st ed. GRAZ, AUSTRIA: Springer, 2007.

[10] Popp, T.; Oswald, E.; Mangard, S., "Power Analysis Attacks and Countermeasures," Design & Test of Computers, Vol. 24, pp. 535-543, 2007.

[11] M. Bucci; L. Giancane; R. Luzzi; A. Trifiletti, "Three-Phase Dual-Rail Pre-charge Logic," Cryptographic Hardware and Embedded Systems (CHES), pp. 232-241, 2006.

[12] T. Popp; S. Mangard, "Implementation aspects of the DPA-resistant logic style MDPL," International Symposium of Circuits and Systems (ISCAS), pp. 2913-2916, 2006.

[13] Zhimin Chen; Y. Zhou, "Dual-Rail Random Switching Logic: A Countermeasure to Reduce Side Channel Leakage," Cryptographic Hardware and Embedded Systems (CHES), pp. 242-254, 2006.

[14] Kulikowski, K.J.; Karpovsky, M.G.; Taubin, A., "Power attacks on secure hardware based on early propagation of data," 12th IEEE International On-Line Testing Symposium, 2006.

[15] Elena Trichina; Tymur Korkishko; Kyung Hee Lee, "Small Size, Low Power, Side Channel-Immune AES Coprocessor: Design and Synthesis Results," Advanced Encryption Standard (AES), Vol. 3373, pp. 113-127, 2005.

[16] W. Fischer; Berndt M. Gammel, "Masking at Gate Level in the Presence of Glitches," Cryptographic Hardware and Embedded Systems (CHES), pp. 187-200, 2005.

[17] T. Popp and S. Mangard, "Masked Dual-Rail Pre-charge Logic: DPA-Resistance Without Routing Constraints," CHES, pp. 172-186, 2005.

[18] Shengqi Yang; Wolf, W.; Vijaykrishnan, N.; Serpanos, D.N.; Yuan Xie, "Power attack resistant cryptosystem design: a dynamic voltage and frequency switching approach," Design, Automation and Test in Europe (DATE), Vol. 3, pp. 64-69, 2005.

[19] Golic, J.D.; Menicocci, R., "Universal masking on logic gate level," IEEE Electronics Letters, Vol. 40, pp. 526-528, 2004.

[20] Tiri, K.; Verbauwhede, I., "A logic level design methodology for a secure DPA resistant ASIC or FPGA implementation," Design, Automation and Test in Europe (DATE), Vol. 1, pp. 246-251, 2004.

[21] D. Suzuki; M. Saeki; Tetsuya Ichikawa, "Random Switching Logic: A Countermeasure against DPA based on Transition Probability," IACR Cryptology ePrint Archive, Report 2004/346, 2004.

[22] Ratanpal, G.B.; Williams, R.D.; Blalock, T.N., "An on-chip signal suppression countermeasure to power analysis attacks," IEEE Transaction on Dependable and Secure Computing, Vol. 1, pp. 179-189, 2004.

[23] Tiri, K.; Verbauwhede, I., "Charge recycling sense amplifier based logic: securing low power security ICs against DPA," Solid-State Circuits Conference, pp. 179-182, 2004.

[24] Trichina E., "Combinational logic design for AES SubByte transformation on masked data," IACR Cryptology ePrint Archive, Report 2003/236, 2003.

[25] H. Saputra; N. Vijaykrishnan; M. Kandemir; M. J. Irwin; R. Brooks; S. Kim; W. Zhang, "Masking the energy behavior of DES encryption," Design, Automation and Test in Europe (DATE), pp. 84-89, 2003.

[26] L. Benini; E. Omerbegovic; A. Macii; M. Poncino; E. Macii; F. Pro, "Energy-aware design techniques for differential power analysis protection," Design, Automation and Test in Europe (DATE), pp. 36-41, 2003.

[27] Tiri, K.; Akmal, M.; Verbauwhede, I., "A dynamic and differential CMOS logic with signal independent power consumption to withstand differential power analysis on smart cards," Solid-State Circuits Conference, pp. 403-406, 2002.

[28] Paul Kocher; J. Jaffe; B. Jun, "Differential Power Analysis," Advances in Cryptology (CRYPTO), vol. 1666/1999, 1999.

زیرنویس‌ها

[1] Simple Power Analysis
[2] Differential Power Analysis
[3] Correlation Power Analysis
[4] Dual-Rail Pre-Charge Logic
[5] Hamming Distance
[6] Hamming Weight
[7] Difference of Means
[8] Correlational Coefficient
[9] Hiding
[10] Masking
[11] Wave Dynamic Differential Logic

[12]Sense Amplifier based Logic
[13]Three-Phase Single-Rail Pre-Charge Logic
[14]Three-Phase Dual-Rail Pre-Charge Logic
[15]Adiabatic Logic Style
[16]Secure Differential Multiplexer Logic using Pass Transistors
[17]Early Propagation
[18]Exclusive-OR
[19]Random Masking
[20]Random Pre-Charging
[21]Masked Dual-Rail Pre-Charge Logic
[22]Dual-Rail Random Switching Logic
[23]Overhead
[24]Glitch

روش‌های کشف و مقابله با تروجان‌های سخت‌افزاری

مهدی کیخا[1]، مهدی فاضلی[2]، احمد اکبری[3]

[1] دانشجوی کارشناسی‌ارشد، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران
Kaykha@comp.iust.ac.ir
[2] استادیار گروه سخت‌افزار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران
M_fazeli@iust.ac.ir
[3] دانشیار گروه سخت‌افزار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران
Akbari@iust.ac.ir

چکیده

گرایش نوظهور برون‌سپاری نیازهای طراحی و ساخت به تسهیلات خارجی، همراه با اتکای در حال افزایش به مالکیت‌های معنوی شخص ثالث[1] و ابزارهای اتوماتیک طراحی سخت‌افزار موجب شده است تا مدارات مجتمع[2] در مراحل مختلفی از چرخه حیات‌شان، بیش از پیش در برابر حملات تروجان‌های سخت‌افزاری آسیب‌پذیر شوند. در این پژوهش ما ابتدا به بررسی مفهوم تروجان سخت-افزاری و خصوصیات آن و دسته‌بندی‌های ارائه شده می‌پردازیم و سپس روش‌های مختلف کشف آن را با توجه به یک دسته‌بندی برای روش‌های کشف مورد مطالعه قرار خواهیم داد. سپس به تشریح روش‌های مطرح برای مقابله با این معضل امنیتی خواهیم پرداخت و در انتها راه‌کار پیشنهادی با توجه به مشکلات و مسائل مطرح در کشف تروجان در مرحله آزمون مدارات مجتمع ارائه خواهد شد.

کلمات کلیدی

تروجان سخت‌افزاری، کشف، کانال جانبی، امنیت سخت‌افزار، آزمون، نظارت

۱- مقدمه

استفاده از سامانه‌های رایانه‌ای و الکترونیکی در طی چند دهه‌ی اخیر رشد چشمگیری داشته است تا جایی که اکثر جنبه‌های زندگی روزانه نیز از مدارات مجتمع کمک یا تاثیر گرفته‌اند. امکان اطمینان به این مدارات مجتمع به جهت انجام صحیح کار مشخص شده برای آنها (و انجام فقط کار تعریف شده و مشخص شده برای آنها) همواره یکی از نگرانی‌های امنیتی بوده‌است. با این نگاه، امنیت سخت‌افزار در سال‌های اخیر به یک موضوع مهم تحقیقاتی در دنیا تبدیل شده است [۱]،[۲] ، [۳].

کاربردها و استفاده‌های فراوان از مدارات مجتمع و سخت‌افزارهای مختلف در حوزه‌های بحرانی گوناگونی همچون صنایع نظامی، اقتصادی، ارتباطات، پزشکی و غیره (جاهایی که عواقب یک حمله موفقیت‌آمیز می‌تواند بسیار جدی و پرهزینه باشد)، اهمیت مطالعه و پژوهش در این موضوع را بیشتر نموده است. از طرف دیگر سخت‌افزار به عنوان پایین‌ترین لایه هر سکوی[3] محاسباتی، بالاترین سطح دسترسی را داراست و سوءاستفاده از آسیب‌پذیری[4]های امنیتی احتمالی در این لایه با مکانیزم‌های امنیتی موجود در سطح نرم‌افزاری و برنامه، قابل شناسایی و پیش‌گیری نیست. این سطح این استعداد را دارد که کنترل کامل بر سامانه هدف را، به حمله‌کننده بدهد [۴].

کاربردهای بحرانی ذکر شده در محیط‌های با خطرپذیری بالا و ظرفیت حملات مبتنی بر سخت‌افزار، نیاز به ساخت سکوهای محاسباتی با امنیت بالا را بیش از پیش روشن می‌سازد.

یکی از آسیب‌پذیری‌های مطرح شده و مورد پژوهش در سطح سخت‌افزار، ورود تروجان سخت‌افزاری به مدار مجتمع مورد استفاده است. تروجان‌های سخت‌افزاری مدارات آلوده‌ای هستند که می‌توانند عملکرد[5] و اطمینان‌پذیری[6] سامانه‌های سخت‌افزاری را تحت تاثیر قرار دهند. این مدارات می‌توانند در هر مرحله‌ای از طراحی تا ساخت، در طراحی و مدار اصلی به صورت مخفیانه جاسازی شوند تا کار تعریف شده برای آنها را انجام دهند [۵]. لذا اشراف یافتن بر پدیده‌ای با نام تروجان سخت‌افزاری هنگام توسعه نسل آینده مکانیزم‌های امنیتی برای توسعه و بکارگیری صنعت الکترونیک در حضور تهدید تروجان-های سخت‌افزاری بسیار حیاتی است [۶].

نگرانی احتمال وجود تروجان‌های سخت‌افزاری در مدارات مجتمع مورد استفاده در صنایع بحرانی به چند دلیل افزایش و قوت پیدا نموده است:

الف- استفاده از مالکیت‌های معنوی شخص ثالث

امروزه استفاده از طرح‌های مالکیت معنوی شخص ثالث به‌دلیل افزایش سرعت طراحی و بهبود کیفیت آن بسیار مرسوم است. بلوک‌های مالکیت معنوی توسط صدها سازنده در سراسر جهان تولید می‌شوند و در طرح‌های بزرگ مورد استفاده قرار می‌گیرند. با توجه به تجاری بودن این ماژول‌ها برای طراح ماژول، مخفی ماندن طرح ماژول بسیار مهم است [۷] و با توجه به همین گرایش و عدم اطلاع از طرح موجود در مالکیت معنوی، شائبه و تردید امکان وجود تروجان سخت‌افزاری در مالکیت‌های معنوی شخص ثالث افزایش می‌یابد.

ب- استفاده از ابزارهای طراحی اتوماتیک شخص ثالث

ابزارهای سنتز و طراحی اتوماتیک مختلفی برای تبدیل سخت‌افزار توصیف شده توسط زبان‌های توصیف سخت‌افزار به طرح نهایی برای تولید وجود دارند. منبع باز نبودن ابزارهای تجاری به دلایل اقتصادی، از علل تقویت این نگرانی است. چرا که تغییر بداندیشانه توسط طراح نرم‌افزار و یا حمله‌کننده‌ی متخاصم در آن برای تغییر طرح نهایی از دیگر موارد ممکن برای واردسازی تروجان در سخت‌افزار است که حمله‌کننده را از دسترسی فیزیکی به طراحی نیز بی‌نیاز می‌سازد [۸].

ج- برون‌سپاری[۷] نیازها

جهانی شدن صنعت نیمه‌هادی‌ها و مزایای بازار رقابتی تولید محصولات نیمه‌هادی با کیفیت بالاتر و هزینه کمتر، شرکت‌های طراحی را مجبور به برون‌سپاری مراحل تولید نموده است. به‌طوری که امروزه حتی کشورهایی که از سازندگان اصلی این مدارات بوده‌اند امروزه بخش قابل توجهی از تولید طراحی‌های مورد نیازشان را توسط شرکت‌های بزرگ تولیدکننده تامین می‌نمایند.

علی رغم فواید ذکر شده برای برون‌سپاری تولید مدارات مجتمع و مدارات الکترونیکی، تغییر در طراحی و ساخت این مدارات در زمان تولید به یک نگرانی فراگیر برای صنایع مصرف‌کننده این محصولات تبدیل شده است. از سوی دیگر، هنگامی که طراحی‌ها پیچیده‌تر و بزرگ‌تر می‌شوند، احتمال این خطر بیشتر می‌شود؛ چرا که جاسازی یک تروجان سخت‌افزاری در مدارات بزرگ‌تر کاری ساده‌تر و شناسایی آن به مراتب کاری سخت‌تر خواهد بود.

مواردی از وجود این تروجان‌ها گزارش شده است که به چند مورد آن به‌طور خلاصه اشاره می‌کنیم.

- در سپتامبر ۲۰۰۷، جنگنده‌های اسرائیلی به یک نیروگاه هسته‌ای در شمال‌غرب سوریه حمله می‌کنند که در این حمله رادارهای سامانه پدافندی سوریه هیچگونه هشداری را در مورد رخداد این حمله گزارش نکردند. پس از مدتی وبلاگ‌نویسان نظامی و فنآوری نتیجه‌گیری کردند که این حادثه به دلیل یک جنگ‌افزار الکترونیک بوده است [۴]، [۳].
- بر اساس گزارش یک پیمانکار نظامی امریکایی، بنابردرخواست پیمانکاران دفاعی فرانسه، شرکت‌های سازنده تراشه اروپایی اخیرا درون ریزپردازنده‌هایشان یک کلیدکشنده[۸] قرار داده‌اند تا در صورتی که این تجهیزات در آینده در دست دشمن قرار بگیرد، فرانسوی‌ها بتوانند آنها را غیرفعال نمایند [۳].
- پژوهشگران اخیرا (۲۰۱۰) وجود یک کلمه عبور به صورت سخت‌افزاری کد شده که امکان دسترسی به برخی ثبات‌های وضعیت ماشینِ ذکر نشده را فراهم می‌کند، را افشا کردند که در برخی پردازنده‌های AMD قرار داده شده و البته هنوز روشن نیست چه سطحی از کنترل با دسترسی به این ثبات‌ها قابل دستیابی‌ست [۴].
- بر اساس اطلاعاتی که در سال ۲۰۰۷ منتشر شد، هاردهای خارجی شرکت Seagate حاوی تروجان‌های از پیش نصب شده‌ای بودند که پسوردها را به یک دشمن دور ارسال می‌کردند [۹].
- شرکت اینتل نسخه‌ای از پردازنده‌های تولیدی‌اش (Sandy Bridge) را ارائه کرده است که برای پیشگیری از سوء استفاده از اطلاعات کاربران از یک فنآوری ضد سرقت در آن استفاده شده است که در هنگام سرقت رایانه با استفاده از یک کلید کشنده که در پردازنده آن تعبیه شده است، کاربر می‌تواند آن را از راه دور خاموش کند [۱۰].

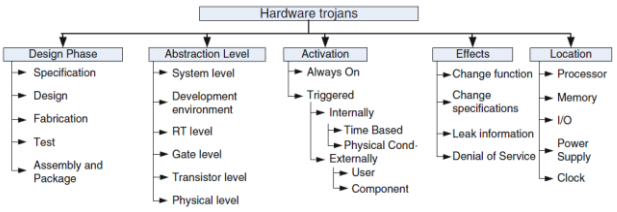
۲- تروجان سخت افزاری

تروجان‌های سخت‌افزاری می‌توانند به صورت تغییرات سخت‌افزاری در مدارات مجتمع خاص منظوره (ASICs)، بخش‌های تجاری عام منظوره (COTS)[۹] و یا به صورت دست‌کاری‌های میان‌افزار[۱۰] در جریان بیتی FPGA پیاده‌سازی گردد [۱۱].

تروجان‌های سخت‌افزاری را از نقطه نظرات متفاوتی می‌توان دسته بندی نمود و دسته‌بندی‌های مفهومی مختلفی نیز ارائه شده است [۱۲]، [۱۳]، [۱۴]، [۵]. دسته‌بندی ارائه شده در [۱۳]، تروجان‌های سخت‌افزاری را بر اساس سه خصوصیت عمده‌ی فیزیکی، نحوه تحریک و عملکرد آنها دسته‌بندی نموده است. در بخش فیزیکی چهار زیردسته بر اساس نوع، اندازه، توزیع و ساختار تروجان آمده است. تروجان‌ها از لحاظ نوع به دو دسته عملکردی و پارامتری تقسیم می‌شوند. تروجان‌های عملکردی تروجان‌های هستند که با افزودن یا حذف دروازه یا ترانزیستور به مدار محقق شده‌اند. تروجان‌های پارامتری تروجان‌هایی‌اند که با تغییر خصوصیات فیزیکی سیم‌ها و منطق موجود پیاده‌سازی می‌شوند. تغییراتی همچون باریک کردن سیم‌ها، ضعیف کردن ترانزیستورها یا هر تغییری در هندسه فیزیکی طراحی شده برای خرابکاری در اطمینان‌پذیری مدار.

تروجان‌های سخت‌افزاری عملکردی معمولا از دو بخش محرک[11] و بار[12] تشکیل شده‌اند. مدار محرک ذاتا یک مدار غیرفعال با کاربری نظارتی است بدون اثر عملکردی بر کارایی مدار. محرک رخداد شرایط خاص و معمولا نادری که برای آن تعریف شده است را جستجو می‌کند و با رخداد آن شرایط یا واقعه، بخش بار تروجان را فعال می‌نماید. بخش بار تروجان سخت‌افزاری مسئولیت پیاده‌سازی وظیفه تعریف شده برای تروجان را بر عهده دارد. در بخش‌های آتی به عملکردهای ممکن این نوع تروجان به صورت مشروح خواهیم پرداخت.

ما در این سمینار از دسته‌بندی نسبتا جامع ارائه شده در [۵] استفاده می‌کنیم و بر اساس آن خصوصیات تروجان‌های سخت‌افزاری را شرح می‌دهیم. این خصوصیات شامل: فاز تزریق تروجان به مدار، سطح انتزاع تروجان، مکانیزم فعال‌سازی تروجان، تاثیر و عملکرد تروجان و مکان تروجان است. در ادامه به بررسی هر کدام از این موارد می‌پردازیم.



شکل ۱ - دسته‌بندی تروجان‌های سخت‌افزاری [۵]

۲-۱- فاز طراحی تروجان سخت‌افزای

تروجان سخت‌افزاری می‌تواند در طول چرخه توسعه و تولید تراشه وارد آن گردد. برهمین اساس تروجان‌های سخت‌افزاری را می‌توان بر اساس فاز ورود به تراشه دسته بندی نمود. این دسته بندی عبارت است از:

۲-۱-۱- خصوصیات

در فاز خصوصیات، ویژگی‌های سیستم تعریف می‌شود. ویژگی‌هایی همچون محیط هدف سیستم، عملکرد مورد توقع، اندازه، توان، تاخیر و غیره. در زمان توسعه و کار بر روی مدار مجتمع در این مرحله، ویژگی‌های عملکردی یا دیگر محدودیت‌های طراحی می‌تواند تغییر یابد. برای مثال، محدودیت‌های زمان‌بندی سخت‌افزار ممکن است توسط تروجان تغییر داده شود.

۲-۱-۲- فاز طراحی

در فاز طراحی محدودیت‌های عملکردی، منطقی، زمان‌بندی و فیزیکی تعریف شده به هنگام نگاشت طراحی به فناوری مورد نظر در نظر گرفته می‌شود. در این مرحله ممکن است طراحان از مالکیت‌های معنوی شخص ثالث استفاده نمایند.

۲-۱-۳- ساخت

در فاز ساخت، مجموعه الگو[13] تولید می‌شود و ویفرها با استفاده از الگوها تولید می‌گردند. تغییرات ماهرانه‌ی الگو می‌تواند اثرات جدی داشته باشد. در حالت افراطی، دشمن می‌تواند مجموعه الگوی جایگزین خودش را با مجموعه الگوی اصلی عوض کند. یا در حالت دیگر، ترکیبات شیمیایی مورد استفاده در فرایند، به جهت افزایش مهاجرت الکتریکی[14] در ترکیب مداراهای بحرانی همچون منبع تغذیه و شبکه‌های ساعت ممکن است تغییر داده شود؛ که موجب تسریع خرابی می شود.

۲-۱-۴- مونتاژ[15]

در طی فاز مونتاژ، تراشه آزمایش شده و دیگر اجزاء سخت‌افزاری بر روی برد مدار چاپی (PCB)[16] مونتاژ می‌شوند. هر واسطی[17] در یک سامانه که دو یا بیش از دو جزء توسط آن با هم در تعاملند یک مکان مستعد برای ورود تروجان است. حتی اگر تمام مدارات مجتمع مورد استفاده در یک سامانه قابل اعتماد باشند، مونتاژ بداندیشانه[18] می‌تواند موجب پدید آمدن حفره‌ها و نقص های امنیتی در سامانه شود. به عنوان مثال یک سیم بدون پوشش متصل به یک گره در PCB می‌تواند بین سیگنال بر روی برد و محیط الکترومغناطیسی آن، زوج الکترومغناطیسی قابل بهره‌برداری برای نشت اطلاعات و تزریق اشکال، بوجود بیاورد.

۲-۱-۵- آزمون

این فاز برای اعتماد سخت‌افزار مهم است، نه از این جهت که ممکن است در این فاز ورود تروجان صورت پذیرد؛ بلکه به این دلیل که این فاز یک ظرفیت برای کشف تروجان است. آزمون هنگامی برای شناسایی و کشف تروجان قابل استفاده است که خودش قابل اعتماد باشد. برای مثال، دشمن که یک تروجان را در مرحله ساخت وارد ساخته است، می‌خواهد که بر روی بردارهای آزمون کنترل داشته باشد تا مطمئن شود که تروجان در مرحله آزمون شناسایی نمی‌شود. اعتماد در مرحله آزمون یعنی بردارهای آزمون مخفی نگه‌داشته خواهند شد، بردارهای آزمون صادقانه اعمال خواهند شد و اعمال مشخص شده همچون رد و پذیرش صادقانه انجام خواهند پذیرفت.

۲-۲- سطح انتزاع تروجان سخت‌افزاری

مدارات تروجان می توانند در سطوح انتزاع مختلفی وارد شوند. عملکرد و ساختار آنها به شدت وابسته به سطح انتزاعی‌ست که در آن توصیف شده‌اند.

۲-۲-۱- سطح سامانه

در سطح سامانه، ماژول‌های سخت‌افزاری مختلف، ارتباطات و پروتکل‌های ارتباطی مورد استفاده تعریف می‌شوند. در این سطح،

تروجان ممکن است توسط ماژول‌های سخت‌افزار هدف تحریک شوند. برای مثال مقادیر ASCII ورودی‌های صفحه کلید می‌توانند با هم عوض شوند.

۲-۲-۲- محیط توسعه

یک محیط توسعه نوعی، شامل ابزارهای سنتز، شبیه‌سازی، تصدیق و اعتبارسنجی است. ابزارهای و دستورالعمل‌های ابزارهای طراحی به‌کمک رایانه (CAD)[۱۹] می‌توانند برای ورود تروجان مورد استفاده قرار گیرند. تروجان‌های نرم‌افزاری وارد شده به این ابزارهای طراحی به‌کمک رایانه می‌تواند اثرات تروجان‌های سخت‌افزاری را بپوشاند. برای مثال، بخش‌های تروجان یک مدار ممکن است توسط ابزار سنتز به کاربر نشان داده نشود.

۲-۲-۳- سطح انتقال ثبات

در سطح انتقال ثبات، هر ماژول عملکردی توسط ثبات‌ها، سیگنال‌ها و توابع بولی توصیف می‌شود. زمانی که حمله‌کننده کنترل کامل بر عملکرد سخت‌افزار در این سطح دارد، یک تروجان می‌تواند به راحتی طراحی شود و در سطح انتقال ثبات وارد شود. برای مثال یک تروجان پیاده‌سازی شده در این سطح ممکن است تعداد مرتبه‌های اجرای یک الگوریتم رمزنگاری را با تغییر مکانیزم شمارش شمارنده‌ی تعداد اجرا به پله‌های دوتایی به جای تکی، نصف کند.

۲-۲-۴- سطح دروازه

در سطح دروازه، طراحی به صورت ارتباطات بین دروازه‌های منطقی نشان داده می‌شود. این سطح به حمله‌کننده اجازه می‌دهد تا تمام جنبه‌های تروجان وارد شده، از جمله اندازه و مکان را به‌دقت کنترل کند. برای مثال یک تروجان ممکن است یک مقایسه‌گر ساده، تشکیل شده از دروازه‌های XOR باشد که سیگنال‌های داخلی تراشه را تحت نظر دارد. این تروجان‌ها معمولا برای تغییر عملکرد طرح استفاده می‌شوند و بنابراین به آنها تروجان‌های عملکردی نیز می‌گویند.

۲-۲-۵- سطح ترانزیستور

ترانزیستورها برای ساخت دروازه‌ها استفاده می‌شوند. این سطح به طراح تروجان کنترلی بر روی ویژگی‌های مدار از جمله توان و زمان-بندی می‌دهد. ترانزیستورهایی که به صورت تکی می‌توانند حذف یا اضافه شوند، عملکرد مدار را تغییر می‌دهند. برای تغییر پارامترهای مدار اندازه‌های ترانزیستورها می‌توانند تغییر داده شوند. برای مثال، یک تروجان سطح ترانزیستور ممکن است یک ترانزیستور با عرض دروازه کم باشد که موجب تاخیر بیشتر در مسیر بحرانی مدار شود. تروجان‌های در سطح ترانزیستور معمولا برای تغییر کارایی طرح استفاده می‌شوند و بنابراین به آنها تروجان‌های عملکردی نیز گفته می‌شود.

۲-۲-۶- سطح طرح‌بندی[۲۰]

در سطح طرح‌بندی ابعاد و مکان‌های همه اجزای مدار توصیف می‌گردد. این سطح، سطح فیزیکی طرح است، جایی که یک تروجان می‌تواند وارد شود. تروجان ممکن است به صورت تغییر اندازه سیم‌ها، فاصله بین المان‌های مدار یا تخصیص مجدد لایه‌های فلزی وارد شود. برای مثال، تغییر عرض سیم‌های فلزی شبکه ساعت درون تراشه می‌تواند موجب انحراف ساعت[۲۱] شود. تروجان‌های در سطح طرح‌بندی معمولا به صورت تغییر در پارامترهای طراحی فیزیکی مدار مجتمع توصیف می‌شود و بنابراین تروجان‌های پارامتری نامیده می‌شوند.

بر اساس تعداد دروازه‌های وارد شده، تروجان‌های سخت‌افزاری می توانند بزرگ یا کوچک باشند. همچنین بر اساس توزیع تروجان در طرح، تروجان‌ها می‌توانند دارای پیوند محکم باشند یا آزادانه توزیع شده باشند.

۲-۳- مکانیزم تحریک تروجان سخت‌افزاری

برخی تروجان‌ها بگونه‌ای طراحی شده‌اند که همیشه روشن باشند؛ بقیه ممکن است خاموش بمانند تا زمان تحریک شدن‌شان توسط شرایط تعریف شده برای محرک. یک تروجان همیشه روشن، همچنان که از اسمش نیز بر می‌آید در تمام زمان‌ها، تراشه را تحت تاثیر قرار می‌دهد. تروجان‌های پارامتری که اثرشان بر طرح‌بندی فیزیکی است به عنوان تروجان‌های همیشه روشن در نظر گرفته می‌شوند.

یک تروجان تحریک‌شونده برای فعال شدن، نیاز به یک رویداد داخلی یا خارجی دارد. یک‌بار که محرک تروجان را فعال کرد، می‌تواند برای همیشه فعال بماند یا پس از مدتی مجدد به حالت خاموش برگردد.

- **تحریک‌شونده داخلی**

یک تروجان تحریک‌شونده داخلی، توسط یک اتفاق که در درون وسیله هدف رخ می‌دهد، تحریک می‌شود. این اتفاق ممکن است بر اساس زمان باشد یا بر اساس شرط فیزیکی. یک شمارنده در طرح می‌تواند یک تروجان را در یک زمان از پیش تعیین شده تحریک نماید که به آن بمب زمانی گفته می‌شود. تروجان مبتنی بر شرط فیزیکی می‌تواند توسط طیف وسیعی از شرایط فیزیکی از جمله تداخل الکترومغناطیسی، رطوبت، ارتفاع، فشار جو، دما و غیره تحریک شود. همچنین یک تروجان می‌تواند با رسیدن به یک حالت خاص از ماشین حالت فعال شود.

- **تحریک‌شونده خارجی**

یک تروجان تحریک‌شونده خارجی به ورودی‌های خارجی به ماژول هدف نیاز دارد تا تروجان را تحریک کند. محرک خارجی می‌تواند یک ورودی وارد شده توسط یک کاربر باشد یا خروجی یک بخش دیگر سامانه. محرک‌های ورودی کاربر می‌توانند شامل دکمه‌های فشاری، کلیدها، صفحه کلیدها یا کلمات کلیدی و عبارات خاصی در جریان داده ورودی باشند. محرک‌های بخش خارجی ممکن است از هر

کدام از بخش‌هایی که با دستگاه هدف تعامل دارند باشد. برای مثال، یک تروجان می‌تواند توسط داده‌های واردشونده از طریق واسط‌های خارجی همچون RS-۲۳۲ تحریک شود. معمولا تروجان‌های تحریک-شونده خارجی برای دریافت تحریک خارجی، نیاز به مدارات سنجش دارند.

یک دسته‌بندی ممکن دیگر می‌تواند مکانیزم فعال‌سازی تروجان را به دو دسته‌ی تحریک‌شونده سنسوری و تحریک‌شونده منطقی تقسیم‌بندی نماید. تروجان‌های تحریک‌شونده سنسوری بر اساس شرایط فیزیکی از جمله دما، ولتاژ و غیره تحریک می‌شوند. تروجان‌های تحریک‌شونده منطقی با شرایط منطقی از جمله وضعیت فلیپ فلاپ، شمارنده، سیگنال ساعت، داده، دستورالعمل و/یا وقفه‌ها تحریک می‌شوند.

۲-۴- عملکرد تروجان سخت‌افزاری

تروجان‌های سخت‌افزاری بسته به هدف سازنده و مدار مورد هدف تروجان می‌توانند اهداف متفاوتی را دنبال نمایند و شدت اثرات آنها بر روی سخت‌افزار یا سامانه هدف می‌تواند در بازه‌ای از اختلالات ظریف تا خرابی‌های فاجعه‌آمیز سامانه قرار گیرد. دسته‌ای از آنها می‌توانند همیشه فعال باشند و تنها از طریق یک کانال مخفی، موجب نشت اطلاعات سری سامانه شوند. دسته‌ی دیگر می‌توانند با رخداد در شرایطی بسیار نادر موجب تغییر عملکرد سامانه شوند. از سوی دیگر تغییر در خصوصیات ساخت مدار نیز می‌تواند موجب کاهش اطمینان‌پذیری مدار طراحی شده شود.

۲-۴-۱- کاهش اطمینان‌پذیری مدارمجتمع

یک تروجان می‌تواند با تغییر عمدی پارامترهای دستگاه موجب کاهش کارایی شود. این نوع عملکرد عمدتا توسط تروجان‌های پارامتری صورت می‌گیرد. آنها ممکن است خصوصیات عملکردی، واسط یا پارامتری همچون توان مصرفی و تاخیر را تغییر دهند. تروجان‌های پارامتری با تغییر در خصوصیات مدار موجب وارد شدن خرابی‌هایی همچون خرابی‌های Stuck-at یا خرابی اتصال کوتاه و نیل مدار مجتمع به تولید خروجی‌های خطادار می‌شوند و یا می‌توانند موجب بروز شرایط نامطلوب در سامانه شوند. برای مثال، یک تروجان ممکن است تعداد بافرهای بیشتری را در ارتباطات تراشه وارد کند و در نتیجه موجب مصرف توان بیشتر گردد که در نتیجه باطری را به-سرعت تخلیه خواهد کرد.

کشف تروجان‌های با این نوع تاثیرگذاری بر مدارات مجتمع بسیار مشکل هستند چرا که مدار کارایی اصلی‌اش را انجام می‌دهد و هیچ عملکرد اضافه‌ای را اجرا نمی‌کند و تنها کیفیت کار مدار افت خواهد داشت.

۲-۴-۲- نشت اطلاعات محرمانه سامانه از طریق روش و کانال مخفی تعبیه شده

یک تروجان می‌تواند عامل نشت اطلاعات حساس از سیستم باشد. این نوع عملکرد تروجان‌های سخت‌افزاری مستلزم ایجاد تغییرات سخت‌افزاری در مدار اصلی با هدف ارسال اطلاعات حساس از سامانه اطلاعاتی به دشمن، بدون اطلاع یا همکاری سامانه اطلاعاتی یا کاربر سامانه است. این کار می‌تواند از طریق کانال‌های مخفی یا آشکار صورت پذیرد و مکانیزم‌های ارسال می‌توانند از مسیرهای داخلی یا خارجی خود سامانه استفاده کنند یا از طریق کانال‌های جانبی اقدام به ارسال اطلاعات نمایند. ارسال اطلاعات از طریق کانال‌های آشکار می‌تواند از طریق واسط‌های RS-۲۳۲ و JTAG صورت پذیرد و ارسال اطلاعات از طریق کانال جانبی می‌تواند از طریق رسانه‌هایی همچون فرکانس رادیویی، نور، دما، توان مصرفی و زمان‌بندی صورت پذیرد. ارسال‌ها همچنین می‌توانند در حاشیه نویز خصوصیات فیزیکی یا عملکردی مدار مجتمع مخفی شوند. جین و ماکریس در [۱۵] کلیدهای رمزنگاری را از طریق حاشیه‌های دامنه انتقال بی‌سیم یا فرکانسش که بدلیل تغییرات روند[۲۲] بوجود می‌آید نشت داده‌اند و لین و همکارانش در [۱۶] داده‌ها را از طریق یک تکنیک کانال جانبی طیف گسترده در زیر سطح نویز روند CMOS نشت داده‌اند.

۲-۴-۳- تغییر در عملکرد سامانه

یک تروجان می‌تواند در کارایی دستگاه هدف تغییر ایجاد نموده و موجب بروز خطاهای ظریفی شود که ممکن است به‌سادگی قابل کشف نباشند. برای مثال، یک تروجان ممکن است باعث شود یک ماژول کشف خطا ورودی‌هایی که باید رد کند را بپذیرد.

این دسته همچنین شامل تروجان‌هایی می‌شود که خصوصیات طراحی را به‌نحوی تغییر می‌دهند که موجب ایجاد عملکردی متفاوت از عملکرد مطلوب در سامانه می‌شود. مثال تغییر عملکرد رادار سوریه که در مقالات [۳]، [۴] بیان شده از این دسته است.

۲-۴-۴- از کار انداختن سامانه یا حمله تکذیب سرویس

تروجان‌های تکذیب سرویس (DoS)[۲۳] می‌توانند مانع از انجام کار یک تابع یا منبع شوند. یک تروجان ممکن است باعث تمام شدن منابع حیاتی و کمیابی همچون پهنای باند، توان محاسباتی و توان باتری شود. یک تروجان ممکن است موجب تخریب فیزیکی، غیرفعال شدن یا تغییر پیکربندی دستگاه شود. برای مثال یک تروجان ممکن است باعث شود پردازنده درخواست وقفه‌ی یک دستگاه جانبی خاص را نادیده بگیرد.

و در نهایت تروجان‌های سخت‌افزاری همچنین می‌توانند برای کمک به اجرای حملات مبتنی بر نرم‌افزار، همچون حملات ارتقای سطح دسترسی، ورود از طریق درب پشتی[۲۴]، سرقت کلمه عبور و حملات تکذیب سرویس[۲۵] طراحی شوند [۱۱].

۲-۵- مکان تروجان سخت‌افزاری

یک تروجان می‌تواند در یک جزء تنها وارد شود و یا در میان چندین جزء گسترش یابد. تروجان‌ها می‌توانند در واحدهای پردازشی، حافظه، ورودی/خروجی، شبکه توان یا شبکه ساعت قرار گیرند. تروجان‌های توزیع شده در میان اجزاء مختلف ممکن است مستقل از یکدیگر عمل کنند یا به صورت گروهی کارشان را انجام دهند.

۲-۵-۱- واحدهای پردازشی

تروجان‌ها ممکن است به واحدهای پردازشی وارد گردند. هر تروجانی که درون واحدهای منطقی که بخشی از واحد پردازشی‌اند، جاسازی شده باشد، می‌تواند در این گروه قرار گیرد. برای مثال، یک تروجان در پردازنده ممکن است ترتیب اجرای دستورالعمل‌ها را تغییر دهد.

۲-۵-۲- واحدهای حافظه

تروجان‌های قرار گرفته در بلاک‌های حافظه و واحدهای واسط-شان می‌توانند در این دسته قرار گیرند. آنها ممکن است مقدار ذخیره شده در حافظه را تغییر دهند یا دسترسی خواندن یا نوشتن به برخی مکان‌های حافظه را مسدود نمایند. برای مثال، یک تروجان ممکن است محتوای حافظه PROM یک مدار مجتمع را تغییر دهد.

۲-۵-۳- واحدهای ورودی/خروجی

تروجان‌ها می‌توانند در لوازم جانبی تراشه‌ها یا درون PCB ساکن شوند. این واحدها با واحدهای خارجی تعامل دارند و در نتیجه این موقعیت برای تروجان امکان کنترل داشتن بر روی ارتباط داده‌ای بین پردازنده و اجزاء خارجی سامانه را فراهم می‌کند. برای مثال، یک تروجان ممکن است داده‌ای که از طریق گذرگاه RS-۲۳۲ می‌آید را تغییر دهد.

۲-۵-۴- واحدهای منبع تغذیه

تروجان‌ها ممکن است ولتاژ و جریان عرضه شده به تراشه را تغییر داده و موجب خرابی شوند.

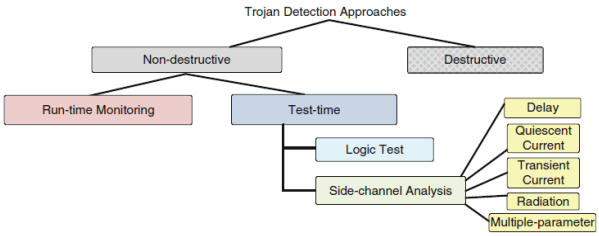
۲-۵-۵- شبکه ساعت

تروجان‌ها در شبکه ساعت، فرکانس ساعت را تغییر داده و یا اشتباهات کوچکی[۲۶] را به ساعت عرضه شده به تراشه وارد می‌سازد که موجب حملات خرابی می شود. تروجان‌های قرار گرفته در شبکه ساعت همچنین می‌توانند سیگنال ساعت عرضه شده به دیگر ماژول-های درون تراشه را متوقف[۲۷] کند. برای مثال، یک تروجان ممکن است انحراف سیگنال ساعت عرضه شده به بخش‌های خاصی از تراشه را افزایش دهد.

۳- روش‌های کشف تروجان‌های سخت‌افزاری

در بخش‌های گذشته به مطالعه چیستی تروجان‌های سخت‌افزاری پرداختیم. در این بخش به مطالعه و بررسی روش‌های کشف و مقابله با تروجان‌های سخت‌افزاری خواهیم پرداخت. بسته به نوع و عملکرد تروجان روش‌های مختلفی برای کشف تروجان‌های سخت‌افزاری تحقیق و ارائه شده است. روش‌های آزمون و تصدیق مرسوم نمی‌توانند برای کشف قابل اتکا و اعتماد تروجان‌های سخت‌افزاری مورد استفاده قرار گیرند. هدف و تمرکز روش‌های آزمون مرسوم بر شناسایی رفتار عملکردی غیرمطلوب ناشی از خرابی‌ها در مدار مجتمع است و قصد کشف و شناسایی عملکردهای اضافه و پنهان در یک طرح که ناشی از تغییرات خراب‌کارانه‌اند را ندارند. از سوی دیگر، یک تروجان هوشمندانه طراحی شده مدار نسبتا محجوبیست که برای گریز از کشف، تنها تحت شرایط نادری یک خراب‌کاری را تحریک می‌نماید.

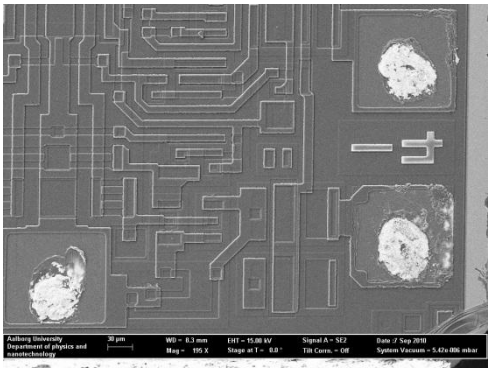
بر اساس یک دسته‌بندی ارائه شده در [۵] برای روش‌های کشف تروجان‌های سخت‌افزاری، به معرفی این روش‌ها و جزئیات مختصری از آنها خواهیم پرداخت.



شکل ۲ - دسته‌بندی روش‌های کشف تروجان [۵]

۳-۱- روش‌های تخریبی

تمرکز و کار روش‌های تخریبی بر روی مدارات مجتمع ساخته شده است؛ که با جداسازی با استفاده از صیقل مکانیکی-شیمیایی و سپس با پویش میکروسکوپ الکترونی (SEM)[۲۸]، تصاویر لایه به لایه مدار مجتمع را استخراج می‌نمایند.



شکل ۳ - نمونه‌ای از یک تصویر SEM

پس از استخراج تصاویر، بازسازی تصاویر و تحلیل برای شناسایی ترانزیستورها، دروازه‌ها و المان‌های مسیریابی انجام می‌شود. سپس یک

روش مهندسی معکوس از پایین به بالا[29] مورد استفاده قرار می‌گیرد که در آن با بهره‌گیری از یک روش تطبیق قالب، ابتدا بلاک‌های ساختاری مدار مجتمع مشخص شده و برای تعیین تمام توابع پیاده-سازی شده، با هم گروه‌بندی می‌شوند.

کاربرد اولیه این تکنیک‌ها برای مهندسی معکوس مدارات مجتمع رقیب جهت استخراج رموز کاری و طراحی بهتر محصولات بوده است. اما این روش‌ها می‌توانند برای تصدیق عملکرد یک مدار مجتمع داخلی که خصوصیات تعریف شده برای آن سالم و بدون تشکیک بوده، جهت تعیین وجود مدار یا عملکرد ناخواسته در مدار مجتمع تولید شده به-کارگرفته شوند. روش‌های مذبور به‌شدت زمان‌گیر و پرهزینه هستند (تحلیل تخریبی یک تراشه چندین ماه وقت خواهد برد [۱۷]) و با افزایش چگالی تجمیع ترانزیستورها نمی‌توانند تغییر مقیاس یابند و ممکن است ناکارآمد شوند. همچنین، نتایج به‌دست آمده از یک آزمایش نمی‌تواند به تمام تولیدات تعمیم داده شود چرا که ممکن است یک دشمن تنها بخش کوچکی از مدارات مجتمع را تغییر داده باشد. بنابراین هر مدار مجتمع برای جلب اعتماد، نیاز به آزمون جداگانه دارد. کاربرد مناسب ممکن برای این تکنیک علی رغم مشکلات موجود در کشف تروجان با این تکنیک، کمک برای بدست آوردن خصوصیات یک مجموعه مدار مجتمع سالم جهت کالیبره‌سازی روند و مقایسه پارمترهای کانال جانبی آنها با دیگر مدارات مجتمع، برای تعیین وضعیت آلودگی آن مدارهاست.

۳-۲- روش‌های غیرتخریبی

روش‌های غیرتخریبی می‌توانند به دو دسته دسته‌بندی شوند: ۱) روش‌های زمان آزمون و ۲) روش‌های زمان اجرا.

یک راه حل ممکن برای افزایش سطح اعتماد در کشف، ترکیب روش‌های مختلف کشف با قابلیت‌های مکمل با یکدیگر است. برای مثال روش آزمون منطقی می‌تواند با تحلیل کانال جانبی ترکیب شود. به طور مشابه، راه حل‌های آزمون ساخت می‌توانند با نظارت برخط ترکیب شوند. در ادامه به بررسی روش‌های قرار گرفته در این زیرمجموعه خواهیم پرداخت.

۳-۲-۱- روش‌های زمان آزمون

روش‌های زمان آزمون به روش‌هایی اتلاق می‌شود که پیش از قرار گرفتن مدار مجتمع مورد نظر در مدار اصلی‌اش برای کشف تروجان به آن اعمال می‌شوند. این روش‌ها می‌توانند مشابه روش‌های طراحی برای آزمون (DfT)[30] توسط مدارات طراحی برای امنیت (DfS)[31] پشتیبانی شوند. این مدارات حساسیت یا پوشش کشف تروجان را به طور قابل ملاحظه‌ای می‌توانند افزایش دهند. اما برای این منظور ضروری‌ست از سالم بودن مداربندی آزمون افزوده شده اطمینان حاصل شود. برای مثال، اگر در مدار اضافه شده یک سیگنال به‌سادگی قابل تشخیص برای فعال‌سازی آزمون وجود داشته باشد، حمله کننده می‌تواند از این سیگنال سوء استفاده نموده و رفتار مخرب تروجان را در زمان آزمون غیرفعال نماید و یا ورودی تغذیه آن را با دروازه‌ها قطع نماید تا مانع از نشت اطلاعات کانال جانبی آن نیز گردد.

روش‌های زمان آزمون نیز خود به دو دسته تقسیم‌بندی می‌شوند: ۱) آزمون منطقی و ۲) تحلیل کانال جانبی.

- **آزمون منطقی**

روش‌های آزمون منطقی بر تولید و اعمال بردارهای آزمون برای فعال‌سازی مدار تروجان و مشاهده اثر مخرب آن در خروجی‌های اصلی تمرکز می‌کنند. این روش‌ها در ماهیت مشابه آزمون خرابی[32] Stuck-at هستند، هرچند مدل‌های تروجان‌ها بسیار متفاوت از مدل‌های خرابی‌هاست. خرابی‌های تولید نوعا توسط خرابی‌های Stuck-at مدل می‌شوند. سختی آزمون این خرابی‌ها عبارت است از تحریک تمام نقاط داخلی به تمام مقادیر ممکن و مشاهده اثر آن در برخی خروجی‌های اصلی. هر چه تعداد گیت‌ها افزایش می‌یابد، تعداد گره‌هایی که آزمون‌شان سخت می‌شود نیز افزایش می‌یابد. از سوی دیگر، تروجان‌ها به صورت مجموعه‌ای از دروازه‌هایی که به صورت هوشمندانه وارد شده‌اند و تحت شرایط نادری تحریک می‌شوند و خراب‌کاری را ظاهر می‌سازند، مدل می‌شوند. تعداد مدارات تروجان ممکن از یک نوع و اندازه خاص تابعی نمایی‌ست از تعداد گره‌های مدار. همچنین، برای تروجان‌های ترتیبی که برای فعال شدن، نیاز به چندین اتفاق نادر دارند؛ مشاهده خراب‌کاری‌شان در خروجی در زمان آزمون، احتمال بسیار کمی دارد. نهایتا، از آن‌جایی که تعداد تروجان‌های ممکن بی‌شمار است، روش‌های مرسوم برای تخمین میزان پوشش کشف خرابی به‌سادگی برای پوشش کشف تروجان به‌کار نمی‌روند.

به‌دلیل مشکلات کشف تروجان توسط آزمون عملکردی و تفاوت آن با روش آزمون خرابی، روش‌های آماری برای تولید بردار مناسب‌تر هستند. یک روش مبتنی بر تصادفی کردن برای مقایسه احتمالاتی عملکرد مدار بعد از پیاده‌سازی با طراحی طلایی (سالم) اصلی در [۱۸] توصیف شده است. تکنیک استفاده شده از بررسی برابری احتمالاتی[33] اقتباس شده است، جایی که هر خروجی مدار، برای ایجاد مقدار منطق "۱" در گره خروجی با احتمال از پیش تعیین شده، می-تواند به یک توزیع احتمال یکتا در مخروط ورودی مرتبط شود. این توزیع برای تولید بردارهای ورودی تصادفی جهت آزمون مدار مورد بررسی و مقایسه خروجی‌های آن با خروجی‌های طرح طلایی استفاده شده است. هرگونه عدم تطبیق در عملکرد مدار تولید شده برای شناسایی مدار تروجان استفاده می‌شود و بردار ورودی که موجب این عدم تطبیق شده است به عنوان اثرانگشت این نوع خاص تروجان تعریف می‌شود. این تکنیک برای تولید بردارهای آزمون، هیچ مدل خاصی از تروجان را در نظر نمی‌گیرد و بر اساس پیدا کردن برابری بین دو مدار است.

با توجه به نکات ذکر شده و سختی تحریک تروجان در مدارات پیچیده با تولید برادرهای آزمون و آزمون منطقی، این روش بیشتر

برای کمک به روش‌هایی همچون تحلیل کانال جانبی به کار می‌رود. بدین صورت که سعی می‌شود تا بردارهایی تولید گردد تا گذارهای مدار تروجان را تا حد امکان افزایش داده و گذارهای مدارهای غیرتروجان را تا حد امکان کاهش دهد، تا بدین‌ترتیب امکان کشف تروجان از طریق تحلیل کانال جانبی را افزایش داده و کار این روش‌ها را ساده‌تر و دقیق‌تر نماید [۱۹].

در [۲۰] یک روش تولید بردار آماری برای کشف تروجان با نام MERO معرفی شده است. این روش یک مجموعه بهینه از بردارهای آزمون را تولید می‌کند که بتواند هر گره نادر و کم تحرک در یک مدار را چندین بار به مقدار نادرش تحریک نماید (N بار که N توسط کاربر تعیین می‌شود). متغیرهای ورودی الگوریتم عبارتند از ندرت و تعداد گره‌های محرک برای مدل تروجان و طبیعت تروجان (ترکیبی یا ترتیبی). فعال‌سازی گره‌های نادر به تنهایی در مقایسه با الگوهای کاملا تصادفی، احتمال فعال شدن تروجانی را که توسط یک ترکیب نادر از گره‌های انتخاب شده فعال می‌شود را افزایش می دهد. روش MERO همچنین از مزیت فلیپ‌فلاپ‌های پویش برای مدارات ترتیبی بهره می‌برد که موجب کاهش بیشتر طول آزمون و افزایش پوشش تروجان می‌شود. پوشش نسبی بالاتر بدست آمده توسط MERO ظرفیت استفاده از این روش در روش‌های کشف تروجان توسط تحلیل کانال جانبی را نیز نشان می‌دهد.

- **تحلیل کانال جانبی**

روش‌های تحلیل کانال جانبی بر این واقعیت استوارند که واردسازی بداندیشانه هر المانی در مدار مجتمع باید حضورش را در برخی پارامترهای کانال جانبی از جمله جریان نشتی، جریان منبع ساکن (IDDQ)[۳۴]، رد[۳۵] توان پویا، خصوصیات تاخیر مسیر، تشعشعات الکترومغناطیسی به واسطه فعالیت کلیدزنی[۳۶] و یا در ترکیبی از این‌ها منعکس نماید [۵].

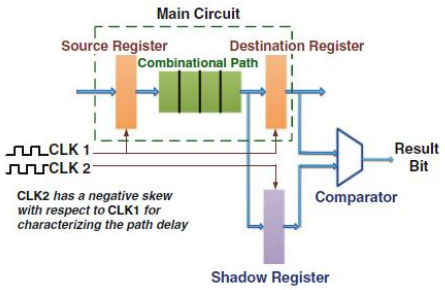
مفهوم تحلیل کانال جانبی برای کشف تروجان از زمینه حملات کانال جانبی که در آن کلیدهای رمز استفاده شده توسط تراشه‌های رمزنگاری با مشاهده یک پارامتر کانال جانبی و برقراری ارتباط بین ردهای موجود برای مقادیر کلیدهای پیش‌بینی شده مختلف تخمین زده می‌شود؛ قرض گرفته شده است.

مزیت مهم استفاده از تحلیل کانال جانبی برای کشف تروجان عدم نیاز این روش به فعال‌سازی کامل تروجان و مشاهده اثر مخرب آن در خروجی سامانه است. همچنین این روش برای کشف تروجان‌هایی که تنها برای نشت اطلاعات سامانه به‌کار رفته‌اند نیز مناسب است در حالی که روش آزمون منطقی قادر به کشف این تروجان‌ها نخواهد بود. همچنین هر چه تروجان بزرگ‌تر باشد کشف آن آسان‌تر خواهد بود چرا که اثر آن بر پارامترهای کانال جانبی واضح‌تر خواهد بود. از آن سو، هر چه طرح اصلی بزرگ‌تر باشد، اثر تروجان راحت‌تر توسط نویز پوشش داده می‌شود.

در عین حال، تغییرات روند بزرگ در فناوری‌های پیش‌رفته نانوتکنولوژی و نویز اندازه‌گیری که به‌خصوص برای تروجان‌های کوچک، می‌تواند اثر مدار تروجان را پوشش دهد، چالش‌های اصلی مرتبط با این روش‌ها هستند.

روش‌های کانال جانبی موجود برای تنظیم روند به سوی روش‌هایی همچون نرمال‌سازی و تنظیم نویز اندازه‌گیری با میانگین‌گیری از چندین اندازه‌گیری برای خلاصی از نویز تصادفی، گرایش دارند. برای رهایی از تغییرات بین قالب[۳۷] و تنظیم سامانه‌ای تغییرات درون قالب[۳۸]، روش‌های مبتنی بر منطقه استفاده می‌شود که در آن اندازه‌گیری‌ها از چندین پایه توان مربوط به فعال‌سازی بخش‌های جداگانه، به مقایسه پارامترهای اندازه‌گیری شده از یک مدار مجتمع تحت شرایط متفاوت، کمک می‌کند.

برای مدارات ترتیبی بزرگ یک روش فعال‌سازی مبتنی بر نواحی مدار برای افزایش حساسیت کشف تروجان توسط تحلیل کانال جانبی مفید است. مدار می‌تواند به نواحی عملکردی مجزا تقسیم‌بندی شود و یا به صورت ساختاری با حداقل هم‌پوشانی بین دروازه‌های هر بخش تقسیم شود. سپس، بردارهای آزمون خاصی با هدف حداکثرسازی گذارها در بخش تحت بررسی و حداقل‌سازی فعالیت در بقیه مدار تولید می‌شود. این کار حساسیت اندازه‌گیری جاری برای کشف تروجان در ناحیه فعال را افزایش می‌دهد. اثرگذاری روشی همچون تولید آزمون مبتنی بر ناحیه در مقایسه با الگوهای آزمون تصادفی در [۱۹] نشان داده شده است.

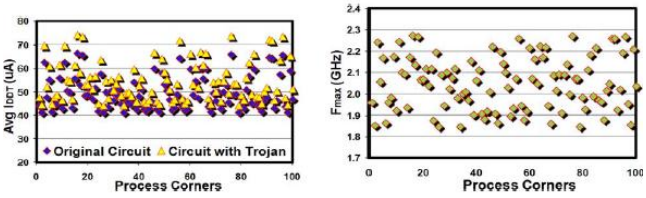


شکل ۴ - روش توصیف تاخیر در سرعت برای کشف هر گونه تغییر تاخیر به دلیل وجود تروجان [۲۱]

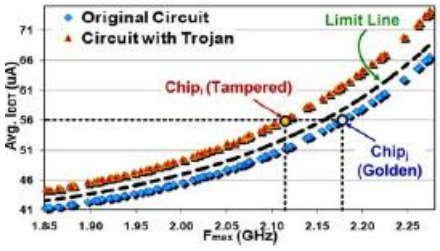
تاخیرهای مسیر اندازه‌گیری شده در پایه‌های خروجی برای یک مجموعه از بردارهای آزمون می‌تواند برای تعیین وجود تروجان مورد استفاده قرار گیرد که بر اساس مسیر اندازه‌گیری شده و تغییر تاخیر آن استوار است. تکنیک توصیف تاخیر در سرعت [۲۱]، ثبات‌های سایه را معرفی می‌کند که همراه با مقایسه‌گرها برای تعیین لختی تاخیر مسیر برای تاخیرهای ثبات به ثبات به کار می روند. این تکنیک DfS از افزایش انحراف[۳۹] منفی در ساعت ثبات سایه در مقایسه با ساعت عملیاتی و مقایسه خروجی ذخیره شده برای یک سری از ورودی‌ها به منظور تعیین توزیع تاخیر استفاده می‌کند. هر تروجان سخت‌افزاری که تاخیرهای اندازه‌گیری شده را افزایش دهد کشف خواهد شد. این

تکنیک نیاز دارد تا با روش‌های تولید بردار آزمون و مدل تروجان مناسب ترکیب شود تا پوشش بالایی را در حداقل زمان آزمون به‌دست آورد. همچنین این روش می‌تواند برای کشف تروجان‌های سخت‌افزاری خاموش که تنها در زمان به کارگیری فعال می‌شوند، در زمان اجرا مورد استفاده قرار گیرد. در [۲۱] نشان داده شده است که این روش قادر به کشف تروجان‌ها در مدار یک ضرب‌کننده آرایه‌ای ۸×۸ با تغییرات روند زیر ۲۰٪± است.

به منظور افزایش دقت کشف در برابر تغییرات روند و نویز اندازه‌گیری روش‌های تحلیل کانال جانبی با استفاده از چند پارامتر کانال جانبی می‌توانند مورد استفاده قرار گیرند. در [۲۲] از ارتباط ذاتی بین جریان گذرا منبع (IDDT)[۴۰] و حداکثر فرکانس اجرا (F_{max}) یک مدار برای کشف تروجان در آن استفاده شده است. روش ارائه شده شامل یک تحلیل تئوری است در ارتباط بین این دو پارامتر و هیچ نیازی به تغییر روال طراحی ندارد و هیچ‌گونه سربار سخت‌افزاری نیز تحمیل نمی‌کند. هر گونه تغییراتی در IDDT که خارج از روال عادی با توجه به تغییرات روند باشد می‌تواند برای شناسایی وجود تروجان مورد استفاده قرار گیرد. این روش همچنین با یک روش تولید بردارهای آزمون مناسب برای تقسیم بندی مدار به نواحی کوچک‌تر جهت کاهش جریان پس‌زمینه و همچنین با روش MERO برای افزایش فعالیت کلیدزنی در مدار احتمالا کم فعالیت تروجان همراهی می‌شود.



شکل ۵ - اندازه‌گیری دو پارمتر کانال جانبی برای دو مدار سالم و آلوده به تروجان به صورت جداگانه در حضور تغییرات روند [۲۲]



شکل ۶ - اندازه‌گیری پارامترهای کانال جانبی و تحلیل ترکیبی آنها برای دو مدار سالم و آلوده به تروجان در حضور تغییرات روند [۲۲]

در روش‌هایی که تاکنون برای تحلیل کانال جانبی ارائه شد، عمدتا به مجموعه‌ای از مدارات مجتمع طلایی به منظور مقایسه نتایج اندازه‌گیری شده و مدرج‌سازی[۴۱] نویز روند و نتیجه‌گیری در مورد وجود یا عدم وجود تروجان نیاز خواهد بود. این نیاز و ضرورت یکی از ایرادات این روش‌ها محسوب می‌شود چرا که در مواردی ممکن است تمام مدارات مجتمع موجود یکسان باشند و نسخه طلایی از آنها وجود نداشته باشد. یک راه‌کار ارائه شده برای این مشکل تحلیل پارامترهای کانال جانبی مجموعه‌ی محدودی از مدارات مجتمع به روش تخریبی و سپس مقایسه دیگر مدارت مجتمع با این مدارات است [۱۱]. یک روش دیگر ارائه شده استفاده از حسگرهای تعبیه شده در زمان طراحی در مدار است [۲۳] که توسط آنها بعد از تولید تراشه می‌توان به سالم بودن تراشه یا آلوده شدن آن پی برد. روش دیگر اندازه‌گیری و مقایسه امضای جریان در پنجره‌های زمانی مختلف از یک مدار مجتمع و مقایسه آنها با یکدیگر جهت حذف اثر نویز روند و تحلیل وجود تروجان است [۲۴].

۳-۲-۲- روش‌های زمان اجرا

کشف جامع و کامل همه انواع تروجان‌های سخت‌افزاری با اندازه‌های متفاوت در زمان آزمون و تصدیق پس از تولید ممکن است دست‌نیافتنی باشد. نظارت برخط بر محاسبات بحرانی می‌تواند سطح اعتماد را با توجه به حملات تروجان سخت‌افزاری به مقدار قابل توجهی افزایش دهد [۵]. روش‌های زمان اجرا برای کشف تروجان‌های سخت‌افزاری مبتنی بر نظارت بر اجرای محاسبات بحرانی هستند تا رفتارهای مخرب خاصی که موجب تحریک تروجان در زمان کار بلند مدت سامانه می‌شود را شناسایی کنند.

روش‌های زمان اجرا نوعا روش‌های تهاجمی هستند که از تکنیک‌های طراحی برای امنیت (DfS) بهره می‌برند [۵]. این روش‌ها برای متوقف ساختن بخش آلوده مدار می‌توانند از افزونگی از پیش تعبیه شده به صورت هسته‌های قابل بازبرنامه‌ریزی استفاده نمایند و بدین ترتیب کار بخش آلوده متوقف شده و در قرنطینه قرار می‌گیرد و کار آن بخش توسط افزونگی تعبیه شده به‌عهده گرفته می‌شود. بنابراین، اطمینان‌پذیری کار مدار حتی در حضور تروجان تضمین شده خواهد بود حتی اگر تراشه‌های آلوده به تروجان دور انداخته نشده باشند.

تکنیک‌های مختلفی برای تامین امنیت سامانه در این سطح ارائه شده است. در [۲۵] یک روش کشف تروجان سخت‌افزاری را شرح داده شده است که از نرم‌افزار و سخت‌افزار بهره می‌گیرد. راهبرد ارائه شده به بررسی رخداد دو نوع حمله می‌پردازد. حمله اول حمله تکذیب سرویس (DoS) است که با استفاده از یک گارد سخت‌افزاری کوچک شخصی‌سازی[۴۲] شده که بر روی مسیر بین حافظه و پردازنده قرار می‌گیرد؛ تشخیص داده می‌شود. سیستم عامل در بازه‌های زمانی مشخص به گارد سخت‌افزاری سیگنال‌های بررسی حیات[۴۳] متناوب ارسال می‌نماید. هرگونه خرابی در ارسال به‌موقع این سیگنال به گارد به عنوان کشف موفقیت‌آمیز یک حمله DoS تلقی می‌شود. حمله دومی که روش ارائه شده قادر به شناسایی آن است حمله‌ی ترکیبی سخت‌افزاری و نرم‌افزاری بالابردن سطح دسترسی است که در آن تروجان سخت‌افزاری حفاظت حافظه را غیرفعال می‌کند و نرم‌افزار تبانی کننده

را بر افزایش سطح دسترسی قادر می‌سازد. این حمله با بررسی امکان دسترسی یک نرم‌افزار با عدم مجوز دسترسی به بخش‌های حفاظت شده حافظه صورت می‌پذیرد. در این روش از یک سیستم عامل مبتنی بر یونیکس تغییر یافته استفاده شده است تا سیستم عامل بتواند با گارد سخت‌افزاری تعبیه شده هم‌کاری نماید.

آبرامویکی و بریدلی در [۲۶] برای پیاده سازی ناظران امنیتی بلادرنگ، منطق قابل پیکربندی DEFENCE[۴۴] را به طراحی عملکردی اضافه نموده‌اند. پس از ساخته شدن مدارات مجتمع منطق قابل پیکربندی با بایسته‌های رفتاری و روش کاری دستگاه برنامه‌ریزی می‌شود. این بررسی‌ها می‌توانند همروند با کار عادی اجرا شوند و هر گونه تغییرات و نوسانات از هنجارهای تعریف شده را کشف نمایند.

در [۲۷] یک روش مرکب سخت‌افزار/نرم‌افزاری با عنوان BlueChip ارائه شده است که شامل یک بخش زمان طراحی و نظارت زمان اجراست. روش ارائه شده تلاش می‌کند تا هرگونه مداربندی بلااستفاده را توسط آزمون‌های تصدیق طراحی شناسایی نموده و به عنوان مشکوک برچسب گذاری نماید. در زمان اجرا مداربندی مشکوک حذف شده و با منطق استثناء[۴۵] که می‌تواند یک استثناء نرم‌افزاری را تحریک کند، جایگزین می‌شود. این مکانیزم به سامانه اجازه می‌دهد تا با انحراف[۴۶] در اطراف تروجان‌های سخت‌افزاری مخرب به کار عادی خود ادامه دهد. این تکنیک برای به دام انداختن تروجان‌های سخت-افزاری طراحی شده است که در هدف مشابه تروجان‌های نرم‌افزاری‌اند. تروجان‌هایی با عملکردهایی از جمله ارتقای سطح دسترسی، دسترسی به بخش‌های محدود و محافظت شده حافظه و غیره که می‌توانند در یک مالکیت معنوی سخت‌افزاری نگاشت شده به FPGA یا به کدهای دستورالعمل اجرا شونده بر روی یک پردازنده تعبیه شده، وارد شود.

در شرایط پردازنده‌های چند هسته‌ای، یک طرح زمان‌بندی داخلی[۴۷] در زمان اجرا می‌تواند پیاده‌سازی شود [۲۸] که در آن نسخه-های نرم‌افزاری یکسان عملکردی بر روی چند هسته پردازنده اجرا می-شوند و توسط زمان‌بندی نرم‌افزاری توزیع شده پویا همراهی می‌شوند. خروجی‌های زیرکارها از هسته‌های مختلف برای ارزیابی سطح اعتماد هسته‌ها به صورت فردی با هم مقایسه می‌شوند که توسط یک روال یادگیرنده در طی چند مرتبه اجرا صورت می‌پذیرد. این طرح قادر است با ارتقای گذردهی در اجراهای متوالی، کارها را در یک محیط آلوده به تروجان با موفقیت کامل نماید. رویکرد این روش مشابه مفهوم محاسبات دارای تحمل در برابر خطاست، که در آن هسته‌های خراب یا با اطمینان‌پذیری کم در طول زمان اجرا، بدون کاهش بهره‌وری تولید دور زده[۴۸] می‌شوند.

در [۲۹] چند راه‌کار برای کشف تروجان و همچنین تامین امنیت سامانه در حضور تروجان ارائه شده است. ابتدا آنها از یک رمزگشای آدرس امن استفاده می‌کنند که در آن مجوز دسترسی ماژول‌های مختلف به آدرس‌های درخواستی‌شان نیز بررسی می‌شود. در صورت درخواست دسترسی یک ماژول به آدرس غیرمجاز سیگنال دسترسی غیرمجاز فعال شده و جز یک ماژول پیش فرض، تمام ماژول‌های کارگر[۴۹] غیرفعال می‌شوند و شناسه ماژول کارفرمای[۵۰] بداندیش در لیست پوشش ماژول‌های بداندیش ذخیره می‌شود. همچنین رمزگشای آدرس امن از دسترسی ماژول‌های سالم به یک ماژول آلوده نیز جلوگیری می‌کند. برای تشخیص تروجان‌های سخت‌افزاری عامل حملات DoS آنها یک مدیر خطوط ارتباطی (داور) در نظر گرفته‌اند که به هر ارتباط زمان مجازی را اختصاص می‌دهد و در صورتی که زمان در اختیار گرفتن خطوط ارتباط توسط یک تراکنش از حد مجاز بیشتر شد سیگنال کشف قفل مخرب (MLD)[۵۱] فعال شده و دسترسی انحصاری کارفرما به خطوط ارتباطی را لغو نموده و ماژول کارفرمای خاطی را در لیست ماژول‌های مشکوک قرار می‌دهد. برای پیشگیری از شنود اطلاعات توسط تروجان، معماری خاصی برای مدیریت خطوط ارتباطی مشترک ارائه شده است که در این معماری تنها فرستنده و گیرنده اصلی اطلاعات قادر به دسترسی به خطوط ارتباطی هستند. در ادامه به راه‌کارهای پس از کشف یک ماژول مخرب و جایگزینی عملکرد آن توسط مدارهای افزونه یا قابل برنامه‌ریزی و پیکربندی مجدد پرداخته شده است.

مطالب این بخش نزدیکی زیادی با بخش مقابله با تروجان‌های سخت‌افزاری دارد، چراکه بیشتر روش‌های کشف در زمان اجرا به گونه-ای طراحی شده‌اند که پس از شناسایی وجود تروجان و رخداد عمل مشکوک توسط یک ماژول به مقابله با آن خواهند پرداخت.

۴- مقابله با تروجان‌های سخت‌افزاری

ترکیب آخرین روش‌های پیش‌گیری از تروجان‌های سخت‌افزاری و مکانیزم‌های کشف پیش از استقرار هنوز نمی‌توانند یقین کامل بر عاری از تروجان بودن مدارات مجتمع تولید شده یا طراحی‌های منطق قابل پیکربندی مجدد، فراهم کنند. با توجه به تعداد و انواع زیاد تهدیدها و فضای حالت گسترده محرک‌های تروجان‌ها، برخی پژوهشگران بر روی مسئله تامین امنیت اجرا در حضور تروجان‌های سخت‌افزاری تمرکز نموده‌اند.

روش‌های مقابله موفق باید به سخت‌افزار اجازه دهند تا به تروجان‌های وارد شده بی‌توجه باشد و حتی اجازه استفاده از بخش‌های تجاری عام منظوره (COTS) در ساخت سامانه محاسباتی قابل اعتماد و مقاوم در برابر تروجان را بدهند. تاکنون یک روش کلی ارائه یا توسعه داده نشده است که بتواند اجازه دهد که یک مدار مجتمع در یک روش قابل اعتماد در حضور یک تروجان دلخواه بتواند به کارش ادامه دهد [۶]. در این بخش به روش‌های مقابله با تروجان‌های سخت‌افزاری خواهیم پرداخت.

- **طراحی برای امنیت (DfS)**

تکنیک‌های DfS می‌توانند با بهبود قابلیت کنترل و مشاهده-پذیری نقاط ممکن برای محرک و بار تروجان، قابلیت آزمون مدارات

بزرگ را افزایش دهند. با افزودن یک ماشین حالت متناهی (FSM)[۵۲] که به‌سختی قابل شناسایی است، ماژول‌های مختلف در یک طراحی پیچیده می‌توانند در حالت‌های شفاف تعریف شده به صورت یکتا به صورت انتخابی مورد آزمون قرار گیرند. با استفاده از سری خاصی از ورودی‌ها، هر ماژول به مد شفاف ویژه‌ای وارد می‌شود [۳۰]، جایی که تمام ماژول‌های دیگر کنارگذاشته می‌شوند. گره‌های داخلی مدار در ورودی ماژول از ورودی‌های اصلی سامانه قابل آزمون‌اند و یک امضای فشرده شده از مقادیر گره خروجی در خروجی‌های اصلی ارائه می‌گردد، که نشان‌دهنده‌ی وجود یا عدم وجود تروجان است.

- **طراحی مبهم**

یک طراحی مبتنی بر ابهام [۳۱] می‌تواند از مهندسی معکوس موفقیت‌آمیز عملکرد هسته‌های مالکیت معنوی درون یک مدار مجتمع توسط حمله‌کننده پیش‌گیری نماید. در نتیجه حمله‌کننده فرض‌های اشتباهی را در مورد پایین بودن احتمال گذار نقاط داخلی مدار به هنگام وارد سازی تروجان در نظر می‌گیرد که موجب می‌شود یا تروجان در مکانی قرار گیرد که به سادگی تحریک و کشف شود و یا در جایی قرار گیرد که هرگز فعال نشود (تروجان خوش‌خیم) که در هر دو صورت هدف حمله‌کننده انجام نشده است.

- **حفاظ‌های[۵۳] داده**

با حفاظ‌بندی داده (از جمله دستورالعمل‌های پردازنده) طراح تلاش می‌کند تا از تحریک شدن تروجان یا دسترسی مستقیم به اطلاعات حساس رمزنشده و بهره‌برداری از آن پیش‌گیری کند. یک حفاظ می‌تواند چگونگی شکل داده‌ها در زمان ذخیره شدن و انتقال داخل یا بین مدارات مجتمع یا ماژول‌های منطقی را کنترل کند؛ که بر روش تعامل تروجان با داده اثر می‌گذارد.

درهم آمیختن[۵۴] خطوط داده یکی از تکنیک‌های مورد استفاده است که برای جلوگیری از رسیدن کد فعال‌سازی به تروجان توسط خطوط داده استفاده می‌شود. این روش برای بخش‌های غیرمحاسباتی که با داده‌ها سر و کار دارند، استفاده می‌شود. یکی از روش‌های ممکن برای این‌کار XOR کردن داده‌های روی خطوط ارتباطی با یک عدد شبه تصادفی و سپس ارسال هر دو این داده‌ها بر روی خطوط ارتباطی است [۳۲]. البته باید توجه داشت که سخت‌افزار پیاده کننده‌ی این روش خود باید مورد اعتماد باشد. یک روش کنترل‌شده‌تر می‌تواند نگاشت تمام ورودی‌ها به یک فضای حالت کاملا تایید شده از لحاظ عملکردی باشد. برای پیش‌گیری از نشت اطلاعات سامانه می‌توان از روش‌های رمزنگاری مناسب‌تر و قوی‌تری استفاده نمود.

برای بخش‌های محاسباتی درون مدار مجتمع مبهم‌سازی داده‌ها بر توانایی واحد محاسباتی در تولید نتایج صحیح با ورودی‌های درهم‌آمیخته اثر می‌گذارد. در [۳۲] یک روش رمزنگاری هومومورفیک مورد استفاده قرار گرفته است که اجازه می‌دهد واحدهای محاسباتی بر روی داده‌های رمزشده کاری را که بر روی ورودی‌های عادی انجام می‌دادند انجام دهند و نتایج قابل رمزگشایی و صحیحی تولید نمایند. هر مداری می‌تواند توسط توابع هومومورفیک مبهم شود اما هزینه‌ی بالایی خواهد داشت که عملا ناکارآمد و غیرقابل چشم‌پوشی خواهد بود.

روش دیگر [۲۹] عدم اجازه دسترسی به ماژول‌هایی‌ست که در یک تراکنش دخیل نیستند که در بخش روش‌های زمان اجرا توضیح داده شد.

- **حفاظ زمانی و ترتیبی**

همچنان که در بخش معرفی تروجان‌های سخت‌افزاری بیان شد برخی تروجان‌ها به صورت بمب زمانی عمل می‌کنند و در زمان خاص یا گذر مدت مشخص از کار سامانه فعال می‌شوند. در [۳۲] یک روش با عنوان حفاظ زمان برای پیش‌گیری از فعالیت چنین تروجان‌هایی در فضای حالت عملکردی تایید شده، ارائه شده است. در این روش یک مدار مجتمع به صورت کامل برای تعداد سیکل‌های مشخصی از لحاظ عملکردی تایید شده است و پس از آن به صورت دوره‌ای خاموش و روشن می‌شود برای اطمینان از این‌که هیچ تروجان مبتنی بر زمان-سنجی نمی‌تواند فعال شود. یک مکانیزم سبک ذخیره‌سازی سابقه برای اطمینان از پیوستگی پردازش مورد استفاده قرار گرفته است. علت طرح این ایده این است که هر تروجان سخت‌افزاری که در تمام فضای حالت آزمون (یک فضای حالت متشکل از زمان و ورودی‌ها) ساکت و خاموش بوده است، در زمان مشابه در طی شرایط کاری‌اش نیز خاموش خواهد ماند. به طور واضح و مشخص این روش در برابر انواع دیگر تروجان‌های سخت‌افزاری کارایی نخواهد داشت و تنها برای این نوع از تروجان پیشنهاد شده است.

حفاظ دیگر معرفی شده برای خنثی‌سازی تروجان‌های ترتیبی که بر اساس یک FSM عمل می‌کنند، پیشنهاد شده است. در این حفاظ ترتیب کارها و وقایع در صف، تا حد امکان به صورت تصادفی تغییر داده می‌شود و یا برخی وقایع ساختگی و زائد در درون سری ورودی-های ماژول‌های مختلف قرار داده می‌شود تا مانع از رخداد یک سری از وقایع خاص شود که ممکن است موجب فعال‌سازی تروجان گردد.

- **تکرار[۵۵]، پراکندگی[۵۶] و رای‌گیری**

یک روش مقابله عمومی در برابر تروجان سخت‌افزاری ممکن است از مواردی همچون: تکرار یا افزونگی منطق و/یا داده‌ها، تقسیم یا پراکندگی منطق و/یا داده‌ها، توزیع منطق و/یا داده‌ها و جمع‌آوری و ترکیب منطق و/یا داده‌ها استفاده کند.

تاثیر این روش‌های مقابله در سه جبهه ظاهر می‌شود: حفاظت در برابر تروجان‌های سخت‌افزاری که اطلاعات حساس سامانه را نشت می-دهند با تقسیم داده‌ها و پردازش آنها با المان‌های منطقی مستقل؛ حفاظت در برابر تغییرات خصوصیات و عملکرد المان‌ها با استفاده از

چندین نسخه از منطق؛ و حفاظت در برابر حملات DoS با فراهم آوردن افزونگی در عمل المان‌های منطقی در طراحی.

در [۳۲] استفاده از چندین نسخه از منطق را پیشنهاد داده شده است؛ بدین صورت که از چند نسخه‌ی یک ماژول یا مدار مجتمع غیرقابل اعتماد که توسط چند طراح مختلف تولید شده‌اند استفاده شود. خروجی ماژول‌ها در هر سیکل ساعت می‌تواند مقایسه شود و خروجی صحیح با رای‌گیری مشخص شود. اما این روش هزینه بالایی را بر سطح سیلیکون مورد نیاز و توان مصرفی وارد می‌سازد.

روش ارائه شده در [۲۸] که در بخش کشف در زمان اجرا نیز مورد بررسی قرار گرفت؛ از یک ایده برای مقابله با بخش‌های مشکوک به تروجان استفاده می‌کند و آن عبارت‌است از این‌که با توجه به سطح اهمیت هر کار، تلاش بر ارجاع کمتر کارهای وارد شونده به صف پردازش به بخش‌های مشکوک خواهد بود. در این کار، نویسندگان پیشنهاد استفاده از یک سامانه پردازش چند هسته‌ای را برای بهره-برداری از مزایای افزونگی داخلی داده‌اند. در این شرایط این امکان وجود خواهد داشت تا هسته‌هایی که غیرقابل اعتماد بنظر می‌رسند را کنار گذاشت. پردازه‌های از لحاظ عملیاتی برابر اما متفاوت، در چندین المان پردازشی کاشته می‌شوند و نتایج آنها مقایسه می‌شود. ممکن است با توجه به الگوریتم‌ها و نحوه پیاده‌سازی‌ها، پردازش در هر المان متفاوت باشد، اما نتایج باید یکسان باشند. اگر تفاوتی در نتیجه وجود داشت المان پردازشی سوم برای ایجاد امکان رای‌گیری، وارد خواهد شد و این روال ادامه خواهد یافت تا زمانی که حداقل دو المان نتیجه محاسباتشان یکسان باشد. سپس المان یا المان‌هایی که نتیجه اشتباه تولید نموده‌اند به صورت پویا تنبیه خواهند شد؛ برای مثال اعتماد به این المان‌ها کاهش خواهد یافت و احتمال استفاده از آنها نیز کاهش خواهد یافت.

۵- طرح پیشنهادی

با مطالعات انجام شده در حوزه امنیت سخت‌افزار و کشف تروجان‌های سخت‌افزاری نکته‌ای که به وضوح می‌توان یافت پیچیدگی کار کشف تروجان در مراحل پس از ساخت در تراشه‌های پیچیده و بزرگ، با توجه به اندازه‌های ممکن بسیار کوچک و انواع مختلف برای تروجان‌ها و امکان مخفی ماندن تروجان‌های هوشمندانه طراحی شده، برای پیش‌گیری از کشف‌شان در مراحل بازرسی است.

با توجه به این مشکلات به‌نظر می‌رسد راه‌کار مناسب برای کنترل آسیب‌پذیری‌های احتمالی و پیش‌گیری از حملات مبتنی بر سخت‌افزار و تروجان‌های سخت‌افزاری در سامانه استفاده از سامانه نظارت سخت-افزاری است که با توجه به شرایط کاری یک سامانه روال عادی کاری آن را می‌تواند استخراج نماید و یا از طریق تعریف قواعد توسط کاربر از حیطه‌های مجاز کاری سامانه مطلع شود و سپس با این اطلاعات بر عملکرد سامانه نظارت نموده و رخداد هرگونه عملکرد خلاف قاعده را به عنوان یک شرایط غیرعادی و احتمال حمله شناسایی نموده و هشدار دهد و حتی توسط مکانیزم‌های امنیتی طراحی شده این رخداد را کنترل نموده و از آسیب رسیدن به روند عادی سامانه پیش‌گیری نماید.

مفهوم پایه‌ای استفاده از واحد سخت‌افزاری یا کمک‌پردازنده[۵۷] برای دستیابی به اجرای امن ریشه‌اش بر می‌گردد به پردازنده‌های رمز[۵۸] مقاوم در برابر مداخله[۵۹]، که برای ذخیره کلیدها و اجرای الگوریتم‌های رمزنگاری استفاده می‌شود [۳۳].

در [۳۴] از یک پردازنده کمکی برای تشخیص نفوذ به سامانه استفاده شده است. کشف نفوذ با استفاده از یک سخت‌افزار اختصاصی که از تدابیر کشف ناهنجاری[۶۰] در سطح پایین استفاده می‌کند، تاثیر چشم‌گیری در ارتقا و بهبود زمان پاسخ و اطمینان‌پذیری در برابر سامانه‌های تشخیص نفوذ نرم‌افزاری خواهد داشت.

یکی از رویکردهایی که سامانه‌های کشف نفوذ بر اساس آن کار می‌کنند عبارت است از کشف نفوذ بر اساس جستجوی ناهنجاری؛ که در آن سامانه کشف نفوذ با اطلاع از شرایط نرمال، هر گونه تغییراتی از این شرایط را به عنوان رخداد نفوذ در نظر می‌گیرد. مسئله اساسی در اینجا تعیین شرایط نرمال و مولفه‌های مورد استفاده برای تعریف این شرایط است.

پیدا کردن مولفه‌های قابل اندازه‌گیری و مناسب برای تعریف شرایط عادی یک سامانه سخت‌افزاری می‌تواند زمینه مناسبی برای طراحی یک سامانه تشخیص نفوذ سخت‌افزاری برای سخت‌افزارهای مختلف بسته به کاربردشان باشد.

۶- نتیجه گیری

تروجان سخت‌افزاری به عنوان یک تهدید امنیتی برای سامانه‌های محاسباتی در چند سال گذشته مطرح شده‌اند و روش‌های متعددی برای کشف‌شان ارائه شده است. در این پژوهش ما به بررسی ماهیت تروجان‌های سخت‌افزاری و روش‌های موجود برای کشف این تهدیدات امنیتی پرداختیم. همچنین به دلایل گرایش پژوهش‌گران به روش‌های کشف تروجان در زمان اجرا اشاره کردیم و راه‌کارها و راه‌بردهای موجود برای این روش را مرور کردیم. در نهایت به ارائه روش پیشنهادی برای دست‌یابی به سامانه‌هایی امن علی رغم حضور المان-های مشکوک به تروجان پرداختیم.

۷- منابع

[۱] F. Koushanfar and M. Potkonjak, "Hardware Security: Preparing Students for the Next Design Frontier," *Microelectronic Systems Education, MSE'۰۷. IEEE International Conference on*, ۲۰۰۷.

[۲] N. Potlapally, "Hardware security in practice: Challenges and opportunities," *Hardware-Oriented Security and Trust (HOST) IEEE International Symposium on*, ۲۰۱۱.

[۳] S. Adee, "The hunt for the kill switch," *Spectrum, IEEE ۴۵,۵*, pp. ۳۴-۳۹, ۲۰۰۸.

[۴] M. Bilzor, T. Huffmire, C. Irvine, and T. Levin, "Security Checkers: Detecting processor malicious inclusions at runtime," *Hardware-Oriented Security and Trust (HOST), IEEE International Symposium on*, pp. ۳۴ - ۳۹, ۲۰۱۱.

[۵] M. Tehranipoor and C. Wang, *Introduction to Hardware Security and Trust*.: Springer, ۲۰۱۱.

[۶] M. Beaumont, B. Hopkins, and T. Newby, "Hardware Trojans – Prevention, Detection, Countermeasures," *Command, Control, Communications and Intelligence Division Defence Science and Technology Organisation, Australian Government*, ۲۰۱۱.

[۷] S. S. Kumar, J. Guajardo, R. Maes, G. J. Schrijen, and P. Tuyls, "The butterfly PUF protecting IP on every FPGA," *Hardware-Oriented Security and Trust, HOST. IEEE International Workshop on. IEEE*, ۲۰۰۸.

[۸] S. Yun, Q. Li, H. Gao, and Z. Ping, "Towards Hardware Trojan: Problem Analysis and Trojan Simulation," *Information Engineering and Computer Science (ICIECS), ۲nd International Conference on*, pp. ۱- ۴, ۲۰۱۰.

[۹] Y. Alkabani and F. Koushanfar, "Extended abstract: Designer's hardware Trojan horse," *Hardware-Oriented Security and Trust, HOST. IEEE International Workshop on*, pp. ۸۲ - ۸۳, ۲۰۰۸.

[۱۰] D. Gomez. (۲۰۱۰) Intel to introduce processor with remote kill switch. [Online]. http://www.tgdaily.com/opinion-features/۵۳۱۰۸-intel-to-introduce-processor-with-remote-kill-switch

[۱۱] L. W. Wang and H. W. Luo, "A power analysis based approach to detect Trojan circuits," *Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR۲MSE), International Conference on*, pp. ۳۸۰ - ۳۸۴, ۲۰۱۱.

[۱۲] R. S. Chakraborty, S. Narasimhan, and S. Bhunia, "Hardware Trojan: Threats and emerging solutions," *High Level Design Validation and Test Workshop, HLDVT. IEEE International*, pp. ۱۶۶- ۱۷۱, ۲۰۰۹.

[۱۳] X. Wang, M. Tehranipoor, and J. Plusquellic, "Detecting malicious inclusions in secure hardware: Challenges and solutions," *IEEE International Workshop on Hardware-Oriented*, pp. ۱۵-۱۹, ۲۰۰۸.

[۱۴] J. Rajendran, E. Gavas, J. Jimenez, V. Padman, and R. Karri, "Towards a comprehensive and systematic classification of hardware Trojans," *Circuits and Systems (ISCAS), IEEE International Symposium on*, pp. ۱۸۷۱- ۱۸۷۴, ۲۰۱۰.

[۱۵] Y. Jin and Y. Makris, "Hardware Trojans in Wireless Cryptographic Integrated Circuits," *Design & Test of Computers, IEEE*, p. ۱, ۲۰۰۹.

[۱۶] L. Lin, M. Kasper, T. Güneysu, Ch. Paar, and W. Burleson, "Trojan Side-Channels: Lightweight Hardware Trojans through Side-Channel Engineering," *Cryptographic Hardware and Embedded Systems - CHES. Lecture Notes in Computer Science*, vol. ۵۷۴۷, pp. ۳۸۲-۳۹۵, ۲۰۰۹.

[۱۷] Inc. Chipworks. (۲۰۱۱) Semiconductor Manufacturing – Reverse Engineering of Semiconductor components, parts and process. [Online]. http://www.chipworks.com

[۱۸] S. Jha and S. K. Jha, "Randomization based probabilistic approach to detect trojan circuits," *High Assurance Systems Engineering Symposium, HASE. ۱۱th IEEE*, ۲۰۰۸.

[۱۹] M. Banga and M. S. Hsiao, "A region based approach for the identification of hardware Trojans," *Hardware-Oriented Security and Trust (HOST) IEEE International Workshop on*, ۲۰۰۸.

[۲۰] R. S. Chakraborty, F. Wolff, S. Paul, C. Papachristou, and S. Bhunia, "MERO: a statistical approach for hardware Trojan detection," *Cryptographic Hardware and Embedded Systems-CHES* , pp. ۳۹۶-۴۱۰, ۲۰۰۹.

[۲۱] J. Li and J. Lach, "At-speed delay characterization for IC authentication and Trojan horse detection," *Hardware-Oriented Security and Trust, HOST. IEEE International Workshop on. IEEE*, ۲۰۰۸.

[۲۲] S. Narasimhan, D. Du, R. S. Chakraborty, S. Paul, and F. Wolff, "Multiple-parameter side-channel analysis: a non-invasive hardware Trojan detection approach," *Hardware-Oriented Security and Trust (HOST). IEEE International Symposium on.*, ۲۰۱۰.

[۲۳] M. Li, A. Davoodi, and M. Tehranipoor, "A sensor-assisted self-authentication framework for hardware trojan detection," *Design, Automation & Test in Europe Conference & Exhibition (DATE), IEEE*, ۲۰۱۲.

[۲۴] S. Narasimhan, X. Wang, D. Du, R. S. Chakraborty, and S. Bhunia, "TeSR: A robust temporal self-referencing approach for hardware trojan detection," *Hardware-Oriented Security and Trust (HOST), IEEE International Symposium on.*, ۲۰۱۱.

[۲۵] G. Bloom, R. Simha, and B Narahari, "OS support for detecting Trojan circuit attacks," *Hardware-Oriented Security and Trust, IEEE International Workshop on*, pp. ۱۰۰ - ۱۰۳, ۲۰۰۹.

[۲۶] M. Abramovici and P. Bradley, "Integrated circuit security: new threats and solutions," *Proceedings of the ۵th Annual Workshop on Cyber Security and*

Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies. ACM, ۲۰۰۹.

[۲۷] H. Matthew, M. Finnicum, S. T. King, M. M. Martin, and J. M. Smith, "Overcoming an untrusted computing base: Detecting and removing malicious hardware automatically," *In Security and Privacy (SP), IEEE Symposium on.*, pp. ۱۵۹-۱۷۲, ۲۰۱۰.

[۲۸] D. Mclntyre, F. Wolff, C. Papachristou, S. Bhunia, and D. Weyer, "Dynamic evaluation of hardware trust," *Hardware-Oriented Security and Trust, HOST. IEEE International Workshop on*, pp. ۱۰۸-۱۱۱, ۲۰۰۹.

[۲۹] L. Kim and J. D. Villasenor, "A system-on-chip bus architecture for thwarting integrated circuit trojan horses," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on ۱۹,۱۰*, pp. ۱۹۲۱-۱۹۲۶, ۲۰۱۱.

[۳۰] R. S. Chakraborty, S. Paul, and S. Bhunia, "On-demand transparency for improving hardware trojan detectability," *Hardware-Oriented Security and Trust, HOST. IEEE International Workshop on. IEEE*, ۲۰۰۸.

[۳۱] R. S. Chakraborty and S. Bhunia, "Security against hardware Trojan through a novel application of design obfuscation," *Computer-Aided Design-Digest of Technical Papers, ICCAD. IEEE/ACM International Conference on.*, ۲۰۰۹.

[۳۲] A. Waksman and S. Sethumadhavan, "Silencing hardware backdoors," *Security and Privacy (SP) IEEE Symposium on*, pp. ۴۹-۶۳, ۲۰۱۱.

[۳۳] D. Arora, S. Ravi, A. Raghunathan, and N. K. Jha, "Hardware-assisted run-time monitoring for secure program execution on embedded processors," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, pp. ۱۲۹۵-۱۳۰۸, ۲۰۰۶.

[۳۴] S. Hart, "APHID: Anomaly Processor in Hardware for Intrusion Detection," *AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF ENGINEERING AND MANAGEMENT*, ۲۰۰۷.

زیرنویس‌ها

۱ - Third Party Intellectual Property(۳PIP)
۲ - Integrated Circuits
۳ - Platform
۴ -Vulnerability
۵ - Functionality
۶ - Reliability
۷ - Outsourcing
۸ - Kill Switch
۹ - Commercial-off-the-shelf
۱۰ - Firmware
۱۱ - Trigger
۱۲ - Payload
۱۳ - Mask
۱۴ - Electromigration
۱۵ - Assembly
۱۶ - Printed circuit board
۱۷ - Interface
۱۸ - Malicious
۱۹ - Computer-Aided Design
۲۰ - Layout
۲۱ - Clock Skew
۲۲ - Process Variation
۲۳ - Denial-of-service
۲۴ - Backdoor
۲۵ - Denial of Service (DoS)
۲۶ - Glitch
۲۷ - Freeze
۲۸ - Scanning electron microscope
۲۹ - Buttom-Up
۳۰ - Design-for-Test (DfT)
۳۱ - Design-for-Security (DfS)
۳۲ - Fault
۳۳ - Probabilistic equivalence checking
۳۴ - Quiescent Supply Current
۳۵ -Trace
۳۶ - Switching Activity
۳۷ - Inter-Die
۳۸ - Intra-Die
۳۹ - Skew
۴۰ - Transient supply current
۴۱ - Calibration
۴۲ - Customized
۴۳ - Liveness Check Signals
۴۴ - DEsign-For-ENabling-SEcurity (DEFENSE)
۴۵ - Exception
۴۶ - Detour
۴۷ - Self-Scheduling
۴۸ - Bypass
۴۹ - Slave
۵۰ - Master
۵۱ - Malicious Lock Detection
۵۲ - Finite State Machine
۵۳ - Guards
۵۴ - Scrambling
۵۵ - Replication
۵۶ - Fragmentation
۵۷ - Coprocessor
۵۸ - Cryptoprocessor
۵۹ - Tamper-Resistant
۶۰ - Anomaly

افزایش قابلیت اطمینان حافظه‌های روی تراشه در پردازنده‌های مدرن با استفاده از حافظه‌های غیر فرار

بهار عسگری[۱]، مهدی فاضلی[۲]، سیدوحید ازهری[۳]

[۱] دانشجوی کارشناسی ارشد مهندسی کامپیوتر- معماری سیستم‌های کامپیوتری، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
asgari@comp.iust.ac.ir

[۲و۳] استادیار گروه سخت افزار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
m_fazeli@iust.ac.ir و azharivs@iust.ac.ir

چکیده

حافظه‌های درون تراشه مانندحافظه‌ی نهان و بانک ثباتها یکی ازمهم‌ترین و چالش برانگیزترین اجزای پردازنده‌ها هستند. اما با کاهش ابعاد فناوری‌های ساخت این حافظه‌ها به سمت مقیاس‌های نانومتری، سطح ولتاژ و فاصله‌ی بین عناصر حافظه کاهش یافته-است. این موضوع سبب شده تا حافظه‌ها به شدت نسبت به برخورد ذرات پرانرژی حساس شوند. لذا نقش مهم این حافظه‌ها درکارکرد پردازنده‌ها، اهمیت تحمل‌پذیر نمودن این قسمت در برابر خطا را به صورت قابل توجهی افزایش می‌دهد. از طرفی دیگر جریان‌های نشتی که باعث تولید توان مصرفی ایستا در مدارات CMOS می‌شوند با کاهش ابعاد فناوری به طرز قابل توجهی در حال افزایش هستند. اخیرا برای برطرف کردن این مشکل‌ها، فناوری حافظه‌های غیرفرار معرفی شده‌اند. به دلیل تفاوت‌های بنیادی که در پیاده‌سازی این نوع حافظه‌ها درمقایسه با سلول‌های حافظه‌ی ایستا، وجود دارد، حافظه‌های غیرفرار در برابر تک و چند رخدادهای واژگونی کاملا مقاوم می‌باشند. ازطرفی جریان نشتی کم و در نتیجه توان مصرفی ایستای اندکِ این حافظه‌ها، ویژگی مثبت دیگری است که طراحان را ترغیب می‌کند که عناصر حافظه‌ی مبتنی بر فناوری‌های نانومتری CMOS را باحافظه‌های غیرفرار جایگزین نمایند.

درکنار این ویژگی‌های مثبت، حافظه‌های غیرفرار، ازمعایبی نظیر محدودیت تعداد نوشتن و تاخیر زیاد عملیات نوشتن رنج می‌برند که بزرگترین مانع برای جایگزینی آنها با حافظه‌های موجود است، لذا مباحث اصلی که در این سمینار مورد مطالعه قرارخواهندگرفت روش‌هایی است که تاکنون برای غلبه برمشکلات حافظه‌های غیرفرار ارائه شده‌است تا بتوان از آنها در سطوح مختلف حافظه استفاده‌ی بهینه نمود. همچنین روش نوینی برای استفاده از حافظه‌های غیرفرار در بانک ثبات ارائه داده‌ایم که در این سمینار به آن پرداخته خواهد شد.

کلمات کلیدی

حافظه‌های روی تراشه، حافظه‌ی نهان، حافظه‌های غیرفرار، قابلیت اطمینان.

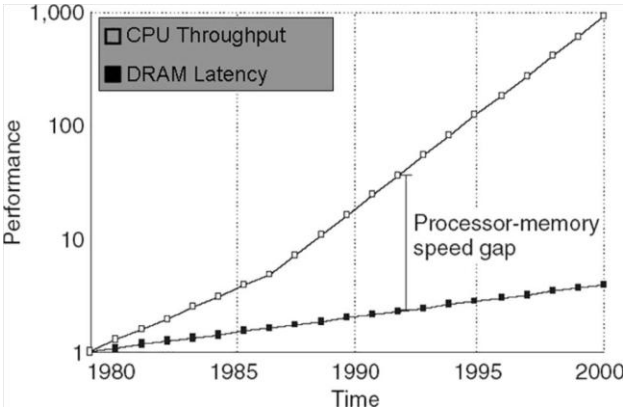
۱- مقدمه

سرعت پردازش پردازنده‌ها در سال‌های اخیر به شدت رو به افزایش می‌باشد. این افزیش به طور متوسط، ۷۵درصد در سال بوده‌است[۱]. با این روند رو به جلو، نرخ دسترسی به حافظه‌های داخل تراشه، از جمله حافظه‌ی نهان، افزایش می‌یابد. اما روند افزایش سرعت دسترسی به حافظه‌ها حدود ۷ درصد در سال می‌باشد که این نرخ بسیار کمتر از روند افزایش سرعت پردازنده است. شکل (۱) نمودار اختلاف کارایی پردازنده و حافظه اصلی [۳]. را نشان می‌دهد. از سال‌های گذشته تا کنون، این موضوع سبب شده تا طراحان سامانه‌های کامپیوتری به‌دنبال راهی برای جبران اختلاف سرعت میان پردازنده و حافظه باشند. از نخستین گام‌هایی که در این راستا برداشته شد، استفاده از حافظه‌هایی با حجم کمتر و سرعت دستیابی بیشتر، در کنار پردازنده بود. از حافظه‌ی نهان می‌توان به عنوان یکی از بارزترین مثال‌های به‌کارگیری این روش یاد کرد.

اما قرار دادن حافظه ی نهان در کنار پردازنده هم به تنهایی نتوانست پاسخگوی سرعت زیاد پردازنده‌های امروزی باشد، لذا برای اینکه

حافظه‌های داخل پردازنده منجر به ایجاد گلوگاه نشوند و کارایی کل پردازنده را تحت تاثیر خود قرار ندهند، احتیاج است تا حافظه‌هایی چگال تر و با سایز سلول کوچکتر و سریع‌تر داشته باشیم.

اما فناوری‌های رایج مانند DRAM، SRAM که هم‌اکنون برای ساخت حافظه استفاده می‌شوند در سایزهای کوچکتر از ۱۶ نانومتر با مشکلاتی نظیر کاهش قابلیت اطمینان و افزایش توان مصرفی ایستا، دست و پنجه نرم می‌کنند. یکی از مهم‌ترین این مشکلات، برخورد ذرات پرانرژی و ایجاد خرابیِ گذرا در این حافظه‌هاست[۲]. از آنجا که این حافظه ها دارای داده‌های مهم و پراستفاده در پردازنده هستند لذا رخداد اشکال در آنها می تواند منجر به رخداد خرابی در نتایج شود.



شکل (۱) نمودار اختلاف کارایی پردازنده و حافظه اصلی [۳].

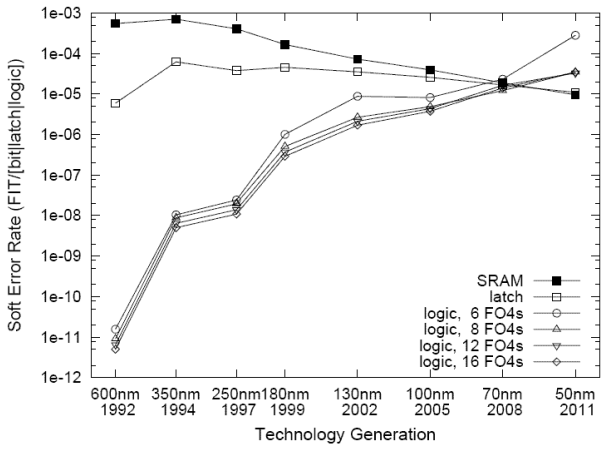
از یک سو، قابلیت اطمینان و تحمل‌پذیری اشکال ویژگی‌هایی هستند که در بسیاری از سامانه‌های نهفته به ویژه آن‌هایی که در کاربردهای بحرانی-امن[۱] به کار گرفته می‌شوند دارای اهمیت بسیار زیادی هستند. از سوی دیگر با کوچک شدن اندازه‌ی ترانزیستورها، در فناوری‌های رایج، توان نشتی نیز به شدت رو به افزایش است[۴]، و از آنجایی که بهینه بودن مصرف انرژی ، همواره به عنوان یک شرط غیر قابل انکار در طراحی سامانه‌های نهفته در نظر گرفته می‌شود لذا طراحان همواره به دنبال حل چالش‌ها در این دو زمینه در سیستم‌های امروزی بوده‌اند. در سال‌های اخیر مطالعات انجام شده به سمتی پیش رفته که از فناوری‌های جدیدتری به نام حافظه‌های غیر فرار[۲] به عنوان عنصر ذخیره‌سازی در سطوح مختلف حافظه، اعم از داخل و خارج پردازنده استفاده شود. این نوع حافظه‌ها در مقابل خطاهایی که بر اثر برخورد ذرات پرانرژی به وجود می‌آیند کاملا مقاوم هستند[۵]. از جمله مزایای دیگر این حافظه‌ها که آن را برای جایگزین نمودن با فناوری‌های جدید مناسب نموده‌است، چگالی بالا و توان نشتی نزدیک به صفر آن‌هاست لذا انواع حافظه‌های غیرفرار، همانند PCM، MRAM و RRAM این امکان را مهیا ساخته‌اند تا بتوان حافظه‌های روی تراشه‌ای با خواص غیرفرار بودن، کم بودن مصرف انرژی، چگالی بالا و زمان دسترسی بالا[۶] داشت.

در ادامه‌ی این مقاله، در بخش دوم با محدودیت‌های فناوری‌های رایج حافظه آشنا می‌شویم، سپس در بخش دوم حافظه‌های غیر فرار که محدودیت‌های این فناوری‌ها را ندارند و می‌توانند جایگزین مناسبی برای آنها باشند را بررسی می‌کنیم، بعد از آن در بخش سوم معایب این نوع حافظه‌ها را خواهیم دید و در ادامه در بخش چهارم روش‌های مقابله با آنها، و نهایتا در بخش پنجم روند معماری پردازنده‌های نوین و نقش حافظه‌های روی تراشه در آن‌ها را مرور می‌کنیم.

۲- محدودیت‌های فناوری CMOS

با حرکت تدریجی فناوری‌ها به سمت مقیاس‌های نانومتری، سطح ولتاژ و فاصله بین عناصر حافظه کاهش یافته‌است. این ویژگی‌ها سبب شده تا حافظه‌ها به شدت نسبت به برخورد ذرات پر انرژی حساس شوند[۲].

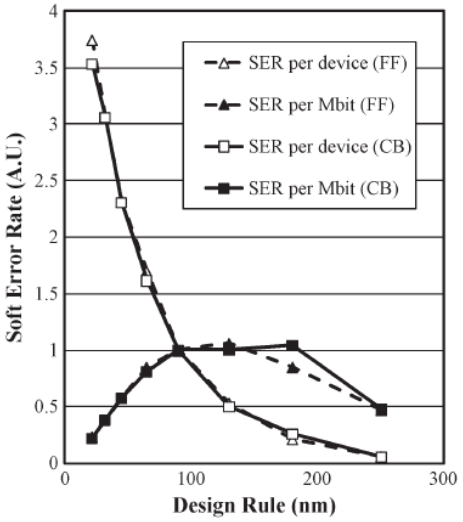
به عنوان مثال هنگامی‌که اندازه‌ی فناوری از ابعاد ۰٫۶μm به ابعاد ۰٫۱μm تغییر می‌یابد، میزان خطای نرم در عناصر حافظه تقریباً ثابت می‌ماند ولی تعداد بیت‌های روی یک تراشه به شکل سهموی در حال افزایش است که در نتیجه تعداد خطای نرم روی یک تراشه با گذر از یک فناوری به فناوری با ابعاد کوچک‌تر در حال افزایش است. شکل(۲)اثرکاهش اندازه ابعاد در فناوری‌های مختلف در نرخ خطای نرم را نشان می‌دهد.



شکل(۲)اثرکاهش اندازه ابعاد در فناوری‌های مختلف در نرخ خطای نرم [۷]

در تحقیقی در[۸] در مورد روند تغییر نرخ خطای نرم در حافظه از نوع SRAM با کاهش ابعاد فناوری نشان داده شده است که با کاهش ابعاد فناوری از ۱۳۰nm به ۲۲nm نرخ خطای نرم به ازای دستگاه[۳] در حافظه SRAM ۶ تا ۷ برابر افزایش می‌یابد .شکل(۳)روند تغییر نرخ خطای نرم در حافظه‌ی SRAM[۸] را نشان می‌دهد. ولیکن نرخ خطای نرم به ازای یک مگابیت از حافظه‌ی SRAM در حال کاهش است. علت این مسئله این است که با کاهش ابعاد، فضای اشغالی برای هر بیت کاهش می‌یابد که به تبع آن فضای حساس در هر بیت حافظه کاهش می‌یابد که باعث کاهش نرخ خطای نرم می‌گردد. ولی، با کاهش ابعاد فناوری ولتاژ تغذیه و بار بحرانی مدار نیز کاهش می‌یابد که باعث

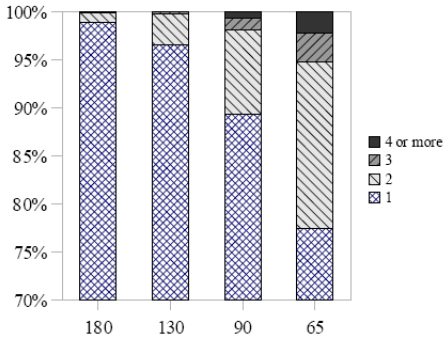
افزایش نرخ خطای نرم به صورت نمایی می‌گردند. از طرف دیگر با کاهش ابعاد فناوری حجم حافظه‌ی SRAM استفاده شده در تراشه یا دستگاه نیز به شکل قابل توجهی افزایش می‌یابد. در حقیقت اثر افزایش کلی حجم حافظه به طرز قابل توجهی بر اثر کاهش فضای اشغالی یک بیت غلبه کرده است. این مسئله باعث شده است نرخ خطای نرم کلی به شکل قابل توجهی با کاهش ابعاد فناوری ساخت افزایش یابد. مسئله دیگری که وجود دارد این است که با کاهش ابعاد فناوری، چالش جدیدی به نام خطاهای نرم چندگانه[٤] به وجود آمده است. خطاهای نرم چندگانه زمانی رخ می‌دهند که ذرات پرانرژی بیش از یک گره حساس را تحت تأثیر قرار دهد و منجر به رخداد چند خطای نرم در مدار شود. با کاهش ابعاد فناوری و متعاقب آن کاهش ولتاژ کاری مدارها، کاهش ولتاژ آستانه، کاهش فاصله بین دو سلول، نرخ خطای چندگانه به شکل قابل ملاحظه‌ای افزایش یافته به طوری‌که در سال‌های اخیر پژوهش‌هایی نیز در این زمینه انجام شده است[۸-۱۰]. در [۸] نشان داده شده است که در فناوری ۲۲nm نسبت نرخ خطای نرم چندگانه به بیش از ۵۰٪ خواهد رسید.



شکل(۳)روند تغییر نرخ خطای نرم در حافظه‌ی SRAM[۸]

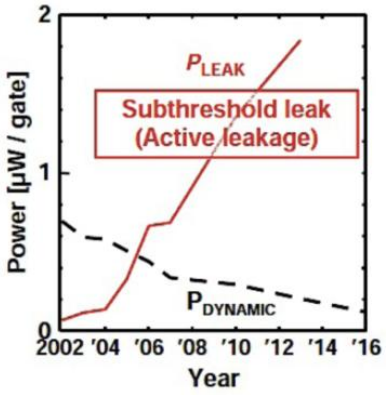
طبق نتایج یک کار پژوهشی انجام شده در [۱۱] که توسط آزمایش‌های شتاب داده شده توسط تشعشعات نوترون در Los Alamos بر روی یک ریزپردازنده انجام شده است، می‌توان گفت، نرخ خطای نرم چندگانه بر اثر کاهش ابعاد فناوری ساخت نقش برجسته‌ای در نرخ کلی خطای نرم پیدا کرده است. شکل(۴) میزان مشارکت خطاهای چندگانه در نرخ خطای کلی[۱۸] را نشان می‌دهد.

اما مشکل استفاده از حافظه تنها محدود به کاهش قابلیت اطمینان آن نیست، در حقیقت حافظه‌ها را می‌توان مهمترین منبع مصرف انرژی در یک سامانه دانست، این موضوع به خصوص برای سامانه‌های نهفته که دارای محدودیت در منابع انرژی هستند، اهمیتی دو چندان دارد[۱۲].



شکل(۴) میزان مشارکت خطاهای چندگانه در نرخ خطای کلی[۱۱]

اما از جمله دیگر مشکلاتی که فناوری CMOS در مقیاس نانومتر با آن روبروست، توان مصرفی ایستا است. شکل(۵) روند میزان توان مصرفی فناوری CMOS در سالهای آینده [۴] را نشان می‌دهد. همان طور که این شکل نشان می‌دهد، علارغم اینکه در سال‌های گذشته، بخش بسیار کمی از توان مصرفیِ یک گیت مربوط به توان نشتی بود، در سال‌های آینده مصرف توان ایستا به شدت افزایش خواهد یافت و کاهش آن از جمله نگرانی‌هایی است که نظر طراحان را به خود جلب نموده‌است.



شکل(۵) روند میزان توان مصرفی فناوری CMOS در سالهای آینده[۴]

۳- حافظه‌های غیر فرار

در سال‌های اخیر انواع مختلفی از حافظه‌های غیرفرار با ویژگی‌های مقاوم بودن در برابر خطاهای نرم و توان مصرفیِ ایستای ناچیز، معرفی شده‌اند. این حافظه‌های خود ویژگی‌های متفاوتی از نظر سرعت و چگالی دارند. لذا طراحان سامانه‌های کامپیوتری بسته به محل قرارگیری این حافظه‌ها در سلسله مراتب حافظه و ویژگی آنها از جمله زمان دسترسی، انتخاب‌های متفاوتی می‌توانند داشته باشند. در ادامه سه نوع حافظه‌ی تغییر فازی، مقاومتی و مغناطیسی، که تا کنون متداول‌ترینِ آنها برای استفاده در داخل تراشه بوده‌اند بررسی خواهند

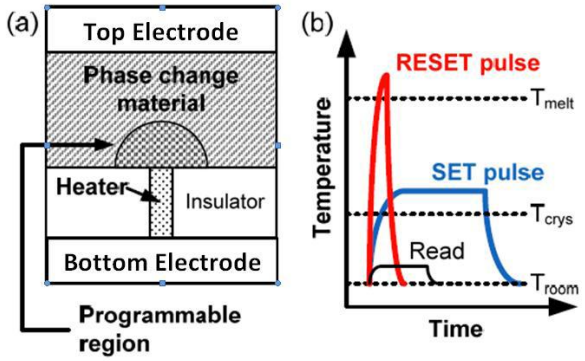
شد و با توجه به ویژگی‌های هر یک خواهیم دید برای کدام سطح از حافظه مناسب‌تر خواهد بود.

۳-۱- حافظه‌های تغییر فازی[۵]

حافظه‌های PCM از اختلاف مقاومتی زیاد موجود میان حالات بی‌نظم و منظم موادی که تغییر فازی در آنها رخ می‌دهد، بهره می‌برند. فاز بی‌نظم در این نوع مواد دارای مقاومت الکتریکی بالا و فاز منظم، دارای مقاومت الکتریکی کم، و در برخی مواد 10^{-3} یا 10^{-4} بار کمتر از حالت بانظم است[۱۳].

برای اینکه یک سلول از PCM مقدار یک را در خود ذخیره کند، یا به عبارت دیگر فاز سلول به حالتی با مقاومت الکتریکی کم (حالت منظم) تغییر شکل دهد، یک پالس الکتریکی به سطح قابل توجهی از آن سلول اعمال شده و دمای آن را تا دمایی بالاتر از دمای کریستاله شدن ماده بالا می‌برد. از این عملیاتِ نوشتنِ مقدارِ "یک" به عنوان عملیات محدود کننده در زمان تاخیر نوشتن یک سلول PCM یاد می‌شود. مدت زمانی که یک پالس الکتریکی به سطح سلول PCM اعمال می‌شود تا مقدار منطقی یک را در آن بنویسد، بطور مستقیم وابسته به سرعت کریستاله شدن ماده‌ی به‌کار رفته در ساخت سلول PCM است.از آنجایی که عملیات تغییر ساختار حافظه‌های PCM در شرایط محیطی عادی، امکان‌پذیر نیست، از PCM به عنوان ساختار حافظه‌ای که سال‌های زیادی داده را بدون مخدوش کردن آن، در خود نگهداری می‌کند، یاد می‌شود[۱۴].

در عملیات نوشتنِ مقدار "صفر" در یک سلول از حافظه‌ی PCM، یک جریان الکتریکی بزرگتر از جریان عملیات نوشتن "یک"، به نقطه‌ی مرکزی سلول اعمال شده و بخشی از مرکز یک سلول را ذوب می‌کند. در صورتی‌که این پالس به مدت زمان مناسبی اعمال گردد، ماده ذوب شده به حالت بی‌نظم تغییر یافته و یک سلول با مقاومت الکتریکی بالا را پدید می‌آورد. عملیات نوشتن "صفر" در داخل یک سلول از دیدگاه مصرف توان و جریان یکی از بخش‌های چالش‌برانگیز استفاده از PCM در سامانه‌ها است. به‌همین دلیل طراحان در انتخاب قطعه‌ای که امکان عبور جریان و توان مناسب برای عملیات نوشتن "صفر" را داشته باشد و حجم آن از حجم سلول PCM بیشتر نباشد، نهایت دقت را دارند. عملیات خواندن از سلول نیز بوسیله‌ی خواندن میزان مقاومت سلول PCM در ولتاژ کم، انجام می‌گیرد و به‌این صورت اطمینان حاصل می‌گردد که خواندن از PCM سبب تغییر حالت آن نخواهد شد. پالس‌های ولتاژ مورد نیاز برای عملیات مذکور و نمایی از یک سلول PCMدر شکل (۶) a- مقایسه‌ی پالس‌های اعمال شده به یک سلول PCMبه منظور انجام عملیات مختلف[۱۴].b- نمایی از یک سلولPCM [۱۵] . مشاهده می‌شود.



شکل (۶) a- مقایسه‌ی پالس‌های اعمال شده به یک سلول PCMبه منظور انجام عملیات مختلف[۱۴].b- نمایی از یک سلولPCM [۱۵] .

برای ذخیره‌ی داده‌ها در سلول‌های PCM، همان‌طور که اشاره شد، اختلاف بین مقاومت سلول را در نظر می‌گیرند، حال در صورتی که این اختلاف‌ها به بازه‌های بیشتری تقسیم شوند، به‌جای یک بیت، تعداد بیشتری داده در یک سلول گنجایش می‌یابد [۶, ۱۶].

از PCM به عنوان یک فناوری مقیاس‌پذیر نام برده می‌شود. در واقع هر چه چگالی قطعات با کاهش مقیاس فناوری‌ها، افزایش می‌یابد، لایه‌ی موجود از ماده‌ی تغییر فاز دهنده نیز کاهش یافته و جریان کمتری به‌منظور عملیات نوشتن نیاز دارد. به همین دلیل هم‌اکنون نمونه‌های اولیه این حافظه در فناوری ۲۰ نانومتر ایجاد گردیده و قصد بر آن است تا پیاده‌سازی PCM ها را تا فناوری ۹ نانومتری ادامه دهند [۱۵]. این موضوع در حالی است که حافظه‌های مبتنی بر فناوری DRAM در مقیاس‌های زیر ۴۰ نانومتر، امکان پیاده‌سازی نخواهند داشت [۶, ۱۶].

میزان پایداری یک سلول PCM به تعداد نوشتن‌هایی بستگی دارد که در آن انجام می‌شود. در واقع هنگامی که جریان به ماده‌ی تغییر فاز دهنده اعمال می‌شود، انبساط و انقباض دمایی باعث فرسودگی محل‌های اتصال شده، و پس از مدتی جریان به‌صورت مناسب به سلول تزریق نمی‌گردد [۱۵]. میزان پایداری[۶] سلول‌های PCM، هم‌اکنون بین 10^4 تا 10^9 عدد نوشتن گزارش شده است، اما عددی که به‌صورت محافظه‌کارانه برای پایداری سلول‌های PCM گزارش شده برابر 10^8 می‌باشد[۱۶].

۳-۲- حافظه‌های مقاومتی[۷]

حافظه‌های RRAM سلول‌هایی با ساختاری شبیه خازن دارند. در این سلول‌ها، با اعمال جریان و ولتاژهای مختلف، در لایه‌ی انتقالی اکسید فلز، خاصیت تغییر مقاومت نمایان می‌شود. به‌صورت کلی دو رفتار مختلف در تغییر مقاومت حافظه‌های RRAM مشاهده می‌شود [۱۷]. تغییر تک قطبی، که در این حالت جهت تغییر به دامنه‌ی ولتاژ وابسته

خواهد بود و تغییر دو قطب، که در این حالت جهت تغییر به مثبت یا منفی بودن ولتاژ بستگی خواهد داشت.

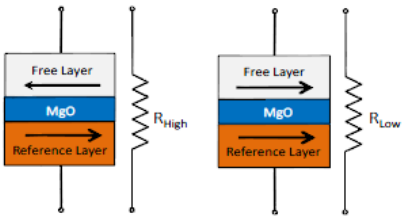
اغلب این نوع حافظه را مناسب برای جایگزینی با حافظه‌ی flash می‌دانند. دلایل اصلی اقبال طراحان به RRAMها را می‌توان در چند موضوع دانست.. نخست مشکل بودن ساخت سلول‌های Flash در فناوری ۱۶ نانومتری و در مقابل آن، ساختار ساده‌ی RRAMها و توانایی تولید آنها در فناوری ۸ نانومتری. از دیگر عوامل تمایل طراحان برای استفاده از RRAMها می‌توان به هزینه‌ی اندک به ازای ساخت هر بیت، توانایی انجام عملیات در ولتاژهای پائین، کم بودن مصرف انرژی، پایداری بالا ، اشاره کرد .

با توجه به اینکه طیف گسترده‌ای از فناوری‌های مختلف برای سلول‌های RRAM موجود است و در این تحقیق امکان پرداختن به تمامی این فناوری‌ها وجود ندارد، در ادامه سلول‌های Memristor، به عنوان یکی از موفق‌ترین ساختارهای ارائه شده برای سلول‌های RRAM، مورد بررسی قرار خواهندگرفت [۱۵].

از مقیاس‌پذیری به عنوان نقطه‌ی قوت سلول‌های این حافظه یاد می‌شود. انتظار می‌رود تا بتوان در چند سال آینده، Memristorها را در فناوری‌های ۴ یا ۵ نانومتری، تولید کرد [۱۸]. انرژی مصرفی برای تغییر حالت Memristor، می‌تواند تا ۲۰ برابر کمتر از سلول‌های Flash باشد.نقطه‌ی چالش‌برانگیز استفاده از Memristorها پایداری آنها است. نمونه‌های اولیه از این نوع حافظه، دارای پایداری برابر 10^5 هستند، این در حالی است که حافظه‌های اصلی به پایداری در حدود 10^{17} نیاز دارند. با این وجود می‌توان از آنها برای کاربردهایی نظیر کاربرهای سلول‌های Flash که در آنها اغلب عملیات مربوط به خواندن بوده و کمتر نوشتن رخ می‌دهد، استفاده نمود. از جمله این کاربردها می‌توان به FPGA ها اشاره کرد [۱۵]. روش دیگر غلبه بر مشکل پایداری آنها استفاده از آنها در ساختارهای ترکیبی با دیگر انواع سلول‌های حافظه، همانند سلول‌های DRAM است.

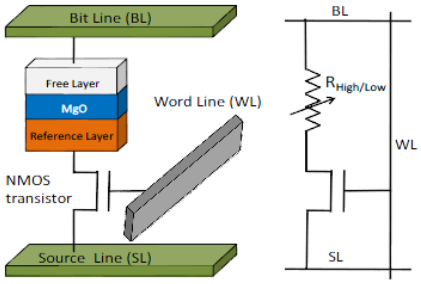
۳-۳- حافظه‌های مغناطیسی[۸]

در میان حافظه‌های غیرفرار، MRAMها خصوصیات قابل توجهی دارند. در واقع حافظه‌هایMRAM این توانایی را دارند تا در آینده به عنوان یک حافظه‌ی همه منظوره، چگال و با سرعت بالا در طراحی سامانه‌های کامپیوتری مورد استفاده قرار گیرند. اگرچه در بین مشکلات موجود در این مسیر، جریان نوشتن در این سلول‌ها باید به‌صورت قابل ملاحظه‌ای کاهش یابد. حافظه‌های MRAMاز عنصری به نام MTJ استفاده می‌کنند. یک MTJ از دو لایه‌ی فرومغناطیسی که توسط یک لایه‌ی اکسیدی همانند MgO از هم جدا می‌شوند، تشکیل شده است. شکل (۷)طرح انتزاعی از MTJ که در آن جهت موازی دو میدان مغناطیسی نشان‌دهنده‌ی منطق صفر است وجهت مخالف دو میدان مغناطیسی نشان‌دهنده‌ی منطق یک است[۱۹]. را نشان می‌دهد.



شکل (۷)طرح انتزاعی از MTJ که در آن جهت موازی دو میدان مغناطیسی نشان‌دهنده‌ی منطق صفر است وجهت مخالف دو میدان مغناطیسی نشان‌دهنده‌ی منطق یک است[۱۹].

یکی از لایه‌های مغناطیسی، لایه‌ی مرجع است که در آن جهت مغناطیسی ثابت می‌باشد. لایه‌ی دیگر، لایه‌ی آزاد است. در این لایه جهت مغناطیسی می‌تواند توسط میدان مغناطیسی و یا جریان پولاریزه شده، تغییر کند. هنگامی که جهت مغناطیسی هر دو لایه یکسان باشد، مقاومت MTJ اندک و در صورتی که جهت مغناطیسی دو لایه معکوس هم باشد، مقاومت MTJ زیاد خواهد بود [۱۵]. نام این پدیده TMR[۹] است[۶]. در شکل (۸) طرح مداری یک سلولMRAM و طرح انتزاعی یک سلول‌مشاهده می‌شود.

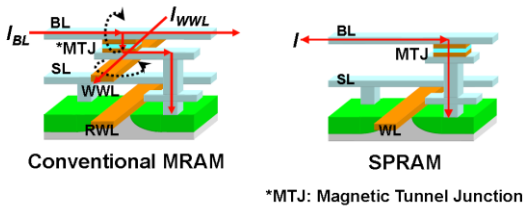


شکل (۸) طرح مداری یک سلولMRAM و طرح انتزاعی یک سلولMRAM [۱۹]

با وجود اینکه فناوری MRAM به بلوغ نسبی رسیده است، پیاده‌سازی‌های موجود از این نوع حافظه‌ها از دیدگاه‌های چگالی و انرژی مصرفی توانایی رقابت با سلول‌های دیگر از جمله سلول‌های DRAM و Flash را ندارند[۱۵]. به همین دلیل در ادامه نوع خاصی از این سلول‌ها به نام STT-RAM مورد بررسی قرار خواهند گرفت.

حافظه‌های STT-RAM در سال‌های اخیر یکی از پراقبال‌ترین ساختارهای حافظه‌های غیرفرار بوده‌اند. دلیل این موضوع مقیاس‌پذیری در فناوری‌های زیر ۱۰۰ نانومتر و جریان نوشتن کمتر در مقایسه با ساختارهای سنتی MRAMها است. در ساختارهای سنتی حافظه‌های MRAM به‌منظور تغییر جهت مغناطیسی لایه‌ی آزاد، از ترکیب جریان نوشتن در خط نوشتن کلمه[۱۰]و خط بیت[۱۱]استفاده می‌شود. این روش دو ایراد اساسی دارد. نخست نیاز به

جریان زیاد برای نوشتن و دیگری تاثیر میدان مغناطیسی بر سلول‌های مجاور سلولی که در آن عملیات نوشتن اتفاق می‌افتد. اما در روش استفاده شده در STT-RAMاز جریان خط بیت، به‌دنبال خط منبع[۱۲] به‌منظور تغییر جهت مغناطیسی در لایه‌ی آزاد MTJ استفاده می‌شود. در واقع در این روش الکترون‌ها با یک جهت چرخشی از داخل لایه‌ی مرجع عبور کرده، یک تونل در لایه‌ی نازک MgO ایجاد نموده و به لایه‌ی آزاد تزریق می‌شوند.شکل(۹)مقایسه دو نوع سلولMRAM، سمت چپ ساختار سنتی و سمت راست ساختار STT-RAM را نشان می‌دهد.



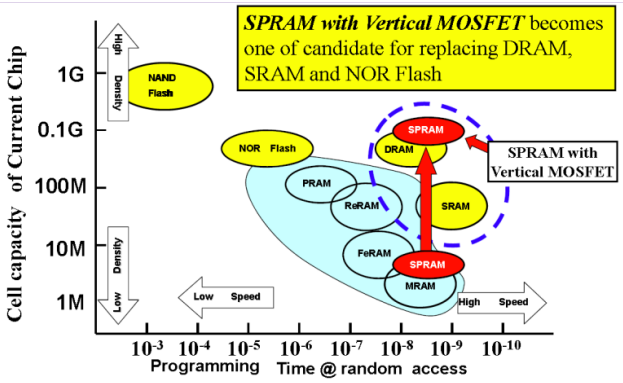
شکل(۹)مقایسه دو نوع سلولMRAM، سمت چپ ساختار سنتی و سمت راست ساختار STT-RAM [۲۰]

در یک سلول STT-RAM به منظور انجام عملیات خواندن، ترانزیستور NMOS روشن شده و یک اختلاف ولتاژ اندک بین خط بیت و خط منبع ایجاد می‌شود. این اختلاف ولتاژ سبب به وجود آمدن یک جریان داخل MTJ می‌گردد، میزان جریان به وجود آمده در این زمان به حالت سلول MTJ بستگی دارد. در ادامه توسط یک حس‌گر جریان میزان جریان ایجاد شده در MTJ با جریان مرجع مقایسه شده و بر مبنای آن تصمیم‌گیری در رابطه با صفر یا یک بودن مقدار ذخیره شده در سلول MTJ صورت می‌گیرد.

به منظور انجام عملیات نوشتن یک اختلاف ولتاژ مثبت زیاد، برای نوشتن صفر، و یک اختلاف ولتاژ منفی زیاد برای نوشتن یک، بین خط بیت و خط منبع ایجاد می‌شود. میزان جریانی که برای اطمینان از انجام صحیح عملیات نوشتن در یک سلول STT-RAM مورد نیاز است را جریان آستانه می‌نامند. این جریان به نوع ماده به‌کار رفته برای ساخت لایه‌ی میانی بین دو لایه‌ی آزاد و مرجع، مدت زمان اعمال پالس نوشتن، و شکل هندسی MTJ وابسته خواهد بود[۲۰].

از دیدگاه میزان پایداری، سلول‌های STT-RAM در میان دیگر فناوری‌های حافظه‌های غیرفرار دارای بیشترین پایداری در برابر عملیات نوشتن هستند. اعداد گزارش شده در این رابطه در بازه‌ی 10^{12}تا 10^{15} [۱۵] قرار می‌گیرند.

ویژگی سلول‌های STT-RAM از دیدگاه مصرف انرژی نیز قابل تامل است. میزان انرژی مصرفی سلول‌های STT-RAM در هنگام خواندن برابر ۰/۰۲۶ نانو ژول و در هنگام نوشتن برابر ۲/۸۳۳ نانو ژول است[۲۱].شکل(۱۰) مقایسه‌ی STT-RAM با سایر حافظه‌های موجود از نظر چگالی را نشان می‌دهد. این نمودار، چگالی حافظه‌ها را نسبت به سرعت دسترسی آن نشان می‌دهد. همان‌طور که ملاحظه می‌شود، SPRAMیا همان STT-RAM، با MOSFET عمودی نسبت به DRAM وSRAM که فناوری‌های رایج برای حافظه‌ی اصلی و حافظه‌های درون تراشه هستند، چگالی بیشتری دارد.



شکل(۱۰) مقایسه‌ی STT-RAM با سایر حافظه‌های موجود از نظر چگالی[۲۰]

۳-۴- مقایسه‌ی انواع حافظه‌ی غیرفرار

در جدول(۱)خلاصه‌ای از ویژگی‌های فناوری‌های موجود برای حافظه آمده‌است.

جدول(۱)خلاصه‌ای از ویژگی‌های فناوری‌های موجود برای حافظه

	SRAM	PCM	RRAM	MRAM
میزان رشد فناوری	ساخت	توسعه یافته، نمونه-های اولیه موجود است	در حال توسعه	توسعه یافته، مدل فیزیکی و ابزار شبیه‌سازی موجود است
اندازه‌ی سلول(F^2)	بزرگتر از ۱۰۰	۸ تا ۴۰	بزرگتر از ۵	۲۱ تا ۳۷
انرژی مصرفی خواندن(pj/bit)	۰,۰۰۱	۲,۴۷	۰,۱۲	۰,۰۰۱
تاخیر خواندن(ns)	کمتر از ۱۰	۲۰ تا ۵۰	کمتر از ۱۰	۲ تا ۲۰
مقیاس پذیری	محدود به ۴۵ نانومتر	وابسته به قطعه‌ی انتخاب سلول	کمتر از ۳۰ نانومتر	امکان پیاده-سازی در فناوری‌های زیر ۱۰۰ نانومتر

۳-۴-۱- استفاده از حافظه‌های غیر فرار به عنوان حافظه-ی برروی تراشه

همان طور که پیش‌تر اشاره شد، برخی از حافظه‌های غیر فرار مانند STT-RAM، PCM و RRAM به عنوان جایگزین‌های مناسبی برای سلول‌های SRAM و DRAM معرفی شده‌اند. علاوه بر کارهای تحقیقاتی و مقالاتی که در مورد حافظه‌های غیر فرار در حال انجام است، بسیاری از شرکت‌های مطرح دنیا شروع به استفاده از حافظه‌های غیر فرار در محصولات خود نموده‌اند. به عنوان مثال شرکت توشیبا به منظور کاهش توان مصرفی، از STT_RAM برای جایگزین نمودن ۵۱۲کیلوبایت از حافظه‌ی نهان سطح دو که قبلا از جنس SRAM ساخته می‌شد، استفاده کرده است [۲۲]. شرکت سامسونگ نیز یک تراشه‌یPCM یک گیگابایتی با فناوری ۵۸ نانومتر تولید کرده است[۲۳]. همچنین شرکت‌های HP وHynix هم اعلام کرده‌اند که فناوری RRAM را در سال ۲۰۱۳ با حافظه‌های فلش، در سال‌های ۲۰۱۴ و ۲۰۱۵ با فناوری DRAM و در سال‌های بعد از آن با فناوری SRAM جایگزین نمایند[۲۴].

به طور کلی با توجه به ویژگی‌هایی که هر یک از انواع حافظه‌های غیر فرار دارند، می‌توانند جایگزین‌های مناسبی برای سطوح مختلف حافظه‌باشند. به عنوان مثال یک حافظه‌ی نهانِ ترکیبی که در آن از سلول‌های PRAM استفاده شده باشد، طول عمری حدود ۴,۷۰ تا ۱۹۶,۱۲ روز می‌توان داشته باشد[۲۵]، حال آنکه طول عمر یک حافظه‌ی نهان ترکیبی با استفاده از سلول‌های STT-RAM بیش از ده سال است. با توجه به این ویژگی، در بسیاری از مراجع، STT-RAM را بسیار مناسب برای استفاده در حافظه‌های درون تراشه و خصوصا حافظه‌ی نهان می‌دانند[۱۹, ۲۶-۲۸]. از طرفی فناوری PCM با توجه به چگالی زیاد و مصرف توان اندک، می‌تواند جایگزین مناسبی برای DRAM در طراحی حافظه‌ی اصلی باشد[۱۶].

البته لازم به ذکر است که با توجه به افزایش مقدار داده‌ها در کاربرد-های اخیر، نقش حافظه‌های داخل تراشه بسیار اهمیت بیشتری پیدا می‌کند و نیاز است که ظرفیت این سطح از حافظه افزایش بیابد، لذا برخی استفاده از PRAM را، به دلیل چگالی بسیار زیادِ آن، برای حافظه‌های نهان در سال‌های اخیر پیشنهاد کرده‌اند[۲۹].

۴- محدودیت‌های حافظه‌های غیرفرار

حافظه‌های غیرفرار در کنار تمام ویژگی‌های مثبت و مزایایی که نام برده شد، معایبی نیز دارند که تا کنون مانع از آن شده که به راحتی آنها را با حافظه‌های موجود جایگزین و از آن به عنوان حافظه‌ی فراگیر استفاده نمود. یکی از این معایب، تعداد دفعات نوشتن محدود است که از آن به اسم پایداری[۱۳] یاد می‌شود. به این معنا که بعد از تعداد محدودی نوشتن، دیگر نمی‌توان چیزی در سلول این نوع حافظه نوشت و تنها قادر خواهیم بود که آخرین مقدار موجود در این حافظه را بخوانیم[۳۰, ۳۱]. پس می‌توان گفت که حافظه‌های غیر فرار، علارغم مقاومت بسیار عالی در برابر خطاهای نرم، از وقوع خرابی در اثر انجام عمل نوشتن به شدت رنج می‌برند و به سرعت فرسوده می‌شوند. اما علاوه بر مساله‌ی فرسایش، حافظه‌های غیر فرار از مشکل دیگری نیز رنج می‌برند و آن مدت زمان عملیات نوشتن است که بسیار طولانی‌تر از فناوری‌های مرسوم می‌باشد. طبق آنچه در [۳۲] ارائه شده‌است در فرکانس ۳گیگاهرتز، مدت زمان عملیات نوشتن STT-RAM با ابعاد ۳۲نانومتر، ۳۳ سیکل ساعت است که حدود ۱۱ برابر SRAM با همین ابعاد است.همچنین انرژی مصرفی عملیات نوشتن هم زیاد و لذا مساله‌ی حائز اهمیتی است. . به طور خلاصه، جدول (۲) مقایسه‌ی میزان پایداری و انرژی مصرفی نوشتن را نشان می‌دهد.

جدول (۲) مقایسه‌ی میزان پایداری و انرژی مصرفی نوشتن

	SRAM	PCM	RRAM	MRAM
پایداری	بیش از 10^{15}	10^{8}تا10^{9}	10^{5}	10^{11}تا10^{15}
انرژی مصرفی نوشتن(pj/bit)	۰,۳۱	۱۴,۰۳تا ۱۹,۷۳	کمتر از ۰,۳۵	۲,۶

۴-۱- روش‌های ارائه شده برای افزایش قابلیت اطمینان حافظه‌های غیرفرار

با توجه به دو مشکل اساسیِ حافظه‌های غیر فرار که در قسمت قبل به آن‌ها اشاره،تحقیقات بسیاری برای چگونگی مقابله با دو مشکل این نوع حافظه‌ها انجام شده‌است که بتوان بدون لطمه زدن به کارایی سیستم، از مزایای بسیار مهم آن‌ها بهره برد. در ادامه شرحی در مورد این روش‌ها آمده‌است.

۴-۱-۱- روش‌های مبتنی بر تشخیص و تصحیح خطا

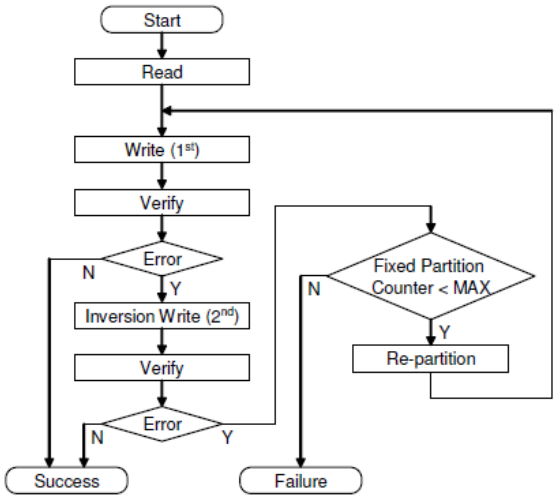
در [۳۰] یک روش سخت‌افزاری جدید برای تصحیح چند خطا برای حافظه‌های RRAM ارائه شده‌است. این روش می‌تواند به صورت ترکیبی با روش‌های سطح‌بندی فرسودگی[۱۴]موجود، استفاده بشود. ویژگی اصلی که این روش از آن استفاده می‌کند این است که یک سلول که که طول عمرش تمام شده‌است، هنوز قابل خواندن است. این موضوع باعث می‌شود که بتوانیم از بیت‌های خراب هم به عنوان یک عنصر ذخیره سازی استفاده کنیم. روش SAFER بلوک داده را به صورت پویا پارتیشن بندی می‌کند به نحوی که مطمئن باشیم که در هر پارتیشن فقط یک بیت خراب وجود دارد و در این صورت می‌توانیم

از روش‌های تصحیح یک خطا برای هر پارتیشن استفاده نماییم. در این مقاله ادعا شده است که روش SAFER با سرآیند سخت‌افزاری کمتری نسبت به روش‌های ECP (اشاره‌گر برای تصحیح خطا) و حتی روش مرسوم کد همینگ، می‌تواند تعداد خرابی‌های بیشتری را بازیافت نماید و طول عمر حافظه را افزایش دهد.

این روش به این صورت است که اگر در یک بلوک داده فقط یک خرابی موجود باشد، داده‌ای که قرار است در این بیت خراب نوشته بشود، مقدار معکوس مقدار موجود را داشته باشد، داده می‌تواند به صورت معکوس ذخیره شود و فقط یک بیت نیاز داریم که نشان بدهد که داده‌ی ذخیره شده در این بلوک مقدار معکوس داده‌ی واقعی است و موقع خواندن داده با توجه به این بیت بتوانیم تشخیص دهیم که مقدار خوانده شده را باید معکوس کنیم یا خیر.

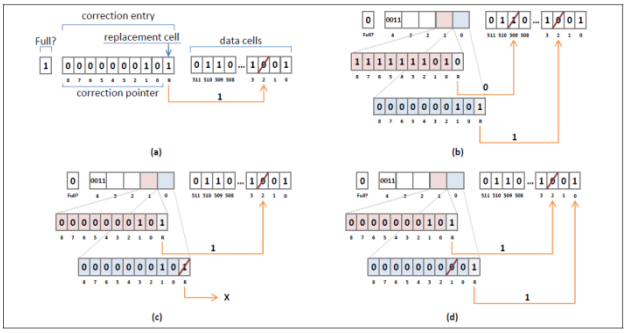
با وجود اینکه سرآیند سخت افزاری این روش فقط یک بیت برای هر پارتیشن است، مشکلی که وجود دارد این است که تصمیم گیری برای معکوس کردن داده و ذخیره‌ی آن بعد از اولین عملیات نوشتنی که با شکست روبرو می‌شود انجام می‌شود و این دو عمل نوشتن به دنبال دارد که خود موجب فرسایش حافظه می‌گردد. برای بر طرف کردن این مشکل از یک حافظه‌ی نهان کوچک برای ذخیره کردن مکان‌های خراب شده و مقدار بیت‌های خراب برای بلوک‌های مختلف، استفاده شده‌است. در واقع یک تاریخچه‌ی کوچک از خرابی بلوک‌ها در این حافظه‌ی نهان ذخیره می‌گردد. شکل(۱۱) الگوریتم کلی عملیات نوشتن درSAFER را نشان می‌دهد.

این روش برای بلوک داده‌ی ۵۱۲بیتی، طول عمر را به تعداد ۲۱,۶میلیون بار نوشتن افزایش می‌دهد.



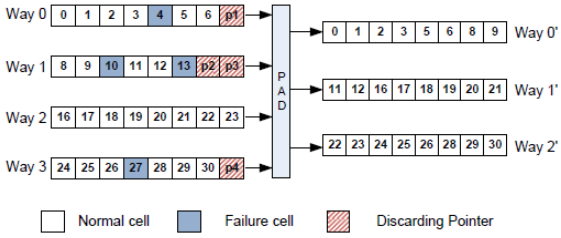
شکل(۱۱) الگوریتم کلی عملیات نوشتن درSAFER [۳۰]

روش دیگر تصحیح خطا، روش ارائه شده در [۳۱] است که از روش‌های افزایش قابلیت اطمینان استفاده می‌کنند که حافظه‌های غیر فرار را در مقابل خطاهای سخت[۱۵] که در اثر تعداد نوشتن زیاد به وجود می‌آید تحمل‌پذیر نمایند. روش ECP[۱۶] برای هر بلوک داده یک اشاره‌گر دارد که آدرس سلول خراب را ذخیره می‌کند و مقدار صحیح آن را هم در بیت دیگری ذخیره می‌نماید. و بلافاصله وقتی که هنگام نوشتن، تشخیص داده شد که در بیتی خرابی رخ داده است، با استفاده از این اشاره‌گرها آن را تصحیح می‌کند. و بدین ترتیب طول عمر حافظه را افزایش می‌دهد. شکل(۱۲) نمونه ای از نحوه‌ی کار ECP را نشان می‌دهد.



شکل(۱۲) نمونه ای از نحوه‌ی کار ECP [۳۱]

همچنین در [۲۴] روش PAD[۱۷] برای تحمل‌پذیر کردن حافظه‌ی نهانِ لایه‌ی آخر در مقابل خطاهای سخت ارائه شده‌است. ایده‌ی استفاده شده در این مقاله دور انداختن بایت‌های خراب به جای تصحیح خطاست. به این صورت که بایت‌های خرابِ هر خط از حافظه‌ی نهان، استفاده نمی‌شود و در عوض بایت‌های سالم از خطوط مختلف در کنار هم قرار می‌گیرند و یک مسیر[۱۸] جدید می‌سازند.دسترسی به یک خط از حافظه‌ی نهان، به صورت سخت‌افزاری و توسط یک شیفت دهنده‌ی چند سطحیِ شامل چند مالتی پلکسر انجام می‌شود. هر مسیر همچنین شامل چند اشاره‌گر هست که آدرس بایت‌های خراب را نگه می‌دارد و مقادیر همین اشاره‌گرهاست که تعیین می‌کند که چند بیت برای رسیدن به داده‌ی درست باید شیفت داده‌شود. ساختار ارائه شده به نحوی بوده که کارایی سیستم را خراب نکند اما باتوجه به اینکه به مرور زمان تعداد مسیر های سالمِ حافظه‌ی نهان کاهش می‌یابند، کارایی کمی کاهش خواهد یافت اما این میزان در مقابل افزایش چشمگیرِ طول عمر حافظه‌ی نهان قابل چشم‌پوشی است. شکل (۱۳) مثال کوچکی از چگونگی کارکرد روشPAD [۲۱] را نشان می‌دهد.



شکل (۱۳) مثال کوچکی از چگونگی کارکرد روشPAD [۲۱]

۴-۱-۲- روش‌های مبتنی بر کاهش دفعات عمل نوشتن

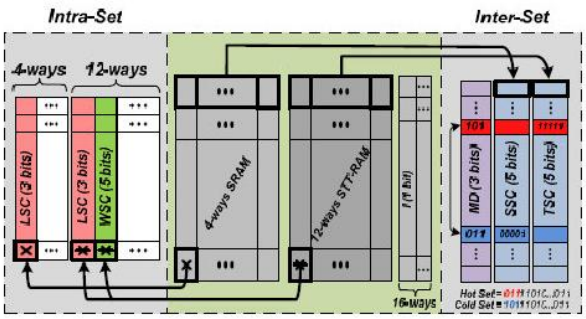
برخی از روش‌هایی که برای کاهش تعداد عملیات نوشتن بیان شده‌اند سعی می‌کنند که از نوشتن‌های غیر ضروری در حافظه‌ی غیر فرار جلوگیری کنند. به عنوان نمونه در روش ارائه شده در [۱۶] سعی شده که فقط خطوطی از حافظه‌ی نهان که در آنها تغییری ایجاد شده‌است نوشته بشوند. قبل از آن هم در روش‌هایی مشابه اما در سطح کوچکتر، یعنی نوشتن بیت‌ها، از این رویکرد استفاده شده‌است. در اغلب این روش‌ها، با استفاده از روش خواندن قبل از نوشتن، سعی کرده اند که قسمت‌هایی که با مقدار قبلی فرق دارد را بیابند و بدین ترتیب تعداد عملیات نوشتن را کاهش بدهند. در واقع می‌توان گفت که این روش‌ها که بر مبنای حذف عملیات غیرضروری نوشتن هستند، مشابه روش-هایی می‌باشند که پیش تر برای کاهش توان مصرفی پویا استفاده می-شد[۳۳].

در [۳۴] ، برای کاهش تعداد عملیات نوشتن روی سلول های حافظه‌ی PCM، از روش معکوس کردن داده ها استفاده شده است. هنگام نوشتن داده ها در یک بلوک حافظه‌ی نهان، ابتدا داده‌ی کنونی آن خوانده می شود و فاصله ی همینگ آن دو حساب می شود که اگر مقدار حساب شده از بزرگتر از نصف اندازه‌ی بلوک حافظه‌ی نهان باشد، داده‌ی جدید را معکوس کرده و بعد در حافظه‌ی نهان ذخیره می‌کنیم.

در دسته‌ای دیگر از روش‌ها که در این بخش به آنها می‌پردازیم، سعی شده‌است تا به منظور کاهش میزان فرسایش سلول‌های حافظه‌ی غیرفرار و نیز جلوگیری ازکاهش کاراییِ سیستم، تعداد عملیات نوشتن در آن کمتر انجام بشود. دربیشتر این نوع روش‌ها یک معماری ترکیبی ارائه شده‌است که بخشی از حافظه، از جنس حافظه‌ی غیرفرار و بخشی دیگر از جنس SRAM است و به طور کلی سعی می‌شود که حجم زیادی از عملیات نوشتن روی SRAM و بخشی کمی از آن روی حافظه‌ی غیرفرار باشد. اما در دسته‌ای دیگر از روش‌ها هم که می‌توان گفت مشابه روش قبلی هستند، سعی شده با بافر کردن داده‌هایی که قرار است داخل حافظه‌ی غیرفرار نوشته بشود، هم کارایی سیستم حفظ شود و هم تعداد دفعات نوشتن کاهش یابد.در ادامه مختصرا به توضیح برخی از این روش‌ها می‌پردازیم.

۴-۱-۲-۱- مدل‌های ترکیبی حافظه

در [۳۵] یک مدل ترکیبی از SRAMو STTRAM برای حافظه‌ی نهان سطح دو[۱۹] ارائه شده است. ایده‌ی کلی که این مقاله از آن استفاده نموده، این است که بلوک‌هایی که در آن‌ها عملیات نوشتن بیش از بقیه انجام می‌شود شناسایی شوند و در این قسمت‌ها از حافظه‌ی SRAM استفاده شود. برای این کار لازم است که تعداد عملیات نوشتن هر خط از حافظه‌ی نهان شمارش بشود. از دو روش نگاشت بین مجموعه[۲۰]و نگاشت داخل مجموعه[۲۱] در پیاده سازی آن استفاده شده است. یکی برای اینکه بلوک هایی که تعداد نوشتن آنها زیاد است در SRAM قرار بگیرند و دیگری برای اینکه عملیات نوشتن در STTRAM ها به صورت یکنواخت توزیع گردد. شمارنده هایی که در این مقاله از آن استفاده شده است همه از جنس SRAM هستند. شکل (۱۴)معماری حافظه‌ی نهان ترکیبی[۳۵] را نشان می‌دهد.



شکل (۱۴)معماری حافظه‌ی نهان ترکیبی[۳۵]

در این معماری از یک شمارنده‌ی اشباع شونده به نام LSC به منظور شمارش تعداد عملیات نوشتن در هر مجموعه از حافظه‌ی نهان در نظر گرفته شده است. این شمارنده‌ها نشانگر چگونگی توزیع عملیات نوشتن در سطح هر بلوک می‌باشند. هر بار که یک نوشتن در یک خط انجام می‌شود، شمارنده‌ی LSC مربوط به آن یک واحد افزایش می‌یابد و اشباع شدن این شمارنده نمایانگر این است این خط بسیار مستعد به نوشتن است و اگر جنس آن STT-RAM است، با یک خط از جنس SRAM که دارای کمترین مقدار شمارنده است، جابجا بشود. در این مقاله نشان داده شده‌است که به طور متوسط، ۷۱ درصد ترافیک انتقال به سمت SRAM و ۲۱درصد انتقال به سمت STT-RAM داریم. این روش تا ۴۹ برابر طول عمر حافظه‌ی نهانِ از جنس STT-RAM را افزایش و تا ۵۰ درصد توان مصرفی آن را کاهش می‌دهد.

مدل دیگری برای حافظه‌ی نهان در[۲۵] ارائه شده است. در این مقاله یک حافظه‌ی نهان ساخته شده از جنس SRAM و حافظه‌ی غیرفرار به صورت ترکیبی، به منظور کاهش توان مصرفی است. همچنین روشی که برای کاهش توان مصرفی از آن استفاده شده‌است، این است که قسمت‌هایی از حافظه که نیازی به آنها نیست خاموش می‌شوند، البته با توجه به اینکه توان نشتی حافظه‌های غیرفرار ذاتا بسیار کم است نیاز

به داشتن سخت افزار اضافه برای خاموش کردن آنها نیست. ایده‌ی استفاده شده در این مقاله به این صورت است که تعدادی از مسیرهای یک مجموعه از حافظه نهان از جنس SRAMو تعدادی دیگر از جنس حافظه‌ی غیرفرار هستند.اما با توجه به سرعت پایین عملیات نوشتن در حافظه‌های غیر فرار، برچسب‌ها[۲۲] به طور کامل از جنس SRAM ساخته می‌شود. حال مساله‌ای که وجود دارد نحوه‌ی مدیریت نوشتن-هاست به طوری که تعداد زیادی عملیات خاموش و روشن کردن بخشی از حافظه کم باشد زیرا بازگشت از حالت خاموش، خود شارژ و دشارژ شدن خازن‌های حافظه‌ی SRAM و مصرف توان را در پی دارد. و از طرفی استفاده از حافظه‌ی غیر فرار نباید کارایی حافظه را خراب کند. مدیریت حافظه‌ی نهان در این مقاله به صورت سخت‌افزاری و با استفاده از چند بیت شمارشگر برای مسیرهای مختلف، پیاده‌سازی شده است. این روش تنها با کاهش ۴ درصدی کارایی، انرژی مصرفی را تا ۲۵ درصد کاهش داده است.

در [۳۶] برای پیاده‌سازی حافظه‌ی نهان سطح دو، از ترکیبی از SRAM و حافظه‌ی غیر فرار استفاده شده‌است. و مانند کارهای قبلی که به آنها اشاره شد، در یک مجموعه[۲۳]، برخی از مسیرها از جنس حافظه‌ی غیر فرار و برخی دیگر از جنس از SRAM هستند. اما نکته-ی جدیدی که در این مقاله وجود دارد این است که برای اینکه بتواند از فضای کوچک نظر گرفته شده برای SRAM به خوبی استفاده نماید از یک ساختار داینامیک استفاده کرده به نحوی که هر مجموعه می-تواند از مسیرهای مجموعه‌های دیگر نیز استفاده نماید. علاوه بر این برای رسیدن به توان مصرفی و کارایی مطلوب، مدیریت نوشتن در این حافظه به نحوی انجام می‌شود که بلوک‌هایی بیش از یک بار در آنها نوشته می‌شود از STTRAM به SRAM بروند و همچنین اگر نوشتن در بلوکی در کارایی تاثیر زیادی ندارد، از SRAM به STTRAM برود. مدل قرار گرفتن حافظه‌های غیرفرار در این مقاله به صورت سه بعدی است. اما از نظر منطقی روشی که ارائه شده است فرقی با حالت دو بعدی ندارد و فقط گفته شده که به دلیل مساحت کمتر، از مدل سه بعدی استفاده کرده‌اند. این روش توان مصرفی را تا ۵۴درصد کاهش و کارایی را ۱,۱۶ درصد بهبود بخشیده‌است.

در [۲۹] هم یک ساختار ترکیبی از SRAM و حافظه‌ی غیر فرار برای حافظه‌ی نهان سطح دو معرفی می‌کند. ایده‌ای که دراین مقاله از آن استفاده شده‌است پیش‌بینی کردن دسترسی ها به حافظه‌ی نهان با توجه به الگوی دسترسی به آن است به نحوی که تعداد دفعات نوشتن در PCM کنترل شود و از فرسایش آن جلوگیری به عمل آید. دو الگویی داده‌های حافظه‌ی نهان از آنها بیشتر تبعیت می‌کنند و در این مقاله از آنها استفاده شده‌است یکی داده‌هایی هستند که هنگام اتفاق افتادن یک عدم اصابت[۲۴] از حافظه‌ی نهان سطح یک به حافظه‌ی نهان سطح دو منتقل می‌شوند و برای بار دوم توسط حافظه‌ی نهان سطح یک نوشته یا خوانده نمی‌شوند و دسته‌ی دوم آنهایی هستند که به تعداد دفعات زیاد توسط حافظه‌ی نهان سطح یک تغییر می‌کنند. که برای PCM این داده‌ها مناسب نیستند چون کارایی را خراب می‌کنند. پیاده‌سازی این روش به صورت سخت‌افزاری و با اضافه کردن چند بیت به حافظه‌ی نهان سطح دو انجام می‌شود تا بر حسب مقدار آنها تصمیم گرفته شود که داده‌ای که از حافظه‌ی نهان سطح دو خارج می‌شود به حافظه‌ی اصلی بازگردد و یا در PCM نوشته شود. با توجه به اینکه قرار دادن فناوری‌های مختلف حافظه‌در کنار هم از نظر فرآیند ساخت پیچیده می‌باشد، PCM در لایه‌ی بالای پردازنده، حافظه‌ی نهان سطح یک و دوی از جنس SRAM قرار می‌گیرد. این روش تعداد نوشتن‌ها را ۸۵,۵ درصد کاهش و با کاهش ۶,۴ درصدی کلاک بر دستور، کارایی سیستم را بالا می‌برد.

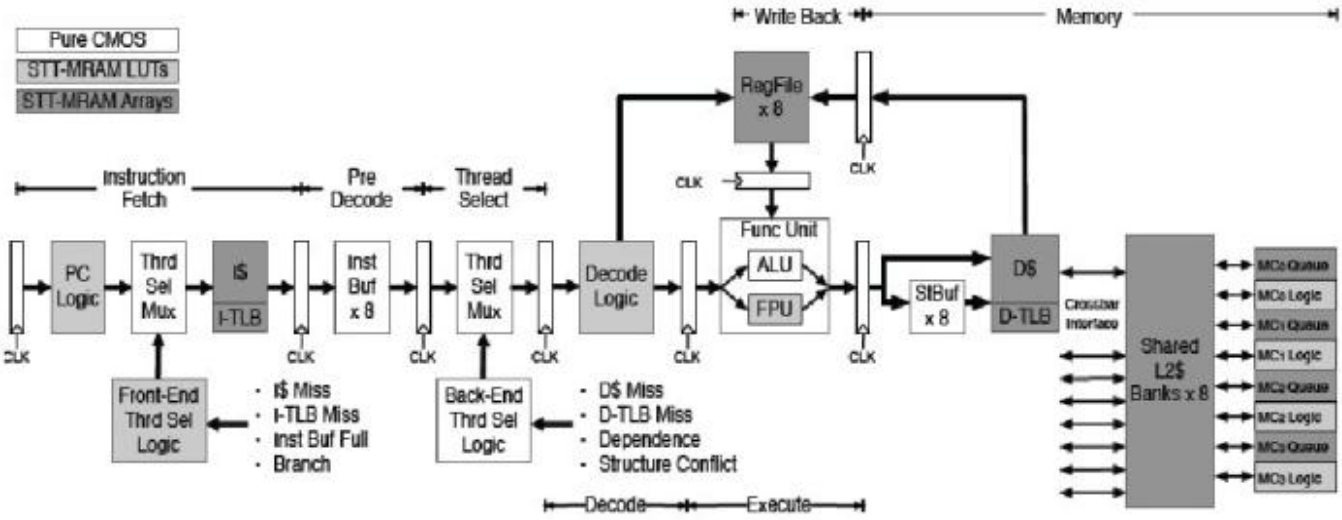
۴-۱-۲-۲- استفاده از بافر نوشتن

در[۳۷] سعی شده است در یک پردازنده‌ی دارای خط لوله از هر یک از ویژگی‌های خانواده‌های CMOS و STT-RAM، به نحو احسن استفاده بکند. یعنی در برخی از قسمت‌ها از STT-RAM و در برخی قسمت‌های دیگر از CMOS استفاده شده است. در شکل (۱۶) ساختار یک خط لوله‌ی ترکیبی از فناوری‌های CMOS و حافظه‌ی غیرفرار [۳۷] نشان داده شده است.

یکی از قسمت‌هایی که برای جایگزین شدن با STT-MRAM انتخاب شده‌است بانک ثبات است. اما با توجه به یک مشکل اساسیِ طولانی بودن زمان نوشتن در STT-MRAM، در استفاده از آن باید همیشه توجه بکنیم که هیچ عملیات خواندنی به دلیل اینکه یک عمل طولانیِ نوشتن در حال انجام است، به تاخیر نیفتد. همچنین این نوشتن‌های طولانی بازدهیِ خط لوله را خراب نکند. یک راه حل برای جلوگیری از این دو مشکل، که در این مقاله از این روش استفاده شده‌است، افراز بانک ثبات است. این افراز به صورت شکل زیر انجام می‌شود:

شکل(۱۵) استفاده از بافر نوشتن در بانک ثبات افراز شده[۳۷]

بافرهایی که در کنار هر یک از قسمت‌ها ملاحظه می‌شود به منظور بالا بردن کارایی عمل نوشتن، بدون افزایش سیم‌کشی[۲۵] است. با وجود این بافر ها می‌توان چند عمل نوشتن را به صورت موازی انجام داد. این بافرها که فقط برای درگاه‌های نوشتن هستند، داده‌ی ورودی آدرس را در خود نگه می‌دارند. در صورت نیاز می‌توان یک داده را از همین بافرها خواند. یک مزیت این روش این است که می‌توان فقط نوشتن-هایی که مقادیر داخل ثبات را تغییر می‌دهند را بنویسیم. به این صورت که هر بار داده‌ی قبلی با بعدی XOR می‌شود و بیت‌هایی که



شکل (۱۶) ساختار یک خط لوله‌ی ترکیبی از فناوری‌های CMOS و حافظه‌ی غیرفرار [۳۷]

متفاوت هستند نوشته می‌شوند. این روش که به نوشتن تفاضلی[۲۶] معروف است می‌توان انرژی مصرفی را کاهش داد.

در هر کلاک، یک دستور که در سر یکی از بافرهای دستور قرار دارد را برای رمزگشایی و اجرا شدن فرستاده می‌شود. یکی از اتفاقاتی که می‌تواند باعث شود که یک ریسمان[۲۷] زمان‌بندی نشود، این است که دو عمل نوشتن بخواهند در یک زیرمجموعه‌ی رجیسترفایل بنویسند. برای مثال اگر زمان نوشتن در بانک ثبات سیزده سیکل باشد، در مدت این سیزده سیکل، یک دستور دیگر نمی‌تواند داخل همین زیر-بانک بنویسد اما یک عمل خواندن دیگر می‌تواند به صورت همزمان انجام بشود. پس اگر مقصد یک عمل نوشتن و یک عمل نوشتن که در حال اجرا است، یک زیر-بانک باشد، آن ریسمان، تا زمانی که مقصدش آزاد شود، زمان‌بندی نمی‌شود. برای افزایش بازدهی و نیز کاهش شانس ایجاد ناسازگاری در یک بانکِ رجیستر، رجیسترهای مورد نیازِ یک ریسمان بین بانک‌های مختلف پخش می‌شود. یک مرحله‌ی دیگر پایپ‌لاین که در آن به رجیسترفایل دسترسی داریم، مرحله‌ی دوباره نوشتن[۲۸] است. در این مرحله امکان وجود ناسازگاری وجود ندارد، زیرا قسمت انتخاب ریسمان، قبلا دستورات را با در نظر گرفتن ناسازگاری‌ها زمان‌بندی کرده است. همچنین نوشتن تفاضلی در این مرحله می‌تواند توان مصرفی را کاهش دهد.

در تحقیق دیگری در [۳۸] ، ساختار استفاده شده برای حافظه‌ی نهان،PCM است و برای بهبود مشکل پایداریِ آن از یک بافر نوشتن از جنس STTRAM استفاده شده است و با سرآیند سخت‌افزاری ۴ درصدی، طول عمر حافظه‌ی نهان را تا ۳۹ درصد افزایش داده است.

۴-۱-۳- توزیع یکنواخت نوشتن‌ها در کل حافظه

همان‌طور که در بخش‌های قبل اشاره شد، برخی از مدل‌های ترکیبی سعی کرده اند برای جلوگیری از فرسایش یک بخش خاص از حافظه، عملیات نوشتن را در کل حافظه به صورت یکنواخت پخش کنند. در این بخش به صورت دقیق تر به توضیح روش‌هایی که در آنها سعی شده تا عملیات نوشتن در حافظه به شکل یکسان توزیع گردد می‌پردازیم. در [۳۳] یک روش نگاشتِ ریزدانه، با کمک کدهای تصحیح خطا و اشاره‌گرهای تعبیه شده، به نام FREE-p ارائه شده‌است. این روش در سطح کنترل کننده‌ی حافظه پیاده‌سازی می‌شود و مستقل از ساختار حافظه است. این روش طول عمر حافظه را تا ۲۶ درصد افزایش داده‌است و کاهش کارایی بسیار کم و در حدود کمتر از ۲درصد بوده است.روش‌های رایج توزیع عملیات نوشتن به این ترتیب هستند که در هر یک از بلوک‌های حافظه می‌توانیم تعداد محدودی عمل نوشتن داشته باشیم و هر وقت به این حد نصاب رسید، این بلوک غیر فرار و به یک بلوک دیگر نگاشت می‌شد. مشکل این روش‌ها این است که درسطح سیستم عامل انجام می‌شوند؛ لذا نگاشت‌ها و در مقیاس بزرگ، یعنی در اندازه‌ی صفحه انجام می‌شود. اما در روش FREE-p نگاشت، در مقیاس بلوک‌های کوچکتری انجام می‌شود. با توجه به اینکه معمولا تعدادی بیت پشت سر هم از یک صفحه خراب می‌شوند و بیت‌های دیگر سالم می‌ماند، این روش بسیار موثرتر از روش‌های قبلی عمل خواهد کرد.

۴-۱-۴- روش پیشنهادی ارائه شده برای استفاده از حافظه‌های غیر فرار در بانک ثبات

با توجه به اینکه بانک ثبات از مهم‌ترین حافظه‌های داخل پردازنده است و بخش زیادی از مصرف توان پردازنده مربوط به آن می‌باشد[۳۹-۴۱] لذا یکی دیگر از مواردی است که می‌تواند از جنس حافظه‌های غیر فرار باشد. همچنین موقعیت رجیسترفایل در داخل پردازنده به نحوی است که با کمک دیگر اجزای داخل پردازنده، مانند بافر تغییر اولویت[۲۹] و جدول تغییر نام[۳۰] با سرآیند سخت‌افزاری بسیار

کم، می‌توان نوشتن‌های درون آن را کنترل نمود و بدین ترتیب مانع فرسایش سریع آن شد. علاوه بر این، داده‌هایی که داخل بانک ثبات نوشته می‌شوند از یکسری الگو تبعیت می‌کنند که با توجه به آن‌ها الگوریتمی برای کاهش تعداد عملیات نوشتن در بانک ثبات ارائه گردیده است. این الگوها به ترتیب زیر هستند:

- بیش از ۸۰ درصد مواقع، مقادیر ثبات‌ها طول عمر کوتاهی دارند و بعد از چند سیکل با مقدار دیگری جایگزین می‌شوند[۴۲].
- حدود ۹۵ درصد از این مقادیرِ با طول عمر کم، یک مصرف‌کننده دارند و یا اصلا مصرف کننده‌ای ندارند.
- در ۸۰ درصد مواقع بین دوبار نوشتن در یک ثبات، عملیات پرش وجود ندارد.

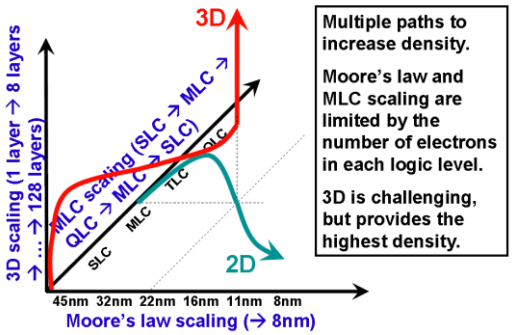
با توجه به سه الگوی فوق می‌توان دریافت که نیازی به نوشتن مقادیری که طول عمر آنها کوتاه است نیست. و مشکل تنها زمانی پیش می‌آید که دستوری بخواهد از مقداری که داخل بانک ثبات ننوشته‌ایم استفاده نماید، اما با توجه به الگوی دوم، که در اکثر مواقع هر مقداری، تنها یک مصرف‌کننده دارد، در مواردی به جز این می‌توان مقدار ثبات را در بانک ثبات ذخیره نمود و چون تعداد دفعاتی که این حالت پیش می‌آید کم است، طول عمر بانک ثبات را زیاد تحت تاثیر قرار نمی‌دهد.همچنین هر بار که پرشی به جایی که پیش‌بینی نشده است رخ می‌دهد، باید هر آنچه داخل بافر تغییر اولیت است تخلیه شود. دراین صورت اگر مصرف کننده‌ای بعد از عملیات پرش در بافر تغییر اولیت باشد، و عملوندش هم قرار باشد در بانک ثبات نوشته نشود، بعد از ورود دوباره به بافر تغییر اولیت، با مشکل روبرو خواهد شد، لذا در چنین مواردی هم باید ثبات در بانک ثبات نوشته شود ولی در این حالت هم با توجه به الگوی سوم، احتمال پیشامد کم است و بنابراین صدمه‌ی زیادی به طول عمر بانک ثبات نمی‌زند. لذا در روش پیشنهاد شده، بدون استفاده از سرآیند سخت‌افزاری زیاد و تنها با اضافه کردن دو بیت به ساختار بافر تغییر اولویت، توانسته ایم داده‌هایی که از الگوهای نام برده شده تبعیت می‌کنند را تشخیص دهیم و مانع از نوشته شدن آن‌ها داخل بانک ثبات بشویم و بدین ترتیب درصد زیادی از نوشتن‌های بانک ثبات را کاهش خواهیم داد.

۵- روند معماری پردازنده‌ها در سال‌های آینده و چالش‌های حافظه در آن‌ها

همواره طراحان پردازنده‌ها به دنبال افزایش سرعت به کمک روش‌های گوناگون بوده و هستند. اما با به اشباع رسیدن میزان افزایش فرکانس کلاک پردازنده‌ها، محققان در سال‌های گذشته برای افزایش سرعت پردازنده‌ها به سمت روش‌هایی نظیر ILP رفته بودند. اما در سال‌های اخیر این روش‌ها هم تا حد امکان پیشرفت کرده‌اند. لذا روند طراحی پردازنده‌ها به سمت استفاده از چند پردازنده در یک تراشه(CMP)، رفته است. اما با کنار هم قرار گرفتن چندین هسته و افزایش سرعت پردازنده، مساله‌ی دیوار حافظه بسیار پررنگ تر و مهم‌تر خواهد بود[۴۳] و چالش‌های عمیق‌تری در مورد حافظه‌های غیر اشتراکی و اشتراکی بیت هسته‌های مختلف، نحوه‌ی به اشتراک گذاشتن اطلاعات بین آنها مطرح می‌شود. همچنین انتخاب حافظه‌ی مناسب از نظر قابلیت اطمینان،کارایی، چگالی و توان مصرفی بسیار حائز اهمیت خواهد بود. برای دستیابی به ظرفیت زیاد حافظه، روش‌های مختلفی تا کنون ارائه شده‌اند. یکی از این روش‌ها طراحی‌های سه بعدی[۳۱] بوده‌بوده‌اند. در این روش سعی می‌شود که حافظه‌ها در لایه‌ای روی هسته‌ها قرار بگیرند و به این ترتیب مشکل پهنای باند دسترسی به حافظه‌را برای سیستم‌های چند هسته‌ای تا حدودی بهبود ببخشند. تحقیقاتی هم تا کنون برای استفاده از DRAM و SRAM در مدل‌های سه‌بعدی انجام شده‌است[۴۴]. همان‌طور که پیشتر اشاره شد، STT-RAM یکی از انواع حافظه‌های غیرفرار است که با توجه به چگالی بالا گزینه‌ی مناسبی برای استفاده در حافظه‌های داخل تراشه از جمله حافظه‌ی نهان است. اما مساله‌ای که وجود دارد نحوه‌ی ساخت و قرار دادن این نوع حافظه در کنار هسته‌های پردازنده به صورت دو بعدی‌است. خوشبختانه، مدل سه بعدی، این امکان را فراهم نموده تا بتوان با صرف هزینه‌ی کمتر، فناوری‌های مختلف را در کنار هم قرار داد[۲۷, ۲۹, ۳۶]. به طور خلاصه، راه‌حل‌هایی که برای افزایش چگالی حافظه‌ها وجود دارد در سه دسته قرار می‌گیرند:

- کوچک کردن ابعاد سلول حافظه با توجه به قانون مقایس‌پذیری مور[۳۲][۲۰]
- سلول‌های چند سطحی[۳۳] و یا چند بیت در یک سلول[۲۰]
- روش‌های طراحی سه بعدی

روش دوم به دنبال محدودیت‌هایی که برای روش اول وجود دارد و پیش‌تر نیز به آنها اشاره شد، مطرح شده‌است. اما به هر حال هر دوی این روش‌ها با محدودیت تعداد الکترون‌ها در هر سطح منطقی مواجه هستند. و البته روشMLC برای برخی از انواع حافظه‌های غیر فرار مثل STT-RAM قابل پیاده‌سازی نیست. و تنها برای PCM و حافظه‌های فلش مناسب می‌باشد. لذا روش سوم که استفاده از حافظه در یک لایه بر روی هسته‌ها می‌باشد، علارغم داشتن چالش‌های بسیار زیاد مخصوصا در مورد نحوه‌ی مدیریت آن، محدودیت چگالی بسیار کمتری دارد و پیش‌بینی می‌شود که در سال‌های آینده بسیار مورد توجه طراحان قرار بگیرد. شکل(۱۷) مقایسه‌ی روش‌های مختلف برای افزایش چگالی حافظه را نشان می‌دهد.



شکل(۱۷) مقایسه‌ی روش‌های مختلف برای افزایش چگالی حافظه[۲۰]

۶- نتیجه گیری

همان‌طور که در این مقاله به آن اشاره شد، تا کنون در کارهای تحقیقاتی بسیاری سعی شده‌است تا برای افزایش قابلیت اطمینان حافظه‌های درون تراشه، از جمله حافظه‌ی نهان، که یکی از مهم‌ترین آنهاست، از حافظه‌های غیرفرار استفاده نمایند.با این رویکرد، حافظه‌ی مورد نظر را کاملا در برابر خطاهای نرم مقاوم کرده‌اند. اما از طرفی با توجه به اینکه این نوع حافظه‌ها توان مصرفی ایستای بسیار ناچیزی دارند، ملاحظه شد که برخی معماری‌های ترکیبی ارائه شده، توانسته اند توان مصرفی را تا حد خوبی کاهش دهند. این تحقیقات نشان می‌دهد که در استفاده‌ی حافظه‌های غیرفرار در پردازنده، مساله‌ی توان مصرفی نقطه‌ی مقابل قابلیت اطمینان نیست و با روش‌های نام برده شده می‌توان از هر دو لحاظ به معماری بهتری دست یافت. اما همان‌طور که اشاره شد، در سال‌های آتی معماری پردازنده‌ها به سمت مدل‌های چند هسته‌ای پیش میرد که در آن‌ها مساله‌ی جایگزین نمودن حافظه‌های غیرفرار چالش‌های جدیدی در بردارد، زیرا در آنها برخی از حافظه‌ها به صورت اشتراکی بین هسته‌ها قرار دارند و مدیریت حافظه‌ی اشتراکیِ از جنس حافظه‌های غیرفرار، با در نظر گرفتن تعداد دفعات محدود نوشتن، باید به نحوی باشد که مصالحه‌ی مناسبی بین کارایی و قابلیت اطمینان برقرار شود. نحوه‌ی چینش حافظه‌های از جنس این فناوری‌های جدید، در کنار پردازنده‌های مدرن از نظر فناوری ساخت، از دیگر چالش‌هایی است که در این پروژه مورد بررسی قرار خواهند گرفت

مراجع

[۱] S. A. McKee, "Reflections on the memory wall," in *CF*, ۲۰۰٤, p. ۱٦۲.

[۲] L. G. L. Li , J. Xue, "Memory Coloring: A Compiler Approach for Scratchpad Memory Management," presented at the PACT, ۲۰۰٥.

[۳] D. P. J. Hennessy *Computer Architecture: A Quantitative Approach*, first ed.: Morgan Kaufmann, ۱۹۹۰.

[٤] T. Sakutai, "Perspectives on power-aware electronics," presented at the ISSCC, ۲۰۰۳.

[٥] C. L. H. Sun, W. Xu, J.Zhao, N. Zheng, T. Zhang, "Using Magnetic RAM to Build Low-Power and Soft Error-Resilient L۱ Cache" *TVLSI*, vol. ۲۰, pp. ۱۹-۲۸, ۲۰۱۲.

[٦] E. d. G. R. Baert, E. Brockmeyer "An automatic Scratch Pad Memory management tool and MPEG-٤ encoder case study," presented at the DAC, ۲۰۰۸.

[۷] P. Shivakumar, "Modeling the Effect of Technology Trends on Soft-Error Rate of Combinational Logic," presented at the DSN, ۲۰۰۲.

[۸] H. T. E. Ibe, Y. Yahagi, K. Shimbo, T. Toba, "Impact of Scaling on Neutron-Induced Soft Error in SRAMs From a ۲٥۰ nm to a ۲۲ nm Design Rule," *IEEE Transactions on Electronic Devices*, vol. ٥۷, pp. ۱٥۲۷-۱٥۳۸, ۲۰۱۰.

[۹] V. Z. N. Seifert, "Assessing the Impact of Scaling on the Efficacy ofSpatial Redundancy based Mitigation Schemes for Terrestrial Applications," in *SELSE*, ۲۰۰۷.

[۱۰] H. P. D. Radaelli, P. Chia, S.Wong, S. Daniel, "Investigation ofMulti-bit Upsets in a ۱٥۰nm Technology SRAM Device," presented at the NSREC, ۲۰۰٥.

[۱۱] A. W. R. Heald, "Trends from Ten Years of Soft Error Experimentation," in *SELSE*, ۲۰۰۹.

[۱۲] M. Verma, Wehmeyer, L., Marwedel, P., "Cache-Aware Scratchpad-Allocation Algorithms for Energy-Constrained Embedded Systems," *IEEE Transactions on Computeer-Aided Design of Integrated Circuits and Systems*, vol. ۲٥, ۲۰۰٦.

[۱۳] G. W. B. S. Raoux, M. J. Breitwisch , C. T. Rettner, Y.-C. Chen , R. M. Shelby , M. Salinga , D. Krebs , S.-H. Chen , H.-L. Lung , C. H Lam "Phase-change random access memory: A scalable technology," *IBM Journal of Research and Development*, vol. ٥۲, pp. ٤٦٥-٤۷۹, ۲۰۰۸.

[۱٤] J. M. W. Geoffrey "Phase change memory technology," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, vol. ۲۸, pp. ۲۲۳-۲٦۲, ۲۰۱۰.

[١٥] A. F. C. T. Perez "Volatile Memory: Emerging Technologies And Their Impacts on Memory Systems," ٢٠١٠.

[١٦] E. I. B. C. Lee , O. Mutlu, D. Burger, , "Architecting phase change memory as a scalable dram alternative," in *ISCA*, ٢٠٠٩, pp. ٢-١٣.

[١٧] C. S. K. M.H. Kryder "After Hard Drives—What Comes Next?," *IEEE Transactions on Magnetic*, vol. ٤٥, pp. ٣٤٠٦-٣٤١٣, ٢٠٠٩.

[١٨] R. Williams, "How We Found The Missing Memristor," *IEEE Spectrum*, vol. ٤٥, pp. ٢٨-٣٥, ٢٠٠٨.

[١٩] X. W. X. Dong , G. Sun ,Y. Xie ,H. Li ,Y. Chen, "Circuit and microarchitecture evaluation of ٣D stacking magnetic RAM (MRAM) as a universal memory replacement," presented at the DAC, ٢٠٠٨.

[٢٠] n. T. R. f. S. (ITRS), "ERD_ERM ٢٠١٠ FINAL Report Memory Assessment," ٢٠١٠.

[٢١] Y. C. H. Li "An overview of non-volatile memory technology and the implication for tools and architectures," presented at the DATE, ٢٠٠٩.

[٢٢] K. Nomura, "Ultra low power processor usingperpendicular-STTMRAM/SRAM based hybrid cachetoward next generation normally-off computers," presented at the MMM, ٢٠١١.

[٢٣] H. Chung, "A ٥٨nm ١,٨V ١Gb PRAM with ٦,٤MB/s program BW," presented at the ISSCC, ٢٠١١.

[٢٤] X. D. J. Wang, Y. Xie, "Point and discard: a hard-error-tolerant architecture for non-volatile last level caches," presented at the DAC, ٢٠١٢.

[٢٥] J. C. Y. Chen, H. Huang, B. Liu, C. Liu, M. Potkonjak and G. Reinman, "Dynamically Reconfigurable HybridCache: : An Energy-Efficient Last-Level Cache Design," presented at the DATE, ٢٠١٢.

[٢٦] J. L. X. Wu, L. Zhang, E. Speight, R. Rajamony, and Y. Xie, "Hybrid Cache Architecture with Disparate Memory Technologies," presented at the ISCA, ٢٠٠٩.

[٢٧] X. D. G. Sun, Y. Xie, J. Li, and Y. Chen, "A Novel Architecture of the ٣D Stacked MRAM L٢ Cache for CMPs," presented at the HPCA, ٢٠٠٨.

[٢٨] D. C. M. Rasquinha, S. Chatterjee, S. Mukhopadhyay, and S. Yalamanchili, "An Energy Efficient Cache Design Using Spin Torque Transfer (STT) RAM," presented at the ISLPED, ٢٠١٠.

[٢٩] Z. L. S Guo, D Wang, H Wang, "Wear-Resistant Hybrid Cache Architecture with Phase Change Memory," presented at the NAS, ٢٠١٢.

[٣٠] D. W. N. Seongy, V. Srinivasanz ,Jude A. ,HsienHsin ,S. Lee, "SAFER, Stuck at Fault Error Recovery for memories," presented at the MICRO, ٢٠١٠.

[٣١] H. G. L. S. Schechter, K. Strauss, D. Burger, "Use ECP, not ECC, for Hard Failures in Resistive Memories," presented at the ISCA, ٢٠١١.

[٣٢] X. D. AK. Mishra, G. Sun, Y. Xie, "Architecting On-Chip Interconnects for Stacked ٣D STT-RAM Caches in CMPs," presented at the ISCA, ٢٠١١.

[٣٣] N. M. DH. Yoon, J. Chang, "FREE-p: Protecting Non-Volatile Memory against both Hard and Soft Errors," presented at the HPCA, ٢٠١١.

[٣٤] Y. Joo, "Energy- and Endurance-Aware Design of Phase Change Memory Caches," presented at the DATE, ٢٠١٠.

[٣٥] M. A. A. Jadidi, and H. Sarbazi-Azad, "High-endurance and performance-efficient design of hybrid cache architectures through adap- tive line replacement," presented at the ISLPED, ٢٠١١.

[٣٦] L. S. J. Li, CJ. Xue, C. Yang, Y. Xu, "Exploiting Set-Level Write Non-Uniformity for Energy-Efficient NVM-Based Hybrid Cache," presented at the ESTIMedia, ٢٠١١.

[٣٧] E. I. X. Guo, T. Soyata, "Resistive computation avoiding the power wall with low-leakage, STT-MRAM based computing," presented at the ISCA, ٢٠١٠.

[٣٨] S. P. Y. Joo, "A Hybrid PRAM and STT-RAM Cache Architecture for Extending the Lifetime of PRAM Caches," *Computer Architecture Letters*, ٢٠١٢.

[٣٩] R. G. R. Nalluri, PR. Panda, "Customization of Register File Banking Architecture for Low Power," presented at the VLSI Design, ٢٠٠٧.

[٤٠] M. F. SN. Ahmadian, NF. Ghalaty, SG. Miremadi, "Value-Aware Low-Power Register File Architecture," presented at the CADS, ٢٠١٢.

[٤١] Y. K. M. O□zsoy, M. Kayaalp, "Dynamic register file partitioning in superscalar microprocessors for energy efficiency," presented at the ICCD, ٢٠١٠.

[٤٢] Y. H. D. She, B. Mesman, "Scheduling for Register File Energy Minimization in Explicit Datapath Architectures," presented at the DATE, ٢٠١٢.

[٤٣] A. Iog, "Cache revive: architecting volatile STT-RAM caches for enhanced performance in CMPs," presented at the DAC, ٢٠١٢.

[٤٤] G. H. Loh, "٣D-Stacked Memory Architectures for Multicore Processors," presented at the ISCA, ٢٠٠٨.

[١] Safety-Critical

[٢] Non-Volatile Memories(NVM)

[۳] Soft error rate per device
[۴] Multiple Cell Upset
[۵] Phase change memories(PCM)
[۶] endurance
[۷] Resistive memories (RRAM)
[۸] Magnetic RAM (STT-RAM)
[۹] Tunneling Magneto- Resistance
[۱۰] Write Word Line
[۱۱] Bit Line
[۱۲] Source line
[۱۳] endurance
[۱۴] wear-leveling
[۱۵] Hard errors
[۱۶] Error Correction Pointer
[۱۷] Point and Discard
[۱۸] way
[۱۹] L۲ cache
[۲۰] intra-set remapping
[۲۱] inter-set remapping
[۲۲] Tag array
[۲۳] set
[۲۴] Cache miss
[۲۵] wiring
[۲۶] Differential write
[۲۷] thread
[۲۸] Write back
[۲۹] Re-Order Buffer(ROB)
[۳۰] Rename table
[۳۱] ۳D
[۳۲] Moore’s law
[۳۳] MLC(Multi Level Cell)

بررسی امنیت در سیستم‌های ذخیره‌سازی مبتنی بر معماری باز

سعید مسلمی نسب[1]

[1]دانشجوی کارشناسی ارشد، دانشکده کامپیوتر، دانشگاه علم و صنعت ایران، تهران،
Mosleminasab@comp.iust.ac.ir

استاد راهنما: دکتر رضا برنگی
استاد مشاور: دکتر احمد پاطوقی

چکیده

با افزایش تقاضا برای مخازن ذخیره‌سازی بزرگ و گسترده، ذخیره‌سازی شبکه‌ای برای ذخیره‌سازی داده‌های انبوه و پردازش آنها، قابل دسترس بودن، کیفیت خدمات و امنیت ذخیره‌سازی داده‌ها، بسیار حائز اهمیت است. در این وضعیت نیاز به ظهور فناوری‌های جدید در زمینه ذخیره‌سازی داده‌ها احساس می‌شود. شبکه منطقه‌ای ذخیره‌سازی، یکی از راه‌حل‌های امیدوار کننده برای رسیدگی به خواسته‌های ذخیره‌سازی سازمان‌های تجاری و مدیریت ذخیره‌سازی داده‌ها است. یکی از وظایف مهم و چالش برانگیز در هنگام طراحی شبکه منطقه‌ای ذخیره‌سازی، رسیدگی به نگرانی‌های امنیتی این شبکه‌ها است و از آنجا که شبکه منطقه‌ای ذخیره‌سازی، اطلاعات را در یک محل متمرکز نگهداری می‌کند، برای حفظ آنها لازم است که اقدامات امنیتی مناسبی را برای مقابله با حملات داخلی و خارجی در نظر گرفت.

کلمات کلیدی

شبکه‌های منطقه‌ای ذخیره‌سازی، تهدیدات، خطرات، امنیت، ذخیره‌سازی امن، شبکه کانال فیبری، وفق‌دهنده گذرگاه میزبان، نام گسترده جهانی، سیستم تشخیص نفوذ.

۱- مقدمه

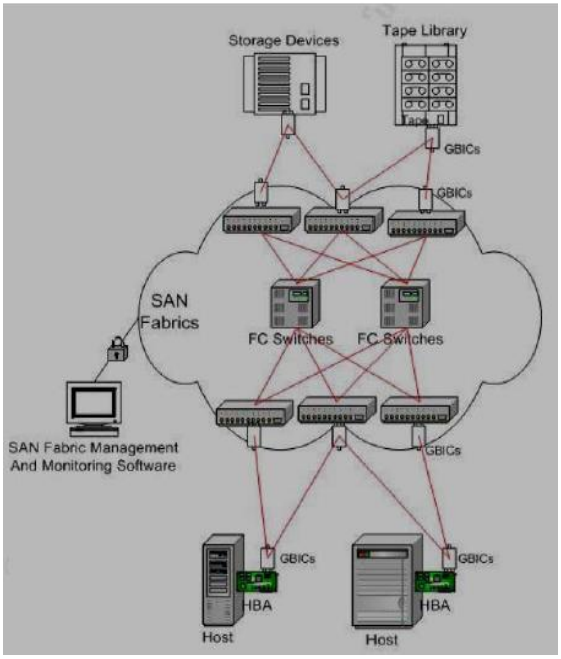
اینترنت یک رسانه جهانی قابل دسترس برای اتصال کامپیوترهای شخصی، سرورها، شبکه‌ها و سامانه‌های ذخیره‌سازی است که به طور فزآینده‌ای نسبت به حملات شبکه‌ای آسیب‌پذیر است [۱]. اطلاعات ارزشمند از جمله مالکیت شخصی، هویت شخصی، و تراکنش‌های مالی در آرایه‌های ذخیره‌سازی قابل دسترس از طریق اینترنت ذخیره می‌شوند و بالطبع بیشتر در معرض تهدیدات امنیتی مختلف قرار می‌گیرند. این تهدیدات می‌توانند به صورت بالقوه به داده‌های مهم آسیب برسانند و خدمات حیاتی را مختل کنند. از این رو تأمین امنیت شبکه‌های منطقه‌ای ذخیره‌سازی[1] تبدیل به یک جزء جدایی‌ناپذیر از فرآیند مدیریت ذخیره‌سازی شده است [۱].

رشد اطلاعات دیجیتال یک واقعیت در زندگی امروز است و منابع اطلاعات دیجیتال روز به روز در حال گسترش هستند و فناوری ذخیره‌سازی نقش مهمی در نگهداری اطلاعات برای استفاده آنها در آینده ایفا می‌کند. سیستم‌های ذخیره‌سازی داده همواره بخش مهمی از زیرساخت فناوری اطلاعات تجاری هستند، چه این سیستم‌ها دسترسی مستقیم به دیسک‌های سخت ذخیره‌سازی یک سرور باشند و چه ذخیره‌سازی شبکه‌ای اشتراکی باشند [۲]. بسیاری از شرکت‌ها برای به اشتراک گذاشتن فایل‌های مدیریت شده به وسیله مدیر یک فایل سرور متصل در یک شبکه دارند که بسیار نسبت به سرقت داده‌های مهم مأموریتی آسیب‌پذیر است. امروزه، راه‌حل‌های شبکه‌ای مبتنی بر ذخیره‌سازی از جمله ذخیره‌سازی متصل به شبکه مانند شبکه‌های منطقه‌ای ذخیره‌سازی بسیار محبوب شده‌اند، که داده‌ها را در یک مسیر متمرکز با یک سیستم امنیتی مستقل حفظ می‌کنند [۴][۳]. شبکه‌های منطقه‌ای ذخیره‌سازی یک نوع معماری ذخیره‌سازی شبکه‌ای هستند که اتصال دسترسی به داده‌ها را بین کامپیوتر میزبان و دستگاه‌های ذخیره‌سازی فراهم می‌کنند. به منظور حفظ یکپارچگی داده‌های ذخیره شده در شبکه منطقه‌ای ذخیره‌سازی، به روش امنیتی نظارت شده سطح بالایی نیاز است تا این داده‌ها از حملات داخلی و خارجی حفظ شوند. اگر چه امنیت شبکه منطقه‌ای ذخیره‌سازی مستلزم بعضی روش‌های تخصصی برای حفاظت اطلاعات ذخیره شده و حفاظت آنها در هنگام انتقال است، زیرا شبکه منطقه‌ای ذخیره‌سازی انتقال داده در سطح بلوک را با برقراری ارتباط بین میزبان و دستگاه‌های ذخیره‌سازی ممکن می‌سازد [۵][۶]. امنیت شبکه منطقه‌ای ذخیره‌سازی بیشتر شامل یک گارد در برابر مهاجم داخلی یا خارجی است که قصد از بین بردن داده‌ها یا سرقت داده‌ها را دارد و همچنین نظارت[2] مداوم بر داده‌های در حال جریان، عملیات‌های پشتیبانی و دسترسی را نیز شامل می‌شود. روش‌های امنیتی خاصی

توسعه داده شده‌اند و در برابر حملات احتمالی در محیط شبکه منطقه‌ای ذخیره‌سازی امتحان شده‌اند و این فرآیند برای بهبود امنیت داده‌های ذخیره شده ادامه دارد [۷].

۲- معماری شبکه منطقه‌ای ذخیره‌سازی [۲۸]

مطابق با انجمن صنعتی شبکه ذخیره‌سازی[۳]، شبکه منطقه‌ای ذخیره‌سازی به عنوان شبکه‌ای که هدف اصلی آن انتقال داده بین سیستم‌های کامپیوتری و عنصر ذخیره‌سازی و میان عناصر ذخیره-سازی است، تعریف شده است [۱۰]. مزیت اصلی شبکه منطقه‌ای ذخیره‌سازی این است که از ذخیره‌سازی در یک محیط خارجی استفاده می‌کند. معماری شبکه منطقه‌ای ذخیره‌سازی نوعی در شکل ۱ نشان داده شده است [۱۳]، که شامل ۴ لایه از اجزاء است و عبارتند از: زیرساخت، سرورها، سیستم‌های ذخیره‌سازی و نرم‌افزار مدیریت. این چهار لایه به هم متصل شده‌اند و وظایف خاصی دارند.



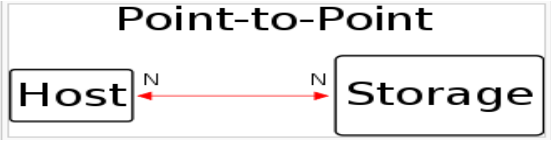
شکل ۱: معماری یک شبکه ذخیره‌سازی نوعی [۱۳].

اولین لایه، زیرساخت شبکه منطقه‌ای ذخیره‌سازی عبارت است از واسطها (مانند کانال فیبری، گذرگاه PCI، دروازه‌ها و غیره)، اتصالات (مانند پل‌ها، سویچ‌ها، مسیریاب‌ها و غیره)، وفق‌دهنده‌های گذرگاه میزبان[۴] و فابریک‌ها. یک وفق‌دهنده گذرگاه میزبان یک تخته مدار یا یک مدار مجتمع وفق دهنده است که پردازش ورودی و خروجی و اتصال فیزیکی بین یک سرور و یک دستگاه ذخیره‌سازی را میسر می-سازد و بین گذرگاه کامپیوترهای میزبان و کانال فیبری حلقوی قرار می‌گیرد و انتقال اطلاعات را بین دو کانال مدیریت می‌کند. به منظور به حداقل رساندن فشار کاری روی عملکرد پردازنده میزبان یا بهبود زمان عملکرد سرورها، بسیاری از عملیات‌های سطح پایین خط اتصال توسط وفق‌دهنده گذرگاه میزبان انجام می‌شود. در مورد فابریک در ادامه توضیح داده خواهد شد. لایه دوم سرورها هستند که با استفاده از واسطها به دستگاه‌های ذخیره‌سازی ناهمگون متصل شده‌اند. لایه سوم دستگاه‌های ذخیره‌سازی هستند که انواع دستگاه‌های ذخیره‌سازی انبوه خواهد داشت.

آرایه افزونه دیسک ارزان[۵] و تنها یک دسته از دیسک‌ها (بدون آرایه افزونه دیسک ارزان)[۶] دو دیسکی هستند که به طور معمول در محیط شبکه منطقه‌ای ذخیره‌سازی استفاده می‌شوند. تنها یک دسته از دیسک‌ها یک آرایه از دیسک‌های سخت است که یک یا چند دیسک سخت به صورت یک حجم واحد با هم ترکیب شده‌اند و یک دیسک تک بزرگتر را تشکیل می‌دهند. به عبارتی بدون آرایه افزونه دیسک ارزان هستند. چهارمین و آخرین لایه از شبکه منطقه‌ای ذخیره‌سازی نرم‌افزار مورد استفاد مدیر در محیط شبکه منطقه‌ای ذخیره‌سازی است، که به عنوان یک میز فرمان مدیریت ذخیره‌سازی در شکل‌دهی، تخصیص و صدور حقوق امنیتی و نرم‌افزار کنترل دسترسی عمل می-کند.

۲-۱- لایه‌های کانال فیبری

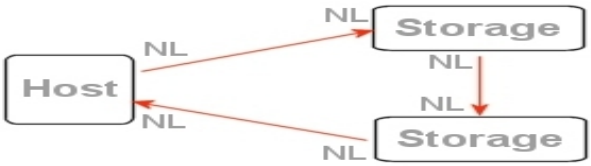
شبکه منطقه‌ای ذخیره‌سازی از کانال فیبری برای ارتباطات خود استفاده می‌کند و کانال فیبری مانند پاکت‌های داخل شبکه‌های IP از فریم بین سیستم ذخیره‌سازی به سرور یا کلاینت‌هایش و بالعکس، استفاده می‌کند. فریم‌های کانال فیبری مانند پاکت‌های IP در محیط یک شبکه آسیب‌پذیر هستند. اکثر پیاده‌سازی‌های شبکه منطقه‌ای ذخیره‌سازی از کانال فیبری حلقوی اختیاری[۷] به عنوان توپولوژیشان استفاده می‌کنند. دو توپولوژی دیگر توپولوژی نقطه به نقطه و توپولوژی فابریک سویچ شده[۸] هستند. کانال فیبری نقطه به نقطه یک توپولوژی کانال فیبری است که در آن دقیقا" دو درگاه (دستگاه) به طور مستقیم به یکدیگر متصل می‌شوند، این توپولوژی ساده‌ترین توپولوژی است و در آن به هیچ آدرس دهی شبکه نیاز نیست زیرا هر پیام فقط یک گیرنده ممکن دارد. پهنای باند در توپولوژی نقطه به نقطه اختصاص یافته است و این توپولوژی از N درگاه استفاده می‌کند.



شکل ۲: نمودار توپولوژی یک اتصال کانال فیبری نقطه به نقطه نوعی

کانال فیبری حلقوی اختیاری یک توپولوژی کانال فیبری است که در آن دستگاه‌ها در یک روش حلقه یکطرفه در یک توپولوژی حلقه[۹] به هم متصل شده‌اند. توپولوژی کانال فیبری حلقوی اختیاری دارای یک معماری سری است، که در آن پهنای باند روی حلقه در میان تمام درگاه‌ها به اشتراک گذاشته می‌شود، و فقط دو درگاه ممکن است در

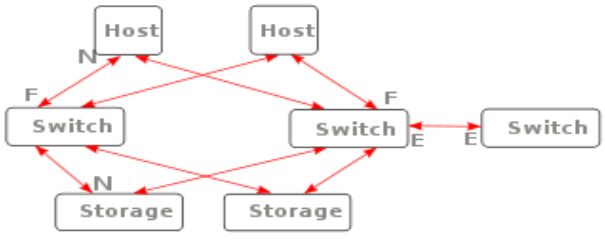
یک زمان روی حلقه ارتباط برقرار کنند. درگاه‌های کانال فیبری با توپولوژی اتصال حلقوی اختیاری، درگاه حلقوی گره[۱۰] و درگاه حلقوی فابریک[۱۱] هستند که در مجموع به عنوان درگاه‌های L نامیده می‌شوند. یک حلقه اختیاری بدون درگاه فابریک (فقط با درگاه‌های حلقوی گره)، یک حلقه خصوصی است و یک حلقه اختیاری متصل شده به یک فابریک از طریق درگاه حلقوی فابریک، یک حلقه عمومی است. درگاه حلقوی گره باید ورود به سیستم فابریک[۱۲] و امکانات ثبت نام را برای شروع ارتباط با گره دیگر از طریق فابریک فراهم کند (آغازگر ارتباط باشد).



شکل ۳: کانال فیبری حلقوی اختیاری خصوصی نوعی

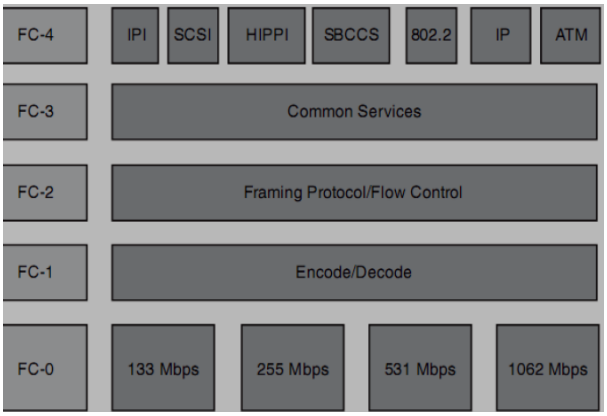
شکل ۴: کانال فیبری حلقوی اختیاری عمومی نوعی با هاب[۱۳]

فابریک سویچ شده یا فابریک، یک توپولوژی شبکه است که در آن گره‌های شبکه به وسیله یک یا چند سویچ شبکه با یکدیگر متصل می‌شوند. فابریک‌های سویچ شده می‌توانند مجموع توان عملیاتی بهتری نسبت به شبکه‌های پخش فراگیر[۱۴] ارائه دهند زیرا ترافیک در سرتاسر چندین لینک فیزیکی منتشر می‌شود. در توپولوژی فابریک سویچ شده کانال فیبری، دستگاه‌ها از طریق یک یا چند سویچ کانال فیبری به یکدیگر متصل شده‌اند. در حالی که این توپولوژی دارای بهترین مقیاس پذیری از سه توپولوژی کانال فیبری است (دو توپولوژی دیگر حلقوی اختیاری و نقطه به نقطه هستند)، فقط یکی از سویچ‌های آن مورد نیاز است که دستگاه‌های سخت‌افزاری پرهزینه‌ای هستند. چندین سویچ در یک فابریک، با دستگاه‌های موجود روی لبه‌های (برگ‌ها) مش معمولا" به شکل یک شبکه مش در می‌آیند.



شکل ۵: توپولوژی یک شبکه کانال فیبری فابریک سویچ شده نوعی

کانال‌های فیبری دارای پنج لایه هستند "شکل ۲"، لایه اول فیزیکی، لایه دوم انتقال، لایه سوم فریم کردن (سیگنال کردن)، لایه چهارم خدمات عادی و لایه پنجم لایه فوقانی پروتکل نگاشت در کانال فیبری بر اساس راه‌حل‌های شبکه منطقه‌ای ذخیره‌سازی است. فریم-های کانال فیبری مانند هر شبکه IP دیگر از لایه فیزیکی به سمت بالا عمل می‌کنند.



شکل ۶: لایه‌های کانال فیبری

۳- تهدیدها و آسیب‌پذیری‌های شبکه‌های منطقه‌ای ذخیره‌سازی

۳-۱- مفهوم خطر سه‌گانه

خطر سه‌گانه بر حسب خطر تهدیدات، دارایی‌ها و آسیب‌پذیری‌ها تعیین می‌شود. خطر هنگامی به وجود می‌آید که یک عامل تهدید (مهاجم) به دنبال دسترسی به دارایی‌ها به وسیله سوء استفاده از آسیب‌های موجود است.

برای مدیریت خطرات سازمان‌ها عمدتا" بر آسیب‌پذیری تمرکز می‌کنند، زیرا آنها نمی‌توانند عوامل تهدید که ممکن است در اشکال و منابع مختلف پدیدار شوند و به دارایی‌ها دست یابند را از بین ببرند. سازمان‌ها می‌توانند به وسیله کاهش آسیب‌پذیری به مقابله بپردازند تا تأثیر یک حمله به وسیله یک عامل تهدید کاهش یابد.

ارزیابی خطر اولین گام در تعیین میزان تهدیدات و خطرات بالقوه در یک زیرساخت فناوری اطلاعات است. این فرآیند خطر را ارزیابی می‌کند و به شناسایی کنترل‌های مناسب برای کاهش یا حذف خطرات کمک می‌کند. به منظور تعیین احتمال انجام گرفتن یک رخداد مضر، تهدیدات وارد به یک سیستم فناوری اطلاعات باید در ارتباط با آسیب-پذیری‌های بالقوه و کنترل‌های امنیتی موجود تجزیه و تحلیل شوند.

شدت یک رخداد نامطلوب به وسیله تأثیری که ممکن است بر روی فعالیت‌های تجاری مهم داشته باشد تخمین زده می‌شود. بر اساس این تجزیه و تحلیل، یک مقدار نسبی از وضع بحرانی و حساسیت را می‌توان به دارایی‌ها و منابع فناوری اطلاعات اختصاص

داد. برای مثال ممکن است به یک جزء خاص سیستم فناوری اطلاعات یک ارزش بحرانی بالا اختصاص داده شود، که اگر یک حمله روی این جزء خاص رخ دهد می‌تواند به طور کامل باعث خاتمه دادن خدمات یک مأموریت مهم شود.

در ادامه به بررسی سه عنصر اصلی خطر سه‌گانه می‌پردازیم. دارایی‌ها، تهدیدات و آسیب‌پذیری از لحاظ شناسایی خطر و آنالیز کنترلی مطرح شده‌اند.

۳-۱-۱- دارایی‌ها

اطلاعات یکی از مهمترین دارایی‌ها برای هر سازمان هستند. دارایی‌های دیگر شامل سخت‌افزار، نرم‌افزار، و زیرساخت شبکه مورد نیاز برای دسترسی به این اطلاعات هستند. برای حفاظت از این دارایی‌ها، سازمان‌ها باید مجموعه‌ای از پارامترها را برای اطمینان از قابل دسترس بودن منابع به کاربران مجاز و شبکه‌ای قابل اعتماد، ایجاد کنند. این پارامترها به منابع ذخیره‌سازی، زیرساخت شبکه، و سیاست‌های سازمانی اعمال می‌شوند.

عوامل متعددی باید در هنگام برنامه‌ریزی برای امنیت دارایی در نظر گرفته شود. روش‌های امنیتی دارای دو هدف هستند. هدف اول این است که اطمینان حاصل شود شبکه به راحتی در دسترس کاربران مجاز است. همچنین این شبکه باید در شرایط ناهمگون محیط زیست و حجم‌های استفاده، قابل اعتماد و با ثبات باشد. هدف دوم این است که در برابر حملات بالقوه برای دستیابی و سازش با سیستم بسیار مقاوم ساخته شود. این روش‌ها باید حفاظت کافی را در برابر دسترسی غیر مجاز به منابع، ویروس‌ها، کرم‌ها، تروجان‌ها و دیگر برنامه‌های نرم-افزاری مخرب، فراهم بیاورند. اقدامات امنیتی همچنین باید داده‌های مهم (بحرانی) را رمزنگاری کنند و خدمات استفاده نشده را غیر فعال کنند تا تعدادی از فضاهای بالقوه امنیتی کاهش یابد. روش امنیتی باید این اطمینان را بدهد که با سیستم عامل و دیگر برنامه‌هایی که نصب هستند به طور منظم به روز می‌شود. ضمنا"، باید افزونگی کافی را به صورت تولید داده‌های همسان و تکراری فراهم آورد تا در صورت وجود یک خرابی غیر منتظره، از از دست دادن داده‌های فاجعه‌آمیز جلوگیری شود. برای اینکه یک سیستم امنیتی بصورت یکنواخت و به راحتی کار کند این اهمیت دارد که اطمینان حاصل شود همه کاربران از سیاست-های حاکم بر استفاده از شبکه آگاه هستند [۱].

۳-۱-۲- تهدیدات

تهدیدات حملات بالقوه‌ای هستند که می‌توانند به زیرساخت فناوری اطلاعات تحمیل شوند. این حملات به دو دسته معلوم[۱۵] یا مجهول[۱۶] تقسیم می‌شوند. حملات مجهول تلاش می‌کنند که دستیابی غیر مجاز به سیستم داشته باشند. آنها تهدیدات مطرح برای محرمانه بودن اطلاعات هستند. حملات معلوم شامل دستکاری داده‌ها، جلوگیری از سرویس[۱۷]، و حملات انکاری[۱۸] هستند. آنها تهدیدات مطرح مطرح برای یکپارچگی و در دسترس بودن داده‌ها هستند.

در یک حمله دستکاری، کاربر غیر مجاز تلاش می‌کند که اطلاعات را برای اهداف مخربش تغییر دهد، که این اهداف می‌توانند بر روی داده‌های در حال انتقال یا بقیه داده‌ها انجام شوند. این حملات یک تهدید مطرح برای یکپارچگی داده‌ها هستند.

حملات جلوگیری از سرویس استفاده از منابع را برای کاربران قانونی دچار اختلال می‌کنند و مانع استفاده کاربران مجاز از شبکه یا وب سایت می‌شوند. به طور کلی این حملات شامل دستیابی به اطلاعات یا دستکاری اطلاعات روی سیستم‌های کامپیوتری نمی‌شوند، در عوض، تهدیدی برای در دسترس بودن داده‌ها هستند.

حمله انکاری، یک حمله در مقابل پاسخ دهی اطلاعات است. این حملات تلاش می‌کنند که به وسیله جعل هویت کسی یا انکار این که یک رویداد یا تبادل اطلاعات صورت گرفته است، اطلاعات نادرستی ارائه دهند [۱].

۳-۱-۳- آسیب‌پذیری

مسیرهایی که دستیابی به اطلاعات را فراهم می‌آورند نسبت به حملات بالقوه آسیب‌پذیرتر هستند. هر یک از این مسیرها ممکن است شامل نقاط دستیابی مختلفی باشد، که هر یک از آنها سطوح مختلف دسترسی به منابع ذخیره‌سازی را فراهم می‌کند. این خیلی مهم است که کنترل‌های امنیتی کافی در تمام نقاط دسترسی روی یک مسیر دسترسی پیاده‌سازی شود. پیاده‌سازی کنترل‌های امنیتی در هر نقطه دسترسی از هر مسیر دسترسی، دفاع در عمق[۱۹] نامیده شده است.

دفاع در عمق حفاظت تمام نقاط دسترسی داخل یک محیط را بیان می‌کند. این کار آسیب‌پذیری را برای یک مهاجم کاهش می‌دهد، کسی که می‌تواند به وسیله دور زدن کنترل‌های امنیتی پیاده‌سازی شده ناکافی در تک نقطه دسترسی آسیب‌پذیر، به منابع ذخیره‌سازی دستیابی پیدا کند. چنین حمله‌ای می‌تواند امنیت دارایی‌های اطلاعاتی را به خطر بیاندازد. برای مثال، یک مشکل در تأیید درست هویت یک کاربر ممکن است محرمانه بودن اطلاعات را به خطر بیاندازد. به طور مشابه، یک حمله جلوگیری از سرویس در مقابل یک دستگاه ذخیره-سازی می‌تواند در دسترس بودن اطلاعات را به خطر بیاندازد.

رویه حمله، مسیر حمله، و عامل کار سه عامل در نظر گرفته شده هستند در هنگام ارزیابی این که یک محیط تا چه حد نسبت به تهدیدات امنیتی آسیب‌پذیر است [۱].

- **رویه حمله**[۲۰]

رویه حمله به نقاط ورودی مختلف اشاره می‌کند که یک مهاجم می‌تواند به منظور شروع حمله استفاده کند. هر جزء از یک شبکه ذخیره‌سازی یک منبع آسیب‌پذیری بالقوه است. همه واسطهای خارجی پشتیبانی شده توسط آن جزء، از جمله واسطهای سخت-

افزاری، پروتکل‌های پشتیبانی شده، و واسط‌های اجرایی و مدیریت، می‌توانند مورد استفاده مهاجم برای اجرای حملات مختلف قرار گیرند. این واسط‌ها یک رویه از حمله برای مهاجم هستند. حتی خدمات شبکه استفاده نشده در صورتی که فعال باشند می‌توانند بخشی از رویه حمله باشند.

- **مسیر حمله**[۲۱]

یک مسیر حمله، یک گام یا یک دنباله از اقدامات لازم برای تکمیل یک حمله است. به عنوان مثال، یک مهاجم ممکن است از یک اشکال (باگ)[۲۲] در واسط مدیریتی برای اجرای حمله اسنوپ (جاسوسی)[۲۳] سوء استفاده کند که به موجب آن مهاجم می‌تواند پیکربندی دستگاه ذخیره‌سازی را تغییر دهد تا خود را مجاز کند که بیشتر از یک میزبان به ترافیک دستیابی داشته باشد. این ترافیک دوباره هدایت شده می‌تواند برای اسنوپ داده در حال انتقال مورد استفاده قرار گیرد.

- **عامل کار**[۲۴]

عامل کار به مقدار زمان و تلاش مورد نیاز برای سوء استفاده از یک مسیر حمله اشاره می‌کند. برای مثال، اگر مهاجمان برای بازیابی اطلاعات حساس تلاش کنند، آنها زمان و تلاشی را که برای اجرای یک حمله روی پایگاه داده مورد نظر نیاز خواهد بود در نظر می‌گیرند. این ممکن است شامل تعیین حساب‌های ویژه، تعیین شمای پایگاه داده، و نوشتن پرس و جوهای[۲۵] پایگاه داده SQL باشد. در عوض، بر اساس عامل کار، آنها یک راه که دارای تلاش و فشردگی کمتری است را برای سوء استفاده از آرایه ذخیره‌سازی به وسیله اتصال به آن به طور مستقیم و خواندن از بلوک‌های دیسک خام در نظر می‌گیرند.

با داشتن ارزیابی آسیب‌پذیری محیط شبکه نسبت به تهدیدات امنیتی، سازمان‌ها می‌توانند در جهت کاهش آسیب‌پذیری به وسیله به حداقل رساندن رویه‌های حمله و به حداکثر رساندن عامل کار، اقدامات کنترلی ویژه‌ای برنامه‌ریزی کنند و گسترش دهند. این کنترل‌ها فنی یا غیر فنی هستند. کنترل‌های فنی معمولا" از طریق سیستم‌های کامپیوتری اجرا می‌شوند، در حالی که کنترل‌های غیر فنی از طریق کنترل‌های فیزیکی و اجرایی اجرا می‌شوند. کنترل‌های اجرایی[۲۶] شامل امنیت و سیاست‌های کارکنان یا پروسه‌های استاندارد در جهت اجرای امن عملیات‌های مختلف هستند. کنترل‌های فیزیکی شامل راه اندازی موانع فیزیکی از جمله گاردهای امنیتی، دیوارها، یا قفل‌ها هستند.

بر اساس نقش‌هایی که مهاجمان ایفا می‌کنند، کنترل‌ها می‌توانند به عنوان پیشگیری کننده، تشخیص دهنده، اصلاح کننده، بهبود دهنده، یا تقویت کننده طبقه‌بندی شوند.

کنترل پیشگیرانه برای جلوگیری از یک حمله تلاش می‌کند، کنترل تشخیص دهنده تشخیص می‌دهد که آیا یک حمله در حال انجام است، و پس از اینکه یک حمله کشف شد، کنترل‌های اصلاح کننده اجرا می‌شوند. کنترل‌های پیشگیرانه سوء استفاده از آسیب-پذیری‌های موجود را دفع می‌کند و از یک حمله جلوگیری می‌کند یا تأثیر آن را کاهش می‌دهد. کنترل‌های اصلاح کننده اثر یک حمله را کاهش می‌دهند، در حالی که کنترل‌های تشخیص دهنده حملات را کشف می‌کنند و کنترل‌های پیشگیرانه یا اصلاح کننده را به راه میاندازند. برای مثال، یک سیستم جلوگیری نفوذ یا تشخیص نفوذ[۲۷] یک کنترل تشخیص دهنده است که تعیین می‌کند آیا یک حمله در حال انجام است، و پس از آن تلاش می‌کند که آن را به وسیله خاتمه دادن اتصالش به شبکه یا با استناد به یک قانون دیواره آتش[۲۸] برای مسدود کردن ترافیک، متوقف کند.

۳-۲- دسته‌بندی تهدیدات امنیتی

تهدیدات امنیتی شبکه منطقه‌ای ذخیره‌سازی به پنج دسته تقسیم می‌شوند: مجهول، فعال، مجاور[۲۹]، داخلی[۳۰] و پخش[۳۱]. حملات مجهول شامل تجزیه و تحلیل ترافیک، نظارت کردن ارتباطات محافظت نشده، رمزگشایی ترافیکی که ضعیف رمزنگاری شده، و گرفتن اطلاعات احراز هویت مانند رمزهای عبور است. حملات فعال شامل تلاش برای دور زدن یا شکستن ویژگی‌های حفاظتی، معرفی کد مخرب، سرقت یا تغییر اطلاعات است. حملات مجاور که در آن یک فرد غیر مجاز در مجاورت فیزیکی نزدیک به شبکه‌ها، سیستم‌ها، یا سایر وسایل برای تغییر اطلاعات، جمع‌آوری اطلاعات، یا محروم کردن افراد مجاز از دسترسی به اطلاعات است. حملات داخلی می‌توانند مخرب یا غیر مخرب باشند. حملات داخلی مخرب قصد دارند که اطلاعات را استراق سمع[۳۲] کنند، به سرقت ببرند، یا به آنها خسارت و صدمه وارد کنند. حملات غیر مخرب به طور معمول از نتیجه بی‌دقتی، عدم آگاهی و دانش، یا عمدا" دور زدن امنیت ناشی می‌شوند. حملات پخش بر تغییر مخرب سخت‌افزار یا نرم‌افزار در طول توزیع یا داخل کارخانه تمرکز می‌کنند. در اغلب موارد شبکه منطقه‌ای ذخیره‌سازی با سه نوع تهدید مواجه می‌شود: تهدیدات خارجی مخرب، تهدیدات داخلی مخرب و تهدیدات داخلی غیر مخرب [۲۳].

اکثر خطرات در محیط شبکه منطقه‌ای ذخیره‌سازی همواره با تهدیدات شروع می‌شوند. این تهدیدات بیشتر به سه سطح دسته‌بندی می‌شوند [۱۶]. سطح اول تهدیدات غیر عمدی، و به دلیل حوادث یا اشتباهات است. سطح دوم تهدیدات یک حمله ساده مخرب است که از تجهیزات موجود استفاده می‌کند و امکان دارد برخی اطلاعات به آسانی بدست بیایند. سطح سوم از تهدیدات، حمله در مقیاس بزرگ است که نیاز به یک سطح غیر معمول از پیچیدگی و تجهیزات برای اجرای حمله دارد. یک حمله سطح سوم معمولا" از یک منبع خارجی است و چه از لحاظ فیزیکی یا از لحاظ مجازی نیاز به دسترسی دارد. در شبکه منطقه‌ای ذخیره‌سازی این حملات احتمالی مختلف هستند که می-توانند به منظور کنترل شبکه منطقه‌ای ذخیره‌سازی برای سرقت اطلاعات رخ دهند [۱۶]: حمله به نام فابریک[۳۳]، حمله با شناسایی

حوزه[۳۴]، حمله تغییر نام اطلاعات سرور[۳۵]، حمله به شماره کنترل ترتیبی دوره[۳۶]، حمله به شناسه‌های ترتیبی دوره[۳۷]، حمله به نام‌های گسترده جهانی مورد استفاده در فابریک[۳۸]، حمله به اطلاعات فریم لایه دوم[۳۹]، حمله به آدرس‌های ۲۴ بیتی[۴۰]، حمله به اطلاعات مسیریابی[۴۱]، حمله به اطلاعات مدیریت[۴۲]، با حفاظت تک تک اجزاء بالا با استفاده از روش‌های مناسب در واقع می‌توان اطلاعات ذخیره شده ارزشمند را در دستگاه‌های ذخیره‌سازی خارجی شبکه منطقه‌ای ذخیره‌سازی محافظت کرد. در یک شبکه منطقه‌ای ذخیره‌سازی به سخت‌افزاری که ایستگاه‌های کاری و سرورها را به دستگاه‌های ذخیره‌سازی متصل می‌کند، فابریک گفته می‌شود. فابریک شبکه منطقه‌ای ذخیره‌سازی، اتصال هر سرور به هر دستگاه ذخیره‌سازی را از طریق استفاده از فناوری سویچ کانال فیبری ممکن می‌سازد. سرویس نام فابریک اجازه می‌دهد هر دستگاه آدرس‌های تمام دستگاه‌های دیگر را پرس و جو کند. یک نام گسترده جهانی[۴۳] یا شناسه گسترده جهانی[۴۴]، یک شناسه واحد مورد استفاده در فناوری‌های ذخیره‌سازی شامل کانال فیبری و غیره است. یک نام گسترده جهانی ممکن است در انواع مختلفی از نقش‌ها مانند یک شماره سریال یا برای قابلیت نشانی پذیری به کار گرفته شود. برای مثال، در شبکه‌های کانال فیبری یک نام گسترده جهانی ممکن است به عنوان یک نام گره گسترده جهانی[۴۵] به منظور شناسایی یک سویچ، یا یک نام درگاه گسترده جهانی[۴۶] به منظور شناسایی یک درگاه تک روی یک سویچ استفاده شود. دو نام گسترده جهانی که به شیء مشابه اشاره نمی‌کنند باید همیشه متفاوت باشند حتی اگر این دو در نقش‌های متفاوتی مورد استفاده هستند، یعنی یک نقش مانند نام گره گسترده جهانی یا نام درگاه گسترده جهانی، یک فضای نام گسترده جهانی جدا را مشخص نمی‌کند. هر نام گسترده جهانی یک شماره ۸ یا ۱۶ بایتی است.

۴- امنیت شبکه منطقه‌ای ذخیره‌سازی

۴-۱- چارچوبی امنیتی برای شبکه‌های منطقه‌ای ذخیره‌سازی

یک چارچوب امنیتی برای شبکه‌های منطقه‌ای ذخیره‌سازی وجود دارد که در ادامه توضیح داده خواهد شد. این چارچوب برای تسکین تهدیدات امنیتی که ممکن است در آینده به وجود آیند، بر مبنای مقابله با حملات مخرب روی زیرساخت ذخیره‌سازی کار می‌کند.

چارچوب اساسی امنیت پیرامون چهار خدمت اصلی امنیت ساخته شده است: پاسخ دهی، محرمانه بودن، یکپارچگی، و در دسترس بودن [۱]. این چارچوب شامل تمام اقدامات امنیتی لازم برای کاهش تهدیدات است که شامل این چهار ویژگی اصلی امنیت است:

۴-۱-۱- سرویس پاسخ دهی[۴۷]

این سرویس به حسابداری برای تمام وقایع و عملیات‌ها که در محل زیرساخت مرکز داده قرار می‌گیرند اشاره می‌کند. سرویس پاسخ دهی یک گزارش (لاگ)[۴۸] از عملکرد وقایع نگهداری می‌کند که در آینده می‌توان این وقایع را به منظور تأمین امنیت، بازرسی[۴۹] یا ردیابی کرد.

۴-۱-۲- سرویس محرمانه بودن[۵۰]

سرویس محرمانه بودن اطلاعات مورد نیاز را فراهم می‌کند و این را تضمین می‌کند که تنها کاربران مجاز به داده‌ها دسترسی دارند. این سرویس به کاربرانی که نیاز به دسترسی به اطلاعات دارند اعتبار می‌بخشد و معمولا" محرمانه بودن هر دو داده‌های در حال انتقال (داده‌های انتقال داده شده بر روی کابل‌ها)، یا بقیه داده‌ها (داده‌های بر روی رسانه پشتیبان یا در آرشیوها) را پوشش می‌دهد. داده‌های در حال انتقال و بقیه داده‌ها می‌توانند رمزنگاری شوند تا محرمانه بودن آنها حفظ شود. علاوه بر محدود کردن کاربران غیر مجاز از دسترسی به اطلاعات، سرویس محرمانه بودن همچنین اقدامات حفاظتی جریان ترافیک را به عنوان بخشی از پروتکل امنیتی اجرا می‌کند. این اقدامات حفاظتی به طور کلی شامل پنهان کردن آدرس‌های منبع و مقصد، فرکانس داده‌های در حال ارسال، و مقدار داده‌های فرستاده شده هستند.

۴-۱-۳- سرویس یکپارچگی[۵۱]

این سرویس تضمین می‌کند که اطلاعات دست نخورده است. هدف این سرویس شناسایی و محافظت از اطلاعات در برابر تغییر غیر مجاز یا حذف آنها است. مشابه سرویس محرمانه بودن، سرویس یکپارچگی با همکاری سرویس پاسخ دهی کار می‌کند تا کاربران را شناسایی و احراز هویت کند. اقدامات سرویس یکپارچگی برای هر دو داده‌های در حال انتقال و بقیه داده‌ها تعریف می‌شوند.

۴-۱-۴- سرویس در دسترس بودن[۵۲]

این سرویس تضمین می‌کند که کاربران مجاز به موقع به داده‌ها دسترسی دارند و همچنین این کاربران به داده‌های قابل اعتماد دسترسی دارند. این سرویس کاربران را قادر می‌سازد که به سیستم‌های کامپیوتری مورد نیازشان، داده‌ها، و برنامه‌های کاربردی موجود در این سیستم‌ها دسترسی داشته باشند. سرویس در دسترس بودن همچنین در سیستم‌های ارتباطی مورد استفاده برای انتقال اطلاعات بین کامپیوترهایی که ممکن است در مکان‌های مختلف اقامت داشته باشند پیاده‌سازی شده است و این تضمین را می‌دهد که اگر یک خرابی در یک مکان خاص رخ دهد، باز هم اطلاعات در دسترس هستند. این سرویس‌ها باید برای هر دو داده‌های فیزیکی و داده‌های الکترونیکی پیاده‌سازی شوند.

۴-۲- روش‌های تأمین امنیت در شبکه منطقه‌ای ذخیره‌سازی

امنیت شبکه منطقه‌ای ذخیره‌سازی مجموعه‌ای از پارامترها و تنظیمات است که باعث می‌شود منابع ذخیره‌سازی فقط در دسترس کاربران مجاز و شبکه‌های مورد اعتماد قرار گیرند و دیگران به آنها دستیابی نداشته باشند. این پارامترها را می‌توان به سخت‌افزار، برنامه-نویسی، پروتکل‌های ارتباطی و سیاست‌های سازمانی اعمال کرد. در واقع، اکثر تهدیدات امنیتی ذخیره‌سازی شبکه بر اساس تهدیدات داخلی ناشی از کارمندان داخلی هستند [۱۴]. در نتیجه بهترین شیوه-های امنیتی در تلاش برای حفظ پنج هدف اساسی یعنی در دسترس بودن، یکپارچگی، احراز هویت، محرمانه بودن و عدم انکار داده‌ها هستند [۱۵].

امنیت شبکه‌های منطقه‌ای ذخیره‌سازی شامل بسیاری از روش‌ها و راهکارها است، به دلیل اینکه این شبکه‌ها به بالاترین سطح امنیت داده‌های ذخیره شده برسند این روش‌ها با یکدیگر در ارتباط هستند. یک مدیریت ذخیره‌سازی مؤثر باید پنج حوزه اساسی امنیت را در هر سطح شبکه منطقه‌ای ذخیره‌سازی پیاده‌سازی کند، که عبارتند از: کنترل حجم دستیابی به آرایه ذخیره‌سازی، کنترل حجم دسترسی روی یک میزبان، کنترل دستیابی به پیکربندی دستگاه، کنترل دستیابی به نرم‌افزار مدیریت ذخیره‌سازی و تشخیص بلادرنگ تجاوز دسترسی، بازرسی و ثبت وقایع[۵۳]. پنج دسته گسترده بالا بیشتر در تکنیک‌های امنیتی زیر به طور جداگانه انجام می‌پذیرند:

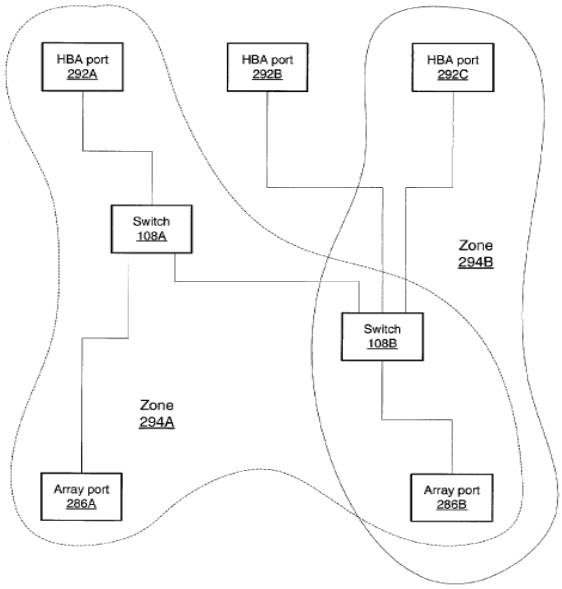
۱- کنترل دسترسی (وفق‌دهنده کردن و مخفی‌سازی شناسه آرایه[۵۴])، ۲- سیستم تشخیص نفوذ[۵۵]، ۳- رمزنگاری (CFS, SFS & EFS) ۴- احراز هویت و صدور مجوز[۵۶]، ۵- امنیت کانال فیبری ۶- تأمین امنیت با استفاده از نرم‌افزار مدیریت شبکه منطقه‌ای ذخیره-سازی.

Fibre Channel
Layer 4. Protocol mapping
LUN masking
Layer 3. Common services
Layer 2. Network
Fibre Channel fabric Fibre Channel zoning Registered State Change Notification
Layer 1. Data link
Fibre Channel 8B/10B encoding
Layer 0. Physical

شکل ۷: برخی از تکنیک‌های امنیتی موجود در لایه‌های کانال فیبری شبکه منطقه‌ای ذخیره‌سازی

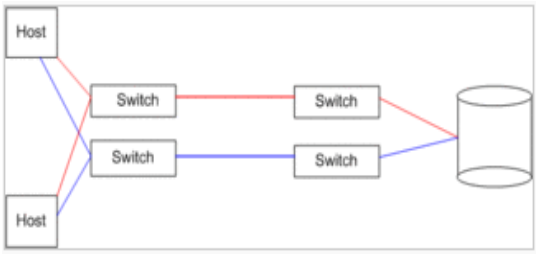
۴-۲-۱- کنترل دسترسی (ناحیه‌بندی کردن [۲۹] و مخفی‌سازی شناسه آرایه[۳۰])

در شبکه ذخیره‌سازی، ناحیه‌بندی کانال فیبری قسمت‌بندی یک فابریک کانال فیبری[۵۷] به زیر مجموعه‌های کوچکتر برای محدود کردن تداخل، اضافه کردن امنیت، و برای مدیریت آسان است. هنگامیکه یک شبکه منطقه‌ای ذخیره‌سازی چندین دستگاه و درگاه را برای یک دستگاه قابل دسترس می‌سازد، هر سیستم متصل به شبکه منطقه‌ای ذخیره‌سازی فقط باید اجازه دسترسی به یک زیرمجموعه کنترل شده این دستگاه‌ها یا درگاه‌ها داشته باشد.



شکل ۸: ناحیه‌بندی سرورها

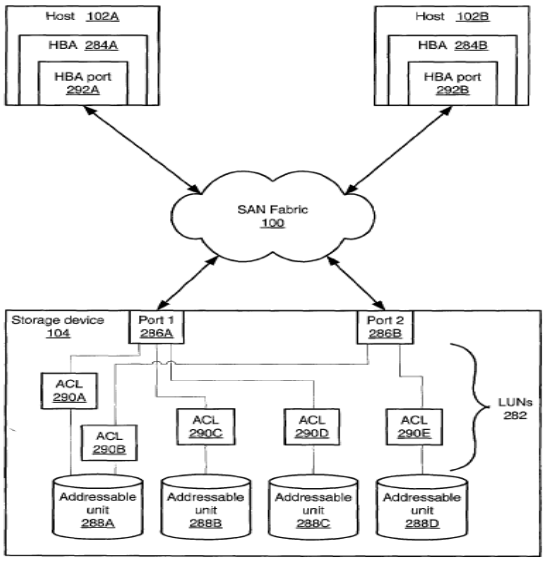
ناحیه‌بندی فقط به توپولوژی فابریک سویچ شده اعمال می‌شود و در توپولوژی‌های کانال فیبری ساده‌تر وجود ندارد. میدان دید در میان دستگاه‌ها (گره‌ها) در یک فابریک به طور معمول با ناحیه‌بندی کنترل می‌شود. اکثر طرح‌های شبکه کانال فیبری دو فابریک مجزا را برای افزونگی بکار می‌گیرند. دو فابریک، گره‌های لبه (دستگاه‌ها) را به اشتراک می‌گذارند، اما در غیر اینصورت به هم متصل نیستند. یکی از مزایای چنین راه اندازی، قابلیت فیل اُور[۵۸] آن است، به این معنی که در مورد یک قطعی لینک یا از کار افتادن یک فابریک، دیتاگرام‌ها[۵۹] (برنامه‌های دارای اطلاعات و داده‌ها) را می‌توان از طریق فابریک دوم فرستاد.



شکل ۹: یک شبکه منطقه‌ای ذخیره‌سازی ساخته شده با دو فابریک سویچ شده مجزا (قرمز و آبی)، برای افزایش قابلیت اطمینان

در مواقع قطعی، لینک اول به صورت اتوماتیک بر روی لینک دوم سویچ می‌شود. در ناحیه‌بندی مسیر از سرور به آرایه ذخیره‌سازی بسته می‌شود. دستگاه‌ها به حوزه‌های تکی یا حوزه‌های اشتراکی محدود می‌شوند. فابریک مبتنی بر ناحیه‌بندی یک روش قوی و مستحکم در برابر دسترسی غیر مجاز است. مسیر برای دیگران بسته می‌شود و فقط برای آنهایی باز می‌شود که حقوق دسترسی به آنها اختصاص دارد. به طور معمول دو نوع ناحیه‌بندی در شبکه منطقه‌ای ذخیره‌سازی فابریک تطبیق داده شده است، به نام حوزه سخت‌افزار که در سطح سویچ انجام می‌شود و حوزه نرم که در سطح نرم‌افزار سویچ انجام می‌شود. حوزه نرم آسیب‌پذیرتر است و به آسانی می‌توان با میل نفوذ به وسیله ترکیب آدرس منبع و مقصد، آدرس فریم را جعل کرد. سرویس نام فابریک اجازه می‌دهد که هر دستگاه آدرس‌های تمام دستگاه‌های دیگر را پرس و جو کند. ناحیه‌بندی نرم فقط سرویس نام فابریک را محدود می‌کند تا فقط یک زیر مجموعه مجاز از دستگاه‌ها را نشان دهد. بنابراین، وقتی که یک سرور محتوای فابریک را نگاه می‌کند، فقط دستگاه‌هایی را که مجاز است ببیند خواهد دید. با این حال، هر سرور هنوز می‌تواند تلاش کند تا با هر دستگاه روی شبکه از طریق آدرس تماس برقرار کند. در مقابل، ناحیه‌بندی سخت ارتباط واقعی در سراسر یک فابریک را محدود می‌کند، این کار به پیاده‌سازی سخت‌افزار کارآمد (فیلترینگ فریم) در سویچ‌های فابریک نیاز دارد، اما خیلی امن‌تر از ناحیه‌بندی نرم است. ناحیه‌بندی می‌تواند به درگاه سویچی که یک دستگاه به آن متصل شده است یا به نام گسترده جهانی روی میزبانی که در حال اتصال است اعمال شود. در نتیجه درگاه مبتنی بر ناحیه‌بندی، جریان ترافیک را بر اساس درگاه سویچ خاصی که یک دستگاه به آن متصل است محدود می‌کند، که اگر آن دستگاه از درگاه جدا شود، دسترسی خود را از دست خواهد داد. علاوه بر این، دستگاه متفاوتی که به همان درگاه متصل شود، دستیابی خود را از هر منبعی بدست خواهد آورد که میزبان قبلی به آن دسترسی داشته است. ناحیه‌بندی نام گسترده جهانی (ناحیه‌بندی نام نیز نامیده می‌شود) دسترسی را به وسیله یک نام گسترده جهانی که دستگاه (میزبان) دارد محدود می‌کند. هنگامیکه نام گسترده جهانی روی میزبان است، میزبان می‌تواند از درگاهی که به آن متصل است جدا شود و به درگاه دیگر متصل شود بدون اینکه دستیابی آن به منابع قبلی از دست برود، به عبارتی از درگاه جدید نیز به همان منابع دسترسی دارد. اتصال یک دستگاه جدید به درگاهی که قبلا" توسط یک دستگاه با حوزه نام گسترده جهانی استفاده شده است، هیچ دسترسی را به منابع دستگاه قبلی ممکن نخواهد ساخت.

مخفی‌سازی شناسه آرایه[۶۰] این اطمینان را می‌دهد که فقط کاربران مجاز به فایل‌هایی که به آنها اختصاص داده شده دستیابی دارند. در ذخیره‌سازی کامپیوتر، شماره واحد منطقی، یک شماره مورد استفاده برای شناسایی یک واحد منطقی است. یک شماره واحد منطقی ممکن است با هر دستگاهی که عملیات‌های خواندن و نوشتن را پشتیبانی می‌کند استفاده شود، مانند نوار چرخان، اما اغلب برای یک دیسک منطقی ایجاد شده روی یک شبکه منطقه‌ای ذخیره‌سازی کاربرد دارد و به آن اعمال می‌شود. که در این تحقیق این شماره واحد منطقی یک شناسه برای آرایه ذخیره‌سازی است. مخفی‌سازی شناسه آرایه یک فرآیند مجوز است که شماره واحد منطقی را برای بعضی میزبان‌ها در دسترس و برای دیگر میزبان‌ها غیر قابل دسترسی قرار می‌دهد.



شکل ۱۰: مخفی‌سازی شناسه آرایه و تعیین مسیر به وسیله لیست کنترل دسترسی[۶۱]

لیست کنترل دسترسی یک لیست از قوانین است که برای ورود و یا خروج ترافیک بر روی روتر قرار داده می‌شود. مخفی‌سازی شناسه آرایه عمدتا" در سطح وفق‌دهنده گذرگاه میزبان پیاده‌سازی شده است. مخفی‌سازی شناسه آرایه پیاده‌سازی شده در این سطح، به هر حمله که با وفق‌دهنده گذرگاه میزبان سازش می‌کند آسیب‌پذیر است و از آن برای جعل آدرس‌های منبع (نام‌های گسترده جهانی) و سازش برای دسترسی استفاده می‌شود. برخی از کنترل کننده‌های ذخیره‌سازی نیز مخفی‌سازی شناسه آرایه را پشتیبانی می‌کنند. هنگامیکه مخفی‌سازی شناسه آرایه در سطح کنترل کننده ذخیره‌سازی اجرا شود، خود

کنترل کننده سیاست‌های دستیابی به دستگاه را به علت امن‌تر بودن آن اجرا می‌کند. با این وجود، عمدتا" به خودی خود به عنوان یک معیار امنیتی اجرا نمی‌شود، بلکه به عنوان یک محافظ در برابر سرورهای بدرفتار است، که ممکن است دیسک‌های متعلق به سرورهای دیگر را فاسد کنند. به عنوان مثال، سرورهای ویندوز متصل شده به یک شبکه منطقه‌ای ذخیره‌سازی تحت برخی شرایط حجم‌های غیر ویندوزها (لینوکس، یونیکس) روی شبکه منطقه‌ای ذخیره‌سازی را به وسیله اقدام برای نوشتن برچسب‌های حجیم ویندوز روی آنها، فاسد خواهند کرد. به وسیله پنهان کردن شماره واحدهای منطقی دیگر از سرور ویندوز، این کار می‌تواند جلوگیری شود، از آنجا که سرور ویندوز حتی متوجه نمی‌شود شماره واحدهای منطقی دیگر وجود داشته باشد.

مخفی‌سازی شناسه آرایه مهم است زیرا ویندوز سرورها به نوشتن برچسب‌های حجیم به تمام شماره واحدهای منطقی در دسترس اقدام می‌کند. این امر می‌تواند شماره واحدهای منطقی غیر قابل استفاده را به وسیله سیستم‌های عامل دیگر ارائه دهد و در نتیجه می‌تواند باعث از دست رفتن داده‌ها شود. یک راه دستیابی به شناسه آرایه توسط وفق‌دهنده‌های گذرگاه میزبانش است که از نام گسترده جهانی موجود با هر وفق‌دهنده گذرگاه میزبان و هر حجم سرور استفاده می‌کند و برای امکان دسترسی به کاربران شامل عرضه می‌شود. این یک خطر بالقوه دیگر بدست گرفتن کنترل سرور برای دسترسی به حجم‌های سرور است. مدیر حجم معمولا" حقوق را اختصاص می‌دهد به کسانی که قصد استفاده از هر حجم را دارد و این کار به وسیله بخشی از سیستم عامل و مدیر حجم انجام می‌شود.

۴-۲-۲- سیستم تشخیص نفوذ [۳۱,۳۲]

امنیت شبکه‌های منطقه‌ای ذخیره‌سازی به نظارت[۶۲] متداوم فایل-ها، ویژگی‌های فایل، بازرسی، و ثبت وقایع برای دسترسی به اطلاعات فایل که در شناسایی مزاحمان بالقوه مفید خواهد بود، نیاز دارد. ذخیره‌سازی مبتنی بر تشخیص نفوذ [۱۷] باید به منظور دیدن اینکه آیا اطلاعات شبکه منطقه‌ای ذخیره‌سازی به وسیله کاربران مجاز خوانده می‌شوند یا خیر، پیاده‌سازی شود. تکنیک تشخیص نفوذ یکی از روش‌های مرسوم امنیت برای تشخیص و جلوگیری از تغییرات مخرب و دستیابی‌های غیر مجاز به داده‌های در حال انتقال روی شبکه است، که یا مبتنی بر اترنت یا بر اساس IP یا داده‌های ذخیره شده است [۲۲][۱۹]. سیستم‌های تشخیص نفوذ مبتنی بر میزبان، مبتنی بر چند میزبان و مبتنی بر شبکه هستند. سیستم مبتنی بر میزبان داده‌ها را از یک میزبان بازرسی می‌کند، سیستم چند میزبان چندین میزبان را بازرسی می‌کند و سیستم مبتنی بر شبکه داده‌های ترافیک شبکه و نیز میزبان را بازرسی می‌کند تا نفوذ را تشخیص دهد [۲۱]. دو روش مورد استفاده در سیستم تشخیص نفوذ وجود دارد، که یکی بر اساس امضاء است، که مطابق با هویت قبلی امضای ذخیره شده در پایگاه داده است و دومی بر اساس رفتار است.

۴-۲-۳- رمزنگاری (EFS و CFS، SFS) [۳۳,۳۴،۳۵]

رمزنگاری فایل سیستم[۶۳] برای تبدیل محتوای داده ذخیره شده در منبع به شکل دیگری به وسیله اضافه کردن رشته اضافی است که رمزنگاری نامیده شده است و هنگامی که داده به مقصد منتقل شد به شکل اصلی درآورده خواهد شد که رمزگشایی نامیده شده است. به طور معمول دو طرح رمزنگاری مورد استفاده برای امنیت اطلاعات وجود دارد، اولی رمزنگاری از قبل محاسبه شده که اطلاعات را به شکل رمز شده با نیاز به کلیدهای دارای طول عمر طولانی ذخیره می‌کند و سرور با رمزنگاری و رمزگشایی از کار نخواهد افتاد، دیگری رمزنگاری سیمی است که در این روش داده‌ها قبل و بعد از فرستادن اطلاعات روی شبکه رمزنگاری و رمزگشایی خواهند شد. رمزنگاری سیمی محتوا را برای انتقال داده کلید می‌کند و هر دو سرور و کلاینت بار را در پردازنده تحمل می‌کنند.

رمزنگاری فایل سیستم در ابتدا در آزمایشگاه AT و T BELL توسعه داده شد، داده رمزنگاری می‌شود به طوری که وقتی کاربر مجاز می‌خواهد به داده دستیابی داشته باشد، با صدور فرمانی داده را رمزگشایی می‌کند و می‌تواند از داده استفاده کند. اخیرا"، رمزنگاری شفاف فایل سیستم[۶۴] امنیت و احراز هویت قدرتمندی برای کاربران فایل سیستم فراهم می‌کند و یک نقطه ضعف آن عملکرد کندش است. فایل سیستم امن[۶۵] مبتنی بر درایور MS-DOS است که قصد دارد تمام بخش‌ها (پارتیشن‌ها) را رمزنگاری کند [۲۴]. داده یکبار رمزنگاری می‌شود، درایور نمای (راهنمای) رمزگشایی از داده رمزنگاری شده را نمایش می‌دهد و به این ترتیب امکان استخراج آسان داده اصلی فراهم می‌شود. اشکال عمده فایل سیستم امن این است که زیاد به MSDOS متکی است و امنیتی مانند سیستم عامل‌های جدید را فراهم نمی‌کند. فایل سیستم رمز شده[۶۶] در هسته ویندوز NT مایکروسافت استفاده می‌شود و یک توسعه از فایل سیستم NTFS است و روش‌های احراز هویت ویندوز و همچنین لیست‌های کنترل دسترسی ویندوز را فراهم می‌کند. رمزنگاری EFS داده‌ها، با استفاده از کلید طولانی مدت و کلیدهای ذخیره شده در جعبه قفل شده روی دیسک است که بیشتر با رمزهای ورود به سیستم کاربر رمزنگاری می-شوند.

۴-۲-۴- احراز هویت و مجوز [۳۶]

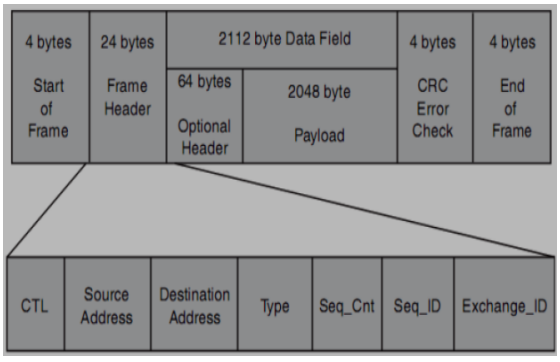
احراز هویت فرآیند تأیید هویت شخصی است که می‌خواهد وارد سیستم شود. استانداردهای خاص جدید ذخیره‌سازی و پروتکل‌های احراز هویت مانند Diffie-Hellman CHAP [۲۷] برای زیرساخت ذخیره‌سازی در حال ظهور هستند. احراز هویت در زیرساخت ذخیره-سازی می‌تواند با استفاده از عوامل زیر انجام شود: ۱- کسی که شناخته شده است (تصدیق پسورد) مانند رمزعبور یا شماره پین[۶۷]، ۲- چیزهایی که اشخاص دارند (عامل دوم احراز هویت) مانند کلید، نشانه رمز، کارت هوشمند، ۳- چیزهایی که در اشخاص وجود دارد (عامل

سوم احراز هویت) مانند انگشت نگاری، بیومتریک (سیستم‌هایی کـه از خصوصیات فیزیولوژیکی و رفتاری انسان جهت شناسایی استفاده مـی-کنند مانند اثر انگشت و شبکیه چشم، الگوهـای صـوتی و چهـره و ...) [۲۵,۳۷].

به طور کلی روش‌های احراز هویت امن‌تـر، قابـل حمـل، انعطـاف پذیر، استفاده آسان و همیشه قابل ارتقاء هستند. دو نوع احـراز هویـت وجود دارد: ۱- احراز هویت کاربر: فرآیند تصمیم‌گیری درباره این است که کاربر همان کسی است که او ادعا می‌کنـد، ۲- احـراز هویـت نهـاد: فرآیند تصمیم‌گیری درباره این است کـه اگـر یـک نهـاد (موجودیـت) وجود دارد همان کسی است که آن ادعا می‌کند. سیستم احـراز هویـت مبتنی بر نقش یکی از روش‌های احراز هویت نهاد است کـه در آن بـه کاربران یک یا چند نقش باز تعریف اختصاص داده شـده اسـت و ایـن نقش‌ها امتیازات کاربران را فراهم می‌آورند. مجموعه‌ای از امتیازات، این اجازه را می‌دهد که کاربران به ناحیه خاصی دسترسـی داشـته باشـند، آنها می‌توانند آن ناحیه خاص را ببینند و دستکاری کنند. مجوز فرآیند کنترل دسترسی و حقوق داخل منابع و فایل سیستم‌ها است.

۴-۲-۵- امنیت کانال فیبری [۳۸]

همانطور که قبلا" گفته شد، در شبکه‌های منطقه‌ای ذخیره‌سازی، کانال فیبری رسانه انتقال است و کانال فیبری دارای معمـاری ۵ لایـه است که در لایه دوم کانال فیبری لایه پروتکل فریم کـردن یـا کنتـرل جریان است که هدف اصلیش رسیدگی به ضعف امنیتی انتقـال اسـت [۱۶]. لایه دوم دارای اطلاعات سرآیند (سرآیند)[۶۸] از هر فریم است که در شکل ۳ نشان داده شده است.

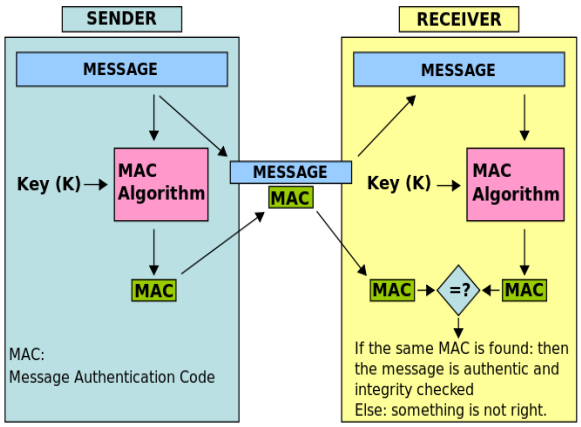


شکل ۱۱: سرآیند فریم کانال فیبری

اطلاعات سرآیند فریم‌ها نقطه مرکزی ضعف هستند که مهاجم می‌تواند از داده‌های منتقل شده در آن سوء اسـتفاده کنـد. سـرآیند شـامل ۲۴ بیت آدرس است که ID پورت گره منبع و آدرس گـره مقصـد نامیـده می‌شود. نهادهای زیـر در سـرآیند فـریم قـرار دارنـد [۱۶]: ۱- آدرس منبع[۶۹] ۲- آدرس مقصد[۷۰] ۳- آیدی ترتیب[۷۱] و آیدی تبـادل ۴- تعـداد ترتیب[۷۲] ۵- آیدی تبادل فرستنده[۷۳] ۶- آیدی تبادل گیرنده[۷۴] ۷- نوع و کنترل مسیریابی[۷۵].

یکی از حمله‌های خاص و مطرح شبکه‌های منطقه‌ای ذخیره‌سازی کـه بسیار خطرناک است حمله مرد در میانه[۷۶] است کـه بـه عنـوان تـلاش برای شکستن امنیت شبکه با جلـوگیری یـا تغییـر داده‌هـای در حـال انتقال از طریق آن تعریف شده است. حمله مرد در میانه که بـه خـوبی شناخته شده است، حمله‌ای است که در آن کاربر محول شده پاکت‌هـا را از شبکه ردیابی[۷۷] می‌کند، سپس این پاکت‌ها دستکاری مـی‌شـوند و آنها را به پشت شبکه اضافه می‌کند [۲۶]. مشابه این کـه چگونـه یـک پاکت IP برای پاکت‌های مسیر استفاده شـده اسـت، آدرس ۲۴ بیتـی برای فریم‌های مسیر از یک گره به گره دیگر استفاده شده است.

دو حمله معمول مرد در میانه در شبکه منطقه‌ای ذخیره‌سازی، نام گسترده جهانی روی وفق‌دهنده گذرگاه میزبان و حمله روی نرم-افزار کنسول مدیریت برای بدست آوردن کنترل نام کاربری و پسوردها از طریق سرویس‌های راهنما است. لایه دوم کانال فیبری در شبکه منطقه‌ای ذخیره‌سازی به طور عمده تحت تأثیر این حمله قرار می-گیرد، که در آن کاربر غیر مجاز کنترل داده‌های در حال انتقال از طریق کانال فیبری را در دست می‌گیرد. برای مدیریت یا جلوگیری از این نوع حملات پروتکل‌های حمل و نقل امنی معرفی شده‌اند از جمله لایه سوکت امن[۷۸]، امنیت لایه انتقال[۷۹] و پوسته امن[۸۰]. لایه سوکت امن یک کلید عمومی رمزنگاری برای پروتکل تونل‌زنی است و خدمات اینترنتی مختلفی می‌تواند از طریق آن لوله انجام شود. پوسته امن یک برنامه اجرایی کنترل از راه دور بر روی شبکه برای اجرای برخی از عملیات‌ها و انتقال داده از یک مکان به مکان دیگر به شیوه‌ای امن است. امنیت لایه انتقال یک پروتکل است که یکپارچگی داده‌ها را از منبع به مقصد حفظ می‌کند. امنیت لایه انتقال از دو لایه پروتکل ضبط (که دو وظیفه رمزنگاری[۸۱] و یکپارچگی داده‌ها را بر عهده دارد) و پروتکل دست‌دهی[۸۲] (که وظیفه توزیع کلید متقارن و نیز محاسبه مک[۸۳] را بر عهده دارد) تشکیل شده است. در رمزنگاری، یک کد تصدیق پیام (مک)، یک قطعه کوتاه از اطلاعات مورد استفاده برای اعتبار دادن به یک پیام و برای فراهم آوردن یکپارچگی و تضمین صحت روی پیام است. تضمین صحت بر اینکه پیام اصل است تأکید می‌کند.



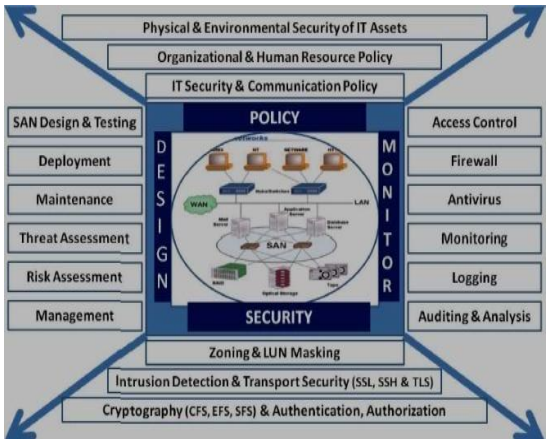
شکل ۱۲: نموداری از نحوه محاسبه کد تصدیق پیام (مک)

۴-۲-۶- امنیت به وسیله نرم‌افزار مدیریت شبکه منطقه‌ای ذخیره‌سازی

نرم‌افزارهای مدیریت ذخیره‌سازی از جمله سیستم عامل یا مدیر حجم، برای محدود کردن دسترسی به داده ذخیره شده در شبکه منطقه‌ای ذخیره‌سازی مورد استفاده هستند. سیستم عامل مبتنی بر قطعه‌بندی[۸۴] یا بخش‌بندی[۸۵] ذخیره‌سازی ممکن است برای محدود کردن دسترسی مفید باشد. سه نوع نرم‌افزار مبتنی بر امنیت که می‌توانند در ذخیره‌سازی‌های مهم پیاده‌سازی شوند وجود دارد، که شامل نرم‌افزار نظارت و مدیریت، نرم‌افزار کنترل دسترسی و امنیت و دسته سوم نرم‌افزار ویژه و خاص امنیت است [۲۰]. مدیر شبکه منطقه‌ای ذخیره‌سازی حقوق را ایجاد و برای گروه‌های مختلف و کاربران مطرح می‌کند. نرم‌افزارهای مدیریت در نظارت، بازرسی، و ثبت وقایع در تمام فعالیت‌های دسترسی به فایل‌ها در داخل ذخیره‌سازی پیش قدم هستند. حداکثر امنیت با بررسی ویژگی‌های فایل‌ها بدست خواهد آمد از جمله اندازه فایل با استفاده از چک سام[۸۶] که عموما" قسمتی از یک فایل است و وظیفه آن حفاظت از کل فایل در برابر تغییرات است، یا تاریخ آخرین تغییر. نظارت مداوم بر هر تجاوز یا هر تلاشی برای تجاوز یا هر گونه تغییرات در فایل‌ها در ردیابی نفوذ مفید خواهد بود. ثبت وقایع از تمام فعالیت‌ها در ذخیره‌سازی و بازرسی این داده‌ها در تأمین امنیت اطلاعات ذخیره شده کمک خواهد کرد. به طور کلی کنسول نرم‌افزار مدیریت توسط مدیر محیط شبکه منطقه‌ای ذخیره‌سازی استفاده می‌شود و به محافظت از داده‌ها با استفاده از خدمات نرم‌افزاری کمک می‌کند.

۵- پیشنهادات

چارچوب امنیتی شبکه منطقه‌ای ذخیره‌سازی شامل بسیاری از سیستم‌های مستقل و به هم پیوسته است که با یکدیگر همکاری می‌کنند تا بالاترین سطح امنیت برای داده‌های مهم مأموریتی در یک سازمان بدست آید. ملاحظات امنیتی شبکه‌های منطقه‌ای ذخیره‌سازی در واقع با چند عامل مورد توجه از طراحی مرحله به مرحله تا تست و اعتبارسنجی این شبکه‌های ذخیره‌سازی شروع می‌شود. شکل زیر روش‌های امنیتی را به عنوان یک چارچوب برای به حداکثر رسیدن امنیت داده‌های مهم تجاری نشان می‌دهد. امنیت جامع شبکه منطقه‌ای ذخیره‌سازی به ۴ قسمت تقسیم شده است که عبارتند از: ۱) سیاست ۲) طراحی شبکه منطقه‌ای ذخیره‌سازی ۳) تکنیک‌های امنیت داده ۴) نظارت، ثبت وقایع و بازرسی.



شکل ۱۳: یک چارچوب از روش‌های امنیتی برای بهبود امنیت شبکه منطقه‌ای ذخیره‌سازی

۱) سیاست:

سیاست سازمانی یک طرح کلی از سیاست سازمان فراهم می‌آورد که شامل IT و سیاست امنیتی مرتبط با امنیت ذخیره‌سازی است. سیاست امنیتی IT، اطلاعاتی جزئی از اینکه چه شخصی به چه چیزی دستیابی داشته باشد مشخص می‌کند. به عنوان مثال یک مهاجم با اطلاعات حساب‌های شرکت نمی‌تواند به بخش تولید دستیابی پیدا کند. منابع انسانی اطلاعاتی در مورد اینکه چه شخصی کجا باید باشد فراهم می‌آورند. ارتباط کاری به وسیله منبع انسانی[۸۷] انجام می‌پذیرد و منبع انسانی، افراد درست را برای کار صحیح شناسایی می‌کند.

امنیت محیط زیستی و فیزیکی شامل مراحلی برای رسیدگی به بلایای طبیعی مانند آب، آتش و سرقت فیزیکی است، همچنین این امنیت شامل تعیین مکان و سیستم‌های هشدار و اهمیت جزئیات طبقه‌بندی دارایی منابع شبکه منطقه‌ای ذخیره‌سازی به عنوان تجهیزات IT که باید خصوصی طبقه‌بندی شوند و کمتر دستکاری شوند، و مردم مرتبط در یک سازمان، نیز می‌شود. در نهایت، سیاست ارتباطی طرح کلی از استفاده مناسب از سیستم‌های ارتباطی مانند ایمیل‌ها و دیگر رسانه‌های الکترونیک را ارائه می‌کند به طوری که اطلاعات شرکت به صورت غیر قانونی با استفاده از این منابع منتقل نخواهد شد.

۲) طراحی شبکه منطقه‌ای ذخیره‌سازی:

از آنجا که هر سازنده، شبکه منطقه‌ای ذخیره‌سازی را با روش‌های امنیتی مختلف طراحی می‌کند، امنیت شبکه منطقه‌ای ذخیره‌سازی از انتخاب خود سازندگان مختلف شروع می‌شود. در طول فاز تجزیه و تحلیل، نیازمندی شبکه منطقه‌ای ذخیره‌سازی هر سازمان باید زمان مناسبی برای ارزیابی انتخاب فروشنده صحیح با توجه به نیاز دقیق یک سازمان صرف کند. روش‌های امنیتی از فروشنده به طراحان فروشنده شبکه منطقه‌ای ذخیره‌سازی مختلف هستند و انتخاب مناسب سیستم شبکه منطقه‌ای ذخیره‌سازی مطابق با محیط سازمان‌ها نیمی از بار موضوعات امنیتی را کاهش خواهد داد. پس از نصب و راه اندازی شبکه منطقه‌ای ذخیره‌سازی، با حفظ و نگهداری از شبکه منطقه‌ای ذخیره-

سازی و انجام اقدام پیشگیرانه به طور منظم، در واقع قابلیت اطمینان شبکه منطقه‌ای ذخیره‌سازی افزایش خواهد یافت. بررسی دوره‌ای تک تک اجزاء مانند سویچ‌ها، کانال فیبری و دیسک‌های سخت داخلی به افزایش طول عمر شبکه منطقه‌ای ذخیره‌سازی کمک می‌کند. تجزیه و تحلیل خطر و تهدید شبکه منطقه‌ای ذخیره‌سازی در شناسایی انواع حملات و روش‌های اصلاح کننده امنیتی در مقابل این حملات کمک می‌کند. مدیریت عملیات شبکه منطقه‌ای ذخیره‌سازی برای عملکرد نرم و ملایم شبکه‌های منطقه‌ای ذخیره‌سازی باید به خوبی طراحی شود.

۳) تکنیک‌های امنیت داده:

تکنیک‌های امنیتی ویژه، به مدیریت تک تک خطرات و تهدیدات داده-های ذخیره شده در شبکه منطقه‌ای ذخیره‌سازی کمک خواهد کرد. هر تکنیک امنیتی خاص، جایی که حملات مؤثر اجرا می‌شوند را ایمن نگه می‌دارد و در حفاظت از اطلاعات مهم مأموریتی ذخیره شده در شبکه منطقه‌ای ذخیره‌سازی کمک می‌کند، که این تکنیک‌های خاص قبلا" توضیح داده شده‌اند.

۴) نظارت، ثبت وقایع و بازرسی:

سیستم کنترل دسترسی، دیواره آتش و آنتی ویروس، سیستم‌های دفاعی برای شبکه منطقه‌ای ذخیره‌سازی هستند تا تهدیدات را به صورت مفید مدیریت کنند. بخصوص دیواره آتش که برای جلوگیری از دستیابی غیر مجاز خارجی کمک می‌کند و آنتی ویروس که در امن نگه داشتن داده‌ها و سیستم‌ها در برابر کرم‌ها و ویروس‌های مخرب کمک می‌کند. نظارت، ثبت وقایع و بازرسی، بسیاری از اطلاعات حیاتی را از تاریخچه دسترسی و جزئیات تغییرات داده‌های ذخیره شده خواهد داد و با این اطلاعات ما می‌توانیم به آسانی اطلاعاتی را در مورد اینکه چه شخصی به چه چیزی دستیابی داشته است بدست بیاوریم. نظارت کانال فیبری شبکه در شناسایی انتقال داده‌های غیر مجاز کمک خواهد کرد. بازرسی دوره‌ای از اطلاعات وارد شده و اقدام لازم برای گرفتن هر تاریخچه دسترسی به داده، راه بسیار امن‌تری برای جلوگیری از سرقت داده‌ها است.

۶- نتیجه‌گیری

تأمین امنیت به کمک روش‌های رمزنگاری کارایی بالایی دارد اما سربارهای زیادی نیز به سیستم تحمیل می‌کند [۳۹]. روش‌های مبتنی بر تشخیص نفوذ سربار کمی روی کارایی شبکه‌های منطقه‌ای ذخیره-سازی دارند و کاهش کارایی توسط آن به دلیل تأمین مطلوب امنیت این شبکه‌ها قابل چشم‌پوشی است. برای افزایش بهره‌وری، به منظور برآورده کردن تقاضای مشتری و پیشروی در صنعت، هر سازمان باید داده‌های ارزشمند داخلی و خارجی خود را ایمن کند. علاوه بر این، داده‌های شرکت نیاز به نظارت مداوم، ثبت وقایع، بازرسی و ارزیابی برای احتمال خطرات و تهدیدات دارد. شبکه‌های منطقه‌ای ذخیره-سازی یکی از فناوری‌های مؤثر برای هر سازمان هستند که نیاز به محرمانگی، یکپارچگی، مقیاس پذیری، در دسترس بودن و کارایی بالای راه‌حل ذخیره‌سازی دارند. از آنجا که شبکه منطقه‌ای ذخیره-سازی داده‌های مأموریتی مهمی را نگهداری می‌کند که برای مهاجمان بسیار آسیب‌پذیر هستند و نیاز به در نظر گرفتن چند عامل برای رسیدگی به مسائل امنیتی دارند، این تحقیق نه تنها به خلاصه‌ای از چارچوب امنیتی شبکه منطقه‌ای ذخیره‌سازی می‌پردازد بلکه روش‌های امنیتی استاندارد و همچنین دیگر عوامل در نظر گرفته شده در هنگام طراحی سیستم امنیتی مؤثر برای شبکه منطقه‌ای ذخیره‌سازی را نشان می‌دهد.

۷- منابع

[۱] "*Information Storage And Management*," Published By Wiley Publishing, Inc.Edited by G. S. AlokShrivastava, Indianapolis, Indiana, ۹۷۸-۰-۴۷۰-۲۹۴۲۱-۵, ۲۰۰۹.

[۲] J. Jordan, Storage Consolidation: "Moving from DAS to SAN/NAS," *Dell White Paper*, April, ۲۰۰۲.

[۳] T. Anderson, M. Dahlin, J. Neefe, D. Patterson, D. Roselli, R. Wang, "Serverless Network File Systems," *ACM Transactions on Computer Systems ۱۴(۱)*, pp. ۴۱-۷۹, February ۱۹۹۶.

[۴] G. A. Gibson, D. F. Nagle, K. Amiri, J. Butler, F. W. Chang, H. Gobioff, C. Hardin, E. Riedel, D. Rochberg, and J. Zelenka, "A cost-effective, high-bandwidth storage architecture," *in ASPLOS-VIII: Proc. of ۱۸th international conference on Architectural support for programming languages and operating systems*, vol. ۳۲, no. ۵. New York, NY, USA: ACM Press, pp. ۹۲-۱۰۳, December ۱۹۹۸.

[۵] E. Riedel, M. Kallahalla, and R. Swaminathan."A Framework for Evaluating Storage System Security," *Proc. of the ۱st Conference on File and Storage Technologies (FAST'۰۲)*, Monterey, CA, January, ۲۰۰۲.

[۶] C. Rhodes, "Security Considerations for Storage Area Networks". *The Infosec Writers Text Library*, East Carolina University, December, ۲۰۰۵.

[۷] W. E. Freeman, E. L. Miller, "Design for a decentralized security system for network-attached storage," *Proc. of the ۱۷th IEEE Symposium on*

[١٩] K. K. Gupta, B. Nath, K. Ramamohanarao, "Network Security Framework. IJCSNS International Journal of Computer Science and Network Security," Vol ٦ No ٧B, pp. ١٥١-١٥٧, July ٢٠٠٦.

[٢٠] B.Aziz, S.Foley, J. Herbert and G. Swart, "Configuring Storage Area Networks Using Mandatory Security," *Journal of Comp*, Volume ١٧, Issue ٢, pp. ١٩١-٢١٠, ٢٠٠٩.

[٢١] D. Denning, "An Intrusion-detection model," *IEEE Transactions on Software Engineering*, vol.١٣, Issue ٢, pp. ٢٢٢-٢٣٢, ١٩٨٧.

[٢٢] H. Debar, M. Becke and D. Siboni, "A Neural network component for an intrusion detection system," *In Proceedings of the IEEE Computer Society Symposium on Research on Security and Privacy*, pp. ٢٤٠, ١٩٩٢.

[٢٣] L. Vancura, "Building a Security Policy Frameworkfor Large," Multinational Company, GSec Practical, Ver١,٤c, option ٢, SANS Institute, InfoSec reading Room, January, ٢٠٠٥.

[٢٤] C. P. Wright, M. C. Martino, E. Zadok: NCryptfs: "A Secure and Convenient Cryptographic File System," *USENIX Annual Technical Conference*, General Track, pp. ١٩٧-٢١٠, ٢٠٠٣.

[٢٥] SafeNet, "Multi-Factor Authentication – Protecting Applications and Critical Data against Unauthorized Access," SafNet Inc., ٢٠٠٨.

[٢٦] B. B. Bhansali, "Man-In-the-Middle Attack," *GIAC Practical Reporsitory*, SANS Institute, ٢٠٠٠ – ٢٠٠٢.

[٢٧] W. Diffie and M. E. Hellman, "New directions in cryptography," *IEEE Transactions on Information Theory*, Volume ٢٢, Issue٦, pp. ٦٤٤-٦٥٤, ١٩٧٦.

[٢٨] Jiwu Shu, Bigang Li, and Weimin Zheng, "Design and Implementation of an SAN System Based on the Fiber Channel Protocol," *IEEE Transactions on Computers*, Vol. ٥٤, No. ٤, April ٢٠٠٥.

[٢٩] L. Lu, D. Hildebrand, R. Tewari, "Zone-Based Data Striping for Cloud Storage," *IBM Mass Storage Systems and Technologies*, pp. ٣٦١–٣٧٣, March ٢٠٠٠.

[٨] H. Gobioff, G. Gibson, and D. Tygar, "Security for network attached storage devices," Technical Report CMUCS ٩٧-١٨٥, Carnegie Mellon, October ١٩٩٧.

[٩] M. RJT & T. BJ, "The evolution of Storage Systems," *IBM Systems Journal*, Volume ٤٢, Issue ٢, *Proquest Science Journal*, pp. ٢٠٥-٢١٧, ٢٠٠٣.

[١٠] S. Duplessie, "Storage and Information Brief," *Enterprise Strategy Group*, September ٢٠٠٤.

[١١] P. Ross, "Design Considerations in Enterprise Storage Networks," Industry Trend or Event. Computer Technology Review, November, ٢٠٠٠.

[١٢] T. Clark, "Designing Storage Area Networks," ٢nd Edition, Addison-Wesley Longman Publishing Co., Inc, ISBN: ٠-٣٢١- ١٣٦٥٠-٠, March ٢٠٠٣.

[١٣] J. D. Coffed, "Security for the SAN Workgroup," *ATTO Technology*, ٢٠٠٠.

[١٤] U. Khan, "Consolidating and Securing Enterprise Storage," *GIAC Security Essentials Certification Practical*, Version ١,٤, Option ١, SANS Institute, ٢٠٠٣.

[١٥] Dr. W. L. McKnight, "What Is Information Assurance?," *The Journal of Defense Software Engineering*, July ٢٠٠٢.

[١٦] H. Dwivedi, "Securing Storage: A Practical Guide to SAN and NAS Security," Addison – Wesley Professional, November ٢٠٠٥.

[١٧] M. Banikazemi, D. Poff, and B. Abali, "*Storage*-Based Intrusion Detection in Storage Area Networks (SANs)," *IEEE/NASA Conference on Mass Storage Systems and Technologies (MSST'٠٥)*, April ٢٠٠٥.

[١٨] J. C Sipior and B. T Ward. "A Framework for Information Security Management Based on Guiding Standards," *A United States Perspective, Information Science and Information technology*, Volume ٥, ٢٠٠٨.

Journal of Research and Development, Vol. ٥٥, No. ٦, Paper ١ November/December ٢٠١١.

[٣٠] Avita Katal, Niharika Gupta, Seepaj Sharma and R.H. Goudar, "Information Storage on the Cloud: A Survey of Effective Storage Management System," *Engineering and Systems (SCES)*, ٢٠١٢.

[٣١] Mohammad Banikazemi, Dan Poff, Bulent Abali ,Thomas J. Watson, "Storage-Based Intrusion Detection for Storage Area Networks (SANs)," *Proceedings of the ٢٢nd IEEE / ١٣th NASA Goddard Conference on Mass Storage Systems and Technologies (MSST'٠٥)*, ٢٠٠٥.

[٣٢] Yacine Djemaiel, Noureddine Boudriga, "Dynamic detection and tolerance of attacks in Storage Area Networks," *Advanced Information Networking and Applications, ٢٢nd International Conference on*, ٢٠٠٨.

[٣٣] Yongdae Kim, Fabio Maino, Maithili Narasimha, Kyung Hyune Rhee, Gene Tsudik, "Secure Group Key Management for Storage Area Networks," *IEEE Communications Magazine*, August ٢٠٠٣.

[٣٤] Charles P. Wright, Michael C. Martino, and Erez Zadok, "NCryptfs: A Secure and Convenient Cryptographic File System," *Appears in the General Track of the USENIX 2003 Annual Technical Conference*, ٢٠٠٣.

[٣٥] Roman Pletka, Christian Cachin, "Cryptographic Security for a High-Performance Distributed File System," *24th IEEE Conference on Mass Storage Systems and Technologies (MSST)*, ٢٠٠٧.

[٣٦] De-Zhihan, Jian-Zhong Huang, "Security for the Storage Network Merging NAS and SAN," *Proceedings of the Fifth International Conference on Machine Learning and Cybernetics*, Dalian, ١٣-١٦ August ٢٠٠٦.

[٣٧] Simon Liu, Mark Silverman, "A Practical Guide to Biometric Security Technology," *IEEE Journals and Magazines*, ٢٠٠١.

[٣٨] Casimer DeCusatis, "Developing a Threat Model for Enterprise Storage Area Networks," *Proceedings of the 2006 IEEE Workshop on Information Assurance*, ٢٠٠٦.

[٣٩] Tao Cai, Shiguang Ju, JunJie Zhao and Wei Zhong, "Performance Study of Cryptographic Storage Area Network," *2007 IFIP International Conference on Network and Parallel Computing*, ٢٠٠٧.

---

١ SAN (Storage Area Network)
٢ Monitoring
٣ SNIA
٤ HBA (Host Bus Adaptors)
٥ RAID (The Redundant Array of Inexpensive Disk)
٦ JBOD (Just a Bunch of Disks)
٧ FC-AL (Fibre Channel Arbitrated Loop)
٨ Fibre Channel switched fabric (FC-SW)
٩ Ring Topology
١٠ NL_port (node loop port)
١١ FL_port (fabric loop port)
١٢ Fabric logon (FLOGI)
١٣ Hub
١٤ Broadcast Networks
١٥ Active
١٦ Passive
١٧ DoS (Denial of Service)
١٨ Repudiation Attacks
١٩ Defense In Depth
٢٠ Attack surface
٢١ attack vector
٢٢ Bug
٢٣ Snoop Attack
٢٤ Work factor
٢٥ Queries
٢٦ Administrative Controls
٢٧ IDS/IPS
٢٨ Firewall
٢٩ Close-in
٣٠ Insider
٣١ Distribution
٣٢ Eavesdrop
٣٣ Fabric name
٣٤ Domain identification
٣٥ Switch name server information
٣٦ Session sequence control number
٣٧ Session sequence IDs
٣٨ World Wide Names used in the fabric
٣٩ Layer-٢ frame information
٤٠ ٢٤-bit addresses
٤١ Routing information
٤٢ Management information
٤٣ World Wide Name (WWN)
٤٤ World Wide Identifier (WWID)
٤٥ WWNN (World Wide Node Name)
٤٦ WWPN (World Wide Port Name)
٤٧ Accountability service
٤٨ log
٤٩ Auditing

٥٠ Confidentiality service
٥١ Integrity service
٥٢ Availability service
٥٣ Logging
٥٤ Zoning & LUN masking
٥٥ Intrusion Detection System (IDS)
٥٦ Authentication and Authorization
٥٧ Fibre Channel fabric
٥٨ Failover
٥٩ Datagram
٦٠ Logical Unit Number (LUN)
٦١ Access control list (ACL)
٦٢ Monitoring
٦٣ The Cryptographic File System (CFS)
٦٤ Transparent Cryptographic File System (TCFS)
٦٥ Secured File System (SFS)
٦٦ Encrypted File System (EFS)
٦٧ PIN
٦٨ Header
٦٩ Source Address (S_ID)
٧٠ Destination Address (D_ID)
٧١ Sequence Id (SEQ_ID)
٧٢ Sequence Count (SEQ_CNT)
٧٣ Originator Exchange ID (OX_ID)
٧٤ Recipient Exchange ID (RX_ID)
٧٥ Type & Routing Control (R_CTL)
٧٦ Man-in-the-middle
٧٧ Sniff
٧٨ Secure Socket Layer (SSL)
٧٩ Transport Layer Security (TLS)
٨٠ Secure Shell (SSH)
٨١ Encryption
٨٢ Handshake
٨٣ MAC (Message Authentication Code)
٨٤ Segmentation
٨٥ Partitioning
٨٦ Checksum
٨٧ HR (Human Resource)