مجله علمی- پژوهشی "پدافند الکترونیکی و سایبری"
سال ششم، شماره ۲، تابستان ۱۳۹۷، ص ۲۰-۱۱

تخمین کور تعداد کاربران در شبکه‌های TDMA با استفاده از اضافات کدگذاری تطبیقی کانال

سید مرتضی احمدیان[۱]، مهدی تیموری[۲*]

۱- دانشجوی کارشناسی ارشد، ۲- استادیار، دانشکده علوم و فنون نوین- دانشگاه تهران

(دریافت: ۹۵/۰۷/۲۴، پذیرش: ۹۶/۰۸/۱۱)

چکیده

انجام یک مهندسی معکوس (یا شناسایی کور) کامل دریک محیط جنگ الکترونیک، منجر به آشکارسازی اطلاعات انتقالی توسط یک سیگنال دریافتی می‌گردد. تمرکز تحقیقات انجام‌شده در حوزهٔ شناسایی کور، بر روی ارتباطات مخابراتی یک‌سویه و غیر شبکه‌ای بوده است که در آنها هدف آشکارسازی اطلاعات ارسالی توسط یک فرستندهٔ مشخص می‌باشد. اولین قدم در شناسایی شبکه‌های مخابراتی، تخمین تعداد کاربران فعال در این شبکه‌ها می‌باشد. در این مقاله، مسئلهٔ تخمین تعداد کاربران در شبکه‌های مبتنی بر دسترسی چندگانهٔ تسهیم زمانی (TDMA) مورد بررسی قرار می‌گیرد. این تخمین می‌تواند با استفادهٔ مستقیم از سیگنال‌های الکترومغناطیسی دریافتی و تحلیل لایه فیزیکی انجام شود. اما به دلایل مختلفی مانند محدودیت‌های سخت‌افزاری و یا عدم حصول نتایج دقیق به دلیل نزدیکی کاربران فعال به یکدیگر، همیشه امکان استفاده از این روش وجود ندارد. در چنین شرایطی، روش دیگر برای تخمین تعداد کاربران، استفاده از تحلیل اطلاعات لایه‌های بالاتر در ساختار شبکه می‌باشد. در این مقاله، با استفاده از اطلاعات اضافیِ تحمیلی توسط کدگذاری تطبیقی کانال، روشی برای تخمین تعداد کاربران فعال شبکه ارائه شده است. نتایج شبیه‌سازی نشان می‌دهد که روش پیشنهادی دارای مقاومت مناسبی در برابر خطای کانال می‌باشد به طوری که در نسبت سیگنال به نویز ۷/۳ دسی‌بل، دقت الگوریتم حدود ۸۰٪ است.

واژه‌های کلیدی: تخمین کور تعداد کاربران، شبکه‌های TDMA، کدگذاری تطبیقی کانال، یادگیری ماشین

۱- مقدمه

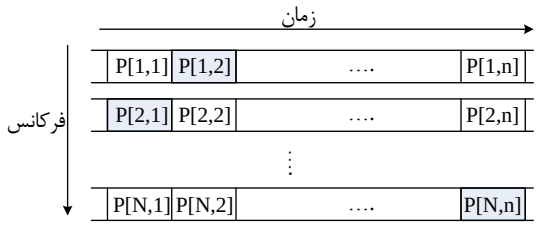
امروزه با طیف وسیعی از سامانه‌های مخابراتی روبرو هستیم که در اطراف ما در حال ارسال و دریافت اطلاعات می‌باشند. با در نظر گرفتن مسئله امنیت لایه فیزیکی، مسائل زیادی در این حوزه دارای جذابیت می‌باشند. این مسائل عموماً می‌توانند در یک یا چند زمینه از زمینه‌های زیر خلاصه شوند: ۱- اختلال شبکه، ۲- کشف ساختار ارتباطی شبکه و شنود اطلاعات ارسالی، ۳- فریب شبکه و کاربران آن، ۴- نفوذ به شبکه و جا زدن خود به عنوان یک کاربر مجاز و ۵- کنترل کامل یا جزئی شبکه. در این مقاله بر روی زمینه دوم متمرکز می‌شویم، هرچند ارتباط تنگاتنگی بین این پنج زمینه وجود دارد. در حقیقت مجموعه دانشی وسیعی باید در کنار یکدیگر مورد استفاده قرار گیرند تا ما را به سمت تمام این اهداف سوق دهد. در یک تصور نسبتاً اشتباه، جنگ الکترونیک به دو بخش مجزای سایبری و تحلیل سیگنال تقسیم می‌شود که معمولاً در بخش سایبری فرض اصلی بر وجود یک شبکه کامپیوتری (مبتنی بر شبکه اینترنت) به عنوان بستر حمله است و در بخش تحلیل سیگنال، مسئلهٔ اصلی عموماً شناسایی یک سیگنال راداری یا یک سیگنال مدوله‌شده است. این نگاه در شرایط واقعی می‌تواند دردسرساز باشد، زیرا برای رسیدن و حمله به اطلاعات لایه‌های بالای یک شبکه، اولین قدم عبور از لایه فیزیکی است که تنها پیچیدگی آن مدولاسیون نیست، بلکه انواع کدگذاری کانال پیشرفته (و معمولاً ناشناخته) می‌تواند سر راه ما قرار گرفته باشد [۱].

از اوایل دهه ۹۰ میلادی با شروع پروژه‌های مختلف در زمینهٔ شناسایی مدولاسیون‌های دیجیتال کار در این حوزه شروع شد [۲، ۳]. اما قاعدتاً این نوع از فعالیت‌ها، اطلاعات بسیار کمی از داده ردوبدل شده در سامانه‌های مخابراتی به ما ارائه می‌دهد، زیرا در بهترین حالت این روش‌ها قادر خواهند بود رشته بیت ارسالی را آشکار کنند که این موضوع با توجه به سامانه‌های پیشرفته امروزی چندان جذاب نیست، زیرا قاعدتاً این رشته بیت‌ها حتی اگر رمزگذاری شده نباشند حاصل عمل مجموعه‌ای از کدهای منبع و کانال هستند که راه را برای رسیدن به محتوای اصلی دشوار می‌کند. با مطرح شدن چنین معضلی و با شروع قرن ۲۱، عملاً پروژه‌های زیادی در این راستا کلید خوردند. تمامی این

* رایانامه نویسنده مسئول: mehditeimouri@ut.ac.ir

پروژه‌ها با فرض اینکه یک و تنها رشته بیت پیوسته در اختیار دارند، سعی در مهندسی معکوس سامانه و رسیدن به محتوا دارند. با فرض عبور از لایه فیزیکی، تحقیقات زیادی در جهت شناسایی و نفوذ به لایهٔ پیوند[1] داده انجام شده است. برای مثال، شناسایی قاب‌بندی [۴، ۵]، شناسایی جایگردان (که برای مقابله با خطاهای متوالی اعمال می‌شود) [۶، ۷]، شناسایی درهم‌ساز (که برای هم‌زمان‌سازی بیتی فرستنده و گیرنده اعمال می‌شود) [۸، ۹] و نهایتاً شناسایی کدگذاری کانال (که برای تشخیص و تصحیح خطا در لایهٔ پیوند داده می‌باشد) [۱۰، ۱۱] موضوعاتی بوده است که مورد توجه محققان قرار گرفته است. ضعف اساسی همه تحقیقات مورد اشاره، فرض وجود یک ارتباط یک‌سویه و غیر شبکه‌ای است.

در عمل و در بسیاری از سامانه‌های تجاری و نظامی، ارتباط (چه بین دو سامانه به صورت نقطه به نقطه و چه بین چندین سامانه در ساختار سلولی و ...) به صورت ناپیوسته (بسته‌ای) و در N باند فرکانس انجام می‌پذیرد (شکل (۱) را ببینید). در چنین حالتی با به کار بردن چند دمدولاتور می‌توان اطلاعات موجود در این فرکانس‌ها را تبدیل به چندین رشته بیت کرد که هر رشته بیت اطلاعات مربوط به چندین نشست را منتقل می‌کند. مثلاً در شکل (۱) می‌توان ساختارهای مختلفی از ارتباطات را متصور شد. به عنوان مثال، این شکل می‌تواند مربوط به شبکه‌ای ارتباطی بین K کاربر باشد. مثلاً می‌توان در یک سناریوی محتمل فرض نمود که برست‌های مشخص‌شده توسط P[1,1], P[1,2],..., P[N,n], ... مربوط به یک ارتباط مشخص (مثلاً بین کاربر x و y) است. در چنین حالتی بسته به نوع کدگذاری به‌کاررفته (که می‌تواند یک کدگذاری مشخص یا یک کدگذاری تطبیقی باشد)، هر چند برست (یعنی همان بسته‌های لایه یک) می‌توانند عناصر تشکیل‌دهنده یک بسته در لایه پیوند داده (که آن را از این پس به اختصار لایه دو می‌نامیم و هدف آن کنترل جریان لایه فیزیکی است) باشند. به عنوان مثال در سامانه GSM و در کانال کنترلی، هر چهار برست نماینده یک بسته در لایه پیوند داده هستند. همچنین هر چند بسته در لایه پیوند داده یک بسته در لایه شبکه[2] را تشکیل دهد (که آن را از این پس به اختصار لایه سه می‌نامیم). لایه چهار که مطابق با مدل هفت لایه‌ای OSI برای شبکه‌های مخابراتی لایه انتقال[3] نام دارد، وظیفه کنترل جریان داده را به عهده دارد؛ در شبکه‌هایی که تک هاپ[4] هستند (یعنی فرستنده و گیرنده مستقیم با هم ارتباط دارند)، عملاً وجود چنین لایه‌ای مورد نیاز نمی‌باشد. سه لایه بالایی را هم می‌توان لایه کاربرد (یا لایه منبع) نامید. این لایه مسئول تفکیک، فشرده‌سازی و رمزگذاری محتوا است. این محتوا می‌تواند صوت، تصویر، داده (مانند پیام کوتاه، فایل، صفحات وب، فرمان کنترل و ...) مربوط به یک کاربر باشد. هدف اصلی یک سامانه شناسایی کور کشف تمامی ارتباط‌ها (مانند ارتباط بین x و y) و رسیدن به منبع اطلاعات ردوبدل شده در هر یک از آنها است. به عبارت دیگر هدف از طراحی چنین سامانه‌ای این است که اولاً بتوانیم تمامی نشست‌های را از یکدیگر جداسازی نماییم و ثانیاً بتوانیم هر نشست را با انجام مهندسی معکوس به محتوا (داده، صوت و تصویر) برسانیم.



شکل (۱): مثالی از یک شبکه مخابراتی دارای فعالیت در N فرکانس.

یکی از متداول‌ترین روش‌های تسهیم منابع در شبکه‌های بی‌سیم استفاده از دسترسی چندگانهٔ تسهیم زمانی (TDMA[5]) می‌باشد (حالت خاصی از شکل (۱) با N=۱ فرکانس فعال). همان‌طور که اشاره شد، روش‌های متداول شناسایی کور نمی‌توانند چنین شبکه‌هایی را شناسایی کنند و باید برای شناسایی این شبکه‌ها و استخراج اطلاعات هر یک از کاربران روش‌هایی جدید ارائه نمود. شناسایی کامل این شبکه‌ها به این معنی است که ابتدا قطعات زمانی مربوط به کاربران فعال را از هم تفکیک کنیم و سپس داده‌های مربوط به کاربران مختلف را با روش‌های متداول و موجود مهندسی معکوس کرده و محتوای اطلاعات ارسالی توسط هر کاربر را شناسایی کنیم.

یکی از مسائل مهم قبل از تفکیک کاربران تشخیص تعداد کاربران فعال شبکه می‌باشد تا بتوان در جهت تفکیک اطلاعات هر یک از کاربران شناسایی بهتر و سریع‌تری انجام داد. یک روش متداول برای تشخیص تعداد کاربران فعال در شبکه استفاده از ویژگی‌های سیگنال (شکل موج) ارسالی در لایه فیزیکی می‌باشد [۱۲]. در این روش با توجه به جهت و توان سیگنال‌های دریافتی در یک محیط، تخمینی از تعداد و موقعیت کاربران فعال در شبکه ارائه می‌شود. اما به دلایل مختلفی مانند محدودیت‌های سخت‌افزاری و یا عدم حصول نتایج دقیق به دلیل نزدیکی کاربران فعال به هم، استفاده از این روش مناسب نمی‌باشد. در

1- Data Link Layer
2- Network Layer
3- Transport Layer
4- Hop
5- Time Division Multiple Access

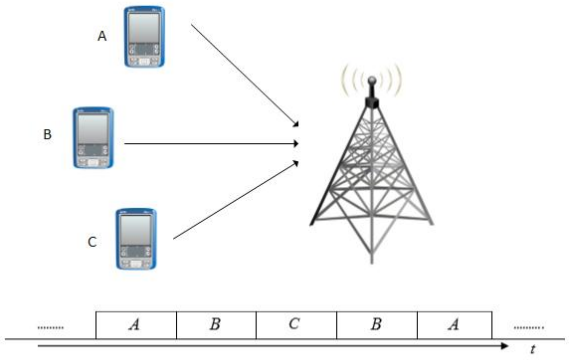
چنین شرایطی، روش دیگر برای تخمین تعداد کاربران، استفاده از تحلیل اطلاعات لایه‌های بالاتر در ساختار شبکه می‌باشد و تا جایی که می‌دانیم در این زمینه تحقیقات انتشاریافته‌ای در دسترس نیست.

با توجه به توضیحات فوق، در این مقاله با استفاده از اضافات تحمیلی توسط کدگذاری تطبیقی کانال روشی ابداعی برای تخمین تعداد کاربران فعال در شبکه‌های مخابراتی TDMA ارائه می‌شود. این روش در نوع خود اولین روش ارائه‌شده است که در تحلیل کور شبکه‌های مخابراتی بسیار کارساز خواهد بود. این مقاله در پنج بخش و به صورت زیر ارائه می‌گردد. در بخش دوم، مدل سامانه ارائه می‌شود. سپس، در بخش سوم روش پیشنهادی برای تخمین تعداد کاربران فعال تشریح می‌گردد. در بخش چهارم، جهت ارزیابی عملکرد روش پیشنهادی، نتایج شبیه‌سازی ارائه و تحلیل می‌شود. در انتها و بخش پنجم، به نتیجه‌گیری می‌پردازیم.

۲- مدل سامانه

یک شبکه مخابراتی را مطابق شکل (۲) در نظر می‌گیریم که از ساختار TDMA برای ارسال اطلاعات استفاده می‌شود. تمامی کاربران با ایستگاه مرکزی در ارتباط هستند و از یک فرکانس استفاده می‌کنند. ارتباطات در مسیر فراسو[1] در نظر گرفته می‌شود. طول هر قطعهٔ زمانی[2] در پنجرهٔ زمانی مشاهده‌شده برابر n بیت می‌باشد. هر پنجرهٔ زمانی مشاهده‌شده شامل چندین قطعهٔ زمانی است. هرچند ممکن است یک کلمه کد در چند قطعهٔ زمانی و یا چند کلمه کد در یک قطعهٔ زمانی ارسال شود؛ اما فرض ما این است که هر کلمه کد در یک قطعهٔ زمانی ارسال می‌شود و طول تمامی قطعات زمانی یا کلمه‌های کد با یکدیگر برابر است (در شکل (۲)، ۵ قطعهٔ زمانی در یک پنجرهٔ زمانی مشاهده شده است که کاربر C در یک قطعه و هر کدام از کاربران A و B در دو قطعه ارسال داشته‌اند).

در یک ساختار شبکه‌ای، لایهٔ پیوند داده اطلاعات مربوط به هر کاربر را جهت تشخیص و تصحیح خطا توسط یک کد کانال کدگذاری کرده و ارسال می‌نماید. بسته به نوع شبکه، خانوادهٔ مختلفی از کدها می‌تواند توسط کاربران استفاده شوند. در این مقاله فرض می‌شود که از خانوادهٔ کدهای BCH استفاده شده است. فرض می‌کنیم تعداد کاربران M است و منابع شبکه (قطعات زمانی) به‌صورت تقریباً یکنواخت بین کاربران تسهیم شده است. همچنین فرض می‌کنیم نرخ کد مورد استفاده توسط کاربران متفاوت و برابر $R_i, i = ۱,۲,...,M$ است. با توجه به طول کلمه‌های کد (همان بیت‌های ارسالی در هر قطعهٔ زمانی)، می‌توان گفت که کاربر i ام $k_i = R_i \times n$ بیت اطلاعات در هر قطعهٔ زمانی ارسال می‌کند. علت تفاوت در نرخ مورد استفاده کاربران مختلف، میزان دوری و یا نزدیکی کاربران به ایستگاه مرکزی می‌باشد (به طور کلی تفاوت در نرخ کاربران به کیفیت کانال ارتباطی بین آن‌ها بستگی دارد). علاوه بر این فرض می‌کنیم که داده ورودی به بلوک کدگذار کانال در فرستنده تصادفی باشد.



شکل (۲): ساختار شبکهٔ TDMA مبتنی بر کدگذاری کانال تطبیقی.

با توجه به توضیحات فوق، در قطعهٔ زمانی مورد استفاده توسط کاربر i ام، $n-k_i$ بیت اضافه بر اطلاعات اصلی ارسال می‌شود. در صورت متفاوت بودن کدهای مورد استفاده توسط کاربران مختلف، می‌توان با استفاده از این اطلاعات اضافی، تخمینی برای تعداد کاربران فعال شبکه به دست آورد. واضح است که اگر دو کاربر از یک کد یکسان استفاده کنند، به دلیل ماهیت یکسان این اطلاعات اضافه، از دیدگاه الگوریتمی که صرفاً از این اطلاعات اضافی استفاده می‌کند، این دو کاربر یک کاربر دیده می‌شوند. به عبارت دیگر، هر الگوریتمی که فقط از این اطلاعات اضافی استفاده نماید و اطلاعات لایه‌های بالاتر و پایین‌تر را مورد توجه قرار ندهد، دو کاربر با کد یکسان را به عنوان یک کاربر می‌بینند. به همین دلیل، چنین الگوریتمی یک حد پایین برای تعداد کاربران فعال در شبکه تخمین می‌زند. بنابراین، با استفاده از اطلاعات لایه فیزیکی (حاصل از تحلیل شکل موج دریافتی) و همچنین اطلاعات لایه‌های بالاتر می‌توان این تخمین را بهبود داد. یکی از سامانه‌هایی که از کدهای متفاوت برای کاربران مختلف در ساختار TDMA استفاده می‌کند، سامانه سلولی GSM است. در این سامانه، در کانال

1- Uplink
2- Time Slot

منطقی TCH/AFS با توجه به کیفیت کانال هر کاربر، کد کانال مورد استفاده توسط هر کاربر می‌تواند متفاوت باشد [۱۳].

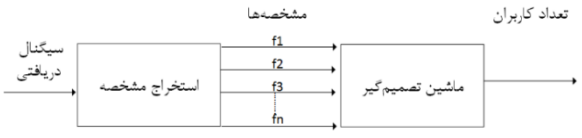
یک فرض ضمنی و اساسی در این مدل این است که عملیات تشخیص نقاط شروع و پایان قطعات زمانی، تشخیص مدوله‌سازی و همچنین دمدولاتوریی به صورت موفقیت‌آمیز توسط تحلیل‌گر سیگنال در لایه فیزیکی انجام شده است. البته همانند یک گیرندهٔ مجاز، خروجی چنین تحلیل‌گری می‌تواند دارای خطا باشد. در این مقاله فرض می‌کنیم که از دمدولاتوریی[1] سخت‌تصمیم استفاده شده و لذا می‌توان یک مدل سادهٔ BSC[2] را برای انتقال داده در نظر گرفت.

همچنین فرض می‌کنیم که پنجره مشاهده جهت تصمیم‌گیری، پنجرهٔ نسبتاً کوچکی است به طوری که کد مورد استفاده توسط کاربران در زمان مشاهدهٔ این پنجره تغییر نکرده است.

۳- روش پیشنهادی

برای حل این مسئله دو رویکرد مختلف را می‌توان در نظر گرفت. یک رویکرد مدل‌سازی و حل دقیق ریاضی این مسئله است. رویکرد دیگر استخراج مشخصه[3] و استفاده از روش‌های هوشمند یادگیری ماشین برای حل این مسئله می‌باشد. این روش‌ها در مسائل پیچیدهٔ ریاضی با ساده‌سازی به حل مسئله می‌پردازند و هرچند جواب دقیقی را به دست نمی‌آورند، اما در بسیاری از موارد جوابی قابل قبول در اختیار ما قرار می‌دهند.

با توجه به پیچیدگی مسئلهٔ تخمین تعداد کاربران، در این مقاله از روش‌های یادگیری ماشین تحت نظارت برای حل آن استفاده می‌کنیم. در شکل (۳) روش کلی کار پیشنهاد شده است. همان‌طور که در این شکل مشاهده می‌شود ابتدا مشخصه‌هایی از سیگنال دریافتی استخراج می‌شوند. سپس، ماشین تصمیم‌گیر (مبتنی بر روش‌های مرسومی مانند شبکه عصبی یا درخت تصمیم‌گیر) با استفاده از این مشخصه‌ها تعداد کاربران شبکه را تخمین می‌زنند. برای آموزش این ماشین تصمیم‌گیر از روش‌های یادگیری تحت نظارت استفاده می‌کنیم.



شکل (۳): روش کلی تخمین تعداد کاربران فعال شبکه.

1- Demodulation
2- Binary Symmetric Channel
3- Feature Extraction

۳-۱- الگوریتم پیشنهادی برای استخراج مشخصه

در ارتباط‌های یک‌سویه که از یک کد BCH برای کدگذاری کانال انتقال داده استفاده شده است، وانگ و همکارانش [۱۰] روشی برای شناسایی این کدها پیشنهاد داده‌اند. در این روش، ابتدا طول کلمه‌های کد و سپس چندجمله‌ای سازنده شناسایی می‌شود. بعد از شناسایی طول کلمه‌های کد، برای شناسایی چندجمله‌ای سازنده مراحل زیر طی می‌شود. فرض کنید با مشاهدهٔ یک پنجرهٔ زمانی، مجموعهٔ کلمه‌های کد $\{C_1, C_2, C_3, ..., C_N\}$ را در اختیار داریم که ممکن است برخی بیت‌های آنها دچار خطا شده باشد. برای هر یک از این کلمه‌های کد، همهٔ انتقال‌های چرخشی[4] را به دست می‌آوریم: $C_j^{(1)}, C_j^{(2)}, C_j^{(3)}, ..., C_j^{(n-1)};\ j = 1, 2, ..., N$. اگر شمارندهٔ مشترک[5] همهٔ انتقال‌های چرخشی هر یک از این کلمه‌های کد مخالف یک باشد (یعنی $f_j(x) = \gcd[C_j^{(1)}, C_j^{(2)}, C_j^{(3)}, ..., C_j^{(n-1)}] \neq 1$)، آن را به عنوان یک چندجمله‌ای سازندهٔ کاندید در نظر می‌گیریم. همچنین با توجه به اینکه همهٔ کلمه‌های کد با وزن زوج دارای عامل اضافی $x+1$ هستند، چنین عاملی را در محاسبهٔ چندجمله‌ای‌های سازندهٔ کاندید (در صورت وجود) حذف می‌کنیم.

فرض می‌کنیم تعداد چندجمله‌ای‌های کاندید متمایز برابر L ($L \leq N$) باشد. این L چندجمله‌ای را به صورت $g_i(x);\ i = 1, 2, ..., L$ و دوگان[6] آنها را به صورت $h_i(x);\ i = 1, 2, ..., L$ نمایش می‌دهیم (می‌دانیم که $g_i(x) h_i(x) = 1 + x^n$). حال هر یک از چندجمله‌ای‌های دوگان $h_i(x)$ را در تک‌تک کلمه‌های کد $\{C_1, C_2, C_3, ..., C_N\}$ ضرب می‌کنیم. سندرم[7] حاصل از این عملیات ضرب را به صورت $r_{ij}(x) = [(h_i(x) C_j(x)] \bmod (x^n + 1)$ تعریف کرده و وزن $\text{weight}(r_{ij}(x))$ را در متغیر w_{ij} ذخیره می‌کنیم (وزن یک چندجمله‌ای برابر تعداد ضرایب غیر صفر آن است). برای همهٔ چندجمله‌ای‌های سازندهٔ کاندید مجموع وزن سندرم‌ها را به صورت $w_i = \sum_{j=1}^{N} w_{ij}$ ($i = 1, 2, ..., L$) محاسبه می‌کنیم. در نهایت چندجمله‌ای سازندهٔ کاندیدی که مجموع وزن سندرم‌ها را کمینه نماید به عنوان چندجمله‌ای سازندهٔ کد معرفی می‌شود:

$$g(x) = \underset{g_i(x), i=1,2,...,L}{\arg\min} (w_i).$$

4- Cyclic Shift
5- Greatest Common Divisor
6- Dual
7- Syndrome

این روش و روش‌های مشابه برای شناسایی کدهای BCH ارتباط را یک‌سویه و کد مورد استفاده را یک کد مشخص و ثابت فرض می‌کنند. اما در مدل پیشنهادی شکل (۲)، چندین کد BCH توسط کاربران مختلف استفاده شده است. لذا با اعمال الگوریتم وانگ بر روی دادهٔ دریافتی از چنین سامانه‌ای، کماکان فقط یک چندجمله‌ای سازنده شناسایی می‌شود. این چندجمله‌ای سازنده در حقیقت متناظر با کد BCH دارای بیشترین بُعد است و هیچ‌گونه تشخیصی در تعداد کاربران شبکه نخواهیم داشت. علت این موضوع آن است که در کدهای BCH تولیدشده بر روی یک میدان گالوای مشخص، تمام کدها زیرکدی[1] از کد BCH دارای بیشترین بُعد (یا بالاترین نرخ کد) هستند. چندجمله‌ای سازندهٔ مربوط به کد BCH با بیشترین بُعد تمامی چندجمله‌ای‌های سازنده مربوط به کدهای BCH با ابعاد کمتر را عاد می‌کند. برای مثال اگر سه کاربر با استفاده از سه نوع کد BCH مختلف دارای طول کلمه‌های کد ۲۵۵ و طول پیام‌های ۱۱۵، ۱۰۷ و ۹۹ بسته‌های خود در شبکه را ارسال کنند، الگوریتم وانگ چندجمله‌ای سازندهٔ مربوط به کد BCH با طول پیام ۱۱۵ را (که دارای بیشترین بُعد است) به عنوان چندجمله‌ای سازندهٔ تمامی کلمه‌های کد معرفی می‌کند و هیچ‌گونه تشخیصی بر ۳ کاربره بودن شبکه فوق نخواهد داشت.

پیشنهاد ما برای تخمین تعداد کاربران فعال شبکه استفاده از مقادیر مجموع وزن سندرم‌ها (مقادیر w_i) می‌باشد. قبل از ارائهٔ روش، ابتدا با ارائهٔ چند مثال ارتباط بین تعداد کاربران و w_iها را نشان می‌دهیم. با در نظر گرفتن مدل شکل (۲)، فرض کنید در یک شبکهٔ تک‌کاربرهٔ بدون نویز، کد BCH با طول کلمه‌های کد $n=۱۲۷$ و طول پیام $k=۷۱$ استفاده شده است. در یک مرتبه شبیه‌سازی، $N=۱۰۰$ کلمهٔ کد را در یک پنجرهٔ زمانی تولید کرده‌ایم و در این شرایط، تعداد چندجمله‌ای‌های سازندهٔ کاندید $L=۸$ شده است (با توجه به طول ۷۱ بیتی پیام‌ها، این تعداد حداکثر می‌توانست ۸ باشد که در این شبیه‌سازی تصادفی چنین چیزی رخ داده است). علت اینکه تعداد چندجمله‌ای‌های سازندهٔ کاندید بیش از یک است آن است که بعضی از کلمه‌های کد دارای عامل‌هایی اضافی هستند که باعث می‌شود به عنوان کلمهٔ کدی از یک کد با نرخ پایین‌تر هم دیده شوند. چندجمله‌ای سازندهٔ کاندیدی که به عنوان چندجمله‌ای سازندهٔ اصلی تشخیص داده می‌شود تمامی کلمه‌های کد دریافتی را عاد می‌کند و تمامی سندرم‌های حاصل از چندجمله‌ای دوگان آن برابر صفر خواهد شد؛ لذا مجموع وزن سندرم‌های متناظر با آن نیز صفر می‌شود.

مجموعهٔ $L=۸$ چندجمله‌ای سازندهٔ کاندید اعداد مرتب‌شدهٔ $\mathbf{w}_۱ \approx (۰,۰/۴۹,۰/۵,۰/۵,۰/۵,۰/۵,۰/۵)$ را برای مجموع وزن سندرم‌های خود نشان می‌دهند. توجه کنید که برای تولید این اعداد از نرمال‌سازی (تقسیم بر $N\times n$) استفاده کرده‌ایم. $\mathbf{w}$ را بردار مجموع وزن سندرم نرمال‌شده می‌نامیم.

حال فرض می‌کنیم در شبکه ۲ کاربر حضور دارند. همچنین کدهای BCH متمایز با طول کلمه‌های کد $n=۱۲۷$ و طول پیام‌های $k_۱=۷۱$ و $k_۲=۷۸$ توسط این دو کاربر استفاده می‌شود. تعداد کلمه‌های کد شبیه‌سازی‌شده را مجدداً برابر $N=۱۰۰$ می‌گیریم. سهم هر یک از کاربران را نیز ۵۰ کلمهٔ کد (یا قطعهٔ زمانی) در نظر می‌گیریم. در این شرایط نیز تعداد چندجمله‌ای‌های سازندهٔ کاندید برابر $L=۸$ می‌شود. چندجمله‌ای سازندهٔ کاندیدی که به عنوان چندجمله‌ای سازندهٔ اصلی تشخیص داده می‌شود مربوط به کلمه‌های کد با بُعد بیشتر (کد با طول پیام‌های ۷۸ بیتی) خواهد بود. تمامی سندرم‌های متناظر با این چندجمله‌ای سازنده برابر صفر خواهد شد و مجموع وزن این سندرم‌ها نیز صفر می‌شود. مجموعهٔ $L=۸$ چندجمله‌ای سازندهٔ کاندید اعداد مرتب‌شدهٔ $\mathbf{w}_۲ \approx (۰,۰/۲۵,۰/۴۹,۰/۴۹,۰/۵,۰/۵,۰/۵,۰/۵)$ را برای مجموع وزن سندرم‌های خود نشان می‌دهند (دقت کنید که مجدداً این اعداد را نرمال‌سازی و مرتب کرده‌ایم). عدد دوم در این بردار متناظر با چندجمله‌ای سازندهٔ کد دارای بُعد کمتر (کد با طول پیام‌های ۷۱ بیتی) است. در حقیقت چندجمله‌ای دوگان متناظر با این چندجمله‌ای سازنده، فقط بر نیمی از کلمه‌های کد دریافتی (یعنی کلمه‌های کد مربوط به کد با طول پیام‌های ۷۱ بیتی) عمود است.

در مثالی دیگر، فرض کنید در شبکه مجدداً ۲ کاربر حضور دارند و کدهای BCH متمایز با طول کلمه‌های کد $n=۱۲۷$ و طول پیام‌های $k_۱=۷۱$ و $k_۲=۸۵$ توسط این دو کاربر استفاده می‌شود. تعداد کلمه‌های کد شبیه‌سازی‌شده و سهم هر یک از کاربران را نیز مانند مثال قبل در نظر می‌گیریم. در یک بار شبیه‌سازی تصادفی تعداد چندجمله‌ای‌های سازندهٔ کاندید برابر $L=۶$ می‌شود. مانند قبل، چندجمله‌ای سازندهٔ کاندیدی که به عنوان چندجمله‌ای سازندهٔ اصلی تشخیص داده می‌شود مربوط به کلمه‌های کد با بُعد بیشتر (کد با طول پیام‌های ۸۵ بیتی) خواهد بود. در این حالت، مجموعهٔ چندجمله‌ای سازندهٔ کاندید اعداد مرتب‌شدهٔ $\mathbf{w}_۳ \approx (۰,۰/۲۴,۰/۲۵,۰/۵,۰/۵,۰/۵)$ را برای مجموع وزن سندرم‌های خود نشان می‌دهند. در چنین شرایطی، دومین وزن مربوط به چندجمله‌ای سازندهٔ کد $(۱۲۷,۷۱)$ است. اما در مورد وجود وزن سوم کوچک (که به شکل معناداری کمتر از

1- Subcodeff

وزن‌های بعدی و تقریباً برابر وزن دوم است) باید گفت که به دلیل وجود عوامل چرخشی اضافی در کلمه‌های کد، چندجمله‌ای سازندهٔ کد $(۱۲۷,۷۸)$ نیز عضو مجموعهٔ چندجمله‌ای‌های کاندید است. توجه کنید که این کد در هیچ کدام از قطعات زمانی استفاده نشده است، با این حال چندجمله‌ای دوگان متناظر با چندجمله‌ای سازندهٔ آن برای تمامی کلمه‌های کد $(۱۲۷,۷۱)$ سندرمی با وزن صفر تولید می‌کند. دلیل این موضوع این است که کد $(۱۲۷,۷۱)$ زیرکدی از کد $(۱۲۷,۷۸)$ است.

واضح است که در سه مثال فوق، اگر دومین مقدار از بین درایه‌های بردار $\mathbf{w}$ در شبکهٔ تک‌کاربره را با مقدار مشابه آن در شبکهٔ ۲ کاربره مقایسه کنیم خواهیم دید این مقدار در شبکهٔ تک‌کاربره (مقداری در حدود ۰/۵۰) و در شبکهٔ دوکاربره (مقداری در حدود ۰/۲۵) تفاوت معناداری با یکدیگر دارند. با مقایسهٔ این دو عدد متوجه می‌شویم مقدار مورد نظر در شبکهٔ ۲ کاربره تقریباً نصف مقدار مورد نظر در شبکهٔ تک‌کاربره است. بنابراین، اگر دومین مقدار از بردار مجموع وزن سندرم نرمال‌شده را به عنوان یک ویژگی برای تشخیص شبکهٔ تک‌کاربره و ۲ کاربره در نظر بگیریم، موفق به تشخیص این دو شبکه از یکدیگر خواهیم شد.

زمانی که تعداد کاربران بیشتر می‌شود مقدار درایهٔ دوم بردار $\mathbf{w}$ کاهش می‌یابد (مقداری در حدود $1/(2U)$ برای زمانی که U کاربر با کدهای مختلف و به صورت تقریباً یکنواخت از شبکه استفاده می‌کنند). در حقیقت، در شرایط کانال بدون خطا، صرفاً با بررسی این مقدار می‌توان عملیات شناسایی تعداد کاربران را با درصد خطای کمی انجام داد. در این شرایط، با استفاده از یک ویژگی (دومین مقدار از بردار سندرم) می‌توانیم تعداد کاربران شبکه را با دقت بالایی (حدود ۹۹% برای ماشین تصمیم‌گیر نزدیک‌ترین همسایه) تشخیص دهیم.

اما در شرایط عملی که کانال انتقال داده دارای خطا است، شرایط کمی متفاوت خواهد شد. برای مثال، زمانی که ۲۰% کلمه‌های کد دریافتی دچار خطا باشند، با تنها استفاده از همین یک ویژگی دقت تشخیص به کمتر از ۲۰% کاهش خواهد یافت. در حقیقت، با وارد شدن خطا در کانال انتقال داده و استفاده از فقط یک مشخصه، توانایی تشخیص تعداد کاربران به میزان قابل توجهی کاهش می‌یابد. بنابراین نیاز است تا مشخصه‌های جدیدی به مشخصهٔ موجود اضافه شوند تا در شرایط عملی کانال دارای خطا نیز بهترین تشخیص ممکن صورت گیرد.

با اضافه کردن مقدار اول و سوم بردار $\mathbf{w}$ به مجموعهٔ مشخصه‌ها (که حالا حاوی سه مشخصه است)، در شرایطی که ۲۰% کلمه‌های کد دچار خطا شده باشند، متوجهٔ بهبود قابل‌توجهی خواهیم شد. در این شرایط دقت الگوریتم به بیش از ۵۰% می‌رسد (توجه داشته باشید که در شرایط نویزی، مقدار درایهٔ اول بردار مجموع وزن سندرم نرمال‌شده دیگر لزوماً برابر صفر نیست). با مشاهدهٔ این بهبود، این ایده به ذهن می‌رسد که شاید اگر چهارمین درایهٔ بردار $\mathbf{w}$ نیز به سه ویژگی قبلی اضافه شود دقت الگوریتم تشخیص بهبود یابد. با اضافه کردن این ویژگی دقت الگوریتم‌های تشخیص به بیش از ۶۵% افزایش می‌یابد. در حقیقت اگر در شبکه U کاربر فعال وجود داشته باشند، دست‌کم U مقدار اول در بردار $\mathbf{w}$ دارای فاصلهٔ معناداری از بقیهٔ مقادیر هستند. بنابراین، باید حداقل به تعداد کاربران، درایه‌های ابتدایی بردار $\mathbf{w}$ را وارد فرایند تصمیم‌گیری نماییم. در تفسیر ویژگی‌های استخراج‌شدهٔ فوق، نیاز به یک ویژگی مهم دیگر هم داریم. این ویژگی مهم همان میزان خطای کانال است. اضافه شدن این ویژگی قطعاً برای تفسیر بهتر مقادیر بردار مجموع وزن سندرم نرمال‌شده و عملکرد بهتر الگوریتم مفید است.

برای اینکه دقیق‌تر صحبت کنیم، باید گفت در صورتی که طول کد برابر n بوده و تعداد کدهای BCH ممکن با این طول برابر $T=T(n)$ باشد، تا حداکثر T چندجمله‌ای کاندید خواهیم داشت (به عنوان مثال داریم $T(۱۲۷)=۱۸$). لذا بردار مجموع وزن سندرم نرمال‌شده حداکثر دارای طول T خواهد بود. در چنین شرایطی حداکثر $T+۱$ مشخصه (شامل T مجموع وزن سندرم نرمال‌شده و ۱ مقدار احتمال خطا) قابل تولید است. توجه کنید که اگر U تعداد کاربران دارای کد BCH متفاوت باشد، لزوماً داریم $U \le T(n)$. اگر F درایهٔ اول از بردار مجموع وزن سندرم وزن‌یافته را به عنوان مشخصه در نظر بگیریم ($M \le F \le T(n)$ که M حداکثر تعداد محتمل کاربران حاضر است)، در شرایطی که تعداد چندجمله‌ای‌های سازندهٔ کاندید کمتر از F باشد، باید به جای سایر مقادیر مشخصه که توسط الگوریتم تولید نمی‌شوند یک مقدار مشخص (و البته غیر محتمل) قرار داد. به عنوان مثال یک گزینه می‌تواند قرار دادن مقدار ثابت ۱ باشد.

با توجه به مباحث فوق، الگوریتم استخراج مشخصه را می‌توانیم به این شکل بیان کنیم:

ورودی: N قطعهٔ زمانی (هر قطعه حاوی یک کلمهٔ کد نویزی)، تعداد مشخصه‌های خروجی ($F+۱$) مورد نیاز و احتمال خطای کانال انتقال داده.

تنظیم مقادیر اولیه: $t=۰$

۱- **حلقهٔ FOR**: برای j از ۱ تا N

۲- j امین کلمه کد را انتخاب و تمامی چندجمله‌ای‌های حاصل از انتقال‌های چرخشی آن را محاسبه می‌کنیم.

۳- شمارندهٔ مشترک همهٔ این انتقال‌های چرخشی را محاسبه می‌کنیم. عامل $x+۱$ را در صورت وجود با استفاده از تقسیم چندجمله‌ای حذف می‌کنیم.

۴- در صورت غیر یک و غیرتکراری بودن شمارندهٔ مشترک، آن را در $g_t(x)$ ذخیره می‌کنیم و همچنین مقدار t را یک واحد افزایش می‌دهیم ($t=t+۱$).

۵- **اتمام حلقه FOR**

۶- حلقهٔ **FOR**: برای k از ۱ تا t

۷- چندجمله‌ای دوگان متناظر با $g_k(x)$ را محاسبه و در $h_k(x)$ ذخیره می‌کنیم.

۸- $h_k(x)$ را در تمامی کلمه‌های کد ضرب می‌کنیم و سندرم‌های به‌دست‌آمده برای هر کلمهٔ کد را در r_{ki} ذخیره می‌کنیم (i از ۱ تا N). وزن سندرم‌های به‌دست‌آمده را محاسبه کرده و در w_{ki} ذخیره می‌کنیم.

۹- برای چندجمله‌ای سازندهٔ کاندید $g_k(x)$ مجموع وزن سندرم $\sum_{i=۱}^{N} w_{ki}$ ($k=۱,۲,...,L$) را محاسبه و در w_k ذخیره می‌کنیم.

۱۰- w_k را در درایهٔ k ام بردار **w** ذخیره می‌کنیم.

۱۱- **اتمام حلقه FOR**

۱۲- **نرمال‌سازی بردار w**: تمامی عناصر بردار را بر مقدار $N \times n$ تقسیم می‌نماییم.

۱۳- **مرتب‌سازی بردار w**: تمامی عناصر به ترتیب کوچک به بزرگ قرار می‌گیرند.

۱۴- **تکمیل بردار w**: اگر بردار **w** دارای کمتر از F عنصر بود، با افزودن درایه‌های ثابت ۱ به انتهای آن، طول آن را به F می‌رسانیم.

۱۵- **خروجی**: F مقدار اول بردار **w** را به همراه خطای کانال به عنوان بردار مشخصه تحویل می‌دهیم.

۳-۲- ماشین تصمیم‌گیر

برای شناسایی تعداد کاربران با استفاده از مشخصه‌های ایجادشده از روش‌های یادگیری ماشین مرسوم استفاده می‌شود. یک شبکهٔ ارتباطی را مانند آنچه در شکل (۲) نمایش داده شده است، در نظر بگیرید. با مشاهدهٔ یک پنجرهٔ زمانی از سیگنال دریافتی از این شبکه، N قطعهٔ زمانی در اختیار خواهیم داشت. با اجرای الگوریتم استخراج مشخصه بر روی این قطعات زمانی به یک بردار مشخصه به طول $F+۱$ می‌رسیم. این بردار مشخصه به همراه تعداد واقعی کاربران فعال در شبکه (که دارای کد BCH متفاوتی نسبت به بقیه هستند) را یک نمونه[1] می‌نامیم که دارای تعدادی مشخصه و یک برچسب (همان تعداد کاربران) است.

در این مقاله از سه ماشین تصمیم‌گیری نزدیک‌ترین همسایه، درخت رگرسیون و شبکه عصبی استفاده شده است. برای آموزش این ماشین‌های تصمیم‌گیری، باید در طی یک فرایند آموزشی تحت نظارت[2]، تعدادی نمونه با برچسب‌های معلوم را به ماشین داد تا ماشین الگوها را فراگیری نماید. بدیهی است که برای آموزش هر چه بهتر، باید تا حد ممکن توزیعی یکنواخت از نمونه‌های محتمل را به ماشین داد. پس از اتمام آموزش نیز می‌توان در فرایند آزمون ماشین، تعدادی نمونه با برچسب‌های نامعلوم را به ماشین داد تا در مورد آنها تصمیم‌گیری نماید. با بررسی تصمیم‌های اتخاذشده توسط ماشین و مقایسهٔ آنها با برچسب‌های واقعی می‌توان پی به دقت ماشین برد.

الگوریتم نزدیک‌ترین همسایه ساده‌ترین روش برای تخمین مورد نظر می‌باشد. این الگوریتم برای تخمین تعداد کاربران، با انتخاب K نزدیک‌ترین همسایهٔ هر نمونه (از میان همهٔ نمونه‌های آموزشی) و انتخاب برچسب دارای بیشترین تکرار در بین این همسایه‌ها، تصمیم‌گیری را انجام می‌دهد. برای انتخاب نزدیک‌ترین همسایه‌ها از معیار فاصلهٔ اقلیدسی استفاده می‌کنیم. معمولاً دقت این الگوریتم با افزایش تعداد مشخصه‌ها کاهش می‌یابد.

برای تولید درخت رگرسیون، از الگوریتمی بازگشتی با رفتار حریصانه (با معیار کاهش آنتروپی) استفاده می‌شود. در هر گره تصمیم‌گیری چنین درختی، تنها از یک ویژگی استفاده می‌شود. بعد از تولید کامل درخت، درخت باید طوری هرس شود که بهترین عملکرد را بر روی دادهٔ آزمون داشته باشد.

برای ساخت شبکه عصبی، یک شبکه عصبی پیش‌خور[3] دولایه‌ای استفاده می‌شود. تابع فعال‌سازی در لایهٔ نخست سیگموید و در لایهٔ خروجی خطی می‌باشد. همچنین برای کدگذاری خروجی شبکهٔ عصبی از کدهای یگانی[4] استفاده

1- Sample
2- Supervised Learning
3- Feedforward
4- Unary Coding

می‌شود. برای سنجش عملکرد در تمامی الگوریتم‌های فوق از عملیات اعتبارسنجی متقابل[1] استفاده می‌شود.

۴- نتایج شبیه‌سازی

در این بخش با استفاده از شبیه‌سازی، کارایی الگوریتم پیشنهادی را مورد بررسی قرار می‌دهیم. برای انجام شبیه‌سازی، از کدهای چرخشی BCH با طول ۲۵۵ استفاده شده است. همچنین طول پنجرهٔ مشاهده‌شده از قطعات زمانی برای هر نمونهٔ آزمایش معادل ۶۴۰ قطعهٔ زمانی (حدود ۲۰ کیلوبایت بیت کدشده) است. فرض می‌کنیم تعداد کاربران فعال حداکثر برابر ۱۰ می‌باشد ($M=10$). همچنین برای ارزیابی روش پیشنهادی خطای بیت را از ۰ تا ۰/۰۰۳۵ تغییر می‌دهیم تا کلمه‌های کد با احتمال‌هایی بین ۰ تا ۰/۹ (با فواصلی به اندازهٔ ۰/۰۵) دچار خطا شوند. با فرض استفاده از مدولاسیون BPSK[2]، این محدودهٔ خطا متناظر با تغییر نسبت سیگنال به نویز به ازای هر بیت (E_b/N_0) بین بی‌نهایت تا حدود ۵/۶ دسی‌بل است. برای هر مقدار ممکن از خطای بیت، ۲۰۰۰ نمونه شبیه‌سازی می‌شود که در هر نمونه به تعداد ۱ تا ۱۰ کاربر (با توزیع احتمال یکنواخت) فعال خواهند بود. همچنین در هر نمونه کدهای متمایز برای کاربران مختلف به صورت تصادفی از بین تمامی کدهای BCH ممکن به طول ۲۵۵ انتخاب می‌شود. بنابراین ۲۰۰ نمونه برای حالت تک‌کاربره، ۲۰۰ نمونه برای حالت دوکاربره و به همین ترتیب ۲۰۰ نمونه برای حالت ۱۰ کاربره شبیه‌سازی می‌شود. با توجه به اینکه الگوریتم پیشنهادی از ورودی‌های تصمیم-سخت استفاده می‌کند، نیازی به شبیه‌سازی سایر مدولاسیون‌های دیجیتال نیست. در حقیقت، در اینجا تنها مقدار خطای بیت در ورودی الگوریتم تعیین‌کنندهٔ نحوهٔ عملکرد الگوریتم است.

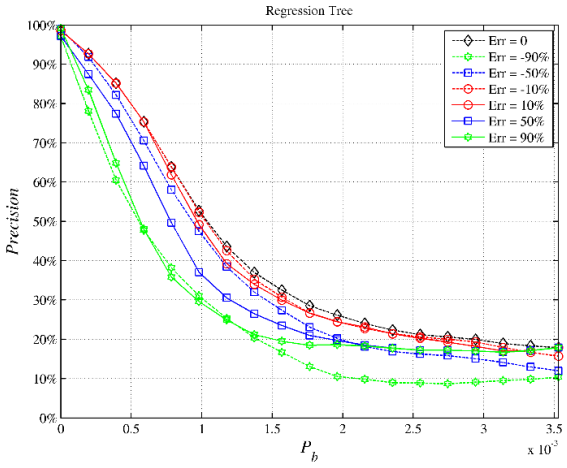
تعداد تکرار در عملیات اعتبارسنجی متقابل را نیز برابر ۱۰ می‌گیریم. در الگوریتم نزدیک‌ترین همسایه $K=10$ می‌باشد. در الگوریتم درخت حداقل تعداد نمونه لازم در برگ برای پایین رفتن شاخه‌های درخت ۱۰ می‌باشد. همچنین در الگوریتم شبکه عصبی، تعداد نورون‌های لایه مخفی برابر ۲۰ تنظیم شده است.

با اینکه ماشین تصمیم‌گیر فرض کرده است که خطای بیت در کانال (مقدار p_b) مشخص است، اما در عمل این‌گونه نیست. خطای کانال توسط بخش دمدولاتور تخمین زده می‌شود. این تخمین (که آن را با $\hat{p}_b$ نشان می‌دهیم) قطعاً دقیق نیست. لذا برای اینکه نتایج شبیه‌سازی تا حد ممکن به واقعیت نزدیک شود، فرض می‌کنیم خطایی به نام عدم تطابق تخمین احتمال خطا وجود دارد که برابر $\mathrm{Err}=\frac{\hat{p}_b-p_b}{p_b}\times 100\%$ است. همان‌طور که در شکل (۴) تا شکل (۶) مشاهده می‌شود، اثر این خطا بر روی سه ماشین تصمیم‌گیر درخت رگرسیون، نزدیک‌ترین همسایه و شبکه عصبی مشاهده می‌شود. خطای منفی به معنی تخمینی کمتر از خطای بیت واقعی و خطای مثبت به معنی تخمینی بیشتر از خطای بیت واقعی می‌باشد. همان‌طور که در این سه شکل مشاهده می‌شود، پایداری درخت رگرسیون در مقابل عدم تطابق تخمین احتمال خطا به نسبت دو ماشین دیگر بیشتر است. همچنین می‌توان دید که در مقادیر مختلف احتمال خطای کانال، شبکه عصبی نسبت به دو ماشین دیگر کارایی کمتری از خود نشان می‌دهد.

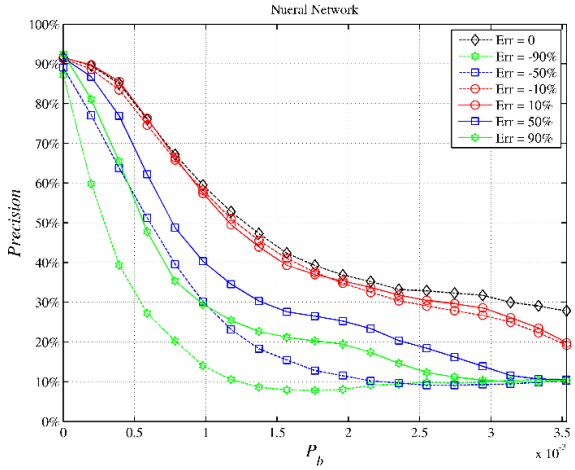
با در نظر گرفتن مقدار صفر برای عدم تطابق تخمین احتمال خطا، در شکل (۷) دقت هر سه الگوریتم در کنار یکدیگر ترسیم شده است. همان‌طور که در این شکل مشاهده می‌شود در خطای بیت حدود $p_b=0/001$ هر سه ماشین بین ۵۰ تا ۶۰ درصد مواقع، تعداد کاربران فعال در شبکه را به درستی تخمین می‌زنند. در خطای بیت حدود $p_b=0/0005$ نیز ماشین‌ها با دقت بیش از ۸۰٪ تشخیص صحیحی از تعداد کاربران را دارند. روش تصمیم‌گیری نزدیک‌ترین همسایه در خطای بیت کم و الگوریتم شبکه عصبی در خطای بیت بالا بهترین عملکرد را دارند. اینکه دقیقاً از کدام الگوریتم برای تشخیص تعداد کاربران فعال شبکه استفاده کنیم بستگی به میزان دقت مورد نیاز و همچنین پایداری مد نظر برای ماشین تصمیم‌گیر دارد. بدیهی است اگر طول کلمه‌های کد بیشتر شود، نسبت سیگنال به نویز مورد نیاز برای رسیدن به یک دقت مشخص نیز افزایش می‌یابد [۱۴]. همچنین توجه داشته باشید که میزان نسبت سیگنال به نویز مورد نیازِ شناسایی کور قاعدتاً بیشتر از میزان نسبت سیگنال به نویز برای کدگشایی صحیح می‌باشد.

در شکل (۸) ماتریس پراکندگی دقت تخمین به تفکیک تعداد کاربران در خطای بیت ۰/۰۰۰۴ نمایش داده شده است. هر سطر از این ماتریس متناظر با یکی از حالت‌های موجود برای تعداد کاربران فعال است. همچنین هر ستون نیز متناظر با یکی از تخمین‌های ممکن برای تعداد کاربران است. برای همان‌طور که ملاحظه می‌شود، دقت تخمین تعداد کاربران برای حالتی که یک یا دو کاربر فعال داریم، ۱۰۰٪ است و هر چه تعداد کاربران بیشتر می‌شود، این دقت می‌یابد.

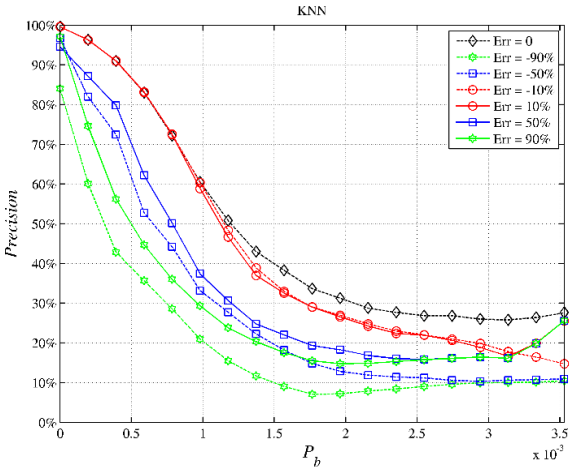
1- Cross Validation
2- Binary Phase Shift Key



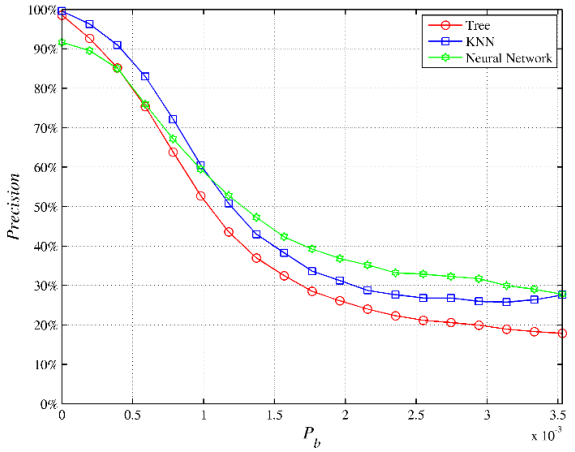
شکل (۴): اثر خطای تخمین بر روی دقت درخت رگرسیون



شکل (۶): اثر خطای تخمین بر روی دقت شبکه عصبی



شکل (۵): اثر خطای تخمین بر روی دقت نزدیک‌ترین همسایه



شکل (۷): مقایسهٔ دقت درخت رگرسیون، نزدیک‌ترین همسایه و شبکه عصبی

تعداد تخمینی کاربران / تعداد واقعی کاربران	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
۱	۱۰۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
۲	۰	۱۰۰	۰	۰	۰	۰	۰	۰	۰	۰
۳	۰	۰	۹۹/۳	۶/۷	۰	۰	۰	۰	۰	۰
۴	۰	۰	۰	۹۲/۹	۷/۱	۰	۰	۰	۰	۰
۵	۰	۰	۰	۵/۶	۷۷/۸	۱۱/۱	۵/۵	۰	۰	۰
۶	۰	۰	۰	۰	۰	۸۰/۶	۱۹/۴	۰	۰	۰
۷	۰	۰	۰	۰	۰	۱۹/۰	۵۷/۲	۲۳/۸	۰	۰
۸	۰	۰	۰	۰	۰	۰	۳۳/۳	۴۷/۶	۱۹/۱	۰
۹	۰	۰	۰	۰	۰	۰	۴/۵	۹/۱	۷۷/۳	۹/۱
۱۰	۰	۰	۰	۰	۰	۰	۰	۷/۷	۷/۷	۸۴/۶

شکل (۸): ماتریس پراکندگی دقت تخمین به تفکیک تعداد کاربران در خطای بیت ۰/۰۰۰۴

۵- نتیجه‌گیری

در این مقاله روشی برای تخمین تعداد کاربران فعال در شبکه‌های مبتنی بر دسترسی چندگانهٔ تسهیم زمانی ارائه شد. در روش ارائه‌شده فرض کردیم که همهٔ کاربران از کدهای BCH برای کدگذاری استفاده می‌کنند. بدیهی است که می‌توان این فرض را در تحقیقات بعدی مورد بازبینی و توسعه قرار داد. در روش پیشنهادی، با فرض استفاده از مدولاسیون BPSK، برای رسیدن به دقت بالای ۸۰% حداقل سیگنال به نویز موردنیاز حدود ۷.۳ دسی‌بل است. این نسبت سیگنال به نویز، برای کدهای BCH با طول ۲۵۵ لازم است. بدیهی است اگر طول کدها بیشتر شود، نسبت سیگنال به نویز مورد نیاز نیز افزایش می‌یابد. روش پیشنهادی نشان داد که استفاده از یادگیری ماشین در حل مسائل پیچیدهٔ حوزه شناسایی کور شبکه‌های مخابراتی می‌تواند بسیار مفید باشد.

یکی از مسائل مهم جهت ارزیابی روش پیشنهادی، مقایسه آن با تحقیقات پیشین در زمینه تخمین تعداد کاربران فعال است. از آنجا که در حوزه تحلیل بیت، مقاله پیش‌رو اولین تحقیق منتشرشده از این نوع است، باید تحقیقات انجام‌شده در حوزه سیگنال جهت مقایسه مد نظر قرار گیرد. این موضوع نیاز به یک چارچوب دقیق و مفصل مدل‌سازی و شبیه‌سازی دارد که به عنوان یک کار در آینده پیشنهاد می‌شود. همچنین ترکیب روش‌های لایه سیگنال با روش پیشنهادی از دیگر زمینه‌های تحقیقاتی است که در آینده می‌تواند مورد توجه قرار گیرد.

۶- مراجع

[1] M. Teimouri and H. Kakaei Motlagh, "Reverse Engineering of Communications Networks: Evolution and Challenges," arXiv preprint arXiv:1704.05432, 2017.

[2] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," IET Commun., vol. 1, no. 2, pp. 137-156, 2007.

[3] Q. An, W. Xia, Z. S. He, and H. Y. Li, "Algorithm for modulation classification of PSK signals," in Proc. 19th International Conference on Digital Signal Processing (DSP), pp. 54-58, Kowloon, Hong Kong, 20-23 Aug. 2014.

[4] R. Imad, G. Sicot, and S. Houcke, "Blind frame synchronization for error correcting codes having a sparse parity check matrix," IEEE Trans. Commun., vol. 57, no. 6, pp. 1574-1577, 2009.

[5] H. Wu, Z. Sha, Z. Huang, and Y. Zhou, "Frame synchronization for DVB-S2 based on scrambling sequence," in Proc. International Symposium on Wireless Personal Multimedia Communications (WPMC), pp. 103-105, Sydney, Australia, 7-10 Sep. 2014.

[6] J. P. Tillich, A. Tixier, and N. Sendrier, "Recovering the interleaver of an unknown turbo-code," in Proc. IEEE International Symposium on Information Theory (ISIT), pp. 2784-2788, Honolulu, HI, 29 June - 4 July 2014.

[7] H. Ryu, J. Lee, H. Hong, and D. Yoon, "Estimation of interleaver period for unknown signals," in Proc. 2nd IEEE International Conference on Network Infrastructure and Digital Content, pp. 678-680, Beijing, China, 24-26 Sep. 2010.

[8] X. W. Wu, S. N. Koh, and C. C. Chui, "Primitive polynomials for robust scramblers and stream ciphers against reverse engineering," in Proc. IEEE International Symposium on Information Theory Proceedings (ISIT), pp. 2473-2477, Austin, TX, 13-18 June 2010.

[9] X. B. Liu, S. N. Koh, X. W. Wu, and C.-C. Chui, "Investigation on scrambler reconstruction with minimum a priori knowledge," in Proc. IEEE Global Telecommunications Conference, Houston, TX, pp. 1-5, 5-9 Dec. 2011.

[10] J. Wang, Y. Yue, and J. Yao, "Statistical recognition method of binary BCH code," Communications and Network, vol. 3, no. 1, pp. 17-22, 2011.

[11] W. Li, J. Lei, L. Wen, and B. Chen, "An improved method of blind recognition of RS code based on matrix transformation," in Proc. 15th IEEE International Conference on Communication Technology (ICCT), pp. 196-200, Guilin, China, 17-19 Nov. 2013.

[12] F. Samsami-Khodadad and F. Nazari, "A Novel and Efficient DS/CDMA Direction of Arrival Algorithm for Multipath Fading Channel," Journal of Electronical & Cyber Defence, vol. 4, no. 3, pp. 27-36, 2016. (in persian)

[13] ETSI, GSM, "Technical Specification 05.03," Digital cellular communications system (GSM Radio Access Phase 3), Channel coding, Version 6.0, 1998.

[14] A. Valembois, "Detection and recognition of a binary linear code." Discrete Appl. Math., vol. 111, no. 1, pp. 199-218, 2001.

Journal of Electronical & Cyber Defence
Vol. 6, No. 2, 2018, Serial No. 22

Blind Estimation of the Number of Users in TDMA Networks Using the Redundancy of Adaptive Channel Coding

S. M. Ahmadiyan, M. Teimouri*

*Faculty of New Sciences and Technologies, University of Tehran, Tehran

(Received: 15/10/2016, Accepted: 02/11/2017)

ABSTRACT

A complete reverse engineering (or blind identification) in an electronic battlefield determines the information conveyed by a received signal. Most of the research in the field of blind signal identification is around one-way and non-network communications in which the goal is to determine the information transmitted by a single transmitter. The first step of signal identification in communications networks is to determine the number of active users. In this paper, estimation of the number of users in a time-division multiple access (TDMA) network is considered. In order to estimate the number of users, a physical layer analysis can be applied to the received electromagnetic signals. However, due to some difficulties such as hardware limitations or closeness of active users, this method cannot always be employed. In these situations, a solution is to analyze the information in the upper layers of the network. In this paper, a method is proposed to estimate the number of active users using the redundant data generated by adaptive channel coding. Simulation results show that the proposed method is quite resistant against channel errors. In fact, the accuracy of the proposed method for signal to noise ratio of 7.3 dB is around 80%.

Keywords: Blind Estimation Of Number Of Users, TDMA Networks, Adaptive Channel Coding, Machine Learning

* Corresponding Author Email: mehditeimouri@ut.ac.ir