Journal Of Electronical & Cyber Defence
Vol. 2, No. 1, 2014, Serial No.5

Detection of Covert Timing Channels Based on Statistical Methods

B. Bayrami[*1], M. Dehghani[2], M. Saleh Esfahani[3]
1,2,3- Imam Hossein Comprehensive University
(Receive: 2013/12/02, Accept: 2014/05/26)

Abstract

Covert timing channels, as a growing threat of network security, provide the possibility of leaking confidential information to an attacker. Thus, detection and countering these channels are important as a defensive measure in computer networks. The attacker uses an appropriate encoding schema to modulate covert information on temporal features of network packet stream. Embedding covert information in network traffic, casuses changes in traffic statistical properties such as distribution, correlation and entropy, which can be used in detection of covert timing channels. In this paper, statistical methods are identified and analyzed to detect these channels, and two encoding schemas L-bit to N-packet and non-detectable are used to implementi covert timing channels and those are evaluated. The results show that by using Kolmogorov-Smirnov, Regularity, corrected Entropy and corrected Conditional Entropy tests, we are able to completely detect L-bit to N-packet channel, and the stealthiness of non-detectable timing channel can be proved in a practical evaluation as well.

Keywords:
Covert Timing Channel, Detection, Entropy, Regularity, Kolmogorov-Smirnov

***Corresponding Author Email:** beyrami60@yahoo.com

مجله علمی- پژوهشی "پدافند الکترونیکی و سایبری"

سال دوم، شماره ۵، بهار ۱۳۹۳، ص۲۴-۱۳

تشخیص کانال‌های زمان‌بندی‌دار پوششی به روش‌های آماری

بهمن بیرامی*[۱]، مهدی دهقانی[۲]، محمود صالح اصفهانی[۳]

۱- کارشناس ارشد فناوری اطلاعات گرایش امنیت ، معاونت فاوای نیروی زمینی ۲- دانشجوی دکتری کامپیوتر گرایش نرم افزار، دانشگاه امام حسین(ع)

۳- دکتری کامپیوتر گرایش نرم افزار، دانشگاه امام حسین(ع)

(دریافت: ۹۲/۰۹/۱۱ ، پذیرش: ۹۳/۰۳/۰۵)

چکیده

کانال‌های زمانبندی‌دار پوششی به عنوان یک تهدید رو به رشد در امنیت شبکه، امکان نشت اطلاعات محرمانه را برای یک مهاجم فراهم می‌سازند. از این‌رو تشخیص و مقابله با این‌گونه کانال‌ها به عنوان یک اقدام پدافندی در شبکه‌های رایانه‌ای از اهمیت ویژه‌ای برخوردار است. مهاجم برای ایجاد کانال زمانبندی‌دار پوششی، با استفاده از یک روش کدگذاری مناسب، اطلاعات پوششی را روی یکی از ویژگی‌های زمانی جریان بسته‌های شبکه سوار می‌کند. با تعبیه اطلاعات پوششی در ترافیک شبکه، تغییراتی در ویژگی‌های آماری ترافیک سالم مانند شکل توزیع، همبستگی و آنتروپی ایجاد می‌شود که می‌توان از آنها در تشخیص کانال‌های زمانبندی‌دار پوششی استفاده کرد.

در این تحقیق، روش‌های آماری تشخیص این‌گونه کانال‌ها مورد شناسایی و تجزیه‌وتحلیل قرار گرفته، و میزان نامحسوسی دو کانال L بیت به N بسته و کانال غیرقابل تشخیص به‌طور عملی مورد ارزیابی قرار می‌گیرد. نتایج تحقیق نشان می‌دهد که با استفاده از آزمون‌های کلموگروف–اسمیرنوف، رگولاریتی، آنتروپی تصحیح‌شده و آنتروپی شرطی تصحیح‌شده، می‌توان کانال L بیت به N بسته را به‌طور کامل تشخیص داد، همچنین نامحسوسی طرح کانال غیرقابل تشخیص، به طور عملی اثبات می‌شود.

واژه‌های کلیدی: کانال‌های زمانبندی‌دارپوششی، تشخیص، آنتروپی، قاعده‌مندی، کلموگروف- اسمیرنوف

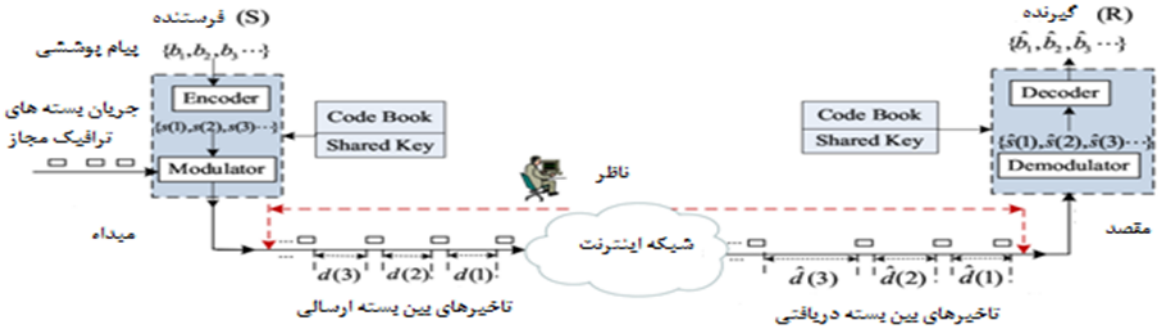
۱. مقدمه

در طول چند دهه گذشته، وظایف و خدمات انسانی به معادل دیجیتالی آنها تبدیل شده است. با وابستگی فزاینده به اطلاعات و فناوری اطلاعات، حفظ محرمانگی داده‌های حساس ذخیره‌شده در رایانه از اهمیت ویژه‌ای برخوردار است. سازمان‌ها برای مقابله با تهدید نشت اطلاعات حساس و محرمانه، سازوکارهای امنیتی شامل نظارت بر تمام ترافیک ورودی و خروجی، رمزنگاری قوی برای تمام ارتباطات داده، استفاده از سیستم‌های تشخیص نفوذ و اجرای سفت و سخت سیاست‌های امنیتی ایستگاه‌های کاری را پیاده‌سازی می‌کنند. با این حال کانال‌های پوششی به عنوان یک تهدید رو به رشد مطرح می‌باشند.

برخلاف رمزنگاری که تنها باعث جلوگیری از دسترسی افراد غیرمجاز به محتویات داده‌ها می‌شود، کانال‌های پوششی، وجود ارتباطات را مخفی می‌سازند. هدف اصلی در کانال‌های پوششی، مخفی کردن اطلاعات محرمانه درون‌رسانه‌ای است که برای انتقال ترافیک مجاز کاربران استفاده می‌شود. بهترین رسانه برای پیام‌های پوششی باید دو ویژگی داشته باشد. اولاً، بایستی رایج باشد. یعنی استفاده از چنین رسانه‌ای خود نبایستی به عنوان یک ناهنجاری در نظر گرفته شود. دوم این که، تغییرات حاصل از به‌کارگیری کانال پوششی در رسانه نبایستی برای شخص ثالثی که از این فرآیند بی‌اطلاع است قابل مشاهده باشد[۱].

برای کانال‌های پوششی، تعاریف مختلفی ارائه شده است. لمپسن در سال ۱۹۷۳ برای اولین بار کانال پوششی را در سیستم‌های رایانه‌ای MLS مطرح کرده است که به عنوان مکانیزمی در نظر گرفته شده که در آن، پردازه‌ای که در یک سطح امنیتی بالاتر قرار دارد، قادر است با بهره‌برداری از منابع اشتراکی مانند CPU، حافظه و یا رسانه ذخیره‌سازی، اطلاعات را به یک پردازه در یک سطح امنیتی پایین‌تر نشت دهد[۲]. از نظر گلیگور، کانال پوششی یک کانال ارتباطی انگلی است که به منظور ارسال اطلاعات بدون اجازه یا آگاهی طراح، مالک یا اپراتور کانال، از پهنای باند آن

* رایانامه نویسنده پاسخگو: beyrami60@yahoo.com



شکل ۱. نمای کلی از یک کانال زمانبندی‌دار پوششی [۶]

استفاده می‌کند[۳]. به طور کلی، یک کانال پوششی یک کانال ارتباطی است که سیاست‌های امنیتی را نقض کرده و مکانیزمی برای ارسال و دریافت داده بین سیستم‌ها است؛ به‌طوری که هیچ هشداری از ناحیه ابزارهای دفاعی شبکه مانند دیواره‌های آتش یا سیستم‌های تشخیص نفوذ اعلام نشود[۴]. دهقانی کانال پوششی را به طور مختصر و گویا « برقراری ارتباط پنهان در پوشش یک ارتباط مجاز » دانسته است[۳].

کانال‌های پوششی به دو دسته «کانال‌های انبارشی پوششی[1]» و «کانال‌های زمانبندی‌دار پوششی[2]» تقسیم می‌شوند. در یک کانال انبارشی پوششی، پردازه فرستنده، داده‌های مورد نظر را در جاهایی مثل هارد دیسک، حافظه، هدرهای بسته و غیره جاسازی و ارسال می‌نماید. پردازه گیرنده، اطلاعات پوششی را شناسایی و بازیابی می‌کند. در یک کانال زمانبندی‌دار پوششی، اطلاعات با تغییر ویژگی‌های زمانبندی ارسال می‌شود. در این کانال‌ها پیام پوششی با تغییر تأخیرهای بین بسته‌ای (اختلاف زمان بین دو بسته متوالی) و یا با تغییر ترتیب تاخیرهای بین بسته بدون در نظر گرفتن ویژگی‌های دیگر بسته (اندازه، نوع و ...) کدگذاری می‌شود[۳].

کانال‌های زمانبندی‌دار، خود به دو دسته فعال و منفعل تقسیم می‌شوند. کانال‌های زمانبندی‌دار پوششی منفعل بر روی ترافیک موجود تکیه دارند. این کانال‌ها ترافیک اضافی برای انتقال پیام پوششی ایجاد نمی‌کنند. از این رو کمتر مستعد تشخیص هستند. در کانال‌های زمانبندی‌دار فعال مهاجم بایستی یک سیستم را در اختیار بگیرد، تا بتواند اطلاعات را با ایجاد ارتباط جدید از سیستم در اختیار گرفته‌شده ارسال نماید. با توجه به این ارتباط جدید ایجادشده، این‌گونه کانال‌ها بیشتر مستعد تشخیص هستند. با این حال، ظرفیت این کانال‌ها از کانال‌های منفعل بیشتر است، چرا که مهاجم می‌تواند نرخ ارسال بسته‌ها را کنترل کند. شکل ۱ نمای کلی از یک کانال زمانبندی‌دار پوششی را نشان می‌دهد.

کانال‌های پوششی در پروتکل‌های شبکه شبیه به پنهان‌نگاری است. در پنهان‌نگاری، هدف اصلی پنهان کردن اطلاعات در محتوای صوت، تصویر یا متن می‌باشد. بنابراین گاهی اوقات کانال‌های پوششی را به‌عنوان پنهان‌نگاری در شبکه نیز در نظر می‌گیرند. در حالی که پنهان‌نگاری نیاز به شکلی از محتوا به عنوان پوشش دارد، در کانال‌های پوششی به پروتکلی از شبکه به عنوان رسانه نیاز است[۵].

روش‌های مقابله با کانال‌های زمانبندی‌دار پوششی را می‌توان بر-اساس روش‌های مبتنی بر پیشگیری و روش‌های مبتنی بر تشخیص دسته‌بندی کرد. هدف اصلی در دفاع مبتنی بر پیشگیری، حذف احتمال وجود کانال یا غیرعملی ساختن ایجاد یک کانال پوششی است. از آن‌جا که عملکرد کانال‌های زمانبندی‌دار پوششی به اطلاعات زمانبندی ترافیک شبکه بستگی دارد، بنابراین روش‌های مبتنی بر پیشگیری با مخدوش ساختن ویژگی‌های زمانبندی کانال، اطلاعات پوششی را در جریان‌های ترافیک از بین می‌برند. بنابراین، تاثیرات ناخواسته‌ای بر روی ترافیک کاربران مجاز داشته و ممکن است به‌طور ناخواسته عملیات عادی کاربران را بر هم زند.

در مقابل، رویکردهای مبتنی بر تشخیص یا آشکارسازی، از این واقعیت بهره‌برداری می‌کنند که کانال‌های زمانبندی‌دار پوششی، باعث ایجاد ناهنجاری در ویژگی‌های آماری ترافیک شبکه می‌شوند. معمولاً کانال‌های زمانبندی‌دار منفعل، تأخیرهایی را در تأخیرهای بین بسته ترافیک سالم اضافه می‌کنند که باعث می‌شوند توزیع این تاخیرها متفاوت از ترافیک سالم باشد. برای تشخیص این‌گونه کانال‌ها می‌توان از شکل توزیع تأخیرهای بین بسته استفاده کرد. در

1. Covert storage channels
2. Covert timing channels

مقابل، در کانال‌های زمان‌بندی‌دار فعال، تأخیرهای بین بسته به‌طور تصادفی انتخاب می‌شوند و تکرار الگوهای قابل مشاهده در ترافیک سالم را دنبال نمی‌کنند. برای تشخیص این‌گونه کانال‌ها می‌توان از قاعده‌مندی توزیع تأخیرهای بین بسته استفاده کرد.

تا کنون روش‌های مختلفی برای تشخیص کانال‌های پوششی مبتنی بر اختلاف‌های آماری در شکل و قاعده‌مندی کانال مطرح شده است. شناسایی تغییرات در شکل می‌تواند با هر آزمون آماری که اختلاف بین دو توزیع را اندازه گیری می‌کنند انجام شود؛ مانند آزمون کلموگرف-اسمیرنوف. کابوک یک روش ابتکاری براساس واریانس تاخیرهای بین بسته برای شناسایی قاعده‌مندی اضافه‌شده در کانال را پیشنهاد نموده است[۷]. جیان‌وچیو استفاده از آنتروپی را برای اندازه‌گیری شکل و قاعده‌مندی یک جریان شبکه ارائه داده است که می‌توان از آن برای تمایز بین ترافیک سالم و کانال پوششی استفاده کرد[۸]. این روش کاملاً موثر بوده و قادر به تشخیص بسیاری از کانال‌های زمانبندی‌دار موجود است.

برک یک روش آماری ساده را برای تشخیص کانال‌های زمانبندی‌دار پوششی مورد بررسی قرار داده است[۹]. در این روش فرض بر آن است که تأخیرهای بین بسته جریان شبکه تقریباً منطبق به یک توزیع نرمال است. بنابراین اگر ما یک هیستوگرام از تأخیرهای بین بسته ایجاد کنیم، به‌طوری که هر بلوک هیستوگرام بازه یکسانی داشته باشد، انتظار می‌رود بلوکی که در وسط مقادیر تأخیرهای بین بسته قرار دارد، بیشترین فراوانی را داشته باشد. وجود یک توزیع دووجهی یا چندوجهی، بیانگر وجود کانال زمانبندی‌دار پوششی خواهد بود.

استیلمن برای تشخیص کانال‌های زمانبندی‌دار پوششی، پیشنهاد یافتن همبستگی بین تأخیرهای بین بسته یک جریان شبکه و محتوای حافظه رایانه به خطر افتاده را مطرح کرده است[۱۰]. برای تشخیص کانال‌های پوششی، به مقدار کافی ترافیک از کانال پوششی نیاز است که باید جمع‌آوری شود، تا بتوان تحلیل و اثبات کرد که یک ارتباط در اختیار یک مهاجم قرار گرفته است. لذا با توجه به عدم دسترسی به دیتاست‌های ترافیک آلوده کانال‌های پوششی، بایستی این‌گونه کانال‌ها در یک شبکه واقعی پیاده‌سازی شوند.

برای ارزیابی کانال‌های پوششی معمولاً سه عامل زیر مورد بررسی قرار می‌گیرد و سعی می‌شود که با توجه به شرایط و کاربردها ، هریک از این عوامل در حد مورد نیاز رعایت شوند. این معیارها عبارت‌اند از:

ظرفیت[1]: بیشینه نرخ انتقال عاری از خطای داده در یک کانال پوششی را تعیین می‌کند. ظرفیت معمولاً بصورت بیت بر ثانیه اندازه‌گیری می‌شود، اما در کانال‌های پوششی شبکه ظرفیت می‌تواند به‌صورت بیت بر بسته نیز بیان می‌شود.

استحکام[2]: میزان سختی حذف کانال پوششی یا محدود کردن ظرفیت آن را از طریق نویز کانال یا نویزی که به‌طور ساختگی توسط یک ناظر ایجاد شده است نشان می‌دهد.

نامحسوسی: میزان سختی تشخیص کانال پوششی از طریق مقایسه مشخصات ترافیک کانال پوششی با ترافیک کانال مجاز را نشان می‌دهد.

در این مقاله، ما دو کانال زمانبندی‌دار پوششی مبتنی بر طرح کانال L بیت به N بسته و طرح کانال غیر قابل تشخیص را پیاده‌سازی کرده و میزان نامحسوسی هر کدام را با استفاده از روش‌های آماری می‌سنجیم. سایر معیارها شامل استحکام و ظرفیت، خارج از حوزه این تحقیق است.

ساختار ادامه مقاله بدین ترتیب است که در بخش دوم، روش‌های آماری تشخیص کانال‌های زمانبندی‌دار پوششی شرح داده می‌شود. در بخش سوم، جزئیات پیاده‌سازی دو کانال پوششی مورد نظر، شامل طرح کانال L بیت به N بسته و طرح کانال غیرقابل تشخیص ارائه می‌شود. در بخش چهارم، نحوه انجام آزمایش‌ها و ارزیابی نتایج روش‌های تشخیص بر روی کانال‌های پوششی مورد ارزیابی شرح داده شده است. در نهایت، جمع‌بندی مقاله و معرفی کارهای آینده در بخش پنجم مطرح گردیده است.

۲. روش‌های آماری تشخیص کانال‌های زمانبندی‌دار پوششی

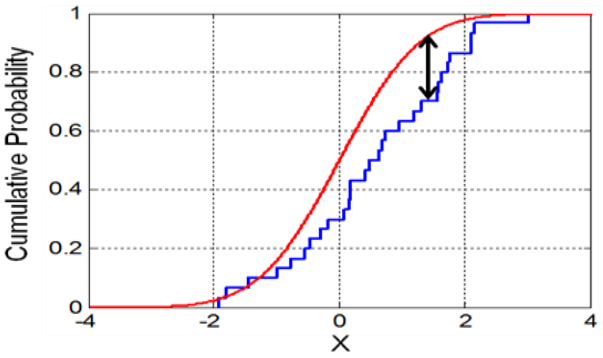
آزمون‌های آماری تشخیص کانال‌های زمانبندی‌دار پوششی، به دو دسته آزمون‌های شکل و آزمون‌های قاعده‌مندی تقسیم می‌شوند [۱۱]. شکل ترافیک با آمارهای مرتبه اول مانند میانگین، واریانس و توزیع توصیف می‌شود. قاعده‌مندی ترافیک یعنی قاعده‌مندی فرآیند در طول زمان که با آمارهای مرتبه دوم و یا بالاتر مانند همبستگی داده‌ها توصیف می‌شود. در ادامه، روش‌های مختلف آماری تشخیص کانال‌های زمانبندی‌دار پوششی تشریح می‌شود.

1. Capacity
2. Robustness

۱.۲. آزمون کولموگرف- اسمیرنوف

آزمون کولموگروف- اسمیرنوف (KS) یک آزمون ناپارامتری مفید و عمومی برای مقایسه این که آیا دو نمونه از یک توزیع یکسان هستند یا نه استفاده می‌شود. چون آزمون KS ناپارامتری است، متکی بر هیچ فرضی در مورد توزیع نیست و به همین دلیل می‌توان برای هر نوع توزیع این، آزمون را به‌کار برد. آزمون KS حداکثر فاصله بین دو تابع توزیع تجربی $s_1(x)$ و $s_2(x)$ متعلق به دو نمونه X1 و X2 را اندازه‌گیری می‌کند که از رابطه (۱) محاسبه می‌شود:

$$KSTEST = \max \mid s_1(x) - s_2(x) \mid \qquad (۱)$$



شکل ۲. نمودار آزمون کلموگروف- اسمیرنف

۲.۲. آزمون رگولاریتی[1]

کابوک و همکارانش سازوکار تشخیصی را براساس واریانس تأخیرهای بین بسته‌ها مطرح کرده‌اند. این روش مبتنی بر این واقعیت است که در ترافیک سالم، واریانس تاخیرهای بین بسته متوالی، در طول زمان تغییر می‌کند؛ در حالی که در یک کانال زمانبندی‌دار پوششی این مقدار نسبتاً ثابت باقی می‌ماند.

این روش، قاعده‌مندی را در جریان شبکه اندازه‌گیری می‌کند و در نتیجه می‌توان آن را به عنوان آزمون قاعده‌مندی در نظر گرفت[۷].

برای اندازه‌گیری قاعده‌مندی، تأخیرهای بین بسته جریان شبکه به پنجره‌های غیرهمپوشان با اندازه M تقسیم می‌شوند. پس از آن، برای هر پنجره، واریانس σ_i محاسبه می‌شود. قاعده‌مندی به صورت انحراف‌معیار اختلاف‌های بین هر زوج σ_i و σ_j برای تمام مجموعه‌ها به طوری که $i < j$ باشد، از رابطه (۲) محاسبه می‌شود:

$$regularity = STDEV\left(\frac{|\sigma_i - \sigma_j|}{\sigma_i}, i < j, \forall i, j\right) \qquad (۲)$$

۳.۲. تشخیص مبتنی بر آنتروپی

در تشخیص مبتنی بر آنتروپی سعی در اندازه‌گیری شباهت نسبی بین ترافیک مجاز و ترافیک مشکوک به کانال زمانبندی‌دار پوششی است. از آن‌جا که در طرح های کدگذاری کانال‌های زمانبندی‌دار، اطلاعات با تغییر در زمانبندی تأخیرهای بین بسته ارسال می‌شود، در نتیجه بر روی توزیع زمان‌های بین بسته‌ها تاثیر می‌گذارد. این روش تشخیص متکی بر استفاده از آنتروپی برای مشاهده چنین تغییری در توزیع است. محاسبه آنتروپی براساس توزیع نمونه‌های تاخیر بین بسته ترافیک سالم انجام می‌شود. هرگونه انحراف از این توزیع آموزشی باعث افت آنتروپی محاسبه‌شده می‌شود و می‌توان از آن در تشخیص کانال‌های زمانبندی‌دار پوششی استفاده کرد.

جیان‌وچیو برای تشخیص کانال‌های زمانبندی‌دار پوششی در یک جریان شبکه، از آنتروپی و آنتروپی شرطی استفاده کرده است[۸]. از آنتروپی می‌توان در تشخیص ناهنجاری ایجادشده در شکل توزیع یک کانال و در مقابل، از آنتروپی شرطی می‌توان در تشخیص ناهنجاری در قاعده‌مندی یک کانال استفاده کرد.

در یک فرآیند تصادفی $X=\{Xi\}$ که به صورت یک دنباله اندیس‌دار از متغیرهای تصادفی است، آنتروپی به صورت زیر تعریف می شود:

$$H(X_1, \dots, X_m) = -\sum_{x_1,\dots,x_m} P(X_1, \dots, X_m) \log P(X_1, \dots, X_m) \qquad (۳)$$

که در آن، $P(X_1, \dots, X_m)$ احتمال توأم متغیرهای تصادفی $X_1, \dots, X_m$ می‌باشد. معادله بالا همان آنتروپی مرتبه اول می‌باشد. از روی آنتروپی یک دنباله از متغیرهای تصادفی، آنتروپی شرطی یک متغیر تصادفی بر اساس دنباله‌ای از متغیرهای تصادفی داده‌شده، به صورت زیر تعریف می‌شود:

$$CE(X_m|X_{m-1}) = H(X_1, \dots, X_m) - H(X_1, \dots, X_{m-1}) \qquad (۴)$$

1. Regularity

نهایتاً نرخ آنتروپی یک فرآیند تصادفی برابر است با:

$$H(X) = \lim_{m\to\infty} H(X_m|X_{m-1}) \quad (۵)$$

نرخ آنتروپی به‌صورت آنتروپی شرطی یک دنباله نامتناهی از متغیرهای تصادفی تعریف می‌شود و نمی‌توان آن را برای نمونه‌های متناهی اندازه‌گیری کرد. بنابراین، نرخ آنتروپی بایستی با استفاده از آنتروپی شرطی برای نمونه‌های متناهی تخمین زده شود.

نرخ آنتروپی برای یک فرآیند ساده دارای توزیع مستقل و یکسان (iid) برابر با آنتروپی مرتبه اول است. یک فرآیند بسیار پیچیده، نرخ آنتروپی بالایی دارد، اما کمتر از آنتروپی مرتبه اول است. یک فرآیند بسیار منظم، نرخ آنتروپی پایینی دارد و برای یک فرایند تکراری محض، نرخ آنتروپی صفر می‌باشد.

در تخمین نرخ آنتروپی، بر اساس روش هیستوگرام، توابع چگالی احتمال با توابع چگالی احتمال تجربی جایگزین می‌شود. داده‌ها در Qبلوک دسته‌بندی می‌شوند. توابع چگالی احتمال تجربی بوسیله نسبت‌های الگوها تعیین می‌شود. نسبت الگو احتمال آن الگو است. در اینجا یک الگو به عنوان یک دنباله از شماره‌های بلوک‌های هیستوگرام تعریف می‌شود. تخمین نرخ آنتروپی یا همان آنتروپی شرطی تصحیح‌شده به صورت زیر محاسبه می‌شود:

$$CCE(X_m|X_1,\dots,X_{m-1}) = CE(X_m|X_1,\dots,X_{m-1}) + perc(X_m).EN(X_1) \quad (۶)$$

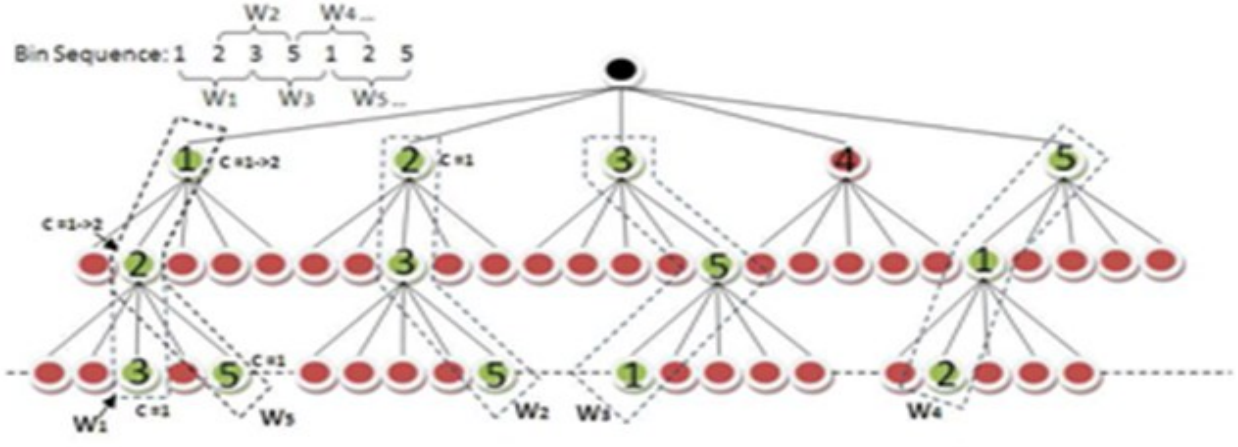
که در آن، $perc(X_m)$ درصدی از الگوهای یکتا با طول m است و $EN(X_1)$ آنتروپی مرتبه اول می‌باشد. بهترین تخمین نرخ آنتروپی، کمینه آنتروپی شرطی تصحیح‌شده به ازای مقادیر مختلف m است.

با توجه به این که در آزمون‌های تشخیص، از نمونه‌های نسبتاً کوچکی از تأخیرهای بین بسته‌ای استفاده می‌شود (در بسیاری از مواقع ۲۰۰۰ تأخیر بین بسته‌ای)، در نتیجه بسیاری از بلوک‌های هیستوگرام خالی می‌ماند. جیان‌وچیو برای حل مشکل نمونه‌های محدود، از آنتروپی تصحیح‌شده به صورت زیر استفاده می‌کند:

$$CEN = H + perc(X_1).\mathrm{H} \quad (۷)$$

که در آن، H آنتروپی مرتبه اول و $Perc(x_1)$ نسبتی از بلوک‌های هیستوگرام است که دقیقاً شامل یک تأخیر بین بسته از نمونه مورد بررسی است.

در تشخیص مبتنی بر آنتروپی، یک درخت با ارتفاع m ایجاد می‌شود. سطح درخت نشان‌دهنده طول الگوها است. به عنوان مثال، فرزندان ریشه نشان‌دهنده الگوهای با طول یک و گره‌های برگ نشان‌دهنده الگوهای با طول m می‌باشد. در آزمون آنتروپی تصحیح‌شده، ارتفاع درخت m=1 و گره ریشه دارای ۶۵۵۳۶ فرزند است. در مقابل، در آزمون آنتروپی شرطی تصحیح‌شده، ارتفاع درخت m=50 و هر گره دارای ۵ فرزند است. هر گره دارای یک شمارنده است که تعداد رخدادهای الگو را نشان می‌دهد.



شکل ۳. درخت محاسبه آنتروپی شرطی تصحیح‌شده [۱۱]

نحوه محاسبه آنتروپی شرطی تصحیح‌شده به این صورت است که در مجموعه متشکل از تأخیرهای بین بسته، مجموعه متناظر با شماره بلوک‌های هیستوگرام ایجاد می‌شود. این مجموعه به پنجره‌های کوچکی با اندازه ۵۰ (بیانگر طول الگوهای مورد نظر) و با شیفت پنجره به اندازه یک واحد تقسیم می‌شود. برای هر پنجره، درخت را پیمایش می‌کنیم. اگر گره در حال حاضر در درخت وجود داشت (یعنی الگو در پنجره قبلی رخ داده باشد)، فراوانی را یک واحد افزایش می‌دهیم، در غیر این صورت یک گره جدید با فراوانی یک، نشان‌دهنده اولین وقوع الگو می‌باشد. پس از پیمایش درخت، مقادیر آنتروپی برای هر سطح از درخت محاسبه شده و حداقل مقدار آنتروپی شرطی تصحیح‌شده در میان تمام سطوح به عنوان آنتروپی شرطی تصحیح‌شده یا همان تخمین نرخ آنتروپی برای آن مجموعه در نظر گرفته می‌شود.

$$EER(X) = \min (CCE(X, m) \mid m = 1, \dots , n) \quad (۸)$$

برای مثال، فرض می‌کنید در یک درخت با ارتفاع سه، دنباله شماره بلوک‌ها به صورت {۱, ۲, ۳, ۵, ۱, ۲, ۵} باشد. شکل ۳ نحوه ایجاد درخت را نشان می‌دهد. ابتدا دنباله شماره بلوک ها به Wi پنجره با اندازه سه تقسیم می‌شود و پس از آن، درخت با به‌روزرسانی فراوانی‌ها پیمایش می‌شود.

۳. طرح کدگذاری کانال زمان‌بندی‌دارپوششی

سلکی دو طرح کانال L بیت به N بسته و کانال غیرقابل تشخیص را برای ایجاد کانال زمانبندی‌دار پوششی ارائه داده است که در آن، بیشتر به نرخ ارسال داده و نرخ خطای ارسال داده در کانال پرداخته شده است[۱۲]. ولی نامحسوسی این کانال‌ها به طور عملی مورد بررسی قرار نگرفته است که این تحقیق به این مهم می‌پردازد. در این بخش، طرح‌های ارائه‌شده معرفی می‌شود و میزان نامحسوسی آنها در بخش ۴ مورد ارزیابی قرار خواهد گرفت.

۱.۳. طرح کانال L بیت به N بسته

در این طرح، از تأخیرهای بین بسته ترافیک کانال مجاز، برای سوار کردن اطلاعات پوششی استفاده می‌شود. بدین صورت که یک رشته L بیتی در دنباله‌ای از N تأخیر بین بسته $T_1, T_2, \dots, T_n$ کدگذاری می‌شود. تأخیرهای بین بسته Ti از رابطه (۸) محاسبه می‌گردد:

$$T_i = \Delta + k_i \cdot \delta \quad (۹)$$

در رابطه فوق، Δحداقل فاصله زمانی بین ارسال دو بسته متوالی است، به نحوی که اطلاعات کدگذاری به دلیل لغزش زمانی شبکه ناشی از تأخیرهای ایجادشده در مسیریاب‌ها از بین نرود. همچنین δ نیز حداقل اختلاف زمانی بین دو کلمه‌کد متفاوت است، به نحوی که تداخلی بین و دو کلمه‌کد به‌وجود نیامده و در گیرنده قابل تمایز باشند. به‌طور مثال، یک طرح ۴ بیت به ۲ بسته دارای جدول کلمه‌کدها، به صورت زیر خواهد بود:

جدول ۱. طرح کدگذاری ۴ بیت به ۲ بسته

۲ تاخیر بین بسته	۴-بیت	۲ تاخیر بین بسته	۴-بیت
[70] [70]	1000	[50] [50]	0000
[50] [80]	1001	[50] [60]	0001
[80] [50]	1010	[60] [50]	0010
[80] [60]	1011	[60] [60]	0011
[60] [80]	1100	[50] [70]	0100
[80] [70]	1101	[70] [50]	0101
[70] [80]	1110	[60] [70]	0110
[80] [80]	1111	[70] [60]	0111

۲.۳. طرح کانال غیرقابل تشخیص

در طرح کدگذاری کانال غیرقابل تشخیص، از یک الگوی ۸ بیت به ۲ بسته استفاده شده است[۱۲]. این طرح که تفاوت‌های عمده‌ای با طرح کانال L بیت به N بسته دارد، کد اسکی ۸ بیتی هر کاراکتر به ۲ تاخیر بین بسته T1، T2 در سه مرحله نگاشت می‌شود. جدول کدهای اشتراکی شامل نگاشت رشته‌های باینری ۸ بیتی به یک بردار دوتایی $(\frac{K_1}{16}, \frac{K_2}{16})$ است که در آن، K_1 و K_2 اعداد صحیح بین ۰ تا ۱۵ می‌باشد. بردار ۲۵۶ تایی $(\frac{K_1}{16}, \frac{K_2}{16})$ برای تمام رشته‌های باینری ۸ بیتی کافی است.

بردار $(\frac{K_1}{16}, \frac{K_2}{16})$ به طور مستقیم به تأخیرهای بین بسته نگاشت نمی‌شود، این تاخیرها بعد از طی سه مرحله به دست خواهند آمد که در شکل ۴ نشان داده شده است. فرض کنید فرستنده می‌خواهد یک پیام شامل n کاراکتر $\{C_1,C_2,\dots,C_n\}$ را انتقال دهد. در اولین مرحله از این طرح، جدول کدها برای هر کاراکتر پیام بررسی شده و بردار (x_{2k-1}, x_{2k}) به‌دست می‌آید که نشان‌دهنده کلمه کد برای کاراکتر C_k است. پیام در مرحله اول، به یک دنباله از اعداد به صورت زیر تبدیل می‌شود:

$$X = \{x_1, x_2, \dots, x_{2n-1}, x_{2n}\}$$

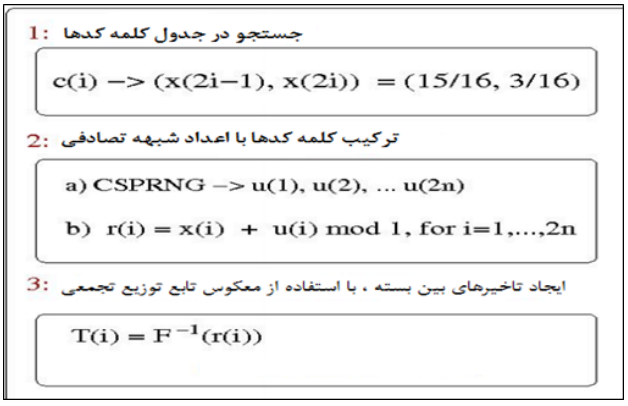
در گام بعدی، از یک مولد تولید اعداد شبه تصادفی برای تولید اعداد تصادفی در بازه (۰,۱) به صورت $u_1, u_2, \dots, u_{2n-1}, u_{2n}$ استفاده می‌شود. مقدار اولیه استفاده‌شده در تولید اعداد تصادفی به عنوان یک کلید مخفی بین فرستنده و گیرنده به اشتراک گذاشته می‌شود. بعد از تولید اعداد تصادفی u این اعداد با بردار x ترکیب شده و بردار جدید $r = r_1, r_2, \dots, r_{2n-1}, r_{2n}$ که حاصل XOR بردار u با x و یا جمع به پیمانه ۱ است به دست می‌آید.

$$r = x_k \oplus u_k \cong x_k + u_k \bmod 1 \qquad (۱۰)$$

در مرحله پایانی از معکوس تابع توزیع تجمعی ترافیک مجاز برای ایجاد تأخیرهای بین بسته استفاده می‌شود:

$$T_k = F^{-1}(r_k) \qquad (۱۱)$$

تابع $F(x)$ تابع توزیع تجمعی تأخیرهای بین بسته ترافیک مجاز است که کانال زمانبندی‌دار پوششی بر روی آن پیاده‌سازی شده است.



شکل ۴. مراحل کدگذاری و تولید تاخیرهای بین بسته

بدین ترتیب، پیام $\{C_1, C_2, \dots, C_n\}$ با استفاده از تاخیرهای بین بسته $T_1, T, \dots, T_{2n-1}, T_{2n}$ ارسال می‌شود. دستیابی به دنباله تاخیرهای بین بسته، بدون داشتن مقدار اولیه‌ای که در تولید اعداد شبه تصادفی استفاده شده حتی با در اختیار داشتن تابع توزیع تجمعی ترافیک مجاز $F(x)$، از لحاظ محاسباتی غیرممکن است.

فرآیند بازیابی پیام در سمت گیرنده، عکس مراحل ارسال در سمت فرستنده است. فرض کنید $R_1, R_2, \dots, R_{2n}$ تأخیرهای بین بسته‌های دریافتی باشد. در ابتدا با استفاده از تأخیرهای بین بسته دریافتی در سمت گیرنده و تابع توزیع تجمعی ترافیک مجاز، دنباله مقادیر r_i^* به دست می‌آید:

$$r_i^* = F(R_i) \qquad (۱۲)$$

در گام بعدی، دنباله اعداد تصادفی $u = u_1, u_2, \dots, u_{2n-1}, u_{2n}$ با استفاده از مقدار اولیه مشترک بین گیرنده و فرستنده تولید می‌شود و مقادیر بردار x_i^* به دست می‌آید.

$$x_i^* = r_i^* \oplus (1 - u_i) \qquad (۱۳)$$

و در گام آخر، مقادیر x_i^* به نزدیکترین مقدار $\frac{K}{16}$ گرد شده و مقادیر x_i^d از رابطه ۱۳ محاسبه می‌شود:

$$x_i^d = \frac{\lfloor 16 . x_i^* + 0.5 \rfloor}{16} \qquad (۱۴)$$

در نهایت با جستجوی در جدول کدها، بردار (x_{2k}^d, x_{2k+1}^d) به عنوان کاراکتر c_k^d کدگشایی شده و کل پیام بازیابی شده عبارت از $c_1^d, c_2^d, \dots, c_n^d$ خواهد شد.

۴. ایجاد ترافیک کانال‌ها و ارزیابی نامحسوسی و تفسیر نتایج

در این بخش به ارزیابی نامحسوسی طرح‌های مطرح‌شده در بخش ۳ می‌پردازیم. در این‌جا نحوه پیاده‌سازی دو طرح کانال پوششی و جمع‌آوری مجموعه داده‌ها، متدولوژی تشخیص کانال و نحوه ارزیابی عملی و تفسیر نتایج تشریح می‌گردد.

۱.۴. نحوه پیاده‌سازی کانال و تهیه مجموعه داده‌ها

مجموعه داده‌های حاوی تأخیرهای بین بسته که در آزمایش‌ها استفاده شده، با ایجاد جریان‌های ترافیک شبکه و نیز از مجموعه داده DARPA جمع‌آوری شده است[۱۳]. به منظور استخراج تاخیرهای بین بسته ترافیک HTTP و Telnet، ابتدا ترافیک ضبط‌شده براساس شماره پورت (برای مثال ۸۰ برای HTTP و ۲۳ برای Telnet فیلتر می‌شود. از آدرس‌های IP مبداء و مقصد برای جدا کردن جریان‌های روی هم افتاده و ایجاد یک دنباله از تأخیرهای بین بسته‌ای مرتبط برای هر جریان مجزا استفاده می‌کنیم و به ازای هر جریان، تاخیرهای بین بسته‌ای را به صورت اختلاف زمان بین دو بسته متوالی محاسبه می‌کنیم. در ادامه این دنباله‌ها را ترکیب کرده تا دو مجموعه مجزا از داده‌ها را ایجاد کنیم.

مجموعه اول به عنوان مجموعه آموزشی در نظر گرفته می‌شود که برای ایجاد بلوک‌های آموزشی استفاده می‌گردد. از مجموعه دوم به عنوان نمونه‌های ترافیک سالم استفاده می‌شود. تعداد تأخیرهای بین بسته در هر مجموعه در جدول ۱ نشان داده شده است.

جدول۱. تعداد تأخیرهای بین بسته‌ای مجموعه آموزشی

نوع ترافیک	نوع مجموعه	تعداد تاخیرهای بین بسته‌ای
HTTP	داده‌های آموزشی	۶۸۰۰۰
	داده نمونه سالم	۶۸۰۰۰
Telnet	داده آموزشی	۵۰۰۰۰
	داده نمونه سالم	۵۰۰۰۰

در پیاده‌سازی کانال‌های زمانبندی‌دار پوششی، فرستنده و گیرنده به عنوان دو سیستم در شبکه اینترنت در نظر گرفته شده که ۱۴گام از هم فاصله دارند و میانگین زمان رفت و برگشت (RTT) آن ۱۳۴ میلی‌ثانیه می‌باشد. اطلاعات پوششی شامل یک فایل متنی متشکل از ۵۱۱ کاراکتر می‌باشد که از طریق کانال پوششی ارسال می‌شود. تعداد دفعات تکرار آزمایشات نیز ۱۰ بار در ساعات و روزهای مختلف هفته می‌باشد. برنامه کانال‌های زمانبندی‌دار پوششی را به زبان پایتون پیاده‌سازی نموده‌ایم که شامل برنامه سرور و برنامه کلاینت می‌باشد و به ترتیب به عنوان فرستنده و گیرنده عمل می‌کند. فرستنده زمان‌های بین ارسال بسته‌های TCP را با استفاده از تابع sleep(T) کنترل می‌کند. مقدار T زمان بین ارسال دو بسته را مشخص می‌کند. گیرنده به طور منفعل زمان‌های دریافت بسته‌های TCP را ثبت می‌کند و با استفاده از جدول کلمه‌کدهای اشتراکی، پیام ارسال‌شده را کدگشایی می‌کند. کانال به صورت یک‌طرفه پیاده‌سازی می‌شود به طوری که فرستنده هیچ بازخوردی در رابطه با زمان‌های دریافت بسته‌ها یا این که آیا پیام به درستی کدگشایی شده، دریافت نمی‌کند. این کار، کارآیی کانال زمانبندی را محدود می‌کند، ولی نامحسوسی کانال افزایش می‌یابد.

در کانال L بیت به N بسته، ما یک طرح ۴ بیت به ۲ بسته براساس ترافیک HTTP با مقادیر مختلف پارامترهای سیستمی D و d پیاده‌سازی نموده‌ایم که مقادیر آنها به صورت زیر می‌باشد:

d =10,20 D==10,20,30,40

در کانال غیرقابل تشخیص، ما یک طرح ۸ بیت به ۲ بسته را بر روی ترافیک HTTP و Telnet پیاده‌سازی می‌کنیم که در آن کانال، الگوهای تاخیرهای بین بسته ترافیک سالم را تقلید می‌نماید. این کانال در واقع ترکیبی از روش کانال مبتنی بر مدل جیان‌وچیو و روش L بیت به N بسته است که در آن از نوعی مولد تولید اعداد شبه تصادفی استفاده شده است[۱۴]. در پیاده‌سازی این کانال از توابع کتابخانه‌ای پایتون برای تولید اعداد تصادفی استفاده می‌کنیم.

توزیع ترافیک مجاز HTTP از لحاظ آماری و بر اساس تخمین حداکثرسازی درست‌نمایی[1] (MLE)، و با حداقل میانگین مربعات ریشه[2] (RMSE)، منطبق بر توزیع ویبول بوده که در آن، مقدار پارامترهای توزیع λ را ۰.۴۲۶ و پارامتر شکل K را ۰.۱۲۵در نظر می‌گیریم. این همان مقادیری است که جیان‌وچیو در کانال مبتنی بر مدل خود برای مدل‌سازی ترافیک HTTP استفاده کرده است[۱۴].

تابع توزیع تجمعی توزیع ویبول و تابع معکوس آن به صورت زیر است:

$$F(x) = P[X \le x] = 1 - e^{-}(\frac{x}{\lambda})^{K} \quad ,x \ge 0 \qquad (۱۵)$$

$$F^{-1}(x) = \lambda((-ln(1-x))^{1/K} \quad ,0 \le x < 1 \qquad (۱۶)$$

با توجه به این که توزیع ترافیک مجاز Telnet از لحاظ آماری منطبق بر توزیع پارتو است، در توابع توزیع تجمعی و تابع معکوس آن مقدار پارمترهای توزیع α را ۱۰۰ میلی‌ثانیه و پارامتر β را ۰.۹۵ در نظر می‌گیریم[۶, ۱۲]. توابع توزیع تجمعی و تابع معکوس توزیع پارتو به صورت زیر است:

$$F(x) = P[X \le x] = 1 - (\frac{\alpha}{x})^{\beta} \quad ,x > \alpha\ ,\alpha,\beta > 0 \qquad (۱۷)$$

$$F^{-1}(x) = \alpha(\frac{1}{1-x})^{1/\beta} \quad ,0 < x < 1 \qquad (۱۸)$$

در پیاده‌سازی این کانال، مقدار اولیه تولید اعداد تصادفی را ۱۰۲۴در نظر می‌گیریم که می‌تواند هر مقدار دلخواه دیگری نیز باشد ولی بایستی بین گیرنده و فرستنده مشترک باشد.

۲.۴. متدولوژی تشخیص

به منظور تشخیص کانال‌های زمانبندی‌دار، ما آزمون‌های آماری تشخیص را بر روی نمونه‌های ترافیک پوششی و ترافیک سالم اجرا می‌کنیم. در ابتدا نرخ خطای مثبت غلط هدف را با مقدار ۰.۰۱ مقداردهی کرده و برای رسیدن به این نرخ خطای مثبت غلط، نمرات آستانه که پوششی یا سالم بودن یک نمونه را تعیین می‌کند، به دست می‌آید. نمرات آستانه به صورت ۱ امین و ۹۹ امین صدک، یعنی کمترین و بیشترین نمرات در آزمایش‌های مختلف بر روی نمونه‌های سالم به دست می‌آید. برای تعیین نمرات آستانه در آزمون‌های آنتروپی شرطی تصحیح‌شده و کولموگروف-اسمیرنوف، ما

1. Maximum Likelihood Estimation
2. Root Mean Squared Error

از صدک ۹۹ ام و برای آزمون‌های آنتروپی تصحیح‌شده و رگولاریتی از صدک ۱ام نمرات به‌دست‌آمده از نمونه‌های ترافیک سالم استفاده می‌کنیم. نمرات آزمون‌ها به شرح ادامه تفسیر می‌شوند.

در آزمون کولموگروف- اسمیرنوف، ما فاصله بین نمونه‌های تست و مجموعه آموزشی که نشان‌دهنده رفتار مجاز است را اندازه‌گیری می‌کنیم. بنابر این، در صورتی که نمره آزمون کمتر از آستانه باشد، حاکی از آن است که نمونه نزدیک به رفتار عادی است. با این حال، اگر نمونه منطبق بر رفتار مناسب نباشد، نمره آزمون بزرگتر از آستانه خواهد شد و احتمال وقوع یک کانال زمانبندی‌دار پوششی را نشان می‌دهد.

در آزمون رگولاریتی، ما انحراف معیار واریانس نرمالیزه‌شده مجموعه‌های ۲۰۰ بسته‌ای را اندازه می‌گیریم. اگر نمره آزمون رگولاریتی کمتر از آستانه قابل قبول باشد، نمونه بسیار قاعده‌مند و منظم بوده و احتمالاً نشان‌دهنده وجود یک کانال زمانبندی‌دار پوششی خواهد بود.

در آزمون آنتروپی تصحیح‌شده اگر نمره آزمون کمتر از نمره آستانه باشد، بیانگر این است که نمونه منطبق بر توزیع مناسب نبوده و به احتمال زیاد از نوع ترافیک پوششی است.

در آزمون آنتروپی شرطی تصحیح‌شده اگر نمره آزمون بالاتر یا خیلی پایین‌تر از نمرات آستانه باشد، نشان‌دهنده احتمال وجود کانال زمانبندی‌دار پوششی است. هنگامی که نمره آزمون آنتروپی شرطی تصحیح‌شده بسیار پایین باشد، نمونه بسیار قاعده‌مند بوده و وقتی که نمره آزمون آنتروپی شرطی تصحیح‌شده بالاتر از نمره آستانه و یا خیلی نزدیک به آنتروپی مرتبه اول باشد، نمونه عدم همبستگی را نشان می‌دهد.

آزمون کولموگرف–اسمیرنوف و آزمون رگولاریتی با استفاده از نرم‌افزار MATLAB پیاده‌سازی شده است. آزمون‌های آنتروپی تصحیح‌شده و آنتروپی شرطی تصحیح‌شده به زبان C پیاده‌سازی شده است. ما آزمون‌های تشخیص را بر روی ۶۸ نمونه ۲۰۰۰ تایی از تأخیرهای بین بسته ترافیک سالم HTTP و ۵۰ نمونه از ترافیک Telnet اجرا کرده‌ایم که نتایج آن در جدول ۲ آمده است. در ادامه، از این نتایج برای استخراج نمرات آستانه استفاده می‌شود که در جدول ۳ نشان داده شده است. مقادیر ۱۰ درصد نمرات آستانه، در شبکه‌های بی‌سیم که نرخ خطای بالایی دارند به‌کار می‌رود.

جدول ۲. نمرات آزمون‌های مختلف تشخیص بر روی ترافیک مجاز

	HTTP		Telnet	
نوع آزمون	میانگین	انحراف معیار	میانگین	انحراف معیار
CEN	17.055	1.12	17.51	1.45
CCE	1.72	0.14	1.51	0.26
KS	0.231	0.111	0.19	0.14
Regularity	8.77	20.91	6.47	8.96

جدول ۳ . نمرات آستانه برای سالم بودن ترافیک در آزمون‌های مختلف تشخیص

	ترافیک HTTP	ترافیک Telnet	
نوع آزمون	نمرات آستانه	نمرات آستانه	نرخ مثبت غلط
CEN	$\leq$ 13.93 CEN	13.79 $\leq$ CEN	1%
	CEN 15.32 $\leq$	$\leq$ 15.19 CEN	10%
CCE	CCE $\leq$ 1.94	CCE 1.9 $\leq$	1%
	CCE $\leq$ 1.90	CCE 1.76 $\leq$	10%
KS	KS $\leq$ 0.5	KS $\leq$ 0.66	1%
	KS $\leq$ 0.4	KS $\leq$ 0.32	10%
Regularity	R $\leq$ 0.22	$\leq$ 0.34 R	1%
	R 0.33 $\leq$	$\leq$ 0.55 R	10%

۳.۴. نتایج ارزیابی نامحسوسی کانال‌های مورد نظر

میانگین و انحراف معیار نمرات آزمون‌های مختلف بر روی کانال L بیت به N بسته که به صورت یک طرح ۴ بیت به ۲ بسته پیاده‌سازی شده، در جدول ۴ نشان داده شده است. این کانال به‌خوبی ناهنجاری را هم در شکل و هم در قاعده‌مندی ترافیک نشان می‌دهد. شکل ناهنجار ترافیک این کانال به‌وسیله طرح کدگذاری ایجاد می‌شود.

جدول۴. نمرات آزمون‌ها بر روی کانال L بیت به N بسته (HTTP)

نمرات آستانه		CEN ≥ 13.93		CCE ≤ 1.94		KS ≤ 0.5		Regularity≥ 0.22	
δ	Δ	mean	std	mean	std	mean	std	mean	std
۱۰	۱۰	۱۱.۵۲	۰.۴۲	۰.۰۲۵	۰.۰۱۱	۰.۶۱۴	۰.۰۰۳	۰.۰۳۶	۰.۰۰۴
۱۰	۲۰	۱۱.۸۴	۰.۲۹	۰.۰۰۶	۰.۰۱	۰.۶۵	۰.۰۰۲	۰.۰۴۳	۰.۰۰۱
۱۰	۳۰	۱۰.۹۳	۰.۱۶	۰.۰۰۹	۰.۰۱۶	۰.۶۵	۰.۰۰۱	۰.۰۱۸	۰.۰۰۶
۱۰	۴۰	۱۱.۱۳	۰.۲۱	۰.۰۰۲	۰.۰۰۳	۰.۶۸	۰.۰۰۱	۰.۰۴۵	۰.۰۰۹
۵	۱۰	۱۱.۷۸	۰.۴۹	۰.۰۲۵	۰.۰۰۸	۰.۶۰	۰.۰۰۵	۰.۰۵۹	۰.۰۲۵
۵	۲۰	۹.۴۵	۰.۲۱	۰.۰۰۹	۰.۰۱۰	۰.۶۵۵	۰.۰۰۱	۰.۱۸۶	۰.۱۴۹
۵	۳۰	۱۰.۸۰	۰.۴۲	۰.۰	۰.۰	۰.۶۵	۰.۰۰۲	۰.۰۳۵	۰.۰۱۶
۵	۴۰	۱۰.۸۵	۰.۰۸۴	۰.۰۰۲	۰.۰۰۳	۰.۶۸	۰.۰۰۱	۰.۰۴۰	۰.۰۱۲

در آزمون‌های رگولاریتی و آنتروپی تصحیح‌شده، نمرات حاصل کمتر از نمرات آستانه بوده و نشان‌دهنده پوششی بودن ترافیک است. در آزمون‌های آنتروپی شرطی تصحیح‌شده و کولموگروف- اسمیرنوف میانگین نمرات به‌دست‌آمده بایستی کوچکتر از نمرات آستانه باشد، تا به عنوان رفتار سالم در نظر گرفته شود. ولی این نمرات در آزمون کولموگروف- اسمیرنوف بزرگتر بوده، از طرفی میانگین نمرات در آزمون آنتروپی شرطی تصحیح‌شده خیلی نزدیک به صفر می‌باشد و قاعده‌مند بودن ترافیک را نشان می‌دهد و بیانگر پوششی بودن ترافیک است.

نتایج آزمون‌های مختلف تشخیص بر روی کانال غیرقابل تشخیص بر اساس ترافیک HTTP و Telnet در جدول ۵ و ۶ نشان داده شده است. همان‌طور که مشاهده می‌شود، هیچ‌یک از آزمون‌های تشخیص، قادر به شناسایی این کانال نبوده به‌طوری‌که در آزمون‌های آنتروپی شرطی تصحیح‌شده و آزمون کولموگروف-اسمیرنوف نتایج به‌دست‌آمده کوچکتر از مقادیر آستانه است و در آزمون آنتروپی تصحیح شده و آزمون رگولاریتی نتایج به‌دست‌آمده، بزرگتر از نمرات آستانه بوده که بیانگر عدم پوششی بودن ترافیک است. بنابراین توزیع این کانال از توزیع ترافیک سالم پیروی می‌کند. مقایسه نتایج، نمرات آزمون‌های مختلف تشخیص بر روی دو کانال مورد ارزیابی، به تفکیک هر آزمون در شکل ۳ نشان داده شده است.

۴.۴. تحلیل آزمون‌های تشخیص

مشکل اصلی در آزمون رگولاریتی، انحراف معیار بالای این آزمون در سنجش ترافیک سالم می‌باشد. این آزمون نسبت به وضعیت ترافیک سالم بسیار حساس است. برای مثال، در عبارت $\frac{|\sigma_i-\sigma_j|}{\sigma_j}$ اگر σ_i به‌دلیل تأخیرهای بین بسته مشابه بسیار کوچک و σ_j بزرگ باشد، در این صورت انحراف معیار در این آزمون بزرگ

جدول ۵ . نتایج آزمون‌های مختلف تشخیص بر روی کانال غیرقابل تشخیص (HTTP)

نوع آزمون	میانگین	انحراف معیار	تشخیص
CEN ≥ 13.93	18.94	0.112	0%
CCE ≤ 1.94	1.83	0.008	0%
(KS) ≤ 0.5	0.429	0.2	0%
Regularity ≥ 0.22	0.24	0.04	0%

جدول ۶ . نتایج آزمون‌های مختلف تشخیص بر روی کانال غیرقابل تشخیص (Telnet)

نوع آزمون	میانگین	انحراف معیار	تشخیص
CEN≥13.79	15.79	1.68	0%
CCE ≤ 1.9	0.78	0.04	0%
KS ≤ 0.66	0.65	0.02	0%
Regularity≥ 0.34	1.53	0.51	0%

خواهد شد. چنین وضعیتی باعث می‌شود که یک نمونه ترافیک آلوده، به‌عنوان ترافیک سالم در نظر گرفته شود. بنابراین، این آزمون دارای ضعف مشهودی در تشخیص ترافیک آلوده است. برای فرار از تشخیص این آزمون بایستی طرح الگوی کدگذاری در طول زمان تغییر نماید.

مزیت آزمون کولموگروف- اسمیرنوف در ناپارامتریک بودن آن است. بنابراین متکی بر هیچ فرضی در رابطه با توزیع نیست و می‌توان آن را برای هر توزیع به‌کار برد، اما محدود به توزیع‌های پیوسته است. ضعف آزمون کولموگروف-اسمیرنوف در این است که به جای اندازه‌گیری اختلاف در کل توزیع، حداکثر فاصله بین دو توزیع را اندازه‌گیری می‌کند. بنابراین، هنگامی که توزیع ترافیک

پوششی بسیار نزدیک به توزیع ترافیک مجاز باشد، این آزمون نمی‌تواند ترافیک پوششی را از ترافیک مجاز متمایز سازد.

آزمون آنتروپی تصحیح‌شده، نحوه توزیع یکنواخت تأخیرهای بین بسته را می‌سنجد. آزمون آنتروپی تصحیح‌شده نسبت به تغییرات کوچک در طول توزیع حساس است. با این حال، اگر توزیع یک کانال زمانبندی‌دار پوششی تقریباً مشابه ترافیک سالم باشد، آزمون آنتروپی تصحیح‌شده با شکست روبرو می‌شود.

در مقابل، آزمون آنتروپی شرطی تصحیح‌شده به جای توزیع، پیچیدگی یا قاعده‌مندی ترافیک را اندازه‌گیری می‌کند. بنابراین در صورتی که همبستگی ترافیک اصلی حفظ و توزیع تغییر یابد، آزمون آنتروپی شرطی تصحیح‌شده قادر به تشخیص ترافیک آلوده نخواهد بود. بنابراین، استفاده همزمان از هر دوی این آزمون‌ها، می‌تواند در تشخیص کانال‌های زمانبندی‌دار پوششی موثر باشد.

از آنجا که در آزمون‌های مبتنی بر آنتروپی از ساختار درختی برای محاسبه آنتروپی متغیرهای تصادفی استفاده می‌شود، بایستی برای فرار از تشخیص، الگوها به طور یکنواخت در طول درخت توزیع شده باشند، به‌طوری‌که احتمال هر الگو یکسان باشد. برای این کار نیاز است با تقلید ویژگی‌های آماری ترافیک سالم، باعث هموارسازی شکل توزیع و حفظ قاعده‌مندی ترافیک سالم شد.

۵. نتیجه‌گیری

در این مقاله چهار آزمون آماری جهت تشخیص کانال‌های زمانبندی‌دار پوششی ارائه شد و با پیاده‌سازی دو کانال زمانبندی‌دار پوششی در شبکه اینترنت، میزان نامحسوسی آنها با روش‌های آماری مطرح‌شده به طور عملی مورد ارزیابی قرار گرفت. آزمون‌های تشخیص تحت شرایط خاص می‌توانند کانال‌های زمانبندی‌دار پوششی خاصی را تشخیص دهند. یکی از دلایل عمده در تنوع بالای ترافیک سالم نهفته است. در تشخیص کانال‌های زمانبندی‌دار پوششی علاوه بر شکل توزیع ترافیک می‌توان از قاعده‌مندی الگوهای ترافیک نیز برای تمایز آنها از ترافیک سالم استفاده کرد. با توجه به نتایج به‌دست‌آمده، هر قدر کانال پوششی قادر به تقلید خواص آماری ترافیک سالم و تولید ترافیک با خواص مشابه باشد، تشخیص آن سخت‌تر می‌شود. فاصله بین بسته‌ها در ترافیک سالم، دارای خواصی از قبیل خودهمبستگی، تغییرات واریانس و حتی تغییر تابع توزیع می‌باشند.

نتایج این تحقیق نشان می‌دهد که طرح کانال غیرقابل تشخیص سلکی، علاوه بر نامحسوسی که از لحاظ تحلیلی توسط طراح آن اثبات شده، به صورت عملی نیز از نامحسوسی کامل برخوردار است و می‌توان از آن در کاربردهای مورد نظر استفاده کرد. در این تحقیق، کانال‌های مورد ارزیابی به صورت فعال پیاده‌سازی شده‌اند. لذا برای کارهای بعدی، پیاده‌سازی کانال‌های زمانبندی‌دار پوششی منفعل را برای افزایش میزان نامحسوسی آن می‌توان پیشنهاد داد. همچنین استفاده از روش‌های دیگر تشخیص مانند مدل‌سازی یا روش‌های داده‌کاوی در تشخیص کانال‌های پوششی نیز می‌تواند موضوع تحقیقات بعدی باشد.

۶. مراجع

[1] Wojciech Mazurczyk and K. Szczypiorski, "Towards Steganography Detection Through Network Traffic Visualisation " Institute of Telecommunications Warsaw University of Technology, 2012.

[2] B. W. Lampson, "A Note on the Confinement Problem," Communications of the ACM, 16, 10 (Oct. 1973), pp. 613-615, 1973.

[3] M. Dehghani and M. S. Esfahani, "Network Covert Channels: An Information Leakage Flow," Passive Defence Quarterly, vol. 9, 2012.

[4] U.S.DoD, "Trusted computer system evaluation criteria, TCSEC, "in National Computer Security Center, WashingtonDec, 1985.

[5] FABIEN A. P. PETITCOLAS, et al., "Information Hiding—A Survey," Proceedings of the IEEE, special issue on protection of multimedia content, 1999.

[6] Y. Liu, et al., "Robust and Undetectable Steganographic Timing Channels for i.i.d. Traffic," in Information Hiding. vol. 6387, ed: Springer Berlin Heidelberg, 2010, pp. 193-207.

[7] S. Cabuk, et al., "IP covert timing channels: design and detection," Proceedings of the 11th ACM conference on Computer and communications security, pp. 178–187, 2004.

[8] Steven Gianvecchio and H. Wang, "An Entropy-Based Approach to Detecting Covert Timing Channels," Dependable and Secure Computing, IEEE Transactions on, vol. 8, 2011.

[9] Vincent Berk, et al., "Detection of Covert Channel Encoding in Network Packet Delays," In Proceedings of FLOCON, 2005.

[10] R. M. Stillman, "Detecting IP covert timing channels by correlating packet timing with memory content," in Southeastcon, 2008. IEEE, 2008, pp. 204-209.

[11] Kush Kothari and M. Wright, "Mimic: An active covert channel that evades regularity-based detection," elsevier computer network, 2012.

[12] S. H. Sellke, et al., "Covert TCP/IP Timing Channels: Theory to Implementation," in Proceedings of The 28th Conference on Computer Communications (INFOCOM), 2009.

[13] DARPA Intrusion Detection Data Sets from 1999. Available: http://www.ll.mit.edu/mission/communications/cyber/CSTcorpora/ideval/data/

[14] S. Gianvecchio, et al., "Model-Based Covert Timing Channels: Automated Modeling and Evasion," Proceedings of the 11th international symposium on Recent Advances in Intrusion Detection, pp. 211–230, 2008.