مجله علمی - پژوهشی "پدافند الکترونیکی و سایبری"
سال چهارم، شماره ۳، پاییز ۱۳۹۵، ص ۴۶-۳۷

روشی برای مدل‌سازی و تولید ترافیک هنجار شبکه مبتنی بر ویژگی‌های اندازه و زمان ورود بسته‌ها با استفاده از قانون زیف

علی نقاش اسدی[۱]، محمدعبداللهی ازگمی[۲*]
۱- کارشناس ارشد، ۲- دانشیار، گروه نرم افزار دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
(دریافت: ۹۴/۱۱/۱۴، پذیرش: ۹۵/۰۵/۱۱)

چکیده

امروزه مدل‌سازی ترافیک شبکه و تولید ترافیک هنجار از اهمیت بالایی برخوردار است. تا به امروز مدل‌سازی‌های زیادی بر روی ویژگی‌های مختلف ترافیک شبکه انجام شده است که تقریباً اکثر آنها از توزیع‌های احتمالاتی استفاده کرده‌اند. در این مقاله، روش جدیدی برای مدل‌سازی ویژگی‌های مختلف ترافیک شبکه معرفی می‌شود که مبتنی بر قانون زیف است. قانون زیف یک قانون تجربی است که رابطه‌ای بین فراوانی و رتبه هر دسته در مجموعه داده‌ها، ارائه می‌کند. در این مقاله نشان داده می‌شود که قانون زیف می‌تواند به‌خوبی ویژگی‌های مختلف ترافیک شبکه را مدل‌سازی نماید. برای این منظور، دو ویژگی مهم ترافیک شبکه، یعنی اندازه و زمان بین ورود بسته‌های TCP و UDP، مورد مطالعه قرار گرفته است. از مدل‌سازی ویژگی‌های مختلف ترافیک شبکه می‌توان در زمینه‌های مختلفی از جمله شبیه‌سازی ترافیک شبکه و تولید ترافیک هنجار استفاده کرد. مزیت استفاده از قانون زیف این است که می‌تواند با کمترین اطلاعات، بیشترین شباهت را ایجاد کند. همچنین قانون زیف می‌تواند ویژگی‌های مختلف ترافیک شبکه را که ممکن است از توزیع ریاضی خاصی پیروی نکنند، مدل‌سازی کند. با توجه به روش ساده‌ای که این قانون ارائه می‌کند، علاوه بر دقت و محدودیت‌های کمتر نسبت به روش‌های پیشین، مدل‌سازی و شبیه‌سازی را در زمان مناسبی نیز انجام می‌دهد.
در این مقاله نشان داده خواهد شد که با دسته‌بندی مقادیر ویژگی‌ها و به‌دست آوردن رتبه آنها، می‌توان مدل‌سازی دقیقی از ویژگی‌ها ایجاد کرد. به عبارت دیگر، رتبه هر دسته، مدل به‌دست‌آمده از مقادیر ویژگی خواهد بود که می‌توان از آن در شبیه‌سازی استفاده کرد.

واژه‌های کلیدی: ترافیک شبکه، قانون زیف، مدل‌سازی، شبیه‌سازی، اندازه بسته‌ها، زمان بین ورود بسته‌ها

۱- مقدمه

امروزه تحلیل ترافیک شبکه و بررسی جنبه‌های مختلف آن، اهمیت زیادی پیدا کرده است. محققان به منظور بررسی عملکرد یک شبکه تازه تاسیس و یا بررسی مقاومت آن در برابر حمله‌های مختلف، نیازمند ترافیکی هستند که به وسیله آن بتوانند، عملکرد و امنیت یک شبکه را ارزیابی کنند. به همین منظور، محققان نیازمند تولیدکننده ترافیکی هستند که به بهترین و دقیق‌ترین شکل ممکن ترافیک واقعی را شبیه‌سازی کند. برای این کار، محققان باید ترافیک واقعی و هنجار را بررسی کرده و الگوهایی را از آن استخارج کنند؛ سپس به وسیله ابزارها و الگوریتم‌های مختلف آنها را شبیه‌سازی کرده و به شبکه وارد کنند. همچنین محققان با شناخت ترافیک هنجار، می‌توانند ترافیک ناهنجار ایجاد کرده و نحوه پاسخگویی شبکه را بررسی کرده و امنیت بهتری را فراهم کنند. بنابراین مدل‌سازی و شبیه‌سازی ویژگی‌های مختلف ترافیک شبکه از اهمیت بالایی برخوردار است.

این مقاله برای مدل‌سازی و شبیه‌سازی ویژگی‌های مختلف ترافیک شبکه، قانون زیف[1] را پیشنهاد می‌کند. مزیت این قانون، ساده و جامع بودن آن است. این قانون می‌تواند به راحتی ترافیک شبکه را مدل‌سازی کرده و به کمک مدل به دست آمده، شبیه‌سازی دقیقی ارائه کند.

در بخش ۲ این مقاله، قانون زیف و دلایل مدل‌سازی و شبیه‌سازی ویژگی‌های ترافیک شبکه معرفی می‌شوند. در بخش ۳، کارهای مرتبطی که در زمینه مدل‌سازی ترافیک شبکه ارائه شده است، معرفی می‌شود. در بخش ۴، روش ارائه شده برای مدل‌سازی و شبیه‌سازی ترافیک شبکه با استفاده از قانون زیف معرفی می‌شود. در این قسمت، مجموعه ترافیک مورد استفاده و روش انجام آزمایش‌ها معرفی می‌شوند. مدل‌سازی و شبیه‌سازی با استفاده از قانون زیف روی اندازه و زمان بین ورود بسته‌های TCP

* رایانامه نویسنده مسئول: azgomi@iust.ac.ir
1- Zipf

و UDP انجام شده و با ترافیک واقعی مقایسه می‌شوند.

در بخش ۵، آزمایش‌های مربوط به مدل‌سازی و شبیه‌سازی مورد ارزیابی دقیق‌تر قرار می‌گیرند. در این قسمت، نحوه تعیین تعداد دسته‌ها در قانون زیف که یک فرآیند مهم در مدل‌سازی ترافیک شبکه است، معرفی می‌شود. در بخش انتهایی نیز نتایج به دست آمده مورد بررسی قرار می‌گیرند.

۲- پیش‌زمینه

۲-۱- قانون زیف

جورج کینگزلی زیف[1]، استاد زبان شناسی دانشگاه هاروارد، در سال ۱۹۴۹ با بررسی کلمه‌های موجود در کتاب‌های مختلف، به نتایجی در مورد کلمه‌ها و میزان تکرار آنها در متن رسید. نتایج او به این صورت بود که اگر تمام کلمه‌های یک کتاب شمارش شده و از فراوانی زیاد به کم مرتب شوند، به این نتیجه می‌توان رسید که رتبه هر کلمه با فراوانی همان کلمه نسبت عکس دارد؛ یعنی تعداد دفعاتی که هر کلمه در متن ظاهر می‌شود، با رتبه همان کلمه در متن، رابطه معکوس دارد. این رابطه، به قانون زیف معروف شده است. بر طبق این قانون، کلمه‌ای که در رتبه ۱ قرار دارد، دو برابر بیشتر از کلمه‌ای که در رتبه ۲ قرار دارد، در متن ظاهر می‌شود. همچنین این کلمه ۳ برابر بیشتر از کلمه‌ای که در رتبه ۳ قرار دارد، ظاهر می‌شود [۲-۱].

این قانون، بین فراوانی Fre و رتبه r رابطه زیر نشان می‌دهد. بر اساس این رابطه، حاصل ضرب فراوانی یک کلمه در رتبه آن در کل متن، تقریباً عددی ثابت برابر با A=0.1 خواهد بود. در این رابطه، N مجموع تعداد کل کلمه‌ها است.

$$\frac{r_i * Fre_i}{N} = A \qquad (۱)$$

با توجه به این‌که $\frac{Fre_i}{N}$ برای هر کلمه، نشان دهنده احتمال وقوع آن کلمه در بین تمامی کلمه‌ها است؛ می‌توان رابطه فوق را به این شکل اصلاح کرد:

$$r_i * Prob_i = A \qquad (۲)$$

این رابطه را می‌توان به صورت لگاریتمی نیز تعریف کرد. شکل لگاریتمی این رابطه و نمودار به وجود آمده توسط آن، از اهمیت بالایی برخوردار بوده و در ادامه مورد استفاده قرار می‌گیرد.

$$\log r_i + \log Prob_i = A \qquad (۳)$$

از این رابطه می‌توان در محیط‌های دیگری نیز استفاده کرد. از نظر محققان بسیار عجیب است که چطور و چرا چنین رابطه ساده‌ای در بسیاری از محیط‌های پیچیده اتفاق می‌افتد. از این قانون به منظور تشخیص تقلب در منابع مالی و حسابداری استفاده شده است [۳]. همچنین این قانون برای شبیه‌سازی بازدید کاربران از صفحه‌های وب، پیشبینی رتبه سایت‌ها، و تشخیص ایمیل‌های ویروسی نیز مورد استفاده قرار گرفته است [۵-۴].

۲-۲-مدل‌سازی و شبیه‌سازی ترافیک شبکه

محققان برای تصور و فهم بهتر ترافیک شبکه تمایل بسیار زیادی دارند. تحلیل ترافیک شبکه، شامل فرآیند ثبت و ضبط[2] ترافیک شبکه و بررسی آن است [۶]. این مفهوم با نام‌های دیگری مانند تحلیل شبکه، تحلیل پروتکل، شنود[3] بسته و تحلیل بسته نیز شناخته می‌شود.

تحلیل ترافیک شبکه با اهداف گوناگونی انجام می‌شود و می‌تواند اطلاعات مهمی را در مورد الگوهای رفتاری کاربران فراهم سازد و مدیران شبکه با استفاده از آن می‌توانند درک بهتری نسبت به شبکه پیدا کنند. یکی از اهداف مهم تحلیل ترافیک شبکه، مدل‌سازی ویژگی‌های مختلف ترافیک شبکه است. مدل‌های ایجاد شده از ویژگی‌های مختلف ترافیک شبکه در زمینه‌های مختلفی قابل استفاده هستند که از جمله آنها می‌توان به تولید شبیه‌ساز ترافیک هنجار و حمله، تشخیص ناهنجاری‌ها و حمله‌ها، بررسی عملکرد سرویس‌های ارائه شده در شبکه، بررسی سیاست‌های امنیتی شبکه، پیش‌بینی نیازمندی‌های شبکه در آینده، و جلوگیری از نظارت مهاجمان بر ترافیک شبکه اشاره کرد.

تولید شبیه‌ساز ترافیک هنجار و حمله، یکی از کاربردهای مهم مدل‌سازی ترافیک شبکه است. به علت نبود مجموعه ترافیک مناسب برای آزمایش‌های مختلف، محققان در تلاش هستند که مولدهای ترافیکی ایجاد کنند که بیشترین شباهت را به مجموعه ترافیک واقعی داشته باشند. همچنین تولید حمله‌های مورد نیاز، یکی دیگر از نیازمندی‌های محققان در این زمینه است که با فهم درست از ترافیک شبکه و مدل‌سازی دقیق آن حاصل می‌شود. مولدهای ترافیک می‌توانند نقش مهمی را در ارزیابی عملکرد خدمات ارائه شده توسط یک شبکه تازه تاسیس و یا بررسی سیاست‌های امنیتی لحاظ شده برای یک شبکه، قبل از راه‌اندازی آن ایفا کنند. محققان مختلف، مولدهای ترافیک شبکه زیادی با ویژگی‌های مختلف ایجاد کرده‌اند که از جمله آنها می‌توان به Harpoon [۷]، D-ITG [۸]، TCPReplay [۹]، Avalanche [۱۰] و غیره اشاره کرد. همچنین تحلیل و مدل‌سازی ترافیک شبکه می‌تواند اطلاعات مناسبی در مورد

1- George Kingsley Zipf

2- Capture
3- Sniff

نیازمندی‌های یک شبکه در آینده ارائه کند و به وسیله این اطلاعات، مدیران شبکه می‌توانند راه‌کارها و برنامه‌ریزی‌های مناسبی برای این نیازمندی‌ها تعریف کنند.

در برخی سیستم‌ها به‌منظور جلوگیری از دسترسی مهاجمان به ترافیک شبکه و به‌دست آوردن اطلاعات حیاتی، اقدام به تولید ترافیک مصنوعی در بین ترافیک واقعی خود می‌کنند تا مانع از دسترسی مهاجمان به الگوهای ترافیک شبکه و اطلاعات حیاتی آن شوند [۱۱]. برای این منظور باید ترافیکی مشابه ترافیک واقعی شبکه ایجاد شود تا علاوه بر پنهان کردن اطلاعات مهم و رفتارهای واقعی شبکه، امکان تفکیک آن از ترافیک واقعی توسط مهاجمان امکان‌پذیر نباشد. در منبع [۱۲]، توضیحاتی در مورد مشکلات کسانی که سعی دارند ترافیک اینترنت را شبیه‌سازی کنند و ویژگی‌های آن را مورد مطالعه قرار دهند، آورده شده است. با این حال برای رفع این مشکلات نیز راه‌حل‌هایی ارائه شده است [۱۳].

از مدل‌سازی ترافیک شبکه می‌توان در جهت تشخیص ناهنجاری‌ها و حمله‌ها نیز استفاده کرد. برای مثال، با مدل‌سازی ویژگی‌های مختلف ترافیک هنجار شبکه و مقایسه آن با ترافیک واقعی، می‌توان هر گونه انحراف از مدل هنجار را به عنوان ناهنجاری در نظر گرفت [۱۴]. بنابراین دقت در مدل‌سازی، در تشخیص دقیق ناهنجاری‌ها و حمله‌ها و کاهش هشدارهای کاذب[1]، نقش مهمی را ایفا می‌کند. تحلیل ترافیک فقط به منظور شناسایی خطا در شبکه به کار گرفته نمی‌شود بلکه به فهم دلیل اصلی وقوع یک خطا و تاثیر آن بر روی ارتباطات بین کاربران کمک می‌کند.

۳- کارهای مرتبط

مدل‌سازی‌های زیادی بر روی ترافیک شبکه توسط محققان انجام شده است. آنها متناسب با اهدافشان، به بررسی ویژگی‌های مختلف ترافیک شبکه پرداخته و برخی از آنها را که قابلیت مدل‌سازی داشته‌اند، با تعریف روابطی، مدل کرده‌اند. برای مثال، در منبع [۱۵] ویژگی‌های مهم پروتکل‌های Telnet، SMTP، Http و FTP مدل شده است. محققان تا قبل از این تحقیقات، اکثر ویژگی‌های ترافیک شبکه را با توزیع پواسون[2] توصیف می‌کردند ولی در این منبع نشان داده شد که به جز ورودی TCP آغاز شده توسط کاربر، بقیه ورودی‌های اتصالات TCP پواسون نیستند.

در منبع [۱۶]، زمان بین ورود بسته‌های TCP و UDP مورد بررسی قرار گرفته است. در این منبع، زمان بین ورود بسته‌ها برای تطبیق با توزیع‌های ریاضی، به روش کولموگروف- اسمیرنوف[3] بررسی شده است. در این منبع، توزیع‌های مورد بررسی شامل توزیع ویبول، پارتو، گاما، لاگ- نرمال و نمایی هستند. میزان مناسب بودن زمان بین ورود بسته‌های TCP و UDP با هر یک از توزیع‌ها به کمک پارامتر KS نشان داده می‌شود که همان معیار کولموگروف- اسمیرنوف است و بیشینه[4] اختلاف را نشان می‌دهد. در نتیجه هر چقدر این پارامتر کمتر باشد، به معنی مناسب‌تر بودن آن توزیع است. این منبع توزیع پارتو و ویبول را به عنوان نزدیک‌ترین توزیع‌ها به زمان بین ورود بسته‌های TCP و UDP معرفی کرده است.

در منبع [۱۷] اندازه بسته‌ها[5] در ترافیک شبکه مورد بررسی قرار گرفته است. در این تحقیق سعی شده تا تابع چگالی احتمالی تخمین زده شود که کمینه[6] اختلاف ممکن را با نمودار اندازه بسته‌ها داشته باشد. در شبکه، بسته‌ها با بیشینه اندازه خاصی قابل انتقال هستند که این اندازه توسط پارامتر MTU[7] تعیین می‌گردد. به همین منظور، بسته‌های بزرگ باید به بسته‌های کوچک تقسیم شوند. این موضوع باعث می‌شود که شبیه‌سازی درستی از اندازه واقعی بسته‌ها صورت نگیرد. چون فراوانی بسته‌هایی که اندازه آنها برابر با MTU است به دلیل قطعه قطعه[8] شدن بسته‌ها، افزایش می‌یابد. این منبع با یکپارچه‌سازی[9] بسته‌های ترافیک، این مسئله را حل کرده است. با استفاده از روش یکپارچه‌سازی می‌توان پارامترهای توزیع اندازه بسته‌ها را تخمین زد. در روش یکپارچه‌سازی بسته‌ها، همه بسته‌هایی که آدرس مبدا و مقصد یکسان دارند و اولین بسته آنها کوچکتر از اندازه MTU نیست، با هم ترکیب می‌شوند.

تا قبل از در نظر گرفتن مفهوم خودهمانندی[10]، توزیع پواسون به عنوان مهم‌ترین توزیع برای مدل‌سازی ترافیک شبکه مطرح بود. مفهوم خودهمانندی به این معنی است که شکل ظاهری و نمودارهای آماری ویژگی‌ها در مقیاس‌های مختلف هیچگاه تغییر نخواهند کرد. در منبع [۱۸] نشان داده شده است که ویژگی‌های ترافیک شبکه ویژگی خودهمانندی دارند و بنابراین نمی‌توانند با توزیع پواسون مدل شوند. بعد از کشف این موضوع، توزیع ویبول به مهم‌ترین توزیع برای توصیف ویژگی‌های مختلف ترافیک شبکه بدل شد. علاوه بر مفهوم خودهمانندی، دنباله بلند بودن اکثر ویژگی‌های ترافیک شبکه باعث شد تا توزیع ویبول از اهمیت بیشتری برخوردار شود؛ چون این توزیع

1- False Alarm
2- Poisson Distribution

3- Kolmogorov-Smirnov
4- Maximum
5- Packet Length
6- Minimum
7- Max Transmission Unit
8- Fragment
9- Defragmentation
10- Self Similarity

می‌تواند برای برخی از پارامترهای شکل و مقیاس، رفتار دنباله بلند از خود نشان دهد. در منبع [۱۹] نقش توزیع ویبول در مدل‌سازی ترافیک اینترنت توضیح داده شده است.

در منابع [۲۰] و [۲۱] نیز سعی شده است که تاخیر شبکه با استفاده از توزیع‌های ریاضی مدل شود. تاخیرها در شبکه می‌توانند کیفیت سرویس ارائه شده را تحت تاثیر قرار دهند. بنابراین با مدل‌سازی تاخیر شبکه می‌توان عوامل به وجود آورنده آن را شناسایی کرد. در این منابع، تاخیر به وجود آمده بر روی بسته‌های شبکه با بررسی تعداد گره‌های موجود در مسیر بسته مورد بررسی قرار می‌گیرند. توزیع‌های ریاضی استفاده شده برای مدل‌سازی تاخیر شبکه، با بررسی این گره‌ها و تعداد آنها می‌توانند تاخیر شبکه را مدل کنند. در این منابع، توزیع‌های مختلفی برای این مدل‌سازی ارائه شده است و مقدار پارامتر توزیع‌ها برای شرایط مختلف، با مقادیر مختلفی تخمین زده شده است.

استفاده توزیع‌های ریاضی برای مدل‌سازی ترافیک شبکه کارآمد است، ولی این روش محدودیت‌هایی را ارائه می‌کند. برای مثال، اگر ویژگی ترافیک شبکه از توزیع ریاضی خاصی پیروی نکند، قابل مدل‌سازی نیست. همچنین در بحث توزیع‌های ریاضی، تخمین پارامترهای آنها، موضوع مهمی است که می‌تواند از یک مجموعه ترافیک به مجموعه ترافیک دیگر، قابل تغییر باشد. به عبارت دیگر، با مدل کردن یک ویژگی با توزیع ریاضی و مقدار پارامتر خاص، نمی‌توان انتظار داشت که تمام ترافیک‌های دیگر هم از همین توزیع و همین مقدار پارامتر پیروی کنند. بنابراین در این مقاله روش دیگری معرفی می‌شود که بتواند با این محدودیت‌ها مقابله کند.

۴- استفاده از قانون زیف در مدل‌سازی و شبیه‌سازی

در قسمت قبل، انواع مدل‌سازی‌های صورت گرفته بر روی ویژگی‌های مختلف ترافیک شبکه ارائه شده است. در اکثر این تحقیقات، محققان ویژگی‌های ترافیک شبکه را با مدل‌ها و توزیع‌های ریاضی مقایسه کرده و در صورت مطابقت، از آن مدل‌ها به عنوان بهترین ارائه کننده رفتار هنجار ویژگی استفاده می‌کنند. ولی این روش محدودیت‌هایی دارد؛ برای مثال ویژگی‌هایی که از توزیع ریاضی خاصی پیروی نمی‌کنند را نمی‌توان مدل‌سازی کرد. همچنین ممکن است مدل به دست آمده از یک مجموعه ترافیک قابل استفاده در مجموعه ترافیک‌های دیگر نباشد و یا در مقدار پارامتر توزیع، نسبت به یکدیگر متفاوت باشند. همچنین بررسی پیروی توزیع‌های ریاضی از مقادیر یک ویژگی و تخمین پارامترهای آنها، نیازمند استفاده از روش‌های برازشی مانند کولموگروف-اسمیرنوف است که می‌تواند زمان‌بر باشد. در این مقاله برای رفع مشکلات فوق و ارائه روشی ساده و دقیق که بتواند انواع مختلف ویژگی‌های ترافیک شبکه را مدل کند، از قانون زیف استفاده شده است. این قانون با رتبه‌بندی[1] دسته‌های مختلف یک ویژگی، می‌تواند رفتار هنجاری از مقادیر آن ویژگی، ارائه کند.

۴-۱- مجموعه ترافیک مورد استفاده

مجموعه ترافیکی که در این مقاله مورد استفاده قرار گرفته است، مجموعه ترافیک MAWI[2] است[۲۲]. مجموعه ترافیک MAWI، یک مجموعه ترافیک عمومی است که از ترافیک آرشیو شده گروه کاری MAWI استخراج شده است. در این مقاله فقط از بخشی از آن که شامل اطلاعات backbone اقیانوس آرام است و سرعتی برابر با 150Mbps دارد، استفاده شده است.

فایل مجموعه داده، شامل همه بسته‌های IP ارسالی بر روی این لینک، از ساعت ۱۴ تا ۱۴:۱۵ روز ۲۰۱۴/۱۱/۰۱ می‌باشد. برای کاهش حجم داده، فقط یک دقیقه از این مجموعه داده مورد بررسی قرار گرفته است و بسته‌های TCP و UDP را در دو فایل جداگانه، از آن استخراج شده است. دو فایل نهایی، مجموعه بسته‌های TCP و UDP در مدت زمان ۱۴ تا ۱۴:۰۱ از مجموعه داده اصلی است که شامل ۳۲۷۷۴۰ بسته UDP در فایل اول و ۳۷۶۳۸۰۶ بسته TCP در فایل دوم می‌باشد. برای استفاده از این مجموعه ترافیک به عنوان مجموعه ترافیک هنجار، به صورت دستی و با استفاده از ابزارهایی مانند Snort، ترافیک‌های ناهنجار و حمله‌ای از این مجموعه ترافیک حذف شده است. مجموعه ترافیک به دست آمده با اطمینان بالا می‌تواند به عنوان مجموعه ترافیک هنجار استفاده شود.

۴-۲- روش انجام آزمایش

برای انجام آزمایش، ابتدا مجموعه داده‌ای از مقادیر ویژگی مورد نظر ایجاد می‌شود. برای انجام این کار، از مجموعه ترافیک MAWI که اطلاعات کاملی از بسته‌ها را در خود دارد، اطلاعات مورد نیاز جدا می‌شوند. سپس برای مدل‌سازی مجموعه داده به وسیله قانون زیف باید تعداد دسته‌ها و بازه آنها مشخص شود. این بخش مهم‌ترین مرحله در قانون زیف است. اگر تعداد و بازه دسته‌ها به درستی انتخاب شوند، مدل دقیقی به دست خواهد آمد. این موضوع سبب می‌شود که شبیه‌سازی به خوبی انجام شود. پس از تعیین تعداد دسته‌ها، مطابق با رابطه (۴) باید بیشینه و کمینه مقدار مجموعه داده از هم کم شده و در تعداد دسته‌ها تقسیم شوند. با این کار، بازه دسته‌ها مشخص می‌شود.

$$Range = \frac{Max(FreReal) - \min(FreReal)}{Category} \qquad (۴)$$

1- Ranking
2- Measurement and Analysis on the WIDE Internet

در ادامه اجرای فرآیند مدل‌سازی به کمک قانون زیف، فراوانی هر دسته با بررسی مجموعه داده به دست می‌آید. بعد از به دست آمدن فراوانی هر دسته، باید رتبه هر دسته مشخص شود. همان طور که اشاره شد، رتبه هر دسته، تنها نماینده از آن دسته است و به عنوان مدل مورد استفاده قرار می‌گیرد. به منظور تعیین رتبه هر دسته، ۳ گام زیر تعریف شده است:

- **گام اول:** ابتدا فراوانی دسته‌ها از بزرگ به کوچک مرتب شده و به ترتیب، به اولین دسته رتبه یک و به آخرین دسته رتبه‌ای برابر با تعداد دسته‌ها داده می‌شود. دسته‌هایی که فراوانی صفر یا یک دارند، رتبه آنها صفر خواهد بود.
- **گام دوم:** دسته‌هایی که فراوانی یکسانی دارند باید رتبه آنها برابر با رتبه پایین‌تر تنظیم شود. برای مثال، اگر دسته‌های چهارم و پنجم که دارای رتبه ۴ و ۵ هستند، فراوانی یکسانی داشته باشند، رتبه هر دو دسته برابر با ۵ خواهد بود.
- **گام سوم:** همان طور که در تعریف قانون زیف اشاره شد، دسته‌ای که دارای رتبه ۱ است نسبت به دسته‌ای با رتبه ۲، باید دو برابر فراوانی داشته باشد و نسبت به دسته‌ای با رتبه ۳، سه برابر فراوانی خواهد داشت. برای رعایت این موضوع از الگوریتم شکل (۱) استفاده می‌شود. برای این منظور از رتبه‌های به دست آمده از گام دوم استفاده می‌شود. برای این کار بررسی می‌شود که اگر فراوانی دسته c که دارای رتبه k است، کوچک‌تر از فراوانی دسته‌ای که دارای رتبه یک است تقسیم بر رتبه k، و بزرگ‌تر از فراوانی دسته‌ای که دارای رتبه یک است تقسیم بر رتبه k+1 باشد، این رتبه برای این دسته مطابق با قانون زیف مناسب است. در صورت برقرار نبودن این رابطه، در هر مرحله k افزایش یافته تا به رتبه مناسب برسیم.

با محاسبه رتبه هر دسته، مدل مورد نیاز از رفتار هنجار ویژگی ترافیک، به دست آمده است. برای نشان دادن میزان دقت این مدل، فراوانی هر دسته با استفاده از قانون زیف و به کمک رتبه‌های به دست آمده محاسبه شده و نتایج آن با فراوانی‌های هنجار واقعی مقایسه می‌شود. به عبارت دیگر، مقادیر فراوانی محاسبه شده برای هر دسته، با مقادیر فراوانی شمارش شده از ترافیک واقعی مقایسه شده و در صورتی که نتایج، فاصله کمی از هم داشته باشند، می‌توان نتیجه گرفت که مدل‌سازی با دقت بالا انجام شده و قانون زیف برای این منظور کارا است. برای انجام مقایسه، از نمودار فراوانی، به دو صورت ساده و لگاریتمی استفاده می‌شود. همچنین از معیار اختلاف SSE[1] برای نشان دادن اختلاف بین فراوانی واقعی و فراوانی محاسبه شده استفاده می‌شود. این معیار با استفاده از رابطه (۵) محاسبه می‌شود.

$$SSE = \sum_{i=1}^{Categoty} (log_{10}^{FreNew(r_i)} - log_{10}^{FreReal(r_i)})^2 \qquad (۵)$$

```
for ( c=2; c<=Category; c++ )
{
   k = Rank(c);
   Flag = true;
   while ( Flag==true )
   {
      if ( FreReal(c)>(FreReal(1)/k))
         Flag = false;
else if ( (FreReal(c)==floor(FreReal(1)/k)) || (FreReal(c)==ceil(FreReal(1)/k)))
      {
         Rank(c) = k;
         Flag = false;
      }
   else if ( (FreReal(c)<(FreReal(1)/k)) && (FreReal(c)>(FreReal(1)/(k+1))) )
      {
         Rank(c) = (k+1);
         Flag = false;
      }
   }
   k++;
}
```

شکل (۱). گام سوم از فرآیند رتبه‌بندی دسته‌ها در قانون زیف

در این رابطه، Category برابر با تعداد دسته‌های تعریف شده است. همچنین FreNew و FreReal به ترتیب، فراوانی محاسبه شده و فراوانی واقعی را نشان می‌دهند. r_i نیز رتبه دسته i را نشان می‌دهد. برای اثبات این موضوع که قانون زیف، شبیه‌سازی دقیقی ارائه می‌کند، باید مقدار SSE نسبتاً کمی توسط هر دسته ایجاد شود.

برای محاسبه فراوانی دسته‌ها، از الگوریتم شکل (۲) استفاده می‌شود. این الگوریتم به این صورت عمل می‌کند که ابتدا باید تعداد بسته‌های مورد نیاز برای شبیه‌سازی، تعیین شود. در اینجا تعداد بسته‌ها برابر با تعداد بسته‌های واقعی در نظر گرفته می‌شود. سپس برای هر دسته، تعداد بسته‌های مورد نیاز، تقسیم بر رتبه هر دسته، به فراوانی آن دسته اضافه می‌شود. در انتها بررسی می‌شود که آیا تعداد بسته‌های مورد نیاز برای شبیه‌سازی تولید شده است یا خیر. اگر تعداد بسته‌ها به حد کافی نبودند، تعداد بسته‌های باقی مانده را در نظر گرفته و مراحل فوق، مجدداً انجام می‌شود. مشاهده می‌شود که در این کد، فقط رتبه هر دسته برای تولید فراوانی آن دسته مورد استفاده قرار می‌گیرد. مقدار ثابت A هم می‌تواند ۰/۱ در نظر گرفته شود. البته برای دقت بیشتر در مقدار A، طبق رابطه (۶) می‌توان فراوانی واقعی هر دسته را در رتبه آن ضرب و بر تعداد کل بسته‌ها تقسیم کرد و مجموع حاصل از آنها را در تعداد دسته‌های غیر صفر تقسیم کرد. این مقدار نیز می‌تواند مقدار A را تشکیل دهد که همواره عددی نزدیک به ۰/۱ خواهد بود.

$$A = \frac{1}{N \times Category\_NonZero} \times \sum_{i=1}^{Categoty} (FreReal_i \times r_i) \qquad (۶)$$

1- Sum of Squared Errors

```
sum = 0;
Remain = NumPacket - sum;
while ( Remain > 0 )
{
   for (c=1; c<=Category; c++)
   {
      if (Rank(c) == 0)
         FreNew(c) = 0;
      else
      {
         FreNew(c) = FreNew(c) + ceil( (A*Remain) / Rank(c) );
         sum = sum + FreNew(c);
      }
   }
   Remain = RowNum - sum;
}
```

شکل (۲). محاسبه فراوانی دسته‌ها به کمک قانون زیف

۴-۳- نتایج آزمایش‌ها

در این بخش برای هر آزمایش ۲ نمودار نشان داده می‌شود. نمودار اول، فراوانی واقعی و محاسبه شده هر یک از دسته‌ها را نشان می‌دهد. محور عمودی فراوانی هر دسته و محور افقی شماره دسته را نشان می‌دهد. در نمودار دوم، لگاریتم فراوانی واقعی و محاسبه شده هر دسته، به ترتیب رتبه دسته‌ها نشان داده می‌شود. در این نمودار، محور عمودی لگاریتم فراوانی هر دسته و محور افقی لگاریتم رتبه هر دسته را نشان می‌دهد.

در مجموع روی دو ویژگی زمان بین ورود و اندازه بسته‌ها، ۴ آزمایش انجام شده است که جزئیات آنها در جدول (۱) آورده شده است. برای انجام آزمایش‌های مربوط به ویژگی زمان بین ورود بسته‌ها، زمان ورود هر بسته از زمان ورود بسته قبلی کم می‌شود. چون مقادیر این ویژگی بسیار کوچک هستند، برای سهولت در محاسبات، همه مقادیر این ویژگی ضرب در عدد بزرگ ثابتی شده‌اند. نتایج حاصل از این ۴ آزمایش در شکل (۳) نشان داده شده است. آزمایش‌ها نشان می‌دهند که فقط با دانستن رتبه هر دسته و تعیین تعداد بسته‌های کل می‌توان فراوانی هر دسته را محاسبه کرد؛ به طوری که فراوانی‌های محاسبه شده با فراوانی‌های مورد انتظار در ترافیک هنجار واقعی تفاوتی زیادی نداشته باشند. با این روش می‌توان تمامی ویژگی‌های ترافیک شبکه را شبیه‌سازی کرد. برای این کار ابتدا باید مقادیر یک ویژگی را به تعدادی دسته تقسیم کرد و سپس رتبه هر دسته را در ترافیک هنجار طبق قانون زیف محاسبه کرد. بعد از این مرحله مدل مورد نظر از ویژگی به دست آمده، و می‌توان با تعیین تعداد بسته‌های مورد نیاز برای شبیه‌سازی، طبق الگوریتم (۲) فراوانی هر دسته را محاسبه کرد. پس از محاسبه فراوانی‌ها باید به تولید بسته پرداخت. تعداد بسته‌های تولید شده باید متناسب با فراوانی هر دسته و مقدار ویژگی در آن دسته تولید شوند. برای مثال، اگر طبق قانون زیف، فراوانی اندازه بسته‌ها با طول ۷۰-۸۰ بایت در بین ۱۰۰۰ بسته برابر با ۴۰۰ محاسبه شود، باید ۴۰۰ بسته با این ویژگی تولید شود.

۵- ارزیابی

۵-۱- ارزیابی دقیق‌تر آزمایش‌ها در تعداد دسته‌های مختلف

در این مقاله از قانون زیف برای مدل‌سازی و شبیه‌سازی ویژگی‌های مختلف ترافیک شبکه استفاده شد. در انتهای این تحقیقات، روش پیشنهادی مورد ارزیابی دقیق‌تر قرار می‌گیرد. ارزیابی‌ها در سیستمی با پردازنده ۴ هسته‌ای و با ۴ گیگابایت RAM انجام شده است.

به منظور ارزیابی مدل‌سازی، از معیار زمان و SSE استفاده می‌شود. به عبارت دیگر، مدت زمان صرف شده برای مدل‌سازی و مقدار SSE (که اختلاف بین فراوانی‌های محاسبه شده از مدل‌سازی و فراوانی‌های واقعی که مدل از آن استخراج شده است را نشان می‌دهد)، معیارهای ارزیابی مدل‌سازی خواهند بود. بدیهی است که هرچقدر مدت زمان لازم برای مدل‌سازی و مقدار SSE، کمتر باشد، آزمایش مورد نظر عملکرد بهتری داشته است.

جدول (۱). جزئیاتی از ۴ آزمایش انجام شده

ویژگی	تعداد بسته‌ها	کمترین مقدار	بیشترین مقدار	تعداد دسته‌ها	تعداد دسته‌های غیرصفر	مقدار A	SEE
زمان بین ورود بسته‌های TCP	۳۷۶۳۸۰۶	صفر	۴۱۶	۵۰۰۰	۲۶۲	۰/۲۵۶۹	۰/۰۰۲۹
زمان بین ورود بسته‌های UDP	۳۲۷۷۴۰	صفر	۳۵۴۳	۱۰۰۰	۵۰۹	۰/۱۳۸۷	۰/۰۲۳۳
اندازه بسته‌های TCP	۳۷۶۳۸۰۶	۶۰	۱۵۱۴	۱۰۰	۱۰۰	۰/۳۶۲۳	۰/۰۰۰۵
اندازه بسته‌های UDP	۳۲۷۷۴۰	۶۰	۱۵۱۴	۱۰۰	۱۰۰	۰/۱۸۳۶	۰/۰۱۸۰

جدول (۲) خلاصه‌ای از نتایج به دست آمده را نشان می‌دهد. همان طور که گفته شد، تعداد و بازه دسته‌ها نقش مهمی را در مدل‌سازی و شبیه‌سازی در قانون زیف ایفا می‌کنند. برای نشان دادن اهمیت این موضوع، آزمایش‌های مربوط به زمان بین ورود بسته‌ها، در دسته‌های ۱۰۰۰، ۵۰۰۰ و ۱۰۰۰۰ تایی تکرار شد. همچنین برای مدل‌سازی اندازه بسته‌ها، آزمایش آنها در دسته‌های ۵۰، ۱۰۰ و ۲۰۰ تایی مورد بررسی مجدد قرار گرفت. همان طور که مشاهده می‌شود با افزایش تعداد دسته‌ها، مدت زمان مدل‌سازی و مدت زمان محاسبه فراوانی دسته‌ها افزایش می‌یابد ولی تغییر زیادی در مقدار SSE ایجاد نمی‌شود. با توجه به اینکه، مدت زمان مدل‌سازی و مقدار SSE در تعداد دسته‌های کمتر مناسب‌تر است ولی با افزایش تعداد دسته‌ها، اطلاعات بیشتری از وقایع رخ داده در ترافیک شبکه به دست می‌آید. برای مثال، اگر فراوانی اندازه بسته‌های ۸۰ تا ۱۰۰ بایت برابر با ۱۰۰۰ باشد، ممکن است فراوانی اندازه بسته‌های ۸۰ تا ۹۰ بایت برابر با ۹۰۰ و فراوانی اندازه بسته‌های ۹۰ تا ۱۰۰ بایت برابر با ۱۰۰ باشد. بنابراین باید متناسب با نیاز، تعداد دسته‌ها را تعیین کرد. در همه نتایج به دست آمده این موضوع قابل مشاهده است که مقدار SSE (که اختلاف بین فراوانی‌های محاسبه شده از مدل‌سازی و فراوانی‌های واقعی که مدل از آن استخراج شده است را نشان می‌دهد) بسیار کم بوده و این موضوع به این معنی است که قانون زیف در مدل‌سازی و شبیه‌سازی ترافیک شبکه کارآمد است.

جدول (۲). ارزیابی ۴ آزمایش انجام شده با تعداد دسته‌های مختلف

	تعداد دسته‌ها	بازه دسته‌ها	مدت زمان مدل‌سازی(ثانیه)	مدت زمان محاسبه فراوانی‌ها(ثانیه)	مقدار SSE
زمان بین ورود بسته‌های TCP	۱۰۰۰	۰/۴۱۶۰	۵۹/۴۷۶	۰/۰۱۲	۰/۰۰۲۹
	۵۰۰۰	۰/۰۸۳۲	۲۴۵/۰۴۹	۰/۰۳۱	۰/۰۰۲۹
	۱۰۰۰۰	۰/۰۴۱۶	۴۷۶/۷۹۰	۰/۰۴۵	۰/۰۰۲۸
زمان بین ورود بسته‌های UDP	۱۰۰۰	۳/۵۳۴۰	۵/۱۲۵	۰/۰۱۵	۰/۰۲۳۳
	۵۰۰۰	۰/۷۰۸۶	۲۱/۶۹۸	۰/۰۲۸	۰/۰۷۴۸
	۱۰۰۰۰	۰/۳۵۴۳	۴۲/۲۶۷	۰/۰۴۵	۰/۰۷۴۸
اندازه بسته‌های TCP	۵۰	۲۹/۰۸	۷/۶۰۷	۰/۰۱۰	۰/۰۰۰۵
	۱۰۰	۱۴/۵۴	۱۱/۹۷۸	۰/۰۰۹	۰/۰۰۰۵
	۲۰۰	۷/۲۷	۲۰/۸۷۲	۰/۰۱۰	۰/۰۰۰۲
اندازه بسته‌های UDP	۵۰	۲۹/۰۸	۰/۶۴۲	۰/۰۰۹	۰/۰۰۳۰
	۱۰۰	۱۴/۵۴	۰/۸۱۹	۰/۰۱۰	۰/۰۱۸۰
	۲۰۰	۷/۲۷	۱/۲۸۸	۰/۰۰۹	۰/۰۴۰۷

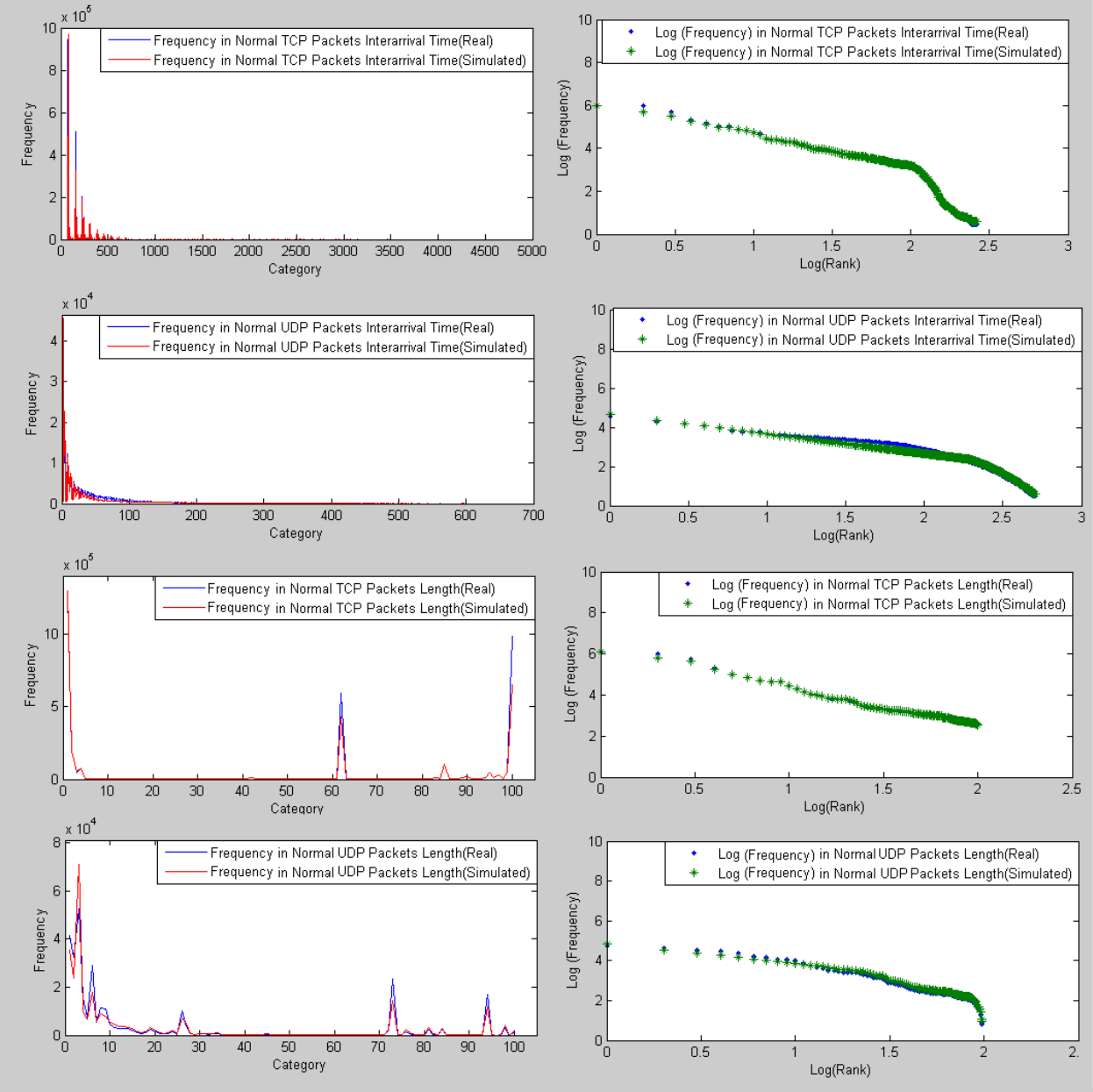
۵-۲- نحوه تعیین تعداد دسته‌ها

مهم‌ترین مرحله در قانون زیف، تعیین درست و دقیق تعداد دسته‌ها است. اگر تعداد دسته‌ها به درستی انتخاب شوند، مدل دقیقی به دست خواهد آمد که باعث می‌شود شبیه‌سازی با عملکرد بالایی انجام شود. تعداد دسته‌ها بر روی سرعت اجرای فرآیندها و دقت نتایج به دست آمده تاثیر می‌گذارند. در ارزیابی‌های انجام شده، به وضوح تاثیر اندازه دسته‌ها بر روی سرعت اجرای فرآیندها و دقت نتایج، نشان داده شده است. برای تعیین تعداد دسته‌ها باید توازنی بین دو شرط افزایش تعداد دسته‌ها و کاهش تعداد دسته‌ها با فراوانی صفر (دسته‌های بی‌معنی) برقرار شود.

افزایش تعداد دسته‌ها باعث می‌شود که حجم محاسبات افزایش یابد و این موضوع باعث می‌شود که فرآیند مدل‌سازی و شبیه‌سازی با سرعت کمتری اجرا شود. همچنین افزایش تعداد دسته‌ها، احتمال به وجود آمدن دسته‌های بی‌معنی و یا با فراوانی صفر را افزایش می‌دهد. برای مثال، بازه ۰/۵ بایت برای اندازه بسته‌ها بی‌معنی است چون اندازه بسته‌ها هیچگاه مقدار اعشاری نخواهند داشت. همچنین اگر بیشینه و کمینه مقدار مجموعه داده فاصله زیادی از هم داشته باشند و داده‌ها به صورت یکنواخت در کل بازه‌ها پخش نشده باشند، تعداد دسته‌های صفر زیادی تولید می‌شود که شاید در مدل‌سازی و شبیه‌سازی نقشی ایفا نکنند. با این حال، افزایش تعداد دسته‌ها، باعث دقت در نتایج به دست آمده می‌شود. برای مثال فرض کنید که فراوانی اندازه بسته‌های ۸۰ تا ۱۰۰ بایت برابر با ۱۰۰۰ باشد. اگر این بازه به دو قسمت تقسیم شود، ممکن است فراوانی اندازه بسته‌های ۸۰ تا ۹۰ بایت برابر با ۸۰۰ و فراوانی اندازه بسته‌های ۹۰ تا ۱۰۰ بایت برابر با ۲۰۰ باشد. این مثال به وضوح تاثیر افزایش تعداد دسته‌ها را در مدل‌سازی نشان می‌دهد. با وجود موارد ذکر شده، افزایش

تعداد دسته‌ها یا کاهش بازه دسته‌ها تا حد امکان، ارجحیت دارد. چون فرآیند مدل‌سازی فقط یک بار انجام می‌شود.

حداقل اندازه بازه دسته‌ها متناسب با نوع ویژگی می‌تواند متفاوت باشد. برای مثال در ویژگی اندازه بسته‌ها، حداقل اندازه بازه دسته‌ها می‌تواند یک بایت در نظر گرفته شود. چون اندازه بسته نمی‌تواند مقدار اعشاری داشته باشد. این در حالی است که در زمان بین ورود بسته‌ها، این مقدار می‌تواند متناسب با دقت و واحد زمان ذخیره‌سازی شده (مثلاً ثانیه یا میلی‌ثانیه)، تا جای ممکن کوچک انتخاب شود.



شکل (۳). نمودار فراوانی واقعی و محاسبه شده برای ۴ آزمایش انجام شده در دو حالت ساده و لگاریتمی

۶- نتیجه‌گیری

در این مقاله به منظور مدل‌سازی رفتار هنجار ترافیک شبکه، از قانون زیف استفاده شد. محققان به منظور مدل‌سازی ترافیک شبکه ابتدا ویژگی‌های مهم آن را شناسایی کرده و در صورت تطابق این ویژگی‌ها با توزیع‌های ریاضی، از آن توزیع‌ها به عنوان مدل رفتار هنجار آن ویژگی استفاده می‌کنند. همان طور که اشاره شد، این روش‌محدودیت‌هایی را ارائه می‌کند که به کمک قانون زیف، این محدودیت‌ها برطرف می‌شوند. در این مقاله نشان داده شد که رتبه دسته‌ها، مدل ارائه شده از رفتار هنجار آن ویژگی است که به وسیله آن می‌توان ویژگی ترافیک شبکه را مجدداً شبیه‌سازی کرد. برای این منظور، اندازه و زمان بین ورود بسته‌های TCP و UDP، مدل‌سازی شدند. همچنین با ارائه

نمودارهایی نشان داده شد که فقط با دانستن رتبه دسته‌ها می‌توان ترافیک شبکه را شبیه‌سازی کرد و شبیه‌سازی‌ها با مقادیر واقعی اختلاف بسیار کمی دارند. مزیت این روش نسبت به روش‌های قبل در این است که هر ویژگی ترافیک شبکه قابل مدل‌سازی خواهد بود و مانند روش‌های قبل به توزیع‌های ریاضی وابسته نیستیم. همچنین چالشی برای تعیین مقدار پارامتر توزیع‌ها وجود ندارد. فقط با تعیین تعدادی دسته و شمارش بسته‌های متعلق به هر دسته و تعیین رتبه دسته‌ها به کمک فراوانی آنها، می‌توان ویژگی‌های مختلف ترافیک شبکه را شبیه‌سازی کرد. در واقع، در این روش، مدل، همان رتبه دسته‌ها است و همان طور که نشان داده شد، فقط با دانستن رتبه هر دسته می‌توان ویژگی ترافیک شبکه را شبیه‌سازی کرد، به طوری که چه از نظر زمانی و چه از نظر دقت، نتایج مناسب خواهد بود. برای مثال آزمایش‌ها نشان می‌دهند که میزان اختلاف بین مدل و ترافیک واقعی همواره کمتر از ۰/۰۱ است و می‌توان با تعیین تعداد دسته‌های مناسب این مقدار را نیز بهبود داد.

۷- مراجع

[1] G. Zipf, "Human behaviour and the principle of least effort: an introduction to human ecology," Linguistic Society of America, vol. 26, no. 3, pp. 394-401, 1949.

[2] A. I. Saichev, Y. Malevergne, and D. sornette, "Theory of Zipf's Law and Beyond," Springer-Verlag Berlin Heidelberg, 2010.

[3] S. Huang, D. Yen, L. Yang, and S. Hua, "An investigation of Zipf's Law for fraud detection," Decision Support Systems, vol. 46, no. 1, pp. 70-83, 2008.

[4] M. Jauhari, A. Saxena, and J. Gautom, "Zipf's Law and Number of hits on the World Wide Web," Annals of Library and Information Studies, vol. 54, no. 2, pp. 81-84, 2007.

[5] L. Adamic and B. Huberman, "Zipf's Law and the Internet," Glottometrics 3, pp. 143-150, 2002.

[6] B. R. Chang and H. F. Tsai, "Improving network traffic analysis by foreseeing datapacket- flow with hybrid fuzzy-based model prediction," Expert Systems with Applications, vol. 36, no. 3, pp. 6960-6965, 2009.

[7] J. Sommers and P. Barford, "Self-Configuring Network Traffic Generation," in Proceedings of ACM Internet Measurement Conference, Italy, pp. 68-81, 2004.

[8] S. Guadagno, D. Emma, A. Pescap and N. Federico, "D-ITG Distributed Internet Traffic Generator," in Proceedings, First International Conference on the Quantitative Evaluation of Systems, Netherlands, pp. 316-317, 2004.

[9] "tcpreplay-Tcpreplay," [Online].Available: http://tcpreplay.synfin.net/wiki/tcpreplay, 2015.

[10] "Network, devices & services testing - Spirent," [Online]. Available: http://www.spirent.com/. [Accessed 24- 07-2015].

[11] W. M. Shbair, A. R. Bashandy, and S. I. Shaheen, "A New Security Mechanism to Perform Traffic," in International Conference on Computational Science and Engineering, pp. 405-411, 2009.

[12] F. Sally and P. Vern, "Difficulties in simulating the internet," IEEE/ACM Trans. Netw., vol. 9, no. 4, pp. 392-403, 2001.

[13] V. Paxson, "Strategies for sound internet measurement," in Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, pp. 264-271, 2004.

[14] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: a survey," ACM Computing Surveys, vol. 41, no. 3, pp. 1-58, 2009.

[15] S. Luo and G. A. Marin, "Generating Realistic Network Traffic for Security Experiments," in Proceedings of the IEEE Southeast Conf., North Carolina, USA, pp. 200-207, 2004.

[16] E. Garsva, N. Paulauskas, G. Grazulevicius, and L. Gulbinovic, "Packet Inter-arrival Time Distribution in Academic Computer Network," Elektronika IR Elektrotechnika, vol. 20, no. 3, pp. 87-90, 2014.

[17] M. Fras, J. Mohorko, and Z. Cucej, "Packet Size Process Modeling of Measured Self-similar Network Traffic with Defragmentation Method," in Proceedings of the 15th International Conference on Systems, Signals and Image Processing, Bratislava, pp. 253-256, 2008.

[18] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "the Self-Similar Nature of Ethernet Traffic (extended version)," IEEE/ACM Transactions on Networking, vol. 2, no. 1, pp. 1-15, 1994.

[19] M. A. Arfeen, K. Pawlikowski, D. McNickle, and A. Willig, "The Role of the Weibull Distribution in Internet Traffic Modeling," in 25th International Conference Teletraffic Congress (ITC), Shanghai, pp. 1-8, 2013.

[20] L. R. Dominguez, D. O. Roman, D. M. Rodriguez, and C. V. Rosales, "Jitter in IP Networks: A Cauchy Approach," IEEE Communications Letters, vol. 14, no. 2, pp. 190-192, 2010.

[21] T. K. Bandhopadhya, M. Saxena, and A. Tiwari, "Jitter's Alpha Stable Distribution Behavior," Computer Technology and Electronics Engineering (IJCTEE), vol. 3, no. 1, pp. 13-16, 2013.

[22] "MAWI Working Group Traffic Archive," [Online]. Available: http://mawi.wide.ad.jp/mawi/. [Accessed 24-07-2015].

Journal Of Electronical & Cyber Defence
Vol. 4, No. 3, 2016, Serial No.15

A Method for Modeling and Generating Normal Network Traffic Based on the Features of Length and Arrival Time of Packets Using the Zipf's Law

A. Nghash-Asadi, M. Abdollahi Azgomi*

*Iran University of Science and Technology

(Received: 03/02/2016, Accepted: 01/08/2016)

ABSTRACT

Today, modeling and generating normal network traffic is a very important. In existing works, the features of network traffic are modeled using probabilistic distributions. In this paper, a new method is proposed for modeling the features of network traffic. The proposed method is based on the Zipf's law. The Zipf's law is an empirical law that provides the relationship between the frequency and rank of each category in data set. In this paper, we will show that the Zipf's law can model different features of network traffic in a good manner. For this propose, two important features of network traffic, i.e., length and inter-arrival time of TCP and UDP packets, are examined. The proposed method for modeling the features of network traffic can use in various applications areas, such as, simulation or generation of the normal network traffic. The advantage of this law is that it can provide high similarity using less information. Furthermore, the Zipf's law can model different features of network traffic that may not follow from probalistic distributions. The simple approach of this law can provide accuracy and lower limits from existing methods. Furthermore, the proposed method can provide good times for modeling and simulation.

In this paper, we will show that by classifying the feature values and obtaining their ranks, we can create an accurate modeling of features. In other words, the rank of each category will be the model resulting from the feature values that can be used in simulation.

Keywords: Network Traffic, the Zipf's Law, Network Modeling, Network Simulation, Traffic Generation

* Corresponding Author Email: azgomi@iust.ac.ir