دوفصل‌نامه علمی منادی امنیت فضای تولید و تبادل اطلاعات (افتا)
سال ۱۴۰۰، جلد ۱۹، شماره ۲ (۶۱-۷۰)
https://monadi.isc.org.ir

ارائه‌شده در هفدهمین کنفرانس بین المللی انجمن رمز ایران در دانشگاه علم و صنعت ایران، ۱۹ و ۲۰ شهریور ۱۳۹۹

ارتقا حریم خصوصی ترافیک شبکه در برابر حمله‌ی دسته‌بندی به کمک یادگیری خصمانه★

محمدرضا کریمی* و رسول جلیلی

دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف، تهران، ایران

اطلاعات مقاله

کلمات کلیدی:
دسته‌بندی ترافیک
مبهم‌نگاری ترافیک
یادگیری عمیق
یادگیری ماشین
شبکه‌ی عصبی
شبکه‌ی عصبی همگشتی

dor: 10.0000/000000000

نوع مقاله: پژوهشی

چکیده

در چند سال اخیر بطور گسترده از معماری‌های مختلف شبکه‌های عصبی عمیق در ادبیات پژوهش‌های دسته‌بندی ترافیک و انگشت‌نگاری وب‌سایت استفاده شده است. دسته‌بندی‌های یاد شده بر روی ویژگی‌های آماری ترافیک مانند طول بسته‌ها و فاصله‌ی زمانی بین بسته‌ها صورت می‌پذیرد. ارتقا حریم خصوصی ترافیک شبکه در برابر حملات دسته‌بندی با الگوریتم‌هایی صورت می‌پذیرد که ترکیبی از افزایش طول (لایه‌گذاری) و شکستن بسته‌ها و اضافه کردن تاخیر در ارسال بسته را انجام می‌دهند. در این پژوهش به جای طراحی چنین الگوریتم‌هایی، با استفاده از روش‌های سنجش و ارزیابی مقاومت شبکه‌های عصبی موسوم به الگوریتم‌های تولید نمونه‌ی خصمانه با اعمال حداقل سربار اقدام به لایه‌گذاری بسته‌های جریان ترافیک شده است. یک دسته‌بند شبکه‌ی عصبی عمیق همگشتی را قبل و بعد از اعمال دفاع بر روی ترافیک، به کمک پنج الگوریتم تولید نمونه‌ی خصمانه، کارلینی-ونگر، جِی.اس.ام.ای، اف.جی.اس.ام، دیپ‌فول و پریشیدگی سراسری، ارزیابی می‌کنیم. هر یک از الگوریتم‌ها با اضافه کردن میزان سربار متفاوت، از دقت و مثبت کاذب دسته‌بندی شبکه عصبی یاد شده می‌کاهند.

۱ مقدمه

کاربران اینترنت همواره سعی در حفظ و ارتقا سطح محرمانگی خود دارند که یکی از مصادیق این محرمانگی مربوط به رفتار ترافیکی کاربران است. در نقطه مقابل دولت‌ها و شرکت‌های ارائه خدمات اینترنت به جهت کیفیت سرویس، تعرفه گذاری بر حسب نوع ترافیک و مهار محتوای ناهنجار سعی در شناسایی ترافیک کاربران و حتی انتساب ترافیک به کاربر، دارند. در طی زمان، شناسایی ترافیک از شناسایی با شماره درگاه و فیلدهای پروتکل‌های لایه‌ی شبکه و انتقال، به شناسایی با وارسی عمیق بسته یعنی تطبیق با الگوی رشته‌ای تا سطح لایه کاربرد، به شناسایی به کمک یادگیری ماشین و شبکه‌ی عمیق عصبی ارتقا یافته‌اند. روش‌های شناسایی ترافیک در دو شکل دسته‌بندی ترافیک و انگشت‌نگاری صورت می‌پذیرند. دسته‌بندی ترافیک عبارت است از وظیفه انتساب صحیح برنامه کاربردی یا پروتکلی که جریان ترافیک به آن مربوط است. انگشت‌نگاری وب‌سایت به انتساب صفحات وب بازدید شده توسط کاربر به جریان ترافیک تونل شده آن وب‌سایت در پروتکل تُر یا ابزارهای درهم‌سازی ترافیک مربوط می‌شود.

همزمان با توسعه‌ی روش‌های شناسایی ترافیک، روش‌های متعدد جلوگیری از دسته‌بندی ترافیک به منظور افزایش سطح محرمانگی طراحی شده‌اند که هر کدام با روشی سعی در از بین بردن امضاها و انگشت‌نگاره‌های شناساننده ترافیک دارند. بطور کلی این روش‌ها به سه دسته‌ی روش‌های مبهم‌نگاری، شبکه‌های گمنامی و شبکه‌های بازشکن

★از کمیته علمی هفدهمین کنفرانس بین‌المللی انجمن رمز ایران برای داوری این مقاله تشکر می‌شود.
*نویسنده مسئول
آدرس‌های رایانامه: mokarimi@ce.sharif.edu (محمدرضا کریمی)، jalili@sharif.edu (رسول جلیلی)

تقسیم می‌شوند. دسته‌ی مبهم‌نگاری خود عبارت است از تصادفی‌سازی، مانندسازی و تونل‌سازی. هر کدام از این دسته‌ها مشتمل بر مقالات متعددی است. امضا و انگشت‌نگاری شناسای پروتکل‌ها را می‌توان به دو انگشت‌نگاره‌ی محتوایی و انگشت‌نگاره‌ی آماری تقسیم کرد. انگشت‌نگاره‌های محتوایی مربوط به الگوهای محتوایی است که در بسته‌های پروتکل مشاهده می‌شود با روش‌هایی مانند رمزنگاری و یا تونل کردن امحا می‌گردند. اما انگشت‌نگاره‌های آماری به توزیع آماری ویژگی‌های ترافیک مانند طول بسته‌ها و فاصله‌ی زمانی بین بسته مربوط می‌شود که با لایه‌گذاری [۱] بسته‌ها، شکستن بسته‌ها، اضافه کردن بسته‌های جعلی [۲] و تاخیر در ارسال بسته، از بین می‌روند. تجربه‌ی سیر پژوهش‌ها نشان داد که برای پوشاندن انگشت‌نگاره‌های محتوایی باید به رمزنگاری اکتفا کرد، آنچه که همه‌ی انواع روش‌های جلوگیری از شناسایی ترافیک، چه روش‌های جلوگیری از دسته‌بندی ترافیک، چه روش‌های جلوگیری از انگشت‌نگاری وب‌سایت همگی به آن آسیب‌پذیر بوده‌اند، باقی ماندن انگشت‌نگاری آماری است، لذا پژوهش‌های اخیر در جلوگیری از شناسایی ترافیک یا دفاع از حریم خصوصی ترافیک همگی مبتنی بر امحای انگشت‌نگاری آماری ترافیک هستند.

انگشت‌نگاری‌های آماری به قدری که مورد توجه پژوهش‌های انگشت‌نگاری وب‌سایت قرارگرفته‌اند، مورد توجه پژوهش‌های دسته‌بندی وب‌سایت نبوده‌اند. یکی از اهداف این پژوهش، توجه بیشتر به ویژگی‌های آماری است. ویژگی آماری مورد توجه در دسته‌بندی فاصله زمانی بین بسته‌ها و اندازه‌ی بسته‌های جریان است. در مورد دفاع از حریم خصوصی ترافیک نیز به لایه‌گذاری بسته‌ها اکتفا می‌شود، چرا که تاخیر در ارسال بسته ممکن است موجب افت کیفیت ارتباط به خصوص در کاربردهایی مانند وُیپ شود؛ لذا با محدود کردن سربار به سربار پهنای باند، از سربار تاخیر اجتناب شده است. بر خلاف رویه رایج در طراحی یک شمای دفاع در برابر دسته‌بندی ترافیک یا یک شمای ضدانگشت‌نگاری، که تابع قواعدی از پیش تعیین شده در شکستن و لایه‌گذاری بسته، اضافه کردن بسته‌های جعلی و تاخیر در ارسال بسته‌ها است، نوآوری این پژوهش در توجه به نقطه ضعف شبکه‌های عصبی عمیق در دسته‌بندی داده‌هایی موسوم به نمونه خصمانه است. نمونه خصمانه داده‌ای است، که از اضافه شدن نویز طراحی شده‌ای موسوم به پریشیدگی [۳]، به یک داده خام اصلی که سابقاً درست دسته‌بندی می‌شده به وجود می‌آید. اضافه شدن این نویز به داده اصلی موجب به خطا انداختن شبکه‌ی عصبی عمیق در دسته‌بندی می‌شود. با الهام از کاربردهای موفق روش‌های تولید نمونه‌ی خصمانه در حوزه صدا و تصویر و متن، برای لایه‌گذاری بسته‌های جریان ترافیک از الگوریتم‌های تولید نمونه‌ی خصمانه بهره گرفته شده است.

ارزیابی این پژوهش در دو بخش دسته‌بندی ترافیک، به عنوان حمله به حریم خصوصی ترافیک و تولید نمونه‌ی خصمانه از جریان ترافیک، به عنوان دفاع از حریم خصوصی تقسیم می‌شود. دادگانی که به جهت ارزیابی مورد استفاده قرارگرفته است، دادگان وی.پی.اِن– غیر وی.پی.اِن آی.اِس.سی.ایکس ۲۰۱۶ [۱] است که مشتمل بر شش دسته‌ی چت، انتقال فایل، جریانی، وُیپ، همتا به همتا و رایانامه است. شبکه‌ی عصبی عمیق همگشتی [۴] دسته‌بند بر روی دو ویژگی طول بسته و فاصله‌ی زمانی بین ۲۵۰ بسته‌ی اول هر جریان به دقت ۸۰٫۰۰ درصد و مثبت واقعی به ۸۰٫۳۳ درصد رسید. پنج الگوریتم تولید نمونه‌ی خصمانه نسبت به داده‌ی آزمون شبکه دسته‌بند اعمال می‌شود، یعنی بسته‌های جریان مطابق الگوریتم لایه گذاری می‌شوند، مدل دسته‌بند با جریان‌های لایه‌گذاری شده ارزیابی می‌شود. الگوریتم اِف.جی.اِس.اِم [۲] با تحمیل ۱۶٫۳۱ درصد سربار پهنای باند موفق به کاهش ۶٫۱ درصدی دقت، الگوریتم کارلینی–ونگر ال۲ [۳] با تحمیل ۵۰٫۱۷ درصد سربار پهنای باند موفق به کاهش ۲۷٫۷۷ درصدی دقت ، الگوریتم جِی.اِس.اِم.اِی [۴] با تحمیل ۷۶٫۸۱ درصد پهنای باند موفق به کاهش ۷۶٫۱۱ درصدی دقت، الگوریتم دیپ‌فول [۵] با تحمیل ۱۰۷٫۸۹ درصدی به کاهش ۷۳٫۰۶ درصدی دقت و الگوریتم پریشیدگی سراسری [۶] با تحمل ۱۱۸٫۵۶ درصد سربار پهنای باند موفق به کاهش ۶۸٫۰۶ درصدی دقت می‌شود.

۲ مرور کارهای پیشین

دامنه‌ی مرور و مقایسه کارهای پیشین این پژوهش شامل مقالاتی است که از یادگیری عمیق برای ارتقا سطح حریم خصوصی ترافیک بهره جسته‌اند. پژوهش پیش رو مربوط به دفاع از دسته‌بندی ترافیک به کمک تولید نمونه خصمانه است، از پژوهش‌های مشابه این پژوهش تنها پژوهش وِرما و همکاران [۷] مربوط دسته‌بندی ترافیک است ولی اثر ایمانی و همکاران [۸] که موسوم به ماکینگ‌بِرد می‌باشد و همچنین لی و همکاران [۹] مربوط به دفاع از دسته‌بندی در برابر انگشت‌نگاری وب‌سایت است. نتایج ارزیابی روش‌های ارائه‌شده در دو مقاله‌ی لی و همکاران و ایمانی و همکاران که مربوط به دادگان انگشت‌نگاری وب‌سایت است، با دادگان این پژوهش که دادگان انواع ترافیک شبکه است، قابل مقایسه نیست، چرا که نه تنها با دادگان یکسانی آزمون و ارزیابی نمی‌شوند بلکه جنس دادگان نیز کاملا متفاوت است. مقاله‌ی وِرما و همکاران نیز که بر روی دادگان عمومی شبکه ارزیابی شده است، ارزیابی شفاف و درستی ارائه نکرده است.

وِرما و همکاران برای دسته‌بندی و دفاع در برابر دسته‌بندی از ویژگی فاصله زمانی بین بسته‌ها و اندازه بسته‌ها از ۲۰ بسته اول هر جریان، ویژگی‌هایی را مشتق می‌کنند و دادگان آزمون از ترافیک یک لینک گیگابیت دانشگاهی از ۶ دسته‌ی وب، حجمی، پایگاه داده، رایانامه، سرویس‌ها، همتا به همتا جمع‌آوری شده است. شبکه‌ی دسته‌بندی ترافیک را از یک شبکه‌ی عصبی کاملا متصل سه لایه تشکیل می‌دهد و بر روی دادگان یاد شده با ویژگی‌های یاد شده به دقت ۹۷ درصد می‌رسد. برای مبهم‌سازی ترافیک صرفا الگوریتم تولید نمونه‌ی خصمانه‌ی کارلینی–ونگر ال۲ را بر روی ترافیک اعمال می‌کند. پس از اعمال کارلینی–ونگر بر روی ویژگی‌های یاد شده در دسته‌ی حجمی ۱۱ درصد، در دسته‌ی پایگاه

[1] padding [2] dummy [3] perturbation

[4] convolution

داده ۱ درصد، در دسته‌ی رایانامه ۲۲ درصد، در دسته‌ی سرویس‌ها ۴۸ درصد و در دسته‌ی همتا به همتا ۳۹ درصد، در دسته‌ی وب ۱۳ درصد کاهش دقت گزارش می‌کند. پیاده‌سازی و ارزیابی این پژوهش در موارد متعددی مبهم و نادرست است. تعداد بسته‌های لازم برای دسته‌بندی هر جریان یا به عبارت دیگر اندازه ورودی شبکه‌ی عصبی ذکر نشده است، این عدد به شکل خوبی توان دسته‌بندی شبکه‌ی عصبی را نشان می‌دهد. هر قدر تعداد بسته‌های مورد نیاز برای دسته‌بندی جریان کمتر باشد یعنی توان شناسایی و دسته‌بندی شبکه بیشتر است. در ارزیابی روش ارائه‌شده اشاره‌ای به سربار پهنای باند و سربار تاخیر وجود ندارد چرا که ارزیابی موفقیت روش دفاعی، با گزارش میزان کاهش دقت، بدون اشاره به سربار تحمیل شده میسر نیست. روش ارزیابی روش دفاعی صحیح نیست، زیرا با وجود اینکه در داده آموزش و آموزن، داده‌های دسته‌ها متوازن نیستند برای دادگان ارزیابی روش دفاعی از هر دسته به تعداد برابر و تصادفی انتخاب می‌کند، روش صحیح ارزیابی این است که ارزیابی روش دفاعی بر روی همان داده آزمونی صورت پذیرد که ارزیابی مدل دسته‌بندی بر آن صورت پذیرفته است، لذا از جهت ارزیابی این پژوهش به کلی غیر قابل استناد است. اما مهمترین اشکال وارد بر این اثر در پیاده سازی روش دفاعی است. پریشیدگی که الگوریتم‌های تولید نمونه خصمانه به یک بردار ویژگی اضافه می‌کنند ممکن است منفی یا مثبت باشد و یا خارج از دامنه‌ی معنادار اعداد بردار ویژگی نباشد. اگر مقداری که به ویژگی زمان بردار اضافه می‌شود منفی باشد، یعنی باید در ارسال بسته تعجیل کرد که تا زمانی که بسته از سمت لایه‌ی کاربردی آماده نباشد، بسته‌ای برای تعجیل در ارسال وجود ندارد و اگر مقداری که به ویژگی طول بسته‌ی بردار اضافه شود منفی باشد، به این معناست که باید از طول بسته ارسالی کاست و بسته را بصورت ناقص ارسال کرد؛ ولی در این مقاله هیچ روشی برای کنترل پریشیدگی افزوده‌شده بر داده‌ی خام ارائه نمی‌کند.

لی و همکاران به کمک شبکه‌های مولد خصمانه روشی ارائه می‌دهند که ویژگی‌های آماری جریان ترافیک مسدود شده و غیر مجاز را به گونه‌ای دگریخت می‌سازد که به ترافیک نامسدود و مجاز دسته‌بندی شود، این دگریخت‌سازی را استتار پویا می‌نامند و به این شبکه‌ی مولد خصمانه‌ی دگریخت‌ساز، را فلوگن می‌گویند. رویکرد این روش ناظر به دسته‌بندی انواع ترافیک است ولی برای ارزیابی روش خود از ترافیک وب‌سایت بهره می‌برد. ضمن ارائه روش دفاعی فلوگن، معیار سنجش جدید ϵ-تمییزناپذیری، را برای سنجش موفقیت روش‌های دفاع و پوششی ویژگی ترافیک ارائه می‌دهد. برای آموزش فلوگن از شش ویژگی مجموع بسته‌های خروجی، مجموع بسته‌های ورودی، مجموعه بایت‌های خروجی، مجموعه بایت‌های ورودی، بایت‌های تجمعی و میانگین فاصله زمانی بین بسته‌ها استفاده می‌کند. برای ارزیابی توان دفاعی فلوگن، آن را نسبت به هشت روش انگشت‌نگاری وب‌سایت مجهز به یادگیری ماشین برای دو معیار ϵ-تمییزناپذیری و اِی.یو.سی می‌آزماید. فلوگن روش خود را با مهاجم‌های دسته‌بند ضعیف مجهز به الگوریتم‌های یادگیری ماشین متداول، ارزیابی می‌کند، در حالیکه روش‌های قدرتمندتر مجهز به یادگیری عمیق مثل دیپ فینگرپرینتینگ [۱۰] را در ارزیابی دخیل نمی‌کند. این موضوع نشان از ضعف ارزیابی و یا ضعف روش پیشنهادی دارد که در برابر دسته‌بندی‌های با دقت بالاتر نتیجه خوبی نمی‌گیرد. این مقاله با اینکه در حوزه‌ی ضد انگشت‌نگاری وب‌سایت است، ولی با معیارهای کارائی این حوزه، یعنی دقت، سربار تاخیر و پهنای باند ارزیابی نشده است. این مقاله معیار ارزیابی ابداعی و جدید ϵ-تمییزناپذیری، را ارائه می‌کند و تمییزناپذیری دفاع ترافیک را برای دسته‌بندهای مختلف اندازه می‌گیرد و از معیار دیگری برای ارزیابی استفاده نمی‌کند.

ایمانی و همکاران در اثر خود موسوم به ماکینگ‌بِرد، با تغییر اندکی در تابع بهینه‌سازی الگوریتم کارلینی-ونگر که برای عکس‌های دو بعدی ارائه شده بود آن را برای رَد[1] های تک بعدی ترافیک تبدیل می‌کند. به جهت مدل انگشت‌نگاری از شبکه‌ی عصبی عمیق دیپ فینگرپرینتینگ استفاده می‌کند. حتی به جهت دادگان نیز از همان دادگان مقاله‌ی دیپ فینگرپرینتینگ ،با اندکی پردازش پسین، استفاده کرده است اما ورودی شبکه‌ی دیپ فینگرپرینتینگ یعنی تعداد بسته در هر جریان، ۵۰۰۰ است که در دسته‌بندی ماکینگ‌بِرد به ۱۰۰۰۰ افزایش می‌یابد. ایده اصلی این است که هر رَد دلخواه ،رَد مبدا، تا حد ممکن به یک رَد در دسته‌ای متفاوت نگاشت شود که این رَد را، رَد هدف می‌نامیم. به این منظور، تا جایی که امکان دارد رَد اولیه به رَد هدف نزدیک می‌شود. فرض کنید یک مجموعه سایت S داریم که می‌خواهیم از آن در مقابل دسته‌بند انگشت‌نگاری f که روی رَد سایت‌های مجموعه S نیز آموزش داده شده، حفاظت کنیم. رَد I_s نمونه‌ای از رَدهای متعلق دسته‌ی سایت s از مجموعه S، را می‌خواهیم طوری تغییر دهیم که به دسته‌ی هدف t، دسته‌بندی شود بطوریکه، $f(I_s) = t$ و $t \neq s$. از جنبه‌های مختلفی ارزیابی مفصلی در مقایسه با آثار پیش از خود دارد و دقت دیپ فینگرپرینتینگ را با ۵۶ درصد سربار از ۹۸ درصد به ۳۵ درصد می‌رساند.

مقاله‌ی دیپ فینگرپرینتینگ برای دسته‌بندی جریان‌های ترافیک وب، ۵۰۰۰ بسته در هر جریان را مشاهده می‌کند در حالیکه این مقاله ۱۰۰۰۰ بسته در هر جریان را پیشنهاد می‌دهد، افزایش تعداد بسته منجر به تقویت مهاجم نشده است و دقت آن را افزایش نداده است. افزایش تعداد بسته مورد نیاز برای مشاهده در هر جریان بدون افزایش دقت دسته‌بند، مهاجم را تضعیف می‌کند، چرا که هر چه مهاجم با دیدن تعداد بسته‌ی کمتر، با دقت بالاتری دسته‌بندی کند، دسته‌بند قدرتمندتری است و بالعکس. این ارزیابی از یک سو، با افزایش تعداد بسته مورد نیاز برای مشاهده در هر جریان مهاجم را تضعیف کرده است و از سوی دیگر دو برابر کردن تعداد بسته، سربار را به صورت صوری و عددی کاهش می‌دهد، چرا که مجموع اندازه‌ی بسته‌ها مخرج فرمول محاسبه سربار را افزایش می‌دهد در حالیکه که صورت کسر افزایش چندانی نمی‌یابد، زیرا اگر دسته‌بند روی ۵۰۰۰ بسته‌ی اول جریان به همان دقتی دست یافته که روی ۱۰۰۰۰ بسته اول دست می‌یابد، یعنی دسته‌بند در دسته‌بندی توجه کمتری به ۵۰۰۰ بسته دوم می‌کند؛ لذا روش دفاعی نیز کمتر ۵۰۰۰ بسته دوم را تغییر می‌دهند و لذا افزودن ۵۰۰۰ بسته بیشتر برای دسته‌بندی، فقط به کاهش صوری

[1] trace

سربار کمک می‌کند.

۳ مفاهیم پایه

به بعضی از مفاهیم پایه مورد نیاز برای فهم این پژوهش در مقدمه پرداخته شد. در این بخش به جهت اختصار به توضیح نمونه‌ی خصمانه بسنده می‌شود. قریب به اتفاق الگوریتم‌های تولید نمونه خصمانه برای داده تصویر و وظیفه دسته‌بندی تصویر ارائه شده‌اند، لذا در اینجا نیز با وظیفه دسته‌بندی تصویر توضیح داده می‌شود ولی منحصر به این کاربرد نمی‌باشد.

برای وظیفه دسته‌بندی تصویر، با دسترسی به یک مدل دسته‌بند تصویر منتشر شده توسط یک شخص ثالث، کاربر یک تصویر دلخواه به دسته‌بند می‌دهد و برچسب پیش‌بینی دسته‌بند برای آن تصویر را دریافت می‌کند. با اضافه کردن پریشیدگی‌هایی به تصویر اصلی، تصاویر خصمانه تولید می‌کند که تشخیص تصویر خصمانه از تصویر اصلی توسط انسان قابل تشخیص نیست و دسته‌بند آن را به غیر از برچسب اصلی تصویر، برچسب می‌زند. با داشتن دسته‌بند تصویر $f(\cdot)$ و نمونه داده اصلی x، تولید تصویر خصمانه x' به عنوان یک مسئله بهینه‌سازی توصیف می‌شود:

$$\min_{x'} \|x' - x\| \quad \text{s.t.} \quad f(x') = l',\ f(x) = l,\ l \neq l',\ x' \in [0, 1], \tag{۱}$$

که l و l' به ترتیب برچسب برای x و x' هستند. $\|\cdot\|$ فاصله‌ی بین دو نمونه را نشان می‌دهد. $\eta = x' - x$ همان مقدار پریشیدگی است که به تصویر اصلی اضافه می‌شود. در این مسئله بهینه‌سازی، هدف کمینه‌سازی پریشیدگی برای به حداقل رساندن فاصله و تفاوت تصویر اصلی و تصویر خصمانه است تا امکان تمییز ناظر انسانی را به حداقل برساند.

۴ مدل تهدید

عامل مدافع D می‌خواهد از ترافیک شبکه T متعلق به برنامه‌ی کاربردی یا پروتکل P را در برابر حمله‌ی استنتاج یا دسته‌بندی که از شبکه مهاجم A عبور می‌کند، حفاظت کند. عامل مدافع D با تغییر T به T^* سعی می‌کند، مهاجم A را فریب دهد تا نتواند به سهولت و صحت ترافیک T^* را دسته‌بندی کند. اگر مدافع D موفق باشد، مهاجم A استنتاج می‌کند که ترافیک T^* متعلق به دسته‌بندی Q که متفاوت از دسته‌بندی P است، تعلق دارد. ترافیک T قبل از ورود به شبکه A از شبکه D عبور می‌کند.

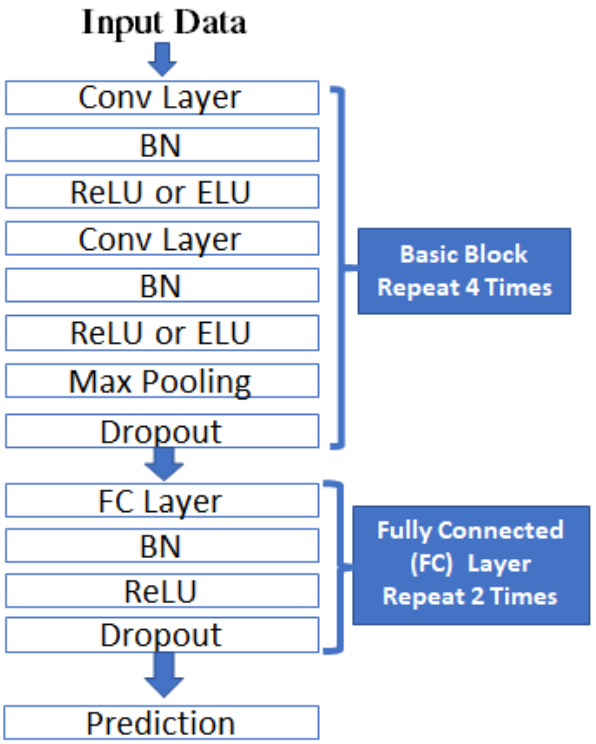
عامل مهاجم در این پژوهش منطبق بر یک شبکه‌ی عصبی عمیق همگشتی است که پروتکل هر جریان ترافیک را برچسب می‌زند. هدف این پژوهش نه ارائه‌ی یک روش دسته‌بندی یا مدل مهاجم به ترافیک، بلکه هدف، ارائه طرح یک مدل مدافع از حریم خصوصی به منظور کاهش دقت و جلوگیری از دسته‌بندی ترافیک است. چنانچه پیشتر در مقدمه اشاره شد، عامل مدافع نیز مبتنی بر یک الگوریتم تولید نمونه‌ی خصمانه است که جریان‌های ترافیک را به گونه‌ای تغییر می‌دهد تا دقت عامل مهاجم را به حداقل رسانند. بهبود و نوآوری در عامل مهاجم و دسته‌بند ترافیک هدف نیست، اما مدل مدافع هم به جهت راه‌اندازی و ایجاد، نیازمند دسترسی به مدل مهاجم است و هم از جهت ارزیابی، دفاع باید با حمله‌ی موثر ارزیابی شود. در ادامه ابتدا به توضیح طراحی مهاجم و سپس به بیان طراحی مدافع پرداخته می‌شود.

۵ طراحی مدل مهاجم

مهاجم دسته‌بند ترافیک یک شبکه‌ی عصبی همگشتی است که طراحی آن شبکه از مقاله‌ای موسوم دیپ فینگرپرینتینگ [۱۰]، هم‌نام با آن شبکه، گرفته شده است. معماری دیپ فینگرپرینتینگ الهام گرفته از شبکه‌ی وی.جی.جی، رِزنت و گوگل‌نت است. این شبکه از دو بخش استخراج ویژگی و دسته‌بندی تشکیل می‌شود. وظیفه استخراج ویژگی مربوط به شبکه‌های همگشتی بلوکی است. بخش بلوکی شبکه حاصل از تکرار بلوک‌هایی است که چندین لایه از شبکه‌ی همگشتی مانند لایه‌ی همگشتی، لایه‌ی ادغام(پولینگ) و لایه فعال‌سازی را در بر دارند. بعد از شبکه‌ی همگشتی، یک شبکه‌ی بلوکی کاملا متصل قرار می‌گیرد که وظیفه دسته‌بندی را بر عهده دارد. شمایی از شبکه‌ی دیپ فینگرپرینتینگ در شکل ۱ آمده است. ریزدانگی دسته‌بندی ترافیک در این پژوهش در حد دسته‌بندی جریان است، شبکه با دیدن چند بسته‌ی اول هر جریان دسته‌ی مربوط به آن جریان را تعیین می‌کند. ویژگی‌های مورد استفاده برای دسته‌بندی دو ویژگی آماری طول بسته و فاصله‌ی زمانی بین ارسال بسته‌هاست. هر بسته به صورت یک بردار ⟨اندازه‌ی بسته، فاصله‌ی زمانی با بسته‌ی قبلی⟩ بازنمائی می‌شود. از درکنار هم قرارگرفتن حداکثر تعداد بسته‌ی مورد نیاز برای مشاهده در هر جریان ویژگی‌های مورد نیاز برای دسته‌بندی اندازه‌ی ورودی شبکه تعیین می‌شود. تعداد دسته‌های ترافیک نیز اندازه‌ی خروجی شبکه را تعیین می‌کند.

۶ طراحی مدل مدافع

مدل دفاعی که از ترافیک در برابر دسته‌بندی دفاع می‌کند از منظرگاه یادگیری ماشین خصمانه، یک حمله به مدل دسته‌بندی تلقی می‌شود. مدل دفاعی این پژوهش از پنج حمله‌ی جعبه سفید، کارلینی–ونگر [۳]، جِی.اس.ام.ای [۴]، اف.جی.اس.ام [۲]، دیپ‌فول [۵] و پریشیدگی سراسری [۶] به کاهش دقت دسته‌بندی شبکه‌ی عصبی ترافیک استفاده می‌کند. حمله‌ی جعبه سفید نیازمند دسترسی کامل به مدل دسته‌بند است، این دسترسی شامل دسترسی به مدل و تمام پارامترهای مربوط به آن است، لذا برای ساخت یک مدل مدافع باید یک مدل مهاجم یا همان دسته‌بند ترافیک را بر روی بخشی از ترافیک آموزش داد، سپس از آن مدل مهاجم یک مدل مدافع ساخت. مدل مدافع یاد می‌گیرد برای افزودن پریشیدگی کمینه با حداکثر تاثیر چطور باید برای جریان‌هایی که حتی مدل مهاجم روی آن‌ها آموزش داده نشده است، پریشیدگی تولید کند. مدل مدافع، پریشیدگی را بر روی جریان ترافیک تولید می‌کند و نسبت به بردارهای ⟨اندازه‌ی بسته، فاصله‌ی زمانی با بسته‌ی قبلی⟩ اعمال می‌کند.



شکل ۱. نمایی از معماری شبکه‌ی دسته‌بند ترافیک [۱۰].

الگوریتم‌های تولید نمونه‌ی خصمانه با هوشمندی بالایی ابعادی از بردار که بیشتر مورد توجه دسته‌بند قرار می‌گیرد را شناسایی می‌کنند، با افزودن کمترین مقدار نویز لازم، بیشترین کاهش دقت را رقم می‌زنند. پس از اعمال پریشیدگی به هر جریان، آن جریان توسط مدل مهاجم دسته‌بند، به هر جریانی غیر از دسته‌ی واقعی خود دسته‌بندی می‌شود و برچسب می‌خورد. پریشیدگی افزوده شده، بصورت مقادیر عددی مثبت و منفی به بعد اول و دوم بردار اعمال می‌شود که هر کدام تفسیر بخصوصی دارد. افزوده شدن پریشیدگی مثبت به اندازه بسته به معنی افزایش طول بسته یعنی نیاز به لایه‌گذاری است. افزوده شدن پریشیدگی مثبت به فاصله‌ی زمانی بین دو بسته‌ی متوالی، به معنی تاخیر در ارسال بسته دوم است. افزودن پریشیدگی منفی به اندازه‌ی بسته، به معنی کوچک کردن اندازه‌ی بسته، یعنی ارسال بسته‌ی ناقص است، که ممکن نیست. افزودن پریشیدگی منفی به فاصله‌ی زمانی بین دو بسته نیز به معنی تعجیل در ارسال بسته است که همانطور که در بخش مرور کارهای پیشین در نقد ارزیابی کار ورِما و همکاران [۷]، اشاره شد، ممکن نیست، لذا باید نسبت به داده‌ی پریشیده شده کنترل صورت گیرد، از همین رو پریشیدگی‌های منفی صفر گرفته می‌شود. کنترل پریشیدگی زمانی نیز نسبت به کاربردهایی همچون نمایش‌های جریانی و وُیپ دشوار است، زیرا اعمال پریشیدگی مستقل از نوع ترافیک به آن اعمال می‌شود و ممکن است منجر به کاهش کیفیت انواع ترافیک جریانی و وُیپ گردد، لذا از اضافه کردن پریشیدگی زمانی نیز امتناع می‌شود. تنها پریشیدگی که به بردار اضافه می‌شود، پریشیدگی مثبت به بعد اندازه بسته است، ولی این پریشیدگی نیز باید به گونه‌ای باشد که اندازه بسته پس از افزوده شدن آن پریشیدگی کمتر از مقدار

جدول ۱. برچسب انواع ترافیک در دادگان آی.اِس.سی.ایکس وی.پی.اِن غیر وی.پی.اِن [۱].

Traffic Type	Content
Email	Email、Gmail (SMPT, POP3,IMAP)
VPN-Email	
Chat	ICQ, AIM, Skype, Facebook, Hangouts
VPN-Chat	
Streaming	Vimeo, Youtube, Netflix, Spotify
VPN-Streaming	
File transfer	Skype, FTPS, SFTP
VPN-File transfer	
VoIP	Facebook, Skype, Hangouts, Voipbuster
VPN-VoIP	
P2P	uTorrent, Bittorrent
VPN-P2P	

حداکثر واحد انتقال شبکه‌ی اترنت، یعنی ۱۵۰۰ بایت باشد. در حالی که مدل دفاعی محدود به اعمال پریشیدگی به بعد اندازه‌ی بسته شده است، مدل مهاجم، دسته‌بندی را روی هر دو بعد اندازه‌ی بسته و فاصله‌ی زمانی بین دو بسته اعمال می‌کند، پس مهاجم امکانات بیشتری از مدافع دارد و مهاجم موثری است، چرا که مدافع با تغییر یک بعد از جریان ترافیک حفاظت می‌کند ولی مهاجم هر دو بعد بسته‌های جریان را تحلیل می‌کند.

۷ ارزیابی

قبل از آموزش مدل مهاجم باید دادگان آموزش، پاکسازی شود، همچنین داده آزمون نیز باید پاکسازی شده باشد. جلوتر ابتدا به بیان روش پاکسازی دادگان پرداخته می‌شود. سپس مدل مهاجم و مدل مدافع مورد ارزیابی قرار می‌گیرد.

۱.۷ دادگان

در این پژوهش از دادگان آی.اِس.سی.ایکس وی.پی.اِن و غیر وی.پی.اِن [۱] که توسط دانشگاه یو.اِن.بی ارائه شده است، استفاده می‌شود. این دادگان دارای هفت نوع ترافیک رمزشده‌ی عادی و هفت ترافیک کپسوله‌شده در یک پروتکل وی.پی.اِن است. این دادگان بصورت فایل خام ترافیک یعنی فایل پیکپ [۱] و فایل ویژگی جریان‌ها است. انواع ترافیک موجود در این دادگان به همراه نوع محتوای آن، در جدول ۱ آمده است.

دادگان یاد شده بالغ بر ۳۶۰ هزار جریان است که عمده این جریان‌ها، جریان‌هایی هستند که تنها شامل دو بسته هستند. تعداد جریان‌های دو بسته‌ای بالغ بر ۲۵۴ هزار می‌باشد که عمده مربوط به ترافیک وُیپ است. جریان‌های دو بسته‌ای ارزشی برای یادگیری شبکه ندارند، زیرا وجود آن‌ها، دادگان را به شدت نامتوازن ساخته و نشان دهنده‌ی یک اتصال ناقص هستند. از این جهت، جریان‌های با کمتر ۳ بسته حذف می‌شوند. جریان‌های دی.اِن.اِس درون هر فایل پیکپ، نباید به برنامه‌ی کاربردی و فعالیت آن‌ها برچسب بخورد، چون پروتکل دی.اِن.اِس رفتار پروتکلی مستقلی دارد، لذا جریان‌های پروتکل دی.اِن.اِس نیز پالایش می‌شوند. پس از اعمال پالایش‌های گفته‌شده، تعداد غیر قابل توجهی جریان، از

[1] PCAP

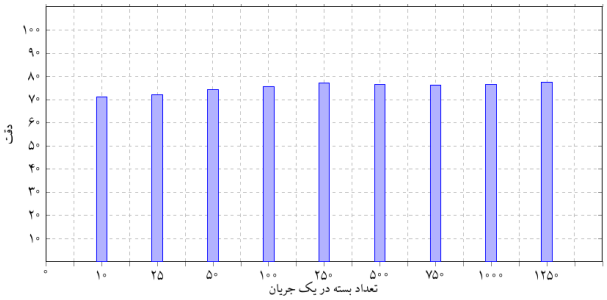
دسته‌ی پروتکل تُر می‌ماند که وجود آن توازن تعداد جریان را بین دسته‌های مختلف بر هم می‌زند. در نهایت، تعداد جریان‌ها از ۳۶۰ هزار جریان، به حدود ۱۸ هزار جریان می‌رسد، که جریان‌های باقیمانده مشتمل بر شش دسته‌ی چت، انتقال فایل، جریانی، وُیپ، همتا به همتا و رایانامه است که همین دسته‌ها نیز نامتوازن هستند. برای برقراری توازن از هر دسته ۶۰۰ جریان برگزیده می‌شود که در مجموع ۳۶۰۰ جریان برای آموزش، آزمون و وارسی باقی می‌ماند که به ترتیب ۸۰ ، ۱۰ و ۱۰ درصد کل داده را تشکیل می‌دهند.

۲.۷ ارزیابی مدل مهاجم

همانطورکه پیشتر نیز اشاره شد، مدل مهاجم یک شبکه‌ی عصبی همگشتی است. برای ارزیابی کارایی یک شبکه‌ی عصبی، معیار دقت، شاخص‌ترین معیار ارزیابی است. به علت کوچک بودن دادگان و واریانس بالا در ارزیابی‌های متعدد از دقت مدل یکسان، اعتبارسنجی متقابل ضرورت می‌یابد. اعتبارسنجی مورد استفاده نوع خاصی از اعتبارسنجی موسوم به اعتبارسنجی متقابل طبقه طبقه است که بخش‌بندی دادگان طوری صورت می‌گیرد که نسبت انواع دادگان در هر بخش تا جای ممکن مشابه نسبت همان برچسب‌ها در دادگان اصلی باشد.

اعتبار سنجی متقابل[1] علاوه بر سنجش میزان تعمیم یافتگی و پیش‌بینی‌گری مدل، به جهت انتخاب مدل نیز، کاربرد دارد. در اینجا، برای انتخاب بهترین مدل، مدل‌ها با ورودی‌های مختلف، با روش اعتبارسنجی متقابل جدا جدا[2] ، ارزیابی می‌شود. در شکل ۲ میزان دقت مدل دسته‌بند بر حسب تعداد بسته‌های در هر جریان آورده شده است. تعداد بسته در جریان بیان می‌کند که برای دسته‌بندی جریان‌ها حداکثر از چند بسته‌ی اول جریان استفاده شده است. بهترین مدل، مدل با اندازه‌ی ورودی ۲۵۰ است، چرا که دقت با ۲۵۰ بسته در جریان نسبت به دسته‌بندهای قبل از خودش، یعنی دسته‌بندهای با کمتر از ۲۵۰ بسته در هر جریان، با اختلاف قابل توجهی، حدود ۵ درصد دقت بیشتری دارد ولی نسبت به دسته‌بند با ۱۲۵۰ بسته در جریان، حدود ۱ درصد دقت کمتر دارد. از آنجا که کارایی دسته‌بند تنها وابسته به دقت نیست و مهم است که دسته‌بند، با چه تعداد بسته می‌تواند عملکرد خوبی داشته باشد، مدل با ورودی ۲۵۰ بسته در هر جریان انتخاب می‌شود، چون در یک بهینه محلی نسبت به ورودی‌های با اندازه‌ی کمتر و بیشتر است؛ در عین حال با بیشینه تعداد بسته نیز تفاوت دقت چندانی ندارد. پایین بودن دقت به دست آمده برای مدل، قابل توجه نیست و این به مساله‌ی محدودیت و کوچک بودن دادگان مربوط می‌شود. دقت بر روی ۲۵۰ بسته اول ترافیک به ۸۰ درصد می‌رسد. مزیت بارز این روش دسته‌بندی نسبت به روش‌های مشابه که تنها روش‌های انگشت‌نگاری را شامل می‌شود، در کمتر بودن چشمگیر تعداد بسته در جریان است. همین تعداد ۲۵۰ بسته ابتدائی هر جریان برای دسته‌بندی و دفاع در این مقاله، در مقاله‌ی ایمانی و همکاران موسوم به ماکینگ‌بِرد به ۱۰۰۰۰ بسته اول هر جریان می‌رسد. کمتر بودن

[1] cross-validation [2] stratified cross-validation



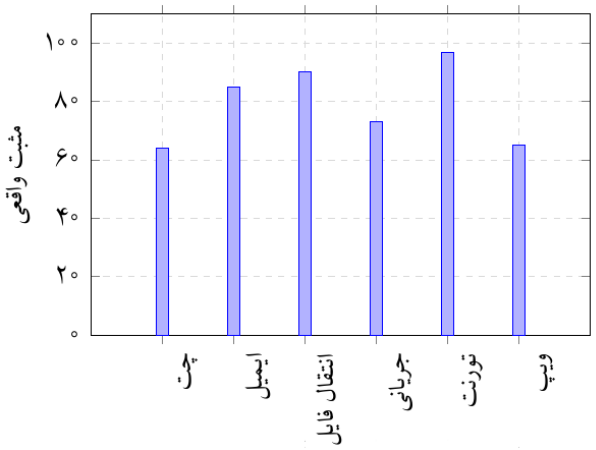
شکل ۲. دقت مدل مهاجم دسته‌بند بر حسب تعداد بسته در هر جریان.

تعداد بسته برای دسته‌بندی نه تنها مزیت مدل مهاجم است که با دیدن تعداد کمتری بسته می‌تواند برچسب جریان را تعیین کند، بلکه مزیت مدل دفاعی نیز هست، زیرا به تغییر تعداد بسته کمتری نیاز دارد.

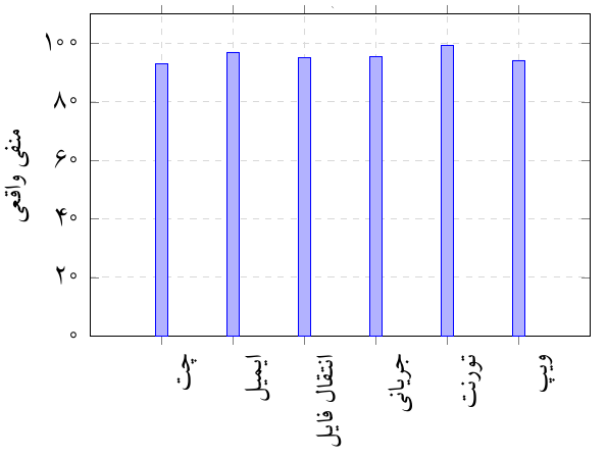
۳.۷ ارزیابی مدل مدافع

مدل دفاعی باید با همان دادگان آزمونی که مدل مهاجم دسته‌بند بخش قبل با آن ارزیابی شده، آزموده شود. همان دادگان مطابق مدل دفاعی و تابع قواعد با افزودن پریشیدگی تغییر داده شد و مجدداً با همان شبکه‌ی عصبی آموزش دیده‌ی مدل مهاجم ارزیابی می‌شود. به پنج روش مختلف که هر کدام از یکی از الگوریتم‌های تولید نمونه‌ی خصمانه استفاده می‌کند، جریان‌های دفاع‌شده از دادگان آزمون تولید می‌شود. الگوریتم‌های مختلف با سربارهای مختلفی، به میزان‌های مختلفی دقت دسته‌بند را کاهش دادند. جدول ۱ دقت دسته‌بند برای ترافیک اصلی و بدون تغییر و ترافیک دفاع شده با هر پنج الگوریتم یاد شده را مقایسه می‌کند. دقت دسته‌بند مدل مهاجم، پس از اعمال دفاع‌های مختلف بدون کنترل قواعد، همچون منفی نبود پریشیدگی اندازه بسته و کمتر ماندن اندازه‌ی بسته از ۱۵۰۰ بایت پس از اعمال پریشیدگی، در دومین ستون جدول آمده است، که در مورد الگوریتم جِی.اِس.اِم.اِی و دیپ‌فول دقت از ۸۰ درصد به زیر ۱۰ درصد می‌رسد. اما پس از اعمال قواعد کنترلی تاثیر کاهش دقت بسیار کم می‌شود. علت این امر آن است که پریشیدگی در این الگوریتم‌ها با توجه به هر دو بعد بردار اعمال می‌شود، لذا وقتی یک بعد بردار پریشیدگی صفر می‌شود و نسبت به منفی بودن پریشیدگی بردار دیگر اعمال فیلتر می‌شود، اثر کاهش دقت پریشیدگی کم می‌شود، ولی همچنان مقدار کاهش دقت صورت گرفته قابل توجه است.

اثر دفاع از ترافیک این است که ترافیک هر دسته به اشتباه به دسته‌ی دیگری دسته‌بندی شود و یا به عبارت دیگر مثبت واقعی آن دسته را کم کند، از این روست که تغییرات این معیار در جدول مورد بررسی قرار گرفته است. پریشیدگی سراسری، یک پریشیدگی عمومی یکسان برای همه‌ی جریان‌ها تولید می‌کند ولی دیگر روش‌ها پریشیدگی موردی تولید می‌کنند. سراسری بودن تولید پریشیدگی، کارایی درهم‌سازی الگوی آماری ترافیک را افزایش می‌دهد، زیرا پریشیدگی سراسری یک بار بر روی کل دادگان آزمون محاسبه می‌شود و یک پریشیدگی واحد تولید می‌کند که افزودن آن به هر جریان از دادگان ترافیک آزمون، باعث درهم‌سازی ترافیک و



شکل ۳. مقایسه مثبت واقعی دسته‌بند برای دسته‌های مختلف ترافیک.



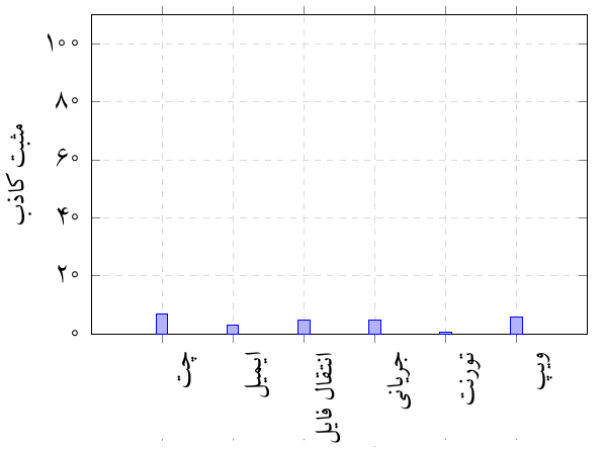
شکل ۴. مقایسه منفی واقعی دسته‌بند برای دسته‌های مختلف ترافیک.

فریب دسته‌بند می‌شود.

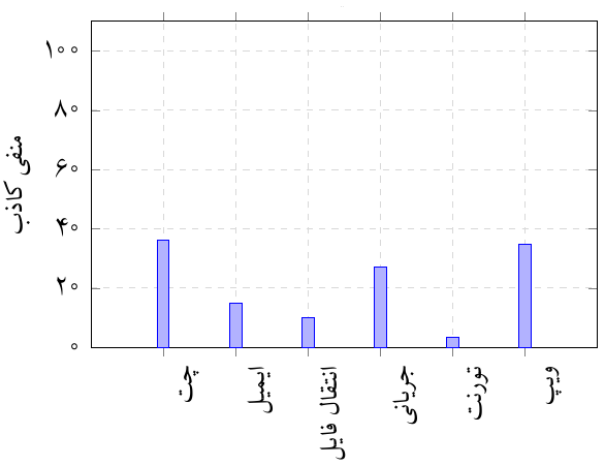
با اینکه نوع دادگان در انگشت‌نگاری وب‌سایت از نوع دادگان در دسته‌بندی ترافیک متفاوت است، لذا نتایج ارزیابی این پژوهش را تا حدودی با مقاله‌ی ایمانی و همکاران غیر قابل مقایسه می‌سازد، ولی کاهش دقتی که دفاع این پژوهش با ۲۵۰ بسته‌ی اول به دست می‌آورد با کاهش دقتی که مقاله‌ی ایمانی بر روی ۱۰۰۰۰ بسته‌ی اول بدست می‌آورد کاملاً قابل مقایسه است. مقاله‌ی ایمانی با ۵۶ درصد سربار، دقت را ۶۳ درصد کاهش می‌دهد، مشابه آن، نتیجه دفاع مبتنی بر جی.اِس.اِم.اِی است که با ۷۶٫۸۱ درصد سربار دقت را ۵۶٫۱۱ کاهش می‌دهد، با دفاع مبتنی بر اِف.جی.اِس.اِم، با ۱۶٫۳۱ درصد سربار دقت را ۱۵٫۵۵ درصد کاهش می‌یابد و دفاع مبتنی بر کارلینی-ونگر ال۲ با تحمیل ۵۰٫۱۷ درصد سربار، دقت را ۲۷٫۷۷ درصد می‌کاهد.

۸ نتیجه‌گیری

در این پژوهش سعی گردید کاراترین روش دسته‌بندی برای ارزیابی روش دفاعی ارائه شود، روشی که از آن بهره برده شد، مبتنی بر حوزه‌ی یادگیری ماشین بود. کارهای صورت گرفته در حوزه یادگیری ماشین به جهت



شکل ۵. مقایسه مثبت کاذب دسته‌بند برای دسته‌های مختلف ترافیک.



شکل ۶. مقایسه منفی کاذب دسته‌بند برای دسته‌های مختلف ترافیک.

جدول ۲. جدول مقایسه‌ی سربار، دقت و مثبت واقعی دسته‌بند در برابر ترافیک اصلی و ترافیک دفاع شده.

حمله	درصد دقت دسته‌بند	درصد دقت دسته‌بند (پریشیدگی کنترل شده)	درصد مثبت واقعی	درصد سربار
بدون حمله	۸۰٫۰	۸۰٫۰	۸۰٫۳۳	–
اِف.جی.اِس.اِم	۵۶٫۱۱	۶۴٫۴۵	۷۳٫۹۰	۱۶٫۳۱
کارلینی-ونگر ال۲	۳۹٫۴۴	۵۲٫۲۳	۶۳٫۹۸	۵۰٫۱۷
جی.اس.ام.ای	۵٫۲۷	۲۳٫۸۹	۳۹٫۰۰	۷۶٫۸۱
دیپ فول	۹٫۱	۲۶٫۹۴	۳۵٫۶۲	۱۰۷٫۸۹
پریشیدگی سراسری	۱۶٫۶۷	۳۱٫۹۴	–	۱۱۸٫۵۶

دفاع از حریم خصوصی ترافیک، انگشت‌شمار است. پس از توضیح مساله دسته‌بندی ترافیک و دفاع از دسته‌بندی ترافیک، ابتدا به مرور این روش‌ها و نواقص موجود در آنها پرداخته شد. سپس راهکار ارائه‌شده در سه بخش مدل تهدید، مدل مدافع و مدل مهاجم ارائه شد. از شبکه‌ی عصبی دیپ فینگرپرینتینگ که برای انگشت‌نگاری وب‌سایت ارائه شده بود برای ارزیابی روش دفاعی استفاده شد. هدف این مقاله ارائه‌ی یک روش دفاعی برای جلوگیری از دسته‌بندی ترافیک بود، لذا شبکه‌ی عصبی دیپ فینگرپرینتینگ به ترافیک شبکه آموزش داده شد. نتایج این پژوهش با کارهای مشابه به دلائلی که گفته شد، قابل مقایسه نیست. چون تنها روش دفاعی مبتنی بر یادگیری ماشین در حوزه دسته‌بندی ترافیک است که آمار و ارزیابی کاملی از آن صورت گرفته است. از امتیازات دیگر این

پژوهش این است که دادگان مورد استفاده، پرارجاع‌ترین دادگان برای دسته‌بندی ترافیک در بین روش‌های دسته‌بندی مبتنی بر یادگیری ماشین است. همچنین در بیان ارزیابی، روش‌های دفاعی مختلف را در دقت، مثبت واقعی و سربار مقایسه کرده است که این شفافیت در ارزیابی با کارهای پیشین قابل مقایسه، مطرح نبوده است.

۹ کارهای آتی

مبنای روش دفاعی ارائه شده در این پژوهش، ریشه در میزان موفقیت حملات یادگیری خصمانه دارد که از این حملات به مثابه دفاع در برابر دسته‌بند، به تعبیری دیگر یک ضد حمله در برابر حمله‌ی دیگر، استفاده شده است؛ لذا کارایی و میزان موفقیت این روش دفاعی به طور مستقیم وابسته به کارایی حملات یادگیری خصمانه است. نقطه مقابل این حملات، دفاع و روش‌های تشخیصی نیز در ادبیات یادگیری ماشین خصمانه ارائه شده است. تقویت دسته‌بند ترافیک به روش‌های دفاعی در برابر حملات یادگیری خصمانه، پایداری و امنیت شبکه را ارتقا می‌دهد. با توجه به وجود روش‌های افزایش پایداری و امنیت شبکه، لازم می‌آید که به عنوان کار آتی ابتدا، دسته‌بند ترافیک را پایدار و مقاوم سازیم و سپس، کارایی روش دفاعی پوشش ترافیک و میزان پایداری دسته‌بند را ارزیابی کنیم.

بکارگیری و افزودن روش‌های دفاعی موجود که مبتنی بر درهم‌سازی محتوایی، تغییر در ساختار پروتکل و اضافه کردن بسته جعلی به ترافیک هستند، مستلزم توافق و همکاری دو طرف ارتباط در برقرار اتصال و تبادل داده است. به این معنا که، منطق روش دفاعی باید بین دو سر ارتباط پیاده‌سازی شده باشد. از طرف دیگر، شکستن بسته‌ها، بر هم زدن الگوی زمانی بسته‌ها از طریق تاخیر، تعجیل و جابجایی معقول، تغییراتی هستند که مدیریت آن تغییر در بطن پروتکل‌های شبکه پیاده سازی شده است. به کمک این دسته از تغییرات می‌توان، روش‌های دفاعی یک طرفه‌ای ارائه کرد که بدون اطلاع طرف دیگر ارتباط و تنها با پیاده سازی در یک طرف ارتباط می‌تواند دسته‌بند را فریب دهد.

۱۰ سپاس‌گزاری

پژوهش پیشِ رو برگرفته از پروژه کارشناسی ارشد نویسنده، تحت نظر جناب آقای مهندس صادق‌زاده است. بر خود واجب می‌دانم زحمات این بزرگوار را خاطر نشان کرده و از توجهشان ممنون و متشکر باشم.

مراجع

[1] Gerard Draper-Gil, Arash Habibi Lashkari, Mohammad Saiful Islam Mamun, and Ali A. Ghorbani. Characterization of encrypted and VPN traffic using time-related features. In Olivier Camp, Steven Furnell, and Paolo Mori, editors, *Proceedings of the 2nd International Conference on Information Systems Security and Privacy, ICISSP 2016, Rome, Italy, February 19-21, 2016*, pages 407–414. SciTePress, 2016.

[2] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[3] Nicholas Carlini and David A. Wagner. Towards evaluating the robustness of neural networks. In *2017 IEEE Symposium on Security and Privacy, SP 2017, San Jose, CA, USA, May 22-26, 2017*, pages 39–57. IEEE Computer Society, 2017.

[4] Nicolas Papernot, Patrick D. McDaniel, Somesh Jha, Matt Fredrikson, Z. Berkay Celik, and Ananthram Swami. The limitations of deep learning in adversarial settings. In *IEEE European Symposium on Security and Privacy, EuroS&P 2016, Saarbrücken, Germany, March 21-24, 2016*, pages 372–387. IEEE, 2016.

[5] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. Deepfool: A simple and accurate method to fool deep neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 2574–2582. IEEE Computer Society, 2016.

[6] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. Universal adversarial perturbations. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 86–94. IEEE Computer Society, 2017.

[7] Gunjan Verma, Ertugrul N. Ciftcioglu, Ryan Sheatsley, Kevin S. Chan, and Lisa Scott. Network traffic obfuscation: An adversarial machine learning approach. In *2018 IEEE Military Communications Conference, MILCOM 2018, Los Angeles, CA, USA, October 29-31, 2018*, pages 1–6. IEEE, 2018.

[8] Mohsen Imani, Mohammad Saidur Rahman, Nate Mathews, Aneesh Yogesh Joshi, and Matthew Wright. Advdwf: Defending against deep-learning-based website fingerprinting attacks with adversarial traces. *CoRR*, abs/1902.06626, 2019.

[9] Jie Li, Lu Zhou, Huaxin Li, Lu Yan, and Haojin Zhu. Dynamic traffic feature camouflaging via generative adversarial networks. In *7th IEEE Conference on Communications and Network Security, CNS 2019, Washington, DC, USA, June 10-12, 2019*, pages 268–276. IEEE, 2019.

[10] Payap Sirinam, Mohsen Imani, Marc Juárez, and Matthew Wright. Deep fingerprinting: Undermining website fingerprinting defenses with deep learning. In David Lie, Mohammad Mannan, Michael Backes, and XiaoFeng Wang, editors, *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS 2018, Toronto, ON, Canada, October 15-19, 2018*, pages 1928–1943. ACM, 2018.