شـــرکت ارتبـاطات زیرساخت وزارت ارتباطات و فناوری اطلاعات	مقاله عنوان: مروری بر تحلیل، اهمیت و تکنیک‌های ترافیک شبکه تهیه کننده/گان: رشته تحصیلی رشته شغلی اداره کل /دفتر تاتینا صدیقی کارشناس مهندسی کامپیوتر نرم افزار فناوری اطلاعات طرح و توسعه شبکه/کارشناس طراحی شبکه دیتا شادی نوری کارشناسی ارشد کامپیوتر نرم افزار فناوری اطلاعات طرح و توسعه شبکه/رئیس اداره طراحی سوئیچ و دیتای کشوری
این قسمت توسط دبیرخانه کمیته علمی تکمیل می‌گردد.	شماره مقاله: ۵۷ حوزه کاربردی: بررسی و تحلیل ترافیک شبکه (با استفاده از ابزارهای نرم افزاری)

چکیده

از زمان انتشار برنامه tcpdump در سال ۱۹۸۸، داده‌های ترافیک شبکه برای تصمیم گیری در مورد امنیت شبکه، جمع آوری، تحلیل مورد استفاده قرار می گیرند. همزمان با تکامل فناوری ها، روش های مختلفی برای تحلیل ترافیک شبکه استفاده می شود. برجسته‌ترین روش ها عبارتند از تکنیک‌های استخراج داده، تکنیک‌های آماری و تکنیک‌های مصورسازی. این موارد در بخش‌های مختلف این مقاله بررسی می شوند. تکنیک‌های یادگیری ماشین نقش شایانی در تحلیل ترافیک شبکه دارند . این فرآیند تحلیل مبنای امنیت شبکه بوده و مسئولیت مهمی برای مجریان شبکه محسوب می‌شود. در انتهای مقاله، خلاصه مقالات بررسی شده ارائه می‌شود.

I. مقدمه

در زمان اتصال به اینترنت، موقعیت های مشترکی برای بسیاری از افراد رخ می دهد: «چرا سرعت شبکه این‌قدر کم است؟»، «چرا نمی‌توانم به ایمیلم دسترسی داشته باشم؟»، «چرا به درایو اشتراکی دسترسی ندارم؟»، «چرا کامپیوتر خوب کار نمی‌کند؟». این سوالات معمولا به مجریان سیستم، مهندسان شبکه و یا مهندسان امنیتی ارجاع داده می شوند. در ادامه، ادمین های سیستم با فرآیند پر درد سر و خسته کننده حل مشکل روبرو می‌شوند. وظیفه اصلی ادمین شبکه در این گونه موقعیت‌ها، تحلیل ترافیک شبکه است. تحلیل ترافیک شبکه به فرآیند دربرگیری ترافیک شبکه و بررسی دقیق آن به منظور تعیین رخدادهای درون شبکه مربوط می شود. بسته‌های داده توسط تحلیل گر شبکه رمزگشایی شده و ترافیک شبکه به شکلی خوانا به نمایش گذاشته می شود. تحلیل شبکه در صورت استفاده پیرامون داده های پیشین، به راهبردی احتمال‌گرا تبدیل شده و مانع از وقوع مجدد نقص های مختلف در امنیت شبکه می شود. تحلیل ترافیک شبکه برای افزایش قدرت زیرساخت شبکه اهمیت زیادی دارد. این کار با شناسایی و جلوگیری از تراکم شبکه و همچنین شناسایی بسته‌های نرمال و مخرب انجام می‌شود. در عین حال، تحلیل شبکه به عنوان شمشیری دو لبه عمل می کند. از یک سو، متخصصان شبکه، سیستم و امنیت از این ابزار برای حل مشکل و نظارت بر شبکه استفاده می کنند، ولی از سوی دیگر، خرابکاران از تحلیل شبکه برای اهدافی مخرب از آن بهره می برند. تحلیل‌گر شبکه نیز مشابه با تم‌امی دیگر ابزارها برای اهداف خوب و مخرب استفاده می شود. تحلیل شبکه در این موارد کاربرد دارد : تبدیل داده‌های دوگانه در بسته‌ها به فرمتی خوانا، حل مشکلات در شبکه، تحلیل عملکرد شبکه برای کشف تنگناها، شناسایی ورود غیرمجاز به شبکه، ثبت ترافیک شبکه برای جمع‌آوری شواهد، تحلیل عملیات‌ها، کشف کارت‌های شبکه اشتباه، کشف م نشا شیوع ویروس یا حملات منع سرویس [۱] (DoS)،

---

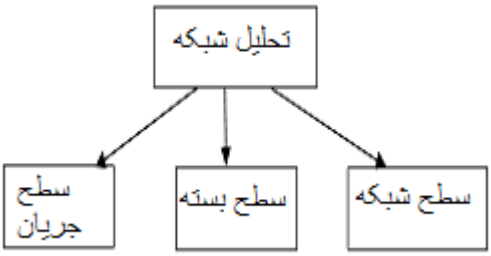
[۱] Denial of Service

شناسایی جاسوس افزارها، برنامه ریزی شبکه برای رفع عیب در مرحله توسعه شبکه، شناسایی کامپیوتر مشکل دار، اعتبارسنجی تبعیت از سیاستگذاری شرکت، فعالیت به عنوان منبع آموزشی پروتکل‌ها، مهندسی معکوس پروتکل‌ها برای نگارش برنامه‌های مشتریان و پشتیبان (۱).

II. پیشینه

شبکه در سطوح مختلفی تحلیل می‌شود: سطح بسته، سطح جریان و سطح شبکه برای مدیریت امنیت (۲).

شکل ۱: انواع تحلیل شبکه



ویژگی‌های مورد نظر در تحلیل سطح جریان عبارتند از مدت جریان، حجم داده، تعداد بسته ها در هر بسته جریان و باقی موارد . ویژگی‌های تحلیل سطح بسته عبارتند از طول بسته، میانگین و واریانس طول بسته، میانگین مربع ریشه و باقی موارد . برای سطح شبکه، ویژگی‌هایی مثل نام میزبان و نام سرور اهمیت دارند. این مسائل در جدول ۱ ارائه شده‌اند.

جدول ۱: مجموعه ویژگی‌ها در سطوح مختلف شبکه

سطح	پروتکل	مجموعه ویژگی
جریان		مدت جریان، حجم داده، تعداد بسته‌ها در هر جریان
	TCP	مقدار بسته‌های مورد نیاز در جریان، زمان رسیدن بسته، تعداد درگاه‌های استفاده کننده از بسته‌ها
	IP	مقصد IP، موقعیت مکانی IP، عدد سیستم مستقل IP
بسته		طول بسته، میانگین و واریانس طول بسته، میانگین مربع ریشه
	TCP	اندازه بسته، درگاه مقصد تعداد بسته‌ها با مجموعه بیت PUSH، تعداد بسته‌های خراب
شبکه	HTTP	نام میزبان، نوع محتوا، ارجاع کننده، کوکی‌ها نوع عامل
	SSL	نسخه SSL، نام سرور، تاریخ انقضای مجوز
	DNS	زمان فعالیت، پرچم‌های پاسخ، تعداد اسامی متعارف، رتبه پرس‌وجو

تحلیل شبکه را با توجه به پیام ها و ایمیل ها درنظر بگیرید. ایمیل‌هایی با حداکثر محافظت و رمزگذاری انتها به انتها هم در معرض تحلیل شبکه و حملات مخرب قرار دارند . در جدول ۲، عناصر پیام ایمیلی که برای سرور واسط SMTP مشخص است ارائه شده؛ این سرور از TLS برای انتقال پیام‌ها و دسترسی در تحلیل شبکه استفاده می‌کند.

جدول ۲: میزان دسترسی عناصر پیام ایمیل

عنصر پیام ایمیل	دسترسی برای تحلیل شبکه
مسیر بازگشت: هدر[۲]	بله
دریافت: هدر	بله
از: هدر	بله
به: هدر	بله
CC: هدر	بله
تاریخ: هدر	بله
موضوع: هدر	بله
پیام- هویت: هدر	بله
تمامی دیگر هدرها	بله
اندازه پیام	بله
زمان دریافت پیام	بله
رمزگذاری آشکار پیام	بله
محتوای پیام (رمزگذاری احتمالی یا حقیقی)	خیر

درکل، تحلیل ترافیک شبکه از داده های NetFlow6 یا پروتکل‌های مشابه استفاده می کند. عناصر داده‌های جریان که برای تحلیل شبکه استفاده می‌شوند عبارتند از:

جدول ۳: دسترسی به عناصر جریان ترافیک شبکه

جریان ترافیک شبکه	آسیب‌پذیر به تحلیل شبکه؟
# بسته‌ها در جریان	بله
# اوکتات‌ها در جریان	بله
شروع/ توقف برچسب زمانی	بله
Src/dst آدرس‌های IP	بله
Src/dst اعداد درگاه	بله
پروتکل IP	بله
ارزش نوع خدمات	بله
پرچم‌ها	بله
(باقی موارد)	بله
محتوای بسته	خیر

[۲] Header

تحلیل ترافیک شبکه را می‌توان در سطوح پارامتری مختلف مطابق با داده‌های ترافیک انجام داد:

A. تحلیل منبع/ مقصد

B. تحلیل حجمی ترافیک

C. تحلیل توالی

D. تحلیل استنتاجی

اطلاعات تحلیلی فوق را می‌توان به صورت زیر خلاصه کرد:

جدول ۴: خلاصه انواع تحلیل ترافیک شبکه

تحلیل	روش	مثال
تحلیل منبع/ مقصد	منبع و مقصد پیام تحت کنترل قرار دارد . محتوای پیام اهمیت ندارد.	۱. ایمیل‌های ارسالی توسط کارمند دولت ۲. تراکنش‌های کارت اعتباری پردازش سیستم ۳. دستورهای شبکه روباتی و میزبان های کنترل ۴. ثبت کاغذی و سفارشات مبتنی بر ردیابی
تحلیل حجمی ترافیک	تعداد پیام ها بین گروه های ارتباطی تحت نظارت قرار دارند.	۱. ارتباطات بین واحدهای مختلف در موقعیتی جنگی ۲. موقعیت خاموشی رادیویی (ارتباطات پیام صفر)
تحلیل توالی	الگوی ارتباطی پیوسته با اندازه داده تقریبا مشابه	۱. پیام‌های ارسالی به اعضای گروه مخالف سیاسی
تحلیل استنتاجی	تحلیل رویدادهای مستقلی که ترافیک را شکل می‌دهند و قادر به پیش بینی رویداد آینده هستند	۱. اطلاعات در مورد ایمیل‌ها از شرکت ثبت رویداد و فاکتورهای فروشگاه شیرینی پزی، کارت تولد غافلگیری

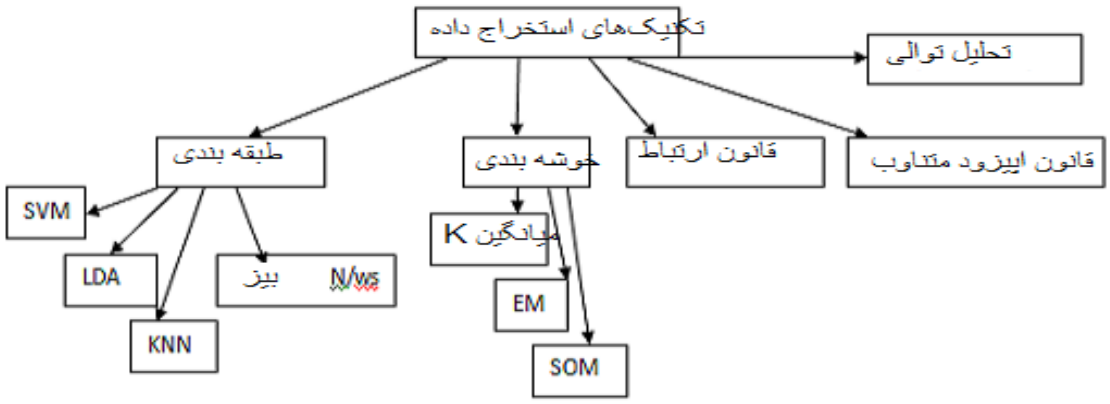
III. تکنیک‌ها

تکنیک‌های مختلف تحلیل شبکه به استخراج داده، آمار و مصورسازی تعلق دارند.

A. تکنیک‌های استخراج داده

داده‌های جمع‌آوری‌شده روند پیش پردازش، فیلتربندی، تغییرشکل و سازماندهی نهایی را برای قرارگیری در مجموعه های اطلاعاتی پشت سر می‌گذارند. این مجموعه‌ها برای شکل گیری الگوها و آشکارسازی الگوهای موقعیتی شناسایی نشده کارایی دارند . تکنیک‌های موفق استخراج داده عبارتند از: استخراج قانون ارتباط، تحلیل طبقه بندی، قانون اپیزود متناوب، رویکرد خوشه‌بندی و تحلیل توالی یا لینک.

شکل ۲: تکنیک‌های استخراج داده



این تکنیک توانایی زیادی در شناسایی نوع خاصی از فعالیت خرابکارانه دارد ولی تعمیم آن برای چندین فعالیت خرابکارانه دشوار است

جدول ۵: دسته‌های مختلف طبقه بندی‌کننده ترافیک شبکه برمبنای استخراج داده

دسته روش طبقه‌بندی	روش مورد استفاده	مزایا	معایب	ملاحظات
طبقه‌بندی مبتنی بر درگاه	استفاده از هدر بسته اپلیکیشن، بارزسی و تطابق آن با عدد درگاه ثبت شده TCP یا UDP و نهاد اعداد انتصابی اینترنت[۳] (IANA). تلاش برای شناسایی اپلیکیشن با استفاده از عدد درگاه	سادگی اجرا	اپلیکیشن‌ها می‌توانند از اعداد درگاه ثبت نشده و همچنین از درگاه تصادفی در کاربردهای نظیر به نظیر استفاده کند . این احتمالات باعث افزایش نتایج منفی طبقه بندی‌کننده می شود. بعضی از اپلیکیشن‌های غیرقانونی خودشان را در درگاه های شناخته شده مخفی می‌سازند تا اینکه دسترسی غیرقانونی به سیستم ها داشته باشند. این احتمالات باعث افزایش نرخ مثبت نادرست طبقه بندی‌کننده‌ها می‌شود چون کاربردهای ناشناخته‌ای وجود دارد. حتی اعداد واقعی درگاه هم در نتیجه انشعاب هدر TCP یا UDP در نتیجه رمزگزاری لایه IP ناشناس باقی می مانند.	میزان دقت تقریبا کم است.
طبقه‌بندی مبتنی بر میزان بار	استفاده از محتوای بسته، بازرسی آنها و تلاش برای تطابق آنها با مجموعه تعیین کننده امضاهای ذخیره شده (الگو)	دقت بسیار بالای نتایج طبقه بندی	بررسی بسته رمزنگاری شده و ابهام پروتکل غیرممکن است، بررسی محتوای بسته باعث نقض سیاست‌های حریم خصوصی می‌شود، این روش هزینه های محاسباتی بالایی دارد، باری را بر ابزار طبقه بندی کننده تحمیل می سازد که این امر به دلیل نیاز به نمونه های مختلف دسترسی به محتوای بسته است . طبقه بندی مبتنی بر میزان بار عملکرد ضعیفی برای ترافیک شبکه‌ای با سرعت بالا و تعداد بالای جریان‌ها دارد.	مقدار زیاد ترافیک رمزگذاری‌شده، فاقد طبقه بندی است.
طبقه‌بندی آماری	استفاده از خصوصیات آماری جریان تراف یک شبکه یا	بهتر از تکنیک های مبتنی بر میزان بار	عملکرد طبقه بندی‌کننده بر ویژگی‌های استخراجی از جریان	استخراج ویژگی باید با دقت زیاد

[۳] Internet Assigned Numbers Authority

	اندازه‌گیری‌های سطح جریان مثل مدت بسته، زمان رسیدن بسته، طول بسته، زمان هدررفت جریان ترافیک برای شناسایی کاربرد . طبقه بندی برمبنای خصوصیات آماری، استفاده از الگوریتم های یادگیری ماشینی برای هدایت الگوهای مختلف ترافیک از مجموعه‌های بزرگ داده . الگوریتم‌های یادگیری ماشینی بسیار سبک وزن بوده و هزینه محاسباتی کمتری نیاز دارند.	چون محتوای بسته‌ها استفاده نشده و قادر به تحلیل ترافیک رمزگذاری شده است.	بستگی دارد. این ویژگی ها به دانش وسیعی نیاز دارند.	انجام شود.
طبقه‌بندی رفتاری	استفاده از الگوهای ترافیک مثل تعداد درگاه های مجزای بررسی شده، پروتکل های لایه انتقال، پیوندها بین نقاط انتهایی برای یافتن نوع اپلیکیشن میزبان . طبقه بندی کننده بر ترافیک شبکه در نقطه میزبان نظارت داشته و تلاش داد تا نوع اپلیکیشن را با استفاده از الگوهای ترافیک شبکه در سمت میزبان هدف شناسایی کند.	نتایج با در اختیار داشتن هزینه محاسباتی کمتر امیدوارکننده هستند.	محدودیت‌های این دسته از تحقیقات ریشه در روش مورد استفاده دارد.	

B. تکنیک‌های آماری

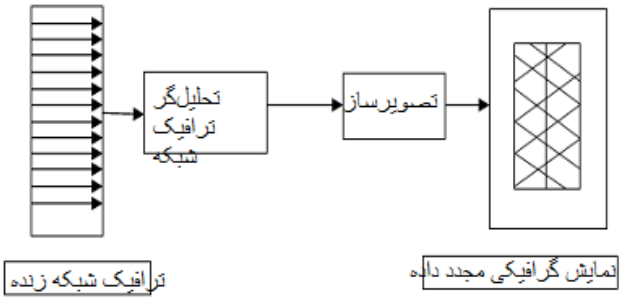
استفاده از تکنیک های آماری برای شناسا یی داده‌های نامنظم و فاقد تناسب است . تکنیک‌های رایج این دسته به حجم و ماهیت داده های ترافیک شبکه اشاره دارند . نوسان فصلی در ترافیک داده اینترنتی وجود دارد، بنابراین بهترین روش های آماری برای پیش بینی عبارتند از تجزیه، هموارسازی نمایی وینتر[۴] و ARIMA.

[۴] Winter`s exponential smoothing

C. تکنیک‌های مصورسازی

تکنیک‌های مصورسازی به تحلیل‌گر انسانی برای شناسایی گرایشات مفقودی در روش‌های خودکار کمک می‌کند. وابستگی به تحلیل‌گر انسانی برای ارجاعات و تفسیرها. تصویرسازی تکنیکی برای نمایش گرافیکی مجموعه داده‌هاست. زمانی‌که داده بزرگ یا انتزاعی باشد، تصویرسازی به درک یا خوانش راحت‌تر داده‌ها کمک خواهد کرد.

شکل ۳: معماری کلی تصویرسازی



جدول ۶: خلاصه تکنیک‌های آماری برای تحلیل ترافیک شبکه

روش آماری	روش مورد استفاده	مزایا	معایب
تجزیه	اصل مبنایی روش تجزیه سری زمانی، تفکیک داده های سری زمانی به چندین الگو، شناسایی سری‌های مجزا شده و سپس کشف آنها به شکل مجزاست. پس از کشف، داده‌ها با هم ادغام می شوند تا پیش بینی انجام شود . مدل افزودنی شامل اجزای زیر است: Yt = گرایش + فصلی + خطا مدل ضربی شامل اجزای زیر است: Yt = گرایش * فصلی * خطا که در آن، مشاهدات Yt به سمت زمان t سوق دارد.	تفکیک داده به بهبود دقت پیش بینی منتهی شده و درک بهتری نسبت به نگرش داده‌های سری زمانی حاصل می‌شود	بیشتر زمان صرف تجزیه و سپس ادغام می‌شود.
هموارسازی نمایی وینتر	دو مدل وجود دارد، ضربی و افزودنی . مدل ضربی حاوی تکثیر بین مولفه گرایش و مولفه فصلی بوده و در زمانی استفاده می شود که داده در فصلی خاص به فصل قبل وابسته است. $Yt = (b1 + b2t)\,St + \varepsilon t$ b1 مولفه دائمی، b2 مولفه گرایش خطی، St فاکتورهای فصلی ضربی، εt مولفه خطا است مدل افزودنی شامل مجموع مولفه گرایش با مولفه فصلی شده و در صورت رسیدن مقدار	یادگیری و استفاده از آن ساده است. پیش بینی های دقیقی را ارائه می دهد. اهمیت بیشتری به مشاهدات اخیر می‌دهد.	پیش بینی‌های این روش تطابق لازم با گرایش واقعی را ندارند . قادر به هدایت مناسب گرایشات نیست.

	تفاوت داده به مقداری ثابت در هر فصل استفاده می‌شود. $yt = (b1 + b2t)\ St + \varepsilon t$ (4) که در آن b1 مولفه دائمی، b2 مولفه گرایش خطی، St فاکتورهای فصلی افزودنی و ε t مولفه خطا هستند.		
ARIMA	روش ARIMA برای تحلیل سری زمانی متشکل از میانگین متحرک (MA) و اوتورگرسیو (AR) استفاده می شود. روش‌های ARIMA با شرایط سری زمانی ثابت استفاده می‌شود که در آن p فرآیند AR است، d فرآیند متمایزسازی برای تبدیل داده به نوع ثابت بوده و در نهایت q در MA پردازش می‌شود.	مدل ARIMA برای پیش بینی دارای کمترین مقدار خطاست	فرآیندی پیچیده و زمان بر

جدول ۷: ابزارهای مصورسازی موجود

ابزار تصویرسازی	مجموعه داده	مالکیت دیداری	تصویرسازی	تعامل	خصوصیت
NfSen	Cisco NetFlow	جریان، بایت، بسته، زمان آغاز	نمودار سری زمانی	فیلترکردن، رمزگذاری رنگی	توان عملیاتی
IDS Rainstorm	Snort Log	آدرس IP، امنیت هشدار، نشان زمان	ترکیب نمودار پراکندگی و نمودار مختصات موازی	زوم کردن معنایی، پرس‌وجوی پویا، رمزگذاری رنگی، تصاویر متحرک	پیوند، هشدار
Rumnit	Pcap	کل	بارش دووجهی، نمودار پراکندگی، نمودار مختصات موازی، ترکیب نمودار پراکندگی و نمودار مختصات موازی	دید چندگانه، فیلترکردن، رمزگذاری رنگی	میزان بار، پیوند، موقتی
SIFT	Argus, NetFlow, NCDA Unified Netflow	آدرس IP	نمودار پراکندگی، نمودار مختصات موازی، هیستوگرام، چشم ماهی، علامت حک شده	پرس‌وجوی پویا، جستجوی محتوای کانونی، زوم کردن	پیوند، آمار
VIAssist	NetFlow	کل	هیستوگرام، نمودار مختصات موازی،	پرس‌وجوی پویا، رمزگذاری رنگی،	آمار، پیوند

			علامت حک شده، لنز جدول، نمودار شبکه	دید چندگانه، انتخاب پیوندی	
IDGraphs	NetFlow	آدرس IP، درگاه، پرچم TCP	نمودار پراکندگی	دید چندگانه، انتخاب پیوندی، رمزگذاری شفاف	پیوند ناقص
InetVis	NetFlow	آدرس IP، درگاه	نمودار پراکندگی سه بعدی	دستکاری فضای سه بعدی، رمزگزاری رنگی	پیوند
AfterGlow	Pcap، تعداد زیادی ثبتی	کل	نمودار شبکه، نقشه درختی	هیچ کدام	پیوند

D. تکنیک‌های یادگیری ماشینی

تکنیک‌های رایج این گروه عبارتند از شبکه بیز، مدلهای مارکوف، شبکه های عصبی، الگوریتم های منطق فازی و ژنتیک . این تکنیک ها به صورت زیر طبقه بندی می‌شوند:

(۱) یادگیری تحت نظارت

یادگیری تحت نظارت عمدتا توسط فعالیت های یادگیری ماشینی استفاده می شود. روند یادگیری تحت نظارت قرار دارد به طوریکه یادگیری الگوریتم از مجموعه داده یادگیری مشابه با معلمی است که بر فرآیند یادگیری نظارت دارد . این روش از متغیرهای ورودی (X)، متغیر خروجی (Y) و یک الگوریتم برای یادگیری تابع نقشه برداری از ورودی تا خروجی استفاده می‌کند. تابع نقشه برداری باید به حدی سازمان‌یافته باشد که در زمان استفاده از داده ورودی جدید (X)، متغیرهای خروجی (Y) به درستی داده را پیش بینی کنند. پاسخ‌های درست شناسایی شده و الگوریتم مرتبا در مورد داده ها پیش‌بینی می‌کند. فرد ناظر هم اصلاحات لازم را انجام می دهد. یادگیری زمانی متوقف می شود که الگوریتم به سطحی قابل قبول از عملکرد برسد.

(۲) یادگیری شبه نظارتی: میانگین‌های k

مسائل یادگیری شبه نظارتی دارای مقدار زیادی داده ورودی (X) بوده و تنها بخشی از آن برچسب (Y) زده می‌شود. این مسائل بین یادگیری تحت نظارت و بدون نظارت قرار می گیرند، برای نمونه بایگانی عکس با چند عکس برچسب خورده (برای مثال، سگ، گربه، انسان) و تعداد زیادی عکس فاقد برچسب. گروه بزرگی از مسائل یادگیری ماشینی واقعی در این طبقه بندی قرار دارند. تکنیک‌های یادگیری بدون نظارت در کشف و شناسایی ساختار متغیرهای ورودی کارایی دارند.

(۳) یادگیری بدون نظارت

مسائل یادگیری بدون نظارت دارای داده ورودی (X) بوده ولی متغیرهای خروجی معادل برای آن وجود ندارد . این روش یادگیری به دنبال مدل‌سازی ساختار یا توزیع مبنایی در داده هاست تا اینکه اطلاعات بیشتری در مورد داده ها به دست آورد. این تکنیک به این دلیل بدون نظارت خوانده می‌شود که در آن پاسخ‌های درست و معلم یا ناظر وجود ندارد. الگوریتم‌ها ساختار مورد نظرشان را در داده‌ها شناسایی می‌کنند.

جدول ۸: طبقه بندی تکنیک تحت نظارت

ماشین بردار پشتیبان	طبقه بندی	تکنیک‌های تحت نظارت
جنگل تصادفی برای طبقه بندی و رگرسیون		
رگرسیون خطی	رگرسیون	

جدول ۹: طبقه بندی تکنیک بدون نظارت

	خوشه بندی طبقاتی	تکنیک‌های بدون نظارت
	خوشه بندی طیفی	
میانگین‌های K	خوشه بندی بخش بندی	
DBSCAN	خوشه بندی مبتنی بر تراکم	
طبقه خودکار	خوشه بندی مبتنی بر مدل احتمال گرا	
	خوشه بندی مبتنی بر گرید	

جدول ۱۰: خلاصه مطالعات مرورشده

نویسنده، سال	پیش پردازش	تکنیک‌ها	مجموعه داده	هدف	متریک‌های ارزیابی
ام بالدی، ای بارالیس، اف ریسو، ۲۰۰۵	بدون اشاره	قانون ارتباط	شبکه پردیس دانشگاه (از جمله حدود ۶۰۰۰ سیستم نهایی)	ردیابی میزبان های فعال	حداقل پشتیبانی
اسرا کاهیا- اوزیرمیدوکوز، علی گزر، سبرایل سیفلیکلی، ۲۰۱۲	تحلیل شناسایی موارد غیرعادی و شبکه‌های کوهونن	مدل درخت تصمیم CART	بایگانی ترافیک گروه کاری MAWI	به تصویرکشیدن داده ترافیک شبکه	ارزیابی ده برابری دقت اعتبارسنجی عرضی
وسام اس بایا، ساود ای الاسدی، ۲۰۱۶	نمونه برداری، کاهش بار، ترسیم نقشه، تجمع	خوشه بندی، طبقه بندی، شناسایی مورد غیرعادی زمانی	DARPA	شناسایی امور غیرعادی در ترافیک شبکه	نرخ‌های مثبت نادرست
ولارد- الواردو، پی، مارتینز- پلاز، آر رویز ایبارا، جی، مورالس- روخا، ۲۰۱۴	پاکسازی رد داد ها، تقسیم به چندین قسمت. فایل‌های ترافیک با ipsumdump ساخته می شوند.	تکنیک‌های استخراج داده و آنتروپی (نرخ عناصر بازمانده یا فضاهای سه بعدی روش‌های آنتروپی)	مبنای اینترنت (شبکه پردیسی)	پروفایل بندی ترافیک شبکه برای شناسایی ورود غیرمجاز	واریانس، آستانه، دونمایی گاس

	تولید جریان انجام می‌شود.				
محمد شفیق، ژیانگژان یو، آصف علی لاگاری، لی یائو، نابین کومار کارن، فودیل عبدسامیا، ۲۰۱۶	انتخاب ویژگی و استخراج از مجموعه داده	ابزار Weka و استفاده از چهار الگوریتم یادگیری ماشینی C4.S، ماشین بردار پشتیبان، بیز نت، بیز ساده برای ایجاد مدل طبقه بندی با استفاده از اعتبارسنجی عرضی ۱۰ پوشه	ترافیک شبکه واقعی برای مدت یک دقیقه استفاده از WWW، DNS، FTP، P2P و Telnet	طبقه بندی ترافیک شبکه	اطلاعات جامع، زمان تمرین و یادآوری دقت
پورناوانسیا، هاویلودین، رینر آلفرد، آخمد فانانی اونلیتا گافر، ۲۰۱۸	داده اصلی برای افزایش سرعت فرآیند شمارش نرمال سازی می‌شود	تجزیه، هموارسازی نمایی وینتر و ARIMA	ترافیک اینترنت در شبکه در دانشگاه مولاوارمان	پیش بینی استفاده از ترافیک اینترنت در دانشگاه مولاوارمان	MAPE MAD MSD
پالاوی سینگال، راجیو ماتور، هیمانی ویاس، ۲۰۱۳	ذکر نشده	یادگیری ماشینی بدون نظارت	مجموعه داده تعادل- تمرین xls متشکل از ۴۹۲ سابقه، ۴ ویژگی و ۶ خوشه. مجموعه داده شیشه- تمرین xls متشکل از ۱۵۸ سابقه، ۹ ویژگی و ۴ خوشه. مجموعه داده iris- تمرین ۳۰ متشکل از ۱۲۰ سابقه، ۴ ویژگی و ۴ خوشه	طبقه بندی ترافیک شبکه	زمان محاسباتی

IV: نتیجه گیری

تحلیل ترافیک شبکه وظیفه مهمی در ارتباط با امنیت شبکه است چون اطلاعاتی را در مورد استفاده از شبکه، آگا هی سرعت داده ها برای دانلود و آپلود، منبع، مقصد، نوع، اندازه و محتوای بسته های داده فراهم می کند. این تحلیل در یافتن تنگناها در شبکه کمک می کند، برای نمونه تخصیص یا مصرف پهنای باند ناکافی. تخصیص پهنای باند با دستیابی به اطلاعات در مورد کاربرد شبکه به حالت بهینه تبدیل می‌شود. این روند به افزایش کیفیت خدمات زیرساخت شبکه کمک می کند. دانش کافی از سطوح استفاده شبکه باعث تسهیل تحلیل نیازهای آینده شبکه می‌شود. این فرآیند بخشی از مهندسی شبکه به شمار می‌رود.

مراجع

[1] http://scitechconnect.elsevier.com/wp-content/uploads/ 2013/09 /Introducing- Network-Analysis.pdf

[2] Manish R. Joshi, A Review of Network Traffic Analysis and Prediction Techniques

[3] Daniel J. Arndt A. Nur Zincir-Heywood , A Comparison of Three Machine Learning Techniques for Encrypted Network Traffic Analysis, 978-1-4244-9941-0/11/$26.00 ©2011 IEEE

[4] https://www.m3aawg.org/sites/default/files/m3aawg_traffic_analysis_2016-06.pdf

[5] M. Baldi, E. Baralis, F. Risso, Data Mining Techniques for Effective and Scalable Traffic Analysis, 0-7803-9087-3/05/$20.00 ©2005 IEEE

[6] Velarde-Alvarado, P., Martinez-Pelaez, R., Ruiz-Ibarra, J. and Morales-Rocha, V. (2014) Information Theory and Data-Mining Techniques for Network Traffic Profiling for Intrusion Detection. *Journal of Computer and Communications*, 2, 24-30

[7] Wesam S. Bhaya, Saud A. Alasadi , Anomaly Detection in Network Traffic Using Stream Data Mining: Review, Research Journal of Applied Sciences 11(10):1076- 1082, ISSN:1815-932X,2016.

[8] BAN Tao and KADOBAYASHI Youki, Data Mining over Network Streams, Journal of the National Institute of Information and Communications Technology Vol. 58 Nos. 3/4 2011

[9] Ankit Naik [1], S.W. Ahmad, Data Mining Technology for Efficient Network Security Management, International Journal of Computer Science Trends and Technology (IJCST) – Volume 3 Issue 3, May-June 2015

[10] Ranju Marwaha, Intrusion Detection System Using Data Mining Techniques– A Review, 2017, IJARCSSE.

[11] Noora AI Khater Richard E Overill, Network Traffic Classification Techniques and Challenges, The Tenth International Conference on Digital Information Management (ICDIM 2015).

F oudil Abdessamia, [12] Muhammad Shafiq, Xiangzhan Yu, Asif Ali Laghari, Lu Yao, N abin Kumar Karn, Analysis Using Machine Learning Algorithms, Network Traffic Classification Techniques and Comparative 2016 2nd IEEE International Conference on Computer and Communications