طراحی و پیاده‌سازی یک چارچوب مبتنی بر اعتماد برای تشخیص ترافیک موبایل در شبکه

مریم ترابی[۱]، حمیدرضا محروقی[۲]، سبحان علی‌آبادی[۳]، هاله امین طوسی[۴]

[۱]کارشناسی ارشد فناوری اطلاعات – مخابرات امن، گروه مهندسی کامپیوتر، دانشگاه بین‌المللی امام رضا (ع)

[۲] استادیار گروه مهندسی کامپیوتر، دانشگاه بین المللی امام رضا (ع)

[۳]کارشناسی ارشد، امنیت فناوری اطلاعات، گروه مهندسی کامپیوتر، دانشگاه بین المللی امام رضا (ع)

[۴]استادیار گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشگاه فردوسی مشهد، amintoosi@um.ac.ir

چکیده

رشد چشم‌گیر دستگاه‌های موبایل باعث شده است تا یکی از مسائل مهم امنیتی امروز، رسیدگی و پاسخگویی به حوادث امنیتی باشد. تحلیلگران امنیتی با اهداف متفاوتی مانند شناسایی برنامه‌های کاربردی موبایلی، شناسایی بدافزارهای موبایلی و شناسایی سیستم‌عامل، تحلیل ترافیک شبکه را انجام می‌دهند. با توجه به ماهیت پویای دستگاه‌های موبایل، تشخیص و تحلیل ترافیک آن‌ها مشکل‌تر از سایر تجهیزات شبکه است. هرچند تحقیقات گسترده‌ای در این حوزه انجام شده است، با این حال اکثر روش‌ها برای تشخیص در ترافیک واقعی کارا نیستند. در این مقاله، چارچوبی جهت تشخیص ترافیک موبایل با رویکردی مبتنی‌بر اعتماد ارائه شده است. چارچوب پیشنهادی از چهار زیرسیستم تشخیص در لایه‌های مختلف TCP/IP و دو زیرسیستم کنترلی تشکیل شده است. زیرسیستم‌های تشخیص وظیفه تشخیص ترافیک موبایل و زیرسیستم‌های کنترلی وظیفه اعتبارسنجی اطلاعات و تشخیص نهایی را برعهده دارند. در این مقاله به‌منظور ارزیابی چارچوب پیشنهادی و بررسی چالش‌های موجود، مجموعه‌داده‌ای واقعی از ترافیک شبکه بی‌سیم (دانشگاه فردوسی مشهد) تهیه و مورد استفاده قرار گرفته است. نتایج حاصل از اجرای سیستم پیشنهادی بر روی مجموعه‌داده نشان می‌دهند که سیستم پیشنهادی قادر است با تجمیع نظرات زیرسیستم‌های تشخیص براساس میزان اعتماد و اطمینان به هریک از آن‌ها، فرآیند تشخیص را با دقت ۰.۹۷ و خطای ۰.۰۹ انجام دهد

کلیدواژه

تحلیل ترافیک شبکه، اعتماد، دستگاه موبایل، بدافزار موبایل، تشخیص سیستم‌عامل، دستگاه‌های اینترنت اشیاء.

مقدمه

دستگاه‌های موبایلی جدید با قابلیت‌های متنوع امکان انجام امور مختلف (مانند امور مالی، خرید برخط[۱]، ذخیره‌سازی اطلاعات شخصی) را برای کاربران فراهم آورده است. کاربران تلفن همراه برای استفاده از اینترنت از شبکه‌های بی‌سیم و یا شبکه‌های سلولی استفاده می‌کنند. کنترل، مدیریت و حفظ امنیت دستگاه‌های موجود در این‌گونه شبکه‌ها، حیاتی به نظر می‌رسد. در این راستا مقالات بسیاری در حوزه تحلیل ترافیک شبکه دستگاه‌های موبایلی ارائه شده است. بنابر دسته‌بندی منتشر شده در کار مروری [۱]، اهداف تحلیل ترافیک شبکه موبایل در سه حوزه کاربران، برنامه‌های کاربردی و دستگاه‌ها جای می‌گیرند و از جمله این اهداف می‌توان به بدست آوردن اطلاعات از خصوصیات شبکه جهت کارهایی نظیر مدیریت شبکه، شناسایی برنامه‌های کاربردی موبایلی، تشخیص بدافزارهای موبایلی، جلوگیری از نشت اطلاعات شخصی کاربران، انگشت‌نگاری کاربران موبایلی و شناسایی سیستم‌عامل موبایل اشاره نمود. هریک از این شاخه‌ها در جایگاه خود از اهمیت بالایی برخوردار است و پژوهش‌های متنوعی در هر یک از این حوزه‌ها انجام شده است. به‌عنوان نمونه، در پژوهش‌هایی مانند [۲] و [۳] شناسایی و جلوگیری از نشت اطلاعات شخصی کاربران با هدف بررسی نحوه رفتار برنامه‌های موبایلی از حیث حریم خصوصی کاربران مورد بررسی قرار می‌گیرد. تشخیص رفتار بدخواهانه یک برنامه موبایلی و یا تعیین میزان امنیت برنامه‌های کاربردی مانند

[۱] online

تاریخ دریافت: ۱۳۹۸/۰۶/۳۰ تاریخ پذیرش: ۱۳۹۸/۱۰/۰۸

کارهای انجام شده در [۴]، [۵]، [۶] و [۷] نیز می‌تواند توسط توسعه‌دهندگان موبایل و یا متخصصین امنیتی مورد بررسی قرار گیرد.

شناسایی برنامه‌های کاربردی موبایلی با هدف مدیریت منابع و کنترل برنامه‌ها در شبکه مانند پالایش[۲] یک برنامه کاربردی خاص بنا به سیاست‌های خاص انجام می‌شود. این اطلاعات می‌تواند در تشخیص بدافزارهای موبایلی و تحلیل رفتاری دقیق‌تر آن‌ها مؤثر باشد. طبق گزارش آماری کار مروری [۱]، شناسایی برنامه کاربردی جزء بیش‌ترین کارهای انجام شده در حوزه تحلیل ترافیک بوده است. از جمله جدیدترین کارهای انجام شده در این حوزه می‌توان به [۸] و [۹] اشاره نمود.

داشتن آگاهی نسبت به سیستم‌عامل هدف، چه برای انجام حملات و چه برای تشخیص و ردیابی حمله‌کننده، به‌خصوص در سیستم‌عامل‌های موبایلی بخاطر ماهیت متفاوت و پویای آن‌ها نسبت به سیستم‌های سنتی به یکی از موضوعات پرچالش تبدیل شده است. در این زمینه نیز کارهای بسیاری مانند [۱۰] و [۱۱] انجام شده است.

تمرکز ما در این مقاله بر روی دو حوزه شناسایی سیستم‌عامل موبایلی و شناسایی برنامه‌های کاربردی با هدف بیان چالش‌های تشخیص در شبکه‌های واقعی است. بسیاری از کارهای صورت گرفته بر روی نمونه‌های ترافیک جمع‌آوری شده در محیط شبیه‌سازی شده و یا به تعداد محدود دستگاه موبایلی هستند. هرچند پژوهش‌هایی نیز مانند [۱۰] و [۱۲] بر روی ترافیک واقعی با حجم بالا ارایه شده‌اند [۱۰] اما چالش اصلی در این کارها تعیین هویت دستگاه‌های موجود در ترافیک جهت ارزیابی دقیق‌تر است.

در دنیای واقعی، شرایط متفاوت شبکه‌ها نظیر معماری، تجهیزات امنیتی و نحوه جمع‌آوری ترافیک در تشخیص ترافیک موبایل بسیار تأثیرگذار خواهد بود. زمانی‌که در شبکه‌ها از انواع سیستم‌های امنیتی نظیر دیواره‌های آتش، انواع ترجمه آدرس شبکه[۳]، انواع تونل‌ها (مانند پروتکل IPsec) و انواع کارپذیرها[۴] (مانند SOCKS) استفاده شود؛ برخی از مقادیر در لایه‌های مدل TCP/IP تغییر خواهد کرد.

یکی دیگر از چالش‌های مطرح در تشخیص ترافیک موبایلی ضعف ابزارهای انگشت‌نگاری است. به طور کلی روش‌های فعال دقت بالاتری نسبت به روش‌های غیرفعال دارند، با این حال روش‌های فعال به دلیل نیاز به تعامل با سیستم‌عامل هدف، باعث ایجاد تغییرات در ترافیک شده و سربار و هزینه زیادی به سیستم وارد می‌کنند. بنابراین امکان استفاده از این روش‌ها در همه شرایط فراهم نیست. پایگاه‌داده‌های ابزارهای غیرفعال معروف مانند P0f، Ettercap و Satori به‌روز نبوده و برای سیستم‌عامل‌های موبایلی مناسب نیستند. با توجه به تنوع بالا و به روزرسانی‌های نسخه‌های مختلف سیستم‌عامل‌ها و برنامه‌های کاربردی و شباهت میان آن‌ها، نیاز به ایجاد مستمر پایگاه داده‌های تشخیص است. رفتارهای مشابه میان سیستم‌عامل‌های موبایلی و رومیزی بخاطر بهره‌گیری از هسته سیستم‌عامل‌های موبایلی از سیستم‌عامل‌های سنتی بیان شده است [۱۳]. در پژوهش [۱۰] پایگاه‌داده‌ای برای تشخیص ترافیک موبایلی تولید شده است، با این حال این پایگاه‌داده کامل نبوده و نیازمند بروزرسانی است.

دستکاری عمدی اطلاعات معمولاً توسط کاربر بدخواه یا بدافزار موجود در دستگاه موبایل انجام شود. بعضی از بدافزارهای جدید مانند بات‌نت‌ها به منظور پنهان نمودن هویت اصلی خود، دست به تغییر اطلاعات لایه کاربرد مانند مؤلفه رشته عامل کاربر[۵] می‌زنند [۱۴]. این مسئله ممکن است باعث پایین آمدن دقت در تشخیص و فریب دادن تحلیلگران امنیتی گردد.

مطالعات اخیر نشان می‌دهند که حمله‌کنندگان برای کاهش دقت روش‌های مبتنی بر یادگیری ماشین، دست به انجام حملات و تغییر داده‌های مربوط به ویژگی‌های پایه‌ای الگوریتم‌های یادگیری می‌زنند. در مقاله [۱۵] در مورد حملات و مسائل امنیتی و حریم خصوصی که روش‌های یادگیری ماشین با آن‌ها مواجه هستند بحث شده است. عدم توجه به این مسائل در تشخیص، عملکرد این‌گونه روش‌ها را بی‌اثر می‌کند.

مقالات متعددی برای رفع مشکلات و چالش‌های مطرح شده در این حوزه منتشر شده‌اند. هریک از روش‌های تحلیل ترافیک از اطلاعات یک یا چند لایه شبکه استفاده می‌کنند. به‌طور کلی روش‌هایی که از اطلاعات چند لایه استفاده می‌کنند، دارای دقت بالاتری هستند. چرا که در صورت نامعتبر بودن اطلاعات هر لایه، امکان استفاده از سایر اطلاعات وجود دارد.

سهم اصلی ما در این حوزه، بررسی چالش‌های مطرح در تشخیص سیستم‌عامل‌ها و برنامه‌های موبایلی است. درواقع در این مقاله چارچوب جامعی برای تشخیص تحلیل ترافیک موبایل در ترافیک واقعی ارائه داده‌ایم. نوآوری مقاله به شرح زیر است:

- در این مقاله چارچوب جامعی برای تشخیص تحلیل ترافیک موبایل بر پایه اعتماد در ترافیک واقعی متشکل از چهار زیرسیستم برای تشخیص در لایه‌های مختلف شبکه ارائه شده است. در جامعه اعتماد چارچوب پیشنهادی، هریک از زیرسیستم‌ها به‌عنوان فرد اعتمادشونده، عملیات تشخیص را به‌صورت مستقل انجام می‌دهند، سپس زیرسیستم تصمیم‌گیر به‌عنوان فرد اعتمادکننده، با تجمیع نظرات

[۲] Filter
[۳] Network Address Translation
[۴] Proxy
[۵] User Agent

برمبنای میزان فاکتور اعتماد و اطمینان محاسبه شده در هر زیرسیستم، تشخیص نهایی را انجام می‌دهد. دو فاکتور اعتماد و اطمینان براساس میزان دردسترس بودن و اعتبار اطلاعات موجود به‌ازای هر کاربر محاسبه شده و باعث بالا رفتن میزان دقت تشخیص نهایی می‌شود.

- از جمله مزایای چارچوب پیشنهادی می‌توان به درنظر گرفتن شرایط و ساختار شبکه در معتبر بودن اطلاعات سطوح مختلف با استفاده از زیرسیستم شناسایی محیط، مدیریت و کنترل بسته‌های ترافیک و تجمیع نظرات زیرسیستم‌ها، عملکرد مستقل و موازی زیرسیستم‌های تشخیص جهت بهبود کارایی سیستم، انعطاف‌پذیری چارچوب جهت قرارگیری در هر شرایط و هر روش تشخیص دلخواه در هر زیرسیستم و امکان ترکیب نمودن آن با دیگر سیستم‌های دسته‌بندی یادگیری ماشین و استفاده از چارچوب در سایر حوزه‌های تحلیل ترافیک به‌خصوص شناسایی اینترنت اشیا اشاره نمود.
- در انتها جهت ارزیابی چارچوب، اقدام به جمع‌آوری ترافیک بی‌سیم دانشگاه فردوسی کرده و براساس ترافیک واقعی ارزیابی انجام شده است. در نهایت نیز به مطالعه موردی چند نمونه از چالش‌های مطرح شده پرداخته‌ایم.

ساختار مقاله در ادامه به این صورت است که بخش دوم برخی از مهم‌ترین کارهای انجام شده در این حوزه بیان شده و در بخش سوم چارچوب پیشنهادی تشخیص ترافیک موبایل معرفی شده و در بخش چهارم ارزیابی چارچوب پیشنهادی انجام شده و در بخش پنجم به بحث‌وگفتگو در مورد چالش‌های مطرح در ترافیک شبکه واقعی پرداخته شده است. در انتها نیز در بخش آخر به جمع‌بندی و ارائه پیشنهادات بیان شده است.

کارهای پیشین

از میان حوزه‌های مختلف تحلیل ترافیک موبایل، تمرکز ما در این مقاله، شناسایی سیستم‌عامل و برنامه‌کاربردی موبایلی است. بنابراین در این بخش به معرفی برخی از کارهای انجام شده در این دو حوزه پرداخته شده است.

برای شناسایی برنامه‌های کاربردی از روش‌های سنتی مانند تشخیص براساس شماره درگاه[۶] استفاده می‌شده است، ولی این روش‌ها به‌علت تغییر رفتار برنامه‌های کاربردی و استفاده اکثر آن‌ها از درگاه‌های HTTP و HTTPS دیگر کارآمد نیستند. در کارهای جدیدتر مانند [۱۶] از بررسی اطلاعات لایه کاربرد و بهره‌گیری از روش‌های مبتنی‌بر درگاه و دسته‌بندهای یادگیری ماشین استفاده شده است. در میان روش‌های ارائه شده تعداد محدودی از روش‌ها مانند [۱۷] از روش‌های مبتنی بر پشته[۷] TCP/IP و کاربرد استفاده کرده‌اند. با این حال این پژوهش روش کاملی به نظر نمی‌رسد، چرا که روشی مبتنی بر امضا بوده و برای ترافیک رمزشده تعمیم‌پذیر نیست.

برخی از کارهای انجام شده فقط بر روی تشخیص یک یا چند برنامه خاص مطالعه داشته‌اند. بنابراین نمی‌توان از آن‌ها برای تشخیص همه برنامه‌های کاربردی بهره گرفت. مقاله [۱۸] روش مناسبی برای تشخیص ترافیک TLS[۸]/SSL[۹] ارائه داده است با این حال تمرکز آن فقط بر روی برنامه پیام‌رسان KakaoTalk بوده است. مقاله [۱۹] طبقه‌بندی ترافیک موبایل را با کمک روش الگوبرداری هستی‌شناسی[۱۰] انجام داده است. در این روش فقط به پنج برنامه موبایلی اشاره شده است. در مقاله [۲۰] نیز تشخیص پنج برنامه موبایلی معروف با استفاده از روش یادگیری ماشین و روش مبتنی بر گراف[۱۱] با استفاده از سرآیند TCP و UDP انجام شده است.

مقاله [۲۱] امکان تشخیص برنامه‌های اندرویدی براساس سرآیند TCP/IP را مورد بررسی قرار داده است. این کار، تشخیص ترافیک رمزشده SSL/TLS را هم شامل می‌شود، با این حال از تشخیص ترافیک IPsec ناتوان است. از دیگر کارهای انجام شده برای شناسایی برنامه‌های کاربردی موبایل مبتنی بر یادگیری ماشین می‌توان به دو کار [۲۲] و [۸] اشاره کرد. در این دو مقاله ترافیک مربوط به ۱۱۰ برنامه‌کاربردی اندرویدی با نسخه‌های متفاوت را جمع‌آوری کرده و با استفاده از الگوریتم‌های دسته‌بندهای پشتیبانی بردار و جنگل تصادفی تشخیص را انجام داده است. این روش را برای سایر برنامه‌های کاربردی غیراندرویدی نمی‌توان تعمیم داد. در مقاله [۱۲] در زمینه بررسی خصوصیات شبکه مستقل از سکو[۱۲] تحلیل را بر روی اطلاعات لایه کاربرد و در بخشی از کار خود از آدرس‌های فیزیکی دستگاه‌ها استفاده کرده است. این کار بر روی شبکه‌های بزرگ که دارای کاربران موبایلی و غیرموبایلی است مورد ارزیابی قرار گرفته است. در مقاله [۲۳] از تکنیک تحلیل بارگذاری صفحات وب برای تشخیص انگشت‌نگاری دستگاه‌های موبایل استفاده کرده است. هرچند این روش فقط بر روی تعداد محدودی دستگاه مورد ارزیابی قرار گرفته است. با این حال این روش از تشخیص ترافیک رمزشده ناتوان است. در کارهای جدیدتر مانند مقاله [۲۴] و [۹] در این حوزه جهت حفظ حریم خصوصی کاربران، از تکنیک‌های جدیدتر

[۶] Port Number
[۷] Stack based
[۸] Transport Layer Security
[۹] Secure Sockets Layer
[۱۰] Ontology Paradigm
[۱۱] Graphlet
[۱۲] Platform-Independent

مانند توزیع احتمالی سایز بسته استفاده شده است. در مقاله [۲۴] در مورد شناسایی برنامه‌های کاربردی ناشناخته صحبت شده است. در این روش با استفاده از دسته‌بند[۱۳] چند سطحی با استفاده از ویژگی‌هایی از جمله اندازه بسته، اندازه محموله بسته، ۱۶ بیت اول محموله بسته استفاده شده است. با این حال این‌گونه روش‌ها در صورت تغییر عمدی اندازه بسته توسط برنامه قابل ردیابی نیستند. در روش [۹] ادعا شده است که در برابر حملات احتمالی از جمله تغییر عمدی اندازه بسته مصون است با این وجود این روش‌ها معمولاً در شبکه‌های بزرگ با برنامه‌های کاربردی متنوع پاسخگو نیستند.

به طور کلی روش‌های تشخیص سیستم‌عامل به دو دسته روش‌های فعال و روش‌های غیرفعال تقسیم‌بندی می‌شوند. در روش‌های فعال، تشخیص با استفاده از تعامل با سیستم هدف انجام می‌شود. از جمله ابزارهایی که از روش فعال برای تشخیص سیستم‌عامل استفاده می‌شود، می‌توان به ابزار Nmap و Xprobe اشاره نمود. این دو ابزار برای تشخیص از بسته‌های TCP، UDP و ICMP بهره می‌گیرند. در مقابل روش‌های غیرفعال، براساس ترافیک دریافتی و بدون تعامل با سیستم هدف، تشخیص را با استفاده از مؤلفه‌های مشخص، انجام می‌دهند. از جمله ابزارهای رایج تشخیص غیرفعال می‌توان به ابزار انگشت‌نگاری سیستم‌عامل P0f اشاره نمود. این ابزار براساس اطلاعات موجود در بسته‌های TCP تشخیص را در دو حالت SYN و SYN+ACK براساس امضاهای از قبل تعیین شده و تطبیق دادن آن‌ها با گزاره‌های بسته تشخیص را انجام می‌دهد [۲۵]. در مقاله [۲۶] میزان مؤثر بودن مقادیر موجود در سرآیند TCP/IP مانند TTL[۱۴]، IP ID، برچسب زمانی TCP[۱۵]، زمان راه‌اندازی سیستم، پایداری فرکانس ساعت[۱۶]، وجود گزینه برچسب زمانی TCP، پیش‌فرض تعیین شده فاکتورهای مقیاس اندازه پنجره TCP با استفاده از روش مبتنی‌بر احتمال برای تشخیص سیستم‌عامل‌های موبایل و تشخیص NAT در شبکه محاسبه نموده و برای تشخیص از آن‌ها بهره گرفته است. در مقاله [۲۷] یک چارچوب فعال/غیرفعال برای تشخیص سیستم‌عامل ارائه شده است. در بخش فعال این روش از بسته‌های ICMP و در بخش غیرفعال از بسته IP مربوط به جریان ویدئو سیستم‌عامل هدف استفاده کرده است. با این حال این روش بر روی تعداد محدودی دستگاه مورد ارزیابی قرار گرفته است. در روش ارائه شده [۲۸] تشخیص سیستم‌عامل براساس زمان‌بندی ترافیک شبکه تولید شده توسط دستگاه موبایل مرتبط به نوع سیستم‌عامل انجام می‌شود. در این کار همچنین شباهت‌های موجود میان نسخه‌های مختلف سیستم‌عامل هم مورد بررسی قرار گرفته است. روش مشابه دیگری که تشخیص سیستم‌عامل‌ها را با تمرکز بر روی ترافیک رمزشده انجام داده است، می‌توان به [۲۹] اشاره نمود. هر چند این روش‌ها برای تشخیص رمزشده مناسب هستند، با این حال در محیط واقعی کارایی خوبی نخواهند داشت، چرا که با هر تغییر عمدی در زمان‌بندی برنامه‌های درحال اجرا در سیستم‌عامل، تشخیص با خطا مواجه خواهد شد. در مقاله [۳۰] از روشی غیرفعال برای تشخیص سیستم‌عامل استفاده شده است که از اطلاعات لایه کاربرد و TCP/IP و بسته‌های TLS برای چندین نشست در یک پنجره زمانی ثابت استفاده می‌کند. تعداد سیستم‌عامل‌های مورد بررسی در این روش محدود هستند. در پژوهش [۳۱] تشخیص به دو صورت فعال و غیرفعال و براساس الگوریتم ژنتیک انجام می‌دهد. تنوع داده‌های جمع‌آوری شده در این روش کم است و تمرکز اصلی آن سیستم‌عامل‌های موبایلی نیست. روش دیگری که با استفاده از ویژگی‌های اطلاعات سرآیند TCP/IP با استفاده از تکنیک ژنتیک دسته‌بندی را انجام شده مقاله [۳۲] است. با این حال در این مقاله نیز تعداد محدودی دستگاه مورد بررسی قرار گرفته است. از جمله جدیدترین روش‌های تک کاوشی می‌توان به مقاله [۳۳] و [۱۱] اشاره کرد، در این روش‌ها تشخیص براساس اطلاعات یک بسته به‌صورت فعال انجام می‌شود.

مقاله [۱۰] یکی از جدیدترین کارهایی که جهت تشخیص سیستم‌عامل‌های موبایلی در شبکه‌های بزرگ بی‌سیم انجام شده است. در این روش، تشخیص براساس ترکیبی از روش مبتنی بر لایه کاربرد و سرآیند TCP/IP انجام شده است. با این حال این روش دارای مشکلاتی است. نویسندگان این مقاله در کار دیگری [۳۴] کارایی چهار الگوریتم یادگیری ماشین را بر روی سه ویژگی TTL، اندازه پنجره TCP و اندازه اولین بسته SYN مورد بررسی قرار داده‌اند. در روش ارائه شده [۱۰] تشخیص براساس سه روش رشته عامل کاربر، دامنه‌های خاص و پارامترهای TCP/IP انجام شده است. یکی از ایرادهای این روش زمانی است که برای یک عامل اطلاعات کافی برای سه روش مذکور نباشد، در این حالت به‌ترتیب اولویت عمل می‌کند. رشته عامل کاربر از اولویت بالاتری نسبت به پارامترهای TCP/IP برخوردار است. این درحالی است که امکان دستکاری این رشته در لایه کاربرد بسیار ساده‌تر است. در بخش ارزیابی نشان داده شده است که پارامترهای لایه کاربرد از جمله رشته عامل کاربر از اعتبار بسیار کمی برخوردار هستند. در کارهای انجام شده اخیر تشخیص انگشت‌نگاری دستگاه‌های اینترنت اشیاء نیز از تکنیک‌های مشابه آنچه برای سیستم‌های موبایلی بیان شده، استفاده شده است. در مقاله [۳۵] اطلاعات موجود در پروتکل لایه کاربرد مانند HTTP، DNS و TLS

[۱۳] Classifier
[۱۴] Time to Live
[۱۵] TCP Timestamp
[۱۶] Clock Frequency

به‌عنوان ویژگی‌های ترافیک انتخاب شده است. در مقاله [۳۶] نیز از اطلاعات سرآیند TCP/IP و اطلاعات لایه کاربرد بهره گرفته است. مقاله [۳۷] از اطلاعات بسته‌ها از جمله اندازه بسته، نوع پروتکل، شماره درگاه جهت استخراج ویژگی‌ها استفاده کرده است. مقاله [۳۸] نیز از مجموع اطلاعات لایه کاربرد و TCP/IP برای تشخیص استفاده کرده است.

همان‌طور که بیان شد، جهت تحلیل ترافیک شبکه از تحلیل اطلاعات ترافیک در لایه‌های مختلف TCP/IP استفاده می‌شود. با این حال هریک از این روش‌ها برای شرایط خاص قابل استفاده هستند. چرا که بسته به شرایط ترافیک ممکن است اطلاعات یک یا چند لایه از ترافیک برای تحلیل کافی نباشد. چارچوب پیشنهادی با هدف پوشش دادن چالش‌های مطرح در تحلیل ترافیک، برای استفاده در کاربردی در ترافیک واقعی ارائه شده است.

چارچوب پیشنهادی

چارچوب پیشنهادی تشخیص را در چهار لایه شبکه انجام می‌دهد. در این چارچوب هر لایه به عنوان یک زیرسیستم تشخیص به صورت مستقل، در مورد موبایل بودن یا نبودن هر عامل تصمیم‌گیری می‌کند. سپس نظرات این چهار زیرسیستم وارد زیرسیستم تصمیم‌گیرنده شده و در آنجا براساس میزان اطمینان به نظرات ارائه شده در هر زیرسیستم، تصمیم نهایی بر پایه اعتماد اتخاذ می‌گردد. هرچند در نگاه کلی، به نظر می‌رسد که بررسی در چهار سطح، سربار محاسباتی دربرخواهد داشت، ولی توجه به چالش‌های مطرح‌شده در بخش قبل، در برخی شرایط نیاز به بررسی در چهار سطح و تجمیع خروجی‌های هر یک اثبات می‌شود.

سیستم پیشنهادی، از دو زیرسیستم کنترلی و چهار زیرسیستم تشخیص تشکیل شده است. نمودار فعالیت چارچوب پیشنهادی در شکل ۱ نشان داده شده است. روند تشخیص به این صورت است که ابتدا در زیرسیستم شناسایی محیط، براساس معماری و ساختار شبکه هدف و نحوه جمع‌آوری ترافیک ورودی، اعتبار هر زیرسیستم مشخص می‌گردد، سپس ترافیک به همراه تصمیم‌گیری‌های انجام شده به زیرسیستم تصمیم‌گیر منتقل می‌شود. در این زیرسیستم که هسته اصلی سیستم پیشنهادی است، براساس نتایج بدست آمده از مرحله قبل، ترافیک را برای تحلیل به زیرسیستم‌های معتبر ارسال می‌شود. هر زیرسیستم پس از دریافت ترافیک به صورت کاملاً مستقل از زیرسیستم دیگر، کار خود را انجام می‌دهد و در نهایت نتایج را به سیستم تصمیم‌گیر برمی‌گرداند. در ادامه به تشریح هر یک از زیرسیستم‌های موجود در سیستم پیشنهادی و نحوه عملکرد آن‌ها پرداخته شده است.

بخش کنترلی

در این بخش، دو زیرسیستم شناسایی محیط و زیرسیستم تصمیم‌گیر قرار گرفته‌اند که در ادامه در مورد هریک توضیح داده شده است:

زیرسیستم شناسایی محیط: همان‌طور که در بخش اول بیان شد، یکی از چالش‌های مطرح در تشخیص، عوامل محیطی هستند، جهت مدیریت شرایط و انتخاب روش مناسب، زیرسیستم شناسایی محیط تعبیه شده است. این زیرسیستم وظیفه اعتبارسنجی زیرسیستم‌های تشخیص را برعهده دارد. پس از بدست آوردن اطلاعاتی پیرامون ترافیک جمع‌آوری شده، در مورد اعتبار اطلاعات هر لایه، تصمیم‌گیری می‌شود. به‌عنوان مثال، در صورت وجود کارگزار HTTP در شبکه، با توجه به ناکارآمدی زیرسیستم کاربرد در این شرایط، به سیستم تصمیم‌گیر اطلاع می‌دهد که مقادیر مربوط به لایه کاربرد نامعتبر هستند، در نتیجه سیستم تصمیم‌گیر داده‌ها را به زیرسیستم کاربرد ارسال نمی‌کند. زیرسیستم شناسایی محیط می‌تواند اطلاعات را به‌صورت دستی، از طریق کاربر و یا به‌صورت خودکار، از ترافیک ورودی بدست بیاورد.

زیرسیستم تصمیم‌گیر: زیرسیستم تصمیم‌گیر که هسته اصلی چارچوب پیشنهادی است، وظیفه مدیریت ارسال ورودی و خروجی به زیرسیستم‌های تشخیص را برعهده دارد. یکی از ساده‌ترین و در عین‌حال معتبرین روش‌ها، تشخیص توسط زیرسیستم پیوند داده است. در این زیرسیستم، از اطلاعات آدرس فیزیکی برای تشخیص بهره گرفته می‌شود. اطلاعات مربوط به این زیرسیستم در صورت معتبر شناخته شدن در اکثر موارد قابل اطمینان هستند. لذا برای عامل‌های با اعتبار قابل قبول، نیازی به تشخیص در سایر لایه‌ها نخواهد بود. در نتیجه زیرسیستم تصمیم‌گیر پس از دریافت اطلاعات از زیرسیستم شناسایی محیط، درصورت معتبر بودن لایه پیوند داده؛ ترافیک را به زیرسیستم پیوند داده منتقل می‌کند و منتظر خروجی این زیرسیستم می‌ماند. پس از دریافت نتایج از زیرسیستم پیوند داده، میزان اعتبار عامل‌ها را مورد بررسی قرار داده و گزارش تشخیص مربوط به عامل‌های با اعتبار قابل قبول را به خروجی منتقل می‌کند. سپس به منظور تشخیص دقیق‌تر سایر عامل‌ها، لیست عامل‌های با اعتبار کمتر را به زیرسیستم‌های تشخیص معتبر دیگر ارسال می‌کند. سپس هر زیرسیستم، روند تشخیص را انجام داده و در نهایت نتیجه را به زیرسیستم تصمیم‌گیر برمی‌گرداند. در نهایت، زیرسیستم تصمیم‌گیر براساس خروجی هریک از زیرسیستم‌ها، در مورد هویت هر عامل، گزارش می‌دهد.

بخش تشخیص

در این بخش زیر سیستم‌های تشخیص براساس لایه‌های شبکه قرار گرفته‌اند. در هر زیرسیستم برحسب نیاز می‌توان چندین تابع (الگوریتم) تشخیص تعریف نمود. عملکرد موازی و مستقل زیرسیستم‌های تشخیص، سبب افزایش کارایی چارچوب پیشنهادی می‌شود.

زیرسیستم پیوند داده: در صورتی که بسته‌های دریافتی از نوع پروتکل IEEE ۸۰۲,۱۱ باشند و در شبکه عمل مسیریابی اعمال نشده باشد؛ می‌توان به آدرس‌های فیزیکی بسته‌ها اعتماد کرد. در این سطح، تشخیص از روی سه بایت اول آدرس فیزیکی انجام می‌شود. سه بایت اول آدرس MAC طبق استاندارد RFC ۱۷۰۰ شامل نام سازندگان دستگاه‌ها است که می‌تواند در تشخیص مؤثر باشد [۳۹]. با توجه به این که برخی از تولیدکنندگان به صورت انحصاری تجهیزات موبایل تولید می‌کنند؛ در این زیرسیستم منوط به معتبر بودن مقادیر در شرایط محیطی و ساختاری شبکه، می‌توان عامل‌های موبایلی و غیرموبایلی را با اطمینان بالا شناسایی نمود.

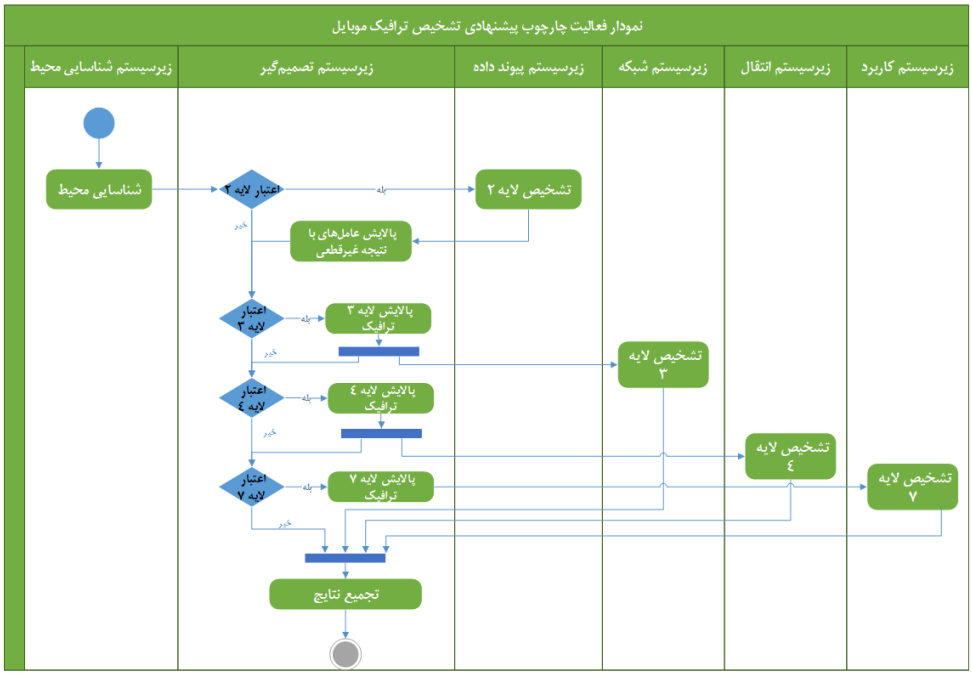
زیرسیستم شبکه: در صورت مؤثر شناخته شدن لایه شبکه توسط زیرسیستم شناسایی، بسته‌های مربوط به هر عامل پالایش شده و وارد این زیرسیستم می‌شود. در این زیرسیستم، در صورت تشخیص مؤثر بودن لایه شناسایی محیط، براساس اطلاعات سرآیند IP و الگوریتم تشخیص تعیین شده در این زیرسیستم تشخیص انجام می‌شود.

زیرسیستم انتقال: در صورت مؤثر شناخته شدن لایه انتقال توسط زیرسیستم شناسایی، همانند لایه شبکه، بسته‌های مربوط به هر عامل را پالایش کرده و وارد زیرسیستم انتقال می‌کند. هرچند در این لایه معمولاً از اطلاعات سرآیند بسته‌های TCP و UDP برای تشخیص استفاده می‌شود. با این حال در این بخش برحسب نوع الگوریتم تعیین شده، تشخیص انجام می‌شود.

زیرسیستم کاربرد: در صورت مؤثر شناخته شدن لایه کاربرد توسط زیرسیستم شناسایی محیط، با توجه به نوع پروتکل‌ها و برنامه‌های کاربردی موجود در ترافیک تشخیص انجام می‌شود.

روابط و فاکتورهای تعریف‌شده

اطلاعات مدنظر در روش‌های تحلیل ترافیک همیشه در ترافیک موجود نیست، به‌عنوان مثال ممکن است ترافیک رمزشده باشد و یا ترافیک جریان[۱۷] در اختیار باشد، در این شرایط محققان از روش‌های آماری با استفاده از دسته‌بندهای یادگیری ماشین استفاده می‌کنند. با این حال در چارچوب پیشنهادی نشان خواهیم داد که برای داشتن روشی با دقت بالا لزوماً نیازی به یک دسته‌بند یادگیری ماشین نیست. در چارچوب پیشنهادی به منظور کاهش چالش‌های مطرح در روش‌های تشخیص، برای تجمیع اطلاعات لایه‌ها از رویکرد جدیدی در این حوزه با بهره‌گیری از مفهوم اعتماد استفاده شده است. البته استفاده از



شکل ۱. نمودار فعالیت چارچوب پیشنهادی

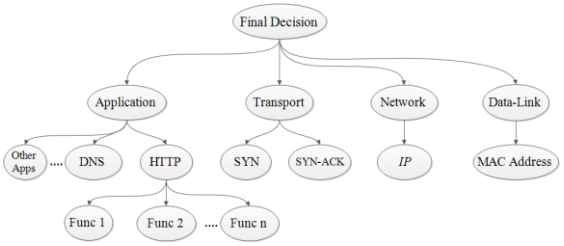
[۱۷] Flow

روابط اعتماد به این معنا نیست که در چارچوب پیشنهادی، نتوان از مزایای روش‌های یادگیری ماشین استفاده کرد. در این چارچوب امکان استفاده از دسته‌بندها وجود دارد. در ادامه به معرفی بیشتر این رویکرد پرداخته شده است.

مدیریت اعتماد در چارچوب پیشنهادی

مدیریت اعتماد شامل روش‌های محاسبه و تصمیم‌گیری افراد، درباره میزان اتکا به تراکنش‌های دارای ریسک است [۴۰]. در چهار زیرسیستم کاربرد، انتقال، شبکه و پیوند داده به عنوان چهار فرد اعتمادشونده، براساس رفتارهای ترافیکی عامل‌های ورودی، تصمیم‌گیری می‌کنند و در نهایت نظر نهایی در مورد عامل‌ها توسط فرد اعتمادکننده که زیرسیستم تصمیم‌گیر است، گرفته می‌شود. اگر از دید کلی به سیستم نگاه شود، جامعه اعتماد چارچوب پیشنهادی، شامل زیرسیستم تصمیم‌گیر به‌عنوان فرد اعتمادکننده و زیرسیستم‌های تشخیص و زیرسیستم شناسایی محیط، به‌عنوان افراد اعتمادشونده تعریف می‌شوند. به‌طور مشابه در هر زیرسیستم نیز افراد اعتمادشونده و اعتمادکننده، تعریف می‌شوند. هر یک از زیرسیستم‌ها، فرد اعتمادکننده و هر یک از توابع موجود در آن زیرسیستم، افراد اعتمادشونده هستند. این سلسله‌مراتب می‌تواند در چند سطح ادامه پیدا کند. ارتباط میان افراد جامعه اعتماد در چارچوب پیشنهادی را می‌توان با شبکه اعتماد[۱۸] نشان داد. نمونه‌ای از اجزای جامعه چارچوب پیشنهادی به صورت شبکه اعتماد در شکل ۲ نمایش داده شده است. همان‌طور که در شکل نشان داده شده است، رئوس گراف، زیرسیستم‌ها و توابع سیستم هستند که به‌عنوان افراد جامعه اعتماد و یال‌ها نشان‌دهنده ارتباط میان آن‌ها است که میزان وزن یال‌ها با استفاده از فاکتورهای اعتماد که در بخش فاکتورهای تشخیص آمده است، محاسبه خواهند شد. به‌عنوان نمونه در شکل ۲، زیرسیستم تشخیص (تصمیم‌گیرنده نهایی) به نظر چهار زیرسیستم کاربرد، انتقال، شبکه و پیوند داده اعتماد می‌کند. هریک از زیرسیستم‌های تشخیص نیز خود توابعی دارند که به‌عنوان رئوس لایه پایینی به آن‌ها متصل شده‌اند. بسته به نوع الگوریتم انتخابی در هر لایه تعداد سطوح و گره‌ها متغیر خواهد بود. به‌عنوان مثال زیرسیستم پیوند داده به تابع تشخیص آدرس فیزیکی اعتماد خواهد کرد.

در یک سیستم مدیریت اعتماد در صورتی که نظر دو یا چند نفر با هم تجمیع می‌شوند، عدم قطعیت کمتر از زمانی است که فقط یکی از نظرات درنظر گرفته شود، لذا در سیستم پیشنهادی، استفاده از تجمیع نظرات چهار زیرسیستم، سبب کاهش عدم قطعیت و افزایش اطمینان به میزان اعتماد محاسبه شده و در نتیجه افزایش دقت تشخیص خواهد شد. در ادامه، هریک از فاکتورهای مؤثر در تشخیص و نحوه محاسبه روابط در هر زیرسیستم معرفی شده‌اند. جدول ۱ لیست متغیرهای تعریف شده در روابط پیشنهادی را نشان می‌دهد.



شکل ۲. شبکه اعتماد چارچوب پیشنهادی.

جدول ۱. لیست متغیرهای تعریف شده.

متغیر	توصیف
S_i	تعداد رکوردهای تشخیص وضعیت i $(i \in \{-1,0,\frac{1}{2},1\})$
S_{ji}	تعداد رکوردهای تشخیص وضعیت i برای تابع j است.
T_i	فاکتور اعتماد محاسبه‌شده برای شاخص i (تابع یا زیرسیستم)
C_i	فاکتور اطمینان محاسبه‌شده برای شاخص i (تابع یا زیرسیستم)
x	شاخص کاربر
M	تعداد کل کاربران
P_{Funx}	تعداد حالات موفق به‌ازای تابع Func و کاربر x
N_{Funx}	تعداد حالات شکست به‌ازای تابع Func و کاربر x
P_{Funm}	تعداد حالات موفق به‌ازای تابع Func و کل کاربران
N_{Funm}	تعداد حالات شکست به‌ازای تابع Func و کل کاربران
WT_{Func}	وزن فاکتور دسترسی‌پذیری به‌ازای تابع Func
n	تعداد کل رکوردهای به‌ازای یک کاربر
WC_{func}	وزن فاکتور اطمینان به‌ازای تابع Func

فاکتورهای تشخیص

خروجی هریک از توابع موجود در زیرسیستم‌های تشخیص، به ازای هر عامل، تک بیتی تشخیص است که شامل یکی از موارد جدول ۲ است.

جدول ۲. لیست موارد بیت تشخیص.

وضعیت	توصیف
S = 1	موبایل بودن عامل
S = 0	غیرموبایل بودن عامل
S = 0.5	مورد مشابه و مشترک یافت‌شده موبایلی و غیرموبایلی
S = -1	عدم تشخیص (احتمالاً به‌علت نبود اطلاعات کافی، ضعف مجموعه‌داده و یا ضعف روش تشخیص)

در هر زیرسیستم تشخیص، زمانی که یک تابع، خروجی تشخیص را به زیرسیستم اعتمادکننده برمی‌گرداند، زیرسیستم اعتمادکننده، یک رکورد شامل مشخصه عامل (آدرس IP و

[۱۸] Web of trust

آدرس MAC)، نام تابع، وضعیت و تعداد تکرار گزارش شده را به جدول ارزشیابی خود اضافه می‌کند. با توجه به‌این‌که در یک شبکه بی‌سیم، پویایی کاربران بالا است و امکان تخصیص یک آدرس IP به چندین کاربر وجود دارد. شناسه هر عامل با زوج آدرس IP و آدرس فیزیکی معرفی می‌گردد. جدول ۳ نمونه‌ای از جدول ارزشیابی زیرسیستم کاربرد را نشان می‌دهد.

جدول ۳. نمونه‌ای از جدول ارزشیابی زیرسیستم کاربرد.

عامل	تابع	وضعیت	تکرار
(172.24.87.30,40:cb:a8:d4:4a:b)	Fuc1	۱	۱۰
(172.24.87.30,40:cb:a8:d4:4a:b)	Fuc1	۰	۴
(172.24.87.30,40:cb:a8:d4:4a:b)	Fuc2	-۱	۲
(172.24.87.30,40:cb:a8:d4:4a:b)	Fuc3	۰,۵	۲

براساس رفتارهای مشاهده شده از هر عامل، با کمک جدول ارزشیابی، دو فاکتور اعتماد با نماد T و اطمینان با نماد C محاسبه می‌شوند. در نهایت براساس نظرات توابع و مقادیر اعتماد و اطمینان، میانگین‌گیری وزنی صورت گرفته و در نهایت خروجی هر زیرسیستم به ازای هر عامل به صورت سه‌تایی (S,T,C) به سیستم تصمیم‌گیر ارسال می‌گردد. در ادامه به معرفی فاکتورهای تعریف‌شده پرداخته شده است.

فاکتور اعتماد: این فاکتور شامل اعتماد به اطلاعات موجود در هر لایه، براساس مقادیر و امضاهای از پیش تعریف شده است که امکان تشخیص دقیق را فراهم می‌آورد. جهت محاسبه این فاکتور از توزیع بتا استفاده شده است [۴۱]. رابطه ۱ نحوه محاسبه فاکتور اعتماد به ازای هر مؤلفه را نشان می‌دهد. در هر زیرسیستم با توجه به مدارک در دسترس، تعداد حالات موفق (P_{Funx}) و شکست (N_{Funx}) به‌ازای تابع مورد نظر (Func) و به‌ازای عامل x محاسبه می‌شود. موفقیت به معنای تشخیص و شامل حالات S_1, S_0, $S_{0.5}$ و شکست به‌معنای عدم وجود مقدار مناسب برای تشخیص و شامل حالت S_{-1} است. در صورتی که در زیرسیستم از چندین تابع استفاده شده باشد، جهت محاسبه مقدار نهایی فاکتور اعتماد، مطابق رابطه ۳ از میانگین‌گیری وزنی استفاده می‌شود. جهت وزن‌دهی هر تابع با استفاده از رابطه ۲ براساس در دسترس بودن هر مؤلفه در کل ترافیک، مقدار وزنی wt_j تابع j محاسبه شده است. با این کار رتبه هر تابع براساس سطح اعتماد هر مؤلفه و تأثیر آن بر اعتماد کلی هر زیرسیستم بدست می‌آید. در رابطه ۲ P_{funm} تعداد موفقیت‌ها و N_{funm} تعداد شکست‌های همه عامل‌های موجود در ترافیک است.

$$T_{funx} = \frac{P_{funx} + 1}{P_{funx} + N_{funx} + 2} \quad (۱)$$

$$wt_{fun} = \frac{P_{funm} + 1}{P_{funm} + N_{funm} + 2} \quad (۲)$$

$$T_{layer} = \frac{\sum_{j \in fun} T_j . wt_j}{\sum_{j \in fun} wt_j} \quad (۳)$$

فاکتور اطمینان: فاکتور اطمینان نشان‌دهنده میزان معتبر بودن و ثبات مقادیر موجود در ترافیک، عدم تغییر مقدار اصلی مبدأ، میزان جعلی بودن و دستکاری کردن مقادیر و در واقع میزان باور به درستی اعتماد محاسبه شده است [۴۲]. اگر فرض شود، که هر عامل موبایلی، تعدادی برنامه کاربردی (با قابلیت اتصال به اینترنت) بر روی دستگاه موبایلی خود داشته باشد، احتمال این‌که هر یک از برنامه‌ها از اطلاعات خاصی برای معرفی عامل خود استفاده کنند، زیاد است. این مسئله باعث عدم ثبات در اطلاعات مربوط به هر عامل به‌خصوص در لایه کاربرد می‌شود. همچنین اگر از این تعداد برنامه کاربردی متصل به اینترنت، تعداد r برنامه سالم و q برنامه مخرب باشند، با فرض این‌که تعداد $q < r$ باشد، در صورتی که برنامه آلوده اقدام به تغییر اطلاعات هویتی خود کرده باشد، با استفاده از تجمیع اطلاعات بدست آمده از هر عامل در اتصالات مختلف می‌توان به صحت اطلاعات کاربر پی برد. این ناسازگاری در لایه‌های دیگر TCP/IP نیز ممکن است اتفاق بیافتد.

در هر زیرسیستم همه توابع می‌بایست به یک نتیجه واحد، مبنی‌بر موبایل بودن یا نبودن هر عامل رسیده باشند. در غیراین صورت، مسئله عدم قطعیت در تشخیص بوجود می‌آید. میزان اطمینان اطلاعات هر عامل به ازای هر تابع مطابق رابطه ۴ تعریف می‌شود. در این رابطه S_0، S_1 و $S_{1/2}$ به ترتیب شامل تعداد رکوردهای تشخیص داده شده موبایلی، غیرموبایلی و موارد مشترک و n تعداد کل رکوردهای مربوط به هر عامل است.

همان‌طور که مقدمه بیان شد، یکی از چالش‌های مطرح در تشخیص تشابه است. در صورتی که یک امضا با دو سیستم‌عامل متفاوت تطبیق یابد، اطمینان به نصف کاهش یافته و برابر با ۰,۵ خواهد شد. در غیر این صورت، مقدار اطمینان برابر با یک درنظر گرفته می‌شود. مطابق رابطه ۶ برای ارزیابی اطمینان کل هر زیرسیستم می‌بایست، اختلاف بین مجموع همه حالات S_0 و تمام حالات S_1 و $S_{1/2}$ را با در نظر گرفتن وزن هر تابع (wc_j)، برای عامل مورد نظر محاسبه نمود. در رابطه ۶، Sj_0 تعداد موارد موبایلی، Sj_1 تعداد موارد غیرموبایلی و $Sj_{1/2}$ تعداد موارد مشترک در تابع j، wc_j وزن محاسبه شده تابع j و n تعداد رکوردهای

موجود به ازای هر تابع است. از آنجایی که هرچه پراکندگی مقادیر S کمتر باشد، اطمینان بیشتر است، مطابق رابطه ۵ برای محاسبه وزن اعتبار هر تابع از متمم واریانس اعتبار به ازای همه عامل‌های موجود در ترافیک (m) بهره گرفته شده است.

$$C_{funx}=\frac{\left|\left|S_0-S_1\right|-\frac{1}{2}S_{1/2}\right|}{n},\ n=S_0+S_1+S_{0.5}+S_{-1} \quad (۴)$$

$$wc_{fun}=1-\sum_{x\in m}(C_{funx}-\overline{C}_{funx}) \quad (۵)$$

$$C_{layer}=\frac{\left|\sum_{j\in fun}S_{j0}.wc_j-\sum_{j\in fun}S_{j1}.wc_j\right|-\frac{1}{2}\sum_{j\in fun}S_{j1/2}.wc_j}{\sum_{j\in fun}n.wc_j} \quad (۶)$$

روابط بیان شده ۱-۶ برای محاسبه میزان اعتماد و اطمینان زیرسیستم‌های تشخیص شبکه، انتقال و کاربرد مورد استفاده قرار می‌گیرد. شرایط محاسبات در زیرسیستم پیوند داده، متفاوت است، از آنجایی که تشخیص براساس آدرس فیزیکی انجام می‌شود، لذا در صورت معتبر شناخته‌شدن این زیرسیستم، در صورت تشخیص کامل عامل موبایل یا غیرموبایلی، اطمینان برابر با یک و در صورت وجود مورد مشترک در میان تولیدکنندگان موبایل و سایر تجهیزات، مقدار اطمینان برابر با ۰٫۵ لحاظ می‌شود. بنابراین همان‌طور که قبلاً هم بیان شد، زیرسیستم تصمیم‌گیر، عامل‌های با مقدار اطمینان ۰٫۵ را برای تشخیص دقیق‌تر، به زیرسیستم‌های تشخیص دیگر ارسال می‌کند. با این کار تعداد عامل‌های کمتری به زیرسیستم‌های دیگر هدایت می‌شوند. این مسئله باعث کاهش محاسبات در زیرسیستم‌های دیگر و کاهش سربار محاسباتی چارچوب پیشنهادی می‌شود.

تصمیم نظر نهایی در هر زیرسیستم: جهت محاسبه تصمیم نهایی بیت تشخیص کافی است حداکثر حالات صفر، یک و ۰٫۵ با احتساب وزن آن‌ها را بدست آورد، هر کدام از مجموعه‌ها که مقدار بیشتری داشته باشند، به عنوان بیت تشخیص انتخاب می‌شود. رابطه ۷ انتخاب بیت تشخیص در این زیرسیستم را نشان می‌دهد.

$$S_{layer}=Max(\sum_{j\in fun}S_{j1}.wc_j+\frac{1}{2}\sum_{j\in fun}S_{j1/2}.wc_j,\sum_{j\in fun}S_{j0}.wc_j) \quad (۷)$$

محاسبه تصمیم نهایی: پس از عمل تشخیص هر زیرسیستم سه‌تایی (S,T,C) که به ترتیب شامل بیت تشخیص، اعتماد و اطمینان محاسبه شده برای هر عامل است، به زیرسیستم تصمیم‌گیر (فرد اعتمادکننده) ارسال می‌شود. زمانی که کلیه اطلاعات مربوط به هر عامل ثبت شده باشد، تصمیم نهایی در مورد آن عامل گرفته می‌شود. در این زیرسیستم مطابق رابطه ۸، با درنظر گرفتن میزان اطمینان به عنوان وزن اعتماد محاسبه شده برای هر نظر، به منظور تجمیع نظرات از میانگین‌گیری وزنی استفاده می‌کند. طبق رابطه ۹ تصمیم نهایی به ازای هر عامل براساس مقدار بیت تشخیص و وزن مربوط به هر یک از زیرسیستم‌ها انجام می‌گردد.

$$T_{final}=\frac{\sum_{i\in subsystem}T_i.C_i}{\sum_{i\in subsystem}C_i} \quad (۸)$$

$$S_{final}=Max(S_7.C_7,S_4.C_4,S_3.C_3S_2.C_2) \quad (۹)$$

ارزیابی

در این بخش به ارزیابی چارچوب پیشنهادی براساس ترافیک واقعی، پرداخته شده است. به‌همین منظور مجموعه‌داده یادگیری‌شده از ترافیک موبایل تهیه شده است. در ادامه به معرفی مشخصات مجموعه‌داده پرداخته و چارچوب پیشنهادی، مورد ارزیابی قرار گرفته است.

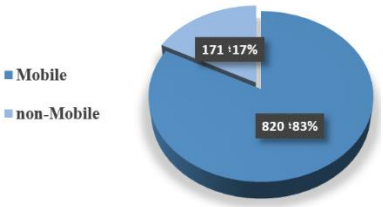
معرفی مجموعه‌داده

به منظور ایجاد مجموعه‌داده مناسب ترافیک واقعی شبکه بی‌سیم خوابگاه دانشگاه فردوسی مشهد با استفاده از روش شماره درگاه SPAN[۱۹] با تنوع تجهیزات موبایل و غیرموبایلی و با حجم ۵٫۸ گیگابایت جمع‌آوری شده است. اطلاعات مربوط به هر چهار سطح در ترافیک شبکه معتبر هستند. با توجه به تعداد بسیار بالای کاربران و عدم اطلاع از هویت موبایلی یا غیرموبایلی آن‌ها، به منظور ارزیابی دقیق و برچسب‌گذاری عامل‌های موجود در ترافیک، از آدرس‌های فیزیکی آن‌ها استفاده شده است. براساس آدرس‌های فیزیکی عامل‌ها، شرکت‌های سازنده هریک از استخراج شده، سپس عامل‌های با اطمینان بالا (C=1) شناسایی و کل ترافیک براساس آدرس‌های فیزیکیِ تفکیک‌شده، پالایش و برچسب‌گذاری شده است. نمودار ۱ آماری از میزان کاربران موبایلی و غیرموبایلی در ترافیک پالایش‌شده را نشان می‌دهد. این آمار نشان می‌دهد که ۸۳٪ کاربران موجود در شبکه، کاربران موبایل بوده‌اند. در این مجموعه‌داده از میان ۹۹۱ کاربر موجود در ترافیک، ۶۶۹ کاربر دارای ترافیک وب و ۸۸۲ کاربر دارای ترافیک TCP/IP بوده‌اند.

[۱۹] Switched Port Analyzer

جهت تحلیل بهتر و بدست آوردن اطلاعات مورد نیاز از پایگاه‌داده‌های معتبر زیر استفاده شده است:

- لیست تولیدکنندگان جهت بدست آوردن نام تولیدکنندگان از روی آدرس فیزیکی آن‌ها [۴۳].
- لیست سیستم‌عامل‌ها از پایگاه‌داده fingerbank برای شناسایی سیستم‌عامل رشته عامل کاربر [۴۴].
- لیست امضاهای لایه TCP/IP ابزار P0f نسخه b۳,۰۹



نمودار ۱. آمار کاربران موجود در ترافیک

پارامترهای ارزیابی چارچوب پیشنهادی

همان‌طور که بیان شد، جهت برچسب‌گذاری عامل‌های موجود در ترافیک از آدرس‌های فیزیکی آن‌ها استفاده شده است، لذا جهت ارزیابی دقیق‌تر چارچوب پیشنهادی، اطلاعات زیرسیستم پیوند داده درنظر گرفته نشده است. با توجه به اینکه مجموعه‌داده ایجاد شده در داخل شبکه WLAN قبل از عبور از هر گونه تجهیز شبکه، جمع‌آوری شده است لذا اطلاعات لایه شبکه، انتقال و کاربرد آن معتبر است. بدین ترتیب امکان ارزیابی چارچوب پیشنهادی در سه زیرسیستم کاربرد، انتقال و شبکه مهیا شده است. نحوه ارزیابی هریک از زیرسیستم‌های تشخیص در ادامه بیان شده است.

زیرسیستم شبکه و انتقال: برای ارزیابی این دو لایه از امضاهای موجود در ابزار P0f استفاده شده است.

زیرسیستم کاربرد: به‌منظور ارزیابی زیرسیستم کاربرد، از تکنیک بازرسی عمیق بسته‌ها و بررسی اطلاعات موجود در سرآیند HTTP، براساس برخی از کلمات کلیدی موبایل، استفاده شده است، عامل‌های مورد بررسی در این زیرسیستم، شامل موارد زیر است.

- **رشته عامل کاربر (UA)**: مقدار این فیلد در اکثر بسته‌های درخواست HTTP یافت می‌شود، لذا به‌عنوان یک عامل مستقیم موثر در این زیرسیستم تعریف می‌شود.
- **مؤلفه‌های غیراستاندارد (NS)**: در برخی از بسته‌های درخواست HTTP، برنامه‌های کاربردی سیستم‌عامل‌ها اطلاعاتی از جمله مشخصات خود را برای سرویس‌دهندگان ارسال می‌کنند. این اطلاعات برای تشخیص موبایل بودن بسیار مفید هستند. مقادیری چون x-wap-profile و x-device-user-agent از جمله مقادیر غیراستاندارد در سرآیند HTTP هستند.
- **تابع GET**: یکی دیگر از مواردی که به احتمال زیاد در بسته‌های درخواست که با روش GET ارسال شده‌اند، اطلاعات موجود در URL بسته‌هاست.
- **تابع POST**: برخی از بسته‌های درخواست از نوع POST، قطعه کدهای CSS، جاوااسکریپت و غیره را همراه با سرآیند HTTP حمل می‌کنند. در این قطعه کدها می‌توان موارد مناسبی از جمله طول و عرض صفحه نمایش، اندازه و نوع قلم استفاده شده و حتی کلمات کلیدی را استخراج نمود.

براساس روابط ۱۰- ۱۲ پارامترهای حساسیت[۲۰]، دقت[۲۱] و میزان نرخ نادرست مثبت[۲۲] محاسبه شده است. توصیف مقادیر موجود در این روابط در جدول ۴ بدست آمده‌اند.

جدول ۴. مقادیر مربوط به روابط ارزیابی.

مقدار	توصیف
TP	تعداد عامل‌های موبایلی که درست تشخیص داده شده
FN	تعداد عامل‌های موبایلی که غیرموبایلی تشخیص داده شده
FP	تعداد عامل‌های غیرموبایلی که موبایل تشخیص داده شده
TN	تعداد عامل‌های غیرموبایلی که غیرموبایل تشخیص داده شده

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (۱۰)$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (۱۱)$$

$$FPR = \frac{FP}{FP + TN} \qquad (۱۲)$$

نحوه ارزیابی

پس از اعمال روش‌های تشخیص در هر زیرسیستم، براساس خروجی مربوط به هر لایه، فاکتورهای اعتماد و اطمینان هر عامل محاسبه شده و در نهایت مقدار بیت تشخیص نهایی و اعتماد نهایی به ازای هر عامل محاسبه شده است.

همان‌طور که در ابتدای این بخش گفته شد، ارزیابی سیستم پیشنهادی در سه لایه شبکه، انتقال و کاربرد انجام شده است. با این حال جهت محاسبه میزان دقت و خطای روش پیشنهادی، از تشخیص زیرسیستم پیوند داده با عامل‌های با میزان اعتماد کامل (T=1) بهره برده شده است. بدین ترتیب عامل‌های با اعتماد

[۲۰] Recall
[۲۱] Precision
[۲۲] False Positive Ratio

کامل براساس موبایل بودن یا نبودن، برچسب‌گذاری شده‌اند و پس از تشخیص نهایی عامل‌ها براساس نظرات سه زیرسیستم تشخیص (کاربرد، انتقال و شبکه)، مقایسه‌ای میان خروجی سیستم پیشنهادی و لیست برچسب‌گذاری شده انجام شده و در نهایت میزان هریک از فاکتورهای خطا براساس مقادیر جدول ۴ محاسبه شده است. در ادامه نتایج ارزیابی آورده شده است.

نتایج ارزیابی و مقایسه با کارهای مشابه

جدول ۵ مقایسه ویژگی‌های دو مورد از مرتبط‌ترین کارهای انجام شده در این حوزه با چارچوب پیشنهادی را نشان می‌دهد. دو کار انتخابی هر دو در حالت غیرفعال تشخیص را براساس پارامترهای لایه‌های شبکه انجام داده‌اند.

ارزیابی‌های صورت گرفته بر روی چارچوب پیشنهادی نشان می‌دهد که میزان دقت برابر با ۰,۹۷۷، میزان حساسیت برابر با ۰,۸۹۹ و میزان نرخ نادرست مثبت برابر با ۰,۰۹ است. این آمار و ارقام به نسبت روش‌های مشابه نتایج بهتری را نشان می‌دهند. نتایج اعلام شده در مقاله [۱۰] برابر با ۰,۸۵۸۲ و کمتر از روش پیشنهادی است. دقت اعلام شده در مقاله [۳۰] برابر با ۰,۹۸ است. مقاله [۳۰] با این‌که دقت بالاتری به نسبت به روش پیشنهادی دارد، با این حال با توجه به این‌که میزان نرخ نادرست مثبت آن اعلام نشده و فقط به مقادیر نادرست منفی و نادرست مثبت اشاره شده و نمی‌توان به برتری آن به نسبت روش پیشنهادی اشاره نمود. همچنین این روش نقاط ضعیفی دارد که در ادامه مورد بررسی قرار گرفته شده است.

دو مقاله مرتبط تشخیص را براساس پارامترهای سه لایه شبکه، انتقال و کاربرد انجام داده‌اند. این در حالی است که در روش پیشنهادی علاوه بر اطلاعات سه لایه، اطلاعات مربوط به لایه پیوند داده را نیز اضافه کرده است. چراکه در صورتی‌که در شبکه مورد ارزیابی، اطلاعات لایه پیوند داده معتبر باشند، برای عامل‌های با میزان اعتماد کامل (T=1) نیازی به محاسبات گسترده در لایه‌های دیگر نیست، استفاده از تشخیص در زیرسیستم پیوند داده در عامل‌های با اعتماد کامل نه‌تنها باعث سربار اضافی در سیستم نمی‌شود، بلکه باعث کاهش فرآیند تشخیص در سایر روش‌ها خواهد شد. این مسئله در روش‌های مشابه در نظر گرفته نشده است.

مسئله بعدی عدم انعطاف‌پذیری سایر روش‌ها در شرایط مختلف است. در دو روش مرتبط، شرایط محیطی ترافیک درنظر گرفته نشده است. همان‌طور که در بخش اول بیان شد، در مورد معتبر بودن اطلاعات هریک از لایه‌های شبکه بسته به نوع پیکربندی شبکه مانند استفاده از کارپذیرها، ممکن است اطلاعات نامعتبر باشند، در این شرایط استفاده از روش‌های مرتبط با خطای زیاد مواجه خواهند بود. روش پیشنهادی با درنظر گرفتن همه لایه‌ها و درنظر گرفتن شرایط محیطی شبکه، از خطاهای احتمالی جلوگیری می‌کند. در واقع می‌توان گفت ارزش عملکرد چارچوب پیشنهادی زمانی خود را نشان می‌دهد که بخشی از داده‌ها معتبر نبوده و به نحوی تغییر یافته باشند و نقاط ضعف هر زیرسیستم، توسط زیرسیستم دیگر پوشانیده شود. مسئله‌ای که در سایر روش‌های مشابه توجه چندانی به آن نشده است.

هر دو روش مرتبط ترافیک رمزشده را درنظر گرفته‌اند. با این حال تمرکز اصلی کار انجام شده [۱۰] بر روی ترافیک رمزشده نبوده است. در چارچوب پیشنهادی نیز این مسئله منوط به انتخاب الگوریتم تشخیص در هر لایه است. با این حال با توجه به ویژگی چندسطحی چارچوب پیشنهادی، در صورت ضعف یک لایه در تشخیص ترافیک رمزشده، می‌توان از اطلاعات سایر لایه‌ها بهره گرفت.

جدول ۵. مقایسه روش پیشنهادی و روش‌های مرتبط.

روش / ویژگی	مقاله [۱۰]	مقاله [۳۰]	چارچوب پیشنهادی
لایه‌های تشخیص مورد استفاده در روش	لایه ۳-۷-۴	لایه ۳-۷-۴	لایه ۲-۳-۴-۷
تشخیص و درنظر گرفتن شرایط محیطی (مانند وجود کارگزارهای لایه کاربرد)	✗	✗	✓
پشتیبانی از ترافیک رمزشده	TLS	TLS	براساس الگوریتم انتخابی در هر زیرسیستم، امکان تشخیص وجود دارد.
مجموعه‌داده واقعی	✓	تنوع کم	✓
بررسی مقایسه‌ای دستگاه‌های موبایلی و غیرموبایلی	✓	✓	✓
دستکاری‌شدن اطلاعات و حملات	✗	✗	✓
بررسی تشابه میان سیستم‌عامل‌های موبایلی و غیرموبایلی	✓	✗	✓
دقت محاسبه‌شده	۰,۸۵۸۲	۰,۹۸	۰,۹۷۷
نرخ نادرست مثبت	ذکر نشده	ذکر نشده	۰,۰۹
حساسیت	۰,۶۰۴۱	ذکر نشده	۰,۸۹۹

در هر سه کار از ترافیک واقعی برای ارزیابی استفاده کرده‌اند. با این وجود ترافیک جمع‌آوری شده کار [۳۰] از تنوع بالای دستگاه‌های موبایلی و غیرموبایلی برخوردار نیست. بررسی تشابه میان سیستم‌عامل‌های موبایلی در مقاله [۱۰] عنوان شده است. ولی در زمان تجمیع روش‌ها در صورت مشاهده مورد مشابه،

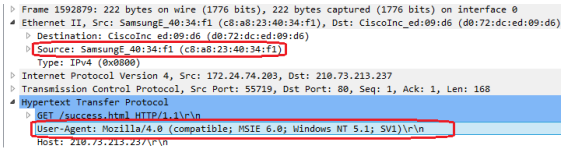
براساس اولویت به‌ترتیب رشته عامل کاربر، دامنه‌های مشخص و پارامترهای TCP/IP درنظر گرفته شده است. این درحالی است که به بحث ناسازگاری داده‌ها و امکان دستکاری اطلاعات در لایه کاربرد توجهی نشده است. با این وجود ممکن است اطلاعات کاربرد یک کاربر در دو نشست مجزا با هم ناسازگاری داشته باشند. استفاده از فاکتورهای اعتماد و اطمینان در چارچوب پیشنهادی باعث درنظر گرفتن این مورد شده است. در بخش بعد چند نمونه از این چالش‌ها مطرح شده‌اند.

بحث و گفتگو

در بخش اول در مورد برخی از چالش‌های مطرح در تشخیص ترافیک موبایل بیان شد. در این بخش چند نمونه از چالش‌های مشاهده شده در ترافیک واقعی و ضرورت کار در این حوزه نشان داده است.

تشابه

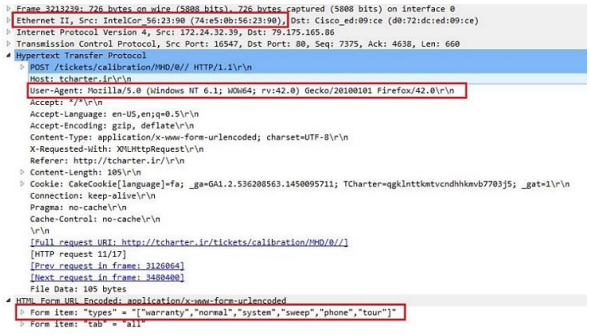
یکی از چالش‌های تشخیص، تشابه میان نشانه‌های سیستم‌عامل‌های موبایلی و غیرموبایلی است. شکل ۳ نمونه‌ای از بسته‌ای در ترافیک مربوط به یک عامل موبایلی را نشان می‌دهد، آدرس فیزیکی این عامل مربوط به یک تولیدکننده موبایل است، با این حال این عامل از یک رشته عامل کاربر مشترک ویندوزی استفاده کرده است. تشابه میان اطلاعات سیستم‌عامل‌های موبایلی و غیرموبایلی در لایه‌های شبکه و انتقال نیز خود را نشان می‌دهد. این مسئله جزء نقاط ضعف ابزار تشخیص است. حدود ۹۰ درصد از کل سیستم‌عامل‌های اندرویدی موجود در مجموعه‌داده تولید شده در ابزار P0f لینوکس گزارش شده‌اند.



شکل ۳. نمونه عامل موبایلی با رشته عامل کاربر ویندوزی

ناسازگاری

ناسازگاری در داده‌های موجود در ترافیک هم در لایه کاربرد و هم در دو لایه انتقال و شبکه خود را نشان می‌دهند. به‌عنوان نمونه شکل ۴ نمونه عامل غیرموبایلی حاوی اطلاعات موبایلی را نشان می‌دهد.



شکل ۴. نمونه عامل غیرموبایلی حاوی اطلاعات موبایلی

مواردی از عامل‌ها از خروجی این ابزار دیده شده است که هرچند در بخش کاربرد، این ابزار عامل را موبایل شناسایی نموده است، با این حال، در لایه انتقال و شبکه این عامل به درستی تشخیص داده نشده است و برعکس. جدول ۶ سه نمونه از این ناسازگاری را در چهار زیرسیستم نشان می‌دهد.

جدول ۶. ناسازگاری در لایه TCP/IP و کاربرد.

مورد	زیرسیستم پیونده داده	زیرسیستم انتقال و شبکه	زیرسیستم کاربرد
۱	غیرموبایل	موبایل	غیرموبایل
۲	غیرموبایل	غیرموبایل	موبایل
۳	موبایل	غیرموبایل	موبایل

دستکاری اطلاعات

یکی از چالش‌های مطرح در زمینه تشخیص سیستم‌عامل‌ها و برنامه‌های کاربردی، دستکاری عمدی اطلاعات است. به این مورد در روش‌های اخیر توجه چندانی به این مسئله نشده است. این مسئله در ترافیک منجر به ناسازگاری در داده‌ها شده و درصورت عدم توجه به این موضوع باعث کاهش دقت روش تشخیص خواهد شد.

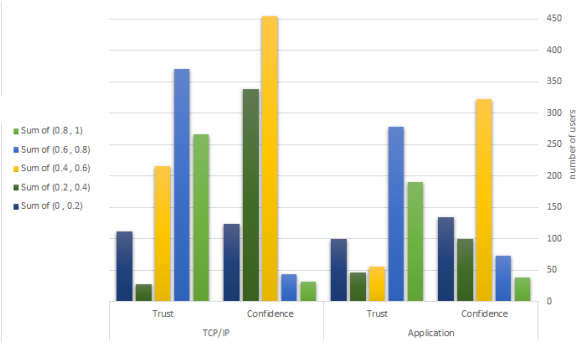
سیستم‌عامل هدف

اکثر کارهای انجام شده با تمرکز بر روی ترافیک دستگاه‌های موبایلی بوده‌اند و در ارزیابی خود مقایسه میان سیستم‌عامل‌های غیرموبایلی و شباهت‌های میان آن‌ها را در نظر نگرفته‌اند. این درحالی است که ممکن است روشی که خاص تشخیص ترافیک موبایل ارائه شده است برای ترافیک غیرموبایلی هم صدق کند. به‌علاوه کارهای اخیر به سمت سیستم‌های اینترنت اشیاء نیز سوق داده شده‌اند. در این مقاله سیستم‌عامل هدف موبایل درنظر گرفته شد چرا که همچنان چالش‌های مربوط به آن باقی است و همچنین فراهم کردن محیط آزمون برای سیستم‌های اینترنت اشیاء در حجم زیاد فراهم نبوده است. با این حال با توجه به‌این‌که دستگاه‌های اینترنت اشیاء نیز از پشته شبکه یکسانی با سایر

سیستم‌ها استفاده می‌کنند، می‌توان از ساختار چارچوب پیشنهادی برای تشخیص سیستم‌های اینترنت اشیاء بهره گرفت.

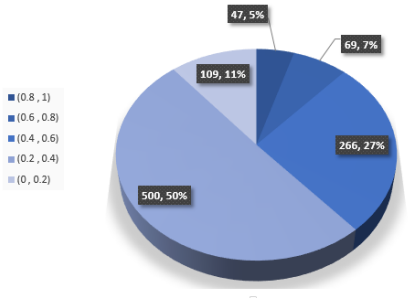
میزان اعتماد و اطمینان به اطلاعات

شمای کلی از نتایج حاصل از محاسبه دو فاکتور دسترسی‌پذیری و اعتبار در سه زیرسیستم کاربرد، انتقال و شبکه را می‌توان در نمودار ۲ مشاهده نمود. محور عمودی بیان‌گر تعداد کاربران شناسایی‌شده و محور افقی نشان‌دهنده میزان فاکتور اعتماد و اطمینان محاسبه‌شده به ازای سه زیرسیستم کاربرد و پشته TCP/IP است. برای نمایش بهتر داده‌ها در نمودار مقادیر محاسبه شده فاکتورها، به‌صورت بازه‌ای و با رنگ‌های مختلف نشان داده شده است. از این نمودار نتایج زیر را استنباط می‌شود:



نمودار ۲. مقادیر فاکتورهای تشخیص در زیرسیستم‌ها

- آمار تعداد عامل‌های تشخیص‌داده شده در لایه کاربرد کمتر از دو لایه دیگر است. این مسئله نشان‌دهنده اطلاعات در دسترس کمتر لایه کاربرد به نسبت دو لایه دیگر است.
- میزان اعتماد محاسبه‌شده در هر سه لایه، برای بیش از نیمی از عامل‌ها، بالای ۰٫۵ بوده است. این بدین معناست برای سایر عامل‌ها (حدود ۳۰ درصد ترافیک) داده‌های مناسب برای تشخیص در دسترس نبوده است.
- میزان اطمینان محاسبه‌شده در دو لایه شبکه و انتقال به نسبت لایه کاربرد بیش‌تر است. این بدین معناست که در ترافیک جمع‌آوری شده ناسازگاری داده‌ها در لایه کاربرد بیش‌تر بوده است. احتمال دستکاری‌شدن داده‌ها در لایه کاربرد بیش‌تر است.
- پایین‌بودن مقادیر این دو فاکتور نشان از چالش‌های عدم در دسترس بودن اطلاعات و ناسازگاری در داده‌های موجود در ترافیک واقعی است. این چالش‌ها به‌طور خاص در روش‌های اخیر مورد توجه قرار نگرفته‌اند.



نمودار ۳. میزان اعتماد نهایی محاسبه‌شده

نمودار ۳ آمار مربوط به محاسبه اعتماد نهایی را نشان می‌دهد. این نمودار نشان می‌دهد که میزان اعتماد نهایی به موجود در ترافیک واقعی، که شامل دسترسی‌پذیری اطلاعات و اطمینان به آن‌هاست، پایین است. بنابراین انتخاب روش تشخیص مناسب بر پایه لایه‌های مختلف شبکه بسیار حائز اهمیت و چالش‌برانگیز است. تاکنون روش‌های متعددی در این زمینه ارائه شده است. ولی نکته مهم در اکثر روش‌های ارائه شده عدم توجه به کارایی روش پیشنهادی در ترافیک واقعی در حجم بالاست. زمانی که حجم ترافیک و تنوع دستگاه‌های موبایلی و غیرموبایلی زیاد شود، دقت تشخیص روش‌ها به مراتب کاهش می‌یابند.

تمرکز این مقاله بر روی انتخاب روش تشخیص نبوده است بلکه هدف اصلی اثبات تجمیع اطلاعات و کاهش اثرات چالش‌های مطرح‌شده در شرایط مختلف و نشان دادن چالش‌های موجود بوده است. ما در این پژوهش توانسته‌ایم با تجمیع نظر زیرسیستم‌ها و درنظر گرفتن فاکتورهای اعتماد و اطمینان چالش‌هایی ذکر شده را کاهش دهیم، این مسئله را می‌توان در میزان دقت نهایی روش و میزان اعتماد نهایی مشاهده نمود.

گرچه در کارهای مرتبط دیگر مانند دو کار [۱۰] و [۳۰] از تجمیع اطلاعات سه لایه استفاده شده است ولی در این دو روش به چالش‌های ذکر شده اشاره‌ای نشده است. در کار [۱۰] از ترافیک واقعی برای ارزیابی استفاده شده است، با این حال با توجه به ضعف الگوریتم تشخیص و عدم درنظر گرفتن چالش‌ها از دقت بالایی برخوردار نیست. به‌علاوه در صورتی‌که ترافیک دستکاری شده باشد، دقت آن کاهش خواهد یافت. همان‌طور که در بخش کارهای وابسته هم بیان شد، در این روش در زمان در دسترس نبودن اطلاعات کافی، به ترتیب اولویت از رشته عامل کاربر و پارامترهای TCP/IP استفاده می‌کند. این در حالی است که نتایج بررسی‌های ما در نمودار ۲، میزان اطمینان کم اطلاعات لایه کاربرد به نسبت لایه انتقال و شبکه را نشان می‌دهد.

در مورد کار [۳۰] نیز با توجه تنوع کم مجموعه‌داده و در نظر نگرفتن چالش‌ها، امکان مقایسه صحیح فراهم نیست با این وجود واضح است که در ترافیک واقعی با حجم بالا، در صورت دستکاری شدن اطلاعات، دقت روش کاهش خواهد یافت.

در روش پیشنهادی چالش‌های مطرح شده با استفاده از فاکتورهای اعتماد و اطمینان کاهش یافته‌اند.

همان‌طور که در بخش معرفی چارچوب پیشنهادی بیان شد، خروجی نهایی سیستم برای هر عامل شامل بیت تشخیص و میزان اعتماد نهایی محاسبه شده برای آن عامل است. این نتایج به محققان کمک می‌کند تا براساس میزان اعتماد محاسبه شده در مورد هر عامل تصمیم‌گیری لازم را داشته باشند. به‌عنوان مثال پایین بودن میزان اعتماد نهایی یک عامل، ممکن است نشان از رفتار غیرعادی آن عامل باشد که بخاطر دستکاری اطلاعات اتفاق افتاده است. این مسئله در هیچ یک از روش‌های مرتبط دیگر دیده نشده است.

جمع‌بندی و کارهای آتی

با توجه به اهمیت تحلیل ترافیک موبایل به‌عنوان یکی از حوزه‌های مهم برای تحلیلگران امنیتی، در این مقاله به بررسی روش‌های تحلیل ترافیک موبایل و چالش‌های مطرح در این حوزه پرداخته شده است. براساس چالش‌های عنوان شده، جهت داشتن سیستمی مناسب برای تشخیص ترافیک موبایل در همه شرایط، چارچوبی ارائه شده است. ساختار این چارچوب از چهار زیرسیستم تشخیص و دو زیرسیستم کنترلی تشکیل شده است. در صورت معتبر بودن اطلاعات هر لایه از پشته TCP/IP تشخیص در زیرسیستم‌های تشخیص انجام می‌شود. هدف اصلی این مقاله بیان چارچوب کلی و قابل استفاده در همه محیط‌ها بوده است. برای بالا بردن دقت تشخیص در این سیستم از فاکتورهای مبتنی بر اعتماد استفاده شده است.

برای ارزیابی سیستم پیشنهادی ابتدا ترافیک بی‌سیم دانشگاه جمع‌آوری و جهت تخمین میزان دقیق سیستم، براساس آدرس‌های فیزیکی آن‌ها برچسب‌گذاری شد. سپس با کمک چندین روش متداول برای تشخیص سیستم‌عامل‌ها در سه لایه شبکه، انتقال و کاربرد استفاده شده است. نتایج حاصل از تجمیع نظرات افزایش دقت تشخیص را نشان می‌دهد. در انتها نیز برخی از چالش‌های مطرح براساس نتایج بدست آمده در ترافیک مجموعه‌داده مورد بحث قرار گرفته است. با توجه به ماهیت سیستم‌عامل‌های مختلف، تحلیل در مورد این سیستم‌عامل‌ها به‌طور مستمر ادامه خواهد یافت.

هرچند ممکن است استفاده از لایه‌های مختلف در روش تشخیص سربار به سیستم ایجاد کند ولی با مقایسه سایر روش‌های مرتبط و چالش‌های بیان شده، در نظر گرفتن لایه‌های مختلف باعث افزایش دقت چارچوب پیشنهادی شده است. به‌علاوه چارچوب پیشنهادی یک روش کلی و جامع است و درآینده می‌توان در هر یک از زیرسیستم‌های تشخیص الگوریتم‌های سبک‌وزنی ارائه نمود تا سربار سیستم را تا حد امکان کاهش یافته و مقیاس‌پذیری آن افزایش یابد.

گرچه هدف اصلی در این مقاله تشخیص سیستم‌عامل موبایلی است با این حال، از این چارچوب می‌توان برای سایر حوزه‌های تحلیل ترافیک استفاده نمود. برای این کار کافی است الگوریتم تشخیص در هر زیرسیستم بسته به نوع هدف تحلیل ترافیک تغییر کند، بنابراین در چارچوب پیشنهادی اهمیتی ندارد هدف از تحلیل ترافیک، شناسایی سیستم‌عامل بوده است یا شناسایی بدافزار. در واقع چارچوب کلی می‌تواند برای هریک از اهداف تحلیل ترافیک مورد استفاده قرار گیرد. بنابراین در کارهای آتی می‌توان روش‌های متناسب با تشخیص در سیستم‌های جدیدتر مانند سیستم‌های اینترنت اشیاء به زیرسیستم‌های تشخیص چارچوب پیشنهادی اضافه نمود و برخی از چالش‌های مطرح مانند حملات احتمالی و دستکاری عمدی به اطلاعات را مورد بررسی بیش‌تر قرار داد.

مراجع

[1] Conti, M., Li, Q. Q., Maragno, A., & Spolaor, R. The dark side (-channel) of mobile devices: A survey on network traffic analysis, IEEE Communications Surveys & Tutorials, 20(4), 2658-2713, 2018.

[2] Continella, A., Fratantonio, Y., Lindorfer, M., Puccetti, A., Zand, A., Kruegel, C., & Vigna, G., Obfuscation-Resilient Privacy Leak Detection for Mobile Apps through Differential Analysis, In NDSS, 2017.

[3] Cheng, Z., Chen, X., Zhang, Y., Li, S., & Sang, Y., Detecting information theft based on mobile network flows for Android users, In 2017 International Conference on Networking, Architecture, and Storage (NAS) (pp. 1-10). IEEE, 2017.

[4] Wang, S., Chen, Z., Yan, Q., Yang, B., Peng, L., & Jia, Z., A mobile malware detection method using behavior features in network traffic. Journal of Network and Computer Applications, 133, 15-25, 2019.

[5] Arora, A., & Peddoju, S. K. Minimizing network traffic features for Android mobile malware detection. In Proceedings of the 18th International Conference on Distributed Computing and Networking (p. 32). ACM, 2017.

[6] Pariwono, E., Chiba, D., Akiyama, M., & Mori, T, Don't throw me away: Threats Caused by the Abandoned Internet Resources Used by Android Apps, In Proceedings of the 2018 on Asia Conference on Computer and Communications Security (pp. 147-158). ACM, 2018.

[7] Huang, X., Zhou, A., Jia, P., Liu, L., & Liu, L., Fuzzing the Android Applications With HTTP/HTTPS Network Data, IEEE Access, 7, 59951-59962, 2019.

[8] Taylor, V. F., Spolaor, R., Conti, M., & Martinovic, I., Robust smartphone app identification via encrypted network traffic analysis, IEEE Transactions on Information Forensics and Security, 13(1), 63-78, 2017.

[9] Chaddad, L., Chehab, A., Elhajj, I. H., & Kayssi, A., Mobile Traffic Anonymization through Probabilistic Distribution, In 2019 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN) (pp. 242-248). IEEE, 2019.

[10] Lastovicka, M., Jirsik, T., Celeda, P., Spacek, S., & Filakovsky, D., Passive os fingerprinting methods in the jungle of wireless networks. In NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium (pp. 1-9). IEEE, 2018.

[11] Shamsi, Z., Cline, D. B., & Loguinov, D., Faulds: A non-parametric iterative classifier for Internet-wide OS fingerprinting, In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 971-982). ACM, 2017.

[12] Wei, X., Valler, N. C., Madhyastha, H. V., Neamtiu, I., & Faloutsos, M., Characterizing the behavior of handheld devices and its implications. Computer Networks, 114, 1-12, 2017.

[13] Ghosh, R. K., Mobile OS and Application Protocols, In Wireless Networking and Mobile Data Management (pp. 217-261). Springer, Singapore, 2017.

[14] Zarras, A., Papadogiannakis, A., Gawlik, R., & Holz, T., Automated generation of models for fast and precise detection of HTTP-based malware. In Privacy, Security and Trust (PST), 2014 Twelfth Annual International Conference on (pp. 249-256). IEEE, 2014.

[15] Pajola, L., Pasa, L., & Conti, M., Threat is in the Air: Machine Learning for Wireless Network Applications, In Proceedings of the ACM Workshop on Wireless Security and Machine Learning (pp. 16-21). ACM, 2019.

[16] Miskovic, S., Lee, G. M., Liao, Y., & Baldi, M., AppPrint: Automatic Fingerprinting of Mobile Applications in Network Traffic, In Passive and Active Measurement (pp. 57-69). Springer International Publishing, 2015.

[17] Yoon, S. H., Shim, K. S., Lee, S. K., & Kim, M. S., Framework for multi-level application traffic identification, In Network Operations and Management Symposium (APNOMS), 2015 17th Asia-Pacific (pp. 424-427). IEEE, 2015.

[18] Park, K., & Kim, H., Encryption Is Not Enough: Inferring user activities on KakaoTalk with traffic analysis, In International Workshop on Information Security Applications (pp. 254-265). Springer, Cham, 2015.

[19] Liu, Z., Wang, R., & Tang, D., Research on mobile network traffic taxonomy, In Computer, Information and Telecommunication Systems (CITS), 2016 International Conference on (pp. 1-5). IEEE, 2016.

[20] Mongkolluksamee, S., Visoottiviseth, V., & Fukuda, K., Combining communication patterns & traffic patterns to enhance mobile traffic identification performance, Journal of Information Processing, 24(2), 247-254, 2016.

[21] Alan, H. F., & Kaur, J., Can Android applications be identified using only TCP/IP headers of their launch time traffic?, In Proceedings of the 9th ACM conference on security & privacy in wireless and mobile networks (pp. 61-66), ACM, 2016.

[22] Chen, Z., Yu, B., Zhang, Y., Zhang, J., & Xu, J., Automatic mobile application traffic identification by convolutional neural networks. In 2016 IEEE Trustcom/BigDataSE/ISPA (pp. 301-307). IEEE, 2016.

[23] Fang, P., Huang, L., Xu, H., & He, Q., Smart Device Fingerprinting Based on Webpage Loading, In International Conference on Wireless Algorithms, Systems, and Applications (pp. 127-139). Springer, Cham, 2018.

[24] Chaddad, L., Chehab, A., Elhajj, I. H., & Kayssi, A., Mobile Traffic Anonymization through Probabilistic Distribution, In 2019 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN) (pp. 242-248). IEEE, 2019.

[25] Michal Zalewski. p0f v3 (version 3.08b), 2016. [Online]. Avaiable:http://lcamtuf.coredump.cx/p0f3/. (Visited on June 14, 2019).

[26] Chen, Y. C., Liao, Y., Baldi, M., Lee, S. J., & Qiu, L., OS Fingerprinting and Tethering Detection in Mobile Networks, In Proceedings of the 2014 Conference on Internet Measurement Conference (pp. 173-180). ACM, 2014.

[27] Malik, N., Chandramouli, J., Suresh, P., Fairbanks, K., Watkins, L., & Robinson, W. H., Using network traffic to verify mobile device forensic artifacts, In 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC) (pp. 114-119). IEEE, 2017.

[28] Ruffing, N., Zhu, Y., Libertini, R., Guan, Y., & Bettati, R., Smartphone reconnaissance:

Operating system identification, In 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC) (pp. 1086-1091). IEEE, 2016.

[29] Gurary, J., Zhu, Y., Bettati, R., & Guan, Y., Operating System Fingerprinting. In Digital Fingerprinting (pp. 115-139). Springer, New York, NY, 2016.

[30] Anderson, B., & McGrew, D., OS fingerprinting: New techniques and a study of information gain and obfuscation, In 2017 IEEE Conference on Communications and Network Security (CNS) (pp. 1-9). IEEE, 2017.

[31] Aksoy, A., Louis, S., & Gunes, M. H., Operating system fingerprinting via automated network traffic analysis, In 2017 IEEE Congress on Evolutionary Computation (CEC) (pp. 2502-2509). IEEE, 2017.

[32] Aksoy, A., & Gunes, M. H,. Operating system classification performance of TCP/IP protocol headers, In 2016 IEEE 41st Conference on Local Computer Networks Workshops (LCN Workshops) (pp. 112-120). IEEE, 2016.

[33] Shamsi, Z., & Loguinov, D., Unsupervised Clustering Under Temporal Feature Volatility in Network Stack Fingerprinting, IEEE/ACM Transactions on Networking, 25(4), 2430-2443, 2017.

[34] Laštovička, M., Dufka, A., & Komárková, J., Machine Learning Fingerprinting Methods in Cyber Security Domain: Which one to Use?, In 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC) (pp. 542-547). IEEE, 2018.

[35] Thangavelu, V., Divakaran, D. M., Sairam, R., Bhunia, S. S., & Gurusamy, M., DEFT: A Distributed IoT Fingerprinting Technique, IEEE Internet of Things Journal, 6(1), 940-952, 2018.

[36] Noguchi, H., Kataoka, M., & Yamato, Y., Device Identification Based on Communication Analysis for the Internet of Things, IEEE Access, 7, 52903-52912, 2019.

[37] Bai, L., Yao, L., Kanhere, S. S., Wang, X., & Yang, Z., Automatic device classification from network traffic streams of internet of things, In 2018 IEEE 43rd Conference on Local Computer Networks (LCN) (pp. 1-9). IEEE, 2018.

[38] Yang, K., Li, Q., & Sun, L., towards automatic fingerprinting of IoT devices in the cyberspace. Computer Networks, 148, 318-327, 2019.

[39] Postel, J., & Reynolds, J. K., RFC 1700 Assigned Numbers. Network Working Group, 1994.

[40] A. Jøsang, C. Keser, and T. Dimitrakos, Can We Manage Trust?, Proceedings of the Third International

[41] Conference on Trust Management (iTrust), Versailles, France, pp. 93-107, 2005.

[42] W. T. L. Teacy, J. Patel, N. R. Jennings, and M. Luck, Coping with Inaccurate Reputation Sources Experimental Analysis of a Probabilistic Trust Model, AAMAS'05, 2005.

[43] A. Twigg and N. Dimmock, Attack-Resistance of Computational Trust Models, Proceedings of the

[44] Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative

[45] Enterprises (WET-ICE'03), pp. 275-280, 2003.

[46] Wireshark manufacturer database, 2019, [Online]. Avaiable:

[47] https://code.wireshark.org/review/gitweb?p=wireshark.git;a=blob_plain;f=manuf (visited on June 14, 2019)

[48] Fingerbank Database, 2019, [Online].

[49] Available: https://fingerbank.inverse.ca/ (visited on June 14, 2019).