دوفصل‌نامه علمی منادی امنیت فضای تولید و تبادل اطلاعات (افتا)
سال ۱۴۰۰، جلد ۱۹، شماره ۱ (۱۰-۱۹)
https://monadi.isc.org.ir

ارائه‌شده در شانزدهمین کنفرانس بین المللی انجمن رمز ایران در دانشگاه فردوسی مشهد، ۶ و ۷ شهریور ۱۳۹۸

ارزیابی عملکرد روش‌های تشخیص شبکه‌های بات در مقابل حملات تقلیدی★

عطیه محمدخانی*، فاطمه فرجی دانشگر و مقصود عباسپور

دانشکده علوم و مهندسی کامپیوتر، دانشگاه شهید بهشتی، تهران، ایران

اطلاعات مقاله

کلمات کلیدی:
شبکه بات نظیر به نظیر
روش‌های تشخیص شبکه‌های بات P2P
حمله تقلیدی
ویژگی‌های آماری
اندازه طول بسته

dor: 10.0000/000000000

نوع مقاله: پژوهشی

چکیده

امروزه شبکه‌های بات به عنوان یکی از مهم‌ترین تهدیدات در امنیت اینترنت مطرح هستند. تاکنون تحقیقات بسیاری برای تشخیص شبکه‌های بات صورت گرفته است. دسته‌ای از روش‌های تشخیص شبکه‌های بات مبتنی بر رفتار، از ویژگی‌های آماری برای تشخیص ترافیک نرمال بات استفاده می‌کنند. در اکثر این روش‌ها، ویژگی‌های آماری مربوط به اندازه طول بسته‌ها و زمان‌بندی جز ویژگی‌های اصلی می‌باشد. در یک شبکه بات، یک مهاجم می‌تواند با دستکاری این ویژگی‌ها، رفتار یک شبکه نرمال را تقلید کند. در این مقاله با تغییر اندازه طول بسته‌های بات بر اساس توزیع نرمال P2P، یک شبکه بات P2P تقلیدی ارائه شده است که در آن توزیع طول بسته‌های بات و ویژگی‌های رفتاری مربوط به اندازه طول بسته‌ها، مشابه با ترافیک نرمال می‌باشد. سپس میزان مقاومت و عملکرد روش‌های تشخیص شبکه بات P2P مبتنی بر رویکردهای آماری موجود در برابر حمله تقلیدی مورد ارزیابی می‌گیرد. نتایج آزمایشات صورت گرفته، کاهش حدود ۲۸ الی ۶۳ درصدی نرخ تشخیص را نشان می‌دهد.

۱ مقدمه

امروزه شبکه‌های بات[1] به عنوان یکی از مهم‌ترین تهدیدات در امنیت اینترنت مطرح هستند. شبکه بات، شبکه‌ای از کامپیوترهای آلوده به کد یک بدافزار است که تحت کنترل یک مهاجم[2] با نام بات مستر[3] یا صاحب بات[4] قرار دارد. عبارت بات[5] از کلمه روبات[6] مشتق شده است و برنامه‌ای است که وظایف را به طور خودکار انجام می‌دهد. برای آنکه یک صاحب بات شبکه بات را فرماندهی کند احتیاج به یک کانال فرماندهی و کنترل[7] (C&C) است که به وسیله آن کانال، دستورات را صادر و حملات را هماهنگ کند. کامپیوترهای آلوده در یک شبکه بات زنبور[8] یا زامبی[9] نامیده می‌شوند. شبکه بات شامل هزاران تا میلیون‌ها کامپیوتر آلوده هستند و دارای قدرت محاسباتی بالا و پهنای باند وسیع هستند که برای انجام حملات در مقیاس جهانی استفاده می‌شوند. شبکه‌های بات پلتفرم توزیع شده بزرگی را فراهم می‌کنند و در آن فعالیت‌های مخربی نظیر انکار سرویس توزیع شده (DDoS)، ارسال هرزنامه[10]، جاسوسی اطلاعات، کلاهبرداری کلیک[11]، کاوش بیت‌کوین[12]، حملات نفوذ وحشیانه به پسورد، آلوده نمودن سرویس رسانه اجتماعی[13]، سرقت موجودیت‌های اطلاعاتی[14] و صیادی[15] صورت می‌گیرد. با توجه خطرات بالقوه شبکه بات در فضای مجازی، مطالعات بسیاری برای تشخیص شبکه‌های بات صورت گرفته است. علاوه بر این، نویسندگان بات همواره

★از کمیته علمی شانزدهمین کنفرانس بین‌المللی انجمن رمز ایران برای داوری این مقاله تشکر می‌شود.

*نویسنده مسئول
آدرس‌های رایانامه: at.mohammadkhani@mail.sbu.ac.ir (عطیه محمدخانی)، fdaneshgar@sbu.ac.ir (فاطمه فرجی دانشگر)، maghsoud@sbu.ac.ir (مقصود عباسپور)

[1] botnets [2] attacker [3] botmaster [4] bot herder [5] bot [6] robot

[7] Command and Control [8] drone [9] zombie [10] spamming [11] click-fraud [12] bitcoin mining [13] compromising social media service [14] identity theft [15] phishing

از روش‌های پیشرفته در طراحی شبکه‌های بات خود استفاده می‌کنند تا فرایند تشخیص را پیچیده و سخت‌تر کنند. یکی از روش‌هایی که برای جلوگیری از تشخیص توسط نویسندگان بات استفاده می‌شود، تقلید رفتار نرمال توسط شبکه بات است.

بسیاری از روش‌های رفتاری تشخیص شبکه‌های بات از خصوصیات آماری برای تشخیص ترافیک شبکه بات از نرمال استفاده می‌کنند. ساز و کار این روش‌ها بر این اساس است که ترافیک شبکه بات دارای خصوصیت‌هایی است که باعث تمایز ترافیک بات از نرمال می‌شود و از این ویژگی‌ها برای مدل‌سازی الگوریتم‌های یادگیری ماشین، استفاده می‌کنند. در اکثر این روش‌ها ویژگی‌هایی نظیر اندازه طول بسته و زمان‌بندی جز ویژگی‌های کلیدی می‌باشند. یک مهاجم می‌تواند با طراحی شبکه باتی که ویژگی‌های رفتاری ترافیک نرمال را تقلید می‌کند باعث به اشتباه انداختن این دسته از روش‌ها گردد.

در این مقاله، با تغییر اندازه طول بسته‌ها منطبق با توزیع نرمال ترافیک برنامه کاربردی P2P، شبکه بات تقلیدی P2P جدیدی را ارائه می‌دهیم، سپس به ارزیابی میزان مقاومت تعدادی از روش‌های تشخیص شبکه‌های بات در برابر این حمله خواهیم پرداخت.

۲ مبانی نظری و پندارهای پایه

۱.۲ معماری‌های مختلف شبکه‌های بات

بر اساس چگونگی ارتباط، شبکه بات‌ها به سه معماری مختلف متمرکز، غیرمتمرکز و ترکیبی تقسیم بندی می‌شوند.

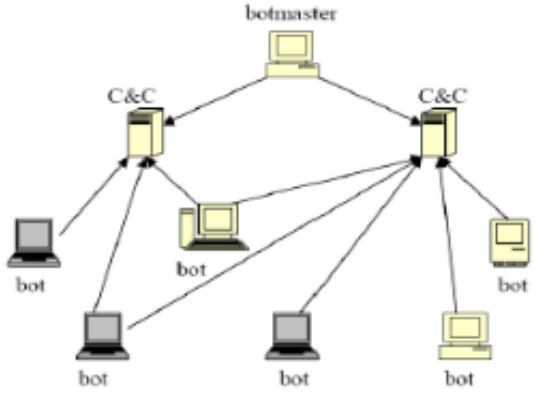
۱.۱.۲ معماری متمرکز

در معماری فرماندهی و کنترل متمرکز بات‌ها برای دریافت دستورات و به روزرسانی، به طور مستقیم به سرور C&C متمرکز متصل می‌شوند. این مدل دارای نقطه شکست می‌باشد، یعنی امکان شناسایی سرور C&C و بلوکه کردن آن وجود دارد. مدل C&C بر اساس پروتکل ارتباطی به دو نوع مبتنی بر IRC و HTTP تقسیم می‌شود.

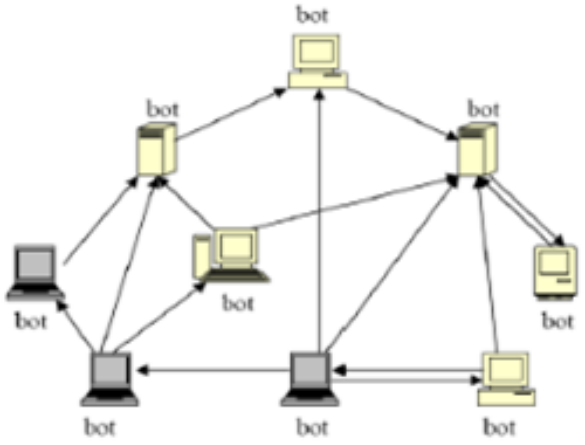
مبتنی بر IRC: IRC یا اینترنت مبتنی بر گفتگو [1] سیستمی است که کاربران آن دارای ارتباط آنلاین یا بلادرنگ هستند. این معماری در اولین نسل بات‌ها بکار می‌رفت که در آن صاحبان بات از سرور IRC و کانال مربوطه برای ارسال دستورات خود استفاده می‌کردند. بات‌های IRC نگرش PUSH را دنبال می‌کردند، بدین معنا که بات IRC به یک کانال مشخص متصل شده و در حالت اتصال باقی می‌ماند [۱].

مبتنی بر HTTP: در این معماری بات از URL و آدرس IP تعریف شده توسط صاحب بات، برای اتصال به وب سرور خاصی استفاده می‌کنند، که نقش سرور C&C را ایفاء می‌کند. بات‌های مبتنی بر HTTP از نگرش PULL استفاده می‌کنند که در آن صاحب بات دستورات را به وب سرور

[1] Internet Relay Chat



شکل ۱. معماری متمرکز



شکل ۲. معماری غیر متمرکز

خاصی ارسال می‌کند، بات‌ها به صورت دوره‌ای به وب سرور مراجعه کرده و دستورات را دریافت و یا خودشان را به روز می‌کنند [۲].

۲.۱.۲ معماری غیر متمرکز

این معماری بر اساس معماری شبکه P2P می‌باشد که به آن معماری توزیع شده نیز می‌گویند که در آن یک کامپیوتر آلوده به طور همزمان به عنوان یک بات و هم به عنوان یک سرور C&C عمل می‌کند. در شبکه‌های بات P2P به جای آنکه یک سرور C&C متمرکز داشته باشیم، هر بات به عنوان یک سرور عمل کرده و دستورات را به بات‌های همسایگان خود ارسال می‌کنند. صاحب بات دستورات را به یک یا تعدادی بات ارسال کرده و آن‌ها نیز دستورات را به سایر بات‌ها ارسال می‌کنند. شبکه‌های بات P2P دارای ساختارهای مختلفی مانند انحصاری [2]، پارازیت [3] و زالویی [4] هستند [۳].

[2] bot-only [3] parasitic [4] leeching

۳.۱.۲ معماری ترکیبی

بات‌های مبتنی بر معماری ترکیبی HTT P2P به منظور گریز از دیوار آتش، از طریق پروتکل HTTP ارتباط برقرار می‌کنند و از ساختار P2P برای حذف سرورهای C&C متمرکز سنتی استفاده می‌کنند. در معماری ترکیبی مبتنی بر AH P2P از تکنولوژی web 2.0 برای پنهان سازی ارتباطات در وب سایت‌های شبکه‌های اجتماعی استفاده می‌کند [۴].

۲.۲ رویکردهای تشخیص شبکه‌های بات

به طور کلی رویکردهای تشخیص شبکه‌های شبکه بات به دوگروه مبتنی بر ظرف عسل و مبتنی بر روش‌های IDS [1] تقسیم‌بندی می‌شوند. روش‌های تشخیص مبتنی بر IDS به دو دسته اصلی مبتنی بر امضاء [2] و مبتنی بر ناهنجاری [3] تقسیم‌بندی می‌شوند. روش‌های مبتنی بر امضاء قادر به تشخیص شبکه‌های بات نوظهور [4] و یا بات‌هایی با الگوی ترافیکی پویا نیستند. رویکردهای مبتنی بر ناهنجاری قادر به شناسائی شبکه بات‌ها بر اساس ویژگی‌های مختلف ترافیک شبکه و ناهنجاری در الگوی رفتاری ترافیک شبکه هستند. در تکنیک‌های مبتنی بر ناهنجاری تعدادی پروفایل ترافیک نرمال ایجاد می‌گردد و با مقایسه ترافیک شبکه با این پروفایل‌ها، در صورت مشاهده انحراف یا ناهنجاری، هشداری ایجاد می‌شود. تکنیک‌های تشخیص مبتنی بر ناهنجاری به طور گسترده‌ای از مولفه‌های یادگیری ماشین [5] (ML) مانند الگوریتم‌های کلاس‌بندی [6] و خوشه‌بندی [7] استفاده می‌کنند. دسته‌ای از روش‌های تشخیص مبتنی بر رفتار هستند و با استفاده از مشخصه‌های رفتاری مربوط به ترافیک شبکه، ترافیک بات و نرمال را تشخیص می‌دهند. گروهی از روش‌های مبتنی بر رفتار از ویژگی‌های آماری نظیر میانگین اندازه طول بسته‌ها و انحراف معیار مدت زمان رسیدن بسته‌ها، برای مدل‌سازی الگوریتم‌های یادگیری ماشین استفاده می‌کنند. مسئله اصلی آسیب‌پذیری این روش‌هادر برابر حملات مختلفی نظیر مسموم‌سازی [8] و یا تقلیدی [9] است. ختک [10] و همکارانش [۵]، رویکردهای تشخیص شبکه بات را از سه جنبه مختلف طبقه‌بندی کردند: تشخیص بات، C&C و صاحب بات. در ادامه روش‌های تشخیص را از جنبه C&C مورد مطالعه قرار می‌دهیم.

۱.۲.۲ تشخیص فعال

تشخیص فعال C&C شامل مشارکت در عملیات شبکه بات می‌باشد، برای مثال تغییر آنلاین جریان شبکه بدست آوردن اطلاعاتی درباره ارتباطات C&C. تشخیص C&C فعال شامل دو رویکرد تزریق [11] یا توقیف [12] می‌باشد.

تزریق مستلزم وارد کردن بسته‌ها به جریان‌های شبکه مظنون است. دریافت پاسخ‌های مشابه به بسته‌های تزریقی نشان می‌دهد که ممکن است جریان مربوط به ارتباط C&C باشد. در توقیف، بسته‌های ورودی/خروجی در جریان شبکه مظنون به دلیل دریافت پاسخ‌های شناخته شده مربوط به ارتباطات C&C، توقیف می‌شوند.

۲.۲.۲ تشخیص غیرفعال

تشخیص غیرفعال C&C شامل مشاهده بی‌صدای ترافیک شبکه و جستجوی ارتباطات C&C می‌باشد. این مکانیزم به دو شیوه نحوی [13] و معنایی [14] انجام می‌شود. تشخیص نحوی این نوع تشخیص با توسعه مدل‌های مبتنی بر امضاء ترافیک C&C کار می‌کند. تشخیص معنایی C&C از یک سری هیوریستیک برای درک رفتار C&C استفاده می‌کند. این رویکرد به سه گروه آماری [15]، همبستگی [16] و مبتنی بر رفتار [17] تقسیم می‌شود.

رویکردهای آماری: این رویکردها برای تشخیص ارتباطات C&C شبکه بات استفاده می‌شود. این رویکرد شامل شناسایی ویژگی‌هایی مانند رنج طول بسته‌ها، زمان رسیدن بسته‌ها [18] و مدت زمان جریان [19] می‌باشد. با استفاده از این ویژگی‌ها، یک طبقه‌بندی‌کننده بر اساس مجموعه داده مرتبط آموزش می‌بیند. سپس قواعد خود را که مربوط به الگوریتم‌های یادگیری ماشین هستند بهبود می‌دهد تا ترافیک شبکه را به دو دسته سالم و C&C طبقه‌بندی کند.

رویکردهای همبستگی: در روش‌های مبتنی بر همبستگی الگوهای ارتباطی مشابه در ترافیک شبکه، می‌توانند مربوط به ترافیک C&C باشند.

رویکردهای مبتنی بر رفتار: در روش‌های مبتنی بر رفتار، ترافیک C&C با مشاهده انحراف [20] با ترافیک نرمال و یا میزان مشابهت آن با مدل‌های رفتاری ایجاد شده از ترافیک C&C شناسایی می‌شود.

۳.۲ حمله تقلیدی

حمله تقلیدی یک کلاس جدید از حملات فریبنده است که متخاصم [21] از استراتژی‌های مختلفی برای فریب دادن سیستم‌های تشخیص استفاده می‌کند. در این نوع حمله مهاجم با تقلید رفتار نرمال شبکه باعث فریب دادن سیستم‌های تشخیص می‌شود و به راحتی می‌تواند به اهداف مخرب خود برسد. حملات تقلیدی ابتدایی، IDSهای مبتنی بر میزبان را مورد هدف قرار می‌دادند. واگنر و دین [22] [۶] اولین بار ایده اصلی حمله تقلیدی را علیه سیستم تشخیص ناهنجاری مطرح کردند. واگنر و سوتو [23] [۷] نشان دادند که چگونه یک مهاجم می‌تواند یک سیستم مبتنی بر ناهنجاری را در سطح فراخوانی‌های سیستمی فریب بدهد.

حمله ترکیبی چندریختی(PBA [24]) حمله‌ای است که هر نمونه حمله دارای کد مخرب یکسان است، ولی از لحاظ ظاهری متفاوت از یکدیگر

[1] Introduction Detection System [2] signature-based [3] anomaly-based [4] zero-day [5] machine learning [6] classification [7] clustering [8] poisoning [9] mimicry [10] Khattak [11] injection [12] suppression

[13] syntactic [14] semantic [15] statistical [16] correlation [17] behavior-based [18] inter-packet arrival time [19] flow duration [20] deviation [21] adversary [22] Wagner and Dean [23] Wagner and Soto [24] Polymorphic Blending Attack

به نظر می‌رسند. از آنجایی که ترافیک این حمله شبیه به نرمال نیست، سیستم‌های تشخیص مبتنی بر ناهنجاری می‌توانند این حملات را شناسایی کنند. فوگلا[1] و همکارانش [۸] یک حمله ترکیبی چندریختی به عنوان زیرکلاسی از حمله تقلیدی ارائه دادند که توانست PAYL[2] (سیستم تشخیص‌دهنده‌ی ناهنجاری مبتنی بر فرکانس بایت پیلود) را فریب دهد. ایده اصلی حمله PBA این است که در هر نمونه حمله، به وسیله تنظیم نمودن فرکانس بایت، از لحاظ ویژگی‌های آماری مشابه با ترافیک نرمال باشد. این نوع حمله قادر است که سیستم‌های IDS مبتنی بر امضاء و نیز مبتنی بر ناهنجاری را فریب دهد. کنیز[3] و همکارانش چندین تکنیک جدید برای حمله تقلیدی ارائه دادند که به صاحب بات اجازه می‌داد تا رویکردهای تشخیص مبتنی بر fast flux را فریب دهد [۹]. ماتا[4] و همکارانش [۱۰] یک شبکه بات با توانایی تقلید از ترافیک عادی ارائه دادند که به طور مداوم الگوهای قابل قبول را از محیط یادگیری می‌کرد. این شبکه بات، با انتخاب پیام از یک فرهنگ لغت شبیه‌سازی شده، الگوهای نرمال را تقلید می‌کرد و به طور مداوم یادگیری انجام می‌دهد تا این اطمینان حاصل شود که نرخ نوآوری به طور معقول می‌تواند پایدار باشد.

در این تحقیق با مطالعه و ایده گرفتن از انواع حملات تقلیدی، یک حمله تقلیدی جدید بر اساس شبکه بات نظیر به نظیر[5] طراحی و اجراء نمودیم. سپس میزان مقاومت تعدادی از روش‌های سیستم‌های تشخیص شبکه‌های بات P2P را در مقابل این حمله مورد ارزیابی قرار دادیم.

۳ پیاده‌سازی شبکه بات P2P تقلیدی

پیاده‌سازی شبکه بات در آزمایشگاه امنیت شبکه دانشکده علوم و مهندسی کامپیوتر دانشگاه شهید بهشتی صورت گرفته است که شامل ۲ فاز می‌باشد: فاز اول شامل اجراء و جمع آوری ترافیک شبکه بات P2P ارائه شده توسط هوارد[6] و همکارانش [۱۱] می‌باشد که بر اساس پروتکل Kademlia است. این شبکه بات از سه جزء اصلی تشکیل شده است:

- (`Botnet.py`): به عنوان گره سرویس‌گیرنده.
- (`Commander.py`): ماژول فرمانده برای ارسال دستورات به نودهای بات
- (`Server.tac`): یک کارگزار Kademlia به عنوان راه‌انداز[7] در شبکه.

در فاز دوم به منظور ایجاد بات P2P تقلیدی، تغییراتی در عملکرد بات P2P ایجاد نمودیم که در ادامه تشریح می‌شود. برای تقلید ترافیک نرمال، ویژگی «اندازه طول بسته» را به عنوان ویژگی که می‌خواهیم به ترافیک نرمال شبیه کنیم، انتخاب کردیم. بدین منظور، از ترافیک نرمال شبکه eMule که نوعی شبکه P2P برای اشتراک فایل می‌باشد، توزیع طول بسته‌ها را به عنوان توزیع نرمال استخراج نمودیم. برای تغییر اندازه طول بسته‌ها ابزارهای مختلفی ارائه شده است [۱۲] که با بررسی این ابزارها، ابزار مشهور NetfilterQueue Scapy [۱۳] را انتخاب کردیم تا تغییر طول بسته به صورت آنلاین و بدون تغییر کد بات انجام گیرد. با استفاده از قواعد دیوار آتش در لینوکس، مطابق زیر

```
Iptables -A OUTPUT -p udp --dport 8468 -j NFQUEUE --queue-num 1
```

ابتدا بسته‌ها را درون صف Iptable نگهداری کرده و سپس طول هر بسته را مطابق الگوریتم ۱ تغییر می‌دهیم. بدین صورت که طول بسته‌ها با انتخاب یک عدد تصادفی از توزیع مرحله قبل، با افزودن بایت اضافی به انتهای پیلود تنظیم شده و ارسال می‌شوند. بنابراین توزیع طول بسته بات‌ها مشابه با توزیع طول بسته ترافیک نرمال eMule خواهد بود.

الگوریتم ۱ دستکاری اندازه طول بسته

Input: C&C packet, File of eMule packets weights;
Output: Mimicry C&C packet;

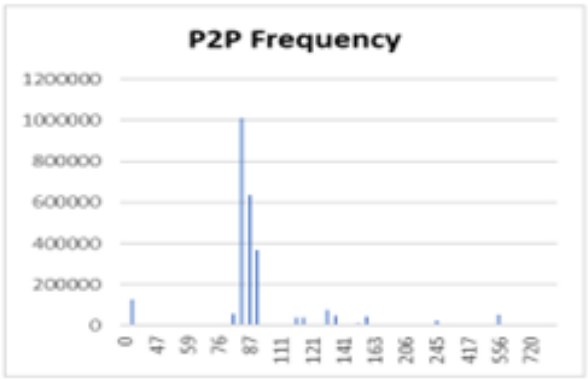
```
packet ← get packet payload
for row in weightsFiles do
    read weights, length
end for
mimcryLen ← randomly select a length from length
add the garbage byte to packet payload matching whit mimcryLen
set packet payload
accept packet
```

شکل ۳ (الف) نمودار توزیع بسته‌های شبکه بات P2P (ب) توزیع بسته‌های نرمال و (ج) توزیع بسته‌های شبکه بات تقلیدی را نشان می‌دهد. همانطور که در نمودارها مشخص است توزیع طول بسته شبکه بات تقلیدی (ب) و نرمال (ج) مشابه یکدیگرند. بنابراین ویژگی‌های رفتاری آماری ترافیک بات که مربوط به طول هستند، مانند میانگین اندازه بسته و تعداد بایت‌های ارسالی/دریافتی ارسالی مانند ترافیک نرمال خواهد بود.

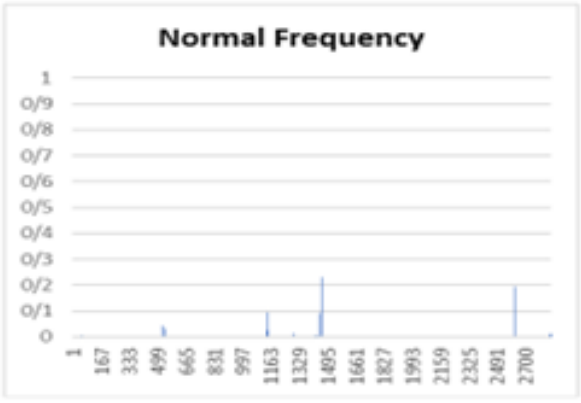
۴ ارزیابی عملکرد روش‌های تشخیص شبکه‌های بات P2P در برابر حمله تقلیدی ارائه شده

به منظور ارزیابی عملکرد روش‌های تشخیص شبکه‌های بات P2P، تعدادی از روش‌های تشخیص مبتنی بر ویژگی‌های آماری را انتخاب نمودیم و در هر مقاله ویژگی‌های آماری که در هر روش برای تشخیص ترافیک بات و نرمال استفاده شده را به طور جداگانه پیاده‌سازی نمودیم. در جدول ۱ روش‌های آماری تشخیص شبکه‌های بات P2P را که ویژگی‌های آنان در مقاله پیاده‌سازی و استخراج شده‌اند، لیست شده‌اند. بعد از استخراج مجموعه ویژگی‌ها، در مرحله بعد مقاومت هر روش را با استفاده

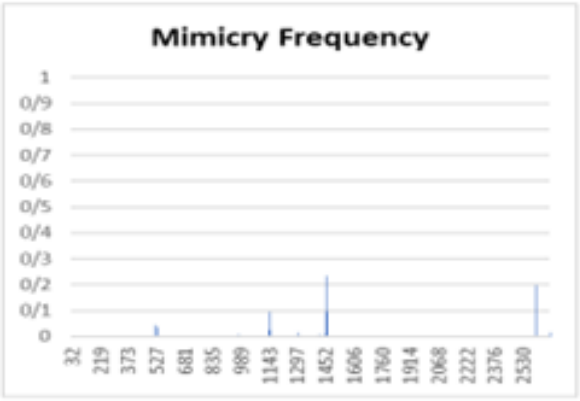
[1]Fogla [2]PAYLoad anomaly detection [3]Knysz [4]Matta [5]Peer-two-Peer
[6]Howard [7]bootstrap



(الف)



(ب)



(ج)

شکل ۳. توزیع طول بسته‌ها: (الف) بات P2P (ب) نرمال (ج) بات تقلیدی

از این ویژگی‌ها با مجموعه‌های داده بات P2P و بات تقلیدی P2P، مورد آزمایش قرار می‌دهیم. ابزار استفاده شده در این مرحله نرم افزار Weka می‌باشد. از آنجایی که روش‌های تشخیص شبکه‌های بات از الگوریتم‌های مختلف طبقه‌بندی و خوشه‌بندی استفاده می‌کنند، تعدادی از الگوریتم‌های مشهوری که در ادوات امنیتی استفاده می‌شوند مانند SVM، J48، Ran-dom forest، Bayesian Net، Naïve Bayes را انتخاب کرده و سپس با مجموعه داده شبکه بات P2P مدل‌های مختلف را ایجاد و آن‌ها را با

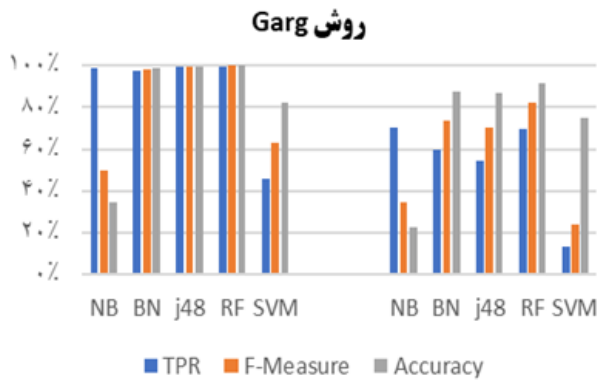
جدول ۱. روش‌های تشخیص شبکه‌های بات P2P مبتنی بر رویکردهای آماری

روش	رویکرد تشخیص شبکه‌های بات P2P	مرجع
۱	Improved Detection of P2P Botnets through Network Behavior Analysis, Shree Garg	[۱۴]
۲	Peer to Peer Botnet Detection Using Data Mining Scheme, Wen-Hwa Liao	[۱۵]
۳	Detecting P2P Botnets through Network Behavior Analysis and Machine Learning, Sherif Saad	[۱۶]
۴	Online Botnet Detection Based on Incremental Discrete Fourier Transform, Xiaocong Yu	[۱۷]
۵	Botnet detection based on traffic behavior analysis and flow intervals, David Zhao	[۱۸]
۶	P2P and P2P Botnet traffic classification in two stages, Wujian Ye	[۱۹]

مجموعه داده Mimicry مورد ارزیابی مجدد قرار دادیم.

۱.۴ روش گارگ

گارگ [1] و همکارانش [۱۴] مدلی را برای تشخیص شبکه بات P2P ارائه دادند که با استفاده از الگوریتم طبقه‌بندی درخت‌های تصمیم با نام جنگل تصادفی بین ترافیک C&C شبکه‌های بات P2P و ترافیک نرمال تمایز ایجاد می‌کرد. در این روش با استخراج ویژگی‌های مختلف جریان شبکه مانند اندازه بسته، مدت زمان جریان و غیره که در جدول ۲ لیست شده‌اند، رفتارها و الگوهای ترافیکی شبکه باتها را مشخص کرده و ترافیک آنان را از ترافیک عادی شبکه متمایز می‌سازند. در جدول ۲ ویژگی‌های آماری روش گارگ لیست شده است. شکل ۴ نمودار ارزیابی مقاومت روش گارگ را در دو حالت قبل و بعد از حمله نشان می‌دهد، همانطور که مشخص است در حالت بعد از حمله نرخ مثبت درست (TPR) در الگوریتم‌های مختلف به شدت کاهش یافته است.



شکل ۴. نمودار ارزیابی روش [۱۴]

[1] Garg

جدول ۲. ویژگی‌های آماری مبتنی بر جریان در [۱۴]

Feature	Des.	No.
IOP_byte	Ratio of incoming over output bytes in the flow	F1
APL	Average packet length in the flow	F2
TBT	Total bytes transferred during whole capture to the number of frames per flow	F3
PPS	Frames per second in the flow	F4
BPS	Bytes per second in the flow	F5
IOP_frame	Ratio of incoming over outgoing number of frame in the flow	F6
From frame	Number of Incoming frames in the flow	F7
Toframe	Number of outgoing frames in the flow	F8
Tobyte	Number of outgoing bytes in the flow	F9
Frombyte	Number of incoming bytes in the flow	F10

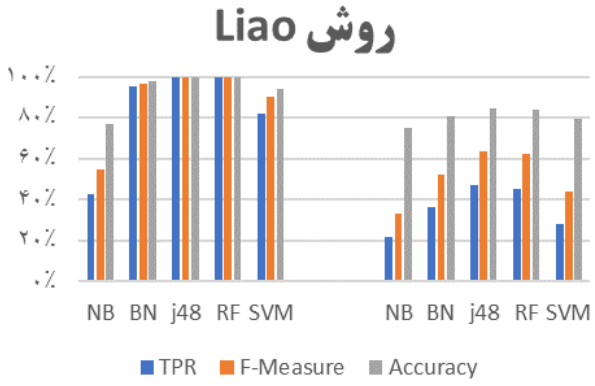
۲.۴ روش لیئو

لیئو [1] و همکارانش [۱۵] از یک متدولوژی مبتنی بر اندازه بسته‌ها استفاده کردند تا ترافیک شبکه بات P2P را از ترافیک قانونی P2P متمایز کنند. آن‌ها سه فرضیه در روش تشخیص خود در نظر گرفتند: اولاً، ارتباطات شبکه‌های بات P2P یک نوع ساختار P2P را تقلید می‌کنند. ثانیاً، به منظور عدم استفاده از اینترنت، یک بات P2P به جای آنکه ارتباطات خود را بعد از تماس گرفتن متوقف کند، اطلاعات جلسات و داده‌های را نگهداری می‌کند. ثالثاً، برای ایجاد اختفاء و اجتناب از کشف شدن، داده‌های ارتباطی میان بات‌ها در سطح پایین منتقل می‌شوند. بعد از مشاهدات بسیار، آن‌ها به این نتیجه دست یافتند که بسته‌های شبکه بات P2P تمایل دارند تا میانگین طول بسته کمتری نسبت به بسته‌های شبکه نرمال داشته باشند. در جدول ۳ ویژگی‌های آماری روش لیئو لیست شده است. شکل ۵ نمودار ارزیابی مقاومت روش لیئو را در دو حالت قبل و بعد از حمله نشان می‌دهد، همانطور که مشخص است در حالت بعد از حمله نرخ مثبت درست (TPR) در الگوریتم‌های مختلف به شدت کاهش یافته است.

جدول ۳. ویژگی‌های آماری مبتنی بر جریان در [۱۵]

Feature	Des.	No.
APL	Average Size of Packets	F1
PPS	Average Packets per Second	F2
PSP	Percentage of Small Packets	F3
NSP	Quantities of Small Packets	F4
NNP	Null Packets	F5

[1] Liao



شکل ۵. نمودار ارزیابی روش [۱۵]

۳.۴ روش سعد

سعد [2] و همکارانش [۱۶] پنج الگوریتم مختلف یادگیری ماشین را مقایسه کردند تا کاربرد آنها را در تشخیص شبکه‌های بات آنلاین P2P مشخص کنند. آن‌ها ۱۷ ویژگی مختلف را در دو گروه مبتنی بر جریان و مبتنی بر میزبان تقسیم بندی کردند. ویژگی‌های مبتنی بر جریان مانند میانگین طول بسته‌ها و تعداد کل بایت کمک به طبقه‌بندی ترافیک P2P و غیر P2P می‌کنند، در حالیکه ویژگی‌های مبتنی بر میزبان مانند تعداد اتصالات به آدرس IP مقصد و نسبت پورت مقصد به پورت مبداء کمک می‌کنند تا میزبان‌ها با الگوهای ارتباطی C&C مشابه را شناسایی کنند. در جدول ۴ ویژگی‌های آماری روش سعد لیست شده است. شکل ۶ نمودار ارزیابی مقاومت روش سعد را در دو حالت قبل و بعد از حمله نشان می‌دهد، همانطور که مشخص است در حالت بعد از حمله نرخ مثبت درست (TPR) در الگوریتم‌های مختلف به شدت کاهش یافته است.

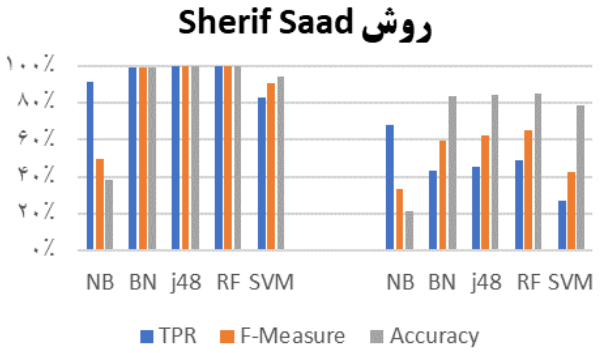
جدول ۴. ویژگی‌های آماری مبتنی بر جریان در [۱۶]

Feature	Des.	No.
PLP	Pack length Payload size in bytes	F1
APL	Average packet length per flow	F2
FPL	The length of the first packet in the	F3
TPC	The total number of packets per flow	F4
TBT	Null Packets	F5

۴.۴ روش یو

یو [3] و همکارانش [۱۷] یک متدولوژی بر اساس DFT [4] ارائه دادند تا الگوهای ارتباطی میان بات‌ها را پیدا کنند. این تکنیک زمان محاسباتی کمی دارد که به عنوان مزیت مهم در اجرای آنلاین تلقی می‌شود. نویسنده از یک فیلتر کاهنده حجم اولیه استفاده می‌کند تا هر نوع ترافیکی را که برای تجزیه و تحلیل نامناسب باشد را حذف کند. مانند فیلتر کردن لیست سیاه، لیست سفید و پروتکلی که برای کانال C&C استفاده می‌شود. آن‌ها

[2] Saad [3] Yu [4] Discrete Fourier Transform



شکل ۶. نمودار ارزیابی عملکرد روش [۱۶]

مفهومی به نام استریم ویژگی [1] برای توصیف ترافیک خام استفاده کردند. اگر این استریم ویژگی میزان مشابهت بالایی را نشان دهد، میزبان مورد نظر مظنون به بات در نظر گرفته خواهد شد. آن‌ها از میانگین فاصله اقلیدسی برای اندازه‌گیری میزان مشابهت استفاده کردند. جریان باقیمانده با استفاده از DFT پردازش شده و یک الگوی ارتباطی گرافیکی بدست می‌آید. با بررسی پی در پی گراف‌های شبکه تولید شده با این تکنیک در بازه‌های زمان دوره‌ای الگوهایی که مشاهده می‌شوند، می‌توانند منجر به بات طبقه‌بندی شوند. در جدول ۵ ویژگی‌های آماری روش یو لیست شده است. شکل ۷ نمودار ارزیابی مقاومت روش یو را در دو حالت قبل و بعد از حمله نشان می‌دهد، همانطور که مشخص است در حالت بعد از حمله نرخ مثبت درست (TPR) در الگوریتم‌های مختلف به شدت کاهش یافته است.

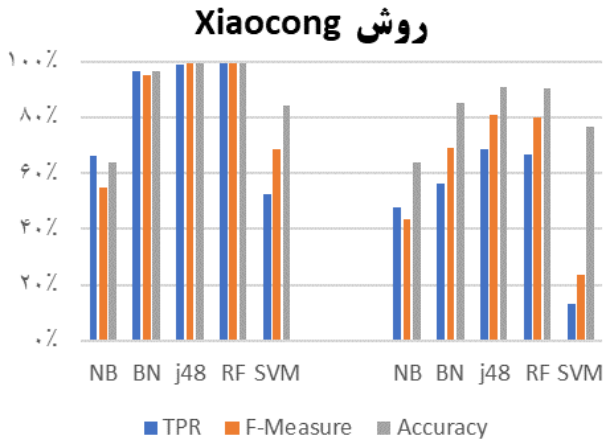
جدول ۵. ویژگی‌های آماری مبتنی بر جریان در [۱۷]

Feature	Des.	No.
TPC	Total pkts exchanged in flow	F1
DUR	Flow duration	F2
TBT	Total Bytes exchanged in flow	F3
ABPP	Average Bytes-per-packet for flow	F4
BitPS	Average bits-per-second for flow	F5
PPS	Average packets-per-second for flow	F6

۵.۴ روش ژائو

ژائو [2] و همکارانش [۱۸] روشی برای تشخیص فعالیت بات مبتنی بر رفتار شبکه ارائه شده است و با استفاده از یادگیری ماشین ترافیک شبکه را طبقه‌بندی می‌کنند. روش‌های تحلیل رفتار شبکه به پیلود بسته‌ها متکی نیست و آن‌ها می‌توانند با پروتکل‌های ارتباطی رمزنگاری شده شبکه کار کنند. آنان مجموعه‌ای از ویژگی‌ها را از جریان استخراج کردند و سپس الگوریتم شبکه بیزی [3] و درخت تصمیم [4] برای طبقه‌بندی ترافیک مخرب و سالم اعمال کردند در جدول ۶ ویژگی‌های آماری روش ژائو لیست شده

[1]Feature stream [2]Zhao [3]Bayes network [4]decision tree

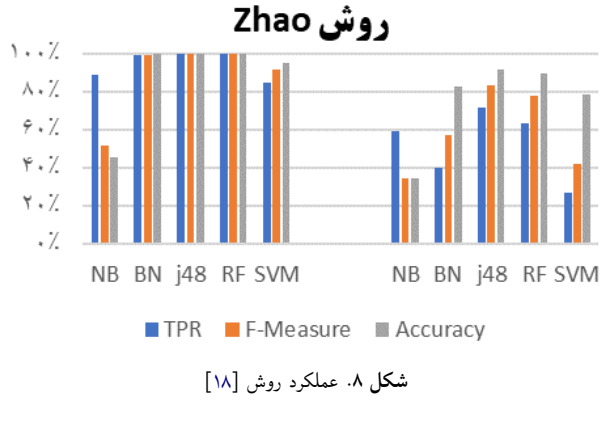


شکل ۷. عملکرد روش [۱۷]

است. شکل ۸ نمودار ارزیابی مقاومت روش ژائو را در دو حالت قبل و بعد از حمله نشان می‌دهد، همانطور که مشخص است در حالت بعد از حمله نرخ مثبت درست (TPR) در الگوریتم‌های مختلف به شدت کاهش یافته است.

جدول ۶. ویژگی‌های آماری مبتنی بر جریان در [۱۸]

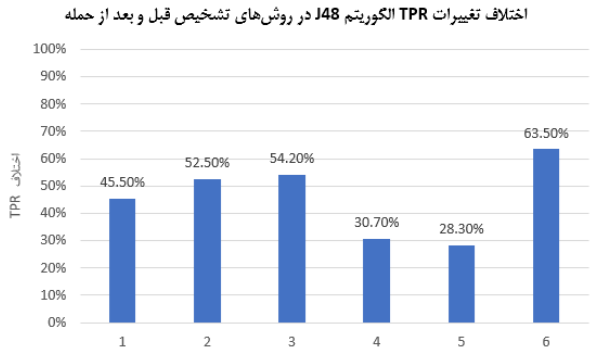
Feature	Des.	No.
TBT	Total number of bytes per flow	F1
DUR	Flow Duration	F2
BPS	Bytes per second in the flow	F3
TPC	The total number of packets per flow	F4
APL	Average Packets per Second	F5



شکل ۸. عملکرد روش [۱۸]

۶.۴ نتیجه‌گیری و کارهای آتی

هدف ما از انجام این تحقیق این بود که با طراحی حمله‌ی تقلیدی P2P روش‌های تشخیص آماری را فریب داده و عملکرد آن‌ها را مورد ارزیابی قرار دهیم. همانگونه که از نمودارها و جداول ارزیابی الگوریتم‌ها در دو حالت مدل‌سازی با مجموعه داده P2P Botnet (Before) و ارزیابی



شکل ۹. مقایسه میزان آسیب‌پذیری روش‌های تشخیص در برابر حمله تقلیدی

مجدد با مجموعه داده Mimicry P2P Botnet (After) مشاهده می‌شود، معیارهای ارزیابی TPR، F-measure و Accuracy در حالت بعد از حمله با کاهش چشمگیری همراه است. نتایج آزمایش‌ها نشان می‌دهد که در این روش‌ها، به دلیل استفاده از مجموعه ویژگی‌های مبتنی بر اندازه طول بسته به عنوان ویژگی کلیدی، در برابر حمله تقلیدی ارائه شده به راحتی آسیب‌پذیر هستند. بنابراین، مجموعه ویژگی‌های هر روش جهت تشخیص شبکه بات مقاوم نبوده و به عنوان نقطه ضعف روش‌های تشخیص تلقی می‌شود.

شکل ۹ اختلاف نرخ TPR الگوریتم قدرتمند J48 را در روش‌های تشخیص شبکه‌ها بات P2P را در برابر حمله تقلیدی در حالت قبل و بعد از حمله نشان می‌دهد که با کاهش حدود ۲۸ تا ۶۳ درصدی نرخ تشخیص همراه است.

با توجه به ارزیابی‌های فوق می‌توان نتیجه گرفت که مجموعه ویژگی‌هایی که در روش‌های تشخیص آماری مبتنی بر رفتار مورد استفاده قرار گرفته‌اند، در برابر حمله تقلیدی طراحی شده آسیب‌پذیر هستند. در تحقیقات آتی، رویکردی را ارائه خواهیم داد که در آن مجموعه ویژگی‌های مقاوم به حمله تقلیدی ارائه خواهد شد و روش تشخیص پیشنهادی خود را با مجموعه‌های مختلف بات‌های موجود و نیز بات تقلیدی ارائه شده مورد آزمایش قرار خواهیم داد.

مراجع

[1] Guofei Gu, Roberto Perdisci, Junjie Zhang, and Wenke Lee. Botminer: Clustering analysis of network traffic for protocol-and structure-independent botnet detection. 2008.

[2] Guofei Gu, Junjie Zhang, and Wenke Lee. Botsniffer: Detecting botnet command and control channels in network traffic. 2008.

[3] Ping Wang, Baber Aslam, and Cliff C Zou. Peer-to-peer botnets. In *Handbook of Information and Communication Security*, pages 335–350. Springer, 2010.

[4] Dae-il Jang, Minsoo Kim, Hyun-chul Jung, and Bong-Nam Noh. Analysis of http2p botnet: case study waledac. In *2009 IEEE 9th Malaysia International Conference on Communications (Micc)*, pages 409–412. IEEE, 2009.

[5] Sheharbano Khattak, Naurin Rasheed Ramay, Kamran Riaz Khan, Affan A Syed, and Syed Ali Khayam. A taxonomy of botnet behavior, detection, and defense. *IEEE communications surveys & tutorials*, 16(2):898–924, 2013.

[6] David Wagner and R Dean. Intrusion detection via static analysis. In *Proceedings 2001 IEEE Symposium on Security and Privacy. S&P 2001*, pages 156–168. IEEE, 2000.

[7] David Wagner and Paolo Soto. Mimicry attacks on host-based intrusion detection systems. In *Proceedings of the 9th ACM Conference on Computer and Communications Security*, pages 255–264, 2002.

[8] Prahlad Fogla, Monirul I Sharif, Roberto Perdisci, Oleg M Kolesnikov, and Wenke Lee. Polymorphic blending attacks. In *USENIX security symposium*, pages 241–256, 2006.

[9] Matthew Knysz, Xin Hu, and Kang G Shin. Good guys vs. bot guise: Mimicry attacks against fast-flux detection systems. In *2011 Proceedings IEEE INFOCOM*, pages 1844–1852. IEEE, 2011.

[10] Elizabeth Stinson and John C Mitchell. Towards systematic evaluation of the evadability of bot/botnet detection methods. *WOOT*, 8:1–9, 2008.

[11] Howard j. python p2p botnet. https://github.com/jhoward321/PythonP2PBotnet. Accessed: 2016.

[12] Shankdhar p. 15 best free packet crafting tool. https://resources.infosecinstitute.com/15-best-free-packet-crafting-tools/#gref. Accessed: 2018.

[13] Fox m. python-netfilterqueue. https://github.com/kti/python-netfilterqueue. Accessed: 2011.

[14] Shree Garg, Anil K Sarje, and Sateesh Kumar Peddoju. Improved detection of p2p botnets through network behavior analysis. In *International Conference on Security in Computer Networks and Distributed Systems*, pages 334–345. Springer, 2014.

[15] Wen-Hwa Liao and Chia-Ching Chang. Peer to peer botnet detection using data mining scheme. In *2010 inter-*

national conference on internet technology and applications, pages 1–4. IEEE, 2010.

[16] Sherif Saad, Issa Traore, Ali Ghorbani, Bassam Sayed, David Zhao, Wei Lu, John Felix, and Payman Hakimian. Detecting p2p botnets through network behavior analysis and machine learning. In *2011 Ninth annual international conference on privacy, security and trust*, pages 174–180. IEEE, 2011.

[17] Xiaocong Yu, Xiaomei Dong, Ge Yu, Yuhai Qin, Dejun Yue, and Yan Zhao. Online botnet detection based on incremental discrete fourier transform. *Journal of Networks*, 5(5):568, 2010.

[18] David Zhao, Issa Traore, Bassam Sayed, Wei Lu, Sherif Saad, Ali Ghorbani, and Dan Garant. Botnet detection based on traffic behavior analysis and flow intervals. *computers & security*, 39:2–16, 2013.

[19] Wujian Ye and Kyungsan Cho. P2p and p2p botnet traffic classification in two stages. *Soft Computing*, 21(5):1315–1326, 2017.