دانشگاه تربیت مدرس
دانشکده برق و کامپیوتر

پایان‌نامه برای دریافت درجه کارشناسی ارشد

رشته مهندسی کامپیوتر گرایش نرم‌افزار

کشف بات‌نت مبتنی بر شبکه‌های نظیر به نظیر با استفاده از رویکردهای یادگیری ماشین

نگارنده

عبد العزیز الیوسف

استاد راهنما

دکتر سعید جلیلی

تیر ۱۳۹۰

بسمه تعالی

تاییدیه اعضای هیات داوران حاضر در جلسه دفاع از پایان نامه کارشناسی ارشد

آقای عبدالعزیز الیوسف پایان نامه ۹ واحدی خود را با عنوان **کشف بات نت مبتنی بر P۲P با استفاده از رویکرد های یادگیری ماشین** در تاریخ ۱۳۹۰/۴/۱۴ ارائه کردند.

اعضای هیات داوران نسخه نهایی این پایان نامه را از نظر فرم و محتوا تایید کرده، پذیرش آنرا برای اخذ درجه کارشناسی ارشد مهندسی کامپیوتر-نرم افزار پیشنهاد می کنند.

عضو هیات داوران	نام و نام خانوادگی	رتبه علمی	امضا
استاد راهنما	دکتر سعید جلیلی	دانشیار	
استاد مشاور	دکتر مهدی آبادی	استادیار	
استاد ناظر	دکتر بهزاد اکبری	استادیار	
استاد ناظر	دکتر سیاوش خرسندی	استادیار	
مدیر گروه (یا نماینده گروه تخصصی)	دکتر بهزاد اکبری	استادیار	

دستورالعمل حق مالکیت مادی و معنوی در مورد نتایج پژوهشهای علمی دانشگاه تربیت مدرس

مقدمه: با عنایت به سیاست‌های پژوهشی دانشگاه در راستای تحقق عدالت و کرامت انسانها که لازمه شکوفایی علمی و فنی است و رعایت حقوق مادی و معنوی دانشگاه و پژوهشگران، لازم است اعضای هیات علمی، دانشجویان، دانش آموختگان و دیگر همکاران طرح، در مورد نتایج پژوهشهای علمی که تحت عناوین پایان‌نامه، رساله و طرحهای تحقیقاتی که با هماهنگی دانشگاه انجام شده است، موارد ذیل را رعایت نمایند:

ماده ۱- حقوق مادی و معنوی پایان نامه‌ها / رساله‌های مصوب دانشگاه متعلق به دانشگاه است و هرگونه بهره‌برداری از آن باید با ذکر نام دانشگاه و رعایت آیین‌نامه‌ها و دستورالعمل‌های مصوب دانشگاه باشد.

ماده ۲- انتشار مقاله یا مقالات مستخرج از پایان نامه/ رساله به صورت چاپ در نشریات علمی و یا ارائه در مجامع علمی باید به نام دانشگاه بوده و استاد راهنما مسئول مکاتبات مقاله باشد.

تبصره: در مقالاتی که پس از دانش آموختگی بصورت ترکیبی از اطلاعات جدید و نتایج حاصل از پایان‌نامه/ رساله نیز منتشر می‌شود نیز باید نام دانشگاه درج شود.

ماده ۳- انتشار کتاب حاصل از نتایج پایان نامه / رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با مجوز کتبی صادره از طریق حوزه پژوهشی دانشگاه و بر اساس آئین نامه های مصوب انجام می‌شود.

ماده ۴- ثبت اختراع و تدوین دانش فنی و یا ارائه در جشنواره‌های ملی، منطقه‌ای و بین‌المللی که حاصل نتایج مستخرج از پایان نامه/ رساله و تمامی طرح‌های تحقیقاتی دانشگاه باید با هماهنگی استاد راهنما یا مجری طرح از طریق حوزه پژوهشی دانشگاه انجام گیرد.

ماده ۵- این دستورالعمل در ۵ ماده و یک تبصره در تاریخ ۱۳۸۴/۴/۲۵ در شورای پژوهشی دانشگاه به تصویب رسیده و از تاریخ تصویب لازم الاجرا است و هرگونه تخلف از مفاد این دستورالعمل، از طریق مراجع قانونی قابل پیگیری می شود.

نام و نام خانوادگی: عبد العزیز الیوسف

امضاء

دانشگاه تربیت مدرس
دانشکده فنی و مهندسی

پایان‌نامه برای دریافت درجه کارشناسی ارشد
رشته مهندسی کامپیوتر گرایش نرم‌افزار

کشف بات‌نت مبتنی بر شبکه های نظیر به نظیر با استفاده از رویکردهای یادگیری ماشین

نگارنده
عبد العزیز الیوسف

استاد راهنما
دکتر سعید جلیلی

استاد مشاور
دکتر مهدی آبادی

تیر ۱۳۹۰

تقدیم به امام زمان (عج)

تقدیم به پدر و مادر عزیزم.

تقدیم به همسر مهربانم که همیشه همراهم بوده‌است.

تقدیم به دختر نازنینم، به خاطر وقتهایی که از او دریغ کردم.

تشکر و قدردانی:

بر خود لازم می‌دانم سپاسگزار عزیـزان و سـرورانی باشـم کـه در لحظـات زنـدگی و دوران تحصیل در کنارم بودند.

قدر دان زحمات اساتید گرانقدر دانشکده مهندسی کامپیوتر، بخصوص جنـاب آقـای دکتـر جلیلی هستم که با صبر و بزرگواری نه‌تنها در انجام این پایان‌نامه بلکه در تمام این دوره مرا یـاری نمودند.

همچنین تشکر و قدردانی می‌نمایم از جناب آقای دکتر آبادی، جناب آقای دکتر اکبـری، و جناب آقای دکتر خرسندی برای داوری این پایان‌نامه قبول زحمت نمودند و وقت گرانبهای خود را در اختیار اینجانب قرار دادند.

به جان منت پذیرم و حق گزارم

این پروژه تحت حمایت مرکز تحقیقات مخابرات ایران می‌باشد.

چکیده

باتنت (Botnet) کلمه‌ای است که معرف شبکه‌ای از بات‌ها است. بات (bot) به کامپیوترهایی اشاره می‌کند که می‌توانند توسط یک یا چند منبع خارجی کنترل شوند. یک فرد مهاجم معمولا کنترل کامپیوتر را با ضربه زدن به آن کامپیوتر توسط یک ویروس یا یک کد مخرب (تروجان یا کرم) بدست می‌آورد و به این وسیله دسترسی فرد مهاجم به سیستم آسیب دیده فراهم می‌شود. بات‌نت‌ها معمولا برای هدایت فعالیتهای مختلفی مورد استفاده قرار می‌گیرند, این فعالیتها می‌تواند شامل: حملات انکار سرویس توزیع شده، هرزنامه، ابزار جاسوسی وغیره باشد.

بیشتر بات‌نت‌ها مبتنی بر IRC (Internet Relay Chat) هستند، که معماری متمرکزی دارند و بسیار بزرگ هستند، که این دو ویژگی کشف این نوع باتنت را تسهیل می کنند. در سال-های اخیر تعداد زیادی روش برای تشخیص بات‌نت‌های مبتنی بر IRC پیشنهاد شده‌اند، به همین خاطر طراحان باتنت سعی کرده‌اند امکانات جدیدی به این نوع حمله اضافه کنند، که مهمترین این امکانات، معماری غیر متمرکز آن می باشد. و در سال ۲۰۰۷ توانستند با استفاده از پروتکل‌های شبکه‌ی نظیر به نظیر (Peer to Peer) یک شبکه از بات‌ها معرفی کنند، که نه تنها برای افزایش سرعت انتشار بات‌ها استفاده می شود، بلکه برای ساخت بات‌نت‌ها با توان بیشتر هم مورد استفاده قرار می‌گیرد. هر بات در این نوع شبکه می تواند نقش فرماندهی وکنترل داشته باشد، بنابراین حمله کننده با برقراری ارتباط با شبکه باتنت به عنوان یک همکار فرمان خود را به بقیه بات‌ها ارسال می‌کند. سپس عمل فرماندهی و کنترل توسط چندین بات انجام می‌شود. در نتیجه کشف آنها سخت خواهد شد.

استفاده از پروتکل شبکه های نظیر به نظیر باعث می‌شود که جریان ارتباطی بات‌ها بین هم شباهت زیادی با جریان ارتباطی شبکه های نظیر به نظیر داشته باشد. علاوه بر آن ارسال فرمان از یک سرور خاص انجام نخواهد شد بلکه همه بات‌ها می‌توانند در ارسال فرمان شرکت کنند و بالطبع مشکل بات‌نت‌های مبتنی بر IRC را نخواهند داشت.

در این پژوهش سعی شده‌است جریان ارتباطی بات‌نت‌ها را از بقیه جریان‌های اینترنت تفکیک کنیم، وبرای این کار یک سیستم دو مرحله‌ای طراحی شده‌است، در مرحله‌ای اول جریان باتنت و شبکه‌های نظیر به نظیر را که با هم شباهت زیادی دارند، از بقیه جریان‌های اینترنت (http, ftp, telnet, ...) با استفاده از الگوریتم‌های دسته بندی تفکیک کردیم که الگوریتم درخت تصمیم، بهترین نتیجه را داشت و دقت آن ۹۹.۲ ٪ بود. در مرحله‌ی دوم به روش تشخیص ناهنجاری عمل کردیم، و برای جریان مربوط به شبکه های نظیر به نظیر یک مدل ساخته شد و هر جریانی که از این مدل پیروی نمی کند، به عنوان جریان بات محسوب می شود. مدل جریان شبکه‌های نظیر به نظیر با استفاده از دو روش GMM و OCSVM ساخته شده است. نتیجه نشان دهنده این است

که الگوریتم GMM از دقت بیشتری برخوردار است و دقت تشخیص آن حدود ۹۸.۱٪ با ۰.۰۱۱ نرخ مثبت کاذب (FP) می‌باشد، در حالی که دقت ساخت مدل با OCSVM (One Class svm)، ۹۶.۳٪ بود، و نرخ مثبت کاذب مساوی با ۰.۰۵۶ می‌باشد.

در این پژوهش بر خلاف پژوهش‌های قبلی که باید منتظر حمله بات بود تا بتوان از جریان ارتباطی وجریان حمله، بات‌ها را تشخیص داد، ما فقط از جریان ارتباطی بات ها با هم و با تعریف خصیصه های مؤثر که هزینه بار محاسباتی را کم می کند، استفاده کردیم. خصیصه‌های مورد نظر در هر دو مرحله، با استفاده از چندین الگوریتم انتخاب خصیصه، و با عمل اشتراک بین خروجی این الگوریتم‌ها، استخراج شد.

در این پژوهش برای ارزیابی روش پیشنهادی، از سه نوع باتنت مبتنی بر P2P (Waledac, Storm, Confiker C) و از سه نوع شبکه نظیر به نظیر که بیشتر برای به اشتراک گذاشتن فایل مورد استفاده قرار می‌گیرند ، استفاده شده‌است.

واژهای کلیدی: بات نت ، شبکه های نظیر به نظیر، تحلیل رفتاری جریان، یادگیری ماشین، دسته بندی، تشخیص ناهنجار، خصیصه.

فهرست مطالب

فهرست شکل‌ها

فهرست جدول‌ها

فهرست اختصارات

ACK	Acknowledgement
BN	Bayesian Network
C&C	command and control
CRT	Click-through Rate
DHT	Distributed Hash Table
DDos	Distributed denial of service
GMM	Guassian Mixture Models
HTTP	Hypertext Transfer Protocol
IP	Internet Protocol
ICMP	Internet Control Message Protocol
IRC	Internet Relay Chat
NAT	Network Address Translation
P2P	Peer-to-Peer
SMTP	Simple Mail Transfer Protocol
SVM	Support Vector Machine
SYN	Synchronize
TCP	Transmission Control Protocol
TTL	Time to live
TDG	Traffic dispersion graph
URL	Uniform Resource Locator

فصل ۱:

کلیات

۱-۱- مقدمه

در چند دهه اخیر شاهد افزایش استفاده از اینترنت و برنامه‌های کاربردی آن هستیم، بصورتی که استفاده از اینترنت به بخش مهمی در زندگی روزمره ما تبدیل شده است. اگر چه اینترنت تسهیلات زیادی را به ما ارائه می دهد، اما با افزایش استفاده از اینترنت چالش‌های امنیتی بزرگی ظاهر شده است، و این امر به مسأله‌ی مهم وحساسی برای بسیاری از اشخاص و شرکت‌ها تبدیل شده‌است.

بسیاری از حملات و فعالیت‌های جعلی در اینترنت از طریق نرم افزارهای مخرب انجام می‌شود، این نرم افزارها شامل ویروس‌ها، اسب‌های تروجان، و کرم‌ها است که در چند سال آخیر بات‌نت‌ها نیز به این مجموعه اضافه شده‌اند. بات‌نت‌ها به یک منبع اساسی برای بسیاری از حملات خطرناک از قبیل پویش[۱]، جلوگیری از سرویس توزیع شده[۲]، هرزنامه[۳]، فعالیت های جعلی[۴]، و غیره تبدیل شده‌اند [۱، ۲].

بات‌نت به عنوان مهمترین تهدید کننده امنیتی در چند سال اخیر شناخته شده است[۳]، که همچنان در حال رشد و توسعه است، کلمه‌ی بات بر گرفته از کلمه‌ی ربات است، که با نام زامبی نیز شناخته می شود، بات‌ها نیز، همانند ربات‌ها برای انجام برخی از عملیات از پیش تعریف شده طراحی شده‌اند[۴]، که به صورت خودکار به انجام فعالیت های خود می‌پردازند. بات‌نت به معنی شبکه‌ای از بات‌ها می باشد، بدین صورت که مهاجم با به تصرف در آوردن تعداد زیادی از میزبان‌ها و نصب برنامه‌ی مخرب خود، میزبان‌ها را به بات تبدیل کرده و شبکه ای از بات‌ها را برای خود راه اندازی می کند، سپس از طریق فرامین ارسالی خود، آنها را کنترل می کند؛ در نتیجه مهاجم می تواند از توان پردازشی میزبان‌های به تصرف درآمده، به صورت توزیع شده، و به نفع خود بهره برداری کند.

همچنین بات‌نت، مهاجم را قادر می سازد تا انواع مختلفی از حملات را به صورت هماهنگ و با قدرت تخریبی بسیار بالا بر روی قربانی سازماندهی کند، و این در حالی است که هویت مهاجم مخفی می ماند. مهمترین خصوصیتی که بات‌نت‌ها را از دیگر بدافزارها متمایز می سازد این است

---

[1] Scanning
[2] Distributed denial of service (DDos)
[3] Spam
[4] fraudulent activities

که نرم افزارهای بات، گرداننده‌ی خود را قادر می سازند تا هر سیستمی را از راه دور کنترل کند، در حالی که قربانیان کاملا از آن بی اطلاع هستند.

بات‌نت‌های خطرناکی که امروزه یافت می شوند، در واقع ترکیبی از حملات گذشته هستند، آن‌ها می توانند، مانند کرم ها منتشر شوند، خودشان را همانند ویروس‌ها از کشف شدن پنهان کنند، مانند خیلی از ابزارهای مستقل هجومی، دارای دستورات مجتمع و سیستم کنترل هستند. مهم‌تر از همه اینکه، امروزه ساخت بات ها یک تلاش مبتنی بر همکاری است.

بر خلاف تشخیص بات‌نت‌هایی با معماری قدیمی، که در سال های اخیر پژوهش‌هایی در مورد آن انجام شده است، پژوهشگران کمی به تشخیص بات‌نت‌هایی با معماری توزیع شده (مبتنی بر ساختار نظیر به نظیر) پرداخته‌اند. چند روش برای تشخیص جریان بات نت پشنهاد شده است که تا به امروز بهترین آنها، روش Gu وهمکارانش (botminer) می باشد [۵]. این روش با استفاده از یادگیری ماشین، جریان‌های ارتباطی مشابه و جریان‌های بد خواه مشابه را خوشه بندی می‌کند. سپس یک همبستگی بین خوشه‌ای انجام می دهد تا میزبان‌هایی را که الگوی ارتباطی مشابه و الگوی فعالیت بد خواهانه مشابه مشترکی دارند، شناسایی کند.

۱-۲- صورت مساله

با افزایش حملات اینترنتی، به دلایلی همچون مسائل سیاسی، اقتصادی، اجتماعی وغیره، وهمچنین شیوع استفاده از بات‌نت‌ها، آنها ابزاری خطرناک که دارای قدرت تخریبی بسیار بالایی هستند تبدیل شده‌اند. همچنین پیشرفت بات‌نت‌ها و انتقال آنها از معماری قدیم (IRC ، Http) به معماری جدید [۶]، که در معماری قدیمی تک نقطه وجود داشت بطوری‌که یک سرور در میان بات ماستر و بقی بات‌ها برای ارسال فرمان وجود دارد، اما با نظارت بر این سرور می توان ارتباط میان آنها را شناسایی‌کرده، آنها را از بین برد. به همین دلیل بات‌نت‌های جدیدی با معماری انعطاف پذیر طراحی شده‌اند، که این معماری از پروتکل‌های شبکه های نظیر به نظیر و تکنیک‌های رمزنگاری برای مخفی سازی بات‌ها و محتوای جریان آنها، استفاده می کند. معماری نظیر به نظیر به بات‌نت این امکان را می دهد که هر بات در آن واحد بتواند یک سرور برای انتقال فرمان بات ماستر به بقیه بات‌ها و یک بات برای حمله به قربانی باشد.

چالش‌های کشف باتنت مبتنی بر P2P عبارتنداز:

- بات‌ها خود را به صورت مخفیانه در سیستم نگه می دارند، و به سیستم صدمه نمی زنند تا از آگاهی کاربر اجتناب کنند. به عنوان مثال بات‌ها از منبع حافظه یا از پردازنده سیستم زیاد استفاده نمی کنند. علاوه بر این می توانند از برنامه‌های آنتی ویروسها با استفاده از روش های Rootkit خود را مخفی کنند[۷، ۸]، به همین دلیل روش‌های مبتنی بر میزبان بی تأثیر می‌باشند، وما در این پایان نامه با استفاده از روش مبتنی بر شبکه به انجام کار می‌پردازیم.
- باتنت بصورت سریع می تواند کد باینری خود را به روز رسانی کند، که به جای ارسال فرمان حمله می تواند، فرمان را بعنوان تغییر کد باینری به کد دیگری، برای باتها ارسال کند. این ویژگی در باتنت‌ها باعث می شود، روش‌های مبتنی بر امضا که بیشتر نرم افزارهای آنتی ویروس از آن استفاده می کنند، در کشف باتنت‌ها، تأثیری نداشته باشند.
- حمله بات نت یک فرآیند چند مرحله‌ای می باشد، که عبارتست از: ۱) مرحله‌ی مسموم کردن سیستم‌ها، ۲) مرحله‌ی برقراری ارتباط با بات ماستر وبقیه بات ها، ۳)مرحله شروع حمله به قربانی. البته می توان جلوی این مراحل را گرفت، اگر باتها از کانال های ارتباطی IRC یا Http استفاده می کردند. اما با پیشرفت روش‌های ارتباطی و انتقال باتها به حالت توزیع شده که در آن هر بات می تواند به عنوان سرور برای ارائه دستور بات ماستر به دیگران عمل کند یا بعنوان بات معمولی در حمله شرکت کند، کشف بات نت مشکلتر خواهد شد، شکل (۱-۱) فرآیند چند مرحله‌ای باتنت را نشان می‌دهد.
- ترافیک بین باتها شباهت زیادی به ترافیک شبکه‌های نظیر به نظیر دارد، واین امر باعث می‌شود که باتنت جریان خود را در درون جریان شبکه‌های نظیر به نظیر مخفی کند، در نتیجه تفکیک این دو جریان از هم مشکل بزرگی است.