



دانشگاه تربیت مدرس  
دانشگاه تربیت مدرس  
دانشکده برق و کامپیوتر

پایان نامه برای دریافت درجه کارشناسی ارشد  
رشته مهندسی کامپیوتر گرایش نرم افزار

## کشف بات نت مبتنی بر شبکه های نظیر به نظیر با استفاده از رویکردهای یادگیری ماشین

نگارنده

عبدالعزیز الیوسف

استاد راهنما

دکتر سعید جلیلی

تیر ۱۳۹۰

سلام الغزالي



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 سترم افزار پیشنهاد می کنند.

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

## 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 پایان‌نامه‌ها / 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 پایان‌نامه / 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 پایان‌نامه / رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با مجوز کتبی صادره از طریق حوزه پژوهشی دانشگاه و بر اساس آئین‌نامه‌های 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 پایان‌نامه / 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 لازم‌الاجرا است و هرگونه تخلف از مفاد این دستورالعمل، از طریق مراجع قانونی قابل پیگیری می‌شود.

نام و نام خانوادگی: عبد العزیز الیوسف



امضاء



دانشگاه تربیت مدرس  
دانشگاه تربیت مدرس  
دانشکده فنی و مهندسی

پایان نامه برای دریافت درجه کارشناسی ارشد  
رشته مهندسی کامپیوتر گرایش نرم افزار

## کشف بات نت مبتنی بر شبکه های نظیر به نظیر با استفاده از رویکردهای یادگیری ماشین

نگارنده

عبدالعزیز 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 بوده است.

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

## تشکر و قدردانی:

بر خود لازم می‌دانم سپاسگزار عزیزان و سرورانی باشم که در لحظات زندگی و دوران تحصیل در کنارم بودند.

قدر دان زحمات اساتید گرانقدر دانشکده مهندسی کامپیوتر، بخصوص جناب آقای دکتر جلیلی هستم که با صبر و بزرگواری نه‌تنها در انجام این پایان‌نامه بلکه در تمام این دوره مرا یاری نمودند.

همچنین تشکر و قدردانی می‌نمایم از جناب آقای دکتر آبادی، جناب آقای دکتر اکبری، 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 معمولاً 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 معمولاً برای هدایت فعالیتهای مختلفی مورد استفاده قرار می‌گیرند، این فعالیتهای 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 می‌کنند. 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 می‌باشد. 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 می‌شود، بلکه برای ساخت باتنت‌ها با توان بیشتر هم مورد استفاده قرار می‌گیرد. هر بات در این نوع شبکه می‌تواند نقش فرماندهی و کنترل 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 شبکه‌های نظیر به نظیر باعث می‌شود که جریان ارتباطی بات‌ها بین هم شباهت زیادی با جریان ارتباطی شبکه‌های 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، و برای 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 شبکه‌های نظیر به نظیر یک مدل ساخته شد و هر جریانی که از این مدل پیروی نمی‌کند، به عنوان جریان بات محسوب می‌شود. 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 و جریان حمله، بات‌ها را تشخیص داد، ما فقط از جریان ارتباطی بات‌ها با هم و با تعریف خصیصه‌های مؤثر که هزینه بار محاسباتی را کم می‌کند، 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.

**واژه‌های کلیدی:** بات‌نت، شبکه‌های نظیر به نظیر، تحلیل رفتاری جریان، یادگیری ماشین، دسته بندی، تشخیص ناهنجار، خصیصه.

## فهرست مطالب

۱	فصل اول: کلیات
۲	۱-۱- مقدمه
۳	۲-۱- صورت مساله
۵	۳-۱- اهداف و نوآوری‌های پژوهش
۶	۴-۱- مروری بر فصول پایان نامه
۸	فصل دوم: مفاهیم پایه
۹	۱-۲- مقدمه
۹	۲-۲- تعریف
۱۰	۳-۲- معرفی شبکه‌های نظیر به نظیر
۱۰	۱-۳-۲- تعریف شبکه‌های نظیر به نظیر
۱۰	۲-۳-۲- ساختار شبکه‌های نظیر به نظیر
۱۴	۴-۲- بات‌نت‌های نظیر به نظیر
۱۸	۵-۲- چرخه‌ی حیات بات‌نت
۱۸	۱-۵-۲- مرحله‌ی ساخت
۲۱	۲-۵-۲- مرحله‌ی نگهداری
۲۱	۳-۵-۲- مرحله‌ی حمله
۲۳	۴-۵-۲- مرحله‌ی بازسازی
۲۳	۶-۲- فرمان و کنترل
۲۴	۱-۶-۲- متمرکز
۲۴	۱-۶-۲- غیر متمرکز
۲۵	۱-۶-۲- ترکیبی نظیر به نظیر
۲۷	۷-۲- جمع بندی
۲۸	فصل سوم: تاریخچه پژوهش در تشخیص باتنت مبتنی بر P2P
۲۹	۱-۳- مقدمه
۲۹	۲-۳- روش‌های مبتنی بر میزبان
۳۲	۳-۳- روش‌های مبتنی بر شبکه
۳۲	۱-۳-۳- روش‌های تشخیص ناهنجاری رفتار شبکه
۳۶	۲-۳-۳- روش‌های ساختاری
۳۸	۴-۳- مقایسات
۳۹	۵-۳- جمع بندی

فصل چهارم: روش پیشنهادی تشخیص باتنت مبتنی بر P2P..... ۴۰

- ۴-۱- مقدمه ..... ۴۱
- ۴-۲- توصیف جریان ..... ۴۲
- ۴-۳- انتخاب خصیصه ..... ۴۲
- ۴-۳-۱- معرفی خصیصه‌های مرحله‌ی اول ..... ۴۲
- ۴-۳-۲- معرفی خصیصه‌های مرحله‌ی دوم ..... ۴۳
- ۴-۴- روش انتخاب خصیصه ..... ۴۵
- ۴-۵- مرحله اول (دسته‌بندی)..... ۴۶
- ۴-۵-۱- روش‌های یادگیری ماشین ..... ۴۷
- ۴-۶- مرحله دوم ( تشخیص ناهنجاری)..... ۴۹
- ۴-۷- جمع‌بندی ..... ۵۰

فصل پنجم: ارزیابی روش پیشنهادی ..... ۵۱

- ۵-۱- مقدمه ..... ۵۲
- ۵-۲- فرآیند تولید ترافیک ..... ۵۲
- ۵-۲-۱- مجموعه ترافیک اول ..... ۵۲
- ۵-۲-۲- مجموعه ترافیک دوم ..... ۵۴
- ۵-۳- فرآیند تولید داده‌های آموزشی ..... ۵۴
- ۵-۴- مدل ارزیابی ..... ۵۵
- ۵-۵- ارزیابی خصیصه‌ها ..... ۵۷
- ۵-۵-۱- ارزیابی خصیصه‌های مرحله‌ی اول روش پیشنهادی ( دسته‌بندی) ..... ۵۷
- ۵-۵-۲- ارزیابی خصیصه‌های مرحله‌ی دوم روش پیشنهادی (نشخیص ناهنجار)..... ۵۸
- ۵-۶- مقایسه‌ی روش‌های یادگیری ماشین ..... ۵۹
- ۵-۷- مدل ارزیابی مرحله‌ی تشخیص ناهنجاری ..... ۶۲
- ۵-۸- ارزیابی مرحله‌ی دوم ( تشخیص ناهنجاری) ..... ۶۲
- ۵-۸-۱- تعداد جریان‌های آموزشی و دقت تشخیص ..... ۶۳
- ۵-۸-۲- تعداد جریان‌های آموزشی و حد آستانه ..... ۶۴
- ۵-۹- بهبود دقت مرحله‌ی دوم ( تشخیص ناهنجاری ترکیبی) ..... ۶۵
- ۵-۱۰- کشف باتنت در مرحله‌ی اول (دسته‌بندی)..... ۶۵
- ۵-۱۱- برتری روش پیشنهادی در مقایسه با سایر روش‌های تشخیص باتنت ..... ۶۶
- ۵-۱۲- جمع بندی ..... ۶۷

فصل ششم: نتیجه گیری ..... ۶۸

- ۶-۱- هدف پایان نامه ..... ۶۹

۶۹.....	۲-۶- نتایج حاصل از پژوهش .....
۷۰.....	۳-۶- بررسی میزان حصول اهداف اولیه پژوهش .....
۷۱.....	۴-۶- پژوهش‌های آینده .....
۷۲.....	مراجع .....
۷۶.....	واژه نامه انگلیسی- فارسی .....
۷۹.....	واژه نامه فارسی- انگلیسی .....
۸۲.....	ضمائم .....
۸۲.....	ضمیمه الف: شرح خصیصه‌های تولید شده توسط ابزار TCPTrace .....

## فهرست شکل‌ها

- فصل اول: کلیات ..... ۱
- شکل (۱- ۱) فرآیند چند مرحله‌ی باتنت ..... ۵
- فصل دوم: مفاهیم پایه ..... ۸
- شکل (۱- ۲) شمای کلی شبکه‌های Overlay ..... ۱۱
- شکل (۲- ۲) شبکه‌های نظیر به نظیر متمرکز ..... ۱۱
- شکل (۳- ۲) شبکه‌های نظیر به نظیر خالص ..... ۱۲
- شکل (۴- ۲) شبکه‌های نظیر به نظیر ترکیبی ..... ۱۳
- شکل (۵- ۲) شبکه‌ی ساخت یافته‌ی Chord ..... ۱۴
- شکل (۶- ۲) رویه کنترل و فرمان باتنت Waledac ..... ۱۵
- شکل (۷- ۲) نمونه از باتنت مبتنی بر P2P ..... ۲۵
- شکل (۸- ۲) نمونه از معماری باتنت ترکیبی نظیر به نظیر ..... ۲۶
- فصل سوم: تاریخچه پژوهش در تشخیص باتنت مبتنی بر P2P ..... ۲۸
- شکل (۱- ۳) نمودار افزایش بسته‌های UDP, ICMP ..... ۳۳
- شکل (۲- ۳) معماری Botminer ..... ۳۴
- شکل (۳- ۳) ارگادیک مارکوف مدل برای باتنت و ماتریس انتقال حالت ..... ۳۵
- شکل (۴- ۳) پدیده IP aliasing ..... ۳۷
- شکل (۵- ۳) پدیده ID aliasing ..... ۳۷
- فصل چهارم: روش پیشنهادی تشخیص باتنت مبتنی بر P2P ..... ۴۰
- شکل (۱- ۴) سیستم پیشنهادی برای تشخیص باتنت مبتنی بر P2P ..... ۴۱
- شکل (۲- ۴) تفاوت تعداد بسته‌های منتقل شده بین شبکه‌های نظیر به نظیر و باتنت‌ها ..... ۴۴
- شکل (۳- ۴) توزیع تجمعی از میانگین حجم جریان بین شبکه‌های P2P و باتنت ..... ۴۵
- شکل (۴- ۴) شاخه‌های از درخت تصمیم ساخته شده در فاز یادگیری مجموعه داده جریان‌های دو گروه P2P و non-P2P ..... ۴۸
- شکل (۵- ۴) فاز آموزش و آزمون در مرحله تشخیص ناهنجاری ..... ۴۹
- فصل پنجم: ارزیابی روش پیشنهادی ..... ۵۱

- شکل (۵-۱) محیط جمع‌آوری ترافیک ..... ۵۳
- شکل (۵-۲) تقسیم‌بندی مجموعه داده اول به سه مجموعه ..... ۵۵
- شکل (۵-۳) دقت تشخیص الگوریتم درخت تصمیم روی مجموعه‌های داده‌ها با خصیصه‌های متفاوت برای انتخاب بهترین مجموعه خصیصه ..... ۵۷
- شکل (۵-۴) دقت تشخیص الگوریتم درخت تصمیم روی مجموعه‌های داده‌هایی با خصیصه‌های متفاوت به منظور انتخاب بهترین مجموعه خصیصه برای مرحله تشخیص رفتار غیر عادی ..... ۵۸
- شکل (۵-۵) مقایسه دقت سه الگوریتم یادگیری ماشین ..... ۵۹
- شکل (۵-۶) مقایسه دقت الگوریتم GMM و OCSVM با استفاده از ۲۰۰۰ جریان برای تولید مدل و ۲۰۰۰ جریان برای تعیین حد آستانه ..... ۶۳
- شکل (۵-۷) تاثیر تعداد جریان‌ها بر دقت ..... ۶۴
- شکل (۵-۸) تاثیر تعداد جریان‌های آموزش مدل بر تعیین حد آستانه و دقت تشخیص ..... ۶۴
- شکل (۵-۹) دقت تشخیص الگوریتم‌های دسته‌بندی روی مجموعه‌های داده‌ها با خصیصه‌های متفاوت ..... ۶۶

## فهرست جدول‌ها

فصل اول: کلیات	۱
فصل دوم: مفاهیم پایه	۸
فصل سوم: تاریخچه پژوهش در تشخیص باتنت مبتنی بر P2P	۲۸
جدول (۱-۳) دقت تشخیص در حالتی که ماکسیمم نسبت درست‌نمایی $\leq 0.2$	۳۵
جدول (۲-۳) مقایسه‌ی روش‌های تشخیص باتنت	۳۹
فصل چهارم: روش پیشنهادی تشخیص باتنت مبتنی بر P2P	۴۰
جدول (۱-۴) تعداد خصیصه‌های بدست آمده با استفاده از الگوریتم‌های انتخاب خصیصه	۴۶
فصل پنجم: ارزیابی روش پیشنهادی	۵۱
جدول (۱-۵) ترکیب ترافیک اول	۵۳
جدول (۲-۵) ترکیب جریانهای پروتکلها در مجموعه‌ی ترافیک دوم	۵۴
جدول (۳-۵) تقسیم نمونه‌ها بر اساس درستی و مثبت بودن	۵۶
جدول (۴-۵) توزیع اشتباهات بین جریان‌های نظیر به نظیر و باتنت با استفاده از C4.5	۶۰
جدول (۵-۵) دقت تشخیص پروتکل‌ها با استفاده از C4.5	۶۰
جدول (۶-۵) توزیع اشتباهات جریان‌ها با استفاده از الگوریتم شبکه‌ی بیزی	۶۱
جدول (۷-۵) دقت تشخیص جریان‌ها با استفاده از شبکه‌ی بیزی	۶۱
جدول (۸-۵) مقایسه‌ی دقت تشخیص وزمان محاسباتی در حالت‌های تشخیص ناهنجاری	۶۵

## 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 می‌دهد، اما با افزایش استفاده از اینترنت چالش‌های امنیتی بزرگی ظاهر شده است، و این امر به مسأله‌ی مهم و حساسی برای بسیاری از اشخاص و شرکت‌ها تبدیل شده است.

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 اخیر باتنت‌ها نیز به این مجموعه اضافه شده‌اند. باتنت‌ها به یک منبع اساسی برای بسیاری از حملات خطرناک از قبیل پویش<sup>۱</sup>، جلوگیری از سرویس توزیع شده<sup>۲</sup>، هرزنامه<sup>۳</sup>، فعالیت‌های جعلی<sup>۴</sup>، 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 می‌شود، بات‌ها نیز، همانند ربات‌ها برای انجام برخی از عملیات از پیش تعریف شده طراحی شده‌اند [۴]، که به صورت خودکار به انجام فعالیت‌های خود می‌پردازند. باتنت به معنی شبکه‌ای از بات‌ها می‌باشد، 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 شبکه‌ای از بات‌ها را برای خود راه اندازی می‌کند، سپس از طریق فرامین ارسالی خود، آنها را کنترل می‌کند؛ در نتیجه مهاجم می‌تواند از توان پردازشی میزبان‌های به تصرف درآمده، به صورت توزیع شده، و به نفع خود بهره برداری کند.

همچنین باتنت، مهاجم را قادر می‌سازد 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 می‌ماند. مهمترین خصوصیتی که باتنت‌ها را از دیگر بدافزارها متمایز می‌سازد این است

---

<sup>1</sup> Scanning

<sup>2</sup> Distributed denial of service (DDos)

<sup>3</sup> Spam

<sup>4</sup> fraudulent activities

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

بات‌های خطرناکی که امروزه یافت می شوند، در واقع ترکیبی از حملات گذشته هستند، آن‌ها می توانند، مانند کرم‌ها 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 بات‌ها یک تلاش مبتنی بر همکاری است.

بر خلاف تشخیص بات‌هایی با معماری قدیمی، که در سال‌های اخیر پژوهش‌هایی در مورد آن انجام شده است، پژوهشگران کمی به تشخیص بات‌هایی با معماری توزیع شده (مبتنی بر ساختار نظیر به نظیر) پرداخته‌اند. چند روش برای تشخیص جریان بات‌نت پیشنهاد شده است که تا به امروز بهترین آنها، روش Gu و همکارانش (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 و غیره، و همچنین 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 بطوری که 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 شناسایی کرده، آنها را از بین برد. به همین دلیل بات‌نت‌های جدیدی با معماری انعطاف پذیر طراحی شده‌اند، که این معماری از پروتکل‌های شبکه‌های 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 عبارتند از:

- بات‌ها خود را به صورت مخفیانه در سیستم نگه می‌دارند، و به سیستم صدمه نمی‌زنند تا از آگاهی کاربر اجتناب کنند. به عنوان مثال بات‌ها از منبع حافظه یا از پردازنده سیستم زیاد استفاده نمی‌کنند. علاوه بر این می‌توانند از برنامه‌های آنتی ویروسها با استفاده از روش‌های Rootkit خود را مخفی کنند [۷، ۸]، به همین دلیل روش‌های مبتنی بر میزبان بی‌تأثیر می‌باشند، و ما در این پایان نامه با استفاده از روش مبتنی بر شبکه به انجام کار می‌پردازیم.
- بات‌نت بصورت سریع می‌تواند کد باینری خود را به روز رسانی کند، که به جای ارسال فرمان حمله می‌تواند، فرمان را بعنوان تغییر کد باینری به کد دیگری، برای بات‌ها ارسال کند. این ویژگی در بات‌نت‌ها باعث می‌شود، روش‌های مبتنی بر امضا که بیشتر نرم افزارهای آنتی ویروس از آن استفاده می‌کنند، در کشف بات‌نت‌ها، تأثیری نداشته باشند.
- حمله بات‌نت یک فرآیند چند مرحله‌ای می‌باشد، که عبارتست از: (۱) مرحله‌ی مسموم کردن سیستم‌ها، (۲) مرحله‌ی برقراری ارتباط با بات ماستر وبقیه بات‌ها، (۳) مرحله شروع حمله به قربانی. البته می‌توان جلوی این مراحل را گرفت، اگر بات‌ها از کانال‌های ارتباطی IRC یا Http استفاده می‌کردند. اما با پیشرفت روش‌های ارتباطی و انتقال بات‌ها به حالت توزیع شده که در آن هر بات می‌تواند به عنوان سرور برای ارائه دستور بات ماستر به دیگران عمل کند یا بعنوان بات معمولی در حمله شرکت کند، کشف بات‌نت 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، و این 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.