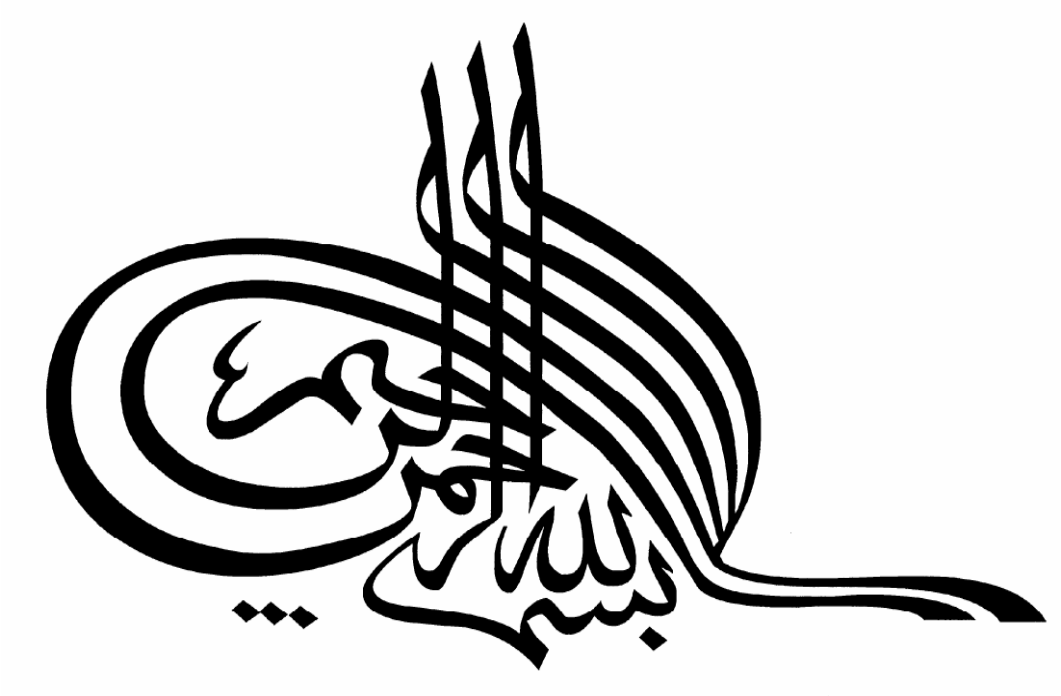
بسم الله الرحمن الرحيم

دانشگاه شهید اشرفی اصفهانی

دانشکده فنی مهندسی

گروه کامپیوتر

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیکز

عنوان پایان نامه

تشخیص بدافزار بر اساس تحلیل رفتار مبتنی بر هم‌افزایی آتوماتای یادگیر و شبکه عصبی

دانشگاه شهید اشرفی اصفهانی

دانشکده فنی مهندسی

گروه کامپیوتر

پایان‌نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیکز ..... تحت عنوان

تشخیص بدافزار بر اساس تحلیل رفتار مبتنی بر هم‌افزایی آتوماتای یادگیر و شبکه عصبی

در تاریخ ..... توسط هیات داوران بررسی و با درجه .... به تصویب نهایی رسید.

۱- استاد راهنمای پایان نامه دکتر ..... با مرتبه علمی .....

امضاء(12)

۲- استاد مشاور پایان نامه دکتر ..... با مرتبه علمی .....

امضاء(12)

۳- استاد داور داخل گروه دکتر ..... با مرتبه علمی .....

امضاء(12)

۴- استاد داور خارج از گروه دکتر ..... با مرتبه علمی .....

امضاء(12)

مهر و امضا مدیر گروه

چکیده:

بخاطر رشد سریع تولید بدافزار و آسیب‌های بالقوه آن‌ها، نیاز به سیستم‌های هوشمند و خودکار برای تشخیص بدافزار نیز رشد کرده است. اگرچه، محصولات آنتی‌ویروس که بر اساس امضا قادر به تشخیص کد بدخواه هستند، موثر جلوه داده شده‌اند؛ اما در برابر حملات جدید و ناشناخته، کارآمدی ندارند. در این پایان‌نامه، یک سیستم تشخیص بدافزار پویای جدید بر اساس کاوش دنباله API ارائه شده است. در روش پیشنهادی پایگاه داده‌ای متشکل از فایل های اجرایی سالم و مخرب ایجاد شد. براساس روش نوین مبتنی بر رفتار فراخوانی‌های سیستمی با اجرای مخرب و غیربدافزار در محیطی کنترل‌شده بدست می‌آید. بعلاوه، یک تکنیک اشکال‌زدایی ارائه می‌شود که قادر به غلبه بر کمبودهای تکنیک‌های اشکال‌زدایی در محیط‌های کنترل شده برای تشخیص بدافزار می‌باشد. به منظور کاهش داده‌ها در مرحله انتخاب ویژگی از الگوریتم ماشین بردار پشتیبان به عنوان تابع هزینه و الگوریتم فاخته به عنوان استراتژی جستجو برای انتخاب زیرمجموعه‌ای بهینه از ویژگی‌ها استفاده شده است. روش پیشنهادی روشی با دقت بالا است که با استفاده از بهینه‌سازی شبکه عصبی توسط آتوماتای یادگیر به دقت ۹۹,۰۶ درصد برای تشخیص بدافزار دست یافته است.

کلید واژه‌ها: بدافزار، تحلیل رفتار، شبکه عصبی، آتوماتای یادگیر

فهرست مطالب

عنوان.....صفحه

فهرست شکل‌ها

عنوان.....صفحه

فهرست جداول

عنوان.....صفحه

فهرست نمودارها

عنوان.....صفحه

فصل اول

کلیات تحقیق

۱-۱-مقدمه

امروزه با گسترش فناوری اطلاعات در زندگی بشر و وابستگی هر چه بیشتر آن به کسب‌وکار، محافظت از اطلاعات، به منزله شاهرگ حیاتی یک صنعت مدرن محسوب می‌گردد. اطلاعات به عنوان یکی از با ارزش‌ترین و حساس‌ترین دارایی‌های سازمان یا فردی بوده و دستیابی به آن و عرضه به‌موقع و مناسب اطلاعات، همواره دارای نقش محوری و سرنوشت‌ساز می‌باشد. حفظ و نگهداری صحیح اطلاعات شرط لازم برای تداوم فرآیند کسب و کارهای اقتصادی می‌باشد. در چند سال اخیر دسترسی‌های غیرمجاز، حملات، تهدیدات کامپیوتری، رخنه به اطلاعات روی دیسک‌ها، کامپیوترها و استفاده غیرمجاز از آن‌ها تبدیل به معضل شده‌است. این تهدیدات و حملات شامل ویروس‌های جدید و یا انواع هک[1] و نفوذ در سیستم‌های کامپیوتری است (فان[2] و همکاران، ۲۰۱۹). انتشار این‌گونه اخبار باعث شیوع اضطراب و نگرانی در بین کاربرانی می‌شود که به‌صورت مستمر از کامپیوتر بهره می‌گیرند و یا اطلاعاتی ارزشمند

[1] Hack
[2] Fan

بر روی کامپیوترهای خود دارند. لذا سازمان‌ها و شرکت‌ها ناگزیر به‌دنبال‌پیاده‌سازی موارد امنیتی می‌باشند (دینگ[1] و همکاران، ۲۰۱۸).

رشد سریع نرم‌افزارهای مخرب یک تهدید جدی در امنیت جهانی محسوب شده است. نرم‌افزارهای مخرب یا بدافزارها، یک برنامه می‌باشند که با هدف آسیب رساندن به کامپیوتر، شبکه، اطلاعات و غیره توسعه داده شده‌اند. از آن‌جا که این برنامه‌ها معمولاً کاربر را آزار می‌دهند یا خسارتی به وجود می‌آورند، مخرب نامیده شده‌اند(کوبر[2]، ۲۰۱۷). برخی از آنان فقط کاربر را می‌آزارند، مثلاً وی را مجبور به انجام کاری تکراری می‌کنند. اما برخی دیگر سیستم رایانه‌ای و داده‌های آن را هدف قرار می‌دهند که ممکن است خساراتی به‌بار آورند. در عین حال ممکن است هدف آن‌ها سخت‌افزار سیستم کاربر باشد. یک نرم‌افزار برپایه‌ی نیت سازنده آن به عنوان یک بدافزار شناخته می‌شود. ضدبدافزار[3] که در اصطلاح عامیانه، آنتی‌ویروس نامیده می‌شود، درصدد تشخیص بدافزارها می‌باشد. این درحالی است که بدافزارنویس‌ها در تلاشند شیوه‌هایی را به‌کار گیرند که ضدبدافزارهایا آنتی‌ویروس‌ها نتوانند مخرب‌بودن برنامه راتشخیص دهند. به‌همین دلیل رقابت تنگاتنگی میان بدافزارها و آنتی‌ویروس‌ها به‌وجود آمده است (سینگ[4] و سینگ، ۲۰۲۱).

به‌دلیل اهمیت داشتن صحت و محرمانگی اطلاعات و عملکرد صحیح سیستم‌ها برای بسیاری از افراد، شرکت‌ها و سازمان‌ها، یکی از مسائل مهم در زمینه‌ی امنیت کامپیوترها و شبکه‌های کامپیوتری، کشف، تحلیل و جلوگیری از فعالیت این‌گونه نرم‌افزارهای مخرب است. تشخیص نرم‌افزارهای مخرب شاخه‌ای ازامنیت کامپیوتر است. این شاخه برای تجزیه‌وتحلیل برنامه‌های مشکوک، تشخیص نرم‌افزارهای مخرب و درنهایت، از بین بردن تهدید در تلاش است.

۱-۲- بیان مسئله

بدافزار به نرم‌افزارهایی گفته می‌شود که باعث هرگونه خرابی در کامپیوتر شخصی، سرویس دهنده یا شبکه‌ی کامپیوتری شوند. روش‌های تحلیل کد در حالت کلی به دو صورت تحلیل ایستا و تحلیل پویا مطرح می‌باشند. در تحلیل ایستای بدافزار، تمامی مسیرهایی از کد که بدافزار ممکن است در هر اجرای خود طی

[1] Ding
[2] Kuber
[3] Anti-malware
[4] Singh

کند، شناسایی و تحلیل می‌گردد. در این روش تحلیل، فایل اجرایی بدون در نظر گرفتن دستورات آن بررسی می‌شود. این روش بسیار سریع و ساده است، اما به شکل قابل ملاحظه‌ای در برابر بدافزارهای پیچیده، ناکارآمد می‌باشد و یک سری ویژگی‌های رفتاری مهم را در نظر نمی‌گیرد. در تحلیل پویای کد، با اجرای باینری بدافزار، رفتار وخصوصیات اجرایی آن شناسایی می‌شوند. این روش برای حذف آلودگی در سیستم و تولید امضاهای موثر خود را درگیر اجرای بدافزار و مشاهده رفتار آن در سیستم می‌کند.

یکی از اولین مراحل در مقابله با بدافزارها تحلیل بدافزار و استخراج نوع و ویژگی‌های آن است. دو رده اصلی از روش‌های تشخیص بدافزار عبارتند از: تشخیص بر اساس امضا و تحلیل رفتار. در روش مبتنی بر امضا با استفاده از یک دیکشنری ویروس که حاوی امضای ویروس‌های شناخته شده است، احتمال وجود ویروس‌های شناخته شده در فایل‌ها مورد بررسی و آزمایش قرار می‌گیرند. مشکل اصلی تکنیک بررسی امضا این است که نرم افزار آنتی ویروس باید قبلا بروز رسانی شده باشد تا بتواند به مقابله و خنثی‌سازی بدافزارها بپردازد و لذا بدافزارهای جدیدی که هنوز شناسایی نشده و به فایل‌های امضا اضافه نشده‌اند تشخیص داده نمی‌شوند. روش مبتنی بر رفتار بر خلاف روش پیشین تنها در تلاش برای شناسایی ویروس-های شناخته شده نیست و به جای آن رفتار تمام برنامه‌ها را نظارت می‌کند. این تکنیک سعی در تشخیص انواع شناخته شده و همچنین انواع جدید بدافزار دارد و این کار را از طریق جست‌وجوی ویژگی‌های عمومی و مشترک بدافزارها انجام می‌دهد.

امروزه افزایش انواع مختلف بدافزارها چالش بزرگی در صنعت آنتی ویروس‌ها ایجاد کرده است. برای محققان این موضوع که چگونه روند رو به رشد نمونه‌های بدافزارها را بطور موثر پردازش کنند و تکنیکی سریع برای محافظت کاربران ارائه دهند، یک زمینه‌ی تحقیقاتی مهم محسوب می‌شود. در این پایان‌نامه روشی ارائه شده که بتواند بدافزار را بر اساس تحلیل رفتار و با استفاده از شبکه‌ی عصبی و آتوماتای یادگیر تشخیص دهد.

۱-۳- اهمیت موضوع

تهدیدها علیه امنیت داده‌ها و اطلاعات و حمله‌های نرم‌افزاری مخرب، فرآیندی پیچیده است. تنوع و تعداد این حملات و تهدیدها نتیجه فراهم‌کردن انواع مختلف روش‌های دفاع در مقابل آن‌هاست؛ اما متأسفانه فناوری‌های تشخیص امروزی برای مقابله با رویکردهای جدید طراحان بدافزار که از آن‌ها برای گریز از ضد مخرب‌ها استفاده می‌کنند، موثر نیستند. محققین تلاش فراوان انجام داده‌اند تا سیستم‌های ضدبدافزار با روش‌های مؤثر تشخیص بدافزار ارائه دهند و از سیستم‌های کامپیوتری محافظت کنند. این تهدیدات سبب بوجود آمدن سیستم‌های تشخیص بدافزار شده است. در واقع یک سیستم تشخیص بدافزار سیستمی است که برای تعیین این مسأله بکار می‌رود، که آیا یک برنامه قصد خرابکاری دارد یا خیر. امروزه فعالیت‌های مخرب وارد مرحله جدیدی شده است که در آن کدهای مخرب نه تنها سیستم‌ها را آلوده می‌نمایند بلکه اطلاعات شخصی کاربران را به سرقت می‌برند. بر همین اساس تعداد و اثرات زیان‌بار بدافزارها به طور روزافزون در حال افزایش است (کوماتوار و کوکار[1]، ۲۰۲۱).

تشخیص به‌موقع این حملات، اهمیت بالایی در حفاظت از سیستم‌های هوشمند دارد. برای تشخیص چنین بدافزارهایی روش‌های مختلفی وجود دارد که از آلوده شدن سیستم کامپیوتری و از دست رفتن اطلاعات ممانعت می‌کند. روش‌های سنتی مختلفی برای تشخیص بدافزارها وجود دارد اما با درنظر گرفتن این‌که بدافزارها با استفاده از روش‌های پنهان‌سازی بسیار پیچیده شده‌اند، نیاز به روش‌های پیشرفته‌تری مانند داده‌کاوی و یادگیری ماشین برای تشخیص آنها احساس می‌شود که در این پایان‌نامه مورد بررسی و استفاده قرار گرفته شده است.

۱-۴- اهداف پایان‌نامه

با توجه به اهمیت مسئله تشخیص بدافزارها و با درنظر گرفتن تاریخچه تشخیص بدافزارها و پس از تحلیل روش‌های پیشین ارائه شده در این پایان‌نامه برای رسیدن به پاسخ این مسئله یک راه‌حل با اهداف زیر ارائه شده است.

[1] Komatwar & Kokare

۱- ارائه روشی که بتواند بدافزار را با دقت بالاتری شناسایی کند.

۲- بررسی اثربخشی و تحلیل خودکار اعمال شده در ارزیابی تشخیص بدافزار

۳- کاهش هشدارهای نادرست در حین شناسایی بدافزارها

۴- بهبود صحت و درستی الگوریتم پیشنهادی نسبت به سایر الگوریتم‌های موجود

۱-۵- نوآوری پایان‌نامه

در این پایان‌نامه روشی نوین ارائه شده است که بتواند بدافزار را بصورت پویا و با استفاده از داده‌کاوی، با دقت بالایی شناسایی کند. در گام نخست مجموعه‌ای از فایل‌های نمونه‌ی بدافزار و سالم گردآوری می‌شوند. سپس برای استخراج ویژگی‌های رفتاری بدافزار نیاز به اجرای آن‌ها می‌باشد و بدلیل اینکه اجرای بدافزار ممکن است آسیب‌های جبران‌ناپذیری به سیستم وارد کند، ما در اینجا برای استخراج ویژگی‌های رفتاری از اجرای بدافزار در ماشین مجازی استفاده خواهیم کرد. پس از آن، پیش‌پردازش‌های لازم و تبدیل فایل‌های اجرایی به گزارش رفتاری و استخراج فراخوانی‌های API توسط نرم‌افزار WinAPIOverride32 صورت می‌گیرد. در مرحله‌ی تشخیص از شبکه عصبی بهمراه آتوماتای یادگیر بهره خواهیم برد. شبکه عصبی مصنوعی یک سامانه پردازشی داده‌ها است. در این شبکه‌ها به کمک دانش برنامه‌نویسی، ساختار داده‌ای طراحی می‌شود که می‌تواند همانند نرون عمل کند. که به این ساختار داده نرون گفته می‌شود. بعد با ایجاد شبکه‌ای بین این نرون‌ها و اعمال یک الگوریتم آموزشی به آن، شبکه را آموزش می‌دهند، تنظیمات شبکه عصبی مانند تعداد لایه‌ها، تعداد نرون‌ها عموما با سعی و خطا انجام می‌شود لذا اگر سیستمی طراحی گردید که تعداد لایه‌ها و تعداد نرون‌ها شبکه را بتواند مشخص نماید، شبکه عصبی کارایی بهتری دارد. در این پژوهش قصد داریم به کمک شبکه عصبی فایل‌های بدافزار را تشخیص دهیم. سپس به کمک یادگیری تقویتی پارامترهای شبکه عصبی تنظیم می‌گردد تا ساختار فایل‌های مخرب با دقت بهتر تشخیص داده شود. الگوریتم مورد استفاده در یادگیری تقویتی الگوریتم اتوماتای یادگیر است که وظیفه آن تنظیم پارامتر شبکه عصبی مانند تعداد لایه‌ها، تعداد نرون، نوع نرون‌ها و غیره است.

۱-۶- ساختار پایان‌نامه

در ادامه پایان‌نامه به‌صورت زیر سازماندهی شده است.

در فصل دو انواع مختلف روش‌های تجزیه و تحلیل بدافزار با استفاده از تکنیک‌های داده‌کاوی بررسی شده است. در این فصل تمامی تکنیک‌های تشخیص بدافزار و نحوه‌ی استفاده از آن‌ها به طور کامل شرح داده شده است. در فصل سوم مروری بر پیشینه‌های پژوهشی و کارهای انجام شده مرتبط با این پایان‌نامه پرداخته شده است و آخرین تحقیقات انجام شده در زمینه بدافزار مورد بررسی قرار گرفته است. در فصل چهارم معماری سیستم پیشنهادی مورد بحث قرارگرفته می‌شود. در فصل پنجم، شبیه‌سازی و ارزیابی روش پیشنهادی مطرح خواهد شد، درنهایت نتیجه‌گیری کلی از پایان‌نامه گرفته‌شده و پیشنهادهایی برای توسعه و بهبود روش‌های پیشنهادی، ارائه می‌شود.

فصل دوم

مبانی و ادبیات تحقیق

امروزه به‌دلیل آسیب‌ها و حملات مخرب ناشی از بدافزارها به سیستم‌های کامپیوتری، نیاز به شناسایی این بدافزارها به امری ضروری تبدیل شده است. علاوه‌بر آن در شبکه جهانی اینترنت، بدافزارها مبنای اصلی حملات امنیتی به‌شمار می‌روند؛ بنابراین نیاز به شناسایی و تحلیل آن‌ها ضروری است. در این فصل، تعاریف پایه، اهداف و طبقه‌بندی‌های مختلف برای بدافزارها ارائه شده است.

۲-۱- بدافزار

تعاریف متفاوتی برای توصیف بدافزار یا نرم‌افزارهای مخرب مطرح شده است. به‌طور مثال، مک‌گرو و موریست بدافزار را هر کدی که عمداً باعث آسیب‌رسانی یا متوقف کردن عملکرد سیستم شود، تعریف کرده‌اند (مک‌گرو و موریست[1]، ۲۰۰۰). امروزه بدافزارها قابلیت ایجاد خرابی در سطح وسیعی را دارند. با توجه به این‌که تولید آن‌ها به‌صورت روزافزون رو به رشد است، شناسایی آن‌ها امری بسیار مهم می‌باشد.

[1] McGraw & Morrisett