



دانشگاه صنعتی خواجه نصیرالدین طوسی
دانشکده مهندسی صنایع

روش‌های تشخیص بدافزار و توسعه شیوه‌های جدید در تشخیص بدافزار

محمود کرمی پور

استاد راهنما: دکتر شهریار محمدی

پایان‌نامه برای دریافت مدرک کارشناسی ارشد

مهندسی فناوری اطلاعات-تجارت الکترونیک

شهریور ۱۳۹۲



دانشگاه صنعتی خواجه نصیرالدین طوسی
دانشکده مهندسی صنایع

روش‌های تشخیص بدافزار و توسعه شیوه‌های جدید در تشخیص بدافزار

محمود کرمی پور

استاد راهنما: دکتر شهریار محمدی

پایان‌نامه برای دریافت مدرک کارشناسی ارشد

مهندسی فناوری اطلاعات-تجارت الکترونیک

شهریور ۱۳۹۲

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

تقدیم بہ پدر و مادرم

کہ از نگاہشان صلابت

از رفتارشان محبت

و از صبرشان ایستادگی را آموختم



تاسیس ۱۳۰۷
دانشگاه صنعتی خواجه نصیرالدین طوسی

بسمه تعالی
تأییدیه هیأت داوران

شماره:

تاریخ:

هیأت داوران پس از مطالعه پایان نامه و شرکت در جلسه دفاع از پایان نامه تهیه شده تحت عنوان :

روش های تشخیص بدافزار و توسعه شیوه های جدید در تشخیص بدافزار

توسط آقای محمود کرمی پور ، صحت و کفایت تحقیق انجام شده را برای اخذ درجه کارشناسی ارشد رشته مهندسی فناوری اطلاعات گرایش تجارت الکترونیک در تاریخ ۱۳۹۲/۰۶/۲۶ مورد تأیید قرار می دهند.

امضاء

جناب آقای دکتر شهریار محمدی

۱- استاد راهنمای اول

امضاء

سرکار خانم دکتر سمیه علیزاده

۲- ممتحن داخلی

امضاء

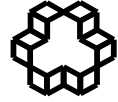
جناب آقای دکتر داود کریم زادگان مقدم

۳- ممتحن خارجی

امضاء

جناب دکتر عماد روغنیان

۴- معاونت آموزشی و
تحصیلات تکمیلی
دانشکده



تاسیس ۱۳۰۷
دانشگاه صنعتی خواجه نصیرالدین طوسی

بسمه تعالی
اظهارنامه دانشجو

شماره:

تاریخ:

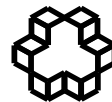
اینجانب محمود کرمی پور دانشجوی کارشناسی ارشد رشته مهندسی فناوری اطلاعات گرایش تجارت الکترونیک دانشکده مهندسی صنایع دانشگاه صنعتی خواجه نصیرالدین طوسی گواهی می‌نمایم که تحقیقات ارائه شده در پایان‌نامه با عنوان

روش‌های تشخیص بدافزار و توسعه شیوه‌ای جدید در تشخیص بدافزار

با راهنمایی استاد محترم جناب آقای دکتر شهریار محمدی، توسط شخص اینجانب انجام شده و صحت و اصل مطالب نگارش شده در این پایان‌نامه مورد تأیید می‌باشد، و در مورد استفاده از کار دیگر محققان به مرجع مورد استفاده اشاره شده است. بعلاوه گواهی می‌نمایم که مطالب مندرج در پایان‌نامه تا کنون برای دریافت هیچ نوع مدرک یا امتیازی توسط اینجانب یا فرد دیگری در هیچ جا ارائه نشده است و در تدوین متن پایان‌نامه چارچوب (فرمت) مصوب دانشگاه را بطور کامل رعایت کرده‌ام.

امضاء دانشجو:

تاریخ:



تاسیس ۱۳۰۷
دانشگاه صنعتی خواجه نصیرالدین طوسی

بسمه تعالی
حق طبع و نشر و مالکیت نتایج

شماره:

تاریخ:

۱- حق چاپ و تکثیر این پایان نامه متعلق به نویسنده آن می باشد. هرگونه کپی برداری بصورت کل پایان نامه یا بخشی از آن تنها با موافقت نویسنده یا کتابخانه دانشکده مهندسی صنایع دانشگاه صنعتی خواجه نصیرالدین طوسی مجاز می باشد.

ضمناً متن این صفحه نیز باید در نسخه تکثیر شده وجود داشته باشد.

۲- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی خواجه نصیرالدین طوسی می باشد و بدون اجازه کتبی دانشگاه به شخص ثالث قابل واگذاری نیست.

همچنین استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مراجع مجاز نمی باشد.

تقدیر و تشکر

سپاس بی‌کران پروردگار یکتا را که هستی‌مان بخشد و به طریق علم و دانش را، نمونه‌مان شد و به بهمنشینی رهروان علم و دانش مفتخرمان نمود و خوشه‌چینی از علم و معرفت را روزی‌مان ساخت.

از استاد محترم جناب آقای دکتر شهریار محمدی برای حمایت‌های بی‌دریغ ایشان کمال تشکر و قدردانی را دارم.

چکیده

بدافزارها و تهدیدات آنها یکی از بزرگترین چالش‌ها در امنیت و یکپارچگی اطلاعات و شبکه‌های کامپیوتری است. دو روش عمده در تشخیص بدافزار وجود دارد؛ روش مبتنی بر امضا و روش مبتنی بر ناهنجاری. نرم افزارهای آنتی ویروس کنونی از روش‌های مبتنی بر امضا استفاده می‌نمایند. روش‌های مبتنی امضا دقیق هستند اما توانایی تشخیص حملات روز نخست و ناشناخته را ندارند. ما می‌توانیم با استفاده از روش‌های مبتنی بر ناهنجاری و تطبیق آنها با روش‌های یادگیری نظارتی توانایی تشخیص بدافزارهای جدید را افزایش دهیم. دو نوع خصیصه وجود دارد که یادگیری نظارتی در تشخیص بدافزار از آنها استفاده می‌کند؛ خصیصه‌های ایستا و خصیصه‌های پویا. خصیصه‌های ایستا از ساختار فایل اجرایی دودویی و خصیصه‌های پویا با اجرای فایل در محیط واقعی یا شبیه‌سازی شده به دست می‌آیند. در این تحقیق روشی مبتنی بر یادگیری ماشین بر مبنای خصیصه‌های ترکیبی ایستا-پویا ارائه گردیده است. تعداد فراوانی از بدافزارهای جدید هر روزه بر روی اینترنت پخش می‌شوند و روشی جهت تشخیص این نمونه‌های جدید و مقاصد آنها نیاز است. تحلیل این بدافزارها توسط عامل انسانی زمان‌بر و پرخطاست و توسعه روش‌های خودکار جهت تحلیل حجم بالای نمونه الزامی است. روش ترکیبی پیشنهادی توانایی تحلیل حجم بالای نمونه اجرایی و تشخیص بدافزارهای ناشناخته را دارد و نیز با توجه به استفاده از خصیصه‌های ایستا و پویا در کنارهم از درصد هشدار اشتباه پایینی برخوردار است. استفاده از خصیصه‌های ایستا و پویا در کنار یکدیگر سبب افزایش درصد تشخیص و دقت تشخیص نسبت به استفاده جداگانه از آنها می‌شود.

کلمات کلیدی: تشخیص بدافزار، روش ترکیبی، روش ایستا-پویا، یادگیری ماشین، دسته‌بندی

فهرست مطالب

۱	فصل ۱: کلیات موضوع
۲	۱-۱ مقدمه
۴	۲-۱ بدافزار چیست؟
۶	۳-۱ انگیزه
۸	۴-۱ ساختار پایان نامه
۸	۵-۱ جمع بندی
۹	فصل ۲: مفاهیم مرتبط با تشخیص بدافزار
۱۰	۱-۲ مقدمه
۱۰	۲-۲ انواع بدافزار
۱۱	۱-۲-۲ ویروس‌ها
۱۲	۱-۲-۲-۱ ویروس‌های رونویسی کننده
۱۳	۲-۲-۲-۱ ویروس‌های مؤخر و مقدم
۱۳	۳-۲-۲-۱ ویروس‌های ساکن در حافظه
۱۴	۴-۲-۲-۱ ویروس‌های سکتوربوت
۱۵	۵-۲-۲-۱ ویروس‌های ماکرو
۱۶	۲-۲-۲ کرم‌ها
۱۷	۱-۲-۲-۲ کرم‌های ایمیل
۱۷	۳-۲-۲ دیگر حقه‌های ایمیلی
۱۸	۴-۲-۲ اسب‌های تراوا
۱۹	۵-۲-۲ در پشتی
۲۰	۱-۵-۲-۲ جاسوس‌افزار و آگهی‌افزار

۲۰DDOS حمله‌های بات‌نت و
۲۲ Rootkit
۲۳ ۸-۲-۲ سربار های بدافزار
۲۳ ۳-۲ فایل‌های اجرایی ویندوز
۲۵ ۴-۲ تحلیل بدافزار
۲۵ ۱-۴-۲ تحلیل ایستا
۲۶ ۲-۴-۲ مشکلات تحلیل ایستا
۲۶ ۳-۴-۲ تحلیل پویا
۲۷ ۱-۳-۴-۲ مشاهده فراخوانی توابع
۲۷ ۲-۳-۴-۲ مشاهده فراخوانی API ها
۲۷ ۳-۳-۴-۲ فراخوانی سیستمی
۲۸ ۴-۳-۴-۲ هوک کردن
۲۸ ۵-۳-۴-۲ تحلیل پارامتر توابع
۲۹ ۶-۳-۴-۲ بررسی جریان اطلاعات
۲۹ ۵-۲ روشهای ضد مهندسی معکوس
۲۹ ۱-۵-۲ ضد اشکالزدایی
۳۰ ۲-۵-۲ مبهم‌سازی
۳۱ ۳-۵-۲ پک کردن
۳۲ ۶-۲ جمع‌بندی
۳۳ فصل ۳: کارهای انجام شده
۳۴ ۱-۳ مقدمه
۳۵ ۲-۳ روش مبتنی بر ناهنجاری
۳۶ ۱-۲-۳ تشخیص مبتنی بر ناهنجاری پویا

۳۸	۲-۲-۳	تشخیص مبتنی بر ناهنجاری ایستا
۴۰	۳-۲-۳	تشخیص مبتنی بر ناهنجاری ترکیبی
۴۱	۳-۳	تشخیص مبتنی بر ویژگی
۴۱	۱-۳-۳	روش مبتنی بر ویژگی پویا
۴۲	۲-۳-۳	تشخیص مبتنی بر ویژگی ایستا
۴۳	۳-۳-۳	تشخیص مبتنی بر ویژگی ترکیبی
۴۳	۴-۳	روش مبتنی بر امضا
۴۵	۱-۴-۳	روش مبتنی بر امضای پویا
۴۵	۲-۴-۳	روش مبتنی بر امضای ایستا
۴۶	۳-۴-۳	روش مبتنی بر امضای ترکیبی
۴۷	۵-۳	جمع‌بندی
۴۸		فصل ۴: روش پیشنهادی
۴۹	۱-۴	مقدمه
۴۹	۲-۴	مجموعه داده
۵۱	۳-۴	تحلیل ترکیبی
۵۲	۱-۳-۴	فاز ایستا
۵۲	۱-۱-۳-۴	بررسی پک بودن نمونه‌ها
۵۴	۲-۱-۳-۴	از پک خارج کردن نمونه‌ها
۵۶	۳-۱-۳-۴	Disassemble کردن فایل‌های دودویی
۵۶	۴-۱-۳-۴	واکشی خصیصه‌های ایستا
۵۹	۵-۱-۳-۴	انتخاب ویژگی
۶۰	۲-۳-۴	فاز پویا
۶۰	۱-۲-۳-۴	هوک کردن APIها

۶۱ اشکال زدا ۲-۲-۳-۴
۶۴ بررسی و انتخاب خصیصه‌های پویا ۳-۲-۳-۴
۶۶ یادگیری ماشین ۴-۴
۶۷ درخت تصمیم ۱-۴-۴
۶۹ SVM بردار پشتیبان ۲-۴-۴
۷۰ شبکه بیز ۳-۴-۴
۷۲ K-NN ۴-۴-۴
۷۳ جمع‌بندی ۵-۴
۷۴ فصل ۵: پیاده‌سازی و ارزیابی
۷۵ مقدمه ۱-۵
۷۵ محیط پیاده‌سازی ۲-۵
۷۶ معیارهای ارزیابی ۳-۵
۷۷ معیار نرخ تشخیص ۱-۳-۵
۷۸ معیار نرخ هشدار اشتباه ۲-۳-۵
۷۸ معیار دقت ۳-۳-۵
۷۹ ROC نمودار ۴-۳-۵
۸۰ جمع‌بندی ۴-۵
۸۲ فصل ۶: نتیجه‌گیری و جمع‌بندی
۸۳ مقدمه ۱-۶
۸۳ نتیجه‌گیری ۲-۶
۸۵ محدودیت‌های تحقیق و کارهای آینده ۳-۶
۸۶ جمع‌بندی ۲-۶
۸۷ لیست مقالات ارائه شده

فهرست مراجع ۸۸

واژه نامه فارسی به انگلیسی ۹۲

واژه نامه انگلیسی به فارسی ۹۳

فهرست جدول‌ها

جدول ۱-۲	هفت نوع بدافزار سالمون	۱۱
جدول ۲-۲	بخش‌های استفاده‌شده در فایل PE	۲۴
جدول ۱-۳	مقایسه بین روش‌های مختلف تشخیص بدافزار	۴۷
جدول ۱-۴	آنتی‌ویروس‌های موجود در سایت virustotal	۵۰
جدول ۲-۴	callgraph	۵۷
جدول ۳-۴	function	۵۷
جدول ۴-۴	instruction	۵۷
جدول ۵-۴	مدل n-gram با طول ۱ تا ۴	۵۸
جدول ۶-۴	فهرست DLL هایی که توابع آنها هوک شده‌اند	۶۲
جدول ۷-۴	Dynamic_log	۶۵
جدول ۸-۴	قسمتی از بردار خصیصه پویا	۶۵
جدول ۱-۵	ارتباط معنایی FN,FP,TN,TP	۷۷
جدول ۲-۵	نتایج ارزیابی با استفاده از معیار درصد تشخیص	۷۷
جدول ۳-۵	نتایج ارزیابی با استفاده از معیار درصد هشدار اشتباه	۷۸
جدول ۴-۵	نتایج ارزیابی با استفاده از معیار دقت	۷۹
جدول ۵-۵	مقایسه روش‌های مختلف دسته‌بندی با استفاده از سطح زیر نمودار ROC	۸۱
جدول ۱-۶	مقایسه روش پیشنهادی با کارهای مشابه	۸۵

فهرست شکل‌ها

- شکل ۱-۱ گزارش امنیتی مکافی در سه ماه اول ۲۰۱۳ ۷
- شکل ۱-۲ طریقه آلوده شدن فایل‌ها توسط ویروس ۱۲
- شکل ۲-۲ یک دیاگرام از فرآیندی که اسپرها از زامبی‌ها استفاده می‌کنند ۲۲
- شکل ۳-۲ ساختار یک فایل PE ۲۴
- شکل ۴-۲ مبهم‌سازی با جاگذاری عملگرهای و کدهای مرده ۳۱
- شکل ۵-۲ ساختار یک فایل PE قبل و بعد از پک شدن ۳۲
- شکل ۱-۳ انواع مختلف روش‌های تشخیص بدافزار ۳۵
- شکل ۲-۳ تعیین ویژگی‌های رفتاری در روش مبتنی بر ناهنجاری ۳۶
- شکل ۳-۳ روش مبتنی بر امضا ۴۴
- شکل ۱-۴ انواع بدافزارهای موجود در پایگاه داده ۵۰
- شکل ۲-۴ سایت Virustotal ۵۱
- شکل ۳-۴ روش ترکیبی پیشنهادی ۵۳
- شکل ۴-۴ شکل ظاهری برنامه PEID ۵۴
- شکل ۵-۴ نتایج حاصل از تشخیص پک‌کننده‌های مختلف ۵۵
- شکل ۶-۴ تعداد خصیصه‌ها به ازای مقدار n ۵۹
- شکل ۷-۴ رجیستر اشکال‌زدا در پردازنده ۶۲
- شکل ۸-۴ نحوه اجرای یک نمونه در اشکال‌زدا ۶۴
- شکل ۹-۴ نمونه‌ای از یک درخت تصمیم در تشخیص بدافزار ۶۸
- شکل ۱۰-۴ الگوریتم درخت تصمیم‌گیری ۶۹
- شکل ۱۱-۴ مثالی از دسته‌بندی با استفاده از ماشین بردار پشتیبان ۷۰
- شکل ۱۲-۴ مثالی از الگوریتم دسته‌بندی KNN ۷۳
- شکل ۱-۵ نمودار ROC برای دو روش دسته‌بندی ۸۰

شکل ۲-۵ مقایسه بین روش‌های مختلف دسته‌بندی با معیار دقت ۸۱

فصل ۱

کلیات موضوع

۱-۱ مقدمه

اینترنت به عنوان بخش اساسی از زندگی روزانه مردم تبدیل شده است زیرا که خدمات مختلف اکنون توسط این شبکه ارتباطی عرضه می‌شوند. اینترنت از یک شبکه ارتباطی به شبکه‌ای از منابع ارائه‌دهنده اطلاعات، شکل جدیدی از تعاملات اجتماعی و بازاری برای فروش محصولات و ارائه سرویس‌ها و خدمات متنوع تبدیل گردیده است. بانکداری اینترنتی و تبلیغات از جمله کاربردهای تجاری اینترنت هستند. گسترش فناوری اطلاعات و تجارت الکترونیک از یک سو فرصت‌های جدید برای محیط‌های کسب‌وکار فراهم نموده است؛ از دیگر سو مسائلی همانند امنیت و محرمانگی اطلاعات در دنیای مجازی و استفاده از اینترنت به عنوان ابزاری جهت گسترش انواع بدافزار و پیدایش شیوه‌های جدید کلاهبرداری مشکلاتی هستند که در دنیای جدید مجازی بروز پیدا کرده‌اند. همانند دنیای واقعی در دنیای مجازی نیز افرادی هستند که تمایلات مجرمانه و شوم دارند. بدافزارها (نرم‌افزارهایی که اهداف بد دارند) به این افراد کمک می‌نمایند که به مقاصد مجرمانه خود برسند. شناسایی بدافزارها نخستین مرحله در پیشگیری سیستم‌های کامپیوتری در مواجهه با از دست دادن بالقوه اطلاعات و به خطر افتادن آنهاست. راه‌های مختلفی برای شناسایی بدافزارها وجود دارد. یک شناساگر بدافزار تلاش می‌کند عملکردها و برنامه‌های نادرست را بیابد. شناسایی بدافزارها مهم‌ترین مرحله در پیشگیری سیستم‌های کامپیوتری در مقابل آلوده شدن، محافظت در برابر از دست دادن اطلاعات و به خطر افتادن آنهاست. روش‌های بسیاری برای شناسایی بدافزارها وجود دارد تا از اجرای آنها جلوگیری شود.

اصلی‌ترین روش‌ها برای شناسایی بدافزارها بر اساس ناهنجاری^۱، امضا^۲، ویژگی‌ها^۳ است. این روش‌ها محدود به شناسایی بدافزارها نمی‌شوند، بلکه برای شناسایی نفوذ در شبکه‌های کامپیوتری و دیگر زمینه‌ها نیز کاربرد دارند.

بیشتر نرم‌افزارهای آنتی‌ویروس، با استفاده از تکنیک امضا و ویروسی عمل می‌کنند که در نتیجه آن به طور پیوسته با مجموعه‌ای از امضاها روبرو هستند. این روش یک روش انفعالی است، به محض آنکه یک امضا برای یک اکسپلویت^۴ ایجاد می‌شود، این اکسپلویت توسط نرم‌افزارهای آنتی‌ویروس به راحتی شناسایی خواهد شد.

روش‌هایی که بر اساس ناهنجاری به وجود آمده در سیستم طراحی می‌شوند بر اساس الگوهای در داده‌ها شناسایی می‌شود که متفاوت از رفتار مورد انتظار است. ناهنجاری‌ها در داده‌ها می‌تواند متناسب با تغییرات مهم در سیستم باشد. اگر یک کامپیوتر با الگوی ترافیکی اینترنتی ناهنجار مواجه شود این امر نشان می‌دهد که مورد حمله واقع شده و اطلاعات در حال فاش شدن است (چاندولا و همکاران^۵، ۲۰۰۹). روند شناسایی یک سیستم بدافزار نشان می‌دهد که چگونه اطلاعات در مورد برنامه جمع‌آوری شده است و روش شناسایی نشان می‌دهد که چگونه این اطلاعات مورد استفاده قرار می‌گیرند. رویکردهای اصلی برای تحلیل و آنالیز بدافزارها شامل آنالیز ایستا، آنالیز پویا و ترکیبی از این دو است.

آنالیز ایستا از اطلاعات ساختاری و ترکیبی از یک برنامه استفاده می‌کند تا ناهنجاری‌های آن را مشخص کند. قبل از اینکه یک برنامه اجرا شود، اطلاعات ایستا قابل اجرا شناسایی می‌شوند مثل داده‌های سرآیند^۶ و بیت‌های پشت سرهم. این کار به این منظور انجام می‌شود که اگر ناهنجاری وجود دارد مشخص شود. آنالیز ایستا قادر است همه مسیرهای موجود در یک برنامه که ممکن است در زمان اجرا قرار گیرد را آنالیز کند، با این وجود معمولاً دربرگیرنده همه حالات ممکن برنامه نمی‌شود چون داده‌های

¹ Anomaly-based

² Signature-based

³ Specification-based

⁴ Exploit

⁵ Chandola

⁶ Header data