فصلنامه علمی-ترویجی پدافند غیرعامل

سال نهم، شماره ۲، تابستان ۱۳۹۷، (پیاپی ۳۴): صص ۱۰۱-۹۵

شناسایی بدافزارها با استفاده از تصویرسازی

هادی رنجی[۱]، سعید پارسا[۲*]

تاریخ دریافت: ۱۳۹۶/۰۳/۱۶
تاریخ پذیرش: ۱۳۹۶/۱۰/۱۹

چکیده

در این مقاله استفاده از راه‌کاری مبتنی بر پردازش تصویر جهت شناسایی بدافزارهای چندریختی مطرح شده است. بدافزارنویسان با ایجاد نسخه‌های مختلف از یک بدافزار در عمل راه‌کار تشخیص ایستای بدافزار براساس امضاء را با مشکل مواجه کرده‌اند. بررسی‌های ما بر روی شباهت تصویرهای تولیدشده از فایل بدافزارهای چندریختی و در جهت ایجاد امکان تفکیک بین بدافزارها و فایل‌های معمولی انجام گرفته است. با توجه به نتایج حاصل‌شده، امکان شناسایی بدافزارها با استفاده از تصاویر آن‌ها میسر شده است. با درنظرگرفتن کد دودویی در قالب یک تصویر، می‌توان ویژگی‌های زیادی جهت تعیین میزان تشابه بین نسخه‌های محتلف یک بدافزار استخراج نمود. براساس این ویژگی‌ها در عمل نشان داده شد که با دقت بی‌سابقه‌ای می‌توان بدافزارهای چندریختی را شناسایی نمود. کد دودویی اغلب بدافزارها به صورت بسته‌بندی شده یا در اصطلاح پک‌شده می‌باشد. می‌توان با استفاده از روش پیشنهادی خود تشابه بین فایل‌هایی که توسط یک ابزار بسته‌بندی یا در اصطلاح پک شده‌اند را استخراج نمود.

کلید واژه‌ها: بدافزار، ویروس، تصویرسازی، چندریختی، پردازش تصویر.

۱- دانشجوی کارشناسی ارشد دانشگاه آزاد اسلامی شبستر
۲- دانشیار دانشگاه علم و صنعت ایران- parsa@iust.ac.ir - نویسنده مسئول

۱- مقدمه

بدافزارها یا کدهای بدخواه یا در اصطلاح عمومی‌تر، ویروس‌های رایانه‌ای برنامه‌هایی با عملکرد مخرب هستند که سازندگان آن‌ها معمولاً اهداف خرابکارانه و سودجویانه‌ای را دنبال می‌کنند. این هدف‌ها می‌تواند از خرابکاری به جهت تفریح تا عملیات جاسوسی منسجم و هدف‌دار متغیر باشد.

انواع مختلفی از بدافزارهای رایانه‌ای براساس نوع عملکرد خود وجود دارند. از این جمله می‌توان ویروس‌ها، کرم‌ها، اسب‌های تروا و باکتری‌ها را نام برد. این تقسیم‌بندی‌ها براساس مواردی مثل راه‌برد حمله، چگونگی تکثیر و روش انتقال اعمال می‌شود [۱].

شناسایی و مقابله با عملکرد بدافزارها یکی از دغدغه‌های جامعه اطلاعاتی ماست. در این راستا، شرکت‌ها و نهادهای مسئول تامین امنیت فضای مجازی به تولید و عرضه محصولات ضد بدافزار روی آورده‌اند.

با پیشرفت روش‌ها و فنون برنامه نویسی، بدافزارهای تولیدشده نیز از ساختارهای پیچیده‌تر و قدرتمندتری نسبت به گذشته برخوردارند و تبعاً برنامه‌های ضد بدافزاری نیز باید همواره خود را برای مقابله با جدیدترین روش‌های خرابکارانه آماده کنند. یکی از روش‌های نوین، استفاده از ساختارهای چندریختی جهت تغییر در ساختار بدافزارها و بی‌اثر کردن روش‌های مبتنی بر امضاء است. بر این اساس، روش‌های متداول شناسایی در مواجهه با این نوع بدافزارها معمولاً دچار شکست می‌شوند. به همین علت، نیازمند روش‌های موثر برای مقابله با این نوع بدافزارها هستیم [۲].

روش پیشنهادی ما براساس تصویرسازی بدافزار و استفاده از روش‌های پردازش تصویر در جهت شناسایی آن است. در این روش، ابتدا یک تصویر از بدافزار تولید می‌شود، سپس، خصوصیاتی از آن تصویر جهت مقایسه استخراج می‌گردد. با استفاده از محاسبه میزان شباهت موجود در بدافزارها و مقایسه آن با نرم‌افزارهای معمولی می‌توان بین بدافزار و نرم افزارهای غیرمخرب تفکیکی در حدود ۹۸٪ ایجاد نمود.

روش پیشنهادی نیازی به عملیات دیس اسمبلی[1] ندارد. در نتیجه روش‌های آنتی دیس اسمبلی[2] تاثیری در عملکرد آن ندارند. همچنین، به‌علت استفاده از ساختار فایل جهت تولید تصویر روش‌های رمزگذاری و فشرده‌سازی مانند پک‌کردن نیز نمی‌تواند در نتایج حاصل‌شده خللی ایجاد نماید.

۲- روش‌های پیشین

روش‌ها و الگوریتم‌های متعددی برای شناسایی بدافزارها ارائه‌شده است که هرکدام مبنای خاص خود را دارد. این روش‌ها یا به‌دنبال ایجاد شناسه‌ای منحصربه‌فرد برای هر بدافزار هستند و یا برای هر دسته از بدافزارها نشانه‌ای را درنظر می‌گیرند. این روش‌ها برمبنای عملکرد خود به دو دسته ایستا و پویا تقسیم می‌شوند.

۲-۱- ایستا

روش‌های شناسایی ایستا مبتنی بر بررسی محتوی دودوئی بد افزار است. این روش‌ها که به شکل سنتی به نام روش مبتنی بر امضاء مشهور هستند و یکی از روش‌های اصلی شناسایی بدافزارها به شمار می‌آیند.

در روش امضاء، معمولاً ضدبدافزار دارای یک بانک اطلاعاتی از امضاءهای تولیدشده از بدافزارهای شناخته‌شده است. این امضاء معمولاً یک رشته تولیدشده توسط الگوریتم‌های رمزنگاری مثل MD5 از محتوای باینری بدافزارها است. درصورتی‌که نرم‌افزار ضدویروس الگوی این امضاء را در فایلی ببیند آن فایل را به‌عنوان یک بدافزار شناسایی می‌کند. این روش با این‌که معمولاً درصد خطای پایینی دارد ولی با ظهور ویروس‌های چندریختی موفقیت زیادی در شناسایی ندارد [۳].

روش‌های بهینه‌تری مانند استفاده از محتوی فایل دودویی در مبنای شانزده به عنوان امضای بدافزار نیز با ترکیب روش‌های عبارات با قاعده[3] در برخی نرم افزارهای ضدویروس به‌کار برده می‌شود [۴].

از دیگر روش‌ها، می‌توان به بررسی توابع API موجود در جدول واردات[4] بدافزار به‌عنوان یکی دیگر از مولفه‌های شناسایی، اشاره نمود [۵].

۲-۲- پویا

روش‌های پویای شناسایی، مبتنی بر عملکرد بدافزار می‌باشند. این روش‌ها مختص بدافزارهای شناخته‌شده نیستند. البته، با این‌که درصد خطای بالاتری نسبت به روش‌های ایستا دارند ولی قابلیت شناسایی بدافزارهای جدید و مواردی که ساختار آن‌ها پیچیدگی زیادی برای اعمال روش‌های ایستا دارند از مزیت‌های این روش‌ها است.

1- Dissasembly
2- Anti Dissasembly
3- Regular Expression
4 - Import Table

روش‌های پویای شناسایی به روش‌های اکتشافی[1] نیز معروف هستند. معمولاً، در این روش‌ها یک مدل از رفتار بدخواه تولید می‌شود و رفتارهایی که متقارن با آن مدل باشند، به‌عنوان رفتاری مشکوک شناسایی می‌شوند [۶].

ارتباط نرم‌افزارها با امکانات سامانه عامل از طریق توابع واسط سامانه‌ای به‌نام API انجام می‌شود. لذا، جهت مدل‌سازی رفتار یک بدافزار، توالی و یا گراف وابستگی بین فراخوانی‌های سامانه‌ای درنظر گرفته می‌شود. البته، در مورد بدافزارها محدوده خاصی از فراخوانی‌های سامانه‌ای درنظر گرفته می‌شود. براساس توالی و یا وابستگی داده‌ای در بین این نوع فراخوانی‌ها، عملیات مضر و خطرناک تشخیص داده می‌شود.

۳- الگوریتم پیشنهادی

برای بررسی عملکرد روش پیشنهادی، تعداد ۱۰۰۰ بدافزار و به همان تعداد فایل اجرایی مربوط به نرم‌افزارهای دیگر براساس الگوریتم ذیل مورد بررسی قرار گرفت که در ادامه هریک از بخش‌های این الگوریتم را مورد تحلیل و بررسی قرار می‌دهیم.

File_Visualizing *(Mlaware, Benign)*

{ Generate_Image (Malware, Benign);

Features _Extraction (Image File);

Features _Selection(Features);

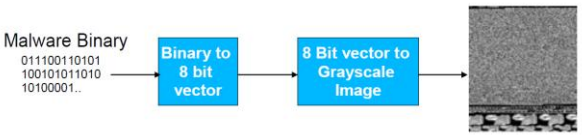
Calculate_Similarity (ReducedFeatures);

Classification_Samples(Similarity Value)}

۳-۱- تصویرسازی

تصویرسازی به معنی ایجاد و مطالعه نمایش بصری داده در شاخه‌های مختلف علوم رایانه‌ای کاربردهای فراوانی دارد و به‌علت انتقال بصری و سریع حجم بالای اطلاعات به‌شکل گسترده‌ای مورد استفاده قرار می‌گیرد.

در مرحله نخست همانند شکل (۱) محتوای دودویی یک فایل به شکل یک بردار ۸ بیتی خوانده می‌شود. سپس، جهت ایجاد امکان تصویرسازی در یک آرایه دوبعدی ذخیره می‌شود و در ادامه از این آرایه دوبعدی جهت تولید تصویر خاکستری شامل دامنه نقاط با رنگ سیاه کامل تا ۲۵۵ و با رنگ سفید کامل استفاده می‌شود. عرض تصویر تولیدشده ثابت گردیده است ولی ارتفاع تصاویر بستگی به اندازه فایل دارد. به‌عبارتی، تمامی تصاویر تولیدشده دارای عرض یکسان ولی به نسبت اندازه فایل دارای ارتفاع متفاوتی خواهند بود.



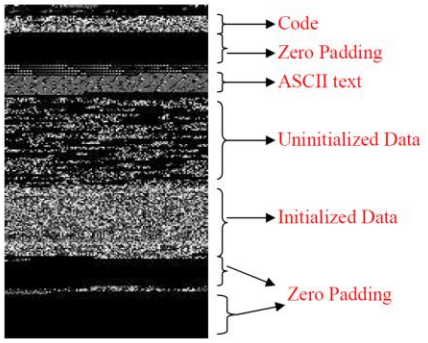
شکل (۱): مراحل تولید تصویر از فایل باینری

تصویرهای تولیدشده از محتوی فایل‌ها همانند شکل (۲) به راحتی امکان تعیین محل بخش بندی‌های منطقی فایل‌های اجرائی PE[2] را به ما می‌دهد.

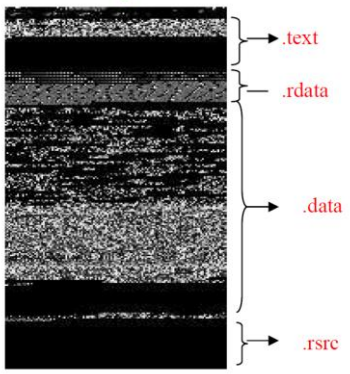
بخش text. محتوی کدهای اجرایی است. همان‌طورکه در شکل (۳) مشخص است قسمت اول این بخش که محتوی کد است به صورت بافت دانه‌دانه مشخص شده است و بقیه قسمت‌های این بخش که به‌صورت سیاه رنگ مشاهده می‌شود، به معنی فضای خالی در انتهای این بخش است.

بخش data. هم شامل داده‌های بدون مقدار اولیه می‌باشد که به صورت سیاه‌رنگ مشخص شده است. این بخش نیز شامل داده‌های اولیه است که به‌صورت دانه‌دانه مشخص شده اند.

بخش rsrc. نیز محتوی تمام منابع فایل ازجمله کلیه شمایل‌ها و تصویرهای استفاده‌شده داخل برنامه است.



شکل (۲): تعیین محل انواع داده های فایل PE



شکل (۳): تعیین محل بخش‌های فایل PE

1 - Heuristic

2- Portable Executable

۳-۲- دسته‌بندی بدافزارها

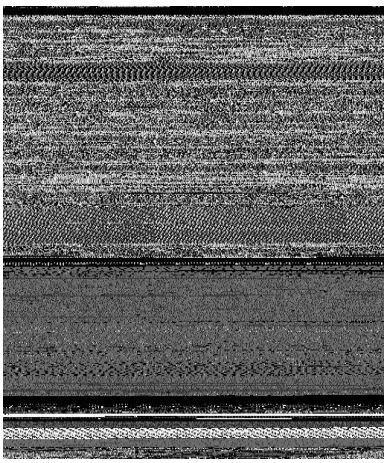
ظهور ویروس‌های چندریختی یا در اصطلاح پلی‌مرفیک که قابلیت تغییر ساختار خود در نسل‌های مختلف را دارا می‌باشند، روش‌های مبتنی بر امضاء را ناکارآمد کرده است. بدافزارنویس‌ها بخش‌های کوچکی از متن کد خود را جهت تولید بدافزارهای جدید تغییر می‌دهند. در عمل مشاهده می‌شود که با تبدیل فایل دودویی بدافزار به تصویر، در عین حفظ ساختار کلی، می‌توان هرگونه تغییری را در تصویر مشاهده نمود. با مقایسه تصاویر تولیدشده از نسخه‌های مختلف یک بدافزار چندریختی، می‌توان به این نتیجه رسید که نسخه‌های مختلف یک بدافزار دارای ساختاری مشابه با یکدیگر، اما متمایز از نسخه‌های مختلف سایر بدافزارها هستند. بر این اساس می‌توان با ایجاد تصویر برای کد اجرایی بدافزارها و محاسبه تشابه بین تصاویر گردآوری‌شده، نسخه‌های مختلف یک بدافزار را شناسایی نمود.

بررسی تصویرهای تولیدشده از بدافزارهای بسته‌بندی یا در اصطلاح پُک‌شده نشان می‌دهد استفاده از بسته‌بندکننده‌های یکسان موجب تشابه تصویر در نمونه‌های مختلف می‌گردد. همچنین، درصورتی که نرخ فشرده‌سازی پایین باشد، تصویر بدافزارهای بسته‌بندی‌شده بسیار شبیه به تصاویر همان بدافزارها درحالتی‌که بسته‌بندی نشده‌اند، خواهد بود [۷].

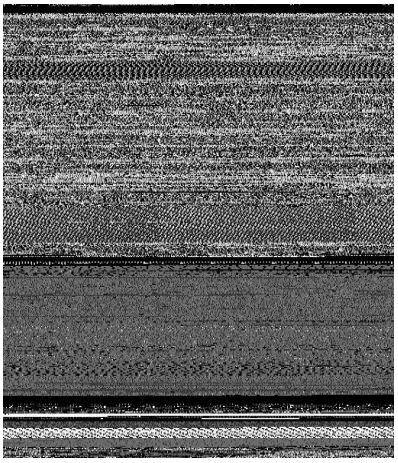
شکل (۴) نمایانگر تصویر تولیدشده از نسخه اولیه بدافزار Dontovo و شکل (۵) تصویر نسخه دوم این بدافزار را نشان می‌دهد. در شکل (۶)، تفاوت تصویرهای این دو نسخه، نمایان شده است. علی‌رغم این‌که شکل (۶) تفاوت پیکسل به پیکسل دو تصویر را نمایش داده است، تفاوت کمی بین دو نسل مشاهده می‌گردد. با این حال، برای مقایسه و محاسبه تفاوت دو تصویر از روش‌های پیشرفته‌تر پردازش تصویر استفاده می‌گردد.

با یک منطق ساده می‌توان این نتیجه را گرفت که دو تصویر زمانی مشابه هستند که اندازه آن‌ها با هم برابر بوده و پیکسل آن دو نظیر به نظیر با هم برابر باشد. این منطق برای مقایسه تصاویری که در وضعیت ایده‌آل نیستند نیازمند بازنگری است. در منطق منعطف‌تر برای یافتن تصاویر مشابه ما باید الگوهایی را بین دو تصویر بیابیم که بیش‌ترین شباهت را با یکدیگر دارند. طبیعتاً این منطق همواره به پاسخ صحیح نمی‌رسد.

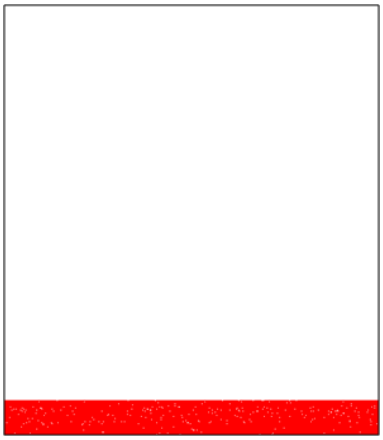
پیشرفت فن‌آوری و گسترش حیرت‌آور اینترنت در خلال سال‌های اخیر، مقوله ذخیره‌سازی و بازیابی اطلاعات و به‌ویژه تصویرها را به یکی از فعال‌ترین حیطه‌ها در توسعه نظام‌های چندرسانه‌ای مبدل کرده است. بازیابی تصاویر مبتنی بر محتوی راه‌کاری است که امروزه جهت محاسبه میزان تشابه تصویرهای مختلف به‌کار گرفته می‌شود. این بازیابی براساس معیارهایی مانند رنگ، طرح‌بندی، بافت، شکل و غیره انجام می‌پذیرد.



شکل (۴): تصویرسازی ویروس Dontovo.A



شکل (۵): تصویرسازی ویروس Dontovo.B



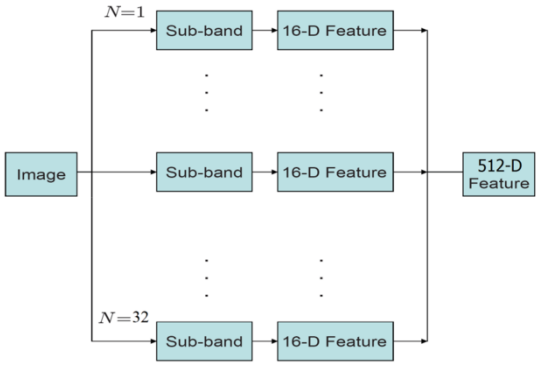
شکل (۶): تفاوت تصویر دو نسل از ویروس Dontovo

۳-۳- استخراج ویژگی

بافت تصاویر شامل مجموعه‌هایی از پیکسل‌ها است که در بخش‌های مختلف دارای ترکیب‌های متنوعی می‌باشند. این ترکیب پیکسل‌ها می‌تواند به‌عنوان شاخصی برای تحلیل تصاویر مورد استفاده قرار گیرد. ویژگی‌های استخراج‌شده مبتنی‌بر بافت تصاویر به شکل گسترده‌ای در نرم‌افزارهای مختلف مورد استفاده قرار می‌گیرد از آن جمله می‌توان به نرم افزا های پردازش تصاویر پزشکی اشاره کرد. ما از ویژگی‌های مستخرج از بافت تصاویر بدافزارها جهت دسته‌بندی آن‌ها استفاده کرده‌ایم.

جهت استخراج ویژگی‌های بافت تصویر از روش توصیف‌گر GIST[1] بهره برده‌ایم. این روش به‌شکل گسترده‌ای در سامانه‌های بازشناسی تصویر مانند دسته‌بندی صحنه‌ها و تشخیص اشیاء به‌کار برده می‌شود [۸].

شکل (۷) مراحل اعمال توصیف‌گر GIST و شکل (۸) نمونه تصویر تولیدشده را با استفاده از مراحل اعمال‌شده به‌ترتیب زیر نمایش می‌دهد:

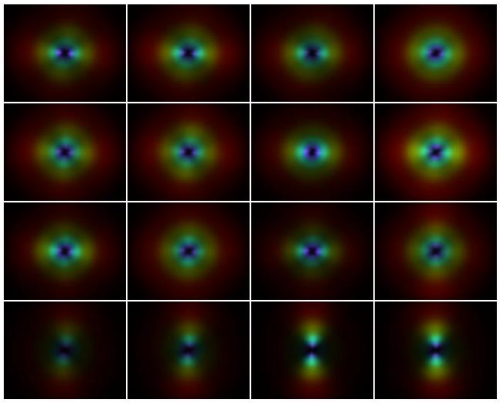


شکل (۷): نمودار محاسبه خصوصیات تصویر

۱- تصویر را با یک فیلتر گابور در ۴ مقیاس و در ۸ جهت همگشت[2] می‌کند و ۳۲ نقشه ویژگی با سایز برابر تصویر ورودی تولید می‌کند.
۲- هر نقشه ویژگی به ۱۶ منطقه تقسیم می‌شود و سپس میانگین مقدار ویژگی در هر منطقه را محاسبه می‌کند.
۳- مقدار میانگین در ۱۶ منطقه در تمام ۳۲ نقشه ویژگی را محاسبه می‌کند که معادل ۵۱۲ ویژگی می‌باشد.

نکته : در برخی کاربردها به‌جای ۳۲ از ۲۰ نقشه ویژگی استفاده شده که موجب تولید ۳۲۰ ویژگی از هر تصویر است.

1- Global Image Similarity Technic
2- Convolution



شکل (۸): توصیف‌گر GIST اعمال‌شده بر روی تصویر

توصیف‌گر GIST با استفاده از فیلترکردن تصویر I(x,y) توسط فیلترهای گابور تولید می‌گردد. این فیلتر مقدار t(x,y) را بر می‌گرداند:

$$t(x,y)=\frac{1}{(2\pi\sigma_x\sigma_y)}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma x^2}+\frac{y^2}{\sigma y^2}\right)+2\pi jWx\right] \qquad (۱)$$

فرمول (۱): محاسبه فیلتر گابور

تبدیل فوریه آن به‌صورت T(u,v) تعریف می‌شود:

$$T(u,v)=\exp[-\frac{1}{2}\left(\frac{(u-W)^2}{(\sigma_u)^2}+\frac{v^2}{(\sigma_v)^2}\right)] \qquad (۲)$$

فرمول (۲): تبدیل فوریه فیلتر گابور

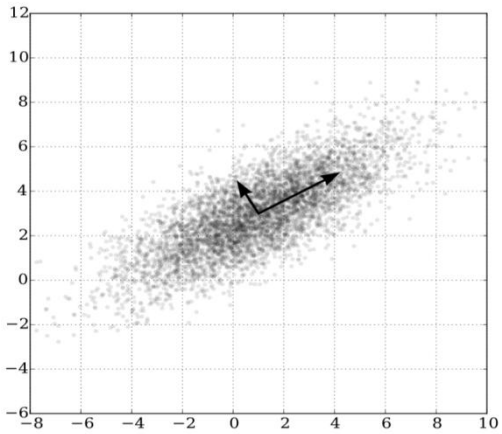
در روابط فوق، $\sigma_u=1/2\pi\sigma_x$ و $\sigma_v=1/2\pi\sigma_y$ است و σ_x و σ_y انحراف استاندارد توابع گاوسی در جهت x و y می‌باشد.

۴-۳- انتخاب ویژگی

ویژگی‌های تولیدشده توسط توصیف‌گر GIST برای هر تصویر شامل ۵۱۲ خصوصیت است که در حجم بالای تصاویر موجب کاهش کارایی و سرعت پردازش اطلاعات می‌شود. بر این اساس، با استفاده از راه‌کارهای کاهش ابعاد می‌توان فضای حالت موجود را به یک فضای کوچکتر نگاشت کرد. در واقع با ترکیب مقادیر ویژگی‌های موجود، تعداد کمتری ویژگی به‌وجود می‌آید. به‌طوری‌که این ویژگی‌ها دارای تمامی یا بخش قابل توجهی از اطلاعات موجود در ویژگی‌های اولیه می‌باشند.

راه‌کار تجزیه و تحلیل اجزای اصلی یا PCA[3] بهترین روش برای کاهش ابعاد داده به صورت خطی می‌باشد. یعنی با حذف ضرایب کم اهمیت به‌دست‌آمده از این تبدیل، اطلاعات از دست رفته نسبت به روش‌های دیگر کمتر است. در این روش، محورهای مختصات جدیدی برای داده‌ها تعریف شده و داده‌ها براساس این محورهای مختصات جدید بیان می‌شوند که در شکل (۹) نشان داده شده است [۹].

3- Principal Component Analysis



شکل (۹): محورهای مختصات جدید روی داده‌ها

مسئله اصلی در تبدیل فضای حالت n بعدی به k، تعیین مقدار مناسب ابعاد فضای حالت جدید است. به عبارت دیگر، درصورتی‌که در کاهش ابعاد صد در صد واریانس داده‌ها حفظ شده باشد در فضای حالت جدید هیچ اطلاعاتی از دست نرفته است. برای این منظور، با محاسبه واریانس حفظ‌شده داده‌ها در تبدیل فضای حالت می‌توان مقدار مناسب ابعاد را تعیین کرد.

$$POV = 1 - \frac{\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)} - x^{(i)}approx\|^2}{\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}\|^2} * 100 \qquad (۳)$$

فرمول (۳): محاسبه درصد واریانس حفظ شده داده ها

محاسبه درصد واریانس حفظ‌شده به ازای مقادیر مختلف k در تکنیک PCA بر روی دیتاست بدافزارها نشان‌دهنده کاهش ویژگی‌های استخراج‌شده به کمتر نصف مقدار اولیه با حفظ ۹۸ درصدی واریانس داده‌ها می‌باشد.

جدول (۱): درصد واریانس حفظ‌شده

#	<K	POV
1	501	100
2	203	99
3	117	98
4	67	97
5	38	96

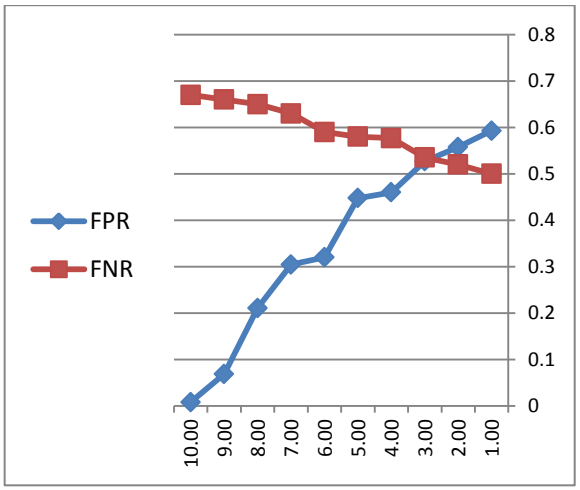
۳-۵- محاسبه تشابه

از فاصله اقلیدسی جهت محاسبه تشابه در نمونه تصاویر بر روی ویژگی‌های استخراج‌شده توسط توصیف‌گر GIST استفاده می‌گردد. پس از محاسبه فاصله اقلیدسی بر روی تمامی بردارهای ویژگی موجود دیتاست از الگوریتم جستجوی نزدیک‌ترین همسایه یا K-NN[1] جهت دسته بندی نمونه مورد بررسی استفاده میگردد.[۱۰]

1- K-Nearest Neighbors

۳-۶- دسته‌بندی

با محاسبه فاصله بین ویژگی‌های مستخرج فایل تحت بررسی و ویژگی‌های موجود در مجموعه داده‌های گردآوری‌شده به ازای تعداد K تعیین‌شده مربوط به بدافزارها امکان اعلام بیشترین مشابهت به نمونه‌های موجود در مجموعه داده وجود خواهد داشت. همچنین، جهت تعیین مناسب‌ترین آستانه برای شناسایی بدافزارها از فایل‌های معمولی، روش فوق بر روی دو مجموعه داده[2] مختلف بدافزار و فایل‌های معمولی اعمال و براساس محاسبه نرخ مثبت کاذب و نرخ منفی کاذب آستانه مناسب برای تفکیک بدافزارها تعیین گردید.



شکل (۱۰): تعیین آستانه دسته‌بندی

۳-۷- ارزیابی نتایج

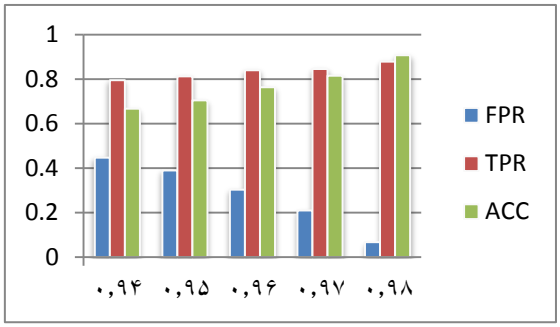
با توجه به نتایج به‌دست‌آمده از روش پیشنهادی، معیارهای ارزیابی همچون دقت را مورد ارزیابی قرار می‌دهیم. روش شناسایی براساس تشابه تصاویر بر روی مجموعه‌ای حاوی ۲۳۹۸ نمونه که ۱۱۲۶ عدد نمونه بدافزار شامل چند دسته از نسخه‌های مختلف بدافزارهای چندریختی و ۱۲۷۲ عدد نمونه فایل معمولی مورد بررسی قرار گرفت و نتایج ذیل حاصل گردید.

معیار آستانه بالا در محاسبه تشابه تصاویر فایل‌های بدافزارهای نسل‌های مختلف نسبت معکوسی با نرخ مثبت صحیح دارد. بنابراین، مقدار ۰/۳ به‌عنوان آستانه مناسب جهت دسته‌بندی نمونه‌ها تعیین گردید. روش پیشنهادی موفق به شناسایی صحیح تعداد ۹۹۰ عدد فایل به‌عنوان بدافزار یعنی دقتی برابر با ۱۱۲۶/۹۹۰=۸۷/۹۲ ٪ و تعداد ۱۱۸۵ عدد به عنوان فایل معمولی یعنی دقتی برابر با ۱۲۷۲/۱۱۸۵=۹۳/۱۶ ٪ گردید.

2- Dataset

جدول (۲): نتایج ارزیابی

#	K	Acc%	TPR	FPR	AUC
1	1	99.04	0.98	0.00	0.035356744
2	2	94.11	0.92	0.04	0.02377079
3	3	97.28	0.96	0.01	0.408445297
4	4	89.48	0.89	0.46	0.223217904
5	5	81.76	0.84	0.20	0.096461775



شکل (۱۱): نمودار نتایج ارزیابی

۴- نتیجه‌گیری

در این مقاله راه‌کاری مبتنی بر پردازش تصویر جهت شناسایی ویروس‌های چندریختی ارائه شده است. در حقیقت می‌توان این‌گونه نتیجه‌گیری کرد که با استفاده از راه‌کارهای پردازش تصویر نسبت به شناسایی بدافزارهای چندریختی به شکل موثری اقدام نمود.

راه‌کار پیشنهادی مبتنی بر تصویرسازی بدافزارها، استخراج ویژگی و دسته‌بندی آن‌ها است. از مزایای این روش می‌توان به سرعت بالا، عدم آلودگی محیط تجزیه و تحلیل به‌علت عدم اجرای بدافزار، عدم تاثیر روش‌های ضد دیس اسمبلی به‌علت عدم اجرای عملیات دیس اسمبلی، دریافت اطلاعات بیشتر از ساختار بدافزار و راه‌کارهای نفوذ بیشتر با استفاده از پردازش تصویر اشاره کرد. از معایب این روش عدم امکان مقابله با حملات روز صفر به‌علت تحلیل روی بدافزارهای موجود و عدم اطلاع از رفتار و عملکرد بدافزار است.

به لحاظ کمی نیز روش پیشنهادی با داشتن دقتی برابر با ۹۹/۰۴ ٪ در مقایسه با روش‌های دیگر شناسایی بدافزارها دارای نتایج قابل قبولی می‌باشد.

۵- مراجع

1. J. Aycock, “Computer Viruses and Malware,” US, Springer 2007.
2. M. Eskandari and S. Hashemi, “Metamorphic Malware Detection using Control Flow Graph Mining,” IJCSNS International Journal of Computer Science and Network Security, vol. 11, no. 12, December 2011.
3. K. Griffin, S. Schneider, X. Hu, and T.-C. Chiueh, “Automatic Generation of String Signatures for Malware Detection,” Symantec Research Laboratories, Springer, pp. 101-120, Berlin, 2009.
4. M. Alazab, R. Layton, S. Venkataraman, and P. Watters, “Malware Detection Based on Structural and Behavioural Features of API Calls,” 1st International Cyber Resilience Conference, pp. 68-73, Perth Western Australia, August 2010.
5. B. Bashari Rad and M. Masrom, “Metamorphic Virus Detection in Portable Executables Using Opcodes Statistical Feature,” Proceeding of the International Conference on Advance Science, Engineering and Information Technology Malaysia, January 2011.
6. L. Wu, R. Ping, L. Ke, and H.-X. Duan, “Behavior-Based Malware Analysis and Detection,” Complexity and Data Mining (IWCDM), pp. 39-42, Nanjing, September 2011.
7. A. Oliva and A. Torralba, “Modeling the Shape of the Scene: A Holistic Representatio of the Spatial Envelope,” International Journal of Computer Vision, vol. 42, no. 3, pp. 145–175, 2001.
8. G. Conti, E. Dean, M. Sinda, and B. Sangster, “Visual Reverse Engineering of Binary and Data Files,” Springer, pp. 1–17, 2008.
9. K. Soo Han, J. Hyun Lim, B. Kang, and E. Gyu Im, “Malware analysis using visualized images and entropy graphs,” Springer, Int. J. Inf. Secur., 2014.
10. K. Soo Han, J. Hyun Lim, B. Kang, and E. Gyu Im, “Malware Analysis Using Visualized Image Matrices,” Hindawi, the Scientific World Journal, vol. 2014, Article ID 132713, p. 15, 2014.

Malware Detection Using Image Visualization

H. Ranji, S. Parsa*

Abstract

In this article a new technique is proposed for detection of polymorphic malware based on image processing. With the proliferation of polymorphic malware, the efficacy of signature-based static analysis systems is greatly reduced. This survey is based on the comparison of the images developed from malware samples binary code. With the advent of image processing applications for binary code analysis, numerous features could be extracted for comparing malware isomorphs. Based on these features, we have been capable of detecting malware isomporphs with an unprecedented accuracy. Most often, malware samples binaries are packed. Using our proposed method, we have been capable of detecting the unique similarity between executables packed with a same packer.

Key Words: *Malware, Viruses, Visualizing, Polymorphic, Image Proccessing*

*Iran University of Science & Technology (parsa@iust.ac.ir) - Writer-in-Charge