مجله علمی - پژوهشی "پدافند الکترونیکی و سایبری"
سال پنجم، شماره ۱، بهار ۱۳۹۶، ص ۸۳-۷۳

تخمین مخاطرات امنیتی نرم‌افزارهای اندروید با استفاده از بهره اطلاعاتی

محمود دی‌پیر[1*]

۱- استادیار، دانشکده رایانه و فناوری اطلاعات، دانشگاه هوایی شهید ستاری، تهران

(دریافت: ۹۴/۱۱/۰۹، پذیرش: ۹۵/۰۸/۱۰)

چکیده

با گسترش روز افزون بدافزارها در اندروید به عنوان پرکاربردترین سیستم عامل همراه، دانستن میزان خطر امنیتی هر نرم‌افزار می‌تواند در اعلام هشدار به کاربر نسبت به استفاده از بدافزارهای احتمالی، مؤثر باشد. مخاطرات امنیتی نرم افزارهای اندروید از طریق مجوزهای درخواستی آنها قابل تخمین است. در این مقاله با توجه به میزان سوء استفاده از مجوزهای درخواستی توسط بدافزارهای شناخته شده قبلی، مفهوم مجوز بحرانی به صورت دقیق‌تری تعریف شده است. بر اساس این تعریف و با تحلیل مجوزهای درخواستی توسط بدافزارها و نرم افزارهای مفید شناخته شده، معیار جدیدی به منظور اندازه‌گیری خطر امنیتی نرم‌افزارهای اندروید ارائه شده است. در این معیار مجوزهایی اثر بیشتری در محاسبه مقدار خطر امنیتی دارند که بهره اطلاعاتی بیشتری در تمایز بدافزارها داشته باشند. آزمایش‌های صورت گرفته نشان دهنده نرخ تشخیص بالاتر و قابلیت تعمیم‌پذیری بیشتر معیار ارائه شده نسبت به معیارهای قبلی است.

واژه‌های کلیدی: خطر امنیتی، اندروید، بدافزار، بهره اطلاعاتی، مجوزهای بحرانی.

۱- مقدمه

امروزه دستگاه‌های تلفن همراه قابلیت‌های متنوعی دارند و در محیط‌های مختلف از جمله محیط‌های حساس و نظامی قابل استفاده هستند. از میان سیستم عامل‌های توسعه داده شده برای تلفن همراه و دستگاه‌های هوشمند قابل حمل، سیستم عامل اندروید فراگیرترین آن‌ها است. برای این سیستم عامل تا کنون نرم‌افزارهای زیادی توسعه داده شده‌اند. بسیاری از این نرم‌افزارها توسط افراد ناشناخته و توسعه‌دهندگان گمنام ارائه شده‌اند. مدل امنیتی نرم‌افزارها در سیستم عامل اندروید بر اساس مجوز[1]هاست. این مجوزها فقط یک بار و در ابتدای نصب هر نرم‌افزار از کاربر پرسیده می‌شوند و پس از آن، کاربر چندان دخالتی در امنیت نرم‌افزار نصب‌شده ندارد. معمولاً خود کاربران هم وقت زیادی به منظور مطالعه و دقت در فهرست مجوزها در صفحه ابتدایی نصب نرم‌افزار، اختصاص نمی‌دهند. علاوه بر این کاربران عادی معمولاً دانش فنی برای شناخت تأثیر استفاده از هر مجوز در نرم‌افزار مورد استفاده را ندارند. بنابراین، این مدل امنیتی کارایی چندانی در ارتقاء امنیت کاربران به منظور حفظ داده‌های شخصی و حریم خصوصی آنها ندارد. نرم‌افزارهای مخرب مانند تروجان‌ها، جاسوس‌افزارها[2] و تبلیغ‌افزارها[3] قادرند با فریب کاربران، خود را در در قالب نرم‌افزار مفید و بی‌خطر[4] نشان داده و اطلاعات حساس شرکت‌ها و مراکز نظامی را به سرقت برند. این‌گونه بدافزارها[5] همچنین می‌توانند با سرقت داده‌های شخصی افراد و انتشار آنها، حریم خصوصی آن‌ها را نقض کنند. بر طبق آمارهای غیررسمی در سال ۲۰۱۵ در بین هر ۵ نرم‌افزار توسعه داده شده اندروید، یکی بدافزار بوده است. بر طبق همین آمارها و با توجه به گستردگی استفاده از سیستم عامل اندروید نسبت به سایر سیستم عامل‌های همراه، ۹۷ درصد بدافزارهای موبایل مربوط به اندروید می‌شوند. بنابراین، شناسایی و مقابله با این بدافزارها، بسیار حائز اهمیت است. تاکنون تحقیقاتی به منظور آگاه‌سازی بهتر کاربران در زمینه امنیت نرم‌افزارها در اندروید صورت گرفته است [۱]. استفاده از عناوین مناسب تر برای مجوزها، دسته‌بندی مجوزها، کاهش تعداد عنوان مجوزها، استفاده از نظرات کاربران علاوه بر مجوزها، نمونه‌هایی از راه‌کارهای ارائه‌شده در این تحقیقات هستند. علاوه بر این، تاکنون معیارهای مختلفی نیز برای سنجش خطر امنیتی یک نرم‌افزار اندرویدی ارائه شده است.

* رایانامه نویسنده مسئول: mdeypir@ssau.ac.ir

1- Permission

2- Spywares
3- Adware
4- Benign App
5- Malware

تعداد مجوزهای حساس درخواستی و تعداد جفت مجوزهای حساس درخواستی نمونه‌هایی از این معیارها هستند [۲]. با استفاده از این معیارها و داشتن یک حدآستانه می‌توان پس از تخمین خطر امنیتی یک نرم‌افزار مشکوک، در صورت بالابودن مقدار خطر آن، هشدار امنیتی صادر کرد. در این مقاله معیار جدیدی به منظور اندازه‌گیری خطر یک نرم‌افزار اندرویدی ارائه شده که نسبت به معیارهای ارائه شده قبلی کارائی بهتری دارد یعنی قادر است درصد بیشتری از بدافزارها را شناسایی کند. در این تحقیق، به منظور محاسبه دقیق خطر امنیتی یک نرم‌افزار با استفاده از مجوزها، بر مجوزهایی تمرکز شده است که سبب تمایز بدافزارها از نرم‌افزارها شوند. به این منظور از نظریه اطلاعات و مفهوم آنتروپی استفاده شده است و مجوزهایی جستجو و استخراج شده‌اند که بهره اطلاعاتی بیشتری داشته باشند. یعنی آنهایی که بیشتر سبب تمایز بدافزارها از نرم‌افزارها می‌شوند در معیار پیشنهادی به کار گرفته شده‌اند.

در بخش بعد برخی از کارهای تحقیقاتی انجام شده مرتبط با امنیت اندروید معرفی شده‌اند. در بخش سوم، صورت مسئله بیان می‌شود. در بخش چهارم معیار جدید پیشنهادی معرفی شده و نحوه محاسبه آن تشریح شده است. در بخش پنجم آزمایش‌هایی به منظور ارزیابی و مقایسه معیار پیشنهادی با معیارهای قبلی، ارائه شده است. در این بخش با استفاده از مجموعه داده‌هایی متشکل از صدها بدافزار و هزاران نرم‌افزار مفید شناخته شده اندروید، معیار پیشنهادی با معیارهای ارائه شده قبلی از نظر اندازه‌گیری میزان خطر و توان تشخیص بدافزارها از نرم‌افزارها مقایسه خواهد شد. در نهایت این مقاله در بخش ششم جمع‌بندی و نتیجه‌گیری می‌شود.

۲- مروری بر تحقیقات گذشته

با توجه به معماری امنیتی خاص اندروید و محدودیت‌های آن، تحقیقات مختلفی در این حوزه انجام گرفته است. مطالعات نشان می‌دهند که کاربران اغلب از بررسی مجوزهای درخواستی نرم‌افزارها در اندروید صرف‌نظر می‌کنند. در برخی از تحقیقات انجام شده اخیر، تلاش شده است بر این مشکل فائق آیند [۶-۳]. فلت و همکاران [۳] روش‌هایی مانند تغییر دسته‌بندی مجوزها، تأکید بر مفهوم خطر امنیتی و چگونگی اختصاص مجوزها را ارائه دادند. در [۷] اطلاعاتی سطح بالا شامل موارد حفظ حریم خصوصی مانند داده‌های شخصی، مکانی و دفترچه تلفن، به جای فهرست مجوزها در صفحه معرفی نرم‌افزار پیشنهاد شد. به منظور کاهش فضای لازم برای نمایش این‌گونه اطلاعات و کمک به کاربر برای تصمیم‌گیری بهتر در انتخاب و نصب، در [۱] فاکتورهای مخاطره و ایمنی[1] که حاصل آنها مقادیر کمی بوده و از روی مجوزها قابل محاسبه هستند، ارائه شدند. با بررسی کاربران مشخص شد که این معیارها تأثیر بیشتری نسبت به اطلاعات متنی دارند. پنگ و همکاران [۸] روشی اساسی بر مبنای مدل احتمالی را به منظور رتبه‌بندی نرم‌افزارهای اندروید بر اساس مجوزهای درخواستی، ارائه دادند. این روش قادر است نرم‌افزارهای موجود در یک فروشگاه نرم‌افزار مانند فروشگاه گوگل را رتبه‌بندی کند. چنین رتبه بندی‌هایی می‌تواند به انتخاب نرم‌افزارهای با امنیت بیشتر توسط کاربران کمک کنند. در مقابل تحقیقات مرور شده بالا، دسته دیگری از تحقیقات وجود دارند که روش‌هایی را به منظور دسته‌بندی نرم‌افزارها و بدافزارهای اندرویدی ارائه داده‌اند. برخی از این تحقیقات با استفاده از مجوزهای درخواستی نرم افزارها، به تشخیص نرم افزارهای مخرب یا مشکوک پرداخته‌اند[۱۱-۹]. برخی نیز از تحلیل ایستای کد نرم‌افزار، توابع برنامه‌نویسی مورد استفاده و مطابقت آن با برخی الگوهای موجود بدافزارها، برای تشخیص بدافزارهای جدید استفاده کرده‌اند [۱۵-۱۲]. تعدادی از محقین نیز روش‌هایی را ارائه داده‌اند که با تحلیل رفتاری نرم‌افزار درحال اجرا، سعی در تشخیص نرم‌افزارهای مخرب اندرویدی دارند [۲۰-۱۶].

باررا و همکاران [۲۱] روشی را برای ارزیابی عملی مدل‌های امنیتی بر اساس مجوز، به کمک نقشه‌های خود سازمانده[2] ارائه داده‌اند. آنها روش خود را به منظور تحلیل توزیع مجوزها بر روی هزار برنامه اعمال کرده و نشان دادند که چگونه استفاده از مجوز با دسته‌بندی برنامه‌ها ارتباط پیدا می‌کند. در [۲۲] تلاش شده است که با دیکامپایل کردن و تحلیل کد به‌دست‌آمده نرم‌افزارها، نشت داده را تشخیص دهند. انک و همکاران [۲۳] سامانه‌ای را توسعه دادند که ترکیب مجوزهای خطرناک را به منظور چگونگی برآورده کردن سیاست های امنیتی به کار می‌برد. در این سیاست‌ها به صورت دستی ترکیب مجوزهای خطرناک مانند FINE_LOCATION و INTERNET در نظر گرفته شده است. این ترکیبات خود با تحلیل بدافزارهای شناخته شده به‌دست می‌آیند. ابزاری بنام MAST در [۲۴] توسعه داده شده که نرم‌افزارهایی که به احتمال زیاد بدافزار هستند را بر اساس تحلیل کد و تحلیل مجوزها تشخیص می‌دهد. ابزار PScout [۲۵] با استفاده از تحلیل ایستای کد اندروید، چگونگی نگاشت مجوز به تابع را در اندروید بررسی می‌کند. این ابزار نشان داد که سیستم مجوزهای اندروید حداقل افزونگی را داشته و این مسئله با توسعه اندروید و ارائه نسخه‌های جدید نیز پایدار باقی مانده است. هدف ما در این تحقیق ارائه معیاری دقیق‌تر و کاراتر به منظور سنجش مخاطره امنیتی نرم‌افزارها در اندروید است. از این نظر، تحقیق ما با تحقیقات ارائه شده در [۲ و ۸] همسو است.

1 -Safety

2 -Self-organizing maps

۳- بیان مسئله

خطر امنیتی، مقداری است که نشان دهنده بدافزار بودن یک برنامه اندرویدی است. هرچه این مقدار بیشتر باشد احتمال بدافزار بودن یک نرم افزار بیشتر است، اما این مقدار خود از جنس احتمال نیست و از قوانین احتمالی تبعیت نمی کند. یعنی هیچ محدودیتی از نظر بازه مقادیر آن وجود ندارد. تنها کافی است برای بدافزارها مقدار بیشتری نسبت به نرم‌افزارها بدهد. نرم‌افزارهای مخرب یا مشکوک معمولاً دارای خطر امنیتی بالایی هستند چون با استفاده از مجوزهایی که دریافت می‌کنند قادرند به منابع سخت‌افزاری و نرم‌افزاری حساس دستگاه همراه، دسترسی داشته باشند. اما بالابودن خطر امنیتی لزوماً به معنای مخرب بودن برنامه نیست زیرا برخی از نرم‌افزارهای مفید به دلیل قابلیت‌هایی که دارند از مجوزهای زیاد و حساسی استفاده می‌کنند. بنابراین، برای تشخیص بهتر بدافزارها از نرم‌افزارها و جلوگیری از تشخیص اشتباه نیاز به معیاری داریم که الگو و رفتار بدافزارها و نرم‌افزارهای فعلی را به‌خوبی به‌کار برده و بتواند به نرم‌افزارهای مخرب، خطر امنیتی بالا و به نرم‌افزارهای مفید، تا حد امکان، خطر امنیتی پایینی نسبت دهد. به این منظور نیازمند تعیین مجوزهای بحرانی یا حساس هستیم. بنابراین، هدف ما در اینجا دسته‌بندی نرم‌افزارهای اندرویدی به دو دسته نرم‌افزار مفید و نرم‌افزار مخرب نیست، بلکه ما به دنبال معیار مناسبی به منظور تخمین مقدار مخاطره امنیتی هستیم که برای هر دو گروه نرم‌افزار مفید و بدافزار معنا دارد.

۴- معرفی معیار پیشنهادی

در تحقیقات گذشته تعاریف مشابهی برای مفهوم مخاطره امنیتی یک نرم‌افزار اندروید با توجه به مجوزهایی که استفاده می‌کند، ارائه شده است. در [۲] هشت معیار مختلف اندازه‌گیری خطر با توجه به مجوزهای درخواستی نرم‌افزار بررسی شده است. برخی از این معیارها در تحقیقات گذشته نیز مورد توجه قرار گرفته بود [۱،۸،۲۳]. هر هشت معیار بر اساس مفهومی به نام مجوز بحرانی عمل می‌کنند. در این تحقیقات، یک مجوز بحرانی، مجوزی است که قبلاً در بدافزارهای اندرویدی شناخته شده، استفاده شده است و یا به منابع نرم‌افزاری و سخت‌افزاری حساس دستگاه دسترسی دارد. از این معیارها می‌توان به منظور اعلام هشدار برای نرم‌افزارهای مشکوک و یا شناسایی بدافزارهای ناشناخته جدید، استفاده کرد. ما به دنبال معیاری برای محاسبه مخاطرات امنیتی نرم‌افزارهای اندروید هستیم که ضمن ساده بودن، توصیف دقیق‌تری از مخاطره امنیتی نرم‌افزار در اندروید را نشان دهد. علاوه‌بر آن، بتواند به نرم‌افزارهای مفید، مقدار خطر امنیتی پایین و به نرم‌افزارهای مخرب، نسبت به نرم‌افزارهای مفید، مقدار خطر امنیتی بالایی تخصیص دهد. ما برای به‌دست آوردن این معیارها، در ابتدا مجوزهای درخواستی ۸۰۸ بدافزار و ۷۱۳۳۱ نرم‌افزار شناخته شده را تحلیل کرده‌ایم ولی در بخش ارزیابی معیار پیشنهادی، نرم‌افزارهای بیشتری بررسی شده‌اند. معیار پیشنهادی ما بر اساس نظریه اطلاعات و مفاهیم آنتروپی[1] و بهره اطلاعاتی[2] مورد استفاده در یادگیری ماشین [۲۶]، ارائه شده است. در این معیار، بهره اطلاعاتی مجوزها را در نظر می‌گیریم. این بهره اطلاعاتی با توجه به مفهوم آنتروپی قابل محاسبه است. در واقع ما با توجه به مفاهیم حوزه نظریه اطلاعات، آنتروپی مجموعه کل برنامه‌ها را از نظر مفید و مخرب بودن محاسبه می‌کنیم و سپس با توجه به نقش هر مجوز در جداسازی اولیه نرم‌افزارها از بدافزارها، بهره اطلاعاتی را برای هر مجوز به‌دست می‌آوریم. اگر ما در یک مجموعه n تایی از برنامه‌های اندرویدی، دو نوع نرم‌افزار مفید و بدافزار داشته باشیم، در این مجموعه با توجه به نوع نرم-افزار، مقدار آنتروپی به صورت زیر قابل محاسبه است:

$$Entropy(all) = -p_b log_2(p_b) - p_m log_2(p_m) \quad (۱)$$

اگر ما در این مجموعه، برنامه‌ای را به صورت تصادفی انتخاب کنیم p_b و p_m به ترتیب احتمال مفید بودن این نرم‌افزار و احتمال مخرب بودن آن است. این احتمالات به ترتیب با تقسیم تعداد نرم‌افزار مفید بر تعداد کل برنامه‌ها و تقسیم تعداد بدافزارها به کل برنامه‌ها، قابل محاسبه هستند. حال فرض کنیم بر اساس یکی از مجوزهای اندروید بنام x این مجموعه را به دو دسته تقسیم‌بندی می‌کنیم، آنهایی که از این مجوز استفاده کرده‌اند (x=1) و آنهایی که از این مجوز استفاده نکرده‌اند (x=0). طبیعی است که دو مجموعه حاصل ممکن است اندازه‌های متفاوتی مانند n_1 و n_2 داشته باشند. حال اگر میزان آنتروپی را در هر کدام از این مجموعه‌های حاصل به صورت جداگانه حساب کنیم و آنها را به ترتیب Entropy(x=1) و Entropy(x=0) بنامیم، می‌توان مقدار کل آنتروپی حاصل را نسبت به نوع برنامه‌ها به صورت میانگین وزنی دو مجموعه به‌دست‌آمده، حساب کنیم:

$$Entropy(x) = \frac{n_1}{n} Entropy(x=0) + \frac{n_2}{n} Entropy(x=1) \quad (۲)$$

حال با توجه به دو مقدار آنتروپی به‌دست‌آمده در روابط (۱) و (۲)، بهره اطلاعاتی مجوز x به صورت زیر قابل محاسبه است:

$$IG(x) \equiv Risk(x) = Entropy(all) - Entropy(x) \quad (۳)$$

مقدار حاصل بیانگر قابلیت متمایز کننده مجوز x در زمینه شناسایی بدافزارها از نرم‌افزارهاست. به همین دلیل ما این مقدار را خطر امنیتی مجوز x می‌نامیم و از آن برای محاسبه خطر امنیتی یک برنامه اندرویدی استفاده می‌کنیم. از بین کل ۱۲۲ مجوز، استفاده از ۱۱۲ مجوز دارای خطر امنیتی است یعنی برای این تعداد از مجوزها، مقدار بهره اطلاعاتی بالاتر از صفر به‌دست آمده است. از بین این مجوزها، تعداد ۲۰ مجوز با بیشترین مقدار

1 -Entropy
2 -Information Gain

بهره اطلاعاتی یا خطر را به‌دست آورده و به همراه درصد تکرار آن‌ها در بدافزارها، با دقت سه رقم اعشار در جدول (۱) نشان داده‌ایم.

جدول (۱). بحرانی‌ترین مجوزهای اندروید بر حسب بهره اطلاعاتی

رتبه براساس بهره اطلاعاتی	نام مجوز	درصد استفاده در بدافزارها	بهره اطلاعاتی
۱	READ_SMS	۰,۶۷۹	۰,۰۲۹۷
۲	WRITE_SMS	۰,۵۶۲	۰,۰۲۴۹
۳	READ_PHONE_STATE	۰,۹۳۱	۰,۰۱۸۶
۴	WRITE_APN_SETTINGS	۰,۳۲۴	۰,۰۱۷۷
۵	ACCESS_WIFI_STATE	۰,۶۷۱	۰,۰۱۷۷
۶	RECEIVE_BOOT_COMPLETED	۰,۵۶۶	۰,۰۱۴۸
۷	RECEIVE_SMS	۰,۴۶۰	۰,۰۱۳۵
۸	SEND_SMS	۰,۴۸۹	۰,۰۱۳۵
۹	WRITE_CONTACTS	۰,۴۱۷	۰,۰۱۲۴
۱۰	RESTART_PACKAGES	۰,۳۴۸	۰,۰۱۰۸
۱۱	INSTALL_PACKAGES	۰,۲۱۸	۰,۰۱۰۶
۱۲	DISABLE_KEYGUARD	۰,۲۹۰	۰,۰۱۰۱
۱۳	READ_LOGS	۰,۲۶۹	۰,۰۰۹۱
۱۴	ACCESS_WIFI_STATE	۰,۶۷۱	۰,۰۰۹۰
۱۵	CALL_PHONE	۰,۴۱۵	۰,۰۰۷۴
۱۶	CHANGE_WIFI_STATE	۰,۲۶۲	۰,۰۰۷۲
۱۷	WRITE_EXTERNAL_STORAGE	۰,۶۴۴	۰,۰۰۵۷
۱۸	READ_CONTACTS	۰,۳۹۲	۰,۰۰۵۶
۱۹	WAKE_LOCK	۰,۴۱۶	۰,۰۰۴۶
۲۰	INTERNET	۰,۹۷۸	۰,۰۰۴۲

همان‌طور که در این جدول دیده می‌شود، مقادیر بهره اطلاعاتی بدست آمده برای کلیه مجوزها به طور کلی پایین هستند که علت آن تعداد کم بدافزار مورد استفاده در تحلیل ماست. علت این مسئله عدم دسترسی نویسندگان این مقاله به بدافزارهای بیشتر است. اما از آنجایی که مقایسه‌ها به صورت نسبی است، این مسئله در تحلیل‌های انجام شده ما چندان تأثیر گذار نیست. مقدار بهره اطلاعاتی نسبی بالا، نشان دهنده اهمیت مجوز iام در تشخیص بدافزار است. این جدول از بالا به پایین بر حسب بهره اطلاعاتی به صورت نزولی مرتب شده است و بر همین اساس مجوزها رتبه بندی شده‌اند. از نظرما هرچه اطلاعات متمایز کننده یا خطر یک مجوز بیشتر باشد، این مجوز بحرانی‌تر است. همان‌طور که در جدول (۱) پیداست، بالابودن فرکانس استفاده از یک مجوز در بدافزارها، دلیل بر اهمیت آن برای محاسبه خطر امنیتی نمی‌شود زیرا ممکن است به همین نسبت در نرم‌افزارها هم استفاده شود و در نتیجه بهره اطلاعاتی کمی داشته باشد. برای مثال، همان‌طور که در جدول (۱) دیده می شود، مجوز INTERNET بیشترین استفاده را در بدافزارها دارد ولی در رتبه بیستم بهره اطلاعاتی قرار گرفته است. به عنوان مثالی دیگر، مجوز WRITE_EXTERNAL_STORAGE نسبت به بسیاری از مجوزهای بالاتر خود در جدول (۱) در درصد بیشتری از بدافزارها به‌کار گرفته شده است ولی بهره اطلاعاتی کمتری دارد. چنین مجوزهایی هم در بدافزارها و هم در نرم‌افزارهای بی‌خطر بسیار مورد استفاده قرار می‌گیرند. بنابراین، به اندازه مجوزهای با رتبه بالاتر خود، باعث تمایز بدافزارها از نرم‌افزارها نمی‌شوند. ما در محاسبه مخاطرات امنیتی نرم‌افزارها باید به مجوزهایی که بهره اطلاعاتی یا خطر امنیتی بیشتری دارند، بیشتر بها دهیم. بنابراین، ما از بهره اطلاعاتی کلیه مجوزهای مورد استفاده در یک نرم‌افزار استفاده کرده و فرمولی برای محاسبه خطر ارائه دهیم. هرچه یک مجوز مانند i در بدافزارهای بیشتری استفاده شده باشد و همین مجوز در نرم‌افزارهای مفید چندان استفاده نشده باشد، مقدار بهره اطلاعاتی آن بیشتر است. بنابراین تعریف جدید و دقیق‌تری از مفهوم مجوز بحرانی را بر اساس بهره اطلاعاتی ارائه می‌دهیم.

مجوز بحرانی: مجوزی است که بهره اطلاعاتی قابل توجهی از نظر تمایز بین نرم‌افزارهای مخرب و نرم‌افزارهای مفید داشته باشد.

به کارگیری چنین مجوزی در محاسبه خطر سبب می‌شود که خطر بالای امنیتی برای بدافزارها و خطر امنیتی پایین‌تر برای نرم افزارهای مفید محاسبه شود. نکته مهم در این تعریف این است که استفاده از یک مجوز ممکن است سبب دسترسی برنامه به داده‌های خصوصی شود ولی تاکنون به هر دلیلی مورد استفاده زیاد از جانب بدافزار نویسان قرار نگرفته است. بدیهی است که مقدار خطر یا همان بهره اطلاعاتی یک مجوز ثابت نبوده و با گذشت زمان تغییر می‌کنند و می‌بایست در هر دوره زمانی مجدداً محاسبه شوند زیرا همواره بدافزارهای جدیدی معرفی می‌شوند و الگوی بدافزار نویسی برحسب استفاده از مجوزها تغییر می‌کند. علاوه بر این، تعداد مجوزهای بحرانی نیز ممکن است تغییر کند. با توجه به این وزن‌ها، معیار *خطر امنیتی بر اساس بهره اطلاعاتی* (*IGR(Information Gain based Risk value*) برای محاسبه خطر یک نرم افزار مانند A را با فرمول زیر ارائه می‌دهیم:

$$IGR(A) = \sum_{i=1}^{122} IG(x_i) \times x_{iA} \quad x_{iA} \in \{0,1\},\ i \in \{1,2,\dots,122\} \quad (۴)$$

در این فرمول منظور از A نرم‌افزاری است که می‌خواهیم خطر امنیتی را برای آن محاسبه کنیم. x_{iA} و $IG(x_i)$ به ترتیب درخواست مجوز i ام توسط نرم‌افزار مورد تحلیل A و بهره اطلاعاتی مجوز iام را نشان می‌دهند. متغییر x_i دودویی بوده و $IG(x_i)$ هم مطابق فرمول (۳) بهره اطلاعاتی مجوز x_i است. بدیهی است که مجوزی، ممکن است دارای بهره اطلاعاتی بالا باشد ولی در نرم‌افزار مورد تحلیل استفاده نشده باشد (x_{iA}=0) و یا در نرم‌افزار استفاده شده باشد ولی بهره اطلاعاتی آن صفر باشد ($IG(x_i)$=0). در هر صورت چنین مجوزی در محاسبه خطر این نرم‌افزار تأثیری نخواهد داشت. اگر تعداد بدافزارهای مورد استفاده

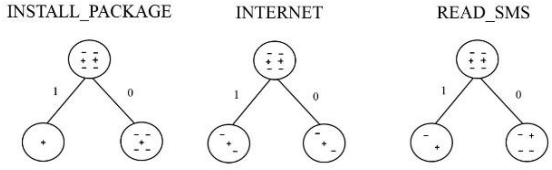
تحلیل ما بیشتر بود و به تعداد نرم‌افزار مفید نزدیک‌تر می‌شد، باز هم مقدار بهره اطلاعاتی مجوزهای حساس نسبت به سایر مجوزها بیشتر می‌شد ولی اعداد بزرگتری به‌دست می‌آمد. مثلاً مجوزهای مربوط به پیام کوتاه مانند SEND_SMS و یا مجوزهای مربوط به مدیریت بسته‌های نرم افزاری مانند INSTALL_PACKAGES که در بسیاری از بدافزارهای اندروید استفاده شده‌اند بیشتر سبب تمایز بدافزار از نرم‌افزار می‌شدند چون از جمله مجوزهای مورد علاقه بدافزار نویسان هستند و در نتیجه مقادیر بهره اطلاعاتی بالاتری برای آنها به‌دست می‌آمد. به منظور توصیف بیشتر چگونگی محاسبه خطر امنیتی برای برنامه‌های اندروید مطابق معیار پیشنهادی، مثال زیر را در نظر بگیرید:

مثال ۱: فرض کنید مجموعه شش نرم افزار مفید و مخرب (بدافزار) مورد تحلیل ما به صورت جدول (۲) باشند. در این جدول برای هر برنامه اندرویدی، شناسه، لیست مجوزهای آن و همچنین مفید یا مخرب بودن آن مشخص شده است.

جدول(۲). اطلاعات مربوط به مثال ۱ نرم افزارهای مفید(-) و بدافزارها(+)

ID	Permissions	Malware
1	INTERNET, READ_PROFILE	-
2	BATTERY_STATS, BLUETOOTH	-
3	BROADCAST_SMS,WRITE_SMS	+
4	INTERNET,INSTALL_PACKAGE, READ_SMS	+
5	READ_SMS, WRITE_EXTERNAL_STORAGE	-
6	BATERY_STATS, INTERNET	-

ما باید برای هر مجوز، بهره اطلاعاتی را با توجه به اطلاعات این جدول محاسبه کرده و از این اطلاعات بر اساس معیار ارائه شده به منظور محاسبه خطر امنیتی برنامه‌های اندرویدی استفاده کنیم. بدیهی است که هر مجوز با توجه به اطلاعات داده شده جدول بالا، بهره اطلاعاتی متفاوتی خواهد داشت و با توجه به این بهره اطلاعاتی، در محاسبه خطر امنیتی یک برنامه جدید مؤثر خواهد بود. برای مثال، بهره اطلاعاتی را برای سه مجوز نمونه با توجه به شکل (۱) محاسبه کرده‌ایم. در این شکل علامت مثبت نشان‌دهنده بدافزار و علامت منفی نشان‌دهنده تست منفی یعنی نرم‌افزار مفید است.



شکل(۱). تأثیر استفاده از سه مجوز نمونه در جداسازی بدافزارها(+) از نرم افزارها(-)

همان‌طور که در شکل (۱) مشخص است، برای هر مجوز، با توجه به استفاده (داشتن مقدار ۱) و عدم استفاده (مقدار ۰)، برنامه‌ها به دو دسته تقسیم می‌شوند و درختی شامل یک ریشه و دو فرزند ایجاد می‌شود. فرزند چپ نشان‌دهنده دسته‌ای از برنامه‌ها است که این مجوز را درخواست کرده‌اند و فرزند سمت راست برنامه‌هایی را نشان می‌دهند که این مجوز را برای اجرای خود نیاز ندارند. در هر دسته ممکن هم بدافزار و هم نرم‌افزار مفید وجود داشته باشد. میزان آنتروپی مجموعه نرم‌افزارها، با انجام دسته‌بندی توسط هر مجوز، می‌تواند کاهش یابد و در نتیجه بهره اطلاعاتی بزرگتر از صفر با استفاده از فرمول (۳) به‌دست آید. برای مثال، بهره اطلاعاتی را برای مجوز INSTALL_PACKAGE محاسبه می‌کنیم. از آنجایی که مجموعه نرم‌افزارهای مورد تحلیل در اینجا ۶ است و از این تعداد ۲ مورد بدافزار است، آنتروپی ریشه درخت یعنی مجموعه کل نرم‌افزارها برابر خواهد بود با:

$$ParentEntropy = -\frac{2}{6}log_2\left(\frac{2}{6}\right) - \frac{4}{6}log_2\left(\frac{4}{6}\right) = 0.7986 \quad (۵)$$

با توجه به مجموعه برنامه‌های مثال ما در جدول (۲)، تنها یک بدافزار از این مجوز استفاده کرده است. یک بدافزار و چهار نرم افزار از این مجوز استفاده نکرده‌اند. بنابراین آنتروپی این فرزندان به ترتیب برابر است با:

$$Entropy(INSTALL\_PACKAGES = 1) = -\frac{1}{1}log_2\left(\frac{1}{1}\right) = 0 \quad (۶)$$

$$Entropy(INSTALL\_PACKAGES = 0) = -\frac{1}{5}log_2\left(\frac{1}{5}\right) - \frac{4}{5}log_2\left(\frac{4}{5}\right) = 0.6429 \quad (۷)$$

میانگین وزنی آنتروپی نودهای فرزند با توجه به تعداد هرکدام به صورت زیر محاسبه می‌شود:

$$Weighted\ Average\ Entropy\ of\ Children = \frac{1}{6}\times 0 + \frac{5}{6}\times 0.6429 = 0.5358 \quad (۸)$$

بر اساس آنتروپی‌های به‌دست‌آمده، میزان بهره اطلاعاتی برای این مجوز بر اساس فرمول (۳) برابر با تفریق میانگین وزنی آنتروپی فرزندان از آنتروپی کل نرم‌افزارهاست:

$$IG(INSTALL\_PACKAGES) = 0.7986 - 0.5358 = 0.2628 \quad (۹)$$

در معیار پیشنهادی ما، این عدد میزان مخاطره امنیتی این مجوز را نشان می‌دهد. میزان بهره اطلاعاتی یا میزان خطر امنیتی دو مجوز INTERNET و READ_SMS به صورت زیر بدست می‌آید، که از محاسبات مربوطه صرف‌نظر کرده و این محاسبات را به عهده خواننده قرار می‌دهیم:

$$IG(INTERNET) = 0 \quad (۱۰)$$

$$IG(READ\_SMS) = 0.0392 \quad (۱۱)$$

با توجه به اعداد می‌توان گفت که در استفاده از مجوز INTERNET هیچ تفاوتی بین بدافزار و نرم‌افزارها نیست. مجوز

READ_SMS اگرچه میزان بهره اطلاعاتی کمتری نسبت به INSTALL_PACKAGES دارد، اما سبب تمایز بدافزارها از نرم‌افزارها می‌شود و به عبارت دقیق‌تر دارای خطر امنیتی کمتری است. میزان خطر امنیتی برای همه مجوزهای اندرویدی با توجه به نرم‌افزارها و بدافزارهای شناخته شده می‌بایست محاسبه شود. ما این کار را برای ۱۲۲ مجوز انجام داده‌ایم. اگرچه ما ایده خود را از یادگیری ماشین و ساختن درخت تصمیم برای دسته‌بندی براساس بهره اطلاعاتی [۲۶] گرفته‌ایم، اما معیار ما هدفی کاملاً متفاوت دارد. نخست این‌که هدف ما در اینجا دسته‌بندی نیست بلکه محاسبه مقدار عددی خطر امنیتی نرم‌افزار براساس میزان بهره اطلاعاتی مجوزها است و ما در اینجا داده‌ها را برچسب نمی‌زنیم. دوم، ما برای همه مجوزها بهره اطلاعاتی را جداگانه حساب می‌کنیم، درحالی که در یک مدل دسته‌بندی، یک درخت تصمیم از یک ویژگی شروع کرده و در سطوح متعددی ساخته می‌شود.

حال فرض کنید با توجه به محاسبات بالا یک برنامه اندرویدی به نام A در اختیار داریم که همه مجوزهای درخواستی آن، سه مجوز بالا است، میزان خطر امنیتی این نرم‌افزار مطابق فرمول (۴) برابر است با:

$$Risk(A) = 0.2628 + 0.0392 + 0 = 0.3020 \qquad (۱۲)$$

این عدد با احتمال بدافزار بودن برنامه مورد بررسی متناسب است. یعنی هرچه مقدار خطر به‌دست‌آمده بیشتر باشد، احتمال مخرب بودن این برنامه بیشتر خواهد بود. البته داشتن خطر امنیتی برای یک برنامه لزوماً بدافزار بودن آن را مشخص نمی‌کند بلکه یک هشدار برای کاربر است و یا می‌تواند به عنوان یک پیش پردازش برای تحلیل‌های جزئی‌تر به منظور شناسایی دقیق بدافزارها به کار رود. علاوه بر این، میزان خطر امنیتی نسبی است و می‌تواند به انتخاب نرم‌افزار بهتر و کم خطرتر کمک کند. یعنی با داشتن دو نرم‌افزار با قابلیت‌های یکسان و مخاطرات امنیتی متفاوت، عقلانی است که نرم‌افزار با خطر امنیتی کمتر استفاده شود.

۵- ارزیابی

به منظور ارزیابی مخاطرات امنیتی نرم‌افزارها و معیار ارائه شده کدهای لازم برای معیار پیشنهادی در نرم‌افزار مطلب توسعه داده شده‌اند. فهرست مجوزهای هر نرم‌افزار اندرویدی از طریق فایل AndroidManifest.xml در پکیج apk مربوط به آن نرم افزار قابل دسترسی است. برای ارزیابی معیارهای ارائه شده، از مجموعه داده مربوط به بدافزارهای اندروید [۲۰] استفاده کرده‌ایم. حدود هشتصد بدافزار این مجموعه داده را بررسی کرده‌ایم. ما معیارها را هم برای بدافزارها و هم برای نرم‌افزارهای مفید محاسبه نموده‌ایم. همان‌طور که گفته شد یک معیار خوب سنجش و تخمین خطر می‌بایست برای بدافزارها عدد بالایی تولید کند و برای نرم‌افزارهای مفید و غیر مخرب عدد کمی به‌دست آورد تا سبب تمایز این دو شود. به منظور ارزیابی معیار ارائه شده بر روی نرم‌افزارهای مفید، از مجموعه داده‌های Market2011 و Market2012 ارائه شده در [۲] استفاده کرده‌ایم. این مجموعه نرم‌افزارها به ترتیب شامل حدود ۷۱ هزار و ۱۳۶ هزار نرم‌افزار مفید مربوط به سایت فروش نرم‌افزار اندروید شرکت گوگل در سال های ۲۰۱۱ و ۲۰۱۲ میلادی هستند. مشخصات مربوط به داده‌های مورد استفاده برای ارزیابی در جدول (۳) خلاصه شده‌اند.

جدول(۳). مجموعه داده‌های مورد استفاده در ارزیابی معیار پیشنهادی

نام مجموعه داده	نوع	تعداد	تعداد مجوز	محتوا
Market2011	نرم افزار مفید	۱۳۶۵۳۴	۱۲۲	نرم افزارهای مفید فروشگاه اندروید گوگل درسال ۲۰۱۱
Market2012	نرم افزار مفید	۷۱۳۳۱	۱۲۲	نرم افزارهای مفید فروشگاه اندروید گوگل درسال ۲۰۱۲
Malwares	بدافزار	۸۰۸	۱۲۲	تعدادی از بدافزارهای شناخته شده اندروید تا کنون

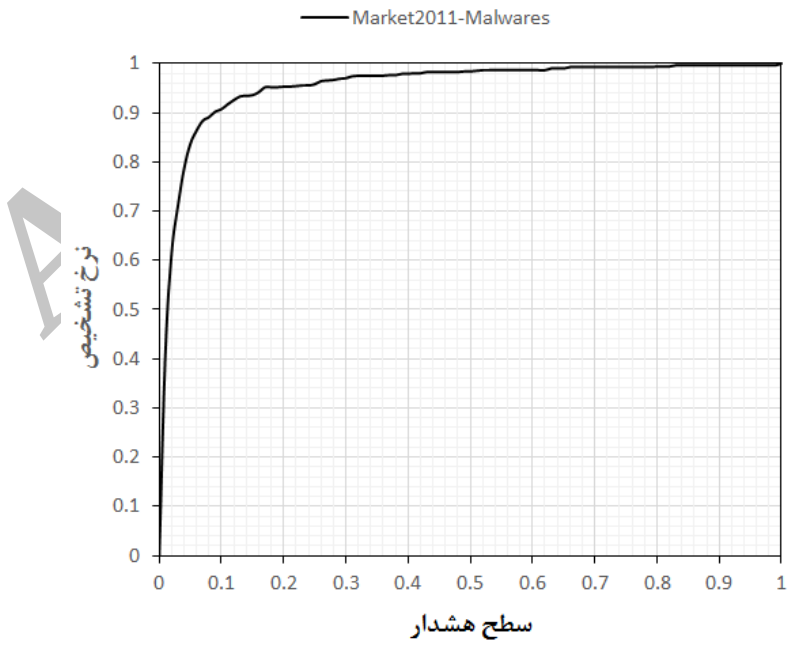
همان‌طور که در بخش مرورکارها گذشته ارائه شد، روش‌هایی مختلفی در منبع [۲] به منظور محاسبه خطر امنیتی بدافزارها ارائه شده است. در این منبع برخی از این روش‌ها کاملاً جدید بوده و برخی نیز حاصل جمع‌بندی کارهای گذشته در زمینه امنیت اندروید هستند که در این منبع به منظور محاسبه خطر امنیتی از آنها استفاده دیگری شده است. برخی از این معیارها از نوع معیار آماری صرف هستند و برخی نیز بر مبنای مدل‌های احتمالی پیشنهاد شده‌اند. ما معیار پیشنهادی خود را با همه این روش‌ها مقایسه کرده‌ایم. به همین منظور جدول (۴) خلاصه‌ای از چگونگی عملکرد آنها ارائه می‌دهد. برای اطلاعات بیشتر در مورد این معیارها، خواننده می‌تواند به منبع [۲] مراجعه کند. اگرچه برخی از این معیارها مشابهت‌هایی با یکدیگر دارند ولی معیار پیشنهادی ما ماهیتی کاملاً متفاوت دارد. در واقع از آنتروپی و بهره اطلاعاتی در هیچ یک از معیارهای قبلی، استفاده نشده است.

ما معیار پیشنهادی خود را با ۸ معیار جدول (۴) که در منبع [۲] معرفی شده‌اند، مقایسه کرده‌ایم. برای انجام آزمایش‌ها از تنظیماتی مشابه منبع [۲] استفاده شده است. کد مربوط به پیاده سازی معیارهای BNB، MNB و HMNB از یکی از نویسندگان منبع [۲] دریافت شد. سایر معیارها را در مطلب و اکسل پیاده‌سازی و محاسبه کرده‌ایم.

جدول (۴). خلاصه ی معیارهای ارائه شده قبلی

نام معیار	توصیف مختصر معیار
RCP (Rare Critical Permission)	مجوزهای بحرانی مورد استفاده در بدافزارها
RPCP (Rare Pairs of Critical Permissions)	جفت مجوزهای بحرانی استفاده شده در بدافزارها
RS (Rarity based risk Score)	محاسبه خطر بر اساس میزان کمیابی مجوزها
RSS (Rarity based risk Score with Scaling)	مجوزهای کمیاب به همراه وزن دهی
BNB (Basic Naive Bayes model)	مدل احتمالی نایو بیز پایه
PNB (Naive Bayes with informative Priors)	مدل احتمالی نایو بیز با دانش اضافه اولیه
MNB (Mixture of Naive Bayes models)	ترکیب مدل‌های نایو بیز
HMNB (Hierarchical Mixture of Naive Bayes models)	ترکیب توارثی مدل‌های نایو بیز

در آزمایش‌ها، تمرکز بر روی توانایی تشخیص[1] این معیارهاست. یعنی معیاری از نظر ما موفق است که بتواند به طور نسبی برای بدافزارها مقدار خطر امنیتی بیشتری محاسبه کند. یعنی اگر برای همه نرم‌افزارها و بدافزارها مقدار خطر را بر اساس یک معیار محاسبه کرده و سپس لیست کلی برنامه‌ها را به ترتیب نزولی مقدار خطر مرتب کنیم، بدافزارهای بیشتری نسبت به نرم افزارهای مفید در بالای لیست قرار گیرند. به این منظور ما در آزمایش اول مجموعه نرم‌افزارهای Market2011 و Malwares را در یک لیست واحد قرار داده و با استفاده از ۹۰٪ لیست حاصل مدل خود را می‌سازیم. سپس با استفاده از ۱۰٪ باقیمانده به تست روش خود می‌پردازیم. به این صورت که با استفاده از مدل به‌دست‌آمده خطر امنیتی آنها را محاسبه کرده وسپس به صورت نزولی مرتب کرده‌ایم. حال در هر بار درصدهای مختلفی را از برنامه‌های بالای لیست مربوطه انتخاب کرده و بررسی می‌کنیم که چه درصدی از بدافزارها در این بخش از لیست قرار گرفته‌اند. به درصدهای انتخاب شده لیست، سطح هشدار[2] و به درصدهای شناسایی شده بدافزارها، نرخ تشخیص گوئیم. اگرچه بازه مقادیر محاسبه‌شده در معیارهای مختلف متفاوت است، با استفاده از این روش ارزیابی، تفاوت در مقادیر تأثیری در مقایسه‌ها نخواهد داشت زیرا سطح هشدار در اینجا عدد خاصی نیست، بلکه نسبی است. بدیهی است که هرچه معیار قوی‌تر باشد، درصد بیشتری از بدافزارها در بالای لیست قرار می‌گیرد و نرخ تشخیص بیشتر خواهد بود. در آزمایش اول با استفاده از مجموعه داده‌های Market2011 و Malwares میزان تشخیص را بر حسب سطح هشدار به‌دست آورده‌ایم. شکل (۲) نمودار *ROC* مربوطه را برای معیار پیشنهادی نشان می‌دهد. در این شکل محور افقی سطح هشدار و محور عمودی نرخ تشخیص بدافزار است. به عبارت دیگر، با انتخاب یک درصد مشخص از کل لیست در محور افقی، محور عمودی درصد بدافزارهای تشخیص داده شده را در این لیست نشان می‌دهد.



شکل (۲). نرخ تشخیص بدافزار با توجه به سطح هشدار انتخابی

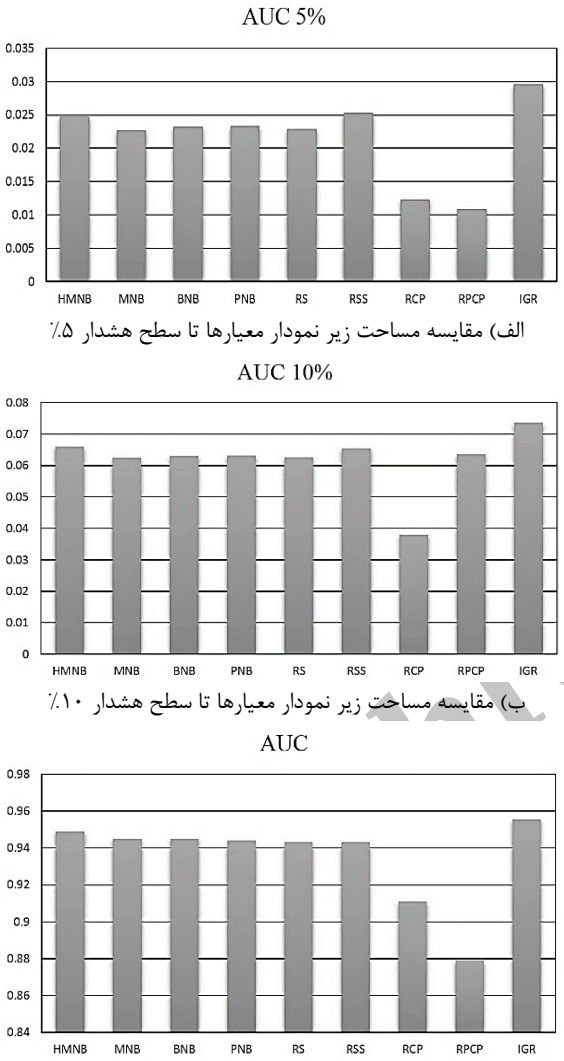
1- Detection Rate

2- Warning Rate

نمودار ROC سایر معیارها به منظور جلوگیری از آشفته شدن نمودار نشان داده نشده‌اند، زیرا مقادیر مربوط به برخی از معیارها به هم نزدیک بوده و قابل تمایز از هم نیستند. همان‌طور که در این شکل دیده می‌شود، مساحت زیر این نمودار به یک نزدیک شده است که نشان دهنده کارایی روش پیشنهادی است. به منظور مقایسه معیار پیشنهادی با سایر معیارها، مساحت زیر نمودار AUC را برای سطوح هشدار مختلف از صفر تا درصدهای ۵ و ۱۰ و همچنین کل سطح زیر نمودار ROC را به‌دست آورده‌ایم. حاصل کار در بخش های مربوط به شکل (۳) نشان داده شده است.

همان‌طور که از این شکل پیدا است روش پیشنهادی مقادیر بیشتری دارد، یعنی میزان تشخیص بالاتری را در مقادیر مختلف به خود اختصاص داده است. دلیل این مسئله، توان متمایز کننده بیشتر معیار ما است که مبتنی بر بهره اطلاعاتی است و پایه نظری قوی‌تری نیز دارد. بنابراین، استفاده از آنتروپی برای محاسبه خطر نسبت به سایر روش‌ها بهتر عمل می‌کند. معیار RSS اگرچه از نظر نرخ تشخیص نزدیک به معیار پیشنهادی است، اما رویکردی کاملاً متفاوت دارد. با استفاده از این معیار مجوزهای نادر در نرم‌افزارهای مفید شناسایی و برای محاسبه خطر امنیتی به کار می‌روند. مجوزهای نادر آنهایی هستند که به ندرت در نرم‌افزارهای مفید مورد استفاده قرار می‌گیرند اما در بدافزارها مورد توجه نفوذگران هستند. علاوه بر این، در RSS به ۶ مجوز حساس‌تر وزن‌های بیشتری هم اختصای داده شده است. در این معیار هرچه یک مجوز در نرم‌افزارها کمتر استفاده شود تأثیر بیشتری در محاسبه خطر خواهد داشت. اما عیب اصلی RSS این است که یک مجوز ممکن است هم در بدافزارها و هم در نرم‌افزارهای عادی نادر باشد که این مسئله سبب محاسبه خطر بالا برای برخی از نرم‌افزارها خواهد شد که در نهایت سبب افت نرخ تشخیص این معیار می‌شود.

به منظور بررسی و مقایسه قابلیت تعمیم‌پذیری[1] روش پیشنهادی بر روی داده‌های دیده نشده و جدید، آزمایش بالا را به شکل دیگری تکرار کرده‌ایم. با توجه به این‌که در آزمایش‌های انجام گرفته معیار RSS نزدیک‌ترین کارایی را با روش پیشنهادی IGR دارد، قابلیت تعمیم‌پذیری این دو معیار با هم مقایسه شده‌اند. در این آزمایش یک بار به کمک مجموعه داده Market2011 و بدافزارها مدل را ایجاد می‌کنیم. به عبارت دیگر برای معیار IGR بهره اطلاعاتی هر مجوز را به‌دست می‌آوریم و برای روش RSS میزان نادر بودن مجوزها را به‌دست می‌آوریم. سپس بر روی مجموعه نرم‌افزارهای مفید Market2012 و بدافزارها، معیار پیشنهادی را با استفاده از مدل به‌دست‌آمده، محاسبه می‌کنیم. به عبارت دیگر از کل داده های Market2011 به منظور آموزش و ازکل نرم‌افزاهای Market2012 که تعداد آنها حدود دو برابر است برای تست استفاده کرده‌ایم. در فازهای آموزش و تست، توان تشخیصی را برای مقادیر مختلف محاسبه کرده و با استفاده از نتیجه به‌دست‌آمده نمودار ROC هر دو فاز را برای معیارهای IGR و RSS به‌دست آورده‌ایم. نتایج کار برای معیار IGR در شکل (۴) و برای معیار RSS در شکل (۵) نشان داده شده‌اند.



الف) مقایسه مساحت زیر نمودار معیارها تا سطح هشدار ۵٪

ب) مقایسه مساحت زیر نمودار معیارها تا سطح هشدار ۱۰٪

ج) مقایسه کل مساحت زیر نمودار معیارها برای همه سطوح هشدار

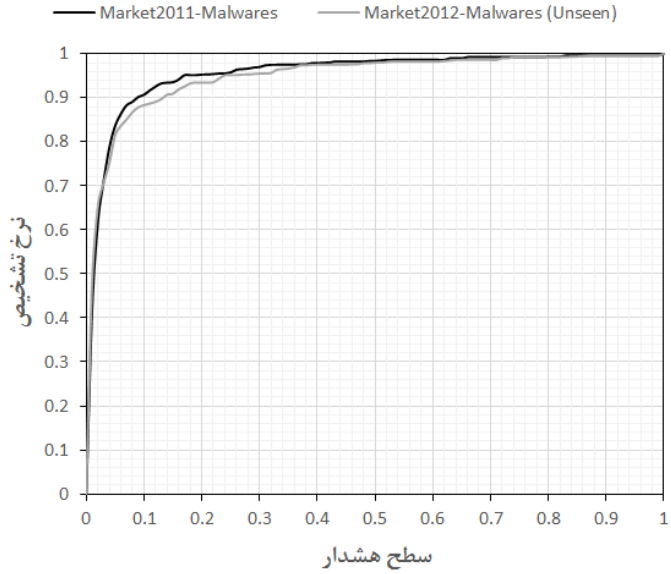
شکل(۳). مقایسه مساحت زیر نمودار منحنی نرخ تشخیص از صفرتا درصدهای مختلف سطح هشدار

در شکل (۴) مشاهده می‌شود که نمودارها در هر دو حالت بسیار به هم نزدیک هستند. توان تشخیصی در حالت داده‌های دیده نشده یعنی Market2012 اندکی افت کرده است که قابل چشم پوشی است. دلیل افت کارایی وجود داده‌ها و الگوهای جدید استفاده از مجوزها است زیرا با گذشت زمان چگونگی استفاده از برخی از مجوزها در نرم‌افزارهای رایج تغییر می‌کند. و طبیعی است که میزان خطر مجوزها هم بر همین اساس تغییر

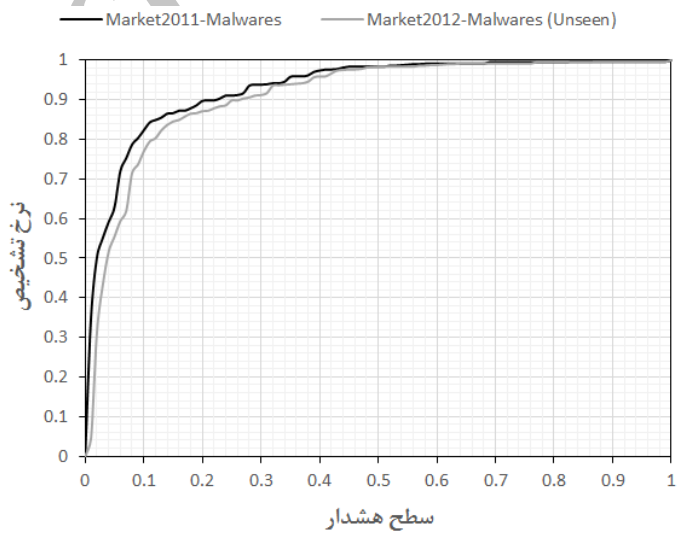
1 -Generalization

مـی‌کنـد. بنـابراین، بـه منظـور محاسـبه دقیـق و درسـت خطـر برنامه‌های اندرویدی نیاز است که مخـاطرات امنیتـی مربـوط بـه مجوزها به صورت دوره‌ای به‌روزرسانی شوند. اما با دقت در شکل (۵) و مقایسه آن با شکل (۴) دیده می‌شود که میزان افت کارایی برای معیار RSS بیشتر است. بنابراین، تعمیم‌پـذیری ایـن معیـار نسبت به معیار پیشنهادی IGR کمتر است. علاوه‌براین، در مقادیر کم سطح هشدار که اهمیـت بیشـتری بـرای کـاربران دارد افـت کارایی معیار RSS نسبت بـه معیـار IGR قابـل توجـه و ملمـوس است.

علت اندک بودن افت کارایی معیار ارائه‌شده بر روی داده‌های جدید این است که این معیار خود از ترکیبی مـدل‌هـای درختـی بسیار ساده‌ای تشکیل شده اسـت کـه نمونـه آنهـا در شـکل (۱) نشان داده شدند. در دنیای یادگیری ماشین ثابت شده اسـت کـه مدل‌های ساده کارایی بسیار زیـادی روی داده‌هـای دیـده نشـده دارند امـا در معیـار RSS عـلاوه بـر شناسـایی مجوزهـای نـادر و اندازه‌گیری مقادیر خطر برای آنها لازم است مجوزهـای حسـاس وزن‌دهی شوند و این وزن‌دهی در توان تشخیصی داده‌های دیـده نشده تأثیر منفی خواهد داشت زیرا وزن‌دهی بر اسـاس داده‌هـای دیده شده قبلی انجام شده است. به عبارت دیگـر در RSS، مـدل به‌دست‌آمده بیش از حد به داده‌های آموزشی تطبیق می‌یابد.



شکل(۴). نرخ تشخیص معیار ارائه شده IGR برای بدافزارهای دیده شده قبلی و دیده نشده



شکل(۵). نرخ تشخیص معیار RSS برای بدافزارهای دیده شده قبلی و دیده نشده

۶- نتیجه گیری

شناسایی بدافزارهای اندرویدی نیازمند معیارهای امنیتی دقیق‌تری است. این معیارها می‌توانند در آنتی ویروس‌های مربوط به سیستم عامل اندروید برای شناسایی اولیه نرم‌افزارهای مخرب و اعلام هشدار در استفاده از آنها به‌کار روند. شناسایی مجوزهای حساس نقش مهمی در محاسبه مخاطرات امنیتی برنامه‌های اندروید دارد. از مجوزهای حساس هم در نرم‌افزارهای مفید استفاده می‌شود و هم در بدافزارها سوء استفاده می‌شود. معمولاً بدافزارها در اندروید خود را در قالب یک نرم‌افزار مفید نشان می‌دهند. در این صورت یک سری مجوز برای کار مفیدشان نیاز دارند و یک سری مجوز هم برای کار مخربشان نیاز دارند. همین مسئله سبب می‌شود که در کل الگوی استفاده از مجوزها در آنها با نرم‌افزارهای مفید متفاوت باشد و در نتیجه خطر امنیتی آنها بیشتر شود. اما ممکن است نرم‌افزارهایی هم وجود داشته باشند که مجوزهای مورد استفاده آنها بسیار شبیه بدافزارها باشد. در این صورت توسط یک معیار امنیتی برای آنها هم خطر امنیتی بالایی محاسبه شود و معیار پیشنهادی هم از این قاعده مستثنی نیست اگرچه نسبت به سایر معیارها کارایی بهتری دارد. برای تشخیص دقیق‌تر یک بدافزار لازم است از روش‌های مکملی مانند تحلیل ایستا و پویای کد و همچنین روش‌های داده کاوی استفاده کرد. معیار پیشنهادی ما که از بهره اطلاعاتی مجوزها به منظور شناسایی مجوزهای حساس استفاده می‌کند، نسبت به معیارهای ارائه شده قبلی بهتر عمل می‌کند. آزمایش‌ها بر روی داده‌های واقعی نشان دادند که معیار پیشنهادی برای بدافزارهای شناخته شده نسبت به نرم‌افزارهای مفید، مقدار خطر بسیار بیشتری را به‌دست می‌آورد. البته تعداد بدافزارهای مورد استفاده در محاسبات ما نسبت به نرم‌افزارهای مفید بسیار کم است و نیاز است که مجموعه بدافزارها تکمیل شود. در صورتی که تعداد بدافزارها بیشتر بود مقادیر بهره اطلاعاتی به‌دست‌آمده در معیار پیشنهادی نسبت به مقادیر فعلی بیشتر می‌شد و مقادیر دقیق‌تری به‌دست می‌آمد. البته این مسئله در توان تشخیصی معیارها و مقایسه آنها چندان مؤثر نیست زیرا مقایسه‌ها نسبی انجام شده است و در هر معیار برای همه بدافزارها و نرم‌افزارهای مفید از مقادیر خطر یکسانی برای مجوزها استفاده شده است. معیار ما اگرچه با استفاده از مجوزها به‌دست آمد، ولی روش مورد استفاده در اینجا قابل اعمال بر روی فهرست توابع مورد استفاده در کد نرم‌افزارها نیز است و این دو می‌توانند مکمل هم نیز باشند یعنی می‌توان معیاری داشت که هم از مجوزها و هم از تحلیل کد دی‌کامپایل شده نرم‌افزارها به منظور تخمین دقیق‌تر میزان مخاطره امنیتی آنها استفاده کند. مدل سازی رفتار انتشاری بدافزارها همانطور که قبلاً در مورد شبکه‌های بی مقیاس ارائه شده است [۲۸]، به عنوان کار آینده می‌تواند در مورد بدافزارهای شبکه تلفن همراه و در سیستم عامل اندروید مورد بررسی قرار گیرد. این مدل سازی می‌تواند در ارائه راهکارهای دفاع در مقابل انتشار اینگونه بدافزارها، مؤثر واقع شود.

۷- مراجع

[1] C. Wilson, "Botnets, cybercrime, and cyberterrorism: Vulnerabilities and policy issues for congress," Washington, DC, 2008.

[2] C. S. Gates, J. Chen, N. Li, and R. W. Proctor, "Effective risk communication for android apps," Dependable and Secure Computing, IEEE Transactions on, vol. 11, no. 3, 2014, pp. 252-265.

[3] C. S. Gates, N. Li, H. Peng, B. Sarma, Y. Qi, R. Potharaju, and I. Molloy, "Generating summary risk scores for mobile applications," Dependable and Secure Computing, IEEE Transactions on, vol. 11, no. 3, pp. 238-251, 2014.

[4] E. Chin, A. P. Felt, V. Sekar, and D. Wagner, "Measuring user confidence in smartphone security and privacy," In Proceedings of the Eighth Symposium on Usable Privacy and Security, ACM, July 2012.

[5] A. P. Felt, K. Greenwood, and D. Wagner, "The effectiveness of application permissions," In Proceedings of the 2nd USENIX conference on Web application development, p. 7, June 2011.

[6] A. P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, and D. Wagner, "Android permissions: User attention, comprehension, and behavior," Tech. Rep. UCB/EECS-2012-26, UC Berkeley, 2012.

[7] P. G. Kelley, S. Consolvo, L. F. Cranor, J. Jung, N. Sadeh, and D. Wetherall, "A conundrum of permissions: installing applications on an android smartphone," In Financial Cryptography and Data Security, Springer Berlin Heidelberg, pp. 68-79, 2012.

[8] P. G. Kelley, L. F. Cranor, and N. Sadeh, "Privacy as part of the app decision-making process," In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, pp. 3393-3402, April 2013.

[9] H. Peng, C. Gates, B. Sarma, N. Li, Y. Qi, R. Potharaju, and I. Molloy, "Using probabilistic generative models for ranking risks of android apps," In Proceedings of the 2012 ACM conference on Computer and communications security, ACM, pp. 241-252, October 2012.

[10] D. Geneiatakis, I. N. Fovino, I. Kounelis, and P. Stirparo, "A Permission verification approach for android mobile applications," Computers & Security, vo49, pp.192-205, 2015.

[11] B. P. Sarma, N. Li, C. Gates, R. Potharaju, C. Nita-Rotaru, and I. Molloy, "Android permissions: a perspective combining risks and benefits," In Proceedings of the 17th ACM symposium on Access Control Models and Technologies, June 2012, pp. 13-22.

[12] L. Cen, C. Gates, L.Si, and N. Li, "A probabilistic discriminative model for android malware detection with decompiled source code," In Dependable and Secure Computing, IEEE Transactions on, vol. 12, no. 4, 2015, pp. 400-412.

[13] A. Desnos, "Android: Static analysis using similarity distance," In System Science (HICSS), 2012 45th Hawaii International Conference on, January 2012, pp. 5394-5403.

[14] A. D. Schmidt, R. Bye, H. G. Schmidt, J. Clausen, O. Kiraz, K. Yüksel, and S. Albayrak, "Static analysis of executables

for collaborative malware detection on android," In Communications, 2009. ICC'09. IEEE International Conference on, June 2009, pp. 1-5.

[15] Y. Zhou, Z. Wang, W. Zhou, and X. Jiang, "Hey, You, Get Off of My Market: Detecting Malicious Apps in Official and Alternative Android Markets," In NDSS, vol. 25, no. 4, pp. 50-52, February 2012.

[16] Y. Aafer, W. Du, and H. Yin, "Droid API Miner: Mining API-level features for robust malware detection in android," In Security and Privacy in Communication Networks, pp. 86-103, 2013.

[17] M. Christodorescu, S. Jha, and C. Kruegel, "Mining specifications of malicious behavior," In Proceedings of the 1st India software engineering conference, ACM, pp. 5-14, February 2008.

[18] K. Rieck, T. Holz, C. Willems, P. Düssel, and P. Laskov, "Learning and classification of malware behavior," In Detection of Intrusions and Malware, and Vulnerability Assessment, pp. 108-125, 2008.

[19] A. Shabtai and Y. Elovici, "Applying behavioral detection on android-based devices," In Mobile Wireless Middleware, Operating Systems and Applications, pp. 235-249, 2010.

[20] I. Burguera, U. Zurutuza, and S. Nadjm-Tehrani, "Crowdroid: behavior-based malware detection system for android," In Proceedings of the 1st ACM workshop on Security and privacy in smartphones and mobile devices, pp. 15-26, October 2011.

[21] Y. Zhou, and X. Jiang, "Dissecting android malware: Characterization and evolution", In Security and Privacy (SP), 2012 IEEE Symposium on, May 2012, pp. 95-109.

[22] D. Barrera, H. G. Kayacik, P. C. van Oorschot, and A. Somayaji, "A methodology for empirical analysis of permission-based security models and its application to android," In Proceedings of the 17th ACM conference on Computer and communications security, pp. 73-84, October 2010.

[23] W. Enck, D. Octeau, P. McDaniel, and S. Chaudhuri, "A Study of Android Application Security," In USENIX security symposium, vol. 2, p. 2, August 2011.

[24] W. Enck, M. Ongtang, and P. McDaniel, "On lightweight mobile phone application certification," In Proceedings of the 16th ACM conference on Computer and communications security, pp. 235-245, November 2009.

[25] S. Chakradeo, B. Reaves, P. Traynor, and W. Enck, "Mast: triage for market-scale mobile malware analysis," In Proceedings of the sixth ACM conference on Security and privacy in wireless and mobile networks, pp. 13-24, April 2013.

[26] K. W. Y. Au, Y. F. Zhou, Z. Huang, and D. Lie, "Pscout: analyzing the android permission specification," In Proceedings of the 2012 ACM conference on Computer and communications security, October 2012, pp. 217-228.

[27] R. Quinlan, "Learning efficient classification procedures," Machine Learning: an artificial intelligence approach, Michalski, Carbonell & Mitchell (eds.), Morgan Kaufmann, pp. 463-482, 1983.

[28] S. Koochaki and M. Abdollahi Azgomi, "A Method for Fluid Modeling of the Propagation Behavior of Malware In Scale-Free Networks," Journal of Electronical & Cyber Defence, vol. 4, no. 4, pp. 1-10, 2017 (In Persian).

Journal Of Electronical & Cyber Defence
Vol. 5, No. 1, 2017, Serial No. 17

Estimating Security Risks of Android Apps Using Information Gain

M. Deypir*

*Shahid Sattari University of Science and Technology

(Received: 27/01/2016, Accepted: 31/10/2016)

ABSTRACT

With the rapid growth of developing malwares in Android platform as the widest used mobile operating system, knowing security risk of an application (app) can be helpful for warning users regarding the use of potential malicious applications. The security risk of an Android app can be estimated using its requested permissions. In this paper, the concept of critical permissions is precisely re-defined according to the abuse of permissions by previously known Android malwares. Based on this definition and analysis of requested permissions of the large numbers of malwares and benign apps, a new criterion is proposed to measure the security risk of the apps. In this criterion, informative permissions have higher impact on the resulting measured security risk values of the apps. Experimental evaluations show the superiority of the proposed criterion with respect to previously proposed ones in terms of detection rete and generalization capability.

Keywords: Security risk, Android, Malwares, Information Gain, Critical Permissions

* Corresponding Author Email: mdeypir@ssau.ac.ir