شماره:
تاریخ درخواست:
پیوست:

این قسمت توسط بخش پر میشود

دانشکده: فنی و مهندسی

بخش/گروه:

بسمه تعالی

طرح تحقیق پایان نامه کارشناسی ارشد/ دکتری عمومی

1- مشخصات دانشجو		
نام: علیرضا	نام خانوادگی: ایرانمنش	شماره دانشجویی: 401155015
دانشکده: فنی و مهندسی	رشته: مهندسی کامپیوتر	گرایش: هوش مصنوعی و رباتیکز
سهمیه قبولی: آزاد	تعداد واحدهای گذرانده شده: 24	معدل: 15.63
پست الکترونیک: alirezairanmanesh78@gmail.com		تلفن: 09031310687
آدرس: شهرک ولیعصر، ولیعصر ۲، پلاک ۱۴۹		

2- مشخصات استاد (اساتید) راهنما			
استاد راهنمای اول	نام: حمید	نام خانوادگی: میروزیری	تخصص اصلی: امنیت و هوش مصنوعی
	رتبه دانشگاهی: استادیار	سنوات تدریس در کارشناسی ارشد:	سنوات تدریس در دوره دکتری:
	پست الکترونیک: h.mirvaziri@gmail.com		تلفن:31323254
استاد راهنمای دوم	نام:	نام خانوادگی:	تخصص اصلی:
	آخرین مدرک تحصیلی:	سال اخذ:	رتبه دانشگاهی:
	سنوات تدریس در کارشناسی ارشد:	سنوات تدریس در دوره دکتری:	
	پست الکترونیک:		تلفن:

3- مشخصات استاد (اساتید) مشاور			
استاد مشاور اول	نام:	نام خانوادگی:	تخصص اصلی:
	آخرین مدرک تحصیلی:	سال اخذ:	رتبه دانشگاهی:
	سنوات تدریس در کارشناسی ارشد:	سنوات تدریس در دوره دکتری:	
	پست الکترونیک:		تلفن:
استاد مشاور دوم	نام:	نام خانوادگی:	تخصص اصلی:
	آخرین مدرک تحصیلی:	سال اخذ:	رتبه دانشگاهی:
	سنوات تدریس در کارشناسی ارشد:	سنوات تدریس در دوره دکتری:	
	پست الکترونیک:		تلفن:

4- عنوان پایان‌نامه		
فارسی	تشخیص قدرتمند بدافزارهای اندروید با استفاده از شبکه‌های عصبی ترنسفورمر	
لاتین	Robust Android Malware Detection using Transformer Neural Networks	
واژگان کلیدی	فارسی	بدافزار اندروید، ترنسفورمرها، تشخیص بدافزار، شبکه‌های عصبی
	لاتین	Robust Android Malware Detection using Transformer Neural Networks
نوع تحقیق	کاربردی ☒	کاربردی بنیادی ☐ بنیادی ☐
تعداد واحد پایان نامه: 6 واحد		مدت اجراء: 12 ماه

5- اطلاعات پایان‌نامه

چکیده

گسترش بدافزارهای اندرویدی، تهدید جدی برای امنیت دستگاه‌های تلفن همراه محسوب می‌شود. روش‌های سنتی تشخیص بدافزار مبتنی بر امضا، در برابر انواع پیچیده و در حال تکامل سریع بدافزارها ناکارآمد هستند. این پیشنهاد، رویکرد نوآورانه‌ای را برای طبقه‌بندی بدافزارهای اندرویدی با بهره‌گیری از شبکه‌های عصبی ترنسفورمر ارائه می‌کند. ترنسفورمرها با مکانیزم توجه متقابل، قادر به درک روابط پیچیده در داده‌های توالی نظیر نمودارهای فراخوانی تابع هستند. در این رویکرد، ویژگی‌های غنی را از کد برنامه‌ها شامل فایل‌های classes.dex و lib.so استخراج می‌نماییم. این ویژگی‌ها را به معماری ترنسفورمر میدهیم تا برنامه‌ها را به مخرب یا سالم طبقه‌بندی کند. انتظار می‌رود این روش، نسبت به روش‌های سنتی و شبکه‌های کانولوشنال، دقت بیشتری در تشخیص بدافزارها داشته باشد. نتایج این پروژه، به پیشرفت تشخیص بدافزارهای اندرویدی و ایجاد اکوسیستم موبایل امن‌تر کمک خواهد کرد.

Abstract:

The proliferation of Android malware poses a significant threat to mobile device security. Traditional signature-based malware detection methods are ineffective against sophisticated and rapidly evolving malware variants. This proposal presents an innovative approach to Android malware classification by leveraging transformer neural networks. Transformers, with their self-attention mechanisms, are capable of capturing complex relationships in sequential data, such as call graphs. The proposed approach extracts rich features from application code, including classes.dex and lib.so files. These features are fed into a transformer architecture to classify applications as malicious or benign. The proposed method is expected to achieve higher accuracy in malware detection compared to traditional methods and convolutional neural networks. The project's outcomes will contribute to advancing Android malware detection and enabling a more secure mobile ecosystem.

مقدمه

امروزه دستگاه‌های تلفن همراه و کاربردهای متنوع آن‌ها، بخش جدایی‌ناپذیری از زندگی روزمره ما شده‌اند. با گسترش استفاده از این دستگاه‌ها و افزایش تعداد کاربران سیستم‌عامل اندروید، امنیت این پلتفرم به یکی از نگرانی‌های اصلی در حوزه فناوری اطلاعات تبدیل شده است. متأسفانه همگام با رشد محبوبیت اندروید، حملات و بدافزارهای هدفمند برای این سیستم‌عامل نیز افزایش چشمگیری داشته‌اند.

بدافزارهای اندرویدی می‌توانند از طریق نفوذ به دستگاه‌ها و سوءاستفاده از آسیب‌پذیری‌های امنیتی، به داده‌های حساس کاربران دسترسی پیدا کرده و خسارت‌های مالی، امنیتی و معنوی جدی را به بار آورند. بنابراین، توسعه روش‌های قدرتمند برای تشخیص به‌موقع و دقیق این تهدیدات امری ضروری است.

روش‌های سنتی مبتنی بر امضا که در بسیاری از آنتی‌ویروس‌ها و سیستم‌های امنیتی به کار گرفته می‌شوند، دیگر کارایی لازم را در برابر انواع پیچیده و در حال تکامل سریع بدافزارهای اندرویدی ندارند. مهاجمان و بدافزارنویسان از تکنیک‌های مبهم‌سازی، فشرده‌سازی و فریب پیشرفته برای دور زدن این سیستم‌ها استفاده می‌کنند. لذا نیاز مبرم به روش‌های نوآورانه و قدرتمندتر برای تشخیص و طبقه‌بندی دقیق بدافزارها احساس می‌شود.

این پژوهش هدف ارائه رویکردی جدید برای طبقه‌بندی بدافزارهای اندرویدی با بهره‌گیری از شبکه‌های عصبی ترنسفورمر را دنبال می‌کند. فرضیه اصلی این است که ترنسفورمرها با مکانیزم توجه متقابل، قادر به درک روابط پیچیده در داده‌های توالی مانند نمودارهای فراخوانی تابع و الگوهای مخرب موجود در کدهای برنامه هستند و می‌توانند این الگوها را بهتر از روش‌های سنتی شناسایی کنند.

نوآوری این پژوهش در به‌کارگیری ترنسفورمرها برای تشخیص بدافزارهای اندرویدی و استخراج ویژگی‌های غنی از کد برنامه‌ها شامل فایل‌های classes.dex و lib.so نهفته است. به این صورت است که ما ویژگی های هر برنامه را با استفاده از این فایل ها بدست میاریم و طبق آنها با ترنسفورمرها شروع به دسته بندی بدافزارها میکنیم. انتظار می‌رود رویکرد پیشنهادی، با توجه به قدرت مدل‌سازی پیچیدگی‌ها در ترنسفورمرها، نسبت به روش‌های سنتی و شبکه‌های کانولوشنال، عملکرد بهتر و دقت بالاتری در تشخیص بدافزارها داشته باشد. دستیابی به چنین سیستم امنیتی کارآمد، می‌تواند به ارتقای سطح امنیت در اکوسیستم موبایل و حفاظت از حریم خصوصی و داده‌های کاربران در برابر تهدیدات بدافزار کمک شایانی نماید.

پیشینه پژوهش

موضوع تشخیص و طبقه‌بندی بدافزارهای اندرویدی از مباحث بسیار مهم و چالش‌برانگیز در حوزه امنیت سایبری و محافظت از دستگاه‌های تلفن همراه است. پژوهشگران متعددی در این زمینه مطالعات گسترده‌ای انجام داده‌اند و روش‌های مختلفی را پیشنهاد کرده‌اند.

در ابتدا، روش‌های مبتنی بر امضا برای تشخیص بدافزارها توسعه یافتند. در این روش‌ها، امضای منحصربه‌فرد هر بدافزار شناسایی و در پایگاه داده‌ای ذخیره می‌شد. سپس برنامه‌ها بر اساس تطبیق با این امضاها به عنوان مخرب یا سالم طبقه‌بندی می‌شدند. اگرچه این روش‌ها برای بدافزارهای شناخته شده مؤثر بودند، اما در برابر بدافزارهای جدید و تغییریافته ناتوان هستند.

در ادامه، روش‌های مبتنی بر رفتار ارائه شدند که در آن‌ها رفتار یک برنامه در حین اجرا رصد شده و با الگوهای رفتاری مخرب مقایسه می‌گردید. اگرچه این روش‌ها بهبودی نسبت به روش‌های مبتنی بر امضا داشتند، اما هنوز دارای محدودیت‌هایی بودند؛ از جمله نیاز به منابع محاسباتی زیاد، آسیب‌پذیری در برابر تکنیک‌های فریب و مبهم‌سازی، و عدم توانایی کافی در تشخیص بدافزارهای پیچیده‌ای که رفتار مخرب آن‌ها به سختی قابل تشخیص است.

با پیشرفت در زمینه هوش مصنوعی، پژوهشگران به سمت روش‌های مبتنی بر یادگیری ماشین روی آوردند[1]. در این زمینه، مطالعات متعددی انجام شد که از تکنیک‌های مختلف یادگیری ماشین برای استخراج ویژگی و طبقه‌بندی بدافزارها استفاده کردند. به عنوان مثال، در این زمینه میتوان مقاله زیر اشاره کرد که از آنتروپی برای مقایسه و رمزگشایی فایل‌های باینری استفاده کرده اند[2]. همچنین Tian و Versteeg از تابع طول برای استخراج ویژگی از کدها و الگوریتم رگرسیون لجستیک برای طبقه‌بندی بهره گرفتند و دقت بالای 90 درصد را گزارش کردند[3]

در سال‌های اخیر، با پیشرفت‌های حاصل شده در زمینه یادگیری عمیق، تمرکز پژوهش‌ها به سمت استفاده از شبکه‌های عصبی کانولوشنی (CNN) و شبکه‌های عصبی بازگشتی (RNN) معطوف شده است. مطالعات Nix و Zhang نشان داد که CNNها می‌توانند با استخراج ویژگی‌های API از کدهای برنامه، عملکرد خوبی در طبقه‌بندی بدافزارها داشته باشند [4]. همچنین Zhu و Zhang با استفاده از گراف‌های کانولوشنی، یک روش بر پایه گراف فراخوانی تابع برای تشخیص بدافزارها ارائه دادند[5].

این روش‌ها به استفاده از داده‌های آموزشی بزرگ، به آموزش مدل‌هایی میپردازند که می‌توانند الگوهای پنهان در ویژگی‌های استخراج شده از برنامه‌ها را کشف کنند. شبکه‌های عصبی کانولوشنی یکی از معروف‌ترین روش‌های این دسته هستند که در تشخیص بدافزار اندرویدی به کار گرفته شده‌اند[6] . این مدل‌ها قادرند ویژگی‌های لوکال نظیر کد باینری برنامه‌ها را به خوبی استخراج کنند. اگرچه این روش‌ها نسبت به روش‌های قدیمی‌تر پیشرفت چشمگیری داشته‌اند، اما همچنان محدودیت‌هایی دارند؛ مانند نیاز به پیش‌پردازش و استخراج ویژگی دستی، عدم توانایی کافی در مدل‌سازی روابط پیچیده در داده‌های توالی نظیر نمودارهای فراخوانی تابع، و اشباع در دقت بالا برای برخی از مجموعه داده‌ها.

در همین راستا، برخی پژوهش‌های اخیر از شبکه‌های عصبی پیچشی (Recurrent) مانند LSTM و GRU برای تشخیص بدافزار اندرویدی استفاده کرده‌اند. این مدل‌ها قادرند روابط سری را در داده‌های توالی مدل کنند و بر روی ویژگی‌هایی مانند نمودارهای فراخوانی تابع عملکرد بهتری نسبت به شبکه‌های کانولوشنی داشته باشند. با این حال، همچنان دارای محدودیت‌هایی در مقیاس‌پذیری و پیچیدگی محاسباتی برای توالی‌های بلند هستند و ممکن است در برابر تغییرات جزئی در توالی‌ها آسیب‌پذیر باشند.

در سال‌های اخیر، مدل‌های ترنسفورمر که بر پایه مکانیزم توجه متقابل استوار هستند، در حوزه‌های مختلف پردازش زبان طبیعی و پردازش سیگنال عملکرد چشمگیری از خود نشان داده‌اند. ترنسفورمرها قادرند روابط پیچیده در داده‌های توالی را به خوبی مدل کنند، از عملکرد و مقیاس‌پذیری بالایی برخوردار هستند و در برابر تغییرات جزئی در توالی‌ها مقاوم‌تر می‌باشند. علاوه بر این، ترنسفورمرها نیاز به پیش‌پردازش کمتری دارند و می‌توانند ویژگی‌های مناسب را به طور مستقیم از داده‌های خام استخراج کنند[7]

با این حال، به کارگیری ترنسفورمرها در زمینه تشخیص بدافزار اندرویدی هنوز در مراحل اولیه قرار دارد و نیاز به مطالعات و پژوهش‌های بیشتری وجود دارد. برخی چالش‌های اصلی در این زمینه عبارتند از: انتخاب معماری ترنسفورمر مناسب، تعیین روش مناسب برای تبدیل داده‌های برنامه به نمایش توالی، بهینه‌سازی مدل ترنسفورمر برای تشخیص بدافزار، و ارزیابی جامع و مقایسه با روش‌های موجود.

همچنین باید توجه داشت که بدافزارنویسان نیز همواره در حال به‌روزرسانی و پیچیده‌تر کردن تکنیک‌های خود هستند. بنابراین، تحقیق در این زمینه یک فرآیند مداوم و پویاست که نیازمند نوآوری و سازگاری مستمر با تهدیدات در حال تکامل است.

روش انجام پژوهش

در این پژوهش از مدل‌های ترنسفورمر برای طبقه‌بندی برنامه‌های اندرویدی به مخرب یا سالم استفاده می‌شود. داده‌های مورد نیاز شامل دو دسته کلی فایل‌های classes.dex از کدهای جاوا و فایل‌های lib.so از کتابخانه‌های برنامه‌نویسی رشته‌ای (NDK) هستند.

برای جمع‌آوری داده‌ها، از پایگاه داده‌های موجود DREBIN برای فایل‌های classes.dex و نیز مجموعه داده‌ای جدیدی متشکل از حدود 250 برنامه اندرویدی حاوی فایل lib.so که با ابزار VirusTotal برچسب‌گذاری شده‌اند، استفاده می‌گردد.

پس از جمع‌آوری داده‌ها، ویژگی‌های غنی مانند نوع توابع، درخواست‌های دسترسی سخت‌افزاری، مجوزها و فیلترهای Intent از آن‌ها استخراج می‌شود. سپس این ویژگی‌ها به مدل ترنسفورمر داده شده و فرآیند آموزش انجام می‌گیرد.

مدل ترنسفورمر شامل لایه‌های کدگذار، دی‌کدگذار و مکانیزم توجه متقابل است که توانایی مدل‌سازی روابط پیچیده در داده‌های توالی را دارد. تابع زیان دوتایی و الگوریتم بهینه‌ساز آدام برای آموزش مدل استفاده خواهند شد.

انتظار می‌رود این روش با مدل‌سازی موثر ویژگی‌های استخراج شده و روابط میان آن‌ها، بتواند برنامه‌های اندروید را با دقت بالاتری نسبت به روش‌های سنتی و شبکه‌های کانولوشنی به دو دسته مخرب و سالم تقسیم‌بندی نماید.

یافته‌های مورد انتظار این است که معماری ترنسفورمر با توجه به قدرت مدلسازی پیچیدگی‌ها، بتواند دقت بالاتری در تشخیص بدافزارهای اندرویدی نسبت به روش‌های موجود ارائه دهد.

منابع :

[1] Y. Lecun, Y. Bengio, and G. Hinton, ‘Deep learning’, *Nature*, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, May 27, 2015. doi: 10.1038/nature14539.

[2] S. Cesare and Y. Xiang, ‘Classification of Malware Using Structured Control Flow’, 2010.

[3] Institute of Electrical and Electronics Engineers., *3rd International Conference on Malicious and Unwanted Software : Malware ’08 to be held at the Hilton Alexandria Mark Center, Alexandria, VA, USA, October 7-8, 2008.* IEEE, 2008.

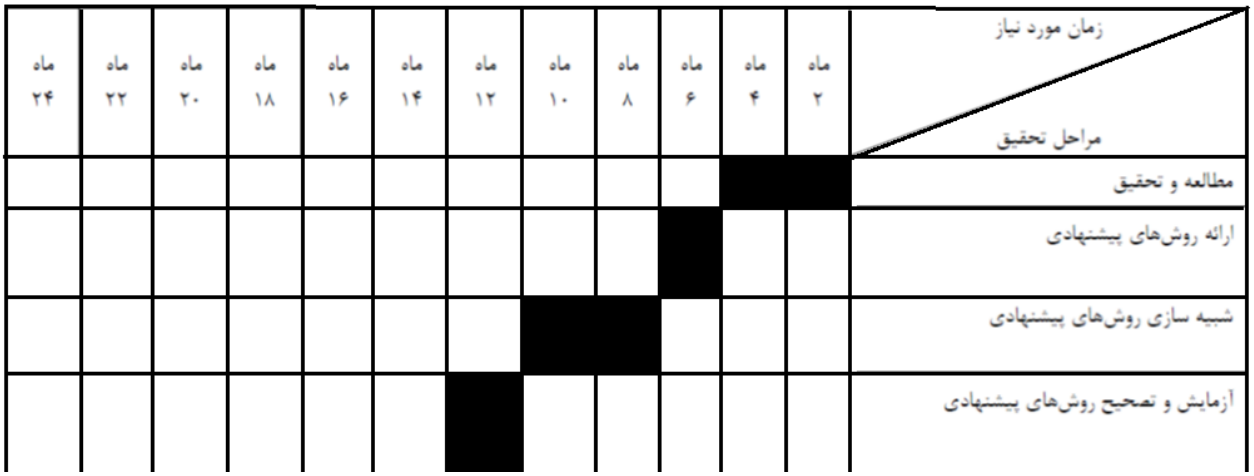
[4] IEEE Computational Intelligence Society, International Neural Network Society, and Institute of Electrical and Electronics Engineers, *IJCNN 2017 : the International Joint Conference on Neural Networks.*

[5] S. Haque, Z. Eberhart, A. Bansal, and C. McMillan, ‘Semantic Similarity Metrics for Evaluating Source Code Summarization’, in *IEEE International Conference on Program Comprehension*, IEEE Computer Society, 2022, pp. 36–47. doi: 10.1145/nnnnnnn.nnnnnnn.

[6] A. Vaswani *et al.*, ‘Attention Is All You Need’.

[7] D.-L. Vu, T.-K. Nguyen, T. V. Nguyen, T. N. Nguyen, F. Massacci, and P. H. Phung, ‘A Convolutional Transformation Network for Malware Classification’, Sep. 2019, [Online]. Available: http://arxiv.org/abs/1909.07227

6- جدول زمان بندی و مراحل انجام تحقیق (از زمان تصویب تا دفاع نهایی) : الگوی مشترک



7- هزینه های تقریبی پایان نامه*

الف – آیا برای این طرح از سازمان های دیگر تامین اعتبار شده است؟ بلی ☐ خیر ☒

در صورت مثبت بودن ،تاریخ تصویب ،میزان اعتبار و نام سازمان را مشخص نمایید.

نام سازمان: تاریخ تصویب : میزان اعتبار :

ب- آیا پایان نامه بخشی از یک طرح تحقیقاتی استادان دانشگاهی می باشد؟ بلی ☐ خیر ☒

در صورت مثبت بودن پاسخ ، عنوان اصلی طرح تحقیقاتی :

نام مجری : دانشکده : تاریخ تصویب نهایی طرح : اعتبار مصوب :

توضیحات:

7-1- هزینه های مواد و وسایل (وسایلی که صرفا از محل اعتبار طرح تحقیق باید خریداری شود)

ردیف	نام مواد و وسایل	مقدار یا تعداد مورد نیاز	مصرفی	غیر مصرفی	ساخت داخل یا خارج	شرکت سازنده	قیمت واحد(ریال)	قیمت کل (ریال)
جمع کل به ریال								

7-2- هزینه های پرسنلی (برای مواردی که در حوزه تخصص و مهارت دانشجو قرار ندارد)

نوع مسولیت	تعداد پرسنل	کل ساعات کار	حق الزحمه در ساعت	جمع به ریال
جمع کل به ریال				

7-3- هزینه های مسافرت

مقصد	تعداد سفر	نوع وسیله نقلیه	هزینه هر سفر	جمع به ریال
جمع کل به ریال				

7-4- جمع کل هزینه ها

ردیف	نوع هزینه	مبلغ(ریال)
1	مواد و وسایل	
2	پرسنلی	
3	مسافرت	
4	متفرقه (تایپ ، تکثیر و تهیه کتاب)	
5	جمع کل	

*هزینه های انجام پایان نامه بر اساس بودجه ومصوبات دانشگاه و در چارچوب قوانین دانشگاه قابل پرداخت خواهد بود.

اینجانب علیرضا ابراهیمی متعهد می شوم که در مدت اجرای پایان نامه دوره کارشناسی ارشد به طور تمام وقت انجام وظیفه نموده و همچنین اطلاع دارم که کلیه نتایج و حقوق حاصله از این پایان نامه متعلق به بخش مهندسی عامپوتر دانشکده فنی مهندسی دانشگاه شهید باهنر کرمان بوده و مجاز نیستم بدون موافقت استاد راهنما اطلاعاتی را در رابطه با پایان نامه به دیگری واگذار نمایم.

امضاء دانشجو:

تائید دانشجو، اساتید راهنما و مشاور		
نام و نام خانوادگی دانشجو: علیرضا ابراهیمی	امضاء	تاریخ :
نام و نام خانوادگی استاد راهنما اول: حمیدرضا پورمیری	امضاء	تاریخ :
نام و نام خانوادگی استاد راهنما دوم:	امضاء	تاریخ :
نام و نام خانوادگی استاد مشاور ۱:	امضاء	تاریخ :
نام و نام خانوادگی استاد مشاور ۲:	امضاء	تاریخ :
اساتید حاضر در جلسه ارائه موضوع تحقیق پایان نامه در تاریخ ۱۴۰۳/۴/۸		
نام و نام خانوادگی: مهدی انتظاری	امضاء	نوع رای: مطلوب
نام و نام خانوادگی: عباس بحرالعلوم	امضاء	نوع رای: موافق
نام و نام خانوادگی: حمیدرضا پورمیری	امضاء	نوع رای: موافق
نام و نام خانوادگی:	امضاء	نوع رای:
نام و نام خانوادگی:	امضاء	نوع رای:
نام و نام خانوادگی:	امضاء	نوع رای:
تائیدیه نهایی موضوع تحقیق پایان نامه		
نام و نام خانوادگی رئیس بخش: وحید محمدی	امضاء	تاریخ صورتجلسه بخش: ۱۴۰۳/۴/۱۰
نام و نام خانوادگی معاون آموزشی و پژوهشی دانشکده:	امضاء	تاریخ صورتجلسه دانشکده: