cases

doc_1		doc_2		decision	id
authors	• Kaleem Nawaz Khan • Najeeb Ullah • Sikandar Ali • Muhammad Salman Khan • Mohammad Nauman • Anwar Ghani	**authors**	• Kaleem Nawaz Khan • Najeeb Ullah • Sikandar Ali • Muhammad Salman Khan • Mohammad Nauman • Anwar Ghani • Shah Nazir	DUPLICATES	100
title	Op2Vec: An Opcode Embedding Technique and Dataset Design for End-to-End Detection of Android Malware	**title**	Op2Vec: An Opcode Embedding Technique and Dataset Design for End-to-End Detection of Android Malware		
publication_date	2022-03-01 00:00:00	**publication_date**	2022-05-19 00:00:00		
source	SupportedSources.INTERNET_ARCHIVE	**source**	SupportedSources.INTERNET_ARCHIVE		
journal		**journal**	Hindawi Limited		
volume		**volume**			
doi		**doi**	10.1155/2022/3710968		
urls	• https://web.archive.org/web/20220530220240/https://arxiv.org/pdf/2104.04798v2.pdf	**urls**	• https://web.archive.org/web/20220524042243/https://downloads.hindawi.com/journals/scn/2022/3710968.pdf		
id	id5546588454221539160	**id**	id-4400818897865448186		
abstract	Android is one of the leading operating systems for smart phones in terms of market share and usage. Unfortunately, it is also an appealing target for attackers to compromise its security through malicious applications. To tackle this issue, domain experts and researchers are trying different techniques to stop such attacks. All the attempts of securing Android platform are somewhat successful. However, existing detection techniques have severe shortcomings, including the cumbersome process of feature engineering. Designing representative features require expert domain knowledge. There is a need for minimizing human experts' intervention by circumventing handcrafted feature engineering. Deep learning could be exploited by extracting deep features automatically. Previous work has shown that operational codes (opcodes) of executables provide key information to be used with deep learning models for detection process of malicious applications. The only challenge is to feed opcodes information to deep learning models. Existing techniques use one-hot encoding to tackle the challenge. However, the one-hot encoding scheme has severe limitations. In this paper, we introduce; (1) a novel technique for opcodes embedding, which we name Op2Vec, (2) based on the learned Op2Vec we have developed a dataset for end-to-end detection of android malware. Introducing the end-to-end Android malware detection technique avoids expert-intensive handcrafted features extraction, and ensures automation. Some of the recent deep learning-based techniques showed significantly improved results when tested with the proposed approach and achieved an average detection accuracy of 97.47%, precision of 0.976 and F1 score of 0.979.	**abstract**	Android is one of the leading operating systems for smartphones in terms of market share and usage. Unfortunately, it is also an appealing target for attackers to compromise its security through malicious applications. To tackle this issue, domain experts and researchers are trying different techniques to stop such attacks. All the attempts of securing the Android platform are somewhat successful. However, existing detection techniques have severe shortcomings, including the cumbersome process of feature engineering. Designing representative features require expert domain knowledge. There is a need for minimizing human experts' intervention by circumventing handcrafted feature engineering. Deep learning could be exploited by extracting deep features automatically. Previous work has shown that operational codes (opcodes) of executables provide key information to be used with deep learning models for the detection process of malicious applications. The only challenge is to feed opcodes information to deep learning models. Existing techniques use one-hot encoding to tackle the challenge. However, the one-hot encoding scheme has severe limitations. In this paper, we introduce (1) a novel technique for opcodes embedding, which we name Op2Vec, and (2) based on the learned Op2Vec, we have developed a dataset for end-to-end detection of Android malware. Introducing the end-to-end Android malware detection technique avoids expert-intensive handcrafted feature extraction and ensures automation. Some of the recent deep learning-based techniques showed significantly improved results when tested with the proposed approach and achieved an average detection accuracy of 97.47%, precision of 0.976, and F1 score of 0.979.		
versions		**versions**			