cases

doc_1		doc_2		decision	id
authors	• Joshua Saxe • Konstantin Berlin	**authors**	• Joshua Saxe • Konstantin Berlin	DUPLICATES	77
title	Deep neural network based malware detection using two dimensional binary program features	**title**	Deep Neural Network Based Malware Detection Using Two Dimensional Binary Program Features		
publication_date	2015-08-13 00:00:00	**publication_date**	2015-08-13 01:22:13+00:00		
source	SupportedSources.SEMANTIC_SCHOLAR	**source**	SupportedSources.ARXIV		
journal		**journal**	None		
volume		**volume**			
doi	10.1109/MALWARE.2015.7413680	**doi**			
urls	• https://www.semanticscholar.org/paper/1eb6449ae040f051120e4d44348a0f68af9c36e8	**urls**	• http://arxiv.org/pdf/1508.03096v2 • http://arxiv.org/abs/1508.03096v2 • http://arxiv.org/pdf/1508.03096v2		
id	id-531791081912409734	**id**	id300024926243860313		
abstract	In this paper we introduce a deep neural network based malware detection system that Invincea has developed, which achieves a usable detection rate at an extremely low false positive rate and scales to real world training example volumes on commodity hardware. We show that our system achieves a 95% detection rate at 0.1% false positive rate (FPR), based on more than 400,000 software binaries sourced directly from our customers and internal malware databases. In addition, we describe a non-parametric method for adjusting the classifier's scores to better represent expected precision in the deployment environment. Our results demonstrate that it is now feasible to quickly train and deploy a low resource, highly accurate machine learning classification model, with false positive rates that approach traditional labor intensive expert rule based malware detection, while also detecting previously unseen malware missed by these traditional approaches. Since machine learning models tend to improve with larger datasizes, we foresee deep neural network classification models gaining in importance as part of a layered network defense strategy in coming years.	**abstract**	Malware remains a serious problem for corporations, government agencies, and individuals, as attackers continue to use it as a tool to effect frequent and costly network intrusions. Machine learning holds the promise of automating the work required to detect newly discovered malware families, and could potentially learn generalizations about malware and benign software that support the detection of entirely new, unknown malware families. Unfortunately, few proposed machine learning based malware detection methods have achieved the low false positive rates required to deliver deployable detectors. In this paper we a deep neural network malware classifier that achieves a usable detection rate at an extremely low false positive rate and scales to real world training example volumes on commodity hardware. Specifically, we show that our system achieves a 95% detection rate at 0.1% false positive rate (FPR), based on more than 400,000 software binaries sourced directly from our customers and internal malware databases. We achieve these results by directly learning on all binaries, without any filtering, unpacking, or manually separating binary files into categories. Further, we confirm our false positive rates directly on a live stream of files coming in from Invincea's deployed endpoint solution, provide an estimate of how many new binary files we expected to see a day on an enterprise network, and describe how that relates to the false positive rate and translates into an intuitive threat score. Our results demonstrate that it is now feasible to quickly train and deploy a low resource, highly accurate machine learning classification model, with false positive rates that approach traditional labor intensive signature based methods, while also detecting previously unseen malware.		
versions		**versions**			