Digital Investigation xxx (2018) 1–9

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

DFRWS 2018 USA — Proceedings of the Eighteenth Annual DFRWS USA

Deep learning at the shallow end: Malware classification for non-domain experts

Quan Le [a, *], Oisín Boydell [a], Brian Mac Namee [a], Mark Scanlon [b]

[a] Centre for Applied Data Analytics Research, University College Dublin, Ireland
[b] Forensics and Security Research Group, University College Dublin, Ireland

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Deep learning
Machine learning
Malware analysis
Reverse engineering

ABSTRACT

Current malware detection and classification approaches generally rely on time consuming and knowledge intensive processes to extract patterns (*signatures*) and behaviors from malware, which are then used for identification. Moreover, these signatures are often limited to local, contiguous sequences within the data whilst ignoring their context in relation to each other and throughout the malware file as a whole. We present a Deep Learning based malware classification approach that requires no expert domain knowledge and is based on a purely data driven approach for complex pattern and feature identification.
© 2018 The Author(s). Published by Elsevier Ltd on behalf of DFRWS. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

In law enforcement agencies throughout the world, there are growing digital forensic backlogs of unimaged, unprocessed, and unanalyzed digital devices stored in evidence lockers (Scanlon, 2016). This growth is attributable to several compounding factors. The sheer volume of cases requiring digital forensic processing extends far beyond digitally executed crimes such as phishing, online sharing of illicit content, online credit card fraud, etc., to "traditional" crimes such as murder, stalking, financial fraud, etc. The volume of data to be analyzed per case is continuously growing and there is a limited supply of trained personnel capable of the expert, court-admissible, reproducible analysis that digital forensic processing requires.

In order to address the latter factor, many police forces have been implementing a first responder/triage model to enable on-site evidence seizure and securing the integrity of the evidence gathered (Hitchcock et al., 2016). These models train field officers in the proficient handling of digital devices at a crime scene enabling the available expert digital investigators to remain in the laboratory processing cases. In this model, the first responders are not trained in the analysis or investigation phase of the case, but can ensure the integrity and court-admissibility of the gathered evidence.

While physical resourcing in terms of hardware, training first responders, and increased numbers of expertly skilled personnel can increase an agency's digital forensic capacity, the digital forensic research community has identified the need for automation and intelligent evidence processing (Sun, 2010). One of the more labor intensive and highly-skilled tasks encountered in digital forensic investigation is malware analysis. A common technique for analyzing malware is to execute the malware in a sandbox/virtual machine to gain insight to the attack vector, payload installation, network communications, and behavioral analysis of the software with multiple snapshots taken throughout the analysis of the malware lifecycle. This is an arduous, time-consuming, manual task that can often span over several days. A survey of digital forensic examiners conducted by Hibshi et al. (2011) found that users are often overwhelmed by the amount of technical background required to use common forensic tools. This results in a high barrier to entry for digital investigators to expand their skillset to incorporate additional topics of expertise, such as malware analysis.

Artificial Intelligence (AI) combined with automation of digital evidence processing at appropriate stages of an investigation has significant potential to aid digital investigators. AI can expedite the investigative process and ultimately reduce case backlog while avoiding bias and prejudice (James and Gladyshev, 2013). Overviews of the applications of AI to security and digital forensics are provided in (Franke and Srihari, 2008) and (Mitchell, 2014). A number of approaches have been implemented to aid digital forensic investigation through AI techniques (Mohammed et al., 2016; Rughani and Bhatt, 2017), automation (In de Braekt et al., 2016), and big data processing (Guarino, 2013).

* Corresponding author.
E-mail addresses: quan.le@ucd.ie (Q. Le), oisin.boydell@ucd.ie (O. Boydell), brian.macnamee@ucd.ie (B.M. Namee), mark.scanlon@ucd.ie (M. Scanlon).

https://doi.org/10.1016/j.diin.2018.04.024
1742-2876/© 2018 The Author(s). Published by Elsevier Ltd on behalf of DFRWS. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article in press as: Le, Q., et al., Deep learning at the shallow end: Malware classification for non-domain experts, Digital Investigation (2018), https://doi.org/10.1016/j.diin.2018.04.024

Contribution of this work

The contribution of this work can be summarized as:

- An overview of existing techniques for malware analysis from a manual and automated perspective.
- An approach to enable malware classification by malware analysis non-experts, i.e., no expertise required on behalf of the user in reverse engineering/binary disassembly, assembly language, behavioral analysis, etc.
- Without using complex feature engineering, our deep learning model achieves a high accuracy of 98.2% in classifying raw binary files into one of 9 classes of malware. Our model takes 0.02 s to process one binary file in our experiments on a regular desktop workstation; this short processing time is of potential practical importance when applying the model in reality.
- Our one dimensional representation of a raw binary file is similar to the image representation of a raw binary file (Nataraj et al., 2011); but it is simpler, and it preserves the sequential order of the byte code in the binaries. The sequential representation makes it natural for us to apply the Convolutional Neural Network - Bi Long Short Term Memory architecture (CNN-BiLSTM) on top of it; helping us achieve better performance than using the CNN model alone.

Literature review/state of the art

There is a growing need for non-expert tools to perform digital evidence discovery and analysis (Sun, 2010; van de Weil et al., 2018). Due to the increasing delays in processing digital forensic evidence in law enforcement agencies throughout the world, there has been a focus in the digital forensic research and vendor communities in empowering the non-expert case detective to perform some preliminary analysis on the gathered evidence in a forensically sound manner (Lee et al., 2010). To this end, the Netherlands Forensic Institute (NFI) have implemented a Digital Forensics as a Service solution to expedite digital forensic processing (Casey et al., 2017). This system facilitates the case officer in uploading evidence to a private cloud-based system. Preliminary preprocessing takes place and the officer is able to browse the evidence to unearth potentially case-progressing information.

Digital forensic backlog

Storage capabilities are increasing exponentially while cybercrime related court cases are being dismissed. According to Ratnayake et al. (2014), the likelihood of a prosecution can be lessened due to the uncertainty in determining the age of a victim portrayed in a digital image. Their work considered a parallel challenge to age estimation which was to scan the sheer surface of disk drives. They are aware of the backlog that is eminent due to the lack of both relevant experts to analyze an offense and a laborious digital forensic process. Per Scanlon (2016), these factors will continuously influence the throughput of digital forensic laboratories; therefore, hinder digital forensic investigators in the future.

Machine learning for malware analysis

Machine learning offers the ability to reduce much of the manual effort required with traditional approaches to malware analysis, as well as increased accuracy in malware detection and classification. In the context of malware analysis, a machine learning model is trained on a dataset of existing labeled malware examples, with the labeling either in terms of *malicious* or *benign* in the case of binary classification, or in terms of the type or family of malware for multi-class classification. In either case, the model learns the differentiating features between the classes and so is able to infer, for a new and previously unseen example, whether it is malicious or benign, or which malware family it belongs to with a certain degree of accuracy.

Of course there are many different types and variations of machine learning algorithms and the training examples can be represented in many different ways, which all influence the classification accuracy of the resulting model. Research in the field generally involves the evaluation of different machine learning algorithms and approaches, in conjunction with different and novel types of features derived from the data. Many different approaches have been proposed and a comprehensive review of the literature is provided by both Ucci et al. (2017) and Gandotra et al. (2014).

In the next section, we focus specifically on approaches based on *deep learning* (a type of machine learning) as these are most related to our work. However, the types of features used and how they are extracted in the general context of machine learning for malware classification is also of key relevance. Machine learning reduces much of the manual effort required with traditional approaches to malware analysis by automatically learning to differentiate between malicious or benign, or different families of malware. However, the analysis and extraction of the features from the data, over which the machine learning model operates, still requires a high level of domain expertise in conjunction with complex and time consuming processes.

There are two families of features used in malware analysis: those which can be extracted from the static malware bytecode, and those which require the malware code to be executed (typically in a sandbox environment). Static features include information such as processor instructions, null terminated strings and other static resources contained in the code, static system library imports, and system API calls, etc. Features derived from executed code capture how the malware interacts within the wider operating system and network and can include dynamic system API calls and interactions with other system resources such as memory, storage and the wider network, e.g., connecting to external resources over the Internet.

Although dynamic features extracted from executed code are generally more time and computational resource consuming to extract than features from the static code, both cases require specialist tools and software environments – not to mention a high level of domain expertise required to understand and extract them. The core benefit of our approach, which we present in detail in the Methodology section, is that our deep learning model requires only the raw, static bytecode as input with no additional feature extraction or feature engineering.

Before moving on to review general deep learning approaches for malware classification in the next section, we first discuss two machine learning approaches which attempt to make use of the raw, static bytecode in a way which has some similarities to our work. Nataraj et al (2011) interpret the raw bytecode as greyscale image data where each byte represents a greyscale pixel, and they artificially wrap the byte sequence into a two dimensional array. They then treat the malware classification task as image classification by applying various feature extraction and feature engineering techniques from the image processing field, and use machine learning over these. Inspired by this approach, Ahmadi et al. (2016) use a similar representation of the data, and they evaluate this technique using the same dataset with which we evaluate our work, however they do not make use of deep learning. We provide a comparison of classification accuracy to our approach in the Results section. The application of image classification techniques to the malware domain however still requires the use of complex feature extraction procedures and domain expertise.

Please cite this article in press as: Le, Q., et al., Deep learning at the shallow end: Malware classification for non-domain experts, Digital Investigation (2018), https://doi.org/10.1016/j.diin.2018.04.024

Download English Version:

https://daneshyari.com/en/article/6884392

Download Persian Version:

https://daneshyari.com/article/6884392

Daneshyari.com