International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT

ELSEVIER SSRN

IJIEMR Transactions, online available on 29th May 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=Issue05

10.48047/IJIEMR/V11/ISSUE 05/45

TITLE: **DETECTING WEB ATTACKS WITH END-TO -END DEEP LEARNING**
Volume 11, ISSUE 05, Pages: 279-282

Paper Authors **S. Spandana1, Velanati Raviteja2, Adumala Haritha2, Jolam Jagadish2,Marpu Sai Charan Reddy2, Beerthi Teja2**

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

DETECTING WEB ATTACKS WITH END-TO -END DEEP LEARNING

S. Spandana[1], Velanati Raviteja[2], Adumala Haritha[2], Jolam Jagadish[2], Marpu Sai Charan Reddy[2], Beerthi Teja[2]

[1]Assistant Professor, [2]UG Scholar, [1,2]Department of Computer Science Engineering

[1,2]Malla Reddy Engineering College and Management Sciences, Medchal, Hyderabad

ABSTRACT

Web applications are often the target of cyberattacks because of their frequent vulnerabilities and network accessibility. An intrusion detection system monitors online applications and notifies users when an attempt at an attack is made. Current implementations of intrusion detection systems frequently extract qualities from input strings or network packets that are explicitly chosen as relevant to attack analysis. However, selecting characteristics by hand requires a great deal of effort and in-depth knowledge of the security area. Moreover, in order to classify normal and abnormal actions, supervised learning algorithms need large volumes of labelled legitimate and attack request data, which are usually prohibitively expensive or difficult to obtain for live online services. Three new perspectives are added to the field of autonomic intrusion detection systems study by this paper. First, we evaluate the practicality of an unsupervised/semi-supervised approach.

Keywords: Web security, Deep learning, Application instrumentation.

1. INTRODUCTION

Emerging trends and challenges. Web applications are attractive targets for cyber attackers. SQL injection [1], cross site scripting (XSS) [2] and remote code execution are common attacks that can disable web services, steal sensitive user information, and cause significant financial loss to both service providers and users. Protecting web applications from attack is hard. Even though developers and researchers have developed many counter-measures (such as firewalls, intrusion detection systems (IDSs) [3] and defensive programming best practices [4]) to protect web applications, web attacks remain a major threat. For example, researchers found that more than half of web applications during a 2015–2016 scan contained significant security vulnerabilities, such as XSS or SQL Injection [5]. Moreover, hacking attacks cost the average American firm $15.4 million per year [6]. The Equifax data breach in 2017 [7, 8] (which exploited a vulnerability in Apache Struts) exposed over 143 million American consumers' sensitive personal information. Although the vulnerability was disclosed and patched in March 2017, Equifax took no action until four months later, which led to an estimated insured loss of over 125 million dollars. Conventional intrusion detection systems do not work as well as expected for a number of reasons, including the following: large demand and relatively low barrier to entry into the software profession, however, many developers lack the necessary knowledge of secure coding practices.

2. LITERATURE SURVEY

large demand and relatively low barrier to entry into the software profession, however, many developers lack the necessary knowledge of secure coding practices. Many intrusion detec- tion systems rely on rule-based strategies or supervised machine learning algorithms to differentiate normal requests from attack requests, which requires large amounts of labeled training data to train the learn- ing algorithms.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

It is hard and expensive, however, to obtain this training data for arbitrary custom applica- tions. In addition, labeled training data is often heav- ily imbalanced since attack requests for custom sys- tems are harder to get than normal requests, which poses challenges for classifiers [10]. Moreover, although rule-based or supervised learning approaches can dis- tinguish existing known attacks, new types of attacks and vulnerabilities emerge continuously, so they may be misclassified.

3. EXISTING METHOD

As discussed, different attacks have dif- ferent characteristics and traditional feature engineering approaches lack a unified solution for all types of attacks. RSMT bypasses these attack vectors and instead cap- tures the low-level call graph. It assumes that no matter what the attack type is (1) some methods in the server that should not be accessed are invoked and/or (2) the access pattern is statistically different than the legitimate traffic. RSMT operates as a late-stage (post-compilation) instrumentation-based toolchain targeting languages that run on the Java Virtual Machine (JVM)

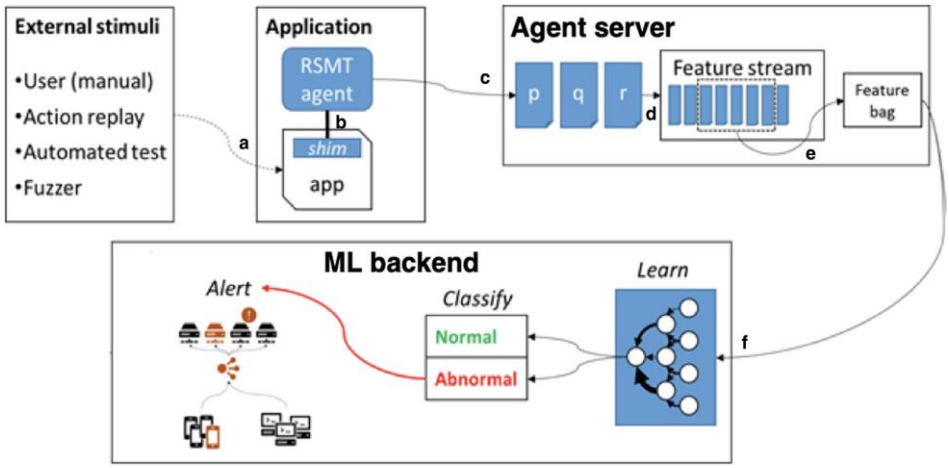

Fig. 1: arbitrarily fine-grained traces of program execution

Figure 1 shows the high-level workflow of RSMT's web attack monitoring and detection system. This sys- tem is driven by one or more environmental stimuli (a), which are actions transcending process boundaries that can be broadly categorized as either manual (e.g., human interaction-driven) or automated (e.g., test suites and fuzzers) inputs. The manifestation of one or more stimuli results in the execution of various application behaviors. RSMT attaches an agent and embeds lightweight shims into an application (b). These shims do not affect the functionality of the software, but instead serve as probes that allow efficient examination of the inner workings of software applications. The events tracked by RSMT are typically control flow-oriented, though dataflow-based analysis is also possible.

As the stimuli drive the system, the RSMT agent inter- cepts event notifications issued by shim instructions. These notifications are used to construct traces of behav- ior that are subsequently transmitted to a separate trace management process (c). This process aggregates traces over a sliding window of time (d) and converts these traces into "bags" of features (e). RSMT uses feature bags to enact online strategies (f), which involve the following two epochs:

4. PROPOSED METHOD

To train our stacked autoencoder we use a pretrain- ing step involving greedy layer-wise training. The first layer of encoder is trained on raw input. After a set of parameters are obtained, this layer is used

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

to transform the raw input to a vector represented as the hidden units in the first layer. We then train the second layer on this vector to obtain the parameters of second layers. This pro- cess is repeated by training the parameters of each layer individually, while keep the parameters of other layers unchanged.

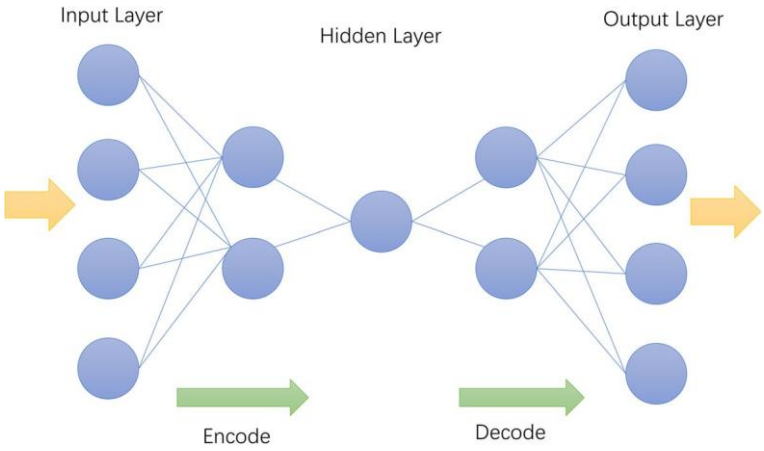

Fig. 2: T-SNE Visualization of Normal and abnormal requests

5. CONCLUSION

This paper describes the architecture and results of apply- ing a unsupervised end-to-end deep learning approach to automatically detect attacks on web applications. We instrumented and analyzed web applications using the Robust Software Modeling Tool (RSMT), which autonom- ically monitors and characterizes the runtime behavior of web applications. We then applied a denoising autoen- coder to learn a low-dimensional representation of the call traces extracted from application runtime. To validate our intrusion detection system, we created several test appli- cations and synthetic trace datasets and then evaluated the performance of unsupervised learning against these datasets. While cross validation is widely used in traditional machine learning, it is often not used for evaluating deep learning models because of the great computational cost. We needed to compare autoencoder approaches with other machine learning methods. To enable a fair comparison, we didn't use cross validation in our experiments.

6. REFERENCES

[1]. Japkowicz N, Stephen S. The class imbalance problem: A systematic study. Intell Data Anal. 2002;6(5):429–49.

[2]. Liu G, Yi Z, Yang S. A hierarchical intrusion detection model based on the pca neural networks. Neurocomputing. 2007;70(7):1561–8.

[3]. Xu X, Wang X. An adaptive network intrusion detection method based on pca and support vector machines. Advanced Data Mining and Applications. 2005;3584:696–703.

[4]. Pietraszek T. Using adaptive alert classification to reduce false positives in intrusion detection. In: Recent Advances in Intrusion Detection. Springer; 2004. p. 102–24.

[5]. Goodfellow I, Bengio Y, Courville A. Deep Learning: MIT press; 2016.

[6]. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2012. p. 1097–105.

[7]. Amodei D, Ananthanarayanan S, Anubhai R, Bai J, Battenberg E, Case C, Casper J, Catanzaro B, Cheng Q, Chen G, et al. Deep speech 2:

[8]. End-to-end speech recognition in english and mandarin. In: International Conference on Machine Learning. New York: PMLR; 2016. p. 173–82.

[9]. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2014. p. 3104–12.

[10]. Sun F, Zhang P, White J, Schmidt D, Staples J, Krause L. A feasibility study of autonomically detecting in-process cyber-attacks. In: Cybernetics (CYBCON), 2017 3rd IEEE International Conference On. IEEE; 2017. p. 1–8.

[11]. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res. 2010;11(Dec): 3371–408.

[12]. Fu X, Lu X, Peltsverger B, Chen S, Qian K, Tao L. A static analysis framework for detecting sql injection vulnerabilities. In: Computer Software and Applications Conference, 2007. COMPSAC 2007. 31st Annual International. IEEE; 2007. p. 87–96.