KIDS: For Key Recovery Attacks and Security of Machine Learning

Pooja Mande[1], Prof. K. S. Kore[2]

[1]Student, ME Computer, [2]Assistant Professor
SPCOE, Department Of Computer Engineering, Otur

***Abstract*: In recent era the use of internet become amplified extremely. Most of people used internet to convey their data and used cloud to save it. There is chance that the data may get scythed and get tainted. Since most current network attacks happen at the application layer, analysis of packet payload is necessary for their detection. To improved security from such unauthorized users various Anomaly intrusion detection schemes are introduced recently. To defeat these troubles one such structure is Keyed Intrusion Detection System is application layer network anomaly detection system which is based on principle which is much same as the working of some cryptographic primitives. By adding mystery component in to plan so that a couple of operations are becomes impractical without knowing the key. Core idea to make evasion attacks more difficult is to add the concept of a "key" which is the secret element used to determine how classification features are extracted from the payload. Key is different for each implementation of the method and is kept secret. Therefore model of normal payload is secret although detection method is public. In KIDS the scholarly model and the irregularity's calculation score are both key-subordinate, a reality which obviously keeps an aggressor from making shirking assaults. In this recovering the key is to marvelously straightforward and require that attacker can collaborate with KIDS and get criticism about examining solicitations. Here present realistic attacks for two different adversarial settings and show that recovering the key requires only a small amount of queries.**

***Keywords*: Anomaly Detection, Intrusion Detection Systems, Machine Learning.**

1. Introduction

In recent years use of internet has been increased tremendously. Most of people used internet to transmit their data and used cloud to save it. There is possibility that the data may get hacked and get misused. For better protection from such unauthorized users various Anomaly intrusion detection schemes are introduced in recent year. Security problem mainly divided into two groups one is malicious and other is non malicious activity. A malicious attack is an attempt to forcefully abuse or take advantage of someone's computer, whether through computer viruses, social engineering, phishing, or other types of social engineering. This can be done with the intent of stealing personal information (such as in social engineering) or to reduce the functionality of a target computer. Malicious Code mostly Hide in Email, Web Content, Legitimate Sites, File Downloads [1]. For example Trojan, Horse, Viruses, Worms, Phishing, Baiting, Spam non- malicious attacks occur due to poor security policies and controls that allow vulnerabilities and errors to take place. An intrusion detection system (IDS) is a device or software application that monitors network or system activities for malicious activities or policy violations and produces reports to a management station. IDS come in a variety of "flavors" and approach the goal of detecting suspicious traffic in different ways. There are network based (NIDS) and host based (HIDS) intrusion detection systems [2]. NIDS is a network security system focusing on the attacks that come from the inside of the network (authorized users).

Some systems may attempt to stop an intrusion attempt but this is neither required nor expected of a monitoring system. So attacker always tries to avoid detection. In terms of network security the evasion attack means bypass a flaw in a security system that allows an attacker to circumvent security mechanisms to get system or network access in order to deliver an exploit, attack, or other form of malware without detection. Evasions are typically used to counter network-based intrusion detection and prevention systems but can also be used to by-pass firewalls [3]. A further target of evasions can be to crash a network security device, rendering it in-effective to subsequent targeted attacks. Few detection schemes are introduced in last decade to protect from such evasion attacks. KIDS (Keyed Intrusion Detection System) one of the scheme to avoid evasion attacks. KIDS first time introduced by Mrdovic and Drazenovic at DIMVA"10. Most current network attacks happen at the application layer, analysis of packet payload is necessary for their detection. Unfortunately malicious packets may be crafted to normal payload, and so avoid detection if the anomaly detection method is known. Model of normal payload is key dependent. Key is different for each implementation of the method and is kept secret. Therefore model of normal payload is secret although detection method is public. This prevents attacks. Payload is partitioned into words [4]. Words are defined by delimiters. Set of delimiters plays a role of a key.

2. Related Work

In this section, the reference are collected from all conferences, sites, articles, books from the internet which helps to implement the project. For development of this project we referred some of the base papers, ideas which helps in development, testing and deployment phase. For good understanding of the advanced authentication system there are some work on the IEEE international journal that we have referenced are:

Can Machine Learning Be Secure?
Machine learning systems suggest unparalleled flexibility in dealing with developing input in a multiple applications such as intrusion detection systems and spam e-mail filtering. Malicious opponent always made target on Machine learning algorithms [4],[10]. System provides a framework for answering the question, “Can machine learning be secure?” System includes taxonomy of different types of attacks on machine learning techniques and systems and multiple defences against those attacks. An analytical model gave a lower bound on attacker’s work function, and a list of open problem.

The Security of Machine Learning:
Machine learning’s ability quickly evolve to changing and complex situations has helped it to become a elementary tool for computer security. That adaptability is also vulnerability that is attackers can exploit machine learning systems. They present a taxonomy identifying and analyzing attacks against machine learning systems. They show how these classes influence the costs for the attacker and defender, and they give a formal structure defining their interaction. They use their framework to survey and analyse the literature of attacks against machine learning systems. They also illustrate their taxonomy by showing how it can guide attacks against SpamBayes, a popular statistical spam filter. Finally, they discuss how their taxonomy suggests new lines of defences [5].

Adversarial Pattern Classification Using Multiple Classifiers and Randomization:
In many security applications an adversarial classification problem is face by pattern recognition system, in which an intelligent, adaptive enemy modifies patterns to avoid the classifier. Several strategies have been recently proposed to make a classifier harder to avoid, but they are based only on qualitative and intuitive arguments [6],[7]. In this work, they consider a strategy made of hiding information about the classifier to the enemy through the beginning of some randomness in the decision function. They focus on an implementation of this strategy in a multiple classifier system is a classification architecture which is broadly used in security applications. They provide a formal support to this strategy, based on an analytical framework for adversarial classification problems which are proposed recently by other authors, and give an experimental evaluation on a spam filtering task to demonstrate their findings [8].

Support Vector Machine Under Adversarial Label Noise:
Malicious adversaries may manipulate data in adversarial classification tasks such as spam filtering and intrusion detection to prevent the outcome of an automatic analysis. Thus as well achieving high quality classification performances then the machine learning algorithms have to be robust against adversarial data manipulation to successfully operate in these tasks. While support vector machines (SVMs) are extremely successful approach in classification problems and their effectiveness in adversarial classification tasks has not been extensively investigated yet [9],[11]. In this system they present a preliminary investigation of the robustness of SVMs against adversarial data manipulation. In particular, they assume that the enemy has control over some training data, and aims to subvert the SVM learning process. Within this assu3mption, they show that this is indeed possible, and propose a strategy to improve the robustness of SVMs to training data manipulation based on a simple kernel matrix correction [12].

Polymorphic Blending Attacks:
A very effective way to evade signature-based intrusion detection systems (IDS) is to use polymorphic techniques to generate attack instances that do not share a fixed signature. Anomaly-based intrusion detection systems offer excellent defence because existing polymorphic techniques can make the attack instances look unlike from each other, but cannot make them same like normal [13]. In this system a new class of polymorphic attacks is introduce which is called polymorphic blending attacks. It can efficiently avoid byte frequency-based network anomaly IDS by watchfully matching the statistics of the mutated attack instances to the normal profiles. The proposed polymorphic blending attacks can be viewed as a subclass of the mimicry attacks [14],[16]. They take a systematic approach to the problem and formally describe the algorithms and steps required to carry out such attacks. They not only show that such attacks are feasible but also analyse the hardness of evasion under different circumstances. They present detailed techniques using PAYL, a byte frequency-based anomaly IDS, as a case study and demonstrate that these attacks are indeed feasible. They also provide some insight into possible countermeasures that can be used as defence [15].

The problem of computing optimal strategies is to modify an attack so that it avoids detection by a naive Bayes classifier [17]. They originate the problem in game-theoretic terms where each change made to an instance comes at a cost and successful detection and avoidance have measurable utilities to the classifier and the enemy respectively. The authors study how to discover such optimally modified instances by adapting the decision surface of the classifier, and also discuss how the enemy respond on this. The setting used in assumes an enemy with full knowledge of the classifier to be avoided. After that how avoidance can be done when such information is not available [18]. They formulate the adversarial classifier reverse engineering problem (ACRE) as the task of learning enough information about a classifier to construct attacks, instead of looking for optimal strategies. The authors use a membership oracle as implicit adversarial model: the attacker is given the chance to query the classifier with any selected instance to find out whether it is malicious or not. Accordingly a reasonable objective is to find instances that avoid exposure with reasonable number of queries. A classifier is ACRE learnable if there is an algorithm that finds a minimum cost instance which avoid detection using just polynomially many queries. Similarly a classifier is also ACRE k-learnable if the cost is not minimum but bounded by k. Then it is proved that linear classifiers with unbroken features are ACRE k-learnable under linear cost functions [19]. Therefore, these classifiers should not be used in adversarial environments. Consequent work by Nelson in generalizes these results to convex-inducing classifiers, viewing that it is normally not required to reverse engineer the decision boundary to construct undetected instances of near minimum cost. In general, some extra works have revisited the task of machine learning in security applications, with exacting prominence on anomaly detection [8], [20].

3. Proposed Work

Our aim is to provide great degree expert, that it is the sensibly easy for an attacker to recoup the key in any of the settings. It is consider that the such an absence of security not protect from anticipate like children from key-recovery assaults. Here claimed the resistance against such assaults is key to any classifier that attempt to hinder avoidance by depending on a mystery bit of data. We have given exchange on this and other open enquiries in the trust of empowering further research around there. The assaults here exhibited could be the forestalled by presenting various impromptu the counter measures of the frameworks, for example, constraining the most large length of words , or including such amounts as order components. Then again, that these variations may in any case be the powerless against some individual assaults. In this manner, our suggestion for future plans is to construct choices in light of hearty standards as opposed to specific fixes. Our aim is enhance the KIDS and try to meet maximum security properties so that it can able to secure stored data in clouds for various healthcare domains. Architecture of proposed system and proposed system module Details: Node Creation & Routing: In this module, authenticated node is created for each user. KIDS system gets all details about user and stored the same for creating the rules. After node creation when user saved files, each files get saved in encrypted format. For each file user get secret Key. Key- Recovery Attacks On Kids: At this point assault can able to attack and get the knowledge about the secret key. Assault able to get user data files and used same information for various reasons [2]. Assault changes the Key and modified the same so it will not available further more to any authenticated user. Implicitly here grey box or black box attacks happened in which secret Key partially or fully modified and then make available to end nodes. Keyed Anomaly Detection and Adversarial Models: Revisited After secret key modification KIDS system alerts to the main station and check the authentication list. If unauthenticated node found then it get blocked by KIDS system. Modified key then recoup and provided to the intended node. Performance Analysis: For performance evaluation following graph can be used Delay, Packet delivery ratio.

System Architecture:

Figure shows system architecture of proposed system. The attacks are extremely efficient, showing that it is reasonably easy for an attacker to recover the key in any of the two settings discussed. In this a lack of security reveals that schemes like kids were simply not designed to prevent key-recovery attacks. However, in this paper system have argued that resistance against such attacks is essential to any classifier that attempts to impede evasion by relying on a secret piece of information.

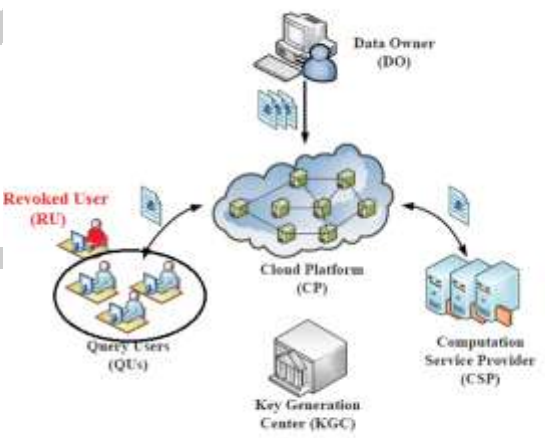


Figure 1. System Architecture

At the point when surveying the security of frameworks, for example, KIDS, one noteworthy issue originates from the nonappearance of broadly acknowledged antagonistic models giving an exact portrayal of the aggressor's objectives and his abilities one such model for secure machine learning and talked about different general assault classes. Our work does not fit well inside in light of the fact that our principle objective is not to assault the learning calculation itself, but rather to recoup one bit of mystery data that, in this way, may be vital to successfully dispatch an avoidance assault.

4. Key Recovery on Gray-Box KIDS

In this attack assume the attacker has access to the anomaly score which is assigned to a selected payload. It is sensible to assume that some normal payloads are identified excessively. Consider p is the one normal payload. A straight-forward strategy to recognize what elements of p belong to the key D consists of feeding KIDS with the first byte of p, then with the first two bytes of p, and so on. When the next-to-the-last byte happens to be a delimiter and KIDS will detect a transition where the left word is same as the word which has been seen during training mode and whereas the right word is often strange. Then at this condition the anomaly score will experience a small decrement. This procedure is repeated conveniently and all the delimiters present in p can be recovered. The main drawback of the naive strategy is that the attacker will only be able to recover those key elements which are present in the normal payloads, which may well be just a fraction of all of them. Moreover the complexity of such an attack is linear in the number of payloads and their lengths. There are different approaches that obtain all the key elements more efficiently and without directly relying on normal payloads [1].

5. Key Recovery on Black-Box KIDS

Here a key-recovery attack is presented when the only information about a actual data an enemy gets from KIDS whether it is normal or anomalous. In some respects this information is less fine-grained than the anomaly score, so it is realistic to expect that attacks which are working under this assumption will be slightly more complex. In this KIDS is used with a normal payload join with a carefully constructed end and this end contains a large number of unseen words divided by the candidate delimiter. If the delimiter does not belong to the key then the entire end will be process as just one word and the anomaly score will be almost similar to that of the actual payload. If this is the case, then the payload will be marked as normal with high probability. On the other hand if the delimiter does belong to the key then the end will be fragmented into a large number of previously hidden words and transitions. This will harmfully impact on the anomaly score and consistently resulting in an abnormal payload [1].

6. Experimental Result

Below table shows experimental result of our proposed system.

Datasize(MB)	Upload Time(in Ms.)	Download Time(in Ms.)
10	125	80
20	210	150
30	380	270
40	460	360

Table 1: Security Performance Comparisons

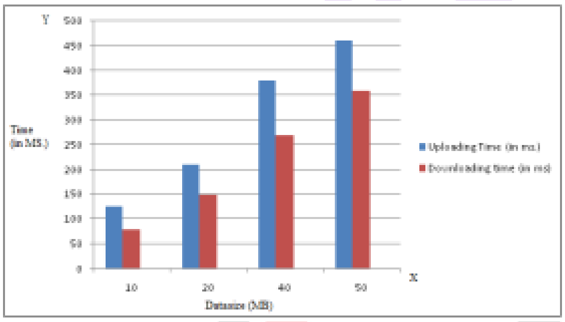
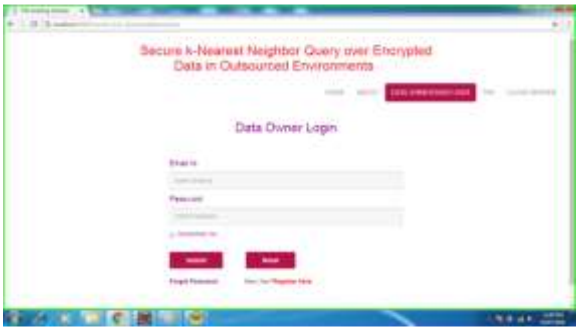


Figure 2. System Performance

In this result the file is uploaded to the different servers or systems then calculating the time (in Millisecond) required for upload the file with different size and also calculate the time required for download the file with different size.

7. Snapshots:



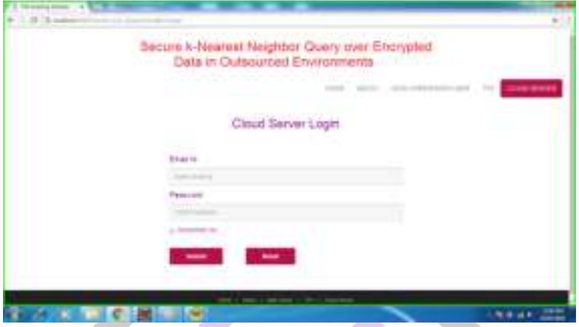
Secure k-Nearest Neighbor Query over Encrypted Data in Outsourced Environments

Data Owner Login

Email Id

Password

Figure 1. Data Owner Login



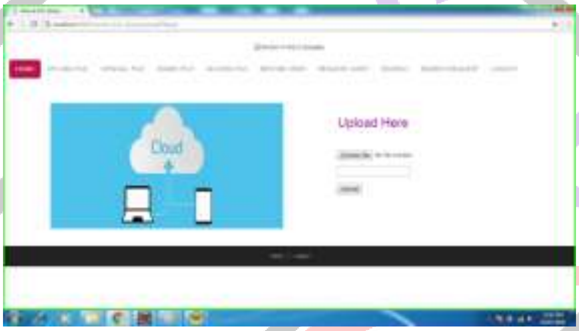
Secure k-Nearest Neighbor Query over Encrypted Data in Outsourced Environments

Cloud Server Login

Email Id

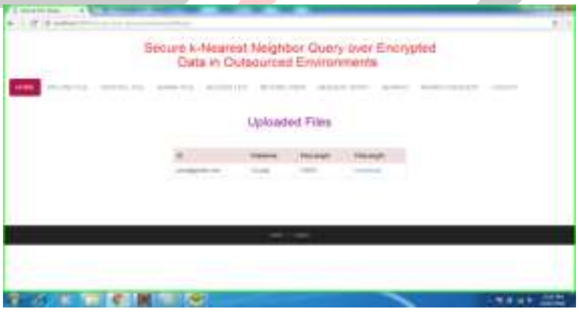
Password

Figure 2. Cloud Server Login



Upload Here

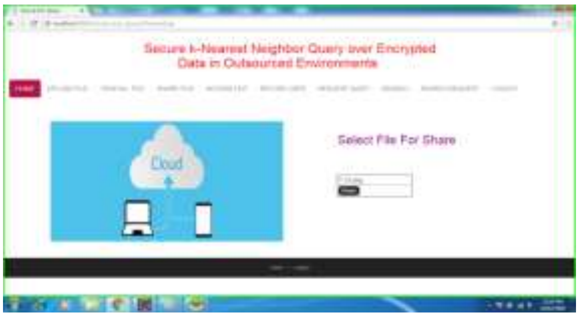
Figure 3. Upload File



Secure k-Nearest Neighbor Query over Encrypted Data in Outsourced Environments

Uploaded Files

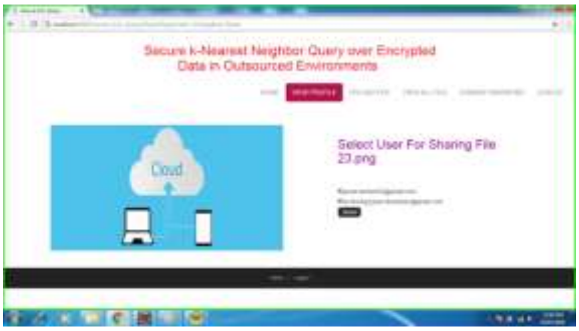
Figure 4. Uploaded Files



Secure k-Nearest Neighbor Query over Encrypted Data in Outsourced Environments

Select File For Share

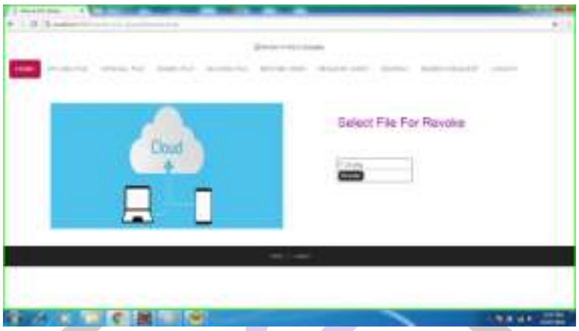
Figure 5. Select File for Share



Secure k-Nearest Neighbor Query over Encrypted Data in Outsourced Environments

Select User For Sharing File 23.png

Figure 6. Select User for Share



Select File For Revoke

Figure 7. Select File to Upload

Download Files

Figure 8. Select File for Revoke

8. Conclusion

We have examined the quality of KIDS against key-recovery assaults. In doing as such, we have adjusted to the irregularity recognition setting an ill-disposed model obtained from the related field of ill-disposed learning. To the best of our insight, our work is the first to exhibit key-recovery assaults on a keyed classifier. Shockingly, our assaults are to a great degree proficient, demonstrating that it is sensibly simple for an aggressor to recoup the key in any of the two settings examined. Such an absence of security may uncover that plans like KIDS were just not intended to avert key-recovery assaults. However, we have argued that resistance against such attacks is essential to any classifier that attempts to impede evasion by relying on a secret piece of information. Our future design is to base decisions on robust principles rather than particular fixes. Going beyond KIDS, it remains to be seen whether similar schemes are secure against key recovery attacks. Our attacks (or variants of them) are focused on keyed classifiers, and we believe that they will not carry over randomized classifiers. We note that, in its present form, KIDS cannot be easily randomized, as choosing a new key implies training the classifier again, which is clearly impractical in real-world scenarios.

9. Acknowledgement

I express my sincere thanks to my project guide Prof. K. S. Kore who always being with presence & constant, constructive criticism to made this paper. I would also like to thank all the staff of computer department for their valuable guidance, suggestion and support through the paper work, who has given co-operation for the project with personal attention. Above all I express our deepest gratitude to all of them for their kind-hearted support which helped us a lot during paper work. At the last I thankful to my friends, colleagues for the inspirational help provided to me through a paper work.

References

[1] Juan E. Tapiador, Agustin Orfila, Arturo Ribagorda, and Benjamin Ramos, "Key-Recovery Attacks on KIDS, a Keyed Anomaly Detection System," IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 12, NO. 3, MAY/JUNE 2015.

[2] M. Barreno, B. Nelson, R. Sears, A.D. Joseph, and J.D. Tygar, “Can Machine Learning be Secure?” Proc. ACM Symp. Information, Computer and Comm. Security (ASIACCS ’06), pp. 16-25, 2006.

[3] M. Barreno, B. Nelson, A.D. Joseph, and J.D. Tygar, “The Security of Machine Learning,” Machine Learning, vol.81, no. 2, pp. 121- 148, 2010.

[4] B. Biggio,G. Fumera, and F. Roli, “Adversarial Pattern Classification Using Multiple Classifiers and Randomisation,” Proc. IAPR Int’l Workshop Structural, Syntactic, and Statistical Pattern Recognition, pp. 500-509, 2008.

[5] B. Biggio, B. Nelson, and P. Laskov, “Support Vector Machines Under Adversarial Label Noise,” J. Machine Learning Research, vol. 20, pp. 97-112, 2011.

[6] P. Fogla, M. Sharif, R. Perdisci, O. Kolesnikov, and W. Lee, “Polymorphic Blending Attacks,” Proc. 15th Conf. USENIX Security Symp., 2006.

[7] Sasa Mrdovic, Branislava Drazenovic “ KIDS – Keyed Intrusion Detection System” Proc. Seventh Int’l Conf. Detection of Intrusions and Malware and Vulnerability Assessment (DIMVA ’10), pp. 173-182, 2010.

[8] C. Gates and C. Taylo, “Challenging the Anomaly Detection Paradigm: A Provocative Discussion,” Proc. New Security Paradigms Workshop (NSPW), pp. 21-29, 2006.

[9] D. Lowd and C. Meek, “Adversarial Learning,” Proc. 11th ACM SIGKDD Int’l Conf. Knowledge Discovery in Data Mining (KDD ’05), pp. 641-647, 2005.

[10] N. Dalvi, P. Domingos, Mausam, S. Sanghai, and D. Verma, “Adversarial Classification,” Proc. 10th ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining (KDD ’04), pp. 99-108, 2004.

[11] B. Nelson, A.D. Joseph, S.J. Lee, and S. Rao, “Near-Optimal Evasion of Convex-Inducing Classifiers,” J. Machine Learning Research, vol. 9, pp. 549-556, 2010.

[12] B. Nelson, B.I.P. Rubinstein, L. Huang, A.D. Joseph, S.J. Lee, S. Rao, and J.D. Tygar, “Query Strategies for Evading Convex-Inducing Classifiers,” J. Machine Learning Research, vol. 13, pp. 1293-1332, May 2012.

[13] Y. Zhou, Z. Jorgensen, and M. Inge, “Combating Good Word Attacks on Statistical Spam Filters with Multiple Instance Learning,” Proc. 19th IEEE Int’l Conf. Tools with Artificial Intelligence (ICTAI ’07), pp. 298-305, 2007.

[14] A. Kolcz and C.H. Teo, “Feature Weighting for Improved Classifier Robustness,” Proc. Sixth Conf. Email and Anti-Spam (CEAS ’09), 2009.

[15] O. Kolesnikov, D. Dagon, and W. Lee, “Advanced Polymorphic Worms: Evading IDS by Blending in with Normal Traffic,” Proc. USENIX Security Symp., 2005.

[16] K. Wang, J. Parekh, and S. Stolfo, “Anagram: A Content Anomaly Detector Resistant to Mimicry Attack,” Proc. Ninth Int’l Conf. Recent Advances in Intrusion Detection (RAID ’06), pp. 226-248, 2006.

[17] K. Rieck and P. Laskov, “Language models for detection of unknown attacks in network traffic,” Journal in Computer Virology, vol. 2, 2007, pp. 243-256.

[18] K. Rieck, “Computer Security and Machine Learning: Worst Enemies or Best Friends?” Proc. DIMVA Workshop Systems Security (SYSSEC), 2011.

[19] R. Sommer and V. Paxson, “Outside the Closed World: On Using Machine Learning for Network Intrusion Detection,” Proc. IEEE Symp. Security and Privacy, pp. 305-316, 2010.

[20] Y. Song, M. Locasto, A. Stavrou, A.D. Keromytis, and S.J. Stolfo, “On the Infeasibility of Modeling Polymorphic Shellcode: Re-Thinking the Role of Learning in Intrusion Detection Systems,” Machine Learning, vol. 81, no. 2, pp. 179-205, 2010.