Prevention of Key Based Attack on Data Sharing System

Mr. SandeshKhaire
PG Student
Department of Computer Science and Engineering
LNCTS, Indore, M.P, India

Prof. Hemant Gupta
Assistant Professor
Department of Computer Science and Engineering
LNCTS, Indore, M.P, India

Prof. Mayank Bhatt
HOD
Department of Computer Science and Engineering
LNCTS, Indore, M.P, India

ABSTRACT

Today security with respect to input data is play as an very important in the networking system. Key recovery is the difficult task in data sharing system. When authenticated user access the file then that user will get the file as well as the key to decrypt that file. But after some time interval if user found less trustworthy then data owner may block that user. The main problem is that user is still having the key so there may be possibility that user can share that key with others so to recover that issue data owner resign the particular file so even though the user try to leak the key then there is no issue of accessing the file. In this paper we proposed two type of key recovery Black box and Gray box key recovery. Many anomaly detection systems depend on machine learning algorithms to derive a model of normality that is later used to detect suspicious events.

Keyword:-*Intrusion Detection System; Anomaly Detection System; Adversarial Classification; SecureMachine Learning.*

1. INTRODUCTION

Most of the computer security problems can be essentially reduced to separating malicious from non-malicious activities. This is, one such a example, in the case of spam filtering, intrusion detection, or the identification of fraudulent behavior. In general defining in a precise and computationally useful way what is harmless or what is offensive is often much complex. To overcome these difficulties, many solutions to such problems have traditionally adopted a machine-learning approach, through the use of classifiers to automatically derive models of (good and/or bad) behavior that are later used to recognize the occurrence of potentially dangerous event.

KIDS idea of learning with secret is not entirely a new. Anagram, another payload-based anomaly detection system was which addresses the evasion problem in same manner was introduced by Wang et al. Here we compare between two broad classes of classifier that make use of key. In the first group, that we term randomized classifiers,the classifier which is entirely public (i.e equivalently, is trained with public information only). However, in detection mode some parameters are randomly chosen every time an instance has to be classified, thus making uncertain for the attacker how the instance will be processed. Note that, in this case, the same instance will be processed differently every time if we choose key is randomly. We emphasize that randomization can also be

applied at training time, although it is sufficiently effective when used during testing only, at least as far as evasion attacks are concerned. KIDS belong to a second group, that we call keyed classifiers.

There are practically various types of cryptanalytic attacks that depends on many factors: Attacks based on few ciphertext are better than attacks that require many ciphertext, known plaintext attacks are better than chosen plaintext attacks, no adaptive attacks are better than adaptive attacks, single key attacks are better than related keyattacks, etc. Since it is difficult to quantify the relative importance of all these factors in different scenarios, we usually concentrate on the total running time of the attack, which is a single well defined number.

2. LITERATURE REVIEW

The difficult problem of computing optimal strategies is to modify an attack so that it evades detection by a Bayes classifier. The problem can be described in game theoretic terms, where each modification made to an instance comes at price and successful detection and evasion have measurable utilities to the classifier and the adversary, respectively. The author study how to detect such optimally modified instances by adapting the decision surface of the classifier and also formulates how the adversary may react to this. The setting requierd in assumes an adversary with full knowledge of the classifier to be evaded after how evasion can be done when such information is unavailable. They formulate the adversarial classifier reverse engineering problem (ACRE) as the task of learning sufficient information about a classifier to construct attack instead of looking for optimal strategies. The author use a membership oracle as implicit adversarial model the attacker is given a chance to query the classifier with any chosen instance to determine whether it is malicious or not. A reasonable objective is to find instances that evade detection with an affordable number of queries. The ACRE is learnable if there exist an algorithm that finds a minimal-cost in-stance evading detection using only polynomial many queries. A classifier is only ACRE k-learnable if the cost is not minimal but bounded by k. Among the results given, it is proved that linear classifiers with continuous feature are ACRE k-learnable under linear cost functions. These classifiers should not be used in adversarial environment. More work by generalizes these results to convex-inducing classifiers, shows that it is generally not required to reverse engineer the decision boundary to construct undetected instances of near minimal cost. For the few open problems and challenges related to the classifier evasion problem. Additional works has revisited the role of machine learning in security application with particular emphasis on anomaly detection.

Barreno et al. [1], [2] has praposed on the risks of applying machine learning algorithms to security domains. In it they introduce a taxonomy which groups attacks on machine learning systems into different categories, depending on whether the adversary influences training or only analyzes an already trained system; whether the goal is to force just one mis-classification, or else to generate too many so that the system becomes unusable; etc. The author described useful discussion on potential counter measures and enumerate various open problems.

Kolesnikov et al. [9] demonstrated that polymorphic mimicry worms, based on encryption and data encoding to obfuscate that their content, are present to evade frequency distribution-based anomaly detectors like PAYL [8]. PAYL models byte-value frequency distributions so detection can be avoided by padding anomalous sequences with an appropriate amount of normal traffic. In order to counteract polymorphic mimicry worms, PAYL authors developed Anagram [8], an anomaly detector that models n-grams observed in normal traffic. Anagram also introduces a new strategy, called randomization, to hinder evasion. There are two possible types of randomization, namely randomized modeling and randomized testing.

3. PROBLEM STATEMENT

In this paper I proposed KIDS for recovering of key. My work shows that recovering the key is extremely simple provided that the attacker can interact with KIDS. Keyed Intrusion Detection System, is a key dependent network anomaly detector that inspects packet payloads. I have provided discussion on this and other open questions in the hope of stimulating further research in this area. The attack here presented could be prevented by introducing a number of ad hoc counter measures the system, such as limiting the maximum length of words and payloads, or including such quantities as classification features. I suspect, however, that these variants may still be vulnerable to other attacks.

4. PROPOSED SYSTEM

To provide user security for file transfer we requires proposed system. As in many roll based access system if the user have the access of the file then user can access the file any time but if the user found unauthorized then there is main challenge is revoking the access of that user .KIDS provide that facility of revoking the access of the user also resignature concept for the particular file.

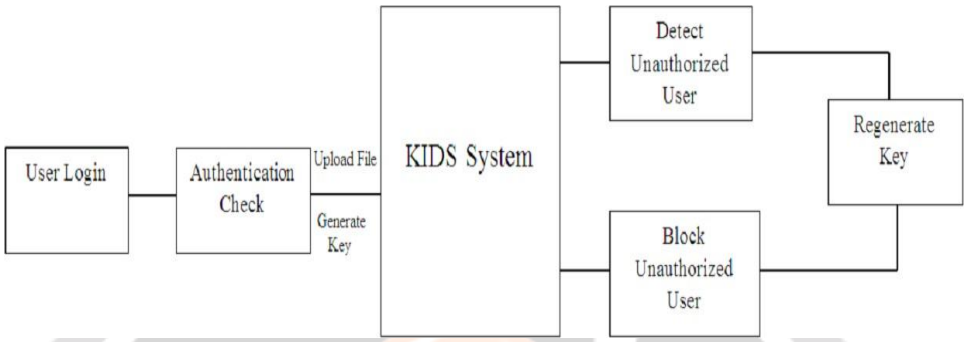

Fig -4.1: Architecture of Proposed System

The above figure shows the Architecture of Proposed System. The system consists of four basic modules which are listed and explain below in detail.

A. Node Creation & Routing:

In this module, first a wireless network is created. All the nodes are randomly deployed in the network area. Our network may be a mobile network; nodes are assigned with mobility. Source and destination nodes are defined. Data transferred from source node to destination node. As we are working in mobile network, nodes mobility is set i.e. node move from one position to another.

B. Key- Recovery Attacks On Kids:

When providing the security of systems such as KIDS, one major problem comes from the absence of widely accepted adversarial model gives a precise explanation of the attacker's goals and his capabilities one such model for secure machine learning and discussed many general categories of attack. Our work does not fit well with in because our main goal is not to attack the learning algorithm itself, but to recover one piece of secret information that, subsequently, may be essential to successfully launch an evasion attack.

C. Keyed Anomaly Detection and Adversarial Models Revisited:

Closely related to the point discussed in above model is the need to establish clearly defined and motivated adversarial model for the secure machine learning algorithms. The assumptions made about the attacker's capabilities are critical to efficiently analyze the security of any scheme but few of them may well be unrealistic for many applications. One debatable issue is that whether the attacker can really receive feedback from the system for instances he chooses. This bears some analogies with Chosen-Plaintext Attacks (CPA) in cryptography. This assumption has been made by many works in secure ma-chine learning, including ours.

D. Unauthorized access:

Once the data owner share the encrypted file with its key to authorized user but after some time interval user found accessing some unauthorized things then there is a problem of key recovery. In this module system will identify the unauthorized access through certain keys. When user press certain keys which predefined by the system then because of this system performance get decrease and that get capture by the server. So to recover the given from the user we apply the resignature concept.

5. ALGORITHM USED FOR IMPLEMENTATION

1. Key-Recovery on Gray-Box

KIDS

1: *D1*←Ø
2: *D2*←Ø
3: for *d* = 0 to 255 do
4: p ← (w1 ‖ d ‖ w2)
5: if S(p) S(w1 ‖ d ‖ w2)∀ d ∈ D1 then
6: D1←D1 ∪{ d }.
7: else
8: D2←D2 ∪{ d }.
9: end-if.
10: end-for
11: q← w2
12: if S(q) S(w1 ‖ d ‖ w2)∀ d ∈ D1 then
13: return D2
14: else
15: return D1

2. Key-Recovery on Black-Box

KIDS

1: for each $q_i \in Q$do
2: $D_i \leftarrow \emptyset$
3: for d= 0 to 255 do
4: p ← (q_i ‖ d ‖ w_2 ‖d‖…. ‖d ‖w_2)
5: if anom(p) = true then
6: $D_i \leftarrow D_i \cup \{ d \}$
7: end-if
8: end-for
9: end-for
10: return $D=\bigcap_{(i=1)}^{T} Di$

6. RESULT AND DISCUSSION

The proposed approach will show the highest accuracy for all type of attacks including internal as well external attacks. We develop the system in java with 5000 external and internal attacks. LOIC is use for remote attack generator. Table 1 shows the proposed system estimated detection rate with existing systems.

Table 6.1: Proposed System Estimated Detection Rate

Approach	Detection Method	Internal Detection Ration	External Attack detection Ratio
IDS using GA	Genetic evaluation with KDD	82%	56%
IDS using FGA	Fuzzy Genetic	86%	89%

IDS using data mining approach	K- means clustering	88%	85%
KIDS (Proposed Estimated)	Black box key recovery approach	95%	92%

All the attack stated allow us to recover some key bytes (we expect that they can be extended to full key-recovery staying within the same complexity).

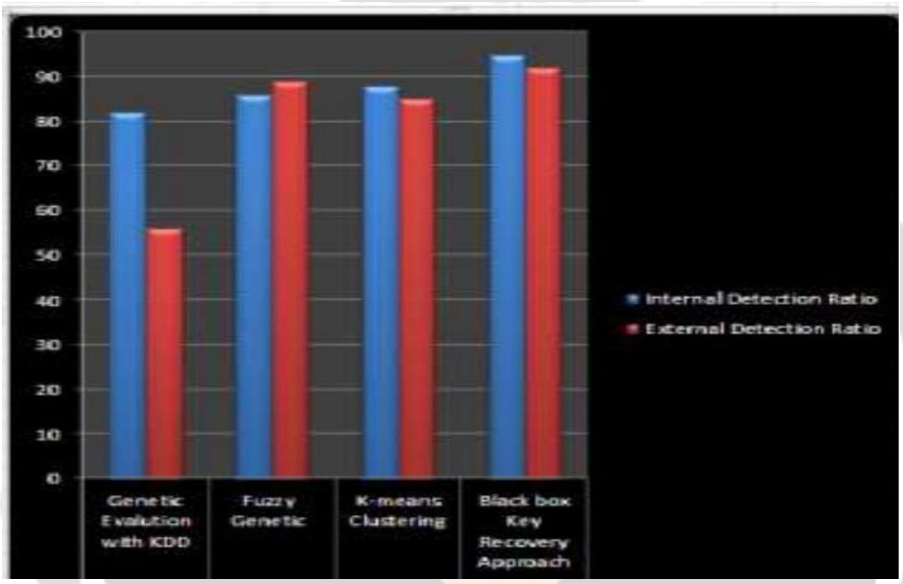

Fig -6.1: Proposed System Estimated Detection Rate

Table 6.2: Result Analysis

Sr.No	File Name	Precision	Recall	F-measure
1	kkk.txt	0.343028	0.656272	0.450713
2	xyz.txt	0.308566	0.628568	0.485326
3	abc.txt	0.506894	0.513682	0.658942
4	aaa.txt	0.308566	0.628568	0.485326
5	abcd.txt	0.259864	0.852645	0.725468

From the Table 6.2 we can show the result analysis graphically for this analysis parameter. Precision it the degree to which repeated measurements under unchanged condition shows the same result.

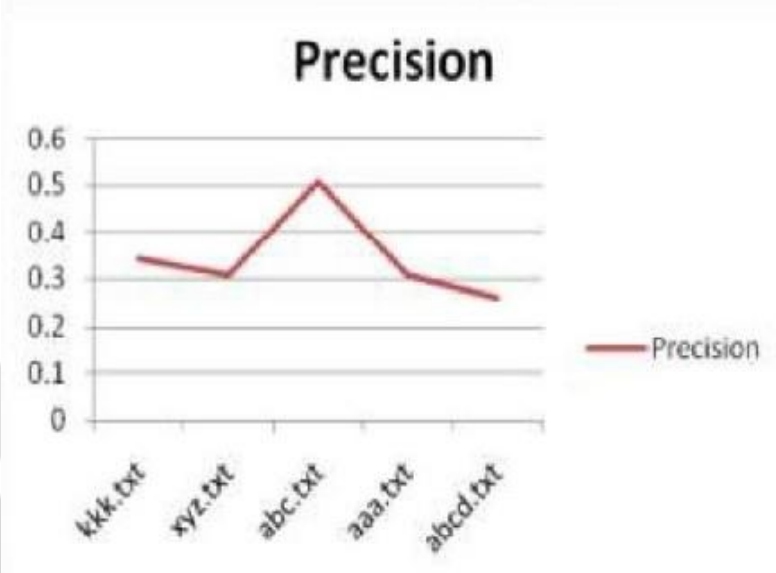

Fig 6.2 shows Precision Graph.

Recall is the fraction of the document that are relevant to the query that are successfully retrieved. This is shown in fig 6.3.

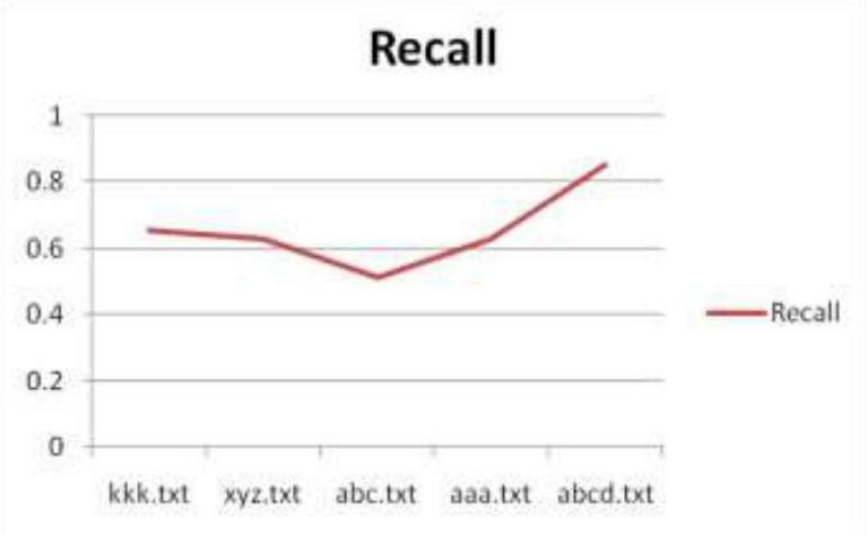

Fig -6.3: Result Analysis(Recall)

F-measure is the measure that combines precision and recall. F-measure is shown in fig 6.4.

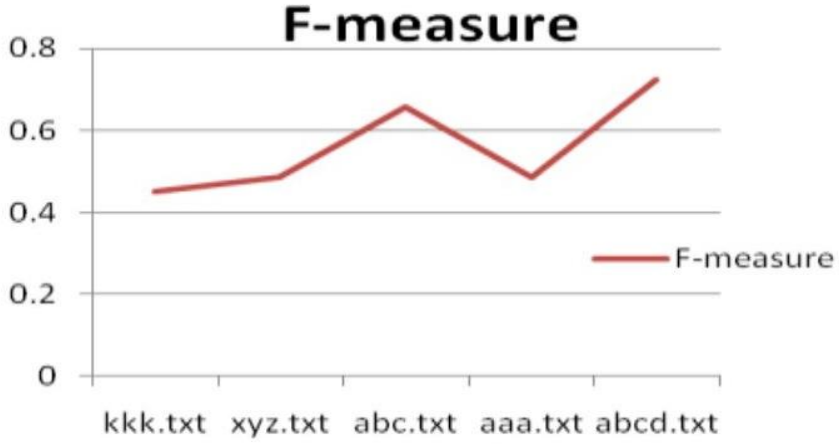

Fig -6.4: Result Analysis(F-measure)

7. CONCLUSION

This system is based on Key-Recovery on Black-Box KIDS & Key-Recovery on Gray-Box KIDS. KIDS system will offer a good platform to prevent information from leakage by regenerating key. This system will detect unauthorized user, and recover or regenerate key & Block respective unauthorized user. The focus in this work has been on recovering the key through efficient procedures.

8. ACKNOWLEDGEMENT

With all respect and gratefulness, I would like to thanks all people who have helped me directly or indirectly for the paper presentation. I am grateful to my guide, Prof. Hemant Gupta, for his guidance and support. I wish to express my sincere thanks to theto Prof. MayankBhattH.O.D. Department of Computer Science and Engineering) for their support. Lastly I would like to thank staff member ofLNCTS, Indore, M.P, India for making all the requirements possible and simply available whenever required.

9. 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 COMPUTINGVOL. 12, NO. 3 (MAY/JUNE 2015): 312-325.

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

[3]. M. Barreno, B. Nelson, R. Sears, A.D. Joseph, and J.D. Tygar. ”Can Machine Learning be Secure?” Proc. ACMSymp. Information, Computer (2006): 16-25.

[4]. B. Biggio, B. Nelson, and P. Laskov. ”Support Vector Machines Under Adversarial Label Noise.” J. MachineLearningResearch,vol. 20 (2011): 97-112.

[5]. B. Biggio, G. Fumera, and F. Roli. ”Adversarial Pattern Classification Using Multiple Classifiers and Randomisation.” Proc. IAPR Intl Workshop Structural, Syntactic, and Statistical Pattern Recognition (2008): 500-509.

[6]. B. Nelson, A.D. Joseph, S.J. Lee, and S. Rao. "Near-Optimal Evasion of Convex-Inducing Classifiers." J.Machine Learning Research,vol. 9 (2010): 549-556.

[7]. B. Nelson, B.I.P. Rubinstein, L. Huang, A.D. Joseph, and J.D.Tygar. "Classifier Evasion: Models and Open Problems." Proc. Intl ECML/PKDD Conf. Privacy and Security Issues in Data Mining and Machine Learning (PSDML 10) (2011): 92-98.

[8]. 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 (May 2012): 1293-1332.

[9]. K. Wang, G. Cretu, and S. Stolfo. "Anomalous Payload-Based Worm Detection and Signature Generation."Proc. Eighth Intl Conf. Recent Advances in Intrusion Detection (RAID 05) (2005): 227-246.

[10]. Meek, D. Lowd and C. "Adversarial Learning." Proc. 11th ACM SIGKDD Intl Conf. Knowledge Discovery in Data Mining (KDD 05) (2005): 641-647.

[11]. N. Dalvi, P. Domingos, Mausam, S. Sanghai, and D. Verma. "Adversarial Classification." Proc. 10th ACMSIGKDD Intl Conf. Knowledge Discovery and Data Mining (KDD 04) (2004): 99-108.

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

[13]. P. Fogla, M. Sharif, R. Perdisci, O. Kolesnikov, and W. Lee. "Polymorphic Blending Attacks ."Proc. 15thConf. USENIX Security (2006).

[14]. Paxson, R. Sommer and V. "Outside the ClosedWorld: On Using Machine Learning for Network Intrusion Detection." Proc. IEEE Symp. Security and Privacy (2010): 305-316.

[15]. R. Perdisci, D. Ariu, P. Fogla, G. Giacinto, and W. Lee. "McPAD: A Multiple Classifier System for Accurate Payload-Based Anomaly Detection." Computer Networks, vol. 5, no. 6(2009): 864-881 .

[16]. Rieck K. "Computer Security and Machine Learning: Worst Enemies or Best Friends?" Proc. DIMVAWorkshop Systems Security(SYSSEC) (2011).