|| Volume 3 || Issue 7 || July 2018 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

Secure k-NN Query Over Encrypted Data in KIDS

Pooja Mande[1], Prof. B. S. Kurhe.[2]
Student, ME Computer, SPCOE, Department Of Computer Engineering, Dumbarwadi[1]
Assistant Professor, SPCOE, Department Of Computer Engineering, Dumbarwadi[2]

Abstract— **-Nearest neighbours (k-NN) query aims at identifying k nearest points for a given query point in a dataset. In the past few years, researchers have proposed various methods to address the security and privacy problems of k-NN query on encrypted cloud data. Now a days, various schemes have been presented to support k-NN query on encrypted cloud data. However, prior works have all assumed that the query users (QUs) are fully trusted and know the key of the data owner (DO), which is used to encrypt and decrypt outsourced data. We present a novel scheme for secure k-NN query on encrypted cloud data with multiple keys, in which the DO and each QU all hold their own different keys, and do not share them with each other; meanwhile, the DO encrypts and decrypts outsourced data using the key of his own. Our proposed scheme is constructed by a distributed two trapdoors public-key cryptosystem (DT-PKC) and a set of protocols of secure two-party computation, which not only preserves the data confidentiality and query privacy but also supports the offline data owner. We have conducted extensive experiments to theoretical and experimental evaluations demonstrate the effectiveness of our scheme in terms of security and performance.**

Keywords*: k-Nearest Neighbours (k-NN), Distributed Two Trapdoors Public-Key Cryptosystem (DT-PKC).*

I INTRODUCTION

KIDS is an application layer network anomaly detection system that extracts a number of features from each payload. The impossibility for an attacker to recover the key under any reasonable adversarial model.Strictly speaking KIDS of "learning with a secret" Identifying Security problems.

These systems provide security for files by recovering the key. Also this system enables anomaly detection in Keyed Intrusion Detection System (KIDS). A few detection schemes proposed over the last few years have attempted to incorporate defences against evasion attacks. One such system is keyed intrusion detection system (KIDS) [1], introduced by Mrdovic and Drazenovic at DIMVA'10. KIDS is an application-layer network anomaly detection system that extracts a number of features ("words") from each payload. The system then builds a model of normality based both on the frequency of observed features and their relative positions in the payload. KIDS 'core idea to impede evasion attacks is to incorporate the notion of a "key", this being a secret element used to determine how classification features are extracted from the payload. The security argument here is simple: even though the learning and testing algorithms are public, an adversary who is not in possession of the key will not know exactly how a request will be processed and, consequently, will not be able to design attacks that thwart detection.

In this work, we make the following contributions:

1. We argue that any keyed anomaly detection system (or, more generally, any keyed classifier) must preserve one fundamental property: The impossibility for an attacker to recover the key under any reasonable adversarial model. We deliberately choose not to analyze how difficult is for an attacker to evade detection if the classifier is keyed. We believe that this is a related, but different problem.
2. We pose the key-recovery problem as one of adversarial learning. By adapting the adversarial setting proposed by Lowd and Meek [10] in a related problem (reverse engineering of a classifier), we introduce the notion of gray- and black-box key-recovery attacks.
3. We present two instantiations of such attacks for KIDS, one for each model. Our attacks take the form of query strategies that make the classifier leak some information about the key. Both are very efficient and show that KIDS does not meet the fundamental security property discussed above. Furthermore, we have implemented and experimentally confirmed the correctness of our attacks.
4. Building on related work in the broader field of secure machine learning , we pose some additional questions and provide constructive discussion about the suitability, limitations, and possible structure of keyed classifiers.
5. We construct a secure *k*NN scheme with multiple keys. And we show that the proposed scheme is secure under the standard semi-honest model [19]. Also, we demonstrate the practical applicability of our solution through extensive experiments using a real world dataset. The remainder is structured as follows. Section 2 reviews the related work. Section 3 define the problem definition of our proposed system. Sections 4 introduce the proposed work of system, followed by the conclusion and future work in Section 5.

II RELATED WORK

In existing system, the problem of computing optimal strategies to modify an attack so that it evades detection by a Bayes classifier. They formulate the problem in game-theoretic

terms, where each modification made to an instance comes at a price, and successful detection and evasion have measurable utilities to the classifier and the adversary, respectively. The authors study how to detect such optimally modified instances by adapting the decision surface of the classifier, and also discuss how the adversary might react to this.

A. Classifier Evasion and Adversarial Learning

Dalvi et al. explored in [5] the problem of computing optimal strategies to modify an attack so that it evades detection by a Na€ıve Bayes classifier. They formulate the problem in game-theoretic terms, where each modification made to an instance comes at a price, and successful detection and evasion have measurable utilities to the classifier and the adversary respectively. The authors study how to detect such optimally modified instances by adapting the decision surface of the classifier, and also discuss how the adversary might react to this. The setting used in assumes an adversary with full knowledge of the classifier to be evaded. Shortly after, P. Fogla [6] studied 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 attacks, instead of looking for optimal strategies. The authors use a membership oracle as implicit adversarial model: the attacker is given the opportunity to query the classifier with any chosen instance to determine whether it is labeled as malicious or not. Consequently, a reasonable objective is to find instances that evade detection with an affordable number of queries. A classifier is said to be ACRE learnable if there exists an algorithm that finds a minimal-cost instance evading detection using only polynomially many queries. Similarly, a classifier is ACRE-k learnable if the cost is not minimal but bounded by k. Among the results given in [10], it is proved that linear classifiers with continuous features are ACRE k-learnable under linear cost functions. Therefore, these classifiers should not be used in adversarial environments. Subsequent work by Nelson et al. generalizes these results to convex-inducing classifiers, showing that it is generally not necessary to reverse engineer the decision boundary to construct undetected instances of near minimal cost.

For the interested reader, Nelson et al. [13] have recently surveyed some open problems and challenges related to the classifier evasion problem. More generally, some additional works have revisited the role of machine learning in security applications, with particular emphasis on anomaly detection [14], [15].

B. Strategies to Thwart Evasion

Kolesnikov et al. [9] demonstrate that polymorphic mimicry worms, based on encryption and data encoding to obfuscate their content, are able to evade frequency distribution-based anomaly detectors like PAYL [12]. PAYL models byte-value frequency distributions (i.e., 1-grams), 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, 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 kinds of randomization, namely randomized modeling and randomized testing. In the former, packets are split into several substrings using a randomly-generated bitmask. Substrings coming from the same packet position are modeled and tested separately. Since the bitmask is kept secret, an attacker only succeeds if he manages to craft an attack vector such that the data is normal with respect to any randomly selected portion of a packet. This clearly makes evasion harder, but substantially increases the overhead of the IDS. Alternatively, randomized testing also partitions packets randomly into several chunks, but tests each of them against the same classifier, which does not incur any substantial overhead. Randomization and/or using an ensemble of classifiers have also been proposed in the context of spam detection. For example, Biggio et al. [3],[4] studied how to introduce randomness in the design of the classifier, preventing the adversary from having exact knowledge about one or more system parameters. A similar approach was presented by Gates et al. in [7]. The work in [3] uses multiple classifiers and randomly chooses the weights assigned to each classifier in the decision. The task for the attacker is much harder then, since he can never guess the detector's configuration. The main problem of this strategy is that it can influence negatively the overall detection performance, particularly increasing the false positive rate. A. Kolcz et al. [8] presented similar strategies to thwart good-word attacks on spam filters. Their scheme transforms each email into a bag of multiple segments (instances), and then applies multiple-instance logistic regression to the bags. An email is classified as spam if at least one instance in the corresponding bag is spam; otherwise it is marked as legitimate. This bags-of-words strategy performs better than single-instance learners such as support vector machines (SVMs) or Na€ıve Bayes. A similar approach was explored in [11] to detect masquerade mimicry attacks.

C. Secure Machine Learning

Barreno et al. [2] have pondered on the risks of applying machine learning algorithms to security domains. They introduce a taxonomy that groups attacks on machine learning systems into different categories, depending on whether the adversary influences training or just analyses an already trained system; whether the goal is to force just one misclassification or else to generate too many so the system becomes unusable; etc.

The authors also provide useful discussion on potential countermeasures and enumerate various open problems.

III PROBLEM STATEMENT

We propose a novel scheme for secure *k*-NN query on encrypted cloud data with multiple keys, in which the DO and each QU all hold their own different keys, and do not share them with each other; meanwhile, the DO encrypts and decrypts outsourced data using the key of his own. Our scheme is constructed by a distributed two trapdoors public-key cryptosystem (DT-PKC) and a set of protocols of secure two-party computation, which not only preserves the data confidentiality and query privacy but also supports the offline data owner. Our extensive theoretical and experimental evaluations demonstrate the effectiveness of our scheme in terms of security and performance.

A proposed system KIDS for recovering of key. Our work shows that recovering the key is extremely simple provided that the attacker can interact with KIDS and get feedback about probing requests. Keyed Intrusion Detection System, a key dependent network anomaly detector that inspects packet payloads.

IV PROPOSED WORK

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. I believe that such a lack of security reveals that schemes like kids were simply not designed to prevent key-recovery attacks. However, in this paper I 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. I have provided discussion on this and other open questions in the hope of stimulating further research in this area. The attacks 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. Thus, our recommendation for future designs is to base decisions on robust principles rather than particular fixes.

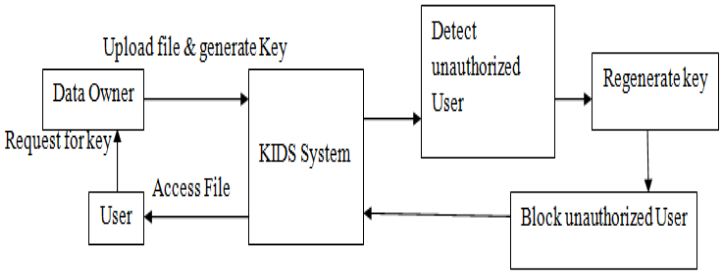

Figure 1 System Architecture

Advantages of Proposed System:

- Attacks are extremely efficient, showing that it is reasonably easy for an attacker to recover the key in any of the two settings discussed.
- It provides more security than previous system by recovering keys of file.
- Prevent leakage of information.

V CONCLUSION AND FUTURE WORK

This system is based on Key-Recovery on Black-Box KIDS & Key-Recovery on Gray-Box KIDS. This 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 recommendation for future designs is to base decisions on robust principles rather than particular fixes. The focus in this work has been on recovering the key through efficient procedures, demonstrating that the classification process leaks information about it that can be leveraged by an attacker.

ACKNOWLEDGEMENT

I express my sincere thanks to my project guide Prof. B. S. Kurhe who always being with presence & constant, constructive criticism to made this paper. 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." IEEE Transactions On Dependable And Secure Computing Vol. 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, pp. 121- 148, 2010.

[3] 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.

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

[5] 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.

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

[7] C. Gates and C. Taylo, "Challenging the Anomaly Detection Paradigm: A Provocative Discussion," *Proc. New Security*

Paradigms Workshop (NSPW), pp. 21-29, 2006.

[8] A. Kolcz and C.H. Teo, "Feature Weighting for Improved Classifier Robustness," *Proc. Sixth Conf. Email and Anti-Spam (CEAS '*09*),* 2009.

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

[10] 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.

[11] Metasploit Framework, www.metasploit.com, 2013.

[12] S. Mrdovic and B. 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.

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

[14] 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.

[15] 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.