Recognize and evaluate security framework of classifier under attack

Miss. Kale Tai
Dept. of Computer Engineering
DGOIFOE, Bhigwan
Pune, India

ABSTRACT—Pattern classification is a part of machine learning that focuses on recognition of patterns and Pattern classification system are used for the biometric authentication, spam filtering, and network intrusion detection. The biometric authentication, spam filtering, and network intrusion detection is an adversarial applications. Biometric system is a tool for person identification and verification. The paper proposed, evaluation the security of pattern classifiers that formalizes and generalizes the main ideas proposed in the literature and give examples of its use in three real applications. The paper proposes a framework for evaluation of pattern security, model of adversary for defining any attack scenario. The paper presents a design of a system for identification and design security evaluation framework of classifier under attack using biometric system, biometric authentication System, spam filtering, and network intrusion detection.

Keywords-*Adversarial Classification, Adversarial Scenario, Pattern classification, Security Evaluation*

I. INTRODUCTION

Pattern classification is a part of machine learning algorithms and it is used for security related application like biometric authentication, network intrusion detection and spam filtering. There are various types of attacks like spoofing attack, Modifying packet at the time of Data transmission, Spam attacks. The Spoof attacks consist of a give fake input biometric traits to biometric systems, and use the benefit of the system [1][2].

Adversarial machine learning is an important research field that lies at the intersection of machine learning and computer security [1][2][9]. It aims to enable the safe adoption of machine learning techniques in oppositional settings like spam filtering, malware detection and biometric recognition. Examples include: attacks in spam filtering, where spam messages are obscured through misspelling of bad words or insertion of good words; attacks in computer security, e.g., to complicate malware code within network packets or mislead signature detection; attacks in biometric recognition, where fake biometric characters [5][6] may be exploited to impersonate a legitimate user (biometric spoofing) or to compromise users' template galleries that are adaptively updated over time.[8] To understand the security properties of learning algorithms in oppositional situations, one should address the following main issues:

1. identifying potential vulnerabilities of machine learning algorithms during learning and classification;
2. developing suitable attacks that correspond to the identified threats and evaluating their effect on the targeted system;
3. Proposing countermeasures to improve the security of machine learning algorithms against the considered attacks.

The paper presents a design of a system for recognizing and evaluating security framework of classifier under attack using biometric system, biometric authentication System, spam filtering, and network intrusion detection.

II. RELATED WORK

Biometric systems are useful tools for person identification and verification [1][3]. The biometric System is any physiological of behavioral trait of a people that can be used to different that people from other people. Biometric system is basically use for the identification of user. The multimodal Fusion Vulnerability to Non-Zero Effort (Spoof) Imposters paper is describe the contribution of multi-biometric system and fusion algorithms [3]. Multi-biometric system is nothing but its identify user using multi-model database like face, iris, and fingerprint match scores. The using Multiple Modalities with Enhanced Fusion, the performance is Medium and security is high. This paper proposes a method for determining the best practices for using multimodal fusion to

minimize spoof attacks. A new performance measure, SFAR (spoof false accept rate)[3], is introduced to represent conditions of a partial spoof attack. It is shown that after a system assessment based on SFAR (spoof false accept rate) is conducted, a calculated adjustment of the operating point can ensure for a more secure system, at a cost of decreased FRR (False reject rate), performance. While intuitively expected, the paper quantitatively demonstrate how to assess the tradeoff [1][2][3].

P. Fogla, M. Sharif, R. Perdisci, O. Kolesnikov, and W. Lee" Proposed [4] the describe in "Polymorphic Blending Attacks" that A very effective means to evade signature-based intrusion detection systems (IDS) is to employ polymorphic techniques to generate attack instances that do not share a fixed signature. Polymorphic attack is an attack that is able to change Data at time of data transmission. There are three component of polymorphism attack: Attack vector, Attack body, polymorphism Decryption. An attack vector is used for exploiting the vulnerability of the target host. the code that performs the intended malicious actions after the vulnerability is exploited. Polymorphism decrypted contains the part of the code that decrypts the shell code. Anomaly-based intrusion detection systems provide good defense because existing polymorphic techniques can make the attack instances look different from each other, but cannot make them look like normal [3][4]. The paper introduce a new class of polymorphic attacks, called polymorphic blending attacks, that can effectively evade byte frequency-based network anomaly IDS by carefully 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. We take a systematic approach to the problem and formally describe the algorithms and steps required to carry out such attacks. [3][4] We not only show that such attacks are feasible but also analyze the hardness of evasion under different circumstances [3][4][6]. We present detailed techniques using PAYL, a byte frequency-based anomaly IDS, as a case study and demonstrate that these attacks are indeed feasible. We also provide some insight into possible countermeasures that can be used as defence [3][4][5].

Z. Akhtar, B. Biggio, G. Fumera, and G. Luca this author describe the some concept about the pattern classifier in Robustness of Multi-modal Biometric Systems under Realistic Spoof Attacks against All Traits. In that Spoof attacks is give the input i.e fake biometric traits to biometric systems.[1][2] In this Paper Multi- biometric System is implementing using the face and a fingerprint matcher. The realistic spoofing attacks provided evidence of two common beliefs about the robustness of multi-modal biometric systems. First, they can be more robust than each corresponding mono-modal system, even in the case when all biometric traits are spoofed. Second, their performance under a spoofing attack against all traits is still unacceptable for security applications[1][2]. In other words, they can be cracked by spoofing all the fused traits, even when the attacker is not able to fabricate an exact replica of the genuine user's traits [1][2].

Existing System is basically based on the adversarial learning system and it can be categorized according to the two main steps, the **pro-active arms** race and the **re-active arms** race. In this type, the classifier designer reacts to the attack by analyzing its effects and grows the countermeasures. In Pro- Active Arm race the classifier designer and the adversary's attempt to accomplish their aims by behaving to the changing component of competitor. In this type, the classifier designer can anticipate the adversary by simulating the potential attacks, evaluating their effects and developing the countermeasures if necessary. The 'Re-active' approaches, neither anticipates the new security vulnerabilities, nor they bid to forecast future attacks. Computer security guidelines accordingly advocate a 'Pro-active' approach in which the classifier designer also attempts to anticipate the adversary's stratagem by (i) repeating this process before system deployment, (ii) devising proper countermeasures ,when required, and (iii) identifying the relevant threats.

III. PROPOSED SYSTEM

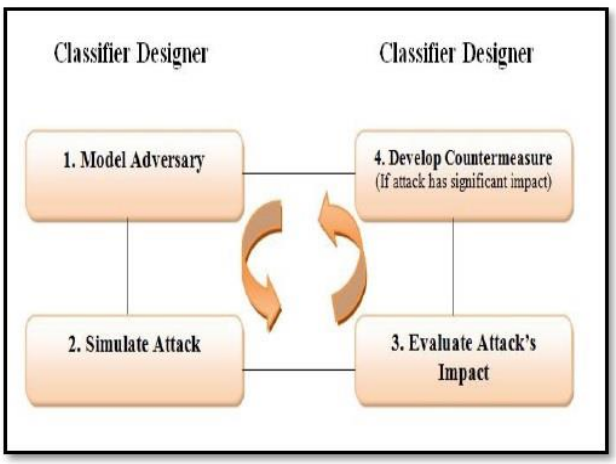

Fig 1: System Architecture

.

Modules:

1. Attack Scenario and Model of the Adversary
2. Pattern Classification
3. Adversarial classification:
4. Security modules

1. Attack Scenario and Model of the Adversary:

Even if the definition of attack scenarios is eventually an application-specific matter, it is possible to give general guidelines that can help the designer of a pattern recognition system. Here we propose to specify the attack situation in terms of a theoretical model of the opponent that includes, joins, and extends different thoughts from previous work. Our model is based on the assumption that the opponent acts rationally to attain a given goal, according to her knowledge of the classifier, and her capability of manipulating data. This allows one to derive the corresponding best attack strategy [1].

2. Pattern Classification:

Multimodal biometric systems for personal identity recognition have received great interest in the past few years. It has been shown that combining information coming from different biometric traits can overcome the limits and the weaknesses inherent in every individual biometric, resulting in a higher accuracy. Moreover, it is commonly believed that multimodal systems also improve security against Spoofing attacks, which consist of claiming a false identity and submitting at least one fake biometric trait to the system (e.g., a "gummy" fingerprint or a photograph of a user's face). The reason is that, to evade multimodal system, one expects that the adversary should spoof all the corresponding biometric traits. In this application example, we show how the designer of a multimodal system can verify if this hypothesis holds, before deploying the system, by simulating spoofing attacks against each of the matchers [1].

3. Adversarial classification:

Assume that a classifier has to discriminate between legitimate and spam emails on the basis of their textual content, and that the bag-of-words feature representation has been chosen, with binary features denoting the occurrence of a given set of words [1].

4. Security modules:

Intrusion detection systems analyze network traffic to prevent and detect malicious activities like intrusion attempts, ROC curves of the considered multimodal biometric system under a simulated spoof attack against the fingerprint or the face matcher. Port scans, and denial-of-service attacks. When suspected malicious traffic is detected, an alarm is raised by the IDS and subsequently handled by the system administrator. Two main kinds of IDSs exist: misuse detectors and anomaly-based ones. Misuse detectors match the analyzed network traffic against a database of signatures of known malicious activities. The main drawback is that they are not able to detect never-before-seen malicious activities, or even variants of known ones. To overcome this issue, anomaly-based detectors have been proposed. They build a statistical model of the normal traffic using machine learning techniques, usually one-class classifiers, and raise an alarm when anomalous traffic is detected. Their training set is constructed, and periodically updated to follow the changes of normal traffic, by collecting unsupervised network traffic during operation, assuming that it is normal (it can be filtered by a misuse detector, and should) [1].

Advantage:

- Proposed system oppose to developing novel methods to assess classifier security against these attacks.
- The presence of an intelligent and adaptive adversary makes the classification problem highly non-stationary.

IV. CONCLUSION

This paper presented an overview of work related to the security of pattern classification systems with the goal of imparting useful guidelines on how to improve their design and assess their security specific attacks. Also the paper focused on innovative security evaluation of pattern classifiers that deployed in adversarial environments. Main contribution is a framework for verifiable security evaluation that construes and establishes the notion from previous work, and can be utilized to different classifiers, learning algorithms, and classification tasks.

In the future, clustering methods can be integrated with the existing technique in order to get better results. Further, this approach can be applied to the

application which makes the classification problem highly non-stationary.

V. ACKNOWLEDGMENT

I want to thank all people who help me in different way., I thank to "IJRAR" who have given opportunity to present my paper.

VI. REFERENCES

1. B. Biggio, G. Fumera, , and F. Roli, "Security Evaluation of Pattern Classifiers under Attack" Ieee Transactions On Knowledge And Data Engineering, Vol. 26, No. 4, April 2014
2. R.N. Rodrigues, L.L. Ling, and V. Govindaraju, "Robustness of Multimodal Biometric Fusion Methods against Spoof Attacks," J. Visual Languages and Computing, vol. 20, no. 3, pp. 169-179, 2009.
3. P. Johnson, B. Tan, and S. Schuckers, "Multimodal Fusion Vulnerability to Non-Zero Effort (Spoof) Imposters," Proc. IEEE Int'l Workshop Information Forensics and Security, pp. 1-5, 2010.
4. P. Fogla, M. Sharif, R. Perdisci, O. Kolesnikov, and W. Lee, "Polymorphic Blending Attacks," Proc. 15th Conf. USENIX Security Symp., 2006.
5. G.L. Wittel and S.F. Wu, "On Attacking Statistical Spam Filters," Proc. First Conf. Email and Anti-Spam, 2004.
6. D. Lowd and C. Meek, "Good Word Attacks on Statistical Spam Filters," Proc. Second Conf. Email and Anti-Spam, 2005.
7. A. Kolcz and C.H. Teo, "Feature Weighting for Improved Classifier Robustness," Proc. Sixth Conf. Email and Anti-Spam, 2009.
8. D.B. Skillicorn, "Adversarial Knowledge Discovery," IEEE Intelligent Systems, vol. 24, no. 6, Nov./Dec. 2009.
9. D. Fetterly, "Adversarial Information Retrieval: The Manipulation of Web Content," ACM Computing Rev., 2007.
10. R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification. Wiley-Interscience Publication, 2000.
11. N. Dalvi, P. Domingos, Mausam, S. Sanghai, and D. Verma, "Adversarial Classification," Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 99-108, 2004.
12. 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), pp. 16-25, 2006.
13. A.A. C_ardenas and J.S. Baras, "Evaluation of Classifiers: Practical Considerations for Security Applications," Proc. AAAI Workshop Evaluation Methods for Machine Learning, 2006.
14. P. Laskov and R. Lippmann, "Machine Learning in Adversarial Environments," Machine Learning, vol. 81, pp. 115-119, 2010.
15. L. Huang, A.D. Joseph, B. Nelson, B. Rubinstein, and J.D. Tygar, "Adversarial Machine Learning," Proc. Fourth ACM Workshop Artificial Intelligence and Security, pp. 43-57, 2011.
16. M. Barreno, B. Nelson, A. Joseph, and J. Tygar, "The Security of Machine Learning," Machine Learning, vol. 81, pp. 121-148, 2010.
17. D. Lowd and C. Meek, "Adversarial Learning," Proc. 11th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 641-647, 2005.
18. P. Laskov and M. Kloft, "A Framework for Quantitative Security Analysis of Machine Learning," Proc. Second ACM Workshop Security and Artificial Intelligence, pp. 1-4, 2009.
19. NIPS Workshop Machine Learning in Adversarial Environments for Computer Security, http://mls-nips07.first.fraunhofer.de/, 2007.
20. Dagstuhl Perspectives Workshop Mach. Learning Methods for Computer Sec., http://www.dagstuhl.de/12371/, 2012.
21. A.M. Narasimhamurthy and L.I. Kuncheva, "A Framework for Generating Data to Simulate Changing Environments," Proc. 25th Conf. Proc. the 25th IASTED Int'l Multi-Conf.: Artificial Intelligence and Applications, pp. 415-420, 2007.
22. S. Rizzi, "What-If Analysis," Encyclopedia of Database Systems, pp. 3525-3529, Springer, 2009.
23. J. Newsome, B. Karp, and D. Song, "Paragraph: Thwarting Signature Learning by Training Maliciously," Proc. Ninth Int'l Conf. Recent Advances in Intrusion Detection, pp. 81-105, 2006.
24. A. Globerson and S.T. Roweis, "Nightmare at Test Time: Robust Learning by Feature Deletion," Proc. 23rd Int'l Conf. Machine Learning, pp. 353-360, 2006.