DATA POISONING

group 213

April 21, 2017

1. Nafuna mereth Gambwa 14/u/23611/eve 211007767

2. Ajwang Mirriam Semmy 14/u/4621/eve 214007408

0.1 Introduction

With the advent of the modern internet, the number of inter connected users and devices, along with the available number of services, has tremendously increased. This has not only simpliïňĄed our lives, through accessibility and ease of use of novel services but it has also provided great opportunities for attackers to perform novel and proïňĄtable malicious activities such as data poisoning [1]âĂŞ[6].

0.2 Background to the Problem

As advocated in a recent workshop [6], poisoning attacks are an emerging security threat for data-driven technologies, and could become the most relevant one in the coming years, especially in the so called big data scenario dominated by data driven technologies. From a practical perspective, poisoning attacks are already a pertinent scenario in several applications. For instance, in collaborative spam ïňĄltering, classiïňĄers are retrained on emails labeled by end users. Attackers owning an authorized email account protected by the same anti-spam ïňĄlter may thus arbitrarily manipulate emails in their inbox, i.e., part of the training data used to update the classiïňĄer. Some systems may even ask directly to users to validate their decisions on some submitted samples, and use their feedback to update the classiïňĄer (see, e.g., PDF Rate, an online tool for detecting PDF malware designed by Smutz [9]. Furthermore, in several cases obtaining accurate labels, or validating the available ground truth may be expensive and time consuming; e.g., if malware samples are collected from the Internet, by means of honeypots, i.e., machines that purposely expose known vulnerabilities to be infected by malware [8], or other online services, like VirusTotal,1 labeling errors are possible.

0.3 Problem Statement

The problem this project will address is data poisoning, which is an emerging security threat for data driven technologies and could become the most relevant one in years to come hence the need to investigate how anonymous credential and blind authorization schemes may be applied to prevent spam and enforce accountability in large scale network security data collection and come up with a technology that will help solve this in real life.

0.4 Objectives

0.4.1 Main Objective

To develop an application that will reduce on data poisoning by detecting any anonymous credential and blind authorization on any data driven technology for the benefit of improving on the privacy, integrity and availability of data.

0.4.2 Other Objective

- To collect and analyze data about how anonymous credential and blind authorization schemes may be applied to prevent spam and enforce accountability in large scale network security data collection.

- To design a technology that will detect any malicious malware in the data so as to address the problem of data poisoning.

- To implement an application that will prevent data poisoning.

- To test and validate the prototype technology.

0.5 Methodology

To achieve our objectives, we shall use; qualitative and quantitative methods. Qualitative method shall be used to manage qualitative data for understanding and explaining social phenomena for example data from interviews, documents, observations. Quantitative methods shall be used to study natural phenomena. For example Numerical methods

0.6 Outcomes

Since the attackerâĂŹs goal is deïňĄned in terms of the desired security violation, which can be an integrity, availability, or privacy violation, and of the attack speciïňĄcity, which can be targeted or indiscriminated [4], [7]. Therefore, at the end of this project, will be able to address the security violations on data since, Integrity is violated if malicious activities are performed without compromising normal system operation, Availability is violated if the functionality of the system is compromised, causing a denial of service and Privacy is violated if the attacker is able to obtain information about the systemâĂŹs users by reverse engineering the attacked system.

0.7 References

[1] Sahami,M., Dumais,S., Heckerman,D., and Horvitz,E. A Bayesian approach to ïňĄltering junk e-mail. AAAI Technical Report WS-98-05, Madison, Wisconsin, 1998.

[2] Rubinstein, Benjamin I.P., Nelson, Blaine, Huang, Ling, Joseph, Anthony D., Lau, Shinghon, Rao, Satish, Taft, Nina, and Tygar, J. D. Antidote: understanding and defending against poisoning of anomaly detectors. In 9th Internet Meas. Conf., pp. 1âĂŞ14, 2009. ACM.

[3] Barreno, Marco, Nelson, Blaine, Joseph, Anthony, and Ty- gar, J. The security of machine learning. Machine Learning, 81:121âĂŞ148, 2010.

[4] SrndiÂťc, Nedim and Laskov, Pavel. Detection of malicious pdf ïňĄles based on hierarchical document structure. In NDSS. The Internet Society, 2013.

[5] Huang, L., Joseph, A. D., Nelson, B., Rubinstein, B., and Tygar, J. D. Adversarial machine learning. In ACM Workshop on ArtiïňĄcial Intell. and Sec., pp. 43âĂŞ57, Chicago, IL, USA, 2011.

[6] Joseph, Anthony D., Laskov, Pavel, Roli, Fabio, Tygar, J. Doug, and Nelson, Blaine. Machine Learning Methods for Computer Security (DagstuhlPerspectivesWork- shop 12371). Dagstuhl Manifestos, 3(1):1âĂŞ30, 2013.

[7] Barreno, Marco, Nelson, Blaine, Sears, Russell, Joseph, Anthony D., and Tygar, J. D. Can machine learning be secure? In ASIACCS, pp.16âĂŞ25,NY,USA, 2006.ACM.

[8] Spitzner, Lance. Honeypots: Tracking Hackers. Addison- Wesley Professional, 2002.

[9] Smutz, Charles and Stavrou, Angelos. Malicious PDF de- tection using metadata and structural features. In 28th Annual Computer Security Applications Conf., pp. 239âĂŞ 248, 2012. ACM.