INSO - Industrial Software
Institute of Computer Aided Automation | Faculty of Informatics | Vienna University of Technology

Expose for Master Thesis

Detecting VoIP Attacks using Data Mining and active Monitoring

Advisor: Thomas Grechenig

Simon Steyskal
0828067
066 926
Business Informatics

October 31st, 2012

.....
Thomas Grechenig

Expose for Master Thesis

1 Problem Description

The number of Voice over IP (VoIP) services is increasing rapidly. More and more users get from public switched telephone networks and other traditional mobile networks to the cheaper and more convenient internet telephony, which therefore gains more and more popularity. The drawback of that popularity is the respective interest of hackers for VoIP and its services.

To protect VoIP networks and its users it is necessary to reduce the amount of damage, which malicious network traffic and respective attacks perform against VoIP systems.

The major problems someone is faced with if he wants to obtain the goal of a secure VoIP network is to recognize an attack in the first place and do that as fast as possible. The more time an attacker has to infiltrate a network the more damage he is able to do.

For this purpose a new approach to predict attacks on a VoIP network in an early stage of the attack would be necessary, for example by using data mining and machine learning techniques.[2]

The question whether such an approach using the aforementioned technologies for detecting VoIP attacks is realizable, should be explored in this work.

2 Expected Result

Hypothesis is, that it is possible to use data mining and machine learning techniques to detect malicious behavior in the traffic of a VoIP network especially in an early stage of such an attack.

The aim of this thesis is to develop a general concept for using data minig and machine learning techniques to detect threats of VoIP networks and evaluate the applicability of such a concept for selected attacks. The real-world sample traffic, which was gathered using a VoIP Honeynet ensures an adequate sample of attack patterns.

The resulting concept can then be used to define classification patterns for various VoIP attacks. Thus, with these patterns it is possible to create a new or extend existing Intrusion Detection Systems(IDS) to detect intruders in an early stage of their attacks by monitoring the traffic of a VoIP network.

3 Method and Approach

After doing extensive research in the main topics of this master thesis like VoIP, Honeynets and data mining, which also will be briefly described in the first chapters of this work, VoIP datapackets as well as VoIP traffic will be analyzed to find classifiable attributes and informations for applying data mining and machine learning techniques. For this purpose, traffic of *Real-World Attacks* gathered in a VoIP Honeynet will be used.[3]

To define a general concept for using data mining and machine learning techniques to detect VoIP attacks, the next step is to perform research on classification and clustering algorithms as well as machine learning strategies to find suitable techniques for this purpose.

Using that concept, forecasting patterns for selected threats will be defined to prove the concepts applicability.

In the end an evaluation of the aforementioned forecasting patterns by using real-world sample traffic is performed and its success rate especially in terms of false positive results will be checked.

4 State of the Art

In the last few years several techniques were investigated to form frameworks for Network Intrusion Detection Systems (NIDS).

Especially to face the fast changing of networks and occurance of unknown intrusion patterns. For this purpose heuristic analysis models [6] as well as neural networks, decision trees and support vector machines were probed [10, 7]. But also database centric solutions are conceivable as shown in [1].
Another approach are signature- or rule-based IDS which compare the behavior of network participants with a database of stored malicious behavior patterns. Dynamic and adaptive solutions to learn signatures for network intrusion detection using a supervised learning classifier system based on genetic algorithms are published in [9, 5, 4].

There has been also some research done in the field of using data mining with honeynets.

An interesting IDS which is able to handle known and unkown attacks is discussed in [8]. It splits the attacks while entering the honeynet, known attacks are led to an intrusion detection module which is based on misuse and can process in time, whereas the unkown attacks are processed by the data mining based classification analysis algorithm module, which makes new rules through learning and updates the rule for misuse detection.

5 Connection to Business Informatics

There are several lectures in the bachelor and master curriculum of Business Informatics, which have their focus on topics related to those of this master thesis.

Topic: Data Mining
Lecture: Business Intelligence(188.429)
General introduction into data mining, applications of data mining and several data mining techniques.

Topic: Machine Learning
Lecture: Machine Learning(184.702)
Principles of supervised and unsupervised machine learning, including pre-processing and data preparation, as well as evaluation of learning systems. Furthermore the discussion of various machine learning models including e.g. *Decision Tree Learning, Bayesian Networks, SVM, HMM,..*

Topic: Security
Lecture: Several Security Lectures
Gainining a basic understanding about security issues, as well as *TCP/IP security, distributed systems security, firewalls and traffic filtering, IDS, network discovery/vulnerability scanning,...*

6 Time Schedule

The primarly research in the main topics should be finished within the first weeks of November. After that the introduction chapter will be written and should not take more than one and a half months to finalize.

The main part of this thesis, the concept of using data mining for detecting VoIP attacks as well as the evaluation will need most of the time but has to be finished within the winter term and the following semester break.

In the end the reviewing phase is estimated with four weeks, followed by the administrative actions necessary to finish the diploma thesis.

Estimated working time: **5-6 months**

7 Contents

References

[1] R Chetan and D V Ashoka. "Data mining based network intrusion detection system: A database centric approach". In: *2012 International Conference on Computer Communication and Informatics* (2012), pp. 1–6.

[2] Deepthy K Denatious and Anita John. "Survey on data mining techniques to enhance intrusion detection". In: *2012 International Conference on Computer Communication and Informatics* 04329022 (2012), pp. 1–5.

[3] Markus Gruber et al. "Security Status of VoIP Based on the Observation of Real-World Attacks on a Honeynet". In: *2011 IEEE Third Intl Conference on Privacy Security Risk and Trust and 2011 IEEE Third Intl Conference on Social Computing* (2011), pp. 1041–1047.

[4] Mohammad Sazzadul Hoque et al. "An Implementation of Intrusion Detection System Using Genetic Algorithm". In: *International Journal of Network Security Its Applications* 4.2 (2012), pp. 109–120. ISSN: 09752307.

[5] R Rangadurai Karthick, Vipul P Hattiwale, and Balaraman Ravindran. "Adaptive network intrusion detection system using a hybrid approach". In: *2012 Fourth International Conference on Communication Systems and Networks COMSNETS 2012* (2012), pp. 1–7.

[6] Wu Kehe, Ding Tao, and He Jianping. "The research of intrusion detection technology based on heuristic analysis". In: *2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference* 1 (2011), pp. 164–166.

[7] Yu-Xin Meng. "The practice on using machine learning for network anomaly intrusion detection". In: *2011 International Conference on Machine Learning and Cybernetics* 2 (2011), pp. 576–581. ISSN: 2160133X.

[8] Dong Nanping, Zhou Guanling, and Wang Yuping. "The Optimal Application of the Algorithms of Detection and Data Mining in Honeynet". In: *2009 IITA International Conference on Control Automation and Systems Engineering* (2009), pp. 13–16.

[9] Kamran Shafi and Hussein A Abbass. "An adaptive genetic-based signature learning system for intrusion detection". In: *Expert Systems with Applications* 36.10 (2009), pp. 12036–12043. ISSN: 09574174.

[10] Hua Tang and Zhuolin Cao. "Machine Learning-based Intrusion Detection Algorithms". In: *Neural Networks* 6 (2009), pp. 1825–1831.