



## Discovering fuzzy association rule patterns and increasing sensitivity analysis of XML-related attacks

Gaik-Yee Chan<sup>a,1</sup>, Chien-Sing Lee<sup>a,\*,2</sup>, Swee-Huay Heng<sup>b,3</sup>

<sup>a</sup> Faculty of Information Technology, Multimedia University, Cyberjaya, Malaysia

<sup>b</sup> Faculty of Information Science & Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia

### ARTICLE INFO

#### Article history:

Received 8 June 2011

Received in revised form

9 July 2012

Accepted 14 November 2012

Available online 12 December 2012

#### Keywords:

Intrusion detection

Intrusion prevention

Fuzzy association rules

Business intelligence

e-commerce

### ABSTRACT

Most active research in Host and Network-based Intrusion Detection (ID) and Intrusion Prevention (IP) systems are only able to detect and prevent attacks of the computer systems and attacks at the Network Layer. They are not adequate to countermeasure XML-related attacks. Furthermore, although research have been conducted to countermeasure Web application attacks, they are still not adequate in countering SOAP or XML-based attacks. In this paper, a predictive fuzzy association rule model aimed at segregating known attack patterns (such as SQL injection, buffer overflow and SOAP oversized payload) and anomalies is developed. First, inputs are validated using business policies. The validated input is then fed into our fuzzy association rule model (FARM). Consequently, 20 fuzzy association rule patterns matching input attributes with 3 decision outcomes are discovered with at least 99% confidence. These fuzzy association rule patterns will enable the identification of frequently occurring features, useful to the security administrator in prioritizing which feature to focus on in the future, hence addressing the features selection problem. Data simulated using a Web service e-commerce application are collected and tested on our model. Our model's detection or prediction rate is close to 100% and false alarm rate is less than 1%. Compared to other classifiers, our model's classification accuracy using random forests achieves the best results with RMSE close to 0.02 and time to build the model within 0.02 s for each data set with sample size of more than 600 instances. Thus, our novel fuzzy association rule model significantly provides a viable added layer of security protection for Web service and Business Intelligence-based applications.

© 2012 Elsevier Ltd. All rights reserved.

### 1. Introduction

Both the Internet and eXtensible Markup Language (XML)-based Web Services (WS) have revolutionized the Information Technology (IT) industry due to their many attractive features such as platform independence, interoperability, ease of use and ability to transport huge amount of information over the World Wide Web. Thus, more and more software applications, especially Business Intelligence (BI) or e-commerce applications are built on the Internet-enabled Web Service (WS) platform. Consequently, the Application Layer is open to various types of threats such as Structured Query Language (SQL) injection, XML injection, XML content and parameter tampering, Simple Object Access Protocol (SOAP) oversized payload, coercive parsing, and recursive payload leading to XML Denial-of-Service (DoS) attack.

WS-Security, an important specification addressing the security needs of WS applications exists, but it is still not 100% dependable. WS-Security is used to preserve the *integrity*, *confidentiality* and *availability* of a WS system, but it does not define any direct countermeasures for DoS attacks. Moreover, according to Jensen et al. (2009), DoS attacks on WS can be conducted with much less resource effort than against non-WS systems. Furthermore, XML Encryption can mask message content from being inspected. Thus, although using WS-Security on WS provides confidentiality to sensible data, the encrypted content can still conceal attacks such as oversized payload, coercive parsing or XML injection. Jensen et al. (2009) therefore, suggest using schema validation as a countermeasure. However, this will incur heavy CPU load and memory consumption as the system is tied up during XML and cryptographic processing for message decryption. Hence, to address this security problem, there is a dire need to have an added layer of protection at the Application Layer especially for WS-based e-commerce applications.

#### 1.1. Motivation

Over the past decades, active research in Host and Network-based Intrusion Detection (ID) and Intrusion Prevention (IP) systems are only able to detect and prevent attacks of the

\* Corresponding author. Tel.: +886 03 4227151x35417.

E-mail addresses: [gychan@mmu.edu.my](mailto:gychan@mmu.edu.my) (G.-Y. Chan), [cslee@cl.ncu.edu.tw](mailto:cslee@cl.ncu.edu.tw) (C.-S. Lee), [shheng@mmu.edu.my](mailto:shheng@mmu.edu.my) (S.-H. Heng).

<sup>1</sup> Present address: Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Malaysia. Tel.: +60 3 8312 5215.

<sup>2</sup> Present address: Graduate Institute of Network Learning Technology, National Central University, 300, Jhongda Road, Jhongli City, Taoyuan County 32001, Taiwan, ROC. Tel.: +88603422715135417 (direct Line).

<sup>3</sup> Tel.: +60 6 252 3485.

computer systems and attacks at the Network Layer. They are not adequate to countermeasure XML-related attacks mentioned above. Furthermore, although research have been conducted to countermeasure Web application attacks, they are still not adequate in countering SOAP or XML-based attacks. For example, [Ye \(2008\)](#) has designed a scheme to authenticate and validate a service request when the system is suspicious of being under XML DoS attack. However, their experiments show that the time taken to authenticate and validate SOAP messages increases as the SOAP size increases. This is due to the fact that more time is taken for the system to digest and decrypt larger SOAP messages; similar to the constraints of schema validation mentioned in [Jensen et al. \(2009\)](#). In another study by [Thakar et al. \(2010\)](#), requests for Web service are simulated on honey-pots and the support vector machine-based semi-supervised classifier used is able to intercept SOAP request to identify, for example, SQL injection and XML DoS attacks only.

### 1.2. Objectives

These security threats are especially true for WS and BI-based applications, where sensitive and a huge amount of business-related confidential information are being transported over the Internet. Hence, in this study, we have developed an adaptive and predictive fuzzy association rule model (FARM) to effectively validate inputs and countermeasure SQL injection attacks, unauthorized access, buffer overflows, XML-based attacks such as oversized payload, coercive parsing, recursive payload, XML content and parameter tampering.

### 1.3. Significance

The significance of our FARM is that, at this point, the ID/IP systems are mainly network-based ID/IP systems, which are not addressing SOAP and XML-related attacks. The main contribution of this study is the discovery of fuzzy associative patterns mapping the attributes (such as User ID, password, service request's input values, input size and SOAP size), with the consequent output, i.e., the decision outcome of the intrusion datasets. The discovered fuzzy

associative patterns are categorized into certainly allow access, certainly deny access or probably deny access decision class.

A series of sensitivity analysis conducted on this model has discovered 20 interesting rules with at least 99% confidence. Additionally, the detection and prediction accuracy of almost 100% and false alarm rate of less than 1% are obtained. Compared to other classifiers, our model's classification accuracy using random forests achieves the best results with RMSE close to 0.02 and time to build the model within 0.02 s for each data set with sample size of more than 600 instances.

The advantages of our FARM are first, identifying known attack signatures or patterns and segregating the anomalies from the normal help the security administrator to counter both signature and anomaly-based XML-related attacks at the Application Layer. Second, classification enables frequently occurring features to be determined from the set of interesting rules. Consequently, this helps the security administrator to prioritize which feature to focus on in the future thus addressing the feature selection problem. Third, our ID/IP model is adaptive (the anomaly database is dynamic whereby new attacks or variants are being identified continuously), accurate (evidence indicated above), sensitive to different datasets with varied associative patterns or instances and extensible (additional rules can be generated by manipulating support and confidence when in search of interesting rules for different contexts). Hence, our adaptive and predictive FARM contributes towards an added layer of security protection for Internet-enabled WS- and BI-based applications.

This paper is organized as follows: [Section 2](#) reviews related works; [Section 3](#) describes sample attacks and the proposed mitigation techniques in our fuzzy association rule model; [Section 4](#) presents the experimental design with the simulation of datasets and evaluation of the model; [Section 5](#) summarizes and compares results and [Section 6](#) concludes and discusses future work.

## 2. Related work

Basically, there are two approaches to intrusion detection and prevention, namely, signature-based and anomaly-based. The

**Table 1**  
ID/IP researches and techniques.

ID/IP researches with data mining techniques	
<b>MADAM ID</b> (Mining Audit Data for Automated Models for Intrusion Detection)	<b>MADAM ID</b> uses data mining algorithms to compute activity patterns from system audit data and extracts predictive features from the patterns.
<b>ADAM</b> (Audit Data Analysis and Mining)	<b>ADAM</b> is a data-mining-based, online-network ID system.
<b>STAT</b> (State Transition Analysis Tool)	<b>STAT</b> uses state transition diagrams, written to correspond to the states of an actual computer system, to form the basis of a rule-based ID expert system.
<b>USTAT</b> (Unix-based STAT)	<b>USTAT</b> is Unix-based State Transition Analysis Tool implemented in UNIX.
<b>DT</b> (Decision Tree)	<b>Decision tree</b> , an induction classification algorithm in data mining, is inductively learned to construct a model from the pre-classified data set.
<b>Predictive pattern generation</b>	<b>Predictive pattern generation</b> uses a rule-based system of user profiles defined as statistically weighted event sequences to predict future events from past events.
<b>Pattern matching</b>	<b>Pattern matching</b> technique encodes known intrusion signatures as patterns that are then matched against the audit data. The matched pattern is detected as intrusion.
ID/IP researches with AI techniques	
<b>IDES</b> (ID Expert System)	<b>IDES</b> uses the expert system concept to adaptively learn the behavior of users for detection of misuse intrusions.
<b>ANN</b> (Artificial Neural Network)	<b>ANN</b> consists of a collection of processing elements that are highly interconnected and transforms a set of inputs to a set of desired outputs.
<b>ANFIS</b> (Adaptive Neural Fuzzy Inference System)	In <b>ANFIS</b> , knowledge regarding intrusions are expressed in the form of linguistic rules for building the FIS. The FIS is then fed with data in order to build the ANN.
<b>LGP</b> (Linear Genetic Programming)	<b>LGP</b> , variant to GP, can hasten up the evolution process by considering the fact that computer programs evolved at the machine code level.
<b>SVM</b> (Support Vector Machine)	<b>SVMs</b> classify data by using support vectors which are members of the set of training inputs that outline a hyper plane in feature space.

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات