Collaborative Outlier Mining for Intrusion Detection

Goverdhan Singh*, Florent Masseglia*, Celine Fiot *, Alice Marascu *, Pascal Poncelet**

*INRIA Sophia Antipolis, 2004 route des lucioles - BP 93, 06902 Sophia Antipolis
Prenom.Nom@sophia.inria.fr
**LIRMM UMR CNRS 5506, 161 Rue Ada, 34392 Montpellier Cedex 5, France
poncelet@lirmm.fr

Résumé. Intrusion detection is an important topic dealing with security of information systems. Most successful Intrusion Detection Systems (IDS) rely on signature detection and need to update their signature as fast as new attacks are emerging. On the other hand, anomaly detection may be utilized for this purpose, but it suffers from a high number of false alarms. Actually, any behaviour which is significantly different from the usual ones will be considered as dangerous by an anomaly based IDS. Therefore, isolating true intrusions in a set of alarms is a very challenging task for anomaly based intrusion detection. In this paper, we consider to add a new feature to such isolated behaviours before they can be considered as malicious. This feature is based on their possible repetition from one information system to another. We propose a new outlier mining principle and validate it through a set of experiments.

1 Introduction

Protecting a system against new attacks, while keeping an automatic and adaptive framework is an important topic in this domain. One answer to that problem could rely on data mining. Actually, Data Mining for intrusion detection aims to provide new tools in order to detect cyber threats (Luo, 1999; Dokas et al., 2002; Bloedorn et al., 2001; Manganaris et al., 2000; Wu et Zhang, 2003). Among those data mining approaches, anomaly detection tries to deduce intrusions from atypical records (Lazarevic et al., 2003; Eskin et al., 2002). The overall principle is generally to build clusters, or classes, of usage and find outliers (*i.e.* events that do not belong to any class or group identifying normal usage). Actually, outlier detection aims to find records that deviate significantly from a well-defined notion of normality. It has a wide range of applications, such as fraud detection for credit card (Aleskerov et al., 1997), health care, cyber security (Bloedorn et al., 2001) or safety of critical systems (Fujimaki et al., 2005).

However, the main drawback of detecting intrusions by means of anomaly (outliers) detection is the high rate of false alarms since an alarm can be triggered because of a new kind of usages that has never been seen before (and is thus considered as abnormal). Considering the large amount of new usage patterns emerging in the Information Systems, even a weak percent of false positive will give a very large amount of spurious alarms that would be overwhelming for the analyst. Therefore, the goal of this paper is to propose an intrusion detection algorithm that is based on the analysis of usage data coming from multiple partners in order