Chapter 16
Guidance for Selecting Data Collection Mechanisms for Intrusion Detection

Ulf Larson
Ericsson AB, Sweden

Erland Jonsson
Chalmers University of Technology, Sweden

Stefan Lindskog
Karlstad University, Sweden

ABSTRACT

This chapter aims at providing a clear and concise picture of data collection for intrusion detection. It provides a detailed explanation of generic data collection mechanism components and the interaction with the environment, from initial triggering to output of log data records. Taxonomies of mechanism characteristics and deployment considerations are provided and discussed. Furthermore, guidelines and hints for mechanism selection and deployment are provided. Finally, this chapter presents a set of strategies for determining what data to collect, and it also discusses some of the challenges in the field. An appendix providing a classification of 50 studied mechanisms is also provided. This chapter aims at assisting intrusion detection system developers, designers, and operators in selecting mechanisms for resource-efficient data collection.

INTRODUCTION

Collection and analysis of audit data is a critical component for intrusion detection. Previous research efforts (Almgren et al., 2007; Axelsson et al., 1998; Kuperman, 2004; Lundin Barse & Jonsson, 2004; Price, 1997; Zamboni, 2001) have concluded that by carefully selecting and configuring data collection mechanisms, it is possible to obtain better and more accurate analysis results. However, data is required to be correct and to be delivered in a timely fashion. The data should also be sparse to reduce the amount of resources used to collect and store it. Since production of audit data

DOI: 10.4018/978-1-4666-7381-6.ch016

directly depends on the deployed data collection mechanisms, adequate mechanism knowledge is thus a critical asset for intrusion detection system (IDS) developers, designers, and operators.

This chapter consists of a theoretical part that introduces the basic concepts of data collection, and a practical part where guidelines and hints for mechanism selection are discussed. The theoretical part discusses the basics of data collection from several perspectives. The components and operation of a generic IDS is described followed by an in-depth discussion of the components and operation of a generic data collection mechanism. Then, two taxonomies are presented, discussing mechanism characteristics and deployment considerations, respectively. Thereafter, the practical part discusses operational considerations and outlines a deployment strategy. Finally, future challenges are discussed, followed by some concluding remarks and an appendix providing a classification of 50 studied data collection mechanisms and techniques.

Both the appendix and the guidelines can be used when selecting mechanisms. They can also assist when a specific type of data collection is desired. For example, it is easy to find out what mechanisms collect samples for execution profiling, and what mechanisms that can be reconfigured without the need for restart. This is a valuable source of information that removes the need to browse multiple manual pages and white papers to find the desired mechanism. Furthermore, by using the selection guidelines, we can obtain a more resource efficient data collection and obtain a more accurate data analysis.

RELATED WORK

Anderson (1980) proposed to use data collection and analysis as a means of monitoring computer systems for detection of different types of intruders. Denning (1986) proposed An Intrusion-Detection Model and pointed out specific log information that is useful for intrusion detection. Price (1997) then derived the audit data needs of a number of misuse detection systems and investigated how well conventional operating systems (OSs) collection mechanisms met these needs. It was clear from her report that the collection mechanisms lacked useful content. Axelsson et al. (1998) investigated the impact on detection by carefully selecting a set of system calls as input to the detector. Their paper showed that the detection rate improved when a selected set of data was collected. Wagner & Soto (2002) further showed that if insufficient data is recorded, an attack might well be treated as normal behavior.

Kuperman (2004) investigated in his PhD thesis the log data needs of four different types of computer monitoring systems and showed that when log data was carefully selected, the detection rate was improved. Killourhy et al. (2004) discussed the impact of attack manifestations on the ability to detect attacks. Attack manifestations are information items that are not present during normal execution and can thus be the key to reveal attacks. Furthermore, Almgren et al. (2007) investigated what impact the use of different log sources had on detection of web server attacks. It was concluded that the properties of the log sources affect the detection capability. Finally, taxonomies regarding data collection mechanisms in general have also been proposed (Albari, 2008; Delgado et al., 2004; Larus, 1993; Schroeder, 1995). Fessi et al. (2010), discusses a network based IDS, and also provides a comparison of different types of IDS.

Log data requirements for security logging have also been proposed in several whitepapers and reports from renowned industry-centered research institutes. The SANS consensus project (SANS, 2006) proposes several log sources, such as network data, OS data and applications. Furthermore, the SANS top 20 critical security controls (SANS, 2013) discusses maintenance, monitoring, and analysis of audit logs. In National Institute of Standards and Technology (2013), a set of guidelines for security log management was