Chapter 7
Towards Intelligent Data Processing for Automated Determination of Information System Assets

Andrey Fedorchenko
https://orcid.org/0000-0002-5727-653X
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Russia

Elena Doynikova
https://orcid.org/0000-0001-6707-9153
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Russia

Igor Kotenko
https://orcid.org/0000-0001-6859-7120
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Russia

ABSTRACT
The chapter discusses how the intelligent data processing techniques, namely, the event correlation, can be used for automated discovery of information system assets. The authors believe that solving of this task will allow more accurate and detailed construction of the risk model for cyber security assessment. This task is complicated by the features of configuration and operation of modern information systems. The chapter describes different types of event analysis, including statistical, structural, dynamic, applied to this task. The authors propose a technique that incorporates determining the types of object characteristics, the types of objects (system assets), and their hierarchy based on the statistical analysis of system events. Significant attention is given to the stage of source data pre-processing. In addition, the developed technique has the broad application prospects to discover the inappropriate, dubious, and harmful information. The case studies and experiments that demonstrate an application of the developed technique are provided and analyzed.

DOI: 10.4018/978-1-7998-1290-6.ch007
INTRODUCTION

Well-known challenge of the modern information world is related to the analysis of huge amounts of data. It affects almost all research areas. In the chapter the authors consider this challenge on the example of tasks related to discovering of various data objects, or assets, including an inappropriate, dubious and harmful information, and the tasks of security assessment. Security assessment is directly connected with discovering of various objects. Assessment quality significantly depends on the accuracy of identifying the objects of various nature including the components of the analyzed system, vulnerabilities, attacks, attackers, etc.

The chapter focuses on the determination of information system assets. For any organization it is important to discover as primary assets as well as secondary assets that support viability of primary assets. While the assets should be specified on the design stage, this does not always happen. The features of the modern information systems complicate this task. In modern organizations the assets are connected by various data channels and semantic relations, and they can be strongly distributed. In such conditions if the architecture of the target information infrastructure is sufficiently large-scale then the expert inventory of assets requires significant time costs and human resources. Besides, in process of the system operation its infrastructure may change as a result of installing new software and hardware, system updates, countermeasure implementation, etc.

The authors believe that an automation of assets discovery and determination of hierarchy of relationships between them will allow more accurate and detailed construction of the risk model for cyber security assessment. This task is rather challenging. Currently network scanners can detect various network objects, such as services, the ports they use, the hosts on which they are deployed, and communication nodes. However, these tools do not allow automated determination of different types of static and dynamic objects (for example, processes, sessions, users, privileges, operation systems, etc.) and of their hierarchy. This prevents obtaining an actual dynamical model of the information system, which is initially uncertain.

In the chapter the authors propose a technique that incorporates determining the types of object characteristics, the types of objects (or system assets), and their hierarchy based on the statistical analysis of system events.

The proposed technique is based on the methods of correlation of events accumulated in security logs. The authors believe that dynamic analysis of system events and calculation of static and dynamic indexes, including frequency characteristics of event types, variability of event property values, pair utilization rates of event properties, and total utilization rates of the properties will allow one to determine the main event sources (objects) and the hierarchy of object types. Namely, the authors use calculated pair utilization rates for event properties, variability of values of object properties, and the total utilization rates of the properties. It allows determining the object types and their hierarchical interconnections from the point of view of a target infrastructure operation. This process can be realized from root object type “infrastructure” to leaf atomic object types, e.g. process id, taking into account subordination relations. The chapter describes different types of event analysis (statistical, structural, dynamic, etc.) that allows one to do it with maximum accuracy that is limited only by the source data on the assets in the event logs.

Though the approach is limited by the information stored in the event logs, the authors believe that further integration of dynamic models of infrastructure objects with the static data on the software and hardware will eliminate this limitation. Consideration of the conditionally static types of security data (vulnerabilities, weaknesses, attack patterns, etc.) will make possible proactive monitoring of organization security state for the preventive deployment of countermeasures. The developed approach has the broad application prospects to discover the inappropriate, dubious and harmful information.