Optimal Monitoring for Distributed Intrusion Detection System

Adam Piotr Grzech

Institute of Computer Science, Wroclaw University of Technology
Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland
(tel: +48 71 320 35 89; fax: +48 71 320 38 84)
(adam.grzech@pwr.wroc.pl)

Abstract: Distributed intrusion detection systems, which consist of spatially distributed monitoring elements, may be applied to detect intrusions in real-time manner based on the analysis of collected data. This paper is devoted to present and discuss some selected aspects of detection systems architecture and efficiency. In the first part detection capabilities as dependent on distributed computer communication system parameters are discussed. The aim of the second part is to present an idea of hierarchical architecture of distributed intrusion detection systems and to discuss quality of monitoring performed at the lower layer of detection system hierarchical architecture.

Keywords: computer networks, distributed IDS, monitoring

I. INTRODUCTION

Intrusion Detection System (IDS), being a system monitoring a stream of events for attacks and taking countermeasures, is defined as the process of intelligently monitoring the events occurring in a computer system, analysing them for signs of violation of the security policy. The security policies are in general designed and applied as services composed of contemplery parts and organized according applied security measures: avoidance and intrusion detection, restrictive and preventive measures, detection and forensic measures. IDS implemented in gain to protect the availability, confidentiality and integrity of networked systems by misuse or anomaly detection, are defined by both the method used to detect attack, and the placement of the IDS on the network [2,3,7,8].

IDS may be classifeied in different manners, but the most common is to classify them based on analysed IDS events sources, i.e., host-based IDS (HIDS) or network-based IDS (NIDS) and application-based IDS (AIDS). Most of them are distributed intrision detection systems (DIDS) utilizing the concept of specialized distributed agents community representing agents with the same purpose for detecting anomalies [3].

DIDS based on anomalies detection, and commonly called as intelligent IDS (IIDS), often requires extensive training data for artificial learning algorithms and are computationally expensive. In intelligent IDS various machine learning algorithms belonging to stochastic learning, rule-based learning, neural learning and empirical learning classes are frequently applied.

Architectures, structures and effectiveness of DIDS are optimized based on various criteria [5-10].

The paper is devoted to present some concepts relating to anomalies detecting capabilities (section II), an idea of DIDS monitoring system quality (section III) and to discuss the limitations of the proposed solutions, as well as suggest further research (section IV).

II. ANOMALIES DETECTION SYSTEMS SENSITIVITY

The decision about the intrusion detection (traffic monitoring) devices location can be dependable on several aspects: the possibility of detecting a large number of intrusions, expected loss of functionality of the monitored computer system due to intrusions and mutual location of devices. It is assumed that number and location of network traffic monitoring devices impact amount of data available for decision making and cost of the data delivery, and that the latter directly influence functionality of the DIDS measured by proposed indexes.

It is assumed that a topology of system is modeled by directed graph $G=(V,E)$, where the given set of nodes $V=\{1,...,n\}$ represents distinguishable parts of monitored system and $E\subseteq\{(u,v):u,v\in V,u\neq v\}$ is a set of ordered pairs of distinct nodes each representing a communication channel used to exchange messages in the computer system. Each node can send and receive messages that are: the results of typical work of that computer system or the messages that are being intrusions, which can violate confidentiality, integrity, or availability of data in the monitored system.

It is also assumed that in each network node an intrusion detection device, that can monitor both in-coming and out-coming traffic, may be located. Location

of intrusion detection device in node i $(1 \le i \le n)$ is denoted by binary variable x_i, where $x_i = 1$ ($x_i = 0$) means that there is (there is not) an intrusion detection device in node i. The number of available intrusion detection devices in the system is given and equal to m.

The number of devices and their location should guarantee the security policy established for monitored system. In this paper three measures defining the security level and loss of performance of monitored system are proposed: intrusion detection accuracy, speed and the overall overhead caused by the communication of devices with each other.

Intrusion detection accuracy is a measure indicating the percentage value of detected intrusions by DIDS to the overall number of intrusions occurred in the monitored computer system. Intrusion detection speed is measured in the number of hops between a node where intrusion where generated and a node with intrusion detection device where that intrusion was detected. Overall overhead is a measure that specifies the overall number of hops needed for communication among all intrusion detection devices.

Location of intrusion detection devices based on the value characterizing the possibility of detecting a large number of intrusions points out to the desire of detect as many intrusion as it is possible. For any node i in the graph, $1 \le i \le n$, it can be estimated using data gathered during the work of the system (e.g. we can use traffic size passed through that node or number of intrusions in that node), using some subjective characterization of node sets up by security experts or even using degree of the node. Location of detection devices based on the value characterizing the possibility of detecting a large number of intrusions can be presented as:

$$x_1, x_2, ..., x_m \leftarrow \min_{x_1, x_2, ..., x_m} w^T x \qquad (1)$$

where $w^T = [w_1, w_2, ..., w_n]^T$ and $w_i \in [0,1]$, $1 \le i \le n$ characterizes the inverse possibility of detecting large number of intrusions (i.e. the higher the value of w_i, the less intrusions are to be detected in that node).

Expected loss of functionality of the monitored computer system is connected with the administrators and users wishes of protecting the most valuable parts of monitored computer system. Quantity of expected loss can be characterized by the lost performance due to unavailability of some parts of the computer system, full cost recovery of attacked node, etc.

Location of intrusion detection devices which minimizes the excepted loss can be denoted as follows:

$$x_1, x_2, ..., x_m \leftarrow \min_{x_1, x_2, ..., x_m} s^T (\bar{1} - x) \qquad (2)$$

where $s^T = [s_1, s_2, ..., s_n]^T$ and $s_i \in [0,1]$ for $1 \le i \le n$ characterizes the expected loss for node i and $\bar{1}$ is a column vector of size n.

Mutual location of intrusion detection devices affects to the communication cost among intrusion detection devices and the speed of intrusion detection. Both quantities can be defined based on the conception of distance matrix $D = [d_{ij}]$ where $d_{i,j} \in [0,1]$ characterizes the shortest distance between node i and node j (e.g. the proportion of the shortest distance in hops between these two nodes to the longest possible distance in graph G). Then the location of intrusion detection devices which satisfies the minimum value of communication cost among all intrusion detection devices can be described as follows (where m is equal to the number of used intrusion detection devices):

$$x_1, x_2, ..., x_m \leftarrow \min_{x_1, x_2, ..., x_m} \frac{1}{m(m-1)} x^T D x \qquad (3)$$

It is also possible to calculate locations maximizing speed of intrusion detection (e.g. can be measured in hops as a distance traveled from the node where the intrusion was generated to the node where it was detected for the first time). Optimal detection devices location can be defined as:

$$x_1, x_2, ..., x_m \leftarrow \min_{x_1, x_2, ..., x_m} \frac{1}{m(n-m)} x^T D(\bar{1} - x) \qquad (4)$$

Suggested criteria (1) - (4) give possibility to set up the location of detection devices. To indicate relative importance of the criteria coefficients ω_1, ω_2, ω_3 and ω_4 ($\omega_1, \omega_2, \omega_3, \omega_4 \ge 0$, $\omega_1 + \omega_2 + \omega_3 + \omega_4 = 1$) may be applied. Introduction of such coefficients leads to definition of optimization task, i.e., location of intrusion detection devices in network that minimizes the following function:

$$f(x) = \omega_1 w^T x + \omega_2 s^T (\bar{1} - x) + \\ + \omega_3 \frac{1}{m(m-1)} x^T D x + \omega_4 \frac{1}{m(n-m)} x^T D(\bar{1} - x) \qquad (5)$$

subject to:

$$x_1 + x_2 + ... + x_n = m \qquad (6)$$

with respect to $x = [x_1, x_2, ..., x_n]^T$ defining location of intrusion detection devices in the monitored network.

The above binary quadratic programming problem is an optimization NP-hard problem solved effciently by optimal and approximate algorithms or heuristics used.

III. HIERARCHICAL DISTRIBUTED INTRUSION DETECTION SYSTEM

Efficiency of DIDS, in which data are collected and transferred for decision making purposes, depends on number and location of monitoring elements, amount of collected and transmitted data and the location, where the data are processed.

Lower layer of the DIDS is responsible for local monitoring of distributed system and consists of

monitored elements - software applications or hardware systems that collect data about the state of linked network devices. Any of the monitored element is local to the network device which entail negligible delay in exchanging data between them. Monitored elements are next grouped into areas (called monitoring areas). All monitored elements from the same monitoring area send its data in specified time intervals to the same middle layer element for the intrusion detection analysis.

The middle layer of the DIDS consists of monitoring elements that gather data from all monitored elements within their monitoring areas. The data between monitored elements of the lower layer and monitoring elements of the middle layer are sent through the same distributed system communication channels that are used for users' traffic exchange in the distributed system. A monitoring element itself can be a software application or hardware system which analyses data collected from all monitored elements within its monitoring area in order to detect intrusions. The amount of sent data for an analysis, as well as users' data results from the ordinary distributed system functionality impacts on delay of the network traffic. In this paper it is assumed that monitoring element can be located locally to any of the network device.

The highest layer of DIDS consists of a correlation element which is responsible for gathering data from all monitoring elements from the middle layer for the intrusion detection analysis within the whole distributed system. The controlling element can also derive dynamic properties of detected intrusions which can be next used to prevent their spread throughout the distributed system.

Quality of local and global decisions (DIDS higher layers functionality) depends on quality of distributed monitoring system [5].

Hierarchical architecture of the DIDS allows the division of functionality that improves the scalability and reliability of distributed intrusion detection system as well as simplifies the design and implementation phase of such a system comparing to the architecture of centralized intrusion detection system. In addition, such architecture fulfills several important features [4,7], like it imposes the minimum overhead on the distributed system in order to avoid interference with its ordinary functionality and is easy to deploy. Therefore the quality of DIDS depends on the quality of monitoring system.

3.2. Basic notations

A distributed system is modeled as an undirected graph $G(V,E)$, where $V=\{v_1,v_2,...,v_N\}$ represents a set of nodes, where each node $v_i \in V$ $(1 \le i \le N)$ depicts a location in which network devices are localized, and $E=\{e_{ij}\}$ defines a set of communication channels between these nodes with given capacity. Knowledge of networks topological structure, traffic requirements, channel capacities and applied routing algorithm leads to knowledge of traffic flows over all networks channels.

A monitoring devices can be located in any node $v_j \in V$ $(1 \le i \le N)$. For each monitoring element located in node v_i there is defined a corresponding monitoring area Λ_i which is a subset of nodes from which data for an analysis is sent to this monitoring element.

It is also assumed that there exists a finite set of possible network device classes $U=\{u^{(1)},u^{(2)},...,u^{(k)}\}$ possible to locate in any node of the graph. A monitoring devices can be located in any node $v_j \in V$ $(1 \le j \le N)$. For each monitoring element located in node $v_j \in V$ there is defined a corresponding monitoring area Λ_i; all monitoring devices located in nodes belonging to the distinguished are send data to distinguished node $v_i \in \Lambda_i \subset V$ $(i=1,2,...,m)$. The total number of monitoring nodes (m) may be given, or selected in gain to optimise selected performance measure.

The set of all Λ_i $(\Lambda_1 \cup \Lambda_2 \cup ... \cup \Lambda_m = V)$, where set which is a subset of nodes from which data for an analysis is sent to this monitoring element.

The number of network devices belonging to the distinguished classes, located in the particular node $v_j \in \Lambda_i \subset V$ and transferring monitoring data to the node $v_i \in \Lambda_i \subset V$ is $L_{ji}=\{l_{ji}^{(1)},l_{ji}^{(2)},...,l_{ji}^{(k)}\}$. It is the set of all monitoring devices, located in the j-th node, and delivering monitoring data to the i-th node ($v_i \in \Lambda_i$).

Moreover, it is assumed that each network device class has its own data generation intensity $\alpha^{(k)}$ $(1 \le k \le K)$ which quantitively depicts the amount of data generated and relayed to the linked monitored element. The network device which belongs to class k $(1 \le k \le K)$ together with its locally monitored element located in node $v_j \in \Lambda_i$ sends monitoring data to a monitoring node $v_i \in V$; amount of transferred data in time interval τ with the data sending rate $\beta_{ji}^{(k)}$ equals to $\alpha_{ji}^{(k)}(\tau)=\alpha^{(k)}\beta_{ji}^{(k)}\tau$. Total amount of data transferred from the $v_j \in \Lambda_i$ node to the $v_i \in V$ node equals to:

$$\alpha_{ji}(\tau)=l_{ji}^{(1)}\alpha_{ji}^{(1)}(\tau)+l_{ji}^{(2)}\alpha_{ji}^{(2)}(\tau)+...+l_{ji}^{(K)}\alpha_{ji}^{(K)}(\tau).$$

Therefore, the total amount of data generated within the particular monitoring area Λ_i equals to:

$$\alpha_i(\tau)=\alpha_{ji}^{(1)}(\tau)+\alpha_{ji}^{(2)}(\tau)+...+\alpha_{ji}^{(K)}(\tau).$$

3.3. Distributed monitoring system quality

In the hierarchical DIDS, the functionality of higher layers is based on the functionality of lower layers,

therefore the intrusion detection capabilities depends on the monitoring system quality.

The quality of the monitoring system is influenced both by the amount and delay of data collected for intrusion detection analysis: having more data with small enoygh delay it is possible to take a better decision .

Therefore, it is assumed that the impact of data intended for an analysis in amount of $\alpha_{ji}(\tau)$ on the monitoring system quality is complied by a penalty function. Penalty function can be calculated for all the given or calculated monitoring area Λ_i:

$$p(\Lambda_i) = \frac{1}{\alpha_i(\tau)} \sum_{v_j \in \Lambda_i} \alpha_{ji}(\tau) \cdot p_{ji}(\alpha_{ji}(\tau)) \qquad (7)$$

Similarly to penalty function, a collected data delay can be calculated for all the given or calculated monitoring area Λ_i:

$$d(\Lambda_i) = \frac{1}{\alpha_i(\tau)} \alpha_{ji}(\tau) \cdot \sum_{v_j \in \Lambda_i} d_{ji}(\alpha_{ji}(\tau)) \qquad (8)$$

A monitoring system quality which depicts the quality of the lower layer of distributed intrusion detection system depends on the localization and number of monitoring elements, form of monitoring areas and amount of data that is sent for intrusion detection analysis and can be calculated a sum of delay and penalty function:

$$Q\left(\Gamma, \Lambda_i, \beta_{ji}^{(k)}\right) = \sum_{v_j \in \Lambda_i} \left(d(\Lambda_i) + p(\Lambda_i)\right) \qquad (9)$$

The optimization task of monitoring system quality can be formulated, analyzed and solved based on all possible parameters defining monitoring system quality (5), namely localization and a number of monitoring elements, form of monitoring areas, amount of monitoring data determined by data sending rate and amount of users' network traffic.

In the discussed quality of monitoring system, depending on amount of monitoring data and the data communication costs, delivery costs of locally made decisions may be included.

Fot example, the data sending rate optimization task is formulated as below:

$$Q^*\left(\beta_{ji}^{(k)}\right) = \min_{\beta_{ji}^{(k)}, v_j \in \Lambda_i, k=1..K;} \sum_{v_j \in \Lambda_i} \left(d(\Lambda_i) + p(\Lambda_i)\right) \qquad (9)$$

In this case the monitoring system quality depends on the data sending rate.

IV. CONCLUSIONS

In this paper some selected issues concering optimal architecture of DIDS are discussed; intrusion detection capabilities depending on parameters describing distributed computer communication systems and quality of distributed monitoring system.

Quality and amount of data,collected and transferred for decision making purposes in DIDS, may be important to control quality of decisions produced by the system. It is also important to manage the trade-off between cost of collecting and transferring monitoring data and the quality of detection.

As in the paper many additional assumptions were made, in the future work more detailed analysis is going to be provided as well as the analysis of real appliance for distributed and intrusion detection systems will be presented and the impact of monitoring system quality on the distributed intrusion system quality will be examined thoroughly. The presented concept will be further enhanced for more complex dynamic computer networks.

The optimisation problem of monitoring devices localization and amount of data exchanged among monitoring and monitored data as well as quality of the distributed monitoring issues, discussed in the paper, is a part of work devoted to investogate interdependencies between quality of monitoring systems and two-stage decision making system. The gain of the latter is to produce decisions, quality of which are usually measured by the probability for undetected attacks.

REFERENCES

[1] Bar-ilan J., Kortzars G., Peleg D. (1993), How To Allocate Network Centers, Journal of Algorithms (15/3).
[2] Bashah N., Shanmugam B., Ahmed A.M. (2005), Hybrid intelligent detection system, Proceedings of World Academy of Science, Engineering and Technology (vol. 6): 291 - 294.
[3] Benattou M., Tamine E. K. (2005), Intelligent agents for distributed intrusion detection system, Proceedings of World Academy of Science, Engineering and Technology (vol. 6): 190 - 193.
[4] Chebrolu S., et.al. (2004), Feature deduction and ensemble design of Intrusion Detection Systems, Computer & Security.
[5] Grzech A. (2006), Anomaly detection in distributed computer communication systems, Cybernetics and Systems (37/6).
[6] Hentea M. (2007), Intelligent system for information security management: Architecture and design issues, Issues in Information Science and Information Technology (vol. 4): 29 – 43.
[7] Neumann P., Porras P. (1997), Event Monitoring Enabling Responses to Anomalous Live Disturbances.
[8] Ozsoy F., Pinar M. (2006), An exact algorithm for the capacitated vertex p-center problem, Computers and Operations Research (33/5).
[9] Peddabachgari S., Abraha A., Grosan C., Thomas J. (2005), Modeling intrusion detection system using hybrid intelligent systems (www.sciencedirect.com).
[10] Snapp S., et.al. (1991), Distributed Intrusion Detection System – Motivation, Architecture and an Early Prototype, NCSC.