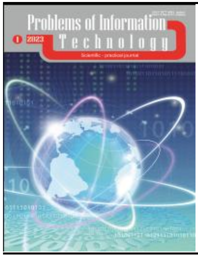

Available online at www.jpit.az

14 (1)
2023

Conceptual model of intelligent monitoring system for computer networks

Ramiz H. Shikhaliyev

Institute of Information Technology, B. Vahabzade str., 9A, AZ1141 Baku, Azerbaijan

ramiz@science.az

0000-0002-8594-6721

ARTICLE INFO

http://doi.org/10.25045/jpit.v14.i1.04

Article history:
Received 01 August 2022
Received in revised form 27 September 2022
Accepted 01 November 2022

Keywords:
Computer networks
Intelligent monitoring
Log file analysis
Machine learning

ABSTRACT

The size and complexity of computer networks (CNs) are constantly increasing, which requires the intellectualization of network monitoring. Undoubtedly, intellectualization will increase the effectiveness of monitoring the CNs. To ensure the intellectualization of the CNs monitoring, it is necessary to use machine learning (ML) methods. The use of ML methods enables to create an intelligent monitoring system. This article proposes a conceptual model of a system for CNs intelligent monitoring. The model is based on the analysis of log files using ML methods. The proposed model will make it possible to monitor the CNs in a targeted manner, which can increase the efficiency of network monitoring and management in terms of the use of network resources.

1. Introduction

Over the past decades, the scale of computer networks (CNs) has increased, that is, the number of nodes, users, services, applications, etc. has increased, leading to a change in the nature and scale of the environment for monitoring and managing CNs. At the same time, due to the complication of the nature of monitoring and managing the CNs, the task of monitoring and managing becomes even more difficult. This is due to the difficulty of making decisions on monitoring and managing the CNs. Traditional decision making for monitoring and managing CNs is based on the personal experience and intuition of network administrators, and therefore cannot meet the requirements of more complex network management decisions.

Monitoring is one of the main means of management decision support, as well as on ensuring the security of the CNs. All tools for managing and ensuring the security of the CNs rely on the data of the monitoring system. At the same time, the main purpose of monitoring is information support of the control systems of the CNs in the decision-making process. This requires continuous and comprehensive monitoring of nodes, services, applications and user activity, etc. However, continuous and comprehensive monitoring, especially active monitoring of large and complex CNs, can reduce the overall performance of the network. Subsequently, the Quality of Service (QoS) of the CNs may decrease. This is because active monitoring introduces additional measurement traffic into the network and uses channel resources of the CNs (Shikhaliyev, 2022). Moreover, CNs monitoring systems are often located in the same network that they monitor, and if the CNs fails or slows down, the monitor itself fails. Therefore, a new approach is required to monitoring the CNs, and an approach is proposed that consists in the intellectualization of monitoring based on the mining of log files (for example, syslog). This will ensure the completeness and timeliness of solutions for monitoring and managing the CNs and reduce the share of network resources (system, computing, channel resources and memory) used for monitoring. As a result, the security and efficiency of monitoring, management and control of the CNs can increase.

CNs automatically generates log files that contain very valuable information about the behavior and state of the network, as well as information about system failures and security breaches. At the same time, various network devices, such as routers, switches, firewalls, intrusion detection and prevention systems, servers, etc., generate log files. Typically, log files are large and complex data format, and therefore it is extremely difficult to manually analyze them and extract valuable information. In fact, the analysis of log files of large and complex CNs is a big data problem that is difficult to carry out with analytical methods. Therefore, it becomes very difficult to accurately identify the states of the CNs and make appropriate decisions on monitoring and managing the CNs. Therefore, for intellectual analysis of CNs log files, it is necessary to use machine learning (ML) methods (Abdalla and Jumaa, 2022) and monitor in accordance with the results of this analysis. At the same time, information about the state of the CNs can be effectively extracted by in-depth analysis of some current and retrospective parameters of the CNs, and subsequently the future state and behavior of the CNs can be predicted. Bandwidth, throughput, time-delay, packet arrival rate, packet inter-arrival time, etc. can be used as parameters. At the same time, some parameters can be taken as important indicators of anomalies in the behavior of the CNs. The choice of parameters for monitoring depends on the specific requirements of the tasks of monitoring and controlling the CNs. The CNs state can be defined as normal, overloaded, failed, safe, attacked, etc. Decision-making on the choice of methods and strategies for monitoring the CNs can be carried out on the basis of the knowledge obtained about the influence of network parameters on the state of the CNs. Thus, the use of ML methods makes it possible to ensure the intellectuality of monitoring the CNs.

The purpose of this article is to develop a model of a system for intelligent monitoring of the CNs based on the analysis of log files. The proposed model will provide intelligent information support for network administrators in monitoring and managing CNs.

The rest of the paper is organized as follows. Section 2 describes a literature review. The model of an intelligent monitoring system for CNs is given in Section 3. Finally, the concluding remarks and future work are discussed.

2. Related works

Due to the very large number of entries in the log files, it is difficult to analyze those using traditional methods. Thus, data mining methods are used to analyze the data of log files.

An analysis of works on the use of ML methods in the analysis of log files and in monitoring systems and intelligent decision support showed significant limitations in the areas of their use.

In (Brandao and Georgieva, 2020), the authors propose an intrusion detection system based on the mining of log files. To predict attacks, log files from various sources are combined into one information panel and the most distinctive features are identified. The features are determined using approaches such as the Optimal Feature Selection algorithm and Factor Analysis. Attack prediction is tested using methods, such as Decision Tree, K-Nearest Neighbors, and Neural Network. The evaluation of the proposed system is carried out using the KDD Cup 1999 public log file dataset. As a result of the experiments, the decision tree shows the best result.

In (Mandagondi, 2021), the author considered the problems of analyzing log files, in particular, detecting anomalies in log files. The main goal is to search for important events in the log files, and for this, ML methods are used to distinguish between a log file with an anomaly and without an anomaly. The experiment uses algorithms such as Local Outlier Factor, Random Forest and Term Frequency Inverse Document Frequency, as well as a set of labeled log files and evaluated by analyzing the anomalous events contained in the log files. To detect anomalous events, the author uses the K-Means and PCA (Principal Component Analysis) clustering algorithm. Experimental results show that when detecting anomalous events in data, the Term Frequency

Inverse Document Frequency method performs better than other approaches.

The paper (Kobayashi et al., 2017) proposes a method for extracting failure cases and their causes from network system log file data. The method is based on causal inference, which restores the causal relationship of network events from a set of time series of events. Causal inference reduces the number of correlated events and therefore infers more plausible causal events than the traditional cross-correlation approach. The proposed method is applied to the data of the network system log file of the academic network in Japan for 15 months. The method significantly reduced the number of pseudo-correlated events compared to the traditional method.

The work (Dasgupta and Gonzalez, 2001) proposes an intrusion detection system based on a genetic classifier, which can provide an active detection and automatic responses to intrusions. The system can monitor various activities on the network (i.e., look for changes such as failures, malfunctions, abnormal values of monitored parameters, misuse, intrusions, etc.). In particular, the system simultaneously monitors activities at the user, system, process, and package levels. At the same time, a genetic classifier is used to determine specific security violation actions. The goal is to find a correlation between abnormal values of controlled parameters in order to determine the type of intrusion and generate an action accordingly.

In (Mohammed et al., 2021), a ML method is proposed that can automatically determine the state of the network and localize faults. Decision tree algorithms, and gradient boosting are used to determine the normal state, congestion and network failure. Experimental results show that the proposed method provides up to 99% accuracy for a data set collected through an emulated network.

In (Kotenko et al., 2020), based on artificial neural networks, an approach is proposed for implementing the analytical block of an intelligent support system for a network administrator. The structure of a combined neural network focused on assessing the state of CS elements is considered. Learning methods such as stochastic gradient descent, adaptive learning rate method, and adaptive inertia method are used. The results of the experiments showed a sufficiently high accuracy of the proposed solution, good adaptability and the possibility of its application in a wide range of network configurations.

The work (Mirzaraxmedova and Fozilova, 2019) analyzes the prospects for the technology of intelligent monitoring systems. Various monitoring intellectualization strategies are aimed at intellectual information support of decision makers. Such support can be implemented by building fuzzy linguistic databases/knowledge in conjunction with fuzzy inference subsystems, and information for decision making can be displayed on the automated workplace of the decision maker.

Summing up, it can be mentioned that these works are aimed at solving specific monitoring problems using the analysis of log files. This study proposes a conceptual model of an intelligent CNs monitoring system based on the analysis of log files using ML methods, which will provide intelligent information support for network administrators when making decisions on CNs monitoring.

3. Intelligent monitoring system model

Traditional CNs monitoring systems make it possible to collect data both offline and online, and provide control over various aspects of data collection, such as data sampling frequency, monitoring duration, etc. However, these monitoring systems face the challenge of accurately and efficiently processing big data in real time, which reduces their effectiveness.

The CNs as an object of monitoring is a complex, dynamic system that is difficult to describe analytically and to model. Therefore, making decisions on CNs monitoring by simple methods of numerical analysis cannot satisfy the needs of network administrators, and a model of an intelligent CNs monitoring system is proposed (Fig. 1).

The proposed model of the intelligent monitoring system of the CNs is based on knowledge and is modeled as a real-time system, which consists of the following components (Fig. 1):

1. The CNs is the environment in which effective monitoring must be carried out. CNs

consists of many nodes, services, applications, users, etc.

2. Making decisions on monitoring the CNs begins with the collection of log files. Collection of log files will be done using log collection tools (for example, using ArcSight, RSA Envision, Q1 Labs, Logstash). A log file is a data format widely used in a network event recording system and is a repository for storing collected data. Typically, when services, applications, network security tools and network devices of the CNs run, their log files are created. Thus, log files are the main source of data for monitoring the state of the CNs. Typically, a log file is a string of semi-structured text that often contains a timestamp and a detailed message (e.g., error symptom, target component, IP address, etc.). Unlike other network data collection mechanisms, log files are primarily stored in persistent storage. Typically, log files take up a large amount of memory, have a low information density and different formats, for example, the format of information received from the log file of the router differs from the format of information received from the firewall. Therefore, log collection tools must be flexible enough to work with all network devices and applications.

3. Pre-processing of log files includes parsing (He et al., 2016), cleaning, sampling, normalization, data transformation, feature extraction, etc. Log files must be pre-processed before parsing, as log file data cannot be used as it is stored. Generally, raw log files are partially structured and must be converted to a structured format, i.e. they must be parsed for further analysis.

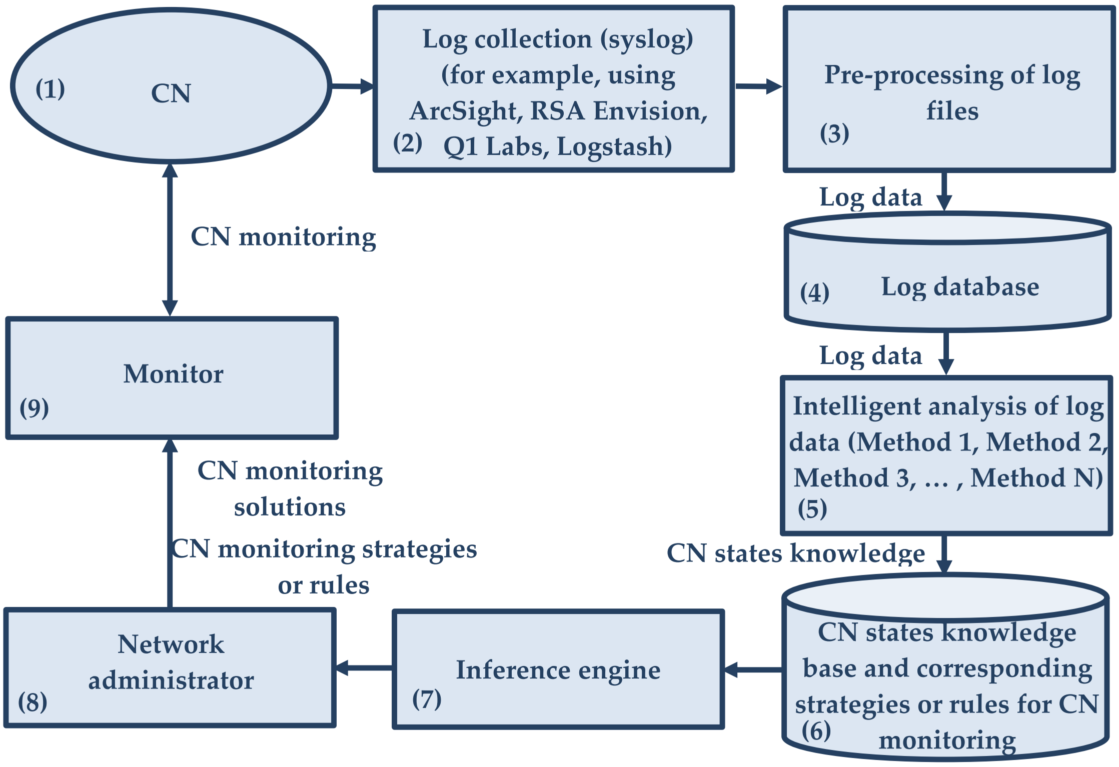
(1) CN
Log collection (syslog) (for example, using ArcSight, RSA Envision, (2) Q1 Labs, Logstash)
Pre-processing of log files (3)
Log data
(4) Log database
Log data
Intelligent analysis of log data (Method 1, Method 2, Method 3, … , Method N) (5)
CN states knowledge
CN states knowledge base and corresponding strategies or rules for CN (6) monitoring
Inference engine (7)
Network administrator (8)
CN monitoring strategies or rules
CN monitoring solutions
Monitor (9)
CN monitoring

Figure. 1. Model of a CNs intelligent monitoring system

4. Log database is a log data storage. After log files pre-processing, the received data about nodes, services, applications, etc. stored in the log database.

5. Log data mining is the most important component, and ML methods will be used that allow detecting unacceptable deviations (or security violations) of the characteristic parameters of the CNs. The levels of monitoring of the CNs are: user activity; network traffic; servers; network applications; network services; network equipment; communication environment, which leads to the multitasking of CNs monitoring (Fig. 2). The multitasking of CNs monitoring, in turn, leads to

the need to use various ML methods. Because some methods are suitable for solving some monitoring tasks, and some methods for other monitoring tasks. At the same time, several ML methods can be used together to solve any monitoring problem. ML methods may include classification, clustering, etc.

6. A knowledge base is a database that stores information about the state of the CNs and strategies or monitoring rules corresponding to the states of the CNs and is an expandable and self-updating knowledge base. In this state, CNs can be defined by the values of some network parameters or by sets of network parameter values. It is clear that a decision support system can provide more effective decision support only if it has a large amount of knowledge.

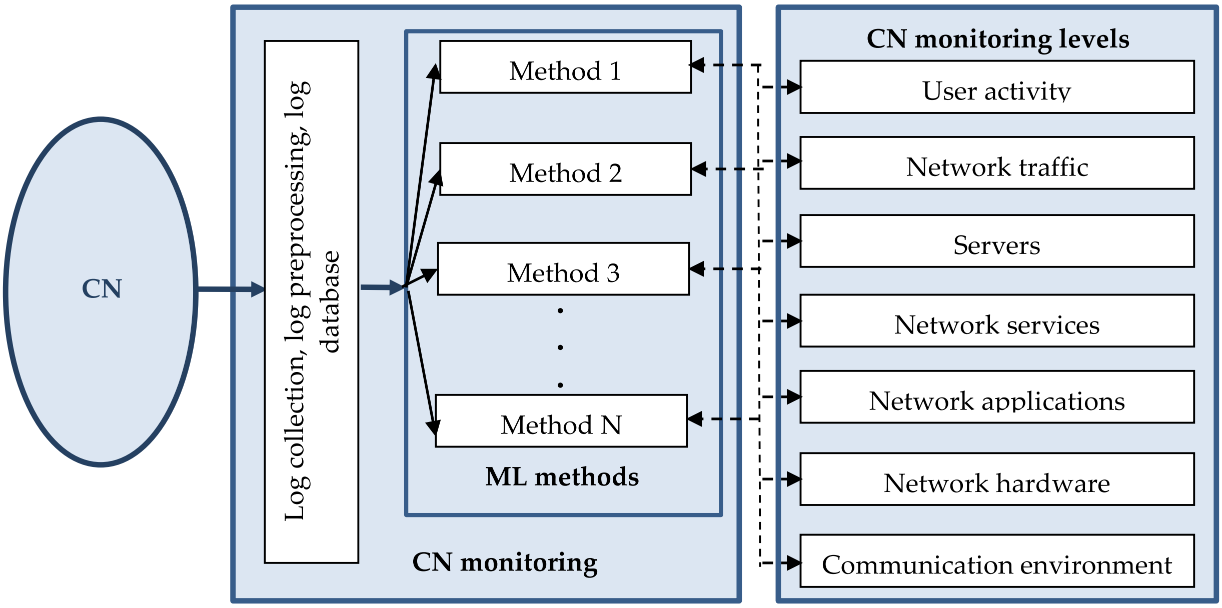

Figure 2. CN multitask monitoring

7. When creating an intelligent system for monitoring the CNs, one of the important tasks is to establish the basic behavior of the CNs for the current architecture and infrastructure. It includes determining the normal behavior of the network, the acceptable range of monitored parameters, types of devices, applications, network services, user behavior, as well as network interactions with external networks. This information is the main one based on which decisions on monitoring the CNs can be made. The task of choosing strategies or rules for monitoring the CS can be considered as an optimization problem, where it is required to choose such strategies or monitoring rules corresponding to the states of the CNs that minimize the impact of the monitoring system on network performance. This task can also be considered as a decision-making task, when the network administrator of the CNs is the decision maker who must decide how best to monitor in order to effectively manage the CNs.

8. The CNs network administrator is a human expert who makes a decision on monitoring the CNs based on the knowledge of the states obtained from the knowledge base and the strategies or rules for monitoring the CNs corresponding to these states. At the same time, depending on the states of the CNs, different methods and strategies or monitoring rules may be required, that is, strategies or rules for choosing certain nodes, users, services, applications, time, etc. for monitoring. Thus, monitoring of the CNs as needed can be implemented as a set of rules from the monitor’s knowledge base.

9. Monitor - these are means of monitoring the CNs. CNs monitoring is carried out based on the decision of the network administrator. Various monitoring methods (active and passive monitoring) and monitoring strategies (policies) can be used to monitor the CNs. The decision on the choice of methods and strategies for monitoring the CNs is based on the intellectual analysis of the log files data.

4. Conclusion

The current level of development of artificial intelligence methods, in particular, ML methods, provides the basis for building a new generation of CNs monitoring systems.

This article proposed a conceptual model of an intelligent system for monitoring the CNs. The proposed model was based on the intellectual analysis of log files.

For intellectual analysis of log files, it was proposed to use ML methods. Using ML methods will ensure the completeness and timeliness of decisions on monitoring and managing CNs. The proposed model is flexible and adaptable to various goals and objectives of monitoring and provides multitasking of CNs monitoring.

Further research is related to the study of mechanisms for accelerating the reading of log files in order to provide an online mode for processing log files, the development of a mechanism for multitasking monitoring of the CNs using various ML mechanisms, the development of a subsystem and a decision-making mechanism for monitoring the CNs.

References

Shikhaliyev R. H. (2022). A method for intelligent planning of computer networks monitoring Problems of Information Technology, 13(1), 42-48. http://doi.org/10.25045/ jpit.v13.i1.05

Abdalla R. R. and Jumaa A.K. (2022). Log File Analysis Based on Machine Learning: A Survey. UHD Journal of Science and Technology, 6(2), 77-84. https://doi.org/10.21928/uhdjst.v6n2y2022.pp77-84

Brandao A. and Georgieva P. (2020). Log Files Analysis for Network Intrusion Detection, Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems, pp. 328-333. https://doi.org/10.1109/IS48319.2020.9199976

Mandagondi L. G. (2021). Anomaly Detection in Log Files Using Machine Learning Techniques, Master of Science in Computer Science, Faculty of Computing, Blekinge Institute of Technology.

Kobayashi S., Fukuda K., Esaki H. (2017). Mining causes of network events in log data with causal inference, 2017 IFIP/IEEE International Symposium on Integrated Network Management (IM2017), pp. 45-53. https://doi.org/10.23919/INM.2017.7987263

Dasgupta D. and Gonzalez F. A. (2001). An Intelligent Decision Support System for Intrusion Detection and Response, Proceedings of the International Workshop on Information Assurance in Computer Networks: Methods, Models, and Architectures for Network Security, 2001, https://doi.org/10.1007/3-540-45116-1_1

Mohammed A. R., Mohammed S. A., Côté D., and Shirmohammad S. (2021). Machine Learning-Based Network Status Detection and Fault Localization, IEEE Transactions on instrumentation and measurement, 70, 3521710, https://doi.org/10.1109/TIM.2021.3094223

Kotenko I., Saenko I., and Skorik F. (2020). Intelligent support for network administrator decisions based on combined neural networks. In 13th International Conference on Security of Information and Networks (SIN 2020), November 04–07, 2020, Merkez, Turkey. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3433174.3433602

Mirzaraxmedova A.X. and Fozilova M.M. (2019). Analysis of Prospects of Technology of Intelligent Monitoring Systems, International Conference on Information Science and Communications Technologies (ICISCT). https://doi.org/ 10.1109/ICISCT47635.2019.9011845

He P., Zhu J., He S., Li J. and Lyu M.R. (2016). An Evaluation Study on Log Parsing and Its Use in Log Mining, 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, pp. 654-661. https://doi.org/ 10.1109/DSN.2016.66