Problems of Information Technology (2022), vol. 13, no. 1, 38-42

Available online at www.jpit.az

13 (1)
2022

A method for intelligent scheduling of computer networks monitoring

Ramiz H. Shikhaliyev

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

ramiz@science.az

0000-0002-8594-6721

ARTICLE INFO

http://doi.org/10.25045/jpit.v13.i1.05

Article history:
Received 10 August 2021
Received in revised form 1 November 2021
Accepted 21 January 2022

Keywords:
computer networks
network monitoring
intelligent scheduling of monitoring
machine learning

ABSTRACT

To ensure the efficiency of management, the security of computer networks (CN), as well as to ensure the required level of quality of service for network applications, accurate and up-to-date information on the state of the CN is required. This information can be obtained through continuous active monitoring of the quantitative characteristics of the CN. Thus, active monitoring becomes an important tool for ensuring the efficiency of management and security of the CN. However, continuous active monitoring, especially of large networks, can lead to congestion of network channels, which can reduce the effectiveness of monitoring the CN. Consequently, with active monitoring of the CN, it is necessary to manage the use of resources (channel and computational) of the network and reduce the load on the network. To solve this problem, this paper proposes a method for intelligent scheduling of monitoring of the CN. Using machine learning algorithms, can be analyzed the state and performance of the CN and acquire knowledge that can be used to determine the most appropriate rules for monitoring the CN. Thus, it is necessary to find such monitoring rules that will ensure the effectiveness of monitoring the CN. The proposed method can reduce the impact of monitoring on network performance, and on the operation of network applications.

1. Introduction

Continuous active monitoring is essential for effective management and security of computer networks (CNs), as well as for ensuring the required quality of service of network applications. Active network monitoring is based on the analysis of data collected as result of inclusion of additional test packages into the network (Shikhaliyev, 2011; Mohan, Reddy, Kalpana, 2011). Moreover, active monitoring includes the measurement of the quantitative characteristics of CN. Through an active monitoring of CN, information on characteristics such as available bandwidth, throughput, delays, round-trip time (RTT), jutter, etc. can be obtained. Active monitoring can be either periodic or on demand. Regardless of the method of active monitoring used, the expenses of consuming network resources become challenging when the network is monitored and they can affect the performance of CN. In terms of traffic flow throughput, performance of CN is defined as the relation of the average flow volume to the average flow duration (Bonald, Roberts, 2007).

Typically, in large CNs, the scale of the monitoring data collection infrastructure and the volume of analyzed data are very high. Conducting active monitoring at this scale through a static monitoring strategy can lead to overload of communication channels, which can reduce the overall performance of the network. Therefore, network resource usage is needed to be managed (network computing power, memory and bandwidth of network channel) in the course of active monitoring, which is a key step in ensuring the effectiveness of CN's monitoring. Moreover, the main task is the effective management of the collection of a large amount of monitoring data. To solve this problem, new intelligent approaches to CN monitoring, that is, methods for intelligent scheduling of CN monitoring have to be developed. Furthermore, the intelligent monitoring scheduling will almost ensure the management efficiency, safety and performance of CN. The goal of CN monitoring scheduling is to minimize the network overload during active monitoring by reducing the volume of monitoring traffic and ensuring the required network performance, as well as meeting the quality of service of the network with application requirements.

Scheduling is defined as the allocation of resources in accordance with tasks over time and is a decision-making process targeted at optimizing one or more goals (Pinedo, 2016). Scheduling is related to the search for optimal plans, taking into account a number of restrictions. However, many methods developed for solving the scheduling problem are not suitable for dynamic systems, which have several constraints and many unexpected failures. Essentially, CN is a dynamic system with various limitations and many unexpected failures. In most real systems, scheduling is a continuous reactive process in which predetermined schedules are continually reviewed and reseat in real time. Therefore, to solve scheduling problems, it is advisable to use artificial intelligence methods, in particular, machine learning algorithms.

The goal of this article is to develop a method for intelligent scheduling of CN monitoring, which minimizes the network load, that is, reduces the consumption of network resources during the CN's active monitoring. Because the consumption of CS network resources varies when different monitoring rules (strategies) are applied. Minimizing the network load will ensure the required network performance during the CN's active monitoring. To solve the problem, it is proposed to use machine learning algorithms. Using machine learning algorithms, the previous state and performance of the CN can be analyzed based on training examples, and knowledge can be acquired. This knowledge can be used to determine the most appropriate rules for current time CN monitoring (for example, monitoring at certain hours, days, as well as of certain servers, nodes, services, network characteristics, etc.). The proposed method for CN monitoring scheduling will reduce the load on both the network and the monitoring system, as result of which the efficiency of the CN monitoring system will increase as a whole.

2. Related works

There are not many works in the literature on the topic of planning the monitoring of the COP. However, let's look at some network monitoring planning solutions. In (Geng et al., 2018), the authors propose an SDN-based (Software-Defined Network) network monitoring and scheduling architecture, in which the management and data levels are distinguished. The network administrator can use the controller to adjust the network control level. It can also program the control level to deploy new network functions. Many available SDN-based methods provide monitoring of the network condition using additional probe packages, polling, etc., which makes network monitoring too expensive. To eliminate high overhead expenditures in SDN networks, it is proposed to use P4 (Bosshart et al., 2014). P4 is a programming language that is primarily used to provide instructions for network equipment (e.g., switches, network cards, firewalls, etc.) to forward data. The essence of the proposed approach is to implement monitoring and scheduling of tracking functions in P4, as well as visualization of information about the network condition. Furthermore, the study proposes and implements a traffic scheduling scheme based on information about the network condition and the P4 routing architecture. The proposed scheduling scheme is evaluated with INT (Inband Network Telemetry), which is designed to enable the collection and transmission of reports on the data layer status without the intervention of control layer (Kim et al., 2016).

(Mathew, 2017) proposes a containerized, decentralized and parallel scheduler for the network performance active monitoring. The proposed scheduler is implemented for the ConMon container monitoring platform and for running distributed containers inside. The goal of this approach is to deploy scheduling containers in hosts, inside which the applications are run. Performing monitoring in this way provides a better understanding of the application's usage of network performance, and the scheduler refine monitoring time better than other distributed and decentralized schedulers. Although the monitoring scheme is close to cycled one, in a dynamically adaptable monitoring cluster may perform scheduling sequentially without any monitoring conflicts.

In (Qin, Rojas-Cessa, and Ansari, 2010), the authors propose a measurement scheduling scheme for network monitoring. The essence of scheduling is to resolve conflicts on the resource consumption for both periodic measurement tasks and on-demand measurement tasks. In this regard, graph coloring is used; it is the ascending order of the sum of the number of clicks and the tasks' degree. In this case, each measurement task is viewed as a graph vertex, and the competition/conflict between the two tasks is an edge connecting these two vertices. The tasks are selected in accordance with the ascending order of the sum of the number of clicks and the degree of conflict task on the conflict graph, which enables performing several tasks of measuring the resource consumption rate simultaneously. The proposed scheme reduces the average latency for the execution of all tasks in the schedule of periodic measurement tasks. For on-demand measurement tasks, the scheme minimizes the latency for the execution of on-demand input tasks while maintaining high temporary space usage. This reduces the total time spent on all tasks.

Network measurement scheduling with limited resources (such as server resources, network bandwidth, etc.) may not be possible due to high measurement query loads. Therefore, requests for network measurements that cannot be scheduled may adversely affect monitoring accuracy. To solve this problem, the study (Calyam et al., 2010) proposes a semantic scheduling algorithm based on the principles of deterministic and heuristic scheduling for measurement query processing. The proposed algorithm achieves the goals of the network monitoring and enables making decisions on adapting resources. The semantic priorities used in this scheduling algorithm are based on user level and resource level policies. This allows dimension requests to be prioritized, which replaces the typical scheduling priorities based on frequency and execution time. The proposed semantic scheduling algorithm is evaluated using metrics such as cycle time and satisfaction coefficient to increase the load of measurement queries.

(Barlet-Ros, 2008) proposes a technique called "load shedding" that allows monitoring systems to efficiently handle overload situations in the presence of multiple arbitrary and competing monitoring applications. The author proposes a predictive load shedding scheme to shed excess load under very high network traffic and preserve the accuracy of monitoring applications within the limits specified by end users, while ensuring optimal distribution of computing resources for various applications. The main novelty of the proposed scheme is that it treats monitoring applications as black boxes with random and highly variable input traffic. Without any knowledge about the internal installations of the application, the proposed scheme extracts a set of features from traffic flows to build a model to predict the resource requirements of each online monitoring application. Further, these features are used to predict congestion situations and manage overall resource consumption by sampling input package streams. Thus, the monitoring system retains a high flexibility, expanding the range of applications and network scenarios used. Because not all monitoring applications are sample tolerant, the load shedding scheme is extended to support load shedding methods for certain end user support. This provides a unique solution for arbitrary monitoring applications. The proposed scheme enables the monitoring system to safely delegate the excess load shedding task to applications.

The study (Shikhaliyev, 2015) proposes a methodology for improving the effectiveness of CN monitoring, which will reduce the time spent on monitoring and ensure the continuity of updating and relevance of monitoring data. The main goal in this case is to optimize the CN nodes monitoring with given network resources. To achieve this goal, it is proposed to use a polling system model, the optimization of which will achieve the optimization of CN monitoring. It is assumed that the main indicator of the polling systems optimization includes the minimization of the average latency for applications in the queue. Therefore, the problem of optimization of the CN monitoring process is reduced to the problem of optimization of the CN nodes polling model by minimizing the average latency for applications in the queue.

3. Machine Learning-based CN monitoring scheduling

CN monitoring scheduling is the implementation of certain network monitoring rules. Moreover, the choice of certain rules depends on the state and performance of the CN, since there are no single ideal rules for CN monitoring. Therefore, during monitoring scheduling, the main problem is the choice of the most appropriate rules for each case, that is, for each state in which the CN can be. Furthermore, when choosing monitoring rules, various real-time information about the CN should be taken into account, and they should be chosen for such a period of time to ensure the timeliness and relevance of monitoring. To achieve these goals, currently applicable knowledge about the relationship between the state of the CN and the monitoring rule should be used. Therefore, the knowledge about the CN monitoring scheduling should be applied to minimize the time and ensure the effectiveness of monitoring of a dynamically changing network. Various machine learning algorithms can be used to obtain the knowledge required for making decisions on CN monitoring scheduling. In this case, training samples and learning algorithm should be chosen so that this knowledge will be useful. Moreover, to get training samples, the attributes chosen are critical for the performance of the scheduling system.

For intelligent CN monitoring scheduling, a Machine Learning-based scheduling model (Figure 1) is proposed, where sample generator generates the training and testing samples for machine learning algorithm. To create different states of the CN and select the optimal monitoring rules for each specific state of the CN, a simulation model is used. Further, the machine learning algorithm acquires knowledge from training samples necessary for the CN monitoring scheduling decisions (on the choice of monitoring rules). Furthermore, determining the optimal size of training samples is a very important task, i.e., an acceptable size of training sample has to be chosen. Moreover, the choice of an adequate monitoring period, that is, the frequency of the CN quantitative characteristics monitoring, which determine the CN performance, is vital for making decisions about changing the monitoring rules. Additionally, the CN monitoring system, using scheduling knowledge, including the information about the CN performance, determines the optimal monitoring rule.

The sample generator uses a non-static knowledge base. Therefore, a procedure capable to automatically change the knowledge is required, i.e., it would clarify the knowledge in the case of significant changes in the state of the CN. The main goals of the knowledge refinement procedure include the identification of the gaps in the knowledge base and addition of appropriate training samples.

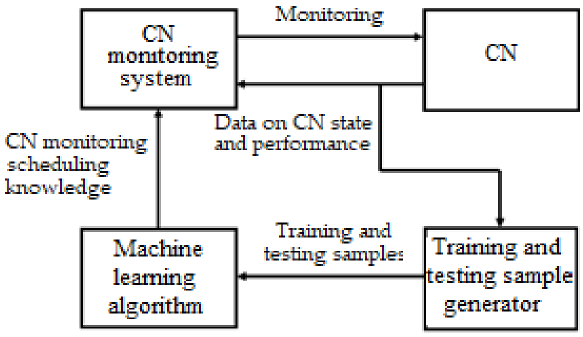

Figure. 1. Machine Learning-based intelligent CN monitoring scheduling.

Various machine learning algorithms can be used to ensure the intelligent CN monitoring scheduling (Pinedo, 2016). Depending on the used type of machine learning algorithm, knowledge-based approaches can be categorized as follows: inductive learning, transductive learning, neural networks, case-based reasoning (CBR), support vector machine (SVM), reinforcement learning, mixed approaches, etc.

To make a decision on changing the CN monitoring rules, an algorithm used by the CN monitoring system is proposed. Figure 2 illustrates the block diagram of the algorithm.

The need to change the CN monitoring rules is determined by the excess of the value $\{V_i\}, (1 \leq i \leq m)$ of the threshold T. In this case, the values

V_i are determined as the difference, i.e., CN performance before monitoring and the values of CN performance $P_i, i = 1, m$, obtained using monitoring rules $R_i, i = 1, m$. The threshold T defines the required level of quality of network service for applications. The proposed algorithm for making a decision to change the CN monitoring rules will find and rank the preferred CN monitoring rules.

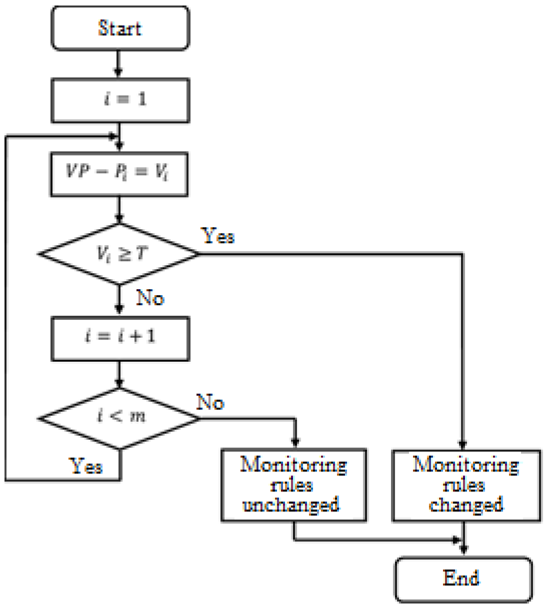

Figure 2. Decision algorithm for changing the CN monitoring rules

4. Conclusion

CN monitoring without scheduling can lead to inefficient consumption of network resources, which can cause overload of communication channels and decrease the overall CN performance. Consequently, the efficiency of the CN monitoring system may decrease. Therefore, the development of a CN monitoring scheduling method is becoming a very urgent problem. The presence of the CN monitoring scheduling method can significantly reduce the network load and, accordingly, provide the required quality of service for applications used in the CN.

This article proposed a method for CN monitoring scheduling based on the use of Machine Learning algorithms. Various machine learning algorithms can be used to schedule the CN monitoring. Machine Learning algorithms will make decisions on the dynamic selection of the most appropriate monitoring rules.

References

Shikhaliyev R.H. (2011). On methods and means of computer network monıtorıng. Problems of Information Society, 2(4), 61-70. (in Azerbaijani)

Mohan V., Reddy Y. R. J., Kalpana K. (2011). Active and Passive Network Measurements: A Survey. International Journal of Computer Science and Information Technologies, 2(4), 1372-1385.

Bonald T., Roberts J. (2007). Scheduling Network Traffic. ACM SIGMETRICS Performance Evaluation Review, 34(4), 29-35. https://doi.org/10.1145/1243401.1243408

Pinedo M.L. (2016). Scheduling Theory, Algorithms and Systems. 5th edition, Springer, 674.

Geng J., Yan J., Ren Y., Zhang Y. (2018). Design and Implementation of Network Monitoring and Scheduling Architecture Based on P4. Proceedings of the 2nd International Conference on Computer Science and Application Engineering, 1-6. https://doi.org/10.1145/3207677.3278059

Bosshart P., Daly D., Gibb G., Izzard M., Mckeown N., Rexford J., Schlesinger C., Talayco D., Vahdat A., Varghese G., and Walker D. (2014). P4: Programming Protocol-Independent Packet Processors. ACM Sigcomm Computer Communication Review, 44(3), 87-95. https://doi.org/10.1145/2656877.2656890

Kim C, Sivaraman A, Katta N, Bas A, Parag D. A. and Wobker LJ. (2016). Inband Network Telemetry. https://www.p4.org, (electronic resource).

Mathew C. (2017). Scheduling Network Performance Monitoring in The Cloud. Degree project, Faculty of Engineering and Science, Uppsala University, 71.

Qin Z., Rojas-Cessa R., and Ansari N. (2010). Task-Execution Scheduling Schemes for Network Measurement and Monitoring. Computer Communications, 33, 124-135. https://doi.org/10.1016/j.comcom.2009.11.005

Calyam P., Kumarasamy L., Ozguner F. (2010). Semantic Scheduling of Active Measurements for meeting Network Monitoring Objectives. Proceedings of the International Conference on Network and Service Management, 4. https://doi.org/10.1109/CNSM.2010.5691256

Barlet-Ros P. (2008). Load Shedding in Network Monitoring Applications Thesis of Doctor in Computer Science, Polytechnic University of Catalonia, Department of Computer Architecture, 160.

Shikhaliyev R.H. (2015). Improvıng the monıtorıng effıcıency of computer networks based on the pollıng system optımızatıon. Information Technology, 1(8), 576-584. (in Russian)