Flow Rate Analysis of Syslog Collected in the Tokyo 2020 Games

HIROSE Masato KANAI Akira

Abstract

In a network for events in which multiple vendor network devices are integrated, it is hard to analyze the semantics of logs or detect anomalies by keywords. We collected device logs from some operational organizations during the Olympic and Paralympic Games Tokyo 2020. We focused on the file size of the logs to follow the scale of the network of the games. And we categorized the logs by device type and installation site, contributing to real-time analysis and discovery of multiple problems.

Keywords : Network monitoring, Log analysis, Tokyo 2020 Games

1. Introduction

For network operators, defining the anomaly or normal state and having the monitoring system notify operators in the event of anomaly states contribute to the early detection of problems. To determine an anomaly or normal state from device logs, it is necessary to define anomaly string patterns according to the contents of the target device logs before the operational term. These preparations are time-consuming tasks for network operators. In addition, considering the scale of the Tokyo 2020 Olympic and Paralympic Games (hereinafter referred to as "Tokyo 2020") network, the ever-changing nature of the events, and the building and operation period of each venue, it is very difficult to define anomaly or normal states in advance [1] proposed to perform log analysis in such an environment. This method uses a Bollinger band graph to analyze the number of lines in the logs of devices in an event network. We decided to conduct a logging analysis using this method and tried to apply it to the environment of the Tokyo 2020 Games. However, it was not possible to prepare the necessary and sufficient compute resources in a short time to implement this method in real-time. A particular challenge was the amount of memory required to count the number of log lines. In the Tokyo 2020 environment, nearly 1 GB of logs were collected in the monitoring environment every minute. Reading these logs into memory and counting the number of lines in real-time required much memory.

2. Log Analysis Environment for Tokyo 2020 Games

To solve this problem, the total size of the log file every few minutes was used in the analysis instead of the number of lines in the log. To obtain the log file size, we checked the file size every five minutes on a log storage server and took the records. In Linux, such as CentOS, it is possible to obtain the target file size by getting the inode information managed by the file system. Therefore, obtaining the file size requires much fewer computing resources than counting the number of lines in the log file.

While performing the operations, a Bollinger band

HIROSE Masato
Innovation Center, NTT Communications Corporation
E-mail masato.hirose@ntt.com
KANAI Akira
Innovation Center, NTT Communications Corporation
E-mail a.kanai@ntt.com
The Journal of The Institute of Electronics, Information and Communication Engineers, Vol.105, No.8, Supplement, pp.67-70, August 2022

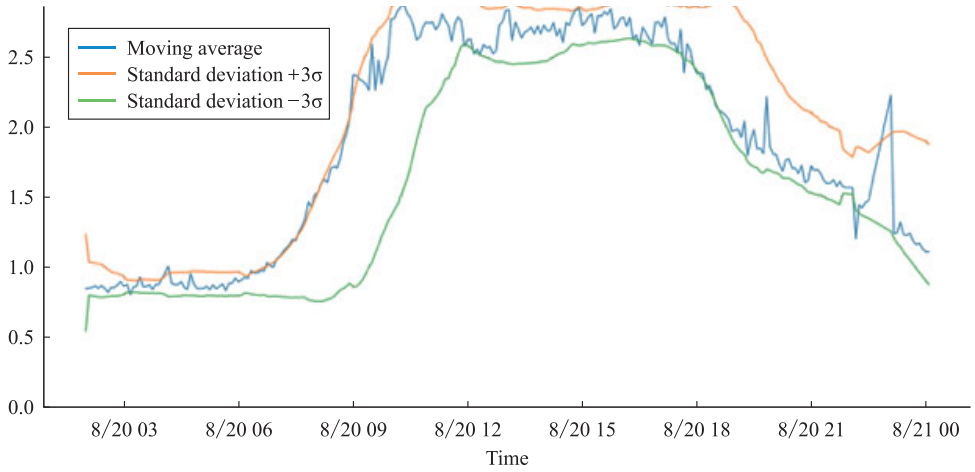

Figure 1 Bollinger Band Graph Generated from Log Flow

graph is generated. The moving average and standard deviation were calculated using data from 12 intervals ago, i.e., from one hour ago.

Figure 1 shows a Bollinger band graph generated based on the log file size increase data for one day during the Tokyo 2020 Games. The blue line shows the moving average, the orange line shows the standard deviation $+3\sigma$, and the green line shows the standard deviation -3σ. This figure shows that the log flow rate gradually increases from 06 : 00 and progressively decreases from 18 : 00. The section between 03 : 00 and 05 : 00 is characterized, as shown in the figure, by an increase in logs generated when logging into the device for periodic backup of settings. The rapid growth in log flow from 06 : 00 was due to the sudden increase in network usage because of the entry of staff and press into each venue. Then, we confirmed that the log flow rate of load balancers and proxies increased rapidly. From around 18 : 00, network usage and the log flow rate decreased.

On the other hand, the flow rate increases in bursts around 22 : 00. When analyzing this graph, the operator was notified when the moving average exceeded the standard deviation of $+3\sigma$ or when it fell below the standard deviation of -3σ. Upon receiving a notification, the operator manually searched the logs for the device that had significantly increased or decreased log flow during the relevant period and escalated the situation if a problem was suspected.

As mentioned above, when a single Bollinger graph is generated for all logs collected, the logs of devices that generally have a higher log flow than those of other devices occupy most of the total graph, and the log flow of different devices is not visible as a trend. Therefore, we categorized the logs of devices by device type, DC, and venue and generated a graph for each. Table 1 shows the actual categorization. This categorization enabled the operators to notice anomalies even in devices with a low log flow and quickly narrow down the list of devices suspected of having problems.

Table 1 Examples of Categorization

Category	Example of division	Number of division
Type of device	Router, Switch, UTM, etc.	40
By DC and By Venue	Primary DC, Secondary DC, National Stadium, etc.	64

3. Results Obtained from the Analysis

The analysis of the Tokyo 2020 Games was conducted using the method described in Section 2.3. The first is external scanning activity and the second is missing security appliance logs.

The first event is related to external scanning activity. We found that the moving average of the log flow on the Unified Threat Management (UTM) log flow graph (Figure 2) exceeded the standard deviation of 3σ. The UTM logs contain many filtered logs of scanned packets during the time in question, and the system detected this event by capturing changes in the log flow rate.

The second event was the log loss of the security appliance. On the graph of the collected logs (Figure 3), the log flow rate decreased by approximately 1 GB. The moving average of the log flow fell below the standard

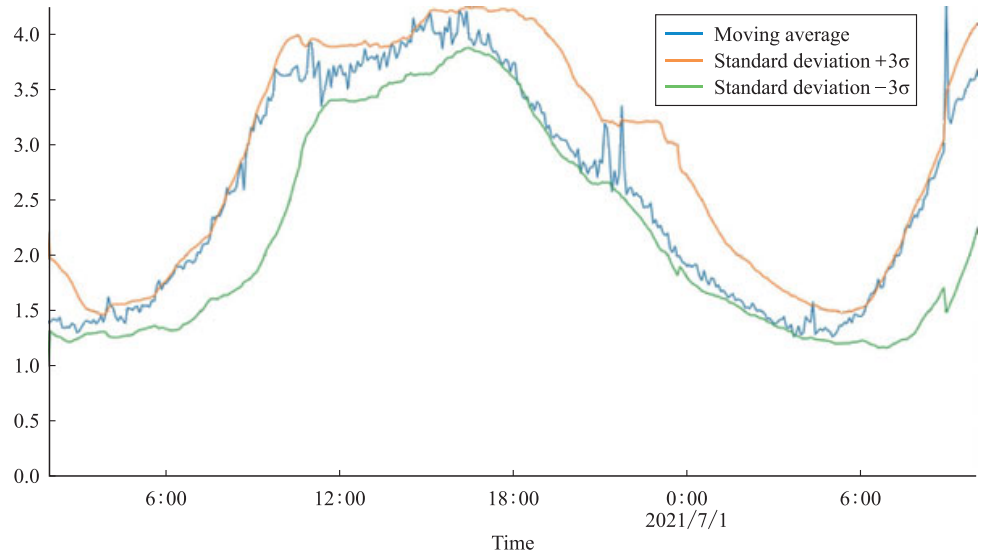

Figure 2　Graph of External Scanning Activity Detected

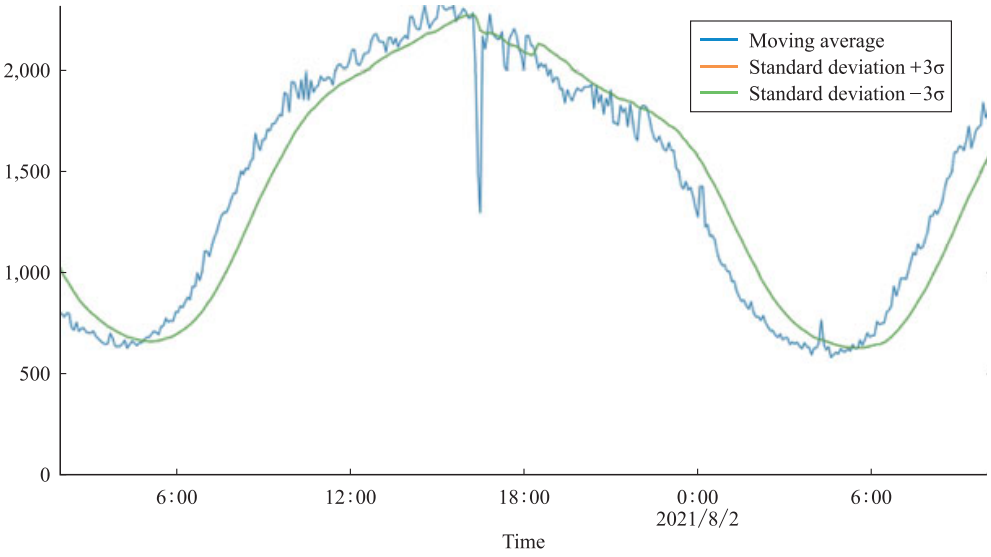

Figure 3　Graph of Missing Security Appliance Logs

deviation of −3σ, which led to the discovery of this event. After a detailed investigation by the team operating the device, it was found that the logs were not being transmitted to the outside world for a certain period due to the performance limitations of the device. Fortunately, the logs for that period were stored inside the device, and there was no impact on business operations. However, it was only by focusing on the logs' flow rate that we could detect and discover anomalies in the logging system of the device.

4. Conclusion

This article describes our analysis of the overall logs data collected from devices across the Tokyo 2020 Games. To keep up with the scale of the network for the Games, we focused on the file size of the logs to realize real-time analysis. In addition, categorizing the logs by device type and installation site made it possible for operators to identify suspected problems quickly. The system contributed to the detection of multiple issues and increased the speed of response.

References

(1) H. Abe, M. Shikida, and Y. Shinoda,"The anomaly detection method analyzing syslog data using bollinger bands algorithm on event network," the Information Processing Society of Japan, vol. 59, no. 3, pp. 1006-1015, 2018.

(Received 28 February 2022 ; Revised 16 March 2022)

HIROSE Masato

Completed the master's program in the Graduate School of Information Science and Technology at JAIST in 2017. He joined NTT Communications Corporation in the same year. Since then, he has been researching and developing Internet technologies and services.

KANAI Akira

KANAI Akira completed the Graduate School of Media and Governance at Keio University in 2009. He joined NTT Communications Corporation in the same year. He is a CISSP, CISA, CISM, and a Professional Engineer (Information Technology).