Contribution ID: **39** Type: **Oral Presentation**

Abnormal Log Flow Detecting System in CNGrid

Wednesday, 26 August 2020 16:00 (30 minutes)

Distributed systems have grown larger and larger since this concept appears, and they soon evolve to environments that contain heterogeneous components playing different roles, e.g. data centers and computing units. CNGrid is a good example of a large distributed environment. It is composed of 19 HPC clusters contributed by many research institutes and universities throughout China. Computer Network Information Center of Chinese Academy of Sciences is the operation and management center of CNGrid, and is responsible for keeping the environment running smoothly and efficiently.

During the maintenance work it is found that logs generated by devices from the environment play a very important role to locate anomalies and system failures, and can help us backtrack root causes of those occurred troubles. In previous report we have proposed the general framework of log based monitoring and diagnosing in CNGrid. Different types of logs are gathered from hosts in the environment and analyzed in various ways. Finally diagnosing reports are expected to be produced, reflecting the running status of the whole environment.

Among these analyses being performed to logs, the abnormal log flow analyzing method has been implemented as a complete anomaly detecting system. The idea of this system assumes that in most time when the environment runs normally and stably, the number of logs for each classified log type should be kept in a nearly stable level in unit times. If in a unit time logs of a certain log type has a dramatic increase or decrease, it is highly possible that in the corresponding host there are something happened, i.e. an anomaly, and the system maintainers should be noticed to make rapid response for any potential threat. The advantage of taking log flow analysis is that most log analyses are based on the content of logs, but some anomalies cannot be captured by words. For example, only one log showing disable to build connection may be caused by just a very short network delay, but a big number of this log continuously happened in a short time may be caused by a misconfiguration in the destination host or a power-off, which is a much more serious problem. Abnormal log flow detecting system can well find anomalies like this.

In this report we will introduce the work process of the abnormal log flow detecting system in details, including how to generate log flow models to classify log flow anomalies, and how to combine different modules to build the system as a whole. We will also demonstrate the real running effect of the system, including the visualization work to make the detecting result much clearer for humans to understand. We believe that the work process of this system can be adapted to many log analyzing methods and systems.

Primary author: Dr ZHAO, Yining (Computer Network Information Center, Chinese Academy of Sciences)

Co-authors: Mr XIAO, Haili (Supercomputing Center, Chinese Academy of Sciences); Mr WANG, Xiaodong (Computer Network Information Center, Chinese Academy of Sciences); Prof. CHI, Xuebin (Computer Network Information Center, Chinese Academy of Sciences)

Presenter: Dr ZHAO, Yining (Computer Network Information Center, Chinese Academy of Sciences)

Session Classification: Network, Security, Infrastructure & Operations Session

Track Classification: Network, Security, Infrastructure & Operations