International Journal of e-Collaboration
Volume 16 • Issue 2 • April-June 2020

An Adaptive Cloud Monitoring Framework Based on Sampling Frequency Adjusting

Dongbo Liu, Hunan Institute of Engineering, Xiangtan, China
Zhichao Liu, Hunan Institute of Engineering, Xiangtan, China

ABSTRACT

In a cloud platform, the monitoring service has become a necessary infrastructure to manage resources and deliver desirable quality-of-service (QoS). Although many monitoring solutions have been proposed in recent years, how to mitigate the overhead of monitoring service is still an opening issue. This article presents an adaptive monitoring framework, in which a traffic prediction model is introduced to estimate short-term traffic overhead. Based on this prediction model, a novel algorithm is proposed to dynamically change the sampling frequency of sensors so as to achieve better tradeoffs between monitoring accuracy and overhead. Also, a monitoring topology optimization mechanism is incorporated which enables to make more cost-effective decisions on monitoring management. The proposed framework is tested in a realistic cloud and the results indicate that it can significantly reduce the communication overhead when performing monitoring tasks for multiple tenants.

KEYWORDS

Cloud Computing, Monitoring Service, Performance Metric, Virtualization

1. INTRODUCTION

In recent years, cloud computing has emerged as a leading paradigm to enable customers deploying their applications in cost-effective and performance isolation manners (Khorshed et al., 2013; Zhang et al., 2016). Accompanying with the new opportunities brought by cloud computing, many challenges also raise when the number of cloud resources are scaled up (Weingartner et al., 2015; Singh & Chana, 2016). Therefore, knowing the online status of various resources (e.g., availability, capability, efficiency) plays a critical role to effective managing nowadays cloud infrastructures from the perspective of cloud providers (Manvi & Shyam, 2014; Canali & Lancellotti, 2014; Lu, et al., 2016). In addition, a full knowledge and control of the current status of those resources is also helpful to improve the delivered QoS or avoid SLA violations from the perspective of cloud users (Serhani et al., 2014; Fatema et al., 2014; Tomanek et al., 2016). As a result, an effective cloud monitoring service has become a necessary infrastructure in current cloud environments (Canali & Lancellotti, 2014; Wang et al., 2017; Mdhaffar et al., 2017).

Generally saying, a monitoring service is to collect and aggregate performance-related metrics from distributed resources (Balis et al., 2011; Fatema et al., 2014). By analyzing monitoring data, cloud providers can understand the runtime status of underlying resources, and then detect, diagnose

DOI: 10.4018/IJeC.2020040102

or solve problems before they happen (Jing & Min, 2011; Pardi et al., 2013). Unfortunately, it is not easy to manage the information of a huge number of resources in a reliable and scalable way due to the enormous scale and complex structure of cloud computing, which is especially true when multiple tenants with different monitoring requirements are taken into account (Fatema et al., 2014; Du & Li, 2017). In addition, the inherent characteristics of cloud systems (such as resource virtualization and elastic resource provision) also increase the difficulties of implementing a non-intrusive cloud monitoring service (Montes et al., 2013; Povedano-Molina et al., 2013). In the past decade, many cloud monitoring tools, or approaches have been proposed, each having its own advantages and disadvantages (Fatema et al., 2014; Wang et al., 2018). Even so, how to mitigate the overhead introduced by monitoring service is still a key issued need to be addressed, since growing scale and complexity of cloud platform make it impossible for simultaneous monitoring all resources without introducing significant overheads on network and storage subsystems (Fatema et al., 2014; Weng et al., 2016). In addition, monitoring data covering all resource elements will be too huge to be efficiently handled, and in most scenarios, it is not necessary. Therefore, a more flexible monitoring service that enables to dynamically adjust or customize monitoring process in an on-demand manner becomes necessary.

In this article, we address the monitoring overhead problem by presenting an adaptive cloud monitoring framework. In the proposed framework, a traffic prediction model is designed to estimate short-term traffic overhead based on historical monitored data, which is then used to dynamically change the sampling frequency of low-level sensors with attempt to achieve better trade-offs between monitoring accuracy and communication overhead. To provide more efficient monitoring service, a monitoring topology optimization mechanism is also proposed, which enables to make more efficient and cost-effective decisions on monitoring management, such as removing duplicated monitoring requests and re-configuring key monitoring parameters. The proposed monitoring framework is highly compatible with existing monitoring solutions, since the introduced monitoring components are designed as independent plug-ins.

The rest of this paper is organized as follows. Section 2 presents the related work. In section 3, we describe the framework of the proposed monitoring framework and its detailed implementation; In section 4, we present our experimental evaluations and analysis on the proposed framework. Section 5 concludes the paper with a brief discussion of our future study.

2. RELATED WORK

Due to the importance of monitoring service to distributed systems, traditional high-performance computing platforms have long incorporated it as a basic service with aiming at improving system efficiency. For instance, the famous LHC Computing Grid (WLCG) has developed a monitoring dashboard system (Andreeva et al., 2010), which provides a user-friendly tool from the perspective of the high energy physics experiments and allows users to observe the online states of allocated resources during their experiments. SECMOL (Baresi et al., 2010) is a general monitoring specification language designed for service-oriented distributed systems, which enables to define separated functions for raw-data collection, data analysis and processing, and user-oriented monitoring query in a highly flexible and scalable way. In European WS-DIAMOND project (Ben-Halima et al., 2010), a large-scale monitoring framework is developed to provide a fine-grained reconfiguration service on the top of Grid5000. MATE (Caymes-Scutari et al., 2010) is parallel system monitoring tool which allows users automatic and dynamic tuning the runtime parameters for their parallel applications. Also, it provides a set of toolkits for fault and bottleneck detection. In (Balis et al., 2011), the authors, the authors real-time monitoring approach, namely Complex Event Processing (CEP), which can be applied in grid-like systems to monitor real-time applications. In (Barlet-Ros et al., 2011), a network monitoring system is proposed to proactively offload traffics so as to maintaining the accuracy of monitored applications without compromising the thresholds defined by networking administrators. VisualGrid (Minutoli et al., 2010) is a simple but effective monitoring tool designed for scientific cloud