Future Generation Computer Systems 56 (2016) 736–745

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Anomaly detection and identification scheme for VM live migration in cloud infrastructure☆

Tian Huang [a], Yongxin Zhu [a,*], Yafei Wu [a], Stéphane Bressan [b], Gillian Dobbie [c]

[a] School of Microelectronics, Shanghai Jiao Tong University, China
[b] School of Computing, National University of Singapore, Singapore
[c] Department of Computer Science, University of Auckland, New Zealand

HIGHLIGHTS

- We extend Local Outlier Factors to detect anomalies in time series and adapt to smooth environmental changes.
- We propose Dimension Reasoning LOF (DR-LOF) that can point out the "most anomalous" dimension of the performance profile. DR-LOF provides clues for administrators to pinpoint and clear the anomalies.
- We incorporate DR-LOF and Symbolic Aggregate ApproXimation (SAX) measurement as a scheme to detect and identify the anomalies in the progress of VM live migration.
- In experiments, we deploy a typical computing application on small scale clusters and evaluate our anomaly detection and identification scheme by imposing two kinds of common anomalies onto VMs.

ARTICLE INFO

Article history:
Received 3 March 2015
Received in revised form
11 May 2015
Accepted 9 June 2015
Available online 30 June 2015

Keywords:
Anomaly detection
Security
Virtualization

ABSTRACT

Virtual machines (VM) offer simple and practical mechanisms to address many of the manageability problems of leveraging heterogeneous computing resources. VM live migration is an important feature of virtualization in cloud computing: it allows administrators to transparently tune the performance of the computing infrastructure. However, VM live migration may open the door to security threats. Classic anomaly detection schemes such as Local Outlier Factors (LOF) fail in detecting anomalies in the process of VM live migration. To tackle such critical security issues, we propose an adaptive scheme that mines data from the cloud infrastructure in order to detect abnormal statistics when VMs are migrated to new hosts. In our scheme, we extend classic Local Outlier Factors (LOF) approach by defining novel dimension reasoning (DR) rules as DR-LOF to figure out the possible sources of anomalies. We also incorporate Symbolic Aggregate ApproXimation (SAX) to enable timing information exploration that LOF ignores. In addition, we implement our scheme with an adaptive procedure to reduce chances of performance instability. Compared with LOF that fails in detecting anomalies in the process of VM live migration, our scheme is able not only to detect anomalies but also to identify their possible sources, giving cloud computing operators important clues to pinpoint and clear the anomalies. Our scheme further outperforms other classic clustering tools in WEKA (Waikato Environment for Knowledge Analysis) with higher detection rates and lower false alarm rate. Our scheme would serve as a novel anomaly detection tool to improve security framework in VM management for cloud computing.

1. Introduction

Virtualization is the fundamental technology that powers cloud computing [1] as a middleware that separates physical infrastructures to create various dedicated resources, e.g. virtual machine (VM). Researchers and engineers have developed distributed computing systems on large clusters of low-cost commodity machines running VMs [2]. Examples of such initiatives include Google's

☆ Fully documented templates are available in the elsarticle package on CTAN.

* Corresponding author.
E-mail addresses: ian_malcolm@sjtu.edu.cn (T. Huang), zhuyongxin@sjtu.edu.cn (Y. Zhu), wuyf0406@sjtu.edu.cn (Y. Wu), steph@nus.edu.sg (S. Bressan), gill@cs.auckland.ac.nz (G. Dobbie).

http://dx.doi.org/10.1016/j.future.2015.06.005
0167-739X/

MapReduce, Hadoop from the open-source community, and Cosmos and Dryad from Microsoft. Applications running in VMs provide easy deployment, flexible resource provisioning, high utilization of server resources, and simpler administration. They can take advantage of different services and capabilities provided by the virtualization infrastructure such as live migration, elastic scale out, and better sharing of physical resources [3–5].

As an underlying technique of virtualization, live migration refers to the process of moving a running VM between different physical machines without disconnecting the client or application. Live migration brings the benefits of workload balancing, elastic scaling, fault tolerance, hardware maintenance, sharing infrastructures and transparency to users. However, the dynamic nature of VMs, especially live migration, makes them difficult to maintain the consistency of security.

For example, VMs' live migration among physical servers could spread security vulnerabilities and human negligence with an ignorant and rapid way. This will be a disaster against a pool of virtualized servers for production use, because there are generally no physical firewalls separating the VMs in a virtual environment [6]. Additionally, VM live migration adds difficulties to anomaly detection because the Inconsistent performance of VM before, in the course of and after the live migration cover anomalies and attacks [7].

Anomaly (or behavioral) detection is concerned with identifying events that appear to be anomalous with respect to normal system behavior [8]. Anomaly based approach involves the collection of data relating to the behavior of legitimate users over a period of time, and then applies statistical tests to the observed behavior, which determines the legitimacy of the behavior. It has the advantage of detecting attacks which have not been found previously. Recently, anomaly detection is extensively researched for cloud computing area [6,9,10] where complete knowledge of security issues are not available because of the heterogeneity and diversity of cloud environments. However existing anomaly detection methods are not ready to cope with the challenges VM live migration brings to the security management of cloud computing [7].

Existing methods detect anomalies but do not identify the root cause of anomalies. As security issues may spread rapidly through live migration, knowing the reason of the security issues and taking action in time will delay the spreading and reduce Mean Time To Recovery (MTTR). The heterogeneity and diversity of cloud environment hinder staff in manually identifying the cause. We believe that along with generating the alarm, Anomaly detection methods ought to provide staff the clues to pinpoint and clear the anomalies.

Besides, existing methods do not distinguish anomalies from the impacts of environmental changes. When VMs are migrated to a destination host whose workload, memory size, network condition, I/O speed and many other infrastructure settings are different from the origin host, the applications provisioned by the VMs may behave differently, making detection methods difficult to distinguish the anomalies and attacks from the impacts of environmental factors.

In this paper, we mitigate these security issues by proposing an anomaly detection and identification scheme for VM live migration. Our scheme incorporate time series analysis methods and Local Outlier Factor (LOF) algorithm to detect and identify the anomalous behaviors through the performance profile of virtual machine. This scheme works based on the fact that an application keeps its behavioral consistency and consumes similar amount of resources in the progress of VM live migration. Live migration and environmental changes impose consistent impacts on the performance profile of VM. Therefore our scheme identifies the legitimacy of behaviors of VM according to the similarity of its performance profile. Our major contributions can be summarized as follows:

- We extend Local Outlier Factors (LOF) [11] to detect anomalies in time series and adapt to smooth environmental changes.
- We propose Dimension Reasoning LOF (DR-LOF) that can point out the most anomalous dimension of the performance profile. DR-LOF provides clues for administrators to pinpoint and clear the anomalies.
- We incorporate DR-LOF and Symbolic Aggregate ApproXimation (SAX) [12] measurement as a scheme to detect and identify the anomalies in the progress of VM live migration.

In experiments, we deploy a typical computing application on small scale clusters and evaluate our anomaly detection and identification scheme by imposing two kinds of common anomalies onto VMs. We compare our scheme with other methods in the scenario live migration. Experiments show that our scheme detects anomalies more accurately and effectively over VM live migration and other environmental changes.

The rest of the paper is organized as follows. In Section 2, we discuss the related work and the background classic LOF and SAX algorithms. In Section 3, we give an overview and present detailed implementation of our scheme. In Section 4, we comparatively analyze the performance of our proposed scheme in contrast with classic LOF and with various clustering algorithms. Finally, we summarize our contribution and results and conclude in Section 5.

2. Related work and background

2.1. Related work

As a classic issue, anomaly detection has been explored by researchers in the areas of intrusion detection, machine learning, web semantics as well as relevant statisticians. The problem arose again with new technical and operational features in cloud computing systems, though there have been well formed methods for operations on classic computer systems.

Many efforts on intrusion detection draw on the experience of machine learning. Classification, clustering and Markov chain [13–16] are well studied for classic datacenters, while anomaly detection methodology in cloud computing system is still in its early stage. For example, Lo and co-authors [17] presented a scheme which successfully reduced the attack of DoS (Deny of Service) by collecting and mining network packets. However, the training phase of the method settles down the dividing line between normal and abnormal data. Therefore the effectiveness highly depends on the long term stability of infrastructure settings, which is contradictory to the characteristics of cloud computing.

Among methods handling anomaly detection in cloud systems [6,9,10], there are a few efforts on security issues in VM live migration. Adamova et al. [7] stated that virtual service migration can adversely affect state of the art anomaly detection techniques, potentially rendering them unusable. Wang et al. [18] proposed a role-based mechanism with remote attestation to leverage Intel vPro and TPM to improve security. VM migration is controlled by specific policies that are protected in seal storage under the mechanism. Xianqin2009seamless and co-authors [19] built a prototype system of the framework based on stateful firewall enabling Xen hypervisor. Alshawabkeh and co-authors [20] focused on feature selection algorithm to enhance detection accuracy in VM, but it did not consider VM migration scenario. Biedermann and co-authors [21] detected live migrations from the inside of an affected VM and measured security status of the VM before the migration finished. Oberheide and co-authors [22] investigated threats to VM migration and discussed strategies to address the deficiencies in virtualization software. These works focused on generic frameworks to detect anomalies and keep VM from attacks. However, none of them gave feasible ways to detect and fix anomalies. As the virtualization technology may incur anomalies that cannot be