Microelectronics Reliability xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Microelectronics Reliability

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

Introductory invited paper

Service reliability modeling and evaluation of active-active cloud data center based on the IT infrastructure

Xiaoyang Li [a,b,*], Yue Liu [a,b], Rui Kang [a,b], Lianghua Xiao [c]

[a] School of Reliability and Systems Engineering, Beihang University, Beijing, China
[b] Science and Technology on Reliability and Environmental Engineering Laboratory, Beihang University, Beijing, China
[c] Data Center of China Life Insurance Company Ltd., Beijing, China

ARTICLE INFO

Article history:
Received 25 December 2016
Received in revised form 27 February 2017
Accepted 10 March 2017
Available online xxxx

Keywords:
Cloud data center
Service reliability
Modeling and simulation
Monte Carlo
Queuing theory
Graph theory

ABSTRACT

As cloud data center has caught the eye of the information-intensive society since its birth, it keeps on a flourishing development due to its advantages such as high availability and resource utilization, rapid elasticity and disaster recovery. However, as a complex system, it means the centralization of failures and risks, which exerts great influence on the service that cloud data center provides to customers. Thus, the service reliability of the cloud data center is always a key concern. In order to evaluate and further improve the service reliability of the cloud data center, modeling analysis is absolutely necessary but difficult to apply because of the complicated cloud control flows, massive-scale service sharing and complex real-world infrastructures. Focused on the active-active data center, which is a typical mode of the cloud data center, a methodology of the service reliability modeling and analysis based on IT infrastructure is proposed. In this methodology, the queuing theory and graph theory are used to formulate the service reliability model, but the Monte Carlo simulation is used for statistical evaluation. During the modeling, the operational process of the active-active data center is divided into two parts—the request stage and execution stage, which are modeled respectively. Then, the modeling and evaluation approaches are applied in a use case, which verifies the applicability and creditability of our approach. Meanwhile, by sensitivity analysis considering the variation of uncertain or key factors both internal and external, several parameters are identified as high-sensitive factors, which can enlighten service providers on service reliability improvement.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

With rapid expansion and mature application of network information technology, the age of "big data" has already sneaked in. As the demand of data processing in all works grows increasingly, data center—a complete, complex and integrated set of facilities focusing on the centralized processing, storage, transmission and exchange of data, has become a communication hub gradually. From 1940s to early 2000s, data center has proceeded through mainframe, minicomputer and internet era. In 21st century, with the advent of cloud computing [1], data center ushers in a brand new cloud era and cloud data center emerges as the times require, due to its superiority in grand scale, service orientation, high density loading, automatic management, great flexibility and agility, on-demand service and low carbon. Serving as the carrier of cloud computing to provide service, the service reliability of cloud data center is always a key concern for both service providers and customers [2,3].

In recent years, sorts of risks and failures, including service interruption, service delay, access error and data loss have always happened to cloud data centers of cloud computing suppliers, such as Google, Amazon and Aliyun, which influenced their service reliability and led to economic damage and undesirable social impacts. Moreover, a technical definition of indexes related to service reliability such as mean time between failures (MTBF), mean time to repair or mean time to recovery (MTTR), percentage of running time, etc. is included in Service-Level Agreement (SLA) [4], which is generally signed between internet service providers and customers in order to ensure the quality of service (QoS). Obviously, it is necessary and essential to model and evaluate the service reliability of cloud data center.

As a complex system, cloud data center usually consists of the three main infrastructures [5]: the power infrastructure, the cooling infrastructure and the information technology (IT) infrastructure. The relationships between each infrastructure are shown in Fig. 1. As the energy source, the power infrastructure supplies conditioned power routing through uninterrupted power supply (UPS) at the correct frequency for both the cooling and IT infrastructures. The cooling infrastructure is composed of air-conditioning units like chiller plants, fans

* Corresponding author at: Science and Technology on Reliability and Environmental Engineering Laboratory, Beihang University, Weimin Building, Room 608, Beijing, China.
E-mail addresses: leexy@buaa.edu.cn (X. Li), liuyue@buaa.edu.cn (Y. Liu).

http://dx.doi.org/10.1016/j.microrel.2017.03.009
0026-2714/© 2017 Elsevier Ltd. All rights reserved.

Please cite this article as: X. Li, et al., Service reliability modeling and evaluation of active-active cloud data center based on the IT infrastructure, Microelectronics Reliability (2017), http://dx.doi.org/10.1016/j.microrel.2017.03.009

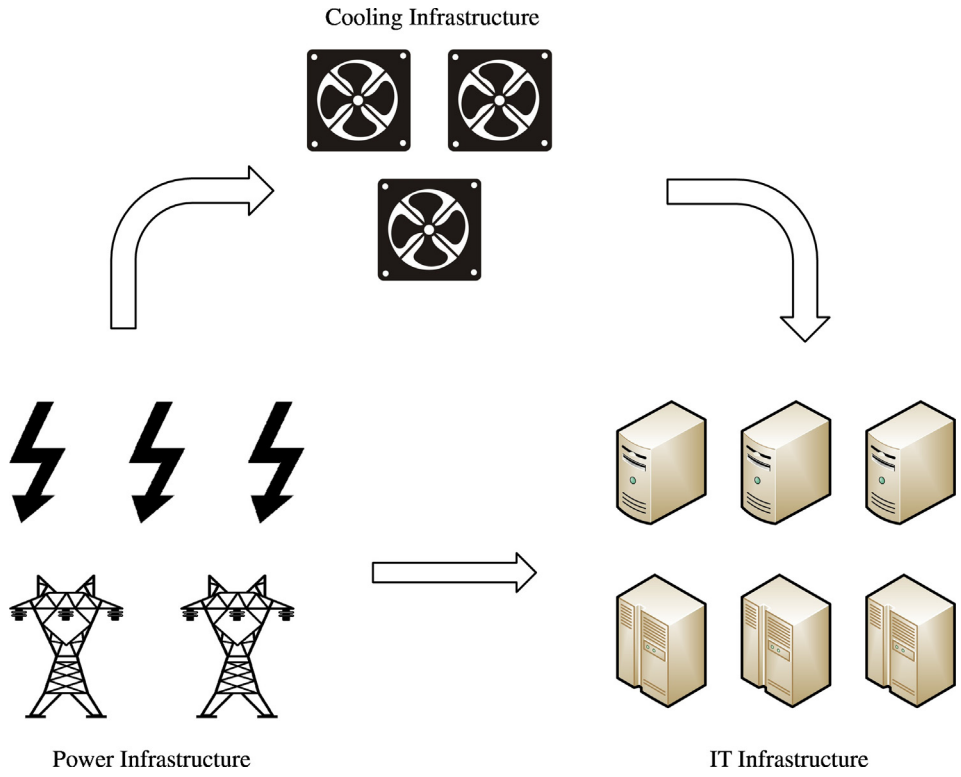

Fig. 1. Cloud data center infrastructure.

and air recirculation systems, which keep the temperature constant. The IT infrastructure refers to the composite hardware, software, network and other resources required for the existence, operation and management of an IT environment. With the sufficient energy and appropriate temperature guaranteed by the power and cooling infrastructure, the IT infrastructure allows the service delivery between service providers and customers. This paper mainly focuses on the IT infrastructure of cloud data center, which takes on the function of providing the service directly.

Focused on the IT infrastructure, the service reliability of cloud data center is affected by many aspects, such as the hardware, software, network, etc. Meanwhile, the service reliability of cloud data center is very critical but hard to analyze due to the characteristics of massive-scale service sharing, wide-area network, heterogeneous software/hardware components and complicated interactions among them. However, little has been done in the existed researches on the service reliability of cloud data center. Arno [6] applied the principles and modeling techniques of reliability engineering to specific examples for each tier from least (Tier 1) to most reliable (Tier 4) and discuss the results, which is a qualitative analysis without quantitative evaluation. Dai [7] concentrated on the cloud service process and used Markov models and Minimal Spanning Tree (MST) to model the service reliability in two stages, which however, is yet to be validated by simulation and real-life data. Zhang [8] considered repair and used MTTF/MTTR to build up the reliability model, which still cannot assess the system time-continuously. Pan [9] used an analytic method based on service process over time to calculate the reliability index through the state transition probability and its distribution, but analytic method has its limitation, like that the probability density function of processing time is difficult to calculate, the modeling is time-consuming and costs a lot in field measurement. Besides, other related researches mostly focus on the power infrastructure or just one part of the IT infrastructure, such as networks or storages. Dionise [10] proposed an evaluation of the power system focusing on the critical components in the utility service and some of the critical electronic loads, which called the power system audit, to analyze the reliability, safety and efficiency. Callou [5] proposed an integrated approach to estimate and optimize the sustainability, availability and dependability of the power system by using reliability block diagrams (RBD), stochastic Petri nets (SPNs), continuous-time Markov chains (CIMC) and energy flow (EFM) tools. Greenberg [11] presented virtual layer 2 (VL2), a practical network architecture to support huge data centers and evaluate the merits of the VL2 design by using measurement, analysis, and experiments. Hsu [12] analyzed the data storage reliability in data center from two aspects: disk array reliability and file system reliability.

Aiming at active-active data center—a typical and most widely used mode of cloud data center, this paper presents a novel method to model the service reliability concerning IT infrastructure based on queuing theory and graph theory, and also proposes its corresponding quantitative calculation and evaluation method based on Monte Carlo method. In the proposed model, various types of failures that influence service reliability in the whole service process have been comprehensively considered.

The organization of the paper is as follows. Section 2 discusses the basic knowledge including the IT infrastructure and operational process of cloud data center; then, the failure analysis is exposited and the service reliability is defined. Section 3 builds up a holistic model for an active-active data center from the request stage to the execution stage, through which the service reliability can be evaluated. Section 4 presents a real-world use case and the corresponding sensitivity analysis. Finally, Section 5 concludes the paper and makes suggestions to the future research.

2. IT infrastructure and failure analysis of cloud data center

2.1. IT infrastructure and operational process

IT infrastructure includes networking equipment, storage, servers and management equipment [13,14]. The communications in data centers are mostly based on networks running the IP protocol suite. Networking equipment contains not only a set of internet connections, switches, routers that transport traffic between servers and the outside

Please cite this article as: X. Li, et al., Service reliability modeling and evaluation of active-active cloud data center based on the IT infrastructure, Microelectronics Reliability (2017), http://dx.doi.org/10.1016/j.microrel.2017.03.009

Download English Version:

https://daneshyari.com/en/article/4971397

Download Persian Version:

https://daneshyari.com/article/4971397

Daneshyari.com