UMEÅ UNIVERSITET

Performance Anomaly Detection and Resolution for Autonomous Clouds

Olumuyiwa Ibidunmoye

Akademisk avhandling

som med vederbörligt tillstånd av Rektor vid Umeå universitet för avläggande av filosofie doktorsexamen framläggs till offentligt försvar i Salens namn eller beteckning, byggnad MA121, MIT-huset, 14:e dagen den december månad, 2017, kl. 13:15.
Avhandlingen kommer att försvaras på engelska.

Fakultetsopponent: Professor, Rolf Stadler,
Network and Systems Engineering, School of Electrical Engineering,
KTH Royal Institute of Technology, Stockholm, Sweden.

Department of Computing Science

Author
Olumuyiwa Ibidunmoye

Title
Performance Anomaly Detection and Resolution for Autonomous Clouds

Abstract
Fundamental properties of cloud computing such as resource sharing and on-demand self-servicing is driving a growing adoption of the cloud for hosting both legacy and new application services. A consequence of this growth is that the increasing scale and complexity of the underlying cloud infrastructure as well as the fluctuating service workloads is inducing performance incidents at a higher frequency than ever before with far-reaching impact on revenue, reliability, and reputation. Hence, effectively managing performance incidents with emphasis on timely detection, diagnosis and resolution has thus become a necessity rather than luxury. While other aspects of cloud management such as monitoring and resource management are experiencing greater automation, automated management of performance incidents remains a major concern.

Given the volume of operational data produced by cloud datacenters and services, this thesis focus on how data analytics techniques can be used in the aspect of cloud performance management. In particular, this work investigates techniques and models for automated performance anomaly detection and prevention in cloud environments. To familiarize with developments in the research area, we present the outcome of an extensive survey of existing research contributions addressing various aspects of performance problem management in diverse systems domains. We discuss the design and evaluation of analytics models and algorithms for detecting performance anomalies in real-time behaviour of cloud datacenter resources and hosted services at different resolutions. We also discuss the design of a semi-supervised machine learning approach for mitigating performance degradation by actively driving quality of service from undesirable states to a desired target state via incremental capacity optimization. The research methods used in this thesis include experiments on real virtualized testbeds to evaluate aspects of proposed techniques while other aspects are evaluated using performance traces from real-world datacenters.

Insights and outcomes from this thesis can be used by both cloud and service operators to enhance the automation of performance problem detection, diagnosis and resolution. They also have the potential to spur further research in the area while being applicable in related domains such as Internet of Things (IoT), industrial sensors as well as in edge and mobile clouds.

Keywords