

# Automated Heterogeneous Service Management in Cloud Systems

Christopher C. Lamb, Pramod A. Jamkhedkar, Gregory L. Heileman, Chaouki T. Abdallah

University of New Mexico

Department of Electrical and Computer Engineering

Albuquerque, NM 87131-0001

{cclamb, pramod54, heileman, chaouki}@ece.unm.edu

**Abstract**—Herein, we examine the problem of a single provider offering multiple types of service level agreements, the implications thereof, and finally simulate scenarios of a hypothetical provider managing these agreements under various stressful situations. We use usage management techniques in tandem with control-theoretic constructs to provide automated control and management of simulated system resources in the scenarios in question. In order to effectively engage with these types of techniques, we first analyze the system from both a usage management and control theoretic perspective, extracting service level objectives and indicators that are then profiled and managed.

## I. INTRODUCTION

The past few years have witnessed unprecedented expansion of commercial computing operations as the idea of cloud computing has become more mainstream and widely adopted by forward thinking technical organizational leadership. This rate of adoption promises to increase in the near future as well. With this expansion has come opportunity as well as risk, embodied by recent major service outages at leading cloud providers like Amazon. These issues promise to become more difficult to control as managed infrastructure expands. This expansion will simply not be possible without large amounts of automation in all aspects of cloud computing systems.

The current state of the art in cloud systems is poorly differentiated and not as customer-focused as it could be. Current providers place the responsibility of monitoring performance and proving outages on the consumer rather than providing more transparent and monitorable infrastructure [REF NEEDED]. Furthermore, providers as a whole only provide one type of service level agreement (SLA) in a loosely-defined one-size-fits-all type of arrangement [REF NEEDED]. This provides strong differentiating opportunities for smaller, second generation cloud system providers who have established the technology required to scalably manage multiple, competing SLAs on the same infrastructure in tandem with clear customer system visibility.

These second generation providers will rely on automated infrastructure management in order to scale, and one key area to automate is resource provisioning and performance management.

## A. Previous Work

## II. ANALYSIS

### A. Scenarios

### B. Service Level Features

### C. Usage Management Modeling

### D. Control Theoretic Modeling

### E. System Integration and Architecture

## III. SIMULATION

### A. Results and Analysis

## IV. CONCLUSIONS AND FUTURE WORKS

### A. Conclusions

### B. Future Works

[1]

## REFERENCES

- [1] Y. Chen, D. Gmach, C. Hyser, Z. Wang, C. Bash, C. Hoover, and S. Singhal, "Integrated management of application performance, power and cooling in data centers," in *Network Operations and Management Symposium (NOMS)*, 2010 IEEE, 2010, pp. 615 –622.