Measured Characteristics of Distributed Cloud Computing Infrastructure for Message-based Collaboration Applications

Geoffrey C. Fox
Informatics Department, Indiana University, 919 E. 10th Street Bloomington, IN 47408 USA

Computer Science Department and Community Grids Laboratory, Indiana University 501 N. Morton St., Suite 224, Bloomington IN 47404 USA

gcf@indiana.edu

Alex Ho, Eddy Chan, William Wang
Anabas, Inc., 580 California Street, Suite 1600, San Francisco, CA 94104, USA

{alex.ho, eddy.chan, william.wang}@anabas.com

ABSTRACT

While the emerging cloud computing systems promise infrastructure resources to support application scalability, there are publications of systematic evaluation of this emerging information technology infrastructure in general, and no obvious publications for some representative collaboration applications in particular. We describe a methodology to study the characteristics of distributed cloud computing infrastructure for message-based collaboration applications. We report our findings of distributed cloud infrastructures in a multitude of dimensions, including performance, scalability and reliability at the network level using standard network performance tools for easy comparison of results with large quantity of literature on classic, non-cloud environments; at the message level using the NaradaBrokering system [1] by the Indiana University Community Grids Laboratory which supports a large number of practical communication protocols; and at the collaboration and communication applications level using the Anabas Impromptu, a message-based Web Conferencing system for synchronous, multipoint data collaboration, Voice-over IP communication, and Video-over

IP conferencing traffics. Specifically, the Amazon EC2 Linux clouds in the U.S. East and Europe West were used because of their broad availability and relative maturity. Our preliminary observed data indicate clouds meet or surpass some stringent systems requirement for large-scaled, message-based collaboration applications.

INTRODUCTION

Cloud is an emerging information technology infrastructure designed for rapid delivery of computing resources. It promises to facilitate new applications deployment by highly efficient virtualized compute, storage, and network resources that can be rapidly scaled up and down in a flexible yet secure way to deliver a high quality of service. Cloud computing could potentially be a cost-effective model for provisioning large-scaled, globally deployed collaborative services such as SensaaS (Sensor as a Service) and XaaS ("X is anything" as a Service)[2] as cloud services while making the infrastructure more responsive to the needs of the services. We devise a methodology to study the characteristics of distributed cloud computing infrastructure at the network, transport messages, and message-based collaboration applications levels. Preliminary observed data indicate clouds offer good performance, scalability and reliability.

SOME PRELIMINARY RESULTS

Specifically, the Amazon EC2 Linux clouds are used as a cloud infrastructure testbed. Both the EC2-US East region and EC2-EU West region are used in our experiments. Each EC2 region is a separate cloud with its own set of components that are specific to that region. We use the following EC2 components in our instrumentation: Images, SSH keys, Elastic IPs, EBS (Elastic Block Storage) volumes, EBS snapshots and S3 (Simple Storage Service).

NETWORK-LEVEL MEASUREMENT

We report summary observation here. Iperf, a commonly used network performance tool for creating TCP and UDP data streams and measuring network throughput is used with our EC2-US East and EC2-EU West images. Bi-directional throughput data across the cloud trans-

Atlantic link was measured and the aggregate throughput for the instance recorded.

In the case when only one instance of Iperf pairs was launched in the EC2-US and EC2-EU clouds, the sub-cases of 1, 2, 4, 8, 16, 32, 64 and 128 connections were measured, accordingly. Figure 1 shows a how the inter-cloud, trans-Atlantic throughput in mbps (megabits per second) scales with the number of Iperf connections.

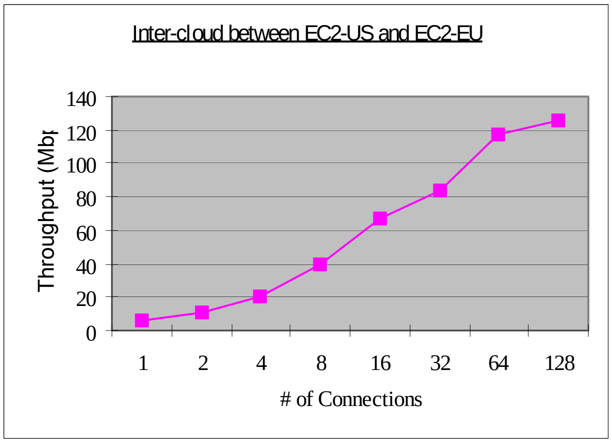

Figure 1. Throughput between 2 trans-Atlantic clouds

The EC2-US and EC2-EU inter-cloud sustains a throughput of 126 mbps at 128 Iperf connection. The maximum sustainable throughput has not been reached.

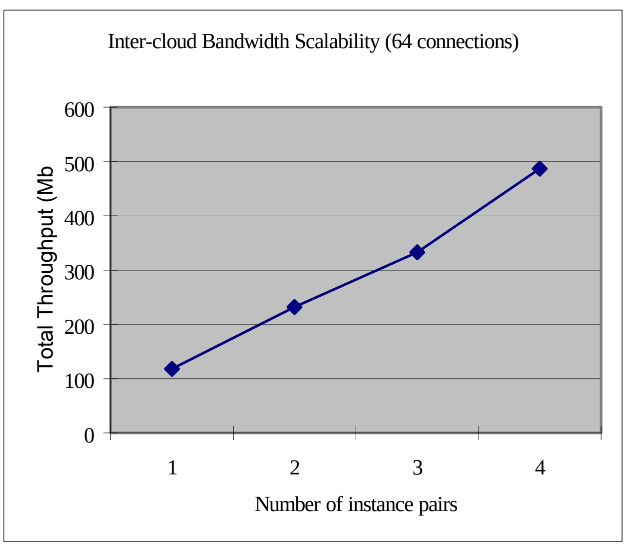

Figure 2. Scalability of total inter-cloud throughput

In another scenario, the number of Iperf connections was fixed at 64 while the number of instance pairs of Iperf was scaled up. Figure 2 shows the scalability plot for up to 4 instances. The virtually linear scalability confirms the isolation of network resource for each instance in the cloud. Even at 4 instances only, the inter-cloud, trans-Atlantic throughput is already measured at nearly 500 mbps. This very large network capacity could enable large-scaled, collaboration applications between the U.S. and Europe.

MESSAGE-LEVEL MEASUREMENT

In one set of experiments, extensive measurements have been made to evaluate cloud scalability, stability and reliability characteristics for an increasingly larger collaboration session by using NaradaBrokering messages. A NaradaBrokering server is launched with our EC2-EU West image while all message publishers and subscribers are launched with our EC2-US image. For each message between a publisher and subscribe, the message needs to traverse a distance twice of that between the EC2-US East and EC2-EU West clouds.

Even though we have an actual Voice-over IP application in the Anabas Impromptu that could be used to generate real VoIP traffic, it is easier to scale the number of publishers and subscribers at the message-level using NB clients than at the application level using real conference people for presenters and participants.

In this segment of the experiments, we mimic certain characteristics of VoIP traffic in a conservative manner at the message level. As reported in [3], small audio packet of the same size evenly distributed in time at 30 ms interval could be used. We used larger size 1 KB packets at a shorter interval of 12.5 ms instead to observe inter-cloud quality of service characteristics. It is important to recall that Cisco's VoIP system deployment guideline requires enterprise networks to be able to sustain at most 300 ms round-trip latency, average two-way jitter less than 60 ms, and packet loss less than 1% [4].

Besides the high-performance requirement of the message broker, the underlying network infrastructure performance is critical to the effectiveness of Voice-over IP sessions. We tested multiple VoIP meetings in the U.S with a fixed number of 200 participants in each meeting, serviced by a single broker in EC2-EU, and the results summarized in Table 1 below.

It is observed that in the inter-cloud link illustrates very satisfactory quality of service characteristics even when the clouds are located across the Atlantic ocean. There was only one case, the case of 11 VoIP sessions, each with 200 participants (a total of 2200 users), that should maximum round-trip latency of 318 ms, which is higher than the Cisco prescribed no worse than 300 ms round-trip latency. The same case also exhibited average jitter of

74 ms, which is higher than the Cisco prescribed no worse than 60 ms. It is worth-noting that the minimum round-trip latency is quite stable, about 90 ms round-trip, for all the test cases.

Table 1: Inter-cloud quality of service

Total Users	Min. 2-way Latency (ms)	Max. 2-way Latency (ms)	Average 2-way Latency (ms)	Average 2-way Jitter (ms)
200	90.15	124	99.51	16.70
400	91.09	133.81	108.38	26.92
600	90.61	155.79	109.80	28.67
800	91.21	183.69	107.56	29.67
1200	91.87	189.82	110.79	35.48
1400	92.18	165.74	106.39	38.69
1600	94.40	235.14	118.94	50.63
1800	93.56	197.89	110.80	33.77
2000	91.25	270.44	110.93	31.98
2200	108.30	318.08	151.66	74.33
2400	93.2	682.01	141.82	57.92

In addition to exhibiting satisfactory latency and jitter characteristics, the inter-cloud connections are very stable, with only 1 packet loss detected by using the ping tool for over a continuous session over 1.5 days, with a ping command issued every second over the duration.

APPLICATION-LEVEL MEASUREMENT

In full paper, we will add application level performance data.

KEYWORDS: synchronous collaboration, sensors as a service, SensaaS, XaaS, distributed cloud

BIOGRAPHY

GEOFFREY C. FOX received a Ph.D. in Theoretical Physics from Cambridge University and is now the chairperson, Department of Informatics and professor of Computer Science, Informatics, and Physics at Indiana University where he is director of the Community Grids Laboratory. He is chief technology officer for Anabas Inc. He previously held positions at Caltech, Syracuse University and Florida State University.

ALEX HO is the CEO of Anabas, Inc. He was a staff scientist with the IBM Research Division and the Caltech Concurrent Computation Program for over ten years. He was the founder and co-founder of several Silicon Valley startups in the areas of collaboration and Internet media technology.

EDDY CHAN is an R&D engineer. He has conducted extensive research in ad-hoc wireless network and Voice over IP, and is focusing on message-based collaboration technology.

WILLIAM WANG heads the Anabas engineering team. He was a co-founder of a startup on ubiquitous voice-enabled information access using speaker-independent speech recognition technology. He was the engineering lead for a mobile application that won the championship of a Nokia S60 Challenge, a global developer competition.

REFERENCES

[1] Shrideep Pallickara, Hasan Bulut, Pete Burnap, Geoffrey Fox, Ahmet Uyar, David Walker. “Support for High Performance Real-time Collaboration within the NaradaBrokering Substrate”, Technical Report May 2005.

[2] Geoffrey Fox, Alex Ho, Rui Wang, Edward Chu and Isaac Kwan, “A Collaborative Sensor Grids Framework”, IEEE International Symposium on Collaborative Technologies and Systems (CTS 2008) at The Hyatt Regency Irvine Irvine, California, USA May 19-23, 2008.

[3] Wenjun Wu, Geoffrey Fox, Hasan Bulut, Ahmet Uyar, Tao Huang, “Service Oriented Architecture for VoIP conferencing”, Special Issue on Voice over IP - Theory and Practice of the International Journal of Communication Systems Volume 19, Issue 4 , Pages 445 - 461 Edited by John Fox, P. Gburzynski.

[4] T. Szigeti, C. Hattingh, “Quality of Service Design Overview”, Cisco Press, 2004.