Stratecast *A Division of* FROST & SULLIVAN

Can Cloud and High Performance Co-Exist?

May 2011

CAN CLOUD AND HIGH PERFORMANCE CO-EXIST?

INTRODUCTION

Could the Cloud be too good to be true? The answer is yes.

On the positive side of the ledger, the Cloud combines the time-honored concepts of shared facilities (e.g., data centers and servers) and economies of scale with a dash of rapid provisioning to deliver pay-as-you-need outsourced IT resources. Because needs vary, Cloud services are offered in three models—Infrastructure as a Service, Platform as a Service, and Software as a Service— spanning raw computing capacity on which the Cloud subscriber overlays its unique workload environment to on-demand applications offered on a per-seat basis. At the surface, Cloud services represent an attractive means for organizations to increase their IT flexibility and optimize IT expenditures.

However, an inevitable ah-ah moment occurs regarding the question of performance. Will end-users experience the same responsiveness, the same degree of immediacy, with an application hosted in the Cloud as they do when that same application is executed locally on their PCs or hosted in the enterprise's Local Area Network (LAN)-connected data center? The answer is no. Distance and how communication networks and applications function inject time into end-users' Cloud sessions.

Will end-users experience the same responsiveness, the same degree of immediacy, with an application hosted in the Cloud as they do when that same application is executed locally on their PCs or hosted in the enterprise's Local Area Network (LAN)-connected data center? The answer is no.

Far from unique, performance degradation is also prevalent as organizations consolidate their applications into centralized data centers that they own or in space they rent in third-party data centers (frequently referred to as 'private' Clouds). Similar to the 'public' or Internet-connected Cloud, end-user application engagements in private Clouds are no longer on their PCs or over a LAN but over the lengthier, bandwidth-constrained connections of a Wide Area Network (WAN).

This injection of time—or visually, the uninvited on-screen hourglass image—robs end-users of productivity and elongates computing and information-dependent workflows. For many organizations, these outcomes are inconsistent with their objective to be operationally time efficient and, for some, to establish a competitive advantage. Furthermore, with communication networks forming the circulatory system connecting end-users with private and public Clouds, organizations must also consider the additional bandwidth required to support operations and applications consolidated into a few private data centers or re-located to the public Cloud.

Essentially, the Cloud's appeal—flexibility and optimized IT expenditures—is chipped away by sub-optimized end-user sessions and a potential uptick in bandwidth expenditures. Organizations justifiably exercise caution in choosing what they host in the Cloud (e.g., customized workloads and applications) and consume from the Cloud (e.g., on-demand applications).

In this article, we describe why this Cloud performance inhibitor exists and offer our perspective on what is needed to remove it.

WHY SO SLOW?

It is an inescapable fact that the Internet plays a prominent role in the public Cloud. By design, Cloud services are accessible from anywhere and there is no better way to accomplish this than being an Internet destination. But it is also a well-known fact that the Internet, the ultimate in a shared public network, is subject to performance issues, notably packet delivery delay stemming from several architectural attributes.

Congestion

The Internet consists of multiple backbone networks exchanging traffic at peering points. Equivalent to the peering points in mass transit systems—highway on-ramps and major intersections—traffic congestion occurs. When it does, there are delays in the movement of packets from point of origination to point of destination. Also, like mass transit systems, unplanned and exogenous factors hinder packet movement, such as severed cables (road closures), downed routers (accidents), and traffic bursts (rush hour traffic and seasonal travel).

Hops and Distance

While the Internet has built-in mechanisms to work around points of congestion, these mechanisms were designed with the objective to faithfully route packets from point to point without regard to delivery time. Consequently, the distance travelled and the number of hops among peering networks is not optimized to be the shortest or the fewest.

Regulated Pace

Third, the communication protocol used in the Internet was designed for packet delivery resiliency. To accomplish this objective, the pace at which packets are sent automatically adjusts for factors that impact packet delivery time, such as distance, number of network hops, and network congestion. As distance or the number of network hops increases or as congestion intensifies, the pace at which packets are sent is reduced. For end-users, pace reduction equates to slower response times with transactional and chatty applications and lengthier webpage rendering and file download times.

Retransmissions

Even with adjustments in the pace at which packets are sent, some packets still get dropped and retransmissions are required. This, too, negatively impacts the end-user experience.

When viewed collectively—traffic congestion, distance, number of hops, regulated packet sending pace, and retransmissions—the Internet has rightly earned the distinction of being a "best efforts" communications network.

The Cloud's appeal—flexibility and optimized IT expenditures—is chipped away by sub-optimized end-user sessions and a potential uptick in bandwidth expenditures.

♦ ♦ ♦

When viewed collectively—traffic congestion, distance, number of hops, regulated packet sending pace, and retransmissions—the Internet has rightly earned the distinction of being a "best efforts" communications network.

IMPOSSIBLE TO DECOUPLE THE CLOUD FROM THE INTERNET

If the Internet contributes to degradation in Cloud performance, can the use of the Internet be downsized? The answer is not entirely as described in the following circumstances.

Remote End-Users to Public Clouds

A primary reason for enterprise adoption of Software as a Service is to support remote access. As web applications, Software as a Service is a natural fit for browser-based access by remote and mobile end-users. However, these same users encounter the pains of the Internet's best efforts performance as they connect to applications hosted in Cloud data centers.

For many organizations, they direct remote end-users' Cloud sessions through a network gateway situated at the edge of their private WAN. Equipped with larger pipes to the Internet than end-users have individually, the notion is that performance issues between the gateway and the Cloud can be mitigated. To an extent this is possible, but far from complete.

- ***Network hops and distance increase*** – Remote end-users now have two Internet sessions, one that backhauls their Cloud-destined traffic to the gateway and the other from the gateway to the Cloud. Depending on location and Internet routing, the distance packets travel varies greatly. For example, a west coast telecommuter's request of a Cloud application is backhauled to the company's east coast gateway and then boomerangs to a Cloud data center situated on the west coast. The response from the Cloud application follows this same circuitous path in reverse. In this example, the distance this telecommuter's request and response travelled ballooned from hundreds of miles to thousands of miles.
- ***Bandwidth and throughput are not synonymous*** – Geographic distance does not change when an organization expands its Internet bandwidth. Inherently, distance is the same at 100 Mbps as it was at 10 Mbps. Since packet delivery delay is correlated with distance and the design of the Internet communication protocol automatically adjusts the pace that packets are sent (e.g., slower on longer distance routes), a 10x bandwidth increase will not equate to a 10x reduction in the time to complete an online task, such as downloading a file. While larger amounts of bandwidth at the gateway site will serve to reduce traffic congestion at that gateway location and assist the organization in supporting more simultaneous users (i.e., a bigger gate to pass through), improvement in individual end-users' Cloud performance is not assured.
- **Performance guarantees limited to the ISP's footprint – Internet** Service Providers (ISPs) do not quote performance metrics (e.g., latency, packet delivery, and availability) on other ISP's networks; they can only stand by what they have control over, their own networks. Consequently, individual ISP

performance guarantees are irrelevant to the actual end-to-end, cross-ISP performance for Cloud transmissions that traverse multiple ISPs.

HQ and Branch End-users to Public Clouds

This community of Cloud end-users is not as performance-challenged as remote and mobile end-users. In the backhaul-to-a-gateway scenario, the backhaul would typically occur over a private WAN allowing the organization's network administrator to possibly manage performance characteristics. These end-users, however, do not escape all the performance issues faced by remote end-users. There still is distance-induced delay (branch office to gateway and gateway to Cloud) and the transmissions between the gateway and the Cloud are subject to ISP performance variations and traffic congestion at peering points. In addition, the organization may need to upgrade the bandwidth at its branch locations and at the gateway to support higher traffic volume for Cloud applications.

The Cloud is Not a Single Destination

One of the virtues of Cloud applications is that the number of applications is expanding. For organizations, they benefit by having more application selections and, potentially, Cloud Service Providers. However, choice contributes to the distributed geographic footprint of Cloud applications and the reliance on the Internet as the universal means to connect. There is no escaping the Internet as part of the Cloud service delivery infrastructure for a growing number of applications.

APPLICATIONS ALSO IN PRIVATE DATA CENTERS

The current reality is that very few organizations rely exclusively on applications hosted in the public Cloud. Instead, applications reside in a mix of private data centers and, increasingly, the dynamic Cloud. As a result, end-users' connectivity to essential business applications is also a complex mix: WAN (branch locations to private data centers); WAN and Internet (WAN backhaul to gateway and then Internet); and even just the Internet (Internet backhaul to gateway and then Internet to the public Cloud).

The current reality is that very few organizations rely exclusively on applications hosted in the public Cloud. Instead, applications reside in a mix of private data centers and, increasingly, the dynamic Cloud.

Another page of reality is that the WAN is not immune to application performance issues. Whether packets traverse over a WAN connection or over the Internet, distance adds time to packet delivery. Also, congestion occurs on the WAN. Few organizations have unlimited budgets to unilaterally increase WAN bandwidth at each of their locations necessary to accommodate the growth in bandwidth demand as applications move out. Doing so eats into the cost savings that organizations expect by consolidating applications to fewer private data centers and subscribing to Cloud-based applications; and, as previously stated, increasing bandwidth is not a silver bullet in performance improvement. Last, the escalating information-intensive nature of business has contributed to greater use of sophisticated and intensely interactive applications and the exchange of large data files. Distance and congestion take an even deeper cut into performance than they do for less sophisticated applications and smaller data files.

Viewed collectively, performance matters—regardless of where applications are located and connectivity type. The challenge ahead is how to consistently ensure the appropriate level of performance as applications and application delivery diversify.

Viewed collectively, performance matters—regardless of where applications are located and connectivity type. The challenge ahead is how to consistently ensure the appropriate level of performance as applications and application delivery diversify.

It is our expectation that the technologies that have been successful in improving the performance of non-Cloud Internet transmissions and transmissions in private networks will find new life in organizations' distributed and dynamic application environments.

Stratecast

The Last Word

Are High Performance Cloud Experiences Possible?

We believe the answer is yes, but not yet. What needs to be recognized with this Cloud phenomenon is that it is in the early stage of development but also a stage that is experiencing strong interest and growing levels of adoption. Driving this, we believe, are the benefits of IT outsourcing that are facilitated through the Cloud as a flexible delivery mechanism. These benefits are just too compelling for organizations not to give the Cloud consideration. Nevertheless, the Cloud in its current form has shortcomings—performance being one.

It also bears recognition that the performance issues highlighted in Cloud use are not new. Reliance on the Internet for performance demanding applications is not new. Neither are performance issues that arise among distributed end-users and private data centers accessed through a Wide Area Network. Furthermore, the desire to continuously push the envelope on performance will never end. Being faster never grows old.

Therefore, it is our expectation that the technologies that have been successful in improving the performance of non-Cloud Internet transmissions and transmissions in private networks will find new life in organizations' distributed and dynamic application environments. While invention of new technologies is good, there is great benefit in re-aiming battle-tested technologies in new but similar scenarios.

The recently announced partnership between Akamai and Riverbed is a promising development on two fronts.

First, these companies have complementary technologies in Internet optimization and WAN optimization. Their collection of weapons to attack the performance issues of distributed applications is materially expanded.

Second, proximity to the application hosting environments is critical in performance improvement and bandwidth utilization. Again, the companies are complementary. Akamai, with its globally distributed EdgePlatforms, provides a new beachhead for Riverbed's symmetrical WAN optimization and application acceleration technologies for SaaS-delivered applications. Riverbed, with its traditional position where the 'WAN meets the LAN' and its virtual machine product instances for deployment in IaaS and PaaS environments, provides a new beachhead for Akamai's technologies.

If successful in their announced intentions, one plus one will be more than two.

Michael Suby

VP of Research

Stratecast (a Division of Frost & Sullivan)

msuby@stratecast.com

CONTACT US

- Beijing
- Bengaluru
- Bogotá
- Buenos Aires
- Cape Town
- Chennai
- Delhi
- Dubai
- Frankfurt
- Kolkata
- Kuala Lumpur
- London
- Manhattan
- Melbourne
- Mexico City
- Milan
- Mumbai
- Oxford
- Palo Alto
- Paris
- Rockville Centre
- San Antonio
- São Paulo
- Seoul
- Shanghai
- Singapore
- Sydney
- Tel Aviv
- Tokyo
- Toronto
- Warsaw

Silicon Valley
331 E. Evelyn Ave., Suite 100
Mountain View, CA 94041
Tel 650.475.4500
Fax 650.475.1570

San Antonio
7550 West Interstate 10, Suite 400,
San Antonio, Texas 78229-5616
Tel 210.348.1000
Fax 210.348.1003

London
4, Grosvenor Gardens,
London SWIW ODH,UK
Tel 44(0)20 7730 3438
Fax 44(0)20 7730 3343

877.GoFrost
myfrost@frost.com
http://www.frost.com

ABOUT STRATECAST

Stratecast assists clients in achieving their strategic and growth objectives by providing critical, objective and accurate strategic insight on the global communications industry. As a division of Frost & Sullivan, Stratecast's strategic consulting and analysis services complement Frost & Sullivan's Market Engineering and Growth Partnership services. Stratecast's product line includes subscription-based recurring analysis programs focused on Business Communication Services (BCS), Consumer Communication Services (CCS), Communications Infrastructure and Convergence (CIC), OSS and BSS Global Competitive Strategies (OSSCS), and our weekly opinion editorial, Stratecast Perspectives and Insight for Executives (SPIE). Stratecast also produces research modules focused on a single research theme or technology area such as Connected Home (CH), IMS and Service Delivery Platforms (IMS&SDP), Managed and Professional Services (M&PS), Mobility and Wireless (M&W), and Secure Networking (SN). Custom consulting engagements are available. Contact your Stratecast Account Executive for advice on the best collection of services for your growth needs.

ABOUT FROST & SULLIVAN

Frost & Sullivan, the Growth Partnership Company, enables clients to accelerate growth and achieve best-in-class positions in growth, innovation and leadership. The company's Growth Partnership Service provides the CEO and the CEO's Growth Team with disciplined research and best-practice models to drive the generation, evaluation, and implementation of powerful growth strategies. Frost & Sullivan leverages 50 years of experience in partnering with Global 1000 companies, emerging businesses and the investment community from more than 40 offices on six continents. To join our Growth Partnership, please visit http://www.frost.com.