NONTRIVIAL Q_OS. SWITCHING THE WEB TO THE GRID

Mircea Mare
"Petru Maior" University of Targu Mures
mircea.mare@gmail.com

ABSTRACT

The Internet provides no QoS, and from the looks of it, never will. We turn to the grid for a solution. In terms of QoS, existing grid technologies can bring a refreshing feel to the way we use the Web. The Internet evolved by interconnecting existing equipment (telephone lines) while quality and security measures were later added to it's agenda, after the issues occurred. Grid architecture is being designed in a more proactive mindset. Grids have started being used not only for scientific purposes but also for communication, collaboration and transferring data. Switching to a grid-based Web and adding nontrivial QoS to the equation is a natural step forward. The paper compares the evolution of the Internet in parallel to the evolution of grids, providing Grid QoS solutions and arguing in favor of the switch to a grid-based Web.

Keywords: grid, internet, qos, www, grid-based internet

1. Introduction

In S. Keshav's words[3]: The Holy Grail of computer networking is to design a network that has the flexibility and low cost of the Internet, yet offers the end-to-end quality-of-service guarantees of the telephone network.

Ian Foster's "three point checklist"[1] lists the following attributes of a grid: (1) computing resources are not administered centrally, (2) open standards are used and (3) nontrivial quality of service is achieved. What's the current Internet, in it's vastest meaning, missing from being entitled a grid? Let's take a look at the first attribute.

The Internet works as a large geographical distributed computer network, made up of many voluntarily interconnected networks. Although ICANN, by coordinating the assignment of unique identifiers, gets close to being a governing body, there is no central administration of the Internet. It can function well without ICANN, only less convenient.

The World Wide Web Consortium (W3C) pursues a process that encourages the development of open standards on the Internet and ensures that it's standards can be implemented royalty-free. HTML, CSS, HTTP and XML are just a few of the well known open standards they promote.

Tackling the 3rd attribute, related to QoS, will serve as the main focus of the paper, and to properly address it, we should start with the very beginning.

2. The birth of the Internet vs. the birth of the "global grid"

As a first striking similarity, the primordial cause for the birth of the Internet came from the scientific community. It's predecessor, Arpanet was a distributed computing network that had the purpose of sharing CPU-intensive tasks, which required resource brokerage, extensive monitoring and analysis. Surprisingly, it also gave birth to e-mail and FTP. The grid initially evolved from a desire to connect and convert supercomputers used for scientific research into "metacomputers" in order to be remotely controlled and administered.

It's a well known fact that that most of the Internet network still uses telephony cable, lacking the ability to send large amounts of data. By contrast, grid projects are being built with dedicated fiber optic cables and modern routing centres. CERN, responsible for the LHC (Large Hadron Collider) project uses over 200,000 servers needed for analyzing the huge amounts of data they produce.

The computing power was still not sufficient so they decided the only viable solution was to build a new network, parallel to the Internet, using fiber optic cables to connect 11 research centers scattered throughout the world[2], setting the outline for a global grid.

3. Internet Qos vs Grid QoS

Internet QoS is almost non-existent. The Internet operates on a "best effort" model. No guarantees for data delivery, service level quality or priority are given. Every user obtains unspecified variable bit rate and delivery time, depending on the current traffic load.

In the mid nineties universities and research centers began outgrowing the Internet's bandwidth limitations and higher demands for compute power appeared. This lead to the birth of advanced networking consortiums relying on heavy usage of IPv6, VPNs and dynamic circuit networks. To name a few: the very-high-performance Backbone Network Service (vBNS – which later evolved into Internet2) in the US, and in Europe: Kennisnet and SURFNet located in The Netherlands. These networks rely on the so called "bandwidth-on-demand" model.

Efforts have been made to implement QoS specifications in the advanced network consortiums. The Internet2 Network set out to deploy the QBone Premium Service (QPS), an interdomain virtual leased-line IP service built on traffic shaping and differentiation primitives. The endeavor never succeeded and has been suspended indefinitely.

The conclusions reached[3] on the difficulties of deploying QoS standards on the Qbone Premium Service apply not just to the QPS, but to any IP QoS architecture. The most prominent issues are: QPS' poor incremental deployment properties, intimidating new complexity for network operators, missing functionality on routers, and serious economic challenges[3].

Commercial ISPs connected via Internet exchange points seeking to provide services for as many users as possible, regardless of the QoS provided. On the other hand, research networks connected into large sub-networks aiming for one goal: more compute power. From a grid perspective, the DAS[1] came to life in a similar manner, employing internal interconnection based on light paths.

Surprisingly, up until today there is no wide-spread agreement on how a massive upgrade of the Internet, with QoS in mind, should take place. Any work done on Internet QoS will just help for a little while until traffic shaping and differentiation mechanisms become too congested, then the QoS algorithms will simply do more harm then good, and won't scale to novel transport technologies, nor new implementations of IP standards.

In a grid perspective, quality of service translates into resource reservation mechanisms instead of the achieved service quality. The reservation and brokerage middleware deals with the problem of allocating resources to whom and where they are needed the most. The middleware provides guarantees to the users of their reservation status and resource quality. These are specified by Service Level Agreements (SLAs). One of the most successful middleware projects that provides such coordination is the well known Globus project. In terms of security, the Grid is being designed from the ground up with security in mind, unlike the Internet which added security measures to it's agenda, after the issues occurred. Additionally, existing security implementations in distributed computing deal with client-server issues and do not address the issue of creating N-way security contexts, large user and resource sets, or local mechanism/policy heterogeneity. In contrast, the Grid Security Infrastructure (GSI), provides specifications for secure communication between grid applications[4], since, for example, parallel computations require security relationships between potentially hundreds of compute nodes across many administrative domains.

Several grid-based architectures, including Grido[5], MOB[6] (running of top of Ibis and taking advantage of the SmartSockets library for custom routing of data) can facilitate QoS requirements, while providing a long sought-after feature on the Internet: multicasting.

4. Enforcing QoS in a Grid-based World Wide Web

Applications in grids are evolving. Grids are becoming more than just platforms for compute-intensive tasks. They are not only connecting compute nodes, but people. More advanced – collaborative - applications are starting to be deployed on grids, like holographic conferencing, tele-immersion, virtual reality and interactive learning environments. These require data to be processed in a timely manner, adhering to stricter QoS requirements.

To address this problem various – rather new – implementations which aim at achieving the desired QoS are proposed. G-QoSm[7] – Grid Quality of Service Management - is a QoS-aimed grid architecture integrated with the Java CoG Core kit, which offers QoS tweaking facilities based on specific application criterias such delay, jitter, throughput, packet-loss rate, network latency, CPU performance, storage requirements etc.

QoS aware services can be provided for grid conferencing service so that they can be adaptively contextualized for each participant. These services have been successfully integrated into existing learning environments, the results[8] are promising. This approach can be summarized as "adapt rather than assume the best and then suffer the worst ".

Advance reservation of grid resources (for e.g., the 3-layered negotiation protocol which implements smart, offer-generation algorithms[9]) along with SLA management entities (for e.g. SNAP[11] - Service Negotiation and Acquisition Protocol) which support grid QoS[8] can play a crucial role in enabling Grid middleware to deliver on-demand resource provision (including deadline-bound architectures like VAS[12]) with greatly improved QoS.

5. Conclusion

A Global Grid will render the current Internet obsolete. The Web simply has too many underlying design flaws. While learning from the lessons of the past, a Global Grid provides the ideal platform for communication, collaboration and data transfer with a quintessential piece: guaranteed quality-of-service.

Furthermore, by comparing the two models (best effort vs. QoS-oriented approach) in a grid test environment[7], the QoS centered approach performs better than a best effort model.

1 http://www.cs.vu.nl/das3/

From a developer's perspective, writing applications without worrying about the inherent shortcomings of the Internet that we've been taking for granted so far sounds very promising.

References

[1] Foster, I. - *What is the Grid? A Three Point Checklist*, GRIDToday, July 20, 2002

[2] Hoffmann, U. - *Grid Computing, E-Science and Applications in Industry*. Wirtschaftinformatik Proceedings, 2005

[3] Teitelbaum, B., Shalunov, S. - *Why Premium IP Service Has Not Deployed (and Probably Never Will)* - Internet2 QoS Working Group Informational Document, May 3, 2002

[4] Foster, I., Kesselman, C., Nick, J., Tuecke, S. - *A Security Architecture for Computational Grids*, Proceeding CCS '98 Proceedings of the 5th ACM conference on Computer and communications security ACM New York, NY, USA, 1998

[5] Das, S., Nandan A., Parker, M. G., Pau, G., and Gerla, M. - *Grido - An Architecture for a Grid-based Overlay Network*, Proceedings of Second International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, QShine, Orlando, FL, USA, August. 2005

[6] Den Burger, M., Kielmann, T. - *MOB: Zero-configuration High-throughput Multicasting for Grid Applications*, Proceedings of The 16th IEEE International Symposium on High-Performance Distributed Computing (HPDC), Monterey, CA, USA, June 27-29 2007

[7] Al-Ali, R.J, Amin, K., von Laszewski, G., Rana, O. F., Walker, D. W., Hategan, M., Zaluzec, N. - *Analysis and Provision of QoS for Distributed Grid Applications*, WiCOM'09 Proceedings of the 5th International Conference on Wireless communications, networking and mobile computing IEEE, Press Piscataway, NJ, USA, 2009

[8] Bateman, M., Allison, C., Ruddle, A., Nicoll, R. - *An Advisory QoS Service for GRID Based Learning Environments*l, 4th International LeGE-WG Workshop - Towards a European Learning Grid Infrastructure: Progressing with a European Learning Grid, Stuttgart, Germany, April 2004

[9] Siddiqui, M., Villazon, A., Fahringer, T. - *Grid Capacity Planning with Negotiation-based Advance - Reservation for Optimized QoS*, SC '06 Proceedings of the 2006 ACM/IEEE conference on Supercomputing, ACM, New York, NY, USA, 2006

[10] Sahai, A., Graupner, S., Machiraju, V., van Moorsel, A., - *Specifying and Monitoring Guarantees in Commercial Grids through SLA*, Cluster Computing and the Grid, 2003. Proceedings. CCGrid 2003. 3rd IEEE/ACM International Symposium on In Cluster Computing and the Grid, 2003

[11] Czajkowski, K., Foster, I., Kesselman, C., Sander, V., Tuecke, S. - *SNAP: A Protocol for Negotiating Service Level Agreements and Coordinating Resource Management in Distributed Systems*, Proceedings of the 8th Workshop on Job Scheduling Strategies for Parallel Processing, 2002

[12] Keahey, K., Motawi, K. - *The Taming of the Grid: Virtual Application Service*, Argonne National Laboratory Technical Memorandum No. 262, May 2003.