Over-provisioning vs QoS
John McCabe

Abstract

Certain communication requires high bandwidth, while other communication requires low latency. Quality of Service can aid computer networks to direct traffic such that the perceived performance is within these constraints. What is the best method to providing Quality of Serivce?

The author takes the position that more experimental research is needed to see if Quality of Service is possible of unprecedented workloads. Packet level QoS and over-provisioning are too limiting in common non-adversarial workloads.

This paper will provide evidence to suppport such a position, and motive why this needs to be researched.

Introduction

Computer networks delivery news, entertainment, financial agreements, and many other daily communication events. This is so commonplace that humans are seemingly dependent on these networks to orient themselves in reality. For instance, many Americans check hourly the stock market trends to see if their pensions are in good health. They do this as the mean real wage has decreased over the past 40 years while GDP has doubled. Of course, reloading a webpage will not influence this fact. However, many humans appear to have learning disabilities when confronted with reality.

Many humans also require a certain Quality of Service (QoS) per application. The most demanding, with respect to QoS, network application is gaming. If a game is too slow, it may be impossible to play. One of the least demanding network applications is file transfer.

QoS can provide better service to selected network traffic [FER1998]. Three metrics define the behavior of the particular QoS system: bandwidth; jitter (i.e., latency); and, acceptable frame loss. Guaranteeing selected network traffic a certain level of priority should not break the network. In other words, the 'unselected' network traffic may be slower with QoS but should not fail due to QoS.

If cost is not an issue, why not purchase twice the bandwidth, links, routers, etc. to ensure quality performance? This is called over-provisioning. It is quite common for ISPs to use this method [ATK2001].

QoS and over-provisioning are complimentary. QoS works within the constraints of the network bandwidth. If more bandwidth exists, the stress on QoS is decreased. If a major event (e.g., 9/11) occurs over-provisioning by itself will not solve the problem.

Different applications and workloads use different networks. For example, a user may use VoIP on a wireless local area network to talk to a friend down the hall. QoS of protocols for the Internet's backbones will have different objectives than for 802.11. This fact will be discussed in the Section Unlimited QoS.

Position Statement

It is the position of the author that protocol level QoS and over-provisioning cannot realistically guarantee minimal quality of service for an unprecedented workload. According to the end-to-end principle [SAL1984], each problem must be solved at the proper level. Since peaks in traffic are from users, the problem should be solved at the user level. No amount of over-provisioning is enough, assuming a finite budget. No protocol level QoS can make near-perfect use of the network given a high workload. In fact, protocol level QoS is likely to be more expensive than over-provisioning.

Cost

The most important parameter in providing Quality of Service is the budget. If we cannot buy extra bandwidth, it is likely

we cannot afford network administrators to begin with. If the budget is non-positive, a different problem must be solved. So, we assume the budget is positive.

Due to naming, over-provisioning may sound rather expensive. It is. However, for protocol level QoS is work correctly all applications, OSes, and network devices must support the same QoS protocol [INT2008]. This may be very expensive. Imagine the nightmare this is for roommates who share a wireless LAN.

In fact, there is a belief in the network business that over-provisioning is always cheaper. Ran Atkinson [ATK2001] wrote, "[a]s to the costs of over-provisioning, once one has access to dark fibre, it is not particularly more expensive (in capital costs) to light it with Gigabit Ethernet than 10baseF or 100baseFX ... It seems more common that big providers have access to dark fibre than the smaller providers, so this might mean that smaller providers have more incentive to deploy QoS; I'm not sure."

Assume, smaller providers have smaller budgets. Why would they spend money on a protocol level QoS, which may or may not require massive infrastructure upgrades to be compliant with a particular protocol level QoS? If the ISP can afford neither protocol level QoS nor over-provisioning then users may leave for a better ISP. If a fraction of the users leave, the performance of the network should increase. It is not possible to fully understand the nature of this example without more information [GLE1988]. However, assume this small ISP can affording their bandwidth. It is possible that if they one day acquire packet level QoS, the packet level QoS is work with a greater bandwidth constraint.

Unlimited QoS

For only I had unlimited time.

The story in this section is to progress from a magical world of dynamically physically redesigning the entire network (including all applications and OSes) toward a plausible solution where changing the protocol is possible. This plausible solution also requires admission control.

Acknowledges

I would like to thank John-Austen Francisco for a fun discussion on QoS. Also I would like to thank the PLDI committee for their early November submission deadline.

References

[ATK2001] Atkinson, R. QoS vs Bandwidth Overprovisioning. End-to-End mailing list (April 2001). http://www.postel.org/pipermail/end2end-interest/2001-April/000615.html retrieved on Oct. 24th, 2008.

[FER1998] Ferrari, D. The Tenet experience and the design of protocols for integrated-services internetworks. Multimedia Systems Vol. 6 Issue 3. 1998.

[GLE1988] Gleick, James. Chaos: Making a New Science. Penguin. December, 1988.

[INT2008] Intel. Providing QoS in WLANs: How the IEEE 802.11e Standard QoS Enhancements Will Affect the Performance of WLANS. http://www.intel.com/network/connectivity/resources/doc_library/white_papers/303762.htm retrieved Oct. 24th, 2008.

[LOP 2007] Lopez-Ortiz, A. Valiant Load Balancing, Capacity Provisioning and Resilient Backbone Design. Lecture Notes in Computer Science Volume 4825. Combinatoral and Algorithmic Aspects of Networking (2007).

[SAL1984] Saltzer, J., D. Reed, and D. Clark. End-to-end Arguments in System Design. ACM Transactions on Computer Systems, Vol. 2, No. 4, November 1984.