Are IP Networks Suitable for Voice?

Bruce Threewitt, Vice President of Marketing, Omnivergent
(bruce@omnivergent.com)

Introduction

The IP network architecture uses a connectionless and stateless approach to provide any-to-any connectivity with the reliability of a meshed network of peers over a wide area (hence the name "WAN" or Wide Area Network). These networks use in-band control with distributed, autonomous control logic centered on routers. By managing the packet traffic at Layer 3 (of the ISO stack[1]) many of the interworking problems arising from incompatible networking hardware systems were overcome, and the Internet became very widely adopted for system-to-system communications. More and more applications are utilizing the Internet for readily available connectivity at prices lower than dedicated connections. That's the good news.

The installed base of routers was engineered to carry bursty traffic whose delivery was not time-critical. Attempting to retrofit this burst-oriented network to transport stream-oriented applications is ill-advised. Constant-bit-stream applications, such as voice over IP (VoIP) and interactive video, require predictable network performance. Because the network comprises a collection of peers (Autonomous Systems), limited to in-band communication across the Bearer Plane, effective control cannot be imposed from inside the network. Consequently, routes and performance are unpredictable, preventing or impeding IP networks from adequately supporting streaming data applications, which will dominate expected hyper-growth in traffic over the Internet. That's part of the bad news.

Carriers often vastly over-provision (or under-utilize) their networks to offset some of these performance-related issues, a strategy that cannot be economically sustained. This strategy raises Capital Expenses (CAPEX) and Operating Expenses (OPEX) without offering an offsetting increase in revenue to sustain the cost increases. The industry needs to raise income and lower expenses to avoid having so many incumbents (80% of major U.S. carriers) teetering on the edge of bankruptcy.[2] This unviable economic model comprises a second part of the bad news.

As more and more of the Nation's economic and communications infrastructure moves to IP networks like the Internet, we find ourselves vulnerable to an insidious, hard-to-detect and counter, and potentially devastating kind of attack. Network (a.k.a. Cyber) attacks are increasing in frequency and sophistication. The base of attackers is growing since the tools on which many attacks are based are readily available on the Internet. Hacking the network is not just a hobby anymore. Cyber-terrorism has the potential to disrupt the economy and wreak havoc with the physical infrastructure, such as the Power Grid.[3] The IP network architecture has gaping fundamental security flaws. Even if the performance limitations and economic flaws of the Internet could be patched, the network itself and the data it carries are not sufficiently secure. This problem is raising its ugly head regularly in the news. That's the third piece of bad news this article will discuss.

The Communications industry has invested billions of dollars in vain attempts to remedy these flaws without reconsidering the fundamental architecture of the network. Example "remedies" include Multi-Protocol Label Switching (MPLS), Route Optimization, Deep-packet Discovery, flow-based routing, *et alia*. None of these technologies has succeeded in solving all the technical problems or addressing the fundamental control theory issues.

So, read on as we discuss three problem areas in IP network architectures that must be addressed:

- Performance Issues
- Economic Issues
- Security Issues

[1] See for reference, http://www.archivists.org/glossary/term_details.asp?DefinitionKey=1438, an overview of ISO7498.
[2] "The 'Insolvency Zone': the Bankrupting of the U.S. Telecom Sector", Scott Cleland, Precursor Group®, May 20, 2002
[3] See for reference, http://www.crime-research.org/articles/Cyber_Terrorism_new_kind_Terrorism/

Performance Issues

Routers analyze every packet at every hop, making routing decisions based on various aspects of link availability, while being unaware, in a timely enough manner, of other key factors such as network congestion (beyond the next hop), latency, required security levels, Carrier business models, etc. In effect, the control system for this network has a very long feedback loop (seconds to minutes [IGP] and minutes to days [BGP]), relative to much faster events (tens of milliseconds) that must be controlled in such applications as VoIP. This trait earns the label "stateless" for IP networks, the converse of which would be "stateful". The Public Switched Telephone Network (PSTN) is a "stateful" network. Absent radical over-provisioning, a stateless network cannot route traffic with sufficiently predictable delays (latency) to assure Quality of Service, especially if more than one Carrier is involved in the transport.

Multi-Protocol Label Switching (MPLS) attempts to add determinism to a route (a virtual connection). This connection-oriented protocol has several limitations, however. The effect of distributing labels to the routers is like establishing a static "pipe" through which the traffic is sent across the network. This method alleviates the latency issue somewhat within a given Autonomous System (AS), but does not work across multiple Carriers. Further, the "pipe's" capacity cannot be modified quickly enough to account for varying traffic profiles, so the bandwidth carried by the "pipe" can be wasted. In other words, one pays for the "pipe" whether it is fully utilized or not. Other limitations prevent MPLS from addressing the economic and security issues raised below.

A basic maxim of control theory is that one cannot control what one does not (or cannot) measure. Routers have no means to adequately measure the performance of the links they employ for a given route. At best the router might be aware of a broken link to a router at the next hop. Information about broken links or failed routers farther downstream is not available quickly enough to be of use. Since the network has no universal time reference, no means currently exist to measure link performance with sufficient accuracy. Latencies in modern routers are estimated by using a "ping", a round-trip, in-band control probe packet. The round-trip time is divided by two to estimate the one-way latency. This method is not accurate enough because the trip out and the trip back are not necessarily over the same links (routes).

Since there is no stateful coherent control of a router-based IP network, routing decisions cannot be made in the context of Carrier business models. In other words, the classes of service a given Carrier wishes to offer cannot be matched against the current network transport resources. Optimal performance of the network would require such matching to avoid wasteful use of high-quality links on applications that do not require premium performance. There is no viable means to differentiate one traffic type from another, e.g., voice versus email. Several attempts, e.g., DiffServ, have been made to remedy this problem, but none have fully addressed the issue. Although related to performance optimization, this facet of router-based networks also adversely affects the economic picture.

Performance problems aside, there are some functional issues in using VoIP to replace PSTN voice services. For example, regulators deem it necessary to maintain power to the telephones during power outages. Further, voluntary compliance with the Communications Assistance for Law Enforcement Act (1994) or CALEA will earn Carriers a certification, boosting their competitive posture. IP networks cannot offer CALEA compliance due to their connectionless nature and a lack of coherent control.[4]

[4] See "DoJ: VoIP providers avoiding CALEA mandate" dated September 10, 2004 at http://www.americasnetwork.com/americasnetwork/article/articleDetail.jsp?id=121788

Economic Issues

The two economic factors of interest are revenue and cost. Cost can further be divided into Capital Expense (CAPEX) and Operating Expense (OPEX). Let's treat these individually.

Router-based Networks (remember that they are connectionless and stateless) have no way to track usage or separately value various types of premium services. The IP transport solutions offered to Carriers by the equipment providers, including MPLS, are not capable of generating transaction records, making usage-based or premium service billing all but impossible. Today's IP Networks do not afford a viable economic model for Carriers. Absent a means to apply value-based pricing to the wide variety of traffic on a converged network, the Carrier must resort to flat fee (subscription-based) billing. Worse, if price is the only parameter on which competition can be based, the industry as a whole will spiral toward zero revenue until all Carriers except the "last man standing" have gone out of business. The telecommunications industry reported worldwide revenue in 2004 in excess of $1 Trillion.[5] In order to support a revenue base of this size, each of the estimated 950 Million worldwide telephone subscribers[6] would need to spend just under $100 per month. This flat-fee subscriber model is clearly not feasible. A healthy telecommunications industry will require a means to offer new, profitable, usage-based revenue-generating services.

We have already asserted that even approximating acceptable QoS, especially for voice, within a given Carrier's network requires radical over-provisioning. Carriers, as a rule, do not maintain peak loads for their networks beyond about 40% of capacity. Actual typical utilization is well below 10% in most Carriers' networks. The reason for this policy relates to the so-called Price of Anarchy [7] (ratio of total latency of a Nash-equilibrium flow to the total latency of a minimum-latency flow) where packets suffer an arbitrarily large latency penalty as instantaneous traffic levels approach maximum capacity. This greater-than-10X over-provisioning represents a very heavy economic burden on Carriers who wish to utilize IP networks for any part of their infrastructure. This cost is only the CAPEX part of the total cost equation.

For each network element Carriers adds to their infrastructures, additional OPEX also accrues due to maintenance and training required for these complex elements. Hence, the OPEX grows as network traffic expands at a rate that is the reciprocal of the network's efficiency (overall utilization). These expenses are incurred every year over the life of the network elements deployed. The average revenue per user (ARPU) unfortunately stays flat or declines over time due to competitive forces (remember Carriers are stuck with competing solely on price). Anyone can see that this model is a recipe for economic ruin.

Additional OPEX occurs due to the manner in which services are deployed (provisioned) and supported. If each new service requires new support resources, and little or no automation is available, time-to-breakeven for new revenue opportunities is too long. The time to provision new IP network resources, largely a manual, labor-intensive process, can be scheduled with a calendar. Carriers need computer-automated provisioning processes to sufficiently reduce OPEX for a healthy profit picture.

Security Issues

Connectionless transport and in-band stateless control are some of the IP network traits that contribute to security flaws. Some years ago, the PSTN was vulnerable to fraudulent use by so-called "phone-phreaks."[8] The basic technique for hacking telephone security at that time in order to obtain fraudulent free long-distance service involved spoofing the signaling network via mimicking the tones used to communicate control information among the control nodes. Such attacks were only possible because

[5] See Information Technology Association of America's E-Letter of February 2004 at http://www.itaa.org/isec/pubs/e20042-04.pdf.
[6] See remarks of Alexander Gray, Lucent Technologies, at the SCO 1999Forum.
[7] See T. Roughgarden, *The Price of Anarchy is Independent of the Network Topology*, Proceedings of the 34th Annual ACM Symposium on the Theory of Computing, May 2002.
[8] For explanatory and historical background on "phreaking", see http://www.telephonetribute.com/phonephreaking.html.

the control signals were “in-band” with the bearer (voice) traffic. The Carriers eliminated this flaw by establishing a separate “out-of-band” signaling overlay using a packet-based network and a protocol called X.25. Consequently, users had no access to the signaling network via a handset. Hence the PSTN became much harder to hack. IP networks rely on in-band control protocols. At last count, there are more than 200 such control protocols, each of which represents a security vulnerability that hackers can and do exploit. Imagine your consternation if a router that should be conveying your 911 call is undergoing an ICMP Echo Request Denial of Service (DoS) attack!
Security professionals are all concerned that much of the Nation’s critical infrastructure is migrating toward the Internet or other IP networks. A CTO of a major global Carrier once lamented that a teenaged hacker in his basement in his pajamas hacking a key softswitch on a Friday night could ruin the CTO’s whole weekend! Such concerns increase as the portion of voice traffic using VoIP grows. Even private IP networks are not immune. A wide range of studies suggests that over 66% of security breaches in infrastructure systems are perpetrated from inside the organization.[9]

A distributed DoS (DDoS) attack is even more sinister. A DDoS attack is launched from a number of surrogate attack agents from all over the network, even utilizing multiple Carriers. The autonomous distributed control of the IP networks prevents even detecting, never mind analyzing or mitigating, this kind of activity until it is too late. A stateful, coherent control system would be required to prevent this type of attack.

Unfortunately, the data communications industry is focusing on Applications layer (Layer 7 in the ISO stack) where encryption and passwords reside. Critical infrastructure users of IP networks are more concerned about (software) patch management than about attacks directed at routers. Intrusion detection and prevention programs are Applications. They may not prevent or detect attacks on the network itself. As always, service providers must constantly balance risks of attacks against costs associated with prevention. Because security flaws are fundamentally a part of the IP network architecture, one must again ask, are these networks suitable for handling critical voice traffic?

In summary, IP networks, such as the Internet, have faithfully provided the features and services for which they were originally designed: ubiquitous, robust any-to-any connectivity that transports non-time-critical traffic among benign user-peers. Now that we are asking such networks to handle a much wider variety of data transport services, such as VoIP, we must take a step back and ask if this architecture will ever be able to provide the necessary performance, economics, or security.

Editor’s Note: To learn more about Omnivergent’s views on the future of networking, please contact Bruce Threewitt at bruce@omnivergent.com or 650 962 0463.

[9] See, for example, Vogon, a global data recovery firm, article at http://www.vogon-international.com/data%20recovery/vision-10/vision-03.htm.