Qualifying the Core of Your Network for IMS

Bill Kine, Product Manager, Spirent Communications
April 2006
P/N 79-000722 Rev.A 0406

Introduction

Communications industry pundits have been predicting the advent of converged networks and services for years. During the past decade, this concept, known as convergence, has been the focus of several infrastructure technologies, including time division multiplexing (TDM), asynchronous transfer mode (ATM), and resource reservation protocol (RSVP). Now that we are all older and wiser, we can look back at the previous attempts to deploy converged networks and learn from their follies.

There is no doubt that the technology has changed over the years. Additionally, much of the rationale behind the convergence initiatives has also evolved. Initially, voice and data services were supposed to be combined over common links between an organization's major enterprise sites for the sole purpose of circumventing long-distance charges. Since that time, toll charges (at least for domestic calls) have nearly disappeared—most cellular services offer virtually unlimited long-distance calling. Many corporate contracts now have toll charges as low as 1 cent per minute—almost free; this of course negated (or at least postponed) much of the original marketing justification for convergence.

The next area of focus became the combined implementation of voice, video, and data over the same access network. Digital subscriber line (DSL) and cable television service providers were the main proponents for this. The rallying cry over the past few years has been "triple play"—all three services over a common infrastructure. These service offerings have usually been tailored toward the edges of the network so that they can be delivered directly to the subscribers. The core technologies and supporting network infrastructure are typically taken for granted in most triple-play discussions. In fact, the core of the network is generally represented in most triple-play network diagrams as some sort of amorphous stylized cloud. That brings us to the latest incarnation of convergence technology known as the IP multimedia subsystem (IMS). This is the first approach to convergence that takes a holistic view of the entire network. The fundamental assumption of the IMS architecture, as the name implies, is that all services will be delivered over an open IP network, regardless of the underlying access or transport protocols (e.g., cable modems, DSL, fiber-to-the-x [FTTx], cellular services). This end-to-end approach is extremely useful for enabling voice (voice over IP [VoIP]), video (IPTV), and data, as well as other emerging IP–based applications. Furthermore, this takes advantage of the fact that IP is the only truly universal networking protocol—IMS does not try to thrust another backbone (e.g., ATM) upon us; instead, it builds upon the existing prevalent infrastructure.

IMS Advantages

Why should IMS succeed where the previous convergence technologies have floundered? There are several valid answers to this question, and when taken altogether, these individual reasons produce a very compelling case for IMS–based convergence. In short, IMS provides a solid solution that benefits both end users and service providers—a true win-win solution.

End-user benefits—Every discussion about the benefits associated with converged networks seems to focus on the financial aspects. These benefits are quite real. From the consumer perspective, one communications solution for voice, video, and data purchased from a single service provider should certainly be more cost-effective than subscribing to two or three networks. The cost of hardware (customer-premises equipment) should also be reduced if this unified solution is supported with a single device—and since that device will be based on the standard Internet protocol, it too should be rather inexpensive. Furthermore, for businesses and other large enterprises, shared lines, equipment, and infrastructure will generate noticeable cost savings.

The benefits inherent to IMS extend far beyond simple cost reductions. The fact that an all–IP infrastructure is used means that thousands of off-the-shelf applications will be available above and beyond just the three basic tenets of triple play. IP–based video surveillance, music downloads, "real" radio, on-line full-motion multiplayer gaming, and many others are all examples of applications that can be easily enabled in an IMS environment. In fact, integrated applications that blur the traditional lines between voice, video, and data and truly revolutionize users' communications behaviors will become possible. For example, users may be able to choose to read their voice mail and listen to their e-mail or instant messages. Or the user may be able to access remote corporate data facilities via a voice interface on a cellular phone. And as for video, in addition to watching the standard broadcasts of the local cable networks, users can now request on-demand video services from anywhere in the world, complete with the ability to interactively stop, start, or replay portions of the show as desired.

Service provider benefits—Most large service providers have already constructed their IP backbone networks. However, this state-of-the-art infrastructure tends to be highly underutilized. Capacity assessments from some of the major carriers tend to suggest that only approximately 10 to 35 percent of their IP network capacity is being used. This means that the service providers already have the ability, for the most part, to implement a new revenue-generating IMS at a minimal cost. The fact that IP will be used end to end also provides some significant benefits to the service providers. To begin with, the dynamic routing and resiliency of IP networks will be a huge advantage. Self-healing networks (especially when accompanied by the many "high-availability" features provided by modern routers) will reduce network downtime without requiring any operator intervention. IMS networks will be constructed with a consistent set of stratified protocols and standards. The same routing and signaling paradigms can be used for all types of services. Furthermore, the standardization of these common specifications such as session initiation protocol (SIP) for telephony (and possibly video) will help ensure multivendor and multi-provider interoperability. This will help guarantee that a cable TV operator, for example, can seamlessly interoperate with a wireless or DSL–based service provider. This standardization will ultimately promote the commoditization of the network elements, thus leading to additional cost savings for service providers.

An end-to-end IP network will also result in simplified provisioning for service providers. When users request a new feature or additional bandwidth, the network operators can dynamically provision the desired service in real time. No truck rolls or house calls will be necessary. Self-service Web-enabled provisioning by the end users may also be possible. This means that customers will receive prompt assistance, and the service provider can start charging for new features immediately. That leads us to bottom-line financial benefits for service providers. Today, service providers generate revenue from traditional voice and data services. Customers often split their business between providers for voice and data. Yet another vendor altogether might get the user's video (cable or satellite TV) revenue. IMS allows a single unified service provider to offer a package of all of these services for incrementally higher revenue. Furthermore, this kind of economic bundle may prevent customers from using applications such as Skype that allow them to circumvent some of the standard telephony charges.

What's the Catch?

The IMS architecture provides more services for end users and generates additional revenue at a minimal cost for service providers. In short, it's a perfect solution! However, there is one huge catch to IMS—in order for users to adopt this new technology (and pay for it), it must work at least as well as their existing systems. In other words, they must experience minimal delays in their voice conversations, and the new IMS–based video quality must be every bit as good as their current cable TV programs. Just imagine how many irate sports fans will immediately cancel their IMS service contracts and revert back to the old ways if a Super Bowl broadcast is marred with IP transmission problems!

Quality is job one—The Internet's version of quality assurance continues to be known as "best effort." This means that nothing is guaranteed. In fact, on average, the Internet somehow tends to lose approximately 2 percent of all of the data packets in transit. This rate is reasonably acceptable for typical data communications applications such as e-mail or Web surfing, since the upper-layer protocols detect the losses and request retransmissions. There is, of course, a degree of additional latency associated with this process, but no one seems to care if an e-mail is delivered in three seconds or 10. Furthermore, intermittent delays in receiving Web pages are accepted by most users as routine occurrences. However, these minor annoyances in the data world are absolutely catastrophic for real-time voice and video applications.

A 2 percent packet loss for video will significantly corrupt the quality of a video stream, to the point of becoming entirely unacceptable. The only way that IMS can be successful is if the service provider can guarantee the end-to-end quality of the connection. In fact, service providers can probably use their quality of service (QoS) as a competitive differentiator in order to attract and retain lucrative IMS customers. IMS and triple play are not just edge technologies; they are entirely dependent upon the core of the network as well as the edge.

The term "quality" is an all-encompassing expression covering several critical facets. Most users and service providers interpret the term to mean actual IP–based QoS—this is probably more appropriately known as type of service (ToS) or differentiated services (DiffServ), both of which are popular IP traffic classification paradigms. Another significant aspect of quality is the resiliency and recovery characteristics of the network. And a third and equally important attribute of quality is the actual implementation of the network protocols themselves.

The good news for service providers is that the IMS architecture provides a solid foundation from which they can address all of the quality issues. An end-to-end IP–based solution enables the service provider to take advantage of a consistent set of industry hardened tools for network protocols, QoS, and resiliency.

Look Before You Leap

Users' perception of the quality of IMS services is what really matters. Unfortunately, "perception" is highly subjective and varies greatly from person to person. Therefore, the service provider must strive for the best possible service to satisfy the greatest quantity of customers. Or alternatively, the service provider may be able to offer several tiers of services at separate price points. Either way, the resulting service level(s) must also remain consistent, regardless of users' quantities, applications, or locations. These service levels must be fully tested in a lab environment—under worst-case conditions—prior to deploying IMS in a service provider's production network. A service provider's likelihood of successfully deploying IMS is directly related to the level of testing that it accomplishes in the lab.

Previous triple play and IMS testing has always focused on the edge of the network. After all, the last mile has always been perceived to be the bottleneck of a carrier's bandwidth. While this certainly is extremely important, edge testing overlooks the vast majority of the network. For example, a fiber-to-the-home (FTTH) solution for network access is wonderful, but if the core is congested, the user will still be subjected to bandwidth constraints. This in turn impacts the quality of the user's voice, video, and data applications. Furthermore, if all of the users at the edge of a network are viewing different full-motion video programs (and that's one of the primary marketing promises of IMS), imagine the resulting bandwidth requirements in the core of the network.

IMS provides a consistent end-to-end architecture designed to facilitate advanced end-to-end user services, so it must be fully tested in an end-to-end configuration.

IMS Testing

QoS—The first objective for quality testing, as the name suggests, is directly related to the end-to-end QoS that the user experiences. This, in essence, verifies that there is sufficient bandwidth throughout the network, under normal operating conditions, as well as during atypical, severely strained situations, to fully support all of the users' applications. Carriers just are beginning to discover that bandwidth is no longer infinite. Network usage has increased considerably over the past few years, while network growth and expansion has been quite minimal. Furthermore, some emerging IMS applications such as IPTV or video on demand require substantial amounts of bandwidth; this will rapidly cut into the surplus capacity left over from the massive network build-outs of 2000 and 2001. Service providers can address these requirements by either dramatically increasing their core capacity (the "brute force" method) or by allocating their existing bandwidth more judiciously. The latter option is probably the more sensible solution. Altering the current bandwidth allocations implies that the service providers will need to deploy some degree of IP–based QoS and traffic engineering in their networks. This is much easier said than done. The service providers will first need to carefully analyze the impact of QoS on their overall network scalability, performance, and resilience. They will also need to identify their own unique criteria for QoS—will they have tiered services for users (at stratified price points), or will they implement service levels based upon the needs of specific applications (for example, always giving time-sensitive video traffic preference over routine e-mail packets)? Routers and switches all have their own output policing and queuing mechanisms. Each equipment manufacturer individually develops solutions that best fit its particular products and markets. All products have different queue depths, latency, and policy rules. This means that the effects of QoS will vary from product to product; it may even be detrimental to the performance of some devices. Therefore a service provider must fully test, characterize, and optimize its proposed QoS implementation in a lab prior to deploying this in a production network. QoS testing involves generating traffic that will be classified at various QoS levels and validating that the device under test (DUT) can process the traffic correctly. Throughput, loss, and latency will be measured for each traffic stream. Links should also be oversubscribed in order to determine the device's response to this likely situation. If a device drops or changes the priority of (i.e., "re-colors") the traffic that violates the configured policy specifications, this process may impact the users' applications, so it also should be fully characterized. QoS traffic should be tested running over a single device, as well as multiple devices (especially if different vendors are used) throughout the simulated network in the lab. It is unlikely that a service provider will treat all users or all applications equally. Therefore the tester must be able to simulate, track, and measure performance data for individual traffic streams under all kinds of conditions. The inputs for the test should be mixed data types and QoS values. The

DUT will then need to classify, prioritize, and police the data streams. The output should be normalized distributions based on the established QoS contracts. The test results should indicate the throughput, loss, and latency for each stream and traffic class.

Network resiliency and recovery—Network outages happen. No network is immune to them. The larger the network, the greater is the likelihood that some component or other will fail. Lines, power supplies, software, and equipment all break down from time to time. Additionally, human errors such as basic configuration mistakes can also lead to partial outages or suboptimal performance. However, if these network anomalies can be hidden from the end user, the overall service level will continue to be perceived to be at a consistently high level. Self-healing networks and protocols are one of the keys to concealing failures from the users. These internal network corrections are designed to minimize downtime. It is important to note that they cannot prevent downtime altogether, but they reduce it—note that every network operator strives for 99.999 percent uptime; none of them are ever bold enough to even mention the possibility of 100 percent network availability! It is quite apparent that user satisfaction is inversely proportional to the amount of network downtime. In other words, the less downtime, the happier the users will be. Therefore a service provider should model, test, and tune the various configurations, protocols, and recovery mechanisms in the lab to determine the optimal solution for their particular network. For example, the BGP graceful restart protocol is one tool that can be used to reduce downtime incurred due to a router or component failure. In one particular case, the system's recovery time from a router failure with normal BGP was 76 seconds. However, when graceful restart was enabled, the recovery time was dramatically reduced to only 18 seconds—and additional protocol and architectural improvements might be able to reduce this even further. Other types of network and component failures should also be induced and measured so that the service provider can develop the appropriate backup and recovery plans without significantly impacting their users.

Protocol performance tests—IMS services will need to be extremely responsive to network changes and user requests. This level of responsiveness extends far beyond just using SIP for call or video signaling and registration. For example, IPTV viewers may want to surf through all of the available channels just as they would with cable or satellite television. With traditional TV, this is a simple process since broadband cable TV or satellite links (or even good old-fashioned television airwaves) carry all of the programs simultaneously and users simply focus on the specific spectrum of data that interests them. However, bandwidth limitations make it impossible to deliver hundreds of concurrent IP–based TV channels to every set, so instead an interface to the network must request the desired channels one at a time. The basic operational mechanics associated with this process can result in a lot of overhead operations and network reconfiguration for active channel surfers. High-bandwidth IP applications such as television that need to be delivered to many simultaneous destinations require advanced multicast protocol support. The other alternatives (unicast or broadcast traffic) are not practical or scalable for large quantities of subscribers. Therefore the service providers will need to implement traditional multicast routing protocols—protocol-independent multicast (PIM) and Internet group management protocol (IGMP). These protocols provide solid foundations for IMS services, but they will also require a considerable amount of tuning and optimization in order to effectively satisfy the demands of millions of television viewers.

Most multicast networks use PIM sparse mode (PIM–SM). This protocol is ideally suited for large Internet environments. It is not particularly chatty, and it is designed to replicate the data as close to the end users as possible, thereby reducing the load on the core of the network. PIM–SM is an on-demand or a "pull" protocol, meaning that users must first request information, and then they will be added to a multicast group. Therefore, if an IPTV channel is showing a major sporting event, users can join that particular multicast group from anywhere in the network. Once they join, they will then receive the appropriate information content. If the user switches to another IPTV channel, the sporting event stream will be discontinued and a new video program from a different multicast group will be delivered to their television set.

The operational aspects of multicast delivery mechanisms can be quite complex. The process starts with a request from the user's network interface module; for television sets, typically this would be known as some sort of set-top box. This device will generate an IGMP "join" message, indicating that the user wants to receive a particular multicast transmission. This "join" request (also known as an IGMP "report" message) is then forwarded to the local router. The router must update its state tables to indicate which users are watching which channels. If the router is already receiving information from the multicast group associated with the requested channel, it can simply add the new user to the group. However, if that channel is not currently present on the local router, it must in turn request the information from an upstream source; this is accomplished by issuing a PIM–SM "join" request (not to be confused with the IGMP "join" initiated by the end user). When an upstream router (meaning a device nearer to the source of the multicast transmission) receives a PIM–SM join request, it will forward the requested traffic stream to the new member of the group. This, of course, will use additional bandwidth on the link between the devices. Furthermore, there may be a bit of latency associated with this entire process as well. As this latency increases, it will become noticeable to the end users. The general rule of thumb seems to suggest that impatient television channel surfers will only tolerate a two-second delay between the time they push a button on their remote controls and when they actually receive the requested content. With this in mind, the service providers must carefully measure, track, and constantly optimize their multicast network performance. Unfortunately, many service providers choose to focus exclusively on the latency associated with the IGMP portion of the channel-switching process. This only takes place between the set-top box and the first router, so it only represents a fraction of the delay that user actually experiences. Instead, latency must be measured for the entire process across multiple devices. And it should be measured under congested or other adverse conditions. IMS services will require the support of the entire network, and the users will judge the carriers based upon their end-to-end responsiveness, so all of these variables must be included in any bandwidth or latency calculations.

Scalability also figures heavily into the overall performance of a multicast network. Up until now, service providers have feared the proliferation of thousands of multicast groups, each with just a few subscribers, such as for videoconferences or distance-learning applications. This kind of a realistic scenario could very quickly inundate the state tables of the routers and other network elements. Any related scalability problems will impact the traditional data network users in all environments, including IMS. However, the IPTV component of IMS changes the scalability dynamics altogether. In a television scenario, there will be a limited quantity of groups and channels (perhaps a couple hundred groups at the most), but each will have a multitude of subscribers. This scenario could also overload the state tables of the routers, although all of the weight will now be on a different axis.

Together these two issues, representing both the data and video components of IMS, will foster some major multicast scalability challenges for service providers. The good news, however, is that it is easy to model these scenarios, with live data, in a lab environment prior to rolling out multicast services in an actual production network.

In addition to multicast protocol performance and functionality and scalability tests, service providers should also verify their traditional protocols. All of the basic routing for data and VoIP packets will still need to take place concurrently with the multicast, QoS, and other functions necessary to support a full suite of IMS applications. Therefore, all of the traditional BGP or MPLS testing and modeling, for example, are still necessary in an IMS environment. In fact, this testing should be done in conjunction with all of the other real-world IMS control protocols and applications.

Tying It All Together

The IMS architecture adds a lot of power and functionality to service providers' IP networks. Many new value-added applications can be offered to their users, all of which will generate incremental revenue. IMS allows service providers to grow beyond the boundaries of their traditional voice and data services. Video gets added to their portfolio, as well as many new multimedia applications. For example, a user might be able to use IP to transmit photos from a cellular camera phone while carrying on a simultaneous voice conversation. More complex applications could include full-motion videoconferencing and gaming or other on-demand video services. These new services do not require the wholesale replacement of carriers' networks. Instead, they build on the solid IP foundation that has already been constructed. However, due to the high bandwidth requirements and the massive quantities of users, this building must take place in a carefully planned and controlled manner. IMS applications, solutions, and infrastructure components will test the limits of existing network elements and designs. Therefore, all possible permutations must be thoroughly modeled and evaluated before being deployed in production networks. Users will experience IMS applications that may originate anywhere on the network. In fact, the location of these applications should be entirely transparent to the users. The information will likely traverse the core of the network, as well as the edges. The service providers will be judged by their users (who vote with their money) based on the end-to-end perceived quality of these services. Therefore, the service providers must expand their visions of IMS—it represents a total solution, not just an edge technology. The success or failure of service providers' IMS implementations will depend upon the quality of the whole network, so all IMS testing must also focus on the entire end-to-end infrastructure.

Content Provided by: Bill Kine, Product Manager, Spirent Communications

About the Author

Bill Kine has two decades of experience in the communications industry. He has designed and implemented networks for small, medium, and large enterprises, as well as service providers. An expert in advanced communications topics, Mr. Kine has written several articles for industry journals and has conducted many classes and seminars. In his current role as Spirent Communications' product marketing manager, Mr. Kine works closely with service providers and network equipment manufacturers to help them with their complex network testing.

Inspired Innovation

Spirent Communications, Inc.
15200 Omega Drive
Rockville, Maryland 20850, USA
Spirent Communications
26750 Agoura Road
Calabasas, California 91302, USA

E-mail: sales-spirent@spirentcom.com
Web: http://www.spirentcom.com

Americas

+1 800.SPIRENT
+1 818.676.2683

Europe, Middle East, Africa

T: +33 1 6137.2250

Asia Pacific

T: +852 2511.3822

Copyright