The
# Internet Protocol Journal

March 2001 Volume 4, Number 1

*A Quarterly Technical Publication for Internet and Intranet Professionals*

## In This Issue

You can download IPJ
back issues and find
subscription information at:
**www.cisco.com/ipj**

## From The Editor

The rapid growth of the Internet has led to numerous changes to the underlying technologies. In the early days, host names and their corresponding IP addresses were kept in a flat text file (“**HOSTS.TXT**”), updated weekly by the Network Information Center at SRI International. In the mid 1980s it became clear that this method of name/address mapping would not scale, and a new distributed lookup mechanism was designed and deployed. This new method, known as the *Domain Name System* (DNS), has proven successful even in the face of millions of Internet hosts.

Another result of Internet growth is the potential for depletion of the IP Version 4 (IPv4) 32-bit address space. In the early 1990s, this became a matter of great focus for the Internet Engineering Task Force (IETF). The “short-term” fix for this problem was to abandon the original concept of A, B and C address classes and introduce *Classless Interdomain Routing* (CIDR), which consumes addresses in a much more efficient manner—that is to say, more slowly. Address consumption has also been slowed by the use of *Network Address Translation* (NAT) and private address space. Predictions for when the Internet will finally run out of IPv4 addresses varies. The long-term solution is to replace IPv4 with IPv6 which uses 128 bits for addressing.

One area of Internet growth that is currently causing some concern among ISPs is the growing size of the routing table that each router participating in the *Border Gateway Protocol* (BGP) must keep in memory. Our first article, by Geoff Huston, is a detailed look at this problem. Geoff takes an historical look at the BGP routing table, and discusses ways to address some of the issues.

In our March 2000 issue, Geoff Huston wrote an article entitled “Quality of Service—Fact or Fiction?” that discussed the prospects for achieving QoS on an Internet-wide scale. In this issue, Bill Stallings looks at QoS in the LAN environment, which is generally easier to control than the Internet as a whole. LAN QoS has been standardized in IEEE 802.1D which is the subject of this article.

We apologize for the delay in getting our online subscription system up and running. It should be available in the very near future. Meanwhile, please continue to use **ipj@cisco.com** for any subscription questions or to give feedback on anything you read in this journal.

*—Ole J. Jacobsen, Editor and Publisher*
**ole@cisco.com**

# Analyzing the Internet BGP Routing Table

*by Geoff Huston, Telstra*

The Internet continues along a path of seemingly inexorable growth, at a rate that has, at a minimum, doubled in size each year. How big it needs to be to meet future demands remains an area of somewhat vague speculation. Of more direct interest is the question of whether the basic elements of the Internet can be extended to meet such levels of future demand, whatever they may be. To rephrase this question, are there inherent limitations in the technology of the Internet—or its architecture of deployment—that may impact the continued growth of the Internet to meet ever-expanding levels of demand?

Numerous potential areas can be searched for such limitations, including the capacity of transmission systems, the switching capacity of routers, the continued availability of addresses, and the capability of the routing system to produce a stable view of the overall topology of the network. This article examines the Internet routing system and the longer-term growth trends that are visible within this system.

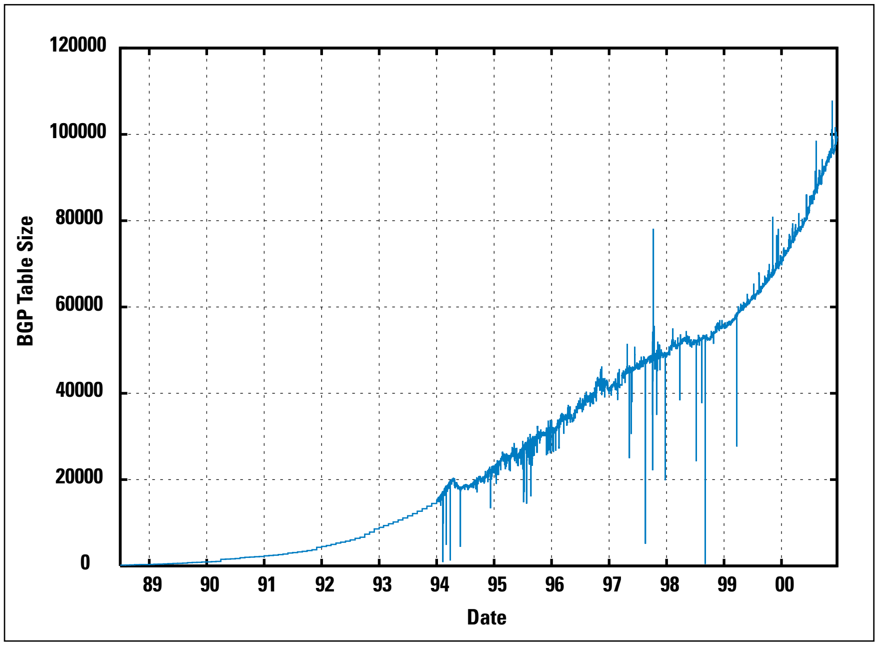
The structure of the global Internet can be likened to a loose coalition of semi-autonomous constituent networks. Each of these networks operates with its own policies, prices, services, and customers. Each network makes independent decisions about where and how to secure the supply of various components that are needed to create the network service. The cement that binds these networks into a cohesive whole is the use of a common address space and a common view of routing. Integrity of routing within each constituent network, or *Autonomous System* (AS), is maintained through the use of an interior routing protocol (or *Interior Gateway Protocol,* or IGP). The collection of these networks is joined into one large routing domain through the use of an inter-network routing protocol (or *Exterior Gateway Protocol,* or EGP).

When the scaling properties of the Internet were studied in the early 1990s, two critical factors identified in the study were, not surprisingly, routing and addressing[1]. As more devices connect to the Internet, they consume addresses, and the associated function of maintaining reachability information for these addresses implies ever-larger routing tables. The work in studying the limitations of the 32-bit IPv4 address space produced many outcomes, including the specification of IPv6, as well as the refinement of techniques of *Network Address Translation* (NAT) intended to allow some degree of transparent interaction between two networks using different address realms. Growth in the routing system is not directly addressed by these approaches, because the routing space is the cross product of the complexity of the topology of the network, multiplied by the number of autonomous domains of connectivity policy multiplied by the base size of a routing-table entry. When a network advertises a block of addresses into the exterior routing space, this entry is generally carried across the entire exterior routing domain of the

Internet. To measure the characteristics of the global routing table, it is necessary to establish a point in the default-free part of the exterior routing domain and examine the *Border Gateway Protocol* (BGP) routing table that is visible at that point.

Measurements of the size of the routing table were somewhat sporadic in the beginning, and many measurements were taken at approximately monthly intervals from 1988 until 1992 at Merit[2]. This effort was resumed in 1994 by Erik-Jan Bos at Surfnet in the Netherlands, who commenced measuring the size of the BGP table at hourly intervals at the start of that year. This measurement technique was adopted by the author in 1997, using a measurement point located at the edge of AS 1221 in Australia, again using an hourly interval for the measurement[6]. The result of these efforts is that we now have a detailed view of the dynamics of the Internet routing-table growth that spans 13 years (Figure 1).

*Figure 1: BGP Table Growth 1988–2000*



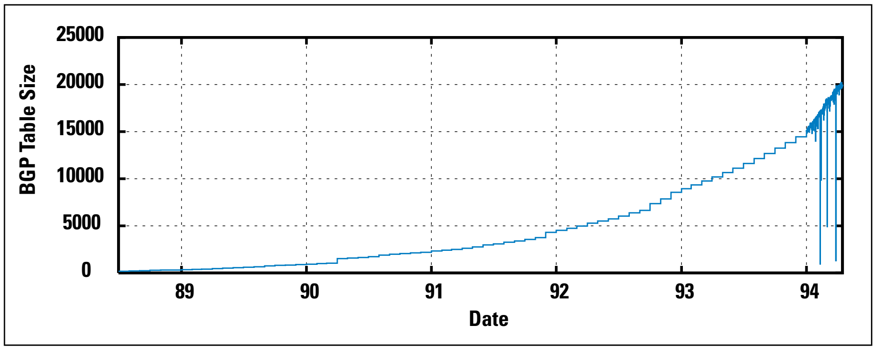
### BGP Table Growth

At a gross level, there appear to be four distinct phases of growth visible in this data.

### Pre-CIDR Growth

The initial characteristics of the routing-table size from 1988 until April 1994 show definite characteristics of exponential growth (Figure 2). Much of this growth can be attributed to the growth in deployment of the historical Class C address space (/24 address prefixes). Unchecked, this growth would have lead to saturation of the BGP routing tables in nondefault routers within a few years. Estimates of the time at which this would have happened vary somewhat, but the overall observation was that the growth rates were exceeding the growth in hardware and software capability of the deployed network at that time.

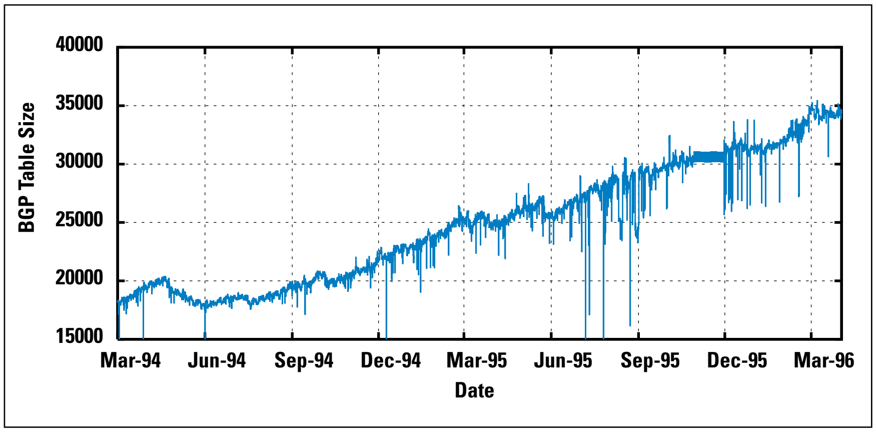
*Figure 2: BGP Table Growth 1988–1994*



## CIDR Deployment

The response from the engineering community was the introduction of routing software that dispensed with the requirement for the Class A, B, and C address delineation, replacing this scheme with a routing system that carried an address prefix and an associated prefix length. A concerted effort was undertaken in 1994 and 1995 to deploy *Classless Interdomain Routing* (CIDR), based on encouraging deployment of the CIDR-capable version of the BGP protocol, BGP4. The effects of this effort are visible in the routing table (Figure 3). Interestingly enough, the efforts of the *Internet Engineering Task Force* (IETF) CIDR Deployment Working Group are visible in the table, with downward movements in the size of the routing table following each IETF meeting.

*Figure 3: BGP Table Growth 1994–1995*



The intention of CIDR was one of supporting an address architecture termed "provider address aggregation," where a network provider is allocated an address block from the address registry, and announces this entire block into the exterior routing domain. Customers of the provider use a suballocation from this address block, and these smaller routing elements are aggregated by the provider and not directly passed into the exterior routing domain. During 1994, the size of the routing table remained relatively constant at approximately 20,000 entries as the growth in the number of providers announcing address blocks was matched by a corresponding reduction in the number of address announcements as a result of CIDR aggregation.

### CIDR Growth

For the next four years until the start of 1998, CIDR proved remarkably effective in damping unconstrained growth in the BGP routing table. While other metrics of Internet size grew exponentially during this period, the BGP table grew at a linear rate, adding about 10,000 entries per year. (Figure 4). Growth in 1997 and 1998 was even lower than this linear rate. Although the reasons behind this are somewhat speculative, it is relevant to note that this period saw intense aggregation within the *Internet Service Provider* (ISP) industry, and in many cases this aggregation was accompanied by large-scale renumbering to fit within provider-based aggregated address blocks. During this period, credit for this trend also must be given to Tony Bates, whose weekly reports of the state of the BGP address table, including listings of further potential for route aggregation, provided considerable incentive to many providers to improve their levels of route aggregation[4].

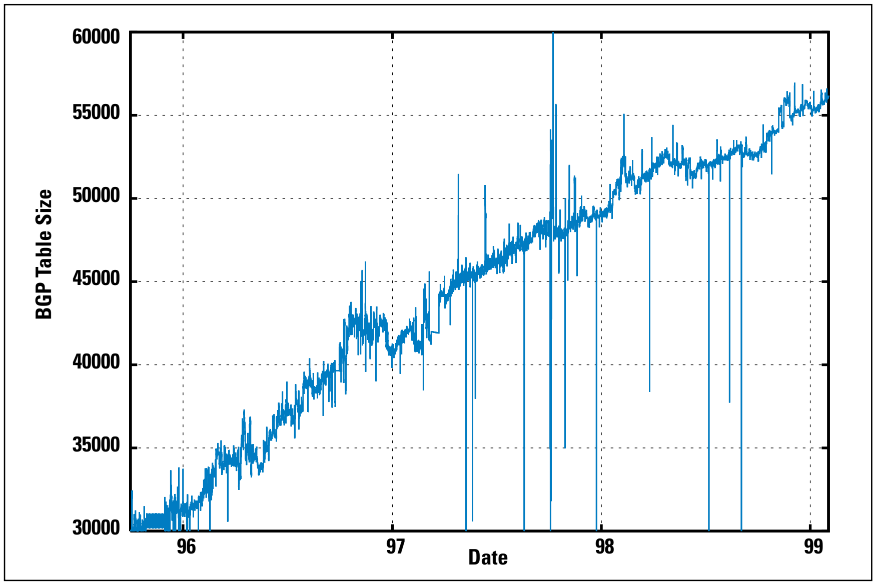


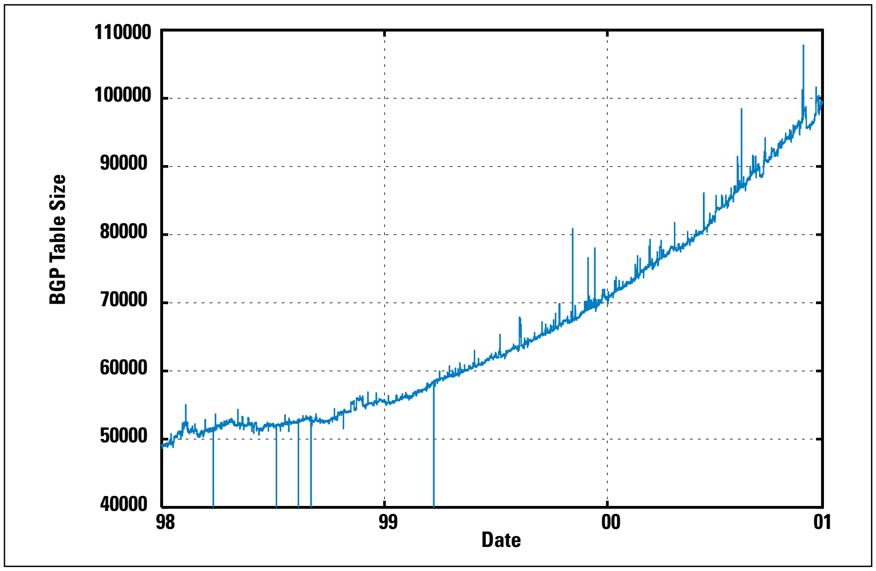
*Figure 4: BGP Table Growth 1995–1998*

A close examination of the table reveals a greater level of stability in the routing system at this time. The short-term (hourly) variation in the number of announced routes decreased, both as a percentage of the number of announced routes and in absolute terms. One of the other benefits of using large aggregate address blocks is that an instability at the edge of the network is not immediately propagated into the routing core. The instability at the last hop is absorbed at the point at which an aggregate route is used in place of a collection of more specific routes. This, coupled with widespread adoption of BGP route flap damping, has been every effective in reducing the short-term instability in the routing space. It has been observed that whereas the absolute size of the BGP routing table is one factor in scaling, another is the processing load imposed by continually updating the routing table in response to individual route withdrawals and announcements. The encouraging picture from this table is that the levels of such dynamic instability in the network have been reduced considerably by a combination of route flap damping and CIDR.

### Current Growth

In late 1998, the trend of growth in the BGP table size changed radically, and the growth for the past two years is again showing all the signs of a reestablishment of exponential growth. It appears that CIDR has been unable to keep pace with the levels of growth of the Internet. (Figure 5). Once again the concern is that this level of growth, if sustained, will outstrip the capability of hardware, or current capability of the BGP routing protocol, or possibly both.

*Figure 5: BGP Table Growth 1998–2000*



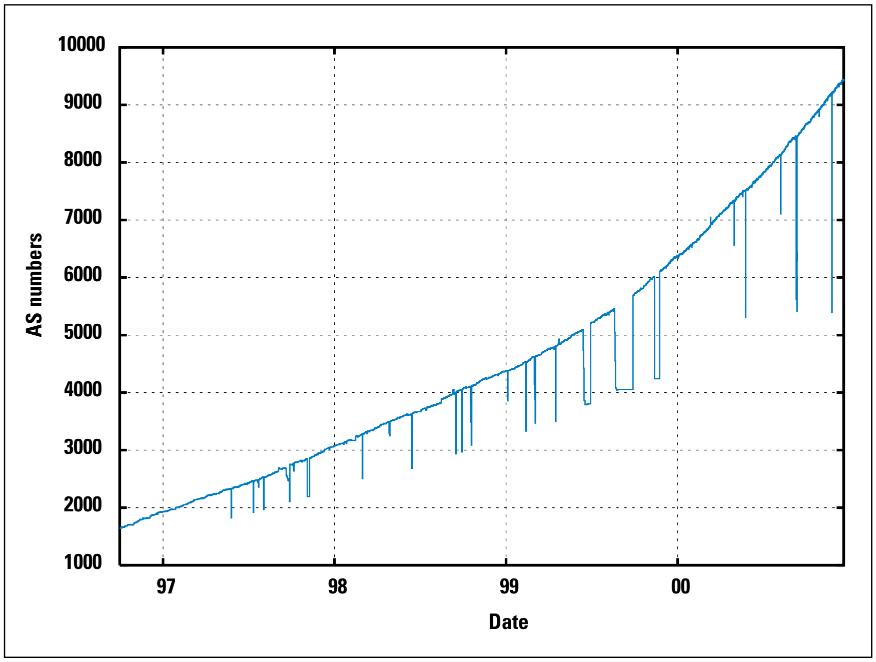
### Related Measurements Derived from BGP Table

The level of analysis of the BGP routing table has been extended in an effort to identify the reasons for this resumption of exponential growth. Current analysis includes measuring the number of ASs in the routing system, and the number of distinct AS paths, the range of addresses spanned by the table, and the average span of each routing entry.

### AS Number Consumption

Each network that is multihomed within the topology of the Internet and wishes to express a distinct external routing policy must use an AS to associate its advertised addresses with such a policy. In general, each network is associated with a single AS, and the number of ASs in the default-free routing table tracks the number of entities that have unique routing policies. There are some exceptions to this, including large global transit providers with varying regional policies, where multiple ASs are associated with a single network, but such exceptions are relatively uncommon. The trend of AS number deployment over the past four years is also exponential (Figure 6). The growth in the number of ASs can be correlated with the growth in the amount of address space spanned by the BGP routing table. At the end of 2000, the span of advertised addresses is growing at an annual rate of 7 percent, while the number of ASs is growing by 51 percent. Each AS is, on average advertising smaller address ranges. This points to increasingly finer levels of routing detail being announced into the global routing domain, a trend that causes some level of concern.

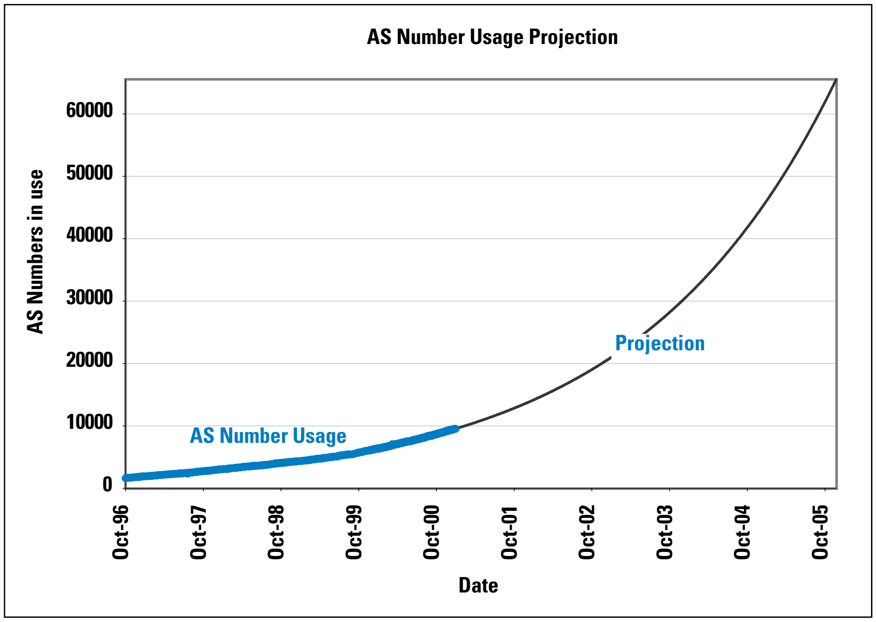
*Figure 6: AS Number Deployment*



This is a likely result of an increasingly dense interconnection mesh, where an increasing number of networks are moving from a single-homed connection into multihoming and peering. The spur for this may well be the declining unit costs of communications bearer services.

If this rate of growth continues, the 16-bit AS number set will be exhausted by late 2005 (Figure 7). Work is under way within the IETF to modify the BGP protocol to carry AS numbers in a 32-bit field[5]. Although the protocol modifications are relatively straightforward, the major responsibility rests with the operations community to devise a transition plan that will allow gradual transition into this larger AS number space.

*Figure 7: AS Number Projections*

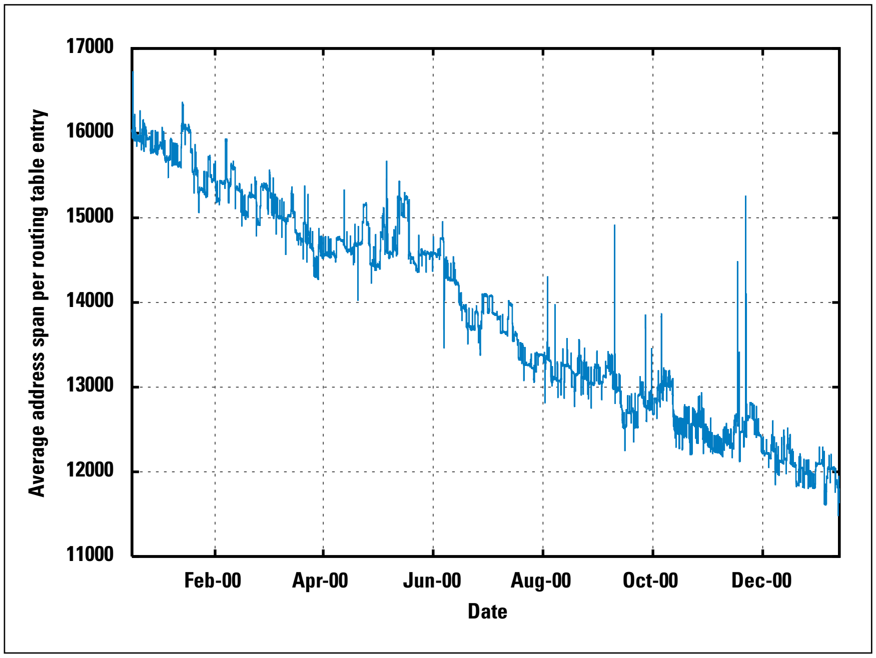


### Average Prefix Length of Advertisements

The intent of CIDR aggregation was to support the use of large aggregate address announcements in the BGP routing table. To check whether this is still the case, researchers have tracked the average span of each BGP announcement for the past 12 months. The data indicates a decline in the average span of a BGP advertisement from 16,000 individual addresses in November 1999 to 12,100 in December 2000 (Figure 8). This corresponds to an increase in the average prefix length from /18.03 to /18.44. Separate observations of the average prefix length used to route traffic in operation networks in late 2000 indicate an average length of 18.1[8]. Again, this trend is cause for concern because it implies the increasing spread of traffic over greater numbers of increasingly finer forwarding-table entries. This, in turn, has implications for the design of high-speed core routers, particularly when extensive use is made of cached forwarding entries within the switching subsystem.

One potential scenario is that the size of the advertisement continues to decrease. With the widespread use of address translation gateway systems, such as NAT, and the continued concern over the finite nature of the IPv4 address pool, this is certainly a highly likely scenario. Projections of the average prefix length of advertisements using current trends in the number of BGP table entries and the total address span advertised in the BGP table indicate a lengthening of the average prefix length of advertisements by 1 bit length every 29 months. This has implications in the lookup algorithms used in routing design, depending on the space/time trade-offs used in the lookup algorithm design. This trend implies that either lookups need to search deeper through the prefix chain to find the necessary forwarding entry, requiring faster memory subsystems to perform each lookup, or the lookup table needs to be both larger and more sparsely populated, increasing the requirements for high-speed memory within the router forwarding subsystem.

*Figure 8: Average Span of BGP Advertisement*



### Prefix Length Distribution

In addition to looking at the average prefix length, the analysis of the BGP table also includes an examination of the number of advertisements of each prefix length.

An extensive effort was introduced in the mid-1990s to move away from extensive use of the Class C space and to encourage providers to advertise larger address blocks. This has been reinforced by the address registries who have used provider allocation blocks of /19 and, more recently, /20. These measures were introduced when there were approximately 20,000 to 30,000 entries in the BGP table. It is interesting to note that five years later, of the 96,000 entries in the routing table, about 53,000 entries have a /24 prefix. In absolute terms, the /24 prefix set is the fastest-growing prefix set in the entire BGP table.

The routing entries of these smaller address blocks also show a much higher level of change on an hourly basis. Although a large number of BGP routing points perform route flap damping, there is still a very high level of announcements and withdrawals of these entries in this particular area of the routing table when viewed using a perspective of route updates per prefix length. Given that the number of these small prefixes is growing rapidly, there is cause for some concern that the total level of BGP flux, in terms of the number of announcements and withdrawals per second, may be increasing, despite the pressures from flap damping. This concern is coupled with the observation that, in terms of BGP stability under scaling pressure, it is not the absolute size of the BGP table that is of prime importance, but the rate of dynamic path recomputations that occur in the wake of announcements and withdrawals. Withdrawals are of particular concern because of the number of transient intermediate states that the BGP distance-vector algorithm explores in processing a withdrawal. Current experimental observations indicate a typical convergence time of about 2 minutes to propagate a route withdrawal across the BGP domain[7]. An increase in the density of the BGP mesh, coupled with an increase in the rate of such dynamic changes, does have serious implications in maintaining the overall stability of the BGP system as it continues to grow.

The registry allocation policies also have had some impact on the routing-table prefix distribution. The original registry practice was to use a minimum allocation unit of a /19, and the 10,000 prefix entries in the /17 to /19 range are a consequence of this policy decision. More recently, the allocation policy now allows for a minimum allocation unit of a /20 prefix, and the /20 prefix is used by about 4000 entries; in relative terms, this is one of the fastest-growing prefix sets.

The number of entries corresponding to very small address blocks (smaller than a /24), although small in number as a proportion of the total BGP routing table, is the fastest growing in relative terms. The number of /25 through /32 prefixes in the routing table is growing faster, in terms of percentage change, than any other area of the routing table. If prefix length filtering were in widespread use, the practice of announcing a very small address block with a distinct routing policy would have no particular beneficial outcome, because the address block would not be passed throughout the global BGP routing domain and the propagation of the associated policy would be limited in scope. The growth of the number of these small address blocks, and the diversity of AS paths associated with these routing entries, points to a relatively limited use of prefix-length filtering in today's Internet. In the absence of any corrective pressure in the form of widespread adoption of prefix-length filtering, the very rapid growth of global announcement of very small address blocks is likely to continue.

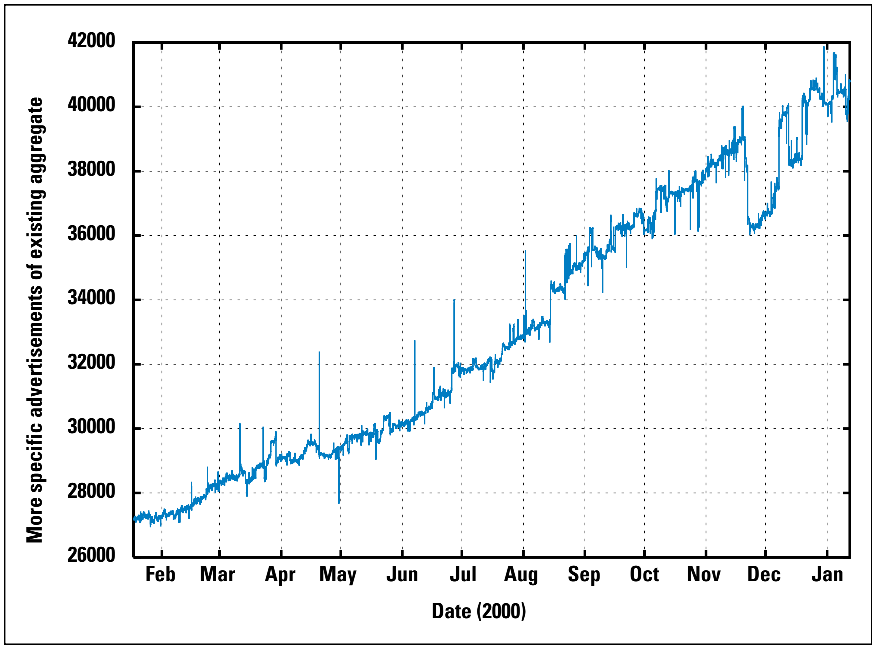
### Aggregation and Holes

With the CIDR routing structure, it is possible to advertise a more specific prefix of an existing aggregate. The purpose of this more specific announcement is to punch a "hole" in the policy of the larger aggregate announcement, creating a different policy for the specifically referenced address prefix. Another use of this mechanism is not to promulgate a different connectivity policy, but to perform some rudimentary form of load balancing and mutual backup for multihomed networks. In this model, a network may advertise the same aggregate advertisement along each connection, but then advertise a set of specific advertisements for each connection, altering the specific advertisements such that the load on each connection is approximately balanced. The two forms of holes can be readily discerned in the routing table—while the approach of policy differentiation uses an AS path that is different from the aggregate advertisement, the load balancing and mutual backup configuration uses the same AS path for both the aggregate and the specific advertisements.

Although it is difficult to understand whether the use of such specific advertisements was intended to be an exception to a more general rule or that it was not intended to be within the original intent of CIDR deployment, there appears to be very widespread use of this mechanism within the routing table. Approximately 37,500 advertisements, or 37 percent of the routing table, is being used to punch policy holes in existing aggregate announcements (Figure 9). Of these, the overall majority of about 30,000 routes use distinct AS paths, so that once more we are seeing a consequence of finer levels of granularity of connection policy in a densely interconnected space.

Although long-term data is not available for the relative level of such advertisements as a proportion of the full routing table, the growth level does strongly indicate that policy differentiation at a fine level within existing provider aggregates is a significant driver of overall table growth.

*Figure 9: More Specific Advertisements*

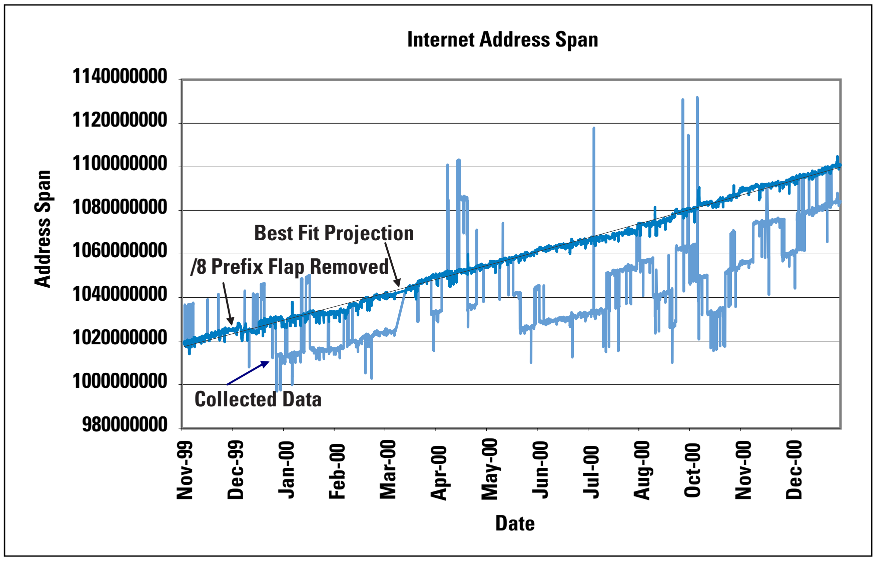


## Address Consumption

A decade ago there were two major concerns over scaling of the Internet, and of the two, the consumption of address space was considered to be the more immediate and compelling threat to the continued viability of the network to sustain growth.

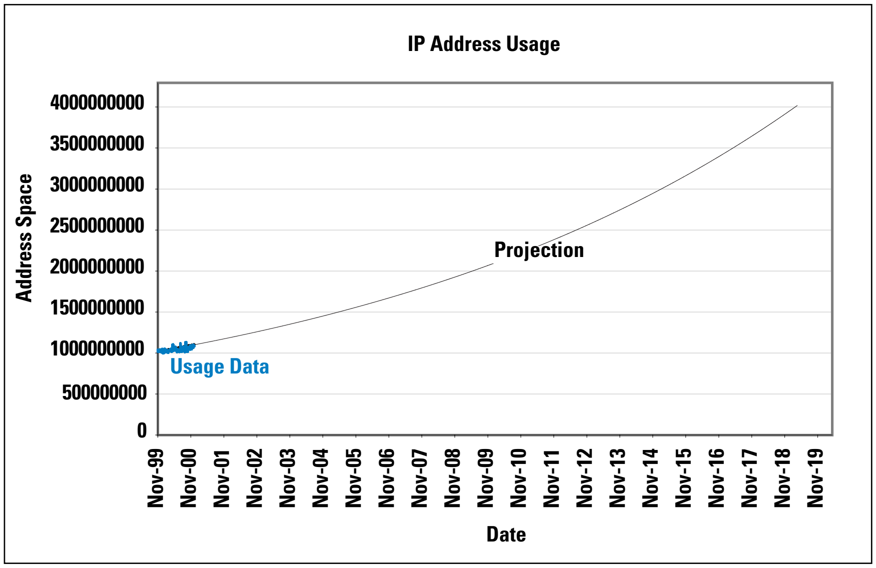
Within the scope of this exercise, it has been possible to track the total span of address space covered by BGP routing advertisements. Over the period from November 1999 until December 2000, the span of address space has grown from 1.02 billion addresses to 1.06 billion. However, numerous /8 prefixes are periodically announced and withdrawn from the BGP table, and if the effects of these prefixes are removed, the final value of addresses spanned by the table is approximately 1.09 billion addresses (Figure 10).

*Figure 10: Total Address Space*



This is an annual growth rate of a little less than 7 percent, and at that rate of address deployment, the IP Version 4 address space will be able to support another 19 years of such growth (Figure 11). Compared to the 42-percent growth in the number of routing advertisements, it would appear that much of the growth of the Internet in terms of growth in the number of connected devices is occurring behind various forms of NATs. In terms of solving the perceived finite nature of the address space identified just under a decade ago, the Internet appears so far to have embraced the approach of using NATs, irrespective of their various perceived functional shortcomings[3]. This observation also supports the observed increase of smaller address fragments supporting distinct policies in the BGP table, because such small address blocks encompass arbitrarily large networks located behind one or more NAT gateways.

*Figure 11: Address Space Projection*



## Anomalies

A common space such as the inter-provider domain is not actively managed by any single entity, and various anomalies appear in the routing table from time to time.

One notable event occurred in late 1997, when some large prefixes were deconstructed into a massive set of /24 prefixes and this set was inadvertently passed into the inter-provider BGP domain. The BGP table graphs show a sudden upswing in the number of routing table entries from 50,000 entries to about 78,000 entries. It could have been higher, except that a commonly used routing hardware platform at the time ran into table memory exhaustion at that number of table entries, and further promulgation of additional routing entries ceased. Numerous other anomalies also exist in the table, including the presence of a /31 prefix and several hundred /32 prefixes.

Although many of these anomalies can be attributed to configuration errors of various forms, the underlying observation is that there are no universally used strong filters on what can broadcast into the BGP routing space. Considering the distributed nature of this table and the critical role that it plays in supporting the global Internet, this can be considered a significant current vulnerability. One potential response is to make more use of authentication measures. A validity check could be a precondition to accepting any route advertisement, allowing the receiver of the advertisement a means to check that the origin AS intended to advertise this route. This would create greater resiliency against inadvertent leaks of large sets of advertisements into the broader interdomain space. It would also improve the resiliency of the BGP domain against some forms of deliberate attack.

### Conclusions

There are strong parallels between the BGP routing space and the condition commonly referred to as "The Tragedy Of The Commons." The BGP routing space is simultaneously everyone's problem, because it impacts the stability and viability of the entire Internet, and no one's problem, in that no single entity can be considered to manage this common resource.

In other common resource domains, when the value of the resource is placed under threat because of damaging exploitative practices, the most typical form of corrective action is through the imposition of a consistent set of policies and practices intended to achieve a particular outcome. The vehicle for such an imposition of policies and practices is most commonly that of regulatory fiat. In a globally distributed space such as the BGP table, it is a challenging task to identify the source and authority of such potential regulatory activity.

### Multihomed Small Networks

It would appear that one of the major drivers of the recent growth of the BGP table is that of small networks multihoming with numerous peers and numerous upstream providers. In the appropriate environment where numerous networks are in relatively close proximity, using peer relationships can reduce total connectivity costs, as compared to using a single upstream service provider. Equally significantly, multihoming with numerous upstream providers is seen as a means of improving the overall availability of the service. In essence, multihoming is seen as an acceptable substitute for upstream service resiliency.

This has a potential side effect: When multihoming is seen as a preferable substitute for upstream provider resiliency, the upstream provider cannot command a price premium for proving resiliency as an attribute of the provided service, and, therefore, has little incentive to spend the additional money required to engineer resiliency into the network. The actions of the multihomed network clients then become self-fulfilling.

One way to characterize this behavior is that service resiliency in the Internet is becoming the responsibility of the customer, not the service provider.

In such an environment resiliency still exists, but rather than being a function of the bearer or switching subsystem, resiliency is provided through the function of the BGP routing system. The question is not whether this is feasible or desirable in the individual case, but whether the BGP routing system can scale adequately to continue to undertake this role.

### A Denser Interconnectivity Mesh

The decreasing unit cost of communications bearers in many part of the Internet is creating a rapidly expanding market in exchange points and other forms of inter-provider peering. The deployment model of a single-homed network with a single upstream provider is rapidly being supplanted by a model of extensive interconnection at the edges of the Internet. The underlying deployment model assumed by CIDR assumed a different structure, more akin to a strict hierarchy of supply providers. The business imperatives driving this denser mesh of interconnection in the Internet are irresistible, and the casualty in this case is the CIDR-induced dampened growth of the BGP routing table.

### Traffic Engineering via Routing

Further driving this growth in the routing table is the use of selective advertisement of smaller prefixes along different paths in an effort to undertake traffic engineering within a multihomed environment. Although considerable effort is being undertaken to develop traffic-engineering tools within a single network using *Multiprotocol Label Switching* (MPLS) as the base flow management tool, inter-provider tools to achieve similar outcomes are considerably more complex when using such switching techniques. At this stage, the only tool being used for inter-provider traffic engineering is that of the BGP routing table, further exacerbating the growth and stability pressures being placed on the BGP routing domain.

The effects of CIDR on the growth of the BGP table have been outstanding, not only because of their initial impact in turning exponential growth into a linear growth trend, but also because CIDR was effective for far longer than could have been reasonably expected in hindsight. The current growth factors at play in the BGP table are not easily susceptible to another round of CIDR deployment pressure within the operator community. It may well be time to consider how to manage a BGP routing table that has millions of small entries, rather than the expectation of tens of thousands of larger entries.

We started this journey over ten years ago when considering the scaling properties of addressing and routing. It is perhaps fitting that we tie the two concepts back together again as we consider the future of the BGP inter-provider routing space. The observation that the BGP growth pressures are largely due to an uptake in multihoming and the associated advertisement of discrete connectivity policies by increasingly smaller networks at the edge of the network has a corollary for address allocation policy. In such a ubiquitous environment of multihomed networks, we will also need to review how address blocks are allocated to network providers, because the concept of provider-based address allocation that assumes a relatively strict hierarchical supply structure is becoming less and less relevant in today's Internet.

## References

[1] D. Clark, L., Chapin, V. Cerf, R. Braden, R. Hobby, "Towards the Future Internet Architecture," RFC 1287, December 1991.

[2] V. Fuller, T. Li, J. Yu, and K. Varadhan, "Supernetting: an Address Assignment and Aggregation Strategy," RFC 1338, June 1992.

[3] T. Hain, "Architectural Implications of NAT," RFC 2993, November 2000.

[4] T. Bates, "The CIDR Report," updated weekly at:
**http://www.employees.org/~tbates/cidr-report.html**

[5] E. Chen, Y. Rekhter, "BGP Support for Four-Octet AS Number Space," work in progress, currently published as an Internet Draft:
**draft-chen-as4bytes-00.txt**, November 2000.

[6] "BGP Table Report" updated hourly at
**http://www.telstra.net/ops/bgp**

[7] C. Labovitz, A. Ahuja, "The Impact of Internet Policy and Topology on Delayed Routing Convergence—Update to This Work," ISMA Winter 2000 Workshop, CAIDA, December 2000.

[8] Peter Lothberg, personal communication.

GEOFF HUSTON holds a B.Sc. and a M.Sc. from the Australian National University. He has been closely involved with the development of the Internet for the past decade, particularly within Australia, where he was responsible for the initial build of the Internet within the Australian academic and research sector. Huston is currently the Chief Scientist in the Internet area for Telstra. He is also a member of the Internet Architecture Board, and is the Secretary of the Internet Society Board of Trustees. He is author of *The ISP Survival Guide,* ISBN 0-471-31499-4, *Internet Performance Survival Guide: QoS Strategies for Multiservice Networks,* ISBN 0471-378089, and coauthor of *Quality of Service: Delivering QoS on the Internet and in Corporate Networks,* ISBN 0-471-24358-2, a collaboration with Paul Ferguson. All three books are published by John Wiley & Sons. E-mail: **gih@telstra.net**

# LAN QoS

*by William Stallings*

A typical organization's on-premise network configuration has multiple *Local-Area Networks* (LANs) connected by bridges or Layer 2 switches. The LANs may all be of one type (for example, Ethernet) or may be of mixed types (for example Ethernet, Token Ring, wireless). In either case, the issue of *Quality of Service* (QoS) arises.

## User Priority and Access Priority

The first attempt to deal with LAN QoS in a standardized fashion appears in the original version of IEEE 802.1D, which is a specification that defines the protocol architecture for bridges and Layer 2 switches, which operate at the *Media Access Control* (MAC) level. IEEE 802.1D deals with the interconnection of LANs with the same MAC protocol and with LANs with different MAC protocols. In addition to passing MAC frames from one LAN to another across the bridge, the bridge is able to pass parameters from software that controls the incoming port to the software that controls the outgoing port. Two of these parameters are *user_priority* and *access_priority*.

The *user_priority* and *access_priority* parameters relate to the problem of how to handle priorities. In the case of IEEE 802.3 (Ethernet) and 802.11 (wireless LAN), priority is not supported. Other 802 LAN types support up to eight levels of priority. The *user_priority* value provided to the MAC- layer entity at the incoming port is derived from the incoming MAC frame; in the case of an incoming frame with no priority value, a value of *unspecified* is used. The *user_priority* value issued to the MAC entity at the outgoing port is to be placed in the outbound MAC frame for LAN types that provide a priority field. The *access_priority* refers to the priority used by a bridge MAC entity to access a LAN for frame transmission. We may not want the *access_priority* to be equal to the *user_priority* for several reasons:

- A frame that must go through a bridge has already suffered more delay than a frame that does not have to go through a bridge; therefore, we may wish to give such a frame a higher access priority than the requested user priority.
- It is important that the bridge not become a bottleneck. Therefore, we may wish to give all frames being transmitted by a bridge a relatively high priority.

The rules for handling priorities can now be summarized. The *user_priority* is determined from the priority field of the incoming frame and placed in the priority field of the outbound frame. Priorities are not used to transmit 802.3 and 802.11 MAC frames, and the frames themselves have no priority field. Therefore, if the outbound frame is 802.3 or 802.11, any incoming priority field (from a frame that has such a field) is ignored. If the incoming frame is 802.3 or 802.11 and the outbound frame requires a priority field, then the priority field in the outbound frame is set to a default *user_priority* value. If both incoming and outbound frames carry a priority field, then the priority field in the outbound MAC frame is set equal to the priority field in the inbound MAC frame.

The *access_priority* is also determined from the priority field of the incoming frame. For incoming 802.3 and 802.11 frames, a *user_priority* of 0 (lowest priority) is assumed. Table 1 shows the access priorities assigned to outgoing MAC frames for each of the LAN types, as a function of incoming user priority value. For 802.3 and 802.11, there is no access priority mechanism and, therefore, a priority of 0 is used. For 802.4 and 802.6, there are eight available access priorities, so the incoming user priority is mapped to the outgoing access priority using equality. IEEE 802.12 permits only two priority levels; half of the possible user priority values are mapped into each of these levels. For the two Token Ring types (802.5 and Fiber Distributed Data Interface [FDDI]), although eight priority levels are available, the highest priority (level 7) is not used in bridge forwarding. The reason for this restriction is that the token-passing protocol reserves priority 7 for its use in transmitting frames needed to manage the token-passing process, such as recovering from a frame loss.

**Table 1: Outbound Access Priorities**

| User Priority | Outbound Access Priority per MAC Method | | | | | | |
|---|---|---|---|---|---|---|---|
| | 802.3 | 802.4 | 802.5 | 802.6 | 802.11 | 802.12 | FDDI |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |
| 2 | 0 | 2 | 2 | 2 | 0 | 0 | 2 |
| 3 | 0 | 3 | 3 | 3 | 0 | 0 | 3 |
| 4 | 0 | 4 | 4 | 4 | 0 | 4 | 4 |
| 5 | 0 | 5 | 5 | 5 | 0 | 4 | 5 |
| 6 | 0 | 6 | 6 | 6 | 0 | 4 | 6 |
| 7 | 0 | 7 | 6 | 7 | 0 | 4 | 6 |

| | | |
|---|---|---|
| 802.3 | = | CSMA/CD |
| 802.4 | = | Token bus |
| 802.5 | = | Token ring |
| 802.6 | = | DQDB (Distributed Queue, Dual Bus) MAN |
| 802.11 | = | Wireless LAN |
| 802.12 | = | Demand priority (100VG-AnyLAN) |
| FDDI | = | Fiber Distributed Data Interface (token ring) |

### Traffic Classes

These rules, summarized in Table 1, are effective in communicating a priority requested by a user and in obtaining access to a LAN in competition with other devices also attempting to transmit on that LAN. However, the rules do not directly provide guidance concerning the relative priority with which frames are to be handled by a bridge. For example, consider a bridge connected to a Token Ring on one side and an Ethernet on the other, and suppose that the bridge receives a large volume of traffic from the Token Ring so that a number of frames are buffered waiting to be transmitted onto the Ethernet. Should the bridge transmit these frames in the order in which they were received, or should the bridge account for the user priority of all waiting frames in determining which frame to transmit next? Consideration of this issue led to the development of a new concept, *traffic class,* which is incorporated in the 1998 version of IEEE 802.1D. This new material is sometimes referred to as 802.1p in the literature. This was the designation when the traffic-class standard was in draft form. In the 802 scheme, a lowercase letter refers to a supplement to an existing standard and an uppercase letter refers to a base standard. Thus 802.1D is a base standard defining bridge operation, and 802.1p is a supplement to the earlier version of 802.1D. With the publication of the 1998 version, the traffic-class supplement was incorporated into 802.1D, and the designation 802.1p is no longer used.

The goal of the traffic-class addition to 802.1D is to enable Layer 2 switches and bridges to support time-critical traffic, such as voice and video, effectively. In the remainder of this article, we begin with an overview of the use of traffic classes in bridges. Next, we examine the mapping of user priorities into traffic classes. Finally, we look at the larger issue of QoS in an internet that includes bridges as well as routers and other Layer 3 switches.

The 1998 version of IEEE 802.1D distinguishes three concepts:

- *User priority:* The user priority is a label carried with the frame that communicates the requested priority to downstream nodes (bridges and end systems). Typically, the user priority is not modified in transit through bridges, unless a mapping is needed for the use of a different number of priority levels by different MAC types. Thus, the user priority has end-to-end significance across bridged LANs.
- *Access priority:* The access priority is used, on LANs that support priority, to compete for access to the shared LAN with frames from other devices (end systems and other bridges) attached to the same LAN. For example, the token-passing discipline in a Token Ring network enables higher-priority frames to gain access to the ring ahead of lower-priority frames when frames from multiple stations are waiting to gain access. When both the incoming and outbound LAN are of the same MAC type, the bridge assigns an access priority equal to the incoming user priority. Otherwise, the bridge must perform a mapping as defined in Table 1.

- *Traffic class:* A bridge can be configured so that multiple queues are used to hold frames waiting to be transmitted on a given outbound port, in which case the traffic class is used to determine the relative priority of the queues. All waiting frames at a higher traffic class are transmitted before any waiting frames of a lower traffic class. As with access priority, traffic class is assigned by the bridge on the basis of incoming user priority.

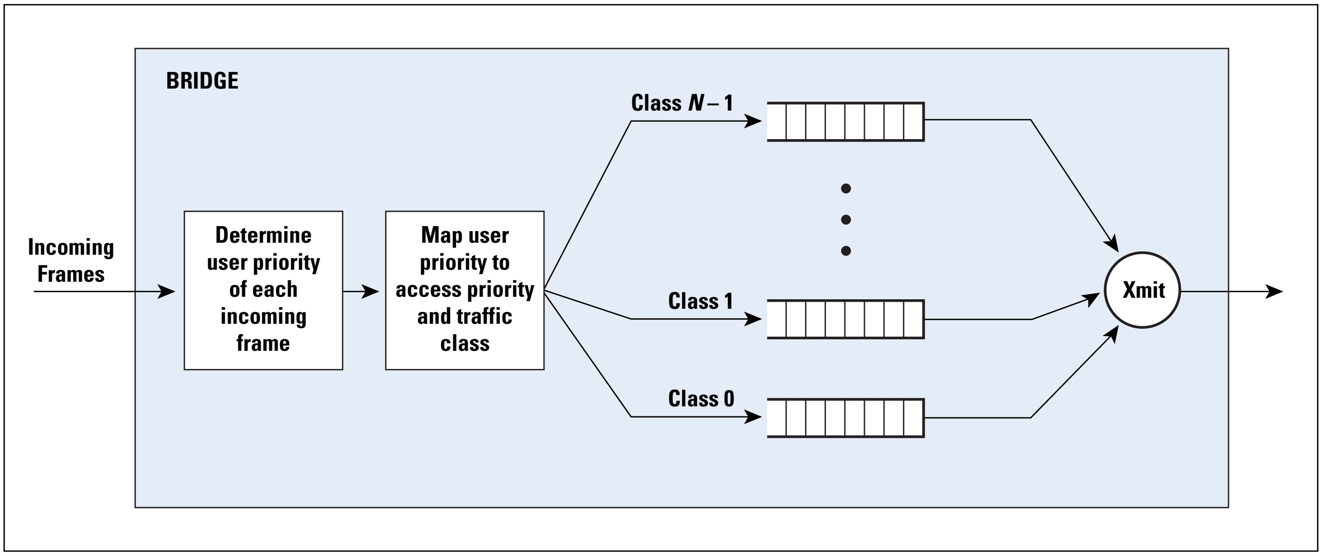
The significance of traffic classes can be seen by recognizing that a frame experiences two types of delay at a bridge:

- *Queuing delay:* The time that a frame waits until it becomes first in line for transmission on the outbound port. This delay is determined by the queuing discipline used by the bridge. The simplest scheme is first-in, first-out (FIFO). Traffic classes permit more sophisticated schemes.
- *Access delay:* The delay that a frame experiences waiting for permission to transmit on the LAN, in competition with frames from other stations attached to the same LAN. This delay is determined by the MAC protocol used (for example Token Ring, Carrier Sense Multiple Access Collision Detect [CSMA/CD]).

The total delay experienced by a frame at a bridge is the sum of its queuing delay and its access delay.

Figure 1 illustrates the mechanism used to support traffic classes at a bridge. A bridge may support up to eight different traffic classes on any outbound port by implementing up to eight distinct queues, or buffers, for that port. A traffic-class value is associated with each queue, ranging from a low of 0 to a high of *N* – 1, where *N* is the number of traffic classes associated with a given outbound port (*N* ≤ 8).

*Figure 1: IEEE 802.1 D Traffic Class Operation*


BRIDGE
Incoming Frames
Determine user priority of each incoming frame
Map user priority to access priority and traffic class
Class N – 1
Class 1
Class 0
Xmit

On a given output port with multiple queues, the rules for transmission follow:

1. A frame may be transmitted from a queue only if all queues corresponding to numerically higher values of traffic class are empty. For example, if there is a frame in queue 0, it can be transmitted only if all the other queues at that port are currently empty.
2. Within a given queue, the order of frame transmission must satisfy the following: The order of frames received by this bridge and assigned to this outbound port shall be preserved for:
    - Unicast frames with a given combination of destination address and source address
    - Multicast frames for a given destination address

In practice, a FIFO discipline is typically used. Thus, a strict priority mechanism is used. It follows that during times of congestion, lower-priority frames may be stuck indefinitely at a bridge that devotes its resources to moving out the higher-priority frames.

### Mapping of User Priority to Traffic Class

IEEE 802.1D provides guidance on the mapping of user priorities into traffic classes. Table 2 shows the recommended mapping. We can make two comments immediately:

1. The mapping is based on the user priority associated with the frame, which, as was mentioned earlier, has end-to-end significance. However, the 802.3 and 802.11 frame formats do not include a priority field, meaning that this end-to-end information could be lost. To address this issue, the bridge is able to reference the priority field contained in a tag header defined in IEEE 802.1Q, which deals with virtual LANs. The 802.1Q specification defines a tag header of 32 bits that is inserted after the source and destination address fields of the frame header. This tag header includes a 3-bit priority field. Thus, if 802.1Q is in use by Ethernet and wireless LAN sources, a user priority can be defined that stays with the frame from source to destination.
2. Outbound ports associated with MAC methods that support only a single access priority, such as 802.3 and 802.11, can support multiple traffic classes. Recall that the traffic class deals with queuing delay, while the access priority deals with access delay.

To understand the reason for the mappings recommended in Table 2, we need to consider the types of traffic that are associated with each traffic class. IEEE 802.1D provides a list of traffic types, each of which can benefit from simple segregation from the others. In descending importance, these types include:

- Network control (7): Both time critical and safety critical, consisting of traffic needed to maintain and support the network infrastructure, such as routing protocol frames.

- Voice (6): Time critical, characterized by less than 10-ms delay, such as interactive voice.
- Video (5): Time critical, characterized by less than 100-ms delay, such as interactive video.
- Controlled load (4): Non-time-critical but loss sensitive, such as streaming multimedia and business-critical traffic. A typical use is for business applications subject to some form of reservation or admission control, such as capacity reservation per flow.
- Excellent effort (3): Also non-time-critical but loss sensitive, but of lower priority than controlled load. This is a best-effort type of service that an information services organization would deliver to its most important customers.
- Best effort (2): Non-time-critical and loss insensitive. This is LAN traffic handled in the traditional fashion.
- Background (0): Non-time-critical and loss insensitive, but of lower priority than best effort. This type includes bulk transfers and other activities that are permitted on the network but that should not impact the use of the network by other users and applications.

Only seven traffic types are defined in IEEE 802.1D. The standard leaves as spare an eighth type, which could be used for traffic of more importance than background but less importance than best effort. The numbers in parentheses in the preceding list are the traffic-class values corresponding to each traffic type if there are eight queues and hence eight traffic classes available at a given output port.

**Table 2: Recommended User Priority to Traffic Class Mapping**

| | | Number of Available Traffic Classes | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| User Priority | 0<br>(default) | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 2 |
| | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | 3 | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 3 |
| | 4 | 0 | 1 | 1 | 2 | 2 | 3 | 3 | 4 |
| | 5 | 0 | 1 | 1 | 2 | 3 | 4 | 4 | 5 |
| | 6 | 0 | 1 | 2 | 3 | 4 | 5 | 5 | 6 |
| | 7 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

We can now address the issue of the mapping between user-priority and traffic-class value. If eight traffic class values are available (eight queues at this output port), the obvious mapping would be equality; that is, a user priority of $K$ would map into traffic class $K$ for $0 \leq K < 7$. This obvious mapping is not desirable because of the treatment of default priorities. For 802.3 and 802.11, which do not use priorities, the default user priority is 0. For other MAC types, such as 802.5, if the user does not specify a priority, the MAC level assigns a default value of 0. The 802.1D standard points out that using a different default value would result in some confusion and probably a lack of interoperability. However, the logical default traffic type is best effort. The solution proposed by 802.1D is to map a user priority of 0 to traffic-class value 2. When there are eight traffic class values available, then user-priority values 1 and 2 map to traffic-class values 0 (background) and 1 (spare value), respectively.

This solution is reflected in Table 2, which shows the mapping of user priority to traffic class when there are eight available traffic classes. The table also shows the mapping when there are fewer traffic classes. To understand the entries in this table, we need to consider the way in which 802.1D recommends grouping traffic types when fewer than eight queues are configured at a given output port. Table 3 shows this grouping. The first row in the table shows that if there is only one queue, then all traffic classes are carried on that queue. This is obvious. If there are two queues (second row), 802.1D recommends assigning network control, voice, video, and controlled load to the higher-priority queue, and excellent effort, best effort, and background to the lower-priority queue. The reasoning supplied by the standard follows: To support a variety of services in the presence of bursty best-effort traffic, it is necessary to segregate time-critical traffic from other traffic. In addition, further traffic that is to receive superior service and that is operating under admission control also needs to be separated from the uncontrolled traffic. The allocation of traffic types to queues for the remaining rows of the table can be explained similarly.

**Table 3: Suggested Traffic Types**

| | | **Traffic Types** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Number of Queues | 1 | **BE** (EE, BK, VO, CL, VI, NC) | | | | | | | |
| | 2 | **BE** (EE, BK) | | | | **VO** (CL, VI, NC) | | | |
| | 3 | **BE** (EE, BK) | | | | **CL** (VI) | | **VO** (NC) | |
| | 4 | **BK** | | **BE** (EE) | | **CL** (VI) | | **VO** (NC) | |
| | 5 | **BK** | | **BE** (EE) | | **CL** | **VI** | **VO** (NC) | |
| | 6 | **BK** | | **BE** | **EE** | **CL** | **VI** | **VO** (NC) | |
| | 7 | **BK** | | **BE** | **EE** | **CL** | **VI** | **VO** | **NC** |
| | 8 | **BK** | — | **BE** | **EE** | **CL** | **VI** | **VO** | **NC** |
| | | **1** | **2** | **0** | **3** | **4** | **5** | **6** | **7** |
| | | **User Priority** | | | | | | | |

Note: In each entry, the boldface type is the traffic type that has driven the allocation of types to classes.
BK = Background VI = Video (<100 ms latency and jitter)
BE = Best Effort VO = Voice (<10 ms latency and jitter)
EE = Excellent Effort NC = Network Control
CL = Controlled Load

## Internet Traffic Quality of Service

The user-priority and traffic-class concepts enable MAC-level bridges and Layer 2 switches to implement a traffic-handling policy within a bridged collection of LANs that gives preference to certain types of traffic. These concepts are needed because these bridges and switches cannot see "above" the MAC layer and hence cannot recognize or utilize QoS indications in higher layers such as IP. However, it is often the case that traffic from a bridged set of LANs must cross Wide-Area Networks (WANs) that make use of QoS functionality. An example of this is an ATM network, which provides for user-specified QoS. Another example is an IP-based internet, which can provide IP-level QoS. Some means is needed for mapping between traffic classes and QoS for such configurations. This is an evolving area of technology and standardization, but a general picture can be provided.

In the case of IP-based internets, the IP *Type-of-Service* (ToS) field provides a way to label traffic with different QoS demands. The ToS field is preserved along the entire path from source to destination through, potentially, multiple routers. Fortunately, the mapping from traffic class to ToS is straightforward. The ToS field includes a 3-bit Precedence subfield. A router connecting a LAN to an internet can be configured to read the Layer 2 Traffic-Class field and copy that into the ToS Precedence field in one direction, and copy the 3-bit Precedence field into the User Priority field in the other direction.

In the case of an ATM connection, a bridge or Layer 2 switch might be connected to a LAN on one side and an ATM network on the other, using the ATM network to link to other remote LANs. For local LAN traffic arriving at the bridge, the bridge must match the user priority level with the appropriate ATM service class and other ATM parameters. For this purpose, the bridge can consult a mapping table whose settings have been predefined through the policy controls of network management software. An appropriate virtual connection is used to carry the traffic. If the traffic exits the ATM network at another LAN, the bridge on that end can map incoming traffic from each virtual connection into the appropriate traffic class and user priority.

## References

A more detailed discussion of bridges, Layer 2 switches, and IEEE 802.1D is contained in [1]. The IEEE 802.1 working group is at
`http://grouper.ieee.org/groups/802/1/index.html.`

[1] Stallings, W., *Local and Metropolitan Area Networks, Sixth Edition,* Prentice Hall, 2000.

WILLIAM STALLINGS is a consultant, lecturer, and author of over a dozen books on data communications and computer networking. He has a PhD in computer science from M.I.T. His latest book is *Local and Metropolitan Area Networks, Sixth Edition* (Prentice Hall, 2000). His home in cyberspace is **WilliamStallings.com** and he can be reached at **ws@shore.net**

# Book Reviews

### E-mail Books

*Essential Email Standards: RFCs and Protocols Made Practical* by Pete Loshin, ISBN 0-471-34597-0, John Wiley & Sons, Inc., 2000. **`www.wiley.com`**

*Internet Email Protocols: A Developer's Guide,* by Kevin Johnson, ISBN 0-201-43288-9, Addison-Wesley, 1999. **`www.awl.com`**

Deciding when to write a book about an exciting new technology is pretty easy. At first issuance of the standards for it, or emergence of a market for it, out will come the requisite texts. In 1993, when the commercial Internet started to surface, Marshall Rose produced *The Internet Message: Closing The Book With Electronic Mail* [Prentice Hall, 1993]; it's an excellent introduction to the core e-mail services. As the market grew, Rose and David Strom issued a more operations-oriented effort, *Internet Messaging; From Desktop to the Enterprise* [Prentice Hall, 1998]. For anyone serious about e-mail technology and operations, it remains required reading.

But what about straight technology exposition when the standards that have been in use for more than 20 years keep getting modified? In the case of Internet mail, this dilemma has been exacerbated by an extended recent effort to coalesce documentation for the service, compiling and clarifying the contents of many independent *Internet Engineering Task Force* (IETF) documents into two, one for the transfer service and one for the mail object definition. The best time to publish a book on the subject would be at the issuance of the two revisions. Unfortunately, the IETF effort has taken perhaps 3 years longer than expected, and Wiley and Addison-Wesley decided the market needed these books earlier. Hence the authors were faced with a juggling act, referring to original specifications, with appropriate nods to the new—but unstable—drafts.

### Comprehensive Introductions

This tactical caveat notwithstanding, Peter Loshin's *Essential Email Standards: RFC and Protocols Made Practical* and Kevin Johnson's *Internet Email Protocols: A Developer's Guide* are credible and reasonably thorough. They introduce the reader to the technical details of Internet mail. Loshin adds detail about the standards culture that produced the specification. Johnson adds a bit of programming detail. No textbook on a technology should be used as the primary reference by someone building products, of course; and these are no exception. These are comprehensive introductions.

With such books, the criteria are simple. I look for helpful overall organization, clear language, and accurate content. These two books qualify. They summarize and restate the basic descriptions of services, data formats, protocol commands, and responses associated with the various standards.

Extra points are assigned when a book comes with commentary that provides some insight into the technical philosophy or operational pragmatics of the technology. Pleasantly, both books have a bit of these extras, too. Such texts typically also have minor technical errors; and these fit that profile, too. Since the reader is not using the book as an implementation reference, the occasional, small errors cause no harm.

Loshin's effort is 330 hardbound pages. Johnson's is about a third longer, softbound. Both books cover the core services of *Submit, Simple Mail Transfer Protocol Service Extensions* (ESMTP), the *Post Office Protocol* (POP), the *Internet Mail Access Protocol* (IMAP), RFC 822, and *Multipurpose Internet Mail Extensions* (MIME), that is, posting, relaying, and accessing e-mail, as well as description of the e-mail object. Both also discuss security. Submit is a recent spinoff from SMTP, for local user-relay posting. It began as a clone of ESMTP, but on a different port, and will permit service-to-service relaying functionality to diverge from the local, first-hop posting process. The market treats POP and IMAP as essentially competitive protocols, and both books explain their details adequately. I wish they had made the very simple architectural point that POP does last-hop delivery, to the user's PC-based message store, whereas IMAP is primarily for user access to a message store on a remote system. That is, one is for simply dumping an entire message queue onto the waiting user machine, whereas the other is for ongoing and interaction with portions of message data. On the other hand, an example of Loshin's extra credit is for noting that ISPs are reticent to support IMAP—they have not yet discovered that they could make money being a small business' back-office data store—whereas corporations like IMAP because it is an open standard that permits replacing proprietary workgroup message stores.

E-mail address resolution can be a bit tricky, requiring general understanding of the Domain Name Service and specific cleverness with MX "routing" records. Johnson devotes a useful, but very terse 2+ pages to the topic. Loshin allocates a 8+ pages.

### Security

As with every other aspect of Internet standards making, e-mail security is problematic because no IETF-originated security protocol has yet gained wide deployment and use. Oddly continuing the peculiarity of security as a topic, both books are a little off-beat, albeit differently. Johnson provides a relatively extensive introduction to basic security technology, including descriptions of various algorithms, as well as a listing of the types of security attacks that can occur. He also discusses enhancements to the basic e-mail protocols for invoking security mechanisms. Loshin has a more functional systems orientation concerning overall e-mail security architecture. Although Loshin does not usually spend much time on ancient history, for some reason in this chapter he discusses two IETF failures of *Privacy Enhanced Mail* (PEM) and *MIME Object Security Services* (MOSS).

Both discuss *Pretty Good Privacy* (PGP), and PGP is certainly the long-standing popular choice among the technical community. Johnson discusses it in some detail; Loshin's coverage is minimal. *Secure MIME* (S/MIME) has support from major industry software vendors. Loshin treats it equally as tersely as he treats PGP. Johnson barely mentions it.

### Standards

Loshin spends the first 50 pages on the Internet standards community, process, and documents. His book also covers Internet News (NNTP) and some work involving standard data for business cards (vCard) and calendaring and scheduling (iCalendar). Besides being interesting topics, these last two were probably included because the Internet Mail Consortium acquired intellectual property rights to the precursor work and highlights the topics on its Web page. Loshin also ends with a chapter about the future, where he adds the topics of instant messaging and message tracking, based on continuing IETF standards work. An included CD-ROM contains a copy of the book, with Web links to cited documents such as RFCs.

Johnson's forays beyond the core services discuss messaging filtering and mailing-list processing, UNIX file issues, and generic, terse descriptions of some programming languages. He also discusses the *Internet Message Support Protocol* (IMSP), the *Application Configuration Access Protocol* (ACAP), and the *Lightweight Directory Access Protocol* (LDAP), protocols for accessing user configuration data. Obviously he intends that the reader take seriously the "Developer's" reference in the book title.

### The Differences

Perhaps it is the programmer's orientation that caused Johnson to be so thorough with his discussions. This includes discussion of e-mail protocols that are not standards and not in use. Loshin in far more selective and reflective. And therein lies the easy distinction between the two efforts. Loshin gives an understanding of a portion of application space, providing the basic technical details tidbits of useful insight. Johnson is more mechanical and more detailed; in effect he chooses to be less selective and more detailed in what he dumps on the reader, letting the reader decide what is useful.

*—Dave Crocker, Brandenburg InternetWorking*
**`dcrocker@brandenburg.com`**

## Wireless and Mobile Network Architectures

*Wireless and Mobile Network Architectures*, by Yi-Bing Lin and Imrich Chlamtac, ISBN 0-471-39492-0, John Wiley & Sons, 2001.

Paging through this book, my first impressions are that it uses very little math and that it is a comprehensive standards-based overview of practical wireless systems. The authors' multidisciplinary tack—systems, networks, and services—is evidenced by their conceptual approach to engineering design issues and their straightforward explanations of implementation issues. The primary concern of the book as a whole is: "How does it all fit together?"

### Organization

The authors divide the book into five major units. The first three units covered their topics well and enhanced my understanding of wireless communications. However, the final two units fell short of my expectations. Coverage of the *Wireless Application Protocol* (WAP) and other up-and-coming issues in wireless networking was patchy and unbalanced.

The "PCS Network Management" section provides an overview of the concepts, definitions, and procedures used in current wireless network implementations. Basic roaming concepts including handoff geometry, detection, and queuing schemes are briefly discussed. An understanding of foundational engineering concepts is assumed as the authors provide detailed algorithmic descriptions of hard and soft handoff message flows.

The "IS-41 Mobile Systems" section provides an introductory overview of *Signaling System 7* (SS7) as a supporting protocol for the IS-41 mobile communications protocol. The importance of integration between these two protocols is presented in practical example format. Intersystem handoff and authentication techniques applicable to IS-41 are then discussed. Included in this section is a functional overview of network signaling for *Personal Access Communications* (PACS) networks as related to IS-41. However, a general understanding of the PACS radio system is assumed.

### GSM

*Global System for Mobile Communication* (GSM) systems are the largest focus of this book. A full ten chapters are dedicated to the concepts and applications of this technology. The section appropriately starts with a high-level overview of the GSM system architecture and moves through mobility management and roaming. Here, the authors present several alternative roaming concepts aimed at reducing the cost of roaming service. Additionally, mobile number portability mechanisms and costs are also addressed. Likewise, significant attention is given to the technical aspects of GSM networks and their integration with data networks. Full chapters are dedicated to describing the GSM network signaling software platform (MAP), operations, administration, and management functions, Voice over IP integration, and General Packet Radio Service over GSM.

For the student, *Wireless and Mobile Network Architectures* is a capstone reference that ties together several courses worth of technical information with a practical focus toward real-world applications. For professional IT managers, engineers, and software developers, it is a practical and handy tutorial for getting up-to-speed on second-generation wireless and mobile technologies.

### Questions

Each chapter ends with a set of very open-ended and thought-provoking analysis and design questions. Reading the chapter does not necessarily prepare you to do in-depth design; rather, you gain enough knowledge to sketch out a basic approach to solving the problem. It is obvious that many of the problems would require interdisciplinary collaboration to arrive at a tenable solution. Members of such a team would contribute different perspectives based on their particular area of expertise.

### Worthwile Reference

This book assumes that the reader has mastered the basics in the field of mobile communications and is seeking to implement a practical design. Throughout the book are many easy-to-follow algorithmic or flow-chart explanations of various wireless communications processes. However, the information gleaned from these treatments tended to be more about functionality than design. Although a worthwhile reference, this book is by no means "all you need to design and implement a mobile services network."

—*Albert C. Kinney*

**kinney@ieee.org**

---

### Would You Like to Review a Book for IPJ?

We receive numerous books on computer networking from all the major publishers. If you've got a specific book you are interested in reviewing, please contact us and we will make sure a copy is mailed to you. The book is yours to keep if you send us a review. We accept reviews of new titles, as well as some of the "networking classics." Contact us at **ipj@cisco.com** for more information.

# Call for Papers

*The Internet Protocol Journal* (IPJ) is published quarterly by Cisco Systems. The journal is not intended to promote any specific products or services, but rather is intended to serve as an informational and educational resource for engineering professionals involved in the design, development, and operation of public and private internets and intranets. The journal carries tutorial articles ("What is...?"), as well as implementation/operation articles ("How to..."). It provides readers with technology and standardization updates for all levels of the protocol stack and serves as a forum for discussion of all aspects of internetworking.

Topics include, but are not limited to:

- Access and infrastructure technologies such as: ISDN, Gigabit Ethernet, SONET, ATM, xDSL, cable, fiber optics, satellite, wireless, and dial systems
- Transport and interconnection functions such as: switching, routing, tunneling, protocol transition, multicast, and performance
- Network management, administration, and security issues, including: authentication, privacy, encryption, monitoring, firewalls, trouble-shooting, and mapping
- Value-added systems and services such as: Virtual Private Networks, resource location, caching, client/server systems, distributed systems, network computing, and Quality of Service
- Application and end-user issues such as: e-mail, Web authoring, server technologies and systems, electronic commerce, and application management
- Legal, policy, and regulatory topics such as: copyright, content control, content liability, settlement charges, "modem tax," and trademark disputes in the context of internetworking

In addition to feature-length articles, IPJ will contain standardization updates, overviews of leading and bleeding-edge technologies, book reviews, announcements, opinion columns, and letters to the Editor.

Cisco will pay a stipend of US$1000 for published, feature-length articles. Author guidelines are available from Ole Jacobsen, the Editor and Publisher of IPJ, reachable via e-mail at **`ole@cisco.com`**

# Fragments

### ICANN Launches At-Large Membership Study

The *Internet Corporation for Assigned Names and Numbers* (ICANN) recently announced that it was commencing a comprehensive study of the structure of its At Large membership. The study will be conducted by an *At Large Membership Study Committee* that will make recommendations to ICANN's Board of Directors on how individuals can effectively participate in ICANN's policy development, deliberations and actions for technical coordination of the Internet.

Mr. Carl Bildt, the former Prime Minister of Sweden and noted United Nations envoy, will serve as Chair of the nine member Study Committee. An international statesman and information technology advisor, Bildt's current duties include Special Envoy of the Secretary General of the United Nations to the Balkans, Member of Parliament of Sweden, and Advisor and Board Member of several Internet and technology-related corporations.

"The Board's approval of the Study Committee and Carl Bildt's selection as Chair is a demonstration of ICANN's commitment to finding an effective way for the perspectives of individuals in every country to be heard and given due consideration," said Vint Cerf, Chairman of the ICANN Board of Directors. "We are extremely fortunate to have someone with Carl Bildt's international consensus building experience to lead this critical effort."

The Committee, which is chartered to seek input from all interested parties and to work toward a broad consensus on ICANN's At Large membership, will use multiple mechanisms for input, including public forums, mailing lists, and a public website. The Committee will encourage the participation of organizations and individuals worldwide, including the development of independent studies and analyses from across the global Internet's constituencies.

"ICANN's actions affect the whole world's Internet users, and I look forward to the challenging task of forging a consensus on the best method for representing this ever-growing constituency," said Bildt. "This will be an international cooperative effort, and I am counting on the participation of a diversity of Internet stakeholders that have an interest in ICANN to help us deliver a workable solution."

The Board invited Charles Costello and Pindar Wong to serve as the Committee's Vice-Chairs. Costello is director of the Carter Center's Democracy Program, and served as an outside monitor for ICANN's At Large elections held last year. Wong served as an ICANN Director and Vice Chairman of the Board during 1999–2000. He also is an active Internet policy leader in the Asia Pacific Region, and Chairman of Verifi (Hong Kong) Ltd., an Internet infrastructure consultancy. The remaining members of the committee are Pierre Dandjinou, Esther Dyson, Oliver Iteanu, Ching-Yi Lu, Thomas Niles, and Oscar Robles.

ICANN also announced the appointment of Denise Michel as the Committee's Executive Director. Ms. Michel has extensive experience in both private and public sector technology policy development, having served previously on the staff of the U.S. National Science Foundation, the American Electronics Association and the U.S. Department of Commerce. From 1993–1995, she was Sr. Technology Advisor to the Secretary of Commerce, Mr. Ronald Brown.

Following public comment, the Board also adopted a charter for the study to ensure a consistent base of expectations on the scope and details of the study committee's work. ICANN has posted the charter on its website at:
**`http://www.icann.org/committees/at-large-study/charter-22jan01.htm`**

For more information about the At Large Membership Study Committee, see: **`http://www.atlargestudy.org/`**

### Correction

In the article "The Trouble with NAT," which appeared in our previous issue, a table of private nonroutable IP addresses taken from RFC 1918 was shown. The table contained an error, as pointed out by a couple of our readers. The correct table appears below.

| Class | Private Address Range |
|---|---|
| A | 10.0.0.0 ... 10.255.255.255 |
| B | 172.16.0.0 ... 172.31.255.255 |
| C | 192.168.0.0 ... 192.168.255.255 |

### Upcoming Events

The Internet Society (ISOC) will hold its annual conference INET in Stockholm, Sweden, June 5–8, 2001. For more information, see:
**`http://www.isoc.org/inet2001/`**

Just before INET, The Internet Corporation for Assigned Names and Numbers (ICANN) will hold its meeting in the same venue. The dates are June 1–4, 2001 and you can find more information at:
**`http://www.icann.org/calendar.htm`**

The Internet Engineering Task Force (IETF) will next meet in London, England, August 5–10. For more information, see:
**`http://www.ietf.org`**

This publication is distributed on an "as-is" basis, without warranty of any kind either express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. This publication could contain technical inaccuracies or typographical errors. Later issues may modify or update information provided in this issue. Neither the publisher nor any contributor shall have any liability to any person for any loss or damage caused directly or indirectly by the information contained herein.

**The Internet Protocol Journal**

**Ole J. Jacobsen**, Editor and Publisher

*The Internet Protocol Journal is published quarterly by the Chief Technology Office, Cisco Systems, Inc.*
*www.cisco.com*
*Tel: +1 408 526-4000*
*E-mail: ipj@cisco.com*

CISCO SYSTEMS

The Internet Protocol Journal, Cisco Systems
170 West Tasman Drive, M/S SJ-10/5
San Jose, CA 95134-1706
USA

ADDRESS SERVICE REQUESTED