Ascend

TECHNICAL OVERVIEW

A PRACTICAL IP SWITCH FOR NEXT-GENERATION NETWORKS

August 26, 1996

ASCEND

TABLE OF CONTENTS

The Relentless Demand for Bandwidth 1

IP Switching 2

The First Practical IP Switch 3

The Internet Infrastructure: An Urgent Need for IP Switching 5

© 1996, Ascend Communications, Inc. Confidential

A PRACTICAL IP SWITCH FOR NEXT-GENERATION NETWORKS

The Relentless Demand for Bandwidth

Internet Protocol (IP) traffic levels have grown exponentially over the past decade for both the Internet itself, and more recently, for Intranets. The underlying LANs and WANs have grown accordingly in both size and complexity. Whenever growth reached a level where overall performance degraded, the network was partitioned into smaller segments, which in turn were re-linked by new, modified IP internetworking equipment. This dance of scale continues today: a relentless cycle of growth, partitioning and re-linking as network architects race to keep pace with the exponential growth in demand.

But in several environments, particularly in large-scale public and private backbones, traffic levels have reached the point where conventional routers are overwhelmed, and proposed routing enhancements are almost instantly rendered ineffective.

The core of nearly all performance problems in modern networks lies in the conventional router's basic design. The specific areas responsible are outlined briefly in Table 1.

Table 1 — Conventional Router

Design Element	Conventional Design Choice	Resulting Problem (Particularly in Backbones)
IP Forwarding Connection Fabric		• Aggregate bus bandwidth is shared by all media: - Adding ports degrades overall performance - Performance normally peaks well before all available slots are used - Sustained packet forwarding rate is far short of theoretical • Limits aggregate packet forwarding rates to unacceptably low levels • Emerging 622 mbps standards (OC-12) exceed maximum bus capacity
Type	• Shared Bus	
Bandwidth	• 533 mbps	
Aggregate Bandwidth	• Up to 2 Gbps with multiple buses in parallel (no single, shared path greater than 533 mbps)	
Latency	• Unpredictable: ranges from milliseconds to infinity	
Route Determination Table and Lookup		• Actual performance is a small fraction of design's potential performance • Packets delayed as destination address is converted to next-hop by full route table look-up • When cache hits drop from 100% to 50% (still an optimistic rate) performance drops by 90% • CPU saturates when cache hits drop below 50% • Saturated CPU causes dropped packets and severely diminishes overall throughput • Limited table size does not meet expected address growth, denying some users network access
Architecture	• Centralized, software-based function shared by all media cards	
Lookup Method	• High-speed cache attempt followed by much slower full table lookup	
Lookup Latency	• Latency low for cache, but high for full table lookup	
Cache Hit Rate	• Typically much less than 50%, making full lookup the norm	
Route Table Size	• Limited by software table lookup speed	
Router Scalability		• Aggregate throughput is limited • Chassis size is disproportionately large for actual performance • Increases cost per port
Port Density	• Typically single port per card	
Maximum Size	• Limited by shared bus bandwidth	

© 1996, Ascend Communications, Inc. Confidential

When used in large IP networks, some of the adverse effects of conventional router design choices are:

- Escalating traffic demands are continually unmet
- Some users are denied access as network resources saturate
- All users experience performance degradation from lost packets and retransmissions
- Unpredictable latency and poor throughput restrict network expansion and prohibit use of new quality of service or integrated service options, or high bandwidth media
- Access providers and major corporations with large networks are forced to over-engineer the IP backbone–at great expense

Simply put: The inherent design limitations of conventional routers are being outpaced by the growing bandwidth demands and complex route topologies of IP-centric networks. And over-engineering with more of the same is only a temporary–and costly–patch.

With these inherent router problems universally recognized and increasingly understood, a variety of next-generation solutions is beginning to appear. While these proposed solutions vary, depending on the vendor's installed-base solutions or engineering capabilities, all involve an inevitable evolutionary step: IP switching.

IP SWITCHING

Just what is IP switching? At one level, the phrase is an oxymoron. IP exists at Layer 3 of the protocol Reference Model defined by the International Standards Organization (ISO). ISO Layer 3 is the network layer, and internetworking at the network layer is called "routing" by definition, while switching is more of a physical Layer 1 "cross-connect" function. Yet the industry now accepts the concept of LAN switching that occurs at the media access control (MAC) Layer 2, which was previously relegated to "bridging" exclusively.

One possible definition of IP switching is LAN switching (at the MAC layer) moved up a notch to the IP network layer. Such an IP switch would employ LAN switch technology with store-and-forward or cut-through designs. But this approach is inherently flawed: the "bridging" function performed by a LAN switch is very different from "routing" function, which must be performed–and performed fast–by an IP switch.

Routing in a Switch

The role of routing remains the same even when implemented using high-speed switching technology. Network layer routing:

- Supports complex topologies not available with MAC layer switching
- Provides stability with a more predictable, scalable and dependable hierarchical network topology
- Maximizes utilization of shared and/or limited resources, especially in the backbone and in relatively low bandwidth WAN links
- Facilitates transparent connectivity among diverse network types and protocols
- Adds optional security and firewall protection safeguards
- Enables static and dynamic virtual LANs (VLANs)

Simply put: IP switching must maintain high bandwidth and low latency while handling all routing functions, including route table lookup and packet forwarding. Such a challenge is easier said than done, which is why there is so much confusion in the marketplace about IP switches.

© 1996, Ascend Communications, Inc. Confidential

Ascend Communications believes an IP switch is a full-fledged router that uses a fast internal switching engine that runs over any media — not just ATM — combined with equally fast route table lookup to achieve wire-speed performance. No tricks. No gimmicks. No complicated "flows" or unrealistic cache "hits." Just a fast, scalable routing switch.

THE FIRST PRACTICAL IP SWITCH

Ascend established ten essential design objectives for its IP switch project:

- Compatibility with existing network infrastructures, including interoperability with conventional routers and LAN switches
- Full compliance with industry standards to eliminate any need for proprietary gateways or special clients software
- IP next-hop address lookup fast enough to take advantage of a switching fabric's low latency and high throughput
- Sustainable throughput that is independent of traffic characteristics (flows and cache hits)
- Wire-speed performance for all external LAN/WAN ports
- Support for a wide range of popular LAN and WAN media
- Support for ATM without an architectural dependence on ATM
- Linear scalability within each IP switch and in a network of IP switches
- Packaging in a small chassis, able to fit into the limited space in network points-of-presence (POPs)
- An order of magnitude improvement in price/performance

All ten objectives were achieved in GRF 400, Ascend's IP Switch product offering. A block diagram of the GRF 400's design is shown in Diagram 1.

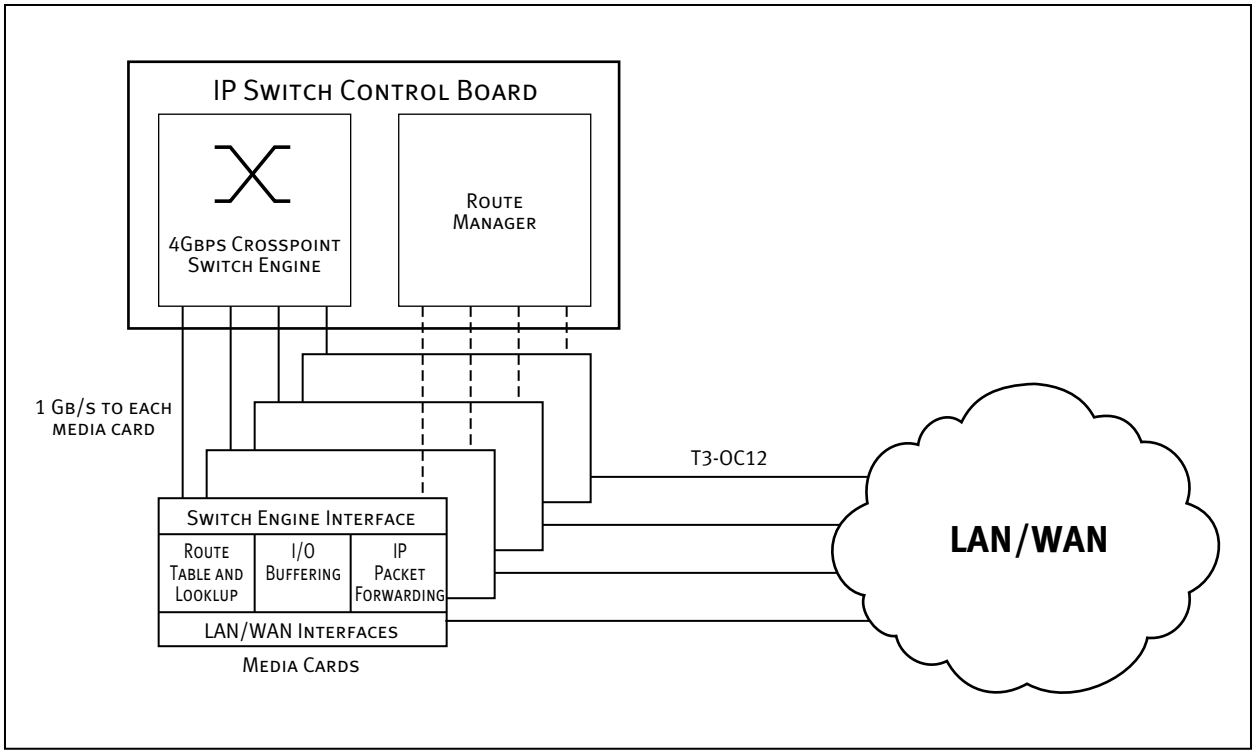

Diagram 1 — The GRF 400 offers wire-speed performance for up to 32 LAN/WAN interfaces

© 1996, Ascend Communications, Inc. Confidential

At the heart of the GRF 400 is a 4 Gbps crosspoint switch fabric. This switch fabric has four data paths, each with a non-blocking, bidirectional capacity of 1 Gbps. The high speed, non-blocking nature of the switch engine permits multiple data paths to operate simultaneously. Path setup time is less than 250 ns (0.25 µs).

External LAN/WAN interfaces are provided by an assortment of plug-in IP forwarding media cards (see the GRF 400 At-A-Glance sidebar for a list of media cards). Each media card has its own dedicated 1 Gbps path in the crosspoint switch fabric, thereby assuring linear scalability to the switch's capacity of 4 Gbps. Up to four media cards can be inserted into the GRF 400, for a total of up to 32 external LAN/WAN interfaces.

Every media card is effectively a full IP router attached to the internal switching engine. Each media card contains 8 MB of input/output buffering, a complete next-hop route table, hardware-assisted route table lookup and an IP forwarding engine.

Key to the GRF 400's wire-speed performance is Ascend's unique hardware-assist table lookup technology. Route lookup times range from 1-3 µs with up to 150,000 next-hop routes in the table. (The largest tables on the Internet today have only around 50,000 next-hop routes.) Performance is enhanced even further with parallel processing of table lookups occurring on each media card, another technique that helps assure linear scalability. The route manager on the controller board, which also contains the switch fabric, maintains the master route table and distributes updates simultaneously to all installed media cards.

The high-throughput switch fabric combined with the fast hardware-assist table lookup give the GRF 400 its unparalleled performance capabilities. A theoretical exercise might be to calculate the packet-per-second (PPS) performance of the GRF 400 based on its unique architectural attributes. But PPS ratings for routers or switches are often like MIPS ratings for computers: a Meaningless Indicator of Performance or Speed. The true test of performance is the ability to maintain wire-speed throughput on all LAN/WAN interfaces, independent of favorable traffic characteristics or ideal network topologies. And the GRF 400 easily handles a full complement of 32 Fast Ethernet LANs, each operating at its maximum 100 mbps data rate.

The GRF 400 At-A-Glance

The GRF 400 IP Switch is ideal for large-scale public and private backbone applications. Packaged in a 5" high, rack-mountable chassis, the GRF 400 is configurable with four slots that accept plug-in media cards for external LAN/WAN interfaces. Available IP forwarding media cards include:

- 8-port 10/100 mbps Ethernet
- 2-port 52 mbps/45 mbps DS-3 HSSI (High-Speed Serial Interface)
- 4-port 100 mbps FDDI (Fiber Distributed Data Interface)
- 2-port 155 mbps OC-3 ATM (Asynchronous Transfer Mode UNI 3.0/3.1)
- 2-port 155 mbps OC-3 SONET (with frame relay and PPP Framing)
- 1-port 800 mbps HIPPI (High Performance Parallel Interface)
- 1-port 622 mbps OC-12 ATM (available Q1-97)

The GRF 400 delivers wire-speed performance for any mix of LAN and WAN interfaces. Its full routing capabilities include support for RIP1/2, BGP4, EGP, OSPF, IS-IS (Q1-97) and IGMP (multicast). The reliable design maximizes system up-time. The GRF 400, built for continuous operation, offers a robust suite of availability features including redundant load-balancing power supplies, hot-swappable media cards and power supplies, and remote SNMP-based management.

Table 2 compares conventional router designs with the GRF 400 IP Switch.

© 1996, Ascend Communications, Inc. Confidential

Table 2 — Conventional Router vs. GRF 400 IP Switch.

Conventional Router Design/Problem	GRF 400 IP Switch Design/Solution
Shared Bus • All media cards contend for theoretical aggregate packet forwarding rate - adding media cards degrades performance - aggregate forwarding rate not achieved - maximum port counts may not be attainable • Chassis capacity may not be achieved • Causes unpredictable latency • Limited scalability yields poor price/performance	**Crosspoint Switch** • Non-blocking design with 250 ns path setup • Independent 1 Gbps data paths with hardware-assisted parallel route table lookups ensure that: - 4 Gbps aggregate forwarding rate is achievable - rate scales linearly as new cards are added - wire-speed performance available on all 32 LAN/WAN interfaces • Consistently low latency • Full scalability yields excellent price/performance
Layer-3 (IP Network Layer) Decision Handling • Shared cache used for fast lookup • Cache hits are typically well below 50% • Each cache miss requires a high latency, software-based lookup in a single route table shared by all media cards • Results in drop to 10% of maximum performance	**Layer-3 (IP Network Layer) Decision Handling** • Hardware-based route table lookup • Route table replicated on each media card • Lookup (with 150,000 routes) in less than 3 µs assures wire-speed performance • Caching is rendered unnecessary • Throughput independent of traffic characteristics

THE INTERNET INFRASTRUCTURE: AN URGENT NEED FOR IP SWITCHING

The need for IP switches is most profound today in the world's largest network: the Internet. Industry analysts estimate there are some 30 million Internet users worldwide currently, with a monthly growth rate of 15 percent!

Already the Internet is showing signs of strain caused by the conventional router's inability to handle the traffic volume: system-wide congestion, regional brownouts, local blackouts. Every Internet user has experienced the problem first-hand in the form of "Host Address Not Found" Web error messages. Problems with the Internet have a rippling effect, as well, with many corporations beginning to build Internet-based Intranets and Virtual Private Networks.

Redundancy and resiliency protect the Internet from catastrophic failure — at least for the time being. But such over-designs are enormously expensive. A better way is urgently needed.

The performance and scalability of Ascend's GRF 400 offers an enduring solution for the Internet infrastructure. Diagram 2 shows an improved end-to-end Internet configuration using the high-performance IP switching.

© 1996, Ascend Communications, Inc. Confidential

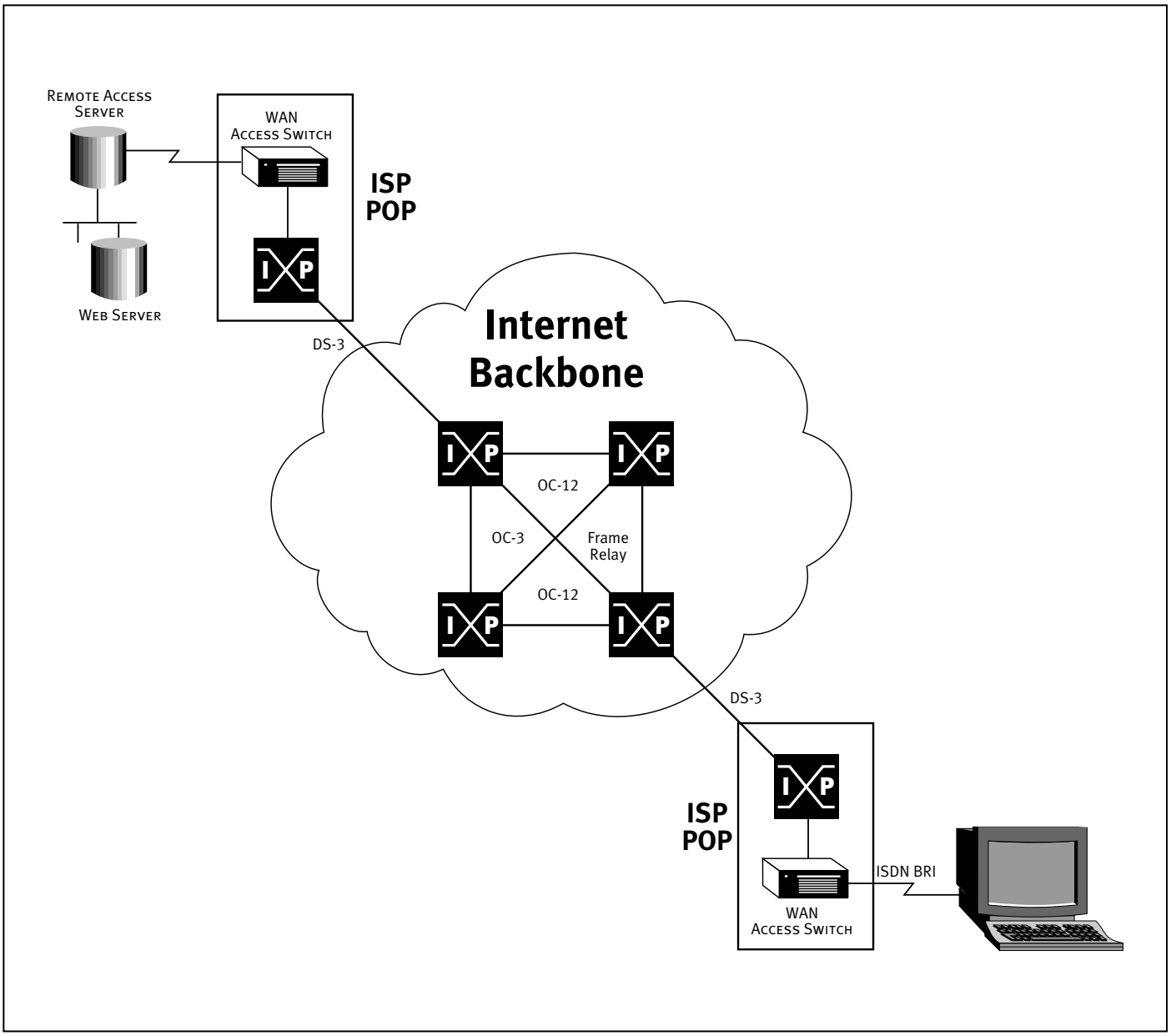

Diagram 2 — The Internet backbone configured with IP switches for maximum price/performance

The Internet backbone shown employs IP switches in a mesh topology, interconnected with high-speed WAN lines. ISPs attach to the edge of the backbone, also over high-speed WAN lines. Traffic that originates at one end traverses a WAN access switch at the ISP (over relatively low-speed dial-up WAN links, such as the ISDN Basic Rate Interface), passes through the ISP's IP switch over a high-speed LAN connection, and then on to the carrier's IP switch at the edge of the Internet backbone. Traffic is passed from switch to switch within the backbone until reaching the edge at the destination end, where the carrier's IP switch directs traffic to the ISP's IP switch, which in turn passes it through the WAN access switch for final delivery to the recipient.

Routers, assisted by high-speed ATM switches, constitute the end-to-end Internet today. IP switches are the next-generation replacement for these separate routers and switches. Of course, conventional multiprotocol routing and ATM switching still have a vital role to play in many networking environments. But for the best in price/performance in any large IP network, the IP switch is soon to displace the alternatives as the building block of choice.

© 1996, Ascend Communications, Inc. Confidential



Ascend. Remote Networking Solutions That Work.™

Ascend Communications, Inc.
1275 Harbor Bay Parkway
Alameda, CA 94502, USA
TEL: 510.769.6001
FAX: 510.814.2300

E-mail: info@ascend.com
Toll Free: 800.621.9578
FAX Server: 415.688.4343
Internet Home Page:
http://www.ascend.com