The AXI 540 router and the public IP network edge

Gordon Saussy

In April 1999, Ericsson acquired Torrent Networking Technologies, a USA-based start-up manufacturer of high-end Gigabit routers for public IP networks. The team at Torrent—now the Ericsson IP infrastructure group within the Datacom Business Unit—began rolling out its new AXI 540 Edge Aggregation Router during the fourth quarter of 1999. The AXI 540 router, which has been optimized to serve the edge of new public IP networks, is a natural fit in Ericsson's expanding portfolio of data products.

In this article, the author discusses the requirements for routers at the edge of the new public IP network, and presents the key technologies and architectural capabilities of the AXI 540 router, which was built to fulfill these requirements.

From the Internet to the public IP network

Fueled by the popularity of the World Wide Web (WWW) and the increased use of intra-company as well as inter-company e-mail, the growth of the Internet during the past five years has been phenomenal—to the point of shaking the foundations of the data communication and telecommunications industries. These industries are aggressively moving toward a common vision of IP convergence that is based on the Internet protocol (IP) and a packet-switched infrastructure.

Today, the Internet is an *ad hoc* overlay built on top of existing telecommunications and wide-area data communication networks. For the most part, entrepreneurial Internet service providers (ISP) have built the Internet around existing carrier networks, with little participation from carriers (network operators). The Internet scores points for its ubiquity—it is accessible from virtually anywhere on the planet via modem and over high-speed telecommunications services in all major markets—but it falls far short of meeting the requirements for reliability and performance associated with traditional wide-area data communication services. Consequently, the vision of IP convergence, and the many bold predictions regarding IP telephony, electronic commerce and the transformation of the workplace will not be fulfilled by the Internet in its present state.

Nonetheless, new public IP networks are already being built to supplant the Internet along with major elements of the telecommunications network. The public IP network is not being built simply as an overlay by independent ISPs; instead, it is an integral part of the carrier network that is built and delivered by a hybrid of innovative, entrepreneurial carrier-service providers. The public IP network will maintain the ubiquity of today's Internet, augmented with the reliability of telecommunications service and performance enhancements that enable the most aggressive applications to operate seamlessly.

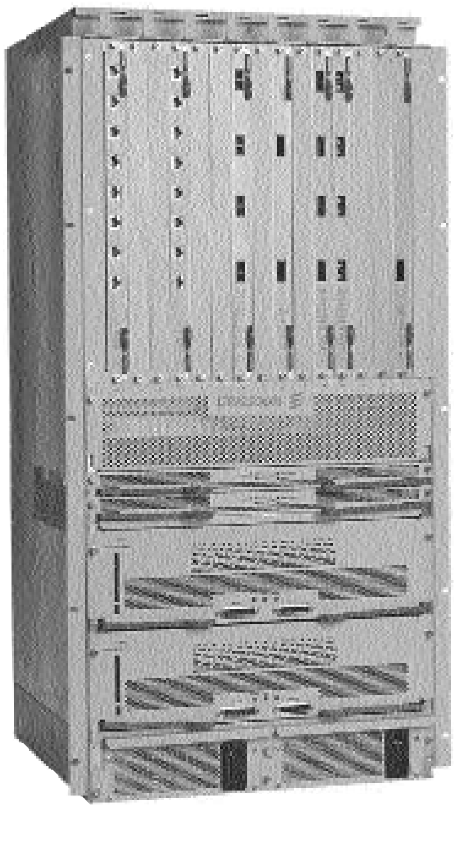


Figure 1
Rear panel view of the AXI 540 router.

Converged IP services

To some, the term IP convergence describes the transition from present-day telecommunication infrastructure technologies (SONET/SDH TDM and ATM/Frame Relay switching) to new, IP-centric solutions. This concept, which could be termed the IP infrastructure convergence, is technologically fascinating and will certainly have a major impact on carrier backbones during the next five to ten years. To others, IP convergence describes a transformation of the current multiprotocol service model to one of bundled services based on IP; that is, IP service convergence. This field is evolving very quickly and will have a highly visible impact in the near term.

The notion of IP service convergence is a simple one. Today, the typical business contracts with a variety of providers of diverse communications services. For instance, a medium-sized corporation might have different providers for basic telephony (POTS, fax), videoconferencing service, private wide-area networking (WAN), remote local area network (LAN) access, and Internet access. Each of these services potentially requires its own WAN connection and customer premises equipment (CPE). Similarly, each service probably requires subscribers to master some level of user expertise. By contrast, IP service convergence promises a simplified model that delivers all these services (and many more yet to be imagined) via a single (or optionally redundant) connection to an IP network. Consequently, subscribers will be able to access a full menu of communication services via a wireline or wireless connection from one service provider.

The Internet as we know it today can be used to prototype new services, but it is not equipped to deliver converged IP services on a large scale. The routers used to forward IP traffic within the Internet and the IP protocol suite that runs as a distributed operating system within these routers are weak in two key areas: performance and topology. The new public IP network will rely on innovations in each of these areas to enable a bundled service model.

Poor performance issues in the Internet today are easy to observe: anyone who has used a browser to surf the Web has firsthand experience of them. Statistical multiplexing of bandwidth is an inherent characteristic of packet-switched networks, and contention for use of the network creates variations in latency ("jitter") and data loss. Present-day IP routers make no effort to improve on these issues for any type of traffic; that is, all service is unreserved, best-effort and connectionless. The converged service model will require more intelligent handling of different classes of traffic within the public IP network, thereby ensuring that certain applications (such as voice) are given priority over other classes of traffic.

Topology is currently a more subtle issue than performance in the Internet. Present-day routers maintain a single topology database (routing table) for the global Internet. From any specific router to any point in the Internet, there is one unique path, dynamically maintained through the exchange of topology information between routers. The protocols currently used to exchange this information

- do not accommodate private traffic (such as a corporate WAN backbone);
- are unable to reroute traffic around congestion points; and
- do not easily allow topologies to be constrained by commercial factors.

The public IP network will require enhancements to be made to topology management, by introducing IP-based virtual private networks (VPN), topologies based on class of service (CoS), and simpler implementation of commercial constraints (for example, the preference of one backbone alternative over another due to lower transport costs).

With enhanced quality and class of service (QoS/CoS) capability and more flexible topologies that support IP-based VPNs, the public IP network will be able to deliver the full range of telecommunications and data communication services, with public Internet access as a bundled feature. New architectures, products and standards are being developed that will be used to build the public IP network and make the convergence of IP services a near-term reality.

BOX A, ABBREVIATIONS

ALU	Arithmetic logic unit
ARP	Address resolution protocol
ASIC	Application-specific integrated circuit
ATM	Asynchronous transfer mode
BGP	Border gateway protocol
CoS	Class of service
CPE	Customer premises equipment
DA	Destination address
DS	Differentiated Services ("DiffServe")
DSCP	DiffServe code point
DVMRP	Distance-vector multicast routing protocol
DWDM	Dense wavelength-division multiplexing
EBGP	Exterior BGP
FR	Frame Relay
IBGP	Interior BGP
IETF	Internet Engineering Task Force
IP	Internet protocol
IS-IS	Intermediate system-to-intermediate system
ISP	Internet service provider
LAN	Local area network
LSP	Label-switched path
MPLS	Multiprotocol label switching
NTP	Network timing protocol
OC-48	Optical carrier 48 (2.5 Gbit/s)
OSPF	Open shortest path first
PIM-D	Protocol-independent multicast, dense mode
PIM-S	Protocol-independent multicast, sparse mode
PoP	Point of presence
POTS	Plain old telephone service
PPP	Point-to-point protocol
PVC	Point-to-point virtual circuit
QoS	Quality of service
RIP	Routing information protocol
RISC	Reduced instruction set comprocessor
SA	Source address
SDH	Synchronous digital hierarchy
S-G	Source address-to-multicast group
SONET	Synchronous optical network
STM-16	Synchronous transport module 16
TDM	Time-division multiplexing
ToS	Type of service
VPN	Virtual private network
WAN	Wide area network
WWW	World Wide Web
xDSL	Digital subscriber line (various types)

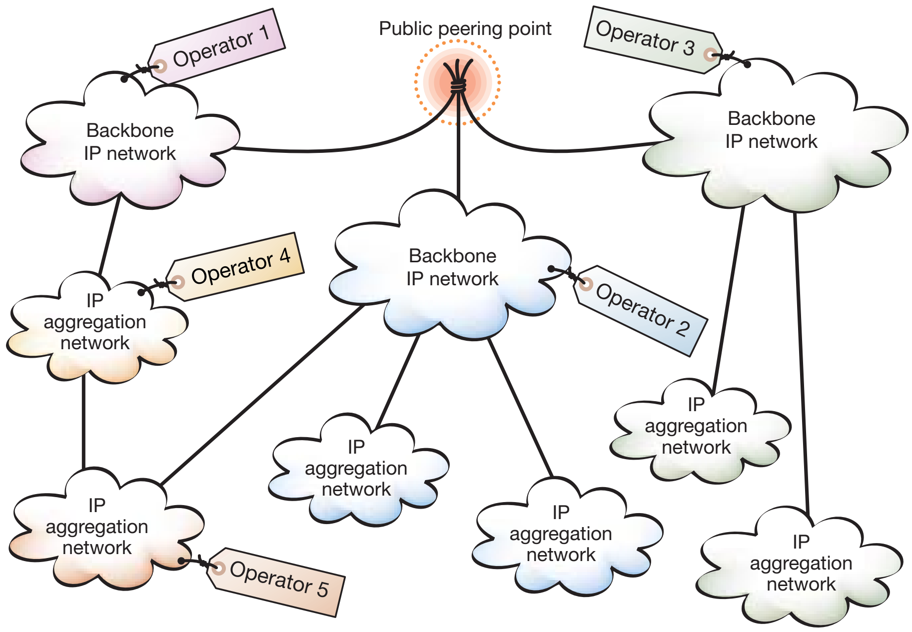


Figure 2
The public IP network: multiple tiers, multiple operators.

Building the public IP network

As with the Internet (as well as telecommunications networks), the public IP network will not be a single network; instead, it will be composed of a mesh of parallel networks that are interconnected at major peering points. Each of these parallel networks, which will be an independent IP network, may in turn be broken down into different elements that are owned and operated by different carriers. The structure of each parallel IP network can be broken down into two major elements:

- backbone networks; and
- aggregation networks.

Backbone networks

Backbone networks will be operated nationally and internationally by large carriers and service providers. They will always run over large fiber plants, typically with a dense wavelength-division multiplexing (DWDM) layer that enables Terabits per second (Tbit/s) of aggregate bandwidth in the network. These networks are always built as a mesh of interconnected IP routing nodes that are richly interconnected by point-to-point links. There are, however, numerous ways of layering the IP network over the fiberoptic backbone. These solutions can be loosely grouped as follows:

- IP-over-ATM multiservice backbones—these networks use asynchronous transfer mode (ATM) switching to multiplex IP traffic with other traffic across the backbone. Routers interconnect via point-to-point virtual circuits (PVC) over ATM backbones. In some cases, the ATM switches also function as active routing nodes using multiprotocol label switching (MPLS) technologies. The advantage of this architecture is its ability to support existing backbone traffic (non-IP data, non-data) alongside IP, which greatly facilitates migration toward the next-generation Internet. However, it also carries a price in terms of loss of throughput, due to ATM overhead, and in terms of ATM network management. Not all network architects believe this price is offset by the benefits of ATM.
- IP-over-SONET/SDH backbones—these networks eliminate the ATM layer, implementing point-to-point links between IP routers directly over SONET/SDH rings (which, in turn, run over DWDM). If present, non-IP traffic is carried over separate point-to-point connections in the same SONET/SDH structure. As with the IP-over-ATM approach, this solution can be implemented today, based on proven standards.
- IP-over-DWDM backbones—networks of this kind are still purely theoretical. The idea is to replace the SONET/SDH layer with a new, lightweight physical

layer that maps IP traffic directly to DWDM fibers. Proponents of this solution argue that because much of the structure of SONET/SDH is optimized for circuit switching, a simpler approach that is optimized for IP packets will result in better price-to-performance.

The vision of IP infrastructure convergence calls for rapid migration to IP-over-DWDM backbones. In reality, however, all three backbone solutions will evolve, and the migration toward a common solution in the backbone will take many years.

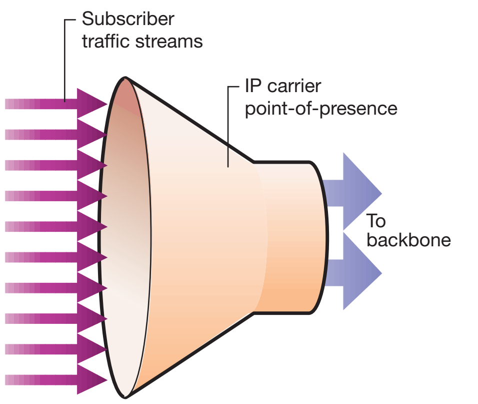


Figure 3
IP aggregation in public IP networks.

IP aggregation networks

Carriers (large and small) and ISPs will operate IP aggregation networks within service areas, which might encompass an industrial park, a few states, or even an entire country. In terms of logic, an IP aggregation network will function like a funnel: thousands or tens of thousands of subscriber connections will be transported via carrier switching and multiplexing networks into an IP aggregation point, where powerful aggregation routers map streams of subscriber traffic to backbone connections. The functionality provided by these new IP aggregation networks can be viewed in three domains: subscriber access termination, subscriber traffic processing, and backbone integration.

Subscriber access termination

The logical connection between the subscriber and the IP aggregation point will be a point-to-point IP connection. Carrier networks will use diverse Layer 1 and Layer 2 multiplexing and switching technologies to deliver thousands of these connections within a service area. Alternatives will range from high-speed leased lines and ATM or Frame Relay PVCs to IP/PPP/ATM connections over various digital subscriber line (xDSL) networks to wireless and cable-modem networks. The IP aggregation routers will thus need to offer tens of thou-

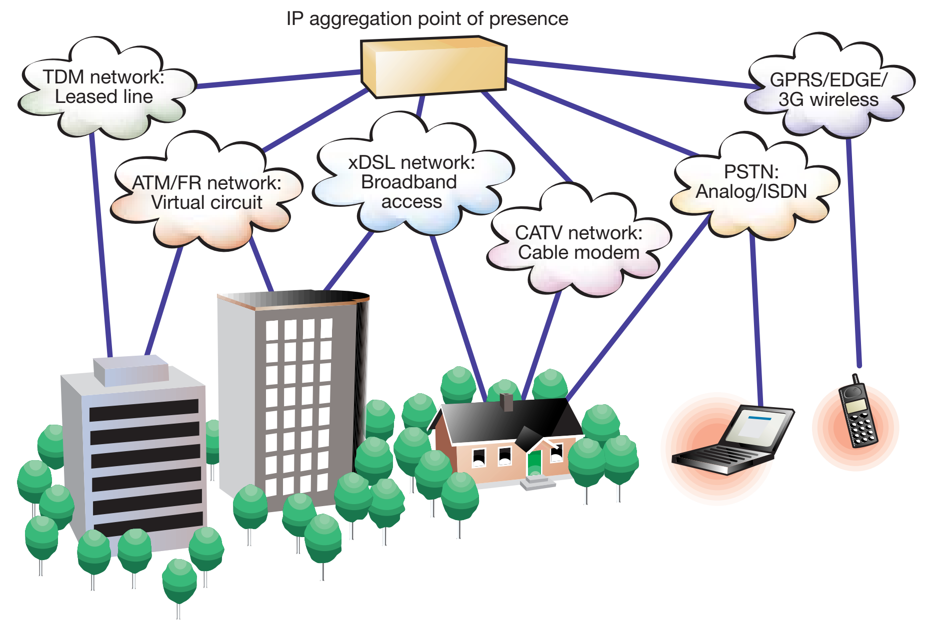


Figure 4
IP subscriber access: diverse alternatives.

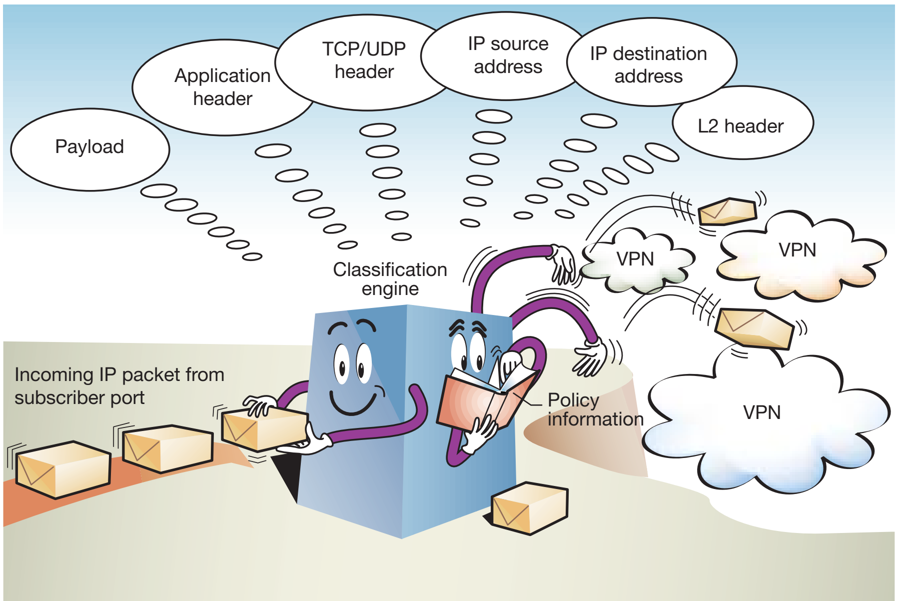


Figure 5
Classifying and handling IP subscriber traffic.

sands of virtual IP interfaces on a variety of physical ports, to ensure easy integration with diverse networks. Below the IP aggregation router, the carrier's access network can operate much as it does today—as a Layer 1 or Layer 2 switched network with minimal IP awareness.

Subscriber traffic processing

Many thousands of subscriber streams are terminated and aggregated at the IP aggregation point. Thus, the IP aggregation router must quickly be able to classify received packets according to predefined policies, per subscriber and application. This extends beyond simple classification into a finite pool of priorities—each traffic class may require mapping to a different VPN, and may require unique shaping and prioritization.

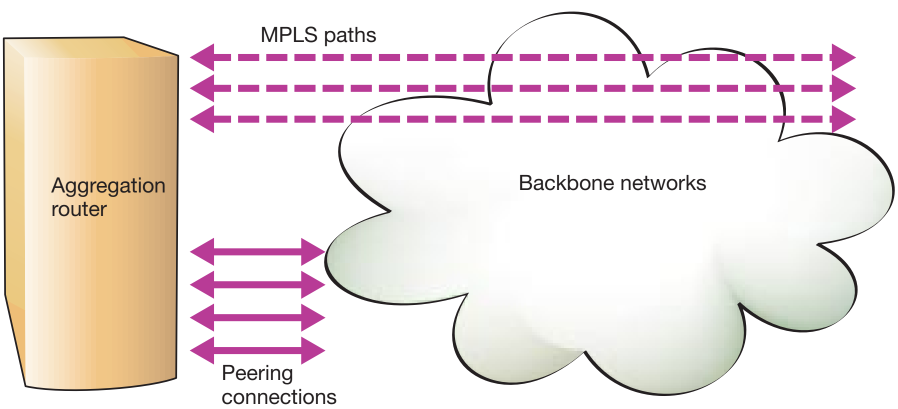
Backbone integration

A primary function of the IP aggregation routers is to route aggregated traffic onto IP backbone networks. This demands comprehensive support for routing protocols (OSPF, IS-IS and BGP-4) as well as the ability to map traffic to supported backbone class-of-service levels and traffic-engineered topologies. Two new standards for IP networks will play a key role: DiffServ and MPLS.

The new Differentiated Services, or DiffServ, standard allows IP traffic to be marked for preferential handling by the network. DiffServ redefines a byte in the existing IP header (the type of service, or ToS, byte) to include a six-bit "DSCP" (DiffServ code point) that indicates the service requirements for the packet. DiffServ-capable nodes examine the value of this field and perform forwarding operations accordingly. Of the 64 possible DSCP values, the Internet Engineering Task Force (IETF) intends to define up to 32 as "global" service classes. The remaining 32 service classes will be left open to network-specific definition. Nodes can also rewrite DSCP values in transit. The DSCP value allows certain packets to be prioritized ahead of others at each network node, thereby reducing jitter and increasing "goodput" (throughput of payload packets) for specific traffic streams (albeit at the expense of less important traffic).

The multiprotocol label switching initiative, which is much broader than DiffServ, has evolved into a family of standards within the IETF. The basic concept of MPLS is to prepend an extra label to the IP packet. Thus, instead of performing a complex lookup to determine the packet's destination, intermediate nodes in an IP network can simply process the label to determine the packet's egress path. Since forwarding within an MPLS "cloud" is solely based on the label, MPLS readily supports IP-based virtual private networks. Moreover, because the process that manages the labels in an MPLS network is decoupled from the basic topology processes of the network, MPLS can be used to augment the basic topology with new paths—a capability called "traffic

Figure 6
Integrating with IP backbones.



engineering." By engineering extra label-switched paths (LSP) for certain classes of traffic, and by using MPLS LSPs to optimize the use of a complex mesh, network operators can generally improve goodput and reduce jitter.

While MPLS and DiffServ represent enhancements to basic IP networks, neither makes any presumptions about the underlying protocol sandwich. Standards have been defined to allow MPLS to run directly over ATM or Frame Relay backbones as well as over any network that carries IP traffic in point-to-point protocol (PPP) frames. Because DiffServ operates strictly within the IP packet, it can be used in any IP network. Infrastructure convergence is not a requirement for benefitting from these new standards—the only criterion is that the IP routing nodes within the infrastructure be capable of supporting the new standards.

Ericsson's portfolio of solutions for new public IP networks is rich. Indeed, Ericsson has taken a leadership position with regard to the new MPLS and DiffServ standards. Ericsson's product range spans from the IP "core" all the way through the aggregation layer and into the diverse access networks (xDSL, wireless, TDM, ATM/FR, and so on) that deliver subscriber traffic to the network. Ericsson's infrastructure solutions for new IP backbone and aggregation networks are particularly powerful.

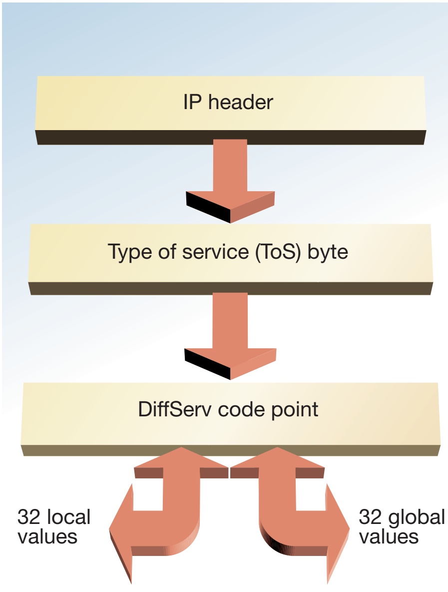


Figure 7
The DiffServ code point (DSCP) in the IP header.

Backbone networking solutions

The AXI 520 IP core router, which can be used to build IP core networks at speeds up to OC-48/STM-16 (2.5 Gbit/s) over SONET/SDH connections, fully implements the new MPLS and DiffServ standards. The AXD 301 IP/ATM switch sup-

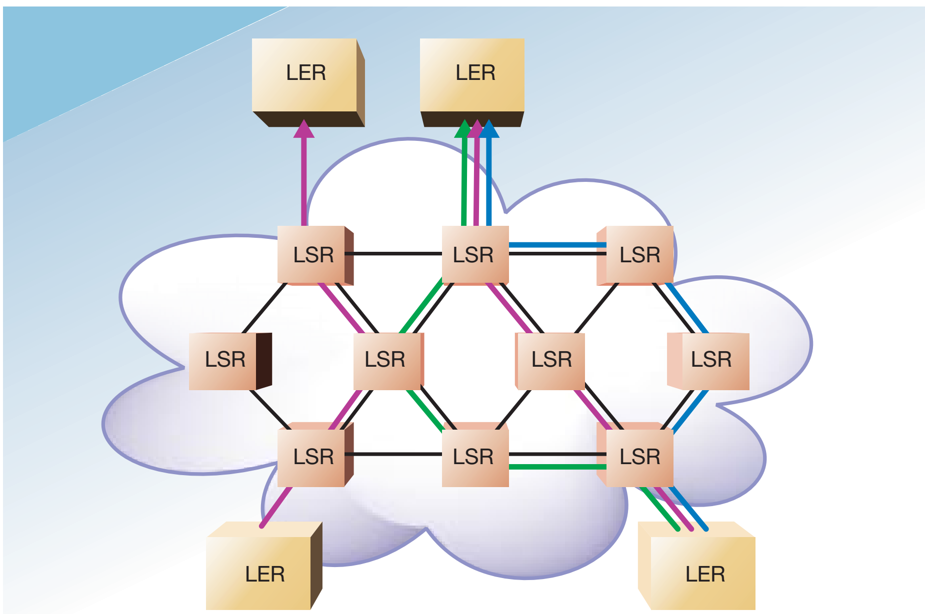


Figure 8
Multiprotocol label switching (MPLS) in the backbone.

BOX B, IPACTION ROUTING SOFTWARE

The AXI 540 router runs all routing software on a centralized and redundant "route processor," which is essentially a single-board computer that has been optimized to handle all the background operations for the router. For optimum resiliency and smooth integration of new features, the routing software runs as a series of independent processes over a streamlined UNIX operating system. Processes with the IPaction software suite include unicast routing, multicast routing, QoS handling, SNMP management, NTP service, and others.

Any router's most important characteristic is the breadth and depth of its routing software. The IPaction suite (currently at revision 2.1, the fifth production release) is rich with features and supports all popular routing protocols. It has been tested extensively in major ISP networks around the world, and demonstrated stability and interoperability in production networks. As the engineering team continues its work on future IPaction releases, this software will evolve into one of the industry's most comprehensive IP routing code bases.

ports ATM cell switching and multiprotocol label switching, and scales to over 100 Gbit/s of switching capacity.[1] Ericsson's DWDM solutions can be used to build metropolitan-area backbones for IP as well as other telecommunications technologies.

Aggregation network solutions

Ericsson offers a broad set of solutions for new IP access and aggregation networks. Optimized access nodes serve narrowband access, xDSL access and new wireless IP access systems. With the new AXI 540 Edge Aggregation Router, Ericsson also addresses the market for high-speed fixed access and point-of-presence (PoP) aggregation.

The AXI 540 Edge Aggregation Router

Service providers and carriers who build the public IP network have very specific criteria for their IP aggregation routers:

- Can the product route?
- Will the product scale?
- Is the product reliable?
- Will the product deliver advanced IP services?

Ericsson's AXI 540 Edge Aggregation Router and its IPaction routing software have been optimized to address these criteria.

Can the product route?

Ericsson's IPaction routing software supports a complete range of routing protocols (Box B):

- Routing information protocol (RIP);
- Open shortest path first (OSPF);
- Intermediate system-to-intermediate system (IS-IS);
- Interior or exterior border gateway protocol (IBGP/EBGP);
- Distance-vector multicast routing protocol (DVMRP); and
- Protocol-independent multicast, dense and sparse modes (PIM-D and PIM-S).

The software suite also implements a comprehensive range of advanced border gateway protocol features, including route mapping, communities, route flap dampening and route reflectors, plus a broad set of value-added features, such as BOOT-P, DHCP relay, proxy address relay protocol (ARP) and network timing protocol (NTP). Major Internet service providers around the world have tested the IPaction software suite and demonstrated full interoperability with existing router backbones.

Will the product scale?

The architecture of the AXI 540 router supports scaling in every important dimension. It scales throughput to over 20 Gbit/s or to more than 20 million packets per second, aggregated from as many as 40,000 virtual IP interfaces per system. It also supports as many as 400 interior or external BGP (IBGP/EBGP) peers and can accept over six million routes from these peers. The active routing table maintained at each port can hold as many as 400,000 network prefixes or 64,000 source address-multicast group (S-G) pairs for multicast routing. The overall system supports up to 50,000 classification filters at each port, mapping traffic to as many as 120,000 independent queues across the system's switch fabric. The inherent scalability of the AXI 540 router architecture addresses every key requirement for long-term growth.

Is the product reliable?

Ericsson's AXI 540 routers deliver fully resilient routing at the hardware, software, system and network levels. The AXI 540 platform uses redundant power supplies, redundant fan trays, redundant 20 Gbit/s switching fabrics and redundant route processors to ensure maximum availability at the hardware level. Ericsson's IPaction software partitions all major capabilities into separate software tasks that run in protected memory over a multitasking operating system, thereby ensuring that a fault in one task cannot affect the operation of others. What is more, the overall design of the AXI 540 system uses SONET/SDH alarms and multipath routing features to identify faults in the network and quickly reroute around them.

Will the product deliver advanced IP services?

Delivering routing that works is a matter of software development and testing; scalability and reliability require a solid architecture and implementation. But to truly enable the convergence of IP service in the new public IP network, a product must incorporate visionary technological breakthroughs. The AXI 540 Edge Aggregation Routers use a custom silicon-based forwarding path that embodies new, patented techniques for route lookup and service management in IP networks.

Every physical port on the AXI 540 router has a local copy of the entire routing table—up to as many as 400,000 destination network entries—and a dedicated ASIC that searches the table and performs packet-processing operations. Within this ASIC, specialized circuits implement a new algorithm (developed by Torrent's founding engineering team) for ultra-fast, ultra-efficient "longest-prefix match" searches in a large routing table. Since every port has its own dedicated ASIC, the system easily scales to higher and higher speeds—every new card added to the AXI 540 router carries all the "horsepower" it needs to forward packets at wire speed to any destination in the global public IP network. Ericsson was recently issued a patent on the techniques developed for this route lookup technique, called the ASIK algorithm (ASIK is an anagram of the inventor's initials).

Advanced IP services require more than just high-speed routing. The custom per-port silicon used in the AXI 540 also performs various simple table lookups, to compare each packet's entire IP header to a list of up to 50,000 filters. The packet's IP destination address (DA), IP source address (SA), Layer 4 protocol and port numbers, and type-of-service label (that is, DiffServ CodePoint) can be matched against installed filters to classify a packet for special handling. Once classified, the AXI 540 silicon-based forward path can

- discard packets immediately—this is useful if you have traffic that you do not want to forward (in implementing IP security, however, router-based packet filtering is not a substitute for a full-fledged firewall);
- police packets to a specified byte-per-second rate—the AXI 540 router allows you to implement either
  - strict policing (the immediate discarding of packets that match a filter when a specified rate has been exceeded); or
  - soft policing (the marking of out-of-profile traffic for early discard when the output port is congested);
- prioritize and queue traffic in the switch fabric—the AXI 540 switch fabric supports four priority levels and up to 8,192 independent queues at every slot; traffic that matches a specified filter can be queued independently of other traffic;
- remark the DiffServ label—at the egress port, the AXI 540 silicon-based forwarding path can change the value of the DiffServ label on the packet; this allows traffic from different filters to be aggregated for core bandwidth handling; and
- originate MPLS tunnels—at the egress port, the AXI 540's silicon-based forwarding function will prepend specified packets with MPLS labels and originate label-switched paths into an MPLS-enabled backbone network.

The ASIC technology used in the AXI 540 router is neither a full-fledged RISC microprocessor nor a simple set of state machines. It is best described as a custom arithmetic logic unit (ALU) with a unique set of built-in IP-specific operations that runs a short set of microcoded instructions on every packet. This approach, which allows some functionality to be enhanced over time through new software, maintains the wire-speed routing and classification capability that is essential at the edge of the new public IP network.

With wire-speed silicon routing for all types of traffic, the AXI 540 router ensures that every application from every subscriber sees the performance it needs. This powerful set of capabilities, together with the scalability and reliability of the system and the power of IPaction routing software, makes the AXI 540 router the clear technology leader for new IP aggregation applications.

Conclusion

The rapid evolution of the Internet during the past five years has shaken the foundations of industries, created and eradicated fortunes, and catalyzed social and economic change. But this is only the beginning. Ahead lies a new era of IP-based communications—the era of the public IP network.

The public IP network will transform the communications industry and enable the convergence of a variety of services to an IP-centric model. It will require powerful "core routers" for IP backbones and a new class of IP aggregation routers for consolidating subscribers and the delivery of services.

Ericsson, the leader in mobile communications worldwide, is delivering a comprehensive set of solutions for new public IP networks, from the core to the edge. The AXI 540 Edge Aggregation Router is a key member of this powerful family of next-generation networking solutions.

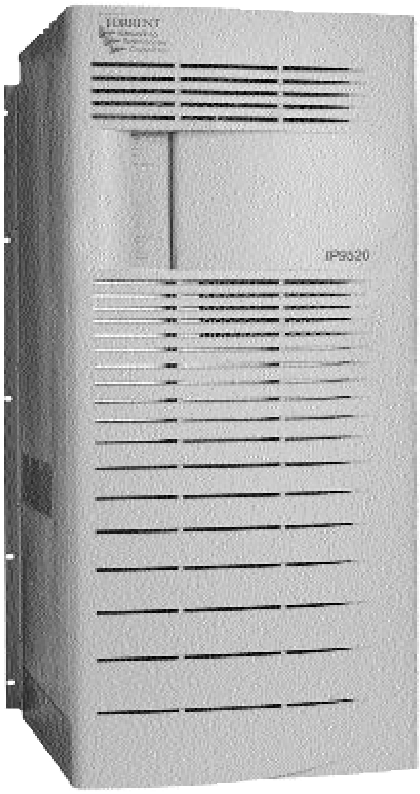


Figure 9
AXI 540 router (front view).

REFERENCES

1 Blau, S. and Rooth, J.: AXD 301—A new generation ATM switching system. Ericsson Review, Vol. 75 (1998):1, pp.10-17.