No. 116
Technology
White Paper

Weathering the Routing Flood in the Optical Metro Service Provider Backbone with Riverstone's Distributed Hardware Routing Feature

Introduction

Whether as a content provider or as a content peruser, doing business on the Internet has evolved from being a novelty to being a necessity. This has caused a massive influx of high bandwidth connectivity to the Internet. Content providers are building out in colocation facilities, businesses are getting higher speed connections, and consumers are migrating to DSL and cable modem technologies as fast as suppliers can build the products. The nature of the traffic on the Internet is also changing. Traditionally, traffic consisted of stable, long-lived connections between a client and a server, or server to server, transferring files or emails. The new access uses the growth of the World Wide Web with bursty traffic and with clients potentially bouncing between many servers in a short amount of time. The change in traffic patterns puts a tremendous amount of load on routers making routing decisions. This situation is aggravated by bandwidth increases in the Internet core. This means more traffic is able to go out of and come back into Metro Area Providers' networks faster than ever.

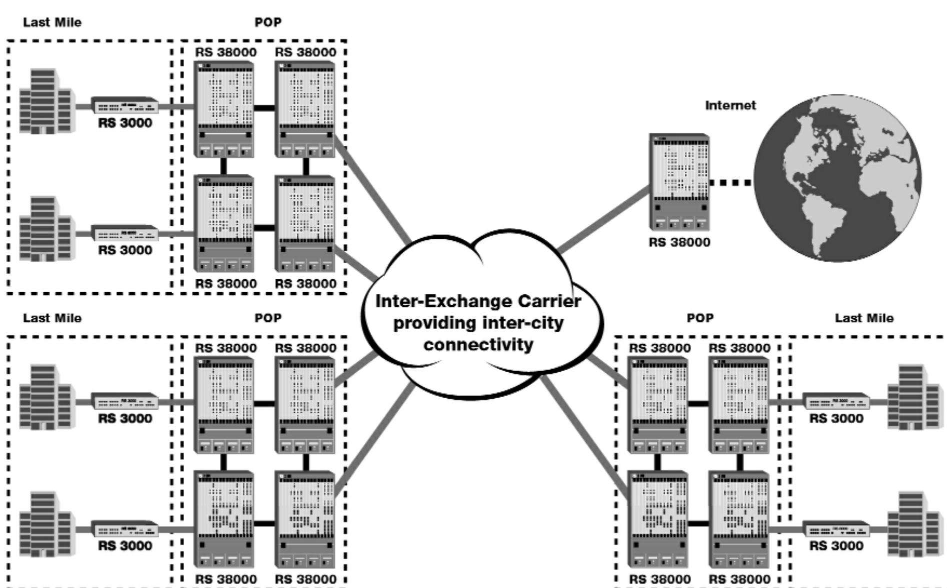

Fig.1 – Speeding up the Edge and Core

River STONE NETWORKS™

The dynamics of routing paths are very fluid. As Figure 1 shows, the availability of high bandwidth access technology for today's Internet customers and high bandwidth access to the Internet together create a potential throttle point when routing traffic from customers to the Internet. Riverstone Networks Layer 3 Switch Router architecture alleviates the problem by completely separating the background task of generating the route tables, and the real-time task of processing the packet to create flow information. This is accomplished through Distributed Hardware Routing. Riverstone's unique method of distributing the routing tables to each line card allows the line card to make real time forwarding lookups as the first packet in a stream arrives and then send it directly to the best next hop on the way to its ultimate destination. The ever-changing packet flow of the Optical Metro Backbone is transported at wire speed without taxing the Riverstone Switch Router.

Speeding up the Process with Hardware-based Routers

Just a few short years ago, Internet routers were strictly software based with T1/E1 connections as high-speed access links and T3 as a heavy-duty backbone. Software routers process every packet coming through the router using the CPU on the control module. Along with the packet processing, the control module would also perform all routing protocol functions, and create and maintain the tables used to make the routing decisions on the packet. The performance of software-based routers were (and still are) bound by the processing power of the CPU. When the access and aggregation rates were relatively low speed, the limitations of software based routers were tolerated. Now these devices are best relegated to low-speed tributaries to the Internet, making way for state-of-the-art hardware multi-layer switch routers.

Hardware-assisted routers speed up the process greatly by splitting the functionality. The first packet goes to the control module to determine where to route the packet based on information found in a Forwarding Information Base (FIB), and forwards the packet to the next hop accordingly. At the same time that the packet is being forward, the control module writes the forwarding information for the flow into the ingress line card along with any Quality of Service and access control policy information associated with the flow. When subsequent packets arrive, the ASICs on the line card look to see if the forwarding information is available locally and sends the packet along based on the forwarding and policy information that the control module wrote into the line card's memory. Hardware routers effectively split the process. Line cards process packets and forward them at line rate. The control module focuses on

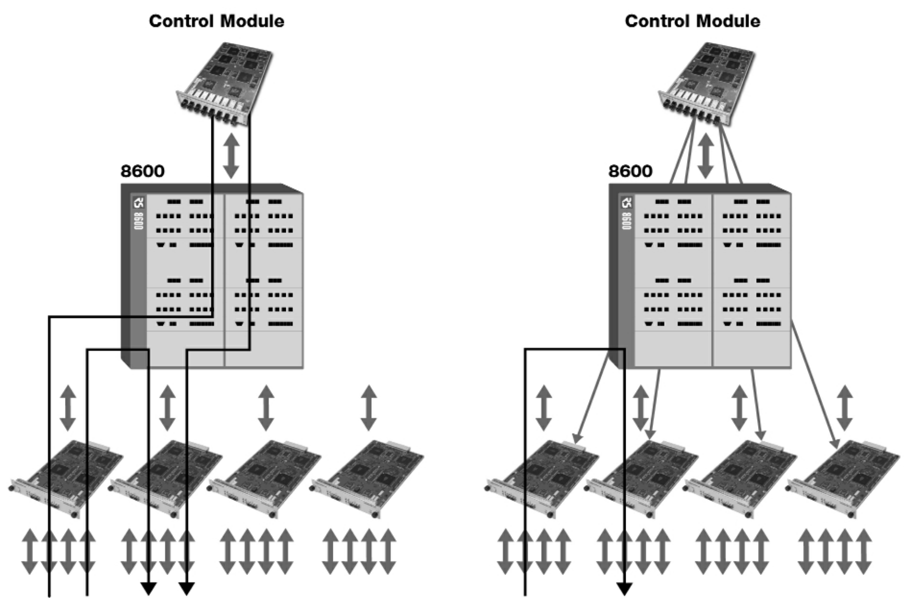

Fig.2 – Packet Paths Comparison

Riverstone Networks

optimizing routing and other control plane functions. Their development has revolutionized the usefulness of routers in the Internet.

Riverstone Networks has taken this concept to the next level. Instead of waiting for the first packet to arrive (and having the control module CPU look up the flow entry in the FIB) to write the information on the line card, Riverstone Switch Routers distribute the routing FIB to the line cards as it is created by the control module. The Riverstone ASICs on the line cards look up the routing information and policy information based on the source address as the first packet arrives. The appropriate Layer 3 routing information is read with the policy applied and the packet is sent along to the appropriate next hop as desired. This is all done at line rate, without interrupting the control module. Even with a continuous stream of packets with changing sources and destinations, the Riverstone Switch Router is able to use the Distributed Hardware Route tables to send packets at the highest rate without taxing the control module.

This state-of-the-art hardware is supported by world-class software. Riverstone's robust set of Carrier-class routing protocols required for Optical Metro Network infrastructures include BGP-4, OSPF, and IS-IS, along with RIP and static route tables. These have been refined and tested over that last four years and blessed by various service providers as Carrier Grade routing code – meaning Tier 1 and 2 service providers have certified Riverstone Networks routing code to interoperate with Cisco GSR series and Juniper M series core routers. The control module is able to run any combination of the above protocols, depending on your network topology. The control module creates a Routing Information Base for each routing protocol it runs. The best routes are selected from the RIB tables, buffering out any route flaps, and are collected into a FIB route table. The resulting FIB route table is distributed to each of the line cards with hardware route lookup support. When a packet requiring routing enters a line card, the route table is accessed to determine if a route exists or if the packet needs to go to the control module for further analysis. When a route is found in the table, the policy tables are examined and the filters are applied to the packets. If a packet passes through the filters, the packet is sent along to the next hop on its path to its ultimate end destination based on the information in the Distribute Hardware Route table. All of this processing is done in hardware on the line card. This is a true advance over software-based routers that need the CPU to look at every packet that passes through the router – a potentially fatal process in today's networking environment. As soon as a new route is learned, or a change in the status of a current route is seen, the route tables on the line cards are instantly updated by the control module.

The revolution in edge access bandwidth provisioning and the terabit routing optical core demand new ways of thinking in routing technologies. The Riverstone ASIC's ability to make the real time routing decisions without accessing the control module answer the demand for high performance switch routing. This moves the routing paradigm into a new level enabling MSPs to build next generation network infrastructures that maximize revenues for the MSP and satisfaction for their customers.

Distributed Hardware Routing

Riverstone Switch Routers have the ability to enhance the hardware routing mode with a more granular service offering on a port-by-port basis. When a port is used for aggregating edge access connectivity, it is important to apply traffic grooming features such as flow filtering, flow-rate limiting, and other provisionable features. These features utilize the assistance of the control module in configuring the service when the first packet is received, then the line card implements the service in hardware for the remaining packets in a flow. This has a minimal impact on edge traffic, and allows for offering tiered services to customers. The aggregated traffic on the optical backbone performs better with the Distributed Hardware Routing mode enabled. Riverstone Switch Routers are able to determine which mode is optimal at the time of configuration and automatically select the correct mode of operation.

The Distributed Hardware Routing mode is designed for use in the aggregation and distribution router. Edge access aggregation switch router interfaces perform the traffic grooming and filtering before reaching the optical backbone ports of the Riverstone Switch Router. By applying the service features on the tributary interfaces of a Riverstone Switch Router, the optical aggregation interfaces are able to handle the large fluctuations in data flow more freely. The Riverstone Networks design allows ports or interfaces to act in either edge or aggregation modes, applying specific services to traffic flows, or Distributed Hardware Routing of large amounts of policed data with the ability to apply screening on specific packet source information as desired. Other hardware routers offer one or the other, but not the flexibility that the Riverstone Networks' router offers, which allows for the tailoring of the router to meet the needs of the network administrator.

At the time of configuration, the Riverstone Switch Router control module looks at features enabled for a specific port or interface, and determines the optimal mode of operation – Distributed Hardware Routing mode or Flow Services mode. When traffic arrives on a line card port, the ASICs begin processing the packet. If a flow already exists for the traffic, then the flow entry information is used for determining what to do with the packet. If there is no flow entry, and the port of entry for the packet is set for Flow Services mode, then the packet is sent to the control module for processing. If the port of entry is configured for Distributed Hardware Routing mode, then the ASIC looks up the appropriate next hop information for the destination address, and looks up the filtering associated with that port. The filters are applied, and if the data is okay to go, the packet is forwarded to the appropriate line card and sent on its way to the next hop on its journey.

An example configuration would be an RS 38000 with channelized T3 and 10/100 Fast Ethernet ports aggregating edge traffic, and Gigabit Ethernet long-reach ports interconnecting aggregation routers. The Flow Services mode would be enabled on the channelized T3 and Fast Ethernet ports, and the Distributed Hardware Route mode would be enabled on the Gigabit Ethernet ports.

In Figure 3, as traffic enters the RS 38000 from the edge ports, the packets are filtered, classified, and prioritized as they are sent to the aggregation links. Each of the RS 38000s is grooming the traffic as it comes in from the edge. The traffic on the aggregation mesh enters each RS 38000 and is routed using the hardware route tables. This configuration optimizes the usage of the hardware. The data in the aggregation mesh has already been pre-screened at the edge; the hardware route table instantly looks up the next hop and sends it along. The queuing features of the output ports are also utilized since the packets have all been classified on the incoming edge ports. This allows control and high priority traffic to go through before medium and low priority, while utilizing strict priority or Weighted Fair Queuing techniques on sending out the traffic and Weighted Random Early Detection to manage the queue buffers.

When Flow Services mode is enabled, the statistics for each flow are maintained on the line card. In Distributed Hardware Route mode, the card maintains information on the number of packets and bytes sent to a specific next hop entry. This allows for getting granular information concerning the specific customers for accounting and SLA fulfillment documentation at the edge, while seeing higher-level changes in your overall network traffic flow patterns using the Distributed Hardware Route packet statistics.

The overall benefit of Distributed Hardware Routing is the optimization of your network. Since the aggregated traffic has already been groomed at the edge, there is no reason to send traffic to the control module when the line card can route it. This frees up the control module to perform routing protocol functions and other background tasks optimizing overall network performance

River STONE NETWORKS™

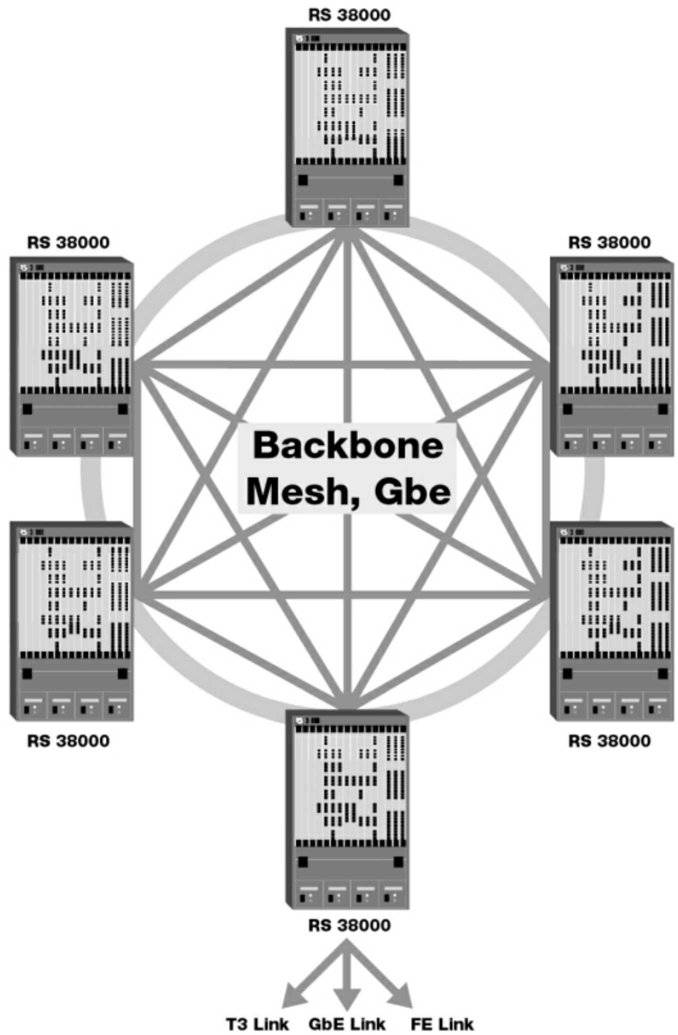

Fig.3 – Edge Access and Optical Backbone Aggregation Traffic

Summary

The explosion of growth in and the fast changing nature of traffic traversing the Internet require bold new ways of processing the traffic and routing it on its way. The Riverstone Networks Switch Routers include state-of-the-art ASICs meeting today and tomorrow's needs in the metro networking market. The Riverstone ASICs offer robust Flow Services features for grooming and policing traffic during Edge Access Aggregation and a Distributed Hardware Routing mode of operation for meeting the demanding needs of Optical Backbone Routing. The unique design of the Riverstone Switch Router allows for one Carrier-class software set to operate across the line of products. This allows for ease of deployment, configuration, management, and operation of the Riverstone Switch Router anywhere in the metro network topology.

River STONE NETWORKS™

Acronyms

ACL	*Access Control List*
ANSI	*American National Standards Institute*
ASIC	*Application-Specific Integrated Circuit*
ASP	*Application Service Provider*
ATM	*Asynchronous Transfer Mode*
CBR	*Constant Bit Rate*
CWDM	*Coarse Wave Division Multiplexing*
DS1/DS3	*Digital Signal, Level 1 (1.54 Mbps) or 3 (44.7 Mbps)*
DSL	*Digital Subscriber Line*
DWDM	*Dense Wave Division Multiplexing*
E1/E2	*European Trunk 1/2 (2 Mbps/34.3 Mbps)*
ERP	*Enterprise Resource Planning*
HSSI	*High Speed Serial Interface*
ISP	*Internet Service Provider*
ITU	*International Telecommunications Union*
LAN	*Local Area Network*
LEC	*Local Exchange Carrier*
MAC	*Media Access Control*
MAN	*Metropolitan Area Network*
MDU	*Multiple Dwelling Unit*
MLPPP	*Multi Layer Point-to-Point Protocol*
MPLS	*Multiple Protocol Label Switching.* *See "MPLS in Metro IP Networks,"* *http://www.riverstonenet.com/technology/mpls.shtml*
MTU	*Multiple Tenant Unit*
OC-3/OC-12	*Optical Carrier 3/12 (155 Mbps/622 Mbps)*
PDH	*Plesiochronous Digital Hierarchy*
POS	*Packet over SONET*
PPP	*Point-to-Point Protocol*
PVC	*Private Virtual Circuit*
QoS	*Quality of Service*
RED	*Random Early Discard*
SONET	*Synchronous Optical NETwork* *See http://www.techguide.com/comm/sec_html/sonet.shtml*
SLA	*Service Level Agreement*
SPE	*Synchronous Payload Envelope*
SRP	*Spatial Reuse Protocol* *See RFC 2892*
T1	*Trunk 1 (1.544 Mbps)*
TCP/IP	*Transport Control Protocol/Internet Protocol*
TDM	*Time Division Multiplexing*
UBR	*Undefined Bit Rate*
VBR	*Variable Bit Rate*
VLAN	*Virtual LAN*
VoD	*Video on Demand*
WAN	*Wide Area Network*
WDM	*Wave Division Multiplexing*
WRED	*Weighted Random Early Discard*

Riverstone Networks, Inc.
5200 Great America Parkway, Santa Clara, CA 95054 USA

408 / 878-6500 *or* ***www.riverstonenet.com***

Printed in the USA

v 1.1 1/01