White Paper

Optimizing Multiplay with Distributed IP Routing

Alan Sardella

Scott Shoaf

Juniper Networks, Inc.
1194 North Mathilda Avenue
Sunnyvale, CA 94089 USA
408 745 2000 or 888 JUNIPER
www.juniper.net

Part Number: 200132-002 May 2006

Contents

Executive Summary

With the advent of advanced IP services (IPTV, VoD, gaming, telephony, security) in broadband architectures, fundamental networking principals as to how to best deliver those services have come to the forefront. As best practices for implementing these advanced IP services continue to evolve, options for the physical placement of IP application nodes in the network are not yet well defined. Application nodes such as web farms, email servers, or download sites have been traditionally positioned globally, nationally, or regionally based on mass-market legacy access to Internet content. As the industry shifts to all applications becoming IP and the applications offered becoming an integral part of the service providers portfolio, these legacy models for centralized application node placement continue to shift to more distributed implementations placing content closer to the user. Examples of this type of distributed placement may be due to the high bandwidth consumption of unicast IP video, or the placement of IP voice elements in distributed locations based on existing voice best practices or overlay requirements.

This is a case where Layer 3 routing, with its ability to organize the network hierarchically, and to find the best path to route subscriber traffic based on its destination IP address, offers a great advantage over Layer 2 forwarding. Application node placement in a Layer 3 routing model becomes trivial. Fundamental IP principles apply so that new content locations, distributed or centralized, are placed into the network and their subnet pushed into IP routing tables. The subnet information is then distributed between routing devices using standard routing protocols. Even the ability to build multicast trees with dual-homed routers can be based on the best IP path towards the source using the calculated routing information.

Attempts to roll out intelligent Layer 2 access networks with advanced service delivery have shown the limitations of a Layer 2 switched distribution architecture. The network becomes restricted to pure backhaul models representative of legacy TDM and ATM networks. If new service models or network utilization dictate distributed placement of IP application nodes, the network can no longer leverage the hierarchical features inherent to IP networks. Instead, packet forwarding is limited to MAC table lookups to find locally attached devices that are part of a flat Layer 2 domain. These types of networks have traditionally been proven to have scaling and flexibility limitations.

Subscriber management, policy, and Quality of Service (QoS) are also major components of broadband access, and the IP edge needs to include these capabilities along with basic packet forwarding. As an alternative to an integrated edge router (with both IP routing and subscriber management), Layer 2 switches attempt to mimic this capability by dividing the two required functions into a service aggregator and a service router. The idea is that the aggregator handles the subscriber management and the router handles the L3 routing requirements. This results in a dual node "faux BRAS" architecture.

The aggregator in this "faux BRAS" model might have some IP intelligence for multicast replication, policy, and QoS requirements, but does not terminate subscriber sessions or provide IP packet forwarding of subscriber traffic. The aggregator essentially acts as an intelligent multiplexer providing limited subscriber functions downstream towards the home and backhauling all subscriber traffic upstream to the services router.

To offer the benefits of flexible IP application node placement and intelligent edge services, a Broadband Services Router (BSR) that provides both IP routing and subscriber management capabilities becomes a fundamental element in a broadband services edge network. The Juniper Networks E-series product family provides the needed BSR functions to deliver a broad array of revenue generating IP applications to broadband subscribers.

Introduction

Prior to the rapid growth of broadband Internet access, service providers were using ATM as the de facto technology for transporting traffic between locations. This traffic consisted of point-to-point logical connections, or *virtual circuits*, predominately replacing legacy point-to-point physical connections such as T1 and DS3 circuits or dark fiber. The ability to multiplex various traffic types over a single network infrastructure lead to cost savings in terms of capital and operational expenditures.

As Internet access technologies in the mass market migrated from dial access to broadband, the common decision was made to leverage existing Points of Presence (POPs) that had been constructed as the entry point into the Internet network. The dial RAS systems were now complemented with "Broadband RAS" or BRAS products to terminate sessions from broadband subscribers. The BRAS could utilize PPPoE or DHCP for subscriber management functions. The data path between the DSLAM (in the CO) and the BRAS (in the PoP) leveraged the existing ATM networks that had become widely deployed in service provider networks. This type of backhaul model is shown below in *Figure 1.*

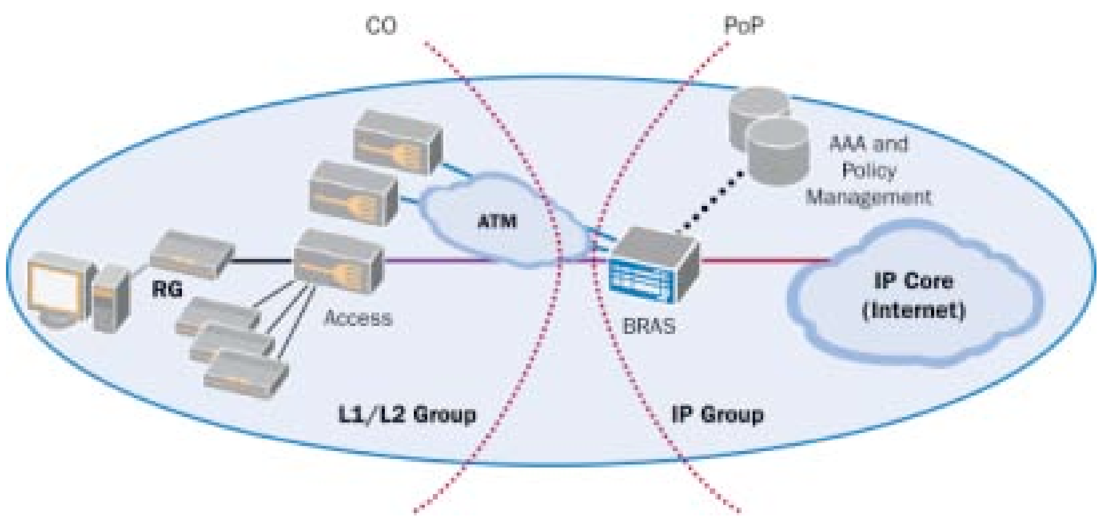

Figure 1: Early Broadband ATM Backhaul Network

Industry best practices involved the clean separation of the Layer 1/Layer 2 planes (DSLAM and ATM) and the IP routing group (a BRAS front-ending the Internet). This simple backhaul solution was sufficient for best-effort Internet applications but did not allow service providers to become part of the value chain and offer enhanced content and services to their broadband subscribers. When competition arising from cable operators led to price pressures as a basic "plumbing company," service providers (through the aid of the DSL Forum) established the TR-58 and TR-59 specifications for a value-add network solution. The focus was to build a smarter network that could leverage ASP relationships to bring richer offerings to broadband customers. These solutions also aimed to offload ATM networks that were rapidly becoming congested due to the rapid growth of broadband, easily outpacing the ability to increase capacity in the backhaul network.

With the current generation of all communications and entertainment applications migrating to IP as the network layer, the service providers have become challenged with no longer isolating IP as an Internet service and are now building a robust IP-capable network to handle a full array of services including video, voice, gaming, audio, and traditional Internet services. Instead of backhauling to Internet PoPs, the subscriber access point into the IP network must be moved closer to the consumer. This results in the IP network becoming more distributed and using standard routing technologies to easily move traffic around the network.

A more robust broadband service edge (*Figure 2*) provides a full suite of IP services to subscribers. In this type of implementation, the access network becomes focused on packet forwarding with limited IP functionality, the exception typically being IP multicast optimization. This allows the service provider to build a common IP edge that supports multiple access technologies such as xDSL and PON without impacting the marketed portfolio to consumers.

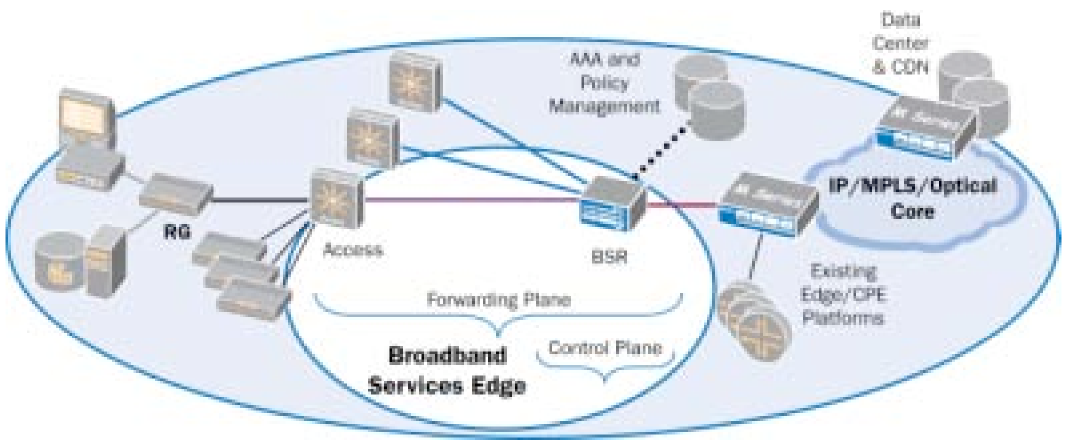

Figure 2: Broadband Services Edge

This solution relies on a third-generation Broadband Services Router (BSR). A BSR is a multi-service broadband router with large bandwidth capacities and advanced QoS capabilities, making it the ideal platform for delivering multi-media broadband services such as IPTV, Video on Demand, and IP telephony to a large number of residential customers, while supporting the most stringent SLAs for business customers. Providing IP routing functionality while acting as the aggregation point for customer access into the IP domain, the BSR allows authorized users to gain access to the bandwidth, QoS, services and features for which they have subscribed. As such, it is the key place in the network to control, deliver and account for customer services.

Interestingly, the system view of this architecture resembles a carrier grade router, with its clean separation of the control and forwarding planes. This model is built on the TR-59 and WT-101 architectures from the DSL forum.[1] The BSR acting as the edge control plane can provide a more centralized point in the network for policy control, lawful intercept, troubleshooting, accounting, and subscriber access. The forwarding plane elements that may include OLTs, DSLAMs, aggregation shelves or switches, and any other access technology are rapidly deployed with limited configuration and lowest cost providers in mind.

The BSR uses a single customer VLAN (C-VLAN) for connectivity to the subscriber. With this architecture (*Figure 3*), the service provider has the operational advantage of being able to pre-provision the customer VLAN and the IP interfaces. This C-VLAN becomes the subscriber's entry point into the IP network with the access elements creating user isolation by acting as

[1] For more information on these standards, see the Juniper Networks white paper *Optimizing Broadband Service Delivery: Centralized Subscriber Intelligence and a Consolidated Service Edge* at www.juniper.net. See also www.dslforum.org.

simple cross connects.

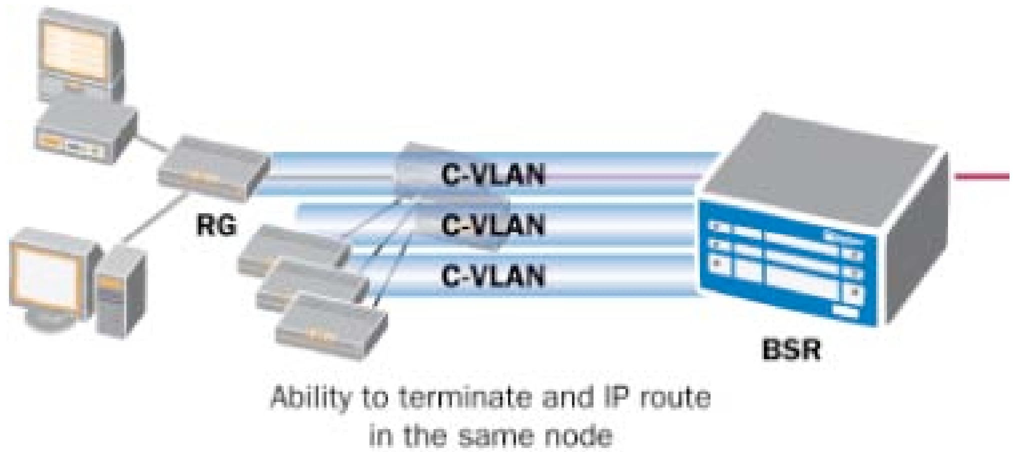

Figure 3: A Unique Service Delivery Point Per Subscriber

The C-VLAN architecture can either support a single IP address per home in situations where the focus is on a homogeneous home network architecture with a routing gateway (RG) capable of NAT and firewall functions or may allow for multiple IP interfaces per home while retaining subscriber isolation.

In addition to the C-VLAN mapped between the BSR and subscriber, the BSR should also have the ability to act in networks utilizing access network multicast optimization. In this situation, the BSR will only send a single replication of each multicast group into the access network based on C-VLAN IGMP state. It is then the responsibility of the access network through the use of IGMP snooping to properly replicate and forward multicast frames to each subscriber.

Alternative Architectures

As networks began to move to Ethernet in lieu of ATM and video became a focal point for IP services, it became apparent that the current Internet model would not suffice, due to the capacity and cost limitations found in the current generation of BRAS products. Instead of waiting for the next generation of IP edge products to alleviate the issue, the industry looked to alternative solutions to rapidly push video out to subscribers in an effort to compete with the cable operators currently offering a triple play solution. The following sections outline attempts to resolve this issue with a heavy focus on video. As service providers look to include a full entertainment and communication offering to their customers, these solutions begin to converge back to the original Broadband Services Edge model.

The first attempt was the IP DSLAM. This solution was operationally challenging in large scale networks with the service control plane pushed too deep in the network with massive numbers of small Fiber to the Node (FTTN) units. As this type of solution falls by the wayside, the next option is to devise ways of leveraging an Ethernet backhaul network to offer multiplay solutions. These Ethernet backhaul models are the focus of this discussion.

The Layer 2 MultiService Edge: One Logical Connection per Service

The first attempt to offer triple services was to utilize the existing concept of point-to-point backhaul networks traditionally used with ATM. By creating "service layers" from the device in the home to the application domains it became trivial to simply backhaul traffic per service by using simple VC or VLAN techniques. The phone would be mapped into the voice domain and the STB mapped into the video domain.

This architecture (*Figure 4*) poses some management difficulties because of the fact that there are multiple circuits per home. The provisioning overhead in the access network became non-trivial. Depending on the customer's service subscription, the number and class of connections into the home would vary. The access network had to be continually modified as subscribers modified their service portfolio.

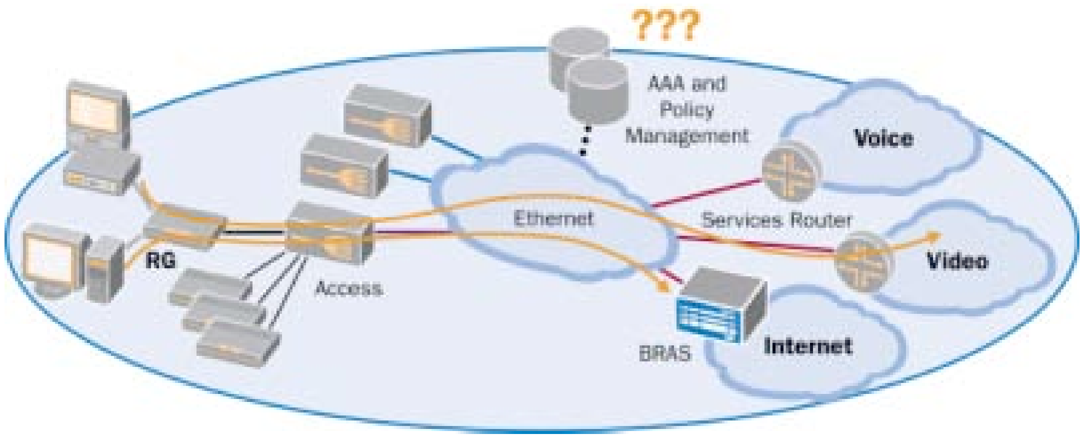

Figure 4: Multiple VC to Home, No Consolidated Service Edge

This model also has no central policy enforcement point or policy management system as found with the Broadband Services Edge; there is no single funnel point to handle QoS and no dynamic policy management. This is antithetical to the intention of TR-59 and it leads to possible congestion in the distribution network. QoS is attempted by prioritizing by service class (e.g., the 802.1p bit marking), but this assumes one class per service, which is an incorrect assumption as found in certain video offerings that may support three or more service classes for broadcast TV, video on demand, command and control, and other background operations.

Lastly, this solution also breaks the consumer electronics view of broadband services: there is no creation of a home network. The evolving concepts of multiplay services focus ultimately on the ecosystem in the home as a single entity for communications and entertainment. In the multiservice edge architecture, each device in the home is delegated to a single service function.

Collapsed Layer 2 MultiService Edge: One Device Per Service

The next evolution of the multiservice edge responded primarily to concerns regarding multiple connections per home. The goal was to build a single VC or VLAN per home so that the consumer connection was only modified once; any change to the service portfolio would not impact the access network. It also created a single policy enforcement point in the network for basic policy control.

Instead of making forwarding decisions into the proper service domains based on the Layer 2 circuit information, the access network relied on Ethernet header information. This decision point was typically pulled from the DSLAM and placed into "Smart Ethernet Switches" that could make forwarding decisions using non-conventional source MAC address or Ethertype policy rules.

The collapsed multi-edge architecture (*Figure 5*), because of its Layer 2 forwarding techniques, relied heavily on the visibility to each device in the home, and each device was still delegated to a single service offering. The Set Top Box (STB) MAC address was inspected and passed to the video domain, the phone MAC address inspected and frames pushed into the voice domain. In this model, Internet support is handed off to the BRAS by either using a PPPoE Ethertype or all other MAC addresses not matching specific service domains. This coarsely patched architecture creates the concept of service routers and service aggregators. Although policy is provided in a single edge node, the routing terminations for each device are pushed into each domain without a single routing edge point.

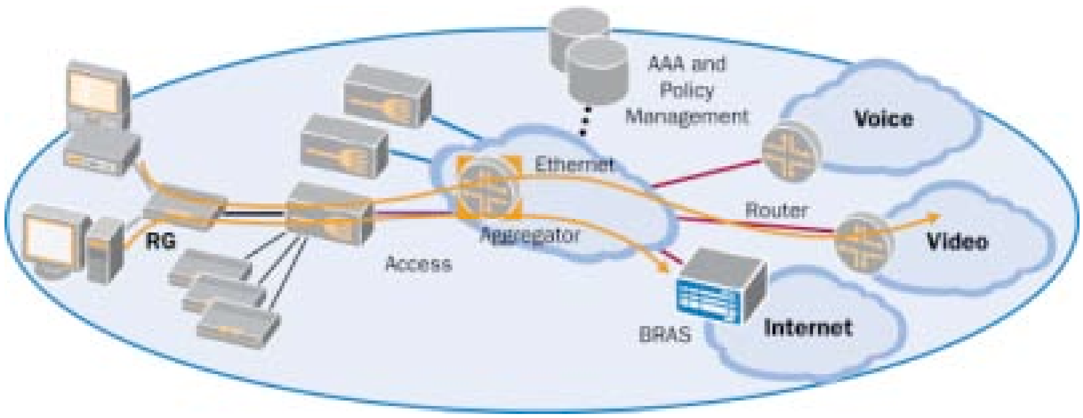

Figure 5: Collapsed Multi-Edge Architecture

The aggregator is essentially an intelligent Ethernet switch with Layer 3 inspection capabilities, but does not utilize Layer 3 forwarding. Unable to perform all the duties of a proper BSR, the aggregator offloads all routing functions to each services domain router. The routing functions typically include DHCP relay, subnet allocations and peer to peer routing. The ambiguity is that you don't know where the true Policy Enforcement Point (PEP) is located, since policy and routing are split between multiple network elements and domains. This can lead to great difficulties in traffic management and QoS control.

Layer 2 Consolidated Edge: Distribution of Policy and Routing

In an effort to support a single MAC address per home (the routing gateway) and a single entry point into the IP network for all services, the next architecture evolution collapses the multiservice edge into a single services router. This solution (*Figure 6*) evolves the roles of the services aggregator and the services router.

The aggregator acts as the role of "policy box" to terminate each C-VLAN and provide QoS and policy per VLAN using Layer 3 inspection. Instead of MAC forwarding per device, the Ethernet frames are forwarded upstream to the router and sent into the IP domain. This is a dual node solution. The aggregator (usually in the CO) acts as the PEP on a per-subscriber basis emulating most of the policy and QoS functions found in a true BSR, and the router is located in the PoP. The assumption is that the services will be available "behind" the router.

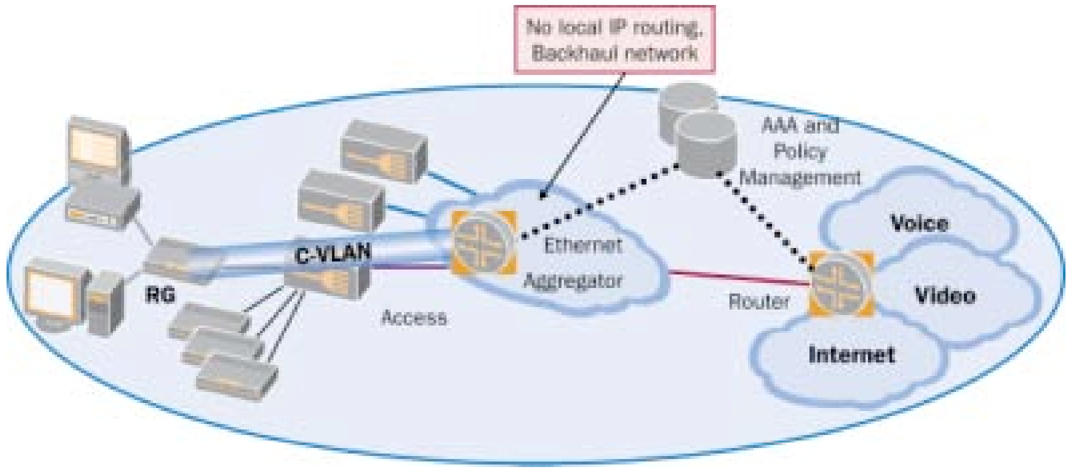

Figure 6: Attempting to Mimic a Consolidated Edge with a Dual-Node Solution

This solution addresses the concerns regarding a single VC or VLAN per home coupled with the need to have a home network. By distributing the subscriber management functions of policy and routing between the two nodes, a "faux BRAS" network is constructed.

The missing element of this architecture is the ability to provide local routing in the aggregator. Depending on the network topology and service endpoints, it is possible that the video servers or voice Session Border Controllers may be collocated with the aggregator. In order to prevent traffic unnecessarily going to the PoP and then back to the CO (where the service resides), the aggregator is now required to provide Layer 3 routing functionality to forward packets not based on destination MAC addresses, but IP subnet information.

This leads to the eventual need to combine the aggregator and router into a single device, which is discussed in the following section.

Layer 3 Consolidated Edge: Combined Policy and Routing

To provide more IP service and routing flexibility, the industry will most likely converge the functionality of the aggregator and router into a single platform. This logical next evolutionary step of the broadband architectures will result in the creation of the Broadband Services Edge: a full-function BSR with a combination of routing and policy. This type of capability is demonstrated today with the Juniper E-series platforms capable of BSR functionality with up to 320G of forwarding capacity.

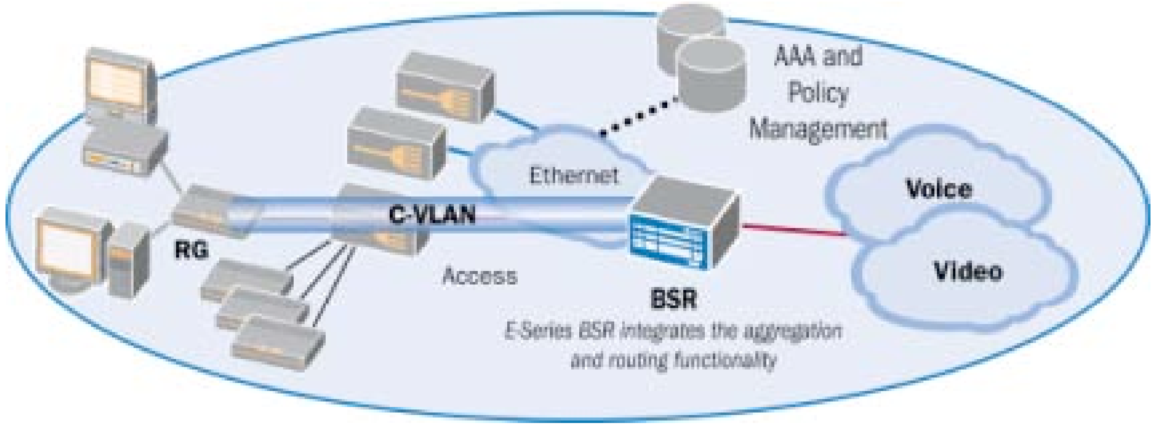

Figure 7: E-Series -- Consolidated Service Edge with Centralized Intelligence

Comparing a True BSR to Dual-Node Systems

As networks look to compare the requirement for a fully functional BSR to multi-node solutions, it becomes helpful to fully understand these platforms and how they either provide (in the case of the true BSR) or emulate (in the case of the dual node system) BRAS functionality.

The Single Platform BSR: Policy and Routing

The creation of a single platform for aggregation, routing and policy requires a purpose-built hardware platform, coupled with software features that leverage the underlying hardware for both subscriber management functions and general IP routing. The logical diagram of the internal workings of the E-Series is shown in *Figure 8*.

Each C-VLAN is terminated in the BSR and mapped to a unique IP interface. Each IP interface is a routable interface with spoofing prevention based on DHCP and PPPoE transactions. The assigned subscriber host route is placed into the routing tabled only after AAA negotiations are complete. Source IP and MAC address validation is used to prevent unauthorized addresses from being sourced from a particular C-VLAN.

The IP interfaces are all contained in the same subnet space based on the mapping to a logical unnumbered interface. This interface subnet is placed into the routing table and propagated through the network using standard IGP or BGP routing protocols. Conversely, as a full fledged routing element, the upstream peer routers can send routing information to the BSR allowing it to make local forwarding decisions based on the destination subnet. The destination subnets may be local to the BSR or in remote locations. Since standard routing principles are used, the subnet only dictates the next hop or egress interface for packet forwarding.

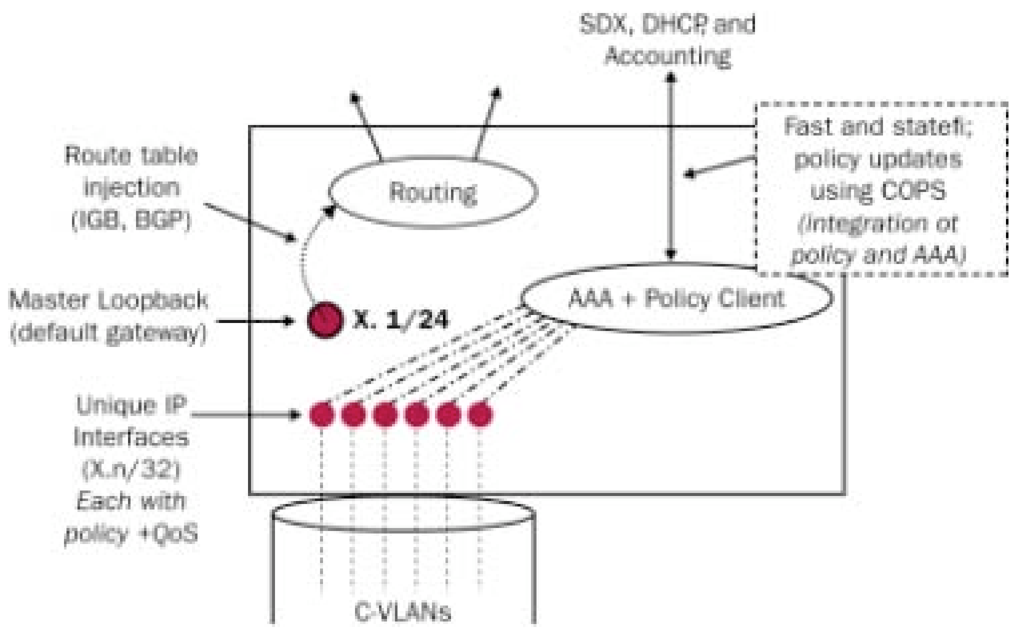

Figure 8: E-Series Contains Both Routing and Policy Management

Since the BSR is the Policy Enforcement Point (PEP) in the network, it must contain not only EMS and AAA interfaces for policy control but also be enhanced for real-time policy changes that are becoming required in the next generation of broadband offerings. These types of policy systems typically utilize Common Open Policy Service (COPS) for rapid and stateful transactions between the policy server and PEP. As policy changes are required (including QoS requirements) the policy server will push policy requests over the COPS interface. These transactions are acknowledged providing for stateful feedback since many of the new generation policy requirements are related to granular and reliable accounting data. Simple SNMP-based or CLI-based policy updates are not deemed acceptable, since they provide little or no feedback mechanisms to ensure and validate policy attachments and have scaling limitations in real-time transaction networks.

The BSR subscriber management system is designed to integrate login, IP address assignment, routing, policy, and QoS at the network edge. This provides operational simplicity for trouble resolution and policy control coupled with the flexibility of local IP routing.

The Emulated BSR: Dual Node BRAS Systems

Basic IP routers or switches are not built with subscriber management functionality in mind. They lack the host routing capabilities to dynamically map a unique IP address per interface. and tend to have very slow policy handling as there is no fast and stateful policy engine; this is rather like a static Element Management System (EMS).

Essentially, a dual node solution (*Figure 9*) is required to loosely mirror the functionality of a single BSR, creating a kind of “faux BRAS” solution. An aggregator is used for per-subscriber policy and QoS while a router node provides Layer 3 functionality. This results in a distributed IP control plane between the two nodes increasing operational overhead in the network.

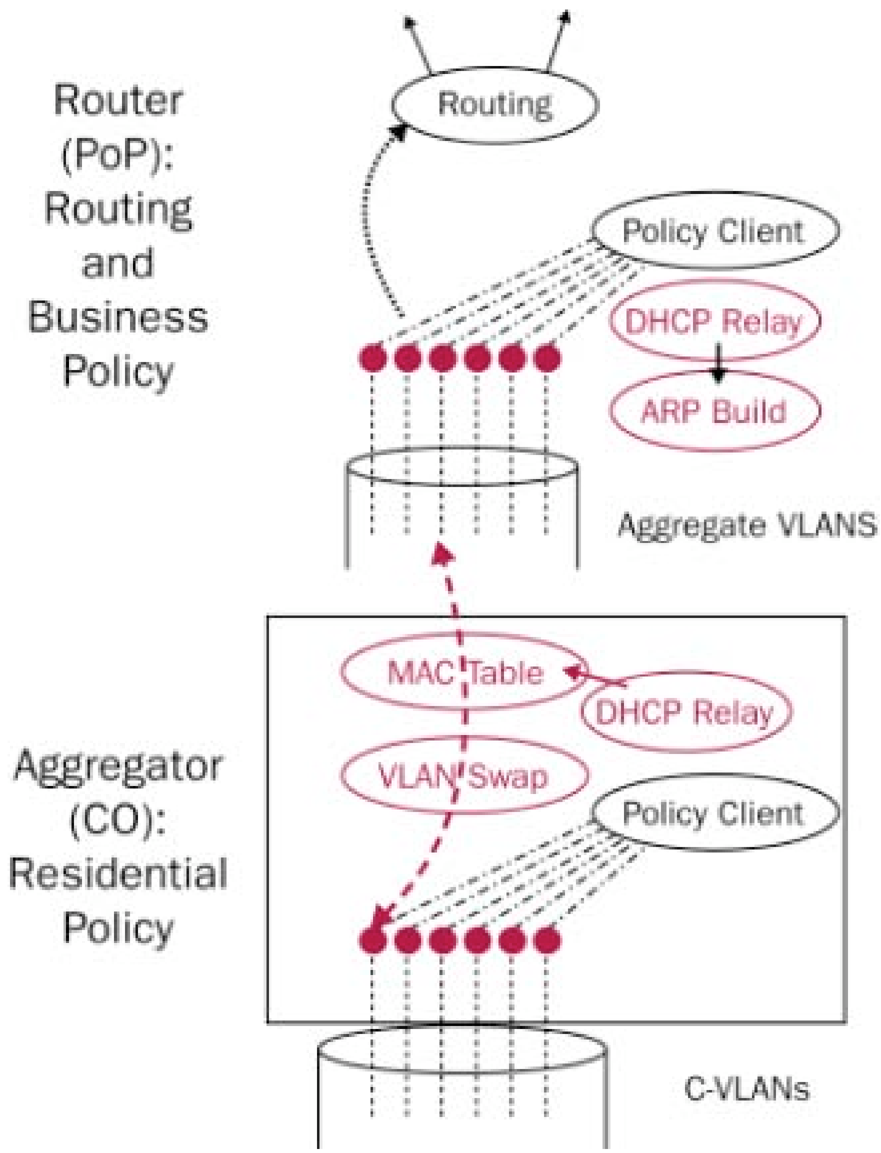

Figure 9: Two Box Solution as Faux BRAS (Dual Node with Aggregator and Router)

This paradigm has a number of disadvantages. There is a separation of routing and policy, resulting in a need to create and share subscriber state between the two nodes for proper Layer 2 and Layer 3 forwarding of packets.

One weakness of the faux BRAS system comes to light when local routing is needed, and Layer 2 backhaul results in unnecessary congestion in the access network. With pure Layer 2 forwarding in the access network, there are no IP lookup tables built (except perhaps for LSP path construction, if MPLS is being used in the metro). This is rather like PNNI in an ATM network, which is used for path construction and not packet forwarding. This model is built to bring in an Ethernet frame, do VLAN swapping (and perhaps VPLS label wrapping), and push upstream.

The faux BSR architecture is typically built around ARP and MAC entries populated in the forwarding nodes based purely on DHCP transactions. This state information must remain synchronized between the aggregator and router for proper network functionality. Only the router uses IP level information to reach the end subscriber while the aggregator is limited to standard MAC forwarding. As a result, there is no mechanism for IP application nodes (video server, session border control, etc.) local to the services aggregator to be placed into content specific subnets for local routing.

This type of dual node system with segmented routing and policy must continually evolve its capabilities to mimic the well-known functionality found in a single IP edge platform. Extensive packet snooping, proxies, policy forwarding and other mechanisms must be configured and managed to override the lack an integrated IP routing capability in the policy edge system.

The E-Series Service Built Edge Portfolio

The E-Series service-built edge portfolio, including the new E-320, is shown in the following figure.

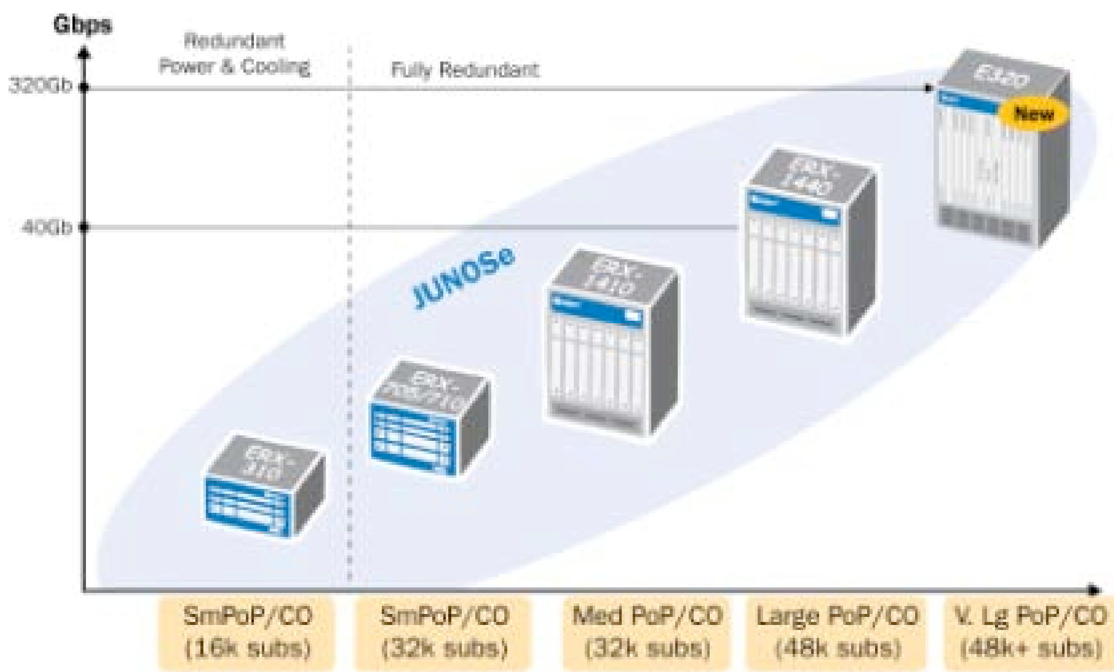

Figure 10: E-Series Service Built Edge Portfolio

Juniper Network's E320 is the first BSR designed specifically to deliver Multiplay services to both corporate and residential subscribers. The E320 enables service providers to differentiate between their service offerings and those of their competitors through a unique blend of both hardware and software features that combine with enhanced QoS capabilities and substantial system capacity to enable both bandwidth scalability and subscriber density.

Conclusion

Ideally, a single IP edge architecture should be used for all IP application services to provide proper routing, QoS, and policy control within the network. The premier architecture to achieve this capability is centered on the E-series family as the Broadband Services Router (BSR). It provides multicast routing, subscriber management, robust quality of service, policy management, security, and operational ease all in a single platform.

Although alternative architectures that are perceived as optimal and more efficient are being investigated, they may in fact be short lived and only part of the evolutionary tract that continues to lead from a traditional L2 multiservice edge network to a single IP edge consisting of edge nodes containing both policy and routing. These “faux BRAS” or “BRAS-less” architectures suffer from the hard fact that the combination of IP routing, policy, and advanced subscribers services as a fundamental service offering requirement will lead to back to a single IP edge system.