RIVERSTONE NETWORKS
ADVANCED TECHNICAL PAPER SERIES

The Importance of Inbound Rate Limiting

Tony Fallows, Riverstone Networks

ABSTRACT

This paper discusses Rate Limiting in general and Inbound vs. Outbound Rate Limiting in particular. The differences between the two are discussed along with certain advantages of Inbound Rate Limiting. In addition, general specifications of a Riverstone solution that was presented to a customer to solve real issues are shown as a concrete example of a real world implementation of Inbound Rate Limiting in a competitive business environment.

Riverstone Networks, Inc.
5200 Great America Pkwy, Santa Clara, CA 95054 USA
(877) 778-9595, (408) 878-6500, www.riverstonenet.com

Version 1.0, 5 December 2001

Introduction

This short paper is an adaptation of a paper presented to a Riverstone customer. The customer's name has been replaced by "GRATE_Provider" to protect their identity.

The requirement, the solution and the synopsis serve as generic examples.

The Requirement

GRATE_Provider has provided the following requirement.

> "….outbound rate limiting, the scenario is that a number (1 or more) of trunk interfaces connect the switch to our national core network, and a number of trunk interfaces connect it to our international core network. We have a service where we deliver a certain national and a certain international bandwidth to our customers (e.g. 10 Mbit/s national + 5 Mbit/s international). In addition to this we will deliver different traffic classes to the customer, e.g. 5 Mbit/s "gold" traffic, 5 Mbit/s "silver" and the rest best effort. The traffic classification may be done on type of traffic, source or destination address etc."

Synopsis

Riverstone can provide this functionality using in-bound rate limiting and the BURST-SAFE feature.

Outbound rate limiting will:

- result in customers' traffic affecting other customers in times of congestion.
- not allow GRATE_Provider to make the guarantees required.
- not scale well resulting in a requirement to deploy additional switches or additional trunk interfaces, if additional trunk interfaces are added then very complex routing is required.

Inbound rate limiting will:

- provide the necessary isolation of customers' traffic.
- allow GRATE_Provider to commit to the bandwidth guarantees and SLAs.
- allow a two or three tiered model to be deployed as stated in the requirements as well as other exceed action options.
- scale well and allow simplified routing and provisioning

Outbound Rate Limiting Discussion

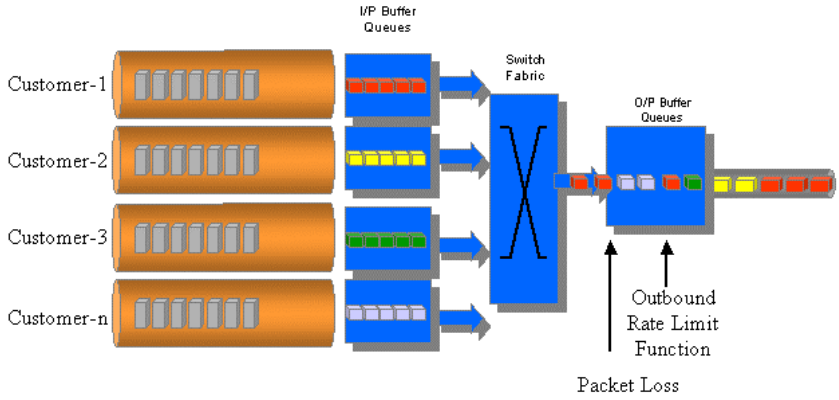

Figure-1 – Outbound Rate Limiting Architecture

Consider the scenario where N (100Mbps or GbE) customers on ACCESS links are being switched to M GbE TRUNKS, typically M will be 1 or 2 and N will be a much larger figure. This represents the GRATE_Provider solution, i.e. the ratio of ACCESS:TRUNKS is MANY:FEW.

The traffic from all N customer access links must pass across the switch fabric before being rate limited. This allows all N customers to burst up to the wirespeed across the fabric. The problem is that the outbound buffer queue will be oversubscribed and packet loss will occur. This will be a common event as data traffic is bursty in nature and all customers could burst to the wirespeed. It would only take 10*100Mbps-connected customers bursting simultaneously to oversubscribe the GbE trunk. The important aspect here is that the outbound buffer queue is the point at which outbound rate limiting occurs and because the output is over subscribed then traffic cannot be rate limited. The result is that the 11th customer would not be able to transmit any data. This results in customers' traffic affecting one another and does not allow GRATE_Provider to make bandwidth guarantees.

Further to this, if the outbound queue only supports X number of policies then the number of customers that can be supported on the switch is limited to X. This is a major scaling factor with outbound queuing. A possible way to work around this is to add more trunk interfaces and then map X customers to each trunk port. This is generally not a feasible solution as complex policy routing would be need on a per customer basis – i.e. how do you get customers 1-16 to traverse trunk-1 and customers 17-32 to traverse trunk-2? Provisioning becomes very complex.

Typically, when the rate limiting is performed at the output the switch has very little time to perform any functions on bandwidth levels that exceed the given rate.

As a result a simple drop action is normally specified. The ability to re-prioritize traffic, mark the IP-TOS field or provide multiple tiers of rate limit policy are generally not available for output rate models. These functions provide incremental revenue opportunity for the Service Provider.

Inbound Rate Limiting Discussion

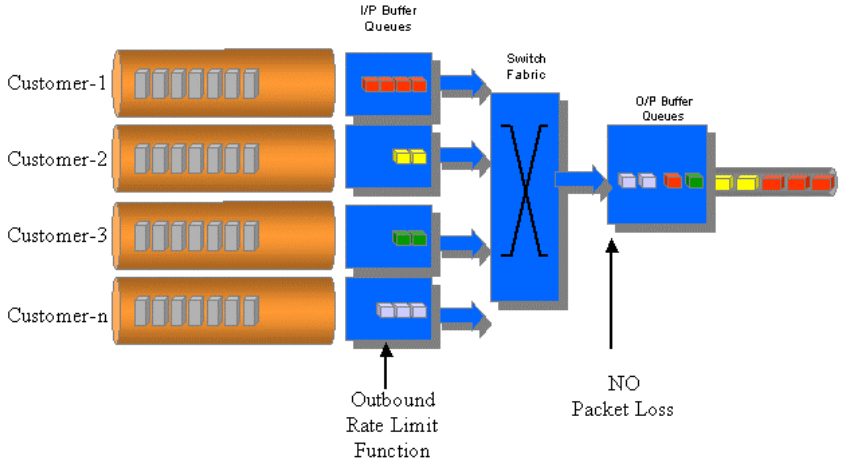

Figure-2 –Inbound Rate Limiting Architecture

Consider the same scenario where N (100Mbps or GBE) customers on ACCESS links are being switched to M GbE TRUNKS.

The traffic from all N customers' access links is now rate limited at the input and unnecessary traffic does not traverse the switch fabric. The output port is not oversubscribed as long as the sum of all the assigned rate limits does not exceed the physical bandwidth of the trunk. The bursty nature of the customers' traffic does not affect other customers and now guarantees can be made.

If every port is able to be rate limited at the input then the rate limiting policies are being distributed and are more scalable. For example, if a rate limit could be applied to every port then the limit of the number of rate limit policies is not determined by the output trunk, but is now a function of how many ports you have in a switch, i.e. potentially hundreds. Also there is no need to tie customers to using certain trunk ports. Multiple trunk ports can be used for redundancy purposes. Again, this means the entire solution is more scalable.

As the rate limiting is performed at the input of the switch, the switch can now provide more sophisticated exceed policies, such æs drop, re-prioritize, TOS marking and provide multiple tiers of services.

Output Rate Limiting "Masquerading" as Input Rate Limiting

Having established the benefits of inbound rate limiting careful consideration should be given to the implementation of inbound rate limiting. Some vendors have implemented only outbound queue manipulation in their switch architecture and have no inherent inbound rate limiting. These vendors have therefore manipulated the outbound architecture to look like an inbound rate limit. This is extremely bad practice as it results in all the same problems associated with outbound rate limiting. Here is how they achieve this.

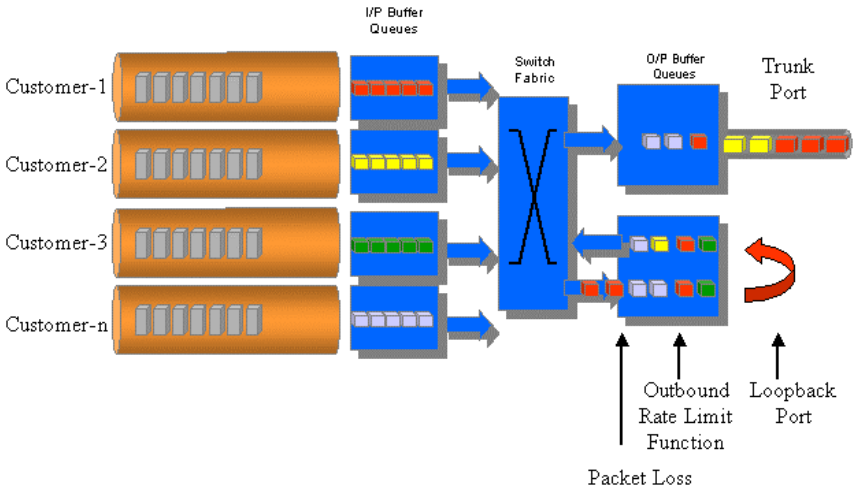

Figure-3 –Outbound Rate Limiting Masquerading as Inbound Rate Limiting

As can be seen from the diagram, traffic is sent to a port that loops traffic back internally and then to the trunk port. Aside from the obvious inefficiencies of this implementation over subscription will be seen in the same way as detailed above.

The obvious aspect to look for with such a switch implementation is the need to put a port into loopback in order to get inbound rate limiting. Typically this port can no longer be used for normal traffic.

The Riverstone Implementation

The subtleties of rate limiting make a difference between a successful implementation and a poor implementation. The ability to offer these services and meet SLA's is dependent upon a solid implementation.

The following provides an overview of the Riverstone architecture, which offers a more scalable, flexible and robust operation:

- Wirespeed input streams with L3 and L4 classification of traffic for queuing and rate limiting.
- Class of Service mapping for all classified traffic.
- Inbound rate limiting based on port – scales to 1 per port.
- Inbound rate limiting based on Aggregate (ACL) policy – scales to 24 per line card (i.e. 16 port 10/100 line cards with one customer per port provides at least one policy per customer).
- Rate limiting from 1Kbps to 1GbE in byte increments.
- Multiple exceed actions, including drop packets, re-prioritize packets, TOS/Diff-Serve marking.
- BURST-SAFE feature providing a dual action rate limit to allow two rate thresholds with specific exceed actions per threshold, e.g. an initial Committed Access Rate (CAR) rate limit of 5 Mbps with an exceed action to lower priority, coupled with a BURST rate limit of 5 Mbps with an exceed action to drop or further lower the priority.
- Outbound queues managed by Weighted Random Early Discard (WRED) for congestion avoidance.
- Outbound port rate limiting.
- Queuing mechanisms that include Strict Priority and Weighted Fair, selectable on a per port basis.

At the time of writing the “Generally Available” line cards from Riverstone support 24 rate-limit “buckets”, this gives the aforementioned 24 rate limits per card. The soon to be released V4 ASIC cards support 1024 rate limit buckets on the RS8x00 and 2048 on the RS38000. This further increases the scalability of the solution.

The GRATE_Provider Proposed Solution

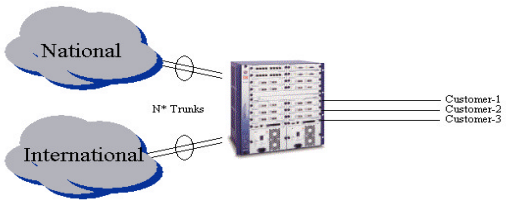

Figure-4 –Physical Connections

Rate Limiting/QoS Operation

Consider the first scenario, i.e. 10 Mbps for National and 5 Mbps for International:

1. Classification could be based on SIP, DIP, TOS, SP, DP or combinations of these parameters.
2. For the GRATE_Provider example, classification would be based on the customer subnet and the destination address. As GRATE_Provider own the National network address space (and hence the IP address range is known) then it is relatively easy to provision a range of addresses corresponding to the National network. The remaining addresses must therefore be destinations on the International network.
3. Rate limiting would be AGGREGATE inbound rate limiting.

So the following is possible:

```
Customer 1 – source subnet 1.2.3.0/24 to destination-range <national_network> - 10Mbps
Customer 1 – source subnet 1.2.3.0/24 to destination-range <any_thing_else> - 5Mbps

Customer 2 – source subnet 4.5.6.0/24 to destination-range <national_network> - 50Mbps
Customer 2 – source subnet 4.5.6.0/24 to destination-range <any_thing_else> - 20Mbps

Customer 3 – source subnet 7.8.9.0/24 to destination-range <national_network> - 34Mbps
Customer 3 – source subnet 7.8.9.0/24 to destination-range <any_thing_else> - 2.048Mbps
```

Benefits

Every customer can have customized rate limits – allowing GRATE_Provider great flexibility in service definition.

This is a scalable and robust solution as the rate limiting is inbound. Service guarantees are possible.

Assuming two GbE ports to the National and two GbE ports to the International network then the RS8600 can scale to accommodate 192 customers. The RS38000 could accommodate over 400 customers. These figures depend on the degree of trunk redundancy required. If an outbound rate limit model is used this would require 192 or 400 policies on the outbound port – Riverstone doesn't believe this is feasible.

If a route change occurs, and International routes are visible from the National network then the rate limit policy still applies as it is inbound. Consequently inbound rate limiting is independent of routes. Using outbound rate limit policies are tied to an outbound port and so are dependent on routes.

Consider the second scenario, i.e. different traffic classes to the customer, 5 Mbit/s "gold" traffic, 5 Mbit/s "silver" and the rest best effort.

1. Classification could be based on SIP, DIP, TOS, SP, DP or combinations of these parameters.
2. For the GRATE_Provider requirement, classification would be performed in the same manner as described for the first scenario.
3. Rate limiting would be AGGREGATE inbound rate limiting with dual policy. All traffic would be classified in the high queue, the first exceed action would be to re-prioritize and the second action would be to re-prioritize.

Using the Riverstone RS8600, with every customer having a two level rate-limit policy then 144 customers could be supported. The RS38000 could support over 400.

Riverstone Networks, Inc.
5200 Great America Pkwy, Santa Clara, CA 95054 USA
(877) 778-9595, (408) 878-6500, www.riverstonenet.com

Version 1.0, 5 December 2001