Dynamic IP Traffic Management using Dyband

1 Introduction

Network Managers faced with the challenge of cutting costs while still improving performance are turning to WAN optimization tools.

Service providers' pipes are strained to capacity, hard-pressed to handle subscribers' insatiable demand for bandwidth. Media-rich content, driven by broadband access technologies, is compounding the congestion problem.

Additional bandwidth alone will not solve today's congestion problems. It circumvents the issue without providing a cure. It is cost-prohibitive to engineer networks to handle peak loads that only occur for a few hours a day. It requires upgrades to network equipment, adds greater network complexity and cost, yet delivers negative returns.

Network operators recognize that no matter how much bandwidth is added, networks still face congestion problems. Dyband is a unique, software-based IP traffic management system that meets this challenge, providing:

- Real-time, intelligent control over IP traffic, managing traffic by IP address
- Unlimited service levels
- Insight into network performance, both real-time and historical
- A single management interface across all transport mediums
- A scalable system accommodating business growth

Dyband manages IP traffic for individual subscribers and also for aggregate traffic points (subnets, gateways, backbone interfaces, and user-defined groups of traffic points). This approach supports distribution networks ranging from local service providers to national providers serving thousands of subscribers. It also satisfies the needs of subscribers, allowing individual choice in service level and providing consistent, reliable service.

Dyband is a software-based system. The traffic-shaping component must be installed on a dedicated system located between the Internet router and the distribution network (cable, wireline, wireless, satellite), where it shapes all traffic passing through it. The other components, enabling system configuration, performance monitoring, archiving, report generation, and global service profiles, may be located anywhere, on single or multiple systems, as long as there is appropriate IP connectivity.

Dyband maximizes the efficiency of Internet connections by responding dynamically to congestion. This paper focuses on the mechanisms, benefits, and implementation of the dynamic features in Dyband's architecture.

2 Real Time IP Traffic Management

The topology of a distribution network is represented in Dyband as a tree structure. The physical devices and logical entities in the network are the nodes in the tree, each with an assigned bandwidth capacity. Endpoints in the tree represent individual traffic points; nodes higher in the tree represent aggregate traffic points.

Dyband achieves real-time control over IP traffic at all levels of the tree through the following design features.

2.1 Rapid Cycle Time

Every 10 ms, Dyband assesses traffic flow, both inbound and outbound, throughout the topology tree. If the demand for bandwidth exceeds the assigned capacity at any point in the tree, Dyband responds to the congestion in the next 10-ms cycle, using two powerful controls: dynamic rate controls and dynamic priorities, described below.

2.2 Dynamic Rate Controls

When congestion is detected, Dyband reduces transfer rates, but only at the points of congestion and only for the affected traffic direction. As soon as the lowered rates have the desired effect (i.e., congestion is relieved), the rates are allowed to return to their normal limit. Since traffic conditions are reassessed every 10 ms, the rapid toggling between normal and congested rates makes full use of available bandwidth.

In Dyband, rate limits are defined for both normal and congested conditions -- that is, when assigned bandwidth can and cannot meet demand. With rate limits in effect even under normal conditions, subscribers experience consistent service. In systems that allow rates to burst when network traffic is light, user expectations are raised to unrealistic levels, leading to frustration when bandwidth competition increases.

Normal and congested rate limits, along with dynamic priorities (described below), make up the basic parameters in service level policies. Policies are assigned to individual and aggregate traffic points, either manually or via an efficient inheritance mechanism that simplifies the configuration process.

2.3 Dynamic Priorities

In Dyband, access to bandwidth is determined not by static priorities but by dynamic priority ranges, operating as follows. Each service policy defines minimum and maximum priority settings for both inbound and outbound traffic. For a given direction, a user who has been denied bandwidth in one 10-ms cycle will be granted increased priority, up to his assigned maximum, in the next cycle; a user whose transfer request has been serviced will drop to his minimum priority in the next cycle. This rotation system ensures that no user is completely starved of bandwidth during periods of congestion, unless the policies have been deliberately configured to do so (see section 2.3.3).

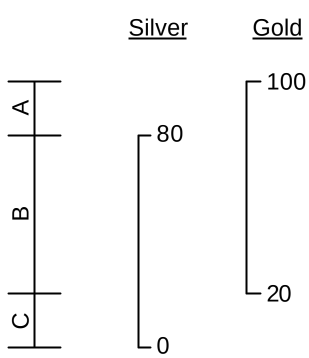
2.3.1 Equal Access

With Dyband, subscribers who have paid for the same priority level will be given equal access to bandwidth. This is achieved simply by assigning identical minimum and maximum values to their policies' priority ranges.

2.3.2 Preferred Access

Dyband gives preferential service to users who have paid for higher priority service (i.e., policies with higher priority settings).

An example of priority settings that will provide preferential service is shown below:



In this example, the size of A indicates the extent to which a Gold level subscriber can have access to bandwidth, to the potential exclusion of a Silver level subscriber, unless there is sufficient bandwidth to service both.

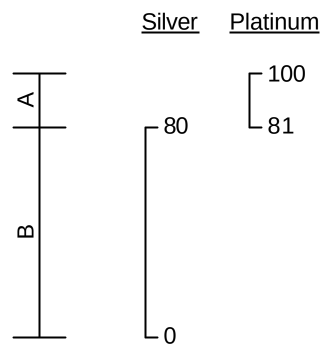
The size of B indicates the extent to which the Gold and Silver level subscribers are prioritized together for access to bandwidth.

The size of C indicates the extent to which a Silver level subscriber must "climb" the priority queue to gain access to bandwidth if Gold level subscribers are requesting all the available bandwidth.

2.3.3 Guaranteed Access

Dyband does not create bandwidth; it provides efficient use of existing bandwidth. It enables service providers to offer guaranteed minimum transfer rates, as long as priority ranges are appropriately defined and there is sufficient bandwidth to accommodate these high-priority users.

An example of priority settings that will allow guaranteed service is shown below:

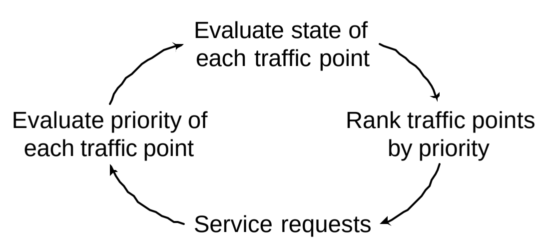


In this example, an active Platinum-level subscriber will always have access to bandwidth, as long as there is sufficient bandwidth to accommodate all active Platinum level subscribers at their guaranteed rate.

The active Silver level subscriber will be serviced only if there is surplus bandwidth after the demand by all active Platinum level subscribers has been met.

2.4 Shaping Cycle

Dynamic rate controls and dynamic priorities come into play during each 10-ms shaping cycle. The four steps comprising a Dyband shaping cycle are described below.



2.4.1 Evaluate Bandwidth State

At the start of a shaping cycle, each traffic point in the Dyband topology tree is classified according to its bandwidth state: Normal or Congested. The determination is based on supply and demand during the previous shaping cycle, as follows:

Normal: when bandwidth capacity was sufficient to meet demand at all the aggregation points above it in the topology tree

Congested: when there was insufficient capacity at any aggregation point above it in the tree

This determination is critical, since it dictates whether the transfer rate for that traffic point will be held to its Normal or Congested Limit in the current cycle.

2.4.2 Rank Traffic Endpoints

All the traffic endpoints (terminal points in the topology tree) are ranked based on their current priority, from highest to lowest. The current priority of each endpoint is the sum of its own priority and the priorities of all the aggregation points above it in the tree.

2.4.3 Service Requests

Once the endpoints have been ranked, their transfer requests are serviced until either the list of endpoints is exhausted or all the interface bandwidth has been consumed. The transfer rate for each endpoint is the lower of:

- The requested transfer rate
- The policy's transfer rate limit, based on its state (normal or congested)
- The highest rate limit supportable by the rate limits of all the aggregation points above it in the tree

2.4.4 Evaluate Priorities

As discussed earlier, for each traffic point whose request is serviced, the priority is set to the minimum value specified in its policy. For each traffic point whose transfer is not serviced, the priority is increased, unless it has already reached the maximum value specified in its policy.

2.5 Controlling Aggressive Users

To prevent aggressive users from causing prolonged periods of congestion, Dyband gives service providers an additional control option: rate ramps. A rate ramp, triggered by sustained high usage, progressively lowers rate limits until usage drops to an acceptable level or until a specified final rate limit is reached. ("Acceptable usage" refers to average transfer rate; burst rate is limited only by the normal and congested limits described above.)

In Dyband, usage is conceived of as a pool that fills and drains as follows:

- The pool fills based on the amount of data transferred at a rate exceeding the acceptable average rate.
- The pool drains based on the amount of data transferred at a rate below the acceptable average rate.

As the pool fills, the transfer rate is increasingly limited; as it drains, the rate is allowed to increase.

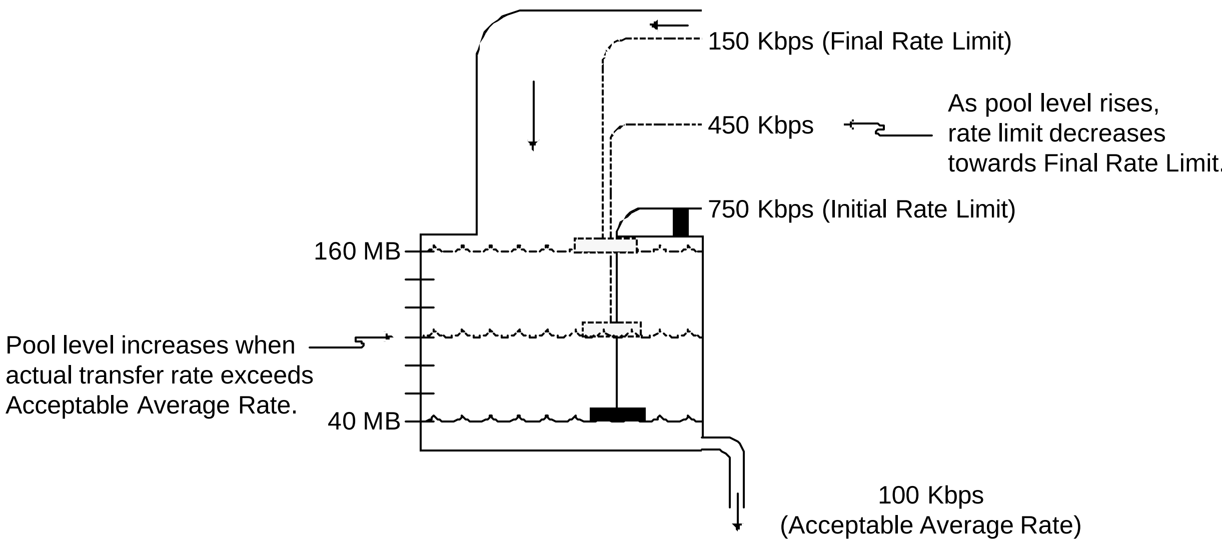
Consider, for example, the rate ramp settings shown in the table below. The rate ramp will be triggered when 40 MB of data transfers at a rate above 100 Kbps and will reach its most restricted level (with a Final Rate Limit of 150 Kbps) when 160 MB of data has transferred at a rate above 100 Kbps.

Ramp Trigger	40 MB
Ramp End	160 MB
Initial Rate Limit	750 Kbps
Final Rate Limit	150 Kbps
Acceptable Avg. Rate	100 Kbps

When the pool is reduced to 40 MB (the initial trigger value), the rate limit will return to the policy's maximum (either normal or congested rate, depending on bandwidth state). Thus, a subscriber who slows or stops transferring data will eventually be able to transfer again at the maximum rate allowed by the policy.

Note: The diagram below illustrates the meter that defines rate limits while the rate ramp is in effect. It does ***not*** illustrate data flow.

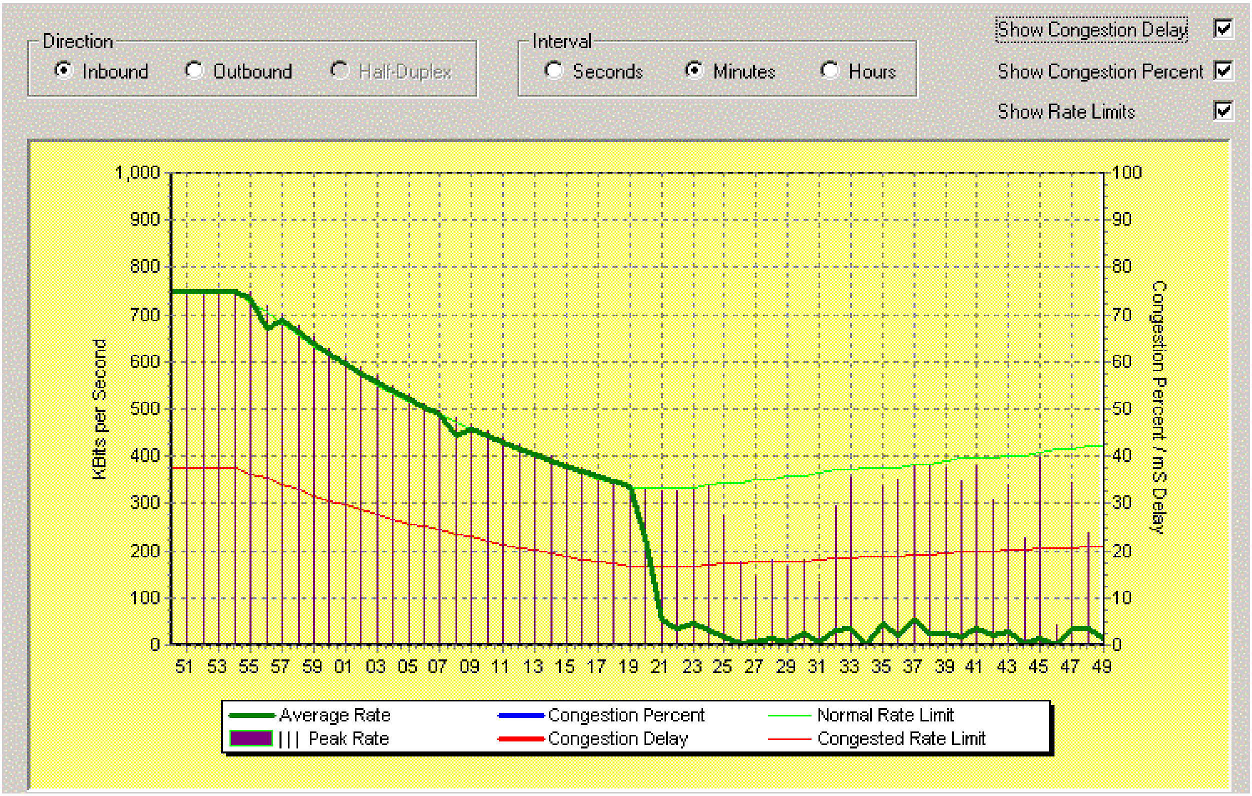
Rate Ramp Meter



The real-time performance graph below, viewable through Dyband's user interface, shows a rate ramp in action. The ramp was triggered at minute 55 after continuous high usage—i.e., the average transfer rate, shown by the thick line, was continuously at the rate limit, shown by the upper thin line (not visible until minute 19). Dyband responded by progressively lowering the rate limit, using the same rate ramp settings used in the example above: Initial Rate Limit of 750 Kbps, Final Rate Limit of 150 Kbps, Acceptable Average Rate of 100 Kbps.

The duration of a rate ramp (seconds, minutes, hours) depends on the settings chosen by the service provider *and* the subscriber's behavior. In this example, the turning point in the rate ramp was reached at minute 19, before the Final Rate Limit of 150 Kbps was ever reached, because at that point, the aggressive behavior ceased. When the subscriber's average transfer rate fell below the Acceptable Average Rate of 100 Kbps, Dyband responded by progressively raising the rate limit. Eventually, barring a recurrence of aggressive behavior, the rate limit would return to the Initial Rate Limit of 750 Kbps.

The graph shows clearly that peak transfer rates (shown as vertical lines) are permitted up to the rate limit without penalty, as long as the average transfer rate remains at an acceptable level.



3 Reducing Packet Loss

A key benefit of Dyband's quick response to surges in network load is a significant reduction in packet loss. The following sections describe how this is achieved.

3.1 Reducing Retransmission Requests

In an access network that is not dynamically managed, a few subscribers may be able to monopolize bandwidth. Subscribers who do not receive fair access may have to request retransmission of data that has been only partially received before a timeout (stall). This is inefficient because the backbone must carry the same data more than once. The duplicated data may saturate the queue at the backbone provider's router (see flow diagram opposite), and packets may have to be dropped.

Dyband's system of dynamic priorities, by providing subscribers with equitable access to bandwidth, achieves a reduction in retransmission requests and thus reduces packet loss at upstream routers.

3.2 Delaying New Requests

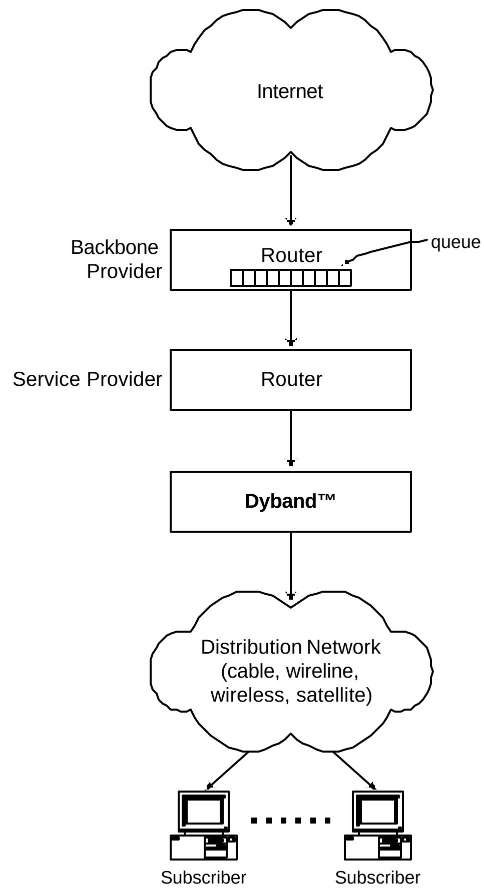
Dyband further reduces packet loss by leveraging the closed loop system of TCP/IP. In TCP/IP, a new request is not issued until a response has been completely received or a timeout occurs. Dyband, using real-time enforcement of configured rate limits, can slow the completion of a given TCP/IP exchange and thus hold off additional traffic that would otherwise be requested and received at the upstream router. In this way, queue saturation and packet loss at the upstream router can be significantly reduced.

3.3 Effects of Reduced Packet Loss

One of the clear benefits of reduced packet loss is the effect on the subscriber's perception of service. Stalls due to packet drops are obvious to users; in contrast, the slowing produced by Dyband's enforcement of rate limits is too brief to be perceptible.

In addition, when packets at upstream routers are dropped, they are dropped randomly. This undermines the ability of the service provider to guarantee preferential service to higher-paying subscribers. By reducing packet loss at upstream routers, Dyband improves service for all subscribers and also gives service providers true control over the service levels that they market.

Inbound Traffic on Simple Access Network



4 Conclusions

For service providers and their subscribers, the dynamic mechanisms at the core of Dyband's IP traffic management system provide unique benefits:

- The most efficient use of network bandwidth
- The most consistent network performance for subscribers
- Control over aggressive users
- Unlimited service levels supporting equal, preferred, and guaranteed access

For further information on Dyband, and how it can benefit your firm, contact us at
sales@dyband.com
or visit us at
www.dyband.com

Dyband Corporation
215 Stafford Road West, Unit 103
Ottawa, Ontario K2H 9C1
Canada
(613) 820-3677