

Quality of Service Pilot: Implementation

Jonathan J. Couzens

Principal Network Engineer

JANET Network Operations & Service Centre

jjc@nosc.ja.net

+44 (0)20-7692-1316



Agenda

- Classes of Service
- Implementation
- Policy
- Backbone Rollout
- Results so far

Agenda

- **Classes of Service**
- Implementation
- Policy
- Backbone Rollout
- Results so far

Classes of Service

- Historically, the Internet has treated all traffic equally:
 - A so-called *Best Efforts* service
 - the network will try its best to deliver data (*packets*) but there are no guarantees
 - of delay
 - of order
 - of delivery
 - no discrimination between applications

Classes of Service

- This *equality-of-service* is no longer seen as being sufficient for all applications' needs:
 - Voice-over-IP
 - » needs bounds on latency
 - H.323 videoconferencing
 - » needs bounds on latency and jitter (the variation in latency)
 - Very large bulk data transfers
 - » moving terabytes of data around could be deemed to be 'anti-social'...

Classes of Service

- New services require that the network treats some data flows as being more equal than others....
 - Low latency (and jitter) require preferential queueing to avoid congestion impacting on the application
 - Very large data transfers may behave ‘altruistically’ and require that other traffic takes precedence under congestion

Classes of Service

- The needs of each application can be quantified by its Quality-of-Service (QoS) requirements
 - each distinct group of requirements forms a *Class of Service*
 - the network needs to be *QoS-aware* in that the network components (routers, switches etc.) need to be able to recognise and process each defined Class of Service appropriately

Classes of Service

- Three Classes defined:
 - Best Efforts (*BE*)
 - » ‘normal’ Internet traffic
 - Less than Best Efforts (*LBE*)
 - » reliable but defers to *BE* traffic under congestion
 - » projected use: ‘altruistic’ bulk data transfers
 - *Premium*
 - » requires low latency and low jitter (consistent latency)
 - » projected use: H.323 videoconferencing

Agenda

- Classes of Service
- **Implementation**
- Policy
- Backbone Rollout
- Results so far

QoS Implementation

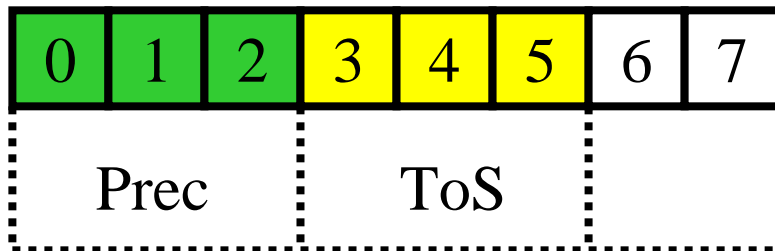
- Various proposals to the IETF to allow multiple services to be implemented across the Internet
- In particular,
 - Differentiated Services (*Diffserv*)
 - hop-by-hop, connectionless
 - Integrated Services (*Intserv*)
 - end-to-end, connection-oriented
 - resource reservation

QoS Implementation

- Diffserv
 - Each packet marked to indicate which service the data flow is using by setting the Differentiated Services Code Point (DSCP)
 - At each hop within the domain, the way each packet is processed based solely on the DSCP
 - » *per-hop behaviour*
 - The DSCP is set at the edge of the ingress domain but may be reset whenever the packet enters a new domain

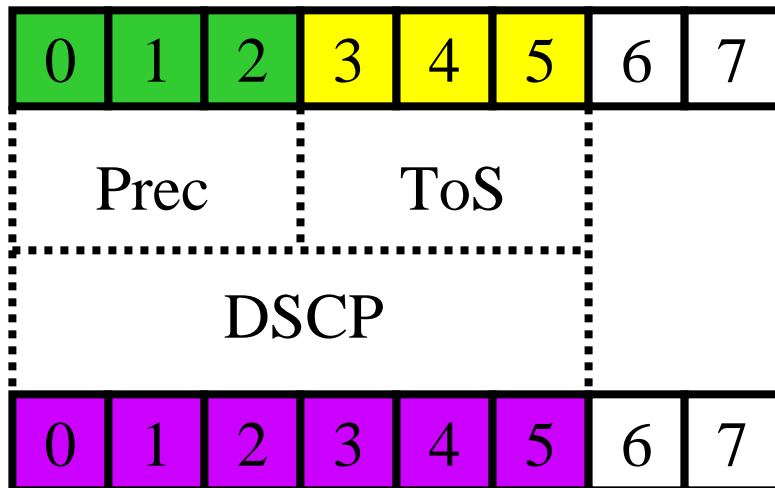
QoS Implementation

- IPv4 precedence and Type of Service fields (RFC791: Sep-1981):



QoS Implementation

- Differentiated Services Code Point (RFC2474: Dec-1998):



QoS Implementation

- Each CoS mapped into a DSCP:
 - *BE*: DSCP 0
 - *LBE*: DSCP 8
 - *Premium*: DSCP 46
- and
 - *Near-premium*: DSCP 40-45, 47

QoS Implementation

- Each CoS maps to a specific DSCP
 - the chosen DSCPs map to different precedences to cater for routers that do not understand DSCPs but are able to use the old IP precedence to differentiate between traffic classes
 - *BE*: Precedence 0
 - *LBE*: Precedence 1
 - *Premium* and *Near-premium*: Precedence 5

Agenda

- Classes of Service
- Implementation
- Policy
- Backbone Rollout
- Results so far

Forwarding Policy

- Queueing/scheduling
 - no congestion: no queues
 - prioritising some traffic above other traffic
 - » multiple queues separate effects
 - » priority queueing
 - selective dropping under resource congestion
 - » queue lengths affect latency
 - » queueing discipline: WRED, Tail Drop?
 - implementation is very platform-specific....

Forwarding Policy

- Best Efforts (*BE*)
 - queue lengths determined by egress bandwidth
 - queueing discipline: WRED
- Less than Best Efforts (*LBE*)
 - same as BE but short fixed-length queues
 - if multiple queues are possible, use a distinct queue
- *Premium*
 - separate queue with strict priority scheduling
 - short queues (large queues cause increased jitter)
 - queueing discipline: Tail Drop

Forwarding Policy

- The mere fact that some traffic is prioritised over other traffic gives us a potential problem:
 - what's to prevent anyone marking their traffic as *Premium* and getting a better service at the expense of others?
 - worse, what's to prevent Mr. Nasty doing so with a Denial-of-Service attack?

Edge Policy

- Classification
- Marking/remarking
- Policing
 - enforce allowable traffic levels
- Needs to be deployed around the entire edge of a management domain

Edge Policy

- Different policy for participants in the QoS pilot to non-participants
 - » in particular, for *Premium* traffic
- Some applications already set either IP precedence or DSCP.....
 - *When the Edge Policy is fully implemented, potentially, both participating and non-participating sites will be affected!*

Edge Policy

- Criteria for classification
 - Examine every packet entering the backbone from another domain to determine the DSCP
 - » includes all external traffic
 - » means instrumenting every interface in the backbone that connects to a router in a different management domain (regional networks, peer networks, sites....)
 - The DSCP is the sole criterion for differentiation
 - » responsibility for setting the DSCP devolved to the Regional Networks and/or sites

Edge Policy

- *BE* or *LBE*
 - just allow it in unchanged
- *Near-premium*
 - remarked and treated as *BE*
 - » needed to cater for precedence-only routers
 - » applies to all edge connections, both participant and non-participant
- Anything else that isn't *Premium*
 - just allow it in unchanged



Edge Policy

- *Premium*
 - police the traffic level and, if it exceeds the permitted level, drop or remark it
 - » if traffic is dropped, this has an impact on the connectivity
 - » required to protect the network
 - otherwise, allow it in unchanged
 - level of policing and the action taken depends on the connection type and bandwidth

Edge Policy

- *Premium* Policing

- External networks

- if providing *Premium* as part of a peering agreement (for example, GÉANT)

- » need to check DSCP value for *Premium* traffic that the peer is using (may not be the same as we use)

- » police to 50% of inbound bandwidth

- » remark to our *Premium* DSCP if DSCP mismatch

- otherwise

- » remark all ‘*Premium*’ traffic to BE

Edge Policy

- *Premium Policing*
 - Participating Regional Networks and sites
 - police to 5% of the ingress bandwidth or to 2Mbps, whichever is greater
 - if traffic exceeds the permitted level, the packet will be dropped

Edge Policy

- *Premium* Policing
 - Non-participating Regional Networks and sites
 - treat as *Near-premium*
 - » remark to *BE* and allow in
 - » no policing of traffic level
 - permitted traffic level is 0Mbps
 - » but we remark, not drop

Agenda

- Classes of Service
- Implementation
- Edge Policy
- **Backbone Rollout**
- Results so far

Backbone Rollout

- Phase 1: Edge instrumentation
 - Examine the traffic entering the backbone from the connected sites and networks
 - No effect on the existing service
 - Needed so that we know what, if anything, will be affected by the next phase

Backbone Rollout

- Phase 2: Edge policy enforcement
 - Examine the traffic entering the backbone from the connected sites and networks
 - Police, remark, pass or drop according to the Policy
 - » potential effect on the existing service for both participants and non-participants
 - » traffic failing to conform to the policy will certainly be affected

Backbone Rollout

- Phase 3: Service Activation
 - Implement the forwarding policy on all routers
 - Potential effect on the existing service
 - » if a lot of *Premium* traffic, possible worse service for *BE*
 - Cannot be started until Phase 2 (Edge Policy Enforcement) has been fully completed
 - » cannot afford to allow unregulated *Premium* traffic into the network once the forwarding policy is activated

Backbone Rollout

- Edge Instrumentation
 - Phase 1
 - started early October 2002
 - now complete

Backbone Rollout

- Edge Policy Enforcement
 - Phase 2.1 - remarking
 - simply change the instrumentation rules to remark all *Near-premium* (and *Premium* for non-participants) to *BE*
 - » do not drop traffic
 - » allows us to see any unexpected effects caused solely by remarking
 - » affects all edge routers

Backbone Rollout

- Edge Policy Enforcement
 - Phase 2.2 - policing
 - simply change the instrumentation rules for *Premium* for participants to drop packets if the traffic contract is exceeded
 - » allows us to see any unexpected effects caused solely by dropping traffic (rather than remarking)
 - » should appear to the end user as virtual congestion
 - » affects only edge routers connecting participants

Backbone Rollout

- Edge Policy Enforcement
 - Phase 2.1
 - scheduled to start 29th April 2003
 - estimated to be complete by the end of June 2003
 - Phase 2.2
 - probable start at end of June 2003
 - probable completion by end of July 2003
 - » just in time for Clearing!
- All changes will be announced by trouble tickets

Backbone Rollout

- Service Activation
 - Phase 3
 - scheduled to start at the beginning of September 2003
 - complete by the end of the year
 - except for tuning....
 - » queue thresholds?
 - » permitted traffic levels?
 - » etc.

Agenda

- Classes of Service
- Implementation
- Edge Policy
- Backbone Rollout
- Results so far

Instrumentation Results

- CAR rules:
 - Look at DSCPs 0 and 8 for *BE* and *LBE*
 - Look at DSCP 46 for *Premium*
 - Look at DSCPs 40-45 and 47 for *Near-premium*
 - Look at DSCPs 48 and 56 for Network Control
 - Summarise traffic for all the other DSCPs
 - » not worried about differentiating these

Instrumentation Results

- Overall traffic profile:

–	
– <i>BE</i>	97.752 %
– <i>LBE</i>	0.042 %
– <i>Premium</i>	0.004 %
– <i>Near-premium</i>	0.150 %
– Network Control	0.074 %
– Other	1.978 %

- most individual connections are similar...

Instrumentation Results

- Anomalies....

–	ClydeNet	RAL
– <i>BE</i>	99.981 %	99.827 %
– <i>LBE</i>	0.000 %	0.000 %
– <i>Premium</i>	0.000 %	0.000 %
– <i>Near-premium</i>	0.000 %	0.000 %
– Network Control	0.019 %	0.173 %
– Other	0.000 %	0.000 %

Conclusions

- A multiservice Internet is coming....
 - It's taking a long time
 - about 15 months for the backbone alone
 - 58 routers involved, 126 edge interfaces
 - » required to be reconfigured on up to 4 separate occasions
 - » need to interleave changes with other work....IPv6!
 - Ongoing effects
 - more complex monitoring required
 - more scope for human error

Quality of Service Pilot: Implementation

Jonathan J. Couzens

Principal Network Engineer

JANET Network Operations & Service Centre

jjc@nosc.ja.net

+44 (0)20-7692-1316

