VoIP Quality of Service - Basic Theory

PacNOG5 VoIP Workshop
Papeete, French Polynesia. June 2009

Jonny Martin - jonny@jonnynet.net
Intro

• What is Quality of Service (Qos)?
  • QoS and the PBX
  • Traffic Types

• VoIP
  • Traffic characteristics
  • VoIP concerns

• Operation and implementation of QoS
QoS

- When people say QoS, they are normally referring to the end result
  - i.e. I want my VoIP calls to always be of a high quality

- QoS can be used for good or for bad though!
  - ISPs shaping bit-torrent is uses QoS tools to *reduce* the quality of that traffic

- QoS tools allow us to manage traffic flows in a network
  - Through queuing, shaping, and dropping traffic

- All about *instantaneous* traffic loads
Why do we need QoS

• Well, we don’t always.

  • Only need QoS when there is not enough network resource to carry the traffic offered to it at any point in time
Legacy PABXs and QoS

- Even legacy TDM PABXs include QoS features
- TDM PABXs are circuit switch
  - A call has guaranteed bandwidth across a circuit end to end
  - No other traffic to worry about on the link
  - When there aren’t enough circuits to carry the load, the call cannot take place
    - Call Admission Control (CAC)
    - Call Blocking
Legacy PABXs and QoS... ctd

- TDM PABXs generally utilise a separate signalling and media paths.
- No other traffic to ever worry about
VoIP Traffic

• Just a particular type of data traffic

• One that is very sensitive to underlying network performance
  
  • Users complain about bad quality calls

• Almost always operating on a ‘converged’ network which supports general IP traffic also
Data Traffic Characteristics

• Variable and inconsistent traffic patterns
  - Sometimes a little bit of bursty traffic, e.g. web browsing
  - Sometimes continuous high bit rate traffic - bittorrent and the likes

• May be symmetric or asymmetric traffic loads

• General non-real time data
  - Delays in traffic delivery (within limits!) not really an issue
  - Most protocols have built in re-transmission functions
Voice Traffic Characteristics

- Very consistent traffic flows
  - High PPS
  - Small packets
  - Different codecs present more or less of the above
- Always Real-Time traffic
  - We’re dealing with real, real-time traffic
  - No re-transmissions
  - Delay and delay variance (jitter) affect the ‘quality’ of the call
Voice Traffic Characteristics... ctd

• Generally runs on an existing ‘data’ network
  
  • Need some form of QoS to ensure traffic gets the priority you consider it needs

• VoIP - voice runs over your IP network
  
  • If your IP network is broken, then you can’t expect your VoIP devices to magically fix this!
Network concerns

- Queuing delays
- Congestion
- Serialisation delays
- Packet drops and performance on third party links (ISP, Carrier links, etc.)
  - over subscription of bandwidth
  - high latency and or jitter
VoIP Concerns

• Latency
  • High latency is bad for voice call ‘quality’ - we can live with it though

• Jitter
  • Jitter is really bad for voice call ‘quality’ - jitter buffers attempt to fix it

• Packet drops
  • Packet drops are really, really bad for voice call ‘quality’ - really hard to recreate a packet that we don’t receive!
Latency Sources

- Codec delay
- Processing delay
- Serialisation delay
- Queuing delay
- Propogation delay
- Jitter buffer delay
How much latency is too much?

• Generally 150ms one way is fine

• Any longer and conversations start to feel half-duplex
  
  • ‘Hi Dan, it’s Jonny here... over.’

  • ...

  • ‘Oh hai Jonny!... over’

  • ...

  • ‘Wow, only one of us can talk at a time!’

• Just like satellite based calls
Jitter

- Jitter is variance in inter-packet arrival times
  - variable delay

- Latency itself doesn’t garble voice traffic. Jitter does

- Normally caused by serialisation delays
  - small packet gets caught waiting for a large packet to transmit

- Or multiple traffic forwarding paths for the same flows resulting in packets taking different paths
  - per packet load balancing a bad idea
Dropped Packets

- If there is not enough bandwidth on a link for the presented load, traffic will be dropped
  - may be due to interface congestion
  - may be due to policers enforcing service parameters
- Bad physical layer can cause lost / corrupt packets
Prioritisation

• QoS based on providing priority to some traffic (VoIP, in our case) over others during times and at points of congestion

• Tell out network to prefer VoIP packets and treat them a little better than general data packets
QoS Components

Classify and Mark

Police and Shape

Queue and Schedule
Classify and Mark

- Inspect traffic and assign it to a ‘class’
  - L2/L3/L4 and higher information can be used
  - IP Addresses, protocol, port...
- Different platforms and different vendors provide different capabilities for this
- Markings
  - Class of Service (CoS) - Layer 2
  - Differentiated Services Code Point (DSCP) - Layer 3
  - EXP (experimental) bits - MPLS
Marking traffic

• Can be done as a result of classification on routers

• Set by hosts

• Set by VoIP devices themselves

• Need to decide on trust boundaries, and whether or not to trust incoming marks at any point in a network
Traffic Policing / Shaping

• Traffic shaping queues/buffers traffic and send it on at a uniform rate
  • a ‘soft’ approach
  • Provides higher link utilisation, at the expense of additional variable latency
• Traffic policing simply drops packets once the offered rates exceeds a threshold
  • a ‘hard’ approach
  • No additional latency or jitter, at the expense of lower link utilisation
Available tools

• Committed Access Rate (CAR)
  • provide rate limits on interfaces

• Low Latency Queuing (LLQ)
  • Always send voice packets first before other traffic

• Class Based Fair Weighted Queuing
  • Enhancement of LLQ that avoids starving other traffic types of bandwidth and a ‘fairer’ queuing mechanism
  • Provide a minimum bandwidth allocation for VoIP, can be used by other classes when not being used
Queuing / Scheduling

• Historical queueing methods
  • First In First Out (FIFO)
    • Just like in McDonalds
  • Priority Queuing (PQ)
    • E.g. airline checking. First/Business have separate queue
  • Custom Queuing (CQ)
  • Weighted Fair Queuing (WFQ)
    • Airline example, but First/Business counters occasionally take Economy
Modern Queuing

- Class Based Fair Weighted Queuing (CBWFQ)

- Weighted Round Robin (WRR)
  - Weights determine frequency of queue service (WFQ, CBWFQ) or ratio of queue serving (WRR)
Packet Scheduling

- Priority Queuing for WFQ (PQ-WFQ)
  - older way of doing things
- Priority Queuing for CBWFQ (LLQ)
  - newer way of doing things
Packet Dropper Techniques

• Tail Drop

• Weighted Random Early Detection (WRED)
  • can be distributed (dWRED)
  • or flow based (FBWRED)
WRED

1. RED with a Slope

2. RED without a Tunable slope
WRED... ctd

Incoming packets → Classify → Discard test → Transmit queue → Outgoing packets

Discard test based on:
- Buffer queue depth
- IP Precedence
- RSVP session

Queueing buffer resources

FIFO scheduling
Where to implement QoS?

• Where ever there is a congestion point
  • WAN edge
  • Slow links
  • Aggregation points
Where to Apply QoS
Cisco QoS support

• We’ll only look at the very basics here

• Modular QoS command line (MQC) on IOS

• Lots of knobs you can twiddle with...
  
  • WFQ, CBFWQ, LLQ, LFI, DSCP, CoS...