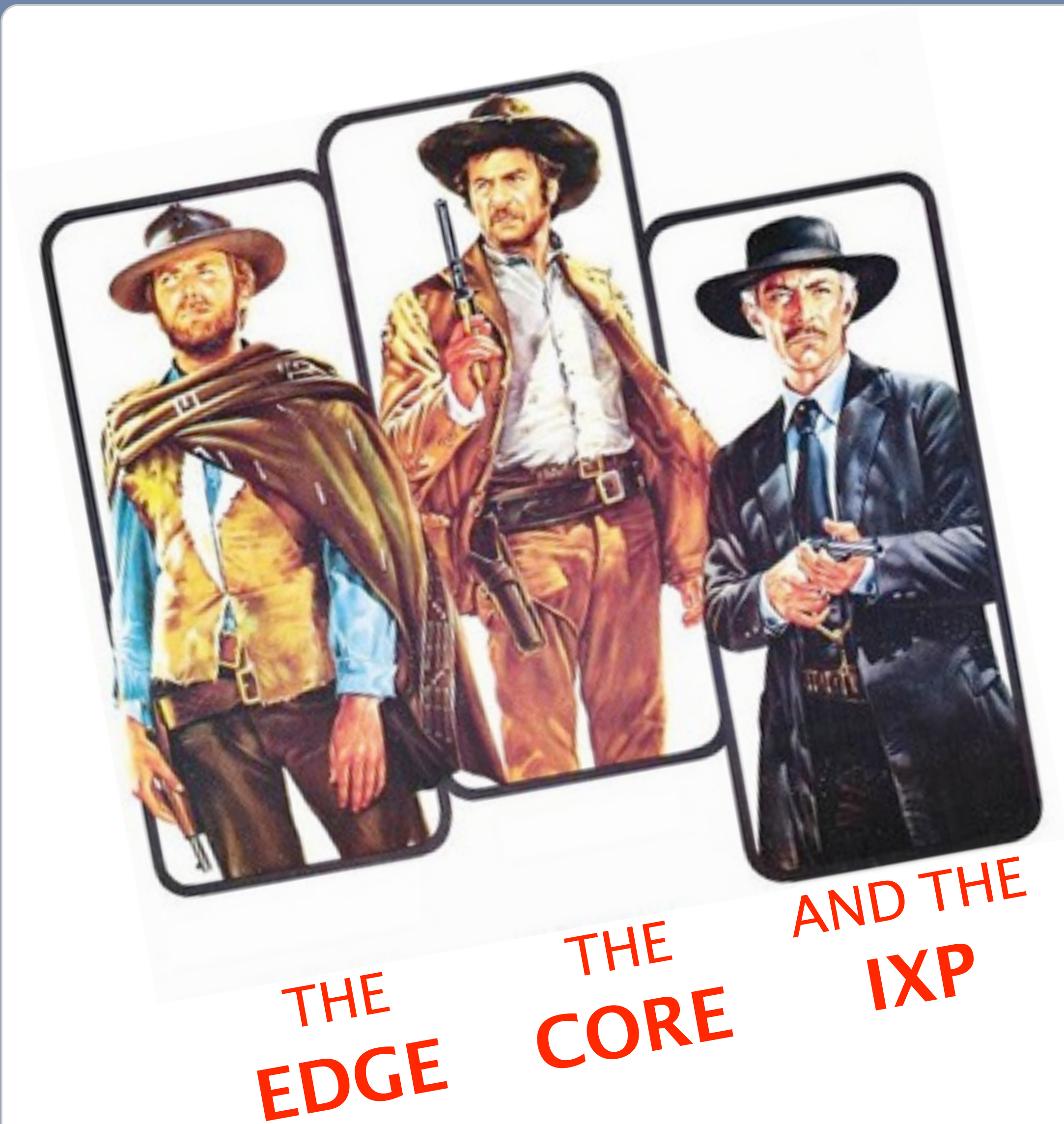
VOIP QOS



ITSPA - Autumn Seminar

11th October 2012

LEEDS

Thomas Mangin

Technical Director

IXLeeds

AGENDA

NO AGENDA

Agenda are good to let you known when to doze off

There is under 30 slides

you can do it !

QOS Definition : Telephony

In the field of telephony

defined by the ITU in 1994 : E.800

Terms and definitions related to quality of service and network performance including dependability

requirements on all the aspects of a connection

response time, loss, signal-to-noise ratio, cross-talk, echo, interrupts, frequency response, loudness levels, ...

Cares about customer experience

The Quality of Service ...

QOS Definition : Networking

In the field of networking

resource reservation control mechanisms rather than the achieved service quality

ability to guarantee a certain level of performance to

data flows, applications, users, ...

Has no concept of end user experience “Quality”

QOS Definition :
Let's speak the same language ..

Explaining IP challenges related to VoIP

- Packet-switched networks issues / constraints ..
  - The problems QOS tries to solves
  - RSVP (MPLS) vs DiffServ
  - QOS across different networks
- End to End network design for your services
  - AND why you should care about peering

I am speaking of that QOS

To improve call quality

- The QOS your customers care about

To achieve that one

Why is QOS required in packet switching networks ?

iX Leeds

Most of the challenge are with RTP (Audio)

- Signaling is much more resilient
  - But mostly still uses UDP

OFF-TOPIC

What can go wrong in a IP network ?

- Errors (hardware issues)
  - Bad configuration (duplex mismatch)

Only money can help

- Bandwidth congestion
  - Dropped packets

Money will help
QOS may help

- Latency / Jitter
  - Bad call quality

QOS may help

Requirements :
VoIP is a sensitive service

VoIP requires regular packets

- RTP has no retransmission
- VoIP buffers cause audio delay

VoIP is very sensitive to packet loss

- dropping packet is the worst thing which can happen
- Missing packet means an audio drop-out occurs for the speaker

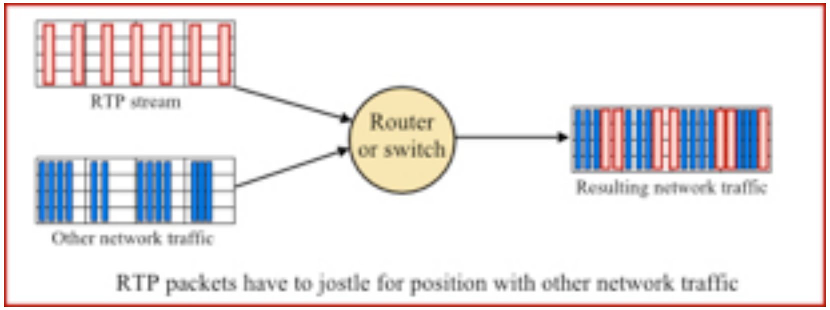
VoIP requires a network which “cares”

- Some networks are “optimised” for data
  - throughput using deep buffers VS jitter
  - buffering is good for data and bad for VoIP

Jitter: What is it ?

Jitter is a variation of Latency

- variation in delay between packets
- can cause out of order packets



Jitter is the biggest VoIP challenge

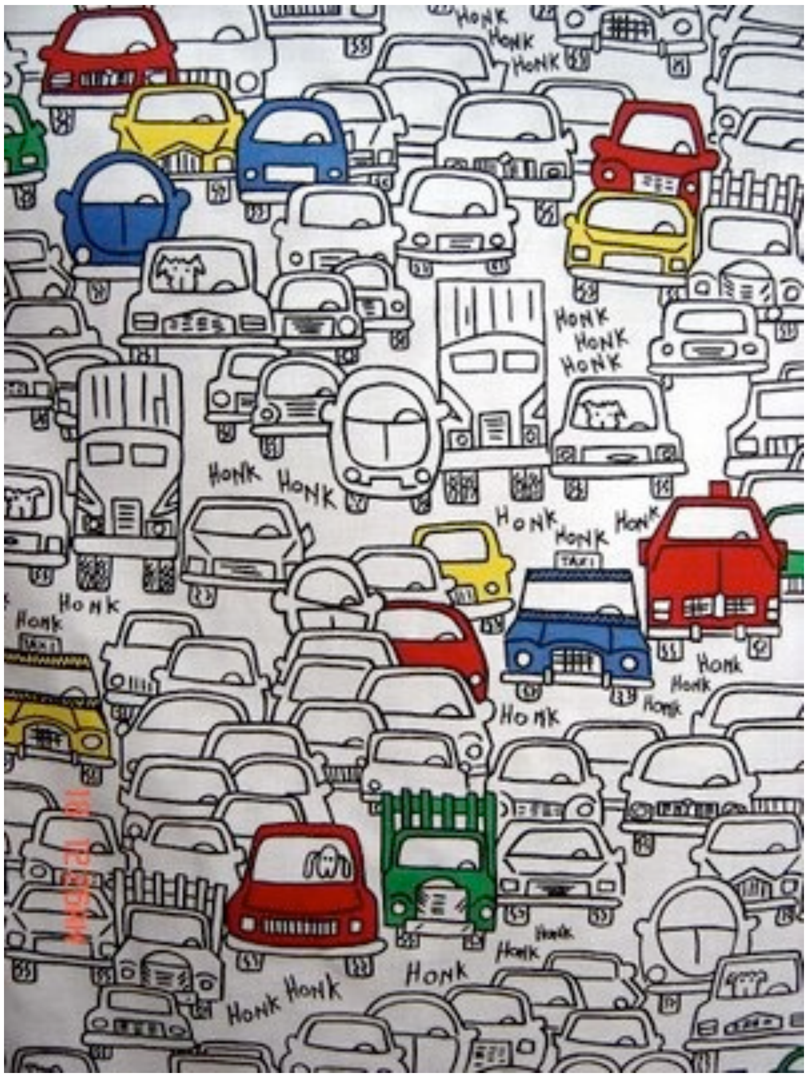
- For an network engineer

Jitter can happen in well provisioned network

- Multiple paths to a single destination, worse with unequal path length
- use of multiple line with L2/L3 load balanced (need to be per flow)
- Can be caused outside the network
  - transcoding on a very busy machine !

Jitter :
The "real world" analogy

A VoIP packet is like a car carrying audio tapes ...



take a very large audio book on tapes

and send it to a friend living abroad

send a tape a day with a car courier

    packet network

require your friend to listen a tape a day too

    (real time constraint)

ask him to only keep two tapes with him

    (small buffer constraint)

Jitter :
The "real world" analogy

What if ..

you could tell the driver which road to use

- using the same road, mean encountering the same traffic jams
- car more likely to arrive in the right order
- **predictable paths reduce the chances of jitter**

The listener can keep more tapes

- more likely to be able to listen a tape a day
- it may delay the starting day
- **correct buffering reduce chances of jitter**

Jitter :
The “real world” analogy

What if ..

The listener moves to nearer to your house (from Asia to Europe)

shorter paths reduces chance of jitter

your friend moves back to “mainland”

less chance of hitting a french ferry/eurostar strike

not using third party infrastructure reduces the chances of jitter

Only use motorways

less chance of being stuck behind a long slow truck on a small road

uncongested path reduces the chances of jitter

Jitter :
What can cause it ?

Every router between the caller and the callee

- small RTP packets (200 bytes) “stuck” behind “larger” (1,500 bytes) data packets
- high serialisation times on slow speed circuit

How much of a delay ?

- 1500 byte packet at 100Mbps takes 0.12ms
- 1500 bytes at 2Mbps = 6 ms
- 1500 bytes at 250kbps = 48 ms

On a DSL router at 250k

- sending a large mail will cause at least 48 ms delay for many packets
- a RTP packet behind behind two data packet will have 96 ms of delay
- 48 ms alone may be enough to exhaust a low de-jitter buffer

Latency / Jitter : End to End transfer time

How much Latency can a VoIP call sustain ?

- ITU G.114 recommendation for end-to-end delay is 150ms max
  - at 400ms two way conversation becomes very difficult.

Source of buffering

- codec delays (converting to and from raw audio)
- network
- jitter buffer

Correct jitter buffering

- typically 30-50 ms worth of data
- Over-buffering can cause latency issues

Bandwith congestion :
The "real world" analogy

Congestion ...

- trying to pass more cars than the motorway can take ..
  - A best everyone arrives late (latency / jitter)
  - Some cars may not reach their destination (packet drops)

Only BIG difference with IP Network ..

- no one will see his car vaporised in a traffic jam to reduce it ...

Never good to any IP network

- Never run any link at more than 70% average capacity
  - micro-burst - will cause bad packet loss for VoIP
  - monitor peak (5 minutes average sampling is a LONG time)

Bandwidth congestion : QOS can / can't

It CAN ..

Provide resource reservation

- provide your VoIP with a "special virtual network" (RSVP/MPLS) with guaranteed bandwidth even if some big FTP transfer is going on

Provide "flow" prioritisation

- make sure all UDP packet from and to to AQL RTP servers pass before anything else

It can NOT ..

More than 100 Mbps worth of RTP down a 100 Mbps circuit !

on a 1Mbps circuit carrying 960kps of RTP perfectly (12 calls)

a single 80 kbps call will cause audio break-out for ALL users

QOS Types :
The car analogy

MPLS (RSVP)

The bandwidth is reserved for VoIP...

But lost for other services

The bus lane

Proritisation (Most often DiffServ)

The flow are prioritised

The police/ambulance siren

Traffic MUST still be “constrained”

Often three levels of prioritisation

best effort	(data)
real-time	(voip, video)
network control	(ISP protocol / remote administration)

QOS :
How does it works

1 - Classify

find the traffic you want to treat specially

2 - Mark

add information to the IP packet telling how it should be treated

3 - Queue

apply the differentiation on every router ...

QOS :
How does it works

1 - Classify

find the traffic you want to treat specially

- using L3 packet header (source IP, port, ...)
  - using ACL on edge / core devices
- using Deep Packet Inspection (looking at the content of the packet)
  - example Cisco NBAR

Agreed between the client and the network provider

- some equipment may be using non-standard port / settings
- can be to boost what you know is VoIP (often ACL)
- can be to slow down what you do not know (often DPI)

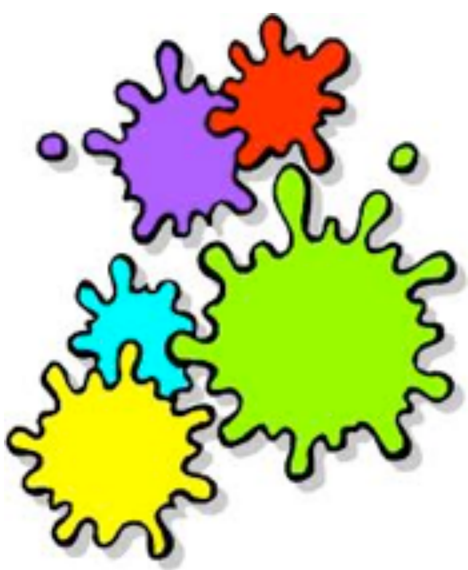
Must be done with any VoIP service provider too

- traffic from and to the PSTN gateway must be prioritised

QOS :
How does it works

2 - Mark

- add information to the IP packet telling how it should be treated
  - used for later action
- can be done at Layer 2
  - ethernet CoS field inside of 802.1q framer (3 bits)
  - information lost when the packet is untagged
- or at Layer 3
  - IP header has 1 byte TOS field for IP Precedence
    - TOS was never widespread
    - Used as DS field (6 bits) + ECN (2 bits)
  - 6 bits Diff Serv (Differenciated Services) + 2 bits ECN
- Ideally done one when the packet enter the core
- unmark traffic marked by customers when (not) appropriate
  - do not want a VoIP DOS by having junk prioritised



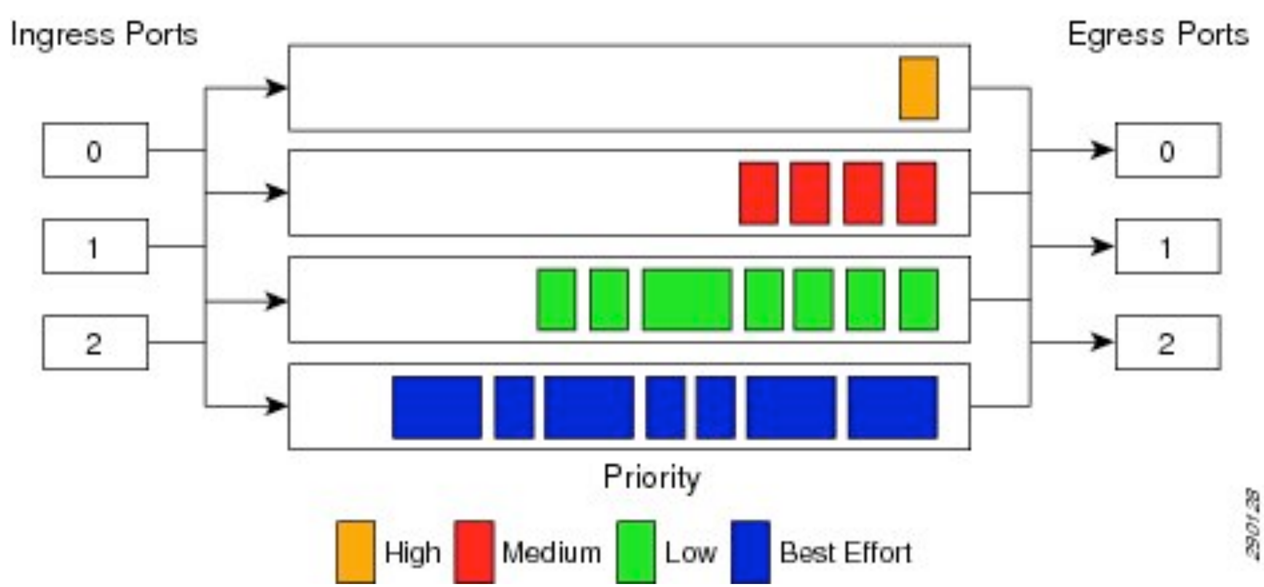
QOS :
How does it works

3 - Queue

routers can only affect OUTGOING traffic

queuing is only important when there is output congestion on an interface

packet loss occurs when interface buffers are full, before : latency and jitter



http://www.cisco.com/en/US/docs/solutions/Verticals/CPwE/CPwE_chapter8.html

QOS :
How does it works

3 - Queue

- prioritise traffic based on marking
    - software queues on small routers
        - default a single FIFO
        - can be split in multiple ones, each with different priority and policy
    - hardware queues on most ASIC accelerated hardware
        - often called TX ring or TX Queue
        - here to make sure there is always data ready to be serialised

- different hardware have very different features
    - for the same router, line-cards can matter enormously

What should be done on the EDGE ?

The edge is the most usual congestion location

- Data burst enormously, VoIP must be protected
- **EDGE QOS protects VoIP packets over normal data**
- **EDGE routers mark VoIP packet as more important**
- until the customer pays for a upgrade :D

The edge suffers from high serialisation times

- Lower bitrate links mean higher serialisation time.
- **use rate-limited over-provisioned product**

What should be done on the EDGE ?

- QOS can only affect outgoing packet
  - **QOS must be applied the customer router for outgoing**
  - **QOS must be applied on your EDGE toward your customer**

- If your customer need to pick an ISP
  - Make sure your customer's ISP has some form of VoIP QOS enabled
  - see if the ISP is willing to QOS your VoIP packet toward the customer

What should be done on the CORE ?

Core networks do see many micro-burst

- **every router on the core should have the same QOS policy configured**
- QOS mitigate issues caused by regular / large micro-burst

Serialisation delays are generally better due to faster line rate

- 1000 x 1500 byte packets on 1Gbps circuit “only” adds 12ms of jitter
- a 10 Gb link causes 10 times less jitter as 1 Gb one
- **over-provision the core as much as possible**
- **use faster interface when it is financially viable**

Enable but try to avoid to have to use QOS altogether

What should be done beyond the CORE ?

- **The QOS mechanisms should be applied end-to-end**
  - not always possible if you do not provide connectivity to your customer
  - **most often not possible with transit providers**

- **VoIP providers should seek to reduce latency as much as possible**
  - with their customers suppliers
  - with their providers
  - **Should use the most direct connections**
    - PI or
    - Peering

QOS :
The IXP

Peering provides a direct connection with your business partners

- **improve network visibility**
  - more direct connection / no third party
  - **European IX are member owned, it is an extension of your network**

- **reduce latency / jitter**
  - **peering provide direct network to network connection**
  - IXes provides the same advantages as PI (no hidden congestion)
  - but allow to reach more networks (**good backup for PI**)

- **join a technical community**
  - **Learn lots about the service provider marker**

Questions ?

mailto/xmpp: thomas@ixleeds.net

Thank you for your attention

More information :

http://thomas.mangin.com/data/pdf/Linx 65 - Halfpenny -VOIP QOS.pdf