End-to-end IP Service Quality and Mobility
- Lecture # 11-

Special Course in Networking Technology
S-38.215
vilho.raisanen@nokia.com

Planned contents & draft schedule

1.	Introduction	Jan 13th
2.	Characteristics of mobile applications	Jan 20th
3.	Service quality requirement characterizations	Jan 27th
4.	Challenges of mobile environment	Feb 3rd
5.	Mobility and QoS in GPRS	Feb 10th
6.	Mobility and QoS in 3GPP systems	Feb 17th
7.	Mobility and QoS with Mobile IP	Feb 24th
8.	Mobile IP QoS enhancements	Mar 3rd
9.	Edge mobility and SIP	Mar 10th
10.	Inter-system mobility	Mar 17th
11.	**End-to-end service quality provisioning**	Mar 31st
12.	Summary	Apr 7th

Agenda

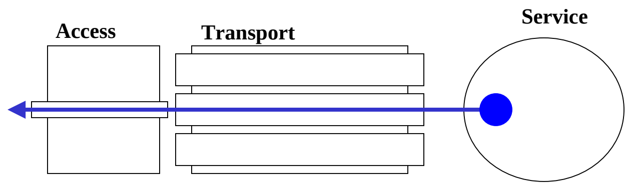
- Service quality provisioning.
- Service Level Agreements.
- Traffic Engineering process.
- Measurements.
- Dynamical end-to-end service quality.

Goals of the lecture

- Understanding of the principles of service quality provisioning.
- Understanding of concepts associated with Service Level Agreements.
- Understanding of the tasks associated with traffic engineering process.
- Assumption: SLA-based end-to-end provisioning.
  - Dynamical mechanisms will be referred to at the end of the lecture.

Service quality provisioning

- Service-centric viewpoint: end-to-end service quality provisioned for instantiations of aggregate service.
- Transport network operator viewpoint: service quality support provisioned for traffic aggregates.
- Access network operator viewpoint: instantiate service quality support for service instances.



Provisioning principles

- Optimal way of end-to-end provisioning depends on the service quality support capability of the access network.
- **Case 1**: access network can instantiate service quality support per session.
  - Session specific SQS signalling.
  - Fine-grained admission control.
  - Support for mobility easier, SQS renegotiation.
- **Case 2**: non-session based service quality support.
  - Allocate service quality level based on user class, application type.
  - Effect of mobility on service quality level determined solely by the network.

Provisioning principles, cont'd

- **Access network:**
  - Case 1: fine-grained SQ levels in access network, negotiated with MN.
  - Case 2: coarse-grained SQ levels in access network, less constraints for e2e SQ.
- **Interface to transport operator:**
  - SLAs covering different traffic aggregates to different PoPs.
  - Traffic conditioning agreement at peering points.

Why optimise provisioning?

- Access network operator wants to make best use of built capacity.
- Low-delay, low-jitter, low-loss service is expensive to provide.
  - Other traffic flows may need to be translated in time or throttled down to make room for high-priority traffic.
- Consequence: high-bandwidth priority traffic expensive to support in the access network.

$\Rightarrow$ In some cases operator may try to use high-priority transport traffic aggregate and save capacity in access network.

- For low-bandwidth high-priority streams, high priority in access network + lower priority in transport may be cheaper.

Example 1: simple MIP-based AN

- Service quality allocated according to user class.
  - E2e SQ defined loosely for application classes.
- Service quality support instantiated in the access network.
  - Traffic conditioning for user traffic or flows.
    - May take into account application class.
  - Particular parameters for access network QoS: DSCP, …
  - Particular traffic aggregate of transport operator used.
- SQ level changes due to mobility (intra- or inter-system) or network load.
  - No change in conditioning or SQ parameters.
  - No change in traffic aggregates towards transport network.

Example 2: future MIP-based AN

- Per-session SQS signalling.
  - E2e SQ precisely defined like in 3GPP QoS model.
- Service quality support instantiated in the access network.
  - Traffic conditioning for session.
  - Particular parameters for access network QoS: DSCP, …
  - Particular traffic aggregate of transport operator used.
- SQ higher due to inter-system mobility or network resources.
  - Tariffing modified.
  - Traffic conditioning reconfigured.
  - SQS related access network parameters may change.
  - Different transport traffic aggregate may be selected.

Summary: rôle of SQS control

- When per-session service quality support signalling used, service quality support control can optimise resource usage for access network.
  - Mapping to different traffic aggregate in transport operator domain (use different SLA components).
  - Map session to different SQS parameters in access network.
  - Renegotiate end-to-end SQ.
- If no per-session SQS control available, renegotiation not possible.
  - Engineering for end user SLAs need to account for effects of mobility on end-to-end SQ.

SLAs

- SLA is a contract between customer and a provider.
  - Definition of agreed-upon service quality level.
    - Preferably using technology-independent terminology.
  - Definition of reporting.
  - Means of verifying service quality level.
  - Definition of procedures for exceptional situations.
  - Common language.
  - Business agreement:
    - Terms of applicability.
    - Validity period.
    - Definition of liabilities.

[Räisänen, ch. 6]

SLA and DiffServ

- **SLA** defines the forwarding service a customer should receive across a DiffServ domain.
- **Traffic Conditioning Agreement (TCA)** specifies the conditioning of traffic arriving at the peering point. **Service Level Specification (SLS)** defines detailed parameters of a DiffServ SLA.
  - Per-Domain Behaviour (PDB) definitions specify the expected performance domain-wide.
- **Traffic Conditioning Specification (TCS)** specifies individual parameters of a TCA.
  - classifier rules
  - metering rules
  - related actions

[RFC3260]

Vilho Räisänen

SLA example

- **Applicability**: 1st of January 2003 – 31st of December 2003.
- **Specifying**: service quality between peering points “Walburg” and “Otaniemi”.
- **Revisioning**: terms of the agreement will be reviewed during the last week of Q1/03 and can be revised by mutual agreement of the parties.
- **Conditioning agreement** for ingress traffic: …
- **Service quality level** (measured over 60-second periods in accordance with RFC 3432):
  - EF PDB:
    - Delay: 90% percentile: <10 ms, 95% percentile <12 ms.
    - Packet loss: percentage < 0.1%.
  - AF PDB:
    - Delay: 90% percentile: <20 ms, 95% percentile <25 ms.
    - Packet loss: percentage < 1%.
- **Verification**: …
- **Reporting**: weekly…, montly …, quarterly …
- **…**

Vilho Räisänen

Traffic Engineering process

- Thus spoke IETF:

  [TE is] the application of technology and scientific principles to the measurement, characterization, modelling, and control of Internet traffic.
- IETF Traffic Engineering process (RFC 3272):
  - Definition of control policies.
  - Feedback mechanisms.
    - Includes measurements.
  - Analysis of network state.
  - Performance optimisation.
    - Includes configuration.

[RFC3272]

Traffic engineering / 2

- Two parts of traffic engineering: capacity management and traffic management.

- **Capacity management.**
  - a – Capacity planning.
  - d-m – Routing control.
  - h-m – Resource management.
    - Link capacities.
    - Buffer spaces
    - CPU resources.

Timescale: flows or packets

- **Traffic management.**
  - Per-node traffic control
    - Traffic conditioning.
    - Queue management.
    - Scheduling.
  - Other traffic control means

[RFC3272]

Measurements

- **Traffic engineering measurement types:**
  - Traffic characterization.
    - Traffic patterns.
    - Distributions.
    - Trends
  - Network monitoring.
    - Service quality monitoring.
    - Determination of network operational state.
  - Traffic control.
- **Measurement technologies:**
  - Passive measurements.
    - One-point.
    - Two-point.
  - Active measurements.
    - Generic measurements.
    - Application emulation.
  - Piggybacking measurements.

[draft-ietf-tewg-measurement-03.txt] Vilho Räisänen **[Räisänen, ch. 7]**

Analysis

- Analysis of network state determines using measurement results whether network parameters need to be changed.
  - Criteria can be based e.g. on PDB thresholds or link utilization levels.
- Analysis can be
  - Proactive.
  - Reactive.
- Analysis can be off-line or on-line.
  - Automated optimisation of network state potentially dangerous.
- Means of obtaining new parameters: analytical modelling, simulations, ad hoc rules.

Vilho Räisänen

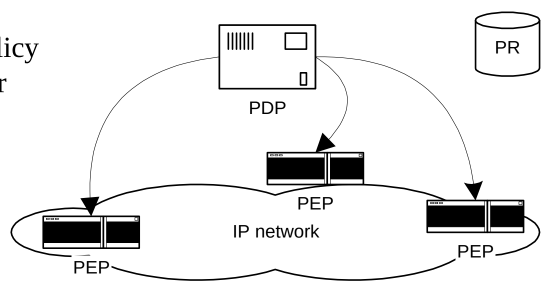
Configuration

- Efficient optimisation of network parameters requires that reconfiguration can be done sufficiently easily.
- IETF and DMTF have adopted policy based management framework as the paradigm for future management:
  - Automation of network management.
  - Means for raising abstraction level of network management.
  - Structured representation of management information.
    - Can express rules for reacting to traffic conditions.
- Two modes:
  - Outsourced mode: decision requested.
  - Provisioned mode: decision "pushed" to network element.

Vilho Räisänen [Räisänen, ch. 4]

Policy management example

- Policy Decision Point (PDP).
- Policy Enforcement Point (PEP).
- Protocol variants:
  - Common Open Policy Service (COPS) for outsourced mode.
  - COPS-PR for provisioning.



Vilho Räisänen

Bandwidth brokers

- Thus far, static inter-domain SLAs have been assumed.
  - May still include multiple traffic aggregates.
- Alternative: dynamic allocation of end-to-end SQ across domains.
- Two basic schemes:
  - Service domain(s) negotiate end-to-end SQ level and allocate it to transport domains.
  - *Bandwidth brokers* (BBs) in different network domains negotiate end-to-end SQ allocation.
- Tasks of a bandwidth broker [RFC2638]:
  - Keep track of resource allocations
  - Configure edge treatment
  - Manage resource allocations to other domains

QBone Bandwidth Broker

- QBone is developing an architecture for dynamic SLAs in a DiffServ-based multi-operator environment.
- Protocols:
  - User/application protocol.
  - Intra-domain protocol.
  - Inter-domain protocol.
- Data available to BB:
  - Routing tables.
  - Aggregate level PDB-like measurements
- Tools used by BB:
  - Aggregation of flows into core tunnels.
  - Negotiation of SQ with end user and other domains.

[qbone.internet2.edu/bb/bboutline2.html]

[Räisänen, ch. 8]

SQS and QBone

- SQS is instantiated using the following parameters:
  - Start and end times (for in-advance reservations).
  - Source and destination.
  - MTU size.
  - Peak rate.
- Service models:
  - QBone Premium Service (QPS)
    - Low-delay, low-jitter, low-loss service.
  - Alternative Best Effort (ABE)
    - Support multiplexing of adaptive real-time applications with data transmission.

[qbone.internet2.edu/bb/bboutline2.html]

Vilho Räisänen

[Räisänen, ch. 8]

Summary

- End-to-end provisioning alternatives:
  - Static
    - Per-session SQS
    - Aggregate SQS
  - Dynamic
    - Service level negotiation.
    - Transport level negotiation: bandwidth brokers.
- Service Level Agreements.
- Traffic Engineering:
  - Measurements.
  - Configuration.
  - Modelling.

Vilho Räisänen