Nokia Siemens Networks

Quality of Service requirements in tomorrow's connected world

A technology guide and vision

Resume

Differentiation solutions are a key component of the Nokia Siemens Networks vision of tomorrow's connected world. In the context of this paper the term is used for solutions that encapsulate different media types, different QoS profiles and different network operator requirements over different time frames. These new solutions are being created in order to address the complex, multi-faceted issues of that multimedia world: issues such as a 100-fold increase in traffic by 2015, 5 billion connected subscribers, and a marketplace with high Quality of Experience (QoE) expectations.

Differentiation solutions provide QoS functionality in three key areas: control, enforcement and management and they allow network operators to fulfill several different objectives simultaneously. This enables the rapid creation of different business models, which in turn enables the delivery of a wide range of services that meet the diverse needs of the marketplace. Differentiation solutions therefore ensure optimum use of precious network resources.

Although this paper focuses on QoS, it should be seen as a synergistic component of tomorrow's connected work. For example, smart connectivity - another innovative development - will enable seamless access to services and applications using different networks and network technologies. Service continuity will be ensured when the transitions take place and a consistent, high QoE experience will be retained smart connectivity is employed in combination with a differentiation solution.

Table of Contents

1. Executive summary

This paper outlines the Nokia Siemens Networks vision of the emerging need for Quality of Service (QoS) solutions that recognize the complex requirements of networks that transport multimedia traffic and enable that traffic to be given different QoS media and subscriber priorities. This need is mandatory. By 2015 there will be an abundance of Internet based services and applications and video related services will set very high QoS requirements for the network and a very high Quality of Experience (QoE) for subscribers. Moreover, by that date there will be 5 billion connected subscribers and traffic volume will have grown 100-fold.

For convenience, we have used the term "differentiation solutions" as a way of encapsulating the complex set of network mechanisms that will meet these objectives. "Differentiation" comprises different media types (voice, video and data); different QoS profiles (premium, best effort, etc); and different network operator requirements over different time frames.

In some cases over-dimensioning the network is a feasible QoS choice for the operator. This allows bandwidth to be segregated and allocated to different traffic types. In other cases capacity upgrades are either too expensive or even impossible due to the regulation. In such cases QoS differentiation provides a wide range of options that allow operators to allocate and manage the network resources as well as enhance the performance monitoring capabilities.

Meeting the needs of a marketplace whose demands are increasing all the time may also require changes and diversity in an operator's business models and strategy. QoS differentiation solutions will provide tools that not only support the requisite changes, but also allow operators to anticipate upcoming requirements. Thus, attractive subscription packages can be created and marketed before the competition.

Each operator and market will be different. This paper can therefore only provide a visionary overview of the various QoS differentiation offers that Nokia Siemens Networks is developing and which will be implemented in line with operator and market requirements.

2. Introduction

2.1 What is QoS differentiation in practice?

QoS differentiation enables operators to segregate the flow of traffic and this allows the performance of the different streams to be monitored on an individual basis, as illustrated in figure 1. Without QoS differentiation, the various traffic streams would be treated in the same way and therefore they would be monitored as a single, aggregated flow.

In this paper the term "QoS profile" does not refer to any particular technology or mechanism. It simply indicates that a particular technology/mechanism is employed that allows the different traffic streams to be identified so that they can be processed and monitored in different ways. The number of QoS profiles that are supported is determined by the requirements of the operator and they can be changed in line with changing requirements.

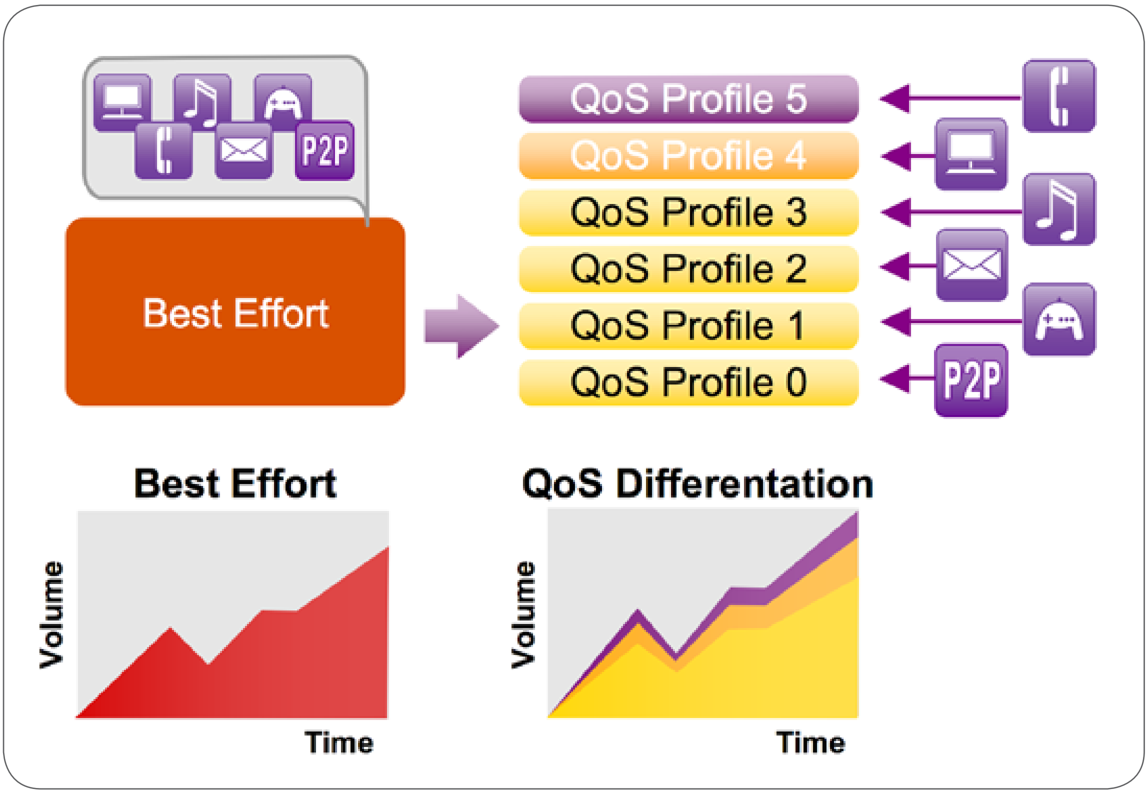


Figure 1. Best effort network and QoS differentiation. Different QoS profiles have different production costs, dimensioning rules, treatment mechanisms and performance counters.

2.2 What are the benefits of QoS differentiation?

The benefits and objectives of QoS differentiation depend on the individual needs of the network operator and operators have different requirements to those of service providers. For example, operators may need to fulfill several different objectives simultaneously. However, most operators are concerned about realizing these common benefits:

- High (or sufficient) QoS
  This is particularly important for real-time (synchronous) applications, which require minimal end-to-end delay and jitter but are not bandwidth intensive. Asynchronous applications, on the other hand, may require high throughput. Fulfilling these different needs will typically involve differentiation in the way the streams are handled.
  For example, video services require short delay and jitter, especially when delivered "on-demand", and they are also bandwidth intensive.

- Emergency services
  Governments will typically demand circuit-switched reliability for emergency services that have migrated to packet switching, e.g. VoIP emergency calls. In case of network congestion these calls must pre-empt the other traffic.

- More efficient use of resources
  As indicated earlier, over-dimensioning the network can enable a high QoS.
  This means that applications such as email that do not require minimal delay and jitter will get them anyway. However, building and employing an over-dimensioned network is an expensive option. QoS differentiation, with its ability to prioritize urgent traffic or guarantee bit rates, is clearly a more efficient way to employ precious network resources.

- Monitoring traffic streams
  QoS differentiation allows operators to monitor individual traffic streams, e.g. ensure that premium services get premium treatment. This allows operators to optimize usage of their network resources and capacity upgrades.

- Flexible support of service offers and business strategies
  Service offers may be based on connectivity having different maximum bit rates with different prices. Alternatively, subscriber pricing can be used to encourage the use of bandwidth intensive applications during off-peak hours. Network operators and service providers may prioritize and otherwise favor in-house applications or those of a partner. Operators may also limit the use of services that would impact unfavorably on profitable services like SMS or circuit-switched voice calls.
  NB: there may be country specific limitations on what operators can do to protect their own services and applications.

2.3 Where will QoS differentiation be applied?

In lightly loaded networks all data streams will experience minimal delay, jitter and packet loss. The need for QoS enforcement mechanisms becomes apparent when there is congestion at the various interfaces, e.g. the access parts of a mobile network. This will typically occur in the access network since this is the most expensive part of the infrastructure. Congestion can be relieved by adding capacity or by investing in QoS mechanisms. However, if the requisite QoS functionality is implemented then expensive capacity upgrades can be postponed. And in some cases country-specific legislation may make the capacity upgrades impossible and leave QoS differentiation as the only viable solution.

The simplest and most mature QoS mechanisms are those that are implemented in backbone networks. Some operators use them already: others have found it cheaper to over-dimension the backbone.

QoS mechanisms are only relevant for shared connections. If resources are dedicated to the end user, e.g. as in the last mile of fixed access networks, then the QoS mechanisms are not required. In the case of mobile networks capacity, e.g. air interfaces, are shared between users. Capacity upgrades in the access part of these networks are more expensive than elsewhere, so this is the area where QoS differentiation provides the optimum solution.

Even though the ability to manage traffic is critically important in the congested parts of the network, similar functionality is desirable in the rest of the network in order to get the maximal monitoring benefits. QoS differentiation identifies traffic types by adding a QoS label to a data packet or logical data stream. Traffic monitoring in a network element will collect traffic statistics for different QoS types and/or combinations. This is important for solving congestion problems as well as the analysis, planning and positioning of services. For example, it can be used to identity which subscribers are using which services. This is critically important since it provides data on the quality of the user's experience and QoE impacts on the price subscribers are prepared to pay for those services.

2.4 Service Level Agreements

The boundaries that used to define the roles of mobile and fixed network operators have blurred. Service providers have entered the picture along with content providers and we also have developments such as fixed mobile convergence. As a result, today's value chains may have several links but that is of no interest to subscribers.

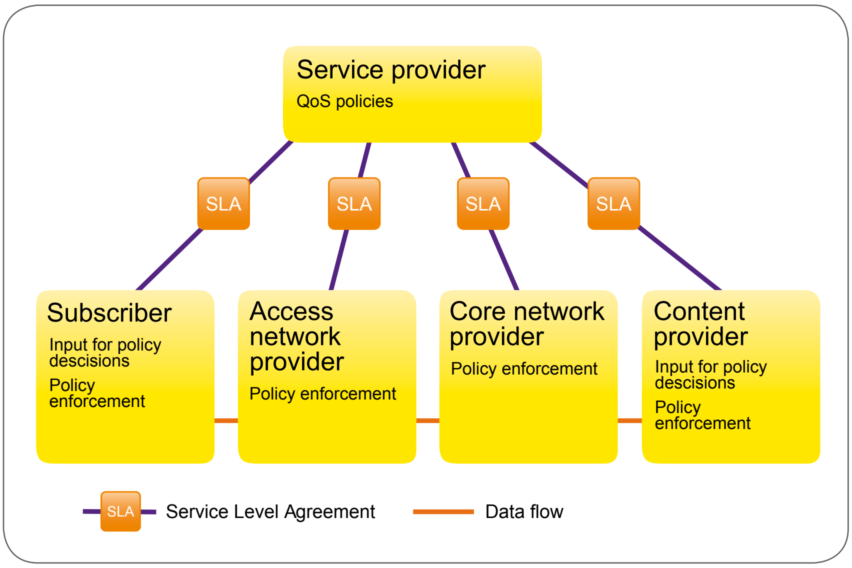
Their need is easy to define: intuitive, ubiquitous access to services and applications and a consistent user experience, i.e. a consistent, high QoS. This can only be achieved when, as illustrated in figure 2, there are Service Level Agreements (SLAs) between the various parties.

SLA may contain Service Level Specifications (SLSs) on the usage and implementation of QoS profiles and labels, as well as Service Level Objectives (SLOs) that define maximum delay and throughput. In addition, different QoS profiles and markings will normally have different tariffs, e.g. services having a guaranteed bit rate quality will cost more than those that don't.

In figure 2 the service provider at the top of the value chain has a business agreement with the subscribers. This SLA may contain some limitations on bandwidth as well as a quota for the consumed data. The agreement may also contain a priority rating related to the other network users, i.e. as in other aspects of life, one pays more for a priority service. The intention of the service provider is to provide the best possible service for the price that the subscriber is prepared to pay. But our hypothetical service provider does not own the networks so there has to be a set of SLAs in order to realize that objective.

SLAs between the service provider and the network providers can contain very detailed agreements on QoS profiles, markings and performance requirements. In order to maximize profits network providers should use the existing capacity as efficiently as possible while fulfilling the requirements of the service provider.

Figure 2. A typical value chain and the requisite SLAs that are needed in order to ensure a consistent QoS for subscribers. Note the fact that QoS mechanisms are needed in order to ensure that agreements are enforced.



3. QoS differentiation scenarios

Standards help define the way that services and applications can be mapped to QoS profiles or QoS markings. However, operators also need to create the mappings that provide the best fit with their business strategy. Differentiation technology has the requisite functionality, but this should not be used to create over-engineered solutions, e.g. use functionality simply because it's available. Relatively simple QoS differentiation scenarios will be widely employed and they will provide optimum solutions. Nokia Siemens Networks will offer consulting services for the operators to help them to select the most suitable scenario.

3.1 Bandwidth limitation solutions

Operators will typically select a bandwidth limitation solution in order to encourage subscribers to use certain applications at times when the network is not heavily loaded, e.g. outside office hours. If the load can be evenly distributed throughout the day then the network will be used in the most efficient way (less need to over-dimension) and this should also lead to a better and cheaper service for subscribers.

As shown in figure 3, operators may define different maximum bit rates for different subscribers (VIP versus Budget). Subscribers who require high maximum bit rates will pay more than those that don't. Many operators already employ this type of simple differentiation in order to increase their revenue.

3.2 Prioritization solutions

Operators will typically select a bandwidth. The key prioritization strategies are illustrated in figure 4. As outlined earlier, network operators can segment their subscribers into different categories and price the service accordingly. The same differentiation model can be applied to the service providers with whom they have SLAs. In addition, different applications have different QoS requirements. Voice and Video over IP are demanding applications; email is not; and applications such as P2P that place heavy loads on the network can be priced so that they are used late at night.

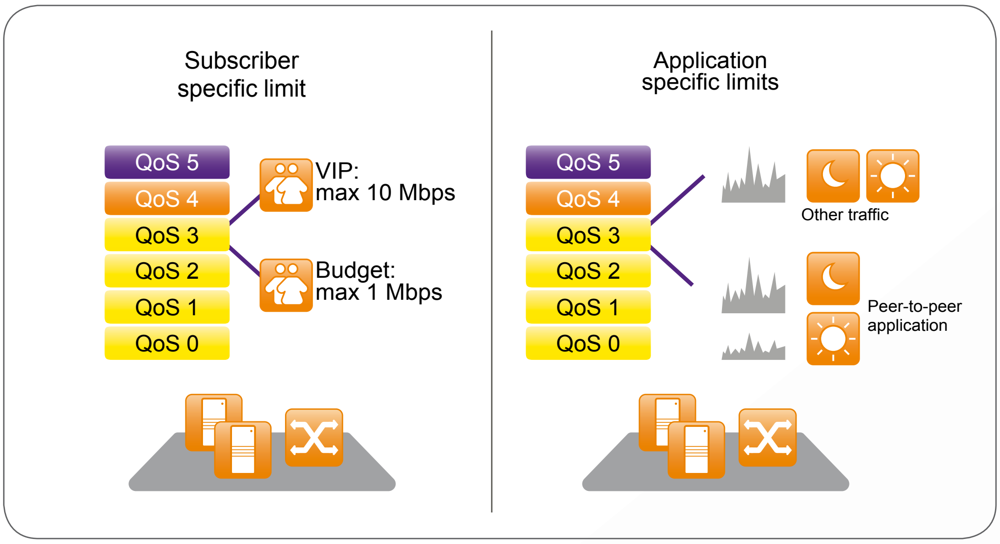


Figure 3. Bandwidth limitation solutions. Business users will typically pay more for high bit rate services that are used in office hours. Other users will pay less when access is made in the evenings or during the night.

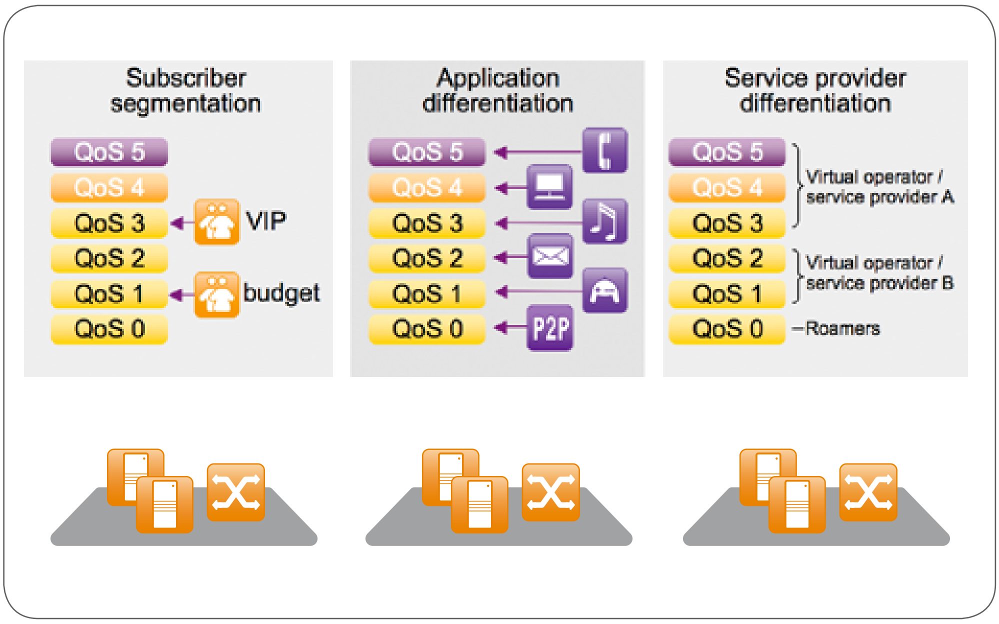


Figure 4. QoS differentiation strategies for subscribers, applications and service providers with whom the network operator has an SLA.

3.2.1 Subscriber segmentation

The most important subscribers, e.g. enterprises and other large organizations, will typically have the most demanding requirements. They will therefore require guaranteed bit rates for their real-time applications and/or obtain a higher priority in case of congestion. The basis of subscriber differentiation is to charge a premium price for premium services, i.e. high priority or high bit rates. And, as outlined earlier, operators can also follow the traffic volumes and throughputs of subscriber segments using the QoS specific performance counters in the access network elements.

Subscriber segmentation scenarios will be normally used in the markets where there are high-income differences between consumers, e.g. in growth markets such as China and India. In mature markets corporate subscribers generate more revenue than the average consumer. If a service provider can charge one segment based on the volume, and offer a flat rate to another segment, then the volume-based subscribers should be prioritized. This subscriber segment should therefore obtain prioritization for all asynchronous connections as well as the guaranteed bit rate for real-time applications since they are paying for a premium service.

3.2.2 Application differentiation

Application and service differentiation enables optimal management of precious network resources. Asynchronous applications tolerate much longer delays than the real-time synchronous applications. The real-time packets can therefore be prioritized when there is congestion.

Application differentiation also enables individual monitoring of the applications as they travel through the network. This allows operators to obtain higher revenues from applications that are prioritized: it also helps retain customers since they see the benefits of better application performance.

Even though operators will be able to differentiate applications from each other with high granularity, operators will typically prioritize key applications that are easy to recognize, e.g. the operator's own VoIP service and their own or a partner's mobile TV service.

3.2.3 Operator differentiation

Network operators may also differentiate the service providers and/or the virtual network operators with whom it partners, i.e. charge more for premium, high-priority QoS services. They can also offer a dedicated QoS profile to an important virtual operator and sell the monitoring statistics for that profile.

4. Technical components of QoS solutions

A complete QoS differentiation solution has the three components, shown in figure 5: QoS Management, QoS Control and QoS Enforcement. The functionality they provide is relevant for all networks, however, the standards that are employed and the way the components are implemented will vary according to the network domain and the operator's specific requirements. For example, in a mobile network the QoS component is typically connection specific, whereas in the backbone and in the fixed access networks QoS can be defined for traffic aggregate.

QoS Control refers to the intelligence that maps the subscribers or applications to the desired QoS. It is also used to prevent the possible attempts of the terminal or the application to misuse QoS, i.e. to try to deceive the network to give too high a priority or a guaranteed bit rate.

QoS Enforcement refers to the mechanisms that treat the connections or flows differently as well as the efficient allocation of network resources. Examples include queuing mechanisms plus shaping, policing and radio resource management algorithms. The network elements are typically not aware of the contents of the data packets: they only obey the QoS that is determined by the functionality of the QoS control component.

QoS Management refers to QoS profile specific monitoring. The network elements maintain QoS profile specific counters and send them to the management tools for visualization and additional network optimization.

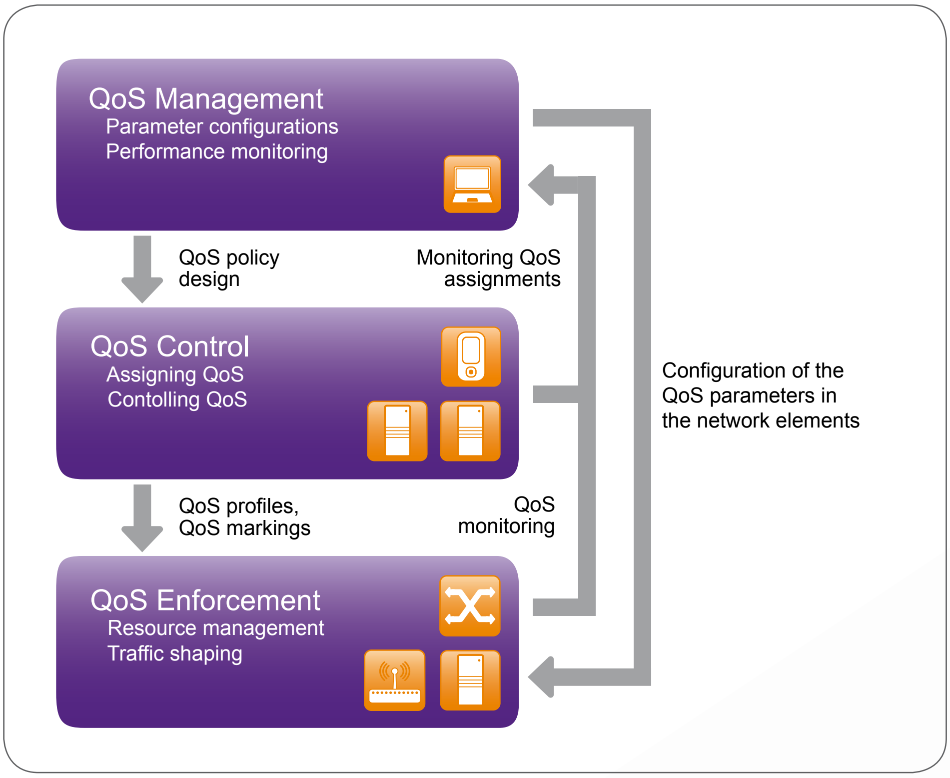


Figure 5. Components of a QoS differentiation solution and their relationships.

4.1 QoS control

QoS control solutions can be implemented in different ways. An application or a terminal may provide input for QoS control, or in certain cases control may rely entirely on terminal QoS requests and QoS markings. However, in most cases QoS control is network-driven because operators cannot trust data coming from terminals. Some operators will apply global policies using IMS, whereas others will rely on local policies in the various networks, e.g. GGSN/ISN, BNG, ASN GW, PDN GW, LTE AGW or routers. Once the appropriate QoS has been assigned the QoS control functionality ensures that the QoS is not misused. For example, the guaranteed bit rates are not used for the applications that do not need them.

Various standards define QoS control mechanisms. 3GPP defines the bearer model with a number of QoS attributes and their values. ETSI has standardized the control for fixed network domains. WiMAX Forum has standardized QoS control for WiMAX and IETF has a set of standards for IP networks.

4.1.1 Terminal and application capabilities

In the future there will be many more types of terminals and a plethora of applications, which means that the diversity of the QoS capabilities will increase. Most of today's terminals cannot indicate the quality requirements of an application or request the requisite quality. Already we can see that subscribers will download an increasing number of different applications to their terminals, which means that software developers in the Internet community will have a significant impact on the QoS capabilities of their applications.

An operator may select and migrate between these two cornerstone approaches:

1. Rely on the terminal requests when devices with the requisite functionality are available. This is feasible when the operator is able to charge more for the higher QoS, or when the operator can control those requests; for example all the subscribers have operator branded terminals with operator-defined applications.

2. Ignore terminal requests or indicated QoS requirements. This is feasible if the operator cannot earn more money when allocating a higher QoS. This applies to a strategy based on flat rate charging and this is a typical near-term business model.

4.1.2 Policy control

QoS policies are rules that define which QoS profile should be allocated to individual applications and subscribers. These rules may depend on times, places and terminal types. The functionality in Nokia Siemens Networks' policy control products will allow a wide range of different policies to be defined and implemented. However, our judgment is that they be used to create simple rules since they will often bring the biggest benefit.

The workflow of policy control is illustrated in figure 6. The operator's first task is to define the rules of the policy control elements. Then, when the connections or traffic appears the rules are recognized and the appropriate policy is matched to the connection or the traffic. Allocating or modifying the QoS profile of the connection or marking the data with a certain QoS enforces the policy. During the connection the rules also ensure that the QoS profile or marking is not misused.

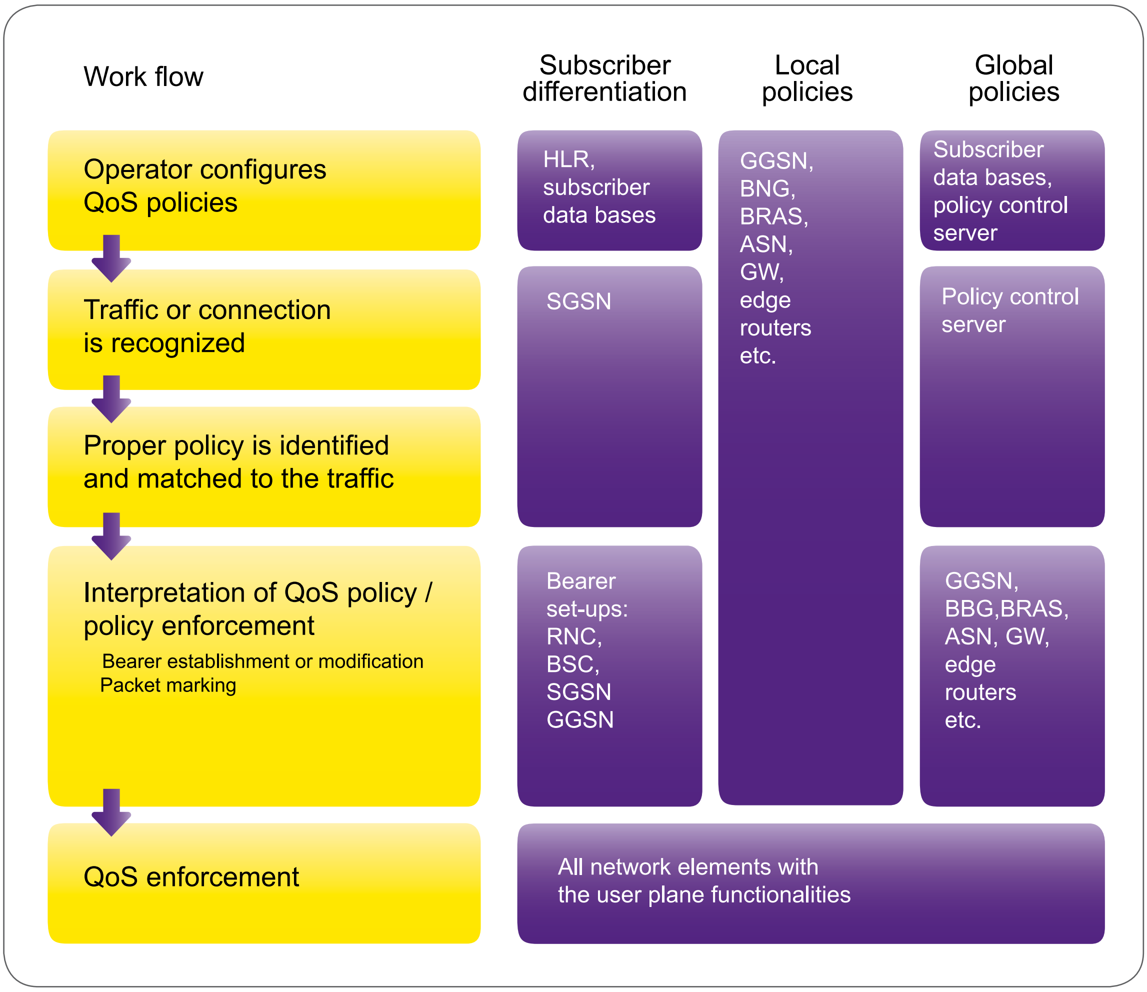


Figure 6. Policy control. The workflow has five steps. The first determines the QoS policies and the last enforces them. The different network elements shown on the right map to these steps, e.g. the policy control server maps to steps 2 and 3.

4.1.3 How does the network recognize the application?

The operator needs to recognize the applications that will be prioritized or have a guaranteed bit rate. Different applications require different recognition mechanisms:

- IMS and SIP applications
  IMS and SIP applications use SIP signaling in session set-ups, modifications and terminations. SIP signaling contains detailed information about the application, e.g. media type, bandwidth requirements and codecs employed.

- Intelligent servers
  Servers offering non-SIP applications can indicate the application characteristics to a policy server within IMS.

- Other servers and the Internet
  Traffic from less intelligent servers can be recognized using deep packet inspection technology. Traffic is categorized using the information in the IP layer and the application layer protocol headers.

- Proprietary peer-to-peer applications
  Proprietary peer-to-peer applications can be recognized using various mechanisms, e.g. pattern recognition.

In business models where the operator has full control over all the terminals and their applications then it is possible to bypass application recognition and allocate the QoS based on terminal requests or markings.

4.2 QoS enforcement

The elements in the different access networks and the backbone network have fairly similar QoS enforcement mechanisms. The basic functions are prioritization and capacity reservations. In IETF standards they are referred to as Differentiated Services (DiffServ) and Integrated Services (IntServ) respectively. Quite often element implementation is a combination of these principles: there will be some resource reservations for parts of the traffic and prioritization will be employed for the scheduling.

Many of the network elements and resource management algorithms contain the following traffic management functions or a sub-set:

- Admission control and bit rate guarantees
  Admission control estimates whether there will be sufficient resources for the new connection or the traffic flow. Admission control functionality is mandatory when implementing guaranteed bit rate connections. Some network elements will also have configurable minimum bit rates for non-guaranteed connections. The network does not need to provide admission control functionality to asynchronous connections that do not have any bit rate guarantees.
  In mobile networks the resource reservations are connection specific: in the backbone and fixed networks it is possible to reserve resources for aggregated traffic.
  MLPS and RSVP are commonly used resource reservation protocols in IP networks.

- Prioritization
  Queuing systems in various elements enforce prioritization. In the air interface there is a packet or frame scheduler that prioritizes the data. In addition the transport resource management algorithms or multiplexing algorithms can be used. The choice will depend on the transmission technology.

- Traffic shaping and policing
  Operators can deliberately limit the throughput of a connection by buffering the data in order not to exceed pre-defined maximum bit rate (shaping), or by dropping packets that would exceed the maximum bit rate (policing). Traffic shaping and policing can also be used within queuing systems as congestion control mechanisms.

- Routing and load balancing
  QoS profiles and markings can be used when selecting a route in ring or meshed topologies.

- QoS profile specific performance counters
  QoS specific counters will typically have the same granularity as the traffic differentiation solution.

4.3 QoS management

QoS management is about monitoring the performance of traffic having different QoS profiles and configuring the QoS differentiation related parameters in the user plane elements. These two functionalities will be combined in the future to form an optimization loop. Based on the monitoring result, the management tools may propose re-configurations to the network elements.

4.3.1 Quality targets and QoS monitoring

There are many dimensions to quality monitoring, as shown in figure 7. QoS differentiation enhances the operator's monitoring capabilities and it provides more accurate measurements. Moreover, when this information is obtained from the many different sources the aggregated result allows operators to predict end-to-end quality and to optimize the network.

The main quality monitoring areas related to QoS differentiation are:

- Connection performance
  Measurements show end-to-end delay, jitter, throughput, packet loss, etc, for individual connections. The quality of a connection or application can be measured with great accuracy using terminal probes. More sophisticated probes can use protocol level information.
- QoS profile specific counters
  The QoS profile specific counters are typically implemented together with queuing systems or resource management algorithms. QoS profile specific counters give aggregated information on the performance in a particular network element. For example, traffic volumes, average throughput and average packet losses can be seen separately for each QoS profile.
- Connection specific control plane counters
  These are available in the networks that rely on handling individual connections such as in UTRAN and LTE. In these cases the network creates Connection Data Records (CDRs) when the connection is established, modified or terminated and also when intersystem handovers occur.
- Network traffic analysis
  These mechanisms produce statistics on the usage of the applications and services. The analysis mechanism is typically based on deep packet inspection, peer-to-peer heuristics, SIP or corresponding signaling.
- Application server statistics
  These statistics contain a lot of useful information on application usage.

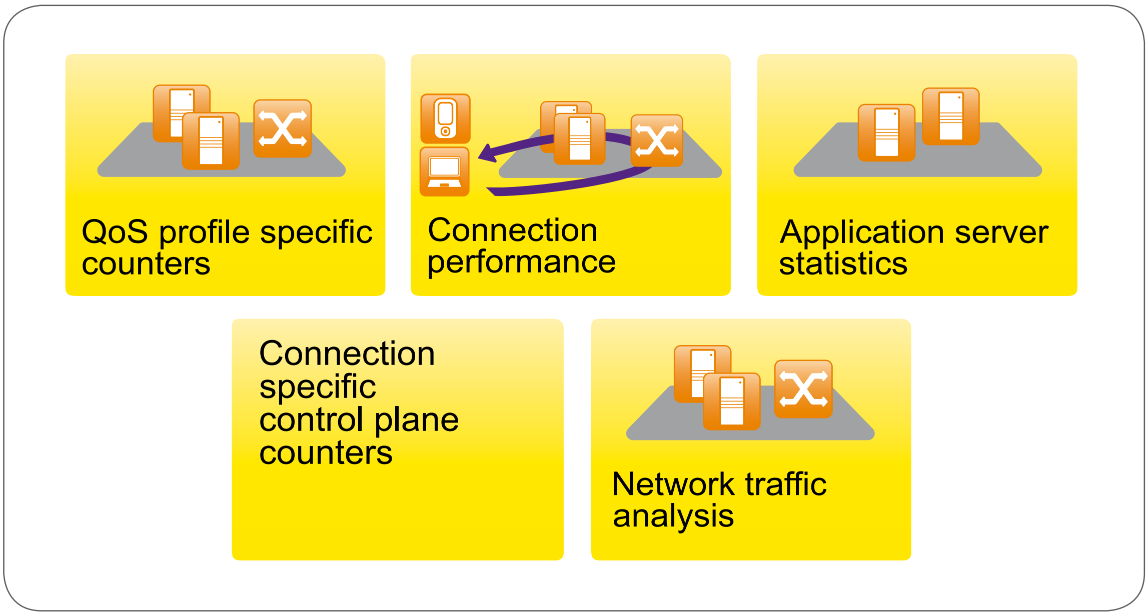


Figure 7. The key dimensions to QoS monitoring. The information that is obtained from these sources is aggregated in order to optimize network performance.

4.3.2 Network QoS optimization loop

The QoS profile specific counters are the most important contribution to QoS differentiation with respect to quality and network resource optimization. The optimization loop (see figure 8) should be run at regular intervals in order to ensure that optimized performance is retained when requirements change. Quality problems should be solved in most cases by tuning the network element parameters or by capacity upgrades. QoS differentiation scenarios and the performance targets should be considered when dimensioning the network. Other important issues to consider are the traffic mix, existing SLAs, terminal capabilities and the network element capabilities.

Once the network dimensioning and the appropriate element configuration have been done the operator can monitor the traffic. The counters and key performance indicators (KPIs) are visualized and checked against the target values. In case the targets are not met, the management system will calculate whether meeting the targets is possible with the existing resources or whether a network upgrade is required. Realizing the resource management challenge can also be achieved by reconfiguring the network element parameters that affect performance and quality. Changing the QoS strategy is also possible and after changes have been made the QoS optimization loop is re-run.

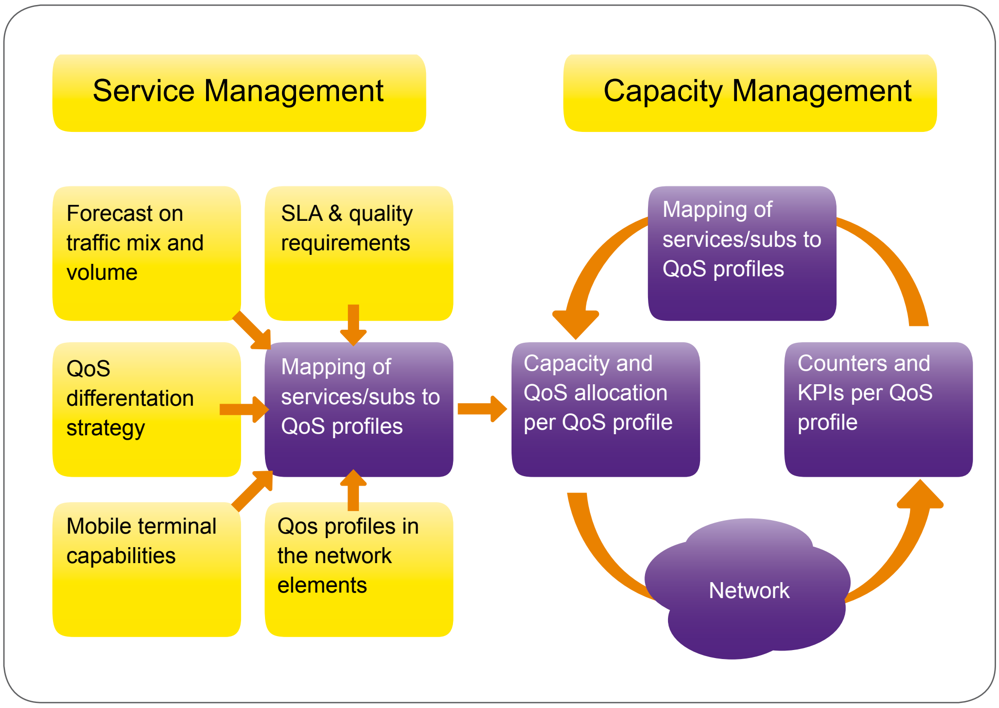


Figure 8. The QoS optimization loop is shown on the right. Service management parameters determine the way services and subscribers map to QoS profiles and the optimization process is based on this information. When this information changes the optimization process is repeated.

5. Conclusions

This paper outlined QoS differentiation solutions that Nokia Siemens Networks is developing and implementing incrementally. Full implementation is anticipated by 2015, but this is also an on-going process. Additional functionality will be provided in line with future needs.

QoS differentiation provides numerous, flexible ways of allocating and managing network resources. It is also used to enhance performance-monitoring capabilities. In addition, QoS mechanisms can be used to support different pricing scenarios and thereby influence subscriber usage and retention.

Video-related services and the abundance of the Internet based applications and content is already driving huge increases in traffic. This creates pressure for network operators to invest in capacity upgrades and to manage existing resources more efficiently. In some cases over-dimensioning the network is the best option. In others capacity upgrades may be too expensive or even impossible due to the regulation etc. In these cases QoS differentiation should be used.

The best QoS solution depends on the operator’s business models and strategy. Nokia Siemens Networks will offer a comprehensive portfolio that will allow operators to build the optimum QoS differentiation solution. The company will also offer consulting services to help operators pick the most suitable components for simple and efficient solutions.

QoS solutions require functionalities in three different areas: control, enforcement and management. QoS control covers the intelligence needed for assigning the desired QoS for a connection or flow. Nokia Siemens Networks has factored in the increasing number of applications and their diversity as well as the increase in terminals (numbers and types) when designing the control mechanisms. QoS enforcement refers to mechanisms that handle traffic in the different elements. The basic enforcement mechanisms are prioritization and resource reservations. QoS management is about visualizing QoS profile specific counters and KPIs, as well as using QoS profile specific information for further network optimization.

Many QoS mechanisms have been standardized and implemented. However, the need for the additional functionality that is provided by differentiation solutions will become more relevant as traffic levels rise. By 2015 a 100-fold increase is anticipated. Carrier Ethernet solutions are being deployed in order to accommodate this additional traffic and Nokia Siemens Networks is a leading supplier of this innovative technology.

6. Glossary

3GPP	Third Generation Partnership Project
AGW	Access Gateway
ASN GW	Access Service Node Gateway
BNG	Broadband Network Gateway
CDR	Connection Data Record
CS	Circuit Switched
DiffServ	Differentiated Services
DSCP	DiffServ Code Point
ETSI	European Telecommunication Standards Institute
GGSN	Gateway GPRS Supporting Node
GPRS	General Packet Radio Service
HLR	Home Location Register
IETF	Internet Engineering Task Force
IMS	IP Multimedia System
IntServ	Integrated Services
ISN	Intelligent Service Node
KPI	Key Performance Indicator
LTE	Long Term Evolution of UMTS Radio Access
MPLS	Multi Protocol Label Switching
NRT	Non Real Time
P2P	Peer-to-Peer
PCS	Policy Control Server
PDN GW	Packet Data Network Gateway
PS	Packet Switched
QoS	Quality of Service
RSVP	Resource Reservation Protocol
RT	Real Time
SAE	System Architecture Evolution
SIP	Session Initiation Protocol
SLA	Service Level Agreement
SLO	Service Level Objectives
SLS	Service Level Specification
TCP/IP	Transport Control Protocol/Internet Protocol
UMTS	Universal Mobile Telecommunications System
UTRAN	UMTS Terrestrial Radio Access Network
VIP	Very Important Person
VoIP	Voice over IP
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network

Nokia Siemens Networks Corporation
P.O. Box 1
FI-02022 NOKIA SIEMENS NETWORKS
Finland

Visiting address:
Karaportti 3, ESPOO, Finland

Switchboard +358 71 400 4000 (Finland)
Switchboard +49 89 5159 01 (Germany)

Code: 11502 - Nokia Siemens Networks – 05/2008 Activeark