Quality of Service Solutions in HSPA RAN

A technology guide and vision on getting the most quality out of the network

Nokia Siemens Networks

Resume

Differentiation solutions are a key component of the Nokia Siemens Networks vision of creating the individual communications experience. Quality of Service helps communication service providers in prioritizing the time and resource critical traffic so that their networks are used optimally. On the other hand, QoS differentiation does not replace the need for having good mobile broadband capacity and coverage, but rather, it minimizes the effect of excessive data usage by end users on the network and helps in occasional busy hour congestion situations, thus improving the end user service experience.

Quality of Service differentiation in HSPA RAN is based on 3GPP standards. The solution is fully geared and self-optimizing and it only requires the operator to specify the Quality of Service policy, i.e. what priorities, guaranteed bit rates and delays are required for different services and users. Nokia Siemens Networks' Quality of Service differentiation solutions enable communication service providers to create different business models, which meet the diverse needs of the market place. This paper focuses on describing the technical components of Nokia Siemens Networks' Quality of Service solution in RAN and the differentiation scenarios for operators.

Table of Contents

1. Executive Summary

HSPA networks today see a tremendous growth in the amount of data traffic and this growth is expected to continue. One powerful solution that can efficiently deal with this increase in traffic volumes is by creating differentiation in the High Speed Packet Access RAN.

Different applications and users have different Quality of Service requirements in terms of guaranteed bit rates, delays, etc. This paper describes the Nokia Siemens Networks' Quality of Service (QoS) solution in HSPA RAN, which uses differentiation, based on 3GPP standards. The solution can work without QoS awareness in the User Equipment (UE) and offers the possibility to support the user and/or service differentiation.

The unique solution is fully geared and self optimizing such that the operator only has to specify the policy, i.e. what priorities, guaranteed bit rates and delays are required for different services and users. The operator can effectively control minimum and maximum service level and relative priority of users and services – even separately for up- and downlink. Radio resource management (RRM) algorithms in the RNC and Node B will then enforce the desired policy.

This white paper concentrates on the RAN HSPA solution, whereas the full E2E QoS requirements are explained in the corresponding white paper entitled "Quality of Service requirements in tomorrow's connected world".

2. Introduction

2.1 Traffic Growth and Applications
According to May 2009 figures from GSA (the Global mobile Suppliers Association), there are 267 commercial HSDPA networks in 114 countries. The amount of data carried in HSPA networks has exploded with the introduction of HSPA and flat rate charging. The number of customers is expected to continue increasing as shown in Figure 1. By the year 2015 there will be 5 billion wireless broadband users and HSPA is expected to carry a large part of this traffic.

This means that the efficient utilization of available capacity is essential to maximize the cell capacity while securing QoS guarantees from bandwidth-greedy applications. One way of achieving this is by exploiting the difference in QoS requirements of the different services, which is done through QoS differentiation. Figure 2 shows an example of different services in a live HSPA network. In the network containing QoS differentiation, the connections can be categorized in different segments and then differentiated. Whereas in a network without QoS differentiation all connections are treated equally and are monitored on an aggregate level (Figure 3).

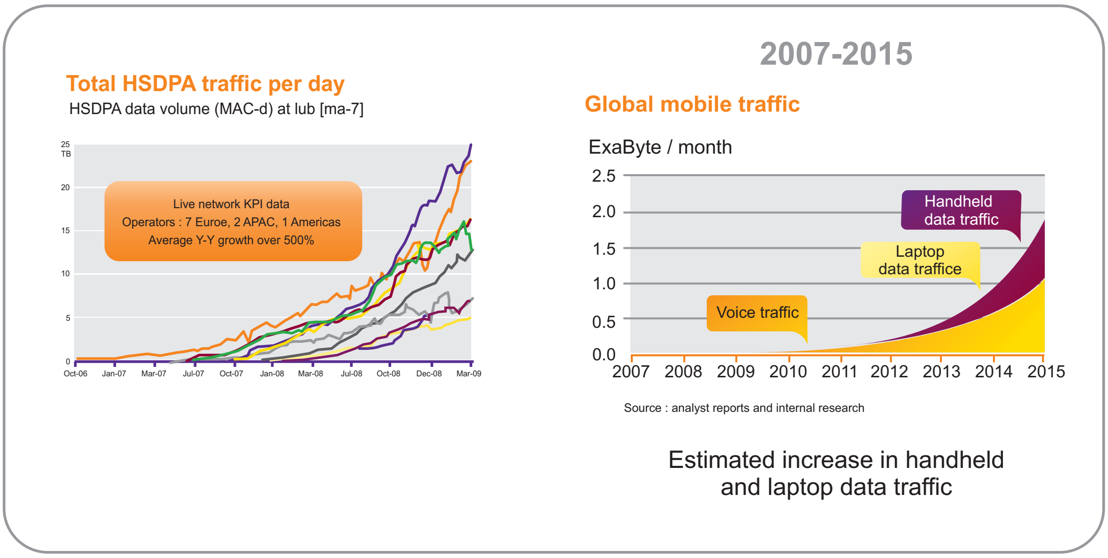

Figure 1. Expected number of customers for the next 7 years

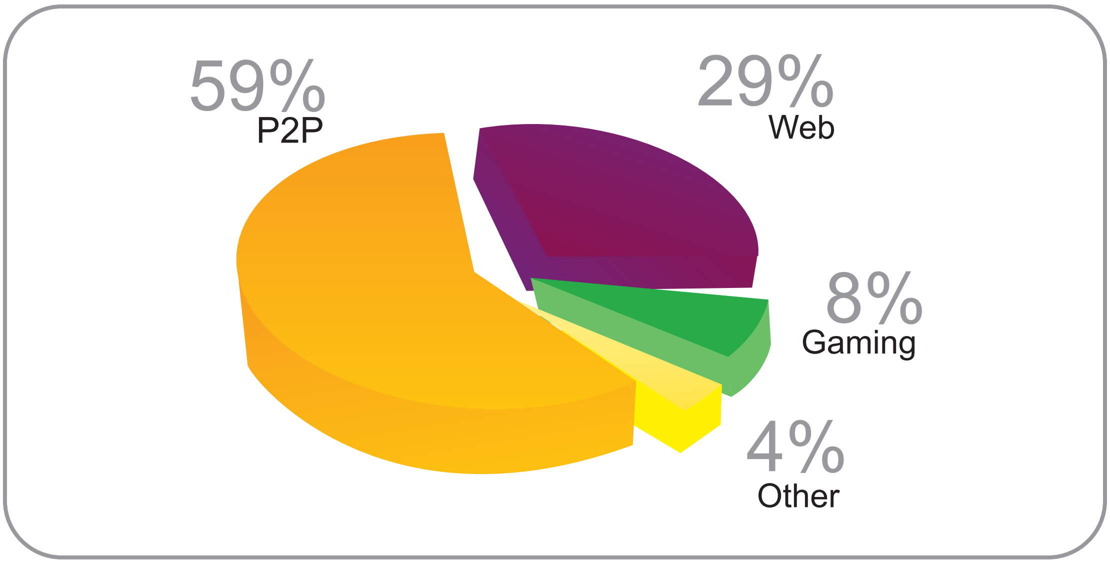

Figure 2. Example of different services in an HSPA network

Quality of Service (QoS) differentiation covers mechanisms to assign different QoS profiles to the connections, treat the traffic differently in case of congestion and monitor different QoS profiles separately. For instance, it can be ensured that streaming and VoIP users always get good quality and ensuring that P2P users are not utilizing resources at the expense of other services. In this document, we concentrate on QoS differentiation in HSPA RAN.

Figure 3. Best effort network and QoS differentiation

2.2 Applications and QoS requirements
Different applications are characterized by different QoS requirements. For instance, real time services like VoIP require tight delay requirements besides a certain guaranteed bit rate, whereas streaming services typically tolerate larger delays, while still requiring a certain guaranteed bit rate. In addition, the services without tight delays and guaranteed bit rate requirements have certain targets in order to meet end user satisfaction, for example, a 'www' user expects a webpage to be downloaded within a limited time. The Table below shows some examples of applications and QoS requirements.

2.3 Targets of QoS Differentiation
There are various targets that can be reached with QoS differentiation solution. Depending on the case, the operator may target one or many of the benefits.

- **Ensuring high (or sufficient) Quality of Service (QoS) to key services.** This is especially important for delay sensitive (real time) applications that typically have a tight delay budget but do not require a large bandwidth. An example of such a service is VoIP. For asynchronous applications on the other hand, high throughput is important, but the delay requirement is much more relaxed. Thus differentiated treatment can be used to guarantee certain services and users, a certain target QoS.

Application	Requirements
Speech	Tight delay per packet (i.e 50-100 ms) Guaranteed Bit Rate
Streaming	Delay requirement according buffer (1-5 s) Guaranteed Bit rate
WWW	Web page downloads within 2-10 seconds
P2P	Pure best effort Downloading typically done in the background

Table 1. Example of QoS requirements and applications

- **User differentiation**
  Prioritizing subscribers does not give any cell level capacity gains but it may support the business strategy well and increase the revenue of the operator. It allows for example, the introduction of different levels of flat fee tariffs to be offered, so that the more expensive subscriptions can receive higher priority, higher peak rates and higher minimum bit rates.

- **Limiting undesired users /applications**
  With a flat rate for data, there are often users who abuse the system and can download hundreds of gigabytes of data per month. A majority of such users are constantly utilizing P2P file sharing applications. It is in the operator's interest to be able to limit either the specific application or the users who extensively consume network resources. The actions towards such users are typically much stronger than a simple user prioritization mentioned above and also include actions taken in the core network.

- **Emergency services**
  With the use of differentiated treatment, emergency services can be given the highest priority and handled in a manner that ensures that the service is guaranteed and reliable.

- **More efficient use of resources**
  With the use of differentiated treatment, one can improve the efficiency of the system and avoid overdimensioning the network. This is illustrated in Figure 4 with a simplified example. On the right side of the figure, three traffic types can be seen. Delay sensitive VoIP traffic, high priority Data 2 and low priority Data 1. When no service differentiation is applied, all the traffic is treated in a similar manner. This means that in order to keep all users happy, all traffic needs to be treated in line with the most delay sensitive traffic such as VoIP. This limits the amount of users that can be served as shown in the right side of Figure 4, since the 'peaks' of the traffic reach the limit of the available cell throughput. When QoS differentiation is used, which is shown on the left side, the low priority data 1 traffic is spread out over time, utilizing the 'gaps' of the other data transmission and this creates room for a new high priority VoIP user. This way, the cell throughput is enhanced by QoS differentiation at the cost of longer delays of delay insensitive traffic.

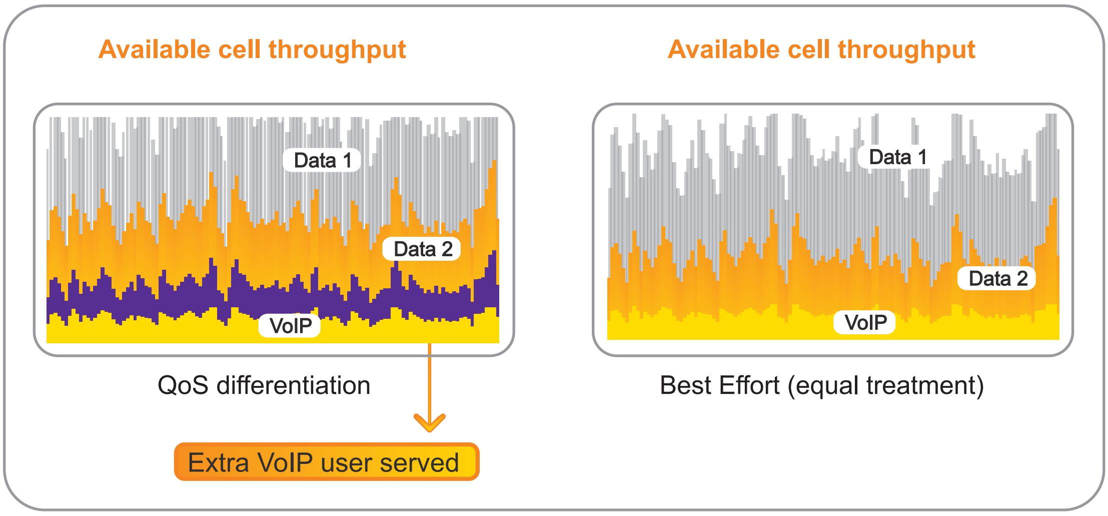

Figure 4. Illustration of cell throughput gains by service differentiation

Summary

By differentiated treatment, the Nokia Siemens Networks' QoS solution in the HSPA RAN ensures sufficient quality for all services while maximizing the number of users served by the available capacity. This includes providing good quality for high value applications, prioritization of premium subscribers over lower priority subscribers and when necessary, overriding lower priority users for time critical applications.

3. Technical Components in RAN

3.1 Overview

QoS differentiation means that different traffic streams are treated according to their requirements. These different streams are identified in the core network. The following parameters are used to notify the Radio Network Controller (RNC) of what traffic is being served:

- **Traffic Class (TC)**
  This can be conversational, streaming, interactive or best effort traffic
- **Allocation Retention Priority (ARP)**
  Can be used to denote user priority with, like for instance Gold, Silver and Bronze users
- **Traffic Handling Priority (THP)**
  Priority per bearer for interactive bearers

The main RRM functionalities for QoS control are located in the RNC and the Node B, as shown in Figure 5. Since the functionalities are distributed over the two functional blocks some information needs to be exchanged.

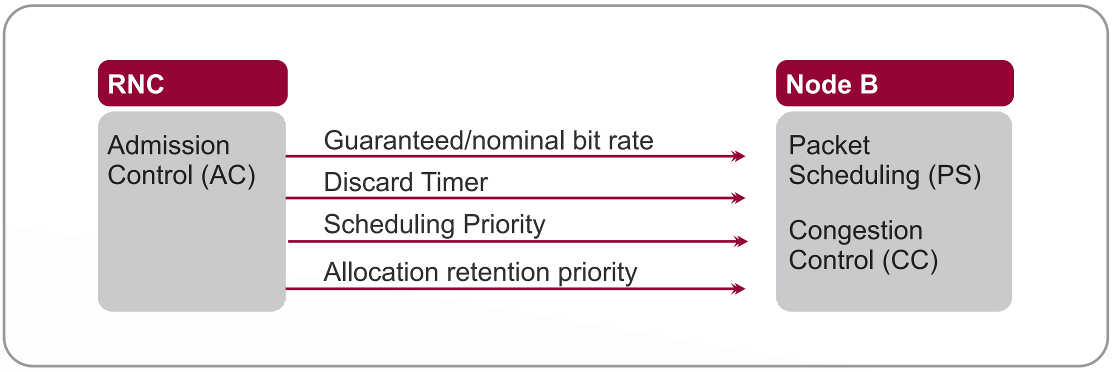

Figure 5. Main QoS funtionalities in RNC and Node B

This information consists of the following:

- **Guaranteed/nominal bit rate**
  requirement, indicating the bit rate which is required for the service

- **Discard Timer**
  The Node B will discard a packet when it is buffered at the Node B for a time which exceeds this timer, i.e. the packet scheduler will try to schedule it before this timer expires

- **Scheduling priority indicator (SPI)**
  priority per bearer, which can be used by the RRM functions in the Node B to prioritize the different bearers

- **Allocation Retention Priority (ARP)**
  priority per user. Can be used for user differentiation

The guaranteed bit rate (GBR) parameter is only applicable for services of the conversational and streaming traffic classes and these can only run on QoS aware terminals, as specified in the 3GPP specifications. In order to have the QoS concept work in non QoS aware terminals as well, the concept of nominal bit rate (NBR) is introduced. The nominal bit rate is a target bit rate for interactive and background users. This way, a non QoS aware terminal can for instance, support a streaming application by setting the appropriate the nominal bit rate, discard timer and priority.

The above-mentioned parameters are set in the RNC, based on the Traffic Class (TC) and Allocation Retention Priority (ARP). An example of settings for a network without QoS aware terminals can be seen in Figure 6, where the different combinations of TC, THP and ARP are mapped to nominal bit rates. After specifying the parameters in a Table, as shown in the example, RRM takes care of enforcing these requirements.

TC+THP+ARP	Priority	UL NBR	DL NBR
Interactive THP1 ARP1	1	128kbps	512kbps
Interactive THP1 ARP2	2	64kbps	256kbps
Interactive THP1 ARP3	3	32kbps	128kbps
Interactive THP2 ARP1	4	64kbps	256kpbs
Interactive THP2 ARP2	5	0kbps	0kbps
Interactive THP2 ARP3	6	0kbps	0kbps
Interactive THP3 ARP1	7	0kbps	0kbps
Interactive THP3 ARP2	8	0kbps	0kbps
Interactive THP3 ARP3	9	0kbps	0kbps
Background ARP1	10	0kbps	0kbps
Background ARP2	11	0kbps	0kbps
Background ARP 3	12	0kbps	0kbps

Figure 6. Example of mapping of TC, THP, ARP to priority, uplink and downlink target bit rates

Besides a nominal bit rate or guaranteed bit rate, a maximum bit rate for uplink and downlink can also be set per user/service in the core network and sent to HSPA RAN in RAB parameters. This maximum bit rate is controlled by the RNC by means of MAC layer mechanism.

The interface between the RNC and the Node B, the Iub interface, has a key role in QoS in cases where air interface is able to provide equal or higher capacity than the Iub. In ATM based solutions, high priority traffic (based on SPI) is mapped to a more stringent ATM class or VCC. In IP based Iub, DSCP code points should be used in line with the SPI to enable harmonized priority treatment over Iub and air interface.

3.2 Admission control

Admission Control is only run for the conversational and streaming traffic class and it is located in the RNC. When a new or a handover call requests a connection, it evaluates how many resources do the current users with a guaranteed bit rate require and estimates the amount of required resources for the new user based on radio measurements. Combining this information provides an answer on whether to admit the call or not.

3.3 RAD Scheduler

The Required Activity Detection (RAD) scheduler is the central RRM functionality in the QoS aware RRM framework and is located in the Node B. The scheduler is able to provide the following tools for the operator:

- Minimum bit rate for high priority users. This ensures that high priority users get at least the specified minimum bit rate, even if the cell is full of lower priority users. Minimum bit rate is important when cells are highly loaded with several users.

- Scheduling weight to determine how capacity is shared between different priority users in low load situations. As an example, high priority user can always be allotted three times the capacity compared to low priority users. Scheduling weight is important when cells only have a small number of simultaneous users.

- Proportional fair scheduling algorithm should still be followed for all users, to ensure that overall cell capacity is not decreasing due to QoS differentiation. According to published simulations, Nokia Siemens Networks' QoS scheduler is more efficient than solutions offered by our competitors.

The RAD scheduler automatically calculates the required scheduling activity for each bearer and schedules it accordingly; eliminating the need to tune any parameters once the QoS parameters per bearer/user are known. The remaining capacity after fulfilling guaranteed and nominal bit rate requirements is distributed among the best effort users according to any operator's policy, i.e. scheduling weights based on the SPI.

3.4 Congestion Control (fast bearer prioritization)

Due to user mobility in the network and since the traffic activity is not easy to predict, there may be some instances where the system is not able to fulfill all of the required QoS of the users in the system. To minimize the effects of this instantaneous overload, the QoS RRM concept includes fast bearer prioritization, which is a kind of congestion control. Fast bearer prioritization at times of congestion, excludes the lowest priority bearers from the group of bearers that can be scheduled. This way, the prioritized bearers can fulfill their required QoS, while only a minimum number of low priority users may temporarily suffer. When the congestion is solved, the throughput of those users is boosted for compensation.

3.5 Performance

The RAD scheduler performance is illustrated in Figure 7 for different traffic mixes of best effort traffic and variable bit rates (VBR) streaming applications with strict GBR constraints. The RAD scheduler strictly serves streaming users with their GBR requirement (128 kbps) regardless of their location in the cell and cleverly distributes the excess capacity among the best effort users. The latter is observed because despite the strict GBR provision reducing the freedom for the packet scheduler, the average cell throughput achieved with only best effort traffic decreases only slightly when more streaming users are present in the cell.

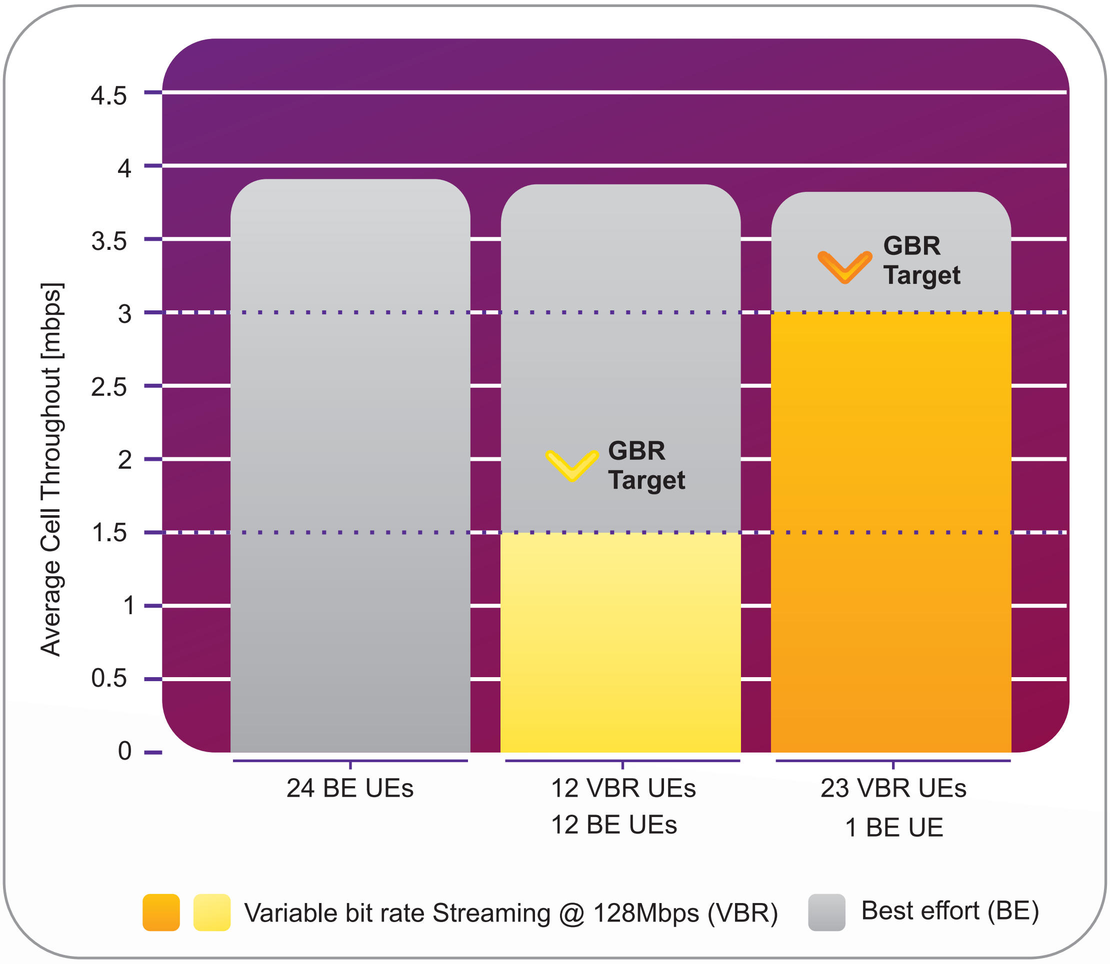

Figure 7. Average cell throughput for different traffic mixes of best effort and VBR streaming traffic.

4. Scenarios

The QoS RRM framework described in the previous pages can be used in many different ways. In the following sections some use cases are described.

4.1 Operator mobile TV quality assurance

In case an operator wants to guarantee a certain minimum quality for mobile TV, the mobile TV bearer needs to be mapped to a high priority bearer and the correct nominal bit rate needs to be set.

4.2 Other operators preferred services

In case of other preferred services, the same procedure can be used; the preferred service needs to be mapped to high priority and the correct nominal bit rate needs to be set. This principle can be used for any service, like for instance push to talk, streaming or VoIP. If the traffic has a certain activity factor, the packet scheduler will automatically take care of this, since the nominal bit rate will only be given when there is data in the buffer.

4.3 Operator Challenge: Increased traffic, but revenue decoupled from capacity requirements

Upon launching HSPA with flat fee, many operators have noticed an exploding traffic growth in their networks. Some operators also have higher revenues from packet data as compared to voice. However, the exploding amount of traffic will not bring a comparable effect on the revenues as flat fees and general price reductions for data usage will apply.

4.3.1 Solution 1: Introduce higher tariff flat fees with higher data volumes

In the core network, the total data volume of each user during the month can be monitored. For all users exceeding a certain data amount, a new lower maximum bit rate and/or a lower priority is set, which is used in the RNC. This approach can be used to sell premium subscriptions, by offering higher amount of data depending on monthly fee. This allows for charging users based on their willingness to pay.

4.3.2 Solution 2: Sell Maximum bit rate parameter

Operators can advertise 500 kbps, 1 Mbps, 2 Mbps etc. subscriptions. For each user, a corresponding value is then introduced for the maximum bit rate parameter in the core network and RNC. The advantage of this is that the higher bit rates may justify higher prices.

4.3.3 Solution 3: Use priority and nominal bit rates in radio access network

Higher priorities and higher nominal or guaranteed bit rates and scheduling weights can be assigned to premium users. Then RRM and especially the scheduling algorithm in the Node B ensures that these users get better service than lower priority users, i.e. they get for example, three times more (operator configurable) resources than the normal user. The nominal and guaranteed bit rates take care that the premium users get a certain minimum service even if the cell is extremely loaded.

4.4 Operator Challenge: How to launch streaming services with no QoS aware terminals?

Streaming services here refer to cases where user is watching a video clip on demand. Video playback starts almost immediately, and content is downloaded as the video proceeds. Bit rate guarantee is needed for this in order to avoid gaps in the video reception.

4.4.1 Solution: Use nominal bit rates to enable streaming in non QoS aware terminals

Operators can define high priority for streaming traffic, regardless of user subscription type (or alternatively only for certain users). For these users, the operator marks a nominal bit rate in the RNC, for example 250 kbps. The Node B scheduler will then ensure that this bit rate is always given to these streaming users in the cell, before any capacity is allotted to other users. This means that HSDPA can be used to offer high bit rate video content to users without actual 3GPP streaming traffic class support in the terminals.

4.5 CS voice/IMS voice over HSPA

The RRM framework for QoS differentiation is prepared to handle CS and IMS voice over HSPA. Mapping the CS bearer to the highest priority and a nominal or guaranteed bit rate and delivering the voice frames within the targeted max delay will enable good CS service over HSPA.

5. Conclusions

This paper outlines the QoS differentiation solution that Nokia Siemens Networks has implemented in the HSPA RAN. This is the most complete QoS solution in the market which supports most of the QoS strategies that operators would like to implement. It can be used to ensure a sufficient end-user experience to premium users or applications (e.g. streaming video). It can also be used to launch premium rate data subscriptions for an enhanced end user experience and higher operator revenues. The solution is ready to support CS voice over HSPA.

QoS differentiation is critical at those interfaces which are most congested. Through Nokia Siemens Networks' QoS solution in the radio interface, operators can accurately define both uplink and downlink distribution of the capacity when different priority users or applications are sharing the cell resources. The paper explains concisely what parameters and RRM functionalities enable full QoS differentiation support in the Nokia Siemens Networks' HSPA RAN solution.
This solution is part of the end to end QoS solution as described in White Paper entitled “Quality of Service requirements in tomorrow's connected world” and consists of packet scheduler, admission control and congestion control. These different functional blocks automatically self adjust their actions based on the input by the operator regardless the load and the heterogeneous traffic in order to support the network. The only input required is the setting of priorities and the services users should have, plus the target bit rate and delay constraints.

The solution is robust and also works with non QoS aware terminals. In the case of non QoS aware terminals, the nominal bit rate is introduced, which is set by the network.

6. Glossary

APN	Access Point Name
ARP	Allocation / Retention Priority
ATM	Asynchronous Transfer Mode
DSCP	Differentiated Service Code Point
DL	Downlink
GGSN	Gateway GPRS Supporting Node
GPRS	General Packet Radio Service
GB	Giga Bytes
GBR	Guaranteed Bit Rate
HSDPA	High Speed Downlink Packet Access
HSUPA	High Speed Uplink Packet Access (same as E-DCH)
HLR	Home Location Register
HSPA	HSDPA + HSUPA
IETF	Internet Engineering Task Force
IP	Internet Protocol
kbps	kilobits per second
Mbps	Megabits per second
NSN	Nokia Siemens Networks
PC	Personal Computer
QoS	Quality of Service
RAB	Radio Access Bearer
RAN	Radio Access Network
RNC	Radio Network Controller
R99	Release '99, refers to non-HSPA traffic
SPI	Scheduling Priority Indicator
SGSN	Serving GPRS Support Node
3GPP	Third Generation Partnership Project
TC	Traffic Class (conversational, streaming, interactive and background)
THP	Traffic Handling Priority
UL	Uplink
WCDMA	Wideband Code Division Multiple Access

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)

C401-00406-WP-200905-1-EN- Nokia Siemens Networks – Alphabet Consulting