Providing Conformance of the Negotiated QoS using Traffic Conditioning for Heterogeneous Services in WCDMA Radio Access Networks *

Frank Yong Li and Norvald Stol
Department of Telematics
Norwegian University of Science and Technology (NTNU)
7491 Trondheim, Norway
E-mail:{frank.li, norvald.stol}@item.ntnu.no

Abstract

In order to provide negotiated QoS to the end users in next generation mobile networks, a traffic conditioner may be deployed to provide conformance of a service agreement. The traffic conditioning is performed by traffic shaping or/and policing. A framework of applying traffic conditioning in radio access network is proposed in this paper. The main idea of our traffic conditioning approach is to employ traffic shaping at each User Equipment (UE) and traffic policing at the Radio Network Controller (RNC). The traffic generated by the application is regulated by the token bucket traffic shaper at each UE, and the conformance of the traffic is policed at the RNC according to traffic policing policy. We investigate the impact of traffic shaping at various system load, as well as the tradeoffs between concerned QoS parameters for heterogeneous traffic classes. A system level simulation model based on the proposed framework is implemented. The numerical results regarding shaping delay, probability of non-compliance and packet loss ratio are also presented in this paper.

1. Introduction

Regardless of which multiple access standard will be widely deployed (e.g., CDMA2000 or UMTS), the next generation radio access system is a Wideband Code Division Multiple Access (WCDMA)-based system, and the external core network could be Internet Protocol (IP)-based. To provide negotiated Quality of Service (QoS) to heterogeneous traffic classes in such a system, traffic control and resource management play an important role.

Regarding the end-to-end QoS provision scenarios, no matter whether it is based on Integrated Service (IntServ) [7], Differentiated Service (DiffServ) [5] or a mixture of them, the user perceived QoS can be negotiated according to FlowSpec. The FlowSpec is usually presented by 5 elements (r, b, p, m, M), i.e., token rate r, bucket size b, peak rate p, minimum policed unit m and maximum packet size M. Using these parameters, a traffic conditioner can regulate the incoming traffic flow with a traffic shaper and check

* This work is funded by Telenor R&D.

the conformance of the shaped packets with a traffic policer, thus providing conformance to the agreed Service Level Agreement (SLA).

The concept of traffic shaping has existed for years [6]. There exists mainly two traffic shaping algorithms, namely Leaky Bucket Algorithm defined by ATM Forum and Token Bucket (TB) Algorithm defined by IETF respectively. The latter has been adopted as a reference algorithm for traffic conditioner in Universal Mobile Telecommunication System (UMTS) by 3rd Generation Partnership Project (3GPP) [1]. Traffic conditioning is well understood in the IP community, but how to apply the idea in 3G mobile networks is still a challenge. Although currently there are lots of research work as well as standardization activities in this area [1], [2], [3], [8], very little literature can be found on how traffic conditioning performs in radio access networks, especially between the UE and the RNC.

This paper addresses this issue based on our recent study in traffic conditioning and QoS management in UMTS. A traffic conditioning scenario is proposed according to our system model in this paper. The main idea is to employ traffic shaping at individual UE and apply traffic policing at the gateway node, i.e., RNC in our case. We study the necessity of traffic shaping at the UE, compare different shaping schemes and discuss the trade-off between concerned QoS parameters for heterogeneous traffic classes. A system level simulation model is implemented and corresponding simulation results are presented.

The remainder of the paper is organized as follows. The system consideration is outlined in Section 2, while the traffic conditioning scenario in radio access network is presented in Section 3. After a description of the simulation model in Section 4, the numerical results from our simulations are given in Section 5. Finally, the conclusions are given in Section 6.

2 System Description

After a brief description of the system model, we present the fundamental characteristics of the concerned traffic classes and show how congestion is calculated for a CDMA channel.

2.1 System Model

A traffic conditioning-enabled Radio Network Subsystem (RNS), which consists of a RNC and one or more Node Bs, is illustrated in Fig. 1 as our system model.The UEs are distributed within each cell with an omnidirectional base station at the cell center. Neither handover nor mobility is considered in this model.

Each user is expected to use either one of the considered traffic classes. A traffic shaper is employed for uplink traffic flow on each UE. The packets received at the Node Bs are forwarded to the RNC and the traffic policing function is manipulated at the RNC. Our work is focused on the user plane and the relevant signalling protocols between the UE and the RNC or between RNCs are beyond the scope of this paper.

2.2 Heterogeneous Traffic Classes

The 3G WCDMA systems support a variety of bearer traffic classes for diverse multimedia applications. Both real-time Variable Bit Rate (rt-VBR) and non-real-time VBR (nrt-VBR) traffic classes [7] are considered in our studied model. The rt-VBR traffic category like video has tightly constrained delay and requires a uniform frame transmission

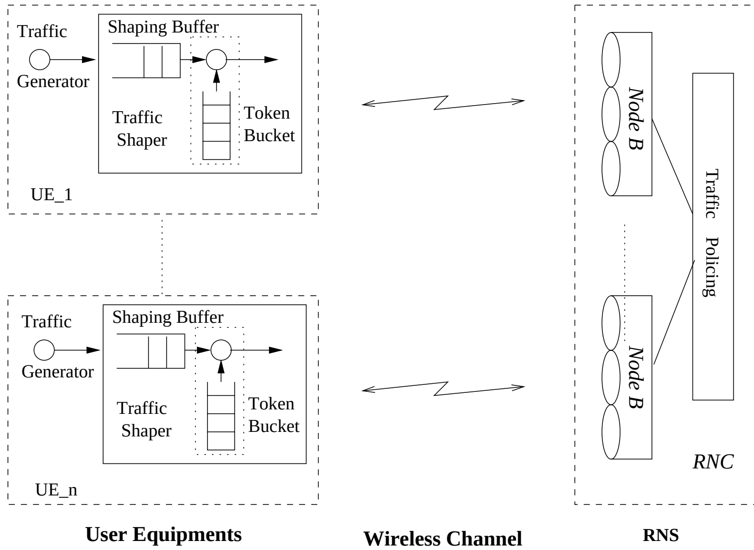

Figure 1. System Model: A Traffic Conditioning-enabled Radio Access Network

rate. A user receiving rt-VBR type of service has stringent requirement on delay, but may tolerate certain amount of packet loss. The nrt-VBR category, on the other hand, does not have tight constraint on delay. The traffic may be more bursty, and allows greater flexibility in delay requirement. But many nrt-VBR services have very strict constraint on packet loss.

We will take these fundamental characteristics of the different traffic classes into account when we investigate the performance of our traffic conditioning scenario in radio access networks.

2.3 ’Congestion’ Calculation on CDMA Channel

One fundamental characteristic of a CDMA-based mobile system is its interference-limited ’soft’ capacity. All users in the cell share the same bandwidth and interfere with each other. The CDMA channel will not really be ’congested’ as load increases, but the quality of all ongoing connections will be degraded by the newly added-on users instead.

As there is no ’congestion’ definition for a CDMA channel, we employ the well known Pole capacity for our ’congestion’ calculation of a CDMA channel.

Assuming perfect power control and negligible background thermal noise, the individual load of service i denoted as η_i can easily be calculated by [4]:

$$\eta_i = \frac{(E_b/N_o)_i}{W/R_i} \cdot \nu_i \cdot (1+f) \tag{1}$$

where $(E_b/N_o)_i$ is bit energy to noise ratio required for desired BER of service i, W is the chip rate (3.84Mcps in 3G WCDMA systems), R_i is the information bitrate of service i, ν_i is the activity factor of service i, f is the interference factor from adjacent cells (i.e., ratio between adjacent cell interference and own cell interference).

The total load η_{total} in a cell is the sum of the individual loads with different services from all existing connections (only uplink is considered here):

$$\eta_{total} = \sum_{i=1}^{N_u} \eta_i \tag{2}$$

where N_u denotes the number of calls currently in service.

We define that the channel is 'congested' if the total load η_{total} exceeds the pre-defined threshold (i.e., the maximum allowed load).

To set up a manageable congestion control mechanism for the CDMA channel in our study, we further simplify equation (1) and define an equivalent bandwidth BW_{equ}^{i} for each individual service type i, as follows:

$$BW_{equ}^{i} = (E_b/N_o)_i \cdot R_i \cdot \nu_i \cdot (1+f) \quad (3)$$

Using this simplification, we define that the channel is congested if $\sum_{i=1}^{N_u} BW_{equ}^{i} > 3840$ Kbps.

3 Traffic Conditioning Scenario in Radio Access Network

A traffic conditioner provides conformance between the negotiated QoS for a service and the data unit traffic. Traffic conditioning is performed by policing or/and by traffic shaping [1]. After a brief review of the token bucket algorithm, this section describes the proposed traffic conditioning scenario in radio access network, i.e., how we employ traffic shaping at each individual UE and traffic policing at the RNC. Two versions of the token bucket algorithm are applied to the traffic shaper at the UE.

3.1 Token Bucket Algorithm

The token bucket algorithm regulates the bursty traffic in such a way that over a long time period the average allowed rate approaches the desired token rate r asymptotically, and over a short time interval the burst size of the traffic is upper bounded by bucket size b.

Figures 2 and 3 depict the *standard* TB algorithm from [1] and our implementation of the token bucket algorithm respectively. Here Token Bucket Counter (TBC) is an internal variable used to record the number of the remaining tokens at any time. When a new packet arrives, three possible measures exist for further decision by the traffic shaper.

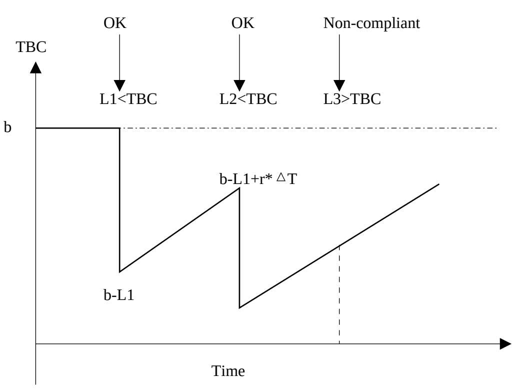

Figure 2: Token Bucket Algorithm: Traffic Tagging Scheme

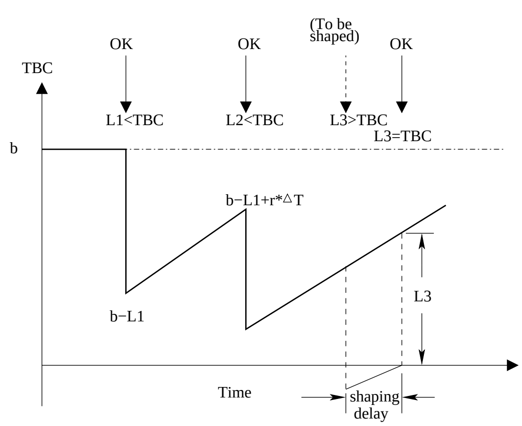

Figure 3: Token Bucket Algorithm: Traffic Conforming Scheme

- Case 1: conformed at arrival. The size of the incoming packet is smaller than the TBC value in the bucket when the packet arrives. The packet is judged as compliant and marked as a compliant packet;
- Case 2: non-conformed. A packet j is deemed as non-compliant if its length L_j is larger than the bucket size b, i.e., $L_j > b$. This type of packet is marked as not-compliant. It will not be dropped at the UE, but left for preferential discarding at the traffic policing point.
- Case 3: shaped to be conformed. The size of packet k is not larger than the bucket size, but larger than the TBC value at arriving instant, i.e., $b \geq L_k > TBC$. In this case, the packet will be enforced to wait by the traffic shaper until there are enough tokens in the bucket. This type of packet is shaped to be conformed and marked as compliant when it is sent out from the bucket. We define the time period when a packet is waiting for tokens in the TB as *shaping delay*, denoted by D_s.

3.2 Traffic Shaping at the UE

According to the token bucket algorithm described in subsection 3.1, the amount of data sent $D(T)$ over any interval of time T obeys the rule [2]:

$$D(T) \leq r \cdot T + b \tag{4}$$

In words, conformance according to a token bucket can be defined as: ”Data is conformant if the amount of data submitted during any chosen time period T does not exceed $(rT + b)$”.

In our UE traffic shaping approach, we introduce two traffic shaping schemes, with different treatment of non-compliant traffic. The scenarios are illustrated in Figures 2 - 4 and explained as follows:

3.2.1 Scheme 1: Traffic Tagging

In this scheme, a packet not matching the relevant token bucket parameters at arriving instant will simply be marked as non-compliant. We refer to this scenario as *traffic tagging*, as shown in Fig. 2. A packet tagged as non-compliant will not be dropped at the UE, but has higher probability of being discarded later at any other network nodes.

In this case, the traffic is not shaped because no decision has been made on those non-compliant-at-arrival packets. But we keep this scheme as one part of our traffic shaping study.

3.2.2 Scheme 2: Traffic Conforming

Another scheme, referred as *traffic conforming*, provides conformance to the incoming packets. Instead of only tagging the non-compliant packets as not matching, the packets is shaped to be compliant. As a result, all packets smaller than bucket size will be sent out as conformed. The flow chart of this scheme is depicted in Fig. 4. This scenario will of course introduce additional traffic shaping delay to specific packets even before they are transmitted. This enforced shaping delay by the traffic shaper will consume part of the total end-to-end delay budget of a service, but will hopefully lead to lower packet loss ratio. One objective of studying this scheme is to investigate the tradeoff between shaping delay and packet loss ratio.

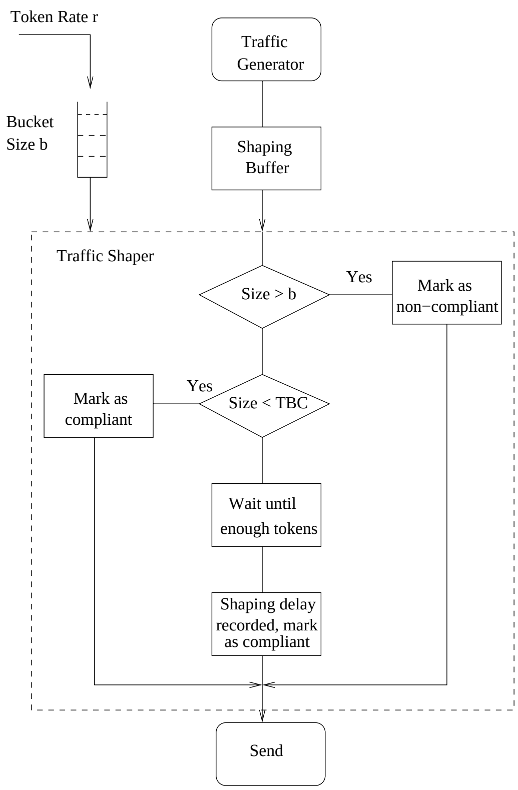

Figure 4: Traffic Shaping (Traffic Conforming) at the UE

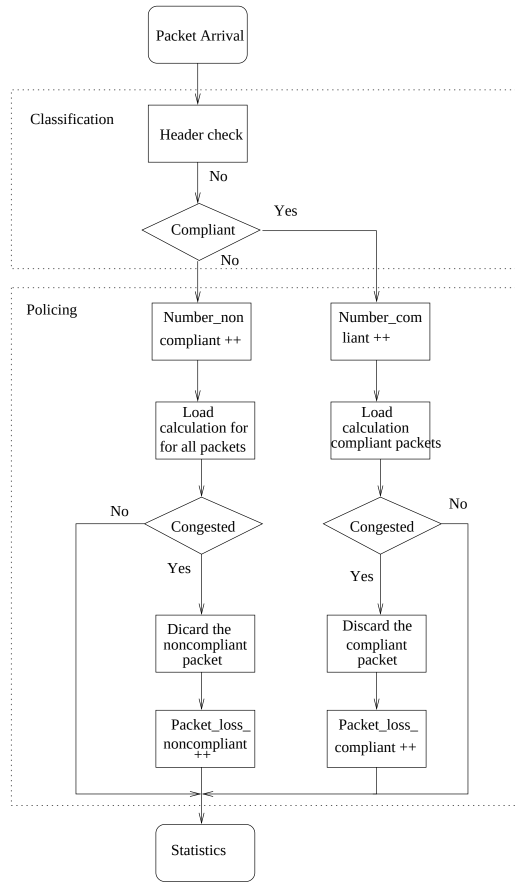

Figure 5: Traffic Policing at the RNC

No matter which traffic shaping scheme is applied, the regulated traffic after shaping will be categorized into two types: compliant or non-compliant packets. They are left for further policing at the RNC.

3.3 Traffic Policing at the RNC

The traffic policing function compares the conformance of the user data traffic with the pre-defined service agreement. In the Bearer Service (BS) manager, policy control is a logical policy decision element which is optional to the UEs and required to the Gateways [2]. Traffic policing policy in practice is a network operator choice. More detailed description of the policy framework can be found in [9].

The traffic policing function considered in our study is depicted in Fig. 5, where the traffic policing element is only embodied at the RNC. We assume that all packets have passed the Call Admission Control (CAC) phase and the policing function used here is composed of traffic classification and policing.

3.3.1 Classification Function

Upon reception, the classification function simply checks the header of every received packet and forwards the corresponding information for further policing, according to its service class as well as traffic conformance status.

3.3.2 Traffic Policing Policy

Depending on the parameters of a packet forwarded from the classification function, the policing element at the RNC will perform corresponding policy on each individual packet. Only very simple policy based on channel congestion status is studied in this paper. More comprehensive polices are left for further study. As shown in Fig. 5, the non-compliant data packets will be acknowledged only if the channel is not congested, i.e., the total interference level by summing of all ongoing calls does not exceed the threshold. Otherwise they will simply be discarded. For compliant data packets, the load calculation will take into account both compliant and non-compliant packets at first. If the load exceeds the threshold, the non-compliant packets will be dropped preferentially. A conformed packet could also be discarded if there are too many concurrent compliant packets on the channel.

4 Simulation Configuration

A system level simulation model has been built on OPNET simulator. The heterogeneous traffic classes include rt-VBR and nrt-VBR as described in subsection 2.2. The equivalent bandwidth is quantified in this section and the relative load used to represent the system load is defined. The TB parameters acquisition algorithm is formulated in subsection 4.3.

4.1 On-Off Traffic Model

The On-Off traffic model applies to both traffic classes in our simulation. We assume that packets are generated at the application layer according to a packet length distribution. Even though the packets could arrive at any time instant according to the interarrival time distribution, they are buffered in the queue and a new packet will not be sent out until the preceding one has finishes its service. In other words, the same UE cannot send more than one packet at the same time. No segmentation happens to the generated packets and a protocol header at IP layer is appended to packets before traffic shaping.

With On-Off traffic model, *On* represents transmission at peak rate R_p. The average rate R_a can be obtained by $R_a = \frac{T_{on}}{T_{on}+T_{off}} \cdot R_p = \frac{T_{on}}{\Delta T} \cdot R_p$, where $\Delta T = T_{on}+T_{off}$ is the interarrival time. For both traffic classes, we assume that the packet size is Pareto(m, α, M) distributed with cut-off M, where m denotes the minimum size and α is the shape factor. The cut-off is set to be equal to well-known MTU size of 1500 bytes (12000 bits). The difference between rt-VBR and nrt-VBR in our model is that the former has constant interarrival time while the latter has exponential interarrival time. The traffic models used in simulation are parameterized and listed in Table 1.

Table 1. Traffic Model used in Simulation

Class	$R_p(Kbps)$	$R_a(Kbps)$	$\Delta T(ms)$	$T_{on}(ms)$	Pareto size (bits)
rt-VBR	28	40	uniform (135.86)	95.10	(1864, 1.7, 12000)
nrt-VBR	30.4	64	expon. (75.64)	35.93	(652, 1.1, 12000)

4.2 Equivalent Bandwidth and Relative Load

Applying the traffic parameters listed in Tab 1 into our equivalent bandwidth definition, the concrete values we use in simulation are tabulated below. Here the interference factor from adjacent cells f is set to be 0 and the maximum users means the maximum number of users in one cell from the same traffic class.

Table 2. Equivalent Bandwidth in Simulation Model

Class	$(R_p)_i(Kbps)$	ν_i	$(E_b/N_o)_i(dB)$	$BW^i_{equ}(Kbps)$	Maximum users
rt-VBR	40	0.699	7.8	241.04	15.93
nrt-VBR	64	0.475	6.7	299.33	12.83

The performance of the concerned system is evaluated at various network load. We introduce a relative load, denoted by $\bar{\eta}$, to represent how heavy the CDMA channel is loaded, defined as follows:

$$\bar{\eta} = \frac{\sum_{i=1}^{N_u} BW^i_{equ}}{W} = \frac{\sum_{i=1}^{N_u} (E_b/N_o)_i \cdot R_i \cdot \nu_i \cdot (1+f)}{W} \tag{5}$$

where activity factor ν_i is decided by $\nu_i = \frac{T_{on}}{T_{on}+T_{off}} = \frac{T_{on}}{\Delta T}$. In simulation, we allow the relative load $\bar{\eta}$ to exceed 1 to a certain degree due to the 'soft' capacity property of the CDMA channel.

Table 3 lists the relationship between the number of users in the cell N_u and the relative load $\bar{\eta}$ for both traffic classes.

Table 3. Relative Load $\bar{\eta}$

rt-VBR	N_u	16	18	20	22	24	26
	$\bar{\eta}$	0.702	0.789	0.877	0.965	1.053	1.140
nrt-VBR	N_u	18	20	22	24	26	28
	$\bar{\eta}$	0.666	0.740	0.815	0.889	0.963	1.037

4.3 Parameters Acquisition Algorithm for Token Bucket Traffic Shaper

Even though token bucket is a well recognized shaping algorithm, determination of the TB parameters is not a simple task. It is application dependent and there is no standardized algorithm for deciding r and b. 3GPP has recommended in [2] to use average bitrate for calculating token rate r and peak bitrate for calculating bucket size b, while a protocol header is included at the same time. Let δ_T denote the sampling interval of the source data (the inverse of δ_T is the number of frames per second), we formulate the algorithm as follows:

$$r = \frac{R_a \cdot \delta_T + L_h}{\delta_T} = R_a + \frac{L_h}{\delta_T} \tag{6}$$

$$b = R_p \cdot \delta_T + L_h \tag{7}$$

Empirically we find the bucket size defined by equation (7) is too conservative and may not be suitable for certain type of traffic. The bursty property of the traffic flow has not been considered. Therefore, we propose to define the bucket size with the following equation:

$$\hat{b} = b \cdot \beta = b \cdot \frac{R_p}{R_a} \tag{8}$$

where β denotes the burstiness of a traffic flow and is defined as $\beta = R_p/R_a$.

The obtained token bucket parameters for both traffic models are listed in Table 4, where we assume a compressed RTP/UDP/IP header (4 octets) for rt-VBR and a TCP/IP header (40 octets) for nrt-VBR.

Table 4. Token Bucket Parameters obtained for Simulation

Class	$R_a(Kbps)$	$R_p(Kbps)$	$L_h(bits)$	$r(Kbps)$	$b(bits)$	$\hat{b}(bits)$
rt-VBR	28	40	32	28.235	5754	8221
nrt-VBR	30.4	64	320	34.631	5161	10865

5 Numerical Results

Based on the simulation model described in the above section, we present a number of numerical results in this section. All simulation results shown below are carried out with 10 random seeds, which give us a 95% confidence interval using student-T distribution. Only mean values are used for Tables 5 and 6. As an excerpt of the statistics, the number of packets generated during the simulation is about 50,000 - 110,000, depending on the traffic load.

Throughout the context, the discarding probability P_d (referred to also as packet loss ratio throughout the context) is defined as the total number of packets discarded at the RNC divided by the total number of packets with the same marking received at the RNC. For example, the discarding probability for non-compliant packets P_d^{non} is the number of discarded non-compliant packets divided by the total number of packets marked by the traffic shaper as non-compliant. The shaping delay D_s is the average value by all generated packets, and π_{non} is the probability of non-compliance among all generated packets.

5.1 Discarding Probability With versus Without Shaping

Our first observation is to inspect the impact of traffic shaping by comparing the discarding probability for both cases. Traffic conforming is applied as the shaping algorithm. The simulation results are depicted in Figures 6 and 7. To make a fair comparison, we set a bucket size of $b = 12000$ bits. This implies, with traffic conforming scheme, that all generated packets will be conformed after shaping. Without shaping scheme means there is no token bucket attached to the traffic generator.

It is shown, for both traffic classes, that the discarding probability with traffic shaping is smaller than P_d without traffic shaping, for all ranges of system load. The heavier the system load, the larger the difference. When the traffic load is light (relative load $\hat{\eta} < 0.7$), very few packets are lost for both cases. This means that for light traffic load, there is no need to do traffic shaping, because the introduced shaping delay does not help with the packet discarding probability. But as the traffic load becomes heavier, the regulated packets are more likely to survive over the wireless channel. The heavier the system load is, the more important it is to use traffic shaping for bursty traffic.

5.2 Comparison of Traffic Tagging and Traffic Conforming

Having illustrated the benefit introduced by traffic shaping, we compare in this subsection the performance of the two shaping schemes described in section 3.

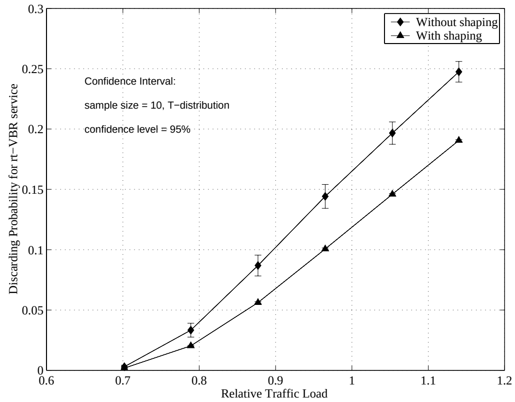

Figure 6: Discarding Probability for rt-VBR Class

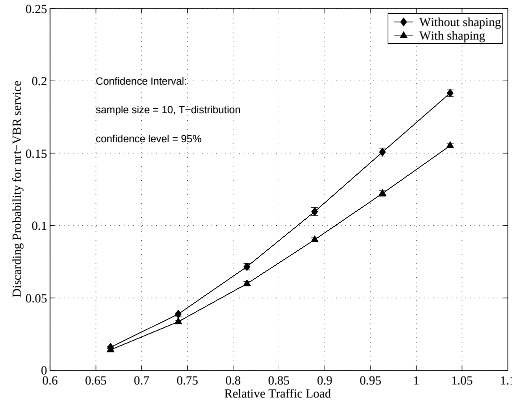

Figure 7: Discarding Probability for nrt-VBR Class

We use token bucket size decided by equation (7) to make the evaluation and the system load is kept closed to 1. The average values of the simulated results are listed in Table 5. As explained earlier, no shaping delay is introduced with traffic tagging, but more packets are classified as non-compliant, i.e., π_{non} is larger with tagging. But in addition to a small shaping delay, traffic conforming also leads to a slightly higher discarding probability. As there are much more non-compliant packets with traffic tagging, from traffic generator's perceptive, there will be more packets possibly dropped during transmission. By comparison, we conclude that traffic conforming is more suitable for loss-sensitive type of traffic (e.g., data) and traffic tagging is more suitable for delay-sensitive type of traffic (e.g., audio).

Table 5. Comparison of Traffic Shaping Schemes

	Scheme	$\bar{\eta}$	$b(bits)$	$\pi_{non}(\%)$	$D_s(ms)$	$P_d^{conf}(\%)$	$P_d^{non}(\%)$
rt-VBR	Tagging	0.965	5754	17.35	0	0.42	25.23
	Conforming	0.965	5754	14.61	1.37	0.74	26.68
nrt-VBR	Tagging	0.963	5161	13.43	0	0.49	26.27
	Conforming	0.963	5161	10.18	1.76	1.01	29.07

5.3 Tradeoff between Shaping Delay and Probability of Non-compliance

Keeping the token rate as defined by formula (6) and allow bucket size b to be variable, we observe the impact of b on probability of non-compliant packets π_{non} and shaping delay D_s in Figures 8 and 9. Again, traffic conforming is applied as the shaping scheme. This means that only packets larger than b will be judged as non-compliant. Based on our traffic model, π_{non} is mathematically [1] available. The simulations show exactly the same results for both traffic classes as depicted in Fig. 8. The figure shows that π_{non} decreases monotonically as b increases. On the other hand, Fig. 9 shows the shaping delay D_s

[1] Given Pareto(m, α, b) distribution with cut-off at b, with probability density function $f_x(x)$, the probability that a packet with length $L > b$, which means π_{non} in our case, can be calculated by $\pi_{non} = Prob(L > b) = \int_b^\infty f_x(x)dx = (m/b)^\alpha$.

increases monotonically as b becomes larger. As π_{non} and D_s are two important QoS parameters, there is always a tradeoff between them. For example, for delay-sensitive application like streaming video, we would like to use a smaller b to have shorter delay despite of a possible larger packet loss. For non-real-time applications, we would prefer to use a larger b for a small packet loss, despite of the fact that a longer delay may be introduced. Fig. 9 also shows that the shaping delay for rt-VBR is more sensitive to token bucket size, i.e., D_s grows faster for rt-VBR as b increases. This means that we have to sacrifice more delay if we use a larger bucket size for rt-VBR traffic.

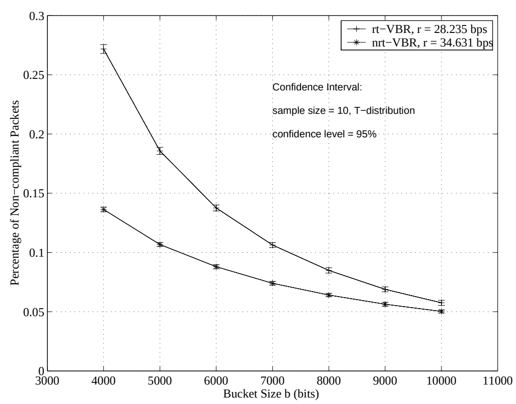

Figure 8: Probability of Non-compliance as b varies

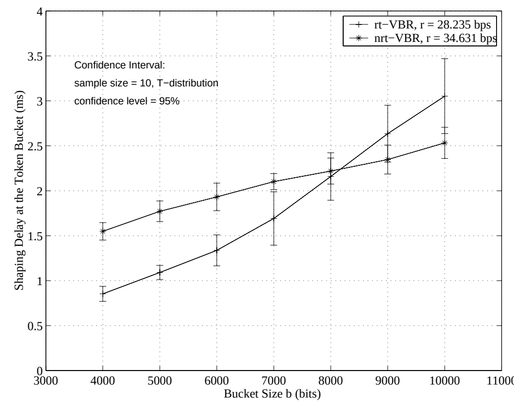

Figure 9: Shaping Delay at the TB as b varies

5.4 Comparison of b and $\hat{b}$

We argued earlier that the bucket size b derived from equation (7) is too conservative and may not be suitable for certain traffic classes. Taking burstiness into consideration we can obtain another bucket size $\hat{b}$. The comparison of b and $\hat{b}$ based on simulation results is tabulated in Table 6.

Table 6. Comparison of b and $\hat{b}$

	$\bar{\eta}$	$r(Kbps)$	$b/\hat{b}(bits)$	$\pi_{non}(\%)$	$D_s(ms)$	$P_d^{conf}(\%)$	$P_d^{non}(\%)$
rt-VBR	0.965	28.235	$b = 5754$	14.61	1.37	0.74	26.68
	0.965	28.235	$\hat{b} = 8221$	8.01	2.20	3.44	29.55
nrt-VBR	0.963	34.631	$b = 5160$	10.18	1.76	1.01	29.07
	0.963	34.631	$\hat{b} = 10865$	4.55	2.43	4.77	30.88

The results generally show that we have achieved a much lower non-compliance probability with $\hat{b}$, at a price of slightly longer delay, larger P_d^{conf} and P_d^{non}. Note there is a difference between the performance of rt-VBR and nrt-VBR. Recall the results in Fig. 9 where D_s is less sensitive to bucket size for nrt-VBR, it means that we have obtained a much lower π_{non} with $\hat{b}$ at a very low cost of D_s. As the non-compliant packets are much more likely to be discarded at any other network nodes, using $\hat{b}$ enables us, especially for nrt-VBR traffic, to have a much lower packet loss ratio. The advantage of using $\hat{b}$ is not so obvious for rt-VBR as for nrt-VBR.

6 Conclusions

We have presented a paradigm of applying traffic conditioning for heterogeneous services in WCDMA radio access network. With the idea of shaping the traffic at the UE and policing at the gateway node, we conclude that traffic conditioning is generally necessary when system load is not very light. The heavier the load is, the more important it is to have traffic control. Using token bucket algorithm as the traffic shaper, there is always a tradeoff between shaping delay and noncompliance probability by adjusting token rate and bucket size. The tradeoff must be made by considering the traffic characteristics of different traffic classes. A smaller bucket size may be suitable for real-time traffic, while for non-real-time traffic, it is advantageous to have larger bucket size which considers also the burstiness of the flow.

Acknowledgment

The work presented here is part of the TURBAN co-operation between NTNU and Telenor AS.

References

[1] 3GPP. QoS Concept and Architecture. http://www.3gpp.org, 3G TS23.107v3.3.0 June 2000.

[2] 3GPP. End-to-End QoS Concept and Architecture. http://www.3gpp.org, 3G TS23.207v1.3.0 March 2001.

[3] R. Koodli and M. Puuskari. Supporting Packet-Data QoS in Next-Generation Celluar Networks. *IEEE Communications Magazine*, 39(2), Feb. 2001.

[4] F. Y. Li and N. Stol. A Priority-oriented Call Admission Control Paradigm with QoS Renegotiation for Multimedia Services in UMTS. *Proc. IEEE Vehicular Technology Conference*, May 2001.

[5] D. McDysan. *QoS & Traffic Management in IP & ATM Networks*. McGraw-Hill, N. Y., 2000.

[6] L. A. Rønningen. Analysis of a Traffic Shaping Scheme. In *Proc. 10th International Teletraffic Congress*, Montreal, Canada, 1983.

[7] W. Stallings. *High-speed Networks: TCP/IP and ATM Design Principles*. Prentics-Hall,Inc., New Jersey, 1998.

[8] Y. G. Sudhir Dixit and Z. Antoniou. Resource Management and Quality of Service in Third-Genration Wireless Networks. *IEEE Communications Magazine*, 39(2), Feb. 2001.

[9] R. Yavatkar, D. Pendarakis, and R. Guerin. A Framework for Policy-based Admission Control. *RFC2753, IETF*, Jan. 2000.