The Throughput Utility Function: Assessing Network Impact on Mobile Services

Markus Fiedler[1], Kurt Tutschku[2], Stefan Chevul[1], Lennart Isaksson[1], and Andreas Binzenhöfer[2]

[1] Dept. of Telecommunication Systems
School of Engineering
Blekinge Institute of Technology
371 79 Karlskrona, Sweden
{markus.fiedler, stefan.chevul, lennart.isaksson}@bth.se
[2] Dept. of Distributed Systems
Institute of Computer Science
University of Würzburg, Am Hubland, 97074 Würzburg, Germany
{tutschku, binzenhoefer}@informatik.uni-wuerzburg.de

Abstract. Based on the need for distributed end-to-end quality management for next-generation mobile Internet services, this paper presents a ready-to-deploy quality assessment concept for the impact of the network on the performance of mobile services. We consider the Throughput Utility Function (TUF) as a special case of the Network Utility Function (NUF). These functions combine the observed network utility at the inlet and the outlet of a mobile network. NUF and TUF capture the damping effect of the network onto user-perceived quality from an end-to-end perspective. As opposed to sometimes hard-to-evaluate QoS parameters such as delay and loss, the NUF is highly intuitive due to its mapping to a simple value between 0 and 100 %, which reflects user perception. We demonstrate the capabilities of the proposed TUF by measurements of application-perceived throughput conducted in a mobile, i.e. GPRS and UMTS network.

1 Introduction

The success of Internet results largely from the *end-to-end (E2E) concept* [1]. Among other benefits, the E2E concept empowers the end-hosts to adapt their data flow autonomously to varying load conditions. By the concept, however, the adaptation is decoupled from the network control. Hence, the network and in particular the network operator is not anymore aware about the requirements of the end-hosts, e.g. their desired throughput. As a result, network management can not address the specific application needs and might disappoint the user.

IP-based *mobile services*, such as mobile streaming, mobile gaming, or mobile file sharing, emerge rapidly due to the advent of highly capable user equipment and increased wireless link capacities. As well as in wireline networks, mobile services need sufficient end-to-end network performance in order to met the users'

Quality of Service (QoS) needs. In mobile networks, however, QoS is typically achieved by network-centric mechanisms, such as *Radio Access Bearer Service* or *CN Bearer Service* in UMTS, [2], or highly influenced by the interconnecting networks. In this way, a gap arises for arbitrary applications between the *user-perceived QoS* and the *network-provided QoS*. In particular, applications which are not designed for signaling their QoS needs or applications which are relayed via non-QoS capable interconnection networks are disadvantaged in obtaining QoS. Unfortunately, this is true for the majority of today's IP-based applications.

As a result, mechanisms are needed which assess the user-perceived QoS between two end hosts for arbitrary application and networks. The concept of a *Network Utility Function (NUF)* [3] constitutes such an approach. Originally, *utility functions* [4] relate the state of applications and networks with the user satisfaction. They are used for rate control and resource allocation [5, 6]. The NUF, used in this work, complements the original concept. It characterizes the change of the utility of the network for a single flow caused by the network behavior, e.g. introduction of (varying) delay and loss or reduction of throughput between two end-hosts. Such a degradation, which may for instance be due to volatile radio conditions, limited capacities or congestion, will be denoted as *damping* in this work. The utility is measured between two end-hosts. It characterizes the quality of a network connectivity with a mapping onto a scale from 0 to 100 %. Thus, the network performance is easy to understand and valuate even for an unexperienced user, which is not necessarily the case for traditional QoS parameters such as delay, delay jitter or loss.

A key characteristic for the utility of a connection is the *perceived E2E throughput*. The perceived throughput is a speed-related parameter [7] and is important both for streaming applications as well as for elastic applications [8]. Throughput difference measurements have turned out to be easy to implement and highly robust since they are based on passive measurements and do not need synchronized clocks [9]. They can easily characterize the change of the perceived throughput between end-hosts. The *Throughput Utility Function (TUF)* makes use of the advantages of throughput difference measurements. It is investigated in this work for mobile services in GPRS and UMTS networks. In [10], we describe a *decentralized QoS monitoring* approach in which the TUF will be embedded.

The remainder of this paper is organized as follows. Section 2 presents the utility-function-based NUF/TUF concept. Section 3 discusses the environment in which the TUF will be investigated. Section 4 develops the TUF for user-perceived throughput in mobile networks, and Section 5 presents a measurement-based case study for UMTS and GPRS. Finally, Section 6 provides conclusions and outlook.

2 The Concept of the Network Utility Function (NUF)

Utility functions are a mathematical tool that is typically used to model the relative preferences of players or bidders in games or auctions. Utility functions reflect the ordering of user preferences regarding the various outcomes of the game by assigning a simple numerical value to each outcome [11]. Utility functions are efficiently applied in network optimization, e.g. [4–6], where individual improvements for users on the throughput and the costs are indented. The NUF concept extends the utility concept by characterizing the damping of the quality and the usefulness of a service caused by the behavior of the network, including network stacks. The network utility is considered between two end-hosts, e.g. a server and a service-consuming client.

Let U_{in} denote the value of the utility function at the sender, e.g. at a server located at the inlet of the network, and let U_{out} describe the utility at the receiver, i.e. at a client located at the outlet. The performance damping of the network due to delay, delay jitter, loss or throughput changes is captured by the *network utility function* U_{Netw} in an E2E view. U_{Netw} defines the relationship between the utilities at the inlet and the outlet as:

$$U_{\mathrm{out}} = U_{\mathrm{Netw}} \cdot U_{\mathrm{in}} \,. \tag{1}$$

The value range of U_{in}, U_{out} and U_{Netw} varies between 0 % in the worst case to 100 % in the best case. Compared to technical QoS parameters such as delay, delay variation and loss, the network utility function is rather intuitive for users, providers and operators [11]. In general, non-specialists cannot necessarily state whether a one-way delay of, let's say, 200 ms represents a problem for a certain application. However, using the NUF, they can rate the perceived service quality on an easy-to-understand scale between 0 and 100 and define thresholds for unacceptability [12].

Service providers and operators can use the utility values to take measures against the network quality problems. For example, they can search for network segments reporting bad conditions. In addition, they can reconfigure the service or the network; or they can compensate affected users; or they can shut down the service for maintenance. Percentage values are also highly appreciated as key performance indicators in business processes, e.g. for demonstrating successful of quality assurance in service provisioning [13].

The network utility function U_{Netw} reaches its best value of 100 % if no network is present or if the network behaves perfectly. The later means that the sent data streams are received instantaneously with no loss and unchanged inter-packet times. In this case, the perception of the quality by the user is that of the application alone, that means for Eqn. 1 that $U_{\mathrm{out}} = U_{\mathrm{in}}$. A lower value of U_{Netw} indicates a disadvantageous change of traffic properties between the corresponding endpoints. In the worst case, the perceived utility U_{out} reaches zero, which can be related either to a very badly behaving network, i.e. $U_{\mathrm{Netw}} \to 0$, or a very bad service quality already on the sender side, i.e. $U_{\mathrm{in}} \to 0$, or a combination of both.

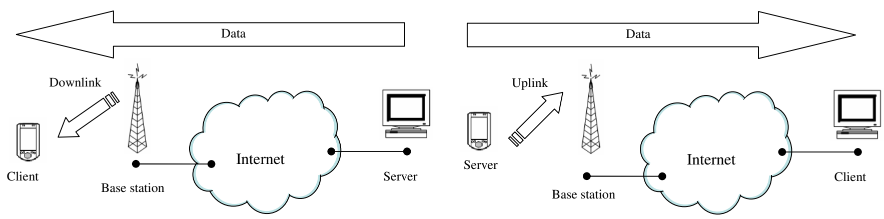
The network utility function U_{Netw} should capture the network impact on a service in such a way that it matches the changes in the user perception and that the same rating applies for the sender and the receiver side. Moreover, the network utility function can capture multiple effects which impact the service quality. In case the influences are rather independent of each other, one can define the network utility function U_{Netw} as a product of specific utility functions $U_{\mathrm{Netw},i} \in [0,1]$:

$$U_{\mathrm{Netw}} = \prod_i U_{\mathrm{Netw},i} \,. \tag{2}$$

For example, if the E2E throughput was considered as the utility of a connection, the U_{Netw} is denoted as *throughput utility function (TUF)*, which is the focus of this paper. In the case of a TUF, the specific utility functions $U_{\mathrm{Netw},i}$ in Eqn. 2 can characterize the change of the throughput, introduced as *m-utility function* in Section 4, or the fluctuation of the throughput, i.e. the coefficient of throughput variation, denoted as *c-utility function*. The m-utility function captures the effect of loss and the c-utility function the effect of delay jitter. A NUF reflecting the impact of one-way delays could be defined as well, but this is a matter of future work.

3 Investigation of the Throughput Utility Function Concept in a Mobile Environment

The applicability of the TUF concept for IP data connections in mobile networks is evaluated by measurements of the perceived end-to-end throughput using a mobile link. Therefore, User Datagram Protocol (UDP) test traffic is sent with constant bit rate (CBR) from a server to a client process, cf. Fig. 1. In the measurements of the *downlink scenario*, cf. part (a) of Figure 1, the client is connected via a mobile link to a base station (BS) and the server is attached to the IP backbone via 100 Mbps Ethernet. In the *uplink scenario*, cf. part (b) of Figure 1, it is vice versa, i.e. the server is connected by a mobile link and the client is attached to the IP backbone with a 100 Mbps Ethernet line.



(a) Downlink scenario. (b) Uplink scenario.

Fig. 1. Mobile scenarios.

During the measurements, the server is trying to send UDP datagrams of constant length L_{A} as regularly as possible. Each datagram contains a sequence number for packet loss detection. In general, the datagrams are sent with a *inter-packet delay* $T_{\mathrm{I,A}} > 0$, i.e. they are not sent back-to-back.[3] Hence, the *offered traffic on application level* is

$$o_{\mathrm{A}} = \frac{L_{\mathrm{A}}}{T_{\mathrm{I,A}}} \,. \tag{3}$$

Considering UDP and IP overhead, the *offered traffic on link level* computes as

$$o_{\mathrm{L}} = \frac{L_{\mathrm{A}} + 28 \text{ Bytes}}{T_{\mathrm{I,A}}} \tag{4}$$

and the *offered link load* to

$$A_{\mathrm{L}} = \frac{o_{\mathrm{L}}}{C_{\mathrm{L}}} \,, \tag{5}$$

where C_{L} denotes the link capacity.

In GPRS networks, the link capacities amount $n \times 9.05$ kbps for Coding Scheme (CS) 1 and $n \times 13.4$ kbps for CS2, where n denotes the number of time slots allocated in up- or downstream direction. The number of allocated slots is typically different for up- and downlink and is upper-bounded by terminal limitations and operator settings. Moreover, the number of available slots can vary significantly in a cell depending on the number and priority of other ongoing telephone calls or data transmissions. Hence, the actual capacity of a GPRS link might deviate largely on short time scales and can not easily be predicted. A similar variability for the actual available throughput can be observed in UMTS networks. Nominally, the UMTS link capacities are determined by the codes used for channel discrimination in CDMA and have typical values of 64, 128 or 384 kbps. The available bandwidth, however, might change due to varying cell load and changing inter- and intra-cell interference.

As a result of the varying link capacity in GPRS and UMTS networks, the sender application does not know about the currently available capacity, i.e. the available bandwidth on a small time scale. Hence, if a sender application generates a packet stream with constant bit rate, the packets might temporally be blocked, the sender might pile up a backlog of datagrams. These are either sent with varying inter-packet time, introducing jitter, or they are dropped if a buffer overflow occurs, introducing packet loss. The "blocked sender" approach helps to discover the network behavior by simply looking at the traffic variations at the sender.

In case of the above mentioned *uplink scenario*, i.e. the server is connected by a mobile link, the UDP traffic generator tries to overcome a possible "blocked sender" state by transmitting packets with shorter inter-packet delay until the cumulative backlog is gone. However, if the offered traffic on link level exceeds the capacity of the uplink, i.e. $A_{\mathrm{L}} > 100$ %, then the backlog becomes permanent. At first, the effective inter-packet delay exceeds the nominal value $T_{\mathrm{I,A}}$, and the average throughput at the server side drops below the (nominally) offered traffic.

[3] A detailed description of the UDP traffic generator is provided in [14]

In the *downlink scenario*, where the server is connected by a 100 Mbps Ethernet link, a bottleneck at the inlet does not exist and the nominal inter-packet time $T_{\mathrm{I,A}}$ can be maintained. A fraction of

$$\ell \simeq A_{\mathrm{L}} - 1 \tag{6}$$

of the traffic will be lost inside the mobile network if the bit rate generated at the sender exceeds the capacity of the mobile link.

4 The Throughput Utility Function for Mobile Services

The adaptation of the NUF concept for the throughput in mobile networks will be described next. First, the parameters for characterizing the throughput utility will be introduced. After that, the components of throughput utility function in mobile environments will be outlined.

4.1 Parameters for Characterizing the Throughput Utility

The characterization of the throughput utility of an E2E connection is based on the concept of the *Throughput histogram Difference Plots (TDP)*, denoted the *bottleneck indicator* introduced in [9]. This concept builds upon the comparison of summary statistics of perceived throughput during an *observation interval* ΔW. Each throughput value denotes the average bit rate perceived during a small *averaging interval* ΔT, typically between 100 ms and 1 s.

In order to apply the concept, the packet streams for an E2E connection are transformed into *throughput time series* $\{R_{\mathrm{A},q}\}_{q=1}^{n}$ at the sender and at the receiver. Each time series contains $n = \Delta W/\Delta T$ throughput values:

$$R_{\mathrm{A},q} = \frac{\sum_{\forall p: T_p \in]T_0+(q-1)\Delta T, T_0+q\Delta T]} L_{\mathrm{A}}}{\Delta T} \quad \text{for} \quad q \in \{1, \ldots, n\} \tag{7}$$

T_p denotes the timestamp of a packet p obtained on application level. For the sender, this timestamp is the instant just before sending the packet. For the receiver, the timestamp is obtained at the instant just after receiving the packet. In this way, the time series capture the whole E2E behavior, in particular they include the behavior of the IP stacks of the sender and the receiver. T_0 is the start time of the time series at the sender respectively at the receiver. The start time is defined by the first packet in a stream both at sender and at receiver, i.e. the first packet triggers the start of the time series. This assumption is motivated by the fact that the receiving application begins to act upon reception of the first packet. The *average throughput* for the whole E2E connection is obtained as

$$m = \frac{1}{n} \sum_{q=1}^{n} R_{\mathrm{A},q}\,, \tag{8}$$

and the *coefficient of throughput variation* for the E2E connection is

$$c = \sqrt{\frac{1}{n-1}\sum_{q=1}^{n}\left(\frac{R_{\mathrm{A},q}}{m} - 1\right)^2} = \frac{s}{m}\,, \tag{9}$$

where s denotes the *standard deviation* of the throughput. Eqn. 9 focuses on the relative variation as compared to the average throughput. The parameters m, s and c are obtained from the observed time series. All three parameters depend on the selection of the observation window ΔW. However, only c and s depend also on the averaging interval ΔT. The values of c and s become smaller as ΔT grows [14].

A condensed form of the *bottleneck indicator* consists of two parameters: the *average throughput* and the *coefficient of throughput variation*. Each of them are observed at the sender and at the receiver:

- sender parameters:
  1. the average throughput at the sender, i.e. the inlet to the network: m^{in};
  2. the coefficient of throughput variation at the sender: c^{in};
- receiver parameters:
  1. the average throughput at the receiver, i.e. the outlet of the network: m^{out};
  2. the coefficient of throughput variation at the receiver: c^{out}.

4.2 Components of the Throughput Utility Function in Mobile Environments

The aim of the throughput utility function is to capture the main influences of mobile networks and to map these influences to a single utility value, cf. Eqn. 1. The mapping is achieved by selecting an appropriate product of specific utility functions, cf. Eqn. 2. The specific utility functions have to be selected such that their parameters describe the change of utility due to problems encountered in mobile networks as accurately as possible. Typically, the following effects can be observed in mobile networks [14–16]:

1. Considerable data loss ($m^{\mathrm{out}} < m^{\mathrm{in}}$) either in the wireless or in the wireline part of the mobile network;
2. Exploding burstiness
   (a) at the receiver ($c^{\mathrm{out}} \gg c^{\mathrm{in}}$) especially when the offered traffic approaches the capacity of the mobile link ($A_{\mathrm{L}} \to 1^{-}$). Such additional burstiness should reflect in reduced TUF values;
   (b) at the sender ($c^{\mathrm{in}} \gg c^{\mathrm{out}}$) in uplink scenarios when the mobile link is overloaded ($A_{\mathrm{Link}} > 1$), which is followed by a strong shaping due to the limited capacity of the channel. As overload implies a mismatch between transportation needs and facilities, the TUF should signal this by displaying very small values.

On this background, we introduce the following utility functions:

1. The *m-utility function*
$$U_m = (1 - \ell)^{k_m} , \tag{10}$$
where
$$\ell = \max\left\{1 - \frac{m^{\text{out}}}{m^{\text{in}}}, 0\right\} \tag{11}$$
denotes *loss* during observation interval ΔW, and k_m is a parameter governing the slope of utility reduction as ℓ increases. The shape of Eqn. 10 has been chosen such as to cushion considerably large loss if $k_m \gg 1$, which was not possible with the initial linear approach discussed in [3].
2. The *c-utility function*
$$U_c = \begin{cases} (\max\{1 - |\gamma|, 0\})^{k_c^-} & \text{for } \gamma < 0 \\ (\max\{1 - \gamma, 0\})^{k_c^+} & \text{for } \gamma \geq 0 \end{cases} , \tag{12}$$
where
$$\gamma = c^{\text{out}} - c^{\text{in}} \tag{13}$$
denotes the absolute change of the coefficient of variation seen from the viewpoint of the receiver. Depending on the sign of that change, we use different parameters to control the slope of U_c, which has the same basic shape as U_m:
   (a) For $\gamma \geq 0$, i.e. $c^{\text{out}} \geq c^{\text{in}}$ equivalent to growing throughput variations, we apply the parameter k_c^+;
   (b) For $\gamma < 0$, i.e. $c^{\text{out}} < c^{\text{in}}$ equivalent to sinking throughput variations, we apply the parameter k_c^-.

The next section will exemplify the impact of these functions on the TUF $U_{\text{Netw}} = U_m \cdot U_c$.

5 Case Study

We present the results of a measurement study carried out in a real UMTS/GPRS network using the same hardware and software in both cases. We apply one observation interval of $\Delta W = 1$ min and an averaging interval of $\Delta T = 1$ s.[4] The packet length in the UMTS case was $L_{\text{A,UMTS}} = 480$ Bytes and in the GPRS case $L_{\text{A,GPRS}} = 128$ Bytes. We chose the following TUF parameters:

[4] Beyond these "snapshots", investigations of the dynamics of the throughput process and thus of the TUF values, by considering different observation intervals are matters of future work. A temporal variation of TUF values might for instance be caused by volatile radio conditions – temporarily bad signal-to-noise ratios are most likely to tear down TUF values due to loss and increased throughput variations – or by resource competition between different traffic streams in a bottleneck [9].

1. For the m-tuility function, $k_m = 10$ in order to make the m-utility function decrease very rapidly as a function of rising loss;
2. For the c-utility function,
   (a) $k_c^+ = 1$ in order to capture additional throughput variations introduced by the mobile network;
   (b) $k_c^- = 2$ in order to capture the overload case implying heavy throughput variations at the sender. As overload reduces the perceived utility dramatically, the decrease of the c-utility function needs to be amplified as compared to case (a), and therefore, $k_c^- > k_c^+$.

We start our investigations by looking at the downlink case.

5.1 UMTS Downlink

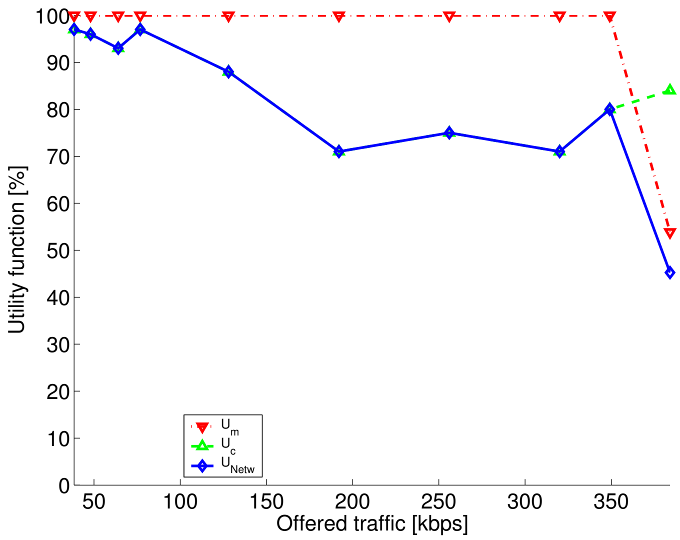


Fig. 2. Utility functions U_{Netw}, U_m and U_c versus offered traffic o_{A} in the UMTS downlink case.

Fig. 2 displays measured values of utility functions in a UMTS downlink scenario. As there is hardly any traffic lost, the m-utility function U_m is close to the optimal value of 100 %. Thus, the TUF U_{Netw} is mainly governed by the c-utility function U_c. We furthermore observe a trend that the traffic variations at the receiver grow as the offered traffic o_{A} increases, which is reflected in decreasing values of U_c.

One particular case deserves some extra attention. Offered traffic on application level $o_{\mathrm{A}} = 384$ kbps implies offered traffic on link level $o_{\mathrm{L}} = 407$ kbps, which exceeds the link capacity C_{L} =384 kbps. The offered link load (5) amounts to 106 %, implying 6 % loss (6). This damps U_m and thus U_{Netw} in a considerable way.

5.2 GPRS Downlink

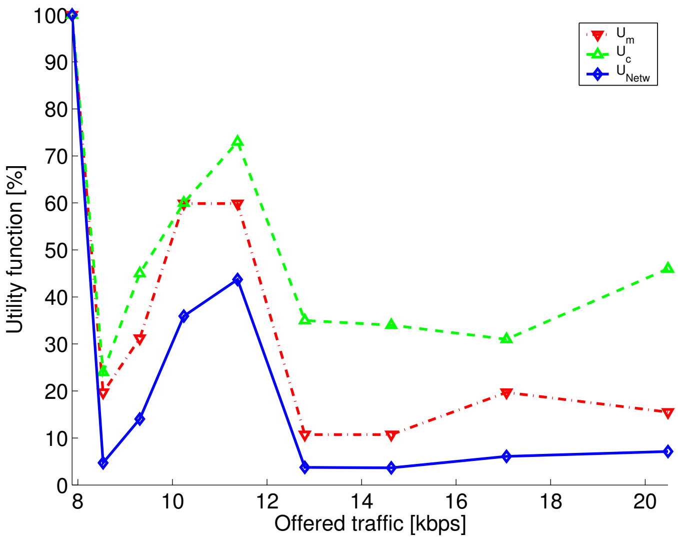


Fig. 3. Utility functions U_{Netw}, U_m and U_c versus offered traffic o_{A} in the GPRS downlink case.

Fig. 3 shows measured values of utility functions in a GPRS downlink scenario. We observe that in come cases, both m- and c-utility functions are low, which indicates considerable loss and growth in variations. As compared to the offered traffic, no real trend can be seen, which is due to extremely volatile conditions in the GPRS network. Detailed studies of the throughput process reveal periods of complete data loss during the observation window ΔW [14]. Consequently, the TUF values are low to very low.

5.3 UMTS Uplink

We now turn our focus to the UMTS uplink case, for which the results are displayed in Fig. 4. As long as the offered traffic is below the critical level of

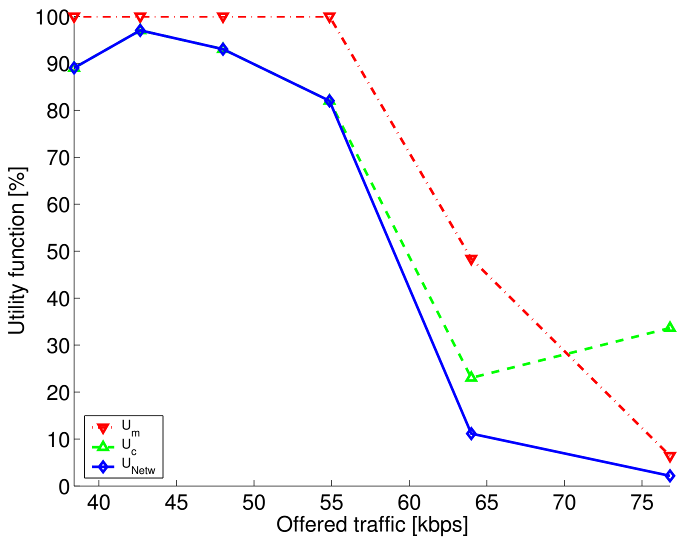


Fig. 4. Utility functions versus U_{Netw}, U_m and U_c versus offered traffic o_{A} in the UMTS uplink case.

59 kbps for which the link load reaches 100 %, the m-utility function signals no loss. Thus, the TUF is governed by the c-utility function. The later signals rather small problems due to variations with a rising tendency as the critial load is approached.

However, for offered traffic $o_{\mathrm{A}} > 59$ kbps, we face overload of the mobile link. We observe heavy throughput variations at the sender and considerable E2E data loss, reflected in low values of the corresponding utility functions.

5.4 GPRS Uplink

Our final investigation deals with the GPRS uplink, cf. Fig. 5. As the m-utility function is close to one despite in some cases with rather little loss, the c-utility function dominates the TUF. However, no real trend can be seen as again, we face quite volatile conditions in the GPRS channel (cf. Sect. 5.2). Interestingly enough, we observe perfect behavior ($U_{\mathrm{Netw}} = U_m = U_c = 1$) in one of the measurements ($o_{\mathrm{A}} \simeq 7.8$ kbps). Here, the GPRS network was able to deliver all packets in a regular fashion. On the other hand, we reach some kind of breakpoint for offered traffic of about 13 kbps. Here, the sender starts to jitter, which indicates overload. For $o_{\mathrm{A}} > 13$ kbps, the corresponding throughput variations are so intense that the c-utility function and thus the TUF are torn down to zero.

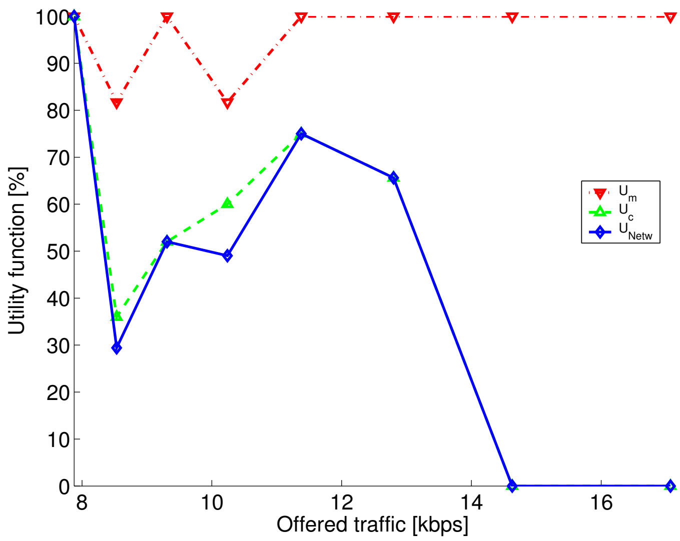


Fig. 5. Utility functions versus U_{Netw}, U_m and U_c versus offered traffic o_{A} in the GPRS uplink case.

6 Conclusions and Outlook

We have described and demonstrated a practicable concept for distributed end-to-end QoS monitoring and assessment on service level, the Network Utility Function (NUF). The NUF relates utility functions at network inlet and outlet and thus captures the damping effect of the network onto user-perceived quality. We investigated a special NUF related to throughput changes imposed by mobile links, the Throughput Utility Function (TUF). The TUF captures changes of throughput averages due to lost traffic, and variations on rather short time scales caused by delay jitter and shaping. Based on measurements of application-perceived throughput via mobile networks (GPRS, UMTS), it was demonstrated that the utility functions proposed in this work are capable of valuating the utility impacts of typical performance problems in mobile networks. In general, this valuation behaves as expected. For instance, UMTS displays a better throughput performance in terms of less loss and variations as as compared to GPRS, which is known from practice. Also, overload is detected correctly in all cases.

Future work will address the adaptation of the TUF, its sub-functions and the corresponding parameters to the needs of specific services and related quantitative ratings by real users. A particularly interesting option is the possibility to determine threshold values regarding user acceptance of particular services. In case the value of the TUF drops below such a threshold, a QoS alarm should be issued. This can happen by sending notifications (e.g. SNMP traps) towards a Service or Network Management System. These notifications may then trigger countermeasures such as adapting the service or the allocation of network resources in order to improve the user perception. Further, the dynamic behaviour of the TUF e.g. due to volatile radio conditions such as hand-overs or traffic conditions such as cross traffic, and the impact of other traffic generation patterns or transport protocols (e.g. TCP) might be studied. And last but not least, the NUF concept might be studied for non-throughput-related parameters as well.

Acknowledgements

The authors would like to thank the Network of Excellence *EuroNGI* for sponsoring this work through the *AutoMon* project and the colleagues from the EuroNGI work package WP.JRA.6.1 *Quality of Service from the user's perspective and feed-back mechanisms for quality control* for valuable input and discussions.

References

1. Jerome H. Saltzer, David P. Reed, and David D. Clark. End-to-end arguments in system design. *ACM Transactions on Computer Systems*, 2(4):277–288, November 1984.
2. 3rd Generation Partnership Project. Technical Specification Group Services and System Aspects; Quality of Service (QoS) concept and architecture - TS 23.107 (Release 5). URL: ftp://ftp.3gpp.org/specs/latest/Rel-5/23_series/23107-5d0.zip.
3. M. Fiedler, S. Chevul, O. Radtke, K. Tutschku, and A. Binzenhöfer. The Network Utility Function: A practicable concept for assessing network impact on distributed systems. In *19th International Teletraffic Congress (ITC19)*, pages 1465–1474, Beijing, China, September 2005.
4. F. P . Kelly. Charging and rate control for elastic traffic (corrected version). *European Transactions on Telecommunications*, 8(1):33–37, Jan. 1997.
5. R. J. La and V. Anantharam. Utility-based rate control in the Internet for elastic traffic. *IEEE Transactions on Networking*, 10(2):272–285, April 2002.
6. M. Dramitinos, G. D. Stamoulis, and C. Courcoubetis. Auction-based resource allocation in UMTS High Speed Downlink Packet Access (HSDPA). In *Proceedings of the First EuroNGI Conference on Traffic Engineering (NGI 2005)*, Rome, Italy, April 2005.
7. ITU-T. Recommendation I.350: General aspects of quality of service and network performance in digital networks, including ISDNs. URL: http://www.itu.ch/.
8. J. Charzinski. Fun Factor Dimensioning for Elastic Traffic. In *Proceedings of the ITC Specialist Seminar on Internet Traffic Measurement, Modeling and Management, Sept. 18–20, Monterey, USA*, 2000.
9. M. Fiedler, K. Tutschku, P. Carlsson, and A.A. Nilsson. Identification of performance degradation in IP networks using throughput statistics. In J. Charzinski, R. Lehnert, and P. Tran Gia, editors, *Providing Quality of Service in Heterogeneous Environments. Proceedings of the 18th International Teletraffic Congress (ITC-18)*, pages 399–407, Berlin, Germany, September 2003.
10. A. Binzenhöfer, K. Tutschku, B. auf dem Graben, M. Fiedler, and P. Carlsson. A P2P-based framework for distributed network management. In *Proceedings of the Second EuroNGI Workshop on New Trends in Network Architectures and Services (WP IA 8.2)*, Loveno di Menaggio, Como, Italy, July 2005. Springer (LNCS).

11. M. Fiedler (ed.). EuroNGI Deliverable D.WP.JRA.6.1.1 – State-of-the-art with regards to user-perceived Quality of Service and quality feedback, May 2004. URL: http://www.eurongi.org/, http://www.ats.tek.bth.se/eurongi/dwpjra611.pdf.
12. H. Hlavacs, J. Lauterjung, G. Aschenbrenner, E. Hotop, and C. Trauner. QoS measurement. Deliverable D-3400, CODIS, November 2003. URL: http://www.ani.univie.ac.at/∼hlavacs/.
13. T. Ingvaldsen, T. Jensen, H. Kjønsberg, H. Abrahamsen, and E. Smørdal. On key performance indicators for end-to-end performance in cellular networks. In *Proceedings of the Seventeenth Nordic Teletraffic Seminar (NTS 17)*, pages 37–48, Fornebu, Norway, August 2004.
14. M. Fiedler, L. Isaksson, S. Chevul, J. Karlsson, and P. Lindberg. Measurement and analysis of application-perceived throughput via mobile links. The Olga Casals Lecture at HET-NETs 05. In *Proceedings of HET-NETs'05*, Ilkley, UK, July 2005.
15. S. Chevul, J. Karlsson, L. Isaksson, M. Fiedler, P. Lindberg, and L. Strandén. Measurements of application-perceived throughput in DAB, GPRS, UMTS and WLAN environments. In *Proceedings of RVK'05*, Linköping, Sweden, June 2005.
16. L. Isaksson, S. Chevul, M. Fiedler, J. Karlsson, and P. Lindberg. Application-perceived throughput process in wireless systems. In *Proceedings of ICMCS'05*, Montreal, Canada, August 2005.