University of Würzburg
Institute of Computer Science
Research Report Series

A Semi-Analytic Model of the UMTS Downlink Capacity with WWW Traffic on Dedicated Channels

Tobias Hoßfeld and Dirk Staehle

Report No. 340 September 2004

University of Würzburg
Department of Computer Science
Am Hubland, D-97074 Würzburg, Germany
[hossfeld,staehle]@informatik.uni-wuerzburg.de

A Semi-Analytic Model of the UMTS Downlink Capacity with WWW Traffic on Dedicated Channels

Tobias Hoßfeld and Dirk Staehle
University of Würzburg
Department of Computer Science
Am Hubland, D-97074 Würzburg, Germany
[hossfeld,staehle]@informatik.uni-wuerzburg.de

Abstract

One of the main advances of 3G networks like UMTS is the ability to support a large variety of different services. These services are subdivided in two domains, circuit-switched services and packet-switched services. The main application expected for packet-switched services is the browsing of the World Wide Web. The web traffic is usually described by quite sophisticated source traffic models and the packet arrivals on IP layer are determined by TCP. On the other hand, the planning process for UMTS networks relies on analytic methods or Monte Carlo simulations that assume the number of users to be Poisson distributed. The intention of this work is to examine if it is possible to apply the existing planning methods to web traffic. We are able to show that the Poisson assumption holds for the number of web users that simultaneously transmit over the air interface and that the resulting NodeB transmit power distribution is valid. We use the Monte Carlo simulation technique to evaluate the web capacity of an example UMTS network.

1 Introduction

One of the main enhancements of third generation (3G) networks is that they allow for a service differentiation and offer a large variety of different services. In the *Universal Mobile Telecommunication System* (UMTS) [1] the services are subsumed in the categories conversational, streaming, interactive, and background that are distinguished by their QoS profiles. The conversational class guarantees a low delay and a low jitter. Typical applications of the conversational class are voice or video telephony that produce a symmetric data volume on the uplink and the downlink. The QoS requirements of the streaming class are less stringent. They sustain a larger delay and tolerate more jitter. The typical applications are audio or video streaming that produce asymmetric traffic. The QoS requirements of the conversational and streaming classes are expressed by the bit rate and the radio link quality, i.e. bit-energy-per-noise-ratio E_b/N_0, of the service. The characteristic of the interactive class is a request response pattern. The most prominent candidate application in the existing and future UMTS networks is the browsing of the world wide web (WWW) which was the dominant kind of Internet traffic for the last decades and only file sharing causes similar traffic volumes in the recent years. Web browsing is also the typical representative of the interactive class while the file sharing falls in the background class. The background class has practically no QoS requirements and uses the bandwidth only when it is available. The web traffic, as a representative of the interactive class, is somehow located between the streaming class and the background class. We distinguish the best-effort web traffic and the QoS web traffic. The best-effort web traffic is close to the background class and utilizes the capacity that remains from the higher classes, i.e. conversational, streaming, and QoS web traffic. On the downlink - which is the relevant link due to the asymmetry typical for web traffic - the best-effort web traffic is transmitted over the downlink shared channel (DSCH), the high-speed downlink shared channel (HS-DSCH), or in the initial phase over a rate-controlled dedicated channel (DCH). Like for the background traffic there is no QoS guaranty for the best-effort web traffic. In contrast, the QoS web traffic is transmitted over a non-rate controlled dedicated channel that guarantees a certain QoS. Like for the conversational and streaming class the QoS requirements of the web service are expressed by the bit rate and the target E_b/N_0.

In this paper we focus on the QoS web service. In particular, we are interested in the number of QoS web users that a NodeB is able to handle in parallel. The capacity analysis of the uplink [2, 3, 4] and the downlink [5, 6, 7] of CDMA and WCMDA systems mainly assume a Poisson distributed number of users per cell and service and a service is defined by its bit rate, target E_b/N_0 and activity factor or mean activity during their sojourn time in the system.

The first objective of this paper is to show by means of a detailed simulations that it is possible to describe a QoS web service by these parameters. This detailed simulation includes a sophisticated web traffic model [8], the implementation of TCP according to 4.4BSD-Lite, and the power control according to the 3GPP standard [9]. As a result of this simulation we obtain the distribution of the number of active web users, where “active” means actively transmitting on the UMTS downlink, and the distribution of the NodeB transmit power. We use a second simulation that does not consider power control to obtain the probability distribution of web session durations and web session activities. In the first instance, we use this distribution in an “activity” simulation to describe the web sessions and show that the results match with the detailed simulation. In a further step, we calculate an offered load from the web session duration and web activity distribution and use it in the Monte Carlo simulation from [7] to determine the NodeB transmit power distribution according to a certain offer web traffic load.

The rest of this paper is organized as follows: In 2 we formulate the objective of this paper more clearly and show how we model the web traffic. In 3 we describe the different simulation models that we use for computing the NodeB transmit power, namely the detailed ON/OFF simulation, the activity simulation, and the Monte Carlo simulation. The results from the three simulation types are compared in 4 where we also evaluate the web traffic capacity of an example scenario. In 5 we summarize our main results.

2 Problem Formulation

We consider a UMTS network that consists of L NodeBs which provide a set of services to the mobile users. These services are either circuit-switched (CS) or packet-switched (PS). The CS service s is defined by a bit rate R_s and an E_bN_0-target ε_s. The users of service s arrive according to a Poisson process with rate λ_s and have an exponential service time with mean $1/\mu_s$. The activity of the CS users is modeled by a Bernoulli random variable with mean ν_s which is referred to as the activity factor.

In the PS domain, a large variety of different protocols and applications is expected. The browsing in the World Wide Web is assumed to be the most important one. WWW traffic is complex to model, since the size and structure of web pages is very heterogeneous. Furthermore, the pages are transported over HTTP and TCP, that determine the packet arrival process on the

UMTS air interface.

In the recent years, a lot of work considered the characterization of web traffic in wireless and wire-line environments. The authors of [8] give an overview of source traffic models that describe web traffic and other applications with some relevance to wireless networks. The web traffic model is mainly based on the measurements of [10].

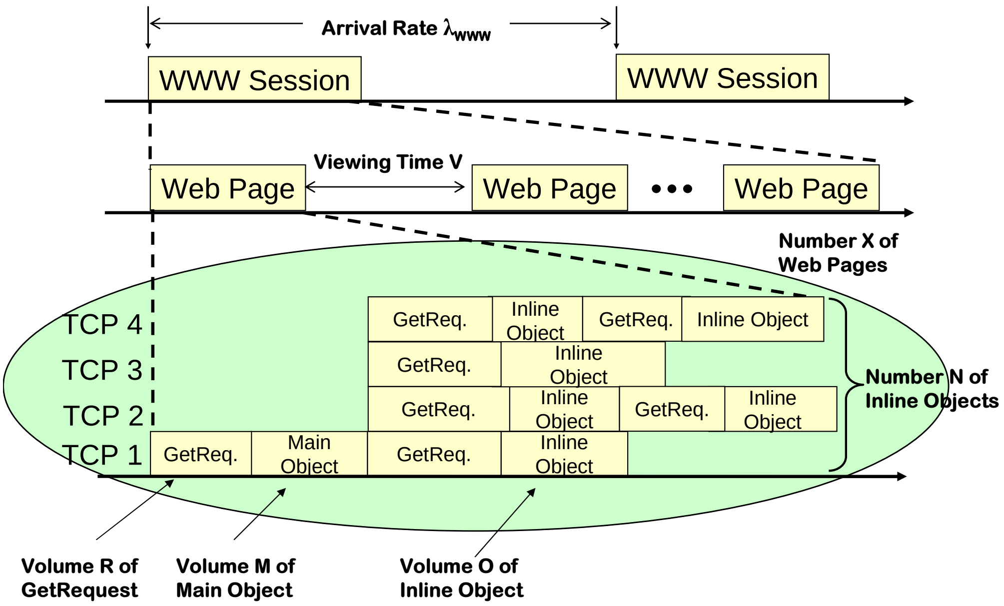


Figure 1: Web traffic model is structured into session, page, and TCP layer

Figure 1 illustrates the web traffic model. We can recognize three layers: session layer, page layer, and TCP layer. On session layer, the users become active according to a Poisson process. In a web session, they download and view a number of web pages.

Table 1: Description of the parameter set of the WWW source traffic model

RV	description	distribution	E[RV]	STD[RV]
I	time between two WWW sessions	exponential	${\lambda_{WWW}}^{-1}$	${\lambda_{WWW}}^{-1}$
X	number of web pages per session	lognormal	25	100
V	viewing time	geometric	5.0 s	30.0 s
M	volume of main object	lognormal	10 kB	25 kB
N	number of inline objects per page	gamma	5.55	11.4
O	volume of inline object	lognormal	7.7 kB	126 kB
R	volume of GetRequest	lognormal	360 B	106 B

A web page consists of the main object and optional inline objects. The main object is the HTML code and an inline object is a file or script and referenced in the main object. HTTP opens a TCP connection and sends a getRequest for the main object. After receiving the main object, the inline objects are requested from the server and sent back in up to four parallel TCP connections. The actual number of TCP connections depends on the HTTP version and the web browsing client.

In our model, we describe the WWW traffic by the session arrival rate λ_{WWW}, the number X of pages per session, and the viewing time V. A page consists of the getRequest size R, the volume M of the main object, the number N of inline objects, and the size O of an inline object. The size of a getRequest for an inline object is also R. All these values are iid random variables which follow distributions obtained through measurements, e.g. [10], and are listed in Table 1.

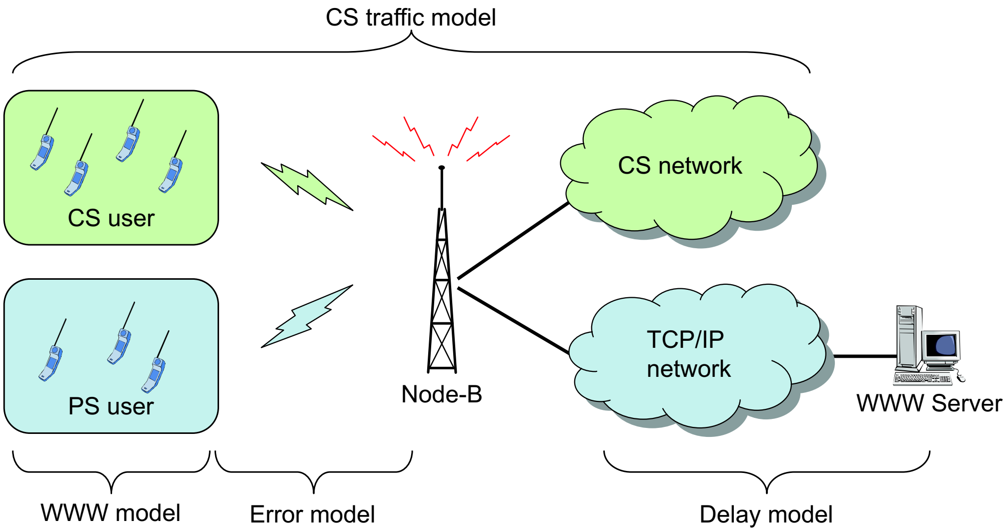


Figure 2: UMTS scenario is described by a given CS traffic and a composed WWW traffic model.

Figure 2 shows the CS and PS domain in a UMTS network. The CS traffic model considers only the air interface, as the behavior on the air interface is independent of the core network and frame errors on the air interface. On the other hand, the WWW traffic is transported over the Internet by TCP/IP. The flow control mechanism of TCP tries to adapt the bandwidth to the congestion in the network. That makes the effective bandwidth depend on the delay in the Internet and lost packets. We model the delay in the Internet by a random variable D for the transmit time between the NodeB and a WWW server in the Internet. The delays for all packets belonging to the same web page are equal. TCP packets get lost with a probability P_{loss} which is assumed to be independent of the system load.

We measured the delay within the GPRS network of a German operator. Therefore, we transmitted periodically ICMP packets from our mobile station to different WWW servers for 12 hours. Figure 3 shows the single delay from the mobile station to the servers. We obtained a mean delay about 400 ms. We simulate the delay time D uniformly distributed between 350 ms and 450 ms.

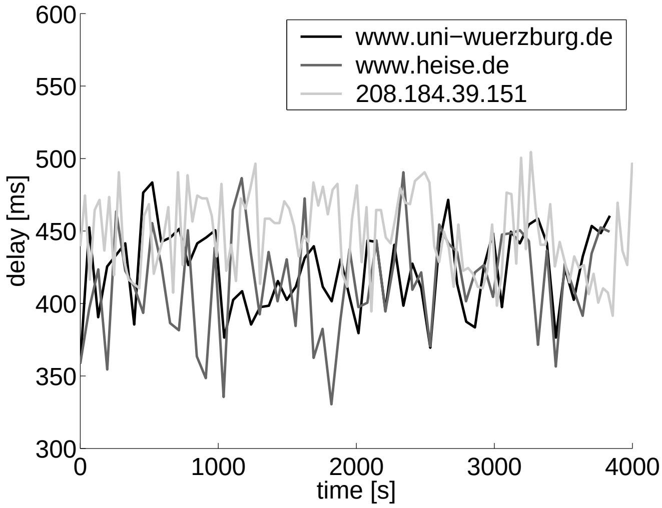


Figure 3: Delay for the transmit time between the NodeB and a WWW server in the Internet.

On the air interface, the CS and PS users compete for the existing resources. The air interface capacity of a UMTS network is on the uplink limited by the multiple access interference and on the downlink by the NodeB transmit power. We focus on the downlink as the traffic a web user produces on the downlink is a multiple of the traffic on the uplink. This shifts the bottleneck to the downlink as long as the users in the CS domain produce symmetric traffic. The 3GPP standard proposes three ways to transport web traffic: on dedicated channels, on shared channels, and in the future on high speed shared channels. We focus on WWW users with a guaranteed quality of service (QoS) that transmit over a dedicated channel with guaranteed bit-rate and E_b/N_0-target.

As the air interface is a scarce resource, power control is employed in order to maximize its capacity. In UMTS, the fast power control is applied on both, the uplink and the downlink. On the downlink, it minimizes the NodeB transmit power while ensuring acceptable service quality. We investigate stationary mobile stations. The power control consists of the inner loop and the outer loop. The inner loop power control adapts the NodeB transmit power in order to meet the E_b/N_0-target by sending one ’‘power-up” or ’‘power-down” command in each time slot. The outer loop power control adapts the E_b/N_0-target in order to maintain a certain frame error rate. We assume a constant E_b/N_0-target throughout the simulation and ignore the outer loop.

The fundamental relation between transmission power $\hat{T}_x$ of NodeB x, connection properties (E_b/N_0-target $\hat{\varepsilon}_k^*$, bit rate R_k, orthogonality factor α), and radio channel conditions (path gain $\hat{d}_{x,k}$ from the mobile station k to the base station x, thermal noise density $N_0 = -174$ dBm/Hz) is summarized in Eq. (1) which is derived by using the bit-energy-to-interference ratio in the downlink direction for an MS k. The transmission power $\hat{S}_{x,k}$ of NodeB x dedicated to MS k is denoted as $\hat{S}_{x,k}$. $W = 3.84$ Mcps is the UMTS chip rate. Assuming perfect power control means that values fulfilling Eq. (1) have to be found.

$$\hat{\varepsilon}_k^* = \frac{E_k^{bit}}{\hat{I}_k} = \frac{\frac{\hat{S}_{x,k}^{rcvd}}{R_k}}{\hat{N}_0 + \hat{I}_{k,other}} = \frac{\frac{W}{R_k}\hat{S}_{x,k}\hat{d}_{x,k}}{W\hat{N}_0 + \alpha(\hat{T}_x - \hat{S}_{x,k})\hat{d}_{x,k} + \sum\limits_{y \neq x} \hat{T}_y \hat{d}_{y,k}} \tag{1}$$

The aim of this paper is to decide whether the UMTS system is stable for a given scenario. A UMTS system is considered as stable, if the transmit power of all NodeBs stays below a given maximum $\hat{T}_{max}$. If the transmit power exceeds this threshold, the NodeB cannot fulfill the E_b/N_0-target requirements and outage occurs. In order to determine the probability that outage occurs, we require the distribution of the NodeB transmit power for a given scenario.

In the following, we present three ways for evaluating this distribution: an ON/OFF simulation, an activity simulation, and an analysis. The ON/OFF simulation reproduces the packet transfer on TCP/IP layer in order to determine when a user is active on the air interface, i.e. a packet is transmitted over the air interface. The TCP layer models the interaction between TCP and HTTP. In our simulation TCP is implemented according to 4.4BSD-Lite. We use the HTTP1.1 version with 4 parallel and persistent TCP connections proposed in [11], too.

The idea of the activity simulation is to approximate the WWW user behavior on web page or session layer by using an activity factor instead of realizing the full TCP stack, as a UMTS user only interferes with other users when being active. For the same reason, an analytical evaluation of a UMTS system requires the approximation of a mobile station by means of activity. Thereby, the activity factor is derived for a given scenario by an ON/OFF simulation.

3 Simulation Description

We study web traffic on page and session layer. For the investigation on page layer, we generate N web users. A web user obtains a location and starts downloading web pages. After one web page is completed, the user immediately requests the next web page such that there is actually no viewing time. We expect the number of active users to be binomial distributed with parameters N and p, whereby p depends on the user activity. We call this type of simulation as *web page simulation*.

In the *session simulation*, we start with no users in the system and then generate web session interarrival times according to the arrival rate λ_{WWW} which we use to scale the traffic intensity. At each arrival event, we determine the location of the new user randomly. We consider a homogeneous spatial traffic distribution, so every position is selected with equal probability. The user keeps this location during the whole session. The web pages in a session are generated according to the model in Figure 2 with the parameters in Table 1.

We have implemented two different simulations how to determine the NodeB transmit power for the web page and the session scenario:

- The ON/OFF simulation has a full implementation of the TCP Reno stack and the WWW source traffic model up to the web page/session layer. We implement the ON/OFF simulation either with or without power control.

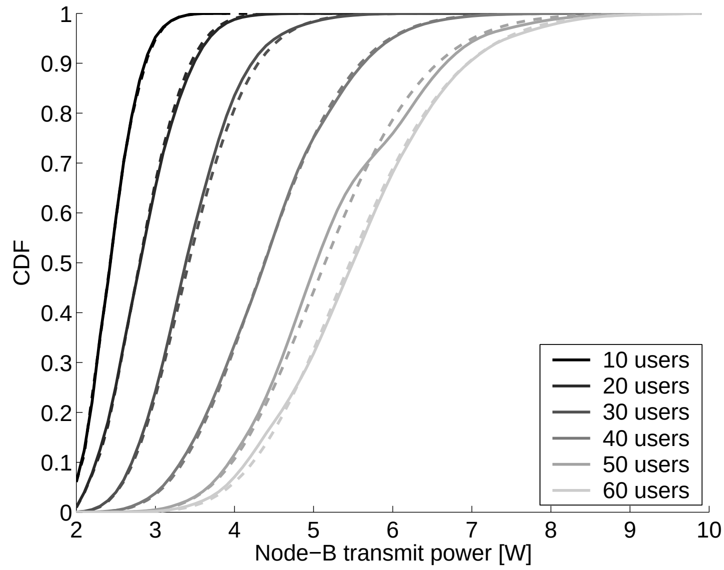


Figure 4: Comparison of the approximative power control algorithm and the exact implementation of the UMTS power control mechanism

- The activity simulation performs snapshots due to the activity factors of the individual users. In order to determine the activity behavior of web users on the air interface for the activity simulation, we use the ON/OFF simulation without power control.

3.1 Implementation of Power Control in the ON/OFF Simulation

The ON/OFF simulation determines the packet arrivals on the air interface exactly according to the TCP/IP protocol stack. This means that for every time instant we know which users are actively transmitting on the downlink. This allows us to determine the NodeB transmit power according to the fast power control algorithm.

However, the fast power control works with a frequency of 1500 Hz, i.e. 15 power updates for each user within a frame of 10 ms. In the simulation, this is very time-consuming, so according to [12] we implement an approximative power control algorithm with only one power update per frame.

In Figure 4, we validate this approach against the exact implementation. We consider 10 to 60 users with fixed propagation gains. The dashed lines show the cumulative distribution function (CDF) of the NodeB transmit power for the approximative power control algorithm and the solid ones for the exact power control. We can see that the curves agree very well.

Approximative Power Control (APC) Algorithm

Input: set $\mathbb{MS}$ of mobile stations transmitting packets on air interface at time t
For each time frame t do

1. Update transmission power and received $E_b N_0$ for all mobile stations

$$\forall k \in \mathbb{MS} \quad \exists x = \text{BS}(k) \in \mathbb{BS} :$$

$$\hat{S}_{x,k}(t) = \omega_k \frac{W\hat{N}_0 + \alpha \hat{T}_x(t)\hat{d}_{x,k} + \sum_{y \neq x} \hat{T}_y(t)\hat{d}_{y,k}}{\hat{d}_{x,k}} \text{ with } \omega_k = \frac{R_k \hat{\varepsilon}_k^*}{W + \alpha R_k \hat{\varepsilon}_k^*}$$

$$\hat{\varepsilon}_k(t) = \frac{\frac{W}{R_k}\hat{S}_{x,k}(t)\hat{d}_{x,k}}{W\hat{N}_0 + \alpha(\hat{T}_x(t) - \hat{S}_{x,k}(t))\hat{d}_{x,k} + \sum_{y \neq x} \hat{T}_y(t)\hat{d}_{y,k}}$$

2. Identify transmission power

$$\forall x \in \mathbb{BS} : \hat{T}_x^*(t) = \hat{T}_{CCH} + \sum_{k \in \mathbb{MS}} \hat{S}_{x,k}(t) \text{ , whereas } x \neq \text{BS}(k) \Rightarrow \hat{S}_{x,k} = 0$$

$$\hat{T}_x(t + \Delta t_{frame}) = \begin{cases} \hat{T}_x^*(t) & \text{if } \hat{T}_x^*(t) \leq \hat{T}_{max} \\ \hat{T}_x(t) & \text{else outage} \end{cases}$$

3.2 Activity Simulation

In contrast to the ON/OFF TCP simulation, the activity simulation does not emulate TCP. It uses an activity factor describing if an user is active or not. The activity factor $\nu = \frac{t_{air}}{t_{total}}$ describes the ratio between the packet transmission time t_{air} over the air interface and the total web page/session download time t_{total}. The web page/session download time is defined as the time between the request of the mobile station to open the first TCP connection and the arrival of the last packet for completing the web page/session.

From the ON/OFF simulation, we obtain a compound distribution for the web page/session download time and the downlink activity of a web page/session. Figure 5 shows this distribution for web pages. The brighter an area is illuminated the higher the probability is for a web page with the given duration and activity. The structure of the compound distribution shows us that web page activity and download time are correlated.

In the activity simulation, we generate the arrival times of web pages/sessions as described above. We determine the download time and the activity factor according to the compound distribution. To determine the distribution of the NodeB transmit power, we generate a system snapshot in regular time intervals. From the compound distribution, each user k obtained an individual activity factor ν_k. This means that at a certain time instant t the user k is active with probability ν_k. The snapshot returns the set of active users. In the next section we describe how to determine the NodeB transmit power for such a snapshot.

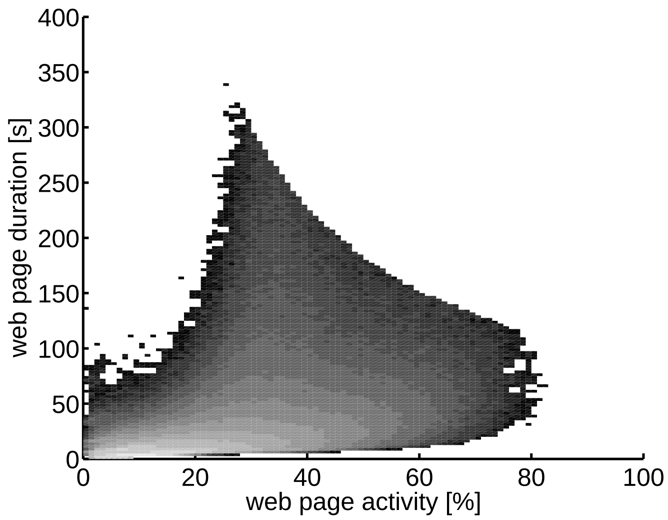


Figure 5: Web page simulation – Compound distribution for download times and activity factors

3.3 Monte Carlo Simulation

With the activity simulation we present a possibility to separate the exact simulation of the web traffic model and the TCP/IP protocol stack and the simulation of the UMTS air interface with the fast power control. However, the simulation is still a time-dynamic simulation and requires the compound distribution of web page/session duration and web page/session activity as input. In contrast to the activity simulation, the Monte Carlo simulation evaluates system snapshots and has no time dynamic. In a snapshot, the number of active web users is described by a binomial distribution in a web page simulation and by a Poisson distribution in the web session simulation. In Section 4 we will validate that the active web users really follow these distribution and also

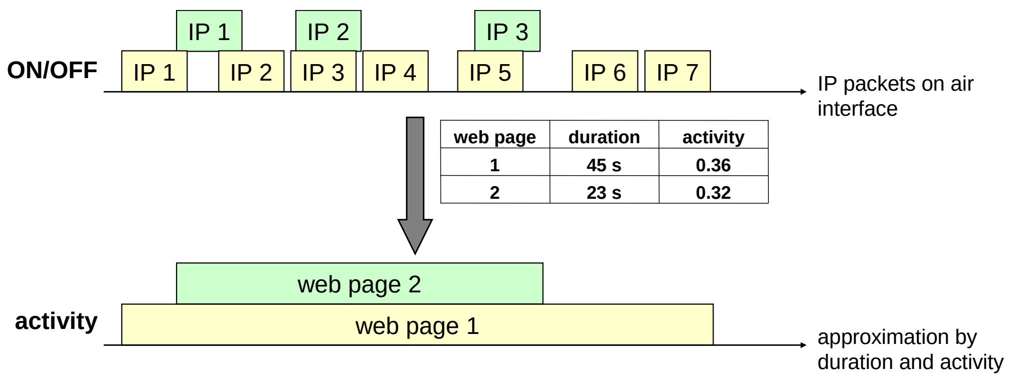


Figure 6: Illustration of the activity simulation

discuss their parameters.

In this section we will describe how to evaluate the NodeB transmit power for a system snapshot. This method and also an entirely analytic method to approximate the NodeB are already published in [7] for general services that are described by their bit rate and target E_b/N_0 value. We formulate the Monte Carlo simulation for a general UMTS network operating with multiple QoS services and one of these is the QoS web service.

A snapshot consists of a set of mobiles with their service and position. The number of users follows from a predefined distribution, in case of the web users either the binomial or the Poisson distribution. The position of each mobile is determined according to the spatial traffic distribution. In our special scenario the traffic distribution is homogenous so every position inside the considered area is chosen with equal probability. The position of the mobile also determines the propagation gain to the different NodeBs.

Consider a mobile k that belongs to NodeB x, or short $k \in x$. Then the transmit power $\hat{S}_{x,k}$ follows from Eq. (1) as:

$$\hat{S}_{k,x} = \left(\omega_{k,0} W \hat{N}_0 + \sum\nolimits_{y \neq x} \omega_{k,y} \hat{S}_y + \omega_{k,x} \hat{S}_x\right) \tag{2}$$

We define the service dependent load of a mobile k as

$$\omega_k = \frac{\hat{\varepsilon}_k^* R_k}{W + \alpha \hat{\varepsilon}_k^* R_k} \text{ and } \omega_{k,y} = \begin{cases} \omega_k \frac{1}{\hat{d}_{x,k}} & \text{if } y = 0 \\ \omega_k \alpha & \text{if } y = x \\ \omega_k \frac{\hat{d}_{y,k}}{\hat{d}_{x,k}} & \text{if } y \neq x \end{cases} . \tag{3}$$

is its service and location dependent load. The correspondent y for the load is either the own NodeB, another NodeB or the thermal noise N_0. The location and service dependent load can be interpreted as the translator between the interference that y causes at the mobile and the transmit power that NodeB x has to spend to overcome this interference. The load of a NodeB similarly comprises the three different kinds of interference. If we sum over all mobiles belonging to NodeB x we speak of the load $\eta_{x,y}$ for NodeB x related to the interference origin y and define this load as

$$\eta_{x,y} = \sum_{k \in x} \omega_{k,y} \text{ and abbreviate } \eta_x = \frac{\eta_{x,x}}{\alpha} = \sum_{k \in x} \omega_k. \tag{4}$$

Again, y is either the NodeB itself, another NodeB, or the thermal noise. The total transmit power $\hat{S}_x$ of NodeB x consists of a constant part $\hat{S}_{x,C}$ required for common channels and the variable part spent for the dedicated channels to the mobiles belonging to x:

$$\hat{S}_x = \hat{S}_{x,C} + \eta_{x,0} W \hat{N}_0 + \sum_{y \in \mathcal{B}} \hat{S}_y \eta_{x,y} \tag{5}$$

These equations for the L NodeBs are written as a matrix equation and solved for the vector $\bar{S} = (\hat{S}_1, ..., \hat{S}_L)^T$:

$$\bar{S} = \bar{S}_C + \bar{\eta}_0 W \hat{N}_0 + \tilde{\eta} \bar{S} \quad \Leftrightarrow \quad \bar{S} = \left(\tilde{\mathcal{I}} - \tilde{\eta}\right)^{-1} \left(\bar{S}_C + \bar{\eta}_0 W \hat{N}_0\right) \tag{6}$$

Note, that a variable $\bar{v}$ stands for a vector and a variable $\tilde{m}$ for a matrix. So $\bar{\eta}_0 = (\eta_{1,0}, ...\eta_{L,0})^T$, $\bar{S}_C = (\hat{S}_{1,C}, ..., \hat{S}_{L,C})^T$, $\tilde{\eta}$ is the $L \times L$-matrix with $\tilde{\eta}(x, y) = \eta_{x,y}$, and $\tilde{\mathcal{I}}$ is the $L \times L$-identity matrix. A reasonable solution exists if the inverse of the matrix $(\tilde{\mathcal{I}} - \tilde{\eta})$ is entirely positive. A sufficient condition for this is that the row sums of $\tilde{\eta}$ are strictly lower than 1.

This condition gives us the means to determine for a snapshot if a power allocation exists such that the E_bN_0-requirements of all mobiles are met. If there is such a solution the NodeBs' total transmit powers follow from Eq. (6) and the power allocated to each mobile from Eq. (2). By generating a series of system snapshots we obtain the moments or the distribution of the transmit powers under the condition that a reasonable power allocation exists. The advantage of the Monte Carlo simulation is that we can easily consider different service combinations, spatial processes, slow fading, and imperfections of power control.

4 Numerical Results

We consider a UMTS network with 19 NodeBs with hexagonal cells that are arranged in two tiers around a central NodeB. The distance between two neighbored NodeBs is 2 km. The mobiles are homogeneously distributed on the considered cell area. For comparison with the ON/OFF simulation the considered area is restricted to the central cell as the simulation of the whole area is too time-consuming. Instead, the NodeBs of the first and second tier transmit with a constant power of 5 W. The parameters of the default scenario are summarized in Table 2.

Table 2: Simulation parameters

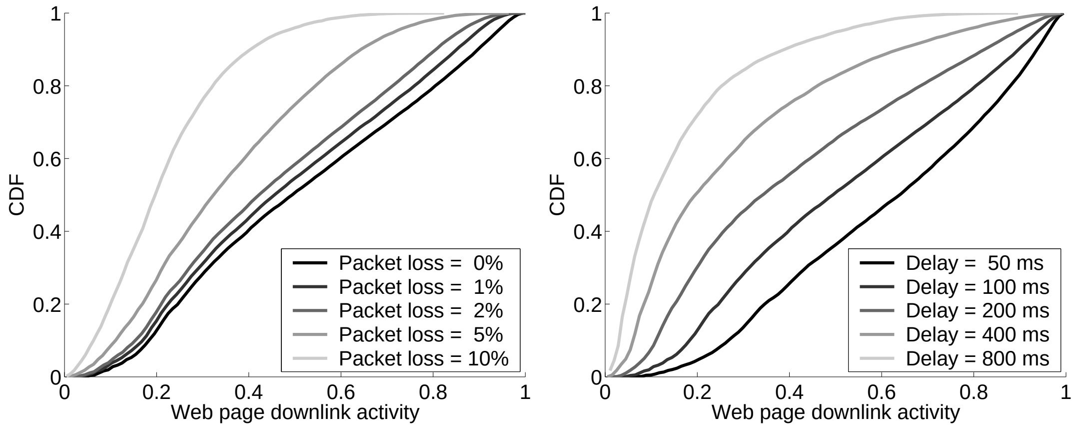
UMTS network parameter	
cell layout	hexagonal
NodeB distance	2 km
number of tiers around the central NodeB	2
propagation gain from NodeB x to MS k according to [13]	$d_{x,k} = -129.4 - 35.2 \cdot \log_{10}(\text{dist}(x, k))$
constant power for common channels	$\hat{T}_{CCH} = 2$ W
power of surrounding NodeBs	$\hat{T}_y = 5$ W
Web traffic parameter	
bit rate of a WWW user	$R = 64$ kbps
target E_bN_0 of a WWW user	$E_bN_0 = 3$ dB
delay between WWW server and NodeB	$D = Uniform(350; 450)$ ms
packet loss probabilty	$P_{loss} = 3$ %

In the following we first investigate the influence of the IP packet loss probability and the fixed network delay between NodeB and web server on the air interface activity of a web user. Afterwards we show that the number of active web users follows a binomial distribution in the web page scenario and a Poisson distribution in the web session scenario. Then, we consider the distribution of the NodeB transmit power and show that the ON/OFF simulation, the activity simulation, and the Monte Carlo simulation yield consistent results. At the end we use the

Monte Carlo simulation to determine the probability that the NodeB transmit power exceeds a maximum of 10 W for different web users and web users arrival rates, respectively.

4.1 Factors Influencing the Air Interface Activity

For investigating the air interface activity while downloading a web page, we use the ON/OFF web page simulation without power control and simulate a single user for 100 hours. Figure 7(a) shows the influence of the packet loss on the web page activity. The larger the packet loss probability is, the smaller is the activity. The same relation holds for the delay time, illustrated in Figure 7(b). A higher delay results in a smaller activity.



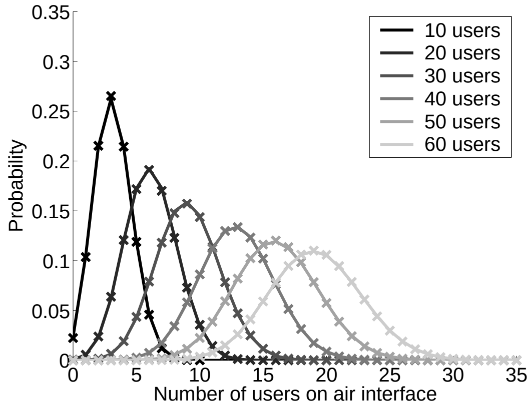
(a) Packet loss (b) Delay time

Figure 7: Web page simulation – Influence factors on the activity factors of web pages

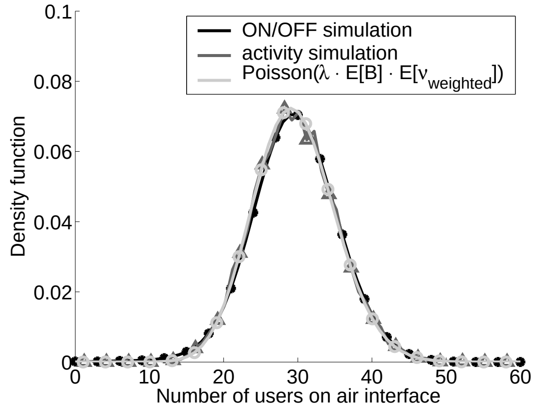
4.2 Distribution of the Number of Active Web Users

We first consider the web page scenario with a constant number N of web users that continuously download web pages. Therefore, we use the ON/OFF simulation without power control and measure the number N_{active} of active users on the air interface. We fit the distribution with a binomial distribution $N_{active} \sim binom(N, \nu)$. Figure 8(a) shows the probability function for $N = 10, 20, ..., 60$ users. The solid line indicates the ON/OFF simulation and the crosses mark the fitted binomial distribution. We can state a good agreement. The parameter of the fitted distribution is $\nu = 0.32$. If we however investigate the mean activity of a random web page, we find that it is only 0.2. The reason for this is that larger web pages with longer download times have larger activities. Therefore, we have to include this correlation in the computation of the parameter and obtain the mean weighted activity as $E[\nu_{weighted}] = \frac{\sum_{web\ page\ i} t_{air\ i}}{\sum_{web\ page\ i} t_{total\ i}}$.

Next, we consider the web session scenario with a session interarrival time of $\lambda_{WWW}^{-1} = 6$ s. The session duration B is a random variable with independent and identical distribution for each session. Thus, we can model the system by an $M/GI/\infty$ queue and obtain that the number of



(a) Web page simulation



(b) Session simulation

Figure 8: Number of active users on the air interface

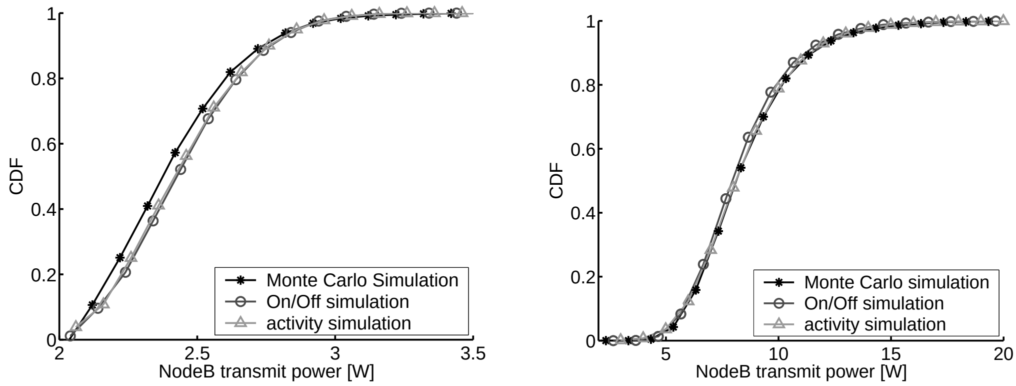
ongoing web sessions N follows the Poisson distribution with parameter $\lambda \cdot E[B]$. If we look at a random time instant on the air interface, we see N sessions and the number of active users on the air interface is $binom(N, \nu)$. We show that despite of the burstiness of the web traffic and the correlation between web page activity and web page download time, the number of user active on the air interface is Poisson distributed:

$$N_{active} \sim binom(N, \nu) = binom(Poiss(\lambda \cdot E[B]), \nu) = Poiss(\lambda \cdot E[B \cdot \nu]). \qquad (7)$$

Figure 8(b) shows that the activity simulation and the ON/OFF simulation match very well. Furthermore, we compare both simulation types with a Poisson distribution. We derive the parameter by computing the weighted mean $E[\nu_{weighted}]$ of the activity factor of the session trace according to the computation of the weighted mean for web page simulations and obtain $E[B \cdot \nu] = E[\nu_{weighted}] \cdot E[B] = 0.267 \cdot 12.08$ min $= 3.22$ min.

4.3 NodeB Transmit Power

We now come to the actual objective of our paper and validate if the Monte Carlo simulation assuming a binomial/Poisson distributed number of users yields the same NodeB transmit power as the ON/OFF simulation. Therefore, we consider both scenarios, the web page scenario and the session scenario. In the web page scenario we consider 10 users located randomly in the central cell. In the ON/OFF simulation and the activity simulation, the users keep their position during the whole simulation period. In the Monte Carlo simulation we generate 10000 snapshots with a binomial distributed number of users that have random positions. In Figure 9(a), we compare the CDFs of the NodeB transmit power obtained from the different simulations. The activity simulation is marked with a triangle, the ON/OFF simulation with a circle, and the Monte Carlo simulation with an asterisk. We see that the activity simulation and the ON/OFF simulation match nearly completely. The Monte Carlo simulation differs as it averages over all possible user locations. In the web session scenario, every new user obtains a random location, so the



(a) Web page simulation (b) Session simulation

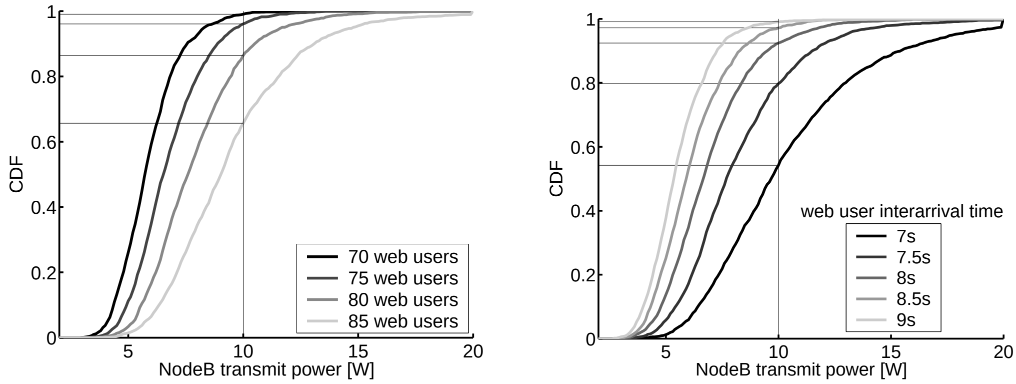
Figure 9: Comparison of ON/OFF, activity, and Monte Carlo simulation

ON/OFF simulation and the activity simulation are not restricted to a single location snapshot. The CDFs of the NodeB transmit power is shown in Figure 9(b) for a session arrival rate of $\frac{1}{6\,\mathrm{s}}$. As expected the three curves show only small differences.

In a UMTS network, the transmit power of a NodeB is technically limited to a maximum $\hat{T}_{max}$ which is typical either 10 W or 20 W. In our scenario we choose $\hat{T}_{max} = 10$ W. If the demand for power exceeds this maximum, the NodeB cannot follow all power-up commands and the mobiles cannot maintain their desired QoS in terms of target $E_b N_0$. This event is called outage and a mobile has to be removed if this situation continues for a longer period of time. One objective of radio network planning is therefore to keep the outage probability below a certain threshold. We obtain the outage probability directly from the CDF of the NodeB transmit power and can thus estimate if the network is capable of carrying the offered traffic or not.

We define the web page/session capacity of a UMTS network as the maximum number of web users/maximum web session arrival rate such that the outage probability stays below a predefined threshold p_{stable}. In the following example, we set $p_{stable} = 5\%$. We use the Monte Carlo simulation to determine the NodeB transmit power. Therefore, we generate system snapshots with mobiles in all 19 cells and evaluate the transmit power of the central NodeB. Figure 10(a) shows the CDF of the transmit power for 70 to 85 web users per cell. We mark the maximum transmit power by a vertical line and indicate the outage probabilities for the different user numbers by the corresponding horizontal lines. The outage probabilities are between 1% for 70 web users and 67% for 85 web users. The maximum tolerable outage probability $p_{stable} = 5\%$ is between the outage probabilities for 75 and 80 web users. So we can conclude that the web page capacity of the example UMTS network is equal to 75 web users.

Figure 10(b) shows the analogous curves for the web session scenario with mean web session interarrival times between 7 and 9 seconds. The web session capacity for $p_{stable} = 5\%$ is reached for an interarrival time in the range of 7.5 and 8 seconds.



(a) Web page simulation (b) Session simulation

Figure 10: Outage occurs if the NodeB transmit power exceeds 10 W

5 Conclusion

The main intention of this paper was to show that we can use well-studied methods for analyzing the UMTS web traffic capacity. These methods are mostly analytic approaches or Monte Carlo simulation that rely on a Poisson distributed number of users per cell and service. By implementing a detailed ON/OFF simulation we could show that the number of web users concurrently transmitting on the UMTS downlink is binomial distributed for the web page scenario with a constant number of web users downloading web pages back-to-back. In the web session scenario where the web users arrive according to a Poisson process and download a random number of web pages the actively transmitting users are Poisson distributed. We also compared the ON/OFF simulation with power control, the simplified time-dynamic activity simulation, and the static Monte Carlo simulation and could state that they all lead to the same NodeB transmit power distribution. Finally, we showed how to use the Monte Carlo simulation to derive the web traffic capacity of a UMTS network with a certain spatial user distribution.

References

[1] 3GPP, "Quality of service (QoS) concept and architecture," Tech. Rep. TS 23.107, 2004.

[2] A. Viterbi and A. Viterbi, "Erlang capacity of a power controlled CDMA system," *IEEE Journal on Selected Areas in Comm.*, vol. 11, pp. 892–900, Aug. 1993.

[3] J. Evans and D. Everitt, "On the teletraffic capacity of CDMA cellular networks," *IEEE Trans. on Veh. Tech.*, vol. 48, pp. 153–165, Jan. 1999.

[4] D. Staehle, K. Leibnitz, and K. Heck, "Fast prediction of the coverage area in UMTS networks," in *Proc. of IEEE Globecom*, (Taipei, Taiwan), Nov. 2002.

[5] W. Choi and J. Kim, “Forward-link capacity of a DS/CDMA system with mixed multirate sources,” *IEEE Trans. on Veh. Tech.*, vol. 50, pp. 737–749, May 2001.

[6] B. Schröder and A. Weller, “Prediction of the connection stability of UMTS-services in the downlink - an analytical approach,” in *Proc. of IEEE VTC Fall*, (Vancouver, CA), Sept. 2002.

[7] D. Staehle and A. Mäder, “An analytic model for deriving the node-b transmit power in heterogeneous UMTS networks,” in *IEEE VTC Spring*, (Milano, Italy), 5 2004.

[8] P. Tran-Gia, D. Staehle, and K. Leibnitz, “Source traffic modeling of wireless applications,” *International Journal of Electronics and Communications (AEÜ)*, vol. 55, 2001.

[9] 3GPP, “Physical layer procedures (FDD),” Tech. Rep. TR 25.214, 2004.

[10] H. Choi and J. Limb, “A behavioral model of web traffic,” in *Protocol 99’ (ICNP 99’)*, International Conference of Networking, September 1999.

[11] D. Staehle, K. Leibnitz, and P. Tran-Gia, “Source traffic modeling of wireless applications,” tech. rep., University of Würzburg, January 2000.

[12] K. Leibnitz, *Analytical Modeling of Power Control and its Impact on WCDMA Capacity and Planning*. PhD thesis, University of Würzburg, 2003.

[13] 3GPP, “Radio frequency (RF) system scenarios,” Tech. Rep. TR 25.942, 2004.