OULUN YLIOPISTO
UNIVERSITY of OULU

KANDIDAATINTYÖ

Joonas Kortesalmi

SÄHKÖ- JA TIETOTEKNIIKAN OSASTO
TIETOTEKNIIKAN KOULUTUSOHJELMA
2009

OULUN YLIOPISTO
UNIVERSITY of OULU

KANDIDAATINTYÖ

Joonas Kortesalmi

Ohjaaja: Juha Röning

SÄHKÖ- JA TIETOTEKNIIKAN OSASTO
TIETOTEKNIIKAN KOULUTUSOHJELMA
2009

Kortesalmi J. (2009) Energy consumption and performance in Finnish UMTS networks. Department of Electrical and Information Engineering, University of Oulu, Oulu, Finland. Bachelor's thesis, 29 p.

ABSTRACT

This bachelor's thesis compares the differences in phone power consumption and performance in Finnish UMTS networks. The latency and the power consumption of packet data connections in UMTS networks is mostly dictated by the RRC state machine. So called inactivity timers control when to transition to a RRC state with lower power consumption. The states with lower power consumption have higher latency so there is a compromise that has to be made by the network operator.

A practical method to control the RRC state and measure inactivity timers with freely available tools was defined using Nokia Energy Profiler on Nokia Series 60 smart phones. The inactivity timers, latency and power consumption of the phone was measured for each RRC state in three different UMTS networks by sending controlled traffic with either ICMP or a custom HTTP server.

In the three networks measured there are two kinds of settings in use. One of the networks uses power saving states aggressively while the others prefer the lower latency states with potentially higher power consumption. The differences in the settings mean that some applications are better suited for the network with aggressive timers to maximize battery life and others may perform better in the other networks, depending on the traffic profile of the application.

Keywords: UMTS, RRC, inactivity timers, power consumption, latency

Kortesalmi J. (2009) Energiankulutus ja suorituskyky kotimaisissa UMTS-verkoissa. Oulun yliopisto, sähkö- ja tietotekniikan osasto. Kandidaatintyö, 29 s.

TIIVISTELMÄ

Tässä kandidaatintyössä vertaillaan puhelimen tehonkulutuksen ja suorituskyvyn eroja kotimaisissa UMTS-verkoissa. UMTS-verkkojen viiveen ja tehonkulutuksen määrää pääasiassa RRC-tilakone. RRC-yhteys siirtyy matalamman tehonkulutuksen tilaan, kun liikenne on tauonnut riittävän pitkäksi ajaksi. Näissä matalamman tehonkulutuksen tiloissa on toisaalta myös suurempi latenssi, minkä vuoksi asetukset valitsemalla operaattorin on mahdollista vaikuttaa tehonkulutuksen ja suorituskyvyn tasapainoon.

Tässä työssä määriteltiin käytännönläheinen tapa mitata RRC:n tila ja mitata RRC:n tilakoneeseen liittyvien ajastinten asetusarvot Nokia Energy Profiler -sovelluksella Nokian Series 60 älypuhelimilla. Ajastinten arvot, latenssi ja puhelimen tehonkulutus mitattiin kolmessa eri UMTS-verkossa lähettämällä määrätynlaista liikennettä joko ICMP-protokollalla tai erityisellä HTTP-palvelinsovelluksella.

Kolmessa mitatun verkon asetukset voidaan jakaa kahteen päätyyppiin. Yhdessä verkoista virtaa säästäviä tiloja käytetään aggressiivisesti, kun taas toisissa verkoissa pyritään pieneen latenssiin vaikka siitä voikin seurata suurempi tehonkulutus. Asetusten erojen vuoksi jotkin sovellukset sopivat paremmin verkkoon, missä puhelimen akku kestää pidempään, ja toiset hyötyvät toisenlaisten asetusten mahdollistamasta pienemmästä viiveestä.

Avainsanat: UMTS, RRC, latenssi, tehonkulutus

TABLE OF CONTENTS

FOREWORD

This bachelor's thesis was written as a part of my work at Oulu University Secure Programming Group (OUSPG).

I would like to thank Christian Wieser for the help with writing my first formal text in English. With his support I was able to quickly build the structure of the thesis and concentrate on the actual content. Big thanks also go to Mika Kuulusa at Nokia Devices R&D for his valuable comments with the initial experiments.

Thomas B. Rücker deserves credit for being able to share his time with me. Throughout this thesis he was able to help me find the right words and gave me new ideas to consider. His interest and support were invaluable for this work.

Last but not least I would like to thank my fiancée Helka Kuorilehto, who was my inspiration during the writing of this thesis.

Writing this thesis was even more challenging than I initially expected. Expressing in written words what was simple in my mind made me realize how complex the reality is. I also discovered that it takes a lot of work to present my ideas and findings in a clear and understandable manner.

ABBREVIATIONS

DCH	Dedicated Channel
DRX	Discontinuous reception
FACH	Forward Access Channel
GSM	Global System for Mobile communications
HTTP	Hypertext Transfer Protocol
ICMP	Internet Control Message Protocol
ISDN	Integrated Services Digital Network
NAT	Network address translation
PCH	Paging Channel
PDP	Packet Data Protocol
QoS	Quality of Service
RRC	Radio Resource Control
RTT	Round-trip time
UE	User equipment
UMTS	Universal Mobile Telecommunications System
URA	UTRAN Registration Area
UTRAN	Universal Terrestial Radio Access Network
W-CDMA	Wideband Code Division Multiple Access

1. INTRODUCTION

Mobile phones have become an integral part of the way we live. Everywhere we go we are expected to be reachable.

The first generation mobile telephone systems were analogue systems. For example the NMT (Nordic Mobile Telephone) and the AMPS (American Mobile Phone System) were notable early systems. The first generation systems did not offer many services beyond phone calls. The second generation (2G) systems like GSM (Global System for Mobile communications) enabled regional or semi-global access to a mobile phone for the first time so that one phone could be used while traveling from one country to another. The second generation systems also provided some data and supplementary services. The third generation (3G) was designed to provide a fully specified standard for a mobile telephone system, backward-compatible with at least GSM and ISDN (Integrated Services Digital Network). Clear added value in all aspects was required from the third generation systems. [1 p.3–8]

The third generation or IMT-2000 (International Mobile Telecommunications-2000) was defined by the ITU (International Telecommunication Union) in a set of ITU Recommendations. There are both backwards compatible standards that are based on the existing second generation systems and new standards that typically require additional spectrum in IMT-2000. The IMT-2000 air interface standards include for example the ”evolutionary” EDGE (Enhanced Data Rates for GSM Evolution) and cmda2000 (Code Division Multiple Access 2000) standards and the ”revolutionary” standards W-CDMA (Wideband Code Division Multiple Access) and DECT (Digital Enhanced Cordless Telecommunications). Many in the industry only consider some of the IMT-2000 standards to constitute as third generation systems, in particular EDGE is usually excluded from 3G statistics. [2]

Due to regional differences, 3G is known with different names in different places. In Europe 3G is considered to be synonymous with UMTS (Universal Mobile Telecommunication System) following the ETSI (European Telecommunications Institute) perspective, while in Japan and US the system is known as IMT-2000. According to 3GPP (3rd Generation Partnership Projectpart), the system is to be referred to as the ”3GPP System” followed by the release number like ”Release 99”. [1 p. 5]

In 2008 the number of mobile broadband subscribers increased by more than 336 000 subscriptions. The yearly amount of data transferred in the mobile networks grew from about 500 terabytes in 2007 to about 4200 terabytes in 2008. There were almost two million UMTS capable devices in Finland by the end of the year 2008. Also in 2008 the Finnish carriers announced the start of HSPA (High Speed Packet Access) upgrades in their 900 MHz networks. [3]

In an usage study about the usage of the cell phones in Finland 18 % of the responders said they used their phone actively for Internet browsing or finding information. Respectively 14 % said they use their phone for e-mail reading or some other e-mail service. Of those who had bought a 3G phone, 44 % reported that getting a 3G phone had increased their Internet and e-mail use. 27 % of cell phone owners had a 3G phone. [3]

The lower energy consumption phones have, the longer the batteries will last. It would be impractical to have our phones constantly connected to a charger. On the other hand the features gaining popularity on mobile devices such as web browsing, always-on email and instant messaging add to the power demands on the battery. Based on the results of the usage study described earlier, Internet access is becoming increasingly important for mobile networks.

1.1. Research question

Different factors such as battery capacity or the services used affect the battery life. For example, the battery life when the phone is used for voice calls is typically significantly lower than the stand-by time.

The aim of this work was to find the differences between networks in terms of energy consumption and battery life. In UMTS, the RRC (Radio Resource Control) inactivity timers can affect the energy consumption significantly. With long inactivity timers it is possible that the phones consume power unnecessarily and get warm or their battery runs out early. On the other extreme, if the inactivity timers are short it is possible that users experience unsatisfactory performance. Poorly chosen values may also lead to reduced network capacity.

Earlier public studies of the inactivity timer values used in Finnish networks were not found with a quick search. It would be interesting to have a method for discovering the settings without operator involvement. Advanced users might be interested in choosing a network which has the settings best suited for their usage.

This thesis answers the following questions:

- What kind of differences are there in the configuration of the RRC inactivity timers in the UMTS W-CDMA networks in Finland?
- How can one measure the relevant parameters without proprietary tools?
- How do the differences affect energy consumption and performance?

2. UMTS

This chapter provides an introduction to the UMTS with the W-CDMA air interface. Special attention is paid to the Radio Resource Control.

2.1. Architechture

UMTS is designed to reuse as much as possible from existing GSM infrastructure. A GSM network can be upgraded so that one core network is used for both UMTS and GSM/EDGE radio access networks. [1 p. 9–11]

A simplified overview of the UMTS overall architecture is shown in the Figure 1 [4 p. 13]. UMTS can be divided into three elements: the UE (User Equipment), the UTRAN (Universal Terrestial Radio Access Network) and the CN (Core Network).

Figure 1. Simplified UMTS architechture

The UMTS terminal device is called UE in literature. It can be divided into two further components, the ME (Mobile Equipment) and the USIM (UMTS Service Identity Module). In the GSM terminology the term MS (Mobile Station) was used to refer to a GSM terminal, like UE in UMTS. In this thesis the word phone refers to an UMTS phone.

The UMTS Core Network provides connectivity to external networks such as other cellular networks, the public telephone network and Internet. It contains vital functions such as access authorization and mobility management in the HSS (Home Subscriber Server). [1 p. 143–148]

The Uu interface connecting the UE and UTRAN is also known as the air interface. The main function of a base station (Node B) is to provide a physical implementation of the Uu interface, W-CDMA, further explained in 2.2. The Iu interface links the RNCs (Radio Network Controller) in UTRAN and the Core Network [1 p. 99–102,146].

The UTRAN is the access network in UMTS (Figure 2 [4 p. 16]). It contains the base stations (Node B) and RNCs. Each RNC has one or more Node Bs under its control. Each Node B has one or more cells where UEs can connect. Each RNC and its base stations form a RNS (Radio Network Subsystem). [4 p. 9,15–16]

2.2. W-CDMA air interface

In cellular networks, the capacity of base stations is limited and it has to be shared by the users. The phones communicate with a base station over an air interface. Each air interface standard defines how the users share the radio channel. There are different methods to make possible this access of multiple users to the same resource.

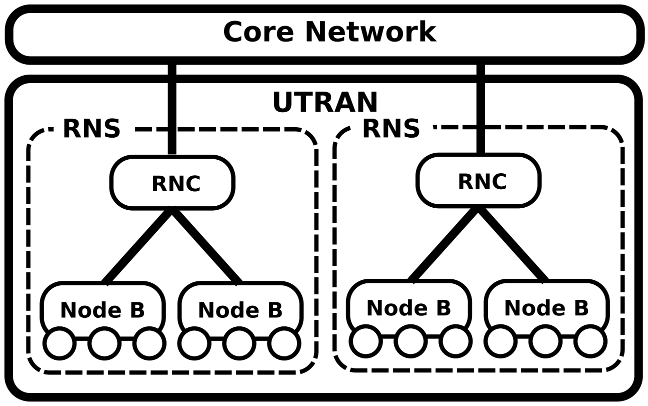


Figure 2. Simplified UTRAN architechture

In FDMA (Frequency Division Multiple Access) each user transmits on a different frequency and in TDMA (Time Division Multiple Access) users transmit in turns on the same channel. CDMA (Code Division Multiple Access) is different and all users transmit on the same frequency simultaneously. Users are differentiated from each other not by frequency or time slot, but by their unique scrambling codes. [5 p. 19–21]

CDMA has some interesting properties for a cellular network. It is efficient in multi-path propagation environment, common to urban areas, where the signal arrives to the receiver through different paths simultaneously. Another key feature is that not only all users in a cell, but in fact all the cells in the network can share the same frequency. This in turn makes possible features like soft handover where a phone can simultaneously use the signal from multiple base stations. On the downside CDMA requires that the transmission power has to be strictly controlled so that the base stations receive signals from all users at equal signal strength. [6 p. 51–55]

2.2.1. Power control

Power control is only described briefly here. For a more detailed analysis it would be useful to consider the differences between different 3GPP releases and the impact of HSDPA.

Power control is used for controlling the transmit power on both the downlink (from base station to phone) and the uplink (from phone to base station) in W-CDMA. For uplink, the point of power control is to make the signal strengths of all phones as even as possible. Because all active users use the same frequency at the same time, the interference level is minimized if all phones and base stations only use the minimum power required to get the required connection quality. There are two different power control methods used: the open loop power control and the closed loop power control. [1 p. 124–126]

In open loop power control, the UE starts with a transmission power estimate based on the received base station signal strength. The stronger the base station signal, the smaller amount of power is used initially. [1 p. 124–126]

Different frequencies have slightly different propagation properties. Closed loop power control is used to improve accuracy of the power control, because the uplink and downlink propagation loss is not the same. The base station sends the phone commands to change the transmit power higher or lower depending on the quality of the signal it receives. The closed loop power control is also used to control the downlink power (base station transmission). In this case the roles are reversed and the phone sends the base station commands to increase or decrease its transmit power. [1 p. 124–126]

2.3. Discontinuous Reception

To conserve energy, UMTS supports DRX (Discontinuous Reception) on the phone. The DRX saves energy by only powering on the UMTS radio receiver periodically. After each sleep cycle the receiver is powered on and messages are received from the network. The network knows when the phone will listen to the paging channel so that it can send a paging message at the right time. [7]

2.4. Packet Data Protocol

The GPRS (General Packet Radio Service) provides the packet switched connectivity for the GSM and UMTS networks. IP access is provided in the form of PDP (Packet Data Protocol) contexts. PDP contexts are the logical connections that are required to send and receive IP packets. [8 p. 323]

A phone can have multiple PDP contexts active simultaneously. The network to connect is selected by choosing an access point name (APN). For example, it is possible that the operator has different access points for services like MMS (Multimeda Messaging Service), WAP (Wireless Application Protocol) and direct Internet access. If a second PDP context is created for the same APN, this is called a secondary PDP context. A secondary PDP context shares the IP address of the primary one but can have a different QoS (Quality of Service) profile. [8 p. 323–325]

The entities providing GPRS access are the SGSN (Serving GPRS Support Node) and the GGSN (Gateway GPRS Support Node). The SGSN connects the RAN to the packet core network and handles both control and traffic-handling functions for the PS domain. The SGSN is responsible for the QoS for the connections. The GGSN provides connectivity to external packet data networks and allocates IP address to the UE. [8 p. 28–29]

A PDP context contains the IP address for the phone during a packet data connection, QoS information other relevant parameters for packet connection. [1 p. 169–170]

2.5. Radio Resource Control (RRC)

The Radio Resource Control is defined in the 3GPP technical specification 25.331 [9]. Its functions include among others establishment, reconfiguration and release of radio bearers that carry the actual payload, cell selection, paging and QoS control. The RRC connection is a bi-directional connection between the UTRAN and the phone. A phone can only have up to one RRC connection. [9 p. 47,49]

There are different types of channels (common, shared and dedicated) that can be used for communications. [10 p. 278]

The common channels are suited for small packets like TCP connection establishment. The latency is typically in the order of 300 ms on common channels. The advantage of common channels is their low setup time. Since common channels do not have a feedback channel, they can not use fast power control, which adds to the interference level. [10 p. 279–280]

Dedicated channels require dedicated resources from the network. Use of high bitrate dedicated channels unnecessarily can lead to resource shortage on the network. [10 p. 280–281]

Shared channels are similar to dedicated channels except that a shared channel is used for multiple users in a time division manner. Shared channels are better suited for bursty traffic than dedicated channels since they support high bit rate but do not require the resources to be dedicated. [10 p. 282]

In W-CDMA, the packet scheduling is controlled both per user and per cell. The RRC state and allocated transport channels are controlled in the user-specific part [10 p. 278]. The cell-specific part controls the sharing of the non-realtime capacity between users using all available cell capacity but maintaining interference below planned levels [10 p. 286–287].

2.5.1. RRC states

The RRC has different states to optimize resource and energy usage. These states are: CELL_DCH, CELL_FACH, CELL_PCH and URA_PCH [9].

The **CELL_DCH** (Dedicated Channel) state provides the best throughput and lowest latency [11]. The energy requirements of the state are also the highest because both transmitter and receiver are on all the time [10 p. 286–287]. In CELL_DCH state both dedicated and shared channels can be used [10 p. 282].

The **CELL_FACH** (Forward Access Channel) state is used when the traffic is low bitrate. The traffic is then transferred on shared channels. The power requirement is approximately half of that of the dedicated channel since the transmitter is not constantly on. [10 p. 282,286]

The **CELL_PCH** (Paging Channel) state provides a low power consumption of 1–2 % of the power usage of the CELL_DCH state. The transmitter and receiver can be powered off for almost all the time and no traffic can be sent or received before first transitioning to another state. [10 p. 286]

The **URA_PCH** is similar to CELL_PCH but the location is known only at the URA (UTRAN Registration Area) level, which reduces number of location updates if the user is moving [12 p. 166].

The RRC Idle mode refers to the state where the RRC connection does not exist at all [12 p. 165–166]. The phone can still remain reachable from the Internet if the PDP context is held open [11]. If there is traffic from the network to the phone, the phone is paged by the network and the RRC connection is created.

DRX is not used in the RRC states DCH and FACH [7].

2.5.2. Inactivity timers

Since there are multiple states with different energy consumption levels, a mechanism for controlling the use of these states is required. The basic idea is that when the traffic levels are low, a state with low power consumption can be used.

A simplified RRC state transition diagram is shown in Figure 3 [9 p. 49][11]. The transitions are based on traffic volumes and inactivity timers T1, T2 and T3 and are managed by the RNC (Radio Network Controller). [11]

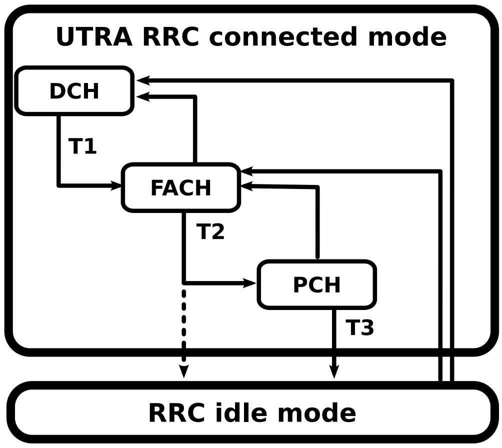


Figure 3. RRC state transition diagram

The **T1** timer controls the transition from DCH to FACH. After there has been no traffic in DCH for T1 seconds, the RRC state transitions to FACH.

The timer **T2** controls the transition from FACH to either PCH or idle mode. If the traffic bitrate in FACH state exceeds a treshold, there is a transition back to the DCH state.

If PCH state is used, the **T3** timer is used to control the transition from PCH to idle mode. Neither PCH or idle mode support traffic sending or receiving so if there is any traffic, the RRC state needs to transition to a higher state.

2.6. Related research

In the study *Energy Consumption of Always-On Applications in WCDMA Networks* [11] the connection between certain RRC (Radio Resource Control) parameters and power consumption in two test networks was analyzed. According to the study, the RRC inactivity timers that control the transitions between various RRC states have a considerable impact on energy consumption. The study found that there are a number of parameters that affect the energy consumption. The main factors for keepalive type of traffic are the inactivity timers, whether the keepalive traffic can be sent in FACH state and the frequency of the traffic.

The energy consumption of web browsing and streaming video was considered in *Impact of Inactivity Timer on Energy Consumption in WCDMA and cdma2000* [13]. In the conclusions an interesting idea of using dynamically adjusted timers, based on user behaviour and battery capacity, was described. However, the work was based on very long timer values ranging from 10 to 160 seconds and the power consumption values in the model were not consistent with those observed in this work.

The whitepaper *Tuning Inactivity Timers in UMTS* [14] describes simulation results based on inactivity timer values of three reference operators. Bursty ON-OFF traffic such as web browsing was used in the simulation model. It was revealed that short inactivity timer values can result in a ”penalty time” caused by delay resulting from RRC state transitions. The state transition delay times used in the simulations ranged from 0.5 seconds (DCH–FACH) to 4 seconds (IDLE–DCH). Significant differences were observed in the power consumption and ”penalty time” with different inactivity timer settings.

None of the works found described actually how to measure the values for the inactivity timers. In the simulations the timers were typically just taken as assumptions. In the *Tuning Inactivity Timers in UMTS* [14] whitepaper it was mentioned that ”radio logging tools” were used. Not enough information was given to reproduce their measurements.

3. EXPERIMENTS

This chapter describes the general test setup and what parameters are being analyzed. First the generic test setup is introduced and then the details for each analyzed variable. The test plan is described for the inactivity timer, latency and power measurements.

3.1. Nokia Energy Profiler

Nokia Energy Profiler is a software available for the Nokia Series 60 phones, 3rd edition Feature Pack 1 and later. It can measure the power consumption of the phone, the signal strength of the network, CPU usage, memory usage and network speed. It is designed to help software developers optimize the resource consumption of their applications on the Series 60 platform [15]. The most interesting feature for this work is the ability to monitor the state of the RRC connection.

The Energy Profiler is a stand-alone application and includes a built-in viewer so the measurements can be analyzed directly on the phone. It is also possible to save the measurement data for later analysis or export it to other formats. The built-in viewer was used extensively to analyze the data in this work. The Energy Profiler was used for RRC state and power measurements.

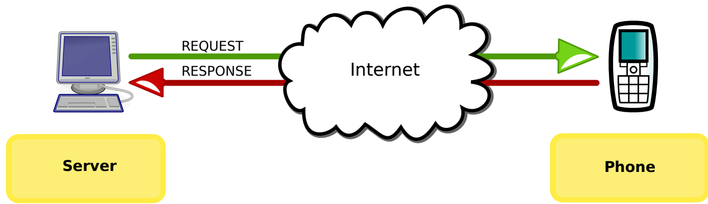
3.2. Test setup

In the first test the inactivity timer values were estimated by generating test traffic and measuring how long the phone remains in each state without traffic flow. In the second test the latency in each of the states was measured to find the impact to performance caused by the states of lower power consumption. Finally in the third test the power consumption in each of the RRC states was measured.

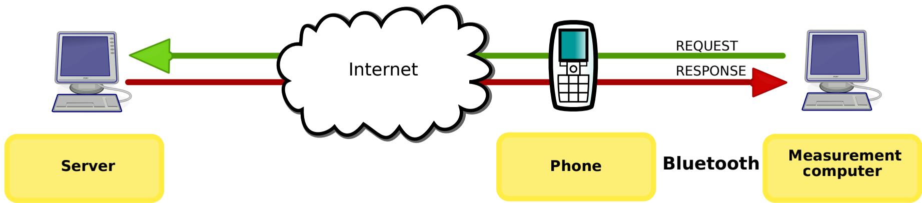
The measurements were performed in three independent commercial UMTS networks. Two different Series 60 phones were used. The second phone was used to check for any obvious differences between the two models. The signal strengths and coverage areas of each network are out of scope of this work. Each of the networks had a strong signal in the location where the tests were done.

Nokia Energy Profiler software was used to read the RRC connection state. Nokia Energy Profiler was set to measure only the network speed, signal levels, 3G timers (UMTS RRC inactivity timers), and energy consumed. Nokia Energy Profiler does not differentiate CELL_PCH and URA_PCH states and these states will be referred to as the PCH state later in this work.

An alternative method for finding the inactivity timer values would be to use the power readings alone and assume that each RRC state results in a different power consumption level. This would be error-prone especially if the times spent in each state were low since the cellular radios are not the only entity consuming power in a phone. This method was not used in this work except to confirm that it supports the RRC state as reported by Energy Profiler.



(a) Internet originated method



(b) Phone originated method

Figure 4. Test setup used in the measurements

The test setup used in this work is shown in Figure 4. The setup consisted of a measurement computer and a mobile phone. There are two methods to control the RRC state.

In the Internet originated method traffic is sent to the phone from the Internet. If the RRC connection is in the idle mode or the PCH state where no traffic can be exchanged, the network will page the phone and its RRC connection will transition to a state that allows data transfer.

In the phone originated method traffic is sent from the phone to the Internet. In that case before the traffic can be sent from the phone it has to request a transition to a suitable RRC state if it is not in such a state already.

To minimize the interference caused by unnecessary traffic, only the necessary software was running on the phone. The HSDPA (High-Speed Downlink Packet Access) support was enabled, because it was enabled by default and is supported by practically all recent devices.

A connection to the public Internet from the cell phone was opened by opening a web page on the web browser software. It was observed that the connection remains open as long as the web browser is running. This fact was used in the Internet originated measurements to keep the connection open during the measurement and find the IP address of the phone.

3.2.1. Discovering the timer values

As described earlier in 2.5.2 on page 13, the inactivity timers control the transition to a state with lower resource consumption when there is no traffic for a certain period. Since inactivity timers are reset on traffic, it is possible to keep the RRC state by sending traffic before the timers expire. It is also possible to elevate the state by sending enough data to exceed the treshold for transition to a higher state.

The inactivity time between last observed traffic and a state transition is measured. If the state transitions are in fact controlled by inactivity timers T1, T2 and T3, the measured times for each type of transition should be constant within a margin of error. If the measured times are not constant, it is likely that some other algorithm controls the RRC state transitions, for example the timers could be dynamically controlled based on user behaviour as suggested in *Impact of Inactivity Timer on Energy Consumption in WCDMA and cdma2000* [13].

3.2.2. Measuring the latency in each state

It is essential for analysis of the inactivity timer configuration to have an understanding of what compromise is made by using the power saving states. It is easy to understand that the states with lower energy consumption also come some drawbacks, otherwise the lower power states would always be used. The latency measurement is most relevant for the states DCH and FACH since those can be used to transfer data. In FACH only limited and shared bandwidth is available. It is expected that the FACH latency is higher than that in DCH state.

No packets can be sent or received directly in IDLE or PCH states so technically it is not possible to measure the latency in these states. The latency is still measured, but the latency then includes an implicit state transition to some other state. The latencies for these states represent how significant a delay is to be expected if these low power states are used aggressively.

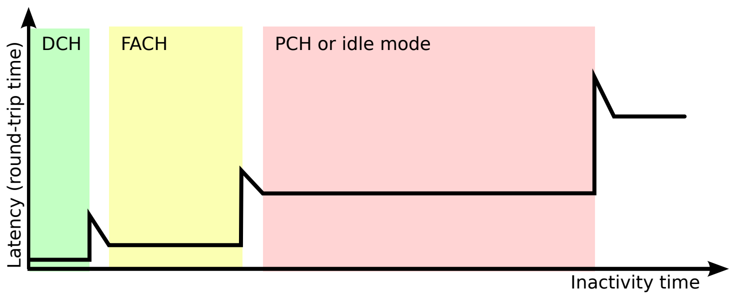


Figure 5. Expected latency as a function of time after inactivity

It is expected that the DCH state has the lowest latency and PCH and idle mode states have the highest latency. The latency after inactivity starting from DCH state is shown in Figure 5. After inactivity in each state there is a transition to a state with lower power consumption. If traffic were to arrive during the transition, the latency would be higher since the transition has to finish before traffic can be transferred.

The latency should only be measured when no state transition is in progress. During the measurement, the RRC state needs to be carefully monitored with the Energy Profiler to make sure the measurements are from the correct state. The phone display backlight should be on during the measurements to prevent the phone from going to sleep [15].

The latency measurement used throughout this work is the RTT (Round-trip time) or in other words the time after sending a request it takes to receive a response from the remote end. It would also be possible to measure one way latency, but the measurements would be considerably more complex.

The host used for measuring latency should have a low latency to each of the networks tested because it adds to the actual latency in the cellular network. Especially important is that the variance of the latency between the measurement host and the cellular operator network is low.

The main limitation of sending traffic from the Internet to the phone is that some operators may have a firewall or a NAT (Network Address Translation) gateway in place blocking inbound connections. The latency may not be identical in Internet originated and phone originated measurements.

3.2.3. Power consumption per state

In this test the screen backlight or background applications could interfere so the backlight should be off and only the absolutely required software should be running while measuring. Additionally, Bluetooth and WLAN should be disabled.

The Nokia Energy Profiler software is used for the power measurements. Despite all the precautions taken, the Energy Profiler software itself consumes some energy. This needs to be taken into account when analyzing the measurement results.

According to the Energy Profiler in-program help file, it can read the built-in current meter once every 0.25 seconds while the voltage meter every 10 seconds. The accuracy listed in the built-in instructions is 0.98 times reality for higher than 400 mW loads. For standby measurements 1 and 5 second intervals are recommended with values roughly 0.7 times reality. The results will be only compared to each other so no corrections are applied to the data.

A reference can be obtained by setting the phone to offline mode and then measuring the energy consumption. This should represent the power consumption of everything other than the UMTS connection and can be compared to the idle mode and PCH measurements.

In this test the 0.25 second measurement interval is used to make sure the RRC state is the one expected even though it was not known for certain if the software also reads the RRC state at the 0.25 second interval or less often. The low measurement interval was selected for all states to get results that can be compared more reliably. The lowest power readings are unlikely to give a reliable absolute value.

3.3. Measurements

The practical measurement details and their results are presented in this chapter.

Initially the Internet originated method was selected and ICMP echo requests were sent from the measurement computer to the phone. This approach worked fine until it was discovered that two out of three operators had a firewall blocking incoming connections from the Internet.

Consequently, the test plan was changed so that the measurements were sent by opening a connection to the Internet via the Bluetooth interface of the phone. This allowed the measuring of the round-trip-time of the connection. However, because Bluetooth also consumes power, this form of RRC state control is not ideal for the power measurements.

Another mechanism had to be devised. HTTP (Hypertext Transfer Protocol) is arguably the most common protocol used on the Internet today. It was realized that a HTTP connection could be used to control the RRC state in place of ICMP packets by writing a custom HTTP server for which it is possible to set the interval and size of sent data. HTTP has the advantage of passing through firewalls easier. None of the networks tested block an outgoing HTTP connection. Another firewall compatible solution would be to use HTTP with TLS, which could possibly be used even with networks with proxies blocking direct communication.

The results for the latency measurements are based on ICMP echo requests. Both the inactivity timers and the power measurements were done with the custom HTTP server based method.

3.3.1. Latency in each state

The host used for these measurements was a server in Espoo. It was confirmed that it had low latency to each of the tested networks. However, all the measurements were performed with the phone situated in Oulu, which means that the traffic had to flow through each of the cellular operator networks between Oulu and Espoo, approximately 540 km [16].

It was verified with traceroute and ICMP echo requests that the latency between the core network of each operator and the measurement host was insignificant, in the order of 1–2 milliseconds or 1–2 % of the lowest latency measured in this test.

Since only the operator A network did not block ICMP echo coming in from the Internet, the phone originated method was used for the measurements below. A laptop was connected to the Internet via Bluetooth on the phone and was used to send the ICMP packets to a server.

The CELL_DCH state latency was measured with a 0.1 second interval. In the CELL_FACH state, an interval of 1 second was used. Only results where the state was constant during the measurement were used.

The measurement for the idle and PCH states differed since it is not possible to send or receive traffic in these states. With these states the echo packet was sent when the RRC state was either idle or PCH. The measured latency then includes the transition to a state where traffic can be transferred.

Table 1. Latency measurement results with phone A showing mean round-trip-time (ms) and its standard deviation (ms)

RRC state	Operator A	Operator B	Operator C
CELL_DCH	**119** / 11 (N=100)	**137** / 19 (N=100)	**132** / 12 (N=12)
CELL_FACH	**334** / 25 (N=25)	**292** / 12 (N=25)	(**308** / 17) (N=5)
PCH	–	–	**875** / 262 (N=9)
Idle mode	**2152** / 102 (N=5)	**2260** / 225 (N=21)	–

Table 2. Latency measurement results with phone B showing mean round-trip-time (ms) and its standard deviation (ms)

RRC state	Operator A	Operator B	Operator C
CELL_DCH	**125** / 13 (N=100)	**139** / 16 (N=99)	**140** / 16 (N=100)
CELL_FACH	**401** / 48 (N=15)	**309** / 24 (N=15)	–
PCH	–	–	**802** / 298 (N=13)
Idle mode	**2128** / 86 (N=7)	**2186** / 181 (N=12)	–

The latency measurement results are presented in the Table 1 for the phone A and the Table 1 for the phone B. The latency measurements and standard deviations are reported in milliseconds. The N represents number of packets that were used in the calculation of these values. Only a limited number of packets were used in the idle mode and PCH measurements because they were considerably slower to measure.

The CELL_FACH latency was very difficult to measure in the operator C network since any traffic in the FACH state triggered a change to DCH and no traffic resulted in a transition to the PCH state. It was measured with the phone A by sending two packets separated by a one second delay. Then this measurement was repeated every 10 seconds. This way the second packet was transferred in the FACH state. It is not expected that it would be fair to compare this result to the other measurements. This measurement was not done on phone B, but like with the other phone, the FACH state was only used briefly.

It is acknowledged that there are differences in the number of packets used in the measurements. With hindsight, it would have been a good idea to have an equal and large number of samples at least for each subcategory, preferrably for all measurements. The results obtained could be improved by repeating the measurements with higher number of samples.

3.3.2. T1, T2, T3 timer measurements

The inactivity timers were first measured by sending ICMP echo packets with different intervals and measuring the time after last traffic for each state. As mentioned, firewalls present in two of the three networks prevented sending ICMP echo requests from the Internet. In this approach only the operating system built-in ping-command was required in addition to the Energy Profiler. The inactivity times between last traffic and a state transition were read from the Energy Profiler.

Since the Energy Profiler did not export the RRC state data into CSV format in the version 1.2, the built-in user interface had to be used to read the RRC state measurements. Further because the Energy Profiler user interface was only able to display data with 1 second resolution, the accuracy of the inactivity times read from the Energy Profiler was ± 0.5 seconds.

The setup was changed to use the custom HTTP server for controlling the RRC state. The initial HTTP request and HTTP headers are big enough to elevate the RRC state to DCH. This made the testing easier and faster. After the headers, the server was set to send 1 byte at a time at different intervals.

If the latency in the connection is not constant, the interval set may not be exactly the same as what the RNC sees, leading to some uncertainty or error in the measurement. The standard deviation of the latency in each state (3.3.1) can be used for estimating the order of the error if necessary.

An increment of 0.25 seconds was used in the measurements for the intervals below 10 seconds and 1 second increment for the longer intervals. The initial hypothesis was that sending data in a HTTP connection like this would be identical to sending traffic with ICMP.

Table 3. Inactivity timer measurement results with the custom HTTP server

	Operator A	Operator B	Operator C
max. interval for a stable DCH	2.75 s	2.75 s	3.00 s
min. interval for a stable FACH	3.00 s	3.00 s	–
max. interval for a stable FACH	40 s	30 s	–

The measurement results in the Table 3 show the maximum interval that resulted in a stable DCH state, the minimum interval that lead to a stable FACH state and the maximum time that could be used to see a stable FACH state for each operator. When the interval was set to a value lower than max interval for a stable DCH, the state remained DCH indefinitely. When the interval was between the two other values, the state FACH remained indefinitely. With a longer interval, there was a transition to idle mode in networks A and B.

In the early experiments it was found that the time after inactivity from FACH to idle mode was 60 seconds in the operator A network. This disproves the earlier hypothesis, because according to the HTTP server based measurements (Table 3) only an interval less than 40 seconds can be used in the network A before the state changes. This means that having a constant interval traffic flow is not equivalent to sending one packet in the desired state and then measuring how long it takes until the transition. It was also verified that the 60 second reading can be obtained in operator A network by stopping the HTTP server so that at first the interval is 40 seconds and then no more packets are sent. This means that it is not specific to the protocol used, HTTP or ICMP. Detailed analysis of this anomaly is left for future work.

The operator C was the only one that transitioned to PCH state instead of idle mode after a long inactivity period. The measurement of the T3 timer controlling how long the PCH state remains was attempted by increasing the interval to infinite and was not found in a reasonable time. If there is an inactivity timer T3, it is quite long. The successful measurement of the T3 timer would provide little added value since there is little difference in the power use and after an inactivity longer than 15 minutes the difference in latency is unlikely to have a significant impact either.

3.3.3. Power consumption per state

As explained in 3.2.3, the phone has to be first configured for minimal energy consumption by disabling everything that is not necessary for the test.

After finding the intervals that lead to stable RRC states (see 3.3.2), it was straightforward to measure the power consumption for each state. The longest interval that can be used for a state was selected and the custom HTTP server was used like before. The Energy Profiler was set to run in the background and record the power use. After the display backlight had turned off, the measurement was left to run for at least 3 minutes to get a reliable average reading. Only the measurements where no anomalies were present in the RRC state during the 3 minute period were used.

Table 4. Power average consumption measurement results (Phone A / Phone B)

RRC state	Operator A	Operator B	Operator C
CELL_DCH	1.18 W / 0.85 W	1.19 W / 0.86 W	1.21 W / 0.89 W
CELL_FACH	0.55 W / 0.41 W	0.56 W / 0.41 W	–
PCH	–	–	48 mW / 29 mW
Idle mode	46 mW / 30 mW	41 mW / 29 mW	–

The power measurement results are presented in the Table 4. The power consumption could be affected by signal strength especially in the DCH state where the transmitter is used. The small differences seen here can easily be due to measurement error.

4. ANALYSIS

The measurement results and the reasons why there are differences are analyzed in this chapter.

4.1. Inactivity timers

Initially it was assumed that only the inactivity time controls the RRC state transition to a lower power state. This seemed to be the case in initial experiments, but with the HTTP based method it was found that at least in operator A network there is a different parameter as well. If there is periodic traffic, the RRC state transitions from FACH to idle after a different time than what it would take if the traffic was simply stopped.

Based on this finding, the logic described for the inactivity timers in 2.5.2 might be an oversimplification and more complex algorithm might be in use instead, at least in operator A network. The measurements still provide a way to analyze and compare different networks. The HTTP based method used still makes sense even though the inactivity timers is not the only variable, because the communication generated there is similar to many protocols with keep-alive. Persistent connections in HTTP, for example, can reduce TCP handshakes, reduce latency and improve performance [17 p. 44].

Based on the measurements, there are two different general kinds of RRC state control in use in the three networks analyzed. The networks A and B use the FACH state whenever the DCH state is not required. The PCH state is used in the network C whenever possible and the FACH state as little as possible. The main difference between the networks A and B is that the network A uses a longer inactivity timer in the FACH state.

4.2. Latency

The lowest latency is provided by the DCH state, followed in performance by FACH, PCH and idle mode. The latency in the PCH state is lower than in the idle mode because the RRC connection does not have to be re-established when traffic arrives.

There are no significant differences in latency for each state. Operator A has a slightly lower latency than the others in the DCH state but because only a limited number of packets were used and the measurement was not repeated at different times of the day at different locations, the difference could be explained by error in the measurement.

What is a significant finding, however, is that only one operator uses the PCH state. Based on the measurements it could be possible to achieve considerably lower latency in the networks A and B by using the PCH state instead of idle mode. On the other hand, in those networks the FACH state was used for at least 30 seconds after inactivity, which reduces the impact of the high latency in the idle mode. A transition from the PCH state to the idle mode was not observed while the connection was open, so the presence of a T3 inactivity timer could not be confirmed.

Having too aggressive inactivity timers can lead to unnecessary transitions, which increase the latency or penalty time in the connection [14]. The transition times were not examined so the detailed analysis of the latency in networks like the operator C network is left for future work.

4.3. Power consumption

The differences in the power consumption between different networks are not significant. The RRC state strongly correlates with the power consumption, the network has little impact.

Two different phones were used in this work. The power consumption between the phones is different, likely due to hardware differences. Both phones are recent smart phones and have similar features. If we calculate the percentages of power consumption in comparison to the DCH state, we see that the FACH and idle mode consume about 46 % and 4 % correspondingly with the phone A and 47 % and 3 % with the phone B. These are a good match to the values presented in 2.5.1.

4.4. Energy consumption

To be able to analyze the energy consumption difference it would be necessary to define a traffic profile that describes how big packets are being transferred and what is the packet interval statistical distribution. It would also be necessary to know what are the triggers to elevate the RRC state, that is how big packets have to be to trigger a transition from FACH to DCH. This was not studied in this work and is left for further work. Then it would be possible to calculate how long each state is used and multiply the power consumption of each state with the time spent in it and get the total energy consumption. The results would be very different depending on the traffic profile, even with the same amount of data transferred.

Power and energy consumption are two related but entirely different concepts. The battery life depends on the energy consumption. The energy consumption is a function of the traffic profile which again depends on how the phone is used.

It is easiest to understand the results of this work through the following examples.

The first user only uses his or her phone to check the bookmarked weather page before walking to work. This only consumes power for a short period of time. If the connection is closed shortly after loading the weather information, it is irrelevant what are the inactivity timers. Short timers do not have a big impact if only one page is loaded.

The second user runs an instant messenger on their phone. The software keeps the connection open all the time. Because many networks utilize sharing of Internet addresses through NAT, it may be necessary to send a short dummy message perhaps even every 30 seconds in order to keep the connection alive. If the inactivity timers are short, high amounts of power is only used for short periods of time.

The third user is using their phone to access the Internet on their computer to browse the Internet and play online games. They have both the computer and the phone plugged in so the battery life is not an issue. If the inactivity timers are short, there might be unwanted state transitions back and forth. This results in higher latency and lower experienced performance.

For the first user the inactivity timers do not make a difference. There is no impact for the battery life or performance. The second user benefits from short inactivity timers since a latency of a few seconds is no problem for an instant messenger. Long inactivity timers may result in unsatisfactory battery life for them. The third user benefits from very long inactivity timers. The more they can stay on the DCH state, the better the performance they experience.

5. CONCLUSIONS

The aim of this work was to compare the RRC inactivity timers in Finnish UMTS networks and analyze the energy consumption of UEs and performance differences. Since no method to analyze the RRC state machine behaviour in the desired way without proprietary tools was found in the earlier works, a method to measure the inactivity timers was defined. The target of the performance analysis was the latency of the connection, not bandwidth.

The firewalls that blocked the measurement packets made it impossible to analyze the latency from the Internet and forced to measure it from the phone side. In retrospect it would have proven more fruitful to directly assume this approach since it is perhaps a more relevant figure.

When analyzing the inactivity timers, it was discovered that the model used could be incomplete. A measurement in early experiments conflicted with later ones and gave reason to believe that there might be more variables than only the inactivity time. Despite this anomaly, the results are useful if interpreted carefully.

Public documentation was readily available from the 3GPP and other sources. Also the literature related to UMTS was extensive and for a short one person project it was impossible to grasp every detail. If the RRC state transitions are described accurately somewhere, it got lost in the very large amount of publications. This formed perhaps the greatest challenge in this work.

As a result of this work it was found that different operators have made different compromises between power consumption and latency. It seems clear based on the measurements that the settings in use in the operator C network aim to provide the maximum battery life, while the other two have strived for low latency. The impact of the differences between networks depends on how phones are being used. The network C had the most aggressive inactivity timer settings. Especially applications such as instant messengers and push e-mail which have bursty traffic are likely to benefit from the configuration in improved battery life.

Dramatic improvements could be gained in each of the networks if the inactivity timers were adjusted depending on the use profile or device. Another possible improvement is to provide two different APNs, one with a strict firewall and one without. This way users could choose to benefit in power savings from the protection of the firewall from unwanted Internet traffic but also gain full connectivity on a device without power limitations. It would be up to the operator to decide which of these would be better as the default for the users. It is the hope of the author that upcoming 3GPP releases provide improvements in the latency and power consumption to make this kind of trick unnecessary.

The state transitions cause additional latency because no traffic can flow during the transition. Future work on this field should study how the transition times could be accurately measured, or depending on which side the reader is, be removed. Another possible approach would be to build the capability to record use profiles from real use and apply the measurements to those scenarios. For the performance analysis it has to be studied what controls the transitions to a higher power state.

Since the inactivity timer model did not match all measurements, a more detailed analysis of the RRC state machine transition triggers also requires future work. With more time it would be possible to use similar methods used here to analyze the triggers of the state transitions. It could be possible to combine measurements with different kinds of test traffic to find ideally all paramters for the state transitions. Of course the most reliable way to get this information would still be from the vendor of the RNC, if the information is provided.

The performance and energy consumption in Finnish networks was analyzed in this work. In conclusion, although it was not possible to tell which network is the best since it depends on the use profile, a practical method to compare different networks was defined and is now available publicly, without the need for proprietary tools or information.

6. REFERENCES

[1] Kaaranen H., Ahtiainen A., Laitinen L., Naghian S. & Niemi V. (2005) UMTS Networks: Architechture, Mobility and Services. John Wiley & Sons, 2nd ed., 406 p.

[2] ITU What really is a third generation (3g) mobile technology. URL: `http://www.itu.int/ITU-D/imt-2000/DocumentsIMT2000/What_really_3G.pdf`. Last accessed: Sep 20 2009.

[3] Markkinakatsaus - Vuosikatsaus 2008. Annual market report, Viestintävirasto, URL: `http://www.ficora.fi/attachments/5fgEgJfk4/mk08_36s_a4_08_090330.pdf`. Last accessed: Sep 20 2009.

[4] 3GPP UTRAN overall description. TS 25.401 V8.2.0, 3rd Generation Partnership Project (3GPP), 53 p.

[5] Novosad T., Soldani D., Sipilä K., Kola T. & Wacker A. (2006) Introduction to wcdma for umts. In: J. Laiho, A. Wacker & T. Novosad (eds.) Radio Network Planning and Optimisation for UMTS, John Wiley & Sons, 2nd ed., pp. 19–92.

[6] Muszynski P. & Holma H. (2005) Introduction to wcdma. In: H. Holma & A. Toskala (eds.) WCDMA for UMTS : Radio Access for Third Generation Mobile Communications, John Wiley & Sons, 3rd ed., pp. 149–184.

[7] Yang S.R. (2007) Dynamic power saving mechanism for 3g umts system. Mobile Networks and Applications 12, pp. 5–14.

[8] Poikselkä M., Mayer G., Khartabil H. & Niemi A. (2004) The IMS: IP Multimedia Concepts and Services in the Mobile Domain. John Wiley & Sons, 419 p.

[9] 3GPP Radio Resource Control (RRC); Protocol specification. TS 25.331 V8.7.0, 3rd Generation Partnership Project (3GPP), 1685 p.

[10] Wigard J., Holma H., Cuny R., Madsen N., Frederiksen F. & Kristensson M. (2005) Packet scheduling. In: H. Holma & A. Toskala (eds.) WCDMA for UMTS : Radio Access for Third Generation Mobile Communications, John Wiley & Sons, 3rd ed., pp. 269–306.

[11] Haverinen H., Siren J. & Eronen P. (2007) Energy consumption of always-on applications in wcdma networks. In: Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th, pp. 964–968.

[12] Vialén J. & Toskala A. (2005) Radio interface protocols. In: H. Holma & A. Toskala (eds.) WCDMA for UMTS : Radio Access for Third Generation Mobile Communications, John Wiley & Sons, 3rd ed., pp. 149–184.

[13] Lee C.C., Yeh J.H. & Chen J.C. (2004) Impact of inactivity timer on energy consumption in wcdma and cdma2000. In: Wireless Telecommunications Symposium, 2004, pp. 15–24.

[14] de Bruynseels L. (2005) Tuning inactivity timer settings in umts. Technical whitepaper, Commsquare, 10 p.

[15] Nokia energy profiler quick start. URL: `http://forum.nokia.com/main/resources/user_experience/power_management/nokia_energy_profiler/`. Last accessed: Sep 20 2009.

[16] Wolfram Alpha Distance between oulu and espoo. URL: `http://www.wolframalpha.com/input/?i=distance+between+oulu+and+espoo`. Last accessed: Sep 20 2009.

[17] Fielding R., Gettys J., Mogul J., Frystyk H., Masinter L., Leach P. & Berners-Lee T. (1999) Hypertext Transfer Protocol – HTTP/1.1. RFC 2616, Internet Engineering Task Force, 176 p.