Improving User Experience of Internet Services in Cellular Networks

Annika Klockar

Faculty of Health, Science and Technology

Computer Science

DISSERTATION | Karlstad University Studies | 2015:18

Improving User Experience of Internet Services in Cellular Networks

Annika Klockar

DISSERTATION | Karlstad University Studies | 2015:18

Improving User Experience of Internet Services in Cellular Networks

Annika Klockar

DISSERTATION

Karlstad University Studies | 2015:18

urn:nbn:se:kau:diva-35334

ISSN 1403-8099

ISBN 978-91-7063-631-8

Distribution:
Karlstad University
Faculty of Health, Science and Technology
Department of Computer Science
SE-651 88 Karlstad, Sweden
+46 54 700 10 00

Print: Universitetstryckeriet, Karlstad 2015

WWW.KAU.SE

Improving User Experience of Internet Services in Cellular Networks

Annika Klockar
Department of Mathematics and Computer Science

Abstract

The Internet has grown enormously since the introduction of the World Wide Web in the early 90's. The evolution and wide spread deployment of cellular networks have contributed to make the Internet accessible to more people in more places. The cellular networks of today offer data rates high enough for most Internet services. Even so, the service quality experienced by the users is often lower than in wired networks.

The performance of TCP has a large impact on user experience. Therefore, we investigate TCP in cellular networks and propose functionality to improve the situation for TCP. We have studied sources of delay and data loss, such as link layer retransmissions, queuing, and handover. Measurements were conducted in a GSM/GPRS testbed. The results indicate that TCP interact efficiently with the GSM link layer protocol in most cases. From experiments of queuing in GPRS, we conclude that with a smaller buffer delay is reduced significantly, but that TCP throughput is about the same as with a larger buffer. Furthermore, we propose an improved buffer management when a connection loses all its resources to traffic with higher priority. We also propose a scheme for data forwarding to avoid negative impact on TCP during handover for WINNER, a research system that was used to test ideas for LTE.

The achievable data rates in cellular networks are limited by inter-cell interference that vary over the cell. Inter-cell interference can be mitigated with Coordinated Multipoint techniques (CoMP), techniques that currently are being standardized for LTE-Advanced. System wide CoMP is, however, not an option, since it would be too resource consuming. In order to limit the required resources for CoMP, we propose an approach to select a subset of users for CoMP that is based on user experience. Simulation results indicate that user experience, represented with application utility, and fairness are improved compared to if only rate is considered in the user selection.

Keywords: Internet, cellular networks, wireless, TCP, ARQ, handover, queuing, CoMP, application utility, user selection, scheduling, fairness

Acknowledgements

First and foremost, I would like to express my gratitude and appreciation to my supervisor Anna Brunstrom for her superior support and encouragement. She has always shown a genuine interest in every detail of my work and given me constructive, well-considered, and apt feedback. I would also like to thank Mikael Sternad who patiently has answered my questions about radio related issues and spent hours discussing research problems with me, especially when I was working on the three last papers. Our discussions have been a source of inspiration for me.

I wish to thank my co-authors: Johan Garica, who, along with Anna Brunstrom, introduced me to research in Computer Science, Jan H. Gustafsson, thanks to whom I got the opportunity to experiment with a real cellular network, Juan Rendón with whom I worked closely when he was a guest researcher at Karlstad University, Albena Mihovska, Jijun Luo, Emilio Mino, and Elias Tragos for insightful discussions at our meetings in the WINNER project, Carmen Botella, who supported me when I learned how to implement radio channels, precoding, etc., in Matlab, and also explained many radio related problems to me, Tommy Svensson, who often helped me to pin down important problems before they went big, and last, but not least, Rikke Apelfröjd, for sharing her knowledge and generously giving her time to discussions. A special thanks go to my colleagues at the Computer Science department. I really enjoy working with them. I would like to thank my colleagues in the DISCO research group for their support, especially when I started my research studies. They helped me with practical problems in the research laboratory as well as gave me advise on scientific writing. I especially would like to acknowledge Katarina Asplund and Stefan Alfredsson.

I would also like to extend my thanks to the people I have been fortunate to work with in various projects, first, my already mentioned co-authors from the WINNER project, then, Bojne Svensson and Jonas Eriksson in the Telia project, the participants in the Wireless IP project, and from the more recent Dynamic Multipoint Wireless Transmission project, I would like to thank Daniel Aronsson, Tilak Rajesh Lakshmana, Jingya Li, Yutao Sui, Behrooz Makki, Thomas Eriksson, Nima Jamaly, and Anders Ahlén.

I wish to express gratitude for financial support from Karlstad University, Telia, the PCC++ graduate school, the Wireless IP project, the WINNER project financed by EU, and the Dynamic Multipoint Wireless Transmission project financed by the Swedish Research Council.

Finally, I wish to express the deepest gratitude and appriciation to my husband Tomas Klockar, both for discussing research and programming problems with me, and for his encouragement at times when I needed it most.

Kista, March 23, 2015 Annika Klockar

List of Appended Papers

I. Annika Wennström, Johan Garcia, Jan H. Gustafsson and Anna Brunstrom. TCP and GSM Link Layer Interactions: Implications for the Wireless Internet. In *Proceedings of IEEE Semiannual Vehicular Technology Conference (VTC 2001 Spring)*, Rhodes, Greece, May 2001.

II. Annika Wennström, Anna Brunstrom, Johan Garcia and Jan H. Gustafsson. A GSM/GPRS Testbed for TCP Performance Evaluation. In *Proceedings of International Workshop on Wired/Wireless Internet Communications (WWIC 2002)*, Las Vegas, USA, June 2002.

III. Annika Wennström, Anna Brunstrom, Juan Rendón and Jan H. Gustafsson. The Impact of GPRS Buffering on TCP Performance. In *Karlstad University Studies 2003:41*, Karlstad, Sweden, 2003.

IV. Annika Klockar, Albena Mihovska, Jijun Luo, Emilio Mino, and Elias Tragos. Network Controlled Mobility Management with Policy Enforcement towards IMT-A. In *Proceedings of International Conference on Communications, Circuits and Systems (ICCCAS 2008)*, Fujian, China, 2008.

V. Annika Klockar, Carmen Botella, Tommy Svensson, Anna Brunstrom and Mikael Sternad. Utility of Joint Processing Schemes. In *Proceedings of the 2010 7th International Symposium on Wireless Communications systems (ISWCS 2010)*, York, United Kingdom, September 2010.

VI. Annika Klockar, Carmen Botella, Mikael Sternad, Anna Brunstrom and Tommy Svensson. Utility as a User Selection Criterion for Coordinated Multi-Point Systems. In *Proceedings of 24th Annual IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC 2013)*, London, UK, 2013.

VII. Annika Klockar, Mikael Sternad, Anna Brunstrom and Rikke Apelfröjd. User-centric Pre-selection and Scheduling for Coordinated Multipoint Systems. In *Proceedings of the 2014 11th International Symposium on Wireless Communications systems (ISWCS 2014)*, Barcelona, Spain, August 2014.

Comments on my Participation

Paper I I implemented the conversion of the RLP frames to the pcap format so that TCP performance could be analyzed with standard tools. Johan Garcia implemented software required to extract RLP frames at the user side, and Jan H. Gustafsson provided software required to extract user data and control information at the network side. I analyzed the results and is the main author of the paper. Section 3.2 was written by Jan H. Gustafsson. I wrote the rest of the paper with help from my co-authors.

Paper II For Paper II the same experimental set up was used as in Paper I. I developed the tool for RLP analysis and analyzed the experimental results. Also in this paper the section about the radio environment was written by Jan H. Gustafsson and the rest by me with my co-authors providing comments.

Paper III The experimental design was a joint work between the authors. Juan Rendón and I analyzed the results. I wrote the report with my co-authors providing comments.

Paper IV Paper VI is a joint work between the authors, but Albena Mihovska and I had the overall responsibility of the paper. I proposed and wrote sections 2, 3a and 3b. The introduction and the conclusions were written by Albena Mihovska and me.

Paper V Carmen Botella implemented and performed simulations of data rates. I extended the simulations with application utilities and analyzed the simulation results. I am the main author of the paper, but Carmen Botella wrote parts of the sections on the simulation setup and the joint processing schemes. The co-authors contributed with comments.

Paper VI I am the principal author of Paper VI, with the other authors providing comments. The simulation code is a joint work between Carmen Botella and me. I proposed and implemented the user selection schemes. Carmen Botella implemented the radio channel model, the precoding and the power allocation. I designed the experiments and analyzed the results.

Paper VII I am the principal author of Paper VII, with the other authors providing comments. I proposed the pre-selection scheme and the time domain scheduler. I implemented the simulation code. For the radio related parts, I often consulted Rikke Apelfröjd and Mikael Sternad. The frequency domain scheduler was based on a scheduler used in a previous paper by Rikke Apelfröjd and Mikael Sternad.

Other Publications

- A. Wennstrom, A. Brunstrom, J. Rendón, and J. Gustafsson. A GPRS testbed for TCP measurements. *Proceedings 4th IEEE Conference on Mobile and Wireless Communications Networks (MWCN 2002)*, Stockholm, Sweden, pp 320 - 324, September 2002.
- A. Wennström and A. Brunström. TCP over GPRS: The Effect of Preemption. *First Swedish National Computer Networking Workshop (SNCNW2003)*, Stockholm, Sweden, September 2003.
- A. Wennström, S. Alfredsson, and A. Brunstrom. TCP over Wireless Networks. Report. *Karlstad University Studies 2004:21*, Karlstad, 2004, ISBN 91-85019-88-7.
- A. Wennström. A Measurement Based Performance Study of TCP over GSM/GPRS. Licentiate thesis. *Karlstad University Studies 2004:22*, Karlstad, 2004, ISBN 91-85019-89-5.
- A. Wennström, A. Brunstrom, J. Rendón, and J. Gustafsson. The Impact of preemptive priority in GPRS on TCP performance: A measurement study. *1st International Conference on E-business and Telecommunication Networks (ICETE 2004)*, Setubal, Portugal, August 2004.
- A. Wennström, A. Brunstrom, and J. Rendón. Impact of GPRS buffering on TCP performance. *IEE Electronics Letters*, Volume 40, Issue 20, pp 1279-1281, September 2004.
- A. Adinoyi et al., IST-2003-507581 WINNER. D3.2 ver 1.1: Description of identified new relay based radio network deployment concepts and first assessment by comparison against benchmarks of well known deployment concepts using enhanced radio interface technologies, February 2005.
- D. Schultz et al., IST-2003-507581 WINNER. D3.5 version 1.0: Proposal of the best Suited Deployment Concepts for the identified Scenarios and related RAN Protocols, January 2006.
- E. Mino et al., IST-4-027756 WINNER II. D4.8.3: Integration of cooperation on WINNER II System Concept, November 2007.
- A. Mihovska, E. Tragos, E. Mino, Ch. Mensing, R. Fracchia, S. Horrich, L. Hui, A. Klockar, and S. Kyriazakos. Requirements and Algorithms for Cooperation of Heterogeneous Radio Access Networks, *Wireless Personal Communications, Springer*, Volume 50, Issue 2, pp 207-245, July 2009.
- A. Klockar, C. Botella, T. Svensson, A. Brunstrom, and M. Sternad. Utility as a User Selection Criterion for Coordinated Multi Point Systems. *IEEE Swedish Communication Technologies Workshop (Swe-CTW)*, KTH, Sweden, October 2011.

- A. Klockar, C. Botella, M. Sternad, A. Brunstrom, and T. Svensson. Utility as a User Selection Criterion for Coordinated Multi-Point Systems. *IEEE Swedish Communication Technologies Workshop (Swe-CTW)*, CTH, Gothenburg, Sweden, August 2013.

- A. Klockar, M. Sternad, A. Brunstrom, and R. Apelfröjd. User-centric Pre-selection and Scheduling for Coordinated Multipoint Systems. *IEEE Swedish Communication Technologies Workshop (Swe-CTW)*, Mälardalen University, Västerås, Sweden, June 2014.

Abbreviations

3GPP	The 3rd Generation Partnership Project
ACK	Acknowledgment
AQM	Active Queue Management
ARQ	Automatic Repeat Request
Bps	Bytes per second
bps	bits per second
BS	Base Station
BSC	Base Station Controller
BSSGP	Base Station System GPRS Protocol
BTS	Base Transceiver Station
C/I	Carrier to Interference ratio
CDMA	Code Division Multiple Access
CJP	Centralized Joint Processing
CoMP	Coordinated Multi Point
CS	Coding Scheme
CSI	Channel State Information
cwnd	congestion window
dB	Decibels
dBm	Decibel milliwatts
DJP	Distributed Joint Processing
DL	Downlink
DTX	Discontinuous Transmission
dupacks	duplicate acknowledgements
ETSI	European Telecommunications Standards Institute
FD	Frequency Domain
FDD	frequency division duplex
FDMA	Frequency Division Multiple Access

GGSN	Gateway GPRS Support Node
GMSK	Gaussian Minimum Shift Keying
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
GTP	GPRS Tunneling Protocol
GW	Gateway
HARQ	Hybrid ARQ
HDLC	High-Level Data Link Control
Hz	Hertz
IETF	Internet Engineering Task Force
IMT-Advanced	International Mobile Telecommunications-Advanced
IP	Internet Protocol
IPCL	IP Convergence Layer
ISDN	Integrated Services Digital Network
ITU	International Telecommunication Union
KB	Kilo Bytes
LAN	Local Area Network
LLC	Logical Link Control
LTE	Long Term Evolution
MAC	Medium Access Control
MIMO	Multiple Input Multiple Output
MS	Mobile Station
MSC	Mobile Switching Center
MSS	Maximum Segment Size
NMT	Nordic Mobile Telephone
NUM	Network Utility Maximization
OFDMA	Orthogonal Frequency Division Multiple Access
pcap	packet capture

PDCH	Packet Data Channel
PDCP	Packet Data Convergence Protocol
PHY	physical layer
PJP	Partial Joint Processing
PPP	Point to Point Protocol
PSTN	Public Switched Telephone Network
QoE	Quality of Experience
QoS	Quality of Service
RB	Resource Block
RED	Random Early Detection
RLC	Radio Link Control Protocol
RLP	Radio Link Protocol
RN	Relay Node
RRC	Radio Resource Control
RRM	Radio Resource Management
RTT	Round Trip Time
rwnd	receiver window
SACK	Selective Acknowledgment
SB	Score-Based scheduler
SDU	Service Data Unit
SGSN	Serving GPRS Support Node
SINR	Signal to Interference plus Noise Ratio
SLA	Service Level Agreement
SNDCP	Subnetwork Dependent Convergence Protocol
SNR	Signal to Noise Ratio
ssthresh	slow start threshold
taps	Transport Services
TBR	Target Bit Rate

TCP	Transmission Control Protocol
TD	Time Domain
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TEMS	Test Mobile System
TU3	Typical Urban 3 km/h
UDI	Unrestricted Digital Information
UDP	User Datagram Protocol
UL	Uplink
UMTS	Universal Mobile Telecommunications System
UT	User Terminal
W	Watts
WCDMA	Wideband Code Division Multiple Access
WINNER	Wireless World Initiative New Radio
WWW	World Wide Web

Contents

Introductory Summary

1 Introduction

Today, the Internet is accessible over cellular networks in most places, and the penetration of smart phones and tablets have increased rapidly over the last years. It is almost hard to remember that there was a time when a mobile phone was a very large piece of equipment only for telephony. The first generation of cellular networks (1G) for telephony was based on analog technology using frequency division multiple access (FDMA) [32] for the radio interface. An example of such a system is the Nordic Mobile Telephone (NMT) system which was introduced in 1981 in Sweden and Norway [1, 43].

With the introduction of the World Wide Web (WWW) around 1990, and then with Global System for Mobile Communications (GSM) [60] shortly after, there was a desire to use cellular networks not only for telephony, but also to access the Internet. The GSM system belongs to the second generation of cellular networks (2G). The systems of this generation are digital and apply time division multiple access (TDMA) [32], which allows multiple users to share a frequency channel by dividing the frequency channel into time slots and allocating time slots to the users. Furthermore, TDMA in combination with frequency hopping in GSM reduces negative impact of fading and makes it harder to eavesdrop the traffic.

About ten years after the introduction of GSM, the General Packet Radio Service (GPRS) [13, 25] system became commercially available. GPRS is often described as a 2.5G system, since is it an extension of GSM, and it is seen as an intermediate step toward the third generation of cellular networks (3G). In contrast to 1G and 2G, which use circuit switched connections both for voice and data transmission, the 2.5G systems provide a packet switched service for data traffic. Many more users can share the available resources, since a channel is only temporarily assigned to a mobile station. With the 3G systems, such as Universal Mobile Telecommunications System (UMTS) [22, 42, 62], the data rates have increased multi-fold. For radio access, wideband code division multiple access (WCDMA) [22, 42] is used. Both circuit and packet switched connections are provided.

Now, we have witnessed the deployment of 4G realized with Long Term Evolution (LTE) [10] and LTE-Advanced [31, 66]. LTE is sometimes denoted 3.9G, while LTE-Advanced is defined as true 4G. LTE-Advanced is defined as true 4G, since LTE-Advanced fulfills the requirements for International Mobile Telecommunications-Advanced (IMT-Advanced) [47, 66] by the International Telecommunication Union (ITU). There are, however, many similarities between LTE-Advanced and LTE, since LTE-Advanced is intended to be backwards compatible with LTE. LTE is an Internet Protocol (IP) network, and, hence, there are only packet switched connections. The data rates are further improved as compared to 3G. In LTE, orthogonal frequency division multiple access (OFDMA) [81] is applied for radio access, and multiple input multiple output (MIMO) transmission with up to four antennas are used. Hence, the radio channels are even more efficiently used than in UMTS. In LTE-Advanced, the data rates are further improved by the use of carrier ag-

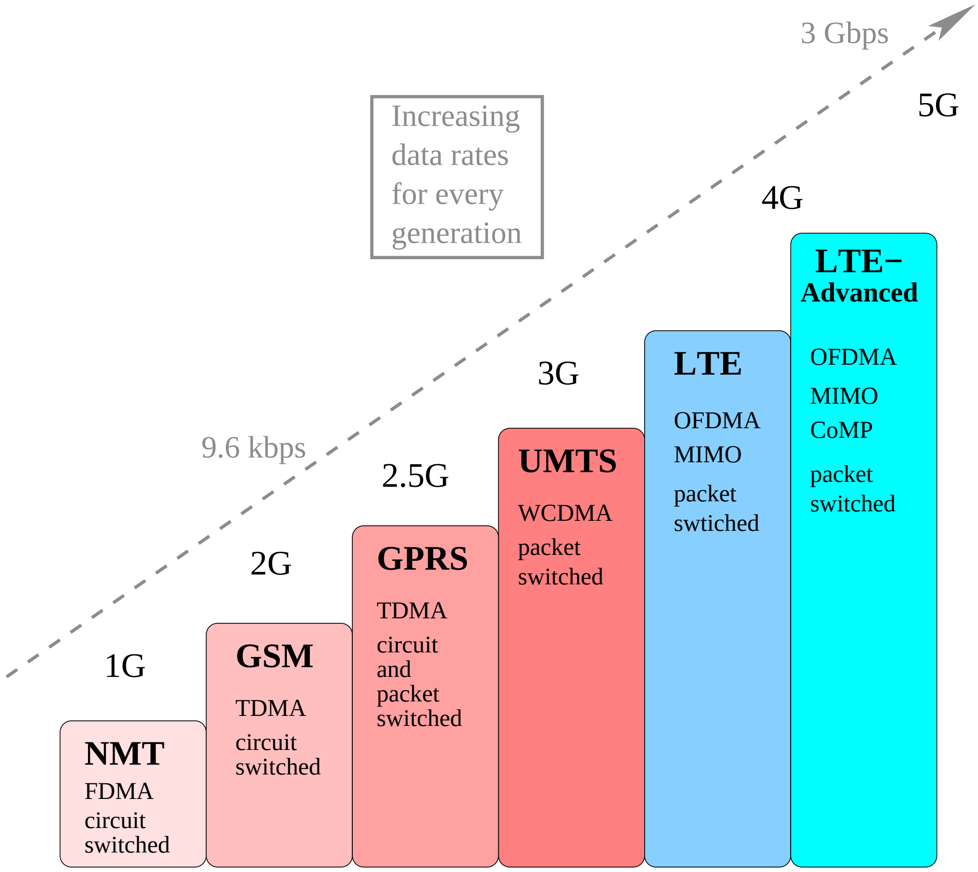

Figure 1: Cellular systems

gregation, relay nodes, and advanced multi-antenna techniques, such as higher order MIMO and Coordinated Multipoint techniques (CoMP), in which base stations (BSs) cooperate to serve the users.

The development of cellular networks has been tremendous from the 1G analog systems for telephony to the 4G multi-antenna systems with advanced inter-networking capabilities for a wide range of services including telephony. A summary of the described cellular networks is illustrated in Fig. 1. Even though the data rates in cellular networks can now be comparable to the data rates in wired networks, there are some inherent differences between cellular and wired networks.

Some of the differences are related to the properties of the medium of the links. Wireless links have lower and more variable quality. The attenuation is much higher over the air than in a wire. A signal is affected by the distance to the base station, and various effects due to objects in the surroundings. The result is that signals are exposed to fading, interference, and noise. Errors are therefore more common over wireless links, and error recovery is required, such as coding to detect and correct errors, and retransmissions by Automatic Repeat Request (ARQ) protocols. Retransmissions introduce delay and delay variations. Another source of delay is queuing delay. As the capacity over the wireless link is variable and, in most cases, also lower than over the backbone,

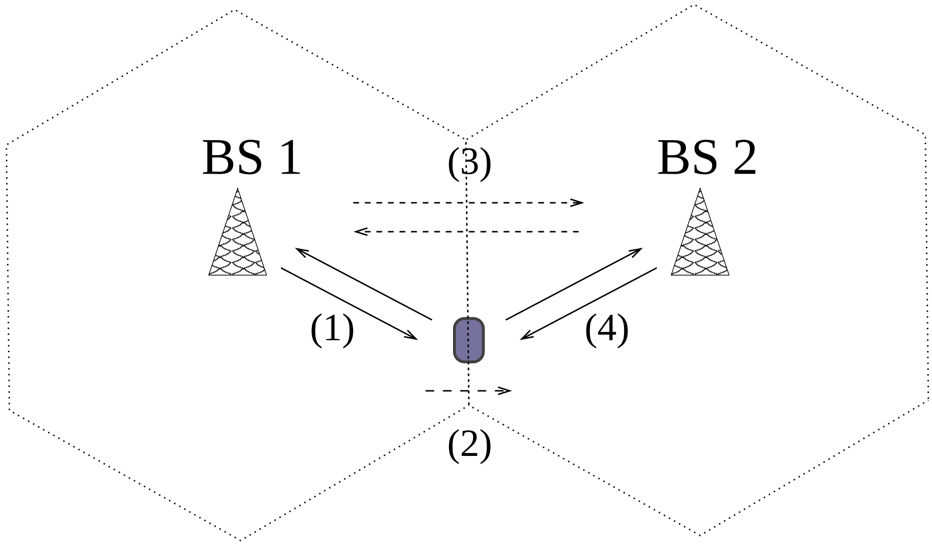

Figure 2: Handover

data need to be buffered before they can be transmitted over the air.

User mobility and handover in cellular networks also result in variations in the communication that are not present in wired networks. Fig. 2 illustrates a handover for a user that moves away from BS 1 to the area served by BS 2. (1) First, the user is served by BS 1. (2) Then, the user moves towards BS 2. (3) The call, or connection, is transferred from BS 1 to BS 2. (4) The users is served by BS 2.

Handover results in delay, and, depending on the type of handover, also in data loss. Buffering is important to reduce data loss during handover and during periods of low radio quality. The amount of buffered data may at times be extensive which results in long queuing delays [7, 18, 35, 37, 40].

The Internet protocols were developed to work over different networking technologies. The link layer is, therefore, not part of the de facto standard for the Internet. The network layer protocol, the Internet Protocol (IP) [68], delivers packets from the source to the destination. The complexity of IP is low, as it does not keep connection states in the routers. This implies that the network is flexible in the sense that a new path can be found to the destination even if some routers go down.

Complex functions, such as error recovery, are expected to be handled by the end nodes. The Transmission Control Protocol (TCP) [69] is the main transport protocol that provides retransmissions in the Internet. After a series of congestion collapses starting in October 1986 and the addition of the congestion control functionality to TCP by Van Jacobson [39], TCP interprets packet loss as a sign of congestion in the network. The transmission rate is also reduced in response to packet loss in order not to contribute more to the congestion. The actions to control congestion work well in wired networks, but when the Internet protocols are used over cellular networks, the assumption of congestion as the only cause of packet loss is incorrect, and the actions to control congestion may reduce performance more than necessary.

In the first four papers of this thesis, we have investigated TCP performance in cellular networks. High TCP performance in cellular networks is important, since many Internet applications rely on TCP for reliable transport, such as applications for web browsing, file transfer, e-mail, and streaming video. We have measured the impact of delay and packet loss on TCP, more specifically the interaction between TCP and the link layer protocol in GSM and the impact of buffering in GPRS. Delay and packet loss may also occur as results of handover. We have studied techniques to improve the handover situation for TCP in the Wireless World Initiative New Radio (WINNER) system [57, 63].

TCP performance may give an idea on how the user experiences an Internet service. However, not all Internet applications use TCP, and requirements on quality of service (QoS) vary depending on the application type, even for applications that use the same transport protocol. Internet applications are often divided into traffic classes depending on their QoS requirements. Some application types, such as interactive video, for example, are especially demanding, since they require both high data rates, low delays and low jitter. The traffic volumes generated by these demanding application types are increasing. In wired networks, over-provisioning of resources is often used to increase the data rates.

In cellular networks, on the other hand, the data rates are limited mainly due to inter-cell interference. In order to increase the data rates, all frequencies are often reused in all cells, which also results in increasing interference levels. CoMP transmission techniques have been proposed to reduce, or even exploit, inter-cell interference [3, 48, 49]. The cost of CoMP is increased overhead due to more feedback from the users in the uplink and a higher load on backhaul links between the base stations (BSs). As CoMP introduces overhead, system wide CoMP is not an option. Therefore, some type of limited CoMP service is often proposed, such as reducing the number of BSs that cooperate and/or the number of users to serve with CoMP. In the last three papers of this thesis, we have examined CoMP for different traffic classes, and suggest that a subset of the users are served with CoMP. We propose and evaluate a user selection algorithm for CoMP that takes user experience into account.

The rest of the thesis is structured as follows. In Section 2, the objective of the research is defined. An overview of the research areas is given in Section 3 before the research questions are stated in Section 4. In Section 5, the research methodology is described. The main contributions of the research in this thesis are indicated in Section 6. In Section 7, the appended papers are summarized. Conclusions and future research are discussed in Section 8. The final part of the thesis consists of the appended papers.

2 Research objective

The overall objective of the research presented in this thesis is to provide the users with a satisfactory service of Internet applications in cellular networks by using the network resources efficiently from the user perspective. By exploiting knowledge about protocols and applications that the end users run, network resources can be allocated to the users according to some user centric objective, e.g., maximizing the level of satisfaction or maximizing fairness, in order to benefit the users more than if only system performance is targeted.

TCP performance is the key topic of the first four papers. TCP has a large impact on user experience, since it is used by many Internet applications. The objective of the research involving TCP is to investigate how problems due to the wireless medium, buffer management and handover affect TCP performance, and how the problems can be solved to gain the users.

The user experience of a service depends not only on the QoS provided by the network and on TCP performance. Also the traffic class of the application has a large impact. In the last three papers, user experience is quantified with utility functions of Internet application types. We evaluate user experience in a CoMP enabled system. The resources for CoMP are assumed to be limited and therefore only a select number of the users can be served with CoMP. We study application utility as a basis for selecting users for limited CoMP, and propose a user selection scheme for CoMP.

3 Research areas

In this section, research areas relevant to the research objective are described. The specific research questions addressed in this thesis are then defined in the next section.

3.1 TCP performance in cellular networks

Many applications in the Internet use TCP for reliable data transmission, including WWW, e-mail, file transfer, video, and audio. The algorithms for error recovery and congestion control in TCP are based on the assumptions that data are lost mainly due to congestion and that data loss due to transmission errors is rare [8, 39].

TCP is a connection oriented ARQ protocol that provides a reliable byte stream to its users. This implies that the TCP receiver delivers the byte stream in the same order as it was transmitted. Before transmission, the TCP sender divides the byte stream into TCP packet data units (PDUs) called segments. Each byte is numbered and the first byte is used as sequence number in the segment. When a segment is received, the TCP receiver responds with a cumulative acknowledgement of the last segment received in sequence.

The TCP sender estimates the round trip time (RTT), which is the time it takes from the transmission of a segment until an acknowledgement comes back. The RTT increases, for example, as intermediate buffers between the sender and the receiver are filled. The transmission rate is increased until a segment is considered lost or until the receiver indicates that the receive buffer is full.

The congestion window (cwnd) limits the number of segments that the TCP sender is allowed to transmit before an acknowledgement needs to be received. The cwnd is an estimation of the instantaneous capacity of the network, including links and buffers, between the TCP sender and the TCP receiver. A segment is considered lost by the TCP sender, if the retransmission timer expires or if three duplicate acknowledgements (dupacks) are received. Data loss is interpreted as a signal of congestion in the network, and, therefore, the TCP sender reduces the transmission rate and the cwnd. The transmission rate and the cwnd are more drastically reduced after a timeout event than after three dupacks. The dupacks indicate that the network is not completely congested, since segments are received and acknowledged by the TCP receiver. TCP also performs flow control. The TCP receiver indicates the available capacity of the receiver buffer in the acknowledgements, the receiver window (rwnd). The transmission rate is limited by the cwnd or the rwnd, which ever is smaller at the instant when segments are transmitted.

The responses to data loss are less efficient in cellular networks, since data loss is often not caused by congestion, but rather by lack of coverage, user mobility, handover, or sometimes even due to transmission errors. By reducing the transmission rate drastically due to such phenomena, TCP performance is degraded more than necessary and bandwidth resources are wasted.

3.1.1 Radio conditions and retransmission protocols

The quality of a wireless link varies on short timescales, since it is affected by fading, interference, and noise. Poor radio conditions result in transmission errors that require error recovery. As the TCP connection typically involves multiple links, it would not be very efficient to leave the responsibility to recover from errors over a wireless link completely to TCP. Therefore, error recovery is required at lower protocol layers in cellular networks. The first step toward error recovery is performed at the physical layer. Error correcting codes, such as Turbo codes are used to reduce the error rate seen by higher layers, e.g., in LTE-Advanced [5]. The cost of error correcting codes is that more bits are transmitted, as compared to error-free transmission, since redundant bits need to be added for error detection and error correction. Unfortunately, error correcting codes can only correct a limited number of bit errors. Therefore, link layer ARQ protocols are typically used in addition to error correcting codes. ARQ protocols retransmit data that cannot be corrected with error correcting codes. By retransmitting data, delay is introduced. Moreover, the delay is variable depending on the number of retransmissions required to recover the erroneous data.

Delay variations may result in reduced TCP performance, since TCP also uses ARQ for error recovery. If the delay due to link layer retransmissions suddenly increases, then TCP may interpret this as a sign of data loss and start to retransmit the same data, which, of course, is not very efficient. Ideally, TCP and the link layer ARQ protocol interact efficiently and do not retransmit the same data. The data link layer would then protect TCP from bit errors that occur over the wireless link, and, hence, would provide a more reliable link to TCP. In cellular networks, the use of ARQ on the link layer is optional, and, for delay sensitive traffic that can tolerate data loss, typically User Datagram Protocol (UDP) [67] traffic, it may be disabled. For TCP, on the other hand, link layer ARQ significantly increases the performance [51, 52, 56, 65]. There are, however, cases when the protocol interaction is less efficient, as shown in Paper I, Paper II, and [33, 34, 65]. Sudden increases in delay trigger TCP to retransmit data regardless of the cause of the delay. In cellular networks, there are also other sources of delay, except from bit errors due to the wireless medium, some of which are discussed in the following sections.

3.1.2 Buffer settings of intermediate nodes

Another source of delay in cellular networks is the buffer before the wireless link. As the quality of a wireless link varies over time with periods of poor radio conditions, the buffer needs to be large enough to store data when the link quality is low without causing unacceptably many packet drops. On the other hand, if the buffer is very large, then the delay experienced by TCP, and by the applications using TCP for transport, may become very long. This problem has been reported for many cellular systems, e.g., for 2.5G GPRS in Paper III and in [18, 35, 52], 3G UMTS [7, 40], 3.5G HSPA+ [7, 40], and 4G LTE [7, 37, 40], even though the problem is less pronounced in 4G LTE [7, 37]. Also in wired networks with fixed hosts, problems with excessive buffers that cause long delays, so called bufferbloat [30], occur e.g., when a large buffer in a router is filled due to congestion on an outgoing link.

For traditional TCP applications such as file transfer and e-mail, which typically generate bulk data transfers, delay is not so important, but the main performance issue is throughput. Today, TCP is used also for delay sensitive interactive traffic, such as video and audio. The queuing delay of intermediate buffers may become unacceptable for interactive traffic, if it arrives at a large buffer that is already filled up with bulk data.

The alternatives to reduce long queuing delays are to use smaller intermediate buffers as in Paper III, to use active queue management (AQM) [27, 30, 72, 73], or to modify TCP behaviour [37, 40]. In Paper III, measurement results with two buffer settings in GPRS, corresponding to one excessive buffer and one smaller, indicate that the throughput was almost the same for the buffer settings, but that the delay was significantly reduced for the smaller buffer. The loss rate was higher with the smaller buffer, but as the TCP segments were dropped from the buffer before the radio interface, no extra downlink resources were used. In most other studies based on measurements, the buffer settings could not be changed, since the measurements were conducted in public networks. AQM, such as Random Early Detection (RED) [27], or some variant for cellular networks [72, 73] could be used to limit the amount of data in the intermediate buffer, but the impact of AQM is hard to predict and it has not been studied extensively [40]. In [40], dynamic adjustment of the advertised rwnd is proposed. The advertised rwnd is dynamically adjusted as the RTT changes. When the RTT is increasing, the advertised rwnd is reduced in order to avoid intermediate buffers to be filled. Decreasing RTTs is an indication that smaller amounts of data is buffered and the advertised rwnd is instead increased. Another proposal to avoid timeout when buffers are filled is to estimate the RTT also for dupacks [37].

3.1.3 Preemption and buffering

In cellular systems, it is not possible to guarantee high QoS to all users at all times, since the available resources vary with the radio quality and due to mobility. Furthermore, it is not an option to use over-provisioning, as the capacity is limited. Therefore, admission control and resource allocation need to be flexible. Preemption is used to provide resources to traffic of high priority when the available resources are scarce. Priority and preemption are especially important for critical communications, e.g., in Public Safety LTE [23], but also in commercial networks. In cellular networks, preemption is typically used when new calls/connections are admitted [19, 46].

If there are incoming connections and no resources are available, then, connections with lower priority may lose all or some of their resources to the new connection, depending on whether full or partial preemption is applied. In the case of full preemption, a connection for high priority traffic is allowed to completely preempt resources from connections with lower priority. In GPRS networks, for example, GSM voice calls are allowed to completely preempt resources from GPRS connections which have lower priority [19].

Full preemption typically results in performance loss for the flows of the preempted connection, also in cases when the preemption period is very short. Partial preemption is an alternative that has a lower impact on the preempted flows, but it is only possible when there are resources enough to avoid full preemption. Loss of all resources, of course, degrades performance more for the preempted flows than partial preemption. In Paper III, our measurements of TCP in GPRS under full preemption indicate that TCP stalls for a longer time than the preemption period. When the TCP sender has transmitted as many segments as allowed by the rwnd it stops transmitting and waits for acknowledgements from the TCP receiver. Data is flushed from the buffer in the Base Station Controller (BSC) due to preemption, which results in consecutive segment losses. The transfer is not restarted again until the TCP sender receives an acknowledgement.

With partial preemption instead, the problem would not occur, but partial preemption is not sufficient in all cases. In [17], a priority-scaled preemption scheme for LTE is proposed in which connections with lower priority lose more resources than the ones with higher priority. A similar idea, a fairness-based partial preemption algorithm for LTE-Advanced, is proposed in [46]. Fairness is taken into account when the data rates of the ongoing connections are reduced until a guaranteed data rate is reached. For both proposals, if even more resources are required, then full preemption is applied.

3.1.4 Handover, data loss, reordering, and duplication

Handover is another situation in which buffers play an important role. In many systems, hard handover is used, which means that the user is only connected to one BS at a time. There is a short time period when the user is disconnected after the link to the source BS is disconnected until the new link to the target BS is set up. In the downlink, the data that arrive to the target BS during the time when the user is disconnected, will be lost, unless the data are forwarded to the target BS. In a soft handover, on the other hand, data is not lost, since the user terminal is connected to both the source and the target BSs during the handover. Even if soft handover could be applied in any system it is only used in Code Division Multiple Access (CDMA) systems, since it would be to costly for TDMA/FDMA.

In order to reach high data rates, more BSs are required, which implies smaller cells. With smaller cells, handover occurs more frequently. In systems of 4G or higher, radio related functionality has moved closer to the radio interface, the BS, instead of being located in the backbone, such as in the BSC in GPRS. As the Radio Link Control Protocol (RLC) connection terminates in the BS, the RLC protocol cannot use ARQ to recover data lost due to handover. As only hard handover is supported, control information about the user, user context, needs to be forwarded from the source BS to the target BS. Data loss can be avoided if the source BS also forwards user data to the target BS during handover. In [70, 71], it was shown that context transfer of RLC service data units (SDUs), in most cases, improves handover performance with respect to TCP throughput, or at least does not have a negative impact [11]. The exceptions are when packets are reordered, which is discussed below. For LTE, data forwarding during handover was specified for LTE as context transfer of Packet Data Convergence Protocol (PDCP) SDUs [2], which would have the same effect as context transfer of RLC SDUs. More recent simulations in [61], support the result that context transfer improves TCP throughput, but that the delay may be increased if the buffer in the BS is large.

TCP performance may also be degraded by reordered and duplicated data, since the TCP receiver transmits dupacks in response and after three dupacks data is considered lost by the TCP sender. Reordering may occur in downlink (DL) due to handover, if new data from the gateway (GW) to the target BS arrive at the user before data forwarded from the source BS. In uplink (UL), retransmissions after handover of data lost before handover may cause reordering at the GW. In [11], forwarded data is proposed to have higher priority in order to avoid reordering. Flow control between the GW and the BS, as proposed in Paper IV, can also be used to prevent reordered data. Handover may also cause duplication of data, e.g., if RLC data is delivered to a higher layer before handover and the acknowledgement is lost, then the same data is retransmitted and delivered again. The PDCP protocol in LTE has functionality based on the PDCP sequence number for detecting reordering and delivering data in-sequence, as well as detection and removal of duplicated data [2].

3.2 User selection for CoMP

The data traffic volumes have increased enormously over the last years and are still increasing. This is also the case with the amount of traffic generated by applications that require high data rates and/or timely delivery. One way to enable for higher data rates is to reuse all frequencies in all cells. The problem with such tight frequency reuse is that it causes high interference between cells. The inter-cell interference needs to be mitigated in some way, otherwise the data rates are limited due to inter-cell interference, especially at the cell edges [15]. Inter-cell interference can be reduced, or even exploited, with techniques for CoMP. The results are higher data rates and more even data rates over the cell areas compared to single BS systems without CoMP. This would be of advantage to many applications, such as VoIP and video, that require relatively constant data rates and delays. As described in the previous section, also TCP performs better with low variations in data rates and delays. With CoMP, a more stable service could be provided over the cell area.

In 3GPP, two types of CoMP techniques have been defined, coordinated scheduling/beamforming and joint processing and transmission [3, 48, 49]. With coordinated beamforming/scheduling CoMP techniques, inter-cell interference is suppressed, since the BS coordinate their transmissions so that low interference occurs. For example, if single antenna BSs are coordinated by scheduling, then at most one user is served by one of the coordinated BSs at a time. Joint processing and transmission CoMP techniques, on the other hand, exploit interference to form useful signals that are transmitted simultaneously from the coordinated BSs to the users. For this CoMP technique, the BSs need to be highly synchronized. For both CoMP techniques, however, information about inter-cell interference is required at the BSs, which implies that the users must transmit more channel state information (CSI) to the BSs than in a conventional single BS system without support for CoMP. Furthermore, user data need to be exchanged between BSs, since user data must be available at all the coordinating BSs. The overhead introduced with CoMP, due to the feedback of CSI in the uplink from the users and due to the forwarding of user data between the BSs, is, however, too large for CoMP to be applied system wide to all users [78, 83]. Selective, or limited, CoMP is a more practically conceivable alternative [78]. Clustering of BSs is used to reduce the overhead, but also since CoMP is not possible between BSs located far apart or far from the users. Other means to limit CoMP is to select or schedule a subset of the users for CoMP, such as the cell edge users [28], or to use only some of the resources for CoMP [12]. In this thesis, we consider the downlink of joint processing and transmission CoMP, which in the following is denoted CoMP, if not otherwise stated.

3.2.1 CoMP schemes

In a clustered CoMP system, the clusters could be fixed and non-overlapping sets of BSs so that all BSs in the cluster cooperate to serve the users. The problem with fixed clusters is that users located in an area between clusters would not gain as much from CoMP and even experience inter-cluster interference. This problem may be solved by introducing dynamic clusters instead [54]. Then, the BSs closest to a user could temporarily form a cluster to efficiently serve the user. The number of BSs that cooperate to serve a user could be varied in order to include only BSs to which the user has high quality channels. An example is the partial joint processing CoMP scheme in [15], which we also apply in Paper V.

The control of a CoMP system could be centralized to a central control unit or distributed between the BSs. If centralized control is applied, then a central control unit, located in one of the BSs or in a separate network node, is responsible for precoding the user data before transmission. Global CSI is required for the precoding, and, therefore, information about all channels between the BSs and the users in the cluster need to be available at the central control unit. In a distributed CoMP system, the BSs are instead responsible for precoding. Each BS operates on local CSI, and, therefore, only information about the channels between the BS and the users is required. The BSs do exchange some information, such as scheduling information [15, 80]. The performance of distributed CoMP is, however, lower than with a centralized solution.

3.2.2 CoMP user groups

The resulting data rates when using CoMP depend highly on the set of users served with CoMP in the same time and frequency resource. User data of the users in the CoMP group are precoded at the transmitter to pre-subtract the inter-cell interference from the signal that will be transmitted. The precoding schemes that are most often applied in the literature are dirty paper coding and zero-forcing [29, 78]. Dirty paper coding more efficiently removes interference than zero-forcing, but has a high complexity. Zero-forcing is often preferred over dirty paper coding, since the performance is almost as high, but the complexity is much lower. In the papers on CoMP included in this thesis, zero-forcing precoding is used. A problem with zero-forcing that may result in reduced data rates, is the problem of power normalization loss [9]. This means that, in order not to cause more interference, the transmission power needs to be adjusted to the user that requires the lowest transmission power to receive a useful signal, typically to the user located closest to a BS.

The user group selected for CoMP has a large impact on the results. A group of users is efficiently served with CoMP if the users are spatially compatible. To compute the CoMP group that would provide the optimal result is a complicated problem, since all possible CoMP groups need to be considered. Therefore, in practice, heuristic algorithms are required to find efficient suboptimal user groups for CoMP, [9, 21, 55, 78, 79]. The problem of user grouping for CoMP is strongly related to scheduling, which is the topic of the next section.

3.2.3 Scheduling for CoMP

In cellular networks, it is necessary to strive for spectral efficiency when resources are allocated [16]. Even though the capacity of cellular networks has increased tremendously with each system generation, the resources are too limited for over-provisioning to be an option. For CoMP, resources in multiple cells are allocated to the users in a CoMP group. Hence, each BS that takes part in a CoMP transmission allocates resources to all the users in the CoMP group, both users in its own cell as well as users in the cells of the other BSs. If the resource allocation is inefficient, then the same resources could have been used in each cell instead to serve more users with single BS transmission.

A scheduling algorithm that maximizes the sum rate would be spectrally efficient, but suffers from the disadvantage that some users may be starved and not served at all. As all admitted users must be served at some point, scheduling algorithms need to consider fairness between the users. There are many definitions and measures of fairness [76]. Proportional fairness [44] of rate is often used for scheduling [16] to impose some degree of fairness and to ensure that all users are served. In order to achieve high spectral efficiency, the degree of rate fairness is typically lower than that of a completely rate fair scheduling algorithm that would provide equal throughput for all users, denoted blind equal throughput in [16], or, e.g., a round robin scheduler that would be resource fair and share the time slots equally between the users without taking their channel strength within each transmission resource into account.

With proportional fair scheduling of rate, users are scheduled when the expected data rate in the next time slot is high related to the past average data rate of the user [55]. Examples of proportional fair scheduling used for CoMP are found in [21, 55, 78, 79]. The objective of scheduling for CoMP is that the users scheduled in the same time-frequency resource would constitute an efficient CoMP group. The crucial part of proportional fair scheduling for CoMP is to determine the expected data rates in the next time slot. For this some simplifying approximation of data rates is required, since the resulting CoMP group is not known at the time when the users are scheduled. A heuristic approach to scheduling for CoMP that results in close to optimal sum rates is presented in [9]. The complexity and the overhead are kept low, since scheduling is performed locally in each cell, as in conventional single BS systems. Hence, the BSs schedule users without considering the resulting CoMP group. This implies that conventional scheduling algorithms, such as proportional fair scheduling, or, as in [9], score-based scheduling [14], can be used. Score-based scheduling has similar performance as proportional fair, but is based on data rate statistics rather than on average data rates. In the next step of the strategy proposed in [9], the users scheduled in the same time-frequency resource form a CoMP group. The main reason to the high performance of this approach is that the scheduled users do not have their strongest channel to the same BS, since each BS schedules the users for which relatively high data rates are expected. The precoding becomes efficient and the power normalization problem is avoided in most cases.

Scheduling for CoMP, such as the approaches described above, typically fall into the category of channel-aware/QoS-unaware strategies in [16], since the QoS requirements of the users are not considered. Hence, the main focus is on reaching a high sum-rate rather than on the QoS requirements of the individual users. Fairness is, however, taken into account in order to ensure that all users are served. The scheduling for CoMP described above is generally intended for allocation of frequency resources, resource blocks, in an OFDMA system.

For users that have strict QoS requirements, e.g., on constant data rates or delay, scheduling strategies that are QoS-aware could be applied. More user centric QoS-aware scheduling strategies could be imposed in the time domain to select the users that the spectrally efficient frequency domain scheduler should allocate resources for in the next time slot. In [58], for example, the users are divided into two sets in the time domain depending on whether their data rate requirements are met or not. Users that require more resources are scheduled according to the blind equal throughput strategy, and the rest of the users with proportional fair.

A step further toward even more user centric scheduling strategies, beyond QoS-aware scheduling, is to take user experience into account and use Quality of Experience (QoE)-aware scheduling. How a user experiences the QoS offered by the network depends on the application that the user runs [75]. Internet applications can be classified according to the characteristics of the generated traffic. In [75], the following application types are characterized: elastic, hard real-time and adaptive real-time. Elastic applications are exemplified by traditional TCP applications such as file transfer and e-mail. The QoS requirements are here relatively low, since elastic applications are capable of providing satisfactory user experience in most cases. Hard real-time applications, such as speech, typically require short delays and relatively low, but constant, data rates. Adaptive real-time applications, such as streaming media, often require higher data rates than hard real-time applications, but have less strict delay requirements, and, hence accept some variations in the data rate. Utility functions of the application types can be used to express the level of user satisfaction. Then, application utility could be used by a QoE-aware scheduler to represent user experience. There are many examples for stationary networks of resource allocation schemes that have the objective to maximize the application utility in the network, as a network utility maximization (NUM) problem, the first of which was proposed for elastic traffic in [44, 45]. It is, however, hard to find efficient algorithms for the optimal solution if also users with adaptive and hard real-time applications are included, since the utility function for adaptive applications is nonconcave and the one for hard real-time applications discrete [20, 36, 77]. The alternatives to solving the NUM problem is either to reformulate the problem in order to avoid finding the maximum of the nonconcave and discrete utility functions [41, 85, 86], or to propose heuristic algorithms that find suboptimal solutions instead [36].

4 Research questions

From the general research objectives and the discussion above of the research areas the following specific research questions are formulated:

1. *How is TCP affected by the radio conditions and link layer retransmissions?*

   The impact of link layer retransmissions on TCP is the research problem addressed in Paper I and in Paper II. Measurements were performed in a GSM network dedicated for test purposes, which was provided by Telia, at the time the largest operator in Sweden. The results indicate that TCP and RLP [24] interact in a constructive way in most cases. But, we observed cases in which the delays increased suddenly due to RLP retransmission and this had a negative effect on TCP performance. A TCP timeout event was triggered and the protocols concurrently retransmitted the same data. However, in most cases the variability in delay was low and RLC retransmissions were completed before a TCP timeout event was triggered. At the time when the papers were written, there was a discussion about TCP and link layer retransmission. Some participants in this discussion had the opinion that link layer retransmissions would take too much time and therefore trigger concurrent TCP retransmission [6]. Yet others [50, 53, 56, 82], confirmed the same results as ours that TCP gain from link layer retransmissions and that concurrent retransmissions occur infrequently. More recent examples of inefficient protocol interaction have been reported also for LTE in [65, 64].

2. *What impact does the buffer management of intermediates nodes have on TCP performance?*

   In Paper III, we studied TCP performance under different parameter settings of the Serving GPRS Support Node (SGSN) buffer in GPRS. Also for these experiments we had access to Telias test network. With the extensive buffers that were used in GPRS, the buffer delay and the RTT became too large for any traffic except for bulk transfer, for which a high throughput is more important than a short delay. In the measurements, bulk traffic was used, but traces of TCP traffic were examined on the segment level and the packet delays could be computed. For TCP, a smaller buffer resulted in shorter delays, but only slightly reduced throughput. TCP performance problems related to delay due to excessive queuing can be observed also in later systems, in 3G and 4G [7, 37, 40], and even in wired networks [30].

3. *How does buffer management interact with preemption and how does this affect TCP?*

   In GPRS full preemption is applied. This means that if there are incoming GSM calls and no resources are available, then GPRS flows may lose all their resources to GSM. In Paper III, we present experiments with preemption in the GPRS testbed. A TCP flow over GPRS was preempted by GSM calls. After the GSM calls were completed and resources became available, the GPRS connection was set up again. Most of the packets buffered during the preemption period were still available, with exception from the packets buffered closer to the air interface, in the Base Station Controller (BSC). The BSC buffer was flushed immediately when the GPRS connection lost its resources due to preemption. Then, after the preemption, a new packet from the TCP sender was required to restart the transmission again and the packets buffered in the SGSN were transmitted to the GPRS user. We propose that the SGSN should start to transmit packets after preemption, i.e., that the GPRS connection should be maintained during the preemption. Furthermore, we propose that the BSC buffer should be flushed first after a short timeout interval. This may not be possible if all the buffer resources are required for the high priority traffic. Our suggestions to improve the preemption handling in GPRS are in line with later proposals to improve the preemption handling in LTE and LTE-Advanced, such as the scalable preemption algorithm proposed in [17] and the fairness-based preemption algorithm presented in [46].

4. *How can handover be improved to avoid packet loss, packet reordering and duplicated packets?*

   In systems that aim to fulfill the requirements of IMT-Advanced, such as WINNER and beyond 3G systems, high data rates are reached partly because of the dense deployment of BSs. With small cells, handover occurs frequently, and, therefore, efficient mobility management is important. In order to reduce delays, radio related functionality is moved closer to the radio interface as compared to previous generations of systems, e.g., the RLC protocol is terminated in the BS instead of in a node in the backbone. Data loss can be avoided if buffered data and other status information is transferred from the target BS to the source BS during handover. In Paper IV, we proposed a scheme for RLC SDU context transfer, and an optional RLC PDU context transfer, of user data for the WINNER system. With RLC PDU context transfer, delay may be reduced in some cases. Due to RLC segmentation, large RLC SDUs may be divided into many smaller RLC PDUs. Assume that only one of the RLC PDUs belonging to a RLC SDU is lost during handover. If only RLC SDU context transfer is performed, then the whole RLC SDU need to be retransmitted after handover, instead of only the lost RLC PDU. Even with a one-to-one mapping between RLC SDU and RLC PDUs delay may be increased, if RLC PDUs are successfully received before handover, but the status message is lost. As packet reordering and duplicates have a negative impact on TCP, we also proposed that the WINNER IP Convergence Layer (IPCL) protocol, which, in contrast to the corresponding PDCP protocol in LTE, is terminated in the backbone, is responsible for in-sequence delivery and duplicate removal. At that time, in 2007/2008, when we wrote the paper, there was still a discussion whether RLC PDU context transfer should be supported in LTE or not [59]. The result was that context transfer of PDCP SDUs is specified along with actions to promote in-sequence delivery and duplicate removal [2].

5. *Which Internet application types would gain more from CoMP?*

   How the users experience a service depends both on the QoS offered by the network and the application type. In Paper V, we study which application types that would gain more from CoMP. User experience represented by application utility is used to indicate the level of satisfaction with a service. The result of this study is then used in the following papers as input to select users for limited CoMP.

6. *How can user experience be taken into account when users are selected for limited CoMP?*

   In a limited CoMP system, the overhead due to feedback in the uplink and information exchange over the backhaul links could be reduced if only a subset of the users were served with CoMP. In Paper VI, we examine user selection based on user experience represented by application utility. We propose and evaluate a heuristic scheme, utility-based user selection, to select users based on application utility. Comparison with the results of an exhaustive search for the maximum utility indicate that the proposed scheme performs close to optimal for the tested cases.

7. *How can CoMP resources be allocated to the users based on user experience?*

   In Paper VII, the utility-based user selection in the previous paper is improved and proposed as long term pre-selection of users. Furthermore, the pre-selection is complemented with short term joint time and frequency domain scheduling. The pre-selection process selects users based on utility gain of CoMP. The time domain scheduler operates on target bit rates that correspond to target utility values set by the pre-selection process. The frequency domain scheduler is similar to the one used in [9]. The proposed user selection improves the total utility in the system, as well as utility fairness as compared to schemes unaware of application utility.

5 Research methodology

The research methodologies used to understand and solve problems related to communication in cellular networks can be classified as experimental and analytical methodologies. If an experimental methodology is used to solve a problem, then experiments can be performed in a real network with real equipment or in a simulator used to represent the parts of the network that are important for the study. With an analytical approach, on the other hand, mathematical models are defined in order to understand various aspects of the problem under study and to find solutions, such as proofs and theoretical bounds.

To study problems and solutions that apply to an existing system, experiments could be conducted in a real network. The main advantage of conducting experiments in real networks is that real nodes and protocol implementations are used. However, the results may not be fully applicable to networks built with equipment from other vendors and different protocol implementations. It may also be difficult and time consuming both to measure and to analyze the results, since factors that are outside the scope of an experiment may have an effect on the results.

To be able to test ideas that would require completely new functionality, adjustments or replacements of existing protocols and algorithms, and maybe even new network nodes a simulation study is the option of choice. If time consumption of the experiments is important, then a simulation study may be preferred over measurements, even if a real system is available for the experiments. With simulations, it is often possible to do an exhaustive search to find the optimal solution to a problem and to compare this solution to the result of a heuristic algorithm. An analytical study may also be desirable, but for complex communication scenarios the problems involved are often NP hard, and therefore there are no efficient algorithms to find the optimal solution. In these cases, we must rely on heuristics to find solutions that could be applied in practice.

For the studies of TCP, we had access to a testbed consisting of a real GSM/GPRS network dedicated for test purposes. An emulated radio environment made it possible to control and repeat the measurements. To test the utility-based user selection for CoMP, we performed simulations. Most evaluations that involve CoMP are based on simulation. There are, however, a few research groups that experiment with CoMP in real networks [38], but this research is mainly concentrated to the physical layer.

6 Main contributions

The main contributions of the research presented in this thesis are:

1. *We apply a user perspective on Internet services in cellular networks.*

   We have a user centric approach on cellular networks, since we study TCP performance in the first four papers and application utility for CoMP in the last three papers. With a user-centric approach, the main objective of system performance is shifted from spectral efficiency and increasing the sum rate per cell towards user experience and satisfying more users. For applications that use TCP for transport, user experience depends on the performance of TCP. TCP performance is, in turn, affected by the interaction between the link layer protocol and TCP, which was studied in Paper I and in Paper II. For interactive traffic, TCP performance may be degraded due to queuing delay if excessive amounts of data is buffered, which was studied in Paper III. Also preemption by traffic with higher priority may lead to long delays, as described in Paper III. In Paper IV, we studied how TCP would be affected by packet loss, duplication and reordering due to handover in the WINNER system. The simulation studies of CoMP in Paper V, Paper VI and Paper VII focused on application utility as a basis for selecting which users to serve with CoMP.

2. *We conducted measurements on TCP over GSM with RLP in acknowledged mode and concluded that TCP performance increases with a reliable link layer protocol.*

   We showed in Paper I and in Paper II that the interaction between TCP and RLP in GSM is efficient in most cases, but that TCP performance is degraded in some cases when there are sudden increases in delay due to RLP retransmissions. In contrast to many other studies on TCP and the link layer in cellular networks, we performed the experiments in a real GSM network with real implementations of TCP and RLP. At that time, when the measurements were conducted, it was not completely clear that TCP should be used with link level retransmissions. Now, there is a consensus on this, but even in LTE there still exist cases when the protocol interaction is less efficient [64, 65].

3. *We showed that TCP suffers from the excessive buffering in intermediate nodes in GPRS and that the delays experienced by TCP could be reduced with maintained throughput.*

   The purpose of the buffer before the wireless link is to store data during periods of poor radio conditions and during handover in order to avoid data loss. If the buffers are too large and filled with bulk data, then the delays may become unacceptable for interactive traffic. In Paper III, we studied TCP performance in GPRS and found that the default setting of the intermediate buffer resulted in excessive queuing. We showed that with smaller buffers, the delays are decreased, and, hence, TCP would work better for interactive applications. The cost of the shorter delays is only a small reduction in TCP throughput. The result is only one, but nevertheless important, piece to solve the problem with TCP efficiency in cellular networks. Delays due to excessive buffering is still a problem for TCP in cellular networks [7, 37, 40], even though this problem was pointed out already for GPRS.

4. *We examined the impact of preemption on TCP in GPRS and suggested a more flexible preemption scheme.*

   If traffic classes with different priorities should be supported in a cellular network, then preemption is required in situations when there are not enough resources to serve all users. In Paper III, we examined a TCP flow over GPRS that was preempted by GSM calls. Instead of full preemption which is used in GPRS and also in LTE, we proposed that the connection over GPRS should remain during the preemption period and that buffered data should be transmitted as soon as possible when resources become available again. We also proposed that the buffer closest to the radio interface should not be flushed immediately, but be guarded with a preemption timer. These ideas would result in a more flexible preemption scheme than full preemption. More recent examples of preemption schemes for LTE and LTE-Advanced are found in [17, 46]. They also propose that the connections are maintained and that full preemption is used only when there are no other alternatives.

5. *We proposed functionality for the WINNER system during handover to avoid negative impact on TCP.*

   Hard handover may result in packet loss, reordered packets, and duplicated packets all of which reduce TCP performance. With systems beyond 3G and the standardization of LTE, functionality for handover moved closer to the air interface, and the RLC protocol was terminated in the BS instead of in the backbone, as in previous systems. This implies that RLC ARQ can not be used to recover data lost due to handover. Therefore, forwarding of data between BSs was proposed to avoid data loss. In Paper IV, we proposed a RLC scheme for context transfer of RLC state and data for the WINNER system in order to reduce the loss rate and improve TCP performance. Context transfer of RLC SDUs was proposed as a base line service and context transfer of RLC PDUs as a value added function. Which context transfer that is used depends on the user profile and on the policy applied for the service class. Handover may not only lead to data loss, but also to duplicated data and data that arrive in another order than the transmission order. Both duplicate detection and in-sequence delivery are important to achieve a high TCP performance. Therefore, we also proposed that the IPCL protocol in the WINNER system, which terminates in the backbone, reduces the number of duplicated and reordered packets. To increase the probability of in-sequence delivery we also suggested a separate service class with high priority for data that is forwarded during handover.

6. *We showed that user experience of CoMP is highly dependent on the application type.*

   In Paper V, we examined the impact of CoMP schemes on application utility through simulations. The results indicate that adaptive applications gain most from CoMP. An increase in available data rate has a large effect on utility for adaptive applications. Elastic applications, on the other hand, require a relatively low data rate to reach a high utility, a data rate that is likely to be reached with conventional single cell transmission. Then, for even higher data rates utility is only marginally increased, which implies that CoMP is not required for elastic applications in most cases. Hard real-time applications gain utility-wise from CoMP only if the constant data rate required by the application is reached. If a priority scheme would be applied for CoMP, then adaptive applications, and, in some cases, hard real-time applications should be prioritized.

7. *We studied application utility as criterion for user selection for CoMP and proposed a heuristic scheme for user selection.*

   As user satisfaction depends on the application, we proposed in Paper VI that user selection for CoMP takes application utility into account. A heuristic algorithm, utility-based user selection, was developed and evaluated through simulation. The sum utility reached with the utility-based user selection was compared to the maximum sum utility that was found through an exhaustive search over all possible CoMP groups. The results achieved with the utility-based user selection are almost as high as the optimal result for the simulated cases.

8. *We developed a utility-based scheme to pre-select and schedule users for CoMP.*

   In Paper VII, utility-based user selection for CoMP including user selection on two time scales is proposed and evaluated through simulation. Users are pre-selected for CoMP based on application utility on a relatively long time scale. The time domain scheduler uses target bit rate values corresponding to target utility values to schedule the users in the time slots. The scheduling of users in frequency resources is performed as in conventional single BS systems. The proposed heuristic approach limits the number of users to serve with CoMP while taking user experience and spectral efficiency into account. The results indicate that the level of user satisfaction is higher than for schemes that do not consider user experience. Also, fairness is improved with the proposed utility-based user selection approach.

7 Summary of papers

Paper I – TCP and GSM Link Layer Interactions: Implications for the Wireless Internet

This paper presents measurements of TCP over GSM conducted in a GSM testbed. The focus is on the interaction between TCP and the link layer protocol used for data transmission in GSM. The results show that the protocols interact efficiently in most cases also when the radio quality is poor. Some inefficient interaction occurs due to sudden variations in the delay. In these cases the protocols perform concurrent error recovery.

Paper II – A GSM/GPRS Testbed for TCP Performance Evaluation

Measurements of TCP over GSM are included also in this paper, but the measurements were conducted with another network configuration. The interaction between TCP and the link layer protocol is described in detail. The analysis of the protocol interaction indicates that TCP retransmits data unnecessarily when the number of retransmitted link layer frames suddenly increases. In the paper we also present the modifications required to extend the GSM testbed with GPRS.

Paper III – The Impact of GPRS Buffering on TCP Performance

In this paper, two aspects of buffering are considered for measurements conducted in the GPRS testbed. The first is how the setting of the buffers between GPRS and the Internet affects TCP performance. The results indicate that if the buffers are reduced, then the delay is shortened by orders of magnitude without any significant degradation in throughput. The second aspect concerns the interaction between buffering and preemption. Measurements of a GPRS data transfer preempted by circuit-switched calls with higher priority indicate that TCP performance is degraded more than necessary, since buffered data is flushed immediately when the GPRS traffic is preempted. TCP performance would improve if data was buffered during preemption and if the buffered data was transmitted immediately after preemption.

Paper IV – Network Controlled Mobility Management with Policy Enforcement towards IMT-A

A framework for mobility management and call handling based on per user service profiles are presented in the paper. In the WINNER system and other beyond 3G systems, radio related functionality has moved closer to the radio interface in order to reduce delay, e.g., RLC connections terminate in the BS. Furthermore, handover occur more often than in previous systems, since smaller cells are required in order to provide the high data rates specified for IMT-Advanced systems. The per user policy is enforced by the network through call setup, handover and other call handling functions, such as flow control. In the paper, some of the mechanisms required to realize policy based mobility management are proposed and described in more detail: RLC context transfer in the user plane, the responsibilities of the IP convergence layer (IPCL) during handover, and policy controlled handover mixture. Policy based flow control is also included, since flow control is efficient to avoid data loss, especially during handover. By avoiding data loss, duplicated packets and reordered packets, TCP performance is improved, and, hence also user experience.

Paper V – Utility of Joint Processing Schemes

In this first paper on user experience and CoMP, CoMP schemes are evaluated through simulation with respect to user experience represented by utility functions of Internet applications. The traffic types considered have elastic, adaptive and hard real-time utility functions. Three CoMP schemes with decreasing overhead are simulated, centralized, partial and decentralized joint processing. The aim is twofold: first, to determine which traffic types that benefit more from CoMP, and, second, to evaluate which CoMP schemes that are more efficient to the user. The results show that what users gain from CoMP is highly dependent on the traffic type. The results show that applications with hard real-time and adaptive utility functions benefit more from CoMP than applications with elastic traffic types. The main impact on the application utility of the transmission scheme is between CoMP and conventional single BS transmission, rather than between the CoMP schemes. Only if the application has high requirements the costly centralized CoMP scheme is required to satisfy the user.

Paper VI – Utility as a User Selection Criterion for Coordinated Multi-Point Systems

User experience of Internet applications, represented by utility functions, is proposed to be taken into account when users are selected for CoMP in systems with limited backhaul capacity. In order to determine the optimal user selection that maximizes the total utility in the system, all possible user selections need to be tested. We propose a heuristic algorithm, utility-based user selection, that has a low complexity. In each step, the utility-based user selection assumes that there is only one CoMP user in the system, even though more users will be selected and served with CoMP eventually. Simulation results indicate that the total utility is higher if utility is considered when users are selected than if users are instead selected with the objective to maximize the total rate. Furthermore, the total utility reached with the utility-based user selection algorithm is close to the optimal total utility.

Paper VII – User-centric Pre-selection and Scheduling for Coordinated Multipoint Systems

The idea of a utility-based user selection for limited CoMP, proposed in the previous paper, is improved and extended in this paper. The proposed approach consists of a pre-selection process and a joint time and frequency domain scheduler. Users are selected for CoMP on two levels that operate on different time scales. Users are first pre-selected for CoMP transmission by the pre-selection process that operates on a long time scale, in the range of hundreds of milliseconds. Then, the joint time and frequency domain scheduler allocates resources to the users on a millisecond basis. By selecting a subset of the users for CoMP the overhead due to feedback of channel state information (CSI) in the uplink is reduced as compared to if all users would be candidates for CoMP in each time slot. Less detailed CSI is required from the users served with conventional single BS transmission. The pre-selection process selects the users that would gain more utility-wise from CoMP. For this, the average utility gain due to CoMP is estimated. Fairness is also considered, since resources are redistributed from the users for which very high utility values could be achieved to the users that require more resources to reach an acceptable user experience. Simulation results indicate that the proposed utility-based user selection approach results in a higher total utility than a similar user selection based on data rate instead. Furthermore, utility fairness is highly improved.

8 Conclusions and future research

In the first generations of cellular networks, telephony was the dominating application. Today, the larger part of the traffic volume is generated by Internet applications. The evolution of cellular networks and the introduction of devices such as laptops, smart phones and tablets, have made it easier to access the Internet. Also, today the Internet is accessible in more places due to the wider deployment of cellular networks and the interworking with wireless local area networks. Even though the data rates offered in cellular networks now are almost in the same range as the data rates in wired networks, the data rates and other QoS parameters, such as delay and jitter, are more variable. The variability may have a negative impact on user satisfaction, partly because the Internet protocols were designed for wired networks with less variable QoS.

In this thesis, we have investigated TCP in cellular networks. TCP provides a reliable transport service to Internet applications by performing retransmissions of erroneous data and by delivering data in-sequence. In cellular networks, the link layer protocol also performs retransmissions to recover from errors, and, therefore, the interaction between TCP and the link layer protocol needs to be efficient. We studied the interaction between TCP and RLP in GSM and concluded that the protocols interacted efficiently in most cases. When TCP is used for delay sensitive traffic, then the delays may become unacceptable due to queuing, if buffers are shared with bulk traffic. We investigated TCP in GPRS with different settings of the intermediate buffer. Our results indicate that the delay sensitive traffic would gain from smaller buffers, while the TCP throughput of the bulk traffic almost remains unaffected. Problems due to bufferbloat have also been reported in later generations of cellular networks and even in the wired Internet. In cellular networks, preemptive priority is applied, in which traffic with higher priority is allowed to preempt resources from traffic with lower priority. In GPRS, data traffic is completely preempted by GSM calls. We studied preemption in GPRS and proposed improvements that would enhance TCP performance. Handover may also have a negative impact on TCP, especially if data is lost during handover. In order to avoid data loss during handover, context transfer of data and protocol state can be used. We proposed a context transfer scheme for the WINNER system and approaches to avoid out-of-order data and duplicates.

An increasing part of the Internet traffic is generated by data hungry applications. The main limitation to higher data rates in cellular networks is inter-cell interference. CoMP has been proposed to combat inter-cell interference, and, is now being standardized for LTE-Advanced [4]. In this thesis, we have studied user selection for CoMP based on QoE represented by utility functions of Internet application types. We proposed a two-level user selection with a long term pre-selection of users and a short term joint time and frequency domain scheduler. By focusing on user satisfaction, resources could be redistributed from users with low requirements to the users with high demands.

Our research on TCP gives an indication about functions and settings in cellular networks that would improve TCP performance and enhance user experience. Other examples include a gradual decrease of the available resources for TCP connections as soon as more users enters a cell [61], and a scheduling scheme that gives priority to ACKs [74]. Another approach to enhance user experience in cellular networks would be to adjust the functionality of the transport layer instead. Most Internet applications use TCP or UDP for transport, but depending on the requirements of the application and the services provided by the cellular network, other transport protocols may be a better choice. A more flexible transport service could be provided to applications, as proposed by the the IETF Transport Services (taps) working group [26, 84].

CoMP is a powerful tool to increase the data rates, but it is not yet completely clear how to use CoMP efficiently. We proposed a user centric allocation of CoMP resources. Even though cellular networks have improved for each generation and the data rates have increased multi-fold since the first cellular networks, the wireless resources are limited. Therefore, a user centric approach may be useful also to other problems related to cellular networks as a means to redistribute resources between users with various requirements. Examples of this are the fairness-based preemption scheme proposed in [46] and the QoS-aware scheduling strategies presented in [16].

References

[1] NMT 30 years. `http://nmt30years.com/`.

[2] 3GPP. TS 36.300 V8.3.0 (2007-12), 3rd Generation Partnership Project; Technical Specification Group; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E- UTRAN); Overall description; Stage 2, Release 8, December 2007.

[3] 3GPP. TR 36.819 v11.2.0, 3rd Generation Partnership Project; Technical specification group radio access network; Coordinated multi-point operation for LTE physical layer aspects, (Release 11);. September 2013.

[4] 3GPP. TR 36.874 V12.0.0 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Coordinated multi-

point operation for LTE with non-ideal backhaul (Release 12). December 2013.

[5] 3GPP. TS 36.212 V12.2.0 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding (Release 12). September 2014.

[6] T. Alanko, M. Kojo, H. Laamanen, M. Liljeberg, M. Moilanen, and K. Raatikainen. Measured performance of data transmission over cellular telephone networks. *ACM Computer Communication Review*, 24(5), 1994.

[7] S. Alfredsson, G. Del Giudice, J. Garcia, A. Brunstrom, L. De Cicco, and S. Mascolo. Impact of TCP congestion control on bufferbloat in cellular networks. *Proceedings of the 14th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2013), Madrid, Spain*, pages 1–7, June 2013.

[8] M. Allman, V. Paxson, and E. Blanton. TCP Congestion Control. RFC 5681 (Draft Standard), Sept. 2009.

[9] R. Apelfröjd and M. Sternad. Measurement based evaluations of coherent JT CoMP - a study of precoding, user grouping and resource allocation with predicted CSI. *EURASIP Wireless Communications and Networking*, June 2014.

[10] D. Astély, E. Dahlman, A. Furuskär, Y. Jading, M. Lindström, and S. Parkvall. LTE: The evolution of mobile broadband. *IEEE Communications Magazine*, 47(4):44 – 51, April 2009.

[11] L. Bajzik, P. Horvath, L. Korossy, and C. Vulkan. Impact of intra-LTE handover with forwarding on the user connections. *16th IST Mobile and Wireless Communications Summit*, July 2007.

[12] Balamurali. A low complexity resource scheduler for cooperative cellular networks. *Proceedings of the 3rd IEEE international conference on Internet multimedia services architecture and applications (IMSAA'09)*, pages 182–187, 2009.

[13] C. Bettstetter, H.-J. Vögel, and J. Eberspächer. GSM phase 2+ general packet radio service GPRS: Architecture, protocols, and air interface. *IEEE Communication Surveys*, 2(3), 1999.

[14] T. Bonald. A score-based opportunistic scheduler for fading radio channels. *European Wireless, Barcelona, Spain*, February 2004.

[15] C. Botella, T. Svensson, X. Xu, and H. Zhang. On the performance of joint processing schemes over the cluster area. *IEEE Vehicular Technology Conference Spring 2010*, May 2010.

[16] F. Capozzi, G. Piro, L. Grieco, G. Boggia, and P. Camarda. Downlink packet scheduling in LTE cellular networks: Key design issues and a survey. *IEEE Communications Surveys & Tutorials*, (99):1–23, 2012.

[17] S. M. Chadchan and C. B. Akki. Priority-scaled preemption of radio resources for 3GPP LTE networks. *International Journal of Computer Theory and Engineering*, 3(5), December 2011.

[18] R. Chakravorty, J. Cartwright, and I. Pratt. Practical experience with TCP over GPRS. In *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '02)*, volume 2, pages 1678–1682, Taipei, Taiwan, November 2002.

[19] W.-P. Chen, Y.-C. Hsiao, J. C. Hou, Y. Ge, and M. P. Fitz. Syndrome: a light-weight approach to improving TCP performance in mobile wireless networks. *The Journal of Wireless Communications and Mobile Computing*, pages 37–57, 2002.

[20] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle. Layering as optimization decomposition: A mathematical theory of network architectures. *Proceedings of the IEEE*, 95(1):255–312, January 2007.

[21] A. Chowdhery, W. Yu, and J. M. Cioffi. Cooperative wireless multicell OFDMA network with backhaul capacity constraints. *Proceedings of the 2011 IEEE International Conference on Communications (ICC), Kyoto, Japan*, pages 1–6, June 2011.

[22] E. Dahlman, P. Beming, J. Knutsson, F. Ovesjö, M. Persson, and C. Roobol. WCDMA - The radio interface for future mobile multimedia communications. *IEEE Transactions on Vehicular Technology*, 47(4):1105 – 1118, November 1998.

[23] T. Doumi, M. Dolan, S. Tatesh, A. Casati, G. Tsirtsis, K. Anchan, and D. Flore. LTE for public safety networks. *IEEE Communications Magazine*, 51(2):106–112, February 2013.

[24] ETSI. Digital cellular telecommunications system (Phase 2+); Radio Link Protocol (RLP) for data and telematic services on the Mobile Station - Base Station System (MS - BSS) interface and the Base Station System - Mobile-services Switching Centre (BSS - MSC) interface (GSM 04.22 version 7.0.1 Release 1998). *ETSI*, 1999.

[25] ETSI. Digital cellular telecommunications system (Phase 2+); General Packet Radio Service (GPRS); Service description; Stage 2 (GSM 03.60 version 6.7.0 Release 1997), 2000.

[26] G. Fairhurst, B. Trammell, and M. Kuehlewind. Services provided by IETF transport protocols and congestion control mechanisms. Working Draft, February 2015. `http://www.ietf.org/internet-drafts/draft-ietf-taps-transports-03.txt`.

[27] S. Floyd and V. Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1:397–413, 1993.

[28] S. Fu, B. Wu, H. Wen, P.-H. Ho, and G. Feng. Transmission scheduling and game theoretical power allocation for interference coordination in CoMP. *IEEE Transactions on Wireless Communications*, 13(1):112–123, 2014.

[29] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu. Multi-cell MIMO cooperative networks: A new look at interference. *IEEE Journal on Selected Areas in Communications*, 28(9):1380 – 1408, December 2010.

[30] J. Gettys and K. M. Nichols. Bufferbloat: dark buffers in the Internet. *Communications of the ACM*, 55(1):57–65, 2012.

[31] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas. LTE-advanced: next-generation wireless broadband technology. *IEEE Communications Magazine*, 17(3):10 – 22, June 2010.

[32] A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

[33] A. Gurtov. Effect of delays on TCP performance. In *Proceedings of IFIP Personal Wireless Communications*, Lappeenranta, Finland, August 2001.

[34] A. Gurtov and R. Ludwig. Evaluating the Eifel algorithm for TCP in a GPRS network. In *Proceedings of European Wireless*, Florence, Italy, February 2002.

[35] A. Gurtov, M. Passoja, O. Aalto, and M. Raitola. Multi-layer protocol tracing in a GPRS network. In *Proceedings of IEEE Vehicular Technology Conference (VTC-02 Fall)*, volume 3, pages 1612–1616, Vancouver, Canada, September 2002.

[36] P. Hande, S. Zhang, and M. Chiang. Distributed rate allocation for inelastic flows. *IEEE/ACM Transactions on Networking*, 15(6):1240–1253, December 2007.

[37] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z. M. Mao, S. Sen, and O. Spatscheck. An in-depth study of LTE: Effect of network protocol and application behavior on performance. *Proceedings of the ACM SIGCOMM 2013 Conference, Hong Kong, China*, pages 363–374, August 2013.

[38] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel. Coordinated multipoint: Concepts, performance, and field trial results. *IEEE Communications Magazine*, 49(2):102–111, February 2011.

[39] V. Jacobson and M. J. Karels. Congestion avoidance and control. *ACM Computer Communications Review*, 18(4):314–329, August 1988.

[40] H. Jiang, Y. Wang, K. Lee, and I. Rhee. Tackling bufferbloat in 3G/4G mobile networks. *Proceedings of the 2012 ACM Conference on Internet Measurement Conference, Boston, Massachusetts, USA*, pages 329–342, November 2012.

[41] J. Jin, A. Sridharan, B. Krishnamachari, and M. Palaniswami. Handling inelastic traffic in wireless sensor networks. *IEEE Journal on Selected Areas in Communications*, 28(7):1105 – 1115, September 2010.

[42] H. Kaaranen, A. Ahtiainen, L. Laitinen, S. Naghian, and V. Niemi. UMTS Networks: Architecture, Mobility and Services. *John Wiley & Sons Ltd, ISBN: 978-0-470-01103-4*, February 2005.

[43] S. Karlsson. The pioneers - a story about modern mobile telephony and the people who developed it. `http://www.teliasonerahistory.com/sources/svenolof-karlssons-the-pioneers/`, 2011.

[44] F. P. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8(1):33–37, JanuaryâĂŞFebruary 1997.

[45] F. P. Kelly, A. K. Maulloo, and D. K. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research society*, pages 237–252, 1998.

[46] M. Khabazian, O. Kubbart, and H. Hassanein. A fairness-based preemption algorithm for LTE-Advanced. *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM 2012)*, 2012.

[47] S. Kumar and N. Marchetti. IMT-Advanced: Technological requirements and solution components. *Second International Workshop on Cognitive Radio and Advanced Spectrum Management (CogART 2009), Aalborg, Denmark*, pages 1 – 5, May 2009.

[48] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana. Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges. *IEEE Communications Magazine*, 50(2):148–155, February 2012.

[49] W. Lee, I. Lee, J. S. Kwak, B.-C. Ihm, and S. Han. Multi-BS MIMO cooperation: challenges and practical solutions in 4G systems. *IEEE Wireless Communications*, 19(1):89–96, February 2012.

[50] R. Ludwig. Eliminating inefficient cross-layer interactions in wireless networking. *Ph.D. Thesis, Aachen University of Technology, Germany*, April 2000.

[51] R. Ludwig and R. H. Katz. The Eifel algorithm: Making TCP robust against spurious retransmissions. *ACM Computer Communication Review*, 30(1):30–37, Jan. 2000.

[52] R. Ludwig, A. Konrad, and A. D. Joseph. Optimizing the end-to-end performance of reliable flows over wireless links. *Fifth Annual International Conference on Mobile Computing and Networks (MobiCom '99), Seattle*, August 1999.

[53] R. Ludwig, B. Rathonyi, A. Konrad, K. Oden, and A. Joseph. Multi-layer tracing of TCP over a reliable wireless link. *ACM SIGMETRICS*, 1999.

[54] P. Marsch and G. Fettweis. On multicell cooperative transmission in backhaul-constrained cellular systems. *Annals of Telecommunication*, 63(5–6):253–269, June 2008.

[55] S. Mehryar, A. Chowdhery, and W. Yu. Dynamic cooperation link selection for network MIMO systems with limited backhaul capacity. *IEEE International Conference on Communications, ICC, Ottawa, Canada*, June 2012.

[56] M. Meyer. TCP performance over GPRS. In *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '99)*, volume 3, pages 1248–1252, New Orleans, Louisiana, USA, September 1999.

[57] W. Mohr. The WINNER (wireless world initiative new radio) project-development of a radio interface for systems beyond 3G. *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2005), Berlin, Germany*, 2:857 – 862, September 2005.

[58] G. Monghal, K. I. Pedersen, I. Z. Kovács, and P. E. Mogensen. QoS oriented time and frequency domain packet schedulers for the UTRAN long term evolution. *2008 IEEE 67th Vehicular Technology Conference-Spring (VTC Spring), Marina Bay, Singapore*, pages 2532–2536, May 2008.

[59] Motorola. Long Term Evolution (LTE): A technical overview. *Technical White Paper*, 2007.

[60] M. Mouly and M.-B. Pautet. *The GSM System for Mobile Communications*. Cell & Sys, 1992.

[61] B. Nguyen, A. Banerjee, V. Gopalakrishnan, S. Kasera, S. Lee, A. Shaikh, and J. Van der Merwe. Towards understanding TCP performance on LTE/EPC mobile networks. *Proceedings of the 4th Workshop on All Things Cellular: Operations, Applications, & Challenges, Chicago, Illinois, USA*, pages 41–46, 2014.

[62] T. Ojanperä and R. Prasad. An overview of air interface multiple access for IMT-2000/UMTS. *IEEE Communications Magazine*, 36(9):82 – 86, 91 – 95, September 1998.

[63] A. Osseiran, E. Hardouin, A. Gouraud, M. Boldi, I. Cosovic, K. Gosse, J. Luo, S. Redana, W. Mohr, J. Monserrat, T. Svensson, A. Tolli, A. Mihovska, and M. Werner. The road to IMT-advanced communication systems: State-of-the- art and innovation areas addressed by the WINNER+ project. *IEEE Communications Magazine*, 47(6):38 – 47, June 2009.

[64] H.-S. Park, J.-Y. Lee, and B.-C. Kim. TCP performance degradation of in-sequence delivery in LTE link layer. *International Journal of Advanced Science and Technology*, 37:27–36, December 2011.

[65] H.-S. Park, J.-Y. Lee, and B.-C. Kim. TCP performance issues in LTE networks. *Proceedings of 2011 International Conference on ICT Convergence (ICTC)*, pages 493 – 496, September 2011.

[66] S. Parkvall, A. Furuskär, and E. Dahlman. Evolution of LTE toward IMT-Advanced. *IEEE Communications Magazine*, 49(2):84 – 91, February 2011.

[67] J. Postel. User datagram protocol. RFC 768 (Standard), Aug. 1980.

[68] J. Postel. Internet protocol. RFC 791 (Standard), Sept. 1981.

[69] J. Postel. Transmission control protocol. RFC 793 (Standard), Sept. 1981. Updated by RFCs 1122, 3168, 6093.

[70] A. Racz, A. Temesvary, and N. Reider. Handover performance in 3GPP long term evolution (LTE) systems. *16th IST Mobile and Wireless Communications Summit*, July 2007.

[71] J. Sachs, B. Khurana, and P. Mahonen. Evaluation of handover performance for TCP traffic based on generic link layer context transfer. *IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, September 2006.

[72] M. Sågfors, R. Ludwig, M. Meyer, and J. Peisa. Buffer management for rate-varying 3G wireless links supporting TCP traffic. In *Proceedings of IEEE Vehicular Technology Conference (VTC-03 Spring)*, volume 1, pages 675–679, Jeju, South Korea, April 2003.

[73] M. Sågfors, R. Ludwig, M. Meyer, and J. Peisaa. Queue management for TCP traffic over 3G links. *Wireless Communications and Networking, WCNC 2003*, 3:1663 – 1668, March 2003.

[74] P. Shang, Y. Zeng, J. Wu, and P. Xiao. A novel resource scheduling algorithm to improve TCP performance for 3GPP LTE systems. *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, China*, pages 1898 – 1902, April 2013.

[75] S. Shenker. Fundamental design issues for the future Internet. *IEEE Journal of Selected Areas in Communication*, 13(7):1141–1149, September 1995.

[76] H. Shi, R. V. Prasad, E. Onur, and I. G. Niemegeers. Fairness in wireless networks: Issues, measures and challenges. *IEEE Communications Surveys and Tutorials*, 16(1):5–24, 2014.

[77] L. Shi, C. Liu, and B. Liu. Network utility maximization for triple-play services. *Computer Communications*, 31(10):2257–2269, June 2008.

[78] S. Sigdel and W. A. Krzymién. User scheduling for network MIMO systems with successive zero-forcing precoding. *Proceedings of the IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall), Ottawa, Canada*, pages 1–6, September 2010.

[79] S. Sigdel, W. A. Krzymién, and M. Al-Shalash. Greedy and progressive user scheduling for CoMP wireless networks. *Proceedings of the 2012 IEEE International Conference on Communications (ICC), Ottawa, Canada*, pages 4218–4223, June 2012.

[80] H. Skjevling, D. Gesbert, and A. Hjorungnes. Low-complexity distributed multibase transmission and scheduling. *EURASIP Journal on Advances in Signal Processing*, 2008.

[81] S. Srikanth, P. A. M. Pandian, and X. Fernando. Orthogonal frequency division multiple access in wimax and lte: a comparison. *IEEE Communications Magazine*, 50(9):153–161, 2012.

[82] M. Taferner and E. Bonek. *Wireless Internet Access over GSM and UMTS*. Springer-Verlag, 2002.

[83] A. Thampi, S. Armour, Z. Fan, and D. Kaleshi. Clustered network MIMO and fractional frequency reuse for the downlink in LTE-A systems. *Proceedings of European Wireless 2014, Barcelona, Spain*, May 2014.

[84] B. Trammell and J. Hildebrand. Evolving transport in the Internet. *IEEE Internet Computing*, 18(5):60–64, 2014.

[85] P. L. Vo, N. H. Tran, and C. S. Hong. Joint rate and power control for elastic and inelastic traffic in multihop wireless networks. *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM 2011), Houston, TX, USA*, pages 1 – 5, December 2011.

[86] W.-H. Wang, M. Palaniswami, and S. H. Low. Application-oriented flow control: Fundamentals, algorithms and fairness. *IEEE/ACM Transactions on Networking*, 14(6):1282 – 1291, December 2006.

Improving User Experience of Internet Services in Cellular Networks

The Internet has grown enormously since the introduction of the World Wide Web in the early 90's. The evolution and wide spread deployment of cellular networks have contributed to make the Internet accessible to more people in more places. The cellular networks of today offer data rates high enough for most Internet services. Even so, the service quality experienced by the users is often lower than in wired networks.

The performance of TCP has a large impact on user experience. Therefore, we investigate TCP in cellular networks and propose functionality to improve the situation for TCP. We have studied sources of delay and data loss, such as link layer retransmissions, queuing, and handover.

The achievable data rates in cellular networks are limited by inter-cell interference that vary over the cell area. Inter-cell interference can be mitigated with Coordinated Multipoint techniques (CoMP), techniques that currently are being standardized for LTE-Advanced. System wide CoMP is, however, not an option, since it would be too resource consuming. In order to limit the required resources for CoMP, we propose an approach to select a subset of the users for CoMP that is based on user experience.

ISBN 978-91-7063-631-8

ISSN 1403-8099

DISSERTATION | Karlstad University Studies | 2015:18