Mobility and opportunistic resource allocation in wireless multimedia networks

VLADIMIR VUKADINOVIĆ

Doctoral Thesis
Stockholm, Sweden, 2010
Abstract

In order to support increasing traffic loads, mobile operators need cost-effective solutions to improve the spectral efficiency of their cellular networks, or to off-load them by diverting some of the load to other networks. Advances in the radio resource management may to some extent reduce the need for costly new deployments. The resource management should not only focus on spectrum efficiency—it should try to meet the service requirements of applications that are expected to contribute large data volumes, such as video streaming. Many of those applications are multicast/broadcast in nature (e.g., mobile TV, data podcasting). Our focus in this thesis is on resource allocation mechanisms that exploit the mobility of users. The mobility induces channel quality fluctuations and creates intermittent connectivity, which both can be used to improve the resource efficiency of wireless multimedia systems. The thesis concentrates on two areas: link-layer resource allocation for video streaming in cellular networks and mobility-assisted content distribution in hybrid cellular/ad-hoc networks.

In the area of wireless video streaming, we study bit-rate allocation, statistical multiplexing, and channel-aware scheduling. The bit-rate allocation should provide a distortion-optimal assignment of source, channel, and pilot bit-rates under link capacity constraints. We derive an analytical model that captures the video distortion as a function of the bit-rates and, based on it, we study various bit-rate allocation strategies and their robustness to varying radio conditions. The statistical multiplexing can be used to smooth out the burstiness of video streams and avoid over-provisioning of transport channels. We study the statistical multiplexing gains of H.264 video streams, both in terms of bit-rate requirements and video quality. When multiple flows are multiplexed on a shared transport channel, multi-user scheduling becomes crucial for the performance. Channel-aware scheduling exploits fluctuations in radio conditions to optimize the assignment of channel resources. We study the impact of channel-aware scheduling on the performance of delay-sensitive applications and possible extensions of channel-aware schemes to multicast scenarios.

In the area of mobility-assisted content distribution, we study the resource efficiency of mobility-assisted podcasting and we propose an analytical model for pedestrian content distribution. The mobility-assisted podcasting exploits random encounters of mobile terminals equipped with short range radios to forward the podcast episodes, thereby relieving the strain on cellular networks. We provide results on the achievable spectrum and energy savings of such scheme. Finally, we introduce the “street model”, the first building block in a conceived library of analytical models that would be used to study the performance of pedestrian content distribution in some common case scenarios of urban mobility. Based on the “street model”, we study how various system parameters and node mobility affect the efficiency of content distribution in a grid of streets that represents a part of Stockholm’s downtown area.
Acknowledgments

I would like to express my gratitude to Prof. Gunnar Karlsson for giving me an opportunity to become a member of his lab, for his continuous support and guidance during the course of my studies, and for inspiring my research with ideas. I would also like to thank my colleagues and professors at LCN: Alberto, Fetahi, György, Ilias, Ioannis, Ljubica, Ognjen, Ólafur, Rolf, Sylvia, and Viktoria for providing me with valuable comments during our seminars, for creating a friendly and enjoyable working atmosphere, and for witty discussions during lunch breaks and “fikas”. Special thanks to my office mates, Jing and Liping, for creating an optimal level of distraction.

Many of the results presented here are outcomes of fruitful collaborations. I would like to thank the people who have contributed to this thesis by co-authoring some of the included papers: György, Ólafur, and Jöerg. I forgive them for editing my perfectly good paragraphs. I am also grateful to the institutions that have funded my research: the Graduate School of Telecommunications (GST), the ACCESS Linnaeus center, Wireless@KTH center, the Affordable Wireless Services and Infrastructure (AWSI) project, and the Simple and Efficient Radio Access Networks (SERAN) project.

Finally, endless thanks to my parents Olga and Vido, and my sisters, Vesna and Ivanka, for their love and support.
# Table of Contents

1. Introduction .................................................................................................................. 1  
   1.1 Motivation ............................................................................................................... 1  
   1.2 Scope and outline of the thesis ............................................................................ 3  

2. Resource allocation for video streaming in cellular networks ......................... 5  
   2.1 Bit-rate allocation .................................................................................................. 5  
   2.2 Statistical multiplexing ......................................................................................... 8  
   2.3 Channel-aware scheduling .................................................................................... 9  

3. Mobility-assisted communication in delay-tolerant networks ...................... 15  
   3.1 Opportunistic content delivery ............................................................................ 15  
   3.2 Pedestrian mobility models ................................................................................ 17  

4. Summary of original work ....................................................................................... 21  
   4.1 Paper A: Trade-Offs in Bit-Rate Allocation for Wireless Video Streaming... 21  
   4.2 Paper B: Statistical multiplexing gains of H.264/AVC video in E-MBMS..... 21  
   4.3 Paper C: Video Streaming Performance under Proportional Fair Scheduling ................................................................. 22  
   4.4 Paper D: Multicast Scheduling with Resource Fairness Constraints ........... 22  
   4.5 Paper E: Multicast Scheduling for Scalable Video Streaming in Wireless Networks ................................................................................................. 23  
   4.6 Paper F: Spectral and Energy Efficiency of Mobility-Assisted Podcasting in Cellular Networks .................................................. 23  
   4.7 Paper G: An Analytical Model for Pedestrian Content Distribution in a Grid of Streets .................................................. 24  
   4.8 Other publications ............................................................................................. 24  

5. Conclusions and future work .................................................................................. 27  

Bibliography .................................................................................................................. 31
1. Introduction

1.1 Motivation

Since the launch of the first GSM networks, digital mobile radio systems, much like their analog predecessors, have been predominantly used to carry voice traffic. However, users’ expectations have evolved with the arrival of powerful handheld devices (“smartphones”) with functionalities that are turning mobile phones into multimedia entertainment devices. Also, in some countries, affordable mobile Internet subscriptions have already caused tremendous changes in usage patterns and service requirements for mobile radio systems. Recent reports on increasing data traffic volumes are the evidence of a big shift from traditional voice towards data services, such as web browsing, electronic mail, multimedia streaming, gaming, and mobile VoIP services. In May 2009, AT&T reported a 5,000% increase in data traffic across the United States in the past three-year period. Around 50% of this increase is generated by iPhone users despite the fact that they represent only a few million of AT&T’s 78 million cellular subscribers. In September 2009, Facebook announced that there are already 65 million active users accessing the social network website via mobile phones, an increase from 20 million just eight months earlier. In its recent report, market research firm Infonetics Research predicts that in 2010 data traffic will surpass voice traffic in mobile networks [1].

To support the increasing loads, mobile operators have to rely on limited spectral resources allocated by regulatory bodies. Most operators have allocations of 10 to 20 MHz or more, but typically have not been using all of their network capacity thus far. At the end of 2009, however, many operators issued warnings of a looming spectrum crisis. Newer cellular technologies, such as HSPA and LTE, do provide increased spectral efficiency, but the rate of data growth in the immediate future is likely to outpace any gains offered by the network technologies. Therefore, mobile operators will need extra spectrum to avert the crisis. Regulators all over the world started considering realignments of spectrum usage to help operators keep up with consumer demand. Most EU countries have just finished their auctions for the 2600 MHz frequency spectrum, which provided some operators with additional 40 to 50 MHz for the deployment of high-speed LTE networks. This is only a temporary relief; severe congestion problems will be difficult to avoid once broadband mobile network access and services become ubiquitous. At the end of 2009, the number of mobile broadband users reached 500 million globally, which is still only 10% of the global population of all mobile customers [2]. ITU predicts that the number of mobile broadband subscriptions will exceed one billion during 2010 and mobile web access via laptops and smartphones will exceed web access from desktop computers within the next five years [3]. A recent report by Ovum Datamonitor predicts that there will be two billion mobile broadband users by the end of 2014 [4]. These users will expect multimedia services and data rates similar to those enjoyed by fixed broadband users today, which will lead to a traffic explosion similar to the one that happened in the wired Internet during the last decade.

The key technical problem in supporting increased data volumes is the low spectral efficiency of mobile radio systems—the number of bits per second that can be delivered
per Hertz spectrum. This is due to the nature of the wireless medium: radio signals are attenuated and subjected to noise and interference as they propagate through the air. Current modulation and signal processing techniques are already so close to the Shannon limit that further advances in signal processing alone will not radically improve the spectral efficiency. The deployment of denser infrastructures would bring cell towers closer to the users to provide a substantial increase in spectral efficiency. This, however, may be too costly for operators who, after the decoupling of traffic volumes and revenues due to the flat-rate pricing, are reluctant to make the large investment that would not necessarily attract new subscribers. Therefore, as long as data traffic is growing out of pace with revenues, the denser deployment will be highly incremental and limited to venues frequented by high concentrations of mobile data users. Operators may turn to new pricing models to compensate for rising network costs, but the point where users are no longer willing to pay per use-time or data volume might have already been crossed. Therefore, operators need readily available and cost-effective solutions to improve the spectral efficiency of their cellular infrastructures or to off-load them by using alternative communication networks.

Advances in the radio resource management may to some extent reduce the need for the costly new deployments. The radio resource management involves algorithms for the allocation of frequency bands and transmission power, as well as channel access and link adaptation schemes. The allocation can be performed independently in each cell, or it can be a coordinated decision for several base stations. The ultimate objective is to maximize the utilization of the radio spectrum and network infrastructure under certain grade-of-service and fairness constraints. The adoption of the orthogonal frequency-division multiplexing (OFDM) in emerging systems offers a great deal of flexibility to perform fast resource allocation based on changing channel and traffic conditions. Such systems may exploit time, frequency, and space diversities of the radio channels to achieve significant increase in spectral efficiency. The resource allocation algorithms, however, should not only be designed to maximize the spectral efficiency, but also to meet service/user requirements: this design objective may span several layers and, therefore, it requires a cross-layered approach. For example, it is expected that some of the applications that will contribute large data volumes in mobile networks will be multicast/broadcast in nature (e.g., mobile TV, data podcasting). Therefore, resource allocation mechanisms should address the specific requirements of multicast/broadcast services and not focus on unicast only.

The increased traffic volumes in cellular systems may also be supported by diverting part of the load to alternative radio networks, since mobile terminals are nowadays often equipped with multiple radio interfaces (e.g., Wi-Fi, Bluetooth). For example, very often there is an overlap between the areas with high density of mobile broadband subscribers and areas that are served by public Wi-Fi networks and hot-spot providers. Integration between the cellular and Wi-Fi networks would provide a significant boost in capacity with only moderate investments from operators. The mobile subscribers may as well connect to each other directly, using Wi-Fi in ad-hoc mode or Bluetooth radios. In that case, cellular capacity can be used to seed the data into an area; the data would then spread among the subscribers through the direct peer-to-peer contacts. The objective of
such hybrid scenarios is to exploit the distinguishing properties of different radio access technologies, especially in terms of their spectral and energy efficiencies.

1.2 Scope and outline of the thesis

This thesis focuses mainly, but not exclusively, on resource allocation mechanisms that exploit the mobility of users to improve the resource efficiency of cellular and hybrid radio networks. When users move, their radio channels vary due to the shadowing and multipath fading. They also encounter other users to which they can connect directly if their terminals are equipped with short-range radios. Therefore, mobility creates opportunities for resource-efficient communication. For example, opportunistic packet scheduling exploits channel variations to assign resources (time slots and frequency carriers) to the users that currently enjoy good radio conditions. Another example is opportunistic content delivery, which exploits contacts between users that arise with the mobility in order to deliver the data in a store-carry-forward fashion and, hence, may off-load the cellular infrastructure. In both examples, the opportunistic schemes introduce certain delay: a user has to wait until channel conditions become favorable to be scheduled in the former example, or it has to wait until it encounters another user to deliver the data in the latter example. Hence, latency requirements of applications must be flexible to some extent in order to avail from such schemes. We consider applications with relatively stringent latency requirements, such as video streaming, but also those that may tolerate long delays, such as podcasting. Within the vast scope of possible resource-allocation problems and application scenarios, our research focuses on two particular topics: link-layer resource allocation mechanisms for video streaming in cellular networks and mobility-assisted content distribution in hybrid cellular/ad-hoc networks. Our work evaluates the performance of some well-known mechanisms, extends the mechanisms to new communication scenarios, and proposes new mechanisms or provides guidelines for designing them. Our performance evaluation method relies on detailed simulation scenarios in some cases, while in the others it abstracts away the details of the communication system that are technology-specific and focuses on performance limits.

The thesis is organized as follows: Section 2 provides an overview of the resource allocation mechanisms for wireless video streaming that are addressed in this thesis. Section 3 addresses mobility-assisted content distribution in delay-tolerant networks and human mobility models. A summary of the original work published in the context of this thesis is provided in Section 4. Conclusions and pointers to some open issues and possible future work are given in Section 5. Finally, the published papers are in the appendices of the thesis.
2. Resource allocation for video streaming in cellular networks

Many believe that video streaming (video on demand, mobile TV) will contribute the largest share of data traffic in future mobile radio systems. Current trends in the wired Internet support this assumption. Several recent reports indicate that video streaming is overtaking peer-to-peer file sharing in terms of broadband network usage. A research by Cisco shows that, although file sharing continues to grow in Sweden, it is becoming dwarfed by online streaming of TV shows and movies, which grows 130% annually. Video is streamed in a variety of bit rates, depending on the picture quality and resolution, and with diverse latency requirements, depending on the particular video application. Besides, streaming traffic produced by video encoders can be extremely bursty over a wide range of time scales. Considering this, and the fact that radio conditions vary over time, it is clear that resource allocation for video streaming in mobile networks has to be extremely flexible, both with respect to the granularity of allocated resource blocks and the time-scale on which the allocation is performed. To meet such demands, modern cellular radio technologies employ a number of advanced features (e.g. orthogonal frequency-division multiple access, adaptive modulation and coding schemes, hybrid-ARQ mechanisms). Some examples of resource allocation functions that require a high degree of dynamism are power control, link adaptation, dynamic channel allocation, and packet scheduling. Very often they are not subjected to standardization, so they become a matter of distinction among equipment manufacturers and network operators.

In this thesis, we address three resource allocation problems for wireless video streaming: bit-rate allocation (link adaptation), statistical multiplexing (channel allocation), and channel-aware packet scheduling. These are some of the link-layer mechanisms that are becoming even more important with the current trend of shifting the traffic to shared transport channels, which provide better spectral efficiency compared to dedicated channels. In addition to providing high spectral efficiency, the mechanism should be optimized to provide a degree of fairness among competing flows, and robustness to varying radio conditions. Given the absence of power control mechanisms in recent high-speed downlink technologies, packet scheduling is becoming the main mechanism of fast resource allocation and, therefore, it is given special attention in this thesis. The main challenges in designing resource allocation strategies for video streaming come from relatively stringent delay-requirements and from difficulties in relating the user experience to link-layer performance measures in a simple and tractable manner.

2.1 Bit-rate allocation

It is difficult to provide reliable transmission over wireless links due to channel noise, interference, and signal attenuation caused by fading and shadowing. One of the central topics in wireless video transmission is the development of error resilience mechanisms to increase the robustness of video streams to transmission errors.
Automatic repeat request (ARQ) error control techniques have been successfully applied in many cases. However, retransmission of corrupted frames introduces additional delay, which might be unacceptable for real-time applications. Forward error correction (FEC) is therefore often proposed for streaming applications, since the delay it introduces is bounded by the length of FEC blocks. FEC comes at the price of a reduced source data rate in capacity constrained channels. The level of protection, hence, the amount of redundancy that must be added into the stream, needs to be adapted to the current channel conditions. The conventional approach to this problem assumes independent design of the source and channel coders. This approach finds support in the famous separation theorem by Shannon [5], which states that, asymptotically with the length of the source data, source coding and channel coding can be designed separately without any loss of performance. The separation theory greatly reduces the complexity of a practical system design. It has led to major advances on source coding (e.g. MPEG-4 and H.264 video coding standards) and to the design of powerful channel codes (e.g. Reed-Solomon, Turbo, and LDPC codes). However, the separation theorem relies on some assumptions (infinite length codes, unbounded delay and complexity) that are not satisfied in real communication scenarios, e.g. in the delay-constrained real-time multimedia delivery. Systems designed based on the separation theorem are not robust to variations in channel quality: they tend to break down when the bit-error rate increases above the error correction capability of the channel code.

Joint source-channel coding (JSCC) has emerged as a promising concept and became a topic of numerous research efforts. JSCC is a rather loose label that encompasses all coding techniques where the source and channel coders are not entirely separated. An overview of various JSCC methods is provided in [6].

In the case of video streaming, most of the JSCC work to date has focused on distortion-optimal bit-rate allocation between source and channel codes. The distortion comes from compression artifacts in the encoder (source distortion) and from residual transmission errors after channel decoding (channel distortion). The available bit-rate on a transport channel is shared between the bits produced by the source and channel encoders so to balance between the source and channel distortions. Some examples of previous research efforts to address this problem can be found in [7]-[19]. The main drawbacks of these works are overly simplified channel models and the use of the means square error (MSE) as a distortion measure. MSE is a simple mathematical measure that does not distinguish between the source distortion due to quantization artifacts, which make the image fuzzy, and the channel distortion due to packet losses, which typically results in “blockiness” (manifestation depends on loss concealment algorithms employed at the decoder). The optimal bit-rate allocation depends on how users perceive different impairments. We tried to address these drawbacks by considering a detailed channel model and by using a perceptual distortion measure that better describes the structural image distortion in a video sequence. However, the choice of a distortion measure that best corresponds to the human perception of video quality still remains an open issue.

In addition to the source and channel coding, the bit-rate allocation problem may encompass other mechanisms that require channel capacity and need to be adapted to varying channel conditions. For instance, in wireless channels, the bit error rate depends
greatly on the channel estimation error. Channel estimation mechanisms commonly use pilot symbols multiplexed into the data stream and hence require a share of the available transmission rate. This share is not negligible, especially in fast fading channels (e.g. at vehicular speeds). Therefore, in addition to the trade-off between source and channel bits, the bit-rate allocation problem can be extended to distortion-optimal assignment of source, channel, and pilot bit-rates. This is an example of a cross-layering problem, where various mechanisms need to work together to render the best quality for users.

In paper A of this thesis, we present a structural distortion model for the video transmission system illustrated in Fig. 1. The model captures the source and channel distortions as a function of source, channel, and pilot data rates. Using the model, we study trade-offs associated with different bit-rate allocation strategies. We show how the bit-rates can be allocated to ensure graceful degradation of video quality when channel conditions vary. This is shown both in principle and in a practical example of channel quality index (CQI) selection in HSDPA systems. The CQI selection is basically a link adaptation problem since the selected CQI determines the modulation and coding schemes that will be used at the base station. In deployed systems, CQI is chosen so to keep the packet error rate below a certain threshold rather than to maximize user perceived quality.

Our goal in the paper A of this thesis was to give an insight into the problems that need to be resolved when designing a practical bit-rate allocation protocol rather than to propose one. In practice, it might be necessary to estimate the distortion at different source coding rates and channel coding conditions in real time. Because of the limited CPU capabilities of mobile devices, the distortion estimation process must not be computationally intensive for the optimization scheme to be practical. Algorithms such as the recursive optimal per-pixel estimate (ROPE) have been proposed for that purpose [20].

Fig. 1. An example of a video transmission system: $R_s$, $R_c$, and $R_p$ are source, channel, and pilot data rates; $D_s$, $D_c$, and $D$ are source, channel, and total distortions; $P_B$ and $P_F$ are block and frame error rates; $\gamma$ and $a$ are channel parameters.
2.2 Statistical multiplexing

Video traffic is inherently bursty, which poses a big challenge for efficient use of scarce resources in mobile networks. For instance, the H.264/AVC video coding standard, which is the preferred choice for mobile streaming applications because of its compression efficiency, produces traffic which is extremely bursty over a wide range of time scales [21]. To avoid the waste of resources by allocating the transmission resources according to the peak rate, various techniques have been developed to smooth out the burstiness in the video streams. Rate control mechanisms in the encoder, for instance, may ensure that the bit-rate of a video stream complies with channel parameters (transmission rate, buffer capacity), but they produce a lower quality video compared to the variable bit-rate encoding of the same average bit-rate. Link buffering may smooth out the short-term rate variations, but it introduces delay, which might be unacceptable for applications such as mobile TV because the maximum switching time between different TV streams should be fairly short (< 1s). A number of smoothing algorithms have been developed to prefetch the data based on feedback from the playout buffer at the client. This way the server can send large frames at a slower rate without disrupting the playout. These algorithms cannot be used in the case of multicast/broadcast applications because of the lack of client feedback. A viable smoothing technique in the case of multicast/broadcast applications, such as the mobile TV, is to statistically multiplex streams that are destined for the same service area. By multiplexing the streams on a shared transport channel, considerable multiplexing gain can be achieved [22]-[30]. The gain can be measured in terms of increase in video quality (for the same average bit rate) or decrease in total bit-rate (for the same video quality) achieved by transporting a stream as a part of a multiplexed bundle instead of allocating a separate transport channel for it. The total bit-rate determines the amount of spectrum resources (i.e. the number of sub-carriers in OFDM) needed to deliver the streams. Hence, the benefit of statistical multiplexing is improved spectral-efficiency.

In the following, we focus on some implementation aspects of the multiplexing function in Multimedia Broadcast Multicast Service (E-MBMS). Standardization bodies of 3GPP have recently decided that the E-MBMS shall support the mapping of multiple video streams onto the same multicast channel (MCH) [31]. The typical application scenario for the multiplexing is mobile TV delivery in MBMS Single-Frequency Networks (MBSFNs). MBSFN is a new feature introduced in 3GPP LTE (Long Term Evolution) specifications, which enables multicast/broadcast from a time-synchronized set of base stations using the same resource block; the block contains a number of sub-carriers in frequency and a number of sub-frames in time domain. This enables soft combining at the receiver and thus, it improves the signal-to-interference-plus-noise ratio (SINR) significantly compared to non-SFN operation. Most likely, the resource block for a MCH will be allocated semi-statically, i.e. re-allocations may take place only when a stream joins or leaves the bundle [32]. Fully dynamic allocation would require excessive signaling overhead; the feasibility of this approach is an open issue. The semi-statically allocated MCH channels are multiplexed in time within a scheduling interval. Each MCH may belong to a different SFN area. Streams of a bundle, which are mapped to the same MCH channel, are dynamically multiplexed both in time and in
A synchronization protocol guarantees that all base stations of an MBSFN area have the same data of a certain video stream available for transmission. The dynamical multiplexing scheme in base stations should follow standardized rules to ensure that identical resource allocation patterns within an MCH are used in all base stations. Each base station should notify the mobile terminals about the chosen resource allocation pattern. Therefore, a user is not required to receive all sub-frames/sub-carriers of an MCH, but only those that contain the stream of interest, e.g. the TV channel that the user is tuned-in to. This is important from the point of power efficiency since terminals would otherwise have to receive the whole resource block allocated to the MCH.

In paper B of this thesis, we evaluate the statistical multiplexing gains of H.264/AVC video in E-MBMS. The evaluation is based on experiments with a rich set of video sequences that we collected. We consider both semi-static and dynamic allocations of the multicast channel. We have introduced and evaluated a number of mechanisms that ensure that the bit-rate requirements of the bundle do not exceed the bit-rate of the allocated channel. This can be achieved by coordinated encoding of multiplexed video streams. If the encoder coordination is not possible, some video frames may need to be dropped at the base station. We consider both cases. The statistical multiplexing gains are measured relative to the case where rate-controlled streams are delivered on individual transport channels.

2.3 Channel-aware scheduling

Recent high-speed cellular technologies such as HSDPA and LTE are entirely packet-switched. Data flows are multiplexed on a shared downlink transport channel and, therefore, packet scheduling policies are crucial to achieve a desired performance. In cellular systems, frequent fluctuations in channel quality due to fading and shadowing give raise to multiuser diversity since all users may experience different channel conditions at any given time [33]-[35]. Hence, at any time, some of the users will enjoy better channel conditions than the others. By scheduling only users with good channel conditions, the shared channel is used in the most efficient manner and the total
system throughput is maximized. This is the basic principle behind channel-aware (opportunistic) scheduling. In multi-channel OFDM systems with frequency selective fading, the channel-aware scheduling may also exploit the frequency diversity by scheduling users on sub-channels on which they experience good radio conditions. This leads to time-and-frequency-domain scheduling; resources being scheduled are blocks in the OFDM matrix. There are several requirements for the deployment of channel-aware scheduling: channel state information for each user must be available at the base station, the scheduling interval should be fairly short to allow timely reactions to varying channel conditions, and the base station should be able to adapt the transmission rate to the channel conditions using power control or adaptive modulation and coding.

The throughput gain of opportunistic scheduling depends on channel dynamics, the number of users and sub-channels in a cell, and traffic demands. It also depends on the *latency window* over which the channel variations are exploited. The gain is maximized if, at each time-slot, a user with the best signal-to-noise (SNR) ratio is scheduled. This is the so-called Max-C/I (maximum carrier-to-interference) scheduling. Clearly, this scheduler may starve users that are further away from the base station or in a deep fade for a prolonged period of time. To avoid this, SNR is often normalized by the average SNR, which is calculated over a number of previous time-slots determined by the size of the latency window. The scheduling decision is then based on the normalized SNR. This guarantees that each user will eventually be scheduled, even if it suffers from unfavorable channel conditions. The average packet delay introduced by the scheduler is proportional to the latency window. One of the challenges when scheduling delay-sensitive video traffic is to choose a latency window that provides an optimal balance between the throughput gain (spectral efficiency) and packet delay (delay jitter). Another challenge is to extend the channel-aware scheduling algorithms, which are tailored to handle unicast flows, to multicast scenarios (e.g. as in the case of a mobile TV service that multiplexes video streams on a common transport channel). In particular, this thesis addresses the following three problems in channel-aware scheduling:

- Throughput-delay trade-offs that the popular proportional fair (PF) scheduler needs to balance when the offered load in a cell consists of a mix of delay-sensitive and elastic flows.
- Integration of elastic unicast and streaming multicast flows on a common transport channel and the extension of the PF scheduler to handle such scenarios.
- Multicast scheduling for layered scalable video streaming, where scheduling decisions involve the choice of a layer to transmit, so that the channel distortion is minimized.

In our work, special attention is given to detailed and realistic channel and mobility models. The performance of channel-aware schedulers depends on the rate and the range of channel fluctuations, which are, to a large extent, determined by mobility. In our simulations, we typically distinguish between vehicular and pedestrian nodes. The mobility of vehicular nodes is captured by the Manhattan mobility model, which
assumes that vehicles move randomly on a grid of streets. The mobility of pedestrian nodes is captured by the UDel mobility model, which assumes that pedestrians move according to empirical activity models [36], [37]. The channel model includes fast fading, shadowing, and propagation loss, which depend on the speed and the distance of a node from the base station. Each simulation run is preceded by a 5 to 10 minutes warm-up period to ensure that the steady-state spatial distribution of nodes is reached. The length of the warm-up period is determined tentatively, based on the visual inspection of mobility traces in a node animator. In most of our simulation scenarios, the coefficient of variation (ratio of standard deviation and mean) in the results is quite low (< 1 %). However, some of the simulation scenarios in papers C to E are extremely computationally intensive since packet scheduling is simulated in a time-slot resolution of 2 ms. In such scenarios, we allow a higher coefficient of variation (up to 20 % in some cases) because the number of simulation runs that we were able to perform was limited.

2.3.1 Throughput-delay trade-offs in PF scheduling

The proportional fair (PF) scheduler [38], [39] is a popular algorithm that provides a good balance between spectral efficiency and fairness. It has been implemented in 1xEV-DO/IS-856 and HSDPA systems and proposed for LTE. The PF scheduler has received considerable attention in the research community (see [40] and references therein). It has been shown that, among all scheduling strategies, PF maximizes the product of throughputs achieved by different users and, therefore, the set of achieved throughputs is proportionally fair. Hence, it is not a heuristic algorithm since it corresponds to the solution of a concrete optimization problem [41].

Suppose that there are $N$ users in the cell and $R_i(k)$ is the achievable rate for user $i$ in time-slot $k$, which depends on the user’s current channel conditions (SNR). Then, according to the PF scheduler, user $J_k$ is chosen for transmission in time-slot $k$ if:

$$J_k = \arg \max_{1 \leq i \leq N} \frac{R_i(k)}{T_i(k)}, \quad k = 1, 2, ..., ,$$

where $T_i(k)$ is the running average of the achieved data rate for user $i$. It is updated at each time-slot as:

$$T_i(k + 1) = (1 - \frac{1}{t_c})T_i(k) + \frac{1}{t_c}R_i(k) \cdot 1_{\{i=J_k\}}$$

where $t_c$ is the memory of the averaging filter that correspond to the latency window and $1_{\{\}}$ is an indicator function that takes value one if the condition in the brackets is fulfilled and zero otherwise. Thus, the algorithm schedules a user when his instantaneous channel quality is relatively high compared to his own average channel conditions over the time-scale $t_c$. By choosing the latency window $t_c$, it is possible to tune the PF scheduler to exploit the multi-user diversity more or less aggressively and, therefore, trade the cell throughput for lower delay jitter and vice versa. When the latency window is equal to one, users are scheduled at random. The delay jitter is then
low, but the throughput gain is zero and the throughput performance is equivalent to that of the round-robin scheduler. The PF scheduler with an infinitely large latency window corresponds to the Max C/I scheduler (assuming that $T_i(k)$ of all users are initialized to the same value). The cell throughput (spectral efficiency) is maximized, but the delay jitter can be substantial, depending on the channel dynamics.

If the latency requirements of all flows are flexible, the PF scheduler can be used to achieve high spectral efficiency. However, for video streaming, the delay jitter may cause underflow in playout buffers and trigger re-buffering events during which the video playout freezes. The performance of the scheduler, which is not easy to quantify, depends on the service mix in the cell. To decide whether to enable PF scheduling in their base stations, operators need a thorough understanding of its impact on the user performance. Using detailed ns-2 simulations, we measured the throughput gains and re-buffering delays achieved with PF scheduling for various latency windows, mobility patterns, and offered loads. Our contributions are presented in the paper C of this thesis.

### 2.3.2 Integration of elastic unicast and streaming multicast flows

In Section 2.2, we considered statistical multiplexing of video streams as a way to reduce traffic burstiness and increase spectral efficiency. We also discussed the trade-offs associated with unicast channel-aware scheduling of elastic and streaming flows. Traffic in a cell is however most often a mix of elastic unicast flows and streaming unicast or multicast flows. Therefore, resource allocation mechanisms should enable the integration of unicast and multicast flows with diverse service requirements on a shared resource, a common transport channel. It has been suggested in 3GPP that LTE should support multiplexing of unicast and multicast/broadcast MBMS traffic on a shared TDM channel [42]. The reason is improved spectral efficiency. This integration requires new multicast scheduling algorithms. The main difference compared to the unicast scheduling is that multicast scheduling needs to address the problem of intra-group fairness: Members of a multicast group typically have diverse channel conditions and therefore, it is not possible to select a transmission rate at the base station that would be optimal for all. If the selected rate is higher than the achievable rate of the weakest user, some of the multicast users will not be able to receive the data. Another issue in the unicast/multicast integration is resource fairness. While it seems reasonable that a large group of users should receive more resources (e.g. time-slots) than a unicast user, such allocation could be considered unfair in terms of transmission opportunities. Hence, the notion of resource fairness becomes somewhat elusive.

Integration of unicast and multicast data in wired networks has been previously addressed in the context of IP multicast. In [43], the authors propose a new scheduler to share bandwidth in a TCP-friendly manner between unicast TCP flows and multicast flows. A bandwidth sharing mechanism that aims to maximize the sum of “satisfactions” of unicast and multicast users is proposed in [44]. Our early work on optimal resource sharing for integration of unicast and multicast flows on TDM channels was, to the best of our knowledge, first to address this problem in the context of wireless networks [45]. Extensions of the unicast PF scheduler to multicast scenarios similar to those presented in the paper D of this thesis have been proposed in [46]. There we consider scenarios when a transport channel is shared by a number of multicast streaming users and unicast elastic users. The satisfaction of each user is given by his utility curve; different users
may have different utility curves. The objective is to find a scheduling that maximizes the product of user satisfaction under certain fairness constraints. In each time-slot, the multicast scheduler needs to select a unicast user or a multicast group and to decide on the transmission rate so that the proportional fairness objective is met.

### 2.3.3 Multicast scheduling for scalable video streaming

Scalability is a highly desirable feature for many multimedia applications including video streaming. It should provide graceful changes in video quality when transmission conditions vary. Scalable Video Coding (SVC) is the extension of H.264/AVC video compression standard that enables a video sequence to be encoded into a number of layers, typically one base layer and several enhancement layers. A subset of layers can be decoded to produce a video of lower spatial resolution, lower frame rate, or lower quality compared to a video decoded from the complete set of layers. When a wireless user suffers from bad radio conditions and his achievable throughput is low, he might be able to receive the base layer to reproduce a video of minimum quality. When his radio conditions are favorable, he might receive additional enhancement layers to improve the quality. Scalable video coding is particularly suitable for multicast because it facilitates video streaming to a set of receivers with heterogeneous channel capacities. When a non-scalable video stream needs to be delivered to all users in a multicast group, it has to be streamed at the rate of the weakest user in the group. This significantly limits the utility of users with favorable radio conditions. With the appropriate scheduling algorithms, scalable coding would ensure that users with good channels receive additional layers and achieve better playback quality. Therefore, it is important to extend unicast schedulers for scalable streaming to handle multicast scenarios.

The unicast scheduling of scalable video streams has been recently addressed in [47], [48]. The proposed algorithms rely on complex packet prioritization methods: Each video packet is assigned a distortion gradient which reflects its importance—a packet is considered more important if its loss would cause larger increase in the video distortion. SVC employs complex motion-compensated prediction algorithms, which makes it very difficult to calculate the expected distortion increase due to the loss of a particular packet. Besides, this increase depends on the history of losses among previously transmitted packets, which is not available at the base station. This makes the scheduling schemes based on explicit packet prioritization too complex to be practical. Our objective in the paper E of this thesis is to propose low complexity channel-aware algorithms for multicast scheduling of scalable video streams. This problem is addressed here for the first time. The target application is mobile TV delivery, where video streams destined for the same service area are multiplexed together and delivered on a shared transport channel. The tasks of the scheduling algorithm are to decide at each time-slot i) which stream should be served, ii) which layer of the stream should be transmitted, and iii) what the transmission rate should be. The low complexity is achieved by associating distortion gradients with video layers, not with individual packets.
3. Mobility-assisted communication in delay-tolerant networks

Delay-tolerant networks (DTNs) [49] enable transfer of data when network nodes are connected only intermittently. Such networks are assumed to experience frequent partitioning and often lack end-to-end paths. The intermittency may stem from a number of reasons (node failures, unreliable communication channels, power management with unsynchronized wake-up schedules, etc.), but most often it is due to the mobility of nodes. This thesis concentrates on disconnected mobile ad hoc DTN networks of portable devices (mobile phones, PDAs, MP3 players) carried by people. These devices are often equipped with short-range radio interfaces (Wi-Fi, Bluetooth), which enable them to communicate directly without a support of infrastructure. A scheme where data is carried by people walking around and relayed from device to device by means of short range radio could be used to form public content distribution networks that span large urban areas. It could also be used to extend the communication beyond the coverage of infrastructure networks, or to provide an alternative transport mechanism in areas where such coverage exists. This type of networks is sometimes labeled as opportunistic because communication opportunities arise sporadically with the mobility [50]. Therefore, the performance in such networks is to a large extent affected by human mobility patterns. In this section, we provide an overview of efforts made by the DTN research community to enable content delivery in opportunistic networks and to capture the effects of human mobility.

3.1 Opportunistic content delivery

Intermittent communication provided by opportunistic networks could be useful for content delivery characterized by a low degree of interactivity, e.g., broadcasting, messaging, and data collection. The “killer” application, however, is yet to be found [51], [52]. The advocates of DTNs often stress that it took 11 years for the Internet to find its killer application in the form of the World Wide Web. Some of the following applications and services may provide a similar boost to opportunistic networks: peer-to-peer file sharing and bulk data transport, wireless podcasting, location-aware advertising, mobile sensing and data collection, mobile pervasive gaming, mobile social networking, emergency communication in disaster situations, and cooperative data backup for mobile nodes. To support a spectrum of application as broad as possible, opportunistic network architectures must provide solutions to a number of technical challenges, such as resource-efficient routing, rapid node/service discovery, intelligent content segmentation and caching, low energy consumption, strong cooperation incentives, and enhanced data security. Most of these challenges are not new—they have been previously addressed in the context of mobile ad hoc networks (MANETs), which can be seen as a subset of opportunistic networks. A comprehensive overview of technical problems in supporting delay and disruption tolerant communication can be found in [50], [53].

The key challenge in opportunistic networks is to develop routing algorithms to forward the data from a source to a destination. Routing protocols proposed for MANETs, such as AODV and DSR, assume that an end-to-end path exists at any
time—this is not the case in delay-tolerant networks. Therefore, opportunistic routing algorithms take a store-carry-forward approach and often replicate the data into many copies to increase the likelihood of successful delivery. When the application is broadcast in nature or the locations of intended receivers are completely unpredictable, the data is typically flooded into the network. Epidemic routing [54] is an example of a flooding algorithm where nodes replicate the data whenever they meet another node that does not possess the data. Epidemic routing maximizes the probability of finding a path to destinations, but it is resource-hungry (storage, battery power) and inherently non-scalable. A number of algorithms have been proposed to limit the replication. Spray and Wait [55] sets a strict upper limit on the number of copies per message allowed in the network. The replication (“spray”) phase ends when the limit is reached and the receiver must wait until it directly encounters one of the nodes that carry the copies. Some of the proposed algorithms maintain a history of encounters and exploit the fact that mobility is often not entirely random to decide when to replicate the data [56]–[58]. For example, the PRoPHET protocol (Probabilistic Routing Protocol using History of Encounters and Transitivity) [56] maintains a set of delivery probabilities to known destinations in each node and replicates a message only when encountered nodes have a good chance of delivering it. The RAPID protocol (Resource Allocation Protocol for Intentional DTN routing) [59] treats replication as a resource allocation problem of optimizing a specific routing metric, such as the worst-case delivery delay or the fraction of packets that are delivered within a deadline. When network nodes are people, routing may exploit the social aspects of human mobility. People belong to different communities and have different degrees of centrality, which measures how often they encounter members of their own and other communities. Bubble Rap [60] is a routing algorithm that selects high-centrality nodes and community members of destination as message relays. Social Selfishness Aware Routing (SSAR) [61] addresses scenarios when nodes are willing to forward the messages for nodes with whom they have social ties, but not for others.

The performance of opportunistic networks depends heavily on the willingness of nodes to cooperate. Some users may decide to act selfishly and refuse to store and forward messages for the benefit of other users to preserve battery, storage and CPU resources. Cooperation incentives have been studied extensively in the context of p2p file sharing in the Internet and message relaying in MANETs. The impact of user cooperation on the efficiency of content distribution in intermittently connected networks has been studied in [62], [63]. A number of incentive systems based on virtual payments, reputation, and social proximity have been proposed in the literature [64]–[66]. Researchers have also studied under what conditions cooperation can emerge spontaneously [67], [68].

Other problems in opportunistic content delivery have received far less attention compared to the routing and cooperation. Time and energy efficiency of node discovery in DTN networks formed by Bluetooth devices has been addressed in [69]. The authors proposed an adaptive node discovery scheme that may trade the energy consumption for shorter discovery delay and vice versa. A search scheme for content discovery in DTNs has been proposed in [70]. Security is an issue in opportunistic networks because data traverses peer nodes, which cannot necessarily be trusted. Addressing security issues has been a major focus of the IETF bundle protocol for the exchange of messages.
("bundles") in DTNs [71]. Contact durations in opportunistic networks are often unpredictable and, therefore, partial message transfers must be supported. In [72], the authors investigate the performance of proactive and reactive message fragmentation strategies, which are part of the IETF DTN architecture [73].

Besides the original IETF DTN architecture [73], several architectures for opportunistic content distribution have been proposed and implemented as experimental testbeds. Haggie [74] is a data-centric software architecture for opportunistic networks that focuses on applications that incorporate social aspects. DieselNet [75] is a mobile testbed composed of wireless nodes deployed on public transportation buses and fixed nodes called “throwboxes”. Mobility traces collected in the Haggie and DieselNet projects have been used by many researchers to study how mobility impacts message forwarding. KioskNet [76] is a network of rural Internet kiosks that provide data services in remote regions. Vehicles with on-board computers ferry the data between the kiosks and gateways connected to the Internet. TACO-DTN [77] is a content-based dissemination system composed of fixed and mobile infostations that allow mobile users to subscribe to certain contents for a period of time. PeopleNet [78] is an architecture for efficient information search in a distributed manner. It uses fixed infrastructure to propagate queries to users in specific geographic locations, called bazars. Within each bazaar, the query is further propagated between nodes via opportunistic contacts. Hikari [79] is publish/subscribe architecture for opportunistic message dissemination in the Paris subway system.

PodNet [80] is an experimental platform that extends Internet podcasts to mobile ad-hoc domains. An Internet podcast is a series of multimedia contents that are released episodically and published through a web feed. A client application known as a podcatcher, which runs on the user's computer or a portable device, can be used to access this web feed and download new files in the series. This process can be automated so that the new files are downloaded automatically. The podcasting has been given a major push by conventional media such as TV and radio broadcasters. Today, the vast majority of top-rated podcasts come from recognizable media entities. A report by eMarketer predicts that the number of Internet users who download at least one podcast per month will grow from 9 % in 2008 to 17 % in 2013 [81]. The podcasting is growing in popularity because it allows users with portable devices to consume contents on the move, while riding public transportation, or waiting in a queue. The extension of podcasting to wireless and ad-hoc domains would eliminate the need to dock the portable devices to download the contents. It would also allow more and more content to be provided by the mobile users themselves. An architecture for mobility-assisted ad hoc podcasting is best described in [82], [83]. In the paper F of this thesis, we compare the spectrum and energy efficiencies of infrastructure-based wireless podcasting and the mobility-assisted ad hoc podcasting. Our study demonstrates how opportunistic content delivery can be used to off-load cellular infrastructures.

3.2 Pedestrian mobility models

In opportunistic networks, people carry data through their own movements and, therefore, their mobility patterns affect the speed, throughput and reliability of the data
forwarding. To perform large-scale real-life experimental evaluations of such networks is cumbersome and expensive, yet each experiment only captures a particular scenario. Therefore, results from one experimental setting are difficult to generalize. Mobility traces available in public trace repositories are often not suitable for the performance evaluation of opportunistic networks due to their coarse time-granularities (see paper G for details). Therefore, it is vital to mathematically model mobility. Simulation models allow us to produce synthetic mobility traces of desired time-resolution. Analytical models can be used to study the connectivity properties directly. Unfortunately, it is rather difficult, if not impossible, to mathematically model human mobility. In the following, we discuss aspects of the mobility that need to be captured in a comprehensive mobility model.

Human mobility can be described at three levels: strategic, tactical, and operational [84]. At the strategic level, people decide on their destinations based on activities that they want to perform. It is extremely difficult to model human behavior at this level. Sometimes people do not have a destination at all, or their destination changes rapidly depending on the environment and local attractors. Destination choice models used in simulations are often empirical. For instance, the workday activity model implemented in the UDel simulator is based on surveys performed by the US Bureau of Labor Statistics [85]. Literature on the activity set generation is very scarce. Analytical models do not address human mobility at the strategic level; the choice of destinations is considered exogenous to the models.

At the tactical level, people decide which route to take between destinations. Route choice is an important aspect of the mobility and literature on this theme is plentiful. Comprehensive overviews of existing route choice models can be found in [86], [87]. It is common to consider route choice as a utility maximization problem. The utility depends on route attributes such as travel time, distance, “attractiveness”, safety, etc. Personal attributes such as sex, age, and cultural background also play significant role. Social interaction influences the route choice as well. It has been observed for instance that in many situations pedestrians favor well established routes and decide to follow other pedestrians. Yet most of the route-choice models are based on simple shortest path algorithms. Also most determine a complete set of route alternatives from the origin to the destination, of which one is chosen. In practice, sequential route choice might be a more common case due to the variations in traffic conditions. In cellular automata (CA) models [88], a route is decided in discrete steps, based on the traffic conditions in neighboring “cells” (space segments in which movement takes place). Therefore, the route choice problem is approached on an operational rather than on a tactical level.

At the operational level, people make decisions on their walking speed and navigate to avoid collisions with obstacles and other pedestrians. Decisions at this level constitute their walking behavior in a narrow sense. Mobility models traditionally used in the networking community typically assume free, unobstructed flow of people. Walking behavior of a person is however very dependent on the actions of his immediate neighbors. For instance, a person slowing down in front is likely to cause others to slow down or change direction. This type of interaction among pedestrians will affect the rate and the duration of contact opportunities that can be used for content delivery. Literature on human walking behavior is abundant [87]. Existing mathematical


models at this level are often classified as macroscopic, mesoscopic, and microscopic. Macroscopic models describe mobility at a high level of aggregation as a flow of people with fluid-like properties in which individual pedestrians become indistinguishable. Typically, these models describe how density and velocity of the flow change over time and location. They show some analogies with models used in fluid mechanics and therefore are often referred to as hydrodynamic models. Gas-kinetic mobility models draw analogies with the kinetic theory of gases. They describe the evolution of a velocity distribution function, which is derived based on the microscopic interactions among pedestrians. Therefore, these models are often classified as mesoscopic [89]. Microscopic models capture the walking behavior of each pedestrian under the actions of surrounding pedestrians. These models draw analogies with Newtonian systems of particles. Some microscopic models are not inspired by analogies from physics. For example, behavioral models based on discrete choice analysis are calibrated on real pedestrian trajectories [90].

Currently, there are no comprehensive analytical or simulation models that capture human mobility at all levels. Here we provide a brief overview of mobility models commonly used in the networking community. Some of these models have been used in this thesis. Random Waypoint (RWP) is a popular model that describes the mobility patterns of independent, non-interacting nodes in an open area. Each node moves along a zig-zag line from one waypoint to the next. The waypoints are uniformly distributed in the area. In each waypoint, a node makes a pause before it moves towards the next waypoint. RWP model is elementary and it is easy to argue about the paths being unnatural. It owns its popularity to simplicity and mathematical tractability: results such as the contact rate and the stationary distribution of nodes have been provided in closed form [91], [92]. The RWP has been used as a basis for many other models. Constrained Random Waypoint model includes geographical restrictions—nodes are restricted to travel between waypoints using pathways. Based on empirical observations, the authors in [93] propose Levy Walk model, which assumes power-law distributions of flight lengths and pause times in the waypoints [93]. The objective is to faithfully capture inter-contact times between nodes. Manhattan mobility model assumes node movements on a grid of streets and intersections. At each intersection, a node randomly selects its future direction of movement [94]. A survey of random walk models, as well as several models with spatial and temporal dependencies, is provided in [95]. There are also many models that have been derived based on observed mobility data or geographical information. UDel mobility model aims to capture a typical day cycle of a working person based on data collected by the US Bureau of Labor Statistics [96]. The activity model includes a number of tasks performed on a realistic city map: traveling from home to work, sitting/working in an office, visiting a restaurant during the lunch-break, traveling back home after work, etc. Working Day Movement model implemented in the ONE simulator has a similar objective [97]. Weighted Waypoint model defines a set of congregation points (destinations), such as cafeterias and classrooms on a university campus. Based on a distribution of pause times for each destination and a transition probability for each pair of destinations, it constructs a Markov chain of nodes’ movements [98]. Urban Pedestrian Flows is a macroscopic model implemented in the MobiREAL simulator. It uses linear programming techniques
to reproduce observed pedestrian densities on a city map [99]. A group mobility model that aims to capture the way people socialize and travel together is presented in [100].

When evaluating the performance of cellular systems, the concern is mostly with the mobility at strategic and tactical levels (macro-mobility), which affects the variations in channel quality over time, and less with the mobility at the operational level (walking behavior). However, when evaluating the performance of opportunistic networks, this might not be the case. The walking behavior is responsible for platooning, creation of virtual lanes and “zipper” formations at bottlenecks, and other characteristics of moving crowds that may affect the rate and the duration of contact opportunities. Elaborate efforts to mathematically describe these phenomena are found in the areas of urban planning, transportation, and traffic engineering. The social force model by Helbing at al. [101] has inspired a number of relatively simple models that aim at capturing behavioral aspects of human mobility. The model assumes that each pedestrian is driven by two kinds of forces: social and physical. The social forces do not have physical source; they reflect the intentions of pedestrians not to collide with other people and obstacles. In response to these forces pedestrians accelerate or decelerate as if they were subjects to external forces. Several commercial multi-agent simulators use social force model to generate mobility, the most notable being VISISM [102] and Legion Studio [103]. Legion Studio is used in urban and traffic planning to design public spaces such as airports, subway stations, and sport stadiums. It aims to faithfully capture speed-distance relations that emerge when people navigate obstacles and other pedestrians. These relations are responsible for the formation of pedestrian crowds—capturing them is of paramount importance when dimensioning emergency exits, stairs, and escalators. Legion Studio is based on detailed models that describe the mechanics of movements and algorithms that pedestrians use to select routes. It has been calibrated based on measurements performed in different locations and among pedestrians with different cultural backgrounds.

Our work focuses on developing a library of analytical models that would be used to study the performance of pedestrian content distribution in some common case scenarios of urban mobility. The library would consist of models for abstract spaces where people move. A “street model”, for example, would describe the rate and the duration of contact opportunities in one-dimensional topologies such as streets, sidewalks, or corridors. A “square model” would describe two-dimensional spaces, such as city squares, parks, parking lots, or airport halls. A “point model” would provide an abstraction of spots where people congregate: crosswalks, bus stops, pubs/cafés, or check-in terminals. The models could be used as building blocks to study the content distribution in larger areas. In the paper G of this thesis, we describe the “street model” and how it can be used to build “grids of streets”. The outputs of the queuing-analytical model are the rate and the duration of contact opportunities, as well as the “content dispersion”: the share of nodes that will obtain a message brought into the area by some of the nodes. To obtain a tractable model, we introduced a number of simplifying assumptions (e.g. Poisson arrivals, uniform speeds, no platooning, etc.). To identify scenarios where the model provides a good approximation of reality, we compare the outputs of the model with the results of simulations performed in the Legion Studio.
4. Summary of original work

4.1 Paper A: Trade-Offs in Bit-Rate Allocation for Wireless Video Streaming

V. Vukadinović and G. Karlsson
Published in *IEEE Transactions on Multimedia*, vol. 11, no. 6, Oct. 2009.

**Summary:** One of the central problems in wireless video transmission is the choice of source and channel coding rates to allocate the available transmission rate optimally. In this paper, we present a structural distortion model for video streaming over time-varying fading channels. Based on the model, we study the end-to-end distortion for various bit-rate allocation strategies and channel conditions. We show that the robustness to channel variations is crucial for the streaming performance when frequent bit-rate adaptations are not feasible. The robustness is achieved at the expense of higher source distortion in the encoder. Our findings are illustrated on a practical problem of distortion-optimal selection of transport formats in an adaptive modulation and coding (AMC) scheme used in HSDPA.

**Contribution:** The author of this thesis formulated the problem, developed the mathematical model, carried out simulation experiments, interpreted the results, and wrote the article under the supervision of the second author.

4.2 Paper B: Statistical multiplexing gains of H.264/AVC video in E-MBMS

V. Vukadinović and J. Hüschke

**Summary:** With the increasing acceptance of H.264/AVC as a video coding standard for mobile multimedia, it becomes very important to control the extreme burstiness of the traffic generated by the H.264/AVC encoder. Statistical multiplexing of video streams can be used to reduce the bit rate variations. In practice, the number of video streams that can be multiplexed together is quite limited and therefore, additional mechanisms are needed to eliminate the residual burstiness and to adapt the bit rate of the bundle to the bit rate of the transport channel. Here we evaluate the potential statistical multiplexing gains with and without the coordination in the encoders. The gains are measured relative to the case where rate-controlled (capped) streams are delivered on individual transport channels.

**Contribution:** The author of this thesis collected and processed the test video sequence, carried out simulation experiments, interpreted the results, and wrote the article under the supervision of the second author. The second author formulated the problem.
4.3 Paper C: Video Streaming Performance under Proportional Fair Scheduling

V. Vukadinović and G. Karlsson
Published in IEEE Journal on Selected Areas in Communications (JSAC), vol. 28, no. 3, Apr. 2010.

**Summary:** In this paper, we study the performance of the proportional fair scheduler, which has been proposed for some of the emerging radio access systems. It maximizes the spectral efficiency of the systems, which is a strong incentive for network operators to use it. Our goal is to investigate how the proportional fair scheduling affects the performance of streaming users. Focus is on the throughput-delay trade-offs associated with proportional fair scheduling in CDMA/HDR systems. Special attention is devoted to defining appropriate performance measures and creating a realistic simulation environment. Our results indicate that the opportunistic scheduling may face difficulties in providing the user-level performance in cases where streaming flows constitute a significant share of the traffic load.

**Contribution:** The author of this thesis formulated the problem, carried out simulation experiments, interpreted the results, and wrote the article under the supervision of the second author.

4.4 Paper D: Multicast Scheduling with Resource Fairness Constraints

V. Vukadinović and G. Karlsson
A version of this paper has been published in Springer Wireless Networks, vol. 15 no. 5, July 2009.

**Summary:** In future cellular networks, multicast and unicast users may share a common transport channel in order to improve the resource efficiency. This requires new and efficient scheduling mechanisms. These mechanisms need to be optimized to provide the best possible trade-off between resource efficiency and fairness. In this paper, we consider a case where streaming multicast users are multiplexed together with elastic unicast users on a common time-slotted channel. We derive a system model to study the user performance under proportional and resource fairness constraints. Fairness is defined in terms of the users’ utilities rather than of the throughputs they are assigned to. We also describe an extension of the well-know unicast proportional fair scheduler to the multicast scenario. Through extensive simulations we demonstrate the performance of this scheduler for various traffic loads and multicast group sizes.

**Contribution:** The author of this thesis formulated the problem, developed the mathematical model, carried out simulation experiments, interpreted the results, and wrote the article under the supervision of the second author.
4.5  Paper E: Multicast Scheduling for Scalable Video Streaming in Wireless Networks

V. Vukadinović and G. Dan

**Summary:** In this paper, we study how relatively simple extensions of popular channel-aware schedulers can be used to support scalable multicast of video streams. To support the evaluation, we first describe a model of the channel distortion of scalable video coding and validate it using eight commonly used test sequences. We use the distortion model in a detailed simulation setup to compare the performance of six schedulers, among them the Max-Sum and Max-Prod schedulers, which aim to maximize the sum and the product of streaming utilities, respectively. We investigate how the traffic load, user mobility, layering structure, and users’ aversion of fluctuating distortion influence the streaming performance. Our results show that the Max-Sum scheduler performs better than other considered schemes in almost all scenarios. With the Max-Sum scheduler, the gain of scalable video coding compared to non-scalable coding is substantial, even when users do not tolerate frequent changes in video quality.

**Contribution:** The author of this thesis formulated the problem, developed the model, formulated the scheduling algorithms, carried out simulation experiments, interpreted the results, and wrote a part of the article. The second author helped to formulate the problem, to interpret the results, and he wrote a part of the article.

4.6  Paper F: Spectrum and Energy Efficiency of Mobility-Assisted Podcasting in Cellular Networks

V. Vukadinović and G. Karlsson
Submitted to *IEEE Transactions on Mobile Computing*.
A shorter version of this paper has been published in *Proc. ACM International Workshop on Mobile Opportunistic Networking (MobiOpp)*, Pisa, Italy, Feb. 2010.

**Summary:** In this paper, we consider a wireless podcasting service for mobile users. That can be provided by mobile operators in their coverage areas. Our focus is on dense urban environments. The latency and throughput requirements of the podcasting service are flexible and, therefore, podcast contents do not need to be delivered to all users via the operator’s network: they can be relayed from one user to another in a store-carry-and-forward fashion. Hence, mobility of users can be used as a supplementary transport mechanism to off-load the operator’s network and relieve strain on scarce spectrum resources. In this paper, we evaluate the achievable throughputs and spectrum savings of the mobility-assisted wireless podcasting for various mobility patterns and user densities in an area.
**Contribution:** The author of this thesis formulated the problem, developed the mathematical model, carried out simulation experiments, interpreted the results, and wrote the article under the supervision of the second author.

### 4.7 Paper G: An Analytical Model for Pedestrian Content Distribution in a Grid of Streets

V. Vukadinović, O. R. Helgason, and G. Karlsson  
A shorter version of this paper has been published in *Proc. International Conference on Simulation Tools and Techniques (SIMUTools)*, Rome, Italy, March 2009.

**Summary:** We consider a communication system where multimedia content is carried by people walking around and relayed from device to device by means of short range radio. The transport mechanism is the flow of people and it can be studied but not engineered. The question addressed in this paper is how well pedestrian content distribution may work in urban scenarios. We answer this question by modeling the mobility of people on a grid of streets. Our contributions are both the queuing analytic model and the results on the feasibility of pedestrian content distribution. We validate the results using Legion Studio, a pedestrian mobility simulator that captures speed-distance relations that emerge in dense crowds. We also use the simulator to study mobility scenarios that cannot be described by the analytical model.

**Contribution:** The author of this thesis developed the mathematical model, interpreted the results, and wrote a part of the article. The simulation experiments were carried out by a master student supervised by the second author, who wrote a part of the article. The third author formulated the problem and wrote a part of the article.

### 4.8 Other publications

In addition to the papers that are in this thesis, the author has published the following papers in the context of this thesis:


5. Conclusions and future work

In this thesis, we addressed the resource efficiency of multimedia distribution in cellular and hybrid radio systems. Our focus has been on opportunistic mechanisms that exploit mobility to achieve resource-efficient communication. The mobility of nodes induces channel quality fluctuations and creates intermittent connectivity, which may not necessarily impair, but can rather be used to enhance the performance of wireless multimedia systems. The following summarizes the main contributions of this thesis:

- We described a mathematical model of perceptual video distortion in wireless channels. The model captures the dependence of source (rate) and channel (loss) distortions on the bit-rate allocation chosen for transmission. Based on the model, we discussed the trade-offs between source, channel, and pilot rates that need to be made in order to minimize the total distortion when channel conditions vary. We illustrated those trade-offs on a practical example of CQI selection in HSDPA, which is usually seen as a problem of keeping the block error rate in the channel below certain threshold, rather than as a cross-layered distortion-optimization problem.

- We considered the statistical multiplexing of video streams as a way to improve resource utilization in multicast/broadcast systems. Our experiments have shown that non-negligible gains in the transport capacity of radio bearers can be achieved in the case of coordinated encoding. The multiplexing gains are measured relative to the rate-controlled streams. This way of measuring gains gives us a metric of a higher practical relevance compared to the studies where the gains are measured relative to the peak rates of variable bit-rate streams. Our results indicate that most of the H.264/AVC streams can be capped close to their average bit rates with a moderate loss in quality.

- We studied the performance of channel-aware scheduling for video streaming applications. Channel-aware schemes schedule users who currently enjoy good channel conditions. This increases throughput, but it may introduce large delay jitter. We demonstrated the throughput-delay trade-offs associated with the time-domain proportional fair scheduling in CDMA/HDR systems. We also studied the multiplexing of unicast and multicast flows on a common transport channels and its implications on resource allocation mechanisms. Our results show that a relatively straightforward extension of the unicast proportional fair scheduler may provide a good trade-off between resource efficiency and fairness for such channels. Encouraged by this, we looked at how similar low-complexity scheduling schemes can be combined with the scalable/layered video coding to further improve the performance of multicast services, such as the mobile TV service. We have found that a Max-Sum scheduler may provide good performance both in terms of channel distortion and in terms of fairness measured by the playout outage probability. We studied how the total bit-rate of a video sequence should be distributed among the layers and showed that thin base layers fare best in terms of channel distortion and playout outage probability. Finally, we quantified the video quality gains of scalable streaming compared to non-scalable streaming.
We considered the mobility of people as a supplementary transport mechanism to provide wireless podcasting service in cellular networks. We presented a method to evaluate the performance of mobility-assisted podcasting in two limiting regimes: quasi-static and fluid. Our results indicate that the mobility-assisted podcasting uses only a fraction of the spectrum used by the infrastructure-based podcasting to provide the same throughput. The achievable throughput, however, depends on the mobility and the number of podcast users in an area. In a case study that considers a hybrid 3G/Bluetooth network we showed that, in some cases, the mobility-assisted podcasting may also be more energy-efficient than infrastructure-based podcasting.

We developed an analytical model to study the connectivity properties of pedestrian mobility in a street segment. We considered low arrival rates of the nodes into the street, which is the critical case for the performance of the system. The model is then used as a building block to model larger areas, grids of streets. The results obtained from the model give an insight into various system parameters and how they affect the content distribution. We have found that the content spreads with high efficiency in a large number of common-case scenarios. We used a state of the art pedestrian mobility simulator, Legion Studio, for validation. The simulation results are compared to the analytical results with good agreement.

In the following, we summarize some of the research topics that we consider highly relevant for the performance of wireless multimedia transmission, and some of the lesser issues that would complement the research presented in this thesis.

Most of the results presented in papers C to E pertain to single-cell scenarios—inter-cell interference and cell coordination have not been considered. Coordinated transmission from multiple base stations may significantly improve the capacity and resource-efficiency of mobile radio systems, especially in multicast scenarios. Architectures for spatial multiplexing of signals arriving from different cell towers/antennas, such as Single Frequency Networks (SFN) and Cooperative Multipoint Transmission (COMP), have been proposed and evaluated by researchers. These architectures should be optimized not only to suppress interference and improve network capacity, but also to address the service requirements of multimedia applications. Some of the questions that stem from our work are: How to deliver scalable video using coordinated transmission? How to decide (e.g. in a distributed manner) which base stations should be coordinated in unicast/multicast/broadcast scenarios? When do the flexibility and simplicity of single-cell multicast transmission outweigh the benefits of the coordination?

Wireless networks are becoming increasingly diverse. Portable devices with various processing capabilities and screen sizes are nowadays equipped with a variety of radio interfaces. This provides an opportunity to achieve resource-efficient communication by exploiting the distinguishing properties of different technologies, but it also makes the design of communication protocols and algorithms extremely complex. We have scratched the surface of the problem by considering the mobility-assisted podcasting in a hybrid cellular/ad-hoc network. In addition to the base stations and mobile terminals, the network may include a
number of other nodes, such as femtocells, Wi-Fi access points, and fixed stand-alone infostations. In our future work, we would explore not only the attainable spectrum and energy efficiencies of such scenarios, but also legislative, business, and security/privacy aspects that might affect the design choices.

- A thorough understanding of user behavior may help us to design more resource-efficient, reliable, and secure networks. This is particularly true in networks that are technologically diverse and highly mobile. In this thesis, we have addressed human mobility to some extent as one of the factors that affect the performance in such networks. However, we have not considered how people use their portable all-in-one communication-and-entertainment devices in their everyday lives and how they search, select, and consume multimedia contents. Possible extensions of our work would address publication, subscription, and consumption patterns of multimedia podcasting and viewing patterns of mobile TV that affect the number and spatial distribution of multicast users. These patterns would certainly have implications for the design of resource allocation schemes. The analytical mobility model described in this thesis could be extended to include group mobility and attraction points where people congregate, which are characteristics of the human behavior that may have a significant impact on opportunistic content distribution.

The scope and the complexity of resource allocation problems in wireless systems are vast. They are increasing with the availability of various radio interfaces on the same device and the ability to allocate radio resources dynamically on time-scales of channel variations. In this thesis, we addressed some of these problems, yet the possibilities to further improve the proposed solutions using cross-layer and cross-system optimization seem to be endless.
Bibliography


