Research Challenges Towards the Future Internet

Marco Conti, IIT-CNR, Pisa, Italy
Song Chong, KAIST, Korea
Serge Fdida, Universite Pierre et Marie Curie, France
Weijia Jia, City University of Hong Kong, Hong Kong
Holger Karl, Universitat Paderbon, Germany
Ying-Dar Lin, National Chiao Tung University, Taiwan
Petri Mähönen, RWTH Aachen University, Germany
Martin Maier, INRS - University of Quebec, Montreal, Canada
Refik Molva, Eurécom, France
Steve Uhlig, TU Berlin/Deutsche Telekom Laboratories, Germany
Moshe Zukerman, City University of Hong Kong, Hong Kong

Abstract. The convergence of computer-communication networks towards an all-IP integrated network has transformed Internet in a commercial commodity that has stimulated an un-precedent offer of novel communication services that are pushing the Internet architecture and protocols well beyond their original design. This calls for extraordinary research efforts at all levels of the protocol stack to address the challenges of existing and future networked applications and services in terms of scalability, mobility, flexibility, security, etc. In this article we focus on some hot research areas and discuss the research issues that need to be tackled for addressing the multiple challenges of the Future Internet. Far from being a comprehensive analysis of all the challenges faced by the Future Internet, this article tries to call the attention of Computer Communications readers to some new and promising research areas, identified by members of the journal editorial board and to stimulate further research activities in these areas.

1. Introduction

In the last years, all communications media are converging towards the use of the Internet platform. This has stimulated an un-precedent offer of new (ubiquitous) IP services ranging from interactive IPTV and social media to pervasive urban sensing, which are pushing towards a continuous increase in the number of Internet users, and their demand for ubiquitous, reliable, secure and high-speed access to the Internet. This generates several challenges, and hence research opportunities, at all layers of the Internet protocol stack.

To cater the increasing bandwidth demand, network technologies with higher capacity are introduced both in the wired and wireless Internet. Indeed, in the wired part of the network, we observe an increased adoption of the optical networking technologies, both inside the Internet core (i.e., ISP networks) and at the network edges (i.e., fiber at home). A similar trend is observed also in the wireless part of the network, where there is a joint effort of industry and academy for increasing the capacity of the wireless-network technologies. In the wireless field the networks’ capacity is constrained by the limited spectrum, and hence research efforts are mainly devoted to increase the efficiency in the spectrum usage.

The new ubiquitous and multimedia communication services are radically changing the Internet nature: from a host-to-host communication service to a content-centric network, where users access the network for finding relevant content and possibly modifying it. The user-generated-content (UCG) paradigm is further pushing towards this evolution, while the social platforms have an increasing role in the way users access, share and modify the content. The radical departure from the objectives that driven the original Internet design is now pushing towards a re-design of the Internet architecture and protocols to take into account the new design requirements that were outside the original Internet design. Security and privacy is clearly one of the new key requirement, but several other requirements must be taken into account in the Future
Internet design, such as supporting users’ mobility, efficiently handling of multimedia and interactive services, tolerating network partitioning and/or node disconnections, etc. In particular, energy efficiency is emerging as a key design requirement to make Internet sustainable as several reports indicate that the energy consumption due to Internet technologies is already high and, without paying attention to it, the problem will become critical while the Internet role in the society expands.

This paper presents an in-depth analysis of key research areas towards the Future Internet. We wish to remark that the paper is far from being a comprehensive analysis of all the challenges towards the Future Internet, but presents a selection of topics on which we expect/solicit more research contributions in the near future.

The paper is organized according to a layered network organization, in the first three parts we focus on network technologies, Internet architecture and protocols, and applications issues, respectively. The fourth part is dedicated to two cross-layer research issues, i.e., energy efficiency and security. More precisely, in Section 2 we analyse the major challenges to build a scalable and robust (wired and wireless) network infrastructure by discussing the research challenges in the optical (Section 2.1) and wireless (Section 2.2) networking field. In Section 3, we analyze the challenges related to the Internet architecture and protocols. Then, in Section 4, we discuss the role of the mobile-phone technology for delivering multimedia services to ubiquitous users. Section 5 is devoted to two research areas that involve all the layers of the protocol stack: energy efficiency (Section 5.1) and security (Section 5.2). Section 6 concludes the paper with a brief review of other important research areas that have not been covered in detail in this paper.

2. Network Technologies

The development of a broadband and ubiquitous Internet is mainly based on optical network technologies for building high capacity transport and access networks, and on wireless network technologies for providing ubiquitous Internet accesses. Therefore in Section 2.1 we review and discuss the optical networking research challenges, while, Section 2.2 is devoted to analyze and discuss some research challenges for building broadband and scalable wireless networks.

2.1 Optical Networking

During the Internet bubble, the expected bandwidth requirements were hugely overestimated and way too many optical networks were built, flooding the market with unneeded capacity. As a consequence, prices for dark fiber became so low that customers, e.g., banks and corporations with large data transfer needs, started to buy up low-cost dark fibers and run their own optical links and networks. Similarly, the prices for monthly leases of optical fiber connections decreased significantly. For instance, the prices for monthly leases on 10 Gb/s links between Miami and New York City fell from around $75,000 in 2005 to below $30,000 at the end of 2007, and prices for 10 Gb/s connections between New York and London fell by 80% from 2002 to 2007 [MM1]. To make things look even dimmer, it is worthwhile to note that about 80-90% of the world’s installed fiber is unlit, i.e., is not used, and only 18% of the world submarine fiber is lit [MM2]. Given these huge amounts of affordable unused capacity in already installed and heavily overbuilt fiber infrastructures, one naturally asks what the research challenges in optical networks are, if any.

The major problem with dark fiber is the fact that it is abundantly installed in some areas, while it is not available at all at other places. Optical networks are commonplace in wide and metropolitan areas, but in today’s access networks fiber has just started to pave all the way with glass to individual homes and businesses, giving rise to fiber-to-the-home/business (FTTH/B) networks.

---

1 By Martin Maier (University of Quebec, Canada)
According to the latest broadband-related statistics of the OECD (Organisation for Economic Co-operation and Development) broadband portal, FTTH/B connections are still a minority in almost every OECD country, accounting only for 9% of the 271 million fixed wireline broadband subscribers worldwide. However, it is important to note that both digital subscriber line (DSL) and cable networks, the two currently mostly deployed wired broadband technologies, rely on so-called deep fiber access solutions that push fiber ever deeper into the access network [MM3]. While copper will certainly continue to play an important role in current and near-term broadband access networks, it is expected that FTTH/B deployment volume will keep increasing gradually and will eventually become the predominant fixed wireline broadband technology by 2035 [MM4]. FTTH/B networks not only alleviate the notorious first/last mile bandwidth bottleneck, but also give access to the ever-increasing processing and storage capabilities of memory and CPU of desktops, laptops, and other wireless handhelds. While current desktop and laptop computers commonly operate at a clock rate of a couple of GHz with a 32-bit wide backplane, resulting in an internal flow of 2-8 Gb/s with today’s limited hard drive I/O, future desktops and laptops are expected to reach 100 Gb/s [MM5]. In fact, optical buses can now be built right onto the circuit board that will unveil computer systems 100 times as fast as anything available today [MM6].

Recently, the convergence of optical broadband access networks with their wireless counterparts has been receiving an increasing amount of attention [MM7]. These hybrid optical-wireless networks have the potential to provide major cost savings by providing wired and wireless services over the same infrastructure. A lot of research activities focused on the optical generation of radio frequencies and remote modulation schemes in order to build low-cost remote antenna units (RAUs) for radio-over-fiber (RoF) networks. RoF networks have been studied for decades and are well suited for access solutions with a centralized control station, e.g., WiMAX and cellular networks. However, wireless local area network (WLAN)-based RoF networks suffer from a limited fiber length of less than 2 km due to the acknowledgment (ACK) timeout value of the widely deployed distributed coordination function (DCF), which is set to 9 µs and 20 µs in IEEE 802.11a/g and IEEE 802.11b WLAN networks, respectively. These shortcomings can be avoided in recently proposed radio-and-fiber (R&F) networks, where access to the optical and wireless media is controlled separately from each other by using in general two different medium access control (MAC) protocols, with protocol translation taking place at the optical-wireless interface [MM8]. R&F networks pose a number of new challenges and opportunities. Among others, these challenges involve the design and investigation of hybrid MAC protocols, integrated path selection algorithms, integrated channel assignment and bandwidth allocation schemes, optical burst assembly and wireless frame aggregation techniques, as well as flow and congestion control protocols to address the mismatch of optical and wireless network data rates [MM9].

Another important area of ongoing research is the migration from current Gigabit-class ITU-T G.984.x gigabit passive optical network (GPON) and IEEE 802.3ah Ethernet passive optical network (EPON) to next-generation PONs (NG-PONs). NG-PON technologies can be divided into the following two categories [MM10]:

- **NG-PON1**: This type of technologies allows for an evolutionary growth of existent Gigabit-class PONs and supports their coexistence on the same optical distribution network (ODN).

- **NG-PON2**: This category of enabling technologies envisions a revolutionary upgrade of current PONs, giving rise to disruptive NG-PONs without any coexistence requirements with existent Gigabit-class PONs on the same ODN.

NG-PON1 technologies include a number of performance-enhancing options, most notably XG-PONs (the Roman numeral X stands for 10 Gb/s), wavelength division multiplexing (WDM) overlay of multiple XG-PONs, or the deployment of reach extenders to enable long-reach PONs. NG-PON1 will be gradually replaced by NG-PON2 solutions after resolving a number of issues related to the research and development of advanced optical network components and enabling technologies such as dense WDM (DWDM), optical code division multiplexing (OCDM), or orthogonal frequency division multiplexing (OFDM) [MM10].
As broadband access with access rates of 100 Mb/s per subscriber becomes increasingly commonplace in most developed countries, the power consumption of the Internet is estimated to rise to more than 7% of a typical OECD country’s national electricity supply, resulting in a power consumption of several TWh and expenses of millions of dollars and tons of carbon gas emissions per year. The power consumption of today’s Internet is largely dominated by its access networks, which account for roughly 70% of the total power consumption. PONs have been shown to provide the lowest energy solution for broadband access, clearly outperforming alternative fiber, copper, or wireless access solutions based on optical point-to-point Ethernet, DSL, or WiMAX technologies [MM11]. While energy efficiency and low-power techniques have been widely considered important design criteria for wireless networks due to the limited battery life of mobile terminals, more research on energy-aware wired networks is desirable after the first IEEE standard 802.3az for Energy Efficient Ethernet (EEE) has been ratified not until September 2010.

Due to the ever-increasing speed of optical access networks, the bandwidth bottleneck will move from the first/last mile toward metropolitan and wide area networks. To provide higher bandwidth efficiency, current optical metro and core wavelength-switching networks based on reconfigurable optical add-drop multiplexers (ROADMs) could be required to resort to more efficient switching techniques at the subwavelength granularity in the near- to mid-term. A wide variety of optical switching techniques have been investigated, including waveband switching (WBS), photonic slot routing (PSR), optical flow switching (OFS), optical burst switching (OBS), and optical packet switching (OPS) [MM12]. According to [MM13], however, there is little evidence that optical subwavelength switching techniques will become competitive with electronic switches in terms of energy consumption related to operation and heat dissipation. It was shown in [MM14] that photonic technologies are significantly more power hungry than CMOS and, except for very simple signal processing subsystems, CMOS will continue to be the most power efficient technology. In response to these issues, a promising solution toward practical OPS networks might be the use of mostly passive wavelength-routing components, e.g., athermal arrayed-waveguide grating, and replace fast optical packet switching with fast tuning lasers [MM15]. Despite the fact that small optical recirculation loops might be a viable solution for optical core routers [MM16], many open issues remain to be addressed with regard to power consumption and functionality of future optical switching equipment. Instead of mimicking their electronic packet-switching counterparts, research efforts should explore novel optical forwarding paradigms as the forwarding engine along with the power supply fans and blowers are the two most energy-consuming building blocks of today’s high-end electronic routers [MM11]. This is of particular importance since electronic routers might evolve from packet switching to flow switching devices [MM17] and the sum of all forms of video (TV, video on demand, Internet, and P2P) is expected to account for over 91% of global consumer traffic by 2014, whereby Internet video alone will account for 57% and 3D/HD Internet video will comprise 46% of all consumer Internet video traffic by 2014, respectively [MM18]. These video streams should be treated as flows rather than individual packets, which despite all the decades of research on advanced optical switching techniques somewhat ironically calls for the old-fashioned yet widely deployed optical circuit switching, which over the past has successfully shown to provide the desired network simplification and cost savings by reducing the number of required optical-electrical-optical (OEO) conversions by optical bypassing electronic core routers.

In summary, FTTH/B networks are poised to become the next major success story of optical networking technologies, whereby PONs will play a key role toward their evolutionary or revolutionary upgrade to NG-PONs by leveraging on a number of new enabling technologies, e.g., OCDM and OFDM. In alignment with the recently standardized EEE, PONs will be also at the heart of energy-efficient optical and bimodal fiber-wireless (FiWi) broadband access networks, which pose a number of new challenges and opportunities at the MAC layer, e.g., design and investigation of hybrid MAC protocols in R&F based FiWi networks as well as integration of optical burst assembly and wireless frame aggregation techniques. In optical metro and wide area
networks, research efforts should explore novel optical forwarding paradigms apart from OBS and OPS, which despite year and decades of research remain questionable to be widely deployed in real-world networks not only due to their technological complexity and increased power consumption but also due to the fact that the vast majority of global consumer traffic will be based on video streams.

2.2 Wireless Networks
The wireless communications world is rapidly and dramatically changing, and entering new and uncharted territory. Mobile data traffic is growing at an unprecedented rate well beyond the capacity of today’s 3G networks. Many researches [SC1, SC2] forecast that by 2014, an average mobile user will consume 7GB of traffic per month, which is 5.4 times more than today’s average user consumes per month, and the total mobile data traffic throughout the world will reach about 3.6 exabytes per month, 39 times increase from 2009.

The increasing number of services and users are generating tough challenges that need to be met by the research community. So far the physical layer research and development have been able to cater for the increasing appetite for capacity. However, the need to increase spectral efficiency and to consider overall energy consumption of the systems, combined with our closeness to Shannon limit, indicate that the networking community and “upper OSI-layers” research have to take a leading role to find research solutions for these needs and solve grand challenges in research. In Sections 2.2.1 and 2.2.2 we analyze two approaches to cope with bandwidth scarcity in wireless networks. Specifically, in Section 2.2.1 we discuss the cognitive network paradigm, while in Section 2.2.2 we discuss the spatial reuse of the spectrum by adopting smaller and smaller cells.

The use of wireless channels is highly affected by interference, thus to increase the efficiency of wireless communications we need to cope with channels’ interference. In Section 2.2.3 we present and discuss the research challenges for exploiting cooperative diversity to combat the fading, and thus increasing the channel efficiency.

2.2.1 Cognitive Networks
Cognitive radio networks and various other cooperative communications methods are proposed as approaches to tremendously increase spectral efficiency and to provide radically new and more efficient wireless access methods [CR1, CR2]. Cognitive radios (CR) have been so far mostly studied in the context of dynamic spectrum access (DSA). The introduction of DSA concept has produced an avalanche of new ideas on how to break the gridlock of spectrum scarcity. Particularly the USA thanks to the initiatives of FCC (Federal Communications Commission) and a former DARPA XG research program, has led the pact on reconsidering also spectrum regulations. Regardless of the advances, a lot needs to be still done even in the domain of DSA capable cognitive radios and networks. One of the challenges is to combine interdisciplinary research approaches in a fruitful and meaningful way. Traditionally the engineering and computer science community, particularly in the context of radio networks, has taken the regulation and business context as fixed boundary conditions. However, in order to speed up progress we need to consider implications and possibilities that regulatory changes can and should induce. This is a tremendous challenge since only few network and wireless researchers are educated or knowledgeable in regulatory issues and network economics. Nevertheless, one can argue that it is especially the cross section area of technology, regulation and economics that can provide new and powerful insights for research and future development. Thus one of the key problems for network researchers is that in order to advance we need to embrace interdisciplinary research topics. Moreover, we need to use and develop more detailed radio network models – but this needs to be done without losing a certain level of abstraction so that our models and tools stay general

2 By Petri Mähönen, (RWTH Aachen University, Germany)
This tension between the case specific, detailed modelling or simulations, and generalized abstractions is likely to increase in coming years.

Even larger challenge in cognitive radio networks is in the domain of self-organization and machine learning based optimization. It is sometimes forgotten that Mitola’s original vision on cognitive radios was not limited to dynamic spectrum access, but focused on general context sensitivity and environmental adaptation. Thus Mitola’s cognitive cycle shares many aspects with cooperative networking, autonomous networks, and adaptive radios [CR1, CR3]. Fulfilling this research vision has proven to be harder than expected. It is still not clear even what sort of network architecture would be best suited to support so called cognitive cycle. In fact, it is not even proven that machine learning methods in general can provide optimization gains that pay off the increased complexity of networks. It seems that new approaches are required to understand fundamental possibilities and limits in this area [CR4, CR5]. Combining methodologies and tools from networking, control theory, machine learning, and decision theory communities is most likely the best approach, but we need clearly stated research problems and a generation of researchers who are familiar on combining knowledge from different disciplines. In general, the communications research itself is in the danger of becoming more and more fragmented and our articles focus on increasingly narrower problems, but at the same time great breakthroughs seem to require more interdisciplinary approaches with great deal of knowledge and abstractions from different fields. Perhaps apart from research problems, we are also faced with educational and research methodology challenges.

One of the great challenges in the radio networks research is also a partial lack of hard data. For example, in the case of development of wireless networks and DSA systems, we are often forced to use anecdotal trend data from various white papers. At best we may have some small-scale measurement campaigns done by universities, which often do not even provide raw data to the community to use and validate the obtained results. One can argue that fundamental theoretical work does not need such data, but certainly more practically oriented research work, e.g., on DSA, scheduling and interference avoidance, become hardly tenable or at least we lack convincing justification for our research problems and conclusions. And it is sometimes hard to do some fundamental research unless you first know what the problem is. It seems that in the future we should require more measured facts instead of being content with visions and general motivational statements. There are some weak signals that this approach may become more dominant in the coming years as some research groups have recently started to share their data more openly or at least have based work on measurement campaigns.

One example in the need to combine knowledge from different fields and having measurement-based facts is radio network topologies. A lot of work was, and indeed is, done by network researchers by assuming extremely simplified propagation models for radio systems – so called disc model is still used far too often. This abstraction may be useful in some cases, but often it can lead to highly misleading quantitative – or even qualitative – results. Similarly we should pay more attention to understand the topology of radio networks. So far most of our theoretical and simulation methods are based on assumption that radio nodes are randomly (Poisson) distributed over the operating area. Recent measurements and general arguments show that this is hardly true in all spatial scales. Thus we need to reconsider our models and previous conclusions that might have been drawn under such assumptions [CR6, CR7]. More generally, it is still not clear how we should abstract various propagation and network topology issues to be analytically tractable and reasonable to simulate without losing quantitative and qualitative prediction power of the models.

Certainly the largest research challenge is the management of the increasing complexity of systems. This challenge is evident both in practical design of systems, and theoretical modelling of increasingly heterogeneous networks. The same pain is also shared by industrial R&D departments on standardization and development of actual deployments. It seems that as the complexity increases, we have to be careful especially with introduced non-linearity and
parameter sensitivity at the system level. Cross-layer design has been often proposed as a solution for many optimization problems and it is strongly linked also to cognitive networking paradigm. However, as this approach can also lead to increased non-linearity and complexity one should heed the recent warning by Kawadia and Kumar that cross-layer design is not always the most efficient overall approach [CR8]. One of the great research challenges is to find out where the sweet spot for cross-layer design is, and what would be an ideal network architecture that allows both simplicity and cross-layer design.

Finally, it may be useful to end this musing with some less philosophical and more specific topics on the future research opportunities and challenges. First more practical item to emphasise is the challenge to provide better experimental research tools. The academic research community used to lead development of new system concepts. Lately an increasing amount of systems are beyond the reach of academic research and can be handled only by few large and well-funded companies. One needs just to think about the development of routers, wireless access systems etc. However, it does not need to be so. Innovative approaches to develop open research platforms and having well defined interfaces so that efforts can be shared could provide again capability for academic groups to experiment and if not to lead, at least to contribute. Recently introduced open platforms such as OpenFlow (Stanford), WARP-boards (Rice) and USRP/gnuRadio are proof that a lot can be done if the talented people and the community put their forces behind common research platforms. Where we need research is on how to develop open APIs and interfaces to handle different hardware and protocol stacks that ensure that development effort is minimized and research opportunities maximized [CR9]. There has been some effort to develop open interfaces and definition methods, but research work in this domain is still lacking momentum.

Another and final research challenge to mention in this article is to predict that research in Medium Access Control layer and scheduling in general is likely to have a renaissance. As wireless access is becoming dominant method to access data, and especially if DSA and CR networking approaches will become mainstream, it is clear that we have to revisit fundamental concepts of, and design approaches for MAC-layers and scheduling. MAC-layers should become most likely more aware of underlying network topology and physical layer limitations. In the context of cognitive radios, the MAC layer should often be able to distinguish between primary and secondary users. Many of these concepts have not been considered to be a part of MAC-layer knowledge, and thus there remains a lot to be done. Scheduling itself is one of the approaches that could provide rapid efficiency gains in heterogeneous networks. For example, selecting scheduling priorities so that different applications and heterogeneous networks are taken into account is likely to provide a lot of research challenges for the next decade.

The next decade for networking research looks indeed promising. The increasing amount of data communications and emergence of new networks including machine-to-machine communications are generating a demand for new results to ensure increased efficiency. We will have no lack of challenges both in theoretical and practical domains. The real challenge is to have the community well educated and ready, and fight against fragmentation of our knowledge and keeping our focus on worthy big problems.

2.2.2 Spectrum Spatial Reuse

There are several approaches to meet the wireless networks explosive traffic growth, one of which is upgrading today’s 3G networks to a next-generation cellular network with enhanced PHY (Physical layer) technology. However, the enhancement of PHY technologies approaches its theoretical limit and may not scale well with the explosive growth rate of mobile data traffic. According to Cooper’s law shown in Fig. 1, the number of voice calls carried over radio spectrum has been increased by a million times since 1950, and Cooper also predicted that this would continue for the foreseeable future [SC3]. Of that million times improvement, roughly 25 times

---

3 By Song Chong (KAIST, Korea)
was from using more spectrum, 5 times was from using frequency division, and another 5 times was from the enhancement of PHY technologies. But the lion’s share of the improvement – a factor of about 2700 – Cooper suggested was the result of spatial reuse of spectrum in smaller and smaller cells. Cooper’s law tells us that despite being close to the Shannon limit, there is no end for practical increases in wireless capacity if we are prepared to invest in an appropriately dense infrastructure.

![Cooper’s law compared to observed trends.](image)

**Fig. 1**

The small-cell gain, however, comes at a high cost. As the infrastructure becomes denser with the addition of smaller cells, inter-cell interference (ICI) inevitably becomes higher and more complex to manage. Thus, a key technical challenge in scaling wireless capacity by increasing the density of cells is how to effectively manage ICI in such a complex cellular deployment. Another important technical requirement for small-cell networks is self-*x* capability of cells where *x* includes configuration, optimization, diagnosis, healing etc., since small-cell base stations would be less reliable and in many cases their deployment/removal and on/off would be done by individual subscribers in an ad hoc manner, not by operators in a pre-planned manner. The self-*x* capability is an enabler for fully distributed autonomous network management that can realize the small cell gain without suffering much from exponentially growing complexity of network management.

ICI management problem can be tackled at PHY layer and also at upper layers such as MAC, routing, transport layers. Techniques such as Successive Interference Cancellation, Interference Alignment are the examples of PHY-layer ICI management techniques. Mathematically, ICI management problem at upper layers and self-*x* problem can be tackled in the light of stochastic network utility maximization (NUM) problem with queue stability constraint [SC4-SC6], which is generally given by

$$\max_{R \in \Lambda} \sum_s U_s(R_s)$$  \hspace{1cm} (1)$$

where $R = [R_s]$ is the vector of long-term average rates $R_s$ of all users $s$ in the network, $\Lambda$ is the unknown long-term rate region of the network that can be shown to be always convex if the randomness in wireless channels has a finite set of states and the sequence of states forms an irreducible Markov chain with stationary distribution, and $U_s$ is a concave utility function of user $s$. Assume that exogenous arrivals to the network follow a stochastic process with finite mean and each wireless link is equipped with a transmission queue.
It is known that the above NUM problem can be asymptotically solved by solving the following MAC-layer problem in (3) in conjunction with the transport-layer algorithm in (2) (here we assume that route is fixed for all flows for simplicity) [SC4-SC6]. At time $t$, each source $s$ independently determines its instantaneous data rate $r_{s}(t)$ by

$$r_{s}(t)=U_{s}^{'s}\left(\sum_{l\in L(s)}q_{l}(t)\right)$$

where $L(s)$ is the set of links on the route of flow $s$, $U_{s}'$ is the first derivative of $U_{s}$, and $\sum_{l\in L(s)}q_{l}(t)$ is the end-to-end queue length of flow $s$ at time $t$. Note that this form of source congestion control can be easily implemented at transport layer in a fully distributed manner by the help of end-to-end signalling to carry queue length information to the source. In fact, the necessity of end-to-end signalling can be removed without losing optimality if each flow has separate queue at every link.

The key technical challenge lies in the MAC-layer problem, expressed by the following network-wide weighted sum rate maximization problem

$$\max_{P} \sum_{s}q_{s}(t) \cdot r_{s}(t,P)$$

where $q_{l}(t)$ is the queue length of link $l$ at time $t$, $P=[P_l]$ is the vector of power allocations of all links in the network and $r_{l}(t,P)$ is the achievable rate of link $l$ at the power allocation $P$ given network-wide channel state at time $t$. This problem is indeed a core problem that arises in any wireless networking problem, for instance, ICI management problem in a cellular network, modelled by a stochastic NUM problem, is nothing but finding $P$ repeatedly at each time $t$ from (3). Note, however, that the problem not only requires centralized computation using global information but it is also computationally very expensive. As an illustrative example, consider an important special case of the problem that each $P_l$ can take either 0 or its maximum value and, furthermore, the choice of $P$ is restricted not to activate any two interfering links simultaneously for conflict-free transmission. Then, the original problem is reduced to so called max-weight scheduling problem [SC7, SC8] that is a central research theme of multi-hop wireless networking research community. The max-weight scheduling problem is a NP-hard problem since it involves a weighted maximum independent set problem of NP-hard complexity. As another example, consider ICI management problem in a multi-carrier, multi-cell network [SC9, SC10]. The corresponding MAC-layer problem turns out to be a centralized joint optimization problem of user scheduling and power allocation, which is computationally very expensive since the user scheduling involves multiple NP-hard problems, and the power allocation involves nonconvex optimization.

There have been several works to find low-complexity, distributed algorithms for the max-weight scheduling problem and the dynamic ICI management problem. In [SC8], a randomized algorithm, called as pick-and-compare algorithm, has been proposed. The algorithm asymptotically solves the max-weight scheduling problem with linear complexity but the reduction of complexity comes at the cost of slow convergence and increased delay. In [SC11, SC12], distributed maximal/greedy scheduling algorithms have been studied but they yield approximate schedules losing throughput optimality. Recently, in [SC13-SC15] it is shown that CSMA algorithms can asymptotically solve the max-weight scheduling problem if the product of back-off time and packet transmission time is adjusted as an exponential function of the weight $q_{l}(t)$ and the first prototype implementation on a real 802.11 hardware has been reported in [SC16]. Nevertheless, finding and prototyping low-complexity, distributed max-weight scheduling algorithms is still an open problem that has many issues to be resolved, one of which is delay issue. The max-weight scheduling intrinsically suffers from large delay incurred by queue build-up and thus how to reduce delay while minimizing loss in throughput optimality is one of the top priority research issues. On the other hand, research on low-complexity, distributed algorithms for
the ICI management problem has received relatively less attention from networking community and there are only a few notable works [SC9, SC10] available in the literature. In [SC9], a concept of reference user has been introduced to decentralize the network-wide optimization and to lessen the involved computational complexity but the algorithm cannot guarantee throughput optimality. Proof of concept through prototype implementation, for instance, prototyping and evaluation on real 802.11 hardware, is also an important research direction in this area. The key question to be answered there is how much capacity gain one can actually achieve by adding low-complexity, fully distributed ICI management functionality in a massively and arbitrarily deployed WiFi access points environment.

In summary, the MAC-layer problem in (3) is a core problem that inevitably arises and needs to be solved in any wireless network whose objective is to maximize network-wide sum utility. ICI management in small-cell networks is an important special case of the general problem. Development and experimental validation of low-cost, fully distributed algorithms for the problem is a very challenging research issue and the key step to realize self-x small-cell networks that are believed to be the most effective way to scale wireless capacity continuously without known limit. The theory suggests that source congestion control to be in the form of (2) but in reality TCP does source congestion control. Other interesting questions are how TCP interacts with the MAC-layer problem and what modification is necessary for TCP.

Network greening is a rapidly emerging research area. From a radio resource control point of view, network greening is in a loose sense a dual problem of the stochastic NUM. Maximizing network capacity for a given power budget is reciprocal to minimizing power consumption for a given capacity requirement. Thus, study on small-cell networks from a network greening point of view would be another important research direction.

2.2.3 Cooperative diversity

One of the distinguishing features of a wireless communication system is the stochastic nature of the wireless channel. It is mainly due to the basic property of multi-path propagation of electromagnetic waves, causing a transmitted wave to interfere with copies of itself, arriving over different paths of different lengths at the receiver. As soon as there is movement – of the sender, the receiver, or even just of nearby objects – this interference situation will change, possibly rapidly and in unforeseeable ways. This is called fast fading and one of the defining characteristics of wireless transmission [HK1]. The main means to combat fast fading is to use different communication resources, formalized as a channel (not to be confused with a particular frequency band of the wireless spectrum). Channels can be orthogonal to each other in that they do not influence each other; if the transmission qualities of such orthogonal channels are stochastically independent, distributing transmissions over several such channels reduces the chances of communication outage. If done properly, this reduction is exponential in the number of used channels, providing a considerable gain through the use of these diverse channels – hence the common term diversity gain.

Typical examples for orthogonal channels are different time slots, different frequency bands spaced sufficiently far apart, or different codes in a code-multiplexing scheme. Also, different propagation paths can be used as orthogonal channels, as demonstrated by multi-antenna, multi-input/multi-output (MIMO) schemes. In the absence of multiple antennas, multiple users with single antennas can cooperate to create a similar situation; it is usually called cooperative diversity -- introduced in [HK2], while a survey can be found, e.g., in reference [HK3]. Despite being superficially similar to multi-hop networks, it is substantially different in the way communication channels are used and in the way a receiver processes received signals. Alternatively, channels can be allowed to interact (e.g., overlapping transmissions from different base stations at the same time) and still be useful to reduce fading; these schemes, however,

4 By Holger Karl (Universität Paderbon, Germany)
typically require careful control (compare the various forms of coordinated multipoint transmission in LTE-Advanced; for a brief survey, see reference [HK4] and references therein). Nonetheless, even in such non-orthogonal channels, similar options as in orthogonal channels exist.

Cooperative diversity, be it over orthogonal or non-orthogonal channels, has received considerable attention in the last few years, but mostly from wirelessly oriented researchers and mostly for cellular systems. The integration with protocol stacks is on going, but there is still a lot of work to do here. The following paragraphs highlight some areas that are still in need of research and new ideas.

**Energy-efficient MAC protocols for cooperation diversity.** Cooperative diversity requires, in its simplest form, the cooperation of three nodes, commonly referred to as a source, an assisting relay, and a destination. To actually exploit diversity effects (and not just to build an inferior multi-hop system), all three nodes must be awake and listening to the channel at the same time. This might be a non-issue for cellular systems, but to use cooperative diversity in ad hoc, mesh, or sensor networks, the integration with sleeping cycles is necessary. It is pointless to create a cooperative diversity system when it cannot be assured by the MAC protocol that all relevant nodes are awake at the right point in time. And this is indeed a tougher problem than it sounds like, as waking up a node (be it sender- or receiver-initiated) requires some form of communication, but diversity systems are intended for the situation when there is no reliable link in the first place. Hence, to communicate, we need to wake up those nodes with which we cannot really communicate – a challenging catch for which only some initial work exists so far [HK5].

**Extend cooperation diversity to other communication primitives.** Most cooperative diversity protocols suffer substantially from multiplexing loss, i.e., the need to use orthogonal channels. But in some settings, this not really a loss but has to happen anyway – the prime example is a broadcast into a wireless network. Here, the same data has to be transmitted by many nodes anyway. Hence, these repeated transmissions can also be exploited to realize diversity gains. Only some few first results are known so far, showing that in general this problem (and variations of it) is NP complete [HK6].

**Integrate cooperation diversity with multi-user diversity techniques.** Is it possible to combine cooperative diversity techniques with existing multi-user diversity techniques? A typical multi-user diversity technique is OFDMA: in a cellular setting, the subchannels of an OFDM system are allocated to different users, e.g., to maximize system capacity by choosing that user for a subcarrier that has the best link; this scheme essentially exploits frequency diversity across multiple users. One appealing approach would be to combine cooperation diversity into this scheme by also choosing subcarriers based on how data can be cooperatively forwarded. Some first results exist [HK7, HK8], but practically, a number of questions are still open. For example, how should this be controlled, what are the maximum possible capacity gains?

**Dealing with limited, outdated channel state information.** How should the relay selection process really work? A lot of analytical work is available, but practical schemes that do consider the actual signalling overhead and limited validity of channel state information are still rare [HK9]. How is the tradeoff between source-based selection vs. relay-based or destination-based selection, using channel state information explicitly or relying on opportunistic schemes, what are the resulting robustness properties?

**Cross-layer aspects.** Is it possible to integrate cooperative diversity techniques even with application layer techniques? As the source/channel separation theorem fails, the question of looking at the source and channel jointly comes up [HK10]. Various techniques could be
considered – for example, network coding has been shown to provide benefits in content distribution applications. When applying this to a mesh network, what is the relationship to diversity gains available via the wireless channel?

**Cooperative diversity in non-wireless settings.** Last in this list, a wild speculation. Cooperative diversity is currently perceived as a wireless-only technique, resting on the random nature of the wireless channels. But even fixed-network channels (rather thought of as a routing path) randomly fluctuate, and fixed networks also possess path diversity. Is it possible and profitable to apply these techniques to wired networks? At which timescales should this happen?

Overall, cooperation diversity is a very powerful tool, but a tool that must be used wisely and with proper consideration of the present scenario, the communication primitive, the user data, and the acceptable trade-offs. There still seems a considerable amount of work before cooperation diversity schemes will be used as a matter of course in all kinds of wireless systems.

3. Internet Architecture and Protocols

The evolution of the Internet is of utmost importance to our economy and our society just because it has been playing a central and crucial role as the main enabler of our digital era. However, the Internet is also a victim of its own success as it should remain stable and robust and therefore develop a natural resistance to revolutions. This is a reason why the main innovation currently comes from the edge with the explosion of wireless technologies and overlay architectures. However under the push of novel services also the Internet structure is changing. Specifically, as discussed in Section 3.1, the emergence of content distribution networks is driving toward a content-centric Internet. In Section 3.1 we discuss how this evolution is changing the way Internet is structured and the associated research challenges. A long-term view about the (re)evolution of the Internet architecture is then discussed in Section 3.2 where we discuss two key aspects of the Future Internet architecture: virtualization and federation.

3.1 A Content-centric Internet

Today’s Internet [SU1] differs significantly from the one that is described in popular textbooks [SU2], [SU3], [SU4]. The early commercial Internet had a strongly hierarchical structure, with large transit Internet Service Providers (ISPs) providing global connectivity to a multitude of national and regional ISPs [SU5]. Most of the content was delivered by client-server applications that were largely centralized. With the recent advent of large-scale content distribution networks (CDNs), e.g., Akamai, Youtube, Yahoo, Limelight, and One Click Hosters (OCHs), e.g., Rapidshare, MegaUpload, the way the Internet is structured and traffic is delivered has fundamentally changed [SU1].

Today, a few “hyper-giants”, i.e., CDNs and OCHs, often have direct peerings with large ISPs or are even co-located within ISPs and rely on massively distributed architectures based on data centers to deliver their content to the users. Therefore, the Internet structure is not as strongly hierarchical as it used to be [SU1].

These fundamental changes in content delivery and Internet structure have deep implications on how the Internet will look like in the future. Hereafter, we describe how we believe that three different aspects of the Internet may lead to significant changes in the way we need to think about the forces that shape the flow of traffic in the Internet. Specifically, we first describe how central DNS has become as the battlefield between content providers and ISPs. Next, we discuss how split architectures may change the ability of many stakeholders to influence the path that the traffic belonging to specific flows will follow across the network infrastructure. Finally, we discuss how

---

5 By Steve Uhlig (TU Berlin/DT Labs, Germany)
the distributed nature of current content delivery networks will, together with changes within the forwarding/routing, enable much more complex handling of the traffic, on a much finer granularity compared to the current Internet.

**DNS and Content Redirection.** The Domain Name System (DNS) was originally intended to provide a naming service, i.e., one-to-one mappings between a domain name and an IP address. Since then, DNS has evolved into a highly scalable system that fulfills the very stringent needs of applications in terms of its responsiveness [SU6], [SU7], [SU8]. Note that the scalability of the DNS system stems from the heavy use of caching by DNS resolvers [SU9].

Today, the DNS system is a commodity infrastructure that allows applications to map individual users to specific content. This behavior diverges from the original purpose of deploying DNS [SU10]. Given the importance of DNS for end-user experience and how much the DNS system has changed over the last decade, understanding how DNS is being deployed and used both by ISPs and CDNs is critical to understand the global flow of traffic in today’s Internet.

For example, recent DNS measurements of DNS resolvers’ performance [SU11] have shown that the DNS deployment of commercial ISPs sometimes leads to poor DNS latency.

Different third-party resolvers, e.g., GoogleDNS or OpenDNS, do not perform particularly better in terms of responsiveness compared to ISPs resolvers. A key aspect of DNS resolvers is not only latency, but also how well they represent the end-host for which they do the DNS resolution. Third-party DNS resolvers do not manage to redirect the users towards content available within the ISP, contrary to the local DNS ones.

While more work is necessary to pinpoint the exact reasons for this behavior, we strongly expect that the explanation has to do with the fact that third-party DNS resolvers are typically outside ISPs and cannot indicate the IP of the original end-host that originates the DNS query [SU12]. The current advantage of DNS resolvers inside the ISP of the end-host is their ability to represent the end-user in terms of geographic location and its vicinity to content.

**Opening the Network Infrastructure.** Content is not the only place where an Internet (r)evolution is taking place. Thanks to a maturing market that is now close to “carrier grade” [SU13], [SU14], [SU15], [SU16], [SU17], the deployment of open source based routers has significantly increased during the last few years. While these devices are not competing with commercial high-end switches and routers available with respect to reliability, availability and density, they are fit to address specialized tasks within enterprise and ISP networks. Even PC-based routers with open source routing software are evolving fast enough to foresee their use outside research and academic environments [SU18], [SU19], [SU20].

The success of open-source routing software is being paralleled with increasing virtualization, not only on the server side, but also inside network devices. Server virtualization is now followed by network virtualization, which is made possible thanks to software-defined networking, e.g., OpenFlow [SU21] that expose the data path logic to the outside world. The model of network devices controlled by proprietary software tied to specific hardware will slowly but surely be made obsolete. Innovation within the network infrastructure will then be possible. A decade ago, IP packets were strictly following the paths decided by routing protocols. Tomorrow, together with the paths chosen by traditional routing protocols, a wide range of possibilities will arise to customize not only the path followed by specific traffic, but also the processing that this traffic undergoes. Indeed, specific actions that are statically performed today by specialized middleboxes placed inside the network, e.g., NAT, encryption, DPI, will be implemented on-path if processing capabilities happen to exist, otherwise the traffic will be dynamically redirected to close-by computational resources. This opens a wide range of applications that could be implemented almost anywhere inside the network infrastructure.
Towards a New Business Model for the Internet. As content is moving closer to the end-user for improved quality of experience and the infrastructure opens up to unprecedented control and flexibility, the old business model of hierarchical providers and customer-provider relationships is hardly viable. Nowadays, content delivery is a very profitable business while on the other side infrastructure providers struggle to provide the necessary network bandwidth for hungry multimedia applications at reasonable costs. The consequence of more and more limited ISP profit margins is a battle between content providers and the network infrastructure to gain control of the traffic.

This battle stems from fundamental differences in the business model of content delivery networks and ISPs. Today, the operators of content delivery networks, for example through DNS tweaking, decide about the flow of the traffic by properly selecting the server from which a given user fetches some content [SU8], [SU22], [SU23]. This makes content delivery extremely dynamic and adaptive. On the ISP side, most of the traffic engineering relies on changing the routing configuration [SU24], [SU25], [SU26]. Tweaking existing routing protocols is not only dangerous, due to the danger of mis-configurations [SU27], routing instabilities [SU28] and convergence problems [SU29], [SU30], but is simply not adequate to choose paths at the granularity of content. ISPs need therefore new mechanisms to regain control of the traffic.

This can be achieved for example by exploiting the diversity in content location to ensure that their network engineering is not made obsolete by content provider decisions [SU31]. Another possibility is to leverage the flexibility in network virtualization and making their infrastructure much more adaptive than today’s static provisioning [SU32].

The deep changes we discussed in this section create unprecedented opportunities for researchers to propose and evaluate new solutions that will address not only relevant operational challenges, but also potentially business-critical ones. The ossification of the Internet protocols does not mean that the Internet is not evolving. The Internet has changed enormously over the last decade, and will continue to do so. What we observe today is simply a convergence of content and infrastructure that questions a model of the Internet that is not appropriate anymore. Content is not just king in the Internet, it is the emperor that will rule all its subjects.

We believe that the three research areas above need critical input from the community in order to enable a truly content-centric Internet. First, even after more than two decades of deployment and evolution, the DNS is still poorly understood. The DNS is much more than a naming system: today it is a central point in the content distribution arena. The way DNS resolvers are used and deployed is a rather open field, which might lead to significant improvements in flexibility and performance for content and application providers, ISPs, as well as end-users. Second, software-defined networking opens a wide range of possibilities that would transform the current dumb pipes of the Internet core into a flexible and versatile infrastructure. For the first time, researchers are able to inject intelligence inside the network. Finally, as content is moving closer to the end-user, the very structure of the Internet is reshaped. This leads to fundamental questions about the possible directions in which the Internet might be going, not only at a technical level, but also from a business perspective.

3.2 Federation and Virtualization in the Future Internet

In the Future Internet, we foresee various concurrent networks being deployed and customized to provide their specific service. An increased diversity and functionality of the networks and their components require a revision of the Internet architecture to support their interoperability and continuous deployment. We can see some potential developments such as:

---

6 By Serge Fdida (Universite Pierre et Marie Curie, France)
- The emergence of virtual worlds, sensing environments, interactions between the physical and virtual worlds.
- Networked systems, embedded systems, vehicular communications
- Developments in Digital life with all related applications and usage to assist the well-being of citizens

The above examples illustrate some different shapes and requirements that the network could take in the future. Each of these evolutions is addressing a given environment where the objects and constraints are quite particular. The objective to design an architecture with an Hourglass model will come at cost to accommodate the diversity of its numerous components. At which level should the interconnectivity be provided? What is the definition of a managed network? How can we support interconnectivity? Should we embed economics in the protocol design from scratch? What are the incentives to share and how they can be evaluated? Is there a reasonable transition methodology and scenario?

We claim that the future Internet will therefore be polymorphic to allow several networking environments, each with their own features and strengths, to be deployed and coexist on a permanent basis. We expect that virtualization and federation are the pillars of such architecture.

At the end, the network at large should be seen as a global shared resource, which is virtualized and made accessible at scale. Virtualization is therefore a strategic component to accommodate different instances of the network into a single framework. In other words, virtualization is an enabler of diversity. In the extreme case, virtualization could contribute this way in the development of parallel global networks run by certain big players. An opposite force is required to ensure global connectivity and competition: the glue to manage and secure the Polymorphic Internet will be provided by the Federation principle. Federation could be horizontal or vertical involving different levels of cooperation between independent organizations. ISP interconnection is a typical example of a nowadays federation implemented through a set of bilateral agreements.

A subsequent question is related to the ability to interoperate virtualized infrastructures supporting heterogeneous protocols and services. The goal is to achieve increased coverage at different layers and/or enable resource sharing between independent domains. Federation is more than interconnection. It covers API, policies, governance, trust and economics. Of course, interoperability should be achieved at different levels, such as naming, service discovery or resource management.

Federation will govern the interoperability of independent networks managed by a given authority. Alike the domain concept in the current Internet, a similar environment will be defined as it is unlikely that a single entity will deploy alone the concurrent networking environments mentioned above. A domain is considered as an independent set of resources providing services managed by a trusted administrative authority. Therefore, a domain has a value by itself but will also often benefit for being associated with other domains in order to achieve scale or heterogeneity. Users of a domain have access to its resources but would, in certain circumstances, benefit from accessing services offered by other domains. The governance of the global shared resources provided by the federation of domains is therefore distributed. It requires local policies to control local resource access but also external policies to grant access to external users. Different federation architectures can be considered, ranging from bilateral agreements to more scalable peering models. Key issues are to enforce a federation model that supports incentives mechanisms for sharing and rewarding policies that favour access to resources and services. Cloud federation can be seen as an example of such evolution. A consequence is that we might want to extend the waist of the reference model, introducing the concept of federation instead of a homogeneous abstraction. From an economic point of view one needs to understand under which assumptions federation is beneficial for the involved parties and the network as whole and what types of federation agreements could help the system reach the desirable equilibrium. Federation economics will have to address the heterogeneity and polymorphism of the network domains.
involved and their complex interactions. How to compare and value multi-dimensional resources, to what extent the future Internet economy should be regulated or be designed as a free market in order to achieve a globally efficient allocation and provision of resources are only some of the questions that need to be answered.

4. Ubiquitous Computing Services  

The Mobile Phone (MP for short) is the key device in accessing *Ubiquitous Computing Services* (UCS), i.e., a variety of emerging ubiquitous multimedia and data services, including mobile cloud computing services. UCS access requires integrated use of wireless, mobile and Internet networks for stable and secure transmissions. To realize MPs access to UCS, the first step is to look at how the different available technologies will integrate and work with each other. A special attention must be deserved at using off-the-shelf mobile phones to accessing UCS through today’s integrated and heterogeneous wired and wireless networks. Specifically, from this analysis it emerges that several fundamental challenges need to be addressed:

**System challenges**: MPs usually have small display, limited size memory/storage, limited processing power, etc. On the other hand, applications usually require high processing, large memory space, and large screen display. In addition, among the system challenges, the limited battery lifetime (also referred to as *energy challenge*) represents one of the most critical constraint.

**Communication challenges**: A stable connection and high-quality communications are required for accessing UCS. However, stable/secure connections with sufficient bandwidth cannot be always guaranteed to MPs, especially to high-speed moving MPs. Indeed, MPs often has to cope with shortage of bandwidth, frequent disconnections, and fluctuating wireless channels, etc.

**Security challenges**: MPs are vulnerable to various security threatens with can affect the communication links, the data access, the storage, etc.

In the following we will focus on *System* and *Communication* challenges, while *Security* challenges are discussed in the next section.

**System Challenges.** The current trend is to use MPs to access richer ubiquitous services than simple phone calls. In addition, mobile cloud computing is promising to bring a lot of new and rich applications to MPs. However MP hardware and software constraints highly limit this evolution. Indeed, today there are about 3.5 billion MPs worldwide that are low-end phones which, for using mobile computing services, need to get over their system gaps such as inadequate computational capability, lack of storage, unstable and slow communication links, and above all the energy constraints (the extra computation and networking activities required for accessing advanced UCS, consume the battery power at the speed exceeding that MPs are designed for.) Therefore, bringing high-end cloud computing services to low end MPs is not an easy task. A key challenge is the code portability across millions of off-the-shelf low-end MPs. Currently we have a broad range of hardware platforms and operating systems, making the interchange of data and applications between devices difficult. Currently, most MPs can only support a few built-in applications shipped with the MP itself. This prevents developers and users to add more advanced applications to the MPs. Therefore, making low end MPs programmable and enabling code portability is a crucial incentive for attracting a broader set of developers to provide state-of-the-art applications for those platforms. Virtual machines constitute a good solution for code mobility,
providing a virtualized processor architecture that is implemented over MP architectures, which allows installing and running extra applications over closed systems. However, virtual machines do not solve the resource-constraint problems because they contend for the MP limited computational resource. Therefore the question is “how to exploit virtual machines without further reducing the limited MP resources.” Forwarding a part of computation to resource-rich and cost-effective clouds could be a good way to reduce low end MPs computation burden. Some research works already exist about offloading the computation tasks to the cloud [WJ1] [WJ2], but they are mainly applicable to smartphones. On the other hand, overcoming system and energy gaps of low-end devices requires novel design approaches. In particular the energy gap is a critical one, as offloading the computations to the cloud requires additional energy for the increased computing a networking activities. It is widely known that energy is limited for a large majority of mobile devices. For a typical low end MPs, assuming reasonable voice talk and very little web-browsing time, a fully charged battery is able to satisfy the MP energy requirements for several days. However, adding extra functionalities to the MPs will probably increase the users’ connection time resulting in shorter battery lifetime. As a consequence, people will need to recharge their MPs more often than as usual. Shortening battery life is very undesirable and annoying to users, particularly when there is no recharge station nearby. Therefore, energy aware features should be carefully addressed.

Some may argue that the low-end phone's capability will soon grow up to achieve the same capabilities of current high end MPs (Moore's Law), and hence solutions focusing on low end devices are unnecessary. However, we argue that though hardware is growing according Moore's Law, software requirements are also growing. Moreover, from the economic point of view, in some world regions (e.g., developing or third world countries) people only look for affordable MPs, i.e. the cheaper the better. Moreover, as for the energy saving, we believe this will continue to be a major research challenge for battery driven mobile devices.

Communication challenges. Accessing in a cost efficient way richer multimedia UCS calls for novel networking approaches for an effective usage of the available heterogeneous communication technologies. To clarify this aspect, let’s use the following example. Some MP users on a bus are lunching bandwidth hunger applications such as video streaming and surveillance applications. To implement such applications they can use the cellular network that, however, can be very expensive, charged by airtime or traffic volume. On the other hand, there may be WiFi WLAN networks nearby with high bandwidth available (e.g., 54Mbps) and unlimited free access. However, even with the available WLAN bandwidth from a nearby public access point, the MPs might be unable to access UCS as: (1) the MPs are unable to access to the heterogeneous networks; or (2) they are unable to access the right services because the MPs do not have necessary access support; or, even when the applications/services can be accessed, (3) the services may be disconnected frequently, resulting in an unacceptable quality of services.

The above challenges can be addressed by aggregating/bundling the available heterogeneous wireless links by converging into a unique ubiquitous networking the modern 3G networks (e.g., W-CDMA, TD-SCDMA, CDMA2000, HSPA), the WLANs and the Internet. To achieve the heterogeneous wireless link bandwidth aggregation, three grand challenges must be addressed: (1) the link-interface heterogeneity -- end-users need to access different types of mobile links; (2) the link-communication interruption due to end-user mobility, unstable radios, and limited coverage; and (3) the link-access vulnerability -- mobile links are highly vulnerable to attacks. To summarize, we need to investigate novel algorithms and protocols for providing MPs with secure, stable and cost-effective bandwidth. For example, data channels can be aggregated on demand or adaptively. How the aggregated links will be managed for downloading and uploading transmissions is a very intriguing challenge.
In the mobile access to UCSs, special challenges occur in accessing Internet for passengers of large/long size vehicles, such as long distance trains [WJ3], fleets or cruise on maritime communications [WJ4]. In these cases the users often suffer from annoying service deterioration due to fickle wireless environment. For example, consider a chained wireless access gateway on a train which consists of a group of interlinked routers with a wireless connectivity to the Internet; the protocol handling the mobile chain system should exploit the spatial diversity of wireless signals to improve the Internet access as the routers do not measure the same level of radio signal. Specifically, a high-speed train can be viewed as a virtual “gateway” long several hundred meters that spans across the train and seeking for the best signal quality. An intelligent protocol will re-route the traffic toward the routers experiencing in that point the best quality signal. This kind of research has to tackle two fundamental issues: 1) to reduce average temporary communication blackout (i.e. no Internet connection), and 2) to enhance the aggregate throughput the system.

In the analysis of the communication challenges, performance evaluation studies of the wireless network QoS (e.g., 3G+ and LTE standards) constitute another fundamental research area. The aim is to study the performance of heterogeneous mobile/wireless networks and to investigate the effectiveness of the novel communication strategies built on top of them, with special attention to the impact of user mobility on the performance of wireless networks. Such study, therefore, will require extensive investigations of all possible mobile scenarios in urban areas, including subways, trains, offshore ferries and city buses. [WJ5].

Before concluding this section, it is worth noting that MPs’ cooperation can be exploited for tackling the system and communication challenges. G. Shen et al. [WJ6] envision a new better-together mobile application paradigm where multiple mobile devices are placed in a close proximity and study a specific together-viewing video application in which a higher resolution video is played back across screens of two mobile devices placed side by side. K. A. Li et al. [WJ7] design a buddy proximity application for mobile phones, in which mobile phones can be useful agents for their owners by detecting and reporting situations that are of interest. SmartSiren [WJ8] is a collaborative virus detection and alert system for smartphones. In order to detect viruses, SmartSiren collects the communication activity information from the Smartphones, and performs joint analysis to detect both single-device and system-wide abnormal behaviours. Multiple-party video conferencing and online gaming also help to stimulate cooperation and collaboration among 3G phones. A kind of online game targeting at augmented reality [WJ9] is developed to allow simultaneous connection and game participation from many different users. Furthermore, cell phones with GPS component are utilized to help cell phones without GPS to locate themselves [WJ10].

5. Cross-Layer Issues

This section focuses on two research areas that affect several layers of the protocol stack: energy efficiency and security. Specifically, in Section 5.1 we present and discuss the research topics emerging when we include the energy efficiency as a major constrain in the Internet design (this is also referred to a Green Internet or Green Networking), while Section 5.2 is devoted to discuss the research challenges that are emerging when we consider the security requirements both at the data and system level.

5.1 Green Internet

For half a century the research field of computer communications, has contributed to the design, and optimization of computer and telecommunications networks, wireline and wireless, with the aim to meet quality of service requirements at minimal cost. Although in wireless

---

8 By Moshe Zukerman (City University of Hong Kong, Hong Kong)
communications, power control and optimization has always been an important consideration, as excessive power by one user may interfere with the reception of another, the cost of energy in wireline networks has not been a key consideration in the traditional teletraffic research. This is changing now. There is an increasing recognition in the importance of energy conservation on the Internet because of the realization that the exponential growth of energy consumption that follows the exponential increase in the carried data is not sustainable. One example that signifies this realization is the GreenTouch™ [MZ1] consortium with membership that includes, among many others, major players such as Bell Labs, Huawei, and Samsung, in which “industry leaders and diverse global talents come together to create an energy efficient Internet through an open approach to knowledge sharing”. The consortium is “dedicated to creating a sustainable Internet through innovation and collaboration — increasing ICT energy efficiency by a factor of 1000” within five years “to fundamentally transform global communications and data networks.” This ambitious goal of three orders of magnitude efficiency improvement is justified by an analysis reported in [MZ1] that indicates a potential of four order of magnitude reduction. This is consistent with the analysis of Tucker [MZ2] that evaluates the minimum energy requirement to be lower by over three orders of magnitude of what the Internet consumes today.

Although much of the improvement will be achieved by research in areas beyond the scope of this journal, such as transmission and circuit design, there are many opportunities to contribute to improve energy efficiency through better network architectures, new protocols and traffic engineering. There is no reason, for example, that a large chunk of data, e.g. a movie download, is chopped into many small IP packets each of which is treated individually (e.g., performing table lookup and queue management), in various routers along the way before reaching its destination. Significant energy savings can be achieved if a lightpath is set up for such a burst end-to-end or edge-to-edge using optical bypass [MZ3] and avoiding treatment of individual packets. This means partially returning to circuit switching (CS) [MZ7] which is also the basis of MIT’s Optical Flow Switching [MZ6]. Another CS option is WDM/TDM that is suitable for connections of subwavelength rates [MZ8]. Small flows (mice) with the same edge-router destinations can be aggregated together to use a low energy lightpath tunnel and possibly to save energy versus the current approach of using IP routers. All these indicate that routing and choice of layers and technology of various traffic types based on service demands and data rate at various locations can be optimized [MZ4],[MZ14]. This can be done is such a way that the cost function includes a substantial component of carbon tax which is necessary for sustainability [MZ4]. It is important that such optimization is done in a scalable way and this normally means in a network such as the Internet that the optimal (or near optimal) solution is obtained in a distributed way. As discussed in [MZ4], the optimization may lead to an optimal solution whereby certain layers (e.g. the IP layer) at certain parts of the network may be redundant and thus can be switched off either permanently or to a sleep-mode.

In general, the realization that the reduction of ICT energy consumption is important and that the “business as usual” [MZ2] is clearly unsustainable is already widespread. This is evidenced by many government and industry projects and researchers that report studies on how to reduce the use of electricity and to adapt the energy consumption to the demand by turning off idle network resources during period of low traffic demand, or to manage the traffic so it uses greener resources or technologies (see e.g. [MZ9], [MZ10], [MZ11], [MZ12], [MZ13], [MZ14] and reference therein). However, despite the effort there is a need for solutions that are scalable and deployable in the real world.

5.2 Communications and Networking Security
Security issue and solutions can be broadly subdivided between data/information security and system/network security. In the former case, discussed in Section 5.2.1, we focus on protecting
private data transmitted on the network so that the data will not be eavesdropped or faked, while in the later case (see Section 5.2.2) the focus of the security solutions is about protecting networks and systems from outside.

5.2.1. Data and Communications Security

To protect the data increasingly sophisticated encryption and authentication algorithms have been developed to either increase the difficulty to crack the algorithms or to accelerate the computation. Different solutions have been devised depending on the networking environment and the type of data to secure. It is well known that mobile wireless networks are generally more vulnerable to information than fixed wired networks as broadcast wireless channels easily allow message eavesdropping and injection (vulnerability of channels). Furthermore, among mobile networks, self-organizing networks (also referred to as mobile ad hoc networks) brings new security challenges due to the lack of infrastructure which makes the classical security solutions relying on security infrastructure based on on-line security servers not applicable. For this reason in the last years extensive research activities concentrated on self-organizing networks. This does not mean that all wired-network security issues have been addressed. Indeed relevant security challenges still exists inside the legacy Internet and novel and hot challenges are emerging with on-line social networking platforms and applications. In this section we briefly present and discuss some relevant challenges.

Security in self-organising networks. Research on communications security in the last decade intensively focused on self-organized infrastructure-less communication systems such as mobile ad hoc networks and sensor networks. Yet efficient methods to start-up security associations and trust in such systems still are lacking. Since the seminal Diffie-Hellmann key exchange protocol [RM1] various self-organized techniques for authentication, key management [RM2][RM3] and the integrated versions thereof as part of basic communication mechanisms like routing [RM4] have been proposed. On the one hand most of existing solutions suffer from unrealistic complexity and on the other hand some crucial problems like Sybil attacks still aren’t properly addressed. New approaches exploiting physical layer features [RM5] for key management in wireless communication and leveraging on multipath communications as a source of randomness [RM6] are quite promising research directions towards efficient solutions for security in self-organizing communication systems.

Security in content-centric networks. One of the concepts aiming to revolutionize the Internet architecture is the Content-Centric Networking (CCN) paradigm; see Section 3.1 and [RM7]. The CCN raises several interesting security requirements that can hardly be addressed with existing communications security mechanisms. Assuring basic requirements such as data confidentiality, integrity and user privacy in the original setting of the CCN is much harder due to the collapsed nature of communications in CCN whereby the separation between the data and network control vanishes. CCN thus calls for new mechanisms to protect data while enabling basic networking functions like routing and forwarding that operate on protected data. Preliminary results in this field have been achieved [RM8][RM9] as part of the HAGGLE project and future work can be inspired by work on publish-subscribe security [RM10][RM11].

Security of Internet infrastructure. On a much more practical basis, the security of the Internet infrastructure itself is a promising potential research topic. The security of the inter-domain routing infrastructure based on the BGP protocol has also been the focus of various studies and standardization activities. Despite the common assumption about well-known attacks such BGP-hijacking and a number of papers as to how to prevent it there is not sufficient public information

---

9 By Refik Molva (Institut Eurecom, France)
about the actual status of vulnerabilities and the security of the inter-domain routing in the Internet. Experimental research trying to evaluate the attacks against BGP and the Internet infrastructure would provide valuable evidence that is much needed to assess the actual security requirements for the Internet and pave the way for further research in designing countermeasures to the attacks. Further, another critical function of the Internet infrastructure that is highly exposed to malicious attacks is the Domain Name System (DNS). Exploited by attacks commonly known as DNS hijacking, DNS spoofing or DNS redirection, the main vulnerability of DNS is due to the lack of authentication in the basic request-response protocol. Despite several attempts in standardization of security features to combat these attacks, DNS still severely suffers from large scale deployment of security mechanisms to counter the attacks. New approaches to secure the DNS protocol and to prevent coordinated attack scenarios capitalizing on basic DNS vulnerabilities seem to be a promising research avenue.

Security of Internet applications. Turning to higher layers of the communication system, new applications like social networks provide new resources for research in communications and security alike. By leveraging on social characteristics of users that are made available by social networking applications, research can shed a new light on some “old” communication security problems such as anonymous communication, key management, and trust establishment. A trusted communication network can thus be brought up by building on some trust relationship that is inherent to the social networks [RM12], or by taking advantage of social behaviour of the users of the communication system. Furthermore, new security challenges are raised in the application layer by the accelerating outsourcing trend in computing whereby essential data storage and computing functions have been moving from the proprietary or private environments towards networked platforms operated by third parties. Having started with service oriented architectures this trend took momentum with web-based applications to reach an extreme through cloud computing. Even though justified by important economic factors such as reduced cost of ownership, outsourcing scenarios like cloud computing come with increased security exposures for the end-users due to the inherent lack of control over data and communications. The main security objective for outsourced environments is to provide users with security guarantees over the outsourced data and communication resources that are equivalent to the ones available in proprietary environments. The main challenge for security research in this context is to assure such equivalent security through security mechanisms that are executed in the untrusted runtime and communication environments provided by the outsourced platforms such as the cloud.

The main research challenges are thus raised by new communication paradigms such as self-organized, opportunistic networks and content-centric communications, and by the security problems of the Internet infrastructure that are still calling for countermeasures despite the coming to age of the Internet itself. In the applications and distributed systems arena, recent hot topics such as social networks and cloud computing including both processing and storage aspects also raise very interesting problems for security research.

5.2.2 System Security

System security focuses on protecting the boundaries of an organization network by keeping out intruders and prevent/react to attacks. A first step in system security is therefore associated with controlling access to the internal and external networks based on the policy of an organization or an Internet service provider. That is, it decides who can access what. This is done by either TCP/IP firewalls that check IP addresses and port numbers of the packets, or application firewalls that examine application headers or payloads. The second step is about protecting networks and

---

10 By Ying-Dar Lin (National Chiao Tung University, Taiwan)
systems so that they are not vulnerable to attacks from outside (e.g., from the rest of the Internet). An attacker may try to find and exploit vulnerabilities of a system to intrude into that system for various purposes, such as stealing critical information, controlling that system to launch another attack, disabling an important service, and so on.

The driving force behind various network attacks has been switched from deliberately abusing the Internet to making a profit [YL1]. The economic profit may be earned from distributing spam (through emails and social networks) or selling personal data (accounts, passwords, etc.) [YL2]. To achieve these goals with high scalability, attackers turn to infect a large number of hosts to form a botnet by attacks through various strategies, such as worms [YL3], emails [YL1] and Web sites [YL4]. It is difficult to defend these strategies with traditional methods, such as firewalls and VPNs, because the attacks are embedded in network traffic that should be allowed. That is why many security devices are equipped with the capability of deep packet inspection. The devices include intrusion detection systems (IDSs), anti-virus systems, anti-spam systems, and Web filtering systems.

The key problem in system security is how to identify the various types of attacks and defend the systems against them. Identification includes checking for attacking signatures or discovering anomalous behaviours. Checking signatures may miss unknown attacks (i.e., false negatives) but anomaly analysis may lead to false positives if normal traffic behaves unusually. So there is a trade-off between false positives and false negatives.

Furthermore, attackers want to evade the detection in a stealthy way. For example, packet splitting was intended to evade the detection of IDSs [YL5], code packing can evade the detection of anti-virus systems [YL6], spam templates can evade the detection of anti-spam systems [YL7] and so on. Therefore, the defenders have to restore the original semantics of the suspicious content, and find out the malicious content within it. Worse yet, an attacker may leverage cryptography to protect the attacks, e.g., encrypting a malicious program or malicious content, making effective detection a bigger challenge than ever. In those cases, the detection should be based on the network behaviour [YL8] or system behaviour [YL9] of a malicious program, rather than the content signatures, which might be evaded with polymorphism.

The struggles between attackers and defenders are endless. Several researchers infiltrate the botnets to see how they work [YL2], and get the idea on how to detect them. In the meantime, social networks have recently become a medium for malicious codes, such as spam and links to drive-by-download Web sites. The security and privacy for online social networks therefore have become a hot topic recently [YL10]. We foresee more research efforts on detecting stealthy system and network behaviour of malware, and studying how the attackers leverage the Internet for profit.

Even though we could design a sophisticated approach for deep packet inspection or behaviour analysis, we should also care about speeding up the processing as the volume of Internet traffic increases rapidly. Therefore, speeding up intrusion detection, virus scanning, or anomaly detection with hardware accelerators or multicore processors for multi-gigabit-per-second traffic is also a trend [YL11].

In summary, the following questions still call for better answers from the research communities:

- Is there a better way to integrate signature-based identification and behaviour-based identification so that both false negatives and false positives could be reduced? In our discussion above, we know that signature-based schemes could lead to more false negatives, while behaviour-based schemes have potentially higher false positives. The latter is more serious than the former. It thus limits the applicability of the behaviour-based schemes. A promising direction is to jointly utilize both to compensate each other.
- How well can we detect stealthy attackers on the network and the host? Do we have good enough countermeasures for evaded or encrypted attack traffic? How about the backdoor programs residing on host machines? In the discussion above, detection based on network
or system behaviours was suggested. But both cost and false positives for that would be high.

- How scalable are our countermeasures, as the attacks have scaled up through botnets and online social networks? Detecting or stopping one single malicious packet or malware does not alleviate the problem much. Most solutions developed so far do not address this scalability issue. Solutions that could scale to subnets, domains, or even the global Internet scale are demanded.
- Can we afford wire-speed deep packet inspection or behaviour analysis as the traffic we handle grows to 10Gbps and beyond? Compared to switching and routing, the processing here is much heavier and requires hardware support, e.g. accelerators or multicore processors. But a 10Gbps hardware would become wasted when integrated with a heavy software component running on a slower processor. The hardware software co-design issues need to be addressed.

6. Further Readings

In this paper we have analysed and discussed some hot research challenges in the evolution/revolution towards the Future Internet. The research challenges that have been discussed in this manuscript represent a notable, but not exhaustive list, of the research opportunities for our community generated by the new role of Internet as a complex techno-socio-economic system which is aimed at becoming a content-centric infrastructure [FR1] able to mediate almost all the socio and economic interactions of our society. In this section, we complement the paper discussions with a brief survey of other relevant research areas and we provide some references (with special attention to those recently published in Computer Communications), which may constitute a starting point for those interested in these research areas.

We start this discussion from the evolution of Internet role in the last ten years and the implications on its architecture (which have already been partially discussed in Section 3). The convergence of all communication media towards the Internet has shown the great flexibility of its original design but, at the same time, has pointed out the limitations of current Internet in terms of security and privacy (including communications privacy [FR2a]), mobility/multimedia support, energy efficiency, etc. To cope with these challenges two main approaches (which are not necessarily at odds) are currently taken for adapting Internet to its new role: an evolutionary/incremental approach and a clean-slate approach [FR2b].

The evolutionary approach is based on the consideration that, given the current Internet scale, only small and incremental changes are possible without any fundamental change to the underlying best-effort IP network. This approach is based on adding middleboxes (e.g., caches, proxies, etc.) into the network or modifying the network at its edges (e.g., p2p overlays). The evolutionary approach results in a stretching of the original Internet design in which new mechanisms introduced to solve emerging problems interfere in an unpredictable way with existing ones possibly leading to the emergence of new problems. For these reasons the scientific community is pushing towards a clean-slate re-design of the Internet architecture and protocols. This means a re-design of the Internet according to disruptive design principles without being constrained by the current Internet. The research on novel architecture and protocols is a hot research topic toward the Internet of the Future, and therefore we solicit more research efforts in this direction.

An in-depth discussion about the proposed approaches for designing the Future Internet is presented in [FR3]. In this work the authors analyse the ongoing research in the field presenting the major research initiatives with a special attention to the research carried out on security,

11 By Marco Conti (IIT-CNR, Italy)
content distribution, challenged/opportunistic networks, internetworking and management. In addition the authors discuss the role of experimental research in the development of the Future Internet. The experimental research has always been a fundamental element in the design and evolution of the Internet. Measurement-based research activities have been extensively used to better understand the Internet properties and analyzing the evolving Internet structure. This has stimulated intensive research activities to develop methods, tools and testbeds for supporting passive and active measurements. Passive measurements (e.g., see [FR4] for a review) use the existing network traffic, while active measurements create and send ad hoc probe packets. Active measurements are often used for estimating the network QoS that can be offered to an application -- e.g., [FR6] presents and compares the available techniques and tools for bandwidth estimation. Some challenges, which occur when applying active measurement techniques, are discussed in [FR5].

Internet measurements are a basic tool for studying the Internet tomography, which is otherwise unknown. For example, available datasets indicate that the current structure of the Internet, as discussed in Section 3.1 is not as strongly hierarchical as it should be according to classical textbooks. This is due to the increasing role of Internet exchange points in the traffic forwarding among the autonomous systems [FR7].

While measurements are extensively used in the study of the current and Future Internet, there is a lack of common standards to perform and validate these studies. To overcome this limitation, in [FR8] the authors discuss common problems in measurement studies and present their Socratic approach to obtain reliable datasets that can be reused by other researchers.

In the Future Internet broadband mobile and wireless networks will have a key role. Indeed the number of people accessing Internet through a mobile device is continuously growing at a fast rate, and we can easily estimate that mobile users will be highly predominant in the Future Internet and they will use bandwidth intensive applications such as video streaming and/or IPTV [FR9]. For these reasons, there is a great interest in the research and industry communities to develop effective broadband technologies for the ubiquitous access to the Internet. In particular the Third Generation Partnership Project (3GPP) has promoted in 2008 the first release of LTE (Long Term Evolution) specifications (immediately followed by a new release in 2009) that include a long-term evolution of the radio access technology (EUTRAN: Evolved Universal Terrestrial Radio Access Network), and the optimization of the core network for IP-based traffic (EPC: Evolved Packet Core). Computer Communications has devoted a special section to survey the LTE technology [FR10] with a special focus on the LTE radio interface and radio network [FR11], the LTE security architecture [FR12], and the LTE media coding [FR13].

In the future Internet, cellular technologies will be complemented by other wireless technologies to provide the ubiquitous access to the Internet. Among these, the WiMAX technology, based on the IEEE 802.16 standard is one of the most promising solutions for broadband wireless metropolitan area networks [FR14], while the WiFi technologies, based on the 802.11 standard family, are the de-facto standards for the nomadic Internet access. The interconnection of WiMAX and WiFi technologies is a very promising solution for providing a high-speed wireless access inside a city to offload the traffic from the congested cellular networks. Therefore, designing effective mechanisms for guaranteeing seamless vertical handoffs among these technologies is a very important and hot research issue [FR15]. The effectiveness of vertical handoff mechanisms between WiMAX and WiFi networks can be measured in terms of energy efficiency [FR16], and/or in the ability to support applications with QoS requirements (e.g., VoIP) while the mobile device changes the network it is connected to [FR17].

The WiFi technology, which has been introduced during the 90s, is still under continuous evolution for adapting it to the emerging scenarios such as, for example, new multi-hop ad hoc networks like vehicular networks (802.11p), mesh networks (802.11s), etc. Indeed, WiFi is also the enabling technology for the development of mobile/multi-hop ad hoc networks for civilian applications. This network paradigm is based on the idea to build a network in any area with no
pre-existing communication infrastructure by exploiting the ability of the users’ devices to self-organize into a temporary network where the source-destination traffic is forwarded via a sequence of intermediate devices [FR18]. This paradigm has been often identified with the technologies developed inside the MANET IETF working group. However, as discussed in ([FR19] and [FR20]), while the MANET paradigm (due to a lack of realism in the objectives and in the design) does not have a major impact on computer communications, the multi-hop ad hoc networking paradigm has been successfully applied in several classes of networks that are currently penetrating the mass market. The mesh-network paradigm is a meaningful example of this by identifying from its initial design a set of application scenarios (i.e., providing a flexible and “low cost” extension of the Internet) to drive/motivate its design, and by reducing the MANET complexities with the introduction of a fixed-router backbone, which limits the impact of node mobility to the last hop [FR21]. Mesh network is already a quite consolidated technology for a low cost extension of the Internet with few hops wireless links. However several aspects of this technology are still under intensive investigations to make this technology more robust and able to support more advanced services. Open research issues include novel routing paradigms ([FR22], [FR23], [FR24]), QoS support ([FR24], [FR25]), security [FR26], and experimental vs. simulation studies [FR27].

Vehicular Ad hoc NETworks (VANETs) are another notable example of a successful networking paradigm that is emerging as a specialization of (pure) MANETs. VANETs research is well motivated by the socio-economic value of the transportation sector, which motivates the development of advanced Intelligent Transportation System (ITS) aimed at reducing the traffic congestions, the high number of traffic road accidents, etc. Advanced ITS systems require both vehicle-to-roadside (V2R) and vehicle-to-vehicle (V2V) communications. In V2R communications a vehicle typically exploits infrastructure-based wireless technologies, such as cellular networks, WiMAX and WiFi, to communicate with a roadside base station/access point. The planning of a cost-effective roadside infrastructure able to provide a good coverage of an urban area is a complex problem. For example, in [FR28], the authors tackle this problem by proposing simple heuristics that (assuming that the characteristics of vehicular mobility are known) provide near-optimal coverage of the vehicles moving in an urban area.

Understanding the performance of the V2R communication channel as a function of the vehicular traffic parameters (e.g., vehicles’ speed, vehicles’ density, road capacity, etc.) is a very important research topic to determine the throughput available to moving vehicles for accessing the Internet. This problem has been investigated mainly through experimental studies [FR29], while only recently there have been some attempts to develop theoretical models to characterize the QoS experienced by moving vehicles [FR30].

V2V communications exploit a new class of multi-hop ad hoc networks, named VANETs. Specifically, according to the multi-hop ad hoc networking paradigm, the vehicles on the road dynamically self-organize in a VANET by exploiting their wireless communication interfaces (e.g., 802.11p). The V2V research field inherited MANET results related to multi-hop ad hoc routing/forwarding protocols [FR31], which have to be tuned and modified for adapting them to the peculiar features of the vehicular field [FR32]. A special attention has been reserved to the development of optimized broadcasting protocols as several multi-hop applications developed for vehicular ad hoc networks use broadcast communication services [FR33] [FR34]. However the high level of vehicles’ mobility and the possibility of sparse networking scenarios, which occur when the traffic intensity is low, make inefficient the legacy store-and-forward communication paradigm used in MANET and push toward the adoption of the more flexible and robust store-carry-and-forward paradigm adopted by the opportunistic networks [FR42]. Specifically, according to this paradigm (which is also referred to as delay tolerant or challenged networks), nodes can physically carry buffered data while they move around the network area, till they get in contact with a suitable next-hop node, i.e., until a forwarding opportunity exists. In this way, when a vehicle does not have a good next hop to forward the data it simply stores the data locally.
without discarding it, as it would happen in the MANET. In addition, with the opportunistic paradigm, data can be delivered between a source and a destination even if an end-to-end path between the two nodes never exists by exploiting the sequence of connectivity graphs generated by nodes' movement. The opportunistic paradigm applied to vehicular networks has recently generated a large body of literature mainly on routing protocols and data dissemination in vehicular networks (e.g., [FR35], [FR36]). However, there are still several interesting and challenging issues to be addressed (e.g., privacy [FR37]); a special attention should be reserved to develop realistic models to characterize the mobility of the VANET nodes [FR38], and to analytically study the VANET performance [FR39].

V2R and V2V communication systems can support a large plethora of applications including safety applications (e.g., collision avoidance, road obstacle warning, safety message disseminations, etc.), traffic information and infotainment services (e.g., games, multimedia streaming, etc.). An extensive survey of the vehicular applications is presented in [FR40], while [FR41] presents a vehicular platform -- that integrates into the Android platform the Open Gateway Service Initiative Vehicle Expert Group framework -- which provides an open environment for the development of automotive telematics applications.

The opportunistic networking paradigm, which has been successfully applied in the context of vehicular networks, is indeed one of the most interesting generalisations of the MANET paradigm. Indeed MANET represents an engineering approach to develop routing protocols, which mask the node mobility by constructing “stable” end-to-end paths as in the wired Internet. On the other hand, opportunistic networks do not consider the node mobility as a problem (to mask) but as an opportunity to exploit. In opportunistic networks the mobility of the nodes creates contact opportunities among nodes, which can be used to connect parts of the network that are otherwise disconnected. Therefore, this paradigm constitutes a generalization of the legacy Internet paradigm (where communications can occur only if and end-to-end path exists), and it seems very suitable for the communications in pervasive environments where the environment is saturated of devices (with short-range wireless technologies) that can self-organize in a network for users’ interactions and content exchange. In these scenarios, the network will be generally partitioned in disconnected islands, which might be interconnected by exploiting the nodes’ mobility.

Opportunistic networking is an area of growing interest with several challenging issues. Routing in opportunistic networks is surely one of the most compelling challenges, due to the scarce knowledge of the topological evolution of the network. This has already generated intense research activities in the area, which has produced several proposals for routing and forwarding in opportunistic networks [FR42]. Among these, the most innovative and promising class of routing protocols is represented by protocols that try to exploit the nodes’ social context such as HiBOp [FR44], Bubble Rap [FR45a] and SimBet [FR45b]. Specifically, HiBOp infers social relationships between nodes from the context information dynamically gathered by each node, and identifies good forwarders by comparing the social context of the forwarder and destination. On the other hand, both Bubble Rap and SimBet exploit social-network properties. The basic idea is to forward packets toward a more central node, i.e., a node that is better connected and hence offers more forwarding opportunities. For example, Bubble Rap assumes that nodes are clustered in “social” cliques and that nodes belonging to disjoint cliques can communicate through nodes, which are shared among cliques (i.e., nodes belonging to more social communities). The dynamic identification of the social communities a node belongs to is currently a hot research problem in the framework of social-aware protocols for opportunistic networks [FR46].

While routing in opportunistic networks is a well-investigated area, other areas, such as data dissemination and security and privacy, still need more intense research activities. Data dissemination is a natural follow-up of research on forwarding algorithms. One of the most interesting use cases for opportunistic networks is indeed the sharing of content available on
mobile users’ devices. For these reasons, content dissemination is now a hot research area where some interesting results can be found in [FR47-FR49].

Privacy is currently one of the main concerns in opportunistic networks as the context information exchanged among nodes (for selecting the best forwarder) might include sensible information. Very promising results to tackle the problem are presented in [FR50]. Security is a hot and key challenge for opportunistic networks, as mobile users operate on the move in open, possibly adversary, public environments. A preliminary discussion on encryption, and robustness against DoS attacks to the operations of opportunistic protocols can be found in [FR51]. Another network security issue is related to preventing uncontrolled resource hogs (i.e., individuals whose message generation rate is much higher than the average), which may significantly reduce the network performance [FR52].

Inside the opportunistic-network research it is worth remembering the research activities carried out inside the Delay-Tolerant Networking Research Group (DTNRG). DTNRG is an IRTF research group12, which is developing architecture and protocols to extend the Internet protocol stack in order to cope with frequent partitions, which may destroy the behaviour of legacy Internet protocols, e.g., TCP. To this end, DTNRG has developed an overlay, named Bundle Layer Protocol, that it is implemented in some network nodes (named DTN nodes) which, during the disconnection phases, use a persistent storage to store the packets to be forwarded [FR53]. The bundle layer is implemented above the transport and below applications and it is aimed to mask the network disconnections to the higher layers. Instead of “small” packets, the bundle layer uses for the data transfer variable-length “long” data units called "bundles". Computer Communications devoted a special issue to present some of the hottest research topics in the DTN research community [FR54]: efficient policies for handling the network disconnections [FR55], routing protocols [FR56], energy consumption/efficiency [FR57], the development of a session-layer approach to augmenting the Bundle Layer Protocol [FR58], and multicast communications ([FR59], [FR60]). The special issue also includes papers analyzing two relevant DTN application scenarios: deep-space networking [FR61] and vehicular networking [FR62].

An opportunistic network exploits the devices mobility for its operations. As humans typically carry the devices, it is the human mobility that generates the communication opportunities. Therefore, understanding and modelling the properties of the human mobility is a key enabler for opportunistic networking. Studying human mobility traces is the key element to understand the properties of the human mobility. The aim is to provide a characterization of the temporal properties of devices/humans mobility with special attention to the contact time, i.e., the distribution of the contact duration between two devices, and the inter-contact time (ICT), i.e., the distribution of the time between two consecutive contacts between devices. In particular the characterization of the ICT distribution has generated a great debate in the scientific community where different research groups have claimed completely different results ranging from heavy-tailed distribution functions -- with [FR63] or without [FR64] an exponential cut-off -- to an exponential distribution [FR65]. In [FR66] has been shown a fundamental result that helps explaining the differences among the ICT distributions claimed by different research groups. Specifically, in that paper the authors derive the conditions under which, by starting from exponential inter-contact times among individual couple of nodes, we can obtain a heavy-tailed aggregate ICT distribution (i.e., the ICT distribution between any couple of nodes). Understanding the properties of the ICT distribution is a critical issue as from this distribution depends the effectiveness of several routing protocols for opportunistic networks. For example, in [FR64] the authors have shown that for a simple forwarding scheme, like the Two Hop scheme, the expected delay for message forwarding might be infinite depending on the properties of the ICT distribution.

---

12 [http://www.dtnrg.org](http://www.dtnrg.org)
Starting from the observed properties of the human mobility, several models have been proposed to provide a synthetic characterization of the human mobility to be used in the performance evaluation studies used for comparing and contrasting the mechanisms and protocols developed for opportunistic networks. Some mobility models, in addition to the inter-contact properties, also represent the impact of social relationships in the human mobility ([FR67a], [FR67b]). An updated survey on human mobility models, with a discussion of the open problems, is presented in [FR67c].

Currently, opportunistic networking is a very active research area. While a consolidated literature exists on routing protocols, additional work is expected in other areas like, for example, dissemination protocols and security. However more contributions are mainly expected on the modelling and performance evaluation in order to develop a better understanding of the basic properties of these networks. Examples of ongoing works include the modelling of (social-aware) routing protocols in heterogeneous settings ([FR68], [FR69]), and new theoretical models for investigating the properties of the connectivity graphs that characterize the connectivity properties of an opportunistic network [FR70].

A further step toward a truly pervasive Internet is represented by the cyber/physical world convergence, where the information about the physical reality (e.g., collected through sensor nodes) is seamlessly transferred into the cyber world where it is elaborated to adapt cyber applications and services to the physical context, and thus possibly modifying/adapting the physical world itself through actuators [FR71c]. The wireless sensor networks (with [FR71a] or without actuators [FR71b]) have therefore a major role in controlling and connecting the physical world from the cyber world [FR71c]. Wireless Sensor Networks (WSNs) represent a “special” class of multi-hop ad hoc networks that are developed to control and monitor events and phenomena. To this end, a number of sensor nodes (with a wireless interface) are deployed inside the monitoring area. If the sensor network is sufficiently dense to guarantee a connected network, the information collected by the sensor nodes is delivered, by following the multi-hop paradigm through the other sensor nodes, to a sink node and through it to the Internet. If the sensor-node density is low, and hence the sensor network is disconnected, mobile elements (also refereed to as data mules or message ferries [FR43]) are used to collect the sensed data and deliver them to the sink. Indeed the design of these networks highly depend on the application scenarios and the requirements of the applications in terms of reliability, timeliness etc. WSNs have been very successful both on the academic and industrial side, as they are developed for solving specific application requirements. Thus, they triggered in the last ten years, intensive scientific activities, which has produced a large body of literature that is addressing all the key WSN research challenges: energy efficiency [FR72], MAC protocols [FR73], routing protocols [FR74], clustering algorithms [FR75], time [FR76] and clock [FR77] synchronization, security [FR78-FR80], coverage and connectivity ([FR81] [FR82]), networks with mobile nodes [FR83], etc. The existing literature leaves a very limited space for producing additional original scientific works on legacy WSN problems like routing, clustering, MAC protocols, synchronization, coverage, etc. On the other hand further works are still expected to address specialized problems ranging from QoS to privacy, security and trust ([FR83], [FR89]), specialized network scenarios [FR90-FR91], or the usage of sensor networks in challenging environments like underwater [FR92-FR93], underground [FR94], industrial environments [FR95], etc. However, the most promising research directions in the sensor network field are related to the new challenges emerging from the use of mobile phones as a human-centric sensing tool [FR96a]. By pushing further this view, we can think to exploit the billions of users’ mobile devices/phones as location-aware data collection instruments under the users’ control for real world observations. In this way we can sense the physical world without deploying an ad hoc sensor networks. This novel paradigm is known as participatory sensing [FR96b].

The participatory sensing paradigm is a part of the more general opportunistic sensing paradigm, which is based on the (opportunistic) exploitation of all the sensing devices available in
the environment to achieve a given sensing task. Indeed in a convergent physical-cyber world, a
wide variety of smart devices spread in the physical world (such as RFID Tags, Sensors and
Actuators, Sensor-rich Smart Phones, or Proximity Sensing Technologies) are leading to the
emergence of a very dense ICT infrastructure for monitoring the physical world, and collecting
information about the user behaviours and requirements. In particular, multi-modal sensors spread
in the environment can be opportunistically exploited to infer precise information about the social
behaviour of the users and the social environment around them. Indeed participatory and
opportunistic sensing offers un-precedent opportunities for pervasive urban sensing [FR97]: to
effectively collect and process the digital footprints generated by humans when interacting with
the surrounding physical world and with the social activities therein. A major goal of these sensing
activities is to investigate the hybrid city, i.e., a city that operates simultaneously in the
cyber/digital and physical realms, to investigate the human behaviour and his socio-economic
relationships. This is a highly challenging and innovative research objective that can bring to the
development of novel urban applications that benefit citizens, urban planners, and policy makers.
Preserving the privacy of the individuals contributing their sensed data is a major challenge for
progressing towards the pervasive urban sensing ([FR98], [FR99]).
References


[CR9] www.ict-aragorn.eu


[SU22] C. Huang, A. Wang, J. Li, and K. W. Ross, “Measuring and Evaluating Large-scale CDNs,” in ACM IMC ’08, paper withdrawn at Microsoft request.


[FR48] PodNET project http://podnet.ee.ethz.ch/


[FR90] Xu Li, Hongyu Huang, Xuegang Yu, Wei Shu, Minglu Li and Min-You Wu, “A new paradigm for urban surveillance with vehicular sensor networks,” Computer Communications Volume 34, Issue 10, 1 July 2011, Pages 1159-1168


