eMobility Technology Platform

Expert Working Group EWG6 on

Broadband Mobile Systems

White Paper

Date: 16 February 2010

Chair: Matti Latva-aho (CWC)

matti.latva-aho@ee.oulu.fi

List of Contributors

Contributors	Company/institute	e.mail address
Behnaam Aazhang	Rice University	aaz@rice.edu
Giuseppe Abreu	CWC	Giuseppe.abreu@ee.oulu.fi
Ulrico Celentano	CWC	Ulrico.celentano@ee.oulu.fi
Marian Codreanu	CWC	Marian.codreanu@ee.oulu.fi
Sudhir Dixit	CWC	Sudhir.dixit@ee.oulu.fi
Markku Juntti	CWC	Markku.juntti@ee.oulu.fi
Matti Latva-aho	CWC	Matti.Latva-aho@ee.oulu.fi
Wei Li	CWC	Wei.lii@ee.oulu.fi
Anna Pantelidou	CWC	Anna.pantelidou@ee.oulu.fi
Rahim Tafazolli	Surrey University	R.Tafazolli@surrey.ac.uk

Table of Contents

1 Rationale

Wireless technologies have experienced major cycles of development in the past twenty years or so. The introduction of digital transmission schemes e.g. in GSM, the numerous attempts to drastically improve the capacity for voice traffic by introducing CDMA to cellular systems, the migration of voice oriented cellular systems to mobile internet for consumer markets in the case of 3G cellular are just a few examples of such development. Most recently, broadband wireless data systems have been introduced for mobile cellular users, both in WiMAX and 3G-LTE based systems. However, during their evolution, these systems soon faced the problem that the radio transmission rates over a link have almost reached the theoretical limit. A promising technology is the IMT Advanced (IMT-A), which aims to provide mobile users with data rates that exceed 100 Mbps. Despite the rapid progress in developing new technologies, the underlying network topology has not changed considerably in the past two decades: The basic cellular network architecture with centralized control is more or less present even in today's 3G-LTE and future LTE-A concepts although quite recently the focus has been towards more decentralized radio access network control.

The ITU-R process to define IMT-A is on-going and there are high expectations to introduce totally novel ideas for radio spectrum sharing, network self-organization and local optimization techniques, improved energy efficient architectures, intelligent radio resource management, user scheduling schemes and so on. In addition, several on going research developments envision to drastically change wireless networking. These include extending various multi-antenna concepts to, the so-called, coordinated multi-antenna transmission and reception, improving transmission range and reducing interference in macrocellular environments by exploiting relay nodes, maximizing local capacity in hierarchical cellular networks through femtocells and enhancing spectrum usage through spectrum sharing and secondary spectrum usage.

In parallel to mobile cellular network development several research initiatives have been launched to study future Internet. The Internet originally was not designed to be a backbone technology for global communications and trade as it is today and its technical solutions were not intended to facilitate connectivity of billions of devices into the ubiquitous information infrastructure of the future networked society. It is obvious that the expected transformation includes great opportunities, but also poses huge challenges. The key technical driver for Internet development in the future stems from the need to account for wireless mobility. As the networks will become more dynamic and aware of their surroundings, all content exchanged in the Internet is expected to become dynamic as well. Under such a setting, server based information acquisition will become very inefficient and user co-operation and information exchange will become more attractive.

Partially related to dynamics in the network utilisation, centralized network control in general needs to be gradually replaced by intelligent context aware network control. This is often called cognitive networking which is a concept where, radio, network and available computation as well as power resources are utilised in the most efficient way in a dynamic network environment. The concept of opportunism should also cover the utilisation of spatial domain, user location and context information, co-existence of several access technologies, knowledge of channel statistical behaviour and traffic pattern that includes long-term monitoring and prediction, network topology awareness in mobile device, and so forth.

The transmitted data volume increases approximately by a factor of 10 every 5 years, which corresponds to an increase of the associated energy consumption by approximately 15 – 20 % per year. Currently, 3 % of the world-wide energy is consumed by the ICT infrastructure. However, if the energy consumption doubles every 5 years, it is natural that it will lead to serious problems. Decreasing the energy consumption of future wireless radio systems is currently being investigated in the framework of green radio networking. Another challenge is to globally reduce the electromagnetic radiation levels to achieve better coexistence of wireless systems and to further decrease human exposure to radiation.

The anticipated major growth in wireless business focuses mainly at emerging markets in China, India, Africa, and South-America. This does not only open new markets for promising technologies, but also brings forth newer technological challenges and opportunities. The need in these market economies is for relatively inexpensive, easily accessible, diversified, and expandable communication services, ranging from basic voice and data services to more content-rich multimedia and broadcasting optimized for the masses. The network set-up should be as simple as "plug and play" procedure that allows a flexible network topology, spectrum usage, co-existence of technologies. Also, the data rate requirements are very diverse depending on the type of services needed. Both CAPEX and OPEX should be low, meaning e.g. auto/self everything, very high energy efficiency in a very flexible wireless network. This may open up a new paradigm in the development of wireless systems relatively soon.

As wireless networks are becoming more complex and diverse, understanding the fundamental behaviour and reasoning of different types of networks is becoming increasingly important. The well-established theories for fixed and often static networks need to be extended to dynamic heterogeneous wireless networks with different types of traffic. Instead of the optimal centralized solutions, practical decentralized networking management schemes must be sought. The cross-layer optimization methods are not yet well understood and need careful studies before they can be successfully applied to the evolution of future networks. All these need careful investigations from the scientific community before they can be used as powerful tools for future wireless network design.

To summarize, the key dimensions and drivers for the development of future wireless networks can be summarized to include at least the following key aspects:

- Evolution of Mobile Cellular Access and Networking Techniques
- Future Internet Architectures
- Position and Context Aware Network Management
- Cognitive Wireless Networking in a Heterogeneous Network Environment
- Opportunistic and Flexible Spectrum Usage
- Energy Efficient and Green Networking
- Special Needs for Emerging Markets
- Development of Fundamental Theories for Wireless Networking

Based on these remarks, the future research agenda on wireless networking are in the following areas:

1. Broadband Mobile Access
2. Mobile Wireless Driven Future Internet
3. Fundamental Enabling Networking Technologies
4. Wireless Technologies and Applications for Emerging Growth Markets

The SRA for eMobility has a broad overview on future mobile technologies. The area 1 in this document covers only the relevant issues not discussed in the SRA. In this whitepaper the area 2 is left mainly out as it is addressed explicitly in a separate white paper. The area 3 can potentially cover many more areas than presented in the present version of this whitepaper; the target has been to show that there exists a major gap between the potential promised by theory and the reality seen by practical systems. The major uniqueness of this whitepaper is presented in the area 4, which can potentially lead to a paradigm shift in the development of wireless systems in a broader scale.

2 Research priorities

2.1 Research priorities in broadband mobile access

The need for higher data throughputs, driven by the high bandwidth demands of a continuously increasing number of multimedia applications, will continue to increase in future wireless networks. Those are enabler for future wireless entertainment, intelligent transport systems, telemedicine, emergency and safety/security applications and more. Furthermore, applications, such as 3D Internet, virtual and augmented reality, and tele-presence are expected to emerge in the future and push the demand of data rates to completely new orders of magnitude. At the same time, it is expected that a trillion of devices will be connected to the Internet of the imminent future, creating great technological challenges in terms of the networking architecture.

To achieve the high data rates needed for enabling the future services, new technologies are needed over the complete radio networking protocol stack. New radio technologies will be operating at significantly higher carrier frequencies. This poses remarkable challenges to the radio frequency (RF) and baseband (BB) hardware, software, and other implementation technologies. More importantly, to make the vision of the futuristic high rate services a reality it will be mandatory to enable concurrent usage of multiple radio frequencies, technologies and heterogeneous networks.

2.1.1 Evolution of mobile cellular systems

The ITU-R process of defining the IMT-A is in progress and there are high expectations for some totally novel ideas for radio spectrum sharing, self-organization and local optimization networking concepts, energy efficient networking architectures, intelligent radio resource management and user scheduling schemes. There are also other developments that are evolving which have the potential to change the wireless networking concepts drastically. Different multi-antenna concepts have been extended to the so-called coordinated multi-antenna transmission and reception, relay nodes have been utilized to improve range and reduce interference in macro-cellular environments, femtocells have been exploited for local capacity maximization in hierarchical cellular networks and spectrum sharing and secondary spectrum use have been suggested to enhance spectrum usage, just to mention few examples.

2.1.2 New radio technologies

Today modern WLAN systems using OFDM based basebands, combined with multiple-input multiple-output (MIMO) techniques, use typically 20–40 MHz of bandwidth and can communicate at a speed that can reach 600 Mb/s, as indicated in IEEE802.11n. New discussions in the IEEE802.15.3c already try to go beyond the 1 Gb/s and achieve 2–6 Gb/s by using the 60 GHz band. Currently, the bandwidth for this speed is 2 GHz, which results in a spectral efficiency of less than 1. Novel physical layer radio interfaces and RF-BB solutions must be designed to improve the spectral efficiency of the transmission schemes and exploit the unutilized frequency resources. To realize data rates of tens or hundreds of Gbit/s on wireless short range links, even a substantial increase of the modulation level, requires a bandwidth in the order of tens of GHz. This can only be found by exploiting higher radio frequencies, e.g., in the bands beyond 60 GHz all the way up 100–500 GHz.

MIMO systems operating in higher frequency bands can leverage small wavelengths, because directive antennae can be implemented onto small size devices. However, this poses also enormous challenges at the signal processing level due to the high signal processing rates. Even more so, the antenna technology and RF parts of the transceivers will require significant R&D to realize this vision.

2.1.3 New network technologies

The evolving novel radio technologies open opportunities for completely new radio resource control, frequency allocation, and multiple access mechanisms. The selection of the best route and space-time-frequency resources poses a complicated multidimensional optimization problem.

The future wireless networks will have novel network topologies to complement the legacy cellular topology (see

Figure 1). Fixed relay stations will be applied in future networks to improve capacity and coverage via improved link budget and increased diversity. Peer-to-peer communications and mesh networking are likely solutions for future high-speed networks. The concurrent full-duplex usage of multiple radio technologies and frequencies will become feasible in the future. This opens a new venue for routing, radio resource control, and usage of the available radio connections in a coordinated manner. For the realization of the promised performance gains, new solutions for network optimization are required.

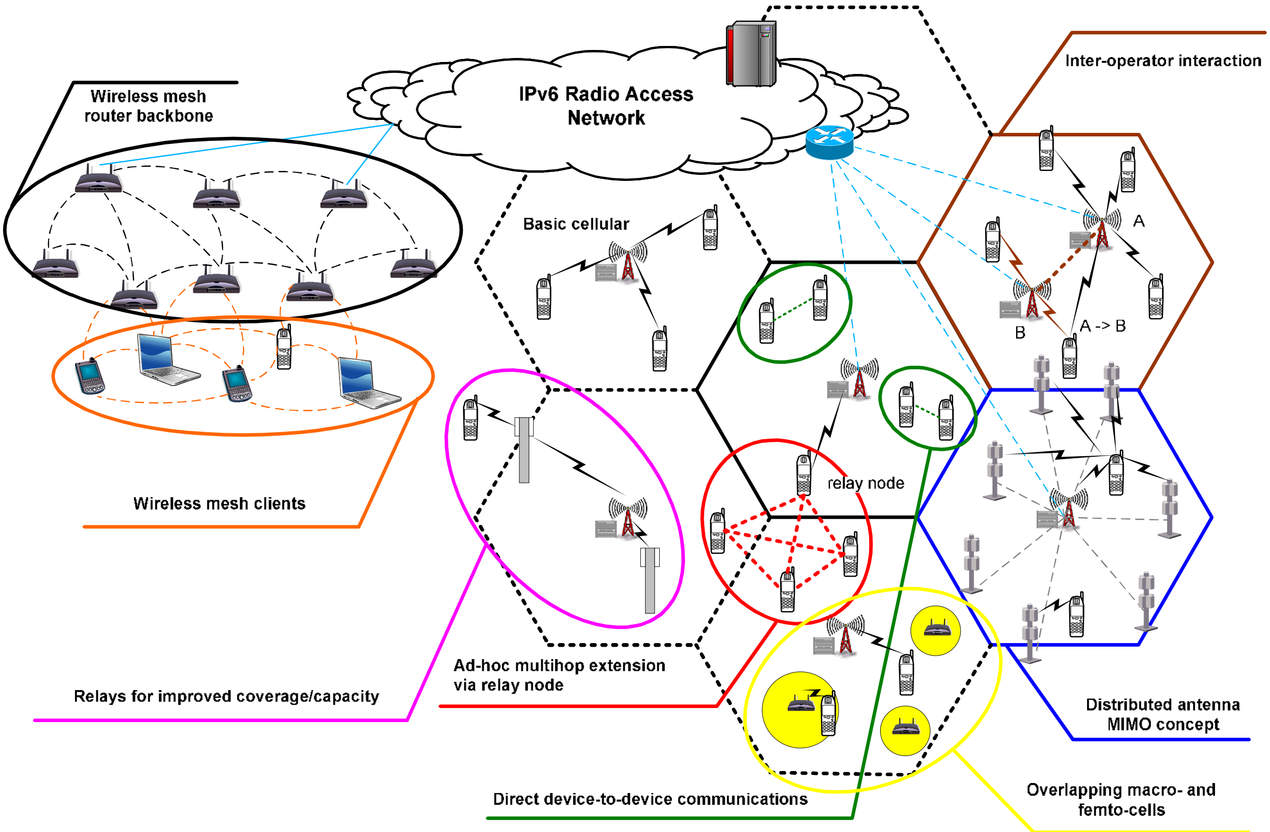

Figure 1. New wireless network topologies.

2.2 Research priorities in fundamental enabling networking technologies

2.2.1 Motivation

Undoubtedly, wireless technology has broadly entered the contemporary era and promises to provide the wireless users with a vast amount of data services. The rapid proliferation in the number of wireless users mandates the efficient utilization of the available spectrum and the achievement of high data rates to support the continuously increasing user demands. Furthermore, traffic in future wireless systems is anticipated to be composed not only of non-real time applications, such as web browsing, messaging or email, but also of real-time traffic such as video, multi-media, and teleconferencing. The latter type of traffic mandates the ability of the network to provide Quality of Service (QoS) guarantees to the wireless users, for instance in terms of (i) satisfying a minimum rate requirement and (ii) maintaining packet delays that with high probability do not exceed a given maximum value. In addition, the heterogeneity between the different traffic types introduces new problems regarding how to support simultaneously non-real time and real time users so that the former receive a maximum possible throughput rate without affecting the QoS requirements of the real-time users.

Finding answers to the above challenging questions is further hampered by the fact that wireless systems require a fundamentally different treatment than the preceding wireline ones. One of the main difficulties in wireless is the observation that there is no “hard” notion for a link. Instead, the notion of a link is “soft” and depends on various parameters such as the transmission power and rate of a transmitting user, the channel quality, and the amount of interference at a receiving user that originates from other concurrent transmissions. Furthermore, adverse effects such as fading, scattering and unpredictable user mobility cause the channel quality to fluctuate with time, which consequently affects the corresponding achievable rates. On the other hand, the wireless medium can be viewed positively from its shared nature; (i) can act synergistically towards achieving better performance as it allows the possibility for opportunistic communication and (ii) offers novel ways of communication that can be exploited such as making use of its broadcast nature to facilitate multicast communication and allow user cooperation.

The Open Systems Interconnection (OSI) architecture is built upon the assumption of the existence of a robust physical layer providing “reliable” links with “stable” capacity. Furthermore, the OSI architecture does not foresee the possibility of exploiting the time-variations and the new modalities that the physical layer has to offer. Instead it views the physical layer purely as a “pipe” of bits. By restricting the ability of the physical layer to influence the layers above, the OSI architecture maintained a well-defined separation between the layers, which modularized and simplified the study of the different functionalities by requiring communication only between directly neighboring layers. Thus, the applications and the higher layer protocols such as protocols related to routing, congestion, and flow control could be designed almost independent of each other. Although this separation was reasonable in wireline networks, it is clear that in wireless a lot can be gained by exploiting the coupling between the physical layer and the layers above it.

2.2.2 Cross-layer design revisited

For wireless systems to fulfill their promise in terms of providing a broad class of services to heterogeneous users at high data rates cross-layer interactions at all layers have to be considered. In particular:

- **Medium Access Control**: Due to the fact that concurrent transmissions interfere among each other appropriate scheduling algorithms need to be introduced to mitigate the effects of unwanted interference. Although scheduling is a decision of the Medium Access Control (MAC) layer, the ability to schedule two or more users concurrently depends on their transmission powers and rates as well as on the underlying channel conditions. Thus, there is an inherent coupling between the MAC and the physical layer. As an example, in the downlink of a base station a scheduler that takes advantage of the channel variations by giving priority to the users with instantaneously better channels, like in the CDMA2000 HDR can do better in terms of performance, e.g., maximizing the total throughput or proportional fairness, as opposed to a blind Time Division Multiple Access (TDMA) scheme. To further exploit the coupling between the MAC and the Physical layers, apart from the transmission powers and rates, parameters regarding space-time codes and signal combining can be exchanged between these two layers with the aim to improve the achievable rates.

- **Network Layer**: Taking the cross-layer issues one step further, the routing algorithms also need to take into account information regarding the physical layer as well as the scheduling decisions at any given time in order to send the packets through the routes that utilize links with the highest capacities. For example, decisions regarding to which relay node the traffic should be forwarded have to take into account the underlying physical layer.

- **Transport Layer**: A cross-layer design is also necessary to support users of real time applications since it is not clear how the higher layers should adapt to the continuously changing wireless environment and traffic conditions. Also, it is non-obvious how to maintain performance guarantees in networks with a large number of connections. For real time applications, apart from the requirement that packets must reach their destinations in short time, in order delivery is also of paramount importance. The prevalent protocol in the internet for reliable end-to-end transmission of data and congestion control is the TCP/IP. This protocol assumes that a packet was dropped by a router due to congestion in case it has a timeout without seeing the identity of this packet. As a reaction to that, it decreases the source rate and thus decreases the network throughput. Although TCP/IP, in its original design, is unable to account for losses due to the unreliability of the wireless medium, various adaptations of TCP/IP have been suggested in the literature. One such adaptation is to introduce the Explicit Congestion Notification (ECN) where an ECN bit is introduced in the packet header. The bit is initially set to zero by the source. In the event of congestion, the router sets the ECN bit to one to indicate congestion. Once the packet reaches the destination, the latter will inform the source that congestion occurred by checking the ECN bit and the source will be able to adapt its transmission rate. Clearly, providing end-to-end reliability is more difficult in the presence of multi-hop transmissions and cognitive radios since the transport protocols run at the end nodes and do not have knowledge of the underlying channel and traffic conditions or congestion of the intermediate nodes.

2.2.3 Systematic cross-layer design for future internet

Future internet requires employing new network topologies, such as multi-hop wireless networks, mobile ad-hoc networks, integrations of wireline and wireless networks. It further calls for advanced routing, scheduling, radio resource management, and radio link control solutions, capable of using the spectrum efficiently and adapting to diverse traffic behaviour. A fundamental problem in designing such complex systems is to derive a network control mechanism comprising of flow control, routing, scheduling, and physical resource management, which can ensure network stability under a large set of service demands and can, at the same time, provide a certain degree of fairness among concurrent sessions (corresponding to different users and/or services). Designing network protocols that achieve these goals requires a holistic network optimization, which cannot be achieved without crossing the boundary between the standard OSI layers. Network protocols for layered architectures have historically been obtained based only on an engineering intuition, and many of the recent cross-layer designs are conducted in the same ad-hoc manner. Despite of the recent progress, there is little understanding of how to divide the network protocol into layers in a systematic (i.e., optimal in some sense) rather than in an ad-hoc way. A completely new approach, based on a rigorous mathematical theory, is required for optimizing the future network architectures.

An important step towards a systematic cross-layer network protocol design was performed in the late nineties, when Kelly et. al. introduced the concept of *network utility maximization* (NUM) for providing fairness in wireline networks. The Kelly's NUM framework has been extended to model, analyze, and design various protocols and resource allocation algorithms. These extensions led to a novel design concept, known as *layering as optimization decomposition* (LOD), which can be summarized as follows: The network protocol is modelled as an optimization program which maximizes a network utility function, defined according to the needs of the main application. Standard optimization decomposition techniques (e.g., primal and dual decomposition) are employed to obtain different decompositions of the original NUM problem, and each one is mapped to a possible layering scheme. Specifically, each sub-problem resulting from a given decomposition scheme corresponds to a layer in the protocol stack. In this way, the different layers become well-defined optimization sub-problems instead of being products of ad-hoc decompositions. In addition, the interfaces between layers can be rigorously designed. The outputs of the layers are the primal and/or the Lagrange dual variables of the sub-problems they solve, and they represent the variables of a master problem. Different layers iterate on different subsets of the decision variables using local information to achieve individual optimality. Coordinated by the master program, these local algorithms attempt to achieve a global objective.

The above LOD approach provides a mathematically rigorous framework for systematic cross-layer network design. It is capable of specifically addressing the application needs (by an appropriate choice of the utility function), providing a globally optimal performance benchmark, and leading to a systematic design of a decomposed solution to attain the benchmark. By exploring the space of alternative decompositions and comparing the resulting algorithms, the network control designer can obtain answers to questions as "how to and how not to" layer. By modifying accordingly the utility functions and the constraint set, the framework can be further generalized to incorporate other degrees of freedom, such as multiple routes, beam-forming patterns, transmit powers, contention probabilities, channel codes, and modulation schemes.

Despite its generality, the standard NUM framework typically assumes that the user population remains static, with each user carrying an infinite backlog of packets that can be treated as a fluid, injected into a network with static connectivity and time invariant channels. While this model proved very successful in the case of fixed networks, it is not accurate enough for the future wireless (or hybrid) networks. In such networks, the links' capacities are time-varying due to the fading in the radio channel and the users arrive with a finite workload and depart when the jobs are finished. In addition, the network connectivity is also time-varying due to the users' mobility, battery limitations, etc. Therefore, creating a general framework, which leads to a systematic design of general, time-varying multi-hop wireless networks, is a challenging, and yet open problem. Specifically, stochastic networks with wireless and/or wireline components, randomly arriving traffic, and time-varying channels with possible disconnections and user mobility should be considered.

The concrete goal is to integrate various network protocol layers into a unified optimization framework, by considering them as distributed over the network components (or sub-problems) of a general utility maximization problem. Coordinated by a master program, these local algorithms attempt to achieve a global objective. By modelling the wireless network as a queuing system with transmission rates that depend on resource allocation decisions and time varying channel states, cross-layer control techniques for the flow control, routing, node scheduling, and radio resource allocation can be derived in order to achieve joint optimal performance. Particular emphasis should be on deriving algorithms and protocols that allow distributed implementation. Different performance criteria, including throughput, delay, and robustness against topology changes, can be investigated.

2.3 Research priorities in wireless technologies and applications for emerging growth markets

The emergence of economies in the Asia-Pacific, Indian sub-continent, Africa, and South America, notably those of Brazil, China, India, South Africa, open not only new markets for technologies, but also brings forth newer technological challenges and opportunities. The economic needs of these markets seek for relatively inexpensive, easily accessible, diversified, and expandable communication services, ranging from basic voice and data services to more rich in content multimedia and broadcasting operation, optimized for the masses at the bottom of the economic pyramid. Various broadband (wireless and wireline, including optical) access techniques for instance have attractive attributes like scalable service offering and flexibility to provide tradeoffs between service quality and availability with costs and service coverage. The challenge however lies in adapting the various technologies available in networking, computing, user interfaces, and content, and possibly creating new ones, and further developing innovative business models to cater to the specific needs of the remaining 4 billion people still untouched by the Internet. The challenges and the opportunities both are astounding.

Some of the key drivers calling for advanced R&D, targeted towards the needs of the emerging economies, are as follows:

- The majority of growth in the networking business is already happening in the emerging economies.
- Competition in these markets is going to be fierce since all manufacturers and operators are increasingly focusing on the same markets.
- True adoption of mobile internet experience will come from these markets since mobile handhelds have already higher penetration than personal computers (PCs), and most people in developing regions never had a PC-based Internet experience. The aggregate ownership of mobile phones (and subscriptions) in the emerging economies is in the order of billions, whereas the ownership of computers (desktops and laptops) is only in the millions. For example, in India, there are over 400M mobile users as compared to about 50M computer users while in China the mobile users exceed 600M as compared to the computer users, which reach 125M.
- Since the wireline infrastructure is very poor and most developing nations have already decided to cap their wireline expenditures in favor of going for broadband wireless, LTE and other non-cellular wireless access technologies (e.g., WiFi, WiMAX) will have a much faster speed of deployment and easier generational skip than in the developed economies.
- There is a huge pent-up demand in rural areas for improved quality of life that networking technologies can meet mainly because of the lack of transport, financial, healthcare, and entertainment infrastructure.
- The broadband wireless Internet will enable a plethora of new useful applications and services for a vast majority of subscribers in the lower middle-class and those at the bottom of the economic pyramid. These considerations call for greater investments in R&D specialized towards creating value propositions through creative business models and technologies tailored to emerging markets.

It is important to first summarize the key demographic, economic, and business reasons why we should pay particular attention to the needs of these markets to "fuel" the economies of the developed markets, especially in the present time of severe recession, and use this research as a driving impetus for the rest of the technical research agenda. We make a case for focused R&D targeted at communication technologies for these new growth markets. To start with, some of the incorrect notions associated with these markets need to be expounded and clarified:

- Solutions that exist for developed regions can, with trivial modifications, be applied for developing regions.
- The scope and necessity for technological innovation in emerging economies is limited.
- Only cheap products are suitable for emerging economies.
- An economic case cannot be made for offering networking technologies in developing regions.

- Users in the rural areas provide low average revenue per user (ARPU) and are not ready for new services and applications.

This will then be followed by determining the specific requirements for the wireless client devices, broadband internet access and services, and where the priorities should be from the perspectives of the network, service, and equipment providers. The identification of the research topics must also meet the challenges of minimizing costs (capex and opex) and sustaining the environment. More specifically, from the networking perspective, the key challenges are in efficient allocation and usage of the spectrum, serving extremely dense urban areas and outlying sparse rural areas, radio access and backhaul technologies that use the vastly deployed but unlit fiber, QoS support for multimedia and broadcast/multicast content, and ease of deployment, management and operations. Also the availability of network access can be discontinuous and highly time and position dependent, which needs to be taken into account at applications layers.

From a business perspective, it is suggested that the focus ought to be on meeting local needs, creating value, and not necessarily on maximizing profit margins in order to boost a faster uptake of the broadband services that will in turn result in higher investments in the infrastructure and equipment. The research will also prove that the solutions (both the business models and technical) for the emerging markets can equally be applied in the developed markets to reduce service costs to the subscriber and to maximize the return on investment for the service provider.

In conclusion, the research outlined above falls in four key areas: **Value Creation with Novel Services and Applications, Innovative System and Network Architectures and Solutions, Innovative Spectrum Usage Solutions, and Innovative Business Models (including public-private partnership).**

3 Recommendations

There are still many unexplored areas for optimisation of radio interface and advanced network architectures requiring major investment and research effort towards achieving efficient ubiquitous personal broadband communications. The key areas requiring research investments are in the development of

- Broadband Mobile Access Technologies
- Mobile Wireless Driven Future Internet Technologies
- Fundamental Enabling Networking Technologies
- Wireless Technologies and Applications for Emerging Growth Markets