Wireless Internet on Heterogeneous Networks

Paul J.M. Havinga
University of Twente, dept. of Computer Science
Enschede, the Netherlands
havinga@cs.utwente.nl

Gang Wu
Communications Research Laboratory
Yokosuka, Japan
g-wu@crl.go.jp

Abstract

The wide proliferation of wireless systems and the use of software radio technologies enables the employment of a heterogeneous network. In this concept services are delivered via the network that is most efficient for that service. The solution is based on a common core network that interconnects access points of various wireless access points. A mobile host can apply multiple different access networks simultaneously to increase capacity or efficiency. A basic access network, separated from other wireless access networks, is used as a means for wireless system discovery, signaling and paging. Quality of Service is of prominent importance due to the heterogeneous environment and the characteristics of the wireless channel. This paper describes the concepts of our architecture, and presents an overview of the architecture.

Keywords*: wireless Internet, Quality of Service, heterogeneous networks, multimedia.*

1. Introduction

The development of mobile devices with built-in wireless access to the Internet will have a major impact on the way we communicate. In particular, two trends are likely to have a major impact on future wireless communication. First, there will be a wide proliferation of various wireless access networks, each having their own preferred type of service, and second, the developments of software radio technologies which allows a mobile terminal to change the radio access mechanisms on demand.

It is expected that in future scenarios users will be able to choose from a wide range of services from various access networks. Today, we already have various first and second-generation wireless and wired networks. Since current access networks are designed for very specific services and applications, these networks usually satisfy the requirement of a certain type of service on a certain type of network. Although it is likely that these latest networks are able to offer a mix of services, they still have some special characteristics. Work on software radios has shown that wireless physical layers can be created dynamically by introducing code into programmable radios with tunable front-ends.

In a heterogeneous network it is possible to use a combination of several networks, each of which is optimized for some particular service. This results in a system that delivers each service *via the network that is most efficient for that service*. Access networks may also be combined to increase the available capacity. Different access networks might also be used for uplink and downlink traffic. This can be useful for user applications like web browsing and e-mail, which in many cases are asymmetrical in nature causing more downlink than uplink bandwidth.

This paper describes an architecture that is able to provide a broad mix of services over a heterogeneous wireless network. Work on this architecture is performed in the MIRAI project (Japanese for *future*, and an acronym of 'Multimedia Integrated network by Radio Access Innovation') at the Communications Research Laboratory in Yokosuka, Japan. The solution is based on a common core network that interconnects access points of various wireless access networks. A basic access network, separated from other wireless access networks, is used as a means for wireless system discovery, signaling and paging.

2. Models of heterogeneous networks

There are several architectures using multiple different wireless networks. The basic models are illustrated in Figure 1 by two wireless networks, network A and network B. The main distinction between these models is the layer on which the wireless networks communicate. Many derivates of these models are possible (see for example [13]).

A. Tunneled network – In this model, a user has a service agreement with operators of several wireless access networks independently. Based on some policy, the optimal network for the requested service is selected. The hybrid core tunnels the traffic across the Internet and the selected access network to the mobile host. This system requires no modification to existing access networks. Main disadvantage is that connectivity between networks is based on relatively high network layers of the Internet (i.e. transport layer), increasing service latency.

B. Hybrid network – In this model we have a hybrid core that interfaces directly between the wireless access networks and the Internet. In this model the wireless networks implement the network layer and below. Advantages are that this in model there will be less duplicate functions, and that it is able to offer advanced services at the network or data link layer (e.g. it can provide a better handover between the access networks).

C. Heterogeneous network – In this model there is a common core network that deals with all network functionality and operates as a single network. Different radio access networks handle only those functions specifically related to a distinct radio access technology (modulation, etc.). In general the wireless access radio incorporates the physical and link layer only. Communication between access networks belonging to the common core is based on lower network layers (link layer or network layer). This reduces the overhead, and improves performance. A major obstacle of this model is that the different access networks must converge, which requires a standardization effort and business commitment to support it.

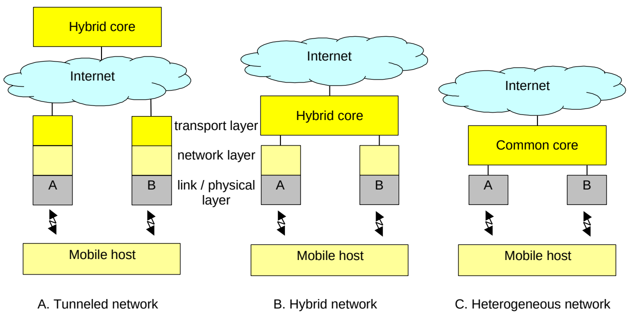

Figure 1: Architectural models.

Notice our differentiation between hybrid and heterogeneous. Often, the various kinds of architectures are all referred to as hybrid [13]. We prefer to call them *heterogeneous* to stress the fact that there can be multiple networks *simultaneously*, all working together. Hybrid networks describe the more traditional view of having multiple networks of which one can be chosen to use.

As described above, existing systems use their own network to provide services to mobile users. Moreover, they all have their own infrastructure for e.g. signaling, handover, and billing. This makes it very difficult for existing network systems to cooperate efficiently. Convergence of current wireless systems is practically only possible at the higher layers of the network protocol stack. The only model suitable is the tunneled home network (model A), in which internetworking between the different systems is based on the Internet, for example based on Mobile IP. The main disadvantages of this solution are that many functions are duplicated, and because of its inefficiency, internetworking between the access networks is not convenient, and seamless handover between different access networks are likely to be impossible.

3. Related work

Future wireless network infrastructures will have to support a wide variety of users, applications, and access needs. High-speed access can be achieved by using small cell sizes. As base station density increases, however, so will handoff rates. Currently, related work mainly is associated with routing and handoff aspects for wireless networks. The Mobile IP protocol [10] supports mobility transparently above the IP level and it allows the nodes to change their location. Mobile IP is generally seen as a *macro mobility* solution, it is less well suited for micro-mobility management, in which a mobile host moves within a sub network. A typical example of micro mobility is a handoff amongst neighbor wireless transceivers, each of which is covering only a very small geographical area. There have been quite a few proposals to support micro mobility (e.g. Cellular IP [3], HAWAII [11]). The differences among all these schemes are related to the mechanisms used to route the packets within a local (home or foreign) domain.

Related work on QoS over Internet is mainly based on Integrated Services [2] and Differentiated Services [1]. Recently, there are some initiatives that specifically relate to heterogeneous networks, but this research just started [13]. Other related research is mainly focusing on hybrid network architectures, or support for macro-mobility [4][9]. Given ATM is able to support QoS, there has been strong interest in developing wireless ATM technologies (e.g. the Magic Wand project [8]).

Current work merely provides solutions to roaming mobile hosts by supporting protocols for mobility. Heterogeneous networks might be used, but more in the traditional sense of selecting one or the other. This is different from our view of heterogeneous networks, in which mobile hosts can communicate over one single access network, as well as over multiple access networks simultaneously.

4. Concepts

The heterogeneous network architecture that we introduce here is intended to be a flexible and open architecture for a large variety of different wireless access technologies, for applications with different QoS demands, and different protocols. A fundamental goal is to make the heterogeneous network transparent to the user. In addition, a goal is to design the system architecture such that it is independent of the wireless access technology. These considerations lead to various requirements like mobility management for seamless handover, energy efficiency, mechanisms to select the most efficient configuration, and QoS mechanisms that are compatible with existing mechanisms. In the MIRAI architecture we try as much as possible to build upon existing protocols to minimize our required effort, and to be compatible with existing protocols and applications.

4.1. Basic entities

Our solution to these requirements is based on three major entities:

- *A multi-service terminal.* The multi-service terminal is equipped with a multi radio system. Any terminal has the *common access component* to communicate with the basic access network. Apart from this radio system, the multi-service terminal is equipped with one or more radio subsystems to access the common core network. These subsystems are essentially (or preferably) based on software-defined radio technologies, which allow the terminal to adapt its radio hardware to the wireless infrastructure available and required.

- *Basic Access Network*. This basic access provides the common access of all multi-service terminals to the common platform. The network is basically used to provide the signaling, wireless system discovery, and location management for all other wireless systems. To support this, it should have a large coverage area, preferably larger than the wireless access networks it supports. The common access network is accessed using the *common access component*.
- *Common core network*. This network provides the common platform through which all multi-service terminals communicate with correspondent nodes residing in external networks. In principal all access points of the wireless access networks are connected to this network. The network provides routing and seamless handovers between the wireless networks. In this way a natural integration of the various heterogeneous networks is achieved. The main functional entity of the common core network is the *Resource Manager*, which coordinates the traffic distribution, and selects the access network. It has a common database for managing users' profiles with entries like authentication, preferred access system, billing, policy, users' terminal capabilities, etc.

4.2. *The network model*

The MIRAI architecture provides communication between mobile hosts and correspondent nodes residing in external networks. Figure 2 gives a conceptual overview of the architecture.

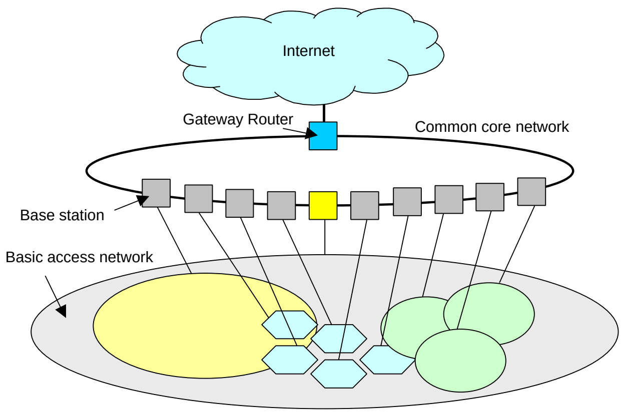

Figure 2: MIRAI architectural concept.

The universal component in our architecture is the *base station*, which serves as the wireless access point and interfaces with the common core network. Common core networks are connected to the Internet via *gateway routers*. A common core network provides services for several wireless access networks. In general the wireless networks will overlap, and a mobile host can have access to several networks at one location. The area covered by these wireless networks can be quite large, covering for example a metropolitan area like Tokyo.

Mobile IPv6 is the envisioned protocol for the connecting the common core networks and providing macro mobility. The common core network with high-speed wireless access with frequent location updates requires micro mobility. Mobile hosts attached to a base station use the IP address of the gateway as their Mobile IP care-of address. Inside the common core network, mobile hosts are identified by their home address. Base stations are connected to (or integrated with) a regular IP forwarding engine. These engines are connected in some network topology that allows packets to be transmitted between the base stations and the gateway.

Note that, although in our concept the base station is equal to a wireless access point, there is no need to dogmatically stay on this. It is quite well possible that some wireless access providers use their own network of interconnected access points, and share one base station to connect to the core network.

An important concept in our architecture is simplicity, thereby enabling low-cost implementation of the network. The concept of common core network and separate basic access network offers providers of wireless services the possibility to setup an infrastructure with little investments. New providers can easily connect to the core network, provided that they use the correct interface. They do not need to have their own infrastructure ready before they can start their business, but instead use the infrastructure provided by the core and basic access network. All they have to do is to develop their wireless service, and concentrate on the wireless access only. The infrastructure that is generally needed to setup a whole new service is already provided by the architecture. This includes both technical issues (like interconnection network between base stations, routing, handoff, providing internet access) as well as business issues (like billing, and managing consumer profiles). They just need to concentrate on how to implement the service they want to deliver as efficient as possible, and then just plug it into the infrastructure. The components they have to build are the base stations, and the access mechanism for the terminal. In general the access mechanism can be a software module suitable to be used in a software-defined radio.

A consumer may have a contract with the common core network provider, and buy various services (provided by a wireless access network) from it. If the consumer has a contract that enables him to use multiple services, then the system and the user is able to select the most appropriate service. Access networks may also be combined to increase the available capacity. Different access networks might also be used for uplink and downlink traffic. This can be advantageous for user applications like web browsing and e-mail, which in many cases are asymmetrical in nature causing more downlink than uplink bandwidth. The result is that each service is delivered via the network that is most efficient (in many perspectives) to support the service. In effect, the consumer is unaware of the wireless technologies used to provide the service.

4.3. Basic Access Network

In our architecture we use two distinct networks: the basic access network for all signaling related traffic, and the common core network for the data traffic itself. We believe that it is very important to separate the signaling network and the real data-traffic network for several reasons.

- First of all, the basic access network can provide *wireless system discovery*. The basic access network delivers the common access; every mobile host is capable of using this basic access network. The network provides the terminal with information about the currently available wireless networks, so that the terminal is not forced to scan all possible variants.
- The Basic Access Network is used as the *signaling network* to provide mobility. Such a dedicated network can provide this service efficiently and securely.
- The basic network is also used for *paging*. In a mobile environment it is very important to be energy efficient since a terminal relies on its batteries to operate. We expect that the wireless IP communicators will be on-line continuously, although not be actively communicating most of the time. In essence, mobile hosts will be in an idle state, but passively connected to the network infrastructure. It is then extremely inefficient to have to scan all networks, and wait for a paging message. Moreover, since the wireless networks are optimized for some special services, it is very likely that they are not efficient for paging messages. A wireless network that is optimized for this kind of traffic is more efficient [7].
- The basic access network can provide the infrastructure to allow mobile hosts to determine their *location*. This information can on its turn be used by the basic access network to provide the mobile the information about the available services in its region. Location management is further important for roaming and paging.
- The basic network is used as the medium for most signaling and control messages. This *simplifies* the design of a new wireless access service, since the signaling is being performed by another entity (the common access component).
- Since we have a heterogeneous architecture in which multiple access networks can be used (semi) simultaneously, we need to have some *network access synchronization mechanism* so that a terminal knows when to tune the software radio to another access network. The basic access network can provide such a service straightforwardly.
- Finally, the basic access network can also be used as *a wireless access service*. It is, however, primarily suitable to provide some very low bandwidth messaging services like paging and broadcasting.

Since the basic access network is mainly used for small messages, the speed is of less importance; the total capacity, however, must be sufficient for a large number of mobile hosts.

5. Quality of Service mechanisms

5.1. End-to-end QoS

Enabling *end-to-end QoS over Internet* is a tough venture, because it introduces complexity starting from applications, different networking layers and network architectures, but also in network management and business models [12]. It becomes even tougher when one is introducing QoS in an environment of mobile hosts, wireless networks, and different access technologies. Yet the need for QoS mechanisms in this environment is greater due to scarce resources, unpredictable available bandwidth and variable error rates. In a heterogeneous network in which we apply different wireless networks with different characteristics, the need for QoS mechanisms is evident.

Within the fixed Internet there are several ways to enable end-to-end QoS. Current work on QoS over IP architectures, i.e. Integrated Services and Differentiated Services seems to leave out mobility support, despite its importance. There are some initiatives that try to integrate the existing QoS over IP architectures with protocols supporting mobility [12], but these are still in development.

The QoS approach can be divided into two parts: core network QoS and fixed network QoS services. In this way, the wireless IP (core) network is compatible with the fixed state-of-the-art QoS solutions. The gateway router provides merely the mapping between the Internet and the core network.

5.2. Differentiating traffic flows

All IP communication is packet based relying on connectionless transmission. The addressing scheme does not enable the system to differentiate traffic flows. The term traffic flow refers to the flow of IP packets that belong to the same connection, i.e. the IP packets that are sent between particular applications (port) and between particular hosts (IP addresses).

Traffic flows within the common core network are differentiated according to its service needs and QoS requirements. There are two main reasons for having such differentiated traffic flows:

1) To enable routing – In our concept a mobile host can have multiple flows for one or more services that use several different wireless access networks. Each access network is used for the kind of service for which it is optimized. Packets of different services between the mobile host and the corresponding node may thus use a different route (i.e. a different base station and a different access network) over the common core network. The mobile host may thus simultaneously use one or more base stations to connect with the common core network. This implies that traffic between the mobile host and the different base stations must be distinguishable according to their required service.

2*) To enable cross-layer interaction* – In a wireless environment it is essential that the lower layer protocols are aware of the traffic characteristics. The Internet is implemented on the basis of the ISO/OSI hierarchy architecture, where the protocols for different layers are independent of each other. Recent research has shown that for wireless Internet, information of other layers may be required in order to improve the overall performance [14]. For example, although the TCP specification contains no explicit reference to the characteristics of the lower layers, implicitly in the timeout and retransmission mechanisms there is the assumption that the error rate is low, and that lost packets occur due to network congestion. TCP has no way of distinguishing between packets corrupted by bit errors in the wireless channel from packets that are lost due to congestion in the network. Another example, for designing a wireless MAC and data link protocol, it is more efficient if the traffic characteristics are known in the MAC and data link layer [6]. Also, in W-CDMA systems, power control can be used to meet the different QoS requirements for different traffic. In other words, they all suppose to be able to know the traffic types even in the physical layer. These examples attest the need to tailor protocols to the environment they operate in. Separating the design of the protocol from the context in which it exists leads to penalties in performance and energy consumption [6] that are unacceptable for wireless, multimedia applications.

5.3. Flow detection

So, differentiating traffic flows is needed and useful. However, it is not trivial how to detect these flows, and how to determine the QoS requirements for these flows. We can distinguish two major classes: *explicit differentiation* by using an application level signaling protocol, or *implicit differentiation* based on the traffic class.

Explicit differentiation – Due to the virtues and future potential IPv6 is selected as a protocol framework. An important IPv6 feature is the introduction of flow labels to enable the labeling of packets belonging to particular *traffic flows* for which the sender requests special handling, such as non-default quality of service or real-time service [5]. It is currently not clear what level of granularity will be provided via the flow label. It is likely that many real-time applications will not employ the flow label, yet they want more than best-effort service. Another issue in this case is to decide what QoS to provide this flow, since no QoS information is provided.

Implicit differentiation – Implicit flow detection can be based on various mechanisms. For example, the Diffserv QoS class can be mapped to the appropriate wireless QoS. IPv6 also has an 8-bit Traffic Class field in its header. This field is available for use by originating nodes and/or forwarding routers to identify and distinguish between different classes or priorities of IPv6 packets. At this moment it is not clear how this field will be employed, but there are a number of experiments underway to provide various forms of 'differentiated service' for IP packets, other than through the use of explicit flow set-up [5]. Alternatively, one could also monitor the transport layer port numbers and forward IP datagrams with WWW or FTP traffic in a non best-effort fashion [8].

5.4. Discussion

Within the common core network we need to have a mechanism to route the flows to the mobile using an efficient wireless network, and a mechanism to exchange information between various layers about the different QoS requirements. Due to the virtues and future potential IPv6 is selected as a protocol framework. Explicit discrimination is probably the only viable solution for a mobile terminal. Applications that do not use explicit discrimination will receive a best-effort treatment. Traffic originating from the external network can be discriminated using either explicit or implicit discrimination mechanisms.

Flow discrimination and routing within the common core network is based on the network layer (i.e. IP). This enables the use of other proposals to support micro-mobility, like HAWAII and Cellular IP. Encapsulation of all datagrams in a new IP datagram within the common core network is likely to be the most appropriate solution, since in this way we can achieve the most uniformly access mechanism, and applications or services residing on the external network do not have to be adapted (albeit they can benefit from applying explicit discrimination mechanisms).

6. MIRAI architecture

In this section we will introduce the functional entities of our architecture and the protocols used. The architecture as depicted in Figure 3 consists of four major building blocks: mobile hosts, wireless access networks, common core network, and external network.

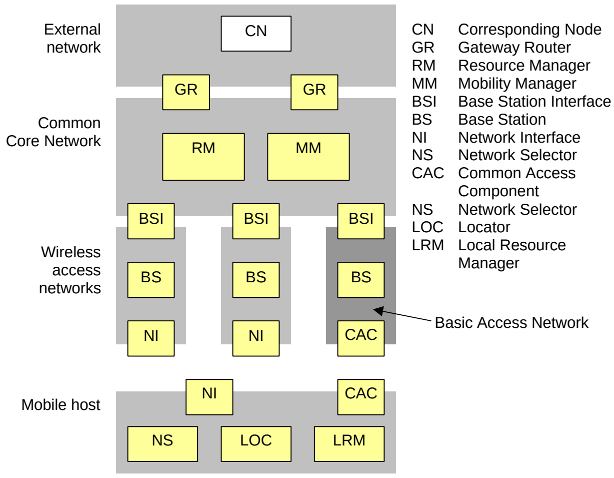

Figure 3: Heterogeneous system architecture.

Within the external network is the *Corresponding Node* (CN). One or more *Gateway Routers* (GR) connects the external network to the common core network. At the external network Mobile IP is assumed. The Gateway Router plays an active role in this as, when tunneled packets arrive at the gateway destined for a mobile host, the gateway detunnels the packets and forwards them toward a base station. Two important functional entities within the common core network are the *Resource Manager* (RM) and the *Mobility Manager* (MM). They are primarily responsible for traffic distribution and mobility related issues.

The common core network supports the communication to the base stations, and thus to the wireless access networks. The *Base Station Interface* (BSI) is primarily used to provide a uniform access mechanism for base stations to the common core network. The BSI can be a component of a base station. The *Base Station* (BS) deals with the normal link layer issues of wireless access and collects status information of the wireless network it supports. It uses the *Network Interface* (NI) to access the network.

A primary component of a mobile host is the *Common Access Component* (CAC) to communicate with the basic access network. Besides this interface, it also has a Network Interface. However, in contrast to the NI of the base station, this interface is in general based on software radio technologies to enable it to use multiple wireless access networks. A *Network Selector* (NS) communicates with the Resource Manager to tune the radio for the wireless access network to use. A *Network Selection Control* protocol is used to enable the proper selection of the access network. The *Locator* (LOC) provides the RM with location information of the mobile host. The *Local Resource Manager* (LRM) deals with the local resources of the terminal, and interacts with the resource Manager at the common core network.

7. Conclusion

The presented architecture shows a novel approach to enable the efficient use of available wireless access networks. The basic concept is that each service is delivered via the network that is most efficient to support the service. The result is that the mobile user receives the requested service at the lowest cost, and that the scarce radio resources are used efficiently. The architecture solves many of the basic problems involved with wireless Internet over heterogeneous networks.

The presented mobility and QoS management schemes are compatible with fixed Mobile IPv6 network techniques. Moreover, these techniques are combined with micro-cellular mobility solutions, and allow interaction between fixed network QoS techniques and wireless QoS techniques (below the network layer). This enables the wireless network to support IP packets with varying IP QoS parameters properly. As the core network QoS and flow management is independent of the deployed IP protocol suite, the system can be enhanced to support alternative IP techniques, such as differentiated services. Since the architecture is built upon existing protocols and architectures for mobile computing, making a prototype of it at a smaller scale than the originally envisioned common core network, should not pose serious obstacles.

References

[1] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Zh., Weiss, W., “An architecture for Differentiated Services”, IETF RFC 2475, 1998.

[2] Braden, R., Clark, D., Shenker, S., “Integrated Services in the Internet Architecture: An Overview”, IETF RFC 1633, 1994.

[3] Campbell A.T., Gomez, J., Kim S., Turanyi, Z., Wan, C-Y., Valko, A: “Design, Implementation and Evaluation of Cellular IP”, IEEE Personal Communications, Special Issue on IP-based Mobile Telecommunications Networks, Vol. 7 No. 4, pg. 42-49, August 2000, http://www.ctr.columbia.edu/~campbell/andrew/publications/papers/pcs2000.pdf.

[4] Daedalus project, Berkeley, http://daedalus.cs.berkeley.edu.

[5] Deering S., Hinden R.: “Internet Protocol, Version 6 (IPv6), Specification”, IETF RFC 2460, December 1998. http://www.ietf.org/rfc/rfc2460.txt.

[6] Havinga P.J.M., Smit G.J.M., Bos M.: “Energy efficient wireless ATM design”, ACM/Baltzer Journal on Mobile Networks and Applications (MONET), Special issue on Wireless Mobile ATM technologies, Vol. 5, No 2., 2000, pp. 147-155.

[7] Havinga P.J.M., Smit G.J.M.: “Energy-efficient wireless networking for multimedia applications”, Wireless Communications and Mobile Computing, Wiley, 2001.

[8] Magic Wand project, http://www.tik.ee.ethz.ch/~wand/.

[9] Monarch project, CMU, http://www.monarch.cs.cmu.edu.

[10] Perkins C., IP Mobility Support, RFC 2002, October 1996, http://www.ietf.org/rfc/rfc2002.txt.

[11] Ramjee R., La Porta T.F., Salgarelli L., Thuel S., Varadhan K., Li L.: “IP-based Access network infrastructure for next-generation wireless data networks”, IEEE Personal Communications, pp. 34-41, August 2000.

[12] Rexhepi, V., Karagiannis, G., Heijenk, G., “A Framework for QoS & Mobility in the Internet Next Generation”, Proceedings EUNICE 2000, Sixth EUNICE Open European Summer School, University of Twente, Enschede, the Netherlands, September 13 - 15, 2000, http://ing.ctit.utwente.nl/WU4/Documents/QoSMobframework.pdf.

[13] Walsh R., Xu L., Paila T.: “Hybrid networks – a step beyond 3G”, third international symposium on Wireless Personal Multimedia Communications (WPMC’00), pp. 109-114, Bangkok, Thailand, November 2000.

[14] Wu G., Bai Y., Lai J., Ogielski A.: “Interactions between TCP and RLP in Wireless Internet”, Global Telecommunications Conference, Globecom’99, pp. 661-666, 1999.