appeared in Proc. of 6th World Wireless Research Forum, London, GB, June 2002

Wireless World Research Forum (WWRF)

(a) Title of the research item:
Concepts for a QoS-enabled, TCP/IP-oriented Link Layer in Beyond 3G Networks

Experts:

Horst Miesmer, Henning Sanneck	Ana Aguiar, Holger Karl
Siemens Information and Communication Mobile - Networks, D-81359 Munich, Germany *{horst.miesmer,henning.sanneck}@icn.siemens.de*	Telecommunication Networks Groups Technical University Berlin *{aaguiar,karl}@ee.tu-berlin.de*

(b) Subject Area: WG4: Spectrum, New Air Interfaces and Ad-Hoc Networking

(c) Objectives of the required research (Why has the topic been chosen? Where will the results be applied?)

Current mobile networks of the second and third generation will at some point in the future – approximately starting in about five years from now – not be able to cope economically with the huge expected growth of the IP traffic volume. This is because they have only a limited capacity concerning the aggregate cell data rate of about 2 Mbit/s at the maximum. Enhanced 3G interfaces like High Speed Downlink Packet Access (HSDPA) aim at an aggregate cell data rate of 10 Mbit/s for the downlink and for the uplink at an aggregate cell data rate far below that. It can be expected that even this will not be sufficient depending on the future user density per cell and overall traffic load.

A lot of international research activities are currently targeted at improving the physical layer of future radio networks and are based on, e.g., multi-carrier technologies like OFDM or Multi-Carrier CDMA. These investigations strive to increase the aggregate cell data rate while allowing a higher degree of mobility compared to previous generations of mobile networks (including current wireless local area networks). Attempting to improve the raw physical bandwidth is certainly necessary, but should in our view be complemented by efforts to:

- better accommodate the properties of TCP/IP traffic
- better take into account the QoS requirements of applications.

These efforts should thus help to better exploit the capabilities offered by the new physical layer radio technologies. We therefore believe that components at the link layer will be crucial for the design of a new cellular radio interface which complements the existing and emerging radio access technologies (see Figure 1).

TCP/IP-orientation

2G and 3G mobile networks have been originally developed from a voice-centric but not from a TCP/IP-centric starting point and consequently, the TCP/IP support in these networks lacks in efficiency. Therefore, in order to support TCP/IP in an adequate fashion for future air interfaces, improved mechanisms should be used where the radio-specific technologies meet Internet-based technologies.

Application- / QoS-orientation

An additional challenge for future mobile radio networks will be the support of "advanced" applications: Traditionally, either real-time applications such as voice or file-transfer-type applications such as electronic mail have been considered; applications such as SMS represent a first step towards some slightly different applications semantics. It is to expected that future mobile networks will carry a much more diverse set of applications, including, but not limited to, networked games or access to distributed file systems. While all these applications could – in principle – be handled by traditional UDP or TCP transport protocols alone, it is unlikely that such a treatment would achieve the best possible application QoS or optimal resource usage without taking the link layer into account.

For simplicity of reference, all these types of applications will be summarized in the remainder of this text as TCP/IP traffic, with the understanding that also traffic types with quite different QoS requirements is included in this notion.

Key issues

The requirements of TCP/IP traffic are faced with the limitations of a highly error-prone channel. Because of the bursty error characteristic of radio channels, not all wireless terminals experience a good channel at the same time. Instead of attempting to average out this channel variability, the channel resource could be assigned to terminals that at this moment observe a good channel. In this sense, transmission scheduling depending on the current state of the various radio channels within a cell is a promising technique.

Therefore, a crucial issue is going to be how to handle the time-variable nature of a wireless channel and match this impediment with the applications's traffic requirements – we conjecture that **prediction-based scheduling** of traffic flows is a pivotal technique. Being able to predict the channel quality for a practically relevant time frame allows to adapt transmission parameters, taking into account the particular requirements of a particular traffic flow. This is essentially a scheduling decision between multiple flows and with a number of degrees of freedom. In particular, the quality of the channel prediction will be important to adequately support packet traffic in general and TCP/IP traffic in particular.

Depending on the flow type and traffic mixture to be supported, the scheduler can take very different actions. For a TCP flow, e.g., the scheduler might take into account the latency budget of a particular packet and the imminent expiration of a peer entities' retransmission timer, resulting in the scheduler assigning a corresponding priority to such a packet. For a real-time flow such as voice data, the importance of a packet for the subjective quality of the transmission could be consulted to decide whether to just drop the packet or try to overcome a currently bad channel situation by brute force. Such mechanisms could considerably improve overall performance, but require a careful system design and **exchange of information between layers**. For example, it would be unadvisable for the scheduler to duplicate and update the entire TCP state of all connections even though it might need to be able to access some of this information.

An air interface compatible with such an approach and capable to support a wide mix of TCP/IP-based applications would likely be quite different from current radio air interfaces. For example, the rather connection-oriented nature of current systems is an ill fit with basic IP properties such as statistical multiplexing: if a particular terminal temporarily experiences a bad channel, the resources (e.g., time slots) assigned to this terminal cannot be easily assigned to and used by another terminal. Additional flexibility here would require deviating from such traditional principles in the hope of better matching the requirements of new types of applications and exploiting sophisticated channel prediction techniques. A likely consequence would be a **connection-less architecture** for a radio interface; it remains to be investigated how this assumption interacts with the support of high mobility.

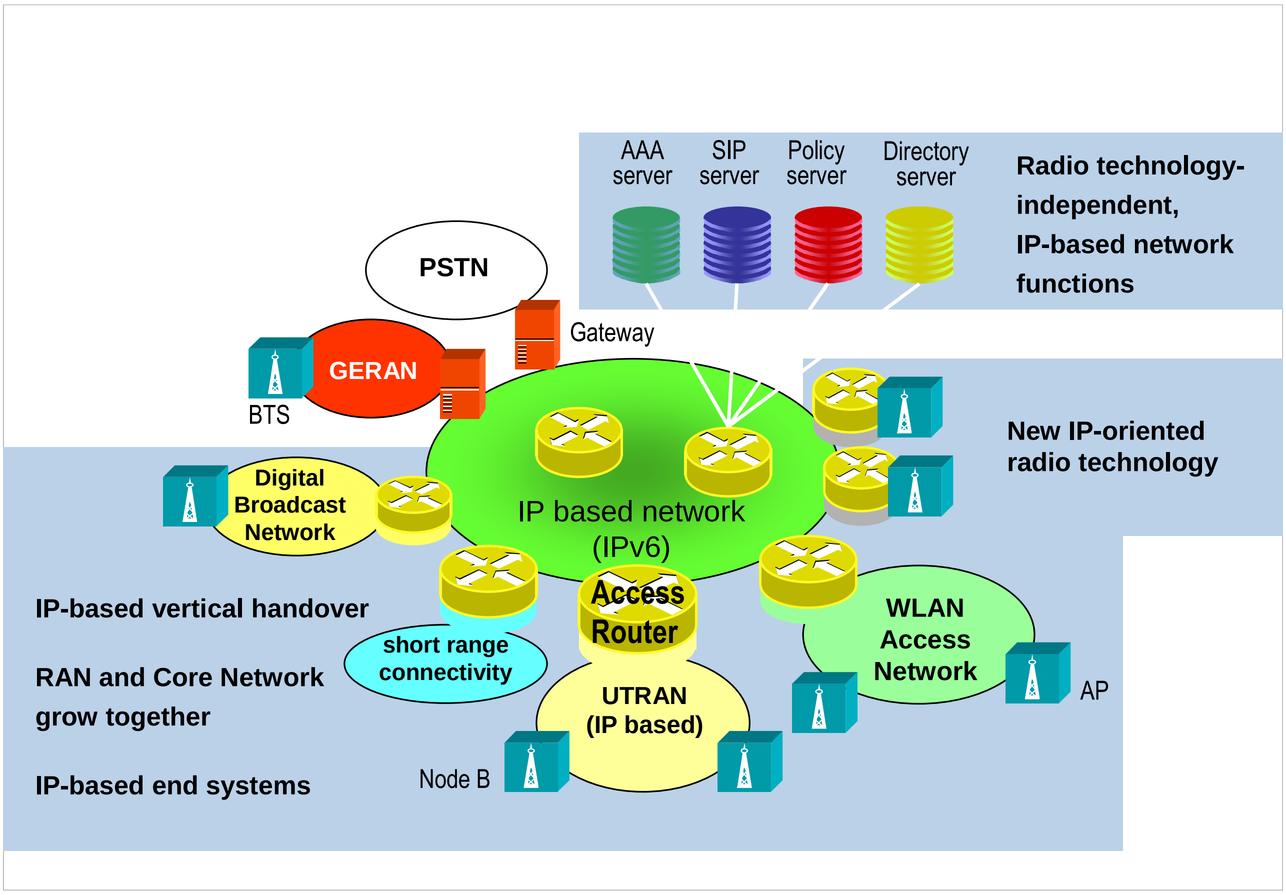

Figure 1 The new TCP/IP oriented cellular radio interface as a complementary building block within the future network of networks beyond 3G

(d) State of the art in the area (including important references)

Previous research to enable TCP/IP- and QoS-oriented system architectures has mainly focused on the adaptation of TCP/IP-based applications to given air interfaces. Two main research directions can be identified:

1. Proxy-type solutions (performance enhancing proxies (PEPs), see Figure 2 , upper protocol stack). This would mean, e.g., to split an end-to-end TCP connection into a wireline and a wireless part.
2. Convergence layer between the TCP/IP stack and the air interface (see Figure 2 , middle protocol stack). The convergence layer [EH00] is employed to shield the TCP/IP stack from specifics of the underlying link layer, yet at the same time to make an efficient cooperation of TCP/IP and the link layer possible. Main tasks of the convergence layer are QoS mapping and negotiation as well as L2/3 address assignment.

Previous research in channel-state dependent scheduling has often considered the goal of fair queueing: guaranteeing a (possibly weighted) fair share of the radio resources to every terminal [LU97, EU98, LU00, FR98, VA00]. The scheduler assures that terminals with an intermittently bad channel state are compensated for once the channel state is good again. Some of this research also has considered the integration with MAC protocols acting as a channel-state feedback; their overhead, however, has not yet been evaluated. In general, the existing algorithms are very complex to the point of rendering them impractical or exhibit undesirable features such as penalizing flows without errors.

The Havanna scheme [GO99] presents an integrated solution to deliver QoS over a wireless channel. It consists of a predictor, a compensator and an adaptor. The first predicts the channel state, which is used to decide on transmitting or not; the compensator uses a credit system, like the one described earlier but with only negative credits, to compensate flows for bad channel periods; and the latter consists on a buffer controller and a regulator, working together to deal with long periods of bad channel state (sort of channel state aware flow control). The performance highly depends on the channel prediction method, which requires a contention-free medium access.

The Wireless IP Framework [ST01, FA01] is also a centralised concept based on accurate prediction of the channel state for up to 10ms in the future. The timeslots are allocated according to the flows’ target error rates, taking channel state and adaptive modulation and coding into account. The chosen scheduling algorithm provides compensation on a (scheduling-)round per round basis, but not over longer timeframes. Besides, there are no bandwidth or delay guarantees. An ARQ protocol combined with adaptive punctured coding and modulation, where only the bits needed for the next coding level are retransmitted, is used to solve the problems of packet errors.

These two schemes are more realistic than the fair-queueing approaches, but their performance is highly dependent on the prediction accuracy. In addition, most of the schemes are centralized and concentrate on downlink traffic. This choice of focus also does not necessarily match with the system and traffic requirements expected to be found in Beyond 3G networks.

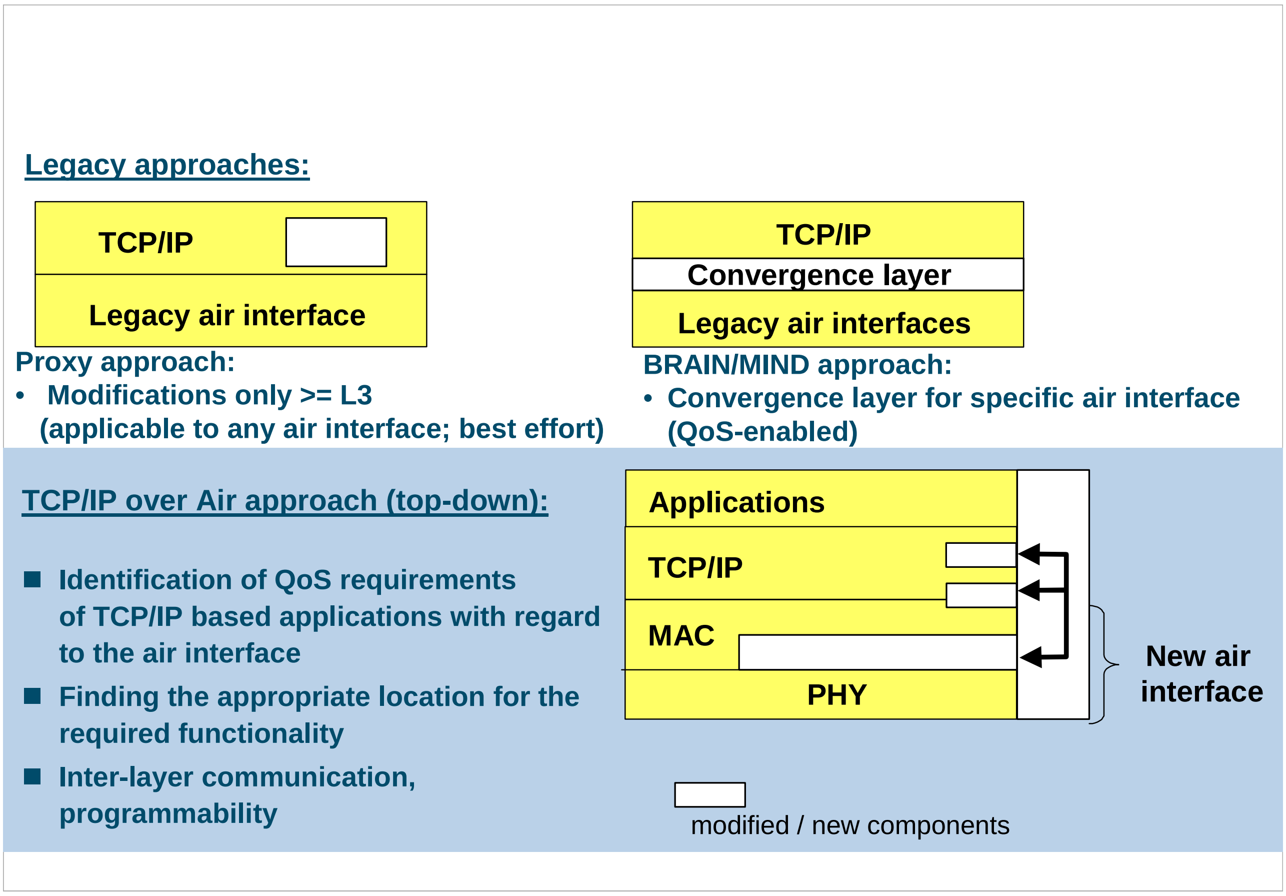

Figure 2 Comparison of legacy and new approaches for TCP/IP support over the air interface

(e) Possible approach

We aim at enabling a top-down approach. First, major requirements of the network up to the application layers should be identified which have an impact on the air interface design. Emphasis should be put on the efficient support of IP-based real-time services (VoIP), particulary with regard to delay, and on applications with growing relevance like network gaming and NFS (network file system). The goal here is to bring the advantages, like statistical multiplexing, of IP to the air. Building on the requirements of such applications, some key functions relevant for TCP/IP-over-air performance should be identified. Such key functions are, e.g., the identification of different flow types, the extraction of relevant protocol state (e.g., of a TCP connection which is about to trigger a retransmission timeout) and the matching of the QoS requirements of the different flows with the current channel state. Then we propose to look for the best location in the protocol stack where such a key function should be enabled (e.g., the identification of flow types is done between layer 2 and 3, extracting TCP protocol state would be done over an API accessing layer 4 and matching the QoS requirements could be done best by a predictor which helps to map information about the current channel state to scheduling decisions for IP packets). Finally, it is important to assure the communication between the key functions by enabling translucent inter-layer communication. That means that all information about a certain flow is concentrated at a single location and can be used by the respective layers.

(f) Expected results

- Definition of application's QoS requirements on lower protocol layers (particularly on the link layer)
- Analysis of predictor and scheduler as a potentially central element of the architecture; evaluation of the impact of prediction on scheduling efficiency
- Development of concepts for inter-layer communication; impact on the system architecture

(g) Time frame to get the expected results

- Requirements on the lower layers: 2003
- TCP/IP-oriented, QoS-enabled link layer (predictor, scheduler): 2004-2005
- System architecture / concepts for inter-layer communication: 2004

List of References

[EH00] P. Eardley, R. Hancock, "Modular IP Architectures For Wireless Mobile Access", BRAIN Workshop, London, November 2000

[EU98] T. S. Eugene Ng, I. Stoica, H. Zhang, "Packet Fair Queueing Algorithms for Wireless Networks with Location-Dependent Errors", in Proceedings of INFOCOM '98, 1998.

[FA01] S. Falahati, N.C. Ericsson, A. Ahlén and A. Svensson, "Hybrid type-II ARQ/AMS and Scheduling using Channel Prediction on Fading Channels", in Nordic Radio Symposium 2001, 2001.

[FR98] C. Fragouli, V. Sivaraman, M. Srivastava, "Controlled Multimedia Wireless Link Sharing via Enhanced Class-based Queueinf with Channel-state-dependent Packet Scheduling", in Proceedings of INFOCOM'98, 1998.

[GO99] J. Gomez, A.T. Campbell and H. Morikawa, "The Havana Framework: Supporting Application and Channel Dependent QOS in Wireless Network", in Proceedings of the 7th International Conference on Network Protocols, 1999.

[LU97] S. Lu, K.W. Lee, V. Bharghavan, "Adaptive Service in Mobile Computing Environments", in Proceedings of IWQoS'97, 1997.

[LU00] S. Lu, T. Nandagopal, V. Bharghavan, "Design and Analysis of an Algorithm for Fair Service in Error Prone Channels", in ACM/Baltzer Wireless Networks Journal, Vol. 6(4), pp. 323-343, Aug. 2000.

[ST01] M. Sternad, A. Svennson, A. Ahlén, T. Ottoson, N. Ericsson, S. Falahati, A.Ewerlid, B. Mielczarek, "PCC Wireless IP - Optimizing Throughput and QoS over Fading Channels", in Nordic Radio Sympusium 2001, 2001.

[VA00] N. H. Vaidya, P. Bahl, S. Gupta, "Distributed Fair Scheduling in a Wireless LAN", in Proceedings of 6th Annual International Conference on Mobile Computing and Networking, 2000.