

Joint optimisation of multimedia transmission over an IP wired/wireless link*

Catherine Lamy-Bergot, Jyrki Huusko, Maria G. Martini, Peter Amon, Cyril Bergeron, Pierre Hammes, Gábor Jeney, Soon Xin Ng, Gianmarco Panza, Johannes Peltola and

Filippo Sidoti[†]

contact @ ist-phenix.org

ABSTRACT

Multimedia applications such as video conference, digital video broadcasting (DVB), and streaming video and audio have been gaining popularity during the last years, growing also among mobile users. Quality of Service requirements for multimedia over wireless raises huge challenges not only concerning the physical bandwidth, but also the network design and services. This has to be addressed by future 4G systems. To this end, joint source and channel coding (JSCC) has provided promising results; however, the proposed mechanisms typically do not provide IP network support. In this paper, we present an architecture for JSCC systems designed for an IP network with end-to-end overall optimisation. All necessary mechanisms for transmitting the joint optimisation control information over wired and wireless network using cross-layer design are also described. Finally, partial simulation results obtained with this system design and H.264/AVC video input are presented.

Keywords

Joint source channel coding and decoding, multimedia transmission, cross-layer design, IPv6, mobility, 802.11 WLAN, medium access control, QoS, robust header compression.

*This work has been carried out within FP6 IST-001812 PHOENIX project, which was partially funded by the European Commission within the EU 6th Framework Programme and Information Society Technologies.

[†]Catherine Lamy-Bergot, Cyril Bergeron and Pierre Hammes are with THALES Communications, Colombes, France. Jyrki Huusko and Johannes Peltola are with VTT Technical Research Centre of Finland, Oulu, Finland. Maria G. Martini is with CNIT/University of Bologna, Italy. Peter Amon is with Siemens Corporate Technology, I&C, Munich, Germany. Gábor Jeney is with Budapest University of Technology and Economics, Hungary. Soon Xin Ng is with University of Southampton, U.K. Gianmarco Panza is with CEFRIEL/Politecnico Milano, Milan, Italy. Filippo Sidoti is with WIND, Italy.

1. INTRODUCTION

The evolution of wireless telecommunication systems can be divided into short term and long term evolution towards a global integrated system meeting the requirements of both users and industrial world, and making efficient use of emerging technologies. Expectations of evolution, whether called Beyond 3G for short term, or 4G for long term, are similar in many ways. For end-users, they include for example good service quality, easy access to applications and services, enhanced security and reasonable cost. Similarly, on the service and network provider side simple quality of service (QoS) and security management, flexibility of system configuration, and maximization of network capacity are on top of the list. Fulfilling these is a challenging task for system designers aiming at producing flexible next generation wireless systems transparently interconnecting a multitude of heterogeneous networks and systems and allowing for a more integrated strategy.

Traditionally, the two encoding operations of compression and error correction are separated from each other, following Shannon's well-known separation theorem [1], which states that, asymptotically with the length of the source data, source coding and channel coding can be designed separately without any loss of performance for the overall system. However, beside further proofs that separation neither necessarily leads to the least complex solution, nor to an applicable one [2], apparently most modern popular applications such as audio/video streaming do not meet the ideal hypotheses of the separation theorem, by often imposing real-time constraints on data transmission, operating with sources of significantly varying encoded data error sensitivity, and requiring to be as simple and low-power consuming as possible. Optimal allocation of user and system resources could be better achieved by the co-operative optimisation of communication system components in different layers.

This paper presents the approach followed by the PHOENIX project, relying on joint source channel coding and decoding (JSCC/D), and aiming at developing strategies, where the source coding, channel coding, modulation, ciphering, and also network parameters are jointly determined to yield the best end-to-end system performance. The system we developed for end-to-end optimisation of multimedia transmission over an IP wired/wireless channel was made flexible enough to be utilized with different video coding standards (e.g. MPEG-4, H.264/AVC [3]), for different net-

work architectures (in particular different transport protocols, e.g. UDP, UDP-Lite, DCCP), and over various different 3G and 4G radio access technologies and modulations (e.g. orthogonal frequency division multiplexing (OFDM), wide-band code division multiple access (WCDMA) with multiple-input/multiple-output (MIMO) and adaptive antennas).

The paper is organised as follows: the overall system model proposed is presented in Section 2, with description of both the model, the underlying network architecture and the extra signalling information introduced to let it work. The translation of this model into a practical complete and integrated simulation chain is proposed in Section 3, together with simulations results showing the interest of the overall adaptation. Finally, some conclusions are drawn.

2. SYSTEM ARCHITECTURE FOR JSCC/D

2.1 Overall System Architecture

Fig. 1 presents the overall system architecture, representing a system for video transmission over an interconnecting network that consists of routers and nodes using wired (e.g. Internet) and wireless (e.g. 4G system) links. It includes signals for transmitting JSCC/D control information used by specific functions such as unequal error protection (UEP) and JSCC adapted channel coding. Those controls are defined both at application level and physical level, and are used to provide the system components with information about user requirements and network and channel conditions, thus supervising their adaptation to changes. In particular, compression rate for video encoder and protection level for the channel coding modules are set by these controllers. Signalling mechanisms for both the transmitter and the receiver side will be detailed in Section 2.2.

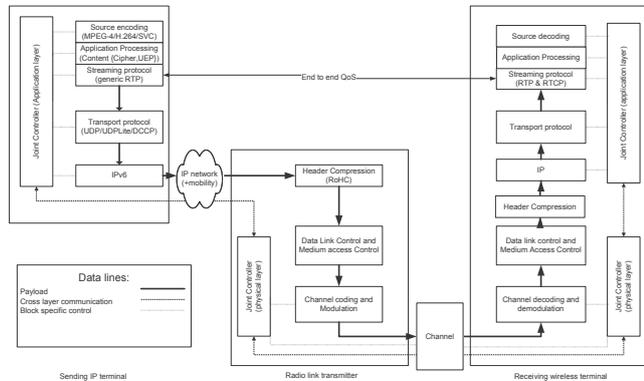


Figure 1: Overall system model and practical set-up.

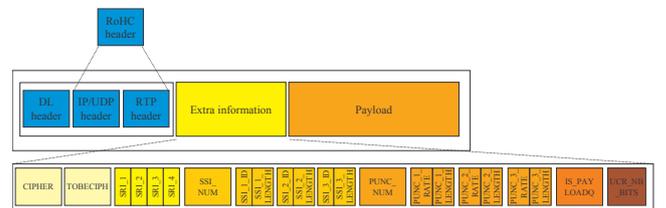
2.2 JSCC/D specific Control Information

Unfortunately, thus far the impact of network and protocol layers has quite often been neglected when considering JSCC systems and only minimal effort has been made to find solutions for efficient inter-layer and network signalling mechanisms. In fact, some work has been carried out in order to provide cross-layer protection strategies for video streaming over wireless network, proposing to combine the adaptive

selection of application forward error correction (FEC) and medium access control (MAC) layer automatic repeat request (ARQ) as presented in [4]. But the cross-layer design of systems is usually still limited to layers 1 and 2, while JSCC/D requires full cross-layering through the whole protocol stack.

As already described in [5], different types of cross-layer control information have been considered, allowing for all possibly relevant information exchange in a JSCC/D system. First is source coding related information, such as source significant information (SSI) and source a-priori information (SRI), generated by the source encoder for either better application protection or better source decoding at the receiver side. Next there is soft information such as Decision reliability information (DRI) and Source a-posteriori information (SAI), which are in practice limited to the last wireless hop in order to avoid network complexity, reduce the signalling load, and enable better tuning of the source decoding process as in [6]. Finally, we also have feedback information that will inform the controllers about the channel state (CSI), network state (NSI) or possibly, for testing purpose, about the final video quality.

Considering that these signals greatly differ in terms of refreshment rate, generation mode, allocation, and bit-rate, different solutions must be considered for their transmission. Take SSI and NSI as an example. The first one carries information to be used by UEP modules, regarding the sensitivity of source bits to channel errors. The second one carries information on availability of network resources across the data path (information such as delay, jitter, packet-loss, ...) that will be exploited by the application controller to better tune the amount of generated rate and coding parameters. SSI needs to be synchronized with payload, while NSI goes through the feedback channel and is not synchronized with the media stream. Solutions considered for transmissions synchronised with payload can either use a specific extra real-time transport protocol (RTP) [7] header, or a newly-introduced extension header for Internet Protocol version 6 (IPv6) [8]. For non-synchronised feedback information (CSI, NSI, quality) we propose to use Internet Control Message Protocol version 6 (ICMPv6) [9] or dedicated sockets. We foresee that soft-input information, due to its large bit-rate, can be transmitted by use of extra IPv6 packets as discussed in [10]. Fig. 2 illustrates the insertion of ciphering, SRI, SSI, puncturing content UEP (PUNC) and soft-input source decoding information into network packets as an extra RTP header.



2.3 Network System Model

With JSCC, the underlying network architecture that is normally invisible to the separate source and channel worlds must in fact be adapted to transparently allow for cross-layer information exchange. This section presents the most important system features and proposes some basic solutions based on IPv6 networks.

Let's consider the layers along the media data route. First, we have the transport layer, which provides streaming video service for the application layer. In our system proposal, the transport protocols use partial checksum to maximize the number of delivered packets since most recent efficient audio/video decoders can use damaged packets to improve error concealment and decoding efficiency. Therefore only the important parts of the packets such as the headers are protected by the partial checksum mechanism, while general audio/video data in the payload remains unprotected. It must be noted that, contrary to IPv4, IPv6 implies a mandatory checksum mechanism at the transport layer, which can however be partial, as proposed by UDP-Lite [11] and DCCP [12]. For media streaming, RTP/RTCP [7] protocol can be used together with UDP-Lite. Furthermore, the RTCP control/signalling information has been identified as a potential NSI information source that, while meeting the overall JSCC/D approach, is compatible with existing networks. With DCCP the use of RTP/RTCP is not necessary, because DCCP itself supports all necessary features provided by RTP/RTCP.

Next follows the network layer, with IPv6 and its mobile IPv6 extension [13] taking care of mobility issues and allowing mobile terminals to possibly change their access points while keeping a continuous connection. The solution implemented in the transmission chain design relies on anycast addresses to identify mobile nodes as they move around performing in-domain handovers [14]. It should also be noted that the network transparent cross-layer solutions proposed in Section 2.2 may lead to modifying the IPv6 header, for instance by inserting a new extension header. Furthermore, the use of a hop-by-hop option to carry SSI information could allow intelligent routers to implement DiffServ services based on the importance of the carried payload data.

Below the IP layer, we introduced a header compression mechanism called Robust Header Compression (ROHC) [15], relying on fixed and known syntax of the various RTP/UDP-(Lite)/IP headers for their reconstruction from partial information. ROHC's principle is to compress the transport and network headers by transmitting only non-redundant information. This allows reducing the bit-rate consumption over wireless links and consequently improves robustness to errors. Note that it can be beneficial to use SSI to identify the most important packets, whose headers should be better protected by inserting unequal error protection at the radio access point [16].

Compressed packets are then handed to the data-link layer, where several critical functions of the whole JSCC/D system are implemented. Mostly related to medium access control, they consist of data-plane services, such as error detection/correction and data delivery mechanisms, and control-plane services, such as link and resource control. In order to

make the system robust, especially protocol headers should always be effectively protected, whereas we intend to let erroneous unprotected audio or video payload go on undiscarded. Consequently, the link layer provides also efficient unequal error detection: a partial CRC is applied on headers only, the payload is left unprotected and no fragmentation is applied. To meet the hypotheses of Beyond 3G to 4G transmission assumed in the PHOENIX project scope, the data link layer is based on the IEEE 802.11 [17] carrier sense multiple access (CSMA) MAC mechanism, with a new data frame format providing identification for the required JSCC/D signals and the new features of partial checksum for multimedia data (including MAC header and (possibly compressed) RTP/UDP(-Lite)/IP headers). This proposal is based on CSMA, but similar mechanisms can be provided for any 4G access technique.

3. SIMULATION RESULTS

3.1 Practical settings and mode of operation

Following the first pieces of work presented in [5], an integrated software chain simulating the different parts of the system model presented in Fig. 1 has been developed. This simulator includes both transmitter and receiver side of the system model and performs real binary packet exchange at the network layer.

The application layer includes the H.264/AVC video source encoder, the content cipher (partial very early assignment (VEA) ciphering and scrambling applied to Intra frames) and the content UEP encoder (not used in the following tests). The resulting binary stream is then fed to the network layer, which performs RTP/UDP(-Lite)/IPv6 packetisation, with insertion of the extra signalling information as detailed in Section 2.2. This follows an IPv6 network emulator (whose parameters have been determined by OPNET simulations), which takes into account the possible packet losses and delay due to possible congestions in a wired IP network, and an IP mobility emulator introducing delay and losses due to the IP wireless mobility. IP packets are then compressed at the base station by the ROHC mechanism and delivered to the data-link layer that uses the enhanced 802.11 MAC header. The radio link follows, with channel encoder, interleaver, modulator, radio channel, demodulator, de-interleaver and channel decoder combined into a single module simulating the physical layer (PHY) driven by the PHY controller, reacting to rapid changes in the channel.

If adaptation is "on", the application layer controller will decide (based on SSI information) on both source coding compression level and radio link protection. To drive the physical layer (i.e. the channel (de)coder, (de)interleaver and the (de)modulator), the application layer (APP) controller also takes into account side information signals (e.g. CSI and NSI continuously fed back to the transmitter side controller), optimizing the average repartition of bandwidth between compression and protection by use of PSNR models for respective channels, as in [16, 19]. In practice, some of the control parameters are negotiated only once during the handshaking phase, while the others are re-negotiated during the optimisation process.

In our simulator, optimisation (i.e. the APP controller update rhythm) and control signalling update is performed on

a fixed time-step of 1 s. Consistently, H.264/AVC picture encoding is done in groups of pictures whose duration is equal to 1 s. The 1 s time-step was chosen as a compromise between the need of regular updates of channel state information at transmitting side, and the compression efficiency, which decreases, when source coding parameters change too often. The network-oriented definition of the H.264/AVC standard is exploited in our scheme by encapsulating different network abstraction layers (NAL) in separate packets, with maximum packet size determined during the handshaking phase. The NAL unit type can then also be used for applying unequal error protection, typically by requesting more protection for Intra frames (e.g. NAL unit type 5). At the radio link level, FEC by means of rate-compatible punctured convolutional coding [18] (to allow different protection rates as decided by the controller at each step, and not for intra-packet UEP protection in the used settings) and single carrier BPSK modulation are considered. It must be noted that the system also works with more complex modulations (e.g. trellis coded modulations, space-time modulations) and other channel coding solutions (e.g. low-density parity-check (LDPC) codes). The PHY controller adapts the protection level to rapid non-selective channel variations, modelled as time-correlated Rayleigh block fading and Log-normal shadowing.

The joint controllers consequently have a key role in the simulation model, because they collect and manage all of the information received from various blocks and send the optimized control information back to the blocks. The application layer controller tries to improve the long-term average received video quality, whereas the PHY controller tries to cope with fast fading effects, accepting the average code-rate decided by the APP controller. Slow fading is assumed constant until the next APP controller update and changes only at the end of each shadowing block. Video quality is measured in terms of peak signal-to-noise ratio (PSNR) to estimate system performance in terms of adaptation gain. This of course cannot be used as a controlling signal fed back to the APP controller: being a full reference method, PSNR assumes knowledge of the original content, not available at the receiver.

Launched with default parameters, the system adapts coding and protection to network and channel conditions, based on updates of NSI and CSI. The latter for now are considered error-free, but future work will include their actual transmission via the system design proposed in Section 2.2.

3.2 Numerical example results

To better illustrate the interest of the optimisation approach, let us consider an example of the gain achievable with the system described. The settings used in the integrated software PHOENIX simulation chain are summarized in Table 1, and the corresponding behaviour of PSNR versus time is shown in Fig. 3. Two sets of curves are shown. The first set corresponds to a non-adaptive (or “fixed”) case, for which a fixed protection rate of 50% of the bandwidth is allocated to protection; the second to an adapted (or “allocated”) case, where the APP controller drives the whole system, with the additional constraint that a minimum of 10% of the bandwidth will be allocated to PHY protection, to ensure minimum protection for network headers (includ-

ing MAC headers). In Fig. 3, the two dotted lines curves show PSNR vs. time in the absence of channel effect, corresponding to coding of the video source according to the APP controller decisions at successive time-steps, and representing the maximum PSNR achievable when channel and adaptation are introduced. The solid line curves show corresponding results obtained for fixed transmission (in triangles) and optimised allocated transmission (in diamonds).

In good channel conditions the two solutions are very close, but, when fading occurs, adaptation provides an improvement with respect to the fixed case. Results obtained for frames 15 to 45 in Fig. 3 provide a very illustrative example of adaptation effect, that shows that when encountering a difficult channel, adapted video is more compressed and consequently received with no errors thanks to its higher protection, whilst video with fixed protection has most of the packets discarded or too corrupted for use. Sample visual results, in accordance with average visual impact, are also proposed in Fig. 4. Here again, the improvement with the adapted schemes (on the left) when compared to fixed classical approach (on the right) is evident. On average over different simulation runs, gains of 6 to 7 dB can be observed in this configuration.

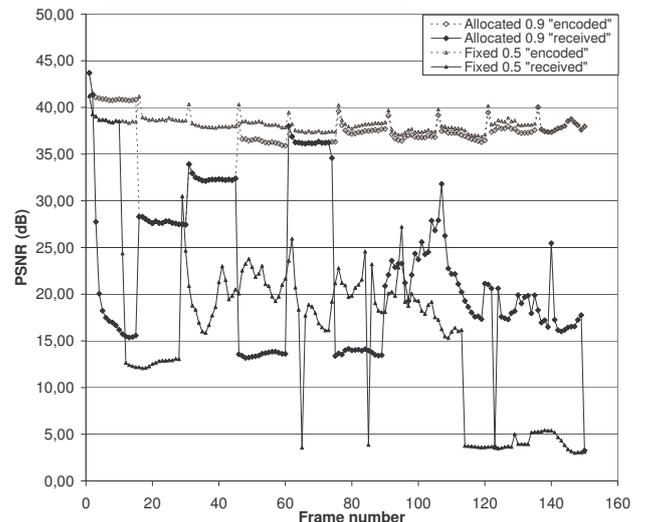


Figure 3: PSNR versus time for simulation with and without adaptation by controller.

4. CONCLUSION

In this paper we propose a novel system architecture for multimedia transmission over an IP-based wireless networks, in which the application world (source coding) and transmission world (channel coding and modulation) interconnect efficiently with the network world (channel access, IP networking and transport services), thanks to JSCC/D. For future systems, it is necessary to find not only solutions finely tuned to given transmission conditions, but also a global solution for efficient multimedia transmission. To this end, we pointed out the several promising mechanisms for the delivery of JSCC/D controls, each best suited for easy adoption into different kinds of IP networks, and not requiring that system administrators and service providers change their whole infrastructure. The proposed JSCC/D signalling



Figure 4: Example of visual results obtained in simulation: adapted (left) vs. non-adapted (right).

mechanisms can thus be adopted in any 4G scenario supporting IPv6 networking. The partial checksum mechanisms can be adopted as well with only minor modifications to link layer entities. The proposed JSCC/D architecture considers the overall transmission chain from application level source coding to wireless and wired channels, the required JSCC/D signalling mechanisms and real network functionality. In particular, we have defined new solutions that effectively deliver JSCC/D control information through the network and through the protocol stack, allowing for the cross-layer design approach we are aiming for.

Finally, this paper presents a software prototype (integrated simulation chain) based on the proposed architecture, which includes not only the source coding and channel coding block, but also the whole protocol stack including link, network, and transport protocols. Sample simulation results obtained using the integrated simulation chain have been presented, showing the improvements of video quality both visually and in terms of PSNR.

5. ACKNOWLEDGMENTS

The authors would like to thank their colleagues, who have participated in IST PHOENIX project and have given valuable work contribution for the development of PHOENIX system and simulation chain integration process.

6. REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July–Oct. 1948.
- [2] S. B. Zahir Azami, P. Duhamel and O. Rioul, "Joint source-channel coding: panorama of methods," in *Proc. of CNES workshop on data compression*, Toulouse, France, Nov. 1996.
- [3] *Final Draft Int. Standard of Joint Video Specification (ITU-T Rec. H.264, ISO/IEC 14496-10 AVC)*, Doc JVT-G050r1, Geneva, Switzerland, May 2003.
- [4] M. van der Schaar, S. Krishnamachari, S. Choi and X. Xu, "Adaptive Cross-Layer Protection Strategies for Robust Scalable Video Transmission Over 802.11 WLANs," *IEEE Journal on Select. Areas Commun.*, vol. 21, pp.1752–1763, Dec. 2003.
- [5] M. Martini et al., "A demonstration platform for network aware joint optimisation of wireless video transmission," *IST Mobile Summit 2005*, Dresden, Germany, June 2005.
- [6] C. Bergeron and C. Lamy-Bergot, "Soft-input decoding of variable-length codes applied to the H.264 standard", *Proc. of the IEEE MMSP'04 conference*, pp.87–90, Siena, Italy, Sept. 2004.
- [7] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," IETF RFC 1889, Jan. 1996.
- [8] S. Deering and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification," IETF RFC 2460, Dec. 1998.
- [9] A. Conta and S. Deering, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification," IETF RFC 2463, Dec. 1998.
- [10] C. Lamy-Bergot and P. Vila, "Multiplex header compression for transparent cross-layer design", *Proc. of the IEEE ICN'04 conference*, pp. 1084-1089, Gosier, French Polynesia, March 2004.
- [11] L-A. Larzon, M. Degermark, S. Pink, L-E. Jonsson and G. Fairhurst, "The Lightweight User Datagram Protocol (UDP-Lite)," IETF RFC 3828, July 2004.
- [12] E. Kohler, M. Handley and S. Floyd, "Datagram Congestion Control Protocol," IETF RFC 4340, March 2006.
- [13] D. Johnson, C. Perkins and J. Arkko, "Mobility Support in IPv6," IETF RFC 3775, June 2004.
- [14] I. Dudas, L. Bokor, G. Zs. Bilek, "Examining Anycast Address Supported Mobility Management Using Mobile IPv6 Testbed," *Proc. MELECON 2004*, 12-15 May 2004, Dubrovnik.
- [15] C. Bormann (editor), "RObust Header Compression (ROHC): framework and four profiles: RTP, UDP, EPS, and uncompressed," IETF RFC 3095, July 2001.
- [16] C. Lamy-Bergot, N. Chautru and C. Bergeron, "Unequal Error Protection for H.263+ bitstreams over a wireless IP network", *Proc. of IEEE ICASSP'06 conf.*, pp.V-377/V-380, Toulouse, France, May 2006.
- [17] *IEEE 802.11 – 1999 edition, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11, 1999.
- [18] J. Hagenauer and T. Stockhammer, "Channel coding and transmission aspects for wireless multimedia," *Proc. of the IEEE*, vol. 87, no. 10, Oct. 1999.
- [19] M. G. Martini and M. Chiani, "Rate-Distortion models for Unequal Error Protection for Wireless Video Transmission," *Proc. IEEE VTC 2004 conference*, pp. 1049-1053, Milan, Italy, May 2004.

Table 1: Recapitulation of the considered simulation parameters.

Test video sequence	
Video sequence:	Foreman
Video format:	QCIF (176x144)
Frame rate:	15 fps
Duration:	10 seconds
Source coding	
Encoded sequence frame rate:	15 fps
Intra frame refreshment period:	15 frames
Initial QP values (I,P):	32, 35
H.264 packet maximum size:	180 bytes
Encoding mode:	standard non DP mode
IPv6 wired network	
IPv6 network nb of nodes:	10
Mean node delay :	3 ms
Mean node packet loss:	100 ppm
Bottleneck rate:	10000 kbps
Buffer size at bottleneck :	100000 bytes
IPv6 mobility	
Delay mean:	10
Delay Sqr StdDev:	4
Ho latency mean:	520
Ho latency SqrStdDev;	100
Ho mean:	820
Ho Sqr StdDev:	34.5
ROHC parameters	
Network headers considered:	RTP/UDP-Lite/IPv6
Compression mode :	unidirectionnal (U)
Compression rate:	average (\simeq 8 bytes, FO_timeout=11, IR_timeout=1000)
MAC layer parameters	
Dynamic partial checksum	MAC+ {ROHC or RTP/UDP-Lite/IP} 36 to 96 bytes coverage, 4 bytes length
Radio Link	
Channel encoder:	RCPC codes
Mother code rate:	1/3
Constraint length:	5
Code generators (in octal):	23; 35; 27
Puncturing period:	8
Code rates considered:	8/9, 4/5, 2/3, 4/7, 1/2, 4/9, 2/5, 4/11, 1/3
Interleaving:	random (done packet by packet)
Modulation:	BPSK
Number of RX/TX antennas:	1/1
Maximum coded bit-rate:	500 kpbs
Radio channel:	non-selective block fading
Median Es/No:	1 dB
Slow fading:	uncorrelated Log-Normal distr., $\sigma = 4$ dB coherence time: 8 s (=slow-fading block duration)
Fast fading:	time-correlated Rayleigh distr. , doppler frequency: 2 Hz ch. gain sample time : 1 ms (=fast-fading block duration)