End-to-end application control of video streaming: implementation and performance evaluation

R. Fracchia, B. Gadat and C. Lamy-Bergot
Thales Communications, 160 boulevard de Valmy, 92740 Colombes, France

Abstract— **This paper presents a solution for multimedia streaming adaptation to the conditions of the network. This scheme takes advantage of the possibility to apply different degrees of compression to video contents, and considers the need of protecting such content from errors and losses by forward error correcting codes (FEC). The FP7 ICT OPTIMIX project aims at optimising the transmission of a video stream to a single user or to a group of clients, connected to the Internet via a portable wireless device. In this context, an application level controlling process has been proposed in [1]. This paper presents the system architecture proposed for the use of this controller and evaluates the performance of the proposed scheme for both point to point and point to multi-point transmissions.**

I. Introduction

In multimedia streaming over IP wireless links, Quality of Service (QoS) requirements raise huge challenges not only concerning the physical bandwidth, but also the network design and services. This has to be addressed by modern communication systems where all users want to be connected dependably and efficiently.

Innovation in the area of sophisticated multimedia source coding schemes aiming at satisfying design criteria and trade-offs in terms of source representation quality, bitrate, delay, encoding/decoding complexity, etc. is a key issue in the modern world where users demand for content anywhere and anytime. Today's approach, relying on traditional layers separation and focusing on services delivered over homogeneous networks, does not allow meeting the on-going demands to maintain the required Quality of Service for the different users, who have different needs and requirements. To this end, joint source and channel coding (JSCC) solutions have been proposed in the scientific literature for several decades and have shown promising results [2]. Until recently, however, they were often considered as purely academic work, not necessarily taking into account practical considerations unavoidable in real deployment. As an example, the lacking for an IP network support in most initial studies resulted in JSCC optimisation strategies with a unique module for source and channel coding, which is hardly compatible with a layered approach.

In this context, the ICT OPTIMIX project develops a scheme including all the elements of major importance in a point to multi-point video streaming chain, like video coding, networking modules, MAC layer and physical layer, efficiently communicating together. In particular, OPTIMIX considers innovative techniques to improve the efficiency of video codecs when used in a wireless multi user environment with respect to robustness, efficient compression and intelligent use of scalability schemes. Furthermore, it develops cross-layer mechanisms to enable the communication between application and transmission worlds through the use of enhanced transport and network protocols.

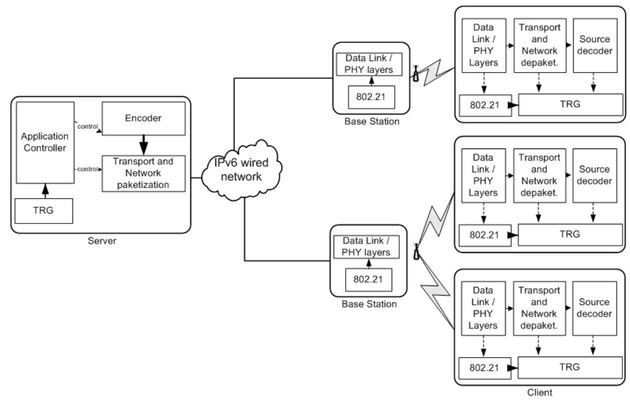

Fig. 1. Application Controller inputs and outputs

In this paper we present and evaluate the adaptation of the multimedia streaming to the transmission conditions for both point to point and point to multipoint video streaming using a complete system simulator.

The rest of the paper is organized as follows. The considered system architecture is presented in Section II of this document. Section III presents the application controller and its behavior, Section IV describes how the input values of the control algorithm are derived while Section V presents the control for a point to multipoint communication. Section VI reports the performance evaluation of the proposed scheme for both unicast and multicast transmissions and Section VII concludes this work.

II. System Architecture

We consider in this work the architecture presented in Figure 1, composed by a server, a wired IP network and one or more wireless networks, with mobile clients connected to base stations or access points. The server is equipped by an application controller driving the video encoder and the transport layer operations for adaptation to the transmission conditions.

Among the several innovations introduced by the project, two works are particularly relevant to this study: i) the RTP Forward Error Correction (FEC) solution and ii) the end-to-end signaling framework.

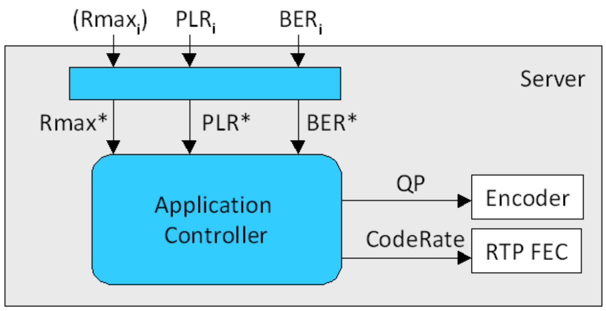

Fig. 2. Application Controller inputs and outputs

RTP FEC provides error correction against both packet losses and bit errors at the transport layer: its implementation is based on the introduction of Reed-Solomon codes in RTP packets and has been presented in [3].

The end-to-end OPTIMIX signaling system is based on the joint use of the Triggering Framework [4] and IEEE 802.21 [5]. The main entity of the Triggering Framework, the Triggering Engine (TRG), is present in server and clients and allows exchanging information, called triggers, between trigger sources and consumers at any layer of the network protocol stack. IEEE 802.21 is used in OPTIMIX clients and BSs to provide timely values of physical and data link layer parameters to the other entities in the network: IEEE 802.21 information is converted into triggers, which are handled by TRG as all other triggers. Although the Triggering Framework allows using filters in the subscriptions, the feedback exchange would cause overhead for the total network traffic. In order to mitigate the feedback overhead, a client-side aggregation mechanism for the Triggering Framework has been developed. Trigger aggregation bundles multiple triggers into one trigger, which is periodically sent to the consumers subscribed to it.

III. The Application Controller

The application controller presented in this paper adapts the source coding parameters (e.g. quantization parameters, bitrate, etc.) and the protection rates of RTP FEC on a relatively slow-rate base according to the video source characteristics and to the state of network and radio channels that are received via the Triggering Engine.

The application controller thus acts as an intelligent streaming pump implementing the controlling strategies and ensures that the compression and protection at high protocol stack level functions are decided jointly and efficiently from the end-user point of view. The controlling process at the video server results in deciding every Group Of Picture (GOP) (i.e., every second) the best compression and protection parameters given the available bitrate and transmission conditions (network and channel state information).

The application controller has not knowledge of the video sequence itself and decides using a set of abacuses providing statistical information on bitrates and video quality curves for given compression parameters and different reference sequences. The semi-analytical approach of the application controller is detailed in [1].

In particular, the application controller bases its decision algorithm on an estimation of the video quality obtained from the Packet Loss Rate (PLR), the Bit Error Rate (BER) and a given Quantization Parameter, as sketched in Figure 2. Assuming to know the PLR and the BER experienced by the clients and a maximum total bit rate R_{max} for data and protection, whose evaluation is detailed in Section IV, the application controller works as follow. Taking a 10% of margin on the maximum bit rate R_{max} to consider the overhead introduced by the headers and determining a new rate R'_{max}, the algorithm selects an abacus among the available ones and for all the possible FEC rates:

- Given the input PLR and BER, the FEC rate and the RS properties, it analytically evaluates the packet loss rate after the RS decoding.
- It determines the amount of bandwidth left after the insertion of the RS codes as: $R = R'_{max} Rate_{FEC}$
- It determines, for that abacus and that FEC rate, the QP providing the highest bit rate not exceeding the threshold R.
- It evaluates the obtained video quality on the reference video corresponding to that abacus, which depends on the output PLR and the selected QP.

The application controller then selects the FEC rate giving the highest video quality and it encodes the stream using the corresponding selected QPs. After the encoding process, it determines the difference between the expected results (PSNR, bit rates, etc.) based on the reference sequence of the considered abacus and the results of the encoding process on the video sequence to transmit. Based on this difference it selects another abacus. We can thus say that the application controller learns from the previous decision, since abacuses showing properties closer to the video sequence to transmit are progressively selected.

This controlling process is based on a trial approach, which is repeated up to three times: after the three iterations, the QPs and the FEC rates generating the encoding sequence with highest PSNR are selected.

As a last note, priorities of the single frames are reported in the DS byte of the IPv6 header and can be used in the wired as well in the wireless network for Quality of Service guarantees.

IV. Application Controller Inputs

As explained in the previous section, the application controller optimizes the encoding and the transmission by considering: i) the maximum bit rate that could be generated, ii) the packet loss rate and iii) the bit error rate.

An indication of the maximum bit rate R_{max} can be obtained at the server side from an estimation of the available bandwidth, which can be derived by using one of the several existing tools (e.g., Spruce [7], Pathload [8], pathChirp [9], Assolo [10], etc.) or from the TCP-Friendly Rate Control (TFRC) adaptation algorithm [11].

PLR and BER are instead feedbacks sent by the clients via the TRG. Packet loss rate is measured by the RTP module every 0.1 sec on the received sequence numbers: this value thus takes into account all the losses occurred along the path from the server to the client, e.g., due to problems in the wired network, to buffer overflow at the Base Station, or to wrong CRCs or checksum at the receiver. BER is instead communicated by the physical layer every twenty packets via IEEE 802.21.

The two values are then sent as an aggregated trigger to the application controller every second. The aggregated trigger contains the average PLR and BER measured on the values obtained since the previous aggregated trigger transmission. Several aggregation schemes, different from the average, have been considered but did not introduced an additional gain: indeed, the application controller can not react to fast changes of the channel conditions due to the relatively slow reactivity of the compression process. A fine optimisation could be realized at the base station, with an intelligent resource allocation and scheduling algorithm exploiting priority information provided by the application controller.

V. Application Controller decision in a point-to-multipoint transmission

The input values described in the previous section allow an accurate parameter selection when encoding a video for a single client. In case of a point-to-multipoint transmission, several triggers are received by the application controller: namely, one for each user. Moreover, each server-to-client communication, which can transit over a different path, through different wireless cells or even different wireless networks, can present very different conditions, resulting in possible very different PLR, BER and bandwidth values. In case of H.264/AVC encoding, a single stream is transmitted to all the receivers. The different input values, received by the different clients, are transformed by the application controller into the requirements $(R^*_{max}, PLR^*, BER^*)$ of a target user U^*, as depicted in Figure 2. The criterion used to determine the target requirements is a choice of the service provider and can depend on the service provider policy, on the kind of scenario and application or on the needs of the clients. We thus decided of not fixing a single aggregation policy but to provide some indications on the possible cost functions that could be used. As an example, if a minimal guarantee of service is required by the clients, e.g., as in case of video surveillance applications, the target requirements could be given by the worst feedbacks. Being R_{max_i} the bandwidth of the communication toward the user i and BER_i, PLR_i the bit error rate and the packet loss rate esperienced by the client i, we can define in this case: $R^*_{max} = min(R_{max_i}), PLR^* = max(PLR_i), BER^* = max(BER_i)$. Conversely, if the service provider aims at guaranteeing a good quality to a subset of "gold users", the target values could be obtained as the average on the feedbacks of the gold users only. Finally, if no particular policies are identified, the definition of the target user could be based on the simple average of all the received feedbacks.

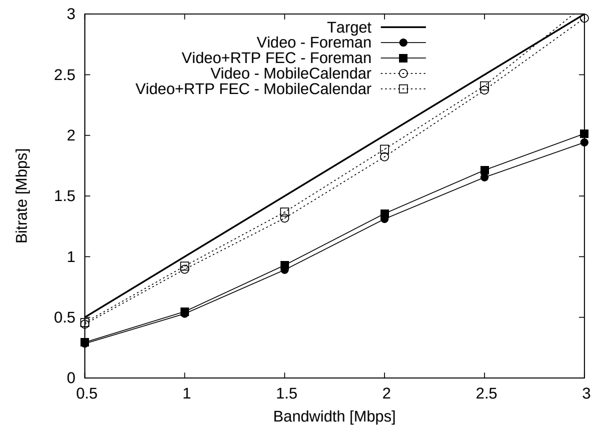

Fig. 3. Bit rate generated by the Master Application Controller as a function of the target bit rate for BER=0 and PLR=0

VI. Simulation Results

This section presents the performance of the application controller for both point-to-point and point-to-multipoint streaming of an H.264/AVC video.

Results are obtained with the simulator developed by the partners of the OPTIMIX project [12] and based on the OMNeT++ framework [13]. This tool simulates the encoding and the transmission - bit by bit - of a H.264/AVC video and accurately models all the OSI layers from the application to the physical layer. Clients integrate an enhanced robust decoder, the Reed-Solomon RTP FEC solution, IEEE 802.21 and the Triggering Engine. The server, in its turn, has been enhanced with the introduction of RTP FEC, of the TRG and of the application controller.

Video streaming is started and controlled by RTSP which allows the end-user requesting the desired content. Every GOP the application controller selects new QPs and FEC rates, used to encode and protect a raw video. At the client side, a robust decoder transforms the received H.264 frame into an uncompressed yuv frame which is then displayed. Below the application layer, RTP fragments each image into packets, introduces correcting codes and handles the image reconstruction at the receiver side. At the transport and network layers, the UDPLite transport protocol and IPv6 are used respectively. At the data link and physical layer, a TDMA transmission with a simple round-robin (RR) policy schedules the users having some data in the transmission buffer; data are transmitted with a rate of 6 Mbps. A Rayleigh block fading channel is used to represent the effects of the radio channel on the wireless transmission.

As video sequences we consider the raw Foreman and the Mobile Calendar CIF reference sequences at 30 Hz, which show different properties and thus differently highlight the behaviour of the application controller.

The first evaluation consists in varying the bandwidth of the system and determining the application controller performance in terms of generated bit rate. Figure 3 reports the bit rate of the encoded video and the total bit rate including video

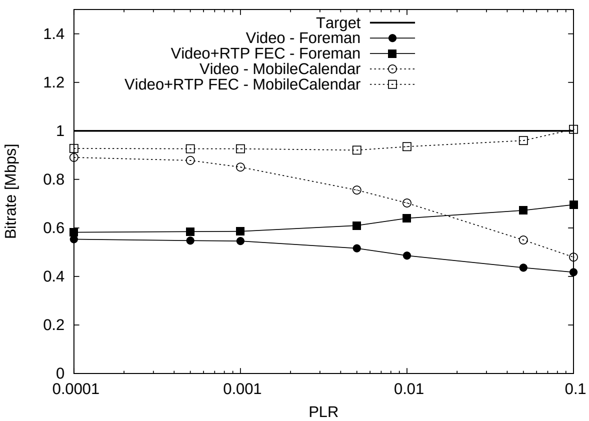

Fig. 4. Generated bit rate as a function of the packet loss rate

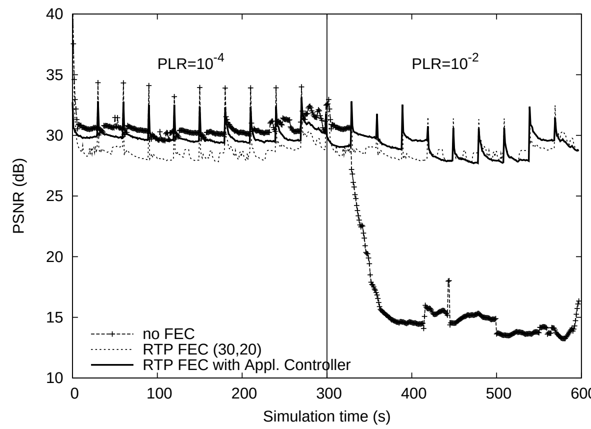

Fig. 5. PSNR versus time (expressed as video frame number) for the Mobile Calendar sequence and different loss rates

and FEC protection as a function of the system bandwidth, for an ideal transmission without packet losses nor bit errors. Since there are no losses nor errors on the transmission, no RTP FEC protection is introduced by the application controller and the small difference in the bit rate at the application and at the RTP layers is given by the fragmentation and the insertion of the RTP header. We can observe first that the generated bit rates never exceed the provided threshold. Second, we can notice that the achieved results depend on the considered video sequence: indeed, the application controller does not work on the video sequence itself, that is considered unknown, but on a set of abacuses corresponding to a set of reference sequences. For the Foreman sequence, the generated bit rate for the selected QP results lower than the bit rate estimated using the abacus. It follows that the obtained bit rate is, on average, lower than the target. For the MobileCalendar sequence, presenting a lot of movement and details, instead, the generated bit rate almost corresponds to the target one. Moreover, we can observe that the generated bit rates increases following the available bandwidth, increasing the received video quality.

We then evaluate the reaction of the application controller to losses in the wired network while assuming very good wireless channel conditions: the client experiences losses in the RTP sequence of received packets and indicates the loss rate in the aggregated feedback. We plot in Figure 4 the bit rate of the encoded video and the bit rate at the RTP layer (i.e., after RTP FEC insertion) as a function of the packet loss rate in the wired network. We consider an available bandwidth of 1Mbps for the Foreman and Mobile Calendar videos. We can notice that the difference between the results obtained on the Foreman and the Mobile Calendar sequences observed in Figure 3 is maintained also in this case. We can moreover see that as the loss probability in the wired network increases, the difference between video and RTP bit rate increases: indeed, the application controller increases the protection rate inserted at RTP. We can finally observe that in order to maintain a total bit rate (i.e., bit rate of video and redundancy) lower than the threshold, the video bit rate has to be reduced: higher QP parameters are thus selected by the application controller in this case.

The gain in terms of video quality on the Mobile Calendar sequence obtained thanks to the application controller decision is presented in Figure 5. This plot compares the PSNR as a function of the simulation time obtained in three different configurations for an available bandwidth of 1 Mbps. First we consider a transmission without FEC: in this case a precoded video of 1Mbps is transmitted to client. Then we consider a fix amount of protection with a RTP FEC solution with a Reed-Solomon code RS(30,20): since one third of the bandwidth is devoted to the RS packets, a precoded video of 666kbps is used. Finally, we consider a dynamic selection of the RTP FEC parameters done by the application controller. At the beginning of the simulation the loss probability is set to 10^{-4} and is then increased to 1% at the time t=10 sec, after the transmission of 300 frames as indicated in the figure. We can observe that, when the loss probability is very low, no gain is introduced by the FEC use and the PSNR obtained with a fixed RS rate (30,20) is the lowest one since a video with an higher compression is transmitted. The PSNR obtained with the application controller is higher, but lightly lower than the one obtained without FEC. The difference is due not to an introduction of FEC but to the 10% margin used by the application controller in its decision: indeed, not knowing the sequence to transmit, the application controller uses a target bit rate of 900kbps instead of 1Mbps, thus resulting in a slightly lower PSNR. The gain introduced by the application controller can be clearly observed when the loss probability increases: indeed, in this case, the PSNR achieved with the application controller is the highest one and still close to 30dB. When no protection is used, the PSNR rapidly decreases to very low values; conversely, results similar to the ones achieved with the application controller are obtained with a RS(30,20). We can conclude that the dynamic selection of the application controller appropriately identifies the parameters to use and guarantees good results in any configuration, independently of the loss probability experienced by the clients.

With a second set of simulations we evaluate the performance of the application controller in a point-to-multipoint streaming of an H.264/AVC video: two clients want to receive the same video and request it to the server via RTSP at two different times (i.e., 20 ms and 500 ms respectively). The two clients are connected to two different base stations and may experience different transmission conditions.

We compare the results obtained with three different definitions of the target user U^*, corresponding to different policies used by the service provider: the target user is defined respectively as the worst user (lowest bandwidth and highest loss and error probability), as the average one and finally as the best user (highest bandwidth and lowest loss and error probability).

We set the bandwidth of the communication toward the first client to 1 Mbps while we vary the bandwidth toward the second client between 1 and 3 Mbps. The radio channel has been modeled as a Rayleigh block fading channel with fast fading and the signal to noise ratio is set for the two users to 38 dB, so no error or losses affect the communication during the simulation. We report in Figure 6 the average PSNR as a function of the difference between the bandwidth experienced by the second and by the first user (U_1 and U_2 respectively, with $U_2 > U_1$), for the three considered definitions of target user. We can notice that, as expected, when the protection and compression parameters are selected considering the user with the lowest bandwidth (i.e., 1 Mbps) both users perceive the same video quality (lines overlap in the figure): indeed, enough bandwidth is available toward the two clients for the transmission of the generated stream and no losses affect the communication. By changing policy, the video quality experienced by the second user increases to the detriment of the video quality of the first user which loses some packets. The highest the difference in the bandwidth, the highest the difference in the PNSR: indeed, while the PSNR of the second user increases since a lowest QP is used, the PSNR of the first user decreases since the transmission exceeds the available bandwidth and packet losses affect the transmission. This difference is more important when the best user policy is used. We can finally point out that the quality of the video received by U_1 decreases more than the increase experienced by U_2: this is due to the PSNR definition and evaluation, which dramatically drops in case of lost packets.

VII. Conclusion

This paper presented a solution for video streaming adaptation to the network conditions that, based on feedbacks on errors and losses experienced by the clients and on an estimation of the available bandwidth, determines video compression parameters and correcting code rates. Performance results showed that the control algorithm allows to experience a good video quality in any condition. It has to be noted that this solution does not pretend to follow fast variations of the wireless channel conditions and it can be seen as a partial degree of optimisation: for a fine tuning of the transmission, an intelligent resource allocator and scheduling algorithm should be used for the wireless transmission. This two-steps optimisation could result particularly interesting for point to multipoint streaming of a non scalable video: indeed, the base station could perform a selective packet drop when the available radio resources are not enough for the transmission of the encoded stream.

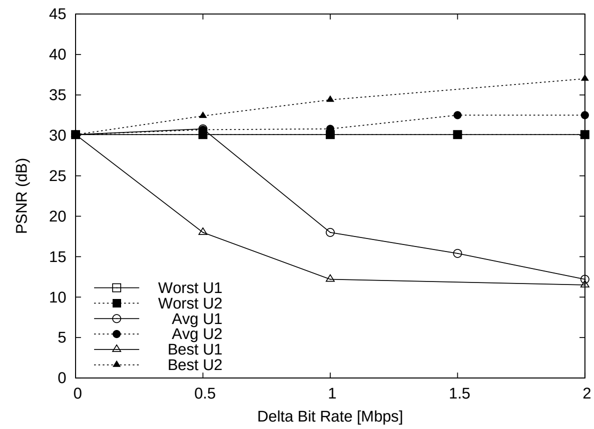

Fig. 6. PSNR as a function of the difference in the available bandwidth toward the clients and different decision policies (Mobile Calendar)

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement no. 214625.

References

[1] C. Lamy-Bergot and B. Gadat, "Application control for fast adaptive error resilient H.264/AVC streaming over IP wireless networks," in Proc. of ICASPP 2011

[2] G. Cheung and A. Zakhor. "Bit allocation for joint source/channel coding of scalable video," IEEE Trans. Image Processing, 9(3):340 356, March 2000

[3] M. Mazzotti et al., "Analysis of Packet-Level Forward Error Correction for Video Transmission," in Proc. of VTC Spring 2011

[4] E. Piri, T. Sutinen and J. Vehkapera, Cross-layer Architecture for Adaptive Realtime Multimedia in Heterogeneous Network Environment, in Proc. of European Wireless, Aalborg, Denmark, 2009.

[5] IEEE 802.21 WG, "IEEE Standard for Local and Metropolitan Area Networks. Part 21: Media Independent Handover Services", IEEE Std 802.21-2008, Jan. 2009.

[6] ICT FP7 OPTIMIX project, D2.3a, Preliminary Scheme of JSCC/D Controller for Point to Multipoint Communication and Optimization Criteria.

[7] J. Strauss, D. Katabi, and F. Kaashoek, A measurement study of available bandwidth estimation tools, in Proc. Of the 3rd ACM-SIGCOMM Conference on Internet Measurement, New York, NY, USA, 2003.

[8] M. Jain and C. Dovrolis, Pathload: A measurement tool for end-to-end available bandwidth, in Proc. of 3rd PAM Workshop, Fort Collins, CO, USA, March 2002.

[9] J V. Ribeiro, R. Riedi, R. Baraniuk, J. Navratil, and L. Cottrell, Pathchirp: Efficient available bandwidth estimation for network paths, in Proc. of 4th PAM Workshop, San Diego, CA, USA, April 2003.

[10] E. Goldoni, G. Rossi, A. Torelli, Assolo, a New Method for Available Bandwidth Estimation, in Proc. Of the Fourth International Conference on Internet Monitoring and Protection, 2009.

[11] I. Kofler, J. Seidl, C. Timmerer, H. Hellwagner, I. Djama and T. Ahmed, Using MPEG-21 for cross-layer multimedia content adaptation, Springer Journal on Signal, Image and Video Processing, vol. 2, pp.355-370. Oct. 2008.

[12] http://www.ict-optimix.eu/index.php/OPTIMIXsim

[13] http://www.omnetpp.org/