Active Concealment for Internet Speech Transmission

Long Le[1], Henning Sanneck[1], Georg Carle[1], and Tohru Hoshi[2]

[1] GMD Fokus, Kaiserin-Augusta-Allee 31,
D-10589 Berlin, Germany
{le,sanneck,carle}@fokus.gmd.de
[2] Systems Development Laboratory, Hitachi, Ltd., 292 Yoshida,
Totsuka, Yokohama, 244-0817, Japan
hoshi@sdl.hitachi.co.jp

Abstract. Recently, active networks have been highlighted as a key enabling technology to obtain immense flexibility in terms of network deployment, configurability, and packet processing. Exploiting this flexibility, we present an active network application for real-time speech transmission where plugin modules are downloaded onto certain network nodes to perform application-specific packet processing. In particular, we propose to perform loss concealment algorithms for voice data streams at active network nodes to regenerate lost packets. The regenerated speech data streams are robust enough to tolerate further packet losses along the data path so that the concealment algorithms at another downstream node or at the receiver can still take effect. We call our approach *active concealment for speech transmission* to distinguish it from concealment performed at the receiver. Our approach is bandwidth-efficient and retains the applications' end-to-end semantics.

1 Introduction

Most real-time multimedia applications are resilient and can tolerate occasional loss of packets to some extent but are sensitive to packet losses which are not in accordance to their flow structure. For example, Internet voice applications that use (sample-based) waveform-coded speech signals can exploit speech properties to conceal isolated losses very well. However, speech quality drops significantly in the occurrence of burst losses [4]. The Adaptive Packetization and Concealment (AP/C) technique successfully demonstrates that speech properties can be efficiently exploited to improve the perceived quality at the application layer [8]. However as AP/C exploits the property of speech stationarity, its applicability is typically limited to isolated, i.e. non-consecutive losses. Under circumstances where the rate of losses that occur in bursts is high, AP/C does not obtain any significant performance improvement compared to other techniques. We believe that this is the point where flexibility provided by active network nodes can be exploited to help applications at end systems to perform better.

In this work, we present an active network application where concealment algorithms are performed at network nodes to regenerate lost packets and to inject them into voice streams. The rest of this paper is structured as follows. First, we briefly review related work and the AP/C algorithm which we download to active network nodes to perform loss concealment for audio streams. We then present our approach of placing active network nodes at certain locations within the networks to leverage the efficiency of the receiver's concealment performance. We perform a simulation study to evaluate the efficiency of our approach. We finally give conclusions of our work and end the paper with an outline of our future work.

2 Related work

Recently, it has been proposed to push more intelligence into the networks to perform application-specific packet processing and actions at network nodes [9]. Significant performance improvements can be gained thanks to network nodes' application-specific packet processing which takes into account the characteristics of packet payload. This is especially true for multimedia data which has a specific flow structure. Typical examples for application-specific packet processing at network nodes are media transcoding [1], media scaling [6], packet filtering [2], or discarding [3] for video distribution on heterogeneous networks with limited bandwidth. Surprisingly, there are very few active network projects that exploit active network nodes' capability of application-specific packet processing to improve the quality of Internet voice or audio transmissions. The only work we are aware of is [2] where active network nodes add an optimal amount of redundant data on a per-link basis to protect audio streams against packet loss. Since most packet losses on the Internet are due to congestion (except for wireless networks), we argue that it can be problematic to transmit redundant data onto a link which is already congested. We propose an approach where application-specific packet processing is performed at an uncongested active network node to regenerate audio packets lost due to congestion at upstream congested nodes.

3 Adaptive Packetization and Concealment

AP/C exploits the speech properties to influence the packet size at the sender and to conceal the packet loss at the receiver. In AP/C, the packet size depends on the importance of the voice data contained in the packet with regard to the speech quality. In general, voiced signals are more important to the speech quality than unvoiced signals. Thus, if voiced signal segments are transmitted in small-size packets and unvoiced signal segments are transmitted in large-size packets and if the packet loss probability is equally distributed with regard to the packet size, more samples for voiced speech are received than for unvoiced speech. Considering the higher perceptual importance of voiced signal segments, this results in a potentially better speech quality when using loss concealment at the receiver. The novelty of AP/C is that it takes the phase of speech signals into account when the data is packetized. AP/C assumes that most packet losses are isolated and that the packets prior and next to a lost packet are correctly received at the receiver. In AP/C, the receiver conceals the loss of a packet by filling the gap of the lost packet with data samples from its adjacent packets. Regeneration of lost packets with sender-supported pre-processing works reasonably well for voiced sounds thanks to their quasi-stationary property.

Regeneration of lost packets works less well for unvoiced sounds due to their random nature. However, this is not necessarily critical because unvoiced sounds are less important to the perceptual quality than voiced signals. Since the phase of the speech signal is taken into account when audio data is packetized, less discontinuities than for conventional concealment algorithms are present in the reconstructed signal.

3.1 Sender Algorithm

In AP/C, an audio "chunk" is defined as a segment of audio data that has the length of the estimated pitch period. In order to alleviate the overhead for protocol header, two audio chunks are copied into an audio packet and transmitted onto the network. When a packet loss is detected at the receiver, adjacent chunks of the previous and the current packet[1] are used to reconstruct the lost chunks. Information on the length of chunks belonging to those packets is transmitted as additional information in the current packet using the RTP header extension to help the receiver with the concealment process ("intra-packet boundary").

In order to estimate the pitch period, the auto-correlation of the audio input segment is calculated. Then the maximum value second to the maximum value at zero[2] of the auto-correlation is searched for. This maximum value, its position, and the auto-correlation itself help to make the decision whether the input segment is voiced or unvoiced. If the input segment is classified as voiced, the position of this maximum is said to be the estimated value of the pitch period because the input segment shifted by that length is most similar to itself. If the input segment is classified as unvoiced, the sender takes an audio chunk that has the length of T_{max} (in AP/C, T_{max} is the correlation window size and is chosen to be 160 samples, corresponding to 20 ms of speech). The found audio chunk is copied from the audio input buffer into an audio packet and the start position of the input segment is moved forward by the length of the audio chunk.

3.2 Receiver Algorithm

The receiver uses RTP message sequence numbers to detect packet loss and applies the AP/C concealment algorithm when an isolated loss is found[3]. RTP timestamp and information on the intra-packet boundary are used to determine the lost chunks' lengths. If silence suppression is enabled and there is a silent period between the lost packet and its adjacent packets, the lost chunks' lengths are not determined correctly. This is because RTP sequence number increments by one for each transmitted packet and RTP timestamp increments by one for each sampling period regardless of whether data is sent or dropped as silent. Thus, only the length of one lost chunk can be determined. Because the chunks' length is smaller than T_{max} and a silent period is

---

[1] The packet carrying the sequence number that allowed the detection of a previous packet loss.

[2] Of course the absolute maximum value of the auto-correlation is found at 0 because a signal without any shift is most similar to itself.

[3] In [7], we presented a scheme that combines AP/C with interleaving to cope with small packet burst loss. However, this scheme suffers from the additional buffer delay which is necessary at the sender.

usually longer than 20 ms (corresponding to 160 μ-law audio data samples), this problem can be easily detected when the length of a lost chunk is larger than T_{max}.

Due to the pre-processing at the sender, the receiver can assume that the chunks of a lost packet are similar to the adjacent chunks. The adjacent chunks (c_{12} and c_{31} in Fig. 1 are resampled in the time domain to match the size of the lost chunks and then used to fill the gap of the lost packet. A linear interpolator as in [10] is used to perform resampling. The replacement signals produced by the linear interpolator have a correct phase, thus avoiding discontinuities in the concealed signal that would lead to speech distortions while still maintaining the pitch frequency at the edges. Due to the pre-processing at the sender, the lost and the adjacent chunks have a high probability to be similar. Thus, the concealment operation introduces no specific distortion in the concealed speech segments. Fig.1 illustrates the concealment operation in the time domain.

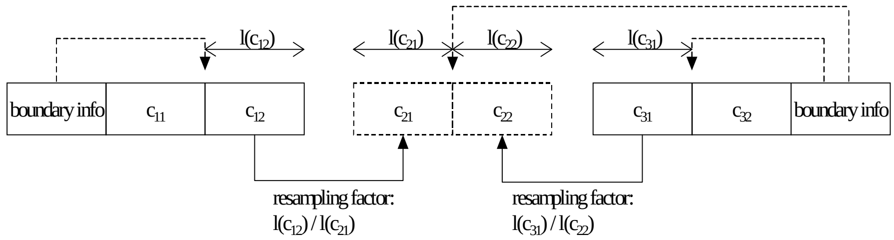

Fig. 1. Concealment operation in the time domain

4 Active Concealment

Since AP/C assumes that most packet losses are isolated, it does not obtain any significant performance improvement compared to other techniques when the rate of burst losses is high. We believe that this is the point where active network nodes' capability of application-specific packet processing can be exploited to help applications at end systems perform better. Since the burst loss rate of a data flow at a network node is lower than at the receiver, the AP/C concealment algorithm works more efficiently and more lost packets can be reconstructed when concealment is performed within the network rather than just at end systems. We thus propose to download and perform the AP/C concealment algorithm at certain active network nodes where the number of burst losses of a voice data stream is sufficiently low to regenerate the lost packets. The regenerated audio stream is robust enough to tolerate further packet losses so that the AP/C concealment algorithm can still take effect at another downstream active network node or at the receiver.

The idea of our approach is demonstrated in Fig. 2. The AP/C sender algorithm is performed to packetize audio data taking the phase of speech signals into account. Along the data path, packet 2 and 4 are lost. Exploiting the sender's pre-processing, the AP/C concealment algorithm is applied at an active network node within the network to reconstruct these lost packets. Downstream of the active network node, another packet is lost (packet 3) which is easily reconstructed at the receiver. In this scenario, active concealment reconstructs six lost chunks (c_{21}, c_{22}, c_{31}, c_{32}, c_{41}, and c_{42}) and clearly outperforms the receiver-only concealment [8] which can only reconstruct at most two chunks (c_{21} and c_{42}) due to the burst loss accumulated along the end-to-end data path.

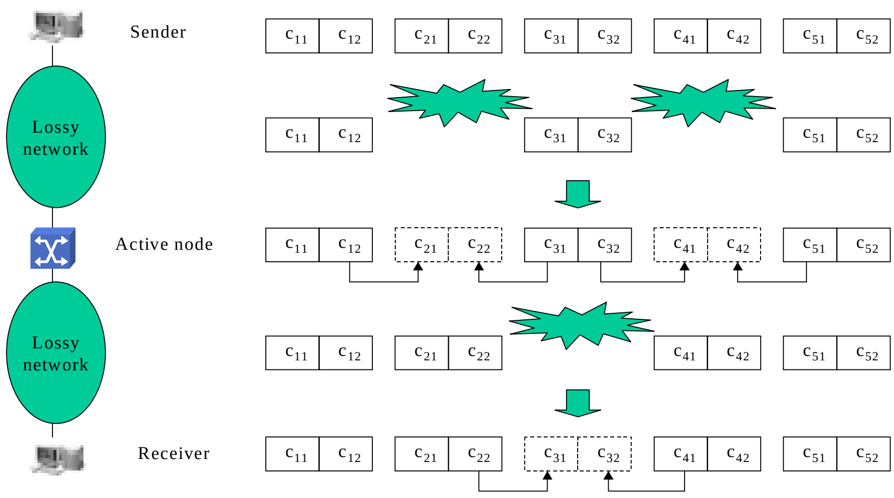

Fig. 2. Active concealment

Our approach is similar to Robust Multicast Audio (RMA) proposed by Banchs et. al. in [2] but it acts in a reactive way upon detection of packet loss in audio data streams. On the contrary to RMA which transmits redundant data on a per-link basis to protect audio streams against packet loss in a proactive way, our approach simply regenerates and injects the lost packets into audio streams and thus is more bandwidth-efficient. Another advantage of our approach is that it does not break the applications' end-to-end semantics and does not have any further demand on the number and location of active network nodes performing the concealment algorithm[4]. RMA, however, requires active network nodes be located at both ends of a link or a network to perform FEC encode and decode operation.

5 Simulations

We perform simulations to study the performance of active concealment at network nodes. As first step towards the transition from traditional to active networks, we assume that there is only one active network node in the path from the sender to the receiver where intra-network regeneration of lost packets can be performed. The logical network topology for our simulation is shown in Fig. 3 where a lossy network can consist of multiple physical networks comprising several network hops. We use the Bernoulli model to simulate the individual loss characteristics of the networks.

[4] Clearly, the number and location of active network nodes influence the performance improvement. However, the applications' functionality is not affected under any circumstances.

The efficiency of the schemes presented in this section is evaluated by using objective quality measurements such as in [5] and [11] to determine the speech quality. Objective quality metrics employ mathematical models of the human auditory system to estimate the perceptual distance between an original and a distorted signal[5]. Objective quality measurements should thus yield result values which correlate well and have a linear relationship with the results of subjective tests. We apply the Enhanced Modified Bark Spectral Distortion (EMBSD) method [11] to estimate the perceptual distortion between the original and the reconstructed signal. The higher the perceptual distortion is, the worse the obtained speech signal at the receiver is. The MNB scheme [5], though showing high correlation with subjective testing, is not used because this quality measurement does not take into account speech segments with energy lower than certain thresholds when speech distortion is estimated.

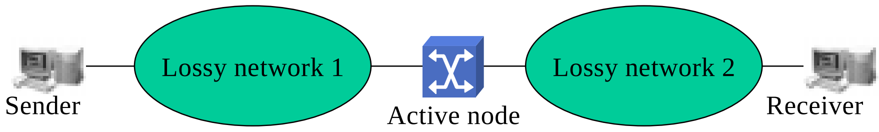

Fig. 3. Simulation topology

The structure of this section is organized as follows. In the first simulation step, we use the same parameter sets for the lossy networks. We then compare the speech quality obtained by the active loss concealment with two reference schemes. In the second simulation step, we vary the parameter sets of the lossy networks and measure the efficiency of the active loss concealment. The parameter sets are chosen in such a way that the packet loss rate observed at the receiver is constant. This simulation step is performed to determine to optimal location of the active network node where the active concealment algorithm can be downloaded and performed.

5.1 Performance comparison to reference schemes

In this simulation step, we compare the speech quality obtained by active loss concealment with two reference schemes: In the first reference scheme, the sender transmits voice data in packets with constant size and the receiver simply replaces data of a lost packet by a silent segment with the same length. Each packet in this scheme contains 125 speech samples, resulting in the same total number of packets as the second reference scheme and the active loss concealment scheme. The second reference scheme is the AP/C scheme applied only at end systems. Packets are sent through two lossy network clouds and are dropped with the same packet drop probability.

The parameters used in this simulation step and the resulting packet loss rate are shown in Table 1.

[5] We use a speech sample that consists of different male and female voices and has a length of 25 s.

Table 1. Parameters and packet loss rate used in simulation for performance comparison

Packet drop probability	0.03	0.06	0.09	0.12
Packet loss rate	0.0592	0.1164	0.1720	0.2257

Fig. 4 shows the results of this simulation step, plotting the perceptual distortion measured by EMBSD versus the network clouds' packet drop probability. The results demonstrate that the higher the packet drop probability is, the higher the perceptual distortion of the schemes and thus the worse the speech quality is.

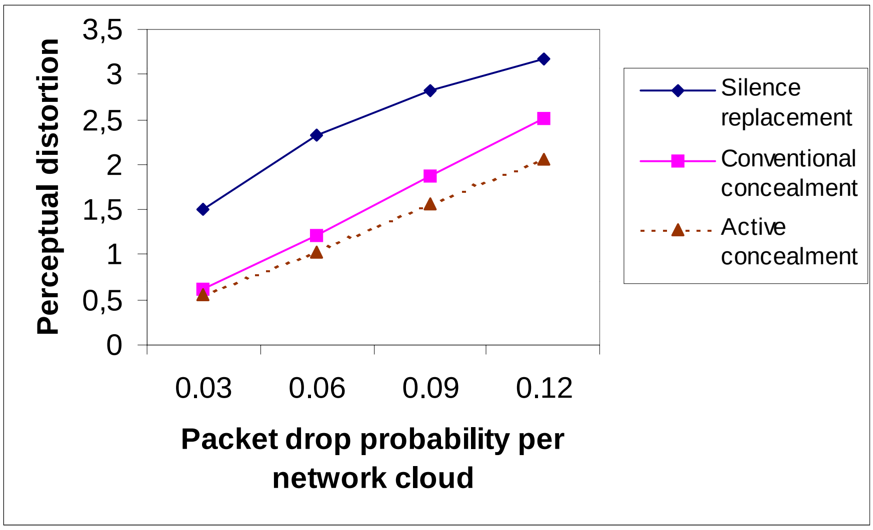

Fig. 4. Performance comparison to reference schemes (simulation step 1)

AP/C performs better than reference scheme 1 which replaces lost packets by silent segments, and the active loss concealment obtains the best speech quality. When the network clouds' packet drop probability is low, the active loss concealment does not gain any significant improvement compared to the AP/C scheme. This is because AP/C performs sufficiently well when the network loss rate is low and the number of burst losses is negligible. However, when the packet drop probability rises and the burst loss rate is no longer negligible, the perceptual distortion obtained with AP/C increases significantly and the active loss concealment achieves a clear improvement as compared to AP/C.

5.2 Optimal active network node location

In this simulation step, we vary the parameters of the lossy network clouds to determine the optimal location of the active network node. This simulation step is intended to help answering the following question: „Given that there are the same loss characteristics along the data path, where is the most effective location to download and perform the active concealment algorithm?"

The packet loss rate p of a data path consisting of two network clouds with packet drop probability p_1 and p_2 is given by

$$p = 1 - (1 - p_1) \cdot (1 - p_2) \qquad (1)$$

Thus, given the packet loss rate p and the packet drop probability p_1 of the first lossy network cloud, the packet drop probability of the second lossy network cloud is determined by

$$p_2 = \frac{p - p_1}{1 - p_1} \qquad (2)$$

The result of this simulation step is presented in Fig. 5 using EMBSD to compute the perceptual distortion of the obtained speech signal at the receiver.

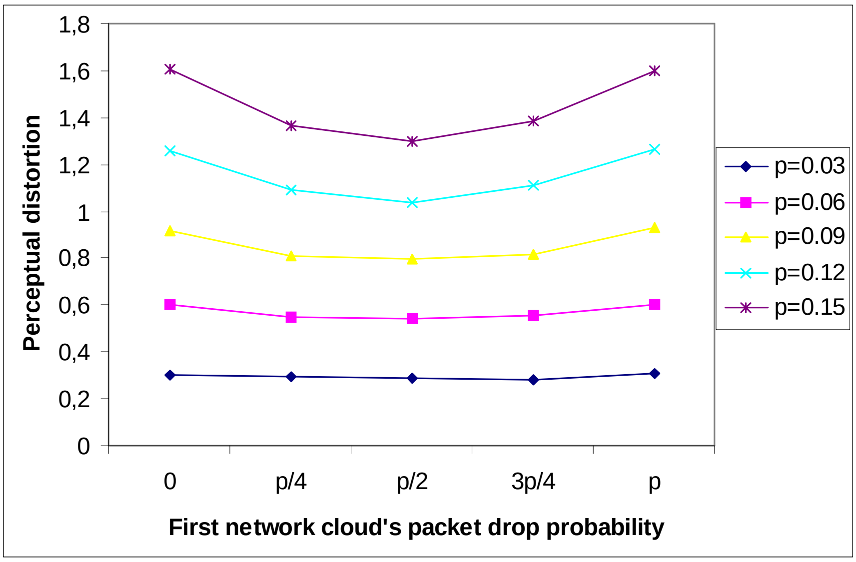

Fig. 5. Optimal active network node location (simulation step 2)

It shows that the optimal location to download and perform the active concealment algorithm is where the packet loss rate from the sender to that location is equal to the packet loss rate from there to the receiver ($p_1 = p_2$). Note that while we show the simulation results for $p_1 = p/2$, the actual minimum is located at $p_1 = p_2 = 1 - \sqrt{1-p} \approx 1 - (1 - p/2) = p/2$. If on one hand the packet loss rate from the sender to the location of the active network node is too high ($p_1 >> p_2$), the active concealment algorithm cannot exploit its advantage in terms of the location as compared to concealment just at the receiver. On the other hand, if the packet loss rate from the active network node to the receiver is too high ($p_1 << p_2$), the concealment algorithm at the active network node cannot be employed efficiently, because the majority of losses happen at subsequent network nodes. This effect is increasingly important when the packet loss rate (and thus the packet drop probability) increases, leading to a higher number of burst losses which causes the “conventional” concealment algorithm to fail.

6 Conclusions

We have presented a new active network application for voice over IP that exploits the flexibility of active networks to perform application-specific packet processing. By taking into account characteristics of the packet payload the efficiency of application-level algorithms has been leveraged. We have performed a simulation study to evaluate the efficiency of our approach. Simulation results have demonstrated that significant speech quality improvements are achieved compared to pure application-level algorithms. We also have run simulations to find the optimal location in a data path to download and perform the active loss concealment algorithm. It has been shown that the optimal location is where the network loss conditions are identical in both the up- and down-stream direction from the active node (considering deployment at only one active network node). An unoptimized software implementation of the active loss concealment reconstructs a lost packet with an average execution time overhead of 220 μs on a PC with a Pentium III 500 MHz CPU and 128 Mbytes RAM. Thus it is obvious that the active loss concealment does not increase significantly the end-to-end packet transmission delay. Since the active network node only performs packet regeneration for a small portion of packets of voice streams, the average consumption of node resources is reasonably low. With an optimized implementation, a significant reduction of additional delay and overhead in terms of node resource consumption can be expected.

While the resource consumption is thus not a problematic issue, the security and deployment implications of our scheme need still to be fully evaluated. For public multicast sessions, the operation is as follows: the active node adds itself as a member to the session. It receives and buffers enough data to perform loss concealment and then injects data on behalf of the sender (i.e. with the sender’s IP address) to the session. For unicast sessions the situation is more complicated as copies of packets of a particular connection need to be diverted to the loss concealment algorithm at a node and again re-created packets need to be sent in a way that pretends that they originated at the original sender. We consider these issues to be severe, however they fall into the general problem area of active network security. Thus, such problems should be solved at the active network platform level, i.e. the entities which provide for the deployment and execution of active code. We employ our algorithm on the BANG (Broadband Active Network Generation, [12]) platform which provides the needed security support.

Additional future work includes the investigation of active network applications where a number of active network nodes can be placed along the data path to download and perform the active loss concealment algorithm. Besides, it is very interesting to attempt to answer the question how well and how many times active loss concealment can be performed in a recursive way. Furthermore, since both application-level Forward Error Correction and application-specific packet processing incur additional consumption of network resources, we plan to compare these two approaches. The result of this comparison might enable an optimal combination of the two approaches to obtain further improvement of speech quality.

References

1. E. Amer, S. McCanne, and H. Zhang. An Application Level Video Gateway. Proceedings of ACM Multimedia 95, San Francisco, CA, November 1995.
2. A. Banchs, W. Effelsberg, C. Tschudin, and V. Turau. Multicasting Multimedia Streams with Active Networks. In Proceedings IEEE Local Computer Network Conference LCN 98, Boston, MA, Oct 11-14, 1998, pp 150-159.
3. S. Bhattacharjee, K. L. Calvert, and E. W. Zegura. An Architecture for Active Networking. High Performance Networking (HPN 97), White Plains, NY, April 1997.
4. J- Gruber and L. Strawczynski. Subjective Effects of Variable Delay and Speech Clipping in Dynamically Managed Voice Systems. IEEE Transactions on Communications, Vol. COM-33(8), August 1985.
5. Objective Quality Measurement of Telephone-Band (300-3400 Hz) Speech Codecs. ITU-T Recommendation P.861, February 1998.
6. R. Keller, S. Choi, D. Decasper, M. Dasen , G. Fankhauser, B. Plattner. An Active Router Architecture for Multicast Video Distribution. Proceedings IEEE Infocom 2000, Tel Aviv, Israel, March 2000.
7. L. Le. Development of a Loss-Resilient Internet Speech Transmission Method. Diploma thesis, Department of Electrical Engineering, Technical University Berlin, June 1999.
8. H. Sanneck. Adaptive Loss Concealment for Internet Telephony Applications. Proceedings INET 98, Geneva/Switzerland, July 1998.
9. D. Tennenhouse, J. Smith, D. Sincoskie, D. Wetherall, G. Minden. A Survey of Active Network Research. IEEE Communications, January 1997.
10. R. Valenzuela and C. Animalu. A New Voice Packet Reconstruction Technique. Proceedings ICASSP, pages 1334-1336, May 1989.
11. W. Yang, K. R. Krishnamachari, and R. Yantorno. Improvement of the MBSD by Scaling Noise Masking Threshold and Correlation Analysis with MOS Difference instead of MOS. IEEE Speech Coding Workshop, 1999.
12. Hitachi SDL, GMD Fokus. BANG (Broadband Active Network Generation) project. Berlin, 2000.