Performance of RTP/UDP/IP header compression in cellular networks

Akin Cellatoglu, Stewart Worrall, Simon N. Fabri, Ahmet M. Kondoz

Centre for Communications Systems Research, University of Surrey

Abstract: *A number of real-time multimedia services are being designed to operate over mobile end-to-end IP networks. Such services require the use of the Real Time Transport protocol (RTP) in order to provide acceptable Quality of Service at the receiver. The resulting RTP/UDP/IP headers however constitute a significant overhead, particularly at low-bitrate speech applications. This paper examines current header compression techniques for multimedia services over IP and examines their performance in mobile channel environments.*

1. Introduction

The Internet Protocol is being used as a Service Interface for a wide range of applications operating over an equally wide range of underlying network technologies. This allows for an independence between the application being run and the access network. For this reason, 3rd Generation mobile networks are being designed to support end-to-end IP service interfaces. This will allow for a wide range of applications to operate over the mobile networks. These applications are currently dominated by delay-insensitive data applications, but are now being joined by delay-sensitive applications such as conversational voice and video over IP, real-time signalling systems and streaming media services.

One of the main weaknesses of IP for the transport of real-time services is that it only offers a best-effort service class. This means that packets may not only be lost en route to their destination, but may also arrive out of sequence. In order to allow a receiving media player to reconstruct a coherent media stream from a sequence of received packets, the Real Time Transport Protocol is used. This protocol provides functions such as timestamping and sequence numbering to facilitate the reconstruction of received media streams.

2. CRTP

The combined headers required for the transmission of real-time multimedia information over IP-based networks totals 40 bytes. This includes the IP header (20 octets), the UDP header (8 octets) and the RTP header (12 octets). If IPv6 is used, this total is increased to 60 bytes. When operating over low throughput links, or when transmitting speech or audio streams which have been compressed to low data rates, this constitutes a considerable proportion of the total throughput, thereby decreasing transmission efficiency. This problem becomes very much more critical if Wireless Voice over IP schemes are to be implemented, as this will constitute a significant waste of radio spectrum. For this reason, a number of combined RTP/UDP/IP compression algorithms have been proposed.

One of these algorithms is the Compressed RTP (CRTP) scheme proposed by Casner and Jacobson [1]. The design of this algorithm is based on a combined TCP/IP header compression scheme described in [2], which reduces the number of bits in the headers by exploiting the fact that for the duration of the connection, half the bytes in the headers remain unchanged. The remaining compression is then achieved by differential encoding of the changing fields between headers of successive packets. In the CRTP scheme, a further gain is achieved from the fact that the difference of several fields between successive packets is constant, thereby representing a second-order difference of zero. This means that the transmitter need only transmit an initial uncompressed header and first-order differences, followed by indications of which fields in subsequent packets have zero second-order differences. The decompressor must therefore maintain a context consisting of the most recent uncompressed header received, which it combines with each received compressed RTP/UDP/IP header to reconstruct the original headers.

This predictive coding approach has a very serious weakness, namely its fragility when subjected to channel errors. If a compressed header is corrupted (this can be detected by employing a 16-bit checksum on the header), then all subsequent headers will be incorrectly reconstructed. The resulting out-of-date context will have to be replaced by a new uncompressed, and preferably undamaged, header. This is achieved by means of a feedback channel from the receiver to compressor, indicating it of the most recently correctly-decoded sequence number. However, if the round-trip delay is larger than the inter-packet spacing, as is often the case for compressed speech, then the decompressor will have to discard all outdated headers it receives before an entire (uncompressed) header arrives. If error detection is employed, an exponential refresh period may be used. When an error is detected at the decompressor, a new uncompressed header is transmitted. This is followed by a single compressed header and another uncompressed header. The period between header refreshes then increases exponentially to 2, 4, 8 etc. until a value of 255 is reached, as long as no errors are detected at the decompressor, in which case the period counter is reset to one. This exponential refresh period allows for the algorithm adapt to varying channel conditions.

3. Experiments

The performance of the RTP transport mechanism in cellular environments was evaluated using the General Packet Radio Service (GPRS) access network as a case study. GPRS is an end-to-end mobile packet radio communication system which makes use of the same radio architecture as GSM [3][4]. Although it was initially designed for use in non-delay-critical data applications, two of its features enable it to be used as a suitable medium for video communications. The multislotting capability of GPRS effectively allows for the throughput capability of a single terminal to be increased simply by allocating more timeslots (or Packet Data Traffic Channels) to a single terminal. In addition, its native IP support will allow for interworking with Internet multimedia applications. GPRS data is transmitted over the Packet Data Traffic Channel (PDTCH) and is protected by four different channel protection schemes [3]. CS-1, CS-2, and CS-3 use convolutional codes and block check sequences of differing strengths so as to give different rates. CS-4 on the other hand only provides error detection functionality. The experiments described in this section employ the CS-2 code at a carrier frequency of 1800MHz and using the TU50 (Typical Urban Scenario, mobile terminal velocity of 50 kph) Multipath model as specified in [5].

Scheme	**Code Rate**	**Radio Block**	**Data Rate kb/s**
CS-1	1/2	181	9.05
CS-2	≈2/3	268	13.4
CS-3	≈3/4	312	15.6
CS-4	1	428	21.4

Table 1. GPRS Coding Schemes

3.1 Protocol Efficiency

The G.723.1 codec was used to compress speech at an output rate of 5kbit/s. The resulting 192-bit frames generated every 30ms were encapsulated into discrete RTP/UDP/IP packets using a one frame-to-one packet mapping. If no header compression is used, it can be seen that the overhead as represented by total number of bits in IP packet divided by the number of speech coding bits, is 2.75. If the Casner-Jacobson scheme with header checksum is employed, this is reduced to 1.207, assuming a context refresh period of 255 packets. These values are independent of the transmission error rate (see Figure 1). However, if the exponential refresh period with feedback is used, it can be seen that the protocol efficiency decreases as the bit error increases. This is caused by the increase in the number of context refreshes that are necessary in error-prone environments. However, in relatively benign transmission environments (C/I>10dB), it can be seen that the difference in efficiency between the schemes with and without feedback is minimal.

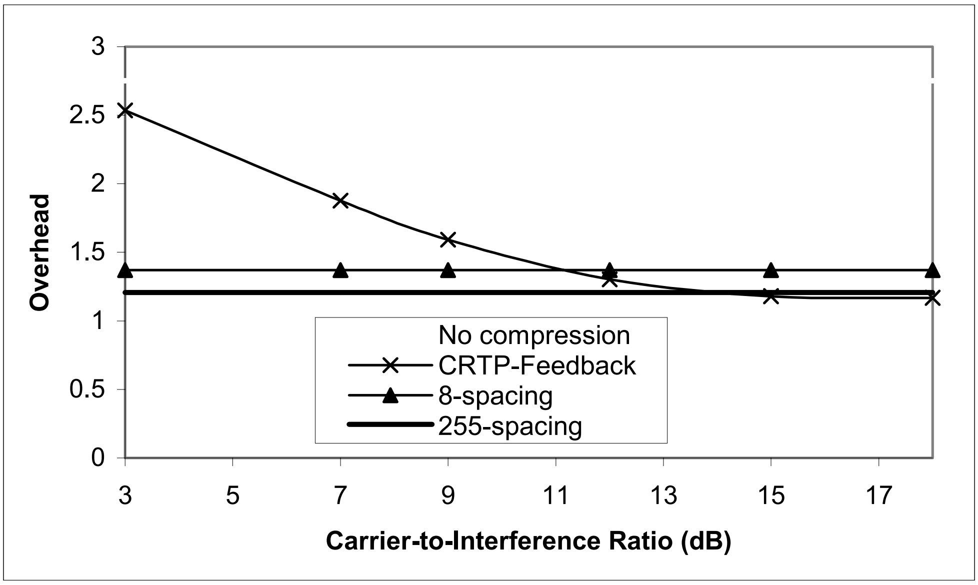

Figure 1. RTP/UDP/IP Header efficiency

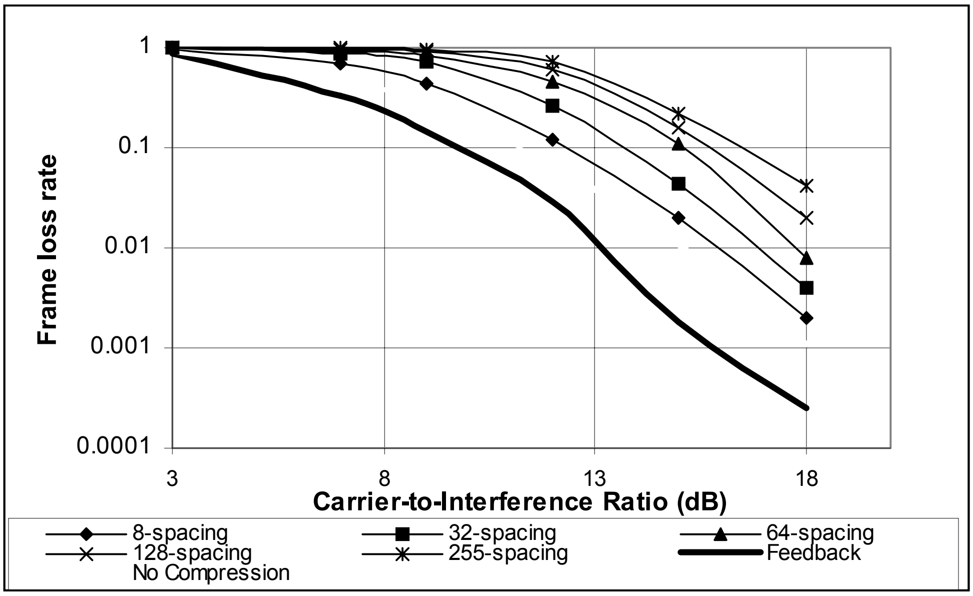

Figure 2. Packet dropping rate for RTP/UDP/IP compression schemes

3.2 Packet Loss Rate

As described above, the CRTP scheme increases the error susceptibility of the joint RTP/UDP/IP headers. As can be seen in Figure 2, when no feedback is used, the CRTP scheme produces higher packet dropping rates at all C/I ratios than for the case with no compression. However, if the feedback scheme with exponential refresh periods is introduced, the robustness is greatly improved, showing a superior performance to the uncompressed headers by about 2dB.

4. ROCCO

Although the CRTP with feedback scheme can achieve considerably better performance to both raw (uncompressed) RTP and CRTP without feedback, it has been shown that a deterioration in channel quality results in a considerable increase in the packet loss rate and in an increase in the packet overhead caused by the requirement to re-initialise the context at the receiver. For this reason, an alternative scheme has been developed [6], known as the robust checksum-based header compression (ROCCO) scheme. Like the CRTP scheme, a checksum is used to detect the presence of errors in the combined compressed headers. A number of different compression profiles are defined to allow for optimal compression performance for different traffic characteristics (such as audio and video) and for different channel conditions. The compressed ROCCO header does not only contain a CRC check, but there is also a code which specifies how the header fields have changed. This information allows the decompressor to carry out a local repair of the header context even if up to 25 consecutive packets have been lost or corrupted. This local repair functionality greatly reduces the need to for the decompressor to request a new (uncompressed) set of headers from the transmitter. This approach has been shown to achieve improved performance.

5. Conclusion

This paper has presented some of the issues involved in carrying out header compression for RTP/UDP/IP in cellular environments. The Casner-Jacobson algorithm was shown to provide efficient compression rates, but can only be used effectively across mobile radio channels if a feedback channel from the decompressor to the source is available. If no such channel is available, then uncompressed headers will result in lower packet loss rates than produced by the CRTP scheme.

References

1. Casner S., and Jacobson V., “Compressing IP/UDP/RTP Headers for Low-Speed Serial Links”, RFC 2508, February 1999.
2. Degermark M., Nordgren B., and Pink S., “Header Compression for Ipv6”, RFC 2507, February 1999.
3. ETSI/SMG, GSM 03.64 1998 “Overall Description of the GPRS Radio Interface Stage 2”, V. 5.2.0., 1998.
4. Brasche G. , and Walke B., “Concepts, Services and Protocols of the New GSM Phase 2+ General Packet Radio Service”, IEEE Communications Magazine, pp.94-104, October 1997.
5. SMG. “GSM 05.03 v6.1: Digital cellular telecommunications system (GSM Radio Access Phase 3); Channel Coding”, January 1998.
6. Jonsson L., Degermark M., Hannu H., and Svanbro K., “Robust Checksum-based header Compression (ROCCO)”, IETF Internet Draft, January 2000.