# Benefits of Header & Payload Compression in 2G, 3G & 4G Networks

August 2020

# Overview

Traffic overhead for mobile backhaul when using microwave or fiber-based networks are typically not a concern for Mobile Network Operators (MNOs). Capacity is relatively inexpensive and the cost and complexity of applying header and payload compression throughout an entire transport network can be quite high.

Over VSAT, the cost of providing bandwidth to carry that excessive overhead is relatively expensive. If MNOs can achieve bandwidth savings of 25 to 30% by eliminating overhead, then that should be an essential consideration for supporting those services over VSAT.

Comtech EF Data occupies a unique position in the VSAT market in that we provide *hardware accelerated payload compression* in our modems. We own AHA (http://www.aha.com/) a company specializing in ASICs for data compression, forward error correction and encryption. We apply AHA technology in our satellite modems through full wire speed lossless GZIP based payload compression in an ASIC, which saves bandwidth over the satellite link while not introducing additional latency, jitter or CPU workload.

While most competing VSAT platforms support header compression, none support low latency, lossless hardware-based payload compression at high data rates like we do. This deficiency has created a market for 3rd party external optimizers that try to achieve the same goal but cannot quite get the same level of efficiency both in terms of performance and cost.

Each mobile standard has its own nuances and the traffic mix is different in all networks, so there is no single answer as to how much bandwidth can be saved. Below, we will look at each standard and examine how our header and payload compression can provide benefits to reduce the amount of data sent over a satellite link.

# Support for 2G

IP 2G BTS has been deployed extensively by MNOs as part of an overall migration to packet-based backhaul and a single RAN topology. Each 2G vendor has its own proprietary way of implementing 2G over IP, which can lead to different amounts of IP overhead. We have observed cases in which, depending on vendor, only single voice packets are loaded into an IP packet (highest IP overhead to payload ratio). Other 2G vendors that more thoroughly understand the overhead issue provide bandwidth efficient backhaul solutions by loading a number of voice packets into single IP packets. Multiplexing a number of voice packets increases the voice traffic to overhead ratio making it more bandwidth efficient.

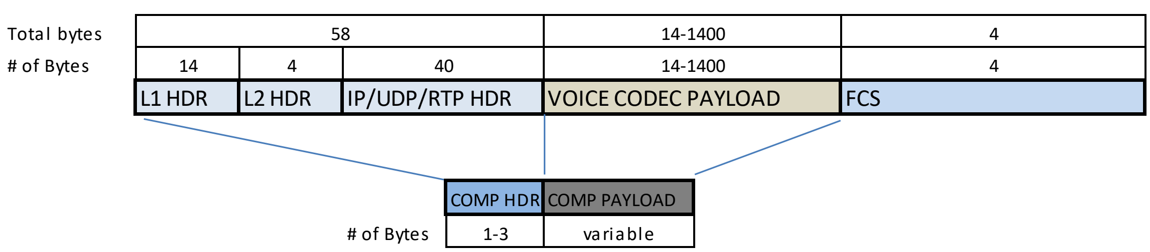


Figure 1: Typical 2G over IP Payload

We will look at two cases:

- Case 1 where the MNO operates the 2G traffic with one voice call per IP packet
- Case 2 where the MNO operates with 2G traffic with multiplexing enabled

**Case 1:** Each voice packet would have its own IP overhead and generate 50 packets per second/voice call. From the above, minimum overhead of 28 bytes (IP/UDP results in 12 kbps overhead) or 40 bytes (IP/UDP/RTP results in 16 kbps overhead) would support a payload of average voice codec of 8 kbps (mix of 12 kbps voice with 30% silence). Clearly, the IP overhead for each packet is then higher than the actual payload, and so optimization based on header and payload compression can significantly reduce the overhead and provide significant savings. In the case of using 3rd party external optimizers, the solution typically includes the following:

- Packet aggregation that multiplexes a number of voice packets together by implementing a wait time whereby voice packets are held back and grouped together (injects additional delay and jitter as the packets are held for a period of time)
- Wrapping the entire multiplexed packet into a new IP packet, which will go between the optimizers
- Header compression on the packets

As voice traffic increases, external optimizers can load more voice packets into a single large packet so that the benefits increase, and savings can reach upwards of 50%. By comparison, our modems' built-in header and payload compression eliminate much of the per voice session overhead and pushes the payload directly into the VSAT framer (with no added delay or jitter), and so there is virtually no overhead being transported across the VSAT and the savings are always maximized regardless of amount of voice traffic. Our modems reduce bandwidth over VSAT to the absolute bare minimum. In this case, observed savings are in the 50% range.

**Case 2**: Some BTS vendors (Ericsson, Huawei, etc.) can perform some level of multiplexing such as combining multiple voice sessions from a BTS into one larger packet, which reduces both the packet per second and overhead of supporting 2G over IP. In this case, solutions such as those provided by external optimizers provide little to no benefit as minimal savings are possible when further multiplexing traffic that a BTS has already multiplexed once. Nevertheless, to achieve a higher level of savings, external optimizers can increase capture timers to increase the amount of time they will collect packets upon which to perform packet aggregation. However, this has the net effect of increasing delay and jitter for minimal savings. With regards to our modem solution, each packet received has header and payload compression performed on it and is then mapped directly to the VSAT framer for transport over satellite (minimal overhead/maximum savings… all the time and with no added delay or jitter).

## 2G Data Services

2G supports two modes of packet data. The first generation is called General Packet Radio Service (GPRS). The size of a packet has a constant length, uses Gaussian Minimum Shift Keying (GMSK) modulation, and is embedded in the same GSM time slot that is used for voice. Hence, the packet overhead is similar to voice. By multiplexing together several time slots, it is possible to achieve up to 85 kbps download. The 2nd generation data services in GSM is called Enhanced Data rates for GSM Evolution (EDGE). It achieves a higher data rate throughput via improved coding (8PSK) and can achieve up to 236 kbps in the downlink direction. Just like with voice services over 2G, our header and payload compression are very effective in removing unnecessary bits across the entire Abis protocol, lowering the bandwidth required over the satellite link. One very telling example is an Ericsson 2G network deployed on a logistics company's vessels for M2M communication with refrigerated containers demonstrating over 50% savings.

## 2G Conclusions

As demonstrated in the cases above, bandwidth savings is dependent on vendor, traffic volume and traffic profile. No vendor should provide a "guaranteed savings" as such a guarantee would imply that those savings are independent of the traffic content and traffic profile.

Both the 3rd party external optimizers and our solution can provide real benefits. However, our solution provides key differentiators, including:

- The modems employ bandwidth optimization mechanisms that provide the lowest possible bandwidth per voice and data call independent of how the BTS encapsulates the traffic.
- The optimization does not add any additional delay nor jitter impact on the traffic.

External optimizers can provide similar benefits but, as the optimized traffic is always encapsulated in IP, it always has higher overhead than our modem solution, which eliminates all IP overhead going as the packet is forwarded to the VSAT framer. Additionally, external optimizers will always inject some additional delay and jitter into the 2G traffic that ultimately effect Key Performance Indicators (KPIs) and voice quality in the network.

# Support for 3G

3G follows an international 3GPP specified Iub interface, which all mobile infrastructure vendors agreed to follow. The Iub interface traffic supports voice traffic via its original Release 99 (R99) while data today is supported typically via 3GPP Release 5, or newer, which has evolved to support High Speed Packet Access (HSPA) and HSPA+.

## 3G Voice Traffic

All voice traffic in 3G follows the R99 formatting and typically carries 55% overhead or more.

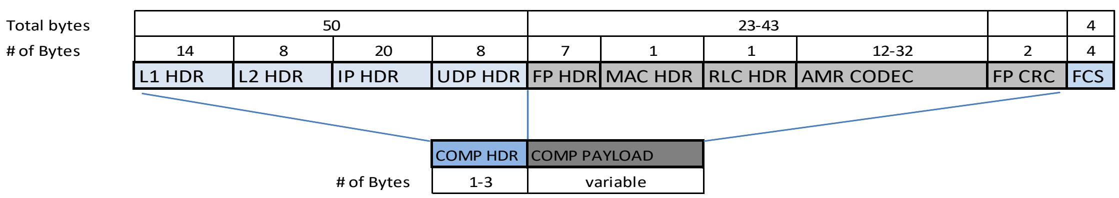


Figure 2: 3G lub Voice Packet Format

As in the case of 2G, our modem solution performs header and payload compression directly on the IP packet and then sends the optimized version across the VSAT framer. Minimum overhead to voice payload is the direct result with our modem solutions, which typically provide savings on the order of 55% (AMR 12.2) to 75% (AMR 4.7). Unlike 2G, there is no voice packet multiplexing in 3G. Therefore, 3G voice always suffers from high overhead, and there are 50 packets per second generated per voice call. 3rd party external optimizers are able to aggregate a number of Iub packets together and, in conjunction with header and payload compression, reduce the amount of overhead before repacking the voice packets into another IP packet. As in the case of 2G, the savings from external optimizers is never as high as our modem solution, and it always comes with the added cost of additional delay and jitter.

## 3G Data Traffic

Unlike voice, there have been many generational changes on how 3G data is supported as 3G evolved from R99 supporting 384 kbps to High Speed Download Packet Access (HSDPA) to improve download speed to 14 Mbps to High Speed Packet Access (HSPA+) that is a combination of High Speed Downlink Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA) supporting data rates in excess of 50 Mbps per user. Despite the transport being over IP, much of 3G still carries overheads which are more commonly associated to the older Asynchronous Transfer Mode (ATM)/TDM transfer modes. Understanding how 3G data packets are created is important to understand why 3G data typically is burdened with approximately 27% overhead.

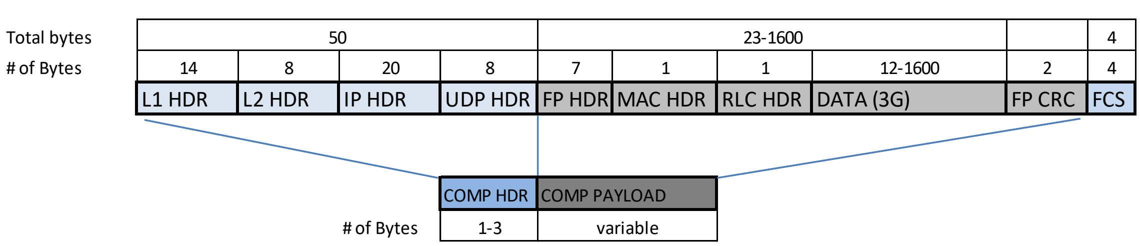


Figure 3: 3G Iub Data format

With our modem solution, the combination of header and payload compression negate much of the overhead such that the amount of bandwidth required to support the Iub interface is typically equivalent to the end user payload itself. On the returns, we typically observe greater than 35% savings which stems from a higher volume of smaller packets that comes from TCP/IP acknowledgments, keep alive messages from applications like WhatsApp, and base station signaling and radio control measurements being sent upstream. Since the packets are smaller, there are more padding and fill bits *within* the IuB protocol in the download direction which increases compressibility.

### 3G Conclusions

With our modem solution, the combination of Iub header and payload compression can typically contribute to 25% or more savings. The savings are performed on each individual packet directly and there is no additional delay or jitter. Furthermore, our modems come with enough Packet Processing per Second (PPS) capability to handle the most challenging 3G networks.

In the case of external optimization, the devices aggregate smaller packets and perform header and payload compression across the traffic. There is no question of the value that external optimizers are providing in reducing the number of PPS in the case of small packets. However, typically only 1/3 of the Iub traffic is small enough (less than 350 bytes) to make packet aggregation worthwhile. Most Iub data packets are large with almost 2/3 of Iub data packets averaging greater than 600 bytes. At this size, external optimizers would only be able to aggregate 2 Iub data packets together. Therefore, any aggregation and header compression benefits would be minimal and less than our modem solution.

3G is extremely sensitive to jitter, even jitter as low as 10 milliseconds can result in severely degraded performance and reductions in Radio Resource Control (RRC) KPIs. Refer to this paper for a discussion on the complications of jitter and latency in 3G. So, while packet aggregation and compression of Iub voice and data via external optimizer devices may seem like a good idea to save bandwidth across the VSAT, the additional delay and jitter result in degraded RRC KPIs for the MNO. Comparing our modem solution to external optimizers, it is clear that any responsible MNO should pick a solution that does not add any additional delay and jitter.

## Support for 4G

As was the case for 3G, 4G (LTE) services are well defined along 3GPP standards. In the case of 4G voice called VoLTE, the voice traffic is carried as Adaptive Multi-Rate (AMR) across LTE GPRS Tunneling Protocol (GTP) as shown below.

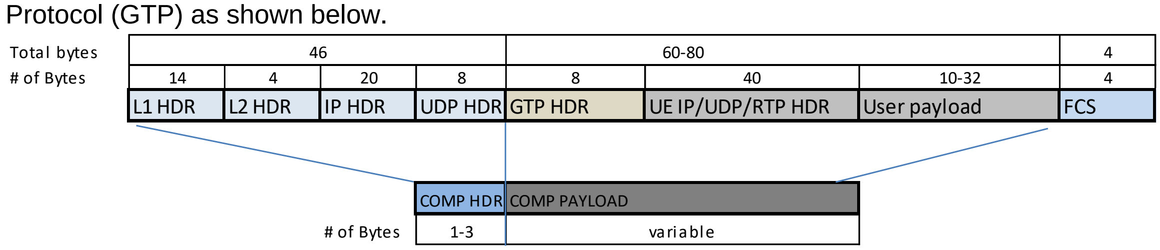


Figure 4

There are two AMR modes defined in VoLTE. One is the AMR Narrowband with a codec rate of 12.65 kbps, and the other is referred to as AMR Wideband or HD-voice at 23.85 kbps. Per 4G voice session, the amount of overhead is significantly higher compared to 2G or 3G due to double levels of header overhead (outer IP associated to the GTP tunnel and then inner IP associated to the end-user IP device). A typical 4G voice session can occupy 46 kbps (including L2 headers) to support an AMR NB voice codec (greater than 2 x overhead to voice codec ratio). Our solution performs header compression on the outer header information and payload compression on the GTP/inner IP header and payload content. We have experienced excellent results from this combination and voice overhead is typically reduced to a few kbps over the voice codec bandwidth (i.e. an AMR NB codec may result in 14-15 kbps across the VSAT). Savings on the order of greater than 60% can be achieved.

With regards to 4G data, the format of the traffic is similar to what was shown above, user payload has adopted standard Internet protocols like TCP (typically http/https), UDP or QUIC within the GTP tunnel. As above, the outer packet header overhead information is compressed via header compression while the inner payload (GTP/inner IP header/payload) is compressed.

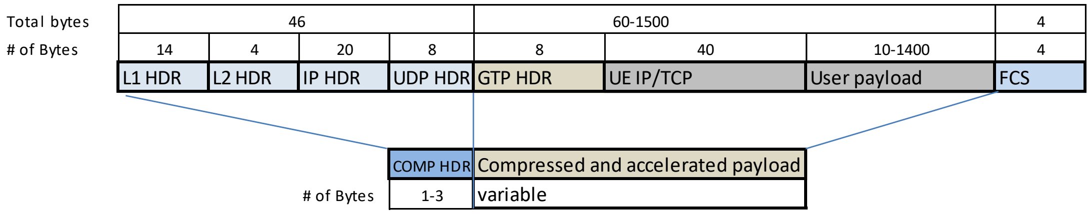


Figure 5

LTE was built from the ground up to be a data-centric service 3GPP decided to split the user plane data (S1-U) from the signaling plan (S1-C) allowing for each protocol to be bandwidth efficient. The overhead ratio of S1-U to user plane data is less than 15%, meaning that the effects of header and payload compression is less than in the other standards simply because there is less padding and framing to consider.

Today, a large percentage of Internet traffic is dominated by large packets (such as video/media, etc.) and/or encrypted traffic which negate some of the benefit of payload compression. While there is an increasing amount of Internet traffic that is being encrypted each year, this varies significantly between markets and mobile applications used. Based on our experience, savings of 10-15% are possible (based on 35% traffic being compressible). Just like in the case of 2G and 3G data, the uplink is always more compressible because of smaller packet size, more keep-alive/signaling messages and overall less encrypted end user traffic.

Perhaps one of the most important innovations in a satellite context that comes with 4G (LTE) is the ability to perform optimization of the end user traffic. We pioneered optimization of user plane traffic within the GTP tunnel to boost the performance of 4G services. This involves providing TCP acceleration and DNS caching benefits directly to end users at the 4G eNodeB, mitigating the effect of the VSAT delay and providing a fiber like service over VSAT and enabling proper function of applications and services which drive the Internet economy. GTP acceleration and DNS caching is included in our Heights Networking Platform, and again there is no need for an external optimizer to enable it.

### 4G Conclusions

Header and payload compression apply to both 4G IP headers as well as the GTP tunnel and its end user payload and in the case of VoLTE, can provide up to 60% savings. However, unlike 3G, there is less overhead in form of framing and padding (less than 15% on average versus 27% for 3G data) which compression algorithms can negate. As such, compression of 4G data traffic is still valuable and comes in at 5-15%, but the focus should also be given to GTP acceleration and DNS caching to improve Quality of Experience for end users.

## Summary

As demonstrated, each mobile standard has its own nuances in how voice and data is packetized and carried over the backhaul interface. However, common for all interfaces is that both header and payload compression is effective and provide an opportunity to save bandwidth over the satellite backhaul link. We are uniquely positioned in that we provide *hardware accelerated lossless payload compression* built-in to our modems that looks **within** the mobile protocols and removes unnecessary bits allowing satellite operators, service providers and mobile operators to achieve a greater net efficiency.

For additional information, please contact us.

**Email:** sales@comtechefdata.com
**Voice: +1.480.333.2200**
**Web:** www.comtechefdata.com