Improving Quality-of-Service of Real-Time Applications over Bandwidth Limited Satellite Communication Networks via Compression

LingSun Tan, SeiPing Lau and ChongEng Tan
Universiti Malaysia Sarawak, Malaysia

1. Introduction

VSAT (Very Small Aperture Terminal) satellite network is one of the widely deployed communication networks for rural and remote communications in today’s telecommunication world. VSAT satellite networks are growing steadily throughout many industries and market segments in many countries. With new applications and shifts in target markets, VSAT based solutions are being adopted at increasingly higher rates since year 2002 (MindBranch, 2011). Up to December 2008, VSAT market statistics show that the total number of Enterprise VSAT terminals being ordered is 2,276,348, the total number of VSATs being shipped is 2,220,280 and the total number of VSAT sites in service is 1,271,900 throughout the world (Comsys, 2008). VSAT satellite network offers value-added satellite-based services capable of supporting the Internet, data, video, LAN, voice and fax communications. VSATs are a single, flexible communication platform which can be installed quickly and cost efficiently to provide telecommunication solutions for consumers, governments and corporations, thus, they are becoming increasingly important.

VSAT satellite network plays an important role in bridging the digital divide and it is the one of the easiest deployment technology and cost effective way to interconnect two networks especially in rural areas, when other wired technologies are practically impossible and unsuitable due to geographical distance or accessibility. In this chapter, a fundamental overview of satellite communication network, with the highlighting of its main characteristics, constraints and proposal on compression technique which can be applied to boost up the Quality of Service (QoS) of the satellite communication services, are provided. VSAT satellite network provides communications support for a wide range of applications, which include point-of-sales transaction, financial management, telemetry & data collection, private-line voice services, virtual private networks, distance education, high speed internet access and more (TM, 2011).

VSAT satellite technology has many advantages. It can be deployed anywhere around the world and it offers borderless communication within the coverage area. Besides, it is cost effective and can be setup in a matter of minutes. VSAT network configuration such as bandwidth, interfaces and data rates can be updated remotely from the central network management system, hence, it provides high flexibility and efficiency. However, like other technologies, VSAT satellite network has its downsides. The limitations of VSAT technology
include the extremely high start-up cost needed for building and launching satellites in the geosynchronous orbit, high round-trip latency of about 500 ms as it utilise the satellites in geosynchronous orbit, and rain attenuation might affect the performance of VSAT communications under rainy conditions (TopBits.com, 2011). Moreover, it provides low and limited network bandwidth resulting in network congestion, reduced Quality-of-Service (QoS) of real-time interactive multimedia applications and also late packet delivery issues. These issues have created some negative impacts on the QoS of communication networks and also user experiences.

Apart from the need for efficient mechanisms for storage and transfer of enormous volume of data, these also lead to insatiable demands for ever-greater bandwidth in VSAT satellite network. In order to strike a balance between the cost and offered satellite bandwidth, some enhancements have to be implemented to reduce the bandwidth requirement of real-time applications that demanding high bandwidth and fully optimize the use of the low speed satellite link. Several techniques have been introduced to further improve the network bandwidth utilization and reduce network traffic especially for wireless satellite networks (Tan et al., 2010). One of such techniques is via compression, which is a technique used to overcome the network packet overhead by eliminating redundancies in packet delivery. By reducing the packet size, more packets can be transmitted over the same communication link at one time and hence increase the efficiency of bandwidth utilization. In this chapter, the concept of data compression is examined in order to know in depth how data compression can actually play a role in improving user experience. After that, the basic concept of packet compression, which consists of header compression and payload compression is also discussed.

Currently, there are many compression schemes, systems and frameworks have been proposed and designed in order to perform efficient data compression for better utilization of the communication channel. However, most of them have their own advantages and limitations, which may not suit for VSAT satellite network environment. For example, the Adaptive Compression Environment (ACE) system which has been proposed might impose additional delays over VSAT satellite network due to computation overhead and large compression time cost of the algorithm used. Besides, the Adaptive Online Compression (AdOC) algorithm which is proposed in the related work might cause the satellite link to be more congested due to the increased network load caused by the algorithm. In addition, some of the proposed compression schemes are designed for a specific aspect, which might create additional issues working under VSAT satellite network. Thus, in this chapter, the performance of several well-known compression schemes are reviewed and evaluated under the context of bandwidth limited VSAT satellite network, in order to highlight important criterions for improving performance over low bandwidth VSAT satellite network. Finally, the proposed enhanced compression scheme will be presented and the performance of the compression scheme will be examined and evaluated through extensive network simulations.

2. Introduction to VSAT communication

VSAT satellite network has become an essential part of our daily lives in recent years. It is used widely in telephony communication, broadband and internet services, and military communication. VSAT is a small satellite dish that is capable of both receiving and sending satellite signals (TM, 2011). It can be used for two-way communications via satellite.
Improving Quality-of-Service of Real-Time Applications over Bandwidth Limited Satellite Communication Networks via Compression

Generally, satellite is a specialized wireless receiver or transmitter that is launched by a rocket and placed in orbit around the earth (DotNetNuke Corporation, 2010). Thus, it is capable of providing coverage over large geographical areas and establishing communication links between various points on earth.

2.1 Basic satellite elements
Satellite communication system is comprised of two main components, namely space segment and ground segment, as illustrated in Figure 1 below. A basic satellite communication system consists of a space segment serving a specific ground segment (Richharia, 1999). The satellite itself is also known as the space segment while the earth stations will serve as the ground segment. The satellite is controlled and its performance is monitored by the Telemetry Tracking and Command (TT&C) station.

Communication can be established easily between all earth stations located within the coverage region through the satellite. The primary role of a satellite is to relay electronic signals. When signals from the earth stations are received by the satellite, the signals are processed, translated into another radio frequency and retransmitted down towards another
desired earth stations after further amplification. Satellite relay can be two way, as in the case of a long distance phone call, and point to multipoint, as in the case with television broadcasts.

2.2 Satellite roles and applications
The most important role of satellite communication network is to provide connectivity to the user terminals and to internetwork with terrestrial networks so that the applications and services provided by terrestrial network such as telephony, television, broadband access and Internet connections can be extended to places where cable and terrestrial radio cannot economically be installed and maintained. Satellite network provides direct connections among user terminals, connections for terminals to access terrestrial networks and connections between terrestrial networks (Mitra, 2005).

Since satellite is capable of providing coverage over a much wider area such as oceans, inter-continental flight corridors and large expanses of land mass, it is used in providing voice and data communications to aircraft, ships, land vehicles and handsets. Besides, satellite allows passengers on an aircraft to connect directly to a land based telecommunication network. Apart from that, it is also used for remote sensing, earth observation, meteorological applications such as weather survey, military communication and global positioning services (GPS).

2.3 Limitations of satellite communication
Three main characteristics and constraints of satellite network are high latency, poor bandwidth and noise (Hart, 1997). High latency is one of the main limitations of satellite network and it is caused by the long propagation path due to the high altitude of satellite orbits. In satellite network, the time required to navigate through a satellite link is longer compared to terrestrial network. Hence, this leads to higher transmission delay.

For geostationary (GEO) satellite communication system, the time required to traverse these distances, namely, earth station to satellite, then satellite to another earth station, is around 250ms (Sun, 2005). Round-trip delay will be 500ms. These propagation times are much greater than those encountered in conventional terrestrial systems. The high latency constraint of satellite link might not affect bulk data transfer and broadcast-type applications, but it will affect those highly interactive real-time applications.

Due to radio spectrum limitations, satellite transmission has a fixed amount of bandwidth (Hart, 1997). Problems like network congestion and packet loss might occur when those real-time interactive applications that consume high bandwidth are running over satellite link. Furthermore, strength of radio signal is in proportion to the square of distance traveled (Hart, 1997). Thus, signals traverse through satellite link might get very weak due to long distance between earth stations and satellite.

3. Data compression
Data compression plays an important role in improving the performance of low bandwidth VSAT satellite network. Among other satellite performance enhancement techniques, data compression is the most suitable and economical way to further improve the user experience of VSAT satellite network. This is because data compression technique is much simpler and can be implemented easily. Currently, a lot of the networking corporations are
Improving Quality-of-Service of Real-Time Applications over Bandwidth Limited Satellite Communication Networks via Compression

providing solutions for improving Internet services over satellite network by using high cost network equipments. These products are very costly and require complicated hardware configuration, while data compression is freely available and no complicated hardware configuration is required. Thus, data compression is adopted in the proposed scheme. Lately, data compression has become a common requirement for most application software as well as an important and active research area in computer science study. None of the ever-growing Internet, digital television, mobile communication or increasing video communication techniques would have been the practical developments without applying compression techniques.

In general, data compression is a process of representing information in a more compact form by eliminating redundancies in the original data representation (Pu, 2006). Due to the presence of redundancies in the original representation, data such as text, image, sound or any combination of all these types such as video is not in the shortest form, thus rendering its compression a possibility. Data compression is adopted in a variety of application areas such as mobile computing, image archival, video-conferencing, computer networks, digital and satellite television, multimedia evolution, imaging and signal processing. It can be divided into two major categories, namely lossless and lossy compression.

3.1 Lossless compression

In lossless compression, the exact original data can be reconstructed from the compressed data without any loss of information (Pu, 2006). Each compress-decompress cycle will generate exactly similar data, hence, lossless compression is known as reversible compression. Lossless compression techniques are used when storing medical images, text and images preserved for legal reason, some computer executable files, database records, spreadsheets or word processing files, where the lost of even a single bit could be catastrophic.

Example of lossless data compression is shown in Figure 2, where the exact input string FFMMMF is reconstructed after the execution of the compression algorithm followed by the decompression algorithm.

![Fig. 2. Example of lossless data compression (Pu, 2006).](./image.png)

3.2 Lossy compression

Lossy compression concedes a certain loss of accuracy in exchanges for greatly improved or more effective compression ratio. Owning to that, it does not allow the exact original data to be reconstructed from the compressed data (Pu, 2006). It usually suffers from information loss as compressing and decompressing the file repeatedly will cause loss of quality gradually. Thus, lossy compression is also known as irreversible compression. Lossy
compression is used frequently in streaming media and telephony applications as it is proven to be effective over graphic images and digitized voice. Lossy compression is not suitable for compressing text file formats due to loss of accuracy. However, it produces a much smaller compressed file than any known lossless compression method as it accepts some loss of data in order to achieve higher compression ratio. Figure 3 shows an example of lossy data compression, where a long decimal number becomes a shorter approximation number after the compression-decompression process.

![Image of lossy data compression](image_url)

Fig. 3. Example of lossy data compression (Pu, 2006).

4. Reviews of existing works

There are numerous compression schemes, systems and frameworks have been proposed and designed in order to improve the performance of communication network. The networking community has approached the problem of compressing network data streams as packets come by (online scenario), while the database community has focused more on applying such techniques to save storage space (offline scenario) (Chen et al., 2008). However, most of them have their own characteristics, which may not be suitable for satellite network environment. This section briefly discusses on packet compression, header compression and payload compression. Several well known compression schemes are also evaluated for their implementation under satellite network environment. These compression schemes can be further divided into three categories, namely packet compression schemes, header compression schemes and payload compression schemes.

4.1 Packet compression

Data compression in packet network is known as packet compression. Normally in computer network, network data will be divided into smaller chunks before transmission and transmitted as packets over the communication channel. Packet compression allows much smaller amounts of packet drops, more simultaneous sessions, and a smooth and fast behavior of applications (Matias & Refua, 2005). Since network packet consists of two parts, namely header and payload, as shown in Figure 4, therefore packet compression can be achieved by either header or payload compression, or the combination of both. Figure 5 depicts a basic packet compression.

![Image of network packet](image_url)

Fig. 4. Structure of a network packet.
4.1.1 Packet compression schemes
Packet compression is proposed by some of the related works, as discussed in the following sections.

4.1.1.1 IPzip
A comprehensive suite of algorithms known as IPzip is presented for network packet headers and payloads compression. IPzip is designed to exploit the hidden intra-packet correlation and inter-packet correlation properties of the data streams (Chen et al., 2008). After that, it produces an efficient compression plan, where the data streams both within and across packets are reorganized to improve the compression ratio. The compression plan is built in an offline phase as reordering of packets and fields is resource intensive. IPzip learns the correlation pattern over a training set, after that generates a compression plan and then compresses the original data set according to the plan. However, the performance of the current compression plan may decrease and new compression plan is needed due to the changes in the intrinsic network traffic pattern. Thus, the effectiveness of the compression plan will be monitored over time. Block compression is introduced, as IPzip aggregates similar packets into a block based on flow information before undergoing compression in order to increase compression ratio.

Unfortunately, IPzip may not suit for real-time processing as it needs to carry out offline training to produce the efficient compression plan. Besides, IPzip may not be able to react if the intrinsic network traffic pattern changes frequently, since the learning process to generate a new compression plan takes time and efforts. Moreover, IPzip will simply cause network congestion if the compression processing speed is slower than the relay processing speed as it compresses all blocks. In conclusion, IPzip is not suitable for satellite network environment.
4.1.1.2 Adaptive packet compression scheme for advanced relay node

This research work presented an adaptive lossless packet compression scheme especially for advanced relay node in the network. This scheme is proposed to mitigate network traffic congestion issue and it is based on the assumption that both the conventional packet delay and additional advanced functions can be performed by the intermediate nodes inside the network using computational or storage resources at the nodes (Shimamura et al., 2009). This scheme compresses the incoming packets adaptively and selectively according to some important metrics. Packet by packet compression is performed to evaluate the potential of adaptive packet compression inside network.

![Fig. 6. Adaptive compression model in advanced relay node (Shimamura et al., 2009).](image)

Figure 6 shows the model of adaptive compression scheme proposed for advanced relay node. Notice that an advanced relay node has two logical queues, which are CPU queue and relay queue. When the advanced relay node receives a packet, this compression scheme determines whether packet compression is effective according to the waiting time in the relay queue, compression processing time, packet size, output link bandwidth and compression ratio. If packet compression is beneficial, then compression will be performed on the packet by the advanced relay node before the packet moves to the relay queue. Simulation results show that this scheme succeeds in reducing the packet delay and packet discard rate. However, the compression ratio achieved via per packet compression is very much lower compared to the one with block compression. Therefore, this scheme will not help much in bandwidth saving when working under low bandwidth satellite link. In addition, massive computation which consume a lot of time needs to be done by the advanced relay node each time it receives a packet. Under a heavy traffic condition which is usually experienced in a low bandwidth satellite link, more and more calculation need to be carried out, thus, more and more delays being created, and finally creates a bad impact on user experience.

4.2 Header compression

The applicability of Internet technology over low speed and high delay links is threatened and reduced by large and repetitive packet headers. Some delay sensitive applications, such as remote login and real-time interactive multimedia applications, need to use small packets (Naidu & Tapadiya, 2009). However, the overhead of large packet headers on small packets can be prohibitive. A natural way to alleviate the problem is to compress packet header as packet header information shows significant redundancy between consecutive packets. Header compression makes more efficient use of link bandwidth in a packet switched network by leveraging header field redundancies in packets belonging to the same packet stream (Taylor et al., 2005).
Most of the header fields such as source and destination address remain constant throughout the duration of a flow, while other fields such as sequence numbers change predictably. Thus, the header size can be significantly reduced for most packets by sending static fields information only initially and utilizing dependencies and predictability for other fields. The reference copies of full headers must be stored at the context of compression and decompression sides in order to communicate and reconstruct the original packet headers reliably.

Initially, a few packets are sent uncompressed and they are used to establish the shared state called context on both sides of the link. The context comprises information about static fields, dynamic fields and their change pattern in protocol headers. The compressor will use this information to compress the packet as efficiently as possible and then the decompressor will decompress the packet to its original state. To correctly decompress the compressed packet header, synchronization between compressor and decompressor is mandatory.

4.2.1 Header compression schemes

In order to overcome packet header overhead, two popular header compression schemes have been developed. They are Van Jacobson header Compression (VJHC) scheme and ROBust Header Compression (ROHC) scheme.

4.2.1.1 Van Jacobson Header Compression (VJHC)

Van Jacobson Header Compression scheme was introduced by Jacobson in year 1990. This scheme is used to improve the interactive terminal response of low speed serial modem links and it is specially developed for Transmission Control Protocol/Internet Protocol (TCP/IP) (Jacobson, 1990). VJHC is commonly used to compress the header of IPv4/TCP packets.

![Diagram of header compression](image)

Fig. 7. The flow context concept in header compression (Suryavanshi et al., 2004).

As shown in Figure 7, the concept of flow context is used during the process of header compression. For each packet flow, the context is built on both the compressor and decompressor side and a unique context identity (CID) is assigned. The flow context information is made up of a collection of field values and change patterns of field values in the packet header.

To establish the context, first few packets of a newly identified flow are sent to the decompressor without compression. This is because the packet header information needed to be stored for future reference. Once the context is formed on both sides, the compressor starts compressing the packets and now only the encoded difference to the preceeding header is transmitted. When the compressed packets reach the decompressor side, the changes contained in the newly received compressed header is applied to the saved header in the context to obtain the uncompressed headers.
VJHC scheme stated that TCP/IP header fields can be grouped into several categories, which are constant, inferred and dynamic (Jacobson, 1990). Constant fields are those field values that remained unchanged between consecutive packets, hence, can be eliminated. Inferred fields are those fields that can be recalculated at the receiving end. For example, ‘total length’ and ‘header checksum’ field. The transmission efficiency can be improved significantly by suppressing inferred fields at the compressor and restoring them at the decompressor. The third group is dynamic fields which do no change frequently at the same time or change slightly, thus it can be omitted in most cases.

VJHC is proven to be effective towards header compression, as it can reduce TCP/IPv4 header from 40B to 4B, which is 10% of its original size (Tye & Fairhurst, 2003). However, the main disadvantage of VJHC scheme is it may lead to error propagation throughout the transmission when a compressed packet is lost on the link. This is due to the inconsistent context which will cause a series of packets to be discarded at the receiver end. Thus, VJHC scheme is not applicable under satellite link with high bit error rate as this will lead to higher packet drop which will cause the satellite link performance to become even worse.

4.2.1.2 ROBust Header Compression (ROHC)

Besides VJHC, ROBust Header Compression (ROHC) scheme is another well known header compression scheme. It is developed by ROHC working group of the IETF (Tye & Fairhurst, 2003). ROHC is used for compressing IP packet headers and it is particularly suitable for wireless network. ROHC scheme allows bandwidth savings up to 60% in VOIP and multimedia communication applications (JCP-Consult, 2008). In this scheme, compression and decompression are treated as a series of states.

Fig. 8. Compressor state diagram (Effnet, 2004).

As shown in Figure 8, ROHC compressor operates in 3 states, which are Initialization and Refresh (IR), First Order (FO) and Second Order (SO) (Effnet, 2004). The concept of flow context is also adopted in this scheme. The states describe the increasing level of confidence about the correctness of the context at the decompressor side. This confidence is reflected in the increasing compression of packet headers. Initially, the compressor will start with the lowest state and gradually moving to higher state. When there is any error occurred, which will be indicated in the feedback packets, the compressor will move to a lower state to resend packets to fix the error.

Similar to the compressor, ROHC decompressor also operates in 3 states, namely No Context, Static Context and Full Context as illustrated in Figure 9 below (Effnet, 2004). In the beginning of the packet flow, the decompressor will start in the first state, No Context as it has no context information available yet. Once the context information is created at the decompressor site, the decompressor will move to higher state, Full Context state. In the case of error condition, the decompressor will move to lower state to fix the error.
The major advantages of ROHC over VJHC are improved efficiency and high robustness. ROHC works well over links with high bit error rates and long round trip times such as cellular and satellite network. Moreover, its framework is extensible and it is designed to discover dependencies in packets from the same packet flow. However, ROHC scheme is very complicated to be implemented as it absorbed all the existing compression techniques. In addition, in ROHC scheme, the decompressor needs to generate feedback packet and send it back to the compressor to acknowledge successful decompression. Besides that, the context updating information is also sent periodically to ensure context synchronization. This will easily lead to network congestion when working under a low bandwidth satellite link with heavy traffic flows as ROHC scheme increases the network load by generating feedback and context information packets from time to time.

4.3 Payload compression

Packet payload is used to store user information and bulk compression method is usually used for compressing packet payload. Bulk compression treats information in the packets as a block of information and compresses it using a compression algorithm (Tye & Fairhurst, 2003). The compressor will construct a dictionary for the common sequences found within the information and then match each sequence to a shorter compressed representation or a key code.

Two types of dictionary, namely a running dictionary which based on the compression algorithm used or a pre-defined dictionary that can be used for bulk compression. In bulk compression, the decompressor must use an identical dictionary which is used during compression and bulk compression is known to achieve higher compression ratio. However, the data dictionary requires larger memory allocation and the dictionaries at both the compressor and decompressor sides have to be fully synchronized.

4.3.1 Payload compression schemes

Apart from packet and header compression, two payload compression schemes have been proposed by other researchers, which are Adaptive Compression Environment (ACE) system and Adaptive Online Compression (AdOC) algorithm.
4.3.1.1 Adaptive Compression Environment (ACE)

Adaptive Compression Environment (ACE) intercepts program communication and applies on-the-fly compression (Krintz & Sucu, 2006). On-the-fly or online compression is mandatory for those real-time interactive applications. ACE is able to adapt to the changes in resource performance and network technology, thus the benefits from using ACE become apparent when the underlying communication performance varies or the network technology changes as in mobile communication network. ACE employs an efficient and accurate forecasting toolkit, which is known as Network Weather Service (NWS) to predict and determine whether applying compression will be profitable based on underlying resource performance.

Short-term forecasts of compression ratio, compressed and uncompressed transfer time is made by NWS using a series of estimation techniques, together with its own internal models that estimate compression performance and changes in data compressibility. After that, based on the end-to-end path information obtained by NWS, ACE will select between several widely used compression techniques, which include bzip, zlib and LZO to perform the transparent compression at TCP socket level. ACE compresses data in 32KB blocks and a 4-byte header is appended to each block to indicate the block size and compression technique used. It is proven to improve transfer performance by 8-93 percent over commonly used compression algorithm (Krintz & Sucu, 2006).

However, ACE may introduce computation overheads due to massive amount of computation are needed during the prediction process. Besides, problem like prediction error which will lead to inaccurate decision may occur and large compression time cost of the compression algorithm such as bzip may impose additional delays. Thus, ACE may not be suitable and may impose additional delays over satellite network.

4.3.1.2 Adaptive Online Compression (AdOC)

This work proposed a general purpose portable application layer compression algorithm known as AdOC. AdOC is an adaptive online compression algorithm suited for any application data transfer and it automatically adapts the level of compression to the speed of the network (Jeannot et al., 2002). Multithreading and First-In-First-Out (FIFO) data buffer are two important features of this algorithm.

In this algorithm, the sender consists of two threads, namely compression thread and communication thread. Compression thread is used to read and compress the data, while communication thread is responsible to send the data. A FIFO data buffer is created to store the data prior to transmission. The compression thread will write the data into the FIFO data buffer, while the communication thread will retrieve the data from it. Thus, the compression level used in the process of compression is depending on the size of the FIFO queue.

To completely eliminate the overhead encountered when data cannot be compressed, AdOC algorithm compresses data into smaller and independent chunks. This made AdOC less reactive to short term changes in bandwidth, but keeping the same compression level for long runs of data also improves the compression ratio (Jeannot et al., 2002). However, too small chunks of data will simply caused overhead of FIFO queue, hence, the size of data chunks need to be determined appropriately. Since AdOC algorithm compresses data into smaller and independent chunks, network load may be increased and network congestion may occur when works under satellite network.
5. Proposed real-time adaptive packet compression scheme

An overview of the proposed real-time adaptive packet compression scheme, with the highlighting of its main concept and properties, is provided in this section. The block diagram of the proposed compression scheme, together with the explanation of each stage involved is also presented.

5.1 Concept of the proposed scheme

Concept of the proposed real-time adaptive packet compression scheme in satellite network topology is shown in Figure 10 below. As stated earlier, the main objective of this research study is to overcome the limitation and constraints of satellite communication link, which are high latency and low bandwidth, therefore the performance of the satellite link has become the main consideration in the proposed scheme. The proposed approach will focus only on the high latency satellite link area, where the proposed scheme will be implemented in both gateway A and gateway B. Both gateways will act as either compressor or decompressor as the communication channel between gateway A and gateway B is a duplex link.

In the proposed compression scheme, the concept of virtual channel is adopted to increase network performance and reliability, simplify network architecture, and also improve network services. Virtual channel is a channel designation which differs from the actual communication channel and it is a dedicated path designed specifically for both sender and receiver only. Since packet header compression is employed in the proposed scheme, thus this concept is mandatory to facilitate data transmission over the link. The duplex link between gateway A and gateway B in Figure 10 will act as the virtual channel, where the rules of data transmission and the data format used are agreed by both gateways.

Fig. 10. Concept of the proposed compression scheme.

The flow of data transmission between both gateways is as discussed in the following. When the transmitted data packets arrive at gateway A, the packets will undergo compression prior to transmission over the virtual channel. When the compressed data packets reach
gateway B, the compressed packets will first undergo decompression before being transmitted to the end user.

Apart from that, adaptive packet compression is mandatory due to the adoption of block compression in the proposed scheme. Although block compression helps to increase the compression ratio, however, it has its downside too. Block compression might impose additional delay when the compression buffer is filled in a slow rate due to lack of network traffic and a fast response is needed. This will further reduce the user experience of VSAT satellite network. Therefore, to avoid this, packet blocks are compressed adaptively when any of the predefined conditions is reached, which will be discussed in details in the following section.

5.2 Strength of the proposed scheme

The proposed real-time adaptive packet compression scheme has several important properties as discussed in the following. Firstly, the proposed scheme is accommodating all incoming packets. To fully exploit the positive effect of compression, the proposed scheme is not restricted to specific packet flow only but is applied to all incoming packets from numerous source hosts and sites. One unique feature of the proposed scheme is the adoption of virtual channel concept, which has not been used in other reviewed schemes. This concept simplifies packet routing and makes data transmission more efficient, especially when packet compression is employed. In the proposed scheme, to facilitate packet transmission over the communication channel, a peer-to-peer synchronized virtual channel is established between the sender (compressor) and receiver (decompressor). Moreover, another important feature, block compression approach is also introduced. Block compression exploits similarities of consecutive packets in the flow and compression is performed on an aggregated set of packets (a block) to further improve the compression ratio and increase the effective bandwidth.

Apart from that, both packet header and payload are being compressed in the proposed scheme. In many services and applications such as Voice over IP, interactive games and messaging, the payload of the packets is almost of the same size or even smaller than the header (Effnet, 2004). Since the header fields remain almost constant between consecutive packets of the same packet stream, therefore it is possible to compress those headers, providing more than 90% (Effnet, 2004) saving in many cases. This helps to save bandwidth and the expensive resources can be used efficiently. In addition to header compression, payload compression also introduces significant benefit in increasing the effective bandwidth. Payload compression compresses the data portion of the transmission and it uses compression algorithms to identify relatively short byte sequences that are repeated frequently over time. Payload compression provides a significant saving in overall packet size especially for packets with large data portions.

In addition, adaptive compression is employed in the proposed scheme. Network packets are compressed adaptively and selectively in the proposed scheme to exploit the positive effect of block compression while avoiding the negative effect. To avoid greater delay imposed by block compression, the set of aggregated packets (block of packets) in the compression buffer is compressed adaptively based on certain conditions. If either one of the conditions is fulfilled, the compression buffer is compressed. Else, the compression buffer will not be compressed. By combining all the features listed above, the performance of the proposed scheme will be greatly improved over other reviewed schemes.
5.3 Overview of the proposed scheme

Figure 11 below demonstrates the main components of the proposed real-time adaptive packet compression scheme. The compression scheme made up of a source node (Gateway A) which acts as the compressor and a destination node (Gateway B) which is the decompressor. A peer-to-peer synchronized virtual channel, which acts as a dedicated path, will be established between Gateway A and Gateway B. With the presence of virtual channel, packet header compression techniques can be performed on all network packets. Data transmission between Gateway A and Gateway B can be divided into three major stages, which are compression stage, transmission stage and decompression stage. Compression stage takes place in Gateway A, transmission stage in the virtual channel while the decompression stage will be carried out in Gateway B. Every data transmission from Gateway A to Gateway B will undergo these three stages.

![Diagram of the proposed compression scheme](image)

Fig. 11. Main components of the proposed compression scheme.

5.3.1 Compression stage

Once the incoming packets reach the Gateway A, the packets will be stored inside a buffer. This buffer is also known as compression buffer, as it is used for block compression, which will be discussed in details in the following section. Generally, in block compression, packets will be aggregated into a block prior to compression. The buffer size is depending on the maximum number of packet which is allowed to be aggregated.

Block compression is employed to increase the compression ratio and reduce the network load. The compression ratio increases with the buffer size, which means that the larger the buffer, the better the compression ratio, as more packets can be aggregated. However, block compression may lead to higher packet delays due to the waiting time in the buffer and also the compression processing time. The packet delay time is expected to increase with the
number of packet to be aggregated. Thus, larger buffer will have higher compression processing latency and also higher packet drops. Therefore, a trade off point is mandatory. Once the whole compression buffer fills up, it will be transferred to the compress module to undergo compression. The compression buffer will be compressed via a well known compression library known as zlib compression library (Roelofs et al., 2010). One apparent drawback of this scheme with block compression is a possible delay observed when the compression buffer is filled in a slow rate due to lack of network traffic and a fast response is needed. To address this shortcoming, the proposed scheme will compress the compression buffer adaptively whenever any of the following conditions are met:

a. The compression buffer reaches its predefined limit or has filled up.
b. A certain time threshold has been exceeded from the time the first packet being stored in the buffer and the buffer contains at least one packet.

After the process of compression, the compressed block will now enter the transmission stage.

5.3.2 Transmission stage
In this stage, the compressed block will be transmitted over the communication link, which is a virtual channel in this scheme, to Gateway B. The compressed block will transit from transmission stage to decompression stage when it reaches the Gateway B.

5.3.3 Decompression stage
The compressed block will be directly transferred to the decompress module once it reaches Gateway B. Decompression will then be performed on it to restore its original form. The original block of packets will be divided into individual packets according to the original size of each combined packet. After that, these individual packets are stored in the decompression buffer while waiting to be transmitted to the corresponding end user or destination node.

5.4 Block compression
Block compression exploits similarities of consecutive packets in the flow, as a specific number of packets are aggregated into a block before undergo compression. Due to the correlation of packets inside the packet stream, the compression ratio is greatly improved. Besides, block compression helps to reduce the heavy network load and avoid network congestion. This is because it reduces the number of packets needed to be transmitted over the communication link by encapsulating a significant number of individual packets into a large packet (block).

An example of block compression, where four network packets are collected in a compression buffer before being compressed and transmitted to the receiver, is shown in Figure 12. As mentioned earlier, one of the shortcoming of block compression is it may potentially add great packet delays, as the packets do not immediately be transmitted but instead stored in the compression buffer. This packet delay time is expected to increase with the number of packet to be combined.

For example, Table 1 below shows the total number of accumulated transmitted packet in 5 unit time for a high latency network with compression scheme (HLNCS) and a high latency network without compression scheme (HLN). Suppose that the number of packet to be encapsulated for the high latency network with compression scheme is 10.
Fig. 12. Block compression.

<table>
<thead>
<tr>
<th>Time</th>
<th>HLN</th>
<th>Time</th>
<th>HLNCS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of packet transmitted</td>
<td></td>
<td>No. of packet transmitted</td>
</tr>
<tr>
<td>1st</td>
<td>1</td>
<td>1st</td>
<td>0</td>
</tr>
<tr>
<td>2nd</td>
<td>2</td>
<td>2nd</td>
<td>0</td>
</tr>
<tr>
<td>3rd</td>
<td>3</td>
<td>3rd</td>
<td>0</td>
</tr>
<tr>
<td>4th</td>
<td>4</td>
<td>4th</td>
<td>0</td>
</tr>
<tr>
<td>5th</td>
<td>5</td>
<td>5th</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>Total</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 1. No. of transmitted packet for HLN & HLNCS.

Note that for HLN, there is no delay in transmitting the packet in each unit time and 5 packets are sent after 5 unit time, while for HLNCS, there is 4 unit time delay and 10 packets are transmitted at 5 unit time. Due to the waiting time in the compression buffer and the compression processing time, packet transmission is delayed. However, the total number of packet transmitted is almost double even though there is a small delay initially. Thus, with tolerable delay, block compression allows more packets to be sent at one time. A trade off value between the packet delay and number of packets to be combined needs to be determined.
6. Results & discussions

In this section, the proposed real-time adaptive packet compression scheme is evaluated and validated by simulations. Two important performance metrics of the scheme, which are packet drop rate and throughput of data transmission, are evaluated, as these two metrics are representing the Quality of Service of satellite link. The performance criteria can be defined as the following. Packet drop rate is the ratio between the total amount of packet loss due to buffer overhead (congestion) and transmission errors, and the total amount of packets being transmitted successfully, in percentage. Throughput is the ratio between the total amount of packets successfully delivered to the receiver and the time of the connection (2000 seconds). A discrete event network simulator known as ns-2 (VINT Project, 1995) has been used in building the simulation model to realize a simulative framework for studying and evaluating the performance of the proposed real-time adaptive packet compression scheme over high latency satellite network environment.

6.1 Simulation setup

This section describes the experimental environment used to present the characteristics and effectiveness of the proposed scheme. Figure 13 below depicts the simulation network topology, where $n$ users are connected to a source node through wired links and the source node is connected to the destination node via the high latency satellite communication link. Each wired link presents a capacity of 10 Mbit/s and a propagation delay of 1 ms. The proposed real-time adaptive packet compression scheme is implemented at both source and destination nodes. Different values of number of user and various satellite link characteristics are simulated to monitor the impact of the proposed scheme over satellite link. TCP continuous traffic flows are used throughout the simulations. All users transmit packets simultaneously to the destination node via the source node and each simulation is run for 2000 seconds.

The effectiveness of the proposed scheme is evaluated by comparing the performance metrics (packet drop rate and throughput of data transmission) of two scenarios: simulation running with proposed scheme and simulation running without the proposed scheme. For the scenario with proposed scheme, packet is compressed in the source node before transmitting over the satellite link and is decompressed when it reaches the destination node. For the scenario without the proposed scheme, normal data transmission is carried out. The packet trace data used throughout the simulations are captured from the research labs in University Malaysia Sarawak (UNIMAS), which composed of normal day-to-day traffic, typical for research purposes. The traces are taken by using a traffic capture utility known as Wireshark (Wireshark Foundation, 1998).

As shown in the Table 2 below, different simulation scenarios are used to evaluate the proposed scheme. Two scenarios, low bandwidth and high bandwidth, are simulated. In each scenario, five different number of user are used to vary the congestion rate of the satellite link, so that the impact of the proposed scheme on link with different congestion values can be examined. The compression rate used in the compression process is also varied for each value of number of user used, as depicted in Table 2. Compression rate is the size of the compression buffer for block compression. For example, compression rate with value 0 means no compression, compression rate with value 1 means packet by packet compression, compression rate with value 5 means 5 packets are to be aggregated in the compression buffer prior to compression, and so on.
Improving Quality-of-Service of Real-Time Applications over Bandwidth Limited Satellite Communication Networks via Compression

Fig. 13. Simulation topology.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Satellite link characteristic</th>
<th>User</th>
<th>Compression rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (low bandwidth)</td>
<td>Uplink bandwidth: 64kbps Downlink bandwidth: 256kbps Round-trip delay: 644ms</td>
<td>5, 15, 25</td>
<td>0, 1, 5 – 1000 (with step of 5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35, 45, 55</td>
<td>0, 1, 5 – 2000 (with step of 5)</td>
</tr>
<tr>
<td>2 (high bandwidth)</td>
<td>Uplink bandwidth: 1024kbps Downlink bandwidth: 2048kbps Round-trip delay: 644ms</td>
<td>5, 15, 25</td>
<td>0, 1, 5 – 1000 (with step of 5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35, 45, 55</td>
<td>0, 1, 5 – 2000 (with step of 5)</td>
</tr>
</tbody>
</table>

Table 2. Simulation scenarios.

6.2 Performance analysis

As discussed in previous section, block compression is employed in the proposed scheme and different sizes of the compression buffer (compression rate) are used in the simulation studies. Block compression helps to improve the packet throughput as more packets can be transmitted over the communication channel at the same time. However, it may lead to higher packet drop rate as the whole packet block will be discarded when it encountered errors or when it is lost in the middle of transmission. This condition is getting worse when a high compression rate is used. Thus, an appropriate compression rate is crucial in achieving a high packet throughput with acceptable packet drop rate. The tolerable value for packet drop rate is depending solely on the application requirements.

From the simulation results, compression rate which yields the highest packet throughput given that the packet drop rate is less than 5%, 10% and 15%, is selected. Thus, the results
can be divided into three cases. Case 1 considers packet drop rate not more than 5%, Case 2 limits the packet drop rate to 10% and packet drop rate less than 15% is considered in Case 3. A communication link with packet drop rate more than 15% is considered as a bad performance link even though the throughput obtained is very high. Therefore, packet drop rate more than 15% is beyond the consideration in this work. The results are presented in the following section.

6.2.1 Best compression rate distribution

Figure 14 & 15 below shows the distribution of best compression rate over the congestion rate for Case 1,2 and 3 in Scenario 1 & 2. Best compression rate is the compression rate that yields the highest throughput, with the condition that its corresponding packet drop rate does not exceed the limit in each case. Notice that in both scenarios, the best compression rate increases with the congestion rate. This means that a larger compression buffer, which can accommodate more packets, is favoured to obtain a higher performance when the link is getting more and more congested. In Scenario 1, due to the packet drop rate constraints that limits the highest throughput that can be achieved, the best compression rate line of Case 1 is slightly lower than the line of Case 2 & 3, while Case 2 & 3 both achieve similar results (overlapped lines). In Scenario 2, all three cases favour the same compression rates.

![Scenario 1 (Low bandwidth)](image-url)

Fig. 14. The best compression rate for each case in Scenario 1.

As shown in the figures, low bandwidth scenario requires higher compression rates (1 - 235) while high bandwidth scenario requires lower compression rates (1 - 15). This shows that the proposed scheme performs better in low bandwidth scenario compared to high bandwidth scenario. This is because high bandwidth scenario has sufficient bandwidth to accommodate heavy flows of traffic, thus compression might not be needed, while in the case of low bandwidth, compression is mandatory as bandwidth limitation problem will cause the communication link to be severely congested.
6.2.2 Packet drop rate distribution

Figure 16 & 17 above shows the distribution of packet drop rate over the congestion rate for the simulation running without the proposed scheme and simulation running with the proposed scheme (Case 1, 2 and 3) in Scenario 1 & 2. Notice that in both scenarios, the packet drop rate of simulation without compression increases with the congestion rate. This means that communication link with higher congestion value has higher packet drops. With the adoption of the proposed scheme, block compression reduces the heavy network load and hence avoiding network congestion. Thus, the packet drop rate can be reduced significantly as no packet being dropped due to buffer overhead at the router.
The proposed scheme succeeds in reducing the packet drop rate by 1 – 90 percent in Scenario 1 and 1 – 82 percent in Scenario 2. In Scenario 1, due to the packet drop rate constraint of 5%, the packet drop rate line of Case 1 is slightly lower than the line of Case 2 & 3, which with the limits of 10% & 15%. Since the best compression rates for Case 1, 2 & 3 are similar as illustrated in Figure 15, thus the corresponding packet drop rate values of these three cases are the same too.

6.2.3 Packet throughput distribution

Figure 18 & 19 below shows the distribution of packet throughput over the congestion rate for the simulation running without the proposed scheme and simulation running with the proposed scheme (Case 1, 2 and 3) in Scenario 1 & 2. The throughput for simulation without compression decrease with the congestion rate in both scenarios. The more congested the link, the more packets being dropped due to buffer overhead at the router, hence the lower the throughput. As shown in Figure 18 & 19, the proposed scheme succeeds in improving the throughput by 8 – 175 percent in Scenario 1 and 5 – 62 percent in Scenario 2. This is because by applying block compression, more packets can be transmitted over the communication link at one time, hence the throughput can be greatly improved.

Notice that the improvement of packet throughput in Scenario 1 is better than in Scenario 2. This also suggests that the proposed scheme is performing much more better in a low bandwidth scenario compared to a high bandwidth scenario. This is because compression might not be necessary in high bandwidth scenario, as there is no bandwidth limitation problem and sufficient bandwidth is provided to accommodate heavy flows. In contrast, applications are competing for the low and limited bandwidth when there are heavy flows in a low bandwidth scenario, thus, compression is required to further improve the network performance.
Fig. 18. The packet throughput distribution in Scenario 1.

Fig. 19. The packet throughput distribution in Scenario 2.

7. Conclusion

In this chapter, a real-time adaptive packet compression scheme for bandwidth limited high latency satellite communication network is presented. The bandwidth limitation problem of high latency satellite network has lead to several crucial network issues, as more and more applications require higher bandwidth allocation. The proposed scheme is intended to improve Quality of Service of real-time interactive applications by increasing the effective bandwidth usage of satellite network. Besides employing both header and payload
compression to achieve maximum bandwidth optimization, this scheme facilitates communication and reduces network processing complication by establishing a virtual channel between sender and receiver. As discussed earlier, virtual channel is a channel designation which differs from the actual communication channel and it is a dedicated path designed specifically for both sender and receiver only. Thus this concept is mandatory to facilitate data transmission over the link, as packet header compression is employed. Block compression is also adopted in the compression scheme to improve the compression ratio and reduce the network load.

To evaluate the performance and effectiveness of the proposed scheme, extensive simulations have been conducted using captured TCP traffic. The proposed scheme is evaluated under two main scenarios: low bandwidth and high bandwidth. Simulation results show that the proposed scheme succeeds in reducing the packet drop rate and improving the packet throughput significantly in both low and high bandwidth scenarios, as shown in Table 3.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Improvement Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Packet drop rate</td>
</tr>
<tr>
<td>Low bandwidth</td>
<td>Up to 90</td>
</tr>
<tr>
<td>High bandwidth</td>
<td>Up to 82</td>
</tr>
</tbody>
</table>

Table 3. Improvement percentage on packet drop rate and packet throughput.

Hence, it is proven that through the introduction of this scheme, the Quality of Service of real-time interactive applications over high latency satellite network can be greatly improved as the main concern of satellite network which is low and limited bandwidth is now not an issue anymore. Real-time interactive applications and software, which have high bandwidth demand, will now gain good user experience and satisfaction over satellite network.

8. Acknowledgement

This research work was funded by Fundamental Research Grant Scheme - FRGS/02(16)/737/2010(23), Universiti Malaysia Sarawak (UNIMAS). The authors would like to thank Faculty of Computer Science and Information Technology, UNIMAS for providing useful equipments and facilities. The authors would also like to thank the anonymous reviewers for their comments which helped in improving the quality of this chapter.

9. References

Improving Quality-of-Service of Real-Time Applications over Bandwidth Limited Satellite Communication Networks via Compression


JCP-Consult. (2008). JCP-C RoHC Headers Compression Protocol Stack, pp. 1-9, Cesson-Sevigne, France


Satellite communication systems are now a major part of most telecommunications networks as well as our everyday lives through mobile personal communication systems and broadcast television. A sound understanding of such systems is therefore important for a wide range of system designers, engineers and users. This book provides a comprehensive review of some applications that have driven this growth. It analyzes various aspects of Satellite Communications from Antenna design, Real Time applications, Quality of Service (QoS), Atmospheric effects, Hybrid Satellite-Terrestrial Networks, Sensor Networks and High Capacity Satellite Links. It is the desire of the authors that the topics selected for the book can give the reader an overview of the current trends in Satellite Systems, and also an in depth analysis of the technical aspects of each one of them.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: