

FM-Delta: Fault Management Packet Compression

Tal Mizrahi^{*†}, Yoram Revah^{*}, Yehonathan Refael Kalim[†], Elad Kapuza[†], Yuval Cassuto[†]

^{*}Marvell Semiconductors, [†]Technion

talmi@marvell.com, revahyo@gmail.com, {srefaelk@campus, eladkap@campus, ycassuto@ee}.technion.ac.il

Abstract—Fault Management (FM) is a cardinal feature in communication networks. One of the most common FM approaches is to use periodic keepalive messages. Hence, switches and routers are required to transmit a large number of FM messages periodically, requiring a hardware-based packet generator that periodically transmits a set of messages that are stored in an expensive on-chip memory. With the rapid growth of carrier networks, and as 5G technologies emerge, the number of users and the traffic rates are expected to significantly increase over the next few years. Consequently, we expect the on-chip memories used for FM to become a costly component in switch and router chips. We introduce a novel approach in which FM messages are stored in compressed form in the on-chip memory, allowing to significantly reduce the memory size. We present FM-Delta, a simple hardware-friendly delta encoding algorithm that allows FM messages to be compressed by a factor of 2.6. We show that this compression ratio is very close to the results of the zlib compression library, which requires much higher implementation complexity.

I. INTRODUCTION

A. Background

Network devices, such as switches and routers, are often required to transmit control-plane messages. The ability to generate and transmit messages has been recognized as an essential building block in network devices, not only in conventional networks, but also in Software-Defined Networks (e.g., [1]). Specifically, the ability to generate packets is important in the context of Fault Management (FM).

Fault detection is essential in large-scale networks, enabling fast recovery and effective troubleshooting; it is one of the key components in Operations, Administration, and Maintenance (OAM) [2]. FM is typically implemented as a combination of proactive and reactive mechanisms for detecting failures. Some of the most commonly deployed FM protocols (e.g., [3], [4], [5]) are implemented using periodic keepalive messages; a fault is detected when no keepalive messages are received from a given source for a given period of time.

B. FM Scaling

The scaling problem of storing FM messages is a real-life problem that some of the authors of this paper encountered while designing packet processor silicons. A switch or a router that runs an FM protocol periodically transmits FM messages. The rate of FM messages varies from 1 packet every ten minutes to 300 packets per second per service [3], [4]. Typical carrier Ethernet devices support tens of thousands of services (e.g., 16k in [6] and 64k in [7]). Due to these large scales, typical devices do not provide FM support at the highest rate for all services simultaneously. As 5G technologies evolve,

network devices will be required to support a larger number of services at a higher bandwidth. Thus, fault detection will be expected with a low detection time, implying a high rate of FM messages. Hence, we expect these scales to continuously increase in the next few years.

Example. The rate of traffic that a network device is expected to generate for 64k services, assuming the FM message length is roughly 100B [3], and assuming 300 packets per second per service [3], is $64k \times 300 \times 100B \cong 16 \text{ Gbps}$. Such high traffic rates cannot be handled by the device's software layer, and thus must be implemented in the device's hardware. A typical implementation (e.g., [8]) uses a hardware engine and an on-chip memory; the engine periodically reads the messages stored in the memory and transmits them to the network. In this case the required memory space for 64k services is $64k \times 100B \cong 6.4 \text{ MB}$. This is a very significant size for an on-chip memory, consuming an expensive area of roughly 18 mm^2 in a 28nm process. As a point of reference, the entire on-chip packet memory of a typical switch is on the order of a few megabytes, e.g., 12 MB in [9].

This example illustrates that as carrier and mobile backhaul networks evolve, the significant size of the FM on-chip memory may become infeasible.

C. FM Packet Compression in a Nutshell

We argue that if FM packets are compressed when stored in the memory, then the expensive on-chip memory can be significantly reduced.

In our approach FM packets are compressed offline, and stored in compressed form in an on-chip memory. Each packet is decompressed by the hardware packet generation engine before it is transmitted to the network. Notably, packets are in compressed form only when stored in memory, and are decompressed before transmission, making the compression transparent to the network.

Why do we focus on FM messages? The solution we propose relies on three properties that are unique to the problem at hand: (i) the entropy of the FM packets stored in the memory is low, as they share common properties, (ii) the FM packet memory is accessed sequentially, since all packets are sent periodically, and (iii) we can determine the *order* of the FM packets in the memory.¹ These three properties significantly improve the effectiveness of the compression.

¹In the FM-Delta system we present we have full control of the order in which packets are stored in the memory, and thus the order in which packets are accessed and decompressed.

D. Related Work

Data compression has been analyzed in the context of network switches and routers, e.g., [10], [11]. Specifically, data-plane packet compression has been widely discussed and analyzed in the literature, e.g., [12], [13]. For example, packet compression is widely used in the Hypertext Transfer Protocol (HTTP) [14]. It has been shown that HTTP compression can be very effective [15], in some cases reducing the data to 20% of its original size. However, it has been shown [12] that packet compression is ineffective for random internet traffic, as this traffic typically has high entropy, and is often already compressed or encrypted. In this paper we focus on control-plane packet compression, and present a use case in which packet compression is highly effective.

E. Contributions

The contributions of this paper are as follows:

- We present a novel approach that uses packet compression to reduce the size of the on-chip memory used in Fault Management (FM) message transmission.
- We introduce FM-Delta, a simple delta encoding algorithm that can easily be implemented in hardware, and present a high-level design of a system that uses FM-Delta.
- We evaluated the algorithm over a synthetically generated FM packet database, and show that it offers a compression ratio that is comparable to state-of-the-art compression algorithms, such as Lempel-Ziv (LZ).² The FM packet database that we used in our experiments is publicly available [17].

II. A SYSTEM DESIGN

A. Overview

Fig. 1 illustrates the design of a system that uses FM-Delta to periodically generate FM packets. The system consists of two main components:

- Software module – responsible for compressing the FM packets and storing them in the on-chip memory of the packet generation engine. The compression procedure is performed offline. When an FM packet needs to be added to the FM packet database, or removed from it, the software layer invokes an *insert* or a *remove* operation, instructing the packet generation engine to perform an incremental update in the FM database.
- Packet generation engine – this hardware module sequentially reads the packets from the on-chip FM packet memory, decompresses them in real-time, and transmits them to the network.

²It is important to note that known LZ implementations (such as zlib [16]) are unlikely to be realizable in the low-level hardware setup of interest here, due to their significant processing and memory requirements.

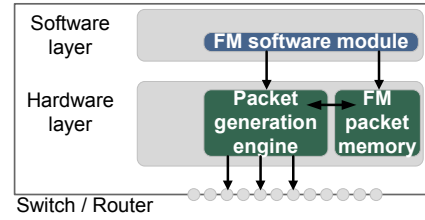


Fig. 1: FM-Delta: a system design.

B. Compression Algorithm

We introduce FM-Delta, a simple delta encoding algorithm that we designed and implemented.

FM-Delta. We present an outline of the FM-Delta algorithm. Given a sequence of N uncompressed packets, the compressed packets are represented as follows.

The first packet is represented in uncompressed form, as shown in Fig. 2.

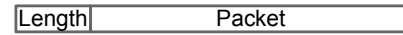


Fig. 2: FM-Delta: the first packet is in uncompressed form.

The i^{th} packet, for $i \geq 2$, is compressed, as shown in Fig. 3.



Fig. 3: FM-Delta: structure of the i^{th} compressed packet.

The compressed packet consists of the following fields (Fig. 3):

- **Length** – represents the length in bytes of the original (uncompressed) packet.

DECOMPRESS(N)

```

1  Read packet  $P_1$ 
2   $L_1 \leftarrow$  Length of  $P_1$ 
3   $U_1 \leftarrow$  uncompressed packet  $P_1$ 
4  for  $i = 2$  to  $N$ 
5    Read compressed packet  $P_i$ 
6     $D \leftarrow$  Delta Bitmap of  $P_i$ 
7     $V \leftarrow$  Values of  $P_i$ 
8     $m \leftarrow 0$ 
9    for  $j = 0$  to  $\min(L_i, L_{i-1}) - 1$ 
10   if  $D[j] = 1$ 
11      $U_i[j] \leftarrow U_{i-1}[j]$ 
12   else
13      $U_i[j] \leftarrow V[m]$ 
14      $m \leftarrow m + 1$ 

```

Fig. 4: Decompression algorithm for N packets. The i^{th} compressed packet is denoted by P_i , and the i^{th} uncompressed packet by U_i .

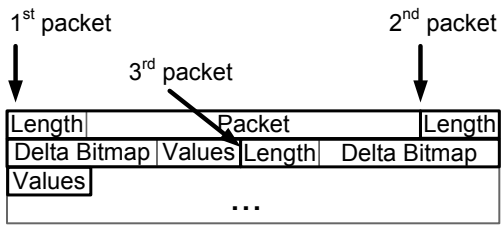


Fig. 5: FM-Delta: the compressed packets in the memory.

- **Delta Bitmap** – indicates the differences between packet i and packet $i - 1$. Every packet is divided into equal-sized words, and each bit in the bitmap indicates whether the respective word in the current packet is equal to the corresponding word in the previous packet. The word size is a parameter of the algorithm. Section III discusses the effectiveness of the algorithm with various word sizes.
- **Values** – the values of the words that differ from the previous packet.

The decompression algorithm is presented in Fig. 4.

Common (de)compression algorithms, such as Lempel-Ziv [18] are not hardware-friendly, as they require a complex iterative indirection over a sliding window. In FM-Delta, every packet is decompressed by comparing it to the previous (decompressed) packet using the Delta Bitmap. This simple comparison does not require an iterative procedure, and thus all the words of the packet can be compared (decompressed) in parallel, making FM-Delta a silicon-friendly algorithm.

C. Insertion and Removal

As described above, the packet database is compressed offline, and accessed by the chip in real-time. An essential question is how the packet database is updated, i.e., how a new packet can be added to the database, or how a packet can be removed from it.

One approach that allows simple insertion and removal is by using a linked list. Each compressed packet in the memory is followed by a pointer that points to the location of the next packet in memory. The linked list approach is simple, but requires some memory overhead for the next pointers, and also may be inefficient due to fragmentation.

We suggest a more efficient approach that allows for simple insertion and removal of packets when using FM-Delta. In this approach (Fig. 5) packets are stored contiguously in the memory. The packet generation engine proceeds by a read-decompress-write procedure; every packet is read from memory, decompressed, and then written back to the memory. There is a cache that stores the previous packet, which enables the engine to decompress the current packet using FM-Delta.

The packet ordering, insertion, and removal are described in further detail in the extended version of this paper [19].

D. Random Access to FM Packets

In our approach we assumed that the FM packet memory is always accessed sequentially. However, in some cases the FM application may require an *urgent* packet P_k to be sent immediately. The delta encoding scheme implies that in order to access the k^{th} packet the packet generation engine must sequentially read all the preceding packets.

In order to allow quick random access, our delta encoding scheme can be extended to include *entry points*. For example, the FM software layer can store packets $P_{10}, P_{20}, P_{30}, \dots$ in uncompressed form. Thus, when the software layer instructs the packet generation engine to access the 21st packet, it also provides the uncompressed P_{20} , and the engine can use P_{21} with one memory access. In this example every packet in the FM memory can be reached in at most 9 memory access operations.

III. EVALUATION

A. Data Set

We evaluated our FM-Delta compression system on 100 sets of synthetically generated packets.³ Each set comprises 100k packets. Our synthesized data sets are publicly available [17].

The data sets consists of two types of packets: Continuity Check Messages (CCM) [4], and Bidirectional Forwarding Detection (BFD) [5] control messages. Each packet type was used on half of the data set. CCMs are defined over Ethernet, while BFDs are over IPv4-over-Ethernet. Each packet included a random number of VLAN tags (either 0, 1, or 2 tags).

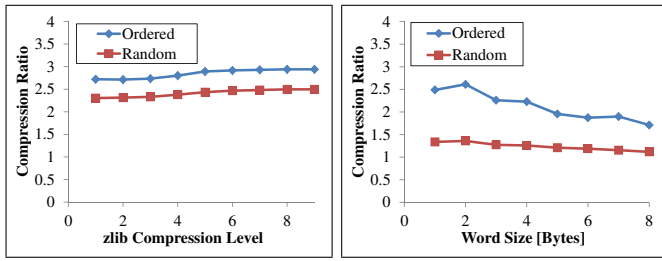
The network topology can significantly affect the extent to which FM packets can be compressed in our setting. For example, if multiple FM packets are sent to the same destination device, then some of the packet fields may be similar in these packets. Moreover, in CCMs [4] the **MEG ID** field is a 48-byte field that has a different value in each Maintenance Entity Group (MEG) [4]. Thus, the number of devices per MEG can significantly affect the similarity among packets.

We assumed that the current device has a set of 32 MAC addresses, and thus the source MAC address of each packet was randomly chosen from the pool of 32 addresses, while the destination MAC address was randomly chosen⁴ without constraints. The IP addresses of each BFD packet were randomly chosen. The VLAN IDs of each packet that included a VLAN tag was also randomly chosen. The MEG ID field in CCM packets was randomly chosen for each set of 3 CCM packets.⁵

³We did not use publicly available packet traces since these traces either do not include FM packets, or do not include packet payloads, thus preventing effective data compression analysis.

⁴Throughout this section the ‘randomly chosen’ refers to a uniformly distributed selection.

⁵We assumed that the number of Maintenance Points (MP) [4] per Maintenance Entity Group (MEG) [4] is 4 on average.



(a) zlib: Compression ratio as a function of the zlib compression level. (b) FM-Delta: Compression ratio as a function of the word size.

Fig. 6: Compression ratio of the two analyzed algorithms. The ‘Ordered’ curve is most interesting, since we have full control over the order of packets in the FM packet memory. In FM-Delta the best compression ratio is achieved when the word size is fixed at 2 bytes.

Our experiments were performed in two modes: **ordered** mode, in which packets were arranged so as to allow a higher compression ratio, and **random** mode, in which packets were ordered randomly. In the ordered mode, the arrangement was implemented according to two criteria: (i) packets were ordered according to their size, allowing alignment between the fields of consecutive packets, and (ii) packets from the same MEG were grouped together.

B. Results

We present a glimpse at some of our experimental results. Fig. 6a illustrates the compression ratio⁶ using zlib. Since zlib supports nine possible compression levels, the graph presents the compression ratio for each of the levels.

Fig. 6b presents the compression ratio of the delta encoding algorithm as a function of the word size. The word size is a design parameter of the FM-Delta algorithm. Specifically, for the FM protocols that were analyzed in this work, if the word size is fixed at 2 bytes, we expect the algorithm to provide the best performance for the FM protocols we analyzed.

Notably, although the delta encoding algorithm is significantly simpler, it provides a comparable compression ratio; zlib provides a compression ratio of 2.9 at the highest compression level, but is not hardware-friendly, while the hardware-friendly FM-Delta provides a ratio of 2.6.

Another significant result is that the ordered message set allows a higher compression ratio, emphasizing the advantage of optimizing the order of the data set. Such an optimization is feasible in light of the relative flexibility of FM specifications as to the order and exact timing of packet transmission.

⁶The compression ratio is the ratio between the uncompressed data set and the compressed data set.

IV. CONCLUSION AND OUTLOOK

In this paper we have shown that packet compression can significantly reduce the required on-chip memory in Fault Management protocol implementations. Surprisingly, delta encoding, which is typically dismissed as an ineffective packet compression approach, is shown to be highly effective for packet compression. We introduce a simple and hardware-friendly delta encoding algorithm that provides a compression ratio of 2.6, and allows to reduce the cost of packet processor silicon.

Potential improvements of FM-Delta can be considered, for example by using a dictionary for common field values. Furthermore, the preliminary results presented in this paper can be further established by experimenting with a hardware implementation of FM-Delta, and by analyzing real-life traces of FM messages. Notably, the concepts we presented may be applied not only to FM messages, but also to other types of control messages that are stored by switches and routers.

REFERENCES

- [1] “OF-PI: A Protocol Independent Layer,” ver. 1.1, ONF, 2014.
- [2] T. Mizrahi, N. Sprecher, E. Bellagamba, and Y. Weingarten, “An Overview of Operations, Administration, and Maintenance (OAM) Tools,” RFC 7276, IETF, 2014.
- [3] IEEE 802.1ag, “Connectivity Fault Management,” 2007.
- [4] ITU-T G.8013/Y.1731, “Operations, administration and maintenance (OAM) functions and mechanisms for Ethernet-based networks,” 2015.
- [5] D. Katz and D. Ward, “Bidirectional Forwarding Detection (BFD),” RFC 5880, IETF, 2010.
- [6] “Ericsson SPO 1400 Family,” tech. rep., 2012.
- [7] “Huawei OptiX OSN 550 and OSN 3500,” tech. rep., 2011.
- [8] T. Mizrahi and I. Yerushalmi, “The OAM jigsaw puzzle,” technical white paper, Marvell, 2011.
- [9] “Pica8 32x40 gbe,” <http://www.pica8.com/wp-content/uploads/2015/09/pica8-datasheet-32x40gbe-p5401.pdf>, 2014.
- [10] O. Rottenstreich, M. Radan, Y. Cassuto, I. Keslassy, C. Arad, T. Mizrahi, Y. Revah, and A. Hassidim, “Compressing forwarding tables for data-center scalability,” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 1, pp. 138–151, 2014.
- [11] G. Rétvári, J. Tapolcai, A. Kőrösi, A. Majdán, and Z. Heszberger, “Compressing ip forwarding tables: towards entropy bounds and beyond,” in *ACM SIGCOMM*, 2013.
- [12] C. S. Tye and G. Fairhurst, “A review of IP packet compression techniques,” *Proc. PGNet*, p. 13, 2003.
- [13] C. Gutwin, C. Fedak, M. Watson, J. Dyck, and T. Bell, “Improving network efficiency in real-time groupware with general message compression,” in *conference on Computer supported cooperative work*, 2006.
- [14] J. Mogul, L. M. Masinter, R. T. Fielding, J. Gettys, P. J. Leach, and T. Berners-Lee, “Hypertext Transfer Protocol – HTTP/1.1,” RFC 2616, Mar. 2013.
- [15] “HTTP Compression,” <http://www.websiteoptimization.com/speed/tweak/compress/>.
- [16] J.-I. Gailly and M. Adler, “zlib,” <http://www.zlib.net/>.
- [17] T. Mizrahi, Y. Revah, Y. Refael Kalim, E. Kapuza, and Y. Cassuto, “The FM-Delta Project,” <https://sites.google.com/site/fmdeltacompression/>.
- [18] J. Ziv and A. Lempel, “Compression of individual sequences via variable-rate coding,” *IEEE transactions on Information Theory*, vol. 24, no. 5, pp. 530–536, 1978.
- [19] T. Mizrahi, Y. Revah, Y. Refael Kalim, E. Kapuza, and Y. Cassuto, “FM-Delta: Fault Management Packet Compression,” technical report, arXiv preprint, 2017.