Binary XML for GMTI Encoding

James McAvoy, CD, BSc, MCS

ThetaStream Consulting

jimcavoy@thetastream.com

April 2014

Introduction

- Past study on GMTI Modelling and Encoding
  - ASN.1 to model GMTI messages
    - Separate specification from encoding
  - Compared ASN.1 BER, XER and GSER encoding with current GMTI message format
  - Results
    - BER (Basic Encoding Rule (i.e. KLV)) 2X > GMTIF
    - GSER (Generic String Encoding Rule) 24X > GMTIF
    - XER (XML Encoding Rule) 33X > GMTIF

Motivation

- Separate specification from encoding for GMTI message format
- Can GMTI message format be specified and encoded using XML based standards and technology
  - Leverage a larger community base
  - Greater availability of tools and applications
  - Reduce cost and complexity
  - Increase interoperability and evolution tempo

Problem

- XML based messages consume more storage and bandwidth than binary encode messages

Study's Goal

- Can binary encoded XML improve compression performance?
  - 1. How binary encoded XML compares with ASN.1 BER (i.e. KLV) and GMTIF
  - 2. What is the reduction?

Fast Infoset (FI)

- ITU/ISO Standard
  - ITU-T Rec. X.891 (14 May 2004)
  - ISO/IEC 24824-1 (4 May 2007)
- Formal specification based on ASN.1
- Uses custom encoding rules via Encoding Control Notation (ECN)
- XML Schema not require to encode/decode FI file
- Parsing FI files is 10X faster than plain XML

EXI (Efficient XML Interchange)

- W3C recommendation (2011)
- Schema Informed to increase compactness
- Reported to have better performance than FI
- Many options to meet different requirements and contexts
  - i.e. Compactness vs. Process Efficiency

Metrics

- Compression Ratio (CR)
  - To measure compression performance of different encoding methods
  - Lower the number the better the compression

$$CR = \frac{Compressed\ Size}{Uncompressed\ Size}$$

- Space Savings ($SSAV$)
  - How much does Binary XML encoders compress an XML document

$$SSAV = 1 - \frac{Compressed\ Size}{Uncompressed\ Size}$$

$$SSAV = 1 - CR$$

- W3C EXI Working Group use both of these metrics to measure “Compactness”

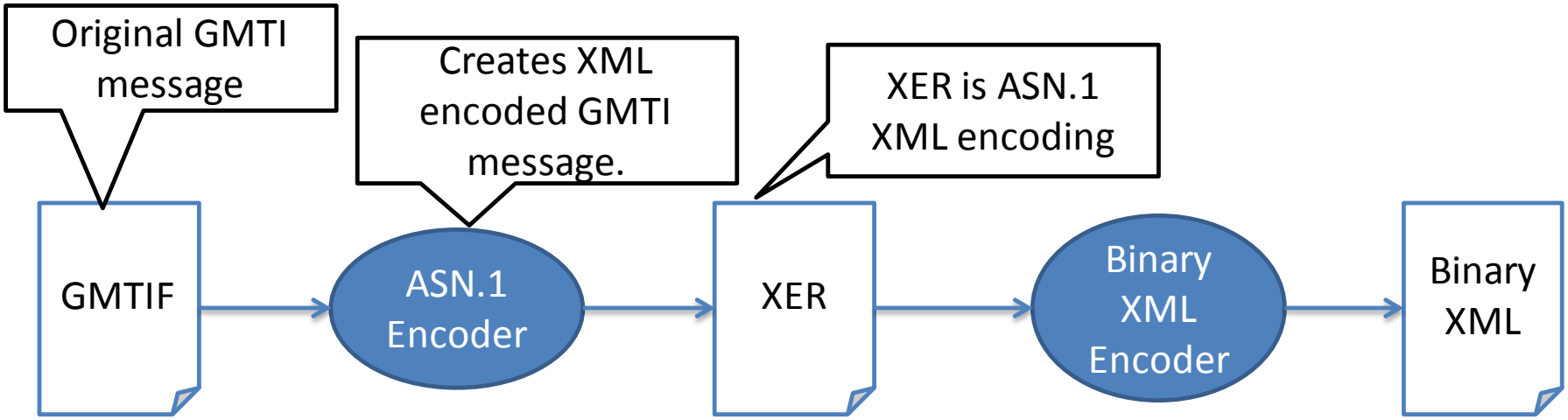
Content Density (CD)

- Percentage of an XML document that is actual data (attribute and element values), in contraposition to the portion that is structure (namespace information, tags, etc)
- Low content density means the XML document carries more XML baggage
- CD = content/file size

```
<?xml version="1.0" encoding="UTF-8"?>
<notebook date="2007-09-12">
    <note date="2007-07-23" category="EXI">
        <subject>EXI</subject>
        <body>Do not forget it!</body>
    </note>
    <note date="2007-09-12">
        <subject>shopping list</subject>
        <body>milk, honey</body>
    </note>
</notebook>
```

content

Experimental Setup



- FI Encoder
  - C# application using Liquid Fast Infoset libraries.
  - http://http://sourceforge.net/projects/fastinfoset/
- EXI Encoder
  - Java application based on the open source EXI library 'EXIficient'
  - http://sourceforge.net/projects/exiprocessor/
- ASN.1 Encoder
  - Custom C++ application using Open Source ASN.1 compiler developed by Lev Walkin
  - http://lionet.info/asn1c/blog/

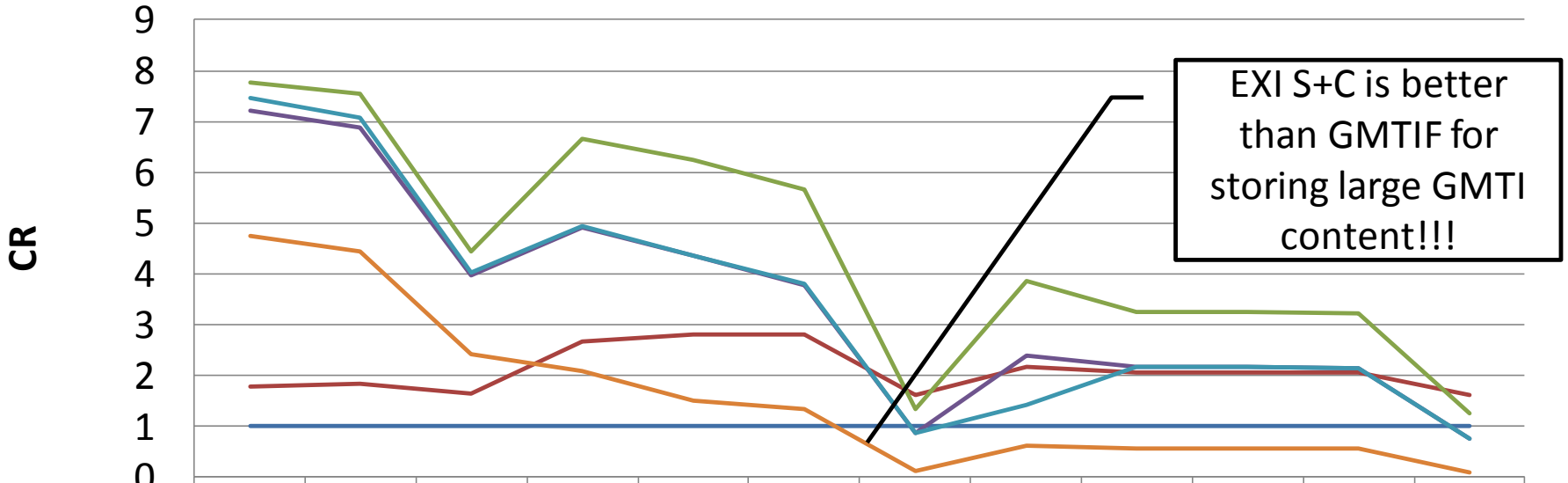
Sample Set

s_n	GMTIF (bytes)	Sensor	Mgs	Targets	Dwells	Content Density
s_1	223	PST Simulator	1		2	9.38%
s_2	249	PST Simulator	1	4	1	9.06%
s_3	443	Live PST	1		6	10.72%
s_4	2,698	Heron Live	1	277	6	9.46%
s_5	42,800	Heron Live	1	5166	24	9.46%
s_6	81,414	Heron Live	1	9814	50	9.47%
s_7	198,370	Live PST	1	380	3929	11.11%
s_8	203,493	French Horizon Helicopter	1	6776	1	4.93%
s_9	1,085,382	NATO Simulator	900			4.97%
s_{10}	1,135,957	NATO Simulator	931			4.95%
s_{11}	1,744,886	UK Sensor	1421			4.95%
s_{12}	5,283,195	Live PST	14168			11.36%

Test Cases

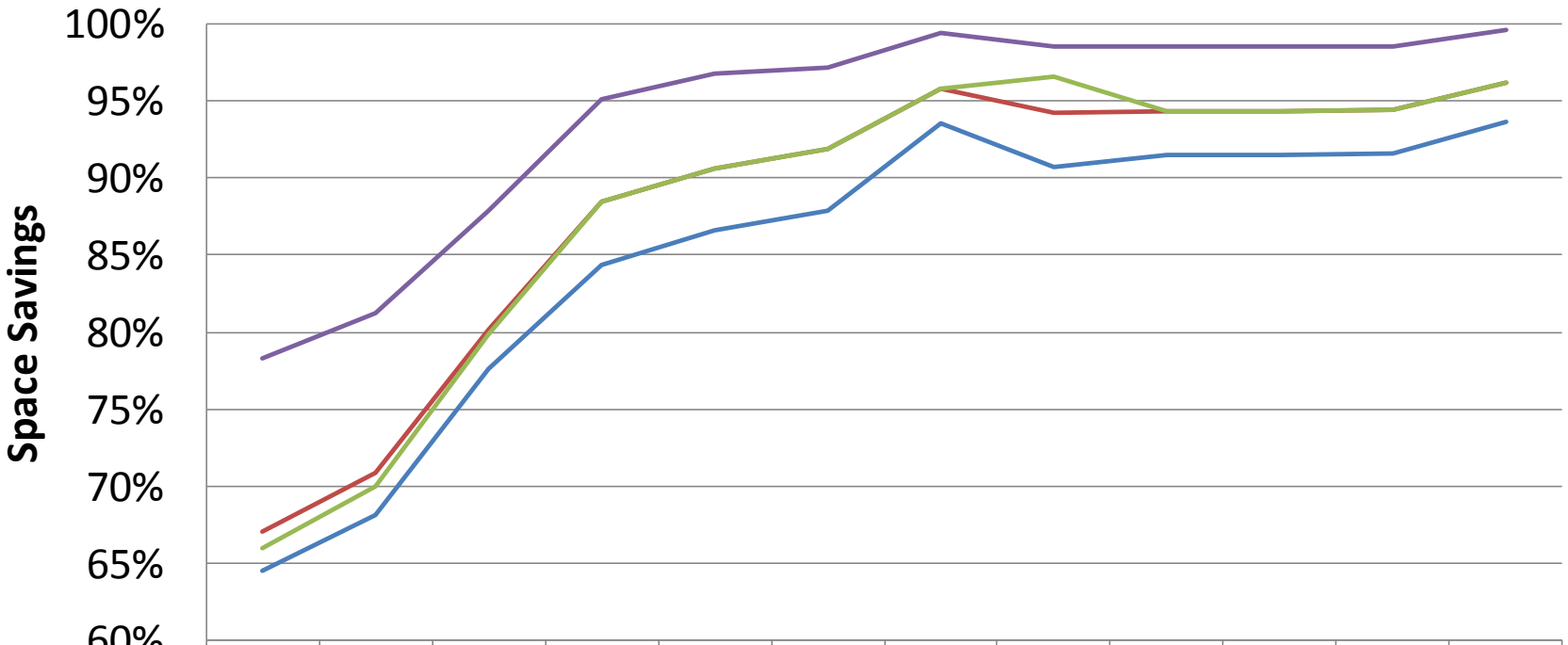
- ASN.1 Basic Encoding Rule **(BER)** i.e. KLV
- Fastinfo Set **(FI)**
- EXI Schemaless **(EXI)**
- EXI Schema Informed **(EXI S)**
- EXI Schema Informed + Compression **(EXI S+C)**
- GMTIF files used as a baseline for comparison

Compression Ratio (CR = C/GMTIF)



	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12
GMTIF	1	1	1	1	1	1	1	1	1	1	1	1
BER	1.77	1.82	1.63	2.65	2.81	2.80	1.60	2.17	2.05	2.05	2.05	1.60
FI	7.78	7.54	4.45	6.65	6.24	5.65	1.34	3.85	3.24	3.25	3.21	1.26
EXI	7.22	6.88	3.95	4.91	4.35	3.79	0.87	2.38	2.16	2.17	2.15	0.76
EXI S	7.45	7.08	4.02	4.93	4.35	3.79	0.87	1.40	2.16	2.17	2.15	0.76
EXI S+C	4.753	4.446	2.422	2.093	1.489	1.329	0.126	0.603	0.547	0.544	0.547	0.078

Compaction Efficiency



	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12
FI	64.4	68.1	77.6	84.4	86.5	87.8	93.5	90.6	91.5	91.5	91.6	93.6
EXI	67.0	70.9	80.1	88.4	90.6	91.8	95.7	94.2	94.3	94.3	94.3	96.1
EXI S	65.9	70.0	79.8	88.4	90.6	91.8	95.7	96.6	94.3	94.3	94.3	96.1
EXI S+C	78.2	81.1	87.8	95.0	96.8	97.1	99.3	98.5	98.5	98.5	98.5	99.6

Result Summary
Compression Performance

- BER 2.1X > GMTIF on average
  - Best for small messages
- FI 4.5X > GMTIF on average
- EXI schemaless 3.5X > GMTIF on average
- EXI schema informed = schemaless
- **EXI schema informed + compression 0.05X < GMTIF on average**
  - **EXI better than GMTIF to store “large” GMTI content!!!**
  - **Large means GMTIF files > 100KB**

Result Summary
Space Savings

- EXI schema informed + compression = 94.13% on average
- EXI schemaless = 88.19% on average
- FI = 85.13% on average
- In general, the larger the XML document the greater the space saving (i.e. compaction)

Conclusion

- BER works better for small messages
  - GMTIF files < 100KB
- Binary XML works better for large messages
  - GMTIF files > 100KB
- EXI is better than FI
- EXI is better than GMTIF when storing or transmitting large GMTI content