

LE-008

汎用 XML 文書符号化方式「XEUS」の性能評価 Performance evaluation of XML document Encoding with Universal Sheet

小林 垂令 松本 一則 井ノ上 直己
Arei Kobayashi Kazunori Matsumoto Naomi Inoue

(株) KDDI 研究所

1. はじめに

XMLは拡張性・可読性の長所を持つ反面、テキストデータのため冗長性が高く、携帯電話網のような伝送速度が低い環境では圧縮伝送を行う必要がある。XML文書圧縮法には、http1.1のようにプロトコルスタックでgzipを用いる手法をはじめ、XMLMill[1],xmlppmのようなXML文書のための圧縮法や、Millauのような対象XML文書のスキーマを利用した圧縮法等、種々提案されている。しかしこれらの圧縮法は、受信側でデコード処理に加えパース処理も行うため、携帯電話のような処理能力が低い端末では負荷が大きくなる。そのため圧縮によって伝送符号量を低減しても、システム全体のパフォーマンスは、XML文書圧縮を用いない場合よりも悪化してしまう。XEUS[2]は、この問題点を解消するためにXML文書のスキーマを利用した圧縮を行うだけでなく、エンコード時にパース処理も行い、受信側の処理負荷を低減させる手法である。本稿では、エンコーダの圧縮率、受信側のデコード処理負荷の観点から他手法と性能比較実験を行い、XEUSの有効性を示す。

2. XEUS(Xml document Encoding with Universal Sheet)

XEUS[2](ゼウス)とは汎用XML文書を対象とした符号化方式で、符号化対象XML文書の「論理構造」「属性値/要素値のデータ型、符号長」「要素名/属性名の符号化テーブル」等を定義した「XEUSシート」に従って符号化を行う。伝送符号量だけでなく、受信側の処理負荷も低減することにより、システム全体のパフォーマンスを向上させることが目的である。

3. 汎用XEUSエンコーダ

XEUSの性能評価のためXEUSエンコーダを開発した。本エンコーダは、XEUSシートによって対象XML文書のスキーマを動的に理解して符号化を行い、中継(ゲートウェイ/プロキシ)サーバとして動作する。

性能評価実験

実験環境は以下の通りである。

httpd : Apache2.0	言語 : C++
CPU:PentiumIII 1GHz	Memory:640M
Network:LAN	対象データ : SVG,XMark[3]

データサイズと符号化時間の関係を表に示す。SVG,XMarkで時間に多少ばらつきはあるが、100KBのXML文書を約300msecでエンコードできている。またエンコード時間は対象文書のデータサイズに比例していることが分かる。XEUSシート展開時間は、SVGが78msec,Xmarkが103msecである。ただしこの処理は、毎回のエンコード時に行われる処理ではなく、名前空間に対して1回のみ行う処理であるためこの処理がボトルネック

クになるとは考えられない。またストレスツールを用いて同時多人数アクセスのシミュレーションを行った。結果を図に示す。SVG,XMark両方とも、同時100アクセスの場合でも、パフォーマンス低下は見られない。

表1: 対象文書のデータサイズと符号化時間

size[Byte] (SVG)	エンコード 時間[msec]	size[Byte] (XMark)	エンコード 時間[msec]
10,385	37.705	5,832	18.938
20,550	72.276	8,947	25.266
30,664	100.873	27,233	67.243
40,805	134.787	39,021	93.267
50,479	167.476	47,811	119.317
102,111	323.294	118,274	276.644

表2: 同時多人数アクセスへの耐性

SVGサイズ:102,111byte, Xmarkサイズ:118274byte

同時 アクセス数 (SVG)	エンコード 時間 [msec]	同時 アクセス数 (XMark)	エンコード 時間 [msec]
1	323.294	1	276.644
10	313.08	10	266.705
100	315.462	100	266.91

他方式との圧縮率の比較

次に他の圧縮方式(gzip,XMLMill)と圧縮率の比較を行った。本実験はコンテンツに対して要素名/属性名の占める割合が高く、数値が多いSVGと要素名/属性名の占める割合が比較的low、文字列値が多いXMarkの2種類を用いた。またコンテンツサイズは2KB-1MBのものを用いている。

以下に実験結果を示す。図から分かるように、XMarkのようにマークアップの占める割合が低い文書は、どの方式も圧縮率が比較的low 30-40%程度である。図からXEUSは他方式と比較して、同等以上の性能であると言える。またSVGのようにマークアップの占める割合が高く、値も数値が多い文書はどの方式も圧縮率が高い。特にXEUSは20-25%前後の圧縮率で、他方式と比べて性能が優れていると言える。

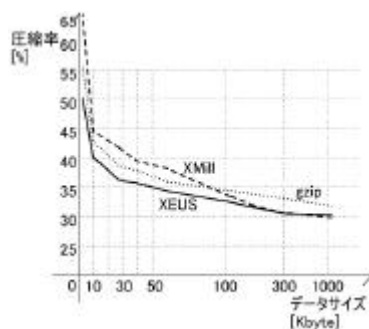


図1: 比較 (Xmark)

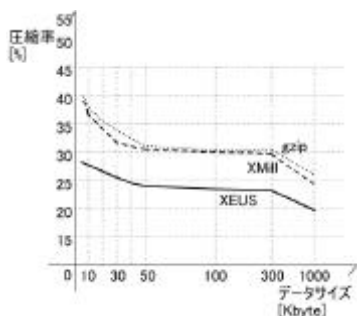


図2：比較 (SVG)

4. XEUS デコーダ

XEUS の性能を、受信側の処理負荷の観点から評価する。

4.1. Java(CLDC)版汎用 XEUS デコーダ

携帯電話で動作可能なプラットフォームである Java(J2ME/CLDC)上のアプリケーションとして本デコーダを開発した。本デコーダは XEUS シートを動的に理解して、XML 文書をデコードする。API は XMLPull API をベースに一部拡張している。

性能評価実験

XEUS デコーダの性能を評価するために、他方式と比較実験を行った。他の圧縮方式のデコーダは、現状携帯電話では動作不可能である。しかし他方式はデコードし XML 文書を復号した後、パース処理を必要とする。そこで携帯電話で高速動作可能な XML パーサと XEUS デコーダを比較することにする[10]。

実験対象パーサ：

MXP1[11],kXML2[12],MinML[13],NanoLite[14],
TinyXML[15],XEUS

サンプル：SVG(10791byte,XEUS は 3310byte)

実験対象端末：A3012CA,A5303H,A5011HMC(KDDI)
P504is(DoCoMo)

測定内容：要素値/属性値の文字、数値の出現数をカウント
(5回平均)

以下に実験結果を示す。表3に各パーサのオブジェクトサイズを示す。どのパーサも20KB以内であり、携帯電話上で動作させるのに十分軽量といえる。XEUS デコーダは他のパーサと比較しても、ほぼ同等のサイズといえる。表4に各パーサのデコーダ初期化時間を示す。XEUS は XEUS シートの展開処理が含まれるため、他のパーサよりも劣る。だがエンコーダ同様、この処理はデコード毎に必要な処理ではなく最初の1回のみ必要となる処理でありボトルネックになるとは考えられない。表5にデコード時間を示す。対象端末によってパフォーマンスのばらつきはあるが、XEUS は他方式と比較して圧倒的に高性能と言える。図3に性能比を示す。最速の XML パーサと言われる MXP1 と比較して、9倍強の性能が得られている。

つまり、XEUS を用いることによって、エンコーダ側では、他の圧縮方式と同等以上の圧縮率を実現し、且つ受信側のデコード処理においては、他のパーサと比べて大幅なパフォーマンス向上が見られ、XEUS の有効性が示せていると言える。

表3：デコーダオブジェクトサイズ(jar)比較

XEUS	MXP1	KXML2	MinML	NanoLite	TinyXML
18.0KB	17.8KB	11.3KB	13.6KB	6.8KB	9.8KB

表4：パーサ起動時間[msec],

上から A3012CA,A5303H,P504is

XEUS	MXP1	KXML2	MinML	NanoLite	TinyXML
4980	2145	735	1145	1225	1210
350	250	75	125	150	125
575	341	108	168	153	143

表5：デコード時間[msec],

上から A3012CA,A5303H,P504is

XEUS	MXP1	KXML2	MinML	NanoLite	TinyXML
2150	16775	23146	21669	44927	25818
85	950	1260	1380	2580	1620
325	2314	3088	2928	5588	3445

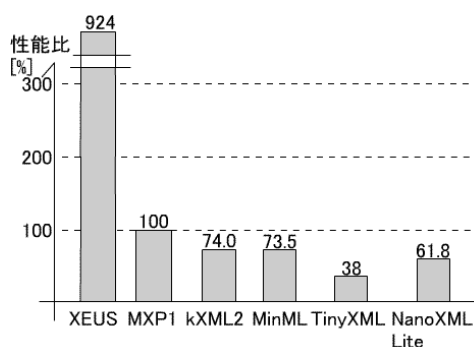


図3：デコード速度相対比較(4機種平均値)

5. まとめ

本稿では、汎用 XML 文書符号化方式「XEUS」の性能評価を行った。エンコーダは他方式と比較して同等以上の性能が得られ、デコーダは他の高速 XML パーサと比較して10倍程度の性能が得られている。また XEUS はエンコーダ/デコーダ共に汎用 XML 文書を対象とし、エンコーダは中継サーバとして動作し、受信側は XMLPull ベースの API としてデコードされる。そのため利用者はコンテンツサーバに XML 文書をアップロードするだけでよく、XEUS を利用することを意識することなくシステム全体のパフォーマンスを大幅に向上させることが可能である。

なおこの研究活動は、TAO (通信放送機構) の国際標準実現型研究開発制度による受託研究「携帯端末向けグラフィックスフォーマットの研究開発」の一環として行っている。

参考文献

- [1]H.Liefke and D.Suciu. Xmill:an efficient compressor for XML data.In Proceedings of the 2000 ACM SIGMOD,pp.153-164,2000
- [2]小林亜令他,“XML 文書汎用符号化方式「XEUS」”,信学技報 DE2001-9
- [3]Xmark; <http://monetdb.cwi.nl/xml/>
- [4]モバイル・組み込み環境での XML ~ 軽量 XML パーサとバイナリ XML;
- [5]MXP1;
<http://www.extreme.indiana.edu/xgws/xsoap/xpp/mxp1/>
- [6]kXML2; <http://www.kxml.org/>
- [7]MinML; <http://www.wilson.co.uk/xml/minml.htm>
- [8]NanoLite; <http://nanoxml.n3.net>
- [9]TinyXML; <http://www.grinninglizard.com/tinyxml/>