

# Will Binary XML Speed Network Traffic?

David Geer

**W**ith its ability to enable data interoperability between applications on different platforms, XML has become integral to many critical enterprise technologies. For example, XML enhances e-commerce, communication between businesses, and companies' internal integration of data from multiple sources, noted analyst Randy Heffner with Forrester Research, a market-analysis firm.

XML use is thus increasing rapidly. Analyst Ron Schmelzer with market-research firm ZapThink predicted XML will rise from 3 percent of global network traffic in 2003 to 24 percent by 2006, as Figure 1 shows, and to at least 40 percent by 2008.

However, XML's growing implementation raises a key concern: Because it provides considerable metadata about each element of a document's content, XML files can include a great deal of data. They can thus be inefficient to process and can burden a company's network, processor, and storage infrastructures, explained IBM Distinguished Engineer Jerry Cuomo.

"XML is extremely wasteful in how much space it needs to use for the amount of true data that it is sending," said Jeff Lamb, chief technology officer of Leader Technologies, which uses XML in teleconferencing applications.

Nonetheless, said Heffner, "XML



adds intelligence on top of data in motion to make that data more manageable across vast technical boundaries. XML is so important that the industry is looking for ways to make its data load more manageable."

Proponents say a thinner binary XML will help. XML currently uses only a plain-text format.

The World Wide Web Consortium (W3C), which oversees and manages XML's development as a standard, and Sun Microsystems are working on binary XML formats.

Some industry observers have expressed concern that multiple formats or proprietary implementations of binary XML could lead to incompatible versions, which would reduce the openness that makes the technology valuable.

## XML'S PROBLEMS

The W3C started work on XML in 1996 as a way to enable data interoperability over the Internet. The con-

sortium approved the standard's first version in 1998.

A key factor driving the standard's development was increased Internet and network usage requiring companies on different platforms to be able to communicate. Many businesses also wanted to make legacy data available to new Web-based applications.

## How XML works

XML is a markup metalanguage that can define a set of languages for use with structured data in online documents. Any organization can develop its own XML-based language with its own set of markup tags. For example, a group of retailers could agree to use the same set of tags for categories of data—such as "customer name" or "price per unit"—on a product order form.

A typical XML file also includes information about a document unrelated to content, such as the encryption used and the programs that must be executed as a result of or as part of processing the file.

The XML document type definition describes a document's metadata rules—identifying markups, stating which elements can appear, and noting how they can be structured—to the applications that must work with it. XML documents are written and stored as text, and documents are read via either text editors or XML parsers.

By enabling cross-platform communications, XML eliminates the need to write multiple versions of documents or to use costly and complex middleware. However, the files contain considerably more information than just the content they are communicating.

XML is the basis for important technologies such as Web services and important standards such as the Simple Object Access Protocol, a way for a program running in one operating system to communicate with a program running in another by using HTTP and XML as the information-exchange mechanisms.

## Performance hit

Standard XML is bigger and, more importantly, less efficient to process than a binary version would be, thereby slowing the performance of databases and other systems that handle XML documents.

For example, IBM's Cuomo said, "You have information in a database that is SQL compatible. You get result sets out of the database and, in our case, you put it into Java Object format, convert it to XML and then to HTML before you send it to the end user." The process must be reversed when the user sends back material, Cuomo explained. "This consumes MIPS," he noted.

Using XML also causes Web services, which are becoming increasingly popular, to generate considerable traffic.

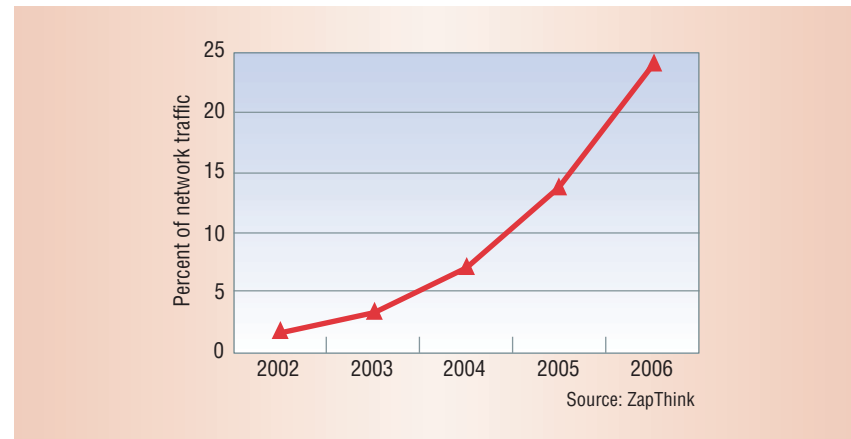
In addition, said Glenn Reid, CEO of Five Across, a Web development firm that works with XML, "You can't really start to process an XML file until you've received the entire thing." Because of the syntax, systems must read to the end of an XML document before determining the data structure. On the other hand, systems can process some file types as they receive them.

## SOLVING THE PROBLEM

One approach to solving XML-related problems is using appliances dedicated to making the documents more manageable. These products—sold by vendors such as DataPower, F5 Networks, Intel, and Sarvega—can preprocess an XML document by applying XSL (Extensible Stylesheet Language) transformations to reorganize its structure so that the host system doesn't have to do all the work.

The appliances can also compress XML files or streamline them by eliminating material—such as spaces or tabs—present only to keep the material in textual, human-readable form.

However, noted Leader Technologies' Lamb, "These appliances are expensive." It would be preferable to make XML itself easier to work with, he said, to reduce costs and enable



**Figure 1. XML usage, as represented by XML's percentage of all network traffic, has grown rapidly during the past few years and is predicted to continue doing so.**

more complex and rich XML-based applications.

Thus, the leading proposal to alleviate XML's performance hit is binary XML, a format that optimizes documents for faster handling.

## W3C specifications

The W3C has formed the Binary Characterization Working Group ([www.w3.org/XML/Binary/](http://www.w3.org/XML/Binary/)) to study binary XML. The working group has issued three recommendations—backed by software vendors such as BEA Systems, IBM, and Microsoft—designed to make handling XML files more efficient.

"All three of these specifications have reached the final stage of the W3C recommendation track process," said Yves Lafon, a W3C XML protocol activity leader who also participates in the working group.

**XML Binary Optimized Packaging.** XOP makes XML files smaller by extracting binary parts such as images, sending them as a separate package with the document, and providing a uniform resource identifier as a link that recipient systems can use to access the extracted material, explained Lafon.

Currently, images and other binary data in a standard XML document must be encoded in base64 to be processed with the rest of the file.

Base64 encodes binary data as ASCII text. The process divides three bytes of the original data into four bytes of ASCII text, making the file one-third bigger.

Using XOP eliminates the need for larger files, as well as the time and effort necessary to conduct base64 conversions.

**Message Transmission Optimization Mechanism.** The W3C has incorporated XOP's method for representing binary data into the MTOM communications protocol. In essence, MTOM implements XOP for SOAP messages. MTOM uses MIME (multipurpose Internet mail extensions) multipart to package the message, after XOP processing, with the extracted binary parts, Lafon explained.

**Resource Representation SOAP Header Block.** RRS HB provides a way for an application receiving an XML message—from which binary parts have been extracted via XOP and packaged with the main file via MTOM—to retrieve the binary parts. In the message's SOAP header, RRS HB references where the binary parts are and how the application receiving the message should access them.

## Sun's Fast Infoset Project

Sun has started the Fast Infoset Project (<https://fi.dev.java.net>), an open

## Industry Trends

source implementation of the International Organization for Standardization's and the International Telecommunication Union's Fast Infoset Standard for Binary XML, used for turning standard XML into binary XML (<http://asn1.elibel.tm.fr/xml/finf.htm>).

According to Sun Distinguished Engineer Eduardo Pelegri-Llopart, the technology encodes an XML document's information set (infoset) as a binary stream and then substitutes number codes for all of the metatags, thereby reducing a file's size. Included in the stream is a table that defines which metatag each number code stands for.

The overall document is generally smaller than a comparable textual XML file, and recipient systems can parse and serialize it more quickly.

In early tests, Sun says, XML applications perform two or three times faster when using software based on its technology.

### CONCERNS OVER INCOMPATIBILITY

According to Leader Technologies' Lamb, XML is currently standardized and interoperable largely because it uses a plain-text format. Moving to binary XML without maintaining standardization, he said, would cost much of the interoperability for which XML was created.

Five Across' Reid expressed concern that the binary XML efforts might lead to incompatible versions of the technology. In addition, he said, different companies could create incompatible binary formats, including some for specific applications such as mobile phones, which have severe processing and memory constraints.

Some industry observers say that future increases in network and processor performance could improve systems' ability to handle standard XML and thereby eliminate the need for binary XML.

However, stated Sun's Pelegri-Llopart, binary XML would offer a badly needed solution sooner than

waiting for adequate network and processor improvements to occur.

And, according to IBM's Cuomo, faster networking won't work or isn't available in many situations, such as in small towns or developing countries in which broadband networking isn't readily accessible or affordable.

**B**ecause binary XML is suitable when network efficiency is important, ZapThink's Schmelzer said, users might decide to work with it only for high-volume applications that demand the best performance, like those in financial transactions, telecommunications, and multimedia.

Even if a single approach is standardized, there will still be applications and systems that can't work with binary XML. In some cases, standard textual XML will be preferable because it is easy to code by hand and is universally understandable.

There is some concern about how well binary XML would work with Web services even if it is standardized. Many Web services models allow intermediate entities—such as an XML security gateway or a policy-enforcement tool—to act on a message during transmission. The overhead involved if intermediaries must code and decode messages could reduce or eliminate binary XML's efficiency.

Nonetheless, Cuomo said, the urgent need for a faster XML that would reduce the burden on CPUs, memory, and the network infrastructure will help ensure its future success. ■

*David Geer is a freelance technology writer based in Ashtabula, Ohio. Contact him at [david@geercom.com](mailto:david@geercom.com).*

Editor: Lee Garber, *Computer*,  
[l.garber@computer.org](mailto:l.garber@computer.org)



## REACH HIGHER

Advancing in the IEEE Computer Society can elevate your standing in the profession.

Application to Senior-grade membership recognizes

- ✓ ten years or more of professional expertise

Nomination to Fellow-grade membership recognizes

- ✓ exemplary accomplishments in computer engineering

GIVE YOUR CAREER A BOOST

UPGRADE YOUR MEMBERSHIP

[www.computer.org/join/grades.htm](http://www.computer.org/join/grades.htm)