The Potential of Extensible Markup Language

David A. Hayes
Space and Missile Defense Technology Center

The Department of Defense should leverage commercial best practices as it upgrades and extends its Battle Management, Command, Control, Communications, and Computers/Intelligence (BMC4I) architecture. The family of technologies associated with the Extensible Markup Language (XML) is the modern foundation of some of these best practices. This article describes XML and its potential application to BMC4I.

The origin of computer markup languages such as Extensible Markup Language (XML) is in the publishing processes that we have employed for centuries. Prior to the invention of computers, markup was simply the editorial comments that a copy editor made on a report or article often in the margins of a paper. The markup added information to the paper's content by specifying format or other meaning. In modern computer science, any information or markup added to a document that specifies the meaning of its content is known as metadata. In turn, any language used to markup a document is known as a metalanguage.

The first computer metalanguage, Standard Generalized Markup Language (SGML), was standardized in 1986, a long time ago in computer-years. SGML provided a standard means to separate content and format in documents of all kinds. Such separation is very useful in applications where information is presented on a variety of media for which no single format is adequate. Early adopters of SGML include the IRS, Patent Trademark Office, SEC EDGAR Database, and Army SGML Registry and Library (ASRL) for its technical manuals.

Next came a series of standards for the HyperText Markup Language (HTML) about a decade ago. Most readers will recognize HTML as the language of the Internet's very popular World Wide Web. The HTML is a simple metalanguage with a rigid syntax designed for the presentation of a common class of office or technical report, with headings, paragraphs, lists, illustrations, etc. In addition HTML has limited support for multimedia. The success of these metalanguages set the stage for the development of the next generation, XML, which is an evolution of SGML without the limitations of HTML.

Extensible Markup Language

The World Wide Web Consortium (W3C) released its proposal for XML in 1996, and approved the standard in 1998. From the start, XML has received an extraordinary amount of attention from public and private industry. Much of this attention did not stem from XML's immediate contributions but from its potential for future contributions.

XML is a standard markup metalanguage for describing, archiving, and communicating digital information. XML is a method for putting structured data in a text file. Therefore, it is readable by both man and machine.

An XML document looks like HTML, but isn't HTML. Both XML and HTML use tags and attributes. Tags are words bracketed by the delimiters, "<" and ">". Tags may contain attributes of the form "name=value." The rigid syntax of HTML specifies what each tag and attribute means and often how the text between them will look in a browser display. XML uses tags to delimit pieces of data, and leaves the interpretation of the data to the application that reads it. As a result, XML-formatted documents can be readily displayed or used by a variety of applications, not just a limited number of Web browsers.

To assist in an application's interpretation of XML, the Defense Information Systems Agency (DISA) provides namespace registry services for XML metadata for BMC4I domains such as Ground Operations, Aerospace Operations, and Geospatial Imagery etc.

The registry diides.ncr.disa.mil/xmlreg/index.cfm enables the consistent interpretation and use of XML, both vertically within a system and horizontally across systems. (The namespace registry is another member of the family of XML technologies.)

Like HTML, XML can use procedural languages to implement applications that further define interfaces, manage data, and permit greater interoperability. Java and JavaScript are two common network-friendly procedural languages used with XML.

Unlike HTML, XML does not contain format information. If XML information is presented, then a style language such as the Extensible Style Language (XSL) defines the presentation format. (The XSL is another member of the family of XML technologies.) This separation of content and format means that an XML-based BMC4I system will not break with each application or new presentation media.

The tags in an XML document delimit and define data. The XML is flexible in that tags can be created to communicate any digital data including text and multimedia. The XML is a metalanguage because its tags may contain metadata, or extra information about message data, with no predefined syntax.

Table 1 is an example of an XML data island. You can observe many qualities of the language in this example. Data elements consist of a start tag (bracketed by "<" and ">"), some data or other data elements, and an end tag (bracketed by "</" and ">"). For every start tag there is an identically named end tag. Other data elements may be nested in as many levels as desired within a data element. Nested data elements must be completely nested within its parent's tags. Metadata may be specified in the start tag. Metadata attributes are completely enclosed in double-quote marks. These qualities are why XML is known as a well-formed language.

Here is an example that illustrates the benefit of metadata. The following United States Message Text Format (USMTF) line simply presents a remark: RMKS/179 248//. The following XML line says the same thing — and more: <location_target>179 268</location_target>. The XML line makes it clear to an extraction routine or human reader what the data between the tags means. The number is not a Canadian zip code or

```
<XML ID="150300 Weather">
<weather-report>
    <date>25 March 2000</date>
    <time>08:00</time>
    <area ZIP="35807">
        <city>Huntsville</city>
        <state>Alabama</state>
        <region>South Central</region>
        <country>USA</country>
    </area>
    <measurements>
        <skies>partly cloudy</skies>

        <temperature>36</temperature>

        <wind>
            <direction>SW</direction>
            <windspeed>6</windspeed>
        </wind>
        <humidity>87</humidity>
        <visibility>10</visibility>
    </measurements>
</weather-report>
    </XML>
```

Table 1. *Example of an XML Data Island*

some other interpretation of the USMTF line. The XML line clearly describes the number as a target coordinate. Thus, metadata gives meaning to XML data.

There is a price for this extra information. The XML metadata in the form of data tags creates a transmission overhead. The metadata imposes an added burden on network capacity and processing to parse the message. Data compression, local computation and manipulation of data, intelligent communication of knowledge as opposed to raw data, and granular updates are mechanisms that mitigate the network load penalty.

An uncompressed XML message is significantly larger than an MTF version of the same message. This is not a problem. The XML can be compressed for more efficient network transmission. Such was the case in the United States-led multinational global command, control, communications, and intelligence 1999 Joint Warfighter Interoperability Demonstration. In that demonstration XML was compressed via a smart compression utility (XMill by AT&T). Dr. Robert Miller of MITRE reports that XML-based Air Tasking Orders (ATO) were actually smaller than the original ATOs in compressed MTF format as indicated in Table 2. Compressed XML and MTF files were found equivalent in size in the 2000 JWID. Therefore, we can conclude that larger XML messages do not stress network capacity more than current formats.

XML does not ensure data compatibility with a given application. Therefore, interoperability standards are required just like previous data exchange methods. Developers of any Defense Information Infrastructure Common Operating Environment such as BMC4I systems must express XML with a standard lexicon and grammar. The previously described XML namespace registry can provide the necessary standardization of XML tags. It is a vital component in the implementation of any metalanguage.

Table 2. *ATO Compression*

File Type	Size (bytes)
MTF-ATO	300K
MTF-ATO (compressed with Pkzip)	72K
XML-ATO	2.2M
XML-ATO (compressed with XMill)	46K

XML, like any text, can be easily encrypted for secure communications. Significant security concerns can be addressed by using the Secure Internet Protocol Router Network , or commercial secure sockets layer technology.

Commercial concerns are rapidly developing the necessary tools and applications to realize the potential of XML in electronic commerce. The capabilities of these commercial tools and applications will drive down the cost of XML development allowing more affordable military-specific implementations.

Many XML applications are inexpensive, powerful, and rapidly evolving. The XML development is driven by the dynamics of the commercial marketplace. As such, XML-based products enjoy an efficiency of effort normally associated with a competitive commercial market. The development of XML-based BMC4I systems can leverage these commercial efficiencies. XML is license free, platform independent, and well supported.

XML Potential for BMC4I

Future XML-based BMC4I systems will likely focus on two areas: The first deals with acquiring information from disparate (often *legacy*) sources used by military systems. The second focuses on distributing dynamic content to users. The Internet community has adopted two words to describe these tasks: aggregation and syndication. The XML's information-exchange prowess applies to both of these tasks.

Aggregation is the process of collecting, organizing, and integrating data from disparate sources. One potential application for XML BMC4I aggregation is in the creation of a Global Family of Interoperable Integrated Air Pictures (GFIIAP) used for Air and Missile Defense. In the case of GFIIAP, legacy databases of satellite orbits can be aggregated with in-theater sensor data to form a more complete and accurate picture. Unexpected deviations of the data stream from any source can be processed based on embedded metadata in XML tags. The metadata can also help clarify issues of time latency, track correlation, and fusion.

An XML aggregation application could be developed to organize references to metadata rather than the actual data itself. The application would collect and build a repository of metadata links, which are in turn archived in XML format. Intelligent agents working with this repository would expedite searches, access, and distribute data as needed. Another set of artificial intelligence agents would gather input from a variety of sources, including databases, XML, Tactical Digital Information Link (TADIL), and Message Text Format (MTF) documents.

Syndication is the process of disseminating data. On a BMC4I network, syndication is often seen in the form of information that participants get from third party information providers. For example, a tactical operations center (TOC) might request information from a weather forecast database to pass on (perhaps aggregated with other information and appro-

priately tagged) to a weapon system. This information is syndicated from this weather database. In this scenario, the weather database is the syndicator or third party, the TOC is the subscriber, and the weapon system is the user.

A BMC4I syndication tool set could be developed to define, extract, and publish syndicated information from large frequently updated databases. The tool set would make information available to subscribers through XML used to describe the structure and content of syndicated information. Information would be distributed by using two complementary components. The first component lets syndicators define the information that the subscriber wishes to receive, i.e. all satellite tracks over a theater of operations in the next hour. The second component reads this definition and performs the actual information extraction and distribution to the subscriber TOC. The TOC could aggregate the information locally and provide it to users.

In addition, XML can be used as a tactical language "Rosetta Stone." XML can translate or encapsulate common military communication languages in use today such as TADIL A/B/J, MTF, and other languages.

Conclusion

The Joint Technical Architecture (JTA) mandates the XML 1.0. W3C Recommendation, 10 February 1998 (Reference: REC-xml-19980210, www.w3.org/TR/1998/REC-xml-19980210) for domain- and application-specific markup languages defined through tagged data applications. The JTA states, "This allows new capabilities to be defined and delivered dynamically."

XML conveys complex information while retaining enough flexibility to accommodate future modifications to the message content in ways that are unforeseen today. Therefore, it is ideal for supporting rapid prototyping, evolutionary, and spiral development of evolving BMC4I systems.

The most important benefit of XML as it applies to BMC4I is that it advances and develops good information management practices that are responsive to modern computer technology and standards. XML is not always the best solution, but it is always worth considering.◆

Notes

1. A "two-line" satellite element set actually has three lines. Line zero is a twenty-four-character name consistent with the name in the North American Air Defense Satellite Catalog celestrak.com/NORAD/documentation/tle-fmt.html

2. Mil-Std 6016A is the TADILJ/Link 16 Message Standard

Additional On-Line Readings

1. diides.ncr.disa.mil/xmlreg/index.cfm
2. diides.ncr.disa.mil/shade/index.cfm
3. www.computer.org/proceedings/meta/1999/papers/21/rdaniels.html

Acknowledgements

The author would like to thank Leo Kimminau, and Dr. Robert Miller for their contributions and assistance.

About the Author

David A. Hayes is cheif of the Computer Technologies Division of the Advanced Technology Directorate of the Space and Missile Defense Technology Center in Huntsville, Ala., where he manages small business and innovative research developments. He is a subject matter expert for BMDO Global Defense BMC4I and the Army's Space Control BMC4I programs. He holds a bachelor's degree in electronics engineering from the University of Tennessee, in Knoxville, and a master's degree in management from the Florida Institute of Technology.

David A. Hayes
U.S. Army Space and Missile Defense Command
Commander U.S. Army Space and Missile Defense Command
Attn: CSSD-TC-TD-AS, David Hayes
P.O. Box 1500
Huntsville, AL 35807-3801
Phone: 256-955-3340
Fax: 256-955-1432
E-mail: David.Hayes@smdc.army.mil
http://users.liveonthenet.com/~dhayes/

Web Sites

Common Vulnerabilities and Exposures

www.cve.mitre.org

The Common Vulnerabilities and Exposures (CVE) site is a list of standardized names for vulnerabilities and other information security exposures. CVE is a dictionary, not a database, which makes it easier to share data across separate vulnerability databases and security tools. CVE content is a collaboration of the CVE editorial board including representatives from security-related organizations, academic institutions, government, and other security experts.

Organization for the Advancement of Structured Information Standards

www.oasis-open.org

OASIS, the Organization for the Advancement of Structured Information Standards, is a non-profit, international consortium that creates interoperable industry specifications based on public standards such as Extensible Markup Language (XML) and the Standard Generalized Markup Language (SGML), and others related to structured information processing. OASIS members include organizations and individuals who provide, use, and specialize in implementing the technologies that make these standards work in practice.

TechNewsWorld

www.technewsworld.com

TechNewsWorld.com is a real-time news service, updated 24-hours a day, in multiple languages. It provides a central resource for technology news links from around the world. In addition to headlines, TechNewsWorld.com also analyzes each news article providing small excerpts and a hyperlink to the original news article, as well as links to related stories from many other sources.