Wei Zhang

Contact Information

Department of Computational Science   Phone: (850)644-8531(office)
Florida State University   Email: wzhang@cs.fsu.edu
Tallahassee, FL 32306   http://ww2.cs.fsu.edu/∼wzhang

Objective

My current career objective is to obtain an internship position in a research and development team in technologies including efficient XML processing, high-performance computing, Web services, grid computing or related areas.

Research Interest

- Efficient XML Processing Technologies
- High-performance Computing
- Service Oriented Architectures and Web Services

Education

Ph.D. candidate in Computer Science, August 2004 - present
Florida State University, Tallahassee, FL, USA
Adviser: Dr. Robert van Engelen
M.S. in Computer Science, August 2004
Baylor University, Waco, TX, USA
Adviser: Dr. Paul Grabow
B.E. in Engineering Telecommunication, July 1991
Chengdu University of Information Technology, Chengdu, P. R. China

Honors and Awards

- Outstanding Teaching Assistant Award, Florida State University (2008).
- Best Student Paper Award, IEEE International Conference on Web Services (2006).
- Outstanding Student Fellowship, Chengdu University of Information technologies (1988-1990).

Full-time Internship (Part of Doctoral Research)

Avaya Research Lab, January 2007 - May 2007
Basking Ridge, New Jersey

- Integrated WS-Security into Web Services Initiation Protocol (WIP)
- Proposed and implemented Role-based Presence for WIP
- Integrated Audio/Video components into WIP

Employment History

Chinese Science & Technology Netwoek at Chengdu, May 1998 - November 2001
Chengdu, P.R. China
- System Architect
- Presided project *CSTNet Chengdu Node Update*
- Proposed solutions of CAS Chengdu Branch Intelligent Broadband Network

Chinese Science & Technology Network at Chengdu, September 1995 - May 1998
Beijing, P.R. China
- Software Engineer
- Developed CSTNet Billing System
- Team member of project *CADS Management Information System*
- Participated in project *CSTNet update*

Tianjin Bureau of Meteorology, September 1991 - May 1998
Department of Computer Network and Technical Equipment
Tianjin, P.R. China
- Assistant engineer
- Developed *Meteorological Data Telecommunication System*
- Team member of project *Radar Model-711 Digitalization*

Research Experience

Research Assistant, Florida State University, Tallahassee, FL. Faculty adviser: Dr. Robert van Engelen, Spring 2005 - present.
- Research on efficient XPath Querying processing
- Research on permutation grammars
- Research on high-performance Table-Driven XML parsing methodology - TDX
- Developed a parser construction toolkit to automatically construct parsers in C code from WSDLs and XML schemas
- Implemented MorphBank LSID authority using LSID-java toolkit
- Investigated and proposed security mechanisms for MorphBank LSID Web services
- Developed Java codes to test the proposed security mechanisms for MorphBank LSID Web services

Research Assistant, Baylor University, Waco, TX. Faculty adviser: Dr. Randall Jean, August 2003 - May 2004.
- Developed FORTRAN codes to calibrate fundamental engineering parameters for waveguide sensors using NIST ODRPACK toolkit

Research Assistant, Baylor University, Waco, TX. Faculty adviser: Dr. Greg Speegle, June 2004 - July 2004.
- Survey on XML query tools

Research Assistant, Baylor University, Waco, TX. Faculty adviser: Dr. Greg Speegle, August 2002 - December 2004.

- Webmaster of computer science department graduate program website

Teaching Experience

Graduate Teaching Assistant, Florida State University, Department of Computer Science, August 2004 - December 2004.

- Instructed recitation sections for *Introduction to Programming with the C Language*

Course Instructor, Baylor University, Department of Modern Foreign Languages, January 2003 - May 2003.

- Instructed course *Elementary Chinese*
- Conducted lectures
- Prepared and graded assignments, quizzes, and exams

Technical Skills

Programming languages and tools: C/C++, Java, SQL, ASP, JSP,PHP, log4J, Junit, ant, cvs, svn, gSOAP, Globus

Web Services & Grid:

- XML, XML schemas, SOAP, WSDL, Web services, XPath, XSLT, XHTML, CSS, Portal.
- WS-Notification, WS-Eventing, WSRF, OGSA, OGSI, GT3, GT4.
- GSI, WS-Security, XML-Signature, XML-Encryption, XKMS, SAML, SSL, PKI, Kerberes.

Databases: MySQL, Oracle

Publications

1. W. Zhang and R. van Engelen. An Adaptive XML Parser for Developing High-Performance Web Services. In *Proceedings of the IEEE Fourth International Conference on e-Science Conference workshop on Advances in High-Performance E-Science Middleware and Applications (AHEMA)*, pages 672-679, IN, USA, December 7-12, 2008.

   - **Abstract:** This paper presents an adaptive XML parser that is based on table-driven XML (TDX) parsing technology. This technique can be used for developing extensible high-performance Web services for large complex systems that typically require extensible schemas. The parser integrates scanning, parsing, and validation into a single-pass without backtracking by utilizing compact tabular representations of schemas and a push-down automaton (PDA) at runtime. The tabular forms are constructed from a set of schemas or WSDL descriptions through the use of permutation grammar. The engine is implemented as a PDA-based, table-driven driver, as a result, it is independent of XML schemas. When XML schemas are updated or extended, the tabular forms can be regenerated and populated to the generic engine without requirement of redeployment of the

parser. This adaptive approach balances the need for performance against the requirements of reconstruction and redeployment of the Web services. Our experiments show the adaptive parser usually demonstrates performance of 5 times faster than traditional validating parsers and performance drop within 20% of the fastest fully compiled traditional validating parsers.

2. W. Zhang and R. van Engelen. "High-Performance XML Parsing and Validation with Permutation Phrase Grammar Parsers." In *Proceedings of IEEE International Conference on Web Services (ICWS 2008)*, pages 286-294, Beijing, China, September 23-26, 2008.

   - **Abstract:** XML has become the de facto standard for exchanging structured information on the Web and for delivering rich content to users and business information processing systems. The extensibility,flexibility, expressiveness, and platform-neutrality of XML delivers key advantages for interoperability. Thus, industrial-strength interoperable XML-based Web services standards are widely adapted by organizations to deploy service-oriented architectures that comprise tens to thousands of loosely-coupled services. However, the interoperability of XML Web services comes at the price of reduced efficiency of message composition, transfer, and parsing compared to simple binary protocols. Parsing and validation of XML against a schema is expensive. This paper presents a high-performance XML parsing and validation technique that is time and space optimal. A schema-specific parsing method is developed that uses a two-stack push-down automaton for single-pass parsing and validation without backtracking. The schemas and validity constraints are packed in a compact parsing table derived from a permutation phrase grammar. This approach reduces both the space and time requirements of XML parsing and validation. Other schema-specific validating XML parsing methods trade efficiency for space (larger code and/or data size) or trade space for efficiency (backtracking). Performance results show that the method is significantly faster than traditional validating and non-validating XML parsers.

3. R. van Engelen, and W. Zhang. An Overview and Evaluation of Web Services Security Performance Optimizations. In *Proceedings of IEEE International Conference on Web Services (ICWS 2008)*, pages 137-144, Beijing, China, September 23-26, 2008.

   - **Abstract:** WS-Security is an essential component of the Web services protocol stack. The WS-Security provides end-to-end security properties (integrity, confidentiality, and authentication) through XML Encryption and XML Signature open W3C standards. End-to-end security properties assures the participation of non-secure transport intermediaries in message exchanges, a key advantage in Web systems. However, compared to point-to-point messaging with TLS, WS-Security has a significant performance penalty. In this paper, we survey several techniques for WS-Security signature performance optimization for message integrity and compare experimental results to determine the overall combined performance impact. We also compare the performance to TLS for point-to-point message integrity and confidentiality.

4. R. van Engelen, and W. Zhang. Identifying Opportunities for Web Services Security Performance Optimizations. In *Proceedings of IEEE Congress on Services - Part I,*

Hawaii, USA, 2008.

- **Abstract:** WS-Security is an essential component of the Web services protocol stack. WS-Security provides end-to-end security properties, thereby assuring the participation of non-secure transport intermediaries in message exchanges, a key advantage in Web-based systems. However, compared to point-to-point secure messaging with TLS, WS-Security has a significant performance penalty. In this paper, we identify several opportunities for optimizing WS-Security.

5. M. R. Head, M. Govindaraju, R. van Engelen, and W. Zhang. Benchmarking XML processors for applications in grid web services. In *Proceedings of SC'06 (Supercomputing): International Conference for High Performance Computing, Networking, and Storage*, Tampa, FL, USA, November 11-17, 2006.

- **Abstract:** Numerous XML processing tools exist today, each of which is optimized for specific features. To make the right decisions, grid application and middle-ware developers must thus understand the complex dependencies between XML features and the application. We propose a standard benchmark suite for quantifying, comparing, and contrasting the performance of XML processors under a wide range of representative use cases. The benchmarks are defined by a set of XML schemas and conforming documents. To demonstrate the utility of the benchmarks and to provide a snapshot of the current XML implementation landscape, we report the performance of many different XML implementations, on the benchmarks, and draw conclusions about their current performance characteristics. We also present a brief analysis on the current shortcomings and required critical design changes for multi-threaded XML processing tools to run efficiently on emerging multi-core architectures.

6. W. Zhang and R. van Engelen. A table-driven streaming XML parsing methodology for high-performance web services. In *Proceedings of IEEE International Conference on Web Services (ICWS 2006)*, pages 197-204, Chicago, IL, USA, September 18-22, 2006, (Best Student Paper Award, acceptance rate:18%, invited for journal extension by International Journal of Web Services Research).

- **Abstract:** This paper presents a table-driven streaming XML parsing methodology, called TDX. TDX expedites XML parsing by pre-recording the states of an XML parser in tabular form and by utilizing an efficient runtime streaming parsing engine based on a push-down automaton. The parsing tables are automatically produced from the XML schemas or a WSDL service description. Because the schema constraints are pre-encoded in a parsing table, the approach effectively implements a schema-specific XML parsing technique that combines parsing and validation into a single pass. This significantly increases the performance of XML Web services, which results in better response time and may reduce the impact of the flash-crowd effect. To implement TDX, we developed a parser construction toolkit to automatically construct parsers in C code from WSDLs and XML schemas. We applied the toolkit to an example Web services application and measured the raw performance compared to popular high-performance parsers written in C/C++, such as eXpat, gSOAP, and Xerces. The performance results show that TDX can be an order of magnitude faster.

7. R. van Engelen, M. Govindaraju, and W. Zhang. Exploring remote object coherence in XML web services. In *Proceedings of IEEE International Conference on Web Services (ICWS 2006)*, pages 249-256, Chicago, IL, USA, September 18-22, 2006.

   - **Abstract:** Object coherence in platform-specific and tightly-coupled systems is achieved with binary serialization protocols to ensure data structures and object graphs are safely transmitted, manipulated, and stored. On the opposite side of the spectrum are platform-neutral Web services that embrace XML as a serialization protocol for building loosely coupled systems. The advantages of XML to connect heterogeneous systems are plenty, but rendering programming-language specific data structures and object graphs in text form incurs a performance hit and presents challenges for systems that require object coherence. Achieving the latter goal poses difficulties by a phenomenon that is sometimes referred to as the "impedance mismatch" between programming language data types and XML schema types. This paper examines the problem, debunks the O/X-mismatch controversy, and presents a mix of static/dynamic algorithms for accurate XML serialization.

8. D. A. Gaitros, W. Zhang, A. Mast, G. Riccardi, and F. Ronquist. A biodiversity semantic associative annotation tool. *Proceedings of International Conference on Internet Computing (ICOMP'06)*, pages 29-35,Las Vegas, Nevada, USA, June 26-29, 2006.

   - **Abstract:** This paper presents a methodology for the creation of an annotation tool for on-line collaboration and sharing of heterogeneous data through a common medium. The annotation tool combines the advantages of highly organized relational database, extensible XML schemas, Life Science Identifiers, and accepted industry ontology using DataGrid technologies to facilitate collection and sharing information on biological specimens. Schematized annotation provides biologists with a flexible framework to perform annotations using their own data models. Structured XML documents enable structure-based semantic retrieval to improve the query accuracy. Retrieval performance can also be improved by combining the relational database and XML documents.

9. W. Zhang and R. van Engelen. TDX: a high-performance table-driven XML parser. In *Proceedings of the ACM Southeast conference*, pages 726-731, Melbourne, FL, USA, March 10-12, 2006.

   - **Abstract:** This paper presents TDX, a table-driven XML parser. TDX combines parsing and validation into one pass to increase the performance of XML-based applications, such as Web services. The TDX approach is based on the observation that context-free grammars can be automatically derived from XML schema. We developed a parser construction tool to automatically construct TDX grammar productions from a schema. Grammar tokens are defined by the specific schema element names, attribute names, and text. Because most of the structural constraints in XML schema are cast as grammar rules, parsing and validation of XML instances are efficiently implemented.

10. R. van Engelen, W. Zhang, and M. Govindaraju. Toward remote object coherence with compiled object serialization for distributed computing with XML web services. In *Proceedings of Compilers for Parallel Computing Workshop*, pages 441-455, 2006.

- **Abstract:** Cross-platform object-level coherence in Web services-based distributed systems and grids requires lossless serialization to ensure programming-language specific objects are safely transmitted, manipulated, and stored. However, Web services development tools often suffer from lossy forms of XML serialization, which diminishes the usefulness of XML Web services as a competitive approach to binary protocols. The difficulty mainly originates from the impedance mismatch between programming language data types and XML schema types. To overcome this obstacle, we propose hybrid static/dynamic algorithms to support lossless serialization of programming-language specific binary-encoded object graphs to text-based XML trees, while staying within the limits imposed by XML schema validation and the XSD type system. This paper presents a compiler-based approach to automatically emit serialization routines for C and C++ data types to XML.

Presentations

1. *High-Performance XML Parsing and Validation with Permutation Phrase Grammar Parsers*, ICWS2008, Beijing, China, September 23-26, 2008.

2. *An Overview and Evaluation of Web Services Security Performance Optimizations*, ICWS2008, Beijing, China, September 23-26, 2008.

3. *A table-driven streaming XML parsing methodology for high-performance web services*, ICWS2006, IL, USA, September 18-22, 2006.

4. *Exploring remote object coherence in XML web services*, ICWS2006, IL, USA, September 18-22, 2006.

5. *TDX: a high-performance table-driven XML parser*, ACMSE2006, Melbourne, FL, March 10-12, 2006.

Selected Recent Projects

- **gTDX Project: Code Generator Tools for Table-Driven XML Parser in C/C++**

  The goal of the this project is to construct a toolkit that automatically generate code for a Table-Driven XML Parser (TDX) in C/C++ from XML schemas or a WSDL description for high-performance applications. TDX is a schema-specific XML parsing methodology that integrates XML well-formedness parsing, validation, and application data handling in one stage. TDX approach is based on the observation that context-free grammars can be automatically derived from XML schemas. XML Well-formedness parsing and structure are verified through grammar productions; CDATA validation are accomplished thought semantic actions; also, application-specific events can be invoked through semantic actions. TDX pre-records the states of an XML parser in tabular form at compile time by exploiting schema information, and utilizes an efficient runtime parsing engine based on a push-down automaton. Because the runtime parsing engine is independent of schemas, TDX parsing tables can be implemented as hot-swap modules, as result, parsing tables can be swapped on the fly when the schema is updated, with no requirement of re-compilation of applications.

- **Streaming XPath Processing Project**

The goal of this project is to create a streaming XPath processing system, using our TDX approach, which extracts XML data on the fly given an XPath expression . The system consists of an XPath Expression Processor (XXP) and an XPath Engine. The XXP takes an XPath expression as input, computes the structural information, and marks parsing tables accordingly. The XPath engine is similar to the TDX parsing engine. This system shares the advantages of TDX parser, such as parsing efficiency, integrating parsing and validation. Furthermore, it can extract XML data on the fly without buffering XML message because the structural information for an XPath expression is incorporated in the parsing tables.

- **MorphBank Project**

    MorphBank is an open web repository of images serving the biological research community, housed at School of Computational Science at Florida State University. It is currently being used to document specimens in natural history collections, to voucher DNA sequence data, and to share research results in disciplines such as taxonomy, morphometrics, comparative anatomy, and phylogenetics.

    I am working on implementation of an LSID authority system for MorphBank. In addition, I am investigating, proposing and implementing message-level security solutions to MorphBank LSID-based Web services.

_____