IDEA GROUP PUBLISHING **ITJ3077**

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

This chapter appears in the publication, International Journal of Web Services Research Vol. 3, Issue 1
edited by Liang-Jie Zhang © 2005, Idea Group Inc.

Bandwidth and Latency Considerations for Efficient SOAP Messaging

Christian Werner, University of Lübeck, Germany

Carsten Buschmann, University of Lübeck, Germany

Tobias Jäcker, University of Braunschweig, Germany

Stefan Fischer, University of Lübeck, Germany

ABSTRACT

Although Web service technology is being used in more and more distributed systems, its areas of application are inherently limited by high latencies and high amounts of protocol overhead. For messaging in environments with user interaction, like Web platforms for business or multimedia applications, the response time of the whole system needs to be kept in tight boundaries. In other scenarios including mobile communication and battery-powered devices, bandwidth-efficient communication is imperative. In this paper, we address both of these issues. First we conduct a detailed latency analysis of different transport mechanisms for SOAP and then we thoroughly investigate their protocol overhead. For both aspects we present a theoretical analysis as well as experimental measurement results. We then will introduce a new transport binding called PURE that significantly reduces the protocol overhead while featuring low latency. Furthermore it enables interesting additional features such as point-to-multipoint communication via IP multicast and broadcast.

Keywords: *SOAP; Web service technology; XML markup*

INTRODUCTION

A major drawback of using Web services for application integration is its enormous demand for network bandwidth. Like all other XML protocols, SOAP suffers from the fact that only a very small part of the transmitted message contains real payload. The rest of it is XML markup and protocol overhead. Comparisons on different approaches for realizing Remote Procedure Calls (RPC) have shown that SOAP over HTTP uses significantly more bandwidth than competitive technologies (Werner Buschmann, & Fischer, 2005; Tian et al., 2003; Marahrens, 2003). Though today's wired net-

works are powerful enough to provide sufficient bandwidth even for very demanding applications like media streaming, there are still some fields of computing where bandwidth is costly. In cellular phone networks (GPRS, UMTS), for example, it is quite common to charge according to the transmitted data volumes.

Another problem, which might become even more severe in the future, is the comparably high latency of SOAP-based communication. Since not all service operation can be processed in parallel, the response time of the whole system increases with the number of involved services. Especially when using several intermediaries between SOAP endpoints, a single services call might take a considerable time. Hence, if an operation with user interaction needs a number of subsequent calls to complete, the responsiveness of the whole system decreases below an acceptable level.

Typical application domains, in which low-latency is particularly important, include high-performance application in the field of grid computing as well as all kinds real-time applications, such as controlling industrial devices and plants. But also all Web service that are used in applications with user interaction have to meet certain latency restrictions: Shneiderman (1984) found that the acceptable response time depends on the user's perception of the complexity of the task the computer system has to solve. For easy tasks, like login procedures, delays up to 200 milliseconds are acceptable. For more complex tasks, that is, search operations on large databases, a computer system should respond within a time interval of 2 seconds. Higher response times lead to decreased usability and unsatisfied users.

Allman conducted a detailed survey on Web service response time using RPC-style Web service calls (Allman, 2003). Under local area network conditions, he measured latencies between 50 and 70 milliseconds for a single Web service call. For more complex test cases, where we have several SOAP intermediaries that are connected over the Internet, the overall response time can easily go up over values larger then 2 seconds. These results show that the optimization of Web service latency is a very important issue. In the past, approaches for improving the response time performance of SOAP services mainly concentrated on the optimization of the used XML parser (van Engelen, 2004; Chiu, Govindaraju, & Bramley, 2002), the SOAP engine (Marahrens, 2003) or the used Web server (Tian et al., 2003; Ghandeharizadeh et al., 2002). But as we will show, the used transport binding also can have a significant impact on the Web service performance with regard to its latency.

Bearing the "talkative" nature of XML in mind, our recent work focused on strategies for compressing SOAP messages. First we will summarize our approach and present the resulting reduction of network traffic. Here we will also discuss how the data rate efficiency and latency can be enhanced further by transport-level considerations. Next we will present a survey on existing transport protocol bindings, again with a strong focus on both data rate efficiency and latency. The following section introduces a new binding based on UDP, which is extremely efficient and versatile at the same time. Furthermore, we will discuss the results of a thorough evaluation of this new binding. Finally, we will draw a conclusion and discuss some future work.

PROBLEM ANALYSIS

Overhead

In our past efforts to enhance Web service efficiency with regard to the generated network traffic we concentrated on SOAP compression. We found that xmlppm (http://sourceforge.net/projects/xmlppm/) is the most effective compressor that can be applied to arbitrary XML documents. It achieves an average compression ratio of 1:2.3. WBXML (http://libwbxml.aymerick.com/) produces even more compact XML representations, but is limited to documents sticking to certain predefined XML languages. Hence, it cannot be applied to SOAP documents.

A Similarity Measure Across Ontologies for Web Services Discovery

Aissa Fellah, Mimoun Malki and Atilla Elci (2019). *Web Services: Concepts, Methodologies, Tools, and Applications (pp. 859-881).*

www.igi-global.com/chapter/a-similarity-measure-across-ontologies-for-web-services-discovery/217867?camid=4v1a

A Decentralized Framework for Semantic Web Services Discovery Using Mobile Agent

Nadia Ben Seghir, Okba Kazar and Khaled Rezeg (2019). *Web Services: Concepts, Methodologies, Tools, and Applications (pp. 530-553).*

www.igi-global.com/chapter/a-decentralized-framework-for-semantic-web-services-discovery-using-mobile-agent/217849?camid=4v1a