Chapter IV

Efficient Transport Bindings for Web Service Messages

Christian Werner, University of Lübeck, Germany
Carsten Buschmann, University of Lübeck, Germany
Tobias Jäcker, EVES Information Technology AG, 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 comprising mobile communication and battery-powered devices, bandwidth-efficient communication is imperative. In this chapter 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.

**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 (RPCs) have shown that SOAP over HTTP uses significantly more bandwidth than competitive technologies (Werner, Buschmann, & Fischer, 2005; Werner, Buschmann, Brandt, & Fischer, 2006; Tian, Voigt, Naumowicz, Ritter, & Schiller, 2003; Marahrens, 2003). Though today’s wired networks 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 systems increases with the number of involved services. Especially when using several intermediaries between SOAP endpoints, a single services call might take 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 of real-time applications like controlling industrial devices and plants. But also all Web services which 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, for example search operations on large databases, a computer system should respond within a time interval of two seconds. Higher response times lead to decreased usability and unsatisfied users.

Allman (2003) conducted a detailed survey on Web service response time using RPC-style Web service calls. 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
Rapid Development of Adaptable Situation-Aware Service-based Systems
[www.igi-global.com/chapter/rapid-development-adaptable-situation-aware/41520?camid=4v1a](www.igi-global.com/chapter/rapid-development-adaptable-situation-aware/41520?camid=4v1a)