An introduction to XML and SOAP

A final year project discussion paper

Table of Contents

About the author

Michael's student number is 964076. He can be contacted at mikal@stillhq.com.

Abstract

As far as Microsoft is concerned, the future of computing is .NET and distributed computing. A large part of their infrastructure for this is XML and a protocol known as SOAP. SOAP is the subject of this discussion paper [1] . SOAP is the Simple Object Access Protocol, and is essentially a XML RPC over HTTP protocol.

The meeting

This report is based on a day long tutorial on XML and SOAP which I attended on 25 April 2002, as part of the *VSLive!* Conference in Sydney, Australia [2] . The speakers at this session were Brian Randall [3] and Ken Getz [4] .

What is SOAP

History

Microsoft has toyed with remote procedure call methods for many years. The most common of these are the moment would have to be DCOM [5] . SOAP (Simple Object Access Protocol) is an attempt to develop a RPC mechanism which will work well over the Internet. Interestingly, SOAP is also one of the few examples of a standard championed by Microsoft which has been published as an open standard [6] , which means developers outside of Microsoft are empowered to develop their own implementations if they want to.

One of the preceived problems with DCOM is that it doesn't work well over the Internet, especially because it requires its own network protocol be allowed transit through the intervening firewalls. SOAP attempts to address with by tunneling its RPC calls through HTTP, which makes it easy for many organisations to set it up within their existing firewall infrastucture. Incidentally, there is some discussion at the moment of whether SOAP should be allowed it's own network protocol as well.

The example

When discussing the use of SOAP, will will use the example of TRIM, which is a heavily RPC orientated client server package developed by TOWER Software in Canberra. It is a prime candidate for conversion to SOAP, as it currently uses DCOM, the previous Microsoft technology for RPC calls over the Internet.

The protocol

SOAP is defined in terms of an XML DTD [7] , which is what has been submitted to the *w3c* as a proposed specification. A RPC call (when using the HTTP transport) takes the form of a request for a given page, which the RPC arguements packed into an XML document which is posted to the web service [8] using the same mechanism that POSTs use under CGI. The result of the RPC call is returned in the form of another XML document, which would be displayed in the browser if the client was a webserver.

It can be seen from this description that the protocol requires a fair bit of overhead. Not only do both the client and web service need to be able to talk HTTP, they also need to be able to generate and parse XML. Web proxies were also not originally intended for the style of traffic that SOAP will impose on them. At the moment if you browse a web site, then a series of connections to the web server are made [9] and data is returned. The data is then rendered for the user. Even on a fast Internet connection this can be quite frustratingly slow. Imagine that those requests are now the challenge and response for a logon to TRIM, which currently takes less than a second.

One possible solution to at least part of this problem is to make client server protocols require less calls to perform common opertations. It should be noted however that competent client server programmers already attempt to minimise the amount of network traffic incurred by their product.

Response to the talk

The talk, as well as the speakers, were well presented. The conference procedings includes over 260 slides for this day long session, even though additional slides were used to further explain topics that the audience had difficulty grasping. The speakers understood how to keep the audience engaged and interested for the entire length of the session as well, and there were plenty of well explained examples during the session.

One criticism of the speakers were that they were very Microsoft focussed in their discussions, with most of the examples being in Visual Basic (the exceptions were in Microsoft's new language, C#). It is not entirely unexpected at a Microsoft development conference.

Notes

1. XML is considered to be outside the scope of this paper, but it is assumed that the reader has some knowledge of the workings of XML
2. The conference website can be found at http://www.vslive.com. Repeats of this conference are heald in Sydney, New York, and Orlando
3. http://www.vslive.com/2002/ny/speakers.asp#randall
4. htto://www.mcwtech.com

5. The distributed COM implementation, which unfortunately has problems dealing with low latency or low bandwith network connections – for instance most Internet connections.
6. It is currently a draft specification with the *w3c* http://www.w3.org/TR/SOAP/
7. Document Type Definition – a way of defining a grammar that your XML will obey for a certain type of transaction
8. The name given to a piece of software which offers RPC facilities under SOAP
9. Although HTTP 1.1 allows for all traffic to flow on one connection