

简单对象访问协议 SOAP 的研究与分析

The Research and Analyse of SOAP

杭州电子工业学院 周必水 倪慧莉 (杭州 310012)

摘 要: SOAP,是由万维网联盟(W3C)制定的一个新通讯协议 Simple Object Access Protocol 的简称,前身为 XML-RPC。文章详细分析了 SOAP 机制的构成原理,并探讨了与其它技术(HTTP 与 DCOM)的差异,展望了用 SOAP 开发网络应用程序的发展前景。

关键词: RPC,XML,HTTP,互操作

1 引言

不同平台之间实现互操作的传统做法是用各自的二进制代码进行转换,但在数据共享方面就受到限制。如何让不同的系统之间能够在 Internet 上透明地通信,便成为网络集成技术的发展目标。SOAP 是一种让不同应用程序之间通过 HTTP 通讯协议、以 XML 格式实现消息互换的通用规范。它为在松散的分布式环境中对等地交换结构化和类型化的信息提供了一个简单的轻量级机制;并且由于采用基于 XML 的消息通信模式,具有平台无关性之外,还是自我描述的和可扩展的。

2 构成原理

SOAP 本身并未定义任何应用语义,如编程模型或特定语义实现,它只是定义了一种简单的机制,通过一个模块化的封装模型和对模块中特定格式数据的编码规则来表达具体应用语义;而系统的实现还需依赖传输协议等其它机制。从协议栈来看,它更近乎表示层,其核心是用于 Web 通信的简单、灵活的数据和命令表示规范。它可以被许多系统扩展到 RPC(远端程序调用)中。

整个规范由相互独立的三部分组成:信封、数据编码集、RPC 描述。“信封”定义整个消息的框架;“数据编码集”定义各种用于交换的数据类型;“RPC 描述”定义与远程调用有关的具体约定。

2.1 结构模型

和许多标准的协议一样,SOAP 采用客服模型的请求—响应机制。由客户程序发出请求、送出参数,而服务器程序作出相应的响应,两种消息均为 XML 格式。消息的基本结构(以 HTTP 为载体)如图 1 所示。无论用什么通讯协议做载体,SOAP 消息的结构框架都是不变的。从图 1 中可以看到,这是典型的一种 XML 描述,Envelope 为根元素,一个可选 Header 和一个必须有的 Body 为子元素。

收稿日期: 2001-09-29

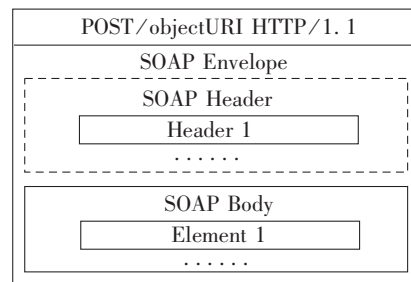


图 1 SOAP 消息体

一个最简单的物理考试查分请求可以写成:

```
POST/examserver/object23 HTTP/1.1
Host: 202.163.2.2
Content-Type: text/xml
<SOAP-ENV:Envelope
xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
SOAP-ENV:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/">
<SOAP-ENV:Body>
<m:query xmlns:m='urn:examdb:ISStudent'>
<theString>physics</theString>
</m:query>
</SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

服务程序在查询结束后返回结果(89分),响应消息为:

```
200 OK
Content-Type: text/xml
<SOAP-ENV:Envelope>
<SOAP-ENV:Body>
<m:queryResponse
xmlns:m='urn:examdb:ISStudent'>
<result>89</result>
</m:queryResponse>
</SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

Envelope 用 encodingStyle 属性指明全局的数据编码规则,即命名空间。应用程序在生成由 SOAP 定

义的所有元素和属性时,必须包含恰当的命名空间(标准命名空间为 `http://schemas.xmlsoap.org/soap/encoding/`,也可以有其它的选择)。

Body 元素由所需交换的数据组成,包含请求或响应消息,以及可能返回的出错信息,它是整个 SOAP 消息的主体部分。

Header 元素是可选的,主要用于模块化地扩展 SOAP 消息的描述能力(典型的扩展例子诸如认证、事务管理以及 B2B 协同等 Internet 扩展业务),而通讯双方并不需要完整的预先的约定。其子元素的 Actor 属性指明处理此业务的中间节点。当消息传送路径中的中间节点接收到消息报文,则判断自己的 URI 是否与 Actor 一致,是则按照子元素内容过滤该消息,否则直接转发到下一节点。

2.2 编码规则

SOAP 编码规则基于一个概括了程序语言、数据库和半结构化数据等类型系统共性的简单类型系统,是 XML 规定的一个子集。它具有以下特点:

(1) 所有类型分为简单和复合两种。简单类型与 XML 语法 2 中规定的内置数据类型相一致,支持枚举、base64 字符编码、联合等基本类型。复合类型中又包括数组和结构体。

(2) 每个值都有局部或全局的存取标识,以便进行序列化。数组元素的存取标识为数据项下标。其余类型的存取标识则多为元素名,但在复合类型中也可能出现数据项下标与元素名并用的情况。

(3) 元素可直接用 `xsi:type` 属性进行自描述。

(4) 被多个存取标识引用的值应编码为一个独立元素,用 ID 属性作为 URI 标志符。而引用它的那些值则为空元素,并用 Href 属性对应被引用值的 ID 属性。

(5) 远程调用中,无论是请求消息还是响应消息,所调用的 URL、方法名、接口、参数等相关信息均封装为单一的结构体。每个输入、输出参数都有相应存取标识。

尽管 SOAP 规范对数据编码格式按照 XML 作了明确规定,但这并不意味着消息传送的双方必须有完全相同的分布式对象结构,事实上,它们只需要能互相理解对方所传送的那部分数据。在 Internet 上经常会有不确定的服务,其服务内容、形式往往随时间等因素发生变化,这些应用程序的编写要求很大的灵活性。采用这种松耦合的消息传递机制,为在开放的 Web 环境下协同工作及相互集成提供了最大范围的支持。

3 比较

在分布式远程调用的程序设计逻辑中,要求建立准确的分布式对象模型以支持双方不同的运行环境,客户程序与系统所提供的服务之间必须使用同一基本结构,即紧密耦合。若服务器端接口发生变化,客户程序也必须重新修改。因此,只适用于某一具体的分布式环境,而对千变万化的 Internet 则有些力不从心。

目前两种主要的 ORPC 技术是 DCOM(分布式组件模型)和 IIOP(Internet Interoperate Orb Protocol)。它们分别用 NDR 和 CDR 描述信息,两者间不能兼容。尽管 IIOP 比 DCOM 更容易实现,但两个协议都有相当多的复杂规则来处理数据排列、类型信息和位操作,并且要依赖于确定的网络运行环境和服务器配置。

与分布式设计逻辑不同,SOAP 与具体的分布式对象模型无关,它只需将要传递的数据打包成有自我描述功能的文本消息,发往对方即可。和 CORBA/IIOP 一样,SOAP 使用基于终端的面向对象的通信方式,一个 SOAP 终端可以看作一个基于 HTTP 的 URL,并用此来识别所调用的目标。具体的调用对象并不是被绑定到一个给定的终端,而是由具体实现程序来决定怎样把终端标识符映射到服务器端的对象。图 2 给出了 SOAP 请求消息与 ORPC 组件的对应关系。服务器端最终采取怎样的映射模式调用对象则依据具体环境而定,并不影响调用的结果和响应消息的返回。如上所述,由于数据编码采用松耦合的结构,服务端程序较之 ORPC 更为轻便和灵活。

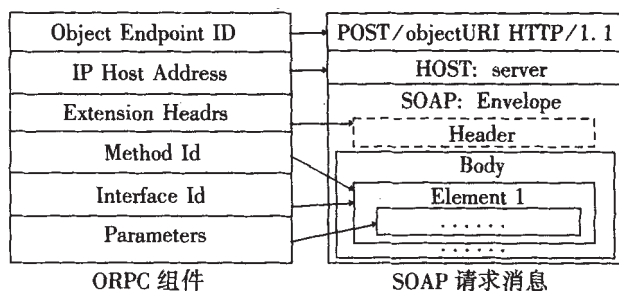


图 2 对应关系图

从功能上看,SOAP 和 IIOP 不象 DCOM 一样具备分布式碎片收集、成批发送消息一类的附加功能。DCOM 非常适用于服务器到服务器的通信,对客户——服机制则过于庞大。而 SOAP 由于其简易性的设计目标,在如何激活所调用的对象及生存期管理机制等方面未加以定义,它更适于实现广泛的 Web 分布式服务。

4 实现

由于利用的是现成技术的集成(如 XML 和 HTTP),套用与开发 SOAP 程序是很容易的。尽管这样会使运行速度略微减慢,但从更便于与其它系统相互沟通的角度却是值得的,事实上 SOAP 本身并不是用来代替原有的低级程序。SOAP 开发工具已被许多开发环境所支持,如 Microsoft SOAP Toolkit 2.0 和 Apache SOAP 2.0,且不受编程语言限制,Java、Visual Basic、Perl 等均可。利用这些开发工具,我们只需编写相应的客户及服务器程序,至于 XML 的解析、底层网络传输机制(如 Socket 连接)的实现等等,已内嵌于开发工具之中。

出于安全的考虑,防火墙技术被广泛应用在 Internet 中,很多防火墙对发送/接收报文所使用的通讯协议加以限制。使用 HTTP 传输则可以畅通无阻。防火墙将通过 HTTP 头部字段 Content-Type (text/xml-SOAP) 和 SOAPAction 识别出 SOAP 报文,并过滤掉接口和方法名不同的报文。

5 结束语

HTTP 并非 SOAP 唯一的选择,尽管 SOAP 1.1 规范只规定了 HTTP 加载 SOAP 协议数据单元的格式,实际上它可与许多其它的通信协议兼容,如 SMTP、POP3 等。IETF 也已于近日提出用 SIP (Session Initiation Protocol) 传输 SOAP 消息的草案。这些通讯协议与 SOAP 一样,具有简易性和可扩展性。

随着全球信息化步伐的迅猛推进,如何实现各种网络服务在 INTERNET 上的跨平台、跨软硬件的

无缝集成,成为 IT 业共同关注的一个问题。用 SOAP 开发网络应用程序有助于实现大量异构程序和平台之间的互操作性,从而使现存的应用能够被最广泛的用户所访问。它将会成为网络服务提供商的新的选择,同时这种机制也会在电子商务等应用领域得到更大的发展。

参考文献

- [1] William Ruh, et al. IIOP Complete, Paul Klinker (Addison Wesley), 1999.
- [2] D Box, et al. Simple Object Access Protocol (SOAP) 1.1, W3C Note, <http://www.w3.org/TR/2000/NOTE-SOAP-20000508/>, 08 May 2000.
- [3] Tom Jepsen. SOAP Cleans up Interoperability Problems on the Web, IT Professional, Jan/Feb 2001.

ZHOU Bi-shui, NI Hui-li (Hangzhou Institute of Electronics Engineering, Hangzhou 310012)

Abstract: SOAP means Simple Object Access Protocol, which began as XML-RPC created in 1998. The structure and principle of it are analysed in this paper. We also compare the schema of SOAP with other mechanisms like IIOP and DCOM. At last, the future of developing web application using SOAP is discussed.

Key words: RPC, XML, HTTP, Interoperate

周必水 男,1964 年生。副教授,硕士生导师。主要从事计算机网络系统、OA 的科研与教学。

倪慧莉 女,1975 年生。硕士研究生,研究方向为多媒体通信网络与信息处理。

(上接第 14 页)

参考文献

- [1] 唐荣锡等. 计算机图形学教程 [M]. 北京:科学出版社,1990,200~242.
- [2] 王永生. 最新 C 语言实用绘图程序与图形. 北京:科学出版社,1996,65~86.
- [3] 江涛,钟宏. 数据结构 [M]. 北京:科学出版社,1995,65~82.
- [4] Reginald C Farrow, Warren K Waskiewicz et al. Alignment Mark Detection in CMOS Materials with SCALPEL E-beam Lithography[J]. SPIE vol 3676, 1999, 3, 223~226.

SHI Qiang, ZHANG Fu-an (Institute of Electrical Engineering, Chinese Academy of Science, Beijing 100080)

Abstract: The alignment technology is introduced. The software of alignment simulation for PELDI is designed. The mathematical model and the arithmetic of the software are described. Two sorts of BSE signals are simulated as examples.

Key Words: Alignment technology, Software of alignment simulation, PELDI

史强 男,1976 年生,硕士研究生。主要从事电子束曝光对准系统研究。

靳鹏云 女,1970 年生,硕士研究生,助理研究员。主要从事电子束曝光研究。

张福安 男,1942 年生,研究员。