

# 基于差分编码和压缩的 SOAP 优化

刘欣, 谢琦

(郑州大学信息工程学院, 郑州 450001)

**摘要:** 针对 SOAP 在序列化和反序列化过程中效率低及消息自身的冗余问题, 提出在 .NET 框架下使用 SOAP 扩展以优化 SOAP 消息。对于 SOAP 请求消息, 采用增加消息模版的方式实现差分编码, 对于 SOAP 响应消息, 根据消息的长短分别采用压缩或差分的方法来优化, 减少带宽占用, 使 Web 服务得到优化。

**关键词:** 简单对象访问协议; SOAP 扩展; 差分编码; 压缩

## SOAP Optimization Based on Differential Encoding and Compression

LIU Xin, XIE Qi

(Department of Information Engineering, Zhengzhou University, Zhengzhou 450001)

**【Abstract】** Aiming at the SOAP serialization of messages and the verbose SOAP messages, this paper provides optimizing the SOAP message by using SOAP extension. For the SOAP request, it realizes differential encoding by adding SOAP message skeleton, for the SOAP response, it compresses long message or realizes differential encoding of the short SOAP message. It can save the bandwidth and optimize the Web services.

**【Key words】** Simple Object Access Protocol(SOAP); SOAP extension; differential encoding; compression

### 1 概述

Web 服务是架构在 XML 技术上, 基于标准体系的组件化松散耦合技术, 采用的核心技术包括简单对象访问协议 (Simple Object Access Protocol, SOAP), WSDL, UDDI 等标准协议<sup>[1]</sup>, 其目标是实现不同系统间跨平台、跨编程语言的互操作。其中, SOAP 用 XML 形式提供了一个简单、轻量的用于在分散或分布环境中交换结构化和类型信息的机制, 是 Web 服务的重要组成部分之一。但 SOAP 的执行效率并不像经典的分布式对象技术如 RMI, CORBA 那样高效<sup>[2]</sup>。SOAP 须额外生成和解析 XML 文档 (即序列化和反序列化的过程), 造成了高昂的 CPU 负载。它作为一种 XML 协议, 在承载少量的有用信息的同时要包含更多用于自解释的 XML 标签载荷。之前的一些优化 Web 服务研究主要采用 Cache 来提高 SOAP 性能, 或利用压缩 SOAP 消息来提高带宽利用率。SOAP 执行的低效不仅来源于 SOAP 消息自身的冗余, 还来自 SOAP 生成和解析 XML 文档的额外开销。

本文提出增加消息模版与 SOAP 请求消息进行比对并形成差分消息, 即差分编码。当 SOAP 响应消息携带数据量较大时, 采用压缩的方式使之形成紧凑有效的 SOAP 消息; 当 SOAP 响应消息数据量较小时, 采用差分比对的方式可更有效地节省带宽。该方法对不同的 SOAP 消息进行区分, 合理且有针对性地优化了 Web 服务。

### 2 对 SOAP 消息的差分和压缩

#### 2.1 基本思想

XML Web 服务的生存期<sup>[3]</sup>如图 1 所示。在阶段 1 中, 通过 1 个 SOAP 消息模版 (skeleton.xml) 与客户端形成的 SOAP 请求消息进行比对, 形成差分 XML 文档 (diffgram.xml), 通过网络进入阶段 2, 在反序列化前, 该 XML 文档与模版文件比对修补成原来的 SOAP 请求消息, 传到服务器端进行处理。如图 2 的上半部分所示。

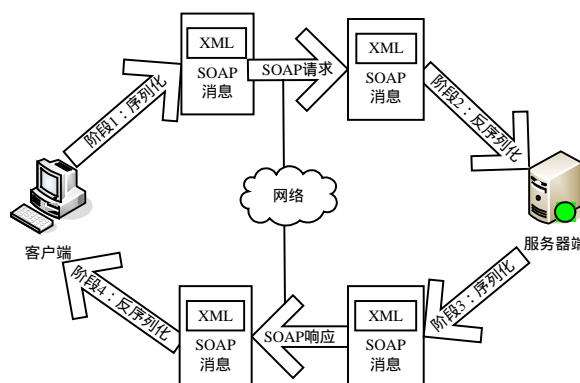


图 1 XML Web 服务生存期模型

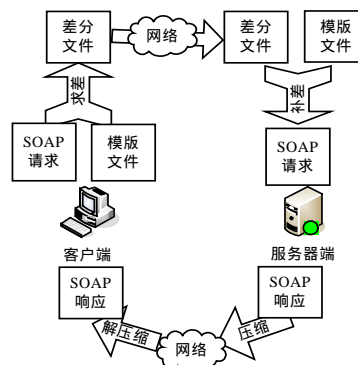


图 2 差分 SOAP 请求消息并压缩 SOAP 响应消息的过程

在图 1 的阶段 3、阶段 4 中, 须根据 SOAP 响应消息的长度做区分。对于消息较长的, 在序列化后压缩消息、反序

**基金项目:** 河南省普通科技攻关计划基金资助项目 (0424220024)

**作者简介:** 刘欣 (1982 -), 女, 硕士研究生, 主研方向: Web 服务; 谢琦, 副教授

**收稿日期:** 2008-05-14 **E-mail:** helloxyndy@126.com

列化前解压缩,如图 2 下半部分所示。对于消息较短的,过程类似于阶段 1、阶段 2,只是这里的 SOAP 消息模版要换成与 SOAP 响应消息相似的 XML 文件。

## 2.2 实现方法

### 2.2.1 SOAP 扩展

.NET 框架提供了 SOAP 扩展以实现对 SOAP 消息的修改,它可用于在客户端或服务器上处理消息时的特定阶段检查或修改消息。生成 SOAP 扩展并让其使用 XML Web 服务需要以下步骤:

- (1)从 SOAP 扩展派生一个类。
- (2)保存对表示将来 SOAP 消息的 Stream 的引用。
- (3)初始化 SOAP 扩展特定的数据。
- (4)在相关的 SoapMessageStage 或多个阶段中处理 SOAP 消息。
- (5)配置 SOAP 扩展以使用特定的 XML Web 服务方法。

### 2.2.2 求差求补工具 XML DiffPatch<sup>[4]</sup>

SOAP 消息也是 XML 文档,XML DiffPatch 是比较 2 个 XML 文档差异的工具。首先比较 SOAP 请求消息(oldstream.xml)和模版消息(skeleton.xml),并将差异结果写入差分消息(diffgram.xml)。

在 Web 服务生存期的第 2 阶段,须将差分 XML 文件(diffgram.xml)和模版文件(skeleton.xml)对比来还原原先的 SOAP 请求消息。与差分算法相关的 XML 报文(包括模版文件、源 SOAP 请求消息、差分文档、合成的 SOAP 请求消息)如下:

```
<soap:Body>
  <addtry xmlns="http://tempuri.org/">
    <a>int</a>
    <b>int</b>
  </addtry>
</soap:Body>
-----
<soap:Body>
  <addtry xmlns="http://tempuri.org/">
    <a>3</a>
    <b>4</b>
  </addtry>
</soap:Body>
-----
<xd:node match="1">
  <xd:change match="1">3</xd:change>
</xd:node>
<xd:node match="2">
  <xd:change match="1">4</xd:change>
</xd:node>
-----
<soap:Body>
  <addtry xmlns="http://tempuri.org/">
    <a>3</a>
    <b>4</b>
  </addtry>
</soap:Body>
```

由此可见,模版和差分文件还原出的 SOAP 请求消息和原来的消息几乎没有差别。对于传输的 SOAP 消息非常类似的 Web 服务而言,在生成和解析 XML 文档时只需要考虑有差别的部分即可,减小了网络带宽占用,提高了运行效率。

具体做法是在 SOAP 扩展的派生类 DiffPatchExtension 中引用 xmldiffpatch.dll。求差算法(Diff)写在客户端的序列化后,补差算法(Patch)写在服务器端的反序列化前,以实现差分的过程。

### 2.2.3 SOAP 报文的压缩

对于数据量比较大的 SOAP 响应消息,压缩数据量往往可有效地减小带宽载荷。本文采用 SharpZipLib 压缩算法。在 SOAP 扩展的派生类 Compression Extension 中引用 SharpZipLib.dll。压缩算法写在服务器端的序列化后,解压缩算法写在客户端的反序列化前。

## 3 实验

由于本文从 SOAP 请求消息和 SOAP 响应消息 2 个方面来优化 SOAP,因此实验结果也从这 2 个方面来说明。本文采用 Microsoft SOAP Toolkit3 来跟踪 SOAP 消息包,以证明实验结果。

### 3.1 SOAP 请求消息

通过与模版消息差分比来实现对 SOAP 请求消息的优化。当模版消息与源 SOAP 请求消息完全一致时,差分文件始终如一。由 SOAP Toolkit 测量可知,此差分消息长度为 239 Byte。源 SOAP 请求消息越大,效益越明显,如图 3 中的虚线所示。以此类推,两者相似度越高,经过 Internet 传送的差分文件越小,所耗工时也越少,如图 3 中的实线所示。

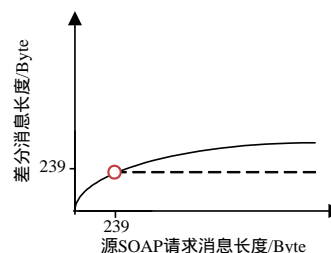


图 3 SOAP 请求消息与差分消息长度对比

### 3.2 SOAP 响应消息

当响应 SOAP 消息携带数据量很大时,压缩此消息可取得明显效果,但是针对数据量很小的 SOAP 响应消息,压缩不但不会减小数据量,还会起到反效果,如图 4 所示。

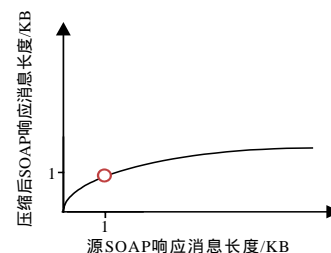


图 4 SOAP 响应消息压缩前后长度的对比

由图 4 可知,当源响应消息长度小于 1 KB 时,压缩后的消息长度反而比不压缩大。此时 SOAP 响应消息采用差分的方法会取得更好的效果,前提是设计的消息模版与响应消息要尽可能相似,可参考 SOAP 请求消息的做法。当源 SOAP 响应消息长度大于 1 KB 时,其长度越大,压缩效果越好。

综合 SOAP 请求消息和 SOAP 响应消息 2 个方面,实验结果表明:当采用差分编码时,模版文件相似度越高,优化效果越好;当采用压缩算法时,携带数据量越大,压缩效果越明显。因此,应根据消息长度区分对待,这样能取得更理想的优化效果。(下转第 146 页)