

基于 XML 远程数据传输协议的研究与实现

张 娜,焦东来,吴子杰,杨 浩

(南京邮电大学 电子科学与工程学院,江苏 南京 210009)

摘 要:物联网应用领域中,应用层的远程数据传输协议与通信程序往往融为一体,当数据传输协议改变时,通信程序一般需要大规模修改甚至重新编写,因而大大降低了通信程序的可重用性。针对这一问题,文中提出了一种基于 XML(eX-tensible Markup Language,可扩展标记语言)结构化方式来描述数据传输协议的方案,并探讨了基于该方案的解析方法。该方案可以统一数据格式,实现异构数据的通用解析。在不修改通讯程序的情况下,通过修改数据传输协议的 XML 文件,即可实现通信数据的接收和解析。最后以电能表数据传输协议的某一子协议为例,对该功能进行测试,验证了方案的可行性。

关键词:可扩展标记语言;远程数据传输协议;解析程序;重用性

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2016)05-0012-05

doi:10.3969/j.issn.1673-629X.2016.05.003

Research and Implementation of Remote Data Transmission Based on XML Protocol

ZHANG Na, JIAO Dong-lai, WU Zi-jie, YANG Hao

(College of Electronic Science and Engineering, Nanjing University of Posts and Telecommunications,
Nanjing 210009, China)

Abstract: In the Internet of Things, remote data transmission protocol and the communication program is closely related. When the communication protocol is changed, the communication program is also to be written, so the reusability of communication program is reduced. In order to solve the problem, based on the XML structure, the transmission protocol of data description is proposed, based on which the analytic method is put forward. The unified data format of the scheme can adapt to the data exchange between heterogeneous systems by modifying the transmission protocol text file with no modification on the program. Finally, given the functional testing, the feasibility of this scheme is verified.

Key words: XML; remote data transmission protocol; analytical procedures; reusability

0 引 言

在物联网应用领域中,前台与后台之间,硬件与软件之间都是通过通信程序进行数据交互。然而在通信程序发展的现阶段,不同终端和采集设备标准各异以及数据传输协议的千差万别导致了通信程序应用中信息孤岛的出现。这使得通信程序的开发门槛较高,阻碍了企业业务应用的创新,降低了面向大众提供大规模应用的可能性。通过对物联网领域数据传输协议分析可知,进行交互的大部分数据为结构化数据。该类数据存在内部逻辑关系,数据由数据项组成,数据项的

长度、内容、顺序都具有一定的规律,是一种模式(schema)强约束的数据组织形式^[1]。在这种结构化数据下,以往程序员都根据该结构对数据进行组织与编写,在编写通信程序之前,必须约定好数据传输协议,包括采集终端与后台服务器的数据传输协议、服务器与 PAD 显示端的数据传输协议等。程序员充分理解数据传输协议后,再根据相应的数据结构编写程序。这样数据传输协议与通信程序已经融为一体。然而,一旦数据传输协议发生改变,即使简单的修改,例如数据传输协议中增加或减少功能,都必须大规模修改甚至重新编写通信程序,这样增加开发人员的工作量,使

收稿日期:2015-06-06

修回日期:2015-09-15

网络出版时间:2016-05-05

基金项目:国家级面上项目(41471329)

作者简介:张 娜(1989-),女,硕士研究生,研究方向为电路与系统;焦东来,硕士生导师,博士,研究方向为空间信息可视化和数据共享与互操作。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160505.0815.030.html>

软件的可重用性与可维护性变差,降低通信效率。

对此,文中提出一种以结构化方式来描述数据传输协议的方案。该方案可在不改动通信程序的基础上适应企业提出的协议修改建议,从而使通信程序对数据的管理与传输通用化,提高了软件的重用性。

1 物联网应用层数据传输协议的特征

国际上,物联网行业中基于TCP层的数据传输协议主要包括MODBUS^[2-3]和IEC62056体系^[4],以及用于远动系统的IEC60870系列。IEC62056系列标准规定了设备间的数据交换协议,其核心部分是设备语言报文规范—DLMS (Device Language Message Specification)和能源计量配套规范—COSEM (Companion Specification for Energy Metering),即DLMS/COSEM^[5]。MODBUS主要被欧美等发达国家采用,DLMS/COSEM主要在东南亚国家使用。通过对协议结构的分析可知,无论数据的发送还是回应,其数据的组织方式都有顺序和条件两种组织形式。在顺序组织结构中,具有实际意义的数据项以顺序方式依次记录数据;在条件组织结构中,具有实际意义的数据按照数据传输过程中的具体条件对数据项进行描述。因此顺序和选择是数据通讯协议中具有的基本结构特征。

为了提高通讯软件的复用程度,宋青山等提出了基于JSON格式的数据描述^[6]。JSON格式是一种公开的数据交换格式,通过JSON描述通讯数据可以实现数据共享^[7-8];但是该格式不能描述协议中的顺序、选择等结构化特征,因此在解析过程中不能完全实现通讯程序的重用。

可扩展标记语言(XML)是一门新兴的面向Internet应用的标记语言,它是由W3C组织(互联网联合组织)于1998年2月制定的一种通用语言规范。XML凭借其良好的扩展性和自描述性、形式与内容分离、遵循严格的语法要求以及对多语种的支持等特点,给跨平台、跨地域异构应用间的协同工作,基于语义的智能数据搜索等重要领域带来了重大突破^[9]。因而,采用基于XML结构化方式来描述数据传输协议的方案是提高通信程序重用性的关键。

2 基于XML的通讯规约描述方法

在描述数据方面,XML提供了较好的描述结构化数据的方法,使用户可以用自定义的标记来描述数据元素,每个元素可以封装复杂或者简单的数据,甚至可以定义一组无限制的XML标记。XML这种结构的描述不是基于只能由软件程序去判读的代码,而是一种简单的、能够用任何编辑器读取的文本。利用这种机制,可以将原来定义在程序中的数据结构按照一定的

规则用XML描述出来。软件程序只需要完成对XML文件的解释,并按照XML的描述来组织和管理数据。当与数据交互的数据传输协议发生变化时,软件程序不需要改动,只需要修改相应的XML文件即可。由于XML文件是文本文件,易读且说明性强,因此XML的修改并不需要专门软件人员来完成,而且对于数据交互的双方,修改一次即可。针对数据传输协议中的顺序、选择、条件等结构化特征,分别定义相应的XML标签,每类标签分别定义若干属性对顺序、选择、条件等结构进行具体约束,如表1所示。

表1 数据传输协议的标签描述

标签名称	标签描述
Structure	数据传输协议中的数据结构块,包含顺序、选择、条件等数据单元,也可以包含下一级的数据结构块
Sequence	数据传输协议中的数据单元或数据块的顺序约束
Element	数据传输协议中的数据单元,其中的属性表示数据的类型、长度、精度等
Choice	数据传输协议中的数据单元或数据块的选择约束,选择的条件可以通过属性描述

Element是数据传输协议中的基本数据单元。该结构中type表示值的类型,包括:INT, String, float, short, long, unsigned, signed, bit, byte, BCD; name表示值的名称; value表示属性值; length表示值的长度; desc表示值在协议中的作用。例如:<element type=“INT” name=“preid” value=“” length=“1” desc=“帧起始符”/>定义一个整型数据单元。

Sequence在数据传输协议中表示顺序约束,它的属性是repeat,当值为“1”表示数据依次出现,当值为“unbound”表示数据循环出现。在通讯数据中数据项顺序出现一次,可以用如下方式描述:

```
<sequence repeat="1">
  <element type="INT" name="functionID" value="" length="2" desc="数据起始地址"/>
  <choice condition="functionID" respondref="=">
    <Structure conditionvalue="01" respondref="1"/>
  </choice>
</sequence>
```

Choice在数据传输协议中表示数据单元或数据块的选择约束,它的属性包括condition和respondref,分别代表选择的条件和该条件下要执行的操作。在数据传输协议中条件选择结构可用如下方式描述:

```
<choice condition="functionID" respondref="=">
  <Structure conditionvalue="01" respondref="2"/>
  <elementconditionvalue="02" type="INT4" name="M_function" value="" length="2" desc="子功能码"/>
</choice>
```

Structure是数据传输协议中的数据结构块,条件

值指向的属性为 id,即结构块对应的结构号,根据该结构号可以找到任何一个 Structure 结构。数据传输协议中当条件值为 1 时,即 id=“1”的结构块可用如下方式描述:

```
<Structure id="1">
<sequence repeat="1">
<element type="String" name="M_data" value="" length=
" " desc="数据区"/>
<choice condition="functionID" respondref="=">
<Structureconditionvalue="03" respondref="3"/>
</choice>
</sequence>
</Structure>
```

3 国内电能计量表数据传输协议分析

文中以国内电能计量数据传输为例展开研究。目前在国内数据采集终端的应用上,除了国家电力部颁布的标准协议 DL/T-645^[10] 规约系列外,《浙江省用电现场服务与管理系统通信规约》(以下简称《浙规》)和《电力负荷管理系统数据传输协议—2004》(以下简称《国电规约》)最具有代表性,受到各电力局的认可,变种非常多。尤其是《国电规约》,由国家电网公司的规模和性质决定,很有可能成为事实上的国家标准^[11],具有代表性的通信帧结构以及 XML 标签之间的映射关系,如表 2 所示。

表 2 《国电规约》通信帧结构及映射关系

代码	字节数	说明	映射结构
68H	1	帧起始符	Sequence 结构
L	4	长度	
68H	1	帧起始符	
C	1	控制符	Choice 结构
A	5	地址域	
DATA	变长	数据域	
CS	1	校验码	Sequence 结构
16H	1	结束码	

《国电规约》通信帧由“帧起始符”、“长度”、“中间帧起始符”、“控制符”、“地址域”、“数据域”、“校验码”、“结束码”组成。而这些组成中,“帧起始符”、“长度”、“中间帧起始符”、“校验码”、“结束码”映射成顺序(Sequence)结构;“控制符”、“地址域”和“数据域”映射成选择(Choice)结构,其中控制符映射成 Choice 结构的条件值(Condition),地址域和数据域是 Choice 结构中的选择结果。在 Choice 结构中选择结果由 Structure 索引号标识,不同的索引号标示不同的 Structure 结构,而该结构所对应的 Element 基本数据单元即为不同条件值所对应的地址和数据。

结合电能表数据传输协议,将其中的顺序、选择等

结构化特征转化成 XML 标签约束。电能表中的某一子协议用 XML 描述如下:

```
<? xml version="1.0" encoding="GBK"? >
<communication id="communication">
<Structure id="0">
<sequence repeat="1">
<element type="String" name="preid" defaultvalue="0x68"
value="" length="1" desc="帧起始符"/>
<element type="INT" name="length" defaultvalue="" value=
"" length="2" desc="数据长度"/>
<element type="INT" name="afterid" defaultvalue="68"
value="" length="1" desc="帧起始符"/>
<element type="String" name="M_data" defaultvalue=""
value="" length="" desc="数据区"/>
<element type="INT" name="CS" defaultvalue="" value=
"" length="1" desc="校验码"/>
<element type="INT" name="tail" defaultvalue="16" value=
"" length="1" desc="结束码"/>
<choice condition="functionID" respondref="=">
<Structureconditionvalue="01" respondref="1"/>
<Structureconditionvalue="02" respondref="2"/>
</choice>
</sequence>
</Structure>
<Structure id="1" desc="MODBUS 请求数据">
<sequence repeat="unbound">
<element type="BCD" name="M_adress" value="" length=
"2" desc="数据起始地址"/>
</sequence>
</Structure>
</communication>
```

4 基于 XML 的数据传输协议的解析

文中采用 DOM4J 作为 XML 文档的解析工具^[12], MINA 为通讯承载服务器,在 MINA 框架下耦合 DOM4J 实现基于 XML 的数据传输协议的解析^[13]。它应用于 Java 平台,而 Java 语言是一种具有简单、跨平台、面向对象等优点的语言,它的这些特性使它较适合作为 XML 编程语言。另外,XML 和 Java 技术还有许多互补特性,两者的结合将形成一个强大的数据共享和处理平台。文中在 XML 文档解析过程中,一方面采用面向对象的技术建立通讯协议结构特征(标签)对应的对象,对象包含了数据的解析方法;另一方面,在 DOM4J 解析过程中,根据标签的名称构建相应的对象,对象的解析方法实现通讯协议中数据块的解析。

4.1 基于 XML 的数据传输协议解析模型

依据通讯协议结构特征(标签),采用面向对象的技术设计对应的对象类,对象类包含了相应数据的解析方法(见图 1)。通信协议中结构之间的关系通过对

象类的递归耦合实现,如 P_Structure,P_Sequence 和 P_element,分别对应数据结构中的 Structure 数据块、Sequence 顺序结构和 element 基本数据单元,类 P_choice 对应数据结中的 choice 结构,它与类 P_Sequence 依赖。一帧数据解析完毕后将结果存入 E_Value 中,它

与类 StructureManage 聚合。在应用过程中,该模型通过 DOM4J 实现与 MINA 的结合,即在解析基于 XML 的数据传输协议文件的同时,实现对象的构建,并提供给 MINA 解析使用。

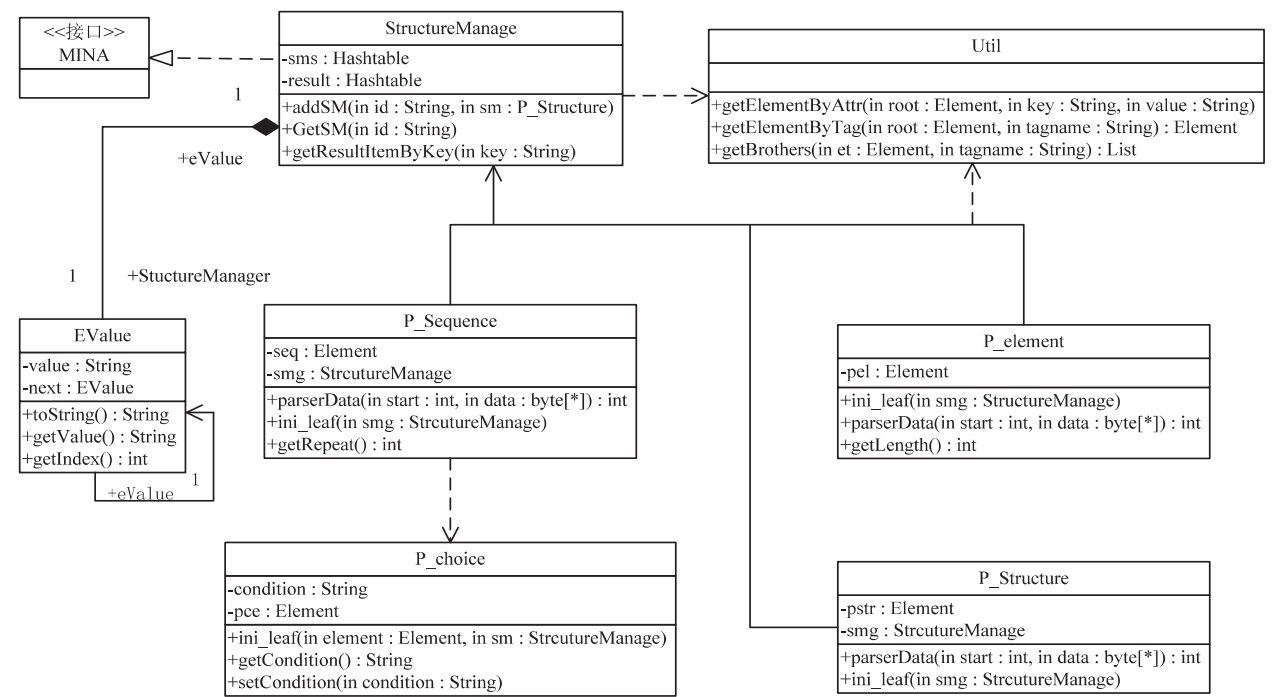


图1 XML 数据传输协议的解析模型

4.2 基于XML的数据传输协议解析过程

文中采用节点递归调用的解析方式。当 MINA 通信程序收到请求后,通过接口将消息转发给 StructureManage 类,该类通过 id 标记 Structure 结构,将遍历得到的值存入 Hashtable 表中。有三种类与 StructureManage 关联,最后将数据返回给 MINA 通信程序,XML 文档解析的逻辑模型如图 1 所示。

XML 解析程序集成在 MINA 通信程序中。通信程序运行后,XML 解析程序将 XML 文档读入内存,当有数据传来,解析程序调用类 StructureManage 接收一

帧数据,数据接收完毕后,首先判断一帧数据是否完整,若不完整则重新接收数据,完整则进行解析。数据解析时首先调用 Util 工具类遍历 XML 文档,当遍历的标签分别为“Structure”、“Sequence”、“element”、“choice”时则分别调用“P_Structure”、“P_Sequence”、“P_element”、“P_choice”类进行相应的解析,最后将解析结果通过元素节点 Evalue 返回给类“StructureManage”,该类再通过接口将数据值传送给 MINA 通信程序,通信程序对其进行后续处理。XML 文档的解析流程图如图 2 所示。

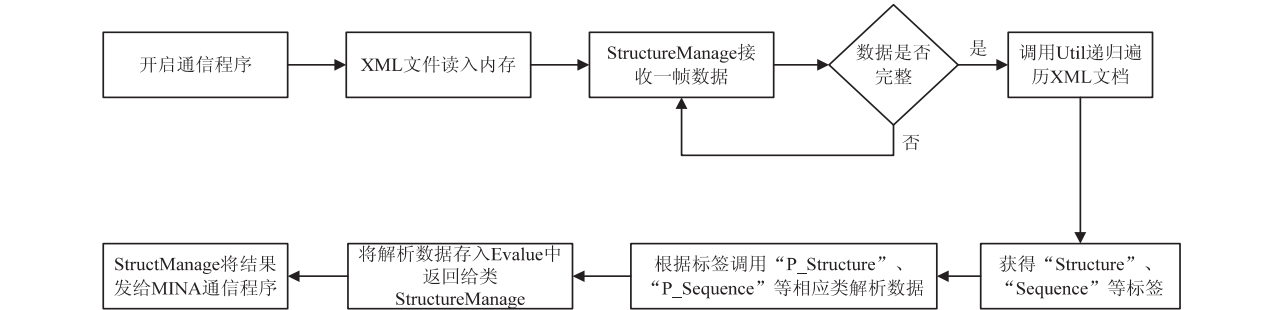


图2 XML 数据传输协议解析流程图

5 不同XML文件的解析结果

当数据传输协议改变时,MINA 通信程序不变,只需修改 XML 文件,即可实现对不同数据帧的解析和

回应。

在实验测试中,使用不同数据传输协议生成两个配置文件 XML1、XML2,它们的数据帧通信情况如图 3、4 所示。

请求协议帧:
68001607010000150000050000000015012711420068000104ff00000002f211102001000001010201100001201016
通信程序回应帧:
68001607010000150000050000000015012711420068000104FF00000001F211102001000001010201100001201016

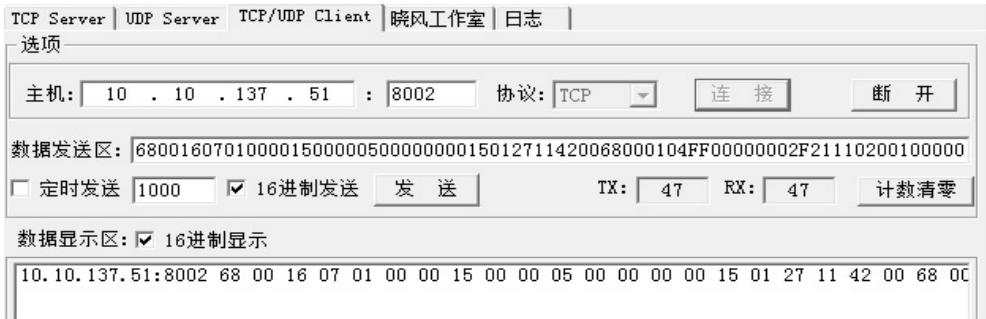


图 3 XML1 的通信结果

请求协议帧:68081234567890111234567890116801FF000000010102070116
通信程序回应帧:6800081234567890111234567890116801FF000000010102070116



图 4 XML2 的通信结果

通过测试结果可以看出,只要按照一定的 XML 格式对数据传输协议进行描述,调用该解析程序,即可实现数据的交互与共享。该解析程序的通用性在国电规约数据传输协议和远程遥测项目中都得到了验证。

6 结束语

文中提出的基于 XML 结构化方式来描述数据传输协议的方案,可以实现异构数据的通用解析。当数据传输协议改变时,不必修改解析程序,只需要修改配置文件 XML,即可实现数据解析。这样提高了通信程序的复用程度,加快了系统的整体开发速度,为企业业务应用的创新提供了可能性。但是解析 XML 文档时,需要从上到下对每个节点进行遍历,增加了时间开销,降低了程序的运行效率,因此在如何提高效率这个问题上仍需要做进一步的研究。

参考文献:

[1] 张源,魏冬邦,康昊. 基于 XML 的测控软件通信协议的研究[J]. 遥测遥控,2010(6):42-46.
[2] 谭文恕. 传输协议现状及国内外相关标准[J]. 电力设备,2004,5(7):11-14.
[3] 陈德仙,郑登峰,俞国勇,等. 基于 MODBUS 协议的智能配

电系统设计[J]. 江苏电器,2007(4):25-28.
[4] 杨洪. IEC60870-5 传输规约的应用及试验[J]. 电力自动化设备,2007,27(11):75-78.
[5] 杨晓西. 计量仪表和抄表系统的设计与 IEC62056 标准[J]. 电测与仪表,2005,42(11):37-41.
[6] 宋青山. JSON 数据格式在企业数据交换中的应用[J]. 天津冶金,2014(2):114-117.
[7] Lin B,Chen Y,Chen X,et al. Comparison between JSON and XML in applications based on AJAX[C]//Proc of international conference on computer science & service system. Nanjing:[s. n.],2012:1174-1177.
[8] 聂晓旭,于凤芹,钦道理. 可配置协议转换方法的设计与实现[J]. 计算机技术与发展,2015,25(4):80-84.
[9] 张富国. 基于 XML 的通用数据交换平台的设计及实现[J]. 科技广场,2005,17(1):100-102.
[10] 刘昕鹏,罗英伟,汪小林,等. 基于 XML 的 WebGIS 通信协议的设计与实现[J]. 软件学报,2004,15(6):899-907.
[11] 王天宇. 电能表及采集终端通信协议通用解析与生成算法研究[D]. 长沙:中南大学,2008.
[12] 张屹峰. DOM4j 解析技术探究[J]. 现代计算机,2011(15):39-42.
[13] 蔚晓娟,冉静,李爱华,等. 基于 DOM 的 XML 解析与应用[J]. 计算机技术与发展,2007,17(4):86-88.