

# 针对异构协议的动态解析器模型<sup>①</sup>

王永娜<sup>1,2</sup>, 赵奎<sup>2</sup>, 王鸿亮<sup>2</sup>, 王俊霖<sup>3</sup>

<sup>1</sup>(中国科学院大学, 北京 100049)

<sup>2</sup>(中国科学院 沈阳计算技术研究所, 沈阳 110168)

<sup>3</sup>(大连理工大学 软件学院, 大连 116024)

**摘要:** 介绍了异构协议报文格式和特点, 通过使用 XML 语言描述可扩展协议报文的方法, 以简单协议解析为基础, 将解析器功能模块化, 从而设计此异构协议动态解析器模型. 应用此模型实现不必重新编译, 只需进行少量代码修改, 即可完成异构协议报文自动模型匹配, 进而完成报文数据解析. 针对两种简单协议进行模型应用, 结果表明使用 XML 语言描述协议报文的方法及可重构的解析器能够有效地描述和解析异构协议报文, 获得可靠的解析结果.

**关键词:** 异构协议; 动态解析; XML; 状态机

## Model of Dynamic Parser for Heterogeneous Protocol

WANG Yong-Na<sup>1,2</sup>, ZHAO Kui<sup>2</sup>, WANG Hong-Liang<sup>2</sup>, WANG Jun-Lin<sup>3</sup>

<sup>1</sup>(University of Chinese Academy of Sciences, Beijing 100049, China)

<sup>2</sup>(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

<sup>3</sup>(School of Software Technology, Dalian University of Technology, Dalian 116024, China)

**Abstract:** This paper firstly describes the format and features of heterogeneous protocol packet. Then it describes the extended protocol packets by using the XML language, and modularizes the parser function to design a model of dynamic heterogeneous protocol parser based on the method of simple protocol analysis. Realizations applying this model can be achieved without recompiling, and only a small amount of code changes is needed to complete the model automatic matching of heterogeneous protocol packets, and then complete the analysis of packet data. We apply the model into two simple protocols, and the results show that the method of using XML to describe the protocol message and the reconfigurable parser can describe and analyze the heterogeneous protocol message effectively, and obtain reliable analysis result.

**Key words:** heterogeneous protocol; dynamic analysis; XML; state machine

## 1 引言

设备管控系统是通过数据通信进行设备采集信息及控制反馈信息的交互, 实现信息采集处理与远程设备管控, 需要按照相关协议格式与底层设备进行数据通信. 由于设备及通信信息种类繁多, 通信协议多样, 存在不同类型的协议帧格式<sup>[1]</sup>, 导致数据预处理的复杂度增大.

基于可扩展低耦合的设计思想, 此协议解析器模型采用具有自描述性语言 XML 表现协议格式与内容, 采纳设计模式思想, 设计出可根据解析对象格式通过

XML 语言自定义协议解析规则, 实现异构协议动态匹配, 使其可在不重新编译情况下, 通过 XML 配置文件和解析模块少量代码的修改, 实现将异构协议动态解析为标准协议. 模型独立性强, 可扩展性好, 可解决数据通信中异构协议数据的交互问题.

## 2 协议解析器模型架构设计

基本协议解析是识别数据报头部和提取数据字段以供后续数据处理的过程<sup>[2]</sup>. 协议解析器模型在此基础上进行扩展, 按照数据的流入、处理、流出, 可分为

① 基金项目: 国家水体污染控制与治理科技重大专项课题(2012ZX07505003)

收稿时间: 2016-04-26; 收到修改稿时间: 2016-06-21 [doi: 10.15888/j.cnki.csa.005533]

三大主要模块：异构数据接入层，译码层，标准数据交互层。

异构数据接入层与数据采集端进行交互，完成 UDP 异构数据报接收<sup>[3]</sup>。译码层进行数据报匹配及解析处理，详细设计分为决策组件和适配组件，决策组件是模型的检测机构，鉴别报文合法性；适配组件将通过检测的报文进行模式适配，匹配 XML 配置文件中定义的解析模板，完成报文解析。标准数据交互层负责将数据发送到管控系统后续的数据处理服务器做进一步的数据存库与推送。

动态协议解析器模型架构设计图如图 1 所示。

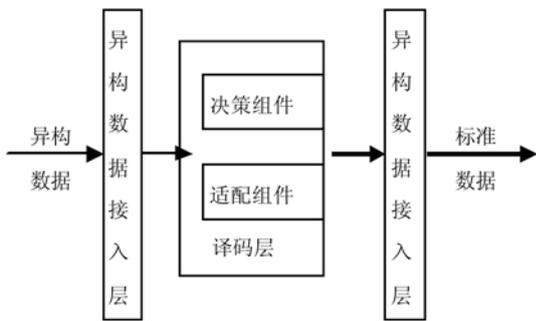


图 1 动态协议解析器模型架构

### 3 协议解析器模型组件详细设计

协议解析的基本流程是解析器读取数据报首部，验证数据报文合法性，通过读取识别信息，确定是否有解析模板可以匹配解析，适配相应解析模板进行异构协议数据的解析，最终将标准数据传送给数据处理服务器做进一步数据处理与数据转存、显示。功能组件详细设计如下。

#### 3.1 异构数据接入层

协议解析器模型要接入的基本信息单元是数据包，将数据包接收阶段定义为不同的状态，将接收新的数据作为外部触发条件，从而改变状态，此异构数据接入层可抽象表现为一个基于有限状态机的网络通信数据接收端<sup>[4]</sup>。基于有限状态机的状态模型图如下：

#### 3.2 译码层决策组件

译码层是整个模型核心部分，解析异构协议首先要判断协议内容是否合法，能否被适配组件解析，因此决策组件协议识别至关重要。协议解析器采用数据报边接收边解析的方式，基于有限状态机进行异构数据接入的同时，进行协议合法性决策判断。

状态机的两个主要组成部分是：事物所处状态、

引发状态变化的外部事件<sup>[5]</sup>。程序实现过程中定义不同数字编号表示不同状态，状态变量值的改变即可描述状态的变化。结合状态图，示例代码实现如下：

```

LONG ReciveInfo::OnCommunication(WPARAM data)
{
    receivedata = (BYTE)data;
    if(receive_state_machine == 0)
    {
        if(receivedata == 0x77) // 接收到起始标志
            receive_state_machine = 1;
        else
            receive_state_machine = 0; // 状态机复位
    }
    else if(receive_state_machine == 4)
    {
        if(receivedata == 0xAA) // 接收到校验位
            receive_state_machine = 5;
        else
            receive_state_machine = 0; // 状态机复位
    }
}

```

将接收数据划分为不用阶段，通过变量值表示状态，接收数据帧和数据监测结果作为状态机触发条件，从而改变状态，最终完成一个数据包接收和校验。

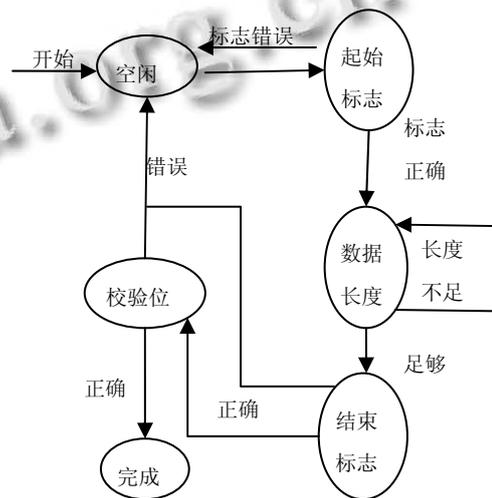


图 2 异构数据接入层状态图

#### 3.3 译码层适配组件

动态协议解析器模型中适配组件的设计是通过分析异构协议变化特点，设计可扩展协议格式，采用具

有自描述性及可扩展性的 XML 配置文档进行封装, 反映通信协议的格式和内容. 经决策组件校验的协议可被适配组件按解析模板进行动态解析. 协议的改变只需要动态修改 XML 描述文件同时符合 XML Schema 验证文档即可. 根据 XML 协议配置文件适配协议解析模板, 通过动态加载 XML 协议数据帧得出字段的物理含义.

解析器模型框架详细组件设计如图 3 所示.

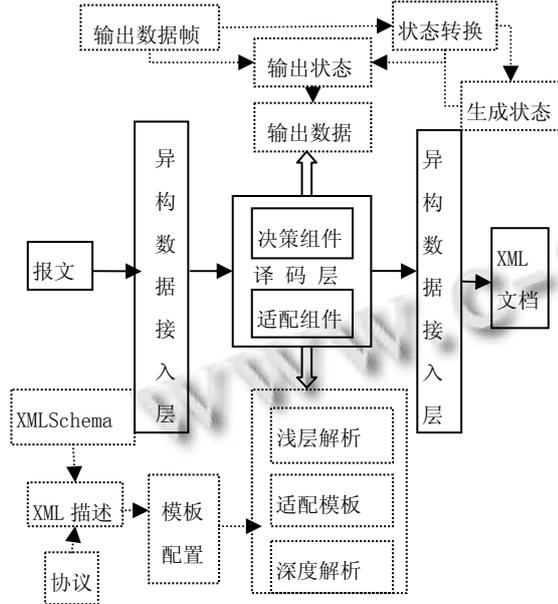


图 3 异构协议动态解析器模型

## 4 动态解析器模型中XML的使用设计

### 4.1 报文的 XML 文件描述

网络通信协议需要用协议帧描述传输信息, 通常采用硬编码方式严格定义协议格式, 针对每个协议格式编写固定代码实现协议解析, 此方式不具通用性, 不便于扩展, 而基于 XML 配置文档的解析方式使得异构协议可扩展, 在软件解析中应用较多<sup>[6]</sup>.

使用 XML 描述协议表现协议字段在数据帧中的位置及所占的字节数等抽象化信息, 可以定义统一的数据传输格式, 协议内容及形式的改变, 只修改 XML 文档中元素的属性值即可, 此文件是解析器浅层解析的模板.

### 4.2 XML 解析模板配置文件

```
<?XML version="1.0" encoding="utf-8"?>
```

```
<Template>
<TypeInfo identifier="X" typeName="X_N">
  <Item frameType="client" temp="CLT">
  <Item frameType="plute" temp="PLT">
</TypeInfo>
```

```
</Template>
</XML>
```

以上配置文件代码描述了帧标识与解析模板号的关系, 协议解析程序对数据帧进行匹配时会识别出帧标识, 与 XML 配置文件进行模板匹配, 最终通过 temp 属性值确定数据帧解析模板号, 完成数据解析.

### 4.3 XML Schema 协议验证文档

为改善硬编码的不可扩展性, XML 协议描述文件也要摒弃严格定义协议格式的方式, 将数据升华为 Item 元素抽象间接描述, 定义一个 Schema 文档规范 XML 配置文件架构<sup>[7]</sup>. XML Schema 示例代码如下:

```
<xml version="1.0"?>
<xsd:schema
xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name=" DataInfo_m ">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="Item" type="xsd:string" />
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
</xml>
```

### 4.4 XML 解析库的使用

常见 XML 解析器有 tinyxml、Xerces、pugixml、libxml 等, 通过对比不同 XML 解析器对编程语言及开发环境等的支持<sup>[8]</sup>, 选择 Xerces-C++应用于本模型.

Xerces-C++ 主要组件是扫描程序, 本解析器模型需要进行验证因此选择 SGXMLScanner<sup>[9]</sup>. 根据线程模型, 主线程调用 Initialize(), 创建其他用于解析的线程, 最后调用 Terminate(). 初始化期间创建解析器池, 可以在运行时使用这些解析器, 从而避免构造和分解解析器的开销.

## 5 测试及结果分析

### 5.1 实验设计

以设备管控系统为背景, 项目中数据网关需要和底层仪器进行通信采集信息, 选取以下两种异构协议报文进行测试.

包头	数据长	设备标识	设备编码	温度	湿度	烟尘	CRC	包尾	
包头	数据长	命令编码	NO	NO <sub>2</sub>	SO <sub>2</sub>	PM2.5	PM10	CRC	包尾

图 4 异构协议格式定义

XML 协议描述示例代码如下:

```
<?xml version="1.0" encoding="utf-8"?>
```

```

<Protocol>
  <DataInfo_m>
    <Data_m ID=" plute_id" type="
plute_type">
      <Item length="1" name="dust">
        </Data_m>
      </DataInfo_m>
    </Protocol>
  </xml>

```

将具体元素抽象为 Item 元素,可承载不同协议内容,避免了固定协议格式的弊端.测试表明:基于 XML 文件的协议描述方式可以很好地描述异构协议,且修改与扩展方便可行.

通过使用编程接口及线程池、网络通信等实现协议数据接收及解析推送.测试程序解析类工厂模式类图设计如图 5.

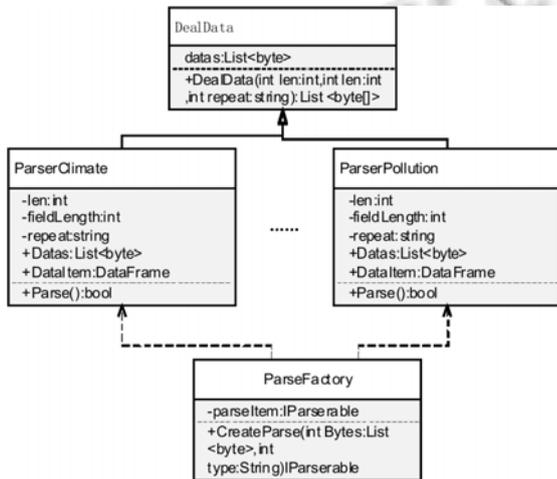


图 5 解析类工厂模式类图

### 5.2 测试结果

测试程序采用多线程编程模拟数据采集器控制数据数量,从而测试解析不同数量数据帧所需要的时间.

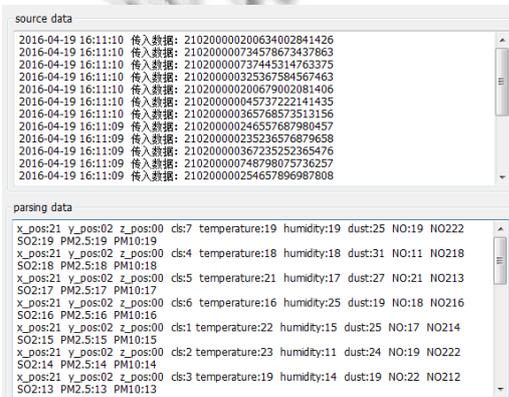


图 6 协议解析器数据接入及解析界面

上图为协议解析器模型测试程序数据接收层接收的数据及经过译码层后解析出的数据演示界面.数据帧解析时间采用软件读取定时计数器值,将不同时刻读取的值进行差值计算得出时间测试值.

结果表 1 所示.

表 1 解析时间记录表

数据帧数量(条)	解析时间(s)
20	0.4527577
200	1.5365547
2000	9.7698875

从测试可以看出:数据帧解析时间为 ms 级别,且从整体看协议解析所需时间未表现出随数据帧数量的增加同比上升,以上数据仅表现出此实验的测试结果,实际应用中结合调度器线程合理分配,进一步改进解析模块程序或将得到更好的时间性能指标.

### 6 结语

该解析器模型使用 XML 描述协议格式与内容,层次关系清晰,协议发生变化或新增协议时,只需适当修改 XML 协议描述文件,或增加新的解析模块,即可实现对数据帧的解析,充分体现了设计考虑的可重构性与可扩展性<sup>[10]</sup>.此模型可靠性强,数据转换效率高,能够对数据格式和映射关系实现灵活定制,可以应用到多种类型设备及工业控制,实现异构数据转换.

### 参考文献

- 1 俞晓华,于凤芹,钦道理.一种能耗数据采集协议描述及生成解析方法.计算机技术与发展,2015.
- 2 Gibb G. Design Principles for Packet Parsers. Stanford University Microsoft Research, 2015.
- 3 王智莉,卜方玲.异构感知数据的动态适配接入方法.传感器与微系统,2015.
- 4 潘准洋,刘彩霞,刘树新.基于有限状态机的网络协议状态机制检测方法.计算机应用研究,2015.
- 5 李莹,贾彬.一种基于状态机的串口通信协议的设计与实现.电子设计工程,2012.
- 6 齐建业,余祥,刘峻宇,李强.协议一致性测试数据包的构造与解析.西南科技大学学报,2013.
- 7 Thaw TZ, Khin MM. Measuring qualities of XML schema documents. Journal of Software Engineering and Applications, 2013.
- 8 <http://lars.ruoff.free.fr/xmlcpp/>.
- 9 <http://xerces.apache.org/xerces-c/>.
- 10 梁彦杰,廉东本.基于消息中间件的数据交换平台传输框架设计.计算机系统应用,2012,21(4):10-13.