一种可扩展的协议分析系统的设计与实现

Design and Implementation of Expansible Protocol Analysis System

崔璐璐　王荣良　（华东师范大学 计算机科学与技术系 上海 200062）

摘 要：本文针对现有协议分析系统价格昂贵，操作复杂，不易扩展等缺点，提出使用独立于计算机架构和编程语言的方法来描述结构化数据，设计并实现了具有良好体系结构的可扩展协议分析系统（以下简称 EPAS），介绍了加速系统运行速度的技术要点，最后通过具体例子论述了 EPAS 的应用前景。

关键词：协议分析　协议描述脚本　协议树

1　引言

协议分析（也称报文分析）是指通过程序分析数据包的协议头和尾，解码并表示出协议用来交互信息的消息格式，从而了解信息和相关的数据包在产生和传输过程中的行为。包含该程序的软件和设备就是协议分析器[1]。协议分析器用于很多系统，具有很好的应用前景。例如网络协议分析系统是监测网络状态的有效工具，给网络管理员提供不同的网络统计数据、网络状况信息和网络出错信息。同样地，通过对捕获到的数据进行解码，获得数据包所承载协议的详细信息还可以为软件开发和协议开发的专业人员提供必要的资料。计算机网络教学体系当中的一个重要环节就是网络协议分析，利用协议分析器可以使学生对抽象的网络知识有直观生动的认识和理解。

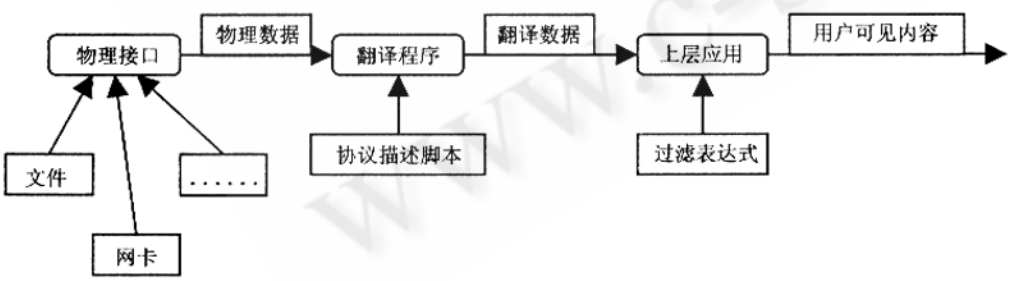

图 1

大多数协议分析器的主要局限是能够分析的协议数量有限，最多几百种，而且要加入新的解析器必须由专门的程序员编写代码。本文针对现存的协议分析系统的这些局限性进行了改进，使系统具有易扩展性。此外，本系统还支持对除网络协议以外的其他协议的分析，如串口协议。

2　EPAS 的框架及技术要点

2.1　EPAS 框架（见图 1）

物理接口模块：主要负责将从不同的数据源（比如说文件、网卡等）得到的数据转换成统一格式的物理数据。这样做的好处是能够提高程序的适用范围，使这个系统又有多输入的特性。物理数据是一段字节流，流的格式如图 2。

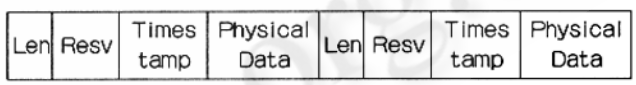

图 2

流中每一个从 len 到 Physical Data 的块称为一个物理数据段，一个流中可能会有很多个物理数据段。流有一个入口和一个出口，新到达的数据从流的入口进入，数据分析时从出口流出。

翻译程序模块：这个模块的主要任务是对从物理接口处获得的物理数据，按照协议描述文件所规定的协议格式和语义进行翻译，将物理数据转换成一棵协议树和一系列的键值对，也即翻译数据，来表示这个物理数据帧中包含的协议以及各个字段的含义。翻译程序模块的结构如图 3。

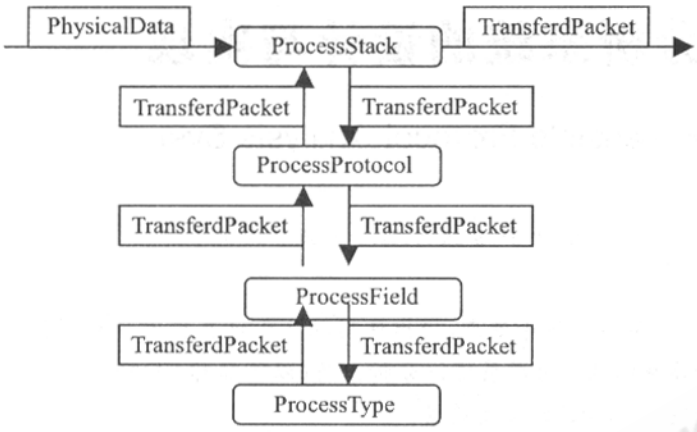

图 3

有一个上层应用模块。在这个模块中，可以利用翻译程序提供的翻译数据，直接取出有用的字段进行处理。

2.2 协议描述语言的定义

协议描述语言是用来规定所要分析的物理数据的结构和含义的文件，它提供了一套正式、无歧义和精确的规则以描述独立于特定计算机硬件的对象结构。本系统使用可扩展标识语言的文件格式，因为它具有良好的结构性以及可扩展性，很多工具都提供支持，这样便于程序的实现和扩展。可扩展标识语言文件主要定义这样几个内容：数据类型的定义、字段的定义、协议的定义、协议栈的定义。他们之间的关系是：协议栈是由若干个协议组成的一棵树；协议是由若干个字段组成；每一个字段对应一种数据类型。协议描述文件可以由用户自己编写，以灵活的扩展出用户自己定义的协议。EPAS 还为用户提供了可视化的脚本编辑器，以方便用户对该文件的编写。文件的基本格式参见图 4。

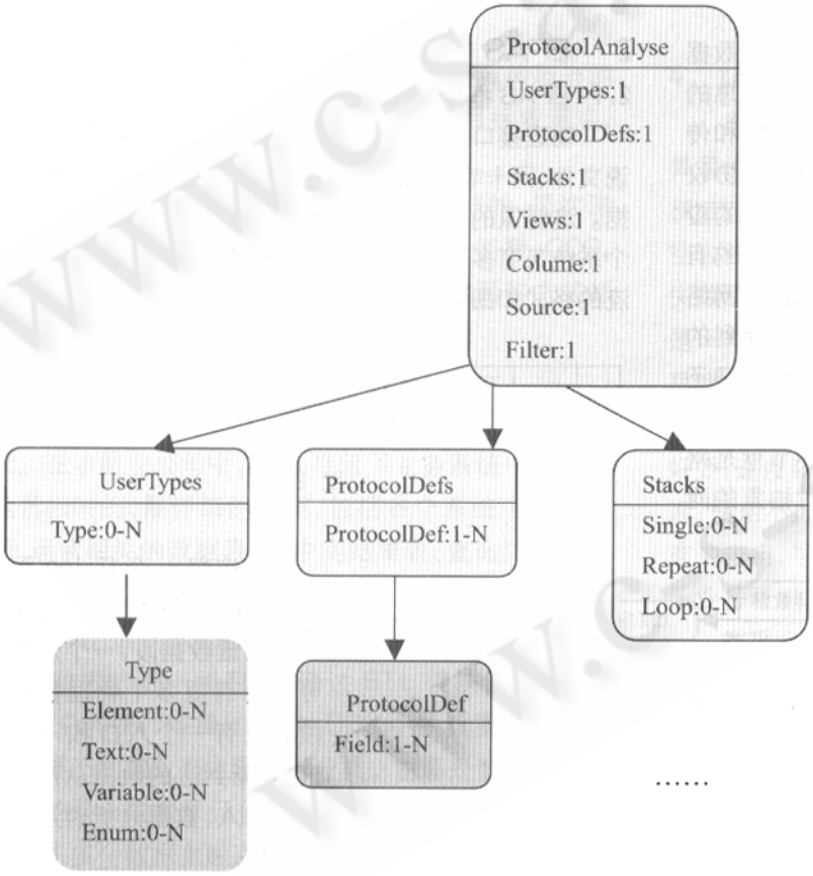

图 4

2.3 表达式预编译技术

协议描述脚本中大量使用了各种表达式，对表达式的优化能够有效的提高翻译程序模块的效率。在 EPAS 中，对表达式的求值分三步进行。

第一步是表达式预编译：当一个表达式第一次被执行到的时候，先将字符串形式的表达式编译成一种类似于后缀表达式的中间代码，存储在一个全局的散列表中。以后该表达式再被执行到时，直接从散列表中取出编译以后的中间代码执行。

中间代码的每一个单元都是一个四元组。

```
public class ExpWord
{
```

翻译数据包括以下几个部分：原始的物理数据；存放变量和相应值的哈希表；分析过的协议组成的一棵树；每一个协议中的各个字段的信息。上层应用模块可以利用这些信息来完成各种指定的应用。

上层应用模块：EPAS 应用广泛，对每一种应用都

public string Name; //该元组对应的表达式中原始的字符串,如果是变量则为变量名

public ExpType Type; //该元组的类型,有常量,变量,操作符等类型

public ExpOperator Opr; //如果单元为操作符,则该元组表示具体的操作类型。(如+ - * /等)

public string Value; //如果该单元为变量或常量,则该元组代表具体的值。

}

例如:1+a/2 经过编译以后变成1,a,2,/,+这五个单元。

第二步是变量替换:如果表达式中含有变量单元,则必须在这一步中把所有的变量单元替换成相应的常量。在 EPAS 中,变量和常量的对应关系被记录在已翻译数据中。随着翻译过程的进行,变量和常量的对应关系也被不断的添加进来。

在上面提到的那个例子中,1,a,2,/,+这五个单元经过变量替换后成为1,4,2,/,+。

第三步是求值:对经过预编译及变量替换处理后的中间代码,利用一个数值栈即可以解释执行了。对于上面提到的那个例子,解释后的结果是2。

2.4 正则表达式匹配模块的设计与实现

现存的网络协议中有很多都是基于文本的,比如说 HTTP, FTP 等。为了能够解析这类协议,EPAS 引入了正则表达式字段类型。即将一个字段的值定义为在协议 buffer 中能够与一个指定的正则表达式匹配的内容。借助.net Framework 中提供的 RegExp 对象,我们能够很方便的实现这一想法。比如说为了找出 HTTP GET HEAD 中的 URL 字段,可以使用这样的表达式 "GET (/[^]*) HTTP 1.1\r\n"。正则表达式语法可以支持表1中的符号:

表 1

符号	说明
\	转意符,可以与某一字符结合表示一个特定的意义
*	匹配 0 次或多次
+	匹配 1 次或多次
?	匹配 0 次或 1 次
()	需要捕获的子表达式
x|y	匹配 x 或者 y
[xyz]	匹配字符集
[a-z]	匹配范围字符集
\d	数字字符匹配。等效于[0-9]。
\D	非数字字符匹配。等效于 [^0-9]。
\f	换页符匹配。等效于 \x0c 和 \cL。
\n	换行符匹配。等效于 \x0a 和 \cJ。
\r	匹配一个回车符。等效于 \x0d 和 \cM。
\s	匹配任何空白字符,包括空格、制表符、换页符等。与 [\f\n\r\t\v] 等效。
\S	匹配任何非空白字符。与 [^ \f\n\r\t\v] 等效。
\t	制表符匹配。与 \x09 和 \cI 等效。
\v	垂直制表符匹配。与 \x0b 和 \cK 等效。
\w	匹配任何字类字符,包括下划线。与"[A-Za-z0-9_]"等效。
\W	与任何非单词字符匹配。与"[^A-Za-z0-9_]"等效。

因此利用正则表达式匹配机制,EPAS 能够比其他的一些协议分析系统更好的理解文本协议。

3 EPAS 的技术参数与现有系统的比较

表 2

系统	EPAS	Ethereal	专业系统
价格	无	无	昂贵
支持现有协议	多	多	单一
可扩展性	可扩展	不易扩展	无
分析速度	1000pps	1000pps	与系统相关
平台相关性	单一	无	单一
使用方便性	使用很方便	使用很方便	使用复杂

4 EPAS 的应用举例

4.1 协议研发

由于网络上协议种类繁多,新的协议层出不穷,EPAS 的可扩展性体系结构恰能适应网络发展的需要,不断加入新的协议解析器。下面例举对新开发的 MRTP 协议的解析结果(图5)。

4.2 网络教学

以 EPAS 为基础结合工具 NISTNET[2] 可开发出一套新颖的计算机网络教学方法。能够非常清晰地呈现出各种情况下的网络活动。这使学生得到实践的机会,同时

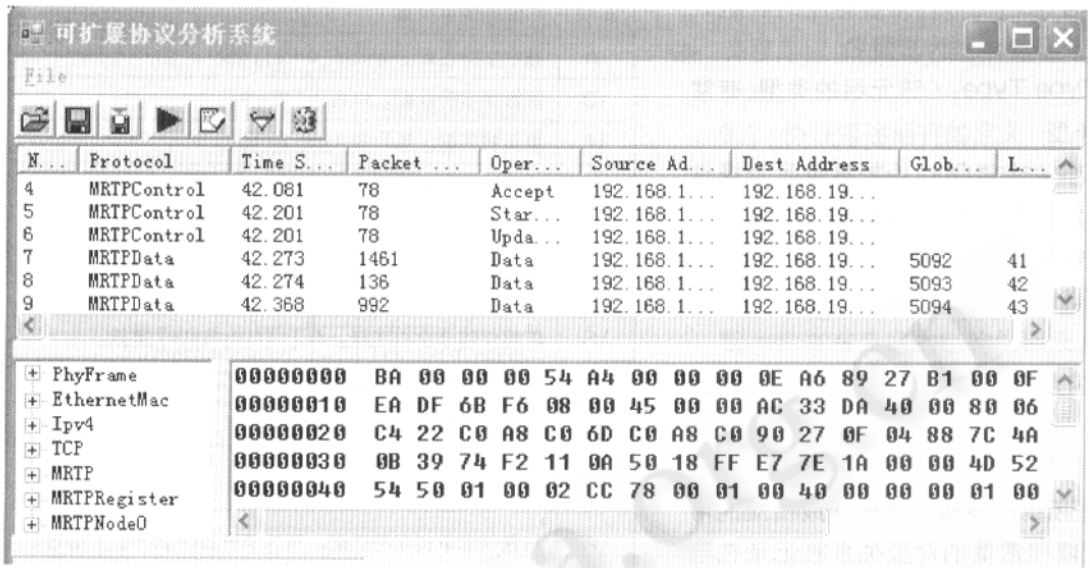

图 5

的操作系统上的网络管理系统的开发平台，并非本地化的界面，使用起来很不方便，使一般的中小企业难以接受。另外，国家明确规定禁止重要部门使用外国的安全产品。所以说，开发协议分析系统有很好的应用前景。

在不需要专用昂贵的网络设备的实验室实验，使学生通过对网络数据包的跟踪、分析和尝试，了解计算机网络技术的基本概念和掌握计算机网络实际技能[3]。

4.3 入侵检测

入侵检测系统需要辨别数据包的协议类型，以便使用相应的数据分析程序来检测数据包。在网络通信中，网络协议定义了标准的，层次化的，结构化的网络数据包[4,5]，利用这种层次性对网络协议逐层分析，取代需要重复进行直到匹配成功的模式匹配方式，可以使所需的计算量大幅度减少，提高分析效率，得到更准确的检测结果[6]。

4.4 网络监控

协议分析系统是了解网络细节的必要手段，是监测网络状态的有效工具[7]。例如通过将数据包 MAC 地址字段的值与用于网元发现的表相对照，若不匹配说明网段上有新的主机加入。这样就完成了新的网络网元的发现。

5 结束语

现在已有可以支持上百种协议的 Network Associate Sniffer Pro，具有多平台性和应用全面性的 OpenView 网络节点管理器和主要针对 IBM 自己的操作系统 AIX 的 TME10 /NetView [8]。但是，这几家大公司的产品或严重依赖相关平台，或价格非常高昂，主要是对其各自

参考文献

1 Jaition. 协议分析和分析器[EB/OL]http://blog. chinaitlab. com/user1/326713/archives/2006/39995.html,2006－01－04/2006－09.

2 Jeanna Neefe Matthews. Hands－on Approach to Teaching Computer NetworkingUsing Packet Traces[DB/OL]. The ACM Digital Library,2005－10/2006－09.

3 Richard Blum. 网络性能开源工具包[M]. 北京：清华大学出版社.2005 年 6 月.

4 Douglas E.Comer 著 林瑶 蒋慧 杜蔚轩等译 谢希仁审校.用 TCP/IP 进行网际互联 第一卷：原理，协议与结构[M].北京.电子工业出版社.2004.

5 Andrew S.Tanenbaum 著 潘爱民译 徐明伟审.计算机网络(第四版)[M].北京.清华大学出版社.2004.290－402.

6 朱红莉 张浩军 程立.基于网络协议分析的入侵检测系统[DB/OL].中国期刊全文数据库.2003/2006－09.

7 陈伯成，范闽，李英杰 利用网络监听维护子网系统安全的一种方法[J]计算机工程与应用.2000 .36(10).133 － 1351.

8 刘芳.网络协议分析实现过程探讨[DB/OL].中国期刊全文数据库.2002/2006－09 .