本发明实施例公开了一种应用协议报文的自动化解析方法和装置。其中，该方法包括：根据应用协议报文的协议特征，添加利用预设脚本语言编辑的具有逻辑性的模板文件，编译各模板文件生成对应的各模板解析类，利用与目标应用协议报文匹配的模板解析类对目标应用协议报文进行解析，最终输出解析结果。本发明实施例在原始应用协议报文的格式发生变化，或需要解析一种新格式的应用协议报文，或需要多提取或删除~~字段内容~~时，无需修改代码或重新发布软件版本，仅通过添加、修改或删除模板文件，即可实现对现有的和新增加的各类应用协议报文的解析，同时实现了应用协议解析过程的可维护性和可扩展性。

下图：



1、一种应用协议报文的自动化解析方法，其特征在于，包括：

根据应用协议报文的协议特征添加模板文件，其中，所述模板文件是利用预设脚本语言编辑的具有逻辑性的报文解析模板；

对各模板文件进行编译处理，生成与所述各模板文件相对应的各模板解析类；

利用与目标应用协议报文匹配的模板解析类对所述目标应用协议报文进行解析，并输出解析结果。

2、根据权利要求1所述的方法，其特征在于，所述模板解析类包括：模板匹配单元、模板校验单元、字段内容提取单元和数据输出单元。

3、根据权利要求2所述的方法，其特征在于，利用与目标应用协议报文匹配的模板解析类对所述目标应用协议报文进行解析，并输出解析结果，包括：

对所述目标应用协议报文的头部进行解析，得到HOST、URL和METHOD三元组；

将解析出的所述HOST、URL和METHOD三元组与各模板匹配单元中的匹配信息进行模板匹配；

如果存在与所述HOST、URL和METHOD三元组匹配的第一模板解析类，获取所述目标应用协议报文中HEAD和BODY的指定内容；

将所述指定内容与所述第一模板解析类的模板校验单元中的校验信息进行匹配校验，利用校验成功的第一模板解析类解析所述目标应用协议报文，并输出解析结果。

4、根据权利要求3所述的方法，其特征在于，利用校验成功的第一模板解析类解析所述目标应用协议报文，并输出解析结果，包括：

利用所述校验成功的第一模板解析类的字段内容提取单元，调用指定解码接口对所述目标应用协议报文进行字段内容提取；

按照所述第一模板解析类的数据输出单元中的预设格式，输出提取的字段内容。

5、根据权利要求1所述的方法，其特征在于，对各模板文件进行编译处理，生成与所述各模板文件相对应的各模板解析类，包括：

利用模板编译器对所述各模板文件进行解析，提取与各模板文件相对应的各字段描述信息集合；

利用所述各字段描述信息集合构造各模板解析类。

6、一种应用协议报文的自动化解析装置，其特征在于，包括：

模板文件添加模块，用于根据应用协议报文的协议特征添加模板文件，其中，所述模板文件是利用预设脚本语言编辑的具有逻辑性的报文解析模板；

模板解析类生成模块，用于对各模板文件进行编译处理，生成与所述各模板文件相对应的各模板解析类；

目标应用协议报文解析模块，用于利用与目标应用协议报文匹配的模板解析类对所述目标应用协议报文进行解析，并输出解析结果。

7、根据权利要求6所述的装置，其特征在于，所述模板解析类包括：模板匹配单元、模板校验单元、字段内容提取单元和数据输出单元。

8、根据权利要求7所述的装置，其特征在于，所述目标应用协议报文解析模块，包括：

三元组解析子模块，用于对所述目标应用协议报文的头部进行解析，得到HOST、URL和METHOD三元组；

三元组匹配子模块，用于将解析出的所述HOST、URL和METHOD三元组与各模板匹配单元中的匹配信息进行模板匹配；

指定内容获取子模块，用于如果存在与所述HOST、URL和METHOD三元组匹配的第一模板解析类，获取所述目标应用协议报文中HEAD和BODY的指定内容；

匹配校验子模块，用于将所述指定内容与所述第一模板解析类的模板校验单元中的校验信息进行匹配校验；

目标应用协议报文解析子模块，用于利用校验成功的第一模板解析类解析所述目标应用协议报文，并输出解析结果；

其中，所述目标应用协议报文解析子模块具体用于：

利用所述校验成功的第一模板解析类的字段内容提取单元，调用指定解码接口对所述目标应用协议报文进行字段内容提取；

按照所述第一模板解析类的数据输出单元中的预设格式，输出提取的字段内容。

9、一种服务器，其特征在于，包括：

一个或多个处理器；

存储装置，用于存储一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求1-5中任一所述的应用协议报文的自动化解析方法。

10、一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时实现如权利要求1-5中任一所述的应用协议报文的自动化解析方法。

**一种应用协议报文的自动化解析方法和装置**

**技术领域**

本发明实施例涉及网络内容审计领域，尤其涉及一种应用协议报文的自动化解析方法和装置。

**背景技术**

在现有网络应用协议种类繁多，协议版本更新换代频繁的情况下，方便快捷的解析各类应用协议的具体内容，对网络安全监控等有着非常重要的作用。

现有的应用协议内容解析方法主要是基于数据包搜索的，该方法是按顺序分析每个应用协议的数据包，从而分析出相应的应用协议，并从应用协议报文中分析出登陆信息，发帖信息、回帖信息、购物订单信息以及网络购票信息等内容。此外，也有将配置文件形式的模板文件和硬编码函数相互配合共同提取应用协议内容的方法。

上述基于数据包搜索的方法由于效率低，工作量繁重等问题，仅适用于网络流量小的情况；将配置文件形式的模板文件和硬编码函数相互配合的方法，需要利用硬编码函数提取应用协议的BODY部分的字段信息，而当原始应用协议格式发生变化，或者需要提取一种新格式的应用协议数据时，以及需要多提取或删除提取字段时，都需要修改、增加对应的硬编码函数，导致应用协议解析过程缺乏可维护性和可扩展性。

**发明内容**

本发明提供一种应用协议报文的自动化解析方法和装置，以实现应用协议解析过程的可维护性和可扩展性。

第一方面，本发明实施例提供了一种应用协议报文的自动化解析方法，该方法包括：

根据应用协议报文的协议特征添加模板文件，其中，所述模板文件是利用预设脚本语言编辑的具有逻辑性的报文解析模板；

对各模板文件进行编译处理，生成与所述各模板文件相对应的各模板解析类；

利用与目标应用协议报文匹配的模板解析类对所述目标应用协议报文进行解析，并输出解析结果。

第二方面，本发明实施例还提供了一种应用协议报文的自动化解析装置，该装置包括：

模板文件添加模块，用于根据应用协议报文的协议特征添加模板文件，其中，所述模板文件是利用预设脚本语言编辑的具有逻辑性的报文解析模板；

模板解析类生成模块，用于对各模板文件进行编译处理，生成与所述各模板文件相对应的各模板解析类；

目标应用协议报文解析模块，用于利用与目标应用协议报文匹配的模板解析类对所述目标应用协议报文进行解析，并输出解析结果。

第三方面，本发明实施例还提供了一种服务器，包括：

一个或多个处理器；

存储装置，用于存储一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如上所述的应用协议报文的自动化解析方法。

第四方面，本发明实施例还提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如上所述的应用协议报文的自动化解析方法。

本发明实施例通过利用应用协议报文的协议特征，添加利用预设脚本语言编辑的具有逻辑性的模板文件，并编译模板文件生成模板解析类，利用与目标应用协议报文匹配的模板解析类对目标应用协议报文进行解析，最终输出解析结果，解决了当原始应用协议报文的格式发生变化，或需要提取一种新格式的应用协议数据，或需要多提取或删除提取字段时，需要重新编码的问题，进而实现了应用协议解析过程的可维护性和可扩展性。

**附图说明**

图1是本发明实施例一中的应用协议报文的自动化解析方法的流程图；

图2是本发明实施例二中的应用协议报文的自动化解析方法的流程图；

图3是本发明实施例三中的应用协议报文的自动化解析方法的流程图；

图4是本发明实施例四中的应用协议报文的自动化解析装置的结构示意图；

图5是本发明实施例五中的服务器的结构示意图。

**具体实施方式**

下面结合附图和实施例对本发明作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释本发明，而非对本发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与本发明相关的部分而非全部结构。

实施例一

图1为本发明实施例一提供的应用协议报文的自动化解析方法的流程图，本实施例可适用于需要对应用协议报文的内容进行解析的情况，该方法可以由应用协议报文的自动化解析装置来执行，该装置可以采用软件和/或硬件的方式实现。如图1所述，该方法具体包括如下步骤：

S110、根据应用协议报文的协议特征添加模板文件，其中，模板文件是利用预设脚本语言编辑的具有逻辑性的报文解析模板。

本实施例中，应用协议报文指的是基于TCP/IP协议体系的互联网应用所产生的网络通信报文，例如可以是HTTP（超文本传输协议，HyperText Transfer Protocol）、FTP（文件传输协议，File Transfer Protocol）、NFS（网络文件系统，Network File System）或DNS（域名系统，Domain Name System）协议报文等。每个应用协议报文可以只包括HEAD部分，也可以包括HEAD和BODY两部分，每个应用协议报文都有其独有的协议特征，这些独有的协议特征都可以通过HEAD部分和BODY部分体现出来。以HTTP应用协议报文为例，该协议下的应用协议报文按照互联网应用类别的不同可以划分成不同的应用协议大类，其中，互联网应用类别可以是购物类、贴吧类、旅游类和地图类等。每一种应用协议大类下还可以包括各个具体的互联网应用协议，其中，各个具体的互联网应用根据所属的互联网应用类别的不同而有所不同，例如购物类的互联网应用类别所包括的具体的互联网应用可以是京东、淘宝、亚马逊、一号店、苏宁和国美等。各个具体的互联网应用对应的应用协议报文的HEAD部分可以包括标准部分和扩展部分，其中，HEAD的标准部分的格式是固定的，其遵循HTTP协议标准，而扩展部分的格式是不固定的，其由各网络应用自行定义；此外，各应用协议对应的报文的BODY部分的格式也是不固定的，并且还可能存在加密和转码等情况。这些不同格式的互联网应用协议报文都会包含某些特定的信息集合，这些特定的信息集合均对应于所属协议大类中可以提取到的字段内容集合。

本实施例中，模板文件是利用预设的脚本语言编辑的具有逻辑性的报文解析模板，其中包括用于解析与之相对应的应用协议报文的解析方式，。每个模板文件都可以根据与之相对应的应用协议报文的协议特征预先添加，每个模板文件分别与一种应用协议报文的协议特征相对应，即各个模板文件互不相同，每一个模板文件都是根据相对应的应用协议报文的协议特征编辑的。

以遵循HTTP应用协议的网络购物类的应用协议报文，且该类应用协议报文均包含HEAD和BODY两部分为例进行具体说明：

网络购物可以包括建立订单、提交订单、查询订单等操作类型，每一个操作类型都对应不同的应用协议报文，每一个应用协议报文都包含HEAD部分和BODY部分。其中，从HEAD部分可以识别出网络购物类协议特征，从HEAD或者BODY中可以识别出其操作类型。

本实施例中，为了解析出在网络购物过程中，用户具体进行的是何种操作类型，可以根据已知的与各个操作类型相对应的应用协议报文的协议特征来预先添加模板文件，例如，建立订单、提交订单、查询订单等均可以对应一个模板文件。模板文件中包括用于解析应用协议报文的解析方式，其中，解析方式根据HEAD部分和BODY部分的特性来确定，优选可以包括字段内容提取的位置、字段内容提取的方法以及对提取字段内容的处理方式等，其可以利用预先设定的具有逻辑性的脚本语言来表示。

S120、对各模板文件进行编译处理，生成与各模板文件相对应的各模板解析类。

本实施例中，由于各模板文件是利用脚本语言来描述应用协议报文的解析过程的，且与各应用协议报文相对应的各模板文件是预先添加的，在模板文件数据库中存在多个模板文件。因此在实际应用中，无法直接获知具体某个模板文件与要解析的目标应用协议报文是相对应的，即使获取到了与目标应用协议报文相对应的模板文件，也无法直接利用模板文件对目标应用协议报文进行解析。因此，需要对模板文件做进一步的处理，以使目标应用协议报文得到解析。

本实例中，由于用于描述应用协议报文解析过程的模板文件是由预先设定的具有逻辑性的脚本语言表示的，为了使各模板文件能够被计算机识别并加以利用，需要对各模板文件进行编译处理，得到与各模板文件相对应的各模板解析类，生成计算机程序。其中，每个模板解析类对应一个模板文件，利用模板解析类即可实现对目标应用协议报文HEAD部分和BODY部分的解析。

S130、利用与目标应用协议报文匹配的模板解析类对目标应用协议报文进行解析，并输出解析结果。

本实施例中，在获取到与各模板文件相对应的模板解析类后，即可利用各模板解析类与目标应用协议报文进行一一匹配，来确定与目标应用协议报文相匹配的模板解析类。当确定了与目标应用协议报文相匹配的模板解析类后，即可利用该模板解析类对与目标应用协议报文进行解析，并输出解析结果。

需要注意的是，为了适应各应用协议版本更新换代日趋频繁的现象，模板文件数据库中的模板文件是可以随时添加、修改或删除的。当产生新的应用协议时，可以根据该新的应用协议的协议特征添加新的模板文件；当原有的应用协议被淘汰时，为了节省存储空间，可以删除与该应用协议相对应的模板文件。这种利用模板文件进行应用协议报文解析的方法，无需修改代码或重新发布软件版本，仅通过添加、修改或删除模板文件，即可实现对各类应用协议报文的解析。

本实施例提供的应用协议报文的自动化解析方法，通过利用应用协议报文的协议特征，添加利用预设脚本语言编辑的具有逻辑性的模板文件，并编译模板文件生成模板解析类，利用与目标应用协议报文匹配的模板解析类对目标应用协议报文进行解析，最终输出解析结果。这种利用模板文件进行应用协议报文解析的方法，解决了当原始应用协议报文的格式发生变化，或需要提取一种新格式的应用协议数据，或需要多提取或删除提取字段时，无需修改代码或重新发布软件版本，仅通过添加、修改或删除模板文件，即可实现对各类应用协议报文的解析，同时实现了应用协议解析过程的可维护性性和可扩展性。

在上述实施例的基础上，

进一步的，对各模板文件进行编译处理，生成与各模板文件相对应的各模板解析类，包括：

利用模板编译器对各模板文件进行解析，提取与各模板文件相对应的各字段描述信息集合；

利用各字段描述信息集合构造各模板解析类。

本实施例中，根据模板文件中的预设脚本语言的词性、语法特征，设计词法规则文件、语法规则文件，并结合FLEX词法分析器和BISON语法分析器生成 自定义的脚本语言编译器，利用该编译器对各模板文件进行编译。本实施例中，各模板文件根据各应用协议报文的协议特征添加，其中记录了相对应的应用协议报文的具体解析方式，以及待提取的协议字段的具体位置，因此，可以利用编译器提取出与各模板文件相对应的各字段描述信息集合，并利用预设方式将各字段描述信息集合构造为与其相对应的各模板解析类。

实施例二

本实施例在实施例一的基础上，提供了步骤S130的优选实施方式，实施例一中的模板解析类包括：模板匹配单元、模板校验单元、字段内容提取单元和数据输出单元。图2是本发明实施例二提供的应用协议报文的自动化解析方法的流程图，如图2所示，该方法包括：

S210、根据应用协议报文的协议特征添加模板文件，其中，模板文件是利用预设脚本语言编辑的具有逻辑性的报文解析模板。

S220、对各模板文件进行编译处理，生成与各模板文件相对应的各模板解析类。

本实施例中，每个模板解析类都记录了相对应的应用协议报文的具体解析方式，以及待提取的协议字段的具体位置。根据应用协议报文的具体解析方式，每个模板解析类都可以包括模板匹配单元、模板校验单元、字段内容提取单元和数据输出单元。其中，模板匹配单元用于对目标应用协议报文的HEAD的标准部分进行匹配，以初步确定选定的模板解析类与目标应用协议是否匹配；模板校验单元用于对目标应用协议报文的HEAD的扩展部分和BODY部分进行校验，再次确定由模板匹配单元选定的模板解析类与目标应用协议确实匹配；字段内容提取单元用于利用模板解析类对目标应用协议进行字段内容提取；数据输出单元用于将提取到的字段按预设形式输出显示。

S230、对目标应用协议报文的头部进行解析，得到HOST、URL和METHOD三元组。

本实施例中，要实现利用模板解析类对目标应用协议报文进行解析，需要确定具体哪一个模板解析类能够对该目标应用协议报文进行解析。由于应用协议报文的HEAD的标准部分包括该应用协议报文遵循的标准协议内容，因此，可以利用模板解析类与目标应用协议报文的HEAD的标准部分进行匹配，以初步确定哪一类模板解析类可以对目标应用协议报文进行解析。

本实施例中，在获取到目标应用协议后，可以先对目标应用协议报文的HEAD部分进行三元组解析，得到用于与模板解析类进行匹配的HOST、URL（统一资源定位符，Uniform Resource Locator）和METHOD三元组。其中，HOST用于指定请求资源的主机名和端口号；URL表示请求对应的地址，其与HOST属性组成完整的请求地址；METHOD是请求方法，例如,GET和POST是最常见的HTTP方法。

S240、将解析出的HOST、URL和METHOD三元组与各模板匹配单元中的匹配信息进行模板匹配。

以模板文件数据库中包括与网络购物、数据搜索以及信息共享相对应的模板文件，目标应用协议报文为提交订单操作类型的应用协议报文为例，对目标应用协议报文的解析过程进行具体说明，其中，网络购物包括建立订单，提交订单和查询订单三种操作类型，每一种操作类型对应一个模板文件。

其中，建立订单模板文件对应建立订单模板解析类，提交订单模板文件对应提交订单模板解析类，查询订单模板文件对应查询订单模板解析类，数据搜索模板文件对应数据搜索模板解析类，信息共享模板文件对应信息共享模板解析类，并且每个模板解析类都包括模板匹配单元、模板校验单元、字段内容提取单元和数据输出单元。在获取到与提交订单操作类型的目标应用协议报文相对应的HOST、URL和METHOD三元组后，可以利用各模板匹配单元分别与目标应用协议报文的三元组进行匹配。由于应用协议的三元组对应的是其遵循的标准协议，提交订单操作类型的目标应用协议报文的三元组对应的是网络购物的标准协议，因此，经过各模板匹配单元与目标应用协议报文的三元组进行匹配后，可以确定与提交订单操作类型的目标应用协议报文相匹配的模板解析类有三个，分别是网络购物下的建立订单模板解析类、提交订单模板解析类和查询订单模板解析类。

S250、如果存在与HOST、URL和METHOD三元组匹配的第一模板解析类，获取目标应用协议报文中HEAD和BODY的指定内容。

上述经过各模板匹配单元匹配后，得到三个可以初步匹配的第一模板解析类，分别是网络购物下的建立订单模板解析类、提交订单模板解析类和查询订单模板解析类。由于三个模板解析类分别对应网络购物下不同的操作类型，显然，不能利用三个模板解析类对目标应用协议报文进行解析。因此，需要利用模板解析类中的模板校验单元对目标应用协议报文进行进一步的校验。

本实施例中，由于应用协议报文的HEAD的扩展部分和BODY部分包括由自身特定功能与作用确定的特定内容，该特定内容与其他应用协议报文相比具有唯一性。因此，要想利用模板校验单元实现对目标应用协议报文的进一步校验，需要获取目标应用协议报文中HEAD的扩展部分和/或BODY部分的指定内容，利用该指定内容与各模板校验单元进行匹配校验，以在上述初步匹配的三个模板解析类中，唯一确定能够对目标应用协议报文进行解析的模板解析类。

需要注意的是，在实际应用过程中，模板文件数据库中可能并不存在与网络购物相对应的模板文件，此时，不存在与目标应用协议报文的三元组相匹配的模板匹配单元，利用各模板匹配单元与该三元组进行匹配时都会失败。在找不到与三元组相匹配的模板解析类的情况下，需要确定模板文件数据库中包含的模板文件是否全面，若确定包含的模板文件全面，则可以放弃对目标应用协议报文的解析；若确定包含的模板文件不全面，则可以在模板文件数据库中添加新的模板文件，并编译生成相应的模板解析类后，重新对目标应用协议报文进行三元组匹配。

S260、将指定内容与第一模板解析类的模板校验单元中的校验信息进行匹配校验，利用校验成功的第一模板解析类解析目标应用协议报文，并输出解析结果。

本实施例中，将获取到的提交订单操作类型的目标应用协议报文中HEAD的扩展部分和/或BODY部分的指定内容，与第一模板解析类的模板校验单元中的校验信息进行匹配校验。由于上述过程确定了三个第一模板解析类，因此需要将指定内容与三个第一模板解析类分别进行匹配校验，以唯一确定第一模板解析类。由于目标应用协议报文为提交订单操作类型的应用协议报文，其指定内容中必然包括与提交订单相关的内容，因此，只有当该指定内容与提交订单操作类型相对应的第一模板解析类进行匹配时，才会校验成功，此时可以确定能够对目标应用协议报文进行解析的第一模板解析类为与提交订单操作类型相对应的模板解析类。之后利用该与提交订单操作类型相对应的模板解析类对目标应用协议报文内容进行解析，并按预设格式输出解析结果。

同样需要注意的是，在实际应用过程中，模板文件数据库中可能并不存在与提交订单操作类型相对应的模板文件，此时，不存在与提交订单操作类型的目标应用协议报文的指定内容相匹配的模板校验单元，利用各模板校验单元与该指定内容进行匹配时都会失败。在找不到与指定内容相匹配的模板解析类的情况下，需要确定模板文件数据库中包含的模板文件是否全面，若确定包含的模板文件全面，则可以放弃对目标应用协议报文的解析；若确定包含的模板文件不全面，则可以在模板文件数据库中添加新的模板文件，并编译生成相应的模板解析类后，重新对目标应用协议报文进行指定内容的匹配校验。

本实施例提供的应用协议报文的自动化解析方法，通过添加与目标应用协议报文相对应的模板文件，利用模板文件生成包括能够与目标应用协议报文进行初步匹配的模板匹配单元、能够与目标应用协议报文进行二次匹配校验的模板校验单元、字段内容提取单元和数据输出单元的模板解析类，能够更加精确的确定能够进行目标应用协议报文解析的模板解析类，实现了目标应用协议报文的精确解析。

实施例三

本实施例在实施例二的基础上，提供了步骤S260的优选实施方式。图3是本发明实施例三提供的应用协议报文的自动化解析方法的流程图，如图3所示，该方法包括：

S310、根据应用协议报文的协议特征添加模板文件，其中，模板文件是利用预设脚本语言编辑的具有逻辑性的报文解析模板。

S320、对各模板文件进行编译处理，生成与各模板文件相对应的各模板解析类。

S330、对目标应用协议报文的头部进行解析，得到HOST、URL和METHOD三元组。

S340、将解析出的HOST、URL和METHOD三元组与各模板匹配单元中的匹配信息进行模板匹配。

S350、如果存在与HOST、URL和METHOD三元组匹配的第一模板解析类，获取目标应用协议报文中HEAD和BODY的指定内容；

S360、将指定内容与第一模板解析类的模板校验单元中的校验信息进行匹配校验，利用校验成功的第一模板解析类的字段内容提取单元，调用指定解码接口对目标应用协议报文进行字段内容提取。

本实施例中，在唯一确定了能够对目标应用协议报文进行解析的第一模板解析类后，可以利用第一模板解析类中的字段内容提取单元对目标应用协议报文中的字段内容进行提取。

本实施例中，字段内容提取单元中包括字段内容提取的原始数据来源、字段内容提取方式、各字段内容提取先后顺序、字段内容提取起始位置、提取字段内容、字段内容提取前的预处理以及字段内容提取后的后续处理等内容。其中，字段内容提取的原始数据来源可以是HEAD的扩展部分，可以是BODY部分，还可以是某个已经提取过的模板字段；字段内容提取方式可以包括START-END、MIME（多用途互联网邮件扩展类型，Multipurpose Internet Mail Extensions）、JSON（JS对象标记语言，JavaScript Object Notation）和URL四种类型；字段内容提取前的预处理以及字段内容提取后的后续处理可以是gzip解压或内容替换等。

根据字段内容提取单元描述的上述内容，调用基础解码库中预先保存的与提取内容相对应的解码接口，对应用协议报文进行相应的字段提取。本实施例中，基础解码库，包括各种基础的解码接口，例如可以是START-END、MIME、URL、JSON格式等各种字段提取接口，也可以是HAS，UNHAS，UNZIP等字段内容提取前的预处理以及提取后的后续处理。其中，各种基础解码接口统一整理封装在基础解码库中，可以根据字段内容提取单元描述的字段内容提取方式调用指定的解码接口以提取相应的内容。对于常见的因为原始数据报文格式发生变化而带来的各种情况，如需要多提取几个字段对应的内容、字段内容提取方式发生变化、在字段内容提取前需要进行预处理、在字段内容提取后还要增加后续处理等，都可以直接通过修改模板文件来实现字段内容的正确解析。

S370、按照第一模板解析类的数据输出单元中的预设格式，输出提取的字段内容。

本实施例中，每个数据输出单元中都描述了对应的模板文件所遵循的协议类型，当字段内容提取单元进行相应的字段内容提取后，数据输出单元可以根据通用数据集模块定义的对应的协议类型的输出格式，输出提取出的字段内容，以完成对目标应用协议报文的解析。其中，通用数据集模块定义了各个协议类型的协议字段输出格式。

本实施例提供的应用协议报文的自动化解析方法，利用校验成功的模板解析类中的字段内容提取单元和数据输出单元对目标应用协议报文进行字段内容提取和输出，可以实现灵活方便的描述用于模板解析的各种方法以及字段内容输出格式；并且当原始数据报文格式发生变化时，也可以通过直接修改模板文件实现正确解析，降低了人力投入，提高了开发效率，增加了系统的可维护性和可用性。

实施例四

图4是本发明实施例四中的应用协议报文的自动化解析装置的结构示意图。如图4所示，应用协议报文的自动化解析装置包括：

模板文件添加模块410，用于根据应用协议报文的协议特征添加模板文件，其中模板文件是利用预设脚本语言编辑的具有逻辑性的报文解析模板。

模板解析类生成模块420，用于对各模板文件进行编译处理，生成与各模板文件相对应的各模板解析类。

目标应用协议报文解析模块430，用于利用与目标应用协议报文匹配的模板解析类对目标应用协议报文进行解析，并输出解析结果。

本实施例提供的应用协议报文的自动化解析装置，通过利用应用协议报文的协议特征，添加利用预设脚本语言编辑的具有逻辑性的模板文件，并编译模板文件生成模板解析类，利用与目标应用协议报文匹配的模板解析类对目标应用协议报文进行解析，最终输出解析结果。这种利用模板文件进行应用协议报文解析的方法，解决了当原始应用协议报文的格式发生变化，或需要提取一种新格式的应用协议数据，或需要多提取或删除提取字段时，无需修改代码或重新发布软件版本，仅通过添加、修改或删除模板文件，即可实现对各类应用协议报文的解析，同时实现了应用协议解析过程的可维护性和可扩展性。

进一步的，模板解析类包括：模板匹配单元、模板校验单元、字段内容提取单元和数据输出单元。

进一步的，目标应用协议报文解析模块430，包括：

三元组解析子模块，用于对目标应用协议报文的头部进行解析，得到HOST、URL和METHOD三元组；

三元组匹配子模块，用于将解析出的HOST、URL和METHOD三元组与各模板匹配单元中的匹配信息进行模板匹配；

指定内容获取子模块，用于如果存在与HOST、URL和METHOD三元组匹配的第一模板解析类，获取目标应用协议报文中HEAD和BODY的指定内容；

匹配校验子模块，用于将指定内容与第一模板解析类的模板校验单元中的校验信息进行匹配校验；

目标应用协议报文解析子模块，用于利用校验成功的第一模板解析类解析目标应用协议报文，并输出解析结果。

进一步的，目标应用协议报文解析子模块具体用于：

利用校验成功的第一模板解析类的字段内容提取单元，调用指定解码接口对目标应用协议报文进行字段内容提取；

按照第一模板解析类的数据输出单元中的预设格式，输出提取的字段内容。

进一步的，模板解析类生成模块420具体用于：

利用模板编译器对各模板文件进行解析，提取与各模板文件相对应的各字段描述信息集合；

利用各字段描述信息集合构造各模板解析类。

本发明实施例所提供的应用协议报文的自动化解析装置可执行本发明任意实施例所提供的应用协议报文的自动化解析方法，具备执行方法相应的功能模块和有益效果。

实施例五

图5为本发明实施例五提供的一种服务器的结构示意图。图5示出了适于用来实现本发明实施方式的示例性服务器设备512的框图。图5显示的服务器512仅仅是一个示例，不应对本发明实施例的功能和使用范围带来任何限制。

如图5所示，服务器512以通用计算设备的形式表现。服务器512的组件可以包括但不限于：一个或者多个处理器516，系统存储器528，连接不同系统组件（包括系统存储器528和处理器516）的总线518。

总线518表示几类总线结构中的一种或多种，包括存储器总线或者存储器控制器，外围总线，图形加速端口，处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说，这些体系结构包括但不限于工业标准体系结构（ISA）总线，微通道体系结构（MAC）总线，增强型ISA总线、视频电子标准协会（VESA）局域总线以及外围组件互连（PCI）总线。

服务器512典型地包括多种计算机系统可读介质。这些介质可以是任何能够被服务器512访问的可用介质，包括易失性和非易失性介质，可移动的和不可移动的介质。

系统存储器528可以包括易失性存储器形式的计算机系统可读介质，例如随机存取存储器（RAM）530和/或高速缓存存储器532。服务器512可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例，存储装置534可以用于读写不可移动的、非易失性磁介质（图5未显示，通常称为“硬盘驱动器”）。尽管图5中未示出，可以提供用于对可移动非易失性磁盘（例如“软盘”）读写的磁盘驱动器，以及对可移动非易失性光盘（例如CD-ROM, DVD-ROM或者其它光介质）读写的光盘驱动器。在这些情况下，每个驱动器可以通过一个或者多个数据介质接口与总线518相连。存储器528可以包括至少一个程序产品，该程序产品具有一组（例如至少一个）程序模块，这些程序模块被配置以执行本发明各实施例的功能。

具有一组（至少一个）程序模块542的程序/实用工具540，可以存储在例如存储器528中，这样的程序模块542包括但不限于操作系统、一个或者多个应用程序、其它程序模块以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块542通常执行本发明所描述的实施例中的功能和/或方法。

服务器512也可以与一个或多个外部设备514（例如键盘、指向设备、显示器524等）通信，还可与一个或者多个使得用户能与该计算机设备512交互的设备通信，和/或与使得该服务器512能与一个或多个其它计算设备进行通信的任何设备（例如网卡，调制解调器等等）通信。这种通信可以通过输入/输出（I/O）接口522进行。并且，计算机设备512还可以通过网络适配器520与一个或者多个网络（例如局域网（LAN），广域网（WAN）和/或公共网络，例如因特网）通信。如图所示，网络适配器520通过总线518与计算机设备512的其它模块通信。应当明白，尽管图中未示出，可以结合计算机设备512使用其它硬件和/或软件模块，包括但不限于：微代码、设备驱动器、冗余处理器、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储装置等。

处理器516通过运行存储在系统存储器528中的程序，从而执行各种功能应用以及数据处理，例如实现本发明实施例所提供的应用协议报文的自动化解析方法，包括：

根据应用协议报文的协议特征添加模板文件，其中，模板文件是利用预设脚本语言编辑的具有逻辑性的报文解析模板；

对各模板文件进行编译处理，生成与各模板文件相对应的各模板解析类；

利用与目标应用协议报文匹配的模板解析类对目标应用协议报文进行解析，并输出解析结果。

实施例六

本发明实施例六提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如本发明实施例所提供的应用协议报文的自动化解析方法，包括：

根据应用协议报文的协议特征添加模板文件，其中，模板文件是利用预设脚本语言编辑的具有逻辑性的报文解析模板；

对各模板文件进行编译处理，生成与各模板文件相对应的各模板解析类；

利用与目标应用协议报文匹配的模板解析类对目标应用协议报文进行解析，并输出解析结果。

本发明实施例的计算机存储介质，可以采用一个或多个计算机可读的介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子（非穷举的列表）包括：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器（RAM）、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括——但不限于无线、电线、光缆、RF等等，或者上述的任意合适的组合。

可以以一种或多种程序设计语言或其组合来编写用于执行本发明操作的计算机程序代码，所述程序设计语言包括面向对象的程序设计语言—诸如Java、Smalltalk、C++，还包括常规的过程式程序设计语言—诸如”C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)—连接到用户计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。

注意，上述仅为本发明的较佳实施例及所运用技术原理。本领域技术人员会理解，本发明不限于这里所述的特定实施例，对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本发明的保护范围。因此，虽然通过以上实施例对本发明进行了较为详细的说明，但是本发明不仅仅限于以上实施例，在不脱离本发明构思的情况下，还可以包括更多其他等效实施例，而本发明的范围由所附的权利要求范围决定。



图1



图2



图3



图4



图5