(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2012 年 6 月 7 日 (07.06.2012)

WIPO | PCT

(10) 国际公布号
WO 2012/071951 A1

(51) 国际专利分类号:
H04L 29/06 (2006.01)

(21) 国际申请号: PCT/CN2011/080795

(22) 国际申请日: 2011 年 10 月 14 日 (14.10.2011)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201010578874.7 2010 年 11 月 29 日 (29.11.2010) CN

(71) 申请人 (对除美国外的所有指定国): **华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人; 及

(75) 发明人/申请人 (仅对美国): **陈建 (CHEN, Jian)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **邹嵘 (ZOU, Rong)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **周洪 (ZHOU, Hong)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **胡新宇 (HU, Xinyu)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **罗之丹 (LUO, Zhidan)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(81) 指定国 (除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明，要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

[见续页]

(54) **Title**: METHOD AND DEVICE USED IN ACQUIRING PARAMETERS FOR GENERAL ANALYSIS OF PROTOCOL AND IN GENERAL ANALYSIS OF PROTOCOL

(54) **发明名称** : 用于通用协议解析的参数获取和通用协议解析方法及装置

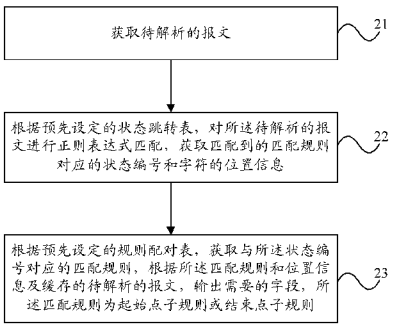


图 2 / FIG. 2

21 ACQUIRE A MESSAGE TO BE ANALYSED
22 PERFORM A HOLOMORPHIC EXPRESSION MATCHING ON THE MESSAGE TO BE ANALYSED ACCORDING TO A PREDEFINED STATE JUMPING TABLE, AND ACQUIRE LOCATION INFORMATION ABOUT A STATE NUMBER AND A CHARACTER CORRESPONDING TO A MATCHING RULE FOUND OUT BY MATCHING
23 ACQUIRE A MATCHING RULE CORRESPONDING TO THE STATE NUMBER ACCORDING TO A PREDEFINED RULE PAIRING TABLE, AND OUTPUT A REQUIRED FIELD ACCORDING TO THE MATCHING RULE AND THE LOCATION INFORMATION AS WELL AS THE BUFFERED MESSAGE TO BE ANALYSED, WITH THE MATCHING RULE BEING A START POINT SUB-RULE OR AN END POINT SUB-RULE

(57) **Abstract**: The present invention provides a method and a device used in acquiring parameters for the general analysis of a protocol and in the general analysis of a protocol. The method includes: acquiring a message to be analysed; performing a holomorphic expression matching on the message to be analysed according to a predefined state jumping table, and acquiring location information about a state number and a character corresponding to a matching rule found out by matching; acquiring a matching rule corresponding to the state number according to a predefined rule pairing table, and outputting a required field according to the matching rule and the location information as well as the buffered message to be analysed, with the matching rule being a start point sub-rule or an end point sub-rule. The embodiments of the present invention can realise a general analysis of protocols.

[见续页]

WO 2012/071951 A1

本国际公布:
— 包括国际检索报告(条约第 21 条(3))。

(57) 摘要:

本发明提供一种用于通用协议解析的参数获取和通用协议解析方法及装置。该方法包括获取待解析的报文；根据预先设定的状态跳转表，对所述待解析的报文进行正则表达式匹配，获取匹配到的匹配规则对应的状态编号和字符的位置信息；根据预先设定的规则配对表，获取与所述状态编号对应的匹配规则，根据所述匹配规则和位置信息及缓存的待解析的报文，输出需要的字段，所述匹配规则为起始点子规则或结束点子规则。本发明实施例可以实现对协议的通用解析。

用于通用协议解析的参数获取和通用协议解析方法及装置

本申请要求于 2010 年 11 月 29 日提交中国专利局、申请号为 201010578874.7、发明名称为“用于通用协议解析的参数获取和通用协议解析方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本发明涉及网络通信技术，尤其涉及一种用于通用协议解析的参数获取和通用协议解析方法及装置。

背景技术

网络中，通信的双方基于标准的协议进行通信。网络协议解析是指通过程序分析网络数据包的协议头和尾，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。从本质上说，网络通信的双方，通信的过程就是一个对网络数据包根据标准的协议进行报文解析的过程。网络设备的协议解析通常是基于协议栈的方式进行的。协议栈是一个分层的解析系统，每一层处理相应的头后，剥离头部数据向上层投递，直到应用层。在应用层根据具体的应用类型，由对应的应用处理模块对应用协议进行字段分析，看是否匹配一些预定的条件，从而提取出一些有价值的字段。

现有协议解析时，通常是采用定位分界符-比较字段-保存内容这一过程。其中，分界符随着协议的不同而不同，例如，HTTP、RTSP 等协议，“\r\n”表示一个字段的结束，而 SIP 中以空格、“;”作为分界符。比较字段是指找出需要的字段，例如在 SIP 报文中需要的字段为 INVITE 和 transport，则需要通过比较找到 INVITE 字段和 transport 字段，之后，将 INVITE 字段和 transport 字段对应的内容保存。重复上述过程直到报文结束或满足预设的结束条件。

基于协议栈的协议解析方法要求对待解析的协议都单独编码处理。由于新的应用协议层出不穷，基于协议栈的解析的方法维护工作量很大，扩展性存在问题，支持新协议解析需要的周期长。非通用的解析方法增加了用硬件方式实现的难度，在性能上存在瓶颈。

发明内容

本发明实施例是提供一种用于通用协议解析的参数获取和通用协议解析方法及装置，用以解决现有技术中对待解析的协议需要单独处理引起的问题，实现对所有协议的通用解析。

一方面，本发明实施例提供了一种用于通用协议解析的参数获取方法，包括：

读入需要匹配的协议字段对应的正则表达式，所述正则表达式中至少包含起始点子规则和结束点子规则；

根据所述起始点子规则和结束点子规则，编译形成状态跳转表和规则配对表，所述状态跳转表中保存输入的字符与跳转到的状态的对应关系，所述规则配对表中保存所述状态跳转表中的接受态与所述起始点子规则或所述结束点子规则的对应关系。

另一方面，本发明实施例提供了一种通用协议解析方法，包括：

获取待解析的报文；

根据预先设定的状态跳转表，对所述待解析的报文进行正则表达式匹配，获取匹配到的匹配规则对应的状态编号和字符的位置信息；

根据预先设定的规则配对表，获取与所述状态编号对应的匹配规则，根据所述匹配规则和位置信息及缓存的待解析的报文，输出需要的字段，所述匹配规则为起始点子规则或结束点子规则。

一方面，本发明实施例提供了一种用于通用协议解析的参数获取装置，包括：

读入模块，用于读入需要匹配的协议字段对应的正则表达式，所述正则表达式中至少包含起始点子规则和结束点子规则；

编译模块，用于根据所述起始点子规则和结束点子规则，编译形成状态跳转表和规则配对表，所述状态跳转表中保存输入的字符与跳转到的状态的对应关系，所述规则配对表中保存所述状态跳转表中的接受态与所述起始点子规则或所述结束点子规则的对应关系。

另一方面，本发明实施例提供了一种通用协议解析装置，包括：

报文过滤器，用于获取待解析的报文；

匹配模块，用于根据预先设定的状态跳转表，对所述待解析的报文进行正则表达式匹配，获取匹配到的匹配规则对应的状态编号和字符的位置信息；根据预先设定的规则配对表，获取与所述状态编号对应的匹配规则，根据所述匹配规则和位置信息及缓存的待解析的报文，输出需要的字段，所述匹配规则为起始点子规则或结束点子规则。

由上述技术方案可知，本发明实施例通过对需要解析的协议字段用正则表达式描述，并根据正则表达式中的起始点子规则和结束点子规则得到用于协议解析的状态跳转表和规则配对表，可以得到要解析报文中匹配起始点子规则和结束点子规则的部分，并进一步得到需要解析的协议字段，无需根据各协议的特点获取对应的分界符，可以实现对协议的通用处理。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图1为本发明实施例一的方法流程示意图；

图2为本发明实施例二的方法流程示意图；

图3为本发明实施例三的方法流程示意图；

图4为本发明实施例四的装置结构示意图；

图5为本发明实施例五的装置结构示意图；

图6为本发明实施例六的装置结构示意图。

具体实施方式

下面通过实施例对本发明的具体实现过程进行举例说明。显然，下面所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

图1为本发明实施例一的方法流程示意图，包括：

步骤11：读入需要匹配的协议字段对应的正则表达式（regular expression），所述正则表达式中至少包含起始点子规则和结束点子规则。

正则表达式描述了一种字符串匹配的模式，可以用于进行文本匹配，文本匹配是指在给定的字符串中寻找与给定的正则表达式相匹配的部分。如有这样一个正则表达式：.*AUTH[0-9]{10}，则表示需要在待匹配的文本中找到这样的一个字符串，其特征是文本中有一个AUTH的字符串，其后直接跟随了10个0~9之间的任意某个数字字符。这

样的话，可能匹配上的字符文本就可以为：http://AUTH2009120901.html/~index，其中的“AUTH2009120901”就是可以匹配上的这个正则表达式的字串。

起始点子规则是描述需要匹配的协议字段的起始位置，结束点子规则是描述需要匹配的协议字段的结束位置；起始点子规则和结束点子规则可以分别描述，例如，将一个字段用两个正则表达式描述，分别对应起始点子规则和结束点子规则；对于起始点子规则和结束点子规则也可以在一个正则表达式中通过添加特殊标记的方法来描述，例如，可以用<表示前面是一个起始点子规则，>表示后面是一个结束点子规则。

步骤12：根据所述起始点子规则和结束点子规则，编译形成状态跳转表和规则配对表，所述状态跳转表中保存输入的字符与跳转到的状态的对应关系，所述规则配对表中保存所述状态跳转表中的接受态与所述起始点子规则或所述结束点子规则的对应关系。

在得到状态跳转表和规则配对表之后，在报文解析时，可以由通用协议解析装置根据该状态跳转表和规则配对表从待解析的报文中获取需要匹配的协议字段。

其中，所述状态跳转表可以采用通常的方法生成，例如，用Perl兼容正则表达式（Perl Compatible Regular Expressions，PCRE）编译器生成，其中保存输入的字符与跳转到的状态的对应关系，根据所述状态跳转表能够输出：输入的待解析的报文中的字符串匹配到起始点子规则或者结束点子规则时，对应的状态编号和字符的位置信息。

所述规则配对表可以采用通常的方法生成后进行形式处理，例如，采用PCRE生成通常的规则配对表后，用新增的参数“起始/结束属性”来分别代表同一个正则表达式对应的起始点子规则和结束点子规则，具体地，如果按照通常的方式，各正则表达式对应的起始点子规则和结束点子规则分别作为独立的匹配规则，而本发明实施例中，一个正则表达式用一个匹配规则对应，该正则表达式对应的起始点子规则或者结束点子规则用“起始/结束属性”来表征。这样可以节省匹配规则的数目，降低规则配对表所需的资源。本发明实施例中的规则配对表中保存状态与规则的对应关系，根据所述规则配对表能够输出：与输入的状态编号对应的匹配规则，使得根据所述匹配规则以及所述位置信息及缓存的待解析的报文能够确定所述需要的协议字段，其中，匹配规则为起始点子规则或者结束点子规则。例如，匹配规则表明为起始点子规则，则需要的协议字段的起始点为缓存的待解析的报文中与所述位置信息对应的字符，具体地，如果缓存的待解析的报文中的字符依次为a、b、c…，如果位置信息为2，则需要的协议字段的起始点为b；对于结束点子规则也可以采用类似的方式确定结束点，之后，将缓存的待解析的报文中的包括起始点和结束点的两点之间的字符作为需要的协议字段。

本实施例通过对需要解析的协议字段用正则表达式描述，并根据正则表达式中的起始点子规则和结束点子规则得到用于协议解析的状态跳转表和规则配对表，可以得到要解析报文中匹配起始点子规则和结束点子规则的部分，并进一步得到需要解析的协议字段，无需根据各协议的特点获取对应的分界符，可以实现对协议的通用处理。

图2为本发明实施例二的方法流程示意图，包括：

步骤21：获取待解析的报文；

其中，可以获取所有接收到的报文作为待解析的报文。也可以对接收的报文进行过滤，过滤后的报文为待解析的报文。具体地，可以通过设置关键字，当接收的报文中包含设置的关键字时，则确定为待解析的报文。

步骤22：根据预先设定的状态跳转表，对所述待解析的报文进行正则表达式匹配，获取匹配到的匹配规则对应的状态编号和字符的位置信息

步骤23：根据预先设定的规则配对表，获取与所述状态编号对应的匹配规则，根据所述匹配规则和位置信息及缓存的待解析的报文，输出需要的字段，所述匹配规则为起始点子规则或结束点子规则。

其中，所述状态跳转表和规则配对表的具体内容可以参见实施例一所述。

图3为本发明实施例三的方法流程示意图，本实施例以解析http get报文为例。假设该http get报文中需要获取的字段为：GET字段、user-agent字段、host字段和cookie字段，每个字段以\r\n结束。

参见图3，本实施例包括：

步骤31：读入模块读入需要匹配的协议字段对应的正则表达式需要匹配的。

以上述的http get报文为例，需要匹配的字段包括GET字段、user-agent字段、host字段和cookie字段。

对应的正则表达式分别为：

1）pcre:/^GET[\x20\x09]<.*>\x2\r\n/is

含义：从净荷（payload）起始开始查找GET字串，后面接一个空格（对应\x20）或tab（对应\09），然后是任意长度字符为GET字段的内容，最后以回车换行结束

2）pcre:/user-agent:<.*>\x2\r\n/is

含义：任意位置匹配user-agent:，然后任意长度的字符为user-agent字段内容，最后以回车换行结束。

3）pcre:/host:<.*>\x2\r\n/is

含义：任意位置匹配host:，然后任意长度的字符为host字段内容，最后以回车换行结束。

4）pcre:/cookie:<.*>\x2\r\n/is

含义：任意位置匹配cookie:，然后任意长度的字符为cookie字段内容，最后以回车换行结束。

上面的4个正则表达式中：

每个协议字段的解析用一个规则来描述，一个规则又分为三部分，第一个部分表示字段的起始点，如^GET[\x20\x09]，第二部分表示字段内容，如<.*>,第三部分表示字段结束,如\x2\r\n。

“Pcre:”和“is”是正则表达式语法属性标记。两个斜杠“/”中间的部分是正则表达式。

尖括号<>和\x2被赋予特殊涵义，它们不是标准正则表达式的一部份，是本发明实施例规则中设定的一些特殊标记，本发明实施例根据它们对规则进行拆分。

<表示前面是一个起始匹配规则，

>表示后面是一个结束匹配规则，

\xn表示字段结束匹配需要回退的字节数为n，例如，\x2表示需要回退2个字符。

步骤32：编译器对该正则表达式进行编译，得到状态跳转表和规则配对表。之后，可以保存在内存，如双倍数据率同步动态随机存取内存（Double Data Rate Synchronous Dynamic Random Access Memory，DDR SDRAM，以下简称DDR）中。

其中，可以采用编译器对正则表达式进行编译，编译器的输入为上述的4个正则表达式，编译器的处理可以分为预处理和处理，其中，预处理可以得到各正则表达式对应的子规则，处理是根据预处理的结果得到状态跳转表和规则配对表。

首先，对该4个正则表达式进行预处理，得到拆分后的子规则，如下表1所示：

表1

规则编号	起始点子规则	结束点子规则
1	pcre:/^GET[\x20\x09]/is	pcre:/\r\n/is，匹配后回退2
2	pcre:/user-agent:/is	pcre:/\r\n/is，匹配后回退2
3	pcre:/host:/is	pcre:/\r\n/is，匹配后回退2
4	pcre:/cookie:/is	pcre:/\r\n/is，匹配后回退2

编码器在上述预处理后，根据表1所示的规则再进行处理，得到状态跳转表和规则配对表。

其中，状态跳转表可以保存在状态跳转表缓存模块中，状态跳转表缓存模块可以为确定有限自动机（deterministic finite automation，DFA），也可以为非确定有限自动机（nondeterministic finite automaton，NFA）。以下以DFA为例，状态跳转表可以如表2所示：

表2

	0	1	2	3	4	…	a	b	c	d	e	f	g	h	…
S1							S2								
S2								S3							
S3（acc）									S4						

该状态跳转表中，纵向的字符（0,1,…a,b,…）表示接收到的报文中的字符，横向的S1、S2、S3表示状态。例如，当前状态为S1，则输入的字符为“a”后，则跳转到状态S2。另外，带标记的状态（具体为S3（acc））是一个接收态，表示匹配上了某条规则，状态跳转表到该状态时，则输出一个匹配结果，具体为状态编号和匹配报文的位置。例如，上述带标记的状态为S3，则当跳转到S3时，输出S3的编号（3）及此时对应的字符（b）的位置（如输入了ab，则位置为2）。

可以理解的是，上述状态跳转表只是示例，并不限定为上述的3个状态，另外，上述跳转的S2、S3、S4也只是示例，上表2中的每个格中均应该有对应的跳转状态。

规则配对表可以如表3所示：

表3

接收态	rule6	rule5	rule4	rule3	rule2	rule1	起始/	回退属

							结束属性	性
S3	0	0	0	0	0	1	0	
S4	0	0	0	0	1	0	0	
S5	0	0	1	0	0	0	0	
…								
S10	1	1	0	0	0	0	1	2
S11	0	0	1	1	1	1	1	2
…								

其中，规则配对表中，rule1~rule6表示匹配的规则，当对应的数字为“1”时表示匹配，为“0”时表示不匹配；

起始/结束属性表示该状态对应的是起始点子规则还是结束点子规则，当“0”时表示起始子规则，“1”时表示结束子规则。

回退属性表示应该对匹配的位置回退多少字节。

例如，接收态为S5时，则匹配的规则为rule4中的起始子规则，如果rule4为表1所示的rule4，则得到需要匹配的字段的起始字段为“cookie:”。

需要说明的是，本发明实施例针对各需要匹配的协议字段是相互独立，不会相互嵌套或者存在重叠区域的情况。此时，当找到某个需要匹配的协议字段的起始点之后，如果找到一个匹配到结束点子规则的结束点，则认为该结束点为已经找到的起始点对应的结束点。在某一需要匹配的协议字段的起始点和结束点都找到之后，才可以获取下一次需要匹配的协议字段。因此，参见图3所示的规则配对表，当某个接受态可能对应多个结束点子规则时，通过位置信息离已经找到的起始点最近的结束点即为需要的结束点。

步骤33：DDR将状态跳转表写入状态跳转表缓存模块中，将规则配对表写入规则配对表缓存模块中。

其中，可以预先编译不同的协议对应的正则表达式，之后，将不同的协议对应的状态跳转表和规则配对表保存在DDR中，之后，在需要解析某一协议时，在DDR中将需要解析的协议的状态跳转表和规则配对表写入状态跳转表缓存模块和规则配对表缓存模块中。

通过步骤31-33可以完成协议解析的准备工作，之后，在接收到报文后可以进行解

析。

步骤34：报文过滤器对接收的报文进行过滤处理，得到待解析的报文。之后，报文过滤器可以将待解析的报文保存在报文缓存模块中。

其中，可以在报文过滤器中预先保存关键字，当接收的报文中包含该关键字时，则确定该接收的报文为待解析的报文。

步骤35：正则表达式引擎从状态跳转表缓存模块获取状态跳转表并根据状态跳转表，对报文缓存模块中的待匹配报文进行匹配处理，输出匹配到正则表达式对应的规则后的状态编号和字符的位置信息。

其中，报文过滤器在过滤处理得到待解析的报文后，可以向正则表达式引擎发送控制信息，指示正则表达式引擎进行上述处理。

正则表达式在匹配过程中，可以对待解析的报文中的字符按照表2所示的状态跳转表进行状态转换。例如，设置初始状态为S1，则待解析的报文中的字符为a时，则跳转到S2。当匹配到规则时，则对应接收态，例如，当输入为“GET\x20”，此时匹配到规则1，则在\x20（对应的字符为空格）时跳转到接收态，假设此时的接收态为S3，则将S3对应的编号“3”输出，另外，还输出“\x20”在整个报文的位置信息，假设报文是按照“GET\x20”依次输入，则该位置信息为“4”。

步骤36：解析器根据规则配对表缓存模块中的规则配对表及正则表达式引擎输出的状态的编号和字符的位置信息，及报文缓存模块中保存的待匹配报文，输出需要匹配的字段。

具体地，通过查找表3找到与该状态的编号对应的规则，例如，输出的状态的编号为3，则查找与S3对应的规则，假设此时对应的规则为规则1，且假设此时的起始/结束属性表明为起始规则，则可以获知匹配到的规则为：规则1的起始规则，之后根据此时的位置信息输出字段，例如，如果此时的位置信息为“4”，则从缓存的报文的第5个字符开始输出。类似的，可以找到结束字符，在结束规则中，由于“\r\n”占用两个字符，有效字符为该两个字符之前的字符，所以，需要回退2个字符，即结束字符为该“\r\n”字符之前的一个字符。可以理解的是，上述的正则表达式引擎可以既完成状态跳转，也完成规则匹配，而解析器根据匹配到的规则编号和位置信息获取需要的字段。即所述正则表达式引擎用于根据预先设定的状态跳转表，对所述待解析的报文进行正则表达式匹配，输出匹配到正则表达式时，对应的状态编号和字符的位置信息，并根据预先设定的规则配对表，获取与所述状态的编号对应的匹配规则；所述解析器用于根据所述匹配规

则和位置信息，输出需要的字段。

本实施例通过对待解析的报文进行正则表达式匹配，无需根据各协议的特点获取对应的分界符，可以实现对协议的通用处理；本实施例的方法具有通用性，将协议的解析工作转化为正则表达式的描述，适用于各种协议的解析，扩展性好，能迅速支持新的协议；正则表达式引擎和解析器本身是稳定的，适合用硬件的方式固化下来，性能大幅提升。

图4为本发明实施例四的装置结构示意图，包括读入模块41和编译模块42；读入模块41用于读入需要匹配的协议字段对应的正则表达式，所述正则表达式中至少包含起始点子规则和结束点子规则；编译模块42用于根据所述起始点子规则和结束点子规则，编译形成状态跳转表和规则配对表，所述状态跳转表中保存输入的字符与跳转到的状态的对应关系，所述规则配对表中保存所述状态跳转表中的接受态与所述起始点子规则或所述结束点子规则的对应关系

本实施例通过对需要解析的协议字段用正则表达式描述，并根据正则表达式得到用于协议解析的状态跳转表和规则配对表，可以无需根据各协议的特点获取对应的分界符，可以实现对协议的通用处理。

图5为本发明实施例五的装置结构示意图，包括报文过滤器51和匹配模块52；报文过滤器51用于获取待解析的报文；匹配模块52用于根据预先设定的状态跳转表，对所述待解析的报文进行正则表达式匹配，获取匹配到的匹配规则对应的状态编号和字符的位置信息；根据预先设定的规则配对表，获取与所述状态编号对应的匹配规则，根据所述匹配规则和位置信息及缓存的待解析的报文，输出需要的字段，所述匹配规则为起始点子规则或结束点子规则。

本实施例通过对待解析的报文进行正则表达式匹配，无需根据各协议的特点获取对应的分界符，可以实现对协议的通用处理。

图6为本发明实施例六的装置结构示意图，本实施例的装置6包括报文过滤器61和匹配模块，其中，匹配模块包括正则表达式引擎62和解析器63。所述正则表达式引擎62用于根据预先设定的状态跳转表，对所述待解析的报文进行正则表达式匹配，输出匹配到匹配规则时，对应的状态编号和字符的位置信息；所述解析器63用于根据预先设定的规则配对表，获取与所述状态的编号对应的匹配规则，根据所述匹配规则和位置信息以及缓存的待解析的报文，输出需要的字段；或者，所述正则表达式引擎62用于根据预先设定的状态跳转表，对所述待解析的报文进行正则表达式匹配，输出匹配到匹配规则时，

对应的状态编号和字符的位置信息，并根据预先设定的规则配对表，获取与所述状态的编号对应的匹配规则；所述解析器63用于根据所述匹配规则和位置信息以及缓存的待解析的报文，输出需要的字段。

本实施例的装置6还可以包括状态跳转表缓存模块64、规则配对表缓存模块65和报文缓存模块66。状态跳转表缓存模块64用于获取状态跳转表，所述状态跳转表中保存输入的字符与跳转到的状态的对应关系；规则配对表缓存模块65用于获取规则配对表，所述规则配对表中保存所述状态跳转表中的接受态与所述起始点子规则或所述结束点子规则的对应关系；报文缓存模块66用于缓存待解析的报文。

其中，状态跳转表缓存模块64及规则配对表缓存模块65中保存的信息可以是从外部模块7获取，该外部模块7包括编译器71和DDR 72，其中，编译器71可以包括图4所示的装置，用于对不同的协议对应的正则表达式进行编译，得到状态跳转表和规则配对表，之后，将不同协议对应的状态跳转表和规则配对表保存在DDR 72中。当需要解析某种协议时，DDR 72可以将对应的该种协议对应的状态跳转表和规则配对表分别写入状态跳转表缓存模块64和规则配对表缓存模块65中。

本实施例的本实施例的装置6可以位于现场可编程门阵列（Field Programmable Gate Array，FPGA）上。

本实施例通过对待解析的报文进行正则表达式匹配，无需根据各协议的特点获取对应的分界符，可以实现对协议的通用处理；本实施例的方法具有通用性，将协议的解析工作转化为正则表达式的描述，适用于各种协议的解析，扩展性好，能迅速支持新的协议；正则表达式引擎和解析器本身是稳定的，适合用硬件的方式固化下来，性能大幅提升。

可以理解的是，上述方法及设备中的相关特征可以相互参考。

本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述，仅为本发明较佳的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到的变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应该以权利要求的保护范围为准。

权利要求

1、一种用于通用协议解析的参数获取方法，其特征在于，包括：

读入需要匹配的协议字段对应的正则表达式，所述正则表达式中至少包含起始点子规则和结束点子规则；

根据所述起始点子规则和结束点子规则，编译形成状态跳转表和规则配对表，所述状态跳转表中保存输入的字符与跳转到的状态的对应关系，所述规则配对表中保存所述状态跳转表中的接受态与所述起始点子规则或所述结束点子规则的对应关系。

2、根据权利要求1所述的方法，其特征在于，所述起始点子规则描述需要匹配的协议字段的起始位置，所述结束点子规则描述需要匹配的协议字段的结束位置。

3、一种通用协议解析方法，其特征在于，包括：

获取待解析的报文；

根据预先设定的状态跳转表，对所述待解析的报文进行正则表达式匹配，获取匹配到的匹配规则对应的状态编号和字符的位置信息；

根据预先设定的规则配对表，获取与所述状态编号对应的匹配规则，根据所述匹配规则和位置信息及缓存的待解析的报文，输出需要的字段，所述匹配规则为起始点子规则或结束点子规则。

4、根据权利要求3所述的方法，其特征在于，所述获取待解析的报文之后，所述方法还包括：

获取状态跳转表和规则配对表；所述状态跳转表中保存输入的字符与跳转到的状态的对应关系，所述规则配对表中保存所述状态跳转表中的接受态与所述起始点子规则或所述结束点子规则的对应关系。

5、一种用于通用协议解析的参数获取装置，其特征在于，包括：

读入模块，用于读入需要匹配的协议字段对应的正则表达式，所述正则表达式中至少包含起始点子规则和结束点子规则；

编译模块，用于根据所述起始点子规则和结束点子规则，编译形成状态跳转表和规则配对表，所述状态跳转表中保存输入的字符与跳转到的状态的对应关系，所述规则配对表中保存所述状态跳转表中的接受态与所述起始点子规则或所述结束点子规则的对应关系。

6、根据权利要求5所述的方法，其特征在于，所述起始点子规则描述需要匹配的协议字段的起始位置，所述结束点子规则描述需要匹配的协议字段的结束位置。

7、一种通用协议解析装置，其特征在于，包括：

报文过滤器，用于获取待解析的报文；

匹配模块，用于根据预先设定的状态跳转表，对所述待解析的报文进行正则表达式匹配，获取匹配到的匹配规则对应的状态编号和字符的位置信息；根据预先设定的规则配对表，获取与所述状态编号对应的匹配规则，根据所述匹配规则和位置信息及缓存的待解析的报文，输出需要的字段，所述匹配规则为起始点子规则或结束点子规则。

8、根据权利要求7所述的装置，其特征在于，所述匹配模块包括：

正则表达式引擎和解析器；

所述正则表达式引擎用于根据预先设定的状态跳转表，对所述待解析的报文进行正则表达式匹配，输出匹配到匹配规则时，对应的状态编号和字符的位置信息；所述解析器用于根据预先设定的规则配对表，获取与所述状态的编号对应的匹配规则，根据所述匹配规则和位置信息以及缓存的待解析的报文，输出需要的字段；或者，

所述正则表达式引擎用于根据预先设定的状态跳转表，对所述待解析的报文进行正则表达式匹配，输出匹配到匹配规则时，对应的状态编号和字符的位置信息，并根据预先设定的规则配对表，获取与所述状态的编号对应的匹配规则；所述解析器用于根据所述匹配规则和位置信息以及缓存的待解析的报文，输出需要的字段。

9、根据权利要求7或8所述的装置，其特征在于，还包括：

状态跳转表缓存模块，用于获取状态跳转表，所述状态跳转表中保存输入的字符与跳转到的状态的对应关系；

规则配对表缓存模块，用于获取规则配对表，所述规则配对表中保存所述状态跳转表中的接受态与所述

起始点子规则或所述结束点子规则的对应关系。

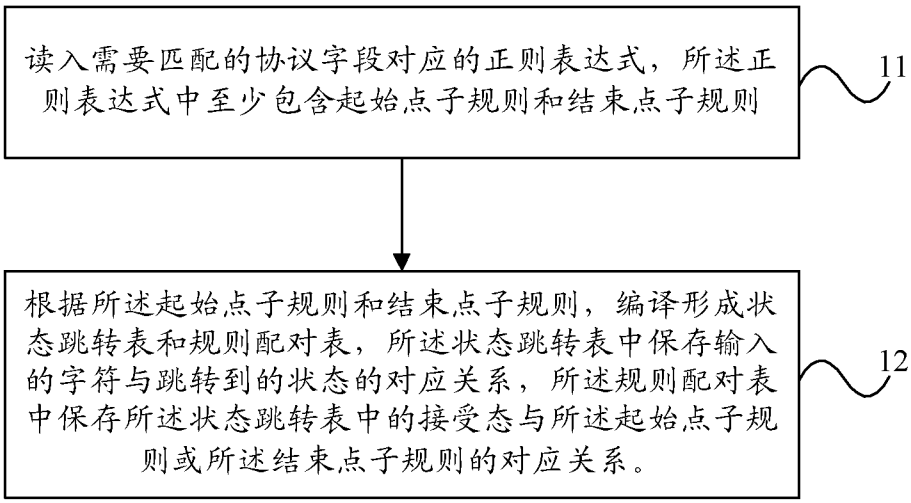

读入需要匹配的协议字段对应的正则表达式，所述正则表达式中至少包含起始点子规则和结束点子规则
11
根据所述起始点子规则和结束点子规则，编译形成状态跳转表和规则配对表，所述状态跳转表中保存输入的字符与跳转到的状态的对应关系，所述规则配对表中保存所述状态跳转表中的接受态与所述起始点子规则或所述结束点子规则的对应关系。
12

图 1

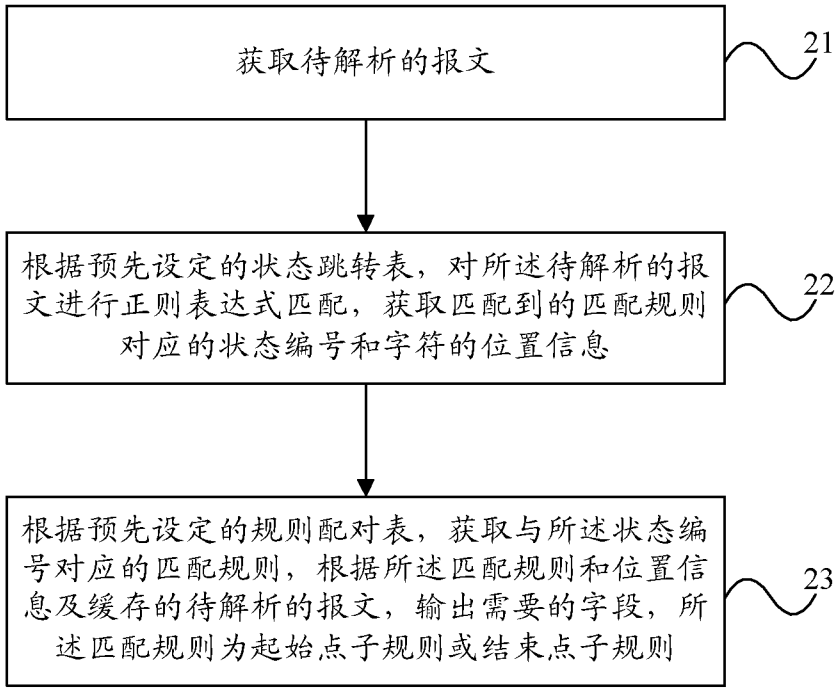

获取待解析的报文
21
根据预先设定的状态跳转表，对所述待解析的报文进行正则表达式匹配，获取匹配到的匹配规则对应的状态编号和字符的位置信息
22
根据预先设定的规则配对表，获取与所述状态编号对应的匹配规则，根据所述匹配规则和位置信息及缓存的待解析的报文，输出需要的字段，所述匹配规则为起始点子规则或结束点子规则
23

图 2

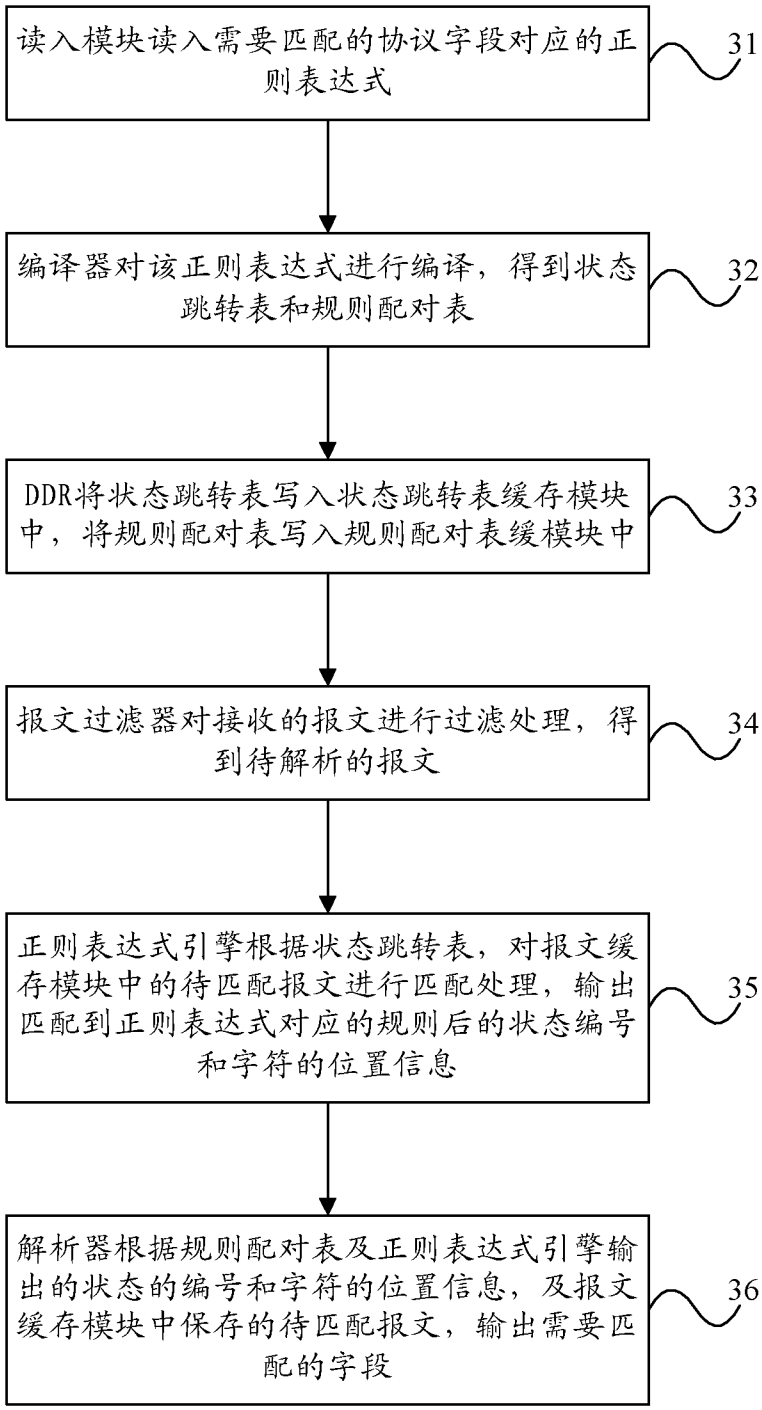

读入模块读入需要匹配的协议字段对应的正则表达式
31
编译器对该正则表达式进行编译，得到状态跳转表和规则配对表
32
DDR将状态跳转表写入状态跳转表缓存模块中，将规则配对表写入规则配对表缓模块中
33
报文过滤器对接收的报文进行过滤处理，得到待解析的报文
34
正则表达式引擎根据状态跳转表，对报文缓存模块中的待匹配报文进行匹配处理，输出匹配到正则表达式对应的规则后的状态编号和字符的位置信息
35
解析器根据规则配对表及正则表达式引擎输出的状态的编号和字符的位置信息，及报文缓存模块中保存的待匹配报文，输出需要匹配的字段
36

图 3

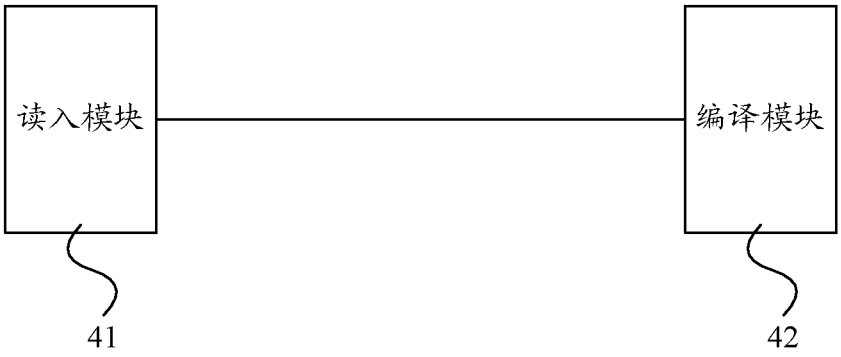

读入模块
编译模块
41
42

图 4

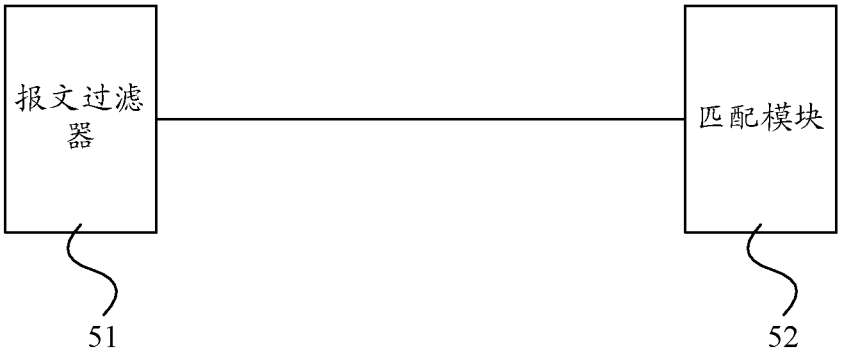

报文过滤器
匹配模块
51
52

图 5

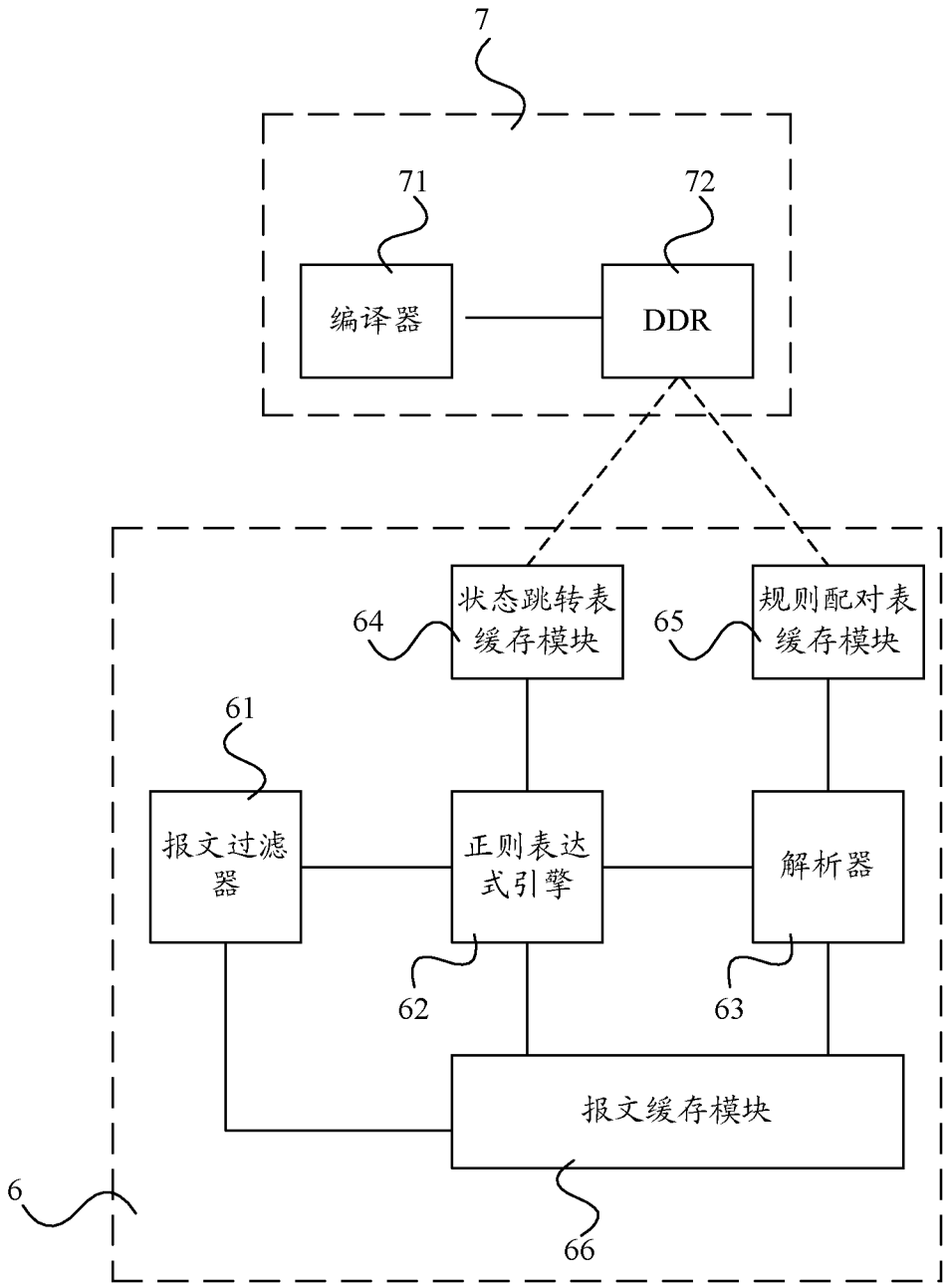

7
71
72
编译器
DDR
状态跳转表缓存模块
64
规则配对表缓存模块
65
61
报文过滤器
正则表达式引擎
解析器
62
63
报文缓存模块
6
66

图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2011/080795

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/06 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC: H04L, H04N, G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
CNABS, CNTXT, CNKI, VEN: protocol, analysis, regular, match, general, universal, parameter, obtain, acquire, regular w expression, DFA, DPA, rule, state w jump w table, rule w pairing w table, rule w matching w table

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN102143148A (HUAWEI TECH CO LTD) 03 Aug. 2011 (03.08.2011) description, Paragraphs 29-41, 94-100	1-9
A	CN101360088A (HUAWEI TECH CO LTD) 04 Feb.2009 (04.02.2009) description, Page 6, line 12-page 8, line 23	1-9
A	US2007265833A1 (NAKAYAMA AKIHIRO) 15 Nov. 2007 (15.11.2007) The whole document	1-9
A	US2010281160A1 (RESERVOIR LABS INC) 04 Nov. 2010 (04.11.2010) The whole document	1-9
A	CN101841546A (HUAWEI TECH CO LTD) 22 Sep. 2010 (22.09.2010) The whole document	1-9
A	CN101820418A (SHANGHAI BOCOM GROUP CO LTD) 01 Sep. 2010 (01.09.2010) The whole document	1-9

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
06 Jan. 2011(06.01.2011)	02 Feb. 2011(02.02.2011)

Name and mailing address of the ISA
State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao
Haidian District, Beijing 100088, China
Facsimile No. (86-10)62019451

Authorized officer
LIANG, Wei
Telephone No. (86-10)010-62412146

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2011/080795

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN102143148A	03.08.2011	NONE	
CN101360088A	04.02.2009	CN101360088B	14.09.2011
		WO2009015603A1	05.02.2009
		EP2184687A1	12.05.2010
		US2010131935A1	27.05.2010
US2007265833A1	15.11.2007	US7684976B2	23.03.2010
US2010281160A1	04.11.2010	WO2010127173A2	04.11.2010
		WO2010127173A3	03.02.2011
CN101841546A	22.09.2010	NONE	
CN101820418A	01.09.2010	NONE	

国际检索报告

国际申请号
PCT/CN2011/080795

A. 主题的分类

H04L 29/06 (2006.01) i

按照国际专利分类(IPC)或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: H04L, H04N, G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用）)

CNABS, CNTXT, CNKI: 协议，解析，正则，匹配，通用，参数，获取，获得，正则表达式，状态跳转表，规则配对表，DFA，DPA，规则; VEN: protocol, analysis, match, regular expression, DFA, DPA, obtain, acquire, rule

C. 相关文件

类　型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN102143148A (华为技术有限公司) 03.8 月 2011 (03.08.2011) 说明书第 29-41 段，第 94-100 段	1-9
A	CN101360088A（华为技术有限公司）04.2 月 2009（04.02.2009） 说明书第 6 页第 12 行至第 8 页第 23 行	1-9
A	US2007265833A1 (NAKAYAMA AKIHIRO) 15.11 月 2007 (15.11.2007) 全文	1-9
A	US2010281160A1 (RESERVOIR LABS INC)04.11 月 2010 (04.11.2010)全文	1-9
A	CN101841546A (华为技术有限公司) 22.9 月 2010 (22.09.2010) 全文	1-9
A	CN101820418A (上海博康智能网络科技有限公司) 01.9 月 2010 (01.09.2010) 全文	1-9

☐ 其余文件在 C 栏的续页中列出。　　☒ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
06.1 月 2011(06.01.2011)	**02.2 月 2012 (02.02.2012)**
ISA/CN 的名称和邮寄地址： 中华人民共和国国家知识产权局 中国北京市海淀区蓟门桥西土城路 6 号 100088 传真号：　(86-10)62019451	受权官员 梁玮 电话号码：(86-10) **010-62412146**

国际检索报告

关于同族专利的信息

国际申请号
PCT/CN2011/080795

检索报告中引用的专利文件	公布日期	同族专利	公布日期
CN102143148A	03.08.2011	无	
CN101360088A	04.02.2009	CN101360088B	14.09.2011
		WO2009015603A1	05.02.2009
		EP2184687A1	12.05.2010
		US2010131935A1	27.05.2010
US2007265833A1	15.11.2007	US7684976B2	23.03.2010
US2010281160A1	04.11.2010	WO2010127173A2	04.11.2010
		WO2010127173A3	03.02.2011
CN101841546A	22.09.2010	无	
CN101820418A	01.09.2010	无	