

(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
2006 年 7 月 27 日 (27.07.2006)

(10) 国际公布号
WO 2006/076869 A1

(51) 国际专利分类号:
H04L 12/26 (2006.01) G06F 17/27 (2006.01)
H04L 29/06 (2006.01)

(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)。中国科学技术大学(UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA) [CN/CN]; 中国安徽省合肥市金寨路96号, Anhui 230026 (CN)。

(21) 国际申请号: PCT/CN2006/000118

(22) 国际申请日: 2006 年 1 月 23 日 (23.01.2006)

(25) 申请语言: 中文

(26) 公布语言: 中文

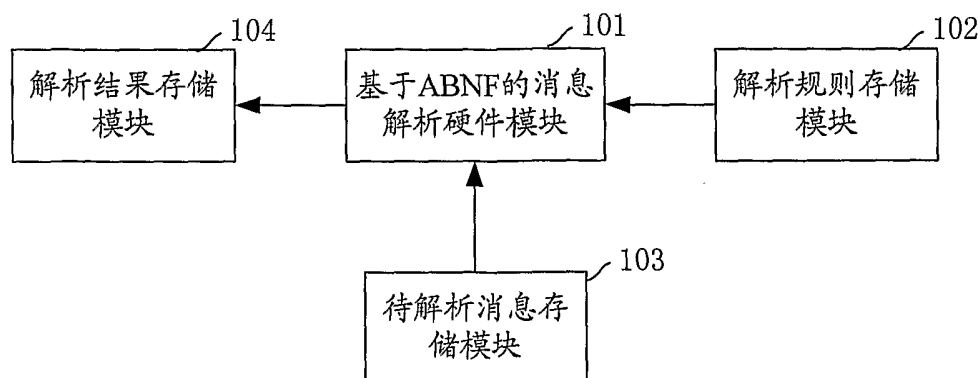
(30) 优先权:
200510001782.1
2005 年 1 月 21 日 (21.01.2005) CN

(72) 发明人; 及
(75) 发明人/申请人 (仅对美国): 赵宝华(ZHAO, Bao-hua) [CN/CN]; 中国安徽省合肥市金寨路96号, Anhui 230026 (CN)。屈玉贵(QU, Yugui) [CN/CN]; 中国安徽省合肥市金寨路96号, Anhui 230026 (CN)。周颢(ZHOU, Hao) [CN/CN]; 中国安徽省合肥市金寨路96号, Anhui 230026 (CN)。田野(TIAN, Ye) [CN/CN];

[见续页]

(54) Title: A TEXT CODING TYPE PROTOCOL PARSER

(54) 发明名称: 文本编码类协议解析器



101... THE MESSAGE PARSING HARDWARE MODULE BASED ON
ABNF
102... THE PARSING RULE STORAGE MODULE
103... THE READY-TO-PARSE MESSAGE STORAGE MODULE
104... THE PARSING RESULT STORAGE MODULE

(57) Abstract: A text coding type protocol parser comprises the parsing rule storage module and the message parsing hardware module. Based on the rule information stored in the parsing rule storage module, the message parsing hardware module produced by the logic chip parses the ready-to-parse message text and outputs the parsing result. Contrasted with the software implementation mode, the invention's parsing efficiency is greatly improved, the efficiency of the product is improved at an advantage and the cost is reduced. Furthermore, the invention may parse based on the ABNF rule, which makes the invention has the good generalization and can be applied to the situation that rule is defined based on ABNF. Consequently, the development cost of the corresponding parser is greatly reduced and the development efficiency is improved.

[见续页]

WO 2006/076869 A1



中国安徽省合肥市金寨路96号, Anhui 230026 (CN)。
王烁(WANG, Shuo) [CN/CN]; 中国安徽省合肥市金寨路96号, Anhui 230026 (CN)。李奇越(LI, Qiyue) [CN/CN]; 中国安徽省合肥市金寨路96号, Anhui 230026 (CN)。吕超(LV, Chao) [CN/CN]; 中国安徽省合肥市金寨路96号, Anhui 230026 (CN)。靳志伟(JIN, Zhiwei) [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)。

(74) 代理人: 北京德琦知识产权代理有限公司(DEQI INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区知春路1号学院国际大厦7层, Beijing 100083 (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, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD,

MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告。
- 在修改权利要求的期限届满之前进行, 在收到该修改后 将重新公布。

所引用双字母代码及其它缩写符号, 请参考刊登在每期PCT公报期刊起始的“代码及缩写符号简要说明”。

(57) 摘要: 本发明公开了一种文本编码类协议解析器, 包括: 解析规则存储模块、和消息解析硬件模块; 所述消息解析硬件模块采用逻辑芯片实现, 其根据规则存储模块中保存的规则信息, 对待解析的消息文本进行解析处理, 输出解析结果。本发明与软件实现方式相比, 解析效率大为提高, 有利于提高产品效率, 降低成本。另外, 本发明可以基于ABNF规则进行解析, 使得本发明具有良好的通用性, 可以适用于各种基于ABNF进行规则定义的场所, 从而极大地降低了相应的解析器的开发成本, 提高了开发效率。

文本编码类协议解析器

技术领域

本发明涉及网络通信技术领域，尤其涉及一种文本编码类协议解析器。

5 发明背景

Internet工程任务组（IETF，Internet Engineering Task Force）组织使用扩展巴克斯范式（ABNF，Augmented BNF）定义了多个协议中的报文格式，例如会话起始协议（SIP，Session Initiation Protocol）等。在实现这些协议时，需要按照ABNF规则对报文进行分析和模式。

10 ABNF是IETF组织在RFC2234中定义的一个字符串模式匹配的文法定义。ABNF是在巴克斯范式（BNF）基础上的扩展，其与标准巴克斯范式的区别包括：命名规则，循环，选择，次序独立以及值域。

目前，常见的ABNF解析器的实现方案为：在通用微处理器上，使用软件来实现编译原理领域内的词法和语法解析算法。例如，常见的LL

15 （1）文法解析算法，该算法是一种自上而下的非递归预测分析器，它通过计算文法的首终结符集合（First集合）和后继终结符集合（Follow集合）从而构造出该文法的预测分析表，进而利用分析表进行推导。LL（1）文法是无二义的文法，且不含左递归。LL（1）中的第一个“L”表示从左到右地扫描输入串；第二个“L”表示产生最左推导；1代表在决定分析器的每步动作时向前看一个输入符号。

20 由于ABNF是基于文本定义的语法，所以在对变长的字符串解析时，其复杂度高于基于二进制的定长数据解析。上述方案的优点在于解析器开发代价小，开发效率高。但是，由于通用处理器并不是专门针对ABNF

解析进行设计的，因而其处理速度、吞吐量等性能参数往往受限。当需要高速处理能力时，例如，SIP软交换系统需要处理百万量级的连接，则为了达到性能要求需要采用更高级别的微处理器，这样，必然提高解析器产品的成本。

- 5 目前，采用的另一种ABNF解析方案是也是在通用微处理器上，针对特定领域内ABNF语法规则的特点设计特定的软件来进行处理。以SIP协议栈为例，常见的实现方式是根据SIP文法规则的特点，通过分割符将SIP报文分割为若干字段，再进行相应的语法规则检查。

10 在这一方案中针对特定语法规则进行设计，在效率上优于前面描述的技术方案。但是由于其针对特定规则进行处理，导致其通用性较差，针对新的语法规则必须重新进行设计，从而提高了解析器开发成本，导致开发效率较低。另一方面，由于其仍然是基于通用微处理器进行的，在需要高速处理能力的情况下仍然会存在前面描述的第一种方案所存在的问题。

15 发明内容

鉴于上述现有技术所存在的问题，本发明的目的是提供一种文本编码类协议解析器，提高解析器的解析效率，并进一步使解析器具有良好的通用性。

本发明的目的是通过以下技术方案实现的：

- 20 一种文本编码类协议解析器，包括：解析规则存储模块、和消息解析硬件模块；所述消息解析硬件模块采用逻辑芯片实现，其根据规则存储模块中保存的规则信息，对待解析的消息文本进行解析处理，输出解析结果。

其中，所述消息解析硬件模块采用的逻辑芯片可以为：现场可编程

逻辑门阵列FPGA或复杂可编程逻辑器件CPLD。

所述的解析规则存储模块可以为规则树存储模块，该模块中保存有基于扩展巴克斯范式 ABNF 方法定义的规则生成的规则树信息。

所述的规则树可以为不包含增式选择、序列组及可选序列结构的规则树，且所述的规则树为以各规则中的一个规则为顶点，采用递归的方式生成的规则树。

所述的消息解析硬件模块可以包含：

主进程处理单元：用于根据栈顶节点的类型控制分别进入解析终结符进程处理单元、匹配首终结符进程处理单元或匹配后继终结符进程处理单元的处理过程；

解析终结符进程处理单元：用于当栈顶结点为终结节点时，由主进程处理单元控制启动，并进行终结节点的终结符的解析处理，解析完成后通知主进程处理单元；

匹配首终结符进程处理单元：用于对当栈顶结点为非终结点时，且为首终结符集合时，由主进程处理单元控制启动，并进行首终结符的匹配处理，判断当前字符是否在节点首终结符集合内，匹配完成后通知主进程处理单元；

匹配后继终结符进程处理单元：用于对当栈顶结点为非终结点时，且为后继终结符集合时，由主进程处理单元控制启动，并进行后继终结符的匹配处理，判断当前字符是否在节点后继终结符集合内，匹配完成后通知主进程处理单元。

所述的消息解析硬件模块的各个进程处理单元可以采用状态机实现。

所述的主进程处理单元对应的状态机可以包含：

循环解析 PARSE_LOOP 状态，在该状态下进行消息报文的解析；

到达最大重复次数 REPEAT_MAX 状态, 在该状态下判断保存着规则树节点的节点栈的栈顶节点的类型是否为终结节点;

匹配终结符 TERM_MATCH 状态, 在该状态下进行终结符的匹配处理, 并启动解析终结符进程处理单元;

- 5 进入首终结符集合判断 IN_FIRST 状态, 在该状态下进行首终结符集合的判断处理, 并在确定存在首终结符集合时, 启动匹配首终结符进程处理单元;

所有的子节点已经处理 ALLSUB 状态, 当 IN_FIRST 状态下所有的子节点处理完成时进入该状态, 且在该状态下返回 PARSE_LOOP 状态;

- 10 后继终结符集判断 JUDGE_FOLLOW 状态, 在该状态下进行后继终结符集合的判断处理, 并在确定存在后继终结符集合时, 进入 FOLLOW 集合判断 IN_FOLLOW 状态, 启动匹配后继终结符进程处理单元, 如果不存在后继终结符集合时进入下一规则 NEXTRULE 状态, 并在节点栈非空时返回 PARSE_LOOP 状态。

- 15 所述的解析终结符进程处理单元的状态机可以包含:

字符串类型终结符 TERM_STRING 状态, 在该状态下对单个字符串循环进行单个字符的匹配处理;

字符编码连接类型终结符 TERM_CODE_LIST 状态, 在该状态下对字符编码连接循环进行单个字符的匹配处理;

- 20 终结符匹配成功 TERM_SUCCESS 状态和终结符匹配失败 TERM_FAIL 状态, 在该状态下通知主进程处理单元匹配处理结果。

所述的匹配首终结符进程处理单元的状态机可以包含:

列表循环处理 LIST_LOOP 状态, 在该状态下判断编码类型为字符串字符编码连接;

- 25 字符串类型 FIRST_STRING 状态, 在该状态下对单个字符串循环进

行单个字符的匹配处理;

字符编码连接类型终结符FIRST_CODE_LIST状态, 在该状态下对字符编码连接循环进行单个字符的匹配处理;

首终结符集合匹配成功FIRST_SUCCESS状态和首终结符集合匹配失败FIRST_FAIL状态, 在该状态下通知主进程处理单元匹配处理结果。

所述的匹配后继终结符进程处理单元可以包含:

列表循环处理LIST_LOOP状态, 在该状态下判断编码类型为字符串字符编码连接;

字符串类型FOLLOW_STRING状态, 在该状态下对单个字符串循环进行单个字符的匹配处理;

字符编码连接类型FOLLOW_CODE_LIST状态, 在该状态下对字符编码连接循环进行单个字符的匹配处理;

后继终结符匹配成功FOLLOW_SUCCESS状态和后继终结符匹配失败FOLLOW_FAIL状态, 在该状态下通知主进程处理单元匹配处理结果。

该解析器可以进一步包含: 存储了待解析报文的待解析消息存储模块; 所述消息解析硬件模块从该待解析消息存储模块中读取消息内容进行解析。

该解析器可以进一步包含: 解析结果存储模块;

所述消息解析硬件模块在解析完成后, 将解析结果存储到解析结果存储模块。

由上述的技术方案可以看出, 本发明提供的这种文本编码类协议解析器, 是基于逻辑芯片硬件实现的, 与软件实现方式相比, 解析效率大为提高, 有利于提高产品效率, 降低成本。

另外, 本发明可以基于ABNF规则进行解析, 使得本发明具有良好

的通用性，可以适用于各种基于ABNF进行规则定义的场所，例如，对SIP协议栈、可扩展标记语言（XML）的处理等等，进一步降低了相应的解析器的开发成本，提高了开发效率。

附图简要说明

5 图1为本发明一较佳实施例的文本编码类协议解析器的结构示意图；

图2为图1所示解析器中消息解析硬件模块101的内部结构和实现原理图；

图3为图2中的主进程处理单元201的状态机示意图；

10 图4为图2中的解析终结符进程处理单元202的状态机示意图；

图5为图2中的匹配首终结符进程处理单元203的状态机示意图；

图6为图2中的匹配后继终结符进程处理单元204的状态机示意图。

实施本发明的方式

为使本发明的目的、技术方案和优点更加清楚明白，下面结合实施
15 例和附图，对本发明进一步详细说明。

本发明的核心是基于逻辑芯片来实现硬件文本编码类协议解析器，通过硬件优化提高解析器的解码效率，并可以通过基于ABNF规则进行解析进一步使解析器具有良好的通用性。

以下举一个较佳实施例对本发明进行详细说明。

20 参见图1，图1为本发明一较佳实施例的文本编码类协议解析器的结构示意图。该解析器包含：基于ABNF的消息解析硬件模块101、解析规则存储模块102、待解析消息存储模块103和解析结果存储模块104。其中：

解析规则存储模块 102 中存储了 ABNF 方法定义的规则生成的未赋值的规则树信息，通常将规则树节点按顺序存储，使基于 ABNF 的消息解析硬件模块 101 一次可以读出整个规则树所有节点。

本实施例中，解析规则存储模块 102 是一种规则树存储模块，其中
5 存储的规则树信息可以按如下方式从 ABNF 规则文本生成：

(1) 对 ABNF 规则进行展平处理，即将各规则进行简化描述，使得其不包含增式选择、序列组和可选序列等复杂结构，具体可以将简化描述后的规则保存于规则表中，每个表项对应一个简化处理后的新的规则，相应的处理步骤如下所述：

10 对于每个规则，在规则表（如哈希表）中添加一条规则表项；

对于包含增式选择的规则，将规则添加的选择内容加入到原规则中，更新原规则表的表项；

对于包含序列组的规则，将序列组内容定义为新规则，并用新规则替换原规则的序列组内容，重复此过程，直到新规则不包含序列组为止；

15 对于包含可选序列的规则，将可选序列内容定义为新规则，并用新规则替换原规则的可选序列内容，重复此过程，直到新规则不包含序列组为止；

(2) 基于所述的规则表生成相应的规则树，具体为指定规则表中的某一规则为顶点，并以此为根节点递归生成对应的规则树；在每个规则
20 节点存储的信息中，都包含了该规则在 ABNF 范式中规定的最大匹配重复次数。

(3) 使用 LL(1) 解析算法中 First 集合和 Follow 集合的生成方法，对每个规则树节点计算其相应的 First 集合和 Follow 集合，并保存 First 集合列表和 Follow 集合列表。

25 待解析消息存储模块 103 中的存储了消息文本，通常可以以字节形

式顺序存放。

基于 ABNF 的消息解析硬件模块 101 从待解析消息存储模块 103 中读取待解析消息文本中的字符串，从解析规则存储模块 102 获得规则树信息，用待解析消息文本中的字符串对规则树进行赋值，即对解析消息
5 进行解析，生成已经赋值规则树，作为解析结果输出给解析结果存储模块 104 存储。

本实施例是基于 ABNF 的文本编码类协议解析器，采用 ABNF 规则进行解析，可以适用于各种基于 ABNF 进行规则定义的场所，例如，对 SIP 协议栈、可扩展标记语言（XML）的处理等等，因此本实施例具有
10 良好的通用性。

本实施例中的基于 ABNF 的消息解析硬件模块 101 为核心处理部分，可以由现场可编程逻辑门阵列（FPGA）或复杂可编程逻辑器件（CPLD）等硬件实现，下面将对该消息解析硬件模块 101 的具体实现方式进行详细说明。

参见图 2，图 2 为图 1 所示解析器中消息解析硬件模块 101 的内部结构和实现原理图。本实施例中的消息解析硬件模块 101，主要包括四个进程处理单元：主进程（main）处理单元 201、解析终结符（parse_term）进程处理单元 202、匹配首终结符进程（match_first_proc）处理单元 203 和匹配后继终结符进程（match_follow_proc）处理单元 204，分别由四
15 20 个状态机实现。

其中，主进程（main）处理单元 201 控制解析终结符（parse_term）进程处理单元 202、匹配首终结符进程（match_first_proc）处理单元 203 和匹配后继终结符进程（match_follow_proc）处理单元 204 这三个子进程处理单元的启动，并实现相应的功能。

25 主进程处理单元 201 根据外部发送的解析控制信号，开始进行解析。

其在判断符合启动相应的子进程的条件时，迁移到相应的状态，向相应的子进程处理单元发送信号启动对应的子进程进行相应的处理，各子进程处理单元处理完成后将结果返回主进程处理单元201，并由主进程处理单元201输出成功或失败的解析结果信号给外部。

- 5 解析终结符进程处理单元202，用于当栈顶结点为终结节点时，由主进程处理单元控制启动，并进行终结节点的终结符的解析处理，解析完成后通知主进程处理单元201。

- 匹配首终结符进程处理单元203，用于对当栈顶结点为非终结点时，且为首终结符集合时，由主进程处理单元201控制启动，并进行首终结
10 符的匹配处理，判断当前字符是否在节点首终结符集合内，匹配完成后通知主进程处理单元201

- 匹配后继终结符进程处理单元204：用于对当栈顶结点为非终结点时，且为后继终结符集合时，由主进程处理单元201控制启动，并进行后继终结符的匹配处理，判断当前字符是否在节点后继终结符集合内，
15 匹配完成后通知主进程处理单元201。

下面将分别对各个进程处理单元的状态机包括的状态及具体处理过程进行说明。

参见图3，图3为图2中的主进程处理单元201的状态机示意图。

- 主进程处理单元201作为主处理进程，由外部输入的时钟上升沿驱
20 动，如图3所示，包含以下状态：

- 空闲 (IDLE)，初始化 (INIT)，循环解析 (PARSE_LOOP)，到达最大重复次数 (REPEAT_MAX)，匹配终结符 (TERM_MATCH)，进行FIRST集合判断 (IN_FIRST)，第一个子节点 (FIRSTSUB)，所有的子节点已经处理 (ALLSUB)，选择子节点 (CHOOSE SUB)，下一个子节点0 (NEXTSUB0)，下一个子节点1 (NEXTSUB1)，取子节点 (GETSUB)，
25 点0 (NEXTSUB0)，下一个子节点1 (NEXTSUB1)，取子节点 (GETSUB)，

FOLLOW 集判断 (JUDGE_FOLLOW), 进行 FOLLOW 集合判断 (IN_FOLLOW), 下一规则 (NEXTRULE), 解析结束 (END_PARSE).

相应的状态迁移处理过程为:

当检测到外部输入的时钟上升沿驱动信号时, 状态机由 IDLE 状态先
5 进入 INIT 状态, 对硬件进行初始化, 并将规则树根的节点, 在压入节点
栈中。所述的节点栈是在主进程实现时的一个节点栈, 保存的内容是规
则树的所有节点。初始化完成进入 PARSE_LOOP 状态。

在 PARSE_LOOP 状态, 从节点栈的栈顶开始, 逐个中读出节点信息,
开始消息报文的解析, 每读出一个节点信息进入 REPEAT_MAX 状态。

10 当没有未匹配报文字符且当前节点是根节点时, 进入 END_PARSE 状态。

状态机在 REPEAT_MAX 状态, 先判断规则节点是否已经达到最大
匹配重复次数, 如果是, 则返回 PARSE_LOOP 状态, 准备对下一个规则
节点进行匹配; 否则判断读出的规则节点的类型, 这里, 节点栈用于跟
踪当前基于相应的规则树的解析进度情况, 并根据栈顶节点的类型确定
15 进入相应的子进程处理单元进行处理, 参照图2所示, 具体如下:

如果是终结节点, 也就是展平规则的叶子节点, 则将终结符匹配信
号 match_term 代入高电平, 启动 parse_term 进程, 等待终结符匹配结果
match_term_result 信号出现高电平后进入状态 TERM_MATCH, 并将信号
match_term 代入低电平以中止 parse_term 进程, 状态机返回 PARSE_LOOP
20 状态;

如果栈顶节点是非终结节点, 则将信号 match_first 代入高电平, 启
动 match_first_proc 进程, 具体的处理过程为:

若 match_first_success 信号出现高电平, 表示当前待解析字符在节点
的 first 集合中, 状态机进入状态 FIRST_SUB, 并把信号 match_first 代入低
25 电平以中止 match_first_proc 进程, 此时状态机根据子节点的 first 集合选取

正确的子节点压入栈中并返回PARSE_LOOP状态。具体来说，就是根据子节点的first集合检查每个子节点，如果当前报文字符内容存在于某个子节点的first集合，则把该子节点压入栈中，尝试匹配。

在FIRSTSUB状态，判断是否有子节点，如果有进入CHOOSESUB
5 状态，否则进入ALLSUB状态。在ALLSUB状态，表示所有子节点都已经处理，返回PARSE_LOOP状态。

在CHOOSESUB状态，状态机按顺序读取一个子节点，依据子节点的first集合判断该子节点是否是正确的子节点。这里判断是为了提高效率，否则进程将尝试匹配每个子节点，直到正确匹配一个或者所有尝试
10 匹配失败，如果是，则进入GETSUB状态；否则进入NEXTSUB0状态。

在NEXTSUB0状态，状态机读取下一个子节点0，判断该子节点是否是正确的子节点，如果是，则进入GETSUB状态；否则进入NEXTSUB1状态。

在NEXTSUB1状态，状态机读取下一个子节点1，判断该子节点是
15 否是正确的子节点，如果是，则CHOOSESUB通过状态进入GETSUB状态；否则返回PARSE_LOOP状态。

在GETSUB状态，状态机将选择的子节点压入栈中并返回PARSE_LOOP状态。

若match_first_fail信号出现高电平，表示当前待解析字符不在节点的
20 first集合中，状态机进入状态JUDGE_FOLLOW状态。

在JUDGE_FOLLOW状态下，信号match_follow被代入高电平，启动match_follow_proc进程，等待match_follow_success信号或者match_follow_fail信号以判断能否进行下一节点的匹配，如果后续报文字符在当前规则节点的follow集合中，进入IN_FOLLOW状态；否则进入
25 NEXTRULE状态。

在NEXTRULE状态下，即检查follow集合失败。这时表示报文解析失败，报文不符合ABNF文法，这时进入END_PARSE状态结束解析，返回解析失败信号。

在IN_FOLLOW状态，如果检查已匹配的的节点的follow集合成功，
5 则返回PARSE_LOOP状态，从栈顶POP已经成功匹配的规则节点，匹配新的栈顶节点，即下一个节点。

参见图4，图4为图2中的解析终结符进程处理单元202的状态机示意图。子进程parse_term为时钟上升沿驱动，用于完成终结节点的匹配，其对应的状态包括：

10 IDLE，启动（START），字符串类型终结符（TERM_STRING），字符编码连接类型终结符（TERM_CODE_LIST），终结符匹配成功（TERM_SUCCESS），终结符匹配失败（TERM_FAIL）。

如图4所示，当match_term信号被主进程代入高电平，状态机由IDLE状态进入START状态，在此状态下状态机从待解析消息存储模块逐字节读入待解析报文，并判断该报文的编码类型，根据不同的编码类型进入不同的状态进行处理：
15

如果是字符串，则进入TERM_STRING状态，循环匹配单个字符，那当然，如果字符串仅包含单个字符编码或者单个字符的有效范围（range）编码，则直接判断是否匹配成功，根据匹配结果进入
20 TERM_SUCCESS状态或TERM_FAIL状态。

如果是字符编码连接，则进入TERM_CODE_LIST状态，循环进行单个字符的匹配。

在TERM_STRING状态或TERM_CODE_LIST状态匹配完后，根据匹配结果进入状态 TERM_SUCCESS 或者 TERM_FAIL，将
25 term_match_result代入高电平通知主状态机。

当主状态机将math_term置为低电平,则parse_term状态机回到IDLE状态,清除term_match_result信号等待下一次匹配。

图5为图2中的匹配首终结符进程处理单元203的状态机示意图。进程match_first_proc为时钟上升沿驱动,用于比较当前字符是否存在于节点first集合中,相应的状态包括:

IDLE,列表循环处理(LIST_LOOP),字符串类型(FIRST_STRING),字符编码连接类型(FIRST_CODE_LIST),FIRST集合匹配成功(FIRST_SUCCESS),FIRST集合匹配失败(FIRST_FAIL)。

如图5所示,当match_first信号被主进程代入高电平,状态机由IDLE状态进入LIST_LOOP状态,在此状态下状态机判断编码类型,并根据编码类型采用相应的处理:

如果是字符串,则进入FIRST_STRING状态,循环匹配单个字符。

如果是单个字符编码或者单个字符的range编码,则直接判断是否匹配,根据匹配结果进入FIRST_SUCCESS状态或FIRST_FAIL状态。

如果是字符编码连接,则进入FIRST_CODE_LIST状态,循环匹配单个字符。

在FIRST_STRING状态匹配完成后,根据匹配结果选择进入FIRST_SUCCESS状态或FIRST_FAIL状态。

在FIRST_CODE_LIST状态匹配完成后,根据匹配结果选择进入状态FIRST_SUCCESS或FIRST_FAIL状态或回到LIST_LOOP匹配下一个列表内容直到发现匹配的first集合元素或者到达列表尾部。

如果没有发现匹配的first元素,则状态机进入FIRST_FAIL状态;状态机在FIRST_SUCCESS和FIRST_FAIL分别把first_match_success和first_match_fail代入高电平通知主状态机。

当主状态机将match_first置为低电平,则match_first_proc状态机回到

IDLE状态，清除match_first_success或match_first_fail信号，并等待下一次匹配。

图6为图2中的匹配后继终结符进程处理单元204的状态机示意图。
进程match_follow_proc时钟上升沿驱动，用于比较当前字符是否存在于
5 节点follow集合中，相应的状态包括：

IDLE，列表循环处理（LIST_LOOP），字符串类型（FOLLOW_STRING），字符编码连接类型（FOLLOW_CODE_LIST），FOLLOW匹配成功（FOLLOW_SUCCESS），FOLLOW匹配失败（FOLLOW_FAIL）；

10 如图6所示，当match_follow信号被主进程代入高电平，状态机由IDLE状态进入LIST_LOOP状态，在此状态下状态机判断编码类型，同样需要根据编码类型进行不同的处理：

如果是字符串，则进入FOLLOW_STRING状态，循环匹配单个字符。

如果是单个字符编码或者单个字符的range编码，则直接判断是否匹
15 配，根据匹配结果进入FOLLOW_SUCCESS状态或FOLLOW_FAIL状态。

如果是字符编码连接，则进入FOLLOW_CODE_LIST状态，循环匹配单个字符。

在FOLLOW_STRING状态匹配完成后，根据匹配结果选择进入FOLLOW_SUCCESS状态或FOLLOW_FAIL状态。

20 在FOLLOW_CODE_LIST状态匹配完成后，根据匹配结果选择进入状态FOLLOW_SUCCESS状态或FOLLOW_FAIL状态或回到LIST_LOOP匹配下一个列表内容，直到发现匹配的FOLLOW集合元素或者到达列表尾部。

如果没有发现匹配的FOLLOW元素，则状态机进入FOLLOW_FAIL
25 状态；状态机在FOLLOW_SUCCESS和FOLLOW_FAIL分别将

follow_match_success和follow_match_fail代入高电平通知主状态机。

当主状态机将match_follow置为低电平,则match_follow_proc状态机回到IDLE状态,清除match_follow_success或match_follow_fail信号,等待下一次匹配。

- 5 基于上述各状态机,则可以利用现场可编程逻辑门阵列(FPGA)或复杂可编程逻辑器件(CPLD)等硬件逻辑芯片来实现消息解析硬件模块,从而实现通用的基于ABNF文法的文本编码类协议解析器。

10 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。

权利要求书

1、一种文本编码类协议解析器，其特征在于，包括：解析规则存储模块、和消息解析硬件模块；所述消息解析硬件模块采用逻辑芯片实现，其根据规则存储模块中保存的规则信息，对待解析的消息文本进行
5 解析处理，输出解析结果。

2、根据权利要求1所述的文本编码类协议解析器，其特征在于：所述消息解析硬件模块采用的逻辑芯片为：现场可编程逻辑门阵列FPGA或复杂可编程逻辑器件CPLD。

3、根据权利要求1所述的文本编码类协议解析器，其特征在于：所
10 述的解析规则存储模块为规则树存储模块，该模块中保存有基于扩展巴克斯范式 ABNF 方法定义的规则生成的规则树信息。

4、根据权利要求3所述的文本编码类协议通用解析器，其特征在于，所述的规则树为不包含增式选择、序列组及可选序列结构的规则树，且所述的规则树为以各规则中的一个规则为顶点，采用递归的方式生成的
15 规则树。

5、根据权利要求1、2、3或4所述的文本编码类协议通用解析器，其特征在于，所述的消息解析硬件模块包含：

主进程处理单元：用于根据栈顶节点的类型控制分别进入解析终结符进程处理单元、匹配首终结符进程处理单元或匹配后继终结符进程处
20 理单元的处理过程；

解析终结符进程处理单元：用于当栈顶结点为终结节点时，由主进程处理单元控制启动，并进行终结节点的终结符的解析处理，解析完成后通知主进程处理单元；

匹配首终结符进程处理单元：用于对当栈顶结点为非终结点时，且

为首终结符集合时，由主进程处理单元控制启动，并进行首终结符的匹配处理，判断当前字符是否在节点首终结符集合内，匹配完成后通知主进程处理单元；

匹配后继终结符进程处理单元：用于对当栈顶结点为非终结点时，
5 且为后继终结符集合时，由主进程处理单元控制启动，并进行后继终结符的匹配处理，判断当前字符是否在节点后继终结符集合内，匹配完成后通知主进程处理单元。

6、根据权利要求5所述的文本编码类协议通用解析器，其特征在于：
所述的消息解析硬件模块的各个进程处理单元采用状态机实现。

10 7、根据权利要求6所述的文本编码类协议通用解析器，其特征在于：
所述的主进程处理单元对应的状态机包含：

循环解析 PARSE_LOOP 状态，在该状态下进行消息报文的解析；

到达最大重复次数 REPEAT_MAX 状态，在该状态下判断保存着规则树节点的节点栈的栈顶节点的类型是否为终结节点；

15 匹配终结符 TERM_MATCH 状态，在该状态下进行终结符的匹配处理，并启动解析终结符进程处理单元；

进入首终结符集合判断 IN_FIRST 状态，在该状态下进行首终结符集合的判断处理，并在确定存在首终结符集合时，启动匹配首终结符进程处理单元；

20 所有的子节点已经处理 ALLSUB 状态，当 IN_FIRST 状态下所有的子节点处理完成时进入该状态，且在该状态下返回 PARSE_LOOP 状态；

后继终结符集判断 JUDGE_FOLLOW 状态，在该状态下进行后继终结符集合的判断处理，并在确定存在后继终结符集合时，进入 FOLLOW 集合判断 IN_FOLLOW 状态，启动匹配后继终结符进程处理单元，如果

25 不存在后继终结符集合时进入下一规则 NEXTRULE 状态，并在节点栈

非空时返回 PARSE_LOOP 状态。

8、根据权利要求6所述的文本编码类协议通用解析器，其特征在于，所述的解析终结符进程处理单元的状态机包含：

字符串类型终结符TERM_STRING状态，在该状态下对单个字符串
5 循环进行单个字符的匹配处理；

字符编码连接类型终结符TERM_CODE_LIST状态，在该状态下对字符编码连接循环进行单个字符的匹配处理；

终结符匹配成功 TERM_SUCCESS 状态和终结符匹配失败 TERM_FAIL状态，在该状态下通知主进程处理单元匹配处理结果。

10 9、根据权利要求6所述的文本编码类协议通用解析器，其特征在于，所述的匹配首终结符进程处理单元的状态机包含：

列表循环处理LIST_LOOP状态，在该状态下判断编码类型为字符串字符编码连接；

字符串类型FIRST_STRING状态，在该状态下对单个字符串循环进
15 行单个字符的匹配处理；

字符编码连接类型终结符FIRST_CODE_LIST状态，在该状态下对字符编码连接循环进行单个字符的匹配处理；

首终结符集合匹配成功FIRST_SUCCESS状态和首终结符集合匹配失败FIRST_FAIL状态，在该状态下通知主进程处理单元匹配处理结果。

20 10、根据权利要求6所述的文本编码类协议通用解析器，其特征在于，所述的匹配后继终结符进程处理单元包含：

列表循环处理LIST_LOOP状态，在该状态下判断编码类型为字符串字符编码连接；

字符串类型FOLLOW_STRING状态，在该状态下对单个字符串循环
25 进行单个字符的匹配处理；

字符编码连接类型FOLLOW_CODE_LIST状态，在该状态下对字符编码连接循环进行单个字符的匹配处理；

后继终结符匹配成功FOLLOW_SUCCESS状态和后继终结符匹配失败FOLLOW_FAIL状态，在该状态下通知主进程处理单元匹配处理结果。

11、根据权利要求 1 所述的文本编码类协议解析器，其特征在于：该解析器进一步包含：存储了待解析报文的待解析消息存储模块；所述消息解析硬件模块从该待解析消息存储模块中读取消息内容进行解析。

12、根据权利要求 1 或 12 所述的文本编码类协议解析器，其特征在于：该解析器进一步包含：解析结果存储模块；

所述消息解析硬件模块在解析完成后，将解析结果存储到解析结果存储模块。

1/4

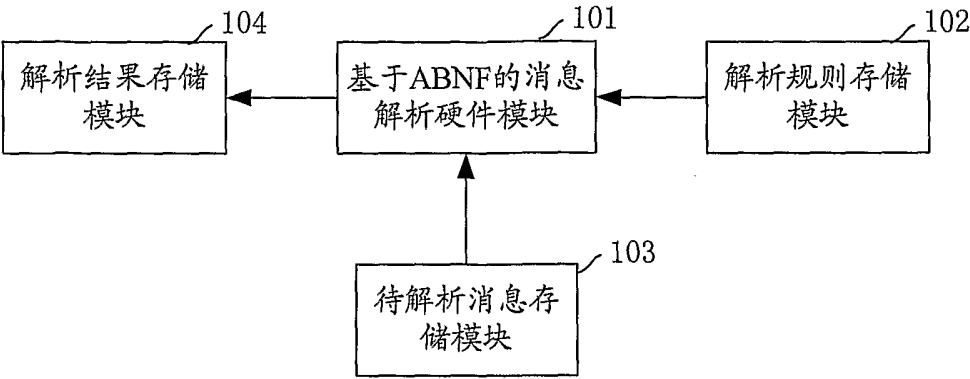


图 1

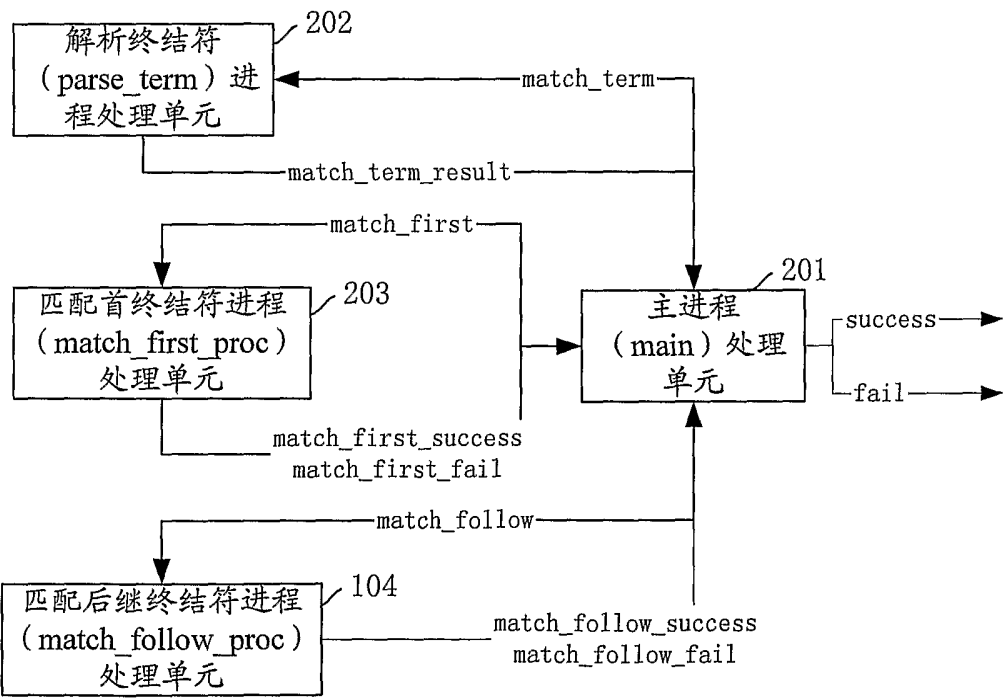


图 2

2/4

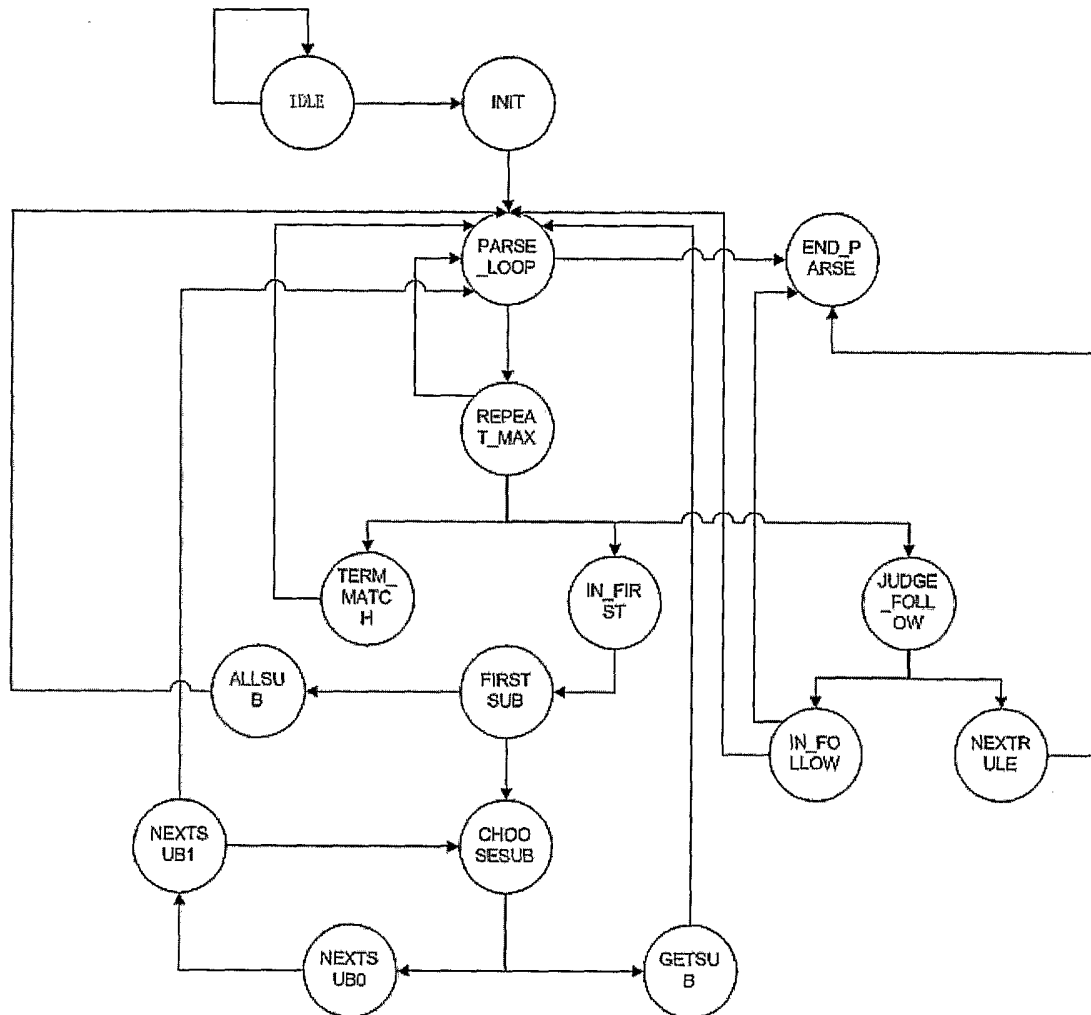


图 3

3/4

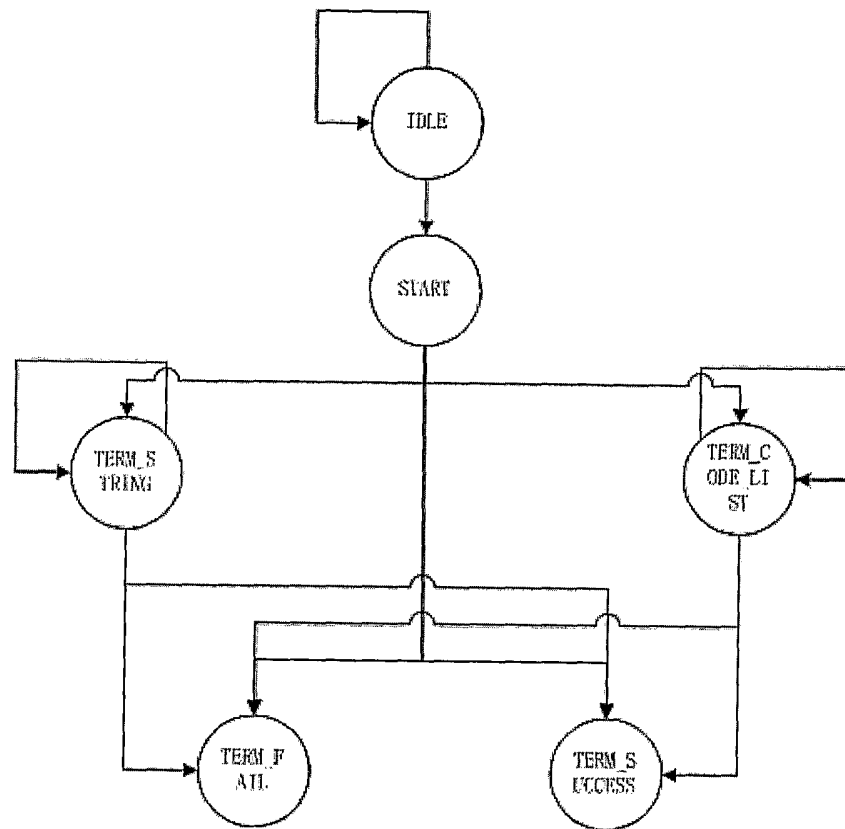


图 4

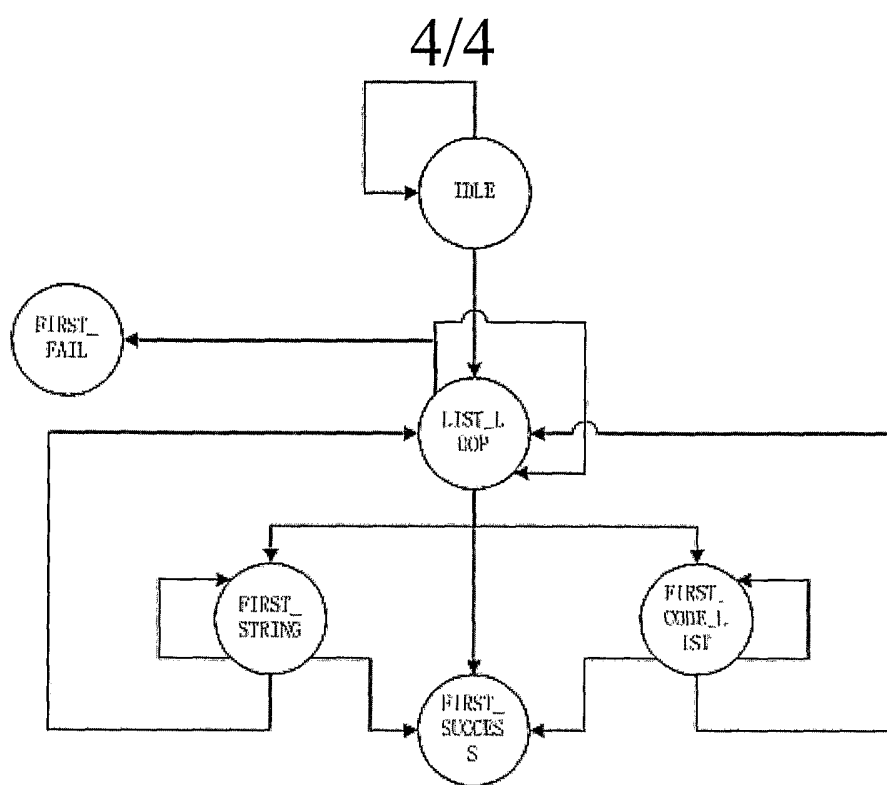


图 5

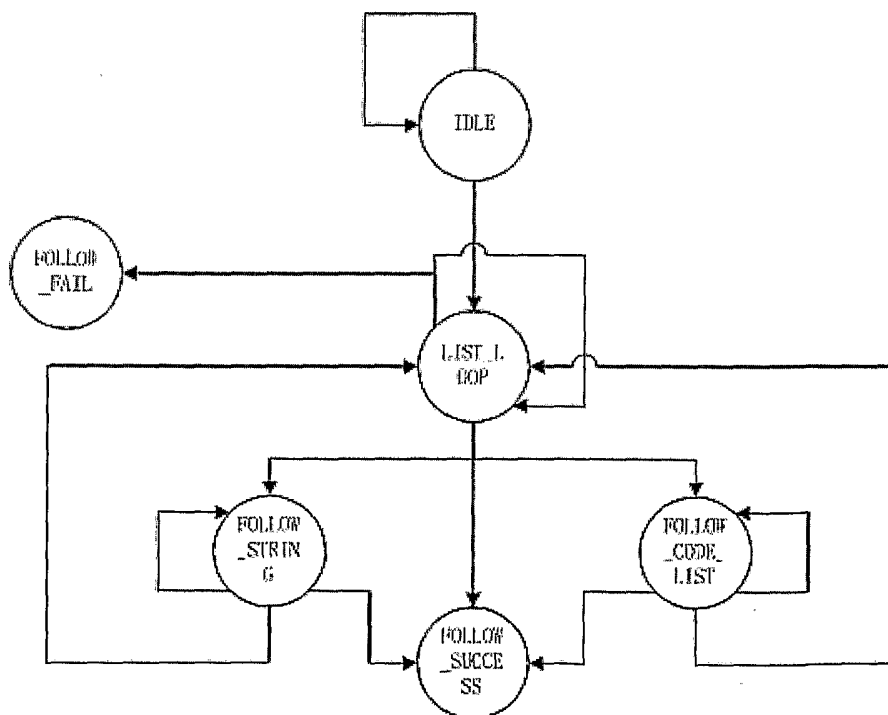


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2006/000118

A. CLASSIFICATION OF SUBJECT MATTER See extra sheet 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) H04L 12 (2006.01) H04L 29 (2006.01) G06F 17/27 (2006.01) 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) WPI, EPODOC, PAJ, CNPAT, CNKI: PROTOCOL W (PARSER OR RESOLVER OR ANALYZER OR ANALYSER) HARDWARE OR (LOGIC W (DEVICE OR CHIP OR CIRCUIT))		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US, A, 5916305 (SHOMITI SYSTEMS INC) 29. Jun. 1999 (29. 06. 1999) see in particular column_6 line_13 to line_35, and column_16 line_42 to line_48	1-4,11-12
A	JP, A, 2002016662 (TEKTRONIX INC) 18. Jan. 2002 (18. 01. 2002) See the whole document	1-12
A	EP, A2, 1420548 (TEXAS INSTR INC) 19. May 2004 (19. 05. 2004) See the whole document	1-12
A	JP, A, 2003163712 (HITACHI MICON SYSTEM KK) 06. Jun. 2003 (06. 06. 2003) See the whole document	1-12
☐ 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 04. Apr. 2006 (04. 04. 2006)		Date of mailing of the international search report 18 · MAY 2006 (18 · 05 · 2006)
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451		Authorized officer WANG Zhiwei Telephone No. 86-10-62084532 

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2006/000118

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP2002016662 A	18. 01. 2002	EP1146690 A1	17. 10. 2001
		US2001032069 A1	18. 10. 2001
US5916305 A	29. 06. 1999	None	
EP1420548 A2	19. 05. 2004	US2004098641 A1	20. 05. 2004
		US2005141432 A	30. 06. 2005
JP2003163712 A	06. 06. 2003	None	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2006/000118

CLASSIFICATION OF SUBJECT MATTER

H04L 12/26 (2006.01) i

H04L 29/06 (2006.01) n

G06F 17/27 (2006.01) n

国际检索报告

国际申请号
PCT/CN2006/000118

A. 主题的分类

参见附加页

按照国际专利分类表(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)

H04L 12 (2006.01) H04L 29 (2006.01) G06F 17/27 (2006.01)

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))

WPI、EPDOC、PAJ、CNPAT、CNKI: 解析器 协议 分析器 PROTOCOL W (PARSER OR RESOLVER OR ANALYZER OR ANALYSER) HARDWARE OR (LOGIC W (DEVICE OR CHIP OR CIRCUIT))

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
X	US, A, 5916305 (SHOMITI SYSTEMS INC) 29.6 月 1999 (29.06.1999) 参见说明书第 6 列第 13-35 行, 第 16 列第 42-48 行	1-4,11-12
A	JP, A, 2002016662 (TEKTRONIX INC) 18.1 月 2002 (18.01.2002) 全文	1-12
A	EP, A2, 1420548 (TEXAS INSTR INC) 19.5 月 2004 (19.05.2004) 全文	1-12
A	JP, A, 2003163712 (HITACHI MICON SYSTEM KK) 06.6 月 2003 (06.06.2003) 全文	1-12

☐ 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

“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

04.4 月 2006 (04.04.2006)

国际检索报告邮寄日期

18.5 月 2006 (18.05.2006)

中华人民共和国国家知识产权局(ISA/CN)

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

王志伟

电话号码: (86-10)-62084532



国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2006/000118

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
JP2002016662 A	18. 01. 2002	EP1146690 A1 US2001032069 A1	17. 10. 2001 18. 10. 2001
US5916305 A	29. 06. 1999	无	
EP1420548 A2	19. 05. 2004	US2004098641 A1 US2005141432 A	20. 05. 2004 30. 06. 2005
JP2003163712 A	06. 06. 2003	无	

主题的分类

H04L 12/26 (2006.01) i

H04L 29/06 (2006.01) n

G06F 17/27 (2006.01) n