

(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

2022 年 5 月 5 日 (05.05.2022)



(10) 国际公布号

WO 2022/088779 A1

(51) 国际专利分类号:

H04L 29/06 (2006.01) H04L 29/08 (2006.01)

(21) 国际申请号:

PCT/CN2021/107642

(22) 国际申请日: 2021 年 7 月 21 日 (21.07.2021)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

202011173763.8 2020 年 10 月 28 日 (28.10.2020) CN

(71) 申请人: 北京锐安科技有限公司 (RUN TECHNOLOGIES CO., LTD. BEIJING) [CN/CN]; 中国北京市海淀区西小口路 66 号中关村东升科技园北领地 B-2 号楼七层, Beijing 100192 (CN)。

(72) 发明人: 孙晓 (SUN, Xiao); 中国北京市海淀区西小口路 66 号中关村东升科技园北领地 B-2 号楼七层, Beijing 100192 (CN)。 谢永恒 (XIE, Yongcheng); 中国北京市海淀区西小口路 66 号中关村东升科技园北领地 B-2 号楼七层, Beijing 100192 (CN)。

Yongcheng); 中国北京市海淀区西小口路 66 号中关村东升科技园北领地 B-2 号楼七层, Beijing 100192 (CN)。 万月亮 (WAN, Yueliang); 中国北京市海淀区西小口路 66 号中关村东升科技园北领地 B-2 号楼七层, Beijing 100192 (CN)。

(74) 代理人: 北京品源专利代理有限公司 (BEYOND ATTORNEYS AT LAW); 中国北京市海淀区莲花池东路 39 号西金大厦 6 层, Beijing 100036 (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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) Title: DEEP PACKET PROCESSING METHOD AND APPARATUS, ELECTRONIC DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 深度报文处理方法、装置、电子设备及存储介质

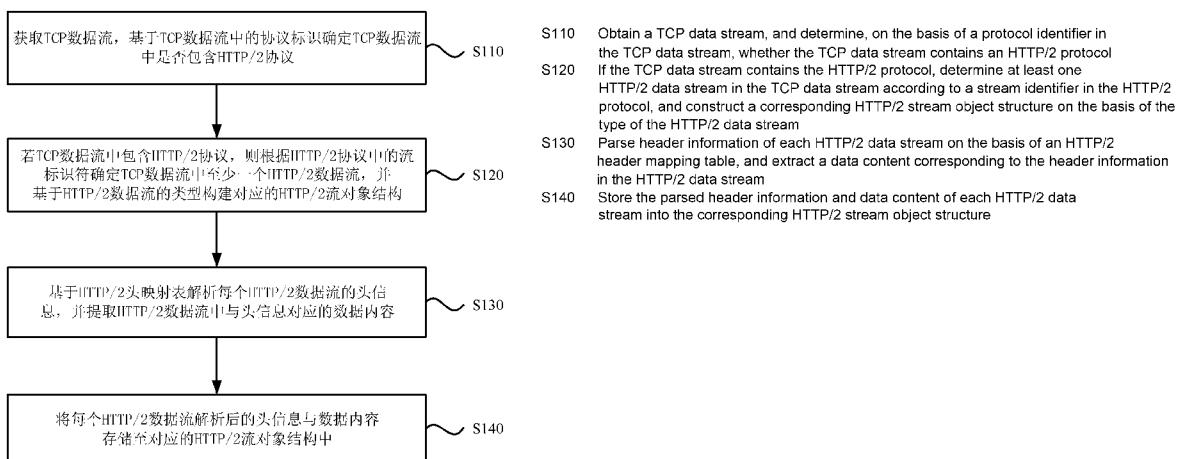


图 1

(57) Abstract: Disclosed in the present application are a deep packet processing method and apparatus, an electronic device, and a storage medium. The method comprises: obtaining a TCP data stream, and determining, on the basis of a protocol identifier in the TCP data stream, whether the TCP data stream contains an HTTP/2 protocol; if determining that the TCP data stream contains the HTTP/2 protocol, determining at least one HTTP/2 data stream in the TCP data stream according to a stream identifier in the HTTP/2 protocol, and constructing a corresponding HTTP/2 stream object structure on the basis of the type of the HTTP/2 data stream so as to obtain a plurality of HTTP/2 stream object structures; parsing header information of each HTTP/2 data stream on the basis of an HTTP/2 header mapping table, and extracting a data content corresponding to the header information in the HTTP/2 data stream; and storing the parsed header information and data content of each HTTP/2 data stream into the corresponding HTTP/2 stream object structure.

ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, WS, 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, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, 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, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 本申请公开了一种深度报文处理方法、装置、电子设备及存储介质。该方法包括：通过获取TCP数据流，基于TCP数据流中的协议标识确定TCP数据流中是否包含HTTP/2协议；在确定TCP数据流中包含HTTP/2协议的情况下，根据HTTP/2协议中的流标识符确定TCP数据流中至少一个HTTP/2数据流，基于HTTP/2数据流的类型构建对应的HTTP/2流对象结构，从而得到多个HTTP/2流对象结构；并基于HTTP/2头映射表解析每个HTTP/2数据流的头信息，并提取HTTP/2数据流中与所述头信息对应的数据内容，将每个HTTP/2数据流解析后的头信息与数据内容存储至对应的HTTP/2流对象结构中。

深度报文处理方法、装置、电子设备及存储介质

本申请要求在2020年10月28日提交中国专利局、申请号为202011173763.8的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

本申请实施例涉及通信技术领域，例如涉及一种深度报文处理方法、装置、电子设备及存储介质。

背景技术

为了协助运营商对网络流量的管控，DPI（Deep Packet Inspection，深度报文检测）方法在网络中被广泛应用。其中的“深度”是和普通报文分析层次相比较而言的，普通报文检测仅分析网际互联协议（Internet Protocol，IP）报文四层以下内容，包括源IP地址、目的IP地址、传输层源端口、传输层目的端口和承载协议类型，而DPI技术基于从二层到七层网络协议的分析，能够实现对网络中数据的精准感知，从而实现对网络现状的精准把握，如业务识别、业务统计、流量管控和网元分析等。

然而，DPI的本质是一种数据报文过滤技术，需要先对应用层有效载荷进行解析，从而根据业务需求对解析后的信息进行匹配过滤。因此，解析出应用层有效载荷信息至关重要。

发明内容

本申请提供一种深度报文处理方法、装置、电子设备及存储介质，以实现针对HTTP/2协议的报文处理，从而为预设检测规则的匹配提供了数据基础。

第一方面，本申请实施例提供了一种深度报文处理方法，包括：

获取TCP数据流，基于所述TCP数据流中的协议标识确定所述TCP数据流

中是否包含HTTP/2协议；

在确定TCP数据流中包含HTTP/2协议的情况下，根据所述HTTP/2协议中的流标识符确定所述TCP数据流中至少一个HTTP/2数据流，并基于所述HTTP/2数据流的类型构建对应的HTTP/2流对象结构；

基于HTTP/2头映射表解析每个所述HTTP/2数据流的头信息，并提取所述HTTP/2数据流中与所述头信息对应的数据内容；

将每个HTTP/2数据流解析后的头信息与数据内容存储至对应的HTTP/2流对象结构中。

第二方面，本申请实施例还提供了一种深度报文处理装置，该装置包括：

协议确定模块，设置为获取TCP数据流，基于所述TCP数据流中的协议标识确定所述TCP数据流中是否包含HTTP/2协议；

结构构建模块，设置为在所述TCP数据流中包含HTTP/2协议的情况下，根据所述HTTP/2协议中的流标识符确定所述TCP数据流中至少一个HTTP/2数据流，并基于所述HTTP/2数据流的类型构建对应的HTTP/2流对象结构；

内容提取模块，设置为基于HTTP/2头映射表解析每个所述HTTP/2数据流的头信息，并提取所述HTTP/2数据流中与所述头信息对应的数据内容；

数据存储模块，设置为将每个HTTP/2数据流解析后的头信息与数据内容存储至对应的HTTP/2流对象结构中。

第三方面，本申请实施例还提供了一种电子设备，所述电子设备包括：

至少一个处理器；

存储装置，设置为存储至少一个程序，

当所述至少一个程序被所述至少一个处理器执行，使得所述至少一个处理器实现如本申请实施例提供的深度报文处理方法。

第四方面，本申请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本申请实施例提供的深度报文处理方法。

附图说明

图 1 为本申请实施例一所提供的一种深度报文处理方法的流程示意图；
图 2 为本申请实施例二所提供的一种深度报文处理方法的流程示意图；
图 3 为本申请实施例三所提供的一种深度报文处理装置的结构示意图；
图 4 为本申请实施例四所提供的一种电子设备的结构示意图。

具体实施方式

下面结合附图和实施例对本申请作详细说明。

实施例一

图 1 为本申请实施例一提供的深度报文处理方法的流程示意图，本实施例可适用于需要解析超文本传输协议 2.0 (Hyper Text Transfer Protocol 2.0, HTTP/2) 数据流，并对至少一个的 HTTP/2 数据流的数据信息进行存储，以统计分析大量 HTTP/2 数据流数据信息的情形，该方法可以由深度报文处理装置来执行，该装置可以由硬件和/或软件来实现，该方法包括如下步骤：

S110、获取 TCP 数据流，基于 TCP 数据流中的协议标识确定 TCP 数据流中是否包含 HTTP/2 协议。

其中，TCP (Transmission Control Protocol, 传输控制协议) 提供一种面向连接的字节流服务，两个使用 TCP 的应用 (如客户端和服务器) 在交换数据包之前，需要建立一个 TCP 连接，在该连接上传输的数据则为 TCP 数据流，每个 TCP 数据流中包含至少一个 TCP 报文，每个 TCP 报文包括报文头部分和有效承载的数据体部分。TCP 数据流可以从传输层中直接获取，或从链路层、网络层等底层数据包中解封装获得，其中，解封装指自下而上的逐层去掉头部或尾部的过程。

在本实施例中，HTTP/2 协议指 HTTP/2 协议报文帧。协议标识指 TCP 数

据流中每个 TCP 报文的数据体部分中所包含的协议信息，其中，协议是指诸如 HTTP 协议、文件传输协议（File Transfer Protocol, FTP）协议或 HTTP/2 等应用层的协议，协议信息指在应用层上发送的至少一个报文帧的组合，其中，应用层协议在交互过程中产生的至少一个报文帧，由传输层根据帧头的标志进行关联和组装，并封装至 TCP 报文的数据体部分中，因此，可以从 TCP 报文的数据体部分中确定应用层协议信息。不同的应用类型通常依赖不同的应用层协议，而每个应用层协议具备相应的特征，如特定的端口、特定的字符串或者特定的 Bit 序列等，基于每个协议的相应特征，通过对 TCP 数据流中的 TCP 报文数据体部分的协议信息进行检测，可以确定出 TCP 数据流所包含的协议。

示例性的，基于 TCP 数据流中的协议标识确定 TCP 数据流中是否包含 HTTP/2 协议，包括：从 TCP 数据流中提取第一个有效承载的报文内容，对第一个有效承载的报文内容进行 Magic 字符串匹配，若匹配成功，则 TCP 数据流中包含 HTTP/2 协议。其中，第一个有效承载的报文内容指 TCP 数据流的第一个 TCP 报文的数据体部分。两个依赖 HTTP/2 协议的应用（如客户端和服务器）在传输至少一个 HTTP/2 报文帧时，第一个报文帧通常为 Magic 帧，因此，将第一个报文帧通常为 Magic 帧作为 HTTP/2 协议的特征，对第一个报文帧进行 Magic 字符串匹配可以判断是否为 HTTP/2 协议。Magic 帧的内容固定为 PRI*HTTP/2.0\r\n\r\nSM\r\n\r\n，因此，将 TCP 数据流的第一个有效承载的报文内容进行 Magic 字符串匹配，可以识别出 HTTP/2 协议。

可选的，若 TCP 数据流不包含 HTTP/2 协议，将 TCP 数据流添加至黑名单，以使在 TCP 数据流的生命周期内不进行 HTTP/2 数据流处理。其中，TCP 数据流的生命周期是指保持 TCP 连接的预设时间段，若 TCP 数据流不包含 HTTP/2 协议，将 TCP 数据流添加至黑名单，以对保持 TCP 连接的预设时间段内的 TCP 数据流均不进行 HTTP/2 数据流处理。

S120、若 TCP 数据流中包含 HTTP/2 协议，则根据 HTTP/2 协议中的流标

识符确定 TCP 数据流中至少一个 HTTP/2 数据流，并基于 HTTP/2 数据流的类型构建对应的 HTTP/2 流对象结构。

其中，HTTP/2 协议中的流标识符用于唯一标识 HTTP/2 的数据流，流标识符为 31 字节，通过流标识符可以确定出至少一个 HTTP/2 数据流。一个单独的 HTTP/2 连接能够包含多个同时打开的 HTTP/2 数据流，每个 HTTP/2 数据流可以看作一个请求，每个 HTTP/2 数据流上传输至少一个 HTTP/2 协议报文帧，不同 HTTP/2 数据流中的 HTTP/2 协议报文帧交错地发送给对方。根据帧头的标志，在传输层将多个不同 HTTP/2 数据流上的 HTTP/2 协议报文帧进行关联和组装，并封装成 TCP 报文。示例性的，根据 HTTP/2 协议中的流标识符确定 TCP 数据流中至少一个 HTTP/2 数据流的过程为：首先，从 TCP 数据流中确定 TCP 报文的数据体，其中，TCP 报文的数据体由多个 HTTP/2 数据流的 HTTP/2 协议报文帧，根据帧头的标志进行关联和组装得到；然后，从 TCP 报文数据体中获取多个 HTTP/2 协议报文帧；最后，根据 HTTP/2 协议报文帧的流标识符确定至少一个 HTTP/2 数据流。示例性的，4 个 HTTP/2 协议报文帧的流标识符分别为 2、4、4、6，则构建 3 个 HTTP/2 流对象结构。

在本实施例中，每个 HTTP/2 流包括一个头帧——headers，用于传输 HTTP/2 流的额外的头部字段，还可以包括至少一个消息体帧——data，用于传输 HTTP/2 消息体。HTTP/2 流对象结构中包括 headers 的存储区域和对应的 data 的存储区域。

S130、基于 HTTP/2 头映射表解析每个 HTTP/2 数据流的头信息，并提取 HTTP/2 数据流中与头信息对应的数据内容。

其中，每个 HTTP/2 数据流的头信息采用 Huffman 编码压缩以减少传输大小，由 HTTP/2 流中的 headers 携带，通过对 HTTP/2 流中的 headers 进行 Huffman 编码解码，再结合头映射表获取 HTTP/2 数据流的头信息。HTTP/2 数据流的传输的两端（如客户端和服务器）需要维护同一份头映射表。

示例性的，HTTP/2 头映射表包括静态头映射表和动态头映射表；其中，基于 HTTP/2 头映射表解析每个 HTTP/2 数据流的头信息，包括：识别 HTTP/2 数据流中未解析头信息，将未解析头信息在静态头映射表中进行匹配，确定未解析头信息对应的解析后的头信息；当静态头映射表中不包括未解析头信息时，调用动态头映射表，确定未解析头信息对应的解析后的头信息。

其中，静态头映射表包含常见的头部名称，以及常见的头部名称与值的组合，头部名称与值组成头部键值对，如表 1 所示，通常预先设置于连接双方（如客户端与服务器）。动态头映射表可以动态地添加内容，如客户端向服务端发送将 cookie: xxxxxxx 添加到动态头映射表的请求信息，以使客户端和服务端可以将整个键值对用一个字符表示。示例性的，HTTP/2 数据流中未解析头信息先通过静态头映射表中进行匹配，若未匹配成功，则再通过动态头映射表进行匹配。

在本实施例中，静态头映射表中包括多个头部键值对和多个头部名称，如表 1 所示。当接收到 HTTP/2 数据流的头信息后，若整个头部键值对都存在于静态头映射表，可以根据索引值直接查询到头部键值，如 HTTP/2 数据流的头信息解码后索引值为 2，则在静态头映射表查询可得对应:method:GET，请求获取 Request-URI 所标识的资源，解码后索引值为 3，则在静态头映射表查询可得对应:method:POST，在 Request-URI 所标识的资源后附加新的数据；若只有头部名称存在于静态头映射表，可以根据索引值查询到头部名称，在将头部值进行解码后，作为该头部名称对应的值，将解码后的头部键值对添加至动态头映射表中，以使后续接收到头信息后，可以根据动态头映射表中的索引值直接确定头部键值对，示例性的，索引值为 32 (100000)，在静态字典中查询可得 cookie，头部值使用了哈夫曼编码，长度是 28 (0011100)；接下来的 28 个字节为 cookie 的值，将其进行哈夫曼解码即可得到 cookie 对应的值，并将该头部名称和对应值添加至动态头映射表中。

表 1

Index	Header Name	Header Value
1	:authority	
2	:method	GET
3	:method	POST
4	:path	/
5	:path	/index.html
6	:scheme	http
7	:scheme	https
.....
32	cookie	
.....

S140、将每个 HTTP/2 数据流解析后的头信息与数据内容存储至对应的 HTTP/2 流对象结构中。

其中，每个 HTTP/2 数据流包含一个头帧，以及优先级帧、ping 帧或至少一个消息体帧等。示例性的，客户端与服务器传输中的某个 HTTP/2 数据流中包括一个头帧和两个消息体帧，则将解析后的头信息和两个消息体帧存储至该 HTTP/2 数据流对象结构中。

本实施例的技术方案，通过获取 TCP 数据流，基于 TCP 数据流中的协议标识确定 TCP 数据流中是否包含 HTTP/2 协议；在确定 TCP 数据流中包含 HTTP/2 协议的情况下，根据 HTTP/2 协议中的流标识符确定 TCP 数据流中至少一个 HTTP/2 数据流，基于 HTTP/2 数据流的类型构建对应的 HTTP/2 流对象结构，从而得到多个 HTTP/2 流对象结构；并基于 HTTP/2 头映射表解析每个 HTTP/2 数据流的头信息，并提取 HTTP/2 数据流中与所述头信息对应的数据内容，将每个 HTTP/2 数据流解析后的头信息与数据内容存储至对应的 HTTP/2 流对象结构中，从而实现了对 HTTP/2 数据流的解析与存储，通过将每个 HTTP/2 数据流存储至对应结构，为预设检测规则的匹配提供了数据基础。

可选的，在将每个 HTTP/2 数据流解析后的头信息与数据内容存储至对应的 HTTP/2 流对象结构中之后，还包括：接收数据检测的匹配规则，将匹配规则分别与每个 HTTP/2 流对象结构中存储信息进行匹配，确定检测结果。

其中，数据检测的匹配规则指根据实际业务需求所确定的规则，可以用于对至少一个 HTTP/2 流对象结构中的每个 HTTP/2 流对象结构存储的信息进行过滤，得到符合业务需求的信息，将过滤结果作为检测结果，从而调用检测结果返回至外部应用，以使外部应用展示或统计检测结果。通过匹配规则分别与至少一个 HTTP/2 流对象结构中的每个 HTTP/2 流对象结构存储信息进行匹配，确定检测结果，实现了 HTTP/2 流对象结构中存储信息的过滤，从而得出业务需求对应的检测结果。

示例性的，匹配规则包括如下至少一项：预设头信息长度、预设数据内容长度和预设字段内容。其中，预设头信息长度和预设数据内容长度可以筛选出固定头长度和数据体长度的 HTTP/2 流对象结构中存储信息。预设字段内容可以是一段预设字符串，预设字符串可以是敏感行为的特征字符串，将匹配出的 HTTP/2 流对象结构中存储信息确定为敏感行为，实现不良网络内容或者是黑客网络攻击行为的检测；预设字符串还可以是各种应用类型对应的特征字符串，用于识别各种应用类型，如 QQ、微信或淘宝等，各种应用类型对应的特征字符串存储于后台特征数据库，以使将至少一个 HTTP/2 流对象结构中的存储信息与特征字符串进行匹配后，得到各个应用类型的检测结果，实现流量流向和业务占比的统计；预设字符串还可以是诸如万维网或无线网等区域的特征字符串或表征时间的字符串，对匹配后的 HTTP/2 流对象结构中存储信息作统计，可以得到 HTTP/2 协议的热点区域和高峰时刻。

在本实施例中，通过预设头信息长度、预设数据内容长度和预设字段内容中的至少一项，分别与至少一个 HTTP/2 流对象结构中存储信息进行匹配，确定检测结果，实现了固定头长度和数据体长度的 HTTP/2 流对象结构中存储信

息的筛选，不良网络内容或者是黑客网络攻击行为的检测，流量流向和业务占比的统计，以及 HTTP/2 协议的热点区域和高峰时刻的探测。

实施例二

图 2 为本申请实施例二提供的深度报文处理方法中的流程示意图，本实施例在上述各实施例的基础上，增加了对 IP 分片报文进行重组，提取重组后的 IP 报文的承载数据，得到 TCP 数据流，并对 TCP 数据流进行排序和去重的步骤。其中与上述各实施例相同或相应的术语的解释在此不再赘述。参见图 2，本实施例提供的深度报文处理方法包括：

S210、接收底层协议信息中的 IP 分片报文，对 IP 分片报文进行重组；提取重组后的 IP 报文的承载数据，得到 TCP 数据流。

其中，底层协议指包含网络层及网络层以下对应的协议，如 MAC (Media Access Control Address, 媒体存取控制位址) 协议、VLAN (Virtual Local Area Network, 虚拟局域网) 协议、MPLS (Multi-Protocol Label Switching, 多协议标签变换) 协议或 IP (Internet Protocol, 网际互联协议) 等。示例性的，若底层协议为 MAC 协议、VLAN 协议或 MPLS 协议，需要对 MAC 协议或 VLAN 协议进行解封装，从而得到 IP 分片报文。IP 分片报文的报头中，包括与分片相关的字段，如：Identification，用于确认不同的分片是否属于同一个 IP 报文；Flags，当 Flags 中的 MF 为 1 时，表示还有分片，此分片为中间分片；Fragment Offset，表示此分片在整个报文中的偏移地址。因此，根据多个 IP 分片报文的报头信息，可以重组出完整 IP 报文，完整 IP 报文的数据体部分包含了 TCP 报文，提取多个完整 IP 报文的数据体部分，可以得到多个 TCP 报文，即 TCP 数据流。

S220、根据 TCP 数据流的序列号对 TCP 数据流进行排序和去重；对于排序和去重后的 TCP 数据流，基于控制信息标志位对 TCP 数据进行释放，其中，用于释放的控制信息标志位包括连接断开和连接重置。

其中，TCP 数据流中的每个 TCP 报文的首部均包含序列号，序列号是 TCP 报文发送的数据组的第一个字节的序号，确保了 TCP 数据流中多个 TCP 报文传输的有序性。示例性的，一个 TCP 报文的序列号为 300，此 TCP 报文的数据部分共有 100 字节，则下一个 TCP 报文的序列号为 400。根据序列号可以对 TCP 报文进行排序和去重，以丢弃冗余 TCP 报文。TCP 数据流中的每个 TCP 报文的首部均包含控制信息标志位，用于表征每个 TCP 报文的性质，如 RST，连接重置标志，用于重置由于主机崩溃或其他原因出现的错误连接，或者用于拒绝非法报文段和拒绝连接请求；FIN，连接断开标志，FIN 为 1 时，表示发送方的数据均发送完毕，关闭本方的数据流；ACK；确认序号标志，为 1 时表示确认号有效，为 0 时表示报文中不含确认信息，忽略确认号字段。若 TCP 报文的控制信息标志位为连接断开或连接重置，则对该 TCP 报文进行释放，不对该 TCP 报文进行 HTTP/2 数据流处理。可选的，建立 TCP 数据流维护表，将排序和去重，以及释放后的 TCP 数据流中的 TCP 报文存储于 TCP 数据流维护表中，以提高对 TCP 数据流进行 HTTP/2 数据流处理的速度。

S230、获取 TCP 数据流，基于 TCP 数据流中的协议标识确定 TCP 数据流中是否包含 HTTP/2 协议。

S240、若 TCP 数据流中包含 HTTP/2 协议，则根据 HTTP/2 协议中的流标识符确定 TCP 数据流中至少一个 HTTP/2 数据流，并基于 HTTP/2 数据流的类型构建对应的 HTTP/2 流对象结构。

S250、基于 HTTP/2 头映射表解析每个 HTTP/2 数据流的头信息，并提取 HTTP/2 数据流中与头信息对应的数据内容。

S260、将每个 HTTP/2 数据流解析后的头信息与数据内容存储至对应的 HTTP/2 流对象结构中。

S270、接收数据检测的匹配规则，将所述匹配规则分别与每个所述 HTTP/2 流对象结构中存储信息进行匹配，确定检测结果。

本实施例的技术方案，在对 TCP 数据流进行 HTTP/2 数据流处理之前，首先接收底层协议信息中的 IP 分片报文，对 IP 分片报文进行重组；然后提取重组后的 IP 报文的承载数据，得到 TCP 数据流，根据 TCP 数据流的序列号对 TCP 数据流进行排序和去重；对于排序和去重后的 TCP 数据流，基于控制信息标志位对 TCP 数据进行释放，实现了对 TCP 数据流的预处理，对预处理后的 TCP 数据流进行 HTTP/2 数据流处理，从而减少了对冗余 TCP 数据流的 HTTP/2 数据流处理，提高了处理效率。

实施例三

图 3 为本申请实施例三提供的一种深度报文处理装置的结构示意图，本实施例可适用于需要解析 HTTP/2 数据流，并对至少一个的 HTTP/2 数据流的数据信息进行存储，以统计分析大量 HTTP/2 数据流数据信息的情形，该装置包括：协议确定模块 310、结构构建模块 320、内容提取模块 330 和数据存储模块 340。

协议确定模块 310，设置为获取 TCP 数据流，基于 TCP 数据流中的协议标识确定 TCP 数据流中是否包含 HTTP/2 协议；

结构构建模块 320，设置为在 TCP 数据流中包含 HTTP/2 协议时，根据 HTTP/2 协议中的流标识符确定 TCP 数据流中至少一个 HTTP/2 数据流，并基于 HTTP/2 数据流的类型构建对应的 HTTP/2 流对象结构；

内容提取模块 330，设置为基于 HTTP/2 头映射表解析每个 HTTP/2 数据流的头信息，并提取 HTTP/2 数据流中与头信息对应的数据内容；

数据存储模块 340，设置为将每个 HTTP/2 数据流解析后的头信息与数据内容存储至对应的 HTTP/2 流对象结构中。

在本实施例中，通过协议确定模块获取 TCP 数据流，基于 TCP 数据流中的协议标识确定 TCP 数据流中是否包含 HTTP/2 协议；基于结构构建模块在

TCP 数据流中包含 HTTP/2 协议时，根据 HTTP/2 协议中的流标识符确定 TCP 数据流中至少一个 HTTP/2 数据流，通过结构构建模块来基于 HTTP/2 数据流的类型构建对应的 HTTP/2 流对象结构，从而得到多个 HTTP/2 流对象结构；并用过内容提取模块来基于 HTTP/2 头映射表解析每个 HTTP/2 数据流的头信息，并提取 HTTP/2 数据流中与所述头信息对应的数据内容，通过数据存储模块将每个 HTTP/2 数据流解析后的头信息与数据内容存储至对应的 HTTP/2 流对象结构中，从而实现了对 HTTP/2 数据流的解析与存储，通过将每个 HTTP/2 数据流存储至对应结构，为预设检测规则的匹配提供了数据基础。

可选的，还包括预处理模块，其中，预处理模块包括报文重组单元和报文排序单元，其中，

报文重组单元，设置为接收底层协议信息中的 IP 分片报文，对 IP 分片报文进行重组；提取重组后的 IP 报文的承载数据，得到 TCP 数据流；

报文排序单元，设置为根据 TCP 数据流的序列号对 TCP 数据流进行排序和去重；对于排序和去重后的 TCP 数据流，基于控制信息标志位对 TCP 数据进行释放，其中，用于释放的控制信息标志位包括连接断开和连接重置。

可选的，HTTP/2 头映射表包括静态头映射表和动态头映射表，内容提取模块 330 还设置为识别 HTTP/2 数据流中未解析头信息，将未解析头信息在静态头映射表中进行匹配，确定未解析头信息对应的解析后的头信息；当静态头映射表中不包括未解析头信息时，调用动态头映射表，确定未解析头信息对应的解析后的头信息。

可选的，上述装置还包括：信息匹配模块，设置为在数据存储模块 340 将每个 HTTP/2 数据流解析后的头信息与数据内容存储至对应的 HTTP/2 流对象结构中之后，接收数据检测的匹配规则，将匹配规则分别与每个 HTTP/2 流对象结构中存储信息进行匹配，确定检测结果。其中，匹配规则可以包括如下至少一项：预设头信息长度、预设数据内容长度和预设字段内容。

可选的，协议确定模块 310 还设置为在 TCP 数据流不包含 HTTP/2 协议时，将 TCP 数据流添加至黑名单，以使在 TCP 数据流的生命周期内不进行 HTTP/2 数据流处理。

本申请实施例所提供的深度报文处理装置可执行本申请任意实施例所提供的深度报文处理方法，具备执行方法相应的功能模块。

值得注意的是，上述系统所包括的各个单元和模块只是按照功能逻辑进行划分的，但并不局限于上述的划分，只要能够实现相应功能即可；另外，各功能单元的具体名称也只是为了便于相互区分，并不用于限制本申请实施例的保护范围。

实施例四

图 4 为本申请实施例四提供的一种电子设备的结构示意图。图 4 示出了适于用来实现本申请实施例实施方式的示例性电子设备 40 的框图。图 4 显示的电子设备 40 仅仅是一个示例，不应本申请实施例的功能和使用范围带来任何限制。

如图 4 所示，电子设备 40 以通用计算设备的形式表现。电子设备 40 的组件可以包括但不限于：至少一个处理器或者处理单元 401，系统存储器 402，连接不同系统组件（包括系统存储器 402 和处理单元 401）的总线 403。

总线 403 表示几类总线结构中的至少一种，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（Industry Standard Architecture，ISA）总线，微通道体系结构（Micro Channel Architecture，MCA）总线，增强型 ISA 总线、视频电子标准协会（Video Electronics Standards Association，VESA）局域总线以及外围组件互连（Peripheral Component Interconnect，PCI）总线。

电子设备 40 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电子设备 40 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 402 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（Random Access Memory, RAM）404 和/或高速缓存存储器 405。电子设备 40 可以包括其它可移动/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，存储系统 406 可以设置为读写不可移动的、非易失性磁介质（图 4 未显示，通常称为“硬盘驱动器”）。尽管图 4 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（例如只读光盘（Compact Disc-Read Only Memory, CD-ROM），数字视盘（Digital Video Disc-Read Only Memory, 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至少一个数据介质接口与总线 403 相连。存储器 402 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申请各实施例的功能。

具有一组（至少一个）程序模块 407 的程序/实用工具 408，可以存储在例如存储器 402 中，这样的程序模块 407 包括但不限于操作系统、至少一个应用程序、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 407 通常执行本申请所描述的实施例中的功能和/或方法。

电子设备 40 也可以与至少一个外部设备 409（例如键盘、指向设备、显示器 410 等）通信，还可与至少一个使得用户能与该电子设备 40 交互的设备通信，和/或与使得该电子设备 40 能与至少一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（Input/Output, I/O）接口 411 进行。并且，电子设备 40 还可以通过网络适配

器 412 与至少一个网络（例如局域网（Local Area Network, LAN），广域网（Wide Area Network, WAN）和/或公共网络，例如因特网）通信。如图所示，网络适配器 412 通过总线 403 与电子设备 40 的其它模块通信。应当明白，尽管图 4 中未示出，可以结合电子设备 40 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、冗余处理单元、外部磁盘驱动阵列、磁盘阵列（Redundant Arrays of Independent Disks, RAID）系统、磁带驱动器以及数据备份存储系统等。

处理单元 401 通过运行存储在系统存储器 402 中的程序，从而执行各种功能应用以及数据处理，例如实现本发实施例所提供的一种深度报文处理方法，该方法包括：

获取 TCP 数据流，基于 TCP 数据流中的协议标识确定 TCP 数据流中是否包含 HTTP/2 协议；

在确定 TCP 数据流中包含 HTTP/2 协议的情况下，根据 HTTP/2 协议中的流标识符确定 TCP 数据流中至少一个 HTTP/2 数据流，并基于 HTTP/2 数据流的类型构建对应的 HTTP/2 流对象结构；

基于 HTTP/2 头映射表解析每个 HTTP/2 数据流的头信息，并提取 HTTP/2 数据流中与头信息对应的数据内容；

将每个 HTTP/2 数据流解析后的头信息与数据内容存储至对应的 HTTP/2 流对象结构中。

当然，本领域技术人员可以理解，处理器还可以实现本申请任意实施例所提供的深度报文处理方法的技术方案。

实施例五

本申请实施例五还提供一种包含计算机可执行指令的存储介质，所述计算机可执行指令在由计算机处理器执行时用于执行一种深度报文处理方法。

本实施例提供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申请任意实施例所提供的深度报文处理方法步骤，该方法包括：

获取 TCP 数据流，基于 TCP 数据流中的协议标识确定 TCP 数据流中是否包含 HTTP/2 协议；

在确定 TCP 数据流中包含 HTTP/2 协议的情况下，根据 HTTP/2 协议中的流标识符确定 TCP 数据流中至少一个 HTTP/2 数据流，并基于 HTTP/2 数据流的类型构建对应的 HTTP/2 流对象结构；

基于 HTTP/2 头映射表解析每个 HTTP/2 数据流的头信息，并提取 HTTP/2 数据流中与头信息对应的数据内容；

将每个 HTTP/2 数据流解析后的头信息与数据内容存储至对应的 HTTP/2 流对象结构中。

本申请实施例的计算机存储介质，可以采用至少一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：具有至少一个导线的电连接、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（Erasable Programmable Read-Only Memory, 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、射频（Radio Frequency, RF）等等，或者上述的任意合适的组合。

可以以至少一种程序设计语言或其组合来编写用于执行本申请实施例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）。

权利要求书

1、一种深度报文处理方法，包括：

获取传输控制协议 TCP 数据流，基于所述 TCP 数据流中的协议标识确定所述 TCP 数据流中是否包含超文本传输协议 2.0 HTTP/2 协议；

响应于所述 TCP 数据流中包含 HTTP/2 协议，根据所述 HTTP/2 协议中的流标识符确定所述 TCP 数据流中至少一个 HTTP/2 数据流，并基于所述 HTTP/2 数据流的类型构建对应的 HTTP/2 流对象结构；

基于 HTTP/2 头映射表解析每个所述 HTTP/2 数据流的头信息，并提取所述 HTTP/2 数据流中与所述头信息对应的数据内容；

将每个 HTTP/2 数据流解析后的头信息与数据内容存储至对应的 HTTP/2 流对象结构中。

2、根据权利要求 1 所述的方法，其中，所述获取 TCP 数据流，包括：

接收底层协议信息中的网际互联协议 IP 分片报文，对所述 IP 分片报文进行重组；

提取重组后的 IP 报文的承载数据，得到所述 TCP 数据流。

3、根据权利要求 1 所述的方法，在获取 TCP 数据流之前，还包括：

根据 TCP 数据流的序列号对 TCP 数据流进行排序和去重；

对于排序和去重后的 TCP 数据流，基于控制信息标志位对 TCP 数据进行释放，其中，用于释放的控制信息标志位包括连接断开和连接重置。

4、根据权利要求 1 所述的方法，其中，所述 HTTP/2 头映射表包括静态头映射表和动态头映射表；

其中，所述基于 HTTP/2 头映射表解析每个所述 HTTP/2 数据流的头信息，包括：

识别所述 HTTP/2 数据流中未解析头信息，将所述未解析头信息在所述静态头映射表中进行匹配，确定所述未解析头信息对应的解析后的头信息；

在所述静态头映射表中不包括所述未解析头信息的情况下，调用所述动态头映射表，确定所述未解析头信息对应的解析后的头信息。

5、根据权利要求 1 所述的方法，在将每个 HTTP/2 数据流解析后的头信息与数据内容存储至对应的 HTTP/2 流对象结构中之后，还包括：

接收数据检测的匹配规则，将所述匹配规则分别与每个所述 HTTP/2 流对象结构中存储信息进行匹配，确定检测结果。

6、根据权利要求 5 中所述的方法，其中，所述匹配规则包括如下至少一项：预设头信息长度、预设数据内容长度和预设字段内容。

7、根据权利要求 1 所述的方法，还包括：响应于所述 TCP 数据流不包含 HTTP/2 协议，将所述 TCP 数据流添加至黑名单，以使在所述 TCP 数据流的生命周期内不进行 HTTP/2 数据流处理。

8、一种深度报文处理装置，包括：

协议确定模块，设置为获取传输控制协议 TCP 数据流，基于所述 TCP 数据流中的协议标识确定所述 TCP 数据流中是否包含超文本传输协议 2.0 HTTP/2 协议；

结构构建模块，设置为在所述 TCP 数据流中包含 HTTP/2 协议的情况下，根据所述 HTTP/2 协议中的流标识符确定所述 TCP 数据流中至少一个 HTTP/2 数据流，并基于所述 HTTP/2 数据流的类型构建对应的 HTTP/2 流对象结构；

内容提取模块，设置为基于 HTTP/2 头映射表解析每个所述 HTTP/2 数据流的头信息，并提取所述 HTTP/2 数据流中与所述头信息对应的数据内容；

数据存储模块，设置为将每个 HTTP/2 数据流解析后的头信息与数据内容存储至对应的 HTTP/2 流对象结构中。

9、一种电子设备，包括：

至少一个处理器；

存储装置，设置为存储至少一个程序，

当所述至少一个程序被所述至少一个处理器执行，使得所述至少一个处理器实现如权利要求 1-7 中所述的深度报文处理方法。

10、一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机

程序，所述计算机程序被处理器执行时实现如权利要求 1-7 中所述的深度报文处理方法。

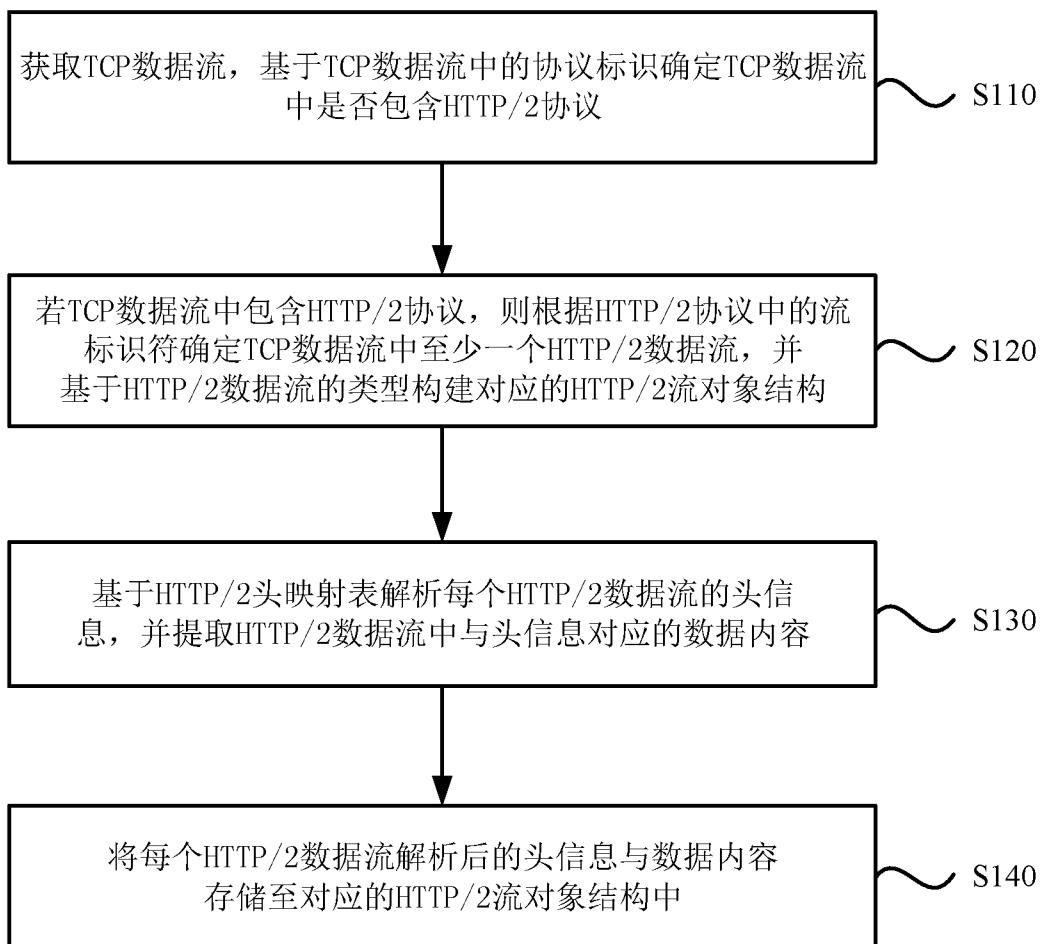


图 1

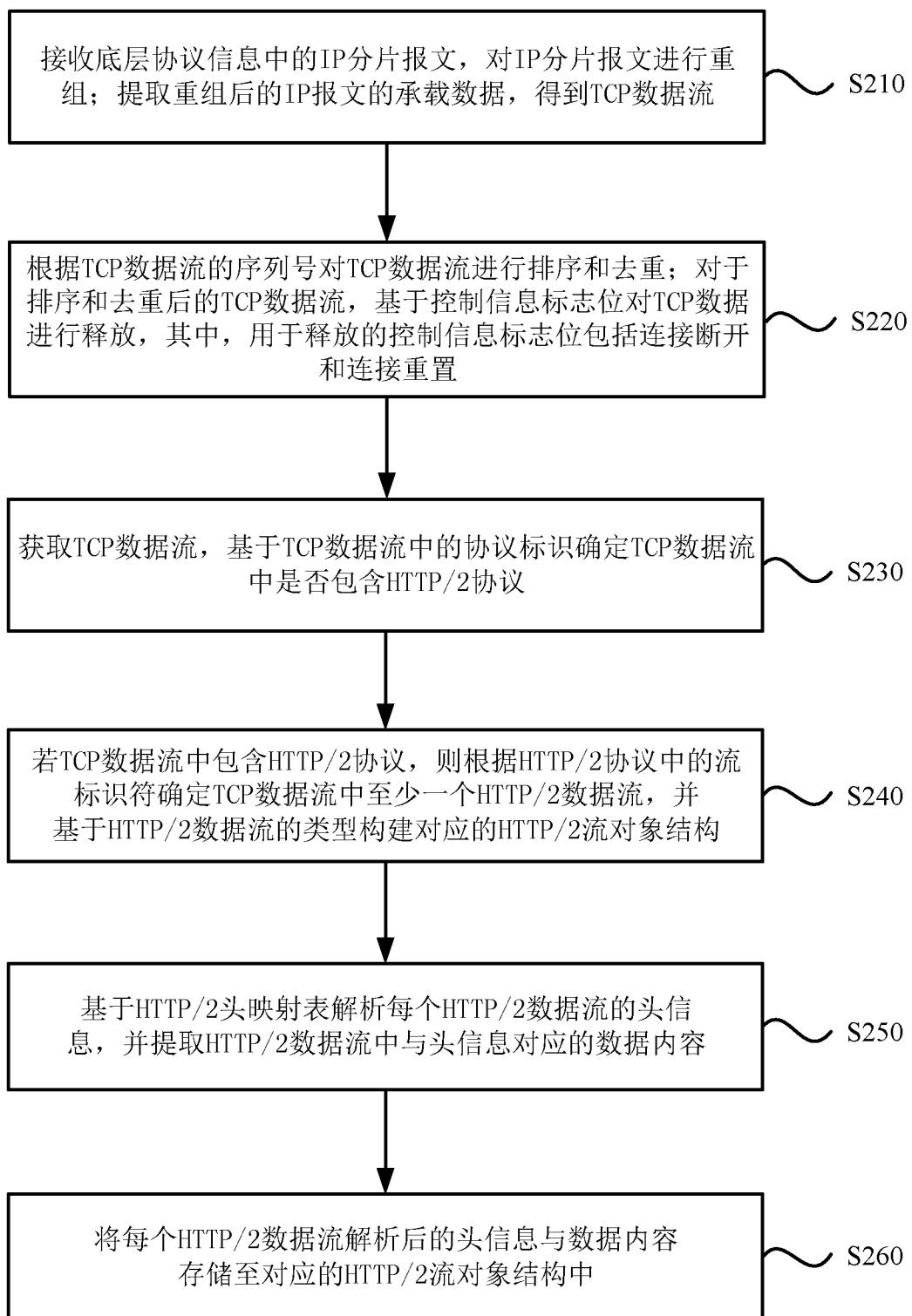


图 2

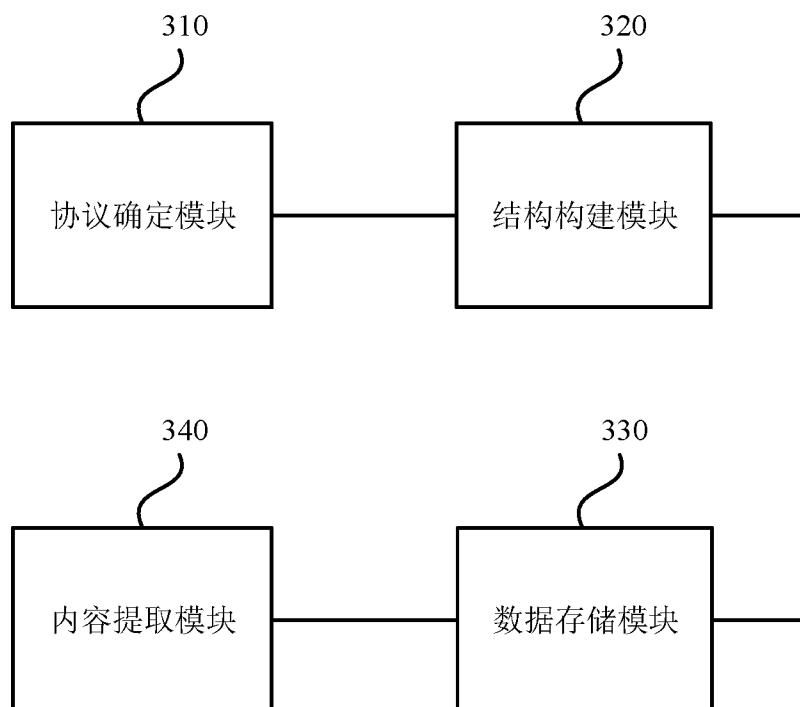


图 3

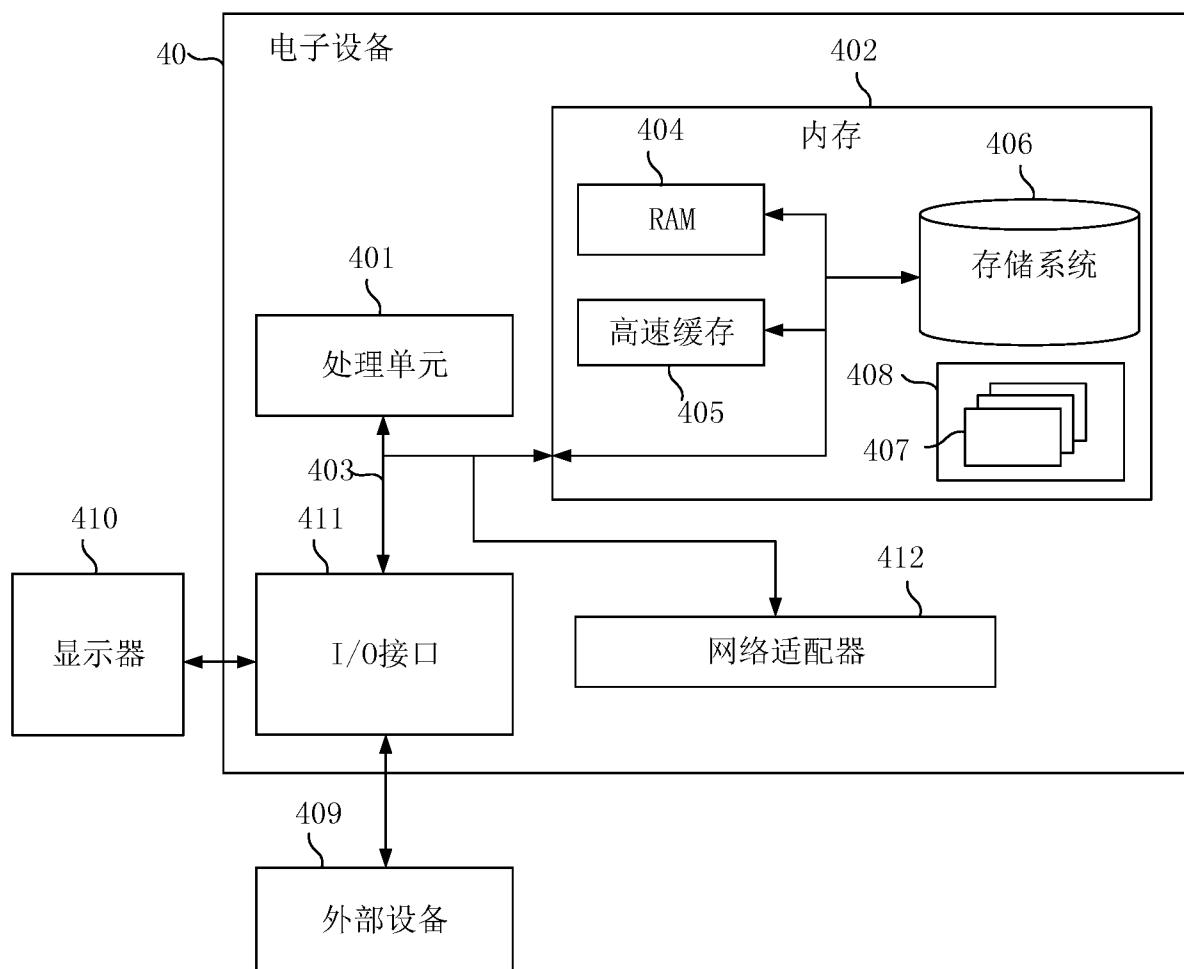


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/107642

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/06(2006.01)i; H04L 29/08(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)

H04L

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; USTXT; WOTXT; EPTXT; 深度, 报文, 分组, 检测, 处理, 超文本传输协议, 2.0, 版本2, HTTP, HTTP/2, 传输控制协议, TCP, 流, 类型, ID, 结构, 头, 数据, 存储, DPI, segment, package, stream, identifier, type, style, structure, header, data, store

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 112311789 A (RUN TECHNOLOGIES CO., LTD. BEIJING) 02 February 2021 (2021-02-02) claims 1-10, description paragraphs [0029]-[0100], figures 1-4	1-10
X	US 10291682 B1 (JUNIPER NETWORKS INC.) 14 May 2019 (2019-05-14) description column 2 lines 49-57, column 8 line 22 to column 10 line 49, figures 1A, 1B, 2, 5	1-10
A	CN 110636151 A (NEW H3C INFORMATION SECURITY TECHNOLOGY CO., LTD.) 31 December 2019 (2019-12-31) entire document	1-10
A	CN 101399843 A (ZTE CORPORATION) 01 April 2009 (2009-04-01) entire document	1-10
A	US 2020162537 A1 (IBM) 21 May 2020 (2020-05-21) entire document	1-10

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

22 September 2021

Date of mailing of the international search report

11 October 2021

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088 China

Authorized officer

Facsimile No. **(86-10)62019451**

Telephone No.

INTERNATIONAL SEARCH REPORT**Information on patent family members**

International application No.

PCT/CN2021/107642

Patent document cited in search report		Publication date (day/month/year)		Patent family member(s)		Publication date (day/month/year)	
CN	112311789	A	02 February 2021	None			
US	10291682	B1	14 May 2019	None			
CN	110636151	A	31 December 2019	None			
CN	101399843	A	01 April 2009	CN	101399843	B	28 November 2012
US	2020162537	A1	21 May 2020	US	10757166	B2	25 August 2020

国际检索报告

国际申请号

PCT/CN2021/107642

A. 主题的分类

H04L 29/06 (2006.01) i; H04L 29/08 (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)

H04L

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; VEN; USTXT; WOTXT; EPTXT; 深度, 报文, 分组, 检测, 处理, 超文本传输协议, 2.0, 版本2, HTTP, HTTP/2, 传输控制协议, TCP, 流, 类型, ID, 结构, 头, 数据, 存储, DPI, segment, package, stream, identifier, type, style, structure, header, data, store

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	CN 112311789 A (北京锐安科技有限公司) 2021年 2月 2日 (2021 - 02 - 02) 权利要求1-10, 说明书第[0029]-[0100]段, 图1-4	1-10
X	US 10291682 B1 (JUNIPER NETWORKS INC) 2019年 5月 14日 (2019 - 05 - 14) 说明书第2栏第49-57行, 第8栏第22行至第10栏第49行, 图1A、1B、2、5	1-10
A	CN 110636151 A (新华三信息安全技术有限公司) 2019年 12月 31日 (2019 - 12 - 31) 全文	1-10
A	CN 101399843 A (中兴通讯股份有限公司) 2009年 4月 1日 (2009 - 04 - 01) 全文	1-10
A	US 2020162537 A1 (IBM) 2020年 5月 21日 (2020 - 05 - 21) 全文	1-10

 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

2021年 9月 22日

国际检索报告邮寄日期

2021年 10月 11日

ISA/CN的名称和邮寄地址

中国国家知识产权局(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

黄淑美

电话号码 86-(512)-88996059

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/107642

检索报告引用的专利文件		公布日 (年/月/日)		同族专利		公布日 (年/月/日)	
CN	112311789	A	2021年 2月 2日	无			
US	10291682	B1	2019年 5月 14日	无			
CN	110636151	A	2019年 12月 31日	无			
CN	101399843	A	2009年 4月 1日	CN	101399843	B	2012年 11月 28日
US	2020162537	A1	2020年 5月 21日	US	10757166	B2	2020年 8月 25日