

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

H04L 29/06 (2006.01)

H04L 12/46 (2006.01)



# [12] 发明专利说明书

专利号 ZL 02829573.0

[45] 授权公告日 2009年12月23日

[11] 授权公告号 CN 100574312C

[22] 申请日 2002.9.6 [21] 申请号 02829573.0

[86] 国际申请 PCT/SG2002/000210 2002.9.6

[87] 国际公布 WO2004/023762 英 2004.3.18

[85] 进入国家阶段日期 2005.3.7

[73] 专利权人 因芬奈昂技术股份有限公司

地址 德国慕尼黑

[72] 发明人 什里达尔·穆巴拉克·米什拉

胡春风

[56] 参考文献

US5748905A 1998.5.5

WO00/52897A2 2000.9.8

CN1112322A 1995.11.22

WO01/33774A1 2001.5.10

US5917821A 1999.6.29

审查员 肖东

[74] 专利代理机构 中科专利商标代理有限责任公司

代理人 朱进桂

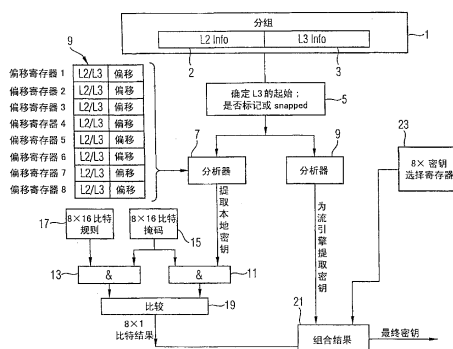
权利要求书3页 说明书7页 附图4页

[54] 发明名称

分析数据分组的分析器

[57] 摘要

一种分析器系统，其被设置成接收具有来自多个不同分组的交叉部分的数据流(1)，并在它到达时从每个部分中提取数据。此分析器系统具有扫描单元，其接收与定义了哪个分组与其有关的数据的每个部分有关的信息，并使用此信息和数据流的属性，以识别(2)层、(3)层和(4)层的位置。此信息被传送到基于此数据和偏移提取数据的分析器单元(7)、(9)。分析器单元(7)用的偏移被存储在用户可编程寄存器(9)。



1、一种分析器系统，具有：

接口，被配置成接收由多个不同分组的交叉部分组成的数据流以及接收关于定义了哪个分组与其相关的每一个数据的部分识别信息，

分析单元，被配置成以一个部分一个部分的方式来处理数据流并用部分识别信息识别和提取数据，所述分析单元包括：

扫描部分，被配置成使用部分识别信息来识别分组的结构特征，所述结构特征包括分组中一个或多个层 2 和层 3 数据的开始位置；

第一分析器，被配置成使用用户可编程偏移信息来提取数据，所述用户可编程偏移信息包括表示是否从层 2 或层 3 数据提取数据的指示以及表示相对于层 2 或层 3 数据的开始位置要提取的数据的位置的指示；以及

第二分析器，被配置成在由所识别的分组的结构特征和预定偏移信息定义的位置处，从分组中提取数据。

2、如权利要求 1 所述的分析器系统，其中扫描部分还被配置成识别 *tagged* 分组。

3、如权利要求 1 所述的分析器系统，其中所述扫描部分还被配置成识别 *snapped* 分组。

4、如权利要求 1 所述的分析器系统，其中所述接口被配置成接收包括 8 个分组的数据流，每一个分组被划分为长度是 8 字节的部分，交叉所述 8 个分组的部分。

5、如权利要求 4 所述的分析器系统，其中所述扫描部分还被配置成针对接收到的 8 个字节中的每一个计算变量集合，所述变量集合包括：指示了迄今为止接收到的字节的数目的长度变量；指示了不大于所述长度变量除以 8 的最大整数的索引变量；以及指示了长度变量模 8 的偏移变量。

6、如权利要求 5 所述的分析器系统，其中所述扫描部分被配置成根据计算的变量集合，识别分组的层 3 的开始以及分组包括 VLAN 标

记还是 SNAP 标记。

7、如权利要求 5 所述的分析器系统，其中所述第一分析器被配置成使用在用户可编程寄存器中设置的用户可编程偏移信息来提取 8 个 16 比特的本地密钥，所述分析单元还被配置成：

对于所提取的本地密钥和 8 个 16 比特掩码执行 AND 运算；

对于所述掩码和 8 个 16 比特规则执行 AND 运算；以及

将 AND 运算的结果进行比较以产生 8 个 1 比特结果。

8、如权利要求 7 所述的分析器系统，其中所述第二分析器被配置成使用在密钥选择寄存器中设置的预定偏移信息来提取 8 比特集合，所述分析单元还包括组合单元，所述组合单元被配置成将所述 8 比特集合与所述 8 个 1 比特结果相组合，以产生最终密钥。

9、一种分析数据流的方法，包括步骤：

接收由多个不同分组的交叉部分组成的数据流；

接收关于定义了哪个分组与其相关的每一个数据的部分识别信息，

以一个部分一个部分的方式来处理数据流，所述处理包括用部分识别信息识别分组的结构特征，所述结构特征包括分组中一个或多个层 2 和层 3 数据的开始的位置；

使用第一分析器，根据用户可编程偏移信息来从分组中提取数据，所述用户可编程偏移信息包括表示是否从层 2 或层 3 数据提取数据的指示以及表示相对于层 2 或层 3 数据的开始位置要提取的数据的位置的指示；以及

使用第二分析器，在由所识别的分组的结构特征和预定偏移信息定义的位置处，从分组中提取数据。

10、如权利要求 9 所述的方法，其中所述识别结构特征的步骤包括识别 *tagged* 分组。

11、如权利要求 9 所述的方法，其中所述识别结构特征的步骤包括识别 *snapped* 分组。

12、如权利要求 9 所述的方法，其中所述接收数据流的步骤包括接收包括 8 个分组的数据流，每一个分组被划分为长度是 8 字节的部

分，交叉所述 8 个分组的部分。

13、如权利要求 12 所述的方法，其中所述识别结构特征的步骤包括针对接收到的 8 个字节中的每一个计算变量集合，所述变量集合包括：指示了迄今为止接收到的字节的数目的长度变量；指示了不大于所述长度变量除以 8 的最大整数的索引变量；以及指示了长度变量模 8 的偏移变量。

14、如权利要求 13 所述的方法，其中所述识别结构特征的步骤包括根据计算的变量集合，识别分组的层 3 的开始以及分组包括 VLAN 标记还是 SNAP 标记。

15、如权利要求 12 所述的方法，其中使用在用户可编程寄存器中设置的用户可编程偏移信息来提取 8 个 16 比特的本地密钥，所述方法还包括：

对于所提取的本地密钥和 8 个 16 比特掩码执行 AND 运算；

对于所述掩码和 8 个 16 比特规则执行 AND 运算；以及

将 AND 运算的结果进行比较以产生 8 个 1 比特结果。

16、如权利要求 15 所述的方法，其中使用所述第二分析器进行提取的步骤包括根据在密钥选择寄存器中设置的预定偏移信息来提取 8 比特集合，所述方法还包括将所述 8 比特集合与所述 8 个 1 比特结果相组合，以产生最终密钥。

## 分析数据分组的分析器

### 相关申请

本规则是具有相同优先权日的一组五个专利申请。申请 PCT/SG02/……涉及具有可配置为 8 个 FE（快速以太网）端口或 1 个 GE（千兆以太网）端口的输入端口的交换机。本申请涉及适合于这样的交换机使用的分析器。申请 PCT/SG02/……涉及适用于将分析器的输出与规则比较的流引擎。申请 PCT/SG02/……涉及用规则和分组的比较结果监控带宽损耗。申请 PCT/SG02/……涉及排列成堆栈的交换机的组合。该组申请中每一个申请各自的主题提出与在其它四个申请中描述的技术的整体内容不同的申请，然而该组中其它申请公开的内容通过参考而被合并。

### 技术领域

本发明涉及一种分析多数据分组以从中提取信息的分析器系统。在一个典型例子中，分析器系统可以被用于象以太网交换机的交换机，以处理收到的数据分组并由此获得通过队列管理和交换结构（switching fabric）处理分组所需要的信息。

### 背景技术

最近互联网技术的提高已经改变了我们交换信息的方式。互联网的普遍使用已经导致了趋同观念。不同的数据类型（例如视频、声音、图像和文字）必须在相同网络中交叉，这引起了为了在具有服务质量支持的单个网络上实时传输和数据通信的协议过剩。

这些协议主要是 DiffServ、IntServ 和 MPLS，每个都需要在收到分组时将其实时地分类（例如由分组的类型决定）。如此分类的第一步是从一个分组中提取相应的字节，“分析”。这在 Andrew S

Tanenbaum, Prentice Hall 2<sup>nd</sup> Ed, 1998 “计算机网络”的第一章得到描述。

分析操作包括确定分组是否包括标记。例如，通常以太网分组会包括一个 VLAN（虚拟局域网）标记——比如指示与分组相联系的 VLAN 的标记。VLAN 标记通常是在以太网帧源 MAC 地址字段和长度/类型字段之间插入的 4 个字节。VLAN 标记的头两字节总是被设置为 0x8100 值，第二个两字节是控制信息（用户优先级字段、标准格式指示符和 VLAN 标识符）。

另一种类型的标记被 SNAP 协议定义（子网接入协议），其被引入以允许更老的帧和协议被压缩成类型 1 LLC 头（Type 1 LLC header），以使任何协议“伪 IEEE 适应”。SNAP 标记（或“snapf 包封”）被直接置于以太网分组的标准长度/类型字段之后（其通常取值小于或等于 1500），并且具有 AA-AA 作为其头两个字节。包含 SNAP 标记的分组被称为 SNAPped。

总之，以太网分组由被称为是层的嵌套数据的级组成。被机器直接解译的数据被称为“层 1”数据或物理层数据、“层 2”或数据链路层，数据是 LAN（局域网）数据，比如唯一确定 LAN 上的一个适配符的 MAC（媒体访问控制）数据。在“层 2”的分组中可能有“层 3”或网络层，在除分组的 IP 源地址和目的地址之外的事物中定义的数据。在“层 3”的分组中可能有“层 4”数据、或传输层数据，例如 TCP（传输控制协议）数据。

考虑到可能相互作用的协议的巨大变化，提供一个更灵活的分析技术将是有益的。

另外，存在一种环境的变化，其将对分析大量同时收到的数据有用。把以太网的应用作为一个例子，这样的环境不仅限于以太网应用，而是以以太网应用为例子，在一个上文提及的未决的申请中，为使容易从 FE 转换到 GE 以太网，本发明者提出一种可配置的以太网交换机，其能起快速以太网和起千兆以太网交换机的作用。一个数据端口作为 8 个 FE MAC 接口或作为一个 GE MAC 接口来操作。在以前的情况，可能会发生 8 个 FE 接口同时接收分组的情况。可能需要在输入端口

提供足够大的缓冲存储器，这样所有的这些分组能在处理任何一个分组开始之前被全部接收，但这增加所需缓冲存储器的成本。若在接收分组的同时（“即时”）处理所有的分组，将很有用。

## 发明内容

概括地讲，本发明的第一个方面的分析器系统被设置成接收具有来自多个不同分组的交叉部分的数据流，并在它到达时从每个部分中提取数据。该分析器系统接收与定义其涉及哪个分组的每个数据部分的信息，并使用此信息识别要从数据流中提取的数据。

本发明的第二个方面涉及一种具有为分析器存储数据的多个可编程寄存器的分析器系统，分析器系统接收数据流并根据存储在可编程寄存器中的偏移信息来从数据流中提取数据。

在本发明的任一个方面中，分析器系统最好包括：扫描部分，其识别数据流中的数据的主结构特征的位置（例如一个数据层开始的位置）；以及至少一个分析器单元，其使用扫描部分的输出和偏移信息（在本发明第二方面情况下至少部分来自可编程寄存器）以提取数据。偏移信息识别从结构特征的位置提取的数据的偏移量。

扫描部分使用接收到的、关于数据流的信息，并检查数据自身以确定数据流中分组的特征。例如，除确定分组中层 2 数据、层 3 数据和/或层 4 数据中的一个或更多的开始位置外，能进一步确定分组是否有 VLAN 标记（tagged）和/或 SNAPed。

可能存在两个分析器单元，一个根据预先确定的偏移量和来自扫描部分的结构数据提取数据，另一个根据来自扫描部分的结构数据和可编程寄存器定义的偏移量提取数据。

## 附图说明

现在仅仅为了说明的目的参考下面的附图描述本发明的优选特征：

图 1 以电路图示出作为本发明实施例的分析器系统的操作；

图 2 示出了如何通过第二分析器和图 1 实施例中组合器实现密钥

的提取；

图 3 示出了要被该实施例分析的 4 种分组的结构；和  
图 4 示出了图 1 的分析器系统的电路图。

### 具体实施方式

首先参考图 1，其以图例示出实施例的操作。

实施例处理具有包括层 2 数据 2（在比特 0 处开始）和层 3 数据 3 实际上也包括层 4 数据的分组 1 的数据流，但为了此实施例的目的，层 4 数据可被简单视为层 3 数据的一部分。层 4 数据的开始位置由作为层 3 数据一部分的字段给定。

数据流作为预定长度的一系列部分进入附图 1。数据流最好由一系列交叉并发分组组成。例如，可以达到 8 个分组，它们被分成部分（例如有时是 8 字节），不同数据分组的部分被交叉。使用关于部分是来自哪个分组的信息，在任何时间针对这些部分执行图 1 的步骤。这意味着当数据流到达时不需要缓冲全部的数据流。

操作的第一步（步骤 5）是为了确定层 3 起始偏移以及数据是否有 VLAN 标志或 SNAPped。在开始，对于每个接收到的 8 字节，算法如下计算不同的变量。首先，它通过向计数变量 *length* 值添加新字节的数目，更新指示目前为止接收到的分组的字节数的计数变量 (*length*)。变量 *index*（索引）被定义为不比被 8 除的 *length* 大的最大整数。变量 *offset*（偏移）被定义为模 8 的 *length*。

图 3 针对 4 中类型的分组示出了变量 *index* 和 *offset*，被标记为 (a)，(b)，(c) 和 (d)，具有在被标记为“字节号 (byte number)”的行所标记的字节位置中显示的数据。

- 分组类型 (a) 没有被 VLAN 标记或 snapped，并且层 3 从字节 14 开始。
- 分组类型 (b) 被 VLAN 标记（所以字节 12 和 13 是 0x8100（1 个 16 进制符号）），并且层 3 从字节 18 开始。
- 分组类型 (c) 被以 snap 包封 SNAPped，所述包封从字节 14 开始，并且层 3 数据从字节 22 开始。在字节 14 到 19 的位置上为



0xAA-AA-03-00-00-00.

- 分组类型 (d) 被以 snap 包封 SNAPped, 所述包封从字节 14 开始, 并且被 VLANtagged (所以字节 12 和 13 是 8100), 并且层 3 数据从字节 26 开始。在字节 18 到 23 的位置上为 0xAA-AA-03-00-00-00.

为了确定 L3 的开始位置。当变量 *length* 使得 *index* 是 1 时, 在某一时间下面步骤执行:

- 检查偏移 (offset) 4 和 5 的字节。如果它们不是 0x8100 并且不小于 1500, 那么字节是类型 (a) 并且层 3 从字节 14 开始。

- 否则, 如果偏移 4 和 5 的字节是 0x8100, 那么分组被标记 (分组必须是类型 (b) 或类型 (d))。设定变量 *tagged* 等于 1。

- 否则, 如果偏移 4 和 5 的字节少于或等于 1500, 而且偏移 6 和 7 的字节是 0xAA-AA, 那么分组一定是 snapped, 设定变量 *snapped* 等于 1。

- 否则, 分组是未知协议的。

当数据分组的下一个部分到达时, 变量 *length* 使得 *index* 值为 2:

- 检查偏移 0 和 1 的字节。如果它们比 1500 大并且 *tagged*=1, 那么分组是类型 (b) 并且层 3 从字节 18 开始。

- 否则, 如果偏移 0 和 1 的字节少于或等于 1500 并且 *tagged*=1, 那么设定 *snapped*=1。

- 如果 *tagged*=1 并且 *snapped*=1 并且偏移 2 到 7 的字节是 AA-AA-03-00-00-00, 那么分组是类型 (d), 并且层 3 从字节 22 开始。

- 否则, 如果 *tagged*=0 并且 *snapped*=1 并且偏移 0 到 3 的字节是 0x03-00-00-00, 那么分组是类型 (c), 并且层 3 从字节 22 开始。

- 否则, 协议是未知的。

再参考图 1, 一旦层 3 (和其它层) 的起始位置已知, 数据流的部分被传送给如下面讨论的第一分析器 7 和第二分析器 9。注意, 针对这些部分之一, 在任何时间执行步骤 5 操作和第一分析器 7 和第二分析器 9 操作。在这种情况下, 步骤 5 的操作使用部分表示信息识别数据的部分属于哪个分组, 以及例如在多个分组的情况下针对这些分组中的每一个保持一组变量 (例如变量 *length*)。在处理来自一个给定分

组的数据流的一部分的过程中，步骤 5 步涉及更新相应分组的变量。不过分析器 1 和 2 不知道此信息。

第一分析器 7 根据被一组寄存器 8 定义的位置从分组中提取数据。例如，当 8 字节要被提取时，8 个寄存器（标记为偏移 reg 1、……、偏移 reg 8）被使用。每个寄存器保留一个指示数据是否从层 2 或层 3 数据中被提取的指示（“L2/L3”），以及指示哪个字节相对于哪些层的此起始位置被提取的偏移。用这种方式，第一分析器 7 能提取本地密钥。被提取的本地密钥在 AND 操作 11 中与  $8 \times 16$  比特掩码 15（8 个寄存器每个提取 16 比特）进行比较。同一个  $8 \times 16$  比特掩码 15 与 8 个 16 比特规则 17 在 AND 操作 13 中进行比较。AND 操作 11 和 13 的结果在步骤 19 被比较以生成  $8 \times 1$  比特结果。

同时，第二分析器 9 接收到同样的数据流和确定的层的起点的结果，并提取一组由 8 个密钥选择寄存器 23 确定的一组 8 比特。第二分析器 9 的输出在步骤 21 与比较操作 19 的结果进行比较。

图 2 示出第二分析器 9 和组合单元 21 的操作。图 2 上面部分示出一个数据分组的通常结构，从层 2 数据（“层 2info”）开始，然后层 3 数据（“层 3info”），然后层 4 数据（“层 4info”）。使用层位置确定运算法则的结果，第二分析器 9 被选择性地提供给如图 2 示出的选择出的字节。根据可编程选择器 23 的输出，MUX 多路复用单元 25 输出它们的输入中的一个。这些被提供给进一步的 MUX 多路复用单元 27、29。MUX 多路复用单元 27 接收数据分组的其它部分和第一分析器 7 的输出。MUX 多路复用单元 29 接收 MUX 多路复用单元 27 的各个输出和 MUX 多路复用单元 25 的各个输出。根据来自可编程寄存器 23 的选择信号 sel[0]31、sel[3]31 控制 MUX 多路复用单元 27、29。其结果是用于流引擎的提取的密钥。

再次参考图 1，进行比较操作 19 的输出与用于流引擎的密钥的组合 21，以生成最终密钥。像两个分析器的提取操作一样，此密钥的使用对本领域的技术人员来说很清楚。

图 4 示出在以太网交换机环境中实现图 1 步骤的分析器系统电路 35 的设计。分析器系统电路 35 每次以 8 字节操作，并具有输入接口

41, 其接收来自以太网交换机的引脚接收分组的缓存 rx\_ififo39 的输入, 和来自提供包括确认与相应分组 (此组成上面讨论的部分识别信息) 联系的分组的种类的索引的控制信息的 MAC 接口 rx\_mac\_ififo37 的输入。图 1 的步骤 5 操作被单元 43 执行, 并且其结果尤其被传输到单元 45, 其用作图 1 的第一和第二分析器并实现图 1 中生成最终密钥的组合。单元 45 还从寄存器文件 47 中接收其它需要的数据, 比如图 2 中寄存器 9、17 和 15 的数据。分析器将每个数据流 (例如每个同时的分组) 的所有信息置于相应分组的描述符中。通过每次每个处理步骤在两个周期操作 8 字节, 分析器能管理 8 个 FE 流。

单元 45 的输出传到分析器 parser\_mem\_iface 输出接口 49, 其按顺序把它传输到以太网交换机的其它单元, 特别是内存管理器 rx\_mem\_mgr51。注意, 所有的分析器系统电路 35 的电路最好在单一集成电路上实现。

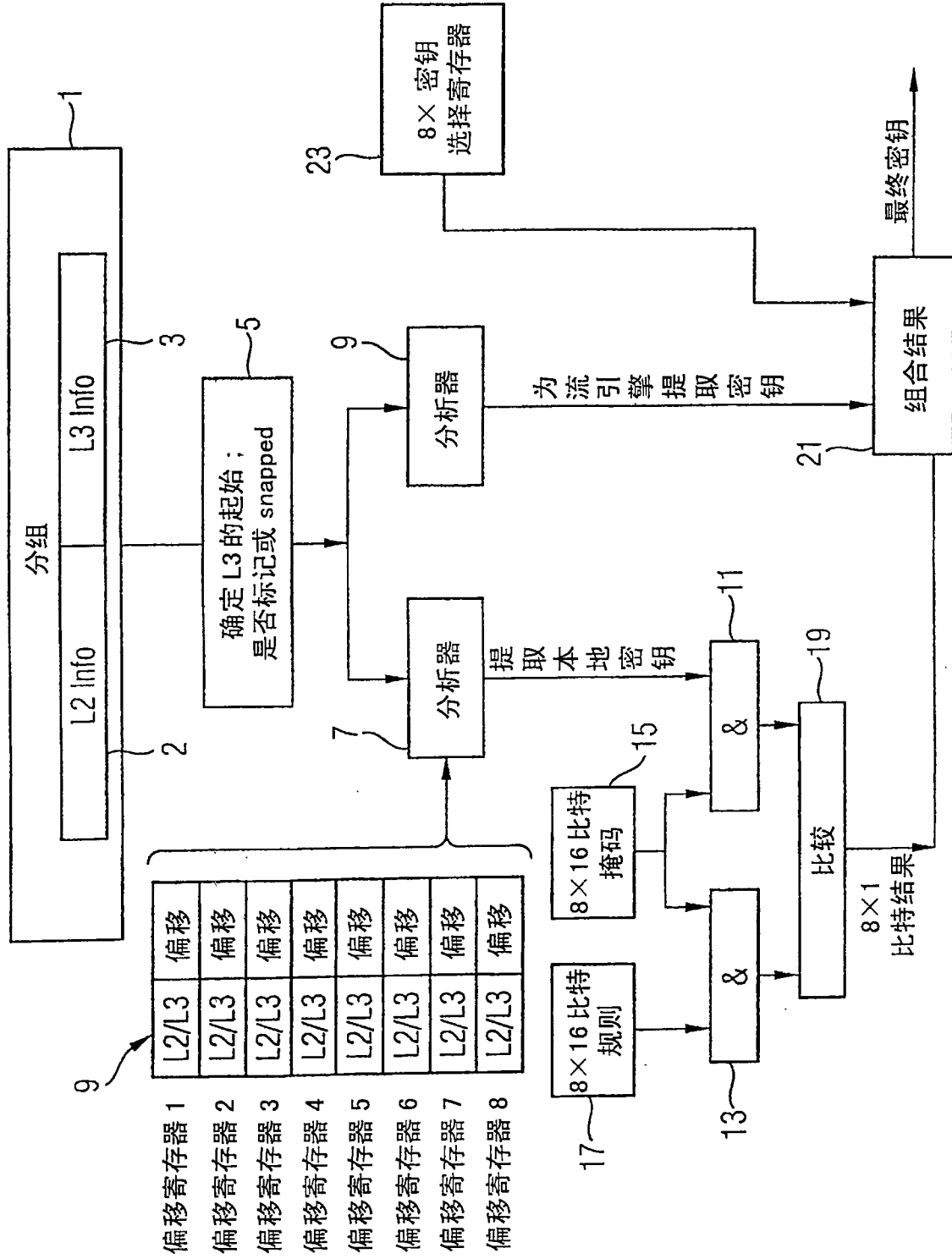


图 1

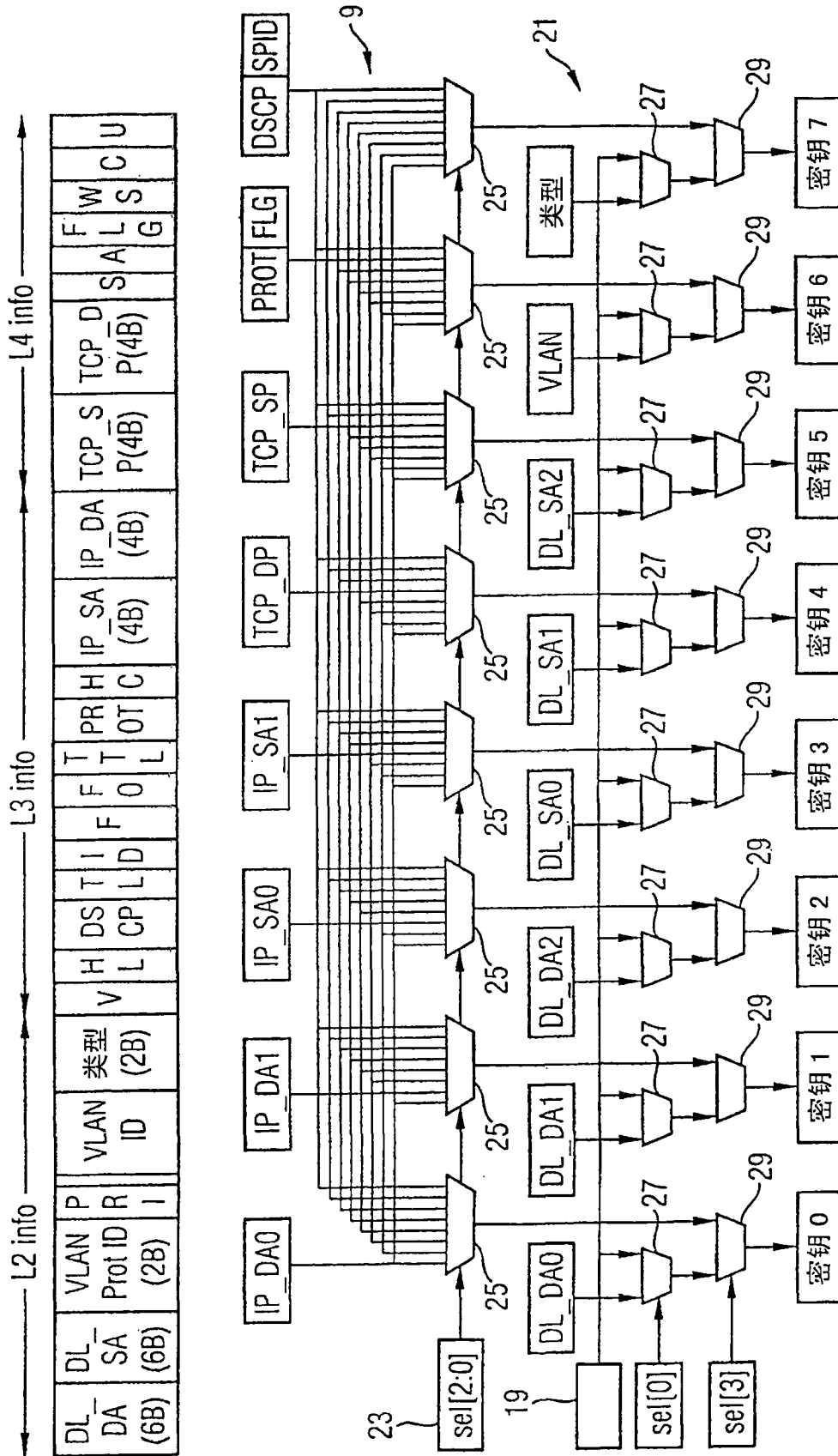


图 2

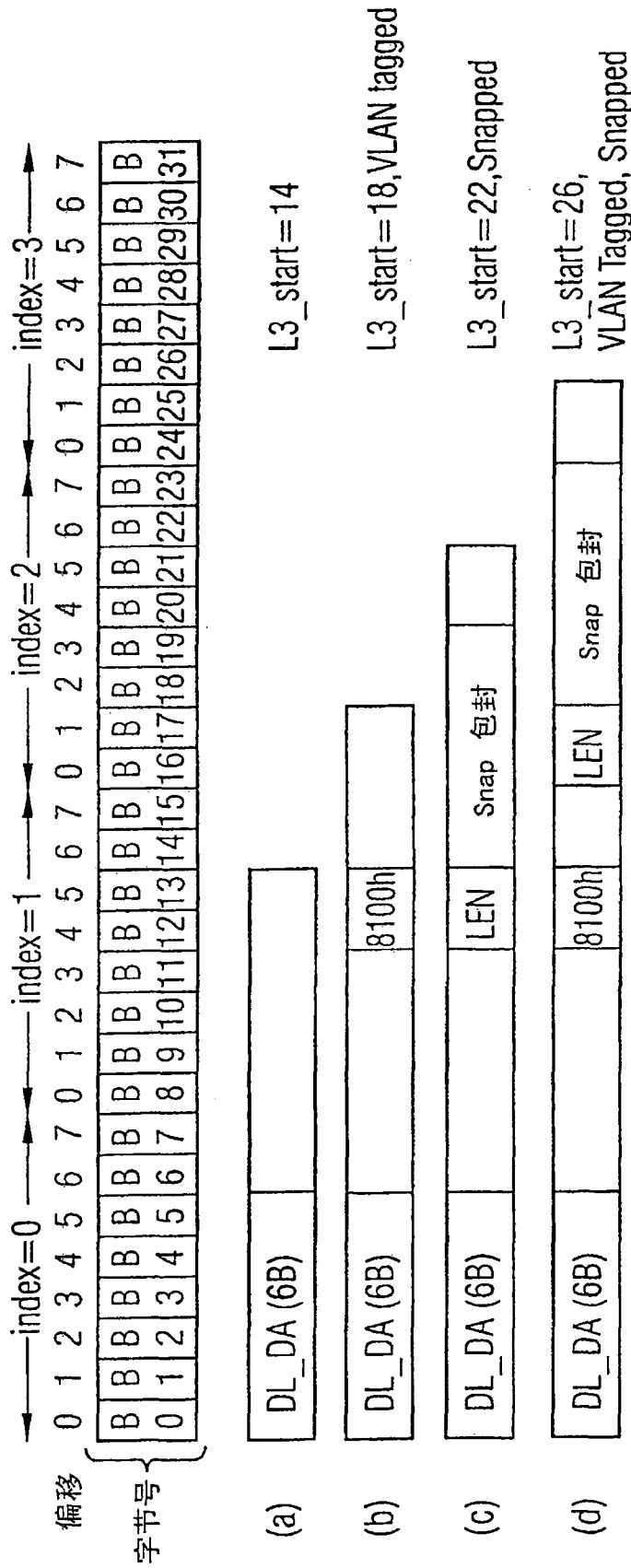
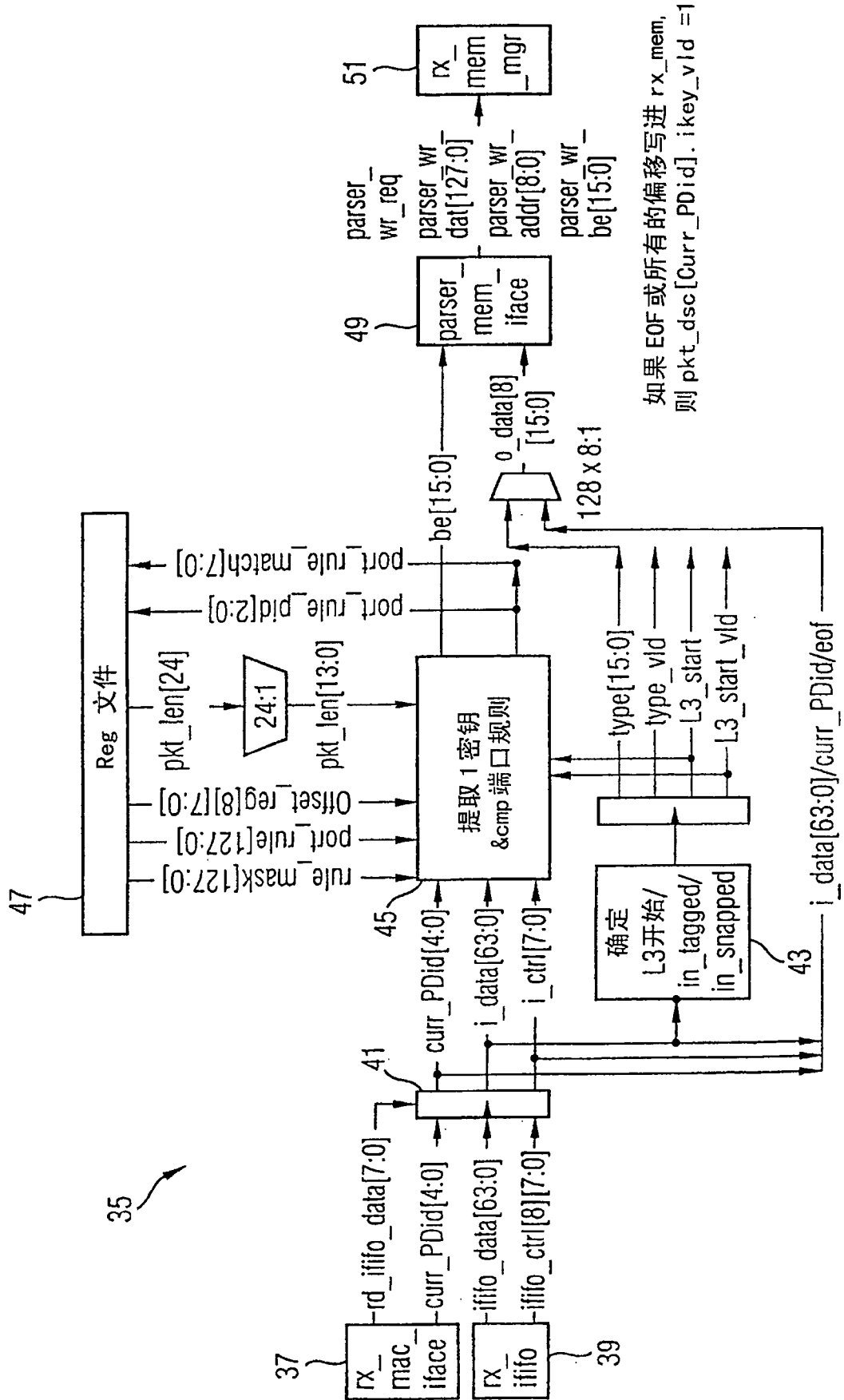


图 3



如果 EOF 或所有的偏移写进 rx\_mem,  
则 pkt\_dsc[curr\_PDId].ikey\_vld = 1

图 4