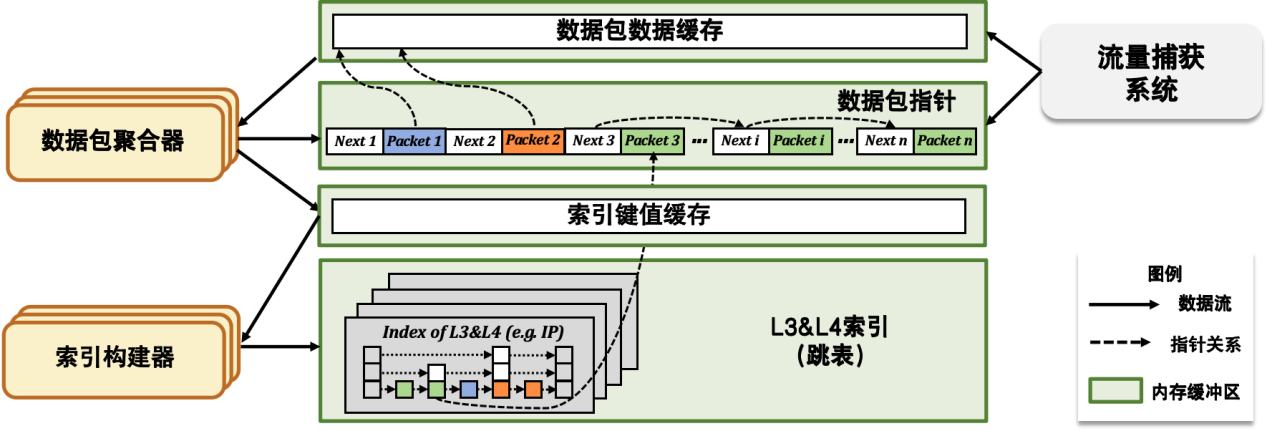
1.流量回溯平台概述

拟实现支持网络全流量捕获、存储和高速检索的流量回溯平台，为安全防御、应用排障和内部监管等流量回溯需求提供支持。该流量回溯平台设计目标在于满足100Gbps高速流量的全镜像捕获与索引建立，并将流量以不低于10:7的压缩比进行压缩存储，同时支持对用户所需任意网络协议字段的TB级流量秒级查询。流量回溯平台预计以四部分组成：**高速流量捕获系统；高并行度索引构建系统；基于相似流聚合的流量数据压缩与存储系统；支持自定义索引建立与检索的用户交互系统。**

* 1. 高速流量捕获系统

该系统设计基于以下见解：网络流量抓取的主要性能开销来自内核协议栈的内存拷贝开销和系统调用开销（约占75%）。由于抓取网络流量时不需要在内核栈对数据包进行协议解析，选择**使用DPDK技术实现内核旁路抓包**。

* 1. 高并行度索引构建系统

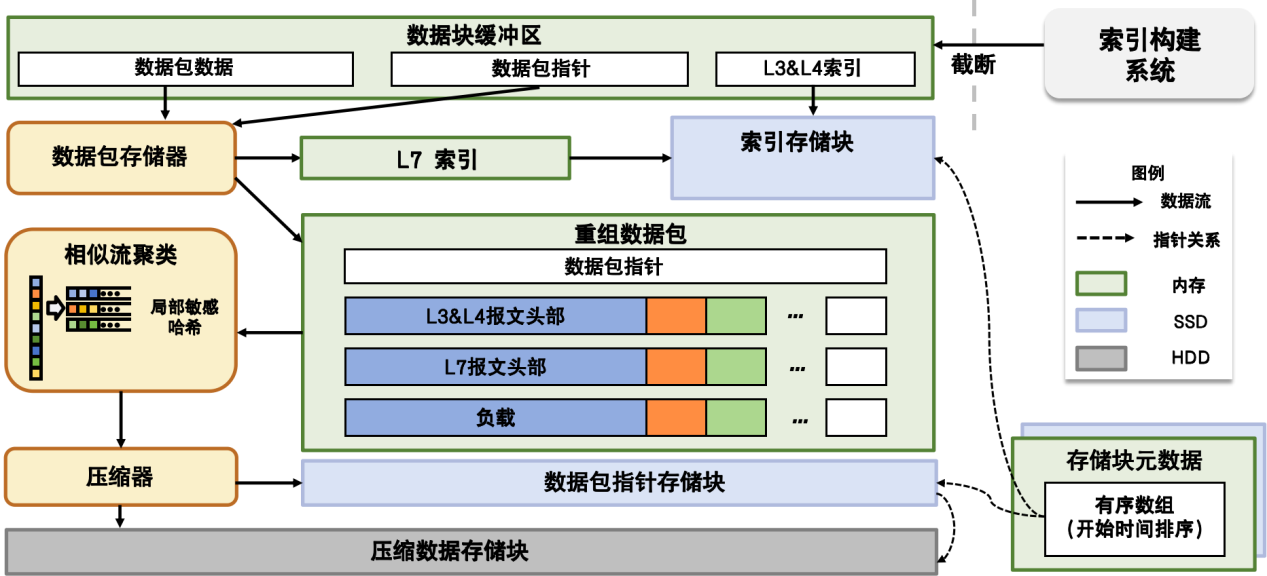


**图1 高并行度索引构建系统**

由于单一线程难以满足100Gbps流量索引（特别是数据包粒度索引）的实时构建，采用可高度并行化的数据结构完成数据包的初步聚合和索引构建。其步骤为：(1)通过特殊设计的**链表结构**，将IP与端口相同的数据包**在负载零拷贝的前提下完成聚合**；（2）基于流量索引写入后不会发生修改的特征，**通过跳表建立索引**，由于跳表在插入操作时只需修改一个节点，相较于基于平衡树的索引可支持高并行度索引构建。

这一设计将数据包与聚合结构连续存储到内存中，尽可能最大限度的利用了内存缓冲区（相较于FLOWSIS对每一个流分配一块固定长度缓冲区），同时对一段时间内的所有抓取流量进行并行索引构建（相较于VAST将数据包分配到不同的节点）。

* 1. 基于相似流聚合的流量数据压缩与存储系统



**图2 基于相似流聚合的流量数据压缩与存储系统**

由于网络流量数据为二进制数据，通用的针对文本数据的压缩方法难以在流量数据中获得较高的压缩比。为此，压缩与存储系统设计基于以下两个见解：(1)网络流量存在大量内容相似的流，如不同用户对同一网页的多次请求等，对其中的相同部分进行整体替换可获得较大的压缩比；（2）更大的数据块可以找到更多的流间相同部分，从而获得更大的压缩比。

因此，我们通过以下步骤对流量数据进行压缩：(1)将1.2中初步聚合的流量进行**整体截断**，形成连续的流量数据块（预计大小在GB级），以流量数据块为单位进行压缩；（2）将流量数据块中**相同流的数据包重新组合**，将路由层、传输层、应用层包头相同字段进行单独压缩，将应用层报文重组为负载数据块；（3）**将协议相同的负载数据块初步聚类，通过局部敏感哈希算法将内容相似的流进一步聚类**；(4)**将流间相同部分使用字典进行整体整体替换**；（5）将压缩后的数据块**整体写入磁盘**中，**将数据块第一个包和最后一个包的时间戳作为数据块索引键**。

* 1. 支持自定义索引建立与检索的用户交互系统

除用于初步聚合流数据的IP地址和端口外，在不同场景需求下，用户还可能对http或DNS协议域名、数据包长度等字段有着快速查询的需求。但是，如果对数据包中每一个协议字段均建立索引，其带来的协议解析时间开销和协议存储开销是难以接受的。为此，我们支持用户对需要的协议进行自定义协议解析与索引建立。该系统参考WireShark的协议解析系统，**用户可自行编写lua脚本定义协议解析过程，并选定任意报文头字段用以建立索引**。由于应用层报文需要在1.3数据包重组后才可获得，自定义应用层索引建立过程也在这一步进行，路由层、传输层索引则与IP、端口索引一同建立。