工程文档编号	
版本号	V1.0
作者	樊 荣
日期	
项目编号	WLCSY-0001
表格号	
模板版本	

10G 网络测试仪数据通道

概要设计

修订记录

版本	日期	作者	备注

目录

1 设计概述

1.1 目的

为在 NetFPGA10G 的 Reference NIC 工程基础上插入用户设计的 IP，实现 packet 的发射、捕获及参数统计等，对用户 IP 的设计进行说明，以方便项目的维护与交流。

1.2 设计依据

系统依据以太网测试的相关标准（具体参考 RFC 相关文档）进行设计，本版依据以太网(二层)测试标准 RFC2544 设计。

1.3 参考资料

- 支持远程可重配置的网络测试仪器研究与实现（BitTester）. 硕士论文，戴硕，2012.6.
- 网络性能测试与分析.林川，施晓秋，胡波，高等教育出版社，2011.8.
- RFC2544 标准 以太网(二层)测试方案 V1.0.彭鹏.
- 10G 网络测试仪方案书 V1.0. 樊荣.
- A Packet Generator on the NetFPGA Platform. G. Adam Covington, Glen Gibb, John W. Lockwood, Nick McKeown.
- An Open-Source Hardware Module for High-Speed Network Monitoring on NetFPGA. Gianni Antichi, David J. Miller, Stefano Giordano.

1.4 术语和缩写词

2 系统架构框图

2.1 Reference NIC 架构解析

由于系统在 NetFPGA10G 的 Reference NIC 基础上进行设计，首先对 Reference NIC 的架构进行介绍，Reference NIC 的核心架构如图 2.1 所示：

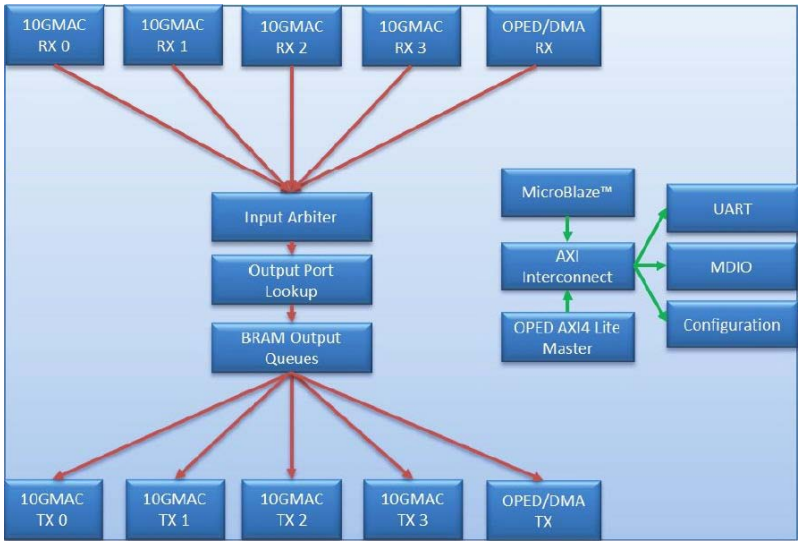

图 2.1 Reference NIC 架构

（1）RX

RX 分为 4 个 GMAC 的 RX 以及带有 DMA 功能的 PCIe 接口 OPED 共 5 个 IP core。RX0~RX3 都包含一个队列，用于缓冲从以太网接口收到的数据报文，并且根据不同的 RX，将入端口号填充到 AXI4-Stream 的 TUSER 通道 SRC_PORT 域中。OPED 为 PC OS 虚拟 4 个 RX，分别对应以太网口的映射关系，四个虚拟 RX 根据 AXI4-Stream 总线中的 TUSER 通道中的 SRC_PORT/DST_PORT 域区分开来。

TUSER 通道中 SRC_PORT 和 DST_PORT 域都占 8bits，分别为源输入端口和目的输出端口号，编码方式采用独热码，4 个 eth 口对应奇数比特，OPED 映射的 4 个口对应偶数比特。例如 eth0 即为 8'b0000_0001，OPED 中与 eth0 相映射的口为 8'b0000_0010。

（2）Input Arbiter 输入仲裁

输入仲裁是主数据通路的起始点，该 IP core 主要负责将 5 个输入 IP core 的并行数据通路转换为串行数据通路。工程采用的仲裁算法为轮询方式，即从 RX0 开始循环，发现某一 RX 队列非空就将数据从队列中拉出，传递给下一模块。

（3）Output Port Lookup 输出端口查找

输出端口查找主要负责读取报文的 TUSER 通道 SRC_PORT 域的内容判断报文来自哪个 RX，将来自 eth 的报文的 DST_PORT 域填充为 OPED 对应的映射口，将来自 OPED 的报文 DST_PORT 域填充为映射的 eth 口，并传递给下一模块。

（4）BRAM Output Queues 输出队列

输出队列是主数据通路的终点，主要负责缓存待输出的数据报文，包含 5 个队列，根据 TUSER 通道中的 DST_PORT，将输出至 eth 口的报文存储到对应的 4 个队列中，输出到 OPED 虚拟 4 个 TX 的报文全部存储到一个队列中。

（5）TX

TX 与 RX 相对应，他们将输出队列的数据拉出，通过对应的 eth 或者 OPED 发出。

2.2 10G 网络测试数据通道架构

通过在 Reference NIC 的数据通道插入 IP 得到 10G 网络测试仪的数据通道架构，如“10G 网络测试仪方案书 V1.0”中图 5.2 所示。通过对“10G 网络测试仪方案书 V1.0”中图 5.2 所示的 10G 网络测试仪硬件结构的数据通道部分进行修改完善，得到图 2.2 所示的数据通道结构。图中，灰色背景部分模块为 NIC 系统中原有的模块，黑色背景部分模块为根据功能需求修改或设计插入的模块。

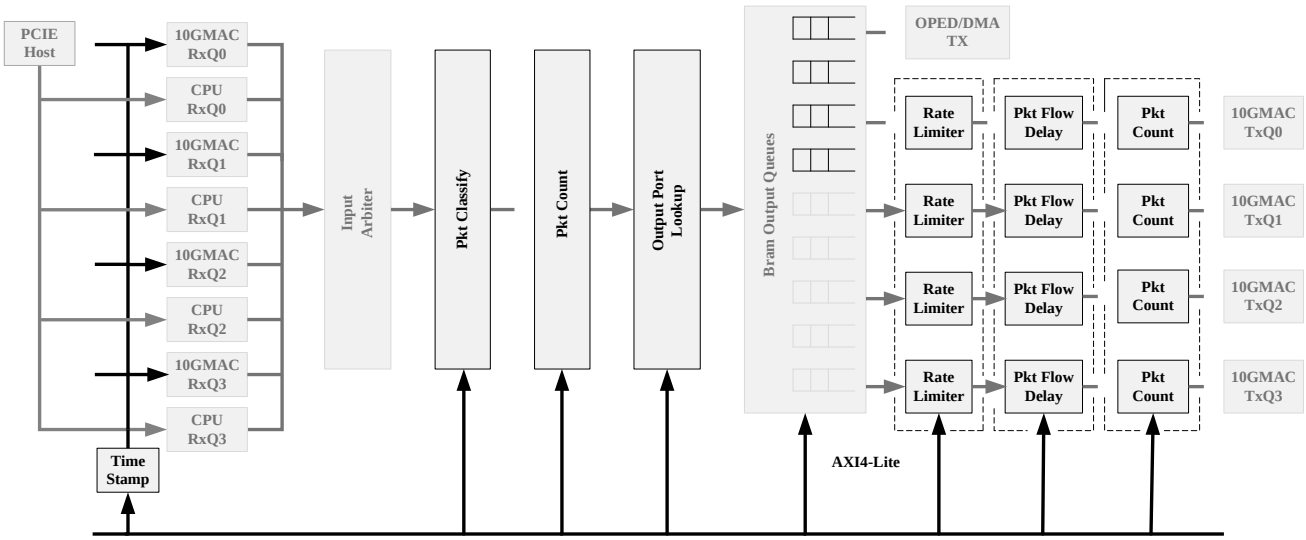

图 2.2 10G 网络测试仪数据通道结构

A．Time Stamp

Time Stamp 模块用于实现报文延时的计算。Time Stamp 模块挂载在 XGMII 接口之后、MAC 接收队列（10GMAC RxQ）之前，使 Time Stamp 模块尽量靠近 MAC，一旦接收到数据包可立即插入时间戳，以减小抖动和时间误差。MAC 核向 rx_queue 输出接收数据报文时，将使能控制信号 rx_data_valid，表明数据接口 rx_data 开始接收到有效数据，Time Stamp 模块内部设置状态机，通过检测 rx_data_valid 标志采样 Time Stamp 模块的计数值，将其作为接收该数据报文的时间标志。如图 2.2，Time Stamp 模块将在 rx_queue 模块内部被调用，设计需要对 rx_queue 及 nf10_axis_converter 模块的接口和逻辑进行修改，将采样到的时间戳插入到 AXI4-Stream 接

口向后续模块传递。时间戳附于 AXI4-Stream 接口的 TUSER 通道，TUSER 通道用于传输用户定义元数据，宽度为 128bit，NIC 系统对 TUSER 通道的格式定义如图所示：

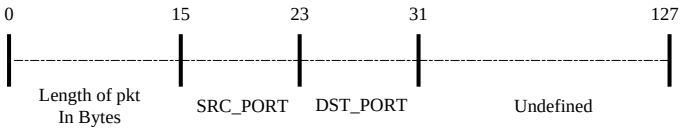

图 2.3 NIC 系统的 TUSER 通道格式定义

Time Stamp 模块内部采用 64 位时间戳计数器，64 位计数值可插入到图 2.3 所示 TUSER 通道的 Undefined 部分的位域。修改后的 TUSER 通道格式如图 2.4 所示，接收数据报文的时间戳插入到 TUSER 通道 32～95 位域。

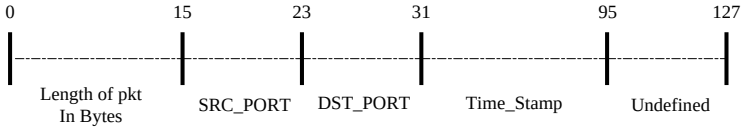

图 2.4 根据时间戳修改后的 TUSER 通道格式

B. Input Arbiter

Input Arbiter 模块直接采用自 NIC 系统，不作修改，模块采用轮询方式对 5 个输入 IP core（共 8 个 RX 队列）进行循环，将非空队列中的数据报文拉出传递给 Output Port Lookup 模块。

C. Pkt Classify

系统数据通道的后续模块对系统发送的数据报文进行参数统计，若接收到其他网络设备发送的数据报文则丢弃，不作统计。因此，需要设计数据报文对比分类模块 Pkt Classify，从接收流量中找出系统发送的测试报文。Pkt Classify 位于 Input Arbiter 模块之后，对 Input Arbiter 模块输出的数据报文进行处理。Pkt Classify 使用 TCAM（三态内容寻址存储器）模块实现，对接收数据报文头的 5 个元组（IP 地址对、协议和端口对）进行滤波匹配处理，匹配的数据报文在后续模块进行统计分析、通过 MAC 口重新发送或发送至 PC 进行进一步分析处理，否则抛弃报文，不作处理。Pkt Classify 通过 AXI4-Lite 总线挂载于 MicroBlaze，在发送测试报文前，MicroBlaze 将匹配表项写入 TCAM。

D. Pkt Count

Pkt Count 模块用于实现：

1） 对接收报文进行计数；

2） 剥离报文在接收时插入的时间戳，并存储于缓冲区等待控制层读取。

Pkt Count 模块对过滤后的数据报文进行计数统计，模块内部分别为 4 个 MAC 口设置计数

寄存器，对来自 4 个 MAC 口的数据报文进行计数统计，MicroBlaze 通过 AXI4-Lite 总线读取计数值。若不使能 Pkt Count 模块，则 Pkt Count 实现硬连线功能，实现前后两个模块的直接互联。模块内部分别为 4 个接收通道设置 FIFO，用于存储时间戳。

E. Output Port Lookup

Output Port Lookup 模块完成两个功能：

1） 报文转发：将 PC 通过 DMA 通道传递的报文从 MAC 口转发出去；

2） 报文采样：在 MicroBlaze 的控制下，采样接收到的报文并转发到 PC 进行分析。

NIC 系统中的 Output Port Lookup 将接收自 MAC 的数据报文通过 PCIE 接口转发到 PC，将 PC OS 的虚拟 RX 通道接收的数据转发至 MAC 口。由于 PC 的处理能力有限，在 4 个 MAC 口都以 10Gbs 速率进行数据发送的情况下，不能将所有从 MAC 接收的数据报文转发至 PC 分析处理。设计采用采样的方法进行处理，即将 Output Port Lookup 挂载于 AXI4_Lite 总线，MicroBlaze 控制数据从 MAC 口至虚拟 RX 通道的转发，Output Port Lookup 模块内部分别为 4 个 MAC 接收通道设置使能和采样计数寄存器，通过 MicroBlaze 使能数据从 MAC 口至虚拟 RX 通道的转发，采样计数寄存器对数据报文的采样间隔 N 进行控制，实现 1/N 的数据采样率（即每接收 N 个报文则采样 1 个转发到 PC）。

F. Bram Output Queues

Bram Output Queues模块在NIC系统的基础上进行修改，NIC系统的Bram Output Queues模块将接收的数据报文写入相应的5个FIFO（其中一个FIFO存储的报文发送到4个DMA虚拟TX通道），然后从相应的TX通道将报文从MAC转发至外部网络或者转发至PC。本系统设计的Bram Output Queues模块增加4个FIFO（共9个FIFO），其中4个FIFO用于存储从MAC口接收的报文（报文接收FIFO），用于循环发送；4个FIFO用于存储待发送的报文内容（报文发送FIFO），用于初次发送测试报文，报文内容由PC通过4个虚拟4个RX写入；最后一个FIFO用于将从MAC口接收的报文转发到4个DMA虚拟TX通道。在测试开始前，由PC通过DMA虚拟RX通道将测试报文写入报文发送FIFO，测试开始后，模块在状态机控制下从报文发送FIFO将报文拉出写入MAC口发送出去，后续发送则分两种情况：1）若报文接收FIFO处于空状态（没有接收报文写入），则从报文发送FIFO读取报文发送；2）在接收FIFO处于非空状态下（有接收报文写入），则从报文接收FIFO读取报文发送。

G. Rate Limiter

Rate Limiter 模块用于控制数据流速，将发送到 MAC 口的报文的速率强制控制在用户给定

的速率范围内。

H. Pkt Flow Delay

Pkt Flow Delay模块控制报文与报文之间的时间间隙。

I. Pkt Count

与接收报文计数模块对应，该模块用于对发送报文进行计数。内部设置发送数目寄存器，用于设定预定发送的帧数量，达到该发送数量时停止对外发送测试报文。

注：图2.2所示结构图中虚线框中的模块将作为一个模块整体实现。

3 系统原理分析

第2节对数据通道的实现架构以及通道中每个功能模块进行了介绍，本节通过介绍数据通道的数据流和控制方法分析其工作原理。

3.1 数据通道数据流分析

（1） 数据通道的初始化

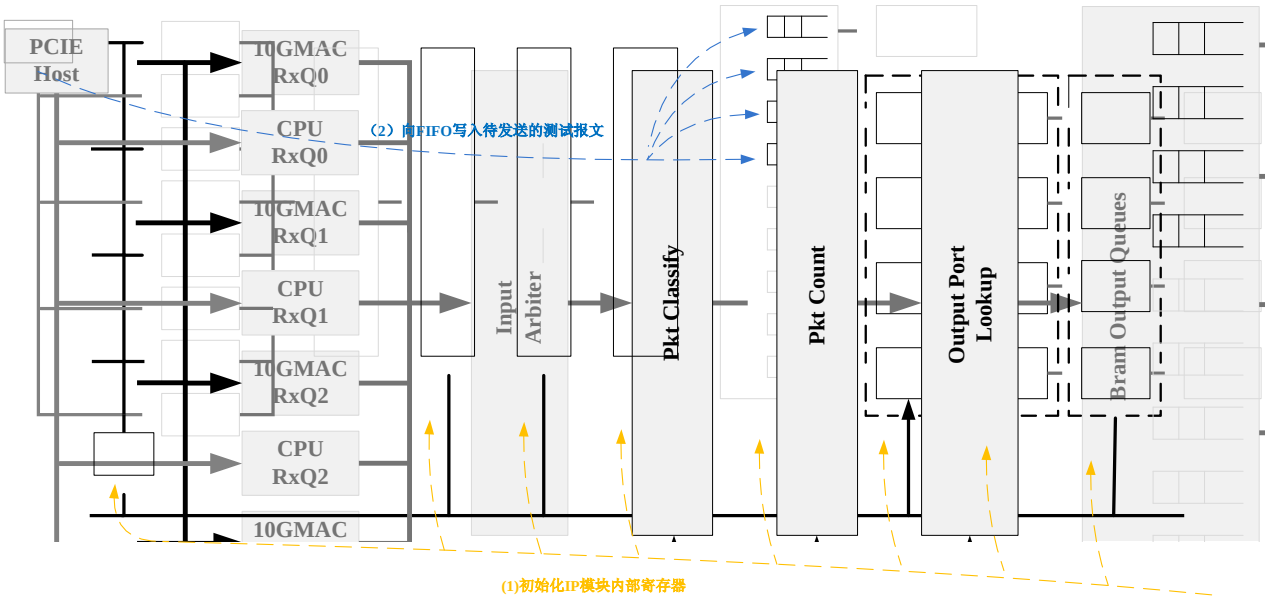

图3.1 数据通道初始化示意图

如图3.1，系统开始工作之前，需要对数据通道进行初始化，初始化分两步完成：

1）在MicroBlaze控制下，通过AXI4-Lite总线配置模块内部寄存器状态，包括时间戳计数器清零、对Pkt Classify模块写入匹配表项、报文计数寄存器清零，以及发送、接收端口配置（具体配置视外部连接方式而定）等；

2）通过PCIE DMA将待发送的测试报文写入FIFO，写相关模块使能寄存器，启动数据通

道工作。

每次测试开始或者重新进行测试时，反复步骤1）和2）。

（2） 系统工作过程数据流分析

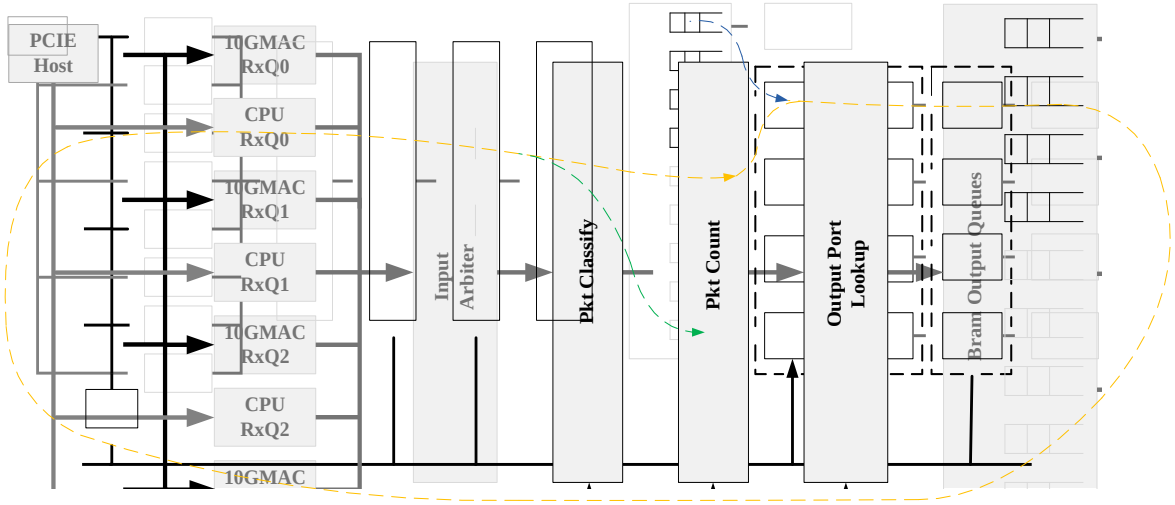

图3.2 系统工作过程数据流示意图

图3.2所示为从MAC0发送测试流、从MAC1进行接收（DUT连接在MAC0和MAC1）的测试示例。在进行正常测试时，系统初始化后首先从发送FIFO将测试报文拉出（图中蓝色箭头线所示）；接收报文经前级模块处理后存储在接收FIFO，用于循环发送（图中黄色箭头线所示）；在采样模式下，被采样的接收报文写入与DMA虚拟TX通道相联的FIFO（图中绿色箭头线所示），通过PCIE接口发送到PC。

3.2 RFC2544测试指标的实现方法

本小节根据“RFC2544标准以太网(二层)测试方案V1.0”中描述的RFC2544测试指标分析其在该数据通道架构下的实现方法。

（1）吞吐量测试实现方法

吞吐量即设备不丢帧情况下单位时间内的最大帧转发数量。该指标测试需要通过3个参数得到：媒质速率(Mbps)、帧长以及帧间隔。首先需要确定测试帧长，该参数由上位机软件生成（测试帧由上位机生成）；媒质速率(Mbps)由限速模块RateLimiter确定（通过寄存器设定）；帧间隔由延时模块PktFlowDelay确定。在上述3个参数确定后发送码流进行测试，控制层通过读取发送帧计数模块和接收帧计数模块Pkt Count判断是否有丢帧情况，若出现丢帧情况，则增大帧间隔开始新一轮测试，若没有出现帧丢失，则减小帧间隔开始新一轮测试，如此反复直到测试出在没有出现帧丢失情况下的最小帧间隔。

（2）（延迟）等待时间测试实现方法

延迟测试确定报文经过DUT传输所需要的时间，延迟测试需测得两个参数：

1） 输入报文的最后一位到达输入端口的时刻；

2） 输出报文的第一位出现在输出端口的时刻。

按照文档“RFC2544标准以太网(二层)测试方案V1.0”中所述的（延迟）等待时间测试指标测试，测试方法为：在测试流中插入带有特殊标记的报文（图3.3），该报文的特殊位置加入特殊标记（Tag），在发送该报文时记录该报文的时间戳A，在接收该报文时记录该报文到达的时间戳B，延迟即时间戳B减去时间戳A。

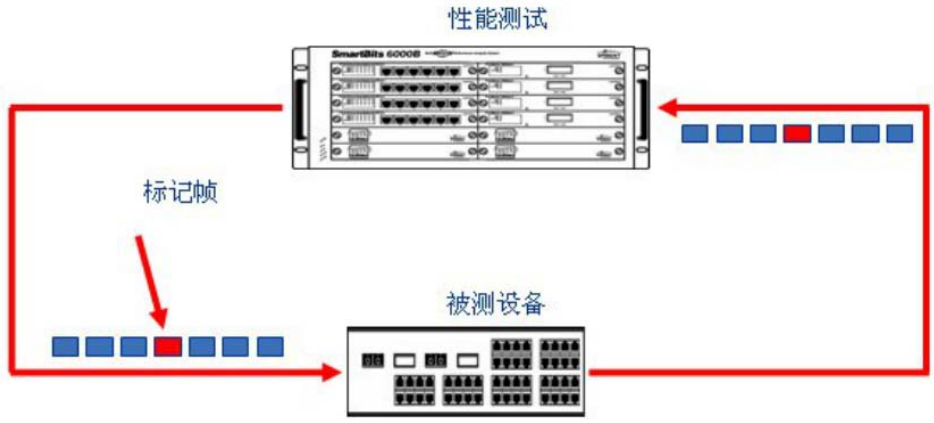

图3.3 RFC2544标准的延迟测试建议方法

该测试方法要求：1）标记报文在传输过程中不能丢失；2）标记报文在被转发的时候网络设备需工作在稳定状态，即标记报文设置在测试流的中间；3）测试设备必须能够识别在报文中得的标记信息。

该测试方法存在某些局限：

1） 在测试流量中，将中间有个标记报文的延迟测试结果作为整个测试的结果；

2） 必须在要在没有丢包的条件下测试，即必须先测试吞吐量；

3） 单次测试结果的偏差可能较大，需要对20次以上的结果进行平均；

4） 系统需要实时辨别出标记报文，实现复杂度较大。

本系统采用如下解决方案：对接收到的所有数据报文的延迟进行统计，需要对图2.2中的数据通道架构进行修改，在MAC发送队列（10GMAC TxQ）之后插入发送时间戳模块（Time Stamp），如图3.4。通过报文发送端和接收端的时间戳模块（Time Stamp），所有报文的发送时刻和接收时刻将被标记。发送时间戳模块内部设置时间戳存储缓冲区（FIFO），用于存储一次延迟测试中发送的所有报文的时间戳。则平均延迟可计算如下：

$$\frac{(B_1-A_1)+(B_2-A_2)+\cdots+(B_N-A_N)}{N}=\frac{B_1+B_2+\cdots+B_N}{N}-\frac{A_1+A_2+\cdots+A_N}{N}=\overline{B}-\overline{A}$$（3-1），分别对所有接收报文时间戳和发送报文时间戳取平均，得到报文平均接收时间戳$\overline{B}$和报文平均发送时间戳$\overline{A}$，，二者相减即延迟平均值。按照式（3-1）计算平均时间戳的前提是发送报文与接收报文数量相同（均为N），即要求传输过程中没有报文丢失，所以需要针对每一特定的报文长度，以不超过吞吐量的发送速率发送测试流，所以该测试需要在吞吐量测试之后进行。

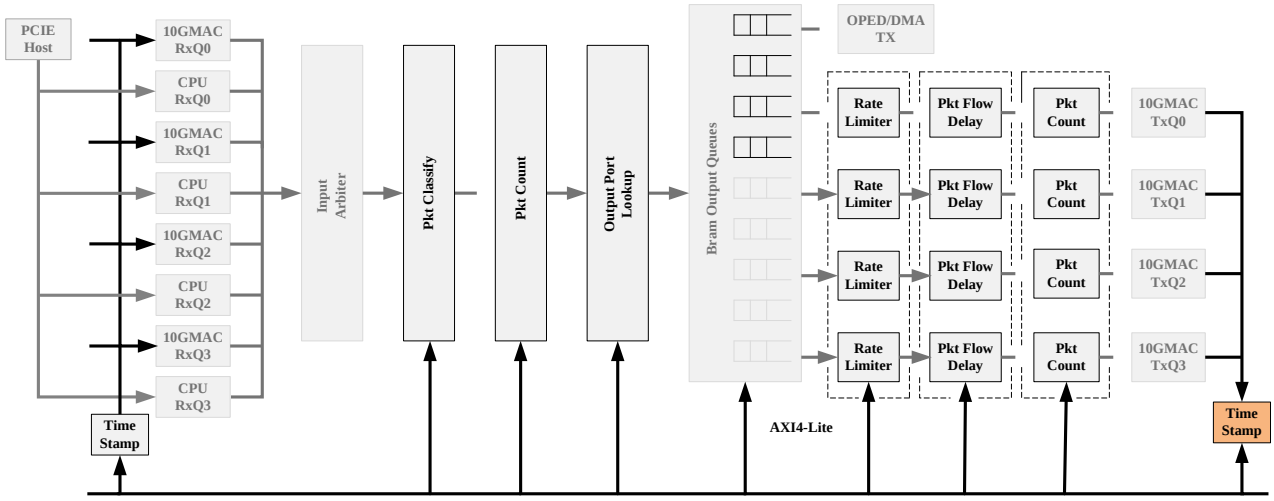

图3.4 延迟测试数据通道结构

（3）帧损失率测试实现方法

帧损失率测试即测试DUT在不同负荷下丢弃包占的比例。不同负荷即从吞吐量到线速（线路上传输包的最高速率），步长使用线速的10%。控制层对RateLimiter模块和PktFlowDelay模块设置好发送速率和报文间隔，启动测试，通过发送Pkt Count模块和接收Pkt Count模块可计算出每次测试的丢帧率。

（4）背对背测试实现方法

背对背测试通过向DUT发送具有合法最小帧间隔的突发数据包，确定DUT在不丢包情况下能够处理的最大数据包数目。测试开始前，对Pkt Count模块设置预发送报文数量，对RateLimiter模块和PktFlowDelay模块设置发送速率和报文间隔，随后启动测试，如果出现报文丢失，则减少报文数量（对Pkt Count模块设置预发送报文数量）重新发送测试，否则增加报文数量。在控制层的控制下，该测试最终要保证：1）发送具有合法的最小帧间隔的突发数据报文的持续时间必须大于等于2S；2）测试应该至少进行50次；3）最终取平均值。