高性能软件 IPv6 网络测试仪及相关研究

题目背景

《计算机组成原理》与《计算机网络原理》联合硬件路由器实验（简称“计网联合实验”）要求同学们在 FPGA 上实现一台硬件路由器。为了对路由器进行功能、性能测试，需要有网络测试仪（由软件或硬件实现）的支持。

路由器测试的主要功能和性能指标为：

- 连通性：接入路由器的主机之间两两连通
- 吞吐率
- （小包）转发速率
- 路由表容量

参考资料：

- 计网联合实验综总述：**https://lab.cs.tsinghua.edu.cn/router/doc/joint/**
- 实验评测技术方案：**https://lab.cs.tsinghua.edu.cn/router/doc/joint/eval/**
    - 2021 年的计网联合实验将改为 IPv6 单栈，在此文档中所有与 IPv4 协议相关的内容均需修改

问题提出

基于硬件的网络测试仪能够胜任此测试任务，但是灵活性不够高，硬件逻辑编译一次需要数十分钟至数小时。因此实现基于软件的网络测试仪是可以尝试的。
计网联合实验的路由器目前有 4 个千兆以太网（IEEE 802.3ab）接口。在性能测试中，这些接口需要同时以线速（64B 小包、千兆）同时收发数据，一共 8000Mbps 吞吐量。这对于软件网络测试仪来说是个挑战：如果直接使用 Linux 提供的 socket 相关系统调用，编写以下 C 语言程序，单线程发送速率仅能达到数十Mbps。
（程序待补充；也请同学们复现）

因此首先需要解决的问题是，如何利用软件实现千兆以上的小包线速收发。
如果此问题能够被顺利解决，那么下一步可以尝试研究：

1. 如何利用软件实现更快的线速收发（如万兆以太网）
2. 如何利用小包线速收发来构建软件网络测试仪（连通性、吞吐率、转发效率、路由表容量）

在上述问题中，主要的困难与挑战可能有：

1. 如果使用 Linux 提供的 socket 相关系统调用收发包，则用户态—内核态之间的上下文切换是否可能成为瓶颈？（1 秒钟能切换多少次？）
2. 即使不存在上下文切换，如果有复杂的网络协议栈存在，各层协议的封包逻辑，和层间的内存拷贝，是否可能成为瓶颈?
3. 如果想绕过 Linux 的 socket 系统调用，绕过 Linux 的网络协议栈，这件事情是否容易直接实现？是否有必要绕过整个 Linux 内核（甚至重写内核）？
4. 在进行小包线速（千兆为 1.488Mpps）的收发过程中，软件网络测试仪不应该受到任何外部中断的影响，因为一旦中断导致漏发或漏收一个包，则测试结果将不正确。如何确保不受中断影响?
5. 网络测试仪测试的是路由器的 **转发效率**，因此在保证线速收发的同时，还需要线速生成将要发给路由器的包，以及线速验证路由器转发出的包是否正确。这里既要验证每个包的格式和内容，也要查找路由表验证路由器转发的出接口是否正确。线速执行上述检查，是否存在性能上的困难?

相关工作和资源

1. “应用程序稳态测试系统”， **https://github.com/JudgeDuck/JudgeDuck-OS** 。该工作基于 MIT 的 JOS 教学操作系统，进行了多项内核修改，实现了用户态程序在 **无中断、无缺页、无系统调用** 的条件下运行，以改进编程竞赛中的程序运行时间测量结果。
2. “测测你的路由器 - 题目 - Judge Duck Online”， **https://duck.ac/problem/router32** 。该题目可测试一个 IPv4 路由表查找算法的运行效率。截至 2021 年 9 月 17 日，该网站上存在 1 秒钟大约能查询 10^7 次路由表（随机被查询的 IP）的算法（如 **https://duck.ac/submission/10642** ）。
3. “计网联合实验” 教学团队：由全成斌老师牵头，谭闻德同学作为主要的助教。从 2019 年秋季学期开始，该团队已带领了两届本科同学挑战计网联合实验。每届约有 20 位同学，三人一组，所有组均实现了硬件路由器。在 2020 年秋季学期所有组均实现了能够（千兆）线速转发的硬件路由器。
4. “The C10M Problem”， **http://c10m.robertgraham.com/p/manifesto.html** 。该问题是指在一个服务器上同时维持 10M（10^7）个网络连接。这一系列技术博客介绍了该问题的主要困难，以及如何解决，并在 Linux 服务器上实现。
5. “The Data Plane Development Kit”， **https://dpdk.org** 。这是一套用于加速网络包处理的框架，其核心思想是在用户态（不经过内核地）运行网络包处理程序。