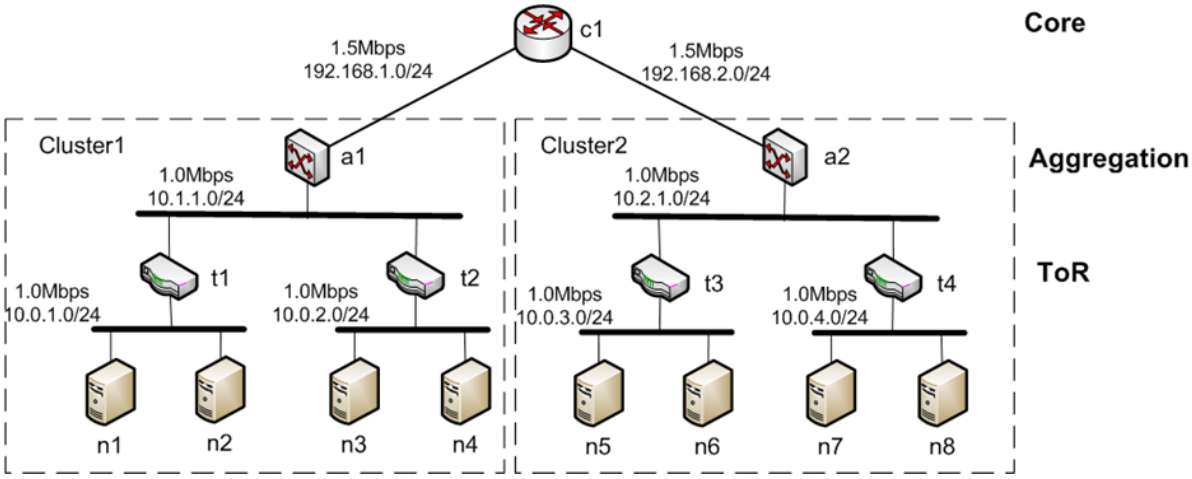
reporter

实验 1： inter cluster traffic

网络拓扑



实验流程

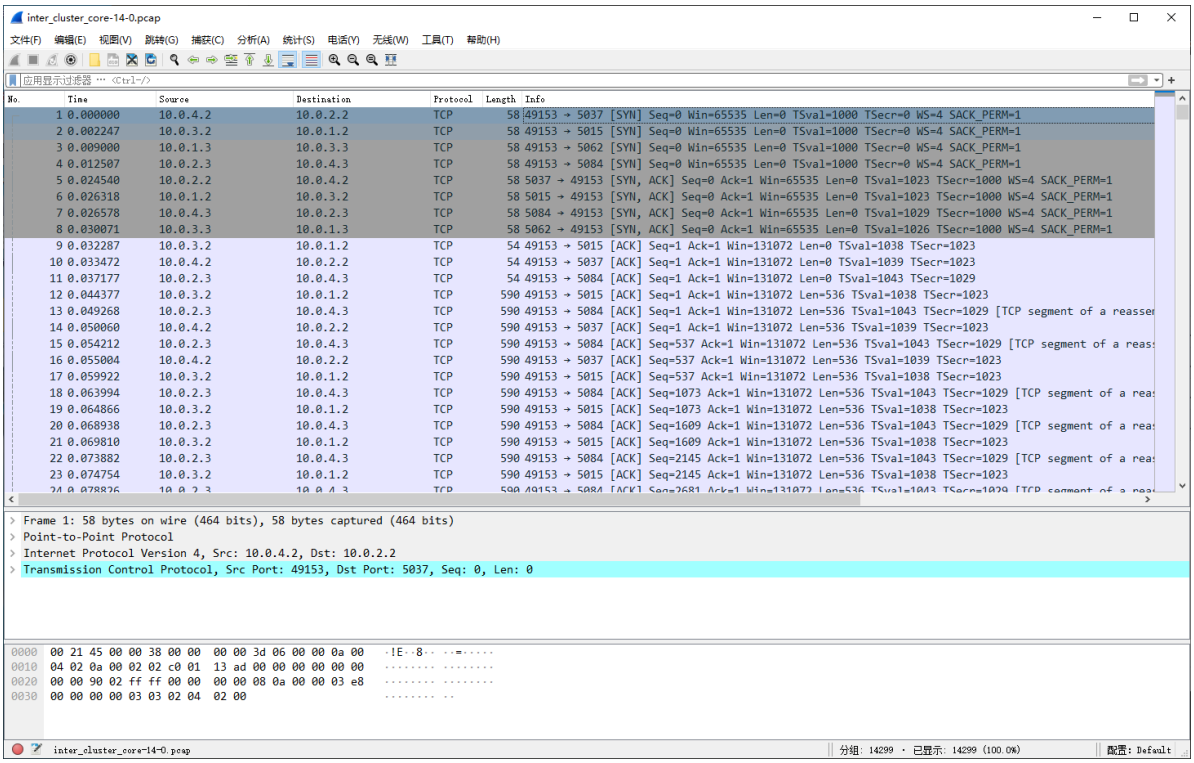
- 初始化各节点，将各节点加入对应的的网段
- 分别给各网段建立相应速率和延迟的信道，对网段上的节点分配网卡设备
- 给所有节点安装协议栈
- 对不同网段的设备分配相应的IP地址
- 建立TCP连接（**server <------> client**）1-5， 6-2，3-7，8-4

TCP	server ip	server port	client ip	client port
n1-n5	10.0.1.2/24	5015	10.0.3.2/24	49153
n6-n2	10.0.3.3/24	5062	10.0.1.3/24	49153
n3-n7	10.0.2.2/24	5037	10.0.4.2/24	49153
n8-n4	10.0.4.3/24	5084	10.0.2.3/24	49153

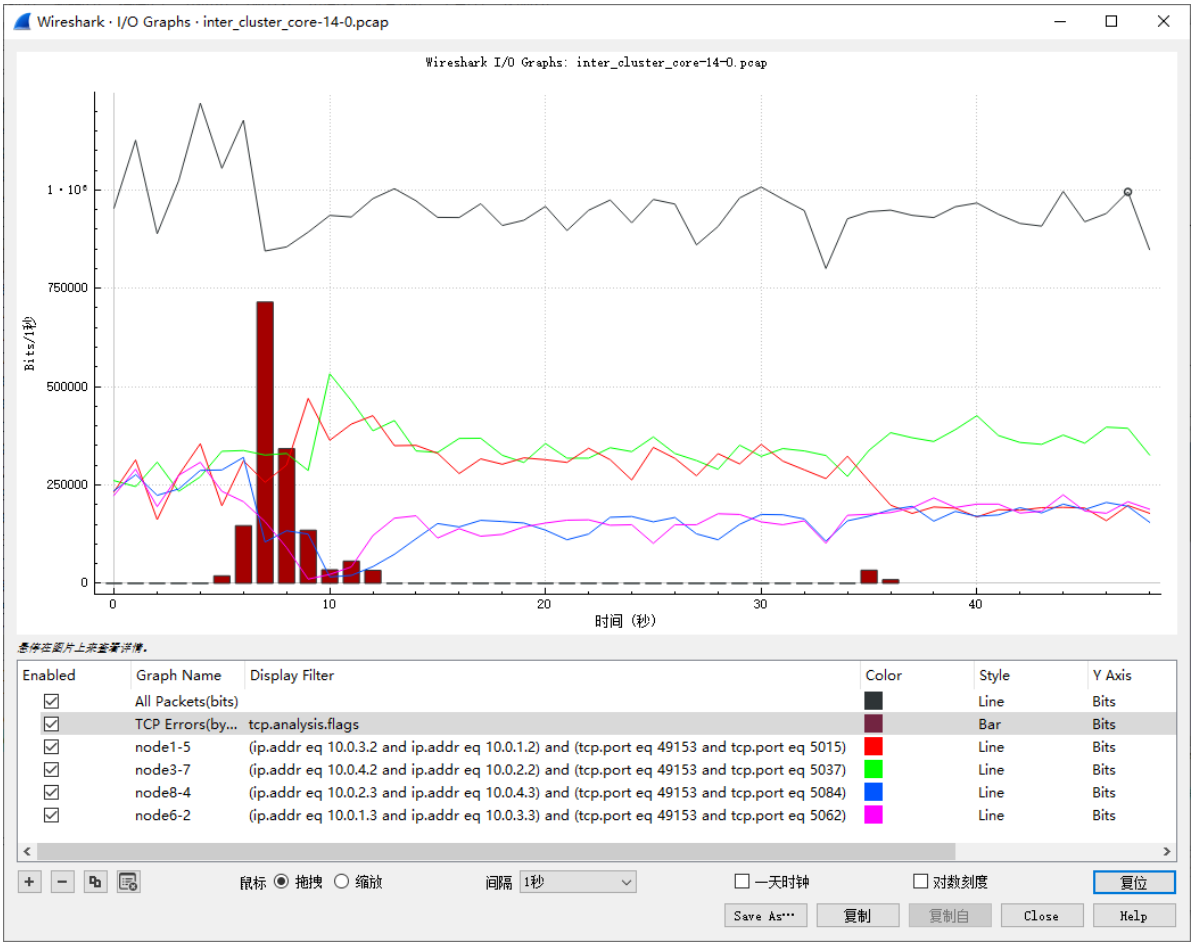
- 设置并保存节点**n1, t1, a1, c1**上的TCP流

结果

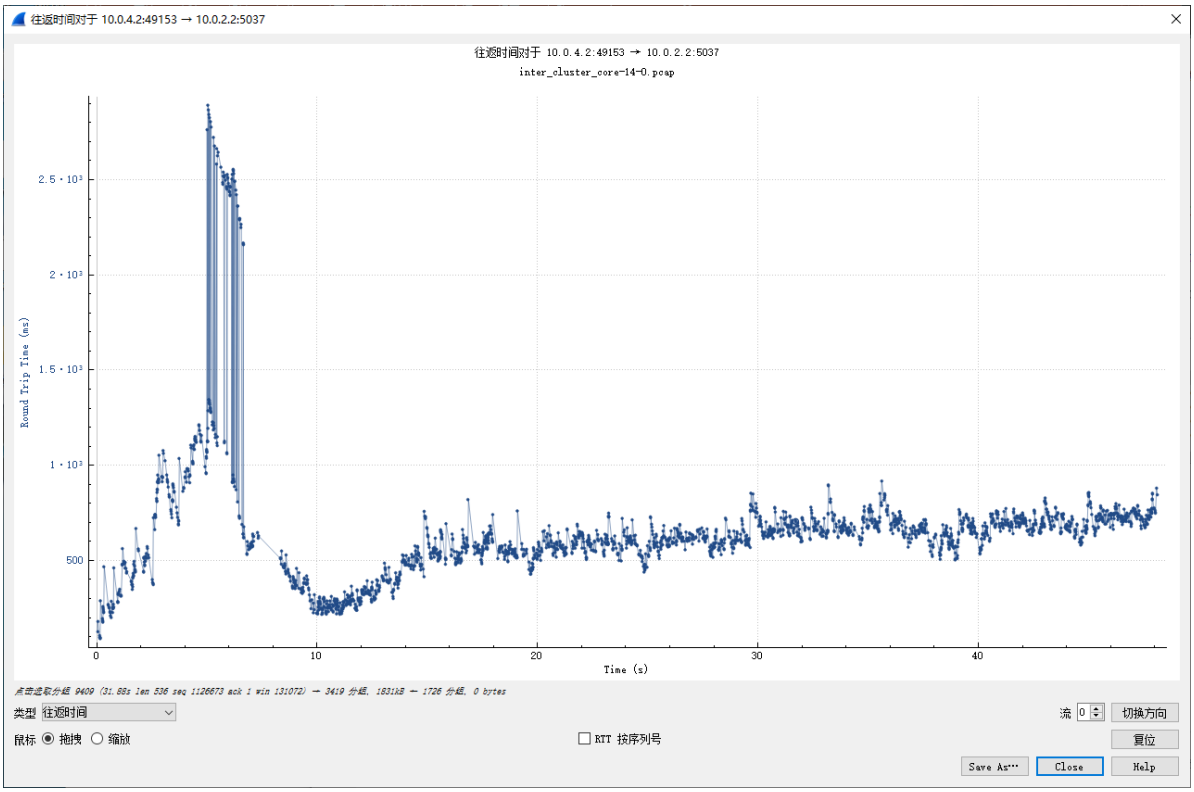
- **wireshark TCP抓包**



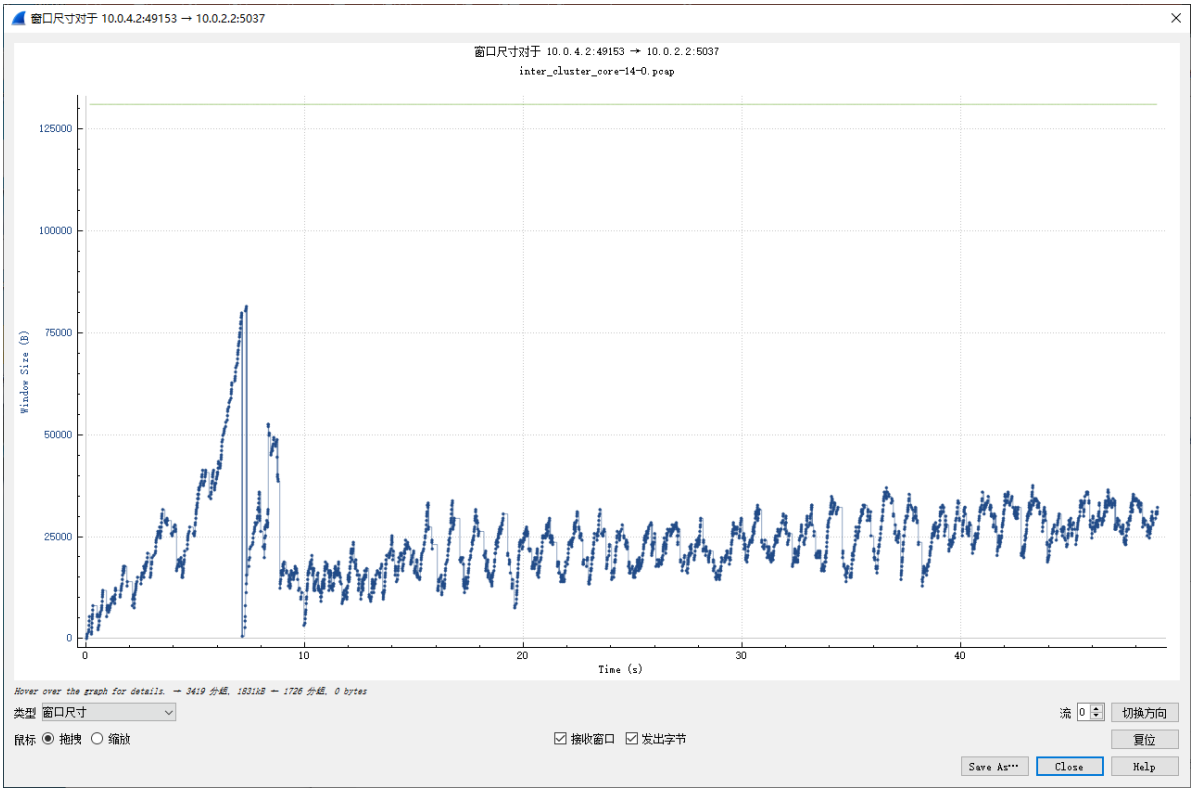
- **核心交换机上的TCP吞吐量**



- **tcp n3-n7上的往返时延：**



- **tcp n3-n7上的窗口变化：**



分析

- **丢包原因：**

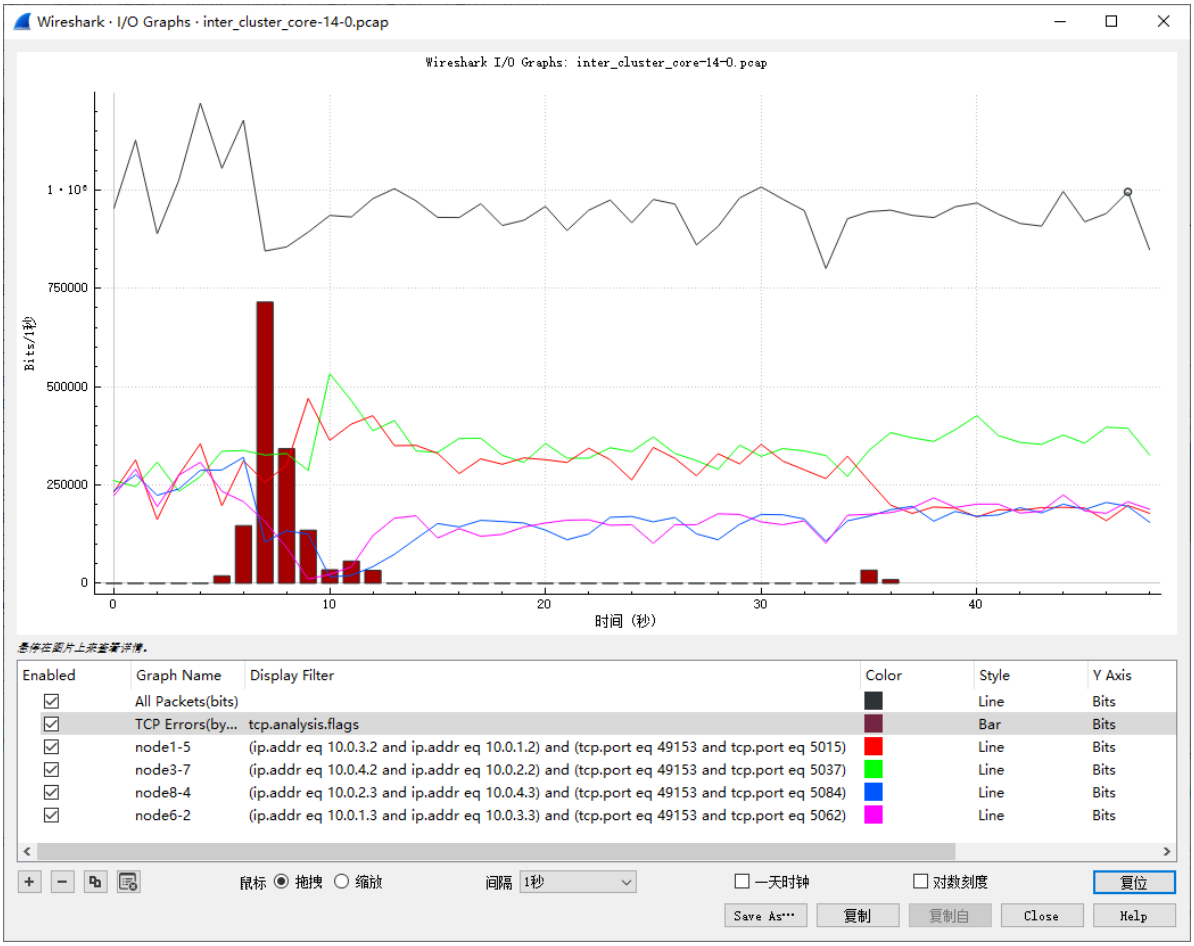
刚开始网络上面没有包，所以各TCP流的窗口大小有序增加。随着信道上的数据包越来越多，逐渐达到信道极限，四个TCP流会同时发生拥塞，所以在第6s-10s各tcp流的窗口大小会急剧减小。同时发生丢包的可能性很低，所以后面各流逐渐趋向稳定。

- **各TCP流平均吞吐率只有0.25Mbps：**

在ToR层，两台服务器（例如10.0.1.0/24网段的n1, n2）竞争1M的带宽，各服务器的平均带宽只有0.5M。而在Aggr层，两台交换机（例如10.1.1.0/24网段的t1, t2）竞争1M的带宽，所以每台ToR交换机分的平均带宽只有0.5M，从而导致**每台服务器的平均带宽只有0.25M**，与实验结果相符。由于**Aggr交换机达到性能瓶颈，最大带宽为1M**，所以**core交换机的两条点对点链路1.5M并没有到达性能瓶颈。**

改进方法

- 增加ToR和Aggr之间链路的带宽：10.1.1.0/24网段和10.2.1.0/24网段所在的链路带宽从**1.0Mbps**增加到**2.0Mbps**



可以看到核心交换机上的数据吞吐率**增加将近一倍。**

实验 2： many to one traffic

网络拓扑

同实验1

实验流程

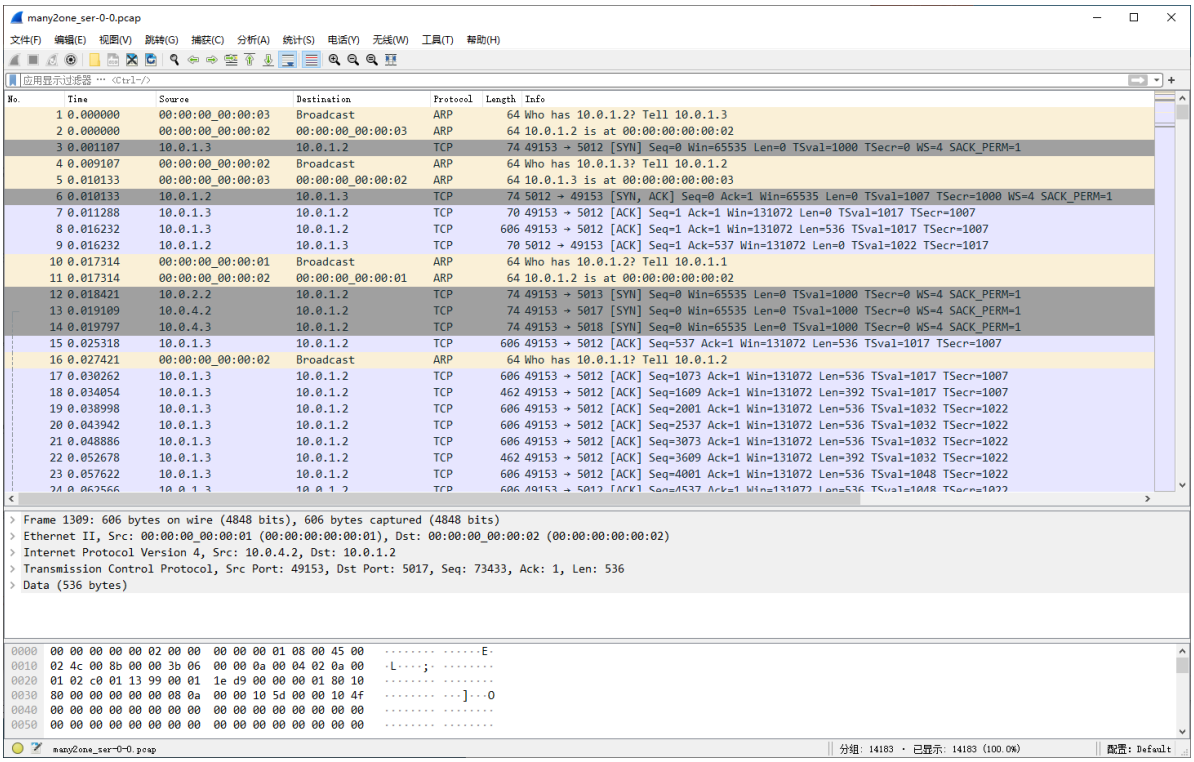
同实验1，仅在建立TCP连接处做了修改。

- 建立TCP连接（**server <------> client**）1-2，1-3，1-4，1-5，1-6，1-7

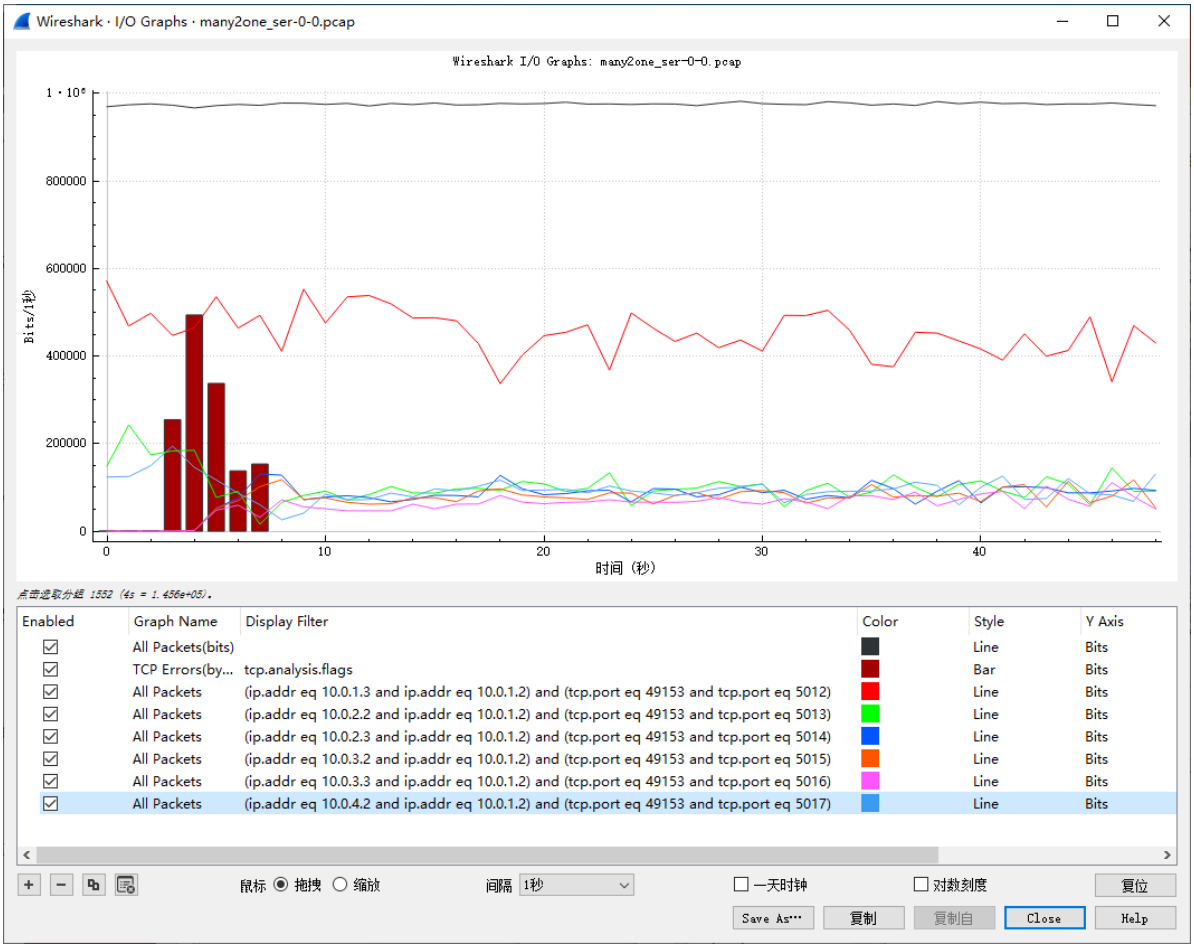
TCP	server ip	server port	client ip	client port
n1-n2	10.0.1.2/24	5012	10.0.1.3/24	49153
n1-n3	10.0.1.2/24	5013	10.0.2.2/24	49153
n1-n4	10.0.1.2/24	5014	10.0.2.3/24	49153
n1-n5	10.0.1.2/24	5015	10.0.3.2/24	49153
n1-n6	10.0.1.2/24	5016	10.0.3.3/24	49153
n1-n7	10.0.1.2/24	5017	10.0.4.2/24	49153

结果

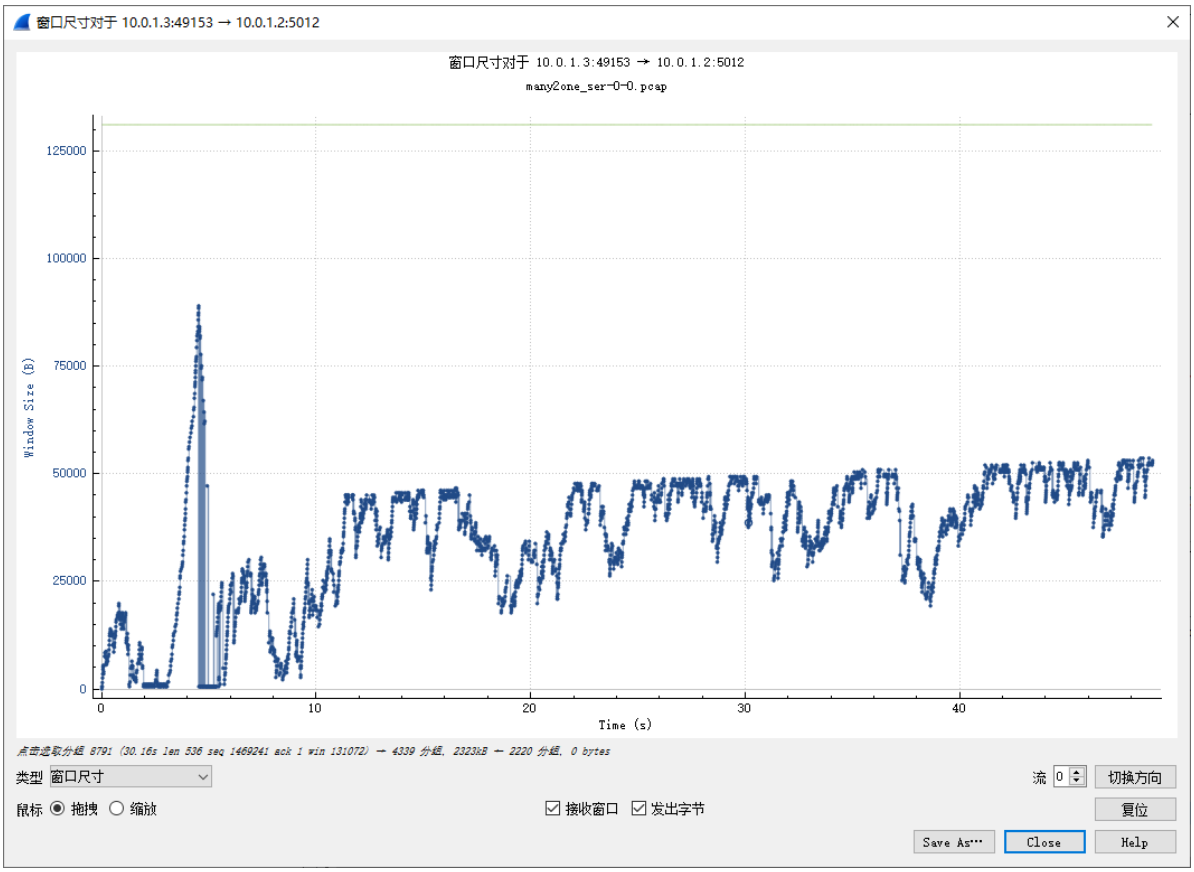
- n1 server上的流量

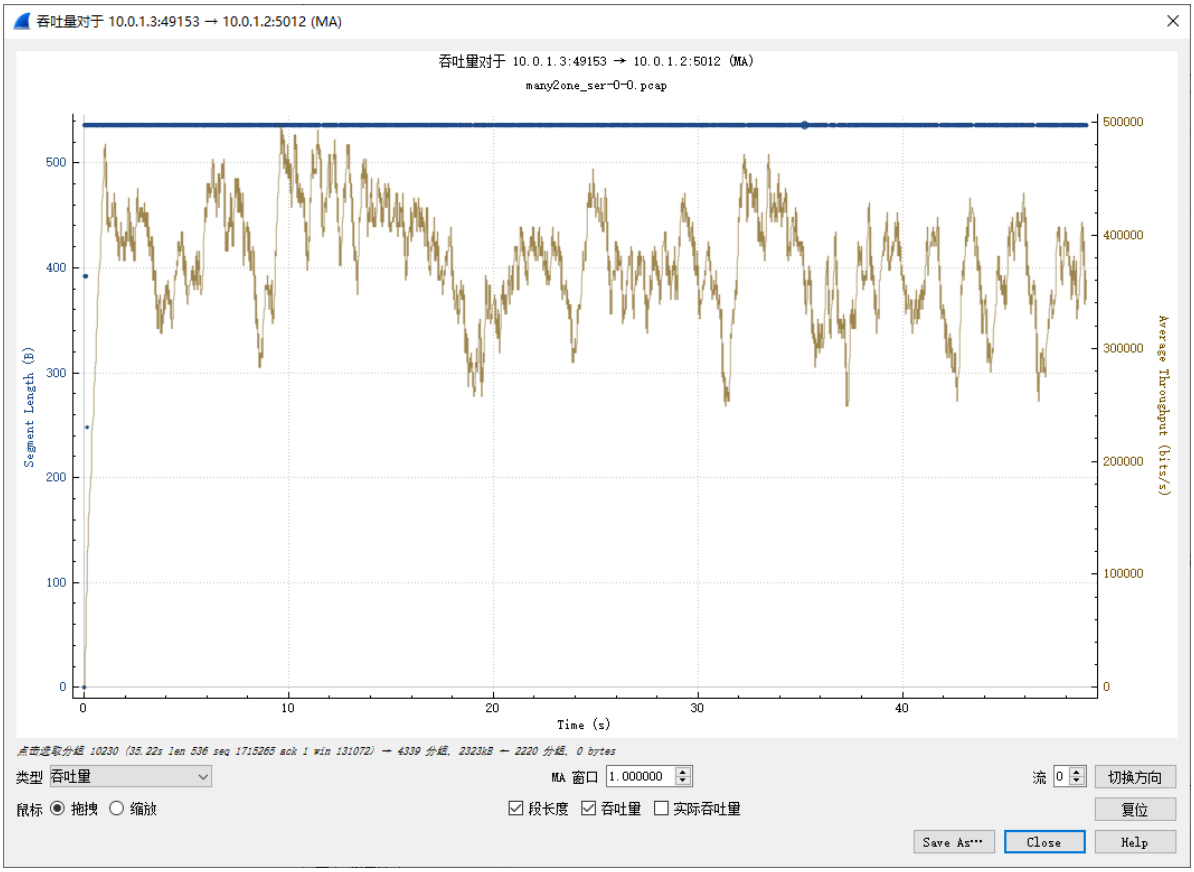


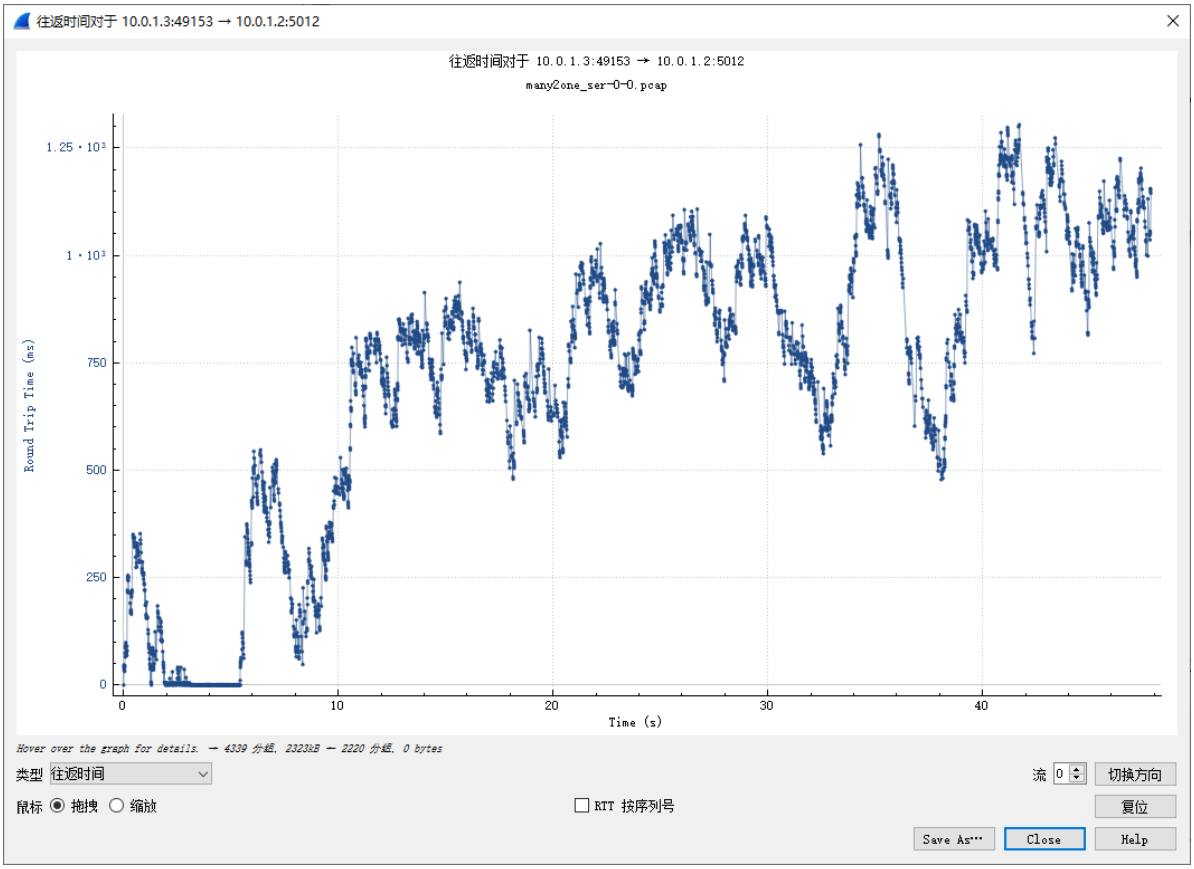
- 各TCP连接吞吐量



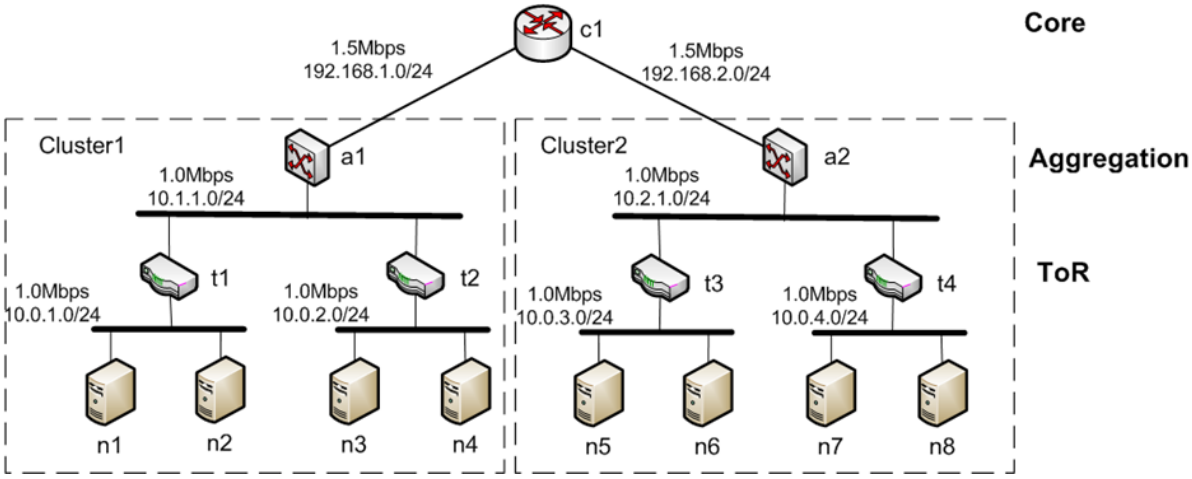
- TCP n1-n2 窗口、吞吐量、时延

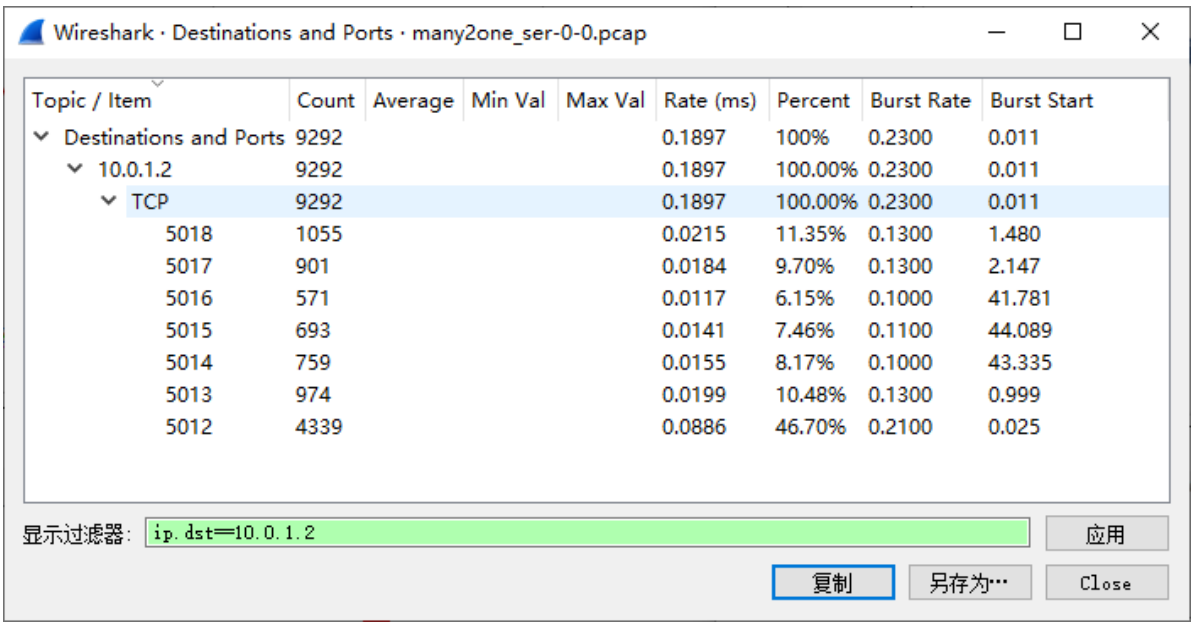



吞吐量对于 10.0.1.3:49153 → 10.0.1.2:5012 (MA)
吞吐量对于 10.0.1.3:49153 → 10.0.1.2:5012 (MA)
many2one_ser-0-0.pcap
Segment Length (B)
Average Throughput (bits/s)
Time (s)
类型 吞吐量
MA 窗口 1.000000
流 0
切换方向
鼠标 拖拽 缩放
段长度 吞吐量 实际吞吐量
复位
Save As…
Close
Help


往返时间对于 10.0.1.3:49153 → 10.0.1.2:5012
往返时间对于 10.0.1.3:49153 → 10.0.1.2:5012
many2one_ser-0-0.pcap
Round Trip Time (ns)
Time (s)
Hover over the graph for details. → 4339 分组, 2323kB ← 2220 分组, 0 bytes
类型 往返时间
流 0
切换方向
鼠标 拖拽 缩放
RTT 按序列号
复位
Save As…
Close
Help

分析





Topic / Item	Count	Average	Min Val	Max Val	Rate (ms)	Percent	Burst Rate	Burst Start
Destinations and Ports	9292				0.1897	100%	0.2300	0.011
10.0.1.2	9292				0.1897	100.00%	0.2300	0.011
TCP	9292				0.1897	100.00%	0.2300	0.011
5018	1055				0.0215	11.35%	0.1300	1.480
5017	901				0.0184	9.70%	0.1300	2.147
5016	571				0.0117	6.15%	0.1000	41.781
5015	693				0.0141	7.46%	0.1100	44.089
5014	759				0.0155	8.17%	0.1000	43.335
5013	974				0.0199	10.48%	0.1300	0.999
5012	4339				0.0886	46.70%	0.2100	0.025

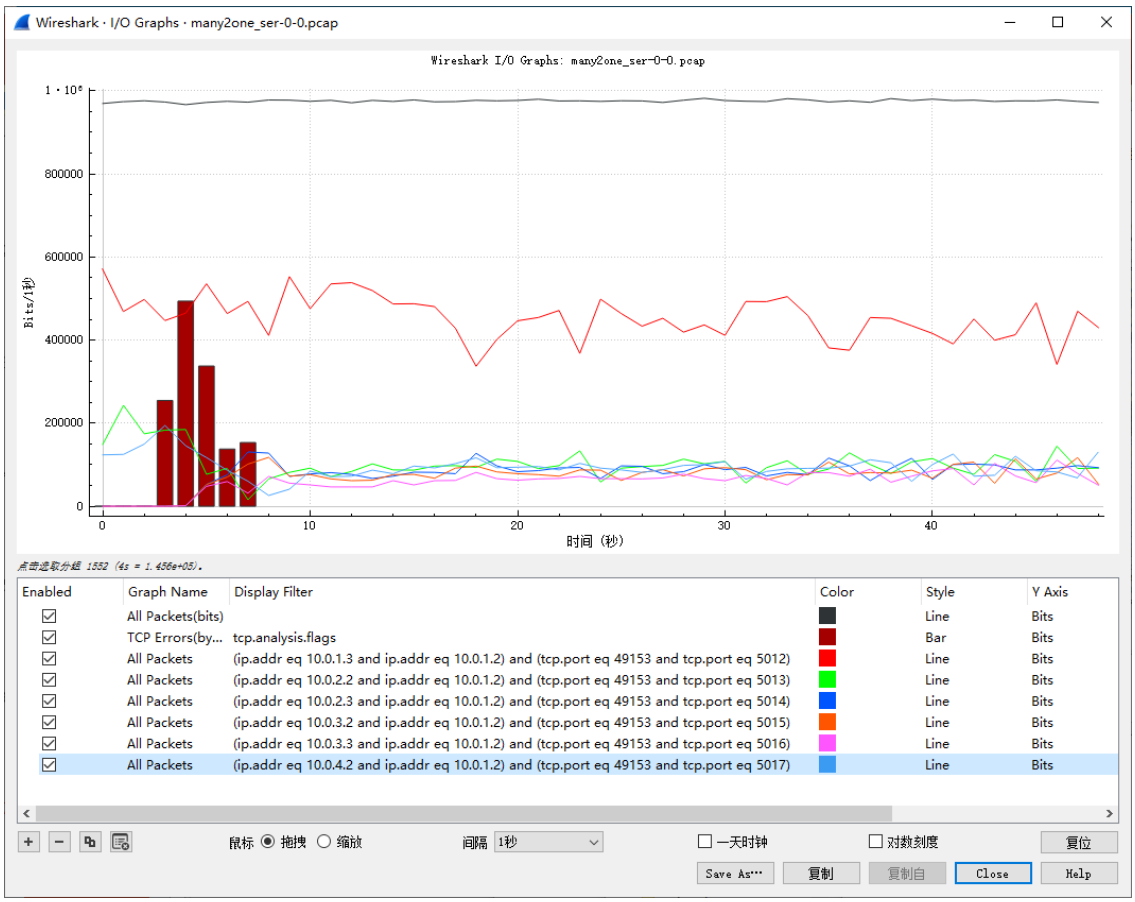
可以看到，在n1服务器与其他服务器建立的TCP流中，n1-n2带宽占比达到46%，剩下的其他6个TCP流平均占比9%。

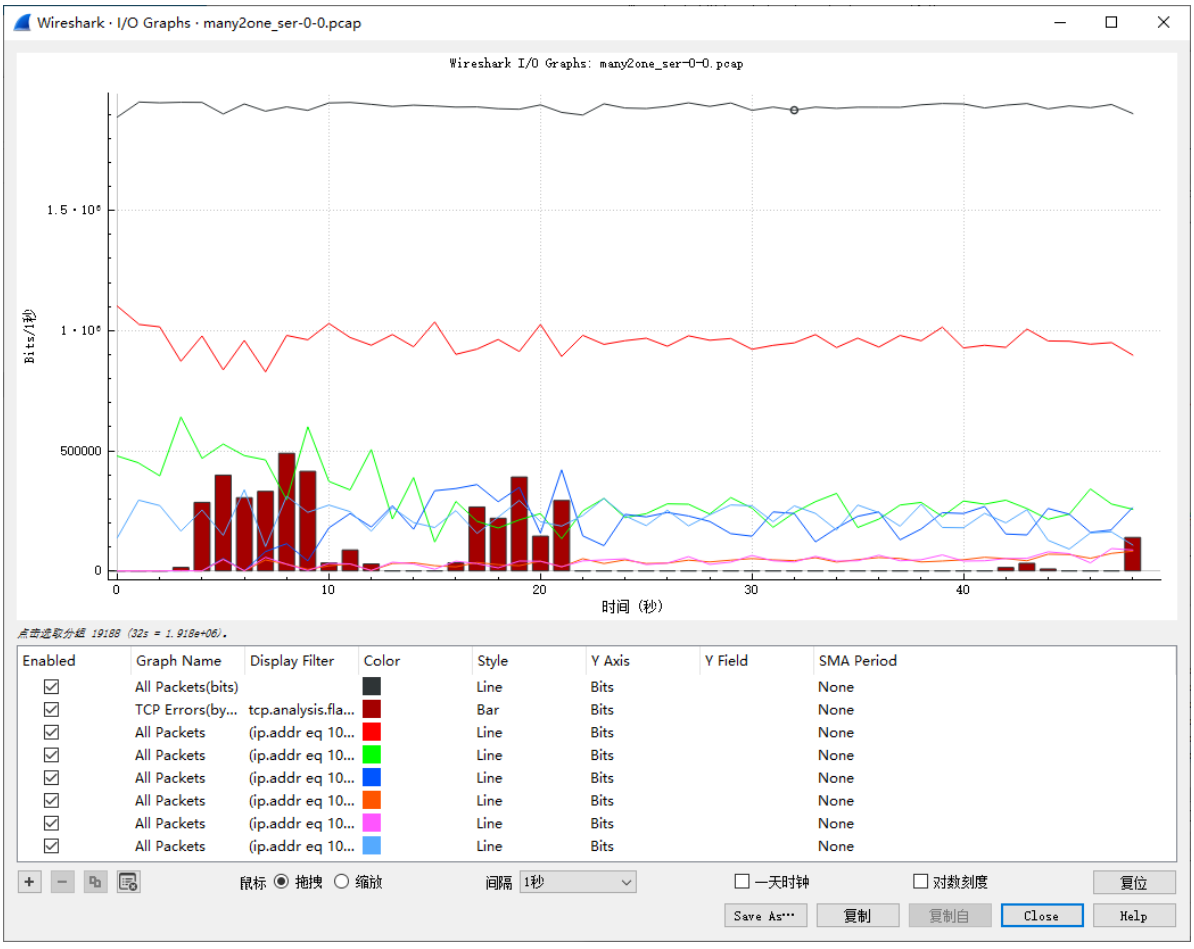
n5-n8会在a2所在的10.2.1.0/24网段竞争1M带宽，此时平均带宽0.25Mbps。n5-n8数据流进入左侧port后，会同n3、n4在10.1.1.0/24网段竞争1M带宽，此时n3-n8的平均带宽0.167Mbps。n3-n8数据流进入n1所在网段后会与n2竞争1M带宽，此时n3-n8的数据率减半到0.083Mbps，n2带宽0.5Mbps，与实验结果相符。

性能瓶颈：**t1、n1、n2所在的10.0.1.0/24网段带宽接近1M，达到瓶颈，其余各网段均未到达理论上的最大带宽。**

改进方法

- 增加链路带宽，将n1所在链路带宽从1Mbps增加到2Mbps

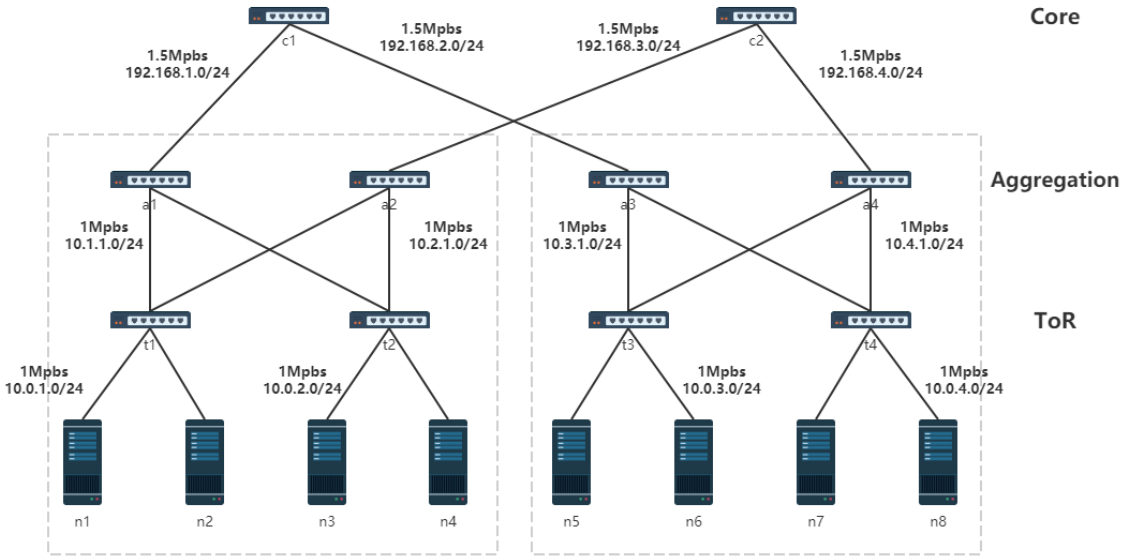




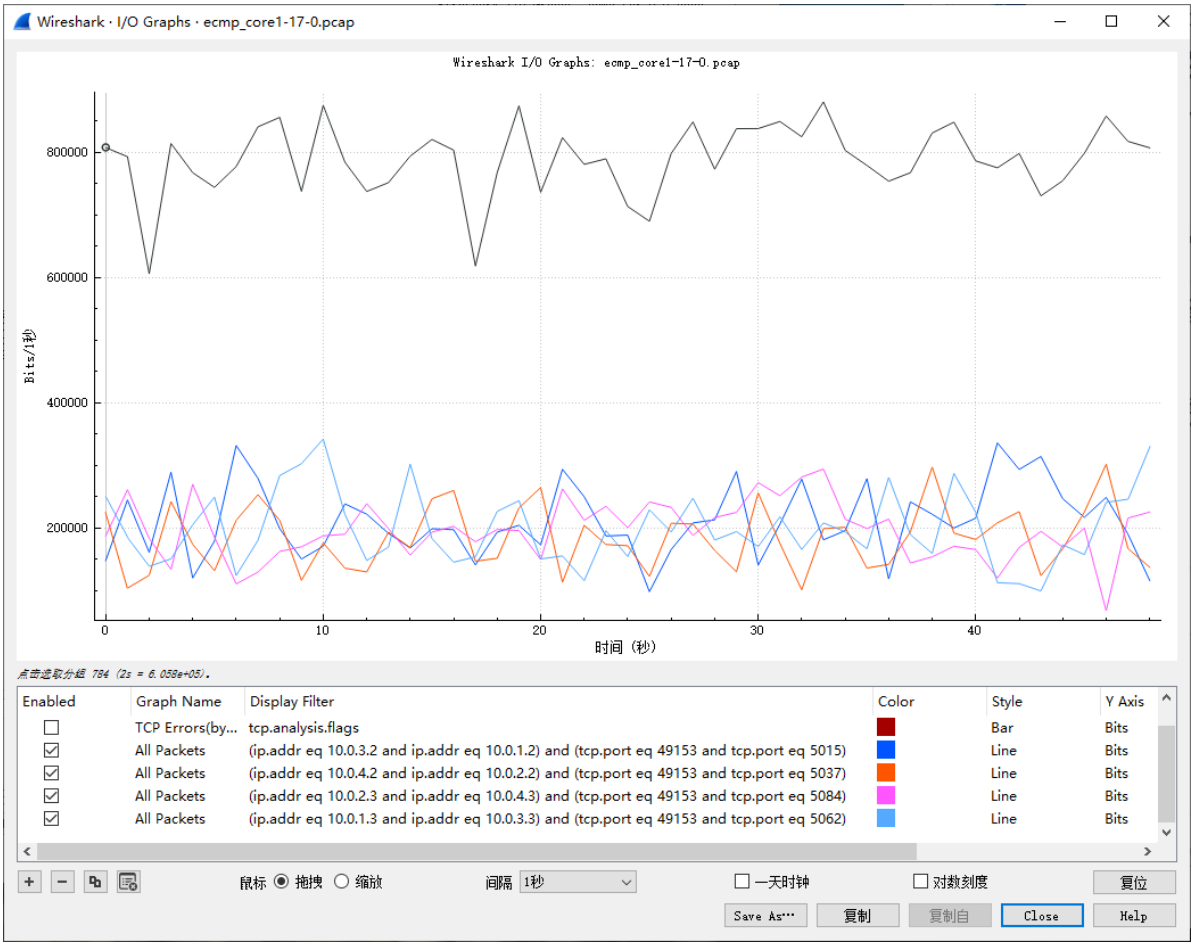
可以看到，n1服务器上的数据吞吐率增加将近一倍

实验3：性能改进实验

增加一个core switch和两个aggregation switch，拓扑如下



enable random ECMP后，进行inter cluster traffic实验，c1，c2的吞吐量如下图



平均**每个**核心交换机上都有0.8Mbps的流量，相比单核心交换机将近1Mbps的流量，**性能提升约60%。**