什么是微突发？如何定位微突发？

文档版本 03

发布日期 2020-07-20

华为技术有限公司

商标声明


和其他华为商标均为华为技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址： https://e.huawei.com

目 录

1 什么是微突发？

微突发（Microburst）是指端口在非常短的时间（毫秒级别）内收到非常多的突发数据，典型的微突发的持续时间通常在1～100毫秒之间，以至于瞬时突发速率达到平均速率的数十倍、数百倍，甚至超过端口带宽的现象。

网管或网络性能监测软件通常是基于比较长的时间（数秒到数分钟）计算网络实时带宽。在这种情况下，看到流量速率通常是一条比较平稳的曲线（如图1-1所示），没有任何的网络异常。但是，一秒钟对于一个高速收发数据包的接口来说是非常长的一个时间段。如果将数据更改为更细粒度（例如毫秒级）进行观察，实际流量中会看到更多突发，这些微突发非常纤细，流量速率很可能是带锯齿的（如图1-2所示）。如果锯齿突变很大，就称为微突发。

图 1-1 宏观流量速率

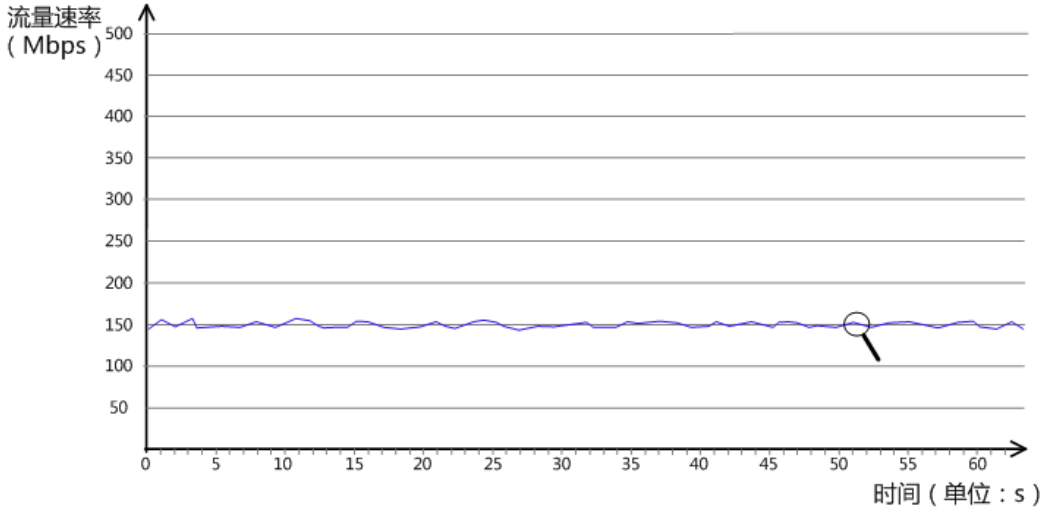
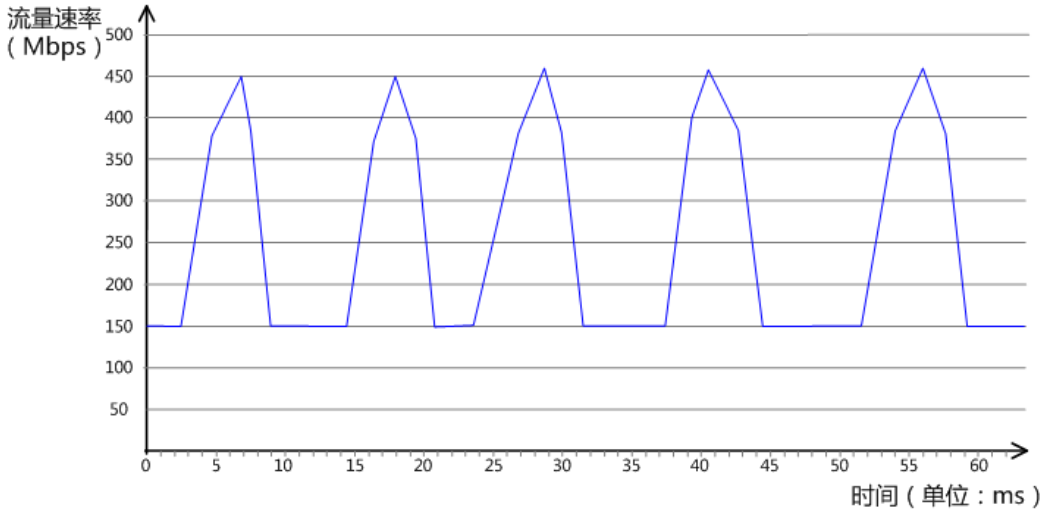


图 1-2 微观流量速率

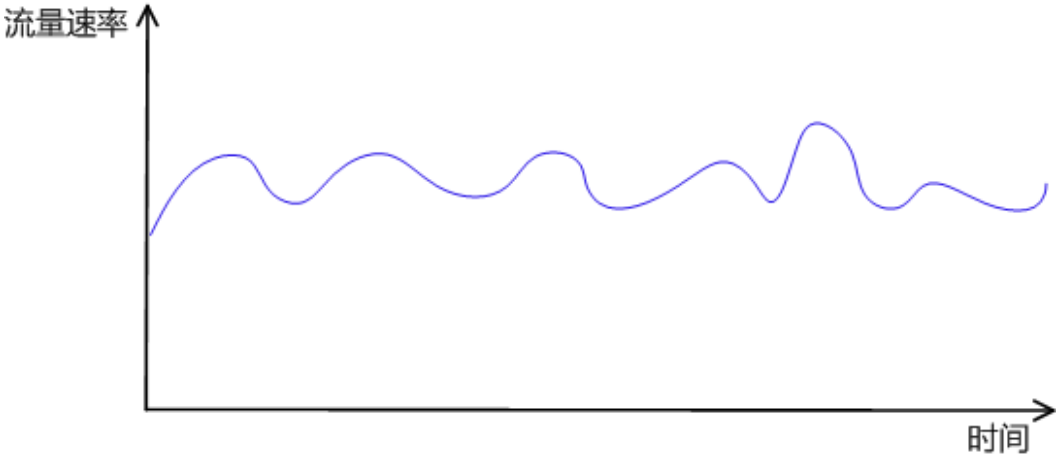


2 微突发产生的原因

网络中产生微突发，主要有如下原因：

1. 业务流量存在波动。很多的业务模型下，用户的请求和服务器的响应是离散出现的，导致业务流量是间歇性的，不稳定。同时对时延和带宽敏感的业务要求尽快发送数据，加剧业务的突发性。

   图 2-1 业务流量存在波动

2. 流量的入端口总带宽超过出端口的总带宽。例如：大带宽的端口向低带宽的出端口发送流量、相同速率的多个入端口向一个出端口发送流量。

图 2-2 大带宽的端口向低带宽的出端口发送流量

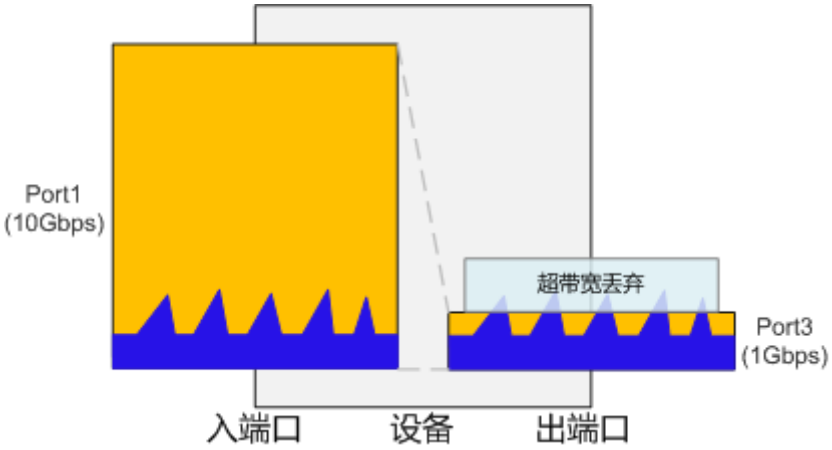
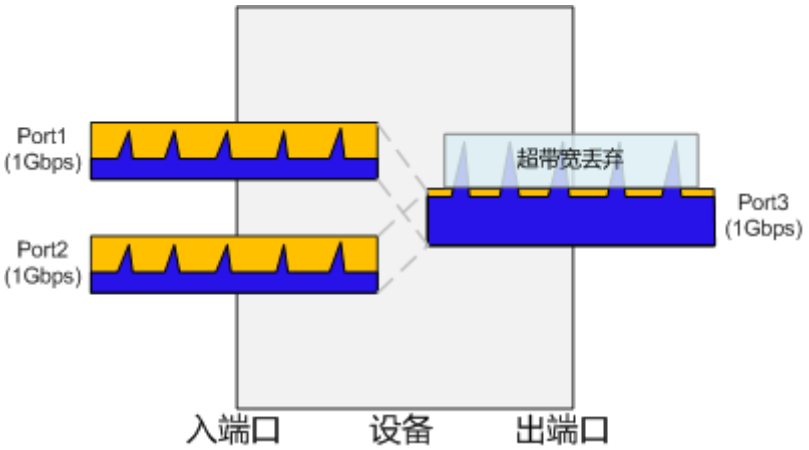
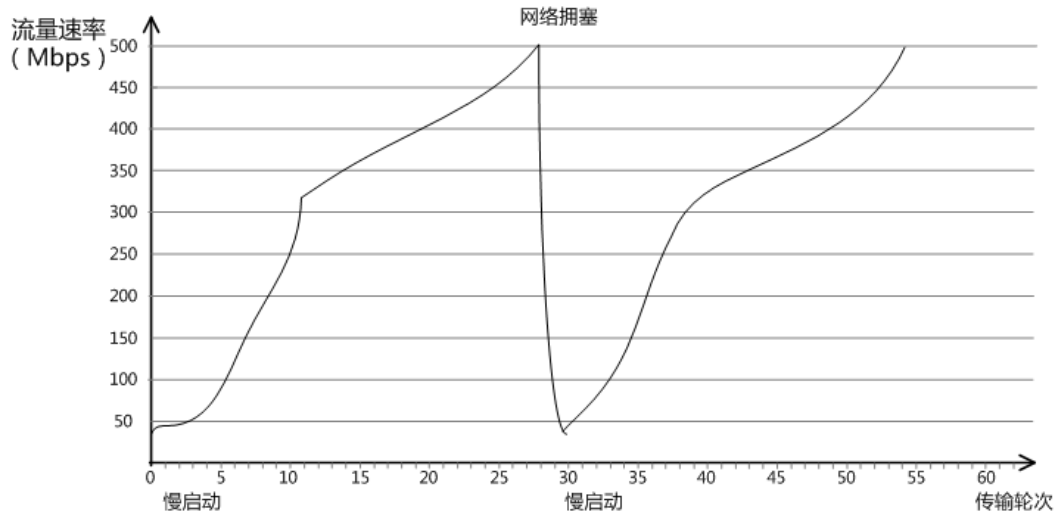


图 2-3 相同速率的多个入端口向一个出端口发送流量



3. 传统的TCP发包原则：通过慢启动和拥塞避免机制，尽快将数据包发送出去。慢启动使得发送速率不会快速上升。当吞吐量达到上限后，TCP滑动窗口减半，速率迅速下降，导致会话流量呈锯齿状，具有突发性。TCP总是期望把发送窗口中的数据尽快发送完，所以会在等待TCP的报文到达确认（ACK）到来后，通过滑动窗口机制再继续发送数据，如此循环，使得发包速率不平缓，突发性强。

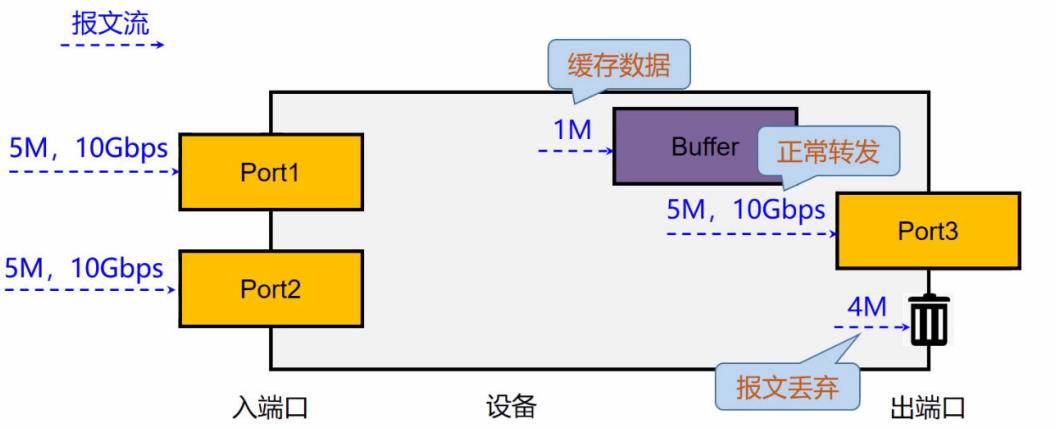
图 2-4 拥塞避免机制下，传统的 TCP 会话流量曲线呈锯齿状



3 微突发的影响及产生过程

当微突发流量的瞬时速率超过交换机的转发能力时，交换机会将突发的数据进行缓存以便稍后发送。如果交换机没有足够的缓存，那么超出的数据只能丢弃，这就产生了拥塞丢包。

如下是一个典型的毫秒级微突发场景。假设Port1、Port2都以10Gbps的线速速率分别向Port3发送5MB的数据，则总发送速率为20Gbps。而Port3的速率为10Gbps，仅为总发送速率的一半，因此只能将一半的数据（5MB）发送出去，另一半数据（5MB）则需要先缓存起来，待Port3有空闲能力时再发送。这时，由于交换机只有1MB的缓存，因此会有4MB的数据由于缓存不足而丢弃。在不考虑帧间隙、前导码、帧校验和、报文头等开销数据的情况下，这个突发持续的时间为5MB/10Gbps = 4ms。



4 如何评估交换机的抗突发能力

RFC4445中定义了衡量数据流的传输质量的关键指标——Delay Factor（简称DF，延迟因子）。DF，表明业务流量的延迟和抖动状况，业务流量抖动越大，DF值越大。当网络产生微突发时，DF可以被测量出来，单位是毫秒。

对于交换机，可以将DF换算为对缓存的需求。具体换算公式为：

$$T = \frac{\left(\mathrm{DF} \times W \times U - \left(\left(\mathrm{DF} \times W \times U\right)/A\right) \times (B - C)\right)}{8}$$

式中：

T—交换机缓存要求

DF—业务流量的延迟和抖动状况

W—出端口带宽

U—出端口带宽利用率

A—突发服务器接口物理带宽之和

B—出接口速率

C—未突发服务器速率和

一般交换机固有的缓存越大，通过如上公式计算出的缓存要求越低，抗突发能力就越强。

5 为什么不能通过网管或者设备观察端口来监控微突发？

客户习惯通过网管或者设备观察端口来监控监控端口流量或者出接口的报文最大速率。但是，这种方法并不能用来监控微突发。

通过网管监控端口流量

网管7*24小时不停地监控端口流量，通过设备的流量曲线判断网络中流量的走势，当网络产生微突发现象（即丢包）时，网管上展现的流量曲线都相对平稳。因为网管主要通过SNMP get方式来获取设备数据，存在如下不足：

- 管理规模有限：通过拉模式来获取设备的监控数据，不能监控大量网络节点，限制了网管的监控规模。
- 监控周期仅能到秒级：网管监控流量的周期通常是30s，最小一般可以设置为5s，部分第三方网管能设置成1s，一般都认为是秒级的精度。而且，网管监控的流量图的精度主要依赖于被管理设备的上报的数据精度，即使网管能计算出毫秒级别精度的速率，但是一般设备上报的数据只有秒级，流量图的精度还是秒级。

那设备为什么不实现毫秒级统计呢？主要因为：交换机端口众多，当设备统计端口报文时，CPU会遍历所有端口，获取端口报文统计性能比较低，无法在微突发的时间段遍历获取所有端口流量信息；而且毫秒级轮询所有端口对CPU性能的消耗非常大，可能会影响到交换机的正常业务，因此设备自身无法捕获微突发信息。

通过设备观察端口的出接口报文的最大速率

当网络产生微突发现象（即丢包）时，设备观察端口的出接口报文的最大速率，远远达不到端口的速率，端口利用率甚至都不到10%。

因为设备转发芯片只记录收到的总报文数，所以要想计算交换机上任何报文速率，都是只能通过总报文数量除以统计时间，所以最终统计精度还是取决于统计周期的长短。对于华为CloudEngine交换机而言，无论峰值速率和瞬时速率，默认都是按照300s的周期计算的平均值，其中峰值速率统计的周期是可配置的，最小为10s，也就是说精度只能达到10s。

6 如何定位微突发？

当出现如下任一现象时，可以认为出现微突发：

- 设备的出接口出现discard计数。
- 端口存在丢包告警：QOS_1.3.6.1.4.1.2011.5.25.32.4.1.11.51 hwXQoSPacketsDropInterfaceAlarm。
- 转发引擎检测到丢包日志：QOS/6/QOS_PACKET_DROP。

当前可以通过如下方式定位微突发：

- Telemetry：判断流量发生微突发主要通过缓存占用和接口毫秒级微观速率综合判断。虽然当前接口速率最小精度只能达到秒级，但接口队列的缓存占用信息的采集精度可以达到100ms，但是对于10ms甚至是1ms的微突发仍然无法监控。日常监控微突发时，推荐使用Telemetry。
- 第三方捕获报文和分析工具：只能对速率为10Gbps以内的报文进行监控。当报文速率超过10Gbps时，将无法监控。长时间使用这种方式来监控微突发，会影响分析工具所在PC/服务器的性能及容量。所以一般在定位微突发问题时，推荐使用第三方捕获报文和分析工具。
- 设备的丢弃报文捕获功能：只能捕获已知单播报文。因为设备将捕获到的丢弃报文上送CPU处理，上送CPU的速率最大为1000pps，所以该功能适用于存在少量丢包的网络，对于存在大量丢包的网络，该功能无法捕获到芯片所有丢弃的报文。所以一般在定位微突发问题时，推荐使用设备的丢弃报文捕获功能。

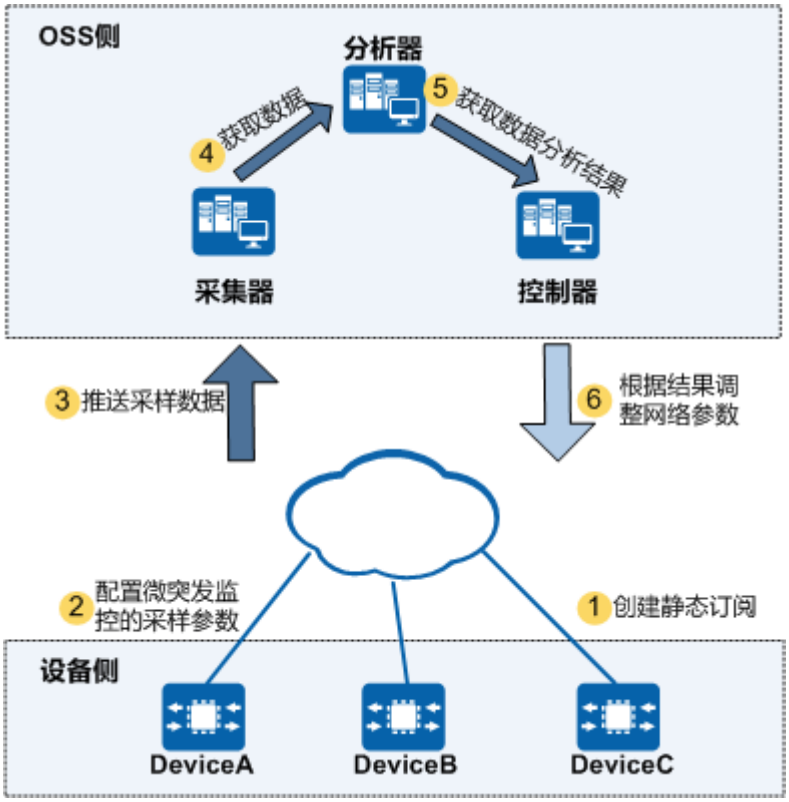
下面将介绍具体的操作方法：

通过 Telemetry 监控微突发

Telemetry是一个闭环的自动化运维系统，分为OSS侧和设备侧，由网络设备、采集器、分析器和控制器等部件组成。其中，网络设备、采集器、分析器、控制器即可以使用第三方的系统，也可以使用华为的系统。华为Telemetry系统中，网络设备对应的产品为：CloudEngine交换机，采集器、分析器对应的产品为：iMaster NCE-FabricInsight，控制器对应的产品为：iMaster NCE-Fabric。

通过Telemetry可以监控网络中是否存在微突发。Telemetry的系统架构和处理流程如图6-1所示。

图 6-1 Telemetry 系统架构及处理流程



如下将详细介绍华为Telemetry系统的详细操作及处理过程：

步骤1 在设备上，创建基于微突发监控的Telemetry静态订阅。

1. 创建采样数据目标采集器所在的目标组，主要包含：目标采集器的IP地址、端口号、推送协议和加密方式。

   system-view
   telemetry//进入Telemetry视图。
   destination-group *destination-group-name*//创建采样数据目标采集器所在的目标组，并进入Destination-group视图。
   ipv4-address *ip-address* **port** *port* [**vpn-instance** *vpn-instance*] [**protocol grpc** [**no-tls**]]//配置目标采集器的IP地址、端口号、推送协议和加密方式。
   commit

2. 配置采样传感器组名称，包括：采样路径和过滤条件。当设置的采样路径的满足过滤条件时，设备会及时上报给采集器。

   如下以配置自定义事件的采样路径和过滤条件为例。

   a. 创建采样传感器组名称。

      system-view
      telemetry//进入Telemetry视图。
      sensor-group *sensor-name*//创建采样传感器组名称，并进入Sensor-group视图。

   b. 配置自定义事件的采样路径，并进入自定义事件视图。

      CloudEngine交换机当前支持的采样路径为队列的缓存占用信息：huawei-qos:qos/qosQueueBufUsageStats/qosQueueBufUsageStat（最小精度为100ms）。

      sensor-path *huawei-qos:qos/qosQueueBufUsageStats/qosQueueBufUsageStat* **self-defined-event**

   c. 创建采样路径的条件过滤器。

      filter *filter-name*//创建采样路径的条件过滤器名称，并进入过滤器视图。
      op-field *field* **op-type** { **eq** | **gt** | **ge** | **lt** | **le** } **op-value** *value*//配置过滤条件。

condition-relation { **and** | **or** }//配置多个过滤条件间的逻辑运算关系。
commit

3. 创建静态订阅。

   该订阅用于关联目标采集器所在的目标组和采样传感器组，并配置gRPC传输协议的相关信息。

   a. 创建订阅。
   system-view
   telemetry//进入Telemetry视图。
   subscription *subscription-name*//创建订阅用于关联目标采集器所在的目标组和采样传感器组，并进入Subscription视图。
   destination-group *destination-name*//关联目标采集器所在的目标组。
   sensor-group *sensor-name*//关联采样传感器组。

   b. （可选）配置gRPC传输协议的相关信息。

   如下配置均有默认值，默认内容已经满足要求，可以跳过此步。
   protocol gRPC [**no-tls**]//配置gRPC传输协议的加密方式。缺省情况下，没有配置gRPC传输协议的加密方式。
   local-source-address ipv4 *ip-address*//配置上送报文的源IP地址。缺省情况下，配置上送报文的源IP地址为Socket选取路由出接口的IP地址。
   dscp *value*//配置上送数据报文的DSCP值。缺省情况下，配置上送报文的DSCP值为0。
   encoding { **json** | **gpb** }//配置上送数据报文的编码格式。缺省情况下，上送数据报文的编码格式为GPB。

   c. （可选）配置Telemetry采样数据时可占用主控板CPU的最大占用率。

   缺省情况下，配置Telemetry采样数据时可占用主控板CPU的最大占用率为5%。
   cpu-usage max-percent *usage*

   d. 提交配置。
   commit

步骤2 在设备上，配置微突发监控的采样参数，主要包括：缓存低门限、缓存高门限、采样周期。

1. 配置缓存低门限和缓存高门限。
   system-view
   interface *interface-type interface-number*
   qos [**queue** *queue-index*] **buffer-monitoring percent low** *low-percent* **high** *high-percent*//配置队列的缓存低门限和缓存高门限。
   quit//退出接口视图。

   当满足如下任意条件时，设备认为出现微突发流量，记录相应的微突发流量信息，包括：微突发流量发生的时间以及发生时对应队列的缓存使用率。
   - 在读取到的队列缓存使用率高于缓存高门限或缓存低门限值配置为0场景下：
     - 读取到的队列缓存使用率与上一次记录的队列缓存使用率的变化率大于2%。
     - 微突发流量发生时间与上一次记录的微突发流量发生时间相比超过1s。
   - 在读取到的队列缓存使用率介于高低门限之间场景下：
     - 读取到的队列缓存使用率与上一次记录的队列缓存使用率的变化率大于2%，且上一次出现微突发流量。
     - 微突发流量发生时间与上一次记录的微突发流量发生时间相比超过1s，且上一次出现微突发流量。

2. 配置采样周期。
   telemetry//进入Telemetry视图。
   subscription *subscription-name*//进入已经创建的订阅，并进入subscription视图。
   sensor-group *sensor-name* **sample-interval** *sample-interval*//配置关联传感器组的采样周期。
   commit

步骤3 当Telemetry数据采样周期到达后，设备会将此采样周期内记录的微突发流量信息按照Telemetry数据格式上送给采集器。

频繁读取端口队列缓存信息对设备的CPU使用率会产生影响。此时设备会根据CPU的使用率来调整微突发监控的采样周期：

- 当CPU的使用率低于或等于60%时，设备按照配置的采样周期读取端口队列的缓存使用率。
- 当CPU的使用率高于或等于80%时，设备停止读取端口队列的缓存使用率。
- 当CPU的使用率高于60%且低于80%时，设备按照10:1的采样比例读取端口队列的缓存使用率，即每10个采样周期读取1次端口队列的缓存使用率。

步骤4 分析器读取采集器（华为iMaster NCE-FabricInsight）收到设备上送的数据后，对数据进行分析。

1. 在设备侧完成iMaster NCE-FabricInsight与设备的对接配置，主要包括如下：
   a. 配置iMaster NCE-FabricInsight采集器集群接入数据中心网络的路由。
   b. 在设备侧配置SNMPv3协议。配置SNMP协议的目的是将设备添加到iMaster NCE-FabricInsight进行正常管理。

```
system-view
snmp-agent sys-info version v3
snmp-agent mib-view included iso-view iso  //iso-view为设置的MIB视图名称。为了保证iMaster NCE-FabricInsight能正常管理设备，MIB视图需要包含iso节点。
snmp-agent group v3 snmpv3group privacy write-view iso-view notify-view iso-view //snmpv3group为设置的用户组；指定写视图名和通知视图名为iso-view。写视图缺省具有读权限，因此不需要设置read-view；通知视图用于限制设备向iMaster NCE-FabricInsight发送告警的MIB节点。
snmp-agent usm-user v3 snmpv3user group snmpv3group //snmpv3user为设置的用户名，与iMaster NCE-FabricInsight的安全名保持一致。用户的安全级别不能低于所在用户组的安全级别，否则无法正常通信。例如设置用户组snmpv3group安全级别为privacy，则所属用户snmpv3user的安全级别必须为认证且加密。
snmp-agent usm-user v3 snmpv3user authentication-mode sha2_512 //设置用户的认证协议和认证密码，与iMaster NCE-FabricInsight的鉴权协议和认证密码保持一致。sha2_512为认证协议，8937561bc为根据设备提示需要输入的认证密码。
Please configure the authentication password (8-255)
Enter Password:
Confirm Password:
snmp-agent usm-user v3 snmpv3user privacy-mode aes256 //设置用户的加密协议和加密密码，与iMaster NCE-FabricInsight的私有协议和加密密码保持一致。aes256为加密协议，68283asd为根据设备提示需要输入的加密密码。
Please configure the privacy password (8-255)
Enter Password:
Confirm Password:
snmp-agent trap enable
snmp-agent target-host trap address udp-domain destip-address source loopback 0 udp-port 10162 params securityname snmpv3user v3 privacy //destip-address为iMaster NCE-FabricInsight数据采集器的IP地址（部署了分析器和采集器的场景下，使用采集器浮动IP地址；仅部署了分析器的场景下，使用分析器南向浮动IP地址）；loopback 0为Trap报文源地址所在接口；10162为Trap报文的端口号，不建议修改为其他值，如果需要修改请寻求技术支持；securityname设置为用户名。
snmp-agent packet max-size 12000  //设置SNMP Agent能接收和发送的SNMP报文的最大尺寸为12000字节；缺省情况下，SNMP报文的最大尺寸为12000字节，为了避免该参数值可能被修改，因此重置该参数值。
commit
```

   c. 配置Syslog协议。目的是将设备的异常日志信息输出到iMaster NCE-FabricInsight进行分析和可视化呈现，并应用于问题分析和辅助排障。

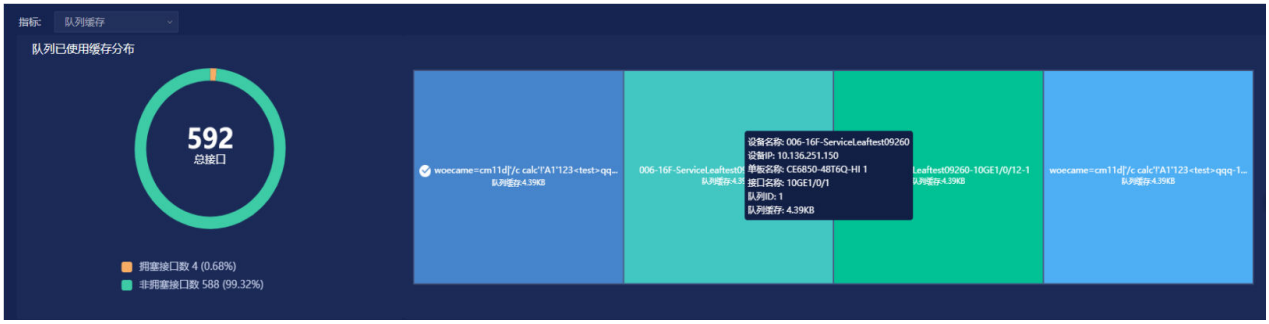
```
system-view
info-center enable //使能信息中心。
info-center channel channel-number name fabricSyslog //配置信息通道，选用的通道号为6、7或8，fabricSyslog为自定义的通道名。
info-center loghost destip-address source-ip source-ip-address { public-net | vpn-instance vpn-instance-name } port 28305 channel fabricSyslog level debugging transport tcp ssl-policy policy1 //配置向iMaster NCE-FabricInsight输出debugging级别的日志信息。
```



```
undo info-center statistic-suppress enable //关闭重复日志抑制。
commit
```

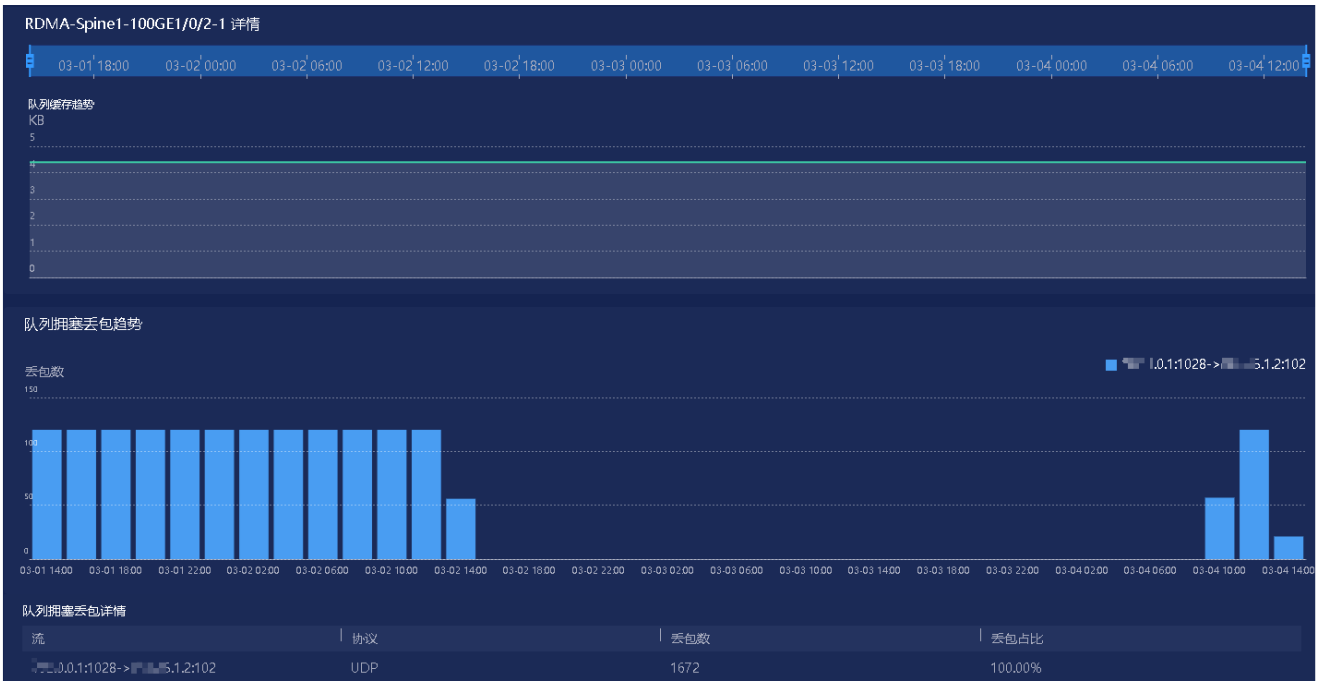
2. 在iMaster NCE-FabricInsight侧完成iMaster NCE-FabricInsight与设备的对接配置，即添加设备到iMaster NCE-FabricInsight。具体操作如下：
   a. 设置SNMPv3协议模板。
      i. 在导航树中选择“资源”，单击“协议模板 > SNMP”。
      ii. 单击“创建”，根据设备侧的SNMP协议信息配置模板中SNMPv3的参数。iMaster NCE-FabricInsight侧参数需要与设备侧SNMP参数保持一致。
      iii. 单击“确定”，完成SNMP协议参数模板创建。
   b. 将设备添加到iMaster NCE-FabricInsight。
      i. 在导航树中选择“资源”，单击“资源 > 设备”。
      ii. 单击“增加设备”。
      iii. 在“基本信息”中，输入设备的“IP地址”和“名称”，并选择对应的“Fabric”和“设备角色”。
      iv. 在“SNMP协议”中，设置SNMP协议参数。
      v. 单击“确定”，完成单个设备添加。
3. 通过iMaster NCE-FabricInsight的如下界面，可以查看到队列性能指标信息。
   a. 在导航树中选择“Telemetry”，并选择“队列”页签。查看队列已使用缓存分布。

图 6-2 队列已使用缓存分布



   b. 查看队列详情，主要包括：队列缓存趋势、队列拥塞丢包趋势、队列拥塞丢包详情。

图 6-3 队列详情



4. 当有微突发流量信息产生时，iMaster NCE-FabricInsight将会根据异常事件，识别为健康度问题。

   a. 在导航树中选择“健康度”，选择“问题”页签。

   b. 当有微突发流量信息产生时，可以通过“交换机端口拥塞导致业务受损”问题，查看队列缓存的丢包趋势和丢包详情，即不同时间点的丢包数量等信息。

步骤5 依据分析结果，通过控制器（华为iMaster NCE-Fabric）对网络进行调整。

1. 在设备侧完成华为iMaster NCE-Fabric与设备的对接配置，主要包括如下：

   a. 配置iMaster NCE-Fabric接入数据中心网络的路由。

   b. 在设备侧配置SNMPv3协议。配置SNMP协议的目的是将设备添加到iMaster NCE-Fabric进行正常管理。具体配置请参考**步骤4.1.b**。

2. 在iMaster NCE-Fabric侧完成iMaster NCE-Fabric与设备的对接配置，即添加设备到iMaster NCE-Fabric。具体操作如下：

   a. 进入“自动发现”页面。

      从菜单进入，在“网络初始化”APP的菜单中选择“物理资源 > 设备 > 设备管理”，进入“设备管理”页面。在“设备发现”的下拉框中，单击“自动发现”。

   b. 配置SNMPV3协议参数。

      选择“自定义”，在页面中填写SNMPV3协议参数。

   c. 单击“开始”，控制器开始发现设备。

      发现成功的设备将会呈现在“扫描成功列表”中。

3. 在华为iMaster NCE-Fabric“故障闭环”APP的菜单中选择“事件管理 ”。

4. 在故障事件页签中，查看“交换机端口拥塞导致业务受损”事件，并根据界面中给出的参考建议处理故障。

图 6-4 “交换机端口拥塞导致业务受损”事件处理建议

----结束

通过报文捕获及流量分析工具监控微突发

通过借助报文捕获及流量分析工具（如Wireshark软件）来监控网络中是否存在微突发。具体操作步骤如下：

步骤1 使用报文捕获功能来捕获报文，并生成报文文件。

1. 在设备出现丢弃报文计数的端口下，配置出方向镜像。
2. 将流量镜像到同等速率的观察口。
3. 在PC或服务器上，使用报文捕获及流量分析工具捕获观察口的报文，并生成报文文件。

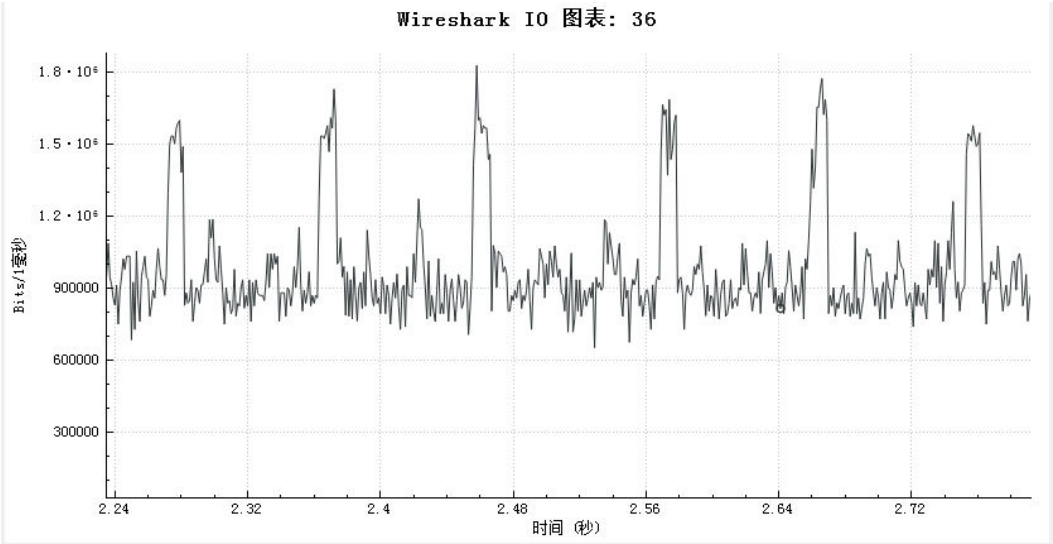
   说明

   受性能影响，PC最多只能捕获1Gbps的接口速率的报文，服务器最多只能捕获10Gbps的接口速率的报文。推荐使用高性能PC，否则可能会影响报文捕获。

步骤2 使用流量分析功能监控微突发。

1. 使用报文捕获及流量分析工具，打开捕获到的报文文件。
2. 在“I/O图表”中，可以看到流量图。
3. 将“IO图表”中的流量单位修改为“Bits”，同时将时间间隔修改为“1毫秒”，这样就能看到毫秒级流量的突发。

图 6-5 Wireshark IO 图表



----结束

通过设备的丢弃报文捕获功能

通过设备的丢弃报文捕获功能来定位微突发。具体操作步骤如下：

步骤1 在发生微突发的设备的端口的出方向上使能丢弃报文捕获功能。

```
system-view
qos capture drop-packet enable tcb egress
```

步骤2 查看捕获到的丢弃报文的信息。

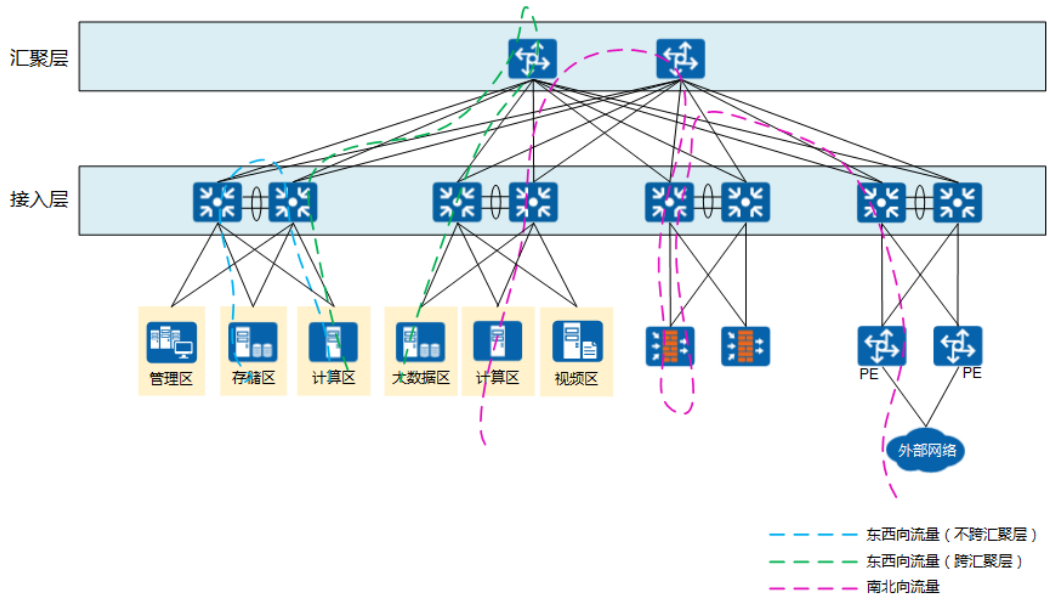
```
display qos capture statistics drop-packet { ip | ipv6 | ethernet } [ tcb ] slot slot-id
```

----结束

7 如何减小数据中心网络中的微突发？

数据中心网络中，运行的业务主要有管理、存储、大数据、计算、视频等业务，典型组网如图7-1所示。

图 7-1 数据中心网络典型组网及流量



- 管理业务：业务流量较小，发生丢包时对业务基本上不会产生影响。
- 存储业务：存储后端，主要使用FC网络，不涉及数据中心网络的优化范围。存储前端，通过以太网访问计算节点，业务流量大，一旦发生丢包，对业务影响比较大。
- 大数据业务：大数据Hadoop等集群内业务突发严重，数据业务本身业务对丢包有一定容忍度，但是一旦集群心跳丢失会导致集群分裂，对业务影响比较大。
- 视频业务：业务流量会存在一定的抖动，一旦发生丢包影响用户业务体验，容易造成业务部门投诉，风险高。

流量分析

1. 接入层

a. 南北向流量：例如每台设备上行2个40GE接口接入汇聚层，下行48个10GE接口接入服务器等设备，则收敛比为6:1，微突发风险较高。

b. 东西向流量：存储、计算集群、Hadoop集群，存在多个服务器同时访问一个服务器的情况，微突发风险较高。

2. 汇聚层

a. 南北向流量：从北向南流量，公网流量小于接入流量，不存在微突发拥塞风险；从南向北流量，分为过墙和不过墙两种场景。南到北流量过墙场景，存在收敛，但是一般数据中心中南北流量仅占总体流量的10%~25%，风险较低，且交换机的转发能力远强于防火墙，瓶颈点在于防火墙。南到北流量不过墙场景与过墙场景类似。

b. 东西向流量：东西向分配的带宽相同，收敛比基本为1：1，突发风险低。

优化措施

最根本的优化措施主要通过优化服务器的流量来减少微突发。网络侧可以通过如下方法，降低微突发的发生几率，缓解微突发的影响，如下以华为CloudEngine交换机为例：

1. 设备开启流量整形功能：

   在延时可控和缓存充足的情况下，在发生拥塞的设备的上游交换机的下行接口通过**qos queue** *queue-index* **shaping** { **percent cir** *cir-percent-value* [**pir** *pir-percent-value*] | **cir** *cir-value* [**kbps** | **mbps** | **gbps**] [**cbs** *cbs-value* [**bytes** | **kbytes** | **mbytes**] | **pir** *pir-value* [**kbps** | **mbps** | **gbps**] [**cbs** *cbs-value* [**bytes** | **kbytes** | **mbytes**] **pbs** *pbs-value* [**bytes** | **kbytes** | **mbytes**]]] }命令开启流量整形功能，削弱流量的瞬时波峰，可以控制突发的程度。需要注意的是，此方法会导致报文转发时延加大。

2. 配置接口的缓存管理为增强模式：

   在发生拥塞的接口下，执行**qos burst-mode enhanced**命令配置接口下缓存管理的突发模式为增强模式，以缓解网络拥塞。

3. 接入层

a. 降低上行收敛比：将接入层设备替换为M-LAG组网，使M-LAG的主备设备上行时形成逐流负载分担，共同进行流量的转发，可以使上行收敛比由6:1降低为3:1。

b. 避免多台服务器访问一台服务器的场景：在网络业务流量规划时，尽量避免多台服务器访问一台服务器的场景，及时扩容突发严重的出端口，消除突发瓶颈。

c. 提高重点业务优先级：在接入大数据区的设备上，通过流分类**traffic classifier**识别业务影响比较大的Hadoop存储集群心跳报文，并通过流行为**traffic behavior**中的重标记内部优先级**remark local-precedence**将该报文的优先级提高，如设置为4，最后通过流策略**traffic policy**应用到接口，保障重点业务、减少微突发的影响。

4. 汇聚层

   固定各业务的优先级：在核心和汇聚的所有设备上，通过流分类**traffic classifier**区分不同业务，并通过流行为**traffic behavior**重标记报文的优先级（内部优先级**remark local-precedence**、重标记VLAN报文802.1p优先级**remark 8021p**、重标记IP报文的DSCP优先级**remark dhcp**），如即将管理业务的优先级设置为0、高性能计算业务的优先级设置为1、视频业务优先级设置为2、存储业务优先级设置为3，最后通过流策略**traffic policy**应用到接口。

8 当端口存在 DISCARD 丢弃，并且端口的带宽使用率较低时，如何判断是否存在微突发？

问题

当端口存在DISCARD丢弃，并且端口的带宽使用率较低时，如何判断是否存在微突发？

适用产品及版本

适用产品：CE6851HI、CE6855HI、CE6856HI、CE6860EI、CE6865EI

适用版本：V200R005C10及之后版本

回答

步骤1 加载V200R005C10SPH025及之后补丁。

步骤2 登录设备，进入用户视图。

步骤3 开启端口的微突发高精度检测功能，并且配置微突发监测时长。

high-precision rate detection interface *interface* **duration** *monitortime*

其中，*monitortime*为微突发监测时长（单位：分，范围1~1440），需要根据现网端口下的DISCARD计数产生频率来进行设置。

具体案例，如下：

```
<HUAWEI>high-precision rate detection interface 10GE 2/0/20 duration 10
Info: High-precision rate detection is enabled, duration 10 minutes.
```

步骤4 当端口下的DISCARD计数有增长，可以通过查看峰值速率，来判断是否存在微突发。

display high-precision rate detection interface *interface*

具体案例，如下：

```
<HUAWEI>display high-precision rate detection result interface 10GE 2/0/12
Interface                    : 10GE2/0/12
Configured detection duration : 10 Min
Detection start time         : 2020-09-28 10:24:19
Detection end time           : --
Average input rate           : 0 Mbits/sec
Average output rate          : 0 Mbits/sec
```

```
Top 10 input rate:
------------------------------------------------
Time                    Input  Rate(Mbits/sec)
------------------------------------------------
2020-09-28 10:24:38.912                      0
2020-09-28 10:24:38.915                      0
2020-09-28 10:24:38.909                      0
2020-09-28 10:24:38.903                      0
2020-09-28 10:24:38.900                      0
2020-09-28 10:24:38.893                      0
2020-09-28 10:24:38.896                      0
2020-09-28 10:24:38.884                      0
2020-09-28 10:24:38.887                      0
2020-09-28 10:24:38.881                      0
------------------------------------------------

Top 10 output rate:
------------------------------------------------
Time                    Output Rate(Mbits/sec) //主要关注“Output Rate”列中数据，当10GE端口的峰值速率达到10000Mbits/sec左右，说明已经达到端口能力上限，即存在微突发。
------------------------------------------------
2020-09-28 10:24:29.345                  10000
2020-09-28 10:24:21.799                  10000
2020-09-28 10:24:31.751                  10000
2020-09-28 10:24:38.648                  10000
2020-09-28 10:24:20.178                  10000
2020-09-28 10:24:24.954                  10000
2020-09-28 10:24:26.718                  10000
2020-09-28 10:24:22.270                  10000
2020-09-28 10:24:22.129                  10000
2020-09-28 10:24:20.767                  10000
------------------------------------------------
```

步骤5 关闭微突发端口高精度检测功能。

undo high-precision rate detection interface *interface*

具体案例，如下：

```
<HUAWEI>undo high-precision rate detection interface 10GE 2/0/12
Info: High-precision rate detection is disabled.
```

----结束

9 微突发的常见误解

误解1：为什么出现微突发时交换机没有上报缓存超限告警呢？

实际情况：这是因为交换机要获取缓存的使用情况，只能依赖CPU的轮询机制，这就面临和端口速率计算类似的问题，CPU不能轮询太频繁，否则CPU利用率会升高，从而导致交换机响应慢，甚至无响应。

误解2：端口当前的利用率不高，速率不大，因此当前的突发也很小。

实际情况：这是错误的理解。平均速率和突发速率的关系，就如同速度与加速度，虽然名字很相似，但并没有简单的线性比例关系。端口的利用率不高，当前速率低，并不表示突发速率就小。

误解3：交换机记录了拥塞丢包计数，所以是交换机引起了突发或拥塞丢包。

实际情况：这是错误的理解。突发流量是由业务终端产生的，除了少量协议报文外交换机不产生其他流量。但是，突发可能在交换机上加剧，例如多端口同时向单端口发送数据，收敛比不合理，导致突发的峰值叠加。所以要从流量来源和组网上着手，寻找突发的源头。

误解4：终端服务器的业务是很随机的，不会有多大的突发。

实际情况：服务器网卡在发送业务报文的时候，会以当前物理速率进行发送，10GE的端口就会按照 10Gbit/s 的速率向外发送，发送完后等待应用层的后续报文。这种原理决定了物理链路在传输报文时，一定是按照链路速率传输，工作一段时间，空闲一段时间。 整体平均后的带宽利用率可能不高，比如20~30% ， 但实际是以 100%~0%~100% 的实际带宽利用率在工作。这种流量在微观层面就是突发，服务器的发包都是按照100%的带宽利用率在发送。

