TCP Probe 说明

1 输出格式

输出文件: `/proc/net/tcpprobe_data`

输出的每一行的字段从左到右分别为 (所有字段均采用十六进制输出):

```
<type>, <timestamp sec>, <timestamp nsec>, <srcaddr> <srcport>, <dstaddr> <dstport>,
<length>, <tcp_flags>, <seq_num>, <ack_num>, <ca_state>, <snd_nxt>, <snd_una>,
<write_seq>, <wqueue>, <snd_cwnd>, <ssthreshold>, <snd_wnd>, <srtt>, <mdev>,
<rttvar>, <rto>, <packets_out>, <lost_out>, <sacked_out>, <retrans_out>, <retrans>,
<frto_counter>, <rto_num>, <user-agent>
```

各个字段的含义如表 1 所示。

2 内核模块参数

内核模块中主要的内核参数如下表所示。

内核模块参数配置方法:

1. 加载内核时配置: `insmod tcp_probe_plus.ko <arg 1>=<value 1> <arg 2>=<value 2> ...`
2. 通过 sysctl 接口配置: `sysctl -w net.tcpprobe_plus.<arg>=<value>`

参数	含义
`port`	要监听连接的 TCP 端口号 (源或目的, 默认值: 0) 0 表示监听所有端口
`full`	0: 只有当拥塞窗口变化时才记录, 1: 任何一个包到达时都记录 (默认值: 1)
`maxflows`	最多同时监听的流数目 (默认值: 1000)
`readnum`	从 proc 文件系统中一次性读取的数据量 (单位: item, 默认值: 10)
`bufsize`	内核模块中 Log 的缓存大小 (单位: item, 默认值: 4096)

3 统计信息

内核模块维护一些实时的统计信息, 这些统计信息可以在文件 `/proc/net/stat/tcpprobe_plus` 中看到:

```
centos@host:~$ cat /proc/net/stat/tcpprobe_plus
Flows: active 4 mem 0K
Hash: size 1000 mem 36K
cpu# hash_stat: <search_flows found new reset>, ack_drop: <purge_in_progress ring_full>, \
conn_drop: <maxflow_reached memory_alloc_failed>, err: <multiple_reader copy_failed>
Total: hash_stat: 0  25877    151    147, ack_drop: 0 0, conn_drop: 0 0, err: 0 0
```

各个统计值的含义如下所示：

- Flows
  - active: 正在监听的连接数
  - mem: Flow table 所占用的内存大小
- Hash
  - size: 哈希表中表项的个数
  - mem: 哈希表所占用的内存大小
- hash_stat
  - search_flows: 哈希表中被找到的流数
  - found: 哈希表中的流数
  - new: 新增加的流表数
  - reset: 因连接关闭而结束的流数
- ack_drop
  - purge_in_progress: 已经弃用，一般为 0
  - ring_full: 因为读取 `/proc/net/tcpprobe_data` 不及时造成数据丢失的数目
- conn_drop
  - maxlfow_reached: 因达到流数过多而不监听的连接的数目
  - memory_alloc_failed: 因分配内存失败而不监听的连接数目
- err
  - multiple_reader: 在写入 `/proc/net/tcpprobe_data` 时，文件正在被多个 reader 读取
  - copy_failed: 无法拷贝数据到用户态

表 1: 输出格式中各字段的含义

字段	含义
type	何时获得这一列数据： 0: 收到数据 1: 发送数据 2: RTO 超时 3: 连接建立 4: 连接关闭 5: 连接移除
timestamp sec	时间戳, 秒部分
timestamp nsec	时间戳，纳秒部分
srcaddr	源 IP 地址
srcport	源 TCP 端口号
dstaddr	目的 IP 地址
dstport	目的 TCP 端口号
length	捕获的包 payload 大小 (单位: Byte)
tcp_flags	TCP 包头中的标志位
seq_num	捕获的包的 tcp 序列号 (相对值)
ack_num	捕获的包的 tcp 确认号 (相对值)
ca_state	拥塞避免状态
snd_nxt	下一个待发数据包的序列号 (相对值)
snd_una	第一个尚未被确认的包的序列号 (相对值)
write_seq	发送缓存中的最后一段数据的位置 (相对值)
wqueue	发送缓存中数据量 (单位，Byte)
snd_cwnd	拥塞窗口大小 (单位：包)
ssthreshold	慢启动阈值 (单位：包)
snd_wnd	接受窗口大小 (单位：包)
srtt	内核估计的 rtt (单位: 8μs)
mdev	RTT 中等偏差 (单位：4μs)
rttvar	RTT 标准差 (单位：4μs)
rto	重传定时器的值 (单位：ms)
packets_out	发送出去的数据量 (单位：包)
lost_out	内核所估计的 (已发送的包中) 的丢包数
sacked_out	被 SACK 的包数
retrans_out	当前所重传的包数
retrans	总共的重传次数
frto_counter	是否发生虚假超时重传
rto_num	发生超时重传事件的次数
user-agent	HTTP 包头中的 User-Agent 字段（可能为空）