

Quality of Service

Tomasz Safuryn

Wirtualna Polska

1 czerwca 2007

tsafuryn@wp-sa.pl

Czym jest QoS?

- ▶ Zbiór mechanizmów, które pozwalają zapewnić określoną jakość transmisji dla różnych klas ruchu.

- ▶ Best Effort
- ▶ Integrated Services
- ▶ Differentiated Services

- ▶ Best Effort
- ▶ Integrated Services
- ▶ Differentiated Services

Pakiety dzielą się na lepsze i gorsze.

- ▶ Dostępność (availability)
- ▶ Pasma (bandwidth)
- ▶ Opóźnienia (delay, latency)
- ▶ Zmiany opóźnień (jitter)
- ▶ Straty pakietów (losses)

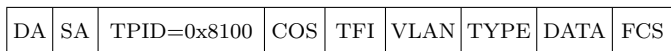
Tak to zazwyczaj wygląda



- ▶ Packet classification
- ▶ Packet marking
- ▶ Congestion management
- ▶ Congestion avoidance
- ▶ Traffic conditioning (shapers and policers)

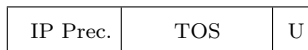
Oznaczanie pakietów

- ▶ 802.1Q (COS)



- ▶ IP (TOS)

- ▶ IP (DSCP)



QoS jest ugodywy

QoS pozwoli Ci się skonfigurować równie łatwo w dobry, jak i zły sposób.

- ▶ włącz QoS na wszystkich urządzeniach
- ▶ nie wszystko musi być priorytetowe
- ▶ pechowy RED

Trzeba dopasować parametry QoS do zarządzanej sieci i ruchu w niej. Czasem należy konkretne urządzenie skonfigurować nieco inaczej niż pozostałe.



Jak to się robi w Linuksie

- ▶ Kolejka (qdisc)
Kolejka jest miejscem gdzie przebywają pakiety przed wysłaniem przez interfejs, różne kolejki w różny sposób zarządzają znajdującymi się w nich pakietami.
- ▶ Klasa (class)
Klasa to wydzielona część dostępnego łącza charakteryzująca się określonymi parametrami.
- ▶ Filtr (filter)
Filtr pozwala skierować pakiety do klasy.

tc – traffic control

```
# tc help
```

```
# tc qdisc help
```

```
# tc qdisc show
```

```
# tc qdisc add dev eth0 root htb
```

```
# tc -s qdisc show dev eth0
```

```
# tc filter add dev eth0 help
```

```
# tc filter del dev eth0 priority 10
```

```
# tc f d dev eth1 p 10
```

- ▶ proste (PFIFO/BFIFO, PFIFO_FAST, RED, SFQ, TBF)
- ▶ złożone (CBQ, HTB, PRIO)

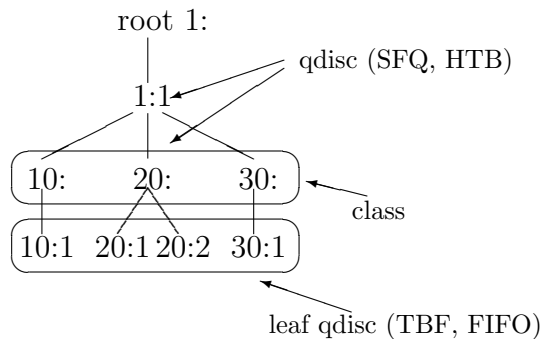
```
# tc qdisc show dev eth0
qdisc pfifo_fast 0: bands 3
    priomap  1 2 2 2 1 2 0 0 1 1 1 1 1 1 1
# tc qdisc add dev eth0 root handle 1: tbf \
    limit 100K burst 20K rate 1024kbit
# tc -s qdisc show dev eth0
qdisc tbf 1: dev eth0 rate 1024Kbit burst 20Kb \
    lat 640.0ms
Sent 0 bytes 0 pkt (dropped 0, overlimits 0 requeues 0)
rate 0bit 0pps backlog 0b 0p requeues 0
```

Klasy i kolejki złożone (klasowe)

- ▶ CBQ – kolejka klasowa

```
# tc qdisc add dev eth0 root handle 1: cbq \  
    bandwidth 1024kbit avpkt 1000 mpu 64  
# tc qdisc show dev eth0  
qdisc cbq 1: dev eth0 rate 1024Kbit \  
    (bounded,isolated) prio no-transmit  
  
# tc class add dev eth0 parent 1:0 classid 1:10 \  
    bounded cbq bandwidth 1024kbit rate 512kbit \  
    weight 5 prio 1 allot 1504b maxburst 10 avpkt 512  
# tc qdisc add dev eth0 parent 1:10 handle 2:1 sfq \  
    quantum 1500 perturb 5
```

Hierarchia



► iptables

```
# tc filter add dev eth0 parent 1: handle 10 \  
    protocol ip fw classid 1:10  
# tc filter add dev eth1 parent 1:0 handle 15 \  
    protocol ip priority 100 fw police rate 1Mbit  
    burst 1024  
  
# iptables -t mangle -A POSTROUTING -d 10.10.10.10 \  
    -j MARK --set-mark 10  
# iptables -t mangle -A POSTROUTING -d 10.10.10.15 \  
    -j MARK --set-mark 15
```

Oznaczanie i klasyfikowanie pakietów – iptables

- ▶ oznaczanie pakietów (zmiany w pakiecie)

```
iptables -t mangle -A PREROUTING -j DSCP --set-dscp 22
iptables -t mangle -A PREROUTING -j DSCP \
    --set-dscp-class EF
iptables -t mangle -A POSTROUTING -j TOS --set-tos 0x8
```

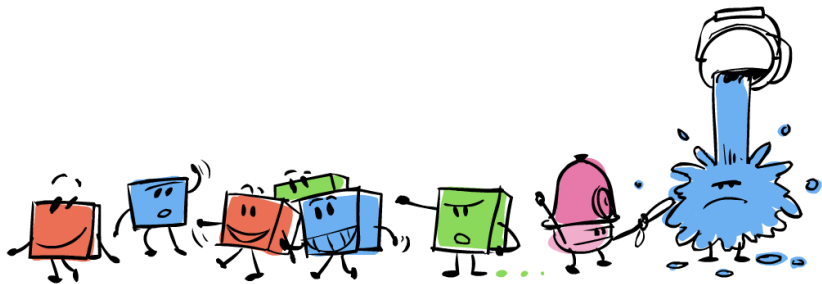
- ▶ klasyfikowanie pakietów

```
iptables -t mangle -A PREROUTING -j MARK --set-mark 1
```

- ▶ albo nawet tak

```
iptables -A INPUT -p dscp --dscp-value 22 -j DROP
```

Rozpoznać i pomalować



- ▶ adres źródłowy, docelowy
- ▶ ToS

```
ip rule add to 10.10.10.10/32 realms 10
```

```
ip rule add from 10.10.10.20/32 realms 20
```

```
ip rule add tos 0x10 realms 30
```

```
tc filter add dev eth0 parent 1:0 protocol ip \  
    prio 100 route to 10 flowid 1:1
```

► nagłówek pakietu

```
tc filter a dev eth0 parent 1: proto ip prio 1 u32 \  
    match u8 0x40 0xf0 at 0 \  
    match u16 0x0006 0x00ff at 8 \  
  
tc filter a dev eth0 parent 1:0 proto ip prio 2 u32 \  
    match ip protocol 0x06 0xff
```

FIFO – First In, First Out

- ▶ FIFO to najprostsza możliwa kolejka, pakiet wchodzący pierwszy do kolejki również pierwszy z niej wychodzi. Kolejka ma jeden parametr — długość. W Linuksie mamy dwa rodzaje takich kolejek `bfifo` i `pfifo`, długość pierwszej jest podawana w bajtach, drugiej w pakietach.
- ▶ `PFIFO_FAST` domyślna kolejka na interfejsach Linuksa, składa się z trzech równoległych kolejek FIFO o różnych priorytetach. Kolejka z niższym priorytetem nie będzie obsługiwana dopóki są pakiety w ważniejszych kolejkach.
- ▶ `PRIO` – kolejka priorytetowa bardzo podobna do poprzedniej, ale z konfigurowalną ilością podkolejek.

TBF – Token Bucket Filter

Kolejka TBF pozwala ograniczyć szybkość przepływu danych przez interfejs i zbuforować chwilowe przeciążenia.

- ▶ limit
- ▶ buffer/burst
- ▶ rate

```
# tc qdisc add dev eth0 root handle 1: tbf \  
    limit 100K buffer 20K rate 1024kbit
```

SFQ – Stochastic Fairness Queuing

SFQ – „sprawiedliwa” kolejka (fair queueing), pakiety widziane jako ciąg strumieni (flows).

- ▶ perturb
- ▶ quantum

```
# tc qdisc add dev eth0 root handle 1: sfq \  
    perturb 10 quantum 1514b
```

RED – Random Early Detection

RED – próbuje zapobiegać możliwym przeciążeniom.

- ▶ limit
- ▶ min
- ▶ max
- ▶ burst
- ▶ avpkt
- ▶ probability

```
tc qdisc add dev eth0 root handle 1: red \  
  limit 30K min 10K max 20K burst 20 \  
  avpkt 1000 probability 0.4
```

CBQ – Class Based Queuing

qdisc

- ▶ avpkt
- ▶ bandwidth

class

- ▶ rate
- ▶ avpkt

```
# tc qdisc add dev eth0 root handle 1: \  
    cbq bandwidth 100Mbit avpkt 1000  
# tc class add dev eth0 parent 1:0 classid 1:11 \  
    cbq prio 1 rate 1Mbit avpkt 1000
```

HTB – Hierarchical Token Bucket

qdisc

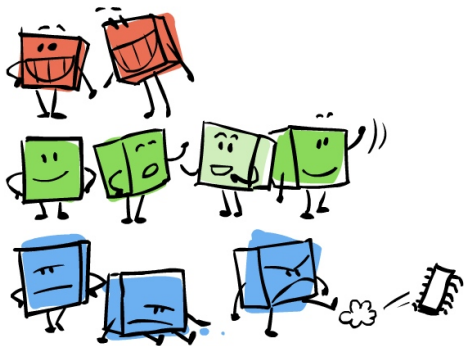
- ▶ default

class

- ▶ rate
- ▶ burst
- ▶ ceil

```
# tc qdisc add dev eth0 root htb default 10
# tc class add dev eth1 parent 1: classid 1:10 \
    htb rate 5mbit
```

Kolejki, kolejki...



Jak to złożyć razem?

```
# tc qdisc add dev eth0 root handle 1: htb default 20
# tc class add dev eth0 parent 1:1 classid 1:1 \
    htb rate 10mbit burst 50k

# tc class add dev eth0 parent 1:1 classid 1:10 \
    htb rate 7mbit ceil 10mbit burst 20k
# tc class add dev eth0 parent 1:1 classid 1:20 \
    htb rate 3mbit ceil 10mbit burst 20k

# TC="tc qdisc add dev eth0"
# $TC parent 1:10 handle 10: sfq perturb 15
# $TC parent 1:20 handle 20: sfq perturb 15

# tc filter add dev eth0 protocol ip parent 1:0 \
    prio 1 u32 match ip sport 25 0xffff flowid 1:20
```

- ▶ ipqosconf(1M)
- ▶ acctadm(1M)

- ▶ ALTQ
- ▶ Dummynet
- ▶ IPfw

