

Aplikacije/programi

Sysctl variable

r = receive (RX)
w = write (TX)

net.core.rmem_max
net.core.rmem_default

net.core.wmem_max
net.core.wmem_default

SKB (sk_buff)
Socket buffer

Linux
Socket

Mrežni stôg

UDP

TCP

Window scaling
Congestion Control
Timers
TCP queue & Limits

IP

RFS

GSO

RPS

GRO

Network scheduler
(qdisc)

(RX) Backlog buffers

TXqueue
(FIFO)
TXqueue len buffer

Upravljački program
(pr. e1000) [eth0]

NAPI

(Ovisno o upravljačkom programu)

Niži sloj upravljačkog programa

RX

TX

(FIFO) Ring buffers
Prstenaste memorije

TCP/UDP/IP
Checksum

LRO

TSO

RSS

aRFS

RX

TX

VLAN

Ethernet
Checksum

RX

TX

Skaliranje TCP prozora (uklj/isklj) je postavljeno u:
net.ipv4.tcp_window_scaling
Odabrani algoritam za nadzor zagušenja postavljen je u:
net.ipv4.tcp_congestion_control
Brojači (timeri) su definirani u:
FIN_WAIT_2 : **net.ipv4.tcp_fin_timeout**
TIME_WAIT= 2 x FIN_WAIT_2
Iskorištavanje veze u Time_wait stanju:
net.ipv4.tcp_tw_reuse
Ograničenje ukupnog broja TCP konekcija u stanju otvaranja: **net.core.somaxconn**
TCP Syn queue memorija
(max broj konekcija koje su u stanju otvaranja za pojedini port):
net.ipv4.tcp_max_syn_backlog

TCP

net.ipv4.tcp_mem
net.ipv4.tcp_rmem
net.ipv4.tcp_wmem

UDP

net.ipv4.udp_mem
net.ipv4.udp_rmem_min
net.ipv4.udp_wmem_min

Mrežne statistike na ovoj razini mreže se mogu dobiti iz datoteka: /proc/softirqs, /proc/net/softnet_stat, /proc/net/dev te /proc/net/netstat i /proc/net/snmp
Odnosno sa sljedećim naredbama:
netstat -i, netstat -s te ss kao i naredbom : nstat

Opseg portova dostupnih za uporabu za sve mrežne konekcije na računalu: **net.ipv4.ip_local_port_range**
Omogućavanje usmjeravanja (routinga) između mrežnih sučelja računala/poslužitelja: **net.ipv4.ip_forward**
Omogućavanje upotrebe IP adrese koja nije lokalna (pr. za VRRP): **net.ipv4.ip_nonlocal_bind, ...**

Druge funkcionalnosti na ovom sloju mreže (potpuno softverske ili djelomično softverske):

- Receive Packet Steering (RPS) - za provjeru i promjenu: **/sys/class/net/eth0/queues/**
- Transmit Packet Steering (XPS) - za provjeru i promjenu (broj niza je XX): **/sys/class/net/eth0/queues/tx-XX/xps_cpus**
- Generic receive offload (GRO) - za uključivanje: **ethtool -K eth0 gro on**
- Generic segmentation offload (GSO) - za uključivanje: **ethtool -K eth0 gso on**

Odabrani scheduler za sva mrežna sučelja je postavljen u: **net.core.default_qdisc**

Za listu svih dostupnih schedulera (pr. code1, pfifo fast, fq, ...) pokrenuti naredbu:

grep '^CONFIG NET SCH' /boot/config-\$(uname -r)

Za konfiguraciju za pojedino mrežno sučelje; pr. eth0 za odabir fq algoritma/schedulera:

tc qdisc add dev eth0 root fq

Međumemorijska za sve opcije koje se snimaju uz mrežne pakete (pr. mrežno sučelje, zaglavlja, greške, ...):
net.core.optmem_max

(RX) Međumemorijska zajednička za sva mrežna sučelja na sustavu: **net.core.netdev_max_backlog**

TXqueue len međumemorijska
Moguće je mijenjati veličinu (xx):
ip link set eth0 txqueue len XX

NAPI Pooling

NAPI Pooling

Broj paketa koji se mogu spremati u međumemoriju prije NAPI dohvaćanja. Ako imamo uključene LRO i GRO onda se to odnosi na agregirane/spojene pakete: **net.core.dev_weight**
Ukupan broj paketa koji se mogu dohvatiti odjednom, zajednički za sva mrežna sučelja na sustavu:
net.core.netdev_budget
Vrijeme u µs dozvoljeno za dohvaćanje (pooling) paketa:
net.core.netdev_budget_usecs

Za mrežne statistike na ovoj razini (za eth0), možemo koristiti naredbe poput:
ip -s link te sar -n EDEV kao i **ethtool -S eth0** za najnižu razinu mreže.

Moguće je gledati statistike i direktno iz datoteka: /proc/net/dev ili /proc/net/netstat ili /proc/net/snmp

Moguće je povećanje/smanjenje ove međumemorije do određene hardverske granice (za RX=YY te za TX=ZZ) sa sljedećom naredbom:

ethtool -G eth0 rx YY tx ZZ

Hardverski ubrzane funkcionalnosti mrežnih kartica (pr. eth0), ako ih imaju:

Receive Side Scaling (RSS) - za povećanje RX i TX nizova na XX i YY: **ethtool -L eth0 rx XX tx YY**

Hardverski Receive Flow Steering (aRFS) - za uključivanje: **ethtool -K eth0 ntuple on**

Large receive offload (LRO) - za uključivanje: **ethtool -K eth0 lro on**

TCP segmentation offload (TSO) - za uključivanje: **ethtool -K eth0 tso on**

Checksum offload - uključivanje i za primanje (RX) i za slanje (TX): **ethtool -K eth0 rx on tx on**

Scatter-gather - za uključivanje: **ethtool -K eth0 sg on**

Hardverska podrška za VLAN (802.1Q) za TX i RX: **ethtool -K eth0 rxvlan on txvlan on**

Postavke mrežne kartice, poput brzine i duplex načina rada; pr. 1000Mbps (1Gbps), full duplex:

ethtool -s eth0 speed 1000 duplex full

Optimizacije signala prekida (IRQ) prema CPU afinitetu (za IRQ broja XY) se podešavaju u:

/proc/irq/XY/smp_affinity

Fizička mrežna sučelja (pr. eth0) se sa TAP sučeljem spajaju preko Linux bridge sučelja (pr. br0) ili OVS bridge sučelja!
VLAN-ovima se postiže izolacija mreže, dok se VxLAN tunelima (OVS) obično povezuju računala (poslužitelji) među sobom, prema standardnoj implementaciji OpenStack-a.
Sva fizička i logička Linux mrežna sučelja se mogu spojiti na vatrozid Linuxa (Firewall).
MTU veličina postoji i može se mijenjati na svakom mrežnom sučelju Linuxa!

tap1 Naziv mrežnog sučelja pod Linuxom tapXY Naziv OpenStack mrežnog sučelja pod Linuxom

GPL licenca
Hrvoje Horvat
v.1.10