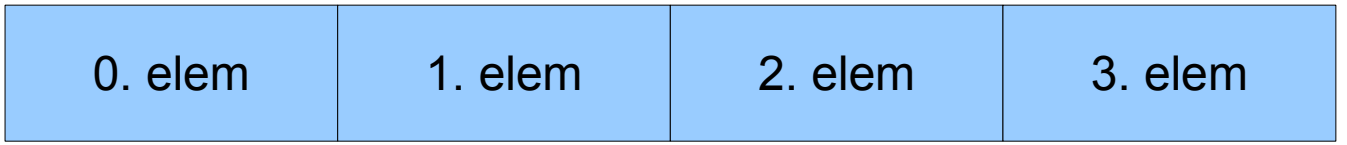
Ipset és essence

Kadlecsik József
MTA Wigner FK
<kadlecsik.jozsef@wigner.mta.hu>

Tartalom

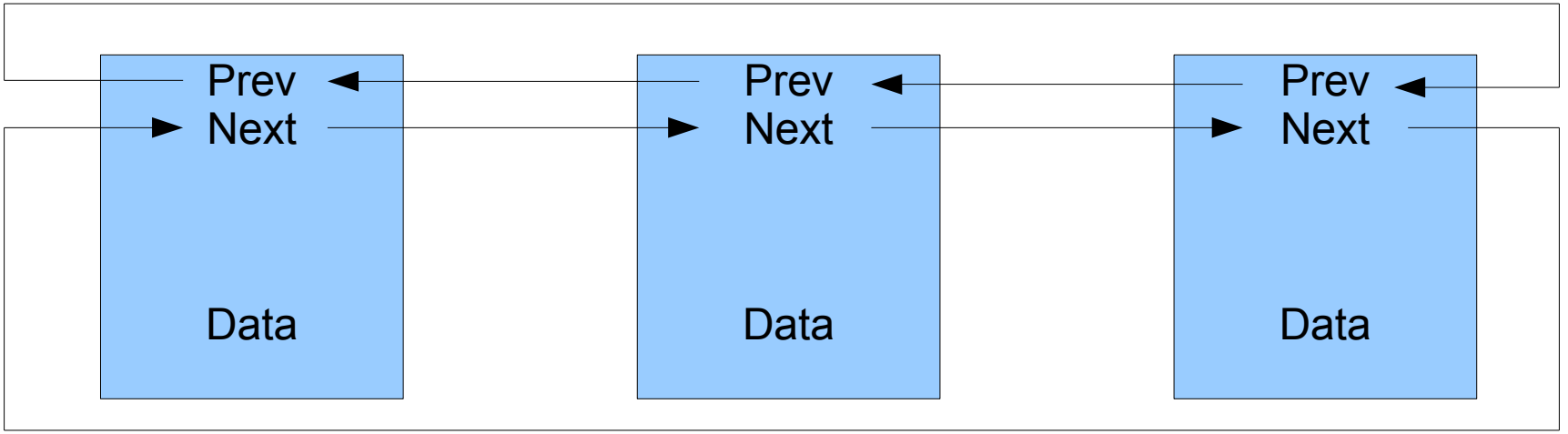
- Kernel tárolási módszerek
- Konkurrens hozzáférés: locking
- Netfilter/iptables
- Ipset
- Essence
- Teljesítmény adatok

Tömbök (arrays)



- Előnyök
  - Gyors hozzáférés bármely elemhez
  - Folytonos a memóriában
- Hátrányok
  - Tudni kell előre az elemek számát
  - Méret nem változtatható

Láncolt listák



- Előnyök
  - Egyszerű módosíthatóság
- Hátrányok
  - Konkurrens hozzáférést koordinálni kell
  - Memóriában szétszórt

Hash táblák

- Tömbök
- A tömbbeli pozíciót az adatból alapján ún. hash függvénnyel számoljuk ki
- Az ugyanazon pozícióra eső elemeket láncolt listában tároljuk

- Locking

Spinlock

CPU0

CPU1

CPU2

Rwlock

CPU0

CPU1

CPU2

RCU

- Szinkronizációs mechanizmus
  - Konkurrens írás lock-al kizárt
  - De írással párhuzamosan olvasás lock nélküli
- Módosított adatokból több másolat, addig nincs törlés, amíg van folyamatban levő olvasó

Netfilter/iptables

- Netfilter: állapottartó csomagszűrő tűzfal keretrendszer a Linux kernelben
  - Iptables/ip6tables: egy lehetséges implementáció
- Iptables/ip6tables:
  - A teljes szabálytábla egyetlen tömb
  - Előnyök… hátrányok
- Conntrack
  - Hash tábla

Ipset

- Csomag egyezés “halmazban” tárolt adatokkal
  - Gyors feldolgozás
  - Gyors módosítás
  - Tömeges iptables/ip6tables szabályok helyettesítése ipset-el
- Halmazhoz elemek dinamikusan hozzáadhatók
- Az irány nincs benne a halmazban
- Különböző tárolási módszerek
  - Bitmap
  - Speciális hash tábla
  - Láncolt lista

Tárolható adatok

- ip
- net
- port
- mac
- set

- ip, port
- net, port
- net, net
- ip, mac
- ip, mark
- net, iface

- ip, port, ip
- ip, port, net
- net, port, net

Elemenkénti kiterjesztések

- Timeout
- Packet/byte counter
- Comment
- Skbinfo
  - skbmark: fw mark
  - skbprio: tc class
  - skbqueue: hardware queue

Essence

- Tűzfal konfiguráló interfész: iptables/ip6tables/ipset
- Egyszerű leíró nyelv
- Automatikus egress-ingress filtering
- Dinamikus tiltás portscannerekre
- Egyetlen perl script néhány segéd fájllal

Példa

```
zone = internet
  interface = eth0
  network = 192.168.112.0/21
  network = 0/0, ::
zone = intranet
  interface = eth1
  network = 192.168.0.0/16, 2001:db8:1234::/48
policy = smtp.server
  ip = 192.168.0.25, 2001:db8:1234::25
  service = smtp, submission, ping
  client = any
```

Teljesítmény adatok

- Jesper Dangaard Brouer
- 10G ixgbe interfész a packet generátor és a tesztgép oldalán
- PREEMPT kernel
- Generált flood:
  - Forrás IP: 198.18.1.x
  - Dest port: 80
- Sender: 12Mpps

DROP a raw táblában iptables-el

```
iptables -t raw -N simple
iptables -t raw -A simple -s 198.18.0.0/15 -j DROP
iptables -t raw -A PREROUTING -j simple
```

- Fogadó teljesítmény: 11.3Mpps

DROP a raw táblában ipset-el

```
echo “create test hash:ip hashsize 65536” > load.set
for x in `seq 0 255`; do
  for y in `seq 0 255; do
    echo “add test 198.18.$x.$y” >> load.set
  done
done
ipset restore < load.set
iptables -t raw -N net198
iptables -t raw -A net198 -m set --match-set test src -j DROP
iptables -t raw -A PREROUTING -j net198
```

- Fogadó teljesítmény: 8Mpps

Performance numbers

```
+   24.65%  ksoftirqd/1  [ip_set]            [k] ip_set_test
-   21.42%  ksoftirqd/1  [kernel.kallsyms]  [k] _raw_read_lock_bh
   - _raw_read_lock_bh
      + 99.88% ip_set_test
-   19.42%  ksoftirqd/1  [kernel.kallsyms]  [k] _raw_read_unlock_bh
   - _raw_read_unlock_bh
      + 99.72% ip_set_test
+    4.31%  ksoftirqd/1  [ip_set_hash_ip]   [k] hash_ip4_kadt
+    2.27%  ksoftirqd/1  [ixgbe]            [k] ixgbe_fetch_rx_buffer
+    2.18%  ksoftirqd/1  [ip_tables]        [k] ipt_do_table
+    1.81%  ksoftirqd/1  [ip_set_hash_ip]   [k] hash_ip4_test
```

DROP a raw táblában RCU ipset-el

```
echo “create test hash:ip hashsize 65536” > load.set
for x in `seq 0 255`; do
  for y in `seq 0 255; do
    echo “add test 198.18.$x.$y” >> load.set
  done
done
ipset restore < load.set
iptables -t raw -N net198
iptables -t raw -A net198 -m set --match-set test src -j DROP
iptables -t raw -A PREROUTING -j net198
```

- Fogadó teljesítmény: 11.3Mpps

Letöltési oldalak

- ipset:
  - http://ipset.netfilter.org
  - https://git.netfilter.org
- Essence:
  - http://git.kfki.hu