Netfilterに関するエトセトラ

2012/12/21
ENOG18
(株)創風システム 外山 文規

CentOS6のNetfilterを中心に

ゆるくふんわりと調べてみました。

Netfilterに関するエトセトラ

- Netfilterの簡単なおさらい
- CentOS6のNetfilter
- CentOS6.1以降のNetfilter
- kernel2.6.32以降のNetfilter

Netfilterの簡単なおさらい

・Netfilterとは

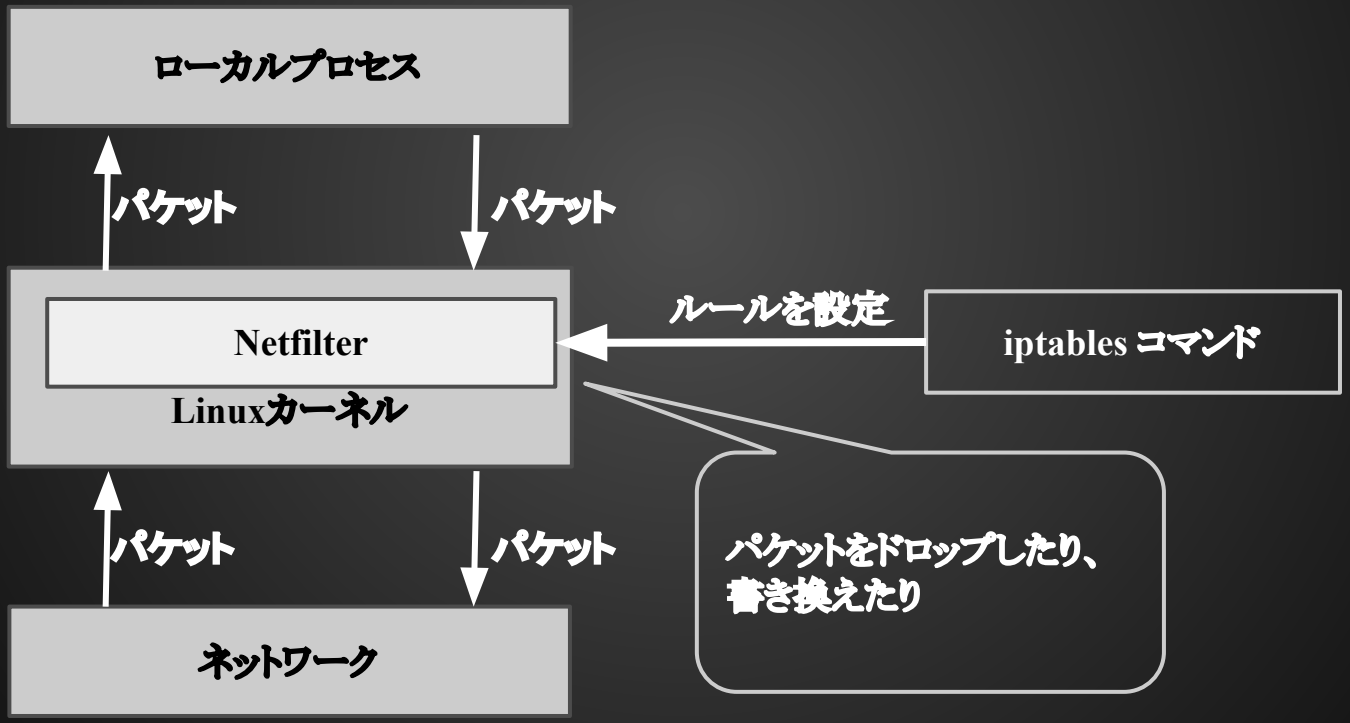
ネットワークパケットをインターセプトして操作するためのLinuxカーネル内でフック処理を提供するフレームワーク

Netfilterの簡単なおさらい

・iptablesとは

Linuxカーネル内のNetfilterのフレームワークを制御するためのコマンドラインツール

図にするとこんな感じ


ローカルプロセス
パケット
パケット
Netfilter
Linuxカーネル
ルールを設定
iptables コマンド
パケット
パケット
パケットをドロップしたり、
書き換えたり
ネットワーク

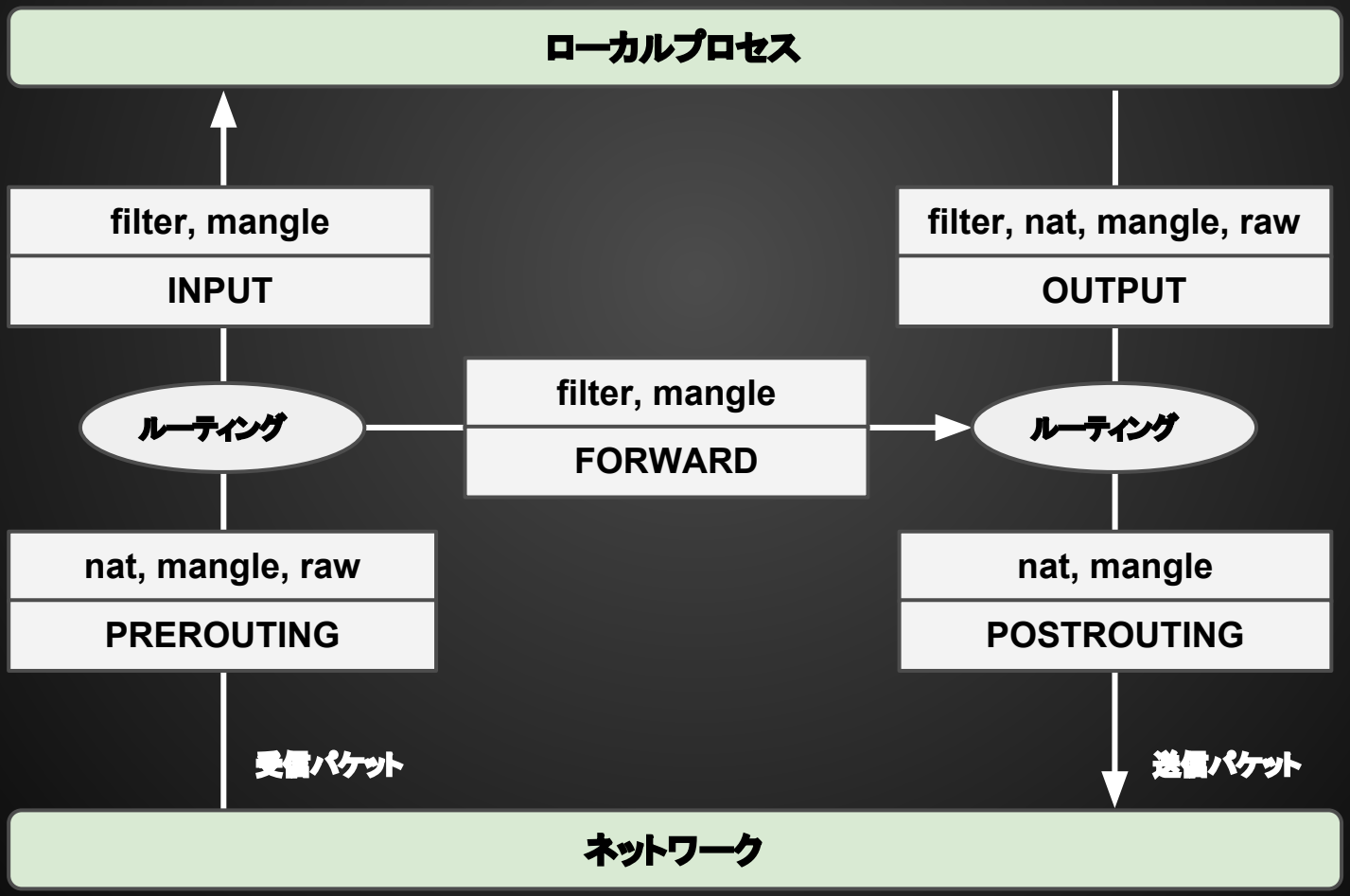
Netfilterできること

- **フィルタリング**
  特定の通信をブロックしたり通過させたり
- **NAT**
  ネットワークアドレス変換したり
- **その他**
  特定のプロトコルのオプションを変更したり

iptablesを設定するときの構成要素

- **テーブル**
  フィルタ、NAT等の役割ごとに使い分ける
- **チェイン**
  ビルドインチェインとユーザ定義の独自チェインがある
- **マッチ**
  ルールを適用するパケットの条件定義
- **ターゲット**
  該当するパケットに対して行うアクション

テーブルとチェインのフロー図



Connection Tracking

- パケットを追跡して状態を監視する機構
- 追跡が難しいプロトコルはヘルパーを用意されている(FTP、IRC、etc...)

CentOS5 -> CentOS6

- バージョンの変化

  kernel 2.6.18 -> 2.6.32

  iptables 1.3.5 -> 1.4.7

CentOS5 -> CentOS6

Red Hat Enterprise Linux 6移行計画ガイドによると

4.3. IP テーブル/ファイアウォール

IPTable には **SECMARK** ターゲットモジュールが含まれています。 SELinux などのセキュリティサブシステムで使用されるパケットに関連付けられるセキュリティマーク値の設定に使用されます。 mangle 表内でのみ有効となります。 使用例については以下を参照してください。

```
iptables -t mangle -A INPUT -p tcp --dport 80 -j SECMARK --selctx \
system_u:object_r:httpd_packet_t:s0
```

CentOS5 -> CentOS6

ん？ これだけ？

調べてみた

- CentoOS5と6のman iptablesの差分
  (1.3.5-9.1 と 1.4.7-3の差分)
- kernel newbiesから
  ※http://kernelnewbies.org
- iptablesのsrpmから
- RHEL 5.x 6.xのリリースノート

調査した結果

- **connection trackingが変わってた**

  ip_conntrack -> nf_conntrack

- **SECMARK以外にも機能拡張されいていた**

  TPROXY target 、TCPOPTSTRIP target、etc

kernel newbies(2.6.19->2.6.32)

Linux 2 6 19(http://kernelnewbies.org/Linux_2_6_19)
なし

Linux 2 6 20(http://kernelnewbies.org/Linux_2_6_20)
- **Add full NAT support for nf_conntrack**
- **Add IRC helper port , FTP NAT helper port , SIP helper port , TFTP helper port , PPTP helper port , H.323 helper port . NetBIOS name service helper port , SNMP NAT helper port**
- **sysctl and /proc compatibility with old connection tracking , tatistics.**
- **x_tables: add port of hashlimit match for IPv4 and IPv6 and add NFLOG target**
- **ebtables: add --snap-arp option**

Linux 2 6 21(http://kernelnewbies.org/Linux_2_6_21)
- **NAT: optional source port randomization support**
- **Add IPv6-capable TCPMSS target support**
- **Add SANE connection tracking helper**
- **Introduces match for Mobility Header (MH) described by Mobile IPv6 specification (RFC3775)**

kernel newbies(2.6.19->2.6.32)

Linux 2 6 22(http://kernelnewbies.org/Linux_2_6_22)

- **Remove IPv4 only connection tracking/NAT**
- **Add support for user mode STP**
- **ipt_DNAT: accept port randomization option**

Linux 2 6 23(http://kernelnewbies.org/Linux_2_6_23)

- **nf_conntrack: UDPLITE support , introduce extension infrastructure , use extension infrastructure for helper , remove old memory allocator of conntrack , use hashtable for expectations , nf_conntrack_helper: use hashtable for conntrack helpers**
- **nf_nat: add reference to conntrack from entry of bysource list , use extension infrastructure**
- **Add u32 match**
- **x_tables: add TRACE target , add connlimit match**

Linux 2 6 24(http://kernelnewbies.org/Linux_2_6_24)

- **Add xt_time match: a "time" match, which allows you to match based on the packet arrival time (at the machine which netfilter is running) or departure time/date (for locally generated packets)**

kernel newbies(2.6.19->2.6.32)

Linux 2 6 25(http://kernelnewbies.org/Linux_2_6_25)

- **Add CONFIG_NETFILTER_ADVANCED option.**
- **x_tables: add TCPOPTSTRIP target.**
- **Merge ipt_tos into xt_dscp. , merge ipt_TOS into xt_DSCP.**
- **IPv6 capable xt_TOS v1 target. , IPv6 capable xt_tos v1 match.**
- **ip_tables: remove obsolete SAME target.**
- **x_tables: add RATEEST target. , add rateest match**
- **ctnetlink: add support for secmark.**

Linux 2 6 26(http://kernelnewbies.org/Linux_2_6_26)

- **nf_conntrack: add DCCP protocol support**
- **nf_nat: add UDP-Lite support , add DCCP protocol support , add SCTP protocol support**

kernel newbies(2.6.19->2.6.32)

Linux 2 6 27(http://kernelnewbies.org/Linux_2_6_27)
- **ebtables: add IPv6 support ,**
- **ctnetlink: add full support for SCTP to ctnetlink**
- **ip_tables: add iptables security table for mandatory access control rules**
- **ip6_tables: add ip6tables security table**
- **accounting rework: ct_extend + 64bit counters**

Linux 2 6 28(http://kernelnewbies.org/Linux_2_6_28)
- **Transparent proxy support**
- **Enable netfilter in netns**
- **xt_recent: IPv6 support**

Linux 2 6 29(http://kernelnewbies.org/Linux_2_6_29)

なし

kernel newbies(2.6.19->2.6.32)

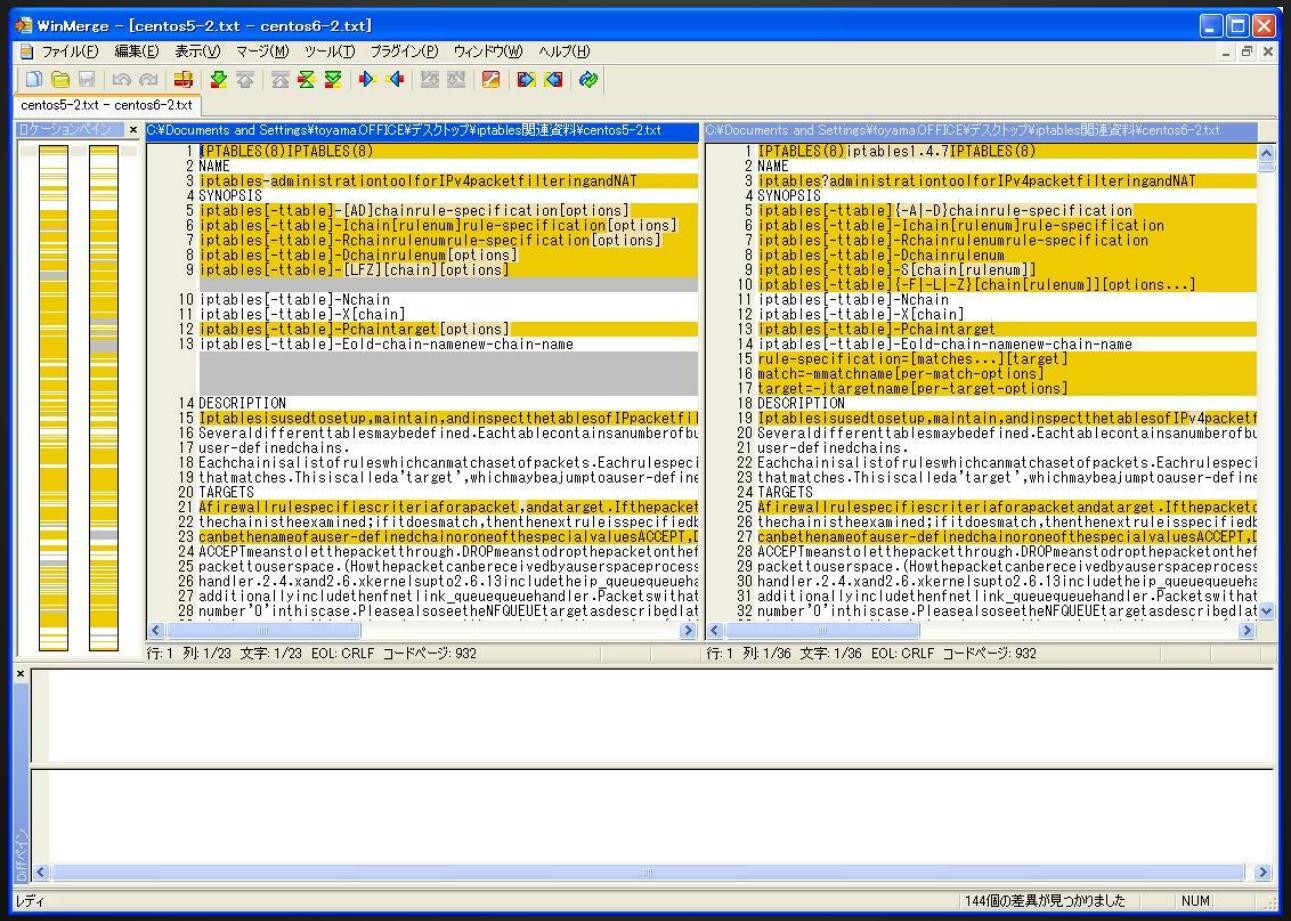
Linux 2 6 30(http://kernelnewbies.org/Linux_2_6_30)
- **Combine ipt_TTL and ip6t_HL source**
- **Combine ipt_ttl and ip6t_hl source**
- **iptables: lock free counters**
- **sysctl support of logger choice**
- **xtables: add cluster match , add LED trigger target**

Linux 2 6 31(http://kernelnewbies.org/Linux_2_6_31)
- **conntrack: add support for DCCP handshake sequence to ctnetlink**
- **conntrack: optional reliable conntrack event delivery**
- **nf_ct_tcp: TCP simultaneous open support**
- **passive OS fingerprint xtables match**
- **xt_NFQUEUE: queue balancing support**

Linux 2 6 32(http://kernelnewbies.org/Linux_2_6_32)
なし

man iptablesのdiff

← CentOS5のman iptables　　CentOS6のman iptables →



ほとんどは、オプションの記述を変えたり、説明文の修正によるもの。

nf_conntrack?

・ip_conntrackの後継
・IPv6への対応強化
・kernel2.6.15より追加
・kernel2.6.22からはip_conntrackが削除

CentOS6で追加された機能

- **TRACE target**

  IPパケットがテーブルとルールをどう通過するのか
  追跡するために使用。

- **TCPOPTSTRIP target**

  TCPパケットからTCPオプションを取り除いて、
  NO-OPSに置き換える。
  Window scaling optionの削除が目的？

CentOS6で追加された機能

- **LED target**

  ルールに一致したらLEDを点灯する。

- **TPROXY target**

  REDIRECT類似した機能で、透過プロキシを行う場合に使用する。Netfilterのコネクショントラッキングや NATに依存しない。

CentOS6で追加された機能

- **NFLOG target**

ULOGの後継。マッチしたパケットについて、ユーザ空間でのロギングを提供してくれる。

- **CHECKSUM target**

checksum offloadに対応してない古いアプリケーション(bootpcなど)の場合にチェックサムを計算してヘッダに書き込む

CentOS6で追加された機能

- **RATEEST target**

  流れているパケットのBPS/PPSを計測、レートの
  比較を行い負荷ベースのマルチパスルーティングを
  実現する

これ以外にも機能改善されています。

```
# Estimate outgoing rates
iptables -t mangle -A POSTROUTING -o eth0 -j RATEEST --rateest-name eth0 \
--rateest-interval 250ms --rateest-ewma 0.5s
iptables -t mangle -A POSTROUTING -o ppp0 -j RATEEST --rateest-name ppp0 \
 --rateest-interval 250ms --rateest-ewma 0.5


# Mark based on available bandwidth
iptables -t mangle -A balance -m conntrack --ctstate NEW -m helper --helper ftp \
-m rateest --rateest-delta --rateest1 eth0 --rateest-bps1 2.5mbit --rateest-gt \
--rateest2 ppp0 --rateest-bps2 2mbit -j CONNMARK --set-mark 1
iptables -t mangle -A balance -m conntrack --ctstate NEW -m helper --helper ftp \
-m rateest --rateest-delta --rateest1 ppp0 --rateest-bps1 2mbit --rateest-gt \
--rateest2 eth0 --rateest-bps2 2.5mbit -j CONNMARK --set-mark 2
iptables -t mangle -A balance -j CONNMARK --restore-mark
```

CentOS6で削除された機能

- **削除された target**
  BALANCE、IPMARK、TARPIT、XOR

- **削除された　match**
  account、childlevel、condition、connrate、
  dstlimit、ipv4options、mport、nth、osf

CentOS6.1以降で追加された機能

- **AUDIT target**

  ルールに一致したらauditのログに記録する

ipsetという新しいツール

・CentOSでは6.3からパッケージが追加

→kernel 2.6.39で追加された機能

ipsetとは？

・newbiesよりipsetの紹介をざっくり訳

1. IPsetと呼ばれるネットワークリソースグループを定義できる。
（ネットワークリソース= IPv4/v6、TCP/UDPポート番号、IPとMACのペア、IPとポート番号のペアなど）

2. IPsetで定義したグループはiptablesのルールとして使う。

3. 普通にiptablesのルールを定義するより、パフォーマンスが向上する。
但し、メモリ消費は多い。

4. 複数のIPアドレスまたはポート番号とのマッチングが必要な場合に
効果がある。

ipsetとは？

・newbiesよりipsetの紹介をざっくり訳

1. IPs
（ネ

2. IP

3. 普
但

4. 複
効果かめる。

BlacklistやWhitelistで
使うとにとても幸せ

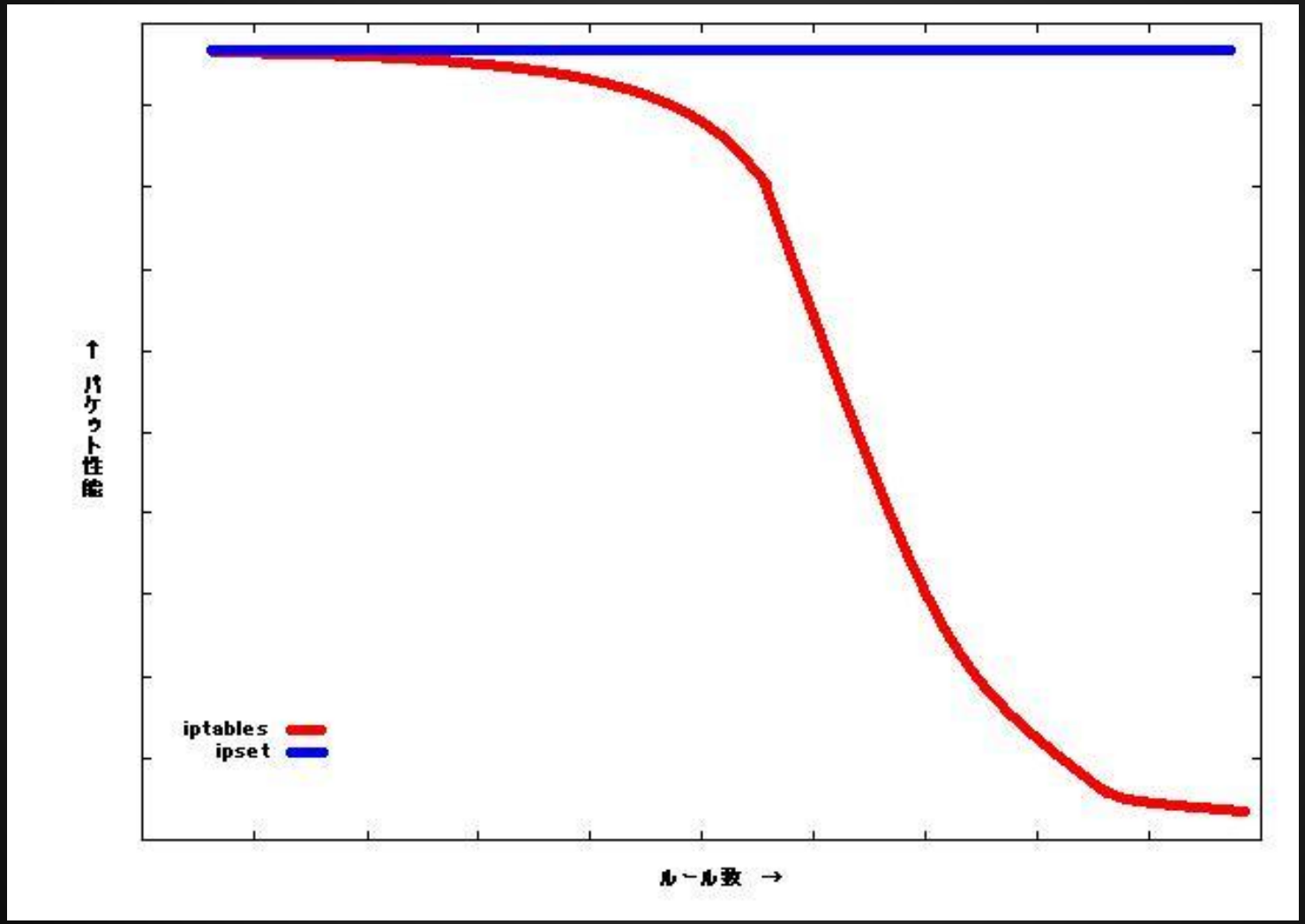
ipsetのパフォーマンス

すいませんベンチマークとってないです。

参考:

- **Mass-blocking IP addresses with ipset**
  http://daemonkeeper.net/781/mass-blocking-ip-addresses-with-ipset/
- **Netfilter Performance Testing**
  http://people.netfilter.org/kadlec/nftest.pdf

ルール数による性能への影響



ipsetのインストール

・CentOS6ならyumで

```
# yum install ipset
```

IPsetの使い方

・設定の流れ

1. ipset create で「セット」を作成
2. ipset add でエントリを追加
3. iptables コマンドで「セット」を使用したルール
   を追加

IPsetでセットを作成

ipset create セット名 セットタイプ [オプション]
ipset add セット名 エントリ [オプション]

例)
ipset create DenyIP hash:ip timeout 0
ipset add DenyIP 192.168.1.10 timeout 30

ipset list DenyIP

セットタイプ

用途に合わせてセットタイプ指定します

セットタイプ (ipset add する時のフォーマット)

- bitmap:ip (ip | fromip-toip | ip/cidr)
- bitmap:ip,mac (ip[,macaddr])
- bitmap:port (port | fromport-toport)
- hash:ip (ip)
- hash:net (ip[/cidr])
- hash:ip,port (ip,[proto:]port)
- hash:net,port (ip[/cidr],[proto:]port)
- hash:ip,port,ip (ip,[proto:]port,ip)
- hash:ip,port,net (ip,[proto:]port,ip[/cidr])
- hash:net,iface (ip[/cid,[physdev:]iface)
- list:set (setname [{ before | after } setname])

セットをiptablesのルールとして適用

マッチの条件として使用(set match)

例) DenyIPセットに一致した送信元をDROP

```
# iptables -A INPUT -m set --match-set DenyIP src -j DROP
```

「セット」内のエントリを変更(SET target)

例) sshにアクセスに来た接続元を DenyIPセットに追加

```
# iptables -A INPUT -p tcp --dport 22 -j SET --add-set DenyIP src
```

ipsetの保存、復元

保存

＃ ipset save > *filename*

復元

＃ ipset restore < *filename*

kernel newbies(2.6.33->3.7)

Linux 2 6 33(http://kernelnewbies.org/Linux_2_6_33)
なし

Linux 2 6 34(http://kernelnewbies.org/Linux_2_6_34)
- **nf_conntrack_sip: add T.38 FAX support, add TCP support**
- **nf_conntrack: support conntrack templates, add support for "conntrack zones"**
- **nf_nat_sip: add TCP support**
- **ctnetlink: add zone support**
- **ebtables: add CONFIG_COMPAT support**
- **xtables: add CT target**

Linux 2 6 35(http://kernelnewbies.org/Linux_2_6_35)
なし

kernel newbies(2.6.33->3.7)

Linux 2 6 36(http://kernelnewbies.org/Linux_2_6_36)
- **Add CHECKSUM target**
- **Add xt_cpu match**
- **ipt_LOG/ip6t_LOG: add option to print decoded MAC header**
- **xtables: idletimer target implementation**
- **xt_ipvs (netfilter matcher for IPVS)**

Linux 2 6 37(http://kernelnewbies.org/Linux_2_6_37)
- **Added IPv6 support to the TPROXY target**

Linux 2 6 38(http://kernelnewbies.org/Linux_2_6_38)
なし

cpu match

- どのcpuコアで処理しているかを判定
  マルチコアに対応してないレガシーアプリケーションをスケールさせる使用する

例)

```
iptables -t nat -A PREROUTING -p tcp --dport 80 -m cpu --cpu 0 -j REDIRECT --to-port 8080
iptables -t nat -A PREROUTING -p tcp --dport 80 -m cpu --cpu 1 -j REDIRECT --to-port 8081
iptables -t nat -A PREROUTING -p tcp --dport 80 -m cpu --cpu 2 -j REDIRECT --to-port 8082
iptables -t nat -A PREROUTING -p tcp --dport 80 -m cpu --cpu 3 -j REDIRECT --to-port 8083
```

idletimer target

・インタフェースが一定時間アイドル状態になっているかを特定するのに使用

kernel newbies(2.6.33->3.7)

Linux 2 6 39(http://kernelnewbies.org/Linux_2_6_39)

- **IPset**
- **Audit target to record accepted/dropped packets**
- **xtable: connlimit revision 1**
- **xtable: speedup compat operations**
- **xtable: "set" match and "SET" target support**
- **xt_addrtype: ipv6 support**
- **xt_CLASSIFY: add ARP support, allow CLASSIFY target on any table**
- **xt_conntrack: support matching on port ranges**
- **ebt_ip6: allow matching on ipv6-icmp types/codes**
- **nf_conntrack: nf_conntrack snmp helper**
- **nf_conntrack_tstamp: add flow-based timestamp extension**

kernel newbies(2.6.33->3.7)

Linux 3.0(http://kernelnewbies.org/Linux_3.0)
- **ipset: SCTP, UDPLITE support added**

Linux 3.1(http://kernelnewbies.org/Linux_3.1)
- **Add SELinux context support to AUDIT target**
- **ipset: support range for IPv4 at adding/deleting elements for hash:*net* types**

Linux 3.2(http://kernelnewbies.org/Linux_3.2)
なし

Linux 3.3(http://kernelnewbies.org/Linux_3.3)
- **Add extended accounting infrastructure over nfnetlink, which aims to allow displaying real-time traffic accounting without the need of complicated and resource-consuming implementation in user-space**
- **Add nfacct match to support extended accounting**
- **Add "rpfilter" reverse path filter match support, allows to match packets whose replies would go out via the interface the packet came in**

nfacct

・nfacctツールと組み合わせることで、
条件一致したパケットのアカウンティング
ができる。

例)

```
iptables -I INPUT -p tcp --sport 80 -m nfacct --nfacct-name http-traffic
iptables -I OUTPUT -p tcp --dport 80 -m nfacct --nfacct-name http-traffic
```

rpfilter match

- **Reverse Path Filterの検証**

  送信元アドレスをルーティングテーブルで見たときに
  同じインターフェイスから出ていくかを検証

kernel newbies(2.6.33->3.7)

Linux 3.4(http://kernelnewbies.org/Linux_3.4)

- **Add timeout extension. This allows you to attach timeout policies to flow via the connection tracking target**
- **ctnetlink: add NAT support for expectations class**
- **ipset: The "nomatch" keyword and option is added to the hash:*net* types, by which one can add exception entries to sets**
- **Merge ipt_LOG and ip6_LOG into xt_LOG**

Linux 3.5(http://kernelnewbies.org/Linux_3.5)

- **Add xt_hmark target for hash-based skb marking**
- **bridge: optionally set indev to vlan**
- **hashlimit: byte-based limit mode**
- **ipvs: add support for sync threads**
- **Remove ip_queue support**

Linux 3.6(http://kernelnewbies.org/Linux_3.6)

- **Add fail-open support**
- **Add user-space connection tracking helper infrastructure**

kernel newbies(2.6.33->3.7)

Linux 3.7(http://kernelnewbies.org/Linux_3.7)

- **Add protocol independent NAT core**
- **Add IPv6 MASQUERADE target**
- **Add IPv6 NETMAP target**
- **Add IPv6 REDIRECT target**
- **Add IPv6 NAT support**
- **Support IPv6 in FTP NAT helper**
- **Support IPv6 in IRC NAT helper**
- **Support IPv6 in SIP NAT helper**
- **Support IPv6 in amanda NAT helper**
- **Add stateless IPv6-to-IPv6 Network Prefix Translation target**
- **Remove xt_NOTRACK**

参考：

- http://www.netfilter.org/
- http://www.at.netfilter.org/patch-o-matic/pom-extra.html
- http://www.frozentux.net/documents/iptables-tutorial/
- http://www.asahi-net.or.jp/~aa4t-nngk/ipttut/index.html
- http://ja.wikipedia.org/wiki/Iptables
- http://member.wide.ad.jp/tr/
- http://kernelnewbies.org/
- https://access.redhat.com/knowledge/docs/Red_Hat_Enterprise_Linux/?locale=ja-JP

ご清聴ありがとうございました。