Oath Moloch Deployments

Andy Wick

Deployments

Oath has three different network types that we monitor, each with their own network design and scale.

- Office - Employees, VPNs
  - 50+ global offices, each with its own egress
  - 10 VPN concentrators
  - Centralized Elasticsearch cluster
- CiC - Backoffice in a data center
  - Each location with its own Elasticsearch cluster
- Prod - Production traffic
  - Each location with its own Elasticsearch cluster
  - Too much Gbps to capture everything
  - Some traffic we don't want to capture

Design

- AOL & Yahoo each had their own take on visibility
  - Combined the best of both for Oath
- Zeek (Bro), Suricata, Moloch and other tools
- Run all tools on each visibility box instead of specialized boxes
- Use a few hardware configurations so easy to reuse
- Use an NPB to load balance traffic
- Watch traffic to/from “internet”
- For production reduce traffic
  - Analyze traffic for less then half
  - Save PCAP for even smaller percent

NPB

- Aggregates, filters, and load balances traffic
- Normal Arista switch, in a special mode
  - Packets flow one direction
  - Still need another switch for standard networking
- Input: Span ports or IXIA optical taps
- Output: Visibility Hosts
- Office/CiC: 7150S-24, 7280SE
- Production: 7508R 13RU, 6 power supplies, max 11,484W

Why use a NPB?

- Easy to add Moloch capacity
- Allows the networking team and security team to act more independently
    - Networking team can add more links at any time, just connect taps to NPB
    - The security team can add more tool capacity at any time, just connect tools to NPB
- Move the traffic filtering from a bpf to purpose built hardware
- Multiple tools can see the same traffic (or subset), again making network team happy they aren’t involved
- Load balancing
- Handles HA issues of packets taking different paths
    - as long as all paths hit the same NPB

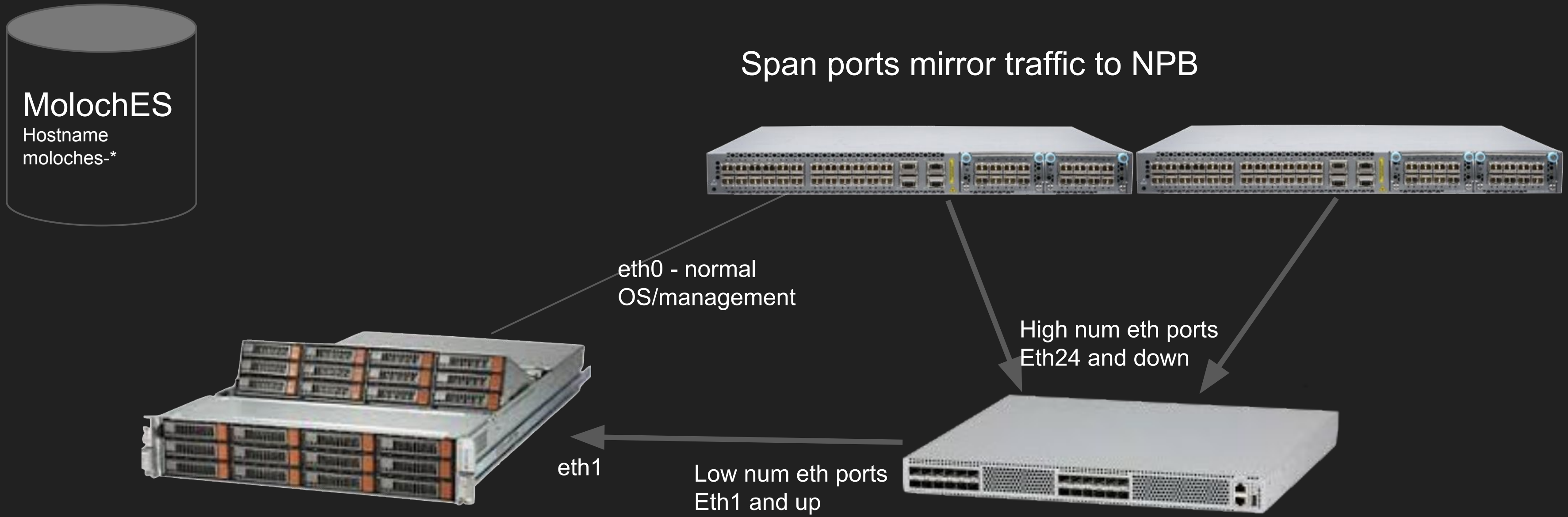
Visibility Hosts

- Bro is a memory/cpu hog
- Use afpacket for everything
  - requires a patch to Bro
- Want enough memory to potential run other tools and scanners in the future
- 2RU for space considerations, however boxes are deeper

Hardware Selected

- Keep number of configurations to a minimum
- Arista NPB
- Visibility boxes
  - New, Supermicro 6028R-E1CR24L
  - 24x10TB 128GB - Office, CiC
  - 24x12TB 256GB - Prod
- Moloches
  - Used, most are 5+ years old
  - 4x10TB 128GB - 1 node - Office, CIC
  - 4x12TB 256GB - 2 node - Prod
  - Session replication

Office/CiC Architecture



MolochES

Hostname
moloches-*

Span ports mirror traffic to NPB

eth0 - normal
OS/management

High num eth ports
Eth24 and down

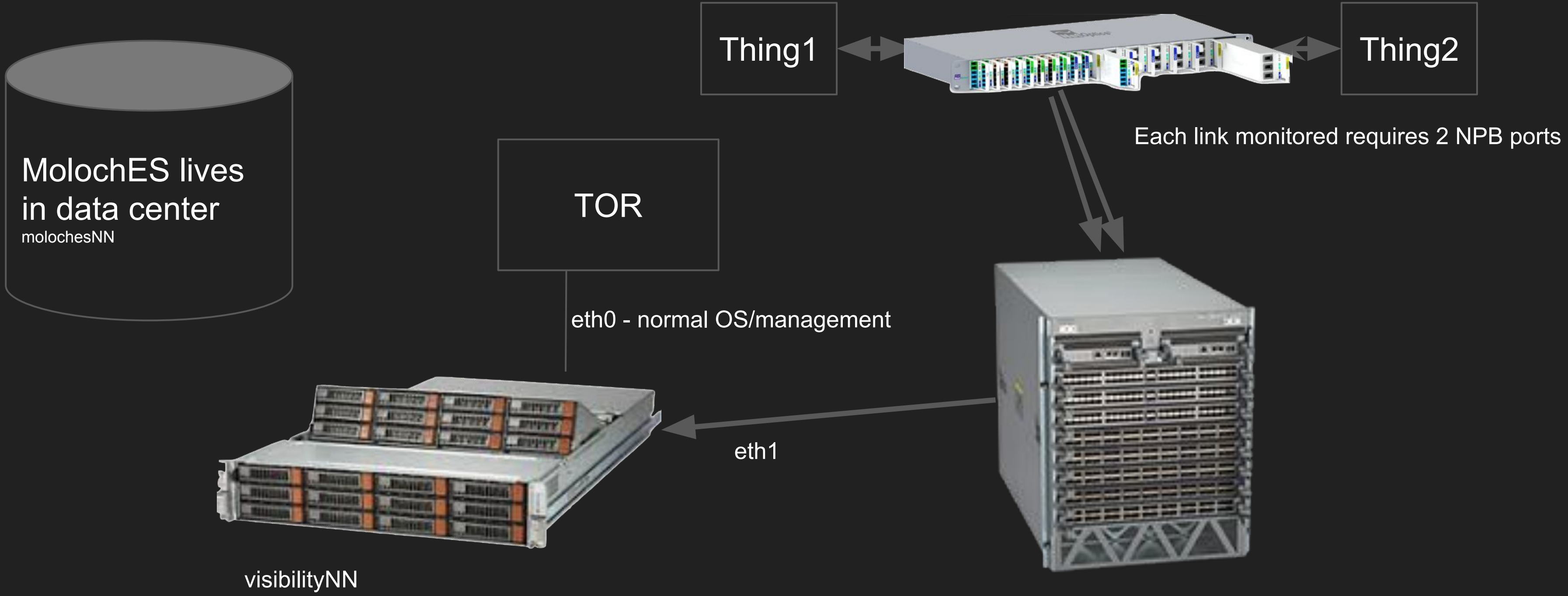
eth1

Low num eth ports
Eth1 and up

Most sites only have 1 or 2 visibility servers

Hostname: visibilityNN

Prod Architecture



MolochES lives in data center

molochesNN

Thing1

Thing2

Each link monitored requires 2 NPB ports

TOR

eth0 - normal OS/management

eth1

visibilityNN

Reality

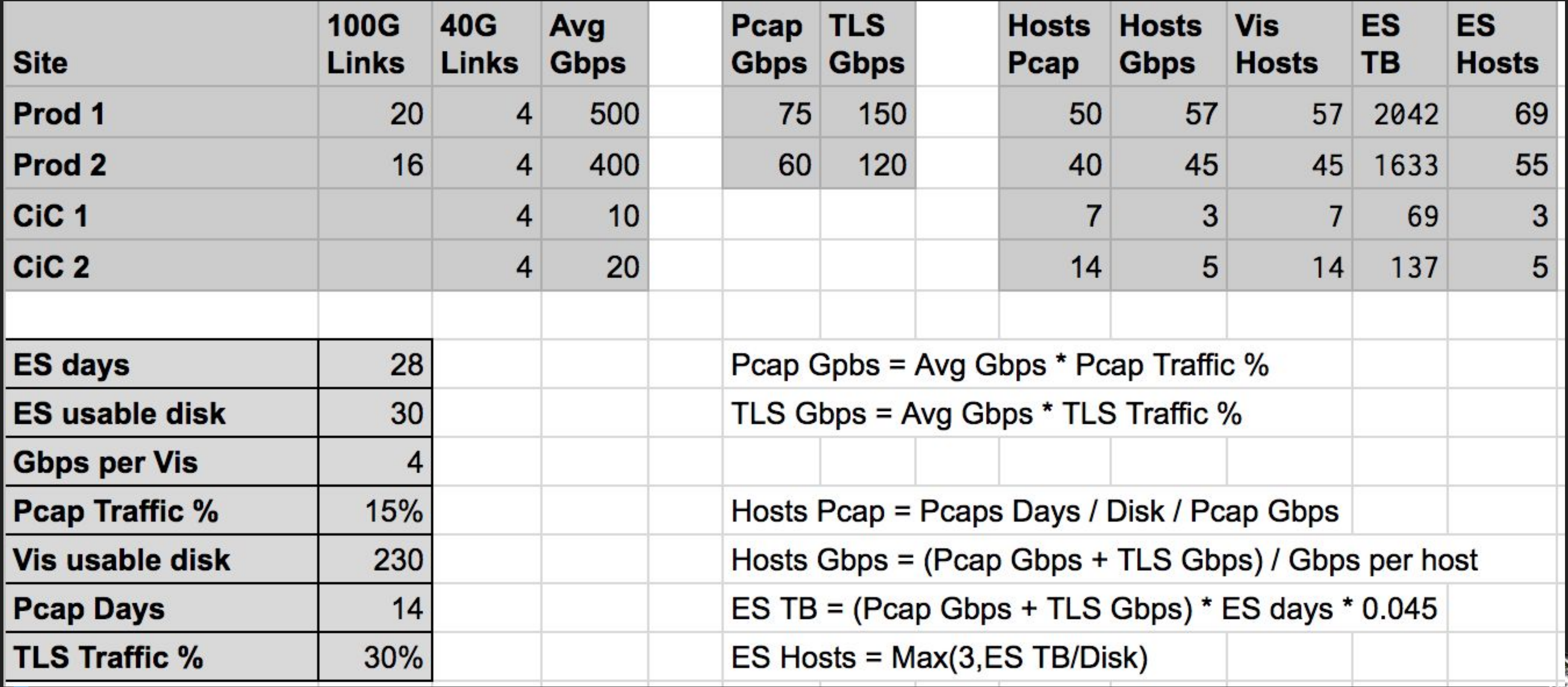
Things to watch for

- Hardware reliability
  - Might require more ES replication
  - Extra capture nodes
  - Extra hard drives on hand
- Configure multiple elasticsearch endpoints to handle failures
- Make sure Elasticsearch is configured with shard awareness
- Increase thread_pool.bulk.queue_size setting in ES
- Use ES 6.4.2 not 6.2.4 if using replication and ES 6.x
- Security, use iptables
- Number of ACLs NPB can handle

Sizing

- Office visibility sizing is done by number of employees.
    - Every site has an Arista NPB
    - Each visibility box can handle ~250 employees for desired retention
    - NPB is used for aggregation
- CiC & Prod sizing is done by avg Gbps
    - Every site has an Arista NPB
    - NPB aggregates traffic
    - NPB is used to drop traffic
    - Moloch rules are used to not save pcap

Example Sizing Sheet



Site	100G Links	40G Links	Avg Gbps		Pcap Gbps	TLS Gbps		Hosts Pcap	Hosts Gbps	Vis Hosts	ES TB	ES Hosts
Prod 1	20	4	500		75	150		50	57	57	2042	69
Prod 2	16	4	400		60	120		40	45	45	1633	55
CiC 1		4	10					7	3	7	69	3
CiC 2		4	20					14	5	14	137	5

ES days	28
ES usable disk	30
Gbps per Vis	4
Pcap Traffic %	15%
Vis usable disk	230
Pcap Days	14
TLS Traffic %	30%

Pcap Gpbs = Avg Gbps * Pcap Traffic %

TLS Gbps = Avg Gbps * TLS Traffic %

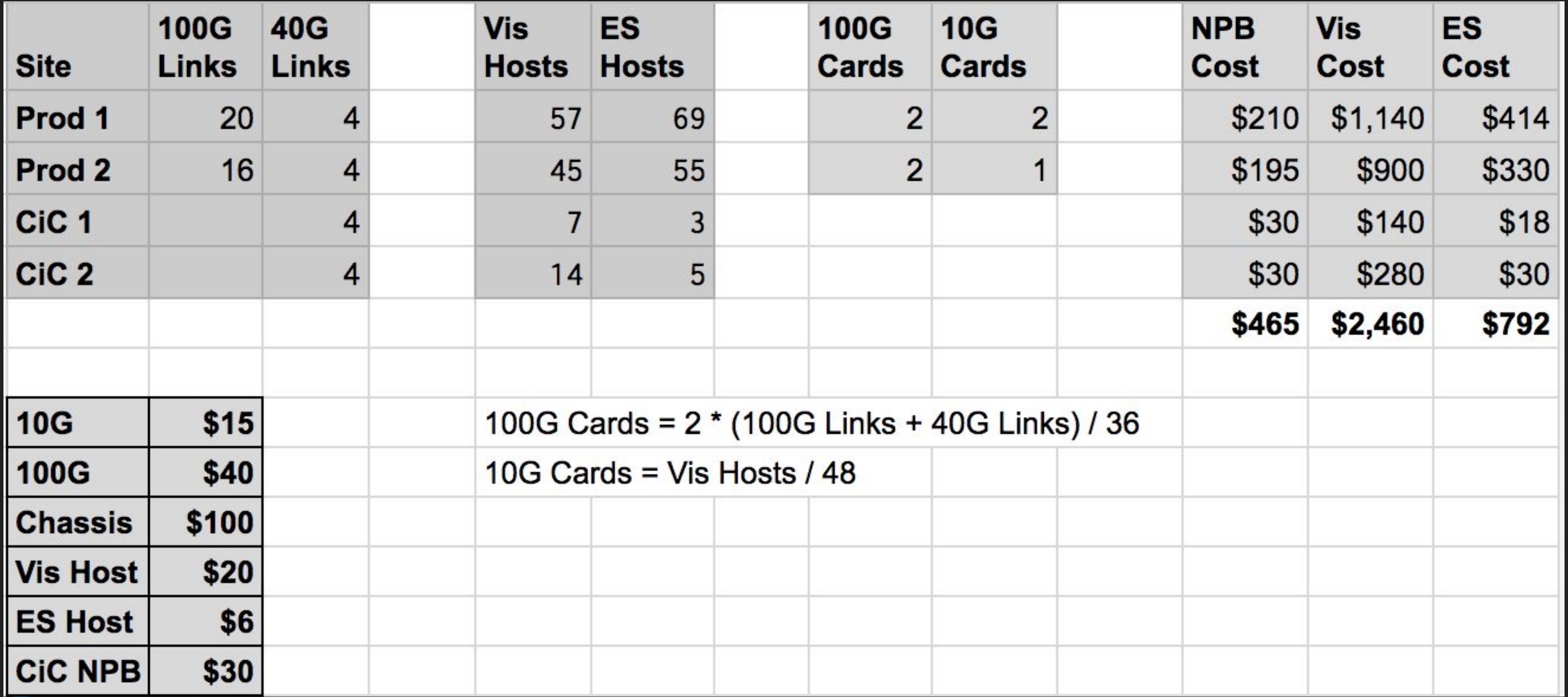
Hosts Pcap = Pcaps Days / Disk / Pcap Gbps

Hosts Gbps = (Pcap Gbps + TLS Gbps) / Gbps per host

ES TB = (Pcap Gbps + TLS Gbps) * ES days * 0.045

ES Hosts = Max(3,ES TB/Disk)

Example Costing



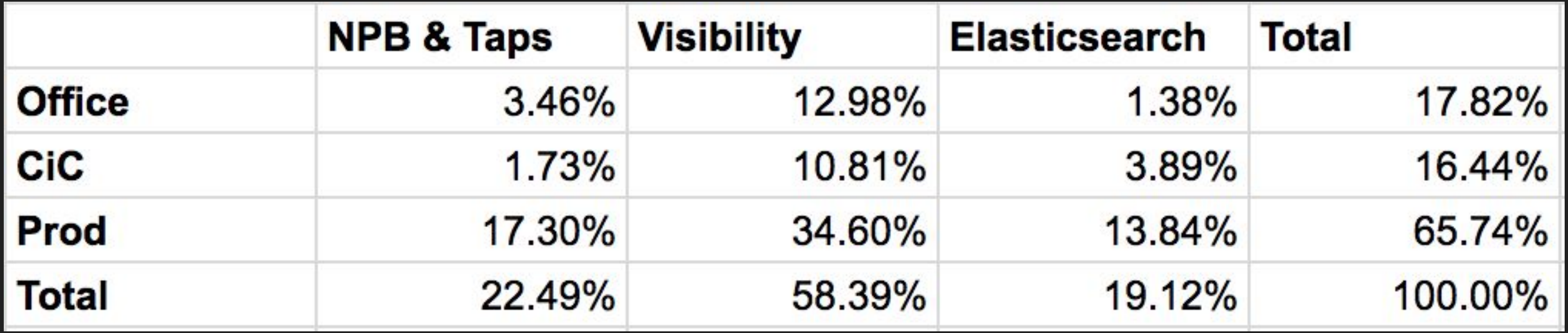
Site	100G Links	40G Links		Vis Hosts	ES Hosts		100G Cards	10G Cards		NPB Cost	Vis Cost	ES Cost
Prod 1	20	4		57	69		2	2		$210	$1,140	$414
Prod 2	16	4		45	55		2	1		$195	$900	$330
CiC 1		4		7	3					$30	$140	$18
CiC 2		4		14	5					$30	$280	$30
										$465	**$2,460**	**$792**

10G	$15
100G	$40
Chassis	$100
Vis Host	$20
ES Host	$6
CiC NPB	$30

100G Cards = 2 * (100G Links + 40G Links) / 36

10G Cards = Vis Hosts / 48

Reality Cost Breakdown



	NPB & Taps	Visibility	Elasticsearch	Total
Office	3.46%	12.98%	1.38%	17.82%
CiC	1.73%	10.81%	3.89%	16.44%
Prod	17.30%	34.60%	13.84%	65.74%
Total	22.49%	58.39%	19.12%	100.00%

Traffic Reduction

- NPB
  - Drop by ip/port
  - Simple perl script generates commands from CMDB
- Moloch
  - Use rules to drop traffic
  - Don’t save all the TLS packets
    - Helps with ES - don’t save file pos
    - Helps with Vis - reduces pcap storage
  - Don’t save SYN scans
  - Don’t save some ad network traffic to clouds

NPB Sample

```
mail-list       file:mail.yahoo.com     tcp     25
^(smtp)
mail-list    imap-a-mtc-a.mx.aol.com tcp     9993 9995
```

```
default ip access-list mail-list
ip access-list mail-list
! file:mail.yahoo.com - ^(smtp):25 ips=100
permit tcp any host 1.2.3.4 eq 25
permit tcp host 1.2.3.4 eq 25 any
permit tcp any host 4.3.2.1 eq 9993 9995
permit tcp host 4.3.2.1 eq 9993 9995 any
```

Prod Rules - Drop TLS after 10 packets

```
- name: "Drop tls"
   when: "fieldSet"
   fields:
     protocols:
     - tls
   ops:
     _maxPacketsToSave: 10
```

Prod Rules - Drop SYN scans

```
- name: "Drop syn scan"
    when: "beforeFinalSave"
    fields:
      packets.src: 1
      packets.dst: 0
      tcpflags.syn: 1
    ops:
      _dontSaveSPI: 1
```

Prod Rules - Drop traffic to cloud

```
- name: "Drop tls by hostname"
    when: "fieldSet"
    fields:
      host.http:
      - ad.doubleclick.net
      - foo.example.com
      protocols:
      - tls
    ops:
      _dontSaveSPI: 1
      _maxPacketsToSave: 1
      _dropByDst: 10
```

Other important high performance settings

```
# IMPORTANT, libfile kills performance
magicMode=basic

# Enable afpacket
pcapReadMethod=tpacketv3
tpacketv3BlockSize=8388608

# Increase by 1 if still getting Input Drops
tpacketv3NumThreads=2

# Start with 5 packet threads, increase by 1 if getting thread drops.  You
do NOT need 24 threads :)
packetThreads=5
```

Pcap Encryption at rest with Moloch

- Each pcap file has its own data encryption key (DEK)
- The DEK is encrypted using a key encryption key (KEK)
- The encrypted DEK, IV, and KEK id used for each file is stored in ES
- The list of KEKs and currently used KEK are stored in the moloch config.ini file

```
[default]
pcapWriteMethod=simple
simpleEncoding=aes-256-ctr
simpleKEKId=kekid1

[keks]
kekid1=Randomkekpassword1
kekid2=Randomkekpassword2
```

QUESTIONS?