Verizon Media Deployments

Andy Wick

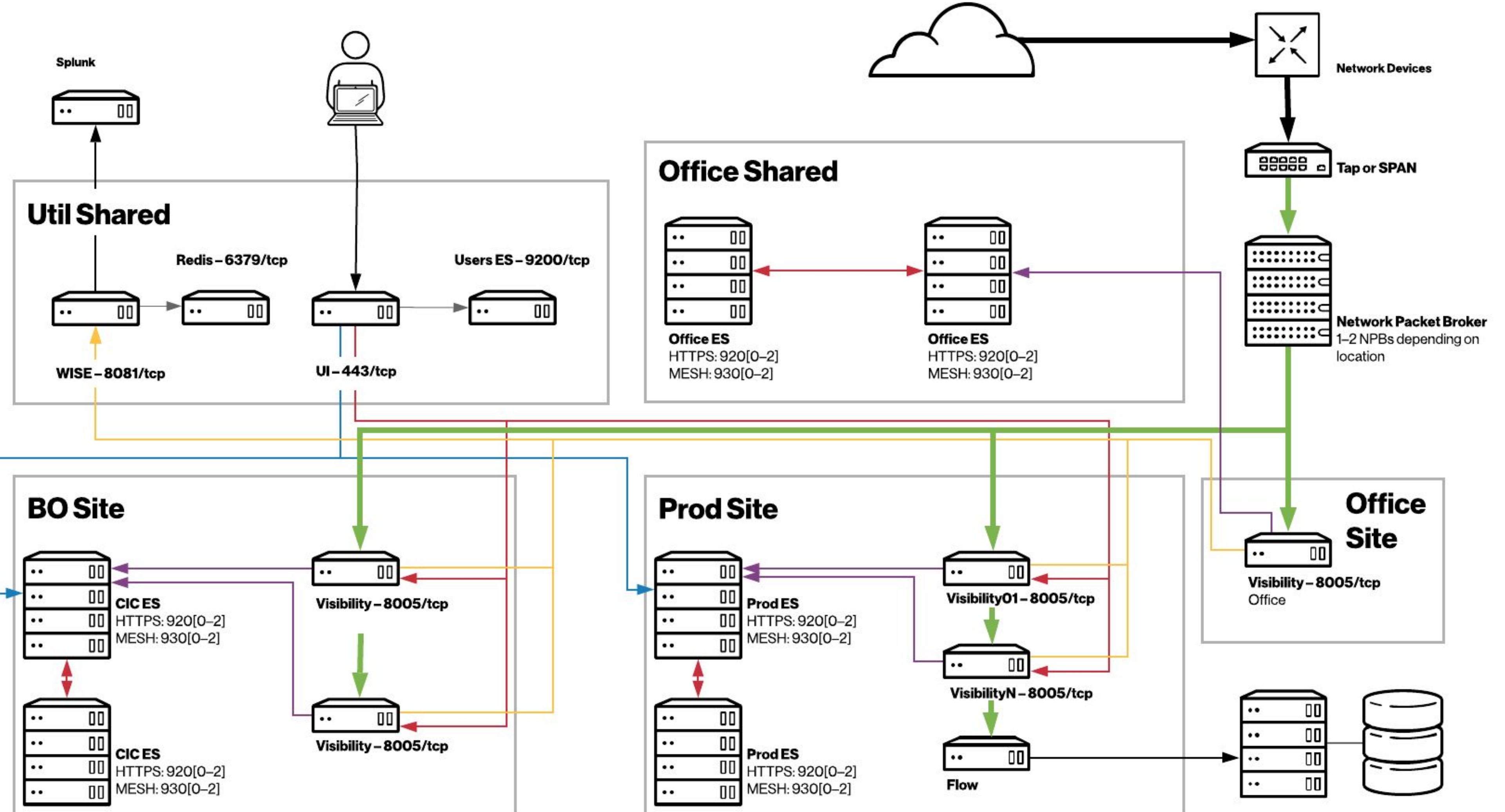
Deployments

Verizon Media 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 local egress
    - 10 VPN concentrators in subset of offices
    - Centralized Elasticsearch cluster
- **BO - Backoffice in a data center**
    - Each location with its own Elasticsearch cluster
- **Prod - Production traffic**
    - Each location with its own Elasticsearch cluster
    - Too much traffic (100Gps to Tbps) to capture everything
    - Some traffic we don’t want to capture

Design

- **Run all tools on each visibility box instead of specialized boxes**
- **Zeek (Bro), Suricata, Moloch and other tools**
- **Use an NPB to load balance traffic**
- **Focus on traffic to/from “internet” and cross “zones”**
- **Traffic reduction**
  - TLS only need handshake
  - Security cameras can produce a lot of traffic
  - Identify PCAP that isn’t needed or wanted

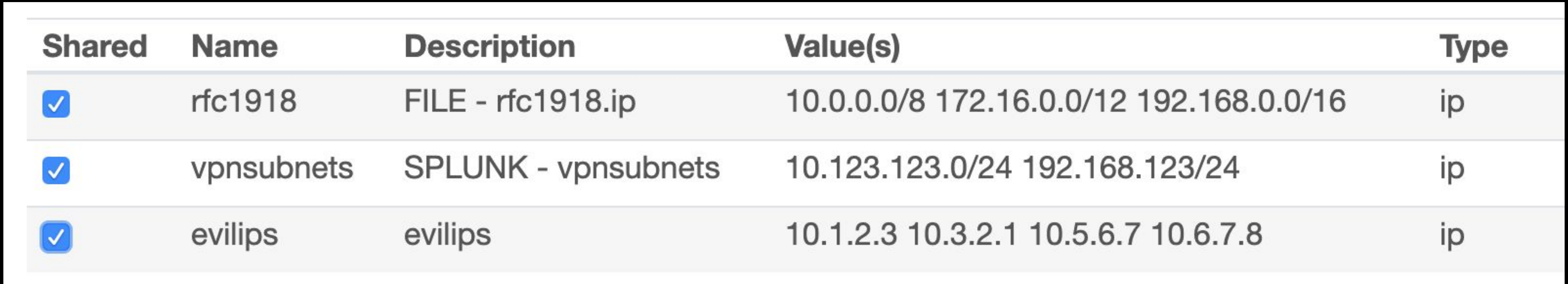

Splunk
Network Devices
Tap or SPAN
Network Packet Broker
1–2 NPBs depending on location
Util Shared
Redis – 6379/tcp
Users ES – 9200/tcp
WISE – 8081/tcp
UI – 443/tcp
Office Shared
Office ES
HTTPS: 920[0–2]
MESH: 930[0–2]
Office ES
HTTPS: 920[0–2]
MESH: 930[0–2]
BO Site
CIC ES
HTTPS: 920[0–2]
MESH: 930[0–2]
Visibility – 8005/tcp
Visibility – 8005/tcp
CIC ES
HTTPS: 920[0–2]
MESH: 930[0–2]
Prod Site
Prod ES
HTTPS: 920[0–2]
MESH: 930[0–2]
Visibility01 – 8005/tcp
VisibilityN – 8005/tcp
Flow
Prod ES
HTTPS: 920[0–2]
MESH: 930[0–2]
Office Site
Visibility – 8005/tcp
Office

Utility Boxes

- **Most secure - Iptables & network ACLS**
- **Load balance incoming requests**
- **WISE**
  - Talks to Redis Sentinel cluster
  - Fetches feeds from splunk, threatstream, ...
- **Moloch UI**
  - Apache Reverse Proxy
  - Use Apache module for authentication
- **Users Elasticsearch**
  - Shared Users ES
- **Suricata/Zeek sigs and rules repo**
  - Sensors periodically check if there are changes and reload
- **Sync Shortcuts from Splunk and files**

Shortcuts

Util boxes have files and splunk config they use to create Shortcuts across all our moloch clusters.



Shared	Name	Description	Value(s)	Type
☑	rfc1918	FILE - rfc1918.ip	10.0.0.0/8 172.16.0.0/12 192.168.0.0/16	ip
☑	vpnsubnets	SPLUNK - vpnsubnets	10.123.123.0/24 192.168.123/24	ip
☑	evilips	evilips	10.1.2.3 10.3.2.1 10.5.6.7 10.6.7.8	ip

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/BO: 7150S-24, 7280SE**
- **Production: 7508R 13RU, 6 power supplies, max 11,484W**

Why use a NPB?

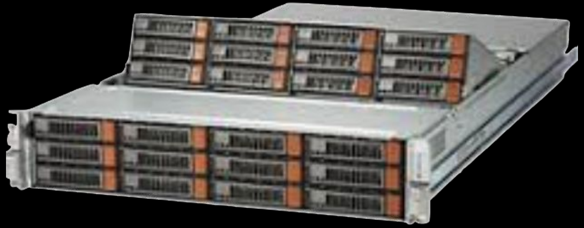
- **Easy to add Moloch capacity**
- **Allows the networking team and security team to act more independently**
    - Networking team can add more links at any time, just connect taps to NPB
    - The security team can add more tool capacity at any time, just connect tools to NPB
- **Move the traffic 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

- **Zeek is a memory/cpu hog**
- **Use AFPacket for everything**
    - requires a patch to Zeek
    - requires newish 3.x or 4.x Kernel
- **Want enough memory to potential run other tools and scanners in the future**
- **2RU for space considerations, however boxes are deeper**

Hardware Selected

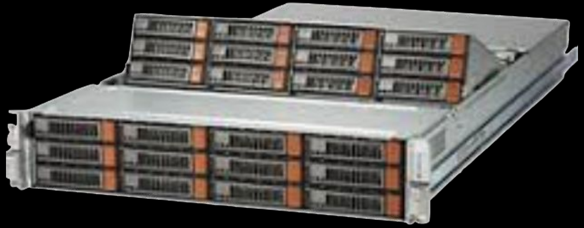
- **Keep number of configurations to a minimum**
- **Arista NPB**
- **Visibility boxes**
  - Supermicro 6028R-E1CR24L or Dell R740XD2
  - 24x10TB 128GB - Office, BO
  - 24x12TB 256GB - Prod
- **Moloches**
  - Used, most are 5+ years old
  - 4x10TB 128GB - Office, BO
  - 4x12TB 256GB - Prod
  - Replication of 1



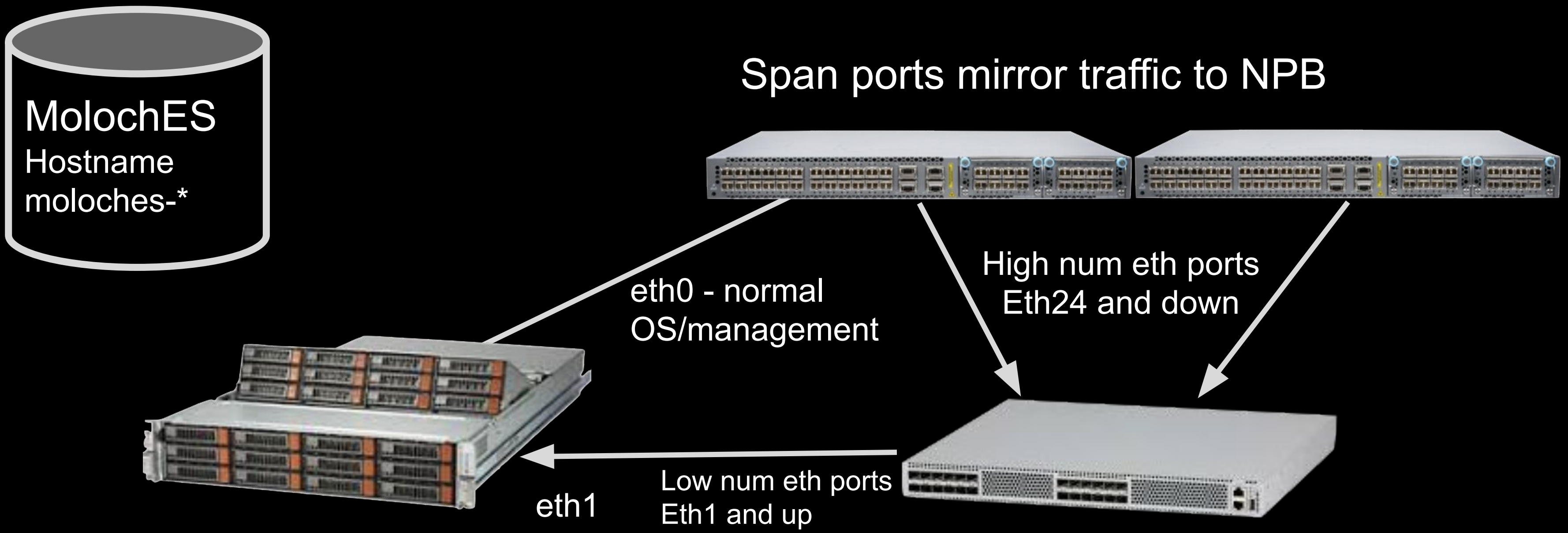
Visibility Cost Example

240TB Raid-6 Usable

12TB Drives	24	$400	$9,600
32GB DIMM	8	$300	$2,400
Box/MB/CPU/RAID	1	$7,000	$7,000
Sub Total			**$19,000**
Brandname			$5,000
Support	15%		$3,600
Total			**$27,600**



Office/BO Architecture



MolochES
Hostname
moloches-*

Span ports mirror traffic to NPB

eth0 - normal
OS/management

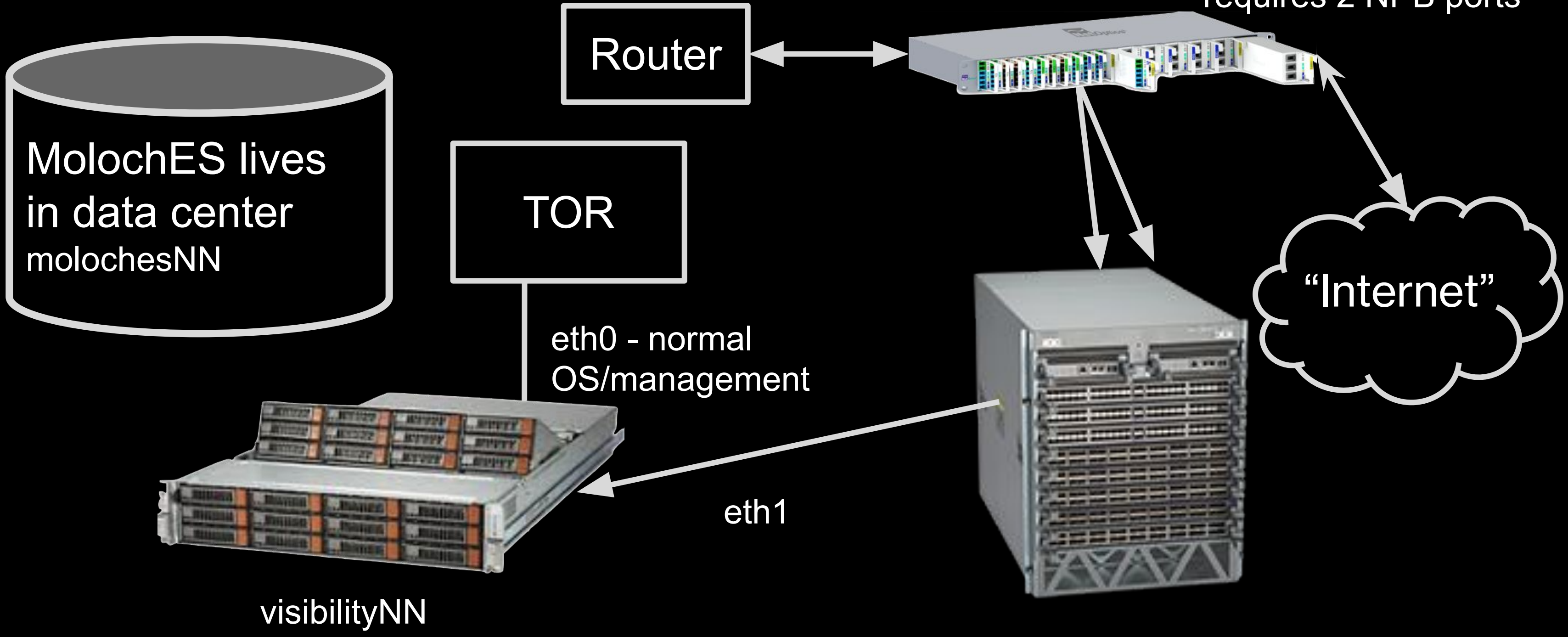
High num eth ports
Eth24 and down

eth1

Low num eth ports
Eth1 and up

Most sites only have 1 or 2 visibility servers
Hostname: visibilityNN

Prod Architecture



Reality

verizon media

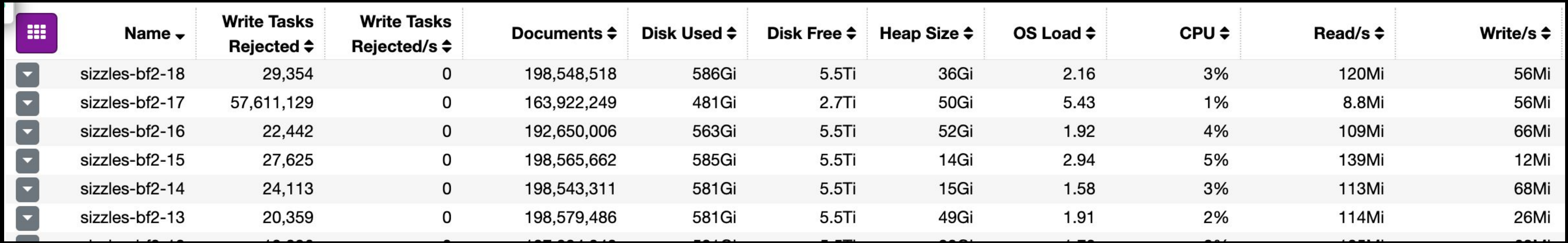
Things to watch for

- **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 6.8.2 or 7.3+**
- **Security - use iptables and/or Elasticsearch Auth**
- **Number of ACLs NPB can handle**

Write Tasks Rejected

Watch for ES nodes with high write tasks reject - sick node

- Ideally should be 0
- Check RAID battery
- Check RAID is in correct write mode
- Check for disk failure

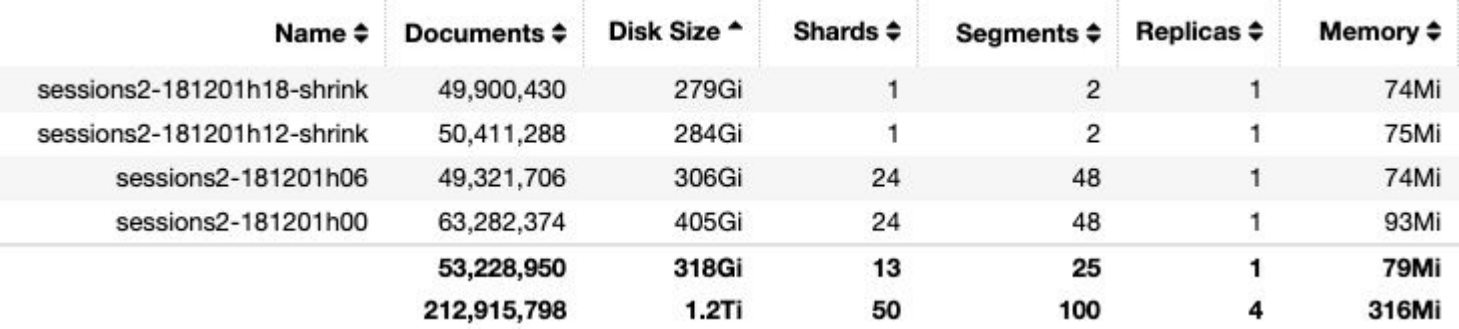


Name	Write Tasks Rejected	Write Tasks Rejected/s	Documents	Disk Used	Disk Free	Heap Size	OS Load	CPU	Read/s	Write/s
sizzles-bf2-18	29,354	0	198,548,518	586Gi	5.5Ti	36Gi	2.16	3%	120Mi	56Mi
sizzles-bf2-17	57,611,129	0	163,922,249	481Gi	2.7Ti	50Gi	5.43	1%	8.8Mi	56Mi
sizzles-bf2-16	22,442	0	192,650,006	563Gi	5.5Ti	52Gi	1.92	4%	109Mi	66Mi
sizzles-bf2-15	27,625	0	198,565,662	585Gi	5.5Ti	14Gi	2.94	5%	139Mi	12Mi
sizzles-bf2-14	24,113	0	198,543,311	581Gi	5.5Ti	15Gi	1.58	3%	113Mi	68Mi
sizzles-bf2-13	20,359	0	198,579,486	581Gi	5.5Ti	49Gi	1.91	2%	114Mi	26Mi

Too many shards

- Elasticsearch works best with lower shard counts
- Aim for 50G-150G per shard - remember replicas are counted in Disk Size column so /(replicas+1)

In this example h00 & h06 have 24 shards, which is way too many

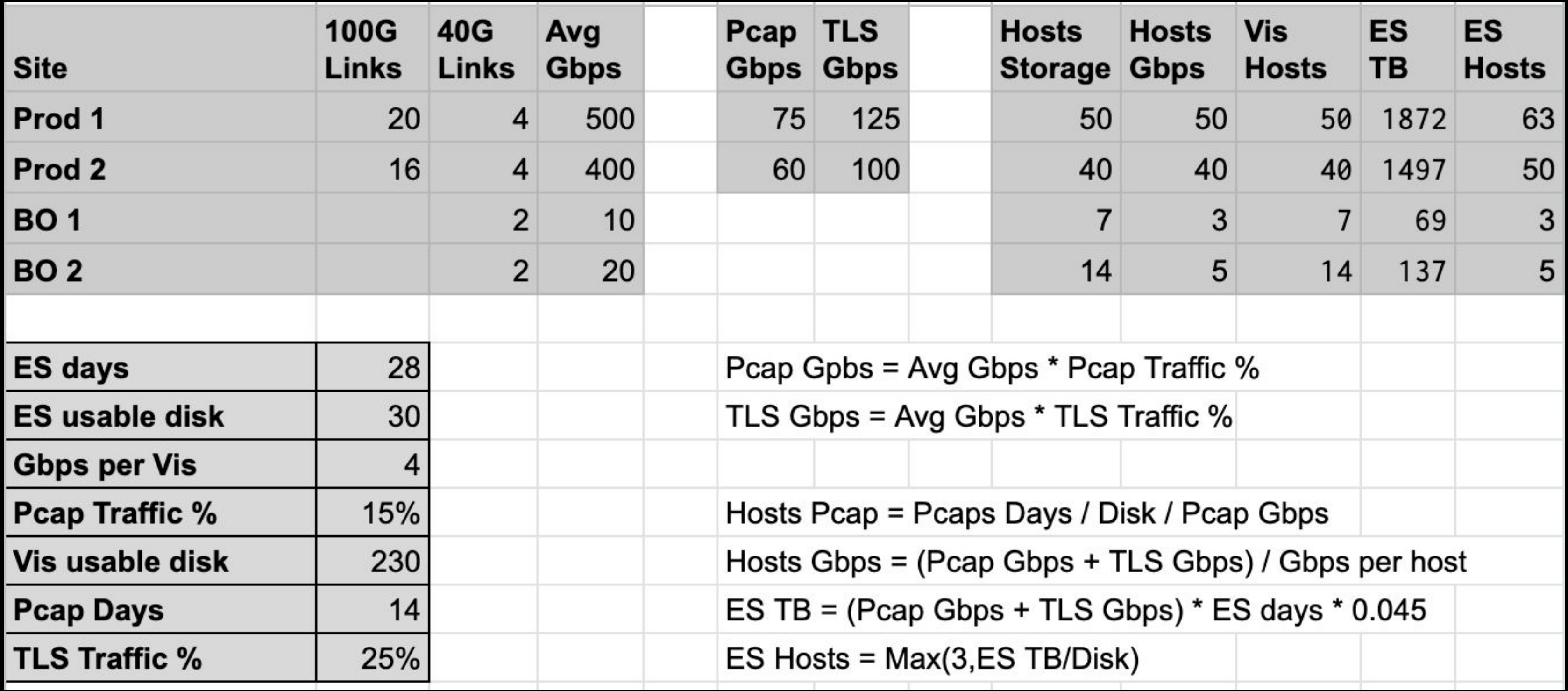


Name	Documents	Disk Size	Shards	Segments	Replicas	Memory
sessions2-181201h18-shrink	49,900,430	279Gi	1	2	1	74Mi
sessions2-181201h12-shrink	50,411,288	284Gi	1	2	1	75Mi
sessions2-181201h06	49,321,706	306Gi	24	48	1	74Mi
sessions2-181201h00	63,282,374	405Gi	24	48	1	93Mi
	53,228,950	**318Gi**	**13**	**25**	**1**	**79Mi**
	212,915,798	**1.2Ti**	**50**	**100**	**4**	**316Mi**

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
  - Moloch rules are used to not save pcap
  - NPB is used for aggregation
- **BO & 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 Storage	Hosts Gbps	Vis Hosts	ES TB	ES Hosts
Prod 1	20	4	500		75	125		50	50	50	1872	63
Prod 2	16	4	400		60	100		40	40	40	1497	50
BO 1		2	10					7	3	7	69	3
BO 2		2	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 %	25%

Pcap Gpbs = Avg Gbps * Pcap Traffic %

TLS Gbps = Avg Gbps * TLS Traffic %

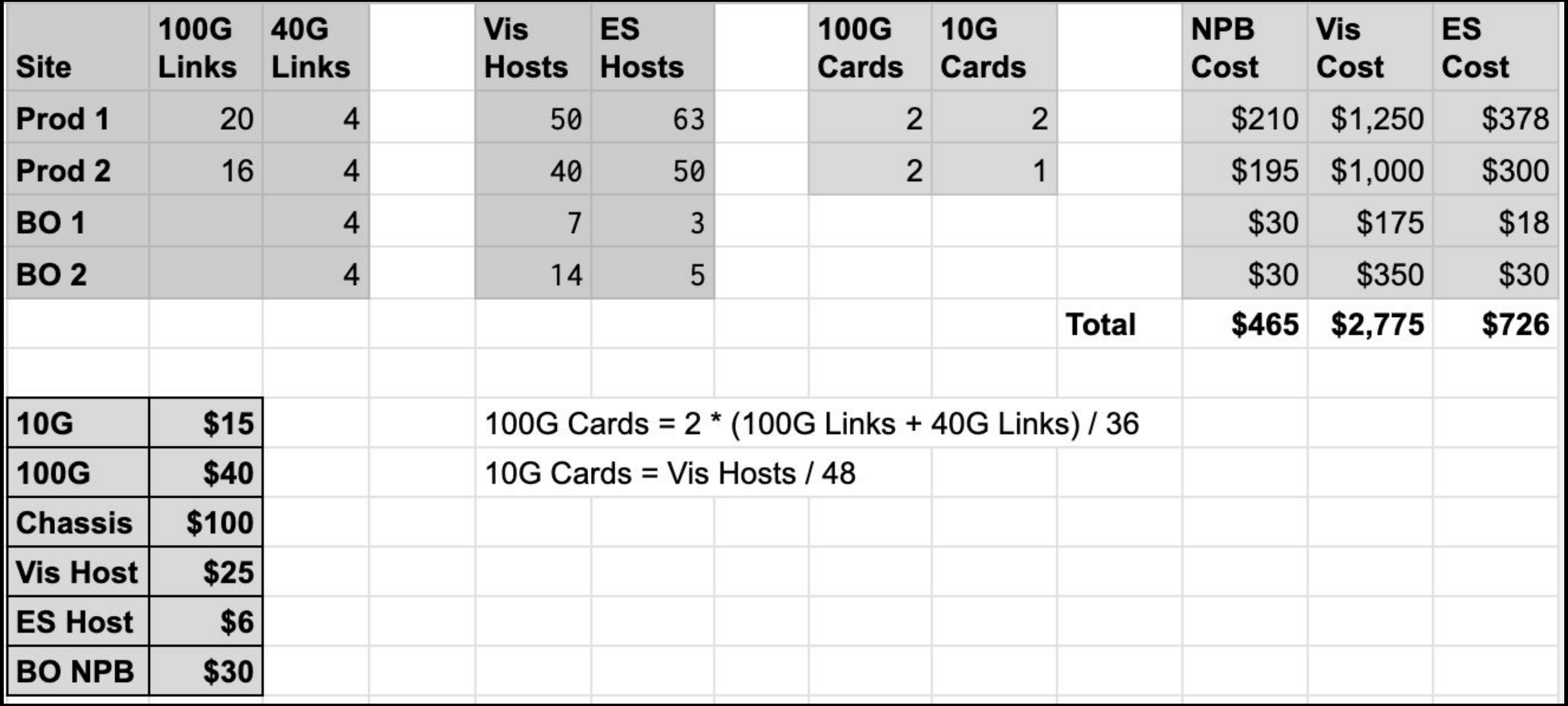
Hosts Pcap = Pcaps Days / Disk / Pcap Gbps

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



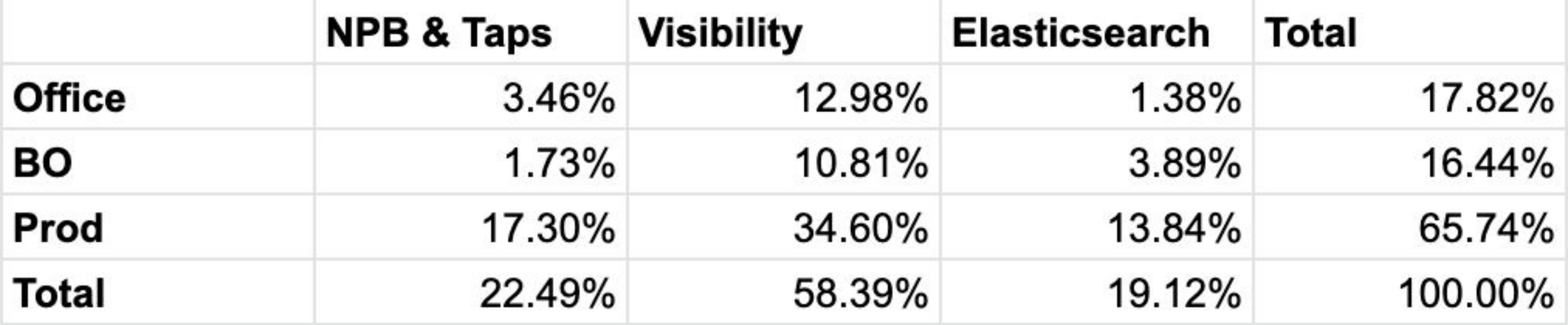
Site	100G Links	40G Links		Vis Hosts	ES Hosts		100G Cards	10G Cards		NPB Cost	Vis Cost	ES Cost
Prod 1	20	4		50	63		2	2		$210	$1,250	$378
Prod 2	16	4		40	50		2	1		$195	$1,000	$300
BO 1		4		7	3					$30	$175	$18
BO 2		4		14	5					$30	$350	$30
									Total	**$465**	**$2,775**	**$726**

10G	**$15**
100G	**$40**
Chassis	**$100**
Vis Host	**$25**
ES Host	**$6**
BO 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%
BO	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
```

Rules - Drop TLS after 20 packets

```
- name: "Drop tls"
   when: "fieldSet"
   fields:
     protocols:
     - tls
   ops:
     _maxPacketsToSave: 20
```

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

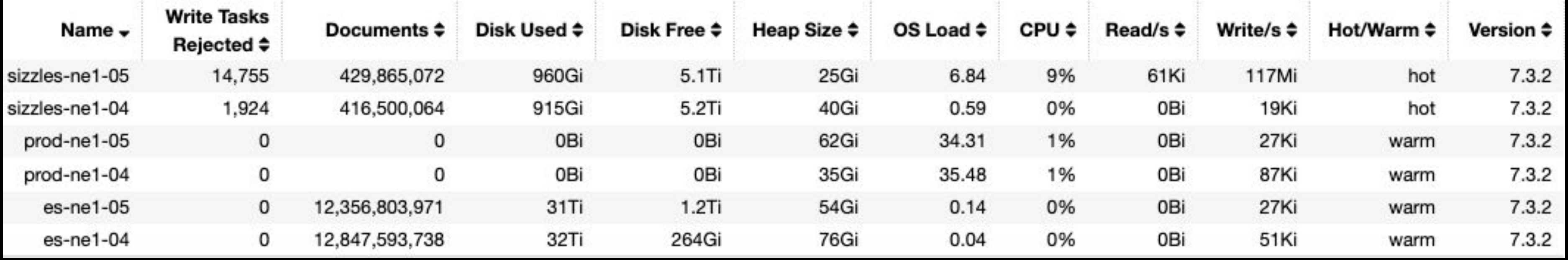
```
- name: "Drop tls to ips by hostname"
   when: "fieldSet"
   fields:
     host.http:
     - ad.doubleclick.net
     - foo.example.com
     protocols:
     - tls
   ops:
     _dontSaveSPI: 1
     _maxPacketsToSave: 1
     _dropByDst: 10
```

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 :) about 1.5 x Gbps
packetThreads=5

# Slightly increase the pcap write size
pcapWriteSize=4194304
```

Hot/Warm Config

- Use a few SSD boxes for HOT nodes
- Moloch supports basic config in db.pl and UI (ILM in future)
- Debate if force merge/shrink should be done on SSDs or SATA
- Still might need to run indexing with replica
- Naming boxes sizzles is tempting the overheating gods



Name	Write Tasks Rejected	Documents	Disk Used	Disk Free	Heap Size	OS Load	CPU	Read/s	Write/s	Hot/Warm	Version
sizzles-ne1-05	14,755	429,865,072	960Gi	5.1Ti	25Gi	6.84	9%	61Ki	117Mi	hot	7.3.2
sizzles-ne1-04	1,924	416,500,064	915Gi	5.2Ti	40Gi	0.59	0%	0Bi	19Ki	hot	7.3.2
prod-ne1-05	0	0	0Bi	0Bi	62Gi	34.31	1%	0Bi	27Ki	warm	7.3.2
prod-ne1-04	0	0	0Bi	0Bi	35Gi	35.48	1%	0Bi	87Ki	warm	7.3.2
es-ne1-05	0	12,356,803,971	31Ti	1.2Ti	54Gi	0.14	0%	0Bi	27Ki	warm	7.3.2
es-ne1-04	0	12,847,593,738	32Ti	264Gi	76Gi	0.04	0%	0Bi	51Ki	warm	7.3.2

Pcap Encryption at rest

Moloch Encryption	**Disk Encryption**
Can't use tools on files directly	Can use packet tools on file
File access isn't enough to copy data	File access is enough to copy data
Password in config file and DEK in ES	Password at boot or TPM
Requires ES	Self contained
Potentially Less Secure Encryption	Potentially More Secure Encryption
Just a config change	More complex to setup

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 configured KEK are stored in the moloch config.ini file

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

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

QUESTIONS?