Elasticsearch应用在数据中心的实时协议分析和安全威胁检测

@Zooboa

zooboa@gmail.com

数据中心面临的挑战

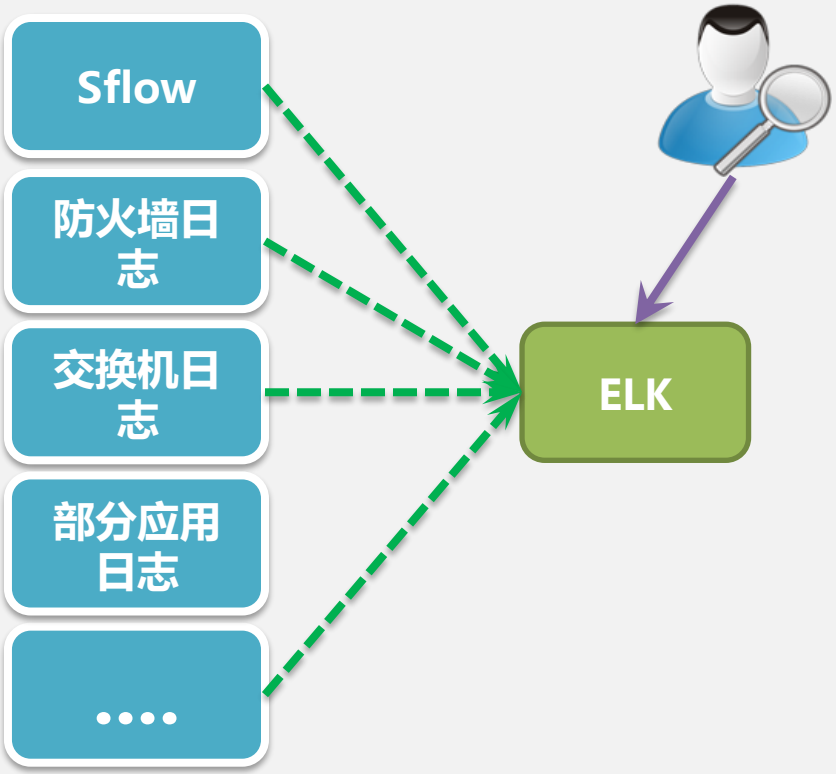
被DDOS攻击	• 网络瘫痪，大面积影响业务
植入后门发包	• 占用带宽资源，消耗成本
运营“黑盒子”	• 无法分辨“好人”、“坏人”
监控粒度粗	• 无法及时响应并定位事件

早期解决方案

原始方案

- Cacti 利用SNMP监控交换机出入口流量
- 交换机推送Sflow流量采样数据，使用Solarwids监控
- 遇到DDOS时，使用手动Sniffer抓包分析

第一期改造后

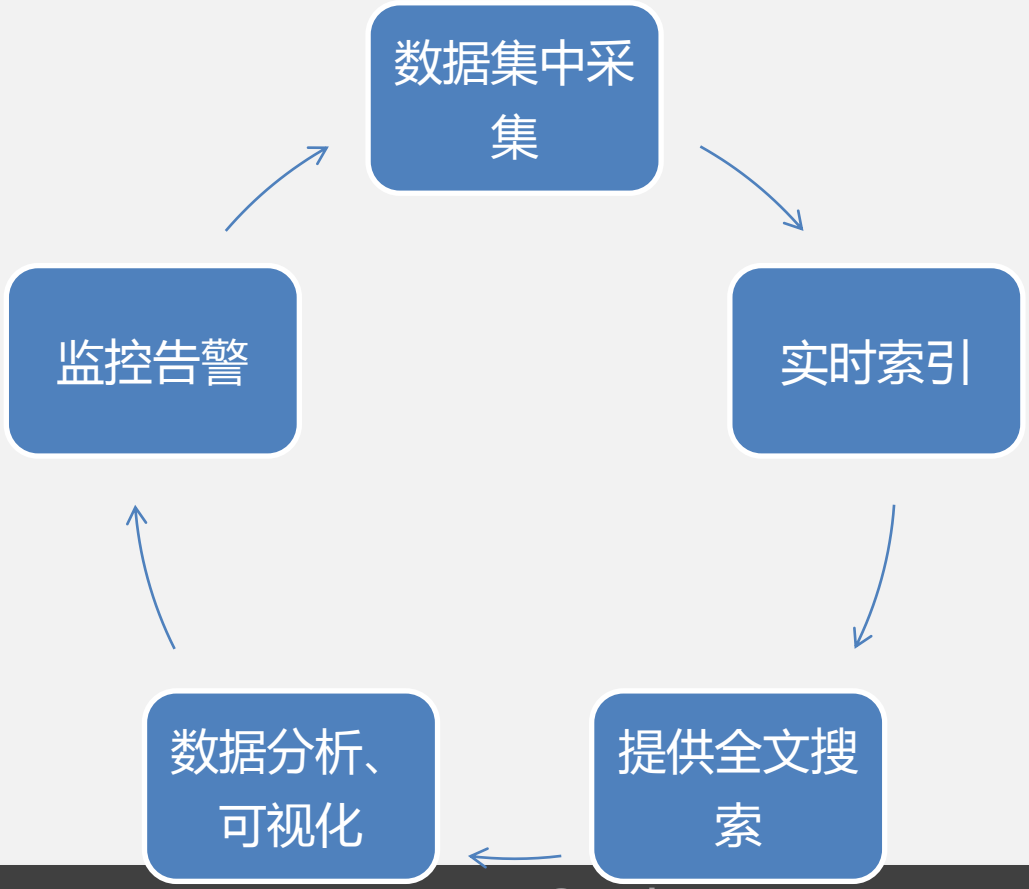


推送Netflow/Sflow 劣势

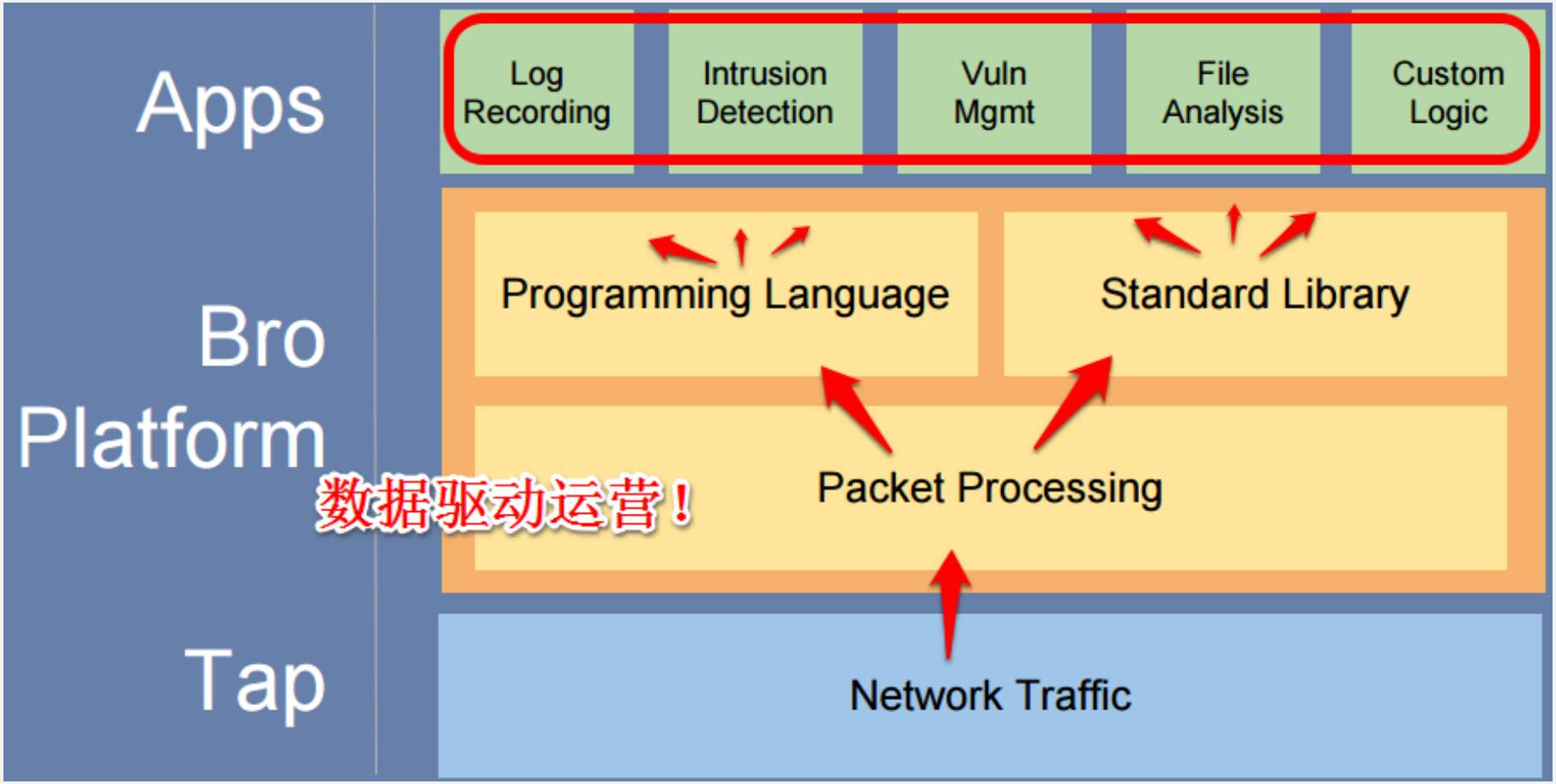
- 消耗路由器CPU资源
- 100-1000：1采样比，监测粒度粗
- 业务和应用识别依赖端口号，无法识别日新月异的业务类型

如何用数据驱动IDC运营

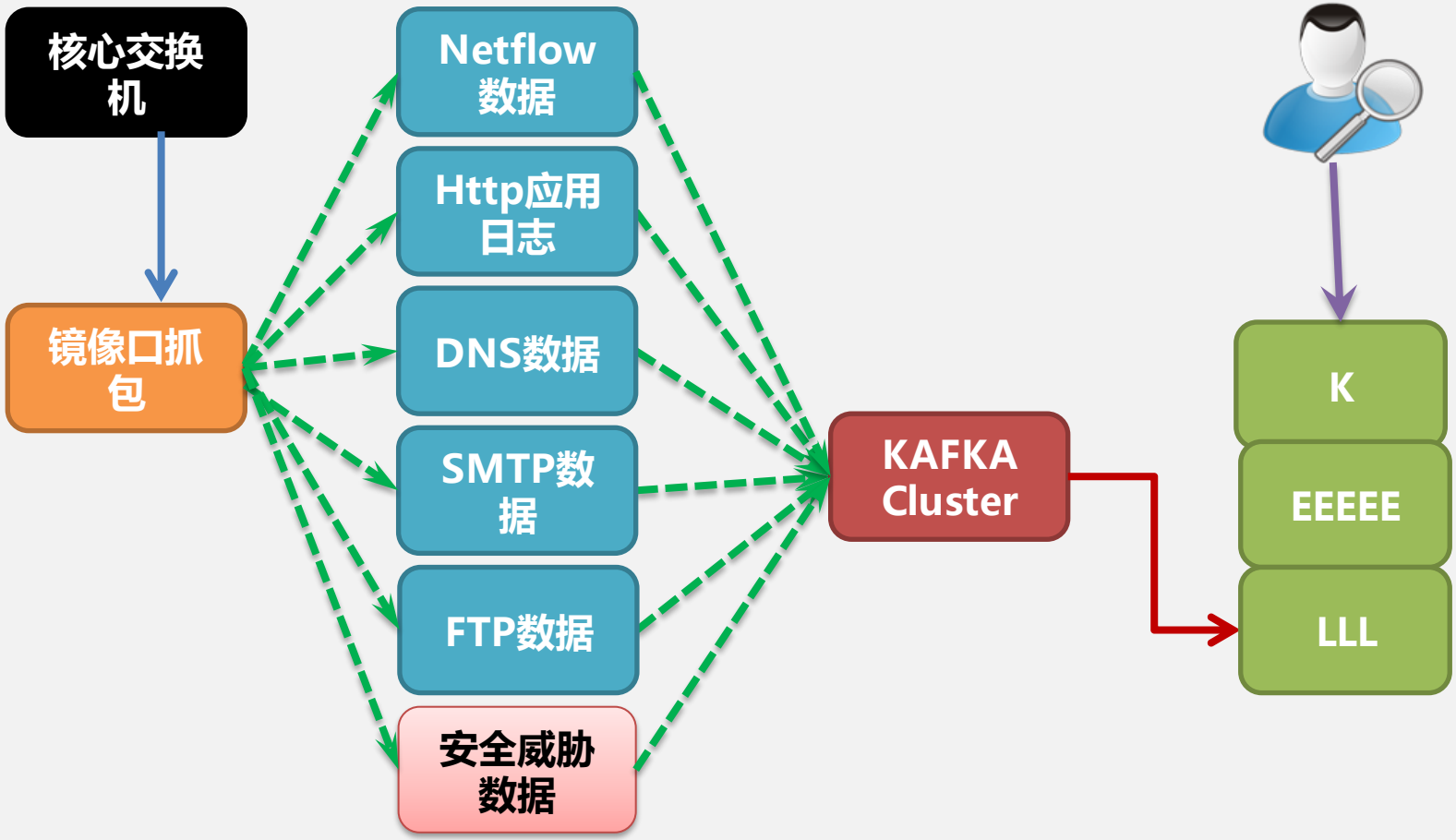
Network Security Monitoring （NSM）



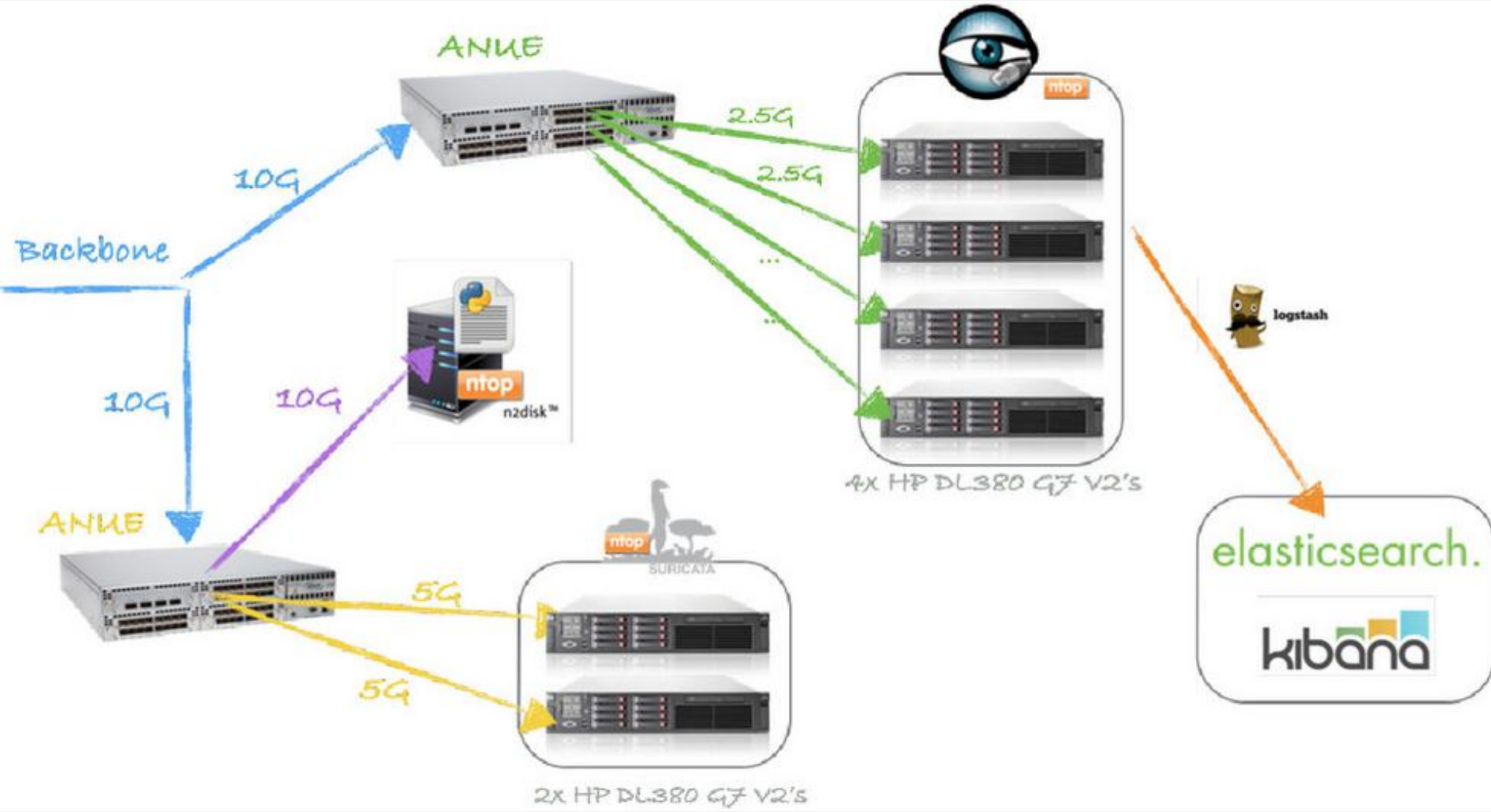
NSM架构设计


Apps
Log Recording
Intrusion Detection
Vuln Mgmt
File Analysis
Custom Logic
Bro Platform
Programming Language
Standard Library
Packet Processing
数据驱动运营！
Tap
Network Traffic

第二期改造后


核心交换机
镜像口抓包
Netflow数据
Http应用日志
DNS数据
SMTP数据
FTP数据
安全威胁数据
KAFKA Cluster
K
EEEEE
LLL

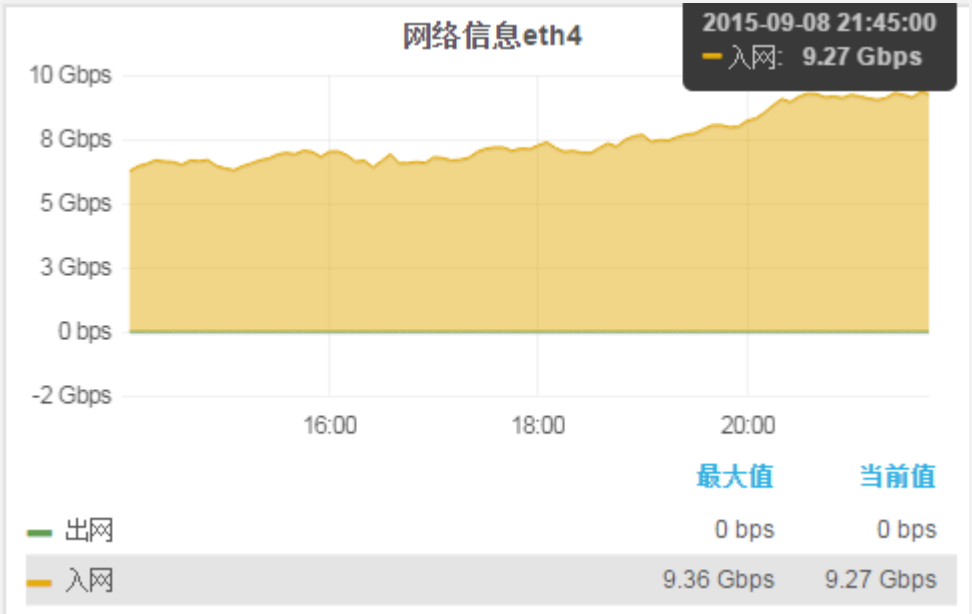
10G下的NSM


ANUE
10G
Backbone
10G
10G
ANUE
2.5G
2.5G
...
...
ntop
n2disk™
4x HP DL380 G7 V2's
logstash
elasticsearch.
kibana
5G
5G
ntop
SURICATA
2x HP DL380 G7 V2's

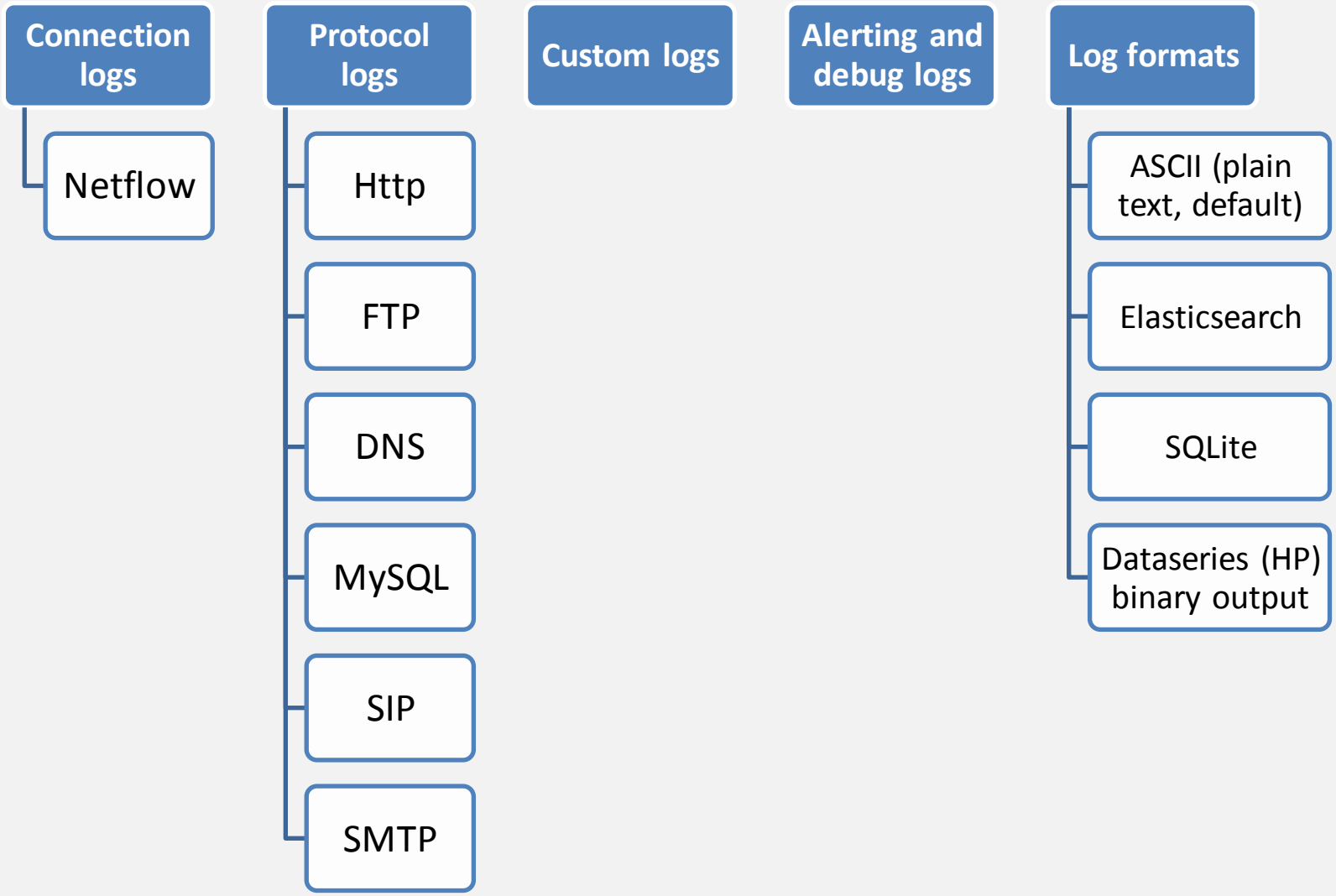
实际效果展示

NSM架构解析


SURICATA
ntop


网络信息eth4
2015-09-08 21:45:00
入网: 9.27 Gbps
10 Gbps
8 Gbps
5 Gbps
3 Gbps
0 bps
-2 Gbps
16:00
18:00
20:00
最大值
当前值
出网
0 bps
0 bps
入网
9.36 Gbps
9.27 Gbps

实时协议分析：Bro日志类型



Flow: 数据格式

字段名	样本	描述
ts	1425429336.809148	UNIX时间戳
uid	ClmuHr1gC6p76JbdVl	唯一ID
id.orig_h	10.1.1.1	源IP地址
id.orig_p	45191	源端口
id.resp_h	207.62.80.166	目的IP地址
id.resp_p	80	目的端口
proto	tcp	协议
service	http	应用服务识别
duration	0.023945	时长
orig_bytes	351	发起字节
resp_bytes	9886	响应字节
history	ShADadfF	状态历史

实时安全威胁检测引擎

Suricata:

- 多线程威胁检测引擎
- Emerging Threats威胁库
- 基于特征

BRO：

- 多节点可扩展分布式引擎
- Intelligence framework
- 基于行为和外部威胁名单库

Suricata Today

- Apply tens of thousands of attack patterns to your traffic
- Detect protocol anomalies in HTTP and others
- Extract files from HTTP and SMTP
- Fast, scalable and stable
- IP reputation
- Lua scripting for advanced detection logic
- Many helpful additions for researchers
- Netflow output (JSON)

实时流量 +ELK + VirusTotal

```
if ( [event_type] == "fileinfo" and [fileinfo][filename] =~
/(?i)\.(doc|pdf|zip|exe|dll|xls|ppt)/ ) {
        virustotal {
            apikey => '77f3d2ef83fc0db26447377cb40c9ce'
            field => '[fileinfo][md5]'
            lookup_type => 'hash'
            target => 'virustotal'
        }
```

virustotal.positives	0
virustotal.resource	7b1a4625706dbc3808a142e43b4f34e
virustotal.response_code	1
virustotal.scan_date	2014-04-10 08:02:14
virustotal.scan_id	8e695613b16e6fa5784663aac72ac3271 839a3b7a-1397116934
virustotal.scans.AVG.detected	false
virustotal.scans.AVG.result	-
virustotal.scans.AVG.update	20140409
virustotal.scans.AVG.version	13.0.0.3169
virustotal.scans.Ad-Aware.detected	false
virustotal.scans.Ad-Aware.result	-
virustotal.scans.Ad-Aware.update	20140410
virustotal.scans.Ad-Aware.version	12.0.163.0
virustotal.scans.AegisLab.detected	false
virustotal.scans.AegisLab.result	-

7b1a4625706dbc3808a142e43b4f34e7

https://www.virustotal.com/file/8e695613b16e6fa
c72ac32719ba37c205639ad5439fc18d839a3b7a/analys
6934/

构建10G+ NSM的几个关键点

1、抓包网卡
2、内核优化
3、驱动与rss
4、PF-Ring_zc
5、ntop、nprobe、ndpi
6、跨数据中心es

流量抓包与网卡

流量分发	主机层分发	引擎	操作系统
• Arista • Brocade • Endace • Gigamon • OpenFlow / SDN	Myricom 10GPCIE28C22 + sniffer 10G drivers • **PF_RING_zc + Intel82599EB** • Packet Bricks + netmap • Endace DAG	• BRO • Suricata • Snort	• Debian • Centos • FreeBSD

ELK部分的关键点

1、用Logstash Kafka input接收数据

2、数据量大，处理结构复杂时：

➔预设Kafka分区

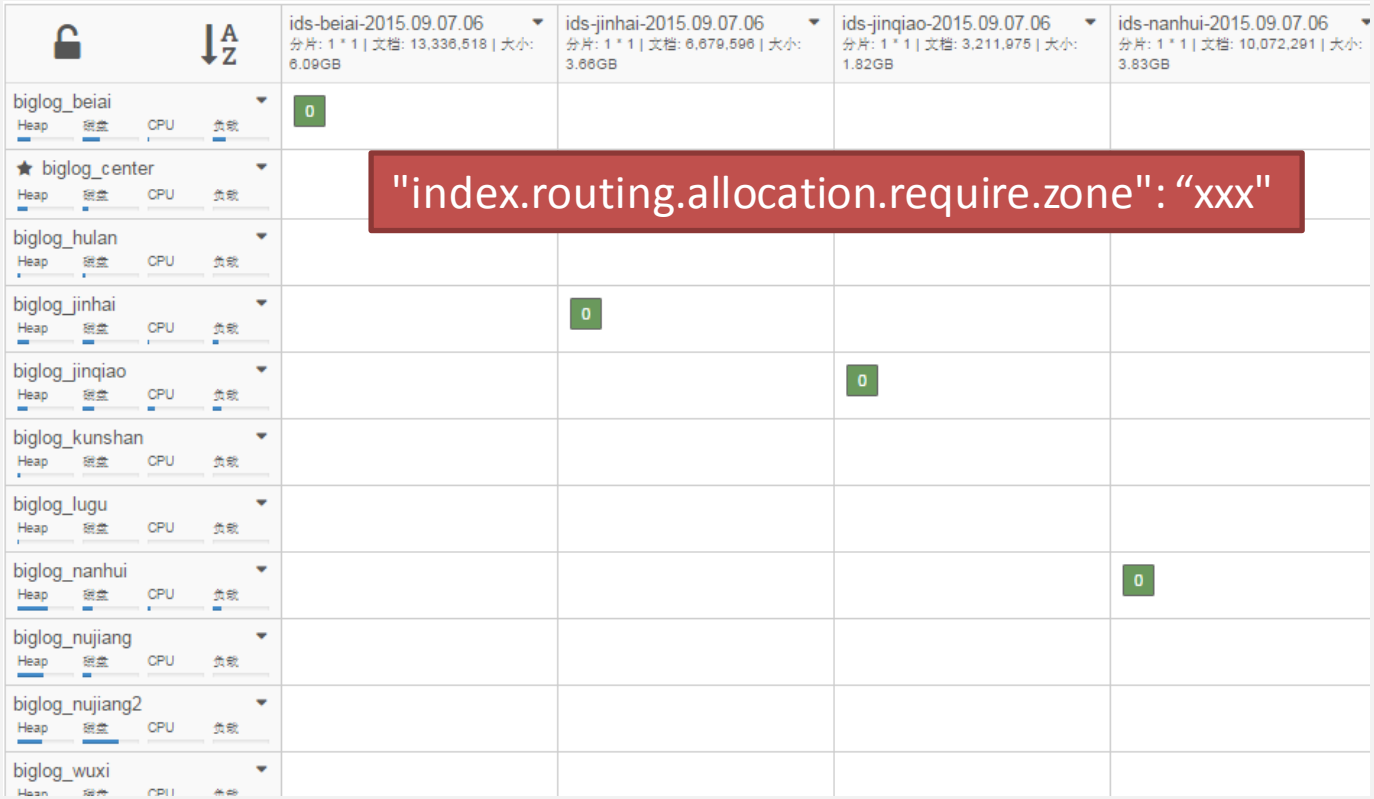
➔开启多个Logstash实例，分别读取Kafka分区数据

➔分别写入不同es节点

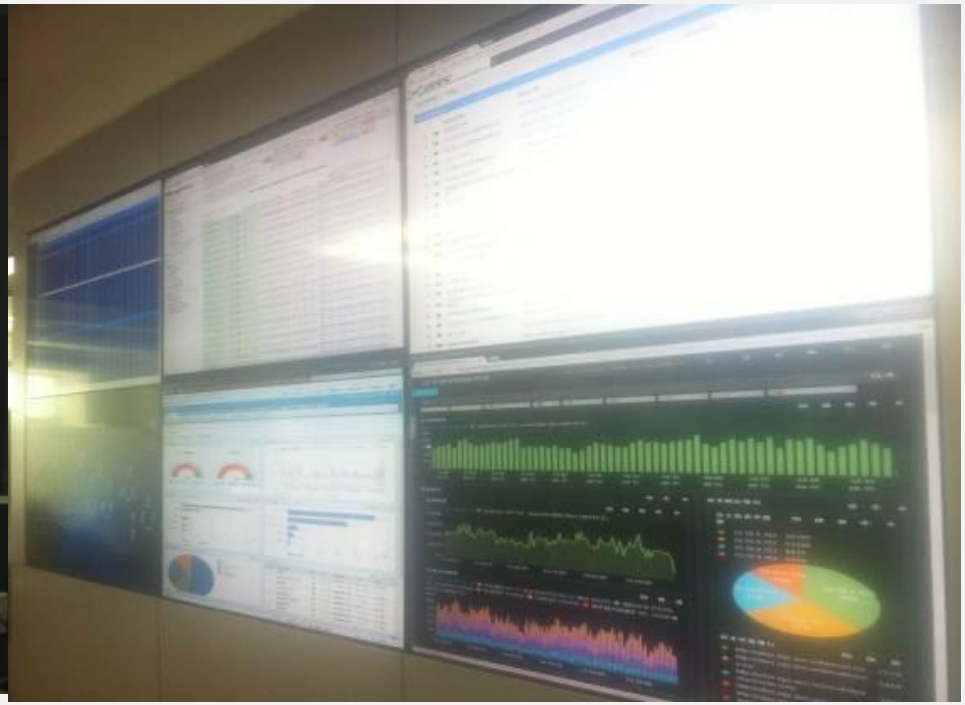
3、多集群互联

跨数据中心es集群

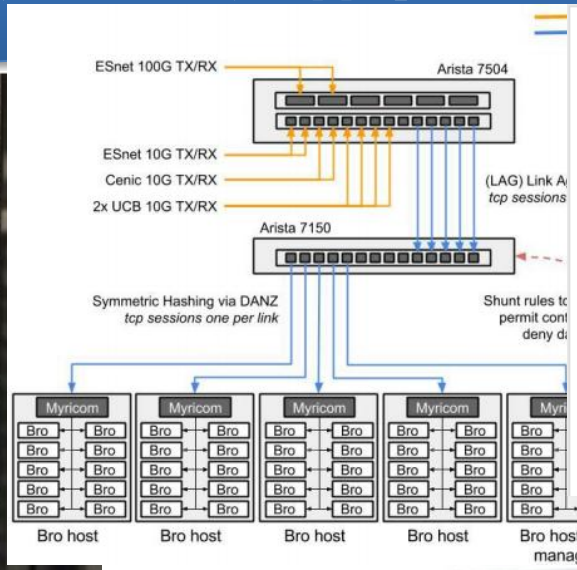
cluster.routing.allocation.awareness.attributes: zone
cluster.routing.allocation.awareness.force.zone.values: xxx
node.zone: xxx

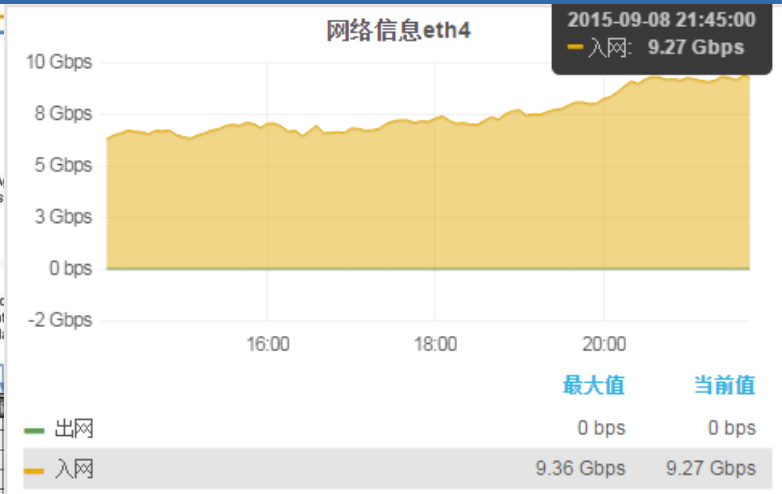


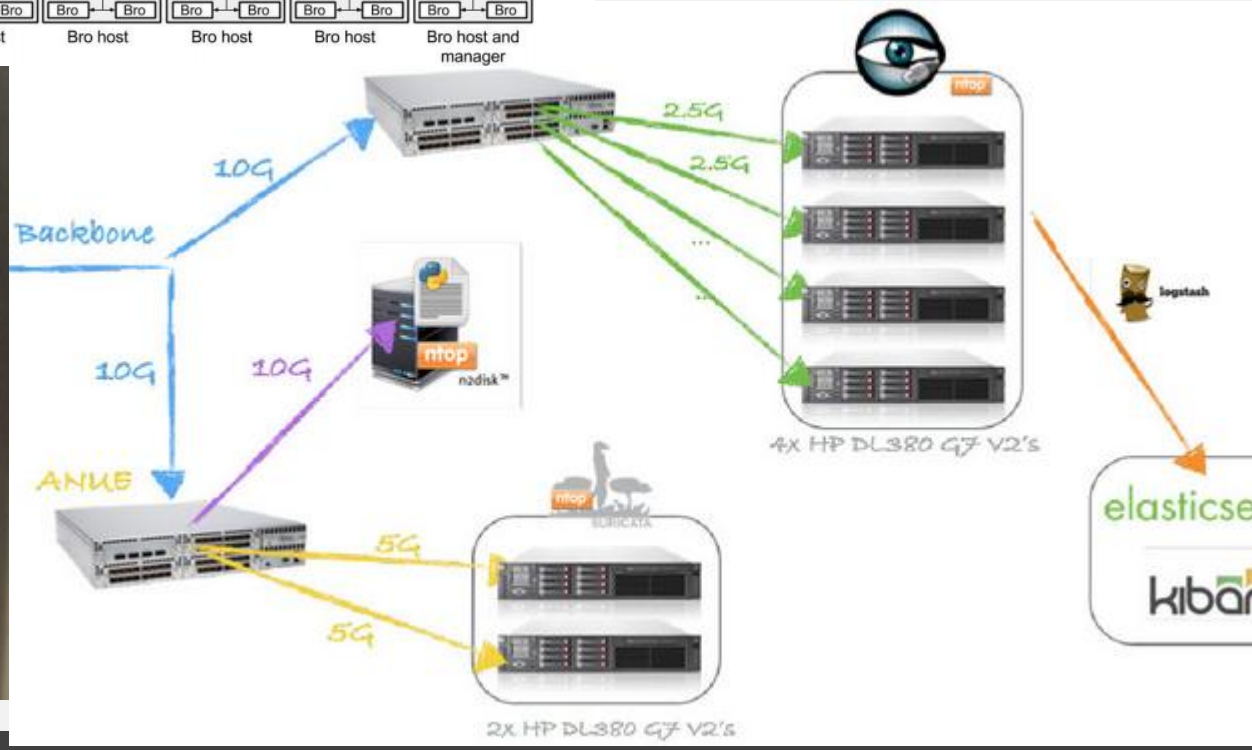
10G NSM平台样例



万兆 实时 安全大数据架构







Thanks

@Zooboa