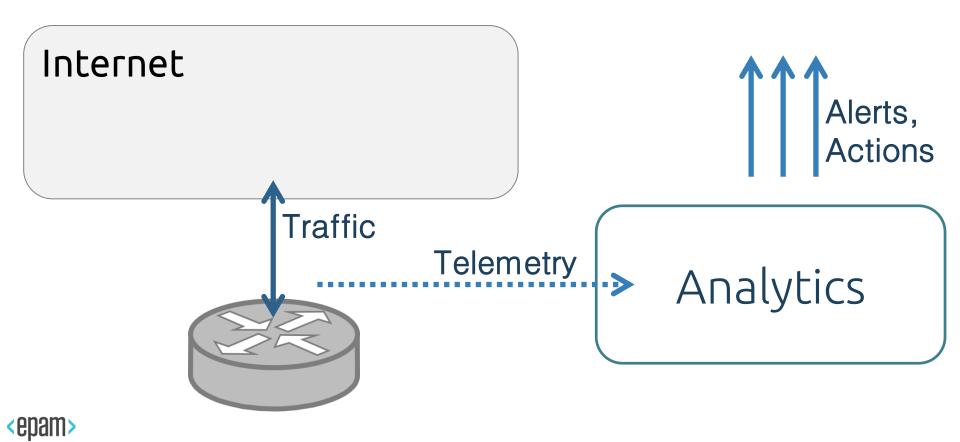


Data Science for Network Security

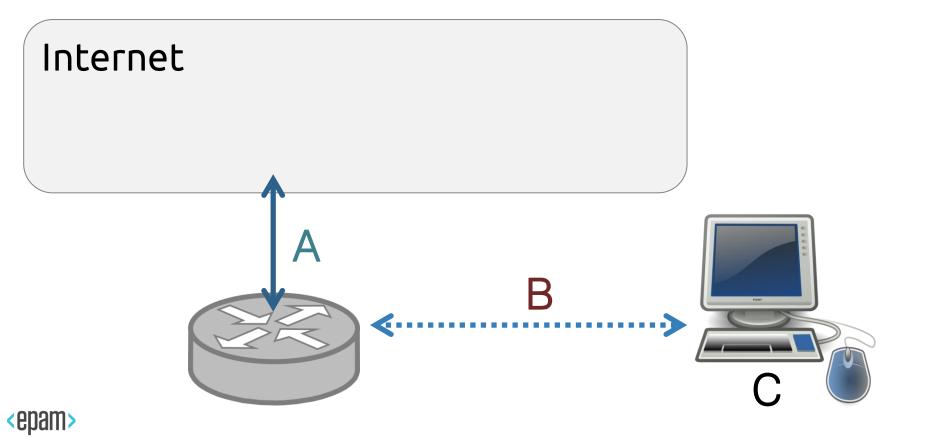
Dmitry Orekhov

Collecting Data

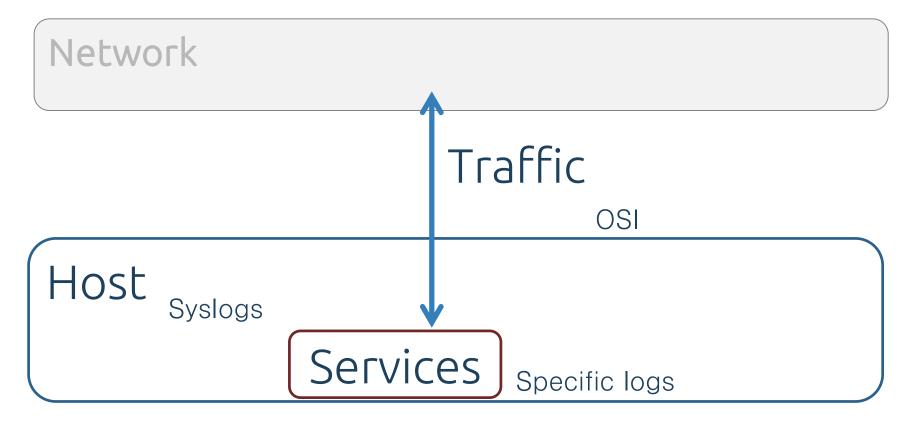
Data Flow



Sensors: Vantage



Sensors: Domain





Transport: NetFlow

- NetFlow is a traffic summarization standard developed by Cisco Systems and originally used for network services billing.
- The heart of NetFlow is the concept of a flow, which is an approximation of a TCP session.
- Plenty of Open Source
 implementations

Header	NetFlow Version 9 Header: 32 bits —>				
First Template FlowSet	Version 9	Count = 4 (FlowSets)			
Template Record	System Uptime				
First Record FlowSet (Template ID 256)	UNIX Seconds				
First Data Record	Package Sequence				
Second Data Record	Source ID				
Third Data Record	Template FlowSet: 16 bits ->		19	← Data FlowSet: 32 bits -	
Second Template FlowSet	Flows	Set ID - 0	it ID - 0		Length =
Template Record	Length =	28 bytes		ID = 256	64 bytes
Template Record	Template	Template ID = 256		192.168.1.12	
Second Record FlowSet	Field Count = 5			10.5.12.254	
(Template ID 257)	IPv4_SRCAD	DR (0x0008)		192.10	58.1.1
Data Record	Length = 4		->	5009	
Data Record	IPv4_DSTADDR (0x000C)			5344385	
Data Record	Length = 4			192,168,1,27	
Data Record	IPv4_NEXT_HDP (0x000E)			10.5.12.23	
	Length = 4			192.168.1.1	
	PKTS:_32(0x0002)			748	
	Length = 4			388964	
	BYTES:_32(0x0001)				
	Length = 4		1	10.5.12.65	
			2	10.5.	Server of the
				192.10	
				65	3
			14	60	3+

Transport: Nmsg

- Developed by Farsight Security for transporting network packets, particulary – DNS
- Low-latency and compactness
- Support binary and presentation forms; protobuf for protocols
- Open impelentation (GNU)

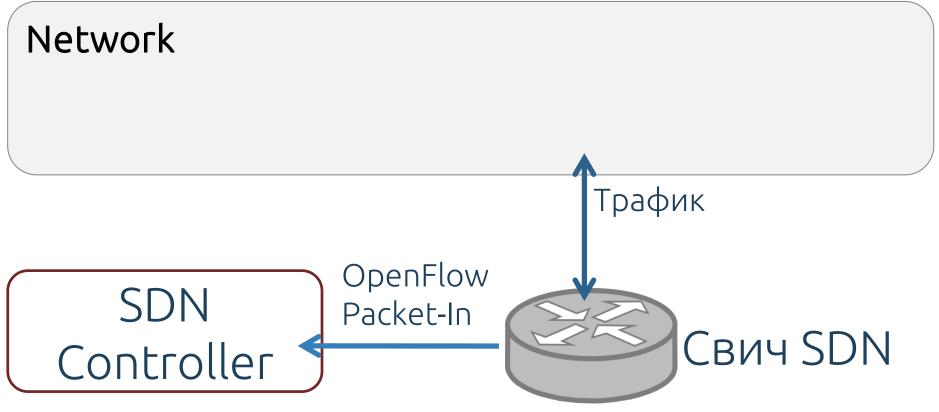
Nmsg

DNSQr

DNS Payload



Transport: OpenFlow





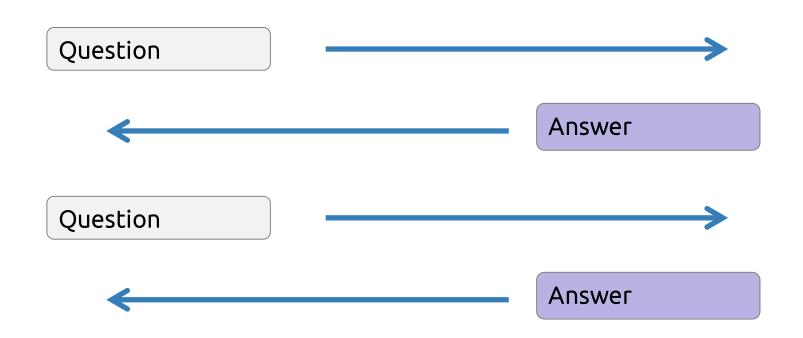
Use Case

DNS Tunneling



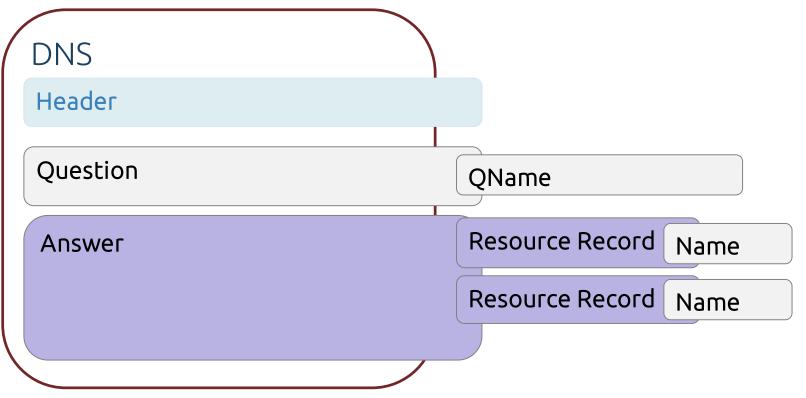








DNS Message: Big Picture







0 A	8 I	16 Q	24 Y
1 B	9 J	17 R	25 Z
2 C	10 K	18 S	26 2
3 D	11 L	19 T	27 3
4 E	12 M	20 U	28 4
5 F	13 N	21 V	29 5
6 G	14 O	22 W	30 6
7 H	15 P	23 X	31 7



Here you are: Tunnel TCP Packet TCP Packet TCP Packet SSH Tunnel TCP Packet TCP Packet 🗲 TCP Packet ► 0 **DNS** Packet **DNS** Packet **DNS** Packet •••



Discovery methods

Payload analysis

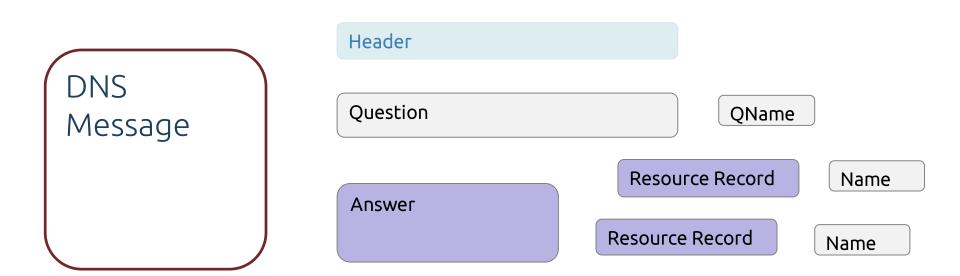
- Size of request and response
- Entropy of hostnames
- Statistical Analysis
- Uncommon Record Types

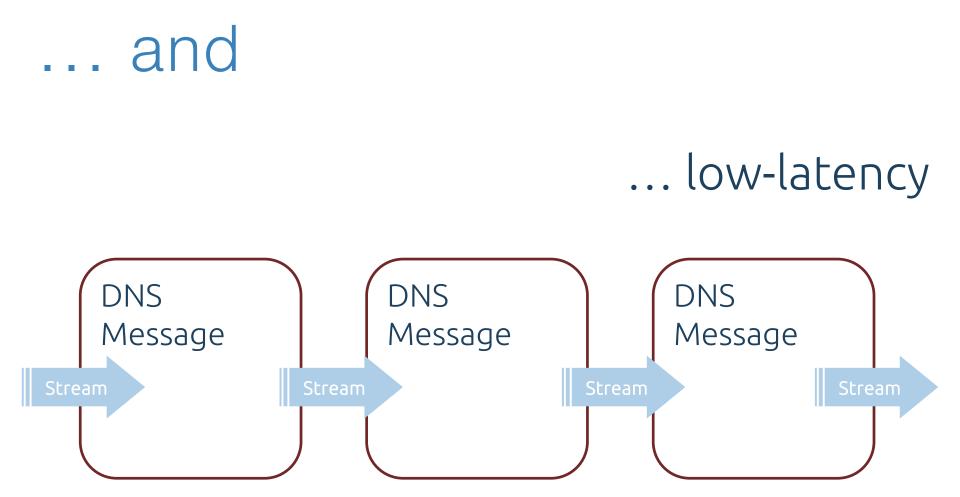
Traffic Analysis

- Volume of DNS traffic per IP address
- Volume of DNS traffic per domain
- Number of hostnames per domain
- Geographic location of DNS server
- Domain history
- Orphan DNS requests

Effective Intrusion Detection

... needs deep packet inspection







Solution: Lambda Architecture

Online model

Streaming Processing (Spark Streaming)





Algorithms

Incremental algorithms Outlier Detection

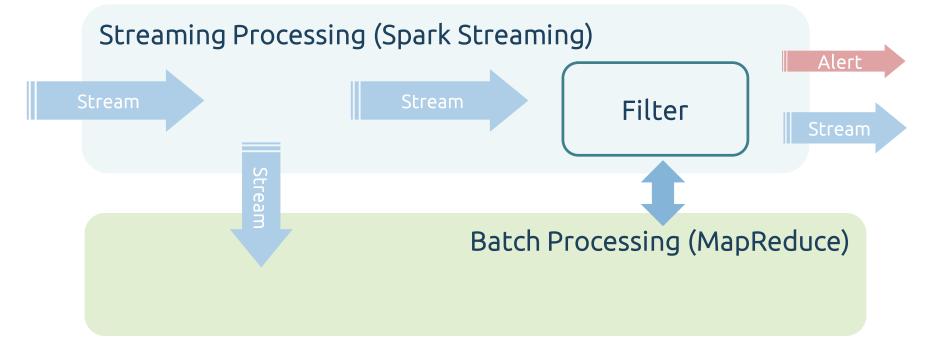
- Median Absolute Deviation, MAD
- Standard Deviation from average
- Standard Deviation from Moving Average

Потоковая классификация

- Incremental decision tree
- Hoeffding Tree (VFDT)
- Half-Space Trees



Offline model





Offline Model – Алгоритмы

<u>Analyze entire data set at once</u> Hypothesis tests

- Simple outlier detection far a period
- Statistical criteria
- Kolmogorov-Smirnov test

Decision Trees

Auto Regressive (AR) Moving Average (MA)



Architecture

