

# ENTRADA: Enabling DNS Big Data Applications

Giovane C. M. Moura - SIDN | NMRG @ IETF96

Berlin, Germany



# What if...

You have many TB's of network data?

And you want to:

1. Store it efficiently
2. Query it efficiently (SQL with interactive response times)
3. Quickly test a large number of hypotheses on your data
4. Continuously keep adding new data

# You could...

1. Convert pcap to text format like csv and use Linux utilities
2. Run Hadoop MapReduce jobs on csv/pcap
3. Store it in a RDBMS
4. ...

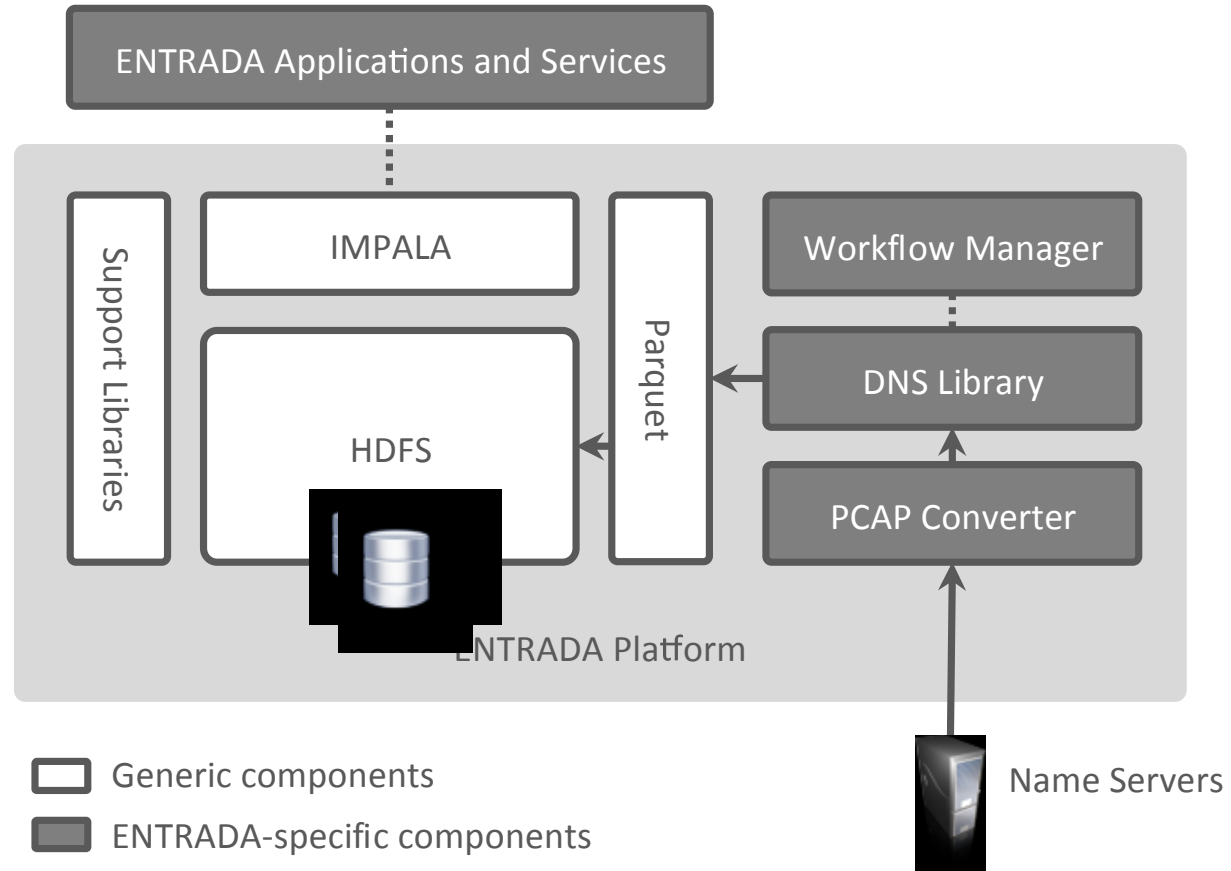
With most options it will be hard to scale and deliver interactive response times

# What to do?

- Build your own data stream warehouse (DSW)
- ENTRADA is our open source Hadoop-based DSW (**[entrada.sidnlabs.nl](http://entrada.sidnlabs.nl)**)
- Analyze 50TB of converted pcap data in under 3.5 minutes using a small cluster
- Our main use case: network (DNS, TCP/IP, ICMP) analytics

# ENTRADA

## ENhanced TOp-Level DOmain REsilience through ADvanced DAta ANalysis



# ENTRADA@SIDN

- We are the TLD registry of the Netherlands (.nl)
- Use ENTRADA to further increase security and stability
- Operational for over 2 years
- Capturing data from .nl name servers
- 160 billion rows (DNS query+response tuple), 21 TB of data

# More ENTRADA details

For design choices and a performance evaluation, see our 2016 NOMS paper:

*“ENTRADA: a High-Performance Network Traffic Data Streaming Warehouse”*,  
IEEE/IFIP Network Operations and Management Symposium 2016 (NOMS 2016),  
Istanbul, Turkey

See: <https://www.sidnlabs.nl/publications>

# Example Use Cases

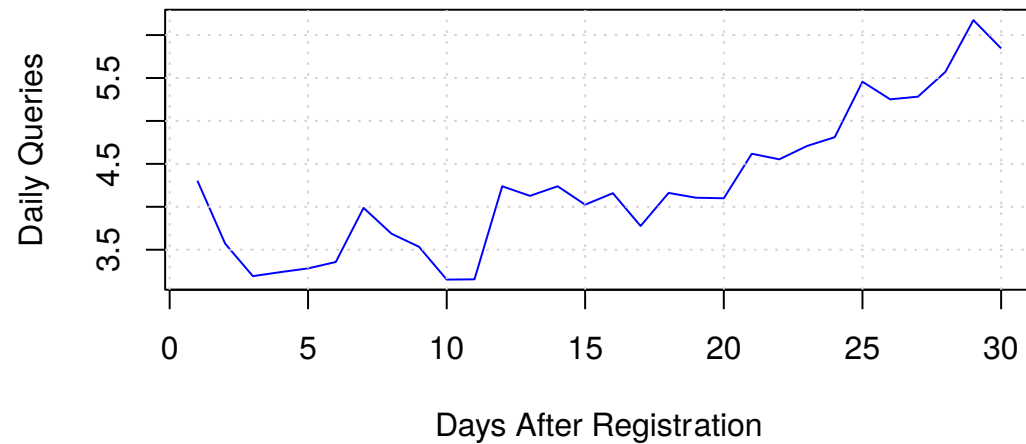
- Statistics (**stats.sidnlabs.nl**)
- Scientific research
- Insight for DNS operators
- **Malicious domain detection**
- **Botnet client detection**
- **Measuring uptake of email security**



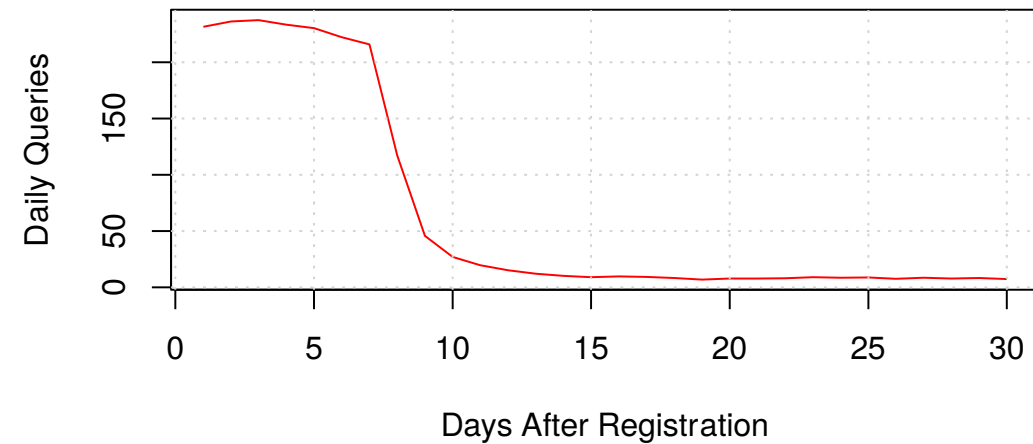
# Malicious Domain Detection (1/2)

**Observation:** New phishing domains have distinct query patterns

**Random Sample Jan—Mar, 2015**



**Phishing**

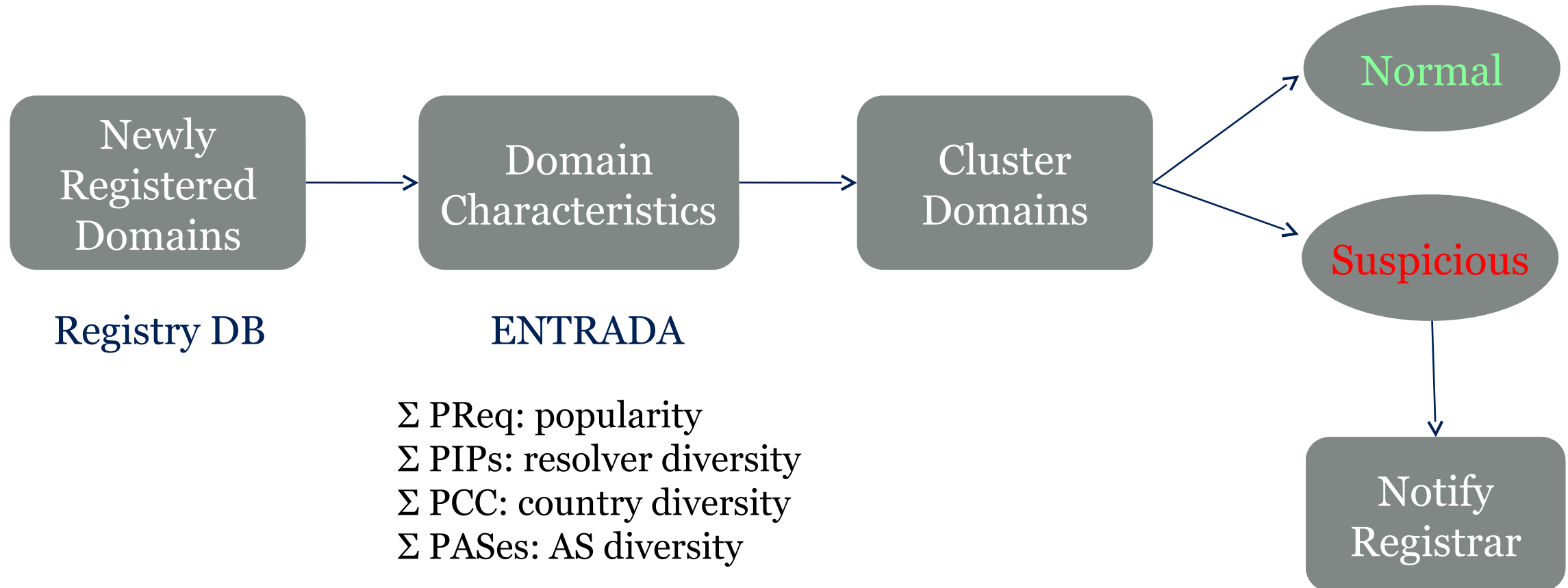


G. Moura, M. Müller, M. Wullink, and C. Hesselman, “nDEWS: a New Domains Early Warning System for TLDs”, IEEE/IFIP International Workshop on Analytics for Network and Service Management (AnNet 2016), <https://www.sidnlabs.nl/publicaties>

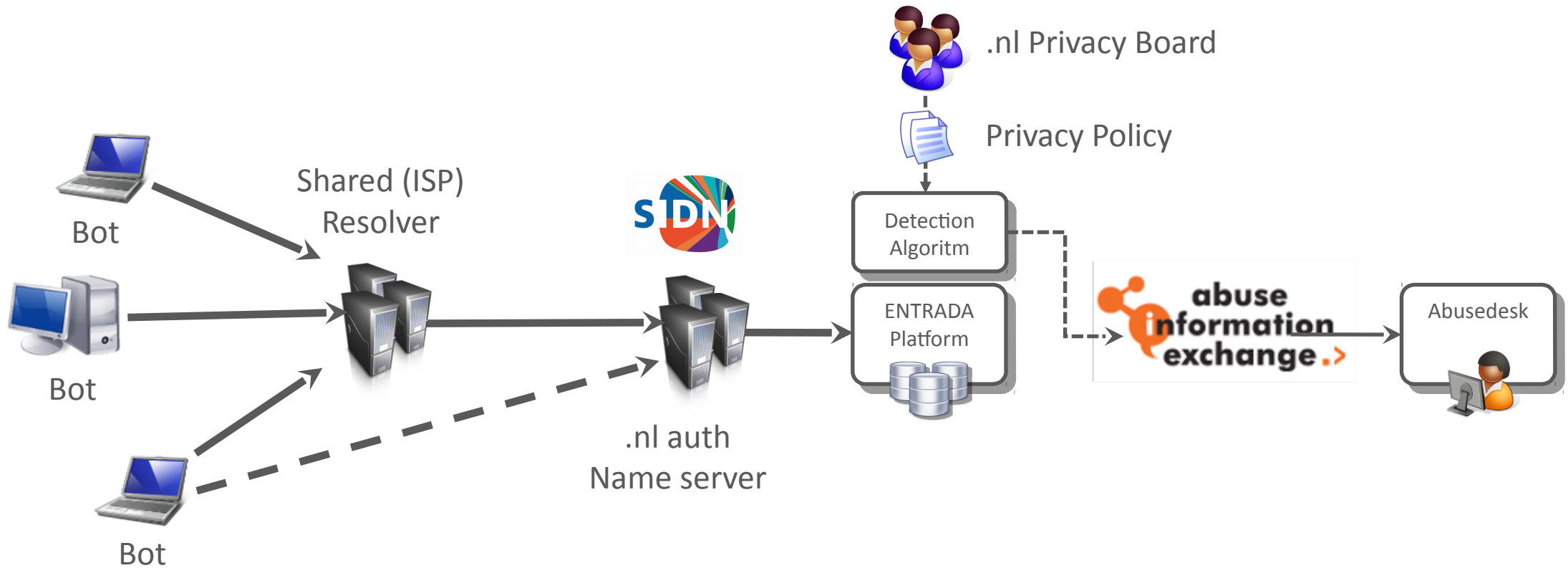


# Malicious Domain Detection (2/2)

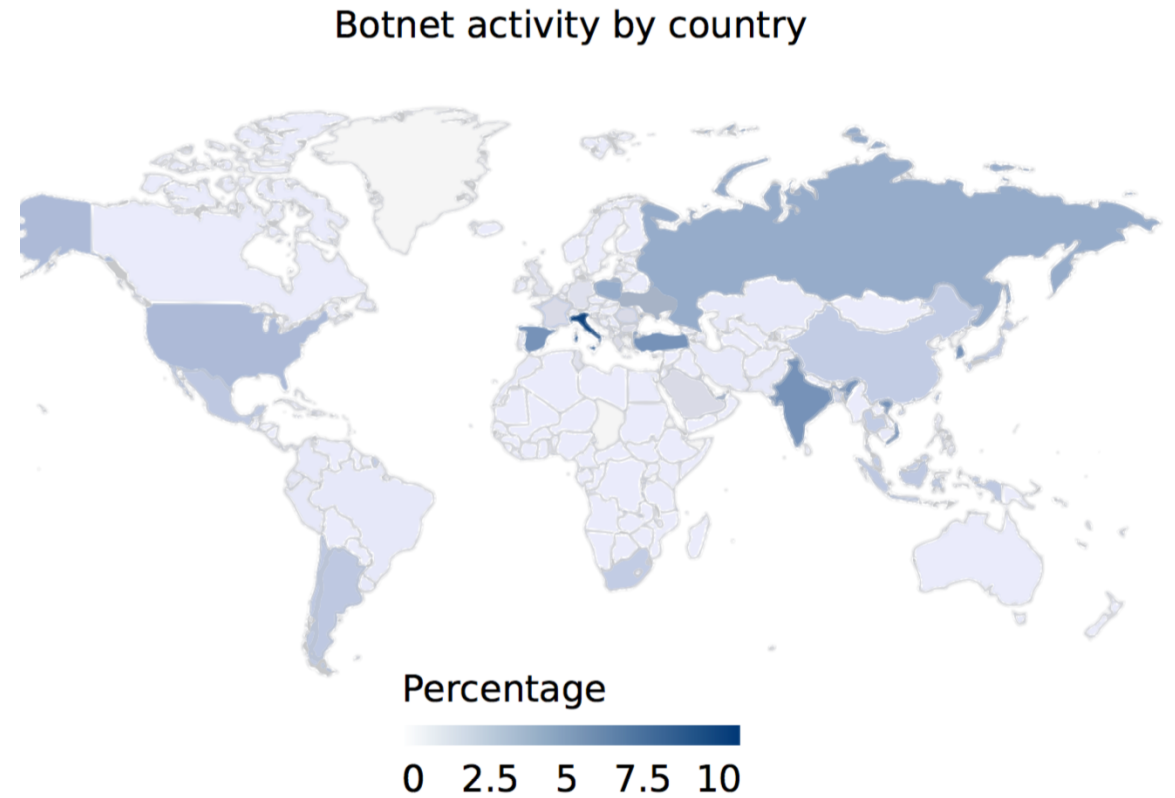
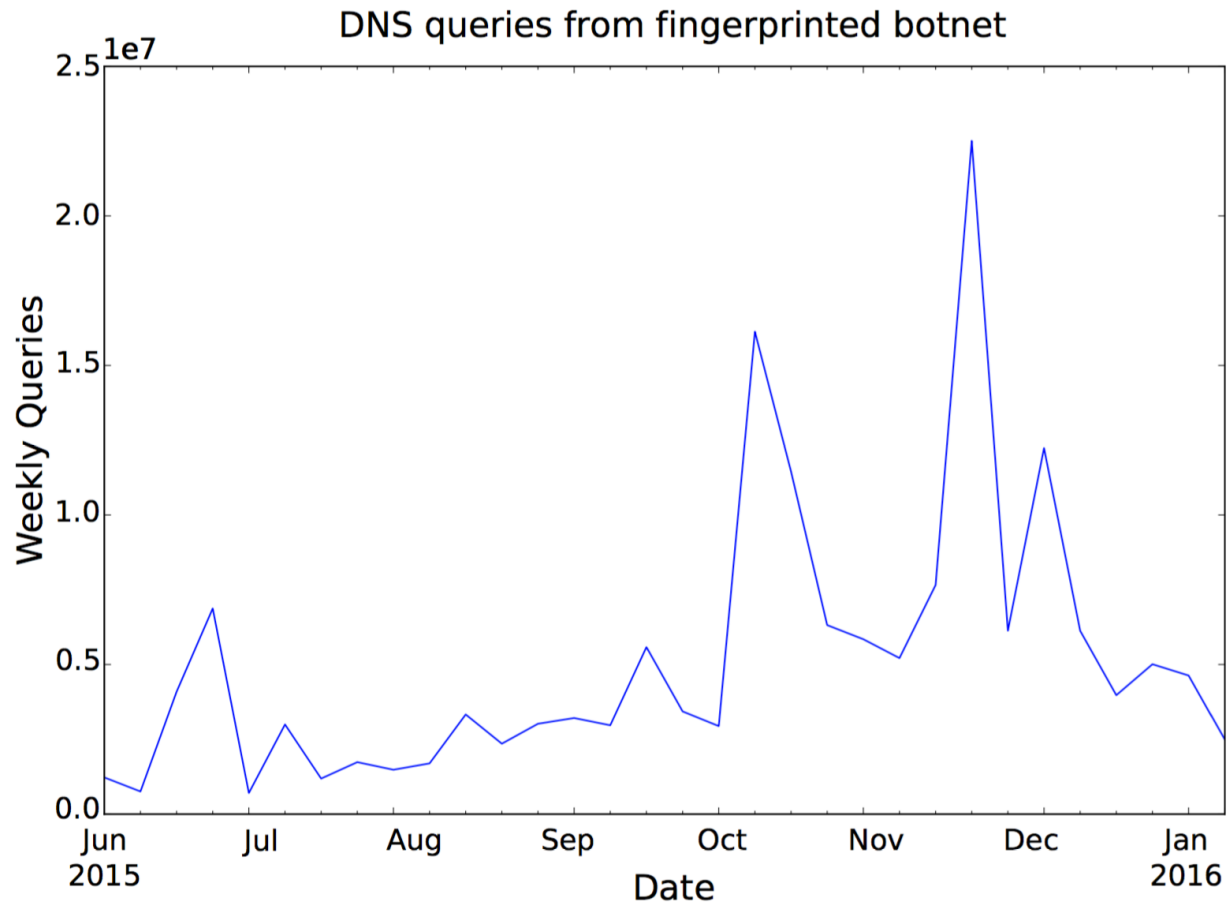
*Every day workflow*



# Botnet Client Detection (1/2)



# Botnet Client Detection (2/2)

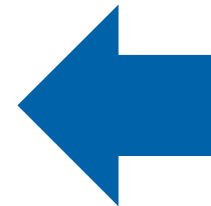


# Uptake of DKIM/DMARC (1/3)

- Email security standards DKIM (RFC 6376) and DMARC (RFC 7489)
- Approach: count standardized labels

Where is DKIM/DMARC used most?

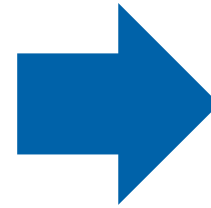
```
select country, count(1) as total
from dns.queries
where qtype=16
and (qname like "%_domainkey.%")
or qname like "_dmarc .%")
and rcode=0
and ((year=2014 and month>6) or
year=2015)
group by country
```



Use standard SQL for analysis

# Uptake of DKIM/DMARC (2/3)

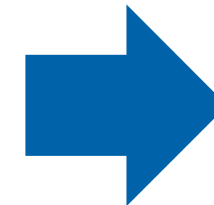
Country	# Queries	Percentage
US	208,533,790	<b>42.60</b>
IE	84,515,235	<b>17.26</b>
NL	79,052,717	<b>16.15</b>
BE	67,963,161	<b>13.88</b>
FI	9,112,053	1.86
RU	7,306,873	1.49
DE	7,119,556	1.45
GB	5,897,734	1.20
CN	5,446,895	1.11
DK	2,958,891	0.60



89.9% of queries originate from top 4 countries

# Uptake of DKIM/DMARC (3/3)

Provider	ASN	# Queries	Percentage
Google	AS15169	302,465,578	<b>61.79</b>
Microsoft	AS8075	51,556,416	<b>10.53</b>
Unknown	UNKN	15,788,699	3.22
AOL	AS1668	12,971,456	<b>2.65</b>
Yahoo	AS36647	11,83,129	<b>2.30</b>
Yahoo	AS26101	10,24,857	<b>2.07</b>
Yahoo	AS36646	9,150,523	<b>1.87</b>
Yahoo	AS34010	4,522,388	<b>0.92</b>
IDC China Tel	AS23724	4,520,819	0.92
Mail.ru	AS47764	3,659,097	0.75



82.13% of queries originate from 4 large e-mail providers

# Summary

- We have shown ENTRADA, a DSW built using open-source “big data” tools
- It enables quick hypothesis testing and application development using SQL
- We have shown real world example use cases
- ENTRADA can be extended to other use cases
- Download and contribute!



# Future Work

- More DNS research in collaboration with research partners
- Develop data-driven applications and services based on ENTRADA
- Facilitate ENTRADA user community