Your world. Our domain.
ENTRADA: Background, Use-Cases and Project Ideas

2017-09-07 | SWITCH Security Tools Hackathon
Agenda

• Technical Background of ENTRADA
• Use Cases
• Ideas for the Hackathon
ENTRADA Architecture

Main components

• Data sources
• Platform
• Applications and services
• Privacy framework
SQL on Hadoop
Best fit for our requirements

Hadoop Node N
  IMPALA
  PARQUET

Hadoop Node N+1
  IMPALA
  PARQUET

Hadoop Node N+2
  IMPALA
  PARQUET

HDFS
# Impala

<table>
<thead>
<tr>
<th>Data formats</th>
<th>Interfaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Web-based GUI</td>
</tr>
<tr>
<td>Hadoop formats</td>
<td>Command line (impala-shell)</td>
</tr>
<tr>
<td>Apache Avro</td>
<td>Python (Impyla)</td>
</tr>
<tr>
<td>Apache Parquet</td>
<td>JDBC</td>
</tr>
</tbody>
</table>

---
Apache Parquet

- Why not just use the PCAP files?
  - Reading (compressed) PCAP data is just too slow
  - Analytical engines cannot read PCAP files
**HDFS**

- Distributed file system for storing large volumes of data
- High availability through replication of data blocks
- Scalable to hundreds of PB's and thousands of servers
Cluster Design

nano sized

location I
management node

location II
data nodes

location III
data nodes

2Gb/s network
Workflow

Query data available for analysis within 10 minutes
Security Use Cases

- nDEWS: Detection of new malicious domain names
- checks for every new domain name:
  - number of queries, unique sources, unique ASes, unique countries
  - uses k-means (k=2) clustering to split domains
Security Use Cases

- DGA detection
- based on lexical features (using tool by SANS ISC)
- and NX queries
- e.g. vufrx4xjje1y5spwle2kp8g4qn5uag2nq636apww9mhyk03k4z.nl
Security Use Cases

• Detect phishes on subdomains
  • e.g. paypal.com.login.example.com
  • filter ENTRADA for keywords in subdomain labels
Security Use Cases

• Detect phishes on subdomains
  • e.g. paypal.com.login.example.com
  • filter ENTRADA for keywords in subdomain labels
• Verify user submissions automatically
  • e.g. from PhishTank or fraud helpdesks
  • features: domain name age, registrar, *DNS query peak*
Security Use Cases

• Detect phishes on subdomains
  • e.g. paypal.com.login.example.com
  • filter ENTRADA for keywords in subdomain labels
• Verify user submissions automatically
  • e.g. from PhishTank or fraud helpdesks
  • features: domain name age, registrar, *DNS query peak*
• Detect botnet infections
  • Cutwail botnet used for sending SPAM
  • Bots use their own, home-brew, recursive resolver <- does weird things
Other Use Cases

- Stats: stats.sidnlabs.nl
- Research, e.g.:
  - How do recursive resolvers select authoritative name servers? ([tech report](#))
  - How to understand and predict changes of anycast catchments? ([tech report](#))
- Adhoc queries, e.g.:
  - Do we see strange queries for a domain name?
  - What else is a resolver querying?
- Policy changes, e.g.:
  - What happens if we change zone file updates from 2h to 1h?
  - What would happen if QNAME minimization gets widely adopted?
Use cases in other organizations

• DNS Magnitude: Measure the popularity of domain names (nic.at)  

• Anomaly Detection

• ...
Use cases in other organizations

- DNS Magnitude: Measure the popularity of domain names (nic.at) [https://ccnso.icann.org/meetings/copenhagen58/presentation-dns-magnitude-13mar17-en.pdf](https://ccnso.icann.org/meetings/copenhagen58/presentation-dns-magnitude-13mar17-en.pdf)

- Anomaly Detection

- ...

- *Your use case here!*
Some Ideas for the Hackathon

1. Platform to exchange ENTRADA queries and related projects
Some Ideas for the Hackathon

1. Platform to exchange ENTRADA queries and related projects
2. Interface to query other ENTRADA instances: e.g. for suspicious resolvers, suspicious sub domains, ...
Some Ideas for the Hackathon

1. Platform to exchange ENTRADA queries and related projects
2. Interface to query other ENTRADA instances: e.g. for suspicious resolvers, suspicious sub domains, ...
3. Resolver profiling
   • e.g. resolvers of eye ball networks, resolvers of domainers, validating resolvers
   • helps to filter for malicious resolvers
Some Ideas for the Hackathon

1. Platform to exchange ENTRADA queries and related projects

2. Interface to query other ENTRADA instances: e.g. for suspicious resolvers, suspicious sub domains, ...

3. Resolver profiling
   • e.g. resolvers of eye ball networks, resolvers of domainers, validating resolvers
   • helps to filter for malicious resolvers

4. ENTRADA + CBOR: Collect queries in lightweight CBOR format at anycast instances and convert it into Parquet
Some Ideas for the Hackathon

1. Platform to exchange ENTRADA queries and related projects
2. Interface to query other ENTRADA instances: e.g. for suspicious resolvers, suspicious sub domains, ...
3. Resolver profiling
   - e.g. resolvers of eye ball networks, resolvers of domainers, validating resolvers
   - helps to filter for malicious resolvers
4. ENTRADA + CBOR: Collect queries in lightweight CBOR format at anycast instances and convert it into Parquet
5. Detect related abuse: Which domain names have the same characteristics as known malicious domains?