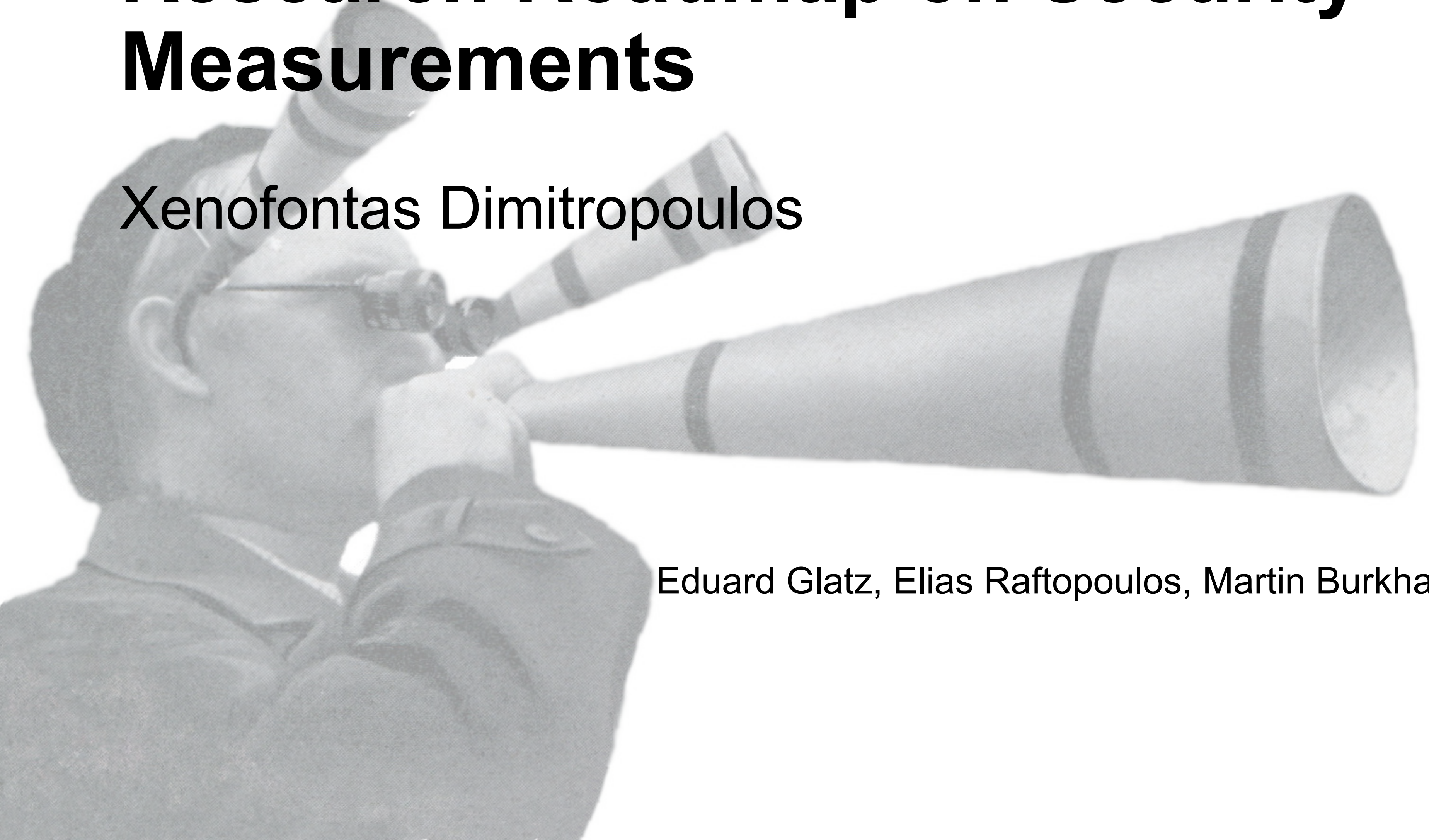
Research Roadmap on Security Measurements

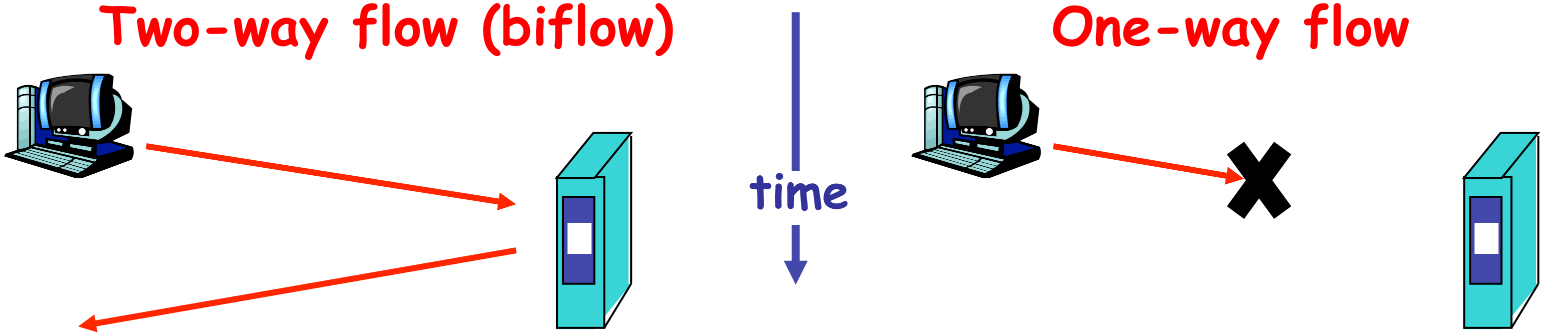
Xenofontas Dimitropoulos

Eduard Glatz, Elias Raftopoulos, Martin Burkhart



What is One-way Traffic?

- Internet traffic can be decomposed into two- and one-way traffic flows.
- One-way flows do not receive any reply, e.g., TCP SYN w/o an ACK.



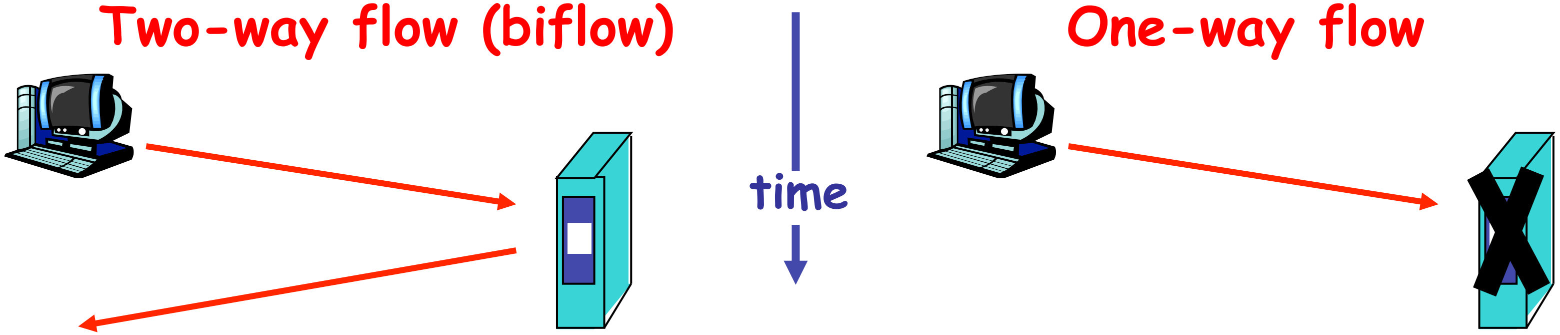
Two-way flow (biflow)

time

One-way flow

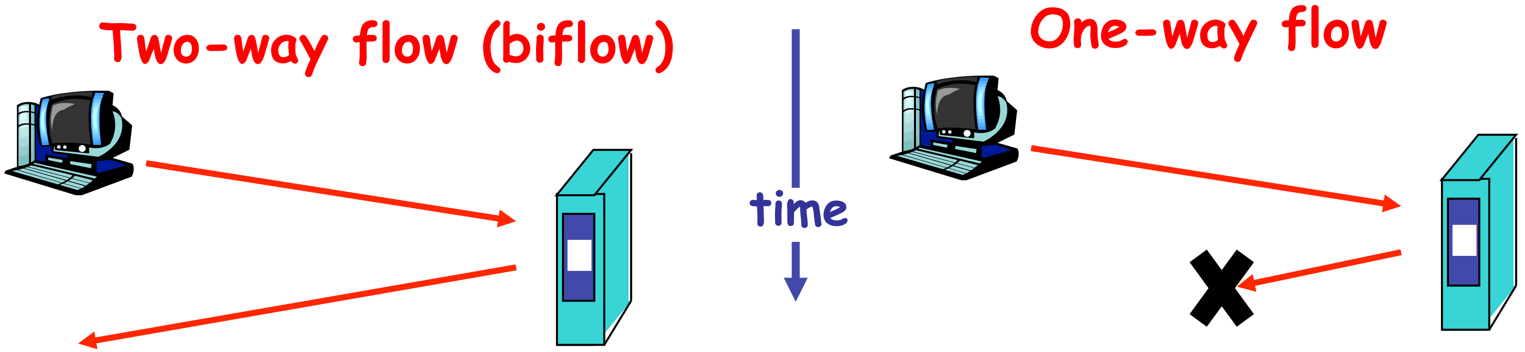
What is One-way Traffic?

- Internet traffic can be decomposed into two- and one-way traffic flows.
- One-way flows do not receive any reply, e.g., TCP SYN w/o an ACK.



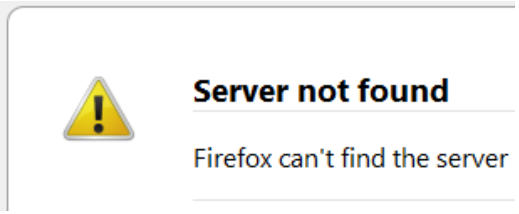
What is One-way Traffic?

- Internet traffic can be decomposed into two- and one-way traffic flows.
- One-way flows do not receive any reply, e.g., TCP SYN w/o an ACK.



Why Should We Care?

- One-way flows are associated with **interesting events** like:
  - Unreachable services
  - Scanning
  - Congestion and routing loops
  - NATs & firewalls
  - Misconfigured port numbers
  - Peer-to-peer applications
  - Prefix hijacking



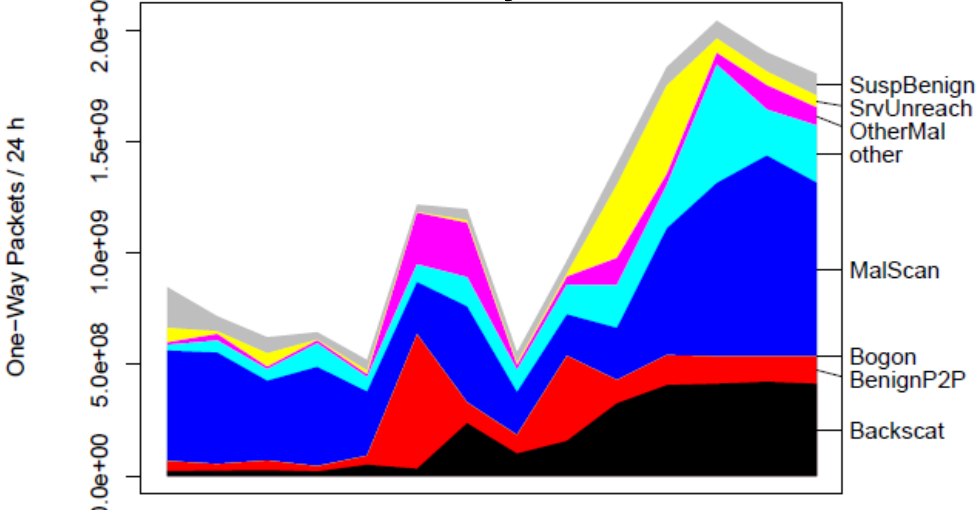
Server not found

Firefox can't find the server

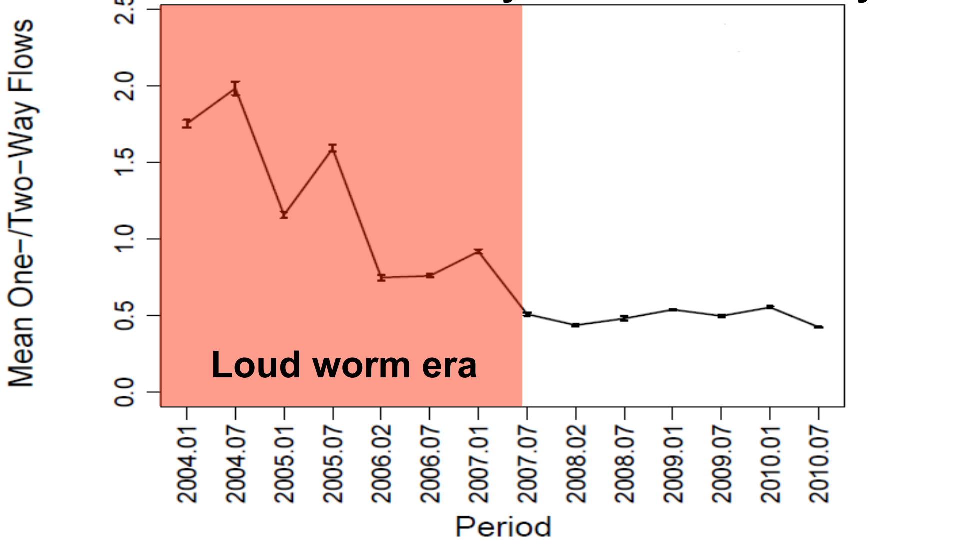
- One-way flows constitute a large fraction of Internet traffic.
- One-way flows have been minimally studied in the past.

What are we doing?

- Introduce techniques to classify one-way traffic into interesting classes.
- Characterize 7.73 petabytes of traffic towards SWITCH between 2004 and 2010.

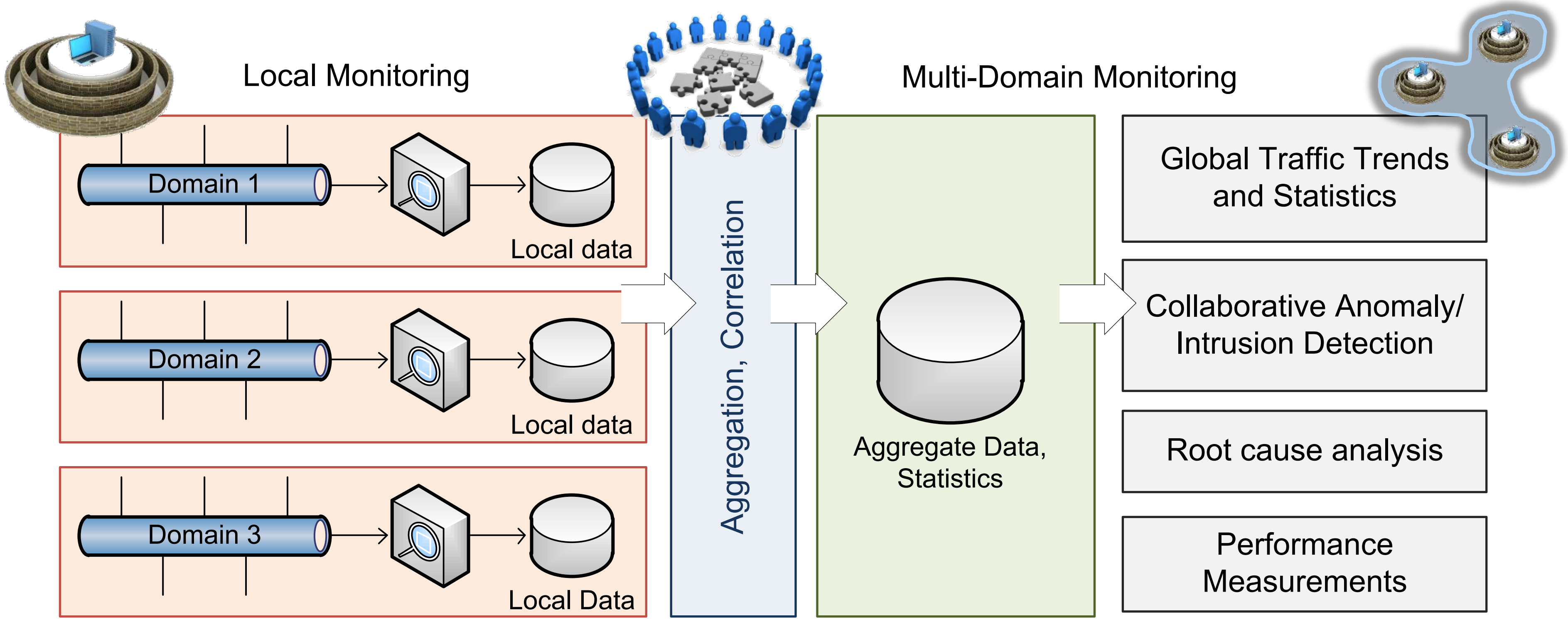


71.8% of the one-way flows are malicious



The share of one-way flows decreased by 73%

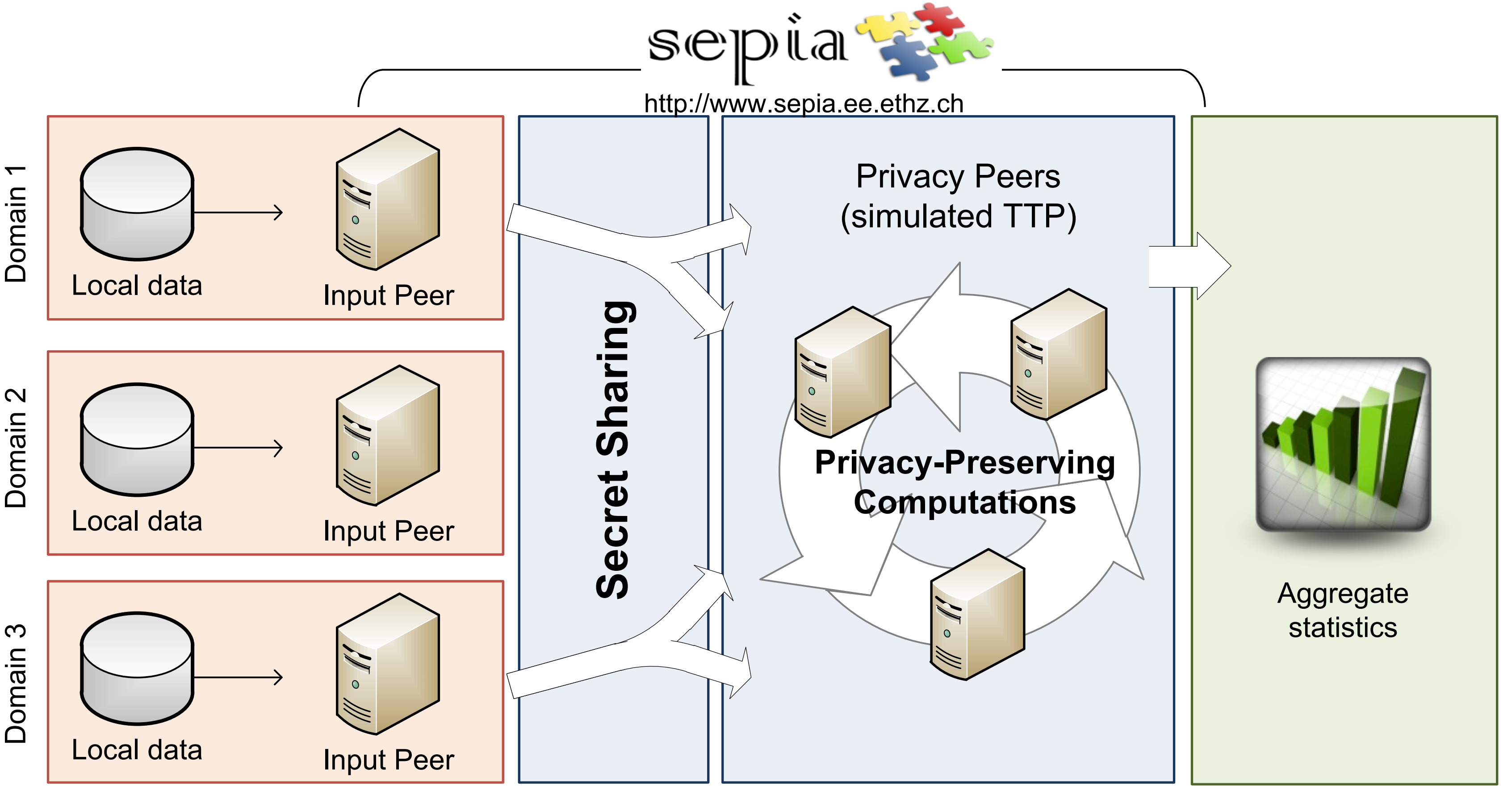
Collaborative Network Security/ Management



Private data
(known to a single domain)

Public data
(known to all domains)

SEPIA Multi-party Computation (MPC) Library



Private data
(known to a single domain)

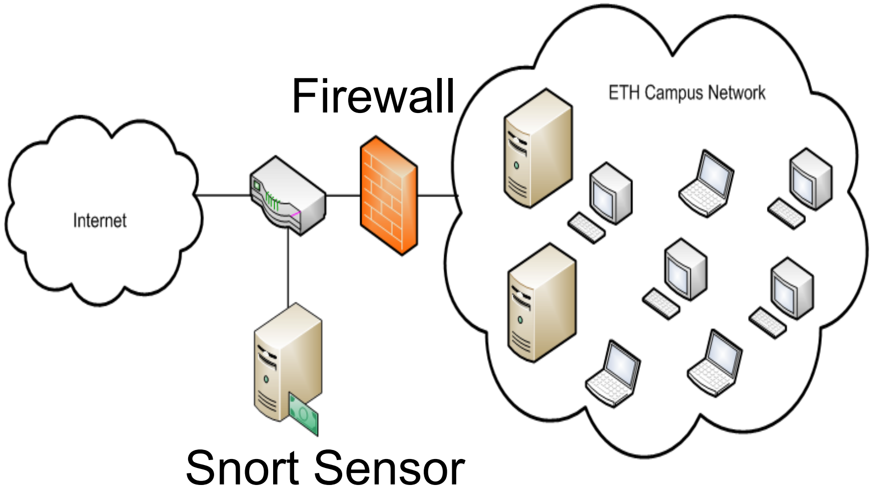
Secret data
(nobody knows)

Public data
(known to all domains)

MPC provides a much better solution to the privacy – utility tradeoff than anonymization

Alert Correlation in a Live Network

- Revisit a good old problem that has lasted the "test of time" without a good solution.
- Analyze an archive of alerts from a live network instead of a test-bed:
  - snort produces on average 3 million alerts per day.
  - the archives include more than 9 months of alerts.

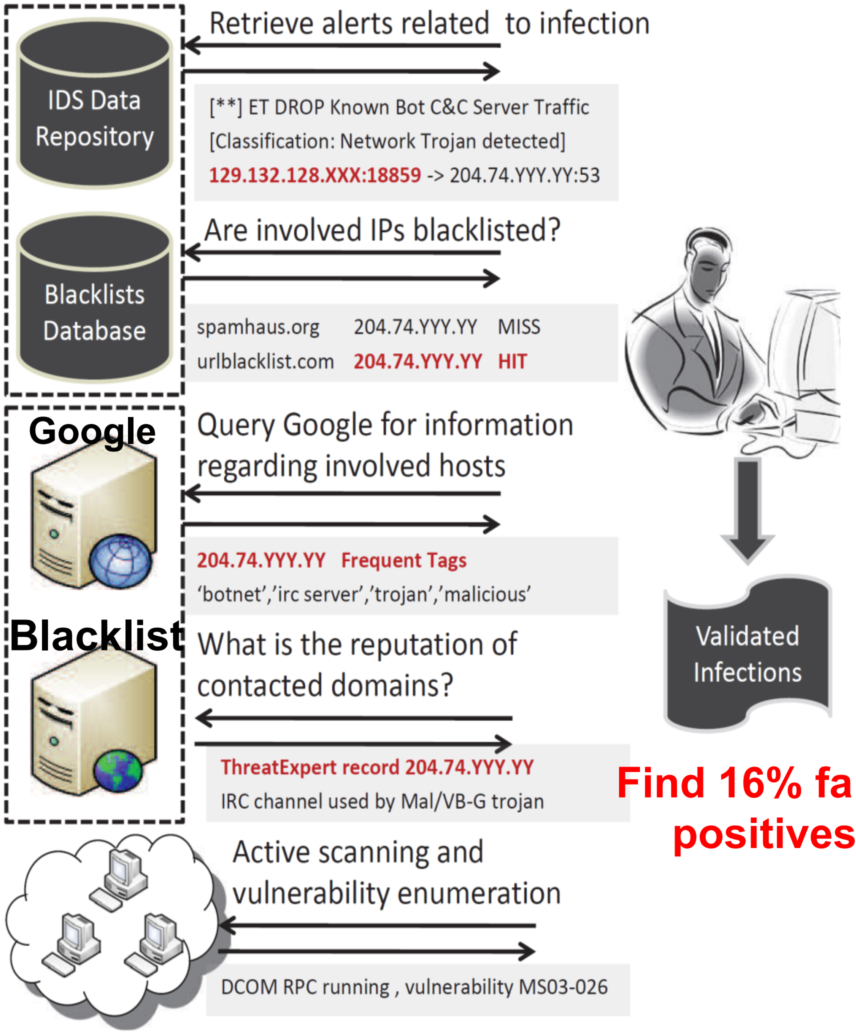


- Build novel alert correlation heuristic to find infected hosts within the network (extrusion detection).
- Characterize 9,163 infected hosts observed over a period of 9 months.

Validate Infected Hosts

- Over a period of one month manually assess 200 live suspected infections.

- Validation methodology:
  - Lunch daily list of suspected infected IP addresses.
  - Collect relevant data from 5 independent sources (see Figure).
  - Use background knowledge about the suspected malware.
  - Connect the dots to make a positive or negative assessment.



IDS Data Repository

Retrieve alerts related to infection

```
[**] ET DROP Known Bot C&C Server Traffic
[Classification: Network Trojan detected]
129.132.128.XXX:18859 -> 204.74.YYY.YY:53
```

Blacklists Database

Are involved IPs blacklisted?

```
spamhaus.org      204.74.YYY.YY   MISS
urlblacklist.com  204.74.YYY.YY   HIT
```

Google

Query Google for information regarding involved hosts

```
204.74.YYY.YY   Frequent Tags
'botnet','irc server','trojan','malicious'
```

Blacklist

What is the reputation of contacted domains?

```
ThreatExpert record 204.74.YYY.YY
IRC channel used by Mal/VB-G trojan
```

Active scanning and vulnerability enumeration

```
DCOM RPC running , vulnerability MS03-026
```

Validated Infections

Find 16% false positives

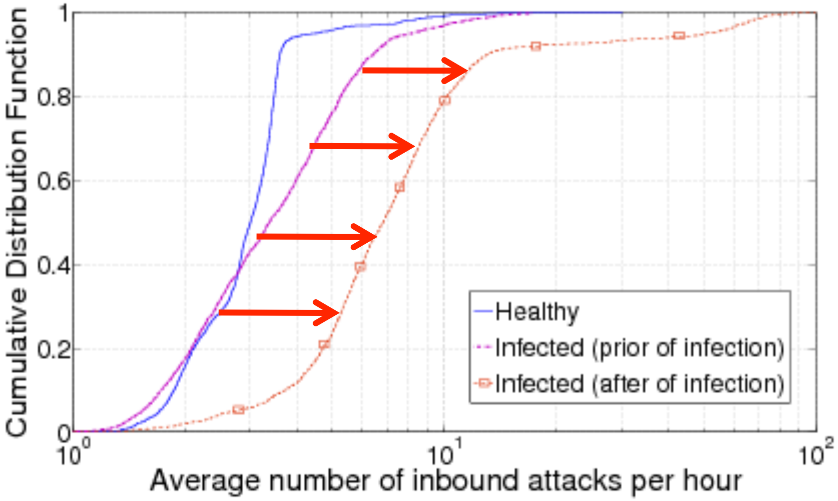
Characterize Infected Hosts

- Characterize 9,163 infected hosts observed over a period of 9 months.

- Find infections in 9% out of a total of 91K hosts.

- Selected observations:
  - The volume of inbound attacks to infected hosts increases rapidly after their infections.
  - Strong spatial correlations: new infections are more likely to occur close to already infected hosts.

Impact of Infection to Inbound Attacks


Cumulative Distribution Function
10^0
10^1
10^2
Average number of inbound attacks per hour
Healthy
Infected (prior of infection)
Infected (after of infection)