# Early Detection of Spam Domains with Passive DNS and SPF

Simon Fernandez, Maciej Korczynski, Andrzej Duda

WAX - 2022

Laboratoire Informatique de Grenoble, Team Drakkar

# Spammers

# Spam domain lifecycle

- Register a new domain (or bulks)
- Configure the domain
- Start the spam campaign and wait for victims
- The domain gets blacklisted and/or taken down
- Restart the procedure with a fresh domain

# Data Source

# Passive DNS of new domains - Farsight SIE and CZDS

**Farsight SIE:** Passive DNS feed of multiple sensors around the world

**CZDS:** ICANN platform to get the full zonefiles of most gTLD

- Stealth measurement
- No interfering
- See the real trafic
- Detect new domains

# Passive DNS of new domains - Farsight SIE and CZDS

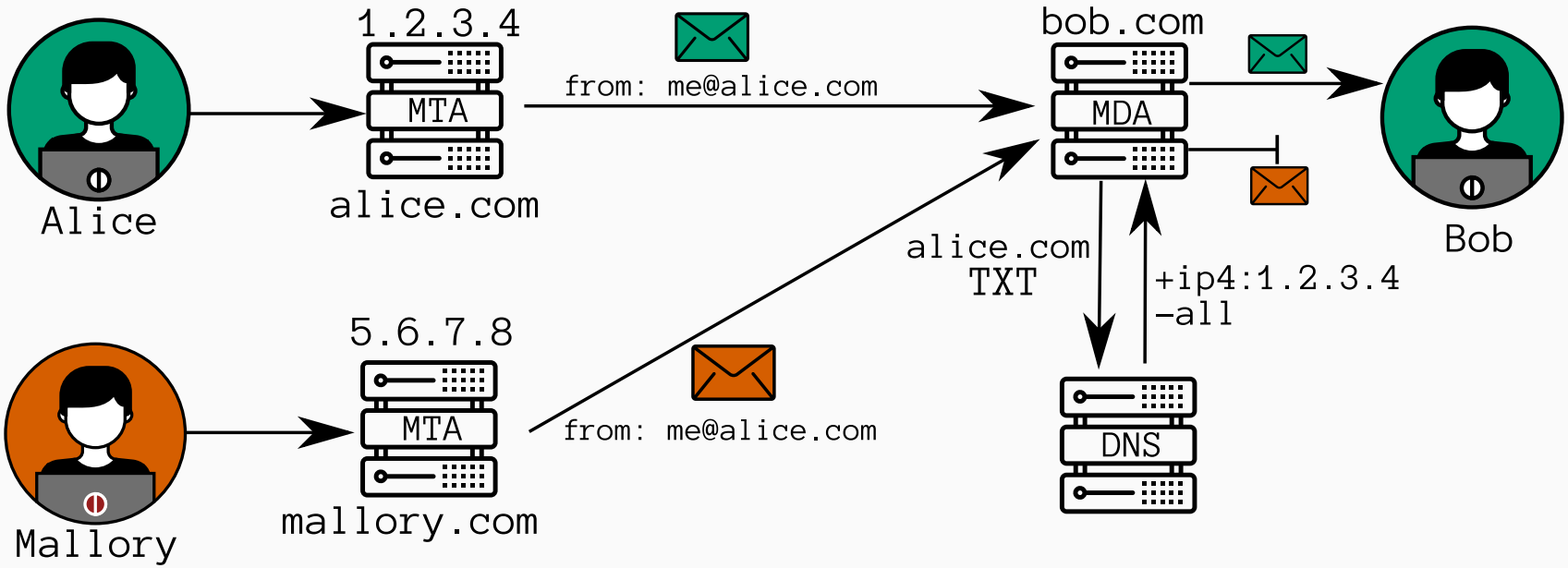
**Farsight SIE:** Passive DNS feed of multiple sensors around the world

**CZDS:** ICANN platform to get the full zonefiles of most gTLD

- Stealth measurement
- No interfering
- See the real trafic
- Detect new domains

# Sender Policy Framework

# The situation with Sender Policy Framework (SPF)



# The SPF rules

<qualifier><modifier>[:<target>]

Qualifier:

- `PASS +`
- `NEUTRAL ~`
- `SOFTFAIL ?`
- `FAIL -`

Modifier:

- With `<target>`
  - `ip4, ip6`
  - `include`
  - `a, mx` (optional)
  - `exists` (optional)
- Without `<target>`
  - `ptr`
  - `all`

`+ip4:1.2.3.0/24 +a -all`

# The SPF rules

<qualifier><modifier>[:<target>]

Qualifier:

- `PASS +`
- `NEUTRAL ~`
- `SOFTFAIL ?`
- `FAIL -`

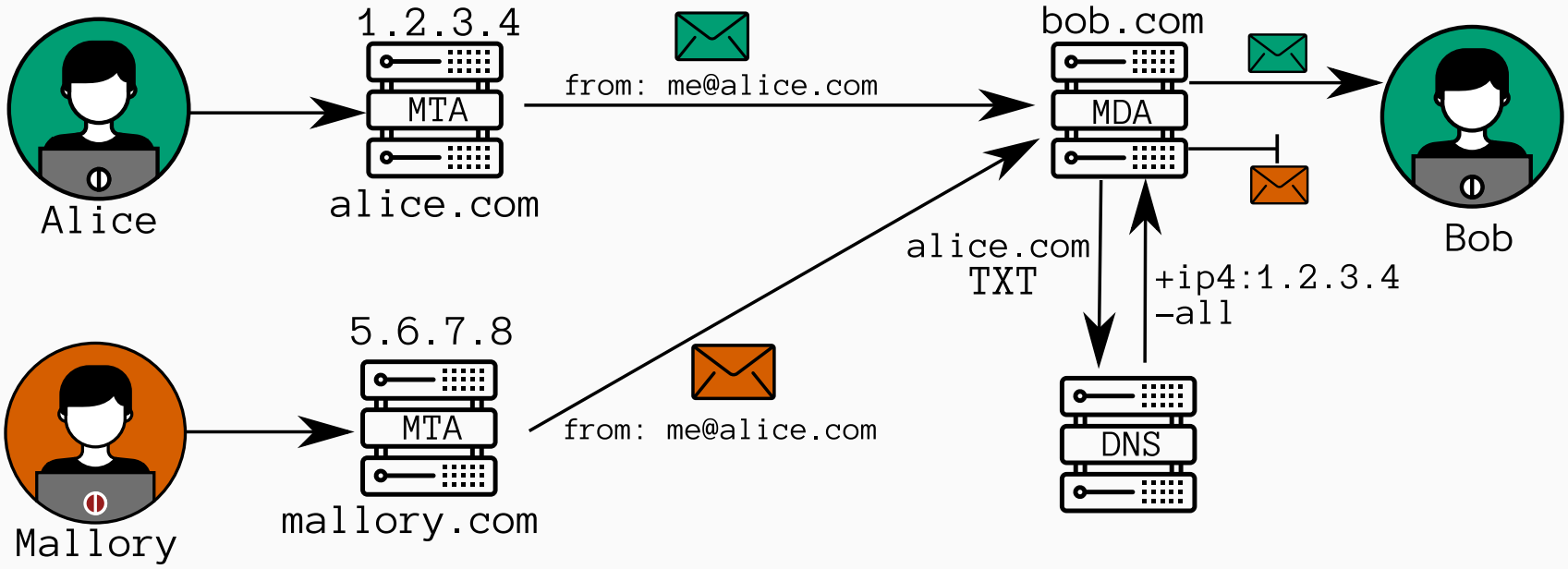
Modifier:

- With `<target>`
  - `ip4, ip6`
  - `include`
  - `a, mx` (optional)
  - `exists` (optional)
- Without `<target>`
  - `ptr`
  - `all`

```
+ip4:1.2.3.0/24 +a -all
```

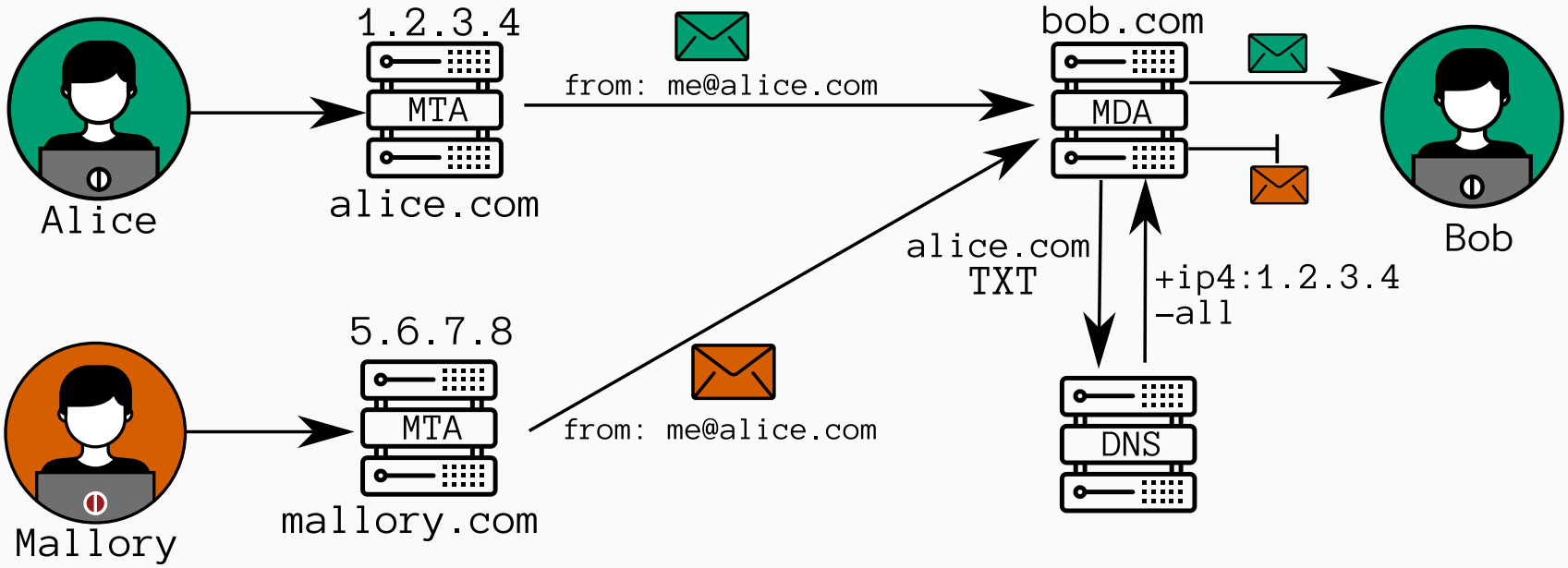
# Features to detect spammers

# Temporal analysis: TXT queries spike




We should see a spike in requests to the `TXT` record of `alice.com`

# Temporal analysis: TXT queries spike




We should see a spike in requests to the `TXT` record of `alice.com`

# SPF configuration

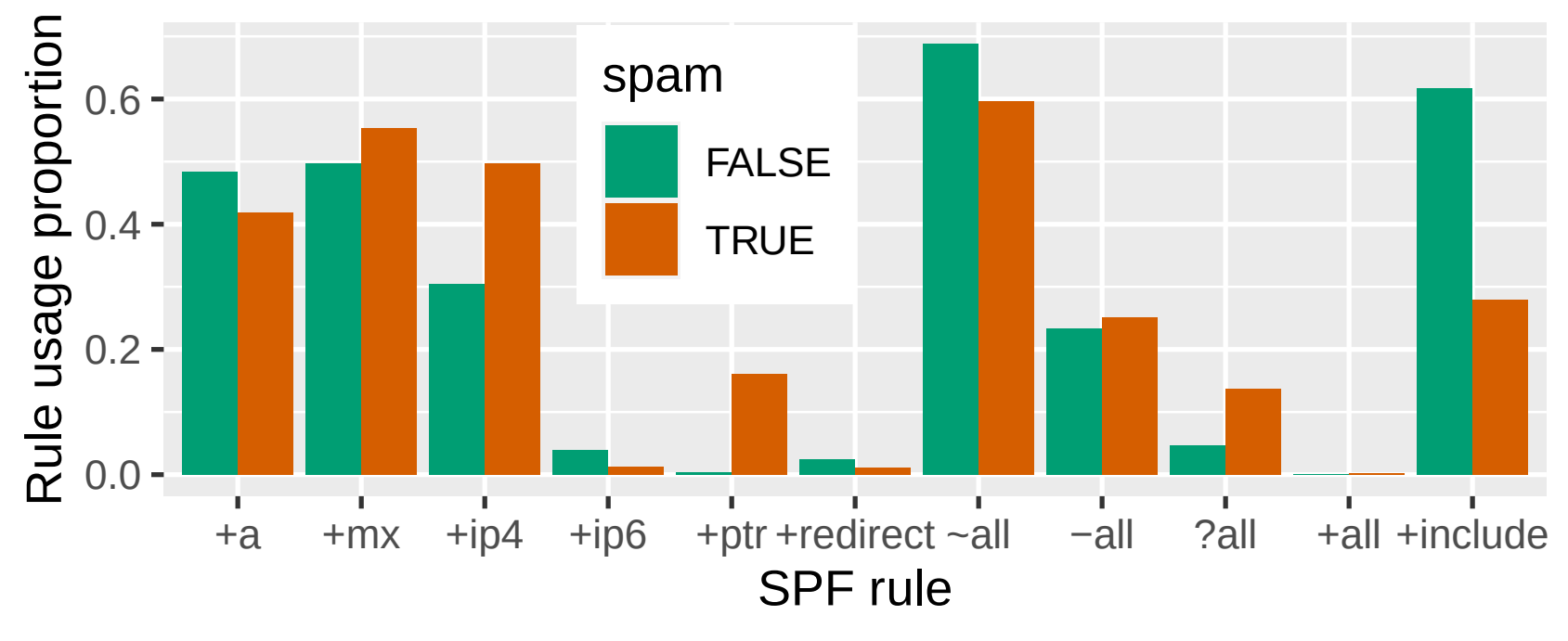



**Figure 1:** Proportion of domains using a given SPF rule

# Graph relations

`a.com IN TXT ip4:1.2.3.4 include:b.com`

Build the SPF relation graph:

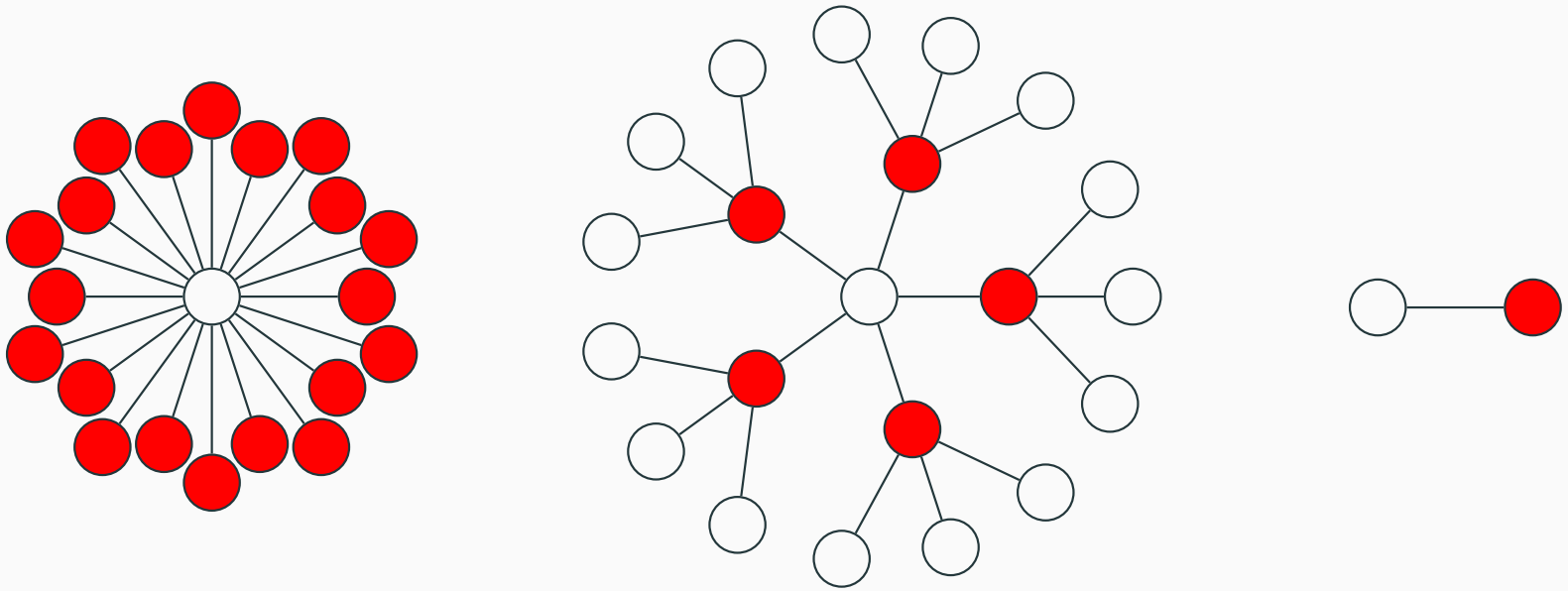
- Nodes: IPs, IP networks, domains
- Edge: A node uses another one as a target in its rules

# Graph relations

`a.com IN TXT ip4:1.2.3.4 include:b.com`

Build the SPF relation graph:

- Nodes: IPs, IP networks, domains
- Edge: A node uses another one as a target in its rules



# Graph metrics

**Mean neighbor degree**
Does the domain use widely deployed rules?

**Toxicity**
Is the target used by known spammers?

**Mean neighbor Toxicity**
Is the domain using targets mainly present in spam domains configurations?

# Results of the Classifier

# Two datasets

Static dataset:

- Only SPF rules and graph
- No time analysis
- Less precise
- Needs a single `TXT` request

Dynamic dataset:

- Uses all properties
- More precise
- Precision increases with time

# Two datasets

Static dataset:

- Only SPF rules and graph
- No time analysis
- Less precise
- Needs a single `TXT` request

Dynamic dataset:

- Uses all properties
- More precise
- Precision increases with time

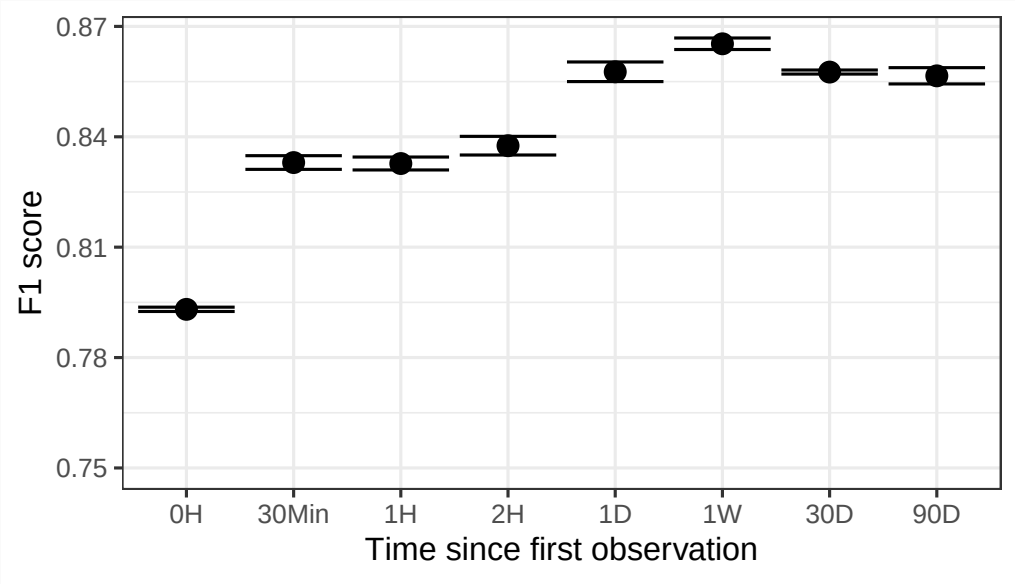
# Two datasets

Static dataset:

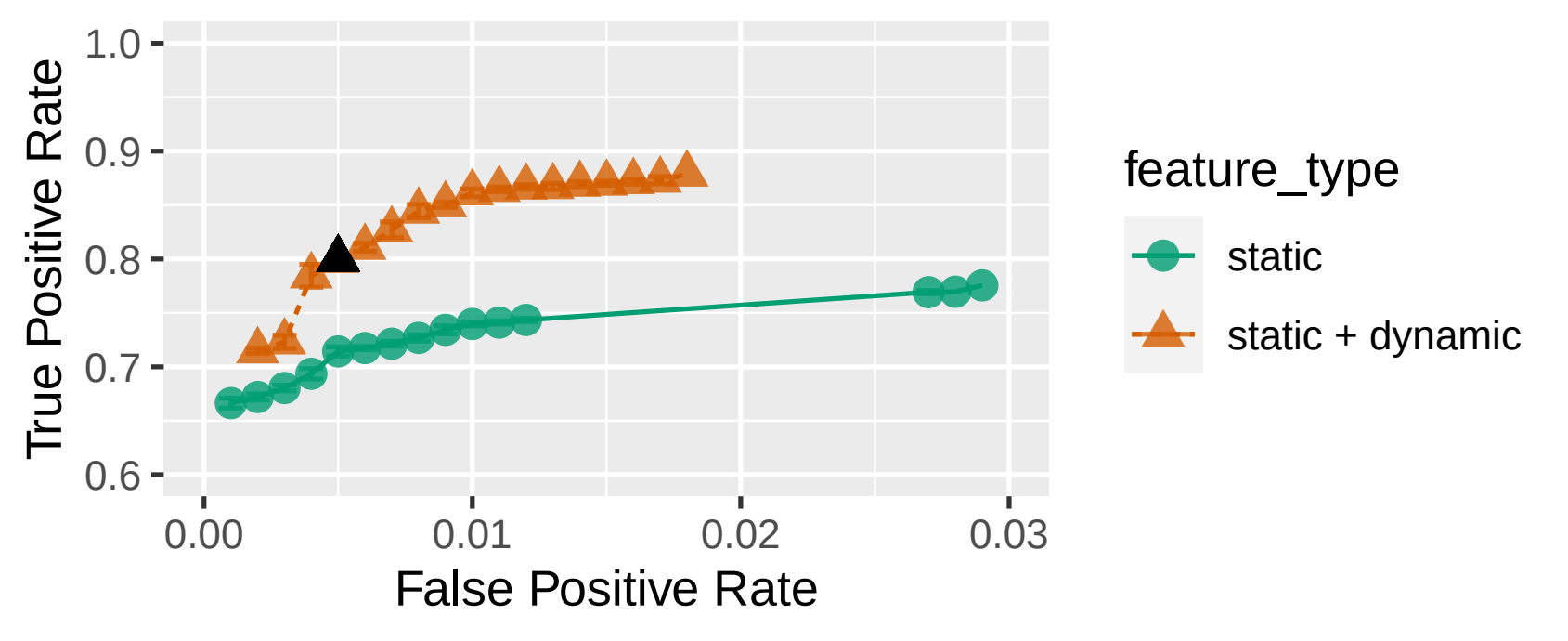
- Only SPF rules and graph
- No time analysis
- Less precise
- Needs a single `TXT` request

Dynamic dataset:

- Uses all properties
- More precise
- Precision increases with time

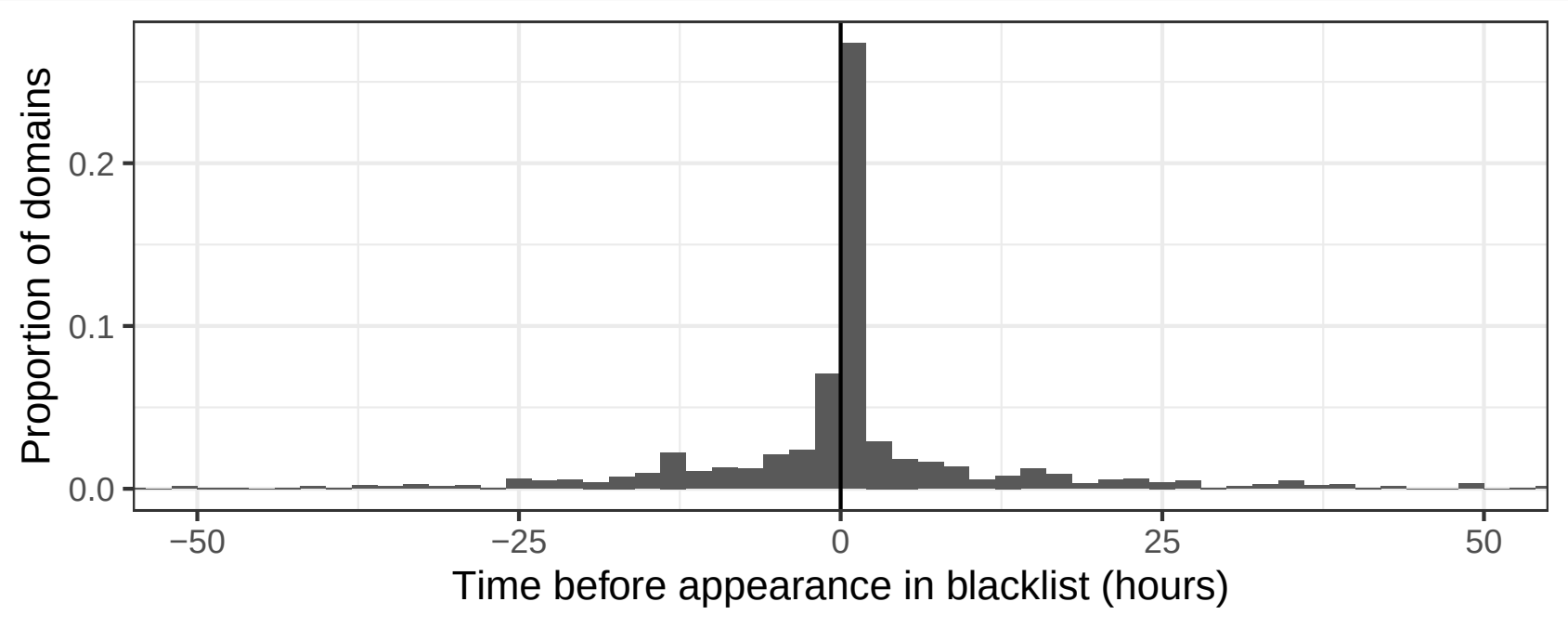


# Precision Results




- True Positive Rate: 80%
- False Positive rate: 0.5%

# Speed Results




- Faster in 70% of cases
- In 30% of cases, we are more than 24h faster

# Conclusion

- Spam domain detection is a race
- Spammers must use SPF to appear legitimate
- We use passive DNS to get the SPF configurations
- Our classifier reaches high detection rates with low false positives
- It can efficiently run on a single `TXT` query and refine its classification with additionnal traffic

# Conclusion

Thank you for your attention

All questions are welcome