CERT.PL
NASK

Phishing detection using various parts of the DNS ecosystem

Piotr Białczak

Michał Hałoń

TLP:CLEAR

cert.pl

Who we are

NASK

National Research Institute

Runs .pl DNS registry

Hosts CERT POLSKA

CERT.PL NASK

Polish National CSIRT

Securing Polish internet

Most common threats

phishing + fraud

98 000 incidents in 2024

= 95% total

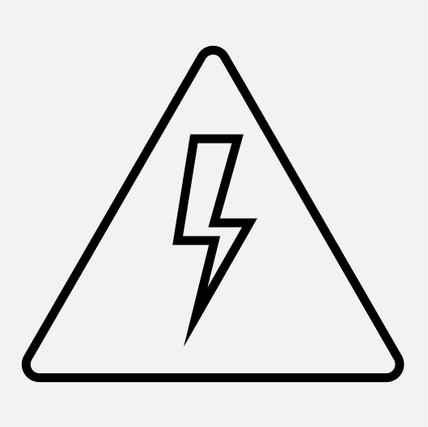
Using DNS to protect our constituency

Block domains at the .pl registry

Proof required

Block domains at resolvers

Proof required



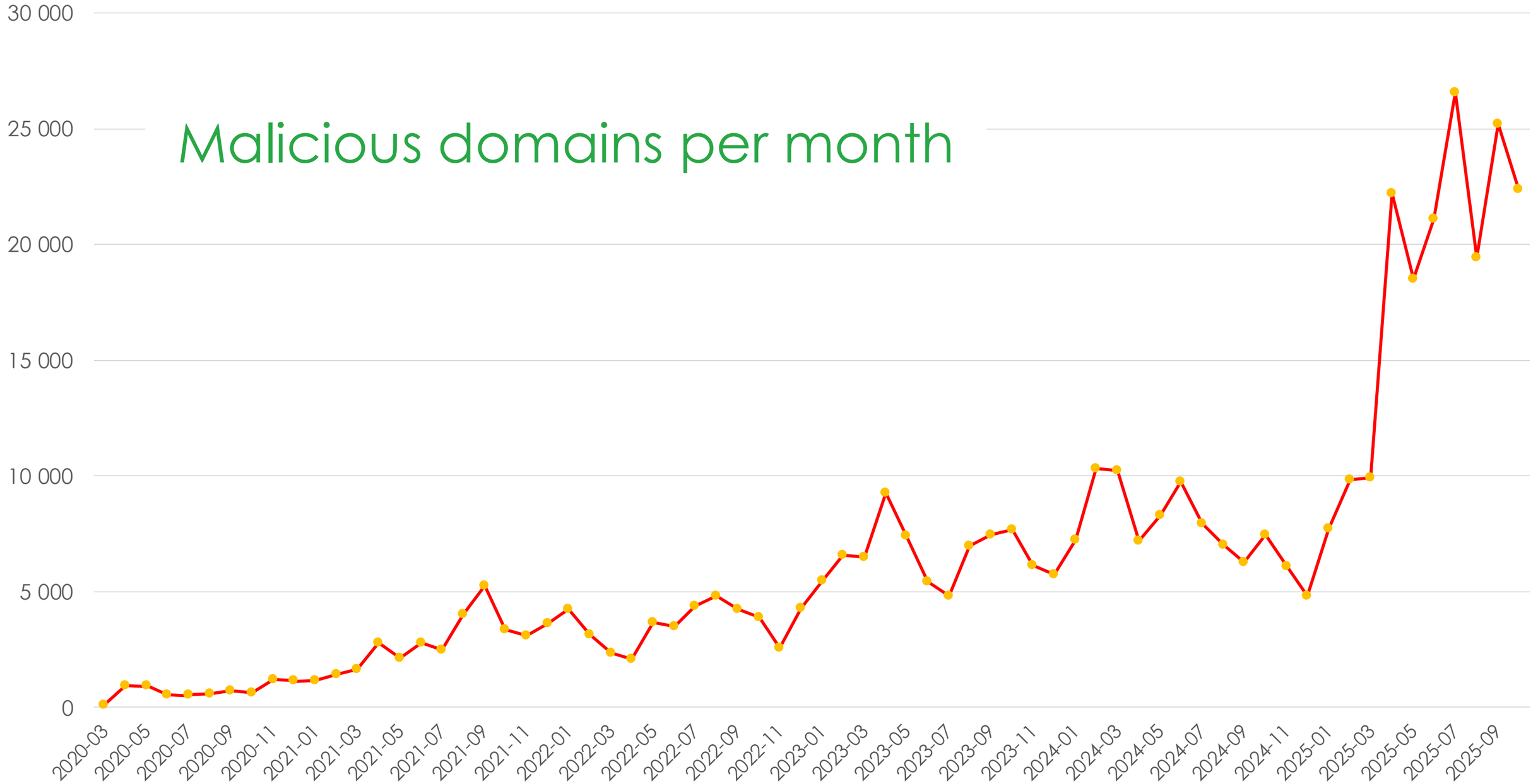
Block domains at resolvers

If likely malicious

Country-level DNS firewall

- Launch: March 2020
- We provide a feed of malicious domains: "Warning List"
- Manually verified
- Updated 24/7
- Data is public
- Blocking is voluntary
- Used by all big Polish telcos

Malicious domains per month



How to identify malicious domains?

1. URLs reported by users (multiple channels)

2. Hunting (CT, pDNS, registry data, etc)

3. ML-powered hunting = main topic of this talk!

DNS4EU project

Project number: 101095329 21-EU-DIG-EU-DNS
Project name: DNS4EU and European DNS Shield.
This project is co-funded by the European Union.

Co-funded by
the European Union

Goals of the European Commission

1. Offer a high-end **alternative to existing dominant non-EU public resolvers**, leading to a more resilient, more secure and diversified DNS resolution offering for EU internet users.
2. **Autonomy** of DNS resolving, diminishing the dependency on major public resolvers established outside the EU, and reducing vulnerability to outages of these resolvers.
3. Complete safeguards for EU internet users that their data and **privacy** are protected and handled according to EU rules.
4. Increased **protection** against malicious activities **based on both global and local (EU) threat feeds** and intelligence.
5. Testing and deploying innovative technologies to enhance internet access **security and privacy**.

DNS4EU

Co-funded by the European Union

ABOUT LEARN EVENTS

LEARNING

June 11, 2025

DNS4EU Goes live: A European Alternative to Google and Cloudflare DNS, Powered by Whalebone

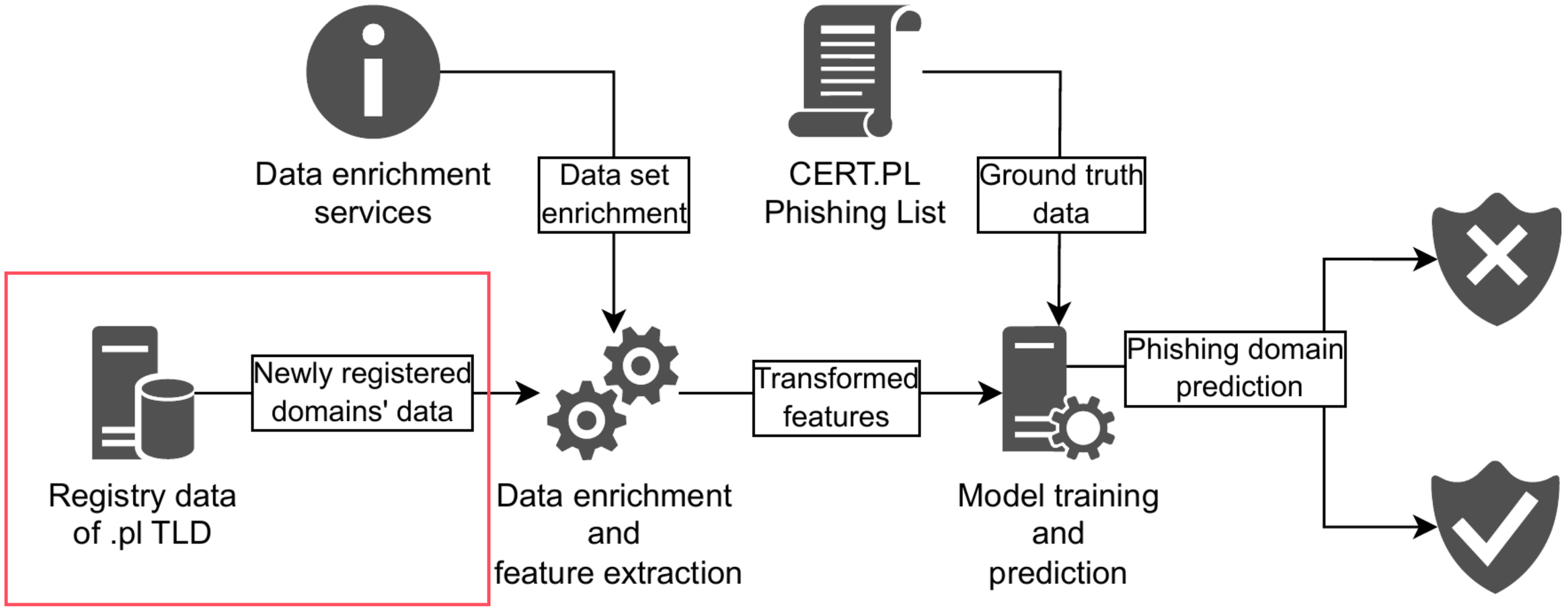
Looking into .pl registry

Approach

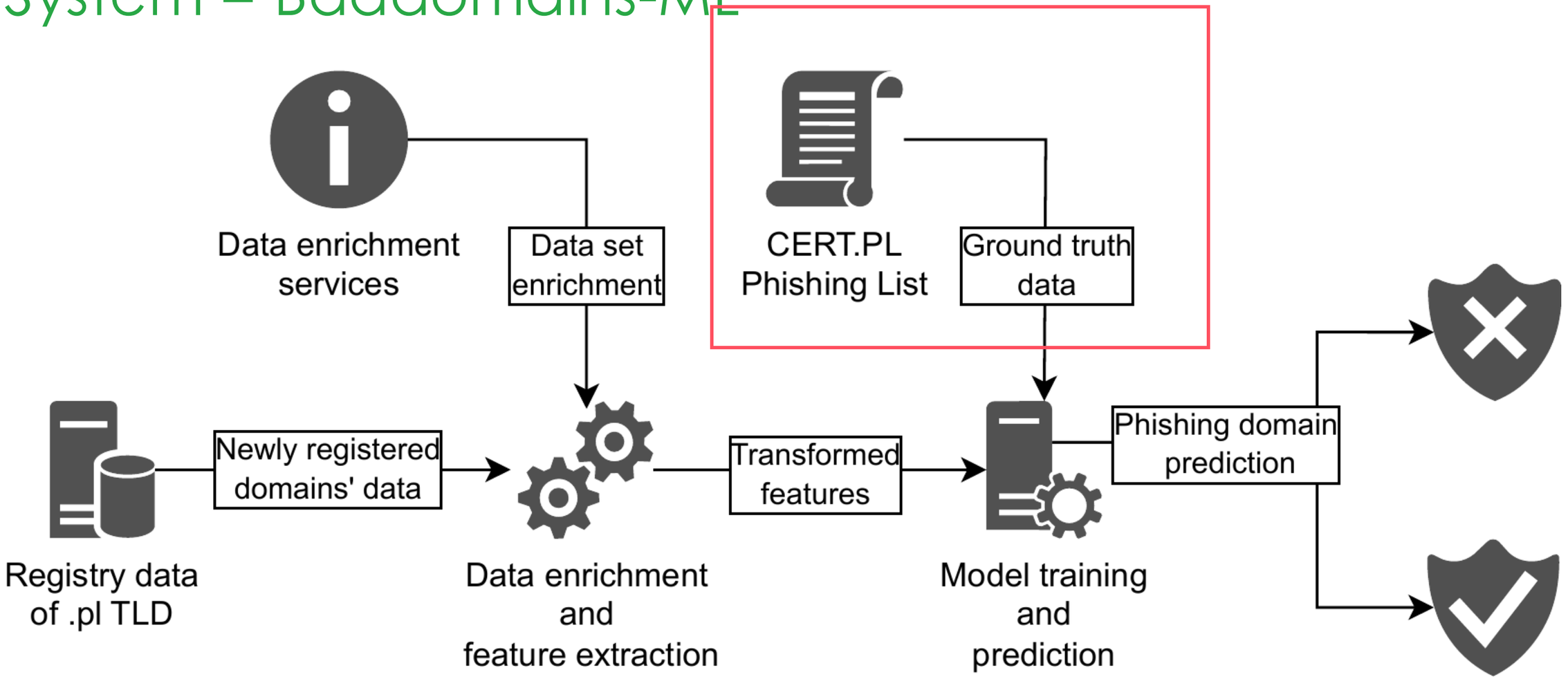
- Phishing domain detection shortly after registration
- Registry data used as model's feature basis: domain name, registrar, registrant's contact information, nameserver, …
- Focus on campaigns targeting Polish internet users

Domain Name	pl-oferta923475.pl
State	Active in DNS [REGISTERED]
Created	2025-06-01 12:41:46
Last modified	2025-06-16 16:44:47
End of the billing period	2026-06-01 12:41:46
Nameservers	pns1.regery.net pns2.regery.net pns3.regery.net

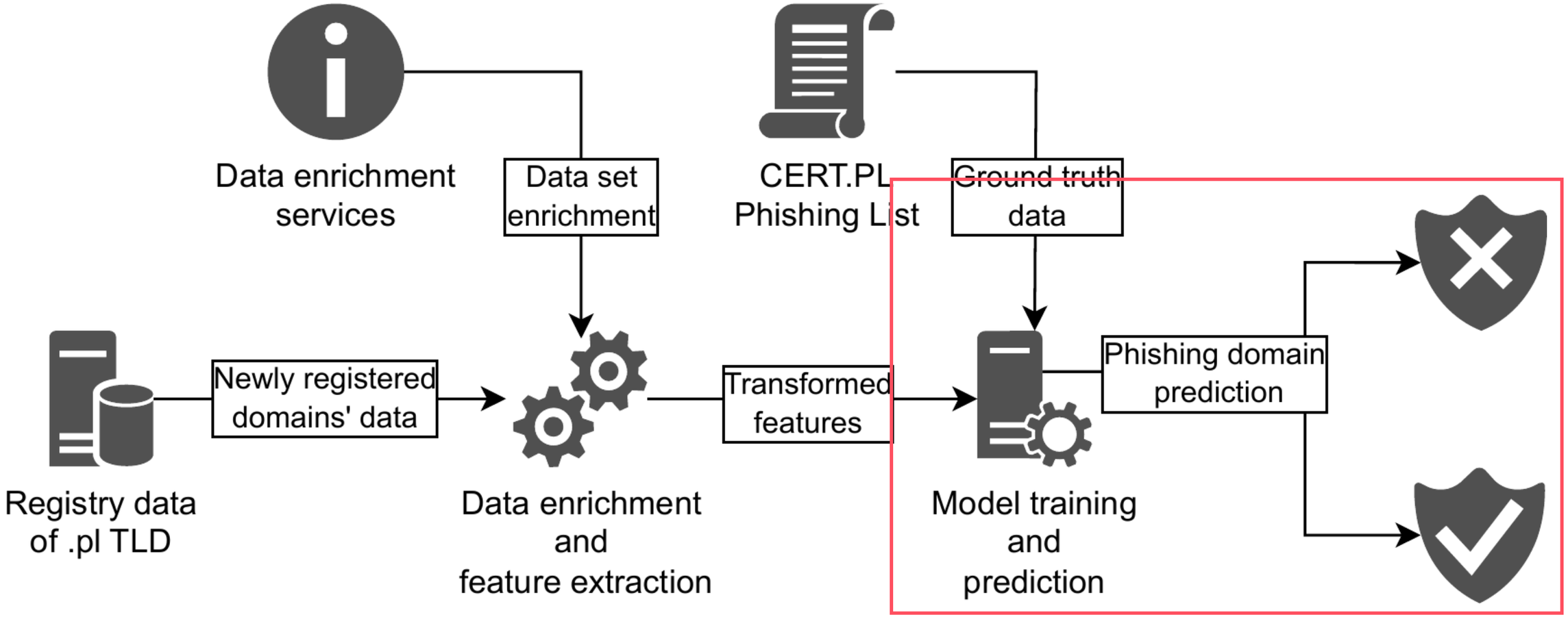
System = Baddomains-ML



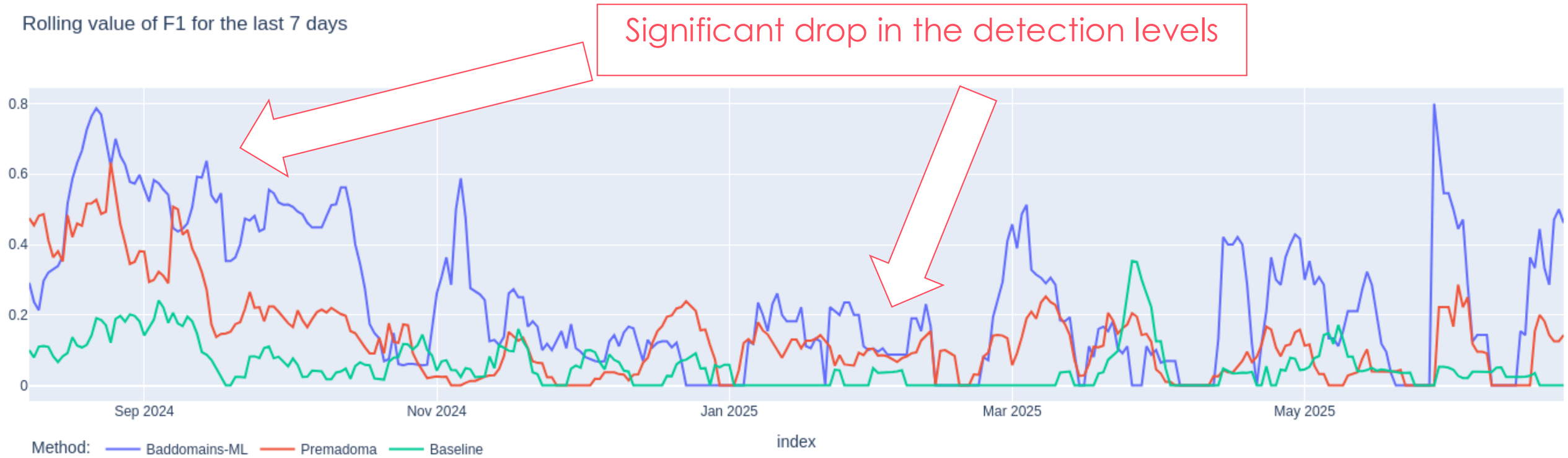
System = Baddomains-ML



System = Baddomains-ML



Detection levels



Drop in numbers: from couple of domains per day to couple per week

Problems with detection

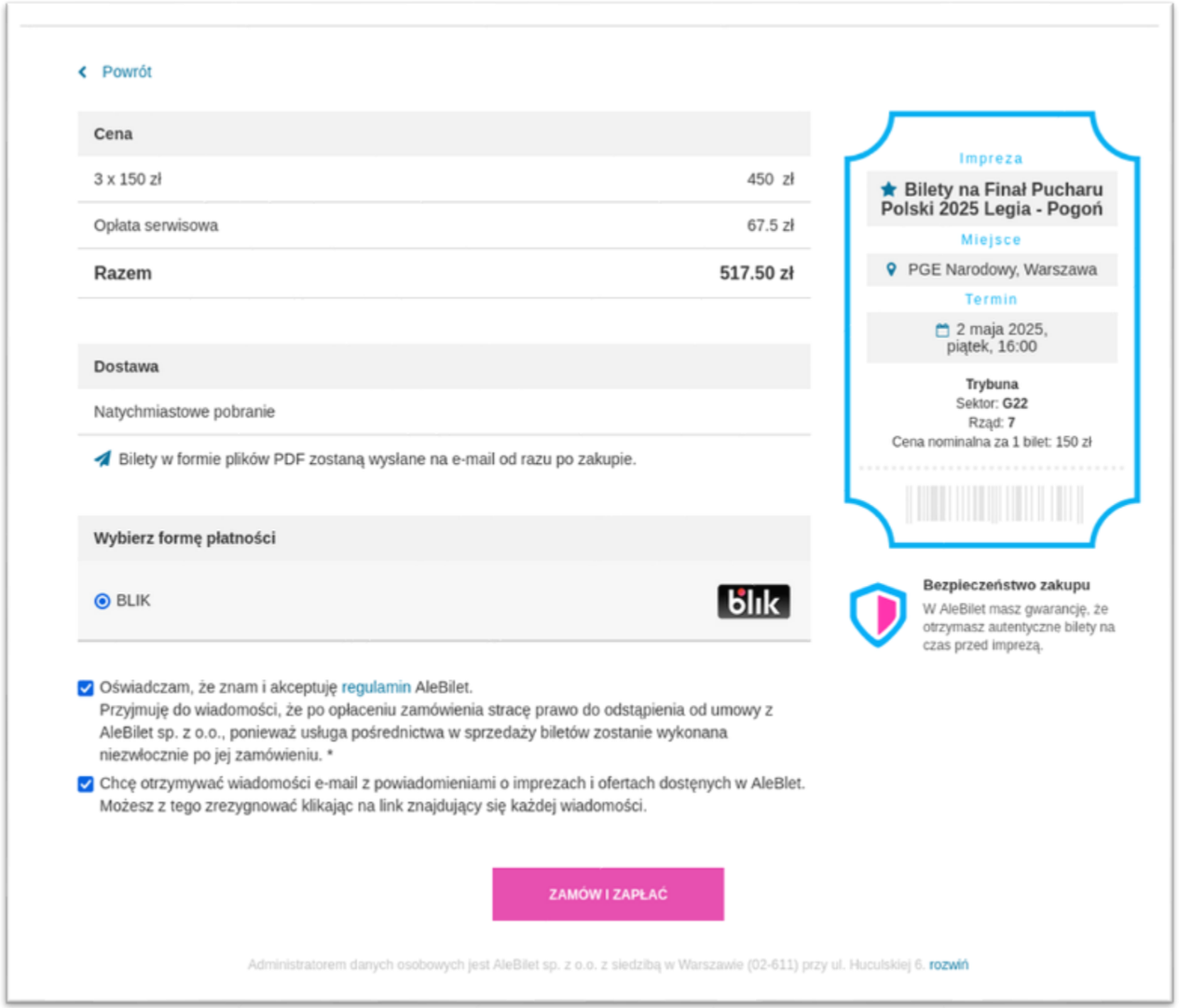
- Predicting domain is not enough
- Phishing payload on subdomains
- Main domains commonly without

  DNS *A* record

id3283[.]pl

Addressing the problems – subdomains

- Subdomains guessing
- Pivoting using known phishing payload URLs
- Hunting in DNS queries



alebilet.id3283[.]pl

Addressing the problems – low number of detections

- Model tweaking
- Result: the number of detected domains is **doubled**
- Only 30% of .pl domains on the Warning List are detected with this method - is it enough given the effort invested?
- Value – **unique detections**
  - Not seen by other non-ML methods (20% of detected domains)
  - Earlier detection than waiting for an incident report (80% of time it was faster)
  - Uncovering FPs of the Warning List

DNS query stream

Monitoring resolvers: why and how

– **Goal**: monitor activity on resolvers to detect phishing

Monitoring resolvers: why and how

- **Goal**: monitor activity on resolvers to detect phishing

- **Change of scope**: all campaigns, not just targeting Poland

Monitoring resolvers: why and how

- **Goal**: monitor activity on resolvers to detect phishing

- **Change of scope**: all campaigns, not just targeting Poland

- **Dataset**: anonymized DNS queries from DNS4EU resolvers

Monitoring resolvers: why and how

- **Goal**: monitor activity on resolvers to detect phishing
- **Change of scope**: all campaigns, not just targeting Poland
- **Dataset**: anonymized DNS queries from DNS4EU resolvers
- Research **in progress**

example.com

Domain names

example.com

Domain names

Time related

example.com

Domain names

Time related

ASN etc.

Other features

example.com

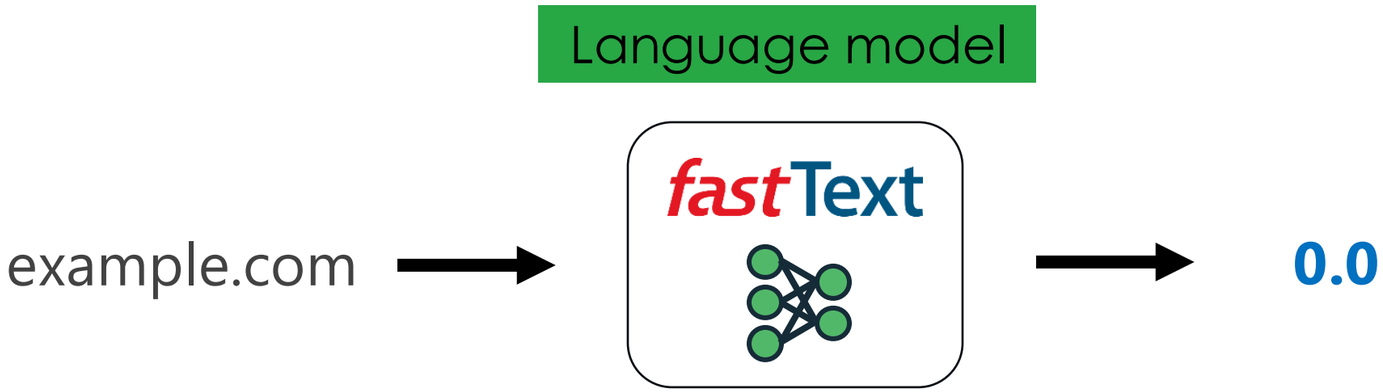
Domain names

Time related

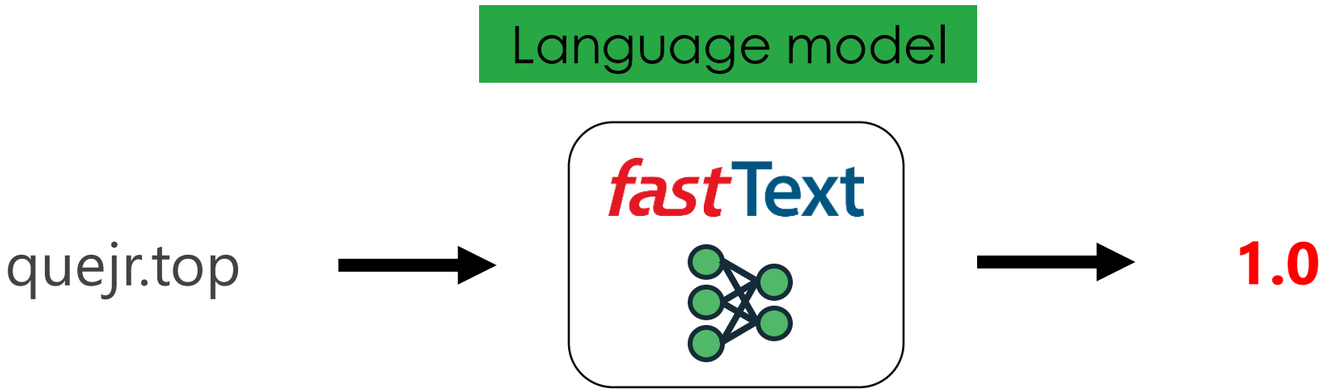
ASN etc.

Other features

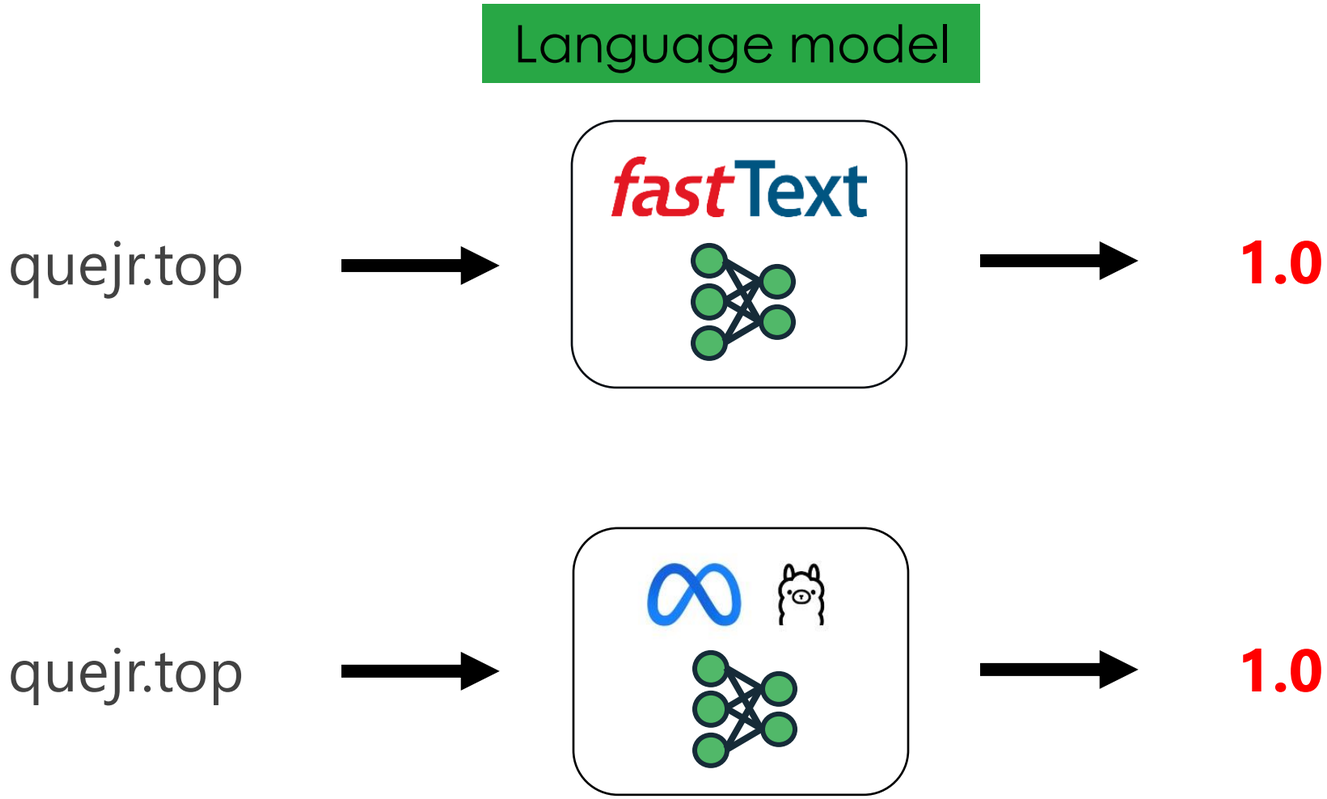
Domain names



Domain names



Domain names



Current processing pipeline

glimzo.com

trexify.net

zuntra.org

google.com

kieab.top

quejr.top

Current processing pipeline

glimzo.com

trexify.net

zuntra.org

~~google.com~~

kieab.top

quejr.top

Current processing pipeline

glimzo.com

trexify.net

zuntra.org

~~google.com~~

kieab.top

quejr.top

filter out domains from Tranco

Research-Oriented Top Sites Ranking
Hardened Against Manipulation

Current processing pipeline

glimzo.com
trexify.net
zuntra.org
~~google.com~~
kieab.top
quejr.top

filter out domains from Tranco + our own whitelist

Research-Oriented Top Sites Ranking
Hardened Against Manipulation

Our whitelist

amazonaws.com

cloudfront.net

microsoft.com

google.com

googleapis.com

apple.com

akadns.net

akamai.net

Our whitelist

reported malicious domains	
93	amazonaws.com
1758	cloudfront.net
0	microsoft.com
0	google.com
1	googleapis.com
0	apple.com
3	akadns.net
0	akamai.net

Our whitelist

reported malicious domains	
93	amazonaws.com
1758	cloudfront.net
0	microsoft.com
0	google.com
1	googleapis.com
0	apple.com
3	akadns.net
0	akamai.net

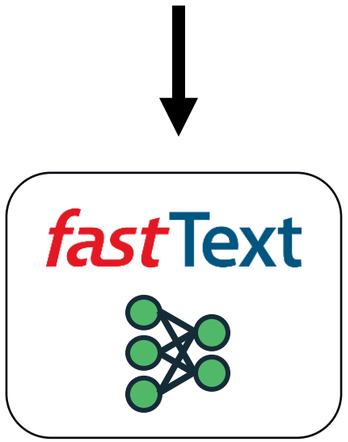
Current processing pipeline

glimzo.com
trexify.net
zuntra.org
~~google.com~~
kieab.top
quejr.top

filter out domains from Tranco + our own whitelist
Research-Oriented Top Sites Ranking Hardened Against Manipulation + other whitelists

Current processing pipeline

glimzo.com
trexify.net
zuntra.org
~~google.com~~
kieab.top
quejr.top

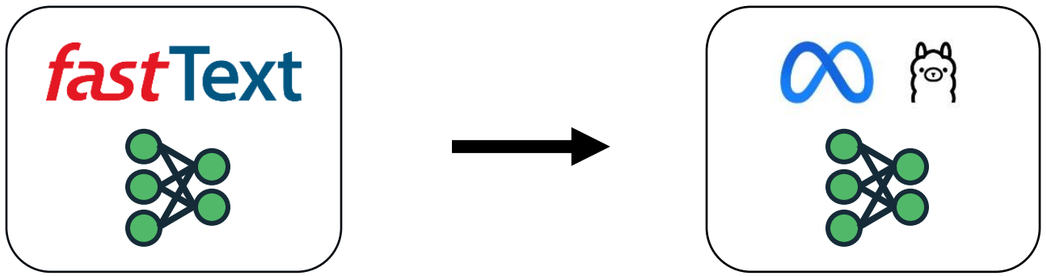


filter out domains from Tranco + our own whitelist
Research-Oriented Top Sites Ranking
Hardened Against Manipulation
+ other whitelists

Current processing pipeline

glimzo.com
trexify.net
zuntra.org
~~google.com~~
kieab.top
quejr.top

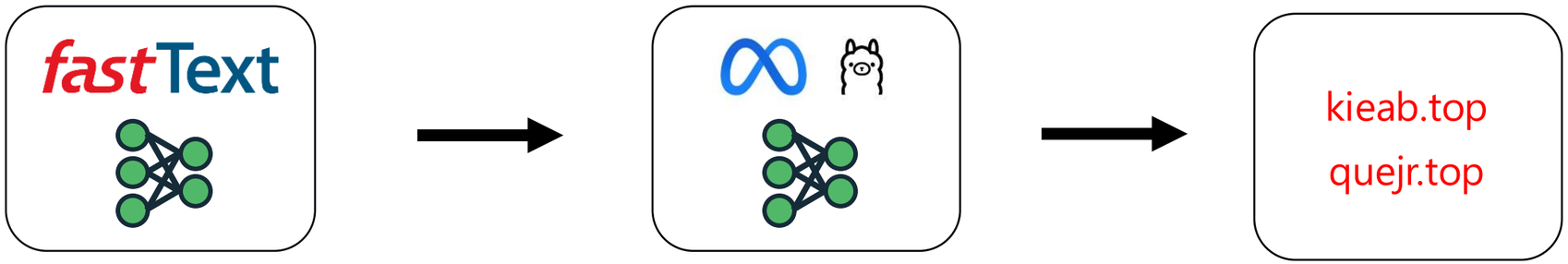
filter out domains from Tranco + our own whitelist
Research-Oriented Top Sites Ranking
Hardened Against Manipulation
+ other whitelists



Current processing pipeline

glimzo.com
trexify.net
zuntra.org
~~google.com~~
kieab.top
quejr.top

filter out domains from Tranco + our own whitelist
Research-Oriented Top Sites Ranking
Hardened Against Manipulation
+ other whitelists



example.com

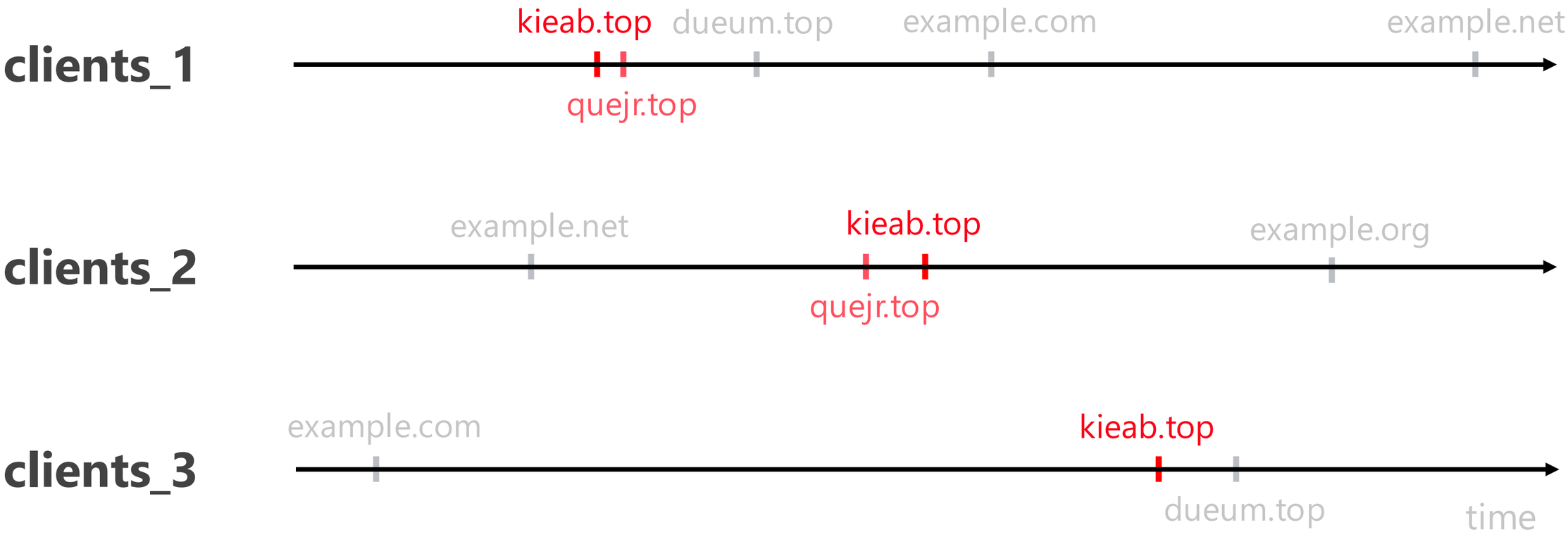
Domain names

Time related

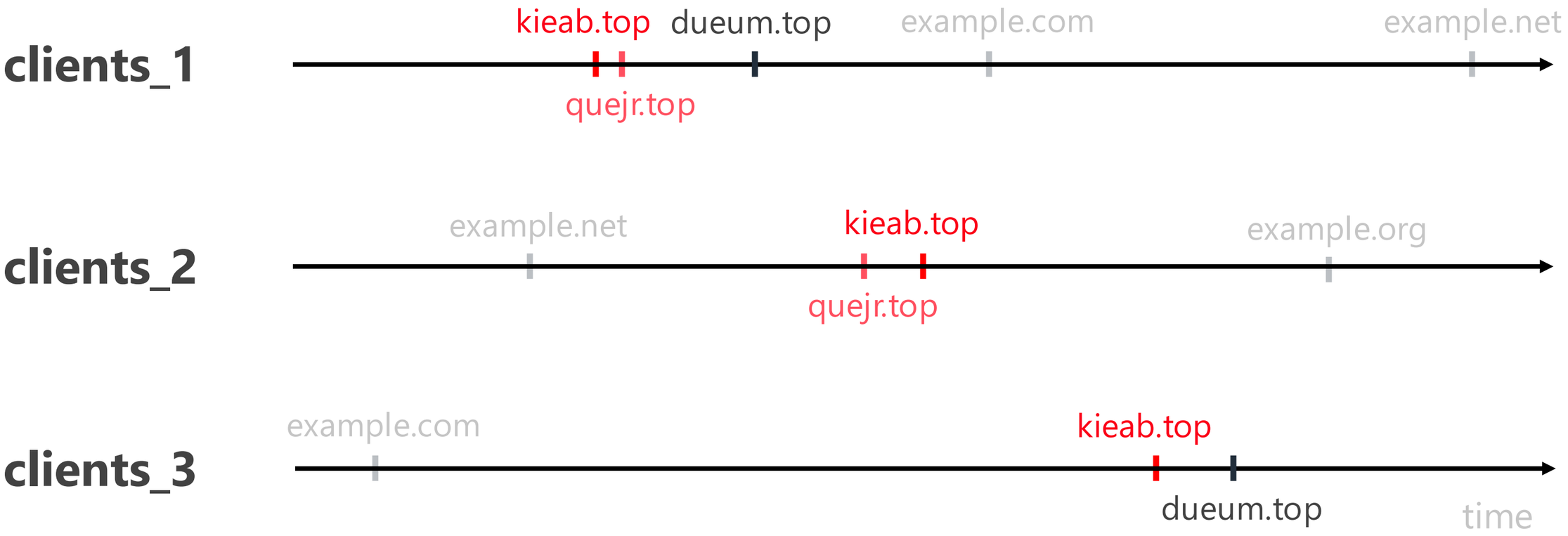
ASN

Other features

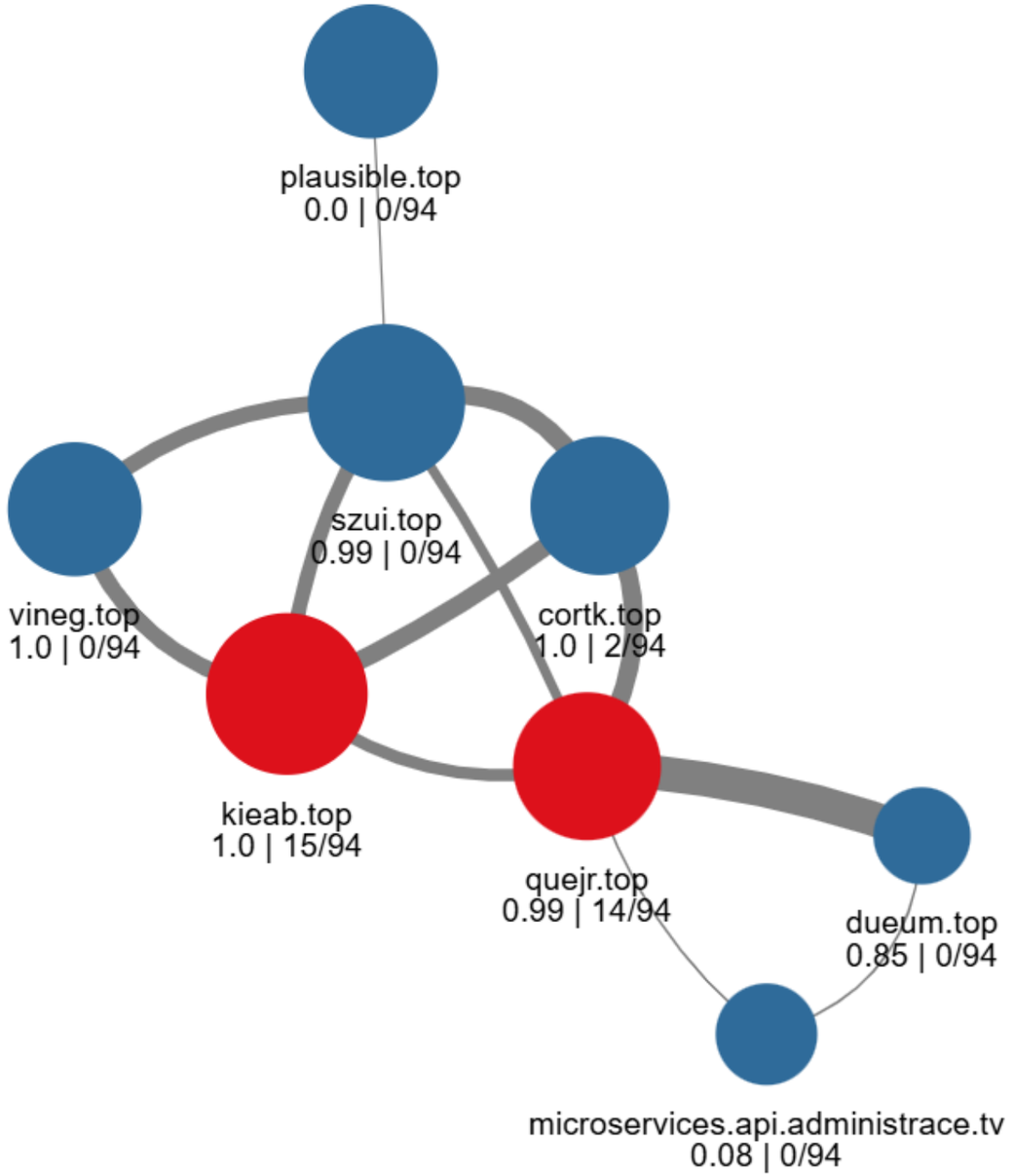
co-occurrence



co-occurrence



co-occurrence



example.com

Domain names

Time related

ASN

Other features

ASN

Autonomous System Number

ASN

Autonomous System Number

ASN **reputation score**

ASN

Autonomous System Number

ASN **reputation score**

work in progres…

Results

3h long analysis (**17.10**, Friday)

Results

3h long analysis (**17.10**, Friday)

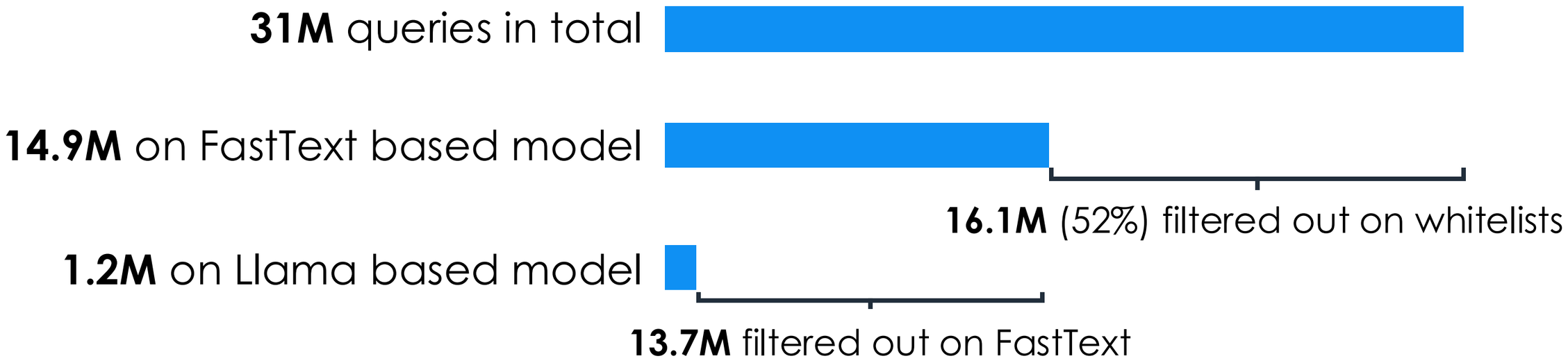
31M queries in total

Results

3h long analysis (**17.10**, Friday)

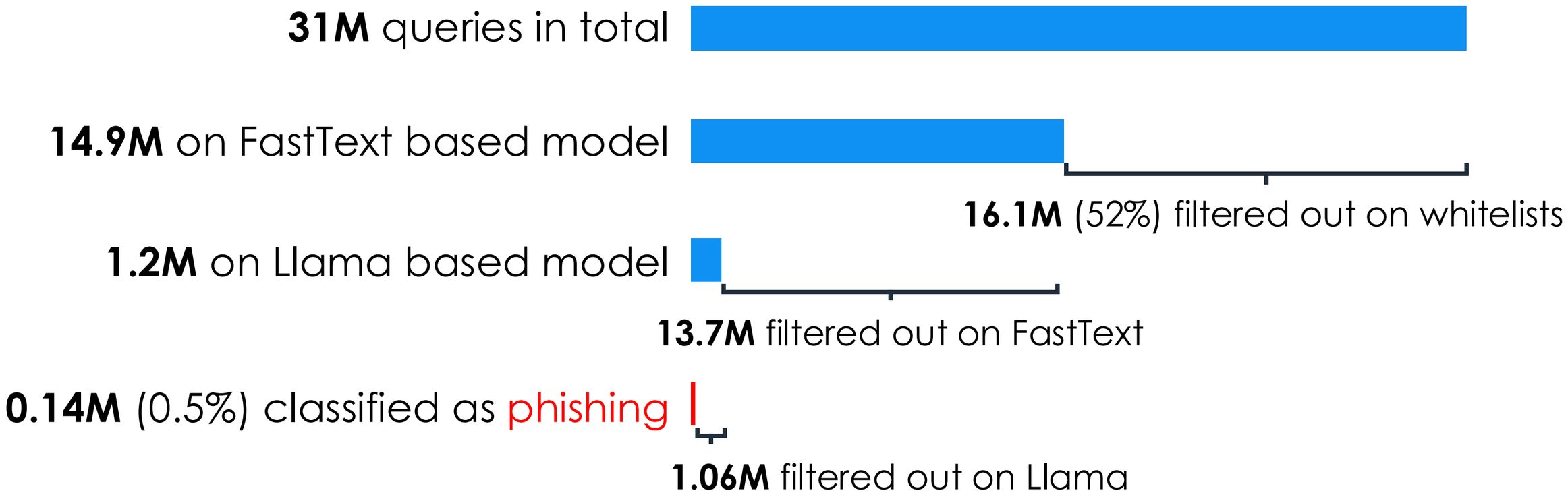
Results

3h long analysis (**17.10**, Friday)



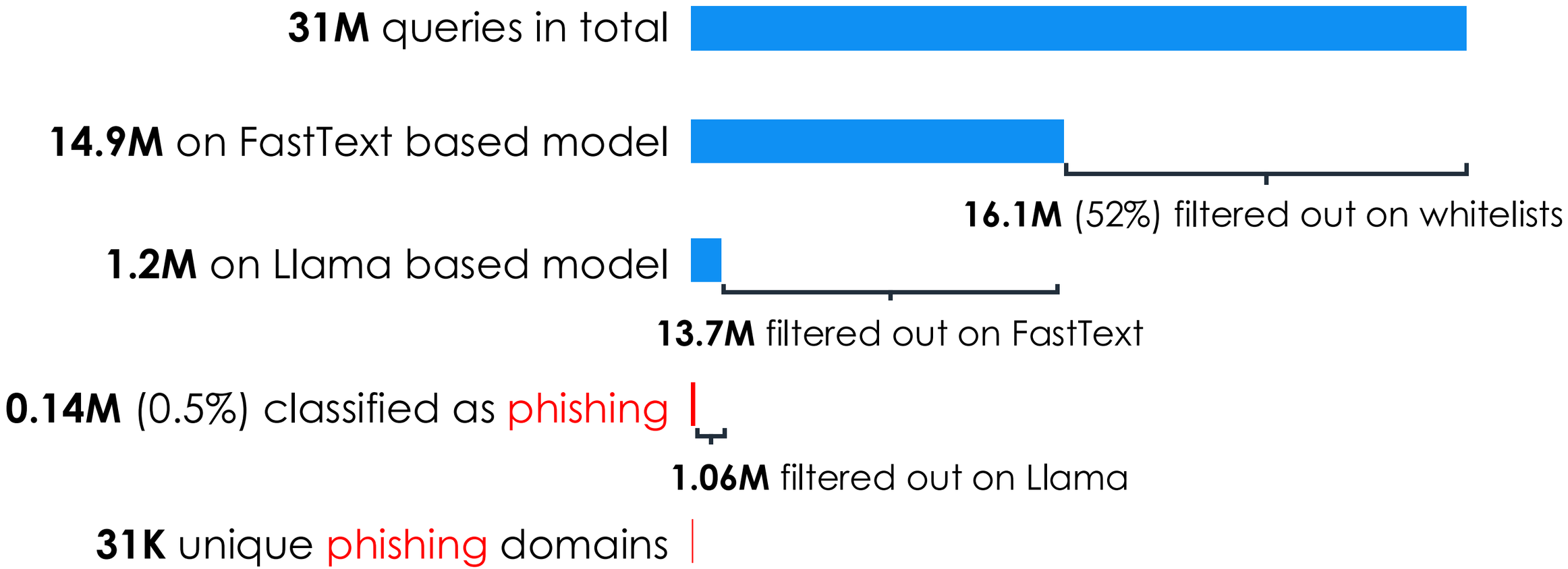
Results

3h long analysis (**17.10**, Friday)



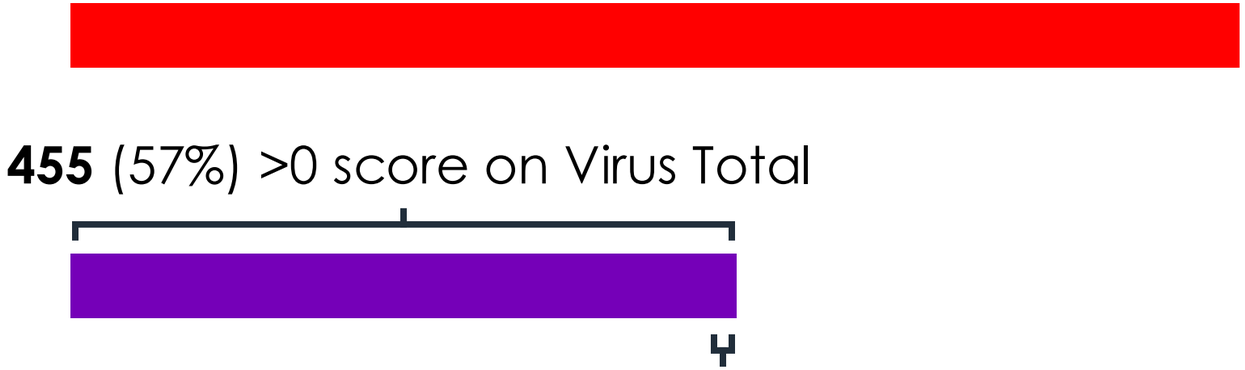
Results

3h long analysis (**17.10**, Friday)



Results

Random **800 elements** sample analysis



455 (57%) >0 score on Virus Total

including **9** in a high confidence
additional phishing domains data source

Challenges

Domain names approach: **TLD balancing**

Non-phishing	phishing
domain1.tld	phish1.tld
domain2.tld	phish2.tld
	phish3.tld
	phish4.tld
	phish5.tld
	phish6.tld
	…

Challenges

Domain names approach: **TLD balancing**

Non-phishing	phishing
domain1.tld	phish1.tld
domain2.tld	phish2.tld
	phish3.tld
	phish4.tld
	phish5.tld
	phish6.tld
	…

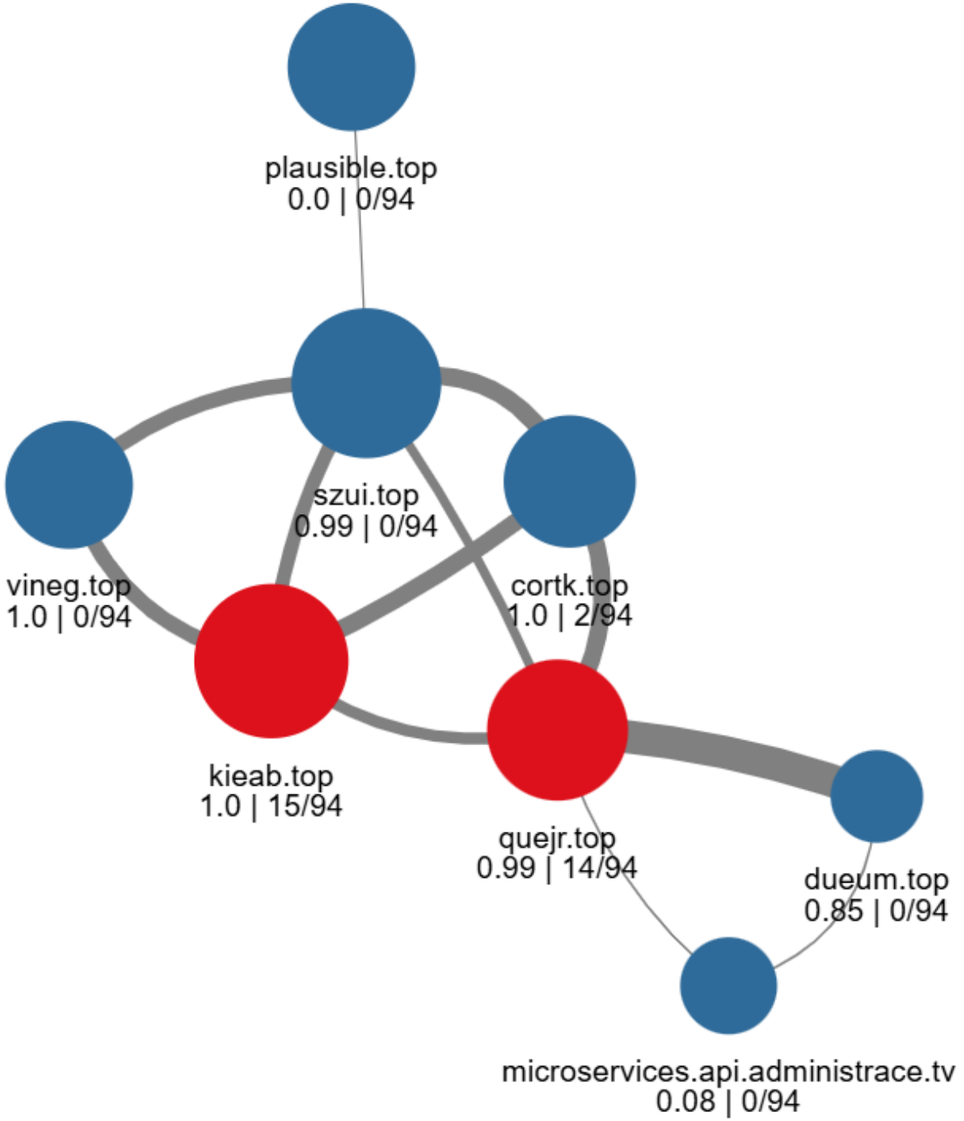
Challenges

Domain names approach: **TLD balancing**

Non-phishing	phishing
domain**1**.tld	phish1.tld
domain**2**.tld	phish2.tld
domain**1**.tld	phish3.tld
domain**2**.tld	phish4.tld
domain**1**.tld	phish5.tld
domain**2**.tld	phish6.tld
…	…

Challenges

Co-occurrence: **difficult to automate**



Challenges

– Dealing with **a lot of** streaming **data**

Challenges

- Dealing with **a lot of** streaming **data**

- Hard to **check** (and proof) whether detected domains are **truly phishing**

Challenges

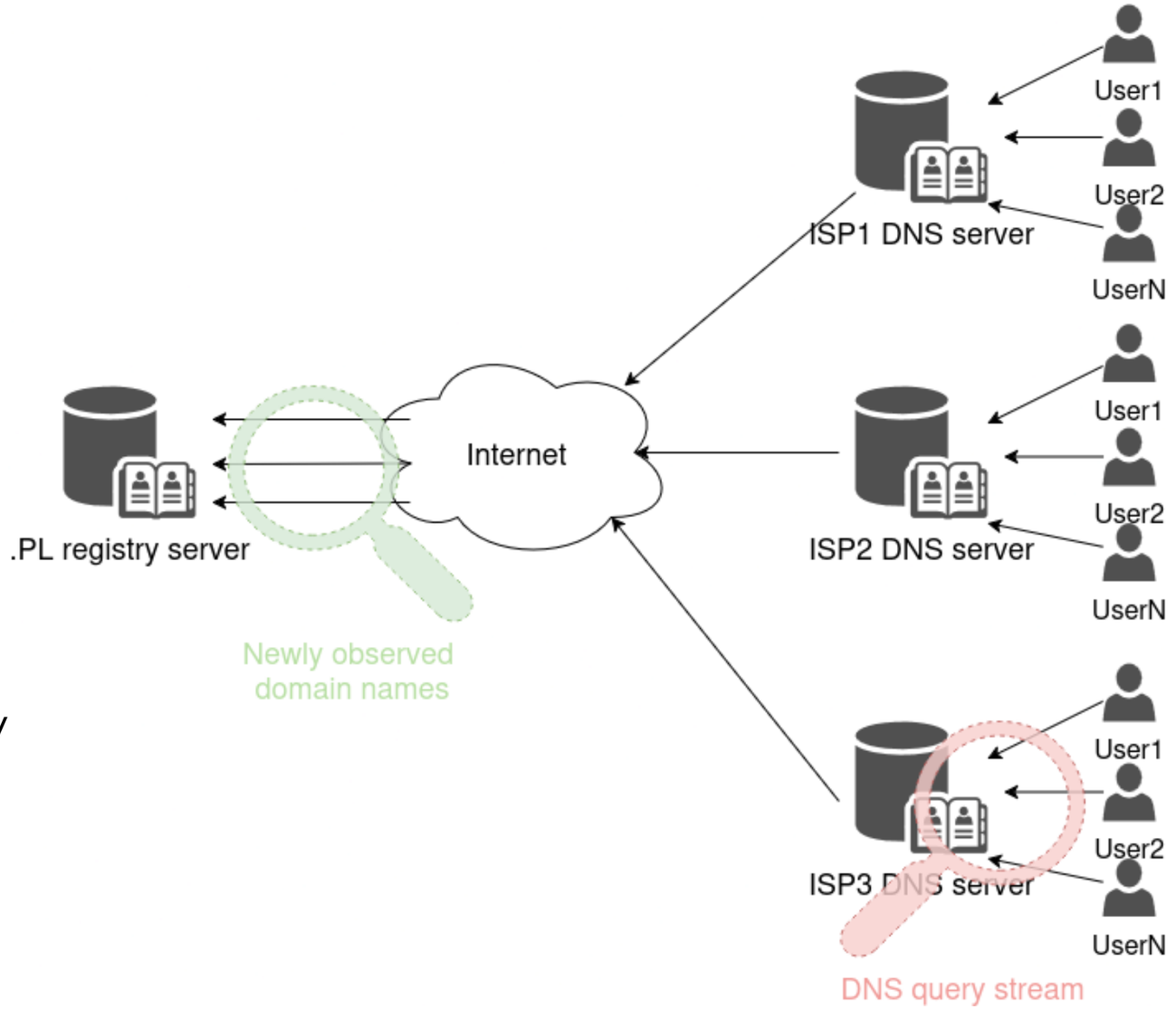
- Dealing with **a lot of** streaming **data**
- Hard to **check** (and proof) whether detected domains are **truly phishing**
- Status: moving to **production environment**

.pl DNS servers

Monitoring DNS requests for newly observed domain names

Approach

- Monitoring DNS requests at .PL registry server level
- Collecting data only about domain names not seen previously
- Advantage: observing subdomains not seen in other sources, e.g., CT log, registry data



Newly observed domain names

- Collecting tuples for all DNS records
- Domain names with various number of sublevels
- Hunting for:
  - keywords, suffixes, regexes
  - subdomains of specific domains
- Cooperation with other CERT.PL systems, incl. Baddomains-ML

FQDN	Record type	First Seen	Last Seen
sub1.example.pl	A	1750765434	1750851789

Newly observed domain names vs CT logs

- No subdomains visibility for wildcard certificates in CT logs

- Newly observed domain names can uncover other subdomains from a hunting set

9527ukfs486yvre.hb20[.]pl

The only subdomain of hb20[.]pl
on the Warning List

7fbn34olxcm6avw.hb20[.]pl
h190uao6wcrkl2i.hb20[.]pl

Subdomains from the NOD monitor
Candidates for blocking
Strong indicator for *hb20[.]pl* blocking

Lessons learned

ML classifier on registry data

- Early detection possible: at registration

- **If significant number of malicious domains: good results**

ML classifier on registry data

- Early detection possible: at registration
- If significant number of malicious domains: good results
- **If low number of malicious domains: problems with getting detections**

ML classifier on query stream

- Just looking at domain names gives quite good results
- Challenge: using other features for detection: work ongoing
- Sufficient scale/coverage needed to see dynamic behaviors

Challenges

- Evaluation: difficult when we go beyond threats targeting Poland

- ML models identify **suspicious** domains

- … but we need proof when using as high confidence source = full URL and defeat

  cloaking (outside scope of this talk)

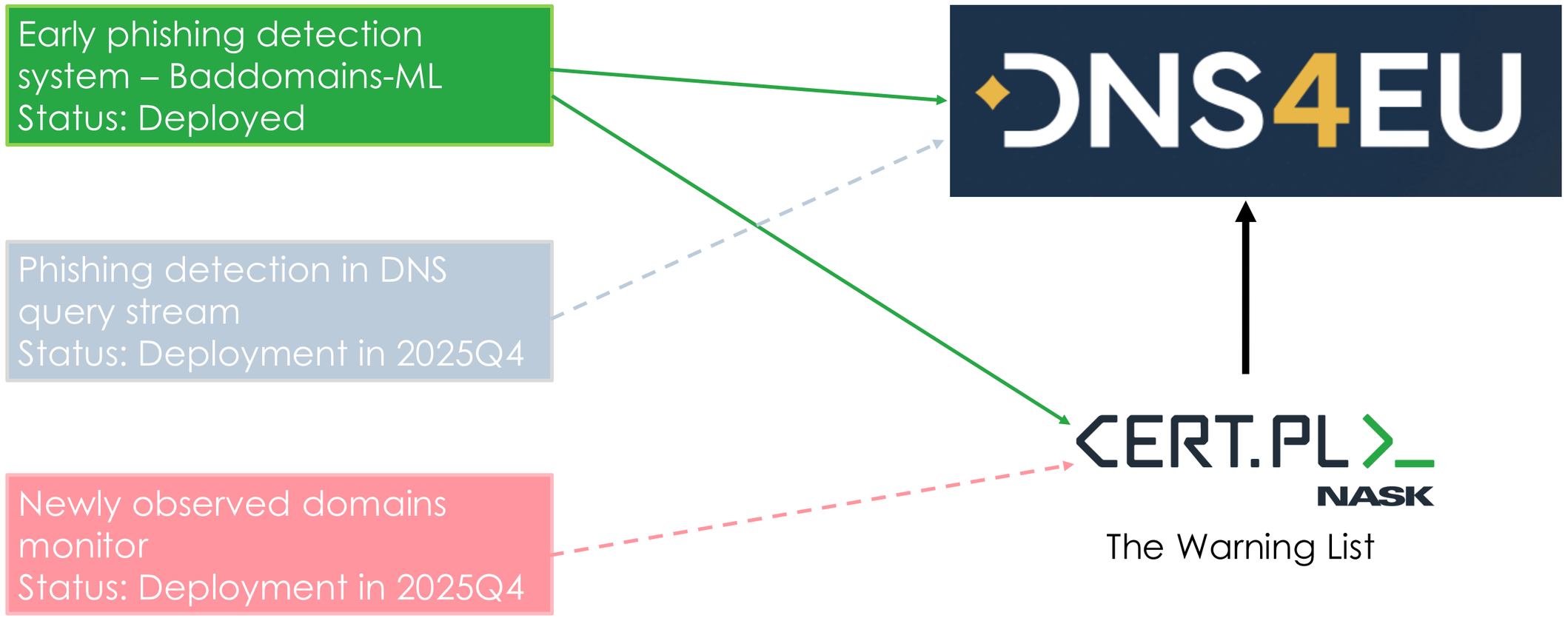
Data from TLD servers

- Underutilized data source for domain hunting

- Not everything has to be a ML model

- Moving to a prototype stage

Current status



CERT.PL
NASK

Questions, feedback & ideas welcome!

piotr.bialczak@cert.pl

michal.halon@nask.pl

joindns4.eu

cert.pl