Evaluating the impact of design decisions on passive DNS-based domain rankings

Victor Le Pochat, Simon Fernandez, Tom Van Goethem, Samaneh Tajalizadehkhoob, Lieven Desmet, Andrzej Duda, Wouter Joosen, Maciej Korczyński

TMA 2024, 23 May 2024

Evaluating the impact of design decisions on passive DNS-based domain rankings

What is a **domain** *(or top sites)* **ranking**?

- Ranking of most popular websites / domain names

```
1,google.com
2,youtube.com
3,facebook.com
4,a-msedge.net
5,microsoft.com
6,netflix.com
7,akamaiedge.net
8,epicgames.com
9,twitter.com
10,instagram.com
```

https://tranco-list.eu/list/LY344/10

What is a **domain ranking**?

- **Essential data source**: sampling the Internet
  - Over a thousand studies rely on them

- Potential **impact** on measurements and findings
  - **Issues**: opaque construction methods, undesirable properties, difficult to reproduce, limited for some use cases

Scheitle et al. A Long Way to the Top: Significance, Structure, and Stability of Internet Top Lists. IMC ’18.
Le Pochat et al. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. NDSS ‘19.
Ruth et al. Toppling Top Lists: Evaluating the Accuracy of Popular Website Lists. IMC ’22.
Ruth et al. A World Wide View of Browsing the World Wide Web. IMC ’22

How do we **rank domains**?

- ***Web traffic***: reported by browsers or in-page scripts
  - Alexa †, Quantcast †, Chrome User Experience Report
- ***Passive DNS traffic***: collected at DNS resolvers
  - Webshrinker DNSFilter, Cloudflare Radar, SecRank, Cisco Umbrella

How do we **rank domains**?

- ***Web traffic***: reported by browsers or in-page scripts

Gradual shift from web to DNS traffic:
- Challenging to recruit users for sharing web traffic
- Privacy challenges for processing browser traffic

- ***Passive DNS traffic***: collected at DNS resolvers

Evaluating the impact of design decisions on passive DNS-based domain rankings

Are DNS-based rankings *appropriate* and *reliable* for Internet/web measurements?

Passive DNS-based ranking **(dis-)advantages**

+ Easier to get large **user base**
+ Diverse range of **providers**
+ Better preserve **user privacy**
+ More willing to be **shared**
+ **Raw data** better available
+ Additional DNS **records**

– **Mix** browser visits with background traffic

– **Selection** of resolvers matters

– Some methods **unavailable**

Evaluating the impact of design decisions on passive DNS-based domain rankings

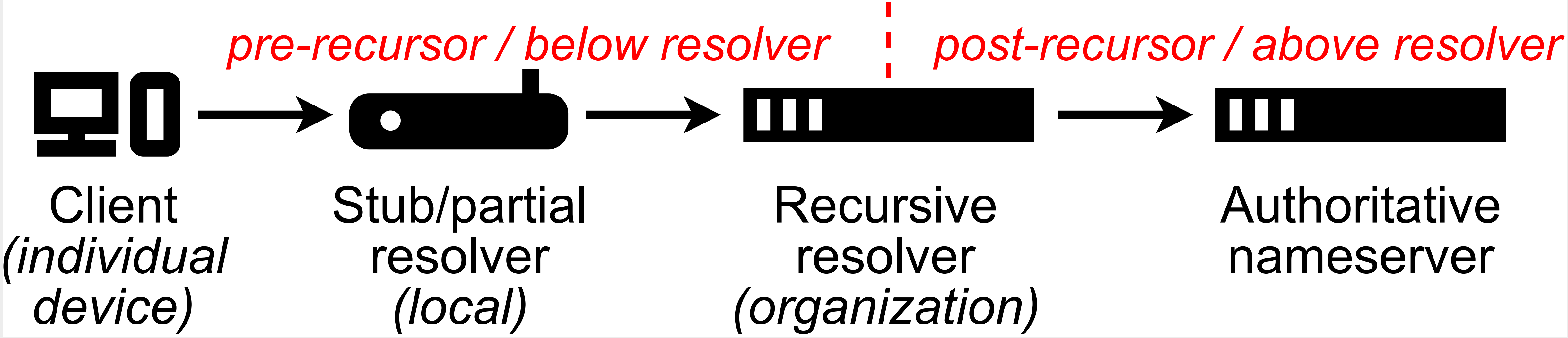
Which design decisions *improve* the *reliability* of (passive DNS-based) rankings?

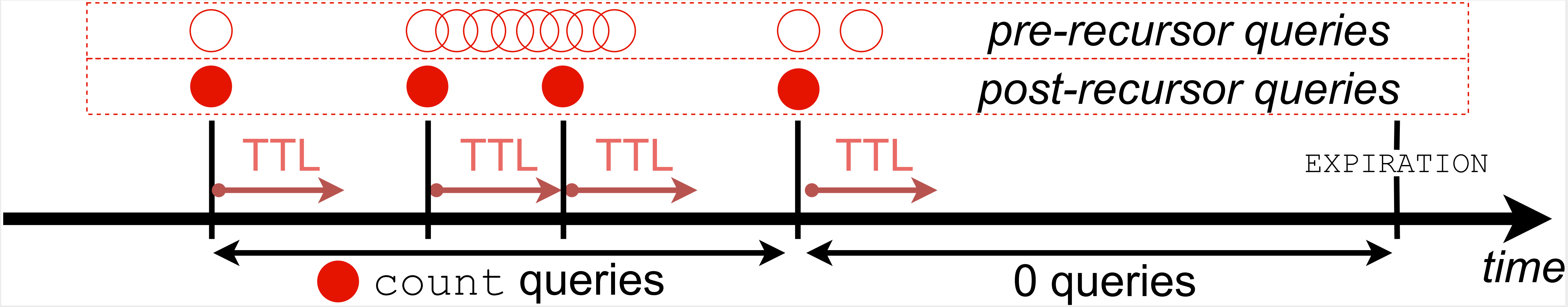
We evaluate the influence of design decisions

- Correcting mechanisms
  - Representativeness → CNAME reverse cache
  - Website *vs.* infrastructure → Service classifier
  - Ranking method → Time-To-Live (TTL)
- Design decisions from recent rankings
  - Individual ranks *vs.* buckets → Bucketing (CrUX, Radar)
  - Time frame of data → Long-term averaging (CrUX, Radar, Tranco)

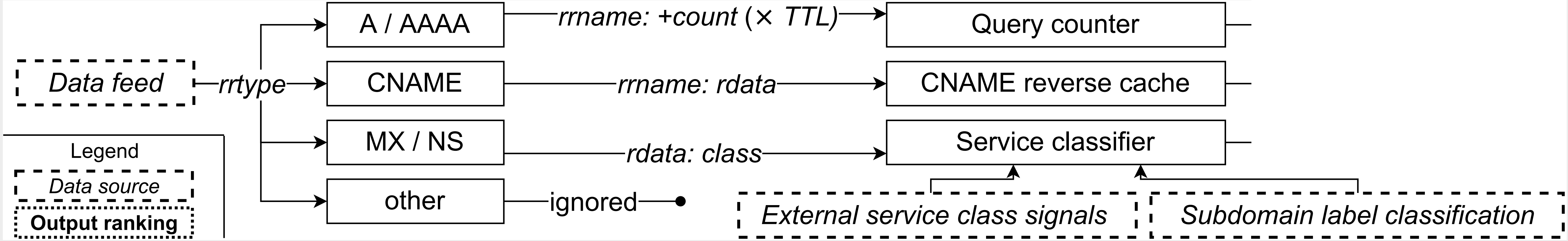
Evaluating the impact of design decisions on passive DNS-based domain rankings

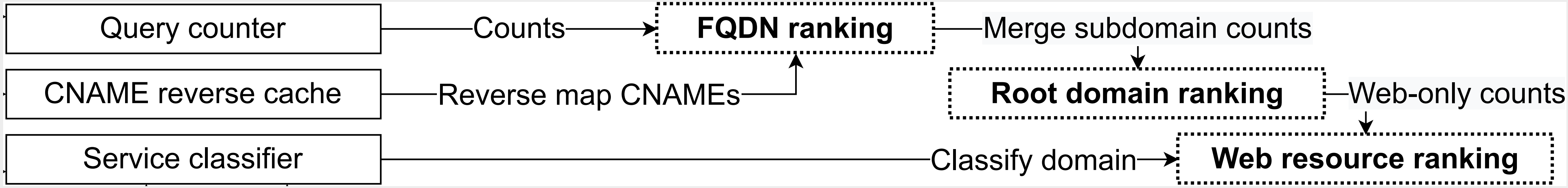
We use **post-recursor** passive DNS data



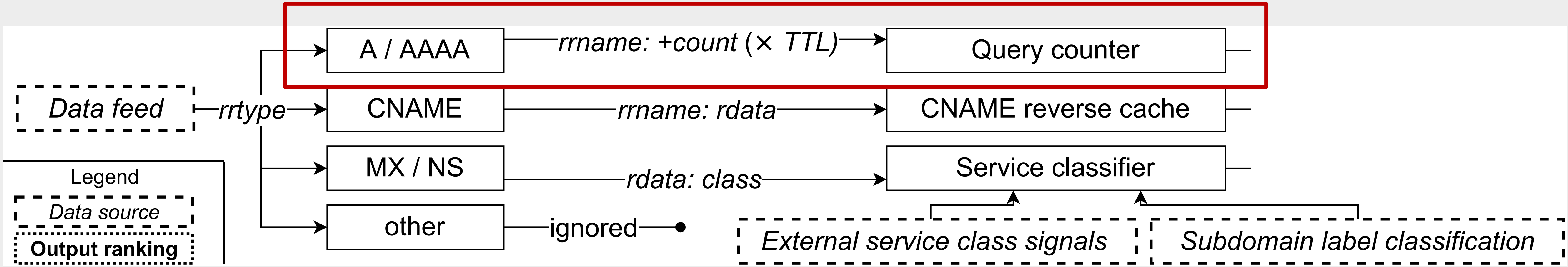


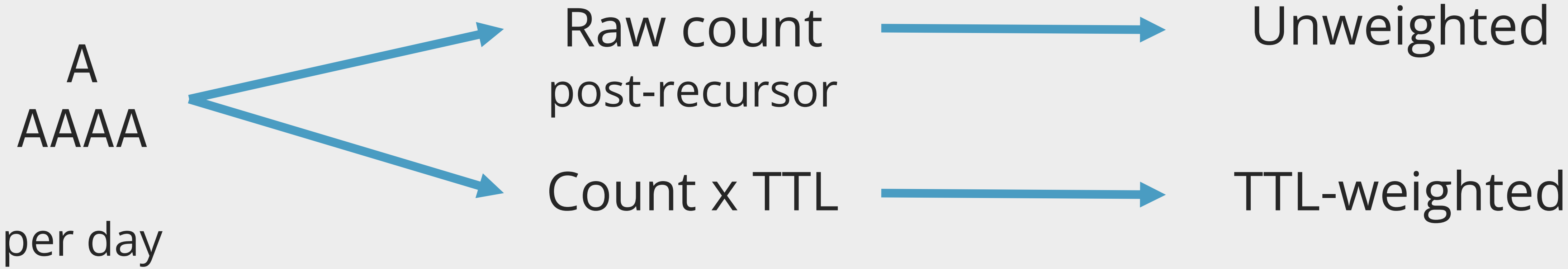
Our own ranking method, to isolate effects



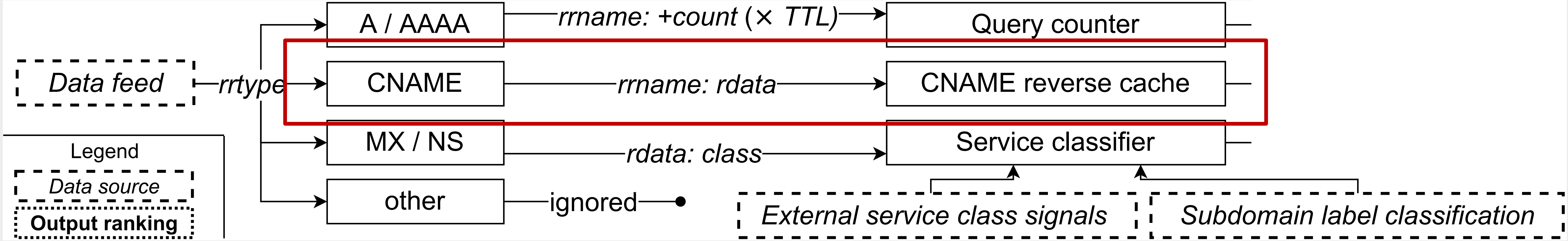


Our own ranking method, to isolate effects





Our own ranking method, to isolate effects



`original.net.   CNAME   example.org.`

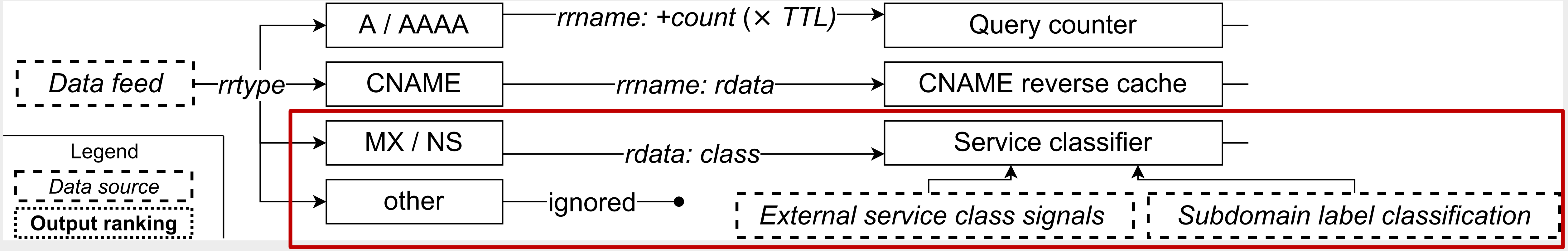
Observed counts

Reverse-mapped to

(if most common mapping)

(if mapping observed enough)

Our own ranking method, to isolate effects



Seen as MX record — *mailserver*

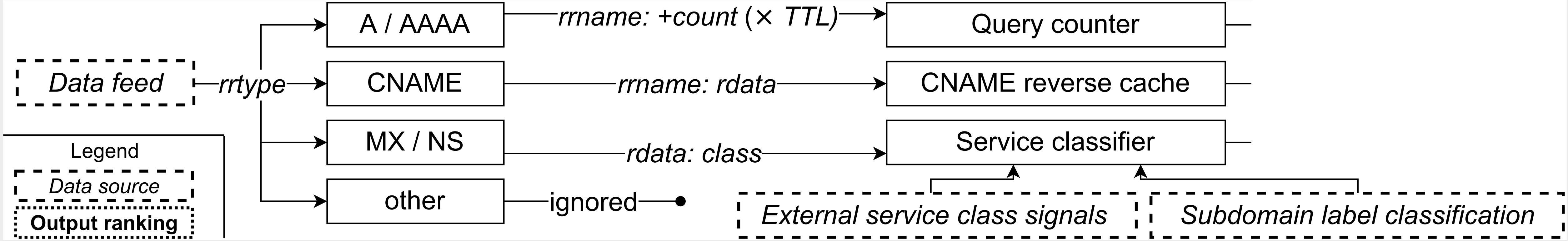
Seen as NS record — *nameserver*

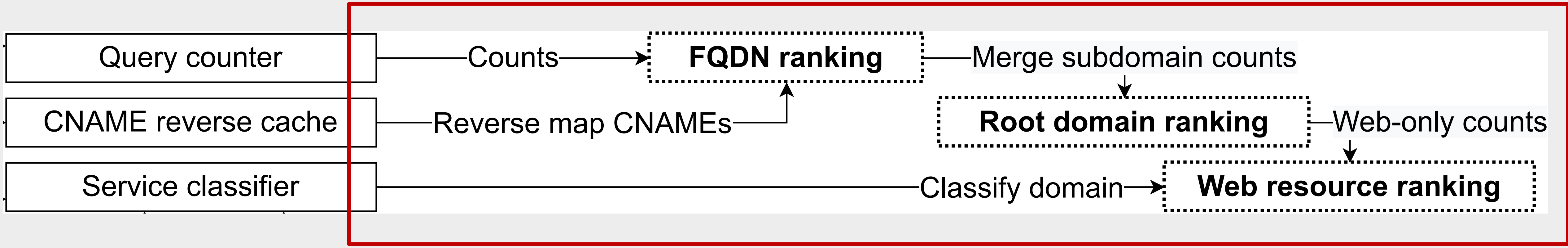
Subdomain label — *www → website*, *ns1 → nameserver*

External sources — *in DBpedia → website*

} Scoring → Most likely class

Our own ranking method, to isolate effects





Limitations

- *Post-recursor:* No true count of client queries / clients
  - Limits available ranking methods
- *Data quality:* Record values are set by domain operators
- *Evaluation:* No ground truth to evaluate accuracy
- *Coverage:* Concrete results only for SIE Europe data

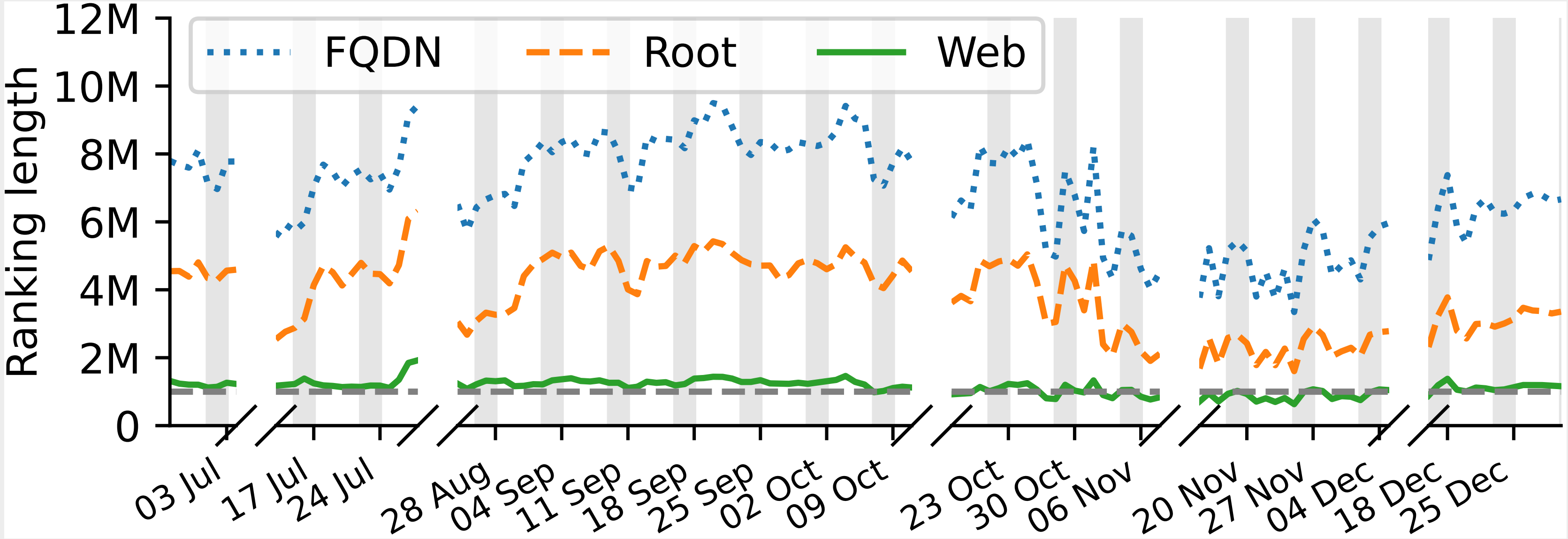
Xie et al. Building an Open, Robust, and Stable Voting-Based Domain Top List. USENIX Security ’22.
Ruth et al. Toppling Top Lists: Evaluating the Accuracy of Popular Website Lists. IMC ’22.

We evaluate the influence of design decisions

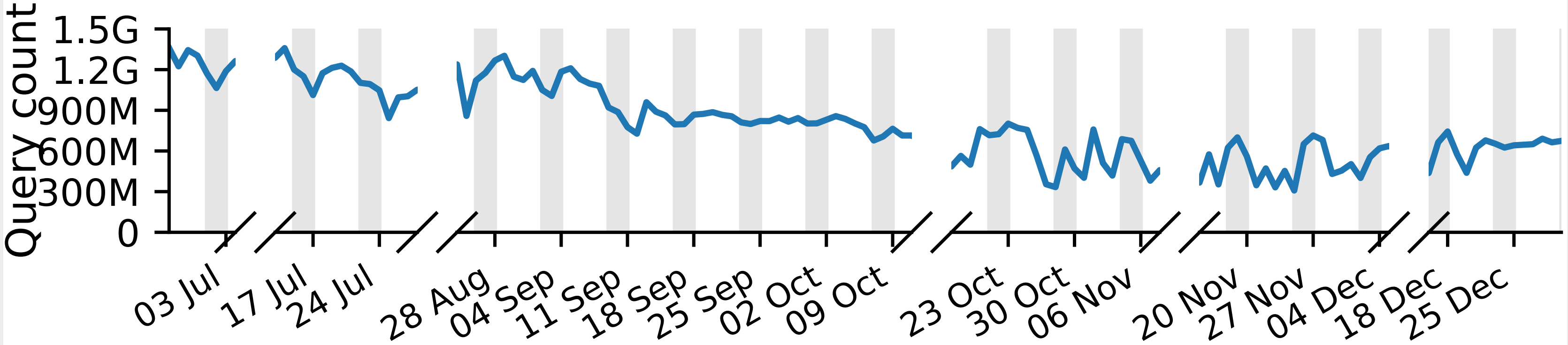
- Correcting mechanisms
  - Representativeness → CNAME reverse cache
  - Website *vs.* infrastructure → Service classifier
  - Ranking method → Time-To-Live (TTL)
- Design decisions from recent rankings
  - Individual ranks *vs.* buckets → Bucketing (CrUX, Radar)
  - Time frame of data → Long-term averaging (CrUX, Radar, Tranco)

Ranking lengths vs. query volumes

1-day

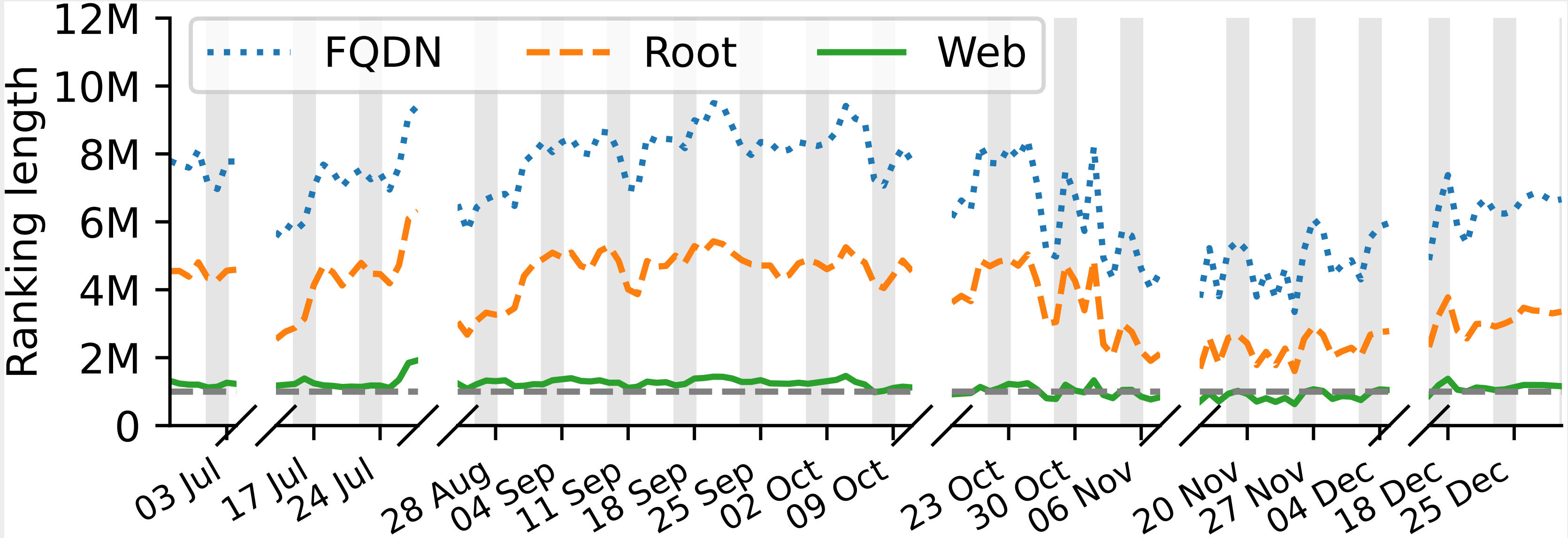


1-day

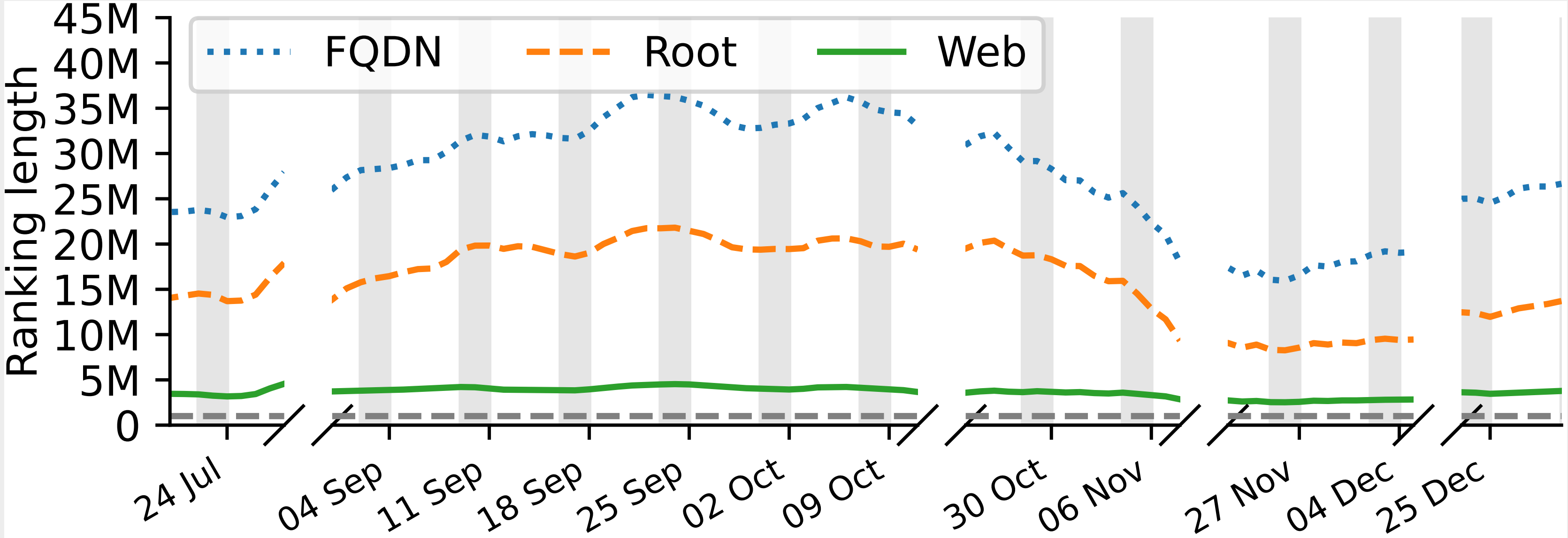


Lengths increase after long-term aggregation

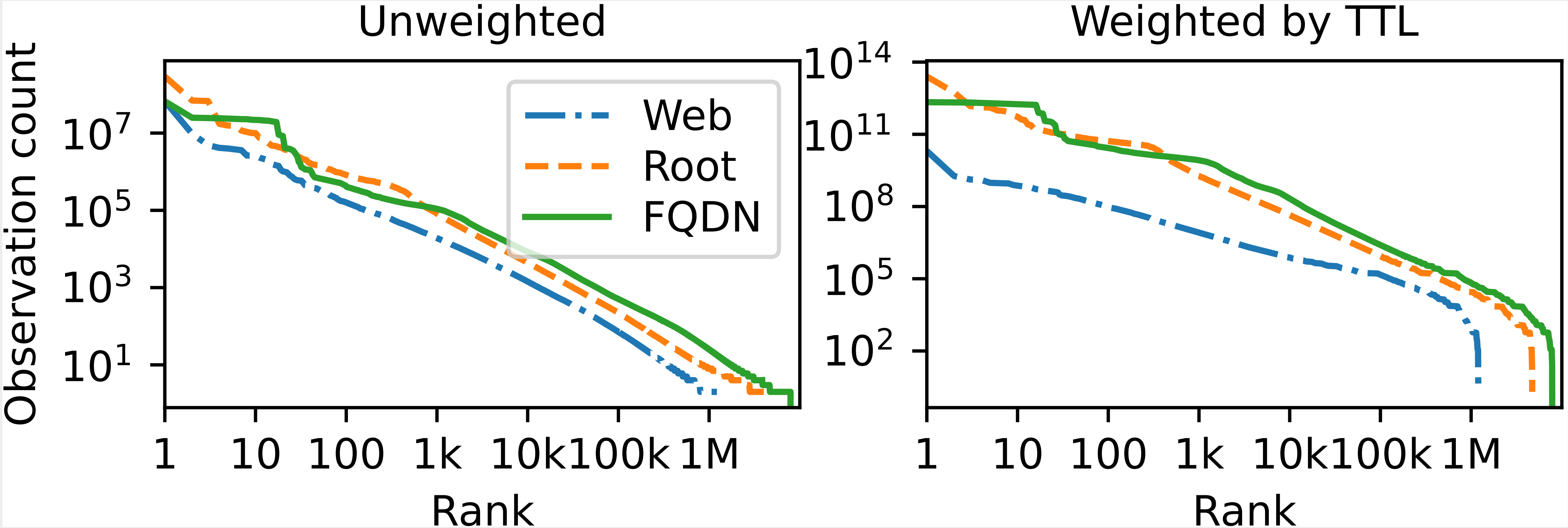
1-day



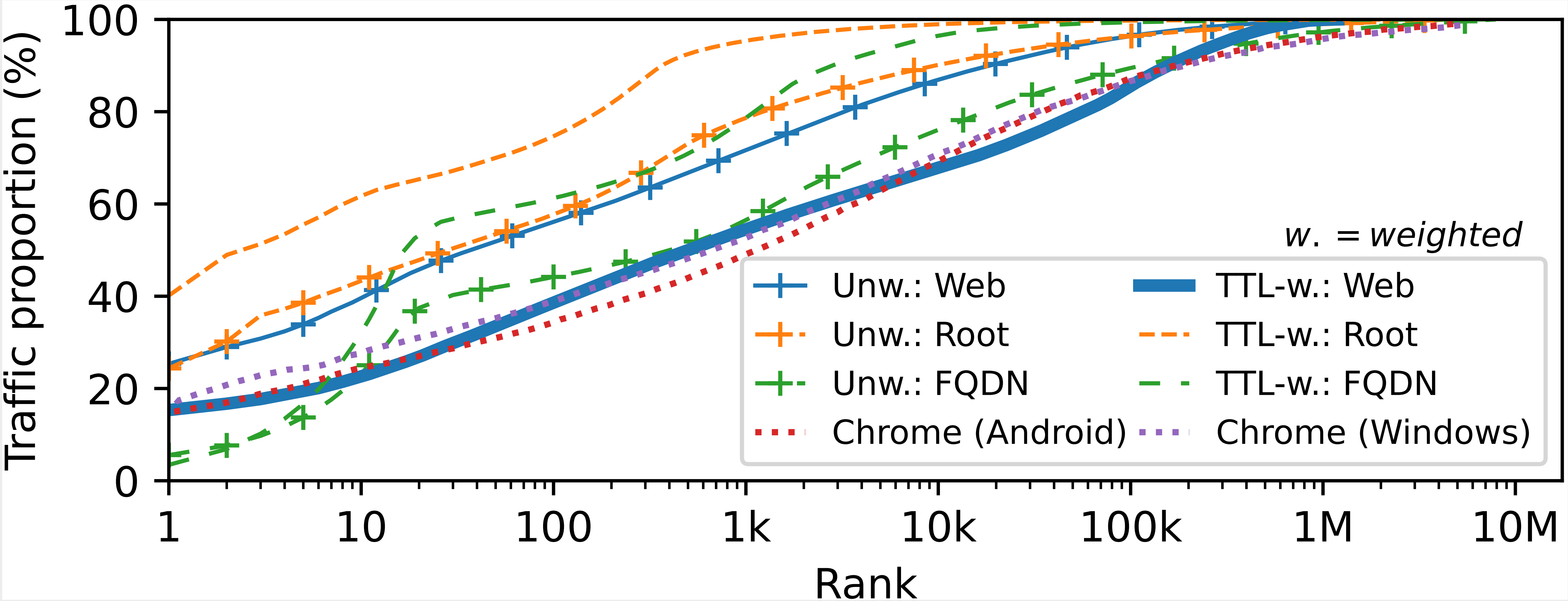
7-day



Observation count distribution follows a power law

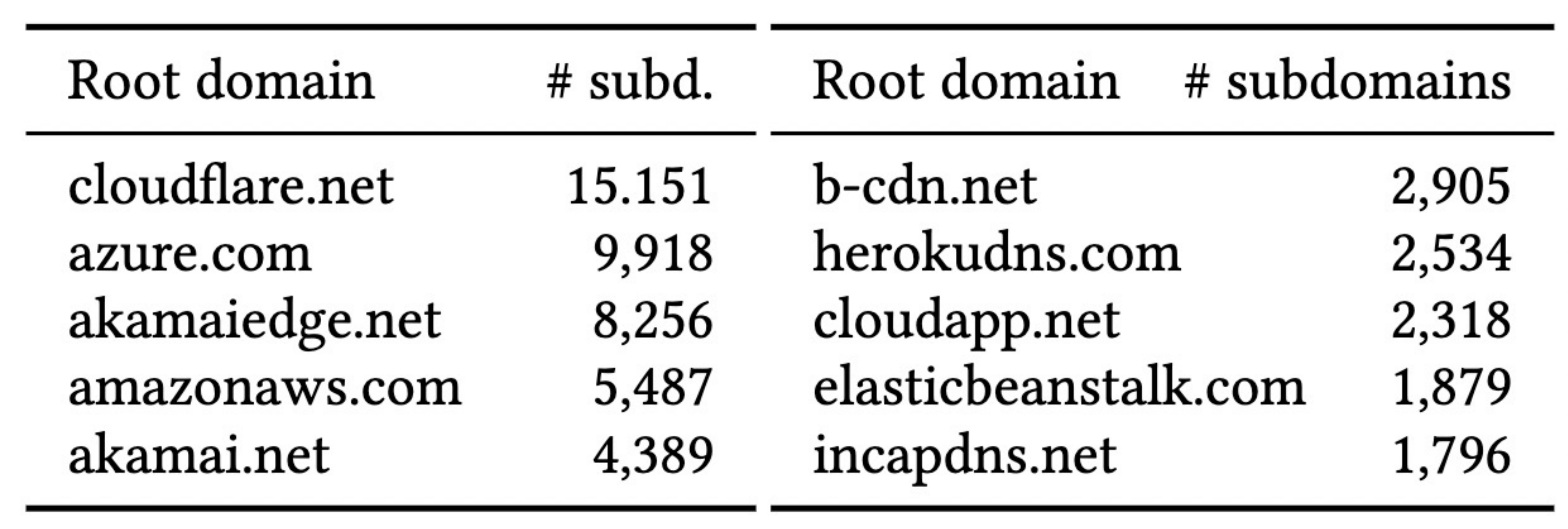


TTL-w. distribution matches Chrome web traffic



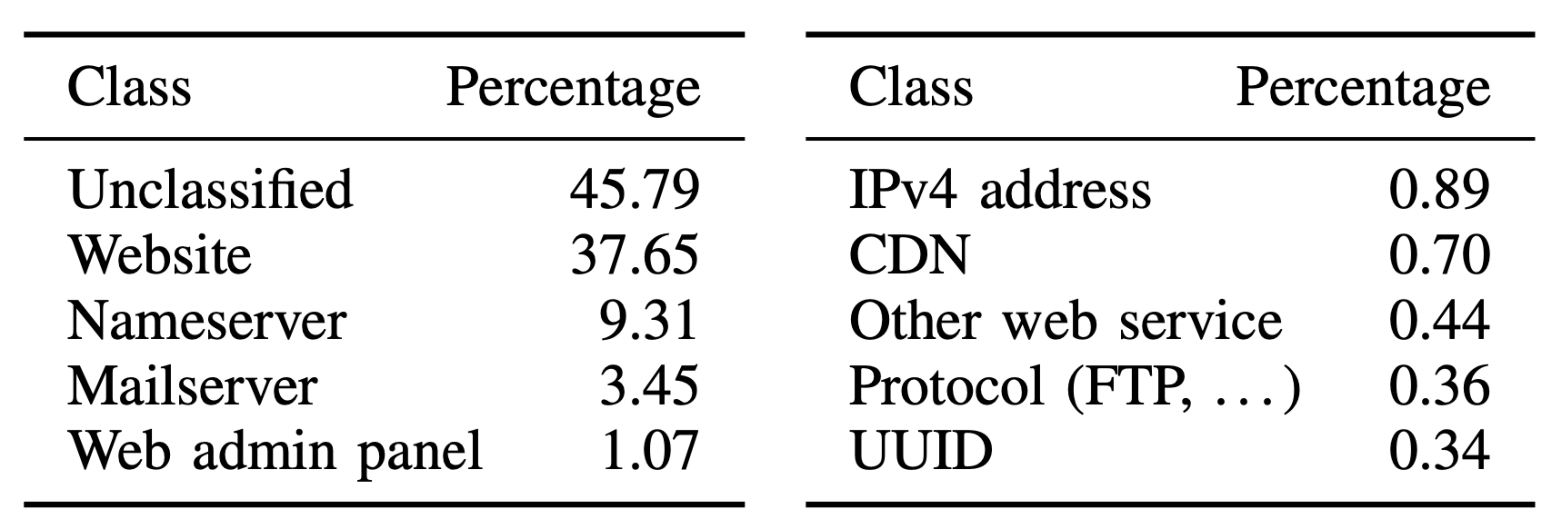
Ruth et al. Toppling Top Lists: Evaluating the Accuracy of Popular Website Lists. IMC ’22.

Correcting mechanisms: CNAME reversal



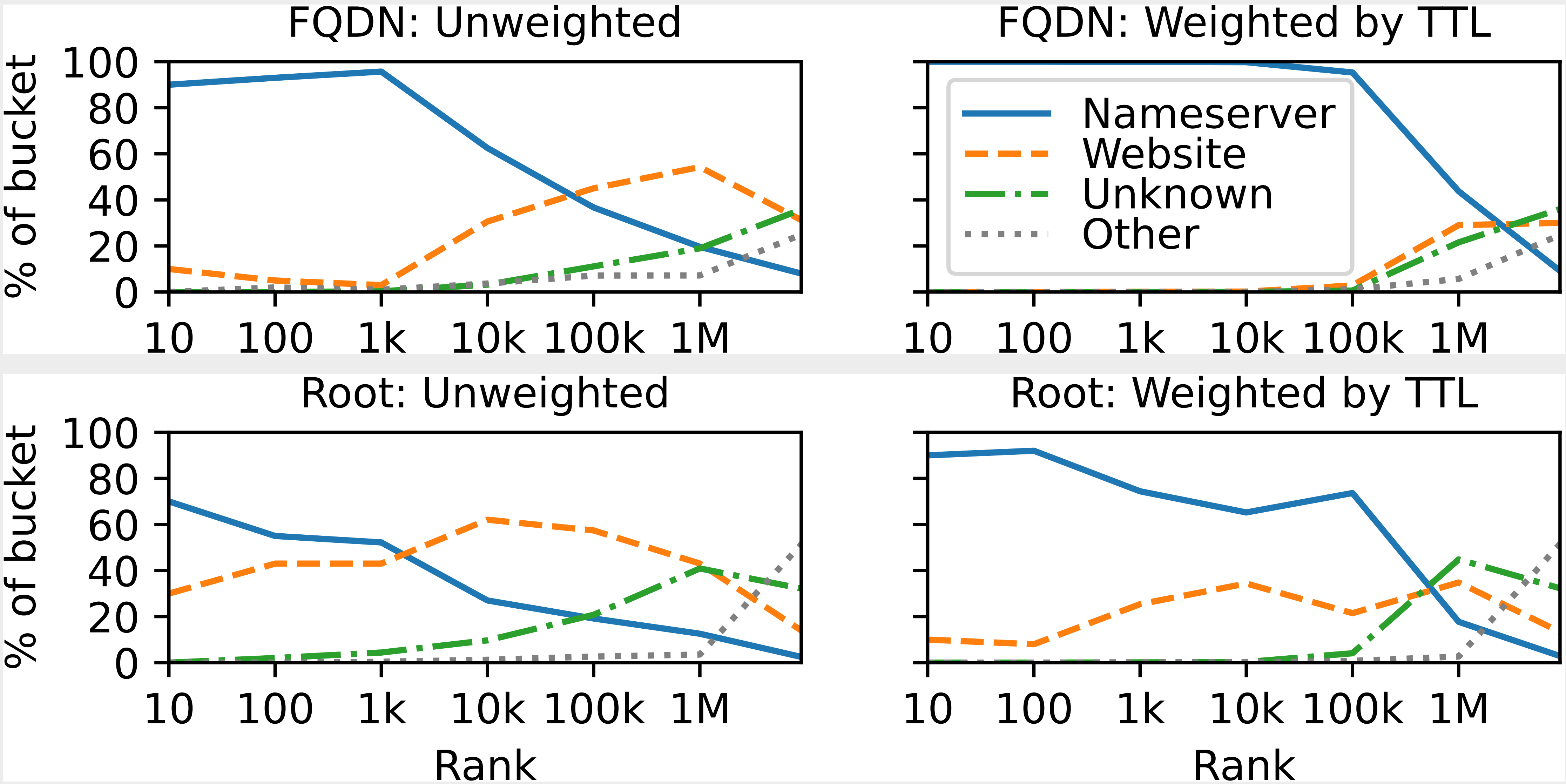
Root domain	# subd.	Root domain	# subdomains
cloudflare.net	15.151	b-cdn.net	2,905
azure.com	9,918	herokudns.com	2,534
akamaiedge.net	8,256	cloudapp.net	2,318
amazonaws.com	5,487	elasticbeanstalk.com	1,879
akamai.net	4,389	incapdns.net	1,796

Correcting mechanisms: Service classification

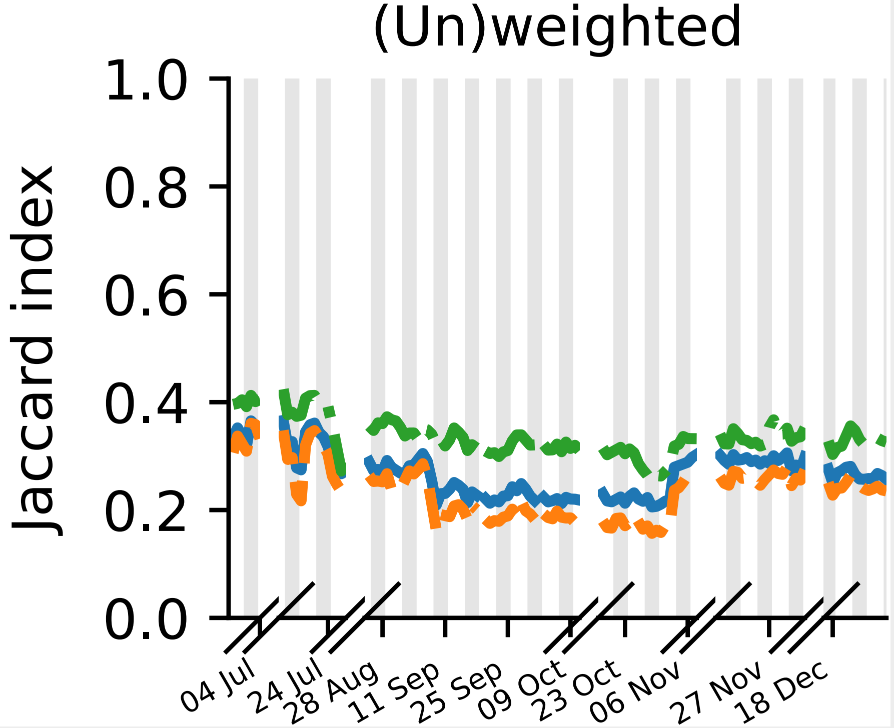


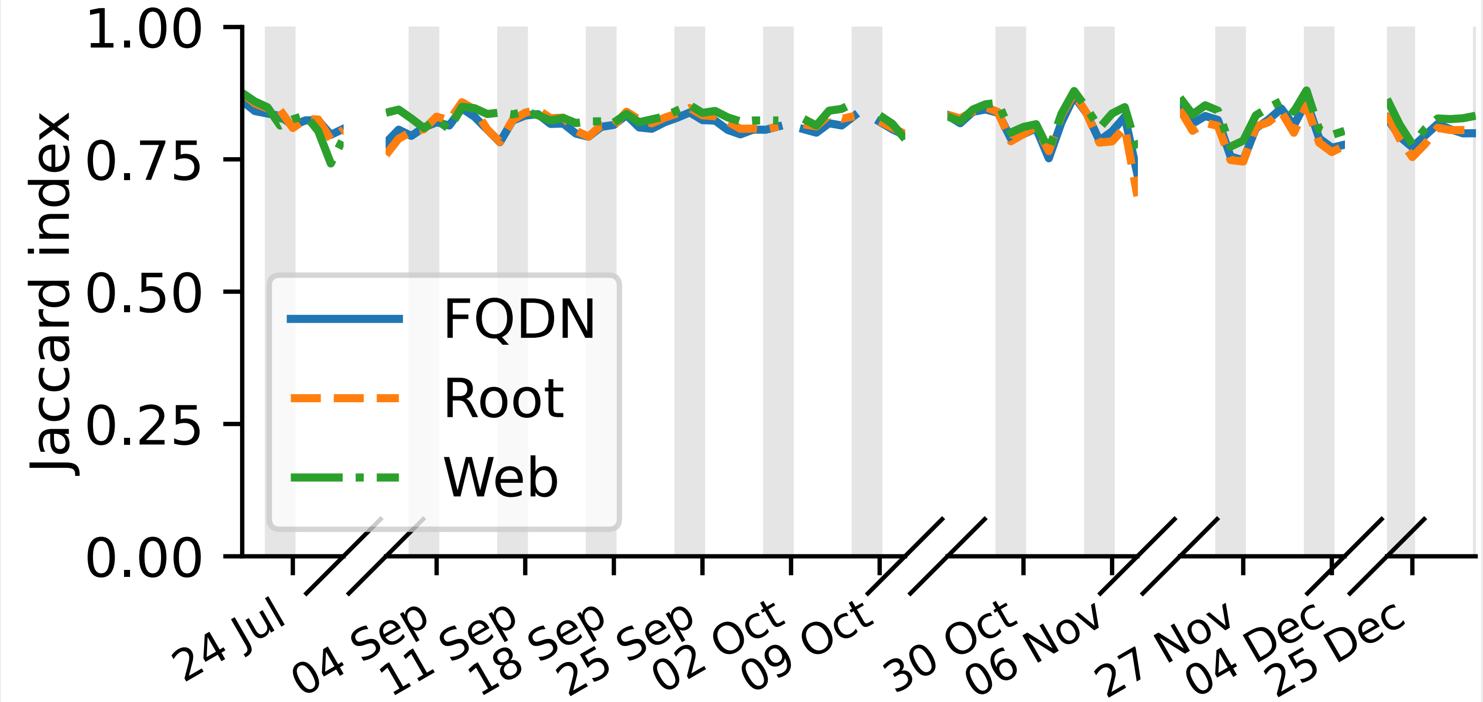
Class	Percentage
Unclassified	45.79
Website	37.65
Nameserver	9.31
Mailserver	3.45
Web admin panel	1.07
IPv4 address	0.89
CDN	0.70
Other web service	0.44
Protocol (FTP, ...)	0.36
UUID	0.34

Nameservers dominate the head of the ranking



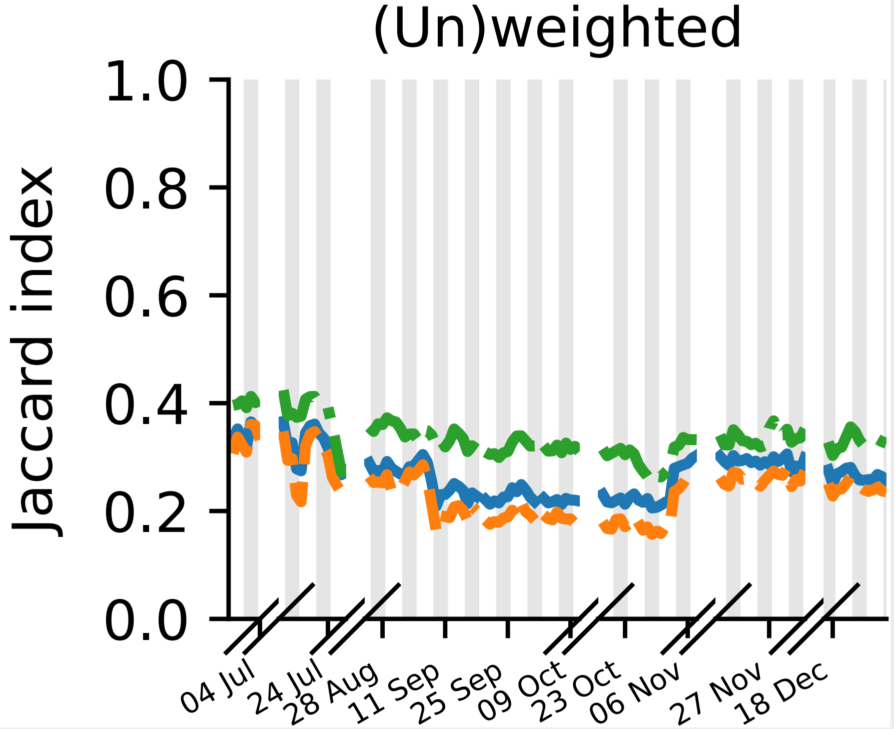
Stability improves with long-term aggregation

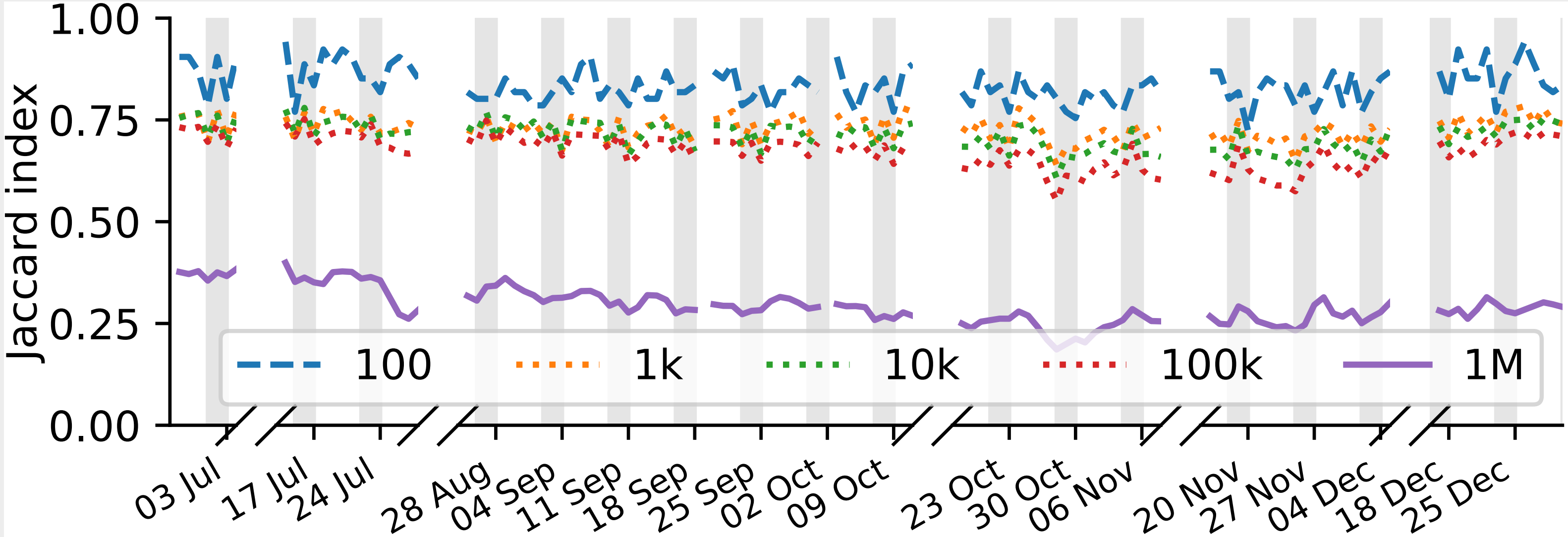




1-day 7-day

Stability improves in buckets at the head





1-day 1-day (buckets)

Evaluating the impact of design decisions on passive DNS-based domain rankings

Discussion & conclusion

- Correcting mechanisms are necessary to avoid dominance
- One design decision can be very impactful
  - Including/ignoring TTL makes a significant difference
  - Reliably comparing rankings across data/methods is challenging
- Buckets & aggregation *(< recent rankings)* improve stability
- **Passive DNS can be used for a reliable (Web) ranking**
- *We should continue evaluating (new) ranking approaches*

Evaluating the impact of design decisions on passive DNS-based domain rankings

https://domain-ranking-design-decisions.distrinet-research.be

victor.lepochat@kuleuven.be https://lepoch.at/
@VictorLePochat