Flow Correlation Attacks on Tor Onion Service Sessions with **Sliding Subset Sum**

Daniela Lopes, Jin-Dong Dong, Pedro Medeiros, Daniel Castro, Diogo Barradas, Bernardo Portela, João Vinagre, Bernardo Ferreira, Nicolas Christin, Nuno Santos

February, 27th, NDSS ‘24

ULISBOA UNIVERSIDADE DE LISBOA

Carnegie Mellon University

UNIVERSITY OF WATERLOO

U.PORTO

People need Tor!

People need Tor!

- **Internet users** face **surveillance** and **censorship**.

People need Tor!

- **Internet users** face **surveillance** and **censorship**.
- **Journalists** and **whistleblowers** need to share information.

People need Tor!

- **Internet users** face **surveillance** and **censorship**.
- **Journalists** and **whistleblowers** need to share information.
- Countries can try to find who they're communicating with.

People need Tor!

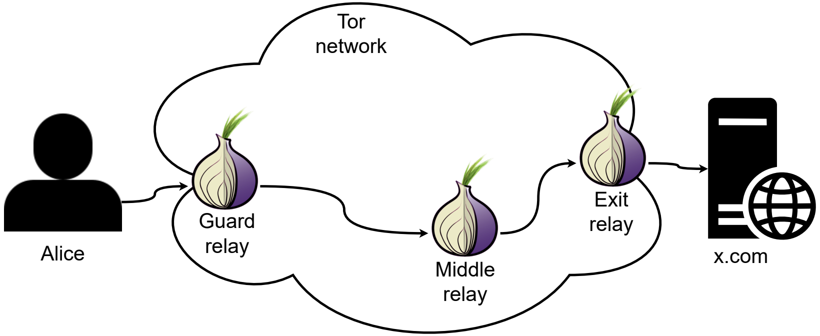
- **Internet users** face **surveillance** and **censorship**.
- **Journalists** and **whistleblowers** need to share information.
- Countries can try to find who they’re communicating with.

- Tor is a network composed of voluntary relays to provide **anonymity**.

Preserving **anonymity** with Tor

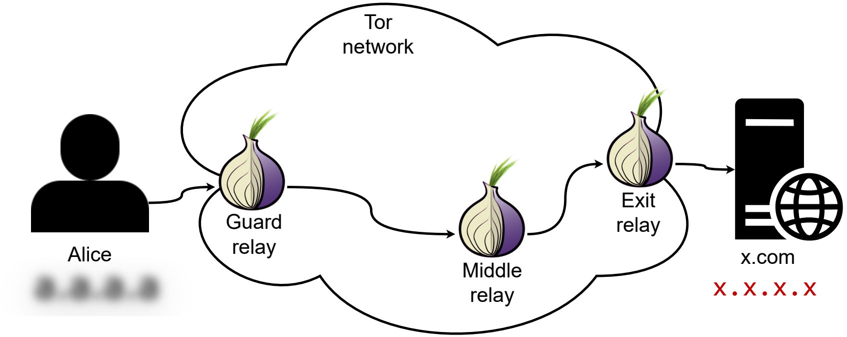
Preserving anonymity with Tor

Circuits to the Internet:



Preserving anonymity with Tor

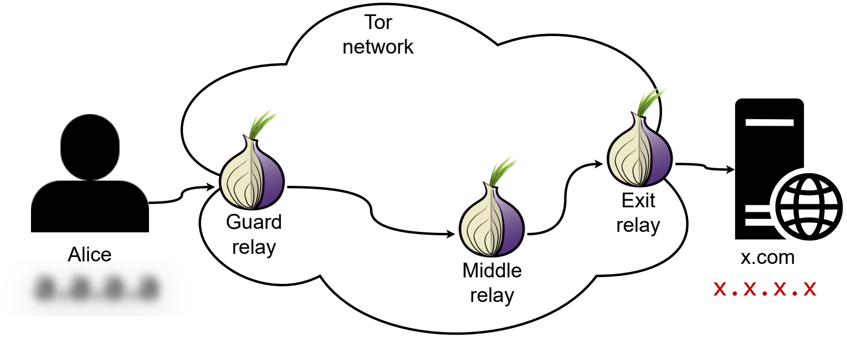
Circuits to the Internet:



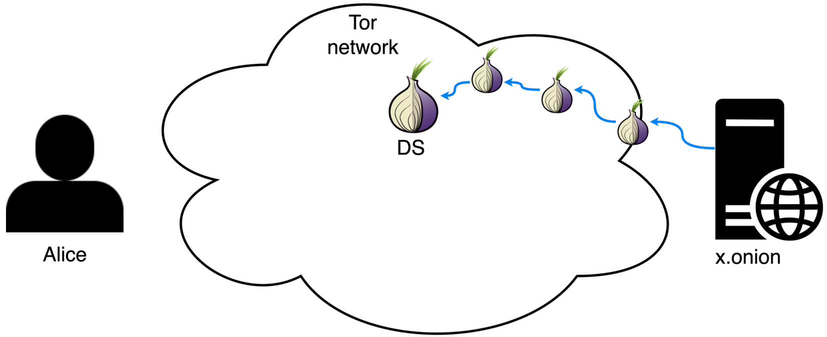
Client-side anonymity

Preserving anonymity with Tor

Circuits to the Internet:



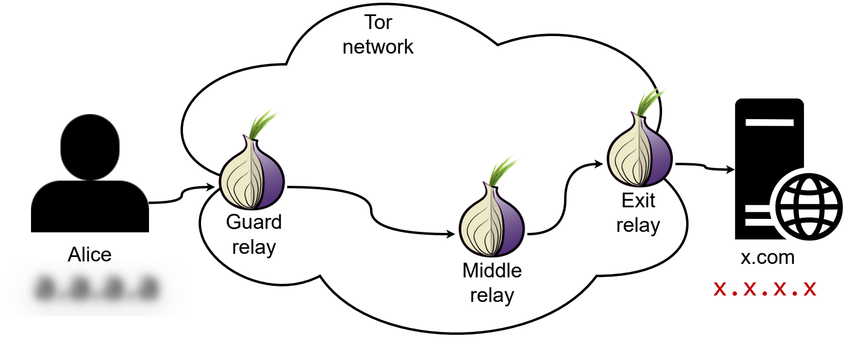
Circuits to onion services:



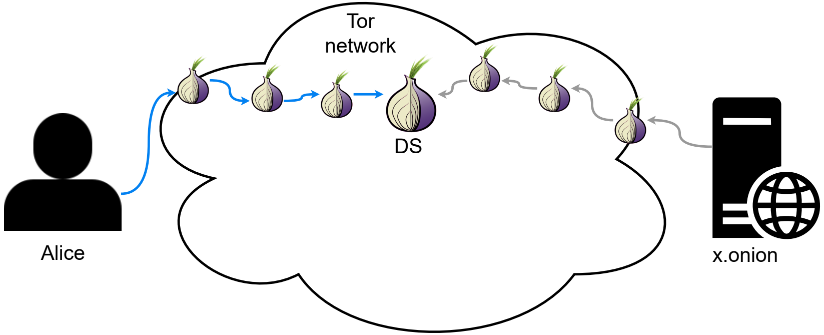
Client-side anonymity

Preserving anonymity with Tor

Circuits to the Internet:



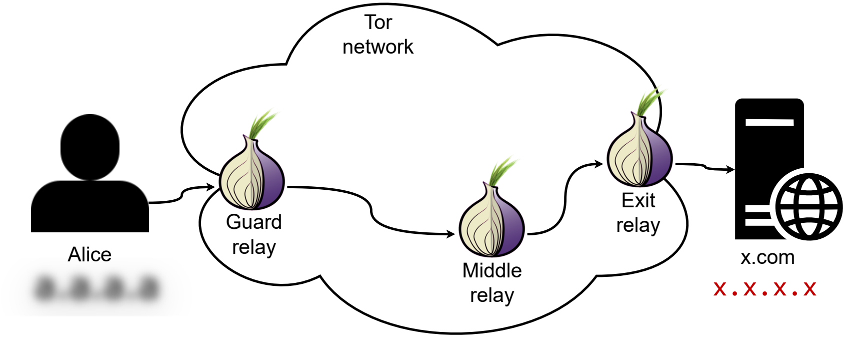
Circuits to onion services:



Client-side anonymity

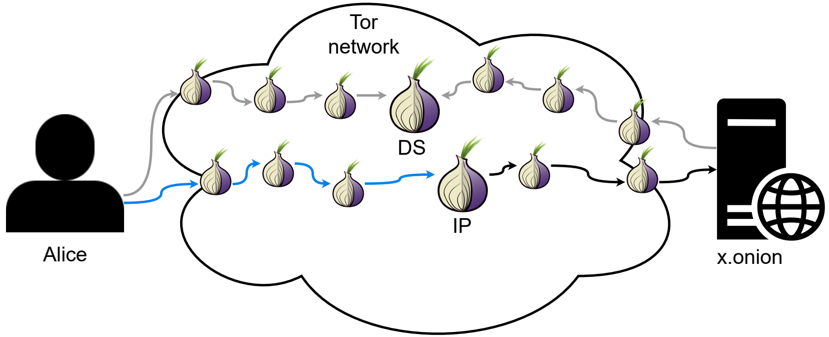
Preserving anonymity with Tor

Circuits to the Internet:



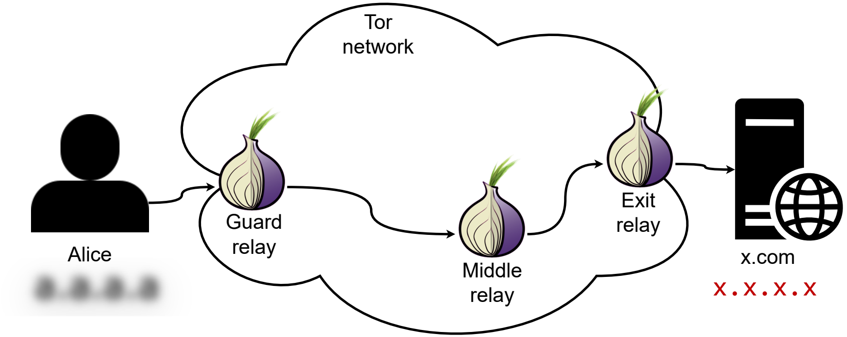
Client-side anonymity

Circuits to onion services:

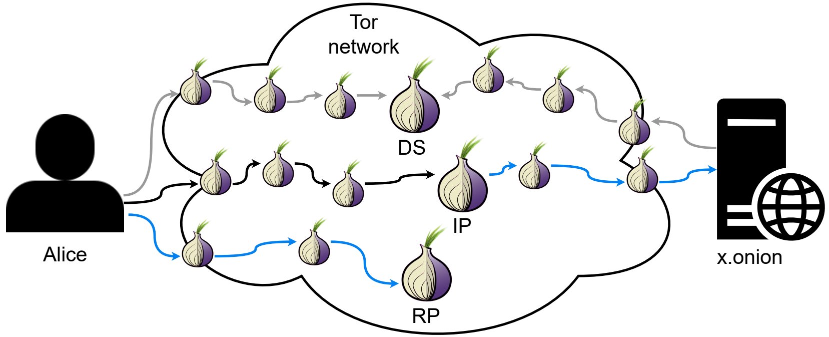


Preserving anonymity with Tor

Circuits to the Internet:



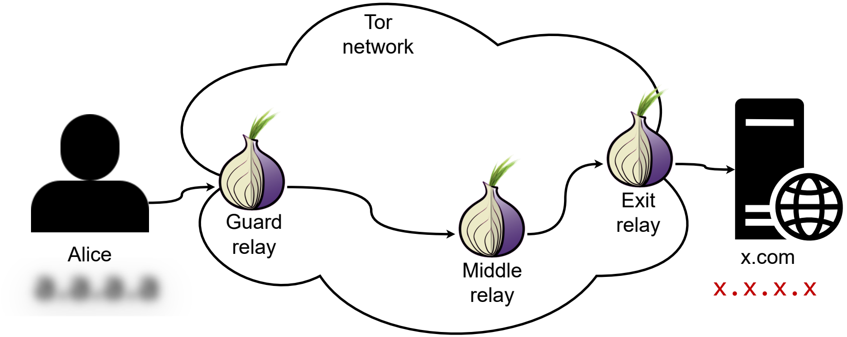
Circuits to onion services:



Client-side anonymity

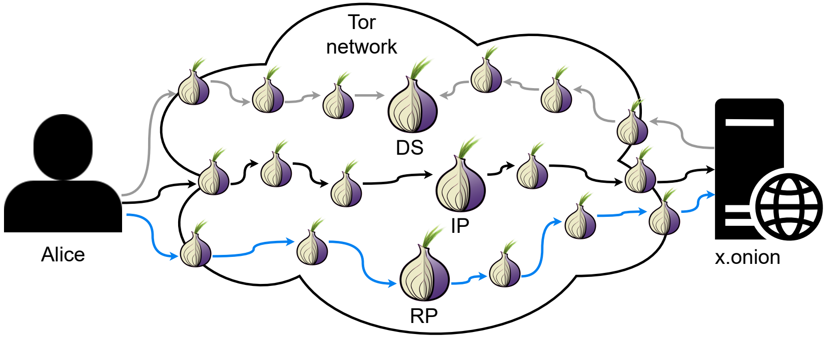
Preserving anonymity with Tor

Circuits to the Internet:



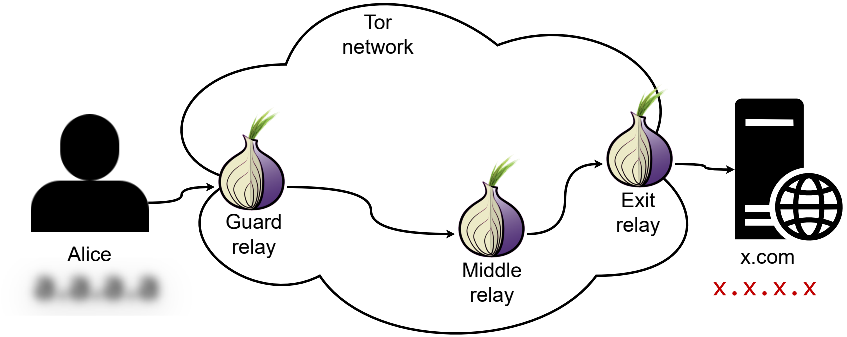
Client-side anonymity

Circuits to onion services:



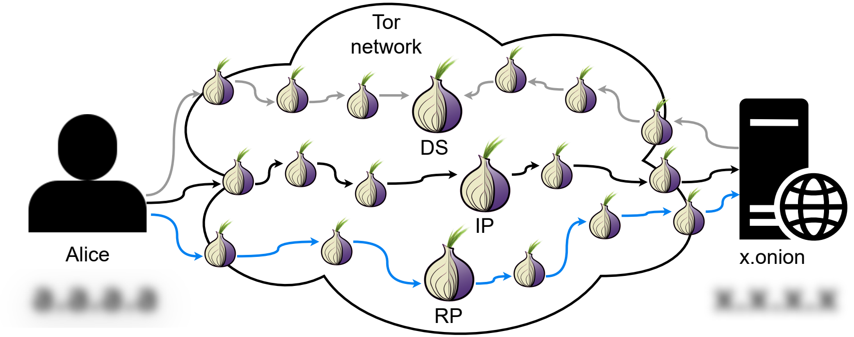
Preserving anonymity with Tor

Circuits to the Internet:



Client-side anonymity

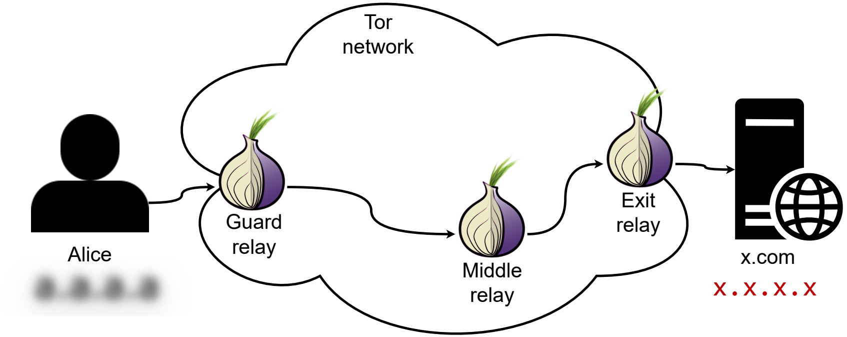
Circuits to onion services:



Client and server-side anonymity

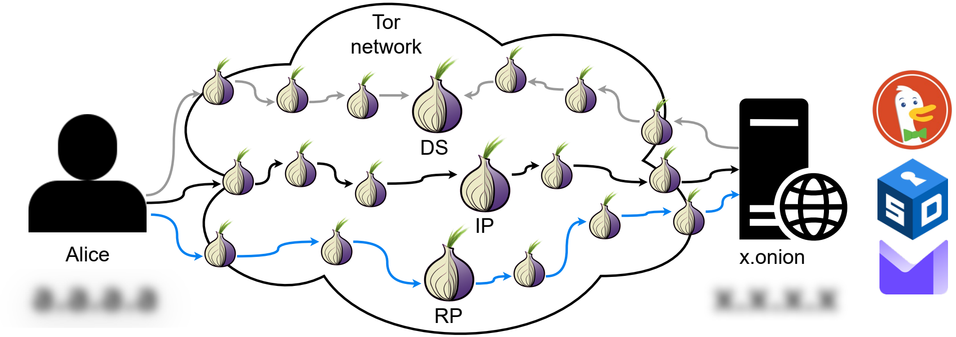
Preserving anonymity with Tor

Circuits to the Internet:



Client-side anonymity

Circuits to onion services:

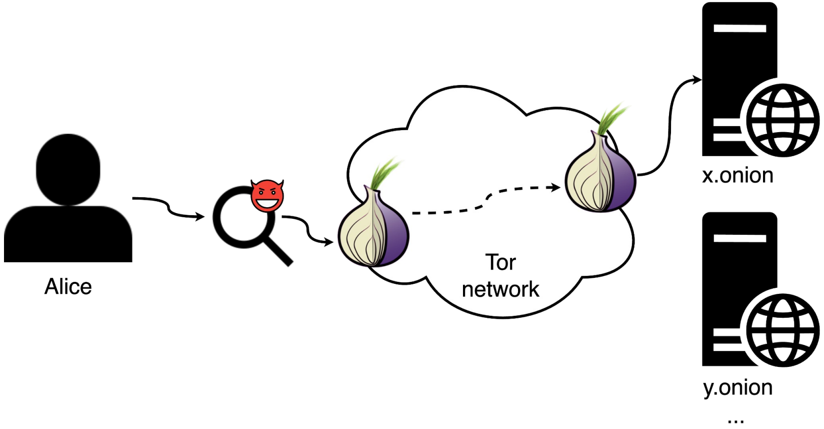


Client and server-side anonymity

Can one **deanonymize** Tor?

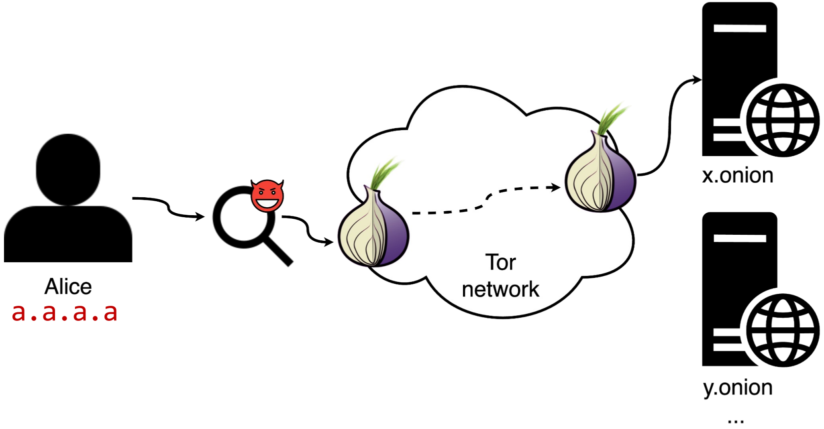
Can one **deanonymize** Tor?

Website fingerprinting:



Can one deanonymize Tor?

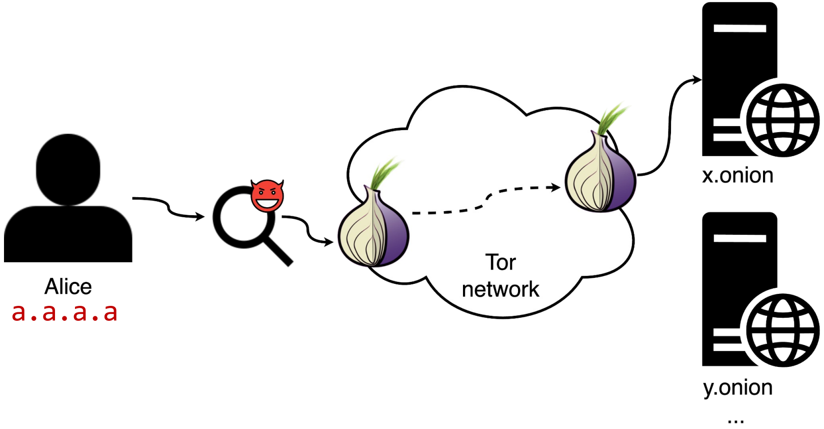
Website fingerprinting:



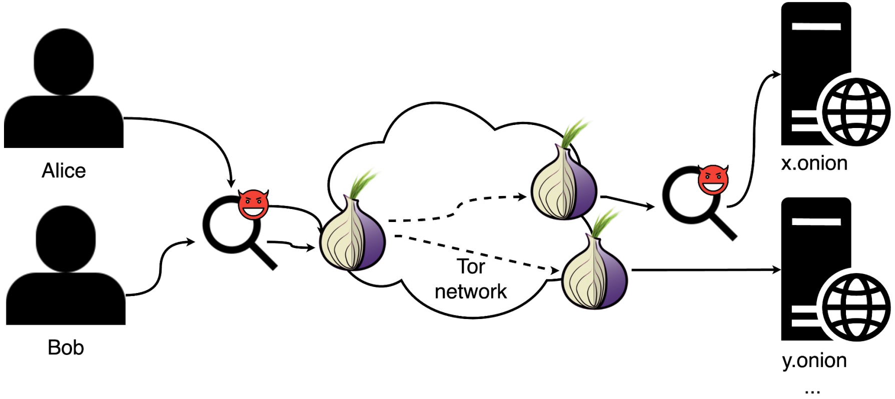
Can't find the service's IP!

Can one deanonymize Tor?

Website fingerprinting:



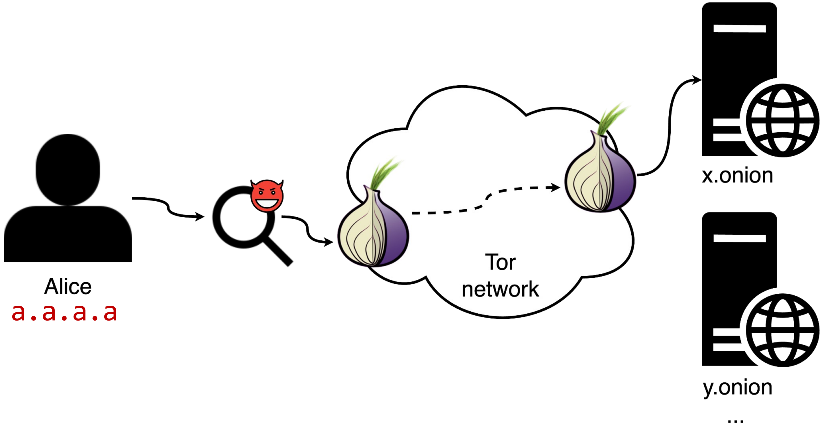
Traffic correlation:



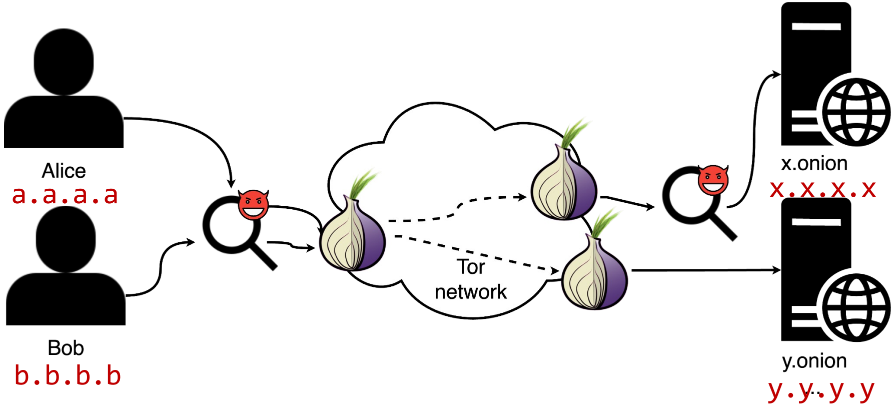
Can’t find the service’s IP!

Can one deanonymize Tor?

Website fingerprinting:



Traffic correlation:

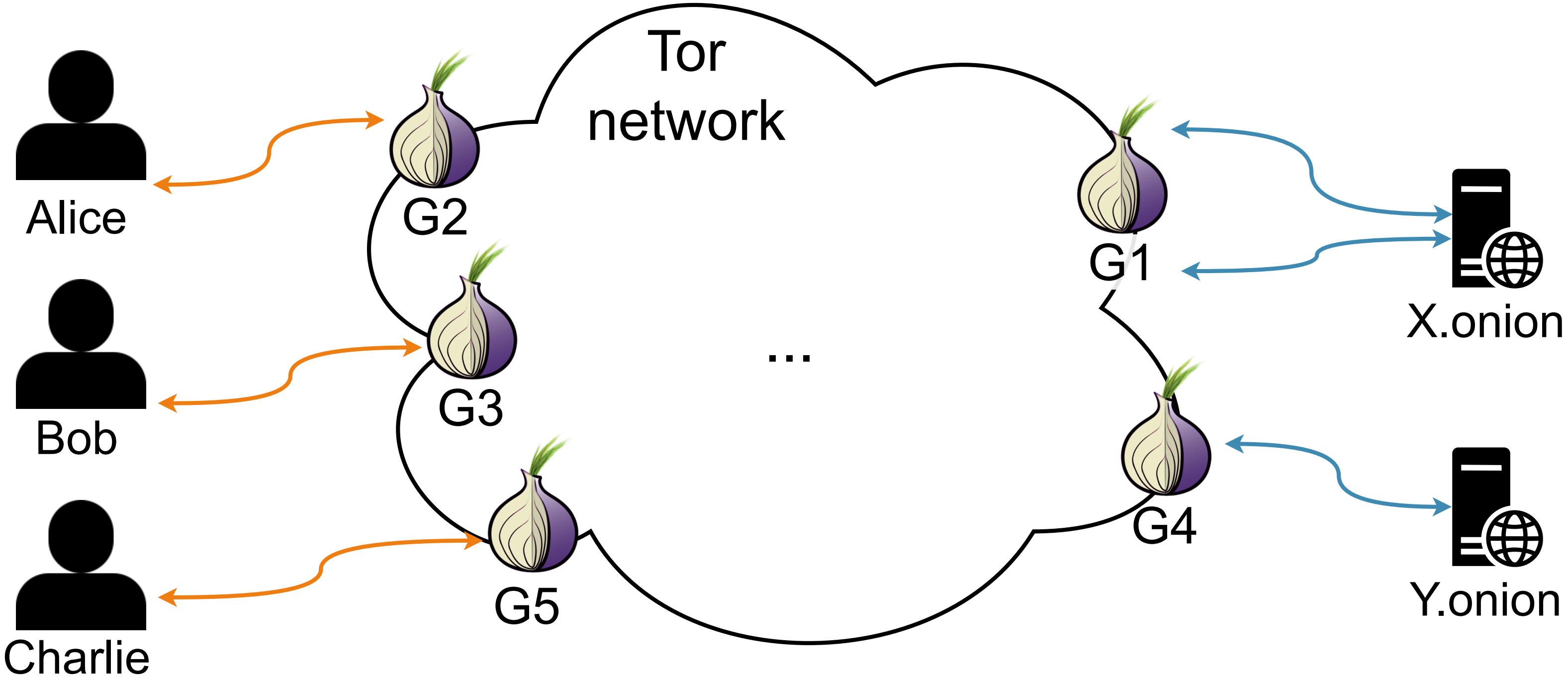


Can’t find the service’s IP!

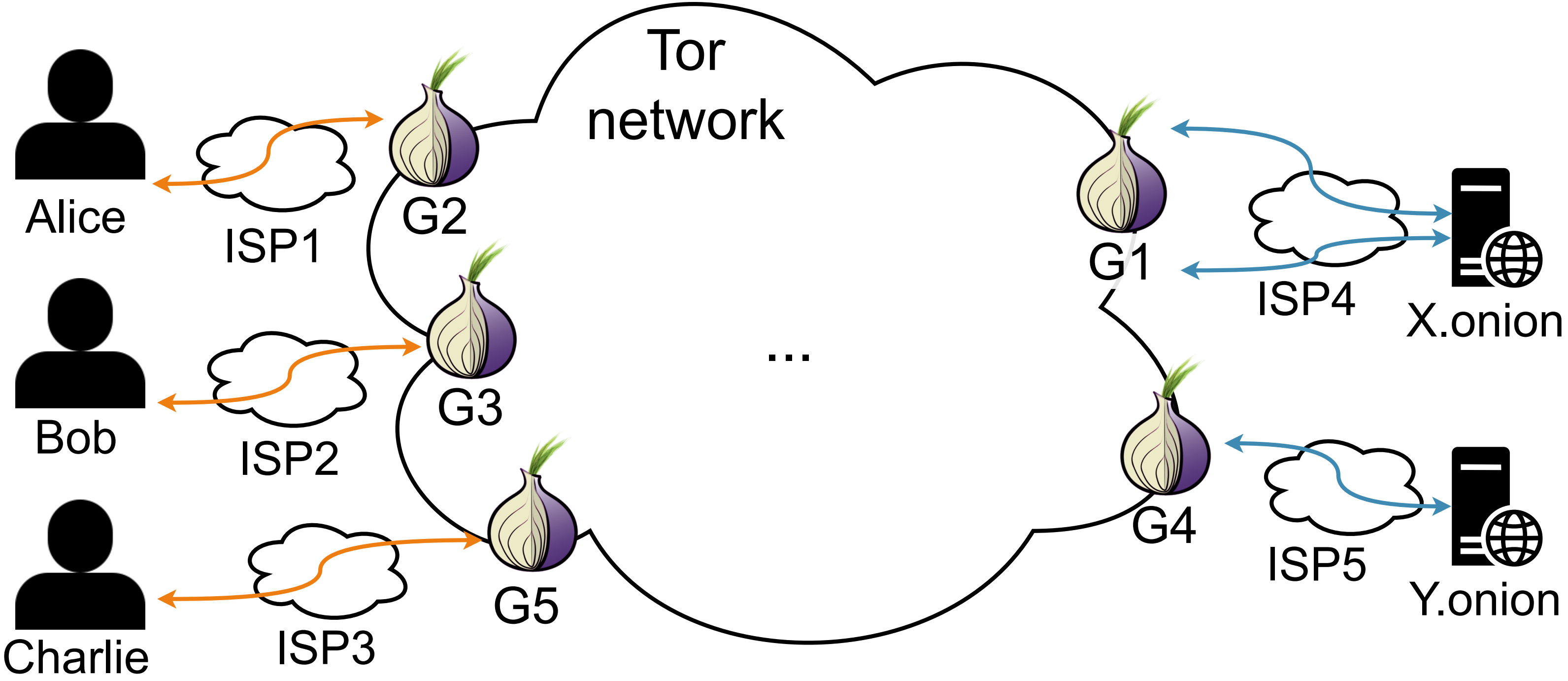
Do existing attacks also work on onion services?

Threat Model

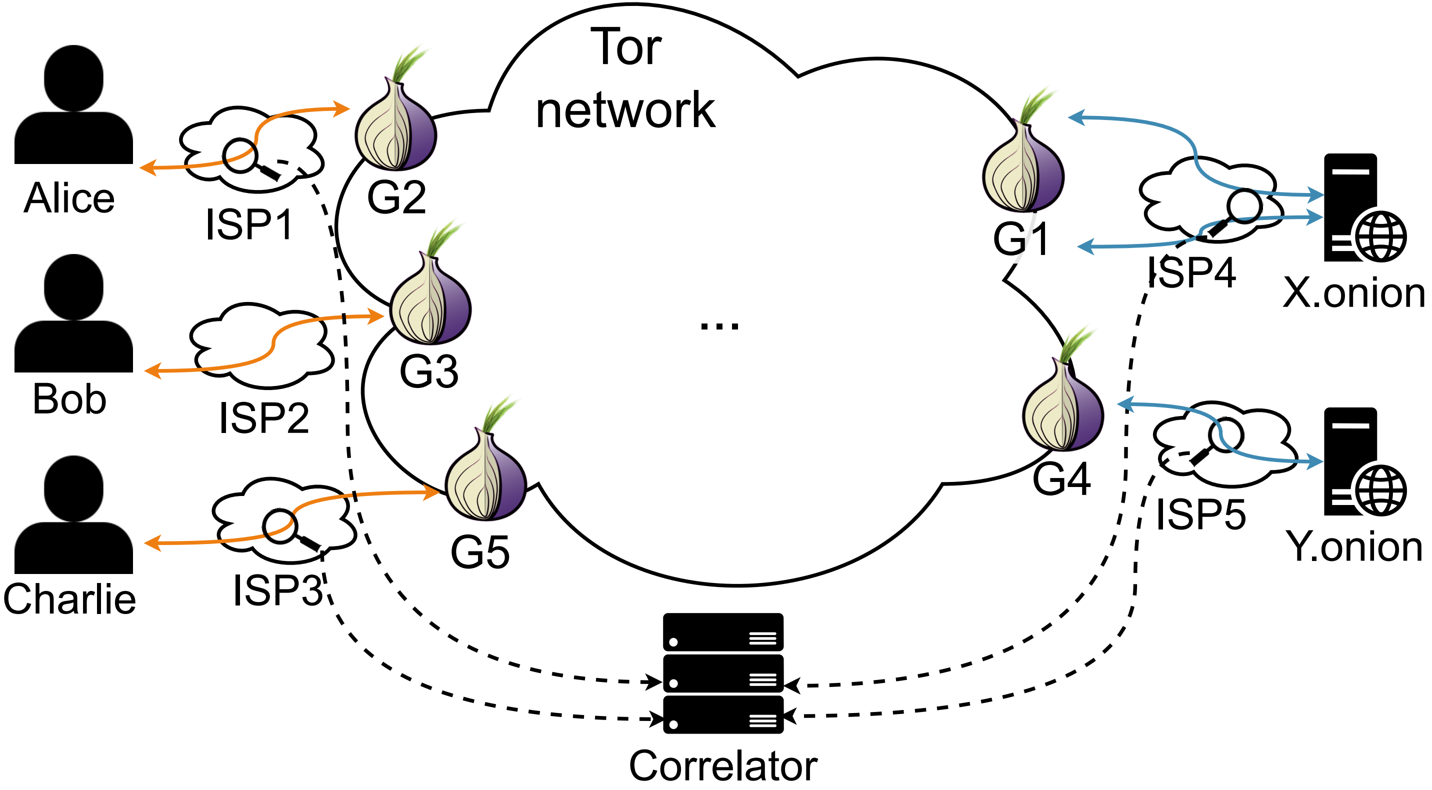
Threat Model



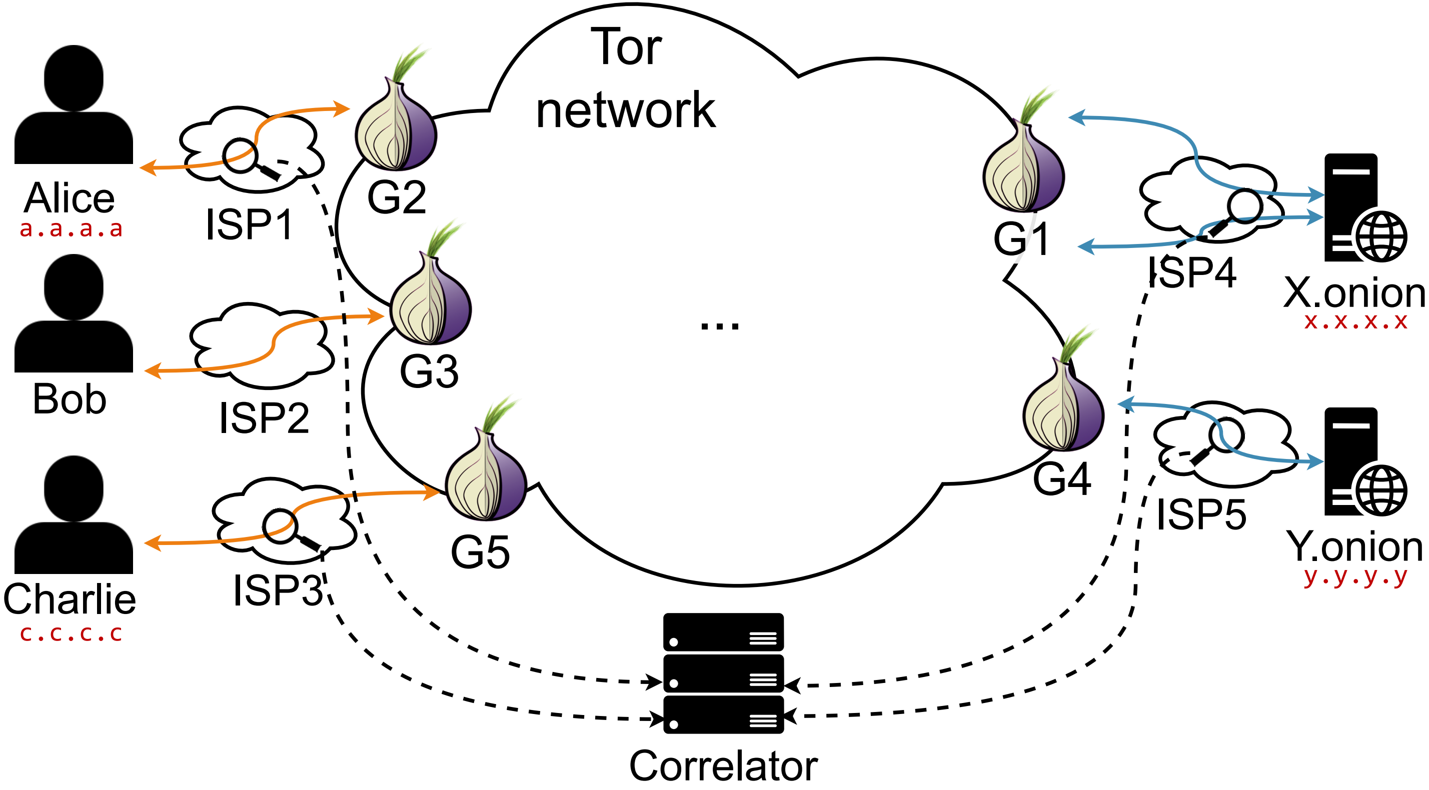
Threat Model



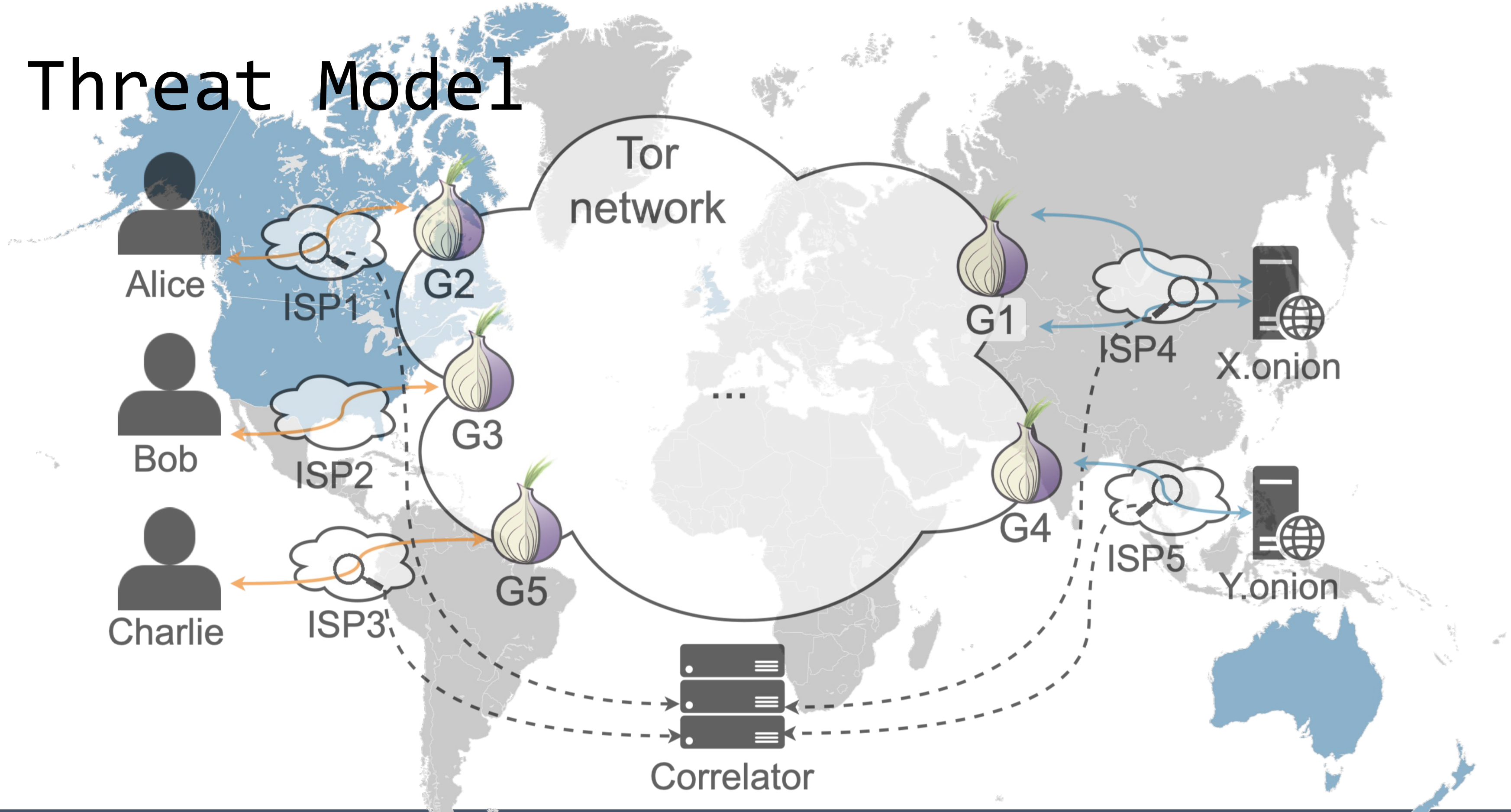
Threat Model



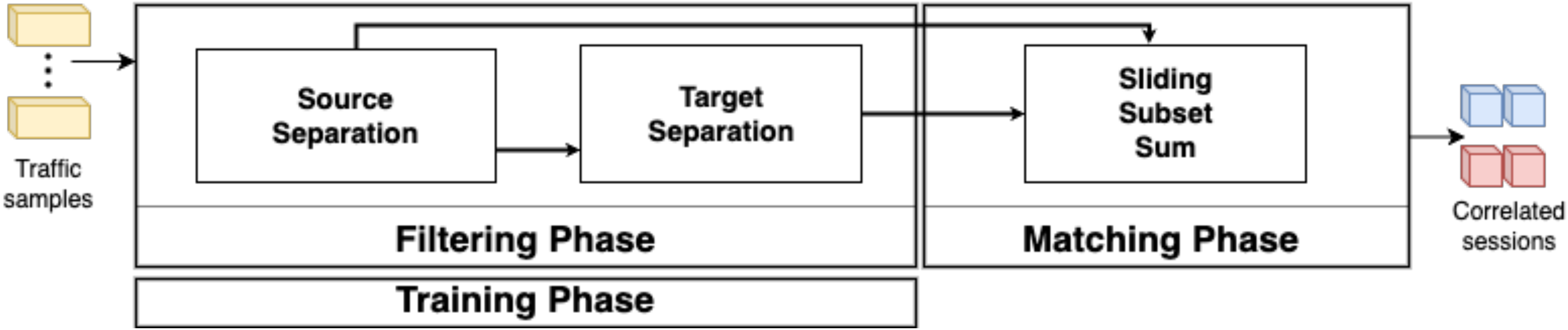
Threat Model



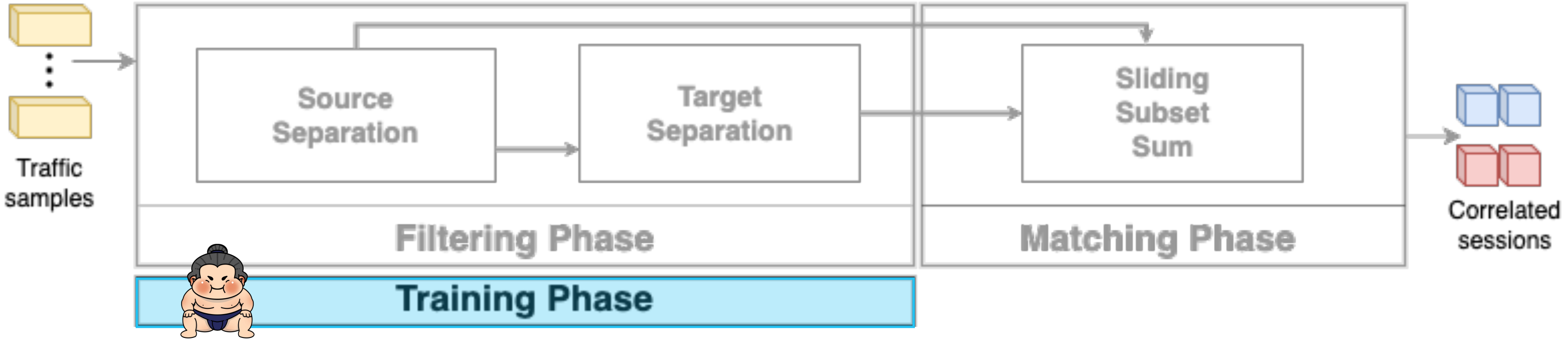
Threat Model



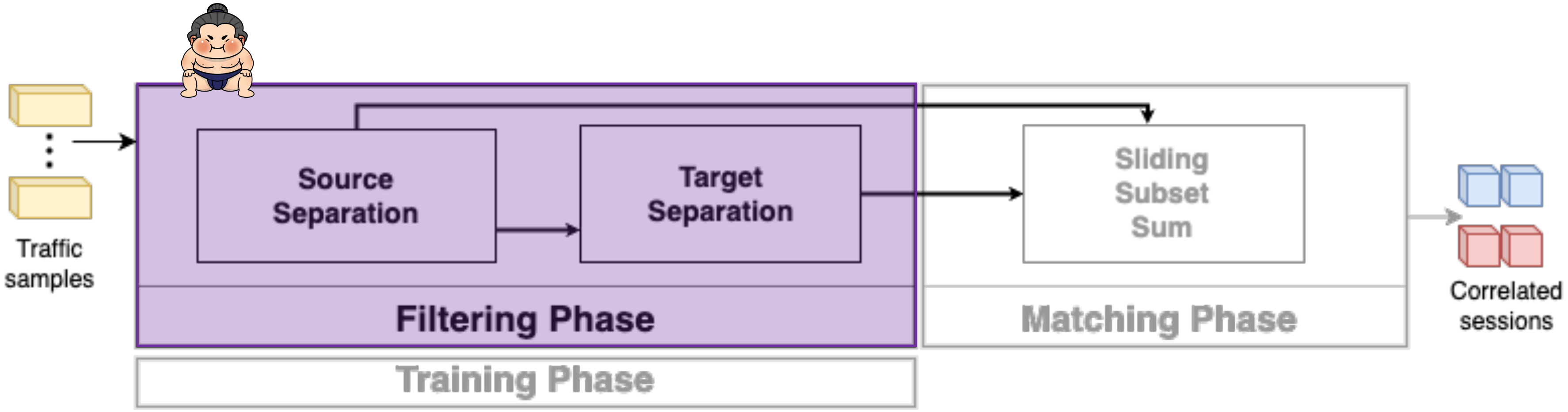
The SUMo Pipeline



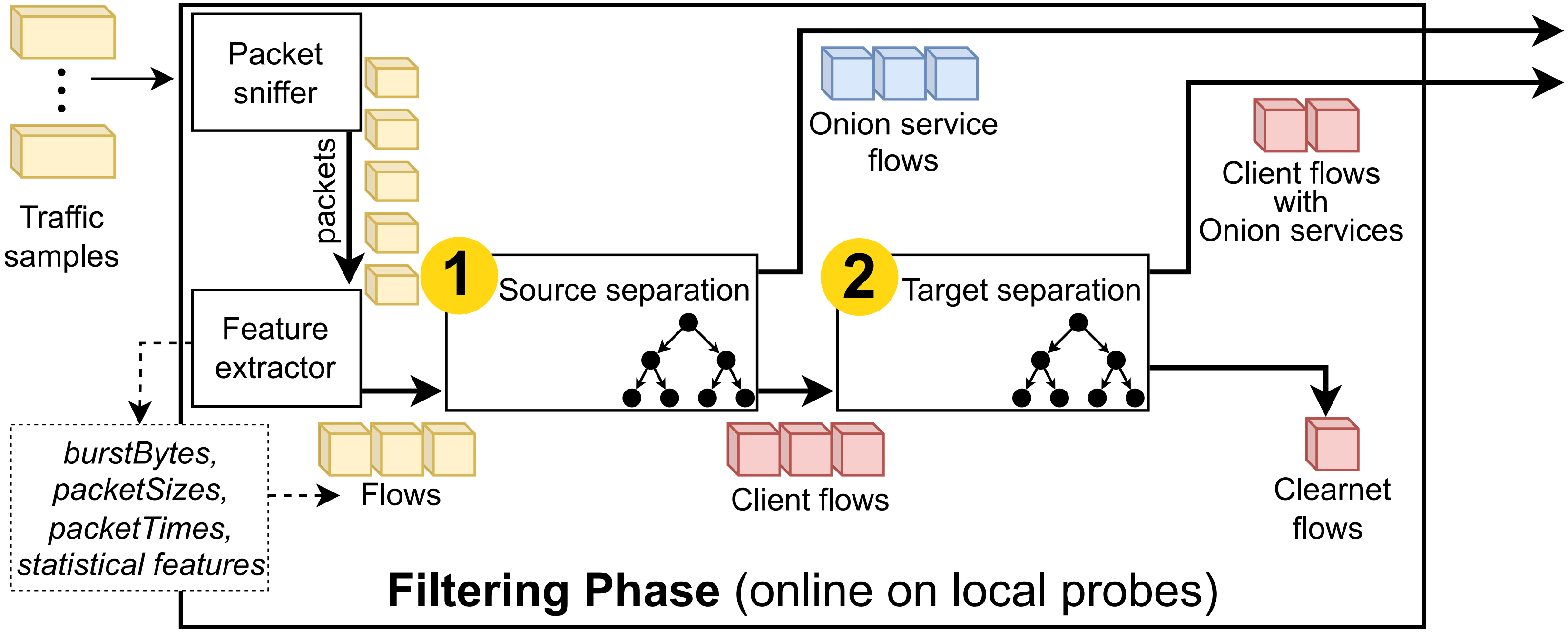
The SUMo Pipeline



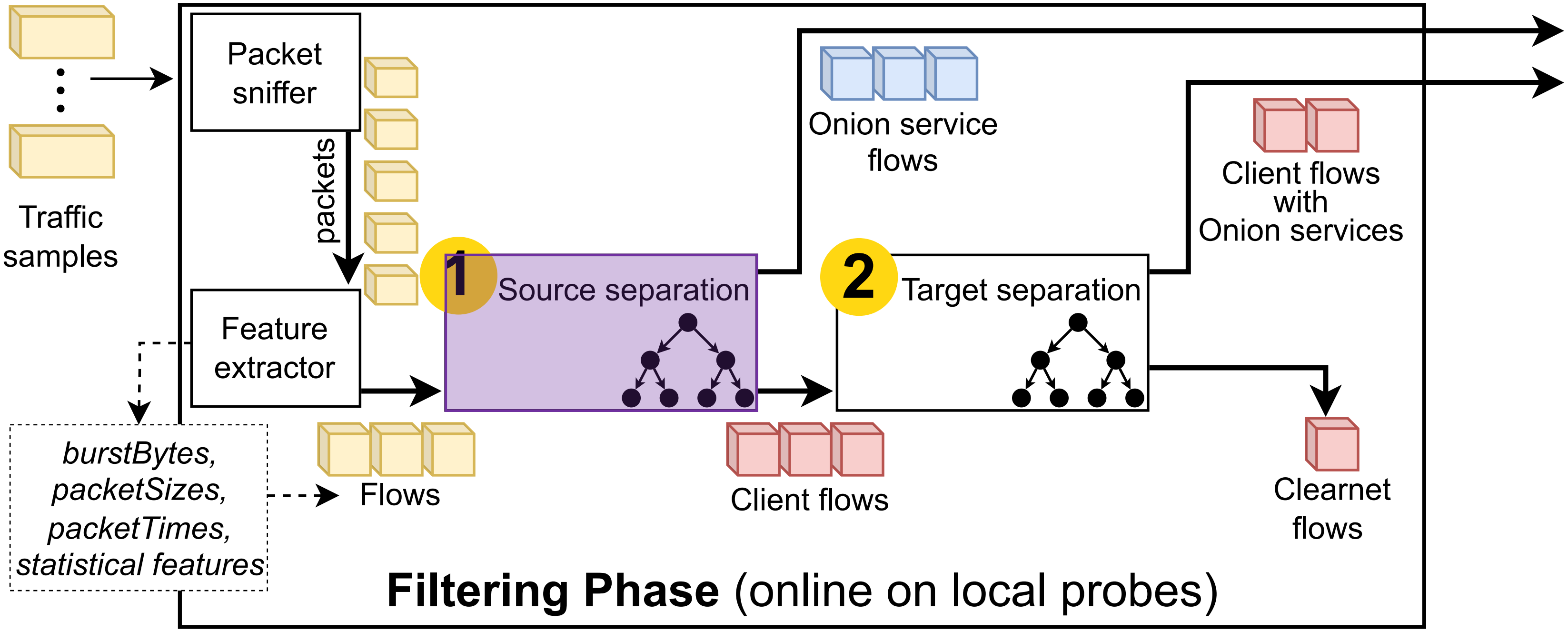
The SUMo Pipeline



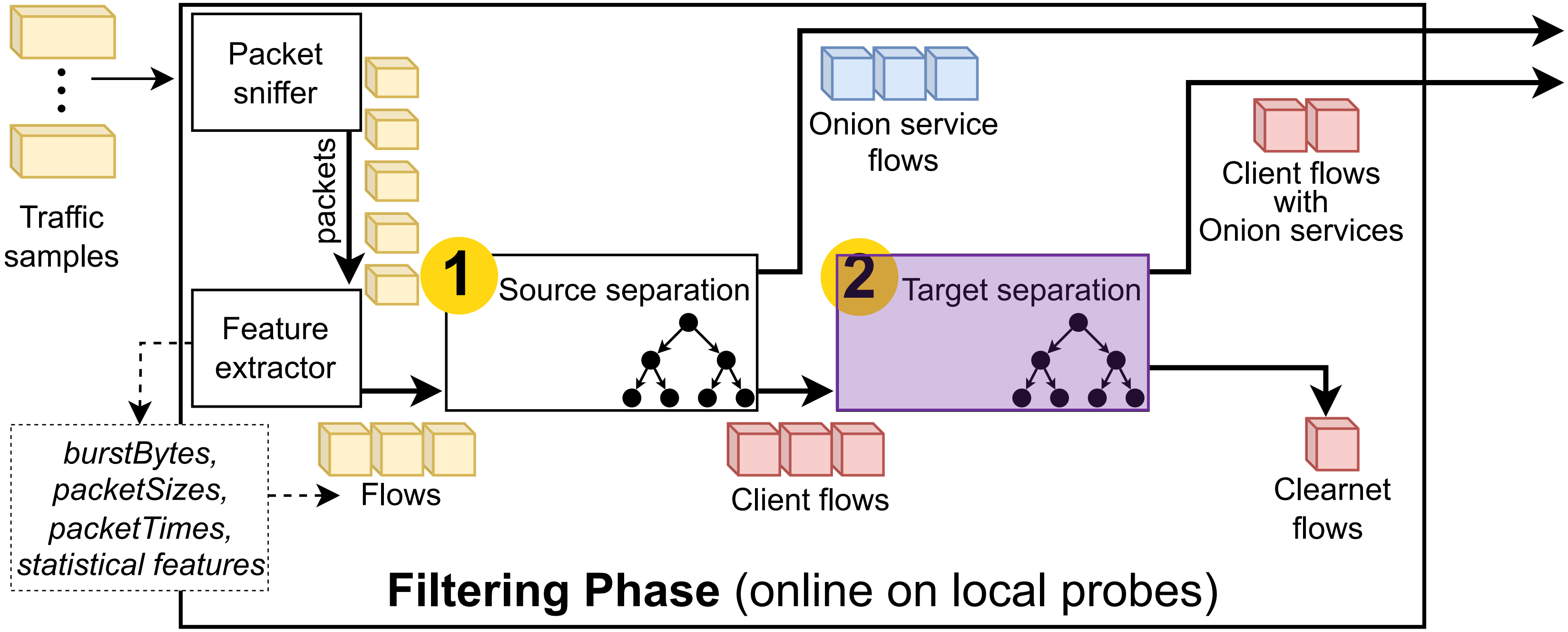
Distinguishing flows



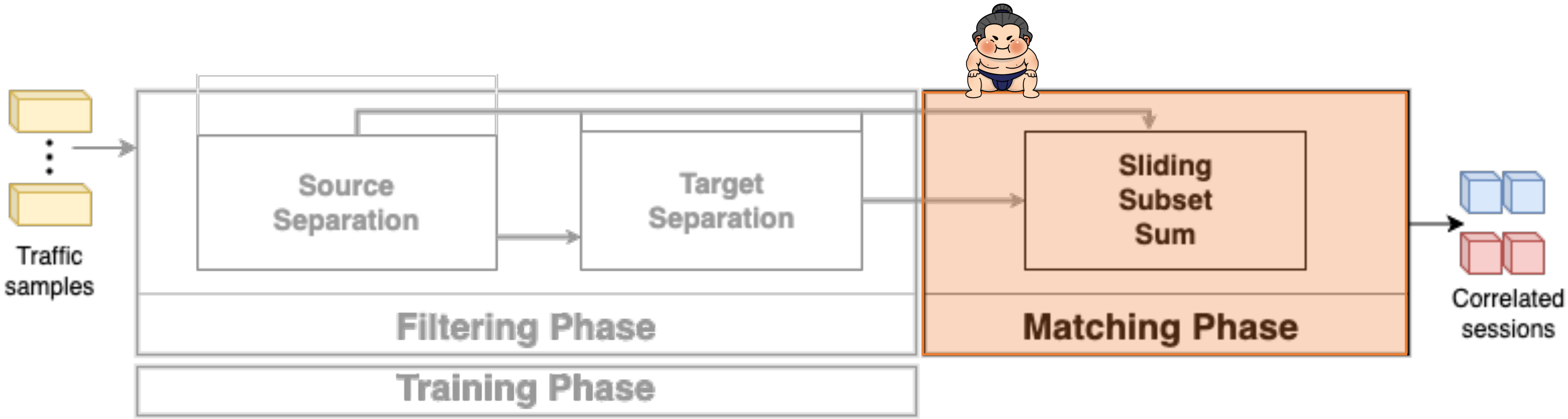
Distinguishing flows by their source



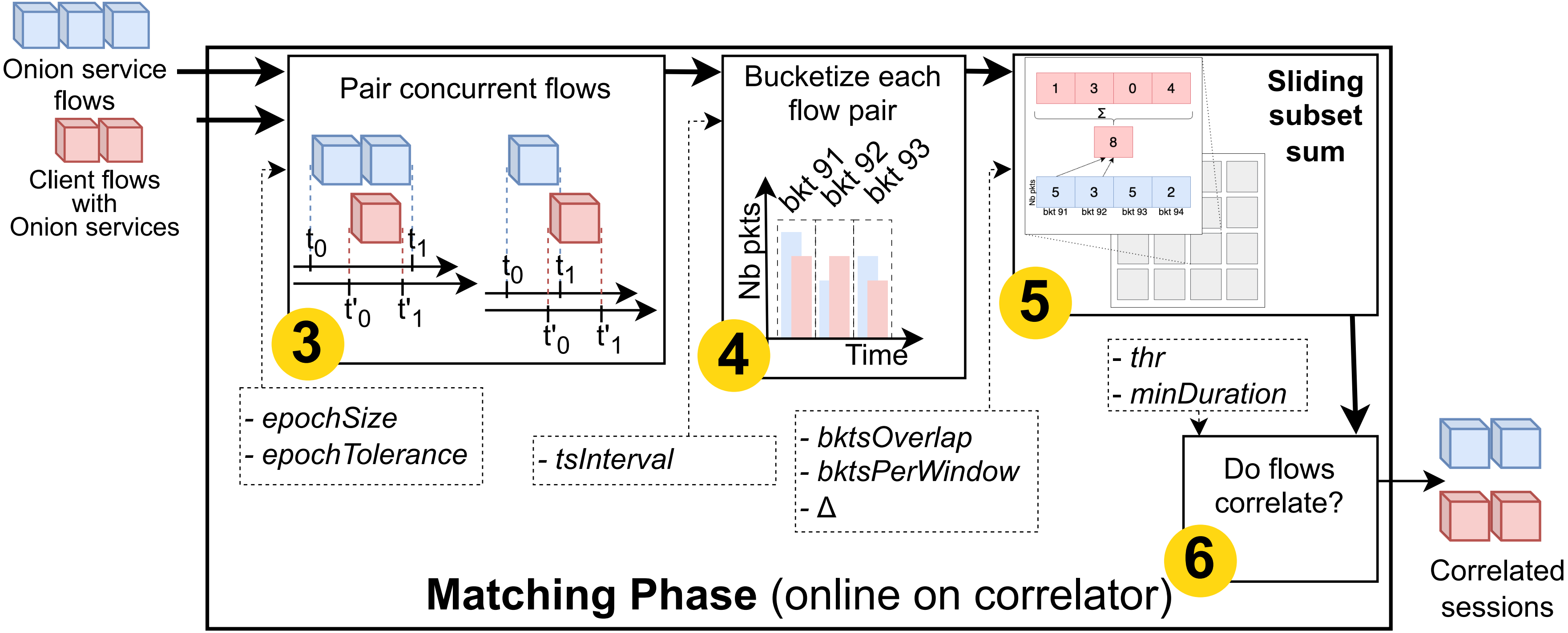
Distinguishing flows by their **destination**



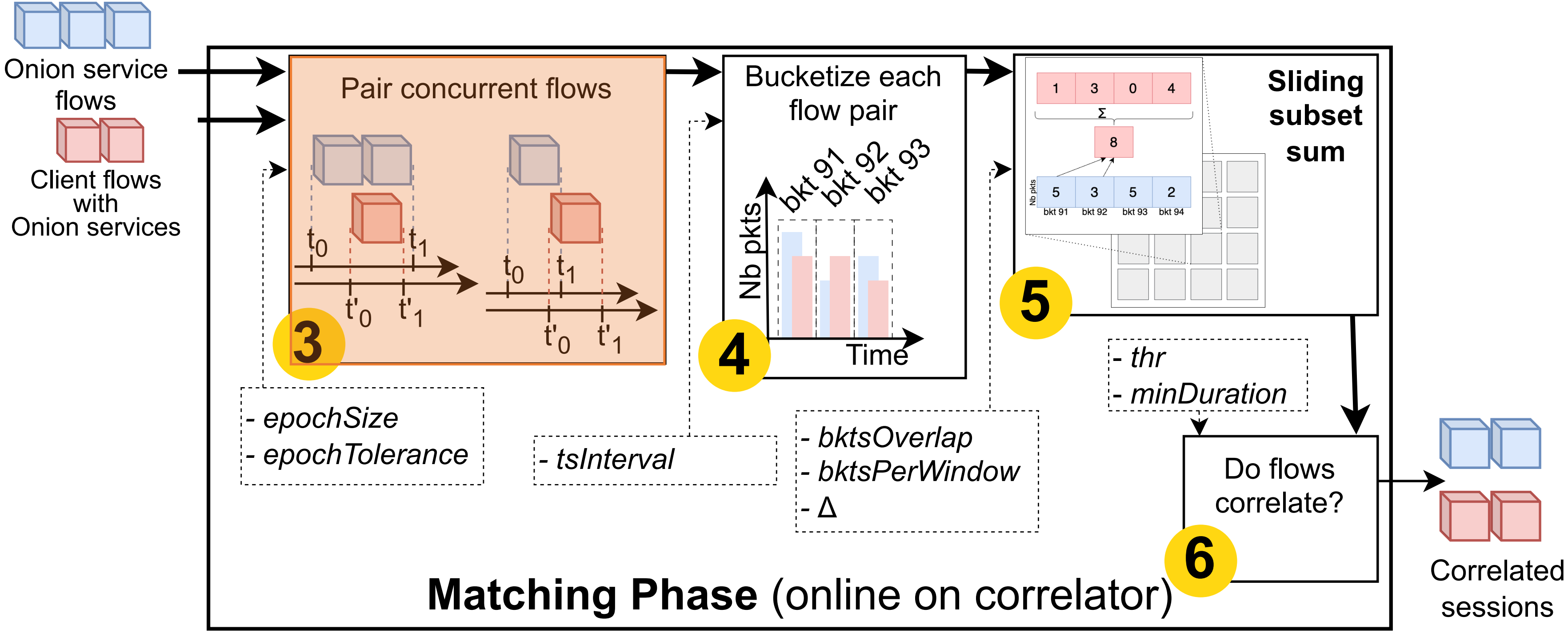
The SUMo Pipeline



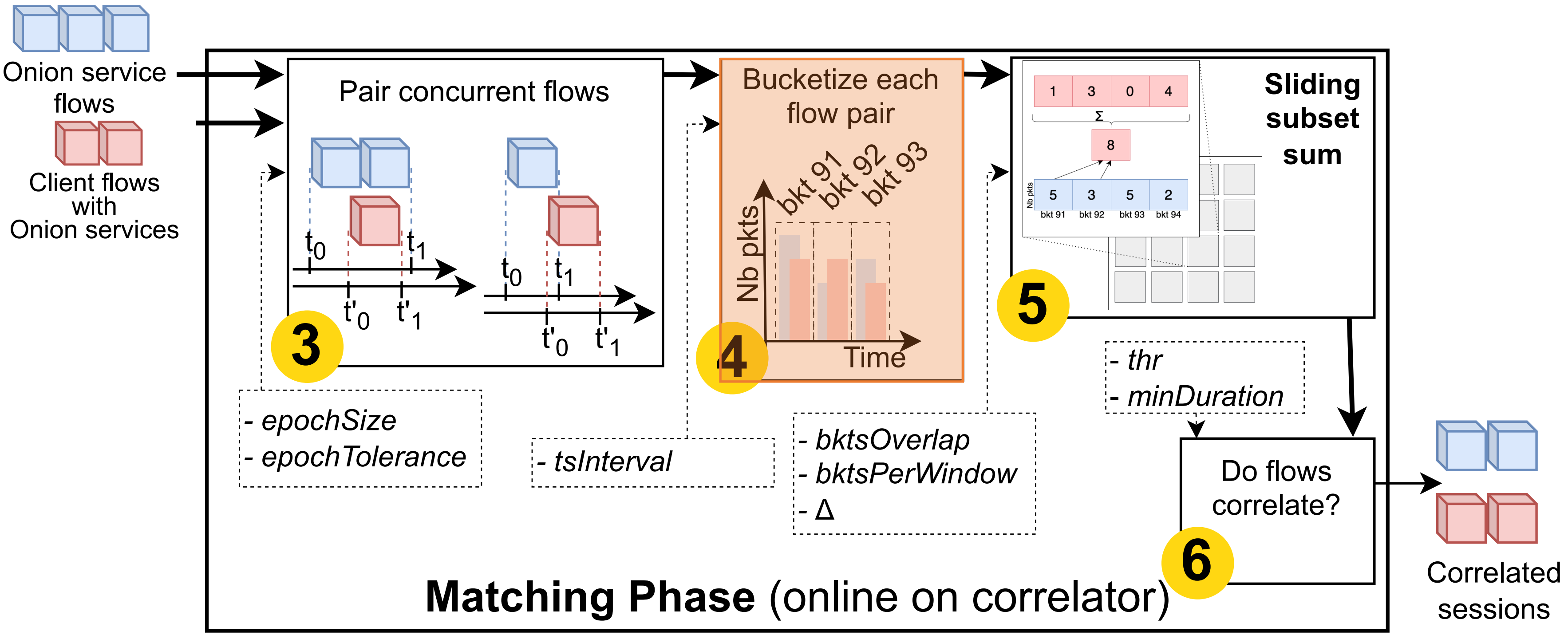
Match clients with onion services



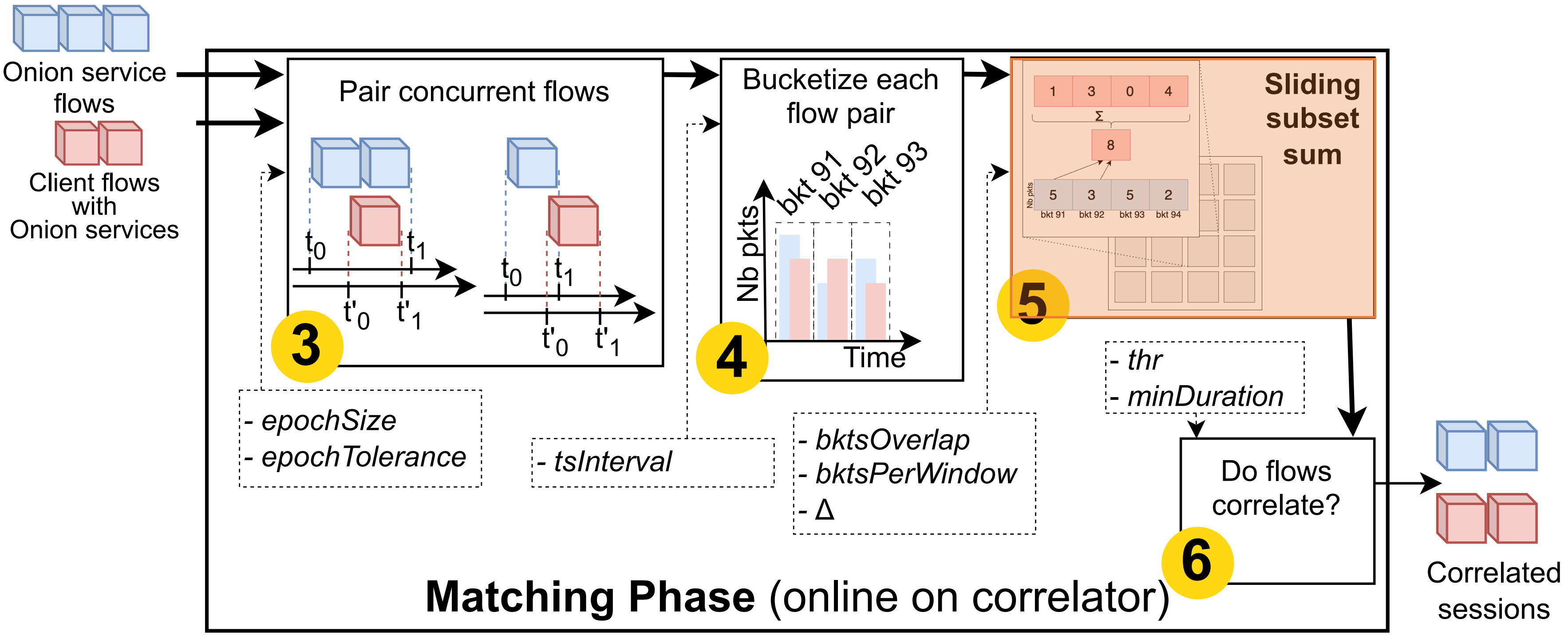
Get possible pair combinations



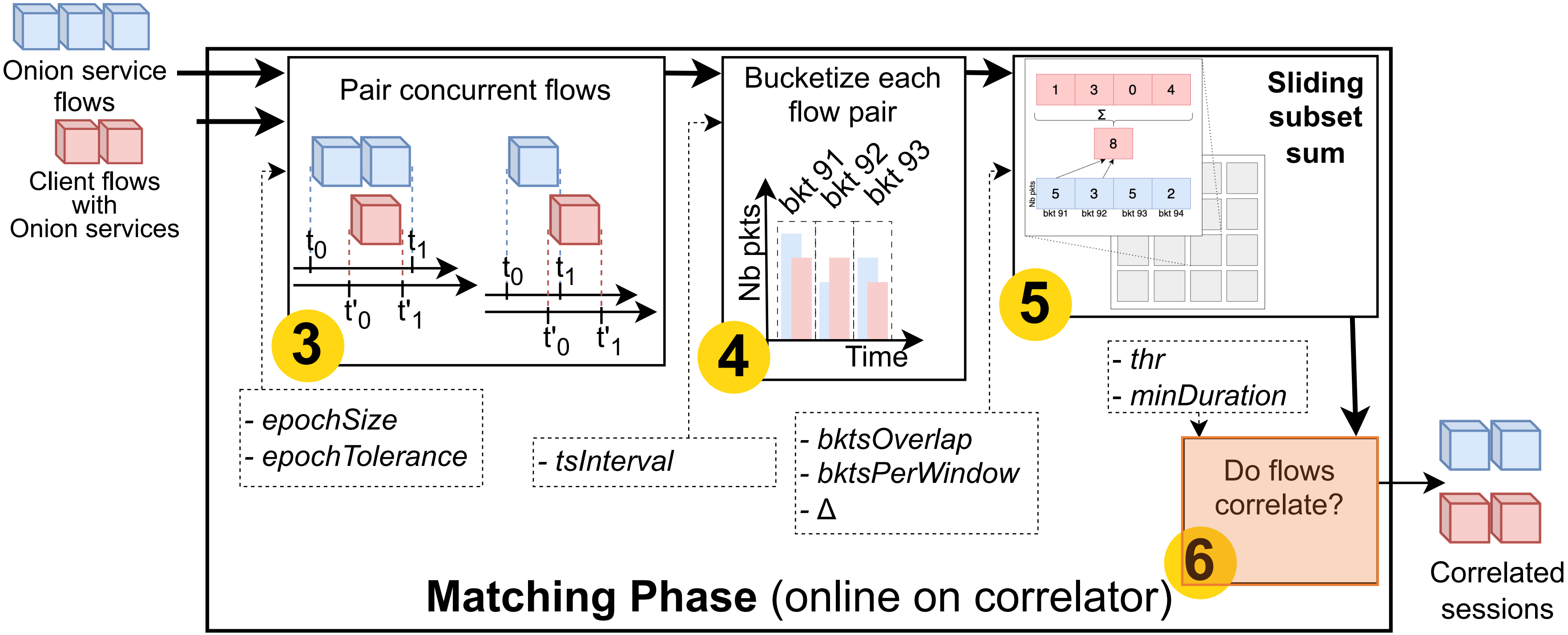
Count packets per time unit



Get similarity scores per window



Group scores to find correlated pairs



Experimental testbed and datasets

Experimental testbed and datasets

- Framework to generate datasets:

Experimental testbed and datasets

- Framework to generate datasets:
  - Geographical distribution.

Experimental testbed and datasets

- Framework to generate datasets:
  - Geographical distribution.
  - Request Concurrency.

Experimental testbed and datasets

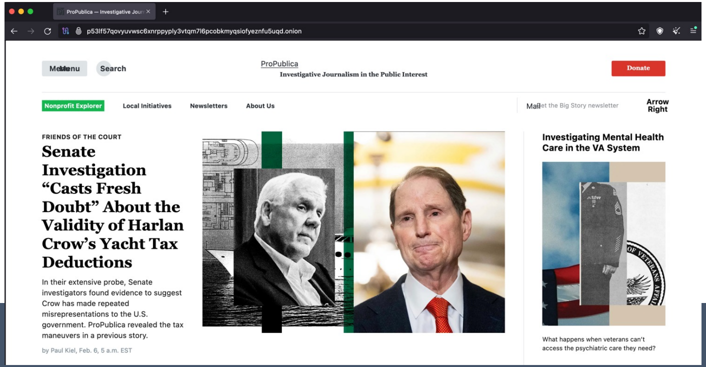
- Framework to generate datasets:
  - Geographical distribution.
  - Request Concurrency.
  - Client-side browsing behaviour.

Experimental testbed and datasets

- Framework to generate datasets:
  - Geographical distribution.
  - Request Concurrency.
  - Client-side browsing behaviour.
  - Host diverse real-world websites.

Experimental testbed and datasets

- Framework to generate datasets:
  - Geographical distribution.
  - Request Concurrency.
  - Client-side browsing behaviour.
  - Host diverse real-world websites.
- Client **sessions** to onion services:


ProPublica — Investigative Journ
p53lf57qovyuvwsc6xnrppyply3vtqm7l6pcobkmyqsiofyeznfu5uqd.onion
MeMenu
Search
Donate
Nonprofit Explorer
Local Initiatives
Newsletters
About Us
MaiGet the Big Story newsletter
Arrow Right

ProPublica
Investigative Journalism in the Public Interest

FRIENDS OF THE COURT

Senate Investigation "Casts Fresh Doubt" About the Validity of Harlan Crow's Yacht Tax Deductions

In their extensive probe, Senate investigators found evidence to suggest Crow has made repeated misrepresentations to the U.S. government. ProPublica revealed the tax maneuvers in a previous story.

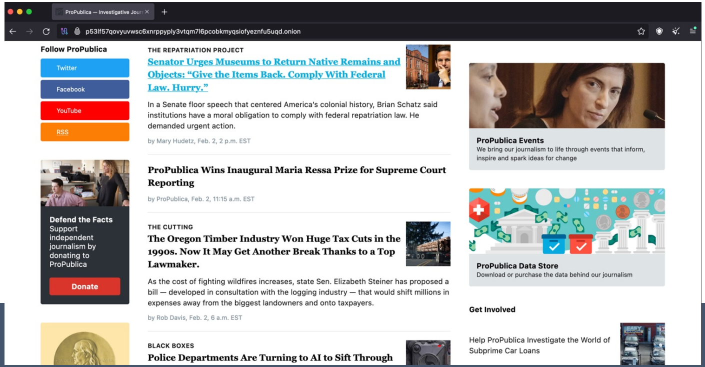
by Paul Kiel, Feb. 6, 5 a.m. EST

Investigating Mental Health Care in the VA System

What happens when veterans can't access the psychiatric care they need?

Experimental testbed and datasets

- Framework to generate datasets:
  - Geographical distribution.
  - Request Concurrency.
  - Client-side browsing behaviour.
  - Host diverse real-world websites.
- Client sessions to onion services:



p53lf57qovyuvwsc6xnrppyply3vtqm7l6pcobkmyqsiofyeznfu5uqd.onion

Follow ProPublica

THE REPATRIATION PROJECT

Senator Urges Museums to Return Native Remains and Objects: “Give the Items Back. Comply With Federal Law. Hurry.”

In a Senate floor speech that centered America’s colonial history, Brian Schatz said institutions have a moral obligation to comply with federal repatriation law. He demanded urgent action.

by Mary Hudetz, Feb. 2, 2 p.m. EST

ProPublica Wins Inaugural Maria Ressa Prize for Supreme Court Reporting

by ProPublica, Feb. 2, 11:15 a.m. EST

THE CUTTING

The Oregon Timber Industry Won Huge Tax Cuts in the 1990s. Now It May Get Another Break Thanks to a Top Lawmaker.

As the cost of fighting wildfires increases, state Sen. Elizabeth Steiner has proposed a bill — developed in consultation with the logging industry — that would shift millions in expenses away from the biggest landowners and onto taxpayers.

by Rob Davis, Feb. 2, 6 a.m. EST

BLACK BOXES

Police Departments Are Turning to AI to Sift Through

Defend the Facts
Support independent journalism by donating to ProPublica

ProPublica Events
We bring our journalism to life through events that inform, inspire and spark ideas for change

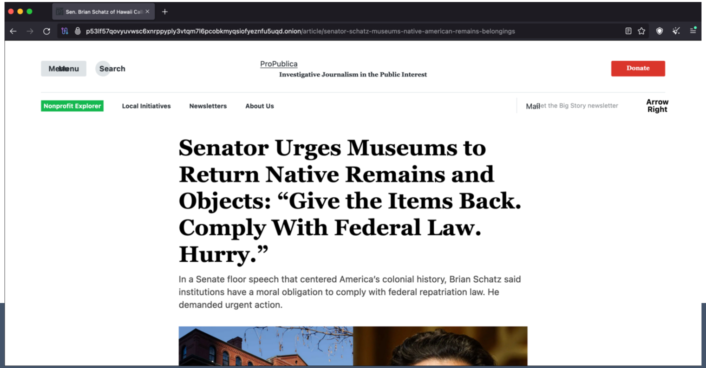
ProPublica Data Store
Download or purchase the data behind our journalism

Get Involved

Help ProPublica Investigate the World of Subprime Car Loans

Experimental testbed and datasets

- Framework to generate datasets:
  - Geographical distribution.
  - Request Concurrency.
  - Client-side browsing behaviour.
  - Host diverse real-world websites.
- Client **sessions** to onion services:

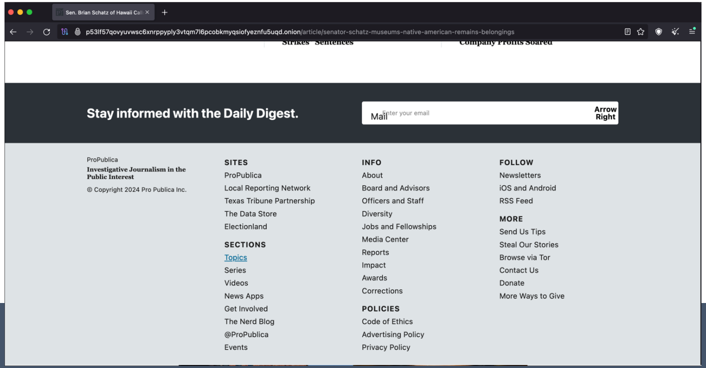

Sen. Brian Schatz of Hawaii Cal
p53lf57qovyuvwsc6xnrppyply3vtqm7l6pcobkmyqsiofyeznfu5uqd.onion/article/senator-schatz-museums-native-american-remains-belongings
MeMenu
Search
ProPublica
Investigative Journalism in the Public Interest
Donate
Nonprofit Explorer
Local Initiatives
Newsletters
About Us
MailGet the Big Story newsletter
Arrow Right

Senator Urges Museums to Return Native Remains and Objects: “Give the Items Back. Comply With Federal Law. Hurry.”

In a Senate floor speech that centered America’s colonial history, Brian Schatz said institutions have a moral obligation to comply with federal repatriation law. He demanded urgent action.

Experimental testbed and datasets

- Framework to generate datasets:
  - Geographical distribution.
  - Request Concurrency.
  - Client-side browsing behaviour.
  - Host diverse real-world websites.
- Client **sessions** to onion services:



Sen. Brian Schatz of Hawaii Cal

p53lf57qovyuvwsc6xnrppyply3vtqm7l6pcobkmyqsiofyeznfu5uqd.onion/article/senator-schatz-museums-native-american-remains-belongings

Stay informed with the Daily Digest.

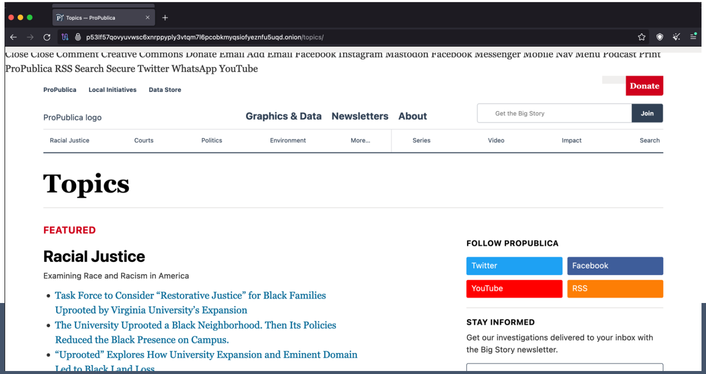
ProPublica

Investigative Journalism in the Public Interest

© Copyright 2024 Pro Publica Inc.

Experimental testbed and datasets

- Framework to generate datasets:
  - Geographical distribution.
  - Request Concurrency.
  - Client-side browsing behaviour.
  - Host diverse real-world websites.
- Client sessions to onion services:


Topics — ProPublica
p53lf57qovyuvwsc6xnrppyply3vtqm7l6pcobkmyqsiofyeznfu5uqd.onion/topics/

Close Close Comment Creative Commons Donate Email Add Email Facebook Instagram Mastodon Facebook Messenger Mobile Nav Menu Podcast Print ProPublica RSS Search Secure Twitter WhatsApp YouTube

Topics

FEATURED

Racial Justice

Examining Race and Racism in America

- Task Force to Consider "Restorative Justice" for Black Families Uprooted by Virginia University's Expansion
- The University Uprooted a Black Neighborhood. Then Its Policies Reduced the Black Presence on Campus.
- "Uprooted" Explores How University Expansion and Eminent Domain Led to Black Land Loss

FOLLOW PROPUBLICA

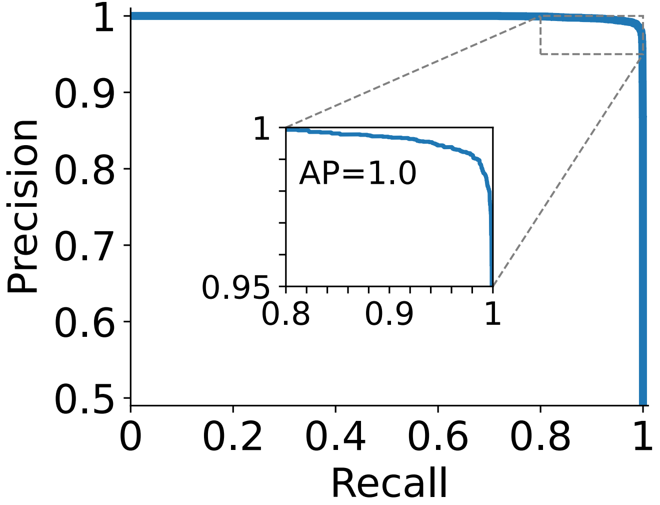
STAY INFORMED

Get our investigations delivered to your inbox with the Big Story newsletter.

Filtering Phase

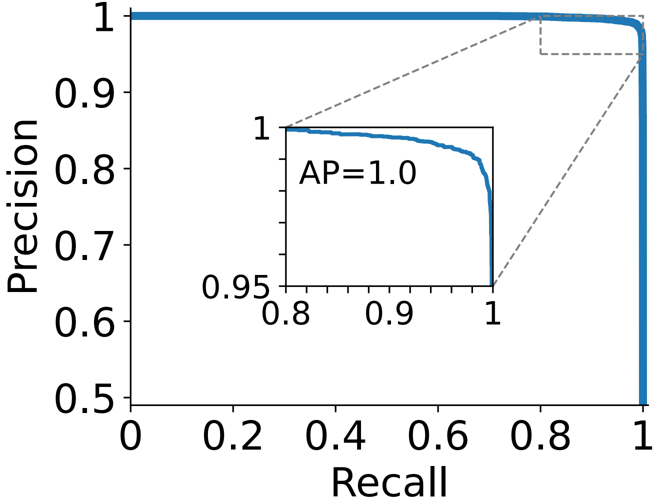
Filtering Phase

Source separation



Filtering Phase

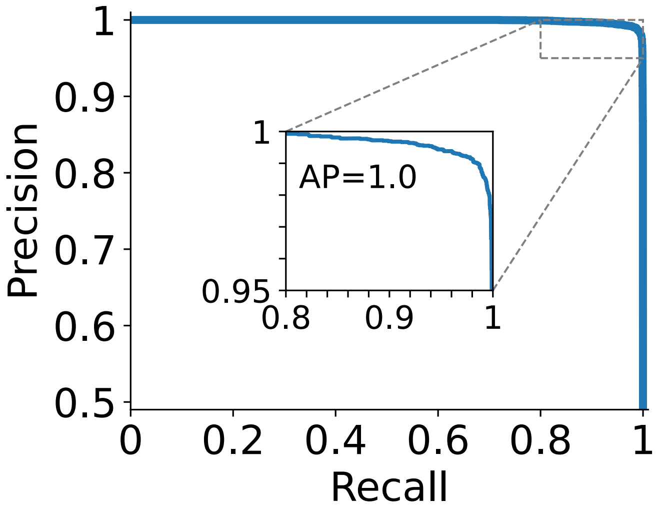
Source separation



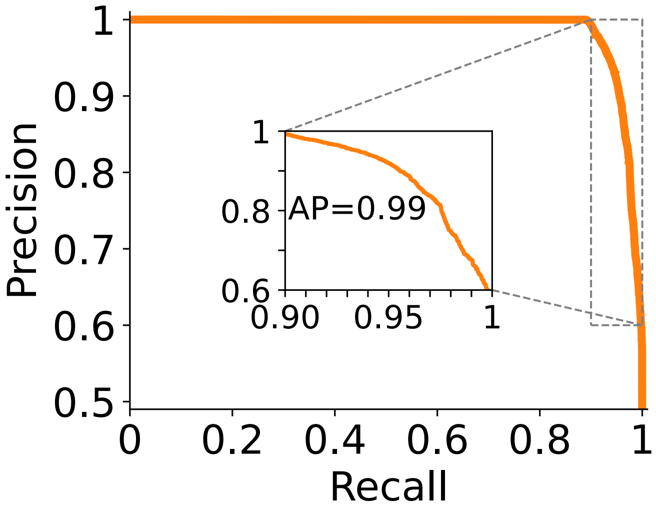
Distinguishes client- from server-side flows!

Filtering Phase

Source separation



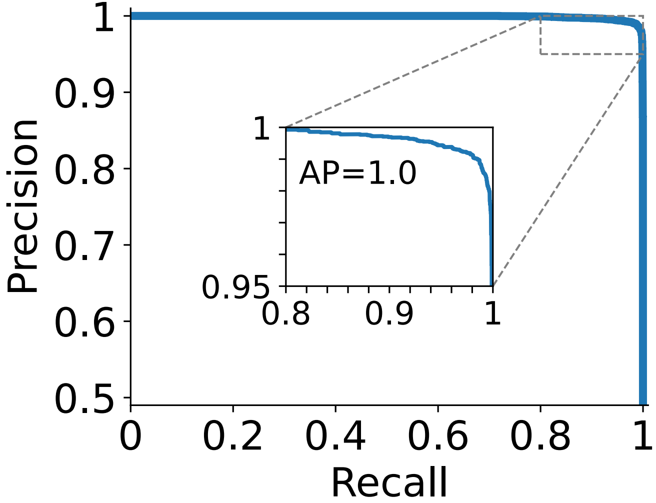
Target separation



Distinguishes client- from server-side flows!

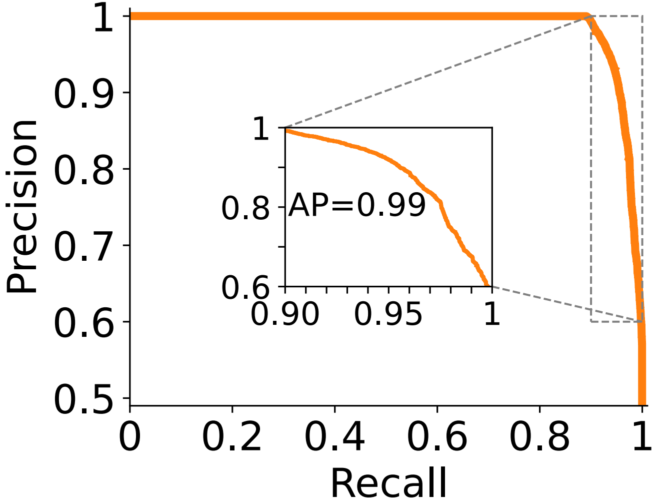
Filtering Phase

Source separation



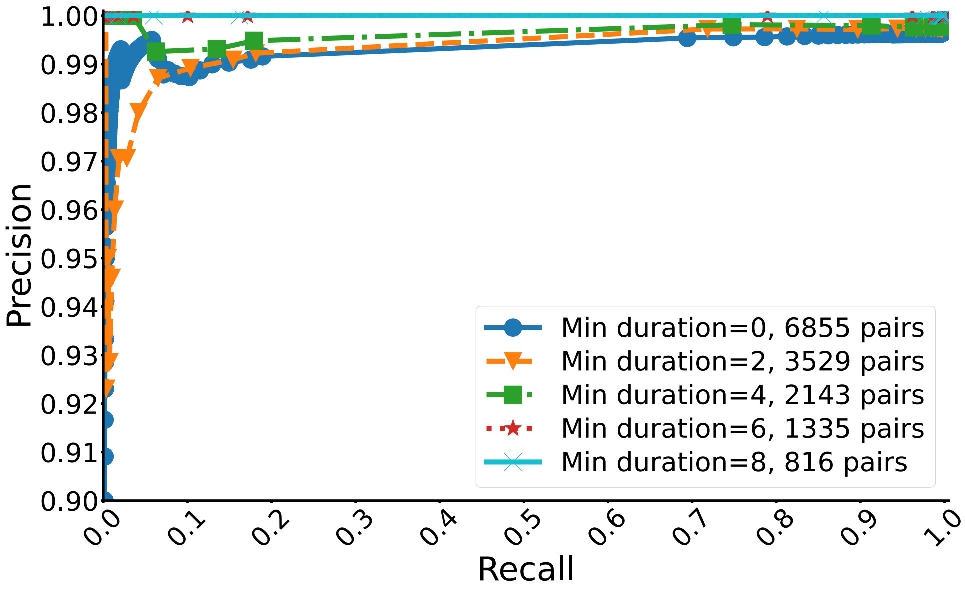
Distinguishes client- from server-side flows!

Target separation



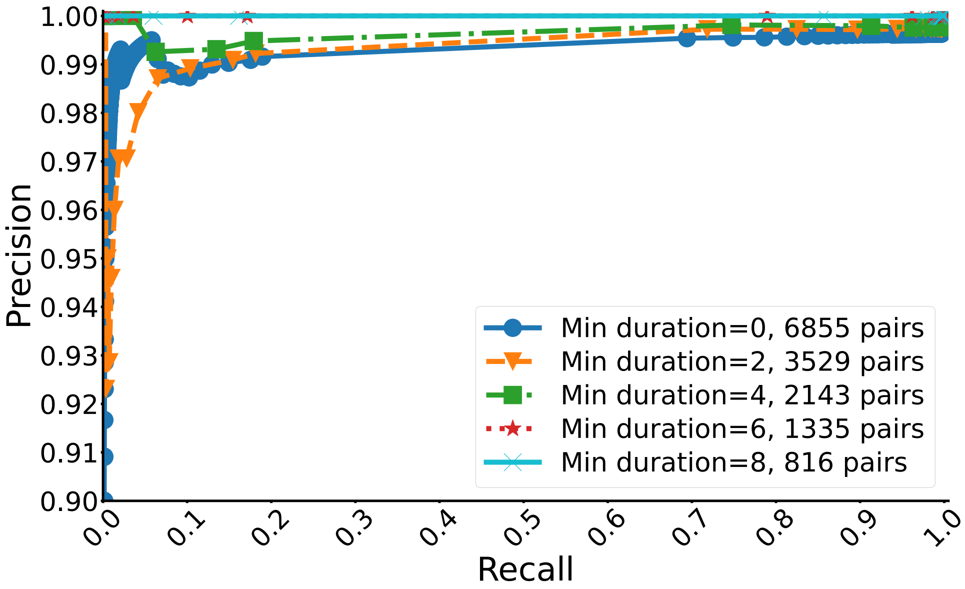
Distinguishes between flows to the Internet and to onion services!

Session Matching



Session Matching

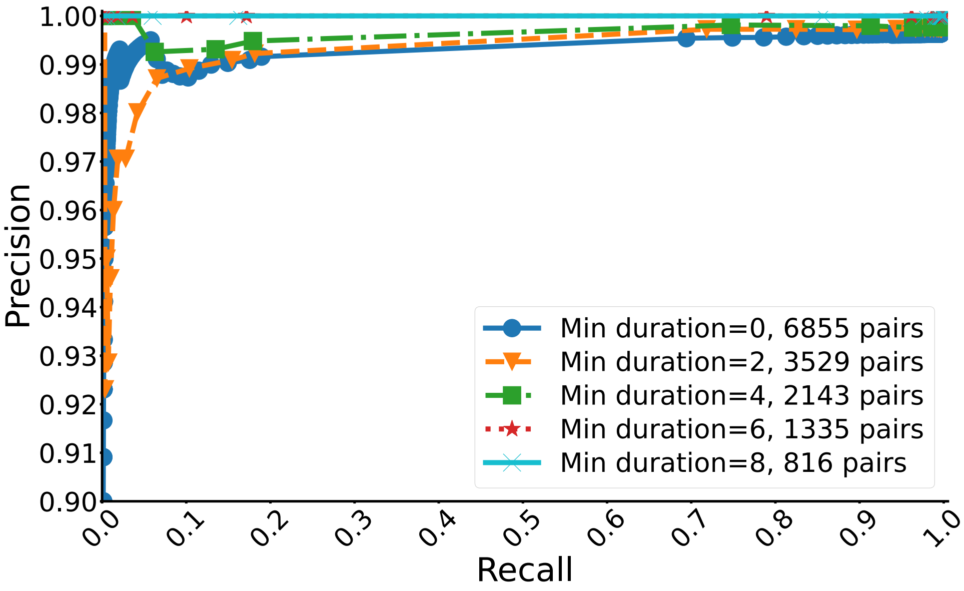
- Over **99.6% precision and recall** for any duration.



Session Matching

- Over **99.6% precision and recall** for any duration.

★ **100% precision** for sessions longer than 6 minutes.

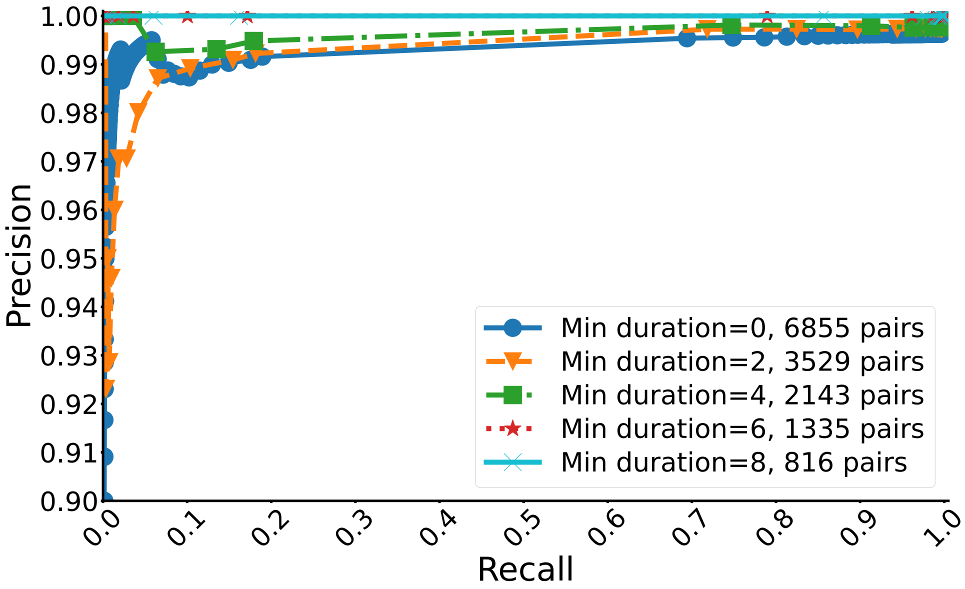


Session Matching

- Over **99.6% precision and recall** for any duration.

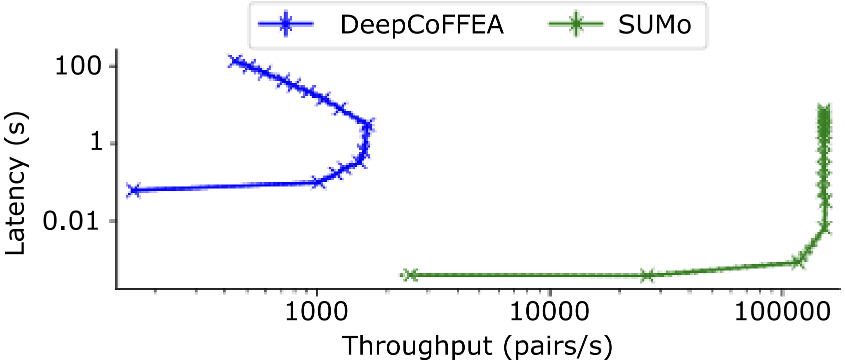
★ **100% precision** for sessions longer than 6 minutes.

- **Imperfect filtering achieves 99.5% precision!**



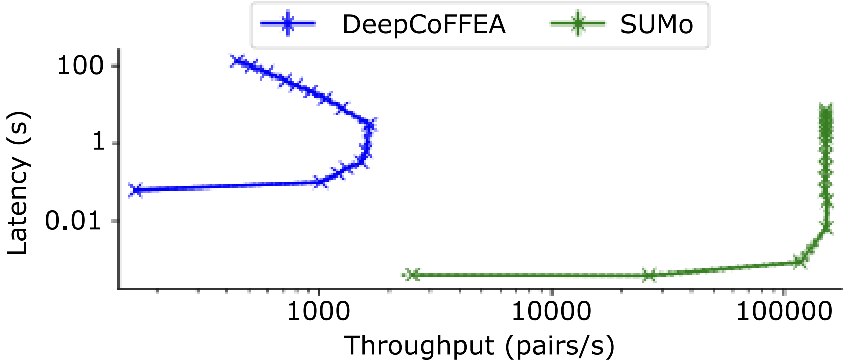
SUMo is fast!

Phase	Stage	Training time	Inference Time
Filtering	Source Separation Target Separation	< 6 seconds total	< 4 µs/flow
Matching	Session Correlation	-	< 6 µs/pair



SUMo is fast!

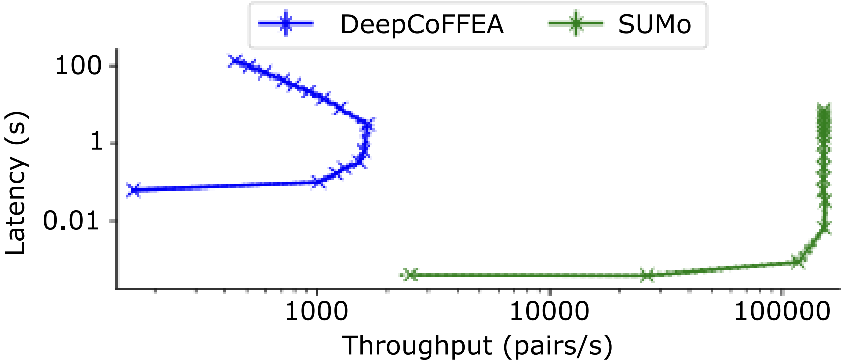
Phase	Stage	Training time	Inference Time
Filtering	Source Separation	< 6 seconds total	< 4 µs/flow
	Target Separation		
Matching	Session Correlation	-	< 6 µs/pair



SUMo is fast!

Phase	Stage	Training time	Inference Time
Filtering	Source Separation Target Separation	< 6 seconds total	< 4 µs/flow
Matching	Session Correlation	-	< 6 µs/pair

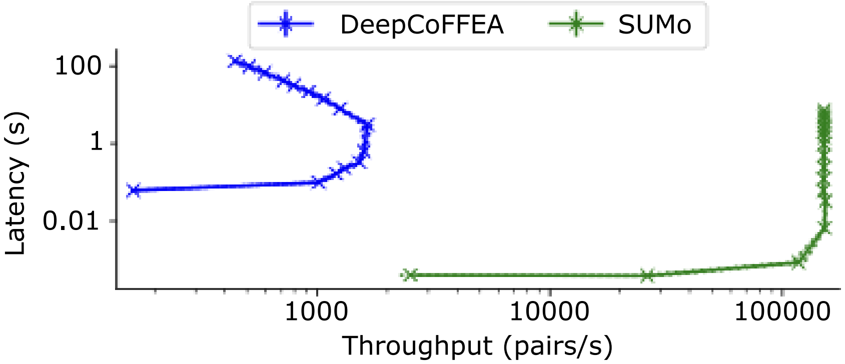
Fast to re-train!



SUMo is fast!

Phase	**Stage**	**Training time**	**Inference Time**
Filtering	Source Separation Target Separation	< 6 seconds total	< 4 µs/flow
Matching	Session Correlation	-	< 6 µs/pair

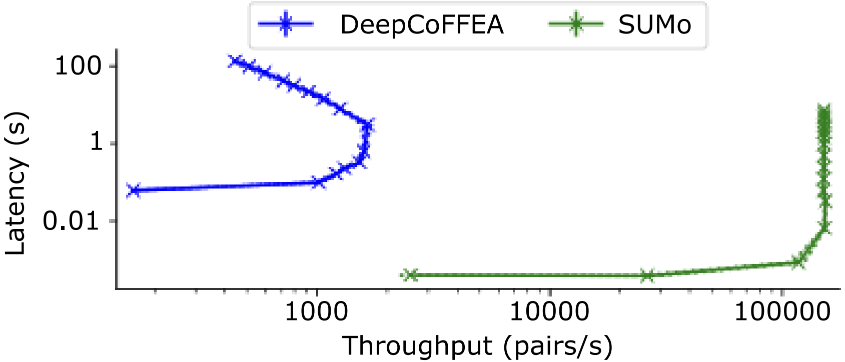
Fast to re-train!



SUMo is fast!

Phase	Stage	Training time	Inference Time
Filtering	Source Separation Target Separation	< 6 seconds total	< 4 µs/flow
Matching	Session Correlation	-	< 6 µs/pair

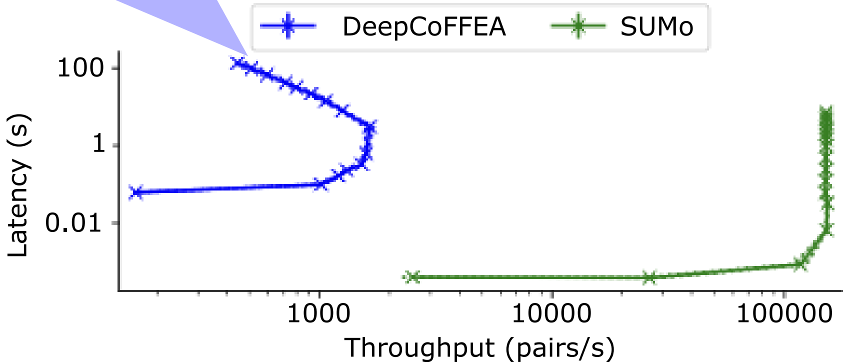
Fast to re-train!



SUMo is fast!

Phase	**Stage**	**Training time**	**Inference Time**
Filtering	Source Separation Target Separation	< 6 seconds total	< 4 µs/flow
Matching	Session Correlation	-	< 6 µs/pair

Deep learning correlation attack of Tor traffic to the clearweb



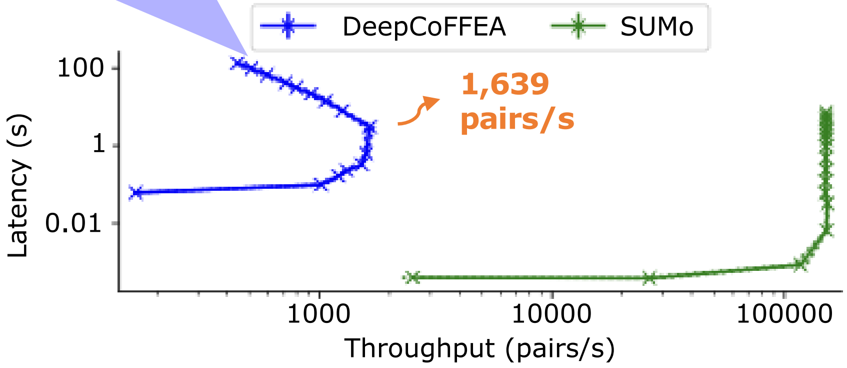
Fast to re-train!

SUMo is fast!

Phase	Stage	Training time	Inference Time
Filtering	Source Separation Target Separation	< 6 seconds total	< 4 µs/flow
Matching	Session Correlation	-	< 6 µs/pair

Fast to re-train!

Deep learning correlation attack of Tor traffic to the clearweb



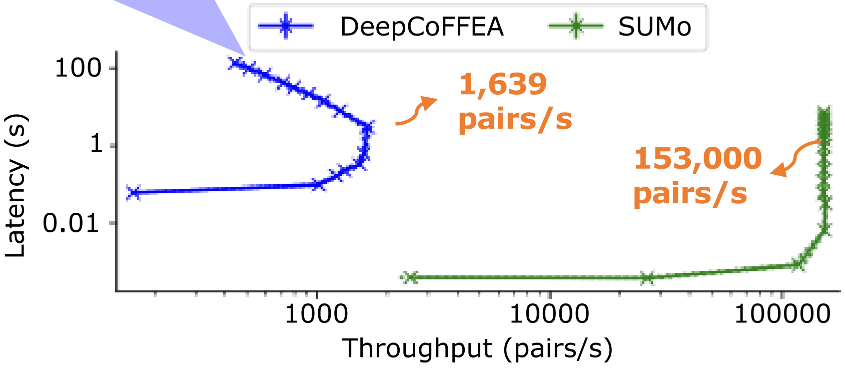
1,639 pairs/s

SUMo is fast!

Phase	Stage	Training time	Inference Time
Filtering	Source Separation Target Separation	< 6 seconds total	< 4 µs/flow
Matching	Session Correlation	-	< 6 µs/pair

Fast to re-train!

Deep learning correlation attack of Tor traffic to the clearweb



1,639 pairs/s

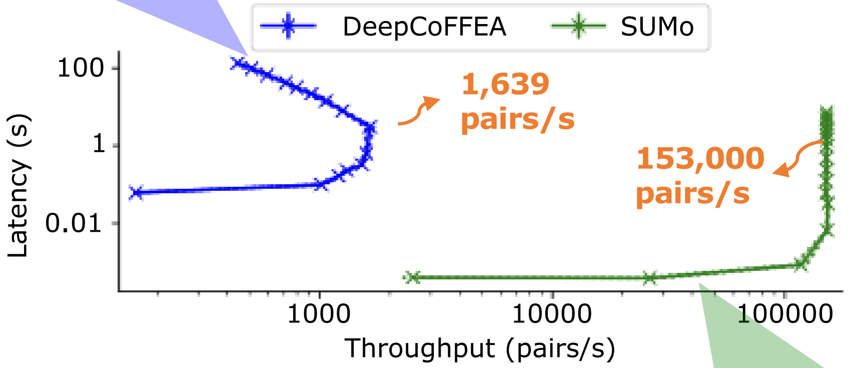
153,000 pairs/s

SUMo is fast!

Phase	Stage	Training time	Inference Time
Filtering	Source Separation	< 6 seconds total	< 4 µs/flow
	Target Separation		
Matching	Session Correlation	-	< 6 µs/pair

Fast to re-train!

Deep learning correlation attack of Tor traffic to the clearweb



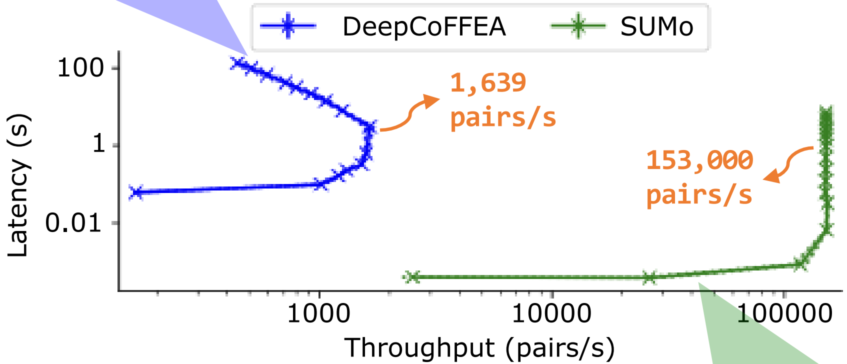
GPU-optimized correlation attack of Tor traffic to onion services

SUMo is fast!

Phase	Stage	Training time	Inference Time
Filtering	Source Separation Target Separation	< 6 seconds total	< 4 µs/flow
Matching	Session Correlation	-	< 6 µs/pair

Fast to re-train!

Deep learning correlation attack of Tor traffic to the clearweb

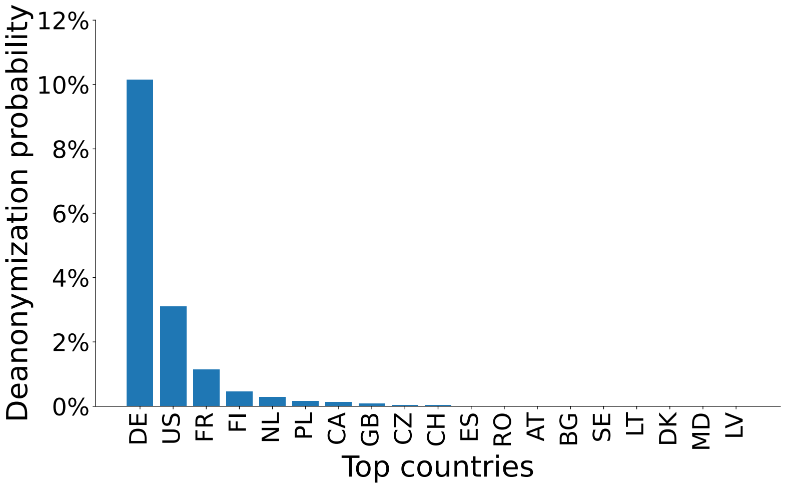


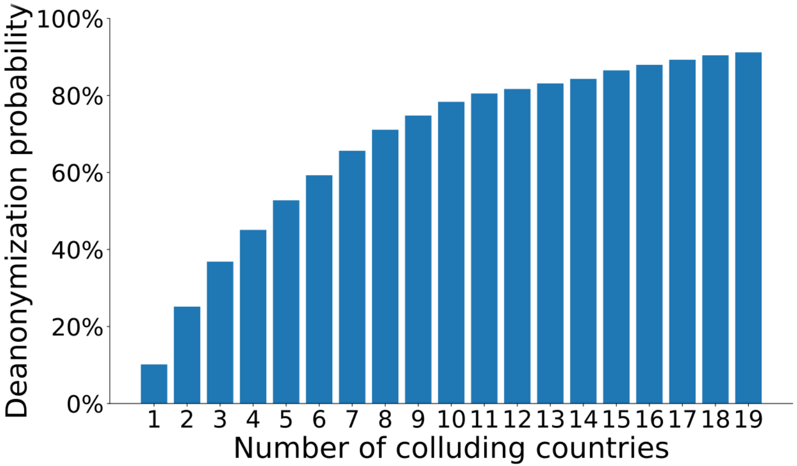
GPU-optimized correlation attack of Tor traffic to onion services

SUMo is 100x faster than the state-of-the-art!

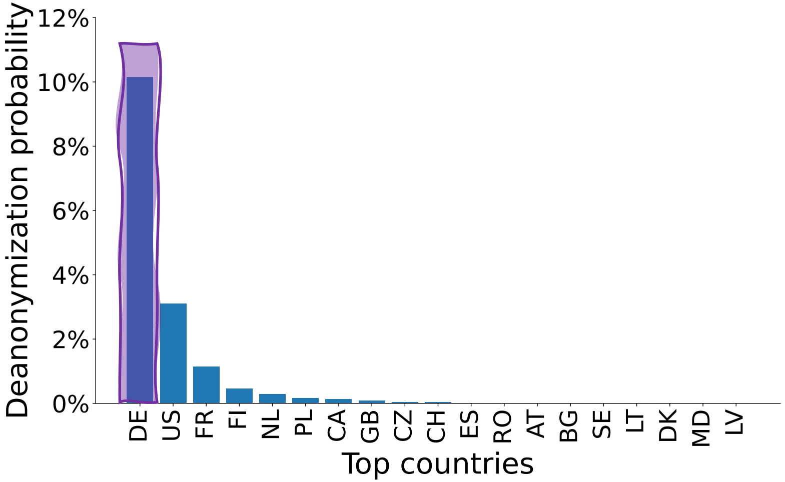
Correlation is a real threat!

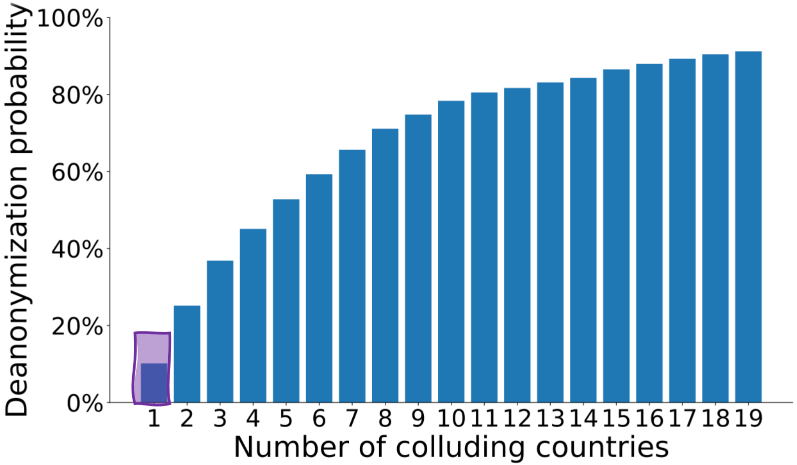
Correlation is a real threat!



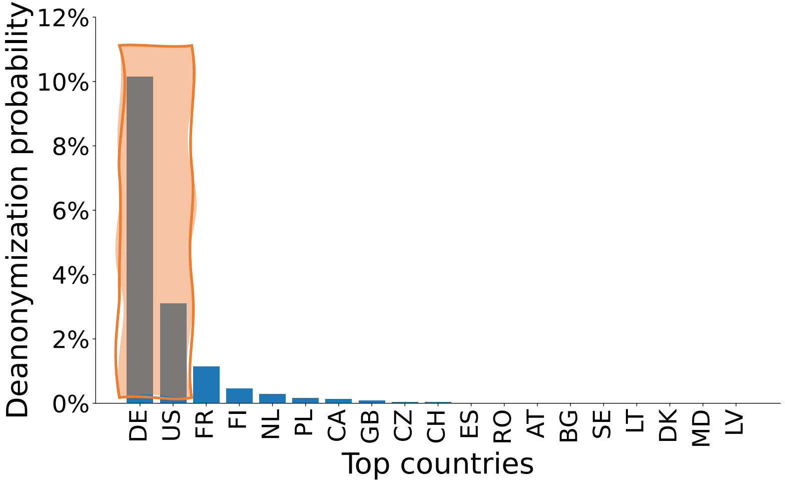


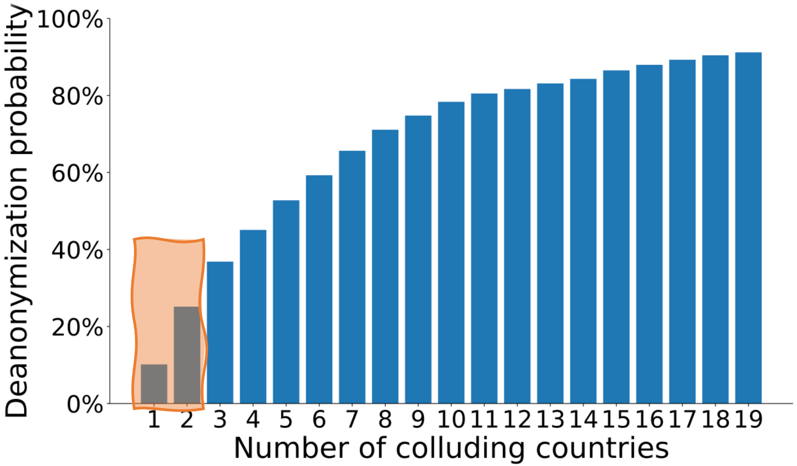
Correlation is a real threat!



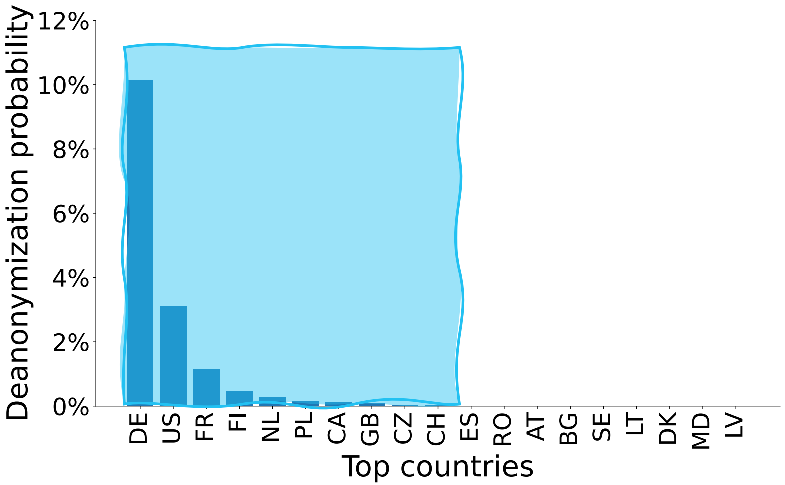


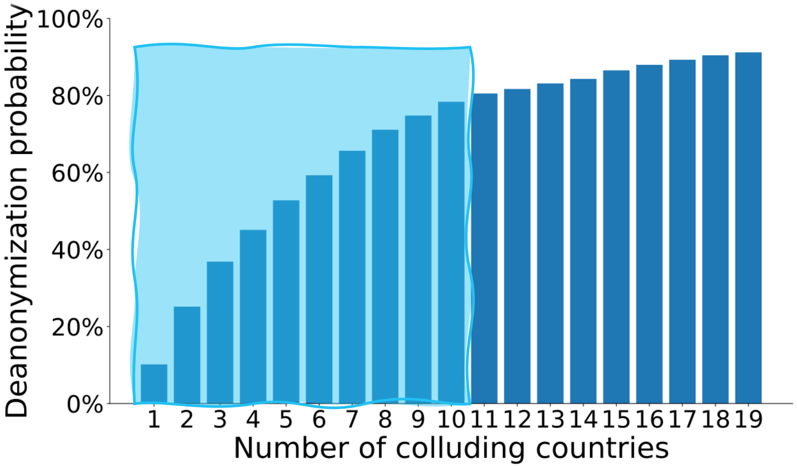
Correlation is a real threat!



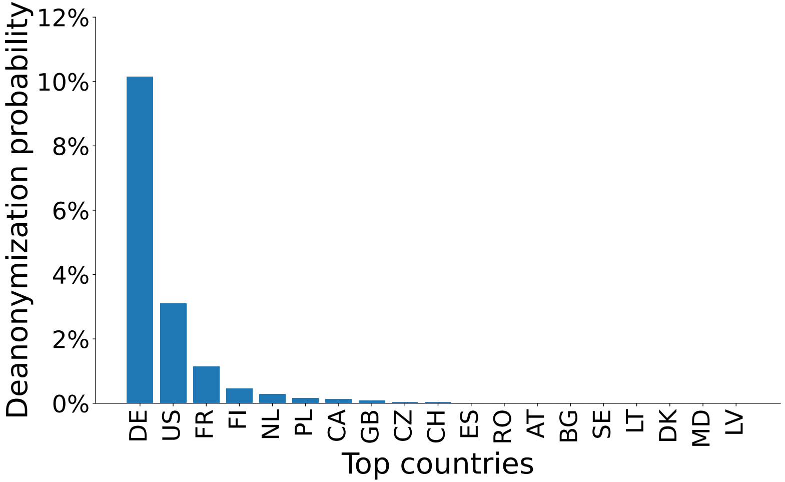


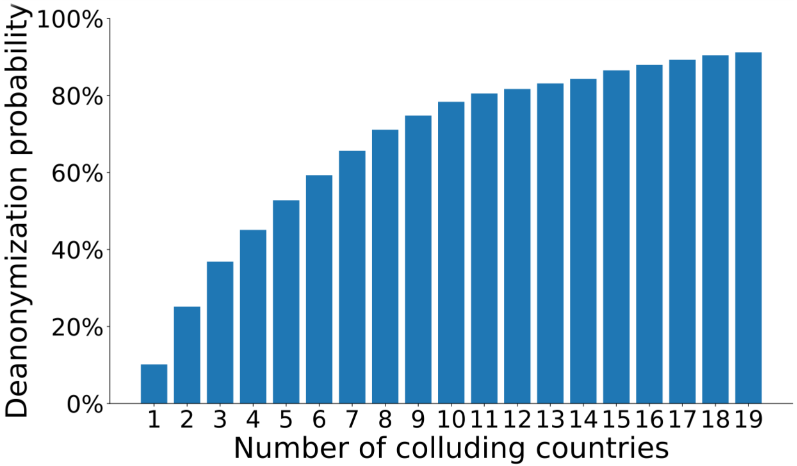
Correlation is a real threat!





Correlation is a real threat!





Guard node attribution is dangerously skewed!

Conclusion

Conclusion

- SUMo is effective at deanonymizing onion services.

Conclusion

- SUMo is effective at **deanonymizing** onion services.
- **Existing entities** can realistically deploy SUMo.

Conclusion

- SUMo is effective at deanonymizing onion services.
- Existing entities can realistically deploy SUMo.
- **Countermeasures:**

Conclusion

- SUMo is effective at **deanonymizing** onion services.
- **Existing entities** can realistically deploy SUMo.
- **Countermeasures:**
  - Pluggable transports (e.g. obfs4).

Conclusion

- SUMo is effective at **deanonymizing** onion services.
- **Existing entities** can realistically deploy SUMo.
- **Countermeasures:**
  - Pluggable transports (e.g. obfs4).
  - Concurrent multitab requests.

Conclusion

- SUMo is effective at **deanonymizing** onion services.
- **Existing entities** can realistically deploy SUMo.
- **Countermeasures:**
    - Pluggable transports (e.g. obfs4).
    - Concurrent multitab requests.
    - Guard geographical diversity.

Conclusion

- SUMo is effective at **deanonymizing** onion services.
- **Existing entities** can realistically deploy SUMo.
- **Countermeasures:**
  - Pluggable transports (e.g. obfs4).
  - Concurrent multitab requests.
  - Guard geographical diversity.

Scan for source code

Conclusion

- SUMo is effective at **deanonymizing** onion services.
- **Existing entities** can realistically deploy SUMo.
- **Countermeasures:**
  - Pluggable transports (e.g. obfs4).
  - Concurrent multitab requests.
  - Guard geographical diversity.

Get in touch!

✉ daniela.lopes@tecnico.ulisboa.pt

Scan for source code