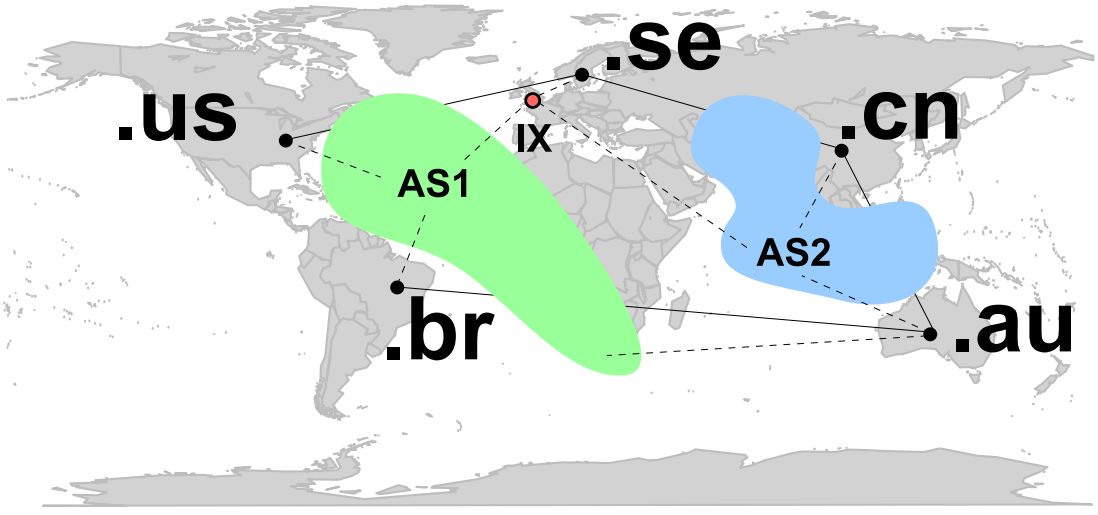
Sampled Traffic Analysis by Internet-Exchange-Level Adversaries



Steven J. Murdoch [1, 2] Piotr Zieliński [1]

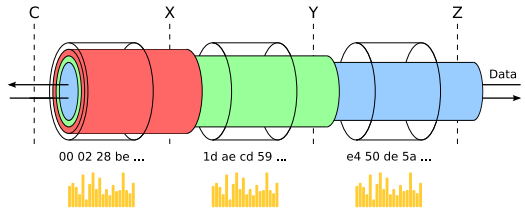
www.cl.cam.ac.uk/users/{sjm217, pz215}

[1] UNIVERSITY OF CAMBRIDGE
Computer Laboratory

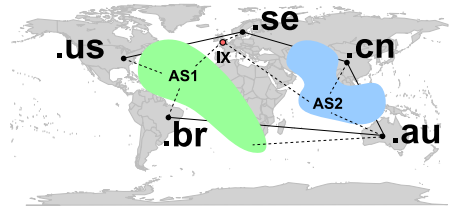
[2] OpenNet Initiative
www.opennet.net

7th Workshop on Privacy Enhancing Technologies, 20–22 June 2007, Ottawa, Canada

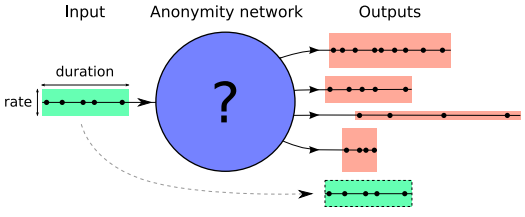
This talk shows the impact of Internet exchanges on anonymity



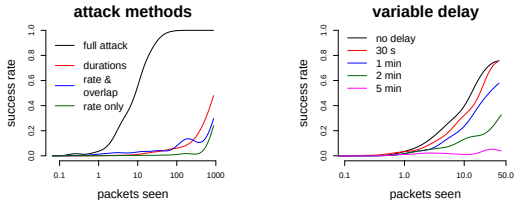
Traffic analysis of low-latency anonymity systems



Internet exchanges as a traffic analysis point

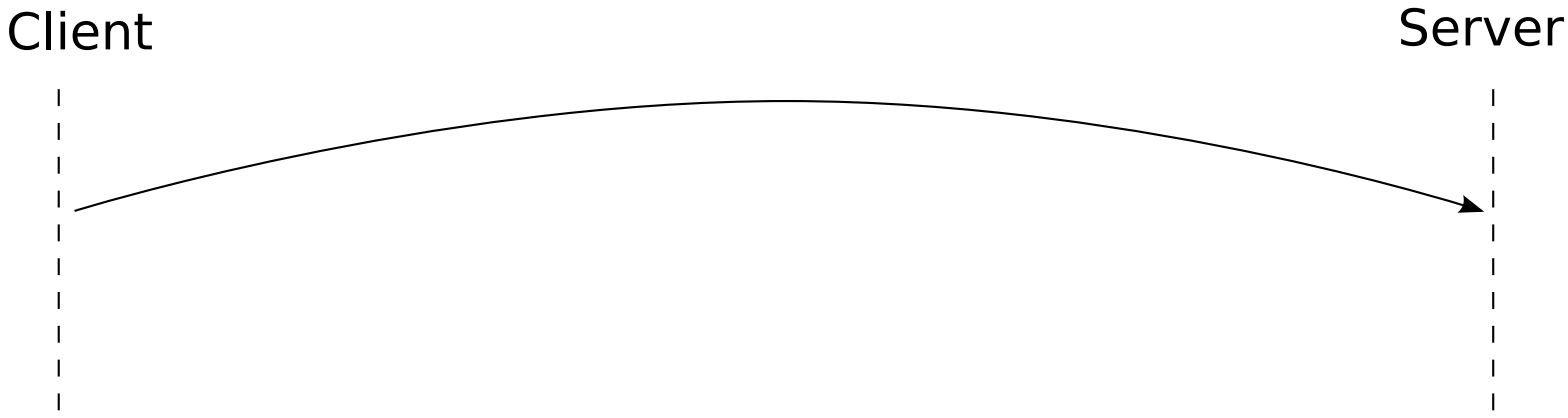


Performing traffic analysis with sampled data



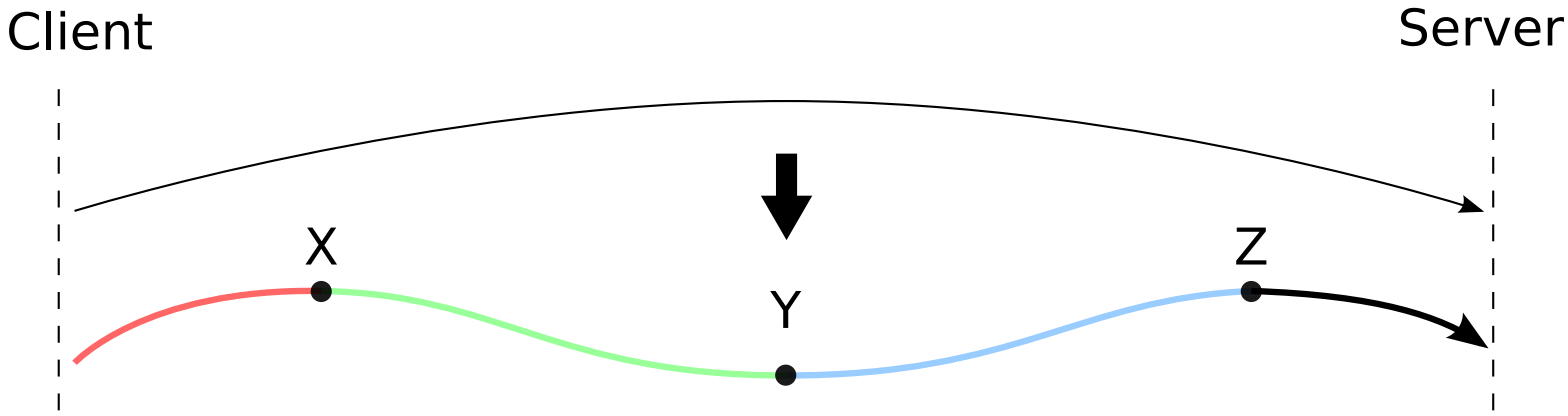
Effectiveness of the attack

Connecting directly to a server leaks information about users' behaviour



Anyone monitoring the client, server or the connection between them can see that the client is accessing that server

Connecting directly to a server leaks information about users' behaviour



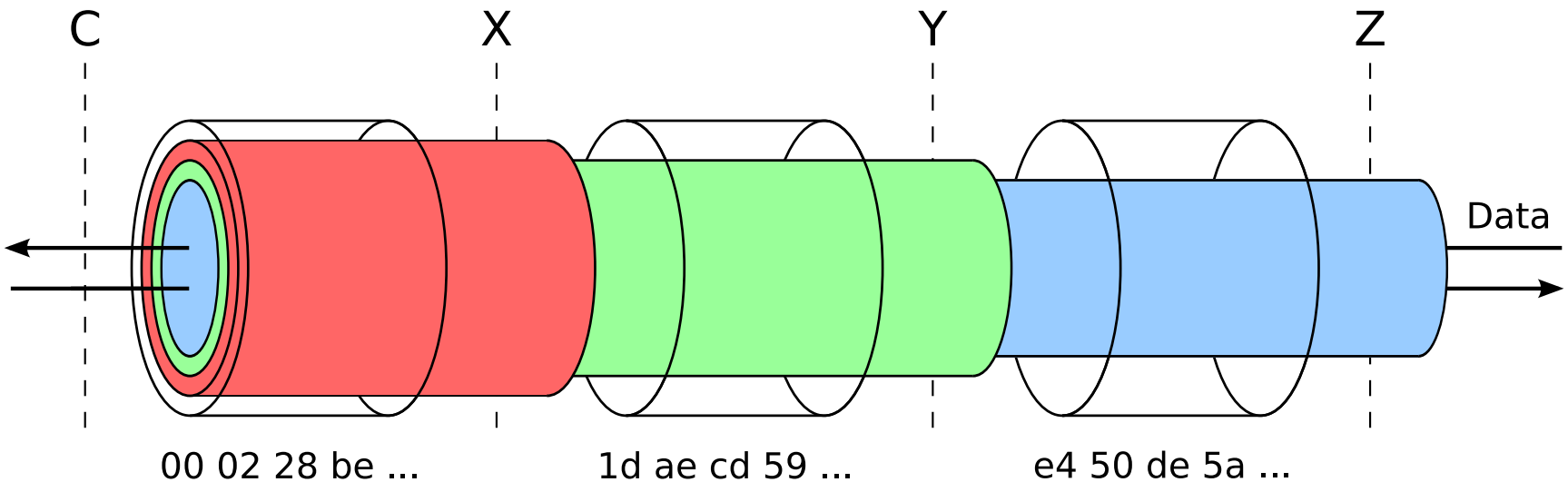
Anyone monitoring the client, server or the connection between them can see that the client is accessing that server

↓

By routing the connection through intermediate nodes, the client's on-line privacy is improved

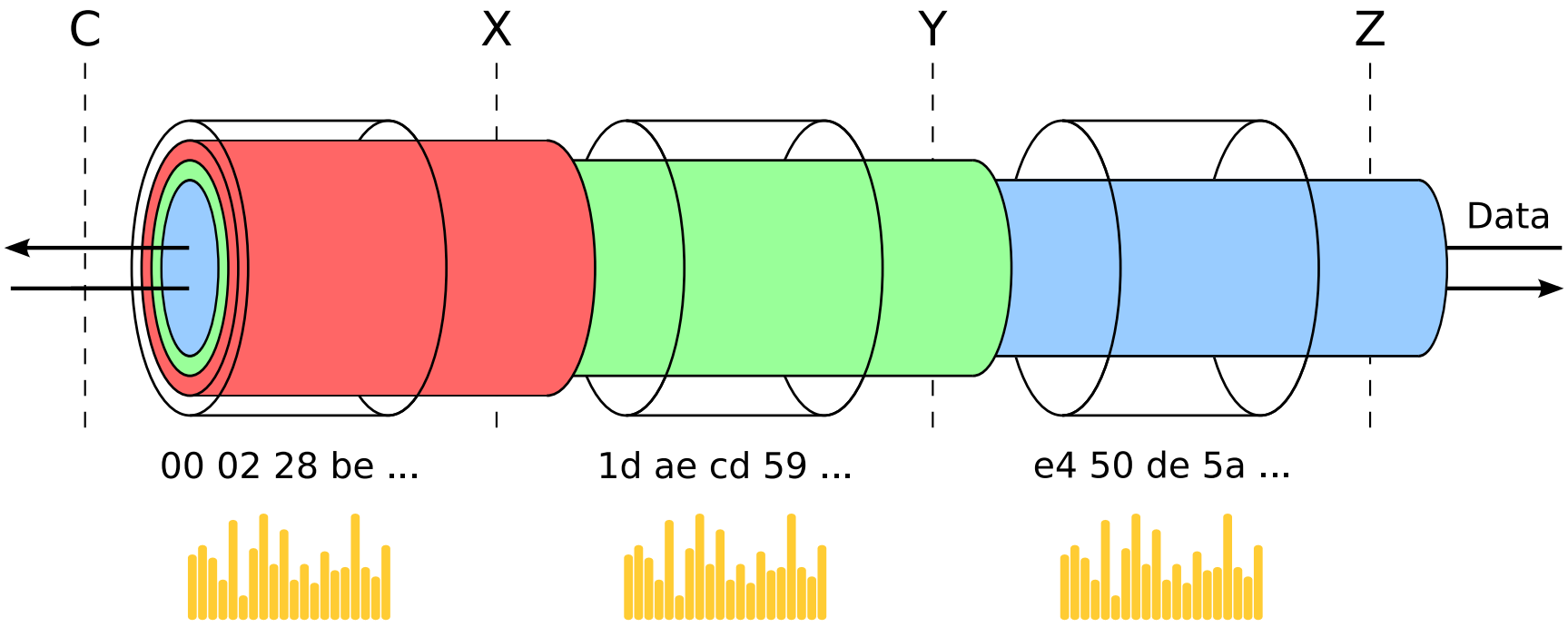
X knows the client's IP address; Z knows the server's IP address, but no node can see both; the server only knows Z's IP address

Tor hides content but not data rate so is vulnerable to traffic analysis



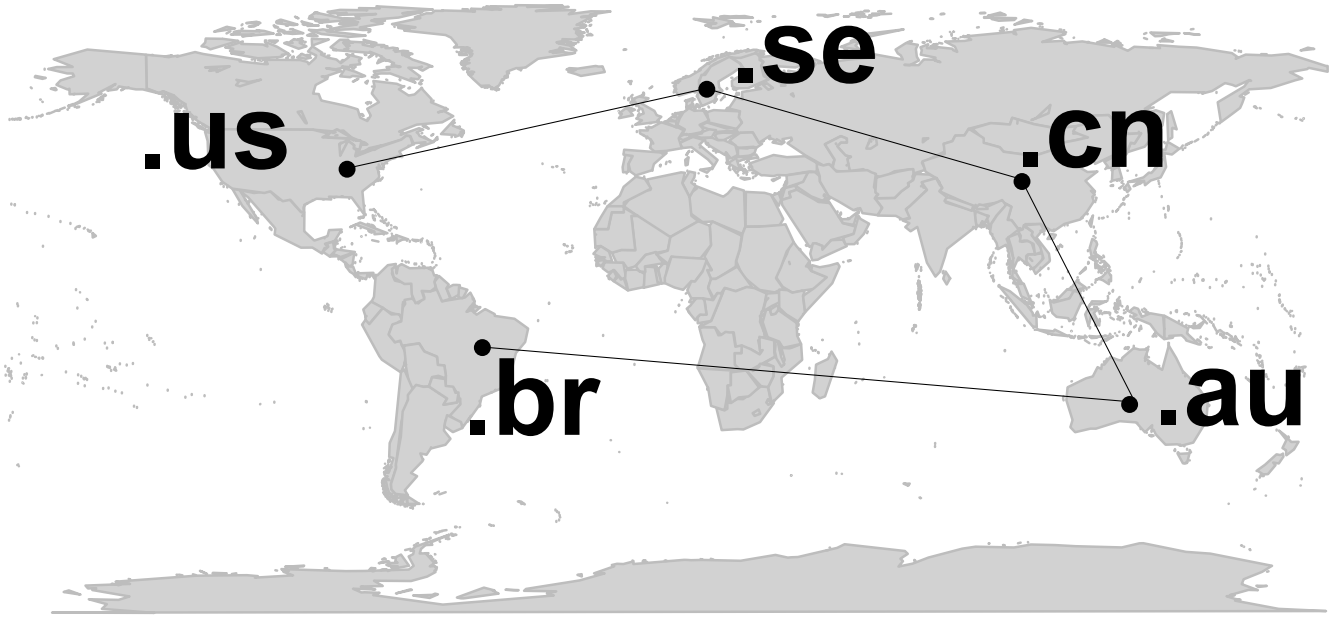
Layered encryption makes data entering and leaving a node unlinkable

Tor hides content but not data rate so is vulnerable to traffic analysis



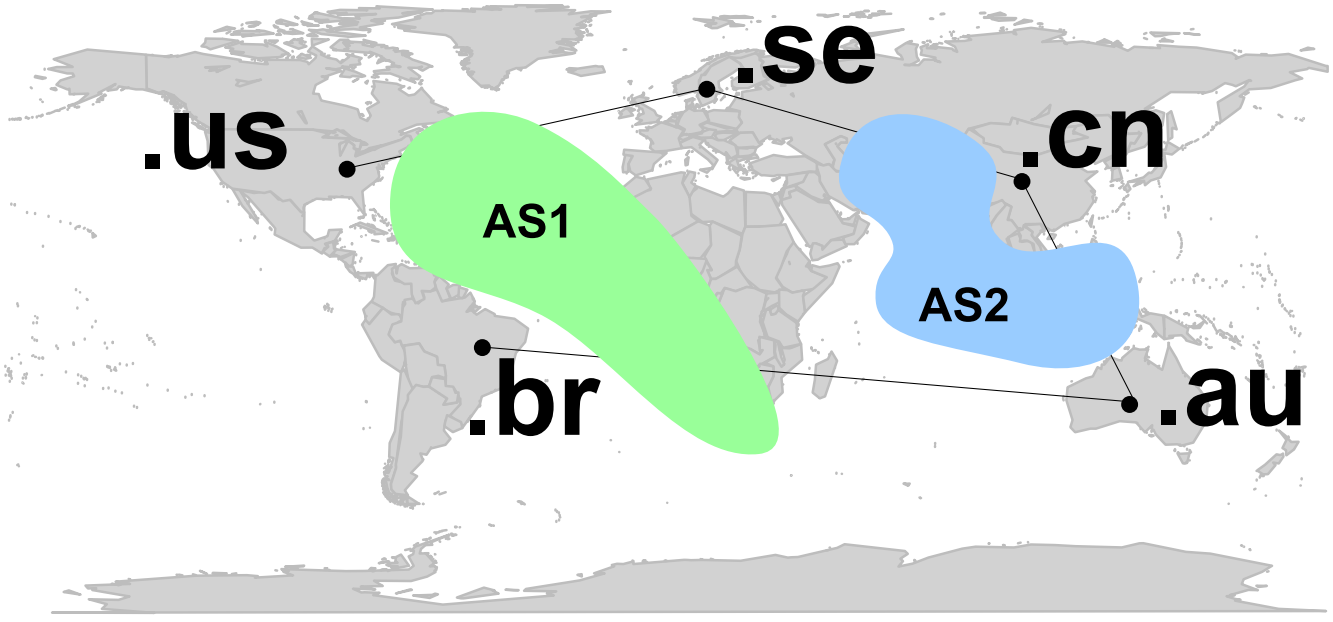
Layered encryption makes data entering and leaving a node unlinkable
But data rate is unchanged so traffic analysis can correlate flows

Location diversity can resist traffic analysis by a partial adversary



Jurisdictional model: attacker can monitor nodes in some countries

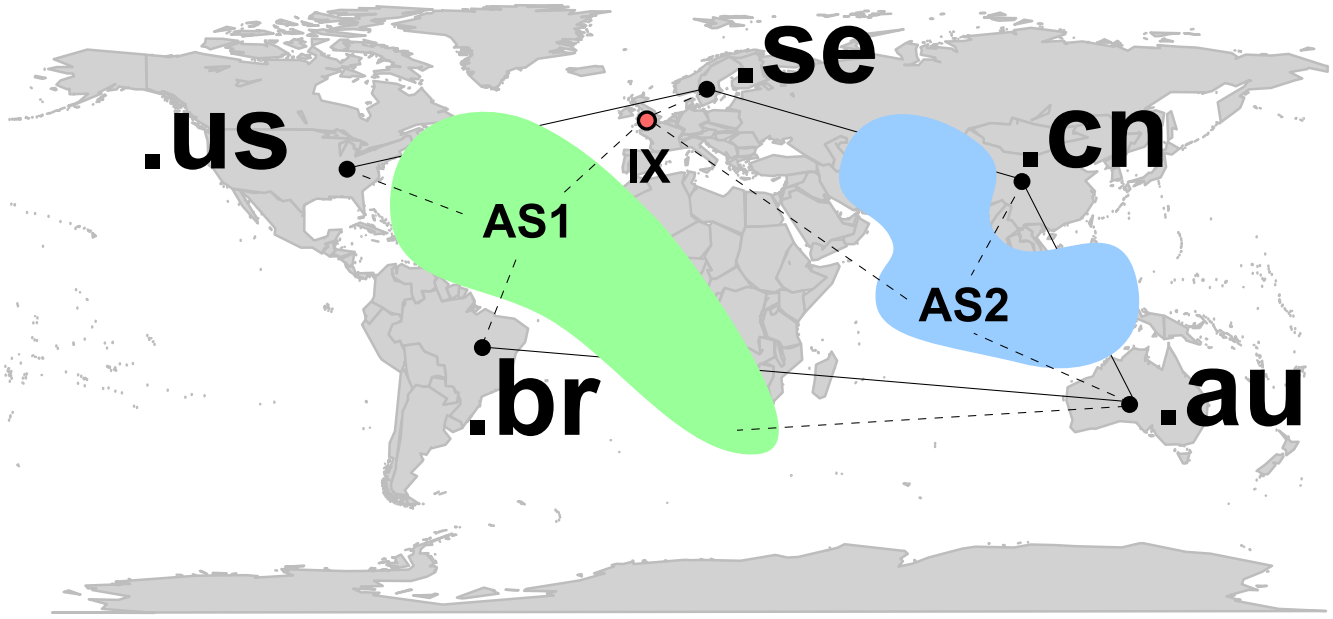
Location diversity can resist traffic analysis by a partial adversary



Jurisdictional model: attacker can monitor nodes in some countries

AS (autonomous system) model: attacker can monitor traffic flowing through some ISPs [Feamster & Dingledine]

Location diversity can resist traffic analysis by a partial adversary



Jurisdictional model: attacker can monitor nodes in some countries

AS (autonomous system) model: attacker can monitor traffic flowing through some ISPs [Feamster & Dingledine]

IX model: attacker can monitor links passing through some points

Internet exchanges are strategically powerful locations for traffic analysis

AS name	Paths	%
Level 3	1 961	**22%**
NTL	1 445	16%
Zen	1 258	14%
JANET	1 224	14%
⋮		

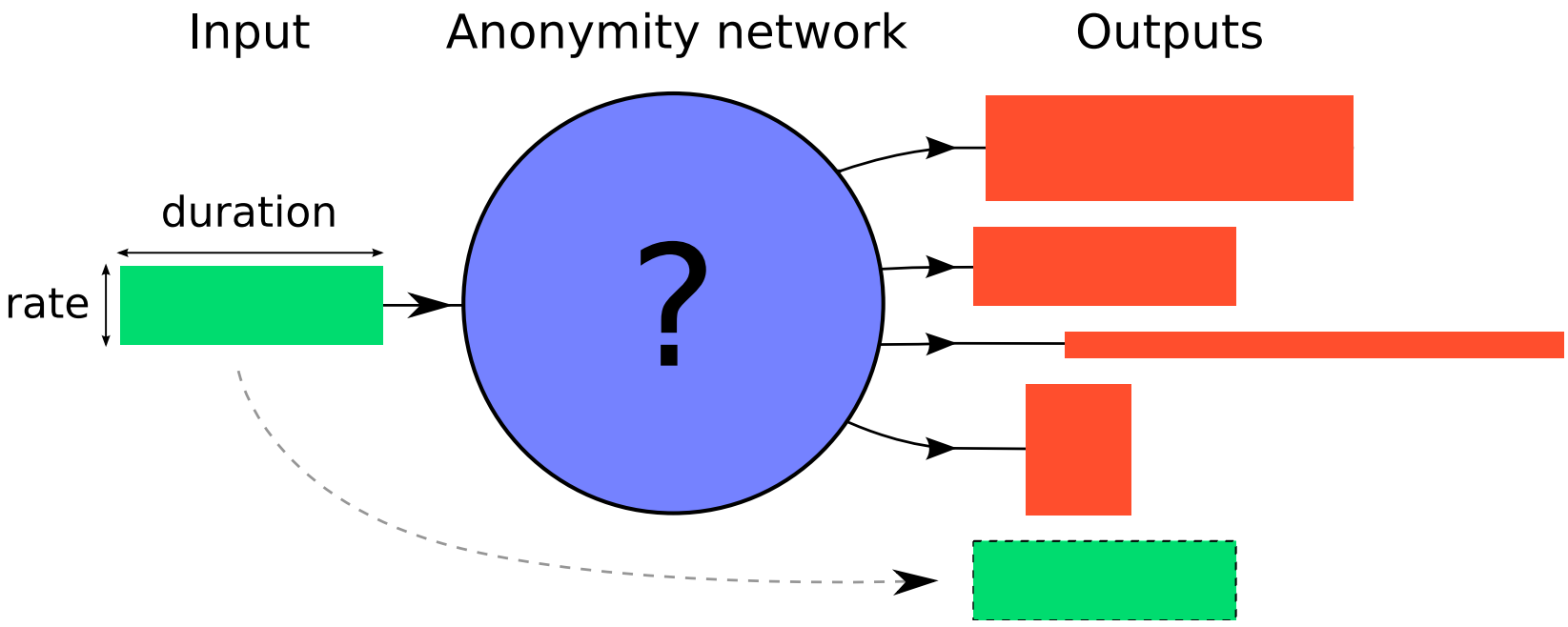
Internet exchange	Paths	%
LINX	2 392	**27%**
DE-CIX	231	3%
AMS-IX	202	2%

For Tor nodes in the UK, the LINX (London Internet Exchange) is on more paths than any other ISP

LINX records and stores (partial) data from some of their core switches, and it is planned to be used for detecting spammers

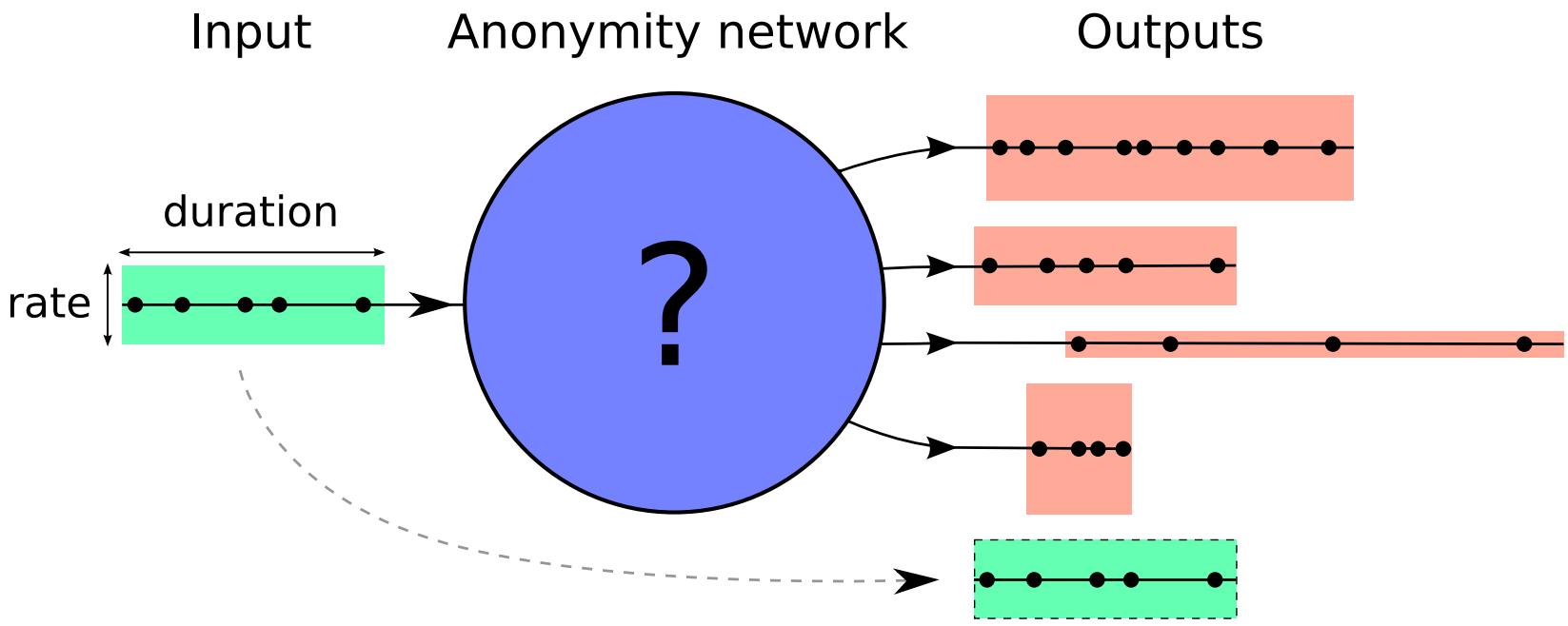
AMS-IX records data too, but only used for generating statistics

Traffic data can be used to link flows, but only sampled data may be available



Attacker's goal is to establish probability that each output flow corresponds to the input flow of interest

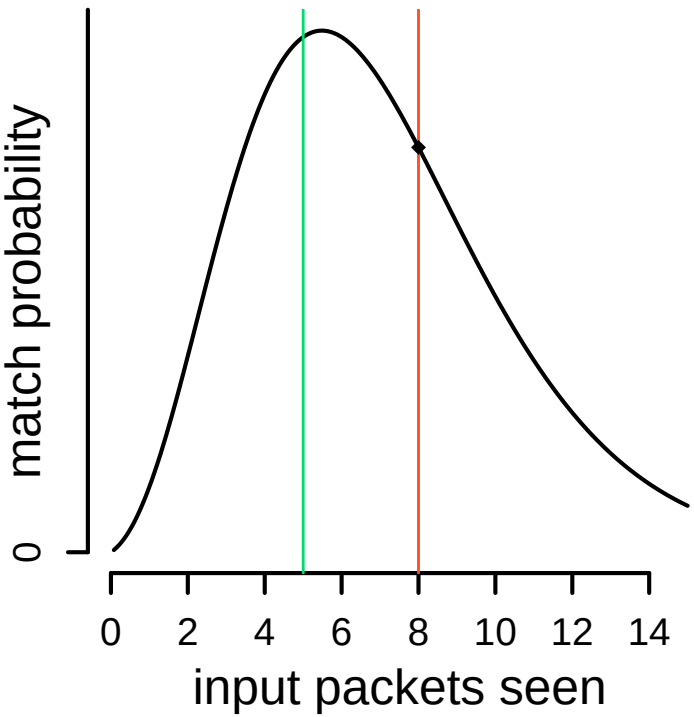
Traffic data can be used to link flows, but only sampled data may be available

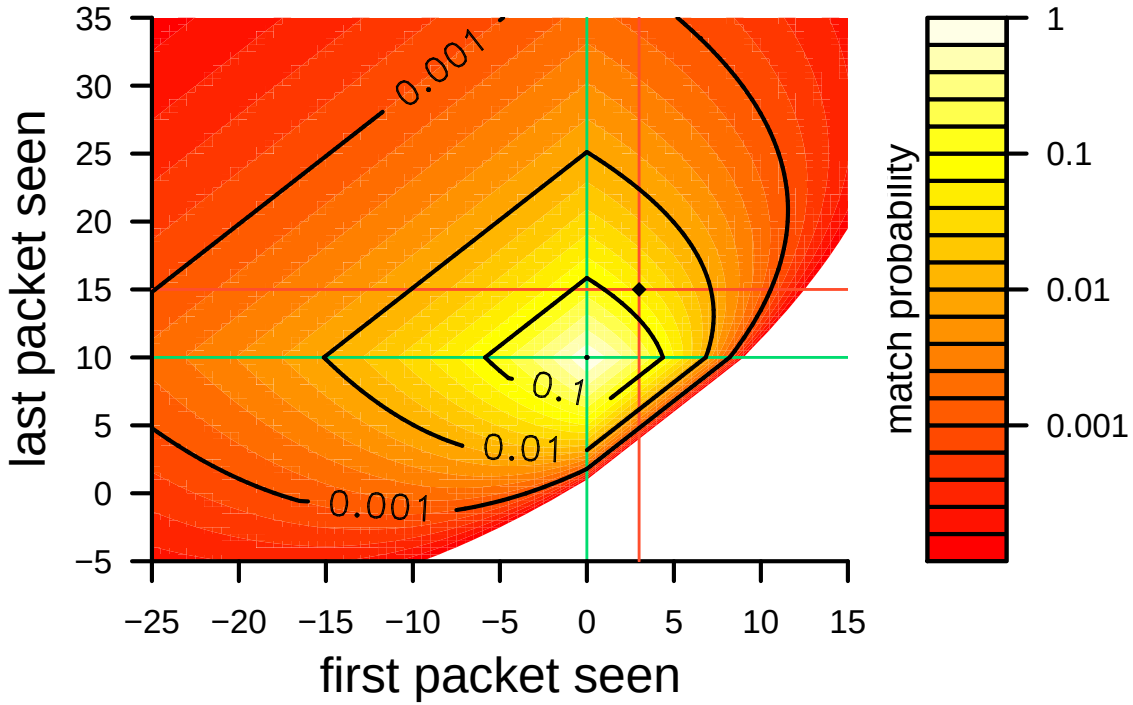


Attacker's goal is to establish probability that each output flow corresponds to the input flow of interest

For fast links only sampled data is available (1 in 2 048 for LINX)

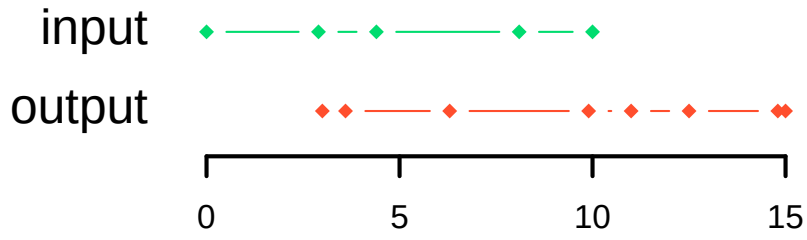
Bayesian analysis shows only flow rates and overlap are significant



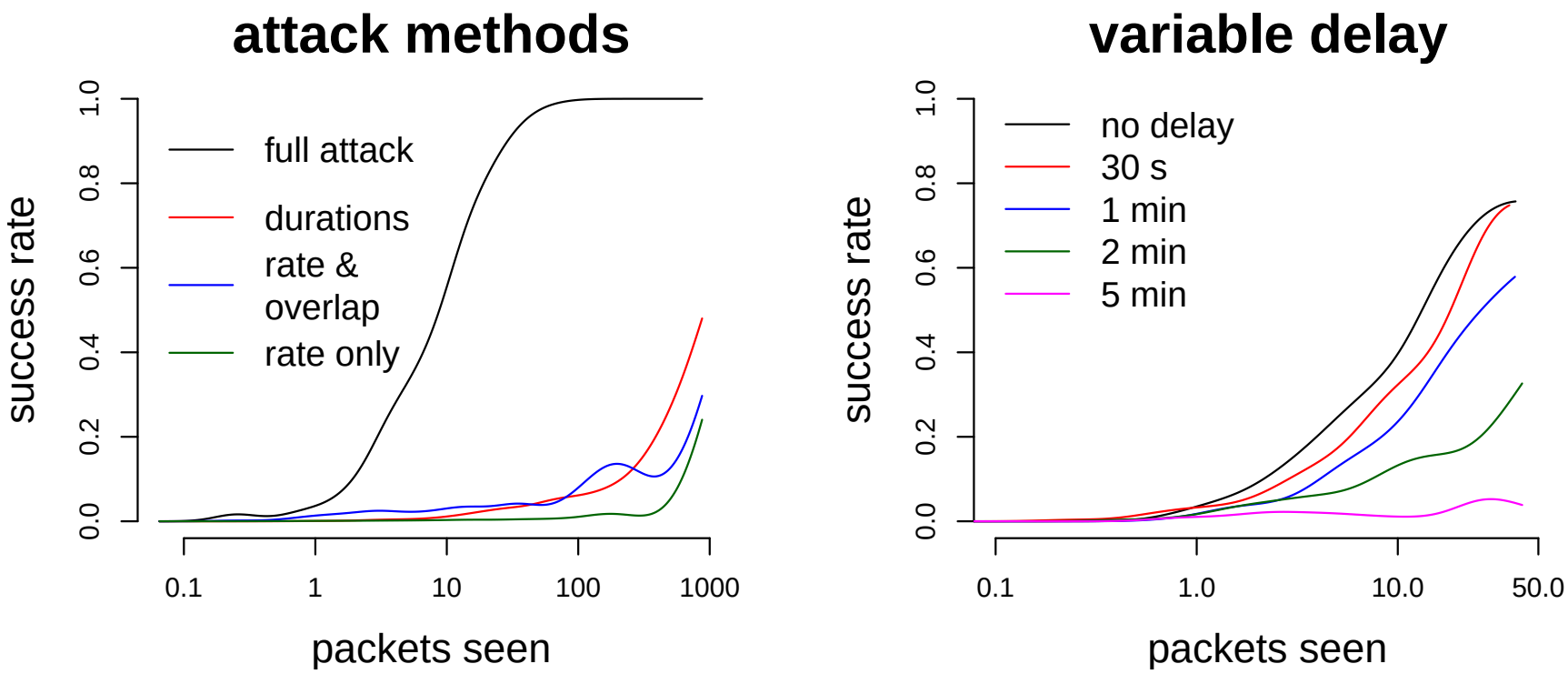


Match probability high when

- input and output **rates similar**
- amount of **overlap high**



Results of analysis show high accuracy and resistance to moderate delay



Using both rate and amount of overlap significantly improves the accuracy of results; (50% success rate after ≈10 MB of traffic)

Introducing up to 30 seconds of latency to flows has no significant effect on the matching algorithm

In summary, Internet exchanges are ideal locations for traffic analysis

- Internet exchanges are present on a high proportion of Internet connections and may have the capability for collecting traffic data
- Sampled data, possible to collect with existing network equipment, is very effective in de-anonymising flows

Future work

- Develop improved defences
  - Because the timing of individual packets is not a relevant factor, introducing moderate latency does not mitigate the attack
  - Dummy traffic is more promising, but comes with a high cost
  - Paths could be selected to maintain Internet exchange diversity
- Refine limits of the attack's effectiveness
  - Simulate with more realistic (non-Poisson) traffic
  - Analyze traffic within the anonymity network
  - Consider more information (e.g. sequence numbers in sFlow)