Invisible Internet Project (I2P)

Tim de Boer and Vincent Breider

Research question(s)

Is it possible for an entity that intercepts network traffic to fingerprint and positively identify hosts that are participating in the I2P network?

Sub-questions:

- How does the I2P network work, how does the protocol operate?
- Can traffic be identified as I2P during the bootstrapping/initialisation phase of the protocol?
- Can traffic be identified as I2P by scraping the netDb distributed hash table?
- Is fingerprinting of the protocol itself possible using statistical analysis based on connection meta-data?

Related work

Bazli et al, investigated how forensic investigation into the I2P network could be conducted, by examining the forensic artefacts of the I2P installer.

Timpanaro et al, performed a study in which they design a distributed monitoring system for the I2P network.

Hjelmvik and John, looked closer on how statistical analysis can be used to identify network protocols.

How does I2P work?

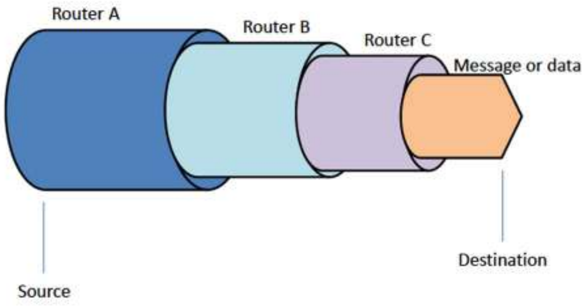
Like TOR it uses Onion Routing and communicates as a mixnet.

However it is decentralised and gathers information on other network participants via the Network Database (netDb) which is implemented as a distributed hash table.

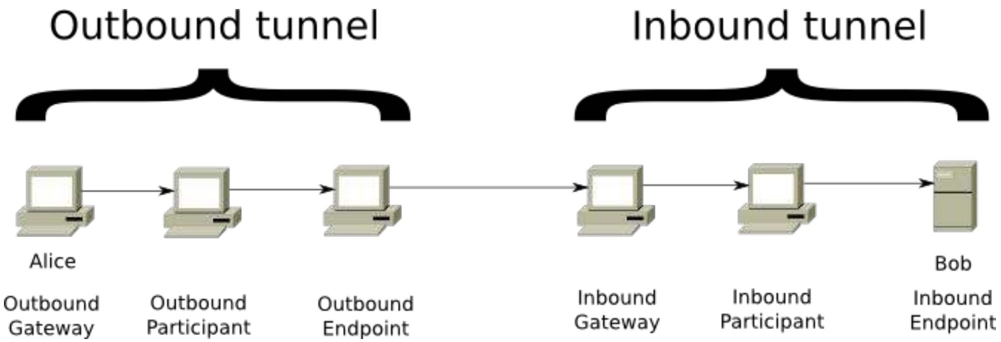
Routers always relay each others traffic, build multi-hop tunnels for anonymity and participate in each others tunnels.

To make statistical analysis harder, routers collect and pack multiple messages in one packet, this is called garlic routing.

How does I2P work? - Onion Routing

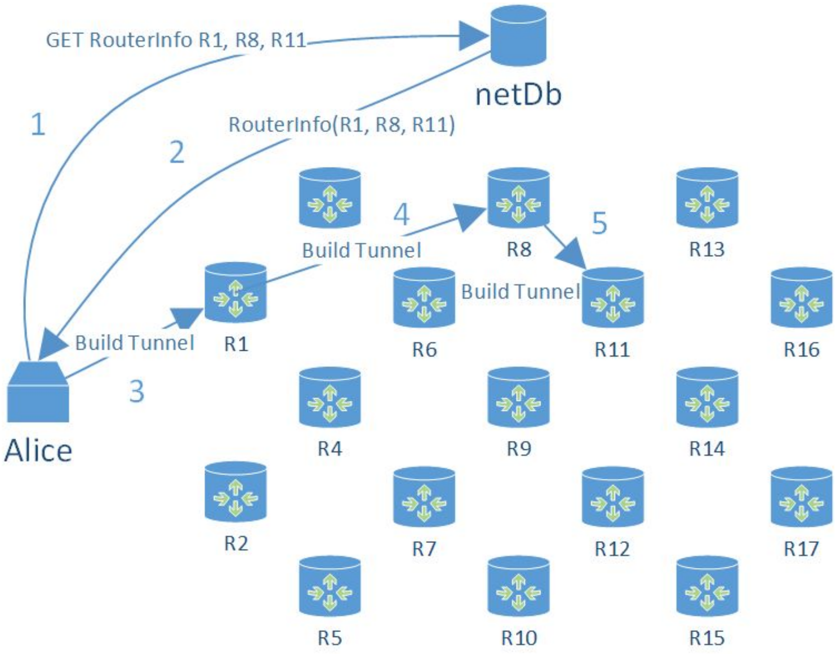


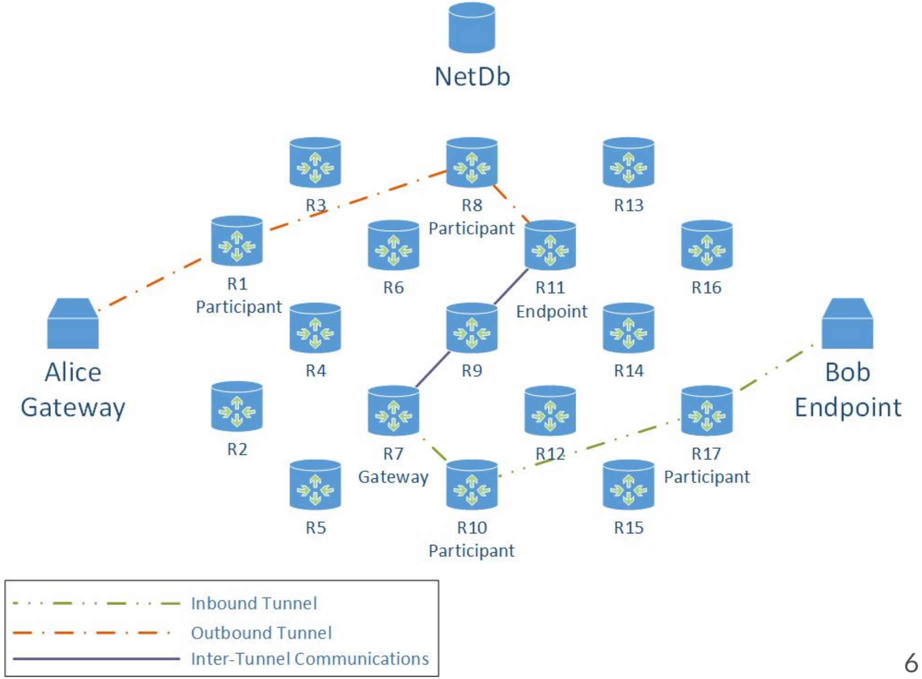
Source: https://1technation.com/tech-savvy-dark-side-onion-router/



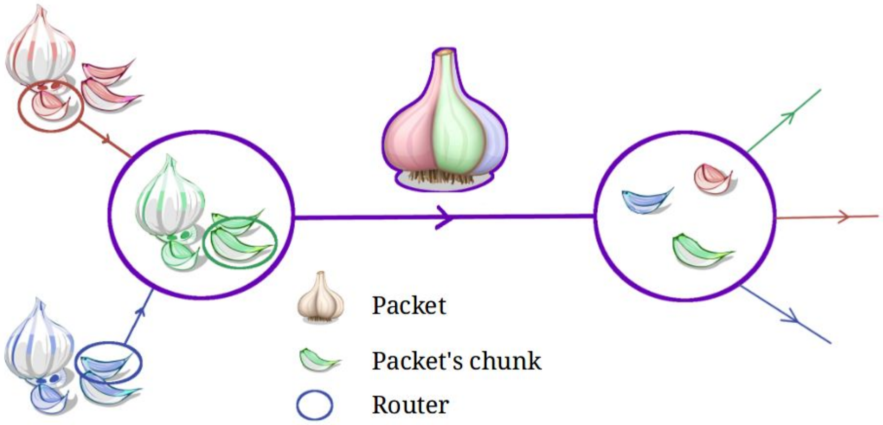
Source: https://geti2p.net/en/docs/how/tech-intro

How does I2P work? - Tunnel Establishment

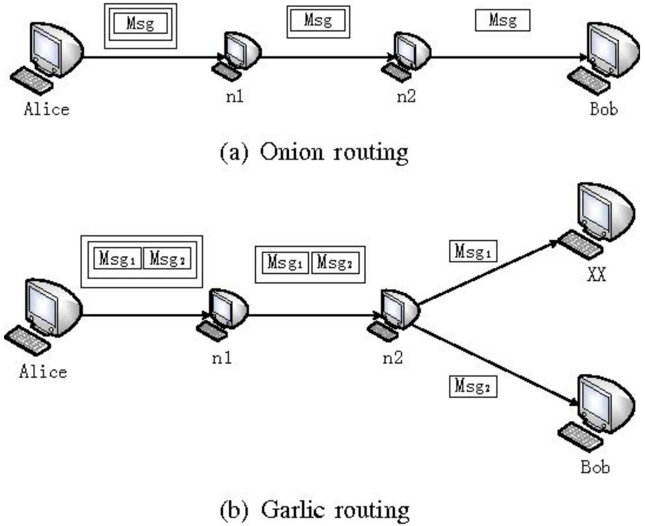




How does I2P work? - Garlic Routing



Source: The Invisible Internet Project - Andrew Savchenko, FOSDEM 2018



(a) Onion routing

(b) Garlic routing

Source: An Empirical Study of the I2P Anonymity Network and its Censorship Resistance - Nguyen et al.

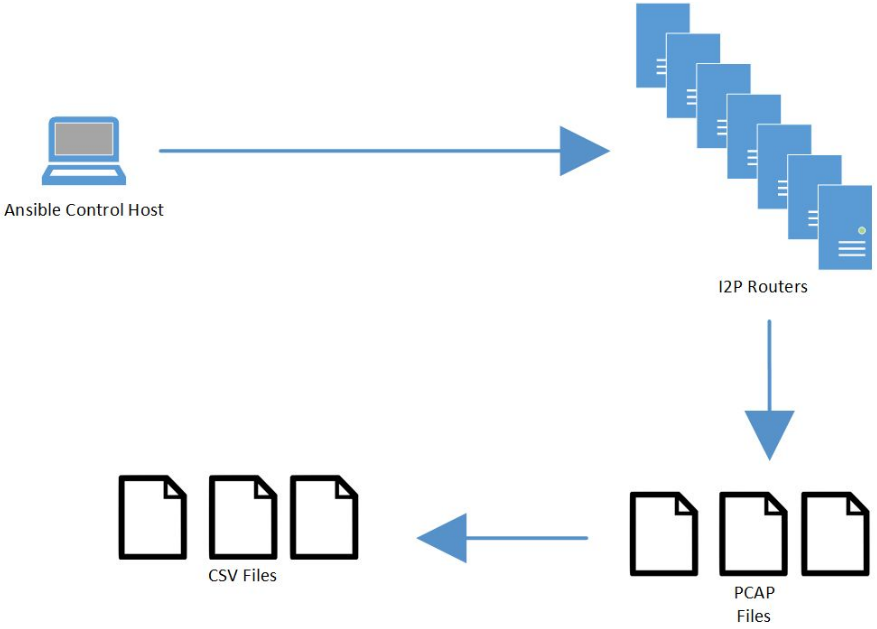
Lab environment

Deployed using Infrastructure as Code with Ansible.

6 VMs with I2P routers participating in the live network.

After a router deployment, network traffic is automatically captured using tcpdump.

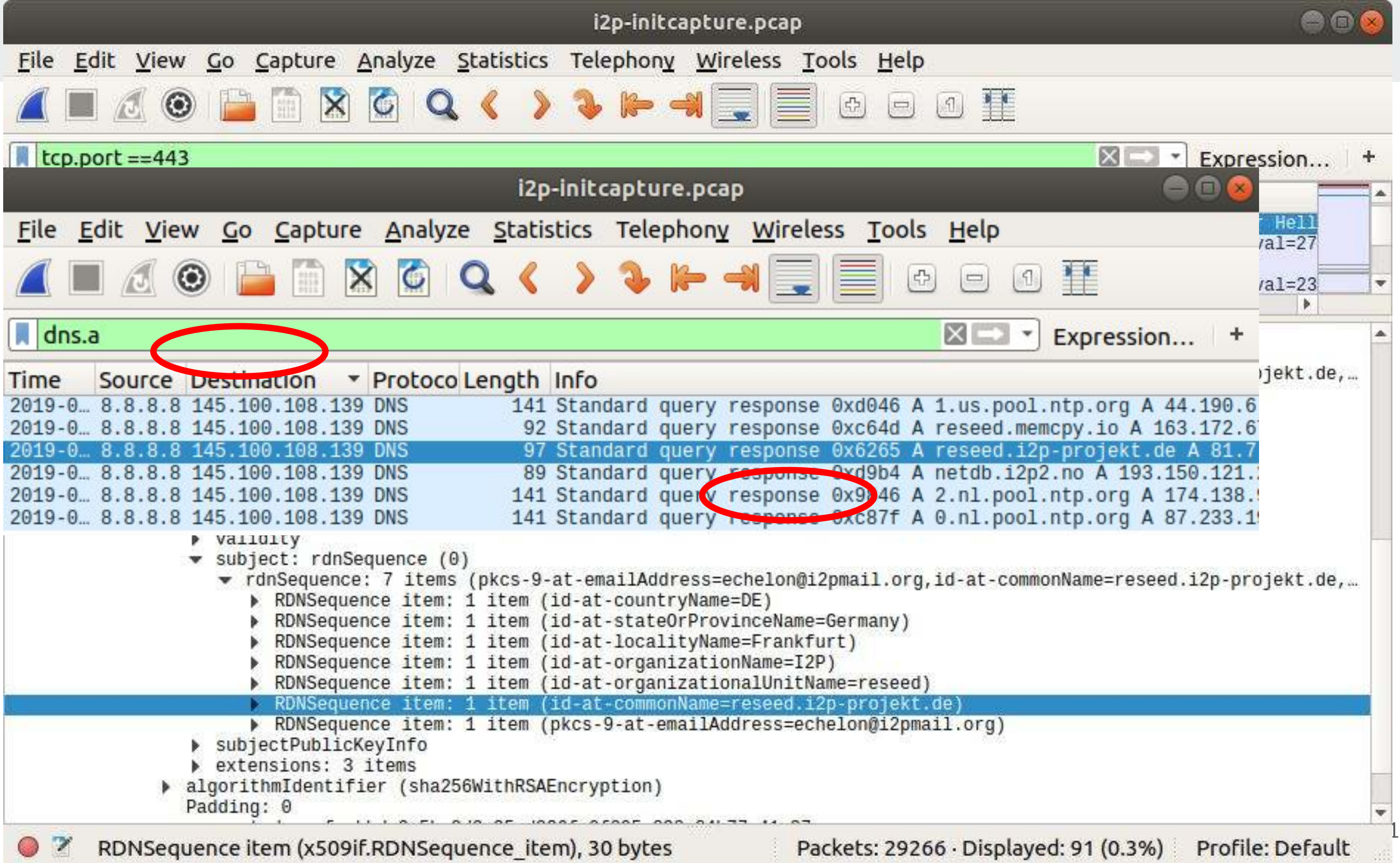
PCAPs are parsed to CSV using Bash. Statistics are extracted and anonymised using Python and R.



Detectability of I2P

Sub-questions:

- Can traffic be identified as I2P by network analysis?
- Can traffic be identified as I2P by scraping the netDb distributed hash table?


i2p-initcapture.pcap
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Expression… +
i2p-initcapture.pcap
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Expression… +

tcp.port ==443

dns.a

```
Time      Source  Destination     Protoco Length Info
2019-0…  8.8.8.8 145.100.108.139 DNS       141 Standard query response 0xd046 A 1.us.pool.ntp.org A 44.190.6
2019-0…  8.8.8.8 145.100.108.139 DNS        92 Standard query response 0xc64d A reseed.memcpy.io A 163.172.6
2019-0…  8.8.8.8 145.100.108.139 DNS        97 Standard query response 0x6265 A reseed.i2p-projekt.de A 81.7
2019-0…  8.8.8.8 145.100.108.139 DNS        89 Standard query response 0xd9b4 A netdb.i2p2.no A 193.150.121.
2019-0…  8.8.8.8 145.100.108.139 DNS       141 Standard query response 0x9946 A 2.nl.pool.ntp.org A 174.138.
2019-0…  8.8.8.8 145.100.108.139 DNS       141 Standard query response 0xc87f A 0.nl.pool.ntp.org A 87.233.1
```

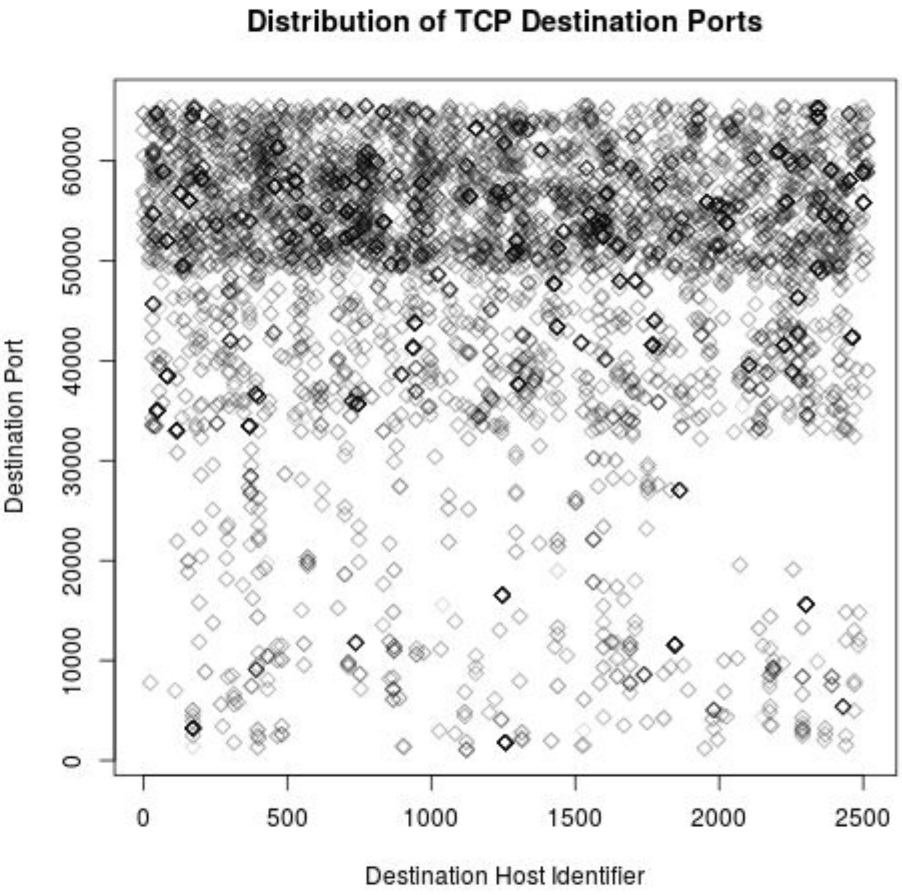
```
Hell
val=27
val=23
            validity
          subject: rdnSequence (0)
            rdnSequence: 7 items (pkcs-9-at-emailAddress=echelon@i2pmail.org,id-at-commonName=reseed.i2p-projekt.de,…
              RDNSequence item: 1 item (id-at-countryName=DE)
              RDNSequence item: 1 item (id-at-stateOrProvinceName=Germany)
              RDNSequence item: 1 item (id-at-localityName=Frankfurt)
              RDNSequence item: 1 item (id-at-organizationName=I2P)
              RDNSequence item: 1 item (id-at-organizationalUnitName=reseed)
              RDNSequence item: 1 item (id-at-commonName=reseed.i2p-projekt.de)
              RDNSequence item: 1 item (pkcs-9-at-emailAddress=echelon@i2pmail.org)
          subjectPublicKeyInfo
          extensions: 3 items
        algorithmIdentifier (sha256WithRSAEncryption)
        Padding: 0
```

RDNSequence item (x509if.RDNSequence_item), 30 bytes   Packets: 29266 · Displayed: 91 (0.3%)   Profile: Default

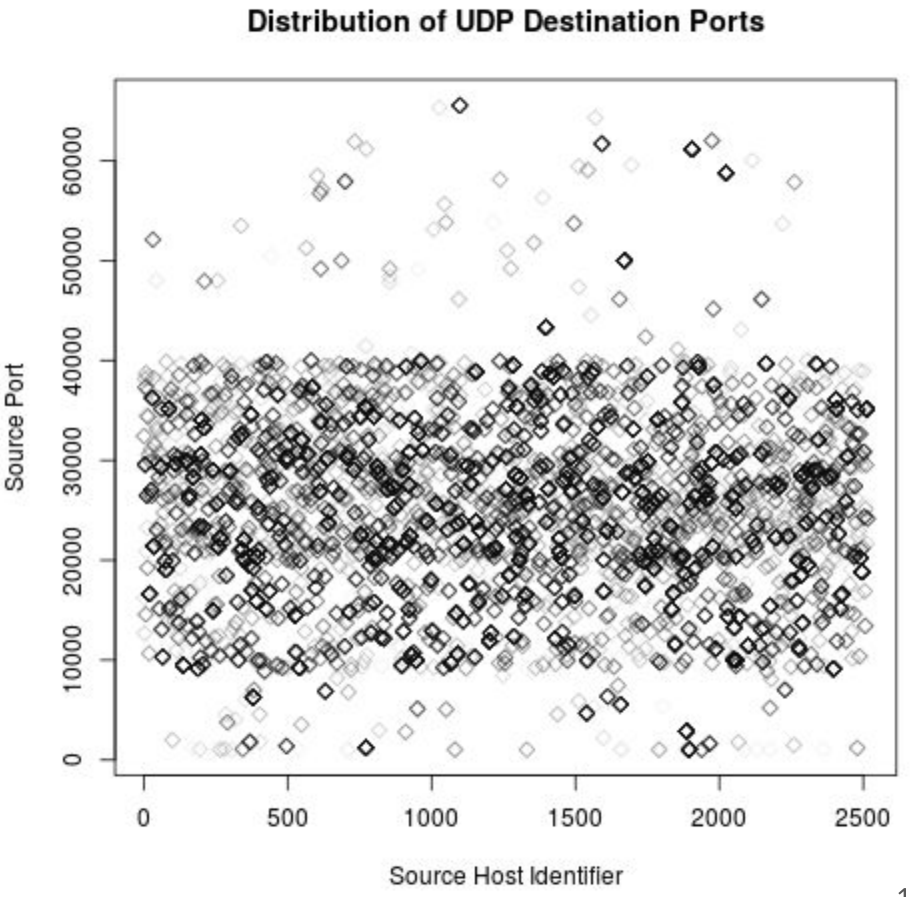
Detectability of I2P

Operational phase

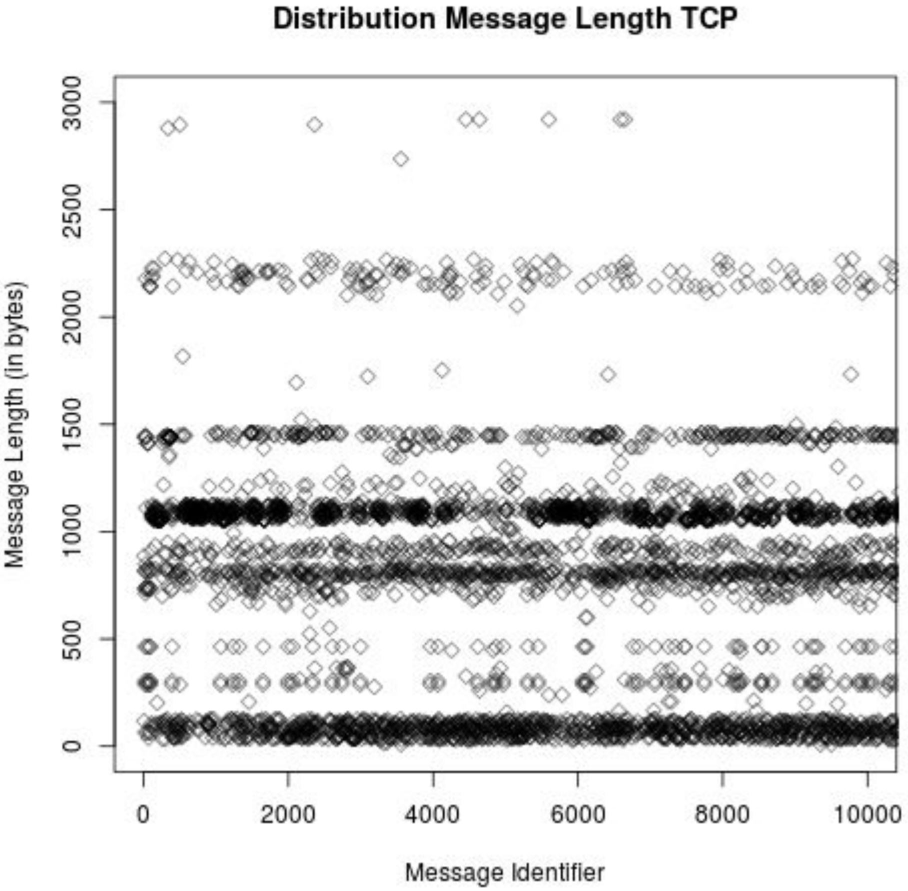
- Packets are completely encrypted without detectable and identifiable constants
- netDb parser -> IDS rules -> every other minute -> not feasible..
- Interesting results with statistical analysis



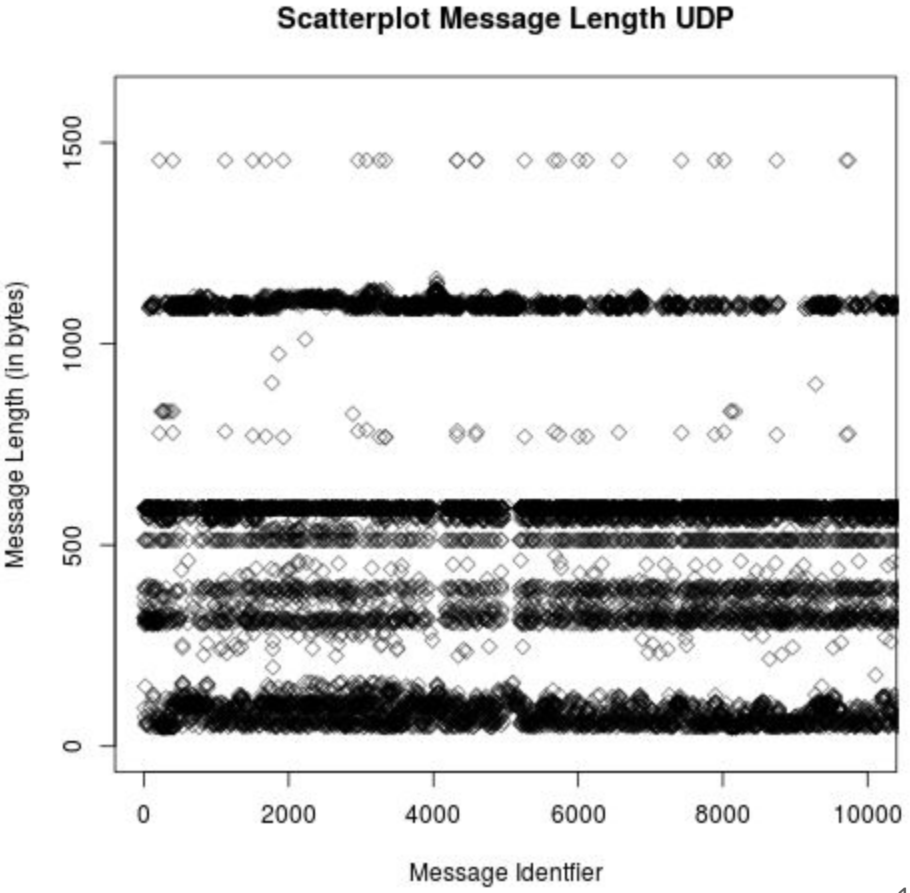
Distribution of TCP Destination Ports



Distribution of UDP Destination Ports



Distribution Message Length TCP



Scatterplot Message Length UDP

Conclusion

Is it possible for an entity that intercepts network traffic to fingerprint and positively identify hosts that are participating in the I2P network?

Initialisation phase -> Yes, under default circumstances this is trivial.

Operational phase -> Theoretically not, but potentially with use of statistical analysis or with a harvested historical netDb.

Discussion

Current patterns are difficult for a traditional IDS:
- potentially possible over a longer period of time
- requires lots of resources for mapping these data

- quickly refreshing "static rules" with IP-Address/Port combinations from netDb entries

Future research

To further investigate the message length, a follow-up study should compare our data:

- With traffic captured from a private I2P network setup, with fixed and known tunnel lengths.
- With traffic captured from other protocols that use Onion Routing, such as TOR.

Is it possible using active probing techniques to discover I2P routers?

Is it possible to exploit an I2P router, forcing it into reseeding?

Questions?