Challenges in building overlay networks: a case study of Tor

Steven Murdoch
Principal Research Fellow
University College London
Who uses Tor?

- Ordinary people
  - e.g. to avoid unscrupulous marketers, protect children, research sensitive topics
- Militaries and law enforcement
  - e.g. to protect field agents and sources, intelligence gathering, decentralise services
- Journalists and their audience
  - e.g. preserve safety of journalists working in hostile regimes, resist Internet censorship
- …
BBC Horizon
Network Topology

Tor user

Entry node

Middle node

Exit node

Web server

Encrypted tunnel

Unencrypted TCP
Network Topology (censored network)
Network Topology (hidden services)
e.g. https://facebookcorewwwi.onion/
Layered encryption prevents linking of input and output flows based on traffic content.
Distribution of network information

- Tor nodes publish their information (descriptors) to the directory authorities.
- Directory authorities negotiate a consensus listing all available Tor nodes, and sign under keys of all participating directory authorities.
- Directory mirrors (most Tor nodes) connect to directory authorities to retrieve consensus.
- Tor clients connect to mirrors to receive up-to-date consensus, or bootstrap by connecting to directory authority directly.
- Tor clients verify digital signatures on consensus.
Number of relays

The Tor Project – https://metrics.torproject.org/

Heartbleed
The Tor Project

Total relay bandwidth

- Advertised bandwidth
- Bandwidth history

≈ 50% utilisation

The Tor Project – https://metrics.torproject.org/
Directly connecting users

The Tor Project – https://metrics.torproject.org/
Time in seconds to complete 5 MiB request

Measured times on all sources per day

- Median
- 1st to 3rd quartile

The Tor Project – https://metrics.torproject.org/
Challenge 1: Source routing

- Clients choose complete path based on information in consensus
- Cryptography enforces that network traffic must follow the path chosen
- If first and last Tor node selected is compromised then user can be de-anonymised
- If attacker has compromised proportion of Tor network bandwidth $p$, then proportion of paths de-anonymised $\approx p^2$ (e.g. 1% of bandwidth de-anonymises 0.01% of paths)
Peer-to-Peer anonymity networks

• Alternative to source-routed is peer-to-peer
• Clients connect to first node in the network, but do not have complete control of the path selected
• Advantage is that load balancing is easier and clients don’t need to know full network
• Disadvantage is route-capture: first malicious node will only route to other malicious nodes
• If attacker has compromised proportion of network bandwidth $p$, then proportion of paths de-anonymised $\approx p$ (e.g. 1% of bandwidth de-anonymises 1% of paths)
Managing consequences of source-routing

• Source routing brings advantages, but has costs
• Clients must know of all nodes
  • Consensus means that only one directory authority is contacted
  • Mini-descriptors reduce size of information downloaded and diff’s have been proposed
• Clients are provided enough information to load-balance correctly, through probing nodes and adjusting probability of selection
• Still a problem for network growth
Challenge 2: congestion control

- Tor network is perpetually congested
  - ~50% utilisation on average
  - IP networks normally 3–5%
- Tor nodes cannot drop packets when congested
  - Counter-mode encryption used
- Links between nodes use TCP, so take advantage of TCP congestion control
  - Down-side is head of line blocking
- End-to-end congestion control achieved through Tor implementing sliding windows
Potential changes in congestion control

- Replace TCP with UDP and bring congestion control explicitly into Tor
  - Also avoids problem of limited number of TCP sockets
  - Use latency-based congestion control (LEDBAT/µTP) to back off before packet loss
  - When multiplexing different circuits on one link, don’t block all circuits when there’s packet loss on only one
  - Also needs new crypto (DTLS?)
Tor over TLS/TCP (current)

Tor over µTP/DTLS/UDP (potential change)
Challenge 3: Clique topology

- Any Tor node can (almost) be at any position in a path selected by a client
- TCP links are kept up unless idle
- As a result any node must be able to connect to any other node, and many nodes will stay connected to almost every other node
- Problems include
  - Tor nodes in censored countries are not useful
  - IPv6 only nodes cannot be used
  - Not optimal for mixing traffic of different users
Alternative topologies: clique

- $O(n^2)$ links
- Node may be at different stage of different paths

Free Route
Alternative topologies: cascade

- O(n) links
- Easy to trace traffic
Alternative topologies: stratified

- $O(n^2)$ links, but less than clique
- Node always at same point in every path
- Optimal design, for some reasonable metrics
- Tor approximately does this already
What is needed next

• More **bandwidth** (both exit and middle)
• More **development effort** on Tor and surrounding projects, e.g.
  • Censorship resistance
  • Safe user experience (browser, chat, mail)
• Principles and tools for **scaling, enhancing performance and security** of source-routed overlay networks
• Techniques for safely **measuring networks** and rolling out **significant design changes**