Route Fingerprinting in Anonymous Communications

George Danezis & Richard Clayton

IEEE P2P2006, Cambridge, UK
6th September 2006
Tarzan “v1”

- Freedman, Sit, Cates, Morris: IPTPS Mar 2002
  - “me relay, you relay”: thousands of p2p participants
  - IP traffic sent hop to hop using encrypted “onions”
  - multiple servers should provide anonymity
  - lack of “network edge” avoids traffic analysis
  - “mimics” provide cover traffic
  - servers located by dipping into a Chord ring
  - route reconstruction from point of failure
Tarzan v2

• Freeman MSc thesis, May 2002
• Freedman & Morris, CCS, Nov 2002
• Same basic design, but notably:
  – servers located by asking peers for neighbour lists; which they learn with a gossip protocol
  – route reconstruction now starts at a random location within the tunnel
• The reason for these changes remain relevant to many other proposed designs!
#1: Node knowledge profiling

- Possible for an eavesdropper to determine which other nodes a Tarzan user learnt about
- So Tarzan learn about several hundred (out of tens of thousands or even millions) and used a few (typically far less than 10) at random
- Unfortunately, this means that their path through the network is likely to be unique!
Node knowledge attack

- Examine traffic at a node the user knows about
- Determine if any traffic \textit{at all} is arriving from another node known to the user and any traffic \textit{at all} is going to another node known to the user
  – if so then the user \textit{may be} using the node
- Unless user learnt about more than 10\% of the nodes in the whole network then \textit{very likely indeed} that the node is being used
  – full explanation and equations are in the paper
#2: Route reconstruction attack

- Tarzan designed to deal with very high rates of churn, so routes will often fail
  - cheap to rebuild around the point of failure because existing tunnel keys remain valid

- But attacker can overload “good” nodes to cause them to fail, until reconstructed route goes through a “bad” node they control
  - if attacker controls fraction $c$ of network and paths are of length $l$ then only $l/c$ attacks needed
Mitigations

• Tarzan v2
  – uses gossip style protocols to learn about nodes
  – rebuilds all of the path from a random position onward so “bad” nodes no longer accumulate

• Mixmaster, Tor etc assume full knowledge of all nodes and rebuild routes from scratch
  – both attacks are avoided
  – works well when nodes are reliable, churn is low and network size is not enormous
Who else is vulnerable?

- MorphMix, Wongoo, SAS let nodes along the path choose the next hop
  - completely side-steps our attacks
  - but this can of course lead to other problems!
- But other systems (AP3, Landsiedel et al, Xiao et al) assume originator will select the path…
  - insufficient details given in these papers to see if problem arises, but talk of selecting nodes “at random” gives some cause for concern
Conclusions

• In an anonymity system with large numbers of participants, peer discovery can be hard work; **but** relying on random selection from small subsets of peers has a significant problem.

• In anonymity systems with high churn, route reconstruction can be an expensive overhead; **but** cutting corners may lay you open to a route capturing attack.
Route Fingerprinting in Anonymous Communications

George Danezis & Richard Clayton

http://homes.esat.kuleuven.be/~gdanezis/
http://www.cl.cam.ac.uk/~rnc1/