Evaluation of the Anonymous I2P Network's Design Choices Against Performance and Security

Juan Pablo Timpanaro, Thibault Cholez, Isabelle Chrisment, Olivier Festor

ICISSP 2015 - Angers, France

10th February 2015

Outline

Introduction

Anonymous Communications

- Low-latency communications are quickly growing
- *Tor* has tripled its user-base in the last 18 months
- *I2P* has doubled its user-base in the last 10 months[a]

[a] Statistics from `http://metrics.torproject.org` and `http://stats.i2p.in/`

Increase use and latest events

- NSA *monitoring programs* put in perspective anonymous systems
- PRISM and MYSTIC programs, among others, bulk-collect non-encrypted Internet data
- Normal, *i.e* non-technical, Internet users are leaning towards anonymous systems
- Efforts, such as the *Tor Button* for Firefox, bring ever closer anonymous systems to Internet users

Introduction

Security Attacks

- Monitoring attacks, passive and active, have been carried out in these networks
- While anonymity is difficult to break, most successful attacks are DoS targeting the *metadata directory* that coordinate the network
- Tor uses a central approach, while the I2P uses a distributed one

Metadata Directories

- Tor trusts a central directory, composed by different *Tor directory servers*
- I2P uses a DHT-based directory to keep network metadata, which includes routing and applications data
- Network metadata allows peers to discover and interact among themselves

Motivations and Contributions

Motivations

- Shall these directories fail, anonymity is not compromised but the system became completely useless
- Even distributed directories can be attacked
- The I2P distributed directory, called the *netDB*, is prone to *Sybil attacks*
- Sybil attacks enables the control of network metadata and Eclipse attacks

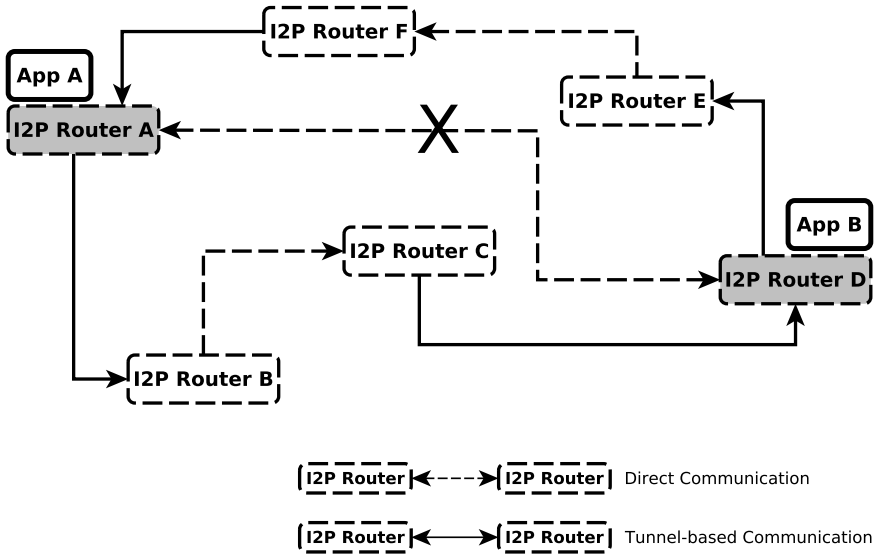
Contributions

- Evaluation of I2P distributed directory's security against another DHT design
- Study of I2P churn and proposal of new parameters

The I2P network

Principles

- Designed as an anonymous layer, mainly focused on anonymous web-browsing and file-sharing
- A closed network, where I2P users deploy *I2P routers* to create *multi-hop* paths, called tunnels, among themselves
- I2P users then deploy applications on top of their I2P routers to communicate with remote applications



The I2P network

Distributed Directory: the netDB

- I2P's network metadata includes *routerinfos* to identify I2P routers and *leasesets* to identify applications
- Routerinfos and Leasesets are stored in the netDB within dedicated nodes called *floodfill nodes*
- Floodfill nodes are I2P routers with high bandwidth and high uptime in the network. Only a reduced number of I2P routers are as well floodfill nodes

A Kademlia-based DHT

- The netDB is a Kademlia-based DHT: XOR-based efficient iterative routing, fully distributed network with untrusted and unverified nodes
- With no central verification authority, any I2P router can become a floodfill node, thus joining the netDB
- Specificity: DHT IDs must be recomputed every day:
  $routing\_id = SHA256(node\_id||yyyyMMdd)$

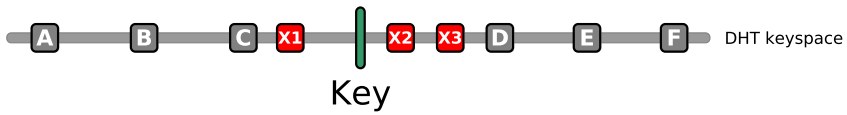
The I2P network security

NetDB design

- The netDB specifies a rather low by-design *replica set* of three nodes, in contrast with the original ten nodes design
- A low replica set tampers with the netDB reliability, specially if network churn is high
- Egger *et al.* attacks take advantage of this low replica set value to deploy easily few *attack nodes*, such as X1, X2 and X3

NetDB Attacks

- Egger *et al.* conducted arguably the sole attacks against the netDB, conducting a *localised Sybil attack*
- Different attack nodes X1, X2 and X3 are placed closer than any other legitimate node A, B or C to a target key



Improving the netDB

Principles

- Replica set was decreased to 3 peers for wrong reasons: to fight against (our) distributed monitoring of netDB
- While monitoring neither breaks anonimity nor reduces the QoS of I2P, reducing the replica set can harm the network
- A low replica set ease the deployment of attack nodes, reducing the *competition* among the replica set
- To increase competition, two ways: increase the number of nodes within the replica set and/or increase the overall number of participating (floodfill) nodes

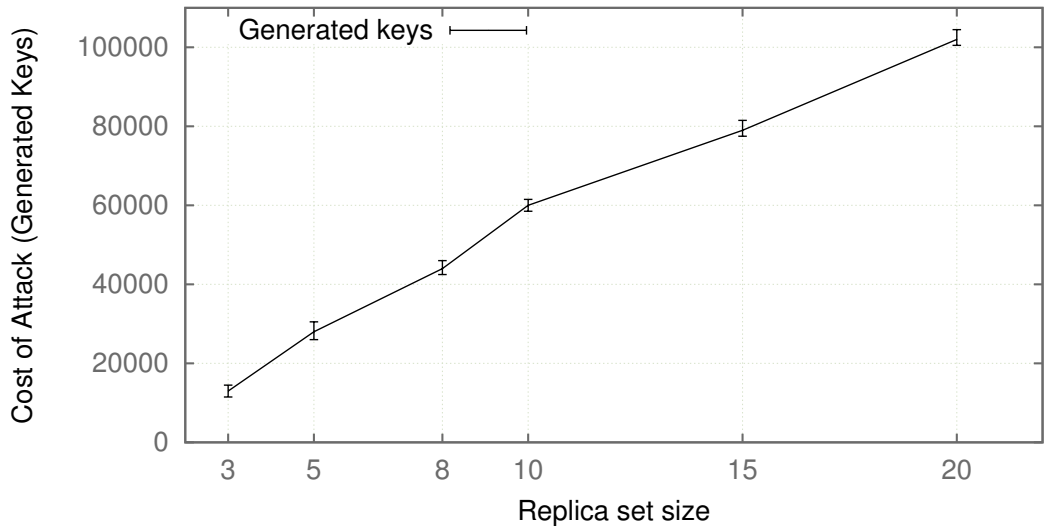
Objective

- Slightly harden the deployment of a localised Sybil attack while keeping backward compatible
- Benefit from increased replica set to reduce republication overhead

Increasing Replica Set Size

Principle

- An attacker will now need to generate only three attack nodes closer than any normal floodfill node, instead of five or ten nodes
- Considering netDB's current design, an attacker needs to generate K x $2^{\lfloor \log_2 N \rfloor + 1}$ fake nodes before finding those closer nodes, where K is the size of the replica set and N is the size of the netDB
- A lager replica set linearly increases the **cost of the attack**



Increasing netDB's size

Principle

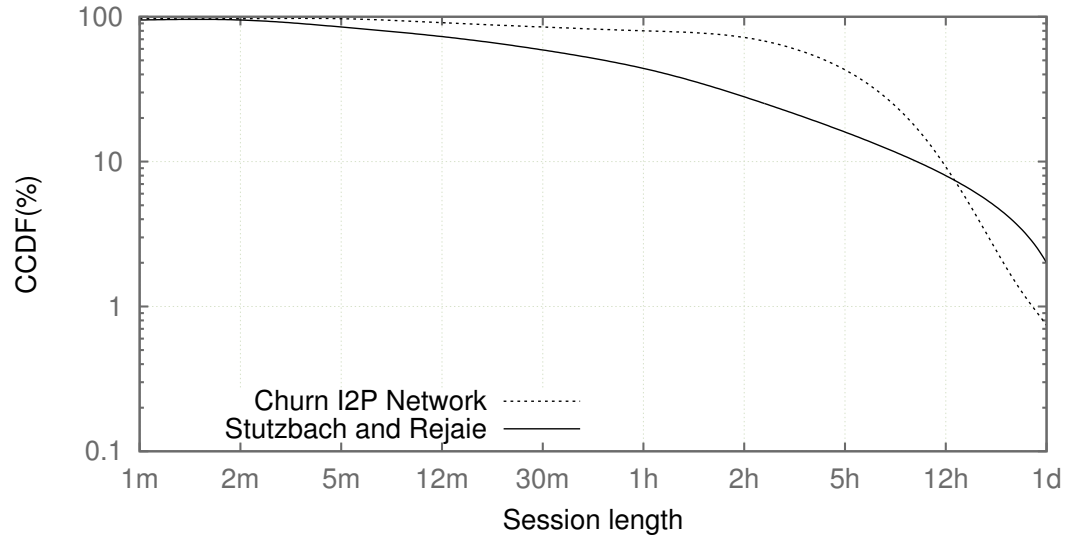
- Allowing every single node in the I2P network to become a floodfill node will enhance netDB's reliability
- Floodfill with a low bandwidth will not affect the netDB, since the bandwidth cost, measured in KB/s, is negligible in nowadays communications
- A replica set of three nodes and enabling every single user to become a floodfill node increases the cost of the attack by a factor of 50

Replica Set	NetDB Size			
	1X ($\sim 4K$)	2X	4X	All ($\sim 55K$)
Size= 3	**13K**	27K	52K	123K
Size= 5	23K	65K	110K	224K
Size= 10	60K	125K	250K	**730K**

Impact on overhead

Principle

- A small replica set implies a high frequency republication of data to fight churn
- We conducted a study on I2P churn for 5 days, measuring session length
- Currently, every 30 minutes: $p_{off}^3 = 0.15^3 = 0.0033$
- We propose: $p_{off}^{10} = 0.57^{10} = 0.0036$
- Result: republication overhead divided by 3 (10 msg per 5 hours vs 30)



Discussion

Proposed Approaches

- Increasing replica set size and netDB's size make increase computation cost of an attack
- Including every I2P user in the netDB and a replica set of 10 peers increases the cost of the attack by a factor of fifty
- However, a moderate attacker can still be able to deploy a localised Sybil attack

Further Approaches

- Douceur previously stated that the only bullet-proof solution against Sybil attacks is a centralised authority
- However, a distributed approach seems a better fit for a DHT-based netDB
- Alternatives solutions can be implemented, such as computational puzzles, which force an attacker to employ high computational resources so as to deal with many fake logical bounded to a single physical entity.
- Crypto-puzzle proof IDs would greatly benefit from increased replica set and DHT size.

Conclusion

Conclusion

- Evaluation of the netDB from a security and design point of view
- Proposal of backward compatible solutions to deal with localised Sybil attacks
- Our proposed mechanisms increases cost of current attacks by a factor of 50

Future Work

- An *iterative* and *flexible* approach to deal with a resourceful attacker
- Based on a flexible replica set value, an I2P user can auto-detect whether it is under attack
- Extending the replica set when an attack is detected will force the attacker to increase the attack nodes, which in turn increases the overall cost of the attack

Questions

Thank you !