Privacy methods.

Simulazione di Sistemi

September 15, 2015

Davide Berardi 0000712698
Matteo Martelli 0000702472

Università di Bologna.

Privacy

Privacy is the right to publish only some informations that we want to publish.
There are a lot of laws and legal issues related to privacy (but some people are just not intrested in laws).

Anonymity

We will talk about anonimy as the propriety of disconnect the user of a service from some basic proprieties:

- Geolocation.
- Association to a face or a name (or to an IP address).

Sometimes we need to reassociate the user with a communication channel or so.

But...who cares?

"I have nothing to hide, who cares about my personal data?"

Surveillance

- Some intelligence organizations obviously do anti-terrorist researches and checks (and that's good!).
- Some intelligence organizations do that in wrong ways, a normal guy searched "Pressure cooking" and "knap sack" on google and FBI knocked at his house.

Figure: the utah NSA data center "Massive Data Repository", 90k-140k m^2

NSA related projects

According to **Edward Snowden** leaks, the NSA organization has a lot of obscure and top secret projects:

- PRISM - Software to collect information about every internet communication (US)
  - Related to some major companies.
- Tempora - British Intelligence auditing software for internet and phone communications.
- MonsterMind - **Automated** reaction to attacks.

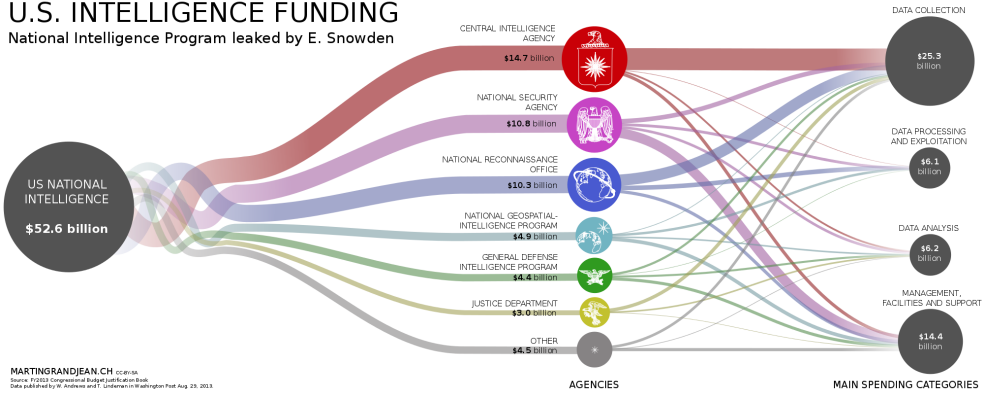
Other stories

- PGP - "munitions export without a license".
- Lavabit - Request to give the public key of the site to the NSA.
- IP-sec - Snowden leaks revealed that NSA breaked (in collaboration with NIST?) the IP-sec suite and his encryption algorithms.
- Truecrypt - US intelligence failed to decrypt some disks encrypted with truecrypt and fifth amendment protected a suspect.

Let's talk about numbers...

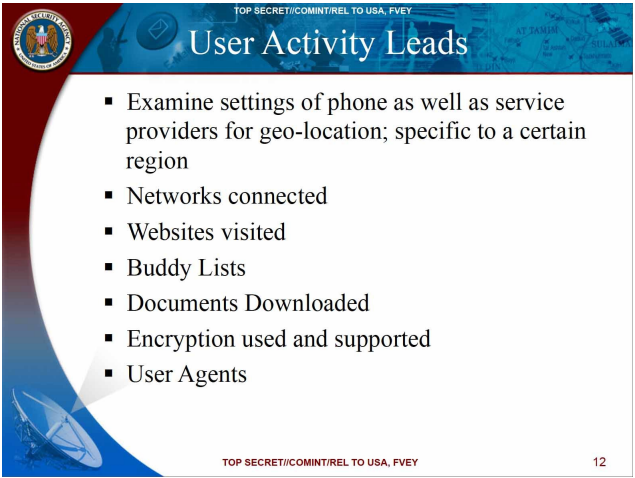
U.S. INTELLIGENCE FUNDING

National Intelligence Program leaked by E. Snowden



MARTINGRANDJEAN.CH CC BY SA
Source: FY2013 Congressional Budget Justification Book
Data published by W. Andrews and T. Lindeman in Washington Post Aug. 29, 2013

...and leaks!



GEO-Obscuration

Some places are sensitive to geo-obscuration (expecially east places like china or japan).

1	Gmail	google.com	gmail.com	Email	English	2014, September (or earlier) to present [2]	BLOCKED
1	Google	google.com	google.com	Search	English	2014, May (or earlier) to 2015, July [3]	BLOCKED
1	Google Maps	google.com	maps.google.com	Maps	English	2014, May (or earlier) to present [4]	BLOCKED
1	Google Docs	google.com	docs.google.com	Sharing	English	2011, May (or earlier) to present [5]	BLOCKED
1	Pornhub	pornhub.com	www.pornhub.com	Porn	English	2012, May to present [6]	BLOCKED
1	Google Encrypted	google.com	encrypted.google.com	Search	English	2011, March (or earlier) to present [7]	BLOCKED
1	Google APIs	google.com	*.googleapis.com	Search	English	2014, September (or earlier) to present [8]	BLOCKED
1	Google+	google.com	plus.google.com	Social	English	2011, July to present [9]	BLOCKED
1	Google Sites	google.com	sites.google.com	Web Hosting	English	2011, March (or earlier) to present [10]	BLOCKED
1	Picasa	google.com	picasaweb.google.com	Sharing	English	2009, July to present [11]	BLOCKED
2	Facebook	facebook.com	www.facebook.com	Social	English	2008, July to present	BLOCKED
3	YouTube	youtube.com	www.youtube.com	Sharing	English	2009, March to present [12] [13] [14] [15]	BLOCKED

SOPA and PIPA

- Stop Online Piracy Act
- Gives the control of the obscuration of a site to the proprietary of the copyright (!) and to the government (!!!).
- In other words...everyone could obscure every site, that use copyrighted contents, from every search engine!
- Legal penalities and fees to the source of the publication.
- Incompatible with DMCA, GPL, etc...
- **Incompatible with VPN, ORs, proxies, etc (!!!).**

Focused ADv

- Some companies can do some users profilation.
- What you've searched, what you say or what you do can be a gold information on who you are and what you're going to do.
- Maybe the company inform you, and you have nothing to hide, but, you really want to say to a company sensible data?
- Imagine if a data leakage occurs and someone learns about your google searches...
- Coff, Coff, Windows 10 keylogger...

Not only powerful adversaries

- Obviously the companies are not the only person interested in our real identity, someone could gain my IP, and/or my geolocation to break into my house when I'm out.
- So we must ensure our anonymity, just to have a form of security in addition to the classical ones.
- On the other hand, there is the identity stealing.

Technologies

What we can do versus controls?
Can we have some privacy even from the companies/government?

Confidentiality and authenticity

We have a lot of programs to protect our data

- PGP
- IPsec
- OTR-based programs
- Protonmail ProtonMail
- TrueCrypt
- **Perfect Forward Secrecy**

Some tool for steganography can help but not too much.

Anonymity

But for anonymity?

- Anonymous networks
- Mix Max networks.
- Anonymous remailers.
- Proxy chains.
- **Onion Routers**

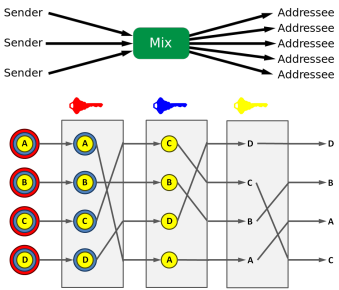
Anonymous network

(mostly) p2p-based networks, no one can identify who put a file on the net.

- FreeNet 
  - OpenNet mode
  - DarkNet mode
- GNUNet 
- Fully Self Contained (like UseNet).
- Search?
- Performance?

Mix Networks

- Model of the 1981.
- Multiple layers of encryption.
- Select different random nodes to deal with controlled nodes.
- **Timing attack?**



Old times

- Anonymous remailers: end to end anonymity.
  - Cypherpunk: remove FROM field and encrypt the mail
  - Mixmaster: Chain of remailers.
  - Mixminion: Mixmaster syntax with replies.
  - nym-server: give a pseudonym to the user detached from his IP.

We'll see that this servers recalls the modern idea of OnionRouting.

Onion Routing

The idea of encapsulate cyphered packets in a chain or an "onion".

- OpenNet? $\rightarrow$ Hidden services.
- New possibility like use a proxy to get to the normal internet.

Problems

- Performance
- DoS resistance.
- Mantain links to the users
- Thrustness of the routers.
- Confidentiality and autenticity.

Onion Routing (2)

- TOR
  - We'll come to this later.
- i2p
  - Done for eepsite(s).
  - Not so much routers/outproxies.
- And what for the low latency? $\rightarrow$ **Timing attacks**.

NSA and Tor

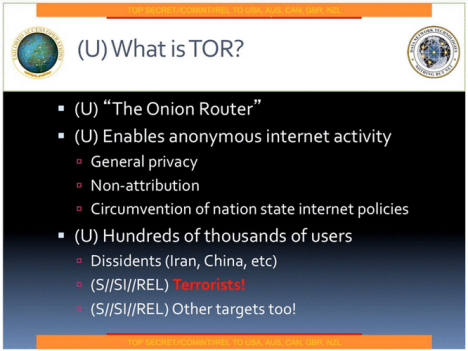
Tor was made from the naval research labs:

- Made for the anonymous control and espionage.
- Tor need a number of exit nodes (and routers) to lead anonymity to an user.
- if an organization use only his exit nodes it's like to not use them at all.

The Tor revolution

Apparently Tor slipped from the hands of the US in the 2004.

- Russia offered $114.000 to identify and deface Tor anonimity.
- NSA now classify TOR as a menace of level *catastrophic*.

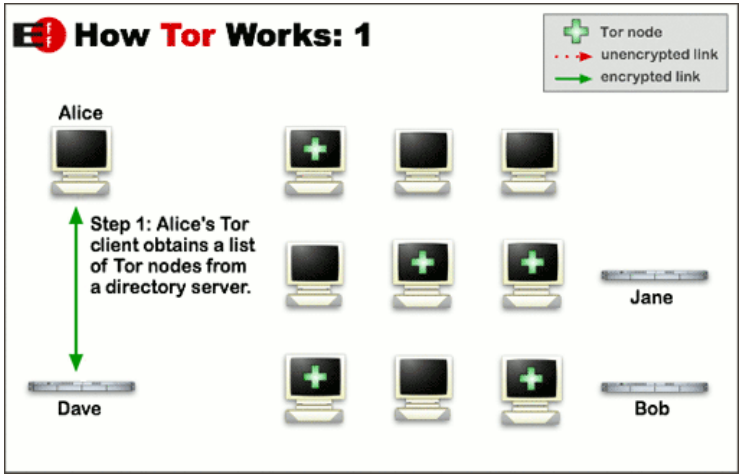


But...nobody knows who and what is hidden under the layer of the onion.

Part 2 – Onion routers and attacks

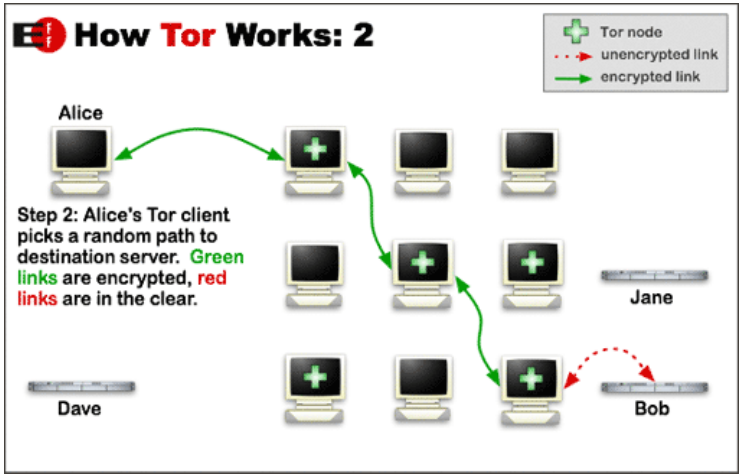
Tor

Tor workings



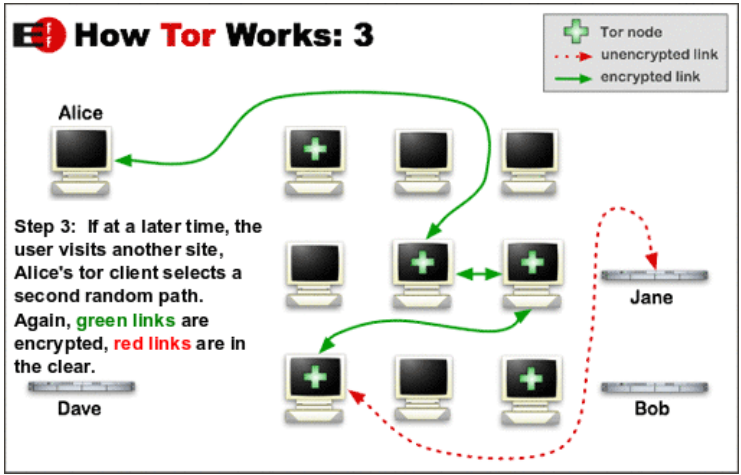
Tor

Tor workings



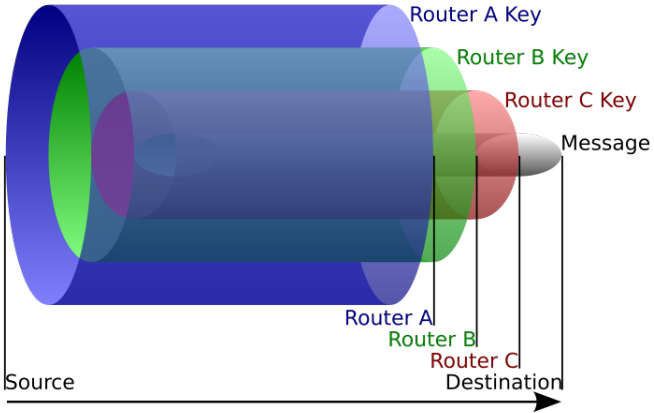
Tor

Tor workings



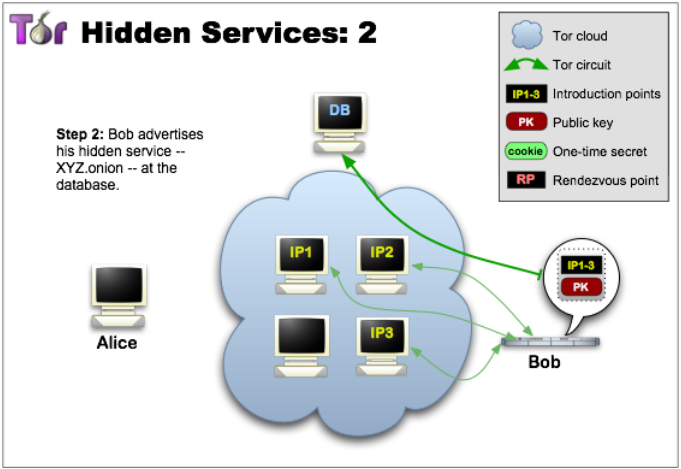
Tor

Tor encryption



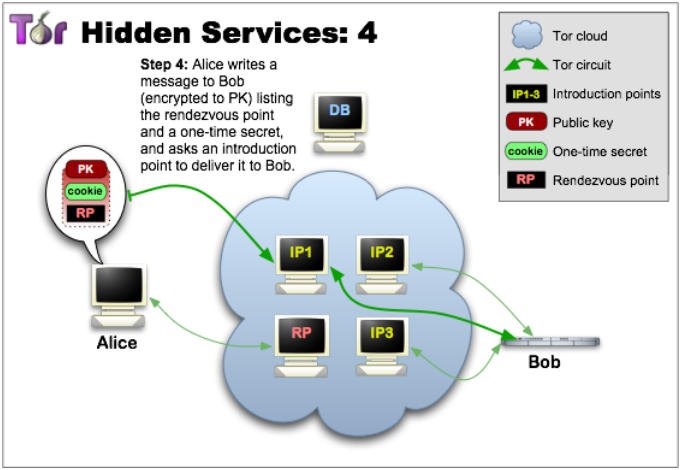
Tor

Tor Hidden Service



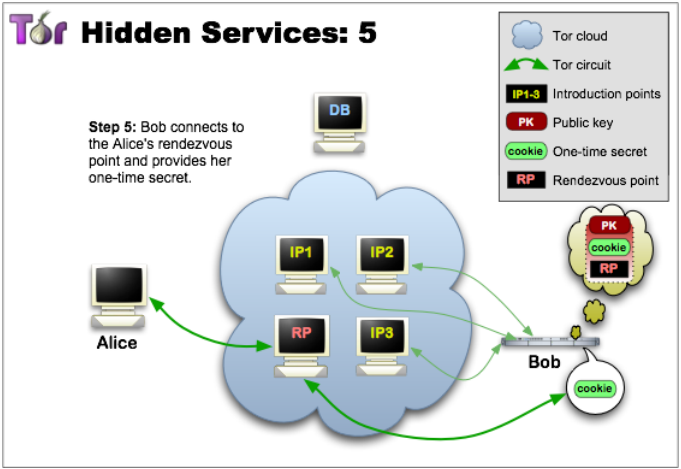
Tor

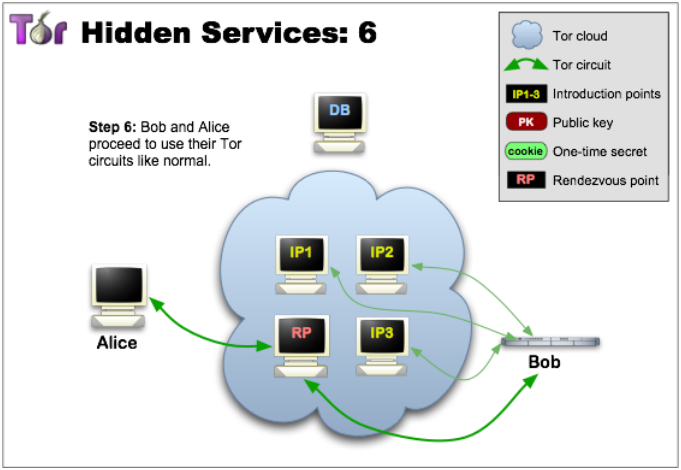
Tor Hidden Service



Tor

Tor Hidden Service





Tor
Summary

- Base: anonimity of clients
- Hidden services: anonimity of client + anonimity of servers

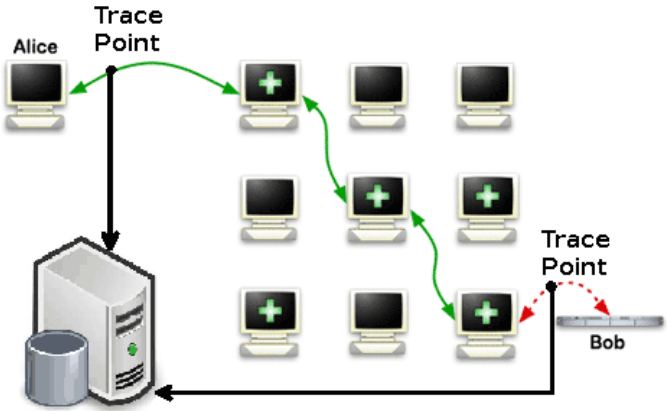
But ... is it enough?"

Time analysis based attacks

> *"Tor does not provide protection against end-to-end timing attacks[...]"*

We can place a tracker after the client node and another before the server node and check for the connection time to profile users and nodes (and later associate IP to users.)

Timing analisys based attacks

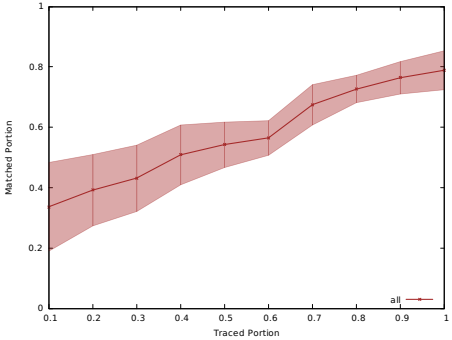


Why simulation?

Simulation help us in a lot of aspects:

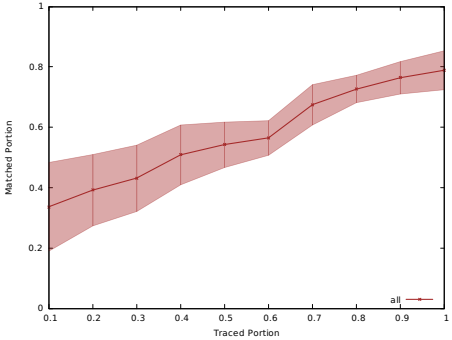
- Compare the performances of two onion routers (p.e. i2p vs Tor).
- To compare effects of changes in the node choice algorithms.
- **Get an idea about the timing attack feasibility and estimate the number of resources needed by an attacker**.

Simulation results



Tracing **half** of the Tor network, an attacker may be able to understand the relationships between the nodes of a **quarter** of the Tor network.

Simulation results



- Timing attack is **feasible**.
- It requires a **LOT** of resources.

Too much?

Who can face such big amount of resources?

NSA timing attack on clients

NSA slides leak

TOP SECRET//COMINT// **REL FVEY**

Technical Analysis: Timing Pattern

(TS//SI)

Send packets back to the client that are detectable by passive accesses to find client IPs for Tor users.

- Current: GCHQ has research paper and demonstrated capability in the lab.
- Goal: Can we expand to other owned nodes?

NSA timing attack on servers

NSA slides leak

TOP SECRET//COMINT// **REL FVEY**

Exploitation: Web Server Enabling

(TS//SI)

Given CNE access to web server modify the server to enable a "timing/counting" attack similar to timing pattern idea.

- Current: GCHQ has a research paper and demonstrated the technique in the lab.

Measuring AS-level adversaries against Tor

- Multiple Autonomous Systems (AS) collude with each other performing time based attacks and collecting asymmetric data.
- Up to 40% of circuits constructed by the current Tor client are vulnerable to AS-level attackers.
- Connections from China were found to be most vulnerable to AS-level attackers with up to 86% of all Tor circuits.

Connections between ASes are negotiated as business arrangements.

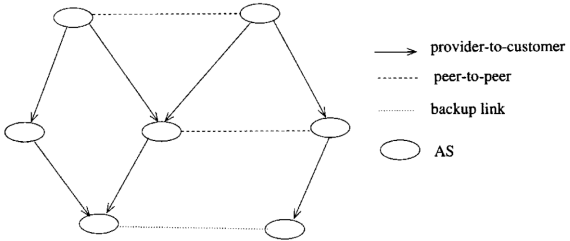


Fig. 3. Hierarchical AS interconnection.

The idea

Building a graph of ASes correlations to identify vulnerable paths.

Astoria

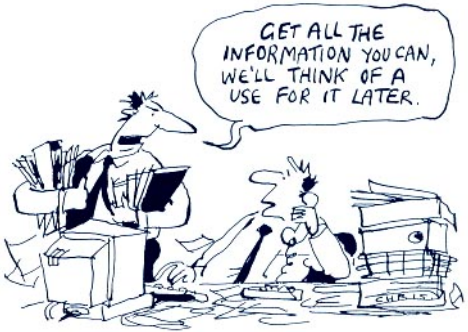
- Use of path prediction to avoid Tor vulnerable paths.
  i.e. The entry node and the exit node may be selected together if their ASes are unrelated to each others.
- Able to perform load-balancing at least as well as the vanilla Tor client.

Open Problems

User awareness on privacy and anonimity.

"I deplore the lack of internet privacy and so do my 5,000 Facebook friends!"

Open Problems



Need of *real* privacy laws and investments on network security solutions. (IPSec, anti surveillance architectures, etc.)

Thank you for your attention.

Diffie-Hellman

- Alice pick a random number a, a prime number p and α as a primitive root of p.
- Alice calculate $k_a = \alpha^a mod p$ and sends $< k_a, p, \alpha >$ over the channel.
- Bob read the packet, pick a random number b, calculate $k_b = \alpha^b mod p$ and sends it to Alice.

Now the shared key $K = k_b^a = k_a^b = \alpha^{ab} mod p$ is known to Alice and Bob[1].

[1] For the little Fermat theorem ($a^p \equiv a \; mod p$) if p is a prime

Perfect Forward Secrecy

- If a key is derived from another with a deterministic method then a leak of the second key can reveal every eavesdropped transmission encrypthed with the first key.
- The immunity to this kind of attacks is called *Perfect Forward Security*.
- Used in Diffie Hellman based TLS, OTR, etc.

Freenet

- A file is encrypthed with his hash and shared over the network.
- It can optionally encrypthed with the public keys of the dark net users (pseudonym-like) (and signed).
- The file is so splitted in chunks and shared over the network.
- Every node can't understand which chunk of which file is processing and what is written in the file itself.

Freenet (2)

- The net is garanted to resemble a small world network (so the max degree is $log(n)$) using the Metropolis Hasting algorithm.
- If a file is not popular and only a chunk is lost, the file is lost forever.