User identification with username and password in structured P2P networks

Rodrigo Fernández
rfernand@inf.utfsm.cl

Universidad Técnica Federico Santa María

12 de diciembre de 2013

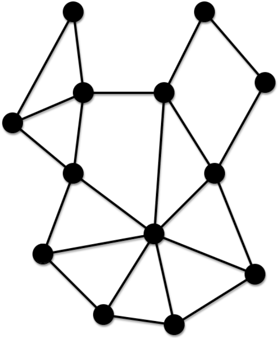
P2P Networks

They can be used for

- Storage and file sharing
- Email
- Money (heard about bitcoin?)
- and a lot more!

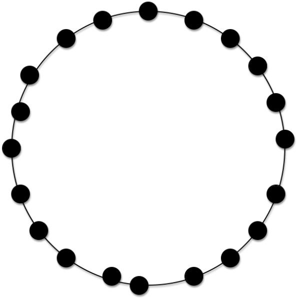
P2P Networks Properties

- Scalable
- Decentralized
- Self-maintained
- Robust



P2P Networks Overlay structure

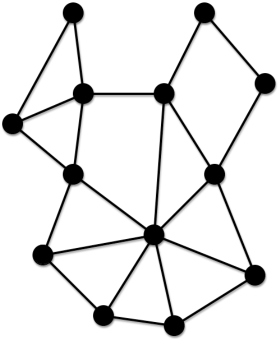
- Structured networks (CAN, CHORD, Pastry)
- Unstructured networks (Gnutella, Bittorrent)



P2P Networks

Unstructured networks

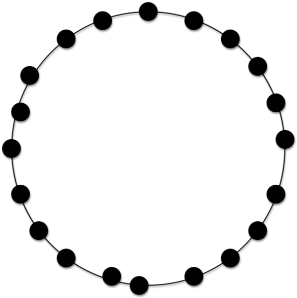
- Usually organized as a hierarchical/plain graph.
- Search is costly and can result in false negatives.
- Can achieve certain grade of node anonymity.



P2P Networks

Structured networks

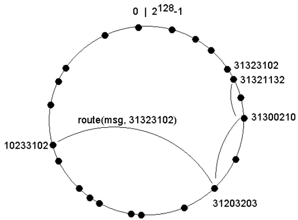
- Strong topology structure, usually ring or mesh based.
- More efficient search, without false negatives.



P2P Networks

Structured network example: Pastry

- Routing algorithm based on SHA-1
- Each node maintains a routing table, leafset and a neighbor set.
- Data search reaches O(log(N)) nodes



Complex systems needs a way to identify an user...

In short:

- Many **complex systems** require authenticated users to work.
- Most of the users has **more than one device**, so identification through multiple devices is needed.
- **User are accustomed** to username/password solutions.

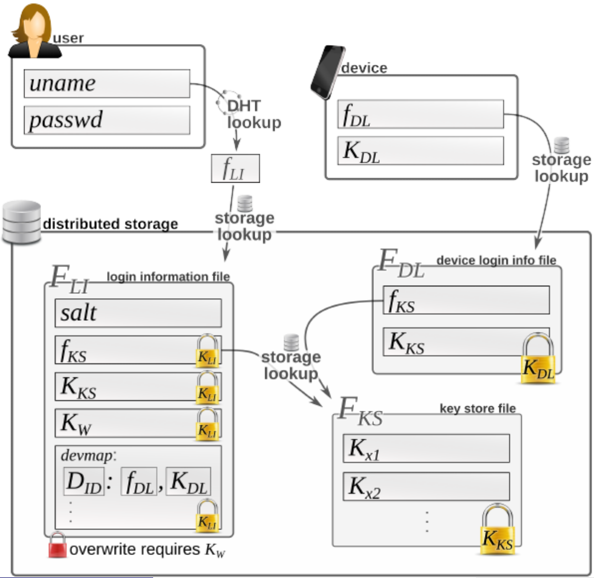
P2P Identification Schemes

Options?

Decentralized schemes distributes the task of public key auth to all participants.

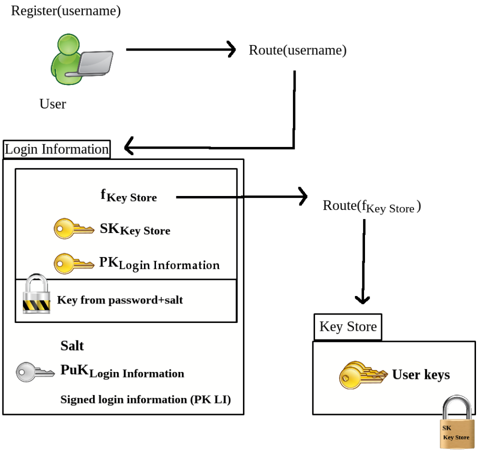
- PGP-like scheme
  Creates web of trust to auth public keys based on their acquaintances opinions.
- Quorum-based scheme
  Multiple independent participants replicate public keys.
- Username/Password scheme

Peerson proposal



Peerson proposal

(simplified version)



Peerson proposal

Protocols

1. User registration
2. User sign in
3. Logout
4. Password change
5. Password recovery

Peerson proposal problem...

It only works in a “perfect” world

1. In reality there are byzantine nodes
2. How to ensure that a node is telling the truth?
3. Resilience against Sybil attacks?

Trust in P2P networks

1. To deliver a valuable service in P2P applications, it is important to trust that the participants will act as requested.
2. We need to be sure that other peers will forward messages, and that the designated peers will indeed save the information correctly so that operations can be successful.

Trust in P2P networks

Building Trust

1. Complex since a P2P network includes untrusted nodes from an open environment.
2. Untrusted nodes may be faulty, malicious, and act together to attack the network.

Trust in P2P networks

Byzantine nodes

Are all nodes that not behave as expected

1. Faulty nodes
2. Malicious nodes
3. Infected nodes

A peer is honest only if the peers executes the protocol faithfully; otherwise, the peer is faulty.

Trust in P2P networks

How to detect byzantine nodes?

1. Reputation systems: Assess the past history of a peer by gathering feedback from nodes with previous interactions with this peer.
2. Accountability: Detects and exposes faulty nodes by creating non-repudiable records of every node's actions.

Reputation systems

CORPS Trust model

1. Every node X has an associated reputation value $R(X)$ which represents the probability that X is an honest node.
2. $R(X)$ is computed using the recommendations emitted by nodes that have completed a transaction with X. Bad recommendations have a stronger effect on $R(X)$ than good ones. It should be more difficult for node to increase its reputation value than to decrease it.
3. For every node X, $R(X)$ is highly available in the DHT.

Byzantine node tolerance

P2P networks can achieve byzantine fault tolerance under 1/3 byzantine peers

Group agreement: The goal is to obtain at least $\frac{L}{2} + 1$ identical answers from a group of L nodes.

Byzantine Generals Problem

Thesis proposal

- Propose and evaluate a new user identification system that can work in the presence of byzantine nodes in real life scenarios.

Working Hypothesis

- It is assumed that the use of a reputation system and trusted nodes management mitigates the effectivity of malicious nodes on identity usurpation attacks.
- It is assumed that encryption schemes available today are sufficient to secure the user's private data in the P2P networks.

Goals

Main Goals

The implementation of a secure *username/password* based user identification scheme in structured P2P networks using:

- Secure routing
- Node trust management
- Encryption schemes

Goals

Specifics Goals

- Study the possibility of password recovery mechanisms.
- Study and use trust management to maintain a secure layer inside the P2P network.
- Use byzantine tolerant algorithms to verify and maintain the system consistency in the presence of malicious nodes.
- Study and use secure routing, search and storage mechanisms in structured P2P networks.

Results

Contributions and Expected Results

- Design of a secure and modern user identification system for structured P2P networks.
- Generate a base system to develop complex projects in P2P networks.
- Reassure that P2P distributed systems have the capabilities to offer complex and high level services.
- Paper publication.

Results

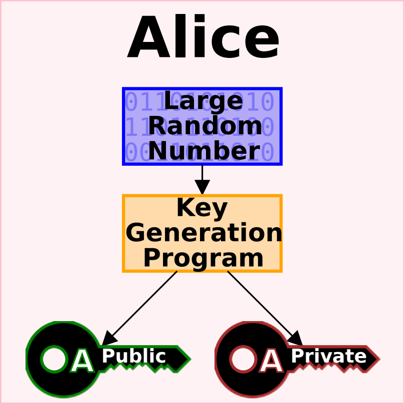
Validation procedures

- Proving that the system will have a minimal probability of error.
- Proving that the system will maintain his consistency in networks with at most 30 % of byzantine nodes.

Questions?

Keys randomly derived

Basic Public-key cryptography



Keys derived from a password

Basic example

$$\mathrm{DK} = \mathrm{KDF}(\mathrm{Key}, \mathrm{Salt}, \mathrm{Iterations})$$

Some functions for this are:

- PBKDF2
- scrypt
- bcrypt

Probability of error

If p represents the probability that a single node is malicious, and N the total number of nodes in the DHT.
With $p = 0{,}3$, for $\frac{L}{2}$ consecutive nodes to be malicious in a leafset.
The probability is given by:

$$P = p^{\frac{L}{2}} \tag{1}$$

Trust in P2P networks

Probability of error

$$P = p^{\frac{L}{2}} \tag{2}$$

	Probability to fail	
Size of Trusted Set (L)	**p = 0,3**	**p = 0,05**
8	0,0081	$6{,}25 \times 10^{-6}$
16	$6{,}56 \times 10^{-5}$	$3{,}9 \times 10^{-11}$
32	$4{,}3 \times 10^{-9}$	$1{,}52 \times 10^{-21}$

Cuadro : Probability of failure in a transaction that needs $L/2+1$ identical answers

Methodology and Working plan

Stage I: Problem definition

1. Study of P2P network systems and P2P search and storage mechanisms. (August 2012)
2. Study the P2P networks capabilities to implement complex systems as seen in centralized systems. (September 2012 - November 2012)
3. Born of the idea. (December 2013)
4. Problem specification, hypothesis and project objectives. (February 2013)

Stage II: P2P Systems definition

1. State of the art of P2P network systems and P2P search and storage mechanisms. (May 2013)
2. State of the art of building of trust between nodes in P2P networks. (April 2013 - June 2013)
3. State of the art of user identification schemes in P2P networks and how to secure the different system protocols. (July 2013)

Stage III: Solution proposal

1. User identification system design for structured P2P networks. (August 2013 - September 2013)
2. Theoretic evaluation of the user identification proposal. (October 2013)
3. Final thesis report development. (November 2013 - January 2013 2013)
4. Paper development. (February 2014)