PROYECTO DE TESIS

☐ Doctorado en Ingeniería Informática

☒ Magíster en Ciencias de la Ingeniería Informática

1. **Título del Proyecto de Tesis**	*"User identification with username and password in structured P2P networks"*
2. **Nombre del Alumno**	Rodrigo Germán Fernández Gaete
3. **Número de Teléfono - Celular**	+56 9 84193413
4. **Correo electrónico**	`rfernand@csrg.inf.utfsm.cl`
5. **Fecha de Ingreso al Programa**	Primer semestre de 2013
6. **Pregrado**	Ingeniería Civil Informática, Universidad Técnica Federico Santa María, 2012
7. **Profesor Guía de Tesis**	Xavier Bonnaire
8. **Fecha Presentación Tema de Tesis**	
9. **Fecha Aprobación Tema de Tesis**	
10. **Fecha Tentativa de Término**	
11. **Comisión Interna de graduación**	Comite académico programa MII

1 Resumen

Los servicios P2P son robustos, escalables y auto-organizados por naturaleza, pero su arquitectura diferente trae nuevos problemas y requerimientos. Tradicionalmente, las redes P2P identifican sólo los nodos que componen el sistema, sin diferenciar a los usuarios detrás de cada uno de ellos. Hoy en día, las personas usan más de un dispositivo para conectarse a la red. Este cambio de comportamiento hace que una identificación por usuario sea necesaria. Además, el usuario normal está acostumbrado al uso de nombres de usuarios y contraseñas para identificarse en éstos sistemas. Mientras que existen propuestas para la implementación de un sistema de identificación por nombre de usuario y contraseña, éstos no contemplan la existencia de nodos maliciosos, fallando su implementación en escenarios vistos en redes P2P reales.

El trabajo a continuación investiga profundamente los requerimientos y características necesarias para un sistema de identificación seguro, junto con los desafíos encontrados para su implementación en redes P2P.

El objetivo principal de este trabajo es desarrollar un sistema de identificación de usuarios basado en nombre de usuario y contraseña en redes estructuradas P2P, usando sistemas de reputación y administración de nodos confiables para protegerse en contra de nodos maliciosos y implementar protocolos seguros en el sistema.

Se realizará un análisis teórico del sistema para asegurar que los protocolos son seguros y no presentan riesgos en ambientes reales.

La arquitectura para desarrollar el sistema será en redes P2P basadas en Distributed Hash Tables (DHT).

Keywords: P2P, identificación de usuario, sistemas distribuidos.

Abstract

P2P services are robust, scalable and self-organized by nature, but have a complete different architecture with new problems and unique requirements. Traditionally, P2P networks used to identify the different nodes that compose the system, but not the user behind each one of them as a different being. Today, the people use multiple devices to join the network. This change of behavior makes the per node identification obsolete when trying to identify the unique user behind them. Also, users are used to username and password identification schemes. While an early approach of username/password identification system exists, it does not contemplate the existence of malicious nodes, lacking the capability to be secure when working in real P2P networks.

We thoroughly investigate the requirements and features of a secure identification scheme, along with the challenges facing a P2P implementation.

The main goal of this work is to develop a secure user identification system based on username/password keywords in structured P2P systems, using reputation systems and trusted nodes management to defend from malicious nodes and implement secure protocols for the system.

Theoretic analysis of the system will be done to ensure that the protocols are secure and do not present security risks in real environments.

The architecture to develop the system will be based on Distributed Hash Tables (DHT) structured P2P networks.

Keywords: P2P, user identification, distributed systems.

2 General formulation of the problem and thesis proposal

2.1 Context and Motivation

A Peer-to-Peer network (from now onward called *P2P*) is a distributed system of big scale. His participants are called *nodes* and they directly share resources and data, acting like clients and servers. They are called big scale because they are made to contain millions of nodes through the Internet. P2P systems are scalable, decentralized, robust and self-organized.

P2P systems are characterized by do not having a central coordination. Each peer is independent and has a local view of the system. The global behavior emerges from the local interaction of its members [1]. P2P networks natural properties make non-profitable services capable of running with the help of the same users that use it.

The most basic P2P systems only provide the structure to anyone in the network to store and retrieve data in it. To provide more complex functionalities, additional logic has to be implemented to them. As P2P systems grow in functionalities, the need to identify users inside the network arose.

The subject of securely establishing stable identities in P2P systems has been previously studied, for instance by Aberer, Datta and Hauswirth [2]. The need for identities mainly arose from technical concerns, such as handling dynamic IP address assignment, or avoiding Sybil attacks [3]. Authentication of a node is done via a signature key, automatically generated and stored on the node. Traditionally, P2P networks identified the different nodes that compose the system, but not the user behind each one of them as a different being.

The increase of number of devices the people have and use to join the network today makes the per node identification obsolete when trying to identify the unique user behind them. An example of when this would be needed, is the backups systems. They need to store important data in the network and then restore it on a different system from where it was backed up.

To store the data safely all approaches build on encrypting backed up content. The schemes to identify the user are basically two: the ones that use keys randomly derived [4] and the schemes that derive keys from a user defined password [5]. While approaches that derive randomly the keys does not have the risk of someone guessing the user keys, they require that the user manually back up the keys. P2P storage systems that implements the use of keyword strings to derive a public-private key pair whose private key is used to sign data and the hash of the public key to identify the data in the storage. Both of

these systems use a keyword string as a seed to a pseudo- random number generator that produces the key pair [6], [7]. Knowing only the memorable keyword string the user can store and retrieve information.

While the identification system can function without the need of forgotten passwords functionalities, they can also be implemented in the system. Related to that, recovery of information in a P2P scenario has been studied by Vu et al. [8] who proposed a combination of threshold-based secret sharing with delegate selection and encrypting shares with passwords. Frykholm and Juels [9] proposed a password-recovery mechanism based on security questions very similar to the one seen in [10].

At last, an initial proposal for passwords in P2P networks has been presented, designed to handle remembered logins, and recovering lost passwords in an idealized network [10], and has not been proved to work in a real environment.

2.2 Problem statement

Most of existing systems for the user identification in P2P networks only consider the use of pre-shared keys to identify the user in the network. While that can be easily implemented, does not provide to the users the flexibility that a username-password based identification provides when using different devices to log in in the system. As the user needs to transfer manually his keys from one device to another, there are many security issues when they are handled without care or the devices (like a cellphone) are lost.

The use of a username and a password means that the user keys needs to be secured inside the identification system. To handle the user keys without compromising the users identity, additional security layers needs to be placed inside the P2P network.

While a solution based in username and password has been proposed before [10], it does not take in consideration the presence of malicious nodes. A malicious or bizantine node is any node that does not behave as expected by the protocol of the system. The presence of this type of nodes can easily break the security and functionalities of the whole system.

2.3 Thesis proposal

Before developing an P2P system with the desired functionalities, an adequate system architecture is needed. We will thoroughly evaluate a new user identification system that can work in the presence of bizantine nodes, with the hopes to reach the desirable functionalities with a minimum probability of failure, to ensure the security of the system in real life scenarios.

2.3.1 System architecture

For the design of the system's algorithms our work will be based on structured networks based on Distributed Hash Tables (DHT), which provide efficient key lookups, high data availability and persistence.

To mitigate the problem of malicious nodes, we can use reputation systems to build trust among the nodes. The key idea of a reputation system is to predict the future behaviour of nodes based on feedback about their past transactions [11]. A transaction is application dependent, for example forwarding a message in the network, buying an item in e-commerce services, share or store files, etc. After a transaction, the client node emits a recommendation that evaluates the behaviour of the other peer. The aggregation of these recommendations leads to a reputation value. A reputation system built on top of a DHT has the ability to compute a global reputation value for every node. Indeed all the recommendations about a single node can be handled consistently at a common location: either by a specific node or by a set of nodes.

To build a group of trusted nodes, the CORPS [12] algorithm presents an efficient solution to to build a scalable trusted ring within a DHT that allows to find reputable peers.

To secure the stored keys, the proposed system uses encryption, indirection and rings of trust inside the network. The system goal is to offer a secure mean to identify an user using only his username/password knowledge taking in consideration the presence of bizantine nodes.

2.3.2 Protocols

For the user identification system, the following protocols will be used as a base for his development, as further security mechanisms will be needed to be added to secure each one from malicious nodes.

Account registration To register a new user account, the user first has to choose a *username* and a *password*.

Considering a key-based authentication, the user creates a *key store file*, containing all the keys used by the P2P application the user wants to log in to. The user generates a cryptographic key to authenticate the write operations that will be made in the file, and store this key along with the others in the *key store file*.

The user then creates a *symmetric key* K_{KS} , encrypts the file content with this key and puts the ciphertext into the storage, obtaining a *file name* f_{KS} . Now, the user creates a *login information file* by creating a random byte string *salt*, deriving a *symmetric key* K_{LI}

from the user *password* and the *salt*. Using the new *symmetric key* K_{LI}, the user encrypts the *file name* f_{KS}, the *symmetric key* K_{KS} and the *cryptographic key to authenticate the write operations* K_W. The salt and the three encrypted values are put into the storage, obtaining a file name f_{LI} . The salt is stored in plaintext, so that the user later can derive the decryption key K_{LI} by only providing the password. Finally, the user performs the write-once operation put on the DHT with username as key and f_{LI} as value.

If the username was taken, the user is prompted for a new username.

Once all operations have succeeded, the user is registered in the system.

Sign-in The user uses his username to find and retrieve his *login information file*. Then, using his *password* and the *salt* included in the *login information file*, obtains the *file name* f_{KS} used to route back to where the *key store file* is stored. Lastly, uses the *symmetric key* K_{KS} to decrypt the *key store file* and recover his user keys.

Logout The system does not have something like a "session" to maintain; the only way to identify an user is by his keys that are obtained by the identification process.

Password Change To change the password, the user has to rewrite his *login information file*.

Before the user can change the password, he must log in using his password to obtain K_{LI} . With this information, the password change can be accomplished: the user is asked for a new password and a new salt is generated. The key-derivation function is used to generate a new key K_{LInew} for the login information file. Then, the content of the key-store file is fetched and decrypted (with the old key). A new key K_{KSnew} is generated and used for encrypting the key-store content again before it is saved to the storage system, obtaining a new filename f_{KSnew}. Finally, the login information file is updated: f_{KSnew}, K_{KSnew}, the write credential K_W is encrypted with the new key K_{LInew}. Together with the new salt, this ciphertext is written to the distributed storage, using the reference f_{LI} and the credential K_W, to authenticate the write operation. Lastly, the keys stored in the key store should be updated by the application using the P2P protocol.

References

[1] K. Aberer and M. Hauswirth, "Peer-to-peer information systems: concepts and models, state-of-the-art, and future systems," *SIGSOFT Softw. Eng. Notes*, vol. 26, pp. 326–327, September 2001.

[2] K. Aberer, A. Datta, and M. Hauswirth, "Efficient, self-contained handling of identity in peer-to-peer systems," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 16, no. 7, pp. 858–869, 2004.

[3] J. Douceur, "The sybil attack," in *Peer-to-Peer Systems* (P. Druschel, F. Kaashoek, and A. Rowstron, eds.), vol. 2429 of *Lecture Notes in Computer Science*, pp. 251–260, Springer Berlin Heidelberg, 2002.

[4] M. Lillibridge, S. Elnikety, A. Birrell, M. Burrows, and M. Isard, "A cooperative internet backup scheme," in *Proceedings of the annual conference on USENIX Annual Technical Conference*, ATEC '03, (Berkeley, CA, USA), pp. 3–3, USENIX Association, 2003.

[5] L. P. Cox, C. D. Murray, and B. D. Noble, "Pastiche: making backup cheap and easy," *SIGOPS Oper. Syst. Rev.*, vol. 36, pp. 285–298, Dec. 2002.

[6] I. Clarke, O. Sandberg, M. Toseland, and V. Verendel, "Private communication through a network of trusted connections: The dark freenet," *Network*, 2010.

[7] K. Bennett, C. Grothoff, T. Horozov, and J. T. Lindgren, "An encoding for censorship-resistant sharing," 2003.

[8] L.-H. Vu, K. Aberer, S. Buchegger, and A. Datta, "Enabling secure secret sharing in distributed online social networks," in *Computer Security Applications Conference, 2009. ACSAC '09. Annual*, pp. 419–428, 2009.

[9] N. Frykholm and A. Juels, "Error-tolerant password recovery," in *Proceedings of the 8th ACM conference on Computer and Communications Security*, CCS '01, (New York, NY, USA), pp. 1–9, ACM, 2001.

[10] G. Kreitz, O. Bodriagov, B. Greschbach, G. Rodríguez-Cano, and S. Buchegger, "Passwords in peer-to-peer," in *Peer-to-Peer Computing (P2P), 2012 IEEE 12th International Conference on*, pp. 167–178, IEEE, 2012.

[11] P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman, "Reputation systems," *Commun. ACM*, vol. 43, pp. 45–48, Dec. 2000.

[12] E. Rosas, O. Marin, and X. Bonnaire, "Corps: Building a community of reputable peers in distributed hash tables," *The Computer Journal*, vol. 54, no. 10, pp. 1721–1735, 2011.

3 Working Hypothesis

The next work is based in the following hypothesis, under the assumption that P2P networks are capable of maintain highly secure services.

- It is assumed that the use of a reputation system and trusted nodes management mitigates the efectivity 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.

4 Goals

4.1 Main Goals

The implementation of a secure username/password based user identification scheme in structured P2P networks using secure routing, building of trust between nodes and encryption techniques.

In particular, the main goals are:

1. Have a minimal possibility of error in the identification process.
2. Use a layer developed scheme that can easily adapt to most commonly used P2P networks.

All this is fundamental to the identification protocol to maintain his desirable properties.

4.2 Specifics Goals

In the development of the present work the following specifics goals are taken in consideration:

- Study the possibility of password recovery mechanisms in the proposed identification scheme.
- Study and use trust management to maintain a secure layer inside the P2P network.
- Use bizantine 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.

Methodology and Working plan

The working plan for the thesis development consists in three stages.

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)*

Results

4.3 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.
- Development of an article to be sent to a distributed systems publisher. The paper will show the results of the user identification system designed in this project.

4.4 Validation procedures

P2P networks presents a big difficulty to be tested in a real environment because of the high number of nodes needed to try it out. Therefore, instead of going after an empiric validation of the proposed identification system, only a theoretical evaluation will be presented. The proposed system will be compared with the other systems available at the moment with a thoroughly analysis of the security of the algorithm used.

Taking that in consideration, the theoretical evaluation will be focused in:

- Proving that the system will have a minimal probability of error, taking in consideration that there are two different types: false positives and false negatives. The first is when the system identifies an user as "foo" and he is not, and the last is when the system says that the user is not "bar" when he really is it.
- Proving that the system will maintain his consistency in networks with at most 30% of bizantine nodes.

5 Resources

5.1 Available Resources

- The books and publications related to distributed systems and P2P networks provided by the *UTFSM Library* will be used.

5.2 Resources Required

No more extra resources are needed.