i3S sophia antipolis · Université Nice SOPHIA ANTIPOLIS · MASCOTTE · CNRS · INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE · INRIA · centre de recherche SOPHIA ANTIPOLIS - MÉDITERRANÉE

Frédéric Giroire
Tel: (+33) (0)4 92 38 50 98
Fax: (+33) (0)4 92 38 79 71
Email: frederic.giroire@sophia.inria.fr

Objet : Subject of the first year internship of the ENS diploma

Title:	Algorithmics of P2P storage systems: Modelisation, Simulation, and Design
Place:	EPI MASCOTTE, INRIA Sophia Antipolis - Méditerranée - FRANCE
Period:	2 to 3 months from June to September 2009
Supervisors:	Stéphane Pérennes, research scientist (CR1 CNRS) and Frédéric Giroire, research scientist (CR2 CNRS)
Prerequisites:	Knowledge and/or interest for networking and probability.

Context. Traditional means to store data are dedicated servers or magnetic tapes. These solutions are reliable but expensive. Recently, hard disks and bandwidth have become cheaper and widely available, allowing new forms of data storage on distributed, peer-to-peer (P2P) architectures. A large number of such systems have been built: These systems are cheap to operate, but their highly distributed nature raises questions about reliability, durability, availability, confidentiality, and routing of the data.

An abundant literature exists on the topic of P2P storage systems. Several efforts to build large-scale self-managing distributed systems have been done, among others, Intermemory, Ocean Store, Freenet, PASTRY, CFS, Total Recall [7, 8, 4, 5, 2]. However, few analytical models have been proposed to estimate the behavior of the system (the data durability, resource usage, e.g., bandwidth) and understand the trade-offs between the system parameters.

In [3], the authors show that a scalable system with high peer dynamicity and high availability is non feasible due to bandwidth limitations. Note that in the backup system that we study here, we do not consider churn as the disks are almost continuously connected to the network. The behavior of a storage system using full replication is studied in [9]. A Markov Chain Model is used to derive the lifetime of the system, and other practical metrics like storage and bandwidth limits on the system. In [1], the authors also use a Markovian analysis, but for a system using ECCs. They analyze the performance of two different schemes (a centralized and a distributed) to recover the data and estimate the data lifetime and availability.

Problem. In all these models, the behavior of a single block is modeled and the block failures are considered independent. We showed that this assumption can lead to severe errors of estimation on the behavior of a system subject to peer failures [6]. Therefore the need of new analytical models to describe the systems.

Internship objectives. This internship will be part of the ongoing work on self-repairing P2P storage systems inside the SPREADS project (`http://golgoth.inria.fr/wiki/Contrats/SPREADS`). The goal is to study and propose models to describe such systems. In particular, we aim at:

- Understanding the trade-offs between different reconstruction policies;
- Modeling and analysing the system behavior in the context of limited network resources;
- Analysing and simulating the system reaction to sharp changes (arrival of new users, multiple failures or flooding attacks).

The intern will be involved in all of the different steps of the project: modeling, mathematical analysis and validation of the models by simulations (programing part of the internship). A preliminary period of reading (in particular [1, 6]) will also be an opportunity for him to acquire knowledge of the networking field.

References

[1] S. Alouf, A. Dandoush, and P. Nain. Performance Analysis of Peer-to-Peer Storage Systems. *LNCS*, 4516:642–653, 2007.

[2] R. Bhagwan, K. Tati, Y. chung Cheng, S. Savage, and G. M. Voelker. Total recall: System support for automated availability management. In *Proc. of NSDI*, pages 337–350, 2004.

[3] C. Blake and R. Rodrigues. High availability, scalable storage, dynamic peer networks: pick two. In *Proc. of the 9th conference on HoToS'03*, 2003.

[4] I. Clarke, O. Sandberg, B. Wiley, and T. Hong. Freenet: A Distributed Anonymous Information Storage and Retrieval System. *LNCS*, pages 46–66, 2001.

[5] F. Dabek, M. F. Kaashoek, D. Karger, R. Morris, and I. Stoica. Wide-area cooperative storage with CFS. In *Proc. of the 18th ACM SOSP*, pages 202–215, 2001.

[6] O. Dalle, F. Giroire, J. Monteiro, and S. Pérennes. Analysis of failure correlation in peer-to-peer storage systems. Technical Report RR-6771, INRIA, Dec 2008.

[7] A. V. Goldberg and P. N. Yianilos. Towards an archival intermemory. In *Proc. of ADL Conf.*, page 147, USA, 1998.

[8] J. Kubiatowicz, D. Bindel, Y. Chen, S. Czerwinski, P. Eaton, D. Geels, R. Gummadi, S. Rhea, H. Weatherspoon, C. Wells, et al. OceanStore: an architecture for global-scale persistent storage. *ACM SIGARCH Computer Architecture News*, 28(5):190–201, 2000.

[9] S. Ramabhadran and J. Pasquale. Analysis of long-running replicated systems. In *Proc. of INFOCOM*, 2006.