Maintaining consistency in large-scale opportunistic networks

Résumé du projet de recherche (Langue 1)

In a distributed system, information sharing is achieved by replicating the data on all the terminals where it is used. This enables a user to access it without any network latency, and even while disconnected from the infrastructure. However, when updating data, the writer must propagate the update to all replicas (opportunistically, when they are reachable). Replicas diverge, and concurrent updates may conflict, i.e., violate an application invariant. The general solution to consistency, used for instance in database clusters, is for the system to impose a total order of operations, and to execute all operations in the same order on all sites. As the system scales up, serialization becomes a bottleneck, and full replication increases the load on all processors. This approach obviously won’t work well in large-scale, intermittently-connected networks, e.g., opportunistic social networks. This suggests using an optimistic replication approach [3]. Regal previously implemented an experimental optimistic platform called Telex [1]. Each terminal executes independently using its replicated local state. Via epidemic communication, sites exchange their actions, augmented with concurrency and conflict constraints. The system ensures eventual commitment, i.e., all sites converge to a state that satisfies application invariants, according to proposals that the different sites exchange with one another. The application invariants, which Telex will guarantee, may be completely arbitrary.

Résumé du projet de recherche (Langue 2)

The current Telex architecture has two limitations. (1) Commitment is a consensus protocol, and requires good communication between sites, whereas an opportunistic network is fluctuating and uncontrolled. (2) Telex is monolithic and requires too much computing resource for small mobile terminals. Therefore, we will study the extension of Telex to a two-tier structure, a small stable core, and a variable-size nebula of terminals that communicate with each other and with the core opportunistically. Nebula terminals run simplified versions of the application; updates propagate opportunistically within the nebula, and to the core. The commitment protocol runs only on the core. Commitment remains a bottleneck: can we get rid of it? Unfortunately commitment is required if some concurrent operations do not commute. Therefore, we propose a complementary approach: to study data structures whose concurrent operations always commute, which we call Commutative Replicated Data Type (CRDT) [2]. This approach, combined with causal-order opportunistic communication, ensures that replicas converge, without any need for concurrency control or for commitment.

Informations complémentaires (Langue 1)

Participation aux conférences internationales. "Internship" dans un autre labo durant la première ou la deuxième année.

Informations complémentaires (Langue 2)

Applicants shall be interested in distributed and/or peer-to-peer systems, and in mobile, opportunistic, or social networking applications; have a good knowledge of distributed algorithms, and good programming skills; and be comfortable with running large-scale experiments. Please include the following information (in French or English) with your application: - A resume or Curriculum Vitae. - The list of courses followed in your last year of study and the corresponding marks. - Name and electronic address of two references (we will contact them ourselves). - A short essay on the proposed topic. It is free form, but here are some suggestions. Using the bibliography and your technical knowledge, motivate your interest in the topic and your qualifications for it; discuss scientific and practical issues, problems and difficulties, possible solutions, relate the issues with your own experience, etc. - Any article, report or thesis, and any code you wrote in the past year. ([Bibliography]) [1] Lamia Bennou?ok, Jean-Michel Busca, Joan Manuel Marquès, Marc Shapiro, Pierre Sutra, and Georgios Tsoukalas. Telex: A semantic platform for cooperative application development. In Conf. Française sur les Systèmes d’Exploitation (CFSE), Toulouse, France, September 2009. [2] Mihai Le?ia, Nuno Preguiça, and Marc Shapiro. CRDTs: Consistency without concurrency control. In SOSP W. on Large Scale Distributed Systems and Midd leware (LADIS), Big Sky, MT, USA, October 2009. [3] Yasushi Saito and Marc Shapiro. Optimistic replication. Computing Surveys, 37(1):42–81, March 2005.