Future Generation Computer Systems 40 (2014) 1–16

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Giving form to social cloud storage through experimentation: Issues and insights

Raúl Gracia-Tinedo *, Marc Sánchez-Artigas, Aleix Ramírez, Adrián Moreno-Martínez, Xavier León, Pedro García-López

Universitat Rovira i Virgili, Department of Computer Engineering and Maths, Av. Paisos Catalans 26, Tarragona, Spain

HIGHLIGHTS

- To define and explore a social cloud for storage.
- Impact of network topology/user availability on a social cloud storage service.
- We implement a hybrid model (contributory storage/cloud service) in FriendBox.
- Trade-offs between service quality and economic cost in a social cloud.
- Alternative mechanisms to make a fairer use of user resources.

ARTICLE INFO

Article history:
Received 8 January 2014
Received in revised form
23 April 2014
Accepted 11 May 2014
Available online 17 May 2014

Keywords:
Social cloud
Social networks
Cloud computing
Storage

ABSTRACT

In the last few years, we have seen a rapid expansion of social networking. Digital relationships between individuals are becoming capital for turning to one another for communication and collaboration. These online relationships are creating new opportunities to define socially oriented computing models. In this paper, we propose to leverage these relationships to form a dynamic "social cloud" for storage. While at first glance, the concept of social cloud looks very appealing, a deeper analysis brings out many problems, particularly in data availability. To overcome this issue, in addition to digital friends, we propose to the members of the social cloud the use of online storage services like Amazon S3 to store data and improve data availability. Through a real deployment in our campus, we study what aspects give form to the definition of social cloud storage and determine the difficulty of realizing this concept in the real world. Our analysis reveals interesting insights of how to reap the full potential of socially oriented storage.

1. Introduction

Online social networks, such as Facebook, Google+ and LinkedIn, are becoming a predominant service today. Catering for people of all ages, gender and class, social networking services have become the primary means of communication between friends, family and colleagues. These digital relationships are creating new opportunities to spur the adoption of socially oriented computing.

One representative example of this trend is the concept of "social cloud" as a means of facilitating resource sharing by utilizing the relationships established between members of a social network [1,2]. A social cloud leverages preexisting trust relationships between users to enable mutually beneficial sharing. This facilitates long term sharing with lower privacy and security requirements than those that are present in traditional cloud environments. For the time being, the cloud accrues massive amounts of private information to provide for instance highly targeted advertisements. Not surprisingly, security breaches, poor judgment, or even the lack of judicial oversight leaves users vulnerable. In this sense, the "social cloud" represents a new form for the users to retake control of the cloud service, avoiding to be tracked or give personal information against their will, or in a way in which they feel uncomfortable. In fact, as pointed out by S. Pearson [3], one of the "top six" recommended privacy practices for cloud systems is to maximize user control, which is one of the outstanding feature of the "social cloud".

Another distinguishing feature of the "social cloud" is that the network comes first. It is not a cloud or middleware extended with a social network; rather, it is a social network extended with

* Corresponding author. Tel.: +34 977558745.
E-mail addresses: raul.gracia@urv.cat (R. Gracia-Tinedo), marc.sanchez@urv.cat (M. Sánchez-Artigas), aleix.ramirez@urv.cat (A. Ramírez), adrian.moreno@urv.cat (A. Moreno-Martínez), xavier.leon@urv.cat (X. León), pedro.garcia@urv.cat (P. García-López).

http://dx.doi.org/10.1016/j.future.2014.05.001

cloud functionality. Users form the basic infrastructure and share resources around their social graphs. Such an organization brings out many benefits. For instance, one of those advantages is usability, since the interface and tools for resource sharing are already familiar to users. Another one is that it allows users to maximize the control of the cloud service by letting users choose how their resources will be used. Giving users the control over their personal information and resources engenders trust, but this can be difficult in a cloud computing scenario. This feature is very interesting for the adoption of the "social cloud", as it permits users to define a series of preferences for the management of their personal data, and take account for that, among other advantages.

1.1. Motivation

The social cloud also carries important deficiencies. The most critical one is that, contrary to commercial clouds, it is not feasible to establish a formal Service Level Agreement (SLA) within a social cloud system. Its operational feasibility is based on the premise that participants are socially motivated and subject to the personal repercussions outside the functional scope of the social cloud. This is primarily due to the existing level of trust that already exists between members. In this context, SLAs or "contracts" should be viewed as a best effort agreement between the social links. This weaker form of agreement translates into a limited availability of resources and capabilities. Although a social cloud system is built upon social incentives, peer pressure, etc., the discontinuous participation of social contacts, or even the abandon of the social cloud, is intrinsic to the nature of social relationships.

In terms of storage, this means that the data stored within the social cloud may be subject to recurrent periods of unavailability. In a social cloud, the percentage of time that data is available is a function of the number of friends contributing their storage space over time. And such a dependence has deep implications for the correct operation of a social cloud, mainly in terms of data availability, understood as the probability to access a data item when needed.

First, while there may be a sizable number of individuals in a social network, typically only an insignificant number can be utilized as a destination for personal data. To inform this argument, over 63% of `Facebook` users have less than 100 friends, and the majority of social interactions occur only across a small subset of them [4]. More specifically, it has recently been observed that only 20% of the social links capitalize 70% of all social interactions [4]. This means that in practice the number of users willing to contribute their storage resources to sustain the social cloud will be small. If in addition to this we add the problem of the temporal correlation in the connection habits of users, the loss of data availability is inevitable. Real measurements from online social networks have detected the presence of strong daily and weekly interaction patterns [5,6]. Very succinctly, this means that the probability of finding simultaneously offline all the social links of a user is high, particularly during night hours, which makes it impossible to maintain data availability even under full replication where a replica is allocated to every member of the social cloud.

Second, the topology of the social network graph plays a central role. As such, it delineates the interaction events that may occur across social links and hence, the amount of resources to be contributed by a member. Although users with many friends have a greater opportunity to store their data with higher availability, they may possibly have to donate more disk space to reciprocate a larger number of friends. Real measurements of social networks [7,4] show that while clustering is very high, the existence of a few users with a large number of friends is characteristic of social interaction. For these users with abnormally high degrees, usually called *hubs* in the graph literature, the contribution of their storage resources may be high for little or no personal gain. In this sense, poor storage fairness may motivate the need for economic or non-economic mechanisms to regulate sharing within a social cloud. Determining the graph properties that have an important bearing on a social cloud is critical to answer questions like: *Is the clustering coefficient a valid indicator of resource contribution? If not, which graph properties determine the obligation to trade storage resources?*

Overall, understanding these factors is a necessary step in determining whether the vision of social cloud is realizable, and therefore, it can really emerge as an alternative to commercial cloud providers. Compared with cloud storage, the information is made only available to *trustable* members of the social network, thus significantly reducing the risk that personal data might be sold on, and without raising suspicions about how commercial storage services are monetized.

1.2. Contributions

However, to truly involve users, we believe that the promise of always available storage is essential. In a recent paper, we demonstrated that this promise cannot be fulfilled today using only social links as discussed above [8]. For this reason, we study in this paper a realistic model for building highly available social clouds. In this model, the storage resources contributed by each user are augmented with an external cloud storage service like Amazon S3. Each member of the social cloud brings out its online cloud storage service to store parts of its data and mask the recurrent, unavailability periods of "friends". Our objective is to improve the resilience of the social cloud to correlated failures and departures.

Our key insight is the following: *Since the central cloud maintains data availability during the time periods where most of the social links are disconnected, we are taking the first step towards the realization of storage as a service atop a social cloud*, i.e., the illusion that users can store their data to a socially motivated cloud and access them anytime from anywhere.

At this point, a natural question that arises is: *What can this storage model offer that more established sites, like* `Facebook` *or a combination of cloud storage plus social network like* `Google Drive` *with* `Google+`*, don't?* The answer is that data is not in possession of these sites and therefore, they cannot generate revenue, for instance, by targeting ads to specific demographics (e.g., single males up to 21 years of age). Specifically, in a social cloud, the control of data, and who can access it, is entirely in the hands of users. The role of the social network site is restricted to connecting and recruiting members for the social cloud through a familiar interface. But the data is out of the control of the social network site operators.

While that sounds good, the use of the cloud also poses a new question: *Does the use of a public cloud such as Amazon S3 carry the danger of undermining the security achieved by the social cloud?* Fortunately, the answer is negative, because our model operates by first encoding, and then distributing, the information between the social contacts and the cloud in such a manner that the cloud cannot recover the original data. In our particular case, we use a non-systematic Reed–Solomon code [9] for that purpose. The code was chosen to be non-systematic in order to make the encoded data not readable at once. Recall that threshold schemes like Shamir's scheme [10] for sharing a secret among multiple participants can be re-formulated in terms of Reed–Solomon codes [11]. As a result, we can blend "the best of both worlds" in a single approach: high data availability and security, the latter thanks to both the maximization of user control and the minimization of the data sent to and stored in the cloud.

To gain a better understanding, this paper contributes to the state of the art by quantifying the influence of the above factors, putting special emphasis on the topological effects, while outlining