Chapter 6

Distributed Social Platforms for Confidentiality and Resilience

Enrico Franchi
University of Parma, Italy

Michele Tomaiuolo
University of Parma, Italy

ABSTRACT

Social networking sites have deeply changed the perception of the web in the last years. Although the current approach to build social networking systems is to create huge centralized systems owned by a single company, such strategy has many drawbacks, e.g., lack of privacy, lack of anonymity, risks of censorship and operating costs. These issues contrast with some of the main requirements of information systems, including: (i) confidentiality, i.e., the interactions between a user and the system must remain private unless explicitly public; (ii) integrity; (iii) accountability; (iv) availability; (v) identity and anonymity. Moreover, social networking platforms are vulnerable to many kind of attacks: (i) masquerading, which occurs when a user disguises his identity and pretends to be another user; (ii) unauthorized access; (iii) denial of service; (iv) repudiation, which occurs when a user participates in an activity and later claims he did not; (v) eavesdropping; (vi) alteration of data; (vii) copy and replay attacks; and, in general, (viii) attacks making use of social engineering techniques. In order to overcome both the intrinsic defects of centralized systems and the general vulnerabilities of social networking platforms, many different approaches have been proposed, both as federated (i.e., consisting of multiple entities cooperating to provide the service, but usually distinct from users) or peer-to-peer systems (with users directly cooperating to provide the service); in this work the most interesting ones were reviewed. Eventually, the authors present their own approach to create a solid distributed social networking platform consisting in a novel peer-to-peer system that leverages existing, widespread and stable technologies such as distributed hash tables and BitTorrent. The topics considered in detail are: (i) anonymity and resilience to censorship; (ii) authenticatable contents; (iii) semantic interoperability using activity streams and weak semantic data formats for contacts and profiles; and (iv) data availability.

DOI: 10.4018/978-1-4666-3926-3.ch006
INTRODUCTION

Nowadays, millions of people of any age and gender regularly access Online Social Networks (OSNs) and spend most of their online time social networking. According to Boyd and Ellison (2008), teenagers have a clear understanding of privacy related issues; however, the same does not apply to some adults that (i) did not even use email and other basic Internet services before the social networking revolution (Stroud, 2008) and (ii) not only have limited computer-related technical skills, but they also lack risk consciousness about privacy issues. Moreover, many people are becoming uncomfortable with the presence of their employers in the same social networking systems, because some personal data may leak in their corporate environment due to privacy configuration errors (Skeels & Grudin, 2009).

However, privacy threats can also come from the service providers. In fact, even if the social networking systems are greatly dissimilar in their user base and functionality, they are almost always centralized systems. Because of their centralized nature, a simple browser-based user experience is possible and, moreover, many algorithms, e.g., friend suggestion, are far easier and more efficient to implement.

A minor drawback is that scaling centralized systems to tens or hundreds of millions of users is not an easy task. At any rate, while we consider this drawback as a minor one from a technical point of view, since the problem can be solved providing enough resources, it becomes a huge social drawback, because for most companies advertisement is the main source of income and, consequently, they have strong motive to make it as precise as possible, typically mining user provided data. This behavior poses serious threats to privacy and data protection issues. In fact, many social networking systems have very demanding terms of service, essentially asking their users a non-exclusive, transferable, sub-licensable, royalty-free, worldwide license to use content that they submit (Facebook, 2011; Twitter, 2011).

Moreover, social networking sites guide their users into “walled gardens,” without giving users full control over their own information because such information constitutes much of their company value (Shankland, 2011; Berners-Lee, 2010).

A second feature of centralized systems is that service providers are in the position to effectively perform a-priori or a-posteriori censorship, or to disclose all the information they have, no matter how private, to other entities. They can perform such actions either motivated by selfish interests or forced under legal terms and other forms of pressure.

Considering that: (i) no single centralized entity can withstand the operative costs of a large scale social networking system without a solid business-plan; (ii) most business plans are based on targeted advertisement; and (iii) even if a service provider would be fair with its user’s data, it would remain vulnerable to legal requests to disclose such data, we favor a P2P approach.

In the first place, P2P systems essentially achieve automatic resource scalability, in the sense that the availability of resources is proportional to the number of users. This property is especially desirable for media sharing social networking systems, considering the exceptionally high amount of resources needed.

Moreover, regarding censorship issues, a P2P system essentially solves them by design. Without a central entity, nobody is in the position of censoring data systematically, nor may be held legally responsible for the diffusion of censorable data: the sole owners and responsible of the data are the users themselves. However, P2P systems, and especially those based on a Distributed Hash Table (DHT), may be liable to attacks meant to disrupt the system functionality (Urdaneta et al., 2011); in this particular scenario the severity of such attacks may be mitigated using the social network itself as source of human trust relation-