Privacy issues in cloud-computing

Kostas Chatzikokolakis
CNRS, INRIA team Comète and LIX, École Polytechnique
kostas@lix.polytechnique.fr

The term "cloud" is one of the most overloaded and vaguely defined terms in information technology. In a strict sense, "cloud computing" has been used to describe computing performed in virtual servers, typically operated by a third party in a remote location, which are provisioned and charged in a dynamic "on-demand" way. More broadly, "cloud" has being used as a metaphor for the Internet. In this case, a "cloud-based" service is a service provided by a third party and delivered to the user through the Internet, which would traditionally require infrastructure operated by the user himself, such as servers or software. Such services are said to run "in the cloud" meaning in a remote location using a third-party's infrastructure. Cloud-based services allows users, individuals or businesses, to access a desired service without the overhead of operating the necessary infrastructure, which also usually implies a lower cost. Moreover, a cloud-based service often provides easier access, higher availability and reliability, better scalability and performance, and lower maintenance needs, compared to a locally operated solution.

Email is one of the best known examples of a cloud-based service. Sending and receiving emails requires a mail server operated by the domain owner. As the cost and burden of operating such a server is high, most private email accounts are hosted by a third-party, typically a global mail provider such as Yahoo and Hotmail. More recently, cloud-based services are being increasingly used also for corporate email, where all email addresses of an organisation are hosted by providers such as Google Apps and Zimbra.

Moreover, tasks that have traditionally been performed using software running in desktop computers can now be performed online using cloud-based services. Word-processing, office productivity, image processing, video processing, customer management, etc, can be now performed online using providers such as Google Docs, pixlr.com, encoding.com, salesforge.com, etc. As more and more tasks are moved from the PC to the cloud, a desktop computer becomes a simple device that accesses the Internet, while all data storage and computation happen in remote servers.

Last, but not least, a good example of the trend for cloud-based services can be found within the scientific community. Conference management systems are used by most academic conferences to facilitate several tasks of the event's organisation, the most prominent one being paper reviewing. The system automates tasks such as distributing of papers, collecting the reviews, ranking the papers, sending notification emails, etc. Systems such as HotCRP and iChair must be installed and maintained by the conference chair in a local server. On the other hand, increasingly popular systems such as EasyChair are entirely

cloud-based, the chair only needs to create an account for the specific conference.

Despite its vague meaning and overloaded use, one thing is certain about the "cloud": it is associated with a tendency to move all data and computation to infrastructure operated by a third party, and this tendency is becoming highly popular. As a result, privacy is becoming a major user concern. First, the cloud provider needs to be completely trusted as he obtains access to all the data hosted in the cloud. Such a trust becomes harder when multiple providers are involved; for example an image processing service, run by provider A, might use a virtualized server of provider B and cloud-based storage of provider C, leading to a situation when 3 different parties need to be trusted. Moreover, privacy violations can happen without a party being malicious. Accidental data loss due to stolen laptops or USB keys has become common. Moreover, even though cloud providers usually employ strict security measures, an intruder's incentive to break into such systems is also much greater.

Cloud-based services offer numerous advantages and are here to stay. This, however, does not imply that user privacy should be completely surrendered. In fact, satisfying privacy guarantees would greatly accelerate the adoption of such services. Legislation and policies are one step, but the ideal solution would be a technical one. Depending on the application, it is feasible to provide certain levels of privacy, while keeping to a great extend the cloud-centric nature of a service. Using PGP for email, for example, can offer great privacy, at the expense of losing search functionality. Similarly, an e-health system could be hosted in the cloud and still protect the patient's anonymity, provided that the link between the patient and his medical record is stored in a protected way, not accessible to the cloud provider. Clearly, as systems become more complex and more functionality is moved to the cloud, satisfying anonymity and privacy becomes a challenging task.

A long-term goal of Comète is the development of protocols and techniques for preserving anonymity and privacy in cloud-based services, as well as methods for formally verifying these protocols. Having a motivating example is important for this goal. As a consequence, we plan to focus in designing and implementing a "privacy-friendly" conference management system that provides the functionality and ease-of-use of cloud-based systems (such as EasyChair) while protecting the users' confidential data. Such a motivating example will allow us to understand in more concrete terms the challenges - both theoretical and practical - of the privacy problem in cloud computing, and at the same time will offer a useful product to the scientific community.