Cloud Privacy: Normative Standards Needed to Foster Innovation

Abstract

The principal thesis of this comment is that the free flow of personal information that respects privacy can fuel and cultivate innovation. The growth of cloud computing has led to a corresponding growth of user data stored on third party servers. While technology has advanced, privacy protection has been slow to evolve. U.S. privacy law has significant limitations in the protection of aggregated data—precisely the type of a substantial amount of cloud computing data. Privacy laws, regulations, and policies will have to be modernized and improved to instill trust in cloud computing, thus providing a foundation to support its growth and innovation.

Introduction

This comment focuses on cloud computing and privacy. As discussed below, cloud computing has become a significant growth area in the Internet economy. It is powering a significant amount of innovation, primarily based upon the collection and dissemination of personal information. As such, “cloud privacy” is an important concept within the Department of Commerce Internet Policy Task Force’s review of the nexus between privacy policy and innovation in the Internet economy.

This comment provides a brief overview of cloud computing and current privacy law, policies and practices. This comment then presents fundamental ideas for modernizing privacy protection for cloud computing users. The ultimate goal of the ideas put forth in this comment is to instill trust in the use of cloud computing, fostering further development of innovations in cloud computing applications.
Cloud Computing and Privacy

With the advent of online search, social networking, e-mail and other web-based applications, cloud computing is growing rapidly (e.g., Diaz 2009). Historically, the law has been slow to adapt to new technologies. This phenomenon is being repeated as the growth in cloud computing outpaces necessary privacy protections for its users. After a brief description of cloud computing and an overview of U.S. privacy law, this section discusses specific privacy protection shortcomings associated with cloud computing.

Cloud Computing

Cloud computing is a software application and data management approach utilizing web-based applications that access and store data on servers provided by third parties (Soghoian 2009). In theory, cloud computing allows developers to deploy and run applications that are highly scalable with a high degree of reliability (Perry 2008). Cloud computing applications include search engines, social networking sites such as Facebook, blog hosting platforms and associated utilities such as traffic monitoring, e-mail services, and application suites which include word processing, spreadsheet, and presentation software. Cloud computing also encompasses online data storage and backup, both independent of and in association with web-based applications. Recent surveys indicate that 69% of Internet users store data online or use web-based applications provided by companies including Amazon, Google, Microsoft, and Yahoo (Horrigan 2008; Soghoian 2009). The result is that cloud computing providers have access to a substantial amount data created by or about end users (Picker 2008). This gives rise to another description of cloud computing: “any computer network or system through which personal information is transmitted, processed, and stored, and over which individuals have little direct knowledge, involvement, or control” (Privacy Rights Clearinghouse 2009).

Individually, each piece of personal information represents a mere pixel of someone’s life, but when pieced together, they present a rather detailed picture of that person’s identity (Ciocchetti 2007). In addition to the surprise of learning that one’s online searches can reveal one’s specific identity, as discussed more fully below, cloud computing users are already beginning to express concern over exactly what is being done with the data collected from online activities (Horrigan 2008). One recent survey finds that a significant majority of individuals (between 73% to 86%) do not want their web activities tracked in order to receive targeted advertising (Turow et al. 2009).

Privacy Law in the United States

U.S. privacy law has a long and complex history, driven primarily by reactions to advances in technology. Instant photography and audio recording devices led to the first calls for formal recognition of legal rights to privacy (Warren and Brandeis 1890; Sprague 2008). By the mid-twentieth century, most states recognized a common law right against highly intrusive invasions of private matters (Prosser 1960; Sprague 2008). Wiretap technology led to the U.S. Supreme Court recognizing an expectation of privacy against government eavesdropping (Katz v. United States 1967). Since then, essentially every common law and constitutional right to privacy evaluation has centered on a person’s reasonable expectation of privacy (Sprague 2008). Unfortunately, there can never be a definitive test to determine what is and is not private (Solove
“We have no talisman that determines in all cases those privacy expectations that society is prepared to accept as reasonable” (O’Connor v. Ortega 1987, p. 715).

Congress’ concern with privacy began in the 1960s with the government’s growing computerized collection and use of personal information (Ciocchetti 2007). Federal privacy legislation has focused on government intrusions and isolated areas of consumer protection, such as movie rental records and health and financial information (Sprague 2008). With the exception of security, the government has relied primarily on private-market solutions to privacy protection through privacy policies (Ciocchetti 2007). From the security perspective, the model has focused on state-level laws that require companies to encrypt data and notify consumers of breaches of unencrypted personal information (Sprague and Ciocchetti 2009).

Privacy law in the United States protects only that information which individuals keep private; it does not protect information which individuals voluntarily disclose to anyone else (Katz v. United States 1967). Courts, for example, assume that a person loses a reasonable expectation of privacy in e-mail messages once they are sent to and received by a third party (Rehberg v. Paulk 2010). In addition, current privacy law does not necessarily protect information derived from the accumulation of data. In other words, when individuals voluntarily relinquish their right to privacy over small, unique pieces of information, an analysis of accumulated data may generate a much fuller profile, which itself is not protected because the underlying data are not protected (Solove 2001). There are very few legal limitations on what can be done with data if its original means of collection did not invade an individual’s privacy.

Google and Facebook

Google and Facebook are good starting points for analyzing cloud privacy. In addition to search, Google offers, among a number of applications, e-mail, an application suite, a calendar, a blogging platform, website hosting services, and a web browser—all of which collect and/or store user-related data on Google’s servers. Internet users in the United States spend an estimated 9% of their time online using some Google service (Helft 2009). Google’s industry dominance, coupled with its mission to organize the world’s information and make it universally accessible and useful (Google Company Overview 2010), raises significant privacy issues (Inside the Googleplex 2007). While Google has an overarching Privacy Policy, it also has privacy policies for more than forty specific applications and services (Google Privacy Center 2010).

Google’s Privacy Policy attempts to explain all of the ways in which users’ personal identifying information may be collected, stored, used, and distributed to third parties. Some commentators are troubled by the vagueness of some of the language within Google’s Privacy Policy, including Google’s obligations regarding requests for user information by law enforcement agencies (e.g., Tene 2008). Google’s Privacy Policy also allows it to share “aggregated, non-personal information” with advertisers, so long as the information does not individually identify users. However, as noted below, it is possible to unveil the identity of a search engine user based on anonymized search results.

More troubling, which exemplifies how many online privacy policies are implemented (Sprague and Ciocchetti 2009), Google reserves the right to change its privacy policies at any
time; and Google’s terms of service state that use of a Google service constitutes an agreement to accept those terms, which include an agreement by users that Google can use their information according to its privacy policies (Google Terms of Service 2007). The result is that every time an individual uses a Google service, that person is accepting Google’s Privacy Policy as it exists at that time (cf. Tene 2008).

Chris Jay Hoofnagle refers to this as a Machiavellian approach to privacy (Hoofnagle 2010). Google initially focused on users’ interests per-transaction, rather than through an analysis of past searches and browsing. But in 2007, Google quietly began behavioral profiling, tracking searches, and, with the acquisition of DoubleClick, nearly all browsing behavior (Hoofnagle 2010). Meanwhile, Facebook offers users substantial control over their privacy, with more than 100 settings. This array of privacy options is so confusing, many users typically choose poorly or not at all while, at the same time, Facebook is making publicly available more and more user information (Hoofnagle 2010). Even after changing privacy settings, certain information will still be shared across Facebook unless users take additional steps. For example, users must change certain “Account Settings” to prevent information from being shared with advertising networks and friends. The only way to prevent some personal data from being shared is to delete it (Bilton 2010).

User Privacy Policies: A Fallacy?

Facebook has recently been at the center of a privacy storm, with near-rebellions by users over Facebook’s continually changing privacy policies (e.g., Nussbaum 2010; Wortham 2010), spurring even a call for a bill of privacy rights for social networks (Opsahl 2010). Facebook’s privacy policy has grown from just over 1,000 words in 2005 to nearly 6,000 words in 2010 (Facebook Privacy: A Bewildering Tangle of Options 2010). Bowing to pressure from various constituencies, Facebook has recently attempted to simplify its privacy settings (Helft and Wortham 2010). And, as noted above, Google publishes over forty separate privacy policies. But recent court decisions have clearly signaled that privacy policies provide no real legal protection for users (Serwin 2008-2009).

For example, in 2004 a number of JetBlue Airways customers sued after JetBlue transferred some 5 million customer records to a data mining company, which then combined the JetBlue data with credit reporting data to create profiles of the customers; all under the auspices of a post-9/11 Homeland Security Agency initiative (In re JetBlue Airways Corp. Privacy Litigation 2005). The claims brought by the JetBlue customers included breach of contract, based on JetBlue’s privacy policy which stated that customer data would not be shared with third parties without the customers’ prior consent. However, a successful breach of contract claim requires the non-breaching party to suffer damages. A loss of privacy is a non-economic loss that is not compensable under contract law (In re JetBlue Airways Corp. Privacy Litigation 2005, p.327-328; Trikas v. Universal Card Services Corp. 2005, p. 46). Further, even though the data may be personal information about a person, it does not belong to that person (In re JetBlue Airways Corp. Privacy Litigation 2005, p.328).
De-anonymization and Data Leakage

By their very nature, many cloud computing applications collect personal identifying information—consider the e-mail application that stores all the e-mail addresses and content of every message sent and received by each user. Similarly, there is concern that Google has the ability to link a reader to every book searched for, browsed, purchased and read through its Google Books application, including which particular pages the user reads and for how long (Privacy Authors and Publishers’ Objection to Proposed Settlement 2009); and Google’s goal is to digitize and make searchable every book in existence (Stross 2008). But even supposedly anonymous data can be traced back to their originator. For example, when America Online published user search data for academics, specific individuals were able to be identified by the content of their searches when cross-linked with other data (Barbaro and Zeller, Jr. 2006). A recent MIT class project revealed that by looking at a person’s Facebook friends, students were able to reportedly determine whether the person was gay (Johnson 2009). Mislove et al. (2010) have developed techniques to infer attributes of online social network members based on the attributes of other members.

Netflix, the online movie rental company, recently awarded a $1 million prize in a competition open to the public to improve its movie recommendation software. Thousands of teams from 186 countries examined a data set with 100 million movie ratings over a two-year period (Lohr 2009). Netflix planned a second competition, again, open to the public, but with a data set with more than 100 million entries that would have included information about renters’ ages, gender, ZIP codes, genre ratings and previously chosen movies (Lohr 2009). However, Netflix canceled the second contest due to privacy concerns (Lohr 2010), losing the potential to enhance the services it provides to its customers. But there was good cause for concern, as Narayanan and Shmatikov (2009a) were able to identify individual Netflix subscribers from the contest’s first data set, uncovering the users’ apparent political preferences and other potentially sensitive information.

A number of academics have recently published research revealing the ability to identify individuals by linking anonymized data with other available data. Narayanan and Shmatikov (2009b) have developed a re-identification algorithm which can recognize up to one-third of verifiable Twitter (the microblogging service) and Flikr (an online photo-sharing site) members with a 12% error rate. Acquisti and Gross (2009) have discovered that personal information from multiple sources, such as data brokers or profiles on social networking sites, can be used to accurately predict social security numbers. Other researchers have identified potential areas within cloud-based applications wherein personal information could be leaked to third parties (Castelluccia, De Cristofaro and Perito 2009; Krishnamurthy and Wills 2009).

Current Privacy Practices and Guidelines

A variety of scholars have analyzed privacy issues associated with aspects of cloud computing: the collection of personal identifying information (e.g., Ciocchetti 2007; Rubinstein et al. 2008; Sprague and Ciocchetti 2009); the effectiveness of privacy policies (e.g., Culnan 2000; Milne and Culnan 2004; Ciocchetti 2008; Oussayef 2008); search engines (e.g., Foley 2007; Grimmelmann 2007; Tene 2008); and e-mail applications (e.g., Goldberg 2005; Yang 2005). Privacy protection within cloud computing is limited primarily to self-regulation, though
there are existing guidelines which can help inform an analysis of evolving standards for cloud privacy.

**Fair Information Practices and European Union Data Protection**

In 1973, the Department of Health Education and Welfare (HEW) created the first set of fair information practices to be issued by the federal government, recommending that the practices be applied to government management of computerized data-collection systems (Ciocchetti 2007). The HEW practices addressed five information principles: (1) openness, (2) disclosure, (3) secondary use, (4) correction, and (5) security. And while the federal Privacy Act of 1974 required all federal agencies to comply with the HEW fair information practices, such a requirement was never imposed on private entities (Ciocchetti 2007).

In 1998, the Federal Trade Commission (FTC) proposed its own set of fair information practice principles to supplement private entities’ self-regulated privacy practices (Ciocchetti 2007). The current FTC fair information practice principles are: (1) notice/awareness, (2) choice/consent, (3) access/participation, (4) integrity/security, and (5) enforcement/redress (Fair Information Practice Principles 2009). As with the HEW principles, Congress has never required private entities to adopt the FTC principles (Ciocchetti 2007).

Since 1995, the European Union has issued a number of data protection directives (European Union Data Protection 2009). Directive 95/46/EC protects individuals with regard to the processing of personal data in areas including: (1) data quality, (2) legitimacy, (3) consent, (4) notice, (5) access, (6) the right to object, (7) confidentiality and security, (8) notification, and (9) remedies. There is already concern that Google’s privacy practices may be inconsistent with European Union privacy protections. For example, the Data Protection Working Party established under Directive 95/46/EC considers IP addresses as data relating to an identifiable person (Opinion 4/2007 on the Concept of Personal Data 2007). Privacy groups have charged that Google’s e-mail server backup policies violate European Union privacy directives by storing messages where users cannot permanently delete them (Yang 2005).

**Cloud Privacy**

As the cloud computing architecture becomes more prevalent, normative standards for “cloud privacy” will need to be examined and developed. Unless users are confident their data will be secure, used only for limited purposes, and often remaining anonymous, they will resist cloud computing. In the United States, end-user privacy within cloud computing is currently dependent upon self-regulated company-specific privacy policies implemented within a legal structure geared toward protecting involuntary disclosures of private information.

Privacy protection in the United States is dependent upon one’s expectation of privacy at any given time for any particular piece of information. While people may have no expectation of privacy regarding isolated bits of data, such as individual postings on a social networking site or occasional browsing of books, they are not necessarily abandoning an expectation of privacy for a more detailed profile of their life that may be aggregated from these bits of information. It is also reasonable to expect that information disclosed for one purpose, such as a trip itinerary, will not be disclosed to others for a different purpose, such as terrorism surveillance (Singel 2009).
To strengthen cloud computing user privacy, changes to current law must address preserving the privacy of accumulated data regardless of the privacy status of underlying bits of information that comprise that data.

Meaningful cloud privacy also requires codifying the FTC fair information practices principles and the European Union data protection directives. Privacy policies will have to be written in plain English, clearly explaining how user information is collected, stored, used, and possibly shared (Sprague and Ciocchetti 2009). Importantly, active user consent should be required for amendments to privacy policies, rather than passive consent implied by continued use (Sprague and Ciocchetti 2009). Users should be provided the opportunity to opt out of certain practices, rather than having to accept a policy in whole or reject services in their entirety. Finally, remedies for violations will have to be strengthened. With the exception of a few FTC settlements for data breaches, there have been no significant legal or enforcement actions against companies regarding their privacy policies and practices (Sprague and Ciocchetti 2009; e.g., In re JetBlue Airways Corp. Privacy Litigation 2005).

While users of cloud computing applications are technically making public information they regard as private—be it social network postings expected to only be read by friends, the content of e-mail messages, pages of books browsed, or the content of searches—they are unwittingly abandoning their right to privacy for that information. The backlash against cloud computing providers is only beginning. Those who initially considered Facebook to be a private, gated community of trusted friends now see it as becoming an increasingly open, public commons of curious strangers (Nussbaum 2010).

While some commentators (e.g., Tim O’Reilly 2010) believe innovation will be spurred by companies, like Facebook, that make privacy-related blunders and then correct them, the current legal boundaries offer very little protection for end users. Not only must companies adopt privacy policies that conform to fair information practices principles and EU data protection directives, but those policies must be legally enforceable. Further, entire notions of privacy must be updated to appreciate that what may be potentially made public in a limited fashion should still be considered private—allowing the originator of that information some level of control over its access, use and dispersion.

Industry norms also need to be similarly changed so the self-regulation model can be consistently and uniformly modified and implemented to achieve the same objectives. A combination of legal and industry changes can result in a sustainable approach to privacy protection for cloud computing users—fostering trust and increased use of and innovation in cloud computing.

Conclusion

End-user privacy within cloud computing is currently dependent upon self-regulated company-specific privacy policies implemented within a legal structure geared toward protecting involuntary disclosures of private information. In addition, cloud computing providers are free to offer all-or-none, vague and confusing privacy policies that can be unilaterally modified at any time with minimal notice to users. Users will resist cloud computing innovations unless they are confident their data will be secure, used only for limited and expected purposes, and often
remaining anonymous. This comment has offered legal and industry-specific approaches to privacy protection which can instill trust in cloud computing applications, which can lead to their continuing growth and innovation.

Very truly yours,

Robert Sprague, J.D., M.B.A.
Associate Professor of Business Law
University of Wyoming College of Business
Department of Management & Marketing
spraguer@uwyo.edu
URL: http://www.uwyo.edu/sprague
SSRN Author Page: http://ssrn.com/author=623074

References


Rehberg v. Paulk, 598 F.3d 1268 (11th Cir. 2010).


