Privacy in the YouTube Era: Evolving Concepts in the Protection of Personal Information

Jeffrey Stanton
Syracuse University

Hamid Nemati
UNCG

Soon Chun
City University of New York

Jengchung Chen
National Cheng Kung University

Follow this and additional works at: http://aisel.aisnet.org/amcis2007

Recommended Citation
http://aisel.aisnet.org/amcis2007/348
PRIVACY IN THE YOUTUBE ERA: EVOLVING CONCEPTS IN THE PROTECTION OF PERSONAL INFORMATION

Jeffrey M. Stanton, Syracuse University, jmstanto@syr.edu
Hamid Nemati, The University of North Carolina at Greensboro, nemati@uncg.edu
Soon Ae Chun, City University of New York, chun@mail.csi.cuny.edu
Jengchung V. Chen, National Cheng Kung University, Taiwan, victor@mail.ncku.edu.tw

Abstract
Justice Louis Brandeis' concept of privacy in the early 20th century was, "the right to be left alone." Yet in the era of YouTube, social networking, and reality TV, it is increasingly common for individuals to seek widespread exposure of their ideas, images, and selves to the broadest possible audience. At the same time, however, the theft and misuse of personal information continues to grow as more and more businesses and governments make their data accessible over the Internet. We are witnessing, in effect, a privacy paradox in which a hunger for greater exposure is opposed by a growing danger of exposure. Resolution of this paradox lies in using a historical perspective on security and privacy to clarify, segregate, and reorder the component concepts underlying these two terms. This paper attempts to begin that task, in part through "unloading" the overloaded concept of privacy.

Introduction and Goals

The mark of distinction in social networking websites is the number of friends you have. Success in blogging requires amassing substantial numbers of incoming links, especially from other bloggers. The most viral videos are the ones that collect tens of thousands of hits and become embedded in hundreds or thousands of other sites. The same general principle holds throughout: exposure equals connections equals success. This exposure principle formerly applied only to politicians and movie stars, but the exponential growth in the use of simple but powerful web applications has made it possible for millions of individuals all over the world to create their own micro-scale media empires. In short, the kind of widespread public exposure that used to be the domain of the rich and famous is now available to, and apparently highly valued by, legions of regular citizens.

Meanwhile, in a thousand dark corners of the net, crackers, scammers, and organized crime organizations plan and implement a dizzying array of methods to steal information, identities, and funds from as many individuals and institutions as they can. While the historical hackers – the Morrises and Mitnicks – did their mischief in service of curiosity and ego, present day e-felons are driven by baser motivations. The same network technologies that facilitate advanced web applications also generate a global scourge of information crimes.

These two trends – the exposure principle and the growth of net crimes – present an apparent privacy paradox. On the one hand, people seem hungry for connectedness and they willingly reveal enormous amounts of information about themselves to promote those connections. On the other hand, the more personal information that is revealed or made net-accessible the easier it becomes for malicious individuals and organizations to exploit that information for criminal purposes. The desire to reveal information and the need to protect information seem in direct opposition to one another. In the YouTube era the old rules of privacy no longer apply: constructing a fortress around private life is no longer feasible or desirable, but it is still just as necessary as it has always been. This paper attempts to unpack this paradox. By examining the space between information privacy and information security, we hope to chart a course towards systems, applications, and practices that promote the exposure of information while preventing its exploitation.
The importance of privacy as an individual right traces its modern connection to Warren and Brandeis (1890), who decried invasive examination of individuals’ personal lives by the press. In the ensuing century, law and philosophy scholars defined privacy and discussed its functions and benefits. In parallel, policy makers have enacted laws pertaining to privacy. The Privacy Act establishes some controls over what personal information is collected by the federal government and how it is used, guaranteeing the rights to see and amend the records about oneself as well as the right to sue the government for permitting unauthorized individuals to read one’s records. The Freedom of Information Act of 1966 and 1996 mandates federal agencies to provide public access to information, but not to disclose protected, private information. Rules issued by the U.S. Department of Health and Human Services implement the privacy requirements of the Health Insurance Portability and Accountability Act of 1996 (“HIPAA”). Ruth Gavison (1980) reviewed privacy case law and legislation, provided a coherent legal definition for privacy, and discussed relationships among the concepts of personal information, control, and freedom. She defined privacy as “a concern for limited accessibility” by other individuals to information about us (including observations of our behavior). As this definition suggests, the common thread through legal and philosophical discussions of privacy is the exploration of its meaning and function within society. While these discussions of privacy provided a definitional foundation for inquiry, they do not provide a psychological account of the thoughts and feelings of people who are potentially affected by privacy invasions or violations. Put differently, legal and philosophical writing teaches what privacy is and what its functions are in society, but not how it works in the minds of individuals who use contemporary, Internet enabled information systems.

Researchers examining the impacts of technology have also drawn on economic, social and political analyses. Attewell and Rule (1984) contrasted economic motives for technological innovations in organizations with the political and power motives of managers who support such innovations. As Ehn (1989, pp. 460-461) described, technology can consolidate and amplify the control that managers have over employees. Milberg, Burke, Smith, & Kallman (1995) examined preferences for regulatory control of privacy in nine international settings. Kling (1985) mapped the analytical and political strategies used by urban planners, including addressing end-user concerns for privacy, onto the processes of implementing computer systems within organizations. Kling’s compilation volume, “Computerization and Controversy” (Kling 1996) collected articles that draw on the disciplinary perspectives discussed above to examine technological transformation of work and connections between privacy and social control. Kling, Clement (1994), and other authors have discussed ethical imperatives incumbent on those who design and implement technology. Clement analyzed the privacy implications of multimedia communication “spaces” and provided a guide -- based in part on the Organization for Economic Cooperation and Development’s (OECD) principles of fair information practice -- which multimedia system designers might use to ensure privacy protections for end users of their systems. One way of summarizing this body of work is to say that privacy and technology are linked by the social context of the work groups, social networks, organizations, and societies that use the technology. Describing the social context of privacy is the major strength of these writings: One missing piece that remains is a psychological understanding of privacy at the level of the individual.

For a psychological perspective on privacy, one can turn to social psychology and its cognate fields of sociology and anthropology. Alan Westin wrote a seminal volume, entitled Privacy and Freedom (1970), that drew on each of these fields. Westin defined four facets of privacy -- solitude, intimacy, anonymity, and reserve -- and four psychological functions of privacy -- personal autonomy, emotional release, self-evaluation, and limited and protected communication. These conceptual elements generated research in human-environment interactions and survey instrument development. For example, Pedersen (1997) and Rustemli and Kokdemir (1993) published factor analyses of Westin's four facets. Priest and Bugg (1991) and Hammitt (Hammitt and Brown 1984; Hammitt and Madden 1989) used the facets and functions to measure of salutary effects of outdoor environments. Duvall-Early and Benedict (1992) used Westin's ideas to examine effects of the physical office environment on job satisfaction. Westin's book was also valuable for focusing public attention on privacy issues and for policy-making, but was never intended as theory. In contrast, Irwin Altman’s (1975) book, entitled “The Environment and Social Behavior,” developed a theory of privacy that generated much subsequent research. Altman synthesized research from anthropology, sociology, and psychology to develop his analysis. Part of the appeal of Altman's formulation is that he presented privacy as a boundary regulation process, rather than as a static event or state. The dynamic nature of privacy, Altman claimed, was a function of the inherent tension and necessary balance between closing/protecting and opening/sharing of the self. The theory was time-based, such that processes of opening and closing personal boundaries varied with the context and changes in one's relationships to others.
The idea of boundaries as a metaphor for privacy regulation was later taken up by Petronio, who developed Communications Boundary Theory, and Stanton, who described the closely related Information Boundary Theory. Both formulations construe privacy as a dynamic regulatory process, intended to maximize expressive and instrumental benefits of individuals, while protecting the individual through the exercise of control over personal information. Both of these theoretical perspectives incorporate three important points from Altman’s work:

1. Privacy is a dynamic process rather than a unitary, unchanging state or condition.
2. The dynamic process of privacy is one of balance: The basic tension between revelation and secrecy is governed by competing desires for benefit (expressive or instrumental) and control.
3. Privacy is limited good rather than a universal good: Most people have an intrinsic desire for some degree of revelation of the self to others at some times and in some circumstances.

Later in the paper, we will use these three key points about privacy to examine the relationship between privacy and security.

Information Security: Ally or Enemy of Privacy?

Information Assurance professionals generally describe three or four goals for information security: Security is concerned with how to store, transmit, and receive information without letting others find out the information, how to prevent modifications of information by unauthorized individuals, and how to keep information services available to those who are legitimately allowed to use them. These goals are often referred to as the “CIA” triad: confidentiality, integrity, and availability. To this basic triad, some also add a fourth goal called non-repudiation. Non-repudiation can be thought of as a record of a transaction (e.g., like a receipt from the purchase of a product) that prevents one or both of the parties involved from later saying that they did not participate in the transaction. Most of the essential technological elements of a network incorporate features designed to support one or more of these goals.

These goals have become increasingly important in the information era as a result of two fundamental trends: action at a distance and virtual anonymity. Telecommunications and network technologies, by their intrinsic nature, compress space to the point of irrelevance. Whereas human behavior used to take place mainly in the context of a local community, information technologies now allow an individual to influence the status of systems, devices, and information that are located hundreds or thousands of miles away. At the same time the current architecture of the Internet and the network technologies upon which it depends facilitate the occurrence of behavior that is not traceable to a particular individual. The old idea of a single identity firmly associated with a single individual no longer holds: Using the Internet people can adopt multiple, simultaneous identities, many or all of which might never be traced back to or connected with the individual who created them.

At first blush, action at a distance and lack of traceability both seem to facilitate privacy. Remembering Westin’s four facets of privacy, these two capabilities seem to serve solitude, anonymity, and reserve. Using the Internet, I can get many things accomplished while keeping my physical being – and by extension my private self – hidden from the view of others. The major problem, of course, is that these benefits also accrue to those individuals with criminal intent. The privacy benefits intrinsic in current Internet architecture support wrongdoers by making their malicious actions invisible to those who might prevent or proscribe the wrongful behavior. This problem explains why identity thieves and other Internet criminals have had so much success of late: Anonymous malicious actions at a distance are difficult for authorities to trace, locate, and assign to their originator. Unlike crimes in the physical world, network crimes carry relatively low risk for their perpetrators.

Security techniques that attempt to ensure confidentiality, integrity, and availability are designed to thwart these malicious trends. As a result, security measures that serve the standard CIA triad tend to work against privacy in certain ways. As common examples, authentication, access control, logging, journaling, and public key infrastructure are all about introducing accountability into network transactions; ensuring that all actions can be traced to their originators; and leaving a detailed audit trail that can be used to investigate and prosecute security violations as well as outright breaches. In this sense, the standard practice of security stands in opposition to privacy. While certain security measures, such virtual private networks, may serve to keep transactions safe from unauthorized eavesdropping, the information technologists who implement and support these measures and the organizations that host them possess a full range of tools for associating data and transactions with the individuals who originate them. Anonymity, one of Westin’s key privacy facets, is intentionally discarded in support of accountability.

One further complication arises when examining the traditional CIA model in light of the goals of YouTube era Internet users. As currently construed, security techniques consider confidentiality and non-repudiation as universal goods. That is, a security technique works best when it prevents release of data to any unauthorized person, and when it unambiguously associates a transaction with the parties to the transaction. As noted above, however, many contemporary web tools focus on sharing data as broadly as possible. Additionally, the possibility of allowing the attribution of a writing or assertion to be generic rather than specific is intrinsic to many new collaborative web tools such as wikis. Conventional security methods are all or nothing – they are generally not designed for providing middle ground either on confidentiality or non-repudiation.
The Privacy-Security Paradox

If one accepts the argument above, then information security serves the needs of privacy only in a very limited and constrained set of circumstances. Top secret military and classified governmental information fit the standard security model well: There is a rigid distinction between insiders and outsiders; insiders must protect information and transactions from outsiders consistently across time and universally across types of transmissions and transactions; any loss of information to an outsider constitutes a failure of security and privacy. In contrast to the clean divisions in these examples, contemporary web tools often attempt to implement many shades of grey. Individuals who originate content (e.g., a MySpace page) usually have the highest degree of control over the visibility and accessibility of their content. These originators may have trusted peers with slightly less control. Below the trusted peers there may be registered users, unregistered users, and blocked users, all with progressively fewer and fewer privileges. Even these categories, however, reflect a conventional way of thinking about security: The intersection of privileges and user types forms the classic matrix model of access control used in many contemporary computer operating systems and applications.

Several problems arise from this matrix approach. First, it tends to be difficult to administer. The allocation of privileges is “blocky” rather than fine grained, but the more fine grained system designers make the privileges, the more tedious they are to administer. Assigning privileges also tends to be a manual process. Whenever something changes about the community of users, the body of content, or the configuration of the system, privileges need to be reassigned by hand. Second, the classes of users tend to be fixed at design time. The distinctions mentioned above – e.g., between registered and unregistered users – are hard coded into the design of the system. Key assumptions are also coded in, for example that there is one and only one type of registration process. Third, in designing security features, system designers focus primarily on metaphors of restriction or protection. One common sentiment among security professionals is that all privileges should be turned off by default (e.g., when creating a new account). This is a sensible, risk averse idea, but it also clearly demonstrates a bias against access, revelation, and exposure. Finally, the matrix model is entirely deterministic: It is assumed that at any moment in time there is a known set of content, a known set of possible transactions on that content, and a known set of users. These sets can change in discreet increments – generally the content is the most malleable of the three – but the whole system is designed either to prevent or resist indeterminacy. While indeterminacy itself is not necessarily a virtue, making indeterminacy difficult or impossible also tends to restrict the dynamism of a system. Transformations of the system, other than those imagined and implemented by system designers, become difficult.

In summary, then, the privacy-security paradox revolves around the three key issues from Altman’s work that we enumerated in a previous section – dynamism, balance, and limited good:

1. Privacy is a dynamic, continuing process, whereas security, as conventionally implemented by contemporary information systems, is static.
2. Security generally favors restriction of privileges and access, whereas privacy focuses on maintaining a balance between revelation and secrecy.
3. The CIA+NR goals of security – and particularly confidentiality and non-repudiation – are considered universal goods, whereas the instrumental and expressive goals that people have for their public communications make privacy a limited good.

Resolving the Privacy-Security Paradox

In reference to the discussion above it is important to note that there is nothing wrong with conventional perspectives on security and privacy per se. Security, as conventionally construed, serves the needs of a classic design for computers and systems that have existed for a long time and will likely persist into the future. Likewise, privacy is a human construct that focuses on the balance between revelation and secrecy – a balance that is driven by human needs rather than technological design goals. Security and privacy – each taken alone – do what they are supposed to do. When security and privacy have met, however, particularly in the context of newly emergent Internet applications, we have encountered problems attributable to the privacy-security paradox described above.

One possible resolution of the privacy-security paradox may come from developing new frameworks for information distribution. Current strategies for information distribution are closely tied to the modality of presentation as well as the commercial interests of content owners. For example, companies that make audio and video players (either standalone devices or software applications for general purpose computers) have implemented a variety of digital rights management (DRM) systems. These DRM systems – which are typically proprietary to each platform – place restrictions on copying and optionally on the lifetime of the information. Generally speaking, no part of the information may be shared with others, modified, customized, or redacted by the individual who has licensed the content. Having purchased a license to the content, however, the licensee generally has the right and capability to watch or listen to the whole product repeatedly. Many media
users — some of the authors of this paper included — find aspects of commercial DRM systems annoying, but the commercial interests that demand them consider them an important bulwark against piracy.

Setting the annoyance factor aside, there are concepts from DRM that might beneficially address the privacy-security paradox. First, DRM systems pair the information content with the rules for distribution (e.g., the notion of a license to go along with your file before you may play it). Conceptually, this packaging helps the content originator to govern the revelation and secrecy of the information object. Second, the concept of information lifetime assists the content originator with the challenges of controlled revelation. One of the balancing acts that people face as they regulate their privacy comes from the knowledge that when personal information is released to another party, it becomes much more difficult to control its subsequent distribution. This factor explains why individuals are may reveal more intimate in a face to face conversation than in writing: The conversation is ephemeral and will eventually be forgotten by the listener whereas the written record may be retained, reproduced, and distributed indefinitely.

Current DRM systems break down, however, in their user models and in their treatment of the data. Generally content originators have their commercial rights protected, authorized licensees have limited access privileges, and everyone else is simply ignored. This is not a tenable approach, however, for the current generation of web-based social networking and content sharing applications. We need new DRM concepts and frameworks that allow for a more fluid conception of users; they must facilitate sharing, annotating, mashing, redacting, and redistribution of information while simultaneously protecting human privacy needs. Current DRM systems also work to preserve the integrity of original data stream, but for privacy purposes, it is sometimes advantageous to perturb the data by introducing controlled, but intentional inaccuracies. For example, in order to preserve privacy in data mining, a randomized perturbation technique can be employed where user’s private data is disguised from the data set such that the central data aggregator can not derive the true identity of the user, and yet conduct useful data mining and filtering. The data is changed in a way that only the broad range of data or scrambled data is known to protect the users’ privacy. Similarly, the location information of mobile users can be perturbed or hidden to partially obscure spatial location and paths. These capabilities could be built into the information objects whose privacy we are trying to protect.

In effect we need a new conception of the CIA security triad: Instead of confidentiality, integrity, and availability, we propose a new security-privacy goal set based on “Control of Intimacy and Anonymity.” An expanded and improved approach to security and privacy would focus on allowing content originators to build controls into information objects of all types that facilitate two interacting forms of control. An intimacy control would afford different degrees of revelation of the information object depending upon the status of the relationship between the content originator and the content audience. Optimally, this control would be adaptive in the sense that as the nature of the relationship changed between these two parties, the revelation of the information object would change correspondingly. In contrast, the anonymity control would afford different degrees of revelation of the authorship and origins of the information object, again depending on the status of the relationship between the content originator and the content audience. These two controls could also interact: An audience might be able to obtain greater access to the information object, but only assuming greater protection of the originator’s identity. Conversely, extensive knowledge of the originator of some content might proportionally limit the audience’s ability to manipulate or redistribute the object. This could address some of the current problems that currently exist with the secondary distribution of databases of personally identifiable information for commercial purposes. Of greatest importance, however, would be the provision of an interface for these controls that made their use intuitive and flexible.

Summary and Conclusions

We have tried to make a case that the advent of new types of social networking and content sharing tools on the Internet has exposed a rift between conventional models of information security and privacy — particularly when privacy is viewed as a dynamic balancing act between competing human needs for revelation and secrecy. The goals of conventional models of information security are not wholly compatible with the way that people actually want to reveal and protect information in social networking and content sharing situations. To resolve this situation we have proposed an extension of existing digital rights management frameworks to incorporate a more varied view of content audiences as well as a shift toward more fluid and flexible control over properties that matter to content originators whose interest is not primarily commercial. We have also advocated an explicit integration of the metaphors of intimacy and anonymity into digital rights management, as these concepts are native to ideas about privacy advanced by Goffman, Westin, Altman, and other scholars.

Challenging work remains for researchers who wish to advance these ideas. The temptation to reduce the social complexities of privacy into mathematical formalisms such as those used in current models of access control might easily lead us into another rigid set of security tools. To resist this temptation, it might be fruitful to begin with user studies, particularly among users of social networking and content sharing tools, to find out what problems they have as well as how they envision the solutions. We believe this is a preferable approach precisely because privacy is a human, rather than a technological, construct. As such, the current individual, social, institutional, and cultural aspects of privacy must be described before system designers can make good decisions about how simultaneously to expose and protect valuable information content.
References