
M. GRANGER MORGAN
ELAINE NEWTON

Protecting Public Anonymity

*The option to
preserve anonymity
will erode unless
designers of new
technologies and
government
policymakers
act now.*

People in the United States have long enjoyed an expectation of anonymity when traveling or performing everyday activities in most public places. For example, they expect not to be recognized, or to have their presence noted and recorded, when making automobile trips far from home, attending large public functions, or visiting a shopping center to make cash purchases. Except for the possibility that they might encounter an acquaintance or violate a law and be asked by legitimate authorities to produce identification, they expect to be able to preserve their anonymity.

A variety of technologies are bringing this situation rapidly to an end. Some of these technologies are responses to heightened concerns about security, but many are simply the natural, if unintended, consequence of swiftly evolving technological capabilities. The society depicted in the recent film *Minority*

Report, in which people are routinely recognized by name wherever they go—and presented with individually tailored advertising—might not be far in the future.

A society in which all people can be located and identified, and their activities and associations tracked in any public space, is not a free society. Such a society would be highly vulnerable to the abuse of power by private or public parties.

Professionals in information technology and the law, groups concerned with civil liberties, and members of the general public should work collectively to preserve and strengthen the concept of public anonymity and strengthen privacy rights.

Already it is impossible to board an airplane, and in many cases even to pay cash for a bus or train ticket, without producing a photo ID. Video systems capture license plates as automobiles enter or leave parking lots, or pass through toll plazas. Some new tires carry electronic transponders (RFID tags) that can be linked to the vehicle. Security cameras capture images of faces in thousands of public locations. The Federal Communications Commission now requires cell phone systems to be able to locate callers when they make emergency 911 calls. Some cell phone systems already have the ability to locate callers.

M. Granger Morgan (granger.morgan@andrew.cmu.edu) is head of the Department of Engineering and Public Policy (EPP) at Carnegie Mellon University, and Elaine Newton (enewton@cmu.edu) is a Ph.D. Candidate in EPP pursuing research in privacy and security policy.

Today, most people remain anonymous much of the time. The bus company's clerks often do not bother to enter passengers' names into their computers, or if they do, the computers do not routinely share that information with other parties. Analogue images of license plates, as well as the thousands of images of faces, are often not subjected to real-time computer processing, recognition, and cross comparison with other databases. But this pattern of benign neglect will likely disappear as advanced automation becomes cheap and ubiquitous.

Does it make a difference if the world knows if someone bought hair dye at the supermarket, flirted briefly with a stranger on the corner while waiting for the light to change, rented an X-rated video, or was denied credit for a car? Surely people can learn to live with such minor embarrassments. Indeed, such matters have been the topic of local gossip in small villages since human civilization began.

But many people in the United States moved out of small villages, or moved west, precisely to escape the strong social control that is inherent in settings in which almost any anonymous public action is impossible and everyone remembers peoples' pasts. If current trends in technology development continue, then everyone in the country soon might find themselves back in the equivalent of a small town. Constant public identification, almost anywhere on the planet, by governments, by firms that want to shape peoples' preferences, by commercial competitors or jealous lovers, might become the norm.

Although preserving a degree of public anonymity is desirable in order to minimize social embarrassment, and important in order to limit social and cultural control, there is a more fundamental reason to resist the erosion of public anonymity. Individuals may not care who knows where they go, who they talk to, or what they do. But, if powerful public or private parties can know where everybody goes, whom everybody talks to, and what everybody is doing, then that creates enormous social vulnerability and the potential for abusive social and political control. The nation's founding fathers adopted a system of government based on checks and balances, arguing that no one in positions of power should be trusted to always act in the public interest, and for the preservation of freedom and civil liberties. That concern remains equally valid today.

Problems at the mall

Many of the issues that arise in preserving public anonymity can be illustrated through the consideration of two innocuous everyday activities: visiting a shopping center and driving an automobile.

Who might want to know that an individual is in a shopping center, whom he or she is talking to, and what he or she is doing there? For starters, there is law enforcement and mall security. Law enforcement personnel want to detect illegal acts. They also might want to screen shoppers to identify wanted persons. Of course, neither of these functions requires that all the individuals who are observed be identified. The only requirement is that illegal acts be identified, after which the persons involved could be identified. Similarly, to screen for wanted persons, one need not identify all persons. One need only identify those who are on a "watch list." However, the default solution for many of the professionals who design surveillance and other information technology systems might be to identify everyone they can and then check those identities against various databases.

The terms "law enforcement" and "wanted" both require clarification. In the narrow legal sense, "wanted" means persons for whom there are outstanding arrest warrants. However, law enforcement personnel also might wish to track suspects in connection with active investigations. Beyond that, things get more complicated. If it were easy to do, then some law enforcement organizations might also want to track all persons with previous police records or all persons who have specific genetic, behavioral, religious, or cultural profiles that suggest they are more likely to engage in unlawful activities. National security authorities might want to screen public places for persons suspected of espionage, terrorist, or other activities, or screen for all persons of a particular national or ethnic origin.

The problem of defining the membership of legitimate watch lists for surveillance performed in public places is a legal question that should be worked out in legislatures and the courts. The key point for system designers is that depending on the design choices that are made, it can be either very hard or very easy for anyone with access to the system to cross legally established boundaries specifying who is to be identified or how the information is to be used.

Next, consider retailers in the shopping center.

Beyond preventing shoplifting and other criminal acts, most retailers would probably like to use surveillance data to perform market research. If their objective is to see what displays or product placements attract attention, then that function could be performed without identifying the individuals involved. If they also need demographics, even that information could be obtained without actually identifying subjects to the users of the system.

Of course, retailers also might want to identify individual shoppers, link their identity with databases on financial resources and previous buying patterns, develop real-time advertising or sales

proposals that are tailored to each individual, target them with focused audio or video messages, and send follow-up messages to the customer's personal digital assistant. Today, all of these uses would be legal in most jurisdictions. Whether it should be legal is a matter that legislatures and the courts should decide. We believe that it should not, because in our view the social vulnerabilities that such widespread identification would create far outweigh the private benefits to retailers, and perhaps occasionally to shoppers. If the law made it legal only for shoppers who opted in, then the system should be designed so that it identifies only those consenting participants and leaves all other shoppers unidentified.

Who else might want access to shopping center surveillance data? Many different parties come quickly to mind, including:

- **Politicians running for office.** "Hello Ms. Newton, I note that you are registered for the other party, but given my strong voting record in support of computer science research, I hope you'll vote for me next week."

- **Detective agencies tracking possibly unfaithful spouses or just trying to drum up business.** "Mrs. Morgan, this is the Ajax Detective Agency. We observed that your husband had lunch with a much younger woman named Elaine Newton at 12:15 p.m. last Thursday. We have a special on this week. Would

There are a number
of promising ways
to promote the
growth of effective
system design
standards without
resorting to
inflexible
government
regulation.

you like us to check them out?"

- **Criminals looking for houses that are empty.** "Hey Joe, we just got the entire Morgan family here in Northway Mall. Here's their address. Don't forget I get a 10 percent cut on the take."

- **Credit agencies and insurance companies.** "Dr. Morgan, we are raising the rate on your health insurance because you regularly order dessert in restaurants."

Once these hypothetical shoppers finish at the mall, they might drive home. The parking system, which used its automated license plate reader to identify their vehicle when they entered, notes when they leave. Along the way, the occupants' cell phones might be

tracked. Or manufacturer-installed transponders on the vehicle or in its tires may be read every few blocks and recorded in a traffic database. License plate readers at intersections may track the vehicle. If there is real-time information being telemetered off the vehicle, then many details about speed, location, vehicle and driver performance, and even occupants' status, such as sobriety, might be available. If the car has biometrically enabled ignition, the driver's identity also could be noted. This partial list of the types of information that could be readily obtained without the driver knowing about it—let alone agreeing to it—should suggest a companion list of outside parties who might want such information.

Solutions by design

Preserving a reasonable degree of public anonymity in the face of the rapid development of advanced technologies presents formidable challenges. To a large extent, success or failure will depend on hundreds of seemingly innocuous design choices made by the designers of many separate and independent systems that collectively determine the characteristics of products and systems.

A simple example will illustrate. For years, we have used a teaching case in Carnegie Mellon's Department of Engineering and Public Policy, in which graduate students are asked to assume that a basic

“smart car” system is about to be implemented. They are asked to consider whether the state should run a pilot study that would implement a number of advanced system functions, such as insurance rates that are based on actual driving patterns, “externality taxes” for congestion and air pollution, and a system for vehicle location in the event of accident or theft. We find that students immediately assume a system architecture that includes real-time telemetering of all vehicle data to some central data repository. Then they become deeply concerned about issues of civil liberty, invasion of privacy, and social control, and often go on to construct arguments that such applications should be banned.

It is often not until students have worked on the problem for several hours that someone finally stumbles on the insight that most of the difficulties they are concerned about result from the default assumptions they have made about the system’s architecture. If information about vehicle location and driving performance is not telemetered off vehicles on a real-time basis, but is instead kept on the vehicle, not as a time series but in the form of a set of simple integral measures (such as a histogram of speeds driven over the past six months), then insurance companies could access it twice a year with all the time resolution they need. If detailed records of who drove where and when are not created, then most of the civil liberty problems are eliminated. Many of the potential concerns raised by other system functions in this teaching case can also be largely or entirely eliminated through careful system design choices.

This example illustrates a fundamental insight. If system designers think carefully about the social consequences of alternative designs before they make their choices, then the potential for negative social consequences often can be dramatically reduced or eliminated.

We suggest a preliminary list of design principles that we believe should be used with systems that collect information about people in public places:

- Identify explicitly the functions that the system is intended to perform.
- Collect only as many measures as required to perform those functions.
- When possible, use measures that integrate information over space and time.
- Use measures that are commensurate with the

function and security level of the task.

- When possible, use technologies that preserve the anonymity of the subjects being observed.
- Avoid unnecessary centralization of information storage and processing.
- In distributed systems, store information as far out in the nodes of the systems as practical and limit access to that information to the performance of legitimate system functions.
- When possible, avoid making data available to system users in real time.
- When possible, avoid the use of record identifiers that could facilitate linking the measures collected to other data sets for purposes other than the performance of the legitimate system function.
- Minimize the sharing of data and share only to the extent that it is required to perform the system’s function.
- Retain data only as long as required for the performance of the function.
- When possible, offer affected parties the opportunity to opt in or out. Set default values for such choices so as to preserve public anonymity.
- Attempt to anticipate and design the system so as to guard against possible unintended system applications and unintended uses of data collected by the system.

We offer these principles as a first suggestion. Refining them will take time and a wide discussion among a variety of communities.

There are a number of tensions implicit in these principles. Perhaps most important is the issue of “function creep.” Once a system has been developed with a rich set of capabilities, inventive people often can find other important, beneficial but perhaps also pernicious ways to use it. It seems unlikely that the initial developers of global positioning systems imagined that the systems would be used to dispatch emergency crews in urban areas, coordinate the operation of trucking fleets, or facilitate precision landings by commercial airliners at rural airports. Similarly, Alexander Graham Bell and his early associates never anticipated caller ID, call forwarding, or automated telemarketing.

Beyond government regulation

Enumerating a set of socially desirable design principles is the easy part. The really difficult part is getting them implemented in a manner that appropri-

ately balances protection with the socially desirable adaptive evolution of system functions.

In the past, society has managed other risks, such as air pollution, through government regulation that assumes one of three broad forms: design standards that specify how a system should be designed (incorporate flue gas scrubbers); performance standards that specify what a system must accomplish (emissions per unit of output must be below a specified level); and market-based standards that require certain actions by the parties involved (polluters must buy a permit for every kilogram of pollution they want to emit).

In the information technology setting, however, each of these approaches has problems. Government design standards are a bad idea for many information technologies, as the standards would almost inevitably lag the current state of the art and thus could kill innovation. Performance standards are less problematic, but implemented narrowly they too would pose serious problems. It is important not to inhibit new socially useful applications that the regulators have never thought of or to force innovators to wait while regulators design a standard for their new product, thus giving competitors time to catch up. However, it might prove possible to devise performance standards that specify the need to comply with a general set of data security, privacy, and anonymity criteria without being specific about the details of the actual software product. Market-based standards are not relevant to most issues involving information technologies.

But all is not lost. There are a number of promising ways to promote the growth of effective system design standards without resorting to inflexible government regulation. Possibilities include:

Best professional practice. Professional societies could develop and promulgate a set of performance standards intended to protect public anonymity and privacy. System designers could be urged to use them as a matter of good professional practice, and

*To a large extent,
success or failure
will depend on
hundreds of
seemingly
innocuous design
choices made by the
designers of many
separate and
independent
systems.*

educational programs could incorporate them into their curricula.

Certification. If such a set of standards were developed to protect public anonymity and privacy, a certification system could then be developed to indicate firms and products that adhered to these standards. Firms might advertise that they comply.

Acquisition specification. Public and private parties could require firm and product certification as a prerequisite to system acquisition, and they could publicize the fact that they impose this requirement.

Legal frameworks. Once they have proven their worth, best professional practice and certification standards can be incorporated into law. Care must be taken, however, to avoid moving too quickly. This

type of evolution has taken place in other domains, such as health and safety. New frameworks should include more detailed limitations on what data can be collected and how and under what conditions the data can be shared (for example, via a trusted third party who can negotiate or broker information exchanges).

Tort and liability. If laws were passed that limit the extent and circumstances under which persons and their actions could be identified via automated systems in public places, and this information shared with others, then this would provide a basis for parties to sue system operators, providers, and designers when abuses occurred. That, in turn, would create a strong incentive on the part of designers to design systems in which abuse was difficult or impossible. The use of a certified product might be made at least a partial defense, requiring, for example, a higher standard of proof of abuse.

Insurance. If firms that supply or provide systems and services are potentially liable for inadequate designs and are subject to liability, then insurance companies will have a strong incentive to require firms to demonstrate that their systems have been designed or acquired in accordance with appropriate standards.

Taxes or fees on uncertified systems. If a system does not conform to accepted design practice,

then state or federal law might impose a fee sufficient to make inappropriate uses economically unattractive.

Widespread adoption of best professional practice and certification standards should, over time, help to create a culture in which system designers routinely think about issues of anonymity and security as they develop systems. For example, in interviews we recently conducted with staff at several data-protection authorities in Europe, we were told that designers in Europe routinely consider issues of privacy when they choose what information to collect and how it will be used. Presumably, this is a consequence of the fact that most countries in the European Union have long had commissioners—and brought in outside auditors—who regularly enforce strict principles of data protection.

Anonymity through technology

Just as advancing technologies are creating headaches, they also can assist in preserving public anonymity. One example involves video surveillance. Latanya Sweeney, a computer scientist at Carnegie Mellon, has defined several formal models to protect databases before they are shared. One such model is known as *k*-anonymity. In this case, each individual record is minimally generalized so that it indistinctly maps to at least *k* individuals. The approach can be extended to images of faces. The magnitude of the parameter *k* can be defined by policy.

Before sharing video surveillance data, the algorithm could be used to “generalize” the images, and each one could be cryptographically locked and labeled. This imagery could then be shared with interested authorities who wish to search for criminal activity but do not have a warrant or know who specifically they are looking for. Once an event is witnessed, a warrant could be requested to restore only the perpetrator’s face to the original. Similarly, if security authorities want to search the images to look for people on a watch list, a cryptographic key could be used to unlock and restore only those faces that match faces on the list, thus assuring anonymity for the general public. In short, anonymity provided by technology need not always be absolute. In at least some circumstances, one would want designs in which, with probable cause and proper legal oversight, anonymity features could be selectively overridden.

A second example uses public key-based digi-

tal certificates. Suppose that after performing a thorough background check, a security authority is prepared to certify Alice as a low-risk domestic airline traveler who requires minimal screening before flights. The following scenario might then unfold, based on ideas first proposed by DigiCash’s David Chaum. The security authority issues to Alice a smart card that incorporates a certificate that she is a low-risk traveler. The card does not carry her name, but rather includes some encoded biometric identifier which can be used to authenticate that the certificate refers to Alice. The certificate is signed and sealed by the security authority with its own secret key and can then be authenticated by anyone using the associated public key.

Alice books and purchases her ticket using anonymous digital cash. The fact that she purchased the ticket is encoded on her smart card. At the airport, the airline’s ticket kiosk confirms that Alice is the person who bought a ticket and issues a boarding pass. Security authorities then verify that Alice is a low-risk passenger requiring minimal screening by using the security authority’s public key to confirm the authenticity of her certificate and comparing her biometric with the one encoded on the smart card. The system also could be designed to check whether Alice has been added to a watch list since the certificate was issued. Such certificates would have to be time stamped and periodically refreshed to assure that the approval remains valid. Tools exist to perform all of these functions without ever creating a record of Alice’s trip either at the airline or with the security screeners.

Updating the law

At the same time that technologists should be paying early and careful attention to their designs, government policymakers also need to be taking action. The nation’s legal system is woefully deficient with respect to anonymity.

The Privacy Act of 1974 requires that federal agencies collecting and maintaining personal records on citizens be responsible for safeguarding the data they collect to prevent misuse. The act, which applies only to federal agencies, builds on a set of fair information practices developed in 1973 by a high-level commission established by the U.S. Department of Health, Education, and Welfare (HEW). The core principles are: There must be no personal-data record-

keeping systems whose very existence is secret; there must be a way for an individual to find out what information about him is in a record and how it is used; there must be a way for an individual to prevent information about him obtained for one purpose from being used or made available for other purposes without his consent; there must be a way for an individual to correct or amend a record of identifiable information about him; and any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuse of the data.

The act directs the Office of Management and Budget (OMB) to "prescribe guidelines and regulations" that federal agencies should use but gives OMB no oversight authority. Currently only one OMB staff member has full-time responsibility and a few others have part-time responsibility for this mandate across the entire federal government. Individuals who believe that their personal records are being collected or used improperly have only one recourse: file suit in federal court against the alleged violator. To win, an individual must prove that actual harm resulted from the violation. The Privacy Act contains blanket exemptions in areas such as national security, law enforcement, and personnel matters.

The situation in the United States stands in marked contrast to how privacy issues are handled in Europe, Canada, Australia, and several other industrialized countries, where privacy rules also apply to private entities. The irony is that the laws and policies in those countries drew heavily on the basic principles developed by the HEW commission. Government authorities charged with protecting data have the authority to audit companies' data practices, have staff with the technical background needed to inspect computer systems and databases during audits, have the power to pursue civil and criminal penalties, and have the power to issue injunctions to temporarily suspend activities thought to be in violation. Although European Union (EU) countries of course have exemptions for national security, many EU data protection offices have a voice in how such exemptions should operate.

Although the U.S. Privacy Act does not cover businesses and other nongovernmental organizations, several federal laws passed since 1974 do provide piecemeal regulation of a few specific sectors of the

economy. For example, the Cable Communications Policy Act of 1984 extends some protection to cable subscribers; the Video Privacy Protection Act of 1988 protects video rental records; the Financial Modernization Act of 1999 addresses consumers' financial data; and the Health Insurance Portability and Accountability Act of 1996 protects health data (but not in the case of health research).

Some of these protections have been at least temporarily rolled back under the USA PATRIOT Act of 2001, which, for example, eases access to personal records by lowering the requirements for search and seizure and expanding the list of record types that can be collected. It allows for searches to be done in secret. Proponents argue that such rollbacks are needed to provide more authority to the attorney general in the fight against terrorism. Clearly in today's world total anonymity and total privacy are not viable. We need legal systems and institutions that balance anonymity and privacy against other legitimate social objectives, assuring that the balance can not be abrogated without adequate legal oversight.

Many states provide more privacy protection than is available under federal law. For example, citizens of California have a constitutional right to privacy. California adds additional protections through legislation, including a recent law (currently being tested in the courts) that protects consumers' financial information. State laws cover many different areas of privacy, including aspects of education, health and genetic information, social security numbers and identity theft, and employee rights. For example, an employer in Connecticut who wishes to monitor employees electronically must provide written notice.

There is evidence that some states are beginning to recognize that surveillance technologies, left unchecked, can encroach on Fourth Amendment protection from search. For example, Texas recently passed a privacy law concerning biometrics, and New Jersey is considering a similar law. In Virginia, legislation proposed in 2004, which passed the House but was tabled in the Senate, would have heavily restricted police use of face recognition technology and video surveillance data.

Current U.S. privacy law does not cover anonymity. However, there are a limited number of court decisions that have begun to address aspects of the vulnerability that might result from the loss of

anonymity. For example, the New Hampshire State Supreme Court recently ruled that companies that collect, combine, and sell data from many sources could be found liable for the harms that result from selling personal information.

Time to act

Rather than responding incrementally to specific problems posed by specific technologies, the United States needs to develop a principled, systematic legal approach to the problems of privacy and anonymity. With this in mind, we believe the time has come to convene a new high-level commission, similar to the HEW panel that laid the foundation for the Privacy Act. This commission should review and evaluate federal and state laws, as well as foreign laws, on privacy and anonymity, and systematically examine the potential effects that current and projected information technologies may have on these matters. The overarching goals should be to refine design guidance of the sort we have outlined; articulate a vision of how best to balance conflicting legitimate social objectives, such as law enforcement and national security, that impact anonymity and privacy; explore the problems of making a transition from the current minimally controlled environment; and develop a set of guidelines that could form the basis of a new set of legislative initiatives by Congress.

It would be best if such a panel were convened as a presidential commission. Alternatively, an executive branch agency could convene such a group. Still another possibility would be for a major private foundation to take on the job with a panel of extremely high-profile participants.

In parallel with the activities of such a panel, the community of information technology professionals needs to develop and disseminate a set of best professional practices for system design that protect public anonymity and privacy. There are several ways in which this could be undertaken. Individual professional societies such as the Association for Computing Machinery, the Institute of Electrical and Electronics Engineers, and the American Association for the Advancement of Science might launch the effort, they might undertake it jointly, or public or private funding might be used to mount an effort through the NRC.

Today, information technologies and systems are being developed with too little consideration of how de-

sign choices will effect anonymity and privacy and how, in turn, that might create more general social vulnerabilities. Unless all parties—in technology, in policy, in law, and in the wider society—join together in seeking creative new solutions, the list of unwelcome “watchers” and the risks of systematic abuse will likely grow. Society cannot afford to take that chance.

Recommended reading

- Advisory Committee on Automated Personal Data Systems, “Records, Computers, and the Rights of Citizens,” DHEW Publication No. (OS) 73-94, Washington, DC: U.S. Department of Health, Education and Welfare, 1973.
- D. Chaum, “Security without Identification: Transaction systems to make Big Brother obsolete,” *Communications of the ACM*, 28 (October 1985): 1030–1044.
- L. Lessig, *Code and other Laws of Cyber Space* (New York: Basic Books, 1999).
- E.M. Newton, L. Sweeney, and B. Malin, “Preserving Privacy by De-identifying Facial Images,” *Transactions on Knowledge and Data Engineering*, in press.
- The Privacy Act of 1974, codified at 5 U.S.C. § 552a; Computer Matching and Privacy Protection Act of 1988 (an amendment to the original 1974 Privacy Act), codified at 5 U.S.C.A. § 552a(o) et seq. (1988); Cable Communications Policy Act of 1984, codified at 47 U.S.C. § 521 et seq.; Video Privacy Protection Act of 1988, codified at 18 U.S.C.A. § 2710 (1988); The Gramm-Leach-Bliley Financial Modernization Act of 1999, Public Law 106-102, codified at 15 U.S.C. § 6801, et seq. and 16 C.F.R. 313, 65 Fed. Reg. 33646 (May 24, 2000); Health Insurance Portability and Accountability Act of 1996, Public Law 104-191 (1996), 45 CFR Part 160 and 164 (2002); Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT) Act of 2001. Public Law 107-56 (2001).
- OECD, *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, available at http://www.oecd.org/document/18/0,2340,en_2649_34255_1815186_1_1_1_1,00.html.
- A.F. Weston, *Privacy and Freedom* (New York: Atheneum, 1967).