Homeland Security

John Yen, *Pennsylvania State University*
Robert Popp, *DARPA/IXO*

Trends & Controversies this issue grows out of a panel discussion at the 2005 AAAI Spring Symposium on AI Technologies and Homeland Security, held at Stanford University in March 2005 ("Reports on the 2005 AAAI Spring Symposium Series," *AI Magazine*, vol. 26, no. 2, 2005, pp. 87–92). This department aims to facilitate the dialogue between policy makers and information security technology developers.

Robert Popp, who gave the keynote speech at the symposium, describes a DARPA initiative for dealing with the 21st-century strategic threat triad: failed states, global terrorism, and weapons of mass destruction proliferation. The new initiative explores innovative quantitative and computational social science methods and approaches that could enable commanders and analysts to understand and anticipate the preconditions that give rise to instability and conflict within weak and failing states. George Cybenko presents a philosophical/strategic viewpoint on national security. He argues that solving the forward problem (model-based predictions) for analyzing adversarial organizations is critical because it can serve as the foundation for solving the inverse problem (understanding organizations on the basis of observables). K.A. Taipale discusses policy implications of using trusted systems for counterterrorism security and how risk management, decision heuristics, and the presumption of innocence relate to such systems. Latanya Sweeney proposes privacy-aware technology (selective revelation) that allows data about people to be shared for surveillance purposes while protecting their privacy. Paul Rosenzweig points out two major changes in privacy protection in the post-9/11 era: the broadening of the approach to generating privacy policy/rules from a purely top-down process to one that includes a bottom-up component in which privacy is protected through institutional oversight, and a change from a focus on rules to a focus on results. The two changes together, he argues, suggest an iterative process in which the oversight institution evaluates technology's efficacy from the perspective of its target results, which might generate further policy changes.

These five articles present a snapshot of the complex interactions between information security and privacy. A comprehensive understanding of such interactions is critical for developing solutions, whether they are technological solutions, political solutions, or both.

—John Yen and Robert Popp

Using Social Science Technology to Understand and Counter the 21st-Century Strategic Threat

Robert Popp, *DARPA/IXO*

During the Cold War era, the strategic threat against the US was clear. The country responded clearly with a policy toward the Soviet threat that centered on deterrence, containment, and mutually assured destruction. To enforce this policy, the US created a strategic triad composed of nuclear intercontinental ballistic missiles, Trident nuclear submarines, and long-range strategic bombers.

Today, however, our security environment is profoundly different. The strategic threat is far more complicated and dynamic. New and deadly challenges—from irregular adversaries to catastrophic weapons to rogue states—have emerged. The 21st-century strategic threat triad, made up of failed states, global terrorism, and WMD proliferation, represents the greatest modern-day strategic threat to our national security interests (see figure 1).

With this new strategic triad's emergence comes the need to craft a new agenda of military and national security priorities. Winning the war against these new threats will require more than victory on the battlefield.

Recently, the US government published a revised national security strategy.[1] It charters our military to reassure our allies and friends, to dissuade future military competition from would-be aggressors, to deter threats against US interests, and to decisively defeat any adversary if preemption and deterrence fail.

To execute the new strategy, it's vital that our military seek to deeply understand these new strategic threats. It's not sufficient to predict where we might fight next and how a future conflict might unfold. We can no longer simply prepare for wars we would prefer not to fight; we must now prepare for those we will need to fight. Our new strategy requires that we make every effort to prevent hostilities and disagreements from developing into full-scale armed confrontations. This, in turn, requires applying political, military, diplomatic, economic, and numerous

1541-1672/05/$20.00 © 2005 IEEE
Published by the IEEE Computer Society

other social options to gain the necessary understanding of potential adversaries' cultures and motivations. Indeed, we must be able to shape entire societies' attitudes and opinions, with predictable outcomes.

Challenges of 21st-century warfare

Recent experience in Iraq and Afghanistan has taught us that military success in post-conflict stability operations requires a deep social awareness of the threat and of the operational environments in which they exist. In fact, successfully managing stability and reconstruction operations has required as much social awareness as military combat savvy.

In this century, our adversaries seek to paralyze US influence by employing unconventional methods and weapons of mass destruction. These new adversaries are asymmetric, transnational terrorists, insurgents, criminals, warlords, smugglers, drug syndicates, and rogue WMD proliferators. They're indistinguishable from and intermingled with the local civilian population. They're not part of an organized, conventional military force, but have formed highly adaptive organizational webs based on tribal or religious affinities. These new adversaries conduct quasimilitary operations using instruments of legitimate activity found in any open or modern society. They make extensive use of the Internet, cell phones, the press, schools, houses of worship, hospitals, commercial vehicles, and financial systems. They don't respect the Geneva Conventions or the time-honored rules of war. They see WMD not as a weapon of last-resort but instead as an equalizer and a weapon of choice. They perpetuate religious radicalism, violence, hatred, and chaos. And, finally, they seek safe haven and harbor in weak, failing, and failed states.

What do I mean by failed states? Failed states facilitate the routine brutalization and repression of their own people. They reject basic human values and are less concerned with international order than with lawlessness, demagoguery, hatemongering, and thuggery. Failed states are internally divided along ethnic, religious, and ideological lines. They're ruled by thugs who act not in the interests of their citizenry, but to settle scores and retaliate against perceived humiliations. Failed states, like the threats they harbor, see the acquisition of WMD technology as empowering and essential to their prestige on the world stage. They provide breeding grounds for terrorists, the narcotics trade, black marketeering, human slavery, weapons trafficking, and other forms of organized crime. Their populations suffer in a climate of fear, institutional deterioration, social deprivation, and economic despair.[2] In today's increasingly interconnected world, they pose an acute risk to US national security.

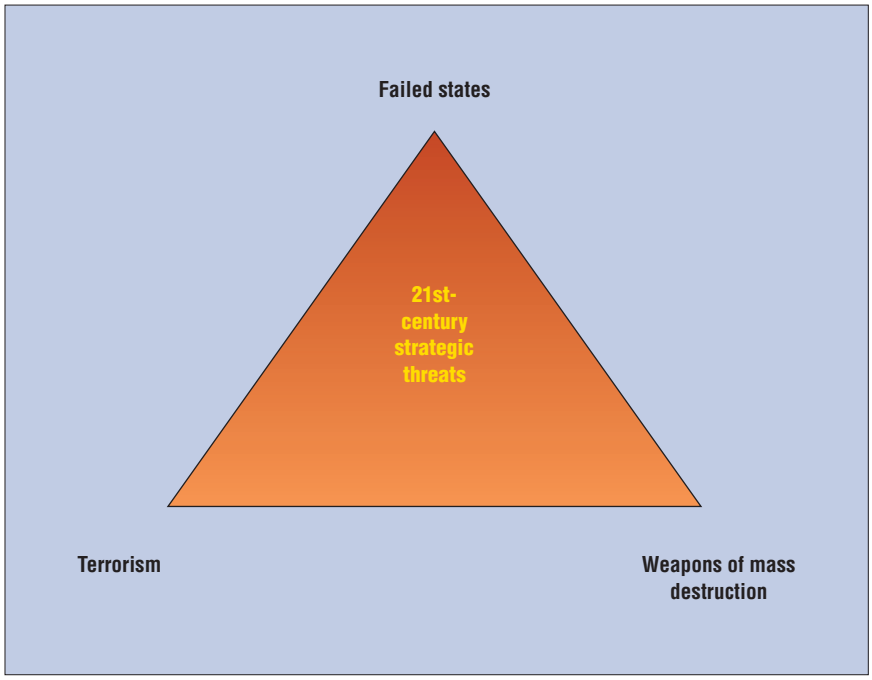


Figure 1. 21st-century strategic threats against the US.

The ballistic missiles and conventional intelligence, surveillance, and reconnaissance systems that were so effective at ending the Cold War are no longer sufficient, nor are they well suited to countering the new 21st-century strategic threats. These new threats—willing to accept almost any degree of risk to achieve their objectives, often under the false pretext of religion—are able to foil our conventional surveillance systems.

In many instances, the decisive terrain in 21st-century warfighting is the vast majority of noncombatants whose support, willing or coerced, is critical to influence. Winning over the local population's hearts and minds by providing aid to improve their lives is as important as, and can no longer be subordinated to, projecting military force or capturing and killing the enemy.

How are we to implement this new national strategy? I believe the way forward is clear. It doesn't involve spending hundreds of billions of dollars procuring more Cold War-oriented, conventional ISR or high-profile weapon systems to gain incremental improvements in precision, speed, or bandwidth. We need a strategy that leads to greater cultural awareness and thorough social understanding of the new strategic triad threats. A commander from the Third Infantry Division explained this need brilliantly while commenting on his march to Baghdad: "I knew where every enemy tank was dug in on the outskirts of Tallil. … Only problem was, my soldiers had to fight fanatics charging on foot or in pickups and firing AK-47s. … I had perfect situational awareness. What I lacked was cultural awareness. Great technical intelligence … wrong enemy."[3]

How social science technology can help

What technologies must we develop to understand and influence nation states, societies, thugs, terrorists, WMD proliferators, and zealots in failed states?

I believe the path to understanding people and their cultures, motivations, intentions, opinions, and perceptions lies in applying interdisciplinary quantitative and computational social science methods from mathematics, statistics, economics, political science, cultural anthropology, sociol-

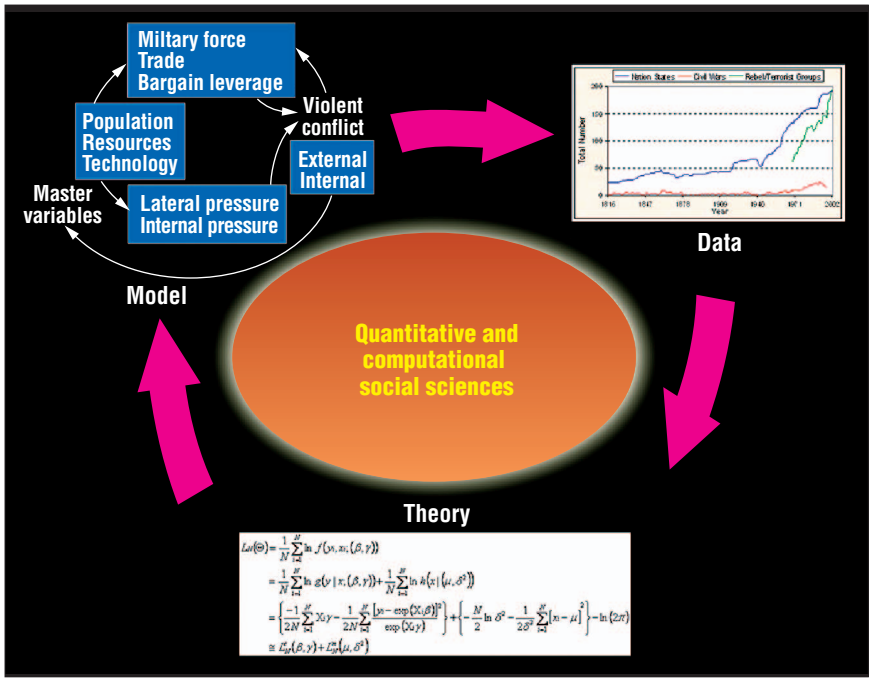


Figure 2. Quantitative and computational social science methods can help the US understand and influence strategic threats.

ogy, neuroscience, and modeling and simulation (see figure 2).

Understanding and countering today's inherently dynamic, socially complex strategic threat isn't easily reduced or amenable to classical analytical methods. It requires applying quantitative and computational social sciences that offer a wide range of nonlinear mathematical and nondeterministic computational theories and models for investigating human social phenomena. These analytic techniques apply to cognition and decision making. They make forecasts about conflict and cooperation at all levels of data aggregation, from the individual to groups, tribes, societies, nation states, and the globe. They use dynamic systems equations and models: models of reactions to external influences, models of reactions to deliberate actions, and stochastic models that inject uncertainties.

Research in these areas is vital. We need good models, good theories, and good tools to apply these technologies. These tools are as critical as any new weapon system. They're central to our war against the new strategic triad threats.

Military commanders need means for detecting and anticipating long-term strategic instability. They have to get and stay ahead of conflicts within and between nation states. In establishing or maintaining security in a region, cooperation and planning by the Regional Combatant Commander is vital. It requires analysis of long-term strategic objectives in partnership with the regional nation states. It also requires a careful balance of finite resources such as humanitarian relief, political and economic outreach projects, infrastructure rebuilding projects, joint military training and exercises, and, when needed, military combat operations.

Innovative tools from the quantitative and computational social sciences will enable military commanders to prevent conflict and manage its aftermath. These tools will allow a greater understanding of the complex political, military, economic, sociological, and demographic landscape associated with nation states. They can also predict the loads and demands placed on the state as a function of its capability to manage the stresses. They will allow alternative shaping options to be generated and evaluated in cost-benefit terms for their ability to mitigate destabilizing events, enhance peacekeeping measures, and influence choices about economies, political systems, rule of law, and internal security.

Because the analysis of conflict and nation state instability is inherently complex and deeply uncertain, no one social science theory or quantitative/computational model is sufficient. An ensemble of models containing more information than any single model must be integrated within a single decision-support framework to generate a range of plausible futures. Robust adaptive strategies—suboptimal ones—that hedge across these plausible futures will provide practical options for the decision-maker to consider.[4] Within the right theoretical framework, these models and decision support tools will provide strategic early-warning capability and actionable options for winning peace, preserving stability, and minimizing deadly conflict.

Quantitative and computational social science has already begun to show promise toward understanding nation states. At DARPA, we've been funding research to model and understand the preconditions that give rise to nation state instability and conflict.[5] In any field of science, the best work is that with the strongest empirical support and explanatory power. This field is no different.

For example, one model using system dynamics has successfully explained how internal and external state pressures can lead to violent conflict (see figure 3). It shows the often unexpected long-term consequences and tipping points that different strategies toward conflict or instability mitigation can have on a nation. Another model involving cellular automata has shown how simple microlevel grievances or preferences from a small number of actors can diffuse and spread in counterintuitive ways. Again, we see surprising macrolevel outcomes. For example, in Schelling's segregation model, even moderately tolerant neighboring groups can produce significant ethnic segregation over time. Another model, based on geopolitical distributions, can show that spatial dynamics, such as the spread of conflict, can differ depending on the scale invariance of subpopulation distributions as defined by political, ethnic, religious, or economic features. These and other theories and modeling paradigms from the quantitative and computational social sciences are making powerful contributions to our understanding of the 21st-century strategic triad threats and to improved policy solutions that can provide strategic and tactical advantages.

Victory in the 21st-century strategic threat environment no longer belongs to the side that owns the best and most sophisticated ISR or weapon systems. It belongs to the side that can combine these cutting-

edge technological marvels with methods, models, and technologies from the quantitative and computational social sciences.

References

1. *The National Security Strategy of the United States of America*, Sept. 2002, www.whitehouse.gov/nsc/nss.pdf.
2. R. Rotberg, "The Failure and Collapse of Nation-States: Breakdown, Prevention, and Repair," *When States Fail: Causes and Consequences*, R. Rotberg, ed., Princeton Univ. Press, 2004.
3. R.H. Scales Jr., "Culture-Centric Warfare," *U.S. Naval Institute Proc.*, Sept. 2004, pp. 32–36.
4. S.W. Popper, R.J. Lempert, and S.C. Bankes, "Shaping the Future," *Scientific American*, Apr. 2005, pp. 66–71.
5. R.L. Popp, "Exploiting AI, Information and Computational Social Science Technology to Understand the Adversary," *Proc. AAAI Spring Symp. AI Technologies for Homeland Security*, AAAI Press, 2005.

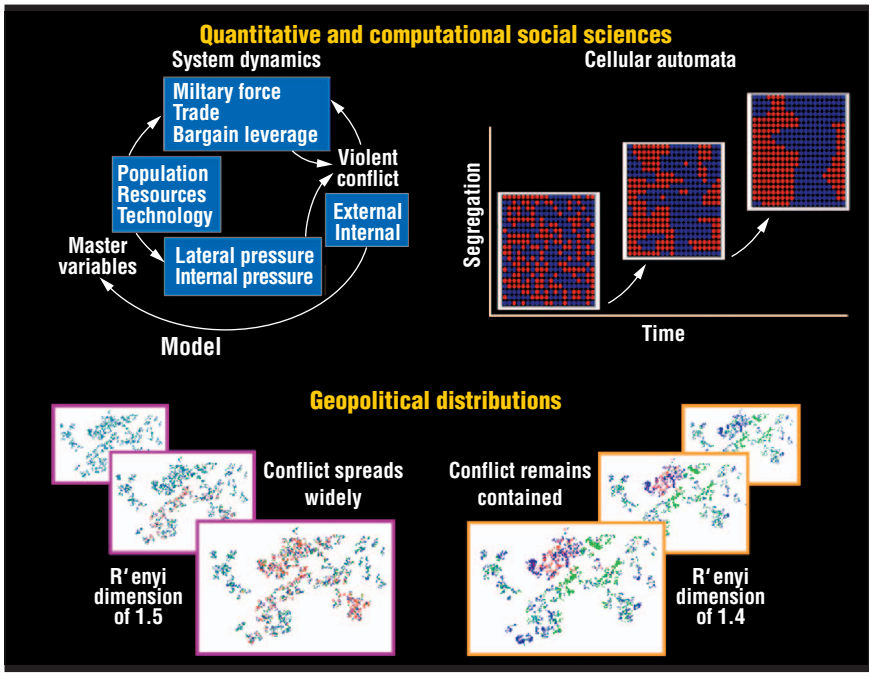


Figure 3. Applying various quantitative and computational social science models to understand nation-state instability has shown great early promise.

AI and the Modern Networked Organization

George Cybenko, *Dartmouth College*

One characteristic of a science is its ability to predict outcomes (deduction) and explain observations (abduction). In many areas of science and engineering, deduction and abduction are manifest as solving forward and inverse problems.[1] Using Maxwell's electromagnetic equations to predict an aircraft's radar image is an example of solving a forward problem. The corresponding inverse problem is automatic target recognition—determining the object that was responsible for an observed radar image. Similar forward and inverse problems arise in speech; speech generation is the forward problem, and speech recognition is the inverse problem.

Let's consider some forward and inverse problems associated with organizations in the context of modern strategic threats. I'm using the word organization in a broad sense, including business, government, military, social, and political enterprises. For this essay's purposes, an organization is a collection of people working toward some common goal. It includes government organizations such as military services and intelligence agencies as well as organized adversarial insurgents and terrorist networks.

Engineering and computer science research hasn't addressed social organizational problems much until relatively recently, and even then the focus has been on selected subsets of the overall problem space. For example, researchers have looked at social networks and the small-worlds phenomenon over the past five years or so.[2] Research on information and decision markets, multiagent systems, and workflow management also represents quantified, computational approaches to investigating organizational theory and practice.[3]

But the problem space is large, and much work remains. AI has traditionally strived to understand the intelligence of individual humans and to harness that understanding in a computational system. I believe that a major research goal over the next few decades will be to understand effective organizations' behaviors more scientifically and then harness that understanding in computational systems.

The forward problem: Designing better organizational structures

Let's consider a few general problems as candidate applications of such scientific foundations. A forward problem in computational organizational science would be to deduce an organization's behaviors and effectiveness using specific information technologies as the means for coordination and information sharing. For example, in the context of eBay auctions, the forward problem would be to determine how effective specific eBay auction mechanisms are in achieving the goal of large-scale, fair, efficient commerce. The goal is to determine this using solely analysis and simulation, not to actually implement and observe the resulting system. In the context of modern strategic threats, an analogous forward problem would be determining how effectively a given organizational and technological structure would solve homeland and national security intelligence problems such as information sharing and analysis.

The current approach to solving forward problems is largely ad hoc and based almost solely on a combination of case studies, previous operational experiences, and educated guesswork. Maybe we can learn something about organizational design from established engineering disciplines, such as bridge design.

Henry Petroski has observed that about every 50 years, there is a catastrophic bridge failure in which the failure is a direct result of pushing a prevailing design paradigm outside the envelope of its applicability.[4] Put another way, over time, bridge designers will push their quantitative model too

far. When designers don't trust their models, they overdesign, with the result that many classical civil engineering projects remain standing today. When they trust their models, they push the design envelope as far as possible, leaving little margin for error or unanticipated operating modes. This can lead to catastrophic failures such as the dramatic collapse of the Tacoma Narrows Bridge in the 1950s.

Many modern organizations are designed to be efficient, minimizing the cost of operation while maintaining some required level of performance. Consequently, we should expect modern organizations, including new companies, social organizations, and government agencies, to collapse with some regularity when subjected to natural but unmodeled perturbations, while older, overdesigned organizations will survive similar perturbations. Consider, for example, the Catholic Church's resilience compared with Enron's.

Any quantitative and computational science for solving forward problems in organization design must take into account the cost/robustness trade-off, especially with respect to critical government services that are meant to counter 21st-century strategic threats. Designing agencies and organizations to be optimally efficient, cost effective, and minimal according to some current theory or model will likely result in catastrophic failures caused by unmodeled phenomena.

We need quantitative and computational social science to help us design better organizational structures, given today's information technology possibilities. But, we should be prepared for catastrophic failures unless we're careful not to overly optimize within the necessarily incomplete modeling paradigm.

The inverse problem: Inferring organizational structure from observation

Inverse problems that arise in computational and qualitative social sciences, and that are highly relevant to modern strategic threats, involve inferring an organization's structure, processes, and goals from indirect observations of the organization's activity. In more concrete terms, the inverse problem is figuring out who the bad guys are and what they're doing on the basis of bits and pieces of observable information.[5]

In many areas of traditional science and engineering, inverse problems are much harder than the corresponding forward problems. For example, predicting an object's radar image is easier than inspecting a radar image and determining what kind of object produced it. In many applications, you solve the inverse problem by repeated applications of the forward problem. That is, you produce a comprehensive list of forward problem input-output relations and then solve the inverse problem using a lookup table type of approach—comparing the observed output with the list's computed outputs, thereby identifying candidate inputs.

It seems likely that solving inverse problems associated with adversarial organizations will involve repeated applications of the forward problem, as is common in other inverse problem areas such as radar imaging. We won't be able to solve the inverse problem until we have good ideas for solving the forward problem: Given a group of people, a common goal, and a collaboration technology, how will they collaborate and with what effectiveness? Solving the forward problem is a key step toward building efficient and effective organizations that are robust and survivable. We should explore this on its own merits, regardless of the inverse problem's importance.

Given a group of people, a common goal, and a collaboration technology, how will they collaborate and with what effectiveness?

I sincerely believe that the scientific study of organization design and behavior will be a major topic of computer science and engineering research over the next 20 years, especially the forward and inverse types of problems that I've outlined here. We need to achieve the kind of progress made in AI over the past 50 years at the next level of the biological hierarchy: the organization.

References

1. P. Neittaanmaki, M. Rudnicki, and A. Savini, *Inverse Problems and Optimal Design in Electricity and Magnetism*, Oxford Univ. Press, 1996.
2. D.J. Watts, *Six Degrees: The Science of a Connected Age*, W.W. Norton, 2004.
3. J. Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*, Doubleday, 2004.
4. H. Petroski, *To Engineer Is Human: The Role of Failure in Successful Design*, Vintage Books, 1992.
5. J.S. Brown and J.R. Cooper, "Intelligence: We've Lost Our Edge," *Washington Post*, 10 May 2005, p. A21.

The Trusted Systems Problem: Security Envelopes, Statistical Threat Analysis, and the Presumption of Innocence

K.A. Taipale, *Center for Advanced Studies in Science and Technology Policy*

> "We need to have a world that is banded with security envelopes, meaning secure environments through which people and cargo can … [with the proper vetting and tracking] move relatively freely from point to point all across the globe with the understanding that those within the security envelope we have a high confidence and trust about so that they don't have to be stopped at every point mechanically and re-vetted and rechecked. And those outside the envelope would be those on which we could focus our resources … to make sure bad people can't come in to do bad things."
>
> *—Secretary of Homeland Security Michael Chertoff (speaking at the Center for Strategic and International Studies, 19 May 2005)*

In response to the threat of potentially catastrophic attacks, governments are under political pressure to preempt terrorist acts. Preempting acts that can occur at any time or place requires optimally allocating limited security resources on the basis of predicted risk rather than perceived vulnerabilities. Security forces simply cannot guard all vulnerable targets at all times or recheck all people or objects at every stage of vulnerability during movement through open systems. Thus, governments are increasingly developing security strategies based on *trusted systems*, exemplified by Secretary Chertoff's call for *security envelopes*. (Trusted systems for the purposes of this essay are systems in which some conditional prediction about the behavior

of people or objects within the system has been determined prior to authorizing access to system resources.)

This essay examines some policy implications of using a trusted-systems model for counterterrorism security. In particular, it discusses certain issues relating to how trusted systems function and fail, how risk management and decision heuristics interact with trusted systems, and how trusted systems relate to the presumption of innocence. This essay is not intended to be a definitive statement of these issues but rather an introductory offering for discussion.

Background

The adoption of preemptive strategies for counterterrorism has blurred the line between reactive law enforcement and preemptive national security methods previously governed by disparate—and often conflicting—doctrinal regimes. The use of advanced information technologies for data collection, aggregation, sharing, and analysis has exacerbated this blurring by allowing information to flow freely between these previously distinct governmental functions. One result has been an acceleration of modern societies' ongoing transformation from a notional Beccarian model of criminal justice based on punishment and deterrence of deviant individuals after they commit criminal acts[1] to a Foucauldian model of general social compliance through ubiquitous preventative surveillance and control through system constraints.[2]

In this emergent model, security services are geared not toward policing through arrest and prosecution but toward risk management through surveillance, information exchange, auditing, communication, and classification. These developments have led to general concerns about individual privacy and liberty—concerns that I've addressed in part elsewhere[3–5]—and to a broader philosophical debate about the appropriate forms of social-governance methodologies that is beyond the scope of this essay. Instead, this essay focuses more narrowly on identifying certain characteristics of the trusted systems compliance model.

Trusted systems: How they work, how they fail

Trusted systems generally depend on two kinds of security strategies—authorization and accountability—to ensure that rules governing behavior within a system are obeyed. Authorization is the process of constraining the terms under which a user can access a system or use its resources. Accountability is the process of associating responsibility to behavior of users or objects within the system.

Accountability strategies are not very effective against suicidal attackers (particularly those without patrons or support infrastructure subject to sanction). Thus, authorization strategies are necessary for keeping vital systems secure and functioning.

However, authorization strategies scale poorly and burden systems with high overhead (that is, they introduce frictions which inhibit functionality). Also, authorization strategies are difficult to manage centrally in complex heterogeneous systems (like global transport) and thus require a federated approach (one composed of trusted partners who reciprocally honor each others' grants and credentialing of authorization on the basis of some agreed minimum vetting standards). Federation, however, introduces a lowest-common-denominator risk—all partners are exposed to the least capable or competent partner's security practices.

While false negatives are a threat to security, false positives are a threat to system functionality because they introduce friction and reduce degrees of freedom.

But, more importantly, any system premised on separating unlikely threats from more likely threats on the basis of trust (that is, based on predictions of future behavior) is prone to two well-known failure modes: false negatives (type II errors in significance testing) and false positives (type I errors). False negatives are people classified as unlikely threats who actually are threats (for example, terrorists wrongly cleared for access despite vetting). False positives are those falsely identified as threats and wrongly denied authorization.

The potential for false negatives requires a layered defense—additional security strategies to supplement access control. Access control alone is a brittle strategy because any perimeter breach provides access to all system resources. Thus, firewalls alone are inadequate to protect technical systems and must be supplemented with code scanners and user monitoring. So, too, border controls are inadequate to protect homeland security and must be supplemented with internal controls such as passenger screening against particular vulnerabilities.

Likewise, systems based on security envelopes will still require some random rechecks within the trusted environment to counter potential false negatives. Furthermore, access authority itself should be limited (individuated to need), dynamic (subject to continuous updating based on new information), and technically easy to revoke or modify. System behavior can then be monitored for conformity to expectations and authorizations adjusted accordingly.

While false negatives are a threat to security, false positives are a threat to system functionality because they introduce friction and reduce degrees of freedom. In addition to true false positives (those wrongly excluded), trusted systems engender another category with similar problems—nonthreats who have not been or cannot be cleared for access because of resource constraints. For example, new market entrants might not have the resources to meet vetting standards or the system might not have sufficient resources (or incentive) to vet all new entrants. If not appropriately accounted for in systems design, such friction can turn a trust-based security system into an unacceptable burden on functionality (or, in the case of security envelopes, into an instrument to consolidate hegemonic, regional, or local trade power).

The ratio of false negatives to false positives in a trusted system is a function of risk tolerance and the degree of certainty demanded in determining the conditional prediction of conforming behavior during the vetting process.

Risk management

Risk management uses decision tools to reduce the probability of negative outcomes within the available resource constraints and the particular risk tolerance. As a practice, it requires continuously assessing and updating risks, determining which risks are most important to address, and implement-

ing strategies to mitigate those risks.

In the context of potentially catastrophic outcomes, the political risk tolerance for false negatives is low. Thus, decision heuristics for counterterrorism policy, including confidence requirements, will bias toward reducing false negatives. Systems design should therefore anticipate a higher false-positive rate and build in adequate compensation mechanisms to manage these. Among other things, this requires ensuring that adequate security resources are available and not overwhelmed (more of a concern with systems designed to isolate suspects than those intended to establish trust) and that vetting or redress mechanisms are not so onerous that they impede functionality. Designing procedures to mitigate potential harms from false positives seems preferable to engaging in recriminations over harms resulting from false negatives.

But risk management has its roots in insurance practice, not security, and the limits in its methodology must be recognized. For example, classical actuarial methods for determining probabilities based on measuring frequency of occurrence are generally not appropriate in the context of counterterrorism where the sample size of actual terrorists or terrorist acts is too small for high degrees of predictive certainty. Instead, a more dynamic view of probability is required.

Bayesian inference is a powerful statistical method for determining the degree of certainty in the truth of an uncertain proposition. In Bayesian systems, new information is constantly evaluated to update the degree of certainty in any particular proposition (to estimate its conditional probability). At any given decision point, a learned critical value of confidence exists above which the system acts as if the uncertain proposition was true, and below which it acts as if the proposition were false. That critical value—the point of significance for decision making—determines the ratio of false positives to false negatives and changes over time according to experience.

The salient point for trusted systems is that vetting and authorization should remain dynamic as well. Thus, authorizations based on investigation and vetting prior to access must be continuously updated with new information generated from actual behavior observed within systems (and other relevant new information). Behavior within trusted systems should be measured against both objective (peer group norms and expert models) and subjective (previous or typical) behavior patterns. Consider, for example, a trusted shipper within a security envelope whose typical pattern is to ship small objects from Europe to Asia. If the shipper suddenly consigns a large shipment from a failed former Soviet state to Washington, it should be flagged in real-time for additional screening regardless of previous vetting.

But, to some observers, using conditional probabilities to allocate security resources seems to counter certain presumptions, including that of innocence.

The presumption of innocence

Fully exposing the presumption of innocence, either as a matter of law or philosophy, is beyond the scope of this essay. Rather, a single narrow question is addressed: Does the use of statistical threat analysis in itself challenge the presumption of innocence?

Presumptions are legal fictions introduced to define the default state or null hypothesis (the presumption that an observation is only coincidence). In the context of criminal justice, the presumption of innocence defines the default state of the accused. The burden of proof then falls to the accuser to present evidence of sufficient weight to meet some level of legal significance—for example, "beyond a reasonable doubt"—at which point the presumption of innocence gives way to a finding of guilt without equivocation. If the burden of proof is not met, the accused remains by default presumed innocent regardless of whether, in true fact, they committed the act.

The analogy in classical statistics is testing the null hypothesis against a level of significance. If the test result is within the level of significance, then the null hypothesis is rejected. If not, the presumption of coincidence stands.

But the presumption of innocence is applicable beyond the narrow confines of criminal justice. In a sense, it defines the relationship between liberal state and individual, requiring the state to meet some threshold of suspicion (that is, some level of significance of the available evidence) before it can exercise any power over the individual (for example, reasonable suspicion to stop or probable cause to arrest). Critics of using probability-based trust systems in counterterrorism argue that probabilities are not particularized to the subject and thus cannot be the basis for (are not evidence of) trust or suspicion. Such a view is counterintuitive, as well as wrong under Supreme Court doctrine. As both the Court and logic would dictate, it is the probative value of the evidence, rather than its probabilistic nature, that is relevant in determining whether it is a sufficient predicate for government action. To argue otherwise is to confuse the presumption of innocence with the probability of innocence.

The importance of design

The threat of potential catastrophic outcomes from terrorist attacks raises difficult policy choices for a free society. Nevertheless, it is clear that we cannot "wait until after the bad guys pull the trigger before we [move to] stop them."[6] Using trusted systems to help allocate security resources on the basis of risk analysis and threat management may offer significant benefits with manageable harms if system designers take the potential for errors into account during development.

Of course, the more reliant we become on probability-based systems, the more likely we are to mistakenly believe in the truth of something that might turn out to be false. That wouldn't necessarily mean that the original conclusions were incorrect. Every decision in which complete information is unavailable requires balancing the cost of type II errors with those of type I. When mistakes are inevitable, prudent design criteria include the need for elegant failures.

References

1. C. Beccaria, *On Crimes and Punishment*, 1764.
2. M. Foucault, *Discipline and Punish: The Birth of the Prison*, 1975.
3. K.A. Taipale, "Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data," *Columbia Science and Technology Law Rev.*, vol. 5, no. 2, 2003; http://ssrn.com/abstract=546782.
4. K.A. Taipale, "Technology, Security, and Privacy: The Fear of Frankenstein, the Mythology of Privacy, and the Lessons of King Ludd," *Yale J. Law and Technology*, vol. 7, 2004, pp. 123–201; http://ssrn.com/abstract=601421.
5. K.A. Taipale, "Designing Technical Systems to Support Policy: Enterprise Architecture, Policy Appliances, and Civil Liberties," *21st Century Information Technologies and En-*

abling Policies for Counter-Terrorism, R. Popp and J. Yen, eds., Wiley–IEEE Press, 2005.

6. "The Limits of Hindsight" (editorial), *Wall St. J.*, 28 July 2003, p. A10.

Privacy-Preserving Surveillance Using Databases from Daily Life

Latanya Sweeney, *Carnegie Mellon University*

As the price of disk storage continues to plummet, the cost of capturing and sharing data approaches zero, making it economical to collect more and more information on individuals' daily lives, often without any particular purpose.[1]

One proposed use for all this information is homeland security (law enforcement and intelligence). When fragments of captured information are combined, they provide person-specific, population-based data for profiling individuals. Database systems might use the data to find behavioral patterns of individuals engaged in illegal activity or planning terrorist acts.

Privacy concerns

American programs that sought to use databases for surveillance include CAPS II (computer-assisted passenger screening) and TIA (Total Information Awareness).[2] Both programs faced serious turmoil over privacy concerns. Such concerns include the following:

- The bulk of people whose information is in the database have done nothing to warrant suspicion.
- Surveillance on databases tends to exasperate privacy expectations and personal protections. While American courts have historically ruled that a person in a public space should have no expectation of privacy,[3] information stored in databases can be so invasive as to remove private enclaves within public spaces. For example, on a crowded bus, you can orient a document to limit what others can see. But limiting what a hidden camera with a zoom lens can see is difficult because its existence and viewing angle are unknown.
- Information in a database can be gathered from private spaces. For example, a private inquiry made on a home phone can become part of a database, making it indistinguishable from inquiries made at a public shop.
- Organizations using databases for surveillance purposes don't tend to implement Fair Information Practices (www3.ftc.gov/reports/privacy3/fairinfo.htm) because they don't want criminals and terrorists to alter their information or behavior. Therefore, no individual whose information is contained in the data has control over his information. The organizations don't seek consent from subjects or give notice to those included. (Arguably, doing so would be impractical.) So typically, subjects don't know their information is being held, and they have no right or means of correcting errors in it.

As the price of disk storage continues to plummet, it becomes economical to collect more and more information on individuals' daily lives, often without any particular purpose.

- No judicial review or impartial oversight exists to weigh societal benefits against individual risks. No independent third party limits fishing expeditions unwarranted inquiries, snooping on friends, or other kinds of "fishing expeditions."

My goal is to guarantee (or at least maximize) privacy protection while making data useful for surveillance. This work introduces a framework that addresses database privacy conditions in surveillance databases such that

- no person whose information is contained in the database can be reidentified without permission,
- investigators can access necessary information contained in the database freely and easily, and
- results from qualified inquiries are equivalent to results found in the absence of privacy protection.

Methods

One way to satisfy these privacy conditions is to model the probable cause predicate in American jurisprudence. A law officer wanting to intrude on a person's private life or affairs needs a search warrant, which a judge can issue. The officer appears before the judge and reports either facts for which he or she has first-hand knowledge or facts that he or she learned through an informant. Typically, the judge uses a two-prong test to make a decision: what is the basis of the knowledge, and is the source believable (see figure 4a)? We can model this process in technology by replacing the officer with anomaly or data-mining algorithms and the informant with data from various sources. We can replace the human judge with a combination of contracts and certifications from the original data collectors and a technology-enforceable policy statement with preset levels that match the identifiability of provided information with the minimal information the algorithm needs (see figure 4b). The technology capable of enforcing the policy is called *selective revelation*.

The first step in constructing a selective-revelation system requires identifying the algorithms to be used and the kinds of data involved. The person setting up the system performs analyses to provably anonymize the data and to verify that the algorithms remain useful with the anonymized data.

Once the initial step is complete, the person maps related regulations, policies, best practices, laws, and data certifications onto the scale of identifiability—from anonymous to identifiable—to specify the authority by which data can be accessed at each status (see figure 5). Finally, boundaries of algorithmic utility are established to identify the algorithmic circumstances under which more identifiable data is necessary.

Figure 5 shows how identifiability maps to investigation status. During normal operation, the surveillance agency uses anonymized data. If the agency encounters unusual activity, as evidenced by algorithmic results, then the system lowers the identifiability of related cases to "de-identified." De-identified data has no explicit identifiers but isn't provably anonymous. As the investigation status shifts downward, the provided information be-

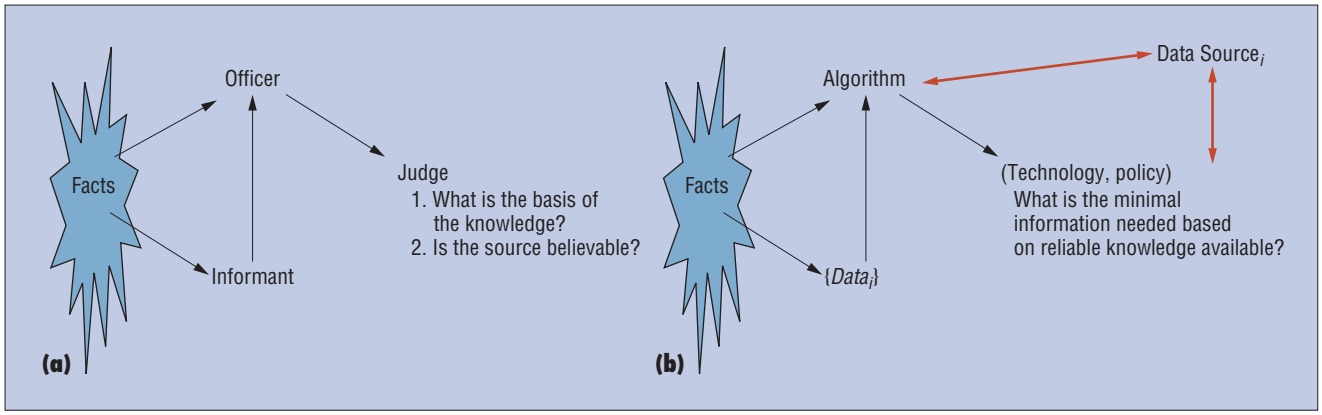


Figure 4. Probable cause predicate as conducted by (a) a human judge and (b) technology.

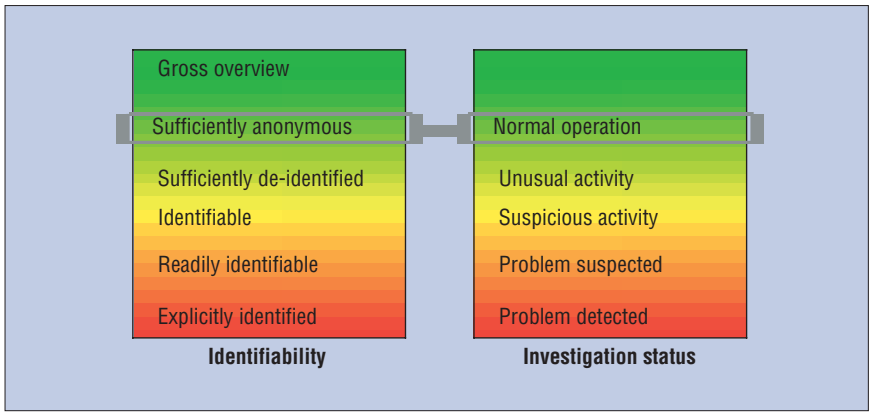


Figure 5. Selective revelation scales matching the identifiability of the data (left) to the operational mode (right).

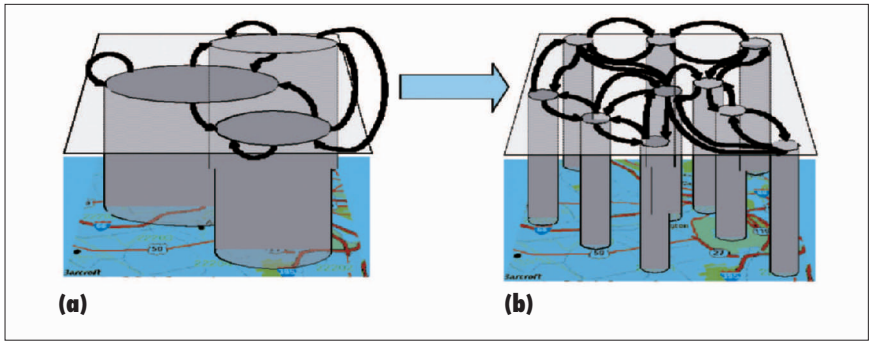


Figure 6. Dynamically augmenting data access as surveillance warrants. (a) Crude relationships are derived from sufficiently anonymous data. (b) More details are revealed using identifiable data.

comes increasingly more identifiable, until the agency meets the criteria for providing explicitly identified data. Figure 6 demonstrates the effect of lowering identifiability.

Example

Earlier, I constructed a selective-revelation system for bioterrorism surveillance was constructed in which hospitals, physicians, and labs provided medical data to a public health agency to determine whether an unusual number of respiratory cases were presented.[4] I anonymized the data under the scientific standard of the medical regulation known as the Health Insurance Portability and Accountability Act (HIPAA) (Privacert Compliance was used; www.privacert.com). The early aberration re-porting system algorithm from the Centers for Disease Control was used with the anonymized data. If it found evidence of unusual activity, the system automatically lowered anonymity. If further evidence emerged that an outbreak was underway, fully identified data under the Public Health Law was provided by the system. This selective revelation system provided impartial, automated oversight to database inquiries. It demonstrates how the American public can enjoy both safety and privacy.

References

1. L. Sweeney, "Information Explosion," *Confidentiality, Disclosure, and Data Access*, P. Doyle et al., eds., Elsevier, 2001, pp. 43–74.
2. "Your Papers, Please," World Net Daily, 16 Jan. 2003; www.worldnetdaily.com/news/article.asp?ARTICLE_ID=30523.
3. Rest. 2d, §652B.
4. L. Sweeney, "Privacy-Preserving Bio-terrorism Surveillance," *Proc. AAAI Spring Symp. AI Technologies for Homeland Security*, 2005.

The Changing Face of Privacy Policy and the New Policy-Technology Interface

Paul Rosenzweig, *Heritage Foundation*

America's rules- and regulation-driven model of privacy protection is undergoing a major transition. Driven partly by needs spawned in the wake of 9/11, the traditional

model is being replaced with a more process-oriented, results-driven set of rules and systems—a model based on oversight and technological protections. Although it's too early to judge the model's effectiveness, it offers, on the whole, a promising new way of merging policy and technology in the service of both privacy and national security.

The traditional way

We can see this change in the contrast between the privacy policy and technology rules systems adopted before 9/11 and their post-9/11 incarnations. For the last 40 years, US privacy law and policy has developed in an ad hoc, organic manner. The US has adopted rules by legislative initiative, with each piece of legislation directed at a specific, unique, narrowly focused problem. Thus, the US enacted the Fair Credit Reporting Act to deal with information in the hands of credit-reporting bureaus. Similarly, HIPAA protects the privacy of personally identifiable medical information; other sector-specific laws deal with banking data and educational information. The list is almost endless. Three distinct themes characterize this traditional approach to privacy policy.

Command and control

First, this approach has traditionally relied on a command-and-control method for defining the rules to protect privacy interests. Thus, the model we've adopted since the 1970s involves a series of statutory prohibitions implemented through a regulatory regime developed by an administrative agency. This method has often proven cumbersome. As with many command-and-control systems, the rationalist exercise of creating rules can produce over-regulation, confusion, and unintended consequences. Moreover, to the extent that broad exceptions exist in these types of regulatory structures, they often swallow the general rule in important cases. To cite one example, a law enforcement and national-security exception exists in almost every regulatory system that, by and large, renders the system inapplicable to the most salient contemporary concerns.

To see this approach's inefficacy, you only need to think, for example, of the regulations promulgated to protect medical information under HIPAA. The initial regulations were so complex and confusing that the federal government revised them even before they were implemented. And, as anyone who has encountered these rules will acknowledge, they've produced a mountain of paperwork and rules with precious little apparent increase in the protection of individual privacy.

Reactive, not proactive

Second, this traditional mode of regulation has been, in many cases, reactive rather than proactive. To be sure, Congress has developed a few individual regulatory systems because of real needs. But equally often, new legislation is the product of political circumstance and publicity. It's no accident that we have a strong series of rules protecting the privacy of video rental records—they were enacted after an enterprising reporter secured Judge Robert Bork's rental records when he was being considered for a Supreme Court vacancy. It's exceedingly odd, however, that the regulation of video rental privacy preceded the protection of medical-record privacy by more than a decade and odder still that, to a significant degree, the protections of the former exceed those of the latter.

> Since 9/11, we've raced to deploy new technologies that offer great promise in combating terrorism and to develop new institutional structures.

A corollary of this organic, reactive method of privacy protection is that our laws are almost invariably sector specific, addressing particular areas of concern. This might be wise in some ways—the best and most appropriate policy answers to adopt for medical-information privacy probably differ from those for financial records. But this method stands in substantial contrast to the European method of regulation, which generally uses a single standard of broad applicability.

Keeping up with change

Finally, the old method of developing privacy policies has been noticeably independent of technology. It has relied largely on non-technology-based rules (such as notice-and-consent regulations) that have proven less effective in practice than in theory. More significantly, legislative rules have often been outpaced by change. For example, the Privacy Act of 1974, once seen as the most significant expression of privacy principles in America, is now largely irrelevant. Its strictures apply only to systems of government-maintained records—that is, centralized government databases. The law utterly failed to anticipate the development of distributed data networks and the possibility of government access to third-party commercial records. Consequently, it's effectively obsolete today and has almost no noticeable role in restraining the antiterrorism technological developments at the forefront of today's debates about liberty and security. To put it colloquially, the Terrorism Information Awareness program was terminated for many reasons—but none of them had anything to do with the Privacy Act's legal limitations.

A new approach

Since 9/11, we've left many of these old methods behind. We've raced to deploy new technologies that offer great promise in combating terrorism and to develop new institutional structures (such as the Department of Homeland Security and the Office of the Director of National Intelligence). At the same time, America has begun to field new privacy-protection methods that we hope will serve in the changing technological environment. More particularly, we're changing to a process-based, results-oriented, technology-driven means of addressing privacy concerns. The change isn't yet complete, to be sure, but we can readily discern the new methods' outlines.

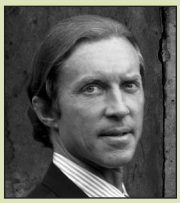
Institutional oversight

First, we're changing from top-down command-and-control rules to a process that protects privacy principally through institutional oversight. To that end, Congress created the Department of Homeland Security with a statutorily required Privacy Officer (and another Civil Liberties Officer). The Intelligence Reform Act, implementing the 9/11 Commission's recommendations, goes further. For the first time, it creates a Civil Liberties Protection Officer residing within the intelligence community. More generally, a presidentially appointed Privacy and Civil Liberties Oversight Board

John Yen is the University Professor of Information Sciences and Technology and the Professor-in-Charge of Penn State's School of Information Sciences and Technology. He is also cochair of the 2005 AAAI Spring Symposium on AI Technologies and Homeland Security. Contact him at jyen@ist.psu.edu.

Robert Popp is the deputy director of DARPA's Information Exploitation Office. He is also cochair of the 2005 AAAI Spring Symposium on AI Technologies and Homeland Security. Contact him at rpopp@darpa.mil.

George Cybenko is the Dorothy and Walter Gramm Professor of Engineering at Dartmouth College. Contact him at gvc@dartmouth.edu.



K.A. Taipale is the executive director of the Center for Advanced Studies in Science and Technology Policy and a senior fellow at the World Policy Institute, where he directs the Program on Law Enforcement and National Security in the Information Age. Contact him at mail05@advancedstudies.org.

Latanya Sweeney is an associate professor of computer science, technology and policy at Carnegie Mellon University. Contact her at latanya@cs.cmu.edu.

Paul Rosenzweig is senior legal research fellow in the Center for Legal and Judicial Studies at the Heritage Foundation and is chairman of the Department of Homeland Security Data Privacy and Integrity Advisory Committee. The views expressed here are his own and not those of the Committee or any other governmental entity. Contact him at paul.rosenzweig@heritage.org.

is to oversee government-wide intelligence and antiterrorism activities.

These institutions serve a novel dual function. They are, in effect, internal watchdogs for privacy concerns. And they also naturally serve as a focus for external complaints, requiring them to exercise some functions of ombudsmen. In either capacity, they're a new structural invention on the American scene—at least, with respect to privacy concerns—and their efficacy has yet to be fully tested.

Focus on results

The second significant change concerning how we address privacy concerns lies in the new focus on results rather than legal rules. We're using that focus to drive and force technological change. The paradigm example of this shift is the Intelligence Reform Act mandate for the creation of an information-sharing environment. That recommendation grew out of work by the Markle Foundation and the 9/11 Commission and recognizes the need for enhanced interconnectivity among federal databases. We must, as they say, connect the dots better.

Under the old paradigm, a detailed set of rules for protecting privacy would have accompanied this mandate. These rules might have mandated state-of-the-art or cutting-edge technological techniques, such as one-way hashes to anonymize data or the use of immutable audit trails. In today's rapidly evolving technological environment, these up-to-the-minute mandates would likely have soon become obsolete.

Recognizing the reality of technological change, Congress took a different tack. It simply defined the results it expected and tasked the Office of the Director of National Intelligence, acting in consultation with the Privacy and Civil Liberties Oversight Board, to issue guidelines and develop a system that protects privacy and civil liberties in developing and using the information-sharing environment. To enhance transparency and oversight, Congress also required that these guidelines be made public unless nondisclosure is clearly necessary to protect national security.

Future expectations

In practice, the two approaches might well mean the same thing in the near term. It's likely, if not certain, that the government will incorporate anonymization, pseudonymous identification, immutable audit trails, automated self-auditing, high-level encryption, and the like in the first iteration of the information-sharing environment.

But notice what's different. Instead of a static set of rules adopted once and for all, we now anticipate an iterative process. The oversight that institutions put in place will evaluate the tools' efficacy. On the basis of that evaluation (and, likely, in light of further technological changes), the information-sharing environment will be dynamically modified as necessary.

We've come a long way since 1974—from mainframes to distributed databases, from fingerprints to biometrics, and now, from a rigid, rule-based system to a dynamic system of results-oriented oversight and review. All the signs point to a better, more responsive, more nimble, more privacy-sensitive system of rules. ■