Background Materials Prepared for:

COMMITTEE ON THE TECHNICAL AND PRIVACY DIMENSIONS
OF INFORMATION FOR TERRORISM PREVENTION
AND OTHER NATIONAL GOALS

APRIL 2006

THE NATIONAL ACADEMIES

SELECTED WRITINGS V.1.6 OF

K. A. TAIPALE

CENTER FOR ADVANCED STUDIES IN SCI. & TECH. POL'Y

ABOUT THE AUTHOR

Kim Taipale, BA, JD (New York University), MA, EdM, LLM (Columbia University), is the founder and executive director of the Center for Advanced Studies in Science and Technology Policy, a senior fellow at the World Policy Institute, and an adjunct professor of law at the New York Law School.

Mr. Taipale is associated with the Markle Task Force on National Security in the Information Age, and serves on the Science and Engineering for National Security Advisory Board of The Heritage Foundation and the Information Policy Forum of LexisNexis. He is a frequent invited speaker and the author of many academic papers, articles, and book chapters on information, technology, and national security issues.

Mr. Taipale was previously the director of new media development for Columbia lnnovation Enterprise at Columbia University, where he was earlier a senior fellow at the Institute for Learning Technologies and also taught communications. Prior to that, he was an investment banker at Lazard Freres & Co., an executive at the Pullman Company, and a lawyer at Davis Polk & Wardwell.

ABOUT THE CENTER FOR ADVANCED STUDIES

The ***Center for Advanced Studies in Science and Technology Policy*** is an independent, non-partisan research and advisory organization focused on information, technology, and national security policy.

The Center seeks to inform and influence national and international policy- and decision-makers in both the public and private sectors by providing sound, objective analysis, insight, and advice; in particular by identifying and articulating issues that lie at the intersection of technologically-enabled change (both opportunities and challenges) and existing practice in public policy, law, and industry.

The Center has ongoing research projects in *Law Enforcement and National Security in the Information Age; Telecommunications and Cybersecurity Policy; Information Operations, Information Assurance, and Operational Resilience (Information Warfare); Environment and Energy Policy;* and *Intellectual Property and Trade*, among others.

More info and contact at www.advancedstudies.org.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

Prolegomenon

Author's Note:

The selected background materials presented in this edition include several short essays as well as excerpts from, or summaries of, longer articles or book chapters published during the period 2003-2006 and relating to technology, security, and privacy.

In particular, the selections and excerpts presented here focus on the technologies of automated data analysis (including "data mining" in both its popular usage and its more technical sense) and their application to counterterrorism and intelligence.

Although each selection provides a slightly different perspective or emphasizes a particular aspect of these issues, several thematic points are repeated throughout, including:

- Security and liberty are not dichotomous rivals to be traded one for another but rather are dual obligations, each to be maximized within the constraints of the other,

- Technology development is not deterministic, but it is inevitable (that is, we face a certain future of more data availability and more sophisticated analytic tools),

- Political strategies premised on outlawing particular technologies or techniques, or seeking to constrain their use through laws alone, are second-best – and ultimately futile – strategies that will result in little security and brittle privacy protections, and

- Technology cannot itself either provide security or insure privacy, however, properly employed, it can help better allocate security resources and, properly designed, it can enable familiar, existing civil liberty protecting mechanisms, procedures, and doctrines (or their analogues) to function.

Providing both security and liberty requires a more informed policy debate in which the nature of the new challenges to traditional civil liberties doctrine is better understood, and where the security strategies to be employed are examined within the construct of the dual obligation to provide both security and liberty. Security and liberty are not competing rivals to be "balanced" but rather vital interests to be reconciled.

K. A. Taipale
New York

THIS PAGE INTENTIONALLY LEFT BLANK

Abstract and Executive Summary of:

Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data

5 COLUM. SCI. & TECH. L. REV. 2 (Dec. 2003)

K. A. Taipale [*]

ABSTRACT:

Official U.S. Government policy calls for the research, development and implementation of advanced information technologies for aggregating and analyzing data, including data mining, in the effort to protect domestic security. Civil libertarians and libertarians alike have decried and opposed these efforts as an unprecedented invasion of privacy and a threat to our freedoms.

This article examines data aggregation and automated analysis, particularly data mining, and related privacy concerns in the context of employing these techniques in domestic security. The purpose of this article is not to critique or endorse any particular proposed use of these technologies but, rather, to inform the debate by elucidating the intersection of technology potential and development with legitimate privacy concerns. It is a premise of this article that security and privacy are dual obligations, not dichotomous rivals to be traded one for the other. "In a liberal republic, liberty presupposes security; the point of security is liberty."

Thus, this article argues that security with privacy can be achieved by employing value sensitive technology development strategies that take privacy concerns into account during development, in particular, by building in rule-based processing, selective revelation, and strong credential and audit features. This article does not argue that these technical features alone can eliminate privacy concerns but, rather, that these features can enable familiar, existing privacy protecting oversight and control mechanisms, procedures and doctrines (or their analogues) to be applied in order to control the use of these new technologies.

[*] *Mr. Taipale is the founder and executive director of the Center for Advanced Studies in Science and Technology Policy, a senior fellow at the World Policy Institute, and an adjunct professor of law at New York Law School.*

The Center for Advanced Studies in Science and Technology Policy is a private, non-partisan research and advisory organization focused on information, technology, and national security policy.

Further, this article argues that *not* proceeding with government funded research and development of these technologies (in which political oversight can incorporate privacy protecting features into the design of these technologies) will ultimately lead to a diminution in privacy protection as alternative technologies developed without oversight (in classified government programs or proprietary commercial programs) are employed in the future since those technologies may lack the technical features to protect privacy through legal and procedural mechanisms. Thus, the recent defunding of DARPA's Information Awareness Office and its Terrorism Information Awareness program and related projects is likely to turn out to be a pyrrhic 'victory' for civil liberties as this program provided a focused opportunity around which to publicly debate the rules and procedures for the future use of these technologies and, importantly, to oversee the development of the appropriate technical features required to support any concurred upon implementation or oversight policies to protect privacy.

Even if it were possible, controlling technology through law alone, for example, by outlawing the use of certain technologies or shutting down any particular research project, is likely to provide little or no security and only brittle privacy protection.

EXECUTIVE SUMMARY:

Vast Data Volumes Exceed Analytic Resources

Recent reports by the U.S. Congress, the National Research Council, the Markle Foundation and others have highlighted that the amount of available data to be analyzed for domestic security purposes exceeds the capacity to analyze it. Further, these reports identify a failure to use information technology to effectively address this problem.

> "While technology remains one of this nation's greatest advantages, it has not been fully and effectively applied in support of U. S. counter-terrorism efforts."

Among the recommendations put forth in these reports are the increased use of data aggregation (information sharing) and automated analysis (in particular data mining) technologies.

Data Aggregation and Automated Analysis

Data aggregation (including data integration and data sharing) is intended to overcome the "stovepipe" nature of existing datasets. Research here is focused on making information available to analysts regardless of where it is located or how it is structured. A threshold issue that has technical, security and privacy implications is whether to aggregate data in a centralized data warehouse or to access information directly in distributed databases.

Automated data analysis (including data-mining) is intended to turn low-level data, usually too voluminous to understand, into higher forms (information or knowledge) that

might be more compact (for example, a summary), more abstract (for example, a descriptive model), or more useful (for example, a predictive model).

> "A key problem [for using data mining for counter-terrorism] is to identify high-level things – organizations and activities – based on low-level data – people, places, things and events."

The application of data aggregation and automated analysis technologies to domestic security is the attempt to "make sense of data" by automating certain analytic tasks to allow for better and more timely analysis of existing datasets in order to prevent terrorist acts by identifying and cataloging various threads and pieces of information that may already exist but remain unnoticed using traditional means, and to develop predictive models based on known or unknown patterns to identify additional people, objects or actions that are deserving of further resource commitment or law enforcement attention.

Compounding the problem in domestic security applications is that relevant data (that is, information about terrorist organizations and activities) is hidden within vast amounts of irrelevant data and appears innocuous (or at least ambivalent) when viewed in isolation. Individual data items – relating to people, places and events, even if identified as relevant – are essentially meaningless unless viewed in context of their relation to other data points. It is the network or pattern itself that must be identified, analyzed and acted upon. Thus, there are three discrete applications for automated analysis in the context of domestic security:

- first, *subject-oriented link analysis*, that is, automated analysis to learn more about a particular data subject, their relationships, associations and actions;

- second, *pattern-analysis* (or data mining in the narrow sense), that is, automated analysis to develop a descriptive or predictive model based on discovered patterns; and,

- third, *pattern-matching*, that is, automated analysis using a descriptive or predictive model (whether itself developed through automated analysis or not) against additional datasets to identify other related (or "like") data subjects (people, places, things, relationships, etc.).

Because spectacular terrorist events may be too rare or infrequent for automated analysis to extract useful patterns, the focus of these techniques in counter terrorism is to identify lower level, frequently repeated events (for example, illegal immigration, money transfers, front businesses and recruiting activity) that together may warrant further attention or resource commitment.

Thus, data aggregation and automated analysis are not substitutes for human analytic decision-making, rather, they are tools that can help manage vast data volumes and potentially identify relational networks that may remain hidden to traditional analysis. If

successful, these technologies can help allocate available domestic security resources to more likely targets.

Privacy Concerns

Because data aggregation and automated analysis technologies can cast suspicion based on recognizing relationships between individually innocuous data, they raise legitimate privacy concerns. However, much of the public debate regarding the potential use of these technologies is overshadowed by simplifications, misunderstandings and misrepresentations about what the technologies can do, how they are likely to be employed and what actual affects their employ may have on privacy and security.
The significant privacy concerns relating to these technologies are primarily of two kinds: those that arise from the aggregation (or integration) of data itself and those that arise from the automated analysis of data that may not be based on any individualized suspicion – the former might be called the *database problem* and the latter the *mining problem*.

The database problem is implicated in subject-based inquiries that access distributed databases to find more information about a particular subject. To the extent that maintaining certain government inefficiencies helps protect individual rights from centralized state power, the primary privacy question involved in aggregation is one of increased government efficiency.

The mining problem is implicated in the use of pattern-matching inquiries, in which profiles or models are run against data to identify unknown individuals. To some, pattern-matching raises privacy issues relating to non-particularized suspicion in violation of the Fourth Amendment.

Additional concerns are that the technology will not work for the intended purpose (providing either a false sense of security by generating false negatives or imposing civil liberties costs on too many innocent people by generating false positives), and that the technology is subject to potential abuse or that it will be vulnerable to attack.
The issue of false positives and false negatives is not insignificant but is an issue of efficacy and requires further research to determine whether an appropriate *confidence interval* for counter terrorism applications can be achieved. The point of the research is to find out if the technologies can work – if they cannot, other privacy concerns are moot since the technologies will not be employed. If they can, then appropriate policies and procedures to manage and compensate for error rates can be developed before implementation.

Building in Technical Constraints

Assuming some acceptable baseline efficacy, it is the premise of this article that privacy concerns relating to data aggregation and data mining in the context of domestic security can be significantly mitigated by developing technologies that enable the application of existing legal doctrines and related procedures to their use:

- First, that *rule-based processing* and a *distributed database architecture* can significantly ameliorate the general data aggregation problem by limiting the scope of inquiry and the subsequent processing and use of data within policy guidelines;

- Second, that *selective revelation* can reduce the non-particularized suspicion problem, by requiring an articulated particularized suspicion and intervention of a judicial procedure before identity is revealed; and

- Finally, *strong credential and audit features* and *diversifying authorization and oversight* can make misuse and abuse "difficult to achieve and easy to uncover".

Further, this article contends that developing these features for use in domestic security applications will lead to significant opportunities to enhance overall privacy protection more broadly in the U.S. (and elsewhere) by making these technical procedures and supporting features available for voluntary or legislated adoption in the private sector. In addition, the development of these technologies will have significant beneficial "spill-over" uses for commercial and scientific applications, including improved information infrastructure security (better user authentication, encryption, and network security), protection of intellectual property (through rule-based processing) and the reduction or elimination of spam (through improved analytic filtering).

Overriding Principles

This article proffers certain guiding principles for the development and implementation of these technologies:

> First, that these technologies only be used as investigative, not evidentiary, tools (that is, used as a predicate for further investigation not proof of guilt) and only for investigations of activities about which there is a political consensus that aggressive preventative strategies are appropriate (for example, counter-terrorism and national security).
>
> Second, that specific implementations be subject to strict congressional oversight and review, be subject to appropriate administrative procedures within executive agencies where they are to be employed, and be subject to appropriate judicial review in accordance with existing due process doctrines.
>
> And, third, that specific technical features that protect privacy by providing opportunities for existing doctrines of due process and reinforcing procedures to function effectively, including rule-based processing, selective revelation and secure credentialing and tamper-proof audit functions, are developed and built into the technologies.

Article Structure

The Prelude and Introduction to this article contextualize the debate about the need for and potential use of these technologies. Part I then provides a more detailed introduction to data aggregation and analysis technologies, in particular, data mining. Part II examines certain government initiatives, including TIA and CAPPS II, as paradigmatic examples of development efforts in these areas. Part III outlines the primary privacy concerns and the related legal framework. Part IV suggests certain technology development strategies that could help ameliorate some of the privacy concerns. And, Part V concludes by restating the overlying principles that should guide development in these technologies.

Article Table of Contents

TABLE OF CONTENTS AND ABSTRACT

TECHNOLOGY, SECURITY AND PRIVACY: THE FEAR OF FRANKENSTEIN, THE MYTHOLOGY OF PRIVACY AND THE LESSONS OF KING LUDD

7 YALE J. L. & TECH. 123;
9 INTL. J. COMM. L. & POL'Y 8
(DECEMBER 2004)

K. A. TAIPALE *

This article was jointly reviewed and edited by the YALE JOURNAL OF LAW & TECHNOLOGY and the INTERNATIONAL JOURNAL OF COMMUNICATIONS LAW & POLICY.

* *Kim Taipale, BA, JD (NYU), MA, EdM, LLM (Columbia University), is the founder and executive director of the Center for Advanced Studies in Science and Technology Policy, a senior fellow at the World Policy Institute, and an adjunct professor of law at New York Law School.*

ABSTRACT

TECHNOLOGY, SECURITY AND PRIVACY: THE FEAR OF FRANKENSTEIN, THE MYTHOLOGY OF PRIVACY AND THE LESSONS OF KING LUDD

K. A. TAIPALE

This article suggests that the current public debate that pits security and privacy as dichotomous rivals to be traded one for another in a zero-sum game is based on a general misunderstanding and apprehension of technology on the one hand and a mythology of privacy that conflates secrecy with autonomy on the other. Further, political strategies premised on outlawing particular technologies or techniques or seeking to constrain technology through laws alone are second-best – and ultimately futile – strategies that will result in little security and brittle privacy protection.

This article argues that civil liberties can best be protected by employing value sensitive technology development strategies in conjunction with policy implementations, not by opposing technological developments or seeking to control the use of particular technologies or techniques after the fact through law alone. Value sensitive development strategies that take privacy concerns into account during design and development can build in technical features that can enable existing legal control mechanisms and related due process procedures for the protection of civil liberties to function.

This article examines how identification, data aggregation and data analysis (including data mining), and collection technologies intersect with security and privacy interests and suggests certain technical features and strategies premised on separating knowledge of behavior from knowledge of identity based on the anonymization of data (for data sharing, matching and analysis technologies) and the pseudonymization of identity (for identification and collection technologies). Technical requirements to support such strategies include rule-based processing, selective revelation, and strong credential and audit.

THIS PAGE INTENTIONALLY LEFT BLANK

The McGraw-Hill Homeland Security Handbook

David G. Kamien, Editor

SECTION TWELVE

Domestic Security and Civil Liberties

CHAPTER 64

Introduction to Section 12

K. A. Taipale
Executive Director, Center for Advanced Studies in Science and Technology Policy, World Policy Institute

The McGraw-Hill Companies

1 2 3 4 5 6 7 8 9 10 DOC/DOC 0 9 8 7 6 5

ISBN 0-07-144665-6

CHAPTER 64

Introduction to Section 12

K. A. Taipale
Executive Director, Center for Advanced Studies in Science and Technology Policy, World Policy Institute

Within the public discourse, concerns about domestic security and civil liberties are often asserted as competing and potentially incompatible policy interests requiring the achievement of some tolerable state of balance. Implicit in this notion of balance is the smuggled assumption of a dichotomous rivalry in which security and liberty are traded one for another in a zero-sum political game. But the notion is misleading, for there is no fulcrum—as is implicit in the metaphor of a balance—at which point the correct amount of security and liberty can be achieved. Rather, security and liberty are dual obligations of civil society, and each must be maximized within the constraints imposed by the other. "In a liberal republic, liberty presupposes security; [and] the point of security is liberty."[1]

Because metaphor affects not just how we communicate but also how we structure our understanding and perception from the outset, challenging the prevailing metaphor of balance is not simply a semantic game. Metaphor has suasive power, particularly in policy debates, because it sets the expectations that can presuppose the outcome. The notion of balance pits security against liberty in a presumed Jacobin antagonism: those seeking to maintain civil liberties can then be said to be against collective security, and those seeking security can be accused of being too easily willing to forgo individual liberty. Often invoked—but rarely parsed—is a comment attributed to Benjamin Franklin: "Those who would give up Essential

Liberty to purchase a little temporary Safety, deserve neither Liberty nor Safety."[2] Inherent in Franklin's rejection of purchasing temporary security by giving up essential liberty, however, is the presumption not of tension between the two, but rather of a duality of concern with security on the one hand, and with liberty on the other.

Nevertheless, in policy debates about security and liberty, the need for achieving balance is often invoked in response to perceived challenges to the doctrinal status quo. The primary challenge to existing doctrine—that is, the presumed imbalance to be righted—in the current debate results from a blurring of the traditional line between reactive domestic law enforcement policies and preemptive national security strategies in response to the changed nature of the threat posed by transnational terrorism.[3] Because the exercise of law enforcement and national security power has previously been governed by disparate—and potentially irreconcilable—doctrines and laws, this blurring requires determining which set of existing principles, or what new principles, will govern in these changed circumstances.

New security policies are necessary because the seed value of potentially catastrophic outcomes to national security has devolved from other nation-states (the traditional target of national security power) to organized but stateless groups (the traditional target of law enforcement power), and the consequences of failing to prevent attacks before they occur have become politically unbearable. Thus a general consensus to take a preemptive rather than a reactive approach to counter these threats has emerged: even the most strident civil libertarians concede the need to identify and stop terrorists before they act. Preemption, however, requires *actionable intelligence*—that is, information useful in predicting and countering future behaviors. Actionable intelligence can generally be obtained only through forms of *surveillance*—the selective observation of precursor behaviors that are usually ambiguous and often resemble lawful behavior. Although surveillance is an accepted national security strategy, when it is applied in the context of domestic security it has the potential to conflict with traditional policy doctrines and legal structures premised in part on protecting individual liberty and based on the presumption of innocence. The policy question becomes when and under what circumstances selective government attention can be properly focused on a particular group or individual, and what standard such intelligence

must meet before it is actionable for counterterrorism sanctions, particularly for sanctions that may not be subject to traditional judicial due process: for example, restrictions on travel or deportation for unrelated infractions.

Reconciling these conflicting needs, however, does not necessitate slighting one for the other if security and liberty are accepted as dual obligations. Indeed, there is no inherent policy conflict at all between security and liberty within the constitutional framework of reasonableness. Strategies that place an unreasonable burden on liberty—for example, demonizing a minority or engendering suspicion of everyone—are not just unacceptable outcomes for liberty but measures that provide little or no security because they are ineffective at identifying terrorists and they undermine the public cooperation and confidence needed for success. On the other hand, liberty incurs responsibility,[4] and unfettered liberty at the expense of security that can potentially result in catastrophic outcomes impinges not just on collective security but also on the very foundation of liberty for all individuals and is itself, therefore, unreasonable. However, *effective* security strategies—strategies that actually help locate, target, and preempt terrorists before they act without unduly burdening the vast majority of innocent people—are by definition not unreasonable, since individual liberty is not synonymous with permitting plotters to commit terrorist acts free from sanction.

Providing both security and liberty requires an informed policy debate in which the nature of the new challenges to traditional civil liberties doctrine is better understood and where the security strategies at hand for resolution are examined within the construct of the dual obligation to provide both security and liberty. The chapters in this section discuss these issues.

In Chapter 65 Paul Rosenzweig examines some of the changing base conditions presented by the transnational terrorist threat and various counterterrorism strategies. Abraham Foxman (Chapter 66) and Laura Murphy (Chapter 67) examine the changing nature of the threat to civil liberties and identify constraints imposed by civil liberty concerns on security strategies. Finally, in Chapter 68 Newton Minow and Fred Cate examine how new technologies challenge privacy but also present new opportunities to protect both civil liberties and security. Together, these chapters help inform a process that can—and must—lead to fulfilling both obligations: improved security and protected liberty.

NOTES

1. Thomas Powers, "Can We Be Secure and Free?" *Public Interest* 151:3 (Spring 2003): 5.
2. Pennsylvania Assembly: Reply to the Governor, 11 November 1755. See Leonard W. Labaree (ed.), *The Papers of Benjamin Franklin*, Vol. 6 (1963), p. 242.
3. U.S. Department of Justice, "Fact Sheet: Shifting from Prosecution to Prevention, Redesigning the Justice Department to Prevent Future Acts of Terrorism" (29 May 2002).
4. See the following: Amitai Etzioni, *The Spirit of Community: Rights, Responsibilities, and the Communitarian Agenda* (New York: Crown Publishers, 1993), 323 pp. George Bernard Shaw, *Man and Superman* and *Maxims for Revolutionists: Liberty and Equality* (Cambridge, Mass.: The University Press, 1903). ("Liberty means responsibility. That is why most men dread it.")

THIS PAGE INTENTIONALLY LEFT BLANK

Excerpts from:

Chapter 8: Why Can't We All Get Along? How Technology, Security, and Privacy can Co-Exist in the Digital Age

***in* CYBERCRIME AND DIGITAL LAW ENFORCEMENT, Yale Information Society Project Book Series, (Jack Balkin, *et al., eds.*, NYU Press, forthcoming 2006)**

K. A. TAIPALE [*]

The public debate that pits security and privacy as dichotomous rivals to be traded one for another in a zero-sum game is based on a general misunderstanding and apprehension of technology on the one hand, and a mythology of privacy that conflates secrecy with autonomy on the other. Further, political strategies premised on outlawing particular technologies or techniques or that seek to constrain technology through laws alone are doomed ultimately to failure and will result in little security and brittle privacy protection.

Security and privacy are not a balancing act but rather dual obligations of a liberal democracy [1] that present a difficult problem for policy makers. Reconciling these divergent needs requires that policy and technology be developed concurrently and designed from the outset to work together. In a technologically mediated information society, civil liberties can only be protected by employing value sensitive technology development strategies [2] in conjunction with policy implementations, not by opposing technological developments or seeking to control the use of particular technologies or techniques after the fact through law

---

[*] *Kim Taipale, BA, JD (New York University), MA, EdM, LLM (Columbia University), is the founder and executive director of the Center for Advanced Studies in Science and Technology Policy. Mr. Taipale is also a senior fellow at the World Policy Institute and an adjunct professor of law at New York Law School.*

The author would like to thank the Yale Information Society Project and the organizers of the Yale CyberCrime Conference 2004. A longer version of this chapter appears in 7 Yale J. L. & Tech. 123; 9 Intl. J. Comm. L. & Pol'y 8 (Dec. 2004).

[1] Thomas Powers, *Can We Be Secure and Free?* 151 PUBLIC INTEREST 3, 5 (Spring 2003).

[2] BATYA FRIEDMAN, HUMAN VALUES AND THE DESIGN OF COMPUTER TECHNOLOGY (1997).

alone. Development strategies that take privacy concerns into account during design and development can build in technical features that enable existing legal control mechanisms for the protection of civil liberties and due process to function. [3]

Code is not law, but code can bound what law, norms and market forces can achieve. [4] Technology itself is neither the problem nor the solution, rather it presents certain opportunities and potentials that enable or constrain public policy choice. Technical features alone cannot eliminate privacy concerns, but by incorporating such features into technological systems familiar due process mechanisms are enabled.

This chapter examines how identification, data aggregation and analysis, and collection technologies intersect with privacy and security, and suggests certain technical features to help ameliorate concerns. It also proposes that strategies premised on separating *knowledge of behavior* from *knowledge of identity* based on the *anonymization* of data and the *pseudonymization* of identity can help protect individual autonomy while still meeting security needs.

While I focus on the intersection of technology and domestic security in the context of the 'war on terrorism,' [5] the analysis presented herein is applicable to law enforcement more generally – subject, however, to certain caveats. In particular, the lesser the crime targeted, the greater the hurdle for any new technology or wider use that implicates those concerns.

It is beyond the scope of this chapter to attempt to delineate precisely where the line between preemptive and reactive strategies is to be drawn. Post hoc analyses of 9/11 have revealed that much relevant information existed but government

---

[3] K. A. Taipale, *Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data*, 5 COLUM. SCI. & TECH. L. REV. 2 (2003); *see also* Paul Rosenzweig, *Proposal for Implementing the Terrorism Information Awareness System,* 2 GEO. J. L. & PUB. POL'Y 169 (2004); ISAT 2002 Study, *Security with Privacy*, Dec. 13, 2002.

[4] LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE 3-8 (1999).

[5] I use the phrase 'war on terrorism' throughout this chapter because it is the prevailing metaphor. *But cf.* Terry Jones, *Why Grammar is the First Casualty of War*, London Daily Telegraph, Dec. 1, 2001 ("How do you wage war on an abstract noun?"). *And see* GEORGE LAKOFF & MARK JOHNSON, METAPHORS WE LIVE BY 3-6 (2003) (discussing how metaphors not only affect how we communicate but actually structure our perceptions and understandings from the outset).

agencies were unable to "connect the dots." [6] It would be an unusual polity that now demanded accountability from its representatives for being unable to connect the dots to prevent terrorist acts yet denied them the available tools to do so, particularly if there were to be another catastrophic event. Therefore, I start with the assumption that there exists a political consensus for proactive investigative strategies intended to prevent future acts of terrorism [7] and that we need to enlist advanced information technology to help counter this threat. [8]

At the same time, however, we must recognize that information technology can be intrusive on certain privacy interests that protect individual freedom and political autonomy, and are core to our political liberties. [9] Further, there is no technological silver bullet that will provide absolute security (nor is there any technical solution that will absolutely protect privacy). Technology alone is not a solution to either problem; but neither are simple laws prohibiting the use of a technology or technique sufficient in themselves. Instead, some complex socially-constructed system combining organizational structures, rules and procedures, and technologies must be crafted. [10]

[PAGES 8-4 through 8-34 DELETED]

---

[6] *See* National Commission on Terrorist Attacks Upon the United States, *Final Report* (July 2004); Joint Inquiry Into the Intelligence Community Activities Before and After the Terrorist Attacks of September 11, 2001 House Permanent Select Comm. on Intelligence & Senate Select Comm. on Intelligence, H. Rep. No. 107-792, S. Rep. No. 107- 351 (2002).

[7] *See* U.S. Department of Justice, *Fact Sheet: Shifting from Prosecution to Prevention, Redesigning the Justice Department to Prevent Future Acts of Terrorism* (2002).

[8] Markle Foundation, *Creating a Trusted Network for Homeland Security: Second Report of the Markle Foundation Task Force* (2003); Markle Foundation, *Protecting America's Freedom in the Information Age: A Report of the Markle Foundation Task Force* (2002); Committee on Science and Technology for Countering Terrorism, National Research Council, *Making the Nation Safer: The Role of Science and Technology in Countering Terrorism* (2002).

[9] *See* ALAN WESTIN, PRIVACY AND FREEDOM (1967).

[10] *See generally* JOHN AUSTIN, THE PROVIDENCE OF JURISPRUDENCE DETERMINED (1832) ("positive law" as social construction); and THE SOCIAL CONSTRUCTION OF TECHNOLOGICAL SYSTEMS (Wiebe E. Bijker et al. eds. 1994).

[PAGES 8-4 through 8-34 DELETED]

The Privacy Divide

Reconciling competing interests in the privacy-security debate requires determining under what circumstances (and following what procedures) identity is to be associated with behavior or behavior with identity. The privacy divide is the point where attribution of behavior and identity occurs. The key to developing new information technologies and systems that can both improve security and protect privacy is to design systems that allow for procedural intervention and control at that point.

The question for both policy and systems development is when and under what circumstances certain data attribution is to occur – simply put, when is an individual to be associated with data representing their behavior or, conversely, when is behavior (whether observed by physical surveillance or within data) to be attributed to a specific individual.

A. Controlling the Privacy Divide: The Privacy Appliance as Metaphor

Conceptualizing and designing the appropriate mechanisms to exert control over the privacy divide is the key issue for protecting privacy in networked information systems. The policy challenge is to determine the rules and procedures governing the divide, and the technical challenge is to build in technical features to execute or enforce those rules and to manage accountability. The overall architecture must include organizational, procedural, and technical features in a framework that integrates these control requirements within business process needs.

The notion of a *privacy appliance* – that is, a technical systems component sitting between the point of access and the data itself to enforce policy rules and provide accountability – has been suggested. [11] Here I consider the privacy appliance as a metaphor – that is, as an analytic device representing the need for policy intervention in technical systems to enforce rules – rather than as a particular technical device or application.

From a policy perspective, it is not necessary *a priori* to decide if the privacy appliance should be a specific piece of hardware, for example, a firewall, or an

[11] Adm. John Poindexter, Presentation at DARPAtech 2002 (Aug. 2, 2002).

application, say an analytic filter, as its actual form of technical implementation is secondary to understanding what business needs are to be supported. The point for policy makers is to understand that intervention can happen at many different points in the technical architecture, and can be subject to varying methods of control. [12]

For the policy maker, the privacy appliance represent the technical object to enforce policy in systems, thus who controls them (for example, the party using the data, the party supplying the data, a trusted or untrusted third party) and how (through direct technical control, automated monitoring, control of audit or logs) and subject to what general oversight and review (for example, executive, legislative or judicial) are the pertinent policy questions.

For the technologist, understanding the policy needs forms the basis for determining technical requirements. Together, policy needs and system design form an enterprise architecture in which the information management, data, systems, and technology architecture all support the overall business process needs [13] – in this case, enabling certain security processes while still protecting privacy interests.

Technical design strategies that emphasize *anonymization* of data for analysis and *pseudonymization* of identity for identification and surveillance systems can provide intervention points where due process procedures can function. Together with strong user authentication and audit controls, these strategies can mediate between distributed data sources (including collection sensors) and the analyst (at any organizational level). Such procedures are dependant on organizational, structural and technical design features functioning together to meet articulated policy needs.

Anonymization of Data

An anonymous record is one in which the data cannot be associated with a particular individual, either from the data itself, or by combining the transaction

---

[12] *See* K. A. Taipale, *Designing Technical Systems to Support Policy: Enterprise Architecture, Policy Appliances, and Civil Liberties*, *in* EMERGENT INFORMATION TECHNOLOGIES AND ENABLING POLICIES FOR COUNTER TERRORISM (Robert Popp and John Yen, *eds*., Wiley-IEEE Press, Mar. 2006) (describing the use of *policy appliance* to enforce rules in technical systems).

[13] *See generally* HOWARD SMITH & PETER FINGAR, IT DOESN'T MATTER, BUSINESS PROCESSES DO (2003).

with other data. [14] Here traditional encryption strategies need to be distinguished from truly anonymous procedures such as *one-way hashing*. With traditional encryption strategies data is encrypted but can be decrypted with the use of a key. An analogy would be handing over data in a locked box. Theoretically, encrypted data is not truly anonymous as the underlying data can be accessed assuming that it is combined with the key.

By using one-way hashes the original data cannot be reconstructed from the hash. Hashing is a process of passing data through a one-way algorithm that returns a digital signature in place of the original data. This digital signature is unique to the underlying data but cannot be turned back into the original data – much like a sausage can be identified as pork but "cannot be turned back into a pig." [15] One-way hash technologies allow for anonymous data processing to occur in a shared environment since the only thing exchanged is the hash. If a match occurs, the processing party still needs to come back to the original data holder to access the underlying data. This allows organizational and procedural structures to be imposed between data matching and disclosure of the underlying data.

Theoretically, hashing is vulnerable to what is known as a "dictionary attack" in which an attacker compiles a list of all potential inputs and computes a hash function for each, then compares the hashed output from the target data set against their own list of hashes computed from all possible inputs to determine if there is a match. To counter the dictionary attack, *salt* is used. Salt is a random string that is concatenated to the original data before it is operated on by the hash function. In order to match the hashed outputs you need to share the salt key. Salt keys can be encoded in hardware or software, can be used to control the domain across which sharing occurs, and can even be used to control data expiration. By controlling the sharing of salt, the kind of processing that the data is subject to can be controlled.

Many security needs for data analysis, including list- and pattern-matching, can be accomplished within an anonymized data framework. Automated data analysis, including some forms of data mining, may also be possible.

---

[14] Roger Clarke, *Identified, Anonymous and Pseudonymous Transactions: The Spectrum of Choice*, Presented at User Identification & Privacy Protection Conference, Stockholm (June 14-15, 1999).

[15] Jeff Jonas, SRD, quoted in Steve Mollman, *Betting on Private Data Search*, WIRED NEWS, Mar. 5, 2003.

A simple example of anonymous list matching follows. Suppose a primary dataset contains traveler data and a second dataset contained suspected terrorists. Before the data is analyzed, both datasets are subject to the same one-way hash algorithm. Now, identifiers for "Joseph K." in the first dataset are represented by encrypted digital hashes that do not reveal the original data but can still be exchanged or matched against the corresponding data in the second set since the matching name or other identifiers in that database would have a matching hash (digital signature). Should a match occur, the government would be required to follow the appropriate procedures prior to being granted access to the raw data corresponding to the match held by the original owner who maintains exclusive control of the actual data throughout.

As noted above, by controlling the sharing of *salt keys* additional policy restrictions can be enforced. Not only can hashed data not be turned back into the original data, it cannot be matched or used for any other purpose without a matching salt key. Thus, control of salt variables allows searches to be restricted to certain data sets or domains.

Pseudonymization of Identity

A pseudonymous record or transaction is one that cannot – in the ordinary course of events – be associated with a particular individual. [16] Pseudonymity is a form of "traceable anonymity" and requires legal, organizational or technical procedures so that the association (or attribution) can only occur under specified and controlled circumstances. Pseudonymity is also referred to as *identity escrow*. [17] A pseudonym can be either *transient* or *persistent*. Persistent pseudonym's develop their own reputational attributes and can be tracked over time or across systems.

Pseudonymity allows for the disclosure of only the particular attribute relevant (and appropriate) for the particular transaction in which an exchange of data is required. For example, in the use of credit cards, the merchant does not actually need the purchaser's name to complete the transaction as only the authorization (that the issuer will pay the amount of the purchase) is relevant to the transaction.

---

[16] Clarke, *supra* note 14.

[17] *See, e.g.,* Joe Kilian and Erez Petran, *Identity Escrow*, in Advances in Cryptology - CRYPTO'98: 18th Annual International Cryptology Conference, Santa Barbara, California, USA, August 1998. Proceedings (H. Krawczyk, ed.).

(Whether the cardholder is the authorized user is an authentication issue unrelated to the transaction specifically and also does not require revealing a name).

Technical means exist to prove authorization without revealing actually identity by using *third party certification* in which a trusted third party certifies an authorization. The holder of the certificate (digital or otherwise) then presents the certificate to the second party (who may still authenticate that the individual is the authorized user), however, the original party does not have to reveal additional identifiers (or identity) in order for the second party to grant the level of authorization certified by the third party. (There are also methods for certifying untraceable authorizations, for example, digital cash).

Technical means also exist to provide accountability without disclosing identity at the point of verification. Identity escrow is a form of third party certification, in which the trusted third party certifies that they know the "true" identity of the user.

For example, a pseudonymous driver's license based on smart card technology (a smart card is any pocket-sized device that contains a processor or microchip that can interact with a reader) could be designed only to authenticate that the driver is authorized to drive (for example, by producing a digital certificate from the DMV certifying the holder's authorized status) without disclosing a common identifier during a traffic stop. The police officer could still run a data match against, for example, a wanted-felon or terrorist watch list (also without revealing a name or common identifier) by reading a hashed identifier keyed (through shared salt) to the felony or watch list database hashing algorithm. If there were an initial data match then additional procedures may or may not be called for, however, without a match, the purpose of the traffic stop could be accomplished without disclosing identity. The same card could be designed to only exchange, for example, age information with a bartender's card reader, or health information with a medical worker, etc.

An important policy consideration in any such system is determining whether such pseudonymous encounters are themselves logged – that is, do they generate their own transaction records. If so, do such queries become part of the data record subject to whatever policy controls are envisioned. As noted earlier, the issue of logs, and who controls them, has policy implications – both for privacy and for oversight.

In the example above, in addition to whether there should be a government database logging the transaction (and who should control it), there is also the

question of whether the card itself should be designed to record the encounter. Arguments in favor would emphasize the empowerment to the individual from an immutable record in their possession of their encounter with law enforcement and a specific record of what queries were run. Arguments against might include that the card record itself becomes a vulnerability point for privacy – that is, recovery of transaction data from the card itself could be used against the individual at a later date.

Pseudonymity gives policy makers an additional method to control disclosure of identity. For example, in *Hiibel v. Nevada*, [18] the issue was whether a suspect could be compelled to give their name during a *Terry* stop. [19] Among the arguments put forward against disclosure of name was that in today's database world, disclosing one's name is the key to unlocking the digital dossier and may lead to "an extensive fishing trip across government databases." [20] One of the arguments in favor of disclosure is the legitimate interest to determine whether the individual is wanted or dangerous. "Obtaining a suspect's name in the course of a *Terry* stop serves important government interests. Knowledge of identity may inform an officer that a suspect is wanted for another offense, or has a record of violence or mental disorder." [21]

Information systems based on pseudonymity, including the use of smart ID cards, could provide another alternative to meet these same needs. As noted in the example above, there are technical methods for an individual to be matched against a watch list (or any other list) without revealing explicit identifying data. Thus, development of a national ID card based on segmented data and pseudonymous identities could improve privacy over existing methods and still meet security needs. [22]

---

[18] Hiibel v. Nevada, 542 U.S. 177 (2004).

[19] Terry v. Ohio, 392 U.S. 1 (1968) (police can detain a suspect for a reasonable period without probable cause to suspect a crime).

[20] *See* Brief of Amicus Curiae of the Electronic Privacy Information Center (EPIC) and Legal Scholars and Technical Experts at 6, Hiibel v. Nevada, No. 03-5554 (S. Ct. June 21, 2004).

[21] Hiibel, *supra* note 18, at 7.

[22] *See generally* Michael Froomkin, *The Uneasy Case for National ID Cards*, YISP CyberCrime 2004 conference paper (2004); *but cf.* Richard Sobel, *The Demeaning of Identity and Personhood in National Identification Systems*, 15 HARV. J. LAW & TECH. 319 (2002).

Towards a Calculus of Reasonableness

Assuming that anonymization and pseudonymization strategies are employed to separate identity from behavior (or data), and control over data attribution is enforced through privacy appliances, the policy issue still remains when and under what circumstances particular methods of inquiry might reasonably be used on specific data sets, and when and under what circumstances data attribution (or de-anonymization) may reasonably occur.

It is my general thesis that procedural mechanisms relate to the concept of *reasonableness* (both in fourth amendment terms and as that term is more generally understood) through a complex policy calculus involving multiple independent and dependent variables that must be understood individually but considered together and in context. Thus, guiding principles, not rigid standards to be determined *a priori* for every conceivable use, condition or context must be derived within which specific administrative procedures, legislative oversight and judicial intervention and review can be fashioned.

Due Process

Due process is the means for insuring fairness in a system [23] and is essentially a function of four factors: the reasonableness of the *predicate* for action, the *practicality* of alternatives, the *severity and consequences* of the intrusion, and the procedures for *error control*. [24]

Determining the appropriateness of *predicate* requires understanding error rates and assessing related confidence intervals – that is, it requires determining the probative weight of the indicia of suspicion. Confidence interval for policy purposes is simply the acceptable error rate for a given application. For example, the confidence interval for a screening application can be viewed as a function of two competing relationships – the number of false positives (innocents identified) adjusted by the severity of the consequences to the individual and the number of the false negatives (terrorists not identified) adjusted by the consequences to security (and by the potential misallocation of resources from false positives). Determining acceptable confidence intervals for any particular application

---

[23] *See generally* RONALD DWORKIN, LAW'S EMPIRE (1988).

[24] Related policy considerations in determining reasonableness are the interaction and effect of authorization and oversight mechanisms (either executive or legislative) and judicial review in the context of a particular use.

requires assessing the probative value of the predicate procedures – for example, determining whether the observed behavior (or data analysis) reasonably justifies the consequences of the action or not.

The Supreme Court has also explicitly recognized that the requirement for individualized suspicion is not absolute and, where the government interest serves a need beyond routine law enforcement, the *practicality* of requiring a warrant or individualized suspicion is also a relevant factor to be considered:

> [O]ur cases establish that where a Fourth Amendment intrusion serves special governmental needs, beyond the normal need for law enforcement, it is necessary to balance the individual's privacy expectations against the Government's interests to determine whether it is *impractical* to require a warrant or some level of individualized suspicion in the particular context. (emphasis added) [25]

The Court has used the same special needs reasoning in upholding the use of sobriety checkpoints, roving border checkpoints, airport searches, and random drug testing of student athletes. Likewise, policy makers should consider the practicality (or impracticality) of requiring specific procedures or individualized predicate in cases of employing information processing systems for particular uses in counter-terrorism.

Another important factor to be considered is the severity of the intrusion and its *consequences*. Thus, where there is an important state interest, and the intrusion minimal and the consequences slight – for example, a brief stop and referral to a secondary inspection or minimal questioning – the courts are likely to find no fourth amendment violation. [26] In upholding roving traffic checkpoints to enforce immigration laws, the Court stated:

> Against this valid public interest we must weigh the interference with individual liberty that results when an officer stops an automobile and questions its occupants. The intrusion is modest. [27]

---

[25] Treasury Employees v. Von Raab, 489 U.S. 656, 665-666 (1989).

[26] United States V. Martinez-Fuerte**,** 428 U.S. 543, 558-561 (1976).

[27] United States v. Brignoni-Ponce, 422 U.S. 873, 879-880 (1975).

Thus, a legitimate inquiry for policy makers is to determine the severity and consequence of a particular intrusion in light of the state interest. Where there is a significant state interest (for example, preempting terrorist attacks), minimal initial intrusion (for example, an automated review of data), and limited consequences (for example, a routine investigative follow-up that may only include cross-checking against additional data), the courts are likely to uphold the use of advanced information systems to screen data.

The final factor to be considered here in assessing due process is *error detection and correction*. Thus, an important policy and system design consideration is to recognize the inevitable potential for error and to build in robust error compensating mechanisms and procedures. Error, however, including falsely identifying suspects (false positives), is not unique to information systems. To the extent possible, therefore, error detection and error correction mechanisms for automated systems should embrace existing due process procedures, including procedures for administrative and, where appropriate, judicial review.

Additional complications arise, however, when one considers systems and uses in which the subject may never become aware of the intrusion or the consequences of the query. For example, in access control situations where permission for access is denied, the individual is usually on notice that their autonomy has been affected and corrective procedures can be triggered. More difficult is the situation where the data subject never becomes aware of the query or its consequence.

Also unresolved is whether *derived data* (for example, the query itself, links, or other information generated by the analysis itself) or *meta-data* (data about the data, for example, data labels) becomes part of the record and whether it also becomes (or should become) subject to applicable laws – including error correcting procedures – that the underlying data may be subject to. [28]

Privacy and Security Information Needs

Foreign intelligence, counter intelligence, and law enforcement information is and should be available for appropriate domestic security purposes. The significant policy challenge lies in determining when and under what circumstance domestic

[28] *See, e.g.,* the Fair Credit Reporting Act (FCRA), 15 U.S.C. § 1681 et seq., in which queries to credit reporting agencies themselves become part of the underlying credit report and subject to the requirements of the Act.

intelligence access should be allowed to routinely-collected government data or to commercially available (or other third party) data. [29]

Unlike intelligence or law enforcement data, government held data collected in the normal course of providing government services is generally subject to restriction for other uses or data matching by the Privacy Act of 1974. However, the Privacy Act has broad exceptions for data matching and inter-agency sharing for national security and law enforcement purposes, thus, for practical purposes there may be no restrictions on secondary uses for domestic security applications. Thus, a threshold question is whether there should be any additional procedural protections or guidelines for secondary uses of routinely-collected data that subsequently comes within the national security and law enforcement exceptions.

A more difficult question, however, involves deciding whether government should have access to, and use of, privately held third party data, particularly data from commercial data sources, and, if so, under what circumstances and what constraints.

That the government should, and will ultimately, have access to this data in some form seems foregone. As already noted, it would be an unusual polity that demanded accountability from its representatives to prevent terrorist acts yet denied them access to available tools or data. Thus, it is the procedures under which such access should be allowed that need to be defined. Here, developing clear goals and concomitant policy guidelines, requiring that the nexus between particular types of information and its use and value for counter-terrorism be clearly identified or articulated, and mandating strict oversight and review procedures, are needed to ensure that appropriate government access to potentially useful information is possible while still protecting civil liberties.

Among the policy tools for dealing with access questions is the use of *categorization* to designate certain information sources or types subject to (or exempt from) particular procedures. [30] However, these procedures may require

---

[29] *See generally* [Daniel Solove, *Digital Dossiers and the Dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083 (2002)]; Second Markle Report, *supra* note 8, at 30-37.

[30] For example, under Homeland Security Presidential Directive/HSPD-6, 39 Weekly Comp. Pres. Doc. 1234 – 1235 (Sept. 16, 2003), certain information is classified as "Terrorist Information" for handling under specific procedures, and under NSA/CSS United States Signal Intelligence Directive 18 ("USSID 18") (July 27, 1993) certain information about U.S. persons gathered during foreign intelligence SIGINT collection is subject to extraordinary procedures ("minimization").

pre-determining what information is relevant or subject to special handling – something that is not always possible in counter-terrorism analysis. Thus, it might ultimately be more practical to control the proposed *use* of data directly through a system based not on categorizing data itself by type but in managing and tracking authorized purpose for use at the time of access, subject to strict technical controls.

Any policy based solely on procedures for pre-designating information as relevant for counter-terrorism or to be managed under special rules needs to recognize three problems: first, it may be difficult to identify certain information as relevant except in the context of an ongoing investigation or in response to a particular threat; second, designations based on place, method of collection or nationality of subject may be outmoded in the context of a worldwide, distributed, networked database and communications infrastructure, and, third, classifying data into categories that do not relate to the purpose of the original data collection may not be possible after the fact. In any case, if categorization strategies are to be used, they require that some technical means for data-labeling be incorporated into systems design.

Specific statutes already exist to protect particular classes of information deemed sensitive. These statutes generally require that use of these types of information conform to particular procedures. For example, census data, medical records, educational records, tax returns, cable television records, video rental, among others are all subject to their own statutory protection, usually requiring an elevated level of predicate, for example, a warrant or court order based on probable cause instead of a subpoena based on mere suspicion, to gain access. Although some of these designations and procedures may need review in the context of domestic security, the general approach of dealing with particularly sensitive personal data by providing additional procedural protections in certain circumstances is workable and could be applied to identification, data aggregation and analysis, and collection systems assuming certain technical features, in particular, for data-labeling, are developed to allow data categorization to be maintained when data is shared or exchanged.

Policy decisions to *a priori* designate higher standards for sensitive information (or declaring certain information off-limits for certain uses) involve the same policy considerations and problems governing pre-designating information or sources as relevant for counter-terrorism analysis. Additionally, there may be a trade-off between the sensitivity of information and its relevance for counter-terrorism that must be taken into account when such designations are made. If it turns out that certain information deemed sensitive by its nature, for example,

financial records, is also quite specifically useful for counter-terrorism (for example, following the money trail), choice of policies (and technical features to support such choice) need to be developed taking both needs into account. Again, strategies based on directly authorizing a particular use at the time and under the circumstances of access, monitored though strict technical controls, may need to be considered rather than continuing to rely solely on pre-determined categorization of data sensitivity or pre-determined usefulness.

Some are concerned that the use of automated data analysis techniques to allocate intelligence and law enforcement resources is not *particular* enough to protect due process or meet fourth amendment requirements. However, as already discussed, there is no absolute constitutional requirement for individualized suspicion and no presumptive constitutional problem with pattern-matching. [31] Nevertheless, for purposes of determining policy, it may be appropriate to address these concerns and recognize that different procedures may be appropriate for different query methods depending, for example, on whether they are subject-, link- or pattern-based.

Subject- and link-based queries generally raise the same issues as outlined above in the general discussion of scope – that is, what data can be accessed and under what circumstance. For some, however, pattern-matching also raises additional concerns relating to individualized suspicion.

The policy question is whether there should be additional technical or procedural protections to be applied for pattern-based queries based on the perception that these methods are potentially more intrusive or problematic. Some have recommended that additional procedures be applied to processes using pattern-analysis derived from data-mining, including specific pre-approvals and stricter oversight. [32] Technical features to support such procedures would be required design features. For example, procedures requiring additional administrative (or judicial) approvals for specific disclosures(for example, of identity) would require technical controls to enforce selective revelation of information.

---

[31] See discussion in *Due Process* above.

[32] *See* Safeguarding Privacy in the Fight Against Terrorism, The Report of the [Department of Defense] Technology and Privacy Advisory Committee (TAPAC) (March 2004).; Rosenzweig, *supra* note 3; Daniel Gallington, *Better Information Sharing and More Privacy in the War on Terrorism – A New Category of Information is Needed* (Potomac Inst. for Pol'y Studies, July 29, 2003).

There is no magic policy formulation that perfectly balances the complex interactions between these many variables. What is called for instead is an analytic framework together with guiding principles – following what have I have called a *calculus of reasonableness* – that can inform the debate as these issues come up in varying contexts as new technologies develop and challenge existing doctrine or precepts in previously unforeseen circumstances. Thus, judicious distinction between when rules (*what* you can do), procedures (*how* you can do something) and guidelines (constraints or *limits* within which you act to accomplish some goal) are appropriate requires understanding these complexities and recognizing the inchoate nature of any solutions given the rapid pace of technological development and the evolving nature of the threat.

Threat Environment and Reasonableness

Reasonableness may also vary depending on the *threat level* and the particular security need. System bias towards more false positives and less false negatives may be reasonable under certain high threat conditions or in applications requiring high security. In other circumstances, system bias towards fewer false positives and more false negatives may be more appropriate.

Policy considerations are also domain dependent. For example, decision heuristics used during the development of traditional defense systems are generally inappropriate for domestic security applications. In designing military defenses, the bias is to eliminate any false negatives by accepting additional false positives. On the battlefield, it is better to have a low threshold for triggering a response then to risk not being prepared. However, in the context of a civilian population, false positives or false alarms may be as destructive of certain values (including security) as are false negatives by undermining trust in the system or creating intolerable burdens. Too many false positives and the resulting misallocation of resources will undermine both popular and political support for security measures as well as impact security itself.

Thus, because of the dynamic nature of the threat and security requirements, no system (technical or procedural) should be contemplated that is either constantly at ease or constantly at general quarters. Flexible systems and policy guidelines that can adapt proportionally to perceived threats faster and more efficiently are required.

It also seems premature to burden either policy development or technical research and development with a requirement to determine *a priori* what policy rules will apply in every conceivable use case. Technical development processes are not

generally amenable to predictable development paths where ongoing research is in its early stages. An iterative process using value sensitive design procedures can help guide technical and policy development to achieve both required outcomes – security and privacy. However, achieving this requires joint participation, not knee-jerk opposition.

Nevertheless, guiding policy principles can be developed even without knowing all the potential technologically enabled opportunities or constraints based on a deeper understanding of these process needs. Policy develops rules of general applicability while judicial review examines cases of specific application; systems design must accommodate both.

Conclusion

The development and use by government of advanced identification, aggregation and analysis, and collection technologies in domestic security applications raise legitimate privacy concerns. Nevertheless, such development and eventual use is inevitable and strategies premised on opposition to research or banning certain uses or deployments through law alone are destined to fail, and, in any case, provide little security and brittle privacy protection. Protecting civil liberties requires that value sensitive development strategies be used to design technical features and systems that enable familiar due process mechanisms to function.

The mythology of privacy and fear of technology should not keep us from opportunities to improve both security and privacy. Reconciling competing requirements for security and privacy requires an informed debate in which the nature of the problem is understood in the context of the interests at stake, the technologies at hand for resolution, and the existing resource constraints. Key to resolving these issues is designing a policy and information architecture that can function together to achieve both outcomes.

THIS PAGE INTENTIONALLY LEFT BLANK

Introduction to:

CHAPTER 23: Designing Technical Systems to Support Policy: Enterprise Architecture, Policy Appliances, and Civil Liberties

***in* 21ST CENTURY ENABLING TECHNOLOGIES AND POLICIES FOR COUNTER-TERRORISM (R. Popp and J. Yen, *eds.*, Wiley–IEEE Press, June 2006)**

K. A. TAIPALE *

I. Introduction.

It has become cliché to describe the relationship between security and liberty as one requiring the achievement of some optimal balance between two competing and irreconcilable needs. But such cliché is metaphorically misleading. There is no fulcrum point – as is implicit in the balance metaphor – at which point the correct amount of security and liberty can be achieved. Security and liberty are not dichotomous rivals to be traded one for another in a zero sum game as the notion of balance suggests or as the *enragés* of the public debate would have. Rather, security and liberty are dual obligations of a liberal republic and each must be maximized within the constraints imposed by the other. [1]

The events and subsequent investigations of 9/11 have highlighted the national security need for better information management, and for new technologies and techniques to improve collection, information sharing, and data analysis in counterterrorism applications. The need to manage vast data volumes and better "connect the dots" is uncontroverted and has been explicitly set out in a series of executive orders, presidential directives, national strategy documents, committee reports, and legislation. [2]

However, emergent information technologies that can enable such improved information management and analysis processes – technologies like those described in this book – also challenge traditional policy doctrines and legal structures premised in part on protecting individual liberty by maintaining *privacy* through the "practical obscurity" arising from inefficiencies in information acquisition, access, management, and analysis. [3] Thus, to some observers, improving the ability of government agencies to

* *Kim Taipale, BA, JD (New York University), MA, EdM, LLM (Columbia University), is the executive director of the Center for Advanced Studies in Science and Technology Policy. Mr. Taipale is also a senior fellow at the World Policy Institute, an adjunct professor of law at New York Law School, and an associate of the Markle Foundation Task Force on National Security in the Information Age.*

"connect the dots" is seen to be in political conflict with the notion of keeping the power to "connect the dots" out of any one hand, particularly that of the central government. [4] The result, as evidenced in the public debate, is a presumed implacable antagonism between security and privacy.

Fortunately, we do not need to resolve this Jacobin discordance in order to design information systems with technical features that can support a broad range of policies to mitigate privacy concerns and still meet security needs. Indeed, there is no inherent *technical* design conflict at all between security and privacy as the technical features required to support privacy policy are in large part the same technologies required to meet operational information assurance and data security needs in national security or law enforcement information sharing applications. Both national security and privacy policy require (i) that shared information be *useful* (that is, that data is accurate, reliable, and timely, and that it can be up-dated or corrected as needed), and (ii) that information be *used appropriately* according to policy rules. Technical features to support these concordant policy needs in information systems include rules-based processing, selective disclosure, data quality assurance, error correction, and strong authorization, logging, and audit functions (to control and record what information goes where, under what constraints, and who has access to it).

This chapter discusses policy-enabling systems design based on an *enterprise architecture* to support *knowledge management* (a lifecycle approach to managing information from production to consumption as a product to support business process needs) and *due process* (procedures to protect civil liberties). This architecture includes *policy appliances* (technical control mechanisms to enforce policy rules and ensure accountability in information systems), [5] interacting with *smart data* (data that carries with it contextual relevant terms for its own use) and *intelligent agents* (queries that are self-credentialed, authenticating, or contextually adaptive). See Figures 1 and 2. It is beyond the scope of this chapter to detail specific technology development or current research avenues in depth, or to exhaustively examine information management strategies or developments. Rather, this chapter provides an overview of the relationship between emerging policy process models and technical design choice in order to better understand the interdependence of technical architecture and policy implementation.

[PAGES 4 through 37 DELETED]

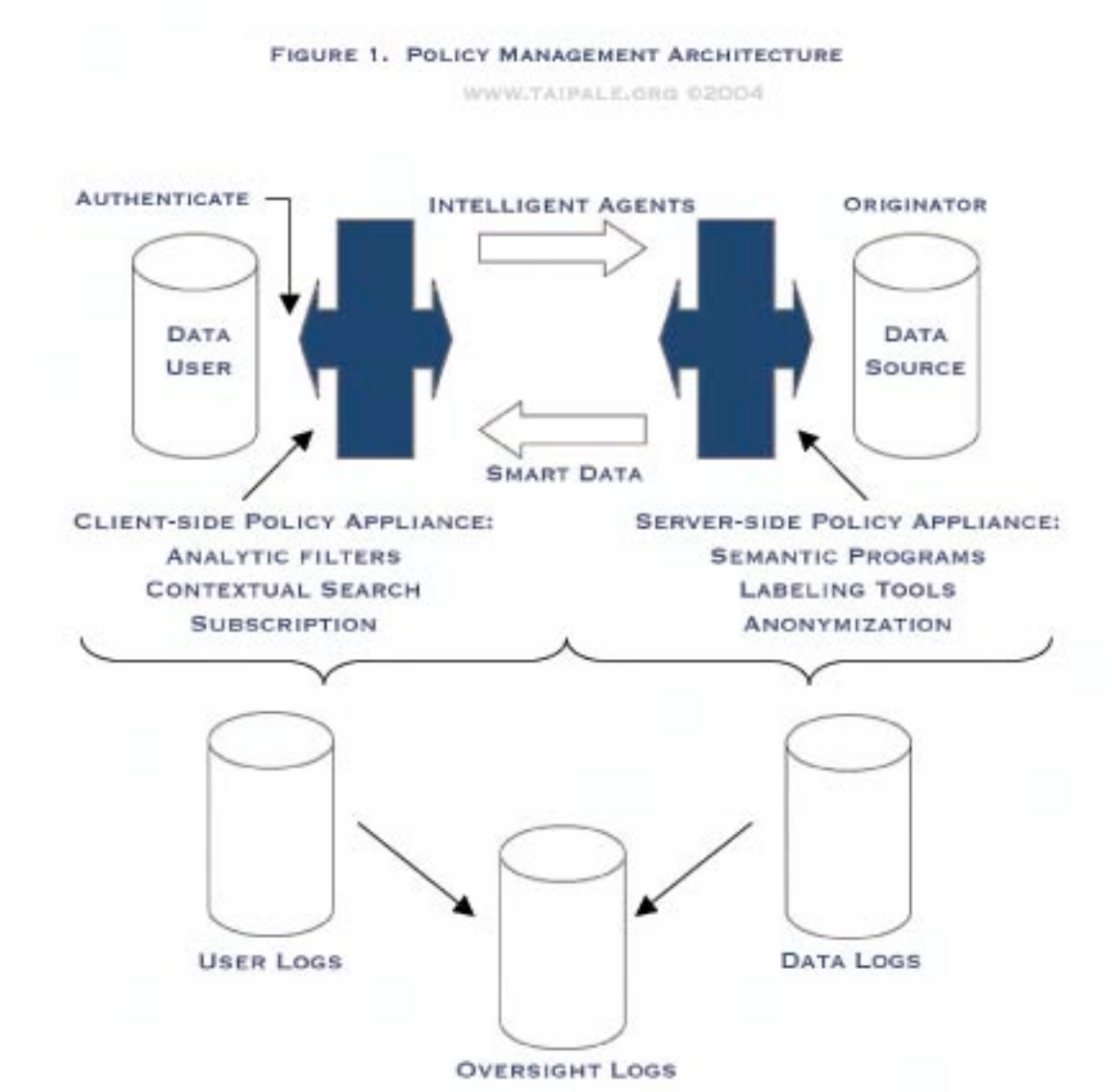


FIGURE 1. POLICY MANAGEMENT ARCHITECTURE. An *enterprise architecture* for *knowledge management* (an information product approach) and *due process* (civil liberties protections) that includes *policy appliances* (technical control mechanisms to enforce policy rules and ensure accountability) interacting with *smart data* (data that carries with it contextual relevant terms for its own use) and *intelligent agents* (queries that are self-credentialed, authenticating, or contextually adaptive).

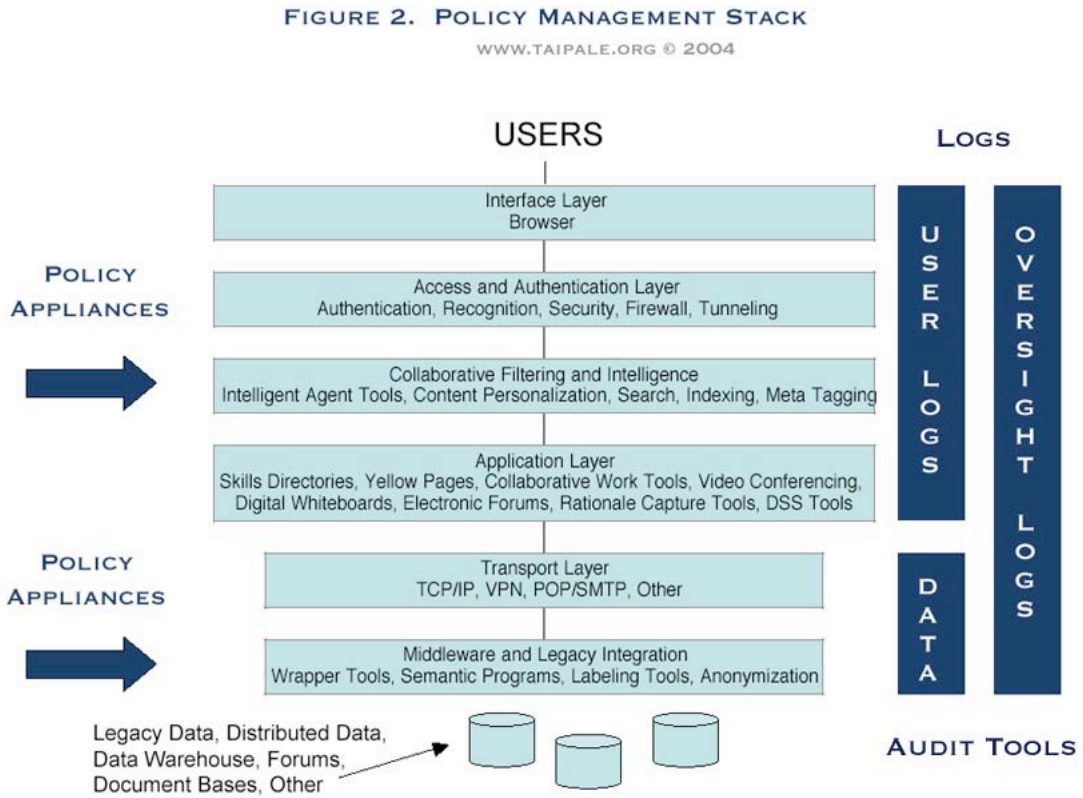


FIGURE 2. POLICY MANAGEMENT STACK. An *enterprise architecture* showing *policy appliances* (technical control mechanisms to enforce policy rules and ensure accountability) and *logging functions* in network stack relationship.

ENDNOTES

---

[1] K. A. Taipale, *Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data*, 5 Colum. Sci. & Tech. L. Rev. 2 at n. 6 (1993) [hereinafter, Taipale, *Data Mining*], citing Thomas Powers, *Can We Be Secure and Free?* 151 Public Interest 3, 5 (Spring 2003); *see also* K. A. Taipale, *Technology, Security, and Privacy: The Fear of Frankenstein, the Mythology of Privacy, and the Lessons of King Ludd,* 7 Yale J. L. & Tech. at 5 (Dec. 2004) [hereinafter, Taipale, *Frankenstein*].

[2] *See, e.g.,* Executive Order 13356 (2004); Presidential Directive, *Strengthening Information Sharing, Access, and Integration B Organizational, Management, and Policy Development Structures for Creating the Terrorism Information Sharing Environment*, June 2, 2005; National Strategy for Homeland Security at 55 (2002); The National Commission on Terrorist Attacks Upon the United States, The 9/11 Report §13.3 (2004); Intelligence Reform and Terrorism Prevention Act of 2004, Pub. L. No.108-458, §1016.

[3] *See Department of Justice v. Reporters Committee for Freedom of Press*, 489 U.S. 749, 780 (1989) (recognizing a legally protected interest in the "practical obscurity" of inefficient paper-based record systems).

[4] *See, e.g.,* Kathleen Sullivan, *Under a Watchful Eye: Incursions on Personal Privacy*, in THE WAR ON OUR FREEDOMS (Richard C. Leone *et al.* eds., 2003).

[5] See Taipale, *Frankenstein*, *supra* note 1 at 56-58 (discussing "privacy appliances" to enforce rules and provide accountability). The concept of privacy appliances originated with the DARPA Total Information Awareness project. *See* Presentation by Dr. John Poindexter, Director, Information Awareness Office (IAO), Defense Advanced Research Projects Agency (DARPA), at DARPA-Tech 2002 Conference, Anaheim, CA (Aug. 2, 2002); ISAT 2002 Study, *Security with Privacy* (Dec. 13, 2002); and *IAO Report to Congress regarding the Terrorism Information Awareness Program* at A-13 (May 20, 2003) in response to Consolidated Appropriations Resolution, 2003, No.108-7, Division M, §111(b) [signed Feb. 20, 2003].

[ENDNOTES 6-38 DELETED]

THIS PAGE INTENTIONALLY LEFT BLANK

Trends & Controversies

K.A. Taipale, *The Trusted Systems Problem: Security Envelopes, Statistical Threat Analysis, and the Presumption of Innocence*

IEEE INTELLIGENT SYSTEMS, pp. 80-83, Vol. 20, No. 5 (Sep./Oct. 2005)

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.

THIS PAGE INTENTIONALLY LEFT BLANK

Whispering Wires and Warrantless Wiretaps: Data Mining and Foreign Intelligence Surveillance

N.Y.U. REV. L. & SECURITY, NO. VII (Spring 2006)

K. A. Taipale *

Introduction.

In the current debate over whether the President has the inherent power to authorize the National Security Agency to monitor international communications with suspected terrorists, one thing is clear—even the most strident opponents concede the need to identify and monitor the communications of terrorists and stop them before they can act. [1]

Preempting terrorist attacks requires uncovering information useful to anticipate and counter future events. [2] Automated data analysis technologies can help by monitoring communications and revealing evidence of organization, relationships, or other relevant patterns of behavior indicative or predictive of potential threats, thus allowing law enforcement or security resources to be focused more effectively on likely targets.

This essay examines certain implications of employing these techniques for foreign intelligence surveillance and suggests that the Foreign Intelligence Surveillance Act ("FISA") [3] is inadequate to address recent technology developments, including: the

* *Mr. Taipale is the founder and executive director of the Center for Advanced Studies in Science and Technology Policy, a senior fellow at the World Policy Institute, and an adjunct professor of law at New York Law School.*

1 *See, e.g.*, Adam Nagourney, *Seeking Edge In Spy Debate*, N.Y. TIMES (Jan. 23, 2006) ("'We all support surveillance … ,' [Senator John] Kerry said."); and Statement released by U.S. Senator Patrick Leahy (Feb. 15, 2006) ("We all agree that we should be wiretapping al Qaeda terrorists").

2 *See generally* U.S. Department of Justice, *Fact Sheet: Shifting from Prosecution to Prevention, Redesigning the Justice Department to Prevent Future Acts of Terrorism* (May 29, 2002); and The National Security Strategy of the United States (Sep. 17, 2002) ("[T]he United States ... will not hesitate ... to exercise our right of self defense by acting preemptively against such terrorists", p. 6).

3 Codified at 50 USC §§1801-1811, 1821-29, 1841-46, and 1861-62. *See also* note 20 *infra* discussing the proposed Terrorist Surveillance Act of 2006 and the proposed National Security Surveillance Act of 2006 (both introduced to the Senate on Mar. 16, 2006). Both bills would provide limited additional statutory authority for electronic surveillance of suspected terrorists in the United States: however, the TSA would allow the President to authorize such surveillance subject to Congressional oversight; while the NSSA would require FISA court approval, authorization, and oversight.

transition from circuit-based to packet-based communications; the globalization of communications infrastructure; and the development of automated monitoring techniques, including data mining and traffic analysis. [4]

Background.

Although this essay discusses how FISA is challenged by technology developments, the suggestion that FISA procedures are inadequate to encompass certain aspects of foreign intelligence surveillance is not new, nor unique to technical developments.

Testifying before the Church Committee in 1975, then-Attorney General Edward Levi suggested that FISA should include provisions for the approval of "programs of surveillance" in foreign intelligence situations where "by [their] very nature [they] do not have specifically determined targets" and where "the efficiency of a warrant requirement would [therefore] be minimal." However, Congress passed FISA in 1978 without including any provisions for such programmatic approvals. [5]

In a recent essay, Judge Richard A. Posner opined that FISA "retains value as a framework for monitoring the communications of known terrorists, but it is hopeless as a framework for detecting terrorists. [FISA] requires that surveillance be conducted pursuant to warrants based on probable cause to believe that the target of surveillance is a terrorist, when the desperate need is to find out who is a terrorist." [6]

FISA is inadequate.

FISA did not anticipate the development of global communication networks or advanced technical methods for intelligence gathering. FISA provides a cumbersome mechanism requiring individual application to the FISA court for authorization to target a specific individual or source based on showing a connection to a foreign power or foreign terrorist group. [7] Although FISA permits such applications to be made after the fact in certain

---

[4] Although details of the NSA program are classified, press reports suggest that data mining and traffic analysis technologies are being used. *See, e.g.,* Shane Harris, *NSA spy program hinges on state-of-the-art technology,* NAT'L J. (Jan. 20, 2006). For an overview of NSA technical capabilities, *see generally* Patrick Radden Keefe, CHATTER (2005); James Bamford, *Big Brother is Listening*, THE ATLANTIC (Apr. 2006). For a general discussion of data mining and counterterrorism, *see generally* K. A. Taipale, *Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data*, 5 COLUM. SCI. & TECH. L. REV. 2 (Dec. 2003); Mary DeRosa, *Data Mining and Data Analysis for Counterterrorism*, Center for Strategic and International Studies Press (Mar. 2004). *See also* James E. Lewis, *Domestic Communications Surveillance: Right Decision, Wrong Rules*, CSIS (January 2006) (describing the differences between the FBI and NSA capabilities and needs in the context of domestic intelligence).

[5] *See* John R. Schmidt, *A historical solution to the Bush spying issue*, CHIC. TRIB. (Feb. 12, 2006) (describing Levi's advocacy of statutory process for programmatic approvals in foreign intelligence). *See also* Sen. Arlen Specter, *Statement Introducing the National Security Surveillance Act of 2006* (Mar. 16, 2006) (recounting same).

[6] Richard A. Posner, *Commentary: A New Surveillance Act*, WALL ST. J. A16 (Feb. 15, 2006).

[7] In the case of a US person, FISA requires probable cause to believe that the target is an "agent of a foreign power," §1801(b), and that the person's activities "involve or are about to involve" a violation of

cases, it does not provide a mechanism for programmatic pre-approval of technical methods like automated data analysis or filtering that may be the very method for uncovering such a connection.

From circuit-based to packet-based communication networks.

To understand the need for applying automated data analysis technologies to foreign intelligence surveillance requires not just recognizing the vast data volumes potentially subject to monitoring—imagine for a moment the capture of an al Qa'ida laptop in the battlefield of Afghanistan containing hundreds or thousands of phone numbers [8] or email addresses—but also an understanding of the nature of modern communications networks.

Thirty years ago when FISA was being drafted it made sense to speak exclusively about the interception of a targeted communication—one in which there were usually two known ends and a dedicated ("circuit-based") communication channel that could be "tapped." In modern networks, however, data and increasingly voice communications are broken up into discrete packets that travel along independent routes between point of origin and destination where these fragments are then reassembled into the original whole message. Not only is there no longer a dedicated circuit, but individual packets from the same communication may take completely different paths to their destination. To intercept these kinds of communications, filters ("packet-sniffers") and search strategies are deployed at various communication nodes to scan and filter all passing traffic with the hope of finding and extracting those packets of interest and reassembling them into a coherent message. Even targeting a specific message from a known sender requires intercepting (i.e., scanning and filtering) the entire communication flow. Were FISA to be applied strictly according to its terms prior to any "electronic surveillance" of foreign communication flows passing through the US or where there is a substantial likelihood of intercepting US persons, then no automated monitoring of any kind could occur. [9]

---

the criminal laws of the United States, § 1801(b)(2)(B), or are activities in preparation for sabotage or "international terrorism" on behalf of a foreign power, § 1801(b)(2)(C).

[8] Note that even a mid-level al Qa'ida operative may carry four or five cell phones with multiple SIM cards supporting several numbers for each phone. Training material describing how to avoid electronic surveillance is widely available on Jihadist web sites. *See, e.g.,* Jeffrey Pool, *Technology and Security Discussions on the Jihadist Forums: Producing a More Savvy Next Generation*, Spotlight on Terror, Vol. 3, No. 10 (Oct. 11, 2005).

[9] The retroactive warrant procedures in FISA also do not work here since those communications 'intercepted' but not selected for further analysis by definition would not meet the requirement for a warrant application (i.e., no probable cause), and there would be no independent predicate for probable cause for those communications selected for follow-up as a result of filtering, unless there was programmatic authorization of the filtering in the first place. Although procedures under FISA allow for the retention and use without a warrant of US person communications "with foreign intelligence value" if it is collected collateral to a legitimate foreign intelligence intercept, in practice such information is not deemed adequate to establish predicate for targeting such a person where the initial intercept is pursuant to a general surveillance program and follow-up investigation is required to determine if probable cause exists.

The globalization of communications.

A further problem arises because FISA is triggered by foreign intelligence collection conducted "within the United States" or against "U.S. persons." [10] Advances in information technology together with the borderless nature of terrorist threats and global communications has made place-of-collection and U.S. personhood an increasingly unworkable basis for controlling the collection of intelligence. Indeed, because of packet-based communication technologies like VoIP and the use of proxy servers, it may no longer even be technically possible to determine exactly when a communication is taking place "within the United States" and no practical means exists to determine if a particular participant is a U.S. person or not until after further investigation. [11] FISA does not account for this. [12]

Automated analysis: data mining and traffic analysis.

Automated screening can monitor data flows to uncover terrorist connections or terrorist communication channels without human beings ever looking at anybody's emails or listening in on their phone calls. Only when the computer identifies suspicious connections or information do humans get involved.

It is beyond the scope of this essay to explore all the different analysis techniques that can be applied to the monitoring of terrorist communications but two examples show the range of activity possible: content filtering and traffic analysis.

Content filtering is used to search for the occurrence of particular words or language combinations that may be indicative of terrorist communications. A simple example of this would be to screen for messages to or from known terrorist sources containing the words "nuclear weapon". Actual search algorithms are, of course, much more complex and sophisticated and can employ artificial intelligence, machine learning, and powerful statistical methods such as Bayesian analysis to identify potential threats.

Traffic analysis is the examination of traffic patterns—message lengths, frequency, paths, etc.—of communications without examining the content of the message (traffic analysis can be used even where content is encrypted). Traffic analysis can reveal patterns of organization, for example, by measuring "betweeness" in email traffic or other communications. By looking for patterns in traffic these techniques, together with social network theory, can help identify organizations or groups and the key people in them. These methods can uncover terrorist organization and reveal activity even if they are

---

[10] *See* § 1801(f).

[11] Indeed, certain networks are specifically designed to conceal such information by chaining proxy servers. *See, e.g.*, the TOR Network, which uses *onion routing* to provide near anonymous communication capability and is specifically designed to avoid traffic analysis monitoring (see text accompanying note 13, *infra*, discussing traffic analysis).

[12] Nor does it account for the fact that even wholly "foreign" communications today may pass through physical nodes located "within the United States."

communicating in code or only discussing the weather. [13]

'Programs of surveillance' are not general warrants.

It is important to remember that we are not contemplating the use of these technologies in an undirected fashion in the manner of a general warrant to examine all communication flows. [14] Rather, we argue for a mechanism for programmatic approval where these technologies are applied in the first instance against known or reasonably suspected foreign terrorist communication sources—that is, against legitimate foreign intelligence targets not subject to FISA and not requiring a warrant [15] —and are used to automate the process of looking for connections, relationships, and patterns for further follow-up investigation.

These technologies are not a general method for finding terrorists by monitoring all global communications with no starting point, nor for determining guilt or innocence. Rather, they are powerful tools to help better allocate law enforcement and security resources to more likely targets. If the initial automated process identifies potentially suspicious connections—including US persons or sources—some additional monitoring or follow up investigation must occur to determine if that initial "suspicion" is justified.

The problem with FISA is that it contemplates only a single threshold for authorizing interception within the US or targeting of US persons—probable cause. In the case of automated monitoring, there must be some approved procedure that identifies potential threats and allows for some limited follow-up—either additional automated monitoring or human investigation—to determine if indeed the initial indicia of suspicion are justified. If so, then existing FISA procedures can be followed to "target" that US person or source.

What is needed then, is the electronic surveillance equivalent of a *Terry* [16] stop—in this case an authorized period for follow up monitoring or investigation of initial suspicion derived from automated monitoring. If the suspicion is not justified on follow-up,

---

[13] *See, e.g.,* Hazel Muir, *Email Traffic Patterns can Reveal Ringleaders*, NEW SCIENTIST (Mar. 27, 2003). Note, however, that certain technologies exist to counter traffic analysis, see, for example, the discussion of the TOR Network in note 11, *supra*. For a discussion of the use of social network theory in counterterrorism analysis, *see* Patrick Radden Keefe, *Can Network Theory Thwart Terrorists?* N.Y. TIMES (Mar. 12, 2006).

[14] Such undirected uses are called "fishing expeditions" by critics and "drift nets" by intelligence professionals. It is the use of general warrants by the English that led in part to the American Revolution, *see, e.g.,* O.M. Dickerson, *Writs of Assistance as a Cause of the Revolution*, THE ERA OF THE AMERICAN REVOLUTION (Richard Morris ed. 1939), and to enactment of the Fourth Amendment, *see* Edward Corwin, THE CONSTITUTION AND WHAT IT MEANS TODAY (1978, 1920).

[15] For example, Abu Musab Zarqawi's cell phone number or a known al Qa'ida communication network in Pakistan, Saudi Arabia, or Hamburg. According to a classified German intelligence report, 206 international telephone calls were known to have been made by the leaders of the 9/11 hijacking plot after they arrived in the United States — including 29 to Germany, 32 to Saudi Arabia and 66 to Syria. *See* John Crewdson, *Germany says 9/11 hijackers called Syria, Saudi Arabia*, CHIC. TRIB. (Mar. 8, 2006).

[16] Terry v. Ohio, 392 U.S. 1 (1968) (holding that police can detain a suspect for a reasonable period without probable cause to suspect a crime).

monitoring is discontinued, however, if suspicion is reasonable then monitoring continues under the programmatic approval. If there is probable cause to suspect that the target is actively engaged in terrorism or is an agent of a foreign terrorist group, then a FISA warrant issues to target that US person or source.

Conclusion: A mechanism for programmatic approvals.

What is needed is a legal mechanism for pre-approving such methods so that legitimate foreign intelligence can be exploited and further threats identified for follow up investigation. With programmatic approval of initial monitoring, existing FISA warrant procedures could then be followed for targeted monitoring of identified US persons or sources in appropriate cases based on the results of the initial automated selection done subject to such programmatic approval.

It is beyond the scope of this essay to recommend particular mechanisms or standards for authorizing such programmatic approvals. It has been argued that courts are ill suited, and may be constitutionally prohibited, from such an oversight role [17] and that a statutory executive [18] or legislative [19] authorization or oversight body should be created.

By all means let us debate who should have the authority to authorize and oversight such intelligence gathering programs, but let us not forget that *someone* must, and the existing mechanisms are inadequate. [20]

---

[17] *See, e.g.,* David B. Rivkin, Jr. and Lee A. Casey, *Commentary: Inherent Authority*, WALL. ST. J. A16 (Feb. 8, 2006) ("The federal courts can only adjudicate actual cases and controversies; they cannot offer advisory opinions.")

[18] *See, e.g.,* Posner, *supra* note 6.

[19] *Compare* the proposed Terrorist Surveillance Act of 2006, *infra* note 20, that would give the President authority to approve the NSA program subject to oversight by special Congressional committees, *with* the proposed National Security Surveillance Act of 2006, *infra* note 20, that would extend FISA court (FISC) jurisdiction to approve, authorize, and oversight "programs of electronic surveillance."

See generally Sheryl Gay Stolberg, *Senate Chairman Splits With Bush on Spy Program*, N.Y. TIMES (Feb. 18, 2006); Shaun Waterman, *Senators to publish bills on NSA wiretap*, UPI (Mar. 8, 2006); Scott Shane and David D. Kirkpatric, *G.O.P. Plan Would Allow Spying Without Warrants*, N.Y. TIMES (Mar. 9, 2006).

[20] *See* K. A. Taipale & James Jay Carafano, *Commentary: Fixing Foreign Intelligence Surveillance*, WASH. TIMES (Jan. 24, 2006).

On March 16, 2006, Senators Mike DeWine (R-OH), Lindsey Graham (R-SC), Chuck Hagel (R-NE), and Olympia Snowe (R-ME) introduced the Terrorist Surveillance Act of 2006, under which the President would be given certain additional limited statutory authority to conduct electronic surveillance of suspected terrorists in the United States subject to enhanced Congressional oversight. *See* Katherine Shrader, *GOP Senators Introduce Eavesdropping Bill*, AP (March 16, 2006). Also on March 16, 2006, Senator Arlen Specter (R-PA) introduced The National Security Surveillance Act of 2006, which would amend FISA to provide FISA court jurisdiction to review, authorize, and oversight "electronic surveillance programs".

Issues & Ideas

FISA's Failings

By Shane Harris

- The law on electronic eavesdropping is behind the times and on shaky legal ground.
- The borderless nature of terrorism and global communications has blurred the foreign/domestic divide.
- Can lawmakers learn enough about the NSA's domestic eavesdropping to control it?

Among the threats facing the National Security Agency are Al Qaeda, the Iraqi insurgency, and eBay. Yes, eBay, the online auction house. Not because its members sell state secrets, but because of a company that eBay purchased last year—Skype.

Skype is an online service that lets people converse through their computers. Its 75 million users place voice calls over the Internet. The calls sound clear. They're free, because phone carriers aren't used. And because of the Internet's diffused architecture and its facility for privacy, Skypesters' identities, their locations, and the substance of their conversations can be undetectable. This is not what the NSA's worldwide eavesdroppers want to hear.

Skype and other widely used Internet communications devices, including e-mail, threaten the NSA's ability to gather intelligence and to do so legally. For more than four years, without warrants and by order of President Bush, the agency has hunted for terrorists by intercepting communications between people in the United States and people abroad possibly connected to terrorism. The legality of that order is being hotly debated in Congress. Bush says that the 27-year-old Foreign Intelligence Surveillance Act, which governs domestic eavesdropping for intelligence purposes, doesn't adequately address Internet-based communications. In the opinion of some legal scholars and intelligence practitioners, lawmakers haven't faced this fact. Until they do, the NSA remains on shaky legal ground and at a strategic disadvantage against terrorists, who may rely on the Internet above all other tools for plotting their attacks.

Technology Law Expert

The FISA law **"did not anticipate** the development of global communications networks."
—Kim Taipale

When FISA became law in 1978, even rudimentary e-mail was years away from use. The law "did not anticipate the development of global communications networks," according to Kim Taipale, a technology law scholar and a member of the Task Force on National Security in the Information Age, a nonpartisan panel supported by the Markle Foundation that has produced widely praised assessments of technology's role in counter-terrorism.

"Thirty years ago … it made sense to speak exclusively about the interception of a targeted communication—one in which there were usually two known ends and a [phone line] that could be 'tapped,' " Taipale writes in an upcoming essay for the *New York University Review of Law and Security*. Phone calls travel over a dedicated circuit, in easily traced paths. But Internet communications are broken down into discrete units, called packets, that swirl through the global network along different, sometimes circuitous routes before being reassembled at their destination. If placing a phone call can be likened to mailing someone a letter, sending an e-mail is like cutting that letter into 50 pieces and dividing them among several couriers, and then asking the couriers to reassemble the letter upon delivery.

To intercept packets, devices called "sniffers" are placed at various communication nodes to scan traffic as it passes, looking for interesting packets and, hope-

fully, reassembling them coherently. If the NSA has an e-mail address to target, catching the message is relatively simple—put a sniffer near the user's Internet service provider.

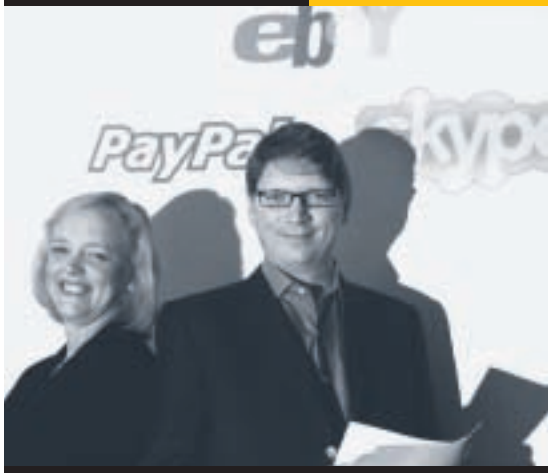
But the NSA's warrantless eavesdropping program also involves looking for suspicious patterns in a sea of communications. The NSA might not know what it's looking for, so it has to examine a lot of data. Put another way, "If you can't find the needle, you have to take the haystack," said Doug Graham, a security expert with BusinessEdge Solutions and a former surveillance-systems operator with the Royal Air Force.

The NSA's program is fundamentally unsuited for FISA, administration officials contend. The law requires the government to develop a reasonable basis to believe that a specific individual or a source is an agent of a foreign power, and then apply to a special court for authorization to pursue that target. It's a "cumbersome mechanism" that doesn't provide for technical methods such as automated packet sniffing, which could uncover suspicious activity in the first place, Taipale writes.

Legislative efforts to address the NSA program have come from Sen. Mike DeWine, R-Ohio, and Senate Judiciary Committee Chairman Arlen Specter, R-Pa., who rejects the Bush administration's argument that the FISA court has no jurisdiction over what the White House calls the "terrorist surveillance program." Specter wants the court to review the program, and DeWine would place limits on NSA's warrantless work. But neither senator proposes making significant amendments to FISA itself, and they provide no new approach that would make new kinds of eavesdropping legal and yet still allow for monitoring by the courts.

"FISA is triggered by foreign intelligence collection conducted 'within the United States' or against 'U.S. persons,' " Taipale writes. But by design, the Internet recognizes no boundaries, and it treats anonymity and deregulation as attributes. Ordinarily, the NSA must answer two questions before it can eavesdrop: Where is the target, and what is his nationality? "The borderless nature of terrorist threats and global communications has made place-of-collection and U.S. personhood an increasingly unworkable basis" for gathering intelligence, Taipale argues. Still, the Bush administration insists that warrantless surveillance occurs only when at least one party to a communication is outside the United States.

Harder to Wiretap



EBAY VIA GETTY IMAGES/SERGIO DIONISIO

Meg Whitman, eBay's president, stands with Niklas Zennstrom, CEO of Skype, a company that transmits telephone calls over the Internet and that eBay purchased in 2005.

But how can the NSA be sure? Graham explained that a terrorist sending an e-mail from Iraq could mask his location by sending the message through a sort of gateway, known as a user agent, which hands off the message to an Internet service provider. If the user agent is based in, say, California, then the service provider thinks the message came from California, not Iraq, Graham said.

It's unclear to what extent Internet service providers are cooperating with the NSA. But in Graham's example, it's possible to see how the agency might think that a message that originated overseas was purely domestic, and hence ignore it. Officials have said that the NSA doesn't intercept communications without warrants when both parties are inside the United States. So, theoretically, a terrorist in Iraq could communicate undetected with a terrorist in California.

In his essay, Taipale writes that FISA "contemplates only a single threshold for authorizing interception"—reason to believe that someone is an agent of a foreign power.

According to officials familiar with the NSA program, the agency broadly monitors information such as an e-mail's route—at least as much as can be traced—and the time of day it was sent to establish patterns of suspicious activity. These patterns can be detected through automated programs and without humans seeing the data. With "some approved procedure that identifies potential threats and allows for some limited follow-up," the NSA program could be regulated, Taipale maintains.

Precedent exists for such limited but legal searches and surveillance. Under a "Terry stop," police officers can briefly detain someone for questioning and conduct a limited pat-down search if they have "reasonable suspicion" to believe the person may be involved in a crime. It is one step short of an arrest, yet it gives police some ability to seek more evidence. (The name stems from the 1968 Supreme Court case, *Terry v. Ohio*, that affirmed its legality.)

"What is needed … is the electronic surveillance equivalent of a Terry stop … an authorized period for follow-up monitoring or investigation of initial suspicion derived from automated monitoring," Taipale says. "If the suspicion is not justified on follow-up, monitoring is discontinued. However, if suspicion is reasonable, then monitoring continues."

Taipale's essay, to be published in June, has attracted some early response. One FBI official called the analysis of FISA's deficiencies "brilliant," and a former government official experienced in intelligence-gathering called Taipale's recommendations "right on the mark." Taipale didn't suggest any new legal standards for conducting limited surveillance. But in the *Terry* case, then-Chief Justice Earl Warren set down the rules of the process: "In justifying the particular intrusion, the police officer must be able to point to specific and articulable facts which, taken together with rational inferences from those facts, reasonably warrant that intrusion."

But can lawmakers learn enough about the NSA's domestic eavesdropping activities to make a well-informed decision on how to better control them? NSA's critics say it's premature to change FISA when members don't understand the program and the Bush administration won't reveal its operational details. Taipale's suggestion, while "insightful," is "based on pure conjecture," said Bruce Fein, an associate deputy attorney general in the Reagan administration. In theory, electronic surveillance might benefit from Terry stops, Fein said, but "what's worrisome … is, we don't have a ghost of an idea what are the criteria to trigger the NSA to target you or me." ■

sharris@nationaljournal.com

Notes:

Notes: