



Georgia Tech Research Institute  
85 5th St. NW, Suite 217  
Atlanta, GA 30332-0760  
Aaron.Massey@gtri.gatech.edu

October 17, 2014

National Privacy Research Strategy  
NCO, Suite II-405  
4201 Wilson Blvd.  
Arlington, VA 22230

To Whom It May Concern,

I am writing on behalf of the Georgia Tech faculty and GTRI researchers who have contributed to this response to your request for information regarding the National Privacy Research Strategy. Should you have any questions concerning this response, you can contact me by email ([Aaron.Massey@gtri.gatech.edu](mailto:Aaron.Massey@gtri.gatech.edu)) or at the above address. Georgia Tech and GTRI look forward to receiving any additional information resulting from this RFI.

Sincerely,

Dr. Aaron Massey

## **Response to the National Privacy Research Strategy RFI**

This document is a response to the request for information (RFI) made by the National Coordination Office (NCO) for Networking and Information Technology Research and Development (NITRD) regarding the National Privacy Research Strategy.

### **About the Georgia Institute of Technology**

Located in Atlanta, Georgia Tech is one of the top public universities in the United States. We are a leading technology- and science-focused research institute known for our commitment to improving the human condition. Georgia Tech is organized into six colleges and contains about 31 departments, emphasizing science and technology. We have more than 100 interdisciplinary research units on and off campus. Georgia Tech ranks among the top 10 in research expenditures among universities without a medical school and is a member of the Association of American Universities, an organization of leading research universities dedicated to maintaining a strong system of academic research and education.

### **About the Georgia Tech Research Institute (GTRI)**

GTRI is the applied research arm of the Georgia Tech, one of the top engineering schools in the nation. GTRI was founded in 1934, as the Engineering Experiment Station, and employs approximately 2,000 expert scientists, engineers and support staff with \$363M in revenues during the last fiscal year (2014). GTRI is headquartered on the Georgia Tech campus in midtown Atlanta, GA with 8 research laboratories (corporate divisions), and 13 field offices located around the nation,. GTRI has a long history of solving complex problems in the areas of electronic warfare, modeling and simulation, materials, radar, sensors, optics, digital media, robotics and unmanned systems, cybersecurity and aerospace technologies. Additional information about GTRI can be found in the appendix to this document.

### **About the Georgia Tech Information Security Center (GTISC)**

The Georgia Tech Information Security Center (GTISC) focuses on research, education, and outreach programs for securing information technology-based systems. The growing scale and sophistication of threats against such systems and our increasing reliance on them creates new challenges that require better understanding of emerging threats and novel ways to counter them. The vision of GTISC is to achieve effective information security in the context of real-world problems. GTISC is an Institute-wide center with affiliated faculty come from several units, including the College of Computing, the School of Electrical and

Computer Engineering, the Sam Nunn School of International Affairs, the Georgia Tech Research Institute (GTRI), and the Office of Information Technology (OIT).

GTISC has developed new research initiatives that span multiple units, and center faculty have worked closely with our industry partners. The research output of GTISC in terms of high quality publications in top conferences and new grants and contracts awarded provide concrete evidence of our leadership position in this important field. At the same time, its education programs continue to grow. The annual GTISC Security Summit and the emerging threats report have been very well received and the summit has become the most visible and effective way for outreach and broader engagement with the information security community.

### Introductory Comments

The NITRD's Request for Information poses several important and challenging questions. The National Privacy Research Strategy (the Strategy) must reflect the nature and relationship of important concepts for information systems, including privacy, security, and trust. It is our view that these concepts are separate yet interrelated. None of these concepts should be viewed as a subset of another. Since the Strategy calls for privacy research in particular, we will highlight potential research adopting that perspective. However, we will also highlight the areas where research in security and trust may ultimately prove beneficial for privacy research.

Privacy is the subjective condition people experience when they have the ability to control information about themselves.<sup>1</sup> It allows people to maintain their individuality and autonomy. Security and privacy must be balanced.<sup>2</sup> Without information security, private information could not be kept reliably secret and control would be impossible. However, without surveillance and other security technologies, detecting, preventing, and investigating criminal activities or espionage would also be impossible. Trust is critical for information sharing and deserving of research in its own right. Patient trust has been a cornerstone of healthcare since the Hippocratic Oath included a confidentiality clause. Without research in establishing, maintaining, and verifying trust, far less information would be shared.

---

<sup>1</sup> J. Harper, "Understanding Privacy -- and the Real Threats to It," *Cato Policy Analysis*, no. 520, p. 20, Aug. 2004.

<sup>2</sup> E. H. Spafford and A. I. Antón, "Controversies in Science and Technology," vol. 2, no. The Balance of Privacy and Security, K. A. C.-H. C. M. Daniel Lee Kleinman and J. Handelsman, Eds. NYC, NY: MaryAnn Liebert, Inc, 2008, pp. 152-168.

## Responses to Specific Requests

### 1. Privacy Objectives: Scenarios and Domains of Interest

**Ensuring software requirements reflect legal obligations:** Laws and regulations are often used to address privacy concerns, but ensuring system requirements reflect legal and regulatory obligations remains challenging. Laws and regulations are not easily or directly specified as software requirements. Simply identifying relevant laws and regulations can be challenging for software developers. Some legal obligations may not fit with existing formal methods for software specification and verification. New approaches for eliciting, specifying, modeling, and evaluating compliance requirements must be developed.

**Meeting ethical obligations:** Software professionals have ethical obligations that go beyond simply meeting minimum legal obligations, but identifying privacy preferences that are not reflected in legal obligations remains challenging, particularly for innovative information systems. Organizations seeking to use new data science techniques to examine information collected from users must be able to understand when the analysis they produce may violate the expectations of their users. Research is needed to improve communication of privacy practices to end users while providing flexibility to software organizations.

**Auditing existing software systems for compliance:** Regulators and developers need methods for assessing compliance with new or updated privacy laws and regulations for existing software systems. In the U.S., laws and regulations are typically reactive, which means that many software systems will be developed prior to relevant regulations. Therefore, it is critical for software developers to be able to accurately audit their existing systems for compliance. Similarly, regulators tasked with enforcing laws and regulations need methods to verify software systems are operating within their legal regulations.

**Compliance with multiple regulations and in multiple jurisdictions:** The United States currently uses a sectoral approach to regulating privacy. Healthcare and finance are two of the most important domains with perhaps the most extensive regulations. The European Union, on the other hand, has taken a broader approach to regulating privacy. Unfortunately, information systems that blur the lines between domains and jurisdictions complicate compliance for this sectoral approach to regulating privacy. Should health data collected by a Fitbit or an Apple Watch be regulated under healthcare regulations? How can international information systems demonstrate compliance with multiple jurisdictions? The Strategy should include research programs to address these questions.

**Evaluating third-party services for compliance:** Developers seeking to use third party software services must be able to properly evaluate how those services protect user information. Methods for aligning and monitoring privacy practices

across software systems must be developed. New practices for evaluating tradeoffs in privacy and security requirements are also essential to this effort.

**Government Surveillance:** Government surveillance has long been identified as necessary for national security and a potentially serious concern in a democracy. Individuals need non-public spaces for discussion, growth, and education. The Warren and Brandeis definition of privacy as the “right to be left alone” remains popular.<sup>3</sup> Government surveillance may be necessary to protect these private spaces, but if taken too far, it may also cause a chilling effect on individuals’ ability to retreat to those spaces. Recent advances in information technologies enable both more effective, cheaper surveillance and more effective, cheaper means of avoiding surveillance. Research is needed to better evaluate the tradeoffs of both surveillance and anti-surveillance technologies.

**Behavioral Profiling:** Previously ephemeral commercial activities, such as shopping, are now regularly tracked to create behavioral profiles that can be used to improve software products or services. For example, Amazon’s popular recommendation service tracks user behavior, including items browsed or purchased, to provide the user with recommendations on similar or complementary products. Other examples of behavioral profiling have proven less popular. When NebuAd, an Internet service provider, rolled out a behavioral advertising program for their users, consumers were outraged, ultimately resulting in congressional hearings and new guidelines for online behavioral advertising from the Federal Trade Commission.<sup>4</sup> Research is needed to better understand and evaluate behavioral profiling technologies.

**Bring Your Own Device:** Employees increasingly demand the ability to use a single device for both work and personal communications. Few employees enjoy carrying both a work phone and a personal phone. Many employers are responding to this by allowing users to use their personal devices for work purposes, and these policies are known as “Bring Your Own Device (BYOD)” policies. Allowing employees to maintain and use their personal devices for work purposes carries both risks and benefits for employee privacy. Research is needed to better understand these risks and benefits as well as the tradeoff between them.

**Internet of Things:** Computing devices are becoming more capable and more prevalent. The Internet of Things (IoT) began with small chips that only provided identification, like an RFID chip in a passport. It is now growing into a vast network of sensors capable of measuring and interacting with their environment, like a network of traffic cameras capable of improving traffic flow in a city. While these devices are capable of providing incredible improvements to society, they are also

---

<sup>3</sup> S. D. Warren and L. D. Brandeis, “The Right to Privacy,” *Harvard Law Review*, vol. 4, no. 5, pp. 193–220, 1890.

<sup>4</sup> NebuAd is mentioned in the FTC report as well:

[http://www.ftc.gov/sites/default/files/documents/one-stops/policy/p085400behavadreport\\_0.pdf](http://www.ftc.gov/sites/default/files/documents/one-stops/policy/p085400behavadreport_0.pdf)

capable of pervasive surveillance. The Strategy should include research on approaches for building privacy and security into IoT devices, networks, and communications protocols. It should also include research on heuristics and methods for monitoring and assessing compliance with regulations.

**Expressing Privacy Preferences:** Online services rarely offer nuanced negotiation of privacy preferences. Users must either opt-in and accept the terms detailed in the service's privacy policy or opt-out and abstain from using the service. This binary approach leaves many consumers seeking an approach that would allow them to express more nuance privacy preferences, allowing them to use only the parts of the service that conforms to their preferences.

**Scaling Privacy Preferences:** Currently, users must specify their privacy preferences on every device or service they use. The increasing variety of devices and services users commonly use makes ensuring that a user's preferences have been accurately specified on every device and service quite challenging. Research is needed to better coordinate or propagate privacy preferences between devices and services.

**Auditing Purpose-based Disclosures:** Users regularly share sensitive information for a specific purpose based on the understanding that their information will only be used for that purpose. Consider a simple example: sharing an email address to receive a receipt for an online purchase. Many users would consider the retailer's sharing of that email address to an advertising agency to be a privacy violation. However, once the information is shared, users typically have no effective means for auditing whether that information has been disclosed to other in violation of the purpose for which it was shared.

## 2. Assessments: Concepts and Methods for Evaluating Privacy

Privacy is challenging in part because of the many valid conceptualizations and methods for evaluating privacy.

**Secrecy:** If privacy is defined as the ability of people to control information about themselves, then perhaps secret-keeping is the ultimate control. Verifiable, reliable, and usable encryption is perhaps the most important research area when privacy is viewed as secrecy.

**Risk-oriented Privacy:** Reactions to data breaches, identity theft and other privacy violations vary widely. Security-based conceptualizations seek to positively verify that information is secure, but an inability to verify that information is secure does not guarantee a privacy violation will occur. A risk-oriented approach to the possible threats resulting from a data breach or identity theft could minimize the impact of these events. One area where this is particularly critical is anonymization and re-identification of individuals' personal information. Recent research has demonstrated that basic methods for anonymization, such as removing direct

identifiers, could theoretically be re-identified by determined attackers using rich databases of linking information. However, these attacks are not known to have proven effective in practice.

**Contextual Privacy:**<sup>5</sup> Individuals may wish for their private information to be disclosed under certain circumstances or used for a particular purpose. When information is disclosed under other circumstances or used for other purposes, individuals may consider this to be a privacy violation. These circumstances are increasingly common as a result of the rapid increase in connectivity. Mobile technologies, ubiquitous computing, cloud computing, and the Internet of Things all reflect increased use of computing in specific contexts. Research is needed to understand the implications for privacy resulting from these developments.

**Privacy in Public:** Recently, the U.S. Supreme Court found that physically installing a GPS device without a warrant on an individual's car was a privacy violation, even if that car was publicly accessible. In a similar case, the U.S. Supreme Court found that using thermal imaging from a public place to capture heat signatures in a private residence without a warrant was also a privacy violation. These cases represent an important shift in the nature of privacy: it can exist in public. The rapid adoption of technologies like surveillance cameras or the Internet of Things make public spaces more easily recorded and searched, and it may also mean we need new methods for assessing privacy. The Strategy should include an investigation of possible methods and the implications of their outcomes. For example, if collection limitations give way to use limitations, then how are uses deemed valid or invalid? New policies for sensitive relationships are another example. If lawyers or spouses have special confidential relationships, perhaps cloud computing providers or telecommunications providers should as well. The Strategy should also examine the legal implications for these new methods for assessing privacy in public, particularly focusing on Third Party Doctrine.

**Privacy Economics:** Personal information collected and organized by private actors may have economic value. It may, therefore, be evaluated using economic and intellectual property analyses.<sup>6</sup> The collection of personal information can also lead to pricing, hiring, and other forms of discrimination in providing goods and services.<sup>7</sup> Additional research into economic perspectives of privacy may also address the disparity between privacy values (i.e. what people say they value) and privacy actions (i.e. how people act regarding their privacy).

---

<sup>5</sup> Nissenbaum, H. *Privacy in Context: Technology, Policy and the Integrity of Social Life*. (Palo Alto, California: Stanford University Press, 2010).

<sup>6</sup> P. Samuelson. Privacy as intellectual property? *Stanford Law Review*, 52(1125), 2000.

<sup>7</sup> R. Calo. Digital Market Manipulation, *George Washington Law Review* 82(995), 2014. Price discrimination is also highlighted by the Obama administration's report on privacy and big data: [http://www.whitehouse.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_5.1.14\\_final\\_print.pdf](http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf)

**Fair Information Practices (FIPs):** The Federal Trade Commission defined and maintained a set of fair information practices that have become an important conceptualization of privacy in the United States. FTC enforcement actions and guidelines based on these practices are extremely important in the marketplace. As a result, researchers cannot ignore this conceptualization of privacy.

**Usable Privacy and Security:** Privacy and security technologies are inherently complex, but that does not mean that they must be complicated to use. Unfortunately, too many of these systems are overly complicated for users, but evaluating and assessing privacy-sensitive systems using methods developed in Human-Computer Interaction (HCI) may allow researchers to better understand and eliminate these complications.

Applying each of these conceptualizations may lead to a rich understanding of privacy for a given domain. Consider Healthcare. By applying these perspectives, researchers could come to a nuanced understanding of why patient information must be kept securely, used only for the purposes for which it was provided, protected from possible use as a factor in discrimination, and maintained according to the FIPs. Furthermore, all of these features must be supported by software systems that are usable.

### 3. Multi-disciplinary Approach: Essential Domain Knowledge

Privacy, like security and trust, is a fundamentally a social concern. Technologies must be built with an understanding of social concerns like privacy, security, and trust. Serious privacy violations will occur if software engineers do not explicitly build-in measures to address these concerns. Unfortunately, software developers are predominately trained in the technical, rather than the social, aspects of building information systems. Software engineers cannot address concerns they do not recognize or understand. To understand how people experience privacy, computer science and software engineering researchers need to work closely with law, public policy, and social science disciplines.

The success or failure of technical solutions to privacy problems will depend heavily on collaborations with non-technical disciplines. At GTRI, we are currently developing the Trustmark Framework and Marketplace. A trustmark is a statement of conformance to a well-scoped set of trust or interoperability requirements. Trustmarks are currently being defined to capture widely adopted sets of requirements for security, privacy, identity assurance, technical interoperability, and business-level identity requirements, but their ability to accurately represent those requirements depends on expertise in many disciplines. A major goal of the pilot is to demonstrate how an organization can be evaluated by a trusted third party and, if found to be in compliance, issued a trustmark for the set of requirements evaluated. This process requires expertise in the specific domain in which the organization operates.

To ensure we are addressing privacy with a multi-disciplinary approach, GTRI analyzed ten prominent national and international privacy frameworks, studied the overlap and differences between these frameworks, and extracted a set of granular, atomic privacy principles. These principles can be used as the basis for defining privacy trustmarks that can be reused in a wide range of international contexts and composed to meet the requirements of any given privacy framework. We are currently building a Web browser plug-in to allow individuals to know whether the web sites they are visiting meet privacy criteria expressed in terms of privacy trustmarks. The results of this work have not yet been published, but it serves as an example of the sort of multi-disciplinary approach we believe to be essential to success for technical approaches to privacy.

Approaches to de-identification and re-identification serve as another excellent example of the need for a multi-disciplinary approach to privacy. De-identification of a dataset with personally identifiable information relies on removing some identifying information, often directly identifying information like names or phone numbers. However, some of the remaining data might allow an attacker to re-identify the individuals. For example, researchers were able to re-identify portions of the anonymized Netflix dataset by correlating reviews with those found on the Internet Movie Database site.<sup>8</sup> Attributes of an anonymized database that could be used in a re-identification attack are called quasi-attributes. They may allow the attacker to link the data to a known, identifiable dataset.

Understanding what datasets are widely available and how they may be used to re-identify an anonymized database obviously requires knowledge from many domains, but research is needed to better characterize the nature of this threat. How likely is it that an attacker would have access to or know about potential linking databases? How damaging would it be if they had access to such a linking database? How practical would it be to leverage information learned from such a database? These are all open, important research questions deserving further analysis.

#### **4. Privacy Architecture: Structures for Success**

Software architectures are critical to success for privacy. A software architecture that closely reflects privacy requirements will be easier to maintain as privacy preferences are changed or evolve. It will also limit developer mistakes resulting in privacy violations or data breaches. For example, privacy-aware software architectures may better support privacy principles, like providing notice to users of privacy practices. New software architectures for data collection, minimization, and use are all potentially transformative areas for researchers to study.

Supporting data provenance with software architectures is a particularly important area for research given the use of big data and newer data science analysis methods.

---

<sup>8</sup> A. Narayanan and V. Shmatikov. Robust De-anonymization of Large Sparse Datasets (How To Break Anonymity of the Netflix Prize Dataset), IEEE Security and Privacy (Oakland), 2008.

Understanding where data was collected and how what analysis was intended for it are often required to prevent privacy violations. The Federal Trade Commission has held that data collected under one privacy policy cannot be used for later analysis under an updated, less restrictive policy. However, maintaining this provenance is extremely challenging with current software architectures.

Compliance with privacy laws and regulations must also be supported as a part of the software architecture for information systems. Research is needed to produce software architectures that are better capable of establishing, maintaining, and demonstrating compliance. Auditors and regulators tasked with enforcing privacy laws and regulations must be able to more easily verify that software systems are actually in compliance. From a technical standpoint, it is easy to build a flashy software demo or prototype that looks impressive, but cannot function at scale or in real-world scenarios. It is therefore critical that auditors and regulators are able to accurately assess compliance and identify software systems that merely appear to comply but cannot handle compliance concerns under real-world conditions.

Technology infrastructure architectures are converging, and this convergence is breaking down traditional privacy barriers. For example, military investigations for national security have traditionally been wholly separate from civilian investigations for criminal activity, but they are now being conducted using many of the same technologies and infrastructure. Similarly, international communications and domestic communications, although previously separate, are now often routed through the same infrastructure. Based on traffic, time of day, and available resources, a domestic-to-domestic email could be sent almost entirely through international servers. These architectural changes represent challenging areas for research. In a world where hardware, software, and networks are used globally, how do we separate privacy expectations in a military context from those in a civilian context?

The economic advantages of cloud computing is one important reason technology infrastructure architectures are converging. Cloud-based services are now an integral part of the computing infrastructures. For example, many users send and receive email using cloud-based email services, like Gmail, or store photos in cloud-based data storage services, like Flickr or iCloud. Data stored in cloud services can be stolen, and user's privacy can be compromised.<sup>9</sup> Research is needed to develop new, user-friendly data privacy protection technologies for the cloud. For example, at Georgia Tech, we have developed a new approach to automatically encrypt user data before it leaves a user's device so that data stored in the cloud is always in encrypted form. Our approach also automatically decrypts user data when it is downloaded to the user's device. We address usability by creating a proxy between the user and the cloud-based app to automatically intercept user input and perform

---

<sup>9</sup> H. Tsukayama. [Apple's iCloud woes come just as it doubles down on the cloud](#), Washington Post, September 3, 2014.

the appropriate encryption or decryption. That is, the proxy displays the clear text to the user but only displays encrypted text to the application.

Cloud-based architectures also affect user behaviors and privacy decision-making. Many cloud-based services are supported by targeted advertisements based on user behavior or user profiles. While many users benefit from these services and targeted ads, the cost to their privacy is not clear. Major services, like Google, Facebook, and Apple, have the ability to collect extremely detailed user behavior profiles, but users are often unaware of how the data is used, how much data is collected, how much of the collected data is used for the service as opposed to the advertisements, or how they can specify their privacy preferences. In short, users are not aware of the benefits and risks of using cloud services that are supported through behavioral advertising. Asking users to make an all or nothing choice is impractical. For many users, cloud-based services are essential for work and personal business. Research is needed to help users understand the risks and benefits of these services. For example, we are performing measurement studies here at Georgia Tech to understand what behavioral data is necessary for targeted ads.

Finally, organizational architectures are at least as important for privacy as software architectures. Privacy, like security and trust, cannot be limited to a single department or group; it must be a core value for each element of the organization. Organizations dedicated to privacy will be structured to reflect that dedication. Research is needed to examine and improve organizational structures in business, non-profit, and government sectors to determine what impact those structures have on privacy values.