
# Testimony before the House Energy and Commerce Subcommittee on Commerce, Manufacturing, and Trade

**Hearing on *"The Threat of Data Theft to American Consumers"***
***May 4, 2011***

**Statement of Eugene H. Spafford**

Professor and Executive Director
Purdue University Center For Education and Research
in Information Assurance and Security (CERIAS)

Chair of the U.S. Public Policy Council
of the Association For Computing Machinery (USACM)


**Summary of Recommendations**

- A Federal mandatory notification law that includes a requirement for informing consumers about redress should be considered..
- Any regulation or statute should incorporate at least the 24 privacy recommendations listed in Appendix A (the USACM Privacy Principles).
- Any regulation or statute should apply equally to government as well as the private sector to maximize the benefit of development of software, training, and requirements, as well as protection of data.
- Our nation needs to invest in cyber forensic technologies to combat cyber crime, to support law enforcement investigation of data breaches, and to bring criminals to trial.
- Entities holding PII data should be required to meet minimum standards of good security, including staying current with software patches. No particular technology use (e.g., encryption) should be held out as a "safe harbor"; some form of appropriate third-party standards and audit should be used.
- There should be considerably more support for both fundamental and applied research in privacy and security technologies by both government and the private sector.
- As a nation, we must strengthen the cybersecurity workforce––federal programs should devote resources to improve computer science and computing education programs in K-12 as well as in higher education.


**Introduction**

By way of self-introduction, I am a professor at Purdue University. I also have courtesy appointments in the departments of Electrical and Computer Engineering, Philosophy, and Communication. At Purdue, I am also the Executive Director of the Center for Education and Research in Information Assurance and Security (CERIAS). CERIAS is a campus-wide multidisciplinary institute, with a mission to explore important issues related to protecting computing and information resources. We conduct advanced research in several major thrust areas, we educate students at every level, and we have an active community outreach program. CERIAS is the largest such center in the United States, and we have been ranked as the #1 such program in the country. CERIAS also has close working relationships with many of other universities, major commercial firms and government agencies.

Along with my role as an academic faculty member, I have served as an advisor to several Federal agencies, including the FBI, the Air Force, the GAO, and the NSA. I have been working in information security for almost 30 years.

I am also the chair of USACM, the U.S. public policy council of the ACM. With over 100,000 members, ACM is the world's largest educational and scientific computing society, uniting educators, researchers and professionals to inspire dialogue, share resources and address the field's challenges. USACM acts as the focal point for ACM's interaction with the U.S. Congress and government organizations. It seeks to educate and assist policy-makers on legislative and regulatory matters of concern to the computing community. USACM tracks U.S. public policy initiatives that may affect the membership of ACM and the public at large,

and provides expert advice to policy-makers. This advice is in the form of nonpartisan scientific data, educational materials, and technical analyses that enable policy-makers to reach better decisions. Members of USACM come from a wide-variety of backgrounds, including industry, academia, government, and end users.

My testimony is as an expert in the field. My testimony does not reflect any official position of Purdue University. My recommendations have been endorsed by USACM.

**General Problem**

Citizen concerns about disclosures of personally identifiable information (PII) held in computer databases is not surprising given the significant — and growing — number of reported breaches each year. Organizations are increasingly collecting data about various groups of people and storing that data in computing systems for their use in various business processes — or simply to warehouse for possible future use. However, those systems are often not adequately protected, and portions of the data are exposed by accident or stolen with criminal intent.

Data may be disclosed in a number of ways. Some disclosures are accidental, as a result of carelessness or flaws in the operation of underlying software (or rarely, hardware). Usually, the disclosures are a result of malicious behavior coupled with inadequate protections and policies. Malicious disclosure may come about from authorized employees (insiders) or customers who are taking or disclosing information, usually for financial gain. These disclosures may occur over a long time. These disclosures are often to confederates who commit the crimes using the information, thus making it more difficult to identify the

source of the disclosure. The resulting problems may be further complicated by delayed response, and inadequate law enforcement follow-up.

A second form of disclosure occurs when an attacker discovers some flaw or misconfiguration in the system, and uses this to gain access to the desired information. One common current method is via *spear phishing*, which occurs when a targeted piece of attack software is sent in email to a victim inside the target company, masquerading as some harmless document or application from a friend or coworker. When the attack code is run, it acts similar to a virus, installing itself on the local machine, and provides remote access for the criminal to access the system.[1] Similar types of attack code also exist that run from web pages that may be visited by employees of the company.

Attacks can also occur by exploitation of flaws in installed software. For instance, the software that drives a web commerce transaction using the SQL database language may improperly check user input given in response to a question about shipping address. A malicious user may be able to take advantage of this by inserting a semicolon followed by SQL instructions to send the entire customer database over the network to a remote site.

Theft of information is not limited to online copying of data — data exists in physical form as well as online. Thus, the fixed, physical copy can be lost or stolen as well as the online version. There are many documented cases of theft or loss of backup media (disks, tapes,

[1] There have been some very high-profile cases of spear phishing in the news recently. Oak Ridge National Labs had to shut down their Internet connection in April when over 500 employees were attacked like this, RSA had some of their security software compromised this spring via spear phishing, and the highly publicized breakins of Google and over 30 other large companies were accomplished with spear phishing from China.

thumb drives, CD-ROMs), theft or loss of laptop computers, and even theft of whole server machines and disks. The theft or loss of paper records may also lead to some of the same forms of disclosure mentioned here — high speed scanners can quickly convert paper documents into database files again; my university has been forced to limit what is printed in our campus phone directory, for instance, because some commercial firms were obtaining copies, digitizing them, and using the results for marketing.

***Growth of the Problem***

One of the more notable incidents occurred in 2005, when the data broker ChoicePoint revealed that fraudulent access to over 140,000 customer records had occurred over the previous two year period, leading to multiple instances of identity theft and fraud.[2] That incident led to investigations by the FTC and SEC, as well as multiple lawsuits.

Despite the publicity of the ChoicePoint case, and the potential for lessons-learned, the instances of disclosure and loss of PII data have only increased in the years since, with hundreds of cases per year in the United States reported — and undoubtedly many more unreported. This year, before this hearing, two very large and troubling exposures of such data were reported by Sony and Epsilon, with potentially over 100 million consumers affected by the combination of incidents.

These two cases are particularly illustrative of the complexities of such incidents. The individuals affected by the Epsilon case had no idea they had records stored with Epsilon, and

[2] See "The ChoicePoint Dilemma", by Paul N. Otto, Annie I. Antón, and David L. Baumer, *IEEE Security & Privacy*, Sep/Oct 2007, pp. 15-23.

likely still have no idea what the extent of their relationship is with that company.[3] In the Sony case, the majority of the victims are likely young people whose sense of risk, privacy and consequence are not yet fully developed, and thus they may also not understand the full ramifications of what has happened. Presumably, both companies are large enough that they could have afforded to spend an appropriate amount on security and privacy protections of their data; I have no information about what protections they had in place, although some news reports indicate that Sony was running software that was badly out of date, and had been warned about that risk.

To put those incidents in a different perspective, the Privacy Rights Clearinghouse keeps a database[4] of *exposed*[5] breaches from 2005 that includes both accidental disclosures and fraudulent accesses. As of the 1st of May 2011, they documented almost 600 million records have been disclosed in 2,459 separate incidents in the United States. That is an average of approximately 100 million records per year. The Sony breaches disclosed in April and May of 2011 alone equal approximately 100 million records. Other firms listed in their database for those months included Blockbuster, several hospitals, the IEEE (Institute of Electrical and Electronics Engineers) and , a restaurant in southwest Indiana, Albright College

---

[3] This is similar to the ChoicePoint breach in that the individuals affected in that incident also did not realize the relationship they had with the company.

[4] Available at http://www.privacyrights.org/data-breach#CP

[5] I emphasize *exposed* because there are undoubtedly many more that are undisclosed, and many that are also simply not discovered. There may be more that are undiscovered than disclosed and undisclosed combined.

in Reading, PA, the Hartford Insurance Company, many doctors offices, US Airways, and Apple iTunes.

Sometimes, a company is involved even though their computers are not the ones breached. Among the more than 50 companies whose customer lists were stolen in the Epsilon data breach were Chase Bank, Hilton, Best Buy, and Target. Customers of those companies should expect to receive emails suggesting that as loyal customers, they can click to receive a valuable coupon. Ironically, some possible fraud may even be in the form of warnings about fraud —customers will receive messages telling them that their email address was stolen and to protect themselves they should click on a link to enter their credit card information, or apologizing for the inconvenience and offering a discount by clicking on a link and signing in, thus disclosing their password to criminals.

It is important to note that data breaches occur in all forms of organizations: retail establishments, financial services, nonprofit entities, health care providers, public utilities, and even computer security firms themselves. Federal and state government agencies are also affected, and are sometimes responsible for disclosure of particularly sensitive material because of their privileged access status under law. A review of the aforementioned list for the last few months reveals disclosures by the IRS, a U.S. District Court, the Social Security Administration, Veterans Affairs, the Oklahoma Department of Health, the Texas Comptroller's Office, the Maine State Prison, and the town of Barton, Vermont (to name a few). Clearly, the problem of properly safeguarding personal information is not limited to the private sector.


Disclosure and theft of PII records has not abated since the ChoicePoint incident in 2005 first prompted Congressional scrutiny. More data is being collected and stored, often for less well-defined purposes. More firms have access to large-scale storage and computing, and thus are now able to store and aggregate data online. Additionally, there are more entities interested in committing fraud online, and their sophistication and reach has grown considerably faster than has that of law enforcement and security personnel in the same time. Their ability to distribute what they take has also increased with the speed and reach of networks.

Nonetheless, the increase in sophistication of attackers, and the growth in data do not totally explain all the incidents. My personal conclusion from reviews of reports in the press and discussions at professional meetings is that operators of these systems — both in government and the private sector — continue to run outmoded, flawed software, fail to follow some basic good practices of security and privacy, and often have insufficient training or support. The most commonly cited reason for these failings is cost. The cost of providing better security and privacy protection is viewed as overhead that is not recovered in increased revenue, and it is usually one of the first things trimmed in budget cuts. Running outdated software and unpatched operating systems exposes citizens to risks and consequences whose cost a company does not bear. Therefore a company does not have an immediate economic incentive to make the investment needed to prevent breaches. There is a risk of real loss if a

breach occurs, however: the cost to a company per record averages $214, and has increased every year since 2005.[6]

As a cautionary note for the future: many companies are eager to move their operations "into the cloud." This will mean that the PII databases may be stored on servers located outside the United States. If those servers are compromised or the media is stolen, it is unclear what legal rights and protections the victims may have.

***Types of Abuse***

It may not be immediately obvious why disclosure of some of this information might be of concern. In some cases, the disclosure might only be of an account name and some password hint, or directory information that might be otherwise easily found in a public directory. However, such information in context or in combination with other information can be quite damaging. The presence of a record in a database is informative — that someone is a customer, patient, or subscriber, for instance. Combining information from several different sources may allow someone to infer much more than from any single source alone (and given the availability of information on social media sites and from other breaches, this is not difficult to do).

It is then how these bits of information are used that are of concern. Certainly, any disclosure poses a privacy concern to some users, but there are additional concerns related more specifically to criminal activities.

---

[6] According to an annual study by the Ponemon Institute: http://www.ponemon.org/blog/post/cost-of-a-data-breach-climbs-higher

***Identity theft.*** If sufficient information is obtained about someone, it is often possible to perform identity theft, thus gaining false identification for employment, obtaining credit, or evading law enforcement.

***Harassment and stalking.*** Information about individuals may be used to harass public officials or celebrities, or stalk victims. Obtaining address information may be used to stalk spouses who have fled abuse, for instance.

***Spear phishing.*** Phishing, the attempt to get someone to click through to a false web site through email or divulge their account information, can be made more effective if the email is tailored somewhat to the victim. This is known as spear phishing. Details from large data bases, such as account names, length of service, addresses, and account options can be used to tailor a phishing message to make it appear legitimate and thus trick someone into divulging their account information.

***Tracking for physical crime.*** It is possible to use data from a database to identify victims for physical crime, although I am unaware of any cases of this yet occurring. This would be instances where the database would indicate something about income level or perhaps that indicated people were away on vacation, and this would be useful to criminals seeking to commit burglaries in an area.

***Extortion.*** The presence of information in a database could be used for extortion. This has occurred in cases of medical information, particularly regarding HIV status. There are many other items of information that might be used, including past criminal violations, past marriages, or even items as simple as what videos and on-line books someone likes to

download. In an extreme case, some individuals open to extortion might be in sensitive positions, and this could then lead to espionage.

***Inference.*** People tend to use the same passwords, and use the same hints for passwords when visiting multiple sites. The trend at sites to use prompts for password recovery such as "Name your first pet" elicit the same (honest) response from most people or they would otherwise not be able to remember all the answers. Thus, gaining the passwords or hint answers for users from one site might be combined with the same user name at other, more valuable sites such as a bank, to provide access for direct fraud.[7]

***Direct fraud.*** Clearly, information containing credit card numbers, ACH numbers, or other financial information may be used directly — and usually is.

**USACM Recommendations**

1. **A Federal mandatory notification law that includes a requirement for informing consumers about redress should be considered.** Mandatory notification of consumers after a breach (possibly) involving their PII, along with information about steps to take to safeguard their identity appears to have some positive value. A study[8] by Romanosky, et al. suggests that state mandatory notification laws provide a small decrease (about 6 percent) in identity theft. Not all states have a mandatory notification law.

---

[7] See, for example, http://www.pcworld.com/article/188763/too_many_people_reuse_logins_study_finds.html or http://www.lightbluetouchpaper.org/2011/02/09/measuring-password-re-use-empirically/

[8] Romanosky, Sasha, Telang, Rahul and Acquisti, Alessandro, Do Data Breach Disclosure Laws Reduce Identity Theft? (Updated) (September 16, 2008). Forthcoming in the Journal of Policy Analysis and Management, 2011. Available at SSRN: http://ssrn.com/abstract=1268926


2. **Any regulation or statute should incorporate at least the 24 privacy recommendations listed in Appendix A.** USACM has developed a set of 24 basic privacy recommendations for use with databases. Those are enclosed as Appendix A to this testimony. We strongly recommend that they be followed for all data sets containing PII, whether government or private, commercial or nonprofit. All of them are important to limit exposure and damage.

3. **Any regulation or statute should apply equally to government as well as the private sector to maximize the benefit of development of software, training, and requirements, as well as protection of data.** We encourage the committee to ensure that any legislation or regulation apply equally to all government data collections as well as private sector data. The dangers and risks apply no matter who collects and holds collections of PII.

4. **Our nation needs to invest in cyber forensic technologies to combat cyber crime, to support law enforcement investigation of data breaches, and to bring criminals to trial.** Law enforcement also appears to be insufficiently supported with resources for forensic investigation of computing incidents. This is another area where resources for research into better tools and technologies would be helpful. So long as the criminals do not fear apprehension, they will continue to attack our systems. There also appear to be too few agents to investigate breaches, and too few resources to ensure prosecutions.

5. **Entities holding PII data should be required to meet minimum standards of good security, including staying current with software patches. No particular technology use**

**(e.g., encryption) should be held out as a "safe harbor"; some form of appropriate third-party standards and audit should be used.**

6. **There should be considerably more support for both fundamental and applied research in privacy and security technologies by both government and the private sector.** There needs to be additional research into privacy-enhancing and privacy-preservation technologies for large data sets. This is a nascent area of research, as is much of security, and the area is under-resourced. Many of the problems being faced might be solved with better tools, software, and understanding of fundamental processes.

**7. As a nation, we must strengthen the cybersecurity workforce––federal programs should devote resources to improve computer science and computing education programs in K-12 as well as in higher education.** As companies increasingly store data in digital formats, a well-prepared cybersecurity workforce is needed. Strengthening computer science and computing education will help address security challenges in the long-run, ensuring that students have adequate knowledge of the field. The education pipeline feeding our current workforce too often focuses on training rather than education and is frequently absent in K-12 education. Expanding this workforce via education is critical and should start at K-12 and extend through our higher education system.

**Acknowledgements**

I wish to acknowledge comments and assistance provided to me in preparing this testimony from David Bruggeman, Cameron Wilson, Annie Antón, Sarah Granger, Emil Volcheck, Travis Breaux, Andy Grosso, Ollie Smoot, Jim Horning, Jeremy Epstein, Aaron


Massey, Paul Otto, and other members of USACM. Despite listing their names here, none of those individuals necessarily agrees with, nor endorses any of my comments or opinions.


# Appendix A

## USACM Policy Recommendations on Privacy

### Background

Current computing technologies enable the collection, exchange, analysis, and use of personal information on a scale unprecedented in the history of civilization. These technologies, which are widely used by many types of organizations, allow for massive storage, aggregation, analysis, and dissemination of data. Advanced capabilities for surveillance and data matching/mining are being applied to everything from product marketing to national security.

Despite the intended benefits of using these technologies, there are also significant concerns about their potential for negative impact on personal privacy. Well-publicized instances of personal data exposures and misuse have demonstrated some of the challenges in the adequate protection of privacy. Personal data — including copies of video, audio, and other surveillance — needs to be collected, stored, and managed appropriately throughout every stage of its use by all involved parties. Protecting privacy, however, requires more than simply ensuring effective information security.

The U.S. Public Policy Council of the Association for Computing Machinery (USACM) advocates a proactive approach to privacy policy by both government and private sector organizations. We urge public and private policy makers to embrace the following recommendations when developing systems that make use of personal information. These

recommendations should also be central to any development of any legislation, regulations, international agreements, and internal policies that govern how personal information is stored and managed. Striking a balance between individual privacy rights and valid government and commercial needs is a complex task for technologists and policy makers, but one of vital importance. For this reason, USACM has developed the following recommendations on this important issue.

## Recommendations

### *Minimization*

1. Collect and use only the personal information that is strictly required for the purposes stated in the privacy policy.
2. Store information for only as long as it is needed for the stated purposes.
3. If the information is collected for statistical purposes, delete the personal information after the statistics have been calculated and verified.
4. Implement systematic mechanisms to evaluate, reduce, and destroy unneeded and stale personal information on a regular basis, rather than retaining it indefinitely.
5. Before deployment of new activities and technologies that might impact personal privacy, carefully evaluate them for their necessity, effectiveness, and proportionality: the least privacy-invasive alternatives should always be sought.

### *Consent*

6. Unless legally exempt, require each individual's explicit, informed consent to collect or share his or her personal information (*opt-in*); or clearly provide a readily-accessible mechanism for individuals to cause prompt cessation of the sharing of their personal

information, including when appropriate, the deletion of that information (*opt-out*). (NB: The advantages and disadvantages of these two approaches will depend on the particular application and relevant regulations.)

7. Whether opt-in or opt-out, require informed consent by the individual before using personal information for any purposes not stated in the privacy policy that was in force at the time of collection of that information.

### *Openness*

8. Whenever any personal information is collected, explicitly state the precise purpose for the collection and all the ways that the information might be used, including any plans to share it with other parties.
9. Be explicit about the default usage of information: whether it will only be used by explicit request (opt-in), or if it will be used until a request is made to discontinue that use (opt-out).
10. Explicitly state how long this information will be stored and used, consistent with the "Minimization" principle.
11. Make these privacy policy statements clear, concise, and conspicuous to those responsible for deciding whether and how to provide the data.
12. Avoid arbitrary, frequent, or undisclosed modification of these policy statements.
13. Communicate these policies to individuals whose data is being collected, unless legally exempted from doing so.

### *Access*

14. Establish and support an individual's right to inspect and make corrections to her or his stored personal information, unless legally exempted from doing so.

15. Provide mechanisms to allow individuals to determine with which parties their information has been shared, and for what purposes, unless legally exempted from doing so.
16. Provide clear, accessible details about how to contact someone appropriate to obtain additional information or to resolve problems relating to stored personal information.

***Accuracy***

17. Ensure that personal information is sufficiently accurate and up-to-date for the intended purposes.
18. Ensure that all corrections are propagated in a timely manner to all parties that have received or supplied the inaccurate data.

***Security***

19. Use appropriate physical, administrative, and technical measures to maintain all personal information securely and protect it against unauthorized and inappropriate access or modification.
20. Apply security measures to all potential storage and transmission of the data, including all electronic (portable storage, laptops, backup media), and physical (printouts, microfiche) copies.

***Accountability***

21. Promote accountability for how personal information is collected, maintained, and shared.
22. Enforce adherence to privacy policies through such methods as audit logs, internal reviews, independent audits, and sanctions for policy violations.

23. Maintain *provenance* — information regarding the sources and history of personal data — for at least as long as the data itself is stored.
24. Ensure that the parties most able to mitigate potential privacy risks and privacy violation incidents are trained, authorized, equipped, and motivated to do so.

USACM does not accept the view that individual privacy must typically be sacrificed to achieve effective implementation of systems, nor do we accept that cost reduction is always a sufficient reason to reduce privacy protections. Computing options are available today for meeting many private sector and government needs while fully embracing the recommendations described above. These include the use of de-identified data, aggregated data, limited datasets, and narrowly defined and fully audited queries and searches. New technologies are being investigated and developed that can further protect privacy. USACM can assist policy-makers in identifying experts and applicable technologies.

(June 2006)