Developing Survey Systems in a Restricted Security Environment

Katherine MASON
Research Computing Division, RTI International, 3040 East Cornwallis Road
Research Triangle Park, NC 27709-2194, USA

and

Nathan SIKES
Research Computing Division, RTI International, 3040 East Cornwallis Road
Research Triangle Park, NC 27709-2194, USA

and

Carol PLACE
Research Computing Division, RTI International, 3040 East Cornwallis Road
Research Triangle Park, NC 27709-2194, USA

ABSTRACT

As massive amounts of data are being collected without the knowledge of the vast majority of citizens, at home and abroad, the U.S. government has taken extraordinary steps in trying to protect the personal data that is electronically collected as the citizenry becomes enlightened and concerned about how their personal data is being stored and used. This paper, developed by computer professionals at RTI International who conduct surveys for the U. S. federal government as well as other levels of government and private sources, explores the methods and techniques as well as pros and cons for helping secure confidential data. Additionally, the paper discusses:

- the basic definitions of the security requirements needed to protect private data;
- handling some of the challenges related to accessibility, implementation of software solutions, and data flow in a high security environment; and
- some of the tradeoffs inherent in this environment.

Keywords: Data, security, identifiers, FIPS, confidentiality, PHI, PII, authentication

INTRODUCTION

In this age of massive amounts of data being collected without the knowledge of the vast majority of people, there has been increasing emphasis by the US government on the privacy and confidentiality of personal data. RTI International conducts surveys for the federal government in a number of different modes: web self-interviews, personal face-to-face interviews, telephone interviews, and paper self-interviews returned by mail. New regulations and restrictions mandate higher levels of protection than ever before for personal data collected in these interviews. This can only increase in the future as citizens become even more concerned about how their personal data is being stored and used. This paper discusses:

- the basic definitions of the security requirements needed to protect private data;
- how our company has handled some of the challenges related to accessibility, implementation of software solutions, and data flow in a high security environment; and
- some of the tradeoffs inherent in this environment.

What personal and private information does RTI capture and store for our many different surveys?

1. Project data such as research files with **direct** identifiers (e.g., Social Security Number (SSN), name, address, telephone number, etc.),
2. Research files with **indirect** identifiers (e.g., Date of Birth, ZIP (postal) code, or some specific demographics),
3. Tracing files used to locate respondents (often containing the SSN), and
4. Contact files purchased from outside sources (e.g., mailing lists).

With such sensitive information being in the possession of RTI coupled with the responsibility and liability that accompanies possession of these data, we are compelled ethically and legally to protect personal information in the following ways:

- Encryption of laptops and other portable media
- Use of separate identifiers from survey data
- Restricting access to files with identifiers
- Limiting copies of files with identifiers
- Using secure transmission methods when sending/receiving files
- Encryption of data files at the highest strength when transferring data
- Destroying files with identifiers at the earliest possible time
- Storing files with identifiers in a secured location if they need to be kept
- Having staff undergo appropriate and regular security training
- Informing staff of the gravity of the use and disclosure of private data (e.g., signing a confidentiality agreement or providing sworn statements of non-disclosure)

DIRECT IDENTIFIERS:

1. Names (of individuals or other entity)
2. Complete Postal Address
3. Telephone numbers
4. Fax numbers
5. Electronic mail addresses
6. Social security numbers
7. Medical record numbers
8. Health plan beneficiary numbers
9. Account numbers
10. Certificate/license numbers,
11. Vehicle identifiers and serial numbers, including license plate numbers
12. Implanted device identifiers and serial numbers
13. Web universal resource locators (URLs)
14. Internet protocol (IP) address numbers
15. Biometric identifiers, including fingerprints and voiceprint
16. Full-face photographic images and any comparable images.
17. Any other unique identifying number, characteristic, or code that could be used to identify an individual.

INDIRECT IDENTIFIERS:

1. City/town, State, ZIP code (without street address)
2. Any elements of date - except year - for dates directly related to an individual (e.g., DOB, date of death, date of admission/discharge, etc.)

SECURITY DEFINITIONS

There are a multitude of government regulations surrounding security but of particular interest to this discussion is the Federal Information Processing Standards (FIPS), which is a series of regulations developed by the US government for all government contractors and non-military agencies. FIPS 199[1], the Standards for Security Categorization of Federal Information and Information Systems, lays out the requirements for three security levels – low, moderate and high. At each level, there are three terms that are essential to understand in the proper application of security. They are confidentiality, integrity, and availability.

- Confidentiality – preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information
- Integrity – guarding against improper information modification or destruction, which includes ensuring information non-repudiation and authenticity
- Availability – ensuring timely and reliable access to and use of information

For the purposes of this discussion, we are primarily concerned with confidentiality at the FIPS moderate level. In addition to the FIPS, RTI uses the National Institute of Standards and Technology (NIST) Special Publication 800-53[2], "Recommended Security Controls for Federal Information Systems," to determine which security controls are needed to comply with minimum security requirements. As a result, RTI developed a hyper-secure network as an environment where private personal data can be captured, stored, analyzed, stripped of identifiers, and transmitted for use by project and client staff.

Two other terms that are important in the realm of personal data security are:
Personally Identifiable Information (PII) and
Personal Health Information (PHI).

- **PII** is information that can be used to uniquely identify a single individual or that can be used with other sources to uniquely identify a single individual such as full name, address, telephone number, e-mail address, Social Security number or other identifying numbers (drivers license number, credit card or medical records numbers).
- **PHI** is similar except that it also includes things such as medical records, medical history, lab results and insurance information, i.e., most any data that is health-related and specific to the individual.

CHALLENGES

There are some definite challenges to working in a hyper-secure environment. In addition to the extra costs associated with increased documents, number of processes, amount of error-checking and a sometimes steep learning curve for staff, there are a number of special considerations such as the following:

PII and Non-PII Data Flow

Data is allowed to flow into the hyper-secure environment unrestricted but as one might expect, data (PII especially) coming out of the secure environment must be carefully monitored and data flow restricted as needed to appropriate users of the data. A form of secure file transport must be used to protect the data. PII vs. non-PII data flow is often handled differently. In a standard survey project, a project team or a system processes all the data in the same way but in the secure environment, there may be different processes to treat the two types of data. Sometimes this leads to redundant network structures – one inside the hyper-secure network and one located outside in a non-hyper-secure network. PII data need to be encrypted using applicable government standards which may require additional software licenses and/or additional training. A given project also must determine if there is a need for their clients to own the same software in order to decrypt the files. Even an incorrect version employed by either the sender of the data or the receiver is enough to prevent decrypting a file.

Accessing /Working in the Environment

We have found that it is a major culture shock for employees (software developers and testers, data analysts, and project managers) when they begin dealing with the burdensome yet truly needed security requirements in the hyper-secure network environment including two-factor authentication and enhanced password restrictions. When we first began working in this environment, the challenge was to remember that one needed to have the token with you at all times or you could not log in to the environment. There are more steps for every task and schedules often need to be adjusted to allow more time because all systems need to be tested twice – once outside the environment and then again once they are copied inside the environment. Communication, additional training and setting expectations are key requirements when working with staff that have not been exposed to this level of data management. Often what is considered normal processing in an everyday network environment is not available in the hyper-secure network environment. One clear example of this can be seen with the "copy and paste" function, a very common Windows feature that many computer users perform multiple times per day. The hyper-secure environment cannot allow this feature because it would be too easy to inadvertently (or purposefully) copy PII data out of its secure environment into a file or window in a non-secure area. Emailing of documents or any file from the environment is forbidden so multiple steps passing encrypted files from one environment to another must be performed to actually send information containing PII. In addition, RTI's hyper-secure network only allows outbound e-mails which are restricted in size to 50K.

Two factor authentication is different from **one factor authentication** in that, for example, one factor authentication is usually the use of a password whereas two factor authentication uses the principle of something you know (such as a PIN) plus something you have (such as a security token with random numbers generated every few seconds that match with the network server).

Software Development

Software developers by nature choose the easiest and fastest steps to generate a complete and robust application as demanded by the system owner and program specifications. When developing applications for use inside a hyper-secure network, programmers can develop the software outside of the restricted security environment and avoid the daily challenges of working within the stringent security framework. IT restrictions for the environment often handicap software development because software development environments, add-ons, DLL development, folder permissions, and other common fundamental software tools and tasks can be considered processes and components that pose a risk at many levels and are often banned. This leads to a constrained debugging environment causing increased development and testing time for software, especially if the full software development life cycle is employed within the hyper-secure environment. Programmers are encouraged to develop their applications outside the secure network in a normal programming environment where it is easier and better to prepare the robust application, test the functionality of the application, and then when completely ready, they can port the compiled application or program code to the hyper-secure network. Once a developed application has been ported inside, tests should be run from both operational and functional standpoints ensuring that the software conforms to all security rules imposed in the environment and that it also runs as expected. Of course, it is expected and required that the software will not make use of files or other software outside the ESN. Sometimes just a tweak is needed to make the software complete and it is tempting to perform the tweak inside the hyper-secure network environment but all software modifications should be conducted outside the environment and transported into the environment. There is often still the need for adjusting

parameters of some software programs once inside the environment to ensure that they conform to the environment and perform as designed and expected.

Documentation

The requirement for documentation for projects needing this level of security is significantly higher than our standard survey project. At a minimum, this set of documents includes a system security plan, a Privacy Impact Assessment and evidence of vulnerability scans in the hyper-secure environment. In addition, there are a number of other formal requirements ultimately culminating in an authority to operate (ATO). The ATO is usually given by the federal government after exhaustive research and discussion of the documents as well as formal testing of the complete system by third parties to ensure that the system conforms to the documents as stated. The documentation not only takes extra time but it requires special knowledge – in fact, RTI has a security group with the required knowledge to specifically address this aspect of the security process.

Laptops and Other Portable Media

RTI often uses laptops, tablets, PDAs or smart phones to conduct face-to-face interviews. These devices present a special problem since they are so portable. One practice we employ to deal with this challenge is to install hard disk encryption software called Pointsec® on all company laptops and all laptops used in field surveys.

PointSec® is a hardware encryption application that protects proprietary and confidential information from loss or theft by using mathematical functions to encrypt hard drive information such as the operating system, resident data, temporary files, deleted files, and any unused space.

If a laptop is lost or stolen, the whole disk encryption provides confidence that the data and any personally identifiable information collected about the survey respondent have not been compromised. Only the individual to whom the laptop was assigned, and a few staff with a special physical token, can access the data on the disk. Another effective method to deal with the need for a higher level of security is to simply avoid the use of portable media for that project.

SUMMARY

Although necessary in this age of identity theft and other crimes, the emphasis on securing personally identifiable information (PII) has made it more challenging and difficult to implement projects consisting of work such as that commonly encountered in multi-mode surveys. There are a number of tradeoffs when providing this increased security, arguably the most important of which are the additional labor costs (up to 25% in some cases) and the requirement to build enhanced and often redundant systems. Every task and process in this environment takes longer to develop and apply and projects need to factor this into their project design and schedules. In addition, the human factor must be accounted for when deploying and working in this new environment. There is an often intimidating learning curve, and because the environment requires a different manner of working, new work habits must be formed. Project staff and clients alike must be aware of the potential pitfalls and plan accordingly.

REFERENCES

[1] Federal Information Processing Standard (FIPS) Publication 199 (http://csrc.nist.gov/publications/fips/fips199/FIPS-PUB-199-final.pdf)

[2] National Institute of Standards and Technology (NIST) Special Publication 800-53 (http://csrc.nist.gov/publications/nistpubs/800-53-Rev3/sp800-53-rev3-final.pdf)