

Protecting Your Data

Dale Plummer

Department of Biostatistics

September 11, 2013

Introduction

Certain kinds of information should be kept private. It is the responsibility of those of us who deal with this information to understand which items should be protected.

This presentation will cover what data should be protected and the policies, practices and tools to that.

Data privacy and protection is a big issue. The unintended disclosure of private information can do harm individuals, projects, researchers and institutions. Such disclosure may expose the institution and responsible persons to bad publicity, legal and civil penalties, and loss of funding.

What data should be protected?

Vanderbilt policy and law says that these categories of data must be protected:

- Protected Health Information (PHI)
- Research Health Information (RHI)
- "personal Information"

<http://privacyruleandresearch.nih.gov/> - This website provides information on the Privacy Rule for the research community.

The document "Summary of the HIPAA Privacy Rule" at http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/privacy_summary.pdf is a more manageable summary of the privacy rule.

What data should be protected?

Protected Health Information (PHI)

1. Names
2. All geographical identifiers smaller than a state, except for the initial three digits of a zip
3. Dates (other than year) directly related to an individual
4. Phone numbers
5. Fax numbers
6. Email addresses
7. Social Security numbers
8. Medical record numbers
9. Health insurance beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers, including license plate numbers;
13. Device identifiers and serial numbers;
14. Web Uniform Resource Locators (URLs)
15. Internet Protocol (IP) address numbers
16. Biometric identifiers, including finger, retinal and voice prints
17. Full face photographic images and any comparable images
18. Any other unique identifying number, characteristic, or code except the unique code assigned by the investigator to code the data

What data should be protected?

Research Health Information (RHI)

"...is a term used by Vanderbilt to identify individually identifiable health information (IIHI) used for research purposes that is not PHI, and thus is not subject to the HIPAA Privacy and Security regulations. RHI is created in connection with research activity and is not created in connection with patient care activity. If a researcher is also a healthcare provider and IIHI is created in connection with the researcher's healthcare provider activities, then the IIHI is PHI and is subject to HIPAA."

<http://www.mc.vanderbilt.edu/root/vumc.php?site=hipaa&doc=12204>).

A lot of our data comes from patients and so is PHI. For our purposes, there is not really difference between PHI and RHI. Both have to be handled basically the same.

What data should be protected?

"personal Information"

...may contain individually identifiable information about patients, employees, students, or research participants. Although not necessarily covered by HIPAA regulations, other regulations and Vanderbilt require that this information be protected as well.

Policies

[Information Privacy & Security Website](#) - The Information Privacy & Security Website for VUMC. Contains links for Privacy (data breach notification, policies, training), Information Security (file transfer application, encryption), HIPAA, and a FAQ.

[Vanderbilt Policy on De-Identification](#) - PHI is considered de-identified if all data elements that identify the individual or of relatives, employers, or household members of the individual are removed.

[Vanderbilt policy on encryption](#) VMC policy stipulates that when a legitimate business purpose exists requiring an individual to maintain identifiable Protected Health Information (PHI) or Research Health Information (RHI) on a device other than a secure network server that device must be encrypted.

State and federal legislation requires public notification when certain person-identifiable information or PHI is lost or stolen unless the device containing the data was known to be encrypted.

See <http://biostat.mc.vanderbilt.edu/ProtectingYourData> for a full set of links.

Policies

Encryption

Policy is pretty clear. If you store PHI/RHI on a mobile device (laptop, flash drive, phone, etc.) then it needs to be encrypted. We believe that the policy requires encryption on desktop computers, too.

Loss Reporting

- VUMC policy and other regulations require notification in the event of any unauthorized disclosure of individually identifiable patient or other personal information.
- Known or suspected incidents involving breach of PHI are reported to the [VMC Privacy Office](#)
- If you lose a laptop or other device, let someone know immediately. Someone on the IT team or the Administrative Officer can help make the appropriate notifications.

Practices

- If you can avoid it, don't store PHI, RHI, or other identifying information on your workstation, laptop, or other device
- Watch out when using cloud storage
- Understand and use de-identification
- Don't use email to transfer data sets
- Use secure data transfer to transfer data sets
- Be careful with email and websites
- Use good passwords

Practices

<http://xkcd.com/936/>

UNCOMMON (NON-GIBBERISH) BASE WORD ORDER UNKNOWN

Tr0ub4dor & 3

CAPS? COMMON SUBSTITUTIONS NUMERAL PUNCTUATION

(YOU CAN ADD A FEW MORE BITS TO ACCOUNT FOR THE FACT THAT THIS IS ONLY ONE OF A FEW COMMON FORMATS.)

~28 BITS OF ENTROPY

$2^{28} = 3 \text{ DAYS AT } 1000 \text{ GUESSES/SEC}$

(PLAUSIBLE ATTACK ON A WEAK REMOTE WEB SERVICE. YES, CRACKING A STOLEN HASH IS FASTER, BUT IT'S NOT WHAT THE AVERAGE USER SHOULD WORRY ABOUT.)

DIFFICULTY TO GUESS: **EASY**

WAS IT TROMBONE? NO, TROUBADOR. AND ONE OF THE 0s WAS A ZERO?

AND THERE WAS SOME SYMBOL...

DIFFICULTY TO REMEMBER: **HARD**

correct horse battery staple

FOUR RANDOM COMMON WORDS

~44 BITS OF ENTROPY

$2^{44} = 550 \text{ YEARS AT } 1000 \text{ GUESSES/SEC}$

DIFFICULTY TO GUESS: **HARD**

THAT'S A BATTERY STAPLE.

CORRECT!

DIFFICULTY TO REMEMBER: YOU'VE ALREADY MEMORIZED IT

THROUGH 20 YEARS OF EFFORT, WE'VE SUCCESSFULLY TRAINED EVERYONE TO USE PASSWORDS THAT ARE HARD FOR HUMANS TO REMEMBER, BUT EASY FOR COMPUTERS TO GUESS.

Tools

- Secure file transfer
 - [Data-Hippo](#)
 - [VUMC Secure File Transfer](#)
- De-identification
 - “How to De-identify Data” by Xulei Shirley Liu (http://www.mc.vanderbilt.edu/crc/workshop_files/2008-03-07.pdf)
 - “Guidance Regarding Methods for De-identification...” <http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveridentities/De-identification/guidance.html#protected>
- Encryption
 - [TrueCrypt](#) (Linux, Windows, Macintosh)
 - [Ubuntu Full Disk Encryption](#)
 - [Check Point Full Disk & Media \(for USB drives\) Encryption](#)
 - [Encfs](#) - <https://help.ubuntu.com/community/FolderEncryption>
 - FileVault (Macintosh)
- Password Management

Tools

<https://data.vanderbilt.edu/data-hippo/>

The screenshot shows a web browser window with the URL <https://data.vanderbilt.edu/data-hippo/>. The page content includes:

- A header message: "This is a tool that is used to transfer files to selected recipients. Using this application, one can place a file on the server and allow only those specified to retrieve that file."
- A "File to upload:" section with a "Choose File" button and the text "No file chosen".
- A "Sender Information" section with the instruction: "Enter your email address, name, and an optional note that describes the file you are depositing. Also you may specify a date after which this file will be deleted." It contains input fields for "Email Address:", "Name:", and "Expiration date:". Below these is an example date format: "Example Date formats: MM/DD/YYYY or MM/DD/YYYY HH:MM:SS" and a note "Default value of 2 weeks from now".
- A text area for "Note that describes the file; which will be included in the recipient notification email".
- A "Recipient Information" section with the instruction: "Enter the email addresses of those who may retrieve the file. Separate the addresses with commas, carriage returns or semicolons. Each email address in the list will be sent a message stating that a file is available to be picked up. Included in the email will be a URL and a password. The recipient will navigate to the provided URL and supply the password and they can then download the file to their computer." It includes an "Email Addresses" input field and a checkbox labeled "Receive download confirmation emails?".
- A "Submit" button at the bottom.

The browser's address bar shows the URL, and the top of the window displays several open tabs: HIPAA R..., www.mc..., delivery..., Guidance..., Protectin..., Make olc..., and Data Hip... The browser's bookmark bar shows various sites like comcast, Gmail, calendar, card, biostat, me, DalesNotes, dbconnect, and NYTimes.

Tools

The screenshot shows a web browser window displaying the 'Secure File Transfer' interface. The browser's address bar shows the URL: <https://accellion1r.mc.vanderbilt.edu/courier/web/1000@c79c7a50a2a8129191d1d5068946723f/wmCompose.html>. The page header includes the Vanderbilt University logo, 'Information Technology', and 'SECURE FILE TRANSFER'. The user is logged in as 'daleplummer@gmail.com (Guest)'. The interface has two tabs: 'File Manager' and 'Send File', with 'Send File' being the active tab. At the top of the 'Send File' section are buttons for 'Send', 'Save Now', and 'Discard'. Below these are input fields for 'To:', 'Subject:', and 'Files:'. The 'Files:' field includes 'Choose File' and 'Choose from File Manager' buttons, along with a note about using a 'Large File/Folder Applet' for files larger than 2GB. A rich text editor is present with the text 'Use Rich Text Formatting »'. An 'Additional Options' section contains a checked checkbox for 'Send copy to myself' and another set of 'Send', 'Save Now', and 'Discard' buttons. The footer of the page features the 'Secured by Accellion.' logo and copyright information: '©2000-2013 Accellion, Inc. All Rights Reserved.'. The browser's taskbar at the bottom shows several open files, including 'password_strength.png' and several 'VUMC-POWERPOIN...' files.

<https://its.vanderbilt.edu/security/secure-file-transfer>

Tools - encfs

```
dalep@biostat666: ~  
dalep@biostat666:~$ encfs ~/.privatetestuff_encrypted ~/privatetestuff  
The directory "/home/dalep/.privatetestuff_encrypted/" does not exist. Should it be created? (y,n) y  
The directory "/home/dalep/privatetestuff" does not exist. Should it be created? (y,n) y  
Creating new encrypted volume.  
Please choose from one of the following options:  
  enter "x" for expert configuration mode,  
  enter "p" for pre-configured paranoia mode,  
  anything else, or an empty line will select standard mode.  
?>  
  
Standard configuration selected.  
  
Configuration finished. The filesystem to be created has  
the following properties:  
Filesystem cipher: "ssl/aes", version 3:0:2  
Filename encoding: "nameio/block", version 3:0:1  
Key Size: 192 bits  
Block Size: 1024 bytes  
Each file contains 8 byte header with unique IV data.  
Filenames encoded using IV chaining mode.  
File holes passed through to ciphertext.  
  
Now you will need to enter a password for your filesystem.  
You will need to remember this password, as there is absolutely  
no recovery mechanism. However, the password can be changed  
later using encfsctl.  
  
New Encfs Password:  
Verify Encfs Password:  
dalep@biostat666:~$ █
```

Tools - encfs

```
dalep@biostat666: ~  
dalep@biostat666:~$ ls -la .privatetestuff_encrypted/ privatetestuff/  
privatetestuff/:  
total 12  
drwx----- 2 dalep dalep 4096 Sep 10 14:06 .  
drwxr-xr-x 100 dalep dalep 4096 Sep 10 14:02 ..  
-rw-rw-r-- 1 dalep dalep 70 Sep 10 14:06 secrets.txt  
-rw-rw-r-- 1 dalep dalep 0 Sep 10 14:05 secrets.txt~  
  
.privatetestuff_encrypted/:  
total 16  
drwx----- 2 dalep dalep 4096 Sep 10 14:06 .  
drwxr-xr-x 100 dalep dalep 4096 Sep 10 14:02 ..  
-rw-rw-r-- 1 dalep dalep 1077 Sep 10 14:02 .encfs6.xml  
-rw-rw-r-- 1 dalep dalep 78 Sep 10 14:06 RVlz9ltqj4iajz1SqXP-LUAo  
-rw-rw-r-- 1 dalep dalep 0 Sep 10 14:05 sKSYq-iqoL01cpiEsvvTTzh0  
dalep@biostat666:~$  
dalep@biostat666:~$ cat privatetestuff/secrets.txt  
111-22-3333  
  
(615) 555-1212  
  
password1  
  
account number 221100  
  
etc.  
  
dalep@biostat666:~$ █
```

Tools - encfs

```
dalep@biostat666: ~  
dalep@biostat666:~$ fusermount -u privatestuff  
dalep@biostat666:~$ ls -la privatestuff/  
total 8  
drwx----- 2 dalep dalep 4096 Sep 10 14:02 .  
drwxr-xr-x 100 dalep dalep 4096 Sep 10 14:02 ..  
dalep@biostat666:~$  
dalep@biostat666:~$  
dalep@biostat666:~$ ls -la .privatestuff_encrypted/  
total 16  
drwx----- 2 dalep dalep 4096 Sep 10 14:06 .  
drwxr-xr-x 100 dalep dalep 4096 Sep 10 14:02 ..  
-rw-rw-r-- 1 dalep dalep 1077 Sep 10 14:02 .encfs6.xml  
-rw-rw-r-- 1 dalep dalep 78 Sep 10 14:06 RVLz9ltqj4iajz1SqxP-LUAo  
-rw-rw-r-- 1 dalep dalep 0 Sep 10 14:05 sKSYq-iqoL01cpiEsvvTTzh0  
dalep@biostat666:~$  
dalep@biostat666:~$  
dalep@biostat666:~$ encfs ~/.privatestuff_encrypted ~/privatestuff  
EncFS Password:  
dalep@biostat666:~$  
dalep@biostat666:~$  
dalep@biostat666:~$ ls -la privatestuff/  
total 12  
drwx----- 2 dalep dalep 4096 Sep 10 14:06 .  
drwxr-xr-x 100 dalep dalep 4096 Sep 10 14:02 ..  
-rw-rw-r-- 1 dalep dalep 70 Sep 10 14:06 secrets.txt  
-rw-rw-r-- 1 dalep dalep 0 Sep 10 14:05 secrets.txt~  
dalep@biostat666:~$  
dalep@biostat666:~$  
dalep@biostat666:~$ █
```