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/privacysummary.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

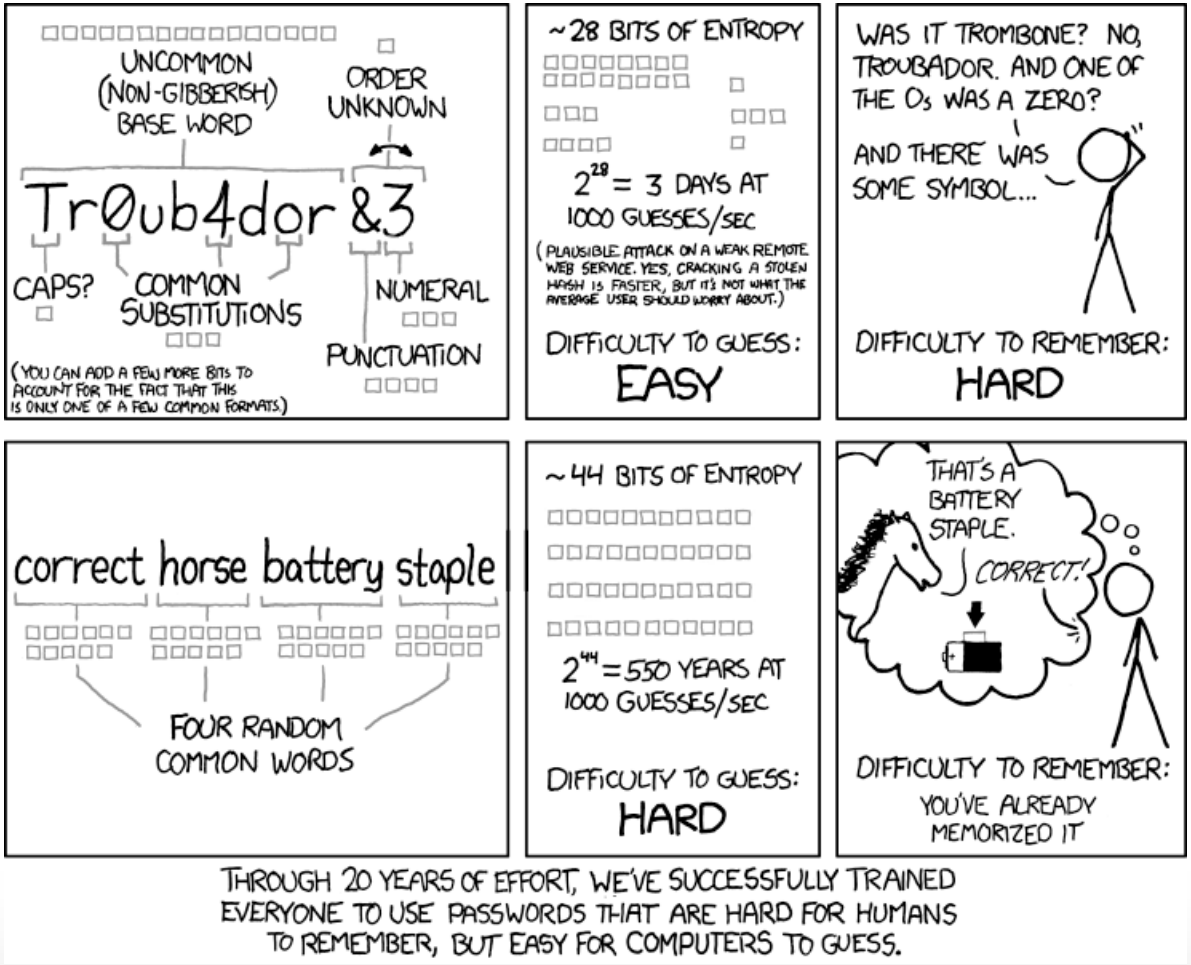
- VUMC policy and other regulations require notification in the event of any unauthorized disclosure of individually identifiable patient or other personal information.
- Known or suspected incidents involving breach of PHI are reported to the VMC Privacy Office
- If you lose a laptop or other device, 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/



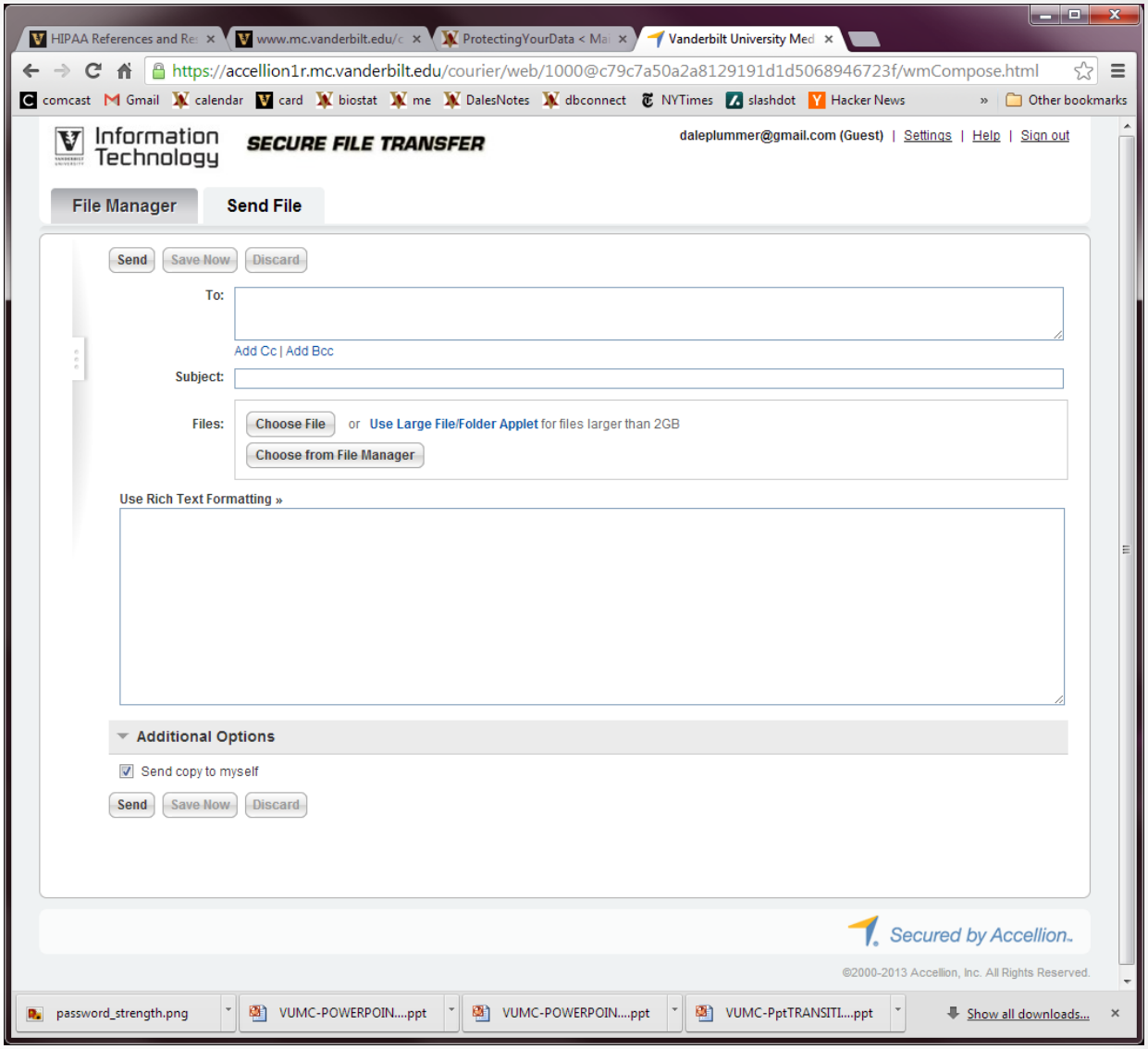
Tools

- 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/coveredentities/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/

Tools

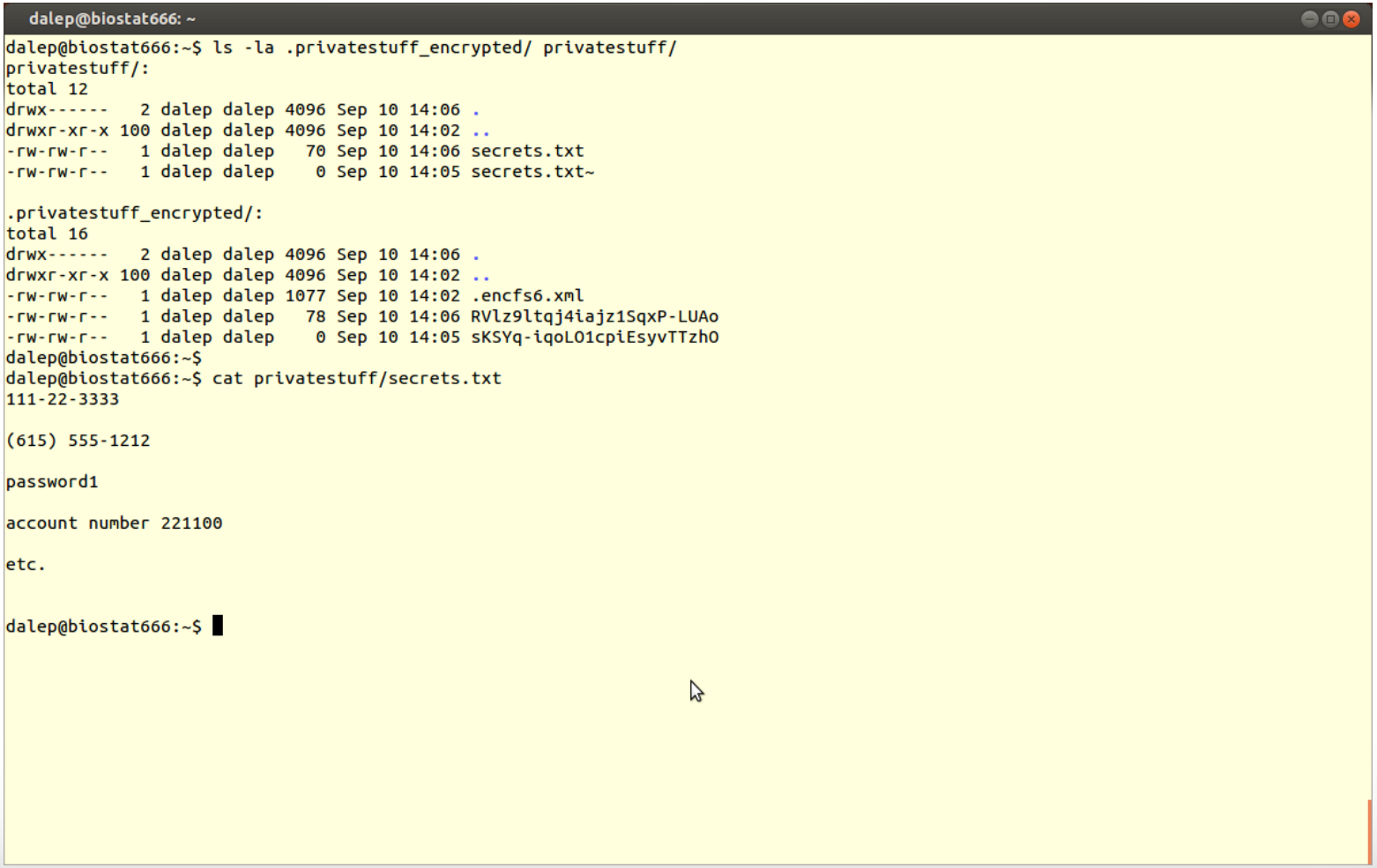


https://its.vanderbilt.edu/security/secure-file-transfer

Tools - encfs

```
dalep@biostat666: ~
dalep@biostat666:~$ encfs ~/.privatestuff_encrypted ~/privatestuff
The directory "/home/dalep/.privatestuff_encrypted/" does not exist. Should it be created? (y,n) y
The directory "/home/dalep/privatestuff" 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 .privatestuff_encrypted/ privatestuff/
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~
.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-iqoLO1cpiEsyvTTzhO
dalep@biostat666:~$
dalep@biostat666:~$ cat privatestuff/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-iqoLO1cpiEsyvTTzhO
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:~$
```