DSCI 519
SEMESTER PROJECT

ADITI KIRAN GOWDA

A report on The Centre's responsibilities for protecting data

COMPONENTS OF THE GIVEN PROBLEM STATEMENT

1. **The Centre:** It is a bioinformatics company which specializes in molecular pathology, clinical genetic labs, and research
2. **Cartagenia:** It is a company that provides it's platform as a service to another company. It has a contract with the Centre to provide a clinical informatics platform that the Center uses for data analysis.
3. **AWS(Amazon Web Service):** AWS provides cloud service to the Centre for managing data.

PRIVACY AND INFORMATION SECURITY POLICY

1. The HIPAA security rule defines any information related to health as Protected Health Information(PHI) using strict policies. The company that has PHI is bound to implement security measures defined by HIPAA. HIPAA security rules implement certain restrictions on usage and sharing of PHI. Violation of HIPAA Rules results in facing financial or criminal penalties. Justifying ignorance of the HIPAA law as a defense is not valid.
2. The HIPAA considers Public health to be the following:
    - Individually identifiable health information created, collected, transferred, or retained by a HIPAA-covered entity in connection with the provision of healthcare, payment for healthcare services, or use in healthcare operations.
    - The provision of healthcare
    - Health information such as diagnoses, treatment information, medical test results, and prescription information
    - Demographic information such as birth dates, gender, ethnicity, and contact and emergency contact information
    - National identification numbers

3. PHI are physical records. ePHI are any records that are created, stored, transmitted or received electronically
4. PHI only applies to information on patients or members of health plans. It excludes information from educational and employment records, as well as health information kept by a HIPAA-covered entity in its position as an employer.
5. The company is expected to protect the HIPAA-covered PHI data and shall segregate it from data that is not considered PHI. When an individual can be identified from the information, it is considered PHI. When all identifiers are removed from health data, it no longer qualifies as protected health information, and the HIPAA Privacy Rule's limitations on uses and disclosures are no longer in effect.
6. PHI is any health information that can be tied to an individual.The company is supposed to store the individual data securely.The company stores the following data items of an individual-
    - Names (Full or last name and initial)
    - All geographical identifiers smaller than a state, except for the initial three digits of a zip code if, according to the current publicly available data from the U.S. Bureau of the Census: the geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and the initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000
    - Dates (other than year) directly related to an individual
    - Phone Numbers
    - Fax numbers
    - Email addresses
    - Social Security numbers
    - Medical record numbers
    - Health insurance beneficiary numbers
    - Account numbers

- Certificate/license numbers
- Vehicle identifiers (including serial numbers and license plate numbers)
- Device identifiers and serial numbers;
- Web Uniform Resource Locators (URLs)
- Internet Protocol (IP) address numbers
- Biometric identifiers, including finger, retinal and voice prints
- Full face photographic images and any comparable images
- Any other unique identifying number, characteristic, or code except the unique code assigned by the investigator to code the data

7. The HIPAA Security Rule mandates that covered entities safeguard PHI against reasonably expected threats. Although HIPAA is not technology specific, covered entities must establish measures to maintain the confidentiality, integrity, and availability of PHI. The particular safeguards that should be implemented are left to the discretion of the covered entity.
8. Safeguards to be implemented by a covered entity are of 3 types:
    - **Physical:** Physical records are kept to ensure accountability. It also includes details such as the date and time the user accessed the PHI. This information is critical for determining whether or not there have been any violations. PHI and ePHI are kept in a secure device that is kept in a secure location with a locking mechanism. Strong authentication techniques, such as a biometrics system, can be used to monitor the premises and authenticate authorized users.
    - **Technical:** Technologies such as encryption software and firewalls are covered under technical safeguards. All the technologies used for data protection are included here.

- **Administrative:**Only a small number of legitimate users have access to sensitive data, owing to access control methods. Individuals or groups who are unable to access the data can also be indicated explicitly. The reception desk, for example, should never have Read-Write access to the PHI. Security awareness training should be provided to all users who have access to the data.

9. When the data is transferred between two entities (local and cloud), checksums should be calculated and confirmed.
10. All clinical data is exchanged within the Hospital's network. When considering cloud storage, Amazon EC2 enabled services are employed. HTTPS is used to connect to Cartagenia. (Cartagenia transfers data at the request of a patient or physician, and de-identified data is archived using Amazon S3 or Glacier storage services.)
11. Physical access of the data is allowed to Director of Bioinformatics and his designees, IT team, pre approved vendors for maintenance and other purposes.
12. Before being sent, data that is shared for research purposes or with vendors for the purpose of debugging software or analysis difficulties is given a PLMUID.
13. Data recovery may be performed by the Bioinformatics Director and his designees. Access to primary archive servers is specified by the Computing Equipment Access Log
14. While on site, you have access to the hospital's network. Offsite access is not permitted, however VPN access is. The Director of Bioinformatics and his designees have network access.
15. Hospital emails are allowed for communication on the internet. Personal emails are not allowed.
16. HIPAA-compliant data and hardware are kept separate from data and hardware that does not belong to the facility.
17. . Access to the hardware is safeguarded by at least three biometric security measures.
18. Dark internet connections are strictly prohibited.

19. Secure data communication can be established using the SSH Protocol. Data transferred outside is encrypted and uses protocols like rsync, SFTP, HTTPS

ACCESS CONTROL POLICIES

Since the Hospital sector is a huge organisation each user might have many different roles and many different user types may have the role. Hence listing access control policies keeping roles in mind is best suited that is Role based access control policy.

The given system can be represented in the Role based Access Control policy (RBAC) as follows:

A. Users:

1. IT Director
2. Bioinformatics Director
3. Bioinformatics Supervisor
4. System Administrator
5. Lab Manager
6. Medical Director
7. Other Employees

B. Objects

1. Clinical data
2. Non- clinical data
3. Research data

C. Roles –

1. Root
2. Bioinfo
3. Bio Infoclin
4. Clinical

5. Research
6. Hla
7. Smbgroup

D. **User to Role assignment**

1. Root – IT Director, Bioinformatics Director, Designee i.e Bioinformatics Supervisor, System Administrator
2. Bioinfo – Other employees
3. Bioinfoclin – Lab Manager
4. Clinical – Medical Director
5. Research – Bioinformatics Director, Designee i.e Bioinformatics Supervisor
6. Smbgroup – Other employees

E. Role to Permission assignment

1. Root – Access and Modify Non-clinical software or data, Clinical data, and Research data
2. Bioinfo – Access and Modify Non-clinical software or data
3. Bioinfoclin – Access and Modify Clinical data
4. Clinical – Access and Modify Clinical Data
5. Research – Access and Modify Research data
6. Hla – Access and Modify Clinical and Research Data
7. Smbgroup – Access Non-clinical software and data

MAC can be implemented with the following security levels with the topmost being the highest priority for security and integrity and followed by others from top to bottom and users can be divided amongst these categories

Directors
Administrators
Users
Researchers

The dom relationship can be written as follows:

Directors **dom** Administrators **dom** Users **dom** Researchers

The **objects** as mentioned above are:

Clinical data
Non-clinical data
Research data

DAC can be implemented with the following Access control lists

1. **Non- clinical**

Object **Subject**	**Non Clinical data**
Root	Read, Write
Bioinfo	Read, Write
Sbmgroup	Read

2. **Clinical data**

Object **Subject**	**Clinical data**
Bioinfo	Read
Bioinfoclin	Read
Clinical	Read
Smbgroup	Read,Write
Root	Read,Write
Hla	Read,Write

3. **Research data**

Object **Subject**	**Research Data**
Research	Read, Write
Hla	Read, Write
Root	Read,Write

OTHER REQUIREMENTS

1. Only integrity and confidentiality can be enforced by access control mechanisms. The availability of data is critical, and it is a factor that is lacking from the policy structure.
2. Storage Policies can be used to determine the data's availability, which might be either immediate or in the future. Data backup is also an important element to consider that ensures availability. Amazon S3 can be used to make the backups that are required for fast recovery. In.S3 ensures availability in the event of a calamity.
3. It is necessary to verify that the company's communications with Cartagenia are secure.There are no man-in-the-middle attacks since the system is secure. Malicious code injections and other types of attacks are also possible. Security features like firewalls and intrusion detection systems can help prevent spoofing. Additional techniques such as intrusion prevention systems and antivirus software can be used.
4. The network can be protected by Network Segmentation and Protected Enclaves. The highly sensitive data can be isolated from less confidential data. In this case, it can help to prevent privilege escalations or network later movements
5. The organization should keep audit records to ensure that those who have access to PHI are accountable.PHI should be kept in a secure location. Multiple security methods, such as biometrics, CCTV cameras, auditing entries, recording access timestamps, and so on, should be used to secure access to the room where the PHI data servers are kept.
6. The Center should implement policies correctly while performing backup and restore options. If there is a gap between the security policy definition and its implementation, it can create loopholes which would be entry points for attackers.
7. Multi Factor Authentication can be implemented by the company to ensure that authentication is carried out properly and there are no loopholes. Techniques of

authentication like smart cards, biometrics, facial recognition, long passwords can be used.

8. If RBAC is used to implement the access control in the system, it should be ensured that no invaders are trying to claim themselves to be the legitimate users of the system. Only the authorized users having the role should be allowed access.
9. Transport Layer Security (TLS) protocol should be preferred over Secure Socket Layer (SSL) as it is safer and better
10. Users of both organizations should be educated about the cyber threats and possible attacks. Users should know how to use the systems responsibly. .

POTENTIAL THREAT TO THE INFORMATION

1. Lack of integrity of origin can enable the attacker to gain unauthorized access to the network. Attackers can misuse this privilege for many reasons like stealing highly sensitive personal data, spreading malicious programs (malicious code injections) in the network and destroying its functioning and so on.
2. Social Engineering attacks like phishing can happen where high priority users like directors and root can be targeted. Once their system is compromised, attackers can gain control over the entire system
3. The Verifiability property of the security policy should be satisfied. Lack of verifiability can result into an attack-prone system with vulnerabilities
4. Subversion of the code can happen if proper security mechanisms are not in place. It should be ensured that security policy is implemented correctly
5. Insider Threats are very critical and can result into unauthorized access of highly sensitive data sets
6. Password hashing algorithms should be used in place of ordinary password systems, which are prone to password cracking attacks like brute-force

7. Weak Encryption Algorithms can result into cryptanalysis. This factor should be considered and advanced encryption algorithms should be used.

THREAT SPACE ANALYSIS

1. Cartagenia makes no mention of the encryption algorithms that were used to transfer the data over the network. Sniffing attacks and cryptanalysis can be caused by weak encryption algorithms. It can utilize AES-512 as an example of an algorithm. Cryptanalysis will be more difficult with a key size of 512 bits.
2. There is a dearth of effective monitoring methods and auditing measures in the Center. Everything that happens in the system should be audited and recorded. There should be a way to figure out how and who committed any frauds.
3. Employees of Cartagenia should not have direct access to PHI data at the Center. The data will be leaked if any adversaries are present in Cartagena, and the Center will be held liable. Between these two entities, there should be transparency, and the Center should be able to see what actions Cartagenia performs with the data during data analysis.
4. For checking integrity in the Center, algorithms like MD5 are utilized, which are rather weak and can be cracked. It is necessary to improve hashing algorithms.
5. There are no requirements for who will preserve the logs or who will monitor them at the Center. Subversion of security policies will occur if there is no auditing system in place. The business can also use surveillance technologies such as cameras and biometrics. Extremely privileged users, such as directors, should also verify logs.

BEST PRACTICES VS HIGH ASSURANCE TRUSTED SYSTEMS

PROS OF BEST PRACTICES

- Current best practices like defense in depth, cross domain solutions or air-gap are highly compatible with the applications used and ensure considerable security. High Assurance systems like Multics are more expensive. Moreover, they tend to slow down the overall system efficiency when integrated with current systems
- Auditing the data for individual accountability is a very important factor of cybersecurity. This feature is absent in the high assurance systems like Multics.

PROS OF HIGH ASSURANCE TRUSTED SYSTEMS

- In High Assurance Systems, isolation is achieved by hardware support that is Protection Rings. The current process ring number and the access bracket decide the access for that process (subject) on the data (object). In this system, subversion is difficult. Hardware does not allow subversion and hence the system is highly secure.
- High Assurance means High Trust. It is assumed that the high assurance system will implement the policies correctly. There will be one to one correspondence between the security policy and its implementation. This is achieved using high assurance systems and it is debatable in the case of best practices

REFERENCES

1. What is considered protected health information under HIPAA.

https://www.hipaajournal.com/what-is-considered-protected-health-information-under-hipaa/

2. HC-C04101 Data Policies
3. National Institute of Standards and Technology - An Introductory Resource Guide for Implementing the Health Insurance Portability and Accountability Act (HIPAA) Security Rule

4. Amazon Web Services – Using AWS for Disaster Recovery (October 2014) 11
5. Amazon Web Services–Architecting for HIPAA Security and Compliance on Amazon Web Services (August 2016)
6. Cartagenia Data Security & Confidentiality Policy