Homework 1

CISC 6644 Privacy and Security in Big Data

Department of Computer and Information Sciences, Fordham University

Total Score 20 points

A. Answer to the following questions (in one or a few sentences). (1*5=5)

1. What are the characteristics of big data?
2. Give three differences between traditional and advanced security approaches.
3. How can Crypto be used for Big Data security?
4. What is RTAP?
5. Tell a few advanced threats for Big Data.

B. Select True or False for the following statements (1*5=5)

6. Traditional security solutions cannot bridge the gaps between data breach protection and compliance.
7. Pattern matching is a function of big data.
8. The storage part of Hadoop is called MapReduce.
9. The goal of designing Hadoop is to process big data with reasonable cost and time.
10. Big Data is designed for access, not security.

C. Select appropriate answer(s) (1*5=5)

11. **Which dimension of big data may correspond to the analysis of data security and intelligence**

    (a) Volume (b) Variety (c) Veracity (d) Value

12. **As organizations started adopting Hadoop at an increasing rate, the future use of Hadoop may need for additional capabilities, including:**

    (a) Improved data storage
    (b) Improved extract, transform and load features for data integration
    (c) Improved data warehousing functionality
    (d) Improved security, workload management, and SQL support

13. **Which statement is wrong in terms of Big data**

    a) No real indexes b) Data is structured
    c) Data can be processed in real time d) It is not just about the volume of data

14. **What does "Velocity" in Big Data mean?**

    a) Speed of input data generation b) Speed of individual machine processors
    c) Speed of ONLY storing data d) Speed of storing and processing data

15. **Which of the following is true?**

    a) Map>Reduce > Combine b) Combine >Reduce>Map
    c) Map >Combine >Reduce d) Reduce >Combine >Map

Homework 3

Department of Computer and Information Sciences, Fordham University

Total Score 15 points

A. Answer to the following questions (in one or a few sentences). (5*2=10)

1. Write down a few methods that are used to steal Big data.
2. Name different types of cryptographic attacks.
3. What are the cryptographic modes of operation?
4. What are the applications of public-key cryptosystems?
5. Describe differential cryptanalysis.

B. Select True or False for the following statements (1*5=5)

6. Conventional security approaches are perimeter-based and advance approaches are data-centric based.
7. Loss of authenticity does not fall into database security category.
8. The hash function can be combined with encryption.
9. Digital signature is used to verify a message comes intact from the claimed sender.
10. RSA security can be broken by brute force key search.

C. Select appropriate answer(s) (1*5=5)

11. What can be also a kind of sources of big data?
    a) Logs b) Network traffic c) Application behavior d) User behavior

12. Which issue does not fall into security category?
    a) Legal and ethical b) Policy c) System-related d) None of them

13. What is the hash function?
    a) Mapping data of arbitrary size to data of fixed size
    b) Indexing and retrieving items in a database
    c) Digital fingerprint of an original message
    d) Taking a message input and returning a fixed-size alphanumeric string

14. Which of the following statements is inappropriate?
    a) People confidence in the security of SHA-3 candidates is very high
    b) SHA-3 candidates are based on new constructions
    c) It is not vulnerable to well-known attacks (e.g., length extension attack)
    d) None of them

15. Which one could be the best in terms of security
    a) ECB b) CTR c) CFB d) OFB

Homework 4

Department of Computer and Information Sciences, Fordham University
Total Score 15 points

A. Select True or False for the following statements (1*5=5)

1. An important concern with a decision maker is whether he can control the sensitivity of the data he gives to others.
2. The operation that de-associates the relationship of attribute by partitioning a set of data records into groups and shuffling their privacy values within each group is called anatomization.
3. A decision maker generally requires to change the data to perform a privacy-preserving.
4. Legal measures can be a privacy preserving approach.

C. Select appropriate answer(s) (1*5=2.5)

5. There are numerous means of data mining methodologies, including...
    a) Mining numerous and new kinds of knowledge
    b) Mining knowledge in multidimensional space
    c) Pattern evaluation and pattern mining
    d) Handling uncertainty, noise, or incompleteness of data

6. What stage of security in data mining is the most critical
    a) Data miner & decision-maker
    b) Data prodder & data collector
    c) Data miner & data collector
    d) Decision maker & data provider

7. What is the outcome of KDD?
    a) Secured data
    b) Information
    c) Truth Discovery
    d) Useful information

8. Who is the responsible for preserving most sensitive information?
    a) Data provider
    b) Decision maker
    c) Data collector
    d) Data miner
9. Which of the following is not a data mining privacy-persevering approach/operation?
    a) Association rule mining
    b) Anonymization
    c) Encryption tool
    d) None of the above

10. What is/are the secure functionalities?
    a) secure comparison
    b) secure polynomial evaluation
    c) secure set intersection
    d) secures sum

D. Matching Questions (1*5 = 5points)

Determine which one is the best match for each problem/statement.

Limit the access	Trade privacy	Cryptographic tools	privacy-preserving decision tree
Secure number	Secure max	Secure Functionalities	Untrustworthy data
Data provider	Data Collector	Data miner	Decision maker
Quasi-identifier	Identifier	Sensitive Attribute	Data provenance

11. In what stage of data mining, an anti-tracking extension should be used for internet environment to protect data.
12. A method can be used to realize the original data and judge the credibility of the decision making outcome.
13. Which stage is the most responsible for adding unsecured data.
14. What can attribute to connect to external data to re-identify individual records.
15. A secure mathematical function.