Are the recommendations from recommender system trustworthy?

Sabah Suhail and Choong Seon Hong
Department of Computer Engineering,
Kyung Hee University,
Yongin, 446-701 Korea
sabah@khu.ac.kr, cshong@khu.ac.kr

Faheem Zafar and Adeel Anjum
Department of Computer Science,
COMSATS Institute of Information Technology,
Islamabad, Pakistan
faheemiiui@gmail.com, adeel.anjum@comsats.edu.pk

Abstract

The sensor data from multimodal data sources is characterized by continuous data streaming which passes through recommendation system curation layers for extensive processing to infer context-aware personalized recommendations to users. Such extensive processing of healthcare big dataset tends to introduce performance overhead during the entire life-cycle of data generation, data processing and data analysis. On the other hand, enforcing security and privacy on individual-based data is a promising requirement of healthcare datasets. However, adding security layer on such complex framework results in another layer of complexity making the performance requirement hard to achieve. Hence, a trade-off between performance and security constraints is required for wellness recommender system. In this paper we have proposed a performance-oriented security solution for a personalized wellness recommender system.

I. INTORDUCITON

Currently, many people are using activity tracker wearable devices to organize their daily fitness activities. The wearable devices monitor and records user's physical activities through the sensors within the devices. The sensor data is then send to a recommender application (usually on smart phone) in order to analyze user's activity pattern and suggest recommendations accordingly for example, user should do exercise or should take balanced diet etc.

Due to connectivity of sensor devices with untrusted Internet, there is high risk of privacy and data breach of user's personal information. To mitigate such risk, the sensor data needs to be secured while it is being communicated to user. For instance, [1-2] have addressed the data trustworthiness issues in IoT devices and the secure data processing by the recommender system utilizing cloud resources. However, security adds a layer of overhead which usually affect the system performance. In this paper we introduce a performance-oriented security solution for the recommender system.

In a personalized wellness recommender system, a multi-tier architecture for the security and privacy of data can be classified at different levels as:

A) Data from sensors
B) Data in transit
C) Data at rest
D) Data sharing

Fig.1 shows the security and privacy threats in a recommender system.

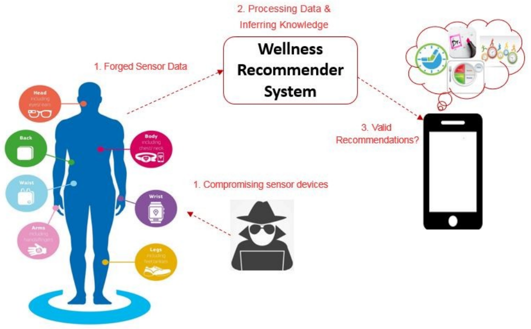
2. Processing Data &
Inferring Knowledge
Wellness Recommender System
1. Forged Sensor Data
Head including eyes/ ears
Back
Body including chest/ neck
Waist
Wrist
Arms including hands/fingers
Legs including feet/ankles
1. Compromising sensor devices
3. Valid Recommendations?

Figure 1: Recommender System: Security and Privacy Threats

A. Data from sensors: Secure communication channel

A malicious user may compromise any sensor node and attempt to forge the sensor data. Under such circumstance, the data arriving at the recommender system is modified data values rather than the actual data values.

We evaluate the trustworthiness of sensor data as it arrives at recommender system. Such abnormal behavior can be detected by integrating provenance in IoT [1-2].

Experimental evaluation: trustworthy sensor data

To illustrate the concept of abnormal behavior of sensor nodes, we have setup RPL-collect (Node1: sink and others: sender nodes) environment at Cooja. The evaluation system in recommender system will check the data behavior and compare with baseline data.

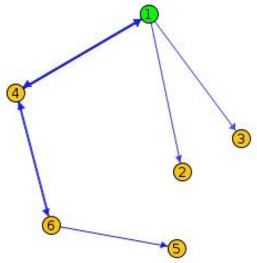

Figure 2: Cooja Simulation

-Results: Fig.3 shows the routing metric for a particular node with respect to time interval. By analyzing such information, we can evaluate the behavior (routing metric and power consumption) of any node (Fig.4 and Fig.5).

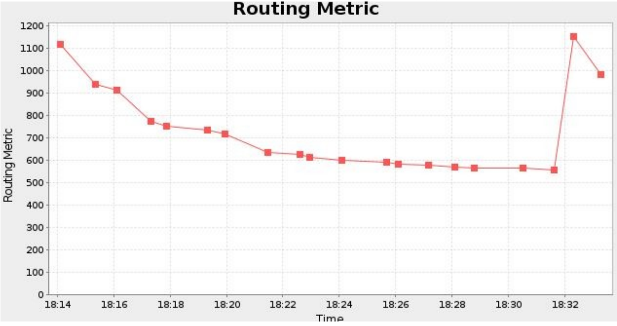

Figure 3: Routing path of a node

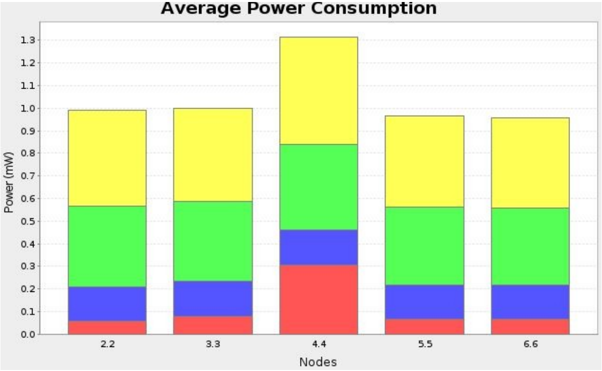

Figure 4: Power consumption of a node

II. PERFORMANCE-ORIENTED SECURITY SOLUTION

B. Data in transit: Secure communication channel

In order to make the communication channel secure, the data flow across layers must be make confidential to thwart attacker from accessing any information. Keeping in view a performance-oriented security solution, recommender system needs to adopt already tested and trusted security mechanisms rather than designing an entirely novel mechanism.

The bi-level secure communication channel used by wellness recommender system is as follows:

Level I: Secure Socket Layer/ Transport Layer Security (SSL/TLS)

RESTful web services can be used for intra-communication among curation layers. They leverage the full capacity of the HTTP protocol which uses cryptographic protocols SSL/TLS to ensure the secure and authentic delivery of data communication over network. Thus HTTPS prevents eavesdropping and man in middle attack as it encrypts all traffic for a session. However; it has multiple loop holes for instance: data is logged in plain-text server logs on the receiving HTTPS server; data is logged in browser history. To overcome this problem, the recommender systems may encrypt input parameters in query string.

Level II: Encrypted parameters

Considering the security threats to SSL/TLS, recommender systems may adopt a light-weight symmetric key encryption algorithm Advanced Encryption Standard (AES) to encrypt the parameter in URL. AES normally works on a fixed block size and hence take approximately the same time independently of input. Thus the time complexity of AES is O (1).

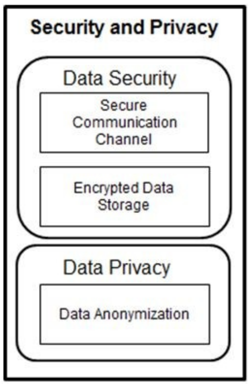

Figure 5: Security and Privacy Module

Experimental evaluation: performance-oriented secure communication

To check for a performance-oriented system solution, we have configured a HTTP client that sends encrypted parameters to a web service on server where they are decrypted.

-Key Generation: For key generation, Java KeyStore (JKS) is used. It is a repository of security certificates for instance, authorization certificates or public key certificates plus corresponding private keys. The private keys generated are used during AES encryption and decryption process.

-Results: The average time required for AES encryption is approx. 364.4 ms and for decryption is 360.8 ms (Fig.7). Hence, the time taken by web service can be computed as:

$$\mathbf{T = T_E + T_{WS}}$$

where T_E is the time required for encrypting and decrypting the parameters and T_{WS} is the time required for a web service to perform required operation.

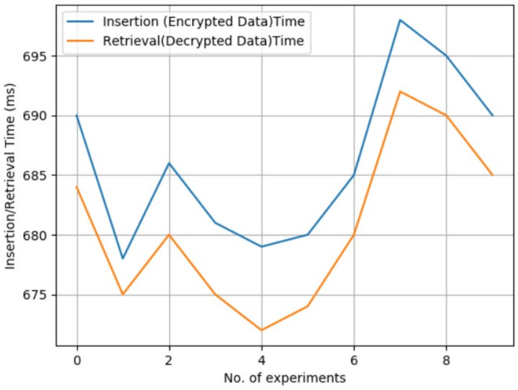

Figure 6: Secure Data Communication Encryption and Decryption Time

C. Data at rest: Secure Storage:

For secure data storage, we have used light-weight method for inserting encrypting data in the tables through procedure calls (Algorithm 1). For retrieval, the procedure will decrypt the requested data.

Experimental Evaluation: performance-oriented secure storage

-Key Generation: To encrypt the parameters of a query string, we need to generate symmetric keys. Using the SQL Server encryption hierarchy, we have opted for the option to secure the encryption keys through certificate.

-Results: The average time to insert encrypted data in table is 686.2 ms while the average time to retrieve the decrypted parameters is 680.7 ms (Fig.8).

```
 Algorithm 1: Procedure for encrypting parameters
CREATE PROCEDURE [dbo].[uspAddEncryptedParam]
–Parameters
@parameter
AS
BEGIN
  OPEN SYMMETRIC KEY SKey –Encryption Key
  –Using Certificate for decrypting the key
  DECRYPTION BY CERTIFICATE Cert;
  Declare @  EncryptedData varbinary(size);
  Set @EncryptedData=       EncryptByKey(KeyGUID('SKey'),
@parameter);
  Insert Into table
  (parameter)
  values
  (@EncryptedData)
   close SYMMETRIC KEY SKey;
      SQL query
```

D. Data sharing: eHealthData publishing

The dissemination of Electronic Health Records can be extremely beneficial as it allows researchers to perform statistical analysis or data mining tasks for decision support. However, privacy preservation on anonymous release shared with researchers-demands a privacy model that must be able to meet the following challenges: it should be able to strike a balance between the privacy and utility of released data set; it should be able to preserve the individual-based privacy.

Considering the performance objective, we refer to [3] as anonymization model in the recommender systems.

III. CONCLUSION

In this paper, we have formulated a performance-oriented security solution for wellness recommender system. The data from a variety of sensor devices is required to be protected from illegal usage at four stages including sensors, transit, storage and sharing. We have proposed light-weight solutions to ensure the security and privacy of individuals participating in wellness recommender systems. However, a better idea for integrating provenance in IoT devices will be part of our future work.

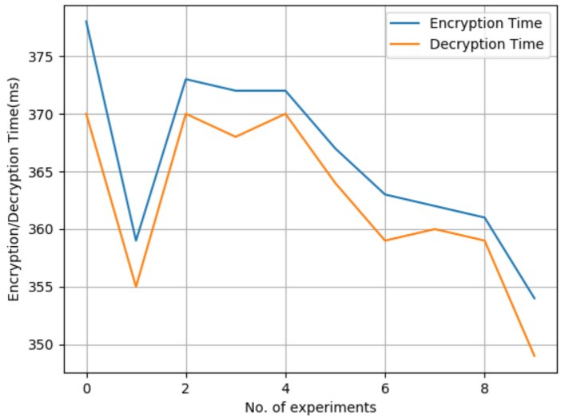

Figure 7: Secure Data Storage Encryption and Decryption Time

ACKNOWLEDGEMENT

This work was supported by the Industrial Core Technology Development Program(10049079 , Development of Mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea). *Dr. CS Hong is the corresponding author.

REFERENCES

[1] Suhail, Sabah, and Choong Seon Hong.” A Secure Provenance-Aware Model for Internet of Things”. (2016): 1154-1156.

[2] Sabah Suhail, Choong Seon Hong, Abid Khan et al. Introducing Secure Provenance in IoT: Requirements and Challenges. The Int. Workshop on Secure Internet of Things) SIOT 2016), held in Conjunction with ESORICS 2016.

[3] Suhail, Sabah, and Choong Seon Hong.” Privacy Preservation in Skewed Data using Frequency Distribution and Weightage (FDW)”. (2016): 769-771.