User authentication using gait and enhanced attribute-based encryption: a case of smart home

Lim Wei Pin, Manmeet Mahinderjit Singh
School of Computer Sciences, University Sains Malaysia, Penang, Malaysia

ABSTRACT

With the increasing popularity of the internet of things (IoT) application such as smart home, more data is being collected, and subsequently, concerns about preserving the privacy and confidentiality of these data are growing. When intruders attack and get control of smart home devices, privacy is compromised. Attribute-based encryption (ABE) is a new technique proposed to solve the data privacy issue in smart homes. However, ABE involves high computational cost, and the length of its ciphertext/private key increases linearly with the number of attributes, thus limiting the usage of ABE. This study proposes an enhanced ABE that utilises gait profile. By combining lesser number of attributes and generating a profiling attribute that utilises gait, the proposed technique solves two issues: computational cost and one-to-one encryption. Based on experiment conducted, computational time has been reduced by 55.27% with nine static attributes and one profile attribute. Thus, enhanced ABE is better in terms of computational time.

Keywords:
Biometrics
Cryptographic
Passwordless
Machine learning
Encryption

1. INTRODUCTION

Smart homes are filled with connected products aiming to make life smarter and easier. However, several challenges confront smart homes in the IoT context, including interoperability issues, where different devices and systems must work together for optimised performance [1]. Real-time data processing should be considered, and existing frameworks are not sustainable [2]. The issues of security and privacy are another challenge given the large number of devices that are susceptible to attack. The data collected are sensitive personally identifiable information. Thus, these data must be handled carefully to prevent any disclosure of information. One preventive measurement besides the usage of encryption in securing data would be employ authentication mechanism. Simple authentication methods that require only username and password combinations are inherently vulnerable to brute force, phishing, key logging and even man in the middle leading to personal data being stolen for unethical purpose. One countermeasure technique is by adopting passwordless authentication method.

Smart homes face serious data security and privacy issues. Various attacks can lead to data privacy issues, such as time, location, data, and user-based privacy attacks. Many smart home devices have sensors that constantly capture user activities. The information transmitted from smart homes to the cloud can be captured by network observers to infer privacy-sensitive activities. Three categories of solutions aim to solve the data privacy issue in smart homes: policy based [3], cryptography based [4], and data anonymisation based [5]. Policy-based solutions create a set of rules or policies to determine data access, such as using the
Jessor language to create a policy to ensure data privacy in the cloud [6]. Cryptography-based solutions use encryptions, such as attribute-based encryption (ABE), to encrypt the data before transmitting them to the cloud. Compared with the two other kinds of solutions, cryptography-based solutions using ABE require no architectural redesign and can be implemented in either a legacy or a new system. However, ABE has limitations, such as computational time and the linear increase in the length of the ciphertext scale with the number of attributes. Therefore, ABE is not suitable for one-to-one-encryption owing to its lack of unique attributes. Gait, on the other hand, can provide the user with a unique identity. In this research, ABE with gait profiling will be explored. The usage of gait or human movement such as walking, standing etc is taken into context. Gait, being a biometric is a suitable candidate for designing passwordless based authentication model. A gait profile will be built and combined with ABE to provide one-to-one encryption and reduce the number of attributes used. This enhanced ABE can improve data privacy in smart homes by providing a unique identity for one-to-one encryption.

The objectives of this research are to; i) design and evaluate an enhanced clustering and classification algorithm to generate a gait profile, ii) propose an enhanced gait profile for user identification, and iii) build an enhanced ABE technique using the generated gait profile. This research focuses on a cryptography-based solution that uses ABE. Moreover, it attempts to create a behavioural profile using gait, which is unique to a user, to substitute for the multiple attributes needed to identify a person. The outline of the paper is as follows. Section 2 present the related work. Section 3 present proposed work and section 4 discuss the experimental results. Finally, section 5 depicts the conclusion and future work.

2. STATE OF ART

With almost 30 billion connected devices globally in 2025, more than 500 billion Gbs of data at rest and in transit will be generated from these devices’ sensors and processing tasks. According to [7], [8], the vast amount of data generated and processed will result in several challenges to the IoT field, including data integration, real-time processing, and security and privacy. Vulnerability exists within smart device applications, such as smart homes due to characteristics of the devices and its operating modes. The features of sensors within a connected application are always ‘on’, attentive to contextual settings, and interconnected as well as operate anytime and anywhere. Next, description on smart home is presented.

2.1. An example of internet of thing-smart home security

A smart home refers to a house where all the electronic devices are connected and capable of communicating with one another. Smart home devices range from temperature sensors, humidity sensors, and smart lights to smart speakers, smart TVs, and smart air conditioners [9]. However, connecting smart home appliances to the internet makes the users vulnerable to malicious attacks. Private information gathered by sensor data–such as health data, shopping preferences, and eating habits–can be stolen by intruders [10]. This risk is particularly true for data gathered in smart homes because these data are sensitive personally identifiable information [11].

A smart hub in a home is designed to control sensors and devices. It also receives information from all the devices connected to it, and if something happens or goes wrong, it immediately notifies its user via phone, SMS, or email in accordance with the user’s preferences. At that moment, all settings are packed in the config.jar file, which the hub then downloads and implements.

What happens when someone can interrupt the smart home’s system and gain control over the home controllers? Figure 1 shows an example of a smart device, a smart bulb that is connected to a Wi-Fi network and controlled over a mobile application.

Based on Figure 1, one of the security threats on the hub is due to the open access of the hub’s firmware, which could be publicly downloaded. Once the firmware is accessed, other files within the firmware could be easily analysed as well. Another threat is the lack of strong cryptographic services protecting the firmware files. The password from the root account in the shadow file is encrypted with the data encryption standard algorithm. As practice shows, this cryptographic algorithm is not secure or highly resistant to hacking. Therefore, an attacker can successfully obtain the hash through brute force and find the ‘root’ password. Another vulnerability in config.jar beyond login credentials is the IP addresses and device IDs for all the devices connected in the home. The manipulation of settings, such as modifying sensitive details, becomes possible. The config.jar file also consists of the owner’s phone number, which is meant to receive alerts and notifications. Another vulnerability point is the smart devices connected within the smart home. To set the smart bulb, a user needs to download the mobile application (iOS or Android), switch on the bulb, connect to the Wi-Fi access point created by the bulb, and then provide the bulb with the SSID and password from a local Wi-Fi network. A hacker can easily launch a man-in-the-middle attack on the device and manipulate the settings as long as the hacker shares the same local Wi-Fi network as the smart home system.
2.2. Attribute-based encryption concept

The concept of ABE was first introduced by [12], who presented a scheme called fuzzy identity-based encryption (FIBE), which is the fundamental of ABE. ABE is a type of public key cryptosystem where the secret key of a user and the ciphertext created are based on a set of attributes. ABE has significant advantages; it is a flexible one-to-many encryption method and is a promising tool for securing fine-grained data sharing and decentralised access control. Almost all existing ABE schemes admit that reasonably expressive decryption policies produce ciphertext whose size depends at least linearly on the number of attributes involved in the policy [13]. However, three challenges are associated with ABE: performance, key revocation, and collusion resistance. In term of performance, data overhead increases with the number of attributes because the attributes are translated into an access tree that grows as a function of attributes embedded in the encryption key and ciphertext. A full performance evaluation of ABE shows that data overhead occurs in ABE [14]. The issue of performance owing to a large number of attributes not only affects the accuracy of ABE but also fails to provide sound security protection. In addition, this variant of ABE may not be able to be employed for the IoT environment which is lightweight based [15]. Based on the original ABE, some variants of ABE for the IoT environment have been proposed. Table 1 lists some of the ABEs suitable for IoT and the benefits they provide.

<table>
<thead>
<tr>
<th>FIBE [12]</th>
<th>√</th>
<th></th>
<th></th>
<th></th>
<th>Scalability</th>
</tr>
</thead>
<tbody>
<tr>
<td>KP-ABE [12, 16]</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td>Scalability</td>
</tr>
<tr>
<td>CP-ABE [17]</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td>Scalability</td>
</tr>
<tr>
<td>HABE [18]</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td>Scalability</td>
</tr>
<tr>
<td>MA-ABE [19]</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td>Scalability</td>
</tr>
<tr>
<td>C-CP-ABE [20]</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td>Scalability</td>
</tr>
<tr>
<td>ECC-ABE [21]</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td>Scalability</td>
</tr>
</tbody>
</table>

Different ABE variants have been devised to solve expressive structure, scalability, and efficiency issues. In targeting ABE in an IoT environment, the schemes focus on either the pre-computation method or algorithm efficiency. However, none focus on attribute selection in ABE. Most schemes require an authority to distribute the attributes.

2.3. Attributes and user and entity behavioural analytics

User and entity behavioural analytics are a type of machine-learning model that uses advanced analysis and aggregate data from logs or other types of information to figure out a certain activity or behaviour. Some user behavioural profiling methods include mouse dynamics, keystroke dynamics, and gait dynamics. The three types of gait recognition are machine vision based, floor sensor based, and wearable sensor based [21]. Machine vision–based technology captures gait using video cameras, while floor sensor–based technology extracts gait through installed sensors on the floor. Wearable sensor–based technology collects gait data using bodywear motion detectors, such as accelerometers and gyroscopes.
challenge of gait dynamics is that it is easily affected by different factors, such as sickness or other physical changes in one’s body caused by injury. A better way would be to combine gait dynamics with other biometric systems. Gait is classified as a human biometric that represents human daily activities [22]. This study will focus on the use of both clustering and classification algorithms to create a profile that is unrestricted to a predefined model for user identification.

3. PROPOSED METHOD

The two most important aspects in this study are the proposed user profiling framework and the integration of the user profile in ABE. In the proposed framework of user profiling, a dataset goes through pre-processing, which includes feature selection. All features go through the ensemble feature selection algorithms.

3.1. Proposed framework of user profiling

The proposed framework can be divided into two parts: the training phase and the verification phase. In the training phase, data are collected first. Users perform activities such as walking, sitting, and lying down. Accelerometer and gyroscope data are collected through smartphones or wearables. The three-axial raw data are then transferred to the data pre-processing steps, where each datum undergoes filtering to filter out noise. Feature extraction is done because raw data cannot give any meaningful information. Each raw datum is sampled in a fixed window size, and feature extraction is carried out by calculating various time and frequency domain variables. Next, after all the features are extracted, feature selection begins. Only features that are closely related to the model are selected. Feature selection is used to reduce the dimensionality of the data. The ensemble feature selection technique [23] is used in feature selection. After selection, the features go through clustering algorithms for activity recognition.

3.2. Proposed method using K-mean algorithm and DBScan algorithm

The next step after feature selection is activity recognition. In a real-world scenario, when users perform various activities, the system cannot know what activities the users are performing precisely. Thus, a clustering algorithm is used to cluster each of these activities. Here, the K-mean algorithm is used to cluster the data. This algorithm has several significant disadvantages. First, the number of clusters K must be specified. Second, the algorithm can only detect hyperspherical clusters that are well separated. Third, it is sensitive to noise and outlier points because these points can affect the centroid being used in the cluster. Fourth, it is highly sensitive to the selection of initial centres. Improper initialisation might cause empty clusters, slow convergence, and bad local minima. A new K-mean initialisation algorithm will be proposed in this research. The proposed initialisation algorithm is a density-based clustering algorithm that groups point near one another. The advantage of the DBSCAN algorithm is that it is robust to outliers.

In standard CP-ABE, the trusted authority learns all the attributes of a person. However, the computational time increases with the number of attributes. These attributes are also static attributes that can be contained by two persons. To solve this problem, human gait profiling is utilised. Figure 2 demonstrates on proposed enhanced attribute-based encryption. Instead of only static attributes, the users first train their gaits, and then the trusted authority can get the gait profile of the user.

Users can include this attribute profile during encryption. During decryption, users present their gait data to a trusted authority, who then compares the gait data with the profile stored in the database.
4. RESULTS AND DISCUSSIONS

The discussion is divided into two parts: the user profile result and the ABE result. As mentioned earlier, the dataset used in this research is ‘Human Activity Recognition Using Smartphones’ from the UCI machine-learning repository [24]. The dataset contains all 561 features that have been extracted. To prevent a curve of dimensionality, the ensemble feature selection technique is applied. With the use of the five feature selection techniques, 16 features are selected based on the consensus list.

4.1. Clustering algorithm result in gait profile creation

After 16 features are selected, they go through clustering algorithms for activity recognition. The experiments are completed according to different clustering algorithms: K-mean, canopy, EM, and farthest first. Five iterations are done on different seeds. Table 2 shows that it is crucial to choose the correct seed number to attain higher accuracy.

<table>
<thead>
<tr>
<th>Seed</th>
<th>Farthest first (%)</th>
<th>EM (%)</th>
<th>Canopy (%)</th>
<th>K-mean (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>47.7</td>
<td>84.4</td>
<td>76.3</td>
<td>98.1</td>
</tr>
<tr>
<td>100</td>
<td>80.3</td>
<td>84.4</td>
<td>79.7</td>
<td>98.1</td>
</tr>
<tr>
<td>10,000</td>
<td>42.3913</td>
<td>84.4</td>
<td>91.1</td>
<td>98.1</td>
</tr>
<tr>
<td>Highest %</td>
<td>81.04</td>
<td>84.4</td>
<td>91.1</td>
<td>98.1</td>
</tr>
</tbody>
</table>

All the experiments are executed using WEKA tool [25], with the ‘classification via clustering’ meta-classifier having the same training and test set. From the result, the K-mean algorithm can provide the highest accuracy, followed by the canopy, EM, and farthest-first algorithms. The EM algorithm gives the most stable result, which is 84.4%, even with different numbers of seeds. Farthest first, canopy, and K-mean are sensitive to seed number because different numbers affect their accuracy. The K-mean algorithm provides the highest accuracy but is also sensitive to the seed number chosen.

4.2. Classification algorithm result in gait profile pattern matching

The results of the classification models are shown in Tables 3 and 4. Three algorithms are tested: support vector machines (SVM), random forest, and K-nearest neighbour (K-NN). The confusion matrix of the walking model is shown in Table 4. In Tables 3 and 4, A is accuracy; TPR is true positive rate; FPR is false positive rate; P is precision; R is recall; F is F-measure; ROC is receiver operating characteristics curve; PRC is precision-recall curve; and K is kappa statistic. The best result is highlighted.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>A</th>
<th>TPR</th>
<th>FPR</th>
<th>P</th>
<th>R</th>
<th>F</th>
<th>MCC</th>
<th>ROC</th>
<th>PRC</th>
<th>K</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-NN</td>
<td>80.7</td>
<td>0.81</td>
<td>0.01</td>
<td>0.81</td>
<td>0.81</td>
<td>0.80</td>
<td>0.90</td>
<td>0.58</td>
<td>0.80</td>
<td>0.1</td>
<td>4.98</td>
</tr>
<tr>
<td>Random forest</td>
<td>87.9</td>
<td>0.88</td>
<td>0.004</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>1.00</td>
<td>0.93</td>
<td>0.87</td>
<td>4.98</td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>50.6</td>
<td>0.51</td>
<td>0.02</td>
<td>0.52</td>
<td>0.51</td>
<td>0.49</td>
<td>0.95</td>
<td>0.42</td>
<td>0.49</td>
<td>1.61</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>A</th>
<th>TPR</th>
<th>P</th>
<th>F</th>
<th>MCC</th>
<th>ROC</th>
<th>PRC</th>
<th>K</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-NN</td>
<td>51.8</td>
<td>0.518</td>
<td>0.521</td>
<td>0.518</td>
<td>0.519</td>
<td>0.502</td>
<td>0.752</td>
<td>0.3</td>
<td>0.5099</td>
</tr>
<tr>
<td>Random forest</td>
<td>71.1</td>
<td>0.711</td>
<td>0.01</td>
<td>0.715</td>
<td>0.711</td>
<td>0.702</td>
<td>0.967</td>
<td>0.754</td>
<td>0.7007</td>
</tr>
<tr>
<td>SVM</td>
<td>8.9</td>
<td>0.09</td>
<td>0.039</td>
<td>0.099</td>
<td>0.09</td>
<td>0.064</td>
<td>0.05</td>
<td>0.675</td>
<td>0.073</td>
</tr>
</tbody>
</table>

Three activity models are tested. Laying on the bed activity result is not being shared. Based on the results shown in Tables 3 and 4 and laying on bed activity result, random forest outperforms the other two classifiers by providing the highest accuracy across all three activities (i.e., 87.9% in walking, 71.1% in sit_stand, and 87% in laying). Among the three algorithms, K-NN requires the least time to build the model, whereas random forest takes the longest time. A trade-off between accuracy and performance can be seen from the different models created.

4.3. Result of enhanced attribute-based encryption

The trusted authority in ABE needs a user profile to build a profiling system. The performance of the enhanced ABE in time and length of the ciphertext is discussed in this section. Table 5 shows the computational time of ABE according to different stages in the algorithm.
Table 5. Computational time of ABE

<table>
<thead>
<tr>
<th>Number of attributes</th>
<th>Setup(s)</th>
<th>Keygen(s)</th>
<th>Encrypt(s)</th>
<th>Decrypt(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0199</td>
<td>0.0533</td>
<td>0.0824</td>
<td>0.0113</td>
</tr>
<tr>
<td>30</td>
<td>0.0197</td>
<td>0.154</td>
<td>0.188</td>
<td>0.0162</td>
</tr>
<tr>
<td>50</td>
<td>0.0197</td>
<td>0.2564</td>
<td>0.2918</td>
<td>0.02144</td>
</tr>
<tr>
<td>70</td>
<td>0.02019</td>
<td>0.3604</td>
<td>0.4014</td>
<td>0.0275</td>
</tr>
<tr>
<td>90</td>
<td>0.0199</td>
<td>0.4611</td>
<td>0.5037</td>
<td>0.0332</td>
</tr>
</tbody>
</table>

The computational time is shown to increase with the number of attributes in all stages except for the setup. The setup stage requires a similar time to compute regardless of the number of attributes. In the decryption stage, the time to decrypt is also higher when the number of attributes increases, but it is negligible compared with those of the keygen and encryption stages. Typically, in an ABE application, around five to ten static attributes are used. However, the use of static attributes cannot identify a unique person. More attributes are needed to increase the accuracy of identifying a unique person. However, this will cause the computational time of ABE to increase significantly, which the enhanced ABE can solve. Using nine static attributes and one profile attribute can reduce the computational time to about 55.27%. Thus, enhanced ABE is better in terms of computational time. To test the security of the enhanced ABE, different case studies in which one case study will be discussed next. The case studies have three users: Alice, Bob, and Trudy.

Case study 1.

Alice's Attributes | Bob's Attributes | Trudy's Attributes |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Student, CS, Security</td>
<td>Lecturer, CS, Security</td>
<td>Student, CS, Security</td>
</tr>
<tr>
<td>Policy</td>
<td>CS and Security and Lecturer</td>
<td></td>
</tr>
</tbody>
</table>

Situation:
Alice encrypts the file with policy and wants Bob to decrypt it. Bob decrypts the file. Trudy tries to decrypt the file, although she is not authorised to do that.

Result:
In case study 1, Alice and Bob can decrypt the ciphertext, not Trudy. This is true because Bob contains attributes that satisfy the policy and is expected to be able to decrypt the file. Alice is able to decrypt the ciphertext, although her attributes do not satisfy the policy because the policy is created using Alice’s profile attribute. Trudy has the same attributes as Alice but is not able to decrypt the file because she does not have Alice’s profile attribute. Next section focuses on discussion.

4.4. Discussions

A comprehensive discussion is presented based on several scenarios:

a. Clustering and classification algorithms

Existing gait profile creation focuses on classification algorithms [26], [27]. However, using only classification algorithms requires a predefined model trained earlier. The proposed method uses clustering algorithms to cluster different gaits into different groups, and then these groups of data will be saved as user gait models. During verification, the unknown data will first be clustered into different groups, and the classification algorithm will be used to find the match of the user gait model created earlier. Thus, user identification is achieved. The advantage of this method is that users are not restricted to a predefined gait model, and thus, it can be used to create a more practical application.

b. Proposed enhanced attribute-based encryption

ABE is a one-to-many encryption algorithm, and thus, achieving one-to-one encryption is impossible in normal ABE. This problem can be solved with the proposed enhanced ABE. Data owners can choose to use one-to-many encryption or one-to-one encryption. In normal ABE, if the policy that the data owner specifies does not match the data owner’s attributes, then they will not be able to decrypt the data. With the enhanced ABE, the use of profile attributes allows the data owner to decrypt their own data.

c. Comparison with normal attribute-based encryption

Normal ABE is a one-to-many encryption system, which means one party can encrypt the data, while multiple parties can decrypt it. This is different from normal public key encryption, which is a one-to-one encryption system. In one-to-one encryption, a key exchange is needed by both parties to enable the decryption process. The advantage of one-to-one encryption is that it allows only the desired party to decrypt the data. The only possible way is to create more attributes to describe the person. However, using more
attributes increases the computational cost. By utilising gait as one of the attributes in ABE, the enhanced ABE allows both one-to-one and one-to-many encryption.

5. CONCLUSION

The proposed framework has several limitations that can be considered in future work. First is pertaining to the gait biometric adopted. Gait is chosen to build the profile. The disadvantage of gait is that posture variation will affect accuracy. This is the current problem of using gait as a profile. Sickness or injuries will affect the gait model of the user. Additional research must be conducted to solve this issue. Next, limitation revolves the method chosen which is K-Means. The initialisation method employs the idea of the density-based cluster algorithm DBSCAN to achieve its purpose. Overall, the proposed enhanced ABE will hopefully be more suitable for use in the IoT world.

ACKNOWLEDGEMENT

This work is supported by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme with project code FRGS/1/2020/ICT07/USM/02/2.

REFERENCES


**BIOGRAPHIES OF AUTHORS**

*Lim Wei Pin* is a Master Student in School of Computer Sciences. His research focus on cryptographic based applications and protocols. He can be contacted at email: limwp.student@usm.my.

*Dr. Manmeet Mahinderjit Singh* is an Associate Professor at the School of Computer Science. She was awarded a PhD Degree in Data Security from the School of Electrical Engineering, from University of Queensland, Australia, in 2012, and an M.Sc. in Computer Science from University of Sains Malaysia. Her research interests are primarily in information security, smart security, and sensor network security where she is the author/co-author of over 70 research publications. She can be contacted at email: manmeet@usm.my.