International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

The Aware Home: Data Analytics for Smart Homes

Sachin Kumar[1], Sahana .V[2]

[1]M.Tech in Computer Science & Engineering, CMR Institute of Technology, Banglore, India

[2]Assistant professor Computer Science & Engineering, CMR institute of technology, Banglore, India

Abstract: ***A System for maintaining security & preserving privacy for analysis of sensor data from smart homes, without compromising on data utility is presented. However the very nature of smart home data analytics is establishing preventive care. Data processing results should be identifiable to certain users responsible for direct care. Storing the personally identifiable data as hashed values withholds identifiable information from any computing nodes .Through a separate encrypted identifier dictionary with hashed and actual values of all unique sets of identifiers. However the level of re-identification needs to be controlled, depending on the type of user accessing the results. Generalization and suppression on identifiers from the identifier dictionary before re-introduction could achieve different levels of privacy preservation. In this paper we propose an approach to achieve data security & privacy throughout the complete data lifecycle: data generation/collection, transfer, storage, processing and sharing.***

Keywords: Big Data, Hash value, Data processing, Security, privacy, Attribute Based Encryption.

1. Introduction

When the number of old age people in industrialized countries they rapidly growing. If old age people they need any of healthcare services are to receive then the same amount quality of help as, the number of professional personnel delivering these services must be double. Oftenly prefers to live at home because they need of some confident and comfortable environment. Promote social interaction and Aging-in-Place becomes for optimize to extend the traditional healthcare services to residential home, using some sensor networks that supported by data analytics and to deliver assistive services. To provide assistive services through data analytic technologies, and usually sensor data has to be collected centrally and effectively and perform knowledge discovery algorithms. The popular solution for storage and processing of the large datasets is Hadoop and it is implemented in the Safer@Home project. Collected sensor data from each smart homes represent a personal and sensitive information. It oftenly disclose the complete living behavior of an individual. At the same time, it is not feasible to perform analytics on data that are to be transformed due the very nature of the solution where in it is very important to be able to identify the individual, to whom preventive care needs to be provide. Encrypted data has ideally analysis and it would be a perfect solution for preserving privacy. It is become very necessary to devise a scheme that can allow execution of data analytic/mining algorithms when monitoring the individuals preserving privacy. Authorized personnel can be provided with personal details of individual in the need of assistance. Finally, computation and storage overhead of the scheme has to be carefully evaluated.

2. Related Work

S. Ruj, M. Stojmenovic, and A. Nayak proposed a new privacy preserving authenticated access control scheme for securing data in clouds. In the proposed scheme, the cloud verifies the authenticity of the user without knowing the user's identity before storing information. Our scheme also has the added feature of access control in which only valid users are able to decrypt the stored information. The scheme prevents replay attacks and supports creation, modification, and reading data stored in the cloud. Moreover, our authentication and access control scheme is decentralized and robust, unlike other access control schemes designed for clouds which are centralized. The communication, computation, and storage overheads are comparable to centralized approaches.

H.K. Maji, M. Prabhakaran, and M. Rosulek introduced a new and versatile cryptographic primitive called Attribute-Based Signatures (ABS), in which a signature attests not to the identity of the individual who endorsed a message, but instead to a (possibly complex) claim regarding the attributes she possesses. ABS offers: – A strong unforgeability guarantees for the verifier, that the signature was produced by a single party whose attributes satisfy the claim being made; i.e., not by a collusion of individuals who pooled their attributes together. A strong privacy guarantee for the signer, that the signature reveals nothing about the identity or attributes of the signer beyond what is explicitly revealed by the claim being made. We formally define the security requirements of ABS as a cryptographic primitive, and then describe an efficient ABS construction based on groups with bilinear pairings. We prove that our construction is secure in the generic group model. S. Moncrieff et al proposed a solution to dynamically alter privacy levels in a smart house, based on environmental context using data masking techniques to decrease the intrusive nature of the technology, while maintaining the functionality.

3. Proposed System

We propose our privacy preserving authenticated access control scheme. According to our scheme a user can create a file and store it securely. This scheme consists of use of the two protocols ABE and ABS.

There are users called as a data collector and data receiver. Data collector receives a token t from the trustee, who is assumed to be honest. A trustee can be someone like the

federal government who manages social insurance numbers etc. On presenting her id (like health/social insurance number), the trustee gives her a token t. There are multiple KDCs (Key Distribution Centre), which can be scattered. For example, these can be servers in different parts of the world. A sender on presenting the token to one or more KDCs receives keys for encryption/decryption and signing. SKs are secret keys given for decryption, Kx are keys for signing. The message MSG is encrypted under the access policy X. The access policy decides who can access the data stored in the data receiver. The data collector decides on a claim policy Y, to prove her authenticity and signs the message under this claim. The ciphertext C with signature is c, and is sent to the data receiver. The data receiver verifies the signature and stores the ciphertext C. When a reader wants to read, the data receiver r sends C. If the user has attributes matching with access policy, it can decrypt and get back original message.

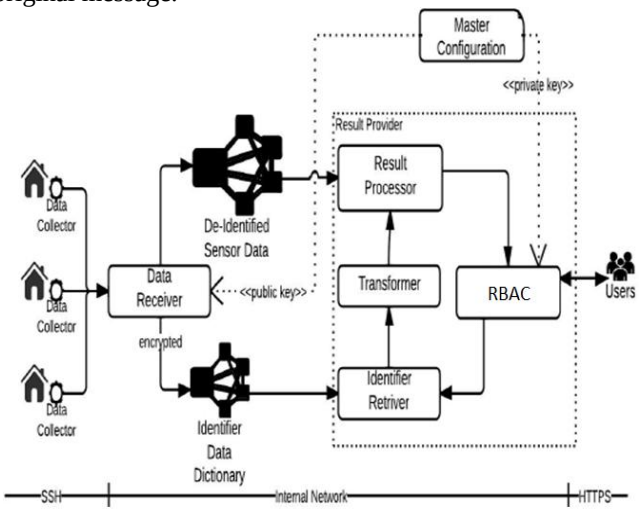

Figure 1: Architecture

Secure data collection architecture framework is given in Fig. 1. It consists of three modules and two storage units. Data collector first module. Data collector present at each smart home and transfers their sensor data to a data cluster at regular intervals of time. Data receiver is second module. Data receiver receives the collected data sent by the data collector and transforms them into two different datasets. The storage unit, de-identified sensor data which stores the actual data with primary/quasi-identifiers values are hashed. The identifier dictionary storage which contains only hashed and actual values for each unique set of primary/quasi-identifiers, and if they do not already exist. Result provider is third module. End users access to data processing results which controls by result provider module. It authorizes the end users and ensures that privacy of any shared results is preserved. Each of these modules is discussed below.

i . Data Collector

It is an application at each smart home. It is responsible for collecting sensor data and then transferring them to the data cluster at regular intervals of time. It can be configurable through a configuration file and controlling every aspect of its functionality. Among others the main aspects it is configures, that are connection to the sensor data sources, and the frequency at which it usually checks for new data, the address to which the data is to be send, the format in which the data is sent after establishing a connection of a protocol use. The data transfers from the data collectors should be automatic, secure, fast and confidential. SSL uses cryptographic automatic session encryption, authentication and integrity protection for transferred data. In contrast to other solutions like GridFTP, kFTP , VPNs and glogin. SSH is easy to install, use, configure and administer. C. SSH has a weakness of speed over WANs for bulk data transfer. However, due to the need of collecting data in real time, but the size of data per transfer remains small and the rate of transfer is frequent SSH is a default transfer protocol for data collector with further evaluation of extensions and additional patches to ensure a secure and high-speed transfer.

ii . Data Receiver

The data receiver module collects the data from the data collectors. Data receiver performs an algorithmic function for separation between the different attributes of the dataset, based on the existing schema definition of the file. Attributes which are classified based on empirical observations, regulations and linkage to public sources.Data processing requirements, a standard process for classification is yet to be established and it would require a separate research focus. The algorithmic transformation function outputs are stored separately to achieve isolation between sensitive and de-sensitized data. The attributes that are primary/quasi-identifiers are hashed using SHA techniques, before encrypting and storing them, and their actual values into the identifier dictionary storage, if they do not already exist. The non-identifiers along with the hashed primary/quasi-identifiers are in de-identified storage they stored. Through master configuration file with a pass phrase added the identifiers before hashing them and to protect against brute force attacks on identifiers with limited value ranges (eg. age, zip).

iii. Result Provider

The previous sub-sections the areas which are securely collecting, storing and processing sensitive data were addressed. In order to realize the benefits of such a system the results from data processing needs to be made which are available to appropriate users. Healthcare providers, social institutions, service providers and researchers may all contribute in improving lives of old age people. Doctors or nurses want to analyze patients health pattern currently or be notified of any patient difficulty to classify. The results provided to the doctors/nurses must be identifiable so that they may provide correct care to right patients. Further researchers or social institutions may want to understand the overall health or lifestyle patterns of elderly in a region. Any information provided to doctors or nurses, should guaranty that the privacy of the data owners. The access control module not only ensure that the right end-users are authenticated and but also verify that they are authorized to access data for the requested patients. It should make sure depending on the role of user, the result provided them are generalized or suppressed. The result provider module can be contains four groups. Access control is the first module, which authorizes, authenticates and for any data share it determines the privacy level. Identifier retriever is the second module. Identifier retriever module queries the identifier dictionary storage and to generate a list of personal/quasi-

identifiers, whose data the end-user requested and is access to authorized. Transformer is the third module using this list generalizes or suppresses the actual personal/quasi-identifier values and dataset is created with the hashed, actual and generalized or suppressed values. Result processor is fourth module. Based on the transformer modules output result processor it starts a data processing job on the de-identified storage and replaces the hashed personal/quasi-identifier values in the result set with generalized or suppressed values. The workflow for the complete module is represented in Fig. 2

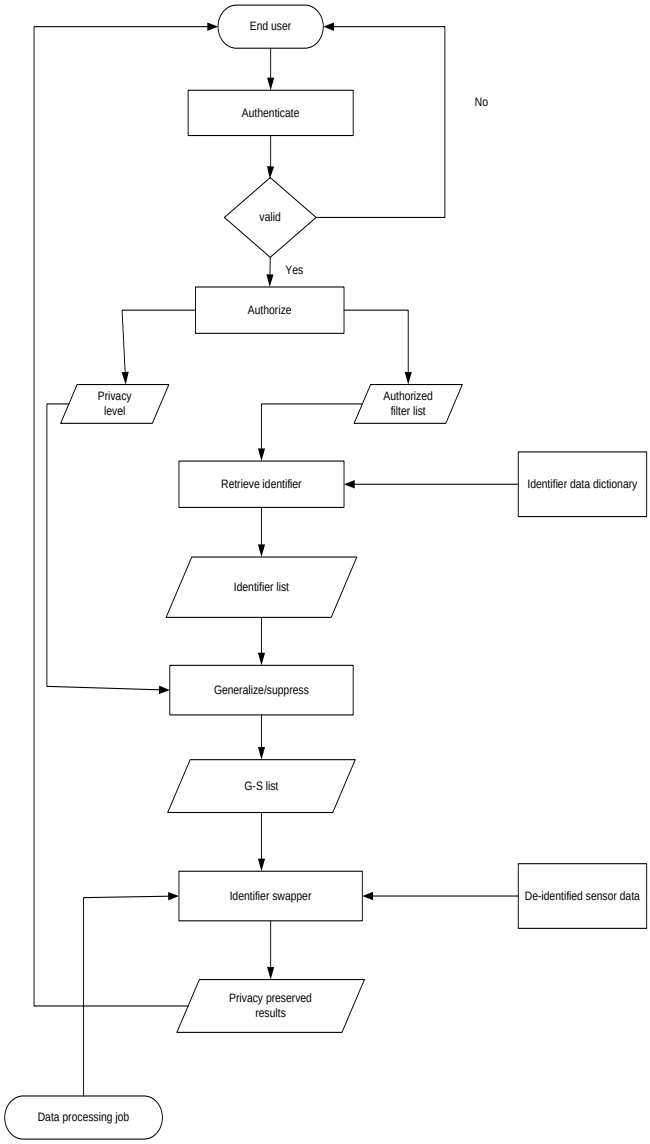

Figure 2: Result provider work flow

1.Access Control

It is providing access to the system that enforce access control requirements through some adequate mechanisms. And along with authentication, it authorize an end-user based on a set of rules and for the shared data it maintain a privacy level. Role based access control (RBAC) concepts which are provide an important high-level organizational constrains and rules. After a user is authenticated, the module based on a set of rules the list of hashed and actual primary identifiers generates whose data that the user requested and is authorized. RBAC also determines the preservation of privacy level shared results must enforce. If a user have the same authorization, then their privacy level could be different. For example a personal doctor having complete authorization for their patients without any requirement for hiding personally identifiable information. The privacy level could be none in such a case. A specialist doctor, the data of a patient is which he is referred and may have complete authorization but personally identifiable information could be protected though a higher privacy level. The same is true for other users such as nurses, researchers. The access rules must not only authorize but also determine the level of privacy based on the role of a user.

2.Identifier Retriever

It is only responsible for preparing a dataset on which generalization or suppression algorithms can perform. It queries the dictionary storage using the authorized personal identifier list as filters. The output which generates provides unique set of decrypted personal/quasi-identifiers with both hashed values and actual values.

3.Transformer

It is responsible for the privacy of shared data. The notion of k-anonymity can be specified for the level of privacy. A dataset can be transformed and satisfies k-anonymity if every combination of values in personally identifiable columns cannot be matched to fewer then k rows. Generalizing or suppressing values in personally identifiable columns achieves a k-anonymized dataset. Having already an existent data dictionary for all combination of personally identifiable information, it becomes a perfect source for performing k-anonymity operations. The generated dataset through an identifier retriever module and the privacy level for end user, is used to perform k-anonymity on all actual values of personal/quasi-identifiers. A list of hashed and k-anonymized values generates from transformer output. For privacy level as none, the k-anonymized values are same as the actual values and with higher privacy level the level of generalization or suppression for k-anonymized output also increase. Although there are several k-anonymization but algorithms only a few are suitable for use in practice. The proper k-anonymization approach and practicality would still need to be evaluated.

4.Result Processer

Result processor module is responsible for swapping the hashed values of results from a data processing job performed on the de-identified storage. For all hashed identifier values from the k-anonymized list the data processing job is executed. These data objects being as requested and authorized, the data processing job performs any analysis or mining for only these sets of personal/quasi-identifiers, thus isolating the operation from any data objects that are not authorized for access. The hashed identifiers from the results are replaced with their respective k-anonymized values, ensuring the privacy of any shared data is preserved.

4. Simulation Results

Following are the simulation results
Scenario 1:

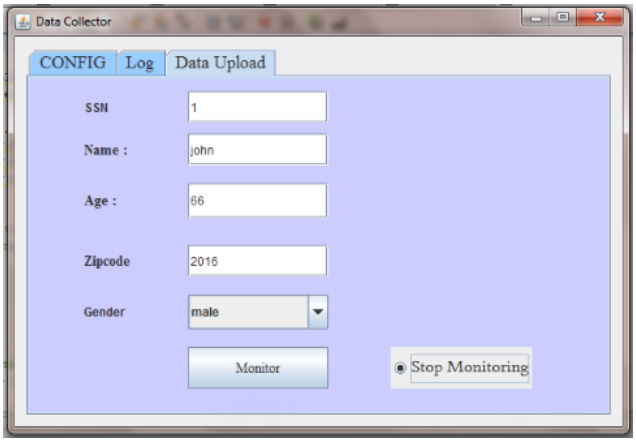

Figure 3: Data collector

Data collector can configure with listen port and after configuring it register into data receiver server then data is uploaded to data collector and monitor the data.

Scenario 2:

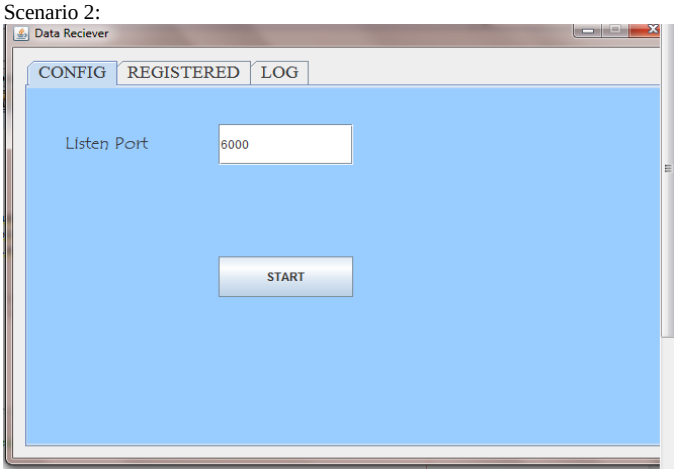

Figure 4: Data Reciever

Data receiver which receives the from data collector after registering the id to data receiver server and server it waits for messages.

Scenario 3:

Figure 5: Encrypted value

After user module running the data which is stored in database with hash values and original values the user getting the right information in web application then encrypted values shows the user has got correct information.

5. Conclusion and Future work

The problem is in the security and preserving privacy for analysis of sensor data from smart homes, without compromising on data utility. In order to solve these issues we propose ciphertext-policy ABE encryption algorithm. In CP-ABE, each user is associated with a set of attributes and data are encrypted with access structures on attributes. A user is able to decrypt a ciphertext if and only if his attributes satisfy the cipher-text access structure. In addition to this we propose k-anonymity algorithm to provide confidentiality. It has advantages such as it provides better access. In future scope we can implement this project using some other protocols and encryption technique for better access.

6. Acknowledgment

Sachinkumar Sugur, thanks to Mrs. Sahana.V, who is always encouraging and motivating me to do research activities. I am also very thankful my families and friends.

References

[1] T. Dalenius, “Finding a needle in a haystack – or identifying anonymous census record,” Journal of Official Statistics, vol.2, no.3, pp.329-336, 1986
[2] R. Sandhu, E. Coyne, et. al., “Role-Based Access Control Models,”IEEE Computer, vol.29, no.2, pp.38-47, Feb. 1996
[3] G. J. Ahn, “The RCL 2000 language for specifying role-based authorization constrains,” Ph.D. dissertation, George Mason University, Verfinia, 1999
[4] H. Gilbert, H. Handschuh “Security Analysis of SHA-256 and Sisters,” Cryptography, vol. 3006, pp.175-193, 2004
[5] M. Chan, E. Esteve, et. al., “A review of smart homes—Present state and future challenges,” Computer Methods and Prigramns in Biomedicine, vol.91. no.1, pp.55-81, Jul. 2008
[6] L. Sweeney, “Achieving k-anonymity privacy protection using generalization and suppression,” International Journal on Uncertainty, Fuzziness, and Knowledge-Base Systems, vol.10, no.5, pp.571-588, 2002
[7] K. Courtney, G. Demiris, et. al., “Needing smart home technologies: the perspectives of older adults in continuing care retirement communities,” Informatics in Primary Care, vol.16, pp.195-201, 2008
[8] R. Bayardo, R. Agrawal, “Data Privacy Through Optimal k- Anonymization,” 21th International Conference on Data Enginnering, pp.217-228, Apr. 2005
[9] E. Dishman, “Inventing Wellness Systems for Aging in Place,” Computer, vol.37, no.5, pp.34-41, May, 2004
[10] G. Abowd, A. Bobick, et. al., “The Aware Home: A living laboratory for technologies for successful aging,” American Association for Artificial Intelligence, 2002
[11] K. Haigh, L. Kiff, “The Independent LifeStyle AssistantTM (I.L.S.A.): AI Lessons Learned,” Innovative Applications of Artificial Intelligence, 2004
[12] G. Demiris, B. Hensel BK, et. al, “Senior residents’ perceived need of and preferences for smart home sensor

technologies," Int J Technol Assess Health Care, vol.24. no.1, pp.120-1024, 2008
[13] S. Moncrieff; S. Venkatesh, et. al., "Dynamic Privacy in a Smart House Environment," IEEE International Conference on Multimedia and Expo, pp.2034-2037, Jul. 2007
[14] S. Meyer, A. Rakotonirainy, "A survey of research on context-aware homes," Australian Computer Society, vol.21, pp.159-168, 2003
[15] S. Bagüés , A. Zeidler, et. al., "Sentry@Home - Leveraging the Smart Home for Privacy in Pervasive Computing," International Journal of Smart Home, vol.1, no.2, Jul. 2007
[16] M. Mozer, "The Neural Network House: An Environment that Adapts to its Inhabitants," American Association for Artificial Intelligence, 1998
[17] G. Drosatos, P. Efraimidis, "Privacy-preserving statistical analysis on ubiquitous health data," 8th International Conference on Trust, Privacy and Security in Digital Business, Springer-Verlag, pp.24-36, 2011
[18] R. Rivest, L. Adleman, and M. Dertouzos, "On data banks and privacy homomorphisms," Foundations of Secure Computation, Academic Press, pp.169–177, 1978.
[19] C. Fontaine, F. Galand, "A Survey of Homomorphic Encryption for Nonspecialists," Journal on Information Security, vol.2004, no.15, 2007.

Author Profile

Sachinkumar received the B.E degree in Computer Science and Engineering from R.E.C.Bhalki in 2010 and M.tech degree in Computer Science and Engineering Pursing in CMR Institute of technology Bangalore, respectively.

Mrs.Sahana.V received the B.E degree in Computer Science and Engineering from MCE Hassan, M.tech Degree in Computer Science and Engineering from PESIT Bangalore. And Working as a Assistant professor of CSE Department in CMR Institute of Technology, Bangalore.