Privacy Preserving Data Analytics for Smart Homes

Antorweep Chakravorty, Tomasz Wlodarczyk, Chunming Rong
Department of Computer & Electrical Engineering
University of Stavanger
Stavanger, Norway
{antorweep.chakravorty, tomasz.w.wlodarczyk, chunming.rong}@uis.no

***Abstract*— A framework for maintaining security & preserving privacy for analysis of sensor data from smart homes, without compromising on data utility is presented. Storing the personally identifiable data as hashed values withholds identifiable information from any computing nodes. 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. Through a separate encrypted identifier dictionary with hashed and actual values of all unique sets of identifiers, we suggest re-identification of any data processing results. 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 through out the complete data lifecycle: data generation/collection, transfer, storage, processing and sharing.**

Keywords—privacy preserving; data security; smart homes; big data.

I. INTRODUCTION

The number of elderly citizens in industrialized countries is growing rapidly and according to estimations by UN is expected to double by 2050. If elderly people in need of healthcare services are to receive the same amount and quality of help as today, the number of professional personnel delivering these services must double. Noticeably, one often prefers to live at home due to the confident and comfortable environment. To optimize resources, prolong independent living and promote social interaction, Aging-in-Place (AIP) becomes a metaphor to extend traditional healthcare services to residential home, using sensor networks supported by data analytics to deliver assistive services. Many researchers work on related technologies [1], [30], [31], [32], [37]. One such example being, the Safer@Home [2] project at the University of Stavanger.

In order to provide assistive services through data analytic technologies, sensor data has to be usually collected centrally to effectively perform knowledge discovery algorithms. One of more popular solutions for storage and processing of large datasets is Hadoop [3] and it is also implemented in the Safer@Home project. However, the collected sensor data from smart homes represent personal and sensitive information and can often disclose the complete living behavior of an individual. At the same time, it is infeasible to perform analytics on data that are transformed due the very nature of the solution wherein it is important to be able to identify individual, to whom preventive care needs to be furnished. Ideally analysis on encrypted data would be a perfect solution for preserving privacy however, it isn't an easy or a cost-free task. Homomorphic encryption [39], tries to address data analytics on encrypted data. C. Fontaine et. al. [40] evaluates the advancements in homomorphic encryption but, current research in encrypted data analytics remain inefficient to be used in practical applications. It becomes necessary to devise a scheme that would allow execution of data analytic/mining algorithms while preserving privacy of monitored individuals. The scheme has to be reversible so that authorized personnel can be provided with personal details of individual in need of assistance. Finally, computation and storage overhead of the scheme has to be carefully evaluated.

Related Work: The main objective of privacy preservation is ensuring that private data remains protected, while processing or releasing sensitive information. Privacy concerns about data from smart homes have been raised in various literatures [26], [28], [33]. However there has been little work, on design of technical solutions protecting privacy through out the complete data lifecycle for smart home analytics. S. Moncrieff et. al. [34] proposes 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. S. Meyer, et. al. [35] demonstrates selected information discloser through a privacy manager module for a context-aware system interacting with a user. S. Bagüés et. al. [36] proposes a framework to control the dissemination of data within the context-aware service interaction chain, based on a set of user defined privacy policies. G. Drosatos et. al. [38] introduces a privacy preserving cryptography approach for distributed statistical analysis of data from wearable sensors. All of the discussed solutions however, are very specific and address privacy concerns required for their solutions. None of them can be easily incorporated or extended to existing or new smart home designs having different data processing needs.

Our Contribution: This paper presents an approach independent of underlying data analytic processes and that can be easily adapted to existing or new smart homes solutions. We present a holistic framework to maintain data utility, ensure security and preserve privacy at different stages of data lifecycle (collection, storage, processing & sharing).

Organization: This paper is structured as follows. Section II discusses the data security and privacy issues. In section III we present a solution to these issues and conclude with section IV.

DOI 10.1109/SPW.2013.22

II. DATA SECURITY AND PRIVACY ISSUES

The concept of privacy varies from countries, cultures and jurisdiction. However in general, privacy is associated with collection, storage, use, processing, sharing or destruction of personally identifiable data. Chen et al. [4] surveys data security & privacy issues around the complete data lifecycle for cloud computing. Based on their framework, we derive four areas to ensure security & privacy for a smart home analytic solution. The areas of data ownership, transfer, storage & processing and access are discussed bellow.

A. Data Ownership

Data generated at smart homes are sensitive, and ownership issues are not always clear. Although a community center, healthcare provider or service providers could own the sensor and network devices, yet the data pertain to the residents of the homes. They should know what kind of data are collected, stored and shared. They should be able to stop the collection as well as ask for destruction of any stored records.

B. Data Transfer

Transmission of the sensor data through unsecure networks should be protected. Confidentiality and integrity should be ensured for any data transfer. Confidentiality is securing sensitive data against a malicious user and integrity is preserving the truthfulness of the data. Cryptography or VPN techniques [5], [6] are some of the commonly used approaches for securely transferring data.

C. Data Storage & Processing

Data stored with personally identifiable information (or identifiers) in an external cluster is a serious threat to data privacy. Personal and quasi identifiers [22] describe personally identifiable information. These attributes can directly or in-directly reveal personal information. Steps to protect privacy are to replace any personally identifiable information with randomized placeholders, introduce noise or swapping values while ensuring that statistical properties and data consistency are maintained [7], [8]. Another alternative approach is using generalization and suppression methods [9], [10], [11]. The processing of smart home data should be independent of sensitive information. Storing the data used for analysis/mining as mentioned above can achieve this. However, the use of transformation challenge is to find the right trade-off between amount of privacy and information loss [9], [10], [11], [12].

D. Data Access

Access to the system should be ensured through proper authentication and authorization. The system should be configurable to assign rights to execute analysis/mining jobs to appropriate users and access the generated results. Among many methods the *role base access control* (RBAC) has been widely accepted because of its simplicity, flexibility in capturing dynamic requirements and support for the principle of least privilege and efficient privilege management [13], [14], [15].

III. THE PROPOSED SOLUTION

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

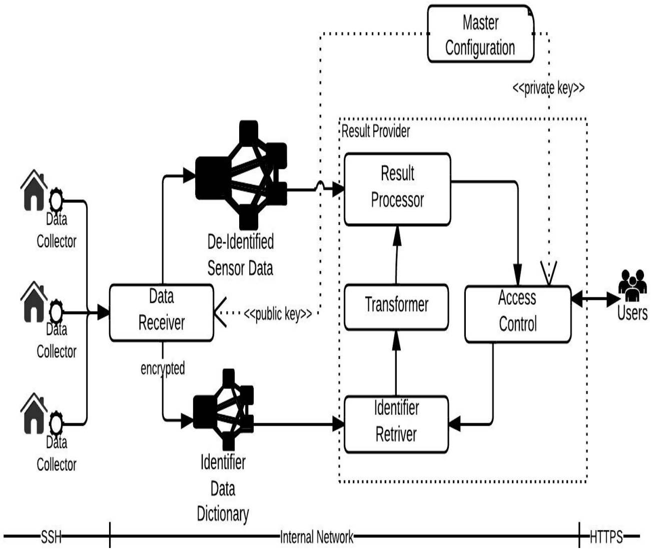


Fig. 1. The Proposed Architecture

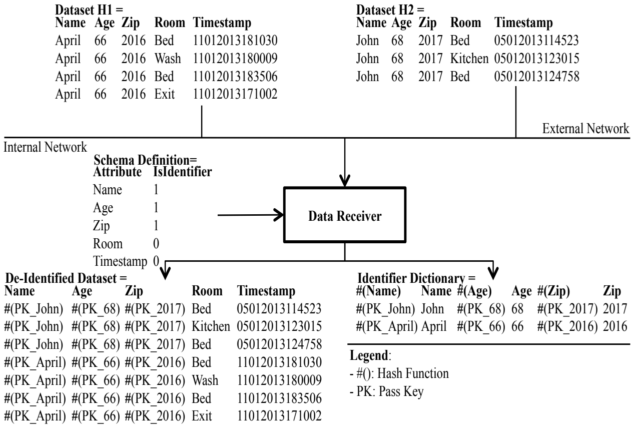
A. Data Collector

The data collector is an application at each smart home. It is responsible for collecting sensor data and transferring them to the data cluster at regular intervals. It is configurable through a configuration file controlling every aspect of its functionality. Among others the main aspects it configures, are connection to the sensor data sources, the frequency at which it checks for new data, the address to which the data is to be send, the protocol using which it establishes a connection and the format in which the data is sent.

The data transfers from the data collectors should be fast, automatic, secure and confidential. SSL uses cryptographic authentication, automatic session encryption, and integrity protection for transferred data [16]. In contrast to other solutions like kFTP [17], GridFTP [18], glogin [19] and VPNs, SSH is easy to install, use, configure and administer. C. Rapier et. al. [20] argues about SSH's weakness of speed over wide area networks for bulk data transfer. However, due to the need of collecting data in real time, the rate of transfer is frequent but the size of data per transfer remains small. The data collector would use SSH as its default transfer protocol, with further evaluation of additional patches/extensions to ensure a secure and high-speed transfer.

B. Data Receiver

The data receiver module accepts inputs from the data collectors. It performs an algorithmic function to make separation between the different attributes of the dataset, based on an existing schema definition file. Attributes are classified based on regulations, empirical observations and linkage to public sources. Being specific to data processing requirements, a standard process for classification is yet to be established and would require a separate research focus. The outputs of the algorithmic transformation function are stored separately to achieve isolation between sensitive and de-sensitized data. Those attributes that are primary/quasi- identifiers are hashed using SHA [21], [25] techniques, before encrypting and storing them, as well as 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 stored into the de-identified storage. The identifiers are concatenated with a pass phrase from the *master configuration* file before hashing them, to protect against brute force attacks on identifiers with limited value ranges (eg. age, zip). Fig. 2. illustrates two datasets {H1, H2} send by the data collectors. The data receiver based on a *master schema definition* uses a hash function *#(<<data-item>>)* to transform the attributes for which "isIdentifier" value is set to 1. If a set of primary/quasi-identifier is not present at the identifier dictionary it stores their hashed values along with the actual values. Thus the identifier dictionary contains only unique sets of data. The de-identified dataset contains all incoming tuples with the primary/quasi-identifiers replaced with their hashed values.



Dataset H1 =

Name	Age	Zip	Room	Timestamp
April	66	2016	Bed	11012013181030
April	66	2016	Wash	11012013180009
April	66	2016	Bed	11012013183506
April	66	2016	Exit	11012013171002

Dataset H2 =

Name	Age	Zip	Room	Timestamp
John	68	2017	Bed	05012013114523
John	68	2017	Kitchen	05012013123015
John	68	2017	Bed	05012013124758

Schema Definition=

Attribute	IsIdentifier
Name	1
Age	1
Zip	1
Room	0
Timestamp	0

De-Identified Dataset =

Name	Age	Zip	Room	Timestamp
#(PK_John)	#(PK_68)	#(PK_2017)	Bed	05012013114523
#(PK_John)	#(PK_68)	#(PK_2017)	Kitchen	05012013123015
#(PK_John)	#(PK_68)	#(PK_2017)	Bed	05012013124758
#(PK_April)	#(PK_66)	#(PK_2016)	Bed	11012013181030
#(PK_April)	#(PK_66)	#(PK_2016)	Wash	11012013180009
#(PK_April)	#(PK_66)	#(PK_2016)	Bed	11012013183506
#(PK_April)	#(PK_66)	#(PK_2016)	Exit	11012013171002

Identifier Dictionary =

#(Name)	Name	#(Age)	Age	#(Zip)	Zip
#(PK_John)	John	#(PK_68)	68	#(PK_2017)	2017
#(PK_April)	April	#(PK_66)	66	#(PK_2016)	2016

Legend:
- #(): Hash Function
- PK: Pass Key

Fig. 2. Data receiver example dataset

The aim of this transformation is to achieve isolation from sensitive information on the data used for any processing. It enables configuration of all attributes that are deemed sensitive and have potential to reveal privacy. The de-identified data although anonymized, still maintains their statistical properties. An attribute hashed would always give the same output. The non-identifiers such as timestamp and value elements remain unhindered. This ensures that the linkage between identifiers and non-identifiers remains intact, although the identifiers are hashed. Proofs for closeness of hashed values are beyond the scope of this paper and won't be addressed here. With maintenance of the separate identifier dictionary storage for only the unique sets of primary/quasi-identifiers, the amount of data stored here is much less than that in the de-identified storage. The de-identified storage contains all data collected from each home, with sizes moving upto tera-bytes of information. Below we demonstrate the amount of information growth on both these data stores:

Let
S_i = Size of a *single* primary/quasi- identifier attribute set
S_n = Size of a *single* non-identifier attribute set
N_l = Total *count* of data transfers from a home
N_h = Number of homes
N_r = Total number of records from all homes per transfer

Data transfer rate from each home: $\overline{N_l} = \dfrac{\sum N_l}{N_h}$

Sizes of the de-identified (D_1) and identifier dictionary (D_2) storage are:

$$D_1 = N_r(S_i + S_n)\sum N_l \qquad D_2 = \frac{2S_i\sum N_l}{\overline{N_l}}$$

From the equations above the rate of growth, in D_1 can be represented as $(\dfrac{S_i + S_n}{2S_i})\overline{N_l}N_r$ times D_2.

Any form of generalization/suppression algorithms performed on such amounts of data as in D_1, would greatly take toll on the performance. The generalization/suppression methods aim at transforming personally identifiable information, in such a way that they can be shared while preserving their privacy. The identifier dictionary store having all unique sets of identifiers, not only acts as reference for re-introduction, but also serves as a great source for any generalization/suppression. The data processing results are replaced with the generalized/suppressed values, thus forgoing any form of information loss for analysis/mining algorithms as well as preserving privacy.

C. Result Provider

Through the earlier sub-sections the areas of securely collecting, storing and processing sensitive data were addressed. However, in order to realize the benefits of such a system the results from data processing needs to be made available to appropriate users. Healthcare providers, social institutions, service providers and researchers may all contribute in different ways at improving lives of elderly. Doctors/nurses may want to analyze the current health patterns or be notified of any anomalies. The results provided to them 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 them, should guaranty the privacy of the data owners. The access control module must not only ensure that the right end-users are authenticated, but also verify that they are authorized to access data for the requested patient(s). It should make sure depending on the role of user, the result provided are generalized or suppressed.

The activities for the result provider module can be classified into four groups. The first is the *access control* module, which authenticates, authorizes and determines the level of privacy for any data share. The second is the *identifier retriever* module. It queries the identifier dictionary storage to generate a list of personal/quasi- identifiers (both actual and hashed values), whose data the end-user requested and is authorized to access. The *transformer* module using this list generalizes/suppresses the actual personal/quasi- identifier values and creates a dataset with the hashed, actual and generalized/suppressed values. The result processor module starts a job on the de-identified storage and replaces the hashed personal/quasi- identifier values in the result set with respective generalized/suppressed values based on the transformer module's output. The workflow for the complete module is represented in Fig. 3.

Access Control: This module aims at providing access to the system through adequate mechanisms that enforce access control requirements. Along with authentication, it would authorize an end-user based on a set of rules and also maintain a privacy level for the shared data. Role based access control (RBAC) concepts provide an important means for laying out high-level organizational rules and constrains [23], [24]. After a user is authenticated, the module based on a set of rules generates a list of hashed and actual primary identifiers whose data the user requested and is authorized. It also determines the level of privacy preservation shared results must enforce. Although a user may have the same authorization, their level of privacy could be different. An example would be a personal doctor having complete authorization for their patients without any requirement for hiding personally identifiable information. In such a case the privacy level could be none. However a specialist doctor, to whom the data of a patient is referred, 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.

Identifier Retriever: This module is responsible of preparing a dataset on which generalization/suppression algorithms can be performed. It queries the personal dictionary storage using the authorized personal identifier list as filters. The generated output provides a concise & unique set of decrypted personal/quasi- identifiers with both hashed and actual values.

Transformer: The transformer module is responsible for guaranteeing the privacy of shared data. The level of privacy can be specified by the notion of k-anonymity [9], [11]. A transformed dataset 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 dataset generated through the identifier retriever module and the level of privacy for end user, is used to perform k-anonymity on all actual values of personal/quasi- identifiers. The output of this module generates a list of hashed and k-anonymized values. For level of privacy as none, the k-anonymized values are same as the actual values and with higher level of privacy the level of generalization/suppression for k-anonymized output also increase. Although there are several k-anonymization algorithms in the literatures [7], [12], [11], [27], [10] only a few are suitable for use in practice. R. Bayardo, R. Agarawal [29] evaluates these approaches and identifies Datafly [27], μ–argus [12], Iyengar-GA [7] and their solution as practical k-anonymization algorithms. K-anonymization itself is a well-accepted solution and suits the secure architecture. However, the proper k-anonymization approach and practicality would still need to be evaluated.

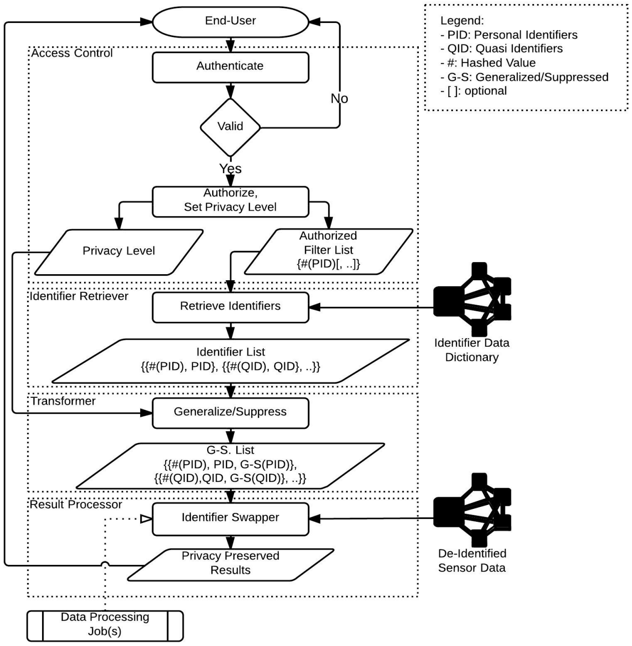


Fig. 3. Result provider workflow

Result Processer: This module is responsible for swapping the hashed values of results from a data processing job performed on the de-identified storage. The data processing job is executed for all hashed identifier values from the k-anonymized list. These data objects being as requested and authorized, the data processing job performs any analysis/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.

IV. Conclusion

In this paper we demonstrated a solution for reliably concealing privacy and ensuring security for analytics of smart home sensor data. The presented approach maintained the data utility by not transforming the stored data. Rather based on cryptographic techniques, we replace the personal/quasi-identifiers of collected sensor data with hashed values before storing them into a de-identified storage. A separate identifier dictionary storage, with hashed and actual identifier values was also maintained as a point of reference for re-introduction of identifiers. We proposed using heuristic-based k-anonymization algorithms based on the end-users privacy level, requirements and authorization on the identifier dictionary storage. The hashed identifiers from outputs of any data processing job on the de-identified store was replaced with their respective k-anonymized value, thus preserving privacy of any presented/shared results.

In future we would present a practical implementation of the framework. The RBAC policies for authorizing and setting privacy levels would be specified and formally validated. Different practical k-anonymization algorithms would be gauged to verify their applicability to our approach. The performance, data utility, uncertainty level and endurance to different data processing techniques would also be measured.

References

[1] D. Cook, M. Youngblood, et. al., “MavHome: An Agent-Based Smart Home,” *Pervasive Computing and Communication*, pp.521-524, Mar. 2003

[2] CIPSI Lab, Department of Compuer and Electrical Engineering, UiS, “Project description - Safer@Home”

[3] Yahoo!, “Hadoop,” *http://hadoop.apache.org*, 2008

[4] D. Chen, H. Zhao, “Data Security and Privacy Protection Issues in Cloud Computing,” *International Conference on Computer Science and Electronics Engineering (ICCSEE)*, vol.1, pp.647-651, Mar. 2012

[5] A.D. Rubin, D. E. Geer, "A survey of Web security," *Computer*, vol.31, no.9, pp.34-41, Sept. 1998

[6] CohesiveFT, “VPNCubed,” http://www.cohesiveft.com/vpncubed/, 2008

[7] V. S. Iyengar, “Transforming data to satisfy privacy constraints,” *Eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp.279-288, 2002

[8] J. Kim and W. Winkler, "Masking microdata files," *Survey Research Methods ASA Proceedings*, pp.114–119, 1995

[9] P. Samarati, L. Sweeney, “Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression.,” *Stanford Research Institute International*, Mar. 1998

[10] L. Sweeney, “Datafly: A system for providing anonymity in medical data,” *11th International Conference on Database Security*, pp.356–381, 1998

[11] P. Samarati, “Protecting respondents’ identities in microdata release,” *IEEE Transactions on Knowledge Engineering*, vol.13, no.6, pp.1010–1027, Nov. 2001

[12] A. Hundepool, L. Willenborg, “μ- and τ- argus: Software for statistical disclosure control,” *3rd Internation Seminar on Statistical Confidentiality*, 1996

[13] R.W. Baldwin, “Naming and Grouping Privileges to Simplify Security Management in Large Databases,” *IEEE Symposium on Computer Security and Privacy*, 1990

[14] K.R. Poland, M.J. Nash, “Some Conundrums Concerning Separation of Duty,” *IEEE Symposium on Computer Security and Privacy*, 1990

[15] J. Joshi, et al., “Access Control Language for Multi-domain Environments,” *IEEE Internet Computing*, vol.8, no.6, pp.40–50, 2004

[16] T. Yloenen, “SSH - Secure Login Connections over the Internet,” *6th USENIX UNIX Security Symposium*, Jul. 1996

[17] J. Kohl, C. Neuman, “The kerberos network authentication service (V5). Request for Comments (Proposed Standard) RFC 1510,” I*nternet Engineering Task Force*

[18] W. Allcook, J. Bester, et al., "*Secure, Efficient Data Transport and Replica Management for High-Performance Data-Intensive Computing*," *Proceedings of the IEEE Mass Storage Conference*, pp.13-28 Apr. 2001

[19] H. Rosmanith, D. Kranzlmuller, “glogin - A Multifunctional, Interactive Tunnel into the Grid,” *grid*, pp. 266- 272, 5th IEEE/ACM International Workshop on Grid Computing (GRID'04), 2004

[20] C. Rapier, B. Bennett, “High speed bulk data transfer using the SSH protocol,” *15th ACM Mardi Gras conference*, 2008

[21] F. Mendel, N. Pramstaller, et. al, “Analysis of step-reduced SHA-256,” *13th International Conference on Fast Software Encryption*, pp.126-143, 2006

[22] 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

[23] R. Sandhu, E. Coyne, et. al., “Role-Based Access Control Models,” *IEEE Computer,* vol.29, no.2, pp.38-47, Feb. 1996

[24] G. J. Ahn, “The RCL 2000 language for specifying role-based authorization constrains,” *Ph.D. dissertation, George Mason University, Verfinia,* 1999

[25] H. Gilbert, H. Handschuh “Security Analysis of SHA-256 and Sisters,” *Cryptography*, vol. 3006, pp.175-193, 2004

[26] 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

[27] 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

[28] 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

[29] R. Bayardo, R. Agrawal, “Data Privacy Through Optimal k-Anonymization,” *21th International Conference on Data Enginnering*, pp.217-228, Apr. 2005

[30] E. Dishman, “Inventing Wellness Systems for Aging in Place,” *Computer*, vol.37, no.5, pp.34-41, May, 2004

[31] G. Abowd, A. Bobick, et. al., “The Aware Home: A living laboratory for technologies for successful aging,” *American Association for Artificial Intelligence*, 2002

[32] K. Haigh, L. Kiff, “The Independent LifeStyle AssistantTM (I.L.S.A.): AI Lessons Learned,” *Innovative Applications of Artificial Intelligence*, 2004

[33] 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

[34] 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

[35] S. Meyer, A. Rakotonirainy, “A survey of research on context-aware homes,” *Australian Computer Society,* vol.21, pp.159-168, 2003

[36] 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

[37] M. Mozer, “The Neural Network House: An Environment that Adapts to its Inhabitants,” *American Association for Artificial Intelligence*, 1998

[38] 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

[39] R. Rivest, L. Adleman, and M. Dertouzos, “On data banks and privacy homomorphisms,” *Foundations of Secure Computation*, *Academic Press*, pp.169–177, 1978.

[40] C. Fontaine, F. Galand, “A Survey of Homomorphic Encryption for Nonspecialists,” *Journal on Information Security*, vol.2004, no.15, 2007.