Computational Intelligence for User and Data Classification in Hospital Software Development

Masoud Mohammadian[1], Dimitrios Hatzinakos[2], Petros Spachos[2]

[1]University of Canberra
Faculty of Information Sciences and Engineering
Canberra, ACT, Australia
Masoud.Mohammadian@canberra.edu.au
[2]University of Toronto
Department of Electrical & Computer Engineering
{dimitris, Petros}@comm.utoronto.ca

Abstract. Lives are saved by utilization and application of the latest technologies in hospitals around the world to improve patient treatments and well being. Secure, accurate, near real time data acquisition and analysis of patient data and the ability to update such data will reduce cost and improve the quality of patient's care. This paper considers a wireless framework based on radio frequency identification (RFID) that uses wireless networks for fast data acquisition and transmission. This paper discusses the development of an intelligent multi-agent system in a framework in which RFID can be used for patient data collection. An approach to make the data communications more secure in a hospital environment is proposed. A new method for data classification and access authorization is also developed which will assist in preserving privacy and security of data.

Keywords: Radio Frequency Identification (RFID), Data Security and Privacy, Patient Profiling, Intelligent Agents, Health Care, Paperless Hospital

1 Introduction

Application of innovative architectures for secure access, retrieval, and update of data in healthcare systems continues to be needed for cost reduction and quality of services in hospitals. To this end the use of Radio Frequency Identification (RFID) has been shown to be a viable and promising technology in the health care industry [1, 5, 7, 8]. RFID technology has the capability to penetrate and add value to many areas of health care. It can lower the cost of some services as well as improve services to individuals and health care providers.

The real value of RFID is achieved in conjunction with the use of intelligent software systems such as intelligent multi-agent systems. The integration of these two technologies can benefit and assist health care services.

This research study considers the use of RFIDs and its potential in hospitals. RFIDs can be more effectively used to collect data at the source thereby providing the

data for monitoring patients' well being in order to provide a higher level of patient health care. There are several areas where using RFID in the role of data collection can have significant positive benefits for hospitals. These include care tracking to provide the right care to the right patient at the right time; quality of care provision; cost of care reduction by providing effective use of available resources to reduce the cost of the resources; and improved service of care by providing timely information to enable a more informed decision about an individual's need for care [7].

RFID tags and readers are commonly associated with inventory and tracking goods in places such as manufacturing and warehousing, but recently hospitals have been starting to apply RFID for new purposes [6].

RFID data can be read through the human body, clothing, and non-metallic materials and have the benefit to work wirelessly. This makes RFID an appropriate technology to fit into the health care environment.

Both research and practical application of the use of RFID technology in hospitals continues to be of importance. For hospitals this has meant the potential of managing inventories in a more efficient manner. Inventories in hospitals take on a variety of different roles than those found in manufacturing.

The nature of the inventory and assets in a hospital can include various types of equipment (that is often very expensive, comes in many sizes, and uses), drugs (that come in a variety of sizes, shapes, color, and governing regulations), beds, chairs, as well as patients and staff.

The management of patients and their condition is paramount in a hospital. RFID technology can assist in asset and personnel tracking, patient care, and billing where unnecessary expenses can be cut, the average length of stay of a patient can be reduced, where more patient lives can be saved due to timely treatments, and patient records being actively continuously updated to provide better patient care [6].

There is a need for more research into applications and innovative architectures for secure access, retrieval and update of data in healthcare systems [1, 5, 7, 8]. This research study considers a framework using RFID and Intelligent Software Agents for managing patients' health care data in a hospital environment. A fuzzy data classification system is also developed to improve the application of regulatory data requirements for security and privacy of data exchange.

Next section describes the patient to doctor profiling and the intelligent software agents developed [7]. Section three discusses the important issue of maintaining patients' data security and integrity and explains the use of RFID. Section four provides an approach to make the data communications more secure in a hospital environment. Last section provides conclusion and further directions for this research study. Part of the research work presented in this paper is already published in journal and conferences in the past few years. The reference to these publications is provided in this paper when appropriate.

2 Patient Profiling

Profiling is combined with personalization, and user modeling. The use of profile in hospitals and healthcare so far has been limited. Patient profiling is useful in a variety

of situations such as providing a personalized service based on the patient and not on the symptoms or the illness to a particular patient, as well as, assisting in identifying the medical facilities in trying to prevent the need for the patient to return to the hospital any time sooner than necessary. Patient profiling can also assist in matching a doctor's specialization to the right patient. A patient profile can assist in providing information about the patient on continuous basis for the doctors so that a tailored and appropriate care can be provided to the patient.

A patient or doctor profile is a collection of information that can be used in a decision analysis situation between the doctor, domain environment, and patient. A static profile is kept in pre-fixed data fields where the period between data field updates is long such as months or years. A dynamic profile is constantly updated as per evaluation of the situation in which the situation occurs. The updates may be performed manually or automated.

The automated user profile building is especially important in real time decision-making systems. The profiling of patient doctor model [7] is based on the patient / doctor information is used in this research study. In the patient / doctor profiling the intelligent agent software [2] will make distinctions in attribute values of the profiles and match the profiles with highest value. It should be noted that the intelligent agents create the patient and doctor profiles based on data obtained from the doctors and patient as described in [7].

The proposed intelligent agent architecture allows user profiling and matching in such a time-intensive critical application. The architecture of the agent profiling systems using RFID and Fuzzy Logic [9] is given in [7]. Due to the important role of intelligent agents in this system [7], it is recognized that there is need for a framework to coordinate intelligent agents so that they can perform their tasks efficiently.

3 User and Data Classification for Operation/Transaction Control

A hospital can be impacted by the corruption, unauthorized access, or theft of its data. A data security breach can impact organization's operations as well as causing large financial, legal implications. It can impact the personal privacy and public confidence in such an organization. With patient data in hospital human lives could be at risk by unauthorized access, corruption and modification of such data. An intelligent agent system is developed in this research study to check the access control from the users to the data stored in the database of a hospital. The intelligent agent system objective is to prevent the unauthorized observation of classified hospital data.

Discretionary Access Controls provide users with permits to access (allow) or disallow other users access to data items stored in a hospital. Such users/group of users with certain access permissions is allowed to access data items. It is possible that authorized users can then pass such data items to other users that may not have permission access to such data [4]. One way to restrict access to data items can be based on an intelligent agent system [2] that is capable to identity user/s data access level to which data can be provided. In this case each data item contains metadata information about its security and privacy. An intelligent agent can then control data access of users by checking metadata information attached to data. In such a case a

data item is provided to users after an intelligent agent checks its access permission. If an authorized user obtains a data item and passes it to an unauthorized user, then the unauthorized user will not be able to access that data item [4]. The data permission is checked for each data item as soon as a user wishes to access a data item. Therefore each data item permits access privileges to their users based on the metadata stored with each data item. Hence access to a data item is left to the discretion of an intelligent agent and the metadata of that given data item. The intelligent agent then can provide access to authorized users based on the metadata values of a data item under their control without the intercession of other authorities such as a system administrator.

Using an intelligent agent provides a higher level of security to hospital data. Intelligent agent access control based on metadata stored with data items can provide another level of security to the already existing role-based access control (RBAC). Data access controls in organizations are determined by user roles [4]. In such a case it is assumed that user/s do not pass data to other users and that they are aware and follow the organizations security, privacy policies and government's security and privacy laws. In hospital the data security and privacy associated with the diagnosis of ailments, treatment of disease, and the administering of medicine are of crucial importance [4].

Our proposed intelligent agent system can restrict access to a data item based on the sensitivity measures (stored with each data item as represented by its metadata) of the information contained with the data and the formal authorization (i.e. clearance) of user/s to access information of such sensitivity.

In such an environment data access policies provides the capability to authorize who can read what data and an unauthorized flow of information from a user with high level access to a user with a low level access is not possible [4]. It is also possible to provide more constraints on who can read or both read and write/update data by adding a metadata to support such constraint on a data item. Therefore each metadata attached to each role provides certain access and privileges to certain data items. Given the new proposed method of access control based on metadata stored the following benefits are obtained:

The security, privacy and confidentiality of data can be enforced further. For example, in a hospital environment a doctor could be provided with read/write access to a prescription data whereas the pharmacist will have the read only access of prescriptions data. Only authorized users can modify data items.

There is a need to provide the minimum disruption when implementing this security and privacy control to an organization. Using intelligent agents technologies and metadata access control information for each data will minimize resources impact without needing to re-design databases in an organization such as a hospital. We can add extra information to each data item by adding metadata information to the attributes of each entity in relational-data bases and domains in classes in object-oriented databases. Consider a simple relational database as shown below:

Patient(PatientID, Name, Address, TelNo, InsuranceID)
Insurance(InsuranceID, Type, InsuranceProviderID)
InsuranceProvider(InsuranceProviderID, Name, Address, TelNo, FaxNo)
Doctor(DoctorID, Name, OfficeNo, TelNo, PagerNo)
PatientDoctor(PatientDoctorID, PatientID, DoctorID, VisitDate, Notes)

The meta-data information could be the value or degree of user roles and related policies for privacy and security for that data item. Metadata values can then be used for adaptation and implementation of access/operation performance identification with each data item in the above database. The meta-data values can be obtained from the knowledge workers of the organization based on organization policies, procedure and business rules as well as government requirements for data privacy and security. Table 1 shows the metadata values for table Patient attributes.

Table 1. Metadata values for table Patient attributes

Patient Attributes	**Meta-data Value of data security access control access based on organization policy for patient data**
PatientID	70
Name	50
Address	29
TelNo	15

Now assume that the following domain metadata linguistic variables for the users (Docror, Nurse, Pharmesists etc.) of data in a given hospital as: TP = "top access user", MD = "medium access user", LO = "low access user" and ZE ="no access to data". For example Table 2 shows the metadata value related to security data access control of several kind of users based on organization's security access policy.

Table 2. Metadata values for different user

userID	**Meta-data value based on organization policy for data access**
DoctorID	52
NurseID	29
PharmacistID	20

The values are in the range of 0 to 70, where seventy indicates the metadata for a user of the hospital data that has top (full) access and zero indicates the metadata for a user that has no access to the data in the hospital. Note that other values are also possible. For simplicity assume that the linguistic terms describing the meta-data for the attributes of entities in the above database have the values: TP = [35,..,70], MD = [25,..,37], LO = [15,..27], ZE = [17,..,0]. Based on each userID metadata value for each user attributes the membership of that attribute to each linguistic variable can be calculated. In this case study Triangular fuzzy set was used to represent the data access classifications. Now assume that metadata value based on organization policy for users are as shown in Table 2. Based on the metadata value for each user the membership of each user to access and perform operation on data item can be calculated. The degree of membership value of the attribute userID based on metadata from Table 3 can then be calculated as follows:

Table 3. Fuzzy membership of metadata value of users as specified in Table 2

μ(USERID)	**TP**	**MD**	**LO**	**ZE**
μ(DoctorID)	0.85	0	0	0
μ(NurseID)	0	0.66	0	0
μ(PharmacistID)	0	0	0.71	0

Now assume that the following access rights exist for each data item. NA = *"no access"*, RD = "*read access"*, WE = "*write access*", RDWE ="*read and write access*", DE = "*delete access*", FA ="*Full access*". Now the data items and users of data items can be classified and categorized into fuzzy sets (with membership value), a process for determining precise actions (access rights) to be applied must be developed. This task involves writing a rule set that provides an action for any data access classification and user classification that could possibly exist. The formation of the rule set is comparable to that of an expert system, except that the rules incorporate linguistic variables with which human are comfortable. We write fuzzy rules as antecedent-consequent pairs of If-Then statements. For example:

IF *Data_Access_Classification* is **TP** and *User_Data_Access_Classification* is **TP Then** *Level_ of_Data_Access_Manipulation is* **FA**

The overall fuzzy output is derived by applying the "max" operation to the qualified fuzzy outputs each of which is equal to the minimum of the firing strength and the output membership function for each rule. Users' metadata and the metadata of each data item can be used to determine data access based on user security level and data security level for each data item.

Table 4. User access knowledgebase system based on patient attributes

Meta-data Value based on organization policy for patient data

μ(USERID)	**TP**	**MD**	**LO**	**ZE**
PatientID	RDWE	RDWE	RD	NA
Name	RDWE	RDWE	RD	NA
Address	RDWE	RDWE	RD	NA
TelNo	FA	RDWE	RD	NA

The precise actions that are allowed or not allowed on that data item by a given user can now be determined. The role of the intelligent agent is to perform data access authorization based on requested data by a user and to allow/disallow access to the data and the operations that can be performed on the data. The knowledgebase is shown in Table 4. is used by the intelligent agent in this research study for its decision making.

4 Secure RFID Data Transmission in Hospital

Due to RFID's inherent broadcasting nature, wireless communications typically pose significant challenges on data security and protection, including susceptibility to unauthorized wireless data interception. Although many privacy related issues can be addressed by security mechanisms, the protection of the source location confidentiality using conventional network security methods appears untenable.

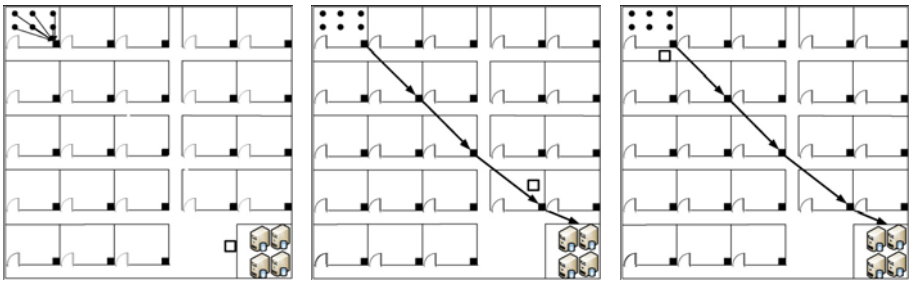

Fig. 1. An illustrative example

Location privacy is an important security issue. Lack of location privacy can lead to subsequent exposure of significant traffic information on the network and the physical world entities. For instance, cardiologic data packet coming out of a hospital in a mesh network enable an eavesdropper to analyze and find out at-risk heart patients, if the source location of those packets can be determined. Toward that goal, a number of source-location communication protocols have been proposed, where the main idea is a mixture of valid and fake messages. Each node transmits either a valid or a fake message, consistently. The main disadvantage of this approach is that the broadcasting of fake messages consumes significant amount of the limited energy in each sensor node. Moreover, because each node has to transmit a packet in every time slot, the effect is increase in number of collisions, and decrease in the packet delivery ratio. Therefore, these approaches are not suitable especially for large scale wireless sensor networks.

Routing based protocols can also provide source-location privacy. In this paper we use the Panda-Hunter model to formalize the problem in sensor networks and propose a phantom routing technique based on both flooding and single path routing. Phantom routing involves two phases: a random walk phase, and a subsequent flooding/single path routing. Random walk is inefficient in making a fake "phantom" source far enough from the actual source. To address this problem, a direct random walk is proposed (see 10, 11, 12 for more details).

This can be achieved by storing direction information in the header of the message. The exposure of the direction information decreases the complexity for adversaries to trace back to the true message source. One of the looming challenges that threaten the successful deployment of the proposed scheme in a WLAN is source-location privacy, especially when a network since the data that are transferred is extremely sensitive. If all the data information – data packets - that are collected in one room follow the same routing path toward the destination, it would be easy for an

eavesdropper to track the direction of the transmitter of each data packet and eventually track the source and reveal the room of the data transmission. Figure 1 shows an illustrative example. The eavesdropper is the square beside the central database room. When a packet reaches the database room, the eavesdropper is moving toward the direction of which the packet came from. With that back tracing technique, the eavesdropper will eventually reveal the exact location of the room that is sending the packets. In order to enhance source location privacy in the above scenario, we are proposing to use the opportunistic mesh networking scheme as given in [10,11].

Opportunistic routing is a multi-hop routing that changes the path between the source and the destination dynamically, according to network conditions in each time slot. We are using four types of packets during the packet relaying process: Request To Send(RTS), Confirm To Send(CTS), DATA and ACK. RTS/CTS are used during the handshake process between neighbor nodes while ACKs are used for verification of DATA delivery.

When a node ***s*** has to transmit a packet, first it broadcasts a RTS packet, in which it includes its own address and the destination address, ***d***, and then node ***s*** keeps listening. All the surrounding nodes which are in the range of s are able to hear this request, conforming a set of candidate nodes $\boldsymbol{E_s}$. There is a subset $\boldsymbol{V_s} \leq \boldsymbol{E_s}$ conformed by any node $\boldsymbol{i} \in \boldsymbol{Es}$ that satisfying the condition $\mathbf{c_{i,d}} < \mathbf{c_{s,d}}$ so, $\mathbf{V_s} = \{\boldsymbol{i} \in \boldsymbol{E_s} | \boldsymbol{c_{i,d}} < \boldsymbol{c_{s,d}}\}$. If a node is in $\boldsymbol{V_s}$ subset and is available for receiving a packet, and there aren't any packets in its buffer waiting to be send, it should send a CTS packet back to the sender node ***s***. In order to prioritize the nodes based on their distance from the destination, each node $\boldsymbol{i} \in \boldsymbol{V_s}$ initialize a timer, with timeout period $\boldsymbol{T_i}$, which is inverse proportional to the difference $\boldsymbol{c_{s,d}} - \boldsymbol{c_{i,d}}$ and can be determined as follows:

$$T_i = \frac{C_0}{c_{s,d} - c_{i,d}} + SIFS, i \neq d \tag{1}$$

where C_0 is a constant and SIFS is the smallest time interval composed of the module processing time and the transceiver RX/TX switch time. In the next step, node i backs off for the period T_i. If the data channel is free after that period, node i sends a CTS to the sender node, otherwise it quits. After that procedure, the sender node s will receive the first CTS from the node which is closer to the destination, and this will be the next hop relay node and it will receive the DATA packet. When the next node receives the DATA packet it replies with an ACK to the sender and follows the same procedure until the DATA packet reaches the destination. In the case that the sender node receives more than one CTS packets simultaneously there are certain mechanisms in the sender node, such as cyclic redundancy check (CRC) that can detect this collision and differentiate the nodes. When a node i sends a CTS it waits for time T_w to receive the DATA packet from the sender node s, otherwise it goes back to its previous mode. T_w is the time needed for the sender to transmit the data to that node and can be defined as:

$$T_w = d(s,i) \cdot D_0 + SIFS \tag{2}$$

where d(s, i) is the distance between the sender and the relay candidate and D_0 is a constant. In the same way, the sender node has to wait for T_c time to get a CTS packet

before it broadcasts a RTS again. T_c is the time needed for a node which is located at the limit of the range R of the sender node, and can be defined as:

$$T_c = R \cdot C_0 + SIFS \tag{3}$$

where R is the range of the sensor. The time that a sender node will wait for an ACK before it retransmits the DATA can be also defined as:

$$T_A = d(s,i) \cdot A_0 + SIFS \tag{4}$$

where A_0 is a constant.

If the set V_s is empty, meaning that there is no available node, within the range of the sender node, which will make process toward the destination, the relay node is not updated. This comes from the fact that in the next time slot there will be other nodes available (because node availability is independently generated) that could provide advancement toward the destination. Finally, after a number of packet transmissions, a number of different paths toward the destination will be discovered, as illustrated in the following Figure 2. A detailed description of the algorithms can be found at [12].

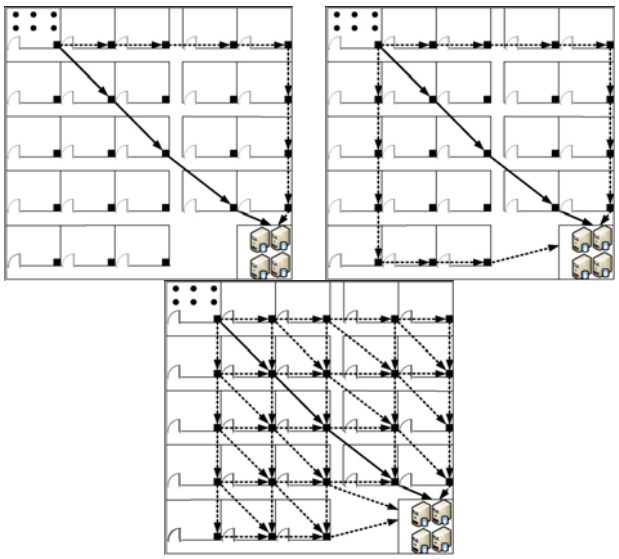

Fig. 2. Different paths toward the destination

5 Conclusion

Managing patients' data wirelessly can prevent errors, enforce standards, make staff more efficient, simplify record keeping and improve patient care. This research in the wireless medical environment introduces new ideas in conjunction to what is already available in the RFID technology and wireless networks. With the reduction in cost of radio frequency identification (RFID) technology, it is expected the increased use of RFID technology in healthcare in monitoring patients and assisting in health care administration. An intelligent agent using fuzzy logic techniques is implemented for

data/user access control and classification to improve the application of regulatory data requirements for security and privacy of data exchange. Finally an approach for passing data packets in the hospital is proposed. This proposed [10, 11, 12] approach will provide more secure data transmission in a hospital environment.

Acknowledgements. The authors would like to acknowledge the initial research work performed on this project at the University of Canberra by Dr Ric Jentzsch, Dr Masoud Mohammadian and MIT students. The authors would like also to acknowledge the research work on Data Classification that was performed initially by Dr Masoud Mohammadian and Professor Dimitrios Hatzinakos at the University of Toronto and used in this research study. Part of the research study presented in this paper is published in a journal and conferences in the past few years. The reference to these publications is provided in this manuscript when appropriate.

References

1. Angeles, R.: An empirical study of the anticipated consumer response to RFID product item tagging, Industrial Management & Data Systems, Vol. 107 No. 4, pp. 461-583 (2007)
2. Bigus, J.P., Bigus, J.: Constructing Intelligent software agents with Java – A Programmers Guide to Smarter Applications, Wiley, ISBN: 0-471-19135-3 (1998)
3. Doan, A-H. Lu, Y. Lee, T. Han, J.: Profile-Based Object Matching for Information Integration, IEEE Intelligent Systems Magazine, pp 54-59. USA (2003)
4. Ferraiolo, D.F., Kuhn D. R.: Role Based Access Control, 15th National Computer Security Conference, Oct 13-16, 1992, pp. 554-563 (1992)
5. Glover, B., Bhatt H.: RFID Essentials, O'Reilly Media, Inc.USA (2006)
6. Kowalke, M.: RFID vs. WiFi for Hospital Inventory Tracking Systems. http://blog.tmcnet.com/wireless-mobility/rfid-vs-wifi-for-hospital-inventory-tracking-systems.asp (2006)
7. Mohammadian, M. Jentzsch, R.: Intelligent Agent Framework for Secure Patient-Doctor Profilling and Profile Matching. International Journal of Healthcare Information Systems and Informatics, 1, 1-10 (2008)
8. Weinstein, R.: RFID: A Technical Overview and Its Application to the Enterprise. IT Professional Magazine, 7(3), 27-33.Whiting, R. (2004), MIT = RFID + Rx. Information Week 988: 16 (2008)
9. Zadeh, L. A.: Fuzzy sets, Information and control, Vol. 8. pp 338-352 (1965)
10. [10] Spachos P., Song L., Hatzinakos D.: Opportunistic Routing for Enhanced Source-Location Privacy in Wireless Sensor Networks", 25th Biennial Symposium on Communications, (QBSC) 2010, Kingston, Canada (2010)
11. Spachos P., Song L., Bui F., Hatzinakos D.: Improving Source-Location Privacy Through Opportunistic Routing in Wireless Sensor Networks", IEEE Symposium on Computers and Communications (ISCC), Kerkyra, Greece (2011)
12. Spachos P., Song L., Hatzinakos D.: Performance Comparison of Opportunistic Routing Schemes in Wireless Sensor Networks, 9th Annual Communication Networks and Services Research Conference (CNSR) , Ottawa, Canada (2011)