Privacy threat warning in eHealth events wireless sensor networks transmissions

Maria de los Angeles Cosio León, Juan Iván Nieto Hipólito, Raymundo Buenrostro Mariscal, and Mabel Vazquez Briseno

Autonomous University of Baja California, Faculty of Engineering Architecture and Design,
Telematic Research Group,
Carr. Tijuana-Ensenada km 103, 22860 Ensenada, Baja California, México
{maria.cosio, jnieto, rbuenrostro, mabel.vazquez}@uabc.edu.mx
http://telematica.uabc.mx

Paper received on 22/09/12, Accepted on 15/10/12.

Abstract. New technological paradigms as Wireless Sensors Networks (WSNs) are proposed to a broad application spectrum to support sensing infrastructures. These kinds of networks gather, process and transmit data with restrictions about energy, computing capabilities and transmission rates. A field where WSNs have found a comfortable application area is in eHealth. Patients wear devices with mobile capabilities sending by consistent time intervals an invariant amount of bits about physiological parameters. Due to data transmission patterns on WSNs, as well as sensed data and considering medical restrictions, privacy issues emerge. In this paper we describe a privacy threat about trajectories built while eHealth events are transmitted. Once, we do a qualitative analysis about the threat, we model it from an adversarial viewpoint by proposing an attack model whose aim is to uncover identifiable personal information and other highly sensitive information.

1 Introduction.

WSNs are a conjoined set of devices used to collect, process and transmit events gathered from their context. In this work, we analyzed a public WSN containing fixed nodes to bring support for transmissions of mobile nodes' events. The latter set of nodes are worn by patients getting in and out from WSN' coverage. The aforementioned interactions leave trails containing rich information about the mobile actors in the scenario. A trail is a trajectory described by a mobile device after an event transmission —referring event as a set of bits produced by sensing devices about physiological parameters—. Event transmissions spread information about the data's sources in near real-time such as a time when a transmission starts, type and length of the event and their trajectory. Interactions among the WSN devices and mobile devices can allow for the building of trajectory history and deduction of personal information from the connection between the mobile device and the person wearing it. Information is strongly related to four basic *Context's* elements: location; identity; activity and time [1].

The aforementioned scenario open an opportunity for a myriad of services based on location as well as activities performed by patients. Hence, unauthorized people could be interested in gaining access to private information. The scenario imposes challenges to protect people's privacy and further security issues. In order to show potential damage that an adversary could cause after gathering enough information of physiological events [2], we do a qualitative analysis about the threat related to primary indices uncovered by the trajectory history. We model it from an adversarial viewpoint, therefore, proposing an attack model whose aim is to uncover identifiable personal information and other highly sensitive information.

Organization Section 2 introduces works resolving privacy threats on transmissions events and the proposal about trajectory reduction. The System Model is described in Section 3, including the network and the data model. Section 4 describes a privacy threat of interest being our main aim in this paper, adversaries and concluding with the Attack model. In section 5 we outline our approach to resolve the threat previously described. Section 6 has a discussion about the privacy issue and our approach; finally, we close this paper in section 7 with Conclusions and future work.

2 Related Work.

Thwarting privacy damage could be accomplished through privacy enhancing techniques (PETs). Cipher is a PET to provide content privacy; in [3] is proposed an instance of these sort of mechanisms. Although it is not enough to provide protection against contextual privacy threats [4] in WSN, as transmissions are an immediately sensed event, they follow routes built by routing algorithms and consider specifically a known restrictions up to sink node (see Figure 2). So an adversary could gather information about existing transmission.

Other kinds of mechanisms achieve privacy through policies that restrict access to data. However, policies are vulnerable to inadvertent or malicious disclosure of private information as described in [5]. With the same aim, there exists anonymity techniques. Their aim is making it impossible to connect data with the data's owner. However, they have limitations; specifically in spatial application domains. Person's identity can be inferred from his or her location. Pseudonymity maintains information restricted for a subset of users. But, anonymity as well as pseudonymity are vulnerable to data mining. Further, anonymity presents a barrier to authentication and personalization, which are a paramount requirement on eHealth services.

A new branch to protect privacy in WSNs is proposed in [2]. In order to protect events observed at different instants that could be and related among them to conclude information about the data's source, (e.g. events caused by the same entity); authors propose a solution considering events with the following characteristics: (i) mobile; (ii) Start from the outermost part of the network —the perimeter—; and (iii) after a given (non predictable) interval of time expires in some place within the network. To measure solution's performance, authors used *communication overhead measure*. Also, communication and computational cost incurred through: (i) the amount of messages produced by the protocol; and (ii) the amount of messages used to forward an event itself (real or dummy) to the base station (BS). The Protocol correctness was measured using Pearson X^2 with the aim to know the association grade between two sets and them the un-observability provided. In [6] authors address the problem of protecting query privacy (e.g., hiding which node matches the query) and data privacy (e.g., hiding sensed data). They introduce a realistic network model and two novel, adversarial models; resident and non-resident adversaries. For each of them propose a distributed privacy-preserving technique and evaluate its effectiveness via analysis and simulation.

In [5], authors use obfuscation techniques, degrading data quality about location in order to protect person's location privacy while he uses location base services (LBS). To prove their proposal, obfuscating data feed a location-based service of acceptable quality. Their experiments show there exists a trade-off between quality of service offered and the privacy level. Key information is the size of the initial obfuscation set; although their results show that obfuscation techniques are adequate to resolve the person's location privacy problem. Authors conclude that there is needed a deepest research about some setup parameters to achieve a good trade-off between privacy and quality of service.

Besides the above mentioned proposals, there exist WSNs applied to know animal habits; they are constantly sending information about animals location. Considering this scenario, authors of [7] propose the question: *How many transmissions are possible to reduce trajectories without losing our solutions intention?*, in reference to reducing energy consumption by these sort of application. They answer the question, tackling the problem through a distributed variant of the Douglas-Peuker heuristic for polyline reduction, from *Computational Geometric literature*, augmented with temporal awareness. To control the quality of their solution, authors settled error boundaries about location of Object of Interest.

Above mentioned works provided us insights to resolve the problem proposed in this paper; though they are not aware of satellite information spread by data transmissions from observed entities (e.g. trajectories performed by low mobility devices or eHealth events appearing in fixed intervals can be related to physiological measures by basic mathematical operations over transmitted packets, including events that expire at any node in the WSNs). Being aware of above mentioned characteristics, an entity has capabilities to deduce personally identifiable information (PII) as well as sensitive information. Therefore, our aim is to show a contextual privacy threat produced by additional information spread by eHealth event transmissions on WSNs.

3 The System Model.

The system model has four basic actors: (i) a mobile node; (ii) a public WSN; (iii) a sink node; (iv) an eHealth system in Internet, besides of interactions allowed among them (see Figure 1 for details). The first actor could have sensing capabilities, as sensing devices installed on it. The mobile device's profile allows to configure sensing intervals, as well as quality (maximal bits by sample). Both features are carefully monitored by the eHealth system. Hence, if an event is not received or accomplished with characteristics aforementioned, an alarm is activated to notify potential emergency or device failure. Also there exists a mechanism to control packet loss parameter up to where there is no information damage. Public WSN is supporting eHealth events transmissions, described in detail at Network Model below. A sink node extends WSN coverage to Internet making a vertical hand-off.

The eHealth system includes a set of control mechanisms to safely interact with the actors in the network. The first barriers are authentication and authorization subsystem managing access to upper layer services such as data storage, data processing, routing and others.

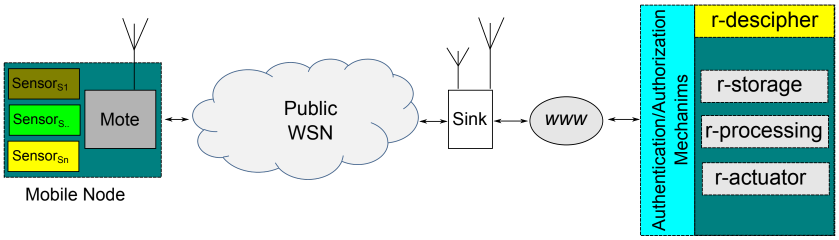


Fig. 1. System model

3.1 The Network Model.

There exists a public WSN that supports eHealth events transmissions from patients wearing devices with mobile capabilities. It is deployed over a finite area forming a grid. So patients could get in or out of the coverage as they perform their activities. We denote the WSN as G. It is defined by the pair $< N, E >$; where N is the set of nodes in G containing four nodes' subsets; $N = \{I, P, M, S\}$. The subset of *inside nodes* defines a *connected dominating subset* in the WSN denoted as I. Another subgroup of nodes is P, corresponding to *perimetrical nodes* in the WSN (graph search and the graph frontier problem). Third subset is M, the group of mobile devices denoted as $M = m_1, \dots, m_{|M|}$. These sort of devices are $M \notin N$, when there are no transmissions requested. The subset's size M is related to t (a time instant), hence $M_t = |M|$, the number of mobiles nodes connected through some node of $(N - (M + S))$. Offering a broad vision about the network, Figure 2, shows nodes in the scenario as well as the trajectory followed by a mobile device performing hand-off processes to transmit an eHealth event, and the set of routes built by a routing protocol to drive packets up to sink node.

Without restrictions in low layers (MAC and physical) to get into the network a mobile device can start to have interactions with nodes in the subset $N - M$. However, restrictions exist in the upper layer to provide services such as routing R. Therefore, an authorization and authentication mechanism should be successfully accomplished by the mobile node $m_i \in M$ as Figure 1 shows. Once m_i fulfils the entire security requirements, the first interaction between $m_i \ and \ r\_Services$ is routing to build and activated routes, getting them ready for transmissions. Lastly, S is the subset of sink nodes, following an uniform distribution in the WSN. The subset S has a size $|S| \geq 1$.

Besides the aforementioned nodes subsets, in G exists E, the set of sensed events. They are produced by M and transmitted by the subset of nodes $(N - (M + S))$.

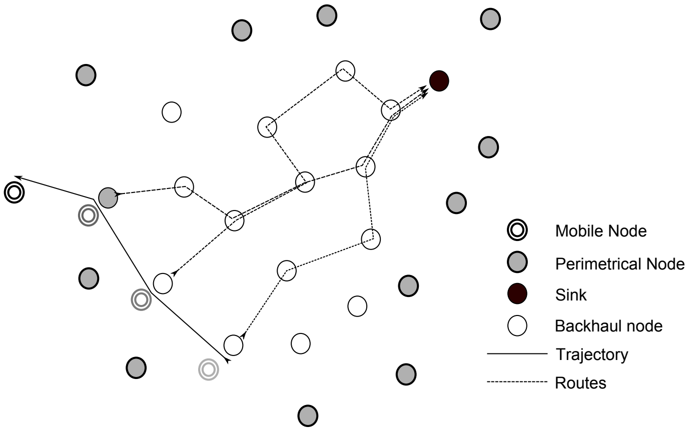


Fig. 2. Trajectory followed by a mobile node, as well as routes carrying packets up to sink node

3.2 The Data Model.

Previous considerations: Mobile devices turn-off the ratio when there is no necessity to transmit data and turn on at fixed time intervals to request possible queries on the eHealth system. About the type of events; we are focusing on ones which: (i) initiate on a random within the WSN (perimetrical or inside nodes); (ii) after a known amount of bits transmitted finishes within the network. It is possible that a mobile device does not finished to send the event due to transmission starts near perimetrical nodes and it losses WSN coverage; (iii) a trajectory in network nodes are a projection of patients' trajectory in physical space.

Data interactions as well as nodes characteristics could be modelled as follows. Mobiles nodes M have at least one sensing device, hence a set of sensing devices are defined as $< st_1, st_2, \ldots, st_n >$; where n is the number of sensing devices allowed in m_i. Sensing tasks are configured off-line, considering medical standards to measure physiological parameters. Sensing interval features allow to define the time interval between two sensed events. Hence this feature related to entire sensing device by node could be defined as follows: $< st_{1t_1}, st_{2t_2}, \ldots, st_{nt_n} >$, and a sensing device $< st_1 >$ will produce an event once a time t_1 is elapsed. Event length is related to a kind of sensing device, so a sample is defined by a triad $< St, T, \mathbb{L} >$ where St is a class of sensing device, T is the interval between samples and $\mathbb{L}$ is event's length in bits. Figure 3 describes a set of events, as well as packets required by the event.

4 Threat: Trajectories.

Based on Kerckhoff Principles [8], adversaries have knowledge about network protocols and privacy mechanisms. Besides the characteristics about people wearing mobile nodes (e.g. old or sick people, walking slow). The adversary performing a passive attack could gather information about hardware and data. Through samples resolution and

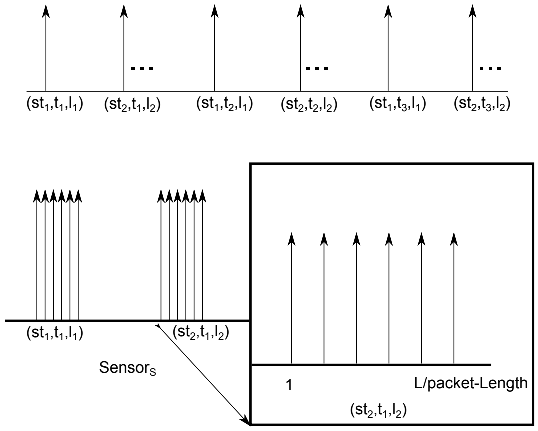


Fig. 3. Samples transmissions

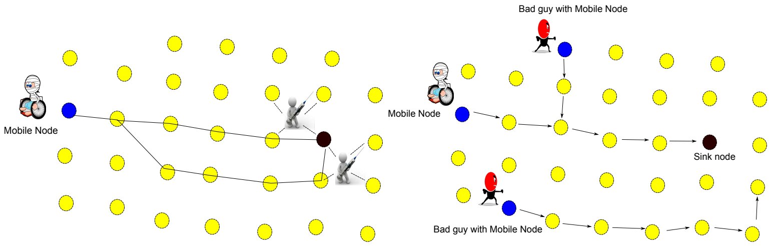
type of event analysis. The same attack class could evolve into active attacks in the network, also in the real world considering that monitoring tasks are carefully configured and monitored. So changes in transmissions patterns could be detected as emergency events or node failures.

Patterns of eHealth events in WSN transmissions could be classified as follows: transmission domains (private or clear); source node's trajectory (trajectory on a set of nodes or in a simple node); time (periodic or single events); and lastly, size of event (time intervals to transmit data or bit numbers). Complete information is a valuable asses for business people. On the other hand, location is inextricably linked to personal safety, so unrestricted access to information about patient location could potentially lead to harmful encounters (for example stalking or physical attacks). We consider a sequence of recorded locations for a person constitutes a quasi-identifier that can be used in combination to identify the user [9]. So trajectories can be used to infer much about a person, even without a name attached to the data. We are aware about it, so our goal is to show a strategy to perform attacks using a trajectory's history and event characteristics with intention to deduce PII (personal identifiable information). In intention to build mechanism to thwart possible privacy damage.

4.1 Attack Model.

In this subsection we define the types of adversaries and kind of damage that could be inferred by them, as a result of related adversarial behaviour and events along the WSN. To uncover vulnerabilities on trajectories, we model events considering them in order to make relations between trajectories and the adversaries.

Adversaries. We are considering two types of adversaries. The first one is an external, passive and global coverage adversary. This means that the adversary is no part of WSN, she has her own device to sniff traffic and to gather all information about events in the network (e.g. node, laptop with capabilities to store trajectory records). Her aim is to know identifiable personal information and other highly sensitive information. To achieve her aim, the adversary uses traffic analysis techniques, such as temporarily on transmission and their length. The second one is an active, external and short coverage adversary. It modifies routes in the scenario by making them unfeasible near the sink devices, through *traffic injection*, in order to modify parameters related to routing algorithm (see Figure 4.a). One strategy of this attacker is to drive the traffic through specific nodes to perform a *sink-hole* attack or accessing to data using less resources. The second strategy is to request by routing services an emergency transmission with the intention to reduce emergency service effectiveness. Figure 4.b shows this strategy. Once the adversary knows how many patients and what class of disease exists, it can perform a *denials of service* attack and reduce service confidence by the users.



(a) Exhausting routes attack (b) Requesting for Routing Services

Fig. 4. Active adversary strategies.

Modelling Events. An event can be described by its length as:

$$e_j = \{< t_x, l_j(t_x) > \ldots < t_{x+\Delta x}, l_j(x + \Delta x) >\}, \qquad (1)$$

where

- l_j is St related and l_j has a maximum length defined by sensor type;
- $l_j : \mathbb{T} \to \mathbb{N}$ is a crescent monotonic function, where its upper threshold is defined $< St, t_{\Delta x}, \mathbb{L} >$;
- t_x is distinct time instant up to $t_{x+\Delta x}$;
- Δx is the elapse time to perform a sensing task.

The trajectory built by a mobile node, whit intention to transmit an event is related to the subset nodes in $(N - (M + S))$ through m_i sending partial or full event's bits. The trajectory along the nodes could be described as:

$$traj_e = \{< n_i, t_x >, \ldots, < n_{i+h}, t_{x+\Delta x} >\}, \qquad (2)$$

where n_i is a node from a mobile device starts to send an event up to n_{i+h}, h hops away from n_i in mobile node's trajectory. It is important to denote that if the node

$n_{i+h} \in P$ and $l_j \notin Ls$, where Ls is the set of event size allowed. Then is possible to conclude that the mobile node lost network coverage. Therefore, trajectories related to unfinished transmissions are not adequate to define the event type. On the contrary, it is useful information to relate physical locations in the scenario at t_x along of events with $l_j \in Ls$. So, a trajectory describes a set of physical locations from a mobile node in a period $T_{0,x}$ while it is transmitting a class of physiological lectures defined through Ls.

The relation ship of equations 1 and 2 can offer valuable information about events, allowing to evolve into natural language expressions as (e.g. Patient is walking in the park while sensing heart rate). Continuous tracking services allow to related location with activities of daily living as an opportunist for sensing relevant information by medical personnel and caregivers or adversaries described above.

Uncovering Information from Trajectories. Authors in [2] propose Equation 3 to acquire events with a length of interest; we will use it with the same goal. Now, given a time instant $t_i \in \mathbb{T}$ and a subset of events E, the function $length_E : \mathbb{N} \times \mathbb{T} \rightarrow \mathbb{N}$ will provide the number of events with ls_i at the t_i:

$$length_E(l_i, t_i) = \sum_{e_j \in E} \epsilon_{e_j}(t_i, l_i), \tag{3}$$

$$where\ \epsilon_{e_j}(t_i, l_i) = \begin{cases} 1 \text{ if } < t_i, l_i(t_j) > \in e_j \\ 0 \text{ otherwise} \end{cases}$$

Selecting a point in the network time scale (elapsing time from the network start-up up to stop) of one event type, we will model this relation to show information discover.

A mobile node m_i will send information to the sink node through the subset of nodes $(N - (M + S))$. Each node in $(N - (M + S))$ have a single network interphase and the communication channel is shared by clients along the network time scale, t_i, so interactions among nodes are restricted to time slots.

The history of event transmissions performed by node m_i is a set of ordered pairs with $m_i = \{< t_0, e_0(l_j(t_t)) > \ldots, < t_n, e_n(l_j(t_t)) >\}$. The history of events in nodes $(N - (M + S))$ is defined by a set of ordered pairs $< t_i, e_x(l_j(t_t)) >$, where t_i is a point in the network time scale, e_x, an event x attended by $node_i$ and $l_j(t_t)$ is the $event_x$'s length on the event time scale (time elapsed by event transmission). The trajectory of an event through nodes $< n_i, n_{i+1}, \ldots, n_{i+h} >$ is defined in Equation 4.

$$te_j = \{< n_i, t_i, e_x(l_j(t_t)) >, \ldots, < n_{i+h}, t_{i+\Delta i}, e_x(l_j(t_t)) >\} \tag{4}$$

Where e_x is an event sensed by a mobile device and transmitted through a subset of nodes, $l_j(t_t)$ is the number of bits transmitted by nodes along the mobile node's trajectory in the time interval $T_{i,i+\Delta i}$. Figure 5 illustrates a trajectory followed by a patient while wearing a device that is transmitting an event.

Considering that each node in N has an unique ID, m_i requires to send its unique ID along with the routing query in intention to identify owner request. Here, the system has enough information to detect who is requesting by a service [10]. On the other hand, the restricted service routing, R, once it is authorized, offers a short path between source and destination nodes. At this point, an adversary could get enough information about

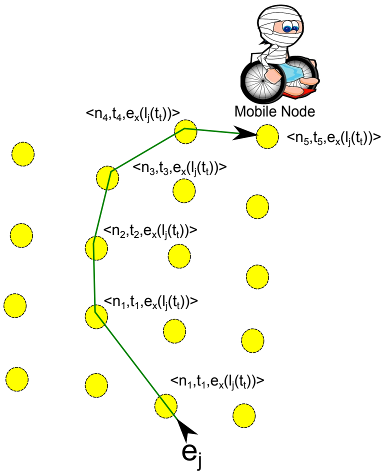


Fig. 5. Event network and physical trajectory

each event and the trial could be defined as $< m_i, St_j, \mathbb{L}, T >$; where $0 < j \leq |St|$. These proposed models allow to relate nodes with events, to filter by event type and to denote m_i locations.

5 Outlining Our Solution

Our system model has security considerations on upper layers outside of the WSN. To control access to remote services (r-services), there exists a r-service provider it includes subservices of: authorization; authentication; decipher; storage; processing and actuator subsystem shown in Figure 1. Data transmissions along the network are performed following routes that build a bioinspired algorithm as AntNET [11]. Due to routes been activated by forwards ants (ants tracking back to path to source node) instead of explorer ants (ants searching the sink node), allows the sink node to have control over the activation process. Routing process activation requires non-reusable tokens from mobile nodes, so an initial token is stored in the mobile device. When a mobile node requests a routing process, it sends hashed token, $h^y(token)$. Once the sink node receives the token, it sends it up to internet system. After this, the system provides a token, $h^x(token)$, and sink nodes wait for the same token from the mobile node to activate a routing process as Figure 6 shows. In other cases, nodes around the source are notified about a possible attack.

We assumed that nodes form a grid. this grid is divided by set regions, A. For each region of a_i, there exists a node called a sink. This node has capabilities to extend Internet services into the WSN. This device has a coverage area defined as: $sink_i =< area_i, s_i >$, where $area_i$ border is CA hops away from $sink_i$. Further, CA defines the maximum coverage area of $sink_i$, and $m_i t$ denotes people in the area at t using n_i sensor node to transmit its data along $area_i$.

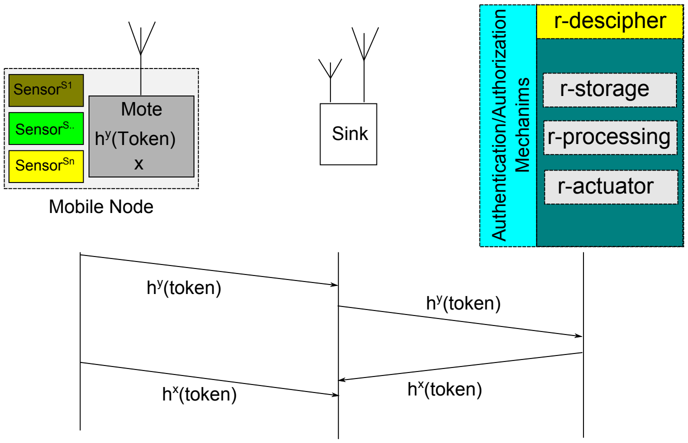


Fig. 6. Chain hash token

Afterwards an event starts and m_i is losing n_i's coverage. It starts a hand-off request, instead to build a new route. Current route is extended using previous knowledge provided by the hand-off mechanism about destination node. However, making a temporal analysis on transmissions, it is possible to know about this change. Thus we propose to introduce a random delay on transmissions in order to prevent adversary knowledge about the change.

6 Discussion.

Information by it self has no privacy requirements, although when it is possible to make a relation between information and the owner, a privacy threat appears. Trajectories allow to build aforementioned relation ship, although providing privacy on trajectories is a paramount task, because additional considerations are necessary to protect them. So the satellite mechanism, in the application layer must be aware of privacy and security requirements; in intention to enhance performance on the lower layer privacy mechanisms. This work proposes aforementioned consideration. Although, it is necesary to define the scope of each mechanism besides of their relation ship. Single solutions cut a path, but a broad opportunity to perform an attack persist. This work has challenges to resolve about the privacy mechanism implementation, metrics to quantify damage and the privacy threat about information spread by transmission length. However this research in progress offers information to resolve them, through mathematical models proposed.

7 Conclusions.

This proposal shows a privacy threat related to eHealth events through WSNs transmissions, adversaries as well as their strategies. The scenario above describes that without privacy considerations, a broad opportunities for unauthorized personnel to gain access to PII is very real. Considering events' length; time; location an adversary has information from at lease two primary indices from context. This increases the likelihood for successful attacks to gather sensitive information by crossing efforts of active and passive adversaries as defined in section 4. Protecting this scenario is a challenging task as once the adversary knows two types of the most important elements of patients' context, the rest can be deduced through them [12]. Our initial effort tries to cover time elapsed by hand-off mechanism. Although events length requires additional techniques that must be thoroughly analyzed for its performance. The future work in this project includes simulation, hand-off delay considerations, as well as metrics to quantify damage.

References

1. B. Schilit, N. Adams, and R. Want. Context-aware computing applications. In *Proceedings of the 1994 First Workshop on Mobile Computing Systems and Applications*, WMCSA '94, pages 85–90, Washington, DC, USA, 1994. IEEE Computer Society.
2. S. Ortolani, M. Conti, B. Crispo, and R. Di Pietro. Events privacy in wsns: A new model and its application. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a*, pages 1 –9, june 2011.
3. David L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Commun. ACM*, 24(2):84–90, February 1981.
4. Trevor Darrell & Daniel J. Weitzner Mark Ackerman. Privacy in context. *HumanComputer Interaction.*, 16(2-4):167–176, 2001.
5. Matt Duckham and Lars Kulik. Simulation of obfuscation and negotiation for location privacy. In Anthony G. Cohn and David M. Mark, editors, *COSIT*, volume 3693 of *Lecture Notes in Computer Science*, pages 31–48. Springer, 2005.
6. Emiliano De Cristofaro and Roberto Di Pietro. Preserving query privacy in urban sensing systems. In *Proceedings of the 13th international conference on Distributed Computing and Networking*, ICDCN'12, pages 218–233, Berlin, Heidelberg, 2012. Springer-Verlag.
7. Goce Trajcevski, Oliviu C. Ghica, and Peter Scheuermann. Tracking-based trajectory data reduction in wireless sensor networks. In *Proceedings of the 2010 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, SUTC '10, pages 99–106, Washington, DC, USA, 2010. IEEE Computer Society.
8. W. Trappe and L.C. Washington. *Introduction to cryptography: with coding theory*. Prentice Hall, 2002.
9. Claudio Bettini, Xiaoyang Sean Wang, and Sushil Jajodia. Protecting privacy against location-based personal identification. In Willem Jonker and Milan Petkovic, editors, *Secure Data Management*, volume 3674 of *Lecture Notes in Computer Science*, pages 185–199. Springer, 2005.
10. Naveen Sastry and David Wagner. Security considerations for ieee 802.15.4 networks. In *Proceedings of the 3rd ACM workshop on Wireless security*, WiSe '04, pages 32–42, New York, NY, USA, 2004. ACM.
11. G. A. Di Caro. *Ant Colony Optimization and its application to adaptive routing in telecommunication networks*. PhD thesis, Faculté des Sciences Appliquées, Université Libre de Bruxelles, Brussels, Belgium, November 2004.
12. Anind Dey, Jeffrey Hightower, Eyal de Lara, and Nigel Davies. Location-based services. *Pervasive Computing, IEEE*, 9(1):11 –12, jan.-march 2010.