Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 9, September 2014, pg.332 – 339

RESEARCH ARTICLE

PROVIDING DATA SECURITY FROM ANOMALOUS USERS WITH DATA AGGREGATION SCHEME

Mohammed Amair[1], Mohd Abdul Wasey Muzakkir[2], M. Saidi Reddy[3]

[1] amer5350@gmail.com, [2] Goal.Muzakkir@gmail.com, [3] Msreddy33@gmail.com

[1]M. Tech, IV semester, Department of CSE, Malla Reddy College of Engineering and Technology, Hyderabad
[2]M. Tech, IV semester, Department of CSE, Alhabeeb College of Engineering and Technology, Chevella R.R Dist
[3]Head of the Department, CSE, Malla Reddy College of Engineering and Technology, Hyderabad

---

Abstract: Data aggregation is a key aspect of many distributed applications, such as distributed sensing, performance monitoring, and distributed diagnostics. In such settings, user anonymity is a key concern of the participants. In the absence of an assurance of anonymity, users may be reluctant to contribute data such as their location or configuration settings on their computer. In this paper, we present the design, analysis, implementation, and evaluation of Anonygator, an anonymity preserving data aggregation service for large-scale distributed applications. Anonygator uses anonymous routing to provide user anonymity by disassociating messages from the hosts that generated them. It prevents malicious users from uploading disproportionate amounts of spurious data by using a lightweight accounting scheme. Finally, Anonygator maintains overall system scalability by employing a novel distributed tree-based data aggregation procedure that is robust to pollution attacks. All of these components are tuned by a customization tool, with a view to achieve specific anonymity, pollution resistance, and efficiency goals. To demonstrate the usefulness of Anonygator, we have used it to prototype three applications, one of which we have evaluated on Planet Lab. The other two have been evaluated on a local testbed.

KEYWORDS: WSN, location privacy, secure routing, context summarization, aggregation

INTRODUCTION:

Data aggregation is a key aspect of many distributed applications. Examples include aggregation of mobile sensor data for traffic monitoring in a city, network performance statistics from home PCs for a network weather service, and machine configuration information for a distributed diagnosis system. In such settings, user anonymity is a key concern of the participants. In some cases, this concern is driven by privacy considerations. For example, a user may be willing to have their GPS-enabled phone report traffic speed information from a particular street so long as the system is not in a position to identify and tie them to that location. Likewise, a user may be willing to have their home PC report the performance of a download from so long as the network weather service they are contributing to is unable to identify and them to accesses to possibly disreputable content. In other cases, the desire for anonymity may be driven by security considerations. For example, a host may reveal local misconfigurations (e.g., improperly set registry keys on a Windows machine) while contributing to a distributed diagnostics system such as Peer Pressure. Some of these misconfigurations may have security implications, which would leave the host vulnerable to attacks if its identity were also revealed. Given such security and privacy concerns, an absence of an assurance of anonymity would make users reluctant to participate, thereby impeding the operation of community-based systems mentioned above. The resource-starved nature of sensor networks poses great challenges for security.

However, in many applications the security aspects are as important as performance and low energy consumption. The security challenges include the extremely large number of interacting devices in a sensor network and the dynamic nature of WSN, that is, frequent changes in both its topology and its membership. Privacy is the ability of an individual or group to seclude them or information about themselves and thereby reveal who they are selectively. As location tracking capabilities of mobile devices are increasing, problems related to user privacy arise, since user's position and preferences constitute personal information and improper use of them violates user's privacy. Our new private data aggregation protocol is based on a special broadcast communication scheme, where the nodes of the cluster organize themselves into a ring and each data packet to be included in the aggregated result is sent around that ring. Note that a broadcast communication scheme is necessary, because in our scheme the nodes do not know the identity of the aggregators, therefore, they can only send data to the aggregators by broadcasting. We chose the ring based broadcast scheme, because our private query protocol exploits its properties. Our new private query protocol allows the aggregator nodes to respond to the queries of the base station without leaking any information about their identity. For this, a query token is passed around the ring, and each non-aggregator node adds some noise to the token, while the aggregator adds noise and the aggregated result.

When the token is returned to the base station, it extracts the aggregated result by removing the noise. The proposed privacy mechanisms represent a promising approach to ensuring user privacy in numerous applications, including cloud services, medical privacy, sensor network aggregation, and smart metering.

RELATED WORK:

A thorough Literature Review of the available papers is done and some of the papers are listed along with the context in which the idea of the paper was studied for the inception of this work. Computational location privacy algorithms treat location data as geometric information, not as general data. Studies show that people are generally not concerned about location privacy, although they are sensitive to how their location data could be used, and their sensitivity may rise with their awareness of privacy leaks. Offering users to control who can read what tags is difficult due to the high number of items and the lack of user interface. The low resources available on passive RFID tags additionally challenge the use of traditional security protocols.

The design of privacy preserving ubiquitous computing context is relative to not only technical implementation but also procedures of services. The proposed way is based that the level of privacy protection ideally should be decided by end user of service. Although it has been shown that it presents some flaws and limitations, and that finding an optimal k-anonymization is NP-hard, the k-anonymity model is still practically relevant and in recent years a large research effort has been devoted to develop algorithms for k anonymity. The concept of location k-anonymity for LBS was first introduced in and later extended in to deal with different values of k for different requests. The underlying idea is that a message sent from a user is k-anonymous when it is indistinguishable from the spatial and temporal information of at least $k - 1$ other messages sent from different users.

WSNs use multi-hop routing and wireless communication to transfer data and hence are vulnerable to routing attacks. There are a lot of approaches to ease routing security. Some secure AODV algorithms have some effects on defending against external attacks. An on demand routing protocol for ad hoc to provide resilience to Byzantine failures can be classified into three successive phases: route discovery with fault avoidance, cryptographic primitives and link weight management. An approach to route recovery with one-hop broadcast to bypass compromised nodes in wireless sensor networks was provided by An and Cam. As for multi-path routing, it is more reliable, though it introduces more communication overheads. Multi-path routing, location disguise, and relocation methods can be used to protect base stations. However, in the environment where the network has a large number of compromised nodes, if the compromised node can modify the routing data, system may involve more security issues. PRSA (path redundancy based security algorithm) uses alternative routing paths for each data transmission call to overcome the sensor network attack. To enhance network reliability, PRSA allows sensor node data to be sent on defined routing paths using various transmission modes including round robin, redundant and selective modes.

Protocol:

The private cluster based data aggregation protocol consists of four main parts. The first part is the initialization, which provides the required communication channel. The second part is needed for the data aggregator election. This subprotocol must ensure that the cluster does not remain without a cluster aggregator. This must be done without revealing the identity of the elected aggregator. The third part is needed for the data aggregation. This subprotocol must be able to forward the measured data to the aggregator without knowing the identifier of that node. The last part must support the queries, where a possibly mobile operator queries some stored data.

Flaws in the Existing Methodologies:

The existing techniques suffer from several major problems including lack of secrecy and privacy, which makes the network vulnerable to adversaries and attacks. Redundancy exists in data, resulting in data and network overload. Mere XOR operations are performed for data aggregation. Hence, the security factor is not that high.

PPP Proposal

The objective of this work is to develop a system that performs data aggregation in the application of wireless sensor police patrol networks focusing on security and routing to increase the efficiency and reduce the communication overhead. To solve the above described problems, the work centres around a solution that constitutes of location privacy, secure routing, aggregation and pattern identification. This approach integrates the following to ensure privacy, secrecy and less overhead.

Location Privacy

Privacy must be insured over the data sent to the nodes in the network. Thus, an efficient algorithm is devised to obfuscate the data to the adversaries and thereby, refraining them from eavesdropping on the content.

Secure Routing

It is apparent that, texts can be broken down by the adversaries using cryptanalysis. Hence the process of routing plays an equal role in the process of security. A randomized multi-level routing is formulated to do the necessary.

Data Aggregation

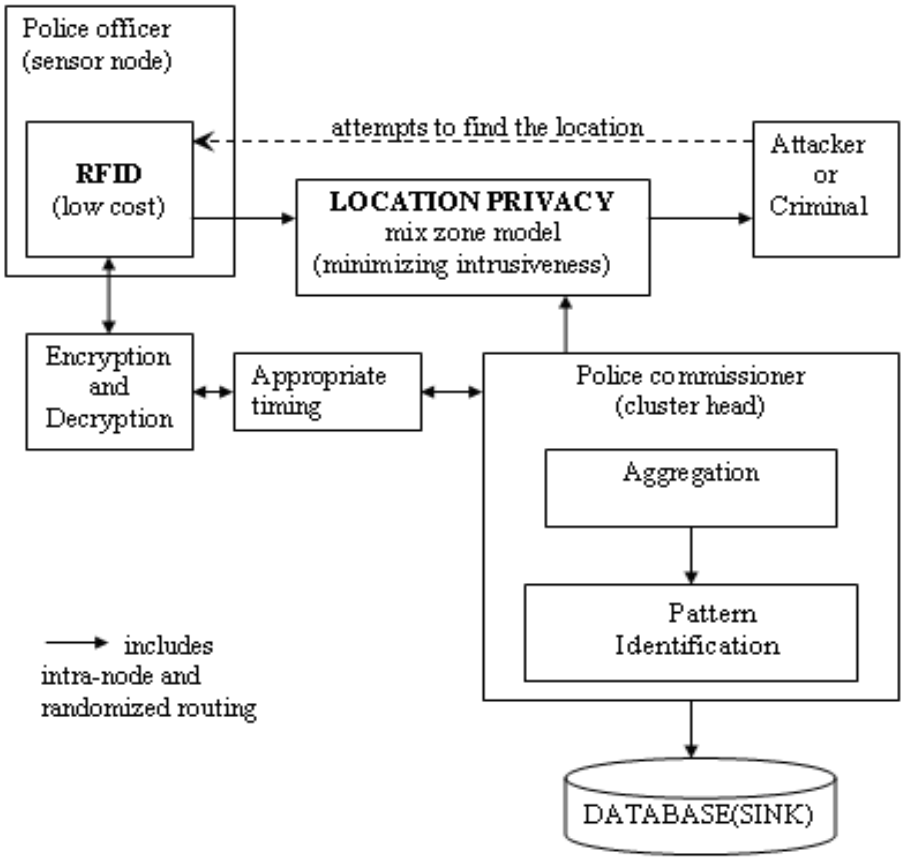
The aggregators collect data from a subset of the police officers, aggregate the data using a suitable aggregation function and then transmit the aggregated result to the next module. This is where the redundancy is removed.

Pattern Identification

The aggregated data can be summarized by identifying general patterns, which is later used to answer related queries. This adds meaningfulness to the police patrol application.

PPP Architecture

The proposed PPP architecture is illustrated in Figure 1. The RFIDs of the police officers are registered to the Police Commissioner's database in prior. The attacker/criminal attempts to read the position of the nearby police officers. The police commissioner recognizes it as a nonregistered ID and replies by obfuscating the positions of the police officers in that zone (Location Privacy module). The police officer sends encrypted data to the police commissioner, which is aggregated to remove redundancy and a pattern, is identified and this knowledge summarized data is stored in the database for further querying.



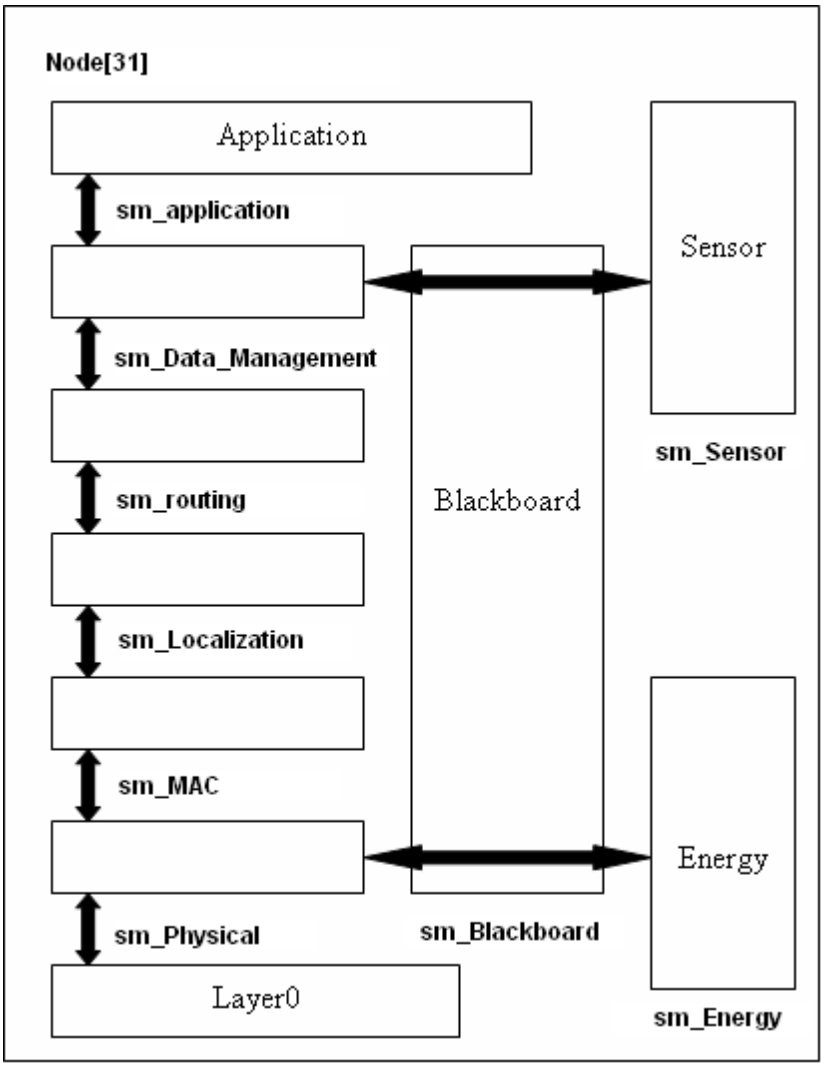
Location Privacy

Location privacy is needed to ensure that no eavesdropping of critical data takes place. The criteria to achieve privacy are trustworthiness, appropriate timing, perceptibility, unobtrusiveness, minimal intrusiveness, flexibility, meaningfulness and low cost.

Radio Frequency Identification (RFID) Tags represent the most prominent ubiquitous computing technology when it comes to privacy issues. Challenges posed by RFID tags are fourfold. The RFID tags are automated. Data acquisition is made much easier by the use of simple reader gates that can easily scan large numbers of tags. The tag identifies the individual serving its purpose and is well integrated so that no criminal can easily spot it. The last challenge is that posed by authentication as a lot of sensitive information is given. The four attributes of RFID applications threaten two classes of individual privacy that being data privacy and location privacy. If a tag ID that is associated with a person is spotted at a particular reader location then the location privacy of that person is threatened.

Secure Routing

In the PPP system, routing comes into picture while information has to be passed securely between the Police Officer and the Police Commissioner. It is a two way communication process that could occur simultaneously. The routing process is categorized into two levels: intra-node routing and randomized routing.



Routing layer implements the randomized routing technique. As for Randomized Routing, the message is split into shares and the message is reconstructed at the aggregator when the minimum needed shares are received.

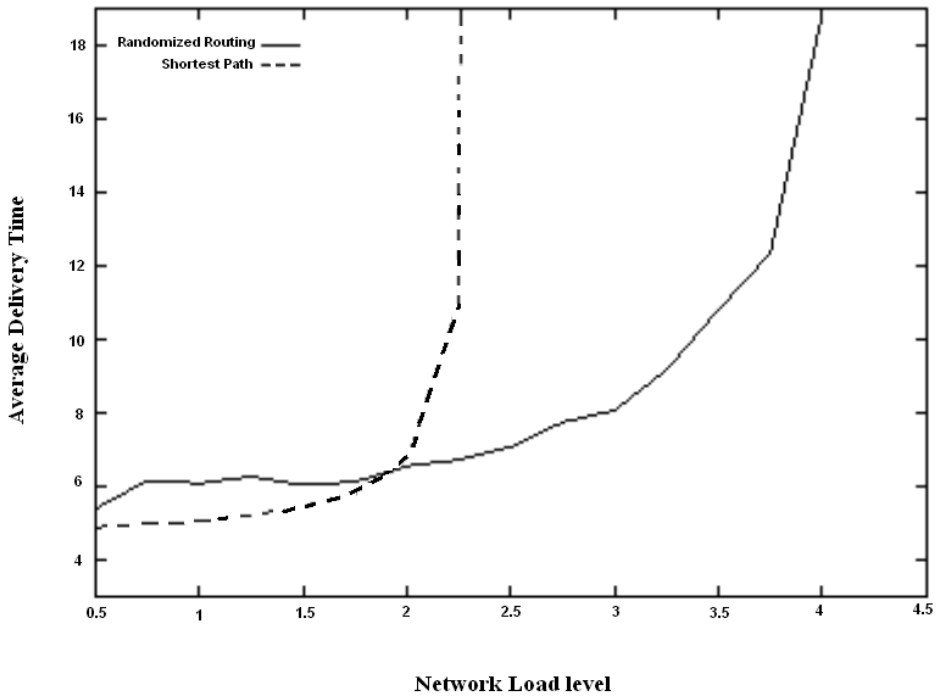
Data Aggregation

Data aggregation is needed in this system so that the database of the Police Commissioner is not overloaded with redundant data of criminals, from the Police Officers. Two or more Police Officers may forward information about the same criminal from the same location of crime to the Police Commissioner. It is enough if only one copy of this information is retained in the database hence saving the storage space and increasing the efficiency of the system.

Pattern Identification

A query that needs only limited information may have to process the entire information to find a simple result. This problem is solved by finding meaningful patterns from the information present in the database and storing them in a summarized manner. Hence a query whose answer matches the stored pattern need not perform the time consuming processing.

RESULT ANALYSIS:



The performance analysis graph shows us that randomized multilevel routing performs better at higher network load levels when compared to mere shortest path routing. This proves that when the number of nodes is high; even though the network may seem to be loaded with obfuscated messages, the messages are delivered efficiently in a faster and a more secure manner using the randomized routing technique.

```
c:\Users\Uttara\Documents\Visual Studio 2008\Projects\Privacy\Debug\Privacy.exe

Detecting RFID:
Enter the requesting ids:501 502 701 503 702

Police Officers detected:       501     502     503
Criminals detected:     701     702
Choose: 1.Process officers' data         2.obfuscate data to criminals  3.Exit
Choice:1
Enter the zone number:0

Police officers assigned to this zone are: 501  502     503
Enter filename: p1.txt
Enter filename: p2.txt
Enter filename: p3.txt

 Data is aggregated.

Using a context aware pattern, criminals details are recorded.
Press a key.
Choose: 1.Process officers' data         2.obfuscate data to criminals  3.Exit
Choice:2

OBFUSCATING DATA - to achieve LOCATION PRIVACY
@ POLICE COMMISSIONER End:
 Enter the anonymity level and the uncertainity: 57 0.001
Parameters:
  K=57, delta=0.001, pi=5, delta_max=0.010, trash_max=10.0%
Loading data...
  Loading objects... Done.
  -> Trajectories: 1697, Points: 96398, Diameter: 34702.982
  -> Removed Trajectories: 16, Removed Points: 7346
Creating equivalence classes...Done.
Processing equivalence classes: Done!  [ 57 equiv. classes ]

Choose: 1.Process officers' data         2.obfuscate data to criminals  3.Exit
Choice:3
```


Any unrecognized RFID is considered to be a criminal. The commissioner will act accordingly and obfuscate the whereabouts of the police officer, using the location privacy technique. The anonymity and the uncertainty levels are set or defined by the police commissioner based on the network load.

Conclusion

In this paper, we proposed a new private aggregator node election protocol for wireless sensor networks that hides the identity of the elected aggregator nodes. We also proposed a private data aggregation protocol and a corresponding private query protocol which allow the aggregators to collect sensor readings and respond to queries of the operator, respectively, without revealing any useful information about their identity. Our protocols are resistant to both external eavesdroppers and compromised nodes participating in the protocol. Our current and future work is concerned with the replacement of the ring based broadcast communication scheme with spanning trees. Trees would provide a better solution, because their existence is guaranteed in connected graphs, unlike the existence of Hamiltonian cycles. Trees can also be constructed much more efficiently. However, switching to trees require some modifications in our query protocol. We are also planning to develop a prototype implementation of our protocols and to analyze their performance.

References

[1] D. J. Bernstein and T. L. (editors). eBACS: ECRYPT benchmarking of cryptographic systems. http://bench.cr.yp.to, accessed 7 March 2011.

[2] D. Boneh, E.-J. Goh, and K. Nissim. Evaluating 2-DNF formulas on ciphertexts.In TCC, 2005.

[3] C. Castelluccia, A. C.-F. Chan, E. Mykletun, and G. Tsudik.Efficient and provably secure aggregation of encrypted data in wireless sensor networks. ACM Trans. Sen. Netw., 5(3):1–36, 2009.

[4] A. Cavoukian, J. Polonetsky, and C. Wolf. Smart- Privacy for the smart grid: embedding privacy into the design of electricity conservation. Identity in the Information Society, 3(2):275–294, August 2010.

[5] C. Dwork. Differential privacy. Invited talk at ICALP, 2006.

[6] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In EUROCRYPT, 2006.

[7] C. Dwork, F. McSherry, K. Nissim, and A. Smith.Calibrating noise to sensitivity in private data analysis In TCC, 2006.

[8] C. Gentry. Fully homomorphic encryption usingideal lattices. In STOC, pages 169–178, 2009.

[9] A. Francillon and C. Castelluccia, "TinyRNG: A cryptographic random number generator for wireless sensors network nodes," in Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops, 2007. WiOpt 2007. 5th International Symposium on, April 2007.

[10] N. Li, N. Zhang, S. Das, and B. Thuraisingham, "Privacy preservation in wireless sensor networks: A state-of-the-art survey," Ad Hoc Networks, 2009.

[11] W. He, X. Liu, H. Nguyen, K. Nahrstedt, and T. Abdelzaher, "Pda: Privacy-preserving data aggregation in wireless sensor networks," in Proceedings of Infocom. IEEE, 2007.

[12] B. Sheng and Q. Li, "Verifiable privacy-preserving range query in two-tiered sensor networks," in Proceedings of Infocom.IEEE, 2008.

[13] J. Deng, R. Han, and S. Mishra, "Decorrelating wireless sensor network traffic to inhibit traffic analysis attacks," Pervasive and Mobile Computing, vol. 2, no. 2, 2006.

Authors Biography:

First Author: Mr. Mohammed Amair

M.Tech in CSE from JNTUH,
Malla Reddy College of Engineering and Technology,
Hyderabad

Second Author: Mr. Mohd Abdul Wasey Muzakkir

M.Tech in CSE from JNTUH,
Alhabeeb college of Engineering and technology
Chevella R.R Dist

Third Author: Prof. Mr. M. Saidi Reddy M.Tech (PHD)

M.Tech in CSE from JNTUH,
Head of the Dept CS&E,
Malla Reddy College of Engineering and Technology,
Hyderabad

