CMPUT 660 literature review: Network Privacy in Wireless Sensor Networks

Peter Atrazhev

I. INTRODUCTION

The privacy challenges of WSN are caused by three major limitations: energy , computational power and communication overhead. These are all a unique consequence of the nature that wireless sensor networks are deployed using many small devices.

There are two categories of privacy in WSN as described in [1]. These categories are:

- User
- Network

User privacy is concerned with the ability to discern a user's personal identity from data. Network privacy is concerned with the security of information and what an attacker can gain from information that is directly accessed from the network.

The focus of this literature review will be network level privacy concerns. These can be further categorized into two major domains each with there own subsets of attack vulnerabilities. These are:

- Content privacy
- Context privacy

Content privacy concerns have to do with the threat of an attacker being able to gain access to data that is being transfered between nodes. Content privacy has two vulnerabilities:

- Aggregated data
- Query

Aggregated data is a product of the hardware limitations that are present in WSN. It is common that data is aggregated by a node before it its sent to the base station. This is done in order to conserve power and bandwidth in the network. The danger of aggregated data is that it becomes easy to compromise a single node that aggregates data in order to gain access to a large amount of sensor data. Attackers can simply compromise all the aggregating nodes to maximize the amount of information that they get.

The query vulnerability occurs when a network is set up in a query-driven approach, in which the user queries information from nodes that are collecting data. This is common when the user seeks to gain information at a specific time of their choosing, while having the sensors sampling at a higher rate. This type of vulnerability can be exploited to gain information on the physical aspects of the network.

Contextual privacy concerns are concerned with an attacker being able to gain information from the features of the communications that are being sent. This type of attack can be used to infer features about the network such as hardware specifications, network topology and type of data that is being sent. This category can be decomposed into the following vulnerabilities:

- Temporal
- Identity anonymity (node)
- Location

Temporal privacy refers to preventing the attacker from knowing the time of creation of sensor packets arriving at a sink [2]. Identity anonymity refers to the obfuscation of the network topology from an attacker, such that the attacker cannot determine which node is a sink or a source. Location privacy refers to being able to hide where a node is in physical space or to be able to hide the location of an event that generates data.

The purpose of this literature review is to examine the current solutions to the above mentioned vulnerabilities and hopefully be able to determine which one is the most dangerous and why. The next section will give short descriptions of papers describing a well established solution to each of the problems. Following that section 3 will argue which vulnerability is of most concern.

II. PAPER DESCRIPTIONS

This section features short description of papers that present solutions to each of the vulnerabilities that were mentioned in the introduction. Context privacy vulnerabilities will be treated first following with content privacy vulnerabilities.

A. Location

The paper that was chosen to be read on the topic of locational privacy was [3]. This paper makes the case that a global eavesdropper is a more realistic attack model than the local eavesdropper. As the cost of devices is getting cheaper and there are scenarios where setting up a sophisticated attacking network is relatively cheap if the attack works. The authors propose two different techniques to prevent the leakage of information within the global attacker framework. These techniques are:

- periodic collection
- source simulation

In periodic collection, the objective is to make network traffic patterns independent of the presence of real objects. To this end, periodic collection has every sensor node send packets at a reasonable frequency regardless if there is real data to send or not. This however makes the network traffic independent of the real sensor values. To combat this, real data is queued in buffers in the nodes and the buffers are checked for data before transmitting. If a node discovers real data, it encrypts it with a pairwise key and sends it off to the next node. If a node does not have queued data, it makes fake data without a key. A node will reject all communications without keys, this ensures that the buffers do not get filled up with useless noise. Periodic collection is very secure in terms of locational privacy, but it requires a large amount of resources in the network to have all nodes transmit.

In Source stimulation, the objective is to create multiple convincing traces of information in the network to hide the traffic generated by a real object. To accomplish this task, a set of virtual objects (tokens) will be simulated in the field with each one generating a traffic pattern similar to that of a real object. The protocol will have token nodes, nodes simulating virtual objects, randomly pick nodes in its neighborhood to receive the token in the next round. The tokens are designed to mimic the signal used for event detection and trigger the appropriate network traffic as if a real event was detected. There is however a single problem with this approach and that is that the signal to pass the token around from node to node can be used to detect the real object. This is solved by the authors by simply having real detections also send an extra message to a random node in each round.

Security analysis: The authors used the entropy of the system ,b as the measure of privacy for both of the techniques[3]. For periodic collection, the entropy of the system was given by:

$$b = \log_2 \frac{N}{C} \tag{1}$$

where N is the total number of nodes in the network and C is the number of real objects generating data.

For Source stimulation, the entropy of the system is given by:

$$b = \log_2(1 + \frac{LP}{C}) \tag{2}$$

where L is the number of "token nodes" and P is the probability of real data being generated in the round.

B. Temporal Privacy

The paper reviewed for the concept of temporal privacy was [2]. This paper proposes a rate-controlled adaptive delaying mechanism. It works by having the buffer generate a random delay time with an exponential distribution. This is done to obfuscate the time of packet generation. A delay time means that the network will now have to buffer packets and this can cause buffer filling near the source. To deal with this, the authors have proposed to selecting and transmitting specific packets, victim packets, immediately rather than dropping packets when the buffer is full. The paper proposes 4 different buffer preemption policies:

- Longest delayed first
- Longest remaining delay first
- shortest delay time first
- shortest remaining time first

Longest delay features a victim packet that is the packet that has stayed in the buffer the longest. This has the effect of ensuring that there is a non zero delay value for each packet. Longest remaining delay features a victim that is the packet that has the longest remaining delay. This has the effect of lessening the buffer load of the system. Shortest delay features a victim packet with the shortest delay time. Shortest remaining time features a victim packet that has the shortest remaining time. The easiest policies to implement are the longest and shortest remaining time policies as current network architectures already keep this information in their records.

The authors found that the longest remaining delay first was the preemption policy that functioned the best. This was found because the policy had the larges number of preemptions. It was found that a larger number of preemptions tends to alter the original distribution less, making it harder to guess the time of packet creation.

C. Identity

In the paper that was reviewed for the identity vulnerability he authors propose an efficient anonymous communication protocol for sensor networks that can achieve anonymity while having small overheads on computation, storage and communication [3].

Important nodes such as a source node of the base station play critical roles in networks and as such their identities on the network should be carefully protected against attack.

Anonymity includes several different aspects:

- sender anonymity
- receiver anonymity
- Unlinkability between sender and receiver

The authors of the paper propose a communication protocol with the aim of addressing the aspects of anonymity. This communication protocol has four elements to it:

- Anonymous data sending
- Anonymous data forwarding
- Anonymous broadcasting
- Anonymous acknowledgment

Anonymous data sending is accomplished using a global anonymous identities when sending messages.This anonymous identity is randomized for every message.

To forward data anonymously nodes employ a one hop anonymous identity that they use to communicate the data between neighboring nodes during the transmission process. This detaches the destination from the path that the data takes

and makes it more difficult for an attacker to gain information as they need to decrypt 2 different identity values. The node path is determined using a probabilistic forwarding node selection scheme.

Anonymous broadcasting comes in two parts: anonymous broadcast and probabilistic latency-based transmission. Anonymous broadcast, when a neighbor node receives a message it will check that it is supposed to receive said message and if so it will add a random delay to that message before relaying it to all nodes in the entire network. Relaying it to all nodes makes it difficult to determine the intended destination while the added delay makes it difficult to identify when a message was actually received.

Anonymous acknowledgment: Messages also come with anonymous ACK identities that are part of the message. This ACK is propagated along all the nodes that make up part of the path of transmission once the message is received. If the sender node does not receive an ACK for its message, it will send the message again after a certain time period.

Security analysis: In the paper, the authors propose two lemmas regarding the security of their methods and then prove them. These lemmas are:

- It it impossible for a node to compute the encryption key between two neighboring nodes.
- It is very difficult to find the source of a transmission.

The first lemma is valid because a node needs to intercept all communications between all nodes in neighborhood to be able to decrypt the keys of any one of the transmissions. The second lemma is valid because the attackers need Z_1 message decryption and the probability of compromising a source node is $\frac{\delta}{N}$.

$$Z_1 = |N_{comp}|\gamma \tag{3}$$

Where γ is the average number of messages that an attacker captures in a captured node. N_{comp} is the number of compromised nodes in the network.

D. Aggregated Data

This paper presented two privacy preserving data aggregation techniques:

- Cluster based Private Data Aggregation (CPDA)
- Slice-Mix-Aggregate (SMART)

In CPDA, the nodes are formed into random clusters to in which nodes do not know the data of their neighbors. CPDA has two main steps:

1) Formation of clusters
2) Calculation within clusters

In the formation phase, nodes will randomly elect cluster leaders and have other member nodes join clusters that have leaders. In the calculation phase, the nodes in each cluster begin by broadcasting shared seed values (non-zero numbers). Along with random numbers known only to each node, the seeds are used to encrypt the data that is being passed through the cluster. Each node's data is encrypted and send between all nodes. The data is decrypted at the cluster leader. The cluster leader data is then aggregated using the Tiny AGgrgation protocol, which uses a routing tree.

In SMART, the data is sliced into pieces and then the slices are sent to aggregation nodes before they eventually get to the sink. The SMART algorithm consists of three steps:

1) Slicing
2) Mixing
3) Aggregation

Slicing involves a node slicing its data into a randomly selected value that is within the set of all possible nodes. A single slice of data is kept at the initial node and the rest of the slices are sent to all nodes that were part of the randomly selected set. The second step consists of each node summing up their received slices. The third step involves all nodes aggregating their data, sending it to the query node. The aggregation is designed using a tree-based routing protocol.

Security analysis:

- CPDA: A node can only crack the private data of a cluster member if it knows the all m-1 keys of a given member.
- SMART: An eavesdropper will crack the private data held by a node s if it breaks all $J-1$ outgoing links and all incoming links of s, where J is the number of slices that have been generated.

E. Query

Data aggregation technique that protects the data from nodes using a rotational scheme. Also reducing Communication overhead

This algorithm is proposed to relay data to sink nodes upon a query request being relayed to the sensor nodes. The algorithm consists of three different phases:

1) Cluster formation
2) intra-cluster rotation
3) inter-cluster aggregation

A cluster consists of a head node and member nodes. In the cluster formation phase there must be at least 3 nodes in each cluster for the algorithm to work properly.

Intra-cluster rotation: Once the clusters have been established the data is passed between all the member nodes on a path that passes by each node at least once. The data is encrypted in this phase as well in order to enhance privacy. The path is determined by a greedy algorithm to conserve time and energy in the network.

Inter-cluster aggregation: After the data has passed through each node in the cluster, the head node sends the data to its parent node, clusters are aggregated together and the intra-cluster rotation repeats until the data gets to its destination.

The authors presented the probability of having a packet overheard and the probability of the network resisting a collusion attack as their security analysis. They found that the

probability that the original data of a head node is disclosed is given by:

$$P'_{overhear} = \frac{1}{\beta}\frac{k}{\alpha} \tag{4}$$

where k is the number of keys from the key pool that have been picked, α is the total number of keys in the pool and β is the number of generated random numbers.

III. Discussion

All of the vulnerabilities that have been discussed in the previous section are of concern for any WSN that will be deployed in a real world application. However there is a hierarchy of danger to these vulnerabilities that is based on the destructive power that an attacker would gain if they compromised the vulnerability. This is as follows:

1) Aggregated data
2) Identity
3) Location
4) Query
5) Temporal

At the top of the list is aggregated data. When an attacker compromises your data encryption and gets access to your data, then they can begin user attacks. Next on the list is network identity. Once an attacker has your network topology, they can target sink nodes and cause lots of other trouble for network maintenance. The result of a network identity hack is that the attacker can now focus their siege on the key points in the network. The third item on the list is a location hack. Once a attacker knows where your nodes are located physically, they can make opportunities to physically interact with the hardware which can be much easier to penetrate any defenses.

To identify which of the vulnerabilities carries with it the most risk, we look to the privacy analysis of the top three items on the list.

The CPDA and SMART algorithm privacy scales with the number of nodes in the network as an attacker needs to have access to data that is being sent to all the nodes in the network in order to break into a single node. The lemmas of the Identity algorithm almost guarantee that the identity of the nodes is safe. The location algorithm privacy differs slightly between the two techniques but both scale with the size of the network.

After considering all of the techniques that were presented, I propose that the most threat is from the methods that scale with the size of the network. This is because if an attacker managed to be able to infiltrate nodes in the network and then shut a vast majority of the network down for a small period of time, the complexity of the privacy system would suddenly be reduced incredibly. This could be used to gain access into more nodes and then repeated until the attacker has control of all the nodes in the network. The assumption that all networks will be large is also not a valid one for practical reasons. Most everyday consumers lack the resources to set up a large network and most certainly lack the knowledge to do so properly. Since the largest Privacy concern is the privacy of the everyday user, any technique that relies on network size will not protect conventional technology users.

IV. Conclusion

Wireless sensor networks are becoming a common paradigm for Iot applications. They bring with them much promise and many privacy concerns and vulnerabilities. Privacy is a realm that is of great concern with the coming of so many interconnected devices all generating data together. This literature review showcased some of the major areas of concern and the current methods of dealing with these vulnerabilities. This literature review has highlighted some of the problems that have been identified and have proposed solutions to them. Following that, it is suggested that the major concern for privacy in a wireless sensor network is any method that relies on the size of the network to assure its privacy.

References

[1] J. Lopez, R. Rios, F. Bao, G. Wang, *Evolving privacy: From sensors to the internet of things*. Future Generation Computer Systems 2017, Vol 75, pp 46-57.

[2] P.Kamat, W. Xu, W. Trappe, Y.Zhang, *Temporal Privacy in Wireless Sensor Networks: Theory and Practice*. ACM transactions on Sensor Networks, November 2009, Vol 5, No. 4, Article 28.

[3] K. Mehta D. Liu, M. Wright, *Location Privacy in Sensor Networks Against a Global Eavesdropper*. IEEE ICNP 2007 publications pp.314-323

[4] X. Zhang, H. Chen, *Rotation-based Privacy-preserving Data Aggregation in Wireless Sensor Networks*. IEEE ICC 2014 proceedings pp.4184-4189

[5] W. He, X. Liu, *PDA: Privacy-preserving Data Aggergation in Wireless Sensor Networks*. IEEE INFOCOM 2007 proceedings pp. 2045-2053.

[6] J. Chen, X. Du, B. Fang, *An Efficient Anonymous Communication Protocol for Wireless Sensor Networks*. Wireless Communications and Mobile Computing, 2011, Published online.