SECURE DATA DISSEMINATION PROTOCOL IN WIRELESS SENSOR NETWORKS USING XOR NETWORK CODING

Jisha Mary Jose[1], Jomina John[2]

[1]*Dept. of CSE, Rajagiri School of Engineering and Technology, Kochi, India*
[2]*Dept. of CSE, Rajagiri School of Engineering and Technology, Kochi, India*

Abstract

Wireless sensor networks (WSN) are basically distributed networks or a collection of sensor nodes which collect information which are used to analyse physical or environmental conditions. WSNs are usually setup in remote and hostile areas and work in extreme conditions. Applications of WSN include habitat monitoring, industrial applications, battlefield surveillance, smart homes etc. Most of them require regular updating of software in sensor nodes through the wireless channel for efficient management and working. So it is necessary to spread data through the wireless medium after the nodes are deployed. This is known as data dissemination or network reprogramming. A good data dissemination protocol must be fast, secure, reliable and energy efficient. To achieve these we can make use of network coding techniques which reduces the number of retransmissions due to any packet drops. But network coding increases the chance of various kinds of network attacks. Also to avoid spreading of malicious code in the network, each sensor node has to authenticate its received code before propagating it further. So here a novel dissemination protocol is introduced based on simple cryptographic techniques which prevents pollution and DoS attacks and at the same time achieves fastness using the technique of network coding.

Keywords: *wireless sensor network; dissemination; reprogramming; network coding; security; pollution attack.*

----------***----------

1. INTRODUCTION

Wireless Sensor Networks (WSN) is one of the major milestones in the field of communication. These networked collection of nodes take us a step closer to obtaining valuable information about the physical world. WSN are used popularly in many applications like remote control and monitoring, construction safety systems, environmental monitoring, health care management, disaster management, surveillance operations, smart homes, habitat monitoring, indoor sensor networks, seismic monitoring of buildings etc [1]. In computer science and communication wireless sensor networks entertain lot of research today.

A WSN is made of sensor nodes used for monitoring and analysis purposes as shown in Fig 1. These sensor nodes pass the information that they collect to a prime location called a base station. In most systems, a WSN communicates with a LAN or WAN through a gateway like medium. The gateway is actually a bridge between the WSN and the various other networks [2]. This allows data to be stored by devices and which can be taken up for processing later. Each sensor node or mote has several parts: a circuit for interfacing with other sensor nodes, a micro controller, a radio transceiver, and a battery for power supply. The topology used can be either a star, ring, grid network or multi-hop wireless mesh network.WSN is used primarily in remote and hostile environments for information collection. Hence it is a major challenge to produce cheap sensor nodes. They must be designed carefully by considering all the different constraints of the environment in consideration.

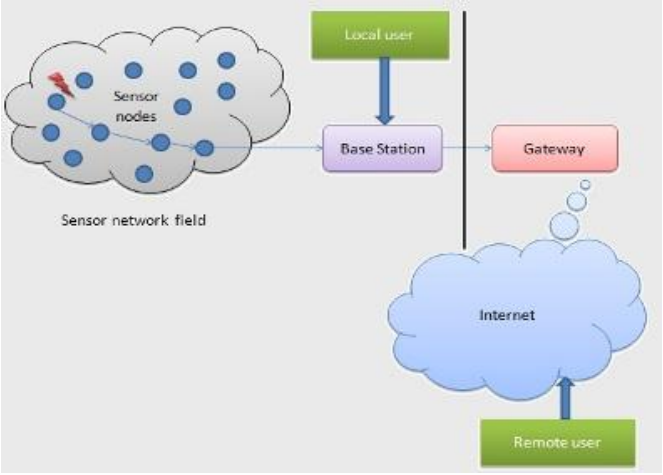

Fig -1: An example wireless sensor network

Wireless sensor networks must be operated for long duration of time and usually don't get any human administration or intervention in between [3]. Evolving conditions and environments can also; cause changes in application features, which hence lead to the need to change the network behaviour by introducing new code or software. But the remote nature of WSN is a disadvantage here. It will require the propagation of new code updates over the wireless medium i.e. over the air as manual updating of such networks will not be possible. This process is known as dissemination or network reprogramming.

But dissemination brings in a lot of challenges [3]. One major challenge is proper and complete dissemination of information to all sensor nodes in the network. This is

difficult since the number of nodes in the network can be huge and the environment is always dynamic, thus the basic topology keeps changing constantly. Secondly the information to be disseminated may be produced at a single node, such as the prime source i.e. the base station, or at the sensor nodes themselves. Thirdly data must be disseminated in a secure way or else adversaries can track out critical data. Also there is a possibility of attackers sending bogus data into the network which must not be received by the sensor nodes as they can cause different attacks like pollution attacks, denial of service attacks and so on. So dissemination of code or program data in wireless sensor networks is an area to be worked in deeply and new techniques need to be introduced to achieve tradeoffs between energy and speed in dissemination. The aim of this work is to develop a novel secure and fast data dissemination protocol for use in wireless sensor networks. This work concentrates on developing a dissemination protocol for dissemination of small data.

Linear network coding is a technique used to achieve fastness and energy efficiency during dissemination [4]. It is a technique that combines packets in the network; increasing the throughput, decreasing energy consumption, and reducing the number of messages transmitted. In traditional systems dropped packets are recovered using retransmissions. But in network coding we can combine packets using mathematical operations and then disseminate so that recovery of lost packets can be achieved without retransmission.

But network coding along with its energy efficiency advantages brings in a lot of headaches. It is highly prone to attacks like pollution, denial-of-service attacks and many others. So to deal with these the proposed system uses simple but efficient cryptographic techniques for data dissemination. This makes sure that we can achieve simple yet secure data dissemination in wireless sensor networks.

Our work is organized as follows. First we focus on the need of data dissemination in wireless sensor networks and some of its related works. Next the design and implementation which concentrates on the dissemination of small values and variables is explained. Then we study the performance of the new protocol through extensive simulation using TinyOS and finally have the conclusion and references.

2. RELATED WORKS

Data dissemination in wireless sensor networks is a critical and vital task. It is based on the concept of traditional communication system, where we have a sender and receiver. The scenario is basically a sender sending out some information, and receiver collecting the information sent, processing it and sending some information back. While in data dissemination, only half of this concept is applied. Some information is sent out and received at the destination, but no reply is given back. The sender sends out information, not to one node, but to many as in a broadcasting system.

Dissemination is used to send code updates or program images to the sensor nodes periodically so as to perform reprogramming of the nodes. This over the air act is required since manual updating of sensor nodes deployed in remote environments is next to impossible in most of the cases. The main aim of a dissemination protocol in WSN is to ensure that all the sensor nodes have consistent data with them always.

There are two kinds of dissemination in WSN [5]:

- Code dissemination - to send program images which are generally bulky data. Usually they are divided into fixed sized pages and packets and then disseminated.
- Data discovery and dissemination - to disseminate small configuration parameters, variables, queries, commands etc in packets.

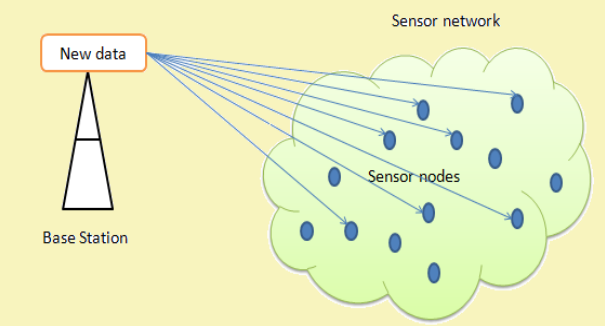

Fig -2: Dissemination process in WSN

2.1 Small Value Dissemination

This work concentrates on data discovery and dissemination protocols i.e. dissemination of small values like variables, parameters and so on. Figure 2 gives a general idea about data dissemination. Traditional protocols available for this include Drip, DIP and DHV. They are all based on Trickle algorithm [6].

Drip proposed by Tolle et. al [7] is the simplest of all dissemination protocols and is based on Trickle algorithm and establishes an independent trickle for each variable in the data. Every time an application wants to transmit a message, a new version number is generated and used. This will cause the protocol to reset the Trickle timer and thus disseminate the new value else the trickle timer interval is incremented.

DIP (DIssemination Protocol) [8] is a data detection and dissemination protocol proposed by Lin et al. It is a protocol based on the Trickle algorithm. It works in two parts: determining whether there a difference in data stored at a node, and then determining which data is different. It is based on the concept of version number and key tuple for each data item. DIP calculates hashes that cover all version numbers of the data. Nodes that receive hashes same as their own know that they have consistent data with respect to their nearest neighbours. If a node gets a hash that differs from its own hash, it knows that a difference exists in the data.

DHV (Difference detection-Horizontal search-Vertical search), [9] is a code consistency maintenance protocol given by Dang et al. It tries to keep codes on different nodes in a WSN consistent and up to date. Here also data items are represented as tuples (key, version). It is based on the observation that if two data items are different, they will only differ in a few least significant bits (LSB) of their version number rather than in all their bits. So only those bits need to be checked. For this steps followed are detection and identification.

There are many code dissemination protocols like Deluge [10] as well. They are used to disseminate large code updates into the network. For this the code is often broken down into pages and then further into packets. Here we have seen some basic data discovery and dissemination protocols. They don't support any techniques which help to reduce packet retransmissions. Also none of these protocols provide security to the data disseminated.

2.2 Network Coding and Data Dissemination

Network coding [4] aims to replace the traditional store and forward technique used in networks; by better routing algorithms that will allow intermediate nodes to transform the moving data. Network coding has become popular due to its properties like robustness and better throughput. It helps to achieve fast data dissemination as it reduces the number of retransmissions that will be needed if there are packet losses. Many dissemination protocols have been developed using the concept of network coding [14] [15].

The advantages of network coding based dissemination protocols are that they achieve energy savings and communication efficiency, especially during increased packet loss or network density. So network coding based protocols can be profitable for reprogramming of WSNs. However we face a potential problem in hostile environments. An adversary may launch pollution attacks, in which a malicious node sends bad encoded packets that consist of bogus data, which leads to incorrect decoding of the original data upon retrieval.

Here we use binary network coding i.e. the mathematical operation used is XOR to combine the contents of packets [4]. Only two packet network coding is done here. Also here we focus on dissemination of small values like configuration parameters, variables, queries, commands etc whose size range from 2-4 bytes and thus is modification of the existing DRIP protocol.

3. ASSUMPTIONS AND THREAT MODEL

3.1 Assumptions

Here assume that the source of the reprogramming variables, i.e., the base station, is a secure location. Also each sensor node has a unique identification number. We assume that while each sensor node is resource limited, it has sufficient memory to store all the security mechanisms of the protocol.

3.2 Threat Model

Here assume that the individual sensor-nodes are unprotected. An adversary may insert its own attacker nodes into the network, or it may capture other nodes. The adversary can attempt to launch pollution attacks to corrupt the data in the network and also to consume the limited resources on sensor nodes.

4. PROPOSED SYSTEM

In this protocol data dissemination is done in a secure and fast way by using the techniques of network coding and cryptography. Network coding reduces the number of retransmissions due to any packet losses happening in the network by combing and sending data. Also data disseminated is always sent as encrypted data. For this nodes first perform node to node authentication and establish session keys. Then the session key is used for encrypted transfer of data. This protocol ensures that the system is free of pollution [13] and Denial-of-Service attacks. The different phases of this protocol include:

4.1 System Initialization Phase

This phase is done before the WSN is deployed in the application field. In this phase the base station generates a master key K_m and a unique random number R_m and stores them safely in each node. Also a list of all the valid node ids is maintained in each node.

4.2 Packet Processing Phase

In this phase the actual data dissemination occurs. Before disseminating data a node will generate a real time key using a key generation algorithm. This includes the generation of two unique random numbers R_1_node and R_2_node. Key generation is done using Trivium-Multilinear Modular Hashing (MMH) as the MAC function and SHA1 as hashing function H(x). The steps are:

1. MAC[i] = R_1_node XOR K[i] (1)
2. a[i] = node_id + MAC[i] (2)
3. h = MMH (a[i]) (3)
4. Key = H (h XOR R_2_node) (4)

Where K[i] is the master key of the MAC function, node_id is the identifier of the corresponding node, XOR is the logical XOR operation. This real time key is broadcast by the node in a packet which will include the node_id and the key. The destination node who receives it will check the node_id with its list of valid nodes and ensure this packet is coming from a valid node. If yes that node will also generate a real time key using the same process as above and send back a reply packet to the sender node which will contain the node_id and the newly generated key.

If this packet is also validated, then the two nodes are ready to generate a session key. The key is generated as:

$$\text{Session key} = K_m \text{ XOR } K_a \text{ XOR } K_b \quad (5)$$

Where K_a and K_b are keys generated at any two nodes A and B. Now this key is used for encrypting the data to be disseminated. The advantage of this scheme is that there is no need of actual exchange of the session key through the network. To encrypt the data we use symmetric encryption techniques preferably Advanced Encryption Standard (AES). So the data packet disseminated from a node will contain the data in encrypted form i.e.

$$\text{Data} = E(d)_{sk} \text{ , where sk is the session key.} \quad (6)$$

Dissemination in wireless sensor networks works on the basis of Trickle algorithm [2]. It takes on the concept of gossiping. Whenever a new data is to be disseminated the trickle timer is reset to 0 and the data is broadcasted. When a node receives a new data it will store it. But if it receives a data which it already is aware of then it will increase the trickle timer interval and suppresses the duplicate incoming data.

To achieve immediate authentication of data packets a onetime hash of the initially generated random number is also calculated and included in each packet. The steps are:

1. Calculate Hash = $H(R_m)$ (7)

2. Result = ADD(Hash) (8)

Where H( ) is SHA-1 and ADD( ) is basic addition operation. The result is included in the packets sent.

4.3 Packet Verification Phase

To achieve immediate authentication of the received packet, the destination node will calculate the hash of R_m stored in its memory and compare it with the value in the received packet. If they match, then the received packet is a valid node. Thus it will be acknowledged ACK by the destination. Otherwise a NACK (negative ack) is sent to the sender.

Next we will have to ensure the integrity of the data. For this first the node checks the id in the received packet. If it is a valid node_id, then it will attempt to decrypt the data using the session key already generated and stored. Every node has an original data and combined data buffer. So the node will check whether it is an original data or combined data. If it is an original data it will be stored and disseminated after a trickle timer fire and if it is a combined data, the node will check whether it is possible to extract any other data from this newly received data using network coding. After that the data will be stored or disseminated out.

So likewise all the data disseminated from the original source node will be distributed to all the nodes and a round of dissemination will be completed. This technique thus makes sure that only valid data is sent out and data is been sent out safely.

5. IMPLEMENTATION AND RESULTS

This protocol has been implemented in TinyOS-2.1.2 simulator TOSSIM [17]. We have considered a network topology consisting of 100 nodes and 25 different data variables are disseminated. The packet size in TinyOS [18] is 29bytes. The sensor node considered for simulation here is micaz.

Cryptographic support has been achieved using hashing algorithms like SHA-1 which generates a 160 bit hash value, MAC functions like Trivium Multilinear Modular Hashing (MMH), and symmetric encryption algorithms like AES which uses a 128 bit key.

The new protocol is found to resist cases of pollution attacks i.e. only valid data packets are received and processed by the intermediate nodes in the network. Also immediate authentication of packets is achieved using the one time hash value generated and stored in the data packets disseminated.

6. SECURITY AND PERFORMANCE ANALYSIS

First we perform and analyze the security offered by this protocol.

- Resistance to pollution attacks- Attackers can't pollute the network with bogus data since data transfer done is always verified using cryptographic techniques.
- Resistance to Denial-of-Service attacks- Immediate authentication of packets is done at each destination, so bogus packets can be discarded and only valid packets pass through.
- Session key agreement- Session keys are used for encryption and decryption. Also this key is locally generated and used, hence not exchanged in the network.
- Real time key generation- No-pre stored keys in nodes; they are calculated at time of data transfer only.
- Light-weight- Only simple yet good mathematical operations and encryptions techniques are used hence no much resource usage in nodes.

Chart 1 gives a comparison graph on the number of data messages disseminated in each protocol namely DRIP [7], CodeDrip [11] and the new proposed protocol. Network coding has reduced total number of messages.

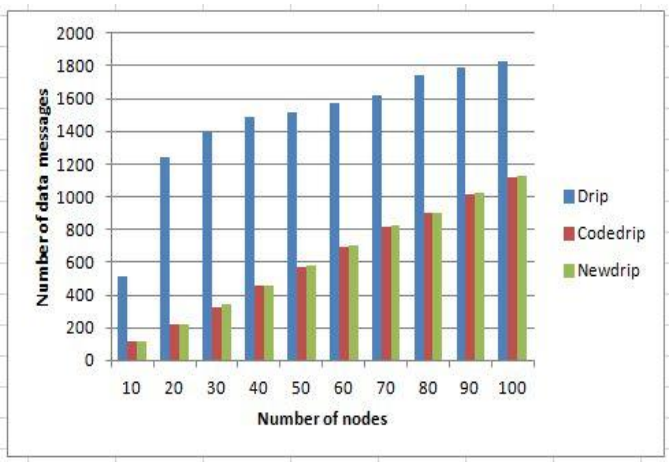

Chart 1: Comparison of data messages disseminated

7. CONCLUSION

This paper proposes a novel data discovery and dissemination protocol for wireless sensor networks which can be used to achieve secure and fast data dissemination especially for small configuration parameters and variables. This technique combines the concepts of network coding and simple cryptographic techniques so as to disseminate data. The advantages of this protocol are that it is resistant to pollution attacks, and achieves immediate authentication of data been disseminated. Session keys are used to encrypt and send data between nodes and there is no need of actual transfer of the session keys through the network. Also only simple mathematical operations are used to calculate keys for encryption of data so not much of resource usage at the nodes. All together it aims to provide a simple yet secure and fast data dissemination protocol for usage in wireless sensor networks. Node compromise by an attacker can be an issue in this protocol. It will be dealt with as part of the future works.

REFERENCES

[1] Mohammad A. Matin, Wireless Sensor Networks: Technology and Protocols: Published by InTech, Croatia, ISBN 978-953-51-0735-4, 2012.

[2] Salvatore La Malfa, Wireless Sensor Networks, 2010.

[3] Jisha Mary Jose, Jomina John, "Data dissemination protocols in wireless sensor networks-a survey", IJARCCE, March 2014.

[4] T. Ho and D. Lun. Network Coding: An Introduction. Cambridge University Press, 2008.

[5] Daojing He, Sammy Chan, Shaohua Tang and Mohsen Guizani, "Secure Data Discovery and Dissemination based on Hash Tree for Wireless Sensor Networks", IEEE transactions on wireless communications, Vol. 12, No. 9, September 2013.

[6] P. Levis, N. Patel, D. Culler and S. Shenker, "Trickle: a self regulating algorithm for code maintenance and propagation in wireless sensor networks", in Proc. 2004 NSDI, pp. 15-28.

[7] G. Tolle and D. Culler, "Design of an application-cooperative management system for wireless sensor networks," in Proc. EWSN, pp. 121–132, 2005.

[8] Lin, K., Levis, P.: "Data discovery and dissemination with dip." In: Proceedings of the 2008 International Conference on Information Processing in Sensor Networks (IPSN 2008), Washington, DC, USA, IEEE Computer Society (2008) 433-444.

[9] T. Dang, N. Bulusu, W. Feng, and S. Park, "DHV: a code consistency maintenance protocol for multi-hop wireless sensor networks", in Proc. 2009 EWSN, pp. 327-342.

[10] Hui, J.W., Culler, D.: "The dynamic behaviour of a data dissemination protocol for network programming at scale." In: Proceedings of the 2nd international conference on Embedded networked sensor systems (Sensys 04), New York, NY, USA, ACM (2004) 81-94.

[11] Nildo dos Santos Ribeiro Junior, Marcos A. M. Vieira1, Luiz F. M. Vieira1 and Om Gnawali, "CodeDrip: Data Dissemination Protocol with Network Coding for Wireless Sensor Networks", In Proceedings of the 11th European conference on Wireless sensor networks (EWSN 2014), Feb. 2014.

[12] Hailun Tan, "Secure multi-hop network programming with multiple one-way key chains", In: Proceedings of the International conference on Embedded networked sensor systems (Sensys 07), Sydney, Australia, ACM.

[13] Yingpei Zeng, Jiannong Cao, Shigeng Zhang, Shanqing Guo, Li Xie, "Pollution Attack: A New Attack Against Localization in Wireless Sensor Networks", IEEE, WCNC-2009.

[14] I-Hong Hou, Yu-En Tsai, T.F. Abdelzaher, and I. Gupta. Adapcode: Adaptive network coding for code updates in wireless sensor networks. In INFOCOM 2008.The 27th Conference on Computer Communications. IEEE, pages 1517–1525, 2008.

[15] Andrew Hagedorn, David Starobinski, and Ari Trachtenberg. Rateless deluge: Over-the-air programming of wireless sensor networks using random linear codes. In Proceedings of the 7th international conference on Information processing in sensor networks, IPSN '08, pages 457–466, Washington, DC, USA, 2008. IEEE Computer Society.

[16] S. Katti, H. Rahul,W. Hu, D. Katabi, M.Medard, and J. Crowcroft, "XORs in the air: practical wireless network coding", ACM SIGCOMM Computer Communication Review, vol. 36, no. 4, pp. 243-25, 2006.

[17] Philip Levis, Nelson Lee, Matt Welsh, and David Culler. "Tossim: accurate and scalable simulation of entire tinyos applications, In SenSys '03, pages 126-137, New York, NY, USA, 2003. ACM Press.

[18] TinyOS: an open-source OS for the networked sensor regime. Available: http://www.tinyos.net/.

[19] Simulating a Wireless Sensor Network, 2010-2013, [Online] Available: http://virtual-labs. ac.in/ cse28/ ant/ ant/8/theory/.