Data Base Security using PGP and ID3

Atul Mishra
M.Tech (ECE)
LPU, Jalandhar
atulmishra005@gmail.com

***Abstract*— In the present scenario, we can understand the need of database security in Wireless Sensor Network (WSN). Database is the most essential part of the network. Wireless Sensor Networks (WSN) are a most challenging and emerging technology for the research due to their vital scope in the field coupled with their low processing power and associated low energy. Today wireless sensor networks are broadly used in environmental control, surveillance tasks, monitoring, tracking and controlling etc. On the top of all this the wireless sensor networks need very secure communication in wake of them being in open field and being based on broadcasting technology. Security of data is the main requirement of any organization. We have proposed an idea of using wireless sensor network for the security of database. For this purpose two algorithms are used PGP (Pretty Good Privacy algorithm) and ID3 (Iterative Dichotomiser 3 algorithm) which will help for security as well as speed of data. ID3 (Iterative Dichotomiser 3) algorithm arranges all the nodes in the form of tree and if it found any unauthorized node accessing then it will block. On the other hand PGP is used to encrypt the data and send data in the form of encrypted data. Now the node that has decrypt key only that can access the data unauthorized can't read it.**

***Key Words*—Wireless Sensor Network, Database Security, Pretty Good Privacy (PGP) algorithm, and Iterative Dichotomiser 3 algorithm.**

I. Introduction

WSN deal with real world environments, in many cases sensor data must be delivered attempt to process as fast as possible and hope that this speed is sufficient to meet deadlines. Some initial results exist for real-time routing. For example, the RAP protocol proposes a new policy called velocity monotonic scheduling. Here a packet has a deadline and a distance to travel. Using these parameters a packet's average velocity requirement is computed and at each hop packets are scheduled for transmission based on the highest velocity requirement of any packets at this node. While this protocol addresses real-time, no guarantees are given, another routing protocol that addresses real-time are called SPEED. This protocol uses feedback control to guarantee that each node maintains an average delay for packets transiting a node. Given this delay and the distance to travel (in hops), it can be determined if a packet meets its deadline (in steady state). However, transient behavior, message losses, congestion, noise and other problems cause these guarantees to be limited. To date, the limited results that have appeared for WSN regarding real-time issues has been in routing. Many other functions must also meet real-time constraints including: data fusion, data transmission, target and event detection and classification, query processing, and security. New results are needed to guarantee soft real-time requirements and that deal with the realities of WSN such as lost messages, noise and congestion. Using feedback control to address both steady state and transient behavior seems to hold promise. Dealing with real-time usually identifies the need for differentiated services, e.g., routing solutions need to support different classes of traffic; guarantees for the important traffic and less support for unimportant traffic. It is important not only to develop real-time protocols for WSN, but associated analysis techniques must also be developed.

II. Security Requirements

A sensor network is a special type of network. It share some commonalities with a typical computer network. Therefore we can think of the requirements of a wireless sensor network as encompassing both the typical network requirements and the unique requirements suited solely to wireless sensor networks.

A. Data Confidentiality

Data Confidentiality is the most important issue in network security. Every network with any security focus will typically address this problem first. In sensor networks, the confidentiality relates to the following:

- A sensor network should not task sensor readings to its neighbors. Especially in a military application, the data stored in sensor node may be highly sensitive.
- In many applications nodes communicate highly sensitive data, e.g., key distribution; therefore it is extremely important to build a secure channel in a wireless sensor network.
- Public sensor information, such as sensor identities and public keys, should also be encrypted to some extent to protect against traffic analysis attacks.

The standard approach for keeping sensitive data secret is to encrypt the data with a secret key that only intended receivers possess, thus achieving confidentiality.

B. Data Integrity

With the implementation of confidentiality, an adversary may be unable to steal information. However this doesn't mean the data is safe. The adversary can change the data, so as to send the sensor network into disarray. For example, a malicious node may add some fragments or manipulate the data within a pocket. This new packet can then be sent to the original receiver. Data loss or damage can even occur without the presence of a malicious node due to the harsh communication network. Thus data integrity ensures that any received data has not been altered in transit.

C. Data Freshness

Even if confidentiality and data integrity are assured, we also need to ensure the freshness of each message. Informally, data freshness suggests that the data is recent, and it ensures that no old messages have been replayed. This requirement is especially important when there are shared key strategies employed in the design. Typically shared keys need to be changed over time. However, it takes time for new shared keys to be propagated to the entire network. In this case, it is easy for the adversary to use a replay attack; Also, it is easy to disrupt the normal work of the sensor, if the sensor is unaware of the new key change time. To solve this problem a nonce or another time related counter can be added into the packet to ensure data freshness.

D. Availability

Adjusting the traditional encryption algorithms to fit within the wireless sensor network is not free, and will introduce some extra costs. Some approaches try to make use of additional communication to achieve the same goal. What's more, some approaches force strict limitations on the data access or propose an unsuitable scheme in order to simplify the algorithm. But all these approaches weaken the availability of a sensor and sensor network for the following reasons:

- Additional computation consumes additional energy. If no more energy exists, the data will no longer be available.
- Additional communication also consumes more energy. What's more as communication increases so too does the chance of incurring a communication conflict.
- A single point failure will be introduced if using the central point scheme. This greatly threatens the availability of the network.

The requirement of security not only affects the operation of the network, but also is highly important in maintaining the availability of the whole network.

E. Self-Organization

A wireless sensor network is a typically an ad hoc network, which requires every sensor node be independent and flexible enough to be self organizing and self healing according to different situations. There is no fixed infrastructure available for the purpose of network management in a sensor network. This inherent feature brings a great challenge to wireless sensor network security as well. For example, the dynamics of the whole network inhibits the idea of pre-installation of a shared key between the base stations and all sensors. Several random key predistribution schemes have been proposed in the context of symmetric encryption techniques. In the context of applying public key cryptography techniques in sensor networks, an efficient mechanism for public key distribution is necessary as well. In the same way that distributed sensor networks must self organize to support multihop routing, they must also self organize to support multihop routing, they must also self organize to conduct key management and building trust relation among sensors. If self organization is lacking in a sensor network, the damage resulting from an attack or even the hazardous environment may be devastating.

F. Time Synchronization

Most sensor network applications rely on some form of time synchronization. In order to conserve power, an individual sensor's radio may be turned off for periods of time. Furthermore, sensors may wish to compute the end-to-end delay packet as it travels between two pair wise sensors. A more collaborative sensor network may require group synchronization for tracking applications etc.

G. Secure Localization

Often the utility of a sensor network will rely on its ability to accurately and automatically locate each sensor in the network. A sensor network designed to locate faults will need accurate location information in order to pinpoint the location of a fault. Unfortunately, an attacker can easily manipulate non secured location information by reporting false signal strengths, replaying signals, etc.

H. Availability

An adversary is not just limited to modifying the data packet. It can change the whole packet stream by injecting additional packets. So the receiver needs to ensure that the data used in any decision making process originates from the correct source. On the other hand, when constructing the sensor network, authentication is necessary for many administrative tasks. From the above, we can see that message authentication is important for many applications in sensor networks. Informally, data authentication allows a receiver to verify that the data really is sent by the claimed sender. In the case of two party communication data authentication can be achieved through a purely symmetric mechanism: the sender and the receiver share a secret key to compute the message authentication code of all communicated data.

III. Data Base Security Using PGP

Pretty Good Privacy (PGP) is a data encryption and decryption computer program that provides cryptographic privacy and authentication for data communication. PGP is often used for signing, encrypting and decrypting texts, E-mails, files, directories and whole disk partitions to increase the security of e-mail communications. It was created by Phil Zimmermann in 1991.PGP and similar products follow the OpenPGP standard (RFC 4880) for encrypting and decrypting data.

A. Design

PGP encryption uses a serial combination of hashing, data compression, symmetric-key cryptography, and, finally, public-key cryptography; each step uses one of several supported algorithms. Each public key is bound to a user name and/or an e-mail address. The first version of this system was generally known as a web of trust to contrast with the X.509 system, which uses a hierarchical approach based on certificate authority and which was added to PGP implementations later. Current versions of PGP encryption include both options through an automated key management server.

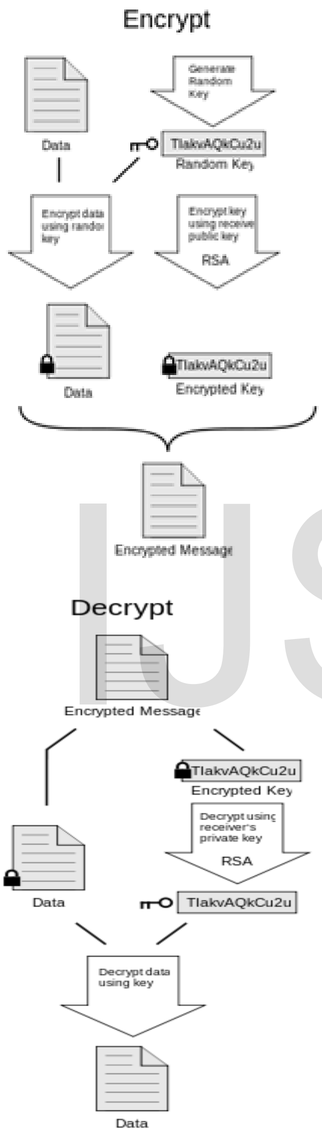

Figure 2: Design for Encrypt and Decrypt.

B. Compatibility

As PGP evolves, PGP systems that support newer features and algorithms are able to create encrypted messages that older PGP systems cannot decrypt, even with a valid private key. Thus, it is essential that partners in PGP communication understand each other's capabilities or at least agree on PGP settings.

C. Confidentiality

PGP can be used to send messages confidentially. For this, PGP combines symmetric-key encryption and public-key encryption. The message is encrypted using a symmetric encryption algorithm, which requires a symmetric key. Each symmetric key is used only once and is also called a session key. The session key is protected by encrypting it with the receiver's public key thus ensuring that only the receiver can decrypt the session key. The encrypted message along with the encrypted session key is sent to the receiver.

D. Digital Signatures

PGP supports message authentication and integrity checking. The latter is used to detect whether a message has been altered since it was completed (the message integrity property), and the former to determine whether it was actually sent by the person/entity claimed to be the sender (a digital signature). In PGP, these are used by default in conjunction with encryption, but can be applied to the plaintext as well. The sender uses PGP to create a digital signature for the message with either the RSA or DSA signature algorithms. To do so, PGP computes a hash (also called a message digest) from the plaintext, and then creates the digital signature from that hash using the sender's private key.

IV. Data Base Security Using ID3

ID3 is a simple decision tree algorithm. A mathematical algorithm for building the decision tree. It builds tree based on the information (information gain) obtained from the training instances and then uses the same to classify the test data. Builds the tree from the top down with no back tracking. A decision tree is defined as a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision. It is commonly used for gaining information for the purpose of decision –making and practical methods for inductive inference. It starts with a root node on which it is for users to take actions. It is a method for approximating discrete-valued target functions, in which the learned function is represented in the form of decision tree. A decision tree is important because it classify correctly new cases. Thus when building classification models one should have both training data to build the model and test data to verify how well it actually works.

The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node. In order to select the attribute that is most useful for classifying a given sets, we introduce a metric information gain.

The advantages of using ID3 is Understandable prediction rules are created from the training data, Builds the fastest tree. Builds a short tree, Only need to test enough attributes until all data is classified, Finding leaf nodes enables test data to be pruned, reducing number of tests and whole dataset is searched to create tree. The disadvantages of using ID3 is Data may be over-fitted or over-classified, if a small sample is tested, Only one attribute at a time is tested for making a decision, Classifying continuous data may be computationally expensive, as many trees must be generated to see where to break the continuum**.**

ID3 algorithm is best suited for: -

- Instance is represented as attribute-value pairs.
- Target function has discrete output values.
- Attribute values should be nominal.

V. Experimental Results

Database security is applied by two algorithms i.e. PGP (pretty Good Privacy) algorithm and ID3 (Decision tree) algorithm. Wireless Sensor Network sense the authorization of nodes. ID3 algorithm arranges all the nodes in the form of tree and if it found any unauthorized node accessing then it will block. On the other hand PGP is used to encrypt the data and send data in the form of encrypted data. Now the node that has decrypt key only that can access the data unauthorized can't read it. The speed of ID3 is less then PGP because the work of ID3 is to check the authorization of each and every node on the other hand work of PGP is to encrypt the data send from one node to another node.

We conclude that in Figure 3 the data variations of our database with respect to error rate. It shows that error rate of Pretty Good Privacy (PGP) is more than ID3 algorithm in this graph PGP shows in pink color and ID3 in blue color and original data shows in red color. And Figure 4 and Figure 5 show the packet delivery ratio of our data base w.r.t error rate and utilizing key size w.r.t data variation respectively.

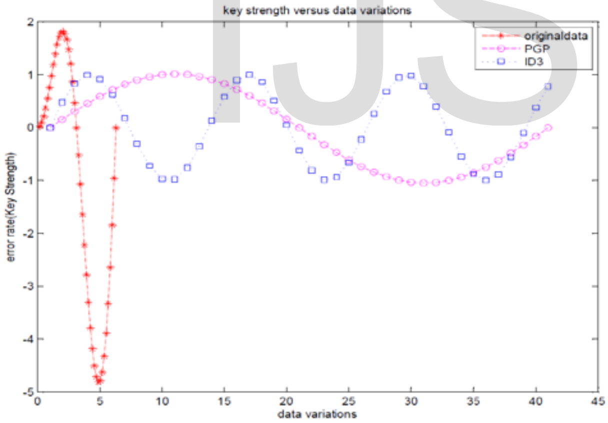

Figure 3: Data variations of our data base with respect to error rate.

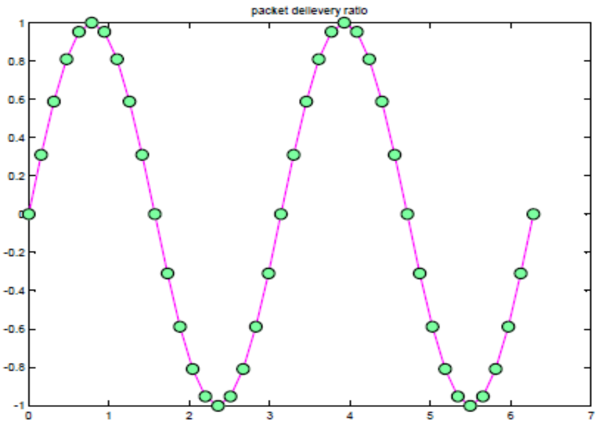

Figure 4: Packet delivery ratio of our data base with respect to error rate.

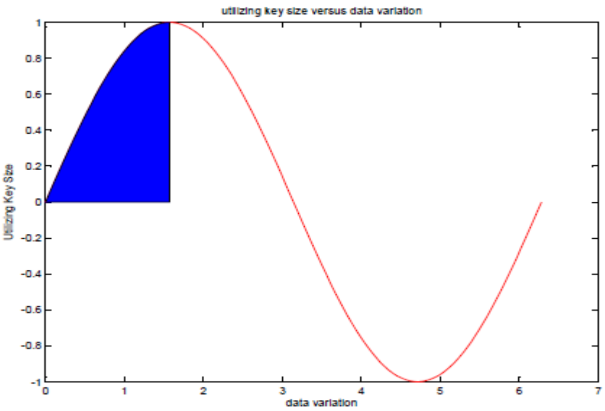

Figure 5: Utilizing key size versus data variation.

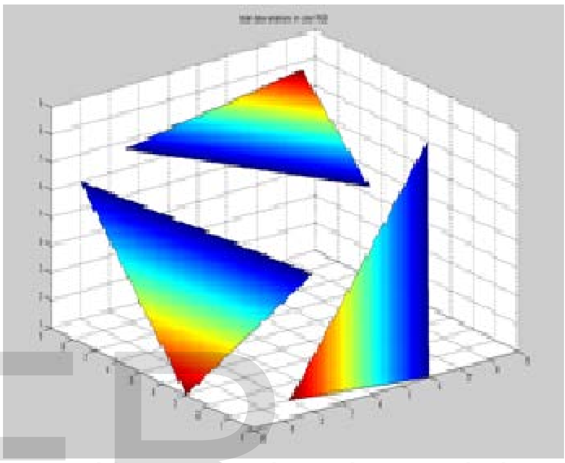

Figure 6: Total data variation in color RGB between our data base using PGP and ID3.

Figure 6 shows that three dimensional axis in which total data variation shown between our data base with respect to error rate using pretty good privacy (PGP) and ID3 algorithm. Figure 7 shows also three dimensional axes in which total data rata shown between our data base with respect to error rate using pretty good privacy (PGP) algorithm. It clears that error rate of PGP is high so data delivery is fast on the other hand graph of ID3 algorithm is denial because ID3 algorithm works on virtual access which makes ID3 process slow.

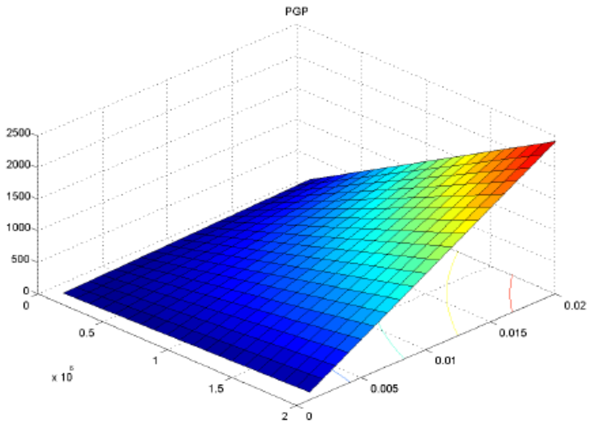

Figure 7: Total rate between our data base using PGP.

We know that database security is applied by two algorithms i.e. PGP (pretty Good Privacy) algorithm and ID3 (Decision tree) algorithm. Wireless Sensor Network sense the authorization of nodes. ID3 algorithm arranges all the nodes in the form of tree and if it found any unauthorized node accessing then it will block. On the other hand PGP is used to encrypt the data and send data in the form of encrypted data. Now the node that has decrypt key only that can access the data unauthorized can't read it. The speed of ID3 is less then PGP because the work of ID3 is to check the authorization of each and every node on the other hand work of PGP is to encrypt the data send from one node to another node.

VI. Conclusion

We conclude that database security in wireless sensor network was described, the specific approaches of database security are characterized and compare two algorithms i.e. PGP (Pretty Good Privacy) and ID3 (Iterative Dichotomiser 3 algorithm). In wireless sensor network when unauthorized person enter in the network or hack the data then ID3 (Iterative Dichotomiser 3 algorithm) algorithm provide security in the form of fault tolerance but ID3 (Iterative Dichotomiser 3 algorithm) process is slow because it works in the form of decision tree and consume more power and packet delivery ratio is less. On the other hand PGP (Pretty Good Privacy) algorithm used less power consumption makes a process fast because it used shortest path for packet delivery then delivery ratio is more.PGP (Pretty Good Privacy) use encryption technique to secure the data. PGP (Pretty Good Privacy) uses route of optimization for packet delivery. At last we says that PGP (Pretty Good Privacy) is much better than ID3 (Iterative Dichotomiser 3 algorithm) algorithm because PGP (Pretty Good Privacy) transfer data with more speed than ID3 (Iterative Dichotomiser 3 algorithm) and it is not easy to decrypt data by unauthorized node. We have also improving the efficiency of the algorithms to reduce the overhead of secure path notification and making use of the secure paths in a local cache or database to detect compromised nodes and perform intrusion detection for WSNs in future experimental work.

References

1) Ejaz Ahmad, Monitoring and analysis of internet traffic targeting unused address spaces. PhD in Computer Science, Queensland University of Technology, Brisbane, Australia, 2010.

2) A. A. Ahmed, H. Shi, and Y. Shang. A survey on network protocols for wireless sensor networks. In Proceedings of the IEEE International Conference on Information Technology: Research and Education, ITRE'03, Network, New Jersey, USA, pages 301-305, august 11-13, and 2003.

3) Modares, H.; Salleh, R.; Moravejosharieh, A. "Overview of Security Issues in Wireless Sensor Networks" Computational Intelligence, Modeling and Simulation (CIMSiM), 2011 Third International Conference on Digital Object Identifier: 10.1109/CIMSim.2011.62.

4) Bertino E.; Sandhu, R. "Database security - concepts, approaches, and challenges" Dependable and Secure Computing, IEEE Transactions on Volume: 2 , Issue: 1 Digital Object Identifier: 10.1109/TDSC.2005.

5) Kumar, D.; Kashyap, D.; Mishra, K.K.; Mishra, A.K. "Security Vs cost: An issue of multi-objective optimization for choosing PGP algorithms" Computer and Communication Technology (ICCCT), 2010 International Conference on digital object identifier.

6) J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. Culler, and K. Pister, System Architecture Directions for Networked Sensors, ASPLOS, November 2000.

7) John A. Stankovic, "Wireless Sensor Networks", University of Virginia, Charlottesville, Virginia 22904, June 19, 2006.

8)

9) F. L. LEWIS," Wireless Sensor Networks" The University of Texas at Arlington 7300 Jack Newell Blvd. S Ft. Worth, Texas 76118-7115.

10) Mainwaring, J. Polastre, R. Szewczyk, D. Culler, J. Anderson, "Wireless sensor Networks for habitat monitoring", in: Proceedings of ACM International Workshop on Wireless Sensor Networks and Applications (WSNA), 2002.

11) Modares, H.; Salleh, R.Moravejosharieh, A. "Overview of Security Issues in Wireless Sensor Networks " Computational Intelligence, Modeling and Simulation (CIMSiM), 2011 Third International Conference on Digital ObjectIdentifier:10.1109/CIMSim.2011.62 .

12) Ben Adida, Susan Hohenberger and Ronald L. Rivest, "Lightweight Encryption for Email", In Proceedings of the Steps to Reducing Unwanted Traffic on the Internet Workshop, August 2005.

13) Axel Poschmann, Gregor Leander, Kai Schramm and Christof Paar, "New Light-Weight Crypto Algorithms for RFID", IEEE International Symposium on Circuits and Systems, May 2007.