MODEL FOR SECURE DATA TRANSMISSION IN DEEP SPACE NETWORKS

By

ARAVIND KUMAR

Master of Science in Information Technology

Anna University

Chennai, Tamilnadu

2004

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
May, 2007

MODEL FOR SECURE DATA TRANSMISSION IN

DEEP SPACE NETWORKS

Thesis Approved:

Dr. Johnson P. Thomas

Dr. Venkatesh Sarangan

Dr. Mathias Schulze
Thesis Adviser

Dr. A. Gordon Emslie
Dean of the Graduate College

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 <u>INTRODUCTION TO SATELLITE SYSTEMS:</u>

A satellite is an object that orbits around another object. For example, the Moon is a satellite of the Earth, and the Earth is a satellite of the Sun. A rocket or cargo bay of the space shuttle is used to carry the satellite into orbit. Many emerging applications will incorporate multiple spacecraft that form communications networks necessary to achieve coverage, latency and throughput requirements.

Satellite systems have the advantage of global coverage and inherent broadcast capability and offer a solution for providing broadband access to end users. Today, satellite technology is all around to bring us live coverage of events from around the world. Satellite networks play an important role in achieving global coverage by providing commercial, civil and military services. Many applications use satellite networks for data delivery. Worldwide communication using internet, telephone, television and radio ride on the presence of backbone satellites. Present day satellite networks enable people to transmit data from/to any part of the globe instantaneously. Compared to geostationary (GEO) satellites, low earth orbit and medium earth orbit satellite networks have shorter round trip delays and lower transmission power requirements. They can also be used to carry signaling and network management traffic as well as data packets.

The two most important elements of satellite networks are the satellites and the Earth stations. Generally, data packets will be transmitted from Earth stations to satellites and vice versa.

1. *Satellites* - A satellite is an object that orbits around another object like earth. Satellites carry equipments like antennas, cameras, radar and transponders. Satellite payload represents all equipment needed to do its job. Communications satellites equipped with antennas and transponders receive the original signal from the transmitting Earth station and re-transmit this signal to the receive stations on Earth. The omni directional antennas that were used in communication satellites were replaced by unidirectional, pointed antennas. Researches concluded unidirectional antennas pointing quite precisely towards the destination outperform omni directional antennas. A weather satellite has cameras included in its payload. The payload for satellites depends on the operations they perform. Inter-satellite links enable inter-satellite communication, while satellite-earth links are used for message exchange with Earth stations. Satellites have processing capabilities and buffers to store information for transmission. Satellites also have rechargeable batteries to supply power when it goes out of the Sun's scope.
2. *Earth Station* – An Earth Station is located on the Earth's surface and is not

mobile. Earth stations transmit or receive data using a relay back bone of satellite networks. Earth stations like satellites have antennas, usually a dish, and are equipped with transmitters, decoders and receivers. In general, the earth stations have high power antennas which enable large coverage distance. The type and size of the antennas used varies with the type of services provided. Earth stations are sink nodes or destinations, for a sensor satellite network. The Application devices of the Earth stations transform radio signals received into information and transfer them to a computer or to a destined device, like a TV if it is a broadcast program. Similarly, this device will transform information to be transmitted into a signal that is suitable for transmission via the antenna, using modulation, amplification and other processing techniques.

A Satellite network, composed of mobile satellites, fixed ground stations and communication links, have characteristics such as: long propagation delays, limited energy and time varying relatively high channel error rates.

a) *Mobility* - Satellites are mobile and their mobility can be pre-computed using Keplerian laws, as they rotate in their orbits. Geostationary satellites move relative to earth and are always stationary above a point on the earth. Satellite mobility balances the resource utilization among the satellites, *Long Propagation Delay* - Satellites communicate using inter-satellite links and use satellite-ground links to communicate with earth stations. Satellites are usually far from one another and from the ground resulting in long propagation delays. Propagation delay for deep-space communication links is variable and extremely long.

Satellite networks are advantageous over terrestrial networks, as they are less affected by congestion; their architecture is scalable and also has coverage at geographical locations where it is hard to have a terrestrial network. Satellite TV, like Direct TV, which is a satellite based application can serve any individual, irrespective of how far is he from the nearest cable TV junction with digital quality television programming. In a country like Japan these services will fit best as it is practically not feasible to lay cable through all its islands.

1.2 PROBLEMS IN EXISTING MODELS:

Little work has been reported on space based security systems. The main thrust of space communications to-date has been to provide secure communications between ground mission control and a single spacecraft. However, with the proliferation of spacecrafts and instruments, the potential for malicious attacks has significantly increased. The security working group of Consultative Committee for Space Data Systems (CCSDS) has published a number of green books as recommendations for security protocols at the different communication layers [8,9]. In addition CCSDS has proposed security architectures for space communications as well as identified potential threats [8,9]. However, currently there is no key management infrastructure for space communications.

Existing terrestrial key management structures are not suited for space due to the high latency and error rates. Key management schemes used previously are used only for general networks and are not suitable for wireless networks due to limited computational abilities of nodes. Predistribution of secret keys for all pairs of nodes is not viable due to large amount of memory used when the network size is large. Furthermore, secure communications during spacecraft emergencies has received a very little attention to-date.

Networking protocols developed so far for mobile ad hoc networks may not be suitable for inter-space networks [1]. These protocols have been developed primarily for scenarios involving links without long propagation delays and for networks where node mobility cannot be pre-determined. While the former prevents the applicability of any such protocol for inter-space networks, the latter does not allow a protocol to take advantage of the predictable mobility patterns found in inter-space networks. Further, the existing protocols also may not incorporate the constraints necessitated by the limited energy-supplies and intermittent connectivity between the orbiters. This further diminishes the applicability of the existing protocols for the given problem scenario. While some protocols have been developed for space-based networks [1], such protocols do not incorporate feedback from the physical layer and hence may result in sub-optimal performance.

1.3 PROPOSED WORK:

Our main objective is to develop an algorithm that can increase the connectivity and security in the path while minimizing overheads such as complex computations, increased end to end delay, etc.

In section 1.2, we have identified certain problems found in existing schemes.

Problem 1) Node compromise is a serious treat to wireless networks deployed in unattended and hostile environments.

Problem 2) Key management schemes used previously are used only for general networks and are not suitable for wireless networks due to limited computational abilities of nodes. Predistribution of secret keys for all pairs of nodes is not viable due to large amount of memory used when the network size is large.

In this thesis we investigate the above problems. The proposed algorithm tries to secure the communications between two nodes by using a two-tiered key management scheme based on location based key management and bloom's scheme. The location based scheme generates a new common or shared key between two nodes or satellites, without the nodes having to exchange the keys themselves. The only information that is exchanged is the location of the nodes and their IDs, Hence, a malicious node that may be eavesdropping or intercepting the communications will not be able to obtain the shared key. Furthermore, only nodes which satisfy certain location constraints are able to generate a shared key. Thus a rogue node which does not satisfy location constraints is not able to generate a shared key. This approach is particularly attractive in space networks since the location of a node is predictable. Hence if the location reported by a node does not match the expected location constraints, a shared key is not generated. However, since authentication is based only on location and ID, a malicious node may be able to generate the shared key because it is within the required proximity. We use Bloom's scheme to ensure that such nodes, although authenticated, cannot generate a pairwise key which is needed for communicating with nodes.

Bloom's scheme is a fast technique for generating a pairwise key between nodes. This key can be used for encrypting messages, thereby providing secure communications. We used Bloom's as a second level of security in our approach. The location based scheme outlined above serves to authenticate a node. Our assumption here is that ID and location are sufficient to authenticate a node. However, it is possible that a malicious node may be able to authenticate itself based on proximity. Bloom's scheme uses the shared keys generated by the location scheme to generate a pairwise key which is used for communications. This may involve communications with the earth station (only if insufficient memory is available at the node). Bloom's scheme prevents a malicious node which may be able to authenticate itself, but it will not be able to generate a pairwise key. In our approach the shared key is needed to generate the pairwise key, hence making our proposed approach more secure.

A second approach is to obtain a secure communication between two nodes. In This approach we use the base topology graph derived from Location based graph and security graph derived from Bloom’s scheme. The proposed scheme helps us in deriving a network graph, which is the edge wise intersection of base topology graph and security graph. The resulting Network graph has the property that two nodes can be connected only if they are within a range and share a common key.

The proposed key distribution scheme outlined in the previous paragraph works because the nodes are assumed to be within a certain range of each other. In our case, we call this range r_0, which is the communication range of a node. We next analyze the probability that if nodes lay within an area or range share a common key between them. A low probability indicates that a node has a low chance of being within the range r_0, and any node that reports to being with this range is therefore likely not to be malicious. On the other hand, a high probability indicates that many nodes may be within range and this increases the chances that there may be malicious nodes within the range. In order to make sure there are no malicious nodes in the network we use the key distribution algorithm to check whether if two nodes within range share a common key between them. If they do not share then one of the nodes is considered to be malicious.

In the proposed algorithm, every satellite is considered to be a node. The satellite that is close to the earth station is called the root satellite and the one near the destination is termed as the terminal node. Adjacent satellites can communicate with each other and share keys to enable secure communication without much overhead. Unlike mobile networks, once the connection is established, they tend to exist unless the satellites move away from each other which will be for a definite period of time. The movement of a satellite is pre-determined and hence proposed algorithm supports this system.

In chapter II we review the literature in the field, Chapter 3 discusses in detail the proposed approach chapter IV analyses our simulation results and finally chapter V concludes the thesis with the main findings and work to be done in the future.

CHAPTER II

REVIEW OF LITERATURE

In this section we review previous work in the field. We first review the popular communication protocols – TCP and UDP. This discussion is followed by a security routing protocol in sensor networks namely Location based authentication which provides perfect network resilience.

In 1945, Arthur C. Clarke first predicted that satellites in orbit approximately 36,000 kilometers above the equator, with a period of 24 hours, could maintain a fixed location as seen from the ground. In this geostationary orbit (GSO), a satellite could receive signals from the ground and transmit them over roughly a third of the Earth's surface. For more than three decades now, GSO satellites have been virtually the exclusive means of providing space-based communications (e.g., TV broadcast, long distance telephone, etc).

Satellite networking, using inter-satellite links, is essential to have continuous access to any part of the globe achieving global coverage and to carryout real time data transmission. A communication satellite is one used to receive and transmit data from and to any part of the globe, while sensor satellites like weather satellites are used to monitor and forecast weather conditions. Satellite sensor networks have sensors to sense the environment of our interest and transmit it to the ground stations. In general, space networks can be classified based on the operations they perform and here are the satellite network types. [2]

2.1 IPSEC (IP SECURITY)[4]:

Protocols operate at the network layer, layer 3 of the OSI model. Other Internet security protocols in widespread use, such as SSL and TLS, operate from the transport layer up (OSI layers 4 - 7). This makes IPsec more flexible, as it can be used for protecting both TCP- and UDP-based protocols, but increases its complexity and processing overhead, as it cannot rely on TCP (OSI layer 4) to manage reliability and fragmentation.

There are two modes of IPsec operation: Transport mode and Tunnel mode.

In Transport mode only the payload (message) of the IP packet is encrypted. It is fully routable since the IP header is sent as plain text; however, it cannot cross NAT interfaces, as this will invalidate its hash value. Transport mode is used for host-to-host communications. A means to encapsulate IPsec messages for NAT traversal have been defined by UDP Encapsulation of IPsec ESP Packets.

In Tunnel mode, the entire IP packet is encrypted. It must then be encapsulated into a new IP packet for routing to work. Tunnel mode is used for network-to-network communications (secure tunnels between routers) or host-to-network and host-to-host communications over the Internet.

2.2 TCP [4]:

The Transmission Control Protocol (TCP) is a virtual circuit protocol that is one of the core protocols of the Internet protocol suite, often simply referred to as TCP/IP. Using TCP, applications on networked hosts can create connections to one another, over which they can exchange streams of data using Stream Sockets. The protocol guarantees reliable and in-order delivery of data from sender to receiver. TCP also distinguishes data for multiple connections by concurrent applications (e.g., Web server and e-mail server) running on the same host.

TCP does not perform well on satellite channels due to high delay bandwidth, high bit error rate and burst errors and thus increases work load on Transport and data link layer for retransmission.

2.3 UDP [4]:

The User Datagram Protocol (UDP) is one of the core protocols of the Internet protocol suite. Using UDP, programs on network computers can send short messages sometimes known as data grams (using Datagram Socket) to one another. UDP is sometimes called the Universal Datagram Protocol.

UDP does not provide the reliability and ordering while TCP does. Data grams may arrive out of order, appear duplicated, or go missing without notice. Without the overhead of checking if every packet actually arrived, UDP is faster and more efficient for many lightweight or time-sensitive purposes. Also, its stateless nature is useful for servers that answer small queries from huge number of clients. Compared to TCP, UDP is required for broadcast (send to all on local network) and multicast (send to all subscribers).

2. 4 SCPS (SPACE COMMUNICATIONS PROTOCOL STANDARDS) [9]:

SCPS is a protocol suite designed allows communication over challenging environments. Originally developed jointly by NASA and DoD's USSPACECOM to meet their various needs and requirements. These protocols have been found to be applicable in meeting the needs of the satellite and wireless communities.

SCPS, a completely open and proven technology, has met the needs of commercial, educational, and military environments. It was designed to meet the following goals:

1. Best possible use of limited bandwidth
2. High link utilization

3. Conservation of power
4. Prioritization of traffic
5. Tolerant of intermittent connectivity
6. High forward/return link asymmetry

2.5 LOCATION BASED AUTHENTICATION[7]:

We use location based authentication in our approach as each satellite resents a potential point of compromise. Once compromising certain nodes and acquiring keying material, adversaries can launch various insider attacks such as they can spoof, alter or replay routing information to interrupt network routing, may launch Sybil attack where a single node presents multiple identities to other nodes, or launch identity replication attack, etc. This situation poses demand for compromise tolerant security design. That is the network should remain highly secure even when a number of nodes are compromised. Moreover, this scheme enables deterministic, secure and efficient establishment of a shared key between any two network nodes be there immediate neighbors or multiple hops away.

Node compromise is a serious threat to wireless sensor networks deployed in unattended and hostile environments. [7] To mitigate the impact of compromised nodes, we propose a suite of location based compromise tolerant security mechanisms. This is based on a new cryptographic concept called pairing and by binding private keys of individual nodes to both their vicinity.

2.5.1 PRE-DEPLOYMENT PHASE:

Assumption: All nodes have the same transmission range R and communicate via bi directional wireless links. Nodes perform a collaborative monitoring of the designated sensor field and report the sensed events to the distant sink, which is the data collection center with sufficiently powerful processing capabilities and resources.

Let p, q be two large primes and E / F_p indicate an elliptic curve $y^2=x^3+ax+b$ over the final field F_p. We denote by G_1 a q-order sub group of additive group of points by E / F_p and by G_2 a q-order subgroup of the multiplicative group of finite field $F^*_{p^2}$

Tasks before network deployment:

a) Generate the pairing parameters, (p, q, E/F_p, G_1, G_2, ệ) and select an arbitrary generator W of G_1.

b) Choose two cryptographic hash functions: H, mapping strings to nonzero elements in G_1, and h, mapping arbitrary inputs to fixed-length outputs, e.g., SHA-1 [7].

c) Pick a random k € Z_q as the network master secret and set $W_{pub} = kW$.

d) Calculate for each node A an ID-based key (IBK for short), $IK_A = k\ H(ID_A)$ € G_1

Each node is preloaded with the public system parameters (p, q, E / F_p, G_1, G_2, e, H, h, W, W_{pub}) and its private IK_A. It is important to note that it is computationally infeasible to deduce from k either (W, W_{pub}) or any (ID, IBK) pair like (ID, IK_A), due to the difficulty of solving the DLP (Discrete Logarithmic Problem) in G_1. Therefore, even after compromising an arbitrary number of nodes and their IBKs (Identity based key), adversaries are still unable to calculate the IBKs of non compromised nodes

2.5.2 SENSOR DEPLOYMENT AND LOCALIZATION:

1. After the pre-deployment phase, the nodes are deployed in various ways - physical installation or random aerial scattering.
2. The nodes are localized using either of two localization techniques namely Range based localization and Range free localization.

a. RANGE BASED LOCALIZATION:

A group of mobile robots are dispatched across the whole sensor field along pre planned routes which have powerful computation and communication capabilities than ordinary nodes.

The robot is equipped with master secret key. In order to localize a node:

i. The mobile robot runs the secure range-based localization protocol mentioned in references to measure their respective distance to node A and the co-determine the location of A.

ii. Then it computes hash function based on its location, master secret key and its id and sends information to A. Here, encrypting message with master key refers to message integrity code (MIC) of message.

iii. Upon receipt of message, node A first uses its pre-loaded keying material to decrypt its location id and then re generates MIC. If both match, then it saves the location id for subsequent use. During subsequent network operations, node addition may be necessary to maintain good network connectivity which will be done in similar fashion.

b. **RANGE FREE LOCALIZATION**:

In this kind of localization technique, there are some special nodes called anchors knowing their own locations. All other non-anchor nodes derive their locations based on information from anchors and neighboring nodes via secure range-free localization techniques mentioned in references.

The nodes are pre-loaded with master secret key, which is used to derive their locations based on information from anchors and neighbors via secure range-free localization with assumption of secure-sensitive environment that the adversary takes time interval t which is more than the time taken to localize node and generation of location based id.

2.5.3 LOCATION BASED NEIGHBORHOOD AUTHENTICATION:

During the post deployment phase, each node is required to discover and perform mutual authentication with neighboring nodes. Each node will think of another node as an authentic neighbor if and only if that node is within its transmission range R and also holds the corresponding LBK (Location Based Key). Suppose node A wishes to discover and authenticate neighboring nodes after obtaining its location and its LBK:

I) Node A broadcasts an authentication request including its ID ID_A and location ℓ_A and some random nonce n_A.

II) Upon receipt of such a request, node B first needs to verify the claimed distance using Euclidian's distance ($||\ell_A - \ell_B|| <= R$) so that adversaries cannot surreptitiously tunnel authentication messages between B and some virtual non-neighbor node. If the inequality holds then B simply discards the authentication request. Otherwise, it continues with step III.

III) It computes a shared key based as follows

$K_{B, A} = \hat{\underset{\cdot}{e}} (LK_B, H(ID_A || \ell_A))$. It then unicasts a reply to node A including its ID and location, a random nonce n_B, and MIC computed as h $K_{B, A} (n_A || n_B || 1)$

IV) Upon receiving the reply, node A also first checks if the inequality $||\ell_A - \ell_B|| <= R$ holds. If so, it proceeds to derive a shared key as

$K_{A, B} = \hat{\underset{\cdot}{e}} (LK_A, H(ID_B || \ell_B))$ whereby to re-compute the MIC. If the result is equal to what B sent, node A considers B as authentic neighbor. Subsequently, A returns to node B a new MIC computed as

h $K_{A, B} (n_A || n_B || 2)$.

V) Upon receipt of it, B uses $K_{B, A}$ to regenerate the MIC and compares the result with what it just received. If they are equal, B regards node A as an authentic neighbor as well.

The above process is valid because, if and only if both A and B have a correct LBK, $K_{A,B}$ is equal to $K_{B, A}$ due to the following equations:

$$
\begin{aligned}
K_{A, B} &= \hat{e}\,(LKA, H\,(ID_B \,||\, \ell_B)) \\
&= \hat{e}\,(k\,H\,(ID_A \,||\, \ell_A), H\,(ID_B \,||\, \ell_B)) \\
&= \hat{e}\,(H\,(ID_A \,||\, \ell_A), k\,H\,(ID_B \,||\, \ell_B)) \\
&= \hat{e}\,(k\,H\,(ID_B \,||\, \ell_B), H\,(ID_A \,||\, \ell_A)) \\
&= \hat{e}\,(LKB, H\,(ID_A \,||\, \ell_A)) \\
&= K_{B, A}
\end{aligned}
$$

In case multiple nodes simultaneously respond to the same authentication request, MAC layer mechanisms like random jitter delay (every node has to wait before answering an authentication request) are used to resolve this problem.

2.6 <u>BLOOM'S SCHEME:</u>

The key sharing scheme model allows any pair of nodes in a network to share a pair wise secret key as long as no more then λ nodes are compromised.

Let $(\lambda + 1) \times N$ be a matrix over a finite field GF(q), where N is the size of the network and q > N. Matrix G is a public information and is shared by all nodes in the network . During key generation phase the base station creates a random $(\lambda + 1) \times (\lambda + 1)$ symmetric matrix D over GH(q) and computes an $N \times (\lambda + 1)$ matrix $X = (D.G)^T$ where $(D.G)^T$ is the transpose of D.G. Matrix D must be kept secret and should not be disclosed to others.

Let K = X.G and $K_{ij} = K_{ji}$ where K_{ij} is the element in the i^{th} row and j^{th} column of K as we use K_{ij} or K_{ji} are the pairwise key between node i and node j. For k = 1,2,...N

- store the k^{th} row of matrix X at node k
- store the k^{th} column of matrix G and node k

When nodes i and j need to communicate they exchange their columns of G and the compute K_{ij} or K_{ji}. Using private rows of X. as G is public information.

2.7 PROBABILITY THEORY:[13][14][15][16]

Probability Theory is concerned with modeling phenomena whose nature involves elements of uncertainty. Probability Theory deals with models of experiments whose outcomes cannot be precisely predicted. Tossing a coin, the duration of the uninterrupted operation of a machine, the proportion of defective articles in a large batch of items and more fall into the category of phenomena which can be studied by means of probabilistic methods.

Probability has become very powerful mathematical tool in understanding those aspects of the world that cannot be described by deterministic laws. The central objects of probability theory are random variables, stochastic processes and events. Probability theory is important in the social sciences because it gives a theoretical background for both sampling and data analysis. The correct application of statistical methods and the interpretation of results also require familiarity with the most important concepts and results of probability theory.

2.7.1 <u>PROBABILITY SPACE:</u>

Let Ω be a set and *A* is a collection of subsets of Ω. *A* is called a σ algebra on Ω if

- Ø Є *A*
- For any A Є *A*, the complement Ω \ A Є *A*
- Whenever A_1, A_2,….,A_n Є *A* then also A_1U A_2 U….U A_n Є *A*

Assume now that A is σ algebra and *A* function

P : A → [0,1] is called a measure if

- P(Ω) = 1
- If a A_1, A_2,….,A_n Є *A* are disjoint then P (A_1U A_2 U….U A_n) = P(A_1) + P(A_2)….P(A_n)
- P(Ω\A) = 1 – P(A) for all A Є *A*.

A probability space is triple (Ω, *A*, P) where Ω is a set, *A* is a σ algebra on Ω and P is a measure on *A*.

2.7.2 <u>RANDOM VARIABLE:</u>

A random variable on a probability space (Ω, *A*, P) with density function *f* is a map such that X: Ω → ℝ.

<u>DISCRETE RANDOM VARIABLE:</u>

Let X be a random variable such that X: Ω → 𝔻. The function p(x) given by p(x) = P(X = x), x Є D. is the probability distribution of X. it is also known as probability mass function.

Suppose a random variable X takes k different values, with the probability that X= x_i defined to be $P(X = x_i) = p_i$. The probabilities p_i must satisfy the following:

1: $0 \leq p_i \leq 1$ for each i

2: $p_1 + p_2 + ... + p_k = 1$.

<u>CONTINUOUS RANDOM VARIABLE:</u>

Let X be a random variable X: $\Omega \rightarrow \mathbb{R}$ that takes values in a uncountable set Ω , which is all or part of the real line $\mathbb{R}$..

The random variable X is said to be continous with density $f(x)$ for all $a \leq b$,

Such that $X^{-1}([a,b]) \in A$ and $P(X \in [a,b]) = P(X^{-1}([a,b])) = {}_a\int^b f(x)\, dx$

The probability density $f(x)$ is called the probability density function (p.d.f) where f (x) has the following properties of densities.

1) $f(x) \geq 0$
2) ${}_{-\infty}\int^{\infty} f(x)\, dx = 1$

2.7.3 <u>UNIFORM DISTRIBUTION:</u>

Uniform Distribution is the simplest continuous distribution in probability. It has constant probability density on an interval (*a*, *b*) and zero probability density elsewhere. The distribution is specified by two parameters: the end points *a* and *b*. We denote the distribution U(a , b). Probability density function (PDF) is given by

$$\varnothing(x) = \begin{cases} \dfrac{1}{b - a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

2.7.4 POISSON DISTRIBUTION:

The Poisson Distribution is a discrete distribution which takes on the values X = 0, 1, 2, 3, ... It is often used as a model for the number of events in a specific time period.

$$f(k\ ;\lambda\ ) = (\lambda^{k}\ e^{-\lambda}\ )/\ k!$$

Where e is the base of natural algorithm, k is the number of occurances of an event, k! is the probability of an event and λ the average number of events for given interval.

2.8 GRAPH THEORY:[17]

Graph theory is a study of graphs. A graph G = (V,E) consists of two sets, a finite set V of elements called vertices and a finite set E of elements called edges. If each edge has unordered pair of vertices, G is called undirected graph and if the edges of a graph G has ordered pairs of vertices, then G is called a directed or an oriented graph. We use v_1, v_2, $v_{3\ldots}v_n$ to represent the vertices and e_1, e_2,e_3,..e_m to represent edges.

PATH:

A path is a sequence of vertices from V_1 to V_u such that there is an edge to the next vertex in the sequence i.e., (V_i, V_j) Є E.

2.8.1 INTERSECTION OF TWO GRAPHS:

Consider two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$. The intersection of G_1 and G_2 denoted as $G_1 \cap G_2$, is the graph $G_3 = (V_1 \cap V_2 , E_1 \cap E_2)$, the vertex set G_3 consists of only those vertices present in both G_1 and G_2, and the edge set of G_3 consists of only those edges present in both G_1 and G_2.

CONNECTEDNESS OF A GRAPH:

Two vertices v_i and v_j are said to be connected in a graph G if there exists a path from v_i to v_j in G. A vertex is connected to itself. A graph G is connected if there exists a path between every pair of vertices in G. In a connected graph any two longest paths have a common vertex.

NODE DEGREE:

The degree of a node u is denoted as d(u), is the number of neighbors of node u, i.e., its number of links. A node degree d = 0 is isolated, i.e., it has no neighbors. The minimum node degree of a graph G is denoted as

$$d_{\min} (G) = \min_{\forall u \in G} \{d(u)\}$$

Average n ode degree of G is

$$d_{\text{mean}} (G) = \frac{1}{n} \sum_{u=1}^{n} d(u)$$

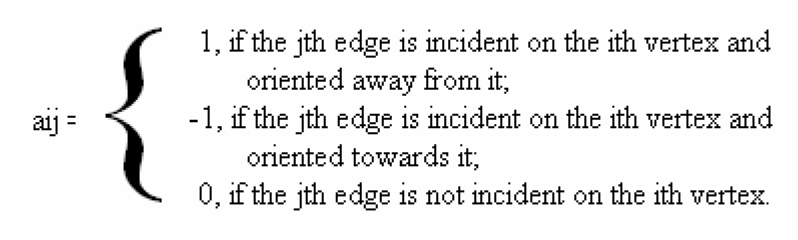
K- CONNECTIVITY:

A graph is *k*–connected if and only if no set of $(k - 1)$ nodes exists whose removal would disconnect the graph. In other words, if $(k - 1)$ nodes fail, the graph is guaranteed to be still connected.

MATRICES OF A GRAPH:

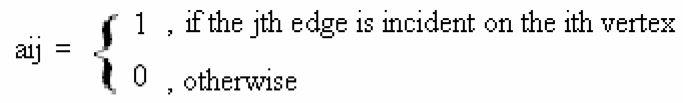
INCIDENCE MATRIX:

Consider a graph G with n vertices and m edges. The incidence matrix A = [aij] of G has n rows and m columns. The element aij of A is defined



$$\text{aij} = \begin{cases} 1, \text{ if the jth edge is incident on the ith vertex and} \\ \quad \text{oriented away from it;} \\ -1, \text{ if the jth edge is incident on the ith vertex and} \\ \quad \text{oriented towards it;} \\ 0, \text{ if the jth edge is not incident on the ith vertex.} \end{cases}$$

G is undirected



$$\text{aij} = \begin{cases} 1 & \text{, if the jth edge is incident on the ith vertex} \\ 0 & \text{, otherwise} \end{cases}$$

ADJACENCY MATRIX:

Let G = (V, E) be a directed graph with no parallel edges. Let V = $\{v_1, v_2, \ldots .v_n\}$. The adjacency matrix M = $[m_{ij}]$ of G is n x n matrix with m_{ij} defines as

$$mij = \begin{cases} 1 \ , & \text{if } (vi, vj) \in E \\ 0 & , \text{otherwise} \end{cases}$$

REACHABILITY MATRIX:

Let D = (V,A) be a digraph, where V = {1, 2, . . . , n}. The adjacency matrix of the digraph D is an n x n matrix A, where aij the entry on the i-th row and j-th column, is defined by

$$aij = \begin{cases} 1 & \text{if } (i,j) \in A \\ 0 & \text{if } (i,j) \notin A \end{cases}$$

The reachability matrix of the digraph D is an n x n matrix R where r_{ij} , the entry on the ith row and jth column, is defined by

$$rij = \begin{cases} 1 & \text{if j is reachable from i} \\ 0 & \text{if j is not reachable from i} \end{cases}$$

CHAPTER III

METHODOLOGY

3.1 PROBLEM DEFINITION:

Our main objective is to develop an algorithm that can increase the connectivity and security in the path while minimizing overheads such as complex computations, increased end to end delay, etc.

Problem 1) Node compromise is a serious threat to wireless networks deployed in unattended and hostile environments.

Problem 2) Key management schemes used previously are used only for general networks and are not suitable for wireless networks due to limited computational abilities of nodes. Predistribution of secret keys for all pairs of nodes is not viable due to large amount of memory used when the network size is large.

To solve the above two main objectives we have four sub problems to be solved

Section 3.3: To use the fusion of Location based algorithm [7] and Bloom's scheme [10] to derive a model for secure communication

Section 3.4: Simulation model to analyze k- connectivity of a network.

Section 3.5: Probability to find a model where two nodes are within a range r_0 and they both share a common key.

Section 3.6: Security graph [10] and Base topology graph [7] are to be combined to form a Network Graph. To analyze whether the network graph has the properties of both security graph and base topology graph.

3.2 <u>CONSTRAINTS AND ASSUMPTIONS</u>

3.2.1 <u>CONSTRAINTS:</u>

1. There is maximum key sharing (Adjacent nodes share a common key which is pre assigned before deployment) between neighboring nodes in a network
2. The probability that key sharing between neighboring nodes in the network is maximized as we use the fusion between Bloom's scheme [10] and location based authentication [7]. One of the main reason for increase in probability that we use the same information from which the bloom's scheme generates symmetric matrix D which is later used for key
3. Each satellite has enough buffers to hold and re-transmit information.
4. Each satellite has very good processing capability.
5. At least one satellite has direct contact with the Earth station.

3.2.2 ASSUMPTIONS

1. All the computations are handled by the ground station, so complexity of initial set up is not considered. Once the network is set up, the individual nodes can compute the keys from key sharing space which can be easily handled by a node in network.
2. The input is a single tree from the root nodes to the destination nodes.

3.2.3 KEY TERMS:

1. **Root satellites**: The satellites that are directly connected to the ground stations
2. **Ground Station:** These are half duplex nodes that are equipped with directional antennas and are fixed. Also, they have more resources compared to satellites and are assumed to be homogenous in nature.
3. **Mobile Satellites:** Satellites rotate according to orbital kinematics and hence their motion can be pre-determined and the root satellites keep changing with time.
4. **Key Graph**: The graph G (V, E), where V is the set of vertices, and E the set of edges of G, each element (vertex or edge) represents the node or link in the network.
5. **Security graph**: This graph is generated by using blooms scheme [10]. A typical network is represented as graph G (V, E). Node set V represents all the routers. E is the set of physical links, i.e. E = (u, v) | u, v **€** V, u and v are connected by a physical link. Physical link represents nodes that share a common key.

6. **Base Topology graph:** This graph is generated using Location based Algorithm [7], considers the shortest path from root node to the terminal node. We denote the Base topology graph as G_L (V_L, E_L). V_L is the set of nodes. $V_L \subseteq V$ and some nodes in V may not have been pre-deployed with an ID (satellites which may be deployed before location based ID scheme was implemented). E_L represents the set of sessions, i.e., E_L = (u, v)|u, v € V_L, u and v share an common key and corresponds to path P_{uv} (Path taken along the nodes to reach the destination)
7. **Network Graph:** Combination of Security and Base topology graph.
8. **Node:** Each Satellite is considered as a Node. A Node is given a unique Id and Secret Key. Nodes also share a common key.
9. **Seed:** A term given for a secret quantity shared between nodes. E.g., Key.
10. **Node degree:** The degree of a node u, denoted as d(u) is the number of neighbors of a node u, i.e., the number of links.

3.3 INTEGRATING BLOOM'S [10] AND LOCATION BASED ALGORITHMS [7]

3.3.1 PREDEPLOYMENT PHASE:

Construct a matrix of size (λ +1) x N over a finite field GF (q) where N is the size of the network and q>N. This matrix G is public information and may be shared by different systems; even adversaries are assumed to know G. Node A is assigned a matrix G_A where G_A has the public information of Node A

During the key generation phase the base station generates random symmetric secret matrices $D_1 \ldots D_w$ of size (λ + 1) x (λ +1) over GF(q). Node A is assigned a secret matrix D_A.

Matrix X_A is product of matrix D_A and matrix G_A.X_A is symmetric hence it holds the below property

$$X.G = (D.G)^T .G = G^T . D^T .G = G^T .D.G = (X.G)^T.$$

Each node carries a random master key *k,* {M}*k* means encrypting message M with key *k* bound to both its ID and geographic location. H is the hash function mapping string to non zero elements. *h* is a hash function mapping arbitrary inputs to fixed length outputs. E.g., when a message of any length < 2^{64} bits is the input, the *h* function produces a 160-bit output this type of process is called a message digest.

ê is a pairing map E.g., Let G1 and G2 be two groups of order q for some large prime q then ê : G1 X G1 → G2 is bilinear if ê (aP,bQ) = ê$(P,Q)^{ab}$ for all (P,Q) Є G1 and all (a, b) Є Z_q^*.

AFTER DEPLOYMENT

STEP 1: Node A broadcasts an authentication request including its ID ID_A, location l_A and a random nonce $n_{A.}$

STEP 2: Node B upon receiving request from Node A checks whether Node A is within its Euclidean distance $||l_A - l_B|| \leq R$.

STEP 3: If Node A is not within Euclidean distance of Node B, it ignores the request. Otherwise B calculates a shared key $K_{B,A} = \hat{e}$ (LK_B, $H(ID_A \,||\, l_A)$. it then unicasts a reply to Node A including its ID and location, a random nonce n_B and MIC computed $HK_{B,\,A}$ (n_A, $||$ n_B $||1$).

STEP 4:Upon receiving the request, node A also first checks if the inequality $||l_A - l_B|| \leq R$ holds if so it proceeds to derive a shared key as $K_{AB} = \hat{e}$ (LK_A, H (ID_B $||$ l_B)) whereby to recompute MIC $HK_{A,\,B}$ (n_A, $||$ n_B $||2$).

STEP 5: The above steps are valid if and only if both A and B has a correct LBK, K_{AB} is equal to K_{BA} which is derived from following equation.

$$\begin{aligned} K_{AB} &= \hat{e}(\, LK_A, H\,(ID_B \,||\, l_B)) \\ &= \hat{e}(\, kH(ID_A \,||\, l_A), H(ID_B \,||\, l_B)) \\ &= \hat{e}(\, H(ID_A \,||\, l_A), kH(ID_B \,||\, l_B)) \\ &= \hat{e}(\, kH(ID_B \,||\, l_B), H(ID_A \,||\, l_A)) \\ &= \hat{e}(\, LK_B, H\,(ID_A \,||\, l_A)) \\ &= K_{BA} \end{aligned}$$

STEP 6: If the result of MIC computed by Node A is equal to Node B, they both become authenticate neighbors.

STEP 7: Assuming Node A and Node B are neighbors and have exchanged the keys. Node A regenerates the matrices G_B and D_{AB}, Node B regenerates matrix G_A and D_{BA}, where G_A and G_B is the public information and D_{AB} and D_{BA} are the secret information exchanged between node A and node B. X_A and X_B are again calculated using new values of G.

$X_{AB} = (D_{AB}. G_A)^T$

$X_{BA} = (D_{BA}. G_B)^T$

STEP 8: A pair wise key between nodes A and B are computed as follows:

When nodes A and B want to establish a pair wise key, they exchange their columns of G_A or G_B after the relation K_{AB} and K_{BA} computed using the private rows of X_{AB} or $X_{BA.}$

For node A

$X_{AB}.G_B = (D_{AB}.G_A)^T. G_B = G_A^T. D_{AB}^T. G_B = G_A^T. (G_B. D_{BA}) = G_A^T. X_{BA} = (G_A. X_{BA})^T = G_A. X_{BA}$

For node B

$X_{BA}.G_A = (D_{BA}.G_B)^T. G_A = G_B^T. D_{BA}^T. G_A = G_B^T. (G_A. D_{AB}) = G_B^T. X_{AB} = (G_B. X_{AB})^T = G_B. X_{AB}$

Now in order to establish a secured communication between node A and node B the pair wise key generated by node A should be equal to the pair wise key generated by node B.

STEP 9: A secured and reliable network is established between Node A and Node B.

EXAMPLE SCENARIOS:

SCENARIO I:

Nodes A and B are within the transmission range.

STEP 1: Node A broadcasts an authentication request including its ID ID_A, location l_A and a random nonce $n_{A.}$

STEP 2: Node B upon receiving request from Node A checks whether Node A is within its Euclidean distance $||l_A - l_B|| \leq R$.

STEP 3: Node B calculates a shared key $K_{B,A} = \hat{e}\ (LK_B, H(ID_A \parallel l_A)$. it then unicasts a reply to Node A including its ID and location, a random nonce n_B and MIC computed $HK_{B,\ A}\ (n_A, \parallel n_B \parallel 1)$.

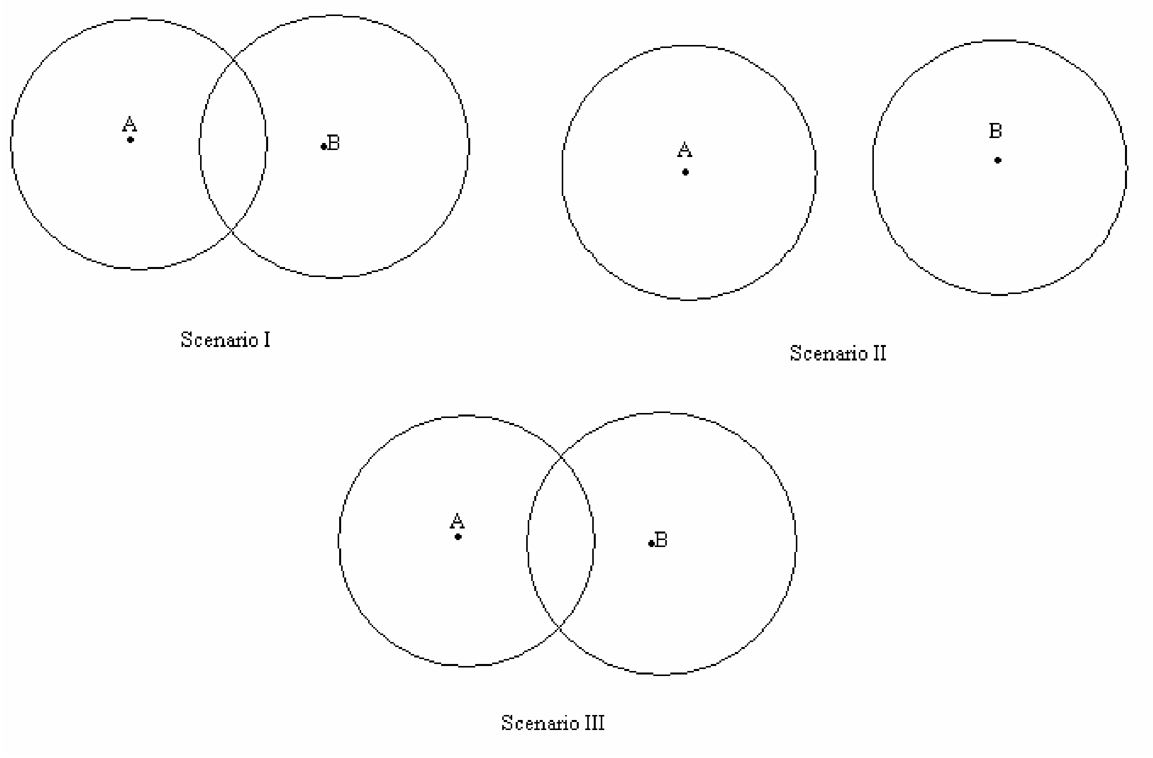


Figure 1: Various Scenarios of Node placement.

STEP 4:Upon receiving the request, node A also first checks if the inequality $||l_A - l_B|| \leq R$ holds if so it proceeds to derive a shared key as $K_{AB} =\hat{e}\ (LK_A, H\ (ID_B \parallel l_B))$ whereby to recompute MIC $HK_{A,\ B}\ (n_A, \parallel n_B \parallel 2)$.

STEP 5: The above steps are valid if and only if both A and B has a correct Location based key, K_{AB} is equal to $K_{BA.}$

STEP 6: If the result of MIC computed by Node A is equal to Node B, they both become authenticate neighbors.

STEP 7: Assuming Node A and Node B are neighbors and have exchanged the keys. Node A regenerates the matrix G_B, and Node B regenerates matrix G_A. X_A and X_B is again calculated use new value of G.

$X_{AB} = (D_{AB}. G_A)^T$.

$X_{BA} = (D_{BA}. G_B)^T$.

STEP 8: A pair wise key between nodes A and B are computed. If the key generated by A is equal to the key generated by Node B then a Pair wise key between Node A and B is established.

STEP 9: A secured and reliable network is established between Node A and Node B.

SCENARIO II

Node B is outside transmission range r0 of Node A, Node B is not malicious and also Node B is not authenticated by Node A.

STEP 1: Node A broadcasts an authentication request including its ID ID_A, location l_A and a random nonce $n_{A.}$

STEP 2: Node B upon receiving request from Node A checks whether Node A is within its Euclidean distance $||l_A - l_B|| \leq R$.

STEP 3: Node A is not within the transmission range of B. So Node B just ignores the authentication request of Node A.

SCENARIO III:

Node A is within the transmission range of Node B, but Node B is malicious, Node A authenticates Node B, but pairwise key not generated.

STEP 1: Node A broadcasts an authentication request including its ID ID_A, location l_A and a random nonce $n_{A.}$

STEP 2: Node B upon receiving request from Node A checks whether Node A is within its Euclidean distance $||l_A - l_B|| \leq R$.

STEP 3: Node B calculates a shared key $K_{B,A} = ê (LK_B, H(ID_A || l_A)$. it then unicasts a reply to Node A including its ID and location, a random nonce n_B and MIC computed $HK_{B,\ A} (n_A, || n_B ||1)$.

STEP 4:Upon receiving the request, node A also first checks if the inequality $||l_A - l_B|| \leq R$ holds if so it proceeds to derive a shared key as $K_{AB} =ê (LK_A, H (ID_B || l_B))$ whereby to recompute MIC $HK_{A,\ B} (n_A, || n_B ||2)$.

STEP 5: The above steps are valid if and only if both A and B has a correct Location based key, K_{AB} is equal to $K_{BA.}$

STEP 6: If the result of MIC computed by Node A is equal to Node B, they both become authenticate neighbors.

STEP 7: Assuming Node A and Node B are neighbors and have exchanged the keys.

Node A regenerates the matrix G_B, and Node B regenerates matrix G_A. X_A and X_B is again calculated use new value of G.

$X_A = (D_{AB}. G_A)^T$

$X_B = (D_{BA}. G_B)^T$

STEP 8: The algorithm checks whether the pairwise key generated by node A is equal to the one generated by Node B. In order to confirm there is no malicious Node between them. If the matrix generated by Node A is not equal to matrix generated by node B there is no pairwise key establishment between Node A and Node B. The malicious node will not be able to calculate its pairwise key as it will not have sufficient information to do the same.

STEP 9: Connection is terminated between nodes A and B considering there is a malicious node between them.

NOTE:

Information exchanged between Nodes consists of their IDs, Location, a random nonce which is a unique number generated only once for a particular node and a secret master key. Using this information each node tries to communicate with the other node to check whether they are an authenticate neighbor

3.3.2 SECURITY ANALYSIS:

To achieve security in wireless networks it is important to encrypt and authenticate messages send between nodes. Keys which are used for performing encryption and authentication must be agreed upon by the communicating nodes.

In the combined scheme first we propose location based keys, in which each node holds a private key bound to both its ID and location. It helps to achieve the desirable goal of localizing the impact of compromise nodes to their vicinity as they check whether the nodes are within its range which is not present in any others previous scheme making it much more secured and reliable. The integrated scheme has strong resilience again node compromise. It guarantee's that as long as there are not more then λ compromised nodes holding a pairwise key which shares the same cell or matrix, adversaries are unable to forge data reports to originate from that particular matrix or cell and escape the filtering by enroute intermiate nodes and the sink.

Adversaries might launch denial of service attacks by trapping legitimate nodes into endless verification of data reports The links in the key sharing graph is secured using a pairwise key computed from the common key space shared by the two nodes. In key sharing setup stage, two neighboring nodes can use the established secure links to agree upon anther random key to secure their communication. Thus the combined scheme is more reliable and secure compared to any other previous scheme.

3.4 SIMULATION MODEL:

The main Objective of this model is to derive a k connected network given its range, number of nodes and its work space width. Our Simulation model consists of rectangular work space where nodes are placed using a random uniform generator. The nodes are placed one after the other in the work space and the transmission range of the satellite are initialized before the nodes are placed, the transmission range of the earth station is also initialized before being placed on the work space. The earth stations or ground stations are placed on the lower part of the work space. The earth stations are assumed to be always interconnected. The earth stations (ground stations) are frequently in contact with the nodes which are within its transmission range.

Once the nodes are placed in the work space, we identify the source and destination nodes between which we need to transmit information. The connectivity algorithm is based on checking whether a directed path exists between the source and destination nodes as well as between all the other nodes in the workspace.

Now, we adopt graphs as a way of identifying the directed paths. In order to do this, we identify the nodes and their interconnections based on the range of the nodes. If the distance between two nodes is within the distance range of the nodes then we consider that the nodes are directly connected otherwise we state that the nodes are not directly connected. Using this premise, we construct a distance matrix (as illustrated in table 2) in which we state the distance between all the nodes in the workspace. For simplicity and computation purposes, if two nodes are not directly connected then we use the value -1 in the distance matrix to indicate that the nodes are not directly connected.

The distance matrix D is created across N x N nodes, that is present in the workspace. From this matrix we construct a binary matrix A, as shown in Table 3, by taking all the elements of the distance matrix D, and apply the transformation that if a cell of D has a positive value then the corresponding cell of the binary matrix A is assigned a value 1. For all other cells, the value 0 is assigned. In linear algebra, the identity matrix or unit matrix of size N is the N-by-N square matrix with ones on the main diagonal and zeros elsewhere. It is denoted by I, or simply by I.

The important property of Identity Matrix I, for any matrix B is that,

$$B \text{ x } I = B \quad \text{and} \quad I \text{ x } B = B$$

From the binary matrix A, we proceed to create the interaction matrix $A_{j.}$ The interaction matrix A_j is defined by the matrix expression $A_j = A + I$, where A is the binary matrix obtained from the distance matrix D and I is the identity matrix corresponding to the binary matrix A as shown in Table 4. In order to establish connectivity between nodes, we try to check if a node x is reachable from node y. This is possible if a directed walk from y to x is possible where walking a graph means traversing an alternating sequence of nodes that are directly connected.

The reachability matrix of the graph A_j is an N × N matrix R where r_{ij}, the entry on the i-th row and j-th column, is defined by

$r_{ij} = 1$ if j is reachable from i

$r_{ij} = 0$ if j is not reachable from i

Satellites	**X coordinate**	**Y coordinate**	**Range**
S1	58905.8	71706.2	24000
S2	94842.9	115704	24000
S3	71794.2	131056	24000
S4	23372.7	119156	24000
S5	105735	89126.7	24000
S6	46585.4	126096	24000
S7	24388.7	7044.21	24000
S8	41553.3	127248	24000
S9	6720.21	8776.27	24000
S10	78846.4	90930.8	24000

Table 1**:** Sample Data

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**
S1	-1	-1	-1	-1	-1	-1	-1	-1	-1	20437.9
S2	5519.51	-1	-1	-1	-1	-1	-1	-1	16915.1	-1
S3	-1	10555.7	-1	-1	-1	-1	-1	-1	-1	-1
S4	-1	-1	-1	-1	-1	-1	5222.74	-1	-1	-1
S5	-1	18590.8	-1	-1	-1	-1	-1	-1	-1	-1
S6	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
S7	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
S8	-1	23746.3	-1	-1	23855.2	-1	-1	-1	-1	-1
S9	-1	-1	-1	-1	-1	-1	-1	-1	-1	9089.5
S10	-1	-1	-1	-1	-1	-1	-1	-1	21693.6	-1

Table 2: Distance Matrix

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**
S1	0	0	0	0	0	0	0	0	0	1
S2	1	0	0	0	0	0	0	0	1	0
S3	0	1	0	0	0	0	0	0	0	0
S4	0	0	0	0	0	0	1	0	0	0
S5	0	1	0	0	0	0	0	0	0	0
S6	0	0	0	0	0	0	0	0	0	0
S7	0	0	0	0	0	0	0	0	0	0
S8	0	1	0	0	1	0	0	0	0	0
S9	0	0	0	0	0	0	0	0	0	1
S10	0	0	0	0	0	0	0	0	1	0

Table 3: Binary Matrix

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**
S1	1	0	0	0	0	0	0	0	0	1
S2	1	1	0	0	0	0	0	0	1	0
S3	0	1	1	0	0	0	0	0	0	0
S4	0	0	0	1	0	0	1	0	0	0
S5	0	1	0	0	1	0	0	0	0	0
S6	0	0	0	0	0	1	0	0	0	0
S7	0	0	0	0	0	0	1	0	0	0
S8	0	1	0	0	1	0	0	1	0	0
S9	0	0	0	0	0	0	0	0	1	1
S10	0	0	0	0	0	0	0	0	1	1

Table 4: Interaction Matrix

A reachability matrix R is a binary matrix with a reflexive and transitive property, i.e., R + I = R and R^2 = R, where I is the identity matrix. A skeleton matrix S for a reachability matrix R satisfies the inequality $S^{K-2} \neq S^{K-1}$ and S^K = R. Using this understanding, we try to arrive at the reachability matrix R for our interaction matrix A_j which serves as the skeleton matrix S. Therefore if we try to establish the equality

$$A_j^{\ N} = A_j^{\ N+1}$$

Then the resultant matrix $A_j^{\ N}$ will be the reachability matrix R. From the definition of reachability between two nodes as given in equation 3.1, we know that if a particular cell has a value 1 then the corresponding column node is reachable from the node at the corresponding row. It follows that to establish connectivity across all nodes all the elements of the reachability matrix should be 1. The algorithm tests this state of the reachability matrix to establish connectivity for the simulation.

The model is then run for a number of times to test for connectivity of the nodes. Each run will generate a new set of location data for the nodes and the earth stations. Satellite range (min and max) specifies the actual distance of transmission of the signal from the node or rather its signal strength in distance. This means that any satellite/station that lies within that range can communicate with this satellite. Transmission of data from one satellite to the other satellite is based on the range of the other satellite.

Satellite density is the number of satellites within a given area. E.g., if total satellites are 100 in a workspace of 1000 x 1000 then for a density area of 100 x 100, we can have probably 2 satellites at the maximum or 0/1 at the minimum.

3.5 PROBABILISTIC MODEL:

The main objective of this probabilistic model is to calculate the probability that two nodes are within range r_0 and also share a common key between them.

3.5.1 BASE TOPOLOGY MODEL:

This Probabilistic model derives a probabilistic expression that calculates probability of nodes which lies within an area A given its range r_0, nodes N, and size of the work area..

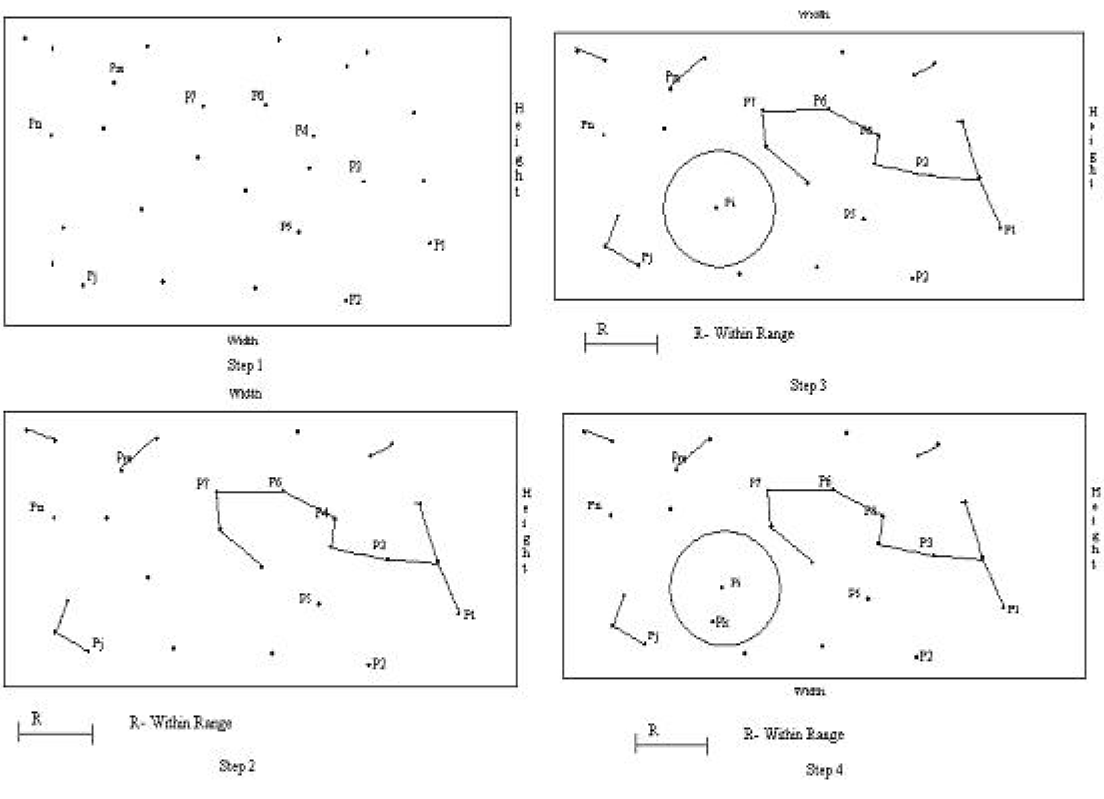


Figure 2: Base Topology Model

STEP 1: Consider p_1, p_2 …p_n Є $\mathbb{R}^2$ are the points randomly uniformly placed in an interval [0,w] and [h,0] on a two dimensional plane.

STEP 2: Connect any two points which are within transmission range r_0

STEP 3: Let $p_i = (x_i, y_i)$ where x_i and y_i are the random variables whose values are in the Disc A. where A is the range of p_i. Let λ denote the probability of a point in A.

Let K denote number of nodes inside the Disc A. ρ is the node density i.e., number of nodes per square area $\rho = N/(w.h)$

$P(p_i \in A) = {}_A\iint \rho \, dA = \rho|A|$

$|A| = \Pi r^2$

$\lambda = \Pi r^2 /(w.\ h)$

STEP 4: Let K denote number of points that are within the disc A. Probability that exactly k points are in disc A.

$$P\ (K=k) = \binom{N}{k} \lambda^k (1- \lambda)^{N - k}$$

Where $\binom{N}{k}$ = N!/((N – k)!k!), Probability to have k points within A is λ^k

Probability to have k points outside Disc A is λ^{N-k}. N is the total number of points.

STEP 5: Probability of finding atleast L points in disc A

$P\ (K \geq L) = P(K = l) + P(K=l +1\ ) +\ldots$

All the above are disjoint

So $P\ (K\geq L) = 1 - (P(K = 0) + (P\ (K=1) +\ldots P(K=L-1))$

$= 1 - ((\lambda^0\ (1 - \lambda)^{N-0} + (\lambda^1\ (1 - \lambda)^{N-1})\ldots\ )$

$= 1 - \sum_{k=0}^{N-1}\ (\lambda^k\ (1 - \lambda)^{N-k}\ )$

3.5.2 KEY SHARING MODEL [10]:

The key sharing probabilistic model allows any pair of nodes in a network to share a pair wise secret key as long as no more then λ nodes is compromised.

Let (λ + 1) x N be a matrix over a finite field GF(q), where N is the size of the network and q > N. Matrix G is a public information and is shared by all nodes in the network . During key generation phase the base station creates a random (λ + 1) x (λ + 1) symmetric matrix D over GH(q) and computes an N x (λ + 1) matrix X = $(D.G)^T$ where $(D.G)^T$ is the transpose of D.G. Matrix D must be kept secret and should not be disclosed to others.

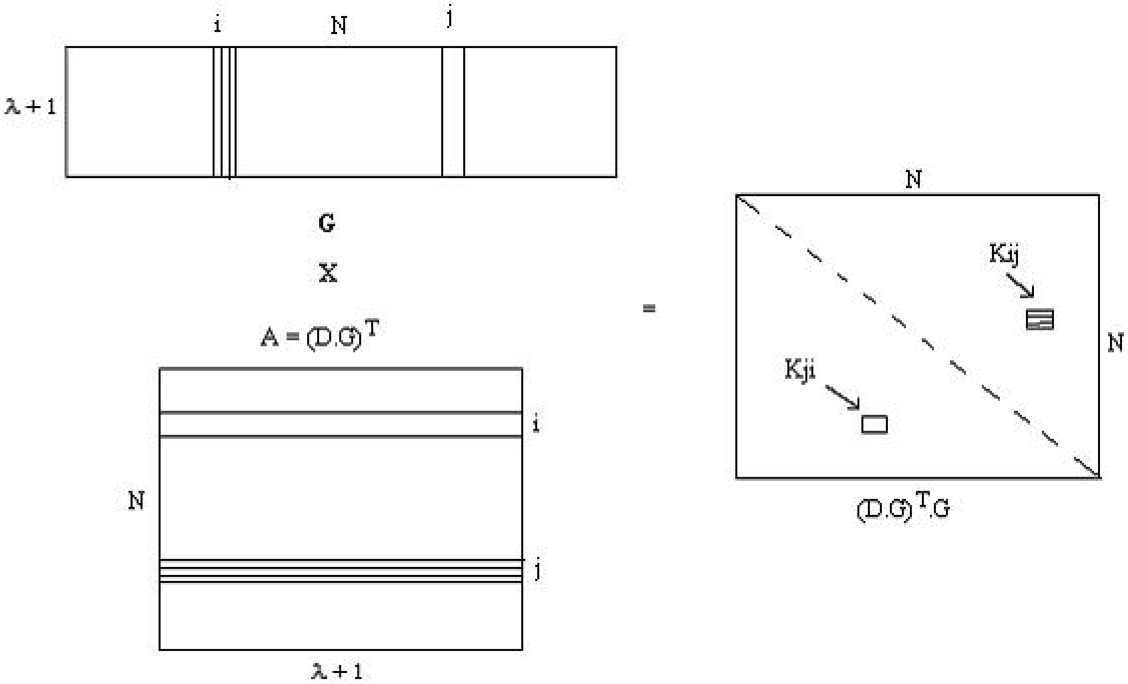


Figure 3: Key Sharing Model [10]

Let K = X.G and $K_{ij} = K_{ji}$ where K_{ij} is the element in the i^{th} row and j^{th} column of K as we use K_{ij} or K_{ji} are the pairwise key between node i and node j. For k = 1,2,...N

- store the k^{th} row of matrix X at node k
- store the k^{th} column of matrix G and node k

When nodes i and j need to communicate they exchange their columns of G and the compute K_{ij} or K_{ji}. Using private rows of X. as G is public information.

During the Key generation phase the base station generates random symmetric matrices. Consider D_1, D_2 …D_n be the secret matrices generated by the base station.

For each node i Є {1,2,…N} the base station chooses a random index di Є {1,2,…m} and assigns the matrix D_{di}.

$P(d_i = j) = 1/m$ where j Є {1,2,…m}

Now fix any two nodes i_1, i_2 Є {1, 2,…N}

Probability that a node Ddi_1 is randomly assigned a key j is given by 1/m and probability that a node Ddi_2 is randomly assigned a key j is given by 1/m.Probability that a node Di_1 and Di_2 is assigned a key j is given by

$Ddi_1 = j = Ddi_2 = 1/m^2$ for any fixed key j.

Probability that the nodes Ddi_1 and Ddi_2 are assigned any random key in j is given by

$$= \sum_{j=1}^{m} \frac{1}{m^2}$$

$$= 1/m$$

3.5.3 COMBINED MODEL:

The combined probabilistic model uses the results obtained from base topology model and security model. Probabilistic model analyzes the probability that nodes if within range share a common key.

Place a point P_{i1} within an area A, where A represents the exact range of P_i. Place a point P_j within the same area A which means that Pi and Pj are within transmission range of each other. Probability that node P_i and Pj are within an area A is given by λ and probability that the node P_i share a common key j with node P_j is (1/m).

The probability that node Pi and node Pj are within range and share a common key is given by λ/m.

Let K denote numbers of nodes that are within the disc A. Probability to have k points within A is λ^k. Probability that k points share anyone of the keys in j is given by (1/m).

Probability that exactly k points are in disc A (within range) and share a common key is given by λ^k/m.

The combined model considers only nodes that are within range and share a common key. We will not consider any nodes that are outside the range.

3.6 NETWORK MODEL:

The main objective of network model is to determine a connected and secure path between nodes. This model intersects two different graphs Base topology graph which gives a reliable path and security graph which gives a secured path to obtain a network graph which is both connected and secure.

3.6.1 BASE TOPOLOGY GRAPH [7]:

The graph generated below uses Location based Algorithm [7] which considers the distance from root node to the terminal node.

Let P_1, P_2 …P_k be the points in $\mathbb{R}^2$ and for some range $R > 0$. A graph is derived using the above input variable as follows, a graph $G_L (V_L, E_L)$ is drawn where $V_L = \{1, 2 \ldots k\}$ and $E_L = \{\{ i, j\} \mid \| Pi – Pj\| \leq R\}$ where $\| Pi – Pj\| = \text{Sqrt} ((X_i – X_j)^2 + (Y_i – Y_j)^2)$ where Pi $=(X_i,Y_i)$.

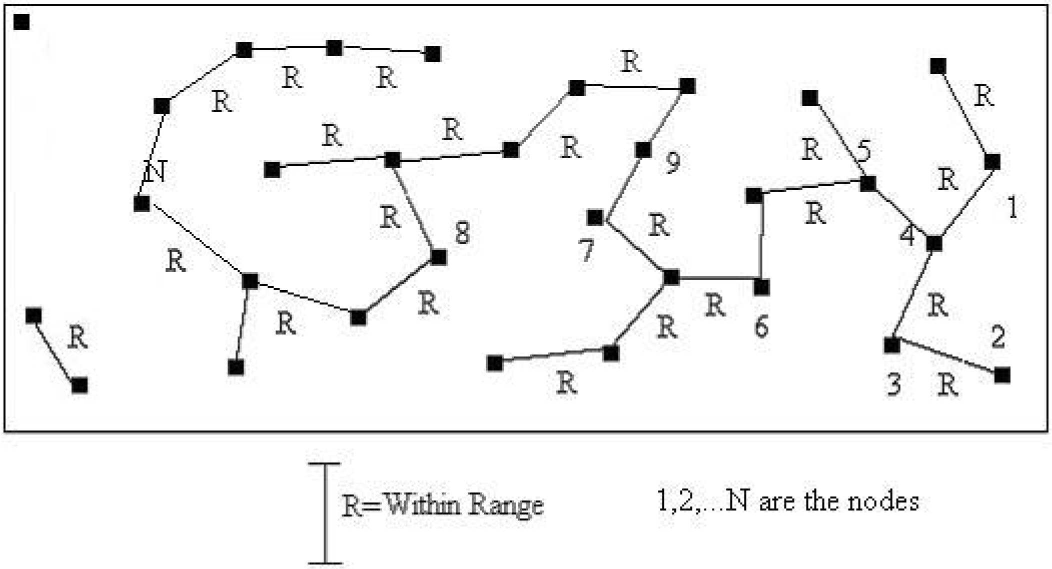


Figure 4: Base Topology Graph

3.6.2 SECURITY GRAPH [10] :

Security graph (Blooms Scheme) considers key sharing algorithm. Nodes which share a common key are connected.

Let K be finite set and K_1, K_2 …K_k be subsets of K i.e., $K_i \subseteq K$. A graph is derived using the above input variables as follows, a graph G_P (V_P, E_P), where $V_P = \{1\ldots k\}$ and $E_P = \{\{i,j\} \mid K_i \cap K_j \neq \emptyset\}$. The output graph connects nodes which share a common key between them.

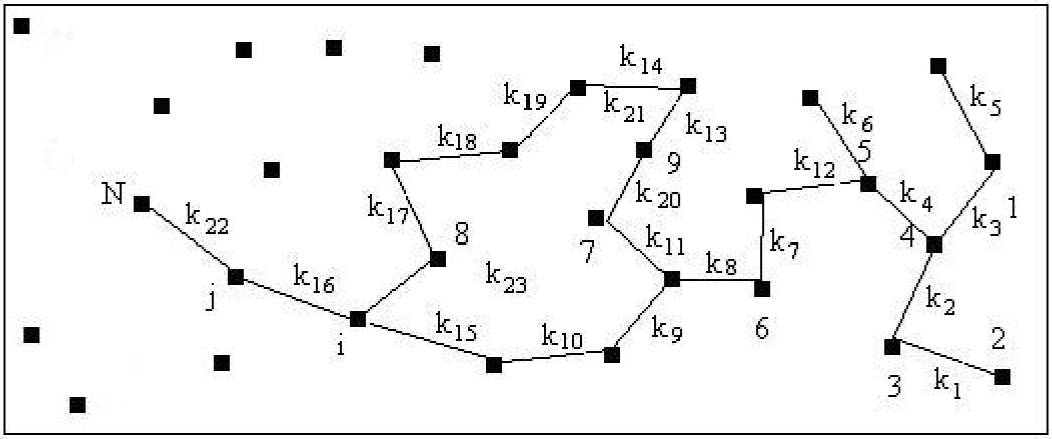


Figure 5: Security Graph

Above diagram depicts a graph where 1,2,..N are the nodes and k = {1,2,...m} are the keys shared between the nodes.

EXAMPLE:

K1 is the key shared between nodes 2 and 3

K2 is the key shared between nodes 3 and 4

3.6.3 NETWORK GRAPH

Network graph is derived by combining Bloom's [10] and Location Based Algorithm's [7]. Bloom's scheme (Security graph) G_P models communication restricted by key sharing such that pair of nodes are adjacent to G_P if and only if they are share a pairwise key between them. The Location based graph (Base topology graph) G_L models logical relationships resulting from the given relation R such that pair of nodes adjacent to G_L if and only if the relation R is true. The Network graph G_N is given by the edge wise intersection of G_P and G_L.

Let (V_L, E_L) and (V_P, E_P) be two graphs derived from Base topology and Security graph respectively. Intersection of these two graphs $G_N = (V_L, E_L) \cap (V_P, E_P)$ given by

$$V_N = (V_L \cap V_P) \text{ and } E_N = (E_L \cap E_P)$$

The resulting Network graph has the property that two nodes can be connected only if they are within a range and share a common key.

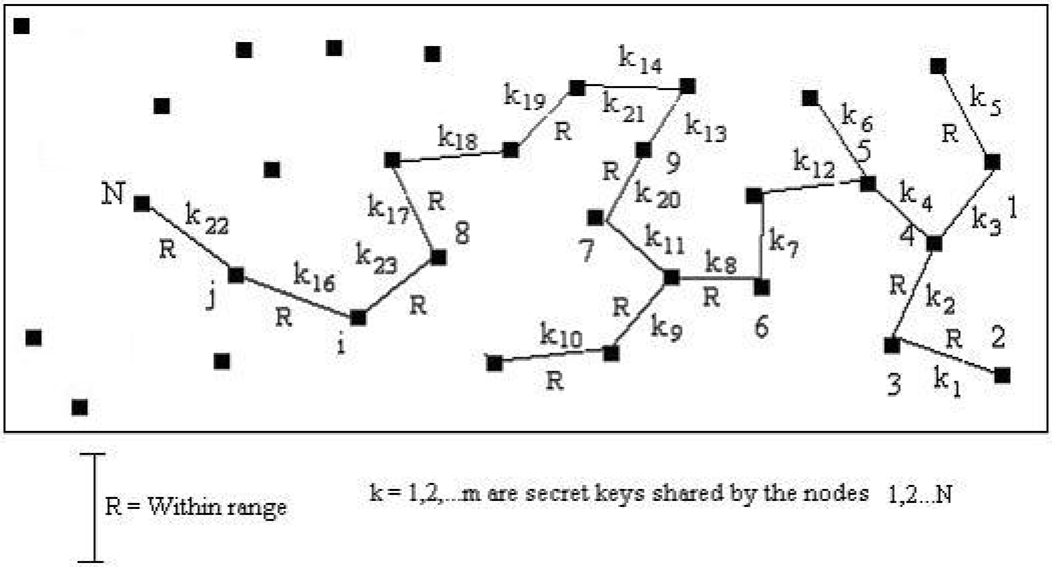


Figure 6: Network Graph

EXAMPLES:

SCENARIO I:

BASE TOPOLOGY GRAPH:

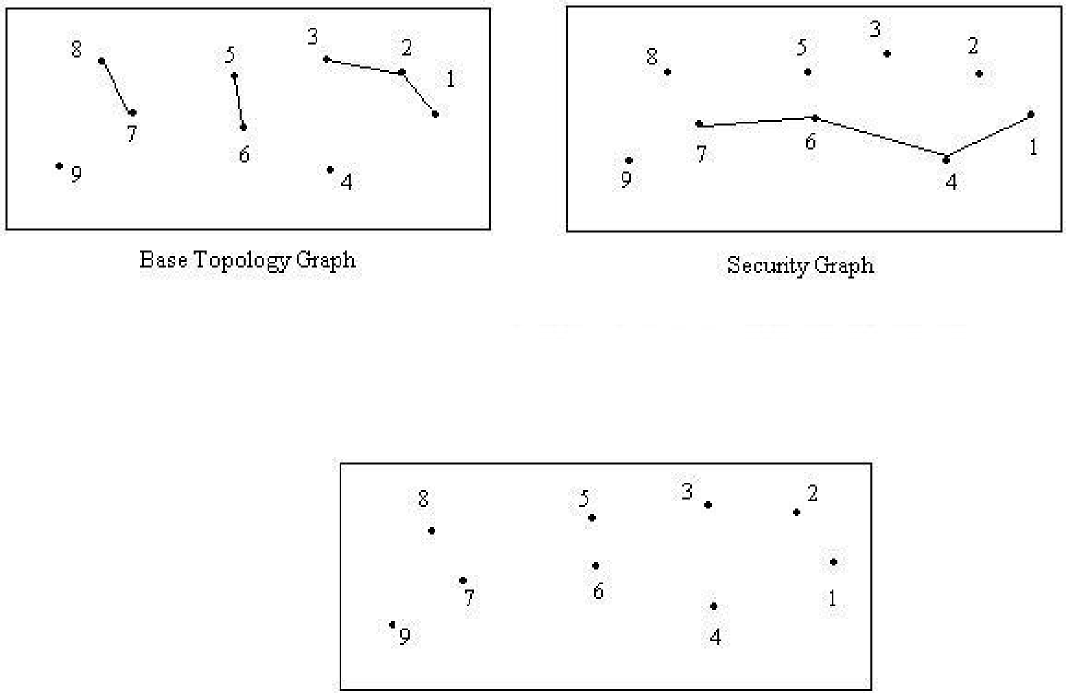
STEP 1: Place points randomly uniformly on the work space.

STEP 2: Connect any two points which are within the range r0.

SECURITY GRAPH:

STEP 3: Place points randomly uniformly on the work space.

STEP 4: Connect any two points which share a common key.



Base Topology Graph

Security Graph

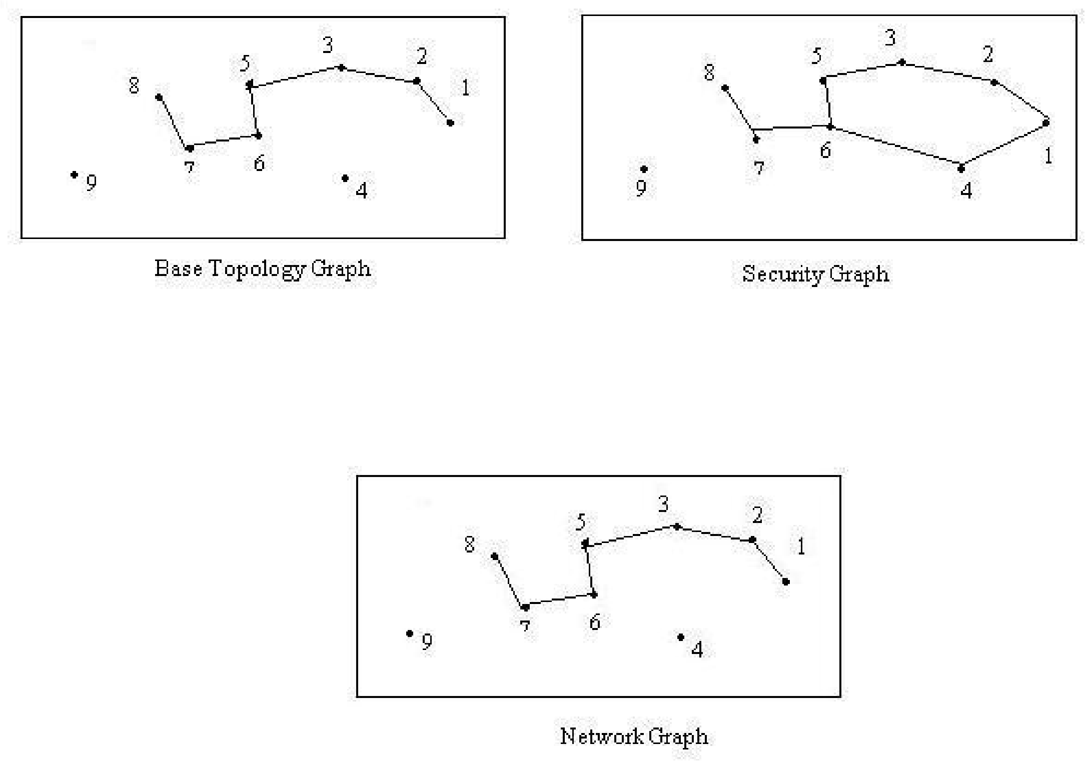
Network Graph

Figure 7: Scenario I

NETWORK GRAPH:

STEP 5: The Edge wise intersection of base topology graph and security graph gives network graph. Scenario I does not have any common links between base topology graph and security graph. There are neither common links nor they share a common key between the two graphs. Network graph cannot be drawn using the above two graphs.

SCENARIO II



Base Topology Graph

Security Graph

Network Graph

Figure 8: Scenario II

BASE TOPOLOGY GRAPH:

STEP 1: Place points randomly uniformly on the work space.

STEP 2: Connect any two points which are within the range r0.

SECURITY GRAPH:

STEP 3: Place points randomly uniformly on the work space.

STEP 4: Connect any two points which share a common key.

NETWORK GRAPH:

STEP 5: The Edge wise intersection of base topology graph and security graph gives network graph. In Scenario II Base topology graph and Security graph share common vertices and they have a path between source to destination. The resulting Network graph has the property that two nodes can be connected only if they are within a range and share a common key.

SCENARIO III:

BASE TOPOLOGY GRAPH:

STEP 1: Place points randomly uniformly on the work space.

STEP 2: Connect any two points which are within the range r0.

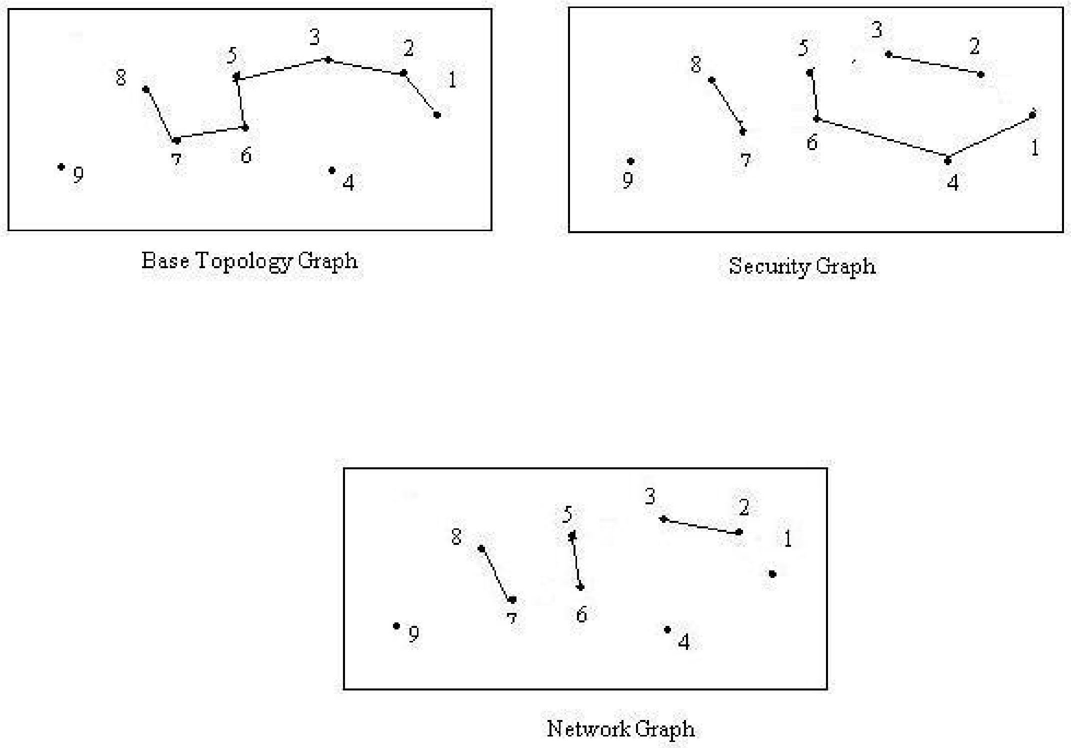
SECURITY GRAPH:

STEP 3: Place points randomly uniformly on the work space.

STEP 4: Connect any two points which share a common key.

NETWORK GRAPH:

STEP 5: The Edge wise intersection of base topology graph and security graph gives network graph. In scenario III Base topology graph and Security graph do have common vertices between them. But they do not have path, there are broken links between them. Network graph obtained is of no use as there is no path between each node to each other node.



Base Topology Graph

Security Graph

Network Graph

Figure 9: Scenario III

SECURITY ANALYSIS:

Network graph is derived from the edgewise intersection of based topology graph and security graph. Network graph is more secure as it has the properties of both base topology graph as well as security graph.

The links in the key sharing graph is secured using a pairwise key computed from the common key space shared by the two nodes. In key sharing setup stage, two neighboring nodes can use the established secure links to agree upon another random key generated using location based to secure their communication. Thus the combined network graph is more reliable and secure compared to any other previous scheme.

CHAPTER IV

SIMULATION

The main objective of this simulation model is to calculate the probability of success to achieve a k connected network given range r0, number of nodes and work space.

A simulation program is designed as explained in section 3.4 in Visual C++. In our simulator a uniform random generator chooses X and Y coordinates of nodes on any given system area. For a given random range r_0, the links between the nodes are created and the min node density of the resulting network is determined along with the probability of connectivity of node with each other node.

The simulation model has following four input parameters

1) Satellite density

2) Earth station density

3) Range of each satellite (min – max)

4) Work area width and work area height

Simulations are carried out by keeping one or two of the above parameters constant and varying the rest of the parameters.

4.1 NO WORK SPACE HEIGHT:

Simulation results for 300 experiments with 200 nodes and the number of earth stations range from 5 to 20 and range r0 ranging 50 to 1000 uniformly distributed on a work space area (70000 X 0).

RESULT:

By varying the ratio of work space, the satellites and earth stations are placed on the same vertical line along x axis as there is no value for y axis. The satellites almost connected if the range r_0>=1000 and number of earth stations are more then 20. The results are taken for r_0<1000. If the range r_0 <=100 and the number of earth stations < 5 the probability of connectedness is almost 0. Now the probability of connectedness increases as the range is increased.

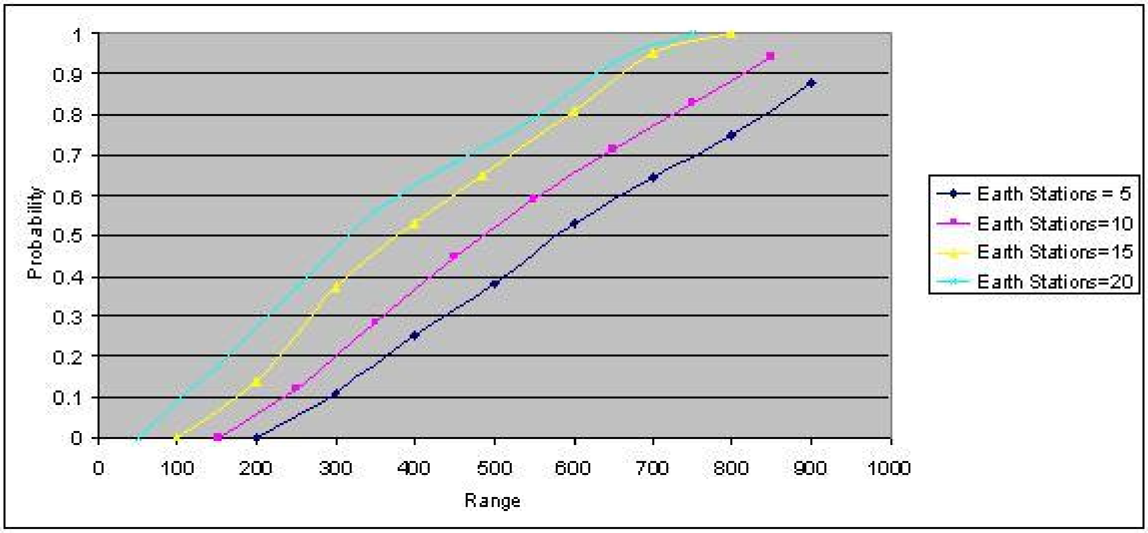


Figure 10: No Work Space Height

4.2 VARYING WORK SPACE HEIGHT AND WIDTH:

MINIMUM HEIGHT MAXIMUM WIDTH:

Simulation results for 300 experiments with 200 nodes by varying both range as well as number of earth stations distributed on work space area (70000 X 10000).

RESULT:

By varying the work space the probability of connectedness of the nodes varies as most of nodes are in contact with some earth station if the number of earth stations are more then 20 and range r0>7000. Probability of success of k connected node is almost 0 if the range is less then 1000 and number of earth stations are less then 5.

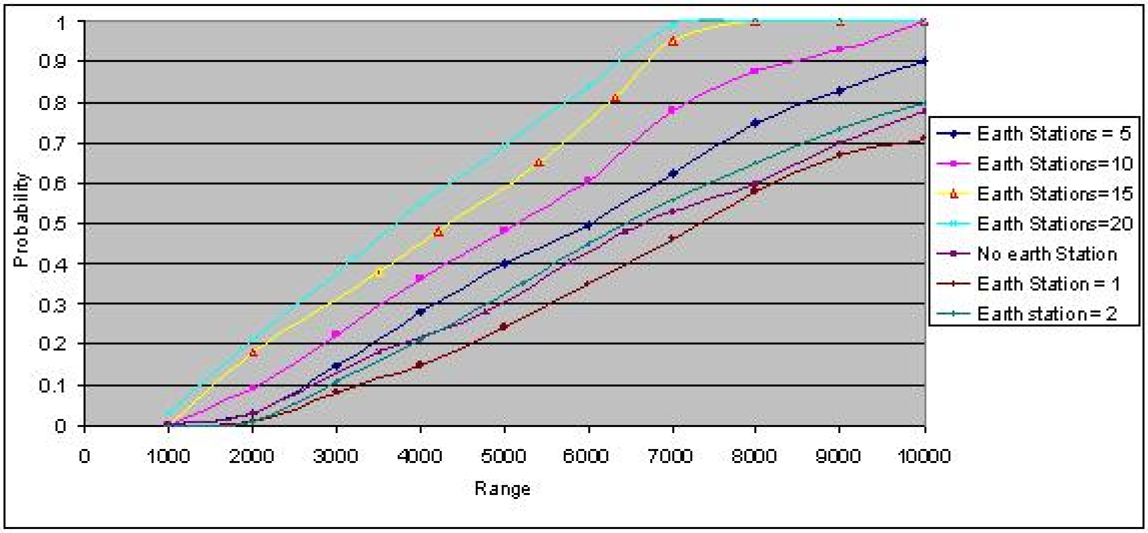


Figure 11: Minimum height Maximum width

MINIMUM WIDTH MAXIMUM HEIGHT:

Simulation results for 300 experiments with 200 nodes by varying both range as well as number of earth stations distributed on work space area (10000 X 70000).

RESULT:

By varying the work space the probability of connectedness of the nodes varies as the number of satellites in contact with the earth station increases. Probability of success of k connected node is almost 0 if the range is less then 1000 and almost 1 if the range is more then 10000 and number of earth stations are greater then 20.

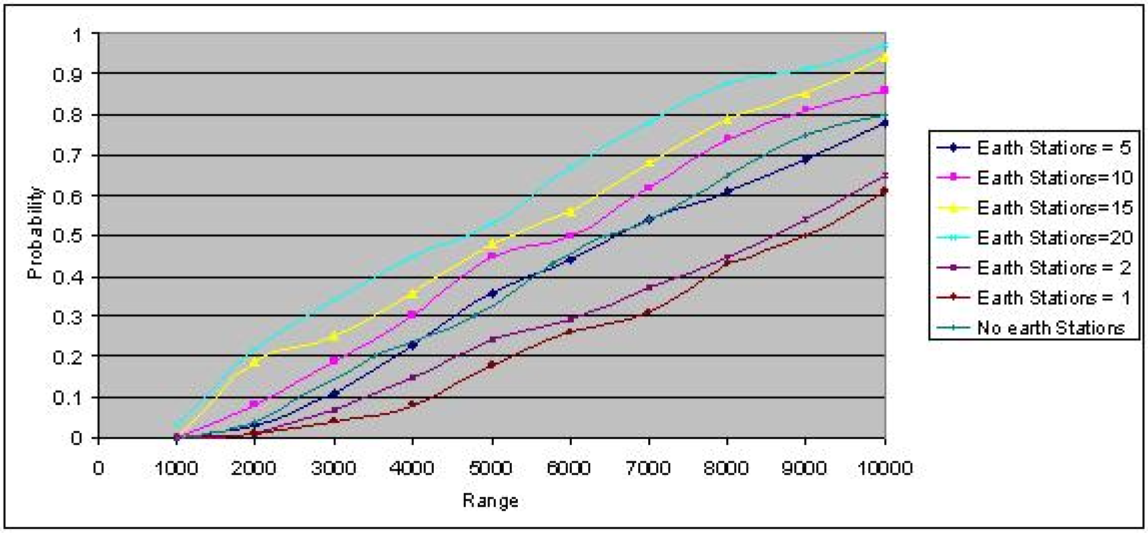


Figure 12: Minimum Width Maximum Height

4.3 VARYING WORK SPACE WIDTH AND HEIGHT WITH NO EARTH STATIONS:

Simulation results for 300 experiments with 200 nodes and number of earth stations is zero uniformly distributed on work space area (10000 X 70000) and work space area (70000 X 10000)

RESULT:

By varying the work space the probability of connectedness of the nodes remains almost the same. As there are no earth stations the probability of connectivity between nodes placed along minimum height and maximum width will remain almost same for nodes placed along maximum height and minimum width.

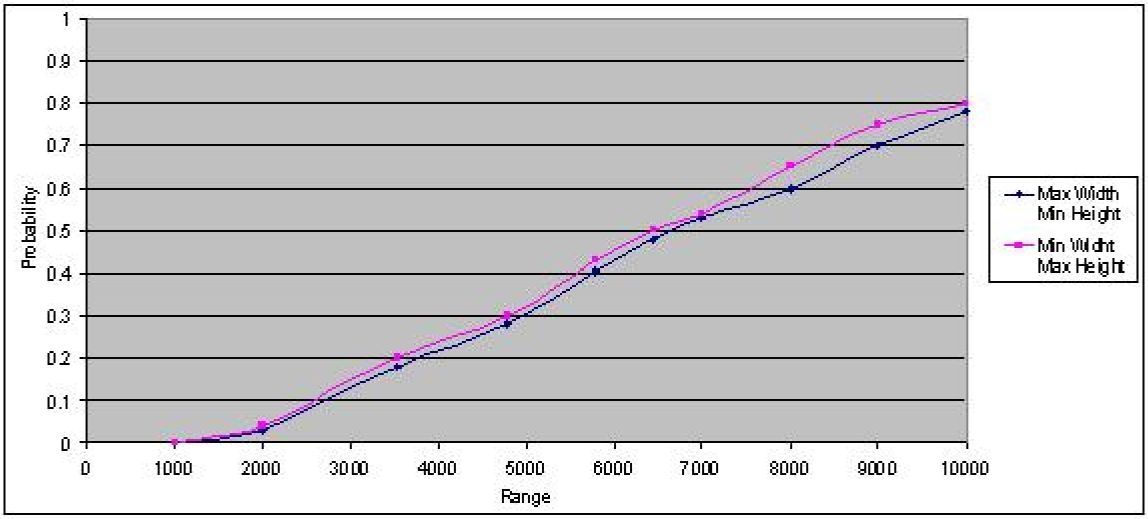


Figure 13: No Earth Stations

4.4 NO WORK SPACE WIDTH:

Simulation results for 300 experiments with 200 nodes uniformly distributed on work space area (0 X 70000). The satellites and are randomly placed along the width of the work space area as the height of the work space area is zero.

RESULT:

By varying the ratio of work space the satellites and earth stations are placed on the same horizontal line along y axis as x axis = 0. The satellites are at most connected if the range r0>1000. The results are taken for r_0>1000. If the range of $r_0 \leq 1000$ the probability of connectedness is almost 0. Now the probability of connectedness increases as the range is increased.

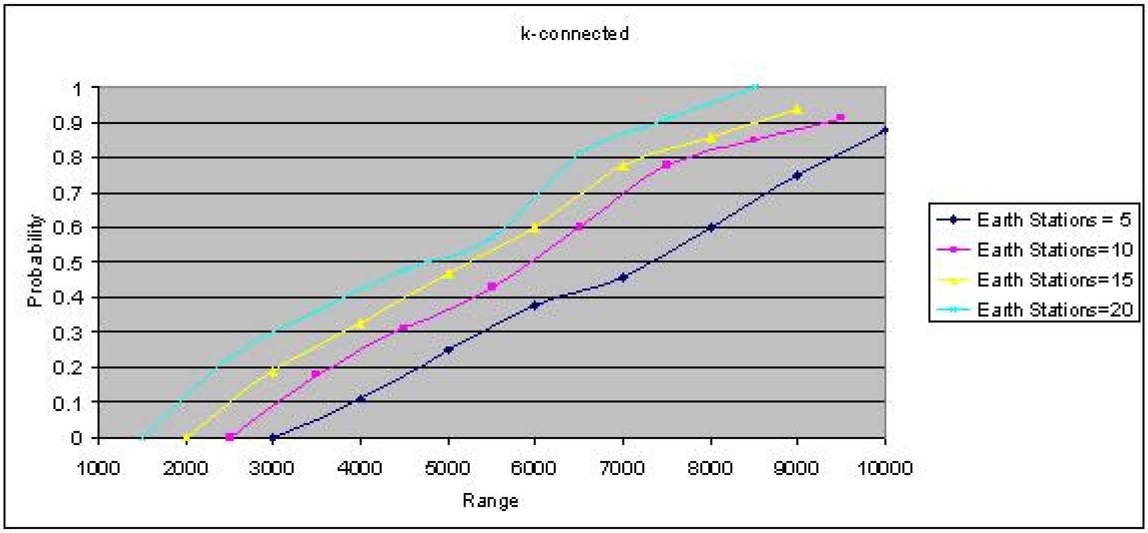


Figure 14: No Work Space Width

4.5 K CONNECTIVITY TEST:

TEST I:

Simulation results for 300 experiments with 200 nodes uniformly distributed on work space Area (70000 X 70000) with range varying from 500 - 2500 checking k connectedness by removing nodes (k≤1 to k=3).

RESULT:

The Probability of being k-connected changes fast from 0 to 1 as r_0 increases. Simulation is run by removing one node at a time; the results may vary as the number of nodes that is being removed varies. The curves below show a significant difference for probabilities below 96%.

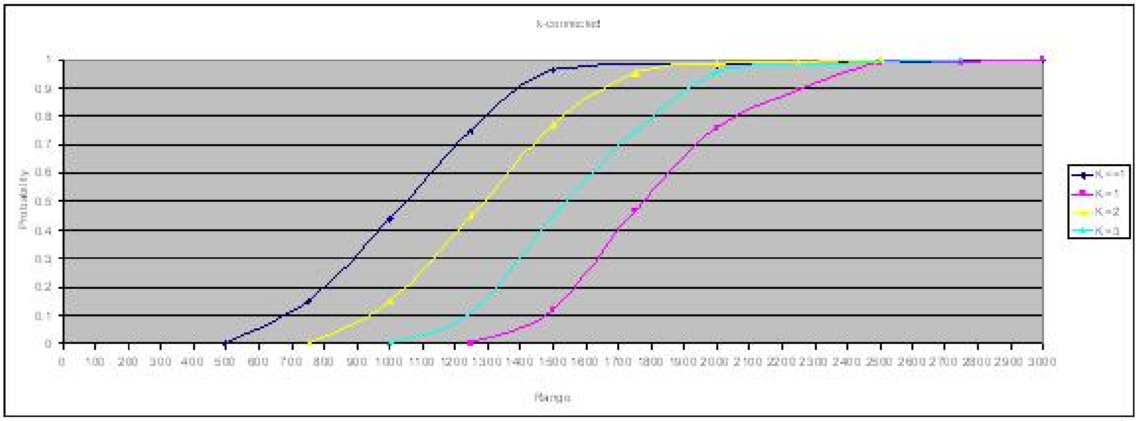


Figure 14: K connectivity test I

TEST II:

Simulation results for 300 experiments with 200 nodes uniformly distributed on work space Area (70000X 70000) m^2 with range r_0 kept constant.

RESULTS:

A network having a constant range r_0 with $n \leq 25$ nodes is almost surely disconnected and a network with $n \geq 100$ has very high probability to have a fully connected network. As the range increases the probability of connectivity between the nodes also increases even with less number of nodes. If the number of nodes n = 200 and range r_0 = 10000 an almost surely connected network is obtained.

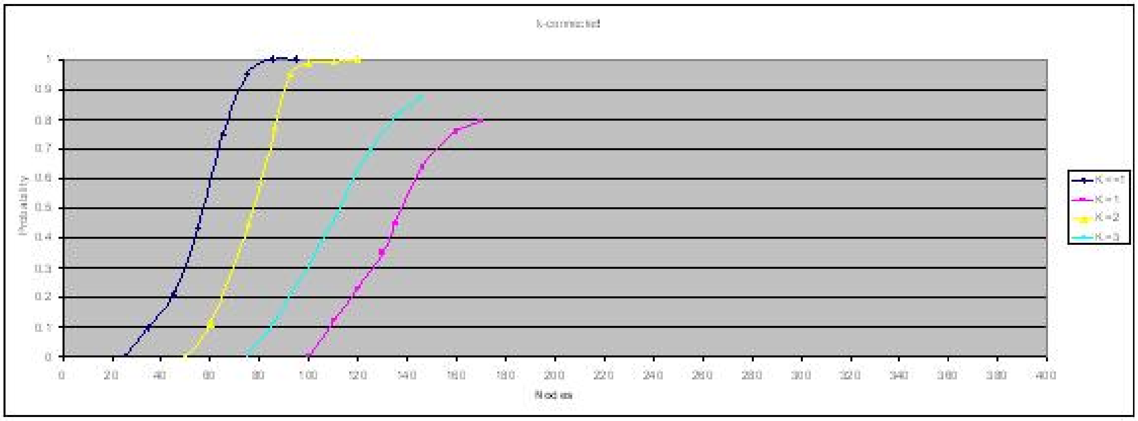


Figure 15: K connected test II

4.6 CONNECTIVITY:

In our thesis, through simulation, we have examined Penrose's theorem on k-connectivity for a geometric random graph. Penrose's theorem [15] states that considering a bounded region in space of d dimensions, with d >= 2, and n uniformly randomly distributed points in it. If P denotes the probability that for a minimum r > 0, a graph created by connecting all points which are r distance apart is k connected, then as n → infinity, P → 1.

Through extensive simulation, for a homogeneous distribution of points, we have examined that in a two dimensional bounded region for a constant distance r and n <= 25, the probability P that the nodes are k-connected → 0 and for n >= 200, the probability P that the nodes are k-connected → 1. We have also examined that as this distance r increases, the probability P that the nodes are k-connected → 1 at lesser values of n (n = 100).

CHAPTER V

CONCLUSION

In this paper we have combined two schemes first is location based algorithm which binds private keys of individual nodes to both their IDs and concrete geographic locations and secondly Bloom's scheme which shares a pairwise key between two nodes if and only if they are authenticate neighbors and share a common key. Our combined scheme has number of appealing properties. First our scheme is scalable and flexible, and nodes do not need to be deployed at the same time, they can be added after initial deployment, and still be able to establish secret keys with existing nodes. Compared to previous predistribution schemes, our scheme is substantially more resilient against node capture.

We have proposed few other models which enables the connectivity of a network with random node distribution and also modeled a network as geometric random graph and derived an probabilistic expression that calculates the probability that if two nodes within range share a common key, given its transmission range r_0, density ρ and number of nodes in the work space. Also if maximum range r0 of the nodes are given, we can find out how many nodes are required to cover a certain area with a k-connected network. Simulations are run for each scenario by modifying input parameters. Results obtained by simulations can be used by researchers and developers who want to further run simulations on these satellite networks. Simulations results can also be implemented in real time satellite systems and also in adhoc networks. Finally the results of this paper can be used to any kind of multihop network.

REFERENCES

[1]. K Hogie, E Criscuolo, R. Parise, “Using Standard Internet protocols and applications in space” (2005) *Computer Networks*, 47 (5), pp. 603-650.

[2]. L P. Clare, J L Gao, E H Jennings, C. Okino, “Space based multi hop networking”, *Computer Networks*, Volume 47, Issue 5, 5 April 2005, Pages 701-724.

[3]. M. Yang, J.-F. Ru, X.R. Li, H. Chen and A. Bashi, "Predicting Internet End-to-End Delay: A Multiple-Model Approach," in Proc. of *IEEE INFOCOM 2005*, Vol. 4, pp. 2815-2819 , Miami, Mar. 2005.

[4].Modadugu,N.,Rescorla,E.,"The Design and implementation of Datagram TLS", Proceedings of ISOC NDSS 2004, Feb2004.

[5]. Robert D.,SCPS, http://www.scps.org/Documents/SCPSoverview.PDF , 6 May 1998.

[6] Vamsi MaramReddy, Osazuwa Amadasun, Venkatesh Sarangan, and Johnson Thomas, **“**Routing in Deep-space networks with lossy links**”** , *2007 IEEE Aerospace Conference, Big Sky, Montana*, March 2007.

[7] Yanchao Zhang, Wei Liu, Wenjing Lou, and Yuguang Fang, “Location-based compromise-tolerant security mechanisms for wireless sensor networks”, *IEEE Journal on Selected Areas in Communications* (Special Issue on Security in Wireless Ad Hoc Networks), vol. 24, no. 2, pp. 247-260, February 2006

[8] Consultative Committee for Space Data Systems, Space Communications Protocol Specification (SCPS)—Rationale, Requirements, and Application Notes, CCSDS 710.0-G-0.3, April 1997**.**

[9]Space Communications Protocol Specification (SCPS)—Security Protocol (SCPS-SP). Blue www.ccsds.org.. May 1999.

[10] W. Du, J. Deng, Y. S. Han, P. K. Varshney, J. Katz, and A. Khalili, “A Pairwise Key Pre-Distribution Scheme for Wireless Sensor Networks,” *ACM Transactions on Information and System Security*, vol. 8, no. 2, pp. 228-58, May 2005.

[11] M. Penrose, “On k-connectivity for a geometric random graph,” *Wiley Random Structures and Algorithms*, vol. 15, no. 2, pp. 145–164, 1999.

[12] C. Bettstetter, “On the minimum node degree and connectivity of a wireless multihop network,” in *Proceedings of the 3rd ACM International Symposium on Mobile Ad Hoc Networking & Computing*. New York:ACM Press, 2002, pp. 80–91.

[13] Probability theory. “Encyclopedia_Britannica”_. 2007. Encyclopedia Britannica Online. 27 Sept. 2007 <http://www.britannica.com/eb/article-9375936>.

[14] A.V. Skorokhod , Basic Principles and Applications of Probability Theory, Springer-Verlag Berlin Heidelberg, 2005.

[15] Vladimir Rotar , Probability Theory, World Scientific Publishing, MA 1997.

[16] D. Stirzaker , Probability and Random Variables: A Beginner’s guide, Cambridge University Press, UK, 1999

[17] M.N.S. Swamy and K. Thulasiraman, Graphs: Theory and Algorithms, Wiley-Interscience, New York, 1992.

VITA

ARAVIND KUMAR
Candidate for the Degree of

Master of Science

Thesis: MODEL FOR SECURE DATA TRANSMISSION IN DEEP SPACE NETWORKS

Major Field: COMPUTER SCIENCE

Biographical:

Education:

1) MS – Computer Science expected day of Graduation May, 2007
2) MS - Information Technology, Anna University, India.(GPA 3.8/4)
3) Post Graduate Diploma in Computer Science, Madras University, India.
4) BS in Mathematics, Madras University, India

Experience:

1) Graduate Research Assistant, CS, Oklahoma State University
2) Graduate Assistant, CEAT Labs, Oklahoma State University

Name: Aravind Kumar Date of Degree: May, 2007

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: MODEL FOR SECURE DATA TRANSMISSION IN DEEP SPACE NETWORKS

Pages in Study: 70 Candidate for the Degree of Master of Science

Major Field: Computer Science

Scope and Method of Study: The main thrust of space communications to-date has been to provide secure communications between ground mission control and a single spacecraft. Little work has been reported on developing a secure mode of communications in a deep space satellite network. The main objective is to develop an algorithm that can increase the connectivity and security in the communication path of the network.

Findings and Conclusions: The proposed model provides a two tiered authentication scheme composed of a first level of location based authentication followed by a pairwise key sharing scheme. We have proposed models which analyze the connectivity of a network with random node distribution. We modeled a network as a geometric random graph and derive a probabilistic expression to determine if two nodes within range share a pairwise secret key between them. Simulations were done to determine the k-connectedness of the network given the communications range r_0 of the nodes.

ADVISER’S APPROVAL: Dr. Johnson Thomas