A Novel Security Framework based on Genetics for Clustered Wireless Sensor Networks

P.V. Ranjith Kumar[1], Sandeep P. Nemagoud[2]
Sandeep Kumar E[3], Vijaya Kumar B.P[4]
Dept. of Electronics & Communication Engg. [1, 2], Dept. of Telecommunication Engg.[3],
Dept. of Information Science & Engg.[4],
M.S Ramaiah Institute of Technology
Bangalore, Karnataka, India

ABSTRACT
Security is of a prime importance in Wireless Sensor Networks (WSN). Because of less human intervention and self- management aspects of these types of networks, they are often prone to security threats and vulnerabilities. In such network scenario, we propose genetics based random key distribution scheme for securing clustered WSNs. The algorithm is simulated in MATLAB. The obtained result proves that the proposed method is energy efficient than other widely used cryptographic techniques and using these types of bio- inspired intelligences provides robust security in the network.

General Terms
Bio- inspired intelligence, security protocols, cryptography.

Keywords
Clustered wireless sensor networks, computational intelligence, random keying technique, genetic computing

1. INTRODUCTION
Research in WSN is gaining lot of momentum in the present scenario. This is because of the widespread applications of these networks. Few of them include military area monitoring, forest fire and pollution monitoring, industrial fault monitoring etc. Majority of the applications are with less human intervention. Hence, these networks are always in the risk of security threats and attacks. In addition, this area is less experimented in WSNs. Still there are many security attacks, which are left unsolved by researchers. In this context, we propose a novel genetics based security framework to combat against spoofing attacks. There are few researches using genetics to solve WSN issues.

Shanthini et al. [1] propose a scheme of using genetic operators to secure the keys generated using biometrics of the user for health care applications. Chien- Lung et al. [2] propose a method of keying technique based on Genetic Algorithm, treating key-generating functions as chromosomes. Rahul et al. [3] propose a method of using Genetic Algorithms for finding optimal positions there by reduction in the power consumption of the network and reducing the effects of malicious nodes for security. Radhika et al. [4] propose an Intruder Detection System based on Genetic Algorithm (GA) for detecting the misbehaviors based on node attributes. Sandeep et al. [5] propose a novel artificial immune system based random keying technique for clustered sensor networks.

There are many issues like clustering, optimal route establishment, localization, node deployment etc. being solved using GA [7] [8] [9] [10] [11].

The proposed protocol is a combination of random key distribution technique with genetics for detecting spoofing packets in the network, trying to alter the normal network operation. The algorithm was simulated in MATLAB and the results prove that the protocol is robust in combating against these types of attacks in WSNs. In addition, the obtained results prove that it is energy efficient than other widely used cryptographic techniques.

The rest of the paper is organized as follows: section 2 deals with genetics related concepts, section 3 deals with the radio model, section 4 describes the proposed methodology, section 5 briefs the attack scenario, section 6 and 7 discusses the simulations and results respectively, section 8 with conclusion and finally the paper ends with few references.

2. GENETIC ALGORITHM (GA)
Genetic Algorithms (GA) is a search heuristic that imitates the natural selection process of nature. They belong to the class of evolutionary algorithms using techniques such as inheritance, mutation, selection and crossover for solving optimal solution problems. The steps in genetic algorithm are explained as follows:

Initialization of Genes: Here the set of chromosomes are considered as population. The size of population depends on the nature of the problem, and the individuals (chromosomes) are called as the solutions in the solution space (pool) and usually pool is generated in random allowing the entire range of possible solutions.

Selection: During each successive generation, portions of existing population are selected to breed a new generation. Individual solutions are selected through fitness-based process, where fittest solutions are more likely to be selected. They are biological imitations of the fittest chromosomes selected for reproduction process.

Crossover and Mutation: These are the genetic operators. They imitate the biological chromosomal crossover and mutation resulting in child solutions. The tuning of genetic operators can be done based on mutation probability, crossover probability, etc. during the process of the crossover the genes exchange takes place in between parental chromosomes and mutation leads to slight deviation from the parental characters biologically.

Termination: the above process stops, if solution found meets the objective.

Genetic Algorithm:

1. Start

2. Randomly generate a population of N chromosomes.
3. Calculate the fitness of all chromosomes.
4. Create a new population:

   a)**Selection:** According to the selection method implemented, select two chromosomes from the population.

   b)**Crossover:** Perform crossover on the two chromosomes selected.

   c)**Mutation:** Perform mutation on the chromosomes obtained.

5. Replace the current population with the new population.
6. Test whether the termination condition is satisfied. If so, stop. If not, return the best solution in current population and go to Step 2.
7. Stop

3. RADIO MODEL

The proposed methodology uses a classical radio model. The sensor node is a transceiver. Hence, this radio model gives the energy consumed for the transmission and reception. The block diagram representation is shown in fig. 1. The radio model consists of transmitter and receiver equivalent of the nodes separated by the distance'd'. Where E_{tx}, E_{rx} are the energy consumed in the transmitter and the receive**r** electronics. E_{amp} is the energy consumed in the transmitter amplifier in general, and it depends on the type of propagation model chosen either free space or multipath with the acceptable bit error rate. We consider E_{fs} for free space propagation and E_{mp} for multipath propagation as the energy consumed in the amplifier circuitry. The transmitter and the receiver electronics depends on digital coding, modulation, filtering and spreading of data. Additional to this there is an aggregation energy consumption of E_{agg} per bit if the node is cluster head.

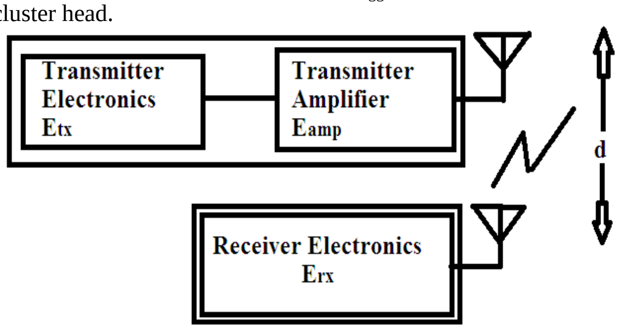

Fig.1 Radio Model

3.1 Energy Consumption

This section describes the energy consumed for communication.

3.2 Packet transmission

$$E_t = (L_P * E_{tx}) + (L_P * E_{amp} * d^n); \qquad (1)$$

Where, $L_P \rightarrow$ is the packet length in bits

$n \rightarrow$ is the path loss component, which is 2 for free space and 4 for multipath propagation.

Suppose a node transmits a packet. Each bit in a packet consumes E_{tx} amount of transmitter electronics energy, E_{amp} amount of amplifier energy. A packet of length L_p, consumes an overall energy of E_t.

3.3 Packet reception

$$E_r = (L_P * E_{rx}); \qquad (2)$$

Where, $L_P \rightarrow$ is the packet length in bits.

Suppose a node receives a packet. Each bit in a packet consumes E_{rx} amount of receiver electronics energy. A packet of length L_p, consumes an overall energy of E_r.

4. PROPOSED METHODOLOGY

This section highlights the method adapted for identifying the spoofed packets. The protocol is designed for clustered WSNs. A single hop network is shown in fig. 2.

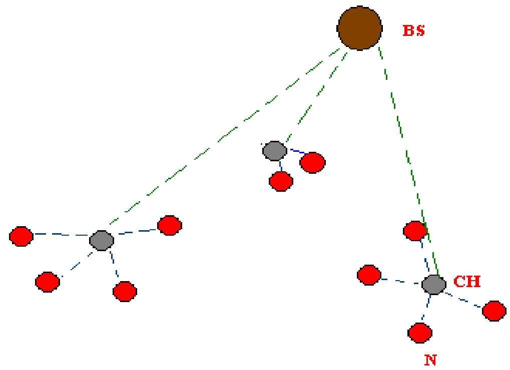

Fig.2 Single Hop Wireless Sensor Network

The identification of spoofing is done according to the notion of genetics. Here, the gene theory is not used for finding any optimal solution, instead modified in such a way that it can be applied to tackle security issues in WSN. Initially at the BS the range is set. Let this be (A,C). The integer numbers within the range indicates the chromosomes. The range is again optimized randomly for a particular round. Out of this pool, again randomly two intergers will be picked and assigned to the CHs. The integers within the range of the numbers sent, at each CH, is called the fittest pool of individuals in genetics paradigm. The same range is broadcasted with the member nodes also. The ordinary node while communicating back, picks two numbers from the pool, does the crossover, mutation and places the result in the trailer and sends the packet to the higher hiearchical node. After the steady phase ends, the BS again dessiminates new ranges to all the CHs at the end of set-up phase of other round, and the above discussed scenario repeats. The step wise description of the process is given below.

Algorithm 1: Normal Communication paradigm

1. Start
2. Define a range in the Base Station (BS). Let the range be (A, C). /* A be the lower limit and C be the upper limit*/
3. Scale down the range at the BS. The re-scaling of range is done using equations (3), which is a novel formula used in the BS.

   A= A+ (B~A)*rand; (3)

   C= B+ (C~B)*rand;

   Where, B= (A+C)/2, which serves to be the mid-key value within the range.
4. Pick two random numbers between (A, C) and assign it to CHs of that particular round. Let this be (x, y). The CHs broadcasts the received range to its member nodes. Step 2 and step 3 happens soon after the set- up phase of the network.
5. Suppose an ordinary node wants to communicate with the CH, then the node picks two random numbers p and q such that p>=x and q<=y. Apply

genetic operator i.e. (g, h) = crossover (p, q) and (g1, h1) = mutate (g, h). p and q will be placed in the header and (g1, h1) will be placed as the trailer of the packet and sent to the CH.

6. Suppose CH wants to communicate with the BS, repeat same process as in step 5 and send the packet to the BS.
7. Stop

Algorithm 2: **Crossover** (num1, num2)

1. Start
2. Input two numbers
3. Convert the decimal numbers to binary
4. Mutual exchange the bits from the center position and are shown in example 1.
   Eg. 1: Let num1= 64 and num2= 84;
   num1 = 0100 0000, num2= 0101 0100
   Cross over is given by:
   0100 0000
   0101 0100

   0100 0100 = 68
   0101 0000 = 80
5. Stop

Algorithm 3: **Mutation** (num1, num2)

1. Start
2. Input two numbers
3. Convert the numbers to binary
4. The binary numbers will complement themselves creating a mutated scenario for input number.
5. Stop

Algorithm 4: Verification at the CH for the packet sent by ordinary node or Verification at the BS for the packet sent by CH.

1. Start
2. Receive the packet
3. Extract header
4. Check header, whether it is in the range that was sent by itself. **Let the received keys in header be m, n and trailer be t1, t2.**
   If (m>= x and n<=y)
   {
   /* packet cleared stage-1*/
   (g, h) = **Crossover** (m, n); /*g and h are results of crossover*/
   (g1, h1)= **Mutation** (g, h);
   If (g1==t1 and h1==t2)
   {
   /* packet cleared stage-2*/
   Else
   /* declare packet is malicious*/
   }
   }
   Else
   {
   /* declare packet is malicious*/
   }
5. Stop

4.1 Packet Description

a. Packet sent from BS to CH/ CH to its member nodes

MAC	x	y

where, MAC→ MAC address of the intended CH node
x, y→ keys randomly picked by the BS for a CH.

b. Packet sent from ordinary node to CH/ CH to BS

This packet consists of the details regarding randomly picked keys by the node and the trailer.

p	q	CRITICAL INFO	g1	h1

where, p, q→ keys randomly picked by the node for communication with its CH and g1, h1 are the trailers after crossover and mutation, CRITICAL INFO→ consists of various fields including, preamble, sync bits, destination address, type, group identity, length of message, counter for message sent, source address, error checking bits, payload.

5. ATTACK SCENARIO

The system relies on confusing the intruder by randomly varying the keys and ranges chosen for selecting the keys at the BS and nodes. The newly deployed malicious attacker may spoof unwanted packets to the CH or the BS. The attack scenarios are shown in the fig. 3 and fig.4.

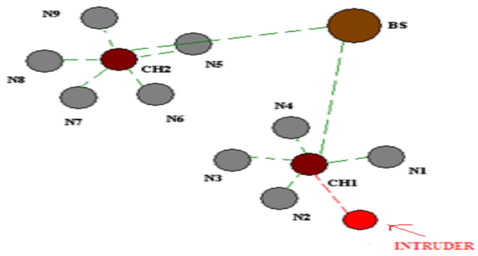

Fig.3 Malicious ordinary node sending a false packet to CH

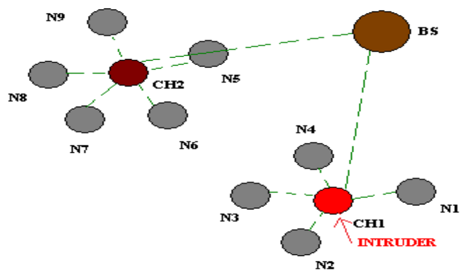

Fig.4 Malicious CH sending a false packet to the Base Station

The new node carefully listens to the network paradigm and assigns its MAC address with that of another node, of which it may start disguising and spoofing packets to the higher hierarchical node. The packets follow the double verification steps and gets identified itself as either a legitimate or a spoofed packet. Suppose, the count of spoofed packets reaches above a pre-fixed threshold, an alarm is sent to the BS for preventing the further epidemic of the infected packet.

The spoofing can also be done at time by the legitimate nodes already deployed. The spoofing in this case can also be detected by the proposed methodology.

Since, the protocol protects the network using randomization concept, the attack not being identified is minimal. One scenario of attack was modeled in this paper, where a malicious node listens to the paradigm of the network and gets to know about the key ranges i.e. the keys are falling within the range (A, C), and puts header of the packet with those numbers and trailers with some random numbers. In this case, there are chances that the packet may pass the first verification stage, but the second stage clearance is difficult since the numbers in the headers has to undergo crossover, mutation and results has to match with the trailers. The results obtained for this scenario of attack is discussed in the fig. 5, fig.6, fig.7 and fig.8. Apart from this case, if the malicious node has to successfully spoof the packet in every attack, then it has to get the algorithmic and mathematical details burnt in the node, which is the case of a node capture attack. The protocol fails if the node undergoes a capture attack and the security details are hacked.

6. SIMULATIONS

The algorithm was executed and tested using MATLAB 2013a on Intel core 5 Duo processor with windows operating system. CH requirement was set to 10% and the algorithm was verified on LEACH protocol for 100 and 500 rounds respectively. Table 1 contains the overhead in packet size due to the proposed security algorithm and table 2 depicts the various key sizes used for simulation. The parameters were set for modeling network environment is shown in table 3. The key sizes of ECC and RSA is shown in table 4, and of which the basic key size of 112 for ECC and 512 for RSA were considered for energy analysis.

TABLE 1. Bits overhead due to cryptographic framework (per communication)

Parameter	Value
Packet sent from BS to CHs	32 bits
Packet sent from CH to ordinary node	32 bits
Packet sent from end node to CH	32 bits
Packet sent from CH to BS	32 bits

TABLE 2. Key sizes used in packets for communication

Parameter	Size
x, y	1 byte each
MAC	2 bytes
p, q	1 byte each
g1, h2	1 byte each

TABLE 3. Radio characteristics and other parameters chosen for simulation

Parameter	Value
Number of nodes	100
Transmitter electronics, E_{tx}	50nJ/bit
Receiver electronics, E_{rx}	50nJ/bit
E_{mp}	0.0013pJ/bit
E_{fs}	10pJ/bit
E_{agg}	5nJ / bit
Length of plot	100 m
Width of plot	100 m
L_{pt} (packet sent from CH to BS)	6400 bits
L_{ctr} (packet sent from ordinary node to CH)	200 bits
Initial energy of the node	0.5 J

TABLE 4. RSA and ECC key length comparison

RSA	ECC
512	112
1024	160
2048	224
3072	256
7680	384
15360	512

7. RESULTS AND DISCUSSIONS

This section deals with the results obtained. The algorithm was tested on LEACH protocol. First six iterations are for analyzing the security, where number of rounds was limited to 100 in every iteration. Next, were six iterations each with 500 rounds. In both cases, after every fifth round a malicious packet was made to spoof into the network, the probability of being identified is checked, and the graph is plotted. The obtained results are plotted in the fig.5 and fig.6. It was observed that in both the cases, the accuracy in identifying the malicious packets was 100%.

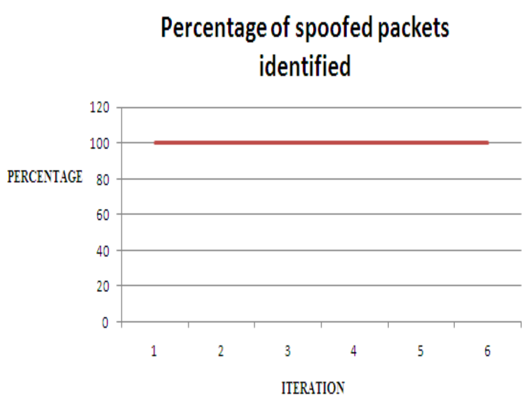

Fig.5 Percentage of spoofed packets identified for six iterations (each for 100 rounds of LEACH)

In addition, in both the cases the number of packets clearing first stage and second stage of verification was plotted separately. It was observed that even though the packets clear first stage, it was likely that they were caught in the second stage of verification; hence, the accuracy was always 100% and shows the robustness of the protocol in identifying the spoofed packets. The energy consumption analysis of our scheme with the existing cryptographic schemes like ECC and RSA was done for 100 rounds of LEACH and the additional overhead in the communication energy consumption is plotted in fig.9.

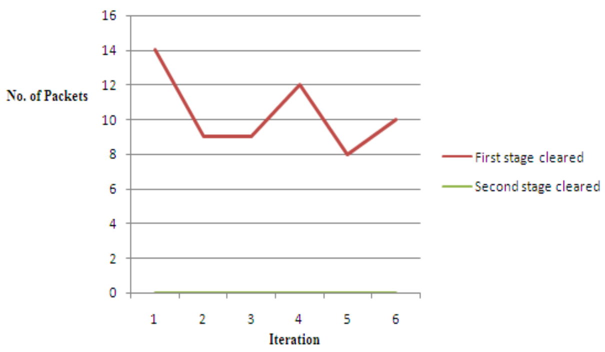

Fig.6 Number of packets cleared first stage and second stage, for six iterations (each for 100 rounds of LEACH)

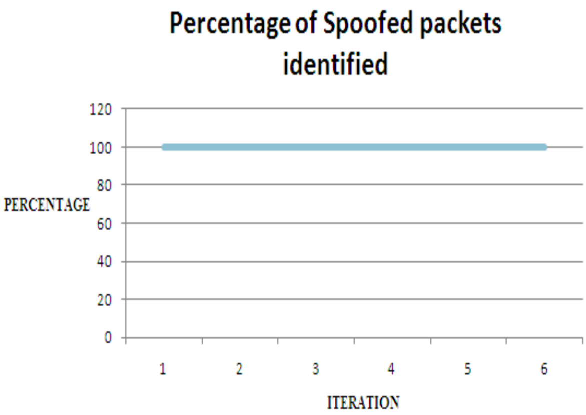

Fig.7 Percentage of spoofed packets identified for six iterations (each for 500 rounds of LEACH)

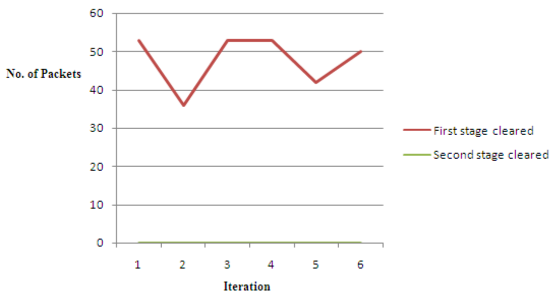

Fig.8 Number of packets cleared first stage and second stage, for six iterations (each for 500 rounds of LEACH)

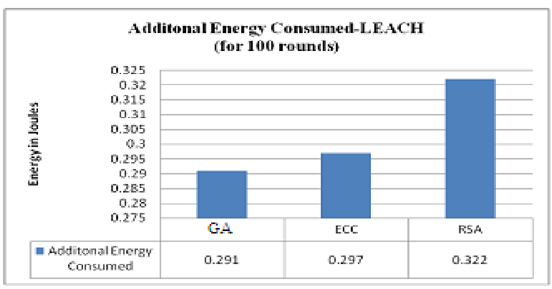

Fig.9 Additional Energy consumed for 100 rounds between various cryptographic techniques

It was observed from fig.9, that the energy overhead due to the keys used for security was more in other keying techniques compared to GA based random keying technique.

8. CONCLUSIONS

In this paper, we propose a novel random keying technique using the concept of genetics. The percentage of malicious packets identified in the network was 100%, since the protocol relies on double verification and randomization process and it was observed that it is energy efficient compared to the other widely used cryptographic schemes. Due to the use of concepts of Artificial Intelligence, there will always be a tradeoff between the security and computational overhead also. Hence, from the simulations it was proved that the proposed methodology could be implemented for the future sensor networks meeting the security concerns of the network.

9. ACKNOWLEDGEMENTS

Authors like to thank Dept. of Information Science & Engg., M.S Ramaiah Institute of Technology for providing lab facilities for conducting the research work. Authors also like to thank Management and Dr. S.Y. Kulkarni, Principal of M.S Ramaiah Institute of Technology, for their constant support to carry the prospective research work.

10. REFERENCES

[1] Shanthini. B and Swamynathan.S, "Genetic- based Biometric Security System for Wireless-Sensor-based Healthcare Systems", International Conference on Recent Advances in Computing and Software Systems (RACSS), pp 180-184, 2012.

[2] Chein- Lung Wang, Tzung- Pei Hong, Gwoboa Hoing, Wen- Hung Wang, "A GA- Based Key- Management Scheme in Hierarchical Wireless Sensor Networks" International Journal of Innovative Computing, Information and Control, vol. 5, pp 4693-4702, 2009.

[3] Rahul Khanna, Huaping Liu, Hsio- Hwa Chen, "Dynamic Optimization of Secure Mobile Sensor Networks: A Genetic Algorithm", IEEE International Conference on Communications (ICC) – 2007, pp 3413-3418.

[4] Radhika Basicar, P.C Kishore Raja, Christeena Joseph, M. Reji, " Node Attribute Behavior based Intrusion Detection in Sensor Networks", International Journal of Engineering and Technology (IJET), Vol. 5(5), pp 3692-3698, 2013.

[5] E. Sandeep Kumar, S.M. Kusuma, B.P. Vijaya Kumar, "A Random Key Distribution based Artificial Immune System for Security in Wireless Sensor Networks", Proceeding of IEEE International Students' Conference on Electronics, Electrical and Computer Science (SCEECS)-2014, 1-2 March, MANIT, Bhopal, Madhya Pradesh.

[6] E. Sandeep Kumar, S.M. Kusuma, B.P. Vijaya Kumar, "An Intelligent Defense Mechanism for Security in Wireless Sensor Networks", Proceedings of IEEE International Conference on Communications and Signal Processing (ICCSP) - 2014, 3-5 April, APEC, Melmaruvattur, Tamil Nadu.

[7] Omar Banimelhem, Moad Mowafi, Walid Aljoby, " Genetic Algorithm based Node Deployment in Hybrid Wireless Sensor Networks", Journal of Computer

Science and Communications, Vol. 5(4), pp 273-279, 2013.

[8] Suneet K, Gupta, Pratyay Kuila, Prasanta K. Jana, "GAR: An Energy Efficient GA- based Routing for Wireless Sensor Networks", Distributed Computing and Internet Technology, Lecture notes in Computer Science, Vol. 7753, pp 267-277, 2013.

[9] Vinay Kumar Singh, Vidushi Sharma, "Lifetime Maximization of Wireless Sensor Networks using Improved Genetic Algorithm based Approach", International Journal of Computer Applications (IJCA), Vol. 57(14), pp 36-40, 2012.

[10] Amol p. Bhondekar, Renu Vig, Madan Lal Singha, C. Ghanshyam, Pawan Kapur, " Genetic Algorithm Based Node Placement Methodology for Wireless Sensor Networks", Proc. of International MultiConference of Engineers and Computer Scientists, Hong- Kong, 2009.

[11] Xinyu WANG, Ziwen SUN, Zhicheng JI, " Genetic Algorithm for Wireless Sensor Network Localization with Level- based Reliability Scheme", Journal of Computational Informational Systems, Vol. 9(16), pp 6479- 6488, 2013.

[12] Wendi Rabiner Heinzelman, Anantha Chandrakasan and Hari Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks", Proc. of 33rd IEEE Hawaii International Conference on System Sciences – 2000.