Security Optimization and Data Classification in Wireless Sensor Networks

Dilek Karabudak

Computer Engineering Department
Bilkent University
Ankara, TR 06530
Email: dilekk@cs.bilkent.edu.tr

Abstract

Large populations of wireless connected nodes, capable of computation, communication and sensing constitute wireless sensor networks (WSNs). Since increasing number of applications have been widely deployed using WSNs, selection of the best type of secure data transmission for WSNs becomes one of the most important issues among other challenges. Besides, there is not any secure sensor network protocols proposed using different encryption algorithms at a time depending on a quality of service (QoS) requirement in the literature. However, there is a need for an alternative that brings the optimum, flexible and efficient solution for secure data transmission. Intelligent optimization algorithms can address this problem. In addition to providing a secure data transmission, efficient data classification is a crucial issue in sensor networks in order to obtain accurate data and reduce the communication overhead. In this paper, previously proposed schemes for secure wireless sensor networks are investigated. Furthermore, an optimization algorithm using genetic algorithms for secure transmission is proposed for WSNs. Besides, different classification algorithms are experimented to find an efficient, fast and accurate sensor data classification algorithm. Experiments are performed for these two type of sensor network issues and the results are compared in terms of time and complexity efficiency. Performance analysis is provided to assess the efficiency of the proposed algorithms.

I. Introduction

The rapid advances in micro-electro-mechanical (MEMS), digital electronics and wireless communication technology have enabled the development of distributed networks of small, inexpensive nodes that are capable of sensing, computation, and wireless communication [2]. They are designed to be deployed for a broad range of environmental sensing applications from vehicle tracking to habitat monitoring. Furthermore, sensor networks of the future are envisioned to develop the paradigm of collecting and processing information in very diverse and heterogeneous environments. However, the limited computing resources, severe energy constraints of the sensors and the need for a secure data transmission especially in adversary environments for military applications, present major challenges for such a vision. All of these challenges are need to be addressed.

One of the key challenges, which needs to be addressed, is secure data communication for Wireless Sensor Networks (WSNs). Security in data transmission is an important issue to be considered while designing wireless sensor networks. Security protocols proposed in the literature only deal with a particular encryption algorithm that encodes the data packets transmitted among the sensor nodes. However, all these schemes do not consider the deployment of more than one encryption algorithms at a time for better security in data transmission. Thus, a scheme considering both the efficiency and reliability of the data transmission and the computational efficiency of the encoding for the data packets used in WSNs should be proposed. To fulfill all the security requirements of WSNs, a optimized security scheme is required. The optimized scheme should consider both the transmission and encoding of the data to address these requirements. Intelligent optimization techniques are an efficient way of solving this problem.

Artificial intelligence techniques are promising for their life-like ability to self-replicate as well as the adaptive ability to learn and control the environment. Among these techniques, genetic algorithms (GAs) have been used in a wide variety of optimization tasks, including numerical optimization and combinatorial optimization problems. There are also several optimization techniques such as simulated annealing, tabu search, etc. other than GAs. However, GAs' ability for parallel searching, fast convergence and fast evaluation distinguish itself from other decision and optimization algorithms.

The other key challenge of WSNs is the data classification and aggregation for the data gathered from the sensor nodes. Since, the sensor nodes are lack of large amounts of power and computation capability, it is a

great waste of resources to monitor, gather, send, receive and process huge amounts of data for WSNs. Hence, a fast, accurate and efficient classification algorithm will help to decrease the number of data transmitted over the network and the communication overhead.

In this paper, an optimized security scheme that addresses the security requirements of WSNs using GAs is proposed and different classifier algorithms are investigated. For the optimized security scheme, a linear cost function is defined consists of different encryption algorithm and sensor network parameters. Then this cost function is optimized using GAs. According to the best solution, an optimum encryption algorithm is selected to encode the data transmitted. The simulation and performance analysis show that the scheme achieves a secure transmission with a very low latency and cost value. In addition to that, for finding the best data classification algorithm, a variety of classifier algorithms are explored and experimented for a large set of sensor data. Instance Based (IB1) algorithm seems to be the most efficient algorithms in terms of time and accuracy.

The remainder of the paper is organized as follows: The previously proposed security schemes are presented in Section II. The background for wireless sensor networks is given in Section III. The optimized security system modeling is presented in Section IV. The data classifier algorithms are briefly described in Section V. The validation of the model, its evaluation results along with the effects of the optimized security algorithm on the network performance and latency and the data classification results are then discussed in Section VI. Finally, conclusions are provided in Section VII.

II. Related Work

A very few proposed data security solutions designed for wireless sensor networks only developed for only cluster-based sensor networks using simple data encryption [3], [4], [7], [10]. [3] and [4] proposes a solution using simple symmetric cryptographic algorithms. It is because, asymmetric cryptographic algorithms are not suitable for providing security on wireless sensor networks due to limited computation, power, and storage resources available on sensor nodes. [7] only covers some implementations of an existing security algorithm proposed for wireless systems. Although these schemes are promising, they do not specifically consider data security as a means to provide a unified and efficient scheme for all types of wireless sensor networks for maximum reliability and security.

Besides, there are several energy efficient data transmission and data aggregation protocols that provide energy efficient solutions [9], [14], [15]. In [15], the authors introduce just a framework that is based on convey tree sequence. To the best of our knowledge, there is not any research in the literature investigates different classifier algorithms to find the optimum and efficient solution for data classification and aggregation.

III. Background

Sensor network refers to a heterogeneous system consist of tiny sensors and actuators with general purpose computing elements. It combines hundreds or thousands of low-power, low-cost nodes, possibly mobile but more likely fixed locations deployed to monitor and affect the environment. There are several types of sensor/actuator nodes manufactured in MEMS technology but especially two of the prototypes are well-known. One of them is Mica mote, a small (several cubic inch) sensor/actuator unit with a CPU, power source, radio, and several optional sensing elements. The processor is a 4 Mhz 8-bit Atmel ATMEGA103 CPU with 128 KB of instruction memory, 4 KB of RAM for data, and 512 KB of flash memory. The CPU consumes 5.5 mA (at 3 volts) when active and two orders of magnitude less power when sleeping. The radio is a 916 MHz low-power radio from RFM, delivering up to 40 Kbps bandwidth on a single shared channel and with a range up to a few dozen meters or so. The RFM radio consumes 4.8 mA in receive mode, up to 12 mA in transmit mode, and 5 μA in sleep mode. The other prototype sensor node is deployed in SmartDust project [8], which has also 4 MHz 8-bit CPU with 8KB instruction flash, 512 bytes RAM and 512 bytes EEPROM. It also communicates at 916 MHz radio with about 10 Kbps bandwidth with 3500 bytes OS code and 4500 bytes available code space.

Sensor networks also have centralized control units as in cellular wireless networks called ”base stations” or ”‘actor nodes”. A base station is simply a gateway node to another network, a powerful data processing and storage center, or an access point for different applications. They are also called as ”sinks” in the literature. The position of sensor nodes do not pre-determined. This allows random deployment in inaccessible terrains or disaster relief operations. Hence, sensor network protocols and algorithms should have self-organizing capabilities [1], [2].

Realization of different types of sensor network applications also needs wireless ad-hoc networking techniques. However, there are several differences between sensor networks and ad-hoc networks. These are:

- The number of sensor nodes in a sensor rework can be several orders of magnitude higher than the nodes in an ad hoc network.

- Sensor nodes are densely deployed.
- Sensor nodes are prone to failures.
- The topology of a sensor network changes very frequently.
- Sensor nodes are limited in power, memory and computational capacities.
- Broadcast communication paradigm is mostly used rather than point-to-point communications in ad hoc networks.

IV. Security Optimization Scheme

This research considers a heterogeneous architecture of sensor networks where data may be routed from sensor nodes to base station (actor node) directly. Base stations interface sensor network to the outside network, the sink. The overall system architecture can simply be demonstrated as in Fig. 1. Sensor nodes are assumed to be immobile and also they do not have a specific architecture when deployed over a specific geographic area.

A. Cost Derivation

In the modeling of the problem, the parameters of the wireless sensor network architecture and encryption algorithms that affect the transmission process such as available bandwidth, network bandwidth, packet size, CPU power consumption are considered in the cost function that have to be optimized. The cost function is linearly formulated. Then a final optimality equation is derived for the optimization and encryption decision process which is implemented by genetic algorithms.

$$Cost\ function \Rightarrow (Cost)_E = \mathcal{F}(Cp_E, Sw_E, Pw_E, Bw_E) \quad (1)$$

where

Cp is the computation to encode the data
Sw is the switching, rerouting of the traffic to another encryption algorithm
Pw is the power consumption
Bw is bandwidth of the network
E is the index, $E = 1...n$ where n is the number of different encryption algorithms

Each parameters in the cost function depends on the wireless network architecture in the system. Power consumption cost, Ψ_{Pw} is fixed with coefficient such as

$$\Psi_{Pw}(E) = \psi_{Pw} = c_{Pw} \quad (2)$$

It is assumed that the bandwidth cost rate ψ_{Bw} depends linearly on all the capacity of the network architecture and inversely with the available capacity as follows

$$\psi_{Bw}(E) = c_{Bw} \cdot \frac{C_E}{C(t)} \quad (3)$$

where C_E is the total capacity of the channel of the network and $C(t)$ is the available capacity of the channel at time t. (i.e. if available capacity of the resource is C_E then the cost will be only the bandwidth coefficient). c_{Bw} is the bandwidth cost coefficient per capacity unit.

The computation cost can also be formulated in the same way.

$$\psi_{Cp}(E) = c_{Cp} \cdot \frac{P_E}{P(t)} \quad (4)$$

where this time P_E is the total packet size to be transmitted and $P(t)$ is the packet size already sent at time t. c_{Cp} is the bandwidth cost coefficient per capacity unit.

Both switching cost coefficient changes with respect to the next encryption algorithm that will be used. Hence, the encryption algorithm decision function can be defined as

$$f_E(E_c, E_x) = \begin{cases} 0 & , E_c = E_x \\ 1 & , E_c \neq E_x \end{cases}$$

where E_c is the current encryption algorithm and N_x is the next encryption algorithm, which the sensor node would probably applies. $f_E(E_c, E_x)$ determines whether the next algorithm, which the sensor node applies, is the same one or not.

$$\psi_{Sw}(E) = c_{Sw}[1 + f_E(E_c, E_x)] \quad (5)$$

where c_{Sw} is the switching cost coefficient.

Then the cost function can be defined as

$$\mathcal{F} = \sum_{1}^{\delta} \Psi_E \quad (6)$$

where δ is the number of cost parameters.

B. The System Solution

In the design of the proposed scheme, first the base stations, the sinks send their QoS requirement to each sensor node that will compute the scheme and decide the optimum encryption algorithm. This requirement covers both the total packet size that the individual sensor node should send to the sink and the total bandwidth available for the network. According to this requirement, the sensor node optimizes the overall cost function derived for the scheme and determines which encryption algorithm that it should apply using GAs. Finally, it encodes and

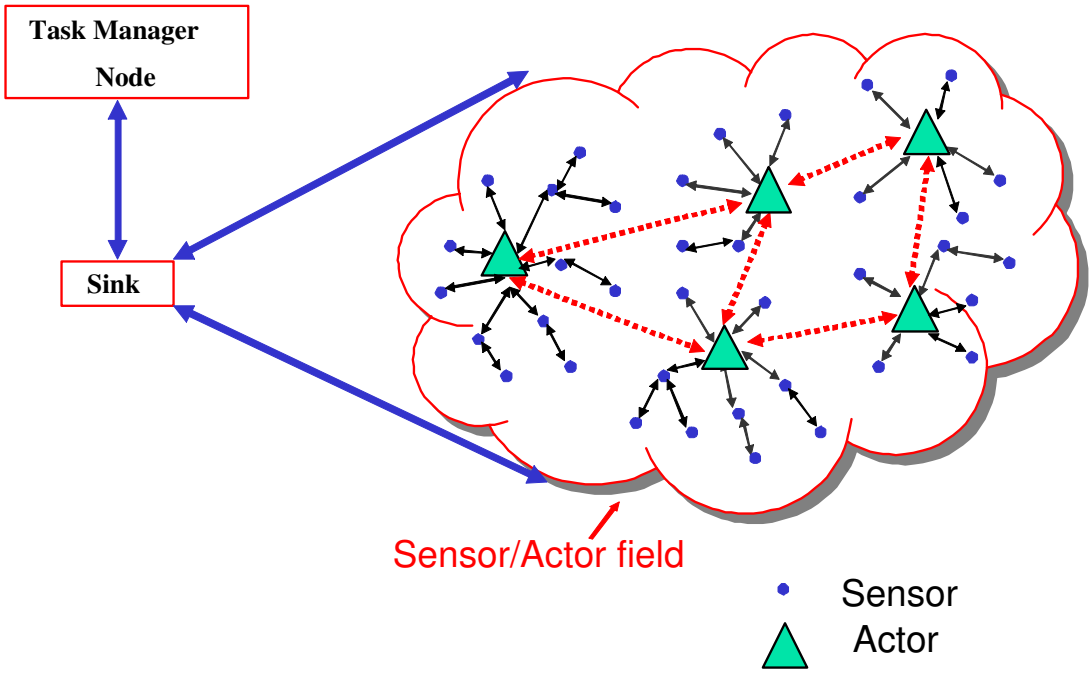

Fig. 1. Wireless Sensor Network Architecture.

transmits data to its sink whether through its actor node or the sink directly.

The key management issue is beyond the scope of this paper. Besides, it is assumed that the initial key is embedded to each sensor node during the manufacturing phase. The other keys used in the algorithms are then generated by the sensor node itself.

C. Encryption Algorithms

During the research, different symmetric encryption algorithms are investigated. For comparison reasons it is assumed that the sensor nodes are capable of processing TEA, RC5, Skipjack and AES encryption algorithms. TEA is a Feistel cipher which uses operations from mixed (orthogonal) algebraic groups. It encrypts 64 data bits at a time using a 128-bit key, the embedded key inside the sensor nodes for this scheme [13]. RC5 is a stream cipher designed by Ron Rivest. The key length can be from 1 to 256 octets. Skipjack encryption algorithm is also a secret key encryption that uses 64 bit blocks and 80 bit keys. Finally, AES provides different block and key size during the encryption process. They can be chosen from 128 to 256 bits. Both encryption algorithms is applied using CBC as the mode of operation for this paper [6].

D. Genetic Algorithms (GAs)

Genetic Algorithms are directed random search techniques used to look for parameters that provide the optimal solution to a problem. They are based on the principles of evolution and natural genetics [5]. As an optimization method, GAs have major differences and advantages over the other optimization algorithms [5]. The notion of genetic algorithms is the *survival of the fittest* of the nature. This implies that the 'fitter' individuals are more likely to survive and have a chance of passing their features to the next generation. In the proposed scheme, GAs is used to solve the final cost function.

The basic operations of GAs are as follows

- **Encoding Scheme** A set of parameters is sought that will give the best solution in optimization. In order to implement GAs, these set of parameters must be encoded into a string so that crossover and mutation operations can be applied [5]. Every encryption algorithm for a sensor node that can be processed represents different search areas for the genetic algorithms. The encoding is not binary, for the simplicity of the solution and to provide more accurate values in a fast manner, the genes have their actual real values. The size of the solution space includes total value ranges of the cost coefficients and the other parameters which constitute the final cost function.
- **Fitness Function Evaluation** The fitness function is used to evaluate the quality of the chromosome [5]. In the proposed scheme, the fitness evaluation function is defined with respect to the cost function.

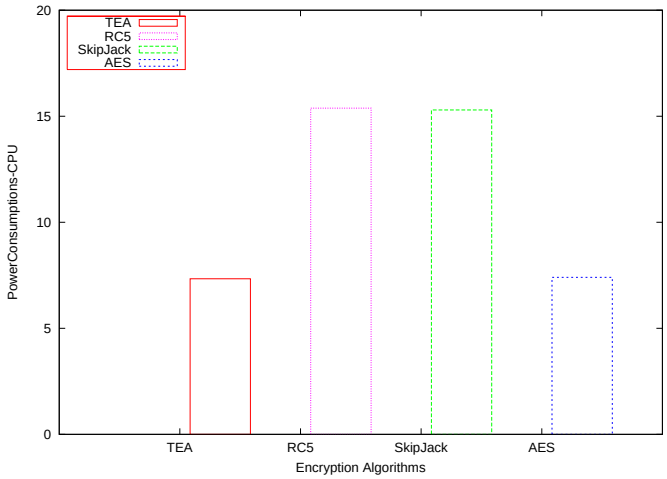

Fig. 2. The CPU power consumptions experienced with different encryption algorithms.

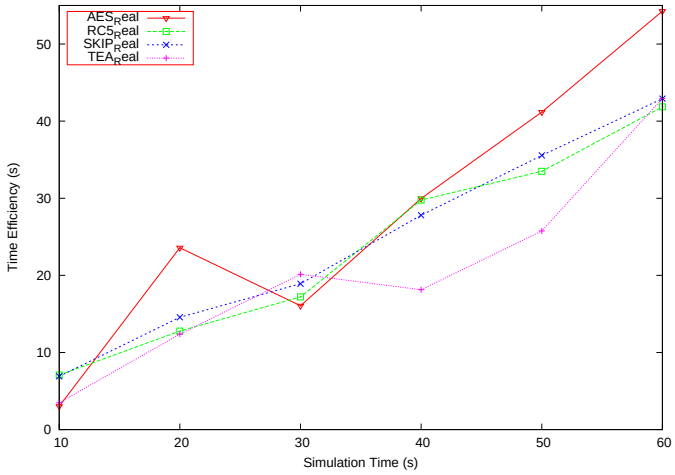

Fig. 3. The real time efficiency is experienced with different encryption algorithms.

As our objective is to find the minimum cost for every encryption algorithm, the fitness function is defined as the inverse of the proposed final cost function.

- **Crossover and Mutation** Crossover is one of the most important operators in genetic algorithms, which creates new candidate solutions for the problem. Another genetic operator is mutation. It introduces new genetic material into a population based on a mutation probability [5]. In the proposed algorithm, the implementations and simulations are performed by the crossover rate, $Xover = 0.7$. This rate is the percent the individual of the new population will be selected randomly and mated in pairs. The algorithm deploys uniform crossover during the operation. This allows the parent chromosomes to be mixed at the gene level rather than the segment level (as with one and two point crossover). Random uniform mutation is used for the proposed scheme. This is achieved by the mutation of the gene to a value chosen from a uniform random variable scaled to the lower and upper bounds of gene range. In our implementations and simulations the mutation rate is set to 0.001 by default.

V. Data Classification

In order to determine the optimum data classification algorithm to deploy to sensor network data, several classifier algorithms are explored. The ones that are mentioned in this paper are

- **ADTree** This classifier is a tree type classifier and also known as alternating decision tree learning algorithm.
- **NBTree** The classifier algorithm forms a Naive Bayes decision tree. The decision tree consists of Naive Bayes classifiers at the each leaf of the tree.
- **PART** PART is a rule base classifier that generates partial decision trees rather that forming one whole

TABLE I

SIMULATION PARAMETERS

Param.	***Value***
# Sim. Time	1000
Area Coord.	2000x2000
NumConn.	2
PacketSize	2000
Interval	100

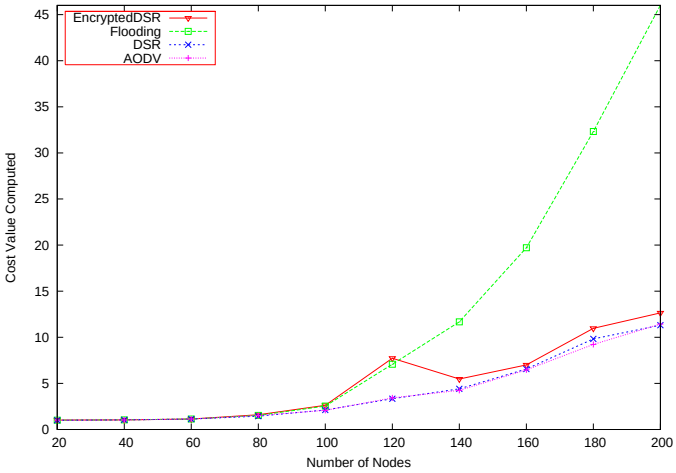

Fig. 4. The cost value experienced with a routing algorithm (DSR) using encryption and decision algorithm using GAs and non-secure sensor schemes using flooding, AODV and DSR routing algorithms for increasing number of sensor nodes.

decision tree in order to achieve the classification.

- **Ridor** The name of the algorithm comes from ”Ripple Down Rule” learner. The algorithm simply based on generating the exceptions for the defined rules and the iteration of these exceptions for the best solution.
- **J48** It is also a rule base classifier algorithm that generates C4.5 decision trees.
- **IB1** This algorithm is known as ”Instance Base Learning Algorithm”. The main idea of this classifier is to use the distance as a metric for the classification.
- **Bayesian Network** Bayesian networks are factored representations of probability distributions. Each attribute in the data is processed independently for this type of classifier algorithm.

VI. PERFORMANCE ANALYSIS

This section demonstrates the performance of the proposed security optimization scheme with simulations and comparison with other non-secure algorithms. The section also gives some comparison results of different classifier algorithms using sensor network data to show data classification performance.

The proposed scheme uses Dynamic Source Routing (DSR) routing algorithm while transmitting encrypted packets over the sensor network. The performance is compared with non-secure sensor network schemes using flooding, ad hoc on demand distance vector routing algorithm (AODV) and DSR routing algorithms. The experiments can not be performed for other security schemes using one encryption algorithm at a time in the literature. This will be the issue for the future work of this research.

The simulations are implemented using a sensor network simulator, TOSSIM, of a sensor node operating system TinyOS for security optimization scheme [11] . In order to investigate the performance, the scheme is deployed for various simulation parameters for varying number of nodes in the sensor network. These parameters are shown in Table I. The classification simulations are experimented using a special software Weka that includes all types of popular machine learning algorithms [12].

A. Security Optimization Performance

The simulation experiments are performed for varying number of nodes as shown in Table I with crossover

TABLE II

GENETIC ALGORITHMS PARAMETERS

Param.	***Value***
# of Population	100
Xover	Uniform
Elitism	no
$Xover_r$	0.7
Mut_r	0.001
Generation	1000

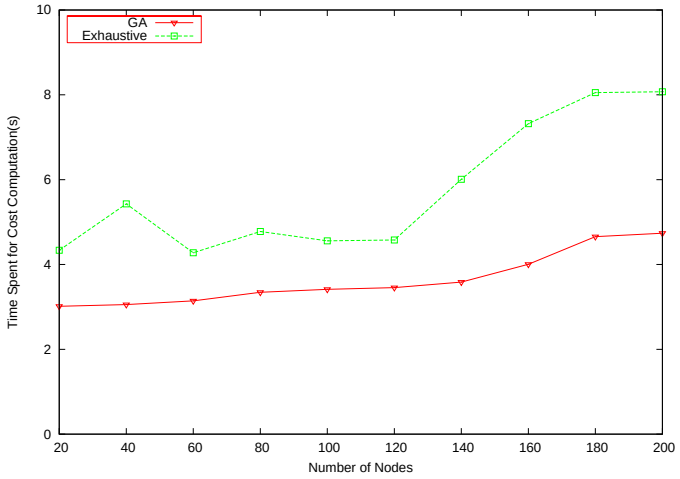

Fig. 5. The latency experienced with a cost computation for encryption algorithm decision using GAs and simple exhaustive search for increasing number of sensor nodes.

probability, $Xover = 0.7$ and mutation probability, $Mut = 0.001$ unless otherwise specified. The other parameter values related to the final optimality function which determine the final cost coefficient values are also shown in Table II. The comparisons of the different encryption algorithms deployed to the sensor data during the simulations in terms of CPU power consumptions and real time efficiencies are shown in Fig. 2 and Fig. 3 respectively. This both gives an idea about the encryption algorithms' capabilities and helps to compute the cost coefficients of the corresponding algorithm in order to use it in the final cost function. For instance, although AES algorithm shows promising power consumption result during the encoding process, it seems not a time efficient algorithm.

As shown in Fig. 4, the proposed scheme gives slightly higher results when compared with other non-secure schemes using AODV and DSR routing algorithms. This is expected since this scheme performs additional encoding operations for encryption. Although under these conditions, the results differs very slightly. This is a trade off of the proposed scheme compromised for a complete secure data transmission. Furthermore, although it has additional computational overhead, the cost value results are much better than the last non-secure algorithm using flooding routing algorithm. This is because of the process time spent and hence the power consumption during the routing of the packets in flooding algorithm. These results demonstrates that the scheme achieves low cost although it also provides a complete secure data communication.

The similar simulation environment is set up to compare the time efficiency of the scheme during the optimization process. Fig. 5 demonstrates the results of the time spent during the cost computation with GAs and a simple exhaustive search. The GAs outperforms the other search algorithm for every cases with increasing number of nodes with respect to the computation latency.

B. Data Classification Performance

These simulation are designed to show to compare a variety of classifier algorithms such as, ADTree, NBTree, PART, Ridor, J48, IB1 and Bayesian Network for a given sensor network data. These experiments are all performed with a real sensor data, which are also experimented in previously proposed sensor network

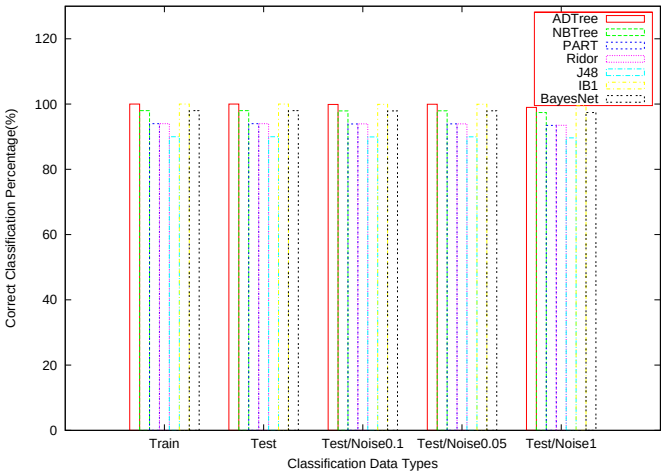

Fig. 6. The correct classification percentage experienced with different classification algorithms for a set of training, correct and noisy testing data.

researches. The data are taken from the environmental monitoring sensors that monitor the temperature of a particular location and determine whether some conditions are satisfied according to the measurements. The training data consists of 50 item set of data. Besides, the test data has number of 8000 different items. During the simulation, test data has been altered with $\%0.01, \%0.05$ and $\%0.1$ probability to demonstrate the noisy data transmitted in the sensor network.

As shown in Fig. 6, the classification percentages are experimented to show the accuracy of the classifier algorithms on sensor data. In this experiments, ADTree and IB1 classifier algorithms outperforms the other algorithms in each type of data with the high correct classification percentages.

The next experiment demonstrates the time efficiency of each algorithm for the test and the training data. As given in Fig. 7, IB1 and Ridor classifier algorithms shows the lowest results during the simulation.

For the overall comparison results, IB1 seems the optimum classifier algorithm that can be applied to sensor network applications for sensor data classification and data aggregation. Since, the experiments are performed for only one sensor application and one type of sensor data, this algorithm may not be the exact answer for every sensor applications. However, this issue is left for future work in this research.

VII. Conclusions

The previously proposed security schemes do not provide a single solution to address the complete security architecture need for Wireless Sensor Networks (WNSs). As current security schemes are inadequate, a new scheme is needed to address this need. To the best of our knowledge, there has been no complete security solution proposed for WSNs.

In this paper, a new security scheme is proposed based on the optimization of a cost function consists of different parameters of the encryption algorithms and sensor networks such as CPU power consumption, transmission overhead and network capacity using GAs. The scheme incorporates a fast, reliable and efficient encryption algorithms to provide high utilization, high reliability and low latency with low overhead. The experimental results showed that the scheme can provide low cost and not very high process time results when compared with other non-secure sensor network algorithms. As a result, the scheme can address the challenges posed by the WSNs and provides reliable, accurate and fast security solution.

Besides providing an efficient solution for security needs of WSNs, this paper also experimented different classifier algorithms in order to address the efficient data aggregation and classification algorithm need. IB1 classifier algorithm gives promising results in terms of time and accuracy.

In the future work, it is planned to improve the overall scheme, to deploy more robust and concrete mathematical models for the security scheme and extend the simulation experiments to obtain better results maybe with real implementations. Besides, it is also expected to use this efficient security scheme and data classifier algorithms to solve other research challenges of WSNs.

References

[1] Y. Sankarasubramaniam, O. Akan, I. F. Akyildiz, “ESRT: Event-to-Sink Reliable Transport in Wireless Sensor,” *ACM MobiHoc 2003*, p. 177-188, June 2003.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, “Wireless Sensor Networks: A Survey,” *Computer Networks Journal (Elsevier)*, vol. 38, p. 393-422, 2002.

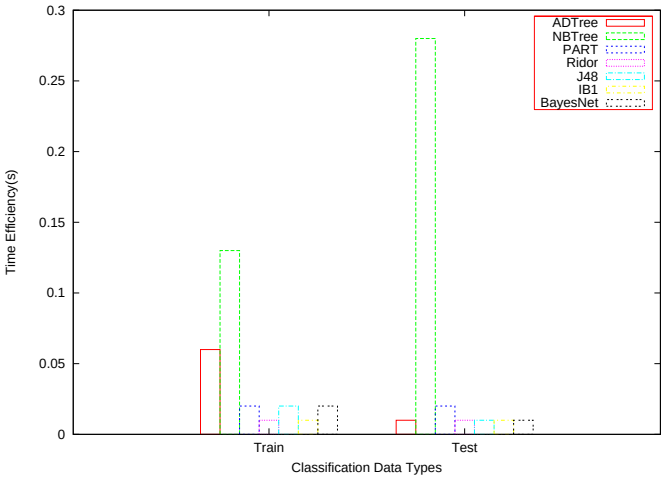

Fig. 7. The time efficiency experienced with different classification algorithms for a set of training and correct testing data.

[3] H. Cam, S. Ozdemir, D. Muthuavinashiappan, P. Nair, "ESPDA: Energy-Efficient and Secure Pattern Based Data Aggregation for Wireless Sensor Networks," *Proc. IEEE*, p.732-736, 2003.

[4] H. Cam, S. Ozdemir, D. Muthuavinashiappan, P. Nair, "Energy Efficient Security Protocol for Wireless Sensor Networks," *Proc. IEEE*, p.2981-2984, 2003.

[5] D. E. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning," *ISBN 0-262-13316-4(HB), 0-262-63185-7(PB)*, Seventh printing, 2001.

[6] C. Kaufman, R. Perlman, M. Spenciner "Network Security: Private Communication in a Public World," *ISBN 0-13-046019-2*, Second edition, 2002.

[7] A. Perrig, R. Szewczyk, V. Wen, D. Culler, J. D. Tygar, "SPINS: Security Protocols for Wireless Sensor Networks," *ACM Mobile Comp. and Net.*, 2001.

[8] K. S. J Pister, J. M. Kahn, B. E. Boser, "Smart Dust: Wireless networks of milimeter-scale sensor nodes," 1999.

[9] S. Rhee, D. Seetharam, S. Liu, "Techniques for Minimizing Power Consumption in Low Data-Rate Wireless Sensor Networks," *Proc. IEEE WCNC 2004*, p. 1727-1731, March 2004.

[10] D. Samfat, R. Molva, "A Method Providing Identity Privacy to Mobile Users during Authentication," *Proc. IEEE*, p.196-199, June 1995.

[11] "TinyOS: Open-Source operating system designed for wireless embedded sensor networks," http://www.tinyos.net/

[12] "WEKA: Machine Learning Software in Java," http://www.cs.waikato.ac.nz/ml/weka/index.html

[13] D. J. Wheeler, R. M. Needham, "TEA, a Tiny Encryption Algorithm," November 1997.

[14] L. Yuan, G. Qu, "Design Space Exploration for Energy-Efficient Secure Sensor Network," *Proc. IEEE Int. Conf. Appl. Sys. Arch. ASAP 2002*, p. 1-10, 2002.

[15] W. Zhang, G. Cao, "An Energy Efficient Framework for Mobile Target Tracking in Sensor Networks," *Proc. IEEE*, p. 597-602, 2000.