International Journal of Engineering & Technology, 7 (2.21) (2018) 443-447

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Smart curiosity sinks node prediction mining algorithm for path optimization in wireless sensor network

A. Kannagi *

Professor, Department of Computer Science and Engineering, Malla Reddy College of Engineering, Tamilnadu, India.
**Corresponding author E-mail: kannagianbu@gmail.com*

Abstract

As of late, data miming and handling for wireless sensor systems (WSNs) have turned into a theme of dynamic research in a few fields of software engineering, for example, the circulated frameworks, the database frameworks, and the data mining. Managing the large amounts of information and efficiently using this information in improved path optimization has become increasingly challenging. In this paper, we will demonstrate how sink node predicted and integrated for path relationships and patterns in linked data, i.e., the interdependencies between data items at the lowest elemental level. The problem of path optimization has been approached using various techniques. The path selection plays the vital role in achieving the quality of service parameter and secure communication. Considering multiple routing, the security can be enforced with various strategies. Using few parameters namely the congestion, delay and hop count would support improve the performance of the network as well as lifetime. With the motivation, an efficient smart curiosity sink node prediction mining algorithm has been presented in this paper. First, both the source and destination nodes maintain information about the routes and network conditions. Based on that a single efficient path has been selected for data transmission. On the other side, the receiver verifies the path being followed, the route available and their conditions. Using this information the delay approximation is performed to decide the legitimate of the path being selected and the traffic incur in the way. The proposed method identifies several network threats and detects the presence of the node in the route. The proposed plan improves the performance of mining data efficiency as well as increases the throughput.

Keywords: *Sink node, path optimization, mining, route estimation, WSN.*

1. Introduction

Machine learning from sensor data has gotten a lot of consideration by the data mining group. Distinctive methodologies concentrating on path optimization, sink node recognizable proof, regular examples, following examples, and characterization have been efficiently utilized on sensor data. Be that as it may, the outline and sending of sensor networks make one of a kind research challenges because of their massive size irregular and unsafe execution, lossy imparting condition, restricted power supply, and high disappointment rate.

These troubles make ordinary mining methodologies inapplicable in light of the fact that usually mining is united and computationally costly. Wireless sensor systems have pulled in critical consideration in numerous applications areas, for example, territory observing, protest following, condition checking, military, catastrophe administration, and additionally shrewd situations. In these applications, constant and solid control is a basic prerequisite. These applications yield the enormous volume of dynamic, geologically circulated and heterogeneous data.

This crude data, if proficiently broke down and changed to usable data through data mining, can encourage computerized or human-prompted strategic/vital choice. Thusly, it is basic to create procedures to dig the sensor data for examples to settle on smart choices speedily.

Subsequently, new calculations have been made, and a portion of the data mining calculations have been altered to deal with the data produced by sensor networks. Plenty of information disclosure procedures. Since data mining is an expansive demonstrate and can be connected to any area data, more broad overviews on data mining systems, where creators analyzed the machine-learning and data mining methods for path optimization amid data move in the network.

Since the order of data mining procedures in this evaluation depends on path design mining, sink node, and it follows, there are adequate of examinations accessible on every one of these methods. The smart curiosity sink node prediction mining procedures focus on design and management of network instead of data disclosure. The sink node expectation mining systems strategies over data stream are given, where the source analyzed conventional characterization methods over data streams.

In any case, inspected data mining methods that importance on path optimization and examination from WSNs data. In correlation with the calculations supposed above and approaches uncommonly intended for WSNs, not just promoting an alternate order, assessment, and exchange on different spaces yet also exhibiting several decisions of a solution. We inspected how data mining calculations will be used to influence the sensor to network applications wise. In this work comprises of sink node forecast of data mining systems for WSNs.

2. Related works

A choice tree is a straightforward and intense type of data investigation which enables us to anticipate, clarify, portray, or order an objective. A choice tree comprises of two kinds of nodes; non-terminal nodes and terminal nodes [1]. A non-terminal node indicates a condition in view of an element. This condition parts data records in light of their qualities. A terminal node,

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

additionally it is known as a leaf node, doesn't have a term since it is utilized to settle on a choice.

Data mining got from the data mining strategies and prepared through scientific models were approved with relating observation data. A basic class of these frameworks in body sensor systems (BSNs) here lightweight installed processors and correspondence frameworks are firmly combined with the human body [2].

Upgraded Mining of High Dimensional Data utilizing Efficient Clustering approach called regulated quality bunching calculation is proposed to enhance the exactness and check the likelihood of the examples. In this technique, speedier recovery of applicable data is made more effective and precise [3]. By utilizing this technique, clients can get exact outcomes and immaterial data misfortune. This strategy shows comes about in light of the high likelihood thickness in this manner giving security to data and decreasing the dimensionality of the data.

BSNs can give specialists, mind suppliers and clinicians access to hugely significant data extricated from data that are gathered. To address this issue, a data mining approach enlivened by the involvement in the regions of content and regular dialect handling [4]. Speak to sensor readings with a succession of characters, called movement transcripts. Transcripts diminish the multifaceted nature of the data altogether while keeping up morphological and auxiliary properties of the physiological signs.

Empowering area mindful quality-controlled access in remote systems structure of provisioning administration separation and decency as indicated by the per-area target stack. In the proposed structure, the heap data is criticism to movement senders and used to modify their sending rate [5], so per-area weight does not digress from a given for every area offer of remote connection limit and lays just average activity on the system in collaboration with different areas.

The highlights of the preparation tests were separated to develop the choice table; the harsh set hypothesis was connected to lessen the choice table. At last, the decreased choice ascribes were utilized to build neural system arrangement data [6]. Fabricated data mining calculation can be incorporated into every sensor organize node. The issues need to understand are changing as the time past, to outline a framework which can be working longer, it must be versatile.

The course of data mining research, at that point portrays the examination on remote sensor organize data mining, outlines the Service Oriented Sensor Network Data Mining System, and incorporates three calculations: dispersed k-implies, multi-target following. In view of the fast data arriving speed and the monstrous size of data set in stream demonstrate [7], novel one-pass calculations are concocted to help data collection on request.

Finish action checking and programmed fall location for indoor and open air utilize. The framework is made out of a portable module worn by the client and call focus spare and examinations the data [8]. Action observing is performed utilizing neural systems. Remote accelerometer sensor module and calculation for determining fall identification. The fall discovery calculation identifies wearer's stance, movement, and fall.

Data administration is the assignment of gathering data from sensors, putting away data in the system, and effectively convey data to the clients (the sinks) [9]. The comprehension of data administration methods executed in today sensor systems is required before tending to abnormal state data mining, design acknowledgment, and circulated theory development. Continuous worldwide situating has been supporting limitation and following in mining for a considerable length of time [10]. An earthbound recipient can decide its situation from a known code grouping originating from at least four satellites of known area, utilizing the synchronization-decided time of flight and the season of transmission and satellite area, which are additionally incorporated into each flag.

3. Materials and methods

The efficient smart curiosity sink node prediction mining algorithm maintains traces about sink node present in each transfer; the process has the number of stages Source Discovery based on Delay Approximation, ID-based Sink Detection, Path Setup for Secure Communication and Distributed Authentication.

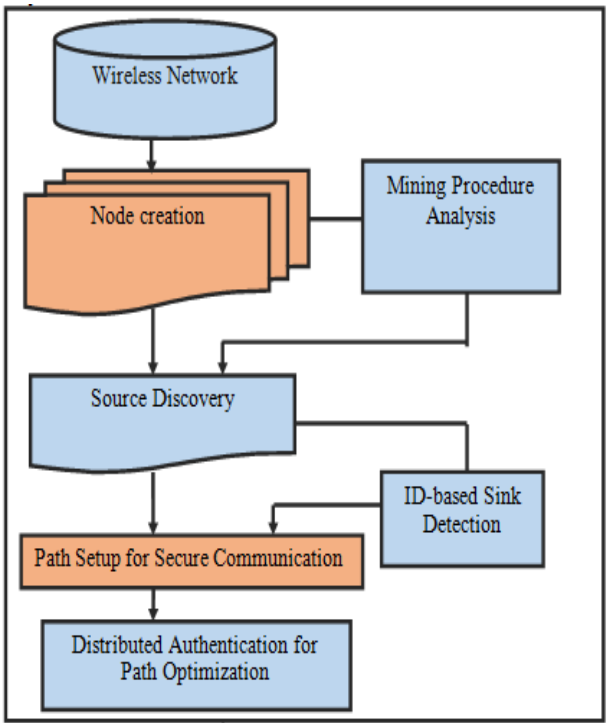


Figure 3.1: Work flow of proposed method

The Figure 3.1, shows the proposed work efficient smart curiosity sink node prediction mining algorithm for path prediction in wireless sensor network and its functional components. We discuss each of the functional components in detail in this given section.

3.1. Source discovery based on delay approximation

The source discovery is the initial process of path optimization, and the method identifies the list of sources which possess the data. The nodes have limited delay, and the way determines the list of nodes which has data to be collected using the neighbor nodes. Initially, the node floods the message and receives the replies from all the nodes of the network. From the response being received the node extracts the list of nodes has information to be collected. By accepting all replies with the message about the data content or to identify the correct sources in the network. Once all the response is received, then the method extracts the list of source nodes, and to collects the delay information from all sources in the network.

Algorithm:
Input: Number of Nodes NN
Output: Source Discovery SD
Start:
Step 1: Read node n
Step 2: Extract Node history Nh
Step 3: Send message to all neighbors nodes
Step 4: If Nh> Payload-Threshold then

Neighbor's nodes receive the floods
Step 5: Else
Extracts the list of source nodes
Step 6: End
Stop:

3.2. ID based sink detection

ID based sink detection uses the node id with mining management to give the end to end authentication. In the network, every node keeps id and its corresponding index value, based on ID and its index value the sink will be chosen quickly. Path optimization scheme is based on route invention packet from the source node and a route reply packet from the destination node. Whenever a source node desires to undertake secure communication with the destination node, the source node forms its sink node and sends to the data in the network and waits for a reply message from sink node. Based on sink node reply message the next step carry on. The list of Sources belongs to the node has been identified as follows:

Node Id set NIS = $\int_{i=1}^{Size(Atr)} \sum Atr(i).Source == Source$

The list of node id belongs to the transmission route has been identified as follows:

Route set Rs = $\int_{i=1}^{size(Atr)} \sum Atr(i).Route \nexists Rs$

The common hostnames can be computed as follows:

Common hosts= $\int \sum Host \in \forall (Routes\ of\ rs)$

According to this, the routing factor can be estimated as follows:

Route factor Rf= $\int_{i=1}^{size(Mt)} \frac{\sum Mt(i) \in \forall CHN}{size(Mtr)}$

Similar to that the number of the malicious packet in the route can be computed as follows:

Nr = = $\int_{i=1}^{size(Mt)} \sum Mt(i).Route == Route$

Using all these measures, the Sink can be computed as follows:

Sink = Rf× Nr

The computed sink node has been used to perform path setup. The algorithm computes the sink detection factor for each route and identifies the moderate hops.

Algorithm

Input: Source Discovery Sd
Output: Sink
Start:
Step 1: Read Source node Id
Step 2: Compute Delay
Step 3: Compute Transmission Weight
Step 4: Compute Source Weight
Step 5: If Transmission and Stream Weights > Threshold
Step 6: If Route Weight> Threshold
Sink node is detected
Step 7: Else
Node id will be checked
Step 8: End
Stop:

3.3. Path setup for secure communication

The path setup is essential one in the network after the sink node reply the source node require not have gone along the way with minimal number of bounces; the slightest jump way may have a higher postponement, either authentically or malevolently showed. For this situation, be that as it may, a no congested, non-slightest jump way is probably going to be liked to a protected correspondence as a result of the lessening in delay. Since sink don't contain a bounce check or particular recorded source course, and in light of the fact that messages are scrambled at each jump, vindictive nodes have no chance to divert activity with the endeavors. After way setup, the goal unicasts an answer parcel back along the turn around way to the source. In the middle of the node, check its scrambled data and pass the reaction message to its next node the other way.

Algorithm:
Input: Sink node id SId,
Output: Correct path
Start:
Step 1: Read Sink node
Step 2: Read Node ID
Step 3: Identify the ID of the source Sid
Sink node Sn= $\int_{i=1}^{size(Sn)} \sum Sn(i).SId = Sources.SId$
Split the sink node into different time window
Tw = $\int_{i=1}^{size(Tw)} \sum Sn(i).Time$
Step 4: Now for each time window
Compute path Cp = $\frac{\sum Sn \in Tw(i)}{Size(ss)}$
Stop:

3.4. Distributed authentication

It's the last stage for path optimization after all the path setup source data will be transferred into a destination using sink node. That time used multiple intermediate nodes in the network, the source node distributed in all intermediate nodes during communication. In the distributed authenticated secure communication scheme, individual nodes of relatively high safety among the sink nodes consisting of the network were set as source Nodes, from which the process of issuing a node ID was started and distributed.

4. Result and discussion

Efficient smart curiosity sink node prediction mining algorithm has been implemented in network simulator NS 2.28 version under various scenarios. The behavior of the suggested approaches is analyzed with a variety of simulation parameters. A node can straightforwardly impart to the nodes that exist in its correspondence extend. In the event that a node needs to speak with a node that isn't straightforwardly inside its correspondence extend, it utilizes middle of the road nodes as switches.

In a node from an area to another area can be empowered utilizing the catchphrase "setdest" in Tool Command Language (TCL) content. Nodes are arranged with the components of the channel, organizing interface, radio spread model, Medium Access Control (MAC) convention, steering convention, interface line, connect layer, geology protest, and reception apparatus compose. Nodes in the system have been exchanged utilizing correspondence display. The developments of portable nodes are kept to a zone of 500m□500m with the respite time of3s. Data transmission is set up between nodes utilizing UDP specialist and CBR activity. These halfway switches forward the bundles created by different nodes to their goal. Our proposed system (Smart Curiosity Sink Node Prediction (SCSNP)) is compared with three existing systems; they are Model-based clustering Algorithm (MBCA) and K-Partitioned Smallest Distance Tree (k-PSDT)

4.1 Throughput ratio

The throughput ratio characterizes the rate of data packets got at a goal as per the quantity of packets created by the source node for a specified period.

Throughput = Received data * 8 / Data transmission period

Table 4.1 demonstrates the throughput proportion examination of the proposed framework with the current frameworks.

Table 4.1: Throughput Ratio

Number of nodes	MBCA in %	k-PSDT in %	SCSNP in %
20	3.6	5.9	20.2
40	18.6	19.2	23.5
60	39.2	42.7	49.2
80	59.3	65.1	75.2
100	65.1	71.6	91.6

Figure 4.1 shows the comparative analysis of throughput ratio of the proposed scheme with the existing, based on the values is given below

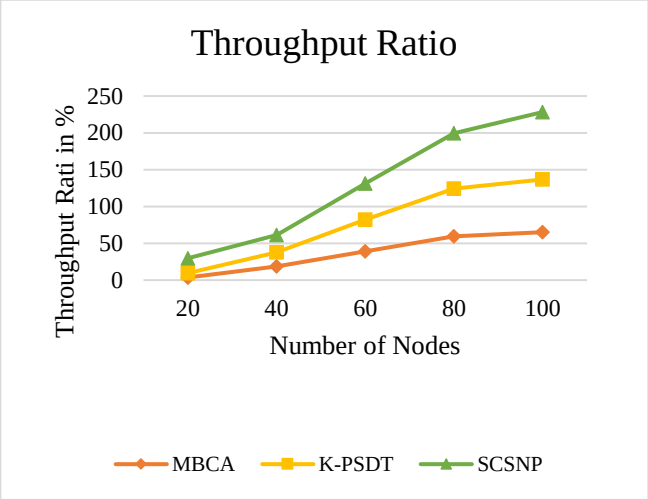


Figure 4.1: Throughput ratio analysis

4.2. Analysis of packet delivery ratio

Packet Delivery Ratio (PDR) is utilized to assess the nature of the system. It defines the ratio between the packets received by the destination and the packets generated by the source. It can be obtained by using awk script, which produces the trace file and the result.

PDR = Received packets/ Generated packets * 100

Table 4.2: Packet Delivery Ratio Comparative Analysis

Number of nodes	MBCA in %	k-PSDT in %	SCSNP in %
20	0	0	0
40	29.75	39.58	47.65
60	33.65	43.65	49.79
80	49.28	61.98	72.89
100	69.78	72.45	89.65

Figure 4.2 gives the relative investigation yield of the proposed plot with the current methodologies. The investigation delineates the accompanying:

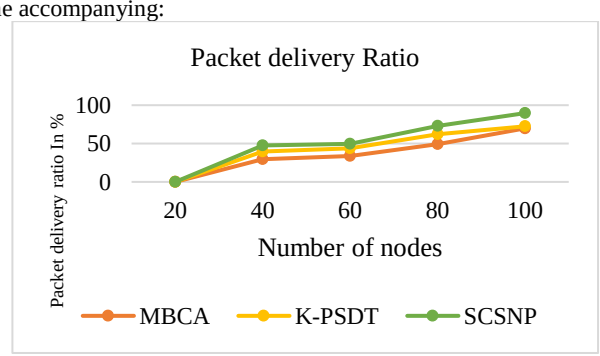


Figure 4.2: Packet delivery ratio comparative analysis

4.3. Mining efficiency ratio

The mining efficiency proportion is utilized to assess the nature of the system gave by the steering plan. The simulation results of efficiency ratio are given in Table 4.3

Table 4.3: Mining Efficiency Performance Analysis with Existing and Proposed Systems

Number of Nodes	MBCA in %	k-PSDT in %	SCSNP in %
20	19.72	21.95	23.65
40	26.98	32.58	39.85
60	39.68	49.85	55.68
80	51.85	59.68	62.98
100	73.65	82.15	91.58

Figure 4.3 depicts the performance of the proposed system SCSNP with other systems. The proposed scheme outperforms the existing one regarding ratio.

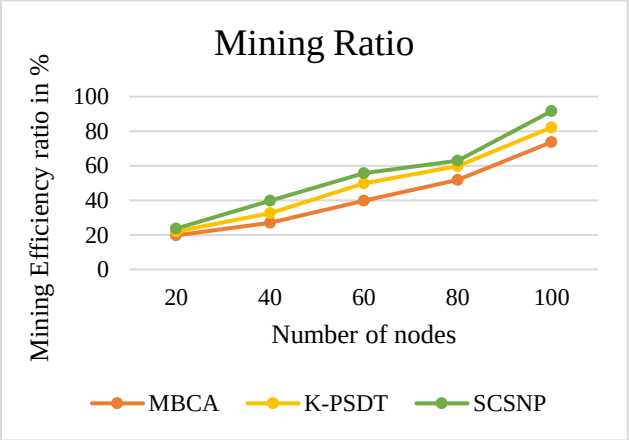


Figure 4.3: Mining efficiency ratio

4.4. Analysis of end to end delay

The time which is taken by the source node to convey the data effectively to the goal is called as a conclusion to end defer. It is the difference between the time at which packets generated by the sender and the time the packets received by the receiver and it can be obtained from the trace file.

End to end delay = Received time – Send time

Table 4.4 shows the end to end delay analysis of the proposed system with the existing systems.

Table 4.4: End to End Delay

Rate of Sending Packets/sec	MBCA	k-PSDT	SCSNP
10	1.23	0.86	0.59
20	5.8689	4.5689	4.3658
30	4.9658	3.8698	2.9875
40	7.5854	7.1025	6.9358
50	7.6549	6.5698	5.9868

Figure 4.4 demonstrates the conclusion to end investigation of the proposed plot with the current frameworks. The postpone computed in light of the rate of sending bundles specifically seconds. Be that as it may, the slight variety is there. The proposed design prompts just endured delay in the system, despite the fact that the parcels/sec. If the destination moves closer to the source node, the suspension will be decreased. Otherwise, the delay will be increased.

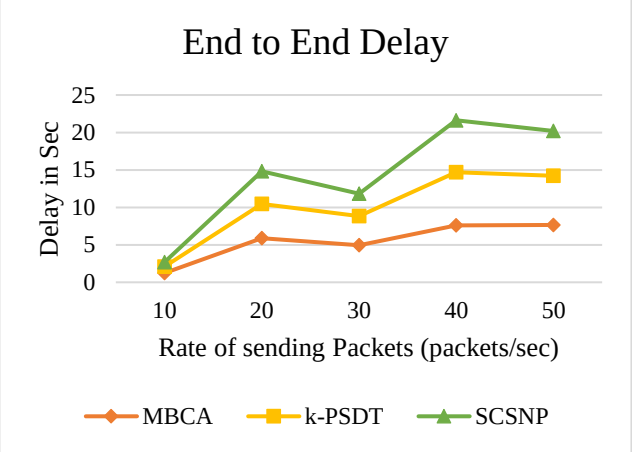


Figure 4.4: End to End Delay Ratio

From Figure 4.4, it is understood that the proposed SCSNP has reduced its end to end delay with the rate of sending packets.

5. Conclusion

The smart curiosity sink node prediction mining algorithm is performed based on the computed value of the sink detection. Here the path optimization of every node is evaluated and calculates the sink of the node. Using this sink node, the path section performed by analyzing, whether, the sink node is available or not. The proposed method after evaluation increase the throughput ratio up to 91.6%, and produce the mining efficiency up to 91.58% in network. Also, our proposed model gives higher delivery ratio compared to all other existing methods in the network.

References

[1] Li L, Li X, Lu Z, Lloret J & Song H, "Sequential Behavior Pattern Discovery with Frequent Episode Mining and Wireless Sensor Network", *IEEE Communications Magazine*, Vol.55, No.6,(2017), pp.205-211.

[2] Kim H, Kim HS, Lee SK, Park EC & Park KJ, "Enabling location-aware quality-controlled access in wireless networks", *EURASIP Journal on Wireless Communications and Networking*, (2011).

[3] Röhm U, "Enhanced Mining of High Dimensional Data using Efficient Clustering Algorithm", I*nternational Journal of Advanced Research in Computer Science and Software Engineering*, Vol.3, No.12, (2013).

[4] Loseu V, Ghasemzadeh H & Jafari R, "A Mining Technique Using $ N $ n-Grams and Motion Transcripts for Body Sensor Network Data Repository", *Proceedings of the IEEE*, Vol.100, No.1,(2012), pp.107-121.

[5] Phung ND, Gaber MM & Rohm U, "Resource-aware online data mining in wireless sensor networks", *IEEE Symposium on Computational Intelligence and Data Mining*, (2007), pp.139-146.

[6] Yuehua H, Shuang X & Huajian W, "Study on distributed data mining model in Wireless Sensor Network", *International Conference on Intelligent Computing and Integrated Systems (ICISS)*, (2010), pp.866-869.

[7] Ye XG, He YH & Wu L, "Service Oriented Distributed Data Mining System in Sensor Network", *Third International Conference on Information and Computing*, (2010), pp.225-228.

[8] Bensheng Wang, Tao Wang, Noufissa.Mikou,"An Efficient Data Streams Mining Method for Wireless Sensor Network's Data Aggregation," IEEE, vol-5, issue-2, 2009.

[9] Shuqin R & Park JS, "Density mining based resilient data aggregation for wireless sensor networks", *Fourth International Conference*, (2008), pp.261-266.

[10] Perolle G, Fraisse P, Mavros M Etxeberria I, "Automatic fall detection and activity monitoring for elderly", *Proc. MEDETEL*, (2006), pp. 65–70

[11] Lee Y, Kim J, Son M & Lee M, "Implementation of accelerometer sensor module and fall detection monitoring system based on wireless sensor network", *Twenty-ninth Annual Int. Conf. IEEE EMBS Cite Internationale*, (2007).

[12] Cantoni V, Lombardi L & Lombardi P, "Challenges for data mining in distributed sensor networks", 1*8th International Conference on Pattern Recognition*, (2006), pp.1000-1007.

[13] Olfati-Saber R & Murray RM, "Consensus problems in networks of agents with switching topology and time-delays", *The IEEE Transactions on Automatic Control*, Vol.49, No.9,(2004).

[14] Kiziroglou ME, Boyle DE, Yeatman EM & Cilliers JJ, "Opportunities for sensing systems in mining", *IEEE Transactions on Industrial Informatics*, Vol.13, No.1,(2017), pp.278-286.

[15] Lian J, Naik K & Agnew G, "Data Capacity Improvement of WSNs Using Non-Uniform Sensor Distribution", *International Journal of Distributed Sensor Networks*, Vol.2, No.2, (2006), pp.121-145.

[16] Huang CJ, Chuang YT & Hu KW, "Using particle swarm optimization for QoS in ad-hoc multicast" *Engineering Applications of Artificial Intelligence, Elsevier Ltd*, Vol.22, (2009), pp.1188-1193.

[17] Bari A, Wazed S, Jacked A & Bandyopadhyay S, "A Genetic Algorithm based Approach for Energy Efficient Routing in Two-Tired Sensor Networks", *Elsevier, Ad Hoc Networks Journal*, Vol.7, No.4, (2009), pp.665-676.

[18] Araújo AFR & Garrozi C, "MulRoGA: A multicast routing genetic algorithm approach considering multiple objectives", *Applied Intelligence*, Vol.32, (2010), pp.330-345.

[19] Lindsey S & Raghavendra CS, "PEGASIS: Power-Efficient Gathering in Sensor Information System", *SIGMOD Record*, Vol.32, No.4, (2003), pp.66-71.

[20] Jeong JK, Kim SO & Song CH, "A solution of search multicast routing tree based on extended simulated annealing algorithm", *WSEAS Transactions on Mathematics*, Vol.4, (2005), pp.218-223.