Multivariate Analysis for Fuzzy Correlated Node Behavior Detection in Wireless Sensor Network

Noor Shahidah[1] and A. H. Azni[2]
[1, 2] Faculty Science and Technology, Universiti Sains Islam Malaysia (USIM), Nilai 71800, Malaysia
[2] Islamic Science Institute, Universiti Sains Islam Malaysia (USIM), Nilai 71800, Malaysia
Email: shahidah_ishak93@yahoo.com; ahazni@usim.edu.my

Abstract **—**Wireless Sensor Network depends highly upon the cooperation among the nodes behavior in transmission of packet data, messages and route discovery. Over open medium environment, nodes are free to move and may change their behavior arbitrarily. In the presence of misbehavior node in some cases, it may instigate its neighboring nodes to compromise with the misbehaved node. Thus, this has resulted to a spreading of correlated node behavior and the impact of this event may result in high severity in network performance. Therefore, fuzzy logic model is proposed to formulate the correlated node behavior in WSN. The formulation of correlated node behavior based on fuzzy logic function of peer nodes real parameter measurement is investigated to determine the status of the node and then the fuzzy neural network will model the correlated node behavior occurrence. The accuracy of the results is established via sensor network simulation. The result of this study is providing a fundamental guideline for network designer to understand the fault-tolerance in network topology.

Index Terms—Fuzzy logic system, correlated node behavior, wireless sensor network, fuzzy correlated function

I. Introduction

Wireless sensor network (WSN) with several detection stations commonly known as sensor nodes connect wirelessly with low deployment, low data usage cost and self-organized system [1]. The availability of WSN is important in many advance monitoring and control applications including telecommunication, biomedical and geographical controller. The operation of the WSN depends highly on the cooperation of nodes during routing, network monitoring and packet forwarding. In harsh environment however, the complex structure of the network and with no central monitoring scheme of the network yield to high vulnerability of security attack in this network. Thus, it is difficult to guarantee for nodes to cooperate consistently and maintain network operation. Early researches found nodes exhibit independent failure, that against the recent studies. For example, Thanakornworakij in [2] found the interruption in network application may cause by only a single node failure. Also, study in [3] found a contagious failures are initiated by the earlier nearby failure. This scenario is called correlated event or correlated node failure. The correlated failure, on the other hand, imposed high severity upon network availability and connectivity. A number of method to detect the correlated node behavior presented previously, meanwhile they did not consider the uncertainties inherited in the network environment and lack in detection accuracy.

Therefore, proposed fuzzy system for correlated node behavior detection model addressed the limitations. The correlated event in this model is determined by the status of nodes behavior and several correlation acts. As the WSN system consist of multiple parameter measurement, it can be used to identify the nodes behavior. Thus, this study aims to analyze the multivariate analysis for both phases; the characterization of nodes behavior and the determination of correlated degree of the event.

The rest of this paper is organized as follows. Section II discussed proposed formulation of fuzzy correlated function. Evaluation metrics presented in Section III. Then, Section IV is presentation of preliminary results. Finally, Section V offers conclusion of the works.

II. Proposed Formulation of Fuzzy Correlated System

This section presents the overview of fuzzy logic system (FLS) and the formulation process of correlated node behavior modeling. Section A explains the process involved in FLS in general while section B presents the formulation process in detail.

A. Fuzzy Logic

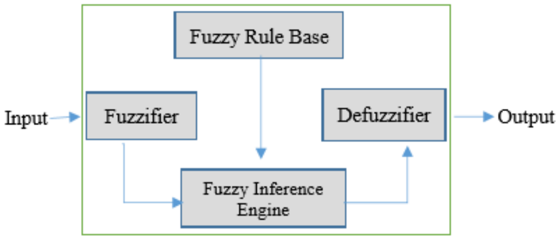


Fig. 1. The structure of fuzzy logic system

Fuzzy logic is first introduced by L.Zadeh in 1965. The implementation of fuzzy logic is varied in numerous fields to improve decision-making, reduce resource consumption and performance enhancement. The system

Manuscript received February 20, 2018; revised July 15, 2018.
This work was supported by the Fundamental Research Grants No: USIM/FRGS/FST/32/51414.
Corresponding author email: ahazni@usim.edu.my.
doi:10.12720/jcm.13.8.430-435

of fuzzy logic consists of fuzzification, inference engine and defuzzification stages. The fundamental structure of fuzzy system is presented in Fig. 1.

The fuzzification is a process to convert the crisp input values into fuzzy linguistic variables according the corresponding membership functions. After the fuzzification process, the inference engine generates if-then statements according to a set of predefined rules derived from knowledge provided by experts. In this stage the inference scheme maps input fuzzy sets to output fuzzy sets. The rules are generated according Mamdani method where the Min (∧) operator is used as AND connective between the antecedents of the rules as follows:

$$\tau_i = A_{i1}(x_1) \wedge A_{i2}(x_2) \wedge A_{i3}(x_3) \qquad (1)$$

where τ_i is denoted as the *degree of firing* of the i^{th} rule for the input values. Next is to determine the individual rule output F_i which is obtained by

$$F_i(y) = \tau_i \wedge B_i(y) \qquad (2)$$

The rules output is aggregated to obtain the overall output system F where the Max (∨) operator is used as OR connector between the individual rules:

$$F(y) = \vee_i F_i(y) = \vee_i(\tau_i \wedge B_i(y)) \qquad (3)$$

Then, the crisp result is computed in the defuzzification stage using Center of area (COA) method:

$$COA = \frac{\sum_{j=1}^{m} F(y_j) \times y_j}{\sum_{j=1}^{m} F(y_j)} \qquad (4)$$

where y_j is a sampling point in the discrete universe of output F and $F(y_j)$ is its membership degree in the membership function.

B. Formulation of Fuzzy Correlated System

In this section, correlated node behavior detection in wireless sensor network is presented according fuzzy association rules. The model consists three steps involved first, node behavior detection followed by attribute correlation and the last is spatial correlation. These three steps which shown in Fig. 2 are then become the input parameters of FLS to determine the degree of correlated events detection in Section C

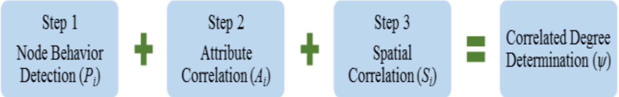

Step 1
Node Behavior Detection (P_i)
Step 2
Attribute Correlation (A_i)
Step 3
Spatial Correlation (S_i)
Correlated Degree Determination (ψ)

Fig. 2. Steps involved in fuzzy correlated behavior model

- *Step 1: Node Behavior Detection (P_i)*

Node behaviors can be classified into cooperative and misbehave node. The misbehave nodes including the malicious, selfish and fail node [9]. Meanwhile in this study, the interest is on the cooperative, malicious and the selfish state of node behavior only. The main parameters considered for the node behavior detection is summarized as bellow:

i. *Packet Loss (PL)*: Packet loss indicates the ratio between the packet sent and the packets forwarded. In WSN, attack against network, dominantly performed by malicious nodes. This misbehaving nodes may attract all the packets using forged Route Reply (RREP) packet to wrongly choose “fake” shortest path towards the destination meanwhile, discard these packets without forwarding them to the destination [10]. Thus, by analyzing the packet loss will assist in malicious node behavior determination in this modeling process.

ii. *Energy Level (EL)*: Energy level for each node defined as the proportion of remainder energy towards its initial energy. As packet forwarding between distant nodes are expected to be relayed by intermediate node which act as routers. However due to energy constraint, the nodes may try to detach from network in order to refrain from cooperating and act selfishly. Thus, *EL* is considered to determine the selfish behavior of the nodes in network operation. According to Kothari et al. [11], node starts to behave selfishly when the energy level is below 25% of its initial energy level.

iii. *Throughput (TP)*: Throughput defined as the number of packets successful transmitted to destination over time. In WSN, throughput is one of variables which can measure the quality of service (QoS) instead of lifetime and delay. This variable is inversely proportional to the packet loss probability such that if the probability of the packet loss is 3%, then the throughput is 97% provided the packets do not collide [12]. Then in the presence of selfish nodes, throughput will decrease as the number of selfish nodes increases and the same cases with the malicious nodes [13].

The protocol converts the numerical values of packet loss, energy level and throughput into fuzzy linguistic variables by using the corresponding predefined variables and membership functions as shown in Table I.

TABLE I: SUMMARY OF MEMBERSHIP FUNCTION

Parameters		Memberships degree		
Inputs	Packet Loss (PL)	Low	Medium	High
	Energy Level (EL)	Low	Medium	High
	Throughput (TP)	Low	Medium	High
Output	Behavior (P_i)	Cooperative	Selfish	Malicious

The membership function for all parameters has triangular shape. For each input and output parameters, it has three membership functions that shows the different degrees of the functions. The degree is determined as low, medium and high for input parameters and the assurance level of the node behavior classified as cooperative, selfish and malicious. The format of rule, R^l can be written as:

R^l : IF PL is low, EL is low, and TP is low;
THEN the node is in Selfish behavior

- *Step 2: Attribute Correlation A_i*

The next step is to determine the attribute correlation between nodes N_i. Attribute correlation A_i can be expressed as the relationship among the parameter measurements such as packet transmission and energy level of multi nodes in the network due to high density of the network topology. Therefore, common relationship should exist among the behavior of neighboring nodes. Fully dependent attribute values should be coherent with each other. For this reason, when dealing with correlated node behavior detection logic, attribute cohesive concepts are added [7]. In this section, each sensor is undergone the behavior detection process according to Step 1. The objective in this Step 1 is that, is to determine the current status of each nodes within the same transmission radius. Then the attribute correlation rule is applied in each of the network cluster. This step is applied to analyze the dependency of attributes measurements. Three variables are declared as low relationship, medium relationship and high relationship. The format of the rules and the membership function described in attribute correlation is as follows:

IF N_1 is cooperative and N_2 is cooperative ….and N_i is cooperative;
THEN Correlation assurance level is high attribute correlation.

The assurance levels of attribute correlation are classified as low attribute correlation (LAC), medium attribute correlation (MAC) and high attribute correlation (HAC) where, LAC denotes less dependency, MAC comprises of semi dependency among attributes and HAC expresses the full dependency relationship among the attribute values of sensor node.

- *Step 3: Spatial Correlation S_i*

In this section, FLS models the spatial correlation between the distances, *d* of nodes N_i within range, r as shown in Fig. 3. To evaluate the spatial relationship between all the correlated events, the dependency any two events increase when the events occur closer in space. Thus, there show a significant relationship between the distance among the nodes and the spreading of node behavior [14]. For instance, a selfish node N_1 tends to drop the packet delivery in the case of energy deficiency and this leads to rerouting path of the other nodes such as N_2, N_3, N_4 and N_5 for the transmission of the data to the sink node. Meanwhile the frequent of rerouting process will lead to extra energy consumption of the nearby nodes and resulting further energy deficiency of the networks players in the specific transmission range [15].

Next, to measure the spatial correlation between the nodes behavior, three fuzzy linguistic variables are declared as low, medium and high. These are to analyze the sensor's farthest distance. The format of the rules can be described as follows:

IF N_1 is H and N_2 is H and N_3 is H…and N_i is H;
THEN Correlation assurance is HSC

The assurance level of the spatial correlation is classified as Low Spatial Correlation (LSC), Medium Spatial Correlation (MSC) and High Spatial Correlation (HSC) where LSC denotes too farthest nodes, MSC comprises of nodes which in between nearer and farthest nodes and HSC comprises of nodes which are too nearer.

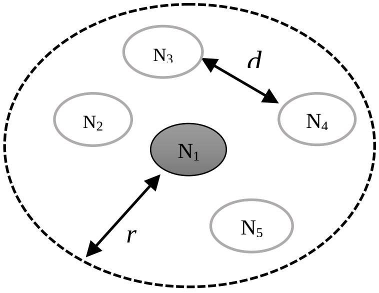
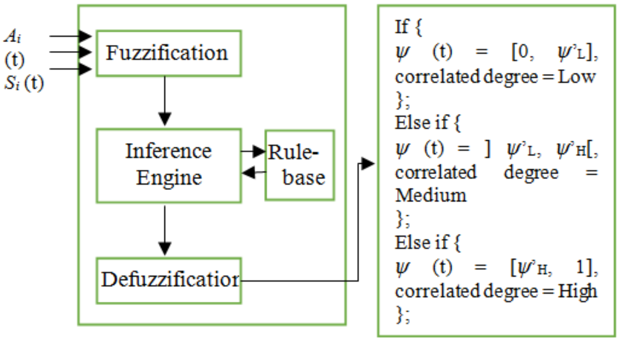


Fig. 3. A cluster of node N_i

C. Correlated Behavior Formulation (ψ)

Network operation such as route discovery and packets forwarding highly influenced by the nodes cooperation. Having misbehave node in the route path, may trigger the multiple failure and thus degrade the quality of service (QoS) of the network. Hence, this section aims to formulate the correlated degree of network behavior by considering all the three steps mentioned in Section B.

Fuzzy logic system process in this section involving three steps; 1) Input, 2) Process and 3) Output. The fundamental of FLS for correlated degree model is illustrated in Fig. 4.



```
If {
ψ (t) = [0, ψ'L],
correlated degree = Low
};
Else if {
ψ (t) = ] ψ'L, ψ'H[,
correlated degree =
Medium
};
Else if {
ψ (t) = [ψ'H, 1],
correlated degree = High
};
```

Fig. 4. Block diagram of fuzzy system for correlated node behavior detection

i. Input: This is the fuzzification stage, where the numerical value of assurance level of each parameters A_i and S_i are converted into fuzzy linguistic variables and the membership functions have triangular shape. Table II summarized the degree of each inputs and output membership function.

ii. Process: The process of fuzzy system or commonly known fuzzy inference system is where the rules are designed based on the expert information. In order to achieve the linguistic results of A_i and S_i, Mamdani-type fuzzy If-Then rules are applied. The rule can be described as follows:

IF A_i between nodes is low and the S_i is low;
THEN the degree of correlated event is low

The IF part is called as antecedent and the THEN part is the consequent. In order to determine the degree of consequent of each rule, the min-max method based on Equation (3) is applied by taking the minimal degree of antecedents of each rule as in Equation (2).

iii. Output: In fuzzy system, the output formation is called as defuzzification stage where the final numeric value of degree of correlated event is determined. The assurance level of the correlated degree event is measured according to Equation (4) and are classified as low, medium and high as shown in Table II.

TABLE II: INPUT MEMBERSHIP FUNCTION OF CORRELATED EVENT

Parameter		Memberships Degree		
Input	Attribute Correlation (A_i)	Low	Medium	High
	Spatial Correlation (S_i)	Low	Medium	High
Output	Correlated Degree (ψ)	Low	Medium	High

III. EVALUATIONS

For analysis purposes, a simulation is conducted in MATLAB version 2015 environment. The experiment considers a network with 10 nodes randomly distributed in a 400 meter x 400 meter area. Each node is free to move following random waypoint mobility model with an average speed 2 meter/second and has 200 meters transmission range. The time step to simulate the scenario is 2 hours. During the simulation, the behavior of the node changes according to the energy resources available for its routing and forwarding packet ratio. In order to calculate the correlated event occurrence, a collection of neighborhood statistics of each node per 5 minutes is needed, together with the number of neighbors and behavior of each neighbors.

A. *Metric for Performance Evaluation with Respect to Correlated Degree*

To analyze the performance of model, two metrics are considered. First the node degree and the second is the mobility. From the simulation, this study aims to identify the impact of both factors towards the performance of correlated degree event.

a) Node Degree, d

The simulation is setup with two cases of node degree, *d=3* and *d=4*. The node degree depicts the minimum and maximum number of deployed nodes for wireless network topology to achieve connectivity. The simulation result represented in Fig. 5 for the malicious node and Fig. 6 for the selfish node. The figures explained that the results of correlated degree increase as the number of neighbor node increases. Both scenario share the equal pattern that is the higher the *d*, the higher the correlated degree. In addition, the correlated degree also accelerated due to the spreading event in the presence of malicious nodes compared the selfish nodes. As the number of malicious nodes in the network increases, the probability of these nodes drops the routing data message increases, leading to a higher packet loss rate [16]. Thus, from the graph, the analyses confirmed that the node degree *d* is significant to determine the speeding of correlated node behavior which is also supports the result found in [9].

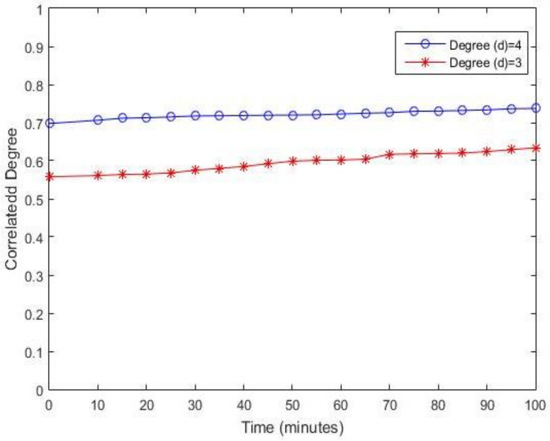


Fig. 5. Malicious node against correlated degree

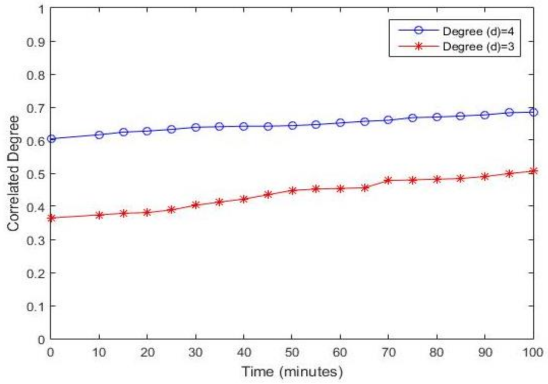


Fig. 6. Selfish node against correlated degree

b) Mobility

Mobility factor indicates the impact of mobility level of one-hop neighbors. In order to evaluate the mobility factor towards correlated event, two difference scenarios with node speeds 2m/s and 20m/s is simulated. According to Azni et al. in [9], the mobility of nodes give significant impact towards the correlated degree as it directly increases the energy consumption as well as the drop ping ratio which affect the infection rate. In the higher node mobility, the impact of correlated degree of node behavior is getting lower thus yields to higher infection rate. Fig. 7 explained the scenario that is the faster the node moves the higher the chances of neighboring nodes getting affected by the misbehavior node.

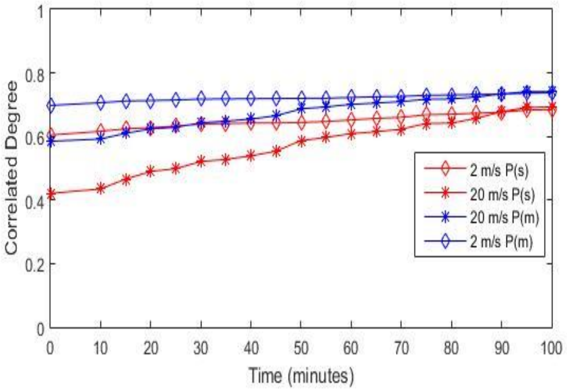


Fig. 7. Selfish and malicious node against correlated degree

B. *Multiple Regression Analysis*

Next, the study evaluates the performance of the model using linear multiple regression analysis. In standard multiple regression, all the variables are included in the model in order to see overall impact of variables in predicting the dependent variable. The analysis of multiple regression in this study conducted in two different phases. First, multivariable behavior detection and second, multivariable correlated degree estimation. For the standard multiple regression, three items of SPSS output are considered for the analysis purposes that are, correlation matrix between independents variables and dependent variable, model summary and covariate estimation.

TABLE III: CORRELATED MATRIX

No	Variables	Behavior	Packet Loss	Energy Level
1	Behavior			
2	Packet Loss	-.550		
3	Energy Level	-.674	.243	
4	Throughput	-.979	.618	.557

The results in Table III indicate a high negative correlation between packet loss, energy level and throughput against nodes behavior detection, respectively. Also, there is high correlation between the throughput against both packet loss and energy level. Meanwhile for the energy level against packet loss, they depict low correlation between them. Next, the model is analyzed in terms of F-statistic. From the Table IV, F-statistic claimed that this model is significant with at least one parameter are significant to determine the behavior of the wireless nodes. It is also supported by the score of R-square, .985 which means the all the parameters packet loss, energy level and throughput are capable to predict the variability of nodes behavior by 98.5%.

The analysis also has estimated the covariate of parameters packet loss, energy level and throughput. From the Table V, it confident that all the parameters are significant to determine the nodes behavior with 95%. Therefore, this study highly recommends a model to determine type of nodes behavior using parameters packet loss, energy level and throughput.

TABLE IV: MODEL SUMMARY OF STANDARD REGRESSION

No	Items	Score
1	F-stat	646.69(p<.001)
2	R^2	.985

TABLE V: COVARIATE ESTIMATION FOR FUZZY SYSTEM OVER NODES BEHAVIOR DETECTION

Model		$\hat{\beta}$	Standard Error (SE)	p-value
1	Packet Loss (β_1)	.061	.001	.000
	Energy Level	-.176	.042	.000
	Throughput	-.919	.000	.000

In addition, covariate analysis toward the correlated node behavior model also conducted in this study. To gain a better understanding to which covariates perform significant impact in predicting the correlated degree of node behavior, the covariates value corresponds to 95% of confidence interval was calculated in SPSS and listed in Table VI. This analysis is to support the objective of this study which is to identify the possible covariates that can significantly affecting the degree of correlated node behavior. From the result in the Table 7, it is found that the attribute correlation and spatial correlation are significant (p<0.001) parameters to address the correlated event detection. As discussed in section 3, the attribute between the neighboring nodes should have significant relationship and thus the correlated event can be identified based on the status of behavior of neighboring nodes. The same happened to the spatial correlation, whereby the misbehave node tends to spread the event over its neighboring nodes in the same transmission range.

TABLE VI: COVARIATE ESTIMATION FOR FUZZY SYSTEM OVER CORRELATED NODE BEHAVIOR DETECTION

Covariates	$\hat{\beta}$	Standard Error (SE)	p-value
Attribute	1.28	0.264	0.000
Spatial	0.745	0.224	0.000

IV. CONCLUSIONS

In this paper, the study proposed the formulation model of fuzzy correlated system for node behavior detection in Wireless Sensor Network. The model basically represents the flexibility of the node behavior determination using fuzzy logic. In addition, the fuzzy logic is capable to confront the uncertainties in the wireless sensor network. The model observes the neighboring node behavior in routing operation between a source and a destination. Therefore, the performance of the network highly dependent on the performance of nodes in a cluster represented by the level of cooperation by the intermediary nodes. For this model, the behavior of node can be regulated by the packet forwarding, energy level, and the throughput. The simulation results show that the correlated node behavior detection based on fuzzy logic system can be modeled by considering both attribute and spatial correlation acts.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Universiti Sains Islam Malaysia (USIM) for the support and facilities provided. This research paper is supported by Fundamental Research Grants No: USIM/FRGS/FST/32/51414.

REFERENCES

[1] J. Sen, "Sustainable wireless sensor networks," *Sustain. Wirel. Sens. Networks*, pp. 279–309, 2010

[2] T. Thanakornworakij, R. Nassar, C. B. Leangsuksun, and M. Paun, "The effect of correlated failure on the reliability of HPC systems," in *Proc. Parallel Distrib. Process. with Appl. Work. (ISPAW), 2011 Ninth IEEE Int. Symp.*, 2011, pp. 284–288.

[3] Y. Xu and W. Wang, "Characterizing the spread of correlated failures in large wireless networks," *Proc. - IEEE INFOCOM*, 2010.

[4] S. A. Khan, B. Daachi, and K. Djouani, "Application of fuzzy inference systems to detection of faults in wireless sensor networks," *Neurocomputing*, vol. 94, pp. 111–120, 2012.

[5] K. Salahshoor, M. S. Khoshro, and M. Kordestani, "Simulation modelling practice and theory fault detection and diagnosis of an industrial steam turbine using a distributed configuration of adaptive neuro-fuzzy inference systems," *Simul. Model. Pract. Theory*, vol. 19, no. 5, pp. 1280–1293, 2011.

[6] S. Shamshirband, *et al.*, "Co-FAIS: Cooperative fuzzy artificial immune system for detecting intrusion in wireless sensor networks," *J. Netw. Comput. Appl.*, vol. 42, pp. 102–117, 2014.

[7] U. B. Nisha, U. Maheswari, N. Venkatesh, and A. R. Yasir, "Improving data accuracy using proactive correlated fuzzy system in wireless sensor networks," *KSII Trans. INTERNET Inf. Syst.*, vol. 9, no. 9, pp. 3515–3538, 2015.

[8] T. V. P. Sundararajan and A. Shanmugam, "Modeling the behavior of selfish forwarding nodes to stimulate cooperation in MANET," *Int. J. Netw. Secur. Its Appl.*, vol. 2, no. 2, 2010.

[9] A. H. Azni, A. Rabiah, M. N. Zul Azri, H. B. A. Samad, and H. Burairah, "Correlated node behavior model based on semi markov process for MANETS," *J. Comput. Sci. Issues*, vol. 9, no. 1, pp. 50–59, 2012.

[10] S. Gopalakrishnan and P. M. Kumar, "Performance analysis of malicious node detection and elimination using clustering approach on MANET," *Circuits & Systems*, vol. 7, no. 7, pp. 748–758, 2016.

[11] H. Kothari and M. Chaturvedi, "Effect of selfish behavior on power consumption in mobile ad hoc network," *Proceedings of the Asia-Pacific Advanced Network*, pp. 91–100, 2011.

[12] E. R. Panwar and S. Malhotra, "Fuzzy based QOS in WSN," vol. 2, no. 5, pp. 457–462, 2014.

[13] M. M. Javidi and M. V. Baseri, "Fuzzy selfish detection ad hoc on-demand distance vector routing protocol (FSDAODV)," *Journal of Computer Science & Computational Mathematics*, vol. 7, no. 1, March 2017.

[14] L. Podofillini, V. Dang, E. Zio, P. Baraldi, and M. Librizzi, "Using expert models in human reliability analysis — a dependence assessment method based on fuzzy logic," *Risk Analysis*, vol. 30, no. 8, 2010.

[15] A. H. Azni, A. Rabiah, and M. N. Zul Azri, "Modeling stochastic correlated node behavior for survivability in ad hoc networks," *Int. J. Cryptol. Res.*, vol. 4, no. 1, pp. 1–16, 2013.

[16] H. Hallani and S. A. Shahrestani, "Mitigation of the effects of selfish and malicious nodes in Ad-hoc networks," *WSEAS Trans. Comput.*, vol. 8, no. 2, pp. 205–221, 2009.

Noor Shahidah graduated her Bachelor's Degree in Financial Mathematics from Universiti Sains Islam Malaysia (USIM) in 2016. She is currently pursuing Master of Science in Information Security and Assurance in Universiti Sains Islam Malaysia (USIM). She has experiences in presenting research paper at international conferences and seminar. Currently, she has published number of journals and proceeding papers and her research interest is on Wireless Security.

A. H. Azni obtained her PhD in Computer Science (Wireless Security) from Universiti Technical Melaka Malaysia (UTeM) in 2014. She has many experiences in presenting research talks and papers at national and international conferences. She has also published tremendous articles in highly esteemed journals. Her research interests are on Wireless Security, IoTs, and Cryptography.