4
Sociopsychological Trust Model for Wireless Sensor Networks

Trust has always been a major concern for a variety of fields such as sociology, psychology, philosophy, computer networks and social networks. It is an elusive notion, mostly because of its subjectivity and context-specificity. In an organisation which incorporates people from diverse backgrounds into its network, each member has to trust others to accomplish one's own personal goals and the collective goal of the organisation. Trust plays a decisive role in establishing and sustaining harmonious relationship among the members and facilitates rational predictions of the dynamics of relationship within the organisation.

In a Wireless Sensor Network, where the sensor nodes coordinate with each other for monitoring environmental conditions and sending the data to the base station, it is essential that trust is established among the nodes so that they would confidently rely on each other and send the data faithfully. Nevertheless, owing to certain hardware and software faults, nodes can behave fraudulently and send fraudulent information. The errors that take place while sending information can decrease the trust ratings of the sensor nodes.

The present chapter provides a novel approach for computing trust of sensor nodes. Section 4.1 explicates the sociopsychological account of trust and analyses the major concepts that are related to trust. Section 4.2 proposes the model that can be used in WSN for generating trust ratings of the sensor nodes. Discussing the proof of concept in Section 4.3, Section 4.4 evaluates the efficiency of the proposed work by comparing it with the machine learning model. Finally, Section 4.5 concludes the discussion.

4.1 SOCIOPSYCHOLOGICAL BACKGROUND

Trust is a subjective phenomenon which anchors on a number of factors that collectively construct the quality of trustworthiness. Merriam-Webster dictionary defines trust as the belief that someone or something is reliable, good, honest or effective [Merriam-Webster, 2014]. Trust, in the primary sense, is a feeling, or an emotion, or an affect. Trusting is a major concern, which is not only urgent and important in one sense, but also comes first and foremost as a matter of personal responsibility [Flores and Solomon, 1998]. Accordingly, trust is largely subject-dependent.

Trust is a trait having congruence between the desired and perceived participation [Driscoll 1978]. It is characterised by hope, faith, confidence, assurance and initiative [Lewicki et al, 1998]. When one person trusts another, he/she has the confidence that the other person has the ability to produce the expected output and the good intention to perform the task. It may be argued that trusting someone or something makes the actions of an individual predictable.

There are two parties, viz. a trustor and a trustee, involved in a trust relationship as shown in Figure 4.1. The person who trusts someone or something is called a trustor and the person who has been trusted is called a trustee. There should be a trusting intention (willingness) as well as a trusting belief between the two parties [McKnight et al, 1998)]. Additionally, some tact and willingness on the part of the trustor and some willingness on the part of the trustee, both to be forgiven and to forgive unfair criticisms, look essential to make

the trust relationship successful [Baier, 1986]. Trust involves a two-way process which counts both giving and receiving. For instance, in a mother-child bond, there exists mutual faithfulness between the two. It means, trust is not something that simply happens or gets intuited. Rather, it is created through interactions which flourish the relationships. The bond so developed is not something that takes place within a few minutes or that has to be limited to happening within a certain time period after birth. With genuine love and care-giving, the bond becomes stronger and stronger over time.

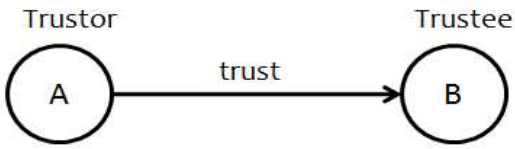


Figure 4.1: Parties Involved in Trust Relationship

4.1.1 Building Blocks for Sociopsychological Model

There are two building blocks, namely, cognition and affection, with which trust is being built. Cognition is the learning which is based on perceptual reasoning. Cognitive learning depends on:

(1) Success of past interactions with the trustee, i.e., past records.
(2) Extent of social similarity, i.e., the group of friends the trustee keeps.
(3) Context considerations, i.e., the credentials owned by the trustee.

Affection is the gentle feeling of fondness or liking. It is based on:

(1) Nature of behavior of trustee.
(2) Frequency of interaction between the trustor and the trustee. (More the frequency of interaction, higher is the trust. This is because interaction makes them comfortable in sharing, which in turn, increases the closeness and reliability.)

If both cognition and affection are high, the ideal level of trust is available. Generally, cognition is followed by affection, and once affection is high, the foundation of cognition-based trust may not be needed. For instance, people trust others initially in the light of learning through cognition, and once cognition is developed, affection increases with time. Likewise, when affection increases, the information attained through cognitive learning is overlooked. Finally, it is to be added that trust is a dynamic and continuous variable, because it is dependent on relationships that are ever changing.

4.1.2 Characteristics of Trust

The following section examines various characteristics of trust, namely, asymmetric property, transitive property, personalised property and context dependent property.



Figure 4.2: Asymmetric Trust

Trust is Asymmetric: Between two parties, the trust level is not identical as shown in Figure 4.2. A may trust B fully; however, B may not necessarily feel the same way about A. B may only trust A partially. For example, in the case of husband-wife relationship, wife may trust husband in most things but she may be skeptical about his ability to cook food.

Trust is Transitive: Arguably, trust can be transitive. Let us say, A and B both know each other very well and are best friends. Now, consider B has a friend named C whom A has never met. But, since A knows B so well and trusts B's choices in making friends, A may trust C to a certain extent even though they might have never met. This is also called as *trust propagation* [Guha et al, 2004*]*. See Figure 4.3.

Now let us say, C has a friend named D whom neither A nor B knows well. Here A could find it hard to trust D. Accordingly, as the link among nodes grows longer, trust level decreases.

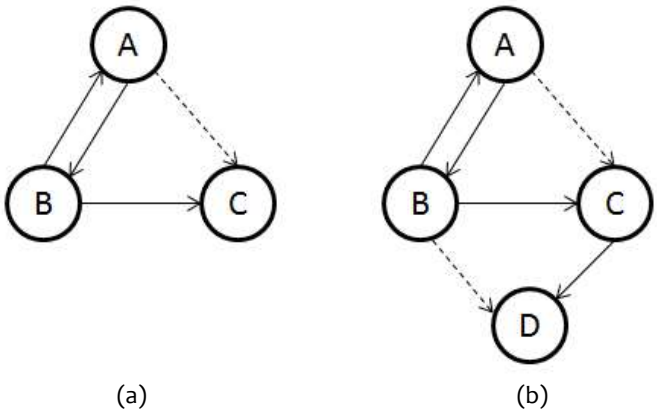


Figure 4.3: Transitive Trust: (a) A trusts B, B trusts C, implying A trusts C; (b) A could find it hard to trust D as the link among nodes increase

Trust is personalized: Any relationship can be conceived either as something positive or as something negative. Consider a social network of a set of three people, i.e., A, B, and C. There can be the following scenarios as shown in Figure 4.4.

(1) Suppose B and C both trust A. It is natural that B and C will also mutually trust each other. This type of relationship is balanced.
(2) Now, consider both B and C do not trust A. In a similar manner, it is fair enough that both B and C can trust each other.
(3) The problem appears when one trusts and other distrusts. For instance, B trusts A and C does not trust A. In this case, B and C cannot trust each other. This type of relationship is unbalanced.
(4) Likewise C trusts A and B does not trust A. In this case too B and C cannot trust each other. Suspicion and doubt occurs in this case.

These conditions are likely to happen in human society when a person changes his/her notion of trust to other person with respect to their relationship with the closed ones. These conditions are desirable in WSN as well.

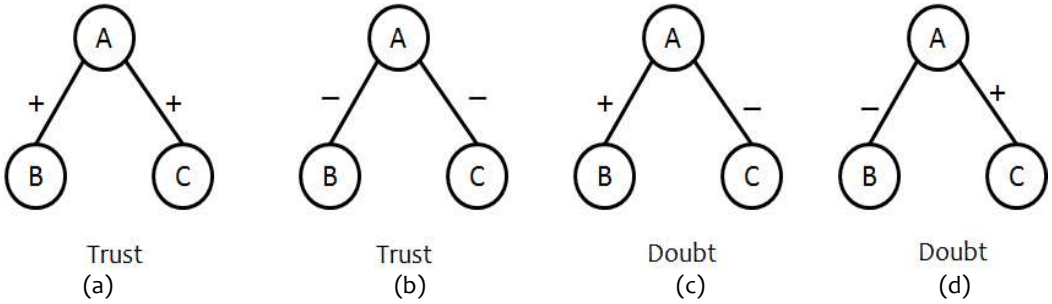


Figure 4.4: Structural Balance

Thus, trust is a subjective phenomenon, because two parties can have very different opinions about the trustworthiness of the same person. For instance, a nation may be divided into many groups one of which strongly supports the political party in rule while other groups strongly disagree with the ruling party.

Trust is context-dependent: Trust is closely associated with overall contexts. Stated otherwise, trust is context-specific [Gray, 2006]. One may trust another enough to lend a pencil, not a laptop.

4.1.3 Trust Model

An in-depth analysis of sociopsychological account unveils that trust is a function of three factors: ability (A), benevolence (B) and integrity (I). The following equation and Figure 4.5 [Mayer et al, 1995] elucidate this idea.

$$Trust = f(A, B, I) \tag{4.1}$$

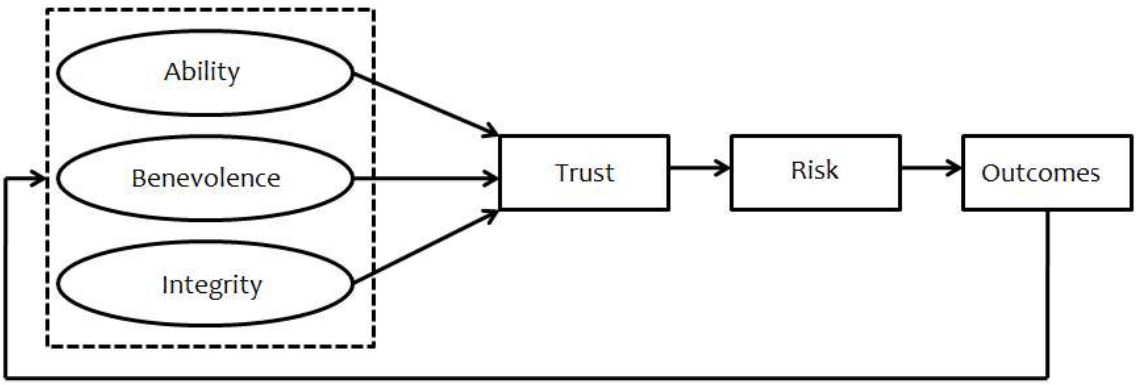


Figure 4.5: Trust Model

Ability denotes the capacity of a person in performing a given task. Higher the ability, higher is the trust level. Likewise, lower the ability, lower is the trust level. However, ability is domain specific, and high ability in one task does not guarantee high ability in other tasks.

Benevolence is the stable disposition which capacitates the trustee to do good for the trustor. Effect of benevolence is likely to increase over time as the relationship between the two parties develops.

Integrity is the virtue, which prompts one to act always in accordance with one's own principles. The role of integrity will be more significant in the early stages of relationships. This is because it is necessary to confirm the consistency and coherence of positive outputs the person had delivered in the past.

Trust is closely linked to the feeling that others will not take undue advantage of the trustor [McAllister, 1995]. Relying on others is difficult when there is uncertainty about their intentions, motives and competencies [McEvily et al, 2003]. In this context, risk is defined as the extent to which there is a probability about whether potentially significant or disappointing outcomes of decisions may be realized [Sitkin and Pablo, 2004]. If this risk is more than the optimal level of trust, one may not get into a relationship. Hence, there is a need to take a leap of faith which is optimum enough that one is confident of the other in not exploiting another's vulnerability. Over-investing or under-investing in trust may misallocate precious resources or miss out the opportunities respectively [Wicks et al, 1999]. The optimum trust appears when there exists a perfect congregation of ability, benevolence and integrity. Refer to Figure 4.6. The weight given to each of these components is not same in reality. In organisation, people tend to give benevolence and integrity factor more than the ability factor.

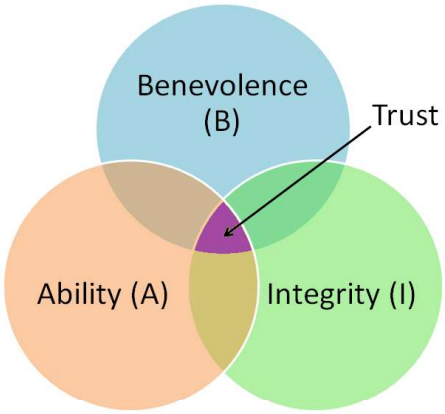


Figure 4.6: Venn Diagram of Three Factors in Building Trust

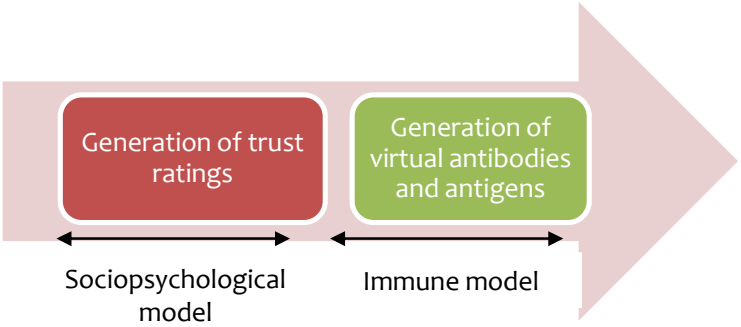


Figure 4.7: Sociopsychological Trust Model

4.2 PROPOSED WORK

The proposed work has two essential blocks, namely, the sociopsychological model and the immune model as shown in Figure 4.7. The sociopsychological model is introduced for generating trust ratings of sensor nodes with respect to three important factors, i.e., ability, benevolence and integrity. Once the trust ratings fall below a particular threshold, the immune model comes into the picture. The immune model is used for removing the fraudulent nodes and thereby generating correct readings to the base station. Precisely, once the trust ratings of the fraudulent nodes fall down, the immune model prompts the fraudulent nodes to give correct readings to the base station. The model simultaneously reduces the readings intake of the suspected fraudulent nodes by decreasing the sampling interval.

4.2.1 System Assumptions and Model

Since the WSN is a densely deployed network, the data is spatially and temporally correlated [Akyildiz et al, 2004].

(1) *Spatial Characteristic*: When the nodes are densely deployed in the field, the data is correlated with the nearby or neighbouring nodes.
(2) *Temporal Characteristic*: Sensor nodes are sometimes required to observe and transmit the sensed event features periodically in some of the WSN applications such as event tracking. There is a correlation among the consecutive sensor measurements sent by the sensor node.

WSNs reliably detect/estimate event features from the collective information provided by sensor nodes. The two major applications of sensor network are environmental monitoring and event detection. The proposed work focuses on the applications of environmental monitoring where there is spatio-temporal correlation among the sensor observations.

A fault is detected when the observations exceed the expected normal behavior. However, what anomalous behavior has been defined in an environmental monitoring system can be said as an alarm or event in case of event detection systems. An assumption is made that all sensor data is forwarded to a central location where the data processing occurs. Likewise, the sensor nodes are not corrupted by any communication fault which could delay the rate at which packets are sent. The uncollected data are not considered as the fault for the system. For the present purpose, focus on fraudulent attacks like sybil attack, sink hole attack etc. are not taken into account as discussed in Chapter 2[Pathan et al, 2006].

The system model comprises of sensor nodes who relay their information to the base station via relay nodes. The relay nodes/base station compute the trust ratings of the nodes with the help of the information provided by the sensor node and its immediate nearby nodes. All the data processing occurs at the base station. The base station has the authority to turn off the fraudulent nodes.

4.2.2 Sociopsychological Trust Model

Paying attention to the sociopsychological concepts of ability, benevolence and integrity, the study finds it fascinating to explore and adopt similar criterion for computing trust of WSN.

Ability of a sensor node points to the efficiency or skill to deliver the desired output faithfully. The sensor node should be capable of gathering, processing and communicating the required sensory information in concordance with other connected nodes in the network. Factors which might affect the ability of sensor nodes are hardware and software constraints. For instance, the power source, the controller, and the external memory should be properly functional. Additionally, a sensor node should be smaller in size and must consume extremely low energy, operate in high volumetric densities, capable to operate unattended, and adaptive to the environment [Akyildiz, 2002]. As wireless sensor nodes are typically very small electronic devices, they can only be equipped with a limited power source of less than 0.5 - 2 ampere-hour

and 1.2 - 3.7 volts. Ability of WSN can be set to 1 if the sensor node is functional otherwise 0. Here 'to be functional' implies that the sensor node is able to send the packet.

Benevolence of a node denotes how good the node is. A benevolent node is the one which sends correct readings in the designed environment. In order to check the goodness of the sensor, the metric is derived as the difference between the actual and the perceived readings (error). It implies that the current reading which the sensor node is sending could be compared with the readings of the neighbouring nodes. More the error between the two, lesser is the benevolence factor. Similarly, the less the error, the more is the benevolence factor of the sensor node.

Integrity denotes how coherently and consistently the sensor node gives correct or positive readings, and it can be calculated by checking the 'quality' of the sensor node. Quality of the sensor node can be evaluated in the light of the number of positive outcomes given by the sensor nodes, that is to say, the number of the positive responses should be higher in comparison to the number of negative outputs. Let p_n refer to the number of positive output values and n_n refer to the number of negative output values. The number of positive outputs could be computed by comparing the error of the readings with a particular threshold i.e. number of times in previous k time steps the error is small i.e., B (benevolence) < δ. Similarly, negative outcomes are calculated using $k - p_n$.

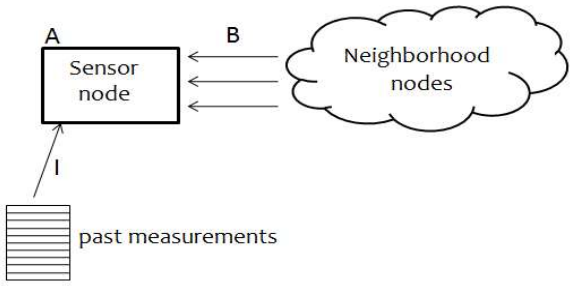


Figure 4.8: Computation of Trust

Table 4.1: Mapping of Benevolence and Integrity to WSN

Benevolence Factor	Integrity Factor	Causes	Psychological Terms	Outcome
Low	Low	Sensor Node is fraudulent: Hardware and software non functional	Mundane trust	Low trust
Low	High	Sensor node is consistently behaving well but produces wrong outcomes due to environmental perturbations	Cognitive trust	Trust ratings less varied
High	Low	Sensor node has good intention to produce output but is not able to do that because of improper calibration or hardware error	Emotional trust	Trust ratings highly varied
High	High	Benevolent node	Ideological trust	Ideal case: High trust ratings

Computation of trust is pictorially represented in Figure 4.8 where *A*, *B* and *I* stands for ability, benevolence and integrity. The figure shows that ability is inbuilt to the sensor node. Benevolence is computed with the help of readings taken from neighbourhood nodes (spatial information) and integrity is computed with the help of readings given by the sensor node in the past (temporal information).

Mapping of benevolence and integrity ratings in WSN nodes is shown in Table 4.1. There may be cases when a node is benevolent but lacks integrity or vice versa. The table gives the psychological term associated with it and explains the reason behind the results along with possible outcomes in WSN scenario.

Now, as stated in the following equation, trust *(T)* of a node n can be defined as the function of *A, B* and *I*, having weights α and β given to benevolence and integrity factors respectively:

$$T_n = A_n(\alpha B_n + \beta I_n) \tag{4.2}$$

where $\alpha + \beta = 1$

$$A_n = \begin{cases} 1 & functional \\ 0 & otherwise \end{cases} \tag{4.3}$$

$$B_n = e^{-(||Error_n||)} \tag{4.4}$$

$$I_n = \frac{p_n}{p_n + n_n} \tag{4.5}$$

where,

$Error = t_n - t_n'$

t_n = current reading from a sensor node x

t_n' = average of readings received from all the neighbours of sensor node x

p_n = number of positive outcomes = number of times in previous k time steps the error is small i.e., $B < \delta$

n_n = number of negative outcomes = $k - p_n$

To compute the number of positive and negative outputs a threshold was set *($\delta = 0.001$)* where:

$$Counts\ the\ number\ of\ responses = \begin{cases} p_n = p_n + 1 & B < \delta \\ n_n = n_n + 1 & B > \delta \end{cases} \tag{4.6}$$

The threshold is set keeping in mind the measurement range to be accepted. In the present study, ±7 measurement variance was acceptable against the genuine measurements. Weighted Averaging and Decrease Sampling interval (WADS) technique explained in Chapter 3 is used to change the measurements and sampling interval.

4.3 PROOF OF CONCEPT

In order to study the variations of three factors, simulations were carried out in LabVIEW, on ten thermocouple sensor nodes of which one sensor node was introduced with fault. It was made intentionally bad after 20 seconds by introducing the faults. The measurement acquired by the sensor node can be represented in the form of $T_{x,y}$ where y denotes the measurement acquired at time x. To consider any node as benevolent, it should be sending the samples within the range of [μ+ σ, μ- σ], where μ is the mean and σ is the standard deviation (SD) of the samples. In the simulations the mean and SD of benevolent nodes is set as 25 and 2 respectively, considering it is a controlled environment setup. Four types of faults were considered for simulations namely short, constant, noise, and drift [Fang and Dobson, 2014].

(1) *Short*: Temperature readings $T_{x,y}$ were replaced by $T_{x,y}$+ $T_{x,y}$× f, where f is a random multiplicative factor; f ranging from *0.1* to *10.0* is used here.

(2) *Constant*: Temperature readings $T_{x,y}$ were replaced by some random constant c and c is randomly selected from *50* to *999*.
(3) *Noise*: Temperature readings $T_{x,y}$ were replaced by $T_{x,y} + x_g$, where x_g was a Gaussian random variable, whose distribution was *N(0,2)*.
(4) *Drift*: Temperature readings $T_{x,y}$ were replaced by $T_{x,y} + a^{t+1}$, where $a > 1$.

4.3.1 Detection of Fraudulent Node

After the introduction of fraudulent node where noise is introduced as the fault, the variation of the three factors in a common time frame is illustrated in Figure 4.9. The other three faults namely short, constant and drift fault show similar trends. As discussed in Section 4.2, ability can be either 1 or 0, depending on whether the sensor node is functional or not. If the sensor node is not functional, then there is no scope for checking benevolence and integrity of the node. In the present case, assumption is made that the sensor node is always able to send the packet (i.e. Ability = 1). Therefore, the function of ability is always a constant equal to 1.

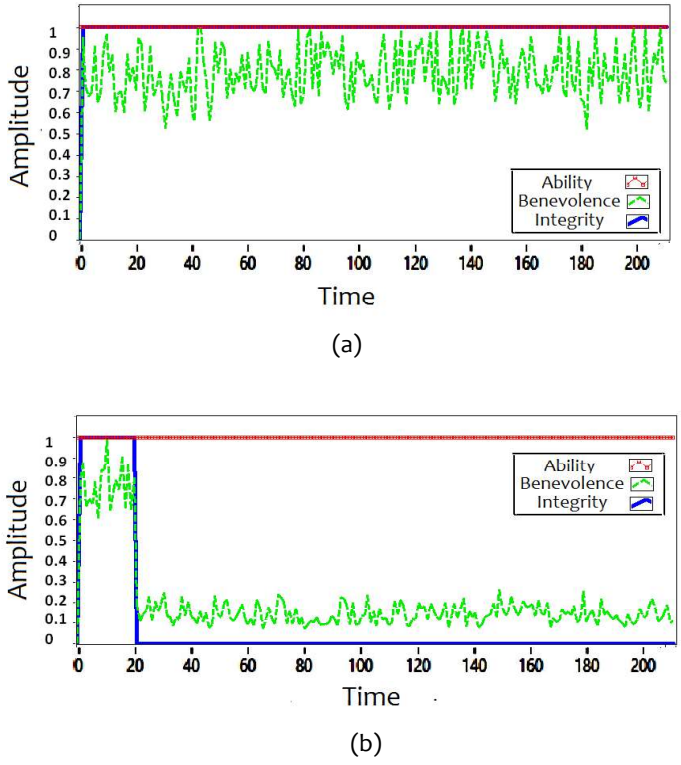


Figure 4.9: Changes in the Three Factors on a Common Time Frame: (a) For a benevolent sensor node the factors are consistent. As the sensor node is able to send packet, ability is equal to 1; (b) Benevolence and integrity factors change in the case of a fraudulent node (noise fault): Initially the node is good; however at t = 20 sec, it starts behaving badly, thereby making benevolence and integrity factors fall

However, the pattern of benevolence on topology of ten sensor nodes showed deviation. To compute the benevolence factor, average of readings from 9 sensor nodes were taken and checked against the current reading received from the selected sensor node. As benevolence is assessed in relation to the readings received from neighbours, the variation is found quite high. In the case of benevolent nodes, the benevolence factor shows variation in the range of [0.6-1] (Figure 4.9). However, when a fraudulent node comes into picture, benevolence factor decreases close to 0. Similarly, the change of integrity of sensor nodes also showed a sudden drop after the

introduction of the fault. It is shown that higher the number of positive outcomes, higher is the integrity. However, when the fraudulent node comes into picture, integrity falls rapidly.

As the three parameters change, the trust factor changes accordingly. To calculate the trust variation Equation 4.2 is used, where the values of α, β are tunable. To study the importance of the three factors in computation of trust ratings, various cases can be taken:

(1) *No ability*: Since ability is an independent term in trust modeling, benevolence and integrity factor do not affect trust modeling. When the node is not able to send packets, its ability becomes zero, thereby making the trust ratings fall down to zero.
(2) High benevolence: α = 0.9, β = 0.1. The pattern as shown in Figure 4.10(a) depicted high variation in the range of 0.6 - 1 and 0.1 - 0.3.
(3) High integrity: α = 0.1, β = 0.9. Here, trust ratings have less variation in comparison to the high benevolence case (See Figure 4.10(b)).
(4) Benevolence and integrity given equal weights: i.e., α, β = 0.5. For optimality, this case is considered as the ideal one, because it stays away both from high variation and low variation (See Figure 4.10(c)).

The final values of α, β are kept as 0.5 because of the trend studied in Figure 4.10 (c).

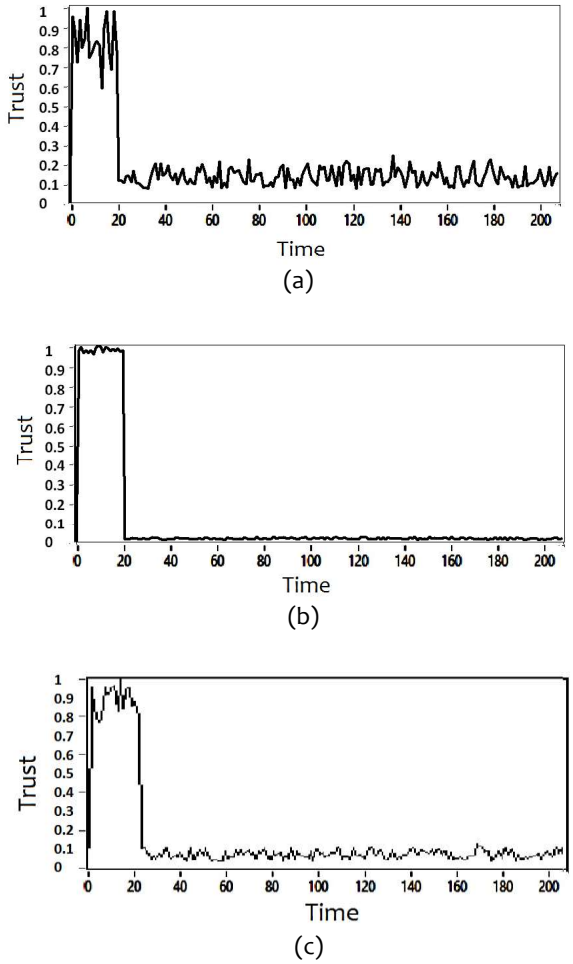


Figure 4.10: Changes in the Trust Ratings of Fraudulent Node on Varying Benevolence and Integrity Factors: (a) Shows high benevolence graph; (b) Shows high integrity graph; (c) Equal weights given to benevolence and integrity: High variation is observed in the case of high benevolence and less variation in the case of high integrity. For optimality, equal weights are assigned to both the factors.

4.3.2 Removal of Fraudulent Node

When the trust ratings fall below the threshold set as T_{trust} = 0.5, the node is considered as fraudulent and eventually immune model is instigated. Figure 4.11(b) represents the raw measurements collected from a fraudulent, unsecured, sensor node which has noise fault.

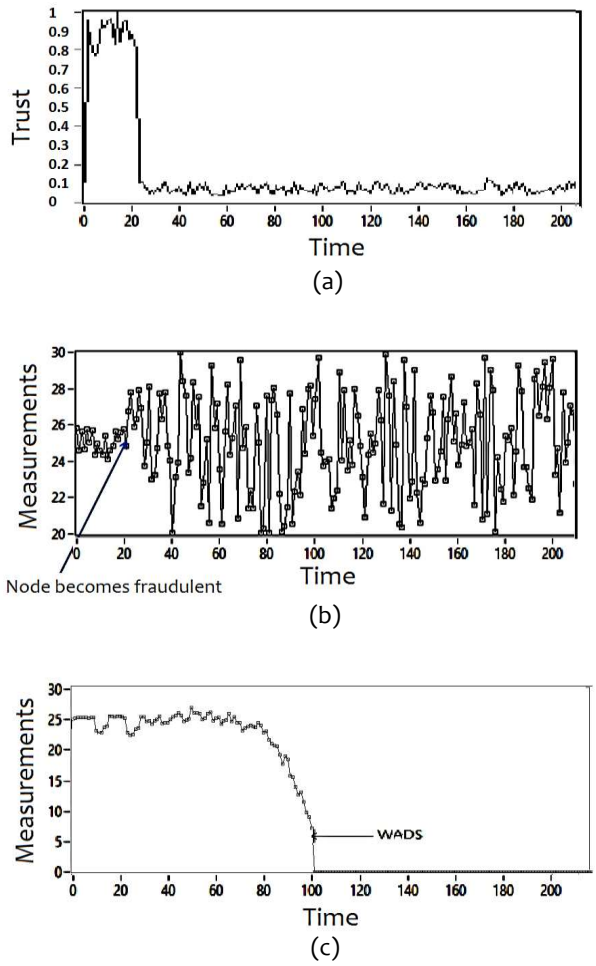


Figure 4.11: Overall Model Representation: (a) Shows the trust ratings of the fraudulent node. After 20 seconds the node becomes fraudulent due to high noise and because of this ratings fall down; (b) Shows the measurements of the fraudulent node (noise fault); (c) Shows measurements of fraudulent node on applying the concept of antigen and antibody. WADS approach (performed at base station) clearly accepts readings for some time even if the node is behaving fraudulently (from 20 to 80 sec).

As studied in Chapter 3, WADS technique was used to change the readings and sampling interval. To implement the proposed work, weights are assigned to the temperature readings whose values are proportional to the antigen quantity taken from the Dibrov's differential equation. The measurements were calculated as per the following equation:

$$T_{new} = \frac{\sum_{i=1}^{k} g_i \times T_{prev}}{k} \quad (4.7)$$

where T_{new} is the new measurement obtained by applying weights, T_{prev} are the previous measurements, k is the history length, g_i are the antigen quantity which can be treated as the weight given to measurements.

The rate at which the samples were collected from the fraudulent node was decreased by taking into account the antibodies' value from differential Eqs.(3.10) and (3.11). The rate at which the sampling interval is decreased is:

$$s_{after_attack} = \frac{s_{prior_to_attack}}{2^{\left\lfloor \frac{a_i}{\frac{a_{max}-a_{min}}{k}} \right\rfloor + 1}} \tag{4.8}$$

where s_{after_attack} is the sampling interval after the fraudulent node is detected and $s_{prior_to_attack}$ is the sampling interval prior to the detection of the fraudulent node, a is the antibody value where a_{max} and a_{min} is fixed to 1 and 0 respectively. k is the number of steps desired to end the influence of fraudulent node.

WADS prove to be an effective method, because even after the node becomes fraudulent genuine measurements are received for some time, thereby giving time to help the network stabilize. Here, fraudulent node's lifetime is 100 sec.

4.4 SIMULATION SETUP AND PERFORMANCE METRICS

The section presents the performance appraisal of the proposed sociopsychological model on various grounds such as efficiency, reliability, detection time etc. Initially the section gives the comparative analysis with the machine learning based model followed by providing detailed comparative analysis with the existing trust models.

4.4.1 Comparative Analysis of Machine Learning with Sociopsychological Model

Broadly there are two groups of methods used for detecting fraudulent nodes: learning based methods and rule based estimation methods [Sharma et al, 2010]. Learning based methods involve machine learning and neural network. The earlier work, i.e., machine learning model uses approaches such as K-means and SVM for identifying fraudulent nodes. However, looking at sensor node constraints, machine learning techniques appear to be significantly expensive. Moreover, these techniques require more time for training since the training data needs regular updating.

The sociopsycholgical model appears to be a better candidate in this regard. It adopts a rule based estimation method which solves simple equations that bring in both the temporal and spatial correlations. Furthermore, the time required for detecting fraudulent sensor nodes is significantly less which in turn reduces energy consumption. Once the detection of fraudulent node is performed, immune mechanism can be used for removing the fraudulent nodes. Accordingly, it simplifies the complexity of the overall method.

One major limitation of the machine learning model is that it uses only temporal information. That is to say, it is dependent only on its own data and no neighbourhood information is required for the detection. In the case of sociopsychological model, both temporal and spatial information are used in the process of detection.

To perform a comparative study between the two models a simulator was developed which creates a wireless sensor network simulation environment. A network generator in the form of grid was built on a field size 20×10 which comprises of randomly placed n benevolent nodes and m fraudulent nodes. Networks of size 10, 20...200 nodes were generated and were given as inputs to the simulator. For each network size, different topology was generated.

Following are the five performance metrics used in this work:

(1) *Detection and recovery time*: The detection time is defined as the amount of time required for identifying the fraudulent nodes. The recovery time is defined as the time to stabilize the network after fraudulent nodes get detected. See Figure 4.12.

$$Detection\ time = t_y - t_x \quad (4.9)$$

$$Recovery\ time = t_z - t_y \quad (4.10)$$

(2) *Reliability*: The reliability of the system signifies how much correct a system is.

(3) *Robustness*: Robustness of a system is defined as the ability to detect maximum fraudulent nodes.

(4) *Scalability*: Scalability is the ability of a system to handle a large number of nodes, viz., how the system responds when the number of nodes is increased in the network. The scalability factor (SF) is a constant or a function on which the output depends on a given input and the results were obtained using a continuous pattern. Therefore, in this case, it is obtained with regard to detection ratio, on *n* (network size), which is calculated using:

$$SF = \left\lfloor \frac{n-1}{2} \right\rfloor \times \frac{1}{n} \quad (4.11)$$

(5) *Space and time complexity*: Total number of steps involved in a solution is the function of the size of the problem, which is the measure of the problem's time complexity. Space complexity is calculated in reference to the amount of computer memory required during program execution, as a function of input size.

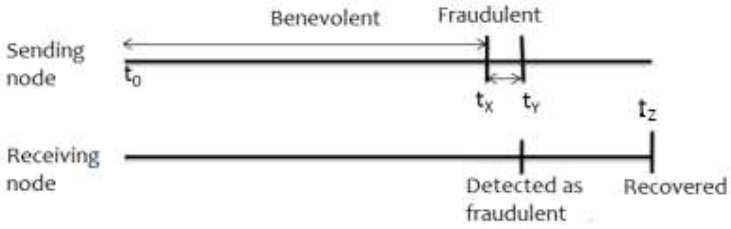


Figure 4.12: Detection and Recovery Time

(a) Detection and recovery time

For the detection time, in the case of machine learning model, the amount of detection depends on the training set. Since the test data compares only the fraudulent data samples, samples can be searched in *O(log s)* time, where *s* is the number of fraudulent data samples. Whereas, in the case of sociopsychological model, the detection time is less which is because of the integrity component. Comparing benevolence *(B)* with small delta (δ) makes the detection time of the fraudulent node a constant time operation. Figure 4.13(c) compares the detection time of machine learning model with the time required in the proposed model. It is evident that the sociopsychological model performs better than the machine learning model.

Figure 4.13(a) presents the graph of detection time vs network size which shows, when the network has one fraudulent node *(m)* the detection time is less for network size > 3. Similarly for $m = 2$, network size > 5 has less detection time. Figure 4.13(b) gives the possible reason behind this behavior. The graph shows the effect of integrity component on networks of different sizes. Furthermore, it explains the probability of a node behaving badly at varying time steps. The rate of change is high when the number of nodes is less and it decreases as the number of nodes is increased in the network. Further, less than 100 time steps are required to reach stability. Similar results can be obtained if the number of fraudulent nodes is increased.

In the proposed model, recovery time is very well managed by the immune system, as shown in Figure 4.11(c). The figure reveals that good readings are obtained even after the node becomes corrupted.

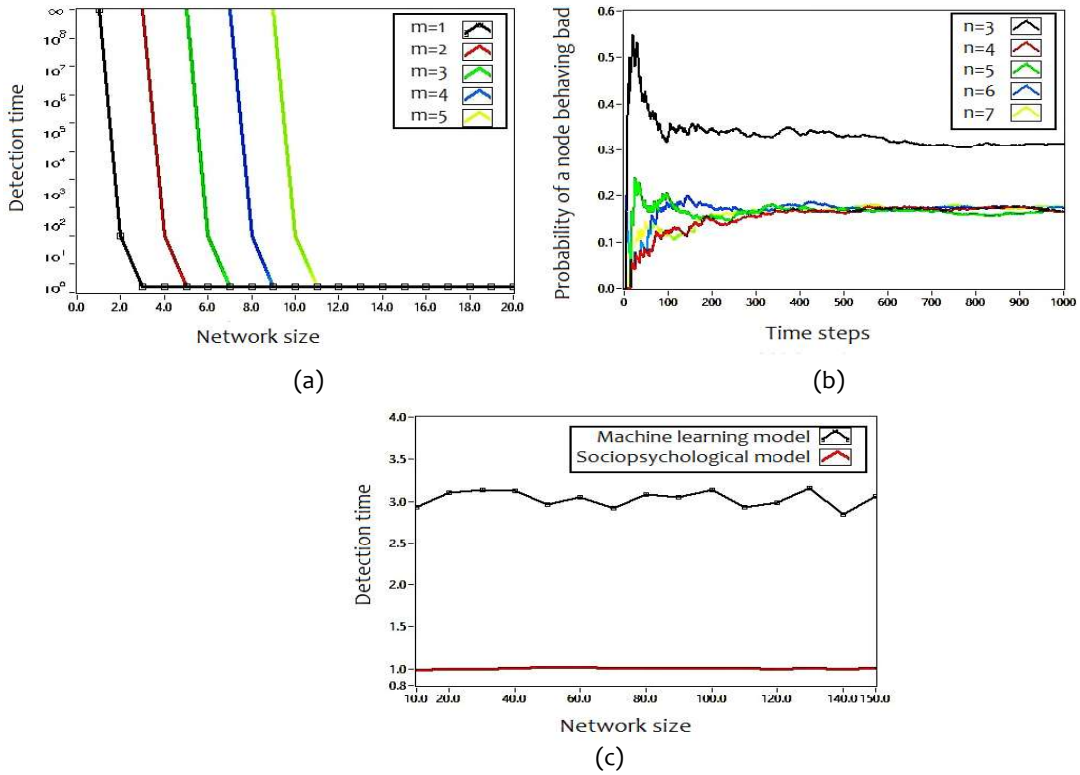


Figure 4.13: Detection and Recovery Time (Noise Fault): (a) Shows the variation of detection time vs network size for sociopsychological model. When the number of fraudulent nodes in a network is 1, with network size > 3, response time is less; (b) The plot gives the variation of probability of a node behaving bad vs time steps on different network sizes for m = 1 (Sociopsychological model): Explains the plot of integrity providing the possible reason for the response time to be less; (c) Compares the detection time of machine learning model with the sociopsychological model

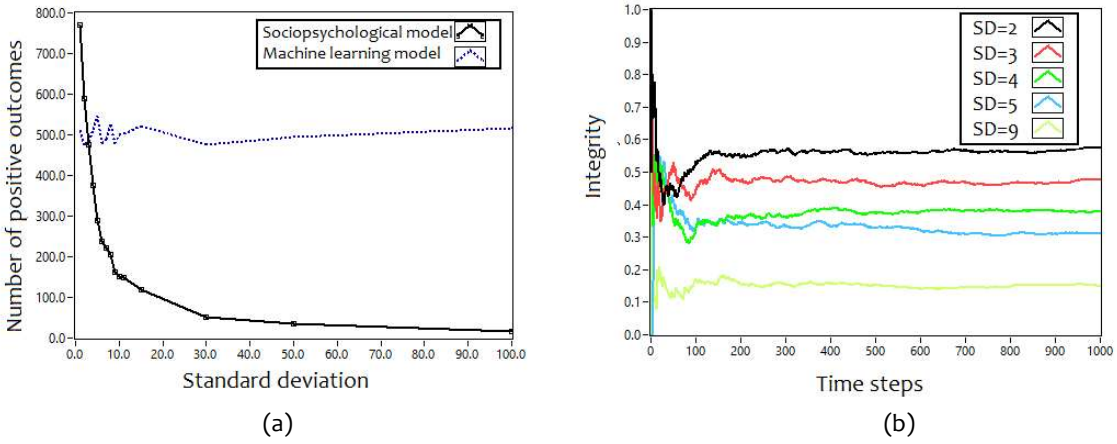


Figure 4.14: Reliability Curves: (a) Shows the plot of number of positive outcome vs standard deviation; (b) Shows the variation of integrity vs time steps for varying standard deviation.

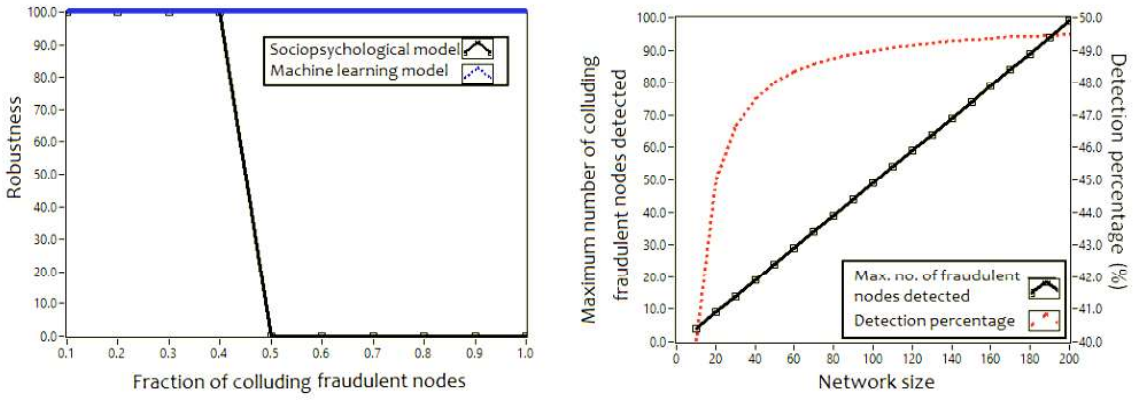
(b) Reliability

To consider any node as benevolent and reliable, it should be sending the samples within the range of $[\mu+\sigma, \mu-\sigma]$, where μ is the mean and σ is the standard deviation (SD) of the

samples. In the simulations the mean and SD of benevolent nodes is set as 25 and 2 respectively, considering it is a controlled environment setup. That is to say the system is capable of taking the readings having SD of 2. In machine learning model, when different values of SD were considered, the system output did not vary, because the K-means has its own way of creating fraudulent and benevolent clusters. Therefore, the number of fraudulent data samples for different values of SD was the same, i.e., around (= 500). Figure 4.14(a) presents the plot of the number of positive outputs from the nodes vs SD. It implies that the changes in SD do not affect the number of positive outputs in case of Machine learning model. In sociopsychological model, the samples sent by fraudulent nodes would consistently fall outside the range $[\mu+ \sigma, \mu- \sigma]$. Consequently, the number of positive outputs would decrease which in turn results in the reduction of integrity factor. It further leads to the reduction of trust ratings. Figure 4.14(b) shows the plot of integrity vs time steps on varying standard deviation. As SD increases, integrity decreases. For instance, for SD > 2, effect of integrity falls below 0.5 which is the bottom limit. It is clear that the probability of node behaving good decreases for a higher SD.

(c) Robustness

The machine learning model presumes that robustness is 100% dependent on the soundness of the data set chosen. However, choosing such an ideal scenario of data set is very difficult. Robustness results for the sociopsychological model are shown in Figures 4.15(a) and (b). Figure 4.15(a) is the plot of robustness vs fraction of colluding fraudulent nodes. The sociopsychological model worked well for independent fraudulent nodes scenario (~99%). However, the scheme works well when the fraction of colluding fraudulent nodes is less than 0.5 of the number of nodes in the network. Stated otherwise, robustness is 100% for less than 0.5 fraction of colluding fraudulent nodes. Figure 4.15(b) brings the number of detected nodes for different network sizes. On a 10 node network, the maximum number of colluding fraudulent nodes that can be detected is 4. The linearity of the graph proves that more number of colluding fraudulent nodes is detected when the number of nodes is increased in the network. Furthermore, the graph shows the result of robustness on y_1 axis and the detection percentage on y_2 axis with a common axis x representing network size. The detection percentage is the maximum percentage of nodes detected on given network size.



(a) (b)

Figure 4.15: Robustness Curves: (a) Shows the plot of robustness vs fraction of colluding fraudulent nodes for the two models; (b) Shows the result of sociopsychological model in terms of maximum number of colluding fraudulent nodes detected on y_1 scale and detection percentage on y_2 scale with a common x scale of network size. Please note: For independent fraudulent nodes the detection percentage is ~99%.

(d) Scalability

With regard to the scalability factor of the system, the sociopsychological model can reach up to the detection rate of *49.5%* for $n = 200$, since the average of readings received from the neighboring nodes was taken. The above advantage of the model is represented by the dashed curve in Figure 4.15(b). The curve reveals that the maximum detectable colluding fraudulent nodes can reach up to *50%* even when the number of nodes was increased in the network. In machine learning model, the scalability issue does not make any change in the detection rate because it takes only the temporal information. Owing to this reason, the detection rate is *100%* depending upon the training data set.

(e) Space and Time Complexity

The machine learning model comprised k-means, SVM and ADE. The time complexity of k-means is $O(sk_c)$, where k_c is the number of clusters and s is the number of samples. Precisely, it is $O(sm_vk_c)$ when there are m_v number of vector space [Yu et al, 2010]. The worst case complexity of k-means on s samples is $2^{\Omega\sqrt{s}}$ where Ω denotes best case lower bound, but this bound might or might not be tight [Arthur and Vassilvitskii, 2006]. On similar grounds, the space complexity of k-means is $O((s+k_c)m_v)$.

SVM is a type of NP-Complete problem [Tsang, 2005]. There are two complexities involved in this regard: at training time and at test time. For linear SVMs, estimation of the vector w and bias b at training time is performed by solving a quadratic problem; at test time prediction is linear in the number of features and constant in the size of the training data. For kernel SVMs, at the training time selection of support vectors is made; and at the test time, the time complexity is linear both for the number of the support vectors (which can be lower bounded by training set size × training set error rate) and for the number of features (since most kernels only compute a dot product; this will vary for graph kernels, string kernels, etc). Solving the quadratic problem and choosing the support vectors is generally very hard. Furthermore, merely testing an optimal solution to the SVM problem involves the order of n^2 dot products alone while solving the quadratic problem directly involves inverting the kernel matrix which has a complexity of the order of s^3 (where s is the size of training set) [Bordes, 2005]. Space complexity depends on the number of training samples stored at each iteration by the solver [Bottou and Lin, (2007)]. Accordingly, space complexity of SVM is $O(s^2)$.

Since the mean and standard deviation are the only concern of ADE, the complexity turns out to be $O(s)$. Therefore, the overall space and time complexity of machine algorithm happens to be very high.

The proposed sociopsychological model has the advantage of having a very low time and space complexity in comparison to the machine learning model. The model has three factors, such as ability, benevolence and integrity, for computing the trust factor. Ability is a constant function which makes the complexity $O(1)$ whereas benevolence is the exponential function. The exponential function, generally has the algebraic complexity $O(\log^2 n)$, which can be evaluated with relative error $O(2^{-n})$ using $O(\log^2 n)$ infinite-precision additions, subtractions, multiplications and divisions [Ahrendt, 1999]. In the present study, for every node (i) the benevolence factor can be computed to be $O(1)$ as follows:

$$B_i = e^{\frac{nt_i - S}{n-1}} \tag{4.12}$$

where $S = \sum_{k=1}^{n} t_k$, n = number of nodes. Total time complexity of calculating benevolence function is $O(n)$ where n is the total number of nodes. Similarly, for computing integrity, for every node (i), complexity takes $O(1)$, because there are fixed number of statements which are executed. Therefore, the time complexity of integrity is $O(n)$ and the overall complexity of the model turns out to be less in comparison to the machine learning model. The space complexity of the sociopsychological model turns out to be $O(nk)$, where n is the number of nodes and k is the history length. Table 4.2 presents space and time complexity of both the models.

Table 4.2: Space and Time Complexity

	Space Complexity	**Time Complexity**
k-means	$O((s+k_c)m_v)$	$O(sk_c)$
SVM	$O(s^2)$	$O(s^3)$
ADE	$O(s)$	$O(s)$
Machine Learning	$\boldsymbol{O(s^2)}$	$\boldsymbol{O(s^3)}$
Ability	-	$O(1)$
Benevolence	$O(n)$	$O(n)$
Integrity	$O(n)$	$O(n)$
Sociopsychological	$\boldsymbol{O(n)}$	$\boldsymbol{O(n)}$

Note: K_c = number of clusters, m_v = number of vectors, s = number of samples, n = number of nodes

The machine learning model requires high processing on sensor nodes, is highly complex and energy consuming. High complexity which is inherent to k-means, SVM and ADE makes it difficult to implement them even on relay nodes/base station. Another major limit of machine learning model is its higher storage requirement mostly because of on-line data processing. On the other hand, the sociopsychological model which uses simple equations should have less energy consumption.

4.4.2 Comparative Analysis of Sociopsychological Model with other Trust Models

Trust models are efficient in detecting fraudulent nodes in the network. Various metrics can be used for comparative analysis among the models. This section presents the comparison of the proposed sociopsychological model with other trust models as stated in Chapter 2. For comparative analysis, a ten node network topology in ring structure was built where randomly nodes send fraudulent measurements.

(a)Node Reliability vs Node Detection Reliability

Bayesian network primarily uses bayes theorem for the formulation of trust. The model uses the beta distribution function, Gaussian function along with Bayesian fusion techniques for modeling of trust as explained in Chapter 2 (Figure 2.9) [Momani et al, 2007]. The proposed sociopsychological trust model is compared to statistical Bayesian network model in terms of reliability as seen in Figure 4.16(a). For simulation purpose, belief and decision network tool was used [Porter et al, 2013] for modeling decision making under uncertainty. The purpose of using this tool was to illustrate the way in which Bayes theorem work, and how probabilities are calculated within them. For comparison purposes, a ten node network topology in ring structure was built since Bayesian network generally prefers a flat structure, which does not allow more nodes to be added in the network [M R Raj et al, 2013]. Nodes were introduced with noise faults to behave fraudulently with some probability. Figure 4.16(a) shows the plot of node detection reliability vs probability of node failure. It can be inferred from the figure, as the probability of nodes failure increases, the detection percentage decreases in case of Bayesian network. However, the detection percentage is higher in case of sociopsychological model. When the nodes fail with a probability of 0.9, the detection is up to 47% in comparison to Bayesian model which is less than 0.5%. In the case of sociopsychological model, the detection percentage is higher because of the integrity factor. As the node sends fraudulent

measurements, the integrity factor decreases which in turn reduces the trust factor. However, detection is less (50%) in case of colluding fraudulent nodes.

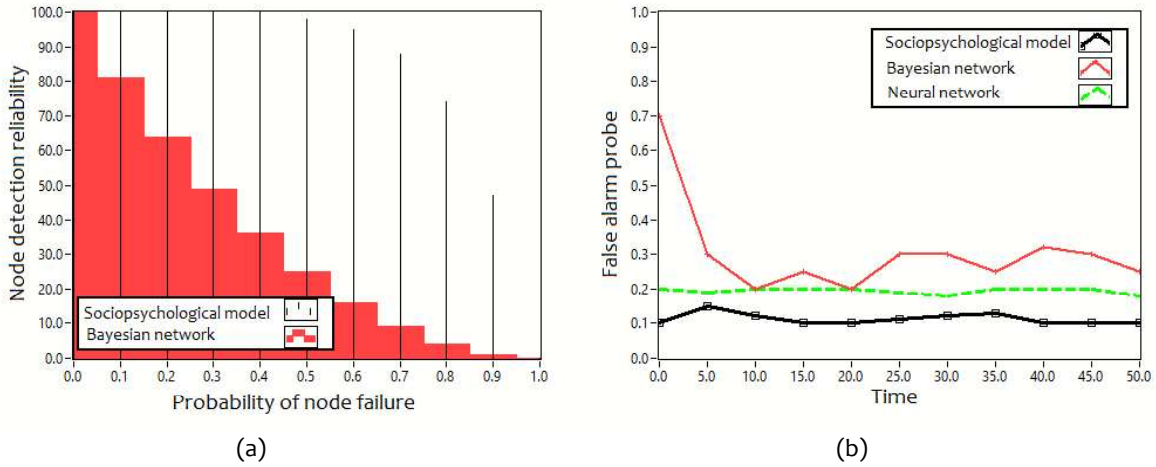


Figure 4.16: Bayesian Network Model vs Sociopsychological Trust Model (a) Comparison of node detection reliability with respect to probability of node failure: It shows that when the probability of node failure is high, then the node detection reliability decreases. However, the decrease is higher for Bayesian network; (b) False alarm probe comparison: It shows the false positive rates for different models. The fault rate is lower in the sociopsychological model.

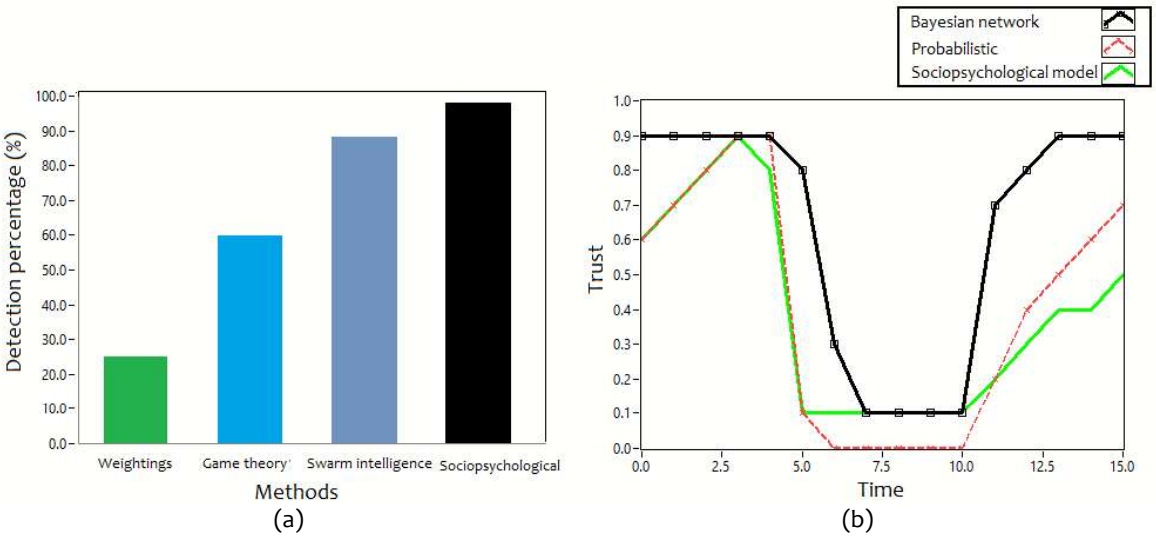


Figure 4.17: Comparative Analysis with other Trust Models: (a) Efficiency graph: High efficiency in case of swarm intelligence and sociopsychological model; (b) Trust reliability curve for ON-OFF attack: Once the node is detected as fraudulent node, the amount of time required to reach optimum level back again is higher in case of the sociopsychological model.

(b) False Alarm Probe

The main disadvantage of Bayesian network technique is that the trust evaluation was based on node's QoS property. Furthermore, the impact of a fixed and a dynamic trust threshold on identifying fraudulent node was evaluated in [Meng et al, 2013]. The work

presented by them states that while using fixed trust threshold mechanism, the false positive rates ranged between 0.31 to 0.68 while the false negative rates ranged between 0.11 to 0.28 against the dynamic trust threshold where the false positive rates ranged between 0.2 to 0.3. On comparing the Bayesian approach to sociopsychological model, the model has further reduced false alarm rates. Additionally, it is noteworthy to compare with the neural network approach which has the output error of 0.2% [Curiac et al, 2007]. Figure 4.16(b) shows the plot of false alarm rate in different models. The sociopsychological model has lower false positive rate due to the integrity factor used in formulation of the trust.

(c) Efficiency in Detecting Fraudulent Nodes

Efficiency is defined in terms of maximum fraudulent nodes detected in the network. The plot for varying efficiency on different models is shown in Figure 4.17(a). The detection percentage in case of the weightings method was lower than 30% [Atakli et al, 2008]. The Game theoretical approach detection was ranged between 30-60% [Shigen et al, 2011]. The Swarm intelligence method is efficient with 90% detection rate of fraudulent nodes [Mármol and Pérez, 2011. Sociopsychological model also has high efficiency in detection of independent fraudulent nodes (~99%). The model has less detection (50%) in case of colluding fraudulent nodes.

(d) Trust Reliability

To check the trust reliability, the proposed model was evaluated for ON-OFF attack. ON-OFF attack is a type of attack where the fraudulent nodes behave good and bad in alternative manner. In case of Bayesian network approach, the trust ratings would raise quickly when the node behaves well after behaving badly since the model was inspired from RFSN [Ganeriwal et al, 2008]. Furthermore, in case of probabilistic approach, the trust ratings would increase slowly based on consecutive cooperation between the measurements [Chen et al, 2007]. It is further reduced in case of socio-psychological approach as seen in Figure 4.17(b). In case of sociopsychological model, the model remembers the bad behavior of previous times which in turn increases the trust reliability. This makes sense, since once the trust lost, takes high time in achieving the optimum trust value back again.

(e) Space and Time Complexity

Space Complexity: Space complexity is defined as the amount of memory required in computation of the model. Ratings, weightings, probabilistic, fuzzy logic methods are less complex having a complexity $O(n)$, where n is the number of nodes. Neural network model has further high complexity $O(RT)$ *per cycle* where R is the total number of neurons and T is the maximum number of activation changes. Bayesian network approach was dependent on the number of components used which in turn generated complexity of $O(cn)$, where c is the number of components. The space complexity increases in case of swarm intelligence and game theoretical approach. In case of swarm intelligence method, the worst case space complexity of maintaining pheromone table is $O(n^2)$ for a fully meshed network of n number of nodes [Bhaskaran et al, 2011].

Time Complexity: Time complexity is defined as the amount of time required for the detection of fraudulent nodes in the network. It was lowest in case of Bayesian network model as Baye's rule was used for detecting the fraudulent nodes. It has the time complexity as $O(1)$. Probabilistic method also has less time complexity of $O(k)$ where k is the history length. Ratings, weightings method have a time complexity of $O(n)$, where n is the number of nodes due to computation on small network topological structure. Neural network has two time complexities associated with it, (a) at training time and (b) at test time. At training time the complexity is of $O(f(\alpha_r))$ where α_r is the learning rate. Once the training is performed, the time complexity of the model turns out to be $O(1)$. Game theory has a time complexity of $O(s_t^3)$ where s_t is the number of states in the game. This is because Nash equilibrium is used for finding the best states. Finding the Nash equilibrium uses matrix row and column operations which is of the $O(s_t^3)$ [Wikipedia, 2013]. Fuzzy logic has a complexity of $O(n+e)$ where e is the number of edges in the

network [Horowitz and Sahni, 1978]. The Swarm intelligence complexity was dependent on the size of the network which in turn limits the application.

The space and time comparative analysis of various trust models is illustrated in Table 4.3.

(f) Miscellaneous

Besides the above metrics, each method has some problems that are independent from other models. While, ratings method was too simple and was applicable only for smaller networks; weightings method was inefficient for weight assignment. It worked under forwarding node notion which is not efficient in comparison to other models. Neural network, despite its high productivity in fraudulent node detection, suffered from high complexity, which in turn is energy consuming and not recommended for resource constrained WSN nodes. Bayesian network approach was applicable only on flat architectures. Thus, the system was not scalable as more number of nodes could not be added in the network. Probabilistic method used only past information for the computation of trust. Fuzzy logic method suffered from memory overhead and inefficiency occurred due to a lot of if-else rules. Swarm intelligence suffered from colluding data node attack.

The comparative analysis of various trust models is illustrated in Table 4.3.

Table 4.3: Comparative Analysis of Different Trust and Reputation Models

Method	Space Complexity	Time Complexity	Miscellaneous
Ratings	$O(n)$	$O(n)$	Applicable for small networks
Weightings	$O(n)$	$O(n)$	Applicable for small networks and weight assignment inefficiency
Neural Network	$O(RT)/cycle$	$O(f(\alpha_r))$ *at training time* $O(1)$ *at test time*	Highly complex and energy consuming
Bayesian Network	$O(cn)$	$O(1)$	Applicable only on flat structure, not scalable
Probabilistic	$O(n)$	$O(k)$	Uses only past information
Game Theory	$O(n^2)$	$O(s_t^3)$	High complexity
Fuzzy Logic	$O(n)$	$O(n+e)$	Memory overhead and inefficiency due to lot of if-else rules
Swarm Intelligence	$O(n^2)$	*Depends on network size*	Cannot capture colluding data node attack
Machine Learning	$O(s^2)$	$O(s^3)$ *at training time* $O(1)$ *at test time*	High complexity
Sociopsychological Model	$O(n)$	$O(n)$	Suffers from colluding data node attack for >50% colluding fraudulent nodes

Note: n is the number of nodes, c is the number of components in Bayesian Network
R is the total number of neurons and T is the maximum number of activation changes,
k is the history length, s is the number of samples
s_t is the number of states, e is the number of edges in the network.

The sociopsycholgical model appears to be a better candidate in this regard. It adopts a rule based estimation method which solves simple equations that bring in both the temporal and spatial correlations. Furthermore, the time required for detecting fraudulent sensor nodes is significantly less which in turn further reduces energy consumption. Once the detection of fraudulent node is performed, Dibrov equations can be used for removing the fraudulent nodes. Accordingly, it simplifies the complexity of the overall method. The model is efficient for detecting up to 50% colluding fraudulent nodes on total number of nodes in the network. It is a light weight model which allows the network to be scalable.

4.5 SUMMARY

Since the three factors such as ability, benevolence and integrity are so crucial in building and retaining trust, it is important that these factors should be optimal for the success of any network of relationships. In a WSN, ability, benevolence and integrity of sensor nodes makes the network trustworthy. The thesis introduces, for the first time, the sociopsychological aspect for establishing trust among WSN nodes. The suggested model is simple yet efficient because it pays due attention to the three major trust factors, such as ability, benevolence and integrity. Assignment of weights to the trust factors substantiates the significance of obtaining the overall trust of sensor nodes. Accordingly, whenever trust ratings fall below the expectable range, fraudulent nodes are to be removed from the system. The method suggested for removing fraudulent nodes follows the antigen-antibody concept of human immune system. The performance metric witnesses the relevance of the proposed model. It shall be interesting to extend the present study to find how the system responds on higher number of colluding nodes scenario.

…