Feature Extraction of Speech Signal and Heartbeat Detection in Angry Emotion Identification

Masnani Mohamed, Lee Chee Chuan and Ida Laila Ahmad

***Abstract*—** Angry is one of emotions that play an essential role in decision making, perception, learning and more. This paper detects the angry emotion by analyzing and recognizing angry speech signal as well as detecting the heartbeat condition. The speech database was uttered by various speakers in different gender and emotions. For the analyzing experiment, several digital signal processing methods such as autocorrelation and linear predication technique was introduced to analyze the features. Then, Artificial Neural Network (ANN) was used to classify each parameter features such as mean fundamental frequency, maximum fundamental frequency, standard deviation fundamental frequency, mean amplitude, pause length ratio and first formant frequency to recognize the emotion. Meanwhile, a heartbeat monitoring circuit was developed to measure the heartbeat. The accuracy of the result has achieved over than 80 percent during emotional recognition test. This method can be used further to recognize angry emotion of patient during counseling session.

***Keywords*—**Artificial Neural Network, digital signal processing, emotions, emotional speech signal and heartbeat.

I. Introduction

ANGER can be defined as any type of changes that causes physical, emotional or psychological strain. It may seriously affect overall health. It can trigger the body's response to perceived threat or danger. During this reaction, certain hormones like adrenalin are released, speeding the heart rate, slowing digestion, shunting blood flow to major muscle groups, and changing various other autonomic nervous functions.

Many researchers prove that assessing emotions is a key to understand human nature, especially for angry emotion. Anger is a state person feels angry, mad and exasperate to a certain condition. This paper is proposed to analyze and recognize the speech signal because it is the most important media of human interaction contains a lot of emotional information [1]. The pitch, tone, timing and energy of speech are all jointly influenced in a nontrivial manner to express the emotional message [2]. In this paper, emotion is roughly classified into two categories which are angry and no angry emotions. The no angry emotion consists of happy, normal and sad speech signal. However the main focus is on angry speech signal. Every emotional speech has its own features, such as fundamental frequency, amplitude, formant frequency, duration and so on. Therefore this project is used to analyze these features and compared the two emotions speech signal in order to obtain the pattern of distribution of different speech.

For the analyzing experiment, in terms of sentences length, environment issue and gender differences are all considered. Generally several methods such as autocorrelation and linear predication technique have been introduced in order to analyze the features, subsequently classify parameter features such as mean fundamental frequency, maximum fundamental frequency, minimum fundamental frequency, standard deviation fundamental frequency, mean amplitude, mean first formant and pause length ratio using Artificial Neural Network in order the obtain relationship between input parameters and output emotions. Eventually a MATLAB GUI has been created by displaying the feature parameters of angry speech signal as well as recognizing the emotion in real time.

It has been widely studied and analyzed that heart rate can be measured by the number of heartbeats per unit of time; an accessible parameter that can be acquired easily [7]. As the heart is beating, an electrical signal is transmitted through the heart muscle and during this moment, the heartbeat can be detected by using infrared sensor. The sensor is placed at fingertips and infrared light is emitted into skin. Basically, the light absorbed is depending on the blood volume [8].

II. Project Methodology

Fig. 1 shows the block diagram of the feature analysis and recognition.

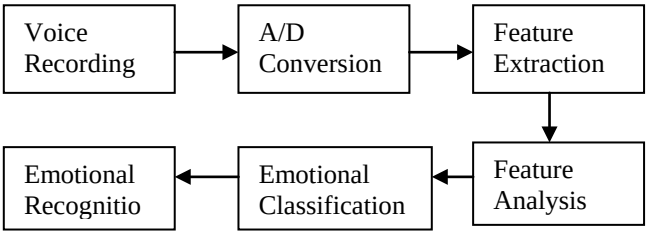

Fig. 1 Block diagram of feature analysis and recognition

Masnani Mohamed is with the FKEE, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia (phone:+60123532854; fax:+6074536060; e-mail: masnani@uthm.edu.my).

Lee Chee Chuan was with the FKEE, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia (e-mail:charles_lee1231@hotmail.com).

Ida Laila Ahmad is with the FKEE, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia (e-mail:ida@uthm.edu.my).

A. Voice Recording

The recording of the voices through microphone were in a quiet room in order to reduce the environment noise. All speakers in different gender uttered the same sentences in different content (angry and non-angry) in order to allow the comparability across emotions and speakers. In order to test the validity of emotion data collected from this experiment, a listening pre-experiment is launched; all speakers are required to tell the emotion type of the sentences and tested by listener.

B. Analog to Digital Conversion

The analog-to-digital converter (ADC) translates this analog wave into digital data that the computer can understand. Voice was undertook A/D conversion into digital signal through personal computer by voice recorder software at 44 kHz sampling rate and saved into Wave file. After that, the entire speech signal in WAV file was analyzed by MATLAB.

C. Feature Extraction and Analysis

Features of the sampled signal such as: pitch, formant frequency, amplitude and duration were extracted individually by using different algorithm. The first stage of analysis process is windowed the speech signal by a hamming window with length of 30 ms and window was shifted to 15 ms so that the overlapping will occur in between windowed [3]. The four selection features are shown in next stages which are pitch, formant frequency, amplitude and duration.

i. Pitch Detection by Autocorrelation Approach Form

Autocorrelation requires that several initial assumptions be made about the set or sequence of speech samples. First, it requires that signal must be stationary and second, it requires that the sequence is zero outside of the current segment. Therefore, by using the windowed signal (stationary and zero outside of current segment), short time autocorrelation has been adopted in order to detect fundamental frequency tone frame by frame.

ii. Formant Detection by Linear Predication Approach

Normally, vocal tract resonances will cause peaks in the observed spectral envelope. Therefore, linear predication analysis was used to find the best matching system by passed through a purely-recursive IIR filter. In order to find formant frequency from the filter, it is necessary to find out the locations of the resonances that make up the filter. This involves treating the filter coefficients as a polynomial and solves the roots of the polynomial. Normally first and second average formant frequency of every frame was used as features parameters [5].

iii. Amplitude

The short-term speech amplitude has been exploited for emotion recognition, because it was related to the arousal level of emotions. The raw data that plotted by time domain was used to analyze amplitude of speech signal. The amplitude shown in time domain graph is energy of the speech signal since the parameter needed to recognize emotion is intensity. Therefore, the energy was converted into intensity. After the intensity of each windowed signal has been calculated, the parameter such as mean intensity of each sentence was taken as features parameter. The reason of calculating the intensity of each windowed signal is to identify the unvoiced signal which may decrease the accuracy of intensity.

iv. Duration

Total uttered duration and total pause duration between speech segments of every emotion sentence has been obtained. By analyzing short term windowed signal, the pause duration between speech segments was calculated. Several assumptions have been considered in order to calculate pause duration such as: low energy, high frequency and low autocorrelation peak [6]. In recognition, the ratio of total uttered duration divided by pause duration of selected sentences was taken as feature parameter.

D. Classification and Recognition

After extracting the desired features, the emotional classification technique was needed to train and test the data. After that, emotions have been recognized. In this project, Artificial Neural Network (ANN) was used to classify each parameter features including mean, maximum, minimum and standard deviation from fundamental frequency, mean amplitude, average sentences duration and first formant frequency and recognize the emotion.

E. Heartbeat Detector Circuit

For the Heart Beat Detector circuit, Light Dependent Resistor (LDR) and infrared sensor were used to detect the heartbeat at fingertips. Basically an infrared light will be emitted into skin. The light absorbed is depending on the blood volume. Backscattered light correspond with the variation of blood volume. The change in blood volume can be determined by measuring the reflected light and using the properties of tissue and blood. The infrared (transmitter) will pick up this signal and then sends an electromagnetic signal containing heart rate data to the LDR (receiver). Fig.2 shows the Heart Beat Detector Circuit simulated by using Proteus.

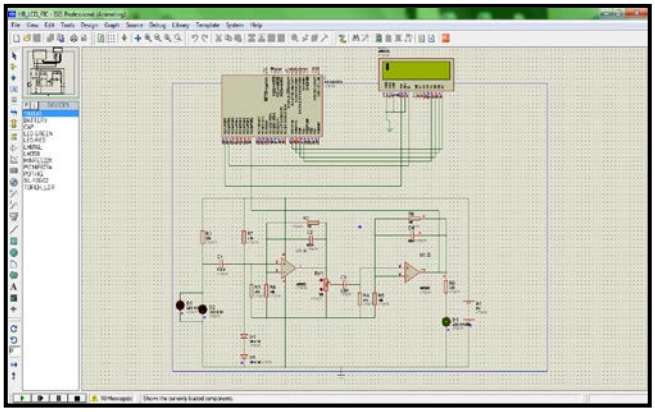

Fig. 2 Heart Beat Detector simulated in Proteus

III. Result and Analysis

For the result and analysis, all the feature parameters of speech signal have been distributed as database of the project. Subsequently, by using Neural Network Pattern Tool from MATLAB, the relationship between input parameter and output emotions were obtained by classified input into set of target categories.

A. Features Parameters Database Distribution

For the acoustic analysis of speech database the features parameters were chosen as representative part of database. By using MATLAB, all the feature extraction method such as autocorrelation and linear predication was applied in order to extract the related parameters features for male and female speakers. Table 1 shows the emotional speech information for 20 speech databases. The speakers were given a sentence to utter with two different emotions which are angry and no angry. Subsequently the recorded speech was analyzed and extracted through MATLAB by obtaining the important features parameters such as mean pitch, maximum pitch, minimum pitch, standard deviation pitch, mean intensity, mean first formant frequency and pause length ratio as in Table 1.

B. Classification using Neural Pattern Recognition Tool

All the features parameters obtained in Table 1 have been used as input databases while emotions were used as targets. A two layers feed-forward network with sigmoid hidden and output neuron was classified arbitrarily well with given enough neurons in its hidden layer. In this case, the number of hidden neuron was set at 20, which is considered as desired number of hidden neuron.

During the classification process, input database was divided into three kinds of samples, which are training sample, validation sample and testing sample. Training samples are presented to the network during training, and the network is adjusted according to its error. While validation samples were used to measure network generalization, and to halt training when generalization stops improving. At the meantime, testing sample provide an independent measure of network performance during and after training due to there have no effect on training. Subsequently, 70% of input data was used as training samples, 15% of input data is for validation samples and 15% input data for testing samples. This percentage considered as desired for classification. After that, network was trained until the satisfied result of all confusion matrixes has been obtained.

Performance of classification by using confusion matrix was divided into three main parts which are training confusion matrix, validation confusion and test confusion matrix, while all confusion matrixes was used to evaluate the overall performance. Inside the confusion matrix, output class and target class played the role in performance of classification. Target class is the emotion output that has been inserted according to input data. Meanwhile the output class is the result of emotional classification.

TABLE 1
EMOTIONAL SPEECH INFORMATION

G	Mean Fo (Hz)	Max Fo (Hz)	Min Fo (Hz)	SD Fo (Hz)	Mean Intensity (dB)	Mean F1 (Hz)	P/L	E
M	136	154	119	26.5	83.8	361	0.12	N
M	134	160	123	37.0	85.3	332	0.17	N
M	164	251	133	49.5	83.5	457	0.54	A
M	244	398	145	34.6	85.7	425	0.13	A
M	212	372	153	29.3	82.6	360	0.09	A
M	205	286	132	38.7	80.0	418	0.27	A
M	154	176	128	15.3	71.6	440	0.40	N
M	117	122	109	2.31	73.2	268	0.13	N
M	213	302	163	25.8	76.5	423	0.12	A
M	111	127	104	15.2	78.8	415	0.58	N
F	228	302	208	15.2	78.4	330	0.16	N
F	257	301	232	27.9	75.1	448	0.22	N
F	350	432	258	34.1	73.2	500	0.16	A
F	241	274	227	13.1	79.4	400	0.21	N
F	259	302	228	35.6	75.4	464	0.29	N
F	308	483	276	67.5	73.1	593	0.17	A
F	264	302	167	58.8	75.5	389	0.17	N
F	403	456	278	29.0	79.1	485	0.30	A
F	201	254	188	18.7	71.8	395	0.33	N
F	336	453	289	21.1	74.5	496	0.11	A

Explanation of Symbols: Max: Maximum, Min: Minimum, F0: Fundamental Frequency, SD: Standard Deviation, F1: First Formant, P: Pause, L: Length, E: Emotions, A: Angry, N: Non-angry.

Fig. 3 shows the confusion matrix of male speakers and Fig. 4 shows the confusion matrix of female speakers. Inside both confusion matrix figures, the number 1 represented angry emotion and number 2 represented no angry emotion, (shown at horizontal and vertical axis of each matrix boxes).

Fig. 3 shows the performance confusion matrix of male speakers, which gives details of the strengths and weaknesses of this system. All confusion matrixes show the results for training, validation and testing experiment which are up to 86.5%. Means that the input parameter obtained have very good relationship with the emotions. Eventually this result was used as reference by recognize emotion in real time or loaded existing emotional speech signal.

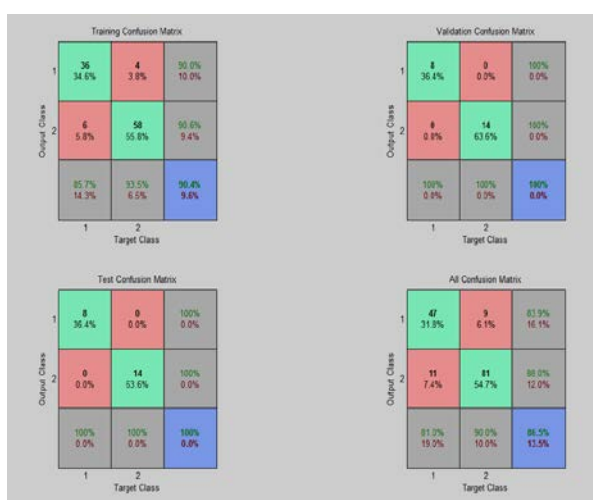

Fig. 3 Confusion matrix of male speakers

Fig. 4 shows the performance of confusion matrix of female speaker. All confusion matrixes show 98.1% of performance classification. This can be concluded that the input parameter have very good relationship between output emotions.

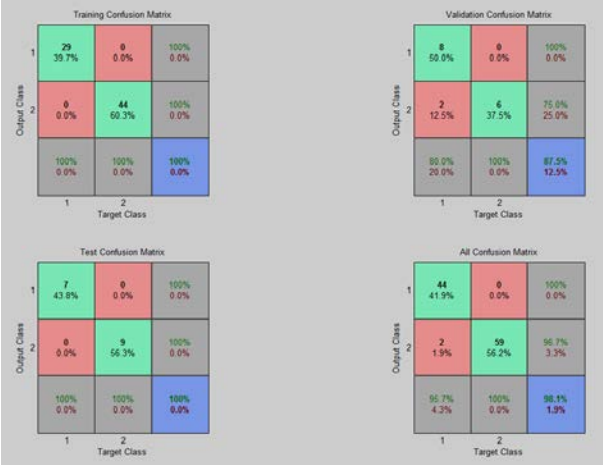

Fig. 4 Confusion matrix of female speakers

C. Emotional Recognition by using MATLAB GUI

MATLAB GUI has been build (Fig. 5) in order to recognize the emotional speech signal. By applied the feature extraction methods which are short time autocorrelation, linear predication coding and time domain presentation, all the features parameters of speech signal have been displayed in MATLAB GUI. After that, result of classification for Neural Pattern Recognition Tool was employed by recognizing the emotion from speech signal.

GUI has several functions, the graph axes were used to display the waveform of speech signal, for the data input section, there are 3 buttons to be chosen. First is “Female” button, which is used to load the existing female speech signal. Second is “Male” button, used to load the existing male speech signal and last is “Record” button, which is used to record the sound of speaker in real time. At the meantime, the feature parameter section shows the output features parameters of speech signal which are mean pitch, standard deviation pitch, maximum pitch, minimum pitch, mean intensity, mean first formant frequency and pause length ratio. For the last section is emotion recognition, which is used to display the emotions of speech signal.

In order to test the emotional speech, first, press the desired button in order to choose speech signal. Once the desired speech file has been chosen and loaded, the waveform of speech signal was shown in the graph axes. Subsequently the features parameters of the speech signal were displayed at the feature parameter output. Finally the emotion was detected and displayed at emotion recognition section.

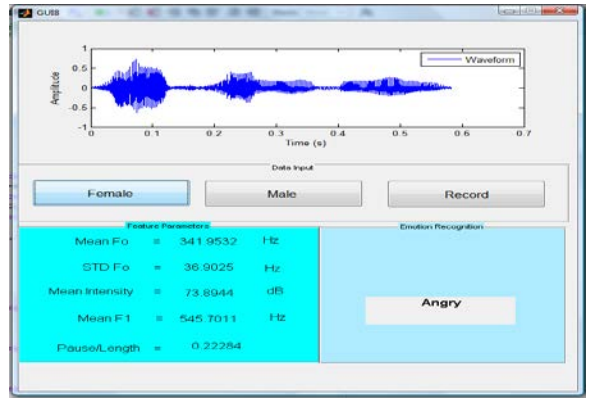

Fig. 5 MATLAB GUI for speech recognition

D. Result of Emotion Recognition

The results of emotion speech recognition has been tested by using existing file and in real time. The “real emotion” meaning the actual emotion of speech signal, whereas the “recognized emotion” is the emotion that has been tested by the system. The “result” shows the outcome of each speech signal that has been tested, if the actual emotion of speech signal match the recognized emotion, then the result will shows “correct”, otherwise it will show “wrong”.

The result of emotion recognition by loading existing file has been achieved 80% of accuracy. However, emotion recognition for real time is only 70% of accuracy.

E. Heartbeat

Table 2 shows the heartbeat taken whenever the speakers uttered the given sentence with two different emotions which are angry and no angry. It shows that when a person is angry, the heartbeat goes higher than normal. In this case, 90 bpm and above are considered as high.

TABLE 2
HEARTBEAT DETECTION

Subjects	Heartbeat (bpm)	Emotion (ANGRY/NORMAL)
1.	89	NORMAL
2.	86	NORMAL
3.	90	ANGRY
4.	90	ANGRY
5.	97	ANGRY
6.	90	ANGRY
7.	84	NORMAL
8.	72	NORMAL
9.	92	ANGRY
10.	88	NORMAL
11.	70	NORMAL
12.	73	NORMAL
13.	91	ANGRY
14.	88	NORMAL
15.	84	NORMAL
16.	98	ANGRY
17.	78	NORMAL
18.	102	ANGRY
19.	88	NORMAL
20.	96	ANGRY

IV. Discussion and Conclusion

A. Discussion

The primary goal of this paper is to extract, analyze, classify and recognize the angry speech signal. Feature extraction and analysis play a most crucial role [4]. In this experiment, feature parameters has been extracted successfully by using different method of digital signal processing, such as, windowing, framing, autocorrelation and linear predication coding.

Classification also shows an outstanding result by employed the neural pattern recognition tool which has been achieved 86.5% for male classification and 98.1 % for female classification. The purpose of separated female and male classification is because, male and female has significant difference in their vocal size means that they have differently pitched voices. Moreover, men generally speaking, have a larger vocal tract, which essentially give the resultant voice higher compared to women. When a person in anger state he will speak faster with bigger volume and raised the tone due to produced changes in respiration and an increase in muscle tension, which influence the vibration of the vocal folds and vocal tract shape, affecting the acoustic characteristics of the speech.

B. Conclusion

In this paper, all the features parameters of speech signal from male and female has been obtained by using autocorrelation and linear predication method in order to obtain mean fundamental frequency, standard deviation fundamental frequency, maximum fundamental frequency, minimum fundamental frequency, mean intensity, first formant frequency and pause length ratio.

Among these features parameters, mean fundamental frequency, standard deviation fundamental frequency and mean intensity show a good result in order to differentiate between angry emotion and no angry emotion. This is because, during high degree emotion (angry), speech was uttered faster with bigger volume and raised the tone due to produced changes in respiration and an increase in muscle tension, which increase the vibration of the vocal folds, and variation of fundamental frequency also increases. Therefore, mean pitch, standard deviation pitch and mean intensity was much higher. However, during in low degree emotion (neutral and sadness), there were no much alteration in the fundamental frequency and the value of fundamental frequency was very low as well as the intensity was much lower compared to high degree emotion.

In addition to the speech analysis, the heartbeat measurement is done to check the level changes while people having different emotion. This is sensed by using a high intensity type Light Emitting Diode (LED0 and LDR. The finger is placed between the LED and LDR. The skin may be illuminated with visible light (red) using transmitted or reflected light for detection. The very small changes in reflectivity or in transmittance caused by the varying blood content of human tissue are almost invisible. Its characteristic is related to human emotion and changes the body condition directly. From the experiment result, it is proved that human heart will beat faster than normal when they are angry.

Acknowledgment

This work has been supported financially by the Office for Research, Innovation, Commercialization and Consultancy Management (ORICC), Universiti Tun Hussein Onn Malaysia and some facilities are provided by Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia.

References

[1] Wang Yu, "Research and Implementation of Emotional Feature Classification and Recognition in Speech Signal," *IITAW'08*, 2008, pp. 471-474.

[2] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor, "Emotion Recognition in Human-Computer Interaction," *IEEE Signal Process. Mag*, vol.18, Jan 2001, pp. 32-80.

[3] Deller, J. R, Hansen, J. H. L. and Proakis, J., *Discrete-Time Processing of Speech Signals.* Wiley-IEEE pres. 2000.

[4] M. Morales-Perez, J. Echeverry-Correa, A. Orozco-Gutierrez and G. Castellanos-Domingue, "Feature Extraction of Speech Signal in Emotion Identification" *EMBS*, 2008, pp.20-24.

[5] Roy C, Snell and Fausto Milinezz, "Formant Location from LPC Analysis Data," *IEEE Transaction on Speech and Audio Processing*, 1993, pp. 129-134.

[6] Keikichi Hirose, Hiroya Fujisaki and Shigenobu Seo, "A Scheme for Pitch Extraction of Speech Using Autocorrelation Function with Length Proportional to the Time Lag," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1992, pp. 149-152.

[7] Takahiro Sunaga and HirooIkehira, "Measurement of the Electrical Properties of Human Skin and the Variation among Subjects with Certain Skin Conditions," *Physics in Medicine and Biology*, vol. 47, no. 1, 2002.

[8] Alberto de Santos Sierra, Carmen Sánchez Ávila, Javier Guerra Casanova, and Gonzalo Bailador del Pozo, "A Stress-Detection System Based on Physiological Signals and Fuzzy Logic," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 10, 2011, pp. 4857–4865.

Trust Based Intrusion Detection System for Mobile Ad Hoc Network

M.B. Mukesh Krishnan[1] and Prof. Dr. P. Sheik Abdul Khader[2]

Abstract – Providing security is a complicated process due to the nature of the network. Restricting the anomaly access becomes the big issue. Most of the anomaly access takes place because of the lack of trust among the nodes. This paper is concerned towards providing trust based intrusion detection system for mobile ad hoc network (MANET). Traditional intrusion detection system detects the intruders through the network based agent and host based agent. The agent process update them self through the adjacent nodes in the range. The issue is the lack of trust among the nodes. This paper proposes a trust model to enable the trust among the nodes in the environment is verified before the intruder is detected. The effectiveness of the system is evaluated by comparing with the Standalone Intrusion Detection System, Distributed and Co-Operative Intrusion Detection System and Hierarchical Intrusion Detection System and the result shows that lack of trust leads to higher number of attacks as well as the false detection percentage is high.

Keywords-- MANET, Trust, IDS.

I. INTRODUCTION

A Mobile ad hoc network is a self created, self organized and self managed network without basic infrastructure. In this type of environment the trust level among the nodes are low. In such an environment, it may be necessary for one mobile host need to establish trust among the adjacent nodes within the cluster.

Security in MANET is an essential component for which the restriction of unauthenticated node access within the cluster should be restricted so the need of the intrusion detection system arises to provide the effective functionality of the intrusion detection system the trust among the nodes should be established become the necessary. To obtain that Trust Based Intrusion Detection System for Mobile Ad Hoc Network was proposed.

In this paper we propose Based Intrusion Detection System for Mobile Ad Hoc Network which ensures the trust among the nodes to obtain reliable data for the intrusion detection. The remaining of this paper is organized as follows: Section 2 discusses on trust formation between the adjacent nodes in MANET. Section 3 formalizes the trust based intrusion detection model for MANET In Section 4, discusses the experimental set up to evaluate the model through simulation and the results are presented , and We finally conclude the paper in Section 5.

M.B. Mukesh Krishnan is Research Scholar , Sathyabama University, Chennai 600 119, India. Email: mukesh_krishnan@yahoo.com

Prof. Dr. P. Sheik Abdul Khader is Professor and Head Department of Computer Application, B.S. Abdur Rahman University, Vandalur, Chennai-600 048,India. Email: psakhaer@yahoo.com

II. TRUST FORMATION BETWEEN THE ADJACENT NODES IN MANET

The formation is a complex task in environments like MANET. To formulate the trust among the nodes it is necessary to identify the malicious, compromised and selfish nodes [1]. The malicious nodes create the malicious activity which also considered as the security threat and will not provide trusted information. The malicious node is the malicious activity can be detected via the intrusion detection system. The intrusion detection system detects the malicious node through the analysis of the node behavior in the environment such as, if the node collects the network information constantly as well as frequently changes the state of the node and influence the other node to change the state. All such nodes are marked as a malicious node and further analysis the traffic flow via that node in such a way whether the node transmit the data as per the routing table. The result of the two analyses will detect whether node is a malicious node or not.

Compromised and selfish nodes cause the trust unattainable because they will not provide the true facts, it leads to the security threat and will not provide the trusted information. The detection of the compromised and selfish nodes is a complex process [2][4][5]. The compromised node can be detected via the intrusion detection through the misbehavior of the node. The selfish node will not create any maliciousness but will not provide any information or will not forward to save power that can be analysis by sending a junk packet to forward if a node keeps on discarding it can be identified as a selfish node [3].

Once the malicious, compromised and selfish nodes the other nodes can be assumed as trusted nodes to ensure the trust among the nodes a trust analysis engine is developed and deployed on cluster head node. The trust main objective of the trust analysis engine is to provide trust among the adjacent nodes. The trust analysis engine keeps on updating the node transaction and routing table information in the network. If the abnormality noticed the node will be isolated from the cluster. Thus the trust formation between the adjacent nodes in MANET is formulated via trust analysis engine.

III. Trust Based Intrusion Detection System for Mobile Ad Hoc Network

The trust based intrusion detection system for MANET is constructed with two stages. In the first stage the intrusion is detected and in the second stage trust among the nodes is evaluated. Through this two stages of evaluation the malicious, compromised and selfish nodes are identified and isolated from the network.