# Malicious Node Detection using Route Prediction based on HMM

**Manoj Kumar G, M. Abdul Rahiman**

***Abstract*****: *Driving route prediction methods based on Hidden Markov Model accurately predicts a vehicle's entire route throughout a trip. Trip history of driver alone cannot be used for predicting the route. Routine history of routes can be modelled and learned for predicting purposes. Driver behavior, another factor of route prediction can be considered as another factor of route prediction. Route recommendation mechanism helps to identify the probability of mobility of vehicles over time. This method can be extended to identify malicious nodes within network traffic. First we define a road network model, the driving routes in a hexagonal coordinate system, build HMM models to predict the movement using a method of training set based on K-means++ technique. The route predicted is taken as input and transmitted along with network data using encrypted headers. A method to identify malicious nodes in VANETs using HMM of prediction about routes helps to identify malicious message from a compromised node. One method of identifying suspicious message is the signal strength which is incompatible with its originator's geographical position. We provide encrypted headers in protocols for detecting suspicious transmissions. Identified malicious node information is disseminated in the network. Evaluation of the detection rate and the efficiency of solution is analyzed using cryptographic methods based on cloud computing. This helps to identify the malicious nodes in the network traffic.***

***Keywords: Hidden Markov Model, Route prediction, malicious nodes.***

## I. INTRODUCTION

The Google map, Baidu Map, etc. uses path algorithms and historic traffic data to give vehicle route recommendation for drivers. The driver could select one of the best path based on personal preference, congestion, path distance, etc. From the above information it is clear that less congested and shortest paths will be preferred compared to other criteria [3]. Live data traffic is used usually to predict the congestion in a route. Understanding these criteria, intruders will create duplicate or spam packets and transmit by choosing the less traffic areas. This makes an illusion that the most preferred areas will be heavily traffic. Thus this sort of attacks can be considered as denial of service attack. To overcome this sort of dos attacks we use encrypted packets [2] along with the existing application layer to check the consistency of the data received. This helps to improve the prediction and to build [4] up a smooth data transfer among vehicles.

This paper is organized as follows. Section 2 describes the architecture of route recommendation and malicious node detection system. Section 3 introduces the construction of a road network model and a network traffic model. The process of building HMM and the method of making route predictions, encrypted communication has been discussed in section 3. In section 4 the experimental results are discussed. Finally section 5 concludes the paper.

## II. ARCHITECTURE OF ROUTE RECOMMENDATION AND MALICIOUS NODE DETECTION SYSTEM

The architecture of route recommendation and malicious node detection system consist of

- **Prediction of route using HMM**: This module predicts the routes to a driver based on HMM. It uses an existing dataset [6] and predicts the extended route data which is important for pre-estimating traffic congestion. Actual taxi data was derived from the existing dataset using the call connections/requests from dataset. Route model was initially generated from existing dataset.
- **Congestion pre-estimation**: This module predicts the congestion of each road. The congestion level of road segment $R_i$ is denoted by driving routes within a time period. Higher values indicate, higher congestion levels.
- **Route recommendation**: The route recommendation system collects the information about just driven road segments, traffic congestion, data parameters, and mobility parameters to analyze better routes. It uses less traffic bi routes also for extended route recommendation system.
- **Correction of routes (HMM)**: The HMM model corrects the routes based on new input driving routes. If more driving routes are considered as input for HMM, the prediction accuracy improves.

## III. CONSTRUCTION OF A ROAD NETWORK MODEL AND A NETWORK TRAFFIC MODEL

The road network model is represented as a set of line segments ($R_i$) as shown in Fig-1. It uses the coordinate of two end points of a line segment say (Xi1, Yi1) and (Xi1, Yi2). The network traffic model is represented as a hexagonal cell, (RNi,1 , RNi,3) , (RNi+1,0 , RNi+1,4) , (RNi+2,1 , RNi+2,3). The algorithm for creating a new route from existing route was derived from [7].

**Revised Manuscript Received on January 15, 2020**

**Manoj Kumar G**, Research Scholar, Department of CSE, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India. Email: manojkumarg@ieee.org

**M Abdul Rahiman**, Research Guide, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India. Email: rehmanpaika@gmail.com

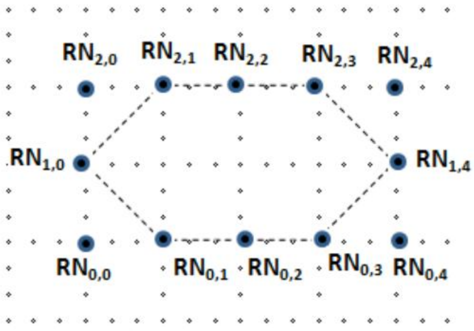

**Fig-1: Representation of a road segment.**

**Algorithm – 1:** Create new route from existing route.

**Input:** A training set, T

**Output:** Extended training set with new routes, $n_{rs}$

```
(1) Coordinate Point set, CP ($P_1$, $P_2$) = Ø;
(2) Extended route set, $n_{rs}$ = Ø;
(3) for each (route, $r_i$ in T)
(4) Initial point, A= $R_{i,j}$ ;
(5) End point, B= $R_{j,k}$ ;
(6) Insert A and B into set CP;
(7) CP'= Filter(CP);
(8) Cluster Set, $C_S$ =K-means++ (CP');
(9) for(int i=1, j=i+1, k=0;( i<n) && (j<n) && (k<CP(i.length)); i++, j++, k++)
(10) for (int l=0 ; l<$C_S$[j] .length; l++)
(11) Insert New_route($C_S$[i][k],$C_S$[j][l] ) into $n_{rs}$;
```

The algorithm for hidden state determination was derived from [7].

**Algorithm – 2:** Hidden state determination.

**Input:** A training set T

**Output:** A hidden state sequence set HS.

```
(1) Hidden state sequence set {HS} = Ø;
(2) for(int i=1; i<m; i++)
(3) Starting point $A_i$ = $R_{i,j}$;
(4) End point $B_i$= $R_{j,k}$;
(5) Vector, $V_i$= B-A;
(6) for (int j=i+1; j<m ; j ++)
(7) Starting point $A_i$=$R_{j,1}$ ;
(8) End point $B_j$=$R_{j,n}$ ;
(9) Vector, $V_j$= B-A;
(10) if (0 <= cos($V_i$,$V_j$) <= 1)
(11) for each ($CP_i$ in $t_i$ )
(12) for each ($CP_j$ in $t_j$)
(13) If ($CP_i$ ==$CP_j$)
(14) Insert $t_j$ into HS;
(15) else
(16) for each ($CP_j$ in $t_i$)
(17) If ( $CP_j$ == '<' or '>' )
(18) Insert symbol X into HS corresponding to A and B;
(19) else
(20) Insert symbol $t_i$ into HS corresponding to each $CP_i$;
(21) Insert {$HS_1$, $HS_2$... $HS_n$} into the sequence set HS.
```

**Algorithm – 3**: Encryption of current route set, message.

**Input:** Current route set, Predicted route set and message

**Output:** Encrypted data set to neighbor node

```
(1) Message, M={$RS_1$, $DP_1$}
(2) Upcoming Route set $U_{RS}$ = {$U_{RS1}$, $U_{RS2}$…$U_{RSn}$}
(3) for each data, $D_i$ from $V_i$
(4) Encrypt, E($D_i$) = {$U_{RSi}$ || M}
```

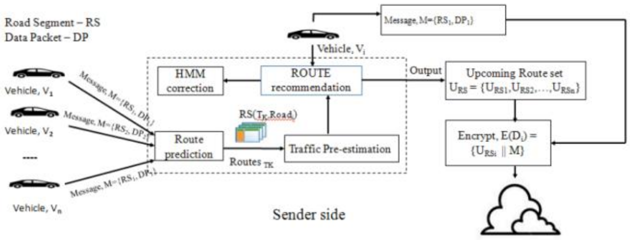

**Fig – 2: HMM estimation and Data transfer**

## IV. EVALUATION OF ROUTE PREDICTION AND DATA TRANSFER

### Experimental Results

Our focus is to track the path, record the path of driving, encrypt the message along with HMM on data set collected and pass the encrypted data to the intended users. We used ns3.26 simulator and SUMO to track the routes for prediction purpose. A dataset of taxi communication from CRAWDAD was taken for analysis [6]. A total of 56 paths where analyzed. Suppose a vehicle has passed through $R_i$ roads then the possible route set R after predicting based on HMM is R={ $R_1$,$R_2$,$R_3$…$R_n$}. The prediction accuracy,

$$P_i = \frac{\sum_{k=1}^{n} D(R_k, U_R)}{\sum_{t=1}^{n} Len \,|\, R_t \,|} *100\%$$

where $U_R$ ε $n_{rs}$(number of route segments in upcoming route) ; duplicate route sets are identified in the numerator part and length of distance for a trip is used in denominator. When the vehicle has passed through a point the prediction accuracy is as follows. The probability that a packet correctly forwarded, P1 is calculated as follows:

$P_1$=(Repeat($R_i$,$R_j$)+Repeat($R_{i+2}$,$R_j$))/(Dist|$R_i$|+Dist|R $_{i+2}$|)*100%

$P(s_i \,|s_{i-1})$ = (Count($S_{D_{i-1}}$,$S_D$) + δCount ($S_{nrs}D_{i-1}$, $S_{nrs}Di$) + 1 ) /

(Count ($S_D$ ) + δCount ($S_{nrs}$) + $m$)

where 0<δ<1. The weight δ is application and network dependent. The value of forwarding probability will be calculated by the source node and compared to the current behavior of each node. This helps to compute the trust level of nodes. The simulation parameters which include parameters from SUMO and ns3 used are represented in Table-1.

**Table-1: Simulation Parameters**

| | |
|---|---|
| Simulator | ns3.26, SUMO |
| Simulation time | 3000 seconds |
| Number of nodes | 150 |
| Number of malicious nodes | 10-15% |
| Network size | 1500m * 1500m |
| Transmission range | 200m |
| Max speed | 2.5m/s – 10m/s |
| Mobility model | Random way point |
| Traffic type | CBR,VBR |
| Number of source/destination pairs | 45% |
| Encryption | RC4(56 bit key) |
| Channel bandwidth | 1 Mbps |

The route prediction method predicts the possible and suitable paths which helps to identify data from neighboring nodes packet timestamps. If time stamps match the encrypted data is verified. From this we can identify the malicious or compromised node data. Such identified malicious nodes are discarded.

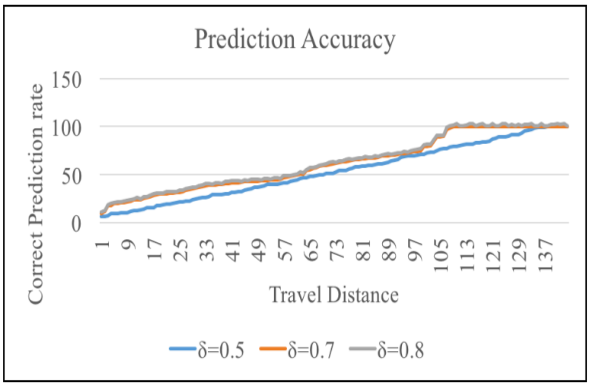

**Fig-3: Prediction Accuracy**

This may in turn reduce the entire throughput and packet delivery ratio of the network for a short instance of time. Once identified the malicious node activity can be discarded which increases the throughput. This directly increase the network efficiency for node data transfer especially in streaming activities. Initial probability distribution considers distance, waiting time for each path based on traffic update through IoT, re-estimation of cost for each path. The equation for initial probability distribution is given below.

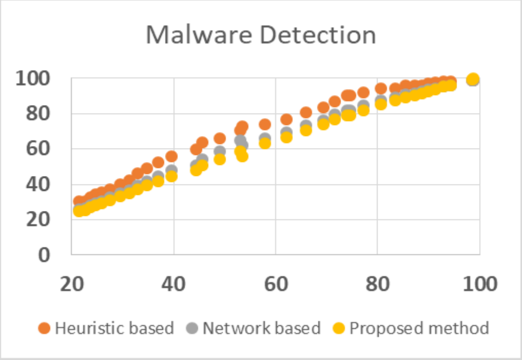

**Fig-4: Malicious Node detection**

Prediction accuracy is plotted from the dataset with the actual distance travelled against the modified HMM for correct prediction rate. The prediction accuracy calculated is plotted in Fig-5.

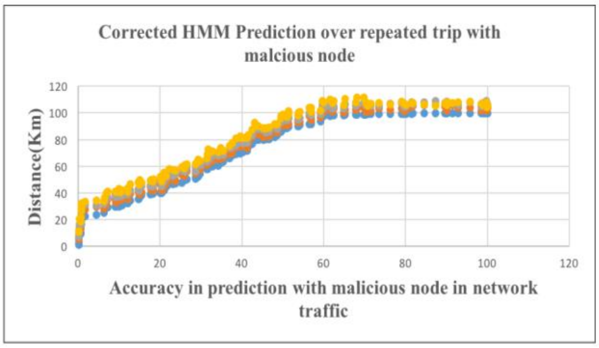

**Fig-5: Corrected HMM prediction**

## V. CONCLUSION

Packet loss occurs in MANET due to several reasons such as interference in links, congestion in node level, overflow of queue, node mobility. Packet loss in MANET can be identified using network parameters, whereas denial of service needs more sophisticated methods. Enhanced route prediction which include bi routes, traffic density, updates from social media or social activities help to increase the identification of malicious activities of nodes, service quality etc. Parameters that could be useful for evaluating the Denial of Service attacks by making use of protocol level steganography helps to authenticate and improve the QoS.

$$\prod_i = \frac{C(D_{hs}i) + \lambda C(D_s i)}{\sum_{j=1}^{n}[C(D_{hs}i) + \lambda C(D_s i)]}$$

Where C ($D_{hs}$i) and C(D$_s$i) represent the number of times the hidden state $hsi$ and state $S_i$ appears in the given and extending data sets, and $\lambda$ represents the weight.

The performance and accuracy in prediction of route with malicious nodes in network traffic analyzed is given in Fig-4. Here the x-axis represents the percentage of malware affected node detected and y-axis represents the percentage of trip completed.

## REFERENCES

1. A. Hamilton, B. Waterson, T. Cherrett, A. Robinson, and I. Snell, “The evolution of urban traffic control: changing policy and technology”, Transportation Planning and Technology,vol. 36, no. 1, pp. 24-43, 2013.
2. A. Karbassi and M. Barth, “Vehicle route prediction and time of arrival estimation techniques for improved transportation system anagement”, in Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 511-516, IEEE, Columbus, Ohio, USA, 2003.
3. J. Krumm, “A markov model for driver turn prediction”, SAE SP 2193(1), 2008.
4. J. Froehlich and J. Krumm, “Route prediction from trip observations”, SAE SP 2193:53, SAE, 2008.
5. R. Simmons, B. Browning, Y. Zhang, and V. Sadekar, “Learning to predict driver route and destination intent”, in Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC '06), pp. 127-132, IEEE, September 2006.

6. Taxi Dataset: {Lorenzo Bracciale, Marco Bonola, Pierpaolo Loreti, Giuseppe Bianchi, Raul Amici, Antonello Rabuffi, CRAWDAD dataset roma/taxi (v. 2014-07-17), downloaded from Https://crawdad.org/roma/taxi/20140717, https://doi.org/10.15783/C7QC7M, Jul 2014.}
7. Ning Ye, Zhong-qin Wang, Reza Malekian, Qiaomin Lin, and Ru-chuan Wang, “A Method for Driving Route Predictions Based on Hidden Markov Model”, Mathematical Problems in Engineering, vol. 2015, Article ID 824532, 12 pages, 2015. doi:10.1155/2015/824532

## AUTHORS PROFILE

**Manoj Kumar G** received his B.Tech Degree in Computer Science & Engineering from Kerala University and M.Tech degree in Computer Science & Engineering from Anna University. Currently, he is working as an Associate Professor in the Department of Computer Science & Engineering, LBS Institute of Technology for Women, Thiruvananthapuram affiliated to APJ Abdul Kalam Technological University. He is currently a research scholar at Karpagam Academy of Higher Education, Coimbatore. His areas of interest are Mobile Computing, Cryptography, Object Oriented Systems and Databases.

Prof. (Dr.) M. Abdul Rahiman is the Managing Director of Kerala State C-apt. He received the Doctor of Philosophy (Ph.D.) degree in Computer Science & Engineering from Karpagam University. He obtained his Master of Technology from Kerala University in 2004, and Bachelor of Technology from Calicut University. He achieved Post Graduate Diploma in Human Resource Management from Kerala University & Master of Business Administration. He is an eminent academician and an able administrator. He was the founder Pro Vice Chancellor of APJ Abdul Kalam Technological University and also served as Director, AICTE, Ministry of HRD, Govt. of India. He was also appointed as Director Vocational Higher Secondary Education to the Government of Kerala. He has also served as a Faculty of Engineering at LBS Institute of Technology for Women, Trivandrum. He specializes in Digital Image Processing & Pattern Recognition and he taught for more than 10 years having a rich teaching experience and current research areas are Image and Computer Vision, Data Mining and Networking. He is also serving as Member of many professional & technical bodies; chaired many Technical Conferences. Also serving in the Editorial board of many International Journals. He was also a Member of Advisory body of Technical Education UT of Daman Diu, which guides the Technical & Higher Education area.