Using trust model to ensure reliable data acquisition in VANETs

Emre Kağan Akkaya

Outline

- Problem
- Related Work
- Assumptions
- Entity-centric model
    - Algorithm
    - Results
- Data-centric model
    - Algorithm
    - Results
- Conclusion

Problem

- The open, dynamic and distributed natures make VANETs face many security-related challenges (dishonest forwarding, false message propagation etc.)
    - Authenticity and reliability of the data exchanged are of great importance!
- Solution: A **trust model** can be used to evaluate the trustworthiness of other nodes thus detect malicious nodes.
    - It must ensure the reliability of the data in **real time** and must provide the **trustworthiness** of a vehicle.

Related Work

- Three kinds of trust model in VANETs:
    - **Entity-centric**: focuses on evaluating the trustworthiness of vehicles to detect the dishonest and malicious nodes and ensure the reliable data delivery.
        - Basic of data trust. Reliable data can enhance the entity trust in turn.
        - Fundamental measure to provide secure routing!
        - Sub-categorized into **direct** and **recommend** trust.
    - **Data-centric/Event-centric**: focuses on evaluating the trustworthiness of the data reported by other vehicles to ensure the applications work securely and effectively.
        - Dynamic and volatile in nature. Time and location closeness, # of reports on the same event and the types of the events are all taken into consideration.
    - **Combined**: makes extensive use of entity trust to evaluate the trustworthiness of data and maintains entity trust over time.

Related Work

- **Huang et al.** compute trust value based on a voting scheme which a static weight is applied to intermediate nodes to prevent oversampling and information cascading.
- **Finnson et al.** proposed multi-faceted (user’s role, location, time, priority, inherent trustworthiness calc. from past interactions, direct/indirect reporting) trust model with majority consensus.
- **Raya et al.** proposed a framework based on collecting multiple reports related to the same data and combine them with their weights to make a decision.
- **Wu et al.** put forward a RSU-aided model by synthesizing the distance from the vehicle to the event, maximum detection range of the vehicle, the number of sensors in the vehicle, and the weight of vehicle to determine the data trustworthiness.
- **Ding et al.** presents an event-based reputation model to filter bogus warning messages by classifying nodes into different roles. Each role has its own trust evaluation mechanism for the incoming traffic message.
- In addition *fuzzy logic, probability or Bayesian inference* are often used which are all based on the previous interaction information.

Related Work

- Common shortages:
  - Entity-centric models are usually based on the past direct interactions but may fail to collect enough information about the neighbor or sender.
  - Direct trust is usually updated periodically but timely updating is more crucial.
  - A static balance coefficient is often used to leverage the proportion of direct and recommend trust but a dynamic coefficient which is variable with the direct trust and context is more beneficial.
  - Take them much time to make a data-centric trust decision. Decisions must be made in real time without any significant delay..
  - Researches on combined trust model for VANETs is relative few.

Related Work

- Trust metrics used are usually common in many existing models:
  - Distance (vehicle-event, receiver-sender, sender-RSU, RSU-event)
  - Time closeness
  - Recommendation from other nodes, # of senders, node experience on data/entity, event/node type etc.
- Focus on following 3 facts:
  - Different data has different impact on traffic
  - Different type of vehicles plays different role and has different authority in traffic
  - The majority of people drive their vehicles locally for their daily commute (most vehicles have predefined constant daily trajectories)

Assumptions

1. **Network model**
   a. Many people’s travel habits always meet a specific distribution, common or similar travel/activity range. Diff. vehicles can meet with a certain probability and make it possible to collect the past interaction experiences to establish entity trust.
   b. All nodes are equipped with smart sensors, computing modules, wireless comm. module, GPS and other devices needed to form VANET.
      i. The range to perceive a traffic event is **20m** and the comm. Radius is **200m**.
   c. Transportation authority (**TA**) organizations issue public certs. during the node registration and checks nodes and their certs. periodically. The Public key of the TA is known to all vehicles in advance. Also, the **vehicle type is signed** by the TA. (more on this later)

Assumptions

2. Application model

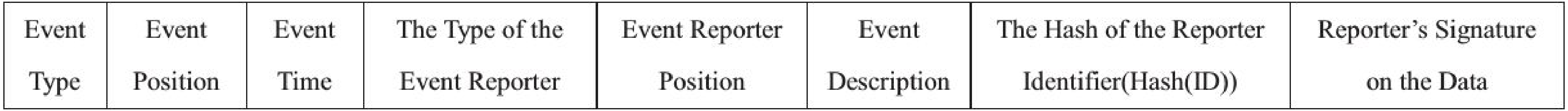
a. Categorize applications into 3:

  i. **Safety application** S: {rear-end accident, serious accident, breakdown, blind spot, icy-road, wet-road, thick foggy, steep slope zone}

  ii. **Efficiency application** E: {congestion, road maintenance, road closed, parking,gas station}

  iii. **Infotainment application** I: {coupon, song, music, scenic spot, restaurant, bar}

b. An event or data description is a subset of these sets. Data format can be seen below (All the items except for “**Event Reporter Type**” are generated automatically by the reporter. Reporter types are previously issued by TA as follows: **Sign(K_M, hash(ID_k)|| T(v_k)**)



Event Type	Event Position	Event Time	The Type of the Event Reporter	Event Reporter Position	Event Description	The Hash of the Reporter Identifier(Hash(ID))	Reporter’s Signature on the Data

Assumptions

3. **Security model**
    a. Entity trustworthiness is relatively slow-evolving and impacted by the past interaction with other nodes.
        i. The direct trust is updated only after the interaction between the two nodes changes.
        ii. Only when the direct trust value less than a threshold, should the recommendation and the comprehensive trust value be calculated.
    b. Data trustworthiness depends on many factors such as the trustworthiness of the reporter, forwarder, time closeness, location closeness and the relations between the data and the reporter as well.
    c. Trust value should be a number in [0,1].
        i. 1 means completely trustworthy, 0.5 means uncertainty and, 0 means untrustworthy.

Entity-centric Model

The trustworthiness of a node is closely related to its authority level and the type of the application data reported by it. So here we define the app. data weight and node weight respectively:

1. **Application data weight**
    a. Different kinds of application data have different impacts on traffic, the public or individual safety.
    b. Application data's weight stands for its importance and any data transmitted should belong to one of the three data types:

$$W_D(x) = \begin{cases} 1, & x = S \\ 0.8, & x = E \\ 0.5, & x = I \end{cases}$$

Entity-centric Model

2. **Node weight**
    a. Categorize various nodes into 3 types according to their authority:
        i. **High level**: police wagon and road side units (RSU)
        ii. **Medium level**: public services such as bus, ambulance, road upkeep vehicles, sanitation trucks etc.
        iii. **Low level**: private car, taxi, freight vehicles which are controlled by individuals
    b. The high level node and the data reported by it are usually with higher trustworthiness than the medium or low level one. A node should belong to one of the three levels.

$$\mathrm{W_N}(x) = \begin{cases} 1, & x = H \\ 0.7, & x = M \\ 0.5, & x = L \end{cases}$$

Entity-centric Model

The comprehensive trust value is co-determined by the direct trust and the recommendation described as the following:

1. **Direct trust**
   a. One node's subjective expectation to the other nodes' future behaviour.
   b. Either two vehicles have history interactive experiences, or they meet first time.
      i. If they have no history, the direct trust value is set to be the **node's weight W_N**.
      ii. Else, the direct trust value is determined by the successful data forwarding rate.
         - But cunning nodes may get high successful forwarding rate by only forwarding the data with low weight (when infotainment data is more than safety and efficiency data…) First, calculate **malicious tendency**!

Entity-centric Model

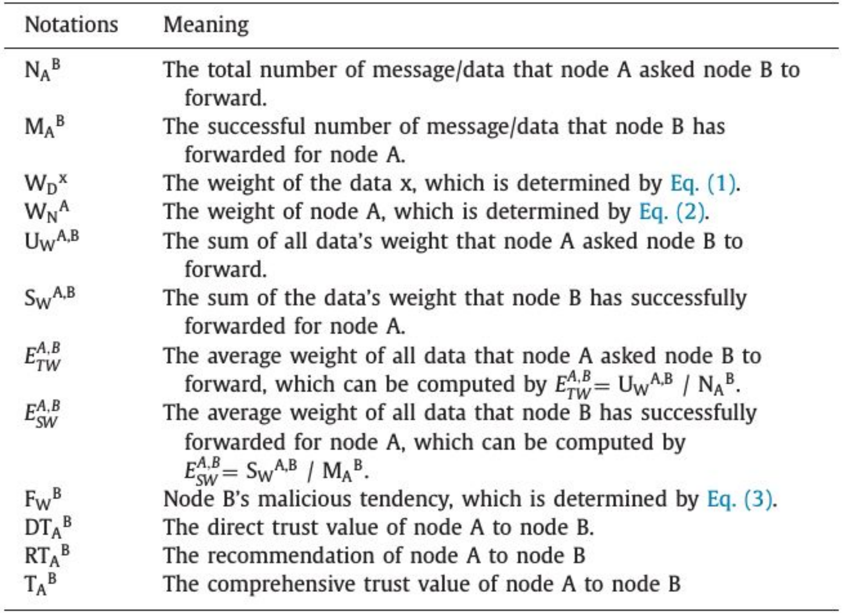
1. **Direct trust (cont’d)**
   a. Malicious tendency of a node is closely related to the average weight of the data failing to be forwarded, node A can evaluate node B’s malicious tendency by $F_W^B = \left(U_W^{A,B} - S_W^{A,B}\right)/\left(N_A^B - M_A^B\right)$
   b. The proportion of the *S, E, I* data are 0.2, 0.4 and 0.4 respectively. The mean weight of the data can be computed by (1x0.2+0.8x0.4+0.5x0.4) = **0.72** which is the threshold for determining whether a node is with malicious tendency or not.
      i. $F_W^B < 0.72$ indicates that the node B has no malicious tendency
      ii. Otherwise, it is considered to be with malicious tendency.

Entity-centric Model

1. Direct trust (cont'd)

a. The node A's direct trust $\mathbf{DT_A^B}$ for node B can be calculated as below (flag is 1 if the msg forward is success, -1 otherwise)

$$DT_A^B = \begin{cases} \dfrac{W_D^x \cdot \left((Flag+1)/2 - DT_A^B\right)}{1 + E_{TW}^{A,B}/E_{SW}^{A,B}} + DT_A^B, \; F_W^B < 0.72 \\ W_D^x \cdot \left((Flag+1)/2 - DT_A^B\right)/4 + DT_A^B, \\ \quad (Flag = 1) and (F_W^B \geq 0.72) \\ W_D^x \cdot \left((Flag+1)/2 - DT_A^B\right) + DT_A^B, \\ \quad (Flag = -1) and (F_W^B \geq 0.72) \end{cases}$$



Notations	Meaning
N_A^B	The total number of message/data that node A asked node B to forward.
M_A^B	The successful number of message/data that node B has forwarded for node A.
W_D^x	The weight of the data x, which is determined by Eq. (1).
W_N^A	The weight of node A, which is determined by Eq. (2).
$U_W^{A,B}$	The sum of all data's weight that node A asked node B to forward.
$S_W^{A,B}$	The sum of the data's weight that node B has successfully forwarded for node A.
$E_{TW}^{A,B}$	The average weight of all data that node A asked node B to forward, which can be computed by $E_{TW}^{A,B} = U_W^{A,B} / N_A^B$.
$E_{SW}^{A,B}$	The average weight of all data that node B has successfully forwarded for node A, which can be computed by $E_{SW}^{A,B} = S_W^{A,B} / M_A^B$.
F_W^B	Node B's malicious tendency, which is determined by Eq. (3).
DT_A^B	The direct trust value of node A to node B.
RT_A^B	The recommendation of node A to node B
T_A^B	The comprehensive trust value of node A to node B

Entity-centric Model

2. **Recommendation trust**
   a. The recommendation is used to avoid the subjectivity/one-sidedness of the trustworthiness or enhance its objectivity
   b. Here **RT_A^B** is defined which synthesizes node A’s direct trust value to other neighbours and other neighbours’ comprehensive trust value to node B.

$$RT_A^B = \frac{\sum_{i=1}^{n} DT_A^{N_i} \cdot T_{N_i}^B \cdot W_N^{N_i}}{\sum_{i=1}^{n} DT_A^{N_i}}, \quad N_i \neq B$$

Entity-centric Model

3. **Comprehensive trust**
   a. Composed of direct trust and recommendation. But how to leverage the shares of them?
      i. When node A is very familiar to node B, A will believe its direct trust to B, and the recommendation is not important.
      ii. When node B is a strange node to A or its direct trust is less than the threshold, the recommendation is very important
   b. Based on these facts, a dynamic coefficient **α in [0,1]** is introduced to adjust their impacts on the comprehensive trust:

$$\alpha = \begin{cases} 1, DT_A^B \in (0.7, 1] \mathrm{or} DT_A^B \in [0, 0.3) \mathrm{or} W_N^B = 1 \\ DT_A^B, DT_A^B \in [0.5, 0.7) \\ W_N^B \cdot DT_A^B, DT_A^B \in [0.3, 0.5) \end{cases}$$

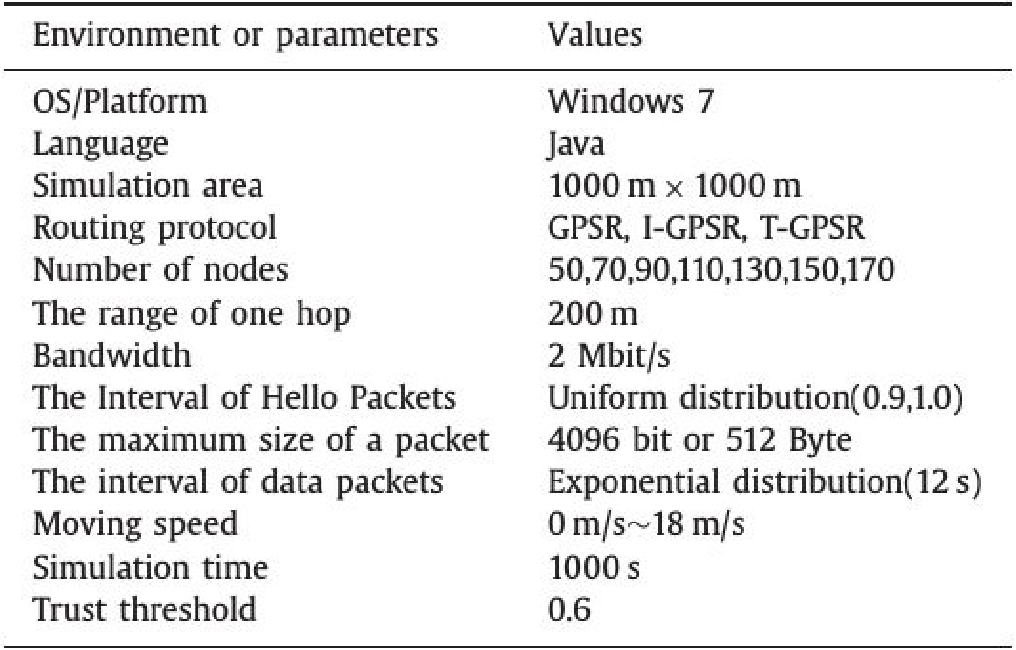
Entity-centric Model

3. **Comprehensive trust (cont'd)**
    a. Based on the coefficient **α**, the comprehensive trust $\mathbf{T_A{}^B}$ which is between [0,1] can be described as:

$$T_A{}^B = \alpha \cdot DT_A{}^B + (1 - \alpha)RT_A{}^B$$

Results

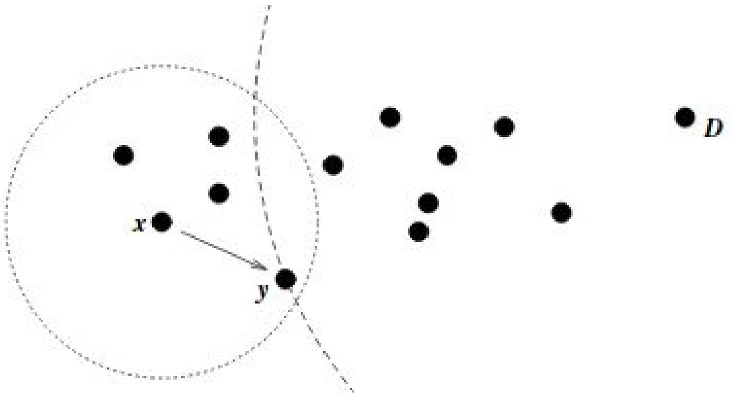
- **MobiSim** is used as the simulation environment and **GPSR** as the routing protocol.
- 3 scenarios are experimented:
  - Without attack
  - Black hole attack
  - Selective forwarding attack



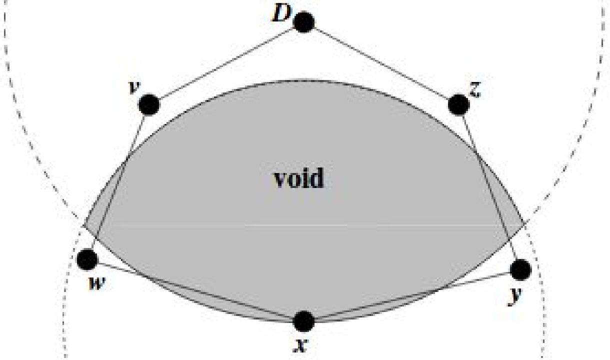
Environment or parameters	Values
OS/Platform	Windows 7
Language	Java
Simulation area	1000 m × 1000 m
Routing protocol	GPSR, I-GPSR, T-GPSR
Number of nodes	50,70,90,110,130,150,170
The range of one hop	200 m
Bandwidth	2 Mbit/s
The Interval of Hello Packets	Uniform distribution(0.9,1.0)
The maximum size of a packet	4096 bit or 512 Byte
The interval of data packets	Exponential distribution(12 s)
Moving speed	0 m/s~18 m/s
Simulation time	1000 s
Trust threshold	0.6

Results Definitions

- **Greedy Perimeter Stateless Routing (GPSR)**: uses *greedy forwarding* to forward packets to nodes that are always progressively closer to the destination.
    - Both advantage & disadvantage: geographic routing use only immediate neighbor information in forwarding decisions. Scalable and independent of the the length of the routes.

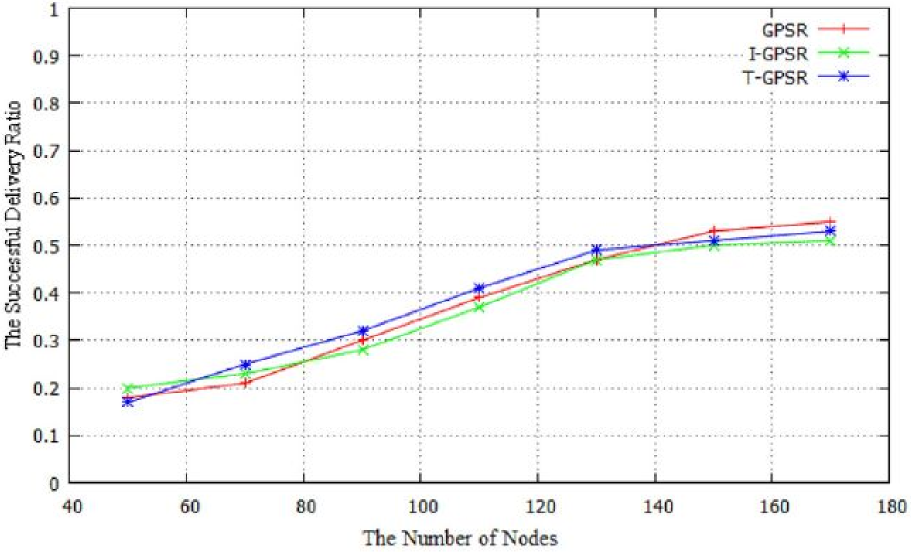


Results Definitions

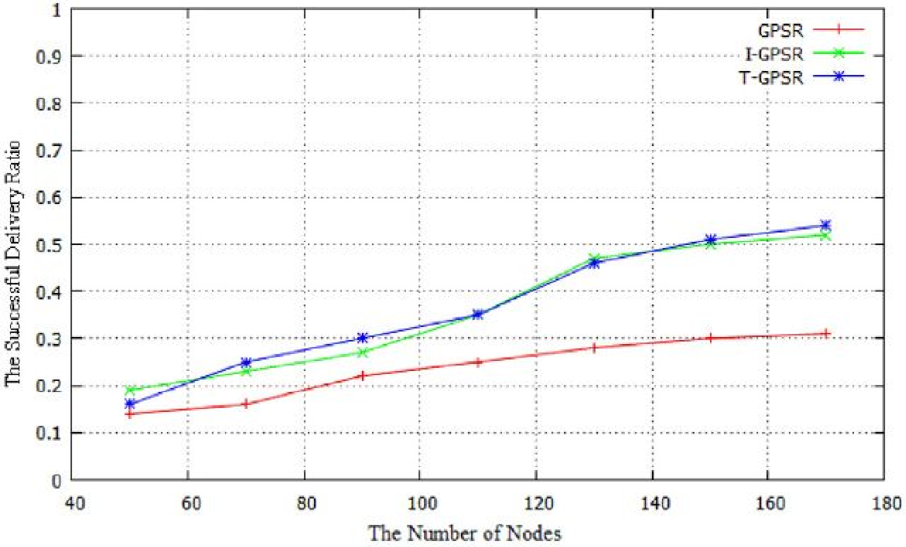


- **GPSR (cont'd):**
  - Where such a greedy path does not exist, GPSR recovers by forwarding in *perimeter mode*, in which a packet traverses successively closer faces of a planar subgraph of the full radio network connectivity graph.
- Proposed solution (T-GPSR) is compared with GPSR and I-GPSR
  - **I-GPSR** (2014): uses a Fuzzy extension of Multi-Entity Bayesian Network in VANETs. How?
- **Blackhole attack:** the malicious node will discard all packets.
- **Selective-forwarding attack:** the malicious node refers to the cunning, dishonest or selfish nodes, who only forward the low-weight, low-cost or the packets in their favor.

Results Normal vs Blackhole attack

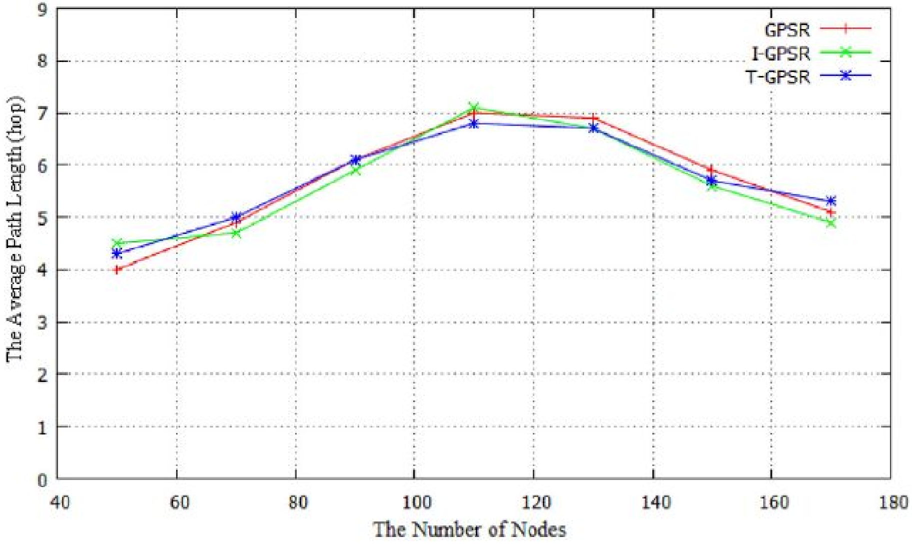


(a) The Data Delivery Ratio

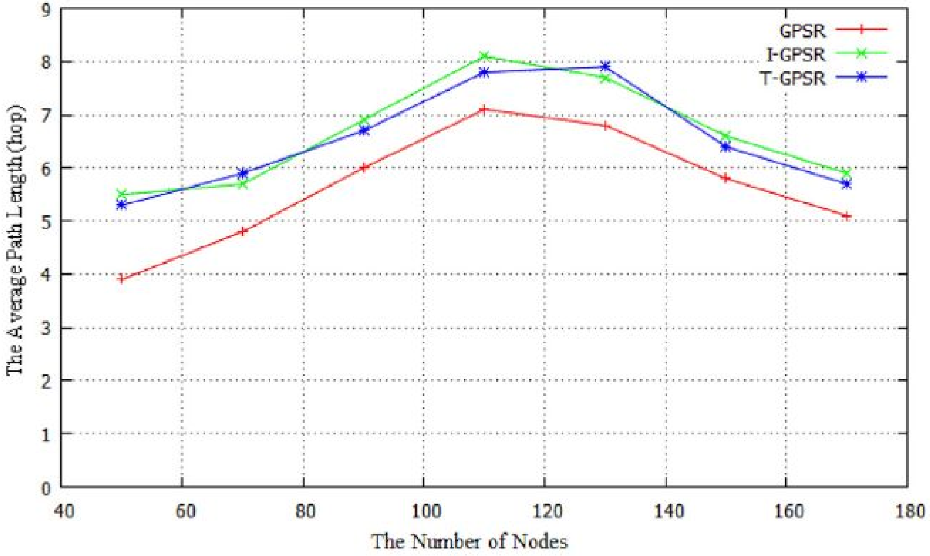


(a) The Packet delivery Ratio

Results Normal vs Blackhole attack

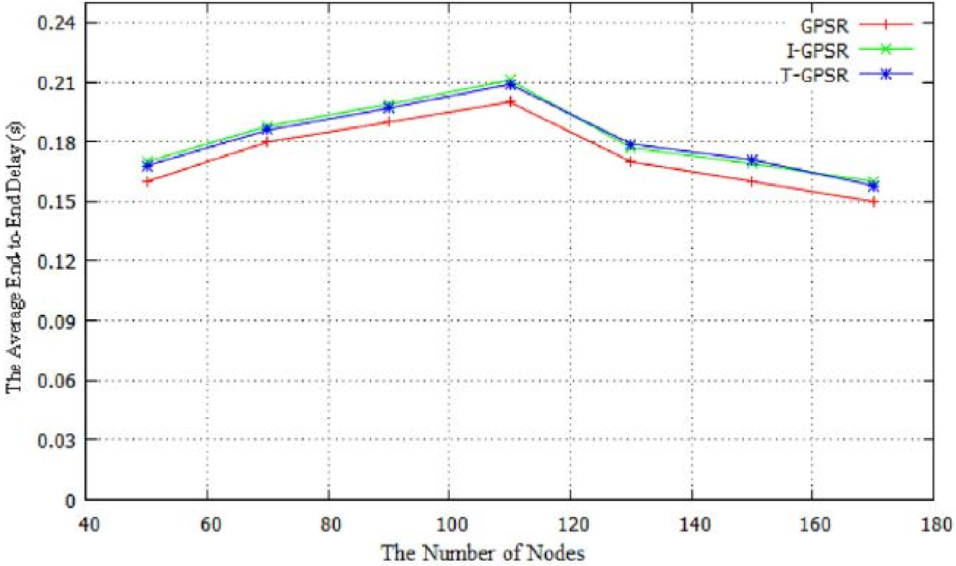


(b) The Average Path Length



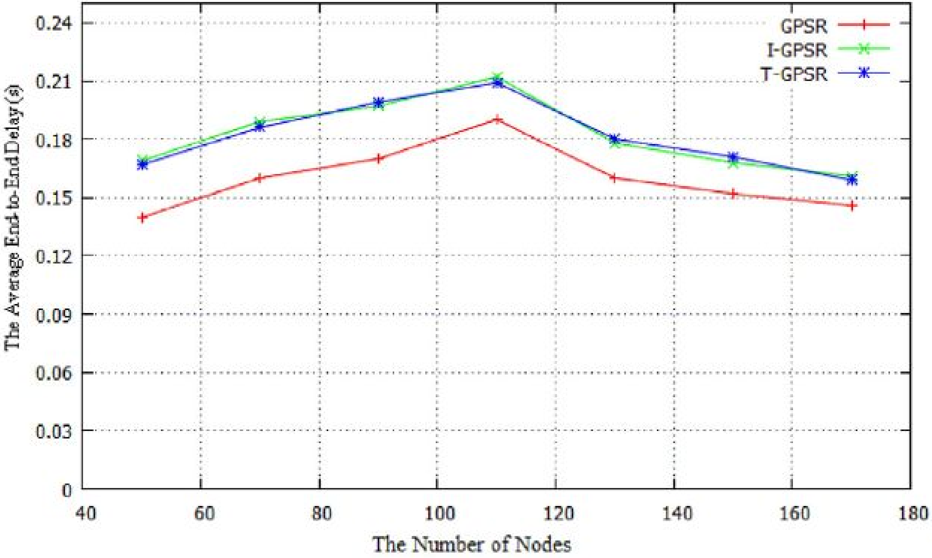
GPSR: The black-hole node is
rarely in the path by chance

Results Normal vs Blackhole attack



(c) The Average End-to-End Delay

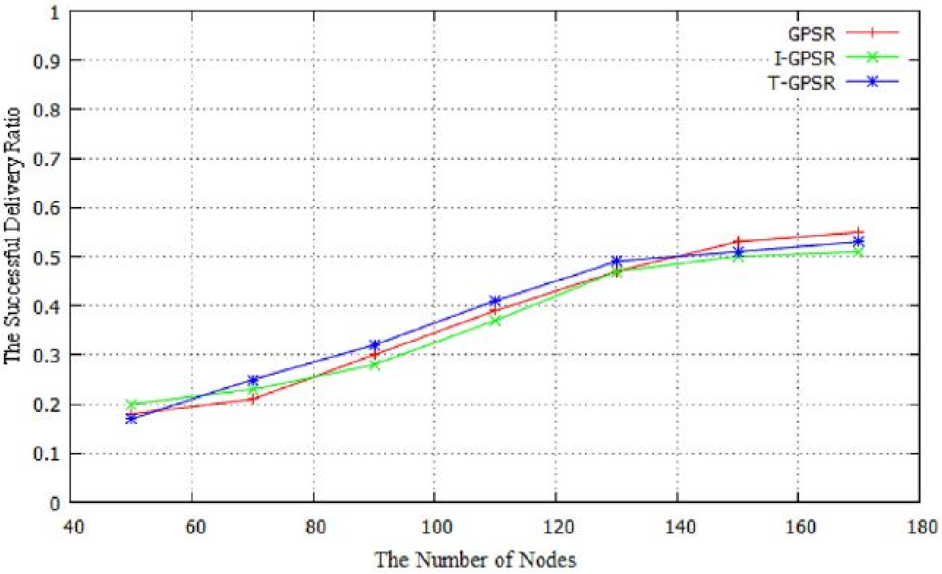
T-GPSR: Additional trust evaluation



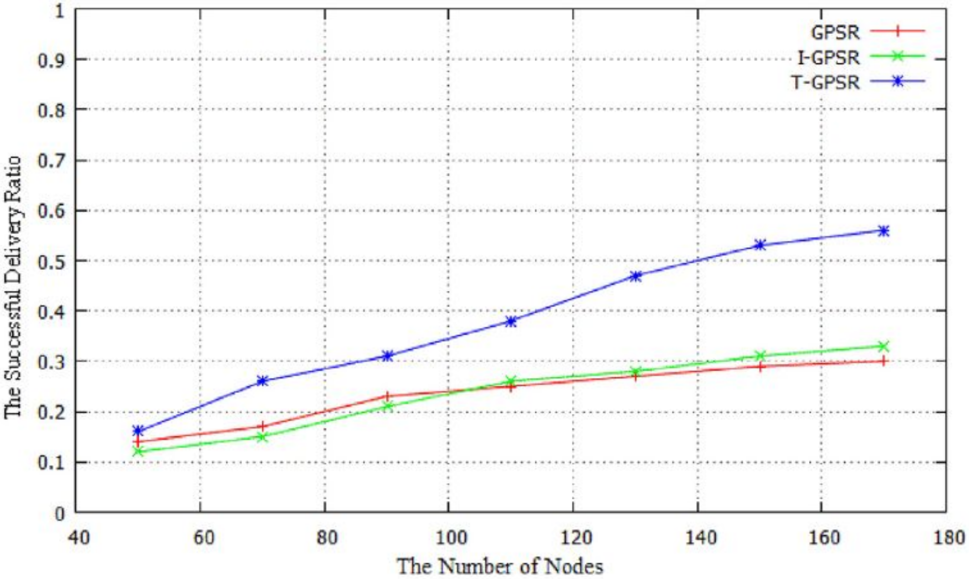
(c) The Average End-to-End Delay

T-GPSR: Same because malicious nodes is not many enough
GPSR: Some failure deliveries caused by the malicious node

Results Normal vs Selective-forwarding attack

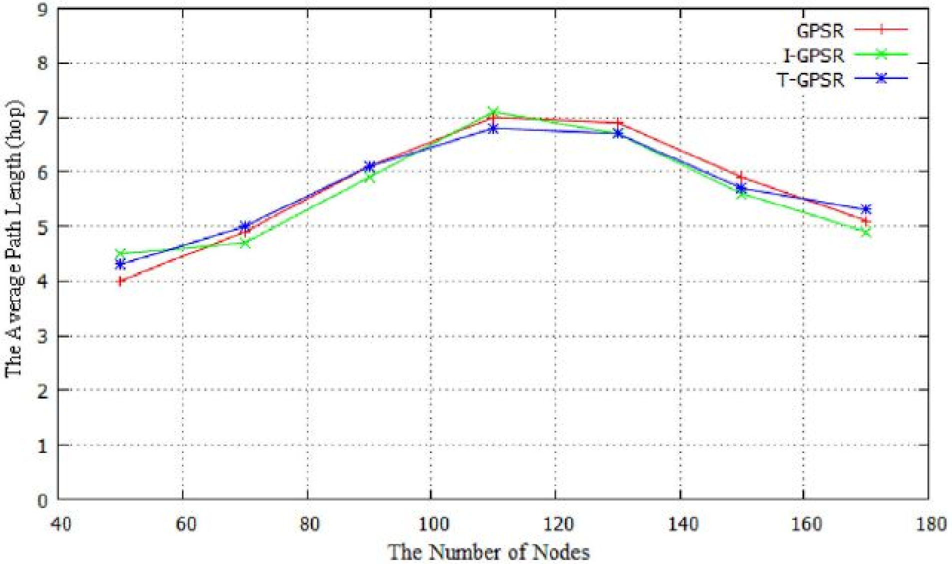


(a) The Data Delivery Ratio

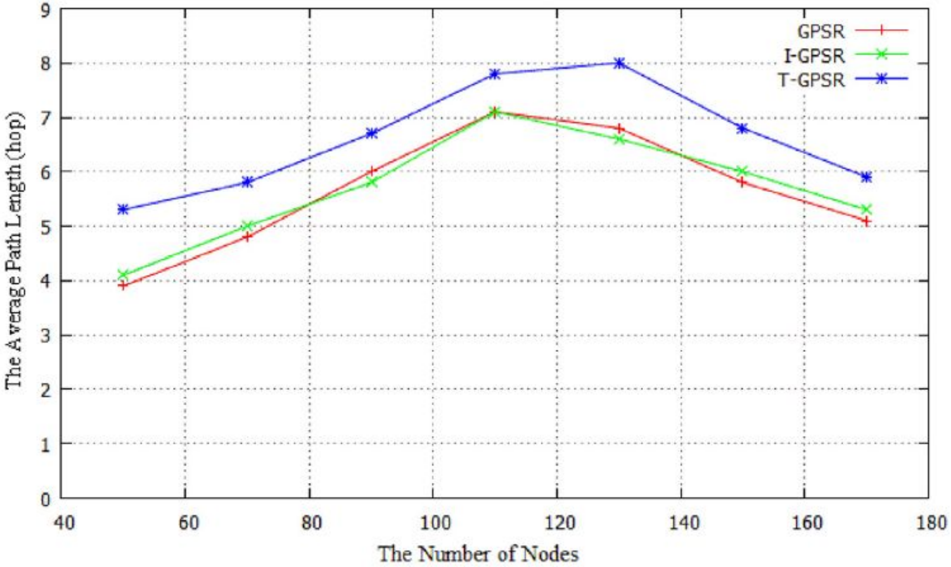


(a) The Packet delivery Ratio

Results Normal vs Selective-forwarding attack

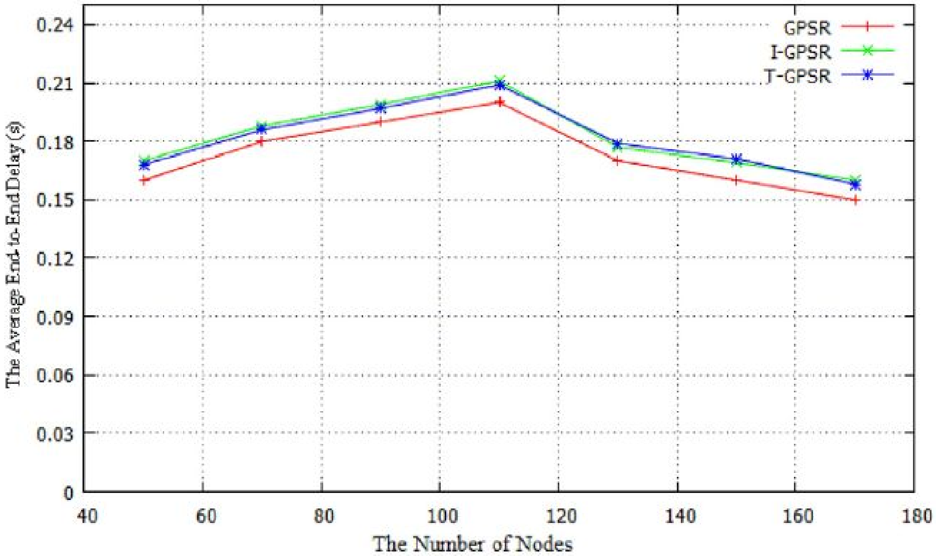


(b) The Average Path Length

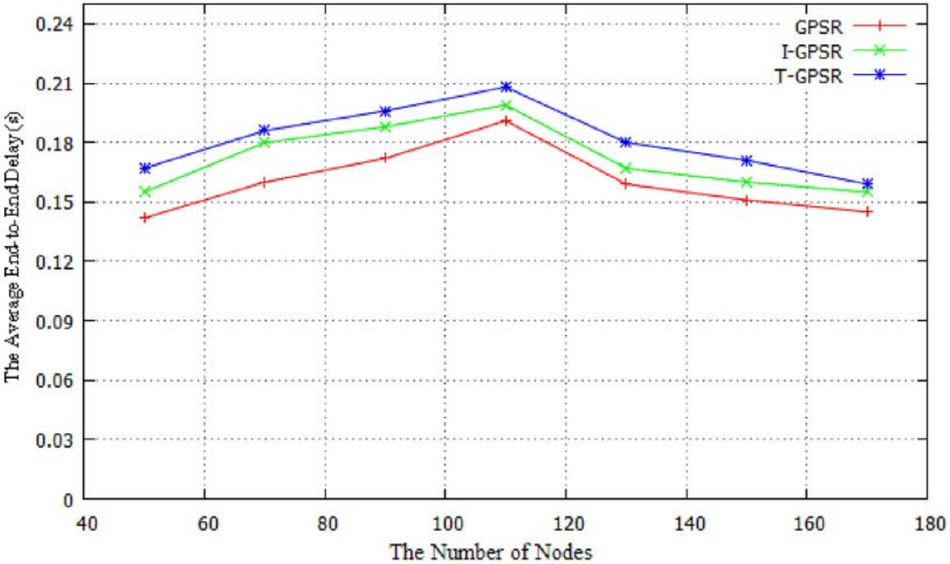


(b) The Average Path Length

Results Normal vs Selective-forwarding attack



(c) The Average End-to-End Delay



(c) The Average End-to-End Delay

Data-centric Model

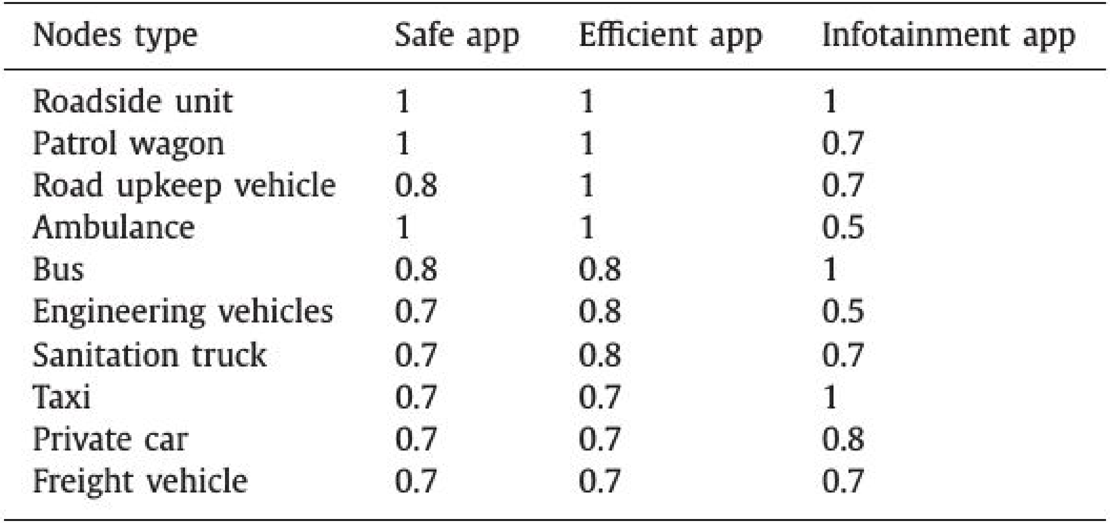
- Data dissemination is required to be as quick as possible, data trustworthiness evaluation method must equally be as simple as possible!
    - So instead of existing complicated trust evaluation techniques, a simple yet effective function must be used.
- Data trust can be determined by many factors. We focus on 4 factors in calculating the trust:
    - The data reporter's trustworthiness
    - The correlative trustworthiness of the event and its reporter.
    - The proximity in geographic location
    - The proximity in time

Data-centric Model

1. **Data reporter's trustworthiness**
    a. Determined by the proposed entity-centric trust model.
        i. T_A^B (the trustworthiness of node A to node B) is used.
    b. Depends on and changes with the trustworthiness of the reporter.
    c. Default value is its weight.

Data-centric Model

2. **The correlative trustworthiness of the event and its reporter**
   a. For a specific event/data, a reporter with low authority may be more trustworthy than the one with high authority; or the vehicles at the same authority level have different trustworthiness.
   b. Further classify nodes into 10 categories defined in a trust matrix ***M(T(v), T(λ))*** :



Nodes type	Safe app	Efficient app	Infotainment app
Roadside unit	1	1	1
Patrol wagon	1	1	0.7
Road upkeep vehicle	0.8	1	0.7
Ambulance	1	1	0.5
Bus	0.8	0.8	1
Engineering vehicles	0.7	0.8	0.5
Sanitation truck	0.7	0.8	0.7
Taxi	0.7	0.7	1
Private car	0.7	0.7	0.8
Freight vehicle	0.7	0.7	0.7

Data-centric Model

3. **The proximity in geographic location**
   a. The closer the reporter is to the location of the event, the more likely it is to have accurate information on the event, which makes the report more trustworthy.
   b. Influence of the distance μ_l between the reporter v's position and the event λ's locality can be seen below:

$$\mu_l(v,\lambda) = \begin{cases} 1 & d \leq 10m \\ 0.9 & 10m < d \leq 20m \\ 0.8 & 20m < d \leq 50m \\ 0.7 & 50m < d \leq 100m \\ 0.6 & 100m < d \leq 200m \\ 0.4 & 200m < d \leq 500m \\ 0 & d > 500m \end{cases}$$

Data-centric Model

4. **The proximity in time**
   a. The shorter interval between the event occurrence and the report generated, the more likely it can reflect the system status.
   b. According statistics, the traffic event usually be solved from **5 mins. to an hour**.
   c. Influence of the interval μ_t between the event occurrence and the report generated by node v can be seen below:

$$\mu_t(v, \lambda) = \begin{cases} 1 & t \leq 5min \\ 0.9 & 5min < t \leq 10min \\ 0.8 & 10min < t \leq 20min \\ 0.7 & 20min < t \leq 30min \\ 0.5 & 30min < t \leq 45min \\ 0.3 & 45min < t \leq 60min \\ 0 & t \geq 60min \end{cases}$$

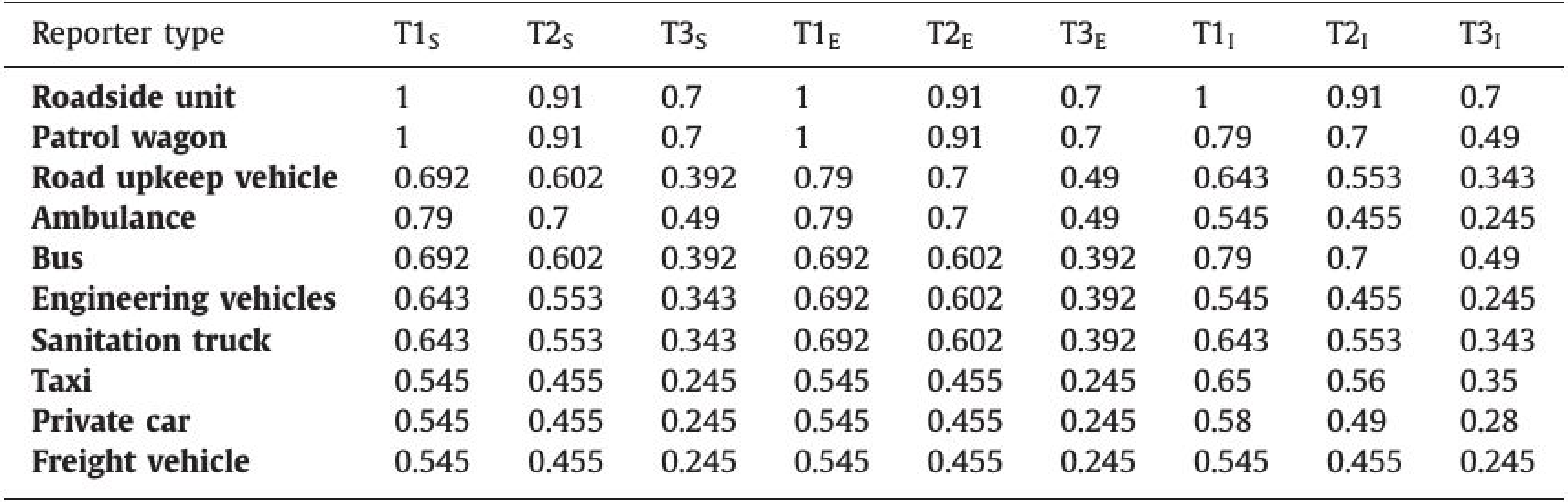
Data-centric Model

When a node A receives a data report λ generated by node v, it can evaluate the data trustworthiness by: $B_{\lambda}^{v} = 0.7 \cdot T_{A}^{v} \cdot M(\tau(v), \tau(\lambda)) + 0.15 \cdot \mu_{l}(v, \lambda) + 0.15 \cdot \mu_{t}(v, \lambda)$

- Data-centric trust model is simple enough to realize fast trustworthiness evaluation due to all the factors being predefined or determined with the data received. Therefore it can be computed in real time.
- Trustworthiness of the same data from different reporters is usually different. The receiver can use the average of several trust values for the same event as the final trust value.

Results

- In order to evaluate the model, 3 kinds of data (safety, efficiency, infotainment) from the 10 types of nodes with different proximity location and time (best, medium and worst cases) is analyzed. It is assumed the trustworthiness of the reporter is its weight.



Reporter type	$T1_S$	$T2_S$	$T3_S$	$T1_E$	$T2_E$	$T3_E$	$T1_I$	$T2_I$	$T3_I$
Roadside unit	1	0.91	0.7	1	0.91	0.7	1	0.91	0.7
Patrol wagon	1	0.91	0.7	1	0.91	0.7	0.79	0.7	0.49
Road upkeep vehicle	0.692	0.602	0.392	0.79	0.7	0.49	0.643	0.553	0.343
Ambulance	0.79	0.7	0.49	0.79	0.7	0.49	0.545	0.455	0.245
Bus	0.692	0.602	0.392	0.692	0.602	0.392	0.79	0.7	0.49
Engineering vehicles	0.643	0.553	0.343	0.692	0.602	0.392	0.545	0.455	0.245
Sanitation truck	0.643	0.553	0.343	0.692	0.602	0.392	0.643	0.553	0.343
Taxi	0.545	0.455	0.245	0.545	0.455	0.245	0.65	0.56	0.35
Private car	0.545	0.455	0.245	0.545	0.455	0.245	0.58	0.49	0.28
Freight vehicle	0.545	0.455	0.245	0.545	0.455	0.245	0.545	0.455	0.245

Conclusion

- A **dynamic entity-centric trust model** based on data and node weight is proposed by correlating data kinds to nodes types with dynamical coefficient to balance the direct trust and recommend trust.
- It is shown that the model can resist **black hole attack** and **selective forwarding attack** at the cost of lowering the performance slightly.
- A lightweight **data-centric trust model** is presented which is simple enough for **timely trust evaluation**.
- Failed to mention **number of malicious nodes** in the simulation results of the proposed entity-centric model
- The data-centric model should be **further optimized in utility parameters** and the default values in future.

Thank you for listening!