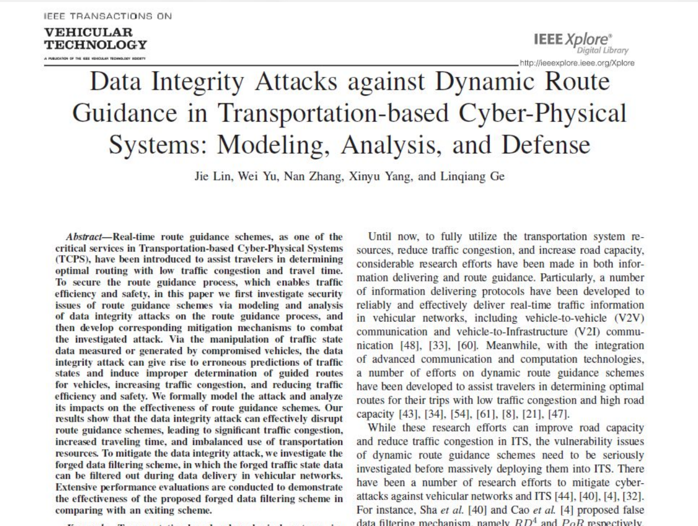
AN ANALYSIS & REVIEW

Data Integrity Attacks against Dynamic Route Guidance in Transportation-based Cyber-Physical Systems: Modeling, Analysis and Defense

Jie Lin, et al. (10.1109/TVT.2018.2845744)

Presented by: Ellis Thompson


IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY
IEEE Xplore Digital Library
http://ieeexplore.ieee.org/Xplore

Data Integrity Attacks against Dynamic Route Guidance in Transportation-based Cyber-Physical Systems: Modeling, Analysis, and Defense

Jie Lin, Wei Yu, Nan Zhang, Xinyu Yang, and Linqiang Ge

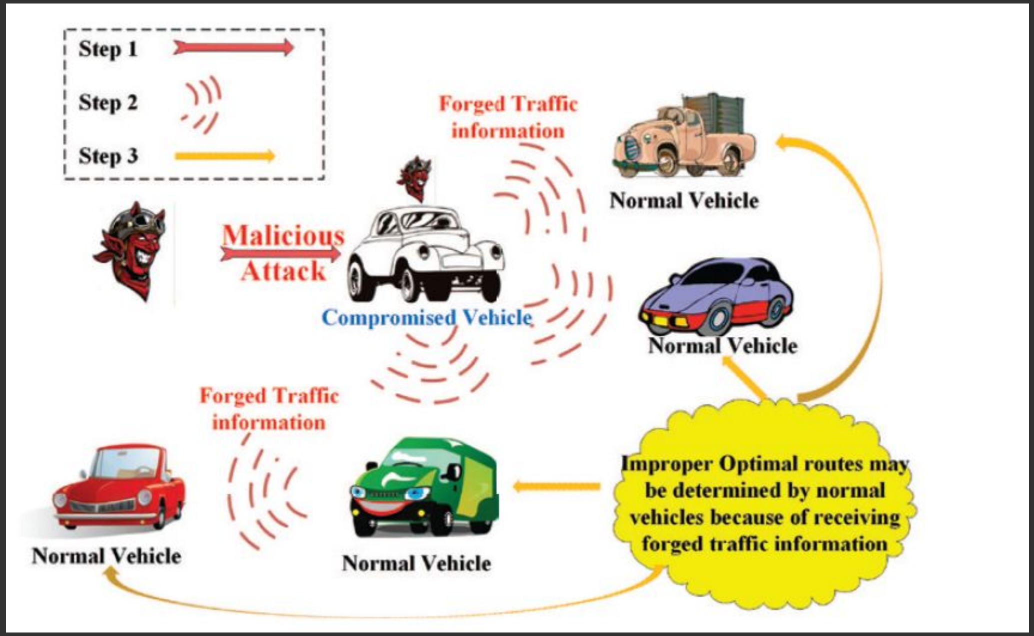
Abstract—Real-time route guidance schemes, as one of the critical services in Transportation-based Cyber-Physical Systems (TCPS), have been introduced to assist travelers in determining optimal routing with low traffic congestion and travel time. To secure the route guidance process, which enables traffic efficiency and safety, in this paper we first investigate security issues of route guidance schemes via modeling and analysis of data integrity attacks on the route guidance process, and then develop corresponding mitigation mechanisms to combat the investigated attack. Via the manipulation of traffic state data measured or generated by compromised vehicles, the data integrity attack can give rise to erroneous predictions of traffic states and induce improper determination of guided routes for vehicles, increasing traffic congestion, and reducing traffic efficiency and safety. We formally model the attack and analyze its impacts on the effectiveness of route guidance schemes. Our results show that the data integrity attack can effectively disrupt route guidance schemes, leading to significant traffic congestion, increased traveling time, and imbalanced use of transportation resources. To mitigate the data integrity attack, we investigate the forged data filtering scheme, in which the forged traffic state data can be filtered out during data delivery in vehicular networks. Extensive performance evaluations are conducted to demonstrate the effectiveness of the proposed forged data filtering scheme in comparing with an exiting scheme.

Until now, to fully utilize the transportation system resources, reduce traffic congestion, and increase road capacity, considerable research efforts have been made in both information delivering and route guidance. Particularly, a number of information delivering protocols have been developed to reliably and effectively deliver real-time traffic information in vehicular networks, including vehicle-to-vehicle (V2V) communication and vehicle-to-Infrastructure (V2I) communication [48], [33], [60]. Meanwhile, with the integration of advanced communication and computation technologies, a number of efforts on dynamic route guidance schemes have been developed to assist travelers in determining optimal routes for their trips with low traffic congestion and high road capacity [43], [34], [54], [61], [8], [21], [47].

While these research efforts can improve road capacity and reduce traffic congestion in ITS, the vulnerability issues of dynamic route guidance schemes need to be seriously investigated before massively deploying them into ITS. There have been a number of research efforts to mitigate cyber-attacks against vehicular networks and ITS [44], [40], [4], [32]. For instance, Sha *et al.* [40] and Cao *et al.* [4] proposed false data filtering mechanism, namely RD^4 and PoR respectively,

MOTIVATION

- Infiltrate and disrupt route planning networks
- Forging traffic information to contrast true observation
- Results in little damage for 1st order neighbours
- Results in sub-optimal trajectories (information propagation)
  - Delays/Congestion for n-order vehicles



REAL-TIME INFORMATION MEASUREMENT

- A single message regarding the state of a single vehicle

$$r_u = \left(u|T|R_d|Sg_{Rd}|W_{Rd}|L_{Rd}|Ds_u^{Sg_{Rd}}\right)$$

- Where:
  - u Vehicle ID
  - T Time Slot Number
  - R_d Road/Street ID
  - Sg_{Rd} Road Segment ID
  - W_{Rd} Road Width
  - L_{Rd} Road Length
  - $Ds_u^{Sg_{Rd}}$ The distance from u to the centre of the segment

REAL-TIME TRAFFIC INFORMATION MESSAGE

- A collection of messages describing the traffic state at some time T

$$NR_T^{Sg_{Rd}} = \left(T|Rd|Sg_{Rd}|W_{Rd}|L_{Rd}\middle|Sd_T^{Sg_{Rd}}\middle|Num_T^{Sg_{Rd}}\right)$$

- Where:
  - T Time Slot Number
  - R_d Road/Street ID
  - Sg_{Rd} Road Segment ID
  - W_{Rd} Road Width
  - L_{Rd} Road Length
  - $Sd_T^{Sg_{Rd}}$ Travel speed in segment Sg_{Rd} at T
  - $Num_T^{Sg_{Rd}}$ Number of vehicles in segment Sg_{Rd} at T

DETERMINATION OF TRAFFIC STATE

- Trust probability (TP) Used to predict traffic state at a future time
- Better describing the movement of vehicles on a given road segment

$$TP_{T_j}^{Rd_i} = P\left(Pin_{T_j}^{Rd_i} \leq Pout_{T_j}^{Rd_i}\right)$$

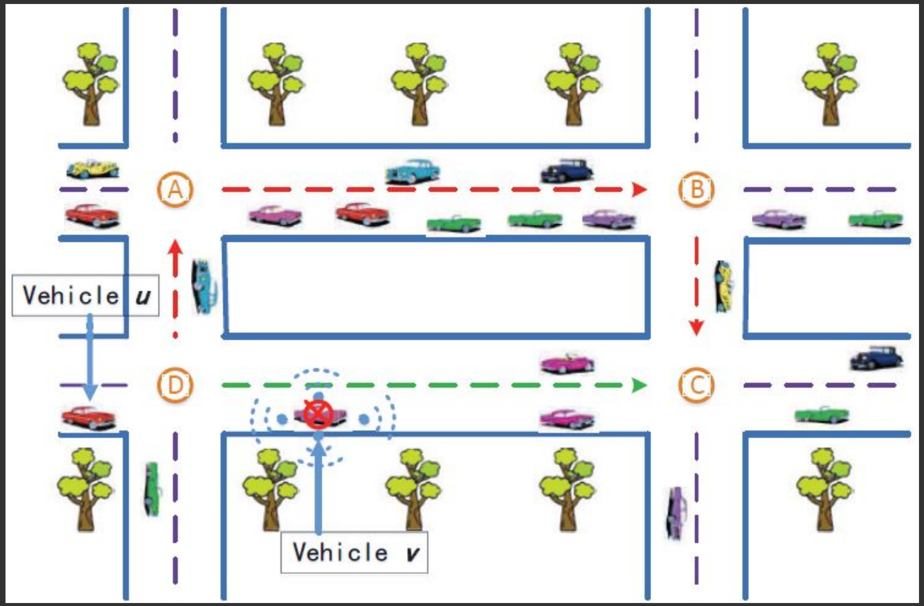
$$TP_{T_j}^{Rd_i} < \phi$$

- "The probability that the # of vehicles entering a road segment ≤ the number of vehicles leaving a road segment at time T"
- "This probability *must* be < some threshold ϕ to be *trustworthy*"

Note: This TP threshold satisfies a clear road $TP_{T_j}^{Rd_i} > \phi$ would satisfy a congested road

DATA INTEGRITY ATTACK

- Cant be random
  - Has to be calculated to cause maximum *damage*
- Congestion
  - Falsify the number of vehicles in a road segment
    - Make this bigger or smaller
- Compromised vehicle v reporting congestion on Rd_{DC}
- Vehicle u takes route $Rd_{DA} \rightarrow Rd_{AB} \rightarrow Rd_{BC}$
- Satisfy $TP_{Tj}^{*Rd_i} < \phi$
- Falsify $Pout_{T_j}^{Rd_i} \rightarrow Pout_{T_j}^{*Rd_i}$

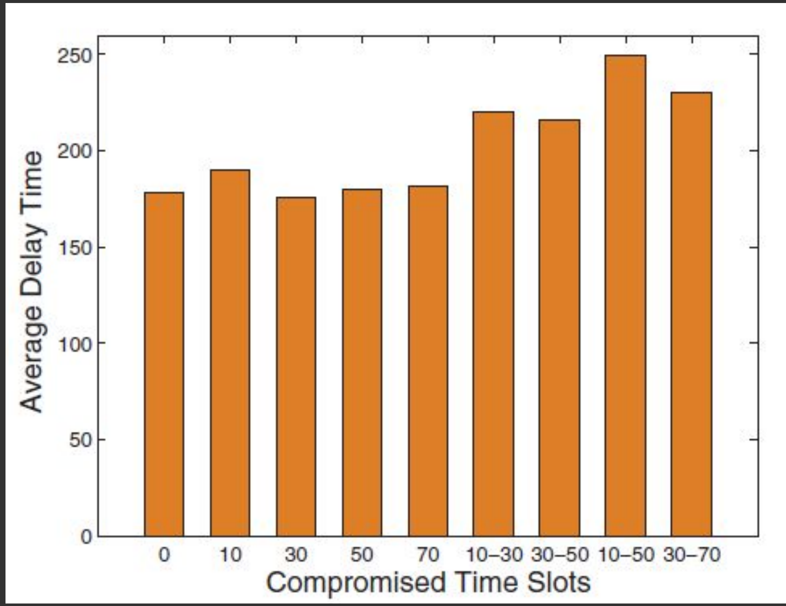


Note: '*' is used to denote a manipulated value

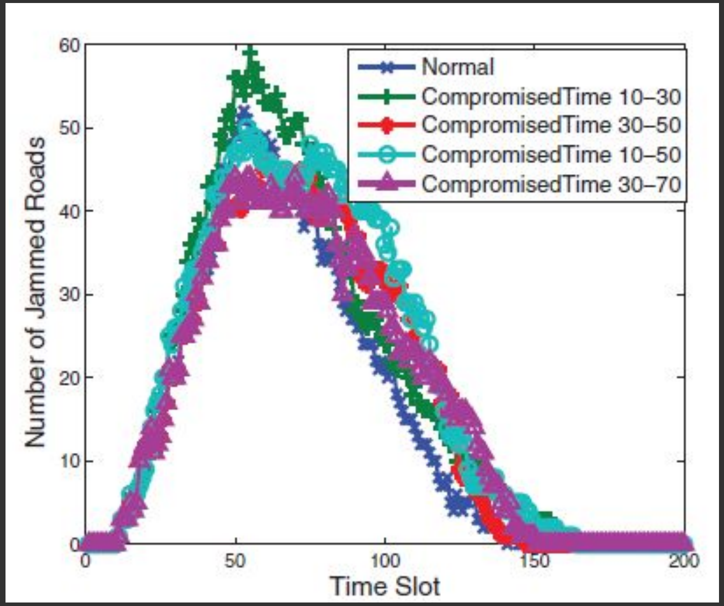
INTERJECTION

- I have evaded explaining modelling and analysis optimal attack locations/strategies
- It’s a lot of maths… You don’t want that… I don’t want that…
- The 411:
  - *Optimal attacking location*
    - Find the road(s) which would cause the *most* congestion. Steer vehicles towards these
  - *Optimal attacking distribution*
    - Concentrated – M roads in a concentrated area (better attack for non-traffic balanced)
    - Distributed – All roads on a network (better attack for traffic balanced)
  - *Optimal attacking coverage ratio*
    - How can we affect the most number of roads
    - Ratio of compromised roads (bigger the ratio higher likelihood of serious congestion)

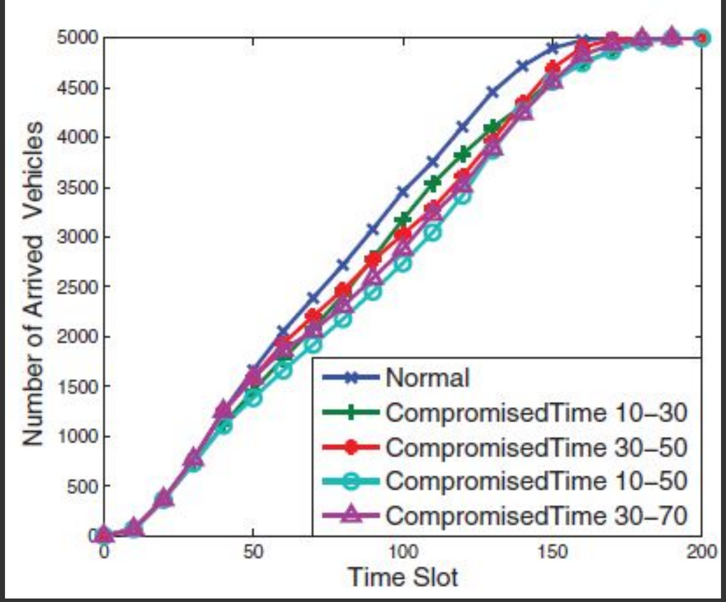
SOME ATTACK STATS



Delay in vehicles reaching their destination based on when roads where attacked (higher is better)



Number of jammed roads for a given timeslot, and how long they are compromised for (higher is better) – Continuous attack



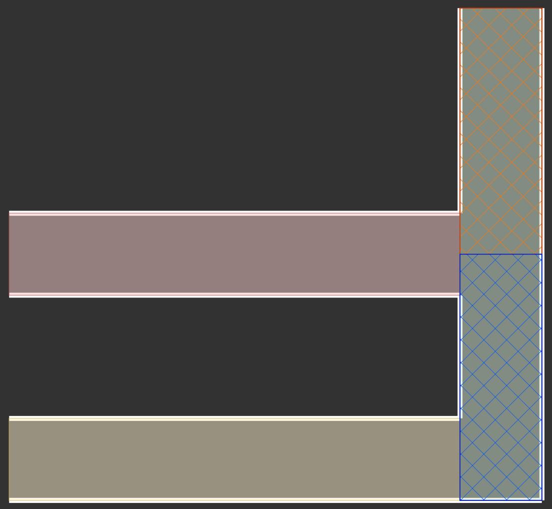
Cumulative number of vehicles reaching their destination (lower gradient is better) – Continuous attack

FORGED DATA FILTERING

FDF – TERMINOLOGY

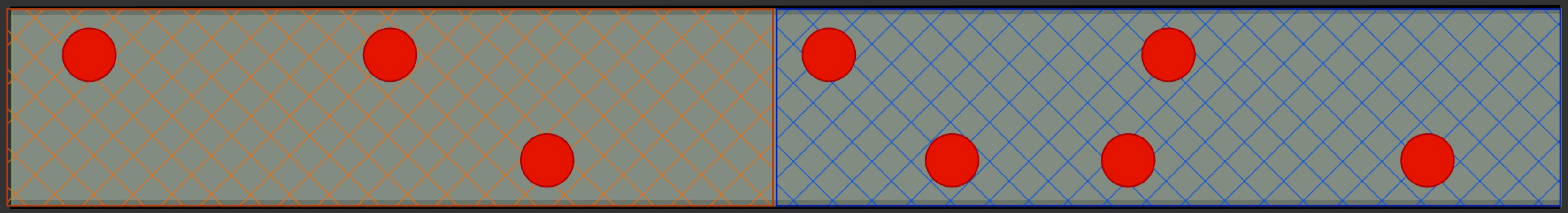
Environment – This contains *all* roads

Roads – exist in the environment colour shaded but identified by an ID

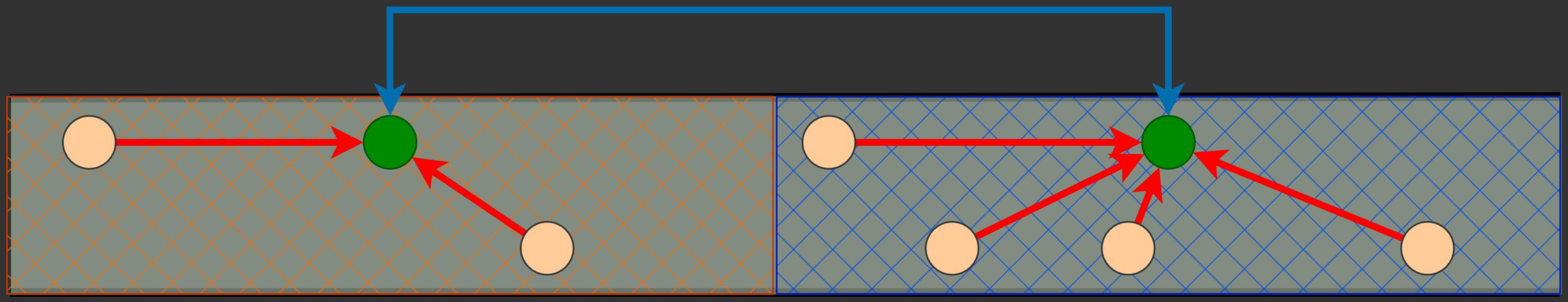


Road Segments – Cross hatch, breaks up a road into smaller segments, these will contain clusters of vehicles

FDF – TERMINOLOGY



Our vehicles exist in segments on the road (direction doesn't matter)



Segments form clusters with the vehicle closest to the centre becoming the cluster head

FDF – SOLUTION

- Local clusters C_i and foreign clusters C_j
- Conditions:
  - **IF** message is stale – Drop Message
    - **IF** our auth check *not in* foreign cluster – Drop Message
    - **IF** the number of MACs (n) < number of vehicles 0n a segment at time T – Drop Message
      - C_j validates all MACs attached to segment
      - **IF** any MAC fails to validate – Drop Message
- Providing all conditions are met the message is sent to the next cluster head

$$NR_T^{Sg_{Rd}} = \left(T|Rd|Sg_{Rd}|W_{Rd}|L_{Rd}\left|Sd_T^{Sg_{Rd}}\right|Num_T^{Sg_{Rd}}\right)$$

Algorithm 1 Forged Data Filtering (FDF)

Input: Real-time traffic information messages (NR_T^{SgRd}) , Cluster-head vehicles C_i and C_j

Output: Forged traffic state data

```
Cluster-head vehicle (C_j) receives the real-time traffic information message (NR_T^{SgRd}) of cluster (C_i)
if (Time is not Fresh) then
    Drop the message;
    Exit
else
    if (Check Polynomial of C_i is included in C_j) then
        Obtain the number of different MACs attached in NR_T^{SgRd}, namely n
        if (n < T) then
            Drop the message
            Exit
        else
            C_j Validates MACs attached on the received message (NR_T^{SgRd}) with stored Check Polynomial in C_j
            if (All MACs are valid) then
                C_j considers message NR_T^{SgRd} is true and send to next cluster-head vehicle
                Exit
            else
                C_j considers message NR_T^{SgRd} is false and drop it
                Exit
            end if
        end if
    else
        C_j sends message NR_T^{SgRd} to next cluster-head vehicle
        Exit
    end if
end if
```

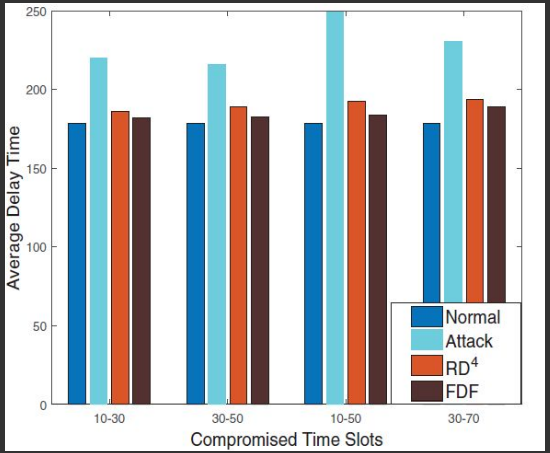
FDF – AUTHENTICATION

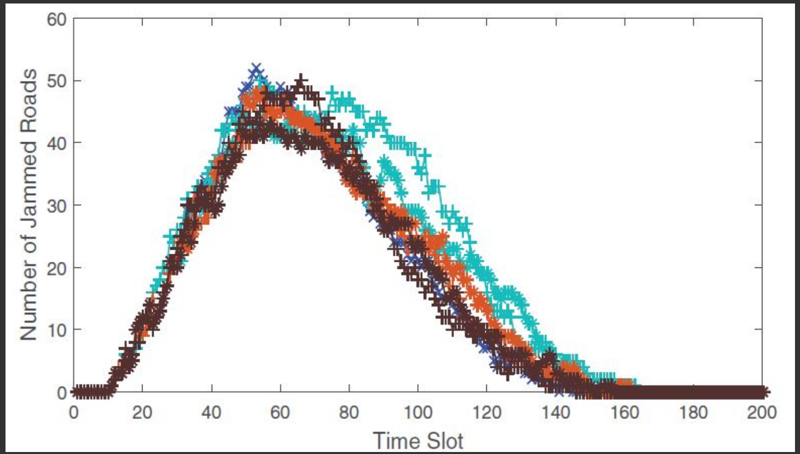
- Message Authentication Code (MAC) for *forwarded* data
- Each vehicle in a cluster can compute the authentication of the local cluster C_i
- u can generate a check for another cluster C_j
- A forwarding vehicle v is responsible of checking the validity of the message from C_i iff it stores the validation already for C_i (assumes validity)
  - If it does then it checks all MACs in in C_i

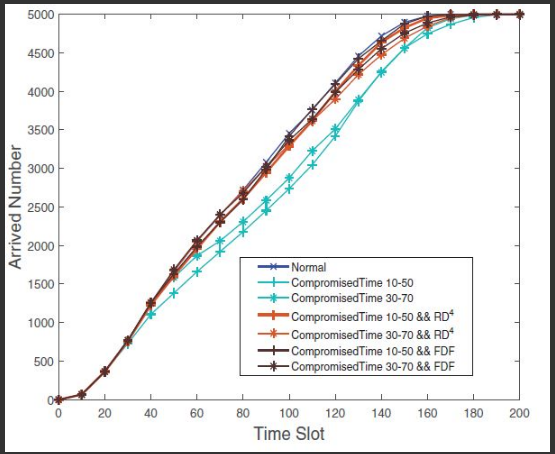
Think of this sort of like public and private keys the authentication and verification can be obtained from the message data and the MAC is there to validate

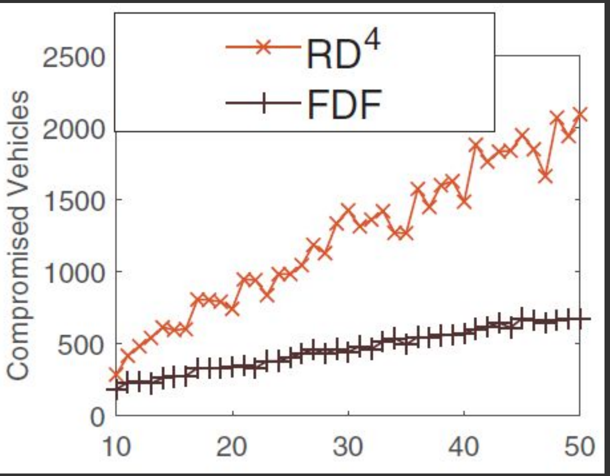
Note: this authentication is based on the concept of Primitive Polynomials it is too complex to explain here

FDF – RESULTS







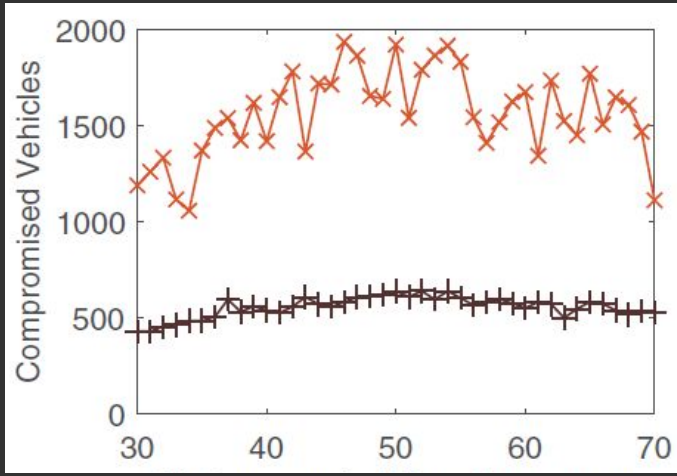


Reduced delay

Fewer jammed roads

Closer to baseline arrivals

Much fewer compromised vehicles



QUESTIONS, CRITIQUE & DISCUSSION

SOME POINTS TO THINK ABOUT

- This approach does require additional equipment to provide frequent changes to MAC generation *polynomial*
  - Could this be a potential issue?
- Nothing was mentioned about computational time inc/dec over RD^4, could this have an impact?
- There's a clear improvement with number of compromised vehicles, but how significant are the other improvements?
- Assumptions are made with regard to vehicle speed, i.e. that the speed of the vehicle will always be the most optimal (quickest) between 2 points. Is it fair to make that assumption?