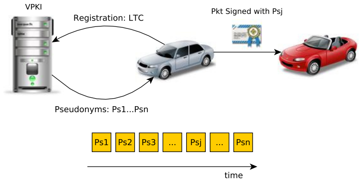
KTH ROYAL INSTITUTE
OF TECHNOLOGY

Cooperative Location Privacy in Vehicular Networks: Why Simple Mix-zones are not Enough

Mohammad Khodaei and Panos Papadimitratos
Networked Systems Security Group (NSS)
www.eecs.kth.se/nss

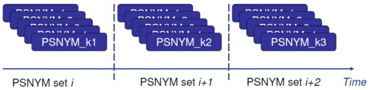
In IEEE Internet of Things (IoT) Journal

Security and Privacy for Vehicular Communication (VC) Systems[1]



Basic Requirements

- Authentication & integrity
- Non-repudiation
- Authorization and access control
- Conditional anonymity
- **Unlinkability (long-term)**



Vehicular Public-Key Infrastructure (VPKI)

- Pseudonymous authentication
- Trusted Third Party (TTP):
  - Certification Authority (CA)
  - Issues credentials & binds users to their pseudonyms

[1] P. Papadimitratos, et al. "Securing Vehicular Communications - Assumptions, Requirements, and Principles," in ESCAR, Berlin, Germany, pp. 5-14, Nov. 2006.
P. Papadimitratos, et al. "Secure Vehicular Communication Systems: Design and Architecture," in IEEE Communications Magazine, vol. 46, no. 11, pp. 100-109, Nov. 2008.

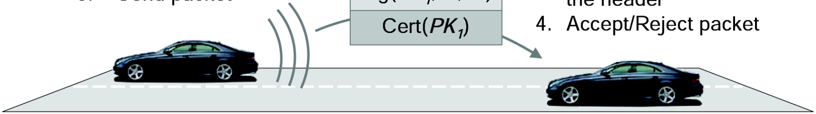
Security and Privacy for VC Systems (cont'd)

Beacon packet

1. Generate signature with SK_1
2. Append certificate
3. Send packet

Header: H
Payload: m
Sig(SK_1, H, m)
Cert(PK_1)

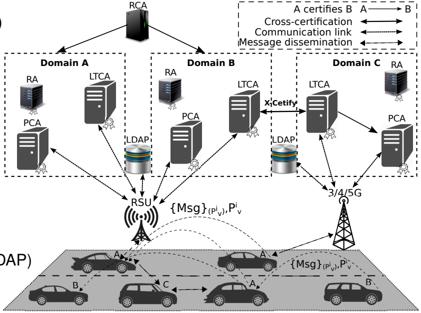
1. Validate certificate (if not previously done so)
2. Validate signature
3. Validate geo-stamp in the header
4. Accept/Reject packet



- Sign packets with the private key, corresponding to the current valid pseudonym
- Verify packets with the valid pseudonym
- Cryptographic operations in a Hardware Security Module (HSM)

Security and Privacy for VC Systems (cont'd)

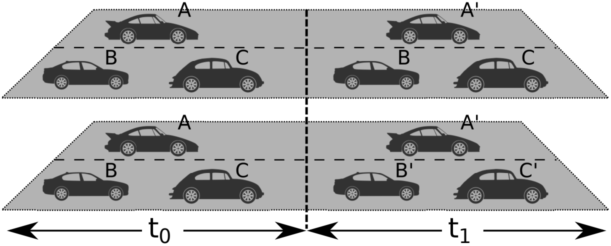
- Vehicular Public-Key Infrastructure (VPKI)
- Root CA (RCA)
- Long Term CA (LTCA)
- Pseudonym CA (PCA)
- Resolution Authority (RA)
- Lightweight Directory Access Protocol (LDAP)
- Roadside Unit (RSU)



- Vehicles registered with one LTCA (home domain)
- PCA servers in one or multiple domains
- Vehicles can obtain pseudonyms from any PCA
- Establish trust among entities with a RCA or with cross-certification
- Resolve (de-anonymize) a pseudonym with the help of an RA

Vehicle Traceability (Syntactic & Semantic Linking Attacks)

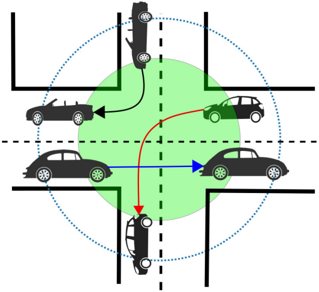
- Leveraging *K-anonymity*, obfuscating Cooperative Awareness Messages (CAMs), or silent period
  - Diminishing situational awareness, thus, affecting operation of safety applications
- Leveraging group signature schemes
  - Computation overhead; only mitigating syntactic linking attack
- Synchronous pseudonym updates
  - Only mitigating syntactic linking attack



Vehicle Traceability (Syntactic & Semantic Linking Attacks) (cont'd)

Cryptographic Mix-Zone (CMIX):

- Mitigating syntactic and semantic linking attacks
- Without affecting the operation of safety applications



- Arrival rates
- Mix-zone geometries
- Physical constraints of the road layout
- Mobility patterns (e.g., velocity, acceleration)
- Vehicle density (e.g., sparse traffic conditions)

Challenges & Motivation

- Mix-zone geometries
- Mobility patterns (e.g., velocity, acceleration, etc.)
- Vehicle density (e.g., sparse traffic conditions)
- Arrival rates
- Physical constraints of the road layout
- Honest-but-curious entities

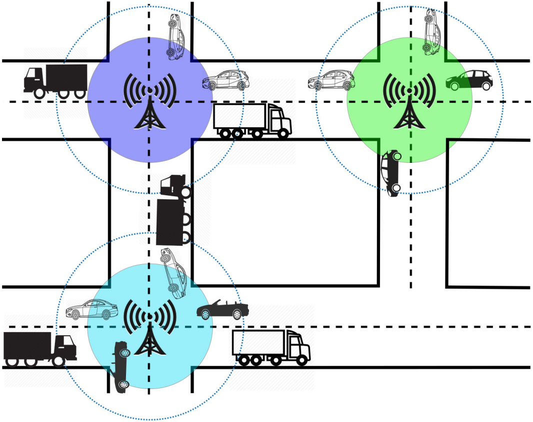
Adversarial Model

- External adversaries with wireless receivers, placed near each mix-zone, eavesdrop communication
- Internal adversaries:
  - Initiating the protocol continuously to impose extra overhead on the system (a DoS attack).
  - Opting in not changing their pseudonyms, or preventing others from changing their pseudonyms.
  - Colluding internal nodes could broadcast CAMs with the same (*"chaff"*) pseudonym from two distinct location.
  - Colluding and sharing information that each of them individually collected, e.g., an *honest-but-curious* RSU with a single VPKI entity.

Requirements

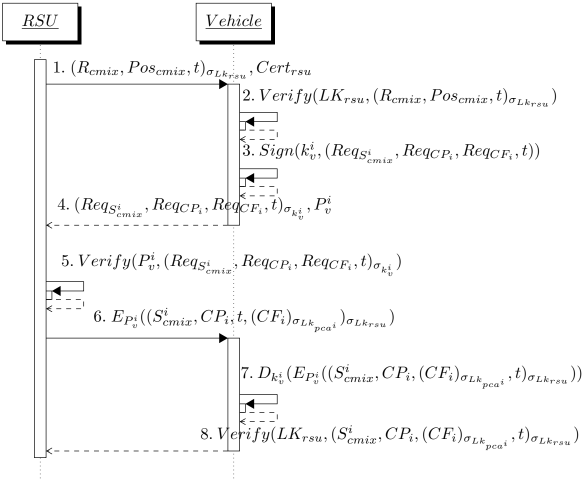
- Privacy (anonymity and unlinkability)
- Availability
- Auditability and misbehavior detection
- Efficiency and scalability
- Notification on CMIX parameters

Mix-zones Construction with Decoy Traffic



- What about safety applications?
  - Dissemination of a signed Cuckoo Filter (CF)

Mix-zones Advertisement and Chaff Pseudonym Acquisition Protocols



Experimental Setup

- OMNET++ & Veins framework using SUMO
- Cryptographic protocols and primitives (OpenSSL): Elliptic Curve Digital Signature Algorithm (ECDSA)-256 and SHA-256 as per IEEE 1609.2 and ETSI standards
- V2X communication over IEEE 802.11p
- Placement of the mix-zones: “highly-visited” intersections with non-overlapping radio ranges

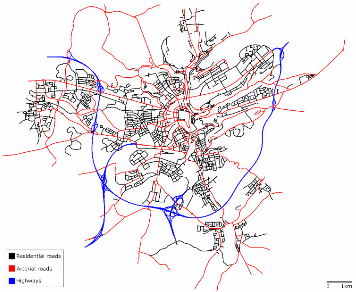


Figure: The LuST dataset, a full-day realistic mobility pattern in the city of Luxembourg (15KM x 15KM) [Codeca et al. (2015)].

Experimental Setup (cont'd)

- One PCA for CF dissemination
- RSUs randomly assign a percentage of vehicles to be relaying ones
- For CF operations (insertion and membership test), we used *PYBLOOM*

- Metrics:
  - Average successful tracking through syntactic and semantic linking attacks
  - Efficiency (latency)
  - Resilience (internal adversaries)
  - Resource consumption (computation/communication)

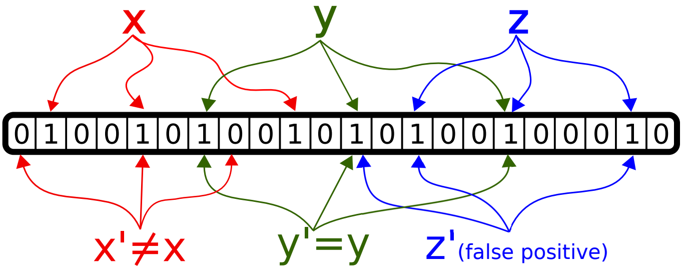
Table: Simulation parameters.

Parameters	Value	Parameters	Value
Beacon TX interval (γ_v)	0.2s, 0.5s, 1s	Number of RSUs	100
Carrier frequency	5.89 GHz	RSUs transmission range	600 meter
TX power	20mW	Number of Mix-zones	25
Physical layer bit-rate	18Mbps	Mix-zone advertisement TX interval (γ_{mz})	0.5s, 1s
Sensitivity	-89dBm	Mix-zone transmission range	100 meter
Thermal noise	-110dBm	Number of eavesdropper	25
Area size	15 KM $\times$ 15 KM	Eavesdropping range	250 meter
Average trip duration	692.81s	Percentage of internal adversaries	10%-50%
Number of trips	287,939	CF distribution bandwidth ($\mathbb{B}$)	50 KB/sec
Number of vehicles	138,259	CF TX interval	1s

Comparison:

- Cryptographic Mix-Zone (CMIX) [**?**] [Win-ITS'07]
- Chaff-based CMIX [**?**] [VNC'18]

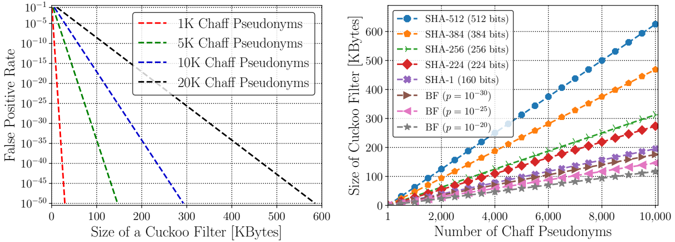
Bloom Filter (BF) and Cuckoo Filter (CF): Construction & Membership Checks



BF/CF features:

- A space-efficient probabilistic data structure
- Fast membership checking
- No false negatives, but false positive matches are possible
- A query returns either *"possibly in set"* or *"definitely not in set"*
- No deletion is allowed in a BF; but CF supports deletion.

Quantitative Analysis

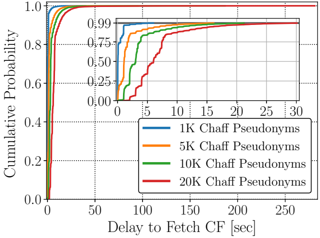


(a) Baseline scheme **(b)** Vehicle-centric scheme

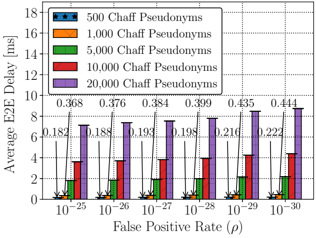
Figure: (a) The size of a CF as a factor of false positive rate. (b) The size of a CF as a factor of chaff pseudonyms numbers.

- For 5,000 chaff pseudonyms with $\rho = 10^{-30}$, the CF size is 87.75 KB.
- By employing SHA-256, the size of a fingerprint for 5,000 chaff pseudonyms becomes 156 KB; while by employing a CF, the size would be 73.13 KB ($\rho = 10^{-25}$).

Quantitative Analysis (cont'd)



(a) Communication Latency



(b) Computation Latency

Figure: (a) Evaluation of end-to-end delay to broadcast CF of chaff pseudonyms to vehicles approaching mix-zones ($\rho = 10^{-30}$, $\mathbb{B} = 50KB/s$).
(b) Computation overhead to validate a chaff pseudonym.

- With 1K chaff pseudonyms, 99% of the vehicles received a CF in 5 sec.
- The latency to validate 1,000 membership check chaff pseudonym with 1K pseudonyms in a CF ($\rho = 10^{-25}$) is ≈0.368 ms, i.e., the average latency to validate one chaff pseudonym is 0.000368 ms.

Quantitative Analysis (cont'd)

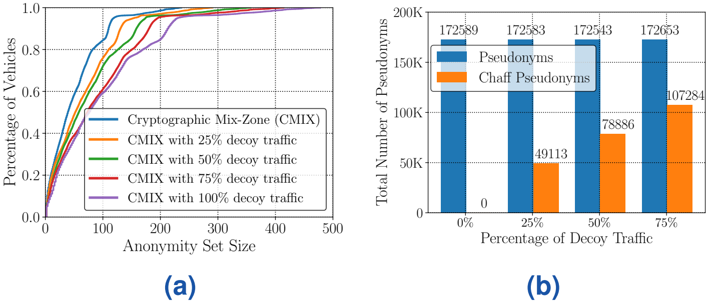
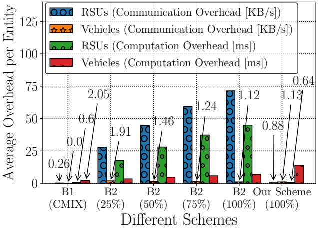
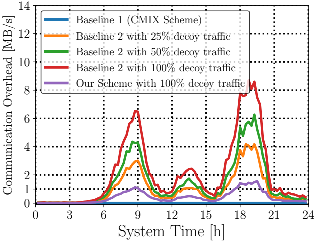


Figure: (a) CDF of anonymity set size for CMIX and our scheme. (b) Total number of disseminated pseudonyms and chaff pseudonyms ($\gamma_v = 0.5s$).

Quantitative Analysis (cont'd)



(a) Overhead Comparison



(b) Communication Comparison

Figure: Comparison among CMIX (B1) [**?**], chaff-based CMIX (B2) [**?**], and our scheme: 1,000 chaff pseudonyms in a CF with $\rho = 10^{-25}$; beacon frequency: $\gamma_{mz} = 0.5$, $\gamma_v = 0.2$. (a) Computation and communication overhead comparison. (b) Communication overhead comparison, averaged every 300s.

Protocol 1 Syntactic and Semantic Linking Algorithm

procedure TRACKINGVEHICLES( )
    Classify eavesdropped beacons based on vehicle length
    Create a list with the first & last seen beacons for each identifier
    Filter out trivially linked pseudonyms (not changing psnyms)
    Latency $\leftarrow$ Estimated time to traverse a Mix-zone
    for Each B_i in BEACON_SET **do**
        B_i^f is the first seen message for beacon B_i
        B_i^l is the last seen message for beacon B_i
        for Each B_{i+1}^f in BEACON_SET **do**
            diff_time $\leftarrow$ time difference between B_{i+1}^l and B_i^f
            if diff_time ≥ 0 && diff_time $\leq$ *Latency* **then**
                if pseudo-id for B_i^l and B_{i+1}^f not seen together **then**
                    if exists a road path from B_i^l to B_{i+1}^f **then**
                        if path $B_i^l \mapsto B_{i+1}^f$ is validated by Kalman Filter (KF) **then**
                            B_i^l and B_{i+1}^f are correlated
                        else
                            B_i^l and B_{i+1}^f are not correlated
                        end if
                    end if
                end if
            end if
        end for
    end for
end procedure

Syntactic and Semantic Linking Algorithm

In order to link two pseudonyms:

- An adversary places wireless receivers near each mix-zone (entry and exit points)
- An adversary tries to link one of the last seen beacon before entering a mix-zone to one of the first-seen beacon exiting the mix-zone
- Filtering out trivially linked pseudonyms
- Estimated time to traverse a mix-zone
- The two pseudonyms have not been seen together
- Considering the physical road layout (exists a path between the two)
- The second beacon (direction) is from an exit points of the mix-zone

Quantitative Analysis (cont'd)

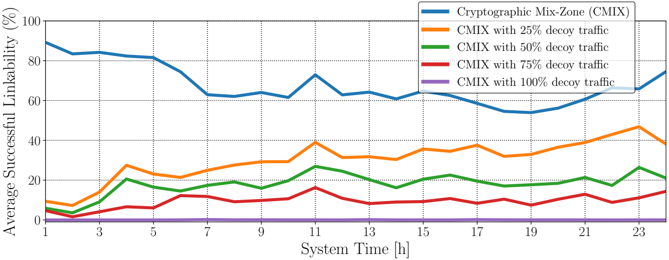
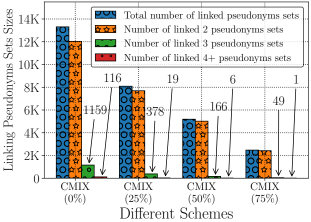


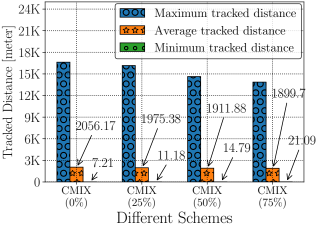
Figure: Average successful linkability comparison with the CMIX baseline scheme through conducting syntactic and semantic linking attacks.

- The probability of linking decreases when the traffic density increases.
- For the baseline scheme, one could link pseudonyms with high probability success rate.
- By introducing decoy traffic for 50% of vehicles, the probability of linking drops from 63% to 17% at system time 7.

Quantitative Analysis (cont'd)



(a)

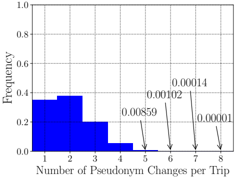


(b)

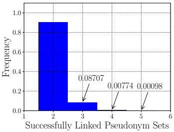
Figure: (a) Linking pseudonym sets for the baseline and our scheme. (b) Successful tracked distance for the baseline and our scheme.

- Successfully linked pseudonyms set size is the number of pseudonyms, linked by the eavesdroppers, corresponding to the same vehicle.
- The higher the percentage of decoy traffic is, the lower the number of linked pseudonyms sets becomes.

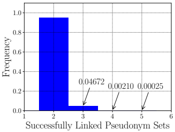
Quantitative Analysis (cont'd)



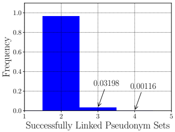
(a) Pseudonym change



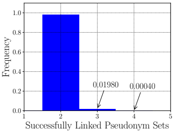
(b) CMIX: 0% decoy traffic



(c) CMIX: 25% decoy traffic



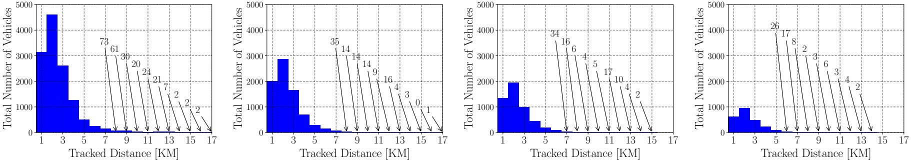
(d) CMIX: 50% decoy traffic



(e) CMIX: 75% decoy traffic

Figure: (a) Histogram of pseudonyms changes. (b) Histogram of successfully linked pseudonym sets for the baseline scheme (b), and our scheme (c-e).

Quantitative Analysis (cont'd)

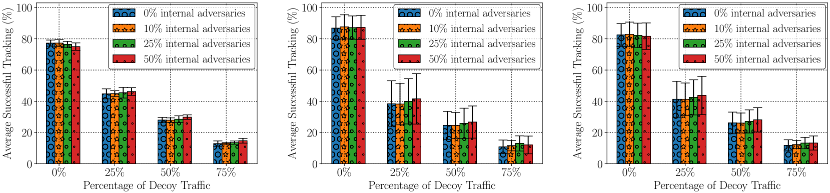


(a) CMIX **(b)** CMIX: 25% decoy traffic **(c)** CMIX: 50% decoy traffic **(d)** CMIX: 75% decoy traffic

Figure: Histogram of tracked distances by eavesdroppers based on the linked pseudonyms sets for the baseline scheme (CMIX) and our scheme.

- By introducing decoy traffic for vehicles exiting the mix-zones, the total number of vehicles, tracked by the eavesdroppers, drastically decreases.

Quantitative Analysis (cont'd)



(a) During Rush Hours **(b)** During Non-rush Hours **(c)** During 24 Hours

Figure: Average successful linkability in the presence of non-cooperative vehicles, not changing their pseudonyms while crossing the mix-zones.

- Non-cooperative vehicles exit the mix-zone without changing pseudonyms; also, if chosen to be relaying vehicles, do not disseminate decoy traffic.
- Selection of such vehicles is independent of selection of relaying vehicles; in each scenario, different sets are selected to be non-cooperative.
- The average successful tracking is not considerably affected in the presence of non-cooperative vehicles.

Conclusions

- A novel scheme to protect user privacy regardless of the geometry of the mix-zones, mobility patterns, vehicle density, and arrival rates.
- Enhancing user privacy protection at the cost of low computation and communication overhead.
- Ensuring the operation of safety applications by the dissemination of decoy traffic.
- Our results show that the deployment of mix-zones can be cost-effective.

Future Works

- Investigating the resiliency of our scheme against a fraction of malicious vehicles or compromised RSUs that covertly send the CMIX symmetric key or the CFs to other (internal or external) adversaries.

- Extending our tracking algorithm towards tracking vehicles based on the physical properties of the wireless radio signals and investigate appropriate countermeasures to mitigate such a vulnerability.

Cooperative Location Privacy in Vehicular Networks: Why Simple Mix-zones are not Enough

Mohammad Khodaei and Panos Papadimitratos
Networked Systems Security Group (NSS)
www.eecs.kth.se/nss

In IEEE Internet of Things (IoT) Journal