CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Antitracking de vehicle connectat mitjançant Intel·ligència Artificial

Sergi Sánchez Deutsch
Sergi Mercadé Laborda
Josep Escrig Escrig
Àngel Martín

i2cat

Cita suggerida. CIDAI-POC-2022-04 // Antitracking de vehicle connectat mitjançant Intel·ligència Artificial per Sergi Sánchez Deutsch, Sergi Mercadé Laborda, Josep Escrig Escrig
Àngel Martín - Fundació i2cat, CIDAI, 2022. Llicencia: CC BY-NC-SA 4.0

1. Resum

1.1. Objectius

El sector automobilístic ha evolucionat enormement en els darrers temps per aconseguir vehicles cada cop més intel·ligents gràcies a la connectivitat i la intel·ligència artificial. Seguint aquesta tendència trobem el concepte de cotxe connectat, que és aquell capaç de comunicar-se amb altres cotxes i elements de la via per on circula.

Dins d'aquest context es troben les connexions V2X (Vehicle-to-Everything), que comprenen totes aquelles connexions sense fils entre un vehicle connectat i qualsevol element que afecti o pugui ser afectat per la mobilitat d'aquest vehicle. És un tipus de connexió que n'engloba d'altres més específiques, com poden ser les de vehicle a vehicle (V2V), vehicle a vianant (V2P) o vehicle a infraestructura (V2I). Amb tot això s'aconsegueix tenir una informació precisa en temps real sobre les condicions de tots els elements involucrats en el trànsit de vehicles connectats, des de les pròpies posicions, velocitats, etc. dels mateixos fins a l'estat de semàfors o carreteres.

Els missatges V2X que s'envien utilitzant aquestes connexions han de ser oberts, de forma que tots els elements que intervenen en el trànsit el puguin rebre i interpretar. Aquest fet planteja una sèrie de reptes de ciberseguretat, ja que cal garantir en tot moment la l'anonimat de les dades que es transmeten.

L'objectiu d'aquest projecte consisteix en minimitzar la probabilitat que un atacant extern pugui fer un seguiment de la posició d'un vehicle capturant els seus missatges V2X.

La feina presentada en aquest document s'ha dut a terme dins del projecte H2020 CARAMEL [1]. Aquest projecte ha rebut finançament del programa Horizon 2020 de recerca i innovació de la Unió Europea sota l'acord de subvenció nº 833611.

1.2. Visió

El hacking de l'automòbil connectat és una part negativa de la digitalització dels sistemes de mobilitat. El que agreuja l'amenaça de la ciberseguretat de l'automòbil és la gran quantitat d'elements susceptibles de ser atacats.

És per això que es considera que el punt de partida hauria de ser la "seguretat des del disseny". Els fabricants d'automòbils han d'incorporar la seguretat des del principi, en lloc de corregir les vulnerabilitats a mesura que són detectades.

1.3. Resum del problema

Els missatges V2X que envien els vehicles contenen dades referents, entre d'altres, a la seva posició i velocitat en un instant de temps determinat. Cada missatge va signat amb el que es coneix com a Authorization Ticket (AT), que identifica inequívocament el vehicle que l'envia. El fet que qualsevol cotxe connectat o element de la via hagi de poder rebre i interpretar en temps real els missatges V2X fa que aquest AT no pugui ser encriptat per preservar-ne la identitat. Això implica que un agent extern maliciós que aconsegueixi capturar els missatges V2X que s'envien en una zona determinada pot aconseguir, mitjançant l'AT, identificar un vehicle concret i traçar-ne la seva posició al llarg del temps, fet que suposa un problema de privacitat.

1.4. Solució

Un vehicle connectat no signa els seus missatges amb un sol AT al llarg de tota la seva trajectòria, si no que pot sol·licitar canviar-lo diverses vegades en un període de temps determinat. La solució per minimitzar la traçabilitat passa per determinar l'instant de temps òptim per fer aquest canvi, de manera que sigui més difícil establir una relació entre missatges consecutius d'un mateix vehicle amb diferent AT. Amb aquest objectiu es proposa un algoritme d'intel·ligència artificial que calculi en temps real aquest instant de temps òptim.

1 https://www.h2020caramel.eu/

2. Descripció del problema que resol

Cada missatge V2X enviat per un vehicle connectat ha d'anar signat amb un Ticket d'Autorització (AT), que identifica un vehicle determinat de la resta d'elements connectats del seu entorn. L'AT és anònim i no revela cap informació sobre el conductor, de forma que un atacant extern (sovint anomenat spoofer) que capturi aquests missatges no pot conèixer directament cap patró que identifiqui a una persona en concret. No obstant, l'atacant pot arribar a traçar la trajectòria d'un vehicle que envia missatges V2X amb el mateix AT i, complementant aquesta informació amb dades obtingudes per altres vies, identificar el conductor. Així doncs podria descobrir, per exemple, que la posició d'inici del trajecte correspon al domicili de l'usuari i la posició final correspon al seu lloc de treball.

Per intentar minimitzar aquesta problemàtica s'assignen periòdicament grups d'ATs a cada vehicle, de forma que es pot anar canviant al llarg de la trajectòria i trencar la relació directa amb missatges anteriors. Aquesta no és una solució definitiva, ja que tot i així l'atacant podria arribar a reconstruir la trajectòria d'una vehicle que ha signat missatges amb diferents ATs. Això és degut a que cada vehicle envia un missatge V2X en períodes que van dels 100 als 1000 milisegons, de forma que la diferència de posició entre dos missatges consecutius pot ser de pocs metres, fet que facilita la relació entre ambdós.

3. Implementació de la solució

3.1. Arquitectura, tecnologia i dades utilitzades

3.1.1. Introducció

En aquesta secció es detalla el funcionament del "Planificador d'AT", l'objectiu del qual és decidir el millor moment per a canviar el Ticket d'Autorització d'un determinat vehicle per a minimitzar el risc de traçabilitat. Aquesta decisió es fa avaluant com de fàcil és seguir la trajectòria del cotxe a partir de la informació continguda en els seus missatges V2X, així com els missatges dels vehicles del seu voltant. El Planificador d'AT està dissenyat per funcionar en un dispositiu de baix consum col·locat dins del vehicle en qüestió.

En termes generals, el Planificador d'AT emmagatzema en un buffer els últims missatges V2X que rep dels cotxes del seu voltant, així com els que envia el propi vehicle. Cada cop que cal enviar un nou missatge, un algoritme d'intel·ligència artificial intenta associar aquest nou missatge amb algun dels missatges antics del buffer. Si aquesta associació és correcta (és a dir, ambdós missatges corresponen al mateix vehicle) i la certesa de la decisió és alta, s'entén que és fàcil fer un seguiment de la trajectòria. d'aquesta forma, és possible calcular periòdicament aquesta facilitat.

El Planificador d'AT està compost principalment per tres mòduls, els quals estan explicats amb més detall a la secció 3.1.2:

- Un seleccionador de missatges candidats, encarregat de seleccionar un grup amb els N missatges del buffer que és més probable que corresponguin al vehicle objectiu.
- Un rastrejador de missatges, que associa el nou missatge del cotxe objectiu amb algun dels missatges candidats amb una certa puntuació.
- Un algoritme de decisió de canvi d'AT, l'objectiu del qual és escollir el millor moment per fer el canvi d'AT en funció d'un buffer de puntuacions, el número d'ATs disponibles i el temps restant fins que el vehicle pugui obtenir un nou paquet d'ATs.

La Figura 1 mostra l'estructura general i com interaccionen aquests tres mòduls.

Figura 1 - Arquitectura general del sistema planificador d'AT.

3.1.2. Mòduls del Planificador d'AT

Seleccionador de Missatges Candidats

El mòdul de selecció de candidats és el responsable d'escollir un grup de missatges passats (candidats) els quals és més probable que corresponguin al vehicle objectiu. Per fer-ho, guarda en un buffer els últims M missatges que s'han rebut dels vehicles del voltant així com els que ha enviat el propi cotxe objectiu. Quan el vehicle objectiu ha d'enviar un nou missatge es calcula, per a cadascun dels missatges del buffer, una sèrie de paràmetres que poden ser útils per associar-los al missatge objectiu. Aquests pa-

ràmetres es calculen fent servir la informació dels propis missatges V2x, i inclouen:

- La diferència entre l'instant de temps t del missatge objectiu i l'instant de temps del missatge candidat.
- La diferència entre la posició del missatge objectiu i la posició que tindria el missatge candidat a l'instant t si mantingués una velocitat constant.
- La variació entre la velocitat del missatge objectiu i la velocitat que tindria el missatge candidat a l'instant de temps t si estigués a la posició del missatge objectiu.

Finalment, els candidats s'ordenen per aquests valors calculats i se seleccionen els N que tenen les menors variacions.

Rastrejador de Missatges

Donat un grup de candidats i les seves variacions de temps, posició i velocitat calculades pel seleccionador, el rastrejador de missatges funciona com un algoritme de regressió que escolleix a quin dels candidats és més probable que estigui associat un missatge objectiu.

El mètode proposat per implementar aquest rastrejador és un Random Forest, el qual és un mètode d'aprenentatge conjunt (ensemble learning en anglès) format per un determinat número d'arbres de decisió. Cadascun d'aquests arbres avalua un subgrup dels paràmetres calculats pel seleccionador de candidats amb un cert llindar per decidir si un missatge candidat correspon al mateix vehicle que el missatge objectiu. Per a cada candidat, cada arbre de decisió genera un valor binari y€{0,1}. Finalment, el Random Forest genera una puntuació Y€[0,1] que representa la mitja de les prediccions de tots els arbres. Computacionalment parlant, els Random Forests necessiten menys recursos que altres sistemes de classificació d'intel·ligència artificial com les Xarxes Neuronals.

Algoritme de Decisió de Canvi d'AT

Per a prendre la decisió de canviar o no el Ticket d'Autorització en un instant determinat, es fa servir un problema matemàtic conegut com a Parada Òptima (Optimal Stopping en anglès). Aquest problema descriu a un agent que observa un sèrie de valors d'entrada que evolucionen en el temps de forma més o menys aleatòria i decideix quin és el millor moment per a dur a terme una acció a partir d'aquests valors. També es defineix un temps límit, ja que en cas contrari s'estaria observant els valors indefinidament. Dins del context que ens ocupa, aquest agent seria el Planificador d'AT i l'acció seria canviar l'AT.

El número de vegades que es pot prendre la decisió de canviar l'AT no és indefinit, si no que depèn del número d'AT diferents dels que disposa el vehicle i el temps restant fins a que en pugui aconseguir de nous. Per a distribuir de manera equitativa els ATs disponibles, s'estableix un temps límit per a fer el canvi d'AT utilitzant la següent equació:

$$Temps\ Límit = \frac{Temps\ fins\ nous\ ATs}{\#AT\ restants}$$

El problema de la Parada Òptima demostra que es pot prendre la millor decisió possible amb una probabilitat major al 36%. L'estratègia a seguir consisteix en dedicar el 37% del temps disponible (Temps Límit) només a observar les puntuacions que arriben des del mòdul rastrejador de missatges i guardar el valor màxim. Un cop esgotat el 37% del temps, es pren la decisió de canviar l'AT quan es rep una puntuació superior a la puntuació màxima registrada durant el temps d'observació. Si, en cas contrari, no es rep cap valor superior, el sistema força un canvi d'AT quan s'esgota el temps límit.

3.1.3. Dades

Per a entrenar el Random Forest (del mòdul rastrejador de missatges) s'ha utilitzat un dataset de missatges V2X sintètics generats pel simulador SUMO i publicat per Uppor et al (2014). Aquest dataset (anomenat Cologne Dataset) representa el tràfic d'un dia laborable habitual a la ciutat de Colònia i cobreix una extensió de 400 quilòmetres quadrats durant un període de 24 hores. Conté dades sobre 700.000 trajectes diferents.

Cada missatge del dataset conté els següents camps:

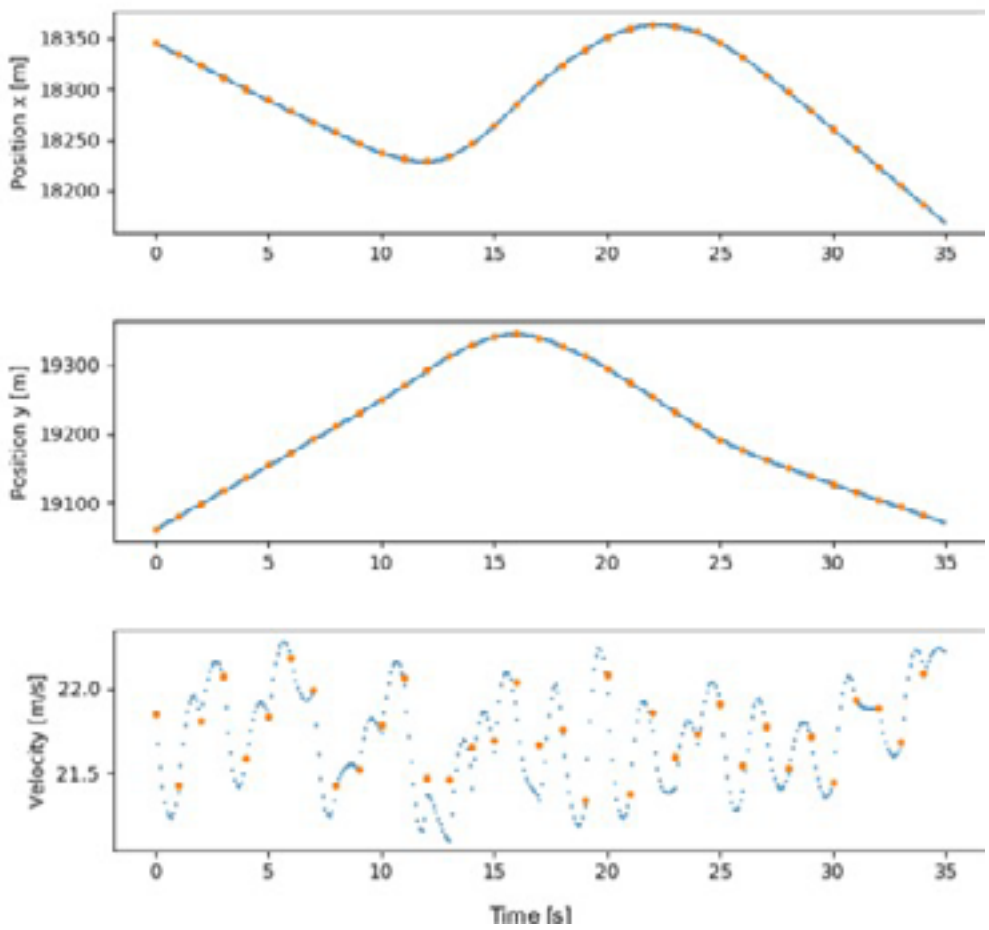
- id: identifica el vehicle que envia el missatge.
- time [s]: instant de temps del missatge.
- x [m]: component x de la posició del vehicle.
- y [m]: component y de la posició del vehicle.
- v [m/s]: velocitat del vehicle.

Tot i que el dataset és extens, per aquest projecte s'ha preferit realitzar un augment de les dades originals. Això és degut, per exemple, a que en el dataset original els missatges s'actualitzen cada segon, mentre que la freqüència d'enviament dels missatges V2X reals oscil·la entre els 10 i els 100 ms. Els missatges reals també contenen informació extra que no apareix en el dataset de Colònia, com per exemples les components x i y de la velocitat i l'acceleració. Amb aquest propòsit, s'han dut a terme dues operacions:

- S'han calculat les components v_x i v_y de la velocitat original.
- S'ha augmentat freqüència d'enviament dels missatges i s'ha afegit un terme aleatori per a tenir missatges cada 100 ± 50 ms
- S'ha fet una interpolació de les posicions i velocitats de cada vehicle per als nous instants de temps creats.

La Figura 2 mostra el resultat d'aquesta interpolació per a les variables de posició i velocitat.

Figura 2 - Resultat de la interpolació (en taronja) del dataset original (en blau) per a les components x i y de la posició i el mòdul de la velocitat.



3.2. Reptes resolts i resultats obtinguts

El sistema Planificador d'AT proposat en aquest document, introdueix dues solucions innovadores:

- Un sistema d'Intel·ligència artificial disenyat i entrenat per a seguir la trajectòria d'un vehicle determinat a partir de la informació dels missatges V2X que envia. L'entrenament s'ha dut a terme utilitzant un extens set de dades sintètiques que simulen missatges reals. Aquest sistema es diferencia d'altres sistemes més tradicionals basats en regles (Rule-based en anglès).
- Un algoritme de decisió de canvi d'AT, el qual implementa una solució matemàtica per a identificar el millor moment per a fer aquest canvi. Mentre una solució convencional té un interval fix per a fer aquest canvi, l'algoritme proposat fa una planificació dinàmica intentant disminuir la facilitat de que el vehicle sigui rastrejat. D'aquesta forma, es tenen en compte no només els Ats disponibles del vehicle, si no l'estat del tràfic que l'envolta.

En les proves realitzades utilitzant les dades de la secció 3.1.3 s'ha observat que el mòdul rastrejador de missatges aconsegueix, de mitja, seguir correctament la trajectòria del 99,8% dels missatges objectiu que li arriben. El fet que entre dos missatges consecutius la posició i velocitat del vehicle tingui variacions molt petites facilita una associació clara. Això provoca que, a la pràctica, l'algoritme de decisió de canvi d'AT força el canvi quan s'esgota el 37% del temps límit, ja que quasi totes les puntuacions que arriben tenen valors similars.

3.3. Limitacions actuals de la tecnologia

El principal factor limitant del sistema proposat és la capacitat de computació del dispositiu que l'executa. El Planificador d'AT ha de poder funcionar en un dispositiu de baix consum incorporat dins del propi vehicle connectat per tal de poder rebre missatges de l'entorn. A més a més, ha de poder funcionar en paral·lel amb altres processos relacionats amb la seguretat del vehicle, per la qual cosa tampoc es disposa del 100% de la capacitat de processament del dispositiu.

En aquest context, s'han realitzat proves executant el sistema en una Raspberry Pi v3 i utilitzant només un nucli de la CPU. El resultat és que es poden processar aproximadament uns 30 missatges per segon, suposant que el vehicle envia missatges cada 100 ± 50 ms. En un entorn real, un vehicle V2X podria arribar a rebre un màxim teòric d'uns 1200 missatges/s abans de que el canal de recepció es saturi i es perdin missatges. Caldria, per tant, aplicar alguna de les següents solucions:

- Que el Planificador d'AT ignori alguns dels missatges que arriben de vehicles externs.
- Disminuir el temps d'inferència de l'algoritme d'intel·ligència artificial (el Random Forest).
- Executar el Planificador d'AT en un dispositiu amb més capacitat de processament.

4. Potencial impacte de la solució

La solució presentada ataca un dels molts problemes de seguretat que presenta l'entorn dels vehicles connectats, i pot ser útil a l'hora d'implementar nous protocols V2X. No obstant això, l'algoritme es podria fer funcionar per a molts entorns diferents al V2X que impliquin dades amb informació rellevant sobre la posició d'un usuari determinat.

Índex

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

C/Bilbao, 72 Edif. A. 08005 - Barcelona
Tel. +34 93 7419 100
info@cidai-catalonia-ai.eu
www.cidai.eu

Generalitat de Catalunya
Departament de la Vicepresidència i de Polítiques Digitals i Territori

Ajuntament de Barcelona
Barcelona Activa

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

CVC
Computer Vision Center

eurecat
Centre Tecnològic de Catalunya

i2cat

Microsoft

NTT DATA

SAP

sdg group

UPC Universitat Politècnica de Catalunya Barcelonatech
Intelligent Data Science and Artificial Intelligence Research Center

IDEAI