PINO: Privacidad en Incidencias Notificadas On-Chain

Carles Anglés-Tafalla,
Jordi Castellà-Roca y Alexandre Viejo
Universitat Rovira i Virgili,
Dept. d'Eng Inf i Mat, CYBERCAT,
43007 Tarragona, Spain
{jordi.castella, carles.angles,
alexandre.viejo}@urv.cat

Jordi Soria-Comas
Universitat Oberta de Catalunya,
Est. d'Inf, Mult i Telecom,
UOC-TECH Research Centre,
CYBERCAT,
08018, Barcelona, Spain
jordi_sc@uoc.edu

Guillermo Navarro-Arribas
y Julián Salas
Universitat Autonoma de Barcelona,
Dept. d'Eng. de la Inf i les Com,
CYBERCAT,
08193 Bellaterra, Catalonia, Spain
{guillermo.navarro, julian.salas}@uab.cat

***Resumen*—El *vehicular crowdsourcing* se ha consolidado como un paradigma clave para la detección y reporte colaborativo de incidencias, generando datos compartidos geolocalizados sobre el estado de la red vial. Sin embargo, los enfoques tradicionales basados en plataformas centralizadas plantean riesgos significativos para la privacidad y la soberanía de los datos. La aparición de modelos descentralizados basados en *blockchain* ha mitigado la dependencia en entidades de confianza, pero introducen nuevos desafíos en seguridad y privacidad al hacer públicas y potencialmente trazables las contribuciones de los vehículos. Este trabajo propone un sistema de reporte de incidencias vehiculares descentralizado y respetuoso con la privacidad, combinando *Blockchain*, *Zero-Knowledge Proofs* y *differential privacy* adaptativa. El sistema garantiza autenticidad y trazabilidad sin comprometer la identidad de la persona participante, ajustando dinámicamente el nivel de perturbación de su localización. Los experimentos muestran que es posible mantener altos niveles de privacidad con una precisión aceptable en la localización agregada, lo que supone una solución prometedora para entornos vehiculares seguros y privados.**

***Index Terms*—Privacidad diferencial, Redes Vehiculares, Blockchain, Vehicular Crowdsourcing.**

I. Introducción

En la última década, el concepto de *vehicular crowdsourcing* ha emergido como un enfoque que aprovecha los datos generados por múltiples vehículos conectados para mejorar la percepción colectiva del entorno vial. En este paradigma, los automóviles actúan como nodos inteligentes capaces de recopilar, compartir o reportar información sobre accidentes, congestión o peligros en tiempo real, contribuyendo de manera activa a la generación de conocimiento compartido sobre el estado de la red vial.

Este modelo de colaboración vehicular ha inspirado la aparición de plataformas basadas en crowdsourcing como Waze[1], que permiten a los conductores compartir información instantánea sobre condiciones de tráfico y eventos en la carretera. A través de simples interacciones con la aplicación, las personas participantes pueden reportar incidentes como accidentes, atascos, obras o vehículos detenidos, generando datos que se agregan y procesan para ofrecer una visión actualizada del estado del tráfico. Más allá de las plataformas comunitarias, fabricantes de automóviles, aseguradoras y autoridades de tráfico han adoptado modelos similares mediante la transmisión periódica de datos vehiculares hacia bases de datos en la nube gestionadas de forma centralizada por terceros [1].

La recolección masiva de datos por parte de estas plataformas ha demostrado ser valiosa en aplicaciones de reporte de incidentes, mantenimiento de infraestructuras viales, análisis estadístico y forense, así como en la gestión del tráfico [2]. El uso y procesado de estos datos facilita la toma de decisiones tanto de conductores como de autoridades y servicios de emergencia, contribuyendo a una gestión más eficiente de la movilidad urbana.

A pesar de su enorme potencial, la aplicación de *vehicular crowdsourcing* introduce riesgos significativos para la privacidad y la protección de datos personales. La recopilación continua de información geolocalizada y telemétrica puede permitir inferir patrones de movilidad, trayectorias frecuentes o lugares sensibles asociados a conductores específicos [3]. Estos riesgos se intensifican en ecosistemas centralizados, donde los datos se concentran bajo el control de una sola entidad, aumentando la vulnerabilidad a accesos no autorizados, correlaciones de datos o usos secundarios no autorizados inicialmente.

Ante estas limitaciones, los modelos descentralizados basados en *blockchain* han surgido como una alternativa prometedora para mitigar la dependencia hacia entidades centrales en los sistemas de reporte de incidentes. Sin embargo, el carácter público y distribuido de esta tecnología introduce nuevos desafíos relacionados con la seguridad y privacidad: las transacciones almacenadas *on-chain* pueden permitir el rastreo sistemático de los vehículos, comprometiendo su privacidad; además, participantes deshonestos podrían intentar aprovecharse de un sistema sin una autoridad central para obtener ventaja o perjudicar a otros participantes.

Teniendo en cuenta estos desafíos, es necesario desarrollar nuevas soluciones que combinen las garantías de seguridad y descentralización propias de la tecnología *blockchain* con técnicas avanzadas de protección de la privacidad.

I-A. Estado del Arte

La preservación de la privacidad en *Location-Based Services* (*LBS*) respaldados con *vehicular crowdsourcing* ha despertado un creciente interés en los últimos años. En este ámbito, las propuestas en la literatura reciente suelen agruparse en *centralizadas* o *descentralizadas*, según su arquitectura y el grado de dependencia hacia una entidad central.

[1]Waze — https://www.waze.com

En la última década, múltiples trabajos han explorado distintas vías con la intención de evitar que proveedores de servicios centralizados accedan directamente a datos geolocalizados de las personas participantes. En [4] se propone un método colaborativo basado en *Multi-Party Computation* (*MPC*) en el que los vehículos calculan de manera conjunta su posición para LBS, evitando revelar sus coordenadas reales al servidor. En un escenario similar, PriTAEC [5] emplea *Ciphertext-Policy Attribute-Based Encryption* (*CP-ABE*) y *Oblivious Transfers* (*OT*) para ocultar a la nube las ubicaciones de los vehículos que participan en redes VANETs. En [6], [7] combinan *Homomorphic Encryption* (*HE*) con *OT* para que los vehículos soliciten información sobre incidencias viales sin revelar su ubicación exacta a terceros. En PRIS [8], lpRide [9] y PSRide [10], en cambio, aplican *Additive HE* para encontrar coincidencias en los trayectos de conductores y pasajeros sin exponer sus puntos de origen o destino al servidor central.

Aunque estos sistemas resultan efectivos para proteger datos sensibles, presentan elevados costes derivados del uso de herramientas como HE o MPC. Por otra parte, su carácter centralizado supone un único punto de fallo o ataque, además de un riesgo de confianza por parte de la entidad supervisora.

Estas limitaciones han motivado la aparición de *alternativas descentralizadas* que prescinden de entidades de confianza, distribuyendo sus operaciones entre vehículos o nodos distribuidos. Usando *MPC*, los sistemas propuestos en [11] y [12] usan protocolos de *Private Set Intersection* combinados con *HE* para que conductores y pasajeros detecten coincidencias en sus trayectos sin revelar sus ubicaciones ni depender de terceros. Aprovechando el modelo distribuido del *edge computing*, VPTS [13] utiliza las RSUs para añadir ruido, por medio de *HE* y *Differential Privacy*, a los datos agregados y garantizar la confidencialidad de los patrones de movimiento en intersecciones urbanas. Tecnologías como el Blockchain también se han incorporado en *vehicular crowdsourcing*. En [14] se presenta *PriSC*, donde los pasajeros cifran sus políticas de localización mediante *Additive HE*, y los vehículos envían sus ubicaciones cifradas a través de RSUs a una Blockchain, que verifica la elegibilidad de los participantes de forma anónima. En [15] se presenta BSDP, un framework basado en blockchain dedicado a monitorizar el tráfico colindante a zonas de parking. Las RSUs ubicadas en el *edge* usan *HE* con *Threshold Decryption* cifrar, agregar y publicar en el blockchain las ubicaciones de los vehículos participantes de forma privada.

Si bien los sistemas descentralizados mejoran la soberanía de los datos y reducen la dependencia en entidades centrales, también presentan desafíos relevantes. La distribución del cómputo entre nodos o vehículos provoca una necesidad de consenso o coordinación que puede introducir latencias adicionales. Además, ante la premisa de privacidad, muchas de estas propuestas recurren a las mismas técnicas criptográficas de agregación o cifrado homomórfico empleadas en los sistemas centralizados, heredando en el proceso sus elevados costes computacionales y de comunicación.

I-B. Contribuciones y plan del artículo

De acuerdo con las limitaciones identificadas en los sistemas de *vehicular crowdsourcing* en entornos descentralizados, este trabajo presenta una propuesta que combina blockchain, *Zero-Knowledge Proofs* y un mecanismo adaptativo basado en *Differential Privacy* para ofrecer un sistema de reporte de incidentes vehiculares descentralizado, privado y seguro.

Las principales contribuciones de este trabajo engloban los siguientes puntos:

- *Privacidad de los testigos:* Los vehículos se autentican usando credenciales verificables, garantizando que cada reporte proviene de un actor legítimo del sistema y permitiendo verificar la autenticidad de los reportes sin revelar la identidad de los conductores involucrados.
- *Privacidad adaptativa basada en densidad de reportes:* El sistema introduce un mecanismo adaptativo que ajusta el nivel de ruido aplicado a las localizaciones en función del número de reportes recibidos. De este modo, los incidentes inicialmente reportados se publican con alta perturbación espacial, acotándose progresivamente conforme más vehículos validan el mismo evento.
- *Reportes descentralizados:* El sistema permite a los vehículos reportar y publicar incidencias de forma autónoma mediante *smart contracts*, aprovechando la inmutabilidad y transparencia de la tecnología *blockchain* para garantizar la trazabilidad y verificabilidad de los eventos, sin depender de una entidad central.

El resto del artículo se estructura de la siguiente manera: la Sección II formaliza los actores y protocolos que sustentan la propuesta; en la Sección III se evalúan las garantías de privacidad y usabilidad de los datos; y finalmente, la Sección IV presenta las conclusiones y líneas futuras de trabajo.

I-C. Requisitos del sistema

Para estar acorde con las contribuciones que presenta nuestra propuesta se traducen en los siguientes detallados requisitos de Seguridad (S), Privacidad (P) y Funcionales (F):

S1) *Descentralización*: La gestión, integridad y almacenado de los reportes deben estar gobernadas por mecanismos distribuidos, evitando que una entidad central actúe como punto único de fallo o confianza.

S2) *Integridad de los reportes*: Las incidencias reportadas deben ser inmutables una vez almacenadas, garantizando que no pueden ser alteradas, eliminadas o falsificadas.

S3) *Autenticidad del vehículo reportante*: Solo los vehículos registrados por la autoridad A_R deben poder generar reportes válidos.

S4) *Resistencia a ataques Sybil*: Un atacante no debe poder crear múltiples identidades falsas para influir en la agregación de incidencias o saturar el sistema con reportes maliciosos.

P1) *No enlazabilidad*: No debe ser posible vincular los diversos reportes generados por un mismo vehículo.

P2) *Anonimato de la posición*: la ubicación desde la que se reporta una incidencia debe mantenerse protegida, y cada persona participante debe poder determinar el nivel de privacidad que desea asociar a dicha ubicación.

F1) *Utilidad*: La localización de los incidentes debe poder precisarse en función del número de veces que se reporte. Es decir, debe existir un mecanismo que permita mejorar progresivamente la precisión de la posición en aquellas incidencias que acumulen un mayor número de reportes.

II. Propuesta en detalle

En esta sección se detalla la propuesta para realizar el reporte seguro y la posterior agregación de incidencias vehiculares. En primer lugar, se describen los tres actores que intervienen en la propuesta, y a continuación se explican en detalle las distintas fases que conforman el funcionamiento del sistema:

- *Vehículo testigo (V):* Nodo participante que detecta un incidente y genera un reporte firmado y verificable. Cada vehículo dispone de una credencial verificable emitida por la autoridad de registro A_R y asociada a un par de claves criptográficas (sk_V, pk_V).
- *Autoridad de registro (A_R):* Entidad de confianza (por ejemplo, una autoridad de tráfico o un fabricante) encargada de registrar, emitir las credenciales verificables y mantener un árbol de Merkle que agrupa los identificadores públicos de los vehículos autorizados. Asimismo, A_R es responsable de desplegar y configurar el *smart contract* $RepSC$, que constituye el núcleo del sistema de gestión de reportes de incidencias.
- *Smart Contract de reporte ($RepSC$):* Componente lógico desplegado en la blockchain que almacena *on-chain* la raíz del árbol de Merkle R_V, y permite el registro y validación descentralizada de los incidentes vehiculares, eliminando la necesidad de una autoridad central o supervisión externa.

II-A. Registro del Vehículo

Este primer protocolo configura la unidad de a bordo (*OBU*) de los Vs con el fin de obtener las credenciales criptográficas para interactuar de forma segura con el resto de actores del sistema. Con este fin, la OBU genera un par de claves (sk_V, pk_V), donde la clave pública pk_V se genera a partir de una función hash segura y resistente a colisiones $\mathcal{H}(sk_V)$. Esta clave pública pk_V se usa como identificador para la credencial verificable asociada al vehículo V. A continuación, V se autentica a la A_R y, a través de un canal seguro (p.e. TLS), se realizan los siguientes pasos:

a) V envía su pk_V a A_R.
b) A_R calcula la regeneración n del árbol de Merkle a partir del conjunto de claves públicas pk_V de los vehículos registrados, generando la raíz R_V^n.
c) A_R publica la raíz R_V^n en el *smart contract* $RepSC$, de modo que sea accesible *on-chain* para posteriores verificaciones.
d) A_R envía a cada vehículo V su ruta del árbol de Merkle, denotada como $\mathrm{MerklePath}_V^n$, es decir, el conjunto de *hashes* hermanos del árbol que lo llevan hasta la raíz.
e) Con los datos recibidos, V construye su credencial verificable: $\mathsf{Cred}_V = \left(sk_V,\, n,\, \mathrm{MerklePath}_V^n\right)$.

Al finalizar el proceso, cada V implicado obtiene una credencial verificable única que puede ser empleada para demostrar, mediante *zero-knowledge proof*, su pertenencia al conjunto de vehículos registrados sin revelar su identidad.

II-B. Reporte descentralizado de incidentes

En el momento en que un vehículo V presencia un incidente, lo registra como la tupla:

$$r = (t, \tau, c, p = (\varphi, \lambda)) \tag{1}$$

donde $t \in \mathbb{T}$ es una marca de tiempo (con precisión horaria), $\tau \in \mathcal{T}$ es el tipo de incidente, $c \in \mathcal{C}$ la criticidad del incidente, y $p = (\varphi, \lambda)$ la posición. La posición se expresa en latitud y longitud, con el sistema de referencia WGS84 [16], $p = (\varphi, \lambda) \in \mathcal{L}$, tal que $\mathcal{L} \subseteq [-90, 90] \times [-180, 180]$.

Para procesar las localizaciones se utiliza una proyección al sistema métrico $\Pi : \mathcal{L} \to \mathbb{R}^2$, mediante UTM (*Universal Transverse Mercator*) dado que la extensión geográfica será relativamente acotada (p.e. a nivel de ciudad o región). Opcionalmente, V puede aportar datos adicionales (p.e. telemetría, imágenes o vídeo) del incidente, que se almacena de manera cifrada en un repositorio distribuido como IPFS, en cuyo caso el identificador de contenido (CID) se incluye en r.

Con estos datos, V genera una *non-interactive zero-knowledge proof* π_V^r, que acredita el conocimiento de los valores $(sk_V, \mathrm{MerklePath}_V^n, pk_V)$ sin revelarlos. Esta prueba garantiza las siguientes propiedades:

(I) $pk_V = \mathcal{H}(sk_V)$, i.e, V conoce la clave secreta que deriva en la clave pública registrada.
(II) $\mathsf{null}_{V,p,t}$ se ha generado a partir de $(sk_V, n, geohash(p), t)$, lo que permite identificar de forma única el evento y evitar reportes duplicados.
(III) pk_V es una hoja válida del árbol de Merkle, y el $\mathrm{MerklePath}_V^n$ reconstruye correctamente la raíz R_V^n publicada on-chain.

Una vez generada π_V, el vehículo V anonimiza su posición p aplicando una perturbación basada en *Differential Privacy*, calibrada de forma inversa a la criticidad c del evento, obteniendo así la posición perturbada $\tilde{x}$. Este proceso se detalla en la Sección II-B1.

Generados estos valores, V publica la incidencia en el Blockchain, invocando el método *submitReport()* del contrato $RepSC$ y enviando como parámetros:

$$(\tilde{x}, t, \tau, \pi_V, \mathsf{null}_{V,p,t}, [CID]) \tag{2}$$

La lógica interna de $RepSC$ realiza las siguientes operaciones para verificar la validez del incidente reportado:

a) Verifica on-chain la validez de π_V a partir de los parámetros recibidos y la raíz R_V^n, confirmando que V posee credenciales válidas, y que no ha enviado previamente un reporte correspondiente al mismo evento, identificado por $< geohash(\tilde{x}), t, \tau >$
b) Si la verificación es correcta, $RepSC$ almacena el incidente *on-chain*, junto con los siguientes atributos: $\tilde{x}, t, \tau, \pi_V, [CID]$.

II-B1. Anonimización de la posición: La posición que cada persona reporta en un incidente se protege mediante la técnica conocida como mecanismo de Laplace planar $\mathcal{M}_{\mathrm{geo}}$ que proporciona *geo-indistinguishability* [17].

Dada una ubicación p y su proyección $x = \Pi(p)$, se genera una localización perturbada $\tilde{x}$ según la siguiente densidad de probabilidad:

$$f_{\mathcal{M}_{\mathrm{geo}}(x)}(y) = \frac{\varepsilon^2}{2\pi} e^{-\varepsilon \|y - x\|_2}, \quad y \in \mathbb{R}^2.$$

El mecanismo $\mathcal{M}_{\mathrm{geo}}$ satisface ε-*geoindistinguibilidad*, es decir, para todas las ubicaciones $x, x' \in \mathbb{R}^2$ y todo conjunto medible, S se cumple:

$$\Pr[\mathcal{M}_{\text{geo}}(x) \in S] \leq e^{\varepsilon \|x-x'\|_2} \Pr[\mathcal{M}_{\text{geo}}(x') \in S].$$

Este mecanismo garantiza que la probabilidad de distinguir dos ubicaciones separadas una distancia $\|x - x'\|_2$ decrece exponencialmente con dicha distancia, proporcionando privacidad diferencial espacial con parámetro ε [17].

II-B2. Elección de ε: El sistema proporciona un marco para la elección de ε, fijando un valor máximo, y permitiendo que las personas participantes determinen el valor mínimo. Es importante tener en cuenta que este valor determinará la privacidad inicial de la localización, que se puede ver degradada al agregar el mismo incidente. Cada persona participante fijará el valor de su ε según la privacidad que desee obtener después de una estimación de su posible agregación. Dicha estimación puede ir variando durante el tiempo en función de la evolución de la densidad de tráfico o reporte de incidencias.

A continuación comentamos algunos aspectos a tener en cuenta para determinar el valor de ε.

- Número de vehículos en el sistema. Si en el sistema hay una estimación de n vehículos, una protección inicial de $\frac{\text{máx}\,\varepsilon}{n}$ fijaría una cota máxima de ε al proceso de agregación en el peor de los casos (recordamos que cada participante solamente puede reportar la misma incidencia una sola vez). Esta sería la opción más conservadora.
- Densidad estimada. Considerar todas las personas participantes del sistema puede ser una estimación poco realista. Alternativamente, podemos estimar la densidad de una localización basada en el historial de incidentes reportados. Podríamos determinar la estimación de densidad (KDE, *Kernel Density Estimation*) en función del espacio y tiempo, de manera que se aplique una ventana temporal o decaimiento a incidencias más antiguas.
- Ajuste al nivel de criticidad. La persona participante puede estar dispuesta a sacrificar su privacidad si está reportando un incidente con un nivel de criticidad muy alto. Es decir, puede determinar ε de forma proporcional al nivel de criticidad. De esta manera, obtendrá mayor privacidad al reportar incidentes poco críticos y menor privacidad en incidentes más críticos, donde la precisión de la posición puede ser de ayuda a servicios de emergencia.

II-C. Agregar incidentes

A la hora de agregar, hay que tener en cuenta que no se puede relacionar incidentes entre ellos a partir simplemente de la información reportada. En general, no se puede saber si dos incidentes reportados corresponden o no al mismo incidente real, pero sí que serán cercanos con cierto margen de error introducido por la anonimización.

Para agrupar los incidentes utilizamos un método de clustering basado en densidad relativa, como HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*) [18]. Al desconocer el ruido utilizado por las personas participantes es difícil establecer un umbral de distancia absoluta para agrupar localizaciones (como haríamos con DBSCAN), por lo que la distancia se determina a partir de las densidades locales de cada punto.

Dado un conjunto de localizaciones anonimizadas $\tilde{L} = \{\tilde{\ell}_1, \tilde{\ell}_2, \ldots, \tilde{\ell}_n\}$, tomamos ρ_i como la densidad local estimada en torno a $\tilde{\ell}_i$, por ejemplo, mediante el inverso del área del círculo que contiene a sus k-vecinos más cercanos:

$$\rho_i = \frac{k}{\pi\, d_k(\tilde{\ell}_i)^2},$$

donde $d_k(\tilde{\ell}_i)$ es la distancia entre $\tilde{\ell}_i$ y su k-ésimo vecino más cercano. Dos puntos $\tilde{\ell}_i$ y $\tilde{\ell}_j$ se consideran *densidad-conectados* si sus densidades son similares, es decir:

$$\frac{|\rho_i - \rho_j|}{\text{máx}(\rho_i, \rho_j)} \leq \delta,$$

para una tolerancia $\delta \in [0, 1)$. Los clusters se obtienen como las componentes conexas máximas del grafo inducido por esta relación de densidad-conectividad.

Dado un clúster $C = \{\tilde{\ell}_1, \ldots, \tilde{\ell}_k\}$ donde cada incidencia ha sido anonimizada con los valores de *epsilon* $\{\varepsilon_1, \ldots, \varepsilon_k\}$, la privacidad final de la localización se ve reducida por el teorema de composición secuencial de privacidad diferencial como:

$$\varepsilon_C = \sum_{i=1}^{k} \varepsilon_i \tag{3}$$

Esto proporciona una localización más precisa y permite situar la incidencia con mayor precisión.

Es importante remarcar que los incidentes se van reportando temporalmente, por lo que esta agrupación irá variando con el tiempo.

III. Evaluación

Como ejemplo ilustrativo hemos tomado 235 localizaciones del mismo tipo de incidencia reportadas en el área geográfica de la ciudad de Barcelona. Estos reportes se corresponden a 50 incidencias, algunas de ellas reportadas diversas veces, tal como muestra la Tabla I.

Incidencia	Número de reportes
A	100
B	50
C	25
D	10
E	5
resto de incidencias	1

Tabla I
INCIDENCIAS REPORTADAS

Las incidencias con sus posiciones reales se muestran en la Figura 1, donde podemos ver la proyección de la localización de dichas incidencias. El grosor de los puntos indica el número de reportes por incidencia.

Las personas participantes reportarían estas incidencias protegidas con el mecanismo planar de Laplace (Sección II-B1). Para simplificar hemos asumido que siempre se utiliza el mismo valor $\varepsilon = 0{,}01$. En la Figura 2 se visualizan las localizaciones reportadas.

Finalmente, al agregar estas localizaciones siguiendo la metodología descrita en la Sección II-C, podemos ver los centroides de los clústeres obtenidos en la Figura 3. Aquí el grosor del punto indica el número de puntos del clúster.

Se puede observar que los centroides de los clústeres grandes se corresponden con las incidencias con un número de reportes altos en la Figura 1.

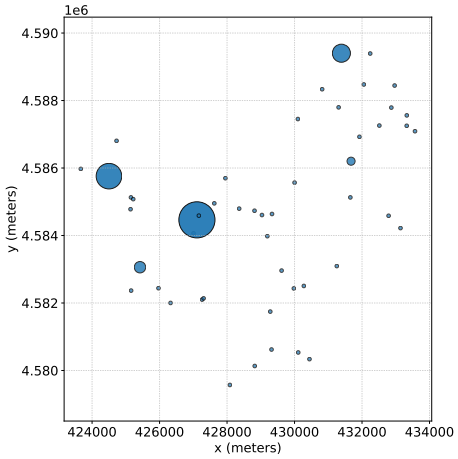


Figura 1. Incidencias reales

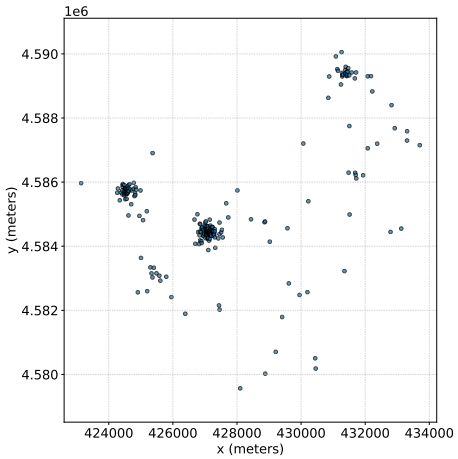


Figura 2. Incidencias reportadas (con localización protegida).

III-A. Utilidad

La utilidad de las localizaciones reportadas de forma anónima se mide cuantificando el grado en que el centroide agregado del clúster se aproxima a la ubicación real de un incidente. Una menor distancia final indica una mayor utilidad para la respuesta al incidente.

$$\mathbf{D}_{\text{Min}}(C) = \min_{i=1,\ldots,n} \|\text{Centroide}(C) - l_i\|$$

Para medir esta distancia hemos considerado la posición real de los incidentes con mayor número de reportes de la Tabla I. Para cada una hemos medido $\mathbf{D}_{\text{Min}}(C)$. Hemos considerado diferentes valores de ε, y dada la naturaleza estocástica de la anonimización hemos repetido cada una 100 veces reportando siempre el valor medio. La Tabla II muestra la distancia mínima media de los 5 clústeres mayores a su correspondiente incidencia.

Podemos ver que en general las distancias son similares y se disparan con la menor ε con una alta variabilidad. Aunque esto nos está indicando que podemos tener niveles de protección relativamente altos con una utilidad aceptable, es importante remarcar que los datos utilizados son pocos y sería necesario realizar un estudio más extenso para poder determinar mejor la influencia de ε respecto a la utilidad estimada de los datos.

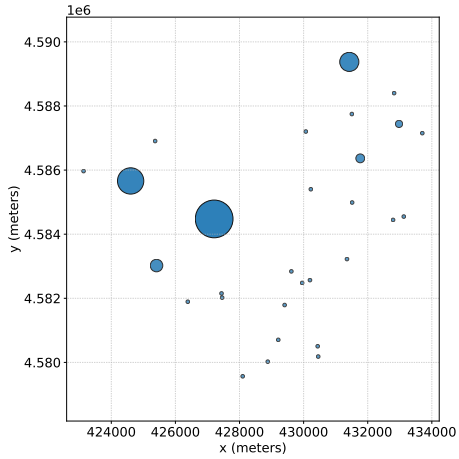


Figura 3. Incidencias agregadas

ε	Distancia media	Std
1	263.289	34.406
0.5	269.398	33.845
0.1	219.020	75.618
0.05	230.363	73.704
0.01	231.275	158.785
0.005	289.511	226.310
0.001	776.785	170.084

Tabla II
$\mathbf{D}_{\text{MIN}}(C)$ PARA LOS 5 INCIDENTES MÁS REPORTADOS.

III-B. Privacidad

El uso del mecanismo de Laplace planar asegura que la probabilidad de distinguir la ubicación real (p) de una ubicación vecina (p') decrece exponencialmente con la distancia $\|p - p'\|_2$, acotada por el parámetro de privacidad ε. Esta garantía es inherente al dato individual y no depende del conocimiento auxiliar del adversario.

Si fuéramos capaces de determinar que k reportes corresponden a la misma incidencia, el nivel de protección se degrada de acuerdo a la composición secuencial (se suma la ε de cada reporte). Ahora bien, cuando hacemos el clustering no conocemos si todos los reportes agregados corresponden al mismo incidente. Por tanto, la suma de ε es el peor caso.

Para cuantificar el impacto práctico sobre la privacidad medimos la reducción de la incertidumbre espacial correspondiente a cada reporte. Esto es, si cambiar el valor reportado por la persona usuaria por el centroide del clúster da lugar a una reducción o un incremento del ruido introducido sobre su posición original.

III-C. Discusión

Respecto a la privacidad del sistema, cada persona participante proporciona su localización con un nivel de privacidad deseado, aunque esta privacidad se verá degradada si se publican más reportes de la misma incidencia. Como se ha visto en la Sección II-B2 determinar el nivel de privacidad inicial es muy importante y debe considerarse cuidadosamente.

Nuestra propuesta protege contra adversarios pasivos globales en el caso general, siempre que exista un número de participantes y reportes mínimo. En caso contrario, se podrían producir reidentificaciones, aunque su posición no pueda llegar a ser determinada con exactitud. Dado que una

persona usuaria solamente puede reportar la misma incidencia una vez, también obtenemos protección contra atacantes internos que puedan forzar el reporte de incidencias para desanonimizar la posición de una víctima. En este sentido, el sistema proporciona una protección determinada por el nivel de privacidad inicial de cada persona usuaria. P.e., si queremos protegernos contra coaliciones de c atacantes que puedan forzar el reporte de la misma incidencia, tenemos que fijar un nivel de privacidad inicial acotado por $\frac{\varepsilon}{c}$.

También es importante remarcar que la información no protegida como el tiempo o el tipo y criticidad de la incidencia podrían dar lugar a revelación de información. En el caso del tiempo, podría protegerse introduciendo un *delay* en el envío de la incidencia; sin embargo, para que dicha protección fuese efectiva, el retraso debería ser lo suficientemente grande, lo que en muchos escenarios haría que el reporte perdiese su utilidad práctica. En cuanto al tipo y la criticidad de la incidencia, su potencial para revelar información dependerá de la granularidad con que se definan, por lo que será necesario considerarlo cuidadosamente al establecer estas categorías.

Un aspecto relevante de nuestro enfoque es que la desanonimización progresiva de la posición ocurre únicamente como consecuencia de un número elevado de personas usuarias que conocen y reportan la misma incidencia. Al mismo tiempo, este proceso incrementa el tamaño del conjunto de anonimato formado por todas las personas usuarias que han participado en el reporte. Esto no implica que la posición de una persona participante concreto no pueda verse comprometida al correlacionar incidentes previos asociados a conjuntos de anonimato más pequeños; sin embargo, sí sugiere intuitivamente que la pérdida de información individual se ve amortiguada por el crecimiento del conjunto de anonimato.

IV. Conclusiones

En este artículo hemos presentado un sistema descentralizado para reporte de incidencias en entornos vehiculares, donde se protege la localización de las personas participantes de forma adaptativa. Cada persona usuaria puede fijar un nivel de protección en su localización. Al mismo tiempo, la localización de incidentes puede desanonimizarse progresivamente a medida que dicha incidencia es reportada por mayor número de participantes. Para ello presentamos una arquitectura que integra tecnología blockchain y *geo-indistinguishability* para ofrecer un sistema de reporte vehicular descentralizado, verificable y con privacidad adaptativa. Como líneas futuras, se plantea ampliar la evaluación con escenarios más complejos y optimizar los mecanismos de agregación para reducir la pérdida de precisión sin comprometer la privacidad.

Agradecimientos

Este trabajo ha sido apoyado por los proyectos estratégicos en ciberseguridad HERMES y DANGER (C062/23), y la cátedra en ciberseguridad ARTEMISA (C057/23) financiados por el Instituto Nacional de Ciberseguridad de España a través de la Unión Europea – NextGenerationEU y del Plan de Recuperación, Transformación y Resiliencia; por los proyectos PROVTOPIA (PID2023-150098OB-I00), SECURING/CYBER (PID2021-125962OB-C31), SECURING/DATA (PID2021-125962OB-C32), SECURING/NET (PID2021-125962OB-C33), SAFE/DATA (PID2024-156914OB-C42), SAFE/BLOCKCHAIN (PID2024-156914OB-C43), SAFE/CYBER (PID2024-156914OB-C41) del Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, y European Regional Development Fund (ERDF); los proyectos SGR2021-00643, 2021SGR 00115 de la Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) de Cataluña.

Referencias

[1] S. Ebbers, S. Gense, M. Bakkouch, F. Freiling, and S. Schinzel, "Grand theft api: A forensic analysis of vehicle cloud data," *Forensic Science International: Digital Investigation*, vol. 48, p. 301691, 2024.

[2] J. Li, Z. Song, Z. Zhang, Y. Li, and C. Cao, "In-vehicle digital forensics for connected and automated vehicles with public auditing," *IEEE Internet of Things Journal*, vol. 11, no. 4, pp. 6368–6383, 2023.

[3] Z. Zhang, Z. Cao, and Y. Wang, "Forensics system for internet of vehicles based on post-quantum blockchain," *Sensors*, vol. 25, no. 19, p. 6038, 2025.

[4] S. U. Hussain and F. Koushanfar, "P3: Privacy preserving positioning for smart automotive systems," *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, vol. 23, no. 6, pp. 1–19, 2018.

[5] Z. Xu, L. Wu, C. Qin, S. Li, S. Zhang, and R. Lu, "Pritaec: Privacy-preserving task assignment based on oblivious transfer and edge computing in vanet," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 4, pp. 4996–5009, 2023.

[6] Z. Tan, C. Wang, M. Zhou, and L. Zhang, "Private information retrieval in vehicular location-based services," in *IEEE 4th World Forum on Internet of Things (WF-IoT)*, 2018, pp. 56–61.

[7] C. Zhou, T. Wang, H. Tian, W. Jiang, and Z. Wang, "A top-k query scheme with privacy preservation for intelligent vehicle networks in mobile iot," *IEEE Access*, vol. 8, pp. 81 698–81 710, 2020.

[8] Y. He, J. Ni, X. Wang, B. Niu, F. Li, and X. Shen, "Privacy-preserving partner selection for ride-sharing services," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 5994–6005, 2018.

[9] H. Yu, J. Shu, X. Jia, H. Zhang, and X. Yu, "lpride: Lightweight and privacy-preserving ride matching over road networks in online ride-hailing systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 10 418–10 428, 2019.

[10] H. Yu, X. Jia, H. Zhang, X. Yu, and J. Shu, "Psride: Privacy-preserving shared ride matching for online ride-hailing systems," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 3, pp. 1425–1440, 2021.

[11] U. M. Aïvodji, S. Gambs, M.-J. Huguet, and M.-O. Killijian, "Meeting points in ridesharing: A privacy-preserving approach," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 239–253, 2016.

[12] P. Hallgren, C. Orlandi, and A. Sabelfeld, "Privatepool: Privacy-preserving ridesharing," in *30th IEEE Computer Security Foundations Symposium (CSF)*, 2017, pp. 276–291.

[13] C. Zhang, L. Zhu, J. Ni, C. Huang, and X. Shen, "Verifiable and privacy-preserving traffic flow statistics for advanced traffic management systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 10 336–10 347, 2020.

[14] J. Zhang, F. Yang, Z. Ma, Z. Wang, X. Liu, and J. Ma, "A decentralized location privacy-preserving spatial crowdsourcing for internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2299–2313, 2021.

[15] C. Zhang, L. Zhu, and C. Xu, "Bsdp: Blockchain-based smart parking for digital-twin empowered vehicular sensing networks with privacy protection," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 9, pp. 7237–7246, 2023.

[16] National Geospatial-Intelligence Agency (NGA), "Department of defense world geodetic system 1984: Its definition and relationships with local geodetic systems," U.S. Department of Defense, Tech. Rep. TR8350.2, 2000, third Edition, amended 2014. [Online]. Available: https://earth-info.nga.mil/

[17] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security (CCS)*. Berlin, Germany: ACM, 2013, pp. 901–914.

[18] R. J. G. B. Campello, D. Moulavi, A. Zimek, and J. Sander, "Hierarchical density estimates for data clustering, visualization, and outlier detection," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 10, no. 1, pp. 5:1–5:51, 2015. [Online]. Available: https://doi.org/10.1145/2733381