Protección de la privacidad en trayectorias para estudiar la propagación de epidemias

Cristina Romero-Tris*, Joan Melià-Seguí*† y David Megías*
* Internet Interdisciplinary Institute
† Estudis d'Informàtica, Multimèdia i Telecomunicació
Universitat Oberta de Catalunya
Email: {cromerotr,melia,dmegias}@uoc.edu

***Resumen*—La modelización de epidemias es un campo de la ciencia que usa herramientas para estudiar el origen, explicar la evolución y predecir el comportamiento de ciertas enfermedades propagables. Por otro lado, el cada vez más extendido uso de dispositivos con reconocimiento de ubicación (e.g., *smartphones*, GPS, etc.), permite generar trayectorias seguidas por los usuarios. Combinando estos dos conceptos, es posible comparar trayectorias que han seguido usuarios infectados por una enfermedad y facilitar así su modelización. Sin embargo, las trayectorias pueden desvelar información privada sobre los usuarios, como por ejemplo la dirección de su vivienda, comportamientos relativos a la hora que entran y salen de ella, etc. Esto puede disuadir a algunos usuarios de compartir sus datos. En este artículo proponemos un método que permita encontrar aquellos lugares donde los infectados hayan coincidido físicamente, sin desvelar información sobre otros lugares que hayan visitado. El sistema usa un protocolo criptográfico que permite a los usuarios cifrar los puntos de sus trayectorias y compararlos con trayectorias de otros usuarios, de forma que la única información obtenida sea aquella relevante para el estudio de la propagación de la epidemia.**

***Palabras clave*—criptografía (*cryptography*), privacidad (*privacy*), propagación de epidemias (*epidemic spreading*), trayectorias (*trajectories*).**

I. Introducción

La epidemiología es una rama de la medicina que estudia la incidencia, distribución, y posible control de enfermedades y otros factores relacionados con la salud [1]. La modelización de epidemias permite estudiar el comportamiento de enfermedades e incluso prevenir su propagación. Como ejemplo a gran escala, el estudio del brote epidémico de ébola de 2014-2016 [2] permitió saber las zonas de infección y contagio. La epidemia apareció en diciembre de 2013 en Guinea y se extendió por Liberia, Sierra Leona, Nigeria, Senegal, Estados Unidos, España, Malí y Reino Unido. La Organización Mundial de la Salud encontró en África occidental el foco de la epidemia. Todos estos datos sobre la propagación permitieron actuar y prevenir el contagio de un mayor número de personas.

Utilizar datos sobre trayectorias seguidas por los enfermos infectados ayuda a la modelización de la epidemia, y puede ser de gran utilidad para controlar su propagación. Estos datos de trayectorias pueden extraerse de numerosas fuentes. Con el uso extendido de dispositivos con reconocimiento de ubicación (e.g., *smartphones*, GPS, etc.) es posible registrar las localizaciones de los usuarios y en qué momentos se produjeron. Un ejemplo de la facilidad de la obtención de trayectorias es el Google Maps Timeline [3] (Figura 1). Con este servicio, es posible visualizar (e incluso exportar los datos para análisis) las localizaciones visitadas en cualquier momento por los usuarios. La aplicación poporciona datos avanzados como la latitud, la longitud y el nombre del sitio visitado (e.g., el nombre de un restaurante o de un hotel).

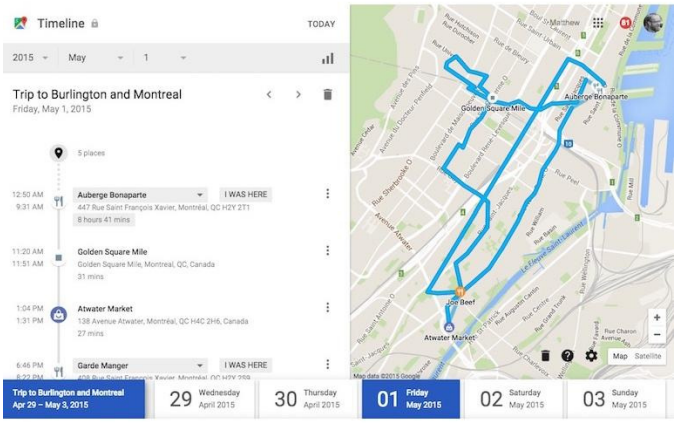

Figura 1. Trayectoria mostrada por Google Maps Timeline

Sin embargo, el problema es que a menudo las trayectorias contienen información que algunos usuarios podrían considerar privada. Siguiendo con el mismo ejemplo, Google Maps Timeline permite generar un *ranking* de las localizaciones más visitadas. Lógicamente, en los primeros lugares de esta lista suelen aparecer las dirección de la vivienda y el lugar de trabajo del usuario, o en cualquier caso puede resultar evidente su clasificación como indican Golle y Partridge en [4]. A partir de ahí es fácil saber qué otros lugares ha visitado y deducir otros datos sobre comportamiento, modo de vida, creencias religiosas, preferencias sexuales, ideología, etc., que pueden amenazar la privacidad de los usuarios.

Una de las soluciones más extendidas en la literatura actual es anonimizar las trayectorias en la base de datos antes de proceder a su análisis, a su publicación, o a ser enviadas a otra

entidad [5], [6], [7]. Esto supone que los usuarios van a querer compartir su información real de trayectorias sin anonimizar. Sin embargo, opinamos que es preferible que los datos hayan sido anonimizados antes de ser enviados al servidor. Nuestra propuesta se basa en que el usuario participe activamente en el proceso de anonimización.

Es importante tener en cuenta que para la modelización de epidemias no es necesario saber la identidad de los usuarios, ni cualquier localización en la que haya estado que no esté relacionada con la enfermedad. Sí que es importante saber dónde se han producido contagios y dónde han estado en contacto con otros enfermos. Como describen Patil et al. [8], los usuarios preferirían decidir qué y cuándo compartir, aunque es necesario que al menos uno de los usuarios comparta todo su historial de eventos si se quiere asegurar que se dispone de todas las posibles coincidencias. Por ello, en este artículo proponemos un protocolo que permite desvelar aquellos puntos de la trayectoria donde al menos dos usuarios infectados hayan coincidido, sin desvelar ninguna otra información.

Este artículo se centra en la aplicación de este protocolo al estudio de epidemias. Sin embargo, muchas otras aplicaciones también serían posibles. Por ejemplo, podríamos agilizar rutas de transporte monitorizando el tráfico para saber cuántos usuarios coinciden en un intervalo determinado en una localización concreta. Otra aplicación, en este caso con fines comerciales, sería un servicio de publicidad adaptado a usuarios que coinciden en varios puntos de una trayectoria, asumiendo que pueden tener intereses comunes.

El resto del artículo se organiza de la siguiente manera. En la Sección II se detalla el estado del arte relacionado con el presente artículo. La Sección III introduce la privacidad centrada en el usuario mediante la definición de un umbral de localización y tiempo, y el método para la compartición de trayectorias de forma distribuida y privada se describe en la Sección IV. Finalmente, la Sección V concluye el artículo y detalla las posibles líneas de investigación a seguir.

II. Estado del arte

En los últimos tiempos, el estudio de trayectorias ha atraído información en la literatura de seguridad y privacidad en los sistemas de información. A continuación, describimos algunos de estos trabajos y la forma que tienen de anonimizar trayectorias.

Abul et al. [5] proponen un modelo llamado de (k, δ)-anonimidad, consistente en agregar k trayectorias dentro de un cilindro de radio δ, de manera que cada trayectoria queda identificada solo por ese cilindro y es indistinguible de las otras $k-1$ trayectorias.

Terrovitis et al. [6] proponen un algoritmo que elimina algunos puntos de la trayectoria. El desafío en este caso es encontrar el conjunto óptimo de puntos a eliminar que minimice la pérdida de información relevante. De manera similar, Pensa et al. [9] proponen eliminar patrones que aparezcan en los datos con una cierta frecuencia. Esto se realiza añadiendo, borrando o sustituyendo algunos puntos de las trayectorias.

Yarovoy et al. [7] usan curvas Hilbert [10] para proyectar las localizaciones (de un espacio multidimensional) a una sola dimensión. El objetivo es encontrar "vecinos" en cada punto de la trayectoria y usarlos para crear grupos de anonimización y generalizar los datos de cada usuario.

Varios artículos en la literatura tratan la compartición privada de eventos. En [11], [12] los autores exploran el problema de compartición de eventos de un calendario de forma privada para organizar una reunión preservando la privacidad de cada usuario (es decir, sin revelar entradas individuales del calendario), y de forma eficiente. En [13], Bilogrevic et al. proponen dos algoritmos para organizar reuniones entre usuarios de forma privada y De Cristofaro et al. propone un método para sugerir reuniones entre usuarios en el futuro de forma privada y en base a la predicción de sus actividades [14]. En [15], Zhong et al. proponen soluciones al problema del "amigo cercano" (determinar si dos usuarios se encuentran a una distancia menor que un determinado radio), mientras que en [16], [17], [18] el problema se traslada a un entorno de celdas.

III. Definición de trayectorias mediante privacidad centrada en el usuario

El estudio de propagación de epidemias tiene una doble componente de sensibilidad en cuanto a la información se refiere. Por un lado se expone informacion privada relativa a cada usuario (trayectorias comunes, lugar de residencia/trabajo, preferencias, etc...) y por otro existe una componente crítica de salud pública alrededor de la compartición de las trayectorias mencionadas. En los siguientes apartados se explora la problemática de definir trayectorias mediante privacidad centrada en el usuario, a partir de la distorsión de los eventos.

III-A. Distorsión de trayectorias mediante umbral de localización y tiempo

Las trayectorias de los usuarios se almacenan en los dispositivos (o en la nube) mediante un formato simple de combinación de eventos, en donde cada evento está representado por la latitud, longitud, fecha y hora. Sin embargo, también podemos proporcionar mayor privacidad al usuario si distorsionamos estos puntos, es decir, si incrementamos el grado de incertidumbre de la posición o el tiempo exacto en el que el usuario se encontraba en ese punto. Una de las opciones es superponer una cuadrícula al mapa, y compartir sólo las coordenadas de la región donde se encontraba el usuario. Esta idea se usa en algunos trabajos previos, como el propuesto en [19].

Otra de las opciones es la que aparece en [20]. En este caso, cada punto de la trayectoria se anonimiza transformándolo en un cilindro. Suponiendo un eje de tres dimensiones como el que se muestra en la Figura 2, el usuario elige un radio (e.g., 1 km) para la base del cilindro, formando un círculo que contiene la latitud y la longitud del punto original. Además, el usuario escoge un intervalo de tiempo, por ejemplo una hora (cf. Sección IV), que formará la altura del cilindro y que marcará el intervalo de tiempo en el que el usuario se

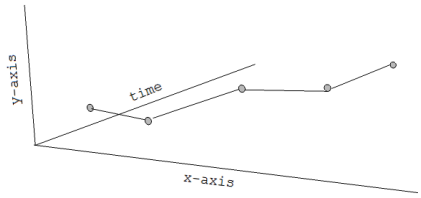

Figura 2. Ejemplo de trayectoria sin anonimizar

encontraba en ese punto. Este sistema permite al usuario decidir el grado de privacidad que desea para cada uno de los puntos de la trayectoria: escogiendo valores superiores para el radio y la altura del cilindro conseguirá mayor incertidumbre sobre su localización exacta. El resultado de un ejemplo de anonimización de la Figura 2 se muestra en la Figura 3.

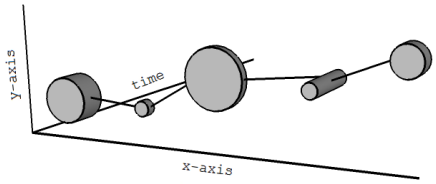

Figura 3. Ejemplo de trayectoria anonimizada

III-B. Propuesta de mejora a partir de actividad en redes sociales

El problema de distorsionar los puntos como se describe en la Sección III-A, es que no tienen en cuenta la distribución real de probabilidades de localización. Si consideramos un caso extremo, por ejemplo una persona que viva en una casa en el campo, donde sus vecinos más próximos se encuentren a unos pocos kilómetros, está claro que la probabilidad de adivinar la posición exacta en los sistemas de [19] y [20] es muy alta. Por ello, una nueva propuesta es utilizar datos extraídos de redes sociales para anonimizar la trayectoria. En concreto, la idea es saber cuántas publicaciones (e.g., tweets en Twitter [21] o estados en Facebook [22]) se han hecho cerca del punto que quiere compartir el usuario. Así, el usuario sólo compartiría aquellos puntos que se encuentran a una distancia inferior al umbral escogido de al menos k publicaciones en redes sociales. La Figura 4 ilustra tres ejemplos en los que a patir de las localizaciones de la red social Twitter, se anonimiza una posición considerando las $k = \{10, 100, 1000\}$ ubicaciones más cercanas desde las que se ha publicado en la red social, en un intervalo de tiempo dado.

IV. Compartición de trayectorias de forma distribuida y privada

En la sección anterior se han introducido diferentes posibilidades para la definición de forma privada de una trayectoria a partir de combinar eventos (localización más umbral de tiempo específico). En esta sección se describen el modelo y el protocolo utilizados para la compartición de dichas trayectorias de forma privada, basado parcialmente en el método descrito en [14]. Este proceso es independiente del formato en el que se encuentre la trayectoria. Por ejemplo, podemos considerar un formato básico en el que la trayectoria está formada por coordenadas GPS exactas, mediante la anonimización de ubicación y tiempo mediante un umbral, a partir de áreas definidas por la actividad cercana en redes sociales (cf. Sección III), o bien por localizaciones con información semántica como en Foursquare [23].

IV-A. Modelo del sistema

Se consideran dos usuarios, Alice y Bob, que han sido infectados por un virus específico. Se quiere comprobar si en el pasado han coincidido en tiempo y lugar (con cierto margen) para conocer el foco de la infección. Se asumen las siguientes características:

- **Intervalos de tiempo:** Alice y Bob acuerdan un conjunto de n intervalos de tiempo consecutivos $t_0, t_1, \ldots, t_{n-1}$, de duración d (por consiguiente, $t_n = t_0 + dn$).
- **Localizaciones:** los *smartphones* de Alice y Bob permiten la localización constante de sus usuarios, almacenándola en una base de datos específica, de modo similar al Google Maps Timeline [3]. En concreto, denominamos $L_{A:i}$ la localización de Alice en el intervalo de tiempo t_i (análogamente, $L_{B:i}$ denota la localización de Bob en el intervalo de tiempo t_i). Se asume que, en este modelo, las localizaciones actúan como identificadores únicos y pueden tener formatos distintos tales como los definidos en la Sección III, lugares preestablecidos (como en Foursquare [23]) o celdas de una cuadrícula [17].
- **Compartición preservando la privacidad:** el objetivo del sistema es permitir a Alice y Bob descubrir $\{t_i \mid L_{A:i} = L_{B:i}\}$, únicamente. Es decir, Alice no aprende ninguna información sobre $L_{B:j}$ si $L_{B:j} \neq L_{A:j}$ (y viceversa).
- **Usuarios local y remoto:** se describe como usuario local el que inicia el protocolo de compartición privada de trayectorias, y usuario remoto el que recibe la petición de inicio del protocolo de compartición privada de trayectorias.

IV-B. Protocolo de compartición distribuida de trayectorias

En el modo distribuido de compartición de trayectorias, los dispositivos de los usuarios pueden comparar las trayectorias (conjuntos de eventos consistentes en la combinación de localización e intervalo de tiempo) mutuos mediante un protocolo seguro y distribuido que solo revela los eventos en los que ambos han coincidido. En este protocolo distribuido, los dos dispositivos pueden cifrar sus eventos mediante una clave de cifrado y comprobar si alguno de los eventos cifrados coincide. Si se encuentra una coincidencia, el dispositivo local puede usar el evento cifrado para determinar cuál de los eventos sin cifrar es común entre el usuario local y el

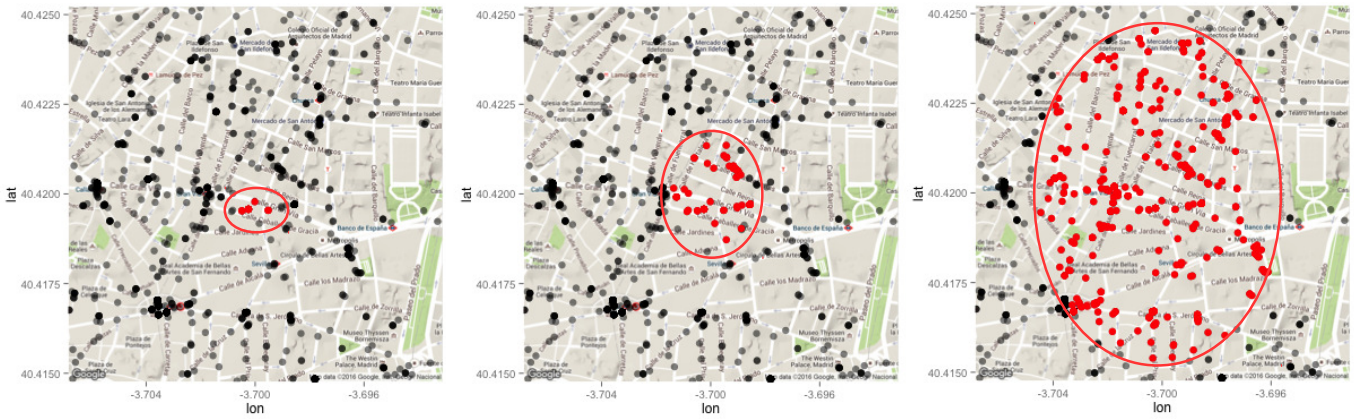

Figura 4. Ejemplo en el que se anonimiza una posición a partir de las $k = \{10, 100, 1000\}$ posiciones más próximas.

usuario remoto. Durante la operación, el dispositivo local obtiene una clave pública criptográfica, desde un repositorio, o directamente desde el dispositivo remoto. Por ejemplo, durante la inicialización del protocolo, el dispositivo remoto puede generar un par de claves pública y privada, y enviar la clave pública al dispositivo local. El dispositivo local puede obtener la clave pública del dispositivo remoto antes de comparar los eventos con el dispositivo remoto por primera vez, y puede guardar la clave pública para usarla durante las siguientes comparaciones.

El dispositivo remoto puede generar valores para las variables (N, e, d), siendo N un número primo. La tupla (N, d) sirve como clave privada para el dispositivo remoto, y el dispositivo remoto proporciona la tupla (N, e) como clave pública para el dispositivo local. El dispositivo local genera un valor aleatorio $R_1 \in \mathbb{Z}_N$, para un grupo acíclico finito $\mathbb{Z}_N$ (por ejemplo $\mathbb{Z}_N = 1, 2, \ldots, N$). En el siguiente paso, el sistema cifra los eventos del usuario local usando la clave local, y envía los eventos cifrados localmente al dispositivo remoto. Por ejemplo, el sistema puede generar en local el evento cifrado α_i usando la Ecuación 1 para el evento A_i.

$$\alpha_i = (H(A_i)R_1^e) \bmod N. \tag{1}$$

En la Ecuación 1, $H(A_i)$ corresponde a un valor de *hash* para el evento A_i, basado en una función de *hash* predeterminada H que es común entre los dispositivos local y remoto. El dispositivo remoto procesa el evento cifrado en local usando la clave privada del dispositivo remoto.

En el siguiente paso, el dispositivo local recibe, desde el dispositivo remoto, los eventos del dispositivo local, cifrados con la clave del dispositivo remoto. El dispositivo remoto genera el cifrado del conjunto de eventos del dispositivo local γ_i usando la Ecuación 2.

$$\gamma_i = \alpha_i^d \bmod N. \tag{2}$$

Expandiendo la Ecuación 2 usando el evento cifrado en local α_i a partir de la Ecuación 1 obtenemos la Ecuación 3.

$$\gamma_i = H(A_i)^d R_1 \bmod N. \tag{3}$$

Cabe recordar que el sistema también puede recibir los eventos del dispositivo remoto de forma cifrada. Estos eventos del usuario remoto, cifrados en el dispositivo remoto, se describen en la Ecuación 4.

$$\beta_i = H(H(B_i)^d \bmod N). \tag{4}$$

En la Ecuación 4, (N, d) corresponde a la clave privada generada por el dispositivo remoto, y $H(B_i)$ corresponde al valor de *hash* para el evento B_i.

El sistema puede ajustar los eventos locales cifrados en remoto (de la Ecuación 3) para convertirlos en el mismo formato que los eventos cifrados del dispositivo remoto de la Ecuación 4. De este modo, comparando los dos conjuntos de eventos cifrados para determinar una coincidencia de eventos cifrados, el sistema ajusta los eventos cifrados de la Ecuación 3 para adoptar la forma descrita en la Ecuación 5.

$$\alpha_i' = H(\gamma_i / R_1). \tag{5}$$

Expandiendo la Ecuación 5 usando el evento cifrado en local α_i de la Ecuación 3, se obtiene la Ecuación 6.

$$\alpha_i' = H(H(A_i)^d \bmod N). \tag{6}$$

Nótese que α_i' en la Ecuación 6 tiene la misma forma que β_i en la Ecuación 4. De este modo, cuando ambos usuarios coinciden en los eventos A_i y B_j, los valores cifrados α_i' and β_j también serán coincidentes.

En el siguiente paso el sistema compara los eventos cifrados del usuario local (Ecuación 6) con los eventos cifrados del usuario remoto (Ecuación 4) para determinar una coincidencia en los eventos cifrados. Si el sistema detecta una coincidencia en los eventos cifrados, el sistema busca una descripción de evento que corresponda al evento cifrado. Continuando con el ejemplo anterior, si el sistema determina que el evento cifrado remoto β_j coincide con el evento cifrado local α_i', el sistema

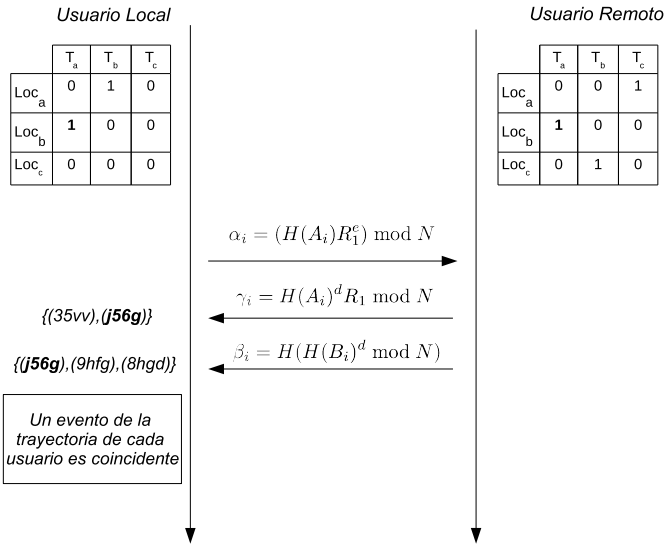

Figura 5. Método distribuido de compartición de trayectorias.

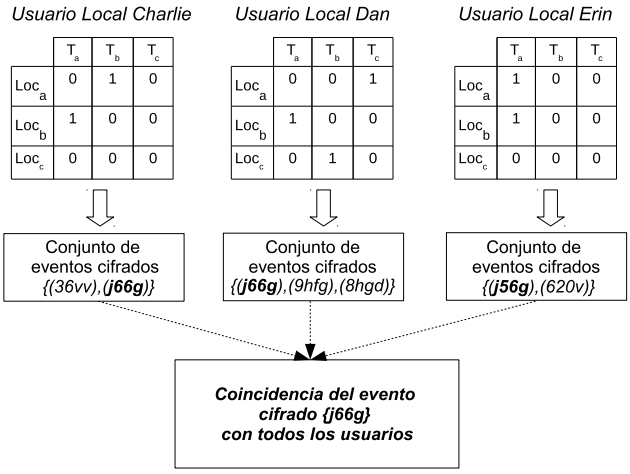

Figura 6. Método centralizado de compartición de trayectorias.

busca el evento A_i que corresponda con el valor cifrado α'_i. Este evento A_i también corresponde con el evento cifrado remoto β_j.

La Figura 5 describe el intercambio de mensajes del protocolo para la compartición distribuida de eventos. Cada dispositivo guarda un conjunto de eventos (localización e intervalo de tiempo) siguiendo un modelo similar al de Google Maps Timeline [3]. El histórico de localizaciones $\{\text{Loc}_\text{a}, \text{Loc}_\text{b}, \text{Loc}_\text{c}, \ldots\}$ se combina en una matriz binaria con los intervalos de tiempo, en un formato de filas y columnas. Cada posición de la matriz se codifica con un 1 si existe ese evento, o un 0 en caso contrario.

En el siguiente paso del protocolo, el dispositivo local cifra los eventos locales y los envía al dispositivo remoto, que los cifra con su clave privada (cf. Ecuación 6). El dispositivo remoto envía el conjunto de eventos cifrados de nuevo al dispositivo local, representado en la Figura 5 como {(35vv),(j56g)}. En paralelo, el dispositivo remoto también cifra los eventos remotos con su clave privada (cf. Equation 4), y los comparte con el dispositivo local, representado en la Figura 5 como {(j56g),(9hfg),(8hgd)}. Nótese que el dispositivo local dispone en este punto de ambos conjuntos de eventos (local y remoto) cifrados con la clave privada del dispositivo remoto.

A continuación, el módulo de compartición de eventos (cf. Figura 5) puede comparar ambos conjuntos de eventos cifrados en el dispositivo remoto para identificar eventos coincidentes para los dos usuarios. Por ejemplo, el módulo de compartición de eventos puede determinar que el evento cifrado {j56g} existe en ambos conjuntos, y que corresponde al evento (Loc_b, T_a).

IV-C. Protocolo de compartición centralizada de trayectorias

El modo de compartición centralizada de eventos se detalla en la Figura 6. En lugar de un protocolo de intercambio de mensajes entre pares de usuarios, un servidor centraliza el intercambio de mensajes permitiendo más de dos usuarios participar en el sistema. En este caso los usuarios son equivalentes al usuario local (a efectos de ejecución del protocolo), mientras que el servidor actúa como usuario remoto.

En la Figura 6, los usuarios Charlie, Dan y Erin disponen de un conjunto de eventos organizados en una matriz, y generan el correspondiente conjunto de eventos cifrados para cada usuario. Cada conjunto de eventos se cifra con una clave pública que ha generado y compartido el servidor central, o una tercera entidad (por ejemplo, un proveedor de servicio).

La Figura 6 representa cómo el servidor central encuentra el evento cifrado coincidente {j66g} entre los usuarios Charlie, Dan y Erin, sin necesidad de descifrar el conjunto de eventos. En el siguiente paso el servidor devuelve el evento cifrado anterior como evento coincidente al conjunto de usuarios. Por último, cada usuario puede comparar el evento cifrado con su conjunto de eventos sin cifrar.

V. Conclusión y trabajo futuro

Los datos sobre trayectorias pueden ser útiles para analizar diversos fenómenos sociales. En este artículo, destacamos su utilidad para el modelado de epidemias, es decir, para descubrir posibles focos de infección y de contagio, y así prevenir su propagación. El método propuesto permite compartir sólo aquella información que es relevante para el modelado de epidemias, protegiendo la privacidad de los usuarios.

Aunque este artículo propone una idea novedosa y su aplicación directa en un campo concreto, el trabajo se encuentra en sus primeras etapas de desarrollo. Existen todavía numerosas cuestiones que deben ser tratadas en un futuro. Nuestro primer objetivo es adaptar el protocolo propuesto para que funcione con trayectorias distorsionadas, no sólo con localizaciones exactas (latitud, longitud). Utilizando datos reales de publicaciones en redes sociales (e.g., Facebook, Twitter, etc.) determinaremos la singularidad de un punto de la trayectoria. Si cerca de ese punto existen más de un cierto número de publicaciones (e.g., k), el punto puede ser

compartido para su posterior análisis, proporcionando así k-anonimidad al usuario. Esta adaptación supone hacer un estudio matemático del protocolo, incorporando homomorfismos que permitan saber si la distancia entre dos puntos es inferior a un umbral, sin revelar ningun otro dato.

Finalmente, el siguiente objetivo es implementar el protocolo y obtener resultados con datos reales de trayectorias relacionadas con epidemias. Queremos analizar los resultados y obtener datos sobre posibles focos de contagio.

Agradecimientos

Los autores agradecen la colaboración de Emiliano De Cristofaro, University College London. Este trabajo está parcialmente financiado por el Ministerio de Economía y Competitividad a través de los proyectos TIN2011-27076-C03-02 CO- PRIVACY, TIN2014-57364-C2-2-R SMARTGLACIS, TEC2015-71303-R SINERGIA y TSI-020602-2012-147 IRIS.

Referencias

[1] Epidemiology. Oxford Dictionaries. Oxford University Press, n.d. Web. 21 April 2016. http://www.oxforddictionaries.com/definition/english/epidemiology

[2] Organización Mundial de la salud, 2016. http://www.who.int/csr/disease/ebola/es/

[3] Google Maps TimeLine, 2016. https://www.google.es/maps/timeline

[4] P. Golle, and K. Partridge. "On the anonymity of home/work location pairs." *7th International Conference on Pervasive Computing*, pp. 390-397, Nara, Japan, 2009.

[5] O. Abul, Osman, F. Bonchi, and M. Nanni, "Never walk alone: Uncertainty for anonymity in moving objects databases," Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on. Ieee, 2008.

[6] M. Terrovitis, and N. Mamoulis, "Privacy preservation in the publication of trajectories," Mobile Data Management, 2008. MDM'08. 9th International Conference on. IEEE, 2008.

[7] R. Yarovoy, F. Bonchi, L.V. Lakshmanan, W. Wang, "Anonymizing moving objects: How to hide a mob in a crowd?", In Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology (pp. 72-83). ACM, 2009.

[8] S. Patil, G. Norcie, A. Kapadia, and A. J. Lee. Reasons, rewards, regrets: privacy considerations in location sharing as an interactive practice. In *SOUPS*, 2012.

[9] R. G. Pensa, A. Monreale, F. Pinelli, D. Pedreschi, "Pattern-preserving k-anonymization of sequences and its application to mobility data mining," PiLBA'08 Privacy in Location-Based Applications, 44, 2008.

[10] D. Hilbert, "Ueber die stetige Abbildung einer Line auf ein Flächenstück", Mathematische Annalen 38.3 (1891): 459-460.

[11] I. Bilogrevic, M. Jadliwala, J.-P. Hubaux, I. Aad, and V. Niemi. Privacy-Preserving Activity Scheduling on Mobile Devices. In *Codaspy*, 2011.

[12] E. De Cristofaro, A. Durussel, and I. Aad. Reclaiming Privacy for Smartphone Applications. In *PerCom*, 2011.

[13] I. Bilogrevic, M. Jadliwala, K. Kalkan, J.-P. Hubaux, and I. Aad. Privacy in mobile computing for location-sharing-based services. In *PETS*, pages 77–96, 2011.

[14] E. De Cristofaro, J. Melia-Segui, R. Zhang, O. Brdiczka, and E. Uzun. "Method and apparatus for performing distributed privacy-preserving computations on user locations." U.S. Patent No. 8,954,737. 10 Feb. 2015.

[15] G. Zhong, I. Goldberg, and U. Hengartner. Louis, Lester and Pierre: Three protocols for Location Privacy. In *PET*, 2007.

[16] Z. Lin, D. F. Kune, and N. Hopper. Efficient Private Proximity Testing with GSM Location Sketches. In *Financial Cryptography and Data Security*, pages 73–88. 2012.

[17] A. Narayanan, N. Thiagarajan, M. Lakhani, M. Hamburg, and D. Boneh. Location Privacy via Private Proximity Testing. In *NDSS*, 2011.

[18] J. D. Nielsen, J. I. Pagter, and M. B. Stausholm. Location privacy via actively secure private proximity testing. In *PERCOM Workshops*, pages 381–386, 2012.

[19] G. Gidofalvi, X. Huang, T. Bach Pedersen, "Privacy-Preserving Data Mining on Moving Object Trajectories," In Proceedings of the 2007 International Conference on Mobile Data Management (MDM '07). IEEE Computer Society, Washington, DC, USA, 60-68, 2007.

[20] C. Romero-Tris, D. Megías, "User-centric Privacy-Preserving Collection and Analysis of Trajectory Data," Lecture Notes in Computer Science. Pag. 245-253, 2016.

[21] Twitter, 2016. https://twitter.com/

[22] Facebook, 2016. https://www.facebook.com/

[23] J. Melià-Seguí, R. Zhang, E. Bart, B. Price, and O. Brdiczka. "Activity duration analysis for context-aware services using foursquare check-ins." Proceedings of the 2012 international workshop on Self-aware Internet of Things, pp. 13-18, 2012.