TÉCNICAS DE ANONIMIZACIÓN PARA DATOS BIOMÉTRICOS CONDUCTUALES: UN ESTUDIO

Realizado por Simon Hanisch[1], Prof. Patricia Arias Cabarcos[2], Dr. Javier Parra Arnau[3] y Prof. Thorsten Strufe[1].

[1] Karlsruher Institut für Technologie, Alemania
[2] Universität Paderborn, Alemania
[3] Universitat Politècnica de Catalunya, España

RESUMEN EJECUTIVO

Hipótesis. La actual transformación digital está impulsando una recopilación cada vez más exhaustiva de datos sobre los ciudadanos, intensificada por la mejora continua de dispositivos electrónicos periféricos como gafas de realidad aumentada y virtual, trajes y guantes de captura de movimiento, relojes inteligentes y otros wearables, que aumentan drásticamente la cobertura y resolución de los datos biométricos y de comportamiento disponibles para su procesamiento. Una parte de estos datos se comparte de forma consciente —al publicar fotos, vídeos u opiniones, compartir la ubicación en redes sociales, participar en foros o valorar servicios—, mientras que una cantidad aún mayor se recopila de manera inadvertida durante la navegación web, el uso de servicios de geolocalización o simplemente al interactuar con espacios inteligentes dotados de asistentes de voz y cámaras. Estos datos conductuales son altamente descriptivos del individuo, ya que revelan rutinas, hábitos, información médica, manías y tics, y pueden incluso permitir la inferencia de condiciones de salud mediante correlaciones conocidas entre rasgos fisiológicos y enfermedades, como la detección de depresión en imágenes faciales o de insuficiencias orgánicas a partir de la coloración de ojos o piel. Además, los datos de comportamiento pueden utilizarse para identificar de forma única a las personas, ya sea a través del análisis de contenidos en redes sociales, de patrones de movilidad y navegación web o de características como la marcha, que también puede revelar atributos como la edad, el género y determinadas condiciones fisiológicas.

Propósito de la investigación. Preservar la privacidad —y, en última instancia, la dignidad— de las personas cuyos comportamientos son captados por estos sensores exige enfoques mucho más sofisticados que la simple eliminación de identificadores directos o cuasi-identificadores intuitivos en las bases de datos, ya que los datos de comportamiento humano presentan dependencias temporales y fisiológicas derivadas de su naturaleza secuencial y de las restricciones biológicas del organismo. Estas interdependencias estructurales, junto con la presencia de información contextual y patrones conductuales estables, cuestionan la eficacia de las técnicas de anonimización basadas en aleatorización o perturbación. En este contexto, un número creciente de estudios aborda la anonimización de datos de comportamiento centrándose en distintos rasgos humanos, como la voz, la marcha, los gestos o el ritmo cardíaco, lo que pone de manifiesto la necesidad de una revisión sistemática que identifique similitudes conceptuales y metodológicas, clarifique las diferencias entre enfoques y destaque tanto sus propiedades como las oportunidades de investigación futuras.

Metodología. En esta edición de los *Premio de Investigación en Protección de Datos Personales “Emilio Aced”*, presentamos un estudio sistemático y crítico del estado del arte de técnicas de protección para datos biométricos conductuales. Nuestro trabajo profundiza en las tecnologías de mejora de la privacidad aplicables a escenarios en los que los datos de comportamiento son recopilados por terceros o compartidos con ellos para llevar a cabo una operación específica o prestar un servicio. Nuestro estudio se centra en la privacidad de los datos más que en su confidencialidad: no abordamos soluciones en las que una entidad cifra sus propios datos para impedir el acceso por parte de terceros. En su lugar, analizamos aquellas propuestas diseñadas para evitar la revelación no intencionada de la información contenida en los datos. Por este motivo, enfoques como la computación confidencial, el procesamiento basado en criptografía homomórfica o técnicas afines —en los que el propietario de los datos es la única entidad que puede inferir información a partir de ellos— quedan fuera del alcance de nuestro análisis.

Resultados. Siguiendo las directrices de Kitchenham, realizamos un estudio sistemático del estado del arte que analizó 142 trabajos seleccionados de un corpus inicial de 364, con el objetivo de identificar las aplicaciones más comunes que procesan datos de comportamiento, extraer métricas de utilidad relevantes y caracterizar las amenazas típicas a la privacidad junto con sus modelos de adversario. En este proceso, definimos dos taxonomías de anonimización específicas para este tipo de datos: una basada en la forma en que las técnicas transforman la información, y otra centrada en los objetivos de privacidad que buscan garantizar. Además, presentamos un resumen estructurado de las soluciones de anonimización, organizado por el rasgo protegido, detallando para cada propuesta las aplicaciones que condicionan su utilidad, las amenazas consideradas, los objetivos y conceptos de anonimización aplicados, así como las evaluaciones realizadas, incluidos los conjuntos de datos utilizados. En conjunto, este estudio contribuye a una mejor comprensión de los riesgos asociados al uso de tecnologías emergentes que operan sobre datos altamente sensibles y de potencial impacto social, y proporciona elementos clave para fomentar su uso responsable y seguro.

Novedad y aplicabilidad. La novedad de nuestro trabajo reside en la visión integral y sistematizada de la anonimización de datos biométricos de comportamiento, un ámbito hasta ahora fragmentado y abordado de forma desigual según el rasgo considerado. La propuesta de taxonomías específicas, junto con el análisis comparativo de aplicaciones, amenazas, modelos de adversario y metodologías de evaluación, permite identificar carencias estructurales comunes y establecer un marco unificador que facilita la comparación y el diseño de nuevas soluciones. Desde el punto de vista de la aplicabilidad, los resultados obtenidos son directamente relevantes para investigadores y profesionales que desarrollan o evalúan sistemas basados en datos de comportamiento, ya que proporcionan criterios claros para seleccionar técnicas de anonimización, evaluar su impacto en la utilidad y anticipar riesgos de privacidad en escenarios realistas. De este modo, nuestro trabajo sienta las bases para el desarrollo de tecnologías más robustas y responsables en contextos de creciente sensibilidad social y regulatoria.

Conclusiones. La anonimización de datos biométricos de comportamiento es clave para proteger la privacidad, pero sigue presentando importantes retos. Nuestra revisión identifica múltiples rasgos de comportamiento y propone una taxonomía para clasificar las técnicas de anonimización según el tipo de técnico de protección aplicada; sin embargo, más allá de la voz (un ámbito ya bien estudiado), la mayoría de estos rasgos han recibido escasa atención, lo que deja su protección como una cuestión de investigación abierta. Además, las técnicas existentes suelen evaluarse de manera limitada y bajo supuestos de atacantes débiles o poco realistas, lo que evidencia la necesidad de mejorar las metodologías de evaluación. Por último, el componente temporal y acumulativo de los datos ha sido en gran medida ignorado, tanto por la escasez de enfoques para el análisis de flujos de datos como por la falta de protección continuada en la mayoría de las técnicas actuales. Creemos que nuestro análisis exhaustivo del estado actual de la técnica puede ser de gran ayuda para los investigadores al ofrecer una visión estructurada de las soluciones existentes, poner de manifiesto sus limitaciones y señalar direcciones claras para el desarrollo de métodos de anonimización más robustos y realistas.