Guía docente
230990 - DP - Privacidad de Datos

Última modificación: 26/05/2023

Unidad responsable: Escuela Técnica Superior de Ingeniería de Telecomunicación de Barcelona
Unidad que imparte: 744 - ENTEL - Departamento de Ingeniería Telemática.

Titulación: MÁSTER UNIVERSITARIO EN CIBERSEGURIDAD (Plan 2020). (Asignatura optativa).

Curso: 2023 **Créditos ECTS:** 5.0 **Idiomas:** Inglés

PROFESORADO

Profesorado responsable: Consultar aquí / See here: https://telecos.upc.edu/ca/estudis/curs-actual/professorat-responsables-coordinadors/responsables-assignatura

Otros: Consultar aquí / See here: https://telecos.upc.edu/ca/estudis/curs-actual/professorat-responsables-coordinadors/professorat-assignat-idioma

METODOLOGÍAS DOCENTES

- Clases expositivas
- Prácticas de laboratorio
- Trabajo individual (no presencial)
- Trabajo en grupo
- Pruebas de respuesta corta (Test)
- Pruebas de respuesta larga (Examen Final)

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

La privacidad de los datos es la adaptación a la Sociedad de la Información del derecho fundamental a la privacidad y la vida privada. Este curso se centra en diferentes aspectos tecnológicos para garantizar la privacidad de la información, que incluyen:

- Presentar las principales métricas de privacidad propuestas en la literatura, tanto aquellas cualitativas como, especialmente, las cuantitativas.
- Evaluar riesgos y analizar tecnologías de privacidad en el campo de control de revelación estadística.
- Evaluar riesgos y analizar tecnologías de privacidad en los sistemas de información personalizados.
- Evaluar riesgos y analizar tecnologías de privacidad en los sistemas de comunicaciones anónimas.
- Comprender el compromiso inherente entre privacidad y usabilidad de los datos.

Por una parte, una aplicación de especial interés es la anonimización de datos, bien sea para cederlos a terceros para que los analicen o incluso publicarlos para que cualquiera pueda realizar los análisis correspondientes. El problema es especialmente relevante cuando los datos originales contienen datos demográficos que pueden servir para identificar a personas, y esta identificación puede conllevar la revelación de información sensible, como salario, afiliación política, religión y estado de salud. Conviene, pues, que los mecanismos de anonimización utilizados minimicen la probabilidad de identificación, a la vez que la perturbación que realicen sobre los datos sea la menor posible para que el análisis posterior de éstos sea significativo.
Por otro lado, en los últimos tiempos estamos asistiendo al surgimiento de una amplia variedad de sistemas de información que adaptan la funcionalidad de intercambio de información para satisfacer los intereses específicos de sus usuarios. La mayoría de estos sistemas de información personalizados capitalizan o se prestan a la construcción de perfiles, ya sea directamente declarados por un usuario o inferidos de la actividad pasada.
Una alternativa para proteger los datos privados de los usuarios consiste en perturbar la información que revelan explícita o implícitamente al comunicarse con un sistema de información. El envío de datos falsos, junto con los datos genuinos del usuario, es un ejemplo ilustrativo de mecanismo perturbador de datos. En este tipo de mecanismos, la perturbación en sí tiene lugar típicamente del lado del usuario. Esto significa que los usuarios no necesitan confiar en ninguna entidad externa como el sistema de recomendación, el proveedor de Internet o los nodos vecinos. Las técnicas de perturbación de datos tienen un coste en la funcionalidad del sistema y la utilidad de los datos, lo que plantea un compromiso entre estos aspectos y la protección de la privacidad.
Una alternativa diferente, que no requiere perturbación de la información y, por lo tanto no tiene necesariamente que afectar a la usabilidad de los datos consiste en la utilización de primitivas criptográficas. Entre estas destaca el cifrado homomórfico y las pruebas de conocimiento cero.
En último lugar se presentan los sistemas de comunicaciones anónimas. En las comunicaciones anónimas, uno de los objetivos es ocultar quién habla con quién frente a un adversario que observa las entradas y salidas del canal de comunicación.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas grupo grande	26,0	20.80
Horas grupo pequeño	13,0	10.40
Horas aprendizaje autónomo	86,0	68.80

Dedicación total: 125 h

CONTENIDOS

1. Introdución

Descripción:
En este primer tema se pretende dar una visión global de la asignatura, haciendo especial hincapié en la motivación y a los requerimientos legales (y también éticos) que sustentan las técnicas de protección de la privacidad.

Objetivos específicos:
- Ofrecer una visión general de la asignatura.
- Motivar la necesidad de proteger la privacidad de los usuarios en los sistemas de información y comunicaciones.
- Introducir el marco legal y las implicaciones éticas que requieren la protección de la privacidad.
- Introducir los campos de aplicación de las tecnologías de privacidad.
- Definir conceptos básicos de la terminología de la asignatura.
- Presentar los modelos de confianza sobre los que se basa la asignatura, identificando los atacantes de los que intentamos protegernos.

Dedicación: 7h
Grupo grande/Teoría: 3h
Aprendizaje autónomo: 4h

2. Anonimización de bases de datos

Descripción:
Este tema trata acerca de la anonimización de bases de datos, posiblemente uno de los campos de aplicación más relevantes actualmente para los estudiantes que se incorporen al mercado laboral. Se presentan técnicas de estimación de riesgo, así como diferentes técnicas y algoritmos de anonimato, tanto para datos categóricos como numéricos. Se hace especial énfasis al compromiso entre privacidad y usabilidad, considerando también métricas de pérdida de usabilidad de los datos anonimizados.

Objetivos específicos:
- Presentar los riesgos de publicación de bases de datos que incluyan información personal de los usuarios.
- Formular el compromiso entre privacidad y usabilidad inherente a las técnicas perturbativas.
- Presentar las principales técnicas de estimación de riesgos de revelado de información privada en bases de datos estadísticas.
- Presentar los algoritmos de anonimato para datos categóricos.
- Presentar los algoritmos de anonimato para datos numéricos.
- Presentar técnicas de generación de datos sintéticos.
- Ser capaz de medir la pérdida de usabilidad de los datos anonimizados.

Dedicación: 24h
Grupo grande/Teoría: 9h
Aprendizaje autónomo: 15h

3. Privacidad en sistemas de información personalizadaTécnologías de privacidrturbad peativas

Descripción:
En este tema se contemplan los mecanismos de privacidad basados en la perturbación de datos que se usan en campos diferentes a SDC. Por una parte, se estudia la problemática de la privacidad en sistemas de información personalizados y en sistemas de recomendación. Por otra parte, se aborda el tema de la privacidad en aplicaciones que tienen acceso a información geo-localizada.

Objetivos específicos:
- Introducir los retos de privacidad en sistemas de recomendación y personalizados, incluyendo sistemas de publicidad en línea.
- Cuantificar el riesgo de privacidad inherente a la creación de perfiles de usuario por terceras partes.
- Presentar diferentes estrategias de generación falsa de consultas, incluyendo TrackmeNot.
- Presentar mecanismos de protección en sistemas basados en anotaciones semánticas.
- Introducir la problemática de privacidad en la localización
- Resaltar el compromiso entre privacidad y usabilidad, inherente a todo sistema basado en perturbación de datos.

Dedicación: 9h
Grupo grande/Teoría: 3h
Aprendizaje autónomo: 6h

4. Sistemas de comunicación anónimos

Descripción:
Aun cuando las comunicaciones vayan cifradas, el atacante puede obtener sensible analizando las cabeceras de los protocolos e infiriendo quien y cuando se está comunicando con quien. Los sistemas de comunicación anónimos, que se estudian en este tema, permiten protegerse frente a estos ataques de análisis de tráfico.

Objetivos específicos:
- Presentar técnicas de análisis de tráfico y justificar la necesidad de comunicaciones anónimas.
- Definir la terminología básica de los sistemas de comunicación anónimos.
- Presentar y analizar los sistemas de comunicaciones anónimos basados en Mix Networks.
- Presentar y analizar los sistemas de comunicaciones anónimos basados en Onion Routing.
- Presentar y analizar los sistemas de comunicaciones anónimos basados en la colaboración de usuarios en redes P2P.

Dedicación: 9h
Grupo grande/Teoría: 3h
Aprendizaje autónomo: 6h

5. Privacidad diferencial

Descripción:
contenido castellano

Objetivos específicos:
En este tema se introducen los principales conceptos relacionados con privacidad diferencial

Dedicación: 23h
Grupo grande/Teoría: 8h
Aprendizaje autónomo: 15h

6. Laboratorio. Prácticas de anonimización de datos

Descripción:
El objetivo principal de las prácticas propuestas es servir de soporte a la asignatura Data Privacy, para que el alumno profundice a través de la experimentación en diferentes aspectos importantes, especialmente los relacionados con la anonimización de datos.

Objetivos específicos:
- Introducción al control de divulgación estadística (SDC)
- Introducción al lenguaje de programación R
- Evaluación del riesgo de divulgación
- Medición de la utilidad de un conjunto de datos.
- Métodos de anonimización para variables categóricas
- Métodos de anonimización de variables numéricas
- Privacidad diferencial

Dedicación: 49h
Grupo pequeño/Laboratorio: 13h
Aprendizaje autónomo: 36h

SISTEMA DE CALIFICACIÓN

- Esta asignatura tiene evaluación de teoría (65%) y de laboratorio (35%).
- La nota de teoría se obtiene mediante dos controles realizados durante el curso. Adicionalmente, hay un examen final para quien no supere los controles.
- La nota de laboratorio se obtiene mediante la realización de los ejercicios prácticos y la entrega de las memorias correspondientes.

BIBLIOGRAFÍA

Básica:
- Torra i Reventós, Vicenç. Data privacy : foundations, new developments and the big data challenge. Skövde: Springer, 2017. ISBN 9783319573564.
- Templ, Matthias. Statistical disclosure control for microdata : methods and applications in R [en línea]. Cham, Switzerland: Springer International Publishing AG, 2017 [Consulta: 05/07/2021]. Disponible a: https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=4855639. ISBN 9783319502724.
- Hundepool, Anco; Domingo-Ferrer, Josep. Statistical disclosure control [en línea]. Chichester, West Sussex, United Kingdom: John Wiley & Sons Inc, [2012] [Consulta: 05/07/2021]. Disponible a: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118348239. ISBN 9781118348239.