Seguridad y Privacidad de Datos

Introducción

Datos de Contacto Profesor

- Nombre: Patricio Galdames
- Departamento Sistemas de Información DSI
- Concepción
- Email: pgaldames@ubiobio.cl
- Fono: 41-3111519

Objetivos del Curso

- **Presentar los desafíos fundamentales involucrados en la privacidad y seguridad de los datos, enfocándose en el objetivo final de permitir análisis e inferencia útil pero buscando minimizar la revelación de información confidencial.**
  1. **Aplicar técnicas y herramientas criptográficas para proteger la privacidad**
  2. **Identificar las limitaciones algorítmicas de la privacidad de los datos, incluido el costo en términos de tiempo y espacio para concluir que algoritmo es el más adecuado en un contexto dado.**
  3. **Proponer un problema de interés investigativo en el área de la privacidad de datos**

Requisitos: Algebra, algoritmos o cursos similares

Evaluaciones

- 2 Certámenes (20 % cada uno, Total: 40 %)
  - Presentación de dos artículos acordados con el profesor, ie., Un certamen => dos artículos.
- Trabajo de Investigación (60 %)
  - Propuesta de Tópico (5%). Descripción de motivación acompañada de al menos un par de referencias (1 pagina)
  - Informe (35%)(12 paginas mínimo, 20 referencias académicas)
  - Presentación oral (20 %)

Posible Fechas Evaluaciones

- Certamen 1: Inicios de Noviembre
- Certamen 2: Inicios de Diciembre
- Presentaciones:
  - Viernes 28 de Diciembre 2018
- Fecha Entrega Propuesta Tema:
  - 15 de Octubre 2018.

Análisis de un articulo

- Presentación oral debe destacar al menos los siguientes aspectos:
  - Objetivo del trabajo
  - Contexto del problema planteado
  - En que se distingue el problema planteado de otros abordados
  - En que se distingue (idea o ideas claves) de la solución planteada
  - Métricas de evaluación empleadas
  - Trabajo futuro

Búsqueda de artículos

- (Algunas)Top conferencias en CS
  - IEEE Symposium on Security and Privacy
  - CCS: ACM Conf on Comp and Communications Security
  - INFOCOM: Annual Joint Conf IEEE Comp & Comm Soc
  - MOBICOM: ACM Intl Conf on Mobile Computing and Networking
  - (DB) VLDB: Very Large Data Base
- Top Journals in CS
  - IEEE Transactions on Information Forensics and Security
  - Elsevier Computers and Security
  - IEEE Security and Privacy
- Web of Science, Scielo, ResearchGate
  - Solicite al profesor articulo que no ha podido descargar gratuitamente.

Lecturas Propuestas

- Descargar artículos desde **ADECCA**
  - Solicitar al profesor registro en plataforma enviando email desde cuenta alumnos.ubiobio.cl (solo si no esta registrado)
- Adecca:
  - Artículos
  - Consultas, discusiones
  - Novedades del curso, etc

Proyecto de Curso

- Informe
  - Mínimo 12 paginas que incluya:
    - Motivación del tema seleccionado
    - Estudio del arte con mínimo de 20 referencias a **artículos académicos** (no valen referencias a paginas web, ni wikipedia)
    - Artículos citados deben ser al menos en 80% de distintos autores o grupos
    - Matriz de resumen señalando métricas de comparación
    - Análisis de la matriz
    - Estudio futuro y posible propuesta
- Presentación oral
  - 15 a 20 minutos
  - Presentar los principales resultados de los papers leídos o seleccionar dos de los artículos de mayor impacto en el área escogida
  - Presentar matriz de comparación de artículos
  - Presentar nuevas ideas de investigación

Análisis de un articulo

- Escriba un resumen de no mas de una plana donde destaque:
  - Objetivo del trabajo
  - Contexto del problema planteado
  - En que se distingue el problema planteado de otros abordados
  - En que se distingue (idea o ideas claves) de la solución planteada
  - Métricas de evaluación empleadas
  - Trabajo futuro

Preguntas a abordar en el curso

- **¿Cómo realizar un análisis de datos consciente de la privacidad?**
  - ¿Como podemos definir formalmente “la privacidad”?
  - ¿Qué limites teóricos existen acerca de la cantidad de información que puede ser hecha publica en una base de datos sujeto a preservar “la privacidad”?

Preguntas a abordar en el curso

- ¿Cómo diseñar algoritmos eficientes que hagan uso de información privada?
- ¿Cómo los agentes económicos debieran de actuar sobre su privacidad?
  - ¿Cómo podemos diseñar subastas y otros mecanismos para clientes conscientes de su privacidad?

Temas del Curso

1. Criptografía
   1. Publica, Privada y de Clave Simétrica
   2. Criptografía sin Certificados, de Identidad
   3. Firma Digital Personal y Grupal
2. Definiciones de Privacidad
   1. Privacidad diferencial, seguridad semántica
   2. K-anonimato, métodos basados en clúster, conjuntos de datos sintéticos
   3. Privacidad Diferencial, Privacidad para Redes Sociales
3. Privacidad en Bases de
   1. PEKS[1], búsqueda en BD cifradas
   2. Métricas de privacidad y seguridad, teoría de medidas, modelado del riesgo
   3. Auditoría y restricción de consultas
   4. Medidas de utilidad para el análisis de datos privados
   5. Privacidad en Bases de Datos con Información Geográfica
4. Recuperación de información privada
   1. Transferencia inconsciente y memoria olvidadiza
   2. Protocolos de Conocimiento Cero.
   3. Protocolos multipartidarios, computación segura, circuitos ilegibles
   4. Criptomonedas

1:Public Key Encryption with Keyword Search

Índice

- Introducción
- Perspectivas de privacidad
- Divulgación de información
- Minería de datos
- Ejemplos de reacción negativa del consumidor

Percepción sobre la Internet

Diseños de Algoritmos

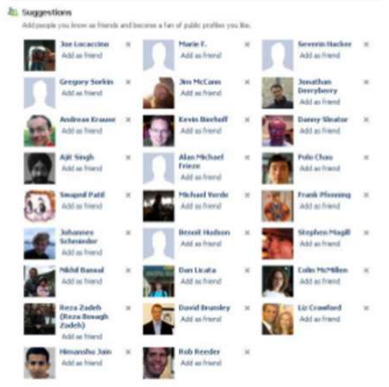
- La computación no es la única restricción
- Debemos enfrentar grandes repositorios de datos
  - Datos que ***pertenecen*** a otras personas
  - Y que debemos proteger su privacidad
  - Debemos convencerlos de dar sus datos con sinceridad

Diseños de Algoritmos

- Use logs de búsqueda para recomendar y completar consultas

Diseños de Algoritmos

- Encuentre componentes conectadas cercanas en una red social



Diseños de Algoritmos

- Decida que avisos comerciales mostrar basado en los datos del usuario y en las búsquedas previas realizadas por otros usuarios

Pero ¿Qué es la privacidad?

Definición de privacidad

- **La privacidad esta relacionada con la noción de acceso**
- Acceso
  - La proximidad física de una persona
  - Conocimiento sobre una persona
- La privacidad es una "zona de inaccesibilidad"
- Las violaciones a la privacidad son una afrenta a la dignidad humana
- Pero ….demasiada privacidad individual puede perjudicar a la sociedad
  - *¿Donde dibujar la línea?*

Daños a la privacidad

- Cubrir actividades ilegales o inmorales
- Sobrecarga el núcleo de una familia
- Familias disfuncionales ocultas
- La gente en la periferia de la sociedad puede ser ignorada

Beneficios de la Privacidad

- Crecimiento individual
- Responsabilidad individual
- Libertad para ser tú mismo
- Crecimiento intelectual y espiritual
- Desarrollo de relaciones amorosas, de confianza, cariñosas e íntimas

Pero ¿Qué No es la privacidad?

- La privacidad no es restringir la realización de consultas solamente a grandes poblaciones
  - "¿Cuál es el salario promedio de un profesor de la UBB?"
  - "¿Cuál es el salario promedio de un profesor de la UBB que no se llame Patricio Galdames?"

¿Que compromiso a la privacidad se genera con estas dos consultas?

Pero ¿Qué No es la privacidad?

- La privacidad no se limita a hechos "comunes o ordinarios"
  - Las estadísticas de los hábitos de consumo de pan de Alice: Por 20 años ella compro regularmente pan y luego nunca mas lo hizo
    - ¿Diabetes tipo 2?

Pero ¿Qué No es la privacidad?

- La privacidad no es anonimización
  - Lograr el anonimato es difícil
  - **Problema:** Ataque de correlación con información auxiliar
    - Se correlacionó la base de datos publica IMDB para re identificar usuarios [Narayanan, Shmatikov 2007][1]
    - El premio Netflix fue cancelado
  - No sabemos lo que un atacante (adversario) sabe, o podría saber en el ***futuro***

1: Narayanan, Arvind; Shmatikov, Vitaly. "How To Break Anonymity of the Netflix Prize Dataset". arXiv:cs/0610105

Casos de Netflix

- Netflix libera datos de clientes, reemplazando nombres con seudónimos (Oct 2006)
  - http://bigdataways.com/2016/05/26/el-revolucionario-concurso-de-netflix/
- Netflix publica un polémico tuit y despierta preocupación sobre la privacidad de sus usuarios (Nov. 2017)
  - http://www.t13.cl/noticia/tendencias/bbc/espectaculos/el-polemico-tuit-en-el-que-netflix-se-burlo-de-algunos-de-sus-usuarios-y-desperto-preocupacion-sobre-la-privacidad

Casos en Chile

- Chile: Servel publicó datos de 13 millones de ciudadanos (Ago 2012)
  - https://www.fayerwayer.com/2012/08/chile-servel-publico-datos-personales-de-13-millones-de-ciudadanos/
- Masiva fuga de datos de más de 14 mil tarjetas de crédito alerta a bancos y al regulador (Jul. 2018)
  - http://www.elmercurio.com/Inversiones/Noticias/Analisis/2018/07/26/Masiva-fuga-de-datos-de-mas-de-14-mil-tarjetas-de-credito-alerta-a-bancos-y-al-regulador.aspx

Pero ¿Qué No es la privacidad?

- La privacidad no es anonimización
  - **El anonimato ayuda pero No es suficiente**
    - La recopilación de registros médicos de un centro de asistencia publica y sus fechas podrían corresponder a un numero reducido de dolencias
    - El conocimiento (¿de un vecino?) que Alice fue a ese centro medico no identifica su registro, pero nos permite concluir que ella sufre de alguna de esas pocas dolencias.

Fuga de Datos de Búsquedas en AOL

- AOL libero 20 millones de términos de búsqueda de sus usuarios (Ag. 2006)
- No hubo filtrado de la información,
  - Muchos términos tenían datos personales identificables.
  - En 3 días AOL los borró, pero los datos ya se habían reproducido en varios sitios más.
  - ¿Costo? 5 mil dólares a cada usuario afectado.

¿Qué es la privacidad?

- Libertad de no sufrir algún daño.
- Definición de privacidad, intento 1:

  Un análisis de un conjunto de datos D es privado si el analista de los datos no sabe más sobre Alice después del análisis de lo que sabía sobre Alice antes del análisis.

¿Qué es la privacidad?

- **Problema:** Es imposible de lograr si hay información auxiliar.
  - Asuma que una compañía de seguros sabe que Alice es fumadora
  - Un analista que revela que fumar y el cáncer de pulmón están correlacionados podría causar que ¡le subieran el costo del seguro a Alice!.
- ¿Fue la privacidad de Alice violada?
  - Este es un problema que afectaría a Alice aun cuando sus datos no estuviesen en la base de datos
  - **Esa es la clase información que nos gustaría aprender.**

¿Qué es la privacidad?

- Definición de privacidad, intento 2:

  Un análisis de un conjunto de datos D es privado si el analista de los datos sabe casi nada más sobre Alice después del análisis de lo que hubiera sabido si hubiera llevado a cabo el mismo análisis en una base de datos idéntica con los datos de Alice eliminados.

  A esta definición se le conoce como privacidad diferencial [Dwork-McSherry-Nissim-Smith06]

Mas adelante volveremos a esta definición

Las Tecnologías de la información Erosionan la Privacidad

- La recolección, el intercambio, la combinación y distribución de la información es más fácil mas que nunca y compromete a la privacidad
- Scott McNealy: "De todos modos, usted tiene cero privacidad. Acéptelo.
- Vamos a realzar el "rastro electrónico" de información que dejamos detrás y lo que otros pueden hacer con esta información

Facebook, Google (Big Data)

- Charla de **Martin Hilbert** (Profesor de la Universidad de California, Davis y experto en comunicación y desarrollo digital)
  - "**Big Data: grandes datos, grandes oportunidades, grandes amenazas**"
  - Se realizo el 10/11/2017 en la sede central de la Contraloría General de la República de Chile.
  - https://www.youtube.com/watch?v=A03fpNk8P50&feature=youtu.be&t=31m52s

Facebook, Google (Big Data)

- "Facebook, Google, Netflix, etc. son empresas de datos, **"su negocio son los datos que tienen"."**
- **Perfil de usuario**
  - Con 100 "Likes" podían describir la personalidad
  - Con 150 "me gusta" el algoritmo podía predecir el comportamiento de esa persona mejor que su pareja
  - Con 250 "likes" de Facebook, el algoritmo conocía su personalidad mejor que él mismo" explica.
- **Objetivos:**
  - Cómo satisfacer las diferentes demandas de los clientes, a dónde están, qué hacen, qué quieren, cuáles son sus hábitos, y hasta sus emociones vinculadas al consumo de tu producto.
  - Predecir con 90% o 95% de precisión dónde vas a estar en dos meses, en qué momento del día, cruzarlos con todos tus datos y saber perfectamente qué venderte, dónde, cuándo y cómo

Facebook, Google (Big Data)

- Mas sobre el tema de la charla:
  - http://www.theclinic.cl/2018/03/27/la-entrevista-the-clinic-anticipo-escandalo-facebook/
  - http://www.theclinic.cl/2018/03/27/caerse-raja-estos-los-datos-le-entregado-facebook-google-llegar/
  - http://www.bbc.com/mundo/noticias-internacional-39511606
  - http://comunicaciones.uc.cl/martin-hilbert-guru-del-big-data-pagamos-con-nuestra-privacidad-pero-debemos-entender-que-nuestra-comunicacion-esta-distorsionada-por-intereses-comerciales/ (2018)

Lectura y Video

- Comentarios
- Cisco “Network Intuitive”
- https://www.youtube.com/watch?v=a_3-pStjHs0