Seguridad, privacidad y aspectos legales

Álvaro López García

Grupo de Computación Avanzada y e-Ciencia
Instituto de Física de Cantabria (IFCA) - CSIC-UC

Máster universitario en ciencia de datos / Master in Data Science

Programa de la asignatura I

Descripción

En esta asignatura el estudiante conocerá los mecanismos básicos para proporcionar un acceso seguro a los recursos en la red, cómo tener en cuenta la debida protección de datos personales, y las condiciones de explotación de datos y software mediante los diferentes tipos de licencias existentes. Asimismo se abordarán aspectos éticos en la ciencia de datos.

Programa de la asignatura II

Bloques principales:

1. Parte I: Introducción a la seguridad.
   - Conceptos generales: privacidad, trazabilidad, anonimización, integridad, repudia.
   - Tecnologías para la protección de la información y privacidad.
   - Identidad digital y acceso a recursos. Autenticación y Autorización.
2. Parte II: Aspectos legales.
   - Protección de datos personales.
   - Licencias y uso de software y de datos.
3. Parte III: Ética en la ciencia de datos.
4. Aplicación en el entorno Open Science.

Información General

Profesores.

- Álvaro López García , CSIC, (aloga@ifca.unican.es).
- Pablo Orviz Fernández, CSIC, (orviz@ifca.unican.es).
- Juan Antonio Losada, CIC.

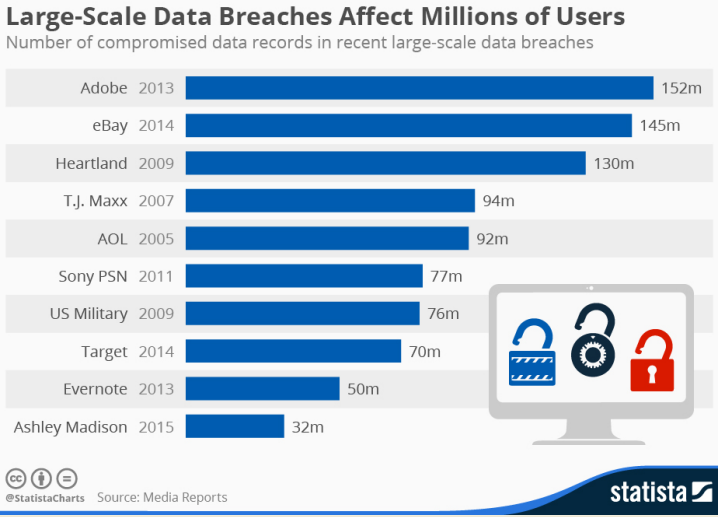
Evaluación.

- Examen: 50 %
- Valoración de informes y trabajos: 30 %
- Seguimiento de actividades presenciales, evaluación continua: 20 %

Calendario

L	M	X	J	V
			16-18h Introducción	16-18h Tecnologías
16-18h Tecnologías	16-18h Tecnologías	16-18h Aspectos legales	16-18h Identidad digital	
16-18h Identidad digital	16-18h Ética	16-18h Licencias	16-18h Ética	
16-18h Aplicación en ciencia				

- Los datos pueden tener un alto valor.
  - Económico (por ejemplo, transacciones financieras),
  - personal (por ejemplo, datos médicos),
  - político (por ejemplo, sesgo para modificar intención de voto),
  - etc.
- La información y los datos representan hoy en día poder.
- Los datos están en riesgo: pueden ser robados, saboteados, manipulados, divulgados sin permiso (de forma consciente o no), etc.
- Los datos pueden ser usados de forma no correcta o no ética.


Large-Scale Data Breaches Affect Millions of Users
Number of compromised data records in recent large-scale data breaches
Adobe 2013 152m
eBay 2014 145m
Heartland 2009 130m
T.J. Maxx 2007 94m
AOL 2005 92m
Sony PSN 2011 77m
US Military 2009 76m
Target 2014 70m
Evernote 2013 50m
Ashley Madison 2015 32m
@StatistaCharts
Source: Media Reports
statista

- ¿Por qué me vigilan, si no soy nadie? de Marta Peirano en TEDxMadrid: `https://www.youtube.com/watch?v=NPE7i8wuupk`

- François Chollet, creador de Keras.
  - Hilo en twitter: `https://twitter.com/fchollet/status/976563870322999296`
  - Hilo completo: `https://threadreaderapp.com/thread/976563870322999296.html`

Parte I

Introducción a la seguridad

Tabla de contenidos

Conceptos Generales

Tabla de contenidos

Definición

Medidas preventivas y reactivas de las personas, organizaciones y sistemas tecnológicos que permiten resguardar y proteger la información manteniendo integras sus propiedades de seguridad (como confidencialidad, disponibilidad e integridad).

- Protección de los activos en un sistema informático.
- Activos: información (datos, en nuestro caso).
- Tipos de protección:
  - Prevención: tomar medidas para preservar la seguridad.
  - Detección: detectar cuando, cómo y quién ha roto la seguridad de un activo.
  - Reacción: mitigar y reparar los daños causados.
- Propiedades de seguridad de los activos: **confidencialidad**, **autenticidad**, trazabilidad, anonimización, **integridad**, **no repudia**, disponibilidad.

Confidencialidad y Privacidad

Confidencialidad

Prevención de la difusión no autorizada (no divulgación) de la información (secreto).

Privacidad

Derecho a no ser objeto de intrusión de terceros (privado).

- Conceptos relacionados pero no similares.
- Se puede preservar una pero no la otra.
- Ejemplo: una empresa cede de forma consciente datos de sus usuarios a un tercero, sin nuestro consentimiento.
  - Se preserva la confidencialidad entre las empresas.
  - No se conserva la privacidad.
- Métodos: autenticación y autorización, cifrado, control de acceso.
- Definir quién puede acceder a y/o modificar qué.
- Conceptos relacionados: Anonimización.

Autenticidad

Definición

Identificación y garantía del origen de la información.

- Asegurar que quién origina la información es quien realmente dice ser.
- Evitar la suplantación de la identidad.
- Relacionado con la integridad.

Definición

No alteración, manipulación, modificación o corrupción (intencional o no) de la información cuando es manejada (transmitida, almacenada, procesada, etc.) por un sistema de información.

- Asegura que la información sea correcta y sin errores.
- Asegura que la información no pueda ser modificada sin permiso
- Asegurar la corrección e invarianza de la información. Confiabilidad
- Método: hash.
- Ejemplo: paquetes linux

Definición

Guardar trazas de todas las acciones susceptibles de comprometer la seguridad de un sistema, para asegurar que se puede identificar al culpable o implicado.

- Es imposible prevenir frente a todo tipo de acciones.
  - Acciones autorizadas pueden violar la seguridad (error de diseño).
  - Acciones no autorizadas.
- El sistema debe guardar trazas de todas las acciones: traza de auditoría.
- Prerequisitos: identificación de usuarios.
- Quién hizo qué.

Definición

Delimitar y suprimir la información concreta que permite identificar a los individuos concretos.

- Anonimato, pseudoanonimato (reversible), anonimato.
- Relacionado con privacidad y confidencialidad.
- Conflicto con autenticidad, trazabilidad y no-repudio

No repudia

Definición

Proveer evidencia de que un evento específico ha ocurrido.

- Comprobar la identidad del emisor de un mensaje.
- Evitar que alguna de las partes de una comunicación niege una acción.
- Métodos: firmas digitales. Servicios de no repudia de entrega (e.g. burofax).

Disponibilidad

Definición

Asegurar que la información está accesible cuando se requiera acceder a ella.

- Prevenir que un atacante impida el acceso a los activos de forma legítima.
- Métodos: redundancia de disco, servidores, localidad, backups.

Ejemplos: Seguridad en un hospital

- Integridad: asegurar que el tratamiento es el adecuado.
- Confidencialidad: prevenir que los datos de los pacientes se hagan públicos.
- Privacidad: asegurar que solo el personal habilitado puede acceder al historial de un paciente.
- Imposibilidad de usar registros anónimos.
- Consentimiento informado: ¿qué se puede hacer con mis datos?

Ejemplos: Seguridad en un entorno científico

- Integridad y trazabilidad: asegurar que los datos de los experimentos son correctos y no son manipulados (datos primarios, secundarios, terciarios).
- Confidencialidad: prevenir que los datos se hagan públicos antes de que finalice un embargo.
- Seguridad en el acceso.
- Disponibilidad.

Ejemplos

- La seguridad se puede ver comprometida aunque tomemos medidas para protegerla, por ejemplo: la anonimización puede no ser suficiente.
- Ejemplo: Resonancias magnéticas del cráneo pueden ser suficientes para identificar a una persona.
- Ejemplo: Publicación de delitos anonimizados, con data aggregation para obtener información de la persona.

Ejemplos: Autenticidad e integridad

- En 2008 un atacante consiguió romper la seguridad de las distribuciones Fedora y RedHat.
  - Se firmaron paquetes falsos como legítimos (OpenSSH).
  - Esto puede comprometer la seguridad de cientos de miles de servidores.
  - `https://lwn.net/Articles/295406/`
- En 2010 un atacante consiguió acceder a varios servidores de `https://kernel.org`.
  - La integridad del kernel no estuvo en riesgo, pero hubo un problema de reputación.
  - `https://pastebin.com/BKcmMd47`

Tecnologías para la protección de la información y privacidad

Tabla de contenidos

Definición

«Campo que se encarga del estudio de los algoritmos, protocolos y sistemas que se utilizan para dotar de seguridad a las comunicaciones, a la información y a las entidades que se comunican.» J. Pastor Franco, M. A. Sarasa López, J. L. Salazar Riaño. Criptografía digital: fundamentos y aplicaciones.

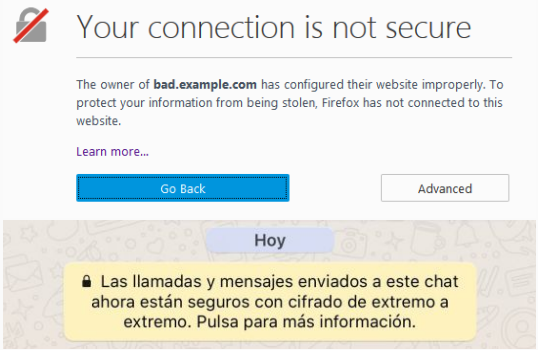
Autenticación asegurar que alguien es quien dice ser.

Confidencialidad asegurar que nadie no autorizado puede acceder a la información.

Integridad asegurar que los datos no pueden ser manipulados.

No-repudia asegurar que nadie puede denegar que haya generado una información.

- Presente en nuestro día a día, aunque de forma imperceptible a veces.
- Transacciones financieras, tarjetas de crédito, etc.
- Certificados digitales y DNI digital.
- Conexiones seguras (https).
- Firma digital de documentos oficiales.
- Transferencias de ficheros.
- Cifrado de discos



A grandes rasgos, los sistemas criptográficos se basan en 3 tipos de algoritmos.

Clave privada o simétrica, donde se utiliza una sola clave.

Clave pública o asimétrica, donde se utilizan un par de claves, una pública y otra privada

Función hash o message-digest.

Los sistemas entran dentro de dos categorías.

- Computacionalmente seguros.
  - No hay recursos y/o tiempo suficientes para romperlos (hoy).
  - Ejemplo: tiempo estimado para romper el cifrado de un certificado SSL de 2048-bit: 6.440 trillones de años (`https://www.digicert.com/TimeTravel/`).
  - Son la mayoría de sistemas de cifrado.
- Seguros de manera incondicional.
  - Nunca se pueden romper.
  - Libretas de un solo uso.

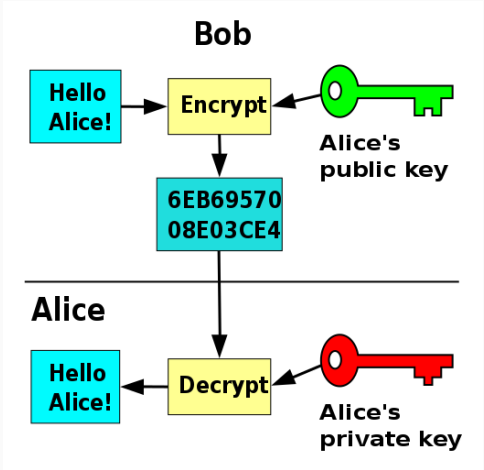
Clave simétrica

- Clave secreta para cifrar la información.
- Misma clave para descrifrar la información.
- La clave tiene que ser conocida por ambos extremos de la comunicación.
- Hay que compartir la clave, debe existir un contacto previo.
- Ventajas:
  - Eficiente.
  - Sencillo de implementar.
- Desventajas:
  - Dificultad para compartir claves.
  - Número de claves elevado.
  - No hay posibilidad de no-repudio.
- DES, Blowfish, AES, etc.

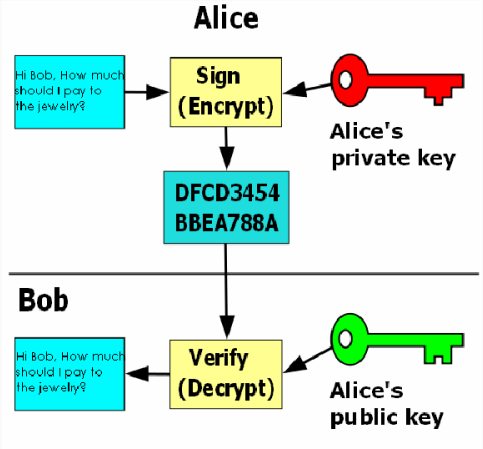
Clave asimétrica

- Se usa un par de claves: pública y privada.
- La clave privada es secreta, la clave pública es pública.
- Ambas claves están relacionadas matemáticamente.
- Una información cifrada con una clave pública solo puede ser descifrada con su clave privada.
- Una información cifrada con una clave privada solo puede ser descifrada con su clave pública (firma).

Cifrado



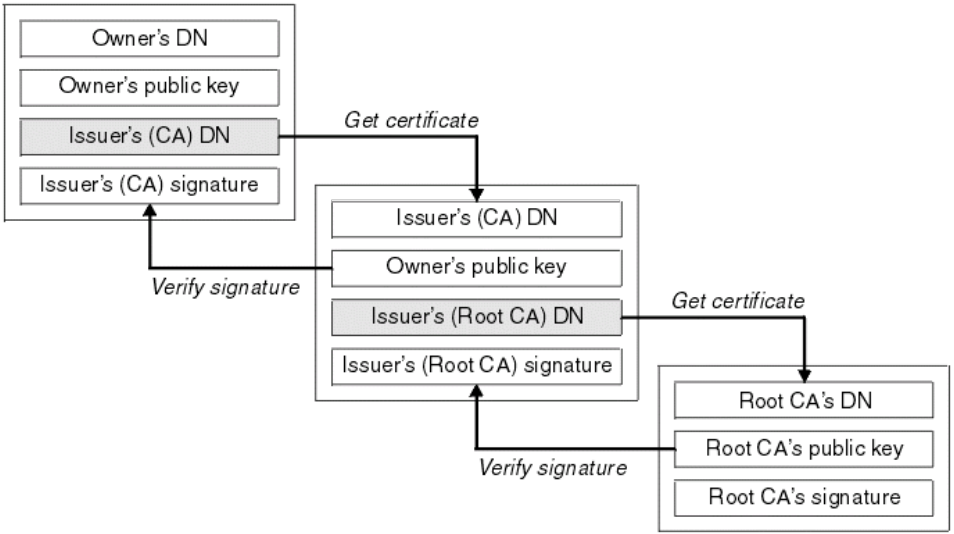
Firmado



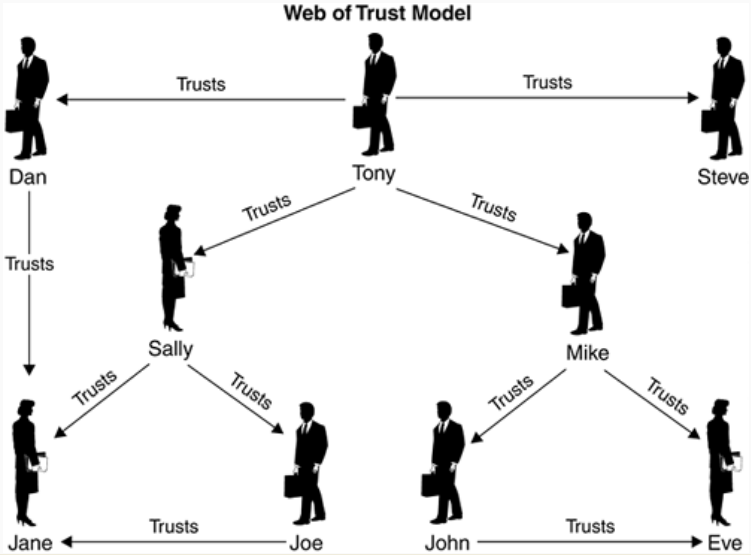
Clave asimétrica

- Ventajas:
  - No hay que compartir ningún secreto.
  - Las claves públicas se puden distribuir libremente.
  - No hay por qué establecer un contacto previo.
  - Firma digital.
- Desventajas:
  - No es eficiente.
  - El tamaño de las llaves tiene que ser más grande.
  - ¿Como asegurar que la clave pública de alguien es realmente suya?
- Sitemas: RSA, Diffie-Hellman, DSA, El-Gamal.

- Entidades encargadas de certificar (firmar) las claves públicas.
- Después de generar un par de claves, el usuario demuestra su identidad frente a una CA.
- El firmado se hace offline.
- La confianza está basada en la autoridad.
- Hay CAs de confianza, instaladas por defecto (navegador, sistema operativo).
- El firmado puede ser jerárquico (root authority).


Owner's DN
Owner's public key
Issuer's (CA) DN
Issuer's (CA) signature
Get certificate
Issuer's (CA) DN
Owner's public key
Issuer's (Root CA) DN
Issuer's (Root CA) signature
Verify signature
Get certificate
Root CA's DN
Root CA's public key
Root CA's signature
Verify signature

- Utilizado en PGP (Pretty Good Privacy).
- No existen autoridades de certificación, sino que son los usuarios los que firman.
- La confianza es transitiva


Web of Trust Model
Trusts
Trusts
Dan
Tony
Steve
Trusts
Trusts
Trusts
Sally
Mike
Trusts
Trusts
Trusts
Trusts
Jane
Trusts
Joe
John
Trusts
Eve

Hash criptográficos

- Message-digest.
- A partir de una información de tamaño arbitrario, se produce una salida de un tamaño fijo (digest o hash).
- Cifrado de un solo sentido.
- Propiedades de los hash criptográficos
  - No se puede conocer la información a partir de hash.
  - No se puede encontrar una información arbitraria que genere un hash determinado.
  - No debe haber colisiones.
  - Sensible a los cambios en la entrada.
- Usos: Firmas digitales, control errores, checksum ficheros, cifrado de contraseñas, etc.
- MD1-5, SHA1, CRC, etc.

```
alvaro@torio:~/w/talks/master/seguridad $ sha1sum talk.pdf
55aade398b78da7091b2d8737b1e199afda5ee4e  talk.pdf
alvaro@torio:~/w/talks/master/seguridad $ md5sum talk.pdf
be3a89d83682f06e31d342caa967e9c7  talk.pdf
```

Preguntas?

Identidad digital: Autenticación y Autorización

Tabla de contenidos

Identidad digital: Autenticación y Autorización

Autenticación

Identidad digital: Autenticación y Autorización

Autorización

Parte II

Aspectos legales

Tabla de contenidos

Protección de datos personales

Tabla de contenidos

Licencias y uso de software y de datos

Tabla de contenidos

Parte III

Ética en la ciencia de datos

Tabla de contenidos

Parte IV

Aplicación en el entorno Open Science.

Tabla de contenidos

Preguntas?