Seguridad, privacidad y aspectos legales

Tecnologías para la protección de la información y privacidad

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

Parte I

Tecnologías para la protección de la información y privacidad

Tabla de contenidos

Fundamentos criptográficos

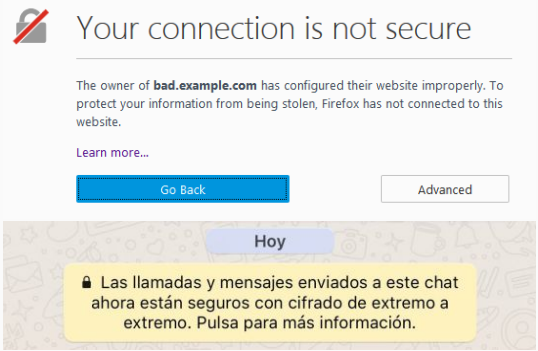
Criptografía I

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.

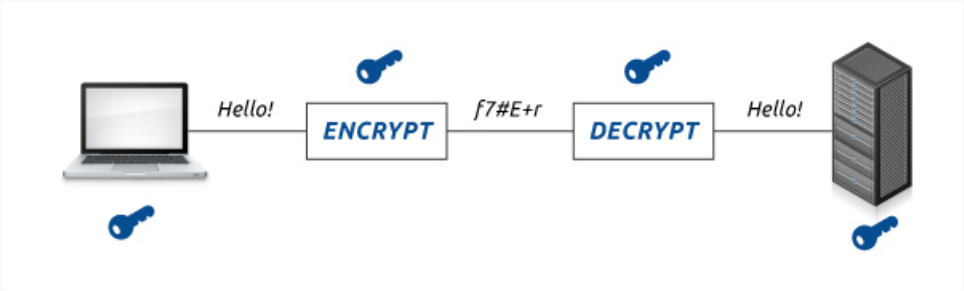
Los sistemas entran dentro de dos categorías.

- Computacionalmente seguros.
  - Se basan en la intratabilidad de ciertos problemas matemáticos (problemas que se pueden resolver en teoría, pero que en la práctica son imposibles de solucionar).
  - 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.
  - 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 descifrar la información.
- La clave tiene que ser conocida por ambos extremos de la comunicación.
- Hay que compartir la clave, debe existir un contacto previo.
- Normalmente tiene una longitud de 128 o 256 bits.
- Ventajas:
  - Eficiente (tamaño de clave pequeño).
  - Sencillo de implementar.
- Desventajas:
  - Dificultad para compartir claves.
  - Número de claves elevado.
  - No hay posibilidad de no-repudio.
- DES, Blowfish, AES, etc.

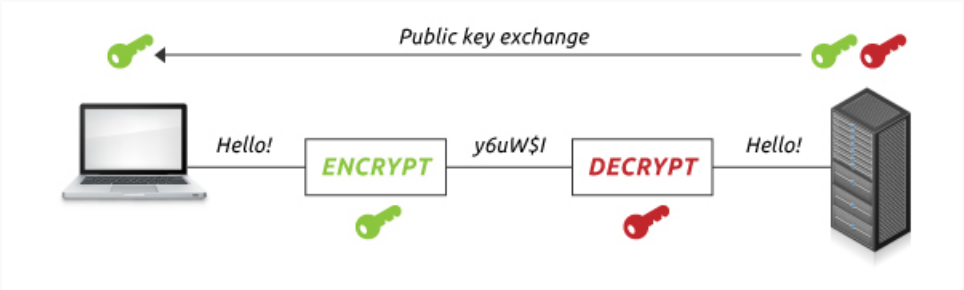
Clave simétrica



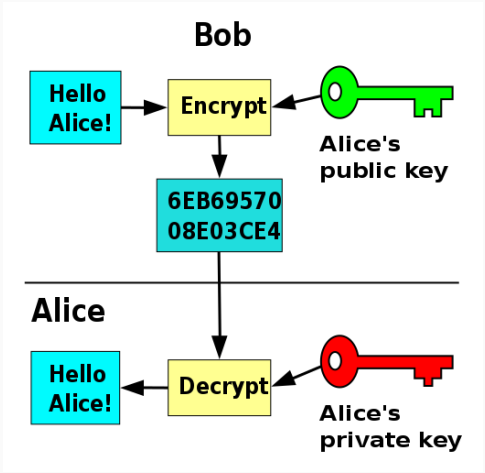
Clave asimétrica

- Se usa un par de claves: pública y privada.
- La clave privada es secreta, la clave pública es pública.
- Ambas claves están relacionadas matemáticamente (RSA o ECC).
- 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).
- Normalmente tiene una longitud de 1024 o 2048 bits.

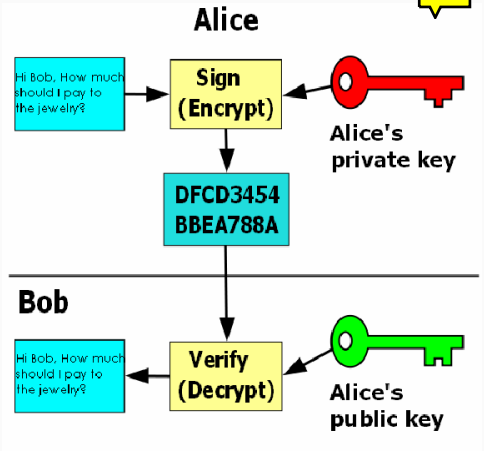
Clave asimétrica



Cifrado



Firmado

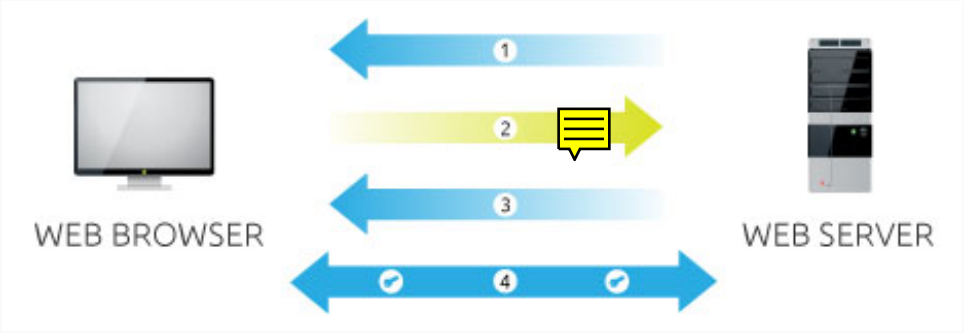


Clave asimétrica

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

- Problema: el cifrado asimétrico es mucho más lento.
- ¿Qué sucede con las comunicaciones con un sistema? (p.e. https)
- Solución: usar una combinación de las dos.
  - Se usa la clave asimétrica para comunicar de forma segura una clave simétrica (generada automáticamente).
  - Se usa la clave simétrica para el resto de la comunicación.

Combinaciones de sistemas



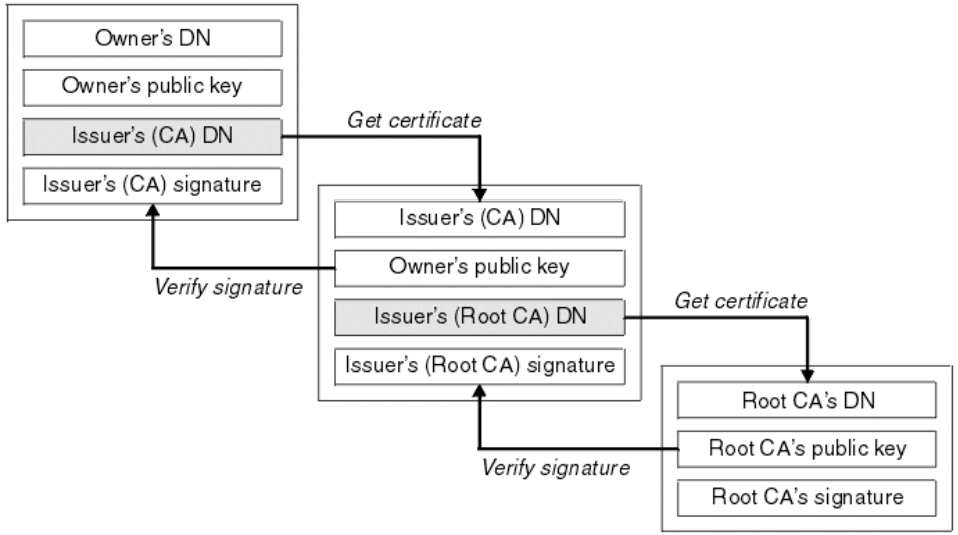
Métodos propagación de la confianza I

- Es necesario establecer la confianza en una clave pública.
- Idealmente, podríamos comunicarla uno a uno (imposible en la práctica).
- Diferentes esquemas:
  - Autoridades de Certificación (CA).
  - Red de confianza.
  - ...

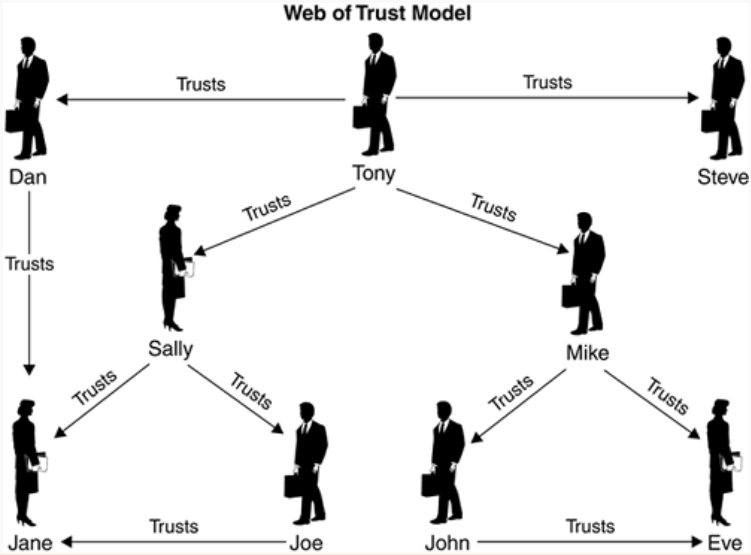
- Entidades encargadas de certificar (firmar) las claves públicas.
- Después de generar un par de claves, el usuario demuestra su identidad frente a una CA.
- El firmado se hace offline.
- La confianza está basada en la autoridad.
- Hay CAs de confianza, instaladas por defecto (navegador, sistema operativo).
  - Ejemplo: FNMT en Mozilla (Firefox): `https://bugzilla.mozilla.org/show_bug.cgi?id=435736`
- El firmado puede ser jerárquico (root authority).
- Estándar: X.509.

Métodos propagación de la confianza II

Autoridades de Certificación (CA)


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).
  - Programa de cifrado que sigue el estándar OpenPGP (RFC 4880).
  - Desarrollado por Phil Zimmermann en 1991.
  - GnuPG: implementación libre de OpenPGP.
- 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
Joe
John
Eve
Trusts
Trusts

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
```

Fundamentos criptográficos

Ejercicios prácticos

Tarea Escribir un mensaje para `aloga@ifca.unican.es`, **cifrarlo, firmarlo** y subirlo a Moodle (Tarea cifrado).

Identidad digital: Autenticación y Autorización

Identidad

- Conjunto de rasgos que hace ser a una persona quien es.
- Conjunto de rasgos que distinguen a un individuo del resto.
- Diferentes ámbitos, diferentes identidades, diferentes características.
  - Ámbito biológico, identidad física: rasgos y características físicas.
  - Ámbito legal, identidad oficial (legal): DNI.
  - Ámbito familiar.
  - Ámbito digital.
- Es un factor clave cuando se trata de establecer una relación entre dos partes.
- Hasta hace algunos años solo era necesario gestionar nuestra identidad con respecto a nuestro entorno físico más cercano.
- Hoy en día no es así: compras por Internet, acceso a recursos on-line, aplicaciones móviles, etc.

Identidad digital I

- Existen diferentes representaciones de un individual en el mundo digital.
- Según el ámbito o dominio una identidad tendrá unas características u otras.
- No tiene por que existir una correspondencia (ni validez) de identidades entre dominios.
- Tipos de identidad, según el dominio donde se va a utilizar.
  - Sector público.
    - Utilización de credenciales que representen a un individuo real unívocamente.
    - Validación de la identidad real, física, del individuo.
    - Objetivo: Validación identidad física e identidad digital. Correspondencia 1:1.
    - Credenciales generadas por autoridades oficiales, con validez legal.
    - Ejemplos: Fábrica Nacional de Moneda y Timbre, Ministerio del Interior.

Identidad digital II

- Identidad corporativa.
  - Credenciales generadas por la corporación (empresa, organismo) de forma interna, ligadas a una identidad física.
  - Pueden incorporar atributos de otras identidades (validados o no) e incluso añadidos por el usuario (sin validar).
  - Objetivo: Validar un usuario en el dominio de la aplicación.
  - No hay correspondencia 1:1. Depende de que se utilice para identificar al usuario (por ejemplo, cuenta bancaria).
  - Validez legal o no.
- Identidad social.
  - Identidad sin correspondencia con una identidad física.
  - Validación no necesaria.

Agentes involucrados

- No solo las personas tienen identidad digital, también hay diferentes agentes que puede tener identidad.
  - Empresas, organismos: Seguridad Social, Agencia Tributaria, Google.
  - Servicios: Gmail.
  - Clientes, servidores: `host01.example.org`
- La identidad digital con respecto a un dominio (por ejemplo Google) no tiene por que ser válida en otro dominio (por ejemplo, el IFCA).
- Diferentes soluciones para solucionar este problema.

Definición

Proceso mediante el cual se declara que alguien (o algo) es alguien (o algo).

- La identificación no es igual a autenticación.
- Ejemplo: llamada de teléfono: "Hola, soy Álvaro".
- No hay pruebas de que quien habla sea realmente Álvaro.

Definición

Proceso mediante el cual se establece que alguien (o algo) es quién (o qué) realmente afirma ser.

- Implica una identificación previa.
- Implica un intercambio de credenciales de forma que se pueda asegurar que la identificación es válida.
- Normalmente basado en alguna de estas tres credenciales (o en una combinación)
  - Algo que se sabe (password).
  - Algo que se tiene (tarjeta de coordenadas, tarjeta criptográfica, certificado digital, etc.).
  - Algo que se es (biométrica).

Definición

Proceso mediante el cual se establece qué nivel de acceso a un sistema tiene un usuario (identificado, autenticado o no) determinado.

- Puede existir autorización sin autenticación (por ejemplo usuarios anónimos).
- Ejemplo: usuarios autenticados en Moodle reciben diferentes roles (alumno, profesor, administrador).

- Más fácil, no hay que recordar contraseñas. Más rápido, no hay que teclear nada.

- Más fácil, no hay que recordar contraseñas. Más rápido, no hay que teclear nada.
- ¿Qué sucede con el lugar donde se almacena esa información? (por ejemplo un teléfono móvil?

- Más fácil, no hay que recordar contraseñas. Más rápido, no hay que teclear nada.
- ¿Qué sucede con el lugar donde se almacena esa información? (por ejemplo un teléfono móvil?
- ¿Qué sucede si esa información se ve comprometida?

- Más fácil, no hay que recordar contraseñas. Más rápido, no hay que teclear nada.
- ¿Qué sucede con el lugar donde se almacena esa información? (por ejemplo un teléfono móvil?
- ¿Qué sucede si esa información se ve comprometida?
- ¿Qué sucede con nuestra información biométrica ya almacenada en otros sistemas?

- Más fácil, no hay que recordar contraseñas. Más rápido, no hay que teclear nada.
- ¿Qué sucede con el lugar donde se almacena esa información? (por ejemplo un teléfono móvil?
- ¿Qué sucede si esa información se ve comprometida?
- ¿Qué sucede con nuestra información biométrica ya almacenada en otros sistemas?
- La información biométrica se puede robar de infinidad de maneras, y es **imposible** de cambiar.

- Más fácil, no hay que recordar contraseñas. Más rápido, no hay que teclear nada.
- ¿Qué sucede con el lugar donde se almacena esa información? (por ejemplo un teléfono móvil?
- ¿Qué sucede si esa información se ve comprometida?
- ¿Qué sucede con nuestra información biométrica ya almacenada en otros sistemas?
- La información biométrica se puede robar de infinidad de maneras, y es **imposible** de cambiar.
- Lecturas:
  - Wired, Biometrics Are Coming, Along With Serious Security Concerns: `https://www.wired.com/2016/03/biometrics-coming-along-serious-security-concerns/`
  - CSO, The dark side of biometric identification: `https://www.csoonline.com/article/3060856/data-protection/the-dark-side-of-biometric-identification.html`
  - Digicert, Biometric Authentication: An Added Layer of Security or Security Risk? `https://www.digicert.com/blog/biometric-authentication-methods/`

Identidad basada en certificados X.509

- Basado en criptografía de clave pública privada (PKI).
- Todos los agentes que necesitan identificarse tienen un certificado.
- Un certificado certifica la posesión de una llave pública asignada a un determinado subject (ya sea persona, host, servicio, etc.).

```
subject=DC = org, DC = terena, DC = tcs, C = ES, O = Consejo Superior de Investigaciones Cientificas
     , CN = ALVARO LOPEZ GARCIA 594796@csic.es
subject=/businessCategory=Government Entity/1.3.6.1.4.1.311.60.2.1.3=ES/serialNumber=Government
     Entity/street=Serrano, 117/postalCode=28071/C=ES/L=Madrid/O=Consejo Superior de Investigaciones
      Cientificas/OU=IFCA/CN=portal.cloud.ifca.es
```

- Un certificado viene firmado por una Autoridad de Certificación (CA).
- La autenticación consiste en:
  1. Comprobar si el certificado ha sido expedido o firmado por una CA de confianza.

     ```
     issuer=DC = es, DC = irisgrid, CN = IRISGridCA
     ```

  2. Comprobar si el certificado ha expirado.

     ```
     notBefore=May  3 11:19:23 2016 GMT
     notAfter=May  3 11:19:23 2017 GMT
     ```

  3. Comprobar si el certificado ha sido revocado (CRL, OCSP).
  4. Comprobar si la otra parte tiene una prueba de posesión.

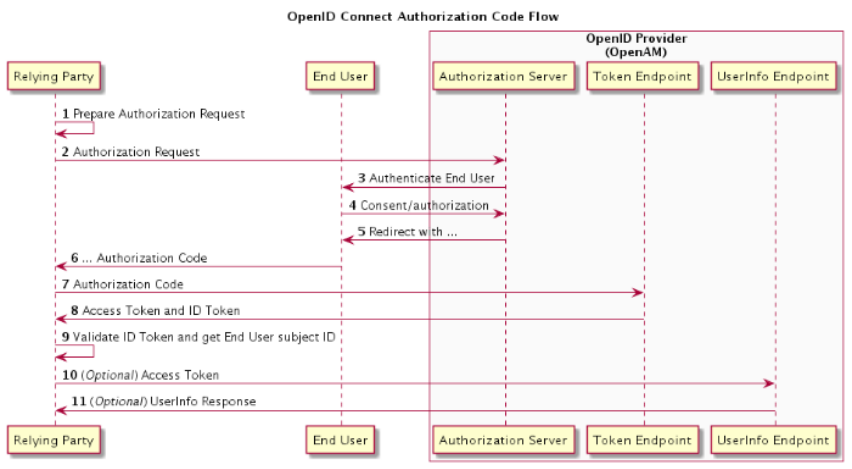
Federación de identidades

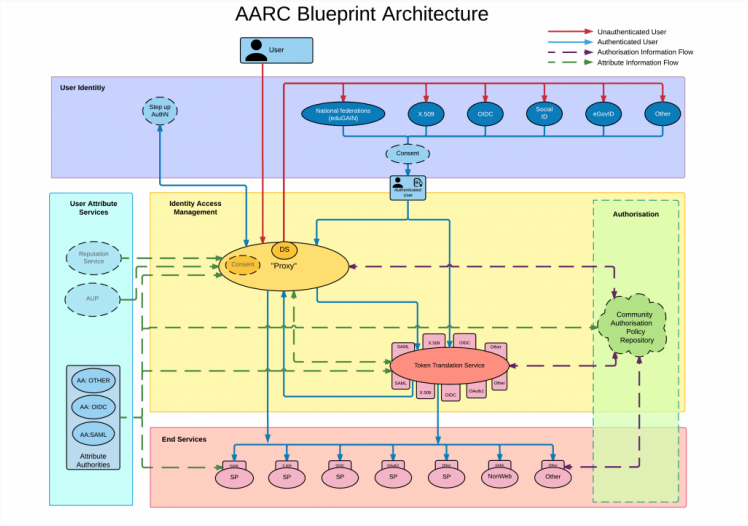
- Un usuario puede tener diferentes identidades, en diferentes sistemas, aislados entre si.
- Hay que manejar varias identidades.
- ¿Cómo hacer que un usuario pueda utilizar y linkar las mimas credenciales en diferentes sistemas?
- Diferente a Single Sign On: utilizar las mimas credenciales en diferentes sistemas (solo autenticación).
- La federación es un problema más amplio.
  - Gestión independiente, pero interoperable, de identidades en diferentes sistemas.
  - Manejo de atributos, traducción de credenciales, etc.

Estándares y tecnologías

- SAML: Security Assertion Markup Language
  - Lanzado en 2005. Basado en XML. Creado por OASIS.
  - Protocolo de autenticación y autorización.
  - Ejemplos: EduGAIN.
- OAuth 2.0
  - Lanzado en 2006. Basado en JSON. Creado por Google y Twitter.
  - Protocolo de autorización, enfocado a autorización entre aplicaciones.
  - Ejemplos: Google OAuth 2.0, Twitter, etc.
- OpenID Connect
  - Lanzado en 2014, basado en JSON, API REST.
  - Protocolo de autenticación, construido sobre OAuth 2.0 (autorización).
  - Un usuario puede utilizar sus credenciales, manejadas por un IdP, para autenticarse en diferentes servicios.
  - Ejemplos: Google OpenID Connect.

OpenID Connect




AARC Blueprint Architecture
Unauthenticated User
Authenticated User
Authorisation Information Flow
Attribute Information Flow
User
User Identity
Step up AuthN
National federations (eduGAIN)
X.509
OIDC
Social ID
eGovID
Other
Consent
Identity Access Management
User Attribute Services
Authorisation
DS
Consent
"Proxy"
Reputation Service
AUP
Community Authorisation Policy Repository
SAML
X.509
OIDC
OAuth
Token Translation Service
AA: OTHER
AA: OIDC
AA:SAML
Attribute Authorities
End Services
SP
NonWeb
Other

Preguntas?