Aretusa, sistema de reputación para BitTorrent

Aretusa, system reputation for BitTorrent

Jaime Pérez Crespo

Resumen

Aretusa es una extensión del protocolo BitTorrent cuyo objetivo es proveer una infraestructura que permita establecer un sistema de reputación mediante el cual los pares de la red decidan qué preferencia otorgan a otros pares. Dicha infraestructura hará uso de la arquitectura de identidad y autenticación proporcionada por eduGAIN.

Palabras clave: infraestructura, arquitectura de identidad y autenticación

Aretusa es una extensión del protocolo BitTorrent

Summary

Aretusa is an extension of the BitTorrent protocol, the purpose of which is to provide an infrastructure that enables the establishment of a reputation system through which network peers can decide the preference they give to other peers. This infrastructure will make use of the identity and authentication architecture provided by eduGAIN.

Keywords: infrastructure, identity and authentication architecture

1.- Descripción general

Para el desarrollo de la siguiente arquitectura partimos de una serie de presunciones y extensiones sobre el protocolo original de BitTorrent:

- El protocolo original no evita el problema del free-riding, o lo que es lo mismo, permite a los pares descargar a la máxima velocidad que permita su ancho de banda sin compartir lo descargado con otros pares de la red. Con esta extensión al protocolo se pretende aliviar este problema, introduciendo sistemas de reputación que permitan identificar de forma inequívoca a los pares que tratan de hacer free-riding.

Para construir un sistema de reputación es necesario contar con una arquitectura de identidad digital que permita identificar en todo momento a un par de la red

- Para construir un sistema de reputación es necesario contar con una arquitectura de identidad digital que permita identificar en todo momento a un par de la red. Dicha arquitectura se construirá utilizando eduGAIN (GEANT Authorisation Infrastructure for the research and education community), lo que proporcionará la capacidad de verificar la identidad de los pares, si bien el ámbito de la autorización de los mismos se dejará en manos del sistema de reputación y del uso y políticas que construyan los clientes alrededor del mismo.

- Aprovechando las funcionalidades que brinda eduGAIN, se aprovechará la capacidad de federación para establecer redes de trackers que confían entre sí, permitiendo esto extender el esquema de confianza propuesto a un nivel superior y que la identidad (y la reputación) de los pares se mantenga de forma ubicua, independientemente del tracker al que se encuentren conectados en la actualidad.

- La identidad de los pares se verificará mediante el uso de certificados digitales y esquemas de criptografía de clave pública. Dichos certificados se utilizarán en el sistema para acreditar la identidad de los pares a la hora de iniciar una descarga o realizar notificaciones acerca del comportamiento de otros pares.

- Los clientes del sistema obtendrán un certificado digital unívoco a su ingreso en el mismo, asociado al tracker al que se conectan por vez primera, y que servirá para acreditar su identidad a lo largo del tiempo en cualquier parte del sistema. Esto significa que el certificado será válido en cualquier momento y en cualquier tracker federado con el sistema, permitiendo el uso ubicuo por parte del cliente.

2.- Conceptos básicos

Los términos utilizados en este artículo coinciden con aquellos utilizados en las especificaciones técnicas oficiales de eduGAIN y la especificación del protocolo BitTorrent no oficial alojada en theory.org. Algunas de las definiciones se han unido para representar los nuevos conceptos propuestos. En particular:

Los clientes del sistema obtendrán un certificado digital unívoco a su ingreso en el mismo

Cliente: se corresponde con un cliente del protocolo BitTorrent situado en el punto de vista del usuario del sistema. El cliente representará siempre las acciones del usuario frente a otros elementos del sistema.

Peer (o par): cualquier cliente del protocolo BitTorrent externo al punto de vista del usuario. Por ejemplo, pares son aquellos que se conectan al cliente local para hacer peticiones de descargas.

Home_Tracker: un tracker BitTorrent que certifica y gestiona el uso del sistema del cliente.

Remote_Tracker: un tracker BitTorrent ajeno al cliente, pero que proporciona información de interés para el mismo (por ejemplo, la reputación de un par).

eduGAIN: es una arquitectura AAI (Authentication and Authorization Infraestructure) en desarrollo por parte del grupo GN2-JRA5, cuyo objetivo es proporcionar una infraestructura de autenticación y autorización común para unificar las infraestructuras ya existentes, como Shibboleth, PAPI, A-Select y otros. Para ello, la arquitectura se define sobre el lenguaje SAML (Security Assertion Markup Language) basado en XML. Al hablar de eduGAIN comúnmente se habla de una infraestructura de confederación o federación de federaciones.

eduGAIN es una arquitectura AAI cuyo objetivo es proporcionar una infraestructura de autenticación y autorización común para unificar las infraestructuras ya existentes

Su arquitectura descentralizada y el uso de un estándar como XML permite virtualmente interconectar cualquier sistema existente de autenticación o autorización.

3.- Protocolos y extensiones

A continuación se definen los protocolos y las extensiones sobre el protocolo original de BitTorrent para la consecución de los objetivos anteriormente descritos.

Interacción entre pares:

La interacción entre dos pares que desean colaborar en una descarga consiste básicamente en el intercambio de un certificado que acredite mutuamente su identidad y permita a cada uno verificar tanto la identidad indicada como su comportamiento mediante el sistema de reputación. Esto requiere en primer lugar una extensión del protocolo. Para ello, en el handshake inicial entre ambos pares se adjunta un campo reserved con el bit menos significativo del séptimo byte a 1. El campo reserved consta de 8 bytes. Es decir:

byte[0]
byte[1]
byte[2]
byte[3]
byte[4]
byte[5]
.....1
byte[7]

De este modo es posible verificar si un par implementa esta extensión de forma rápida y sencilla, leyendo los 8 bytes reservados y verificando si el séptimo es impar. En caso de que ambos pares implementen la extensión, intercambiarán un nuevo tipo de mensaje definido de la siguiente forma:

auth: <len=0001+X><id=10><certlen><cert><signed peer_id>

Donde <len> es un entero de 4 bytes en formato big-endian que indica la longitud total del mensaje, es decir, el tamaño en bytes del payload, los campos <certlen>, <cert> y <signed peer_id> más 1 byte correspondiente al campo <id>; el campo id es un byte indicativo del tipo de mensaje y obtiene el valor 11, exclusivo de esta extensión; tanto el campo len como el campo id son campos comunes a todos los mensajes del protocolo BitTorrent estándar, salvo para los mensajes de tipo handshake. El campo cert es el certificado que acredita la identidad del par en formato ASN.1 DER y el campo certlen que le precede es un entero de 4 bytes que indica la longitud en bytes del certificado; el campo signed peer_id es el resumen MD5 del peer_id enviado en el handshake inicial firmado con la clave privada del par y que se corresponde con la clave pública incluida en el certificado. Dicho peer_id debe corresponderse bien con el CommonName del certificado, bien con un campo SubjectAlternativeName del mismo.

De esta forma, es posible garantizar la correspondencia del peer_id del par con el certificado recibido, y usarlo en lo sucesivo como identificador válido del mismo.

Es posible garantizar la correspondencia del pee_id del par con el certificado recibido, y usarlo en lo sucesivo como identificador válido del mismo

4. Petición de reputación sobre un par

El mero hecho de haber recibido un certificado por parte de un par no garantiza de ningún modo la identidad del mismo, y tampoco proporciona ningún tipo de reputación. Para verificar la identidad de un par que desea una transferencia de bloques con el cliente y obtener información acerca de su comportamiento previo en la red son necesarias dos comprobaciones:

- Verificar que el certificado recibido ha sido emitido por un tracker confiable (autenticación basada en criptografía de clave pública). La arquitectura propuesta (similar a la arquitectura de eduGAIN) posibilita que un par presente un certificado de su home tracker en la red de un remote tracker, de modo que es necesario proporcionar un mecanismo que independice a los pares de su home tracker. En concreto, el cliente aceptará el certificado de otro par si y sólo si:

- El certificado ha sido emitido por el home tracker del cliente:

  1. Extraer la URL del remote tracker.
  2. En caso de que el remote tracker reflejado en el certificado obtenido se corresponda con el home tracker del cliente, se verifica el certificado y en caso de correspondencia, se autentica al par.

Haber recibido un certificado por parte de un par no garantiza la identidad del mismo, y tampoco proporciona ningún tipo de reputación

- El certificado ha sido emitido por un remote Tracker:

    1. Extraer la URL del remote tracker.

    2. Se establece una conexión SSL con el home tracker indicándole el par sobre el que deseamos información y el remote tracker del que procede.

    3. El home tracker verificará la identidad del remote tracker bien buscando en su caché o consultando al MDS (Meta Data Service).

    4. En caso de que el remote tracker se encuentre en caché o sea conocido para el MDS, el home tracker redireccionará al cliente al remote tracker dándole a entender que debe fiarse de él, y el cliente podrá verificar la autenticidad del par contra el remote tracker.

Una vez se ha determinado cuál es el tracker que ha certificado el par y se ha demostrado si es confiable, el cliente le dirigirá una consulta de reputación

* Obtener del tracker adecuado la reputación del par. Una vez se ha determinado cuál es el tracker que ha certificado al par y se ha demostrado si es confiable, el cliente le dirigirá una consulta de reputación.

Se realizará la consulta a la URL de autenticación ("auth") del tracker adecuado (por ejemplo, https://tracker.com/auth) y recibirá el parámetro peer_id cuyo valor será el resumen SHA-1 de la clave pública del par.

El tracker responderá con una respuesta de atributos SAML que incluirá todas las reputaciones conocidas hasta el momento para dicho par, así como el identificador (resumen SHA-1 de la clave pública) del par que genere cada una de ellas.

Es importante notar que el tracker debería configurarse para no responder con información sobre sus pares a clientes que no estén debidamente autenticados (bien por ser clientes del propio tracker, bien por serlo de trackers confiables). En tal caso, responderá con un código o mensaje de error indicativo.

Las notificaciones de reputación se producen cuando el cliente ha comprobado el comportamiento positivo o negativo de un par

5.- Notificación de reputación

Las notificaciones de reputación se producen cuando el cliente ha comprobado el comportamiento positivo o negativo de un par, esto es, ha compartido bloques con él o se ha negado a hacerlo pese a haber descargado del cliente. En tal caso el cliente iniciará una notificación de reputación de forma similar a una petición de reputación, pero cambiando la URL solicitada por la adecuada de la forma https://tracker.com/notify. Adicionalmente, en la notificación enviada al home tracker del par sobre el que se notifica habrá que añadir un parámetro además de los ya explicados en la sección anterior, reputation, consistente en un valor numérico entero descriptivo del comportamiento observado del par.

6.- Interacción entre trackers

Si bien no existe una interacción directa entre los trackers del sistema, si existe a través de otros componentes del mismo, como son el MDS (Meta Data Service) y los propios pares.

MDS

La interacción a través del MDS (componente de eduGAIN) se realiza en términos de la obtención de información sobre los trackers del sistema. Más concretamente, un tracker cualquiera puede realizar una consulta al MDS para averiguar si otro tracker dado pertenece al sistema, y verificar su identidad. El resumen SHA-1 de la clave pública de un tracker servirá como identificador único en el sistema para el mismo, de cara al MDS y al resto de trackers. Esto permite que el ingreso de un nuevo tracker en el sistema sea tan sencillo como la inclusión de su identificador en el MDS, y facilita notablemente la revocación de la identidad o la eliminación de trackers del sistema frente a esquemas más clásicos como la PKI.

Pares

Un tracker puede intercambiar información con otro a través de sus clientes, por medio de redirecciones HTTP. En concreto, un tracker responderá a una petición o a una notificación de reputación con una redirección en caso de que no sea el home
tracker del par referido. Esto es:

- Cuando un cliente desee realizar una petición o una notificación de reputación sobre un par de un remote tracker, se dirigirá a su home tracker. En este caso deberá adjuntar, además del peer_id que identifica al par, un tracker_id que identifique al remote tracker, y que no será sino el resumen SHA-1 de su clave pública. De este modo, una petición de reputación puede ser, por ejemplo:

  https://hometracker/auth?peer_id=[...]&tracker_id=[...].

- Cuando el home tracker responda con una redirección a una petición de uno de sus clientes, lo hará dirigiendo al cliente a la URL adecuada para el remote tracker, y que habrá obtenido del MDS, y añadiendo un par de parámetros al peer_id obligatorio en cualquier tipo de consulta. En primer lugar, un parámetro home_tracker con el resumen SHA-1 de la clave pública del home tracker que permita al remote tracker identificar el origen de dicha redirección, y en segundo lugar, un parámetro peer_auth que contendrá el resumen SHA-1 de la clave pública del cliente firmado por su home tracker. De este modo, el remote tracker puede verificar tanto la identidad del cliente que le hace la consulta, como la del tracker que le redireccionó allí, así como que el tracker autentica a dicho cliente y por tanto es confiable.

Cuando un cliente desee realizar una petición o una notificación de reputación sobre un par de un remote tracker, se dirigirá a su home tracker

7.- Ubicuidad e identidad única

Todos los elementos que componen el sistema constan de una identidad única proporcionada por criptografía de clave pública. Por comodidad se utilizan identificadores de 20 bytes correspondientes con el resumen SHA-1 de las claves públicas. En el caso de los pares, éstos deben contar con un certificado emitido por su home tracker que valide su identidad frente a otros componentes del sistema. En el caso de los trackers, su identidad se registra convenientemente en el MDS de eduGAIN.

Todos los elementos que componen el sistema constan de una identidad única proporcionada por criptografía de clave pública

Para garantizar la seguridad de las comunicaciones así como verificar la identidad de los interlocutores, se utilizará SSL en las comunicaciones que involucren a los trackers. Esto incluye tanto a las peticiones y notificaciones de reputación como a las consultas de identidad al MDS.

Por otra parte, al estar certificado cada par por un único tracker que además es el encargado de informar sobre el comportamiento del mismo en la red y de recoger valoraciones al respecto, se facilita la localización de los recursos manteniendo la ubicuidad del sistema mediante la federación proporciona-

da por eduGAIN. Así, el comportamiento de los pares se mantendrá registrado a lo largo del tiempo y sin importar a qué tracker se conecten. El uso del servicio de metadatos de eduGAIN facilita además la incorporación de nuevos trackers a la federación.

8.- Ponderación de reputaciones

Este proyecto plantea una infraestructura de reputación sobre BitTorrent pero no persigue definir cuál ha de ser su uso, decisión que se deja a las posibles implementaciones. No obstante, se propone un sencillo caso de uso que se utilizará como demostración, consistente en:

El uso del servicio de metadatos de eduGAIN facilita la incorporación de nuevos trackers a la federación

1. Dado que una petición de reputación de un cliente dará como resultado las distintas valoraciones recibidas hasta el momento identificadas por cada par, se hará distinción entre aquellas provenientes de pares conocidos y aquellas que no.

2. Dado que no se pueden hacer inferencias sobre la fiabilidad de la reputación proporcionada por pares desconocidos, sus puntuaciones se tratan con un umbral bajo de confianza (pero no se ignoran, ya que puede no haber información proveniente de pares conocidos).

3. Para las reputaciones proporcionadas por pares conocidos, se examina la reputación conocida del propio peer (dando prioridad a la reputación desde el punto de vista del cliente, de haberla). En caso de resultar negativa, se descarta la reputación obtenida de dicho par. En caso de ser positiva, se le da mayor valor cuanto mayor sea la reputación del par, además de tener en cuenta (de forma positiva) si éste pertenece al mismo home tracker del cliente.

4. El resultado de la ponderación será un número que oscilará entre valores negativos (mala reputación) y positivos (buena reputación). Un valor nulo puede indicar tanto la ausencia de conocimiento sobre el par en cuestión, como la imposibilidad de establecer un juicio en base a las reputaciones obtenidas. Es necesario asumir una reputación nula como si fuese positiva para no bloquear el sistema en su inicio (si nadie deja descargar sin información a priori, nadie podrá descargar y por tanto nadie tendrá reputación).

Investigadores y alumnos podrían beneficiarse de la red BitTorrent a la hora de compartir grandes volúmens de información

9.- Conclusiones

Basándonos en una arquitectura como la aquí propuesta, y promoviendo su uso e implementación, es posible abordar el problema presentado del free-riding y tratar de aliviarlo, permitiendo un uso más eficiente y adecuado de los recursos disponibles.

Así mismo, este tipo de solución basada en el uso de certificados digitales permite tener un mayor control sobre el uso de la red de pares, lo cual fomenta el uso de la misma aún en entornos en los que aún no está muy extendida, como por ejemplo las redes académicas y universitarias. De este modo, tanto investigadores como alumnos, en tal caso, podrían beneficiarse de las redes de pares y más concretamente de la red BitTorrent a la hora de compartir grandes volúmenes de información.

Referencias

[1] Oram, Andy (ed.) Peer to Peer: *Harnessing the Power of Disruptive Technologies*

[2] Hardin, Garrett. *The Tragedy of the Commons*

[3] Cohen, Bram. *BitTorrent Protocol Specification*

Jaime Pérez Crespo
(jaime.perez@rediris.es)
Área de Middleware RedIRIS