Proyecto Fin de Carrera

# Configuración de una infraestructura basada en SAML para el acceso seguro a una federación de sistemas

Escuela Superior de
Ingenieros de Telecomunicación
Universidad de Sevilla

Autor: Álvaro Soto
Tutora: Isabel Román
Área de Telemática
Dpto. de Ing. de Sists. y Automática
Noviembre 2005

# Configuración de una infraestructura basada en SAML para el acceso seguro a una federación de sistemas

Proyecto Fin de Carrera

Álvaro Soto

Noviembre 2005

*A mis padres,*

*gracias por vuestro apoyo durante todos estos años*

# Índice

Capítulo 1

# Introducción

## Contenido



## 1.1 La federación de sistemas

La creciente necesidad de cooperación entre entidades independientes en Internet y el hecho de que en la Red la información se halle ampliamente distribuida, requiere el acceso por parte de determinados usuarios a múltiples sistemas y bases de datos autónomas y heterogéneas. Cuando se trata de información de acceso libre universal, la *World Wide Web* provee ya los medios necesarios para que los propietarios o editores

de dicha información puedan hacerla llegar a sus destinatarios, mediante su publicación en páginas que éstos últimos pueden visualizar con sus *browsers* o navegadores web.

Cuando, por el contrario, la información a compartir requiere algún tipo de control de acceso, por estar destinada sólo a un grupo de usuarios autorizados, la situación se vuelve más complicada.

Un ejemplo de información privada distribuida la encontramos en el ámbito de la práctica médica. Tanto por la movilidad de los pacientes, como por el hecho de que los tratamientos que éstos reciben se administran en instituciones diferentes, es muy posible que el historial médico de una misma persona se encuentre repartido en diversos sistemas informáticos que no forman parte de un mismo dominio de seguridad. Resultaría útil para un profesional de la medicina poder contar con un sistema que le proporcionara acceso a toda la información existente en la Red relativa a un paciente en concreto, se encontrara donde se encontrara dicha información.

De la misma manera, un investigador médico podría encontrar interesante dar acceso a ciertas áreas de su investigación a determinados colegas o instituciones, teniendo la seguridad de que la información será accesible únicamente para las personas autorizadas.

Una respuesta a estas cuestiones puede encontrarse en el concepto de federación de sistemas. Básicamente consiste en un conjunto de sistemas autónomos que desean compartir cierta información entre todos los usuarios federados, cediendo parte del control de acceso a la organización que los engloba, pero manteniendo su autonomía y la decisión última de acceso sobre sus propios recursos. Implementar una federación implica, generalmente, superponer una capa de software (*middleware*) a los sistemas existentes.

Cuando un usuario intente acceder a un recurso electrónico federado, el sistema donde esté ubicado dicho recurso decidirá sobre el acceso basándose en la información (denominada **credenciales** o **atributos** del usuario) existente sobre dicho usuario. Dicha información puede provenir tanto del propio sistema local como de otros sistemas federados. Esto plantea la necesidad de la existencia de mecanismos para el envío de

información de seguridad entre sistemas autónomos. Mediante estos mecanismos, un servidor debe ser capaz de decidir si los datos que recibe de otro componente de la federación son o no fiables.

Esto no significa que esos datos que sirven como criterio para conceder el acceso o denegarlo provengan de una entidad centralizada, sino que la federación proveerá los medios para que los sistemas autónomos puedan obtener dichos datos de fuentes distribuidas y fiables.

La arquitectura de una federación de sistemas depende fuertemente del modelo de confianza elegido. En nuestro caso los sistemas pertenecientes a la federación aceptarán como válidas las credenciales que se envíen desde determinado agente, al que llamaremos **agente de credenciales**. Con este modelo de confianza la autenticación del usuario se realiza una sola vez para toda la federación y el agente de credenciales obtiene las credenciales necesarias para acceder a recursos en cualquiera de los sistemas de la federación.

La federación proporcionará a los usuarios una especie de "acceso primario", mientras que la decisión última sobre el acceso a un recurso particular recaerá siempre sobre el sistema propietario de dicho recurso. La federación, por tanto, respetará por completo la autonomía de cada sistema federado. Dicho esto, podría darse el caso de que uno de los sistemas decidiera conceder acceso a sus recursos a todos los usuarios que superaran únicamente la "barrera" federal, haciendo coincidir los requisitos para acceder a la federación con los necesarios para acceder a sus propios recursos, pero esto constituye únicamente una de las posibilidades por las que puede optar dicho sistema.

Uno de los temas más candentes en los últimos años en el terreno de las comunicaciones digitales es el del respeto a la intimidad de los usuarios. El sistema que posee el recurso deseado debe conocer cierta información sobre el usuario que lo solicita, para poder concederle el acceso y, en principio, podría pensarse que esto implica el revelado de datos confidenciales, como puede ser, por ejemplo, la propia identidad del usuario. En muchas ocasiones, sin embargo, la identidad no es un factor determinante a la hora de otorgar o denegar el acceso a un recurso. A veces lo decisivo es que el usuario que solicita el acceso posea una determinada cualidad o características, que pertenezca a un

grupo en concreto, etc. Con los mecanismos adecuados para intercambiar credenciales de forma fiable, conceder acceso a recursos a usuarios anónimos no debería plantear ningún problema de seguridad.

Una federación debe contemplar, por tanto, la posibilidad de que los sistemas que forman parte de ella tomen decisiones de acceso basadas únicamente en los atributos del usuario, protegiéndose así la identidad del mismo. Además, cada usuario podrá decidir qué datos de los almacenados en su registro local entrega a los sistemas remotos durante el proceso de decisión de acceso, convirtiendo el nivel máximo de privacidad en una decisión personal (el nivel mínimo será establecido por el propio sistema local).

El formato de estos datos personales podrá variar de forma sustancial de un sistema a otro y será la federación la que deba proveer de los medios necesarios para la adaptación de dicho formato entre sistemas diferentes.

Las tareas de seguridad en una federación son realizadas por agentes software especializados. Existen cuatro tipos principales de agentes cuyas funciones, según el modelo de confianza que hemos elegido, serían las siguientes:

- **Agentes de autenticación:** comprueban la identidad del usuario, bien pidiéndole directamente un nombre y una contraseña o por otros medios, como la entrega de un certificado digital.
- **Agentes de credenciales:** se encargan de obtener las credenciales del usuario, realizar un determinado procesado y/o adaptación de las mismas, y de entregarlas al sistema donde reside el recurso al que se quiere acceder.
- **Agentes de decisión de acceso:** basándose en las credenciales recibidas y en políticas de acceso previamente definidas, deciden si el usuario puede o no acceder a un recurso determinado.
- **Agentes de control de acceso:** aplican las decisiones tomadas por los agentes anteriores.

En este Proyecto nos centraremos en los agentes de autenticación y credenciales, incluyéndolos en un esquema general de federación que se presentará en este mismo capítulo.

**Nota sobre terminología:** para evitar confusiones con la nomenclatura que utiliza Eclipse, cuando queramos referirnos a este Proyecto Fin de Carrera usaremos la palabra "Proyecto", para diferenciarlo de "proyecto", todo en minúsculas, que es el conjunto de archivos contenidos en una única carpeta de Eclipse, tales como los proyectos "shib" y "shib-filter" que aparecen en capítulos posteriores.

## 1.2 Objetivos del Proyecto

Como objetivos del Proyecto nos marcamos los siguientes:

- Plantear un modelo de federación que satisfaga todas los requisitos que hemos expuesto anteriormente.
- Investigar el Estado del Arte en el ámbito de la federación de sistemas, comenzando con el estudio de estándares y mecanismos que permitan el intercambio de información de seguridad (nombres de usuario, contraseñas, credenciales en general) entre sistemas autónomos. Dicho intercambio seguro es la base sobre la que es posible construir una federación. A continuación, analizar diferentes alternativas que lleven a la práctica estos mecanismos.
- Implementar un sistema Java siguiendo el modelo de federación que se plantea inicialmente, prestando especial atención a las tareas que deben realizar el agente de autenticación y el de credenciales.
- Proponer los siguientes pasos a dar en el proceso que llevaría a convertir dicho sistema implementado en una federación operativa real.

## 1.3 Materiales y métodos

Para la implementación del sistema Java objeto de este Proyecto se usaron los siguientes elementos:

- Sistema operativo Windows XP Professional Edition
- Servidor de aplicaciones Tomcat 5.5
- Plataforma Java 2 Standard Edition, J2SE 5.0
- Entorno de desarrollo de aplicaciones Eclipse 3.1
- *Plug-in* para Eclipse Sysdeo Tomcat Launcher v31beta
- *Plug-in* para Eclipse XMLBuddy v2.0.62
- Código Java perteneciente al proyecto de código abierto Shibboleth v1.3, que posee licencia Apache 2.0 (la licencia está recogida en los apéndices)

A continuación aparece un diagrama temporal de las actividades llevadas a cabo durante este Proyecto (las fechas y la duración de los intervalos son aproximadas):

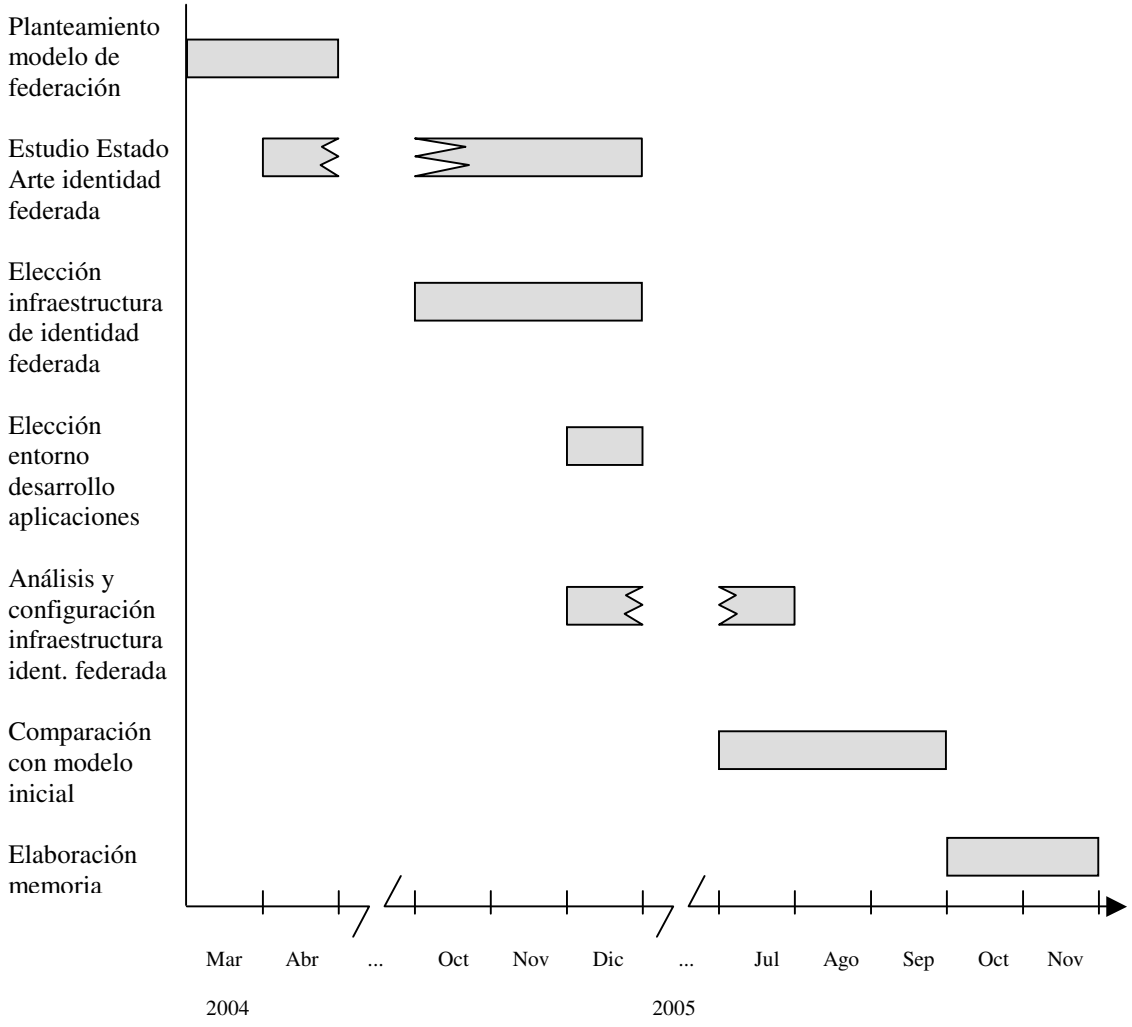


*Figura 1: diagrama temporal de actividades del Proyecto*

Del diagrama anterior se deduce que las actividades que más tiempo han supuesto han sido el estudio del Estado del Arte en identidad federada y el análisis y configuración de la infraestructura de identidad federada.

## 1.4 Modelo de federación

La situación de partida es la siguiente: existen una serie de bases de datos o servidores independientes que desean compartir la información que poseen, de forma prolongada en el tiempo (no se trata únicamente de un intercambio puntual de datos). Cada servidor dispone de un grupo de usuarios registrados, que poseen cuentas en dicho servidor. Por "cuenta" entendemos, como mínimo, un nombre de usuario (el cual constituye la

identidad local de dicho usuario) y una contraseña privada, pero, en principio, suponemos que el servidor almacenará información adicional sobre el usuario. Por ejemplo, en el contexto del personal de un hospital, el servidor podría almacenar el puesto que ocupa cada empleado, el departamento al que está asignado, etc. Esta información constituye los atributos o credenciales del usuario.

El conjunto de servidores y sus respectivos usuarios se unirán para formar una federación. El requisito mínimo para que un usuario pueda formar parte de ella es que tenga, al menos, una cuenta en alguno de los servidores que la conforman. Corresponde a cada sistema federado aplicar o no una política de acceso distinta según el usuario sea local o provenga de otro componente de la federación. Se admite que un usuario tenga cuentas en más de un servidor dentro de la federación, es decir, no se establece ningún límite en el número de servidores en los que un mismo usuario pueda estar registrado.

La federación se basará en las siguientes premisas:

- El hecho de que un servidor se federe no implica que todos sus contenidos estén al alcance de todos los demás miembros de la federación. El control último de acceso a los recursos de un servidor seguirá en manos del propio servidor, es decir, se respetará la **autonomía** de cada sistema.
- Se preservará la **privacidad** de los usuarios. Un usuario registrado en un sistema no entregará más información sobre si mismo a los demás sistemas que la que él mismo desee. Para contribuir a esto, los usuarios dispondrán de una **identidad federada** (y su correspondiente contraseña federada) que servirá para identificarles en el contexto de la federación.
- La federación contemplará la posibilidad de la **firma única** o ***single sign-on*** (SSO), es decir, un solo proceso de autenticación valdrá para toda la federación.
- Deberá garantizarse la **seguridad** en todas las transacciones que se produzcan dentro de la federación y que involucren información sensible. Esto se concretará en los objetivos de autenticación, confidencialidad, integridad y no repudio de los datos (estos conceptos se explicarán en el apartado 2.2).

En la figura 2 aparece el modelo de federación que se propone:

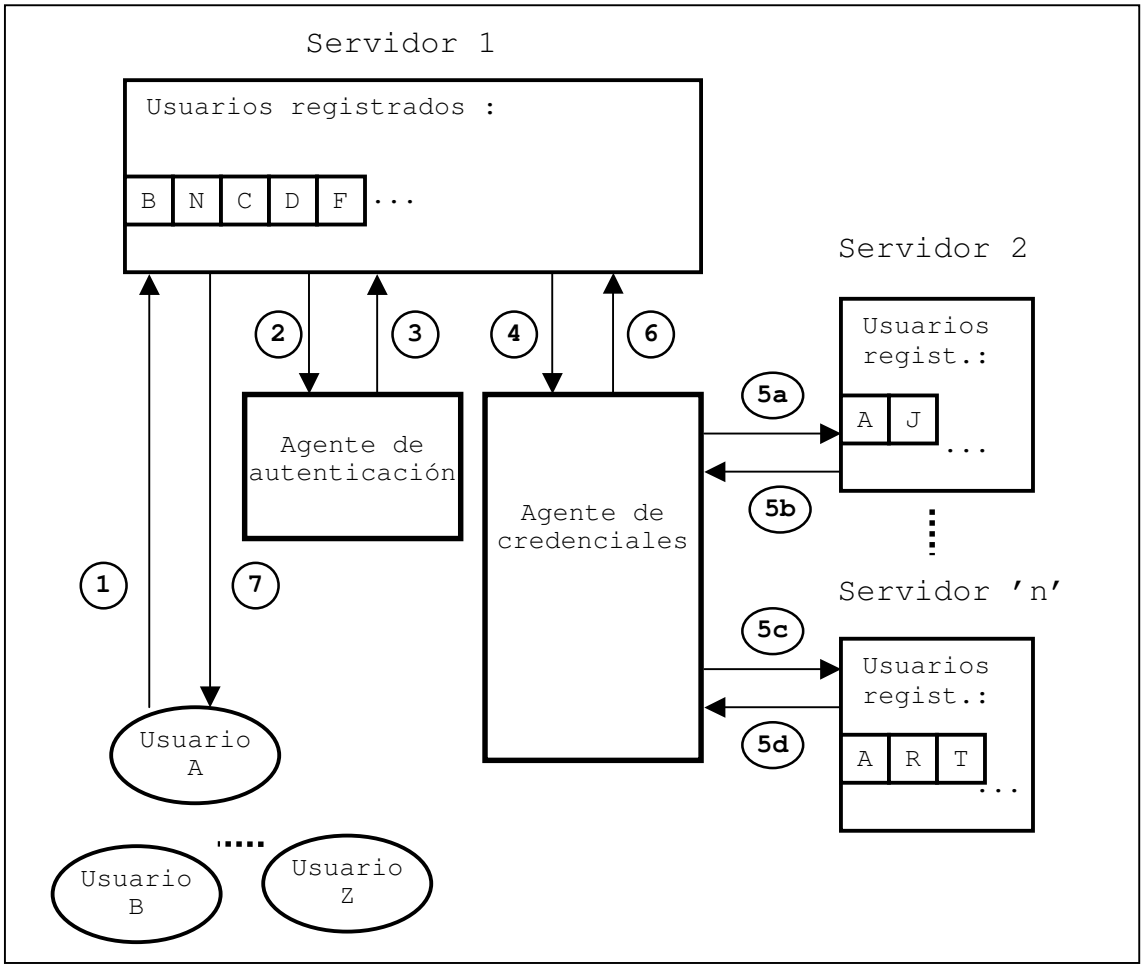


*Figura 2: Modelo de federación*

En la figura, el tamaño de los rectángulos que representan a los servidores no está relacionado con la capacidad ni con el número de usuarios que éstos puedan tener y el lugar privilegiado que ocupa el Servidor 1 no implica ninguna diferencia de jerarquía con respecto a los demás, se dibuja así sólo por claridad. Las letras que aparecen en el interior de los servidores representan a los usuarios registrados en cada uno de ellos (la asignación de determinados usuarios a determinados servidores es totalmente aleatoria en esta figura).

Además de servidores y usuarios, aparecen otros dos elementos: el agente de autenticación y el agente de credenciales. Ambos son entidades centralizadas, esto es, sólo existe un agente de autenticación y un agente de credenciales en cada federación. Los agentes centralizan determinados datos y funciones dentro de la federación. En concreto, los datos mínimos a almacenar serán:

- El **agente de autenticación** dispondrá de una tabla o base de datos donde almacenará las identidades y contraseñas federadas de todos los usuarios de la federación.
- El **agente de credenciales** almacenará una tabla o base de datos donde estará recogida la correspondencia entre cada identidad federada y el sistema (o sistemas) del que proviene.

Adicionalmente, cada servidor dispondrá de una base de datos donde almacenará la correspondencia entre la identidad federada y la identidad local para los usuarios que estén registrados en dicho servidor.

Una transacción típica dentro de esta federación ocurrirá de la siguiente manera:

**1.** El usuario A solicita un recurso situado en el servidor 1.

**2.** El servidor pide al agente de autenticación que proceda a autenticar al usuario. El agente solicita del usuario su identidad y contraseña federadas.

**3.** Si los datos introducidos concuerdan con los que el agente de autenticación posee en su tabla, éste contesta a la petición del servidor 1 enviándole la identidad federada del usuario.

**4.** El servidor envía dicha identidad federada al agente de credenciales y solicita de éste que le proporcione los atributos del usuario.

5. El agente de credenciales consulta la tabla que tiene almacenada para averiguar a qué sistema pertenece dicha identidad federada. Seguidamente contacta con dicho sistema y le solicita los atributos correspondientes, presentando la identidad federada como referencia. Esto ocurre en los pasos 5a y 5b. Si el usuario tiene cuentas en varios servidores de la federación (su identidad federada se corresponde con varias identidades locales), el agente de credenciales contactará con todas ellas secuencialmente hasta conseguir los atributos del usuario en todos los sistemas en los que está registrado. En nuestro caso, el usuario A posee una cuenta en el servidor 'n', además de en el 2, por lo que son necesarios también los pasos 5c y 5d. Previamente, cada usuario habrá decidido, en el servidor donde está registrado, qué atributos deben entregarse como respuesta a estas peticiones (premisa de privacidad). La información sobre el usuario puede provenir de cualquiera de los sistemas de la federación, incluido aquél donde está ubicado el recurso al que se quiere acceder.

6. El agente de credenciales entrega los atributos del usuario al servidor 1, pudiendo efectuar previamente algún tipo de procesado sobre ellos, como adaptar valores y/o esquemas de representación, comparar las credenciales del usuario en diferentes sistemas y generar conclusiones basándose en la comparación, etc.

7. Basándose en estos atributos o credenciales el servidor 1 permite (o deniega) el acceso al recurso solicitado.

En principio, lo más probable es que el acceso del usuario se realice utilizando un **navegador web** (Mozilla Firefox, Opera, Internet Explorer, etc.), aunque no necesariamente tiene que ser así. En cuanto al tipo de recurso solicitado éste puede ser una página web, un archivo PDF, algún otro tipo de documento, una imagen, etc.

La estructura del proceso sería el mismo si el usuario A intentara acceder a un recurso en el servidor 2 (donde sí posee una cuenta), en vez de en el servidor 1 (donde no la posee). La parte del proceso que podría variar sería la decisión última de acceso, que depende exclusivamente de la política de acceso del propio servidor, y aquí sí puede influir que el usuario provenga del propio sistema o de algún otro.

Existen otras alternativas de funcionamiento para el agente de credenciales: podría almacenar toda la información de los usuarios de la federación y la entregue, realizando una adaptación previa, al sistema que la requiera; podríamos no centralizar sus funciones, etc. El modelo tan sólo muestra una de las innumerables opciones posibles.

Para respetar la premisa relacionada con el **single sign-on**, establecemos las dos condiciones siguientes:

- Si un usuario intenta acceder por segunda vez a un recurso, para el cual ya obtuvo permiso anteriormente, el sistema lo detectará y permitirá su acceso sin necesidad de introducción de contraseña ni de obtención de atributos.
- Si un usuario accede con éxito a un recurso de la federación y, seguidamente, solicita otro recurso distinto, no se le requerirá que introduzca de nuevo su clave federada. Se procederá directamente a conseguir y analizar sus credenciales, para decidir sobre su acceso.

La contraseña federada dará, por tanto, acceso general a la federación, mientras que para acceder a un recurso concreto, será la decisión del sistema propietario de dicho recurso la que decida. Se respeta así la **autonomía** de los servidores federados.

Además, al centralizar en los agentes toda la información referente a la identidad federada, se respeta la **privacidad** del usuario. Como información sobre la persona que intenta acceder a uno de sus recursos, un servidor sólo obtendrá su identidad federada y los atributos que ésta decida revelar. Cada servidor conoce la correspondencia entre identidad federada e identidad local, únicamente para sus propios usuarios registrados.

En cuanto a la **seguridad**, es necesario que exista "confianza" mutua entre servidores (que un servidor considere fiable la información que procede del otro) pero las técnicas necesarias para garantizarla están fuera del alcance de este capítulo. Nos ocuparemos de ellas en capítulos posteriores.

## 1.5 Organización de la memoria

Los capítulos del 2 al 6 corresponden a los fundamentos teóricos en los que se basa el Proyecto:

**Capítulo 2:** se proporcionan unas nociones básicas sobre seguridad en redes de ordenadores como son el cifrado y la firma digital.

**Capítulo 3:** damos algunos conceptos básicos sobre aplicaciones Java del lado del servidor, así como de otros elementos que aparecerán a lo largo de la memoria, como son los protocolos SOAP y LDAP.

**Capítulo 4:** entramos en el estudio de la identidad federada propiamente dicha, analizando las especificaciones del protocolo SAML para intercambio de información de seguridad.

**Capítulo 5:** estudiamos el proyecto de software de código abierto Shibboleth. Dicho proyecto se basa en SAML y nos servirá de infraestructura básica para nuestro sistema.

**Capítulo 6:** se realiza un breve análisis de otras iniciativas de identidad federada, como son Liberty Alliance, WS – Federation y Microsoft Passport.

En los capítulos del 7 al 10 se realiza una descripción en profundidad del sistema implementado:

**Capítulo 7:** se realiza una primera descripción de los componentes funcionales del sistema y de cómo se produce la interacción con el usuario.

**Capítulo 8:** entramos a describir detalles más técnicos del sistema, explicando cómo configurarlo, cómo se distribuye el código, etc.

**Capítulo 9:** describimos instrucciones específicas para trabajar con el sistema en el entorno de desarrollo Eclipse, desde su instalación completa hasta la compilación y ejecución del código.

**Capítulo 10:** realizamos una comparación entre el sistema implementado y el modelo inicial de federación. Se discute en qué aspectos nos hemos aproximado más, en cuáles existen diferencias, etc.

La sección siguiente, **"Conclusión y líneas de continuación"**, expone los objetivos alcanzados en el Proyecto, así como las posibles direcciones en las que seguir avanzando para mejorar el sistema.

Finalmente aparecen las referencias bibliográficas y los apéndices, donde se recogen detalles del código y de algunos de los archivos de configuración, así como la licencia del código fuente.