Implantación de un servicio de autenticación basado en Shibboleth en la PUCP - Caso de Estudio

Oscar Díaz Barriga[a], Dennis Cohn Muroy[b], Genghis Ríos Kruger[c]

Pontificia Universidad Católica del Perú, Dirección de Informática Académica,
Av. Universitaria 1801, Lima 32, Lima, Perú
[a] diaz.oa@pucp.edu.pe, [b] dennis.cohn@pucp.edu.pe, [c] grios@pucp.edu.pe

Resumen. La seguridad y gestión de la identidad es uno de los principales problemas que enfrentan las áreas de TI de diversas instituciones educativas. Asimismo, los usuarios se ven en la necesidad de recordar cada vez un mayor número de contraseñas. El presente trabajo, busca dar a conocer el proceso de análisis, diseño y despliegue que se ha seguido en la Pontificia Universidad Católica del Perú con el objetivo de poner en funcionamiento la herramienta de autenticación federada Shibboleth. Dentro de los puntos que se abordarán en el presente documento, se dará a conocer los problemas que fueron encontrados; así como las decisiones que fueron tomadas a lo largo del proceso de despliegue de la herramienta. Adicionalmente, durante la puesta en marcha, se presentó la necesidad de contar con herramientas de apoyo para el proceso de registro de usuarios y activación de cuentas para facilitar la migración de usuarios desde un sistema de autenticación previamente utilizado. Finalmente, se concluye que es posible hacer uso de Shibboleth para desplegar una solución AAI Federada Académica tolerante a fallos bajo un esquema de alta disponibilidad; sin embargo, a menos que los datos de los usuarios puedan ser refrendados, existirá un riesgo a que el sistema pueda presentar un comportamiento anómalo que afecte la percepción de seguridad sobre el mismo por parte de los usuarios.

Palabras Clave: Alta Disponibilidad, Arquitectura, Autenticación, DIA, ECP, Federación, IdP, PUCP, Shibboleth

1 Introducción

La Dirección de Informática Académica (DIA) es el órgano dentro de la Pontificia Universidad Católica del Perú (PUCP) encargado de proveer herramientas tecnológicas a los docentes y alumnos de la comunidad universitaria que les sirva como apoyo dentro de su labor académica.

A lo largo de los años, la DIA ha desarrollado y desplegado diversas soluciones dentro del campus como repositorios de contenido multimedia, una plataforma LMS, soluciones de virtualización de aplicaciones, servicios de videoconferencias, aulas informáticas, entre otros. Cada uno de los sistemas puestos a disposición de la comunidad universitaria, requiere de un usuario y una contraseña para poder autenticarse, estándar de facto en la autenticación web según Haron et al. [4]. Esto genera mayor dificultad para los usuarios quienes, de acuerdo a Florencio et al. [3], utilizan en promedio 8 servicios web diariamente, cada uno con su propia contraseña.

Según Suoranta et al. [11], a fin de ayudar a los usuarios con el proceso de autenticación, diversas instituciones buscan emplear sistemas *Single Sign On (SSO)* que permite a los usuarios autenticarse en distintas aplicaciones haciendo uso de un único usuario y una única contraseña.

En el año 1999, de acuerdo a Morgan et al. [6], Internet2 estableció la iniciativa I2IM a fin de buscar una solución a los problemas de "seguridad y gestión de identidad" que venía afectando a las áreas de TI de los campus universitarios. Como resultado de dicho esfuerzo, se creó el Sistema Shibboleth [4], [5], [11] que dentro de sus beneficios presenta el uso de estándares para la gestión de la identidad en el campus, control de privacidad y autorización basada en atributos. Actualmente, es una de las tecnologías más utilizadas a nivel de las universidades para gestionar la identidad y controlar el acceso de sus usuarios [2].

El presente trabajo presenta los pasos y decisiones tomadas durante el proceso de despliegue de Shibboleth dentro del campus de la PUCP. El contenido se encuentra agrupado como se indica a continuación: la Sección 2 contiene el Marco Teórico del trabajo, la Sección 3 presenta la propuesta de solución y, finalmente, la Sección 4 consolida las conclusiones y trabajos futuros.

2 Marco Teórico

A continuación se detallarán algunos conceptos necesarios para comprender los contenidos expuestos en el presente documento.

2.1. Autenticación

También conocido como acreditación, es el proceso a través del cual se verifica la identidad digital del remitente dentro de una comunicación.

Existen tres mecanismos a través de los cuales se puede llevar a cabo esta validación [1]:

1. Sistemas basados en algo conocido: Son aquellos sistemas que hacen uso de contraseñas para poder acceder.
2. Sistemas basados en algo poseído: Son aquellos sistemas que hacen uso de llaves físicas o digitales que pueden ser utilizadas para garantizar el acceso.
3. Sistemas basados en una característica física: Son aquellos sistemas en los que se requiere alguna característica física del usuario (por ejemplo, su huella dactilar) para poder garantizar el acceso al mismo.

2.2. Single Sign On

El mecanismo de Single Sign On (SSO) le permite a un usuario poder ingresar a

diversos sistemas o recursos informáticos haciendo uso de las mismas credenciales de acceso. En la actualidad, de acuerdo a Pashalidis et al. [8] estos sistemas pueden clasificarse en los siguientes grupos:

- Sistemas seudo SSO locales.
- Sistemas seudo SSO basados en proxy.
- Sistemas verdaderos de SSO locales.
- Sistemas verdaderos de SSO basados en proxy.

2.3. Identidad Federada

La gestión de la identidad hace referencia al conjunto de políticas, procesos y tecnologías que permiten establecer cuentas de usuario y reglas relacionadas a la administración de la información y recursos digitales dentro de la organización.

Dentro de un campus universitario, los recursos digitales pueden ser cuentas de correo, plataformas educativas (sistemas e-learning), bases de datos de las bibliotecas, equipos informáticos, entre otros [6]. Para poder acceder a cada uno de estos recursos, el usuario (ya sea alumno o profesor), requiere autenticarse.

Haciendo uso de un sistema de autenticación centralizado, el usuario es capaz de acceder a cada uno de estos recursos haciendo uso de un mismo usuario y una misma contraseña.

Si extendemos esta funcionalidad a poder garantizar que el usuario haga uso de sus mismas credenciales para poder acceder a recursos compartidos entre distintas instituciones, estaremos definiendo el concepto de Identidad Federada [11] la cual provee beneficios al equipo TI de cada institución en la reducción de tiempo y del número de procedimientos necesarios para poder proveer acceso a un recurso compartido a otras instituciones.

2.4. Shibboleth

Shibboleth [5] es una tecnología Open Source basada en estándares que, de acuerdo a Haron et al. [4], implementa una solución de logueo unificado para sistemas web que puede ser utilizada dentro de una organización o través de diversas instituciones. Shibboleth es un sistema verdadero de SSO basado en proxy [11].

El estándar que utiliza esta tecnología para gestionar la identidad federada de los usuarios, es el Security Assertion Markup Language (SAML); la versión actualmente vigente es la 2.0 (SAML 2)

Es posible identificar 3 elementos dentro de un sistema de autenticación basado en Shibboleth como se muestra en la Figura 1.

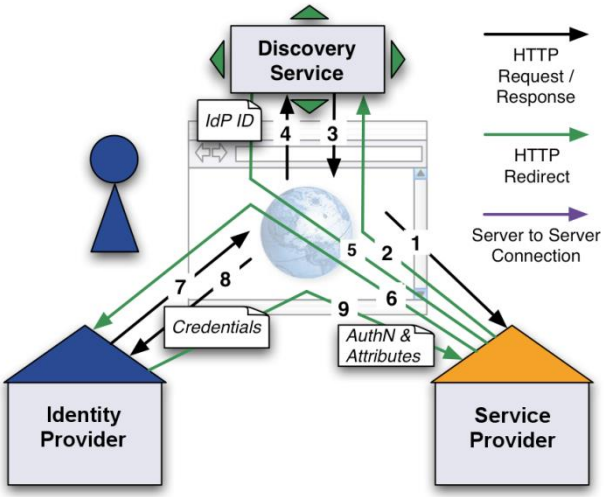

Fig. 1. Sistema de autenticación basado en Shibboleth

1. El Proveedor de Identidad (IdP)

Aprovisiona los servicios de Single Sign On y brinda, en caso de un logueo exitoso, la información necesaria sobre el usuario al proveedor de servicio; a fin de que éste pueda brindar una experiencia personalizada y cuente con la información actualizada del usuario.

2. El Proveedor de Servicio (SP)

Permite integrar los recursos con el sistema de Single Sign On aprovisionado por el Proveedor de Identidad; ya sea de la misma institución, o perteneciente a otra organización dentro de la federación.

3. El Servicio de Descubrimiento (DS)

Permite elegir al usuario el nombre del Proveedor de Identidad que utilizará para iniciar sesión en un Proveedor de Servicio.

Cuando un usuario, mediante un navegador web, intenta acceder a un contenido protegido, el SP redirige al usuario hacia un DS en donde el usuario selecciona el nombre de la organización a la que pertenece. El navegador redirige al usuario al IdP de su organización para que se autentique. Una vez que se autentique de forma exitosa, el IdP de la organización entrega al SP la mínima información necesaria sobre la identidad del usuario que permita al sistema la tomar decisiones sobre el grado de autorización del usuario. En la Figura 2 se muestra el flujo que sigue un usuario para autenticarse en un SP.

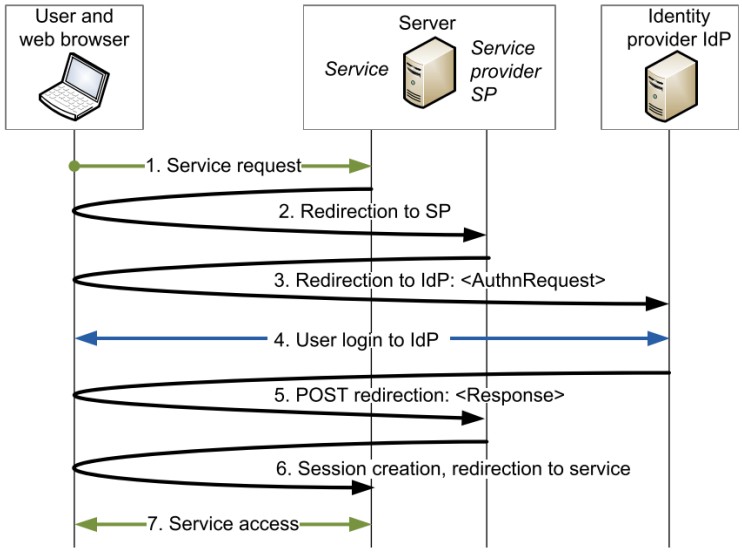

Fig. 2. Proceso de login de Shibboleth sin utilizar un DS por Suoranta et al. [11]

3 Diseño de la Solución

Luego de haber analizado el uso de Shibboleth en diferentes universidades; se determinó la utilidad de esta herramienta como vínculo que permitiera la integración de la PUCP y otros centros de estudio; facilitando el acceso a los alumnos a nuevas fuentes de conocimiento.

Es por ello que, tomando en cuenta ventajas del uso de las Federaciones como herramientas facilitadoras para el proceso de comunicación e intercambio de conocimientos, se establecieron los requerimientos que el sistema de autenticación centralizado de la Universidad debía de soportar.

3.1 Requerimientos del Sistema

Para el despliegue de un nuevo sistema de autenticación centralizada a nivel del campus; éste debía de contar con las siguientes funcionalidades:

- Solución basada en estándares abiertos.
- Integración con Servicios de Directorio de usuarios.
- Integración con el sistema de autenticación de Windows (WNA).
- Integración con Kerberos.
- Soporte de interconexión entre Universidades.
- Capacidad para poder ser desplegado sobre una arquitectura redundante.
- Poder gestionar roles.
- Permitir la gestión de las sesiones de los usuarios.

3.2 Componentes del sistema

Frente a los requerimientos presentados, se optó por hacer uso de la herramienta Shibboleth. Sin embargo, era necesario tomar en cuenta algunas consideraciones al momento de efectuar el despliegue a fin de garantizar la alta disponibilidad del mismo.

Por ello que se decidió contar con los siguientes equipos:

- **Balanceador de Carga (L.B.):** Se cuenta con 1 balanceador de carga, que atiende las peticiones de los usuarios. Este balanceador se ubica frente a los servidores de aplicaciones y, tras analizar el tiempo de respuesta de cada servidor, procede a derivar la solicitud al equipo que presente mejor tiempo de respuesta. Para la presente arquitectura, se optó por utilizar servidores ejecutando el software de balanceo Pound.

  Inicialmente se hicieron pruebas con dos balanceadores y hacer uso de un DNS Round Robin para seleccionar el balanceador de carga que atenderá la solicitud del usuario, sin embargo se ha detectado que bajo dicho esquema no todos los navegadores presentan un comportamiento similar. Algunos navegadores (como el Chrome) tienden a ignorar el valor del TTL entregado por los servidores DNS, lo cual ocasiona (con una probabilidad del 50% en caso se cuente con 2 servidores IdP) que el proceso de inicio de sesión se vea interrumpido, esto hizo que se descarte el uso de dos balanceadores.

- **Servidor de aplicaciones (IdP):** Ubicados detrás de los servidores de balanceo de carga. Son los responsables de la ejecución de los Proveedores de Identidad. Adicionalmente a ejecutar los IdP, cada uno de estos servidores cuenta con un sistema memcache, encargado de gestionar la clusterización de las sesiones entre ambos equipos. Para su adecuado funcionamiento se hace uso del Memcached Storage Service.

- **Servidores Active Directory (A.D.):** Son 2 nodos configurados en modo de sincronización y que funcionan en estado activo-activo. Estos servidores, además de responder a las consultas de autenticación de los IdP, sirven como mecanismos de autenticación que son utilizados por los Servidores de tipo Terminal Servers.

En la Figura 3 se puede apreciar el diagrama de despliegue de los componentes previamente mencionados.

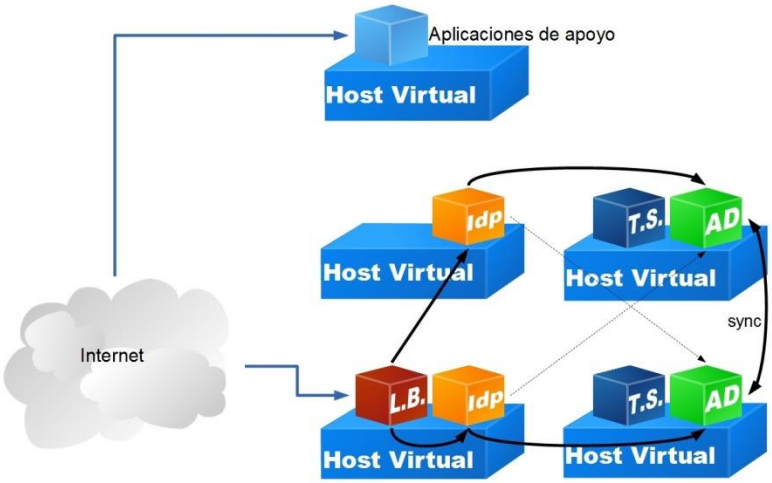

Fig. 3. Diagrama de despliegue

3.3 Integración

a. Aplicaciones Web

Muchas de las aplicaciones que se han desarrollado en el área, se encuentran implementadas en PHP. Por ello, como primer paso para facilitar su integración, se requirió implementar una librería (Figura 4) que permitiera que los desarrolladores pudieran integrar las aplicaciones en el menor tiempo posible.

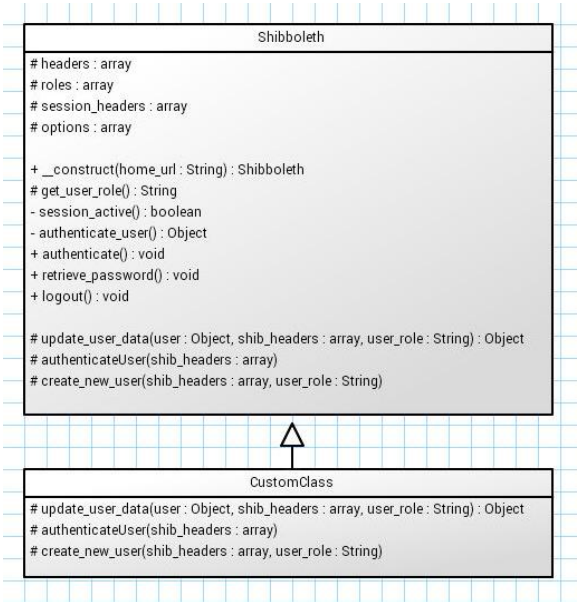

Fig 4. Diagrama de clases de la librería implementada para integrar las aplicaciones en PHP

La librería implementada utilizó como base algunas de las ideas expuestas en el plugin de Shibboleth para wordpress [11]. Para hacer uso de la misma, el programador requiere extender la librería, implementando 3 métodos de acuerdo a los requerimientos de la aplicación:
- authenticateUser(): Crea la instancia del usuario dentro de la aplicación utilizando los datos que el SP ha recibido del IdP
- create_new_user(): Registra al usuario en la aplicación en caso éste no exista.
- update_user_data(): Actualiza los datos de un usuario previamente registrado.

La Figura 5 muestra un diagrama de flujo que representa la comunicación de las 3 funciones previamente implementadas con el IdP y la aplicación web.

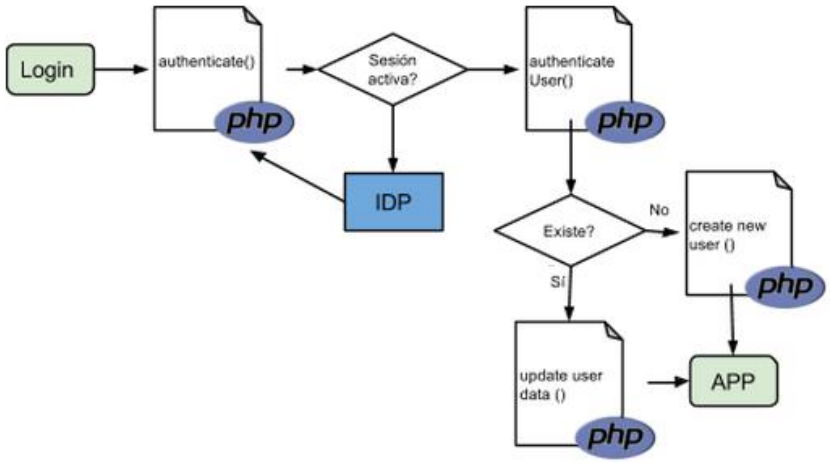

Fig 5. Diagrama de flujo que representa las llamadas a los métodos implementados por la librería de conexión escrita en PHP

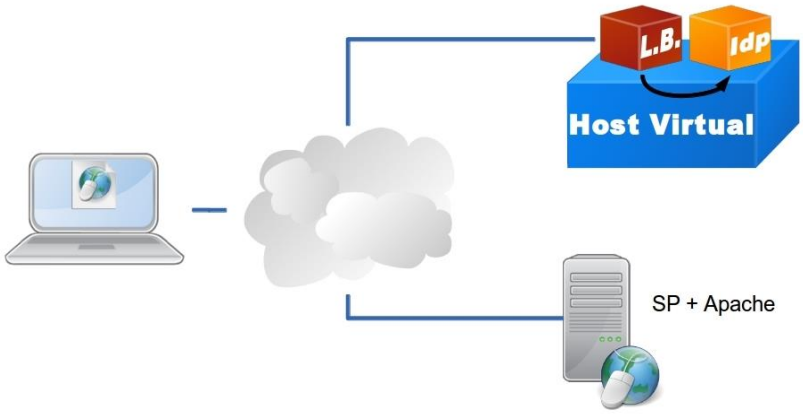

Fig 6. Diagrama de despliegue que muestra los componentes asociados en el proceso de autenticación a través de una aplicación web

b. Aplicaciones Móviles

Además de las aplicaciones web, la unidad ha implementado varias aplicaciones móviles que complementan servicios con fines académicos y que son brindados a los alumnos y profesores.

Para permitir su interconexión con el Shibboleth se habilitó el módulo ECP (Enhanced Client or Proxy) [10], tanto al nivel del IdP como del SP; luego se implementó una aplicación que muestra la información del usuario recibida desde el IdP en formato JSON. Dicha aplicación se desplegó en el SP en una ruta protegida, es decir, requiere una autenticación por Shibboleth para poder ser visualizado.

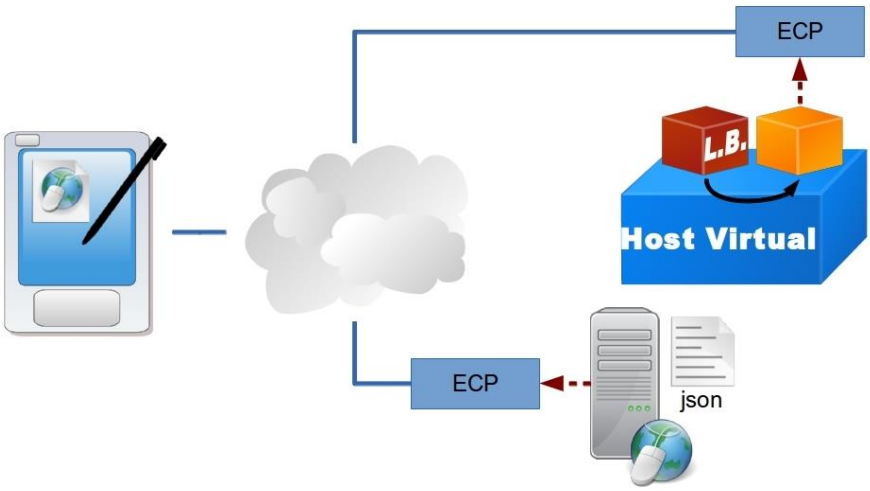

Fig 7. Diagrama de despliegue que muestra los componentes asociados en el proceso de autenticación a través de una aplicación móvil.

3.4 Aplicaciones no compatibles

Es necesario considerar que no todos los servicios que actualmente son utilizados por la comunidad universitaria pueden ser integrados al nuevo sistema de autenticación. En algunos casos, se debe a que dichos sistemas son soluciones a cuyo código fuente, no es posible acceder y no cuentan con un soporte para comunicarse con sistemas SSO.

a. Servidor de aplicaciones virtuales

En este caso en particular, la universidad cuenta con un sistema SaaS, el cual permite ejecutar en línea programas Windows, sin la necesidad de instalarlos, haciendo uso de conexiones RDP a un servidor Windows 2012 Standard.

Por ello, para su adecuado funcionamiento, ha sido necesario que el servicio (un servidor identificado con las iniciales T.S. - terminal server - en la Figura 8) cuente

con acceso de sólo lectura a la base de datos de los usuarios para poder autenticar y autorizar a quienes accedan a la plataforma.

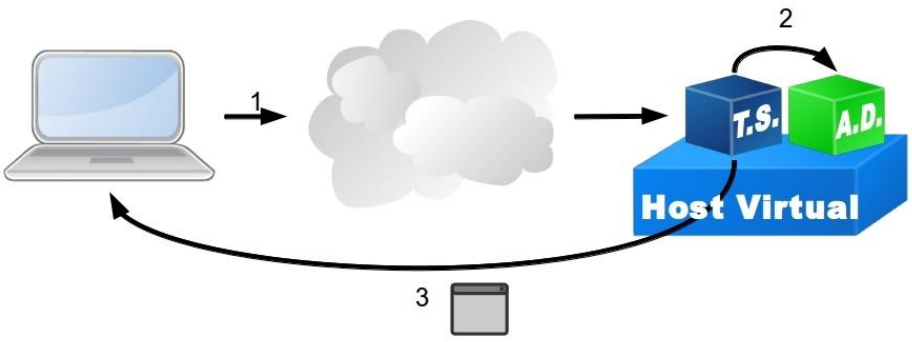

Fig 8. Diagrama de despliegue que muestra los componentes asociados en el proceso de autenticación al servidor de aplicaciones virtuales.

b. Aulas informáticas

Del mismo modo, las aulas informáticas de la universidad (la mayoría cuentan con el sistema operativo Microsoft Windows) requieren que tanto el usuario, como la contraseña sea la misma que la utilizada para autenticarse en las aplicaciones web.

Se procedió a revisar la herramienta pGina, que permite sobreescribir el sistema de autenticación de Microsoft Windows para que éste pudiera validar las credenciales de acceso contra una base de datos (como por ejemplo MySQL) o una herramienta basada en LDAP. Asimismo, dada la apertura del código, brinda la posibilidad de extender las funcionalidades de autenticación, pudiéndose implementar nuevos plug-ins que permiten validarse contra cualquier otro sistema, como el módulo ECP presente en las últimas versiones de Shibboleth.

Sin embargo, dadas las políticas de seguridad que deben ser configuradas en cada una de las PCs de las aulas; así como la flexibilidad de crear políticas ajustadas a las características del equipo en el que el usuario inicia sesión (en nuestro caso, el equipo podría ser tanto una PC como el servicio que aprovisiona acceso a aplicaciones bajo demanda), se optó por hacer uso de la herramienta Microsoft Active Directory.

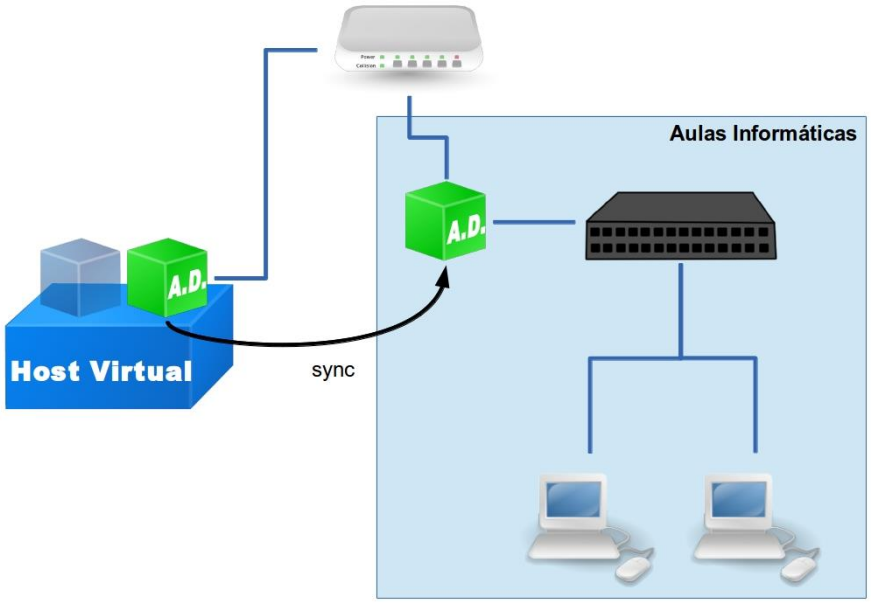

Fig 9. Diagrama de despliegue que muestra los componentes asociados al proceso de autenticación en las aulas informáticas.

A fin de evitar posibles problemas de autenticación a causa de caídas de comunicación entre las Aulas Informáticas y el Datacenter de la Dirección de Informática Académica, se desplegó una réplica del Active Directory en modo de solo lectura en el Datacenter del edificio de las aulas, como se muestra en la Figura 9.

Con ello se redujo el riesgo de que los alumnos no pudieran iniciar sesión en los laboratorios debido a una falla en la red; así como reducir el tiempo de latencia al momento en que los usuarios se autentican en los equipos.

3.5 Sistemas de apoyo

Una vez que se cuenta con la infraestructura necesaria para poder soportar las transacciones que serán procesadas por la aplicación; es necesario cubrir tres funcionalidades que la plataforma Shibboleth no provee.

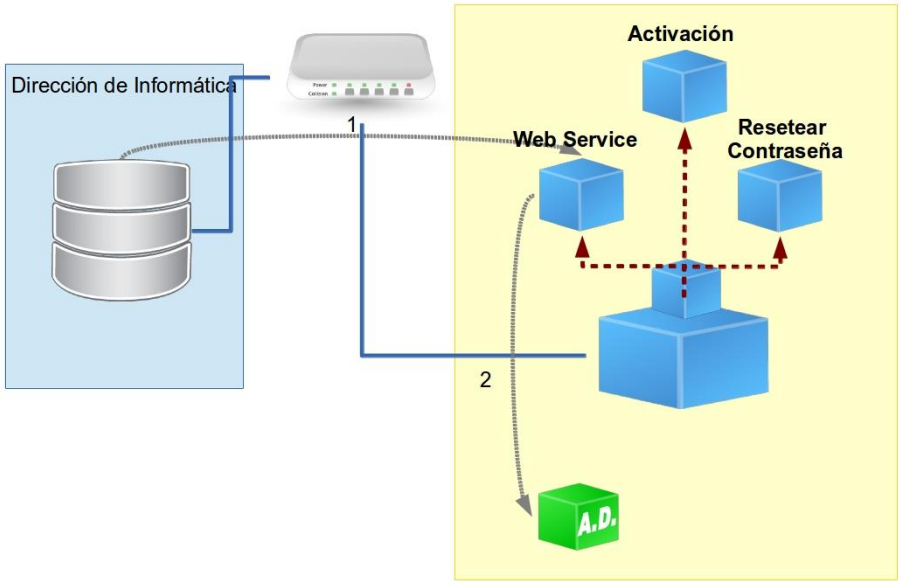

Fig 10. Diagrama de despliegue que muestra los componentes de apoyo que se despliegan sobre la plataforma Shibboleth.

1. Sincronización de los usuarios: Cada vez que un usuario es dado de alta en los sistemas de la universidad o cada vez que sus datos son actualizados, la aplicación recibe una notificación vía Web Service para proceder a actualizar la información almacenada en el Active Directory.

2. Gestionar la inicialización de las contraseñas de los usuarios: No es posible migrar las contraseñas de las cuentas de los usuarios actualmente registrados en los sistemas de la universidad, por ello, se implementó una aplicación que permitiera a los usuarios, activar su cuenta e ingresar una nueva contraseña a ser utilizada dentro del nuevo sistema de logueo centralizado.

3. Brindar al usuario la posibilidad de cambiar su contraseña: Siempre queda la posibilidad que un usuario olvide su contraseña, por ello, se desarrolló una aplicación que permitiera a los usuarios generar una token que les permitiera efectuar una actualización de sus contraseñas.

4 Conclusiones y Planes Futuros

4.1 Conclusiones

Una solución AAI Federada Académica abre oportunidades estratégicas de integración con otras universidades del país y el extranjero para compartir el acceso a sus aplicaciones y con ello facilitar el acceso a nuevo conocimiento a los miembros de la comunidad universitaria.

A pesar de que el IdP no cuenta con módulos precargados que permitan clusterizar su arquitectura es posible incluir librerías externas que faciliten el proceso de clusterización a fin de poder brindar una herramienta tolerante a fallos y que garantice una alta disponibilidad.

Shibboleth ofrece integración con diversidad de aplicaciones y puede funcionar junto a servicios de validación en red como el Active Directory, lo cual conviene para una estrategia de desarrollo BYOD, manejando un único usuario para todos los servicios de la institución.

Sobre la migración e integración de las aplicaciones y servicios web desarrollados por la DIA, no ha habido inconvenientes con esta tarea indistintamente del lenguaje bajo el que hubieran estado escritos (PHP, Java y Python). Para el caso de las aplicaciones móviles, se ha requerido habilitar el módulo ECP. Se espera que para la versión 3 del IdP sea menos complejo la interconexión de aplicaciones móviles con el sistema Shibboleth.

Una parte muy importante en el uso de Shibboleth es contar con los datos de los usuarios los cuales deben ser refrendados, de no ser así, los usuarios pueden tener la falsa percepción de que el sistema está fallando o presenta un comportamiento anómalo. Lo cual, como consecuencia, puede generar malestar a los usuarios y con ello, una desconfianza en el sistema.

4.2 Planes Futuros

Durante el año que el servicio ha estado en funcionamiento, ha sido posible recibir retroalimentación por parte de los usuarios del sistema quienes presentan confusión durante el proceso de activación de cuentas; así como durante el cierre de sesión - caso presentado en el estudio de Suroanta et al. [11]. Es por ello que se propone la necesidad de un estudio de usabilidad a nivel del proceso de activación y cierre de sesiones a fin de reducir problemas de usabilidad que actualmente están presentes en el sistema.

Bajo el esquema de despliegue actual, algunas aplicaciones como el Servidor de Aplicaciones Virtuales o servicios como las Aulas Informáticas se conectan directamente a la base de datos de usuarios de Shibboleth (en este caso el Active

Directory), siendo lo ideal que todo sea hecho con el protocolo SAML2. Por ello se propone evaluar la implementación de un plugin bajo pGina o una solución similar que permita hacer uso del protocolo SAML2 en la autenticación de los usuarios en las aulas informáticas.

Asimismo, de acuerdo a la documentación aprovisionada por SWITCH [12], está entrando en vigencia la versión 3 del servidor IdP. Se propone una validación de dicha tecnología a fin de medir el impacto durante la migración y su posterior puesta en marcha.

Finalmente, en el país está en fase de análisis, el despliegue de un DS, bajo el nombre de Proyecto INCA, que actuará como puente en la autenticación entre los diversos servicios y casas de estudio pertenecientes a la Federación Peruana de Universidades. Dado que es el primer proyecto en el país que busca llevar a cabo una integración a este nivel, es de interés llevar a cabo un análisis en torno a aspectos de seguridad y políticas que pueden ser adoptadas a fin de garantizar la privacidad de los datos.

Agradecimientos

Los autores desean expresar su agradecimiento a los ingenieros Mario Salcedo y Carlos Chuquillanqui por su participación en la configuración y despliegue de los servidores Microsoft Active Directory.

Referencias

1. M. Burnett and D. Kleiman: Perfect password: Selection, protection, authentication. Syngress, (2006).
2. EDUCAUSE: Seven Things You Should Know About Federated Identity Management. En: Educause Learning Initiative, pp. 2 (2009)
3. D. Florencio and C. Herley: A large-scale study of web password habits. En: Proc. 16th Int. Conf. World Wide Web - WWW '07, pp. 657 (2007)
4. G. Haron and D. Maniam:Re-engineering of web reverse proxy with shibboleth authentication. En: Internet Technol. …, pp. 325–330 (2012)
5. Internet 2: Shibboleth, http://shibboleth.net/about/. [Accesado: 23 de Abril de 2015].
6. R. L. B. Morgan, S. Cantor, S. Carmody, W. Hoehn, and K. Klingenstein:Federated Security: The Shibboleth Approach. Educ. Q., vol. 27, no. 4, pp. 12–17 (2004)
7. W. Norris and M. Yoshitaka: Shibboleth Plugin for Wordpress, https://wordpress.org/plugins/shibboleth/. [Accesado: 23 de Apr de 2015].
8. A. Pashalidis and C. J. Mitchell: A taxonomy of single sign-on systems. En: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 2727 LNCS, pp. 249–264 (2003)
9. C. Powell, T. Aizawa, and M. Munetomo: Design of an SSO authentication infrastructure for heterogeneous inter-cloud environments. En: 2014 IEEE 3rd International Conference on Cloud Networking (CloudNet), 2014, pp. 102–107 (2014)
10. Shibboleth: Shibboleth Documentation, https://wiki.shibboleth.net. [Accesado: 23 de Apr de

2015].
11. S. Suoranta, K. Manzoor, A. Tontti, J. Ruuskanen, and T. Aura: Logout in single sign-on systems: Problems and solutions. En: J. Inf. Secur. Appl., vol. 19, no. 1, pp. 61–77 (2014)
12. SWITCH: SWITCH AAI, https://www.switch.ch/aai/. [Accesado: 23 de Apr de 2015].