Where Are You From?

-Shibboleth och WAYF på KIB

Universitetsbiblioteket

Projektgrupp:
Johan Finn
Karin Perols
Henrik Åkerfelt

Innehållsförteckning

1. Bakgrund

Den mest använda och legio metoden för accesskontroll för licensierade e-resurser i den akademiska biblioteksmiljön idag är IP-adresskontroll. Metoden går ut på att användaren ges tillgång till licensierat material genom att förlagets tekniska plattform känner av att användarens internettrafik kommer från en IP-adress som ägs av lärosätet ifråga. IP-adresskontroll är vitt accepterat som den allenarådande metoden för accesskontroll för institutionella licenser. På senare tid har det dock höjts röster både inom akademin och från leverantörshåll om att det finns ett flertal problem med denna metod.

Att registrera ett universitets samtliga IP-adresser för varje förlagsplattform och/eller prenumeration är en tidsödande administration. Då det finns olika format i vilka man kan uttrycka IP-adresser och konsensus om hur dessa bör uttryckas är svag blir detta lätt en källa till missförstånd som kan vara svåra att reda ut. Att erbjuda distansåtkomst genom IP-adresskontroll är inte heller självklart. Då en student eller anställd vid lärosätet t.ex. sitter hemma och jobbar gäller inte längre hemuniversitets IP-adress utan användaren får då en IP-adress som tilldelats av den egna internetleverantören. För att erbjuda distansåtkomst måste biblioteket använda sig av tredjepartslösningar. Idag används på KIB en s.k. URL re-writing proxyserver. URL:er till licensierat material skrivs om för att tunnlas genom KIBs egen server och på så sätt ser trafiken för leverantörens del ut att komma från en IP-adress som ägs av lärosätet. Detta ställer en del specifika krav på både användare och bibliotekets administratörer. Särskilda länkar måste användas för att erbjuda proxytillgång i vilka ett prefix skjuts in som skickar användaren via lärosätets server och därmed dess IP-adress. Länkar som biblioteket inte kontrollerar, t.ex. citeringslänkar mellan utgivare eller direktlänkar i PubMed kan inte hanteras genom URL re-writing proxy. Samtliga URL:er som ska kunna nås genom proxyn måste registreras i proxyserverns konfigurationsfil. Det är en tung administration då URL:er tenderar att byta form med viss regelbundenhet och då även måste uppdateras i proxyns konfigurationsfil. Om användaren måste trafikera genom en brandvägg kan även detta vara problematiskt.

Utöver en omfattande administrativ börda finns även en säkerhetsaspekt på vilken autentiseringmetod man väljer att använda. En användare som autentiseras på basen av den IP-adress som hens trafik kommer ifrån är i stort sett anonym från ett leverantörsperspektiv. Detta kan naturligtvis vara både positivt och negativt. Förlaget kan endast se vilken institution som äger IP-adressen medan användaren bakom endast kan spåras via lärosätets egna webbloggar vilket även med en exakt tidsangivelse kan vara krångligt med tanke på den mängd trafik som passerar genom universitetets servrar. Normalt förfarande vid otillbörlig nedladdning är att hela IP-adressen blockeras av förlaget vilket innebär att ingen som kommer genom denna IP-adress tillåts komma åt det licensierade materialet. Spärren hävs först då webbloggar har analyserats för att identifiera den specifika användaren och användarkontot har spärrats. Detta kan innebära ett stort avbräck för många användare under en längre tid. Proxyservrar är också utsatta för risk för intrång på ett annat sätt än ett internt IP-nätverk. Dessa attacker kan vara svåra att upptäcka. Det kanske mest kända och uppseendeväckande exemplet är från 2002 då mer än 50 000 artiklar otillbörligt laddades ner från JSTOR, genom hackade proxyservrar, innan någon upptäckte intrånget. (Carnevale, 2002)

IP-filtrering är till syvende och sist ett slags autentisering av institution snarare än användare. Det är på institutionsnivå som en användare autentiseras och ingen information förutom vilken institutions nätverk som användaren trafikerar genom skickas vidare till leverantören. Institutionen går på detta sätt i god för användare som varken institutionen eller leverantören vet något om.

Autentisering på institutionsnivå är också otillräcklig för mer personaliserade tjänster. Webbtjänster som bygger på t.ex. personliga konton kommer alltid att kräva ytterligare en autentisering på individnivå. Exempel på sådana tjänster kan vara EndNote, Lynda.com, reSEARCH etc.

Det finns alternativ till IP-adresskontroll och URL re-writing proxy för att erbjuda effektiv hantering av identiteter och autentisering till licensierat material. Det mest lovande alternativet, som också fått viss spridning framförallt i Storbritannien, är Shibboleth.

Av anledningarna ovan har vi sett ett behov av att utföra en förstudie om vilka konsekvenser en övergång från IP-adresskontroll till Shibboleth som huvudsaklig autentiseringsmodell skulle föra med sig för KIB. I denna rapport redovisar vi slutsatserna av nämnda förstudie tillsammans med en grundläggande genomgång av Shibboleth som autentiseringsmodell. Vi för också en diskussion kring vilka förutsättningar, utmaningar och frågeställningar en övergång skulle innebära. Rapporten avslutas med en rekommendation för hur KIB kan gå vidare.

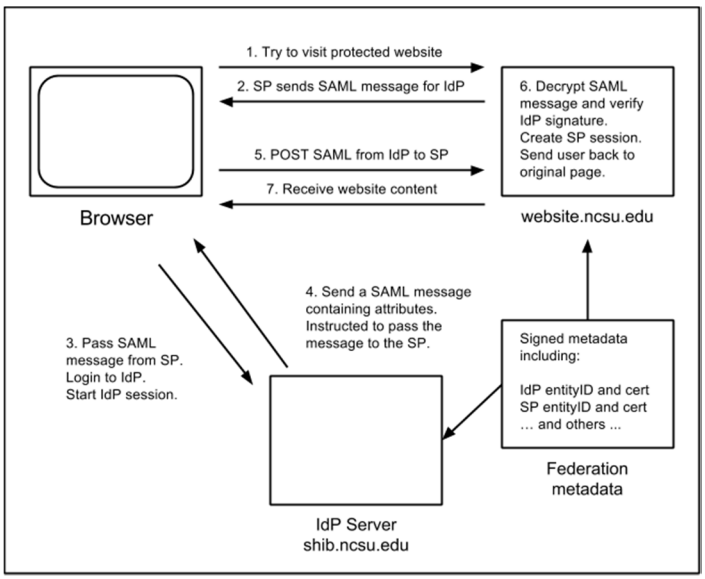
2. Vad är Shibboleth?

Shibboleth är något förenklat ett paket av programvara för identitetskontroll. Shibboleth-programvaran gör det möjligt att ge en grupp användare från olika organisationer tillgång till licensierade e-resurser eller webbtjänster med en elektronisk identitet från någon av organisationerna. Programvaran erbjuds med öppen källkod från Internet2: http://shibboleth.internet2.edu/. (Axelsson, 2007). Den här typen av identitetskontroll kallas för Web Single Sign On och är en implementation av den öppna standarden SAML2. Shibboleth-programvaran är inte nödvändig för att erbjuda Web Single Sign On, huvudsaken är att programvaran man använder sig av bygger på SAML2 standarden. (Lunds Universitet, 2016)

Förutom autentisering kan Shibboleth även överföra information om den specifika användaren genom s.k. attributöverföring. T.ex. kan överföring av personnummer, e-postadresser, roller inom organisationen, samt rättigheter överföras. Vilka attribut som kan eller ska överföras bestäms i överenskommelse mellan de som erbjuder webbtjänsten ifråga och användarorganisationen. Överföringen sker anonymiserat men användaren går att spåra till en specifik session och kan därför blockeras på kontonivå (jmf IP-kontroll där all trafik från en IP-adress blockeras vid överträdelse).

2.1 Teknisk infrastruktur

Den tekniska infrastrukturen för Shibboleth-programvaran kan sägas bestå av ett antal olika komponenter: (1) en Service Provider (SP), (2) en Identity Provider (IdP), som i sin tur är uppdelad i single-sign-on, Authentication Authority respektive Attribute Authority, samt (3) en valbar WAYF-tjänst. När en användare loggar in via en IdP som är ansluten till en SP genom Shibboleth skickas ett POST-meddelande till SP:n, kallat en Authentication Assertion, vilket innebär att IdP:n meddelar SP:n att användaren har autentiserats. Detta är tillräckligt för att användaren ska ges tillgång till tjänsten. Sibboleth är uppbyggt kring detta tillitskontrakt, SP:n litar på att IdP:n utför en tillräcklig autentisering. Om SP:n kräver några särskilda attribut skickas en attributförfrågan tillbaka till IdP:n. IdP:n svarar med ett paket med förutbestämda attributdata om den specifika användaren. Dessa kan användas för att begränsa tillgång till tjänsten till vissa användargrupper inom organisationen eller vissa organisationer inom en identitetsfederation. De meddelanden som skickas mellan IdP:n och SP:n skickas över protokollet SAML 2. Flödet mellan IdP och SP beskrivs nedan i Fig. 1.



(Fig. 1 SAML 2-meddelande i Shibboleth-autentiseringsflöde)

2.2 Federationer och interfederationer

För att förenkla administrationen för alla inblandade i autentiseringsprocessen är det vanligt att liknande organisationer går ihop i en s.k. federation. När en organisation ingår i en federation innebär detta ett kontrakt om tillit. Alla medlemmar i federationen litar på den autentisering som sker inom alla andra deltagande organisationer. Det system eller den tjänst som kopplats mot en Shibboleth-federation för en organisations räkning kan välja att tillåta inloggningar från flera olika IdP:er. Finns IdP:n i federationen accepteras inloggningen. Efter inloggning kan ett paket med information om användaren levereras till systemet ifråga. Ytterligare slagningar mot t.ex. LDAP behövs inte för att få information om namn, e-postadress, organisationstillhörighet osv. (Axelsson, 2007). Behörighet till system kan sedan begränsas genom olika attribut så som organisationstillhörighet, typ av användare (student, forskare eller anställd), institutionstillhörighet osv. Attributen tilldelas i det centrala identitetsregistret. I KIs fall utgörs

det centrala adressregistret av KIMKAT. Datautbytet mellan användarens browser och organisationens IdP är krypterad för att hålla en hög säkerhet och integritet.

För att ytterligare förenkla administrationen av Shibboleth kan flera federationer gå ihop i en interfederation för att dela metadata och utöka tillitskontraktet. En interfederation kan ses som en superfederation, eller en federation av federationer. Om en SP har registrerat sig med någon federation som ingår i interfederationen kan medlemmar från alla andra federationer också logga in i denna SP:s tjänst.

2.2.1 SWAMID och EduGAIN

SWAMID

SWAMID-Swedish Academic Identity Federation (https://www.sunet.se/swamid) är en identitetsfederation som omfattar de flesta lärosäten, forskningsinstitut och andra myndigheter som är relaterade till svensk forsknings- och utbildningssektor. SWAMID drivs av SUNET. Syftet med federationen är att sänka kostnaden för elektronisk identitetshantering. I första hand är SWAMID tänkt att hantera autentisering med elektroniska identiteter mot stora allmänna och vitt spridda system så som Ladok, Project Companion, Box och liknande, men många leverantörer av licensierade informationsresurser har också implementerat Shibboleth som en alternativ autentiseringsmetod. För att upprätthålla tillitskontraktet mellan SWAMIDs olika medlemmar ställs krav om ”Best Practice” för de organisationer som vill ansluta sig. (SUNET, 2016)

EduGAIN

EduGAIN (http://services.geant.net/edugain) är en Europeisk interfederation. Tjänsten har utvecklats i ett samarbete mellan European National Research and Education Network (NREN) och Europeiska Unionen. Interfederationen knyter samman en mängd nationella och regionala identitetsfederationer framför allt inom Europa men också från andra delar av världen så som Australien, USA och Japan. Genom EduGAIN-samarbetet sänks både kostnad och administrativ börda för alla inblandade parter i autentiseringsprocessen. Metadata om SP:er delas mellan medlemsfederationerna. Tjänsteleverantörer behöver inte registrera sig med multipla federationer. För medlemsfederationerna innebär EduGAIN att de kan erbjuda sina användare fler resurser på ett enkelt sätt utan ökad omkostnad. (Géant, 2016)

En stor del av de informationsresurser som KIB licensierar är desamma som licensieras av brittiska universitet. Den brittiska federationen för högre utbildning, UK Access Management Federation (UKAMF), har varit mycket aktiv med att ansluta informationsresurser till sin identitetsfederation. EduGAIN-samarbetet innebär att dessa inte behöver anslutas till SWAMID för Shibboleth-autentisering genom KIs IdP.

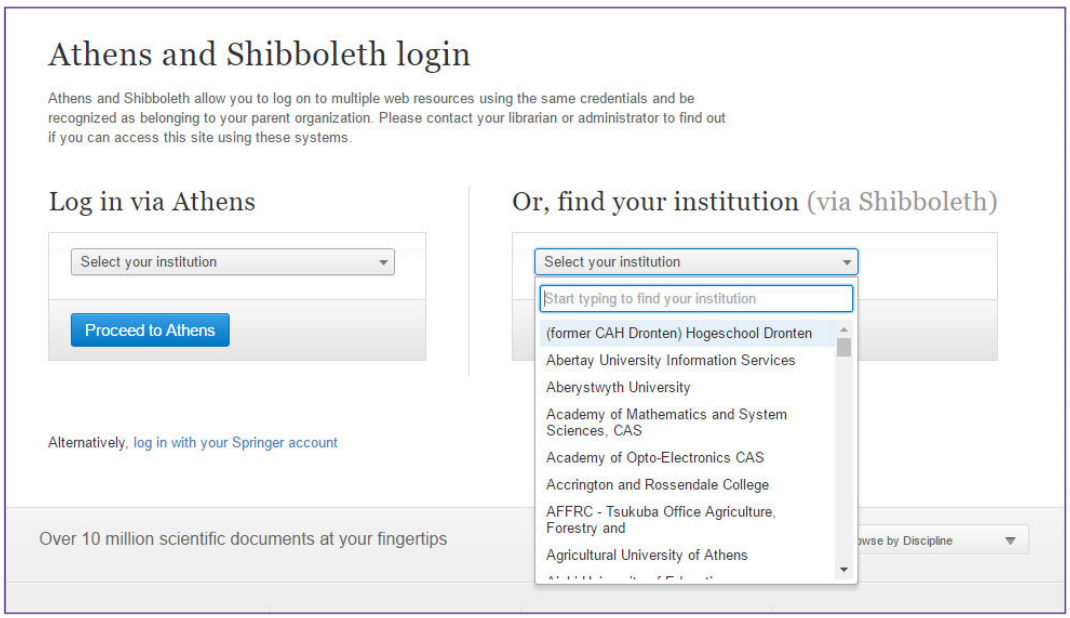
3. Shibboleth och WAYF på KIB

Vi använder redan idag Shibboleth-autentisering på KIB i begränsad utsträckning. Bland annat används Shibboleth för att möjliggöra Single-Sing-On till vår proxyserver och för en del interna administrativa system så som Primula och Agresso. Det vi i projektet ville undersöka är hur Shibboleth skulle kunna fungera som autentiseringsmetod direkt mot förlag och andra leverantörer av licensierade e-resurser. Vi kontaktade i detta syfte ett urval av leverantörer för att sätta upp inloggning via vår IdP. Vi hörde även av oss till BIBSAM för att ta reda på om andra svenska bibliotek använder Shibboleth för nationellt förhandlade resurser. Eftersom vi kände till att Shibbolethanvändning är utbrett i Storbritannien tog vi kontakt med ett par brittiska bibliotek. Vi fick svar från universitetsbiblioteket i York där deras Electronic Resources Co-ordinator gick igenom vilka autentiseringsmetoder de använder sig av och för vilken typ av resurser. Detta var till stor hjälp för oss i vårt fortsatta resonemang.

3.1 Inloggningsförfarande

WAYF

Det finns flera olika sätt att sätta upp inloggningsförfaranden med hjälp av Shibboleth för licensierade produkter. Shibboleth-inloggning kan vara oberoende av administrerade länkar från biblioteket. Det spelar då ingen roll hur användaren hittar till det licensierade materialet eller var användaren fysiskt befinner sig. På de allra flesta förlagsplattformar finns ett alternativ för institutionell inloggning (Fig. 2). Man presenteras då med en s.k. WAYF-prompt. WAYF är en akronym för Where Are You From och användaren förväntas här ange vilken organisation hen vill autentisera sig genom.



(Fig. 2 WAYF-prompt på SpringerLink)

Fördelen med detta inloggningsförfarande är att det inte bygger på att användaren startar sin informationssökning på bibliotekets webbplats. Som vi sett i flera undersökningar, inte minst vårt eget personaprojekt, är bibliotekets webbplats inte den självklara startpunkten för alla våra

användargrupper. Den här autentiseringsmodellen är också bra när tillhörigheten inte är självklar, t.ex. vid samarbetsprogram och utbildningar samt forskningscentra där de anställda kan ha blandade organisationstillhörigheter. Man kan då genom Shibboleth styra behörighet via andra attribut än den fysiska platsen eller organisationstillhörighet. Nackdelen är att få användare är bekanta med förfarandet och det är inte alltid självklart och intuitivt hur man loggar in via sin institutions IdP.

WAYFLess

Ett annat sätt att autentisera genom Shibboleth är att skicka med information om vilken IdP som ska användas i URL:en till själva resursen. Användaren kan då helt kringgå WAYF-prompten och tas direkt till sin organisations IdP-inloggning. Efter inloggning skickas användaren vidare till den resurs hen försöker komma åt. Denna metod kallas vanligen för WAYFLess URL. Denna typ av URL går att skapa för de allra flesta resurser på olika nivå, t.ex. till en tidskrifts startsida, på häftesnivå eller på artikelnivå. Detta förutsätter specialkonstruerade länkar och sålunda också att användaren startar sin informationssökning på bibliotekets webbplats. Metoden är mer intuitiv för slutanvändaren men innebär också ett större mått av administration.

Det finns två olika sorters WAYFless-länkar, SP-side WAYFless URL och IdP-side WAYLless URL. SP-side WAYFless URL skapas och stöds av en SP. IdP-side WAYLless URL skapas på egen hand och måste innehålla en hårdkodad URL till servicekomponenter. SP-side URL:er är stabilare och flexiblare och därmed att föredra då länkarna som är förutsättningen för IdP-side URL:er kan ändras utan förvarning av förlagen utanför institutionernas kontroll.

Vårt bibliotekssystem Alma innehåller WAYFless-länkar till 12 av våra leverantörer (103 s.k. collections) vilket gör det väldigt enkelt att slå på denna funktion, det enda som behöver göras är att ange KIs IdP i ett angivet fält per collection. De flesta av dessa 12 leverantörer är stora förlag vilket gör att huvuddelen av vårt bestånd skulle gå att göra tillgängligt med WAYFless-länkning på ett enkelt sätt. Det går också att skapa egna WAYFless-länkar i ett kalkylblad och ladda upp till Alma men det kräver en större egen arbetsinsats.

Bägge inloggningsmetoder innebär, om de inte kombineras med vanlig IP-adresskontroll, att samtliga användare får logga in via IdP oberoende av vilket nätverk de sitter på. Förfarandet blir konsekvent men tvingar i vissa situationer fram en extra inloggning. En bra sak med Shibboleth är att endast en inloggning behövs så länge användaren har samma aktiva webbläsarsession, även om resurserna är olika. Väl användaren är autentiserad via sin IdP är hen inloggad för alla resurser under hela sin session.

3.2 Walk-in-use

Shibboleth-autentisering bygger på att lärosätena själva tar hand om autentiseringsförfarandet. För att detta ska kunna göras på ett säkert sätt måste inloggningen vara knuten till ett kontrollerat och centralt adressregister. I KIs fall är detta register KIMKAT. I våra licenser med förlag och andra leverantörer har vi förhandlat om walk-in-use. Detta innebär att vem som helst som besöker vårt bibliotek har rätt att använda våra licensierade resurser så länge de befinner sig i våra lokaler. Det här går alldeles utmärkt att administrera genom IP-adresskontroll då våra publika datorer ligger inom de godkända IP-adresser som ska ges tillgång till leverantörens tjänster. Med Shibboleth-autentisering uppstår dock här ett problem. En Shibbolethinloggning

förutsätter att användaren finns registrerad i vår IdP och därmed i vårt centrala adressregister. För att KIB ska kunna uppfylla Walk-in-use-kriteriet måste alltså varje besökare som vill använda sig av någon av våra licensierade resurser registreras i KIMKAT. Även om detta skulle vara möjligt måste ytterligare registrering hos leverantören göras, då dessa personer endast ska ges tillgång i de fall de sitter på vårt publika nätverk. Detta förutsätter dels ett helt nytt attributvärde i EduPersonScopedAffiliation och en dubbelautentisering genom både IdP-inloggning och IP-adresskontroll. Det här är inte gängse sätt att sätta upp Shibboleth och det ligger nära till hands att tro att det skulle bli en omständig administration om det ska göras mot alla våra leverantörer. I våra kontakter med ITA har det också varit ganska tydligt att det inte är önskvärt att registrera biblioteksbesökare i KIMKAT. Då kvarstår alternativet att autentisera mot flera IdP:er. Vi skulle teoretiskt sett kunna sätta upp en Shibboleth-autentisering mot vårt eget AD, där vi redan registrerar besökande användare för att de ska få en datorinloggning till våra publika datorer. Problemet blir här att AD:et i fråga inte är registrerat som IdP varken hos SWAMID eller hos EduGAIN. Det förutsätter dessutom att varje användare gör ett aktivt val mot vilken IdP-autentisering ska ske. Lösningen skulle innebära en tung administrativ börda för biblioteket.

Walk-in-use är inte särskilt utbrett i andra länder (detta gäller i synnerhet Storbritannien) varför det är svårt att hitta exempel och best practice. Vi anser att denna fråga bör utredas vidare, för att om möjligt hitta en lösning för att jämka kravet på Walk-in-Use med Shibboleth-autentisering.

3.3 Uppsättning och administration

I syfte att praktiskt undersöka processen för Shibboleth-registrering gjorde vi ett mindre urval av leverantörer att testa med och utvärdera processen. För att sätta upp Shibboleth-inloggning med en förlagsplattform kontaktar man förlaget i fråga och ber dem skapa en inloggning för ens institution. Som nämnts ovan har många av våra leverantörer redan satt upp autentisering mot UKAMF och via samarbetet i EduGAIN är det en enkel process för leverantörerna att lägga till ytterligare institutioner. Det som då skapas är en WAYF-inloggning.

Inom projektet tog vi kontakt med fyra leverantörer för att sätta upp Shibboleth-inloggning, EBSCO för deras databaser, samt American Medical Association (AMA), MA Healthcare och Cambridge University Press. Med två av leverantörerna, EBSCO och MA Healthcare var det en väldigt smidig process, de visste vilken information de behövde och kunde snabbt sätta upp autentiseringen. MA healthcare upptäckte dock att det fattades information i flödet från EduGAIN vilket vi kunde kontakta SWAMID om och be dem rätta till. Bägge dessa leverantörer har avtal med UKAFM.
Med Cambridge gick det inte lika smidigt och det föll mellan stolarna hos dem då de var mitt uppe i ett plattformsbyte. Efter plattformsbytet gick det dock snabbt att få igång Shibboleth-inloggning vilket var väldigt värdefullt då vår proxyserver inte kan hantera deras nya plattform. Vår fjärranvändning av Cambridge går just nu (september 2016) via Shibboleth. Med AMA har Shibboleth-autentiseringen ännu inte kommit igång. Det har varit svårt att nå fram med informationen om att SWAMID är en del av EduGAIN och att de därför inte behöver registrera sig i ytterligare en federation vilket, enligt uppgift, skulle kosta dem $3000. AMA använder sig av den aggregerade plattformen Highwire.

Av detta något begränsade underlag kan man sluta sig till att det är en tämligen enkel process att sätta upp Shibbolethinloggning med de leverantörer som redan har avtal med UKAFM. De är vana att hantera denna typ av frågor. Det kan vara svårare med leverantörer som har sitt innehåll på aggregerade plattformar så som Highwire, Ingenta o.dyl då de inte styr över infrastrukturen själva. Kunskapen om federationer som EduGAIN verkar variera.

4. Slutsats och rekommendation

Emedan vi ser stora fördelar med Shibboleth som autentiseringsmodell kan vi i dagsläget, av olika orsaker, inte rekommendera att KIB helt går ifrån IP-baserad autentisering. Under projektets gång, när vi praktiskt utforskat Shibboleth tillsammans med ett urval av leverantörer, har det visat sig svårt att förena Shibboleth-autentisering med kravet på walk-in-use. Vi kan heller inte erbjuda stabil SP-side WAFLess-länkning till samtliga våra elektroniska resurser. Även vår URL re-writing proxyserver bör finnas kvar för att erbjuda distansåtkomst då inloggningsförfarandet via WAYF-prompt inte bedöms som tillräckligt intuitivt. Detta borde kunna undersökas vidare genom t.ex. mindre användartester eller annan UX-metod.

Projektgruppen anser trots detta att det är hög tid att börja arbetet med att långsiktigt förbereda för en annan typ av autentiseringsmodell. Infrastrukturen finns redan på plats och det finns anledning att redan nu erbjuda Shibboleth-inloggning till merparten av våra e-resurser som ett alternativ till IP-adresskontroll och URL re-writing proxy. Detta skulle gynna våra användare som hittar våra resurser genom andra verktyg och tjänster än bibliotekets egna. Eftersom proxyhanteringen har visat sig vara en betungande administration med många felkällor vore det också värdefullt att kunna erbjuda alternativ där proxyn av någon orsak fallerar.

Vi rekommenderar därför att tillgängliggörandegruppen på Mediaförsörjning ges i uppdrag att registrera förlagsplattformar för Shibboleth-autentisering där det är möjligt. Webbtjänster som i hög grad bygger på personalisering (så som EndNote webb, Lynda, EBL och liknande) bör prioriteras då vinsterna här är störst ur ett slutanvändarperspektiv. Det är viktigt att vi i detta arbete dokumenterar vilka förlag vi har en fungerande Shibboleth-inloggning med. Förslagsvis kan detta göras i Alma. Det kan även vara bra att dokumentera eventuella WAFLess-länkar till respektive resurs då dessa kan vara e-mediasupporten till stor hjälp vid eventuella avbrott i IP-åtkomst.

Då målet är att Shibboleth-inloggning ska erbjudas som ett alternativ till nuvarande autentiseringsförfarande genom WAYF-prompt kommer detta att behöva kommuniceras ut till våra olika användargrupper så snart en kritisk mängd av leverantörer har anslutits. Lämpligtvis kunde Mediaförsörjnings marknadsföringsgrupp ta fram ett förslag på hur detta ska göras. Särskilt fokus bör här ligga på distansanvändare och i synnerhet forskare.

Det är vidare önskvärt att Shibboleth-autentisering beaktas så som ett krav i förhandling med leverantörer i likhet med t.ex. remote access och COUNTER-statistik. Inköpsgruppen tillsammans med tillgängliggörandegruppen inom Mediaförsörjning bör fundera på hur dylika krav kan formuleras i licensförhandling.

Vad gäller walk-in-use i kombination med Shibboleth skulle det vara önskvärt att detta utreds vidare av DoS tillsammans med ITA i sin roll som IdP-ägare.

References

Axelsson, P. (2007). *Vad är shibboleth?* Swedish Alliance for Middleware Infrastructure.

Carnevale, D. (2002). *Security lapses on campuses permit theft from JSTOR database.* Hämtat från http://chronicle.com/free/2002/12/2002121201t.htm

Géant *About EduGAIN*. Hämtat från http://www.geant.org/Services/Trust_identity_and_security/eduGAIN/Pages/About-eduGAIN.aspx (2016-09-23)

Lunds Universitet. (den 19 aug 2016). *Shibboleth (3)*. Hämtat från Lucatbloggen: http://lucat.blogg.lu.se/files/2015/10/Shibboleth-3.pdf

SUNET. *SWAMID: Swedish Academic Identity Federation*. Hämtat från https://www.sunet.se/swamid/ (den 22 aug 2016).