Autenticazione SSO SAML via reverse-proxy per risorse “legacy” con autorizzazione per IP

Vincenzo Praturlon

Ufficio Coordinamento Centrale Biblioteche, Università degli Studi Roma TRE - vincenzo.praturlon@uniroma3.it

INTRODUZIONE:
Nel corso degli anni sono stati fatti dalle biblioteche con gli editori decine di contratti che prevedono l’accesso (anonimo) autorizzato per le sole sessioni provenienti dai gateway (NAT) di Ateneo.
L’accesso remoto era possibile per gli utenti solo tramite VPN (Virtual Private Network), di difficile configurazione, poco intuitiva e poco sicura.
Inoltre, in caso di abusi o compromissioni, era impossibile risalire all’utente associato alla sessione.

MATERIALI E METODI:
Gli applicativi delle biblioteche comprendevano un motore di meta-ricerca ed un link resolver OpenURL in grado di discriminare l’origine delle sessioni e di selezionare le risorse target che richiedevano l’autorizzazione per IP. Gli applicativo si integravano tramite un sistema SSO (Single Sign On) dedicato PDS (Patron Directory Service). Nonostante la tecnologia fosse di tipo proprietario, era disponibile una API (Application Programming Interface) SAML (Security Assertion Markup Language).

Nel 2009 è stato quindi realizzato un web reverse-proxy e tutti i componenti della infrastruttura sono stati messi in SSO via SAML.
Gli applicativi, per tutte le sessioni remote, e per le sole risorse che richiedono autorizzazione per IP, re-indirizzano le richieste HTTP verso il proxy, riscrivendo la URL.
All’utente una volta reindirizzato sul proxy viene richiesto di autenticarsi (se non l’ha già fatto).
Il proxy a sua volta restituisce le pagine richieste, grazie al fatto che le sue sessioni hanno origine da un IP incluso tra le classi autorizzate.

RISULTATI:
L’utente remoto può accedere alle risorse semplicemente effettuando la ricerca sui portali delle biblioteche e autenticandosi, laddove richiesto, una sola volta.
Inoltre, nonostante l’utente non sia profilato sulla risorsa target, correlando le sessioni è possibile risalire dall’accesso all’identificativo opaco, e da questo alla identità effettiva.
Nonostante la tecnologia datata, l’infrastruttura si è dimostrata funzionale e si pensa di mantenerla in produzione finché sarò necessario autorizzare per IP.

PROBLEMI APERTI:
- non tutto può essere proxato (ssl, problemi con java, risorse non web...)
- la sostituzione di uno dei moduli infrastrutturali deve mantenere le dipendenze illustrate

Nel nostro caso, la migrazione in corso da MetaLib a VuFind potrebbe richiedere una fase di sviluppo.
VuFind infatti può usare un proxy, ma non in base alla origine della sessione. Questa funzione è per ora delegata ad SFX (ma non tutte le richieste passano da SFX).
La “work-around” adottata per ora è quella di reindirizzare tutto il traffico verso il proxy. (sia LAN che WAN) .
EZproxy è infatti in grado di lavorare in modo trasparente per gli IP autorizzati, bypassando sia il re-write delle URL che l’autenticazione.
Nel riquadro Statistiche per utente è stato evidenziato questo contributo di traffico, che appare come “Not Authenticated” in quanto proveniente appunto dalla LAN di Ateneo.

	Platform	Total Journal Requests	Number of Databases	Total Database Searches	Total Database Sessions	Total Federated Searches	Total Federated Sessions
1	0009-8604	2	0				
2	0022-9258	1	0				
3	1461-4103	1	0				
4	1478-422X	1	0				
5	ACS Publications	10,618	0				
6	Allen Press Pinnacle	445	0				
7	AMS Website	309	0				
8	Annual Reviews	1,934	0				
9	APS	13,688	0				
10	Biomed Central	1,937	0				
11	CJO	7,172	0				
12	CSA Illumina		3	19,147		2,441	
13	de Gruyter Reference Global		100				
14	EASI		1	1,046	579		
15	EBSCOhost	5,68	156	436,169	123,465	11,738	
16	Emeroteca Virtuale CIBER	3,99	0				
17	Highwire	329	0				
18	HighWire Press	732	0				
19	IEEE Xplore	494	0				
20	Informaworld	1,850	16	1,138	317	0	0
21	IngentaConnect	81	1	4,477	7,182		
22	Insight	354	5	219	4,010	0	0
23	IOP Electronic Journals	105	426	0			
24	IOPscience	114	8,434	0			
25	JSTOR	2,077	42,381	1	15,449	9,484	0	0
26	mancevalerio	1	2	0			
27	MetaPress	4,685	27,730	0			
28	nature.com	14	13,999	0			
29	Ovid	1	17	16	3,072	2,338	0	0
30	OvidSP	4	84	0			
31	Project MUSE	588	5,519	0			
32	ProQuest	114	251	16	5,963	789	69	69
33	ScienceDirect	2,850	236,412	0			
34	Scitation	38	5,359	2	0	0		
35	Society for Industrial and Applied Mathematics	16	43	0			
36	store	4	4	0			
37	torrossa	57	322	0			
38	Wiley Online Library	2,297	36,788	0			
39	Emdr	1	76	0			

LE RISORSE PIU’ RICHIESTE IN GENERALE APPARTENGONO AL SETTORE SCIENTIFICO E SONO DI AMBITO INTERNAZIONALE

COME FUNZIONA:

DAL PUNTO DI VISTA WEB

Autenticazione:

A partire da SAML2 tutte le transazioni avvengono in SSL nel browser, tramite scambio di asserzioni in XML.
Il servizio viaggia quindi in modo sicuro sulla sola porta 80 senza creare dipendenze con le impostazioni dei firewall.

Autorizzazione:

URL richiesta dall’applicativo a EZproxy:
https://biblio-proxy.uniroma3.it/login?url=http://www.elsevier.com/journals/title/a

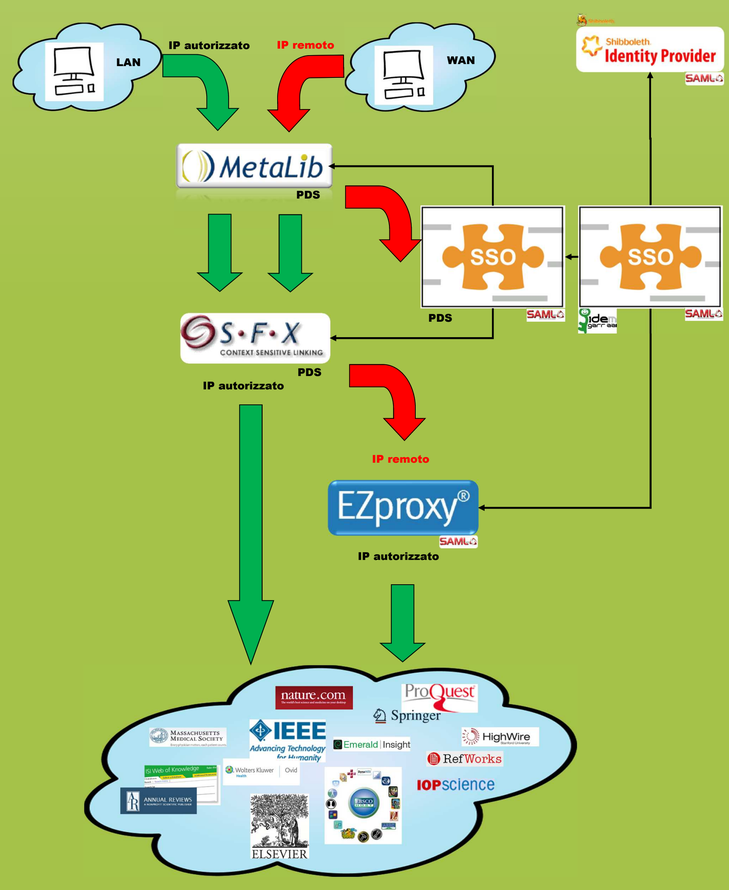
URL originale:
http://www.elsevier.com/journals/title/a

La risorsa vede il proxy come un normale client che apre sessioni HTTP da un IP autorizzato.

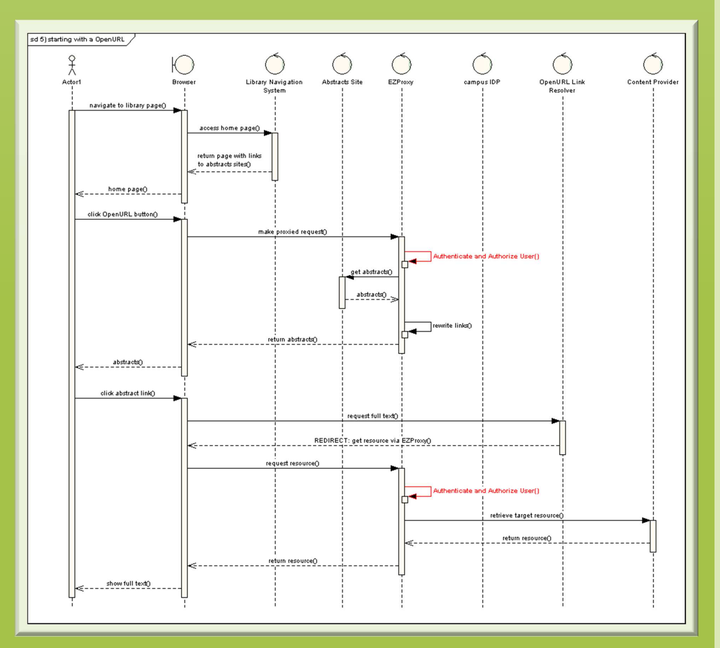
URL restituita da EZproyy al browser utente:
https://www.elsevier.com.biblio-proxy.uniroma3.it/journals/title/a

Il browser utente vede il servizio come un unico dominio biblio-proxy.uniroma3.it, al cui interno sono pubblicate virtualmente tutte le pagine delle risorse.
L’utente è quindi esposto in maniera diretta solo a risorse “interne” (proxy ed IdP).
Il sistema di proxy è “by hostname”, piuttosto che “by port”, per non creare dipendenze con le impostazioni dei firewall.

DAL PUNTO DI VISTA INFRASTRUTTURALE



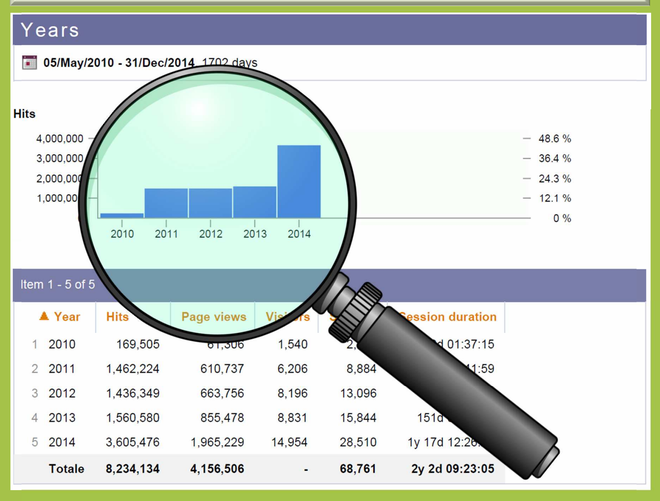
DAL PUNTO DI VISTA SAML



(non sono riportate le transazioni interne al PDS proprietario ExLibris)

QUANTO FUNZIONA:

STATISTICHE DI TRAFFICO

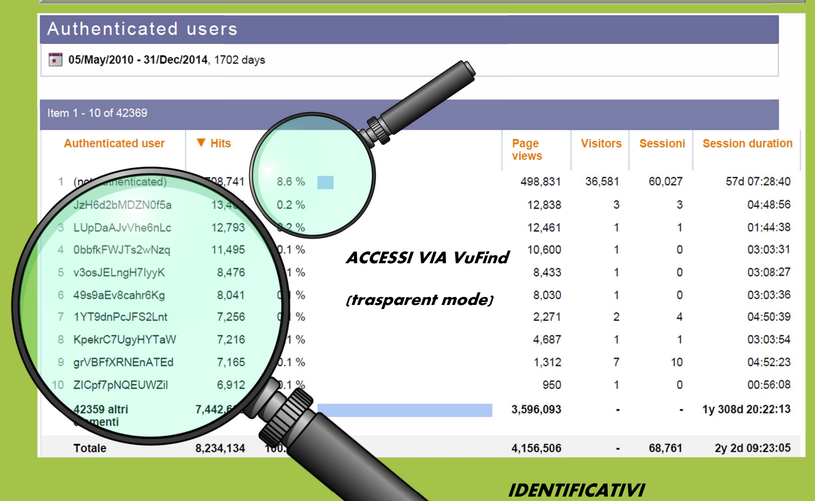


	Year	Hits	Page views	Visitors	Sessioni	Session duration
1	2010	169,505	61,306	1,540	2,[illegible]	[illegible] 01:37:15
2	2011	1,462,224	610,737	6,206	8,884	[illegible]1:59
3	2012	1,436,349	663,756	8,196	13,096	
4	2013	1,560,580	855,478	8,831	15,844	151d [illegible]
5	2014	3,605,476	1,965,229	14,954	28,510	1y 17d 12:26[illegible]
	Totale	8,234,134	4,156,506	-	68,761	2y 2d 09:23:05

GLI ACCESSI AUTORIZZATI PER IP CONTINUANO A CRESCERE

(nonostante IDEM...)

STATISTICHE PER UTENTE



	Authenticated user	Hits		Page views	Visitors	Sessioni	Session duration
1	(not authenticated)	[illegible]98,741	8.6 %	498,831	36,581	60,027	57d 07:28:40
2	JzH6d2bMDZN0f5a	13,[illegible]	0.2 %	12,838	3	3	04:48:56
3	LUpDaAJvVhe6nLc	12,793	0.2 %	12,461	1	1	01:44:38
4	0bbfkFWJTs2wNzq	11,495	0.1 %	10,600	1	0	03:03:31
5	v3osJELngH7lyyK	8,476	0.1 %	8,433	1	0	03:08:27
6	49s9aEv8cahr6Kg	8,041	0.1 %	8,030	1	0	03:03:36
7	1YT9dnPcJFS2Lnt	7,256	0.1 %	2,271	2	4	04:50:39
8	KpekrC7UgyHYTaW	7,216	0.1 %	4,687	1	1	03:03:54
9	grVBFfXRNEnATEd	7,165	0.1 %	1,312	7	10	04:52:23
10	ZICpf7pNQEUWZil	6,912	0.1 %	950	1	0	00:56:08
	42359 altri elementi	7,442,6[illegible]		3,596,093	-	-	1y 308d 20:22:13
	Totale	8,234,134	100.[illegible]	4,156,506	-	68,761	2y 2d 09:23:05

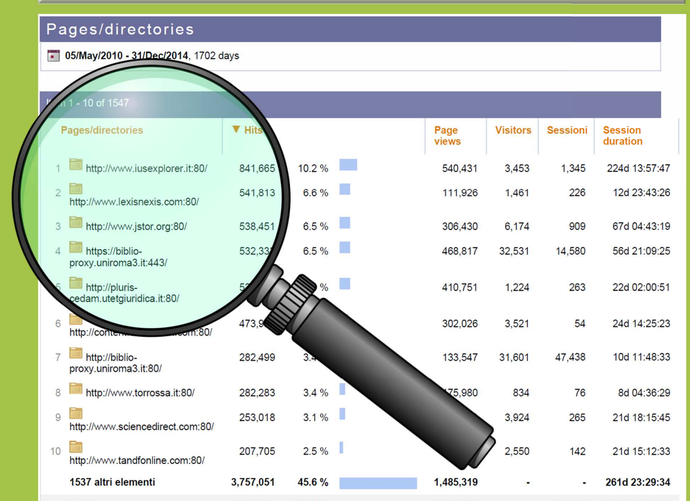
ACCESSI VIA VuFind

(trasparent mode)

IDENTIFICATIVI ANONIMIZZATI

(persistentID)

STATISTICHE PER RISORSA



	Pages/directories	Hits		Page views	Visitors	Sessioni	Session duration
1	http://www.iusexplorer.it:80/	841,665	10.2 %	540,431	3,453	1,345	224d 13:57:47
2	http://www.lexisnexis.com:80/	541,813	6.6 %	111,926	1,461	226	12d 23:43:26
3	http://www.jstor.org:80/	538,451	6.5 %	306,430	6,174	909	67d 04:43:19
4	https://biblio-proxy.uniroma3.it:443/	532,39[illegible]	6.5 %	468,817	32,531	14,580	56d 21:09:25
5	http://pluris-cedam.utetgiuridica.it:80/	5[illegible]	[illegible] %	410,751	1,224	263	22d 02:00:51
6	http://conten[illegible].com:80/	473,9[illegible]		302,026	3,521	54	24d 14:25:23
7	http://biblio-proxy.uniroma3.it:80/	282,499	3.4 %	133,547	31,601	47,438	10d 11:48:33
8	http://www.torrossa.it:80/	282,283	3.4 %	[illegible]75,980	834	76	8d 04:36:29
9	http://www.sciencedirect.com:80/	253,018	3.1 %	3,924	265	21d 18:15:45	
10	http://www.tandfonline.com:80/	207,705	2.5 %	2,550	142	21d 15:12:33	
	1537 altri elementi	3,757,051	45.6 %	1,485,319	-	-	261d 23:29:34
	Totale	8,234,134	100.0 %	4,156,506	-	68,761	2y 2d 09:23:05

LE RISORSE PIU’ RICHIESTE VIA PROXY APPARTENGONO AL SETTORE GIURIDICO E SONO DI AMBITO NAZIONALE.

(scarsa familiarita’ con le tecnologie di accesso, della utenza o dei provider?)

Università degli Studi Roma TRE

Area Telecomunicazioni
Sistema Bibliotecario di Ateneo

Via Ostiense, 161
00154 Roma

Tel.: +39 06 57334391
Posta elettronica: idem@uniroma3.it

Si ringraziano Monica Boccato e Riccardo Valzorio (CILEA) per il supporto sistemistico ma soprattutto per la pazienza...