Řízení přístupu k elektronickým informačním zdrojům

Petr VANDROVEC
České vysoké učení technické v Praze, Výpočetní a informační centrum
vandrove@vc.cvut.cz

Barbora RAMAJZLOVÁ
České vysoké učení technické v Praze, Výpočetní a informační centrum
ramajzlo@vc.cvut.cz

INFORUM 2005: 11. konference o profesionálních informačních zdrojích
Praha, 24. - 26.5. 2005

Abstrakt: *Nezbytnou součástí zpřístupňování elektronických informačních zdrojů je také technické zabezpečení přístupu k těmto zdrojům. Obvykle je používán tradiční mechanismus, tj. přístup ke zdrojům přes povolené rozsahy IP adres dané instituce. Tento způsob je nejjednodušší, ale nese s sebou řadu nevýhod. S využitím moderních informačních technologií existují možnosti skutečného řízení přístupu k informačním zdrojům (jsou to např. autentizační a autorizační* systémy*), s nimiž mají zkušenost řešitelé projektu MŠMT 1N04067 Zajištění klíčových informačních zdrojů a služeb pro technické obory.*

Úvod

Elektronické informační zdroje patří dnes k základním složkám informačního zabezpečení vědeckovýzkumné činnosti, nezřídka i pedagogické činnosti vysokých škol. Přestože se jedná o finančně nákladné zdroje, akademická a výzkumná komunita je poměrně dobře zajištěna, a to díky grantovým programům MŠMT – LI *Informační zdroje pro výzkum a vývoj* (2000-2003) a 1N *Informační infrastruktura výzkumu* (2004-2008). České vysoké učení technické (ČVUT) bylo v prvním období nositelem projektu LI „*Zajištění základních dokumentografických databází pro technické univerzity",*v současné době zastřešuje projekt 1N „*Zajištění klíčových informačních zdrojů a služeb pro technické obory“*, jehož účastníky jsou vysoké školy: ČVUT, Masarykova univerzita (Fakulta informatiky), Technická univerzita v Liberci, Univerzita Pardubice, Univerzita T. Bati ve Zlíně, Vysoké učení technické v Brně, Vysoká škola báňská -Technická univerzita Ostrava, Vysoká škola chemicko-technologická a Západočeská univerzita. Tímto projektem je zajištěn přístup k dokumentografickým databázím *Compendex, Iconda, Inspec*, které jsou sjednocené rozhraním databázového centra Dialog, a k databázové kolekci producenta CSA *Materials Science with Metadex*. Řešitelé projektu získali řadu zkušeností a považují unifikaci řízení přístupu ke zdrojům jako nezbytnou součást zpřístupňování elektronických informačních zdrojů. Proto se tento problém rozhodli řešit jako dílčí cíl uvedeného projektu. V článku seznamují s postupným vývojem řešení.

Základní principy přístupu

Rozhodnutí o tom, zda má uživatel k nějakému informačnímu zdroji přístup, lze rozdělit na dvě fáze – autentizaci a autorizaci. Autentizační proces zajistí identifikaci přistupujícího uživatele – tj., že se ke službě snaží přistupovat konkrétní uživatel (či počítač), autorizaci jsou pak zajištěny informace vázané k této identitě. Při řízení přístupu podle IP adresy lze říci, že se autentizací zajistí to, že určitý počítač (ze kterého chce uživatel přistupovat ke zdrojům) má nějakou IP adresu. Autorizaci provede poskytovatel porovnáním jeho IP adresy s databází předplatitelů služby. Pro autentizaci a autorizaci přístupu k elektronickým informačním zdrojům se používají dva základní principy. Jednak prostřednictvím IP adresy počítače, jednak pomocí uživatelského jména a hesla. Přístup pomocí jména a hesla (uživatelské konto) ještě můžeme dále rozlišit podle toho, kde autentizace a autorizace probíhá – zda jsou informace soustředěny u poskytovatele informací, nebo zda je část autentizace a autorizace svěřena nějakému zprostředkovateli, či přímo zúčastněné univerzitě či fakultě, ze které jsou uživatelé zdrojů.

Vývoj řešení přístupu – zkušenosti

Řešitelské pracoviště (Výpočetní a informační centrum ČVUT) se problematikou zpřístupňování elektronických informačních zdrojů začalo zabývat již v letech 1998-1999, kdy ČVUT přestalo používat dokumentografické databáze na CD ROM, protože získalo možnost přímého přístupu do databázového centra Dialog. V té době neexistovalo příliš možností, jak se k poskytovateli zdrojů připojit. Pro omezený počet uživatelů poskytovatelé nabízeli vytvoření uživatelského konta, samozřejmě za patřičný poplatek, ale pro dvě desítky tisíc studentů a akademických pracovníků neexistovalo jiné řešení než přístup prostřednictvím IP adres počítačů, které byly v síti ČVUT.

Přístup přes IP adresy

Zprovoznění přístupu bylo jednoduché – producentovi byl poskytnut rozsah IP adres ČVUT, takřka okamžitě byl povolen přístup do předplacených zdrojů. V krátké době se ale objevily problémy, které jsou s autorizací přes IP adresy nevyhnutelně spojeny. Většina uživatelů se ve výsledných statistikách užití objevovala pod adresou různých WWW proxy serverů na kolejích či fakultách. Jak se dalo očekávat při takovém množství uživatelů, brzy nastal incident, při kterém bylo pomocí počítače jednoho z pracovníků školy staženo několik desítek tisíc záznamů. Poskytovatel na problém s nadměrným množstvím přístupů upozornil s více než měsíčním zpožděním. Vypátrat s takovou časovou prodlevou na všech proxy serverech, ze kterého počítače byly záznamy neoprávněně stahovány, bylo již samo o sobě značné dobrodružství. Po dohledání počítače se ukázalo, že pracovník, kterému tento počítač patřil, byl dlouhodobě v zahraničí, takže počítač nevyužíval, ale nikdo z daného pracoviště se nechtěl k prohřešku přiznat.

Poučení z tohoto incidentu bylo jednoznačné: pokud se o incidentu nedozvíme okamžitě, nemáme prakticky žádnou šanci vypátrat pachatele neúmyslného porušení licenčních podmínek, o pachateli úmyslného porušení nemluvě.

Protože již v této době byl zájem statisticky sledovat využívání informačních zdrojů, bylo nevýhodné zpoždění přehledů (statistik), které jsme dostávali od poskytovatele. Statistiky byly sumarizovány po měsících, dostávali jsme je od poskytovatele v papírové formě, a to zpravidla kolem dvacátého v měsíci, takže zpoždění se pohybovalo mezi 20 až 50 dny. Jakékoliv další zpracování údajů o využívání informačních zdrojů bylo podmíněno vlastním převedením z papírové podoby do elektronické.

Speciální proxy server

Bylo tedy rozhodnuto najít jiné řešení realizaci speciálního WWW proxy serveru, přes který budou všechny požadavky směrovány. Výhodou bylo, že v té době bylo používání WWW proxy na ČVUT povinné. Nebylo tedy třeba rekonfigurovat všechny počítače na univerzitě, stačilo jen nakonfigurovat všechny univerzitní proxy servery tak, aby byly požadavky určené poskytovateli směrovány na náš speciální proxy server. Ten pak vytvářel detailní logy s informacemi o přístupu. Díky tomu byl získán přístup k informacím o počtu stahovaných článků prakticky v reálném čase, což umožnilo reagovat na případný incident dříve, než se rozvinul do problému, který v krajním případě mohl vyústit ve zrušení licenční smlouvy.

Stále ještě zbývalo několik problémů, které bylo třeba vyřešit. Jednak to byla nutnost speciálního nastavení WWW proxy serverů na ČVUT, nutnost používání WWW proxy serveru pro přístup k webu, a nemožnost zpřístupnit informační zdroj oprávněným uživatelům, kteří jsou mimo kampus – vzdálený přístup. Nejdůležitější byl stále problém s prokázáním, kdo v inkriminovanou dobu seděl u počítače, ze kterého bylo přistupováno ke zdrojům. Pro vyřešení tohoto problému bylo třeba opustit autorizaci přes IP adresy.

Přístup pomocí jména a hesla (jmenná autorizace)

Bylo třeba najít systém, který by umožnil prokázání identity uživatele, který sedí u počítače, a nikoli počítače samotného. Realizace tohoto kroku byla na ČVUT poměrně složitá, protože v té době se na škole nevyskytoval žádný autorizační zdroj s celouniverzitní působností. Tato skutečnost přiměla řešitele k vytvoření modulárního systému, protože každá univerzitní součást (fakulta, ústav) měla jiný autentizační systém a jiné požadavky na uživatele, kteří by měli mít k informačním zdrojům přístup.

Při realizaci autorizace pomocí jména a hesla vyšla na povrch neschopnost některých WWW prohlížečů (a proxy serverů) předat autentizační informace přes první proxy server některé další proxy, například té naší. Následovala tudíž volba: buď překonfigurovat všechny počítače na univerzitě tak, aby se k našemu proxy serveru přistupovalo přímo, nebo opustit koncepci WWW proxy serveru a poohlédnout se po jiném řešení. Dalším problémem autorizace na úrovni WWW proxy byl fakt, že jméno a heslo je posíláno proxy téměř v otevřeném formátu, což je jen velmi obtížně akceptovatelné i ve velmi uzavřených a kontrolovaných prostředích, natož pak na univerzitě plné potenciálních hackerů.

Požadavek na překonfigurování všech počítačů na ČVUT a zasílání hesla v otevřené podobě se ukázalo jako slepá ulička. Koncepce proxy serveru na úrovni protokolu

HTTP nebyla zjevně vhodná, nezbylo než posunout proxy o úroveň výš – a udělat z ní aplikační proxy.

Aplikační proxy

Je třeba stručně objasnit, co to je aplikační proxy. HTTP proxy pracuje tak, že od klienta převezme požadavek na nějakou cizí webovou stránku, připojí se k této stránce, stáhne její obsah a data která tak získá, předá klientovi. Podstatné je si všimnout, že klient říká, kam se připojit – např. když se realizuje připojení přes WWW proxy server.foo.bar na http://www.google.com, do prohlížeče se píše http://www.google.com – WWW proxy server nijak neovlivňuje data, která Google posílá. V případě aplikační proxy komunikuje uživatel přímo s aplikační proxy, místo s cílovým počítačem. Aplikační proxy transformuje jak data, která jsou zasílána z klienta na server, tak i odpovědi serveru klientovi. Zajímavou ukázkou aplikačního proxy serveru je například http://translate.google.com, který se za uživatele připojí na stránku, která mu byla zadaná a obsah stránky uživateli pošle zpět přeložený do jazyka, který uživatel preferuje. Při srovnání s WWW proxy jsou na první pohled vidět rozdíly – do prohlížeče je třeba napsat URL ve speciálním formátu (například URL http://translate.google.com/translate?u=http://www.google.com&langpair=en|de přeloží titulní stránku Google z angličtiny do němčiny). Obsah stránky, který uživatel dostane, se také podstatně liší od originálu – má navíc hlavičku Google a je v němčině, a ne v angličtině. Pro řešitele i uživatele z toho plyne jeden problém. Pro přístup k informačnímu zdroji nelze jednoduše použít URL ve tvaru http://zdroj.poskytovatel.com, ale je třeba použít nějakou jinou formu. Proto na počátku tedy řešitelé zvolili formát, který byl lehce zapamatovatelný – https://brana.cvut.cz/zdroj.poskytovatel.com. Navíc byl na hlavní stránce brány vytvořen rozcestník, takže uživatel ČVUT měl všechny dostupné (a pro jiné neveřejné) informační zdroje k dispozici z jedné stránky, přímo na konečcích svých prstů.

Díky přechodu z WWW proxy na aplikační proxy byl tedy vyřešen problém spolupráce s ostatními systémy používanými k přístupu na Internet na naší univerzitě. Zůstal tedy „jen“ problém autorizace.

Aplikační proxy také umožnila použití šifrované komunikace mezi klientem a aplikační proxy. Tak se vyřešil problém s možností odposlechnutí hesla na síti některým zdatným studentem. Realizace proxy serveru na aplikační úrovni umožnila mnohem širší možnosti identifikace uživatelů. Zatímco autentizace na úrovni HTTP umožňuje pouze zadání jména a hesla, na úrovni HTML lze vytvořit prakticky neomezeně komplikovaný formulář s prakticky neomezenou logikou, která autentizaci a autorizaci řídí. Díky tomu bylo na jedné straně možné uživatelům nabídnout přehledný přihlašovací formulář, na druhé straně se tímto přihlašovacím formulářem značně zkomplikovala možnost stáhnout všechny stránky poskytovatele nějakým automatickým programem na stahování WWW stránek, protože na začátku činnosti je vyžadována interakce mezi uživatelem a WWW serverem.

Zprovozněním jmenné autorizace byl vyřešen problém vzdáleného přístupu a díky použití aplikační proxy, která nevyžaduje žádnou rekonfiguraci na počítači uživatele, začal být přístup k univerzitním informačním zdrojům možný i z dalších míst, tedy i z těch, která povolují pouze protokoly HTTP a HTTPS.

Současné řešení přístupu

Aplikační proxy má i své nevýhody. Tou největší nevýhodou je, že klient musí používat modifikované URL. To znamená nejenom vložit jiné URL do prohlížeče, ale také všechny odkazy v HTML textu stránek, v odpovědích HTTP a v žádostech musí být modifikovány. A nejenom v HTML textu. Na stránkách poskytovatelů jsou často generovány odkazy i prostřednictvím JavaScriptu, i tento kód musí být automaticky upraven tak, aby generoval odkazy na bránu, nikoli přímé odkazy přímo k poskytovateli. Naštěstí lze všechny tyto problémy řešit, s postupujícím časem a větší standardizací HTML a vznikem XHTML se brána pomalu oprošťuje od různých heuristik. Zatímco na počátku realizace naší brány se u poskytovatelů informací odkaz obrázek téměř vždy vyskytoval ve tvaru <IMG SRC=images.poskytovatel.com/logo.gif>, v současnosti nenarazíte na stránku, která by nepoužívala korektní odkaz ve tvaru <IMG SRC=”http://images.poskytovatel.com/logo.gif”>.

I část systému, která se zabývá autentizací má v současné implementaci určité problémy, které jsou umocněny tím, že naši proxy využívají také jiné univerzity. Autentizace probíhá tak, že uživatelovo jméno a heslo je předáno naší proxy, která se zeptá konkrétního univerzitního autentizačního a autorizačního systému, zda zadané jméno a heslo odpovídá některému uživateli univerzity a zda má dotyčný povolen přístup k tomuto informačnímu zdroji. To znamená, že hesla všech přistupujících uživatelů jsou zadávána naší proxy. Je jasné, že tím vzniká velmi velký tlak na bezpečnostní zajištění této proxy. Současně je třeba překonávat nedůvěru uživatelů jiných univerzit, kterým nemusí být úplně po chuti vkládat své heslo, používané do různých univerzitních systémů, také do cizí webové stránky. Je tady ale jiné nebezpečí. Zatímco uživatele lze poučit o bezpečnosti stránek a ochraně jejich hesel, hackeři žádné vysvětlování nepotřebují. Zneužitím jedné proxy a monitorováním jmen a hesel přihlašujících se uživatelů, mají možnost získat pohodlný přístup k informačním systémům hned několika univerzit. Navíc jejich další útok na systémy univerzit zpočátku nezpůsobí žádná varování, neboť budou vybaveni správným jménem a heslem. Řešitelé zatím tuto negativní zkušenost nemají, což neznamená, že nemůže nastat.

Budoucí řešení

Nejenom z důvodů bezpečnosti vzniká stále větší tlak na co možná největší decentralizaci autentizačního systému. V ideálním případě by měla služba, která vyžaduje identifikaci uživatele, odkázat dotyčného na přihlašovací stránku jeho univerzity. Teprve po přihlášení se do systému, plně spravovaného univerzitou a společného pro všechny autorizované zdroje na univerzitě přístupné, pak pokračovat v přístupu k informačnímu zdroji.

V několika posledních letech se rozvíjejí zejména dva systémy pro distribuovanou autentizaci a autorizaci – ve Velké Británii je to systém *ATHENS* (http://www.athens-ams.net), ve zbytku světa pak *SHIBBOLETH* (http://shibboleth.internet2.edu). Projekt Shibboleth se jeví mnohem otevřenější než projekt britský a v našich podmínkách se pro jeho podporu rozhodlo sdružení CESNET. Vzhledem k tomu jsme se rozhodli umožnit v naší bráně autorizaci prostřednictvím autorizačního systému Shibboleth.

Dalším z důvodů, proč jít touto cestou, je skutečnost, že někteří poskytovatelé informací – například vydavatelství Elsevier – poskytují možnost autorizovat se do

jejich systému právě prostřednictvím infrastruktury Shibboleth. Zatím to pro nás nepředstavuje příliš velkou výhodu – detailnost a aktuálnost statistických informací od poskytovatelů snese jen obtížně srovnání s detailností a aktuálností statistik poskytovaných bránou – ale je třeba přijmout, že navržený systém distribuované autentizace a autorizace je perspektivní a lze jej pro naše účely využít.

Závěr

Řízený přístup k elektronickým informačním zdrojům představuje významnou stránku zpřístupňování elektronických informačních zdrojů. K unifikaci tohoto přístupu v rámci vysokých škol, které se účastní výše zmíněného projektu 1N, dochází postupně. Některé zúčastněné vysoké školy zatím vlastní systém autentizace a autorizace vůbec neprovozují nebo ho teprve plánují. Protože není pochyb o tom, že se tyto systémy stanou obecným řešením, je třeba s jistou dávkou trpělivosti vyčkávat, příp. vyvíjet iniciativu ze strany knihoven, které jsou jedněmi z důležitých aktérů tohoto procesu.

Řešitelé projektu 1N sledují maximální využití řízeného přístupu. Již zmiňovaná přístupová brána (proxy server na ČVUT) se stala tzv. Bránou EIZ (stále umístěnou na adrese https://dialog.cvut.cz), pod kterou jednak byla soustředěna nabídka všech komerčních zdrojů, které jsou uživatelům ČVUT dostupné, jednak byl tento přístup poskytnut i účastníkům jiného projektu 1N, a to 1N 04058 „*Informační zdroje na podporu výzkumu v informatice*“. V prvním případě se jedná o příklad řízeného přístupu k elektronickým informačním zdrojům, který by mohl být aplikován i na dalších vysokých školách. V druhém případě se jedná o racionální řešení, protože jádro účastníků je v obou projektech stejné.

Vedle bezproblémového a bezpečného zpřístupňování je tímto způsobem zajištěno dokonalé sledování využívání těchto zdrojů. Vlastním monitorováním přístupů k jednotlivým zdrojům lze získávat aktuální a podrobné statistické údaje o využití konkrétního zdroje na jedné straně, na straně druhé lze zjistit potřeby konkrétního uživatele. Tyto údaje umožňují sledovat informační profil uživatelů, jejich potřeby na navazující služby knihoven, příp. i na další vzdělávání. Všechny tyto údaje sledují jeden cíl – efektivní využití informačních zdrojů jako významné podpory vědecké a výzkumné činnosti.