

厦门大学：基于身份认证的数据库访问服务

文 / 肖铮¹ 陈娟¹ 郑海山²

随着信息科学技术的发展，学术信息出版模式发生了巨大变化，数字资源在师生学习、科研中发挥着越来越重要的作用。2020年的新冠疫情让数字资源成为高校线上教学科研中唯一可获得的学术资源，做好线上资源远程访问的服务保障，是高校图书馆的职责所在。

基于知识产权和版权保护要求，数据库商在与图书馆签订的授权协议中对电子资源的授权用户和授权使用方式做了明确约定，在授权方式上最常见的是提供基于高校校园网IP地址的认证模式。基于该认证方式，高校图书馆目前普遍采用代理服务器、VPN、WebVPN的方式满足用户在校园网外使用电子资源的需求^[1]。受疫情影响，校外使用数据库需求激增，造成代理服务器、VPN、WebVPN用量增大，导致连接不稳定、并发数受限、网络带宽不足等问题。为解决上述问题，厦门大学图书馆和信息与网络中心协力合作，通过加入CARS（中国教育和科研计算机网统一认证与资源共享基础设施）联盟，基于校园网身份认证系统，向用户提供电子资源的合法访问。CARS联盟的身份认证作为在疫情期间校外用户访问数据库电子资源的应急方案，具有部署简单、服务稳定、安全高效的特点，有力保障了校外用户正常使用数据库电子资源，对将来在国内高校推广基于身份认证的数据库访问打下了坚实的基础。

基于身份认证的数据库访问技术基础与应用现状

1. 联盟认证技术标准 SAML



厦门大学

SAML (Security Assertion Markup Language) 是基于XML的一种安全断言标记语言，提供在各种环境下安全地交换数据和身份识别信息，其目标是在多个应用间实现联邦身份，解决联邦环境中如何识别身份信息共享的标准化问题，是一种独立于协议和平台的验证和授权交换机制^[2]。SAML2.0是OASIS组织安全服务技术委员会的产品，作为一种成熟的用户认证授权规范，被广泛应用于基于Web的统一认证和单点登录系统中，SAML通过令牌的方式进行授权数据交换，为保护用户隐私提供了基础条件。

SAML规范包括：（1）断言与协议，定义XML编码的断言的语法和语义，请求和响应的协议；（2）绑定与配置，定义在不同实体间请求与响应的数据包格式，以及如何使用通用的底层通信协议和在不同系统之间交换断言和请求响应消息的协议；（3）一致性规范，设置一种基

本标准，提高互操作性。

SAML框架包含三个实体，主体（Subject）是拥有身份信息的实体；身份提供者IdP（Identity Provider）是为主体提供身份信息的实体；服务提供者SP（Service Provider）是为主体提供服务的实体。

SAML可以使用不同的验证管理方式，如LDAP、AD、RADIUS、CAS、OAuth等。同时，允许使用多种安全方式，如SSL、PKI、Kerberos等。SAML规范在可操作性、兼容性、安全性、保密性方面都十分完备。

2. Shibboleth 与 OpenAthens

Shibboleth是美国Internet2组织的一个开源项目，基于SAML实现Web上的用户身份认证和资源的安全访问和获取。Shibboleth由身份提供者（IdP）、服务提供者（SP）、身份发现中心（DS）组成。IdP提供用户属性信息管理、身份认证服务、属性信息查询、验证服务等功能；SP提供资源管理、处理或重定向用户请求、

处理属性信息等功能; DS 主要功能是为用户提供 IdP 的重定向服务^[3]。

Athens 最早是由英国 EduServ 公司开发的用于提供单点登录服务的身份和访问管理系统, 后来加入了 SAML 标准推出新的 OpenAthens。

Shibboleth 和 OpenAthens 同样是基于 SAML 的身份认证和访问管理系统, 两者在技术实现上没有太大差异。其主要区别在于, Shibboleth 是开源实现, 由社区进行产品维护和支持, 安装部署需要有关单位自行解决, 对负责实施的联盟组织和图书馆有一定的技术门槛。OpenAthens 属于商业产品, 是一套成熟的基于云端的产品解决方案, 由 EBSCO 公司负责其商业化运作并提供服务支持。对于图书馆来说不需要太多技术投入, EBSCO 公司提供了从安装、测试到上线、运维的一系列完整服务^[4]。

3. 国内外学术数据库对身份认证访问模式的支持情况

英国联合信息系统委员会 JISC 下属的身份认证管理联盟, 致力于为教育和科研提供在线资源获取的单一解决方案, 他们推荐使用遵循 SAML 协议的 Shibboleth 和 OpenAthens, 并且对支持以上两种联邦身份认证的服务商进行了注册登记^[5]。经统计共有 107 个数据库商的 243 种数据库提供基于 SAML 的联邦身份认证访问, 包括著名的数据库商 Thomson Reuters、Elsevier、ProQuest、Wiley、EBSCO、Gale、SAGE 和他们旗下的相关数据库。可以看到, 国外主流的数据库均已支持身份认证访问模式, 并且已经得到了较好应用。

在对国内数据库商的调研后发现, 中国知网于 2019 年 11 月开始对基于联邦身份认证的访问方式进行测试, 本次疫情期间, 应众多高校师生校外访问中国知网的需求, 中国知网公司加紧推出了基于 Shibboleth 认证的测试版, 并向各个接入认证的图书馆免费开通使用 4 个月。2020 年 3 月, 万方数据成功加入 CARSi 联盟, 正式上线提供服务。中国知网和万方数据成为目前国内仅有的两家支持联邦身份认

证访问模式的学术数据库。

4. 中国教育和科研计算机网统一认证与资源共享基础设施 CARSi

CARSi (CERNET Authentication and Resource Sharing Infrastructure) 由北京大学发起, 2008 年获得国家发改委 CNGI08 项目支持; 2011 年与 CALIS 合作, 完成 70 余所高校和 7 家数据库商的 60

余项资源服务; 2019 年 5 月成功加入国际身份联盟组织 eduGAIN; 2019 年 10 月, 完成试验环境到产品环境的过渡, 作为教育网的基础服务推出。CARSi 致力于为高校用户建立身份认证和授权基础设施, 整合高校本地校园网统一用户管理和身份认证系统, 面向中国高校和科研机构提供跨区域身份认证和资源共享服务。CARSi 基于 Shibboleth 中间件建设, 符合当前国际主流国家级教育科研网络规范, 为中国高校更便利地获取全球教育科研资源提供了基础。CARSi 作为高校身份认证和授权联盟组织, 加入 CARSi 联盟的机构通过共享建立一种互信机制, 这样可避免作为 IdP 的高校单独与每一个作为 SP 的数据库商建立单独的属性断言, 简化了高校和数据库商的对接工作^[6]。本次疫情期间, 各高校用户对于校外访问在线资源的需求猛增, CARSi 团队迅速组织投入工作, 简化申请流程、缩短审批时间、提高开通服务速度。

厦门大学服务实践

1. 加入 CARSi 联盟

CARSi 联盟接纳两类会员, 第一类是全资格会员, 必须是 CERNET 会员单位, 必须且只能部署一个联盟身份提供者 IdP, 由网络中心或图书馆提交申请; 第二类是服务提供会员, 需要在全资格会



厦门大学图书馆

员推荐下, 对教育科研感兴趣、致力于为 CARSi 资源共享服务或者为联盟会员单位提供服务的服务提供机构可申请成为服务提供会员, 只部署服务提供者软件 SP^[7]。厦门大学作为 CERNET 会员, 由图书馆和信息与网络中心共同申请成为 CARSi 全资格会员, 并组成 CARSi 项目实施小组。信息与网络中心负责 IdP 本地化部署以及与校园统一身份认证系统对接等技术问题, 图书馆负责与数据库商联络, 申请 Shibboleth 服务开通。2019 年 6 月 25 日, 厦门大学通过了 CARSi 联盟对本地 IdP 的测试, 正式启动服务。

2. 本地 IdP 的部署

CARSi 联盟提供了两种本地 IdP 的部署方案, 第一种是使用 CARSi 联盟提供的虚拟机 ova 镜像进行安装和配置, 镜像采用 CentOS 系统, 包含了已安装好的 Shibboleth IdP 及相关组件, 并提供了一定的默认配置。用户只需要在此基础上根据各校实际修改个别定制化配置, 完成对接本地认证系统、属性释放等即可; 第二种是使用 CARSi 提供的安装包, 在 CentOS 服务器上手动安装 Shibboleth IdP 及相关组件, 根据联盟提供的文档进行本地认证系统对接和属性释放等配置。CARSi 提供了详尽的安装部署文档, 整个安装过程只需按照文档执行即可。其中, IdP 与本地认证系统对接, 释放用户属性是较为重要

的两步,需要根据学校自身情况进行配置。

厦门大学信息与网络中心已建立多种形式的校园统一身份认证系统,为加强安全防护和权限管理,采用了两套密码机制,一套密码用于统一身份认证,另一套密码用于校园 Wi-Fi 和 VPN,所有在校教职工和在校生均可使用 VPN。考虑到认证速度问题,厦门大学采用了 IdP 与 LDAP 对接的方式,需要根据本校 LDAP 的目录结构跟字段定义,修改 IdP 的 conf/ldap.properties 文件的具体配置信息,包括 LDAP 服务器地址、用户名、密码,人员所在的分支,登录用户名与 LDAP 哪个用户属性值对应。由于 CARSi 联盟对用户身份的标准取值为: faculty、student、staff、alum、member、affiliate、employee、other 几类,因此需要将本校 LDAP 中获取的用户身份属性映射到 CARSi 联盟的用户身份属性。厦门大学 LDAP 中的用户身份属性值为数字,1 代表教职工、2 代表本专科生、3 代表研究生。因此在 IdP 的 /conf/attribute-resolver.xml 文件中,通过以下代码完成身份属性值的映射,表示将本地 LDAP 中的身份属性为 1 的映射为 CARSi 标准中的 faculty 用户属性,其他 2、3 的映射为 student 用户属性:

```
if(idsCategory.getValues().get(0)==1)
localpart = "faculty";
else localpart = "student";
```

完成 IdP 与本地 LDAP 身份认证系统对接后,再配置好 IdP 的日志管理、隐私保护配置,即可开始 IdP 测试。

3. 向数据库商申请开通 Shibboleth 认证

图书馆作为学校负责数据库订购的单位,与各数据库商有着紧密的联系。已加入 CARSi 联盟的数据库商,可以通过下载 CARSi 网站上各个数据库商关于开通 Shibboleth 服务的申请说明,由图书馆资源采访部门向数据库商提交申请,一般数据库商可在三个工作日内完成 SP 服务开通。疫情开始后,厦门大学图书馆采访部快速响应,向数据库商发出通告,要求疫情期间,数据库商与图书馆共同配合,解决校

外访问当下存在的问题。对支持 Shibboleth 认证但未加入 CARSi 联盟的国外数据库商发出倡议,呼吁其加入 CARSi 联盟,开通 Shibboleth 认证服务。倡议得到多家数据库商的积极反馈,IOP、JoVE 先后加入了 CARSi 联盟,Ovid 在未加入 CARSi 联盟的情况下与图书馆技术人员积极配合,向厦门大学单独开通了 Shibboleth 认证服务。Taylor & Francis、Wiley、Cambridge University Press、SAGE 等出版社积极与 CARSi 联络商议加入 CARSi 联盟细节,并承诺将在不久开通此项服务。

4. Shibboleth 身份认证访问模式的推广

由于在长期使用数据库电子资源过程中形成的习惯,校外用户对校园 VPN 的依赖程度非常高。厦门大学虽然在 2019 年 6 月已推出 Shibboleth 身份认证,但在推出后的一段时期内,用户寥寥。归其原因,一方面支持 Shibboleth 认证的数据库以外文数据库为主,用户需求量小,VPN 已满足了使用需求;另一方面是 Shibboleth 认证的过程较为复杂,且各个数据库商的认证流程页面不统一,用户接

表 1 厦门大学 CARSi 身份认证人数与 WebVPN 使用人数对比(2020.2.5-2020.2.29)

数据库商	CARSi 认证次数	WebVPN 使用人次
中国知网	45389	316369
科睿唯安	1152	51334
Elsevier	644	37750
Springer	272	3775
IEEE	258	3282
Nature	230	4379
EBSCO	222	12309
ProQuest	124	2382
RSC	79	1495
IOP	31	422
Ovid	26	397
JoVE	24	46
Emerald	18	202

受度低。2020 年疫情爆发后,校外用户使用需求激增,校园 VPN 一时无法承载大量用户,造成连接不成功、使用不稳定的问题。图书馆适时加大宣传力度,通过 QQ/微信群转发说明,解答使用问题。

5. 厦门大学 CARSi 认证使用数据分析

经过 CARSi 管理服务平台的数据统计,自 2020 年 2 月 5 日中国知网数据库 Shibboleth 服务的开通以来,身份认证次数不断提高,截至 2 月 29 日已有 41286 人次认证。

对比同期 CARSi 身份认证人数和 WebVPN 使用人数(见表 1),可以看出习惯使用 WebVPN 方式的用户数量仍然较多,尤其是使用外文数据库资源的用户,使用身份认证方式的偏少,说明图书馆仍需继续加强这方面的宣传推广。

问题与对策

基于身份认证的数据库访问模式相较于基于 IP 认证的访问模式,有诸多优点。比如,避免了传统代理服务器模式的安全性问题,分流了用户对 VPN 的使用,缓解了 VPN 的高并发问题。用户在校外无需通过图书馆导航,从搜索引擎或者浏览器收藏的链接访问数据库时,通过机构身份认证即可使用,操作流程便利,隐私保护完善。本地部署实施简单,只需要一台本地 IdP 服务器,与统一身份认证系统对接,充分利用了校园信息化现有的基础设施,投入小、成本低、见效快。此外,通过加入 CARSi 联盟,使用户通过校园身份可访问更多资源,简化了与不同 SP 的技术沟通成本。CARSi 联盟兼容多个国家地区及全球化身份认证联盟协议,可扩展性强。

经过一段时间的使用,我们发现目前基于身份认证的电子资源访问模式还存在一些问题,建议从以下几个方面加以改进,以便更好地促进基于身份认证的数据库访问模式在国内高校的应用。

1. 提高 IdP 服务器的安全性

本地 IdP 服务器用于用户身份认证和

属性释放,系统安全性至关重要。IdP 系统主要功能由开源软件 Shibboleth 提供,整个系统的安全性高度依赖操作系统、Web 应用服务器和 Shibboleth IdP 的安全性。建议采用操作系统的包管理器安装各个软件而不用自行编译的方式安装,操作系统配置安全更新的自动更新机制,订阅各个组件的安全通知邮件,以便保持 Web 应用服务器和 Shibboleth IdP 组件的更新。对操作系统进行安全加固,比如限制严格的防火墙,只允许管理员登录管理端口,限制 root 远程登录,操作时使用普通用户登录后利用 sudo 命令提升权限执行有关操作。Linux 操作系统应当开启 SELinux,启用较小的系统空闲等待时间。对于 Web 应用服务器,比如 Apache 或 Nginx,应当隐藏版本,关闭不需要的模块和 HTTP Method 等。Web 容器 Tomcat 不建议直接暴露在互联网,而应当隐藏在 Web 应用服务器后面。采用 TLS 加密传输,配置 CA 证书或免费的 Let's Encrypt 证书。系统首次安装完毕应对系统进行上线前安全检查,日常定期使用漏洞扫描工具对服务器进行检查,进行系统渗透测试。

2. 增加 CARS 联盟的 SP 数量

目前已加入 CARS 联盟的 SP 仅有 13 个数据库商的 26 种服务,对比在 JISC 份认证管理联盟中登记的 107 个数据库商的 243 种服务,数量严重不足。如果学校单独与数据库商联系开通 Shibboleth 认证服务,由于国外数据库商在国内的机构主要以业务咨询为主,缺少技术支持,对学校来说增加了额外的沟通成本;并且在技术上还需要针对每个单独的数据库商进行额外的配置,也增加了维护成本。另外,国内数据库商目前只有中国知网和万方数据开通了 Shibboleth 认证服务,其他数据库商还未支持,极大地影响了国内高校应用身份认证访问模式的积极性。同时,用户并不清楚哪些数据库可以使用基于身份认证模式访问、哪些数据库只能使用 IP 认证方式使用,在使用数据库的时候,选择何种方式访问,会对用户造成困扰、影响

用户体验。因此,建议图书馆一方面号召国内数据库商能够及早支持身份认证访问模式;另一方面推动更多国外主流数据库商加入 CARS 联盟,共同促进国内联邦身份认证模式的发展。

3. WAYF Less 解决身份认证流程繁琐的问题

在 Shibboleth 认证模式下,由于各个数据库商的机构发现服务在界面和使用方式上各有不同,对用户而言,在使用过程中找到机构认证入口是一项挑战,这直接影响了用户使用身份认证模式访问数据库的态度,在经历一两次使用困难后,用户宁可使用更为直接的 WebVPN,而不是身份认证。为解决这一问题,JISC 身份认证管理联盟推出 WAYF Less 最佳实践^[8],通过将 IdP 地址组合在 WAYF Less 的链接中,从而绕过机构发现的选择页面,使用户可以直达机构认证页面,降低用户使用身份认证的门槛,避免用户在不同数据库中使用不同机构发现页面的复杂情况。目前大多数主流数据库已经支持 WAYF Less 的最佳实践,图书馆只需在数据库使用页面提供针对本校 IdP 的该数据库的 WAYF Less 链接,用户便可通过该链接,打开机构身份认证登录页面,完成认证后,访问数据库电子资源。

4. 拥抱 RA21 最佳实践方案

WAYF Less 的最佳实践虽然极大地方便了用户使用身份认证模式访问数据库,但依然需要用户通过图书馆的数据库导航,找到 WAYF Less 链接,再进行访问。对于不经过图书馆数据库导航的用户,仍需面对复杂的认证流程。

由国际科学、技术和医学出版商协会(STM)和美国国家信息标准组织(NISO)联合发起的 RA21(Resource Access for 21st Century)项目,旨在满足用户随时随地访问所需学术资源的需求,为用户提供一种简单、无缝、可定制、安全的方式获取学术资源。RA21 提出在已被各机构广泛采用的联邦身份认证系统的基础上,通过实验性项目来探索联邦认证的最佳实践^[9]。目前,该项目已通过 NISO 发布了最佳实

践指导意见,在用户隐私安全保护、联邦认证界面统一、认证流程优化等方面进行了改进,以在学术环境中实现一个与目前互联网使用体验相一致的、现代的、基于标准的资源访问模式,使任何设备在任何时间和地点,都能安全便捷地合法获取电子资源,更好地满足用户对学术资源的获取使用^[10]。根据 RA21 最佳实践,SeamlessAccess 已推出面向数据库商、图书馆、研究机构的云服务,对国内基于身份认证的电子资源访问模式的发展提供了借鉴意义。

新冠疫情对人们的生活、学习、工作产生了重要影响,远程办公、在线教育成为高校师生的日常。随着 5G 时代的来临,用户在任何时间、任何地点、任何网络环境下访问获取电子资源的需求必将日益增长,基于 IP 的数据库访问模式将成为用户便捷获取电子资源的阻碍。CEN(责编:项阳)(作者单位 1 为厦门大学图书馆,2 为厦门大学信息与网络中心)

参考文献

- [1] 李瑞芳,赵美泽,马爱芳.我国高校校外访问图书馆电子资源服务现状的调查与建议——以“211 工程”院校为例[J].情报理论与实践,2007(03):366-368.
- [2] 陆志刚,王杰,魏峻.基于 SAML 的真单点登录框架[J].计算机系统应用,2016,25(02):52-57.
- [3] EGGLESTON H, GINANNI K. Simplifying Licensed Resource Access Through Shibboleth[J]. The Serials Librarian, 2009, 56(1-4): 209-214.
- [4] FERGUSON C L. Authentication Issues and Updates[J]. Serials Review, Routledge, 2020, 0(0): 1-7.
- [5] UK FEDERATION INFORMATION CENTRE. Documents / AvailableServices browse[EB/OL]. [2020-03-04]. <https://www.ukfederation.org.uk/content/Documents/AvailableServices>.
- [6] 北京大学计算中心. CARS 服务 - CARS WIKI[EB/OL]. [2020-03-04]. <https://wiki.carsi.edu.cn/pages/viewpage.action?pageId=327683>.
- [7] CARS. 疫情当前, CARS 助力高校停课不停学[EB/OL]. [2020-03-04]. <https://www.carsi.edu.cn/info/1031/1182.htm>.
- [8] RHYS SMITH, JOHN MURISON. Best Practice: WAYF Less Access to Resources[EB/OL]. <https://www.ukfederation.org.uk/library/uploads/Documents/WAYFlessGuidance.pdf>.
- [9] NATIONAL INFORMATION STANDARDS ORGANIZATION. Recommended Practices for Improved Access to Institutionally-Provided Information Resources[EB/OL]. https://groups.niso.org/apps/group_public/download.php/21376/NISO_RP-27-2019_RA21_Identity_Discovery_and_Persistence-public_comment.pdf.
- [10] 吴至艺,林俊伟,肖铮.RA21:网络学术资源访问解决方案的创新与探索[J].图书馆研究与工作,2020(01):29-34+47.