大数据驱动下的满语抢救式数据平台的安全防范研究*

张颢芳[1]，康春颖[1]，张　伟[2]
（1. 黑龙江大学 数据科学与技术学院，黑龙江 哈尔滨 150000；
2. 黑龙江大学 现代教育技术中心，黑龙江 哈尔滨 150000）

摘　要： 为解决大数据驱动下的满语数据平台的安全问题，利用现行的各种安全防范技术，针对满语平台本身的特征，为其搭建平台安全防范模型。综合分析平台的搭建框架，从物理层、数据采集和汇聚层、数据平台建设层、安全服务分析层以及终端层，为数据平台的安全防护提供全面综合的防范体系框架与技术指导。其中重点论述了安全服务分析层，内容涵盖日志分析、平台管理和渗透测试。平台安全防范技术方面则综合了平台防护、检测和恢复三大类手段。实际应用表明，该模型具有一定的安全防范作用，可有效为平台安全的防范工作提供思路和帮助。

关键词： 满语；数据平台安全；安全防范模型；安全防范技术；安全服务分析层
中图分类号： TP309　　**文献标识码：** A　　**DOI：** 10.19358/j.issn.2096-5133.2020.08.006
引用格式： 张颢芳，康春颖，张伟. 大数据驱动下的满语抢救式数据平台的安全防范研究[J].信息技术与网络安全，2020，39(8)：27-32.

Research on the security of Manchu rescue platform based on big data

Zhang Haofang[1], Kang Chunying[1], Zhang Wei[2]
(1.School of Data Science and Technology, Heilongjiang University, Harbin 150000, China;
2.Modern Educational Technology Center, Heilongjiang University, Harbin 150000, China)

Abstract: In order to solve the security problem of the Manchu data platform based on big data, a security prevention model of Manchu platform is built by using various current security technologies and according to the characteristics of the Manchu platform. This paper comprehensively analyzes the framework of the platform, from the physical layer, data collection and aggregation layer, data platform construction layer, security service analysis layer and terminal layer, builds a comprehensive prevention system framework and technical guidance for the security protection of the data platform. It focuses on the security service analysis layer, including log analysis, platform management and penetration testing. In the aspect of platform security technology, platform protection, detection and recovery are integrated. The practical application results show that it has a certain role of security prevention, and can effectively provide ideas and help for the security of the platform.

Key words: Manchu; the security of data platform; security model; security technology; security service analysis layer

0 引言

满语文曾在历史上起过重要作用。随着历史的发展变化，如今满语已成为濒临消亡的语言，对这种语言的记录、保存和研究具有抢救人类濒危文化资源的重要价值与意义。时至今日，满语口语仅在黑龙江省富裕县三家子村、黑河市大五家子一些满族村屯中还有遗存，其中仅有年龄在70岁以上的十几位老人还可以熟练掌握满语，50岁以上满语会话较好者仅有数十人。满语消亡已进入最后阶段，形势紧迫，亟待持续有效的抢救保护与调查研究[1]。在此背景下，满语保护工作者们为抢救满语搭建了一个大数据平台。大数据平台虽然为信息的处理带来许多便利，也会引发新的安全风险与挑战[2]。本文制定一份安全防范流程，并提出了安全防范模型图，为此类数据平台提供一个综合性的安

* 基金项目：国家社会科学基金重大项目(18ZDA300)

全防范框架。

1 安全防范模型

基于满语数据平台自身的特点以及当前有关数据平台安全防范的研究成果，本文首先论述了平台将会使用的安全防范技术手段，为模型的建立提供技术支撑，然后根据平台前期建设及后期维护的工作，设计出安全防范模型。

1.1 主要技术手段

1.1.1 防护手段

安全防范中最基础的防护手段是建立防火墙[3]。平台科学地使用软硬件防火墙，同时将并联使用两个防火墙以增加攻击难度，并对平台权限授予及来访者身份管理进行系统的分级分类。

由于本平台是在大数据驱动下的满语数据平台，平台的数据信息将呈现出体量庞大、种类繁多以及传播速度快等特点[4]，因此平台数据的加密保护将成为防护重点。

1.1.2 检测手段

平台的检测手段有很多，主要分为三种，即漏洞扫描、状态监控和入侵检测。漏洞扫描从内容上大体分为两种，一种是扫描 Web 应用漏洞，另一种是扫描 Web 服务器漏洞。状态监控意味着在平台运行的过程中，管理员需实时监控平台数据库的运行状态及访问者状态等并进行日志记录，以保障平台的安全稳定。

入侵检测则是一种更为积极的主动安全检测手段，其工作内容分为多种：检测用户的行为并进行分析，对平台的异常活动进行拦截与记录，处理内外部攻击，利用安全防范模型对平台进行风险评估等[5]。实施方法上则划分为异常检测法和混合检测法。

1.1.3 恢复手段

数据备份是一种预防式的数据保护模式，可应对平台使用过程中出现的数据丢失及损坏情况[6]。导致平台数据丢失及损坏的原因多种多样，不仅是因为有攻击者入侵，也有可能是由于平台本身存在漏洞、管理员操作不当等。同时，数据备份还可用于对主数据库和重要文件的监控和跟踪，在备份机中生成对应的更新日志，备份系统会根据更新日志，自动对备份磁盘进行数据更新[7]。通过以上操作，最终将实现数据信息的备份，为平台的恢复提供基础保障。

1.2 安全防范模型图

根据满语数据平台的存储结构、系统功能及业务类别等方面的内容，对其采取层次化安全技术防护，构建满语平台安全防范模型。模型具体划分为以下五个层级，如图 1 所示。

图 1 安全防范模型图

2 平台总体框架

2.1 物理层

物理层是搭建平台总体框架的基础，它可以为连接各个设备并为其之间的数据通信提供传输媒介。物理层的内容涵盖软硬件设备接口、平台服务器、数据存储系统基础设施设备和网络传输设备。在数据传输过程中，利用身份认证、权限管理、入侵检测等方法，保障物理层所提供的传输安全。同时为进一步保障物理层安全，将采用软硬件安全隔离以及底层数据加密技术，使物理层不仅提供数据传输，更能保障平台底层安全。

2.2 数据信息的采集和汇聚层

数据信息的采集内容包含阅读整理各类满语文献资料，并需要调查者走访各满族村落，了解满语现状并采访当地人、收集满语使用者的发音音频等。

其次数据信息的汇聚需要管理员将采集到的数据信息进行分类整理，然后以数据流的形式存入后台数据库中。并对关键数据进行加密处理，赋予不同用户访问数据的权限。同时提前对数据进行备份并实时更新数据信息。

2.3 数据平台的建设层

2.3.1 数据整理

在数据平台建设前，需对前期采集的大量数据进行整理。文字类型的数据，可利用相关软件进行数据整理，数据整理包括正确转录、行或列的重置、净化数据、数据审核等。而对于语音和视频类的数据，管理员则需要根据前台功能模块的设计将其分类打包，方便后期以数据流的形式接入数据库。同时利用网闸、防火墙等设备对流入的数据进行安全隔离，并通过数据管控做到对流入数据的溯源和追踪。

2.3.2 前端设计

平台的前台设计主要是为了实现用户界面交互，界面要求功能清晰明了，布局规整[8]。同时需要对用户及其相应的权限进行分级分类管理，设定分级分类名单，并通过数字签名验证身份。

另一方面，平台前端作为交互的窗口，也会遇到各类数据平台前端攻击，如 XSS 攻击、CSRF 攻击等。平台管理员需要针对此做到严格审查用户上传的信息、关注敏感词汇、不将重要文件放在用户都可以触及的地方以及对平台的重要操作进行跟踪验证等。

2.3.3 后台设计

平台数据库的搭建工作是数据建设层的关键。数据库安全的实现依托多方内容，从数据库的功能上划分，分为两个层次：第一层次是内部环境，主要是指访问控制，即不同用户和管理员的访问权限。在这一过程中，平台需要依靠加密技术、数字签名认证技术等实现访问安全。第二层次是外部环境，在外部环境中平台可能会面临病毒入侵、恶意攻击等问题。此时管理员需要对平台进行预保护，比如数据库备份、设立防火墙等，同时要应用入侵检测技术，及时对平台进行安全检测。根据数据库的实际授权情况，对数据库进行相应的加密处理。具体如图 2 所示。

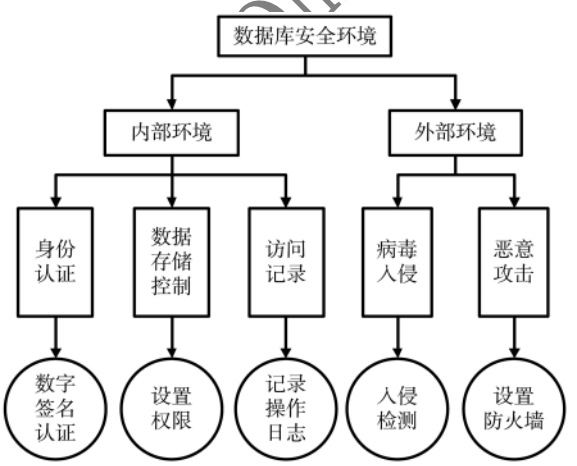

图 2 数据库安全的实现

2.4 安全服务分析层

2.4.1 日志分析

日志是指平台使用者在一定的操作后使系统、软硬件设备产生的相应记录[9]。日志分析主要分为三部分内容：日志数据源、日志分析和日志报告。日志数据的状态则分为：预处理、分析计算、整合和关联。通过收集整理来源于平台各处的日志消息，进行日志分析，从而发现平台中潜在的威胁和安全隐患，及时作出相应的准备和预防工作。经过以上步骤的工作后，最终将生成日志分析报告，具体内容如图 3 所示。

2.4.2 平台管理

数据平台管理主要是指保障本项目所依托的 Web 平台整体的安全，包含用户接口、前台系统、后台系统以及服务器主机等[10-11]。由于数据平台易于攻击，攻击者可选择的方式更是多种多样，如网页

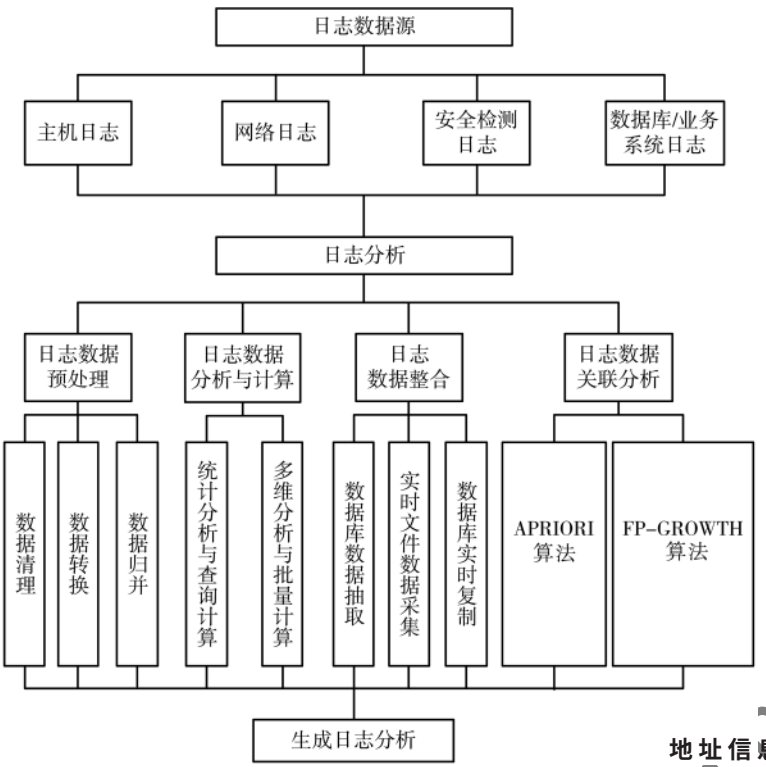

图 3 日志分析工作模块

篡改、网站后门、数据篡改和网络代理等。数据平台管理安全主要包含用户自防御、安全检测、应用防护和防护统计及结果分析几部分的工作。具体如图4所示。

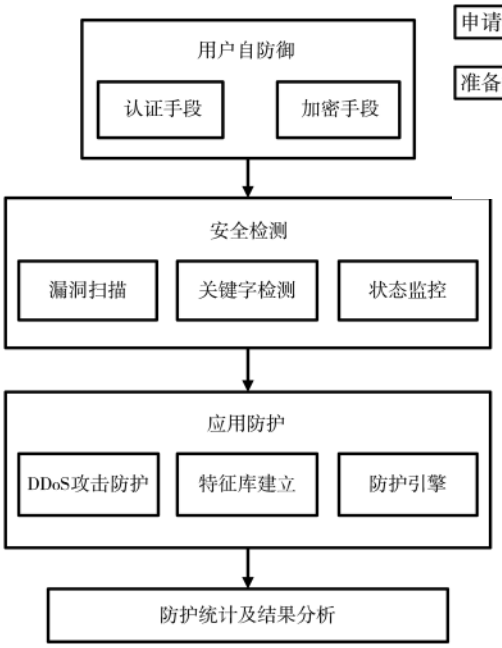

图 4 数据平台安全管理图

2.4.3 渗透测试

渗透测试(Penetration Test)是通过模拟黑客的攻击方法和漏洞发掘技术来评估计算机系统安全的一种评估方法，该方法是从攻击者角度发现分析系统的缺陷及漏洞，进而尝试利用某些漏洞对信息系统实现主动攻击，从而评估系统存在的可能安全风险问题[12]。借鉴文献[13]和[14]，给出渗透测试的操作流程如图5所示。

渗透流程分为前期测试准备、中期测试执行和后期风险评估。测试准备阶段需要确定本次测试的基本情况，如本次测试的目的、方法、参与者和基本操作方案；实验过程中是否有一定的安全措施，能否应对突发情况等。

测试执行需要首先采集平台基础信息，如平台的域名信息、whois信息及IP地址信息；平台的设计框架及工作流程；平台的敏感信息及目录等[15]。然后正式进行渗透攻击。现行的渗透方法有很多，如人工检查、SQL注入、XSS跨脚本攻击、CSRF跨站请求攻击等。再将实验数据进行详细记录和比对，以便于后期解析平台漏洞。

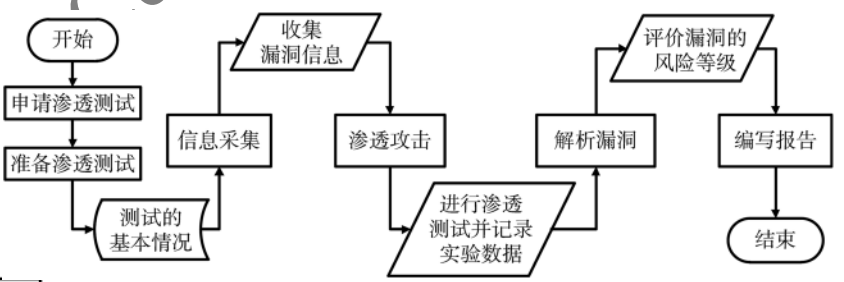

图 5 渗透测试流程图

风险等级评估需要全组实验人员对获取的数据进行综合和分析，进行漏洞解析，并给出大数据驱动下的满语数据平台的风险级别，风险级别一般划分为“低”、“中”、“高”三个等级。测试者最终生成一份渗透报告，达到对平台安全进行全面分析的目的。

2.5 终端层

终端层是指以有线方式或以移动方式连接的终端设备及相关传输网络。用户可在智能手机、台式电脑、平板电脑以及相关移动互联网设备上访问

本平台。传输网络则是指与安全服务分析层的连接，保障网络的正常传输。

终端层的功能是实现人机交互，因而对终端层的管理需要根据来访者的身份进行权限控制，授予用户或管理员不同的读写能力。同时应用数字签名认证相关技术验证来访者身份，保障平台整体的稳定和安全。

3 实验结果与分析

安全防范模型投入满语数据平台应用后，依照模型中安全服务分析层的内容，对满语平台进行安全测试与评估。实验流程分为前期访谈、中期检查和后期测试，即对平台管理者进行工作访谈，了解平台搭建的具体情况；检查平台的相关安全设施和指标；应用现行技术进行安全测试等。实验的名称、内容、结果及改进意见如表 1 所示。

上述实验结束后，针对实验过程中发现的问题及改进意见，生成实验报告。同时根据实验情况，给出本满语平台安全为“中”等级，平台存在部分安全隐患，但总体安全情况良好，依据意见进行改进，有望达到安全“高”等级。应用安全防范模型，对平台进行安全检测，过程条理清晰、内容全面、结果科学有依，表明其可为数据平台的安全检测提供有效方案。

4 结论

本文以满语数据平台为例，针对满语保护的迫切性和相关数据资源的珍贵性，设计出安全防护模型以保障其平台的安全。防护模型结合了当下多方面的防护理念及方法，其中详细论述了安全服务分析层的内容。然而数据平台的安全防护工作不是一蹴而就的，需要平台管理人员和技术人员对平台进行定期检查、坚持记录整理日志信息、不断设计新的防御方案等。未来的工作主要是尝试将本文提出的安全防护模型应用于各类数据平台的保护工作中，在实践中不断调整模型方案，为大数据驱动下的数据平台安全防护研究提供坚实依据。

参考文献

[1] 赵阿平.满语、赫哲语、锡伯语及文化现状调研报告[J].西北民族研究，2018(3)：201－209.

表 1 满语平台安全测试

实验名称		测试内容	测试结果	改进意见
日志分析(日志数据预处理、日志数据分析与计算、日志数据整合、日志数据关联分析)		平台日志是否记录完备；日志消息的内容是否全面；对比以往日志内容，发现是否存在异常日志	日志记录内容全面、日期完整。经过比对，未发现异常日志	日志消息整理分类不够清晰，同时需要对日志消息进行定期更新删除，否则影响平台正常运行
平台管理	用户自防御	身份与登录；认证与授权(遍历各类身份，特别是默认权限与账户)	身份认证成功，权限分配合理	/
	安全检测	漏洞扫描(网络扫描、主机扫描、数据库扫描)	可选：自身平台有可修复性漏洞	可适量安装漏洞补丁
		关键字检测(设定关键字：数据库的插入、删除、更新操作)	关键字工作位置及内容正常	/
		状态监控(比对参数，监控平台状态)	平台运行状态正常并记录下相关参数	/
	应用防护	DDoS 攻击防护、特征库建立、防护引擎	DDoS 防御系统工作良好	/
渗透测试	SQL 注入测试	应用 Sqlmap 工具，对平台进行注入测试	平台存在注入点	对代码进行修改更新，消除注入隐患
	XSS 跨脚本攻击	应用 Beef－xss 工具，对平台发起重定向跨脚本攻击	在平台插入脚本成功并获取了用户信息	对输入和 URL 参数进行过滤，实施白名单和黑名单
	CSRF 跨站请求攻击	应用 CRSF－Tester 抓取浏览器中目标网络，伪造客户端请求	网站服务器接收成功，存在 CSRF 漏洞	阻止不明外域的访问，进行双重 Cookie 验证

[2] 甄海涛,王金玉,杨卓林.基于 Hadoop 大数据平台的数据安全研究[J].自动化技术与应用,2019,38(8):159-161.
[3] 谢林江,杭菲璐.大数据背景下数据治理的网络安全策略[J].科技资讯,2018,16(17):5-6.
[4] 陈保.大数据背景下计算机信息安全处理技术探究[J].南方农机,2019(3):169.
[5] 李妩可,黎元,颜宁.大数据技术在网络入侵检测的应用[J].科学技术创新,2018(26):79-80.
[6] 孟祥富.大数据技术在计算机信息系统中的应用研究[J].办公室业务,2017(24):190,192.
[7] 王小君.网络信息安全防范与 Web 数据挖掘系统的设计与研究[J].电子设计工程,2018,26(12):83-87.
[8] 赵卫东.基于移动设备的实训互动管理平台前端设计与实现[J].重庆科技学院学报(自然科学版),2019,21(3):100-103.
[9] 李刚,陈怡潇,黄沛烁,等.基于日志分析的信息通信网络安全预警研究[J].电力信息与通信技术,2018,16(12):1-8.
[10] 贾迪,黄河滔.对 Web 安全性测试技术的分析[J].信息安全与技术,2014,5(5):68-69.
[11] 吴翰清.白帽子讲 Web 安全[M].北京:电子工业出版社,2014.
[12] 吴兰.Web 应用系统的渗透测试研究[J].电脑编程技巧与维护,2013(6):111-112.
[13] 曲卓,周翰逊,耿天华.信息安全渗透测试流程的研究[J].北方交通,2015(12):112-114.
[14] 徐绍飞,龚家瑜,杨亚萍.医疗行业 Web 应用程序渗透测试实例研究[J].计算机与数字工程,2018,46(1):122-125.
[15] 郎智哲,封筱宇,董齐芬.浅谈 Web 渗透测试的信息收集[J].电子技术与软件工程,2017(8):13-16.

(收稿日期:2020-04-17)

作者简介:

张颢芳(1997-),女,硕士,主要研究方向:网络攻防。

康春颖(1976-),通信作者,女,硕士,教授,主要研究方向:网络攻防。E-mail:kangchunying@hlju.edu.cn。

张伟(1976-),男,硕士,高级工程师,主要研究方向:信息处理。

(上接第 26 页)

设计与实现[C].2009 年中国高校通信类院系学术研讨会论文集,2009.
[9] 张秀岭,万旻,骆建彬,等.Linux 下基于 Squid 的多能代理系统与透明网关解决方案[J].微计算机应用,2004(5):534-539,561.
[10] 杨晓丹.蜜网技术在军队网络安全中的应用研究[J].信息与电脑(理论版),2013(9):127-128.
[11] 谭琼玲.浅谈蜜罐系统的实现技术[J].中国科技信息,2010(7):80-82.
[12] 朱一帅,吴礼发.基于 Sebek 的蜜罐识别机制研究[J].信息技术,2009,33(1):83-86.
[13] BALAS E,TRAVIS G,VIECCO C.A dynamic filtering technique for Sebek system monitoring[C].IEEE Information Assurance Workshop.IEEE,2006.
[14] CHAMOTRA S,SEHGAL R K,ROR S,et al.Honeypot deployment in broadband networks[C].International Conference on Information Systems Security,2016:479-488.

(收稿日期:2020-04-07)

作者简介:

李政达(1994-),男,硕士研究生,主要研究方向:网络安全、工业互联网安全。

周成胜(1982-),男,硕士,工程师,主要研究方向:工业自动化、行业信息化、工业互联网安全。

版权声明