江苏省政务大数据安全管理研究与分析

唐树备　夏　煜　鲁雅楠　韦卯宁　王岩岩
河海大学机电工程学院　江苏常州　213000

摘　要：政务大数据是政府和事业单位从事政务活动或例行社会管理功能、事务处理等一系列活动产生的可存储的数据。伴随着大数据的飞速发展，新的技术架构、支撑平台和大数据软件不断涌现，政务大数据安全技术发展也面临着新的挑战。传统的安全检测技术能够将大量的日志数据集中到一起，进行整体性的安全分析，试图从中发现安全事件。随着大数据系统建设，日志数据规模增大，数据的种类将更加丰富。过多的误判会造成安全检测系统失效，降低安全检测能力。因此，在大数据环境下，大数据安全审计检测方面也面临着巨大的挑战。随着大数据技术的应用，为了保证大数据安全，需要进一步提高安全检测技术能力，提升安全。

关键词：政务大数据；安全监测技术；安全监测能力；日志数据

Research and analysis of government affairs big data security management in Jiangsu Province

Shubei Tang, Yu Xia, Yanan Lu, Maoning Wei, Yanyan Wang
School of mechanical and electrical engineering, Hohai University, Changzhou, Jiangsu 213000

Abstract: Government big data is the stored data produced by a series of activities engaged in government activities or routine social management functions and transaction processing.With the rapid development of big data, new technology architectures, support platforms and big data software are constantly emerging, and the development of government affairs and big data security technology is also facing new challenges.Traditional security detection technology can gather a large number of log data together to carry out overall security analysis and try to find security events.With the construction of big data system, the scale of log data increases, and the types of data will be richer.Too much misjudgment will cause the failure of the safety detection system and reduce the safety detection capability.Therefore, in the environment of big data, big data security audit and testing is also facing great challenges.With the application of big data technology, in order to ensure the security of big data, it is necessary to further improve the security detection technology capability and improve the security.

Keywords: government big data, security monitoring technology, security monitoring ability, log data

1　政务大数据应用背景

近10年间我国电子政务工程取得了显著进展，但在电子政务工程建设中一直存在着“纵强横弱”、“信息孤岛”等问题，制约了部门间的政务协同工作，影响了“数字政府”的建设。与此同时，党中央、国务院高度重视大数据在经济社会发展中的作用，党的十八届五中全会提出“实施国家大数据战略”，党的十九大明确提出要加快推进信息化，建设“数字中国”、“智慧社会”，党的十九届四中全会提出“推进数字政府建设，加强数据有序共享”。国务院要求推进政务服务“一网通办”和企业群众办事“只进一扇门”“最多跑一次”，加快推进“互联网+政务服务”、政务信息系统整合共享、审批服务便民化和建设一体化在线政务服务平台等工作。

各地政府纷纷积极响应，在贵州省依托国家大数据（贵州）综合试验区，形成了集约统一的“云上贵州”数据共享交换平台；在广东组建省政务服务数据管理局，统筹推动“数字政府”建设；在浙江省成立“浙江省大

个人简介：唐树备（1995.2—），男、汉族、江苏省南京市、河海大学机电工程学院机械专业研究生在读，研究方向为智能制造、智能控制，参与过“华为数据安管平台建设”、“英才名匠”等多项大数据安全管理体系研究项目。对大数据安全管理、数据安全与技术、数据应用推进等有一定研究。

数据发展管理局”，统筹管理公共数据资源和电子政务，推进政务大数据综合应用。江苏省相继出台《江苏省大数据发展行动计划》、《智慧江苏建设三年行动计划》等一系列文件，筹划成立了江苏省大数据管理中心，积极推进政务数据的汇聚和应用工作，围绕加强政务、行业、互联网数据共享利用进行了深入研究，目前大数据应用场景已经初见规模。而在业务数据集中的同时，政务大数据安全成为了我们必须要关注的目标。

2 江苏省政务大数据平台的概况、安全现状和管理目标

2.1 江苏省政务大数据平台概况

政务外网大数据平台由大数据交互平台与大数据资源中心组成。大数据交互平台是集目录编制、数据归集、供需对接、共享管理、数据集成、数据治理、运行监控、统计评估、共享开放为一体的大数据共享交换平台，推动省级各部门政务数据资源向大数据共享交换平台迁移集聚，发挥大数据共享交换平台支撑多部门数据和业务协同作业，实现数据资源和相关业务的统一管控。大数据资源中心数据库包括：中心资源库、综合人口信息资源库、综合法人信息资源库、电子证照信息资源库、社会信用信息资源库。

2.2 我省政务大数据平台安全现状

通过对江苏省大数据管理中心政务外网大数据平台的调研和分析了解到，通过采取技术和管理手段对其面临的主要安全威胁采取了相应的安全机制，基本达到保护重要数据资产的作用。

在业务和数据安全方面进行了身份鉴别方面的安全配置，通过用户名和口令的方式进行身份鉴别，基本符合鉴别要求。通过运维运营平台对用户身份进行了划分，实现不同职责用户权限最小化。通过审计系统实现操作日志、系统日志、登录日志的记录汇总，为安全事件的追溯提供数据支撑。在入侵和安全威胁方面由安全服务商定期开展扫描工作和加固工作，基本实现了系统层面的安全可靠。

但是在防止业务系统用户身份冒用方面，数据安全方面的数据分类、分级管控、敏感信息脱敏、数据传输和存储过程的加密等方面还存在明显的问题，需要进行深入的分析和规划设计来解决上述这些问题。

2.3 我省政务大数据安全与开放的管理目标

国家之所以大力开展政务大数据建设，其主要意义就在于通过技术融合、业务融合、数据融合，实现跨层级、跨地域、跨系统、跨部门、跨业务的协同管理和服务。推进政府决策科学化、社会治理精准化、公共服务高效化，利用技术感知社会态势、畅通沟通渠道、辅助决策施政，提高政务数据的科学性、准确性、时效性，提升运用数据服务经济社会发展的能力。政务大数据的安全与开放就是开展政务大数据管理的终极目标。

3 全省政务大数据安全管理体系研究

3.1 大数据安全遇到的挑战

伴随着大数据的飞速发展，各种大数据技术层出不穷，新的技术架构、支撑平台和大数据软件不断涌现，大数据安全技术和平台发展也面临着新的挑战。

大数据技术架构复杂，大数据应用一般采用底层复杂、开放的分布式计算和存储架构为其提供海量数据分布式存储和高效计算服务，这些新的技术和架构使得大数据应用的系统边界变得模糊，传统基于边界的安全保护措施将变得不再有效。如在大数据系统中，数据一般都是分布式存储的，数据可能动态分散在很多个不同的存储设备、甚至不同的物理地点存储，这样导致难以准确划定传统意义上的每个数据集的“边界”，传统的基于网关模式的防护手段也就失去了安全防护效果。同时，大数据系统表现为系统的系统，其分布式计算安全问题也将显得更加突出。

3.2 数据收集安全管理的研究

政务大数据是指政府在推动大数据应用发展的过程中或大数据在公共服务领域的应用实践中产生的大数据。政务大数据是建设新型智慧城市的基础，具体应用场景包括：为政府提供智能办公、智能监管、智能服务、智能决策等大数据服务；帮助政府更好的治理城市，提高政府办公、监管、服务、决策的智能化水平等。

国家实施大数据战略，推进数据基础设施建设，鼓励和支持数据在各行业、各领域的创新应用，促进数字经济发展。各省级人民政府制定数字经济发展规划，并纳入本级国民经济和社会发展规划。为形成政务大数据，需要将各级政府部门、各单位管辖的数据资源汇集起来，实现政务数据的互联互通，并对大量的多源异构数据融合进行大数据综合分析、挖掘，从而帮助政府将现有的数据资源进行转化并创造出价值，有效提升政府管理和决策能力。

3.3 数据存储安全管理的研究

Hadoop逐步完善认证授权等方面的安全机制，尤其是对企业级用户提供了很多内部的安全解决方案，包括在Hadoop生态框架中推出的HUE、Zeus等组件来提供数据权限管理的功能。在系统安全建设的初期以最坏的打算来设计，假使系统被外部成功入侵或攻击发生于内部，存储于HDFS中的明文数据则完全暴露在攻击者面前。因此，对于安全问题的核心保护手段仍然是数据加密。在不影响大数据处理能力的情况下，对重要数据进行不同等级的加密，可以保护数据的核心价值。目前，Hadoop已提供了实现网络传输的数据加密机制，但对HDFS中存储数据的加密还要进一步设计，包括密钥产生方法、密钥持有者所属节点、系统拓扑结构等。

而在众多加密技术之中，透明加密是更加便捷和容

易运用的加密技术之一。透明加密方案的核心思想是加密和解密过程对客户端都是透明的，即客户端不用对程序代码进行任何修改，数据加密和解密操作都由客户端完成，HDFS也不会存储未加密的数据或未加密的数据加密密钥。对于透明加密，向HDFS引入了一个新的抽象：加密区（Encryption Zone，EZ）它是一个特殊目录，其内容将在写入时透明加密，并在读取时透明解密。每个加密区与创建它时指定的单个密钥相关联。加密区内的每个文件都有自己的唯一数据加密密钥（Data Encryption Key，DEK），DEK从不直接由HDFS处理，相反，HDFS只处理加密的数据加密密钥（Encrypted DEK，EDEK）。客户端向KMS发出请求解密EDEK，然后使用后续的DEK读取和写入数据。在HDFS数据节点只能看到加密字节流。如此便实现了大数据系统的加密处理，保证数据在此环境中的安全性。

加密技术是保证大数据安全的核心技术之一，其可以在保证数据机密性的同时，可以实现数据的完整性、可靠性，为大数据系统提供可靠保障。

3.4 数据使用安全管理的研究

数据使用是大数据平台化和全省政务一体化的核心任务，而数据共享使用更是信息泄露的主要途径，信息泄露往往是由内部工作人员、第三方合作伙伴或黑客或非法组织的失误或恶意行为造成的。而信息系统无法加固的安全漏洞，对数据的重要程度不敏感未进行数据脱敏，对安全配置的疏忽大意以及机构本身的防护机制不健全等问题是造成敏感数据会被轻易获取的主要原因。

为了让数据能够被运用和共享交互，必须先对敏感数据进行识别并进行脱敏处理，在保证数据有效性的基础上再交付给外部，防止敏感信息通过特权账号、业务交互以及测试、研究和培训等访问方式外泄。

拥有特权账号的用户是最易将敏感信息外泄的人员，所以对具有特权账号的用户来说，需要将其看到的数据进行动态实时的脱敏，防止了用户对敏感信息的泄露。

4 全省政务大数据安全管理任重道远

4.1 完善大数据运用管理体系

大数据运用应遵循“不危害国家安全利益、不危害企业商业利益、不危害个人信息”三个原则。基于此，在我国现有法律制度的基础上，进一步完善大数据运用的规范管理体系。

一是在《网络安全法》、《数据安全法》确立的网络安全管理框架下，尽快研究制定个人信息及重要数据保护措施，以保障大数据运用中数据主体的合法权利。

二是建立大数据运用的流程体系规范。大数据中心与各组织机构和技术服务机构，共同研究大数据运用的安全管理规范，细化各类数据进行使用过程的策略与规范，明确大数据运用全生命周期的具体措施，加强事前风险研判与监测。

4.2 加强大数据运用监管

一是保障国家安全，严格监管重点领域的大数据运用。对通信、能源、交通、水利、金融、公共服务及电子政务等关键信息基础设施和重点领域的大数据运用提出严格的监管要求，控制总量数据和核心数据流出。

二是制定数据分类分级监管体系，对大数据运用实施分级分类管理。借鉴国际经验，针对人社、金融等涉及关键基础设施的重点领域进行大数据运用进行限制，设立大数据运用监管机构，对整个过程进行实时的风险评估和梯度管理，为大数据运用提供必要的指导。

三是加强大数据安全管理方面的研究，以应对可能面临的安全问题。可联合企业、高校、科研机构等组建专业的研究团队，对如何监管的问题进行全面深入的研究。

加强大数据环境下的网络安全问题研究和基于大数据的网络安全技术研究，建立健全大数据安全保障体系，切实保障数据安全，才能确保大数据“敢共享开放”和“会共享开放”，才能真正促进社会发展。

参考文献：

[1]卫凤林，董建，张群.《工业大数据白皮书（2017版）》解读[J].信息技术与标准化，2017（4）：13-17.

[2]工业互联网产业联盟工业大数据特设组.工业大数据技术与应用实践（2017）[M].北京：电子工业出版社，2017.

[3]王喜文.大数据驱动制造业迈向智能化[J].物联网技术，2019（12）：7-8.

[4]李杰.工业大数据—工业4.0时代的工业转型与价值创造[M].北京：机械工业出版社，2018：57

[5]崔杰.李陶深.兰红星.基于Hadoop的海量数据存储平台设计与开发[J].计算机研究与发展，2017（49）：12-18.

[6] HOSSAIN E, KHAN I, UN-NOOR F, et al. Application of big data and machine learning in smart grid, and associated security concerns: a review[J]. IEEE Access, 2019, 7: 13960- 13988.

[7] HE X, AI Q, QIU R C, et al. A big data architecture design for smart grids based on random matrix theory[J]. IEEE Transactions on Smart Grid, 2017, 8(2): 674-686.

[8] YANG Y, SONG L L, CAO H O, et al. Smart substation secondary system visualization and intelligent diagnosis based on improved SCD model[C] // International Conference on Renewable Power Generation (RPG),2019: 1-5.

[9] Industrial Big Data. Know the future-automate processes. Software for data analysis and accurate forecasting[EB/OL].(2018- 10-23)[2018-3-5].

[10] 李杰.工业大数据—工业4.0时代的工业转型与价值创造[M].北京：机械工业出版社，2017：57.