



大数据环境下电信运营商数据安全保护方案

周胜利^{1,2}, 陈斌³, 吴礼发¹

(1.中国人民解放军理工大学, 江苏 南京 210095; 2.浙江警察学院, 浙江 杭州 310053;
3.华信咨询设计研究院有限公司, 浙江 杭州 310014)

摘要: 随着大数据、云计算技术的应用普及, 电信运营商用户数据向海量化和多样化方向发展, 面临着安全管理、黑客攻击、风险评估等方面的风险。做好数据安全工作对保护电信运营商用户隐私、提升企业竞争力尤为重要。从数据采集传输安全保密、数据分级、数据脱敏、数据加密、数据发布安全防护等方面, 提出了行之有效的电信运营商数据安全保护实施方案, 以保障平台数据的保密性、完整性、可用性、可控性, 并且可在中国电信数据安全保护工程中实施。

关键词: 电信运营商; 数据安全; 隐私保护; 数据加密

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017136

Data security protection schema of telecom operators under the environment of big data

ZHOU Shengli^{1,2}, CHEN Bin³, WU Lifa¹

1.The Chinese People's Liberation Army (PLA) University of Science and Technology, Nanjing 210095, China

2.Zhejiang Police College, Hangzhou 310053, China

3.Huaxin Design Consulting Institute, Hangzhou 310014, China

Abstract: With the popularity of big data and cloud computing technology, data of telecom operators collected is increasingly larger, and much more various. The data faces the risk of broken security management, hacker attacks, and imperfection of risk assessment, etc. The current situation of data security in the environment was analyzed, and a more complete data security implementation plan was put forward. From the two aspects of data security technology and management, several strategies were proposed, including security classification, privacy protection, data desensitization encryption and security protection. The proposed schema has been implemented in China Telecom data security protection project.

Key words: telecom operator, data security, privacy protection, data encryption

1 引言

随着信息化的飞速发展, 社交网络的兴起及云存储、物联网等新兴信息技术的广泛应用, 全球数据飞速增长。为应对海量多样化数据对信息

的有效存储、快速读取、检索提出的挑战, 大数据技术成为电信运营商关注的热点。新技术往往带来新的安全挑战, 电信运营商数据由于集中化等特点, 面临着新的安全风险。如何在大数据环境下确保电信运营商数据共享的安全性, 如何为

收稿日期: 2016-03-22; 修回日期: 2017-05-10



用户提供更为精细的数据共享安全控制策略等问题值得深入研究。

根据电信运营商大数据存储应用实际情况,从数据采集传输、数据安全分级、数据脱敏、数据加密、安全基线配置、数据发布安全保护设计了数据安全保护的关键技术方案,对电信运营商数据安全隐私保护具有非常重要的借鉴意义。

2 研究现状

在电信运营商数据安全保护方面,现有的研究主要集中在基于密码学的数据安全保护研究方面,理论性强,缺乏整体实施架构及安全闭环设计。参考文献[1]从应用的角度介绍了中国联通的大数据平台以及大数据生态系统。参考文献[2]介绍了一个基于 Hadoop MapReduce、公钥加密和 SSL (secure sockets layer, 安全套接层) 协议的安全模型,缺乏模型效率可行性论证。参考文献[3]对数据生命周期的传输阶段的云计算相关的数据安全和隐私保护问题提出了一些解决方案。参考文献[4]结合公钥的同态认证和云审计系统,实现云计算隐私保护。参考文献[5]提出了一种混合云存储架构,以保证混合云中的数据安全。参考文献[6]提出了一种基于素数的动态密码的安全验证方案 (DPBSV), 实现数据共享安全。

3 电信运营商数据安全保护的必要性

3.1 面临黑客攻击的安全挑战

电信运营商存储着海量的用户数据,在用户数据的使用和共享过程中容易成为黑客攻击的目标。用户数据中包含着大量的个人隐私,提高了其潜在的价值,而数据的大量汇集使得一次成功的黑客攻击可以获得更多数据。并且因为电信运营商数据具有低密度特性,很难分析出潜在风险,持续攻击很可能隐藏在海量数据中,从而进一步加大了用户信息泄露的风险和安全分析的难度,导致安全监测偏离应有方向,事故处理难度大大

提升。

3.2 面临管理的挑战

电信运营商数据安全保护面临的管理方面的挑战如下。

- 数据保护管理主体多,缺乏管理体系。我国没有独立的数据保护机关,数据保护管理职责散落在工业和信息化部、商务部、公安部以及各个行业主管部门,数据保护行政管理体系尚未建立。
- 行业自律组织较为欠缺。我国目前尚无受到行业内广泛认可的针对数据保护的行业自律组织或第三方认证机构。
- 数据生态环境中各个角色的权利、义务尚未明确,数据开放和共享过程中的产权保护、权限管理和隐私保护等问题需要解决。
- 尚未建立可管理、维护、可信、易于互操作的数据资源集。

3.3 缺乏安全标准与有效安全评估手段

我国数据安全标准起步较晚,目前逐步受到重视,已有专门针对数据保护的标准,首个关于个人信息保护的国家标准《信息安全技术公共及商用服务信息系统个人信息保护指南》也已于 2013 年 2 月 1 日起实施。2014 年发布电信和互联网行业用户信息保护标准,如《电信和互联网用户个人电子信息保护通用技术要求和管理工作要求》《电信和互联网用户个人电子信息保护检测要求》。

我国至今尚未建立针对数据安全的安全评估体系,针对数据安全及用户信息保护的安全评估尚为空白。一般由政府委托检测机构对安全事故单位开展检测,未形成常态化的数据安全评估体系。针对各个专业网络和系统及新兴技术业务的数据安全要求,分散在各个通信网络安全防护标准及相关标准中,数据安全评估开展落地困难。现有数据安全评估仅针对某个业务系统开展,尚未形成完整的数据安全评估通用体系。

4 电信运营商数据安全实施方案

数据的生命周期^[7]一般可以分为生成、变换、传输、存储、使用、归档、销毁 7 个阶段。根据数据自身特点及应用需求,如图 1 所示,结合上述阶段进行合并与精简,可以得出数据应用过程应划分为采集、存储、挖掘与应用、发布 4 个关键环节。

- 数据采集阶段: 确保数据安全传输并对数据进行分类分级管理。
- 数据存储阶段: 对敏感数据进行特殊保护与脱敏处理,并对数据使用人员进行细粒度的授权管理与访问控制。
- 数据挖掘与应用阶段: 对数据的使用行为进行审计与溯源。
- 数据发布阶段: 对发放数据内容进行安全审计。

4.1 数据采集传输安全

数据传输加密可以选择在链路层、网络层、传输层等层面实现,采用网络传输加密技术保证网络传输数据信息的机密性、完整性、可用性^[8]。一般来说,数据传输的安全要求有如下几点。

- 机密性: 只有预期的目的段才能获得数据。
- 完整性: 数据在传输过程中未受到未经授

权的篡改,即接收到的数据与发送的数据完全相同。

- 真实性: 数据来源真实可靠。
- 防止重放攻击: 每个数据分组必须是单一的,保证攻击者捕捉的数据分组不能重发或者重用。

要达到上述安全要求,一般采用的技术手段如下。

- 目的端认证源端的身份,确保数据的真实性。
- 数据加密以满足数据机密性要求。
- 密文数据后附加 MAC (消息认证码),以达到数据完整性保护的目的。
- 数据分组中加入时间戳或不可重复的标识来保证数据抵抗重放攻击的能力。

VPN (虚拟专用网) 技术将隧道技术、协议封装技术、密码技术和配置管理技术结合在一起,采用安全通道技术在源端和目的端建立安全的数据通道,通过将待传输的原始数据进行加密和协议封装处理后再嵌套装入另一种协议的数据报文中,像普通数据报文一样在网络中进行传输^[9]。只有源端和目的端的用户对通道中的嵌套信息能够进行解释和处理。因此,采用 VPN 技术可以通过在数据节点以及管理节点之间布设 VPN 的方式,满足安全传输的要求。

目前较成熟的 VPN 实用技术均有相应的协议

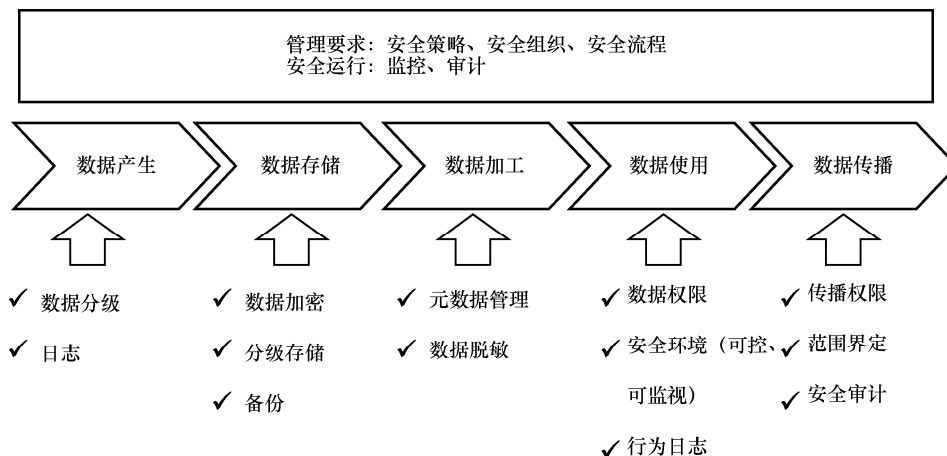


图1 数据安全保障需求



规范和配置管理方法。这些常用配置方法和协议主要包括路由过滤技术、通用路由封装（generic routing encapsulation, GRE）协议、第二层转发协议（level 2 forwarding protocol, L2F）、IP 安全（IP security, IPSec）协议、SSL 协议等，如图 2 所示。其中，SSL VPN 凭借其简单、灵活、安全的特点，得到了迅速发展，尤其在大数据环境下的远程接入访问应用方面，SSL VPN 具有明显的优势。

4.2 数据安全分级

为加强用户个人信息安全保护的精细化管理，按各类用户信息的敏感程度，可将用户信息分为 3 个级别，从低到高分别为一级、二级、三级。一级用户信息主要包含用户的消费信息与业务、合作信息；二级用户信息主要包含用户的身份信息，如用户姓名、证件号码、地址信息、银行账号信息等；三级用户信息主要包含用户通信

信息，如用户通信详单、定位（位置）信息、用户身份鉴权信息等。

因此，需要发现定位和分类企业数据库中的敏感信息（哪些是敏感信息），敏感数据应该分类并加密存储，同时需要跟踪哪些用户下载了敏感数据，控制敏感数据下载的生命周期，二级、三级用户信息必须做脱敏处理。对敏感数据保护，可对数据记录分配数据标签，可对应用用户分配用户标签，使用内置的算法实现对表的透明访问。敏感数据发现与分类如图 3 所示。

4.3 数据脱敏

在开发测试、统计分析、生产环境等过程中，需要对于用户敏感信息要采用模糊化方法进行处理，避免批量信息泄露。

数据脱敏处理指对某些敏感数据通过脱敏规则进行数据的变形、加密或屏蔽，实现数据的隐私保护^[10]。

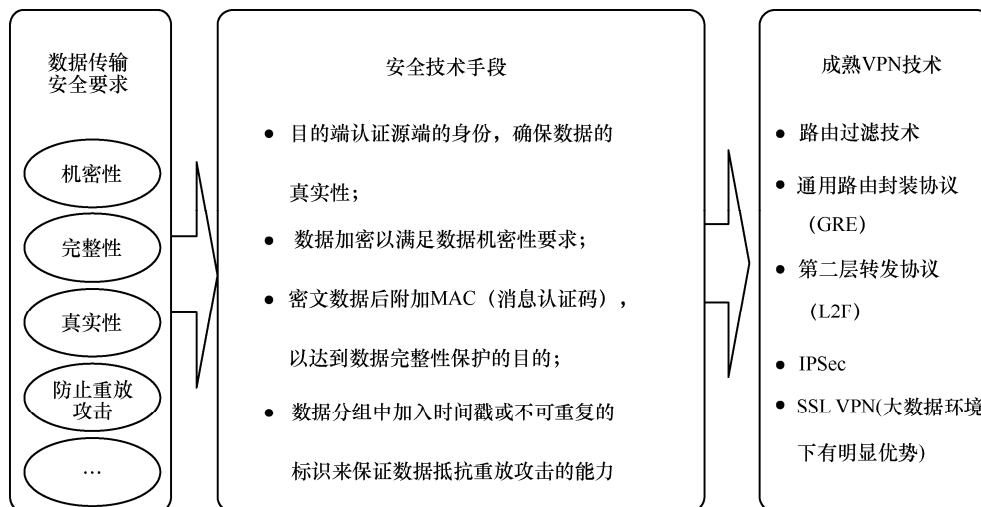


图2 数据采集传输安全

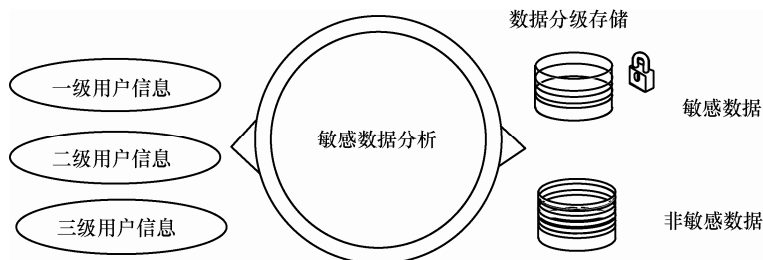


图3 敏感数据发现与分类

数据脱敏的关键步骤具体如下。

- 确定需求：根据业务实际需求，确定数据脱敏的方法以及数据的管理策略等。
- 确定脱敏对象：确定哪些是需要脱敏的敏感数据以及来源。可以通过工具自动探查敏感数据，也可以人工判断。
- 配置脱敏规则：根据不同的数据情况，采用合适的脱敏算法进行配置。
- 测试脱敏规则：通过测试，检验规则的适用度、可靠性、业务适应性以及效率等。如果有新的需求点出现，可以进行迭代开发，不断完善。

数据脱敏的方法主要有：数据替换（截断、加密、隐藏或使之无效）；随机化（以随机数据代替真值）；偏移（通过随机移位改变数字数据）；字符子链屏蔽（为特定数据创建定制屏蔽）；限制返回行数（仅提供可用回应的小分子子集）^[11]。

敏感信息脱敏逻辑部署方案如图4所示。

4.4 数据加密

大数据环境下，可以根据数据敏感特性，对数据进行选择性的加密。敏感数据加密，阻止攻

击者从数据传输过程中获取非授权的敏感信息。

数据加密可分为静态数据（文档、报表、资料等不参与计算的数据）加密和动态数据（需要检索或参与计算的数据）加密。静态数据加密可选择文件或数据块进行加密^[12]。动态数据加密主要考虑数据传输到 Hadoop 系统过程中要进行保护，SASL（simple authentication and security layer，扩充 C/S 模式验证能力的机制）认证框架用于加密流动中的数据。

每个安全级别定义唯一加密密钥，存储在有凭证的 DataNode（数据节点）中。只有授权响应安全级别的用户可以请求对应的密钥，敏感数据也基于安全级别选择加密密钥。当用户运行 MapReduce 作业访问分类的数据集时，要从有凭证的数据节点获取对应的密钥发送给 MapReduce 作业。只有用户通过对应安全级别授权，才可以获取到密钥。在 Hadoop 中存储和获取数据以及在数据流动过程中都要保证数据安全。

4.5 安全基线配置

安全基线是信息系统最小的安全保证，是系统满足最基本的安全要求，在业务系统的整个生命周期的各个环节中对网上的设备以及系

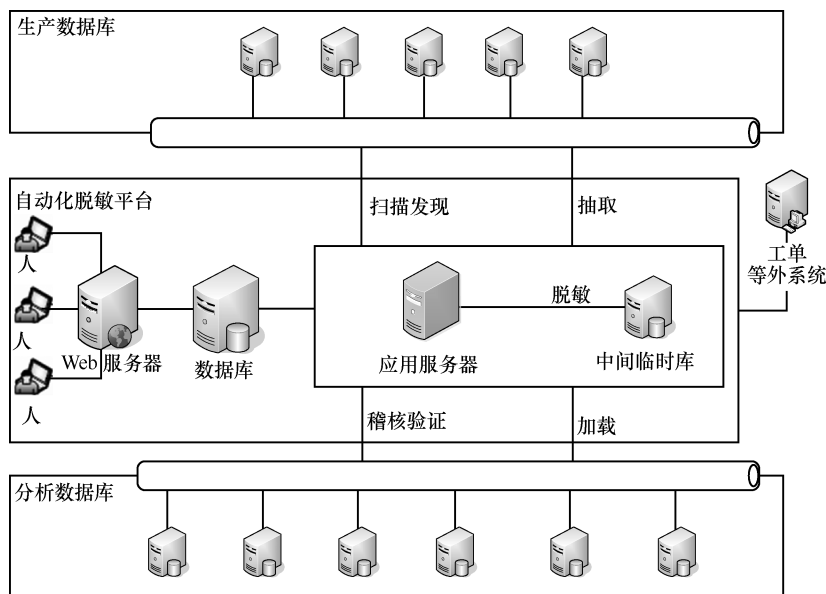


图4 敏感信息脱敏逻辑部署方案示意



统安全配置进行定期检查，而所遵守的最低安全标准^[13]。

安全基线配置包括以下 4 个方面：操作系统、数据库、应用中间件和网络设备^[14]。评估大数据生态系统漏洞和配置缺陷，执行安全加固；保护核心数据安全，建议使用风险评估工具，定期对大数据生态系统组件安全配置进行安全风险检查，以便发现系统组件中的安全隐患，及时进行人工加固；在应用系统上线前，要对应用进行安全风险评估测试，及时发现并修复存在的安全隐患。

对 Hadoop 生态系统组件进行安全配置，所有在 Hadoop 生态系统中运行的服务都要向 KDC（密钥分配中心）进行认证，确保没有恶意进程创建恶意活动。

服务进程需要由 Kerberos（一种安全认证的系统）确保完全安全，运行服务的主机也要限制用户的访问。身份认证至关重要，任何安全失效都会导致恶意用户冒充任何其他合法用户。

加强集群的防火墙策略确保没有非授权访问，需要为 Hadoop 配置安全的 VLAN 并限制所有流量^[15]。防火墙策略确保没有任何来自非法的访问。所有的连接限制建议只能通过网关服务器

发起。

4.6 数据发布安全保护

数据发布前必须进行全面的审查^[15]，确保输出的数据符合“不泄密、无隐私、不超限、合规约”等要求。在数据发布后，一旦出现机密外泄、隐私泄露等数据安全问题，必须要有必要的数据溯源机制，确保能够迅速定位到出现问题的环节、出现问题的实体，以便对出现泄露的环节进行封堵，追查责任者，杜绝类似问题发生。数据发布安全架构如图 5 所示。

5 结束语

大数据、云计算时代，电信运营商逐步把数据和业务搬到云上，面临着新的数据安全挑战。在中国电信云安全实际业务需求的指引下，根据电信运营商数据敏感程度，将数据分为 3 个级别，并且从数据采集传输安全保护、数据脱敏加密、安全基线配置、数据发布阶段安全保护等方面设计数据保护方案，对电信运营商大数据平台数据安全保护机制的新建和完善、安全管理架构的设置、安全管理制度的制定、安全管理流程梳理等有一定的指导意义，并且将纳入中国电信大数据资源池安全加固工程中实施。

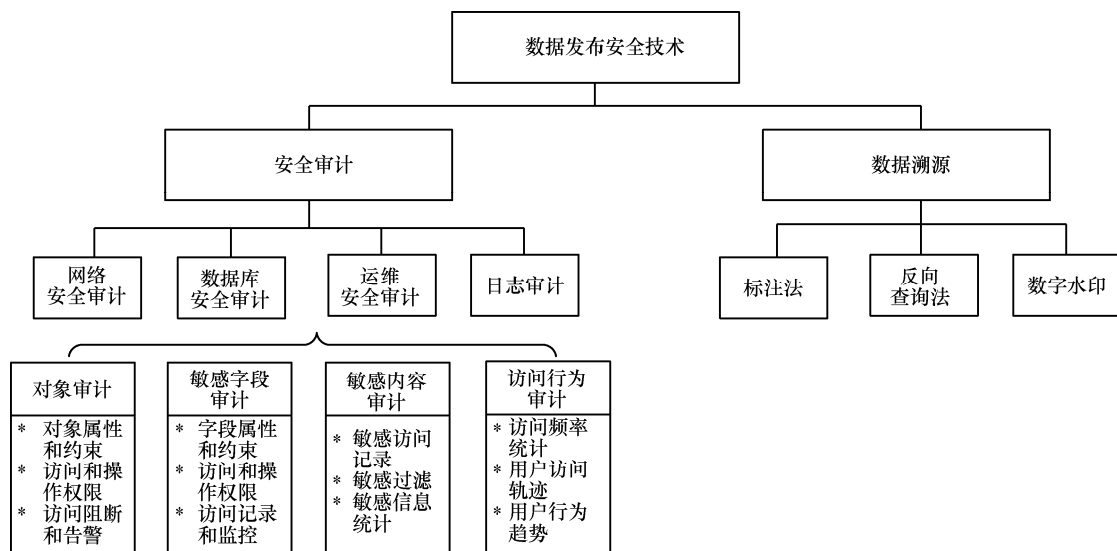


图 5 数据发布安全

参考文献:

- [1] HUANG W, CHEN Z, DONG W, et al. Mobile internet big data platform in China Unicom[J]. Tsinghua Science and Technology, 2014, 19(1): 95-101.
- [2] ZHAO J, WANG L, TAO J, et al. A security framework in G-Hadoop for big data computing across distributed Cloud data centres[J]. Journal of Computer and System Sciences, 2014, 80(5): 994-1007.
- [3] CHEN D, ZHAO H. Data security and privacy protection issues in cloud computing[C]//2012 International Conference on Computer Science and Electronics Engineering (ICCSEE), March 23, 2012, Hangzhou, Zhejiang, China. New Jersey: IEEE Press, 2012: 647-651.
- [4] SUBASHINI S, KAVITHA V. A survey on security issues in service delivery models of cloud computing[J]. Journal of Network and Computer Applications, 2011, 34(1): 1-11.
- [5] ZHOU L, VARADHARAJAN V, HITCHENS M. Achieving secure role-based access control on encrypted data in cloud storage[J]. IEEE Transactions on Information Forensics and Security, 2013, 8(12): 1947-1960.
- [6] PUTHAL D, NEPAL S, RANJAN R, et al. A dynamic prime number based efficient security mechanism for big sensing data streams[J]. Journal of Computer and System Sciences, 2017, 83(1): 22-42.
- [7] 胡坤, 刘镒, 刘明辉. 大数据的安全理解及应对策略研究[J]. 电信科学, 2014, 30(2): 112-117, 122.
HU K, LIU D, LIU M H, et al. Research on Security Connotation and Response Strategies for Big Data [J]. Telecommunications Science, 2014, 30(2): 112-117, 122.
- [8] 张启云. 云计算中数据安全问题的研究 [J]. 计算机光盘软件与应用, 2012, 30(2): 25-26.
ZHANG Q Y. Research on data security in cloud computing [J]. Computer CD Software and Applications, 2012, 30(2): 25-26.
- [9] 卢川英. 大数据环境下的信息系统安全保障技术[J]. 价值工程, 2014, 30(4): 188-190.
LU C Y. Information systems security technology for big data environments [J]. Value Engineering, 2014, 30(4): 188-190.
- [10] 赵祥, 冯华, 韩利凯. 云计算数据安全问题 [J]. 电子世界, 2014, 30(2): 8-9.
ZHAO X, FENG H, HAN L K. Data Security in Cloud computing [J]. Electronics World, 2014, 30(2): 8-9.
- [11] 刘明辉, 张尼, 张云勇, 等. 云环境下的敏感数据保护技术研究[J]. 电信科学, 2014, 30(11): 2-8.
LIU M H, ZHANG N, ZHANG Y Y, et al. Research on sensitive data protection technology on cloud computing [J]. Telecommunications Science, 2014, 30(11): 2-8.
- [12] 王志中, 周城, 牟宇飞. 基于分离密钥的云存储加密解决方案[J]. 电信科学, 2013(11): 51-56.
WANG Z Z, ZHOU C, MOU Y F. A project designed by cloud storage encryption based on separated key[J]. Telecommunications Science, 2013(11): 51-56.
- [13] 唐磊, 马铮. 运营商网络安全基线研究与探讨[J]. 邮电设计技术, 2016(1): 65-69.
TANG L, MA Z. Research and discussion on the network safety baseline of the telecom operators[J]. Designing Techniques of Posts and Telecommunications, 2016 (1): 65-69.
- [14] 张震, 冯伟, 颜金韬. 一种基于云扫描技术的系统安全配置基线核查机制及方法[J]. 电信工程技术与标准化, 2012(4): 31-34.
ZHANG Z, FENG W, YAN J T. New mechanism and method for system security configuration baseline verification based on cloud scanning technique[J]. Telecom Engineering Technics and Standardization, 2012(4): 31-34.
- [15] 陈玺, 马修军, 吕欣. Hadoop 生态体系安全框架综述[J]. 信息安全研究, 2016, 2(8): 684-698.
CHEN X, MA X J, LV X. An overview of Hadoop security framework[J]. Journal of Information Security Research, 2016, 2(8): 684-698.

[作者简介]



周胜利 (1982-), 男, 浙江警察学院工程师, 中国人民解放军理工大学博士生, 主要研究方向为大数据安全管理。



陈斌 (1969-), 男, 华信设计咨询研究院教授级高级工程师, 主要研究方向为云计算、大数据安全。



吴礼发 (1968-), 男, 博士, 中国人民解放军理工大学教授, 主要研究方向为大数据安全。