基于深度内容识别的数据脱敏合规检测平台研究与实现

梁　群

（中通服创发科技有限责任公司，湖南长沙 410016）

【摘要】随着信息化建设的高度发展，信息化给企业生产管理带来了方便快捷，同时也带来了各种信息安全隐患，如何防范数据不被泄露，构筑数据安全应用体系，成为急需解决的问题。文章针对数据在应用和管理过程中可能会出现的问题和风险进行分析，基于深度内容识别技术建设数据脱敏合规检测平台，发现生产运营各个环节是否存在数据脱敏方面的安全隐患，集约化管理，及时减少或杜绝敏感信息泄露的风险，提高安全管理效率。

【关键词】数据安全；内容识别；数据脱敏

【doi:10.3969/j.issn.2095-7661.2020.02.009】

【中图分类号】TP311；TP309　　【文献标识码】A　　【文章编号】2095-7661（2020）02-0032-03

Research and Implementation of DLP-based Data Desensitization Compliance Detection Platform

LIANG Qun

（China Comservice Transfar Technology Co., Ltd. Changsha, Hunan, China 410016）

Abstract: The rapid development of information construction brings convenience and efficiency to enterprise production management, and also exposes it to various information security risks. How to prevent data leaking and how to build data security application system have become urgent problems to be solved. This paper analyses the possible problems and risks in the application and management of data, and with the tentative to build a DLP-based data desensitization compliance detection platform, and detects whether there is a potential data desensitization hidden in every link of production and operation. Through the application of this technology, risks of data leaking will be reduced or diminished, enhancing the efficiency of security management.

Keywords: data security; content recognition; data desensitization

近年来，随着互联网、大数据、物联网和 AI 人工智能等新技术发展和普及，信息化的建设进入了一个新的时代。我国相继出台《关于加强网络信息保护的决定》《电话用户真实身份信息登记规定》《电信和互联网用户个人信息保护规定》等规定，为加强信息保护提供了最基础的法律依据，对数据安全信息保护做了详细要求和规定[1]。运营商通过多年的信息化建设，拥有庞大的信息化建设体系，积累了大量的客户、生产运营、经营决策等方面的业务数据。企业要想持续稳步发展，就必须建设健全完整的信息安全保障体系，企业的信息化程度越高，企业的数据安全也就越重要。这些数据一旦泄露，不仅仅可能给企业的生产经营带来影响，更可能会影响到民生，影响到社会的稳定。面对各种业务系统、海量的业务数据[2]，业务管理和流程是否还存在由于数据防护工作做得不到位而引起的用户隐私数据泄露的安全隐患，以及如何建立有效的敏感数据合规检测机制，已成为企业在信息化建设过程中亟待正视和解决的问题。

1 现状分析

目前运营商在整个业务运营过程中，由于业务流程复杂，用户信息、业务信息等数据在多个业务系统中流转储存，数据脱敏以及数据安全防护工作主要还是依赖于某些独立的系统来进行管理，对于各系统是否按照规范来对业务数据脱敏检测主要还是依托于人工检查，技术手段较缺乏，管理难度大[3]，主要的问题归纳总结如下：

[收稿日期] 2020-01-11

[作者简介] 梁群（1979-），男，湖南衡阳人，中通服创发科技有限责任公司资深网络安全售前专家，研究方向：通信技术、Java 开发、大数据技术。

1)生产运营过程中的管理难点,运营商的信息化体系复杂，系统按照功能域划分为 BSS\OSS\MSS\E-DA\ITSM 等多个域，涉及到的大小系统不下于 100 个,同时,各系统提供了多种对内和对外的渠道接入能力,数据存储的位置和方式也不完全一致,接口错综复杂,敏感数据的脱敏处理大都由各个子系统自行管理,要想通过人工检测来保障敏感数据在整个数据应用的生命周期内不被泄露,工作量和复杂度较大。

2)数据导出过程中的管理难点,为了满足某些业务营销或企业管理的需求,需要导出部分企业财务报表或其它数据,导出的数据是需要根据人员的角色权限、用户敏感信息划分使用级别来进行数据脱敏,导出的数据在生成传输使用过程是否符合安全规范,很难实现对此部分的监管。

3)脱敏数据环境中的管理难点,根据企业内部安全管理的要求,在大数据平台和一些测试平台上保存的数据需要先进行数据脱敏之后才能用于经营分析、决策管理。由于数据存储的地点分散,数据存储的格式和介质不一致[4],在实际的运营过程中,这些数据环境中是否对需要脱敏的数据做过脱敏处理,如何有效地实现自动化的监管,缺少技术和管理手段[5]。

2 典型检测场景与应对措施

针对运营商在运营过程中对用户敏感数据应用的实际情况,我们对脱敏数据应用场景进行了总结和分析,如图 1 所示。

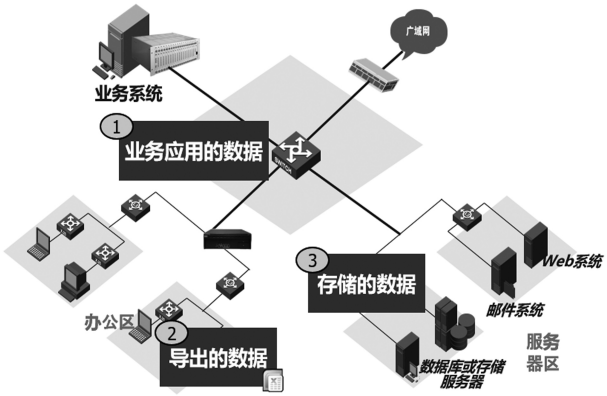

图 1　脱敏数据典型检测场景分析图

2.1 业务应用的数据

业务应用的安全检测场景,主要关注的是在企业运营过程中各系统应用层是否按照相关的要求对用户敏感数据进行脱敏处理,系统间的接口(特别是对外围渠道的接口)传递的数据需要脱敏处理。当前大部分核心运营系统(如:CRM、计费、OSS 等)并没有对系统所需使用的敏感数据进行脱敏处理,即存储在数据库中的用户数据仍采用未脱敏的方式来进行存储的。如图 2 所示,各业务系统需自行根据安全管理规范在服务层对用户数据脱敏,即在接口层或者各系统本身的应用层对数据进行敏感信息处理。

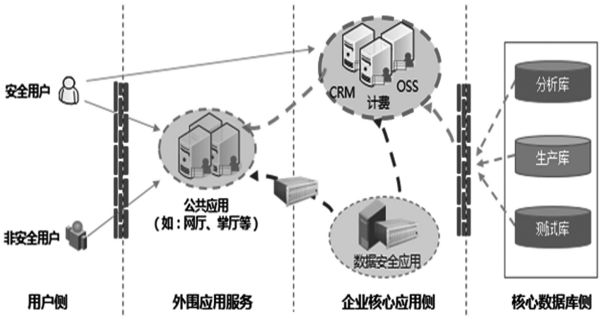

图 2　业务应用场景中脱敏数据的管理图

针对此类情况,我们提供了两种技术手段来对系统是否完成数据脱敏进行检测,描述如下:

1)旁路还原检测:通过旁路流量镜像的方式,还原业务网络上传递的原始业务信息,支持多种网络协议(如:HTTP、SMTP、POP3、SMB、FTP 等)传输的敏感数据进行识别和检测，通过对内容指纹数据的监控,达到对脱敏数据的监控的作用。如:还原网厅中间件应用服务器到 Web 应用服务器间的流量信息，将其还原成最终的页面信息输出到脱敏合规检测平台,通过平台预定义敏感数据指纹信息,动态地识别出信息是否安全脱敏。

2)代理集成检测:代理集成主要是轻量级的代码集成,提供可被集成的 SDK API,外围系统首先识别和收集到需要进行合规性检查监控点,通过代码采用“埋点”的方式进行集成,将应用中的处理数据信息同步发送到检测平台上来,平台基于规则进行合规性判断。如:在客服系统未获取到客户授权前,不能展示客户的脱敏信息,则客服系统需要轻量的集成平台 API 能力,将最终展示的信息也推送到检测平台上,平台再根据预制的规则进行数据脱敏的规范性检查。

2.2 导出的数据

数据导出的安全检测场景,主要关注的是企业在日常的经营生产过程各环节中,往往需要根据业务运营的要求将部分用户信息通过报表或文件的方式提取出来,是否在这个过程中存在用户信息泄露的安全性问题。如:经营分析系统需要根据营销人员的角色权限能获取到不同脱敏级别的用户敏感数据?

针对导出数据合规性检测,也可以通过旁路还原和代理集成的检测技术手段,获取到应用或数据库数据导出的原始信息，将信息汇总至检测平台上来,通过平台的审计能力，完成对数据导出的合规性检测。如:在经分管理平台中下载用户清单需要对非本机号码的数据进行脱敏,可以通过旁路还原的流量信息还原,获取到相关下载文件中的原始信息,并根据规则进行脱敏检查。

2.3 存储的数据

数据存储的安全检测场景,主要指的是监控分析型使用的数据库(如:经营分析系统或培训测试等应用平台)或大数据平台中是否按照要求对用户敏感数据进行脱敏处理,保障该类型系统中存储的用户数据安全合规。如:数据库存储的身份证信息是否进行混淆处理?用户的用户密码是否采用明文来进行保存?如图 3 所示,针对此类场景,目前的安全管理规范要求是在先期的数据准备过程中就对数据库完成预先的敏感信息脱敏处理,再将脱敏后的数据迁移至相应的平台上,用于经营数据的分析和决策。

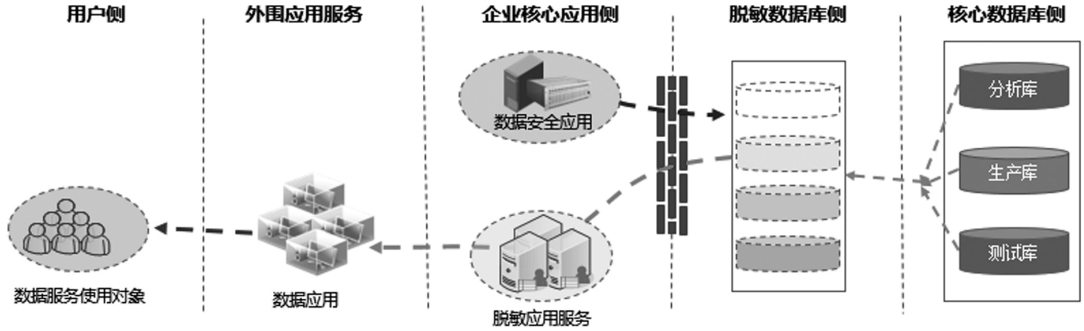

图 3 数据存储场景中脱敏数据的管理图

针对此类情况,需要提供多种类型数据库的动态集成能力,描述如下:

数据存储监测。针对经营分析平台、大数据平台的数据,平台提供众多标准的数据库和非数据库的标准接口适配能力,通过配置各类型的数据监控适配器,平台可针对需要检测的数据制定数据字典,配置相关的元数据信息和指纹信息,采用任务的方式,定时地对这些平台上使用的数据是否达到脱敏要求进行监控。如:通过配置经分系统的数据库连接信息和相关的字典信息,定时扫描相关的存储数据,判断是否满足存放证件号码信息等安全要求。

3 平台概述

数据脱敏合规检测平台是基于深度内容识别技术,依据预先定义或智能机器学习的策略[6],集约化、实时地监控和识别运营过程中对用户数据的使用情况是否符合规范,及时对不合规的行为按照既定的策略执行预警和告警,督促相关的平台或应用做好用户信息安全脱敏的整改,最终建立有效的防止数据泄露的安全防护保障体系[7]。平台的功能架构如图 4 所示。

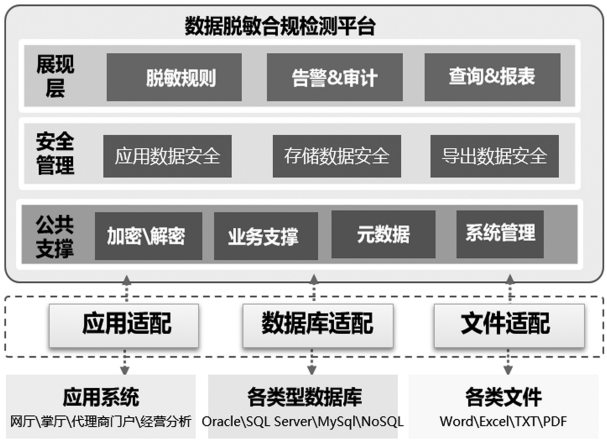

图 4 数据脱敏合规检测平台功能架构图

4 关键技术实现

4.1 深度的内容识别

实现数据的脱敏性安全监测,最基础也是最核心的要求就是能对各类型的数据进行识别,正确获取网络上或应用中需要进行脱敏的用户数据,平台支持的核心能力描述,如表 1 所示。

表 1 已实现的内容识别能力说明表

能力	说明
文件识别	支持目前多种常见的文件和数据格式,如:Word\Excel\ TXT\PDF 等;覆盖 Unix、Linux、Windows 等主流操作系统类型
协议识别	支持多种类型的网络协议,如:HTTP、SMTP、FTP、POP3、IMAP、SMB、NNTP、IM 等
数据识别	支持多种数据的接入识别,支持结构化和非结构化的数据接入;主流关系型数据库:Oracle、SQL Server、MySql、DB2、Sybase、DM、KingBase、Gbase 等;NoSQL 数据库:键值数据库、列存数据库、文档型数据库、图数据库等
内容识别	支持关键字、正则、数据标识符、结构化与非结构化数据指纹、机器学习、OCR 等;国内语言支持丰富:支持简体中文、英文等;字符编码支持丰富,GB2312、GBK、GB18030、Big5、UTF8、ISO8859-1、Unicode 等

4.2 对原有系统“零侵入”

安全检测还有一个重要要求是尽量不影响现有业务的正常运行,数据脱敏合规检测平台支持多种用户数据监听的手段,如:网络旁路内容还原(支持 I-CAP 代理集成、串接路由接入、旁路镜像等多种接入方式)、数据存储检测等手段,均不会影响到原有的 IT 支撑系统的日常使用和运营,通过零侵入技术手段达到用户敏感信息的防范管理效果。同时,平台支持灵活的、规则化、策略化的配置管理,提供元数据配置以及脱敏规则的检测策略配置等功能。(下转第 48 页)

[3]廖忠智,葛滨,张欢迎.高职院校推进校企合作、产教融合长效机制改革的研究与实践[J].湖南邮电职业技术学院学报,2019(1):44-46.
[4]刘桓,陈福明,程艳红.基于产教园的高职院校深化产教融合协同育人的机制探索[J].中国职业技术教育,2018(25):51-56.
[5]吴娜,赵本纲,张红专.职业教育集团化办学与教学管理协同性研究[J].湖南邮电职业技术学院学报,2017(2):74-76.
[6]李增辉.混合所有制下高职院校“企业学院”治理结构及运行机制分析[J].教育与职业,2017(5):57-62.
[7]张艳芳.关于高职混合所有制产业学院的思考[J].职业教育研究,2017(10):15-19.
[8]吴振英,邵利群,倪碧君.产教融合视域下企业学院可持续发展机制研究[J].机械职业教育,2019(9):40-43.
[9]吴显嵘.基于产教融合的高职产业学院建设机理及路径研究[J].中国职业技术教育,2018(29):5-11.
[10]吴金铃.基于产教融合的高职产业学院建设探析[J].教育与职业,2019(18):31-35.

(上接第 34 页)将采集到的数据与指纹数据进行匹配,根据数据对应的场景和检查策略,将数据与检查策略中配置的元数据定义检测规则进行比较,得到数据是否符合脱敏规则的结果,平台动态识别用户敏感数据,无需修改业务系统即可检测敏感数据是否安全合规。

4.3 多元化管理

平台为安全管理人员提供了集约化、可视化、智能化的多元管理手段,实现数据脱敏监控的装载整个流程的自动化、智能化处理,保障数据安全,降低对人员的技术门槛要求,提升管控效率,并提供多维度的数据分析报表,满足审计及监管部门要求。平台采用目前互联网服务化的架构, 支持灵活的动态扩展,特性描述如下:

1)采用 B/S 架构,支持响应式的框架,能支撑多种 PC 和移动终端的接入。

2)平台内置丰富的事件报表、统计仪表,可灵活定制,同时提供丰富的报表格式输出,如:PDF\HTML\Word\PPT\CSV 等。

5 结束语

数据脱敏合规检测平台旨在借助技术手段,将安全管理工作信息化,自动、实时地获取相关信息,将原有传统的被动响应模式转变成主动模式,主动协助管理人员及时地识别出可能存在的安全隐患,通过平台的可视化功能将业务运营过程中对敏感数据运用时可能存在的问题和风险动态进行展示[8],及时预警,规范化管理数据应用流程,保障用户和企业的权益。

【参考文献】

[1]乔宏明,梁奂.运营商面向大数据应用的数据脱敏方法探讨[J].移动通信,2015(13):17-20.
[2]陈天莹,陈剑锋.大数据环境下的智能数据脱敏系统[J].通信技术,2016(7):915-922.
[3]李立高.信息通信建设项目管理中风险认识与风险点确定研究[J].湖南邮电职业技术学院,2018(3):1-3.
[4]邢宇恒,张冰,毛一凡.数据脱敏在海量数据系统中的应用[J].电信科学,2017(S1):8-14.
[5]周倩伊,王亚民,王闯.基于互联网大数据的脱敏分析技术研究[J].数据分析与知识发现,2018(2):58-63.
[6]蒙杨.基于机器学习的数据脱敏系统研究与设计[J].电子技术与软件工程,2019(23):178-179.
[7]吕毅.主动构建数据安全体系,稳步推进数据安全治理[J].中国信息安全,2019(12):54-55.
[8]叶水勇.数据脱敏全生命周期过程研究[J].电力与能源,2019(6):723-727.