动态数据脱敏系统的设计与实现

俞骏豪[1],方 圆[1],宫 帅[1],管建超[1],王靖超[2]
(1. 国网安徽省电力有限公司信息通信分公司,安徽 合肥 230009;
2. 华北电力大学,北京 102206)

摘 要:目前,企业数据集中存储已成为必然趋势,大部分应用采取连接数据中心和数据共享层的方式展开业务。运维人员对业务数据库运维时,敏感数据面临被泄露的风险,存在一定安全隐患。面对组织内部人员造成的数据泄露问题,加强对业务数据的保护是十分必要的。从保护敏感数据的角度出发,针对企业对数据脱敏的需求,研究并设计了一种安全的动态数据脱敏系统。该系统通过定义脱敏方案,结合多维度因素进行组合式的敏感数据访问控制,使不同身份的人员能够访问到相应的生产环境敏感数据。实现了对 MySQL 数据库的实时脱敏操作,提升了数据库运维的安全性。

关键词:动态数据脱敏;MySQL 协议;脱敏算法;数据库代理;运维安全
中图分类号:TP302 **文献标识码**:A **文章编号**:1673-629X(2022)0120-06

Design and Implementation of Dynamic Data Desensitization System

YU Jun-hao[1],FANG Yuan[1],GONG Shuai[1],GUAN Jian-chao[1],WANG Jing-chao[2]
(1. Information Communication Branch of State Grid Anhui Electric Power Co., Ltd.,Hefei 230009,China;
2. North China Electric Power University,Beijing 102206,China)

Abstract:At present,centralized data storage has become an inevitable trend in enterprises. Most applications connect data centers and data sharing layers to implement services. When o&M personnel operate and maintain service databases,sensitive data may be leaked, causing security risks. In the face of data leakage caused by internal personnel,it is necessary to strengthen the protection of business data. This paper studies and designs a secure dynamic data desensitization system to protect sensitive data. By defining desensitization scheme and combining multi-dimensional factors,the system implements combinatorial access control of sensitive data so that people with different identities can access the corresponding sensitive data in production environment. Real-time desensitization of MySQL database is realized,which improves the security of database operation and maintenance.

Key words:dynamic data masking;MySQL protocol;data masking algorithm;database agent;operational safety

0 引 言

当今世界是大数据的时代,数据量的高速增长给如今的数据安全问题带来了很大的挑战,那就是企业和用户的敏感数据保护问题[1]。操作违规事件一般由其内部高权限员工或外包人员造成,但却是所有敏感信息泄露事件中代价最高且最难检测和追责到的事件[2]。因此,敏感数据在运维流程里的安全性必须得到保障。

该文设计了一种数据库脱敏代理方式,以应用于动态数据脱敏系统,使其具备很强的敏感数据访问防绕行能力[3]。在涉及安全、保密等因素及不违反系统规则情况下,通过对用户运维登录时的用户名、密码、IP 地址信息、访问时间和运维工具信息的识别分析,对不同运维用户进行脱敏规则固化配置,按照规则进行数据变形,输出一份符合实际要求的假数据,供运维使用,从而杜绝了数据库通用账户滥用现象,增强了数据保护的稳定性,保证了数据服务质量,提升了数据环境安全[4]。

1 研究现状

传统的静态数据脱敏方式只能针对非生产环境下的数据进行保护,尽管可以提供更好的脱敏效果,但是无法满足实时脱敏的应用需求。常见的静态脱敏技术需要对已有的应用程序进行很大的变更,对性能的影响不可忽略,并且无法提供一种个性化敏感数据管理方式,最终只能呈现单一的保护效果[5]。

由此可见,需要一种与众不同的数据保护方式,既能对不同的运维用户提供有差异的数据保护措施,又

收稿日期:2022-04-07
作者简介:俞骏豪(1990-),男,工程师,研究方向为网络安全与信息通信专业管理。

能保证保护程序的运行对用户是透明的,这就是动态数据脱敏技术[6]。但该技术是存在风险的,需要通过实时的对敏感数据的变形处理,来达到数据实用性和数据保护之间的平衡。如果一味的降低敏感度,会影响数据的可读性和使用价值,但是反之则无法控制用户或企业的隐私泄露问题[7]。此为现有技术的不足之处。

如今,一部分学者早已对动态数据脱敏技术进行了研究,Aleksey I. Baranchikov, Aleksey Yu. Gromov 等在《The technique of dynamic data masking ininformation systems》中就对该技术进行了深入探讨,提出了动态数据脱敏技术在实现过程中的一些难点和脱敏方法[8]。动态脱敏最大的问题就是不能提前知道数据的宏观状况,对脱敏后输出值信息损失量的把控更是困难,这让脱敏后数据呈现的整体效果较静态脱敏会逊色[9]。但是可以有效解决实时数据保护的问题,拥有很明朗的应用前景。尽管现在动态数据脱敏研究仍处在初步发展阶段,但在未来它一定会是一种十分主流的数据安全防护技术。

2　系统关键技术

2.1　两种动态数据脱敏技术

2.1.1　基于结果集处理的脱敏方法

采用基于结果集处理的脱敏方法,可以很好的保护数据的可用性和真实性,维持数据之间的关联关系[10]。图 1 描述了该方式的系统组成。数据库客户端提交查询类型请求至代理程序,此时不做处理直接将其转发至 MySQL 服务端,在代理程序收到带有敏感信息的结果集后,解析 MySQL 数据库协议,根据脱敏规则配置表查询到当前用户下对应的敏感数据库名称、表名称、字段名称和脱敏处理方式,对结果集中的匹配到的敏感数据进行相应处理后,再向数据库客户端返回处理之后的脱敏结果集[11]。

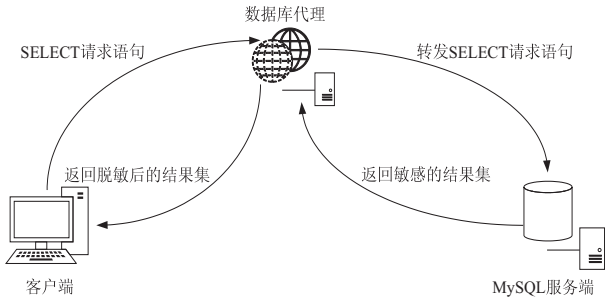

图 1　结果集处理脱敏整体设计

2.1.2　基于 SQL 改写的脱敏方法

采用基于 SQL 改写的动态脱敏方法,其目的是解决对脱敏后数据的关联性、真实性与可用性需求不高的场景下数据保护问题[12]。

图 2 描述了该方式的系统组成。数据库客户端提交查询类型的请求至数据库代理程序,代理程序解析 MySQL 协议,获取到要进行查询访问的数据库名称、表名称和字段名称,与脱敏规则配置表中对该用户的脱敏配置数据进行比对识别,确定需要脱敏的字段和脱敏的方式,按照对应的脱敏函数及参数对 SQL 语句进行修改并回填到请求语句中,最后将修改后的 SQL 语句转发到 MySQL 服务端,服务端直接返回的信息就是脱敏之后的安全信息,一次动态脱敏就完成了[13]。

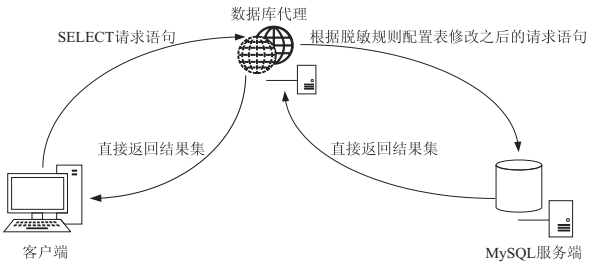

图 2　SQL 改写脱敏整体设计

2.2　脱敏算法

除某些常见特定格式数据的脱敏算法如:地址脱敏算法、姓名脱敏算法、身份证号脱敏算法等,该文还提出一种基于正则表达式的通用脱敏算法[14-15]。

对于像手机号码、邮箱账号和银行卡号码这样的数据类型,它们都具有固定的数据组成结构和范围,所以不可以采用随机数生成函数直接进行处理。但是脱敏系统中各种数据类型之间的元素组成千差万别,若为每一种数据分别设计一种脱敏算法会在很大程度上对脱敏系统造成负担,尤其是对于关联性不强的数据库操作字符型数据,这种代价是不值得的。综上,很有必要设计基于正则表达式的具有通用性的脱敏算法。

正则表达式(Regular Expressions),即正规表达式。正则表达式对字符串数据处理能力十分强大,常用它来描述一类字符串的校验模式,来校验某个字符串里是否具有某类子串或选择符合条件的子串作为替换值。大量的编程语言和软件,都支持使用这种用户可快速编写的表达式。

正则表达式具有很强的灵活性,但是由于它语法相对复杂,所以在实际使用中,并不是所有的运算符都经常用到。在本算法中使用了其中一些常见的字符语法进行组合搭配,完全可以对常见的特定格式数据进行描述。

根据预先设定和用户自定义的正则表达式,反向生成替换数据,这一过程应用有限状态自动机原理,采用先随机输出一个满足语法的字符再转换状态的方式来完成,直到本条正则表达式表达完毕。算法执行步骤描述如图 3 所示。

通用脱敏算法

输入：正则表达式输出：替换字符串
1、 对正则表达式进行解析操作，根据有限状态自动机原理对应的状态转换图，得到状态集 S 与它们之间转换关系集 R
2、 for int i=0 to R.length do
3、 if R 为随机重复打印语法类型 then
4、 随机生成重复次数
5、 else
6、 按语法要求赋值重复次数
7、 end if
8、 for int j=0 to S.length do
9、 从 start 状态开始，随机到达下一个状态，并在这个过程中随机输出可以进行下一个状态转化的字符，直到 end 状态为止
10、 拼接输出的字符为替换字符串
11、 return 替换字符串

图 3 算法执行步骤

由此可见,该脱敏方法支持基本的正则表达式,具有了一定通用性;生成了符合特定格式的替换数据,是可控的;并且生成信息之间并没有关联性,保证了输出值的随机性;因为支持自定义添加规则功能,所以也满足了可定义性。基于正则表达式脱敏算法的引入,在很大程度上增加了脱敏系统的可扩展能力。该文结合特定格式脱敏算法和正则表达式脱敏算法两种方法,各取所长,提升脱敏效率[16]。

2.3 数据库代理

动态数据脱敏主要是通过数据库代理为用户与数据库服务器之间的会话提供代理服务,从而为运维用户提供合规的业务数据展示功能,所以数据库(脱敏)代理模块就是该功能实现的核心[17]。

在整个代理过程中,运维人员登录脱敏系统,点击系统页面中数据库运维客户端 Navicat 进行运维的请求操作,代理模块收到该请求并在合法的情况下连接目标运维设备[18]。

可以把代理分成伪服务端、伪客户端和数据处理单元,数据处理单元处于伪客户端与伪服务端之间,它包括了负责数据传输解析和脱敏的线程类,起到处理两者数据交互的作用。这时,对于 MySQL Client 端来说,代理充当服务器的角色,而对于 MySQL Server 端来说,代理又是客户端的地位。

数据库代理用于启动代理监听以及负责数据传输解析和脱敏的线程类,判断客户端和服务端之间通信的数据,解析和截取有用信息,并记录在相应的日志文件中,以供管理员和对应的审计人员审计。根据运维设备和时间表的设置信息,对本次运维操作进行合规性控制,更大限度保证脱敏系统的安全。

另外,数据库代理对于运维客户端来说是完全透明的,不需要知道代理是如何工作的,增添这个模块不会让其感受到任何的不同,客户端仅需把数据库的连接 IP 配置为本系统的 IP,且连接端口配置为代理的监听端口。

数据库代理通信过程中主要涉及两种数据,客户端发送的数据库请求语句以及服务端响应的结果集。脱敏代理通过 Socket 实现通信的交互过程,mysqlproxy 程序采用多线程的 Socket 进行通信,有几个连接便新建几个连接处理线程,使连接会话之间不受影响。

其主线程流程如图 4 所示。主线程就是监听线程,只负责接受连接并创建处理线程。在主线程里,先应用 socket( )函数创建监听作用的套接字 lis_socket;再调用 bind( )函数绑定主机信息,如 IP 地址和端口号等;然后利用 listen( )函数监听,看是否有 MySQL Client 端的连接请求,当存在合法的请求时,调用 accept( )函数接受连接并创建会话用套接字,之后创建的连接处理线程,仅用于处理本次连接的正常通信。

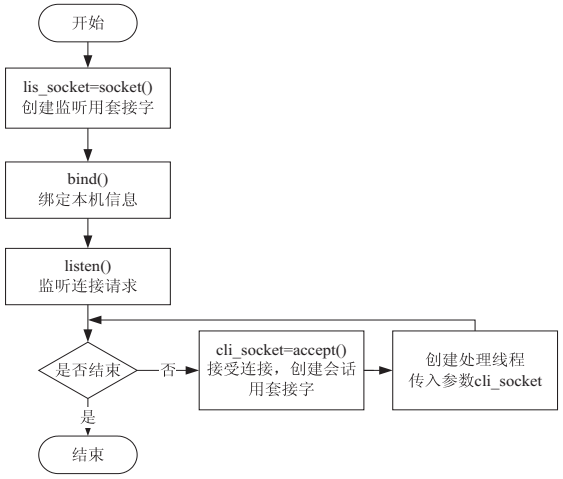

图 4 主线程流程

连接处理线程获取到参数 cli_socket 后,利用该参数与终端通信,这个过程大致可以分成两步:首先检测 MySQL Client 端是否有发来的数据库请求协议包,如果有就接收,并按照相应的配置规则进行解析和改写处理或者直接转发。然后检测 MySQL Server 端是否有发来的应答数据包,如果有则直接转发结果集或者对结果集按规则进行脱敏处理。若 MySQL Client 端与 MySQL Server 端均无有效数据发来,则该连接处理线程将休眠 150 ms,休眠的意义在于,规避因该连接处理线程一直空转而出现的系统资源消耗现象。

3 系统设计

针对企业中对业务数据库运维的实际需求,在对动态数据脱敏关键技术的研究基础上,提出了一种动态数据脱敏系统方案。

本系统主要是为避免出现直连业务数据库时,运维人员越权访问较高级别敏感信息的现象,而设计实现的一种对敏感信息按照提前设定的脱敏算法执行数据相应的变换处理,达到实时返回脱敏后信息的安全

防护方案。该系统基于 MySQL 数据库,同时提供基于 SQL 改写和基于结果集处理的两种脱敏技术路线,可灵活运用不同路线的优点进行脱敏操作。在基于结果集处理的路线下,通过结合替换、遮蔽和模糊化等手段,使脱敏之后的信息可以维持初始信息的逻辑特征、统计特征与一致性特征等,采用代理技术切断用户与业务数据库的直接访问连接。完整的动态数据脱敏系统包括了系统管理模块及动态数据脱敏模块。

本系统建设在生产环境,结合访问权限,不更换初始数据库里字段值,仅把“输入请求”或者“输出数据”做实时脱敏运算,防止敏感信息让未经授权的运维用户查看到。

3.1　总体架构设计

从整体架构上看,该系统采用“物理旁路、逻辑串行”的部署方式,比传统方式有更小的实施成本。有效阻断运维客户端和 MySQL 数据库服务端之间的直连,系统部署时应用 ACL 技术,防止出现敏感数据访问的绕行现象。在不影响数据库正常运维的前提下,通过脱敏代理程序,使运维人员看到符合其安全等级的脱敏后数据信息。本系统的整体架构设计如图 5 所示。

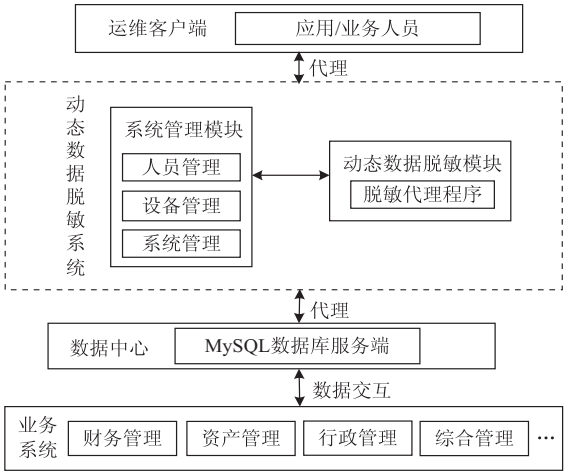

图 5　系统总体架构

3.2　功能模块设计

本系统采用模块化设计思想,保证模块间开发的独立性,提高系统的可靠性和可扩展性。本系统可以分为两个大模块,分别为:系统管理模块,实现了人员及设备的配置与管理;动态数据脱敏模块,结合两种脱敏技术路线,通过代理的方式提供本系统的核心业务,让运维人员透明的访问到目标服务器数据库,从而看到合理的脱敏后的数据信息[19]。

3.3　数据库设计

动态数据脱敏系统主要针对用户信息、设备信息、脱敏方案以及脱敏规则等内容进行管理和设计,在具体系统数据库设计和管理方面不需要太复杂的结构,可梳理出多个数据库实体,其中包括用户数据实体、设备数据实体、工单数据实体、脱敏方案实体、脱敏规则实体。

结合数据实体,明确各数据属性、字段内容,形成系统的具体数据表。

4　动态数据脱敏系统实现

4.1　系统管理模块

系统管理模块给用户提供了可视化的操作界面,为用户自行设置和查看系统相关内容提供了基础和便利。整个管理模块主要分为了三个部分,分别是:人员管理、设备管理和系统服务配置管理。下面将分对这三个模块进行介绍:

(1)人员管理。

动态数据脱敏系统将系统账号分为了四类,分别是:运维账号、审计账号、专责账号和管理员账号,不同类型的账号具有不同的运维权限,四权分立互不干扰,且相互制约。一个普通用户可以拥有一种或者多种类型的账户。

(2)设备管理。

设备管理模块,用户可通过代理连接所有的 MySQL 数据库服务器资源,当用户提出申请访问数据库的工单时,可通过已经配置好的设备信息选择要运维的设备。系统管理员可以通过该模块管理设备组和其中的设备,包括对设备的添加、删除、编辑和导入导出等。

(3)系统管理。

系统管理模块即系统的服务配置,包括系统主页中的公告发布、问题反馈和用户手册等相关文档的发布和下载,也包括对系统参数的配置,比如密码错误次数、密码有效期、网页会话时长等,还包括了对用户权限的配置,用来配置不同身份的人员能够在该系统中操作的权限范围。

4.2　动态数据脱敏模块

4.2.1　实现流程

动态脱敏模块对用户的数据库访问连接进行代理,将整个过程中的相关数据信息做存储和转发操作[20]。动态数据脱敏的实现流程图如图 6 所示。

从其中可以看出,动态数据脱敏不对原始数据做任何处理,实时的在系统用户访问时脱敏输出:首先,动态数据脱敏代理开启,并处于监听状态;代理过滤出 MySQL Client 端发送的数据库查询请求,判断本次脱敏任务的技术实现路线;若为 SQL 改写方式,则根据脱敏方案配置表对 SELECT 语句进行改写,让 MySQL Server 端自行返回脱敏后的数据信息,否则为结果集处理方式,此时数据代理正常转发查询请求,但截取并

解析 MySQL Server 端返回的结果集数据,再根据脱敏方案配置表的规则调用对应的脱敏算法,生成符合要求的随机数据,用生成的数据替换原始数据;返回给 MySQL Client 端脱敏后数据,并向使用者展示输出[21]。至此完成了整个动态数据脱敏的流程。

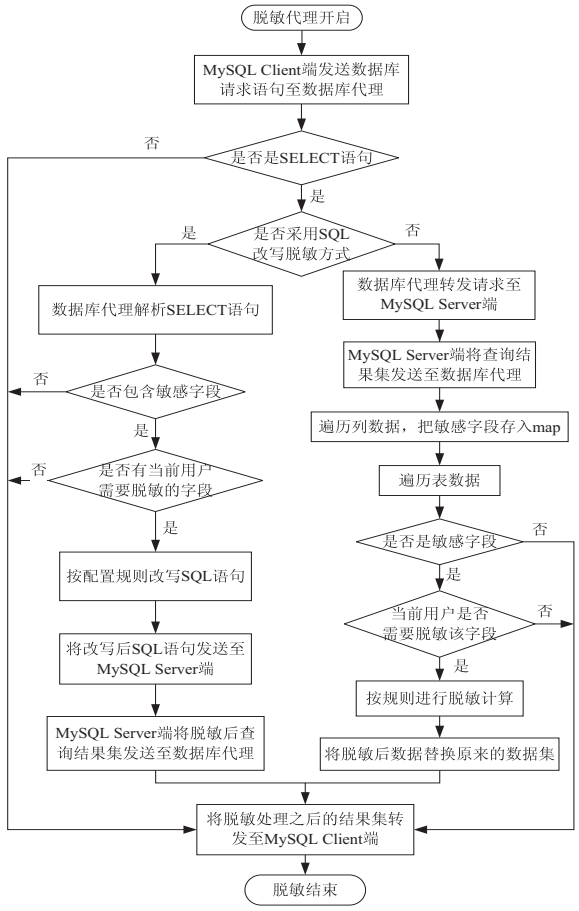

图 6 动态数据脱敏实现流程

4.2.2 子模块

该系统基于数据库代理,实现动态脱敏功能。一次完整的动态脱敏操作需要经过以下五个子模块:脱敏方案配置模块、协议解析模块、脱敏算法模块、数据动态脱敏模块和数据存储模块。

现对动态数据脱敏模块的各个子模块做出如下解释:

(1)脱敏方案配置模块:因为动态数据脱敏不像静态数据脱敏可以提前得到数据源然后做敏感类型的发现,以此确定脱敏方案[21]。因此首先要由管理员在开数据库的运维工单之前配置固化的脱敏方案,包括用户账户、用户 IP 和该账户下需要脱敏的数据库、表、字段等信息。基于敏感信息的配置方法,根据运维身份人员所具备的数据访问权限和数据敏感程度,进行敏感数据信息是否脱敏的判断,如果访问者权限较低或者数据敏感级别较高,则触发动态脱敏进程,反之如果访问者权限较高或者数据敏感级别较低,则并不会触发动态脱敏进程,直接为运维人员展示数据库真实数据。其中“替换” 是通过基于正则表达式的脱敏算法生成的随机数据,“高级替换”是通过特定格式脱敏算法生成的数据,选择“遮蔽”时,可以设定遮蔽的位置和长度。

(2)协议解析模块:动态数据脱敏主要是针对 MySQL Client 端发送的数据库请求报文与 MySQL Server 端响应的结果集报文进行解析和改写。

(3)脱敏算法模块:针对 2.1 节中结果集处理和 SQL 改写两种动态数据脱敏技术路线,分别设计了常见特定格式数据的脱敏方法和基于正则表达式的数据脱敏方法。结合两种方法,因地制宜:在对输出数据的格式和内容没有严格要求的情况下利用正则表达式直接输出随机数据,提高脱敏效率;反之,使用特定格式脱敏算法,输出满足数据真实性和一致性的脱敏结果,提高数据的可用性,各取所长[22]。

(4)数据动态脱敏模块:根据脱敏方案配置表,判断该用户在本次数据库运维工作中是否需要脱敏。

(5)数据存储模块:当匹配到 select 关键词或者结果集信息时,调用数据库接口写入数据库审计表中,用于事后的追责审计工作。

4.2.3 动态数据脱敏测试

图 7 展示的是此次动态脱敏的部分测试数据,并配置脱敏规则姓名列和常驻地址列为“高级替换”,且抖动范围不超过本省,身份证号列为“替换”,手机号列为遮蔽中间四位,性别列不做处理。

ID	姓名	性别	身份证号	手机号	常驻地址
1	周洁	女	1101011990d3d78670	15832609765	山西省朔州市右玉县
2	陈嘉兴	女	230102199007197435	13930765467	河北省秦皇岛市
3	范辉	男	150102201203077131	15076543678	天津市河东区春华街道
4	李静	女	110101199009137678	18822345243	贵州省贵阳市云岩区
5	白可可	女	220102200903075 43X	18930276859	广东省深圳市福田区华富街道
6	王安妮	女	410322201110016524	15100987678	河南省新乡市
7	朱杰	男	110101199009167914	15038768909	上海市庙行镇
8	赵晨义	男	330102200404172054	19203178908	辽宁省大连市
9	李双双	女	120101199003075199	15846576890	北京市昌平区回龙观镇
10	朱珠	女	120111200808036048	18811366578	北京市东城区
11	白如志	男	320102201803071624	18903119087	河北省献县乐寿镇

图 7 脱敏测试数据

按照前文测试规则脱敏后的数据展示如图 8 所示。

ID	姓名	性别	身份证号	手机号	常驻地址
1	周洁	女	1101011990d3d78670	15832609765	山西省朔州市右玉县
2	陈嘉兴	女	230102199007197435	13930765467	河北省秦皇岛市
3	范辉	男	150102201203077131	15076543678	天津市河东区春华街道
4	李静	女	110101199009137678	18822345243	贵州省贵阳市云岩区
5	白可可	女	220102200903075 43X	18930276859	广东省深圳市福田区华富街道
6	王安妮	女	410322201110016524	15100987678	河南省新乡市
7	朱杰	男	110101199009167914	15038768909	上海市庙行镇
8	赵晨义	男	330102200404172054	19203178908	辽宁省大连市
9	李双双	女	120101199003075199	15846576890	北京市昌平区回龙观镇
10	朱珠	女	120111200808036048	18811366578	北京市东城区
11	白如志	男	320102201803071624	18903119087	河北省献县乐寿镇

图 8 脱敏测试结果

由图 8 可以得知,根据脱敏方案可得到正确的数据输出,通过功能测试。

5　结束语

针对企业中对业务数据库运维的实际需求,在对动态数据脱敏关键技术的研究基础上,提出了一种动态数据脱敏系统方案。该系统采用代理的方式提供脱敏服务,创新性的结合了动态数据脱敏的两种技术路线,各取所长,有效提高了业务数据库中数据的安全性,在很大程度上解决了组织内部数据泄露的问题,实现了对运维操作的权限管控。

参考文献:

[1] ADAM N R,WORTHMANN J C. Security-control methods for statistical databases:a comparative study[J]. ACM Computing Surveys,1989,21(4):515-556.

[2] 李　琦. 差分隐私数据发布方法的改进及应用研究[D]. 武汉:武汉理工大学,2018.

[3] 冉　冉,李　峰,王欣柳,等. 一种面向隐私保护的电力大数据脱敏方案及应用研究[J]. 网络空间安全,2018,9(1):105-113.

[4] GUJJARY V A,SAXENA A. A neural network approach for data masking[J]. Neurocomputing, 2011, 74(9): 1497-1501.

[5] 梁浩波,封祐钧,林浩钊. IT 一体化运维管控技术与管理研究[J]. 计算机安全,2014(4):47-51.

[6] 杜　凯. 增强型身份保持的隐私保护方法研究[D]. 桂林:广西师范大学,2017.

[7] 吴行飞. 中小城市商业银行数据脱敏研究[D]. 济南:山东大学,2016.

[8] MURALIDHAR K, SARATHY R. Data shuffling: a new masking approach for numerical data[J]. Management Science,2006,52(5):658-670.

[9] 张　冰. 面向数据发布的隐私保护方法研究[D]. 哈尔滨:哈尔滨工程大学,2015.

[10] 林　山. 中小企业信息安全问题及解决方案[D]. 重庆:重庆大学,2007.

[11] JOHNSON D B,MATYAS S M,LE A V,et al. The commercial data masking facility (CDMF) data privacy algorithm[J]. IBM Journal of Research & Development,1994,38(2):217-226.

[12] 石宏宇. 基于堡垒机技术的运维安全管控系统设计与应用[J]. 中国管理信息化,2016,19(24):44-45.

[13] 叶水勇,刘　琦,陈　明,等. 运维专区管理系统研究与建设[J]. 国网技术学院学报,2019,22(1):40-43.

[14] 李　莹,孙秀胜. 银行数据安全保护关键技术分析与探讨[J]. 科技传播,2018,10(19):110-111.

[15] 韩雨彤. A 供电企业信息安全管理体系研究[D]. 北京:华北电力大学(北京),2018.

[16] 刘哲理,贾春福,李经纬. 保留格式加密模型研究[J]. 通信学报,2011,32(6):184-190.

[17] 江堂碧. 支持挖掘的流式数据脱敏关键技术研究[D]. 成都:电子科技大学,2017.

[18] 蒲　昊. 基于 Http 反向代理的 Web 动态适配平台的设计与实现[D]. 西安:西安电子科技大学,2014.

[19] ZHOU Y,LOUIS T A. A smoothing approach for masking spatial data[J]. Annals of Applied Statistics,2010,4(3):1451-1475.

[20] 李　敏. 保留格式加密技术应用研究[D]. 天津:南开大学,2012.

[21] ARMSTRONG M P, RUSHTON G, ZIMMERMAN D L. Geographically masking health data to preserve confidentiality[J]. Statistics in Medicine,1999,18(5):497-525.

[22] 臧　昊,赵　强,卞水荣. 基于 XML 的电子病历隐私数据脱敏技术的研究与设计[J]. 信息技术与信息化,2017(3):111-114.

(上接第 119 页)

微电子学与计算机,2019,36(3):43-47.

[12] 宋紫华,郭　春,蒋朝惠. 一种基于网络流量分析的快速木马检测方法[J]. 计算机与现代化,2019,5(6):9-15.

[13] 方　澄,殷明瑞,张礼哲,等. 基于 Sketch 数据结构的海量网络流量实时排名系统[J]. 计算机应用,2019,5(1):70-74.

[14] 王　婷,王　娜,崔运鹏,等. 基于半监督学习的无线网络攻击行为检测优化方法[J]. 计算机研究与发展,2020,3(4):791-802.

[15] 黄予春,曹成涛,顾　海. 基于云计算和深度学习的电力电容器故障诊断和识别[J]. 电力电容器与无功补偿,2018,39(4):71-75.

[16] 王　勇,周慧怡,俸　皓,等. 基于深度卷积神经网络的网络流量分类方法[J]. 通信学报,2018,39(1):14-23.

[17] 刘汝媚,黎晓凤. 复杂光纤网络流量数据快速调度模型分析[J]. 激光杂志,2019,40(7):87-91.

[18] 刘纪伟,赵月显,赵　杨. 一种基于统计排序的网络流量特征选择方法[J]. 电子技术应用,2018,44(1):90-93.

[19] 何　婕,赖　敏. 云计算平台中分布式 Hadoop 数据挖掘关键技术研究(英文)[J]. 机床与液压,2018,46(24):150-155.

[20] 赵　英,韩春昊. 马尔科夫模型在网络流量分类中的应用与研究[J]. 计算机工程,2018,44(5):291-295.

[21] 魏松杰,吴　超,罗　娜,等. 移动蜂窝网络流量的时延特征识别方法研究[J]. 计算机研究与发展,2019,56(2):139-150.

[22] 蒲晓川. 大数据环境下的网络流量异常检测研究[J]. 现代电子技术,2018,41(3):84-87.

[23] 龙震岳,艾解清,邹　洪,等. 基于改进灰狼优化算法的网络流量预测模型[J]. 计算机应用研究,2018,35(6):1845-1848.

[24] 董书琴,张　斌. 基于深度特征学习的网络流量异常检测方法[J]. 电子与信息学报,2020,42(3):26-31.