doi:10.3969/j.issn.1002-0802.2016.07.023

大数据环境下的智能数据脱敏系统 *

陈天莹[1]，陈剑锋[1,2,3]

（1. 中国电子科技网络信息安全有限公司，四川 成都 610041；
2. 中国电科网络空间安全技术重点实验室，四川 成都 610041；
3. 保密通信重点实验室，四川 成都 610041）

摘　要：随着大数据时代的到来，大数据中蕴藏的巨大商业价值得以挖掘并面世，同时也带来了隐私、敏感信息保护方面的棘手难题。大数据安全区别于传统信息安全的本质在于数据层面，即如何在实现大数据高效共享、分析挖掘的同时，保护敏感及隐私信息不被泄露。通过对现有数据脱敏技术原理、机制和过程等方面的深入研究，总结当前主流脱敏方法存在的缺点和不足，创新性地提出了大数据环境下的智能数据脱敏系统。该系统能够以集中式、低耦合和高容量的方式，帮助政府、企业等用户解决敏感及隐私数据在共享、交换及使用过程中的难题。

关键词：大数据安全；数据脱敏；信息安全；数据隐私
中图分类号：TP309.2　　**文献标志码**：A　　**文章编号**：1002-0802(2016)-07-0915-08

Intelligent Data Masking System for Big Data Productive Environment

CHEN Tian-ying[1], CHEN Jian-feng[1,2,3]

（1.China Electronic Technology Cyber Security Co.,Ltd, Chengdu Sichuan 610041, China;
2.Cyberspace Security Technology Laboratory of CETC, Chengdu Sichuan 610041, China;
3.Science and Technology on Communication Security Laboratory, Chengdu Sichuan 610041, China）

Abstract：With the arrival of big data era, huge amount of business interest in big data is explored and great potential value mined and utilized. However, this revolution also leads to severe problems of private and sensitive information protection. The main difference of between traditional information security and big data security lies in the data content layer, this means to protect private or sensitive data from being disclosed while retain the ability to effectively share, analyze and distribute such data. Through the in-depth research on current data masking principle, mechanism and process, the shortcomings and deficiencies of existing data masking methods are summed up, and an innovative solution dynamic data masking system suitable for big data productive environment is proposed, which can meet the needs of various enterprises with its service-centric architecture, high throughput capacity and low coupling nature for data masking, exchange and application.

Key words：big-data security; data masking; information security; data privacy

0　引　言

信息技术与经济社会的交汇融合引发了数据的迅猛增长，数据成为国家基础性战略资源。进入大数据时代，企业收集的数据越来越多，数据外泄事件一再发生，企业信息受到严重威胁。为此，企业积极投资于数据隐私和数据安全技术，将不可预见的安全成本最小化并减少风险损失。Ponemon Institute 公司针对美国近年来数据外泄成本进行的

* 收稿日期：2016-03-12; 修回日期：2016-06-12　Received date:2016-03-12;Revised date:2016-06-12

基础研究发现，平均每起数据外泄的成本为 720 万美元，每条外泄记录的成本为 214 美元，其中最高的数据外泄比率来自于内部人员的疏忽，占比为 41% [1]。这意味在业务分析、开发测试、审计监管等使用场合中，敏感数据具有极高的安全风险。如何在这些阶段中确保生产数据的安全，已经成为业界极为关注的问题。

在大数据快速推动国家信息化发展的整体趋势下，第十三个五年规划纲要中明确提出：“实施国家大数据战略，推进数据资源开放共享。”然而，各行业数据资源中往往包含大量的敏感和重要信息，一旦泄露或遭到非法利用，将会给个人甚至是国家带来无法弥补的损失。同时，随着大数据分析的成熟和价值挖掘的深入，从看似安全的数据中还原出用户的敏感、隐私信息已不再困难。如何在数据交换、共享及使用等过程中实现对敏感数据的定向、精准和彻底脱敏，达到数据安全、可信、受控使用的目标，是数据产生者和管理者亟待解决的技术问题。

数据脱敏又称数据去隐私化或数据变形，是在给定的规则、策略下对敏感数据进行变换、修改的技术机制，能够在很大程度上解决敏感数据在非可信环境中使用的问题。Gartner 认为，数据脱敏应成为相关企业在软件开发、数据分析和培训时的强制选项 [2]。目前，数据脱敏的主要实践者包括 IBM、ORACLE 和 Informatica。他们凭借在传统数据库行业较早的进入时间、较深厚的实践经验和技术积累，占据了多数市场份额。相较国内，数据脱敏的研究和应用刚刚起步，银行、通信运营商根据自身需求制订了一些数据脱敏解决方案，但多以静态脱敏为主，设计流程固定，工具能力有限，专用性较强，配置规则复杂，维护困难，不能满足数据交互流量的不断增长和复杂多变的安全处理需求 [3-4]。

论文第一部分将介绍大数据环境下敏感数据面临的风险、已有防护手段和数据脱敏原理，第二部分概括由目标、策略和实现机制构成的数据脱敏过程，第三部分则结合大数据环境的实际需求，设计并实现智能大数据脱敏系统，同时阐述其架构、处理流程、基本功能及运行模式。最后总结全文。

1 数据脱敏的动机

1.1 敏感数据的安全风险

敏感数据又称隐私数据，常见的有姓名、身份证号码、住址、电话、银行账号、邮箱、密码、医疗信息、教育背景等。这些与个人生活、工作密切相关的信息受到不同行业和政府数据隐私法规的管制。如果负责存储和发布这些信息的企业或政府无法保证数据隐私，他们就会面临严重的财务、法律或问责风险，同时在用户信任方面蒙受巨大损失。

敏感数据在其生命周期的各个环节，也即数据的产生、存储、应用、交换等环节中均存在被泄露和攻击的风险。这些风险包括网络协议漏洞、数据库入侵、内部人员越权访问、社会工程学、高级持续性威胁以及合法人员的错误配置等。多数企业将安全工作的重心放在外围安全和终端防护上，往往购买防火墙、反病毒软件，并对网络设备进行安全配置。但是，对于数据这一企业的核心资产而言，这种防护方式实现的能力有限。随着大数据时代信息的价值性越来越突出，企业应当将安全投资侧重于保护数据层面的攻击风险，以取得理想的安全收益。

1.2 当前的数据安全防护手段

数据安全是信息安全的重要一环。当前，对数据安全的防护手段包括对称 / 非对称加密、同态加密、访问控制、安全审计和备份恢复等。

对称 / 非对称加密：加密是对原来为明文的数据按某种算法进行处理，使其成为不可读的乱码，从而达到保护数据而不被非法窃取、阅读的目的。传统加密技术由对称、非对称和散列算法构成，具有极高的安全强度，能够保证数据在传输过程中的机密性和完整性。但是，由于数据在使用时必须完全解密，对最终用户而言，敏感数据依然是明文，因而无法同时满足敏感数据安全性和可用性的需求。

同态加密：同态加密是一种加密形式，允许人们通过精心设计的密码算法对密文进行特定的代数运算并解密，其结果与对明文进行同样的运算结果一致。同态加密能够从根本上解决将数据及其操作委托给第三方时的保密问题，但由于加密后的数据缺乏语义，因而除简单的统计外，无法执行更精细的数据分析、挖掘和价值发现等操作。另外，当前同态的性能也远未达到生产级别数据的处理需求。

访问控制：根据预定义的数据模型和用户角色模型，对数据库、数据表的访问行为进行检测和判断，在必要时阻断查询语句以保护敏感信息的安全。访问控制虽然提供了一定意义上的敏感数据保护能力，但是这种粗粒度的拦截方式难以满足甚至违背了大数据环境下共享交换、综合分析挖掘的需求和

原则。

安全审计：对数据请求进行全时严密监控，对敏感信息的访问者和访问时间进行详细的审核和记录，通过安全分析检测非法行为，并与其他手段联动对违规事件进行处置。安全审计的缺点在于，它是一种事后核查机制，只能在发生数据泄漏问题后才能生效，无法实时对攻击进行拦截和阻断以实现防患于未然。

备份恢复：通过分布式存储、冗余和恢复来实现数据的容灾安全性，是一种可用性机制。

综上所述，这些手段均有各自的优点和适应领域，但它们用于敏感数据防护方面仍有欠缺，无法在不妨碍已有的数据处理、操作及分析过程的同时，实现对敏感数据的针对性保护。

1.3　数据脱敏原理

数据脱敏在保留数据原始特征的条件下，按需进行敏感信息内容的变换。只有授权的管理员或用户，在必须知晓的情况下，才可通过特定应用程序与工具访问数据的真实值，从而降低这些重要数据在共享和移动时的风险。数据脱敏在不降低安全性的前提下，使原有数据的使用范围和共享对象得以拓展，因而是大数据环境下最有效的敏感数据保护方法。

任何涉及敏感信息的行业都对数据脱敏有着天然的需求。其中，金融、政府和医疗行业首当其冲。相关单位在应用开发、测试、培训等活动中普遍使用真实数据，导致数据在暴露期间面临严重泄露风险。在数据脱敏的帮助下，企业能够按照数据使用目标，通过定义精确、灵活的脱敏策略，按照用户的权限等级，针对不同类别的数据以不同方式脱敏，实现跨工具、应用程序和环境的迅速、一致性的访问限制。

数据脱敏通常遵循的几条原则包括 [5]：

（1）数据脱敏算法通常应当是不可逆的，必须防止使用非敏感数据推断、重建敏感原始数据。但在一些特定场合，也存在可恢复式数据脱敏需求。

（2）脱敏后的数据应具有原数据的大部分特征，因为它们仍将用于开发或测试场合。带有数值分布范围、具有指定格式（如信用卡号前四位指代银行名称）的数据，在脱敏后应与原始信息相似；姓名和地址等字段应符合基本的语言认知，而不是无意义的字符串。在要求较高的情形下，还要求具有与原始数据一致的频率分布、字段唯一性等。

（3）数据的引用完整性应予保留，如果被脱敏的字段是数据表主键，那么相关的引用记录必须同步更改。

（4）对所有可能生成敏感数据的非敏感字段同样进行脱敏处理。例如，在学生成绩单中为隐藏姓名与成绩的对应关系，将“姓名”作为敏感字段进行变换。但是，如果能够凭借某“籍贯”的唯一性推导出“姓名”，则需要将“籍贯”一并变换。

（5）脱敏过程应是自动化、可重复的。因为数据处于不停的变化中，期望对所需数据进行一劳永逸式的脱敏并不现实。生产环境中数据的生成速度极快，脱敏过程必须能够在规则的引导下自动化进行，才能达到可用性要求；另一种意义上的可重复性，是指脱敏结果的稳定性。在某些场景下，对同一字段脱敏的每轮计算结果都相同或者都不同，以满足数据使用方可测性、模型正确性、安全性等指标的要求。

2　数据脱敏过程

2.1　脱敏目标确认

数据脱敏通常会带来一定的业务性能开销，其运行和维护过程也需要成本投入。企业应根据自身的业务运行特征、数据资产价值和风险承受能力制订不同的脱敏目标。

脱敏目标中较为关键的部分是数据敏感程度的分级和确认，包括确认原始数据的主观敏感度、在各种使用场景下的关联性、脱敏后数据在系统开发测试方面的可用性等。敏感信息字段名称、敏感级别、字段类型、字段长度、赋值规范等内容，需要在这一过程中明确，以作为脱敏策略制订的依据。

2.2　脱敏策略制订

脱敏策略是在脱敏过程中贯彻的规则、规范、方法和限制的统称。脱敏规则是根据数据及用户的特点制订的全局和个别配置，用以指导脱敏过程的实现；脱敏规范是数据在处理中必须遵循的安全法规及行业标准；脱敏方法是对敏感数据进行具体变换操作的算法及流程；脱敏限制是应用脱敏方法时受到的条件和制约，如时空复杂度要求、时效性要求、接口要求等。

在脱敏策略中，脱敏方法是数据脱敏的重心和难点，包括可恢复和不可恢复两类，原理都是将原始数据转换为“看起来很真实的假数据”。

几种常见的脱敏方法包括[6]：

替换：以虚构的数据代替真值。例如，建立一个较大的虚拟值数据表，对每一真实值记录产生随机种子，对原始数据内容进行哈希映射替换。这种方法得到的数据与真实数据非常相似。

无效化：以 NULL 或 ***** 代替真值或真值的一部分，如遮盖信用卡号的后 12 位。

置乱：对敏感数据列的值进行重新随机分布，混淆原有值和其他字段的联系。这种方法不影响原有数据的统计特性，如最大 / 最小 / 方差等均与原数据无异。

均值化：针对数值型数据，首先计算它们的均值，然后使脱敏后的值在均值附近随机分布，从而保持数据的总和不变。通常用于产品成本表、工资表等场合。

反推断：查找可能由某些字段推断出另一敏感字段的映射，并对这些字段进行脱敏，如从出生日期可推断出身份证号、性别、地区的场景。

偏移：通过随机移位改变数字数据。

FPE：Format Preserving Encryption，即格式维持的加密是一种特殊的可逆脱敏方法。通过加密密钥和算法对原始数据进行加密，密文格式与原始数据在逻辑规则上一致，如都为日期、卡号、结构化值等。通过解密密钥可以恢复原始数据。

基于其他参考信息进行屏蔽：根据预定义规则仅改变部分回应内容（例如，屏蔽 VIP 客户姓名，但显示其他客户等）。

限制返回行数：仅提供响应数据的子集，防止用户访问到全部符合要求的数据。

2.3 数据脱敏实现

按照作用位置、实现原理不同，数据脱敏可以划分为静态数据脱敏（Static Data Masking，SDM）和动态数据脱敏（Dynamic Data Masking，DDM）。随着数据脱敏的应用领域从非生产系统拓展到生产系统，业界的技术需求也逐步从 SDM 过渡到 SDM/DDM 并重。

SDM 一般用于非生产环境。在不能将敏感数据存储于非生产环境的场合中，通过脱敏程序转换生产数据，使数据内容及数据间的关联能够满足测试、开发中的问题排查需要，同时进行数据分析、数据挖掘等分析活动。而 DDM 通常用于生产环境，在敏感数据被低权限个体访问时才对其进行脱敏，并能够根据策略执行相应的脱敏方法。SDM 与 DDM 的区别在于，是否在使用敏感数据时才进行脱敏。这将影响脱敏规则的实现位置、脱敏方法和策略等参数。

目前，在传统关系型数据库中，SDM 依然是重要的数据保护方法，其执行能力、质量和可扩展性较好，适合在数据的时效性需求不高的场合中使用。然而，在大数据环境中，数据的海量、异构、实时处理将成为常态，能够在不影响数据使用的前提下，在用户层面实现数据屏蔽、加密、隐藏、审计或内容封锁的 DDM 具有更强的优势。DDM 基于横向或纵向的安全等级要求，依据用户角色、职责和其他规则变换敏感数据，其能力的发挥对大数据的广泛、合规应用至关重要。

动态数据脱敏目前具有两类实现机制：基于视图的实现机制和基于代理的实现机制。

2.3.1 基于视图的实现机制

在这类机制中，生产数据及脱敏后的数据版本通常存放在同一数据库中，用户能够访问到的数据内容范围取决于其角色的权限等级。在用户访问请求发出时，该请求被与数据库集成的脱敏组件截获，高权限用户获得原始数据的完整版本，低权限用户或未使用指定方式访问的用户获得数据的脱敏版本。由于这种判决是在请求到达时刻完成，用户与权限、脱敏数据视图的对应关系需要预先定义。在敏感数据被脱敏访问时，控制中心将收到一条通知或警告。

基于视图的动态数据脱敏的一种实现方式是编写数据库程序代码，在权限判决后对请求语句进行重写，以寻址原始数据或脱敏数据；另一种方式是建立数据库的真实视图即虚拟数据表，使应用程序如同访问真实数据表一样访问脱敏后的数据。这种方式需要为虚拟数据表构建触发器、存储过程等，以处理数据请求，其原理图如图 1 所示。

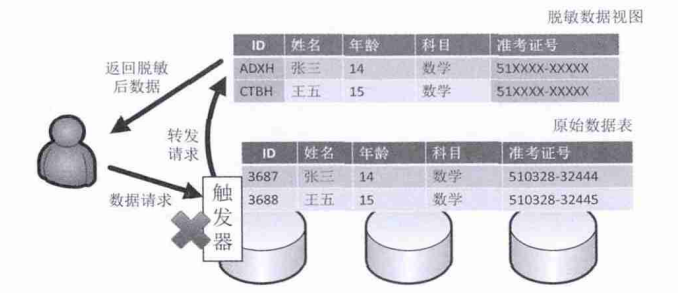

图 1 基于视图的动态脱敏实现机制

2.3.2 基于代理的实现机制

与视图方式相比，基于代理的实现机制适应性

更强，灵活性也更高。用户的数据请求被代理实时在线拦截并经脱敏后返回，此过程对于用户及应用程序完全透明。这种机制与视图方法的不同点在于，脱敏判决是在数据容器外实现，因而能够适用于非关系型数据库，如大数据环境。脱敏代理部署在数据容器的出口处以网关方式运行，检测并处理所有用户与服务器间的数据请求及响应。这种实现机制的好处是，无需对数据存储方式及应用程序代码做出任何更改。

代理实现数据脱敏的具体方法是查询语句或响应语句替换。代理能自动识别目标为敏感数据的查询语句，并将语句改写为不包含敏感字段，或对敏感字段进行变换处理的查询语句。查询结果返回代理时，会被重新计算、修改并包装为与原请求一致的格式交付用户，从而完成一次敏感信息的查询过程，其原理图如图 2 所示。

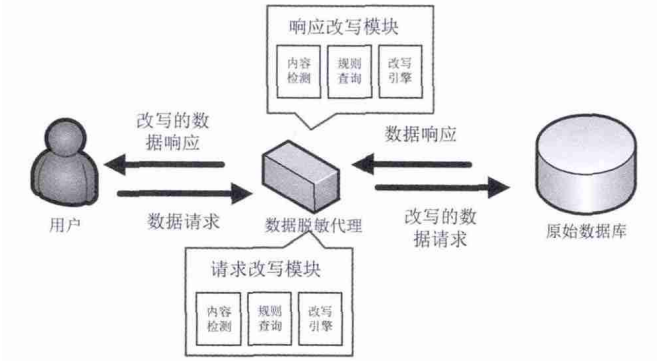

图 2　基于代理的动态脱敏实现机制

就这二类实现机制而言，基于视图的方式尽管效率较高，但需要修改数据库结构及代码，而基于代理的方式又在扩展性和统一管理方面存在欠缺，两者均难以应对大数据环境中数据脱敏的严峻挑战。因此，本文提出了智能大数据脱敏系统，通过分析大数据环境下的敏感数据类型、使用场景等，设计了合理的系统框架及脱敏方式。

3　系统实现

3.1　系统架构

智能大数据脱敏系统架构从底至上由四个层次构成，即资源层、数据层，服务层和应用层。横向包含两大管理功能，即安全管理和运维管理。系统架构图如图 3 所示。

资源层：为数据脱敏服务提供基础性物理资源，包括计算资源、网络资源和存储资源等；

数据层：包括支持系统完成智能敏感数据发现、脱敏的各类数据库、知识库，针对不同敏感数据的脱敏规则库，管理规则及规则集合的脱敏策略库，支持智能敏感数据发现的本体知识库和机器学习所形成的模型库，运维管理和安全管理所需的权限库等。

服务层：以松耦合的方式承载数据脱敏所需的一系列核心服务及中间件，提供数据脱敏、规则化和服务化三大引擎，支撑大数据多元异构敏感数据发现和脱敏操作。

图 3　智能大数据脱敏系统架构

应用层：面向最终用户，按照数据类型，提供数据库脱敏、文件脱敏以及多媒体脱敏；按照业务需求，分为测试和研发过程所需的静态脱敏和生产过程中对敏感数据访问及应用的动态脱敏。

运维管理：包括用户、策略、数据源等系统要素及配置的管理，确保系统的可用性；

安全管理：包括权限、角色和合规性等安全隐私要素及配置的管理，确保系统的对外安全性和自身安全性。它与运维管理的协同，使数据脱敏服务的运行时刻处于严密和安全防护及监控之下。

3.2 系统处理流程

智能大数据脱敏系统主体流程包括脱敏需求配置、敏感数据识别、脱敏策略配置、脱敏服务运行和脱敏状态监控五个环节。

用户需求配置：根据用户的资产重要性和数据价值对脱敏的粒度、强度和目标进行定义和配置。

敏感数据识别：对目标系统的全量数据进行智能识别，获取用户数据源中数据元信息、数据结构等。对数据字段的内容进行分析，对格式和语义进行识别，对主键 / 外键进行处理，识别出系统中存在的敏感数据。

脱敏策略配置：提供两种脱敏策略的配置方式，一种是基于系统内置的敏感数据类型，采用智能推荐方式进行脱敏策略的配置；另一种是支持用户自定义脱敏策略以及更改合适的脱敏算法。

脱敏服务运行：按照用户需求进行静态数据脱敏和动态数据脱敏。

脱敏状态监控：持续对脱敏系统的运行情况进行监控和审计，及时发现异常并做出响应。定期将综合后的运行结果反馈用户，完善脱敏需求配置，提升脱敏效果。

3.3 敏感数据识别方法

敏感数据识别是智能数据脱敏系统中的核心和关键。大数据环境中，非结构化数据占 85% 以上，因而非结构化数据的敏感数据识别、发现、处理是迫切需要解决的问题，否则数据脱敏系统的实用性将大打折扣。图 4 描述了数据库（主要是结构化数据）和文件（主要是非结构化数据）的敏感数据识别方法，其核心技术采用数据特征学习以及自然语言处理等技术进行敏感数据识别。

敏感数据识别分为两个阶段，即数据源注册和数据脱敏任务执行。

（1）数据源注册阶段。数据源注册时，系统将连接注册数据源，一方面验证数据源的联通性，一方面将获取该数据源的元数据和部分样例数据。系统将对样例数据执行一次敏感数据的初步识别。其步骤如下：

①系统识别获取的样例数据，通过其数据类型（字符、数值等）和数据内容进行敏感数据识别。

②敏感数据识别由敏感数据识别引擎完成；敏感数据识别引擎采用规则、知识库以及自然语言处理中的命名实体识别、特征词提取，特征密度计算等方式进行智能识别。

③如果字段属于长字段，则对该字段进行标记。

④如果字段不属于长字段，但无法进行敏感数据识别，此时系统将对其字段描述进行语义分析和理解，补充相关信息后进行识别。

⑤识别出的字段将存储在敏感字段识别库中。

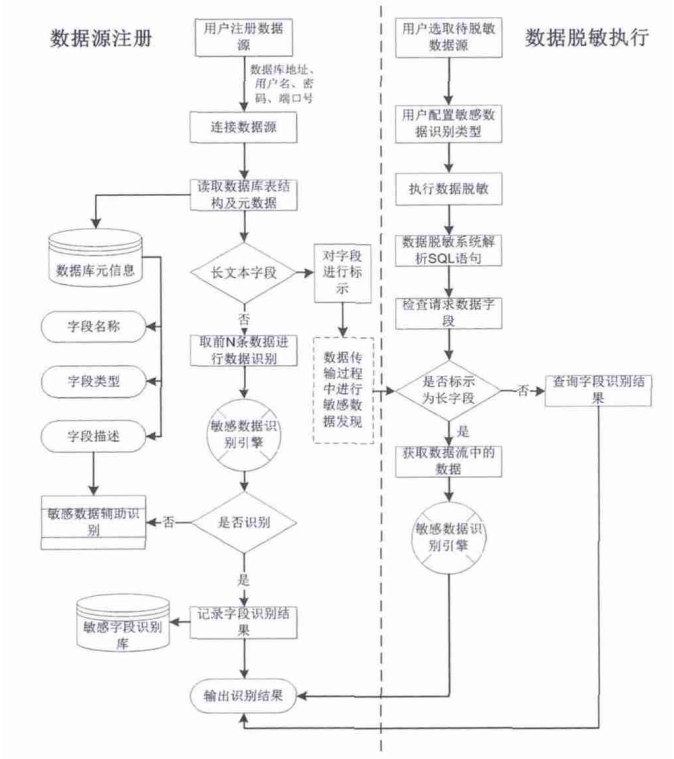

图 4 敏感数据识别方法

（2）数据脱敏任务执行阶段。为提高敏感数据发现以及数据脱敏的效率，在脱敏任务执行阶段，主要对长字段进行识别，步骤如下：

①系统根据用户配置的参数对访问数据库的所有 SQL 语句进行解析，首先在敏感数据字段库中查验哪些属于敏感字段，已识别出的敏感字段按其脱敏策略执行脱敏。

②如果字段为长字段，则获取每一条流经系统的数据，送入敏感数据识别引擎中，作为文本型数据进行识别。文本中可能包含多种敏感数据类型。

③根据识别结果进行脱敏。

3.4　系统主要功能

智能大数据脱敏系统的功能按数据类型划分，主要包括数据库脱敏、文件脱敏、图片及视频脱敏几个主要部分，组成图如图 5 所示。

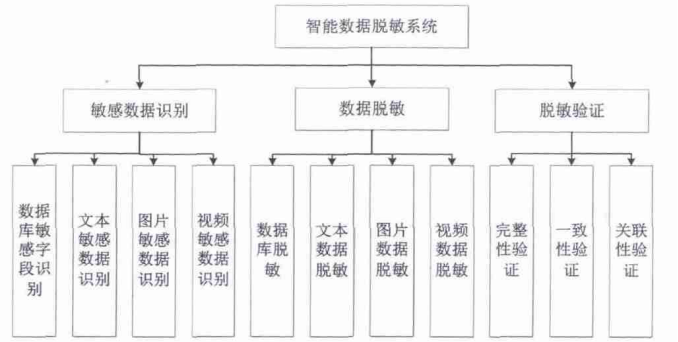

图 5　智能大数据脱敏系统功能组成

敏感数据识别：将针对不同数据的特点，设计敏感数据识别所需的模型、算法、知识库等，以覆盖数据库中敏感字段的识别、文本中敏感数据的识别、图片和视频中的敏感区域识别等；

数据脱敏：将针对不同类型的数据形态，实现不破坏其数据格式和可用性的数据脱敏处理。例如：当对 Word 文件中的数据执行脱敏时，脱敏完成后文件格式依然为 Word。需要注意的是，针对不同的数据类型其脱敏的方式和方法也将会有所不同。

脱敏验证：数据脱敏的本质是通过数据变形来保证对敏感信息的保护，主要目标是安全使用数据。如果脱敏后的数据导致可用性降低或者丧失，将失去数据脱敏的意义。因此，对脱敏后的数据必须在完整性、一致性以及关联性三个方面进行验证。

3.5　服务模式

随着大数据技术的发展和分布式计算技术的成熟，基于大数据平台的脱敏服务为数据安全产品及相关服务设计提供了全新的思路和支撑环境，非常适合数据脱敏这一计算密集、时间敏感型的应用。基于大数据平台的敏感数据智能探测、智能分析与统计、智能处理平台，有望成为数据安全产品的重要发展方向。

按照动态数据脱敏的基本原理和需求，将数据脱敏系统的存储和计算依托大数据平台实现，提供数据脱敏服务 DMaaS（Data Masking as a Service）。它以集中控制和分布代理方式运行，面向政府数据、医疗、教育行业数据和金融数据等，进行按需定制和调用的脱敏服务，如图 6 所示。

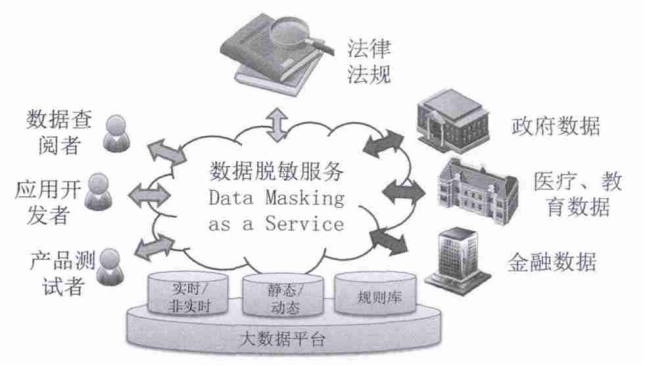

图 6　基于大数据平台的动态数据脱敏服务

基于大数据的数据脱敏平台作为数据拥有者和数据使用者之间的关联途径，承载数据安全隐私保护的重要使命。大数据脱敏平台以集中、松耦合方式进行数据的保护与处理，为企业拥有的敏感和隐私信息提供灵活、实时的服务，不必对应用程序和数据库进行昂贵且耗时的变更，也不会干扰开发、测试及数据使用者履行其各自的职责。

根据应用场景，DMaaS 可以划分为劳务、承包、中转和托管四种应用模式。

劳务模式：SDM 实现方式，按照用户需求将需要脱敏的数据一次性转换完毕，并将结果交付用户。

承包模式：私有化 DDM 实现方式，在用户生产 / 测试环境中搭建 DMaaS，持续运行脱敏功能。

中转模式：公有化 DDM 实现方式。在用户数据环境外搭建 DMaaS，应用程序运行结果在呈现前由脱敏服务处理并交付用户，实现业务流程的灵活调用。

托管模式：公有化 DDM/ 数据仓库实现方式。用户的所有敏感数据存放在 DMaaS 中，业务需要访问数据时调用脱敏服务处理后提交至用户。这种模式有利于数据的集中监管和高强度隐私保护。

4　结　语

数据脱敏是大数据时代企业数据化运行治理的必要安全机制，未来数据脱敏发展的趋势包括精确理解用户需求、更细的粒度、更高的精确度和可用度、更佳的自动化程度、更好的抗破解能力、更强的扩展能力和更友好的方式呈现等，从而满足未来用户多领域的数据交互、共享和融合需求。

参考文献：

[1]　Ponemon Institute.Cost of Data Breach:Global Analysis [EB/OL].(2013-05-28)[2016-05-26].http://www.pone

mon.org/,2015.

[2] Gartner.Gartner 2014 Magic Quadrant Data Masking Report[EB/OL].(2015-12-22)[2016-05-23].http://www.gartner.com.2014.

[3] 姜日敏 . 电信运营商数据脱敏系统建设方案探讨 [J]. 信息科技 ,2014(08):132-133.
JIANG Ri-min.Data Masking System Construction Plans of Telecommunication Operator[J].Information Technology,2014(08):132-133.

[4] 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(11):2-8.
LIU Ming-hui,ZHANG Ni,ZHANG Yun-yong,et al.Research on Sensitive Data Protection Technology on Cloud Computing[J].Telecommunication Science,2014(11):2-8.

[5] Securosis Corporations.Understanding and Selecting Data Masking Solutions:Creating Secure and Useful Data[EB/OL].(2014-03-01)[2016-05-19].http://www.techrepublic.com/resource-library/whitepapers/understanding-and-selecting-data-masking-solutions-creating-secure-and-useful-data/.

[6] Informatica Corporation.Dynamic Data Masking Baseline Deployment[EB/OL].(2013-01-01)[2016-05-22].https://www.informatica.com,2013.

作者简介：

陈天莹（1982—），女，博士，高级工程师，主要研究方向为大数据、信息安全；

陈剑锋（1983—），男，博士，高级工程师，主要研究方向为信息安全、云计算。