

数据脱敏系统白皮书



OLYM 奥联

民族密码 奥联智造

深圳市奥联信息安全技术有限公司

目 录

术语.....	1
1. 前言.....	2
2. 产品简介.....	3
3. 主要功能.....	4
4. 产品特点及关键技术.....	4
5. 技术指标.....	5
5.1. 中间件接口.....	5
5.2. 密码算法.....	5
5.3. 脱敏运算速度指标.....	6
6. 实现效果.....	7

OLYM[®]奥联

术语

- **数据脱敏**：指对某些敏感信息通过脱敏规则进行数据的变形，实现敏感隐私数据的可靠保护。本文数据脱敏是由密钥控制的 reversible 操作，即使用正确密钥可以还原明文。
- **格式保留加密**：是对称加密算法的一种工作模式。明文具有特定的格式，在密钥的控制下计算出的密文保持明文的格式。
- **标识密码**：是采用标识作为公钥的一种非对称密码体制。中国的标识密码算法标准 SM9 支持标识加密、标识签名和密钥交换。

OLYM[®] 奥联

1. 前言

最近几年敏感信息泄露事件频发，网络攻击与漏洞利用正在向批量化、规模化方向发展。根据统计，2017年和2018年均有超过50亿条数据泄露。2018年，有12起数据泄露事件涉及人数超过1亿。2018年最大的数据泄露事件是Aadhaar印度国家身份认证系统泄露事件，涉及11亿印度个人信息，包括国家公民身份号码、姓名、电话号码、电子邮箱、住址及照片等。

信息泄露事件和企业对信息安全的重视程度、安全技术手段的实施完善能力等多个方面相关。伴随信息泄密的规模和影响越来越大，各国都在制定相关法律法规，加大对信息泄露责任主体的处罚力度。中国制定了一系列的行业和国家标准规范并通过立法来规范和认定责任主体，明确处罚力度等。比如：证监会在2013年发布《证券期货业信息系统运维管理规范》明确规定在线数据和离线数据应用于非生产环境时必须进行脱敏；银监会2011年发布《中国银行业“十二五”信息科技发展规则监管指导意见》和《商业银行信息科技风险现场检查指南》，要求银行业完善敏感信息存储和传输等高风险环节的控制措施，对用于测试的生产数据应对相应数据进行脱敏、变形处理，严格防止敏感数据泄露；在政务领域国家标准《信息安全技术 政务信息共享 数据安全技术要求》对政务数据脱敏提出明确要求；《网络信息安全法》明确规定了信息安全的责任主体和法律责任。在欧洲随着GDPR的实施，对个人敏感信息的保护提到了空前的高度。英航因2018年泄露了50万用户的隐私，被处以2.3亿美元罚款，万豪集团也因为2018年11月泄露了3.4亿条酒店住店记录被处1.23亿美元罚款。伴随着法律、法规的实施，企业和机构必将提高对信息安全的重视程度，需要实施真正有效的数据安全保护技术。

数据脱敏技术一般指对某些敏感信息通过脱敏规则进行数据的变形，实现敏感隐私数据的可靠保护，这样就可以在开发、测试和其它非生产环境以及外包环境中安全地使用脱敏后的真实数据集。众多数据脱敏方案也是针对非生产环境的数据脱敏。实际信息泄露事件中大多是攻击者利用系统漏洞攻击生产系统，对数据库拖库，导致生产数据规模泄露。因此仅解决非生产环境中敏感数据的安全性

问题，并不能显著降低数据风险。对于生产系统的数据，数据的可用性是必须保证的，即数据脱敏的过程必须可逆，并且要保证数据从脱敏状态还原的安全性、可控性和高效性。这需要安全性严格的高效、可逆数据脱敏技术和解决方案。

2. 产品简介

奥联数据脱敏系统支持静态/动态数据脱敏，在线/离线数据脱敏。数据脱敏后保持原有数据的格式，支持脱敏数据的在线受控还原，支持生成系统对数据的实时在线处理。支持数据脱敏加密密钥的分隔，允许系统和用户各自拥有加密密钥的分隔向量（分别称为系统密钥和用户密钥），只有用户和系统共同允许才可对相关数据进行脱敏和还原。支持用户密钥采用策略管理，基于策略标识进行基于加密技术的授权访问控制。

奥联数据脱敏系统由数据加密机、安全中间件、标识管理系统构成。其中数据加密机实现对敏感数据的格式保留加密、解密能力、标识密码运算；安全中间件支持灵活的安全策略和批量数据脱敏和还原处理；标识管理系统支持用户的标识管理和标识与策略密钥的分发。如果用户无需对系统中的敏感数据进行策略控制时，则系统无需标识管理系统。

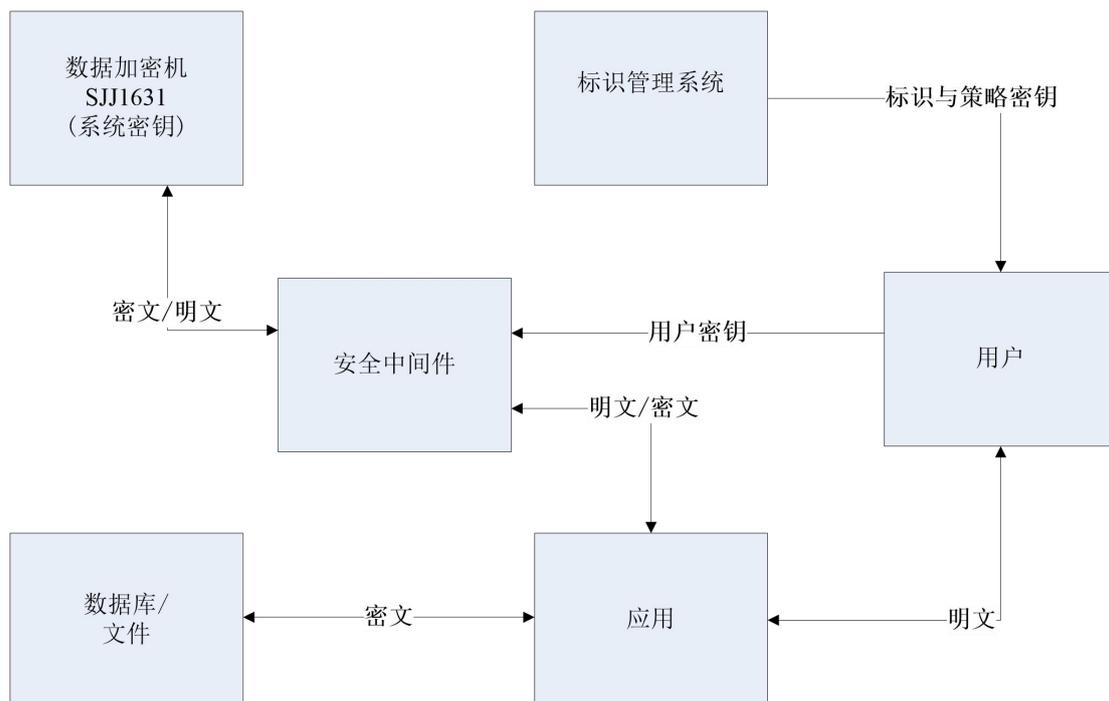


图 1 系统架构图

3. 主要功能

奥联数据脱敏系统具有以下主要功能：

- 数据脱敏：采用加密技术对敏感数据进行加密，加密密文与明文格式相同。
- 数据还原：对脱敏后的数据进行受控解密还原以恢复原始数据，进行使用。
- 操作审计：可对数据脱敏、还原操作进行审计。
- 脱敏密钥分布式管理：支持按用户角色、属性控制数据脱敏过程的加密密钥，以实现按策略对敏感数据的访问控制。

4. 产品特点及关键技术

奥联数据脱敏系统采用两项关键技术：

- 格式保留加密技术：奥联数据脱敏系统支持 NIST.SP.800-38G Recommendation for Block Cipher Modes of Operation:Methods for Format-Preserving Encryption (2019 年更新版本) 中的格式保留加密算法 FF1 和 FF3，以及自有格式保留加密算法，实现对多种数据的格式保留加密功能。
- 标识密码技术：标识密码系统中，用户的标识就可以用做用户的公钥。在这种情况下，用户不需要申请和交换证书，而是直接使用标识进行密码运算，从而解决公钥真实性问题，极大地简化了密钥系统管理的复杂性。这类系统中用户的私钥由系统中的一受信任的第三方（密钥生成中心：KGC）使用标识私钥生成算法计算生成。这样的系统具有天然的密码委托功能，特别适合于对数据恢复有需求的应用环境。应用系统涉及的标识只要具有唯一性特征就可以直接作为公钥进行密码应用。这些标识可以具有任意的格式、属性，例如可以是常用的用户标识如邮件地址、

手机号码、身份证号码，还可以是数据安全控制策略的形式化编码、标识属性集等，特别适合基于标识、基于属性的权限访问控制和数据安全保护。奥联数据脱敏系统支持基于国密SM9算法的属性加密实现对数据加密密钥的授权访问控制。

奥联数据脱敏系统具有以下三个特点：

- 支持多种数据格式的格式保留加密：支持 IP 地址、日期、时间戳、身份证号码、手机号码、固定电话号码、UTF8、GBK、UNICODE、邮件地址、英语字母、英文与数字组合、可打印字符、32 位整数、64 位整数、浮点数、纯中文 UTF8/GBK/UNICODE、中英文混合等 25 种常用数据格式以及自编码格式的加密，加密后密文保持原有格式，系统还支持灵活的自定义数据格式。
- 支持批量数据的高效处理：中间件支持一次调用处理多达 4G 条记录的批量处理（内存允许情况下）。处理支持数据库查询的游标处理，实现应用的无缝调用。批量数据处理高效，单机每秒可处理 1 千万条 IP 地址、日期、时间戳、身份证号码、手机号码、固定电话号码、姓名等常用类型记录。
- 安全性高：采用标准算法 AES、SM4 和安全的格式保留加密模式 FF1、FF3 组合。支持对中间件与密码机间传递的数据进行加密保护，防止内部人员通过网络截取敏感数据。

5. 技术指标

5.1. 中间件接口

中间件支持 C、Java 等语言。

5.2. 密码算法

- 对称密码算法支持：AES、SM4
- 非对称密钥算法支持：SM2、SM9、ECDSA、RSA

- 摘要算法支持：SM3、SHA256、SHA3
- 格式保留加密模式：FF1、FF3、自有模式

5.3. 脱敏运算速度指标

- 运算速度

编号	数据格式(编码)	加解密记录/秒	备注
1	IPv4 地址(3)	XXX/XXX	格式：192.168.1.23
2	IPv6 地址(4)		暂不直接支持,可使用自编码方式 10
3	日期(5)		格式：1981-01-01, 可用于保护生日等
4	数字串(6)		不定长数字串, 从 1 到 10000 个数字。
5	手机号码(6)		11 位数字
6	固定电话号码(6)		8 位数字或 4 位区号+8 位数字
7	英语字母串(7)		IA5String、纯英文字符集
8	字母和数字串(8)		IA5String、英文字符+数字集等
9	可打印字符集(9)		ASCII 表中可打印字符集
10	自编码(10)		自编码的任意格式
11	中英文数字 GKB(11)		支持不定长中文串,性能测试为 8 个中文或中、英、数字等组合
12	中英文数字 UNCODE(12)		
13	中英文数字 UTF8(13)		
14	时间戳(14)		格式：1981-01-01 01:01:01, 可用于保护时序信息
15	身份证号码(15)		格式：18 位数字或者 17 位数字+X
16	小写邮件地址(16)		格式：符合 RFC 5322, 小写邮件地址
17	邮件地址(17)		格式：符合 RFC 5322, 地址大小写敏感
18	纯中文 GKB(18)		支持不定长纯中文串
19	纯中文 UNCODE(19)		
20	纯中文 UTF8(20)		
21	32 位整数(21)		32 位整数
22	64 位整数(22)		64 位整数
23	单精度浮点数(27)		IEEE 754 的单精度浮点数,

			不支持无限大
24	双精度浮点数(28)		IEEE 754 的双精度浮点数，不支持无限大
25	单精度浮点数(29)		IEEE 754 的单精度浮点数，不支持无限大、非正常小
26	双精度浮点数(30)		IEEE 754 的双精度浮点数，不支持无限大、非正常小

6. 实现效果

用户校验

手机号 校验

明文数据

选择	手机号	身份证号
<input type="checkbox"/>	13205883799	481511199801065008
<input type="checkbox"/>	13115093899	461511199601067552
<input type="checkbox"/>	13605493466	45151119950106520X
<input type="checkbox"/>	13635414782	401511199301065215
<input type="checkbox"/>	13004158775	421511199801067415

密文数据

<input type="checkbox"/>	27183693754	592578266666768669
<input type="checkbox"/>	88815606628	606375998067018785
<input type="checkbox"/>	24464107894	67509188127147678X
<input type="checkbox"/>	37672181615	591660139500019639
<input type="checkbox"/>	67086380537	993092468210550856