淘宝网 开放平台 Top
open.taobao.com

数据存储加密接入课程

数据加密产品介绍

常春

2017/05/03

课程提纲

- ✓ 为什么要做数据存储加密？
- ✓ 数据存储加密的技术概要和主要功能
- ✓ 对给定的业务场景，怎样选择适用的方案？

淘宝网 开放平台 Top
open.taobao.com

1

学习目的

学习目的

- ✓ 了解数据存储加密的目标效果
- ✓ 了解数据加密产品的主要功能和产品结构
- ✓ 能根据自身应用的特点确定加密接入方案

淘宝网 开放平台 Top
open.taobao.com

2

前置说明

前置说明

- ✓ 本课程适合开发、测试人员学习

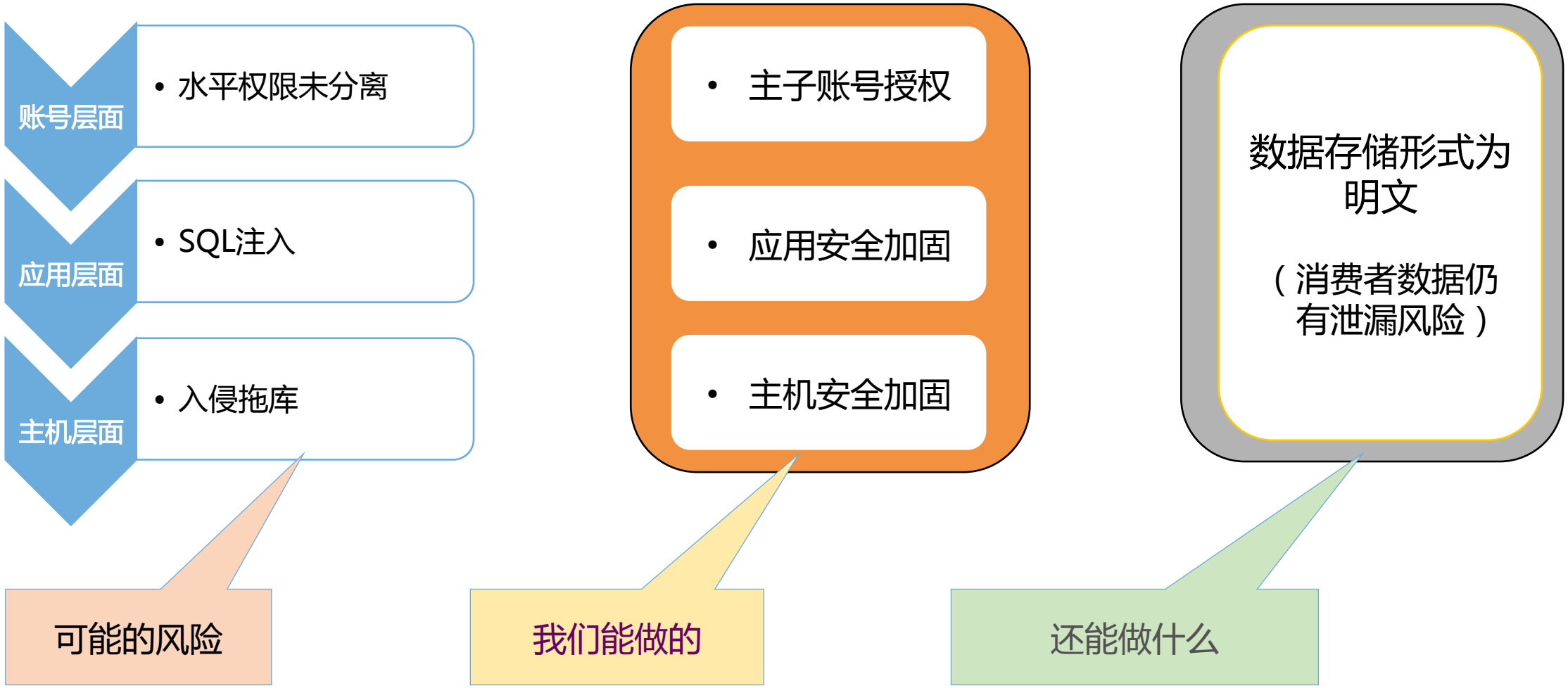
淘宝网 开放平台 Top
open.taobao.com

3

什么是数据存储加密？

服务商安全现状

消费者数据泄漏 - 直接造成商家对服务商的信任危机与公关危机



数据存储加密 – 安全方案介绍

数据存储加密方案致力于解决上述的安全问题：

- **基础安全保障**

加解密从根本上夯实了数据安全性。对敏感字段加密后，可以有效防止数据库内容被直接盗取。且密钥以用户维度隔离，有效解决应用的水平权限隔离问题。

- **数据库内容与密钥存储分离**

ISV的ECS和RDS中只存储加密数据，不保存密钥。只需接入我们提供的SDK，密钥在运行过程中由SDK动态向TOP发起请求，秘钥周期由SDK维护。保证秘钥不落地，只存储在内存，对ISV透明。

- **无缝集成于TOP开放场景、全程加密**

加密服务无缝集成在TOP的订单服务中，一旦开启加密后，从TOP API返回、RDS推送内容都会变成密文，保证数据从源头开始，在传输、存储整个链路都是以密文形式。

数据存储加密 – 消费者敏感字段示意

隐私字段	隐私示例
姓名	涉及消费者‘姓名’的字段，包括但不限于：receiver_name、buyer_name、name、consignee_name、gui_name、relation_name、apply_name、trip_person_name、passenger_name、user_real_name、phone_owner、fetch_ticket_name、rider_name、full_name、traveler_name、contact_name、guest_name等。
Nick	涉及消费者‘Nick’的字段，包括但不限于：buyer_nick、user_nick、nick、owner_nick等。
手机号	涉及消费者‘手机号’的字段，包括但不限于：receiver_phone、mobile、rec_mobile、relation_mobile、travel_contact_mobile、telephone（部分API的手机号）、mobile_bak等。
座机号	涉及消费者‘座机号’的字段，包括但不限于：receiver_phone、contact_phone、phone_no、phone、relation_phone_bak、contact_other等。
身份证号	涉及消费者‘身份证号’的字段，包括但不限于：hk_card_code、cert_card_num、id_number、credentials_code、fetch_cert_number、cert_no、passenger_cert_no等。
支付宝账号	涉及消费者‘支付宝账号ID或账号名’的字段，包括但不限于：alipay_id、pay_account、buyer_alipay_id、account_no、buyer_alipay_id等。
Email	涉及消费者‘Email’的字段，包括但不限于：relation_email、email、buyer_email、travel_contact_mail、contact_email等。
车牌号	涉及消费者‘车牌号’的字段，包括但不限于：et_plate_number。
其他用户ID	涉及消费者‘其他的用户身份ID’的字段，包括但不限于：user_ID、owner_id、buyer_id、taobao_user_id、trip_card_no、trip_person_no等。

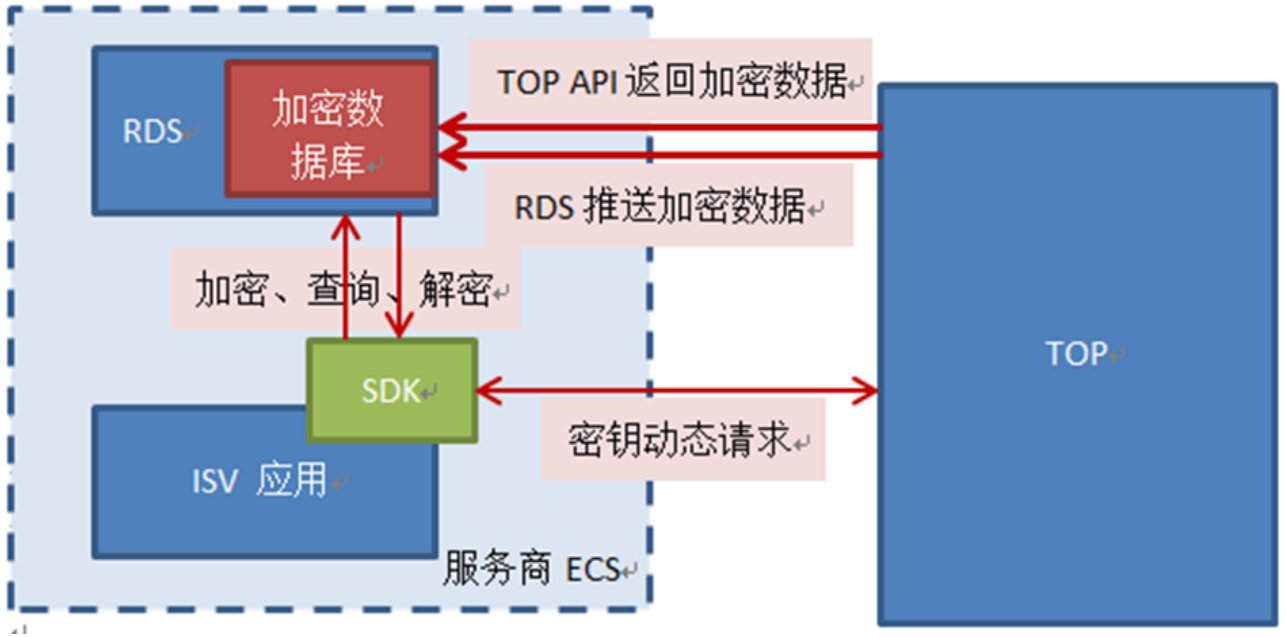
淘宝网 开放平台 Top
open.taobao.com

4

数据加密产品介绍

数据存储加密– 产品说明

解决方案：

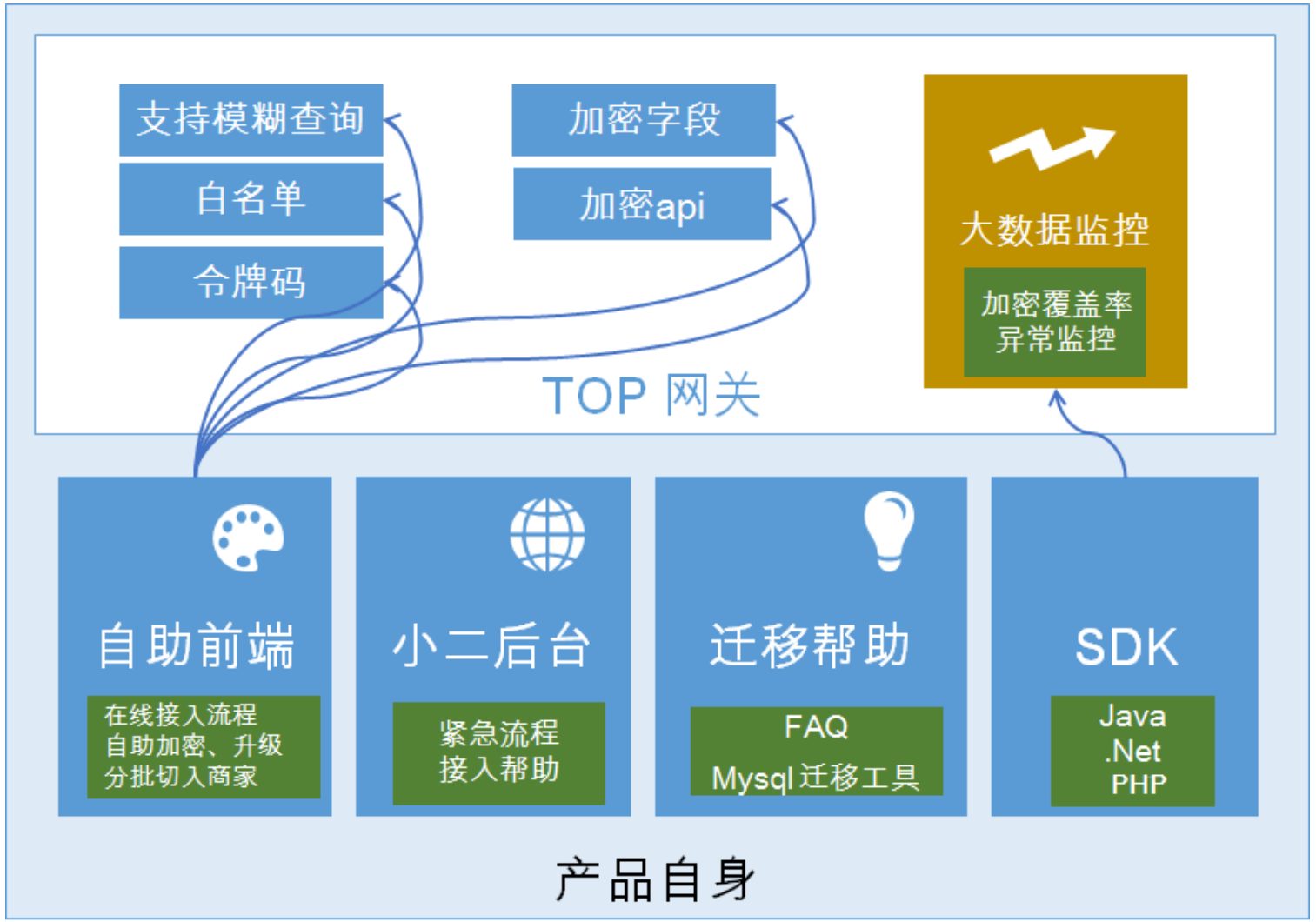


使用我们的加密服务后，从开放平台输出的数据从流出源头开始就是加密的。这包括含敏感字段的TOP api，以及包含订单的RDS系统库。
服务商在自己的数据库(系统推送库和其他数据库)中以密文方式储存敏感信息。

在应用中需要对敏感信息进行检索、展示等操作时，调用SDK中提供的解密函数进行解密。SDK会自动从TOP服务器获取对应的密钥。此外，密钥的网络请求、管理和本地缓存也会自动由SDK完成。

ISV需要根据自己的应用业务逻辑改造代码、接入我们的SDK，通过我们提供的SDK中的加密、解密等api完成敏感数据的加解密转换。

数据存储加密– 架构说明



除了加密服务以外，为了保障ISV能够顺利接入，以及后续能够有效运营其功能、监控服务的健康运行状态，我们的产品提供了以加密为核心的一系列配套服务。我们在开放平台的控制台提供了**自助接入平台，**让ISV可以逐步顺利接入我们的各种语言SDK。对于接入的重点环节，数据迁移，我们提供了迁移工具。

接入加密后，我们后台利用大数据分析手段对SDK的调用行为进行建模、计算加密覆盖率，并对异常行为监控、预警。

数据存储加密 – 技术特点

作为集成于TOP的一款加密产品，我们的加密服务有如下优势：

- **历经优化的高效SDK：**

秘钥是以appkey+session+版本号作为缓存key，对密钥请求做了本地缓存、异步更新等各种优化，在缓存有效期之内不会发起top请求。加解密等操作的性能开销也极小，几乎可以忽略不计。

- **多种加密方案：普通加密方式　vs 支持模糊查询的加密方式**

我们的加密服务支持对长文本（如昵称、姓名）的片段进行模糊查询。分组规则是4位为一个检索块。

例如：nick：test123 拆分成test\est1\st12\t123 4个检索块。~完整密文~dp3dmeY/F4QAoKgSfcSp6QZQq00LU6RQphEvoPZzh1Ec~版本号~~

- **普适性**

我们的加密解决方案不依赖硬件，我们的SDK集成在现有的TOP SDK中，DB字段长度

淘宝网 开放平台 Top
open.taobao.com

5

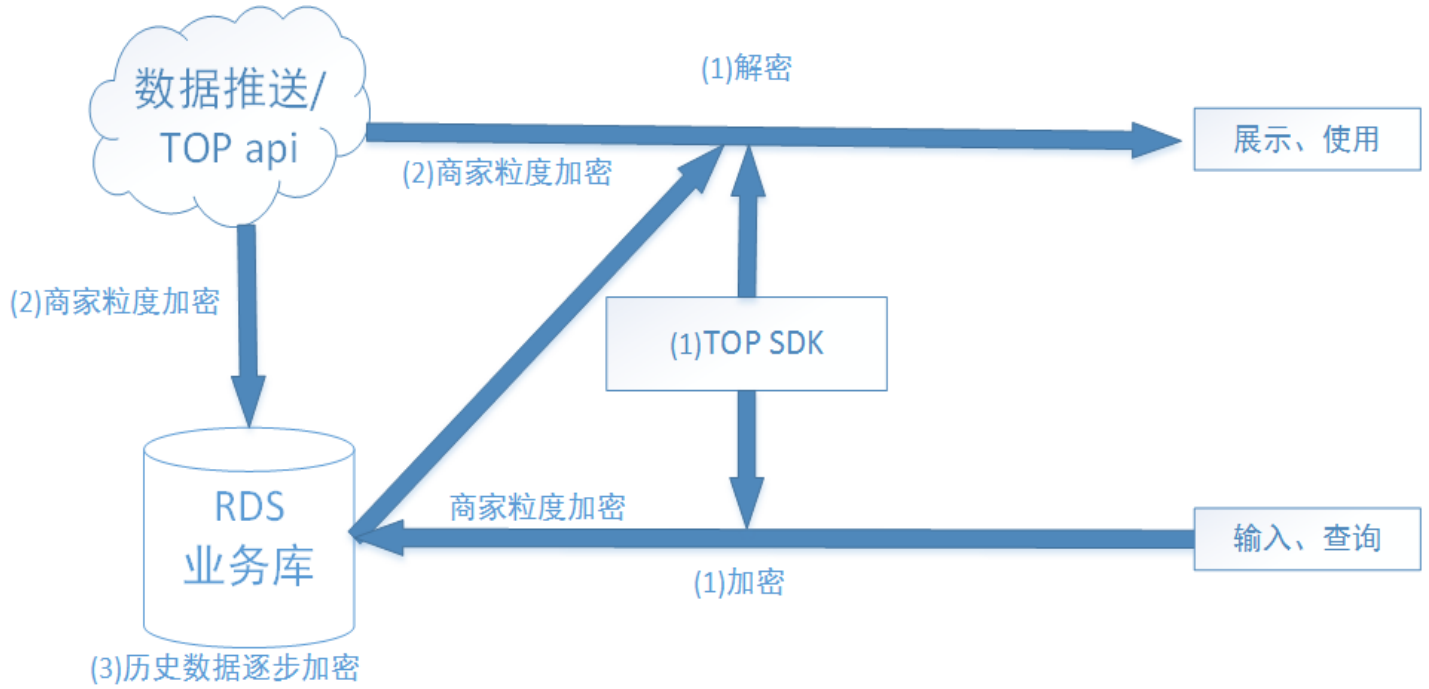
典型场景接入方案

接入数据加密——选择账号方案

在开始加密后，首先会需要选择加密使用的账号方案。

对不同的账号体系我们的加密方案会有所不同。

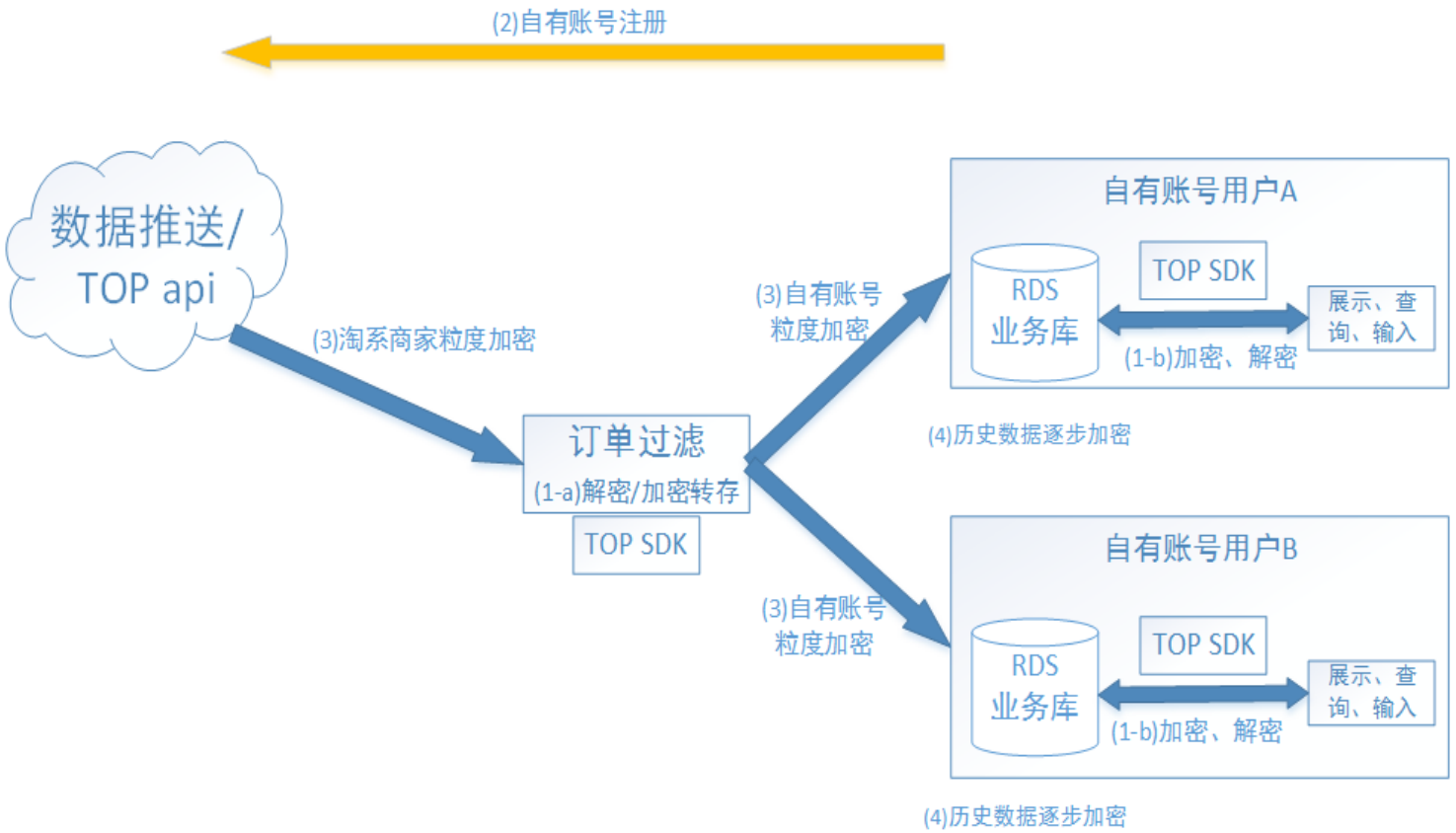
接入数据加密——淘系账号体系



本产品的密钥是商家粒度隔离的，请求加密、解密的时候需要带上用户的授权session。

因此对于淘系账号应用，加密、解密的时候，直接取对应商家的session并且调用sdk就可以了。

接入数据加密——自有账号体系



自有账号体系的应用，通常在应用逻辑中，有一个订单过滤的步骤。根据淘系账号和自有账号映射，将订单信息转存到相应的自有账号下面。淘系账号和自有账号的映射可能是多对一甚至多对多的。

对自有账号的应用，我们可以为自有账号提供加密、解密服务。首先，通过调用一个api注册自有账号，之后这个账号就可以和淘系账号一样被分配到一个自己的密钥了。

在接入加密的过程中，可以选择在订单过滤的地方，将原本用淘系账号对应密钥加密的数据，转而用对应的自有账号密钥加密，并存储在相应的数据库中。

接入数据加密——字段加密检索方式

帮助文档链接：http://open.taobao.com/docs/doc.htm?treeId=1&articleId=106213&docType=1

本产品提供两种加密方式。

默认是普通的加密方式，这种方式加密、存入数据库后，可以精确查询。但是如果有模糊查询的需求(如SQL的like 语句)，就暂时无法满足了。

因此我们还提供一种可以模糊查询的加密方式。但是使用这种方式也有一定代价：

- 支持模糊查询加密方式，产出的**密文比较长**；
- 支持的模糊查询**子句长度必须大于等于4个英文/数字，或者2个汉字**。不支持过短的查询(出于安全考虑)；
- 返回的结果列表中有**可能有多余的结果，需要增加筛选的逻辑**：对记录先解密，再筛选；

如果需要选择模糊查询的加密方式，就在对应字段前面打勾。**推荐用默认的普通的加密方式。**

接入数据加密——分销API加密方案

分销API方案：

分销API分销商和供应商都可以获取同一份订单数据，在rds订单同步中可能是以分销商维度推送、也可能是以供应商维度推送。这样无法确定在RDS中的数据是以那个商家维度加密的。

故分销接口采用app级别秘钥，多用户共享一个秘钥。

涉及分销API：

taobao.fenxiao.orders.get;

taobao.fenxiao.trademonitor.get;

使用手册：java、c#、php sdk升级至最新版本。

兼容场景：若数据源来自分销接口（taobao.fenxiao.orders.get，taobao.fenxiao.trademonitor.get）

加密：调用securityClient.encrypt(String data, String type);

解密：调用securityClient.decrypt(String data, String type);

和securityClient.decrypt(String data, String type, String session)方法进行解密。

注：分销数据加密仅能使用securityClient.encrypt(String data, String type)方式。

分销数据解密可以使用上述两个解密方法（为适配已经改造的用户）。

接入数据加密——Number字段类型加密方案

Number字段类型加密方案：

目前密文通过AES加密之后变成一串Base64的字符串（String），与之前的number类型不匹配。这种情况下SDK转换会报错，而且可能跟db存储的类型不匹配。
例如：taobao.trade.fullinfo.get中alipay_id加密字段是Number类型。包含数字类型的API

加密逻辑变更：

阶段1：加密请求，在返回密文的同时也返回明文，用于开发者做切换（新增一个返回字段 encrypt_XXX ）
当前是加密请求：{"alipay_id"=123, "encrypt_alipay_id"=加密(123)}
当前是不加密请求：{"alipay_id"=123, "encrypt_alipay_id"="123"}

阶段2：加密请求，不返回Number类型明文字段（相当于明文字段作废）。
当前是加密请求：{"alipay_id"=null, "encrypt_alipay_id"=加密(123)}
当前是不加密请求：{"alipay_id"=123, "encrypt_alipay_id"="123"}

阶段1：会为加密数字类型增加一个 encrypt_XX 的String的密文出参。 ISV需要切换到这个 encrypt开头的字段中。
阶段2：将不会再返回 Number类型的明文字段 ，只使用encrypt开头的字段。

ISV需要做的：
1）下载新版本SDK。
2）判断下对应的DB字段（alipay_id等）是否数字类型， 如果是数字类型：
需要新增string类型字段存储，老数据迁移到新字段。 如果DB已是string类型的则忽略。
3）取encrypt_alipay_id字段做为 alipay_id 存储。

接入数据加密——自研SDK方案

自研的SDK方案：

建议：加密解密实现参考其他语言版本的SDK(java/.NET/php)

1. 加密算法要求：AES/CBC/PKCS5Padding。
2. 获取密钥的缓存是否只应该放在内存中，在服务启动后从接口拉取。
禁止存放在数据库、硬盘文件、OSS等持久化存储的地方。

3. 模糊查询部分功能，如不使用，可不实现。

4. 设置密钥的过期时间，过期重新拉取。
秘钥的过期时间，在获取秘钥的接口会返回的，控制权在于top服务端。
秘钥获取接口 ：taobao.top.secret.get .

5. 请回传密钥的加密、解密调用次数。
具体标准：加、解密调用函数每次调用，对应的计数器(各种类型计数器)会+1，5分钟左右同步一次。
异步线程会把计数器同步到top api接口： taobao.top.sdk.feedback.upload.

淘宝网 开放平台 Top
open.taobao.com

6

课程总结

课程总结

- ✓ 接入数据加密可以显著的降低数据泄露风险，提升应用安全。
- ✓ TOP提供的加解密SDK经过了效率优化，并且能适应多种需求和场景。接入数据加密需要ISV根据自己的业务场景对应用进行相应改造。
- ✓ ISV可以通过开放平台的自助接入平台逐步接入数据加密。

淘宝网 开放平台 Top
open.taobao.com

成就你的技术梦想

旺旺群号：1640022674

Thank you