dschain.github.io

DSChain-数据服务区块链

DSChain-数据服务区块链

Data Services Chain

一、简介

DSChain数据服务链从源头入手对数据服务行业上下游产业链实现个人信息脱敏传输、脱敏存储和信息来源与流向追溯。现行系统信息传输和存储方式容易造成信息泄漏、非法留存、来源不明和去向流不明等潜在风险；DSChain使用成熟加密算法将数据源与商户之间的信息通过一户一密的脱敏加密方式进行传输和存储，数据源与商户的供应链体系与脱敏密钥管理采用区块链技术实现。

二、大数据服务背景

1.名词解释

个人信息：是指以电子或者其他方式记录的能够单独或者与其他信息结合识别自然人个人身份的各种信息，包括但不限于自然人的姓名、出生日期、身份证件号码、个人生物识别信息、住址、电话号码等。
非个人信息：是指经过处理无法识别特定个人且不能复原的个人信息。
数据服务：是指依托大数据资源管理与分析的相关服务产业，包括数据交易服务、数据采集服务、数据加工、数据应用服务、数据增值服务等。本方案主要指涉及个人信息的数据服务、征信数据等行业应用。
数据源（包括征信公司）：合法合规拥有个人信息采集、存储、加工和对外服务公司或政府部门指定或审批单位。
商户：是指因业务需求，在取得用户授权前提下，向数据源查询自有用户个人信息的公司，包括：银行、保险、互联网公司等依法经营的公司。
中间商：是指介于数据源与商户之间赚取差价的数据服务公司。
直链模式：商户与数据源直接达成合作，进行商务和技术对接。
间连模式：商户与中间商达成合作、中间商与数据源达成合作，间接实现商户与数据源之间信息技术对接，但商户与数据源之间相互不清楚。
明文：是指个人信息未经任何加密。如手机号13012345678
加密密钥：是特指机构间个人信息字段的加密密钥；机构间通讯时先进行个人信息字段加密，再经通讯加密。

2.行业背景

个人信息安全的严峻形势

2011年12月，CSDN网站600万用户信息泄露事件;
2012年5月，遭“3.15”晚会曝光后，罗维邓白氏中国总部宣布关闭;
2014年1月，2000万酒店开房数据泄漏;
2014年3月，携程被爆涉嫌储存用户信用卡支付信息(卡号和CVV2码等)，开有重大技术漏洞;
2014年12月，13万个12306用户信息遭泄露，其中包括账号、明文密码等;
2015年4月，部分社保数据被泄露，涉及13个省市，3000多万参保用户;
2016年6月，山东临沂，大学新生信息被泄露;
2017年2月，公安部成立由刑侦局牵头乩“2.17”与案组，破获侵犯个人信息重大案件;
2017年3月，京东前离职员工盗取超50万用户数据;
2017年5月，15家大数据公司被介入调查，其中几家公司估值达几十亿元，甚至超百亿元;
2017年8月，徐玉玉被电信诈骗致死案被审理宣判; 。。。。。
2019年9月，“净网2019”专项行动捣毁了一批知名品牌数据公司，因经营个人信息数据涉及违法案件被抓捕和调查。

个人信息保护政策指导

自2017年，国家对网络个人信息安全高度关注，先后颁布了：网络安全法、两高法释、个人信息安全规范等相关法律法规，野蛮发展的大数据服务行业将进入合规整治时期，很多公司再次面临生存危机。

三、传统信息传输与存储

数据服务行业中涉及到个人信息在商户、中间商、数据源间多方传输与存储，传统的信息传输和存储方式容易造成信息泄漏、中间商非法留存、信息来源不明和信息流向不明等潜在风险。

1.传统信息传输方式

明文传输

个人信息明文在商户、中间商和数据源之间传输。

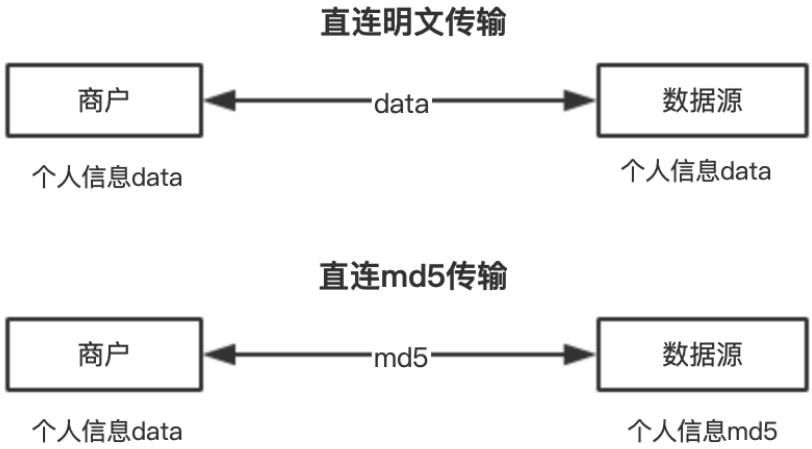
MD5脱敏传输

先将个人信息经哈希算法MD5之后，在商户、中间商和数据源之间传输。

传输流程

直连模式

a.交互流程图



b.各参与方可获取的个人信息情况

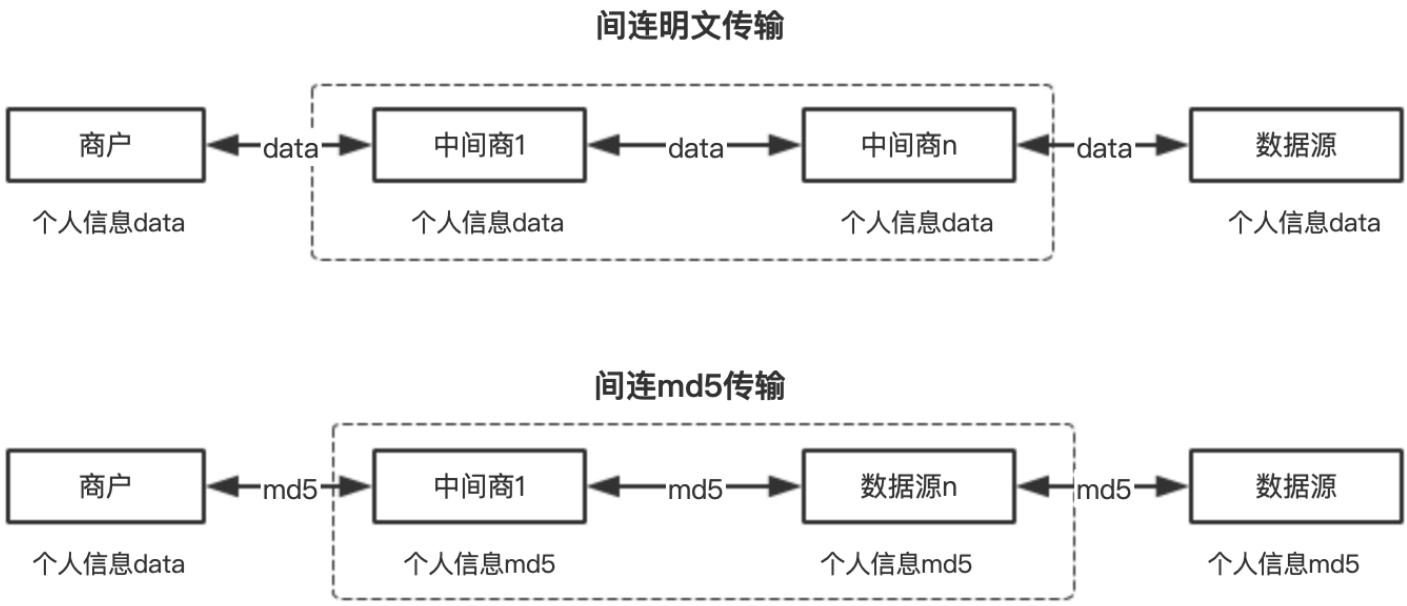
传输方式	机构	可获取个人信息	风险
明文	商户	明文	数据泄漏
-	数据源	明文	数据泄漏
MD5	商户	明文	数据泄漏
-	数据源	2种情况：仅拥有md5、 同时拥有明文和md5	数据泄漏 （md5可撞库）

c.信息流向问题

直连模式商户与数据源直接达成合作，信息来源和信息的使用场景十分明确。

间连模式

a.交互流程图



b.各参与方可获取的个人信息情况

传输方式	机构	可获取个人信息	风险
明文	商户	明文	数据泄漏
-	数据源	明文	数据泄漏
-	中间商	明文	数据留存、数据泄漏、非法获利（数据买卖）、合法获利（精准营销）
MD5	商户	明文	数据泄漏
-	数据源	两种情况： 仅拥有md5、 同时拥有明文和md5	数据泄漏（md5可撞库）
-	中间商	md5	数据留存、数据泄漏（md5可撞库）、非法获利（数据买卖，md5可撞库）、合法获利（精准营销，md5可撞库）

c.信息流向问题

间连模式因多层中间商参与，商户不清楚数据来源，数据源不清楚商户使用场景，双方面临合规风险。

2.传统信息存储方式

明文存储

商户、数据源等公司将个人信息明文存入数据库。

密钥加密存储

商户、数据源等公司自行生成和管理加密密钥或购买硬件加密机，将自有个人信息加密后再存入数据库。

存储方式风险

存储方式	风险
明文	数据泄漏
密文	数据泄漏、密钥泄漏

3.传统信息传输与存储风险原因分析

个人信息泄露和非法买卖等非法商业趋利性行为，虽能通过法律法规规范合法经营企业单位，但无法杜绝非法单位或非法个人进行非法牟利行为。
个人信息泄露是因技术传输方式引起的，随着互联网的发展，技术的缺陷频繁暴露出来。DSChain区块链提供了其中一种完备的技术解决方案。

四、DSChain脱敏传输和存储方式

DSChain从供应链数据体系入手，将个人信息依靠成熟的加密算法，对个人信息进行去标识化脱敏，实现商户、中间商和数据源之间使用非个人信息传输，并在供应链体系下进行非个人信息存储。

1.个人信息脱敏描述

数据源为其下游每一个商户生成各自独立的加密密钥，商户发送交易前，先将个人信息字段通过加密密钥加密，再直接或间接提交到数据源。
个人信息脱敏算法分为两种，非对称加密和对称加密。

2.非对称加密算法脱敏传输

公私钥生成

商户直接或间接（通过中间商）将商户基本信息（名称、地址、营业执照、使用场景、特殊行业牌照等信息）提交至DSChain区块链；数据源从DSChain区块链获取商户基本信息，为其生成独立公密钥对{private_key:public_key}，并将商户pubilic_key发布到DSChain区块链上。

公钥分发

商户或中间商通过DSChain区块链获取商户公钥public_key；

脱敏传输

数据交易时，商户使用public_key将个人信息data加密成public_key(data)，直接或间接（通过中间商）提交至于数据源，数据源通过商户名找到对应private_key,解密取得data，完成业务处理。

3.对称加密算法脱敏传输

密钥生成

商户需具备CA证书，无CA证书的，需向有关单位申请。
商户将CA证书和基本信息（名称、地址、营业执照、使用场景、特殊行业牌照等信息）直接或间接（通过中间商）发布到DSChain区块联盟链。

数据源从DSChain区块链获取商户CA证书和基本信息，并为商户生成对称加密算法密钥key。
数据源使用商户CA证书公钥，将密钥key加密成CA(key)，并将CA(key)发布到DSChain区块链。

密钥分发

商户或中间商通过DSChain区块链获取商户公钥CA（key）；商户通过自由CA私钥解密CA(key)获取加密密钥key。

脱敏传输

数据交易时，商户使用key将个人信息data加密成key(data)，直接或间接（通过中间商）提交至于数据源，数据源通过商户名找到对应key,解密获取data，完成业务处理。

4.脱敏存储

从商户业务使用角度出发，商户存储个人信息明文data，不存储key(data),主要因为自身业务中需要建立和维护用户画像和调用外部数据接口,若不存储明文data将无法进行正常业务。
在实现供应链数据体系后，商户仅存储key(data)就能与外部服务商交互。

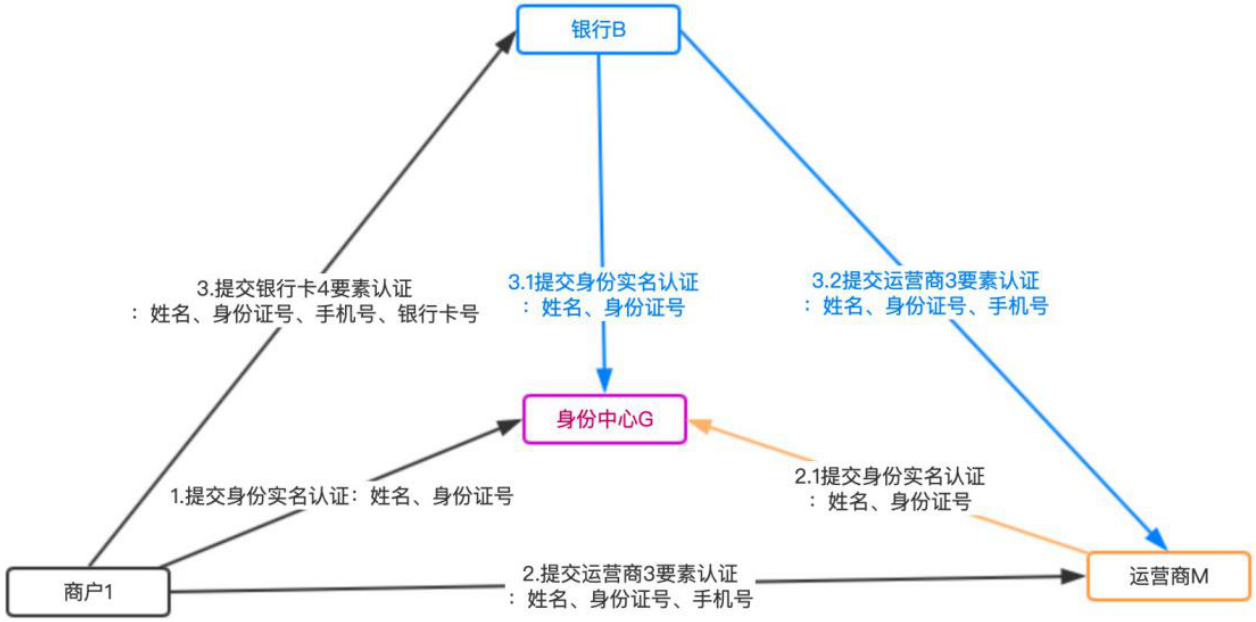
5.脱敏存储案例

数据服务中基本的账户注册和验证操作，涉及到姓名、身份证号、手机号和银行卡号的验证，商户需要与3个外部机构也为交互。同时3个外部机构间也有业务交互。

交互关系

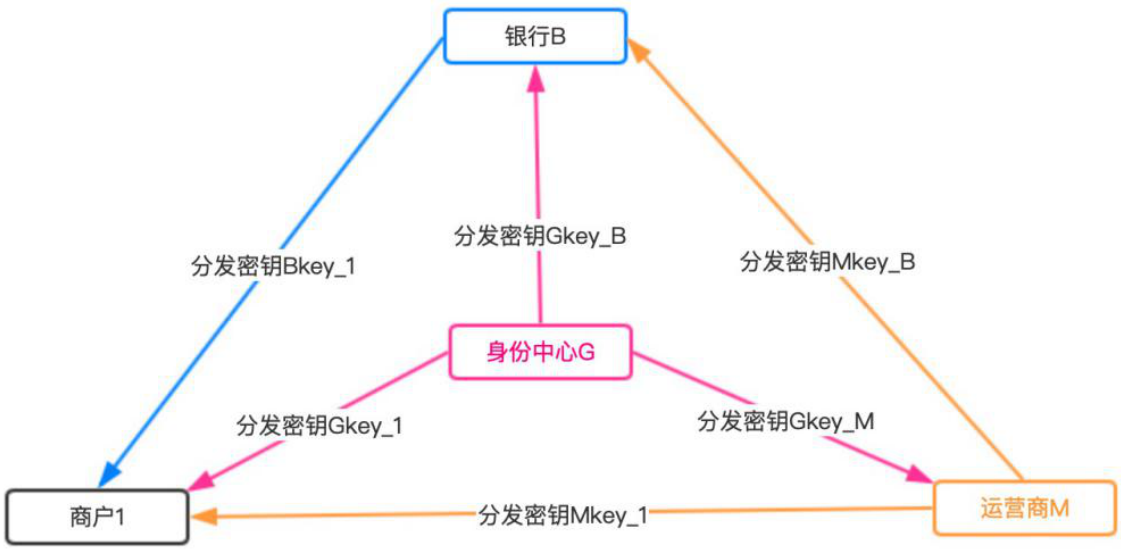
商户与运营商、身份中心、银行3个数据源交互
商户向运营商发送手机验证码短信请求（手机号）；
商户向身份中心发送身份实名认证（姓名+身份证号）；
商户向运营商发送运营商3要素认证（姓名+身份证号+手机号）；
商户向银行发送银行卡4要素认证（姓名+身份证号+手机号+银行卡号）；
运营商与身份中心1个数据源交互
运营商向身份中心发送身份实名认证（姓名+身份证号）；
银行与运营商、身份中心2个数据源交互
银行向运营商发送手机验证码短信请求（手机号）；
银行向身份中心发送身份实名认证（姓名+身份证号）；

银行向运营商发送运营商3要素认证（姓名+身份证号+手机号）。



密钥分发

身份中心向运营商、银行和商户分发加密密钥；
运营商向银行和商户分发加密密钥；
银行向商户分发加密密钥。

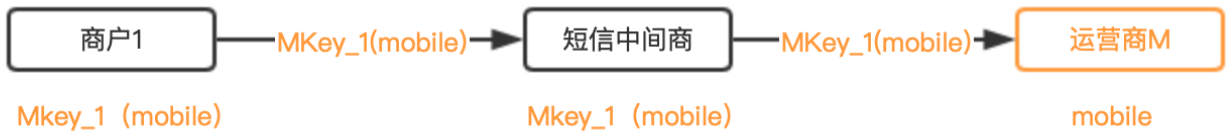


脱敏存储后与交互

手机号账户注册

用户输入手机号mobile，商户1调取运营商M短信验证码接口；注册完成后商户1删除明文mobile，仅存储Mkey_1(mobile)

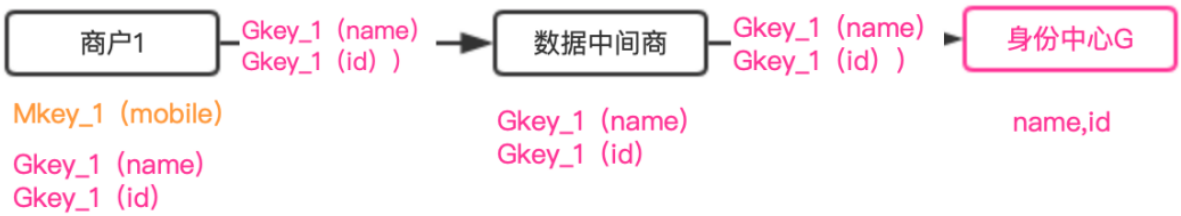
手机号注册账号



实名认证

用户输入姓名name和身份证号id，商户1调取身份中心G实名认证接口；认证完成后商户删除明文name和id，仅存储Gkey_1(name)和Gkey_1(id)。

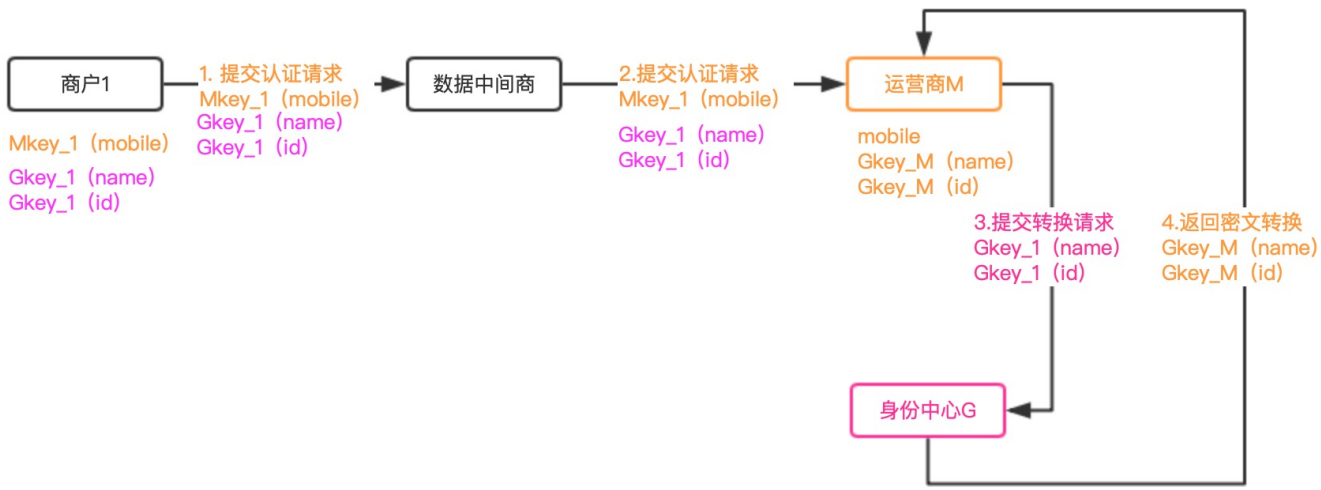
身份实名认证



运营商3要素认证

商户1调取运营商M的3要素认证接口，提交Mkey_1(mobile)、Gkey_1(name)和Gkey_1(id)；
运营商M收到请求，仅能识别Mkey_1(mobile)，无法识别Gkey_1(name)和Gkey_1(id)。运营商M需调用身份中心G接口，将商户1密钥Gkey_1加密的身份信息转换为运营商Gkey_M密钥加密的身份信息Gkey_M(name)和Gkey_M(id）；转换完成后运营商M即能完成3要素认证。

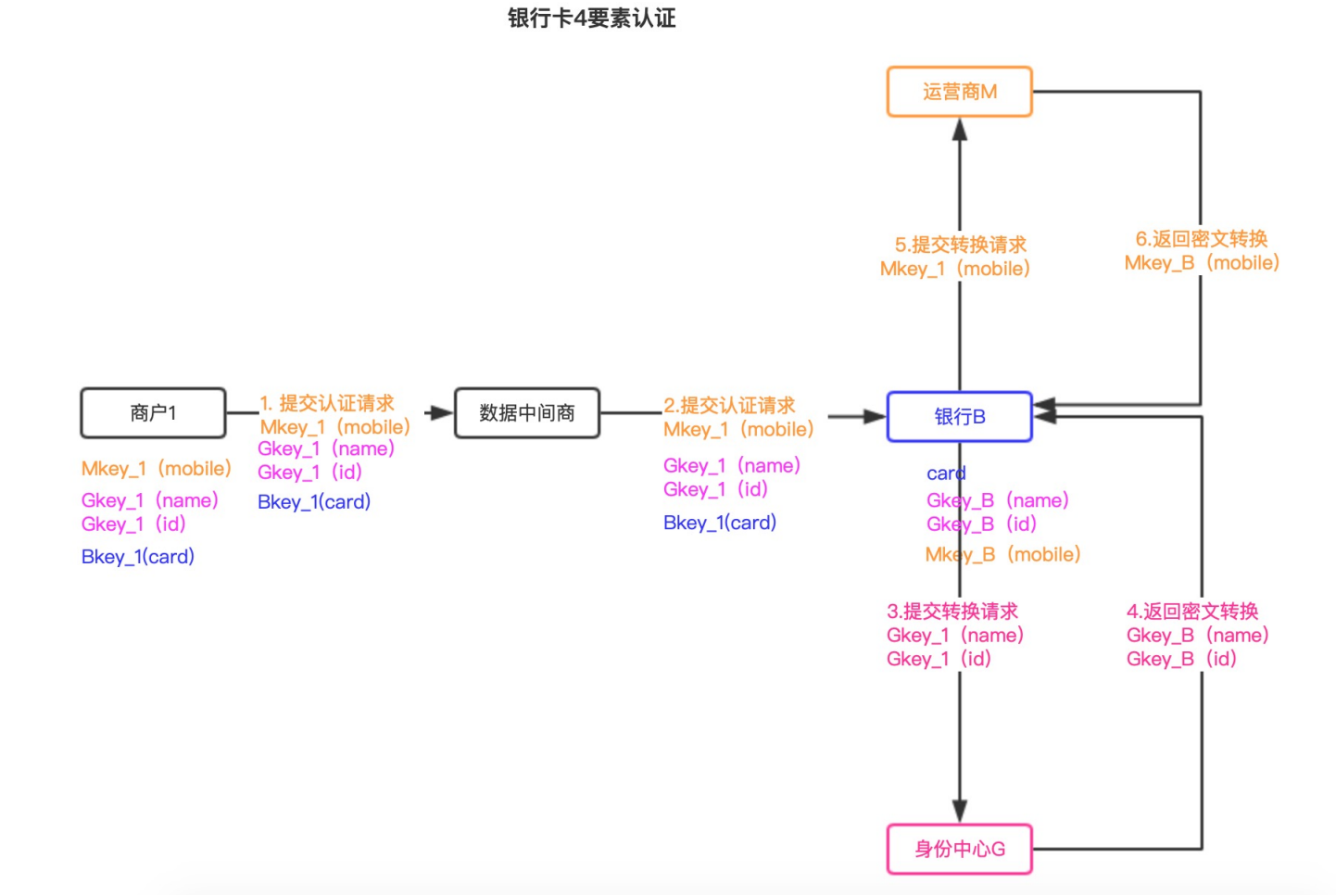
运营商3要素验证



银行卡4要素认证

用户输入银行卡号card，商户1向银行B调取银行卡4要素认证接口，提交Bkey_1(card)、Mkey_1(mobile)、Gkey_1(name)和Gkey_1(id)；
银行B收到请求，仅能识别Bkey_1(mobile),对于Gkey_1(name)和Gkey_1(id)需调用身份中心G接口，将商户1密钥Gkey_1加密的身份信息转换为银行B密钥Gkey_B加密的身份信息Gkey_B(name)和Gkey_B(id）;对于Mkey_1(mobile)需调用运营商M接口，将商户1密钥Mkey_1加密的手机号转换为银

行B密钥加密的手机号Mkey_B(mobile）；转换完成后即能完成4要素认证。



五、DSChain区块链

密钥的分发、同步和维护使用传统的系统交互在落地操作层面很难实现，但通过区块链技术很好的解决。

DSChain在供应链数据体系内，为数据源提供个人信息脱敏加密密钥管理工具，实现商户、中间商和数据源之间的加密密钥管理。

DSChain承担个人信息脱敏加密密钥的分发和同步，不参与机构间的数据交易，使得数据源及其下游每一个商户在个人信息加密脱敏后，完成正常业务处理。

1.非对称加密算法密钥管理

密钥管理分为公私钥对生成、公钥发布、公钥获取、公钥更新、脱敏存储批量更新。

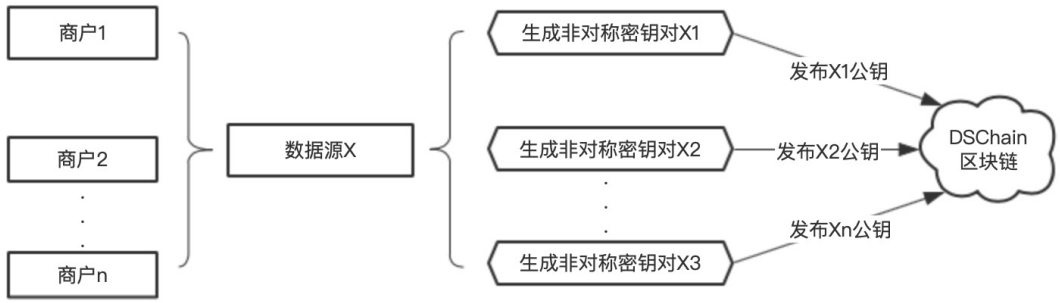
公私钥对生成

商户直接或间接将基本信息提交至数据源，数据源为其生成非对称加密算法密钥对。

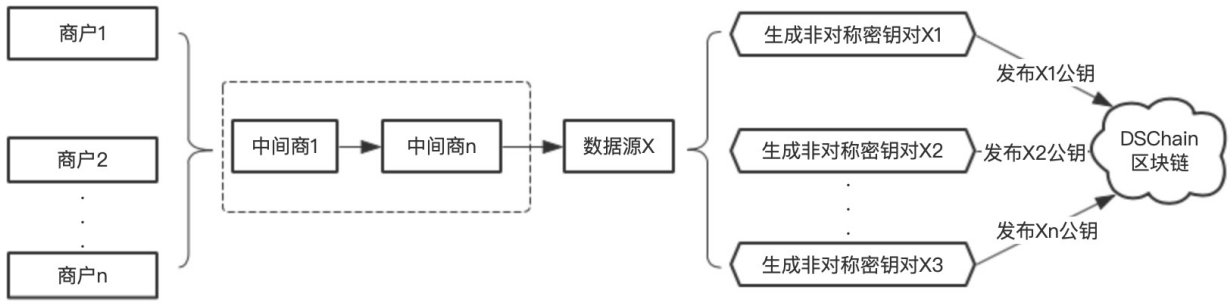
公钥发布

数据源将商户公钥发布至DSChain区块链。

直连模式：商户开通数据源X开公钥



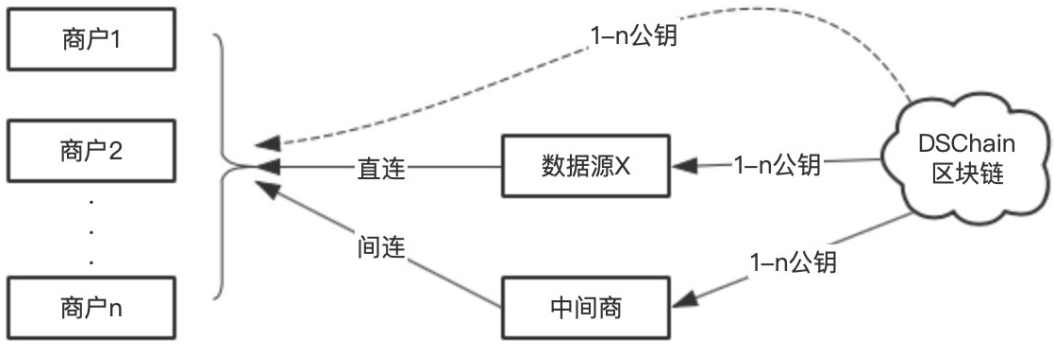
间连模式：商户在中间商开通数据源X公钥



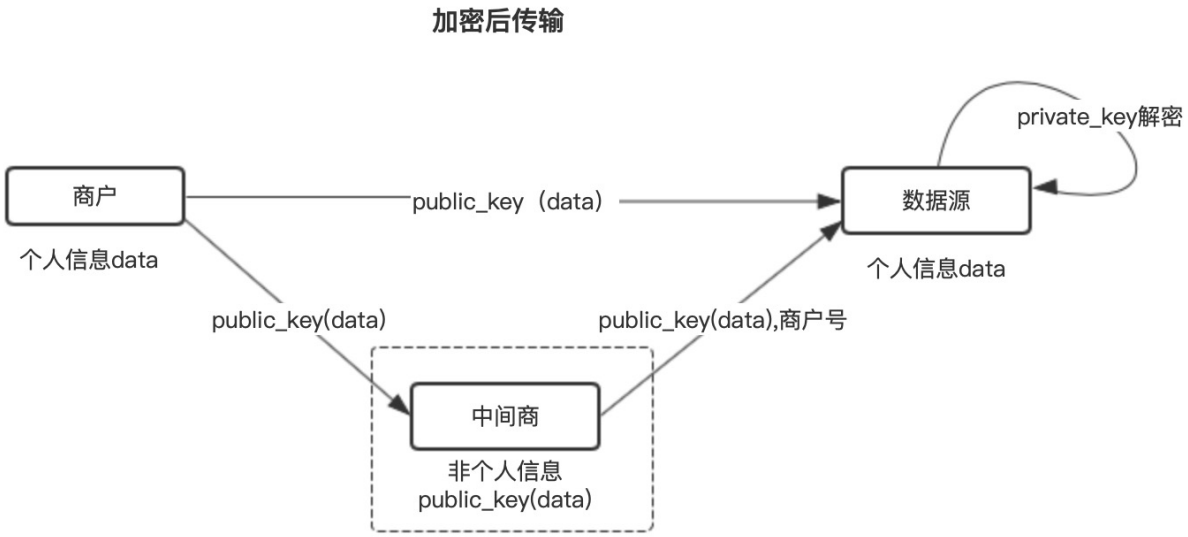
公钥获取

商户直接或间接DSChain区块链获取公钥。

商户获取数据源X公钥



脱敏信息传输



公钥更新

数据源发布新公钥信息（公钥编号，new_public_key,生效日期）至DSChain区块链;
商户直接或间接从DSChain获取公钥信息（new_public_key,生效日期）。

脱敏存储批量更新

商户更新加密公钥后，向数据源批量提交public_key(data)，转换成new_public_key(data)。

2.对称加密算法密钥管理

密钥管理分为密钥生成、密钥发布、密钥获取、密钥更新、脱敏存储批量更新、商户CA更新。

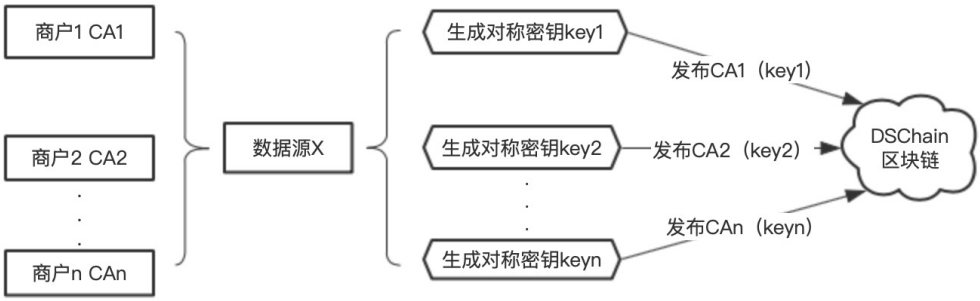
密钥生成：

商户直接或间接将CA证书和基本信息提交至数据源，数据源为其生成对称加密算法密钥key。

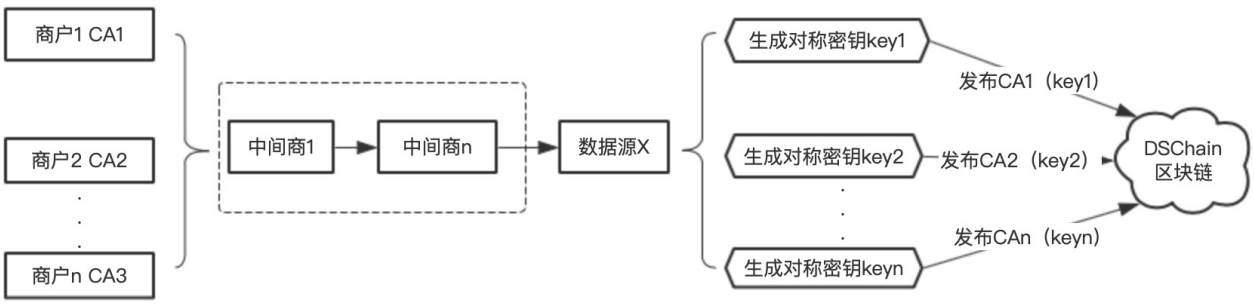
密钥发布

数据源将生成的密钥key通过商户CA证书公钥加密CA(key)，并发布到DSChain区块链。

直连模式：商户开通数据源X密钥



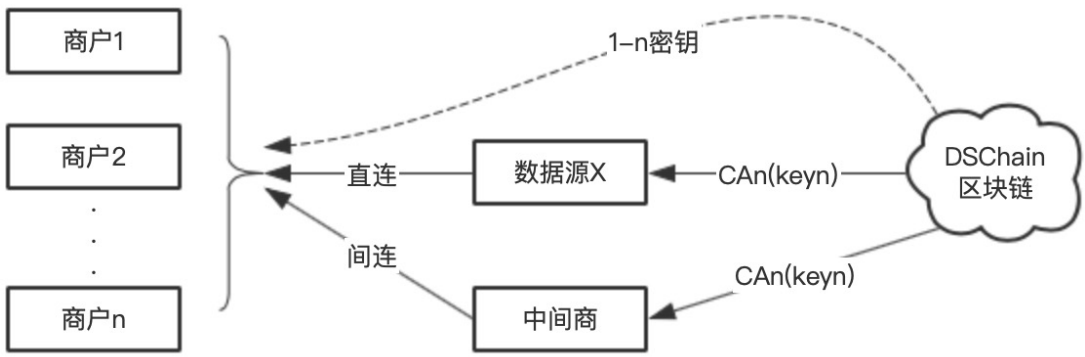
间连模式：商户在中间商开通数据源X密钥



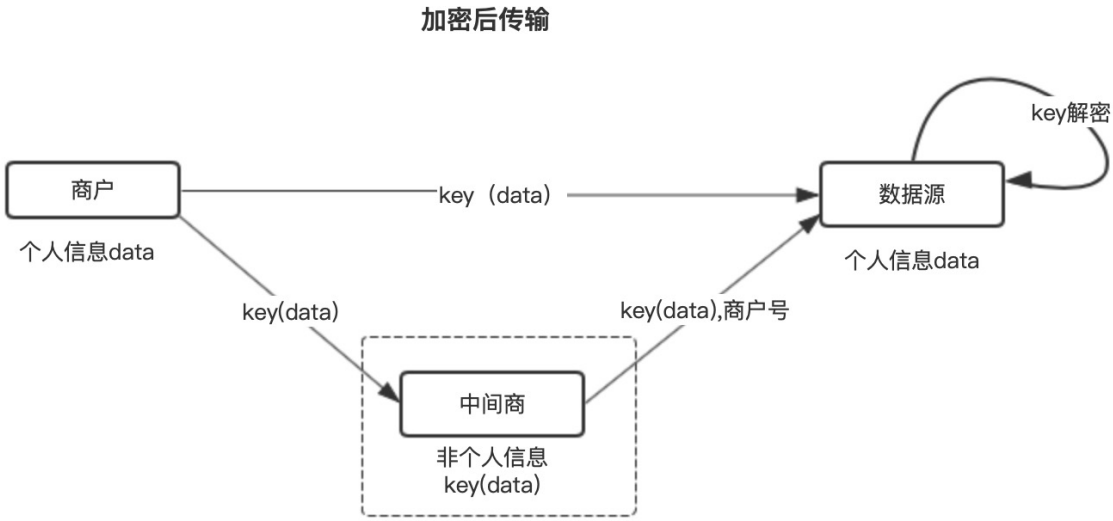
密钥获取

商户直接或间接DSChain区块链获取CA（key）。

商户获取数据源X密钥



脱敏传输



密钥更新

数据源发布新的密钥信息（公钥编号，CA(new_key),生效日期）至DSChain区块链;
商户直接或间接中间商从DSChain获取密钥信息（CA(new_key),生效日期）。

脱敏存储批量更新

商户更新加密密钥后，向数据源批量提交key(data)，转换成new_key(data)。

六、DSChain系统功能设计

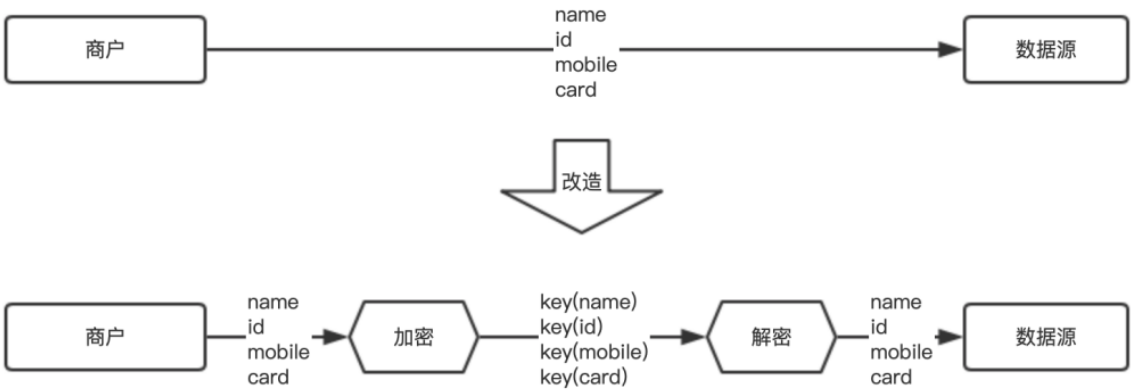
模块	功能	存储信息	操作权限
商户信息	基本信息	版本号、发布方、商户名称、地址、官网、营业执照、使用场景、特殊行业牌照、有效起始日期	数据源、中间商、商户
-	CA证书	版本号、发布方、CA公钥、有效起始日期	数据源、中间商、商户
密钥信息	非对称加密	版本号、发布方、数据源、商户、加密算法类型、public_key、有效起始日期	数据源
-	对称加密	版本号、发布方、数据源、商户、加密算法类型、CA（key）、有效起始日期	数据源

七、DSChain系统改造

DSChain区块链主要应用于加密密钥的管理，相对独立于业务系统，因此系统改造工作少。主要涉及2个层面改造，传输和存储改造。

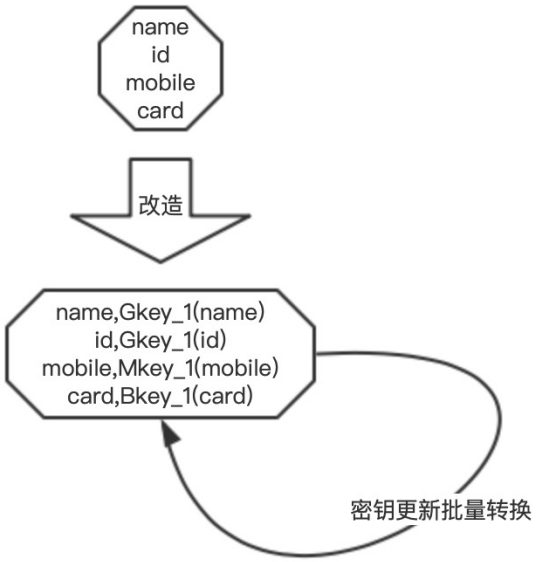
1.商户与数据源传输改造

传输改造工作最小，商户仅做加密操作，数据源做解密操作。

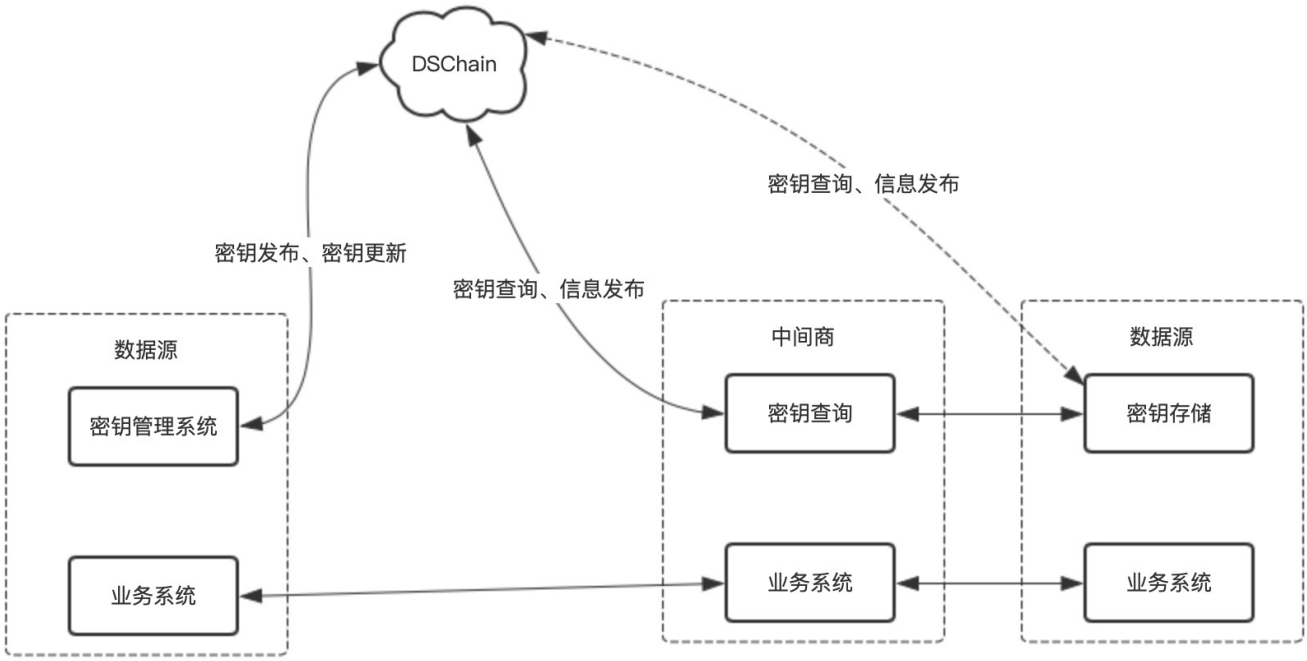


2.商户存储改造

商户同时存储明文和加密后的密文



3.商户、中间商和数据源上链改造



八、优势对比

类型	问题	DSChain方式	传统方式
传输方式对比	中间商留存	无法留存	可留存
-	中间商泄漏	无法泄漏	可泄漏
-	信息流向	信息流向与使用方十分明确	流向不明
-	数据来源	信息来源十分明确	来源不明
存储方式对比	商户留存	留存不可解密文	明文
-	商户泄漏	泄漏无影响	可泄漏
-	数据源留存	明文与密文存储物理隔离	明文存储
-	数据源泄漏	泄漏密文无影响	可泄漏