ICS 35.020
CCS L 70

MR

中华人民共和国市场监管行业标准

MR/T XXXXX—XXXX

市场监管行业数据隐私计算总体要求

General requirements for privacy computing of market supervision industry data

（草案）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

国家市场监督管理总局　　发 布

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国科学院信息工程研究所提出。

本文件由国家市场监督管理总局归口。

本文件起草单位：中国科学院信息工程研究所、中国网络安全审查认证和市场监管大数据中心、河北省市场监督管理局、南阳市市场监督管理局、江西省市场监督管理局、江西省质量和标准化研究院、厦门市市场监督管理局、西安电子科技大学、中央网信办数据与技术保障中心、国家市场监督管理总局竞争政策与评估中心、中国轻工业信息中心、航天信息股份有限公司、京东城市（北京）数字科技有限公司、北京市计算中心有限公司、中移（苏州）软件技术有限公司等。

本文件主要起草人：李凤华、崔琦、李晖、张玲翠、刁毅刚、于大东、朱辉、牛犇、贾洋、宋祁鹏、赵阳、宋若宁、苏义军、赵琉涛、王瀚仪

市场监管行业数据隐私计算总体要求

1 范围

本文件描述了数据隐私保护的目标、隐私信息全生命周期过程的计算操作，给出了隐私计算总体框架和参与者，描述了隐私信息抽取与度量、隐私度量动态调整、隐私延伸控制、隐私按需保护、保护效果评估、存证与取证等框架核心组件的功能。

本文件适用于数据泛在流通与共享过程中隐私信息全生命周期保护、跨平台/跨系统/跨域流通利用的隐私延伸控制、隐私按需保护、保护效果评估等，适用于市场监管部门，还适用于互联网、通信等领域的企业为主体的个人信息处理者、个人信息保护产品提供商、产品评测机构、个人信息保护合规审计评估机构、认证监管机构等，为隐私信息保护、隐私计算服务安全评估提供参考。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 25069-2022 信息安全技术 术语

GB/T 31500-2015 信息安全技术 存储介质数据恢复服务要求

GB/T 35273-2020 信息安全技术 个人信息安全规范

GB/T 37988-2019 信息安全技术 数据安全能力成熟度模型

GB/T 37964-2019 信息安全技术 个人信息去标识化指南

3 术语和定义

GB/T 25069-2022界定的以及下列术语和定义适用于本文件。

3.1

个人信息 personal information

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特定自然人身份或者反映特定自然人活动情况的各种信息，包含个人信息本身及其衍生信息,不包括匿名化处理后的信息。

[来源：GB/T 35273—2020,3.1,有修改]

3.2

标识符 identity

可以明显识别记录主体身份的属性集合，包括姓名、电话号码、身份证号码等信息。

3.3

准标识符 implied identity

组合起来可以识别记录主体身份的属性集合，包括年龄、性别、邮编等信息。

3.4

隐私信息 private information

能通过信息系统进行处理的敏感个人信息，是个人信息记录中的标识符、准标识符和敏感属性的集合。

注：隐私信息包括个人生物特征信息、银行账号、通信记录和内容、财产信息、征信信息、行踪轨迹、住宿信息、健康生理信息、交易信息、14岁以下（含）儿童的个人信息等。

3.5

隐私计算 privacy computing

面向隐私信息全生命周期保护的计算理论、方法和技术，涵盖了收集、脱敏、存储、使用、交换、删除、存证与取证等全生命周期过程的所有计算操作，包含处理视频、音频、图像、图形、文字、数值、泛在网络行为信息流等信息时，对所涉及的隐私信息进行描述、度量、控制、脱敏、使用、评价、删除等处理。

3.6

敏感属性 private attribute

信息载体中含有敏感个人信息的属性，泄露、修改或破坏该属性值会对个人权益产生影响。

注：在潜在的重标识攻击期间需要防止其值与任何一个隐私信息主体相关联。

[来源：GB/T 37964-2019,3.10,有修改]

3.7

隐私信息向量 private information vector

隐私信息提取的输出结果，由一个向量标识和若干隐私信息分量组成，隐私信息分量是指具有一定语义的、彼此之间互不相交的原子隐私信息。

注：向量标识是该隐私信息向量的唯一标识。

3.8

约束条件集合 constraint condition set

由隐私信息分量对应的约束条件向量组成的集合，用于描述在不同场景下实体访问隐私信息分量所需的访问权限，或者使用隐私信息分量时的操作限制要求。

3.9

隐私属性向量 private attribute vector

由隐私属性分量组成，用于量化隐私信息分量及分量组合的敏感度或者期望保护程度。

注：在实际应用时针对不同场景，不同隐私信息分量可进行加权动态组合，隐私属性分量的取值范围可以根据场景定义。

3.10

广义定位信息集合 generalized locating information set

由广义定位信息向量组成的集合，广义定位信息向量是由隐私信息分量在信息中的位置信息和属性信息组成。

注：广义定位信息集合可用于对隐私信息分量在信息中快速定位。

3.11

审计控制信息集合 audit control information set

由传播过程中具体的审计控制向量组成的集合，用于记录隐私信息分量在流转过程中的主客体信息和被执行的操作。

注：当发生隐私信息泄露时，可基于审计控制信息集合进行追踪溯源。

3.12

传播控制操作集合 circulation control operation set

由传播控制操作向量组成的集合，用于描述隐私信息分量及其组合的流转限制要求、可被执行的操作限制要求。

3.13

隐私信息所有者 private information owner

隐私信息所标识或者关联的自然人、组织、设备或程序等实体。

3.14

隐私信息提供者 private information provider

向其他自然人、组织、设备或程序提供隐私信息的实体。

3.15

隐私信息发布者 private information publisher

基于向特定或所有公众自由访问的目的，向其他自然人、组织、设备或程序提供隐私信息的实体。

3.16

隐私信息接收者 private information recipient

接收其他自然人、组织、设备或程序提供的隐私信息的实体。

3.17

隐私信息转发者 private information forwarder

接收其他自然人、组织、设备或程序提供的隐私信息的实体，该实体对接收到的隐私信息经过使用、或迭代脱敏、或保持原样转发给其他隐私信息接收者。

3.18

隐私信息使用者 private information user

对接收到的隐私信息进行统计、加工、模型训练等操作的实体。

3.19

隐私信息收集者 private information collector

从隐私信息所有者、隐私信息提供者或其他公开渠道获取隐私信息的实体。

3.20

隐私信息删除者 private information remover

对持有的隐私信息执行删除操作的实体。

3.21

隐私信息处理者 private information processor

对隐私信息进行收集、存储、使用、加工、传输、提供、公开、删除、脱敏、存证与取证等操作的实体。

3.22

延伸控制 extended control

在数据流通与共享过程中，收集、存储、使用、加工、传输、提供、公开、删除、脱敏、存证与取证等环节的隐私操作迭代控制、控制策略动态调整、控制策略可控传递，以及控制策略执行可信验证。

3.23

脱敏要求 desensitization requirements

待脱敏的隐私信息的脱敏等级、脱敏时机、脱敏算法及其参数选择等约束信息。

3.24

脱敏算法 desensitization algorithm

通过对隐私信息的技术处理，使其在不借助额外信息的情况下，无法识别或者关联隐私信息主体。

注：脱敏算法包括k-匿名、差分隐私等算法。

[来源：GB/T 35273—2020,3.15]

3.25

脱敏效果期望 expectation on desensitization performance

隐私信息执行脱敏操作之后所达到的预期效果。

3.26

按需脱敏 on-demand desensitization

隐私信息处理者按照隐私信息的延伸控制要求进行脱敏的过程。

3.27

可逆性 reversibility

被脱敏掉的隐私信息被复原的可能性。

3.28

泛化 generalization

将一类属性中的特定值用一个更宽泛的值代替，以更概括、抽象的方式描述数据。

注：泛化技术包括替换、取整、K-匿名、模糊化、概化等手段。

[来源：GB/T 37964—2019,A.5.1,有修改]

3.29

抑制 suppression

将某个属性、属性的值或者属性值的一部分进行删除或者以特定的符号代替。

3.30

解耦和置换 anatomization and permutation

去除准标识符和敏感属性间的关联性，而不改变准标识符或敏感属性的值。

3.31

扰动 perturbation

用合成的数据值取代原始的数据值，改变后的数据与真实数据主体失去关联性。

注1：扰动后统计信息不发生显著改变。

注2：扰动化技术包括加噪、数据交换、合成数据生成等。

3.32

差分隐私 differential privacy

通过扰动的方式对个人隐私信息进行脱敏，且扰动添加的噪声类型和参数满足差分的数学定义。

3.33

信息偏差性 information deviation

脱敏算法执行前后，可观测到的脱敏信息与原始信息的偏差。

3.34

信息损失性 information loss

信息被不可逆的脱敏算法作用后，隐私信息损失部分对可用性的影响程度。

3.35

信息可用性 information usability

在隐私信息进行脱敏操作后，其在具体应用场景中的可用程度。

3.36

复杂性 complexity

执行脱敏算法所需要的资源开销。

注：复杂性通常用时间开销和空间开销表示。

3.37

脱敏效果评估 desensitization performance evaluation

隐私信息脱敏后，在可逆性、信息偏差性、信息损失性等方面进行量化评估。

3.38

删除 delete

采用访问控制、消磁、物理破坏等技术或措施，使得信息不能被访问或被检索，或者从物理上去除了信息并保障其难以恢复的操作。

注：删除包括不能被访问或被检索、全部物理删除或部分物理删除。

[来源：GB/T 35273—2020,3.10,有修改]

3.39

删除对象 deleted object

删除操作的客体。

注：删除对象包括个人信息的正本信息、副本信息、正本信息的一部分、副本信息的一部分，以及正本信息与副本信息的全部或者部分的组合。

3.40

删除等级 delete level

删除对象可恢复程度和难度的量化分级。

3.41

数据恢复 data recovery

通过专门的计算机软件、硬件等技术，从删除对象曾经留存过的存储系统或介质中，重建被删除对象的过程。

4 缩略语

下列缩略语适用于本文件。

JSON：JavaScript 对象标记(JavaScript Object Notation)

RAID：独立磁盘冗余阵列(Redundant Array of Independent Disks)

XML：可扩展标记语言(Extensible Markup Language)

5 概述

5.1 隐私计算概述

5.1.1 隐私保护的目标

市场监管数据中包括了大量的个人信息，针对个人信息中隐私信息在多个信息系统中跨系统泛在传播，隐私保护需要达到如下主要目标:

a) 支持全流程一致性保护，隐私信息在多个信息系统保存时，如果由于某个信息系统保护能力偏弱而导致隐私泄露，其他信息系统隐私保护能力再强也失去意义，即存在短板效应和一损俱损的风险。因此，需要确保各个信息系统之间的一致性保护，消除由于个别系统保护薄弱而导致的隐私保护失效现象；

b) 支持多次传播的延伸控制，由于缺乏隐私信息传播的延伸控制，隐私信息接收者对隐私信息不受控的后续传播将导致隐私泄露。因此，需要通过延伸控制机制，确保隐私信息在不同接收者之间的多次传播过程中依然受到有效控制，防止未经授权的传播和使用；

c) 支持隐私按需保护,在数据流通利用过程中,需要对同一隐私信息在同一应用场景的不同阶段，或者同一隐私信息在不同应用场景下进行差异化的脱敏处理或删除。这种按需脱敏或按需删除的机制，能够实现隐私信息利用与隐私保护之间的平衡；

d) 提供基于保护效果评估反馈的保护自适应改进机制，隐私保护与隐私信息挖掘之间存在持续对抗。在这种动态环境中，通过保护效果评估反馈，信息系统可以自适应地调整脱敏算法和参数选择、删除方法选择等。在满足隐私信息利用需求的前提下，提高隐私保护强度，减少因长期不变的隐私延伸控制策略导致的隐私泄露风险。

5.1.2 隐私计算的参与者

隐私计算的参与者包括:隐私信息所有者、隐私信息提供者、隐私信息发布者、隐私信息接收者、隐私信息转发者、隐私信息使用者、隐私信息收集者、隐私信息删除者、隐私信息处理者。隐私信息的参与者在隐私计算的过程中对隐私信息进行相应的隐私信息处理。

5.1.3 隐私计算与特性

数据安全是指在数据泛在流通与共享过程的各个环节中防止用户对数据进行非授权获取或篡改，数据接收方获得的内容与数据提供方提供的内容完全相同，具体的技术包括机密计算、密文计算、安全多方计算等。隐私保护是指隐私信息在泛在流通与共享过程中进行脱敏，使信息产生偏差或去标识化，接收方获得的隐私信息少于提供方的原始隐私信息，具体的技术包括隐私计算和传统的隐私保护技术（如差分隐私、k-匿名等）。

隐私计算是面向隐私信息全生命周期保护的计算理论、方法和技术，涵盖了收集、脱敏、存储、使用、交换、删除、存证与取证等全生命周期过程的所有计算操作，包含处理视频、音频、图像、图形、文字、数值、泛在网络行为信息流等信息时，对所涉及的隐私信息进行描述、度量、控制、脱敏、使用、评价、删除等处理。

隐私计算是对隐私信息跨系统的全生命周期保护、全生命周期的延伸控制和按需保护,如差分隐私、k-匿名等传统的隐私保护技术属于隐私计算范畴，在隐私计算全生命周期的使用环节可以采用数据安全技术防止隐私信息的非法获取或篡改，如实现隐私集合求交、隐匿信息查询、联合统计分析、联合建模等部分功能。隐私计算的主要特性包括：

a) 延伸控制性，是指数据泛在流通与共享过程中全生命周期各环节隐私操作的迭代控制、控制策略的动态调整、控制策略的可控传递、控制策略执行的可信审计。控制策略由隐私信息所有者

的控制意图、当前隐私信息使用者的控制约束和隐私信息接收者的防护能力生成，同一隐私信息的控制策略在全生命周期中是差异化的，且随流转过程同步传递不可分割。迭代控制通过泛在流转过程中的差异化控制策略生成与传递机制实现，贯穿于隐私信息从收集到删除的不同流通与共享过程中；

b) 原子性，在对隐私信息的描述过程中，隐私信息分量是隐私信息的最小度量单位，具有原子性；在此基础上，多个隐私信息分量可以组合构成新的隐私属性，对这些组合属性的敏感性也可进行度量；

c) 一致性，对相同的隐私信息，不同算法的隐私保护效果都使隐私信息分量的敏感度趋近于零，即不同算法在不同系统中隐私保护的趋势保持一致性。例如算法A和B在系统1中的保护能力评估是A>B，在映射到系统2中的评估体系时应仍然保持A>B；

d) 顺序性，隐私保护算法中所有操作必须按照设计的顺序执行，部分操作的顺序不同可能导致隐私保护的效果不同；

e) 不可逆性，隐私计算中使用的脱敏算法对隐私信息的处理应是不可逆的，接收方获得的隐私信息少于提供方的原始隐私信息，且无法通过技术手段从脱敏信息中复原原始隐私信息中缺失的部分。

5.2 隐私信息全生命期过程的计算操作

隐私信息全生命周期过程的计算操作如图1所示，包括收集、脱敏、存储、使用、交换、删除、存证与取证等，这些操作在多个信息系统之间不同的业务流程组合可以涵盖大多数隐私信息全生命周期的操作过程。典型的流程示例如下：

a) 脱敏后发布，隐私信息收集后进行存储，随后对隐私信息进行脱敏，通过交换操作进行发布（流程为③-④-⑩）；

b) 出行场景，对出行服务过程中收集的隐私信息先存储和使用，服务结束后，应对隐私信息进行脱敏后存储，再进行后续使用（流程为③-⑥-⑦-④-⑤），具体示例见附录C；

c) 数据流通交易，收集的隐私信息先通过去标识化等脱敏操作，然后通过交换操作进行交易，此后根据购买方的出价再进行脱敏，进行二次或者部分交易（流程为①或④-⑤-⑪-⑫-⑬-⑯）；

d) 差分统计，对收集的敏感属性值等隐私信息先利用泛化、本地化差分隐私等机制脱敏后再进行存储（流程为①-②-③）；

e) 信用计算，隐私信息所有者存储的隐私信息先交换到隐私信息接收者，隐私信息接收者对其进行脱敏后进行使用，并发布计算结果（流程为⑪-⑫-⑬-⑭-⑮），具体示例见附录D。

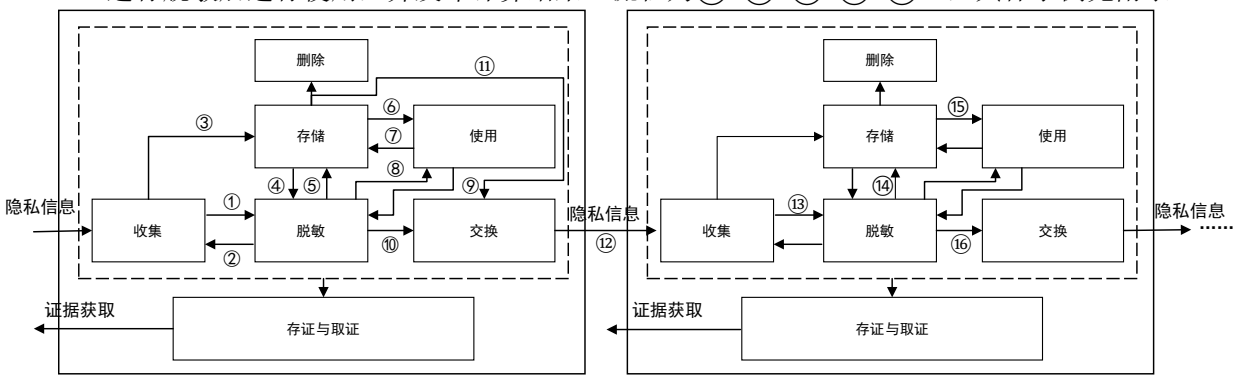


图1 隐私信息全生命周期过程的计算操作

5.2.1 收集

收集操作从隐私信息所有者或者隐私信息提供者采集包含隐私信息的原始信息，收集涉及的功能包括：

a) 针对标识符、准标识符、其他敏感属性值采用关键词匹配、自然语言理解、图像理解等技术措施在多模态数据中感知识别隐私信息，提取不同信息模态敏感属性；

b) 识别出的隐私信息采用多元组形式表示，可包含：隐私信息向量、约束条件集合、隐私属性向量、广义定位信息集合、审计控制信息集合、传播控制操作集合等；

c） 部分场景的收集操作会直接对隐私信息采用泛化、差分隐私等处理后作为收集操作的输出。

5.2.2 存储

存储操作对收集到的隐私信息进行安全高效存储，或对收集的隐私信息进行脱敏后存储，存储涉及的功能包括：

a） 机密性，采用加密技术和访问控制技术，防止存储的隐私信息不被非法获取；
b） 可靠性，采用数据冗余存储技术保证存储隐私信息的可用性；
c） 完整性，采用密码校验或数据签名技术保证存储隐私信息不可篡改。

5.2.3 使用

使用操作包括《个人信息保护法》规定的个人信息处理中的使用和加工，采用的技术包括但不限于数据挖掘与分析、安全计算、模型训练等，使用涉及的功能包括：

a） 可以对未脱敏的隐私信息进行使用操作，也可以对存储的隐私信息经过脱敏后再进行使用操作；
b） 在使用环节中也可以使用机密计算、安全多方计算、同态加密等数据安全技术保护隐私信息不被泄露。

5.2.4 交换

交换操作包括《个人信息保护法》规定的个人信息处理中的传输、提供和公开，是隐私信息在不同隐私信息处理者之间受控流通与共享的过程。交换操作可以具有以下安全机制：

a） 认证性，隐私信息交换时进行单向或双向身份认证；
b） 机密性，采用加密技术保障交换的信息不被非法获取；
c） 完整性，采用密码校验技术保障交换的信息不被篡改；
d） 来源真实性，采用数据签名等技术保障信息来源的不可否认性；
e） 延伸控制性，将延伸控制策略与隐私信息绑定传输，控制隐私信息接收者对隐私信息的后续处理。

5.2.5 删除

删除操作是隐私信息使用者当隐私信息在业务不需要继续使用时，遵循及时、透明的原则，安全可信且完备地删除存储的隐私信息。删除涉及的功能包括：

a） 按需删除，根据隐私信息所有者或前序的隐私信息提供者的删除要求，执行删除操作，履行删除义务；
b） 自动删除，根据隐私信息所有者或前序的隐私信息提供者对隐私信息保留时限的要求，实现到期后，隐私信息使用者自动执行删除操作，履行删除义务；
c） 删除操作，根据延伸控制策略的约束，选择删除方法，保证删除后信息不能通过技术手段恢复数据，或使删除后的信息不能被访问和不能被检索。

5.2.6 脱敏

脱敏操作依据脱敏延伸控制策略对隐私信息选择适当的脱敏算法及其参数进行按需脱敏，脱敏涉及的功能包括：

a） 脱敏操作包括脱敏算法能力评估、脱敏算法选择、脱敏效果评估；
b） 对脱敏后的信息如果没有达到预期脱敏效果，则调整算法及其参数进行迭代脱敏，直至达到预期脱敏效果。

5.2.7 存证与取证

存证与取证操作对隐私信息全生命周期各种操作进行定制化的操作记录生成和存储，并响应用户的证据查询请求，返回生成的证据，支撑隐私保护合规审计、隐私侵权行为溯源与追责，存证与取证涉及的功能包括：

a） 客观完整地记录隐私信息全生命周期的各种操作行为；
b） 采用密码技术保障操作日志记录的安全性；

c） 采用访问控制技术保障操作日志记录受控使用和响应证据服务请求。

5.3 隐私计算技术

5.3.1 功能层次框架

在隐私信息全生命周期过程的计算操作中,隐私计算包含隐私信息抽取与度量、隐私度量动态调整、隐私延伸控制、隐私按需保护、保护效果评估等技术。隐私计算技术用于支撑和实现隐私计算服务的功能，隐私计算的功能组件通过层次框架进行组织，隐私计算功能层次框架如图2所示，包括：

a） 用户层，用于隐私计算的各参与者执行与用户相关的管理功能,访问、使用和维护隐私计算系统；

b） 服务接口层，通过调用隐私计算核心功能层的功能组件,为隐私信息应用系统提供隐私计算服务支撑；

c） 核心功能层，基于基础设施层实现隐私计算相应功能，为服务接口层提供相关功能支持服务，主要包括隐私动态度量、迭代延伸控制、隐私按需保护、保护效果评估、保护量化映射、操作全程存证等；

d） 基础设施层，提供隐私计算系统正常运行所需要的硬件设备之上的运行环境和基础组件，包括网络、计算和存储等；

e） 跨层功能，提供跨越多个层次的功能组件,包括监管、操作全程存证、审计等。

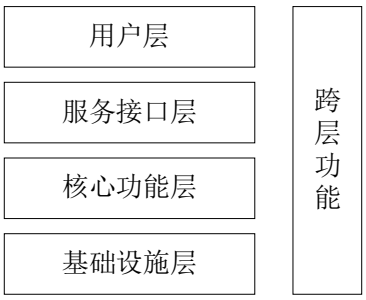


图2 隐私计算功能层次框架

5.3.2 技术功能

隐私计算技术为隐私计算功能提供支撑，具体如下：

a） 隐私动态度量，根据应用场景的不同，对隐私信息分量的敏感度进行动态度量；

b） 迭代延伸控制，根据前序隐私信息处理者的控制策略、后续隐私信息处理者的保护能力等因素动态调整控制策略，并随隐私信息一起向后传递；

c） 隐私按需保护，根据延伸控制策略，对隐私信息进行按需脱敏、按需删除、数据联合利用等处理；

d） 脱敏效果评估，对脱敏算法所达到的效果按照效果评估指标体系进行评估；

e） 保护量化映射，隐私信息跨系统交换或数据联合利用时，对隐私信息提供者和隐私信息接收者的算法保护能力和保护效果进行关联的保护量化映射，以支持跨系统交换的隐私信息一致性保护；

f） 操作全程存证，对隐私信息全生命周期过程的计算操作进行日志保存。

6 隐私计算框架

6.1 隐私计算框架组件

隐私计算框架如图3所示，包括隐私信息抽取与度量、隐私度量动态调整、隐私延伸控制、隐私按需保护、保护效果评估、存证与取证等功能组件，具体如下：

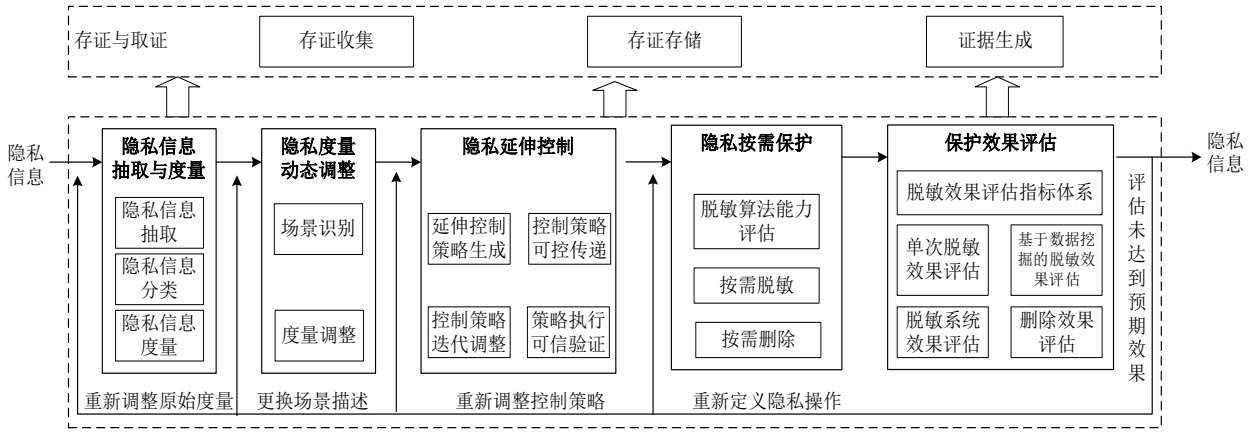


图3　隐私计算框架

a) 隐私信息抽取与度量，通过对采集或接收的信息进行分析，提取不同模态信息中的隐私信息分量，并对隐私信息分量进行分类以及量化隐私信息分量的敏感度或保护程度；当评估未达到预期效果时，还需要重新调整原始度量值。包括隐私信息抽取、隐私信息分类、隐私信息度量；

b) 隐私度量动态调整，通过识别判断隐私信息所属的应用场景，对隐私信息分量的敏感度或保护程度进行针对性的度量调整；当评估未达到预期效果时，还需要重新更换场景描述。包括场景识别、度量调整；

c) 隐私延伸控制，在数据泛在流通与共享过程中，对全生命周期各环节的隐私操作进行迭代控制；当评估未达到预期效果时，还需要重新调整控制策略。包括延伸控制策略生成、控制策略可控传递、控制策略迭代调整、策略执行可信验证；

d) 隐私按需保护，约束隐私信息处理者根据延伸控制策略，对接收到的隐私信息进行按需脱敏、按需删除等处理，提供场景自适应的隐私保护能力；当评估未达到预期效果时，还需要重新定义隐私保护操作；

e) 保护效果评估，根据制定的脱敏效果评估指标体系，对待评估的已脱敏隐私信息的脱敏效果进行量化分析，如未能达到预期效果，则分别视情况从隐私信息抽取与度量、隐私度量动态调整、隐私延伸控制、隐私按需保护等环节进行反馈迭代，直至达到期望的保护效果。包括脱敏效果评估指标体系、单次脱敏效果评估、基于数据挖掘的脱敏效果评估、脱敏系统效果评估、删除效果评估；

f) 存证与取证，对隐私信息抽取与度量、隐私度量动态调整、隐私延伸控制、隐私按需保护、保护效果评估等功能组件的操作行为进行日志记录，并根据证据获取请求生成证据。包括存证收集、存证存储、证据生成。

6.2　隐私计算系统的组件风险

基于隐私计算框架所研制的隐私计算系统的各组件面临着与其他信息系统相似的安全风险，在隐私信息抽取与度量、隐私度量动态调整、隐私延伸控制、隐私按需保护、保护效果评估、存证与取证等方面面临的风险包括但不限于：

a) 隐私信息抽取与度量的风险主要指隐私信息抽取不准或不全、隐私信息分量分类不正确、隐私信息分量敏感度的度量不准确等，会导致隐私延伸控制策略生成、脱敏算法及其参数选择、删除方法选择的不恰当，从而引起隐私信息泄露；

b) 隐私度量动态调整的风险主要指场景识别不准确、隐私信息分量敏感度调整不当等，会导致隐私延伸控制策略生成、脱敏算法及其参数选择、删除方法选择的不恰当，从而引起隐私信息泄露；

c) 隐私延伸控制的风险主要指延伸控制策略生成不准或不全、控制策略传递过程中被篡改或与隐私信息剥离、控制策略迭代调整时出现偏差、控制策略未被正确执行等，会导致数据泛在流通与共享过程中脱敏、删除和使用等操作失控，从而引起隐私信息泄露；

d) 隐私按需保护的风险主要指未能正确选择脱敏算法及其参数、未能正确选择删除方法等，脱敏时隐私信息被过度脱敏影响服务效果，或脱敏不足导致隐私信息泄露，或未能正确删除导致违规甚至隐私信息泄露，多方联合数据利用时未能正确选择所采用的数据使用安全技术、未能正确选择算法协议及配置其安全参数等问题；
e) 保护效果评估的风险主要指保护效果评估指标体系不完备、评估方法不准确等，使得保护效果评估出现偏差，从而会引起隐私信息泄露；
f) 存证与取证的风险主要指存证信息收集不全、存证存储过程中被非法访问或非法篡改、备份措施不足导致的存证信息丢失、证据生成时信息不全、证据生成错误等，会导致不能对隐私侵权行为进行正确追溯；
g) 跨系统一致性的风险主要指隐私信息在多个隐私计算系统保存时，如果由于某个系统保护能力偏弱，其他系统隐私保护能力再强也失去意义，即存在短板效应和一损俱损的风险，会导致隐私泄露。

6.3 隐私计算的互联互通

隐私计算是对隐私信息全生命周期过程的计算操作进行控制。相应地，隐私计算互联互通是体现在隐私信息抽取与度量、隐私度量动态调整、隐私延伸控制、隐私按需保护、保护效果评估、存证与取证等组件的隐私信息描述、隐私信息分类、隐私信息度量、场景识别、度量调整、延伸控制策略生成、控制策略迭代调整、控制策略可控传递、策略执行可信验证、脱敏算法能力评估、按需脱敏、按需删除、脱敏效果评估指标体系、保护效果评估结论等方面的具体方法、算法或量化指标等定义和格式规范上。本文件仅规范这些方面的原则要求。

7 隐私信息抽取与度量

7.1 概述

隐私信息抽取与度量通过对采集或接收的信息进行隐私信息分量识别和抽取，然后对隐私信息分量进行分类和度量。若保护效果评估未达到预期效果，则可能重新执行隐私信息抽取与度量。隐私信息抽取与度量具体包括：

a) 隐私信息抽取，负责对采集或接收到的信息中的隐私信息分量进行识别，根据隐私信息分量的模态进行信息处理，生成初始的隐私信息描述；
b) 隐私信息分类，负责对识别和抽取的隐私信息分量进行分类；
c) 隐私信息度量，负责对抽取的隐私信息分量进行敏感度或保护程度量化。

7.2 隐私信息抽取

隐私信息抽取，对采集或接收到的信息中的隐私信息分量进行定位，确定隐私信息分量模态，然后对多模态隐私信息分量进行处理，生成隐私信息描述，具体如下：

a) 对文本、音频、视频、图像等模态数据，使用合适的技术和方法从信息中识别隐私信息，如关键词匹配、自然语言处理、图像语义理解、模式匹配等；
b) 针对识别的隐私信息生成初始的隐私信息描述，其内容包括隐私信息向量、约束条件集合、隐私属性向量、广义定位信息集合、审计控制信息集合、传播控制操作集合等。具体隐私信息识别示例见附录A。

7.3 隐私信息分类

隐私信息分类，通过对识别和抽取的隐私信息分量进行分类，具体如下：

a) 针对法律法规和应用场景，建立分类分级模板库，以满足不同应用场景的隐私信息分类需求；
b) 可采用知识图谱等技术，结合分类分级模板库，对隐私信息分量进行分类，再结合应用场景对隐私信息分量进行分级；
c) 隐私信息分类分级的结果记录在隐私信息描述的隐私属性向量中。

7.4 隐私信息度量

隐私信息度量，负责对抽取的隐私信息分量的敏感度或保护程度进行量化，具体如下：

a) 可根据隐私信息分类分级，构建粗粒度隐私信息量化模型；
b) 可选择基于信息论、差分隐私、人工智能等多种不同技术，构建细粒度隐私信息量化模型；
c) 选用适合的隐私信息量化模型，对隐私信息向量中不同信息模态的隐私信息分量进行量化；
d) 可对隐私属性向量中的各属性分量进行联合度量，对不同隐私属性向量进行自定义权重的加权平均，获取隐私信息的整体度量；
e) 隐私信息度量结果记录在隐私信息描述中的隐私属性向量中；
f) 可选地，完成隐私信息度量之后，对隐私度量过程开展合规性验证，确保其实施过程符合隐私保护要求。

8 隐私度量动态调整

8.1 概述

隐私度量动态调整，在隐私信息抽取与度量的基础上，识别隐私信息所属的应用场景，并针对性地动态调整隐私信息分量的敏感度或保护程度。若保护效果评估未达到预期效果，则可能重新执行隐私信息动态调整。隐私度量动态调整具体包括：

a) 场景识别，根据提取的隐私信息分量，识别其所属应用场景；
b) 度量调整，结合识别的应用场景，动态调整隐私信息分量的量化结果。

8.2 场景识别

为支持隐私信息分量量化结果的动态调整，场景识别包括但不限于：

a) 采用人工标注或机器学习等技术，对采集的信息进行应用场景分类分级和标注，如：社交网络服务、出行服务、医疗服务等；
b) 采用元数据（例如：信息来源、时间戳、地理位置等）辅助识别隐私信息所属的应用场景；
c) 针对文本类信息，可采用自然语言处理技术，进行分词和词性标注、实体识别等操作，使用主题模型识别主要主题，辅助推断应用场景；
d) 针对视频类信息，可通过内容分析技术，如：物体检测识别、动作识别等，辅助推断应用场景；
e) 针对音频类信息，可通过语音识别、情感分析、背景音分类等技术，辅助推断应用场景；
f) 针对图像类信息，可通过图像语义理解等技术，辅助推断应用场景。

8.3 度量调整

在识别应用场景的基础上，针对不同场景动态调整隐私信息分量的敏感度或保护程度的量化结果，度量调整包含但不限于：

a) 根据不同应用场景及该场景下隐私信息接收者的隐私保护能力，动态调整隐私属性分量的量化值；
b) 根据保护效果评估结论，动态调整隐私属性分量的量化值，并指导隐私信息中隐私属性向量的动态调整；
c) 可综合运用信息论、差分隐私，并交叉融合心理学等主观评价理论，对保护效果评估指标体系进行动态调整。

9 隐私延伸控制

9.1 概述

隐私延伸控制对数据泛在流通与共享过程中脱敏、存储、使用、交换、发布、删除等进行操作约束。隐私延伸控制的核心功能迭代延伸控制详见附录B。若保护效果评估未达到预期效果，则可能重新执行隐私延伸控制。隐私延伸控制具体包括：

a) 延伸控制策略生成，用于根据前序隐私信息提供者和当前隐私信息处理者的意图，准确无误地生成并描述控制策略；

b) 控制策略可控传递，用于保证隐私延伸控制策略在不同隐私信息处理者之间安全可靠地传输到达；
c) 控制策略迭代调整，延伸控制策略在不同隐私信息处理者之间传递时，用于保证控制策略内容根据应用场景、保护效果评估结果自适应地进行调整；
d) 策略执行可信验证，用于保证隐私信息提供者或隐私信息转发者验证延伸控制策略是否被隐私信息接收者完整正确地执行。

9.2 延伸控制策略生成

使用自然语言处理、形式化分析等技术，准确地按照前序隐私信息提供者和当前隐私信息处理者的意图，产生计算机程序可处理的控制策略，具体如下：

a) 针对隐私信息被首次流转的应用场景，支持隐私信息所有者导入延伸控制意图；
b) 针对隐私信息非首次流转的应用场景，支持通过前序的隐私信息所有者、隐私信息提供者或隐私信息处理者的延伸控制策略，获得延伸控制意图；
c) 通过自然语言处理等技术，对获取的延伸控制意图进行解析，并结合隐私信息接收者的隐私保护能力、应用场景等因素，生成延伸控制策略；
d) 采用数字签名等技术，保证延伸控制策略的真实性以及不可篡改；
e) 采用机器可处理的语言形式，例如：JSON、XML等，描述生成的延伸控制策略；
f) 采用形式化分析技术，验证生成的延伸控制策略符合延伸控制意图的要求，且延伸控制策略各项内容彼此没有冲突；
g) 延伸控制策略采用底层系统无关的标准化描述，支持延伸控制策略的跨系统传递。

9.3 控制策略可控传递

采用密码学技术保障延伸控制策略在隐私信息提供者和隐私信息接收者之间传递过程中的完整性、机密性、不可剥离性、不可抵赖性和保护一致性，具体如下:

a) 完整性，利用消息验证码等技术，保证延伸控制策略在传递过程中不可被非授权方式更改或破坏；
b) 机密性，可采用密码技术对延伸控制策略进行加密保护，保证延伸控制策略不可被未授权的第三方解析，避免延伸控制策略被非法获取而导致的隐私泄露；
c) 不可剥离性，采用可信执行环境、密码学等技术将延伸控制策略嵌入被交换隐私信息中，并保证控制策略和隐私信息的关联关系不能被破坏；
d) 不可抵赖性，采用数字签名技术对延伸控制策略进行处理，保证延伸控制策略来源的真实性，避免隐私信息处理者否认其前序隐私信息提供者所生成的延伸控制策略；
e) 保护一致性，隐私信息跨系统交换或采用数据使用安全技术进行数据联合利用时，对隐私信息提供者和隐私信息接收者的算法保护能力和保护效果进行关联的保护量化映射，量化映射关系存于延伸控制策略中。

9.4 控制策略迭代调整

为了支持延伸控制策略在隐私信息提供者和隐私信息接收者之间流转时的动态更新，避免因短板效应导致的隐私泄露，控制策略迭代动态调整包括但不限于：

a) 从接收到的信息中，解析前序隐私信息处理者嵌入的延伸控制策略，生成延伸控制策略集合；
b) 根据后序隐私信息接收者的隐私保护能力、应用场景、数据模态等信息，更新延伸控制策略集合；
c) 更新后的控制集合，连同本级根据控制策略处理过的隐私信息，安全地传递给后序隐私信息接收者；
d) 支持在延伸控制策略集合中，嵌入部分或全部前序隐私处理者信息，应至少包含前一级隐私信息处理者的信息；
e) 可采用可信环境等技术措施，保证延伸控制策略更新调整的安全性。

9.5 策略执行可信验证

为了验证延伸控制策略是否被隐私信息接收者完整正确地执行，策略执行可信验证包括但不限于：

a） 支持隐私信息接收者，验证隐私信息传播链上任一后序隐私信息接收者和隐私信息处理者是否按预期执行其延伸控制策略；
采用日志采集及分析、密码学等技术，保证隐私信息提供者或隐私信息转发者，可验证隐私信息接收者按照预期完整正确地执行了延伸控制策略。

10 隐私按需保护

10.1 概述

隐私按需保护用于隐私信息处理者根据隐私信息所有者或隐私信息提供者的脱敏要求、隐私信息模态以及隐私信息接收者的隐私保护能力等因素，如图4所示，对隐私信息分量进行场景自适应的脱敏和删除操作。若保护效果评估未到预期效果，则可能重新执行隐私按需保护。隐私按需保护包括：

a） 脱敏算法能力评估，对脱敏算法能力从可逆性、信息偏差性、信息损失性、复杂性等方面进行评估；
b） 按需脱敏，隐私信息处理者结合信息模态、业务需求等因素，解析并根据脱敏控制策略，选择脱敏算法集合，进行脱敏处理；
c） 按需删除，隐私信息处理者解析删除控制策略并根据删除控制策略选定删除对象和删除方法，进行删除处理；
d） 按需采用数据使用安全技术,隐私信息处理者根据延伸控制策略，在多方联合数据利用时，隐私信息处理者结合信息模态、业务需求等因素，选择所采用的数据使用安全技术、算法协议以及安全参数，进行多方联合计算。

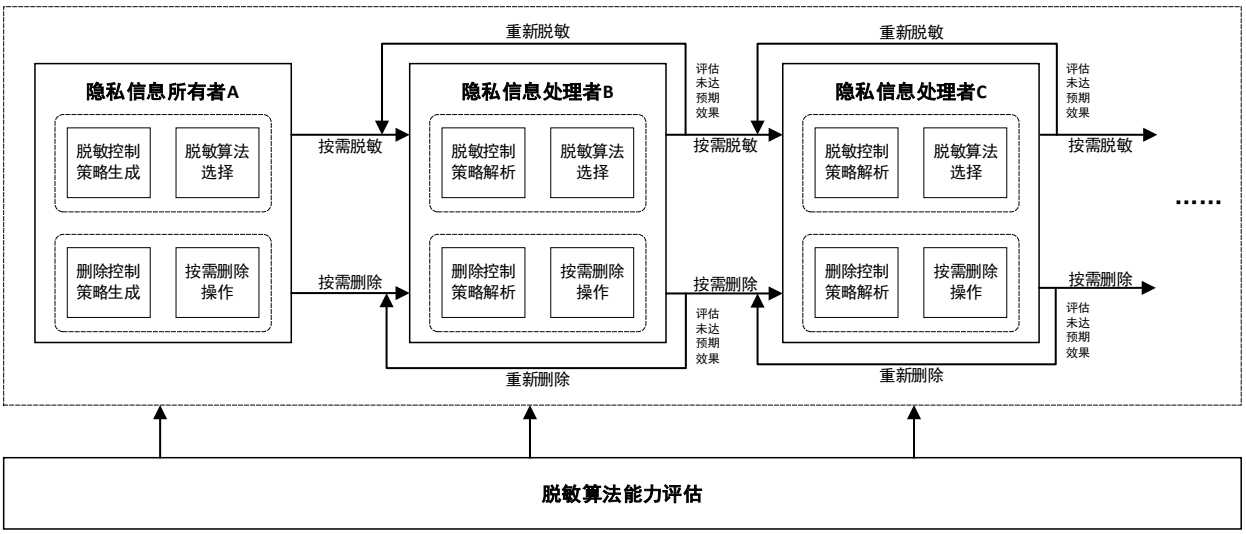


图4 隐私按需保护处理流程

10.2 脱敏算法能力评估

10.2.1 脱敏算法能力评估的指标体系

脱敏算法能力评估的指标体系包括可逆性、信息偏差性、信息损失性和复杂性等四类指标，且基于测评样本基准数据集，对各类脱敏算法进行能力评估。

10.2.1.1 可逆性评估指标

可逆性评估是衡量从脱敏算法处理后信息中复原隐私信息的可能性。由于脱敏旨在保护敏感个人信息，通常情况下脱敏是不可逆的。可逆性度量方法如下：

a） 脱敏算法可逆性，评估隐私信息脱敏使用的是否是不可逆脱敏算法；

b） 脱敏算法参数强度，评估脱敏算法使用的参数强度；
c） 信息还原性，评估通过脱敏后的隐私信息还原出原始隐私信息的程度，例如：恢复信息的准确度、恢复信息的偏差度。

10.2.1.2 信息偏差性评估指标

信息偏差性评估是衡量脱敏算法处理后的信息失真和偏移程度。信息偏差性度量方法如下：
a） 统计偏差性，比较原始数据和脱敏后数据的统计指标，例如：均方差、平均绝对值、KL散度、欧氏距离、余弦距离、峰值信噪比、结构相似性指数、均值、中位数、方差、标准差、最大值、最小值等；
b） 数据分布偏差性，比较原始数据和脱敏后数据的分布差异，例如：分布形状、分位数和累积分布函数等；
c） 模型应用准确性，使用原始数据和脱敏后数据分别构建训练模型，比较模型在验证集或测试集上的效果；
d） 数据随机性分析，评估脱敏算法对隐私信息的随机性影响程度。

10.2.1.3 信息损失性评估指标

信息损失性评估是衡量脱敏算法处理后隐私信息损失部分对可用性的影响程度。信息损失性度量方法如下：
a） 信息熵，信息熵是衡量数据集中信息量的度量指标，通过计算原始数据和脱敏后数据的信息熵，并比较差异；
b） 互信息，互信息是衡量两个随机变量之间相互依赖程度的度量指标，通过计算原始数据和脱敏后数据之间的互信息进行量化评估；
c） 数据分布特征，比较原始数据和脱敏后数据的分布特征的统计指标，例如：均值、方差、分位数等；
d） 数据关联性，计算原始数据和脱敏后数据之间的关联性的度量指标，例如：相关系数、协方差等；
e） 数据可用性，评估脱敏数据在特定应用场景下的可用程度指标，例如：数据分析、模型训练等应用场景。

10.2.1.4 复杂性评估指标

复杂性评估是衡量脱敏算法处理隐私信息所需的资源开销。复杂性度量方法如下：
a） 时间复杂度，用于衡量算法执行所需时间的度量指标，可以通过分析算法中的操作、迭代次数和数据规模等来确定，例如：常数时间$O(1)$、线性时间$O(n)$、对数时间$O(\log n)$、平方时间$O(n^2)$等；
b） 空间复杂度，用于衡量算法执行所需内存空间的度量指标，可以通过分析算法中使用的额外数据结构、变量和递归调用的深度等来确定，例如：常数空间$O(1)$、线性空间$O(n)$、指数空间$O(2^n)$等；
c） 计算资源需求，评估算法执行所需的计算资源，包括CPU执行时间、占用内存等。

10.2.1.5 算法能力综合评估

在脱敏算法能力评估的过程中，需要根据所采用脱敏算法的类别，为每个评估维度设置相应的权重，进行加权计算，得出算法能力的综合评估结果。

10.2.2 脱敏算法可逆性评估

脱敏算法可逆性评估的具体内容如下：
a） 判断脱敏算法支持的数据模态与应用场景，确定算法可逆性对应的权重值；

b） 结合隐私信息数据模态及应用场景，选择10.2.1.1节中的评估指标，衡量脱敏信息的被还原能力，评估内容包括但不限于：恢复信息的准确度、恢复信息偏差度等；
c） 综合考虑算法类别、算法参数、数据模态等因素，设计合理的可逆性评估方案，保证评估结果的准确性和可信性。

10.2.3 脱敏算法信息偏差性评估

脱敏算法信息偏差性评估的具体内容如下：
a） 判断脱敏算法支持的数据模态与应用场景，确定信息偏差性对应的权重值；
b） 结合隐私信息数据模态及应用场景，选择10.2.1.2节中的评估指标，衡量脱敏算法执行前的原始隐私信息与脱敏后的隐私信息之间的偏差程度，评估内容包括但不限于：数据统计、数据应用测试、随机性分析等；
c） 综合考虑数据规模、数据分布、数据模态等因素，设计合理的信息偏差性评估方案，保证评估结果的准确性和可用性。

10.2.4 脱敏算法信息损失性评估

脱敏算法信息损失性评估的具体内容如下：
a） 判断脱敏算法支持的数据模态与应用场景，确定信息损失性对应的权重值；
b） 结合隐私信息数据模态及应用场景，选择10.2.1.3节中的评估指标，衡量脱敏算法执行前的原始隐私信息与脱敏后的隐私信息之间的信息损失程度，评估内容包括但不限于：信息熵、互信息、数据分布特征、数据关联性、信息可用性等；
c） 综合考虑数据可用、数据关联、应用场景等因素，设计合理的信息损失性评估方案，保证评估结果的准确性和有效性。

10.2.5 脱敏算法复杂性评估

脱敏算法复杂性评估的具体内容如下：
a） 判断脱敏算法支持的数据模态与应用场景，确定算法复杂性对应的权重值；
b） 结合隐私信息数据模态及应用场景，选择10.2.1.4节中的评估指标，衡量脱敏算法的执行效率和资源消耗情况，评估内容包括但不限于：时间复杂度、空间复杂度、资源消耗等；
c） 综合考虑平台资源、数据模态、数据规模、数据结构等因素，设计合理的复杂性评估方案，保证评估结果的准确性和有效性。

10.3 按需脱敏

按需脱敏根据数据模态、应用场景、业务需求等要素的不同，进行脱敏控制策略生成、脱敏控制策略解析和脱敏算法选择。

10.3.1 脱敏控制策略生成

脱敏控制策略生成是指，从隐私信息所有者或前序隐私信息提供者处，获取关于相关隐私信息的脱敏要求，并结合应用场景、隐私信息接收者的隐私保护能力等因素，生成适应当前应用场景的脱敏控制策略。脱敏控制策略包括脱敏对象、脱敏等级等，具体如下：
a） 首次脱敏，隐私信息处理者根据隐私信息所有者的脱敏要求，生成脱敏控制策略；
b） 迭代脱敏，隐私信息处理者通过前序隐私信息提供者的脱敏控制策略，并结合当前应用场景要求，生成新的脱敏控制策略；
c） 脱敏控制策略应与脱敏后的隐私信息一起流转。

10.3.2 脱敏控制策略解析

脱敏控制策略解析是指，隐私信息处理者收到脱敏控制策略后，结合隐私信息处理者所处的应用场景，提取当前应用场景所对应的脱敏控制策略，具体如下：

a) 隐私信息处理者解析前验证脱敏控制策略的完整性；
b) 隐私信息处理者依据应用场景识别结果，解析脱敏控制策略后提取当前应用场景所对应的脱敏控制策略。

10.3.3 脱敏算法选择

脱敏算法选择是指，隐私信息处理者根据解析出的脱敏控制策略，对备选脱敏算法集合及其参数进行筛选，具体如下：

a) 隐私信息处理者根据当前应用场景所对应的脱敏控制策略，确定脱敏对象所对应的脱敏等级、脱敏方式等；
b) 根据脱敏等级、脱敏方式，选择脱敏算法及其参数。

10.4 按需删除

10.4.1 删除控制策略生成

删除控制策略生成是指，从隐私信息所有者或隐私信息提供者处，获取关于相关隐私信息的删除要求，并结合数据模态和数据存储方式等因素，生成删除控制策略，包括删除对象、删除方法、删除范围等，具体如下：

a) 删除对象确定，删除对象包括各个隐私信息处理者留存的隐私信息正本或隐私信息副本，根据隐私信息所有者或隐私信息提供者的删除要求，生成由删除对象组成的集合、删除范围的集合；
b) 删除方法确定，根据隐私信息所有者或隐私信息提供者的删除要求，确定删除方法集合；
c) 生成由删除对象集合、删除方法集合、删除范围集合等组成的删除控制策略。

10.4.2 删除控制策略解析

删除控制策略解析是指，隐私信息处理者收到删除控制策略后，提取删除控制策略，具体如下：

a) 隐私信息处理者解析前验证删除控制策略的完整性；
b) 隐私信息处理者解析删除控制策略后提取删除对象集合、删除方法集合、删除范围集合等信息。

10.4.3 按需删除操作

隐私信息处理者根据解析出的删除对象集合、删除方法集合、删除范围集合等内容，执行删除操作，具体如下：

a) 从删除方法集合中选择删除方法，再根据删除方法、删除等级等，构造删除指令；
b) 根据删除对象集合，确定本地删除对象的正本信息、多副本信息、分散存储信息、多备份信息等所处的设备，然后将删除指令发送到对应的设备；
c) 根据删除范围集合，确定其他存放删除对象的正本信息、多副本信息、分散存储信息、多备份信息等所处的设备，然后将删除指令发送到对应的设备。

11 保护效果评估

11.1 概述

保护效果评估是对从脱敏后的隐私信息中恢复损失信息的难度，或者恢复已删除隐私信息的可能性进行评价。如图5所示，若保护效果评估未达到预期效果，则可能重新执行隐私信息抽取与度量、隐私度量动态调整、隐私延伸控制、隐私按需保护。保护效果评估包括：

a) 脱敏效果评估的指标体系，采用可逆性、信息偏差性和信息损失性等评估指标。具体见章节10.2.1.1,10.2.1.2及10.2.1.3；
b) 单次脱敏效果评估，通过分析脱敏算法执行前后的信息，衡量已脱敏的隐私信息分量的可恢复程度；
c) 基于数据挖掘的脱敏效果评估，通过收集特定个人一定时间内的脱敏信息，采用数据挖掘技术试图推算出已脱敏的隐私信息分量；
d) 脱敏系统效果评估，通过收集若干特定个人或所有个人的一定时间内的脱敏信息，采用数据挖掘技术试图推算特定个人的已脱敏的隐私信息分量；
e) 删除效果评估,采用数据恢复技术试图还原被删除的隐私信息分量。

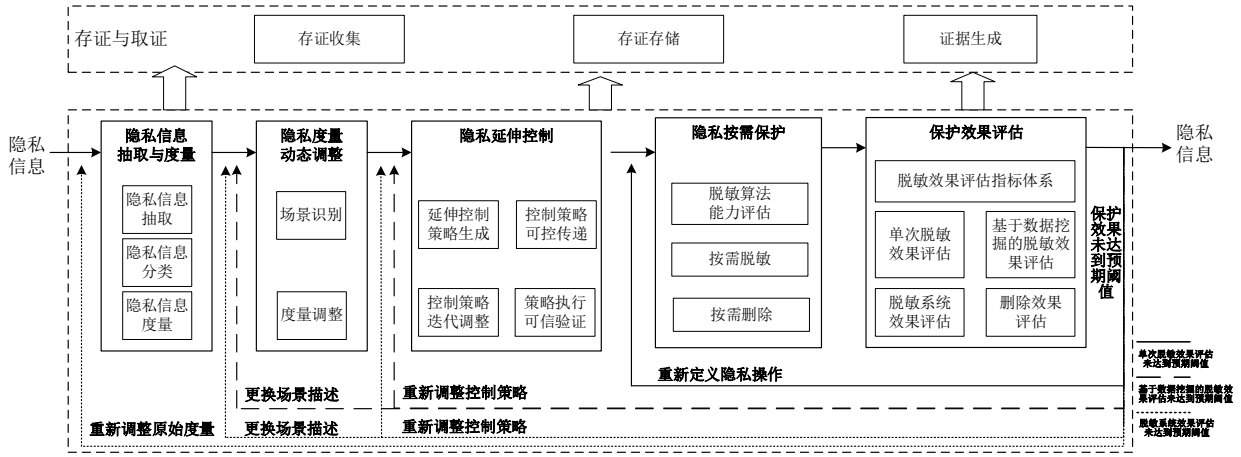


图5 保护效果评估执行策略思路

11.2 单次脱敏效果评估

单次脱敏效果评估，将执行脱敏算法前的隐私信息与脱敏后的隐私信息进行可逆性、信息偏差性和信息损失性的评估，具体如下：

a) 依据执行脱敏算法后的数据模态、数据规模和应用场景，确定可逆性、信息偏差性和信息损失性对应的权重值；
b) 依据数据模态和应用场景，选择10.2.1.1节中的评估指标，评估已脱敏的隐私信息分量的被还原能力；
c) 依据数据规模及统计特性，选择10.2.1.2节中的评估指标，评估脱敏算法执行前的原始隐私信息与脱敏后的隐私信息之间的偏差程度；
d) 依据数据信息量和关联性，选择10.2.1.3节中的评估指标，评估脱敏算法执行前的原始隐私信息与脱敏后的隐私信息之间的信息损失程度；
e) 考虑数据规模、数据分布、数据模态、应用场景等因素，设计合理的单次脱敏效果评估方案，保证评估结果的准确性和可信性；
f) 当单次脱敏效果评估结论的可逆性指标未达到脱敏效果期望阈值，需反馈到隐私按需保护组件，根据10.3.3节要求，重新选择脱敏算法，并设置该脱敏算法的参数；当单次脱敏效果评估结论的可逆性指标符合脱敏效果期望阈值，但其信息偏差性或信息损失性指标未达到脱敏效果期望阈值，需反馈到隐私按需保护组件，根据10.3.3节要求，修改脱敏算法的参数；若多次反馈后，仍未达到脱敏效果期望阈值，需反馈到隐私延伸控制组件，根据9.2节要求，生成调整后的延伸控制策略。

11.3 基于数据挖掘的脱敏效果评估

基于数据挖掘的脱敏效果评估，是指对采用数据挖掘技术分析特定个人一定时间内的已通过单次脱敏效果评估的脱敏信息，以推算出已脱敏的隐私信息分量程度的评估，具体如下：

a) 选取数据挖掘算法，对特定个人一定时间内的脱敏信息进行挖掘分析，推断出特定个人的被脱敏的隐私信息；

b) 评估推断出的特定个人的隐私信息，与其对应的真实隐私信息进行偏差性、损失性的对比分析；
c) 当基于数据挖掘的脱敏效果评估结论的信息偏差性未达到脱敏效果期望阈值，需反馈到隐私延伸控制组件，根据9.2节要求，生成调整后的延伸控制策略；当基于数据挖掘的脱敏效果评估结论的信息偏差性达到脱敏需求阈值，而信息损失性未达到脱敏效果期望阈值，需反馈到隐私度量动态调整组件，根据8.2节和8.3节要求，分别修正场景识别、度量调整机制。

11.4 脱敏系统效果评估

脱敏系统效果评估，是指对采用数据挖掘技术分析若干特定个人或所有个人一定时间内的已通过基于数据挖掘的脱敏效果评估的脱敏信息，以推算出特定个人已脱敏的隐私信息分量程度的评估，具体如下：

a) 选取数据挖掘算法，对若干特定个人或所有个人一定时间内的脱敏信息进行挖掘分析，推断出特定个人的被脱敏的隐私信息；
b) 评估推断出的特定个人的隐私信息，与其对应的真实隐私信息进行偏差性、损失性的对比分析；
c) 当脱敏系统效果评估结论的信息偏差性未达到脱敏效果期望阈值，需反馈到隐私延伸控制组件，根据9.2节要求，生成调整后的延伸控制策略；当脱敏系统效果评估结论的信息偏差性达到脱敏效果期望阈值，而信息损失性未达到脱敏效果期望阈值，需反馈到隐私度量动态调整组件，根据8.2节和8.3节要求，分别修正场景识别、度量调整机制；若多次反馈后，仍未达到脱敏效果期望阈值，需反馈到隐私信息抽取与度量组件，根据7.2、7.3和7.4节要求，分别修正隐私信息的抽取、分类和度量。

11.5 删除效果评估

删除效果评估，是指对被隐私信息处理者删除的隐私信息的可恢复程度的评价，具体如下：

a) 删除控制策略完备性，用于评估隐私信息所有者、隐私信息提供者和隐私信息发布者生成与发送的删除控制策略是否涵盖被删除对象所有副本；
b) 删除操作合规性，用于评估隐私信息处理者的删除方法是否符合删除控制策略的要求；
c) 删除不可恢复性，用于评估隐私信息处理者执行删除操作后被删除对象的不可恢复程度。

12 存证与取证

12.1 概述

存证与取证主要是对隐私计算其他功能组件的运行、隐私信息的处理等情况进行可信记录，以便于开展内部监测、或者按照法律法规接受外部监管者的合规审查、侵权行为追踪溯源的服务请求提供必要的证据和技术接口，包括：

a) 存证收集，用于从各个隐私计算框架的核心组件中收集执行过程和执行效果的事项记录；
b) 存证存储，用于对从各个隐私计算框架的核心组件中收集到的存证信息进行存储；
c) 证据生成，根据证据服务请求，对隐私计算系统的各类存证信息进行多源检索，形成证据链，并将证据返回证据服务请求方。

12.2 存证收集

存证收集主要用于收集隐私计算框架的各核心组件的操作记录，包括但不限于：

a) 存证收集范围，覆盖隐私信息抽取与度量、隐私度量动态调整、隐私延伸控制、隐私按需保护、保护效果评估等组件的执行过程和执行效果；
b) 隐私信息抽取与度量组件的相关存证信息收集，具体如下：
   1) 收集隐私信息抽取过程的相关日志信息，包括但不限于：识别与抽取执行主体、执行时间、识别与抽取结果，支撑审计隐私信息抽取功能的正确性、完备性；
   2) 收集隐私信息分类过程的相关日志信息，包括但不限于：信息模态、所依据的分类标准、分类方法、分类时间、分类结果，支撑审计数据分类的合规性、分类异常问题溯源；

3) 收集隐私信息度量过程的相关日志信息，包括但不限于：量化评估模型及其参数、输入数据和输出量化结果的执行主体、量化结果等，支撑隐私信息度量方法改进、度量异常问题溯源。

c) 隐私度量动态调整组件的相关存证信息收集，具体如下：
1) 收集场景识别过程的相关日志信息,包括但不限于:场景识别执行主体、识别模型及参数、识别执行时间、场景识别结果，支撑审计场景识别功能的正确性；
2) 收集度量调整过程的相关日志信息，包括但不限于：度量调整操作执行主体、度量调整方法及参数、度量调整结果，支撑审计度量调整功能的正确性。

d) 隐私延伸控制组件的相关存证信息收集，具体如下：
1) 收集延伸控制策略生成的相关日志信息，包括但不限于：生成该策略的隐私信息处理者身份标识、延伸控制意图、延伸控制策略内容等，支撑审计延伸控制意图和延伸控制策略之间的匹配性；
2) 收集控制策略可控传递过程的相关日志信息，包括但不限于：隐私信息提供者身份标识、隐私信息接收者身份标识、策略传递会话日志信息、数字签名等，支撑延伸控制一致性判定、控制策略异常问题溯源；
3) 收集控制策略迭代调整过程的相关日志信息，包括但不限于：隐私信息处理者的隐私保护能力、应用场景、数据模态、延伸控制策略更新内容等信息，支撑审计控制策略迭代调整功能的正确性；
4) 收集策略执行可信验证过程的相关日志信息，包括但不限于：验证发起方和被验证方的身份信息、验证方法、验证结果等。如果存在验证结果异常，则记录异常内容，支撑异常行为的检测与问题溯源。

e) 隐私按需保护的相关存证信息收集，具体如下：
1) 收集脱敏控制策略生成的相关日志信息，包括但不限于：生成该策略的隐私信息处理者身份标识、隐私信息接收者身份标识、脱敏对象、脱敏等级等，支撑审计脱敏意图和脱敏控制策略之间的匹配性；
2) 收集脱敏操作的相关日志信息，包括但不限于：执行该策略的隐私信息处理者身份标识、脱敏控制策略解析结果、选择的脱敏算法及其参数、脱敏结果等，支撑审计脱敏操作内容和脱敏控制策略之间的匹配性；
3) 收集删除控制策略生成的相关日志信息，包括但不限于：生成该策略的隐私信息处理者身份标识、删除对象、删除方法、删除范围等，支撑审计删除意图和删除控制策略之间的匹配性；
4) 收集删除操作的相关日志信息，包括但不限于：执行该策略的隐私信息处理者身份标识、删除控制策略解析结果、删除方法、删除指令、删除结果等，支撑审计删除操作内容和删除控制策略之间的匹配性。

f) 保护效果评估的相关存证信息收集，具体如下：
1) 收集单次脱敏效果评估的相关日志信息,包括但不限于:执行该效果评估处理者身份标识、可逆性评估方法与结果、信息偏差性评估方法与结果、信息损失性评估方法与结果等，支撑审计单次脱敏的合规性和有效性；
2) 收集基于数据挖掘的脱敏效果评估的相关日志信息，包括但不限于：执行该效果评估处理者身份标识、特定个人在一定时间内的脱敏信息、信息偏差性评估方法与结果、信息损失性评估方法与结果等，支撑审计特定个人的隐私信息脱敏的合规性和有效性；
3) 收集脱敏系统脱敏效果评估的相关日志信息，包括但不限于：执行该效果评估处理者身份标识、若干特定个人或所有个人在一定时间内的脱敏信息、信息偏差性评估方法与结果、信息损失性评估方法与结果等，支撑审计脱敏系统脱敏的合规性和有效性。

g) 存证信息采用标准描述格式，存证信息的收集采用REST API、gRPC等分布式系统通信技术，提供通用的存证接口，支撑跨系统互联互通。

12.3 存证存储

存证存储是对收集的存证信息进行高效的组织管理，支撑存证信息的高效检索和充分利用，包括：

a) 存储方式，本地自存证，或第三方存证等存证类型；
b) 存证完整性，保证收集的存证信息齐全，确保存证信息被成功存储，在丢失或损坏存证信息的情况下从原渠道补全对应的存证信息；
c) 存证机密性，采用访问控制技术控制存证信息的访问主体、访问路径、访问时间、访问的信息数量；采用密码技术对存证信息进行加密防止镜像拷贝、拖库等非法获取；
d) 存证可信性，保证存证信息不被篡改、不被随意删除；
e) 存证可靠性，采用RAID技术或双活等机制，实现存证信息的冗余灾备。

12.4 证据生成

证据生成是根据证据服务请求对收集到隐私计算框架各组件的存证信息进行检索、数据预处理、证据要素封装、证据链构建的过程，支撑对隐私计算的操作行为进行合规审计、监管、异常事件分析与溯源等，包括：

a) 证据请求响应，接收请求方的证据获取请求，对请求方进行身份验证，解析证据获取事项的要求；
b) 证据完备性，根据证据获取的要求，对各个存证存储系统中的存证信息进行多源证据查找，获取所有相关证据；
c) 多源证据生成，对查找得到的多源存证信息，按照请求事项和时序进行整理并生成证据，采用数字签名技术实现生成证据的不可篡改和不可否认性；如果有机密性要求，采用请求和响应双方预先约定的协议实现密钥协商和传输加密。

附 录 A
（资料性）
隐私信息描述示例

A.1 概述

市场监管行业数据中个人信息所涉及的隐私信息可由隐私信息六元组描述，具体如下：

a) 隐私信息向量，根据文件标识符确定信息的类型，并根据语义特征将信息进行拆分，得到n个在语义上不可分割的隐私信息分量，则隐私信息向量可以表示为$I=(id,i_1,\ldots,i_k,\ldots,i_n)$,其中$k$为取值范围从 1 到$n$的正整数；

b) 约束条件集合，由隐私信息分量对应的约束条件向量组成的集合$\theta=\{\theta_1,\theta_2\cdots,\theta_k,\ldots,\theta_n\}$，用于描述在不同场景下实体访问隐私信息分量$i_k$所需的访问权限。根据隐私信息向量中的隐私信息分量的应用场景，可对隐私信息分量设置相应的约束条件向量$\theta_k=(u_k,t_k,d_k,n_k)$，表示第$k$个隐私信息分量的约束条件，其中，$u_k$表示访问者列表，$t_k$表示访问时间，$d_k$表示访问设备，$n_k$表示网络标识；

c) 隐私属性向量，通过预先标记或隐私保护程度量化操作函数，结合约束条件集合，对隐私信息向量进行隐私度量，度量结果存入隐私属性向量$A=(a_1,a_2,\cdots,a_n,\cdots,a_m)$。可假设初始所有隐私属性分量为 1（假设隐私属性分量的范围为 0 到 1，隐私属性分量越小，其对应的隐私信息分量的保护程度越高）；

d) 广义定位信息集合，分别获取隐私信息分量在信息中的广义定位信息向量γ_1、γ_2、……、γ_n，由此生成广义定位信息集合$\Gamma=\{\gamma_1,\gamma_2,\cdots,\gamma_k,\cdots,\gamma_n\}$；

e) 审计控制信息集合，分别获取隐私信息分量的审计控制信息向量$\omega_1,\omega_2,\cdots,\omega_n$。在初始化阶段，审计控制信息向量可以为空，记录当前持有者对隐私信息分量进行的所有操作，由此生成审计控制信息集合$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$；

f) 传播控制操作集合，针对隐私属性向量和约束条件集合，根据操作判别函数或人工标记生成隐私信息向量及其组合的传播控制操作集合$\Psi=\{\Psi_1,\Psi_2,\cdots,\Psi_n\}$。在初始化阶段，传播控制操作向量可以为空，根据流转过程中对隐私信息分量的传播要求，逐渐添加得到传播控制操作向量$\Psi_2,\cdots,\Psi_n$共同组成传播控制操作集合。

A.2 隐私信息描述六元组生成过程

本附录以文本、图像模态数据为例，介绍隐私信息描述的生成方法和过程，供设计实现隐私信息描述与处理功能时参考。针对待处理的多模态数据，进行隐私抽取，确定隐私信息分量模态，并进行定位，生成隐私信息向量和广义定位信息集合；通过隐私数据分类分级规则对识别和抽取的隐私信息分量进行分类，根据分类结果确定隐私信息向量的约束条件集合；根据约束条件集合以及隐私数据分类分级规则对隐私信息向量进行隐私度量，生成隐私属性向量和传播控制操作集合；记录对隐私信息向量执行的所有操作，生成审计控制信息集合。隐私信息描述六元组生成过程如图A.1所示。

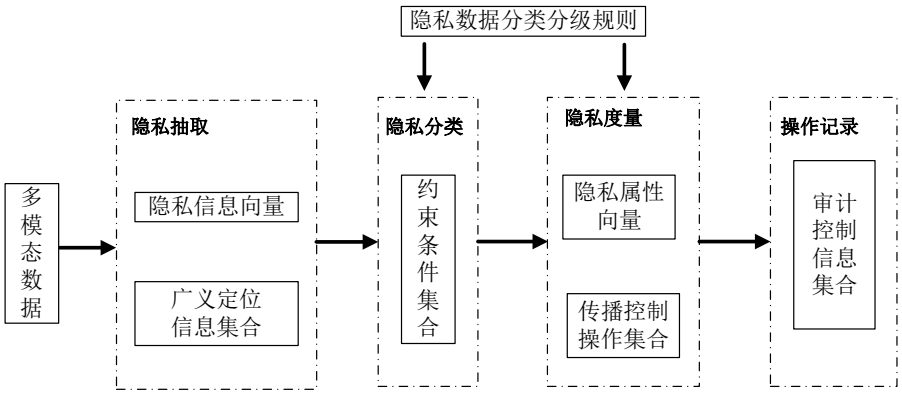


图 A.1 隐私信息描述六元组生成过程

A.3 文本类隐私信息描述生成示例

A.3.1 隐私信息向量

隐私信息向量生成过程如下：

a) 隐私信息处理者根据文件标识符等信息确定信息的类型，并读取其内容。例如：可以通过Apache poi 等工具读取 Word 文档信息、Excel 表格信息；通过 Spire.PDF 等工具读取 PDF 文档信息；

b) 通过正则文本匹配、自然语言处理识别文件中的隐私信息并进行切割。以文本信息“张三和李四去中关村参加活动”为例，使用基于 BiLSTM+CRF 模型的命名实体识别算法对姓名、组织名、地名等实体进行识别；

c) 生成隐私信息向量$I = (id, 张三, 和, 李四, 去, 中关村, 参加, 活动)$。

A.3.2 约束条件集合

约束条件集合生成过程如下：

d) 隐私信息处理者获取文本类数据对应的分类分级规则；

e) 根据当前隐私信息处理者的意愿，生成隐私信息向量对应的约束条件集合。例如：若当前隐私信息处理者仅向 197.224.*.*网段的用户赋予文档读写权限，则非该网段下的用户无法读取该文件；

f) 生成约束条件集合$\theta_1 = (读写, 197.224.*.*)$。

A.3.3 隐私属性向量

隐私属性向量生成过程如下：

g) 隐私信息处理者根据文本类数据对应的分类分级规则和隐私信息分量的约束条件集合，确定待处理文本信息对应的分类分级规则；

h) 根据约束条件集合、分类分级规则以及文档创建者的意愿，通过预先标记或隐私保护程度量化操作函数，依次计算隐私信息向量和隐私信息分量组合对应的隐私属性向量，生成各个隐私信息分量的隐私属性向量；

i) 如度量结果为 0.4，包含X_1至X_n的隐私属性向量可表示为$a_1 = (X_1, \ldots, X_n, 0.4)$。

A.3.4 广义定位信息集合

广义定位信息集合生成过程如下：

a) 隐私信息处理者采用合适的位置标识信息，对隐私信息在文档中的位置进行编码。例如：

1) Word 文档可利用页码、段落、行数、起始位和终止位等表示定位信息；

2) Excel 表格可利用表单号、行号、列号等表示定位信息。

b) 生成隐私信息分量对应的广义定位信息向量。例如：对于 Word、Excel 文档，可使用 Apache poi 工具对隐私信息向量进行索引，统一设置字符在文档中的范围标签和批注引用，生成广义定位信息集合；

c) 生成广义定位信息集合Γ，例如：

1) $\gamma_1 = (X_1, P_{11}, S_3, R_2, 0, 4)$表示记为$X_1$隐私信息分量“张三”，位于文本信息的第 11 页、第 3 段中的第 2 行，起始位为 0，终止位为 4；

2) $\gamma_2 = (X_2, T_1, R_5, C_4)$表示记为$X_2$隐私信息分量“321128202502012921”，位于表格的表单 1、第 5 行中的第 4 列的单元格中。

A.3.5 审计控制信息集合

审计控制信息集合生成过程如下：

a) 隐私信息处理者记录其对文档的隐私信息分量进行的所有操作，并生成审计控制信息集合；

b) 将生成的审计控制信息集合存储于对应的隐私信息描述六元组。

A.3.6 传播控制操作集合

传播控制操作集合生成过程如下：

a) 隐私信息处理者获取文本类数据对应的分类分级规则和隐私信息分量的约束条件集合、隐私属性向量；
b) 根据当前隐私信息处理者流转共享意愿，生成隐私信息向量对应的传播控制操作集合。例如：当前的传播控制操作集合$\Psi = \{\Psi_1, \Psi_2, \Psi_3\}$，假设当前隐私信息处理者只希望隐私信息分量“中关村”被复制转发，则可新增一个传播控制操作向量$\Psi_4 = (\text{转发}, \text{复制})$，生成新的传播控制操作集合$\Psi = \{\Psi_1, \Psi_2, \Psi_3, \Psi_4\}$。

隐私信息六元组的生成结果如图A.2所示。识别文档中的隐私信息实体，例如：“北京市消防局”记为X_1，X_1的位置在文档中的第2页、第2段、第2行上，起始位置是35，终止位置是41。提取外部文本数据分类分级规则，结合使用意愿，得到其对应的隐私属性向量为$\mathrm{a}_1 = (X_1, 0.4)$，表示隐私信息分量$\mathrm{X}_1$的隐私属性度量结果为0.4。约束条件集合$\theta_1$约束该文档只能在197.224.*.*网段中被查看，并且Ψ_1控制后续流转过程中该隐私信息分量只能被转发、复制。传播控制操作集合$\omega_1 = (UID_1, \text{复制}, \text{转发}; UID_2, \text{转发}, \text{修改})$表示隐私信息分量“北京市消防局”先后被唯一标识为$UID_1$和$UID_2$的用户访问和操作过，其中“$UID_1$，复制，转发”表示隐私信息分量“北京市消防局”被用户UID_1执行了复制、转发的操作；当该隐私信息分量传播至用户UID_2，则被执行了转发和修改的操作。

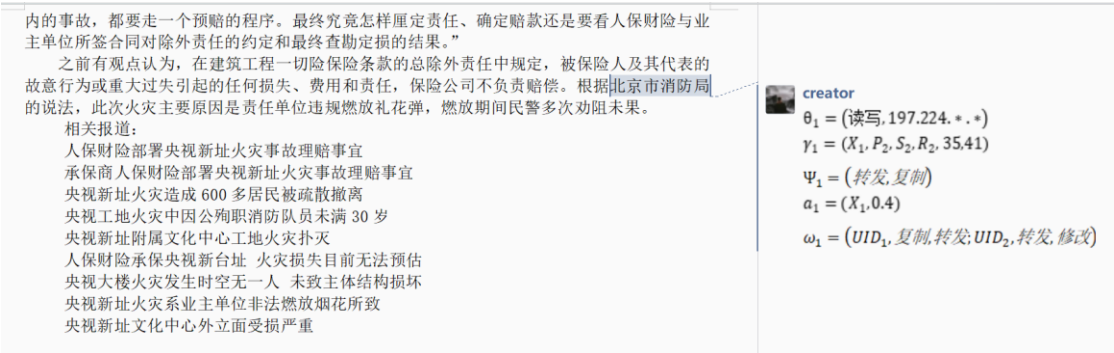


图 A.2 Word 文档隐私信息描述生成结果

A.4 图片类隐私信息描述生成示例

A.4.1 隐私信息向量

图片中的隐私信息向量生成过程如下：

a) 在图片分享过程中，图片持有者使用 ExifTool、exifread 等开源工具/函数库读取图片 EXIF 数据中的拍摄时间、拍摄经纬度、焦距等敏感信息；
b) 通过图像处理算法识别图片像素区域中的敏感内容。例如：一张图片同时出现张三和李四在医院门口下私家车，图片拍摄者为张五。其中，张三人脸、李四人脸、私家车车牌、医院标识即为敏感内容；
c) 将识别的敏感内容作为隐私信息分量（张三人脸、李四人脸、私家车车牌、医院标识）存储在 Exif 信息的 IFD（Image File Directory）结构的自定义区域，命名为隐私信息分量 Privacy Info Entry 中。

A.4.2 约束条件集合

图片中的约束条件集合生成过程如下：

a) 可通过一些外在信息判定敏感信息的约束条件，例如，图片像素的张三人脸区域、李四人脸区域、私家车车牌号区域、医院标识区域，访问者-访问客体信息 V = {<张五, 私家车车牌号区域>, … , <赵六, 医院标识区域>}约束条件；
b) 可通过 ExifTool、exifread 等开源工具/函数库，将约束条件写入 Exif 信息的 IFD 结构自定义区域中的约束条件 Constraint Entry 中。

A.4.3 隐私属性向量

图片中的隐私属性向量生成过程如下：

a) 通过图片内容隐私检测算法、合影人员亲密度检测算法等工具识别私家车车牌号区域、张三人脸区域、赵四人脸区域、医院标识区域的敏感程度；
b) 根据用户的隐私需求设置图片像素的私家车车牌号区域的隐私级别为 Level 1；张三人脸区域、赵四人脸区域的隐私级别为Level 2；医院标识区域为Level 3；
c) 通过ExifTool、exifread等开源工具/函数库，写入在Exif信息的IFD结构的隐私属性向量Privacy Attribution Entry中。

A.4.4 广义定位信息集合

图片中的广义定位信息集合生成过程如下：

a) 通过图像分割算法标记图片像素的张三人脸区域、李四人脸区域、私家车车牌号区域、医院标识区域，将每个区域的坐标、位置等元数据存储的EXIF Entry名称等广义定位信息；
b) 通过ExifTool、exifread等开源工具/函数库，将这些广义定位信息记录在Exif信息的IFD结构的广义定位信息Location Entry中。

A.4.5 审计控制信息集合

图片中的审计控制信息集合生成过程如下：

a) 将用户对图片中张三人脸区域、李四人脸区域、私家车车牌号区域、医院标识区域等隐私信息的复制、转发、滤镜、剪切等操作进行记录；
b) 通过 ExifTool、exifread 等开源工具/函数库，将记录写入 IFD 结构的审计控制信息 Audit Entry 中，审计控制操作 Audit Entry 下可以设置子 IFD，用于动态扩充后续审计控制操作集合；
c) 当出现违规行为时，可通过定制程序读取所有审计控制操作，通过比对隐私信息的约束条件进行违规行为判定。

A.4.6 传播控制操作集合

图片中的传播控制操作集合生成过程如下：

a) 将图片像素的张三人脸区域、李四人脸区域、私家车车牌号区域、医院标识区域的转发、删除、复制等传播操作进行记录；
b) 通过ExifTool、exifread等开源工具/函数库，记录在Exif信息的IFD结构的传播控制操作Control Entry中，传播控制操作Control Entry下可以设置子IFD，用于动态扩充后续传播控制操作集合；
c) 在图片分享过程中，通过定制程序读取Control Entry信息，并根据传播控制信息对图片的传播进行控制；
d) 根据已有的传播控制操作集合和约束条件集合，结合当前文档操作者的意愿，写入新的传播控制操作分量，生成新的传播控制操作集合。例如，当前的传播控制操作集合$\Psi = \{\Psi_1, \Psi_2, \Psi_3\}$，假设当前持有者想要隐私信息分量“医院标识”只能被复制、转发，则可新增一个传播控制操作向量$\Psi_4 = (\text{转发}, \text{复制})$，生成新的传播控制操作集合$\Psi = \{\Psi_1, \Psi_2, \Psi_3, \Psi_4\}$。

附 录 B
（资料性）
迭代延伸控制示例

B.1 概述

迭代延伸控制，通过延伸控制策略的协同，确保不同隐私信息处理者能对接收到的隐私信息提供合适的保护能力，避免因数据流通与共享导致隐私泄露。迭代延伸控制主要包括迭代控制策略生成、控制策略可控传递、控制策略动态调整、控制策略可信验证等环节。

延伸控制策略根据隐私信息所有者或隐私信息提供者的控制意图、当前隐私信息使用者的控制约束和隐私信息接收者的隐私保护能力等因素生成。延伸控制策略在全生命周期中是动态调整的，且随着隐私信息同步流转传递。迭代延伸控制贯穿于隐私信息从收集到删除的全过程。此外，在数据流通与共享过程中，需要对其关联的延伸控制策略进行存证，以支撑迭代延伸控制的合规审查、异常溯源。

数据交易场景下的信息脱敏迭代延伸控制流程如图B.1所示，具体表述如下：

a) 隐私信息提供者A，通过数据交易流程将包含延伸控制策略的隐私信息集合传递给隐私信息接收者B。因为延伸控制策略中脱敏控制约束，B接收到的信息量小于等于A原始信息的信息量；

b) 隐私信息接收者B，根据A传递的延伸控制策略对数据集进行脱敏、使用、存储等操作。当B向隐私信息接收者C和D分别提供隐私信息时，根据C和D的应用场景、隐私保护能力等因素，更新生成新的延伸控制策略，并随数据集交付给C和D。

   1) C和D本地脱敏后信息量一致，但由于延伸控制策略内容的不同，导致可实际使用的信息量不一致；

   2) 由于脱敏策略的不同，C和D本地脱敏后信息量不一致。

c) C和D后续的数据交易流程，以此类推。

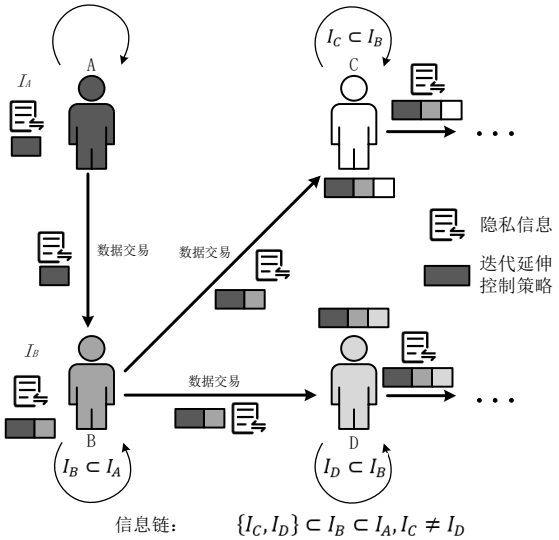


图 B.1 脱敏延伸控制流程示意图

B.2 隐私信息全生命周期中的迭代延伸控制

本节通过数据交易场景示例（如图B.1所示），介绍迭代延伸控制在隐私信息全生命周期中从收集到删除的过程中所发挥的作用。

a) 收集操作，隐私信息收集者A以六元组的形式描述采集原始隐私信息，生成隐私信息描述。隐私信息描述包含初始的延伸控制策略。例如：隐私信息描述六元组中的传播控制操作集合，可对该隐私信息的共享范围进行约束；

b) 脱敏操作，隐私信息收集者A根据隐私信息描述、应用场景等因素，生成延伸控制策略，并据此对原始隐私信息进行场景自适应的按需脱敏；

c） 存储操作，隐私信息收集者A可根据隐私信息所有者要求的存储时间、删除方式等生成删除相关的延伸控制策略，保障隐私信息所有者的删除权；
d） 使用操作，隐私信息收集者A在延伸控制策略中的使用约束下，对隐私信息进行受控使用；
e） 交换操作，隐私信息收集者A和隐私信息接收者B启动数据交易流程。A根据隐私数据模态、应用场景以及隐私信息接收方B的隐私保护能力等因素，调整更新延伸控制策略，并执行控制策略可控传递，将隐私信息及延伸控制策略交付给B。B通过数据交易流程，结合延伸控制策略，再将数据分别转售于C和D。后续的数据交易流程以此类推;
f） 删除操作，隐私信息收集者A收到隐私信息所有者的删除请求，通过迭代延伸控制机制，通知B、C、D，以及其他相关的隐私信息处理者采取合适的删除方法，删除相关的删除对象。

B.3 脱敏延伸控制

脱敏延伸控制关注隐私信息跨域流转场景下，相关隐私信息处理者对接收到的原始隐私信息开展脱敏操作管理和控制工作，包括脱敏延伸控制策略的生成、脱敏延伸控制策略的可控传递、脱敏延伸控制策略的动态调整、脱敏延伸控制策略可信验证等步骤。其过程如图B.2所示。实现脱敏延伸控制的关键环节如下：

a） 脱敏延伸控制策略的生成：
   1） 当个人信息源域接收到用户的脱敏需求时，通过隐私信息所有者或前序隐私信息提供者导入脱敏需求，明确脱敏数据范围、脱敏目的、敏感信息类型等关键要素，并利用隐私信息流转历史记录，查询相关隐私信息的流转历史，以确定脱敏延伸的覆盖范围；
   2） 基于查询结果，生成脱敏延伸控制策略。这一过程中，需明确脱敏算法选择、脱敏的详细要求，并执行初步的数据脱敏处理，为后续的脱敏延伸奠定基础。
b） 脱敏延伸控制策略的可控传递：
   1） 采用消息验证码或数字签名等技术，对流转分发的隐私信息及其相关的脱敏延伸控制策略进行处理，保证在隐私信息提供者和隐私信息接收者之间传递过程中的完整性、机密性、不可剥离性和不可抵赖性；
   2） 当后继隐私信息处理者接收到脱敏延伸控制策略后，需要按照策略要求更新自身的处理策略，并执行相应的脱敏操作。
c） 脱敏延伸控制策略的可信验证：
   1） 采用逐级回传机制，确保每个数据使用实体在正确执行脱敏操作后，能够返回不可伪造的可信验证信息；
   2） 隐私信息所有者负责收集并分析这些验证信息，确保整个脱敏延伸过程的可靠性和有效性；
   3） 一旦完成所有验证，形成此次脱敏延伸控制事件的闭环，确保整个过程的可追溯性和可审计性。

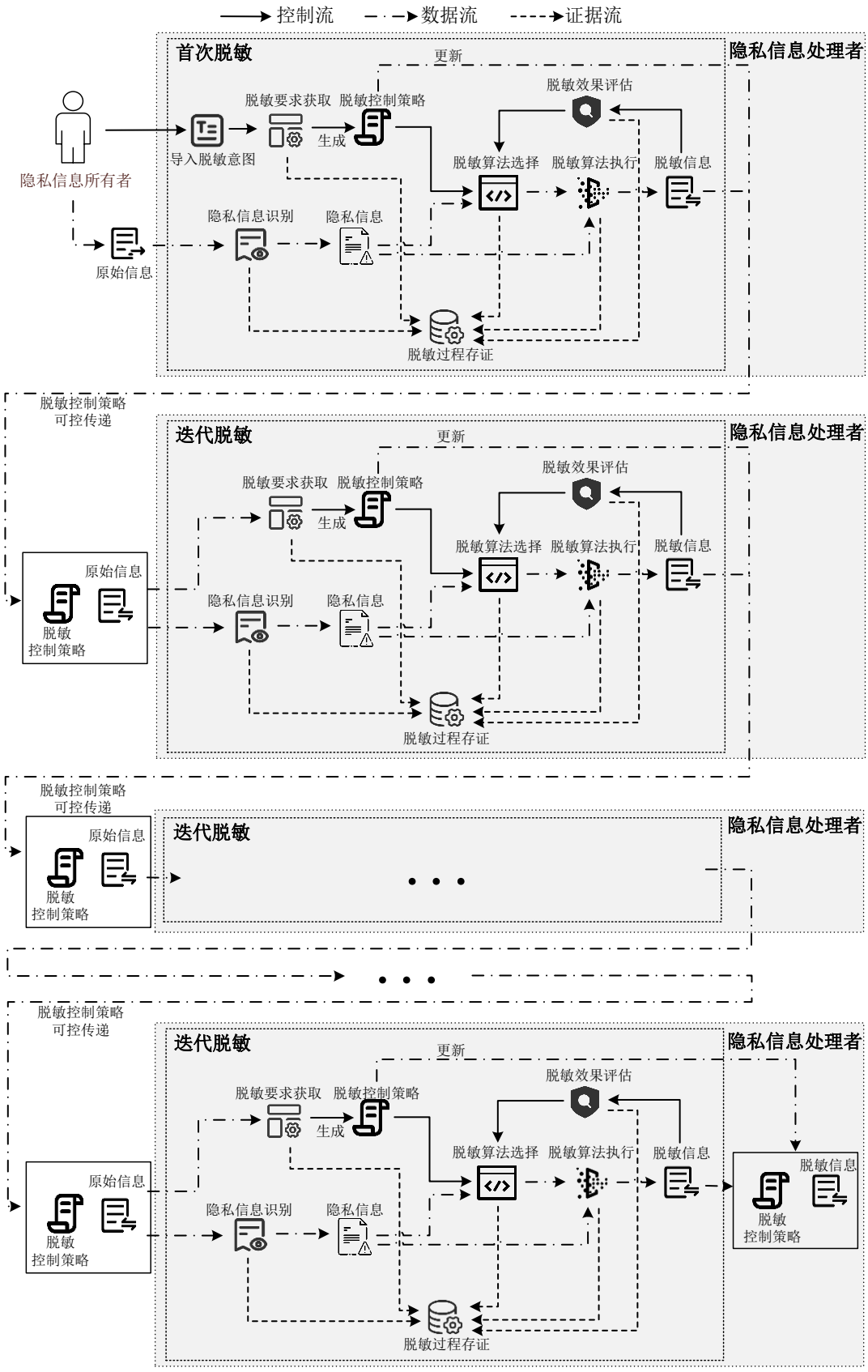


图 B.2 脱敏延伸控制流程示意图

d) 脱敏延伸控制策略的动态调整：

1) 在脱敏延伸控制策略传递和执行过程中，若发现任何异常或不符合预期的情况，如策略传递失败、验证信息回传异常等，需要立即启动动态调整机制；
2) 根据异常信息，对脱敏延伸控制策略进行必要的修改和调整，以确保整个脱敏延伸过程的顺利进行；
3) 通过通知重传、策略更新等手段，确保所有相关数据使用实体能够及时接收到最新的脱敏延伸控制策略，并执行相应的操作；
4) 当继续往后序隐私信息接收者共享信息时，脱敏延伸控制策略的动态调整过程依次类推。

B.4 删除延伸控制

按需删除是实现《个人信息保护法》规定的删除权的技术手段。个人信息源域（即收到隐私信息所有者删除请求的隐私信息处理者），根据隐私信息所有者针对此次隐私信息删除的删除意图，生成隐私信息删除通知，保证所有相关的隐私信息处理者（即存储被删除对象的隐私信息删除者）能安全可信地收到删除通知，并对删除对象执行删除操作，实现隐私信息的完备删除。其过程如图B.3所示。实现删除延伸控制的关键环节如下：

a) 删除延伸控制策略的生成：
   1) 个人信息源域接收用户的删除意图，导入删除意图，利用隐私信息流转历史记录，查询相关删除对象的流转历史（即数据共享情况），确定删除通知范围，确保删除通知在全域内的可达性；
   2) 以删除通知范围为基础生成删除延伸控制策略，选择删除方法，执行删除操作，以实现数据及其副本全面删除。
b) 删除延伸控制策略的可控传递：
   1) 以消息验证码、数字签名等密码学为技术手段，根据删除通知范围生成删除通知流转树，确保隐私信息流转树上的数据使用实体能够正确解密删除通知及其相关信息，并仅能解密当前数据使用实体的后续数据使用实体信息；
   2) 实现隐私信息流转树上数据使用实体的删除延伸控制，同时在删除通知过程中实现通知传递可控；
   3) 后继隐私信息处理者在接收到删除延伸控制策略及删除对象后，更新策略并选择删除方法、执行数据删除等操作。
c) 删除延伸控制策略的可信验证：
   1) 隐私信息所有者设置删除意图，包括但不限于：删除对象、删除时间、删除等级、删除粒度等；
   2) 以删除延伸控制策略生成、可控传递为基础，删除通知可信验证实现删除通知的完备性，删除延伸控制策略的可信验证基于消息验证码、数字签名技术，采用逐级回传机制传递验证信息；
   3) 数据使用实体在正确接收并解密得到相关数据删除信息后，回传不可伪造的可信验证信息，同时将不可伪造验证信息反馈给删除效果评估进行后续操作；
   4) 隐私信息所有者完成删除通知过程全程的验证，形成此次删除延伸控制事件闭环。
d) 删除延伸控制策略的动态调整：
   1) 在删除延伸控制策略传递、删除延伸控制策略执行可信验证回传过程中，以及在隐私信息所有者对可信验证结果未通过的情况下，根据删除延伸控制策略传递异常信息、可信验证信息回传异常信息以及可信验证结果失败信息，对删除通知全生命周期中的异常行为进行动态调整修改；
   2) 通过采用通知重传等手段，确保删除通知过程的可靠性、全面性、安全性和完备性。

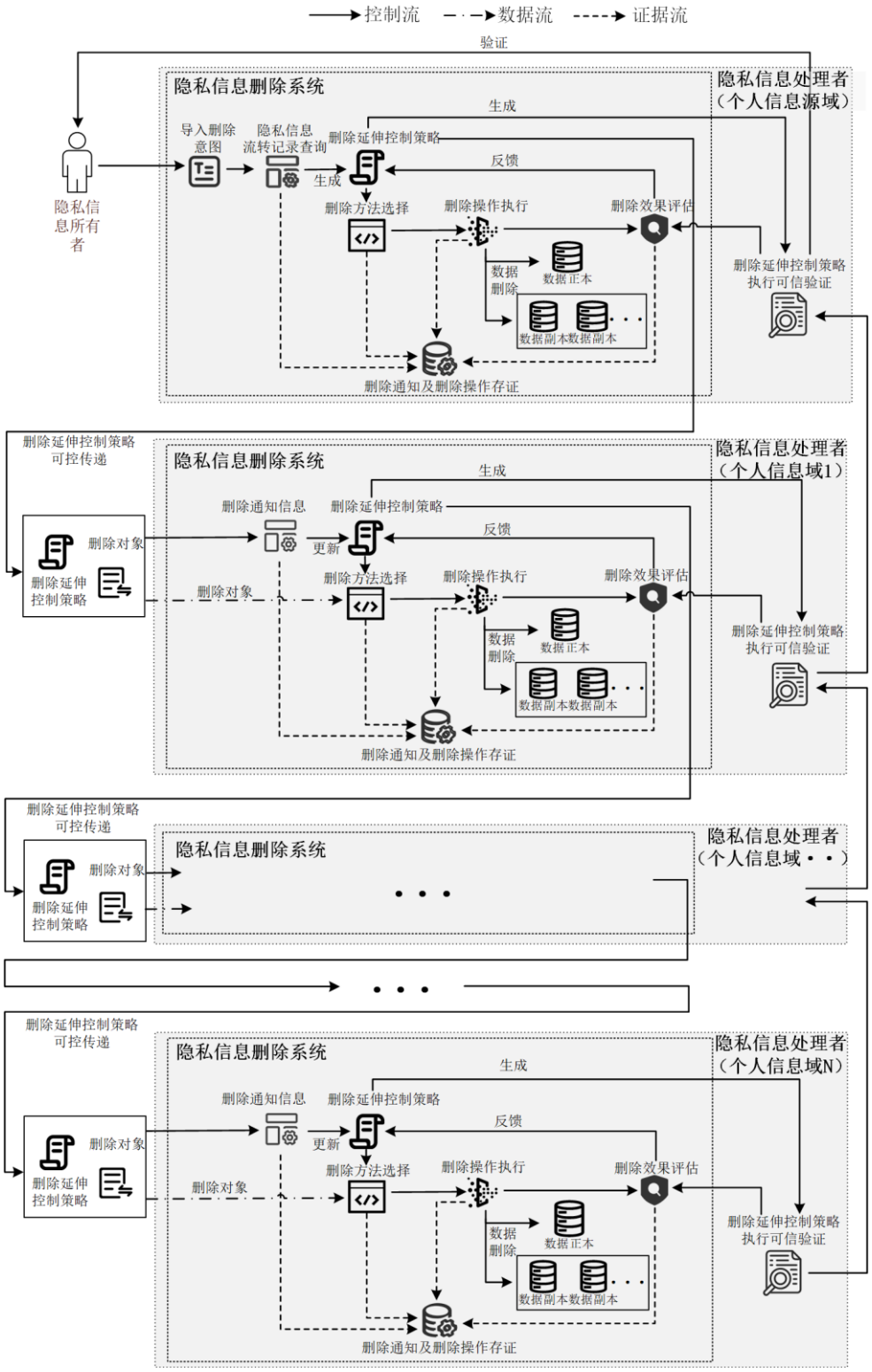


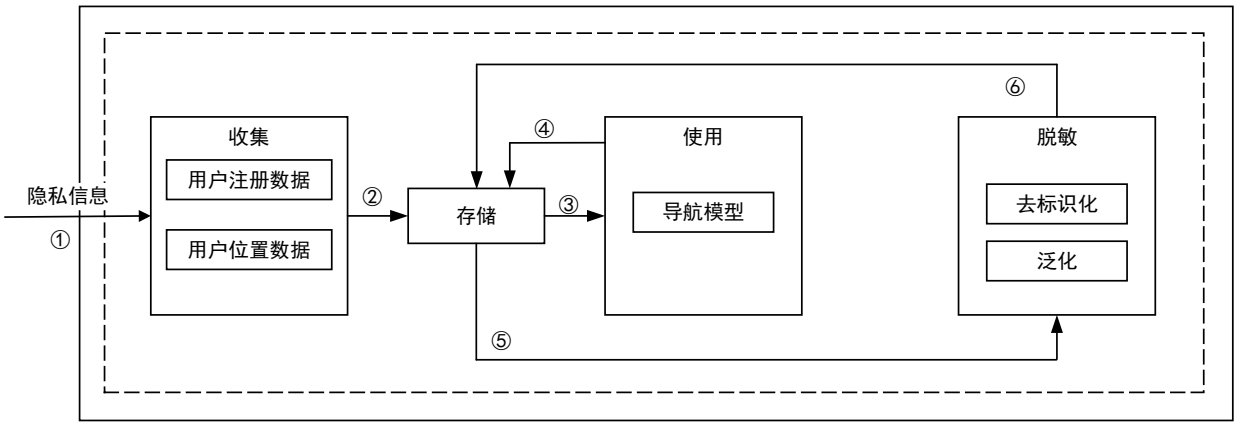
图 B.3 删除延伸控制流程示意图

附 录 C
（资料性）
出行服务应用场景示例

C.1 概述

本附录针对出行服务应用场景，描述了出行服务中隐私信息保护的流程，具体包括隐私保护使用者的收集、存储、使用、脱敏等环节，如图C.1所示。

本示例体现了隐私信息使用者按需脱敏的过程。



图C.1 出行服务应用场景下隐私信息全生命周期的计算操作流程示意图

C.2 出行服务中的隐私信息处理

隐私信息使用者对隐私信息的计算操作流程具体如下：

e) 收集，出行服务场景中，出行 APP 采集的用户数据包括用户位置、用户目的地、用户选路偏好等，如果是注册用户，额外采集的用户数据包括用户名、手机号、亲情手机号、车牌号等，流程步骤对应图 C.1 中隐私信息全生命周期的计算操作流程的步骤①；

f) 存储，这些信息收集后，将存储记录到出行服务商的云服务器，以便后续可以直接按用户偏好和现状提供服务，例如高速优先设置、不同时间段的常去目的地、每日限行提醒、紧急救援通知等等，流程步骤对应图 C.1 中隐私信息全生命周期的计算操作流程的步骤②；

g) 使用，用户打开出行 APP，无论是用户位置、用户目的地等实时采集的用户数据，还是用户选路偏好、用户名、手机号、车牌号等已经存储的用户数据，将送至使用操作，与云服务器上的地图数据一起结合，为用户提供用户定位、周边的信息提醒；如果用户已开启出行服务，将会持续使用用户位置信息，为用户持续播报路况信息、更优的选路信息。流程步骤对应图 C.1 中隐私信息全生命周期的计算操作流程的步骤③；

h) 存储，用户出行过程中会在某些事件节点产生计算结果，例如某时刻途径用户指定的途径点、在区间限速路段的平均时速等等，这些数据需要送往出行服务商云服务器进行数据记录操作，以便后续的统计查询、全程概况回顾。流程步骤对应图 C.1 中隐私信息全生命周期的计算操作流程的步骤④；

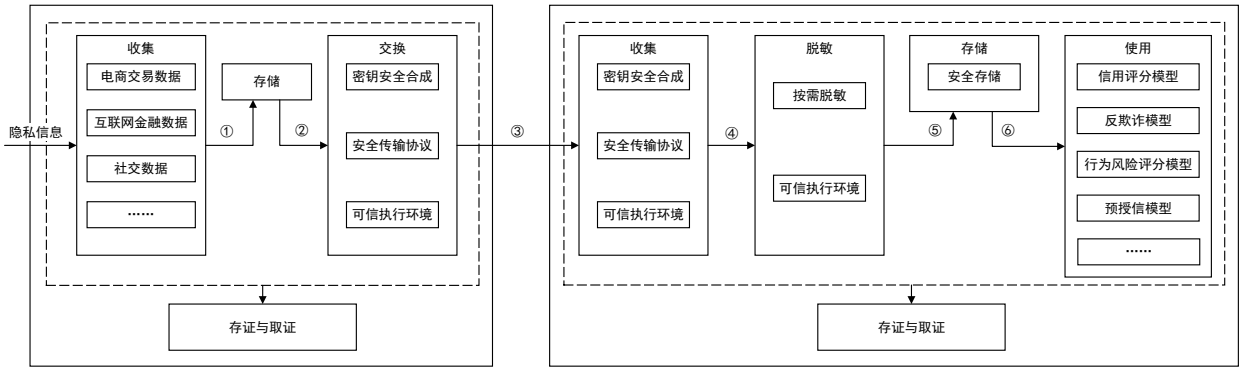
i) 脱敏，出行服务完成后，按《个人信息保护法》的要求去标识化（脱敏），可留存用于行驶途径的路况分析；或者不去标识化，但将记录的数据进行泛化处理（脱敏），可留存用于用户使用情况的统计分析。流程步骤对应图 C.1 中隐私信息全生命周期的计算操作流程的步骤⑤；

j) 存储，出行服务过程记录的信息脱敏后，存至出行服务商的云服务器，流程步骤对应图 C.1 中隐私信息全生命周期的计算操作流程的步骤⑥。

附 录 D
（资料性）
信用计算应用场景示例

D.1 概述

本附录针对信用计算应用场景，描述了跨系统联合的隐私信息保护的流程，具体包括隐私信息提供者的收集、存储、交换、存证与取证等环节，隐私信息使用者的收集、脱敏、存储、使用、存证与取证等环节，如图D.1所示。



图D.1 信用计算应用场景中隐私信息全生命周期的计算操作流程示意图

D.2 隐私信息所有者或隐私信息提供者的计算操作

D.2.1 收集

在信用计算应用场景中，隐私信息提供者收集隐私信息所有者的原始数据（含隐私信息），包括但不限于以下内容：

k） 用户行为数据，包括电商交易数据、互联网金融数据、社交数据、自主提供数据等；
l） 信贷交易数据，包括借贷账户数据、借贷交易还款数据等；
m） 非信贷交易数据，包括电信数据、公共事业缴费数据等；
n） 公共信息数据，包括欠税信息、民事裁决信息、强制执行信息、行政处罚信息等。

D.2.2 存储

隐私信息提供者对收集到的用户原始数据（含隐私信息）进行安全高效存储，保证隐私信息的机密性、可用性、完整性。

D.2.3 交换

隐私信息提供者从存储中获取原始数据（含隐私信息）交换到隐私信息接收者，以确保信用计算的准确性，交换操作可以具有以下安全机制：

o） 可信执行环境，针对跨系统联合的隐私信息保护安全需求，交换操作在可信执行环境中执行；
p） 密钥安全合成，在可信执行环境中，采用多方参与的基于密钥材料与可信执行环境预存秘密的密钥安全合成技术合成传输密钥，用于数据加密，确保隐私信息的机密性；
q） 安全传输协议，确保隐私信息的认证性、完成性和来源真实性。

D.2.4 存证与取证

对上述收集、存储、交换操作进行定制化的操作记录生成和存储，并响应用户的证据查询请求，返回生成的证据，支撑隐私保护合规审计，以及隐私侵权行为溯源与追责。

D.3 隐私信息接收者或隐私信息使用者的计算操作

D.3.1 收集

隐私信息接收者从隐私信息提供者收集用户的原始数据（含隐私信息），收集操作与隐私信息提供者的交换操作相对应，可以具有以下安全机制：

r） 可信执行环境，针对跨系统联合的隐私信息保护安全需求，收集操作在可信执行环境中执行；

s） 密钥安全合成，在可信执行环境中，采用多方参与的基于密钥材料与可信执行环境预存秘密的密钥安全合成技术合成传输密钥，用于数据解密，确保隐私信息的机密性；

t） 安全传输协议，确保隐私信息的认证性、完成性和来源真实性。

D.3.2 脱敏

在数据使用前，隐私信息接收者需对原始数据进行脱敏，脱敏操作需满足以下安全需求：

u） 按需脱敏，按照信用计算场景的脱敏控制策略选择合适的脱敏算法进行隐私信息脱敏。脱敏算法满足用户标识匿名化的安全需求，同一个人在同一信息系统中，同一类型数据在一定时间范围内的相同属性被置换为不同乱码，单向不可逆；

v） 可信执行环境，针对隐私信息接收者不可信的安全需求，脱敏过程在可信执行环境中执行。

D.3.3 存储

隐私信息接收者对脱敏后的隐私信息进行密态数据的安全高效存储。在脱敏数据检索过程中，需考虑满足上述标识化信息随机置乱存储的检索方法，以确保检索数据的准确性。

D.3.4 使用

在使用操作环节，针对信用计算模型输入参数需求，隐私信息使用者对脱敏后的隐私信息进行归一化处理，即使用归一化标识替换随机置乱标识，并基于信用评分模型、反欺诈模型、行为风险评分模型、预授信模型等信用计算模型完成信用评估。

D.3.5 存证与取证

对上述收集、脱敏、存储、使用操作进行定制化的操作记录生成和存储，并响应用户的证据查询请求，返回生成的证据，支撑隐私保护合规审计，以及隐私侵权行为溯源与追责。

参 考 文 献

[1] GB / T 22118-2008 企业信用信息采集、处理和提供规范

[2] GB / T 22120-2008 企业信用数据项规范

[3] GB / T 26819-2011 信用主体标识规范

[4] GB / T 33718-2017 企业合同信用指标指南

[5] GB / T 39440-2020 公共信用信息资源目录编制指南

[6] GB / T 39441-2020 公共信用信息分类与编码规范

[7] 《隐私计算理论与技术》，李凤华、李晖、牛犇著，人民邮电出版社，第 1 版，2021