ICS 35.030

CCS L80

中华人民共和国国家标准

GB/T XXXXX—XXXX

数据安全技术 数字水印技术实现指南

Data security technology — Technical implementation guideline of digital watermarking

（征求意见稿）

（本稿完成时间：2024 年 3 月 22 日）

XXXX - XX - XX 发布　　XXXX - XX - XX 实施

国家市场监督管理总局
国家标准化管理委员会　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。

本文件由全国网络安全标准化技术委员会（SAC/TC 260）提出并归口。

本文件起草单位：

本文件主要起草人：

数据安全技术　数字水印技术实现指南

1　范围

本文件提出了数字水印技术的实现框架、功能、流程、水印算法选择、水印服务封装形式选择等方面的建议，并给出了常见数字水印算法、典型安全场景等相关信息。

本文件适用于数字水印技术的设计、开发、应用和测试。

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 信息安全技术 术语

3　术语和定义

GB/T 25069 中界定的以及下列术语和定义适用于本文件。

3.1

数字水印技术　digital watermarking technology

一种通过在数字内容中嵌入不易察觉的特定信息，用于标识或保护数字媒体的版权、来源或内容的信息安全技术。

注：简称数字水印，本文件所称“数字水印”是指隐式水印（也称“暗水印”、“隐形水印”、“隐性水印”、“不可见水印”），其所嵌入的信息对数据使用者是隐蔽且不可辨识的。

3.2

水印信息　watermark information

通过数字水印技术（3.1）在数字媒体中嵌入的特定信息。

注：常见水印信息包括但不限于版权信息、溯源信息、链路信息、机构/员工 ID 等。

3.3

水印载体　watermark carrier

用于嵌入或携带水印信息（3.2）的文档、图像、音频、视频、网页、数据库等数字内容。

注：简称载体。

3.4

水印编码　watermark encoding

将水印信息（3.2）转换为适合嵌入到水印载体（3.3）的形式或格式的过程。

3.5

水印解码 watermark decoding

将从水印载体（3.3）中提取出的水印信息（3.2）复原为其原始形式或格式的过程。

3.6

水印嵌入 watermark embedding

将水印信息（3.2）嵌入到水印载体（3.3）中的过程。

3.7

水印提取 watermark extraction

从水印载体（3.3）中检测和识别水印信息（3.2）的过程。

3.8

失真干扰 distortion interference

水印载体（3.3）在传输、使用等过程中，因有意或无意的修改或处理，导致其携带的水印信息（3.2）质量下降或损坏，或无法正常提取的现象。

3.9

水印攻击 watermark attack

任何故意试图破译、破坏、移除、修改、伪造水印载体（3.3）中水印信息（3.2）的行为。

注：包括但不限于对水印载体进行裁剪、形变、压缩、滤波、去噪等。

4 缩略语

下列缩略语适用于本文件。

SaaS：软件即服务（Software as a Service）

SDK：软件开发工具包（Software Development Kit）

5 实现框架

数字水印技术实现涉及水印载体、数字水印算法、技术实现流程、水印服务封装等内容。本文件中数字水印技术所适用的水印载体包括文档、图像、音频、视频、网页、数据库等。数字水印算法主要有水印嵌入/提取算法、水印编码/解码算法等，常见数字水印算法见附录 A。技术实现流程主要有水印嵌入阶段、水印载体分发阶段、水印提取阶段。水印服务封装主要有 SDK、SaaS、产品等。数字水印技术主要应用在数据版权保护、数据泄露追踪溯源、生成式人工智能生成内容的水印标识、网络数据分类分级及管理、数据完整性保护等场景，典型安全场景见附录 B。数字水印技术实现框架如图 1 所示。

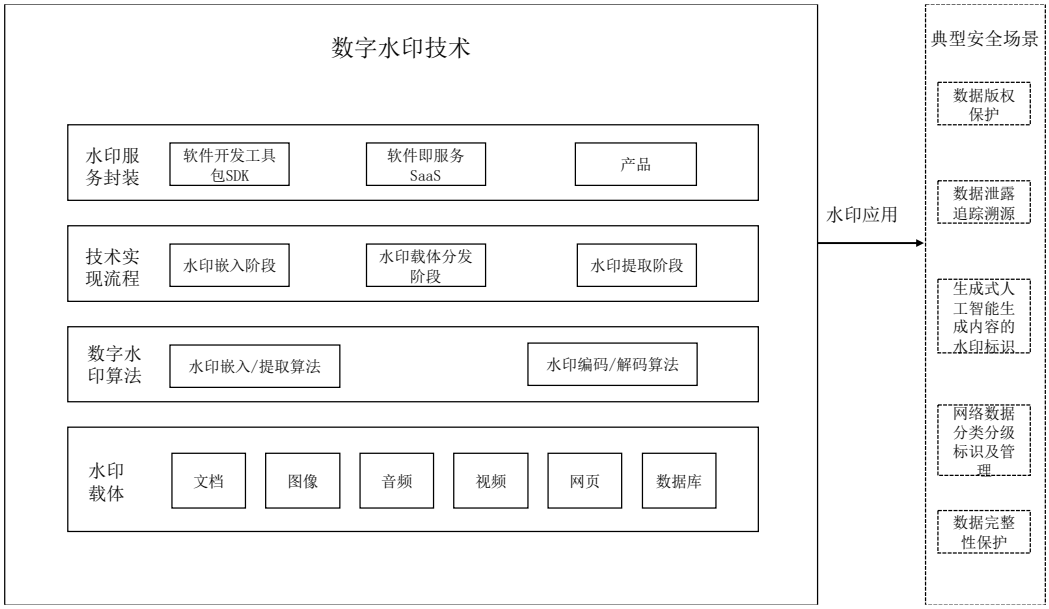

图1　数字水印技术实现框架

6　功能

数字水印技术实现的功能分为：基本功能、增强功能和特定功能。基本功能指数字水印技术基本可用，能够达到预期目的；增强功能指水印载体在遭受失真干扰、水印攻击等情形下，数字水印技术仍然能够达到预期效果；特定功能为满足明确应用场景，数字水印技术需要支持的特定需求。水印技术功能实现情况判定方式见附录 C。

a) 基本功能：
   1) 保证水印信息隐蔽：确保水印的存在难以被载体内容的使用者察觉，且水印信息无法通过视觉、听觉等直观感受识别。该功能是隐式水印与显式标识的本质区别；
   2) 确保载体正常使用：确保嵌入水印的载体能够正常使用，并且完成预期功能和目的；
   3) 支持水印信息提取：确保携带水印信息的载体在未受到任何失真干扰与水印攻击的情况下，水印信息能通过提取算法被完整提取。
b) 增强功能：
   1) 防御水印攻击：确保已嵌入的水印信息难以被破坏或篡改，即攻击者在未知具体提取方法及参数的情况下，难以对水印信息进行有效地毁坏、抹除、窃取、替换等恶意攻击，或是含水印载体遭受恶意攻击后仍然可以完整提取水印信息；
   2) 抵抗失真干扰：确保携带水印信息的水印载体在使用过程中，遭受格式转换、信道噪音、压缩等有损处理后，仍能通过对应的水印提取算法和水印解码算法准确地恢复出水印信息。
c) 特定功能：
   1) 满足大容量需求：在有明确水印容量需求的应用场景中，如版权保护、信息标注，水印算法在目标载体上的信息嵌入量可满足大容量需求，例如，在版权保护场景中，水印载体需要完整地携带对应的版权信息；
   2) 满足实时性需求：在有明确嵌入时效要求的应用场景中，水印算法的嵌入或提取效率需满

足对应的实时性要求，例如在直播场景中，水印嵌入算法的时效性宜与直播流的帧率相匹配。

7 流程

7.1 概述

数字水印技术实现流程通常分为水印嵌入阶段、水印载体分发阶段和水印提取阶段，如图 2 所示。其中，水印嵌入阶段包括嵌入方案设计及预处理、水印编码、水印嵌入等主要环节；水印载体分发阶段，水印载体易遭受失真干扰、水印攻击；水印提取阶段包括提取方案设计及预处理、水印提取、水印解码等主要环节。

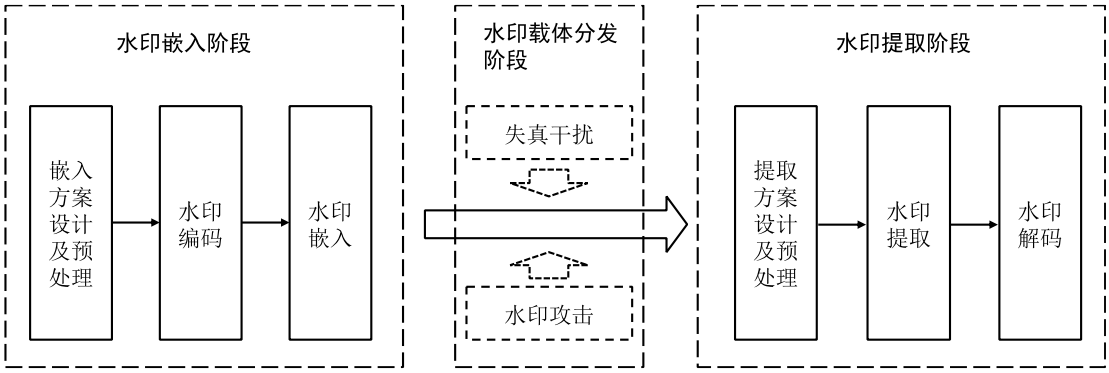

图2 数字水印技术实现流程

7.2 水印嵌入阶段

7.2.1 概述

水印嵌入阶段是将水印信息进行编码并通过合适的策略和算法嵌入到目标水印载体中的过程，包含嵌入方案设计及预处理、水印编码、水印嵌入三个主要环节。

7.2.2 嵌入方案设计及预处理

在此环节，首先需要明确数字水印的使用场景、待嵌入的水印信息及水印载体的基本特性，以明确需要实现的基本功能及特定功能。其次，评估对应场景下潜在的失真干扰和水印攻击，以明确需要实现的增强功能。随后根据上述评估结果对水印嵌入算法、水印编码算法等进行类型选择、细节设计和参数调整等。最后依据所设计嵌入方案对水印载体和水印信息进行必要的初步处理。

a) 对水印嵌入方案进行设计时，需要考虑的因素包括但不限于：
   1) 载体的基本信息：包括但不限于载体类型、载体结构（如是否含有元数据等）、内容编码算法、封装格式、空域尺度信息（如图像的分辨率、位深等）、时序尺度信息（如视频的时长、帧率等）；
   2) 水印信息的类型及容量需求：水印信息的类型理论上可以涵盖所有类型的数字内容，但常见的类型主要包括文本信息（如某公司版权所有、仅供某组织使用）、图像（如企业图标）等；水印的容量需求指在特定环境下能够完整携带水印信息的二进制流的长度，通常以“比特”为单位来衡量；
   3) 水印嵌入强度限制：即水印载体在水印嵌入过程中所允许的内容调整区域和调整幅度。该项因素通常与水印载体的类型及后续的使用场景有关，例如高清影视作品中允许的水印

嵌入强度通常会远小于在线直播视频的嵌入强度；

4) 防御水印攻击功能、抵抗失真干扰功能：即评估水印载体在后续的使用场景中可能会引入的失真干扰或水印攻击带来的损伤程度。例如水印载体后续要通过社交网络进行传输，那么水印技术需抵抗此类传输所导致的失真干扰；
5) 实时性需求：即此应用场景对水印在嵌入和提取阶段所花费时间的限制，通常对实时性有需求的场景包括高并发应用场景、流媒体场景等；
6) 提取准确率需求：多数场景下需要水印的提取准确率趋近 100%，但也有一些场景，例如企业图标图像水印，允许水印提取时在不影响最后的语义读取的前提下有一定比率的错误；
7) 嵌入位置评估：一些内容或场景中，水印仅允许被嵌入在载体的指定或特定位置，或载体的部分区域不适合嵌入水印信息。例如，数据库载体中的个人身份证号码、银行卡余额等高敏感信息不宜用来嵌入水印。

b) 对水印载体和水印信息预处理时，需要考虑的因素包括但不限于：
1) 载体内容解析：将原始载体文件解析到水印嵌入算法可操作的层面，例如在视频帧中嵌入水印需要在预处理阶段对视频载体进行解帧；
2) 载体嵌入容量评估：即根据载体内容计算其在所设计的嵌入方案下的容量范围；
3) 水印信息数字化：将具有现实意义的水印信息构造为数字化表达的过程；
4) 水印信息映射：通过映射函数或映射表将水印信息进行合适变换的过程。例如泄露溯源水印中通常并不会直接嵌入分发渠道的名称，而是将各个渠道映射成具有唯一性的 ID 信息；
5) 水印信息去冗余：对水印信息中的冗长部分进行精简的行为。

7.2.3 水印编码

在水印编码环节，需要根据水印的容量需求、实现功能等进行综合考虑，并有针对性进行编码方案的选择。

a) 对水印信息进行水印编码时，需要考虑的因素包括但不限于：
1) 抗监听、破解：主要以加密、扰动等方式实现；
2) 抗替换：主要通过为水印信息添加校验码等方式实现；
3) 抗失真干扰：通过纠错编码、扩频编码、图形化编码、同步编码等方式实现，该因素的实现可能会损失部分水印容量；
4) 提升水印容量：通过压缩编码或保持二进制明文编码等方式实现，以最大化水印的有效载荷。

b) 文档、图像、音频、视频、网页、数据库等载体的水印编码算法选择参考第 8 章。

7.2.4 水印嵌入

在水印嵌入环节，根据 7.2.2 的水印嵌入方案设计进行，选择合适的水印嵌入策略、水印嵌入算法，对编码后的水印信息进行嵌入。

a) 水印嵌入时还需根据增强功能和特定功能等需求，综合进行嵌入策略的设计，常用的嵌入策略包括：
1) 全局嵌入：在原始载体的所有内容中嵌入水印信息；
2) 局部嵌入：仅选择在原始载体的一部分区域嵌入水印信息；
3) 周期嵌入：以空间或时间为周期，在原始载体中轮番、多次嵌入水印信息；

4) 定点嵌入：仅在原始载体中满足预设条件的位置嵌入水印信息；
5) 自适应嵌入：根据原始载体中各部分内容的不同，动态选择是否嵌入水印信息或动态调整水印的嵌入容量等。

b) 文档、图像、音频、视频、网页、数据库等载体的水印嵌入算法选择参考第 8 章。

7.3 水印载体分发阶段

水印载体分发阶段是将含水印载体分发至对应的目标受众或渠道的过程。在此过程中，水印载体易遭受失真干扰、水印攻击。常见的含水印载体分发方式主要包括：

a) 统一分发：对于同一原始载体，向所有渠道和目标受众分发相同的含水印内容，常用于版权水印、标签水印等场景；
b) 渠道分发：根据分发渠道或内容受众等的不同，在原始载体中嵌入不同的水印信息，再将对应的含水印载体分发至对应的渠道或投放给对应的内容受众。

注：失真扰动和水印攻击，都可能对水印载体及水印信息带来不同程度的损坏，导致水印提取阶段所获取的载体与之前分发阶段的载体存在差异。当此类差异过大时，即使存在抵抗失真干扰的设计，仍可能导致水印失效或无法提取。因此，数字水印技术有其局限性，水印载体也需尽量设计、分发至合适的场景才能起到理想的作用。

7.4 水印提取阶段

7.4.1 概述

水印提取阶段是从含水印载体中检测和识别水印信息的过程，包含提取方案设计及预处理、水印提取、水印解码三个主要环节。

注：水印提取阶段可能包含以下几种结果：1）水印提取完成并解码出正确的信息内容；2）水印提取完成但无法解码出有意义的信息内容，通常是乱码或是错误内容；3）水印无法提取或提取失败。仅第一种情况被认为是水印提取成功。

7.4.2 提取方案设计及预处理

在提取方案设计及预处理环节，首先需要根据相应的水印嵌入算法来决定提取阶段所需要的策略、参数、辅助信息等内容，其次需要评估含水印内容在分发及后续的使用中因失真干扰和水印攻击所引入的噪声对水印信息的影响，最后根据上述评估对水印提取算法、水印信息解码算法等进行设计,并对待提取的水印载体进行初步处理。

a) 对水印提取方案进行设计时需要考虑的因素包括但不限于：
   1) 水印嵌入算法：通常情况下，水印提取算法和水印嵌入算法存在互逆特性，因此，设计水印提取算法时，水印嵌入算法是第一参考要素；
   2) 载体的基本信息：由于失真扰动和水印攻击的存在，待测载体的基本信息可能已经改变，在水印提取算法设计时，需要重新考虑包括但不限于载体类型、载体结构、内容编码算法、封装格式、空域尺度信息、时序尺度信息等；
   3) 水印嵌入的策略：水印的嵌入策略影响着水印提取的区域和策略，例如，采用周期嵌入策略的载体，仅需定位并提取出一个完整周期内的水印信息即可，无需提取整个载体中的水印。
b) 对含水印载体预处理时需要考虑的因素包括但不限于：
   1) 载体内容解析：将原始载体文件解析到水印嵌入算法可操作的层面，例如在视频帧中嵌入水印需要在预处理阶段对视频载体进行解帧；

2) 含水印载体失真分析：即通过对载体的初步处理来评估载体是否经历裁剪、压缩、缩放等有损操作及对应操作的参数范围。

7.4.3 水印提取

在水印提取环节，首先根据 7.4.2 的水印提取方案进行，选择合适的水印提取策略与水印提取算法，对水印载体进行解析及提取。

a) 水印提取策略与水印提取算法紧密相关，常见的水印提取策略包括：
   1) 盲提取：不依赖原始载体内容，直接从含水印载体中提取水印信息；
   2) 非盲提取：以全部或部分原始载体内容为参考才能进行水印信息的提取；
   3) 全局提取：依次提取出载体内容中所有嵌入的水印信息；
   4) 局部提取：定位到载体中含水印的部分并提取其中的水印信息；局部提取适用于局部嵌入和定点嵌入的水印；
   5) 单周期提取：在空间或时间等尺度上定位到水印的一个完整嵌入周期，并提取出该周期内的水印信息；
   6) 多周期提取：定位到水印的多个嵌入周期，提取出其中的所有水印并综合获得最终的水印信息；
   7) 自适应提取：根据载体中各部分内容的不同，动态选择是否提取水印信息或动态调整水印的提取参数等。
b) 文档、图像、音频、视频、网页、数据库等载体的水印提取算法选择参考第 8 章。

7.4.4 水印解码

在水印解码环节，首先根据提取出的水印信息进行解密、分离校验码等各种解码，随后根据所提取的水印信息去验证水印信息的真伪与完整性。

a) 水印解码需要考虑的因素包括但不限于：
   1) 水印信息是否加密：若为加密信息，则进行解密操作；
   2) 水印信息是否存在校验位：若存在，则分离出校验位，并判断校验信息是否正确；
   3) 水印信息是否包含纠错编码：若包含，则对提取出的水印信息进行纠错；
   4) 水印信息是否含有同步码：若含有，则定位同步码，并以同步码为起点定位出完整的水印信息周期；
   5) 水印信息的其他编码方式：采用相应的解码方式进行还原；
   6) 恢复出有实质意义的水印信息：对提取出的水印信息进行解码等相关操作。
b) 文档、图像、音频、视频、网页、数据库等载体的水印解码算法选择参考第 8 章。

水印解码完成后，可通过提取出来的水印信息进行数据版权保护、数据泄露溯源等应用处理。

8 水印算法选择

8.1 概述

不同水印载体类型在结构和内容上存在着一定的共性和差异，载体中可利用的冗余空间及适用的水印嵌入/提取算法也因此存在较大异同点。同时，即使是在相同载体类型上，根据功能的差异，水印嵌入/提取算法的选择也存在明显的差异性。此外，根据载体类型和功能的差异，水印编码/解码算法的选择也存在较明显差异性。

8.2 至 8.7 根据水印载体的特性，针对不同层次的功能给出数字嵌入/提取水印算法和水印信息编码/解码算法的选择建议。当所设计的水印嵌入/提取算法水印编码/解码算法需要同时满足多层次的功能时，宜优先选择对应推荐目标算法的交集。常见水印嵌入/提取算法可参考附录 A1.1，常见水印编码/解码算法可参考附录 A1.2，常见水印编码/解码算法与主要水印载体类型的适配情况可参考附录 A1.3。

注：通常情况下，水印的嵌入算法和提取算法对载体的操作存在较为明显的对称性和互逆性，且同属一类算法，故在本章中，我们将水印嵌入算法和提取算法归并为“水印嵌入/提取算法”进行阐述。同理，我们也将水印信息编码算法和水印信息解码算法归并为“水印信息编码/解码算法”进行阐述。

8.2 文档

文档中的冗余空间与载体本身的特性有密切的联系，选择水印算法时：

a) 对于文档类水印载体，针对基本功能：
   1) 水印嵌入/提取算法宜选择内容水印算法、深度学习水印算法，在文本载体拥有文件结构和页面布局的情况下可选择元数据水印、不可见元素水印等算法；
   2) 水印编码/解码算法宜选择二进制明文编解码、扰动/加密编解码等算法，部分情况下可选择附录 A.2 所述其他所有编码/解码算法等。
b) 在上述 8.2 a）基础上，针对增强功能：
   1) 水印嵌入/提取算法宜选择内容水印、深度学习水印和不可见元素水印等算法；
   2) 水印编码/解码算法宜选择扰动/加密编解码、纠错编解码、同步码、和校验编解码、图形化编码、扩频编解码等算法。
c) 针对特定功能中的大容量需求：
   1) 水印嵌入/提取算法宜选择不可见元素水印算法等，或者在可行的情况下选择元数据水印、内容水印、不可见元素水印算法进行组合使用；
   2) 水印编码/解码算法宜选择二进制明文编解码、扰动/加密编解码、压缩编解码等算法等。

通常，上述文档类水印技术所涉及的水印嵌入/提取算法、水印编码/解码算法均可满足特定功能中的实时性需求。

8.3 图像

对于图像载体来说，元数据、视觉空间、变换域等都是较为常见且理想的水印嵌入位置。运用现代信息处理技术和编码技术等，能够找到这些矩阵中的视觉非敏感部分，从而用水印信息去替代这些部分内容，或者通过深度信号处理去发掘矩阵中的冗余空间来携带额外的水印信息。

a) 对于图像类水印载体，针对基本功能：
   1) 水印嵌入/提取算法宜选择元数据水印、模板水印、变换域水印、直方图水印、最低有效位水印、深度学习水印等算法，在图像拥有透明层等情况下可选用不可见元素水印算法；
   2) 水印编码/解码算法宜选择附录 A.2 所述所有编码/解码算法等。
b) 在上述 8.3 a）基础之上，增强功能：
   1) 水印嵌入/提取算法宜选择模板水印、变换域水印、深度学习水印和不可见元素水印等；
   2) 水印编码/解码算法宜选择扰动/加密编解码、纠错编解码、同步码、图形化编解码、扩频编解码和校验编解码等。
c) 针对特定功能中的大容量需求：
   1) 水印嵌入/提取算法宜选择变换域水印、直方图水印、最低有效位水印或不可见元素水印，或者在可行的情况下选择元数据水印、不可见元素水印与模板水印、变换域水印、直方图水印、最低有效位水印中的一种或多种进行组合使用；
   2) 水印编码/解码算法宜选择二进制明文编解码、扰动/加密编解码、压缩编码等算法等。

d) 针对特定功能中的实时性需求：
   1) 水印嵌入/提取算法宜选择元数据水印、模板水印、不可见元素水印、最低有效位水印等；
   2) 水印编码/解码算法宜选择二进制明文编码、扰动/加密编解码、纠错编解码和校验编解码等。

8.4 音频

宜在不影响原始音频听觉质量的条件下在音频中嵌入水印信息。对于音频载体来说，元数据、时序信号、变换域等都能够携带水印信息。

a) 对于音频类水印载体，针对基本功能：
   1) 对于音频型水印载体，水印嵌入/提取算法宜选择元数据水印、时序水印、变换域水印、直方图水印、深度学习水印和最低有效位水印等算法；
   2) 水印编码/解码算法宜选择附录 A.2 所述所有编码/解码算法等。
b) 在上述 8.4 a）基础之上，增强功能：
   1) 水印嵌入/提取算法宜选择时序水印、变换域水印、深度学习水印等；
   2) 水印编码/解码算法宜选择扰动/加密编解码、纠错编解码、扩频编解码、图像化编解码、同步码和校验编解码等。
c) 针对特定功能中的大容量需求：
   1) 水印嵌入/提取算法宜选择变换域水印、直方图水印、深度学习水印或最低有效位水印，或者在可行的情况下选择元数据水印与变换域水印、直方图水印或最低有效位水印中的一种或多种进行组合使用；
   2) 水印编码/解码算法宜选择二进制明文编解码、扰动/加密编解码、压缩编解码等算法等。
d) 针对特定功能中的实时性需求：
   1) 水印嵌入/提取算法宜选择元数据水印、最低有效位水印等；
   2) 水印编码/解码算法宜选择二进制明文编码、扰动/加密编解码、纠错编解码和校验编解码等。

8.5 视频

图像载体和视频载体有很大的共通性，视频中的关键帧可以被当成一张独立的数字图像来处理。除了元数据、空间视觉内容等与图像载体的共性水印空间外，视频在时间维度上的连续性和冗余度也可以用来携带水印信息。

注：由于无损存储视频开销巨大，几乎所有的视频载体都是经过压缩编码的，例如 MPEG 和 H.26X 系列编码。因此，在考虑时序冗余空间的时候，视频编码的结构、时序的误差传递等特性也要被考虑进去。

a) 对于视频载体，针对基本功能：
   1) 水印嵌入/提取算法宜选择元数据水印、模板水印、时序水印、变换域水印、直方图水印、深度学习水印和最低有效位水印等；
   2) 水印编码/解码算法宜附录 A.2 所述所有编码/解码算法等。
b) 在上述 8.5 a）基础之上，增强功能：
   1) 水印嵌入/提取算法宜选择模板水印、变换域水印、深度学习水印和时序水印等；
   2) 水印编码/解码算法宜选择扰动/加密编解码、纠错编解码、图像化编解码、扩频编解码、同步码和校验编解码等。
c) 针对特定功能中的大容量需求：

1) 水印嵌入/提取算法宜选择变换域水印、直方图水印、深度学习水印或最低有效位水印，或者在可行的情况下选择元数据水印与模板水印、变换域水印、直方图水印、最低有效位水印中的一种或多种进行组合使用；
2) 水印编码/解码算法宜选择二进制明文编解码、扰动/加密编解码等算法等。

d) 针对特定功能中实时性需求：
1) 水印嵌入/提取算法宜选择元数据水印、模板水印、最低有效位水印、深度学习水印等；
2) 水印编码/解码算法宜选择二进制明文编码、扰动/加密编解码、纠错编解码和校验编解码等。

8.6 网页

对于网页内容的各个组成部分来说，可以依据其内容形式、灵活的结构等特性进行水印的添加。

注：由于网页内容可包含视频、音频、图像等几乎所有模态的多媒体内容，为避免描述的重复和冗余，本章节所推荐的适用于网页的相关算法仅指代整体网页渲染页面或网页源码，不包含网页中视频、音频、图像等多媒体内容所适用的算法。

a) 对于网页类水印载体，针对基本功能：
1) 水印嵌入/提取算法宜选择模板水印、深度学习水印和内容水印等算法，在页面上存在时序内容，且支持时序记录的情况下，网页水印宜选择时序水印算法，在以网页源码或页面文本为提取对象时，宜选择不可见元素水印算法；
2) 水印编码/解码算法宜选择附录 A.2 所述所有编码/解码算法等。

b) 在上述 8.6 a）基础之上，增强功能：
1) 水印嵌入/提取算法宜选择模板水印、内容水印、时序水印、深度学习水印和不可见元素水印等；
2) 水印编码/解码算法宜选择扰动/加密编解码、纠错编解码、同步码、扩频编解码、图形化编解码和校验编解码等。

c) 针对特定功能中的大容量需求，数字水印算法的选择建议如下：
1) 水印嵌入/提取算法宜选择内容水印或不可见元素水印，或者在可行的情况下选择内容水印或不可见元素水印相互组合使用；
2) 水印编码/解码算法宜选择二进制明文编解码、压缩编码、扰动/加密编解码等算法等。

d) 针对特定功能中的实时性需求：
1) 水印嵌入/提取算法宜选择模板水印和不可见元素水印等；
2) 水印编码/解码算法宜选择二进制明文编码、扰动/加密编解码、纠错编解码和校验编解码等。

8.7 数据库

数据库的主要形式是结构化数据库。数据库型水印设计宜遵从如下策略：新增数据表的行/列，将水印信息嵌入到新增的内容中；修改数据内容，或采用附加不可见数据内容，或采用附加可见数据水印编码。这些策略不可避免的会修改数据库中的数据，修改的幅度要根据实际情况选用。

注：数据库水印在开发设计时，需要考虑数据污染的问题，若直接修改数据库中数据表的内容，可能会产生风险，例如用户余额表中的数据不宜做任何改动，否则会造成直接经济纠纷。

a) 对于数据库类水印载体，针对基本功能：
1) 水印嵌入/提取算法宜选择内容水印和不可见元素水印等算法，在数据库含有元数据的情况下宜选择元数据水印算法，部分场景下宜选择最低有效位水印算法；
2) 水印编码/解码算法宜选择附录 A.2 所述所有编码/解码算法等。

b) 在上述 8.7 a）基础之上，增强功能：
   1) 水印嵌入/提取算法宜选择内容水印、不可见元素水印等；
   2) 水印编码/解码算法宜选择扰动/加密编解码、纠错编解码、同步码、扩频编解码、图形化编解码和校验编解码等。
c) 针对特定功能中的大容量需求：
   1) 水印嵌入/提取算法宜选择内容水印、不可见元素水印、最低有效位水印，或者在可行的情况下选择内容水印、不可见元素水印、最低有效位水印进行组合使用；
   2) 水印编码/解码算法宜选择二进制明文编解码、压缩编码、扰动/加密编解码等算法等。
d) 针对特定功能中的实时性需求：
   1) 水印嵌入/提取算法宜选择元数据水印、内容水印、不可见元素水印、最低有效位水印等；
   2) 水印编码/解码算法宜选择二进制明文编码、扰动/加密编解码、纠错编解码和校验编解码等。

9 水印服务封装形式选择

9.1 SDK 封装

在本地化或私有化部署的情况下，数字水印技术宜封装为 SDK 形式。采用单个 SDK 或多个独立 SDK 的方式需要侧重考虑整体性或灵活性。

如果侧重于整体性，宜封装为一个 SDK，通过 SDK 提供的多个不同接口提供水印嵌入、水印提取等服务；如果侧重于灵活性，宜封装为水印嵌入、水印提取等多个独立的 SDK，各 SDK 仅提供单一服务。

注 1：SDK 水印嵌入服务的输入接口包括水印载体、水印信息及各功能的控制参数等，输出接口包括嵌入水印信息的载体内容、服务状态等参数；SDK 水印提取服务的输入接口包括携带水印信息的载体、水印提取辅助信息等，输出接口包括水印信息等，部分水印算法（例如可逆水印）也会输出提取水印后的载体内容。若为单个 SDK 模式，输入接口除兼容上述的两类输入以外，接口数据还需包含算法的类型，即指定当前服务是嵌入还是提取。

注 2：SDK 具有较强的私密性，数字水印技术的使用者和服务方之间以技术模块交付为主要交流方式，无需进行数字载体及水印信息的传输和交互。

9.2 SaaS 封装

在远程部署或共享服务的情况下，数字水印技术宜封装为 SaaS 的形式。该服务模式下，水印的嵌入和提取等技术将封装成不同的服务接口。

注 1：SaaS 水印嵌入服务的输入接口包括水印载体、水印信息及各功能的控制参数等，输出接口包括嵌入水印信息的载体内容、服务状态等参数；SaaS 水印提取服务的输入接口包括携带水印信息的载体、水印提取辅助信息等，输出接口包括水印信息等，部分水印算法（例如可逆水印）也会输出提取水印后的载体内容。

注 2：SaaS 具有较强地便捷性，数字水印技术的使用者无需进行数字水印技术模块部署和维护，使用者和服务方之间以数据流形式进行数字载体和水印信息的输入输出。

9.3 产品封装

在本地化或私有化部署且面向业务人员使用的情况下，数据水印技术宜封装为产品形式。该模式下，水印的嵌入和提取等技术将封装成不同的产品功能。

注 1：水印产品根据应用场景需求，适配不同的水印载体、水印信息，同时支持算法选择及参数配置。

注 2:产品具有较强地通用性和易用性,数字水印技术的使用者和服务方之间以数字水印产品交付为主要交流形式，使用者无需关注水印技术的算法及实现细节，直接在数字水印产品上进行操作配置即可嵌入或提取水印。

附 录 A
（资料性）
常见数字水印算法

A.1 水印嵌入/提取算法

水印嵌入/提取算法是水印信息与载体进行融合与分离的主体步骤。通常来说，水印提取是水印嵌入的逆过程，且二者所用算法之间存在较为明显的对称性和互逆性，因此，我们将水印嵌入算法和提取算法归并为“水印嵌入/提取算法”进行阐述。结合具体业务场景选择合适的水印嵌入/提取算法。常见的水印嵌入/提取算法包括但不限于：

a) 基于元数据的水印算法：该类水印算法主要是利用内容载体在文件层面的元数据信息，寻找其中的空白位置、保留位置或者可替换位置用来添加水印信息；
b) 基于最低有效位的水印算法：该类算法的核心思路是将水印信息放置在图像、视频、音频等结构化数据基本单元（例如像素点）的最低有效位中；
c) 模板水印：将水印信息单独设计成与小于等于内容载体大小的模板，再将水印模板套用在载体上，该类型水印技术常用在图像、音频、视频等具有空间或时间尺度的内容载体上；
d) 直方图水印：基于直方图统计结果，通过直方图偏移进行水印信息的嵌入；
e) 变换域水印：先将载体内容通过离散余弦变换（DCT）、离散傅里叶变换 （DFT）、离散小波变换（DWT）等一种或多种可逆变换转换成变换域信号，再对变换域中一些合适的位置进行调制从而达到嵌入水印的目的。可逆变换的算法有很多，不限于上述列举的 DCT、DFT、DWT 算法。由于对变换域的操作带来的视觉效果会被整个载体均匀分摊，变换域水印在一些场景下会拥有更好的隐蔽性；
f) 内容水印：对于文档、网页等可编辑文本载体，通过内容层面的调整来达到嵌入水印的目的，例如，“我把梨子吃了”和“梨子被我吃了”；
g) 空白/不可见元素水印：例如 word 文档中插入不可见字符或者透明图形元素等来携带水印信息；
h) 时序水印：基于原载体内容的时序冗余信息来进行水印嵌入的方法，通常适用于音频、视频等具有时序尺度的内容；
i) 深度学习水印：通过深度神经网络来进行各模态水印嵌入和提取的技术，通常是通过大规模数据训练来实现的，主要包括传统嵌入配合深度提取和端到端深度嵌入、深度提取两种模式。

注：上述所列水印嵌入/提取算法是业界和学术界较为常用的算法类型，各算法并没有明确的边界，也并不在同一个划分维度，故部分算法之间可能存在一定的交集，例如一种算法可以属于变换域水印，同时也属于直方图水印。在数字水印实现过程中，宜根据所推荐算法类型查阅相关学术或技术文档并结合具体应用场景来确定相关细节。

A.2 水印编码/解码算法

水印编码算法是指将待嵌入的水印信息转化成合适的二进制流的算法，常用的编码算法包括但不限于：

a) 二进制明文编码：一些对水印安全性要求较低的场景下，可以直接将水印信息转换成二进制码流后进行嵌入；
b) 纠错编码：纠错编码主要用来提高水印的抗失真干扰，加入了纠错编码的水印信息在提取时候即使出现了若干的错误比特也可以完整恢复出原始内容。不过纠错编码通常会带来水印内容

膨胀，例如 64 比特水印信息进行纠错编码以后可能会变成 256 比特；

c) 加密/扰动编码：在一些高安全性要求的场景下，通过映射、扰动或者加密等操作来增加水印信息的抗攻击能力。一些通用的流加密技术、位置扰动、映射表加密技术等编码技术都适用于此；

d) 校验编码：以一定的规则，通常是较为成熟的校验码算法，为水印信息生成可以检验内容真实性的信息流，并将校验码与水印信息相互融合；

e) 扩频编码：对信息进行扩频调制，增加原始水印信息中每一比特的出现频率和次数。该编码方法以成倍的水印信息膨胀为代价来增加水印的鲁棒性；

f) 同步码：在一些噪声场景下，水印信息可能会面临着被裁剪或者截断的危险。这种场景下加入同步码可以有效帮助寻找水印信息周期的起始位置；

g) 图形化编码：以图形的形式携带原水印信息；

h) 压缩编码：改变（通常是减少）水印信息的长度的编码方式。

水印解码算法是指从提取的水印码流中恢复出有意义水印信息的算法，通常来说是上述编码算法的逆过程。

A.3 常见水印算法与主要类型水印载体的适配情况

常见水印算法与主要类型水印载体的适配情况见表 A.1。

表 A.1 常见水印算法与主要类型水印载体的适配情况

算法类型	算法名称	文档	图像	音频	视频	网页	数据库
水印嵌入/提取算法	元数据水印	√	√	√	√	×	√
	模板水印	×	√	√	√	√	×
	时序水印	×	×	√	√	√	×
	变换域水印	×	√	√	√	×	√
	直方图水印	×	√	√	√	×	√
	最低有效位水印	×	√	√	√	×	√
	内容水印	√	×	×	×	√	√
	不可见元素水印	√	√	×	√	√	√
	深度学习水印	√	√	×	√	√	√
水印编码/解码算法	二进制明文编码	√	√	√	√	√	√
	纠错编码	√	√	√	√	√	√
	加密/扰动编码	√	√	√	√	√	√
	校验编码	√	√	√	√	√	√
	扩频编码	√	√	√	√	√	√
	同步码	√	√	√	√	√	√
	图形化编码	√	√	√	√	√	√
	压缩编码	√	√	√	√	√	√

注： √表示适配，×表示不适配

附 录 B
（资料性）
典型安全场景

B.1 数据版权保护

数据版权保护场景主要是利用数字水印来保护数字内容的版权，在发生相关侵权行为的时候可以利用数字水印来进行鉴权，以维护数字内容拥有者的合法权益，其主要场景如图 B.1 所示：

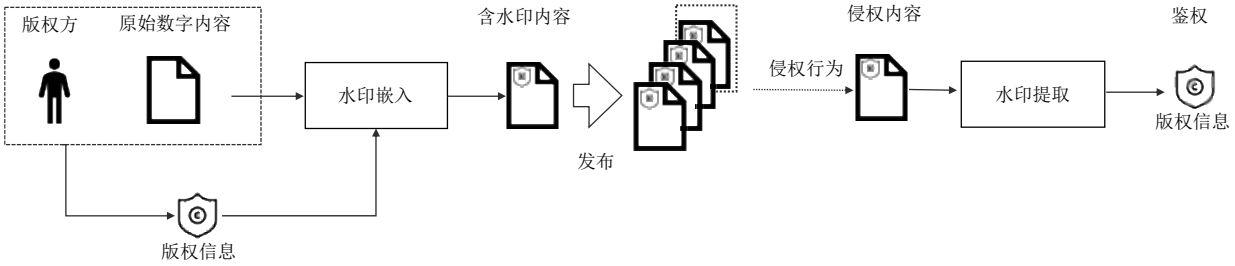

图 B.1 数字内容版权保护场景

在此场景中：

a) 版权方将版权信息以水印的形式嵌入到对应的原始数字内容中，获得含水印的数字内容后再进行发布和传播。其中，水印的嵌入可以依赖第三方可信机构或者采用经认证的公开方法来完成；

b) 当含水印的数字内容在传播过程中遇到盗版、挪用等版权纠纷时，版权方可以通过水印提取算法从对应的侵权内容中提取出版权信息，通过水印所表达的语义信息或通过第三方版权认证&管理机构判定版权归属，从而维护自身权益，达到版权保护的目的；

c) 数据版权保护场景中所使用的版权信息可以是经第三方可信机构认证的。该类版权信息通常是版权号、序列号等非自然语义信息，需要第三方可信机构进行相应的管理。该类版权信息通常适用于原始数字内容的水印容量较小的情况下。版权信息在提取出来以后，需要第三方可信机构的参与来进行鉴权；

d) 数据版权保护场景中所使用的版权信息亦可是具有具体版权声明语义内容的明文水印信息，例如“XXX 有限公司版权所有”。该场景通常适用于水印容量较大，足够携带完整自然语义版权信息的情况下。该类版权信息在提取出来以后，通常可以直接通过原语义信息进行鉴权。

B.2 数据泄露追踪溯源

数据泄露追踪溯源场景主要是利用不同的数字水印信息来对数字内容分发过程中可能会发生的泄露行为进行溯源，以求找到泄露的具体渠道方，其主要场景如图 B.2 所示：

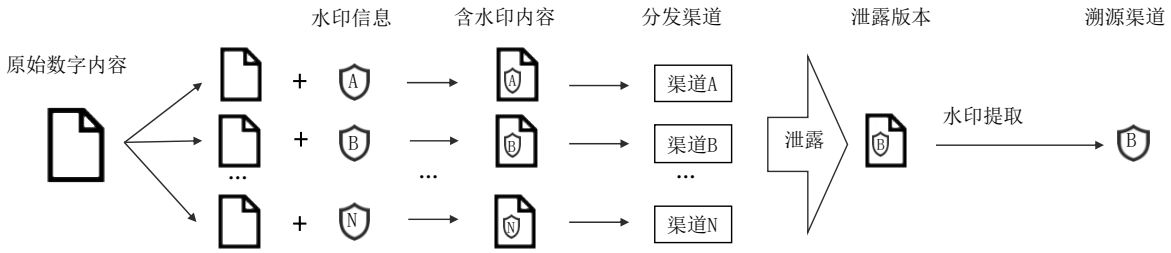

图 B.2 数据泄露追踪溯源场景

在此场景中：

a) 数字内容的拥有者会为所有分发渠道设计各不相同的溯源水印，并维持一个水印与分发渠道的映射关系。通常来说这样的映射关系是一一映射。

b) 在数据分发前，原始数字内容会被拷贝成多份，并在每一份中嵌入各个分发渠道所对应的溯源水印信息，从而获得携带不同溯源水印的数字内容。

c) 在数据分发后，内容相同但是溯源水印不同的各版本数据将分别分发给各对应渠道以供其内部使用。

d) 如果任何一个渠道发生了数据泄露事件，在获取泄露版本后，内容拥有者可以通过提取版本中对应的溯源水印信息来定位到对应的泄露渠道，并根据需要来进行追责等处置。

B.3 生成式人工智能生成内容的水印标识

在生成式人工智能生成内容的水印标识场景中，数字水印被用以注明内容的生成源头等信息。该类应用场景中，数字水印以隐式水印标识方式标明该内容的服务提供者、内容 ID 等众多相关辅助信息。数字水印与生成式内容强绑定，具有携带水印信息容量大、不影响生成式内容二次加工等特点，给人工智能治理带来极大的便利。生成式人工智能生成内容的水印标识目前主要应用于图像、音频、视频等载体，其主要场景如图 B.3 所示：

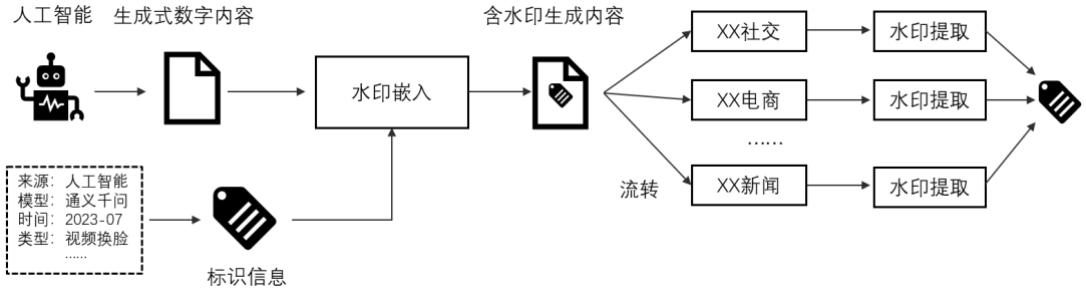

图 B.3 生成式人工智能生成内容的数字水印标识场景

在此场景中：

a) 生成式人工智能通过算法生成数字内容，包括但不限于图像、音频、视频等类型，随着人工智能技术的发展，此类内容已经达到以假乱真的程度。

b) 人工智能生成内容过程中涉及到的相关信息将会被制作成标识信息，标识信息主要包括服务提供者名称，也可包括内容 ID 等其他内容。

c) 上述步骤 b)中所述标识信息以水印的形式被嵌入到上述步骤 a)中对应的生成式内容中，获得含水印的生成式内容。通常来说，该场景下所采用的水印技术宜选择常见的水印算法，以保证其通用可提取性。此后，含水印的生成式数字内容会被正常分发和流转。

d) 含水印的生成式数字内容在流转过程中可能被传播至多种目标受众，各目标受众通过提取内容中的水印信息来判断对应的内容来源。

B.4 网络数据分类分级标识及管理

网络数据分类分级标识及管理场景主要是利用不同的数字水印信息作为数字内容的分类分级标识。数字水印形式的标识可以与对应的内容始终绑定在一起，无需借助多余的数据库、映射表等数据管理系统，具有极大的应用便捷性，其主要场景如图 B.4 所示：

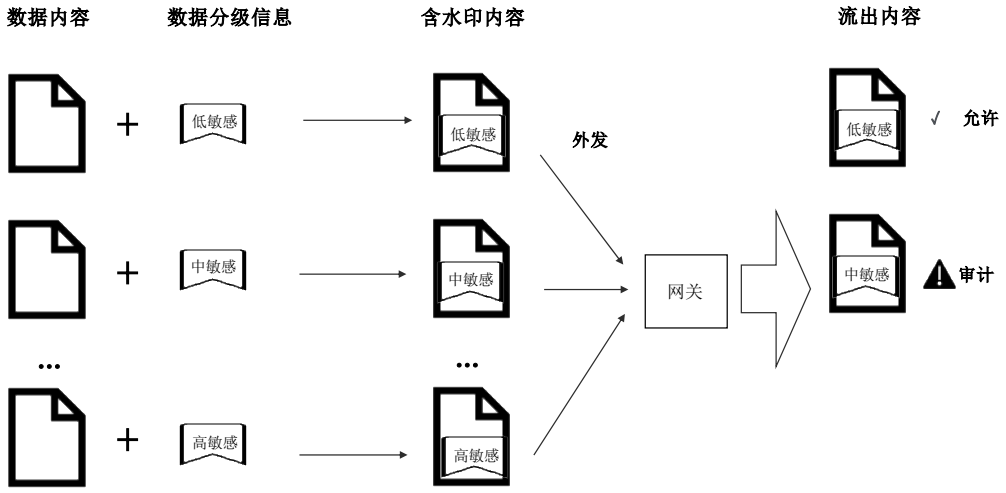

图 B.4 网络数据分类分级标识及管理场景

在此场景中：

a) 不同的数字内容经过数据分类分级系统获得对应的分类分级标识信息。

b) 上述分类分级标识信息以水印的形式被嵌入到对应的内容中，获得含有分类分级标识的数字内容。

c) 上述含水印内容在外发过程中，网关可以通过提取水印获得该内容对应的数据分级信息，从而采取对应的安全措施，如拦截高敏感数据，仅发出中、低敏感数据，同时对中敏感数据进行安全审计。

B.5 数据完整性保护

数据完整性保护场景主要是利用易碎数字水印作为内容完整性的“鉴定器”。易碎水印在载体经过细微的操作后即完全损坏而无法提取，其主要场景如图 B.5 所示：

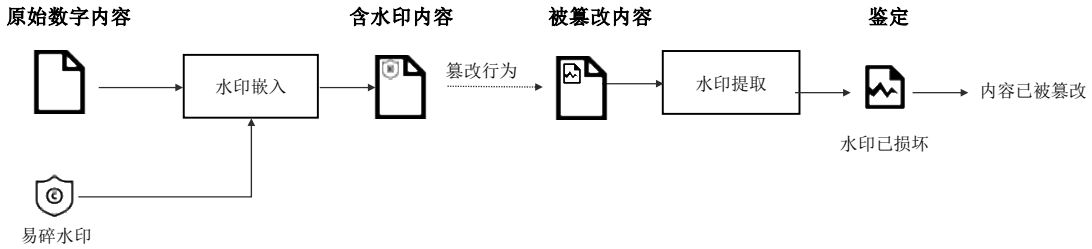

图 B.5 数据完整性保护场景

在此场景中：

a) 易碎水印被添加到受保护的原始数字内容中，获得含水印的数字内容。

b) 上述受完整性保护的含水印数字内容在传播过程中如果经历了删除、修改等攻击，其中嵌入的水印将会被破坏。

c) 内容接收者在获得内容后提取其中的易碎水印进行内容完整性鉴定，易碎水印被破坏的表现形式包括不限于提取失败、提取出的是乱码或空白信息等。

d) 若接收者发现上述水印被破坏的现象，则可鉴定出所接收数字内容已被篡改。

附 录 C
（资料性）
水印技术功能达成情况判定方式

针对典型水印技术的功能，可通过以下建议方式对其达成情况进行判定。

C.1 基本功能

a) 保证水印信息隐蔽

根据算法适用的水印载体，选取典型样本进行水印的编码和嵌入，观察嵌入水印的样本，判断是否可察觉水印的存在。

无法察觉则视为达成功能。

b) 确保载体正常使用

根据算法适用的水印载体，选取典型样本进行水印的编码和嵌入，使用正常方式打开嵌入水印的样本，判断是否影响正常使用。

可正常打开且不影响正常使用则视为达成功能。

c) 支持水印信息提取

根据算法适用的水印载体，选取典型样本进行水印的编码和嵌入后，再进行水印的提取和解码，读取解码后的水印信息，判断是否可解读有实质意义的信息并与原始被编码信息一致。

可解读有实质意义的信息，并且与原始被编码信息一致则视为达成功能。

C.2 增强功能

a) 防御水印攻击

根据算法适用的水印载体，选取典型样本进行水印的编码和嵌入后，使用几何变换（如图像视频的缩放、裁剪、旋转、平移、仿射、透视；音频的频度缩放、裁剪、时长拉伸）、内容篡改（如拼接、替换、内容增删改）、信号处理（如图像视频亮度、对比度、滤波；音频均衡、滤波、降噪）和二次数字化（如打印、扫描、拍照、录屏、录音）等方法进行恶意攻击，再进行水印的提取和解码，读取解码后的水印信息，判断是否可解读有实质意义的信息并与原始被编码信息一致。

可解读有实质意义的信息，并且与原始被编码信息一致的比例达到预期则视为达成功能。

b) 抵抗失真干扰

根据算法适用的水印载体，选取典型样本进行水印的编码和嵌入后，将水印载体选择常用压缩工具进行压缩后解压，使用常用社交工具进行传输后下载，再进行水印的提取和解码，读取解码后的水印信息，判断是否可解读有实质意义的信息并与原始被编码信息一致。

可解读有实质意义的信息，并且与原始被编码信息一致的比例达到预期则视为达成功能。

C.3 特定功能

a) 满足大容量需求

根据大容量水印信息场景需求，明确水印信息包含的具体容量长度，选取典型样本进行水印的编码和嵌入后，再进行水印的提取和解码，读取解码后的水印信息，判断是否可解读出需覆盖的水印内容。

可完整解读出全部水印信息视为达成功能。

b) 满足实时性需求

根据场景需求，在实时业务过程中对样本进行水印的编码和嵌入后，再进行水印的提取和解码，读取解码后的水印信息，判断是否对实时业务产生不良影响。

对实时业务不产生不良影响视为达成功能。