

面向铁路数据安全的不可见字符水印溯源系统设计

齐晨虹, 赵红涛, 尹文志, 许丹亚, 朱明

(中国铁路郑州局集团有限公司 信息技术所, 河南 郑州 450052)

摘要: 随着铁路信息化的发展, 铁路数据呈现爆发式增长, 随之而来的隐私泄露、数据买卖、网络攻击等数据安全问题日益凸显。针对现有铁路信息系统难以有效追踪数据泄露源的问题, 设计并实现了一套基于不可见字符水印的铁路数据溯源系统。该系统采用零宽字符技术构建水印载体, 开发了动态水印嵌入算法和提取工具, 实现了对铁路各类数据的安全标识。该系统由水印生成、数据嵌入和水印提取等模块组成, 针对铁路数据的多源异构性设计了统一的嵌入规则, 并根据不同铁路数据的敏感程度自动调整水印参数, 实现差异化保护。研究表明, 该系统具有良好的不可感知性和鲁棒性, 可有效支撑铁路数据的安全管理与泄露溯源, 为保障铁路信息系统安全提供了新的技术手段。

关键词: 数据安全; 不可见水印; 零宽字符; 溯源分析; 铁路信息系统; 水印算法

DOI: 10.11907/rjdk.241935

扫描二维码阅读全文:



中图分类号: TP31

文献标识码: A

文章编号: 1672-7800(2026)002-0126-09

Design and Implementation of Invisible Character Watermark Traceability System for Railroad Data Security

QI Chenhong, ZHAO Hongtao, YIN Wenzhi, XU Danya, ZHU Ming

(Institute of Information Technology, China Railway Zhengzhou Group Co. Ltd., Zhengzhou 450052, China)

Abstract: With the development of railway informatization, railway data is experiencing explosive growth, which has led to increasingly prominent data security issues such as privacy breaches, data buying and selling, and network attacks. A railway data traceability system based on invisible character watermarking has been designed and implemented to address the problem of existing railway information systems being difficult to effectively trace the source of data leaks. The system uses zero width character technology to construct watermark carriers, develops dynamic watermark embedding algorithms and extraction tools, and achieves secure identification of various types of railway data. The system consists of modules such as watermark generation, data embedding, and watermark extraction. A unified embedding rule is designed for the multi-source heterogeneity of railway data, and watermark parameters are automatically adjusted according to the sensitivity of different railway data to achieve differentiated protection. Research has shown that the system has good imperceptibility and robustness, which can effectively support the security management and leak tracing of railway data, providing new technological means for ensuring the security of railway information systems.

Key Words: data security; invisible watermark; zero-width character; traceability analysis; railroad information system; watermarking algorithm

收稿日期: 2024-12-13

基金项目: 中国国家铁路集团有限公司科技研究开发计划项目(Q2023W002)

作者简介: 齐晨虹(1989-), 女, 中国铁路郑州局集团有限公司信息技术所高级工程师, 研究方向为数据管理和数据安全; 赵红涛(1976-), 男, 中国铁路郑州局集团有限公司信息技术所高级工程师, 研究方向为信息技术和大数据; 尹文志(1993-), 男, 中国铁路郑州局集团有限公司信息技术所工程师, 研究方向为数据治理和数据安全; 许丹亚(1996-), 女, 中国铁路郑州局集团有限公司信息技术所工程师, 研究方向为数据安全和大数据; 朱明(1997-), 男, 中国铁路郑州局集团有限公司信息技术所助理工程师, 研究方向为数据安全和数据分析。本文通讯作者: 齐晨虹。

0 引言

随着铁路数字化转型进程的加快,铁路信息系统已积累海量数据,涵盖业务运营、基础设施和乘客信息等方面,广泛应用于列车调度、设备维护和客运服务等领域^[1]。铁路数据具有覆盖范围广、类型繁杂且体量庞大等特点^[2]。海量数据的互联互通释放了其潜在价值,而数据泄露和非授权访问等安全风险也随之日益突出^[3-4]。铁路数据承载了重要而敏感的业务信息,一旦发生数据泄露、窃取或篡改等事件,将会对铁路运营造成严重影响^[5]。尽管传统的加密和访问控制等安全措施得到广泛应用,但其在数据共享和流转过程中存在取证困难、追责无据等问题。因此,在保障数据高效流通的同时实现有效的安全防护和泄露溯源,已成为铁路信息化建设亟需解决的重要问题之一。

1 相关工作

在铁路数据安全方面,学者们开展了广泛研究。江文彬等^[6]提出基于层次化主题分析的铁路敏感数据智能识别与分类分级方案,实现了对铁路敏感数据的准确识别与定级。李清欣等^[7]提出面向数据安全流通的铁路隐私计算方案。米丽等^[8]从数据安全能力管理、数据全生命周期安全管理和通用安全管理等3个方面构建铁路货运数据安全模型和安全策略架构。李苹^[9]结合数据标识、访问控制、数据加密和脱敏等措施,提出铁路敏感数据流通安全管控体系。然而,目前在铁路数据溯源方面缺少深入研究。

数字水印技术通过对数据进行隐蔽性修改,将特定信息嵌入,同时不影响数据的正常使用和感知,但可在需要时提取水印信息识别^[10]。传统的数字水印技术在多媒体内容保护,如图像、视频和音频等领域,已有广泛应用^[11]。在铁路数据全生命周期中,文本信息在所有信息媒介中具有十分重要的地位和作用。其具有简洁的形式和丰富的含义,所能传达的信息量是其他信息难以比拟的,世界上80%以上的信息是以文本为载体进行传播的^[12]。然而,文本水印技术的发展面临着诸多挑战。相较于其他数字媒体,文本包含的冗余信息较少,限制了水印嵌入的空间,增加了技术实现的难度。此外,各种复制设备的广泛使用也给水印提取带来新的挑战。目前文本水印技术可分为基于文本图像化、基于自然语言处理和基于文档格式的方法3类^[13]。其中,基于文本图像化的方法直接将文本转化为一幅二值化的图片文档,对其进行分块处理后在不同数字块之间进行信息隐藏^[14-15]。基于自然语言处理的方法通过分析文本内容,对部分词句进行同义替换,在确保不改变原文含义且不影响文本阅读体验的前提下实现信息隐藏^[16-17]。基于文档格式的方法与文本的格式内容相关,将

嵌入水印留于文档表层^[18-19]。

基于上述文本水印技术,学者们开发了文本水印系统。例如,董洁^[20]利用基于文档格式的文本水印技术开发了基于零宽字符的XML遥感文件追溯系统,该系统使用零宽字符、MD5算法和海明码算法,结合XML文档特点,将水印信息隐写在XML数据中。许佳君^[21]设计了基于Hadamard变换的文本水印算法对二值文本图像进行字符划分,对获取字符的实际修改量进行相应的像素翻转,基于此开发了可以抵抗手机拍照的文本水印系统。任盼^[22]通过对常见的文本属性列表代表进行视觉与语义冗余评估,设计了基于自然语言处理技术的结构化文本水印算法,开发了结构化数据水印系统。

尽管上述文本水印系统能够在数据流通过程中保障版权所有者权益,但仍存在一些局限性,例如未能充分考虑铁路数据的多源异构特性和不同敏感度的差异化保护需求。铁路数据通常来源多样,包括列车运行数据、客运数据、货运数据、设备维护数据等,每种数据在格式、精度和内容上均存在显著差异。现有水印系统未能针对这种多源异构特性设计专门的水印嵌入和提取策略,难以适应不同类型铁路数据的技术特征。同时,铁路数据的敏感程度存在显著差异,如线路规划、站点位置、关键基础设施等数据的敏感性各不相同,但现有系统采用统一的水印保护模式,无法实现差异化的安全保护,导致资源浪费。因此,开发针对铁路数据多源异构特性和差异敏感度的专门水印系统成为亟待解决的技术难题。为此,本文结合零宽字符水印技术与铁路数据特征,开发了一种适应铁路全域数据的不可见字符水印溯源系统。该系统根据不同铁路数据的敏感程度自动调整水印参数,实现差异化保护;同时考虑了铁路数据的异构性,设计了适用于不同类型数据的水印嵌入策略,有效弥补了现有方法在铁路特定场景下应用的不足。

2 系统设计

2.1 系统框架

面向铁路数据安全的不可见字符水印溯源系统的总体目标为在铁路行业多样化的数据中嵌入不可见的字符水印,实现数据泄露的快速溯源。系统应能高效处理不同类型的铁路数据,包括但不限于文本文档、数据库记录、配置文件等,同时具备跨平台的能力,以适应铁路信息系统的复杂环境。基于这些需求,该系统采用Java语言开发,结合具有良好跨平台兼容性和灵活部署能力的Spring Boot框架。如图1所示,该系统采用经典的三层架构设计,分为表示层、业务逻辑层和数据访问层,各层完成不同的业务需求。

表示层作为系统的人机交互界面,采用Web界面和应用程序接口双模式设计。在Web端,系统提供铁路数据文

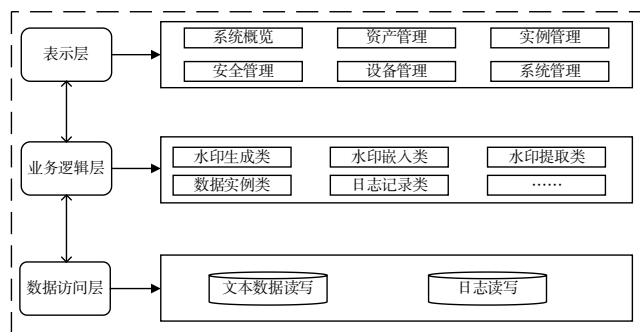


Fig. 1 System framework

图1 系统框架

件处理功能,包括文件上传、水印信息配置和嵌入策略选择等操作。针对铁路行业信息系统的特点,该层设计了标准化应用程序接口,实现了水印处理与数据溯源功能的无缝对接。所有用户输入的处理参数和配置信息均由表示层规范化处理后传递至业务逻辑层。

业务逻辑层作为不可见字符水印溯源系统的核心组件,主要负责实现水印生成、嵌入、提取等关键功能。该层包含水印相关算法、数据实例模块和日志记录模块等。水印算法通过联合编码策略将动态获取的应用系统登录账

户信息、访问终端地址与Web界面设定的固定水印信息联合在一起,生成适合嵌入的不可见字符水印序列,以实现可追踪水印信息的隐蔽嵌入。针对可能泄露的数据,运用水印提取算法进行分析,准确还原水印信息,快速定位泄露源头。该模块具有较强的容错性,针对碎片化数据也能溯源到有效信息。数据实例模块通过代理真实数据源提供虚拟实例和虚拟账户信息,在数据传输前实现水印信息的嵌入。日志记录模块详细记录水印处理全过程,包括操作时间、用户信息和异常告警等,支持后期的追踪分析与责任追究。

数据访问层主要负责系统中铁路数据的读写与存储管理操作,通过统一的数据访问接口实现对多源异构数据的高效读取与写入。系统支持对文本文件、CSV、Excel等多种格式的铁路数据进行处理,可与MySQL、Oracle等主流关系型数据库进行交互。

2.2 系统功能

面向铁路数据安全的不可见字符水印溯源系统采用模块化设计,主要由数据管理模块、水印生成嵌入模块、水印提取检测模块和日志审计模块组成。具体如图2所示。

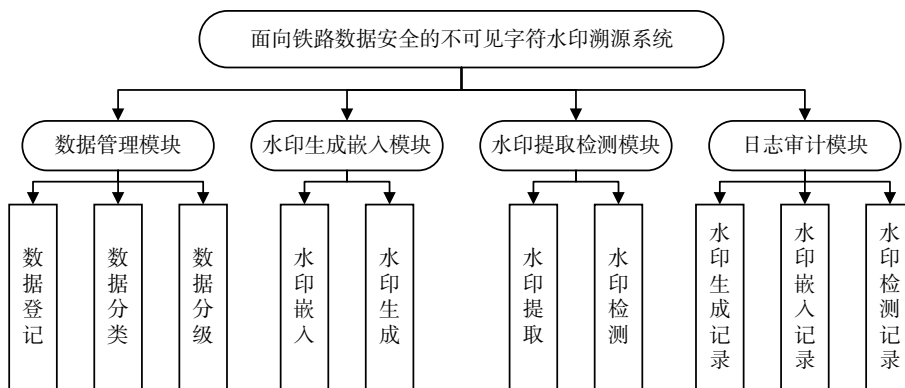


Fig. 2 System function modules

图2 系统功能模块

2.2.1 数据管理模块

数据管理模块是系统的基础功能模块,主要负责铁路数据信息的全面管理与维护。该模块具有数据登记、数据分类和数据分级3个功能。其中,数据登记功能用于铁路各单位对内部各项数据情况进行梳理。系统提供Excel录入、自动识别数据库信息、在线填报等多种方式,方便录入海量数据信息。数据分类功能根据各业务专家和内部分类标准设定多项分类指标,灵活地对铁路数据进行分类。数据分级功能用于评估数据的敏感程度,并作为后续水印嵌入程度的重要指标。采用自评估与智能评估相结合的方法,最终得到各项数据的敏感程度。

2.2.2 水印生成嵌入模块

水印生成嵌入模块主要负责实现水印的生成和嵌入两个关键功能。在水印生成功能中,用户通过页面输入创建固定水印信息,并将其转换为零宽字符序列。为防止大

量水印信息嵌入原数据后造成文本体积膨胀,在转换前采用SM3密码杂凑算法对水印信息进行加密。水印嵌入功能针对目标数据表执行水印嵌入操作,系统根据数据敏感程度自动嵌入固定水印和带有用户标识的混合水印,从而实现数据泄露的溯源追踪能力。

2.2.3 水印提取检测模块

水印提取检测模块主要负责从数据中提取和识别水印信息。系统支持对CSV、XLS、XLSX等多种常用文件格式的处理,用户只需上传可疑文件并点击提取按钮,即可快速识别并还原出隐藏的水印信息,有效实现数据泄露的追踪定位。

2.2.4 日志审计模块

日志审计模块是系统安全管理的重要组成部分,主要负责记录和监控系统中与水印操作相关的各类行为,包括水印生成日志、水印嵌入日志和水印检测日志3个功能模

块。水印生成日志功能模块通过实时记录水印生成操作,支持管理员根据日期、操作用户、水印类型、生成状态等多维度过滤条件进行精确查询,有效追踪水印信息的创建过程。水印嵌入日志功能模块专注于水印嵌入操作的监控,通过记录水印嵌入的目标数据表、虚拟实例、连接账号等信息,并支持按照嵌入时间、目标表名、嵌入状态和操作用户等条件进行筛选,实现对水印嵌入行为的全程追踪。水印检测日志功能模块则重点关注水印提取与检测操作,通过记录可疑文件的检测结果、提取时间、文件类型和检测状态等信息,支持管理员基于检测时间、文件名称、检测结果和操作用户等条件进行查询,以便及时掌握数据泄露的追踪情况。

3 关键技术

3.1 不可见字符水印信息生成

将原始水印转换为不可见字符水印的具体步骤为:①系统接收原始水印信息作为输入;②利用SM3密码杂凑算法对原始水印信息进行处理,生成固定长度的16进制摘要;③根据预定义的编码规则(如表1所示),将16进制摘要转换为不可见字符串,这一步骤是实现水印隐蔽性的关键;④存储原始水印信息及其对应的不可见字符串,建立二者之间的映射关系。

Table 1 Coding rules
表1 编码规则

16进制	替换字符
0	\u202d
1	\u206e
2	\u2060
3	\u200d
4	\u206b
5	\u200e
6	\u206c
7	\u206a
8	\u202c
9	\u202a
A	\u200b
B	\u202b
C	\u202e
D	\u200c
E	\u206f
F	\u206d

为提高水印算法的信息容量和安全性,采用分组策略对生成的不可见字符串进行拆分。分组数 n 的取值范围为 $1 \leq n \leq 15$,分组结构如图3所示。分组策略的具体实现方法为:①单一分组,当 $n = 1$ 时,组头序号为0xF,后接完整的不可见字符串;②双重分组,当 $n = 2$ 时,第一组的组头序号为0x0,后接不可见字符串的前半部分;第二组的组头序号为0x1,后接不可见字符串的后半部分;③多重分组,当 $2 \leq n \leq 15$ 时,不可见字符串被均匀分割为 n 个部分,每部分前面添加相应的组头序号(0到 $n - 1$)。



Fig. 3 Watermark grouping structure
图3 水印分组结构

通过上述步骤生成的不可见字符水印具有高度的隐蔽性和安全性,同时保留了足够的信息量以支持有效的数据溯源。

3.2 不可见字符水印信息嵌入

嵌入算法旨在将水印信息以不可见字符的形式嵌入到铁路数据中,同时保证数据的原始结构和可读性不受影响。水印嵌入流程如图4所示。算法具体步骤为:

- (1)系统通过识别分析数据结构,确定可以嵌入水印信息的数据列。
- (2)系统采用自评估与智能评估相结合的方法对数据进行敏感度分级评估。自评估为用户自行参照数据泄露对业务运行的影响程度,将数据划分核心、重要和一般3个级别。智能评估在自评估的基础上,结合历史同类数据的敏感程度进行加权平均。计算公式为:

$$y = \alpha A + \beta \frac{\sum_{i=1}^n B_i}{n} \tag{1}$$

式中: α 和 β 的值由相关业务专家定期进行综合数据敏感度审查并调整。

核心数据采用嵌入用户完整标识信息、用户访问时间、系统完整标识信息的追踪水印。重要数据采用嵌入用户账号标识信息和系统完整信息的追踪水印。一般数据采用嵌入用户账号标识信息和系统名称的追踪水印。

- (3)根据步骤(1)中列的位置获取对应列的数据。
- (4)系统根据步骤(2)的敏感度评估结果,从不可见字符集中选择相应等级的水印内容。
- (5)将抽取的不可见字符片段插入到步骤(3)得到的列中,更新该列的数据。
- (6)重复步骤(3)至步骤(5),直至当前数据行所有可嵌入的列均被处理,或者所有水印片段均已嵌入。
- (7)如果当前数据行的可用列不能容纳所有片段,系统会在下一行继续嵌入剩余片段,直到一条完整的不可见字符水印被完全嵌入。

(8)系统根据用户配置的嵌入密度,在一条完整的水印嵌入后重复执行步骤(3)至步骤(5),继续嵌入下一条水印信息。

通过以上流程,系统能根据数据的敏感程度选择合适的水印内容进行嵌入,既保证了重要数据的追踪保护,又避免了对普通数据的过度处理,在保证数据完整性和可用性的前提下,有效地将不可见字符水印信息嵌入到铁路数据中。

3.3 不可见字符水印信息提取

提取算法旨在从嵌入水印的铁路数据中识别并提取

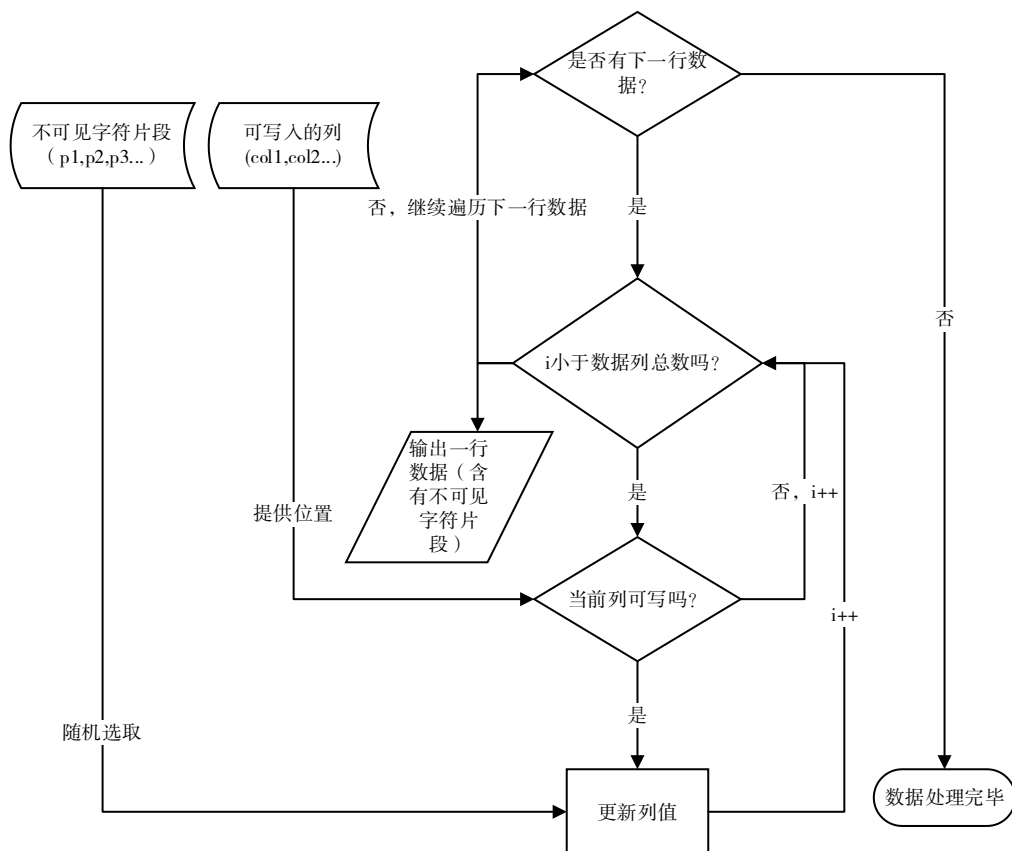


Fig. 4 Watermark embedding process

图4 水印嵌入流程

水印信息,具体步骤为:①系统读取用户上传的数据集文件,支持CSV和Excel等常用格式;②逐行读取步骤①的文件,将读取的内容放入队列,需设置队列的容量限制,避免一次性加载全部数据可能导致的内存溢出现象;③启动多个拆解线程,从队列中获取数据行来分析、识别和提取不可见字符片段;④根据提取的不可见字符片段的序号信息进行重组;⑤重复步骤②至步骤④,直至获取到一条完整的不可见字符串,或者队列中的数据行处理完毕;⑥将步骤⑤得到的不可见字符串与原始不可见字符串进行比对,计算其相似度;⑦系统输出匹配度最高的值所对应的实际水印信息和相似度。

以上流程在确保准确提取的同时亦考虑到实际应用场景中数据不完整或部分损坏的情况,增强了算法的鲁棒性。提取算法、生成算法和嵌入算法共同形成了完整的水印系统闭环,有效实现了数据的安全标记与追踪。

4 系统实现

4.1 主要功能

4.1.1 多源异构数据处理

本文系统采用Spring Boot作为基础框架,集成ETL组件、MyBatis框架等,对结构化、半结构化和非结构化数据进行预处理,便于后续嵌入零宽字符水印。结构化数据通

常存储在关系型数据库中,具有明确的模式和字段定义。系统采用MyBatis框架或JPA提取数据,对其进行必要的转换后批量加载到数据库中。针对半结构化数据,首先通过ETL Engine的Data Analysis模块解析JSON、XML等格式的数据提取工具,然后由Change SQL模块转换为标准数据库操作语句,将数据加载到数据库中。针对非结构化数据,通过ETL Engine的Data analysis进行分析,按照文件模板标识提取CSV、Excel、DBF文件中的增量数据,根据模板中的字段对应关系汇总成增量Change SQL,再使用Change SQL将数据加载到数据库中。各类数据统一输出到预处理器中,采用字符流加载,并对数据进行并行处理。

4.1.2 不可见字符水印信息生成

系统使用SM3密码散列算法对原始溯源信息进行处理,通过Bouncy Castle密码学库提供的SM3实现,将文本信息转换为固定长度的哈希值,确保水印信息的唯一性和不可伪造性。具体实现时,通过MessageDigest类获取SM3算法实例,对输入数据进行哈希计算。

系统建立了16进制值到零宽字符的映射机制,通过预定义的字符映射表将哈希值中的每个16进制数字转换为对应的零宽字符组合。这种映射关系由HashMap数据结构维护,以确保转换过程的高效性和可靠性。在字符串处理过程中,系统采用StringBuilder类进行动态字符串构建,并实现了基于分组的转换策略,有效控制了水印信息

的长度。

为保证水印信息的可靠存储与快速检索,系统采用多级存储架构,将水印映射关系存入 Redis 缓存,实现高速查询;同时通过 Spring Data JPA 框架将数据持久化到关系型数据库中,确保其长期保存。系统引入事务管理和缓存机制,有效提升了数据处理性能。此外,系统还集成了基于 SLF4J 的日志框架,将操作记录转存至 Elasticsearch,以支持后期的审计分析。

4.1.3 不可见字符水印信息嵌入

在不可见字符水印的嵌入过程中,系统通过 WatermarkEmbedService 类初始化嵌入服务,使用自定义的 DataStructureAnalyzer 类分析数据结构,识别适合嵌入水印的数据列。首先,对数据表结构进行解析,并返回可嵌入列的集合对象;其次,通过 DataSensitivityAnalyzer 类对数据进行敏感度分析,包括使用正则匹配识别个人敏感信息、检测业务敏感信息,并根据分析结果划分敏感等级;最后,使用 JdbcTemplate 创建数据操作接口,为后续的水印嵌入作准备。

针对不同的数据类型,数据读写操作需要采用不同的处理方法。对于结构化数据,通过 JdbcTemplate 对象获取数据集的相关信息,例如列名、数据类型、字符长度等,然后使用 ResultSetExtractor 接口获取数据对象,并通过 queryForList() 方法读取数据值。

在水印嵌入阶段,系统首先通过 SensitivityBasedWatermarkStrategy 类,根据数据敏感度等级选择相应的水印策略;其次,根据预设的嵌入密度参数,使用 StringBuilder 类对字符串进行修改,通过 insert() 方法在指定位置插入对应等级的水印信息。

图 5 和图 6 为系统提供的 Web 可视化配置界面,用户可通过该界面灵活设置水印嵌入参数。用户可以根据业务需求选择合适的数据列用于水印嵌入,同时设置水印嵌入密度阈值,实现对水印强度的精确控制。



Fig. 5 Setting fixed watermark content
图 5 设置固定水印内容

4.1.4 不可见字符水印信息提取

水印提取是水印嵌入的逆向过程,首先需要通过 EasyExcel 库的 read() 方法读取待处理的 Excel 和 CSV 格式的铁路数据文件。为了高效处理大规模数据,系统设计了



Fig. 6 Embedding strategy
图 6 嵌入策略

基于 BlockingQueue 接口的数据缓冲队列,用于实现数据的分批读取与处理。在具体实现中,系统通过 LinkedBlockingQueue 存储待处理的数据行。数据入队采用 offer() 方法,出队使用 poll() 方法,并通过 drainTo() 方法实现批量数据处理,有效防止内存溢出。系统构建了基于 ExecutorService 的线程池,支持数据的并行处理,每个处理线程通过自定义的 Extractor 类提取不可见字符水印信息。

在水印信息重组阶段,系统使用 TreeMap 数据结构确保水印片段按序号正确排序,通过 entrySet() 方法遍历,重组完整的水印信息。为提升系统的健壮性,采用 Java 8 的 Optional 类处理空值情况。水验证阶段通过自定义的相似度计算方法 getSimilarity() 评估提取水印的有效性。该方法基于字符串匹配原理计算待验证水印与原始水印的字符相似度。图 7 和图 8 为系统提供的提取界面与结果,展示了快速提取嵌入的水印内容和操作数据的用户标识信息。



Fig. 7 Extraction interface
图 7 提取界面

4.2 实验结果与分析

4.2.1 实验数据

为验证本文系统的有效性和鲁棒性,选取多种包含不同敏感数据类型的铁路文本数据进行实验,包括货运发送货票信息、职工人员主数据信息和收入文件核对信息。具体如表 2 所示。

4.2.2 不可见性分析

不可见性分析的目的为检验算法在直观上产生的差

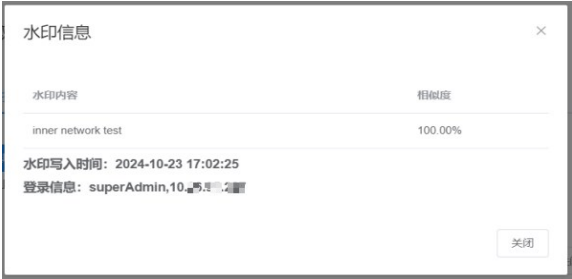


Fig. 8 Extracted results
图8 提取结果

异,即能否通过软件或者肉眼分辨出含有和不含有水印信息的影像。本研究采用未嵌入水印的数据作为原始数据,

Table 2 Test data
表2 实验数据

数据名称	文件格式	数据量/条
货运发送货票信息	Excel	3 000
职工人员主数据信息	CSV	1 000
收入文件核对信息	Excel	2 300

将其通过系统嵌入水印信息,然后在浏览器和办公软件上对结果数据和原始数据进行比较,结果如图9和图10所示(上半图为原始数据,下半图为结果数据)。可以看出,无法直观识别嵌入水印的数据与原始数据的差异,所有文本数据的字符串不可见性均为100%,说明系统具有良好的不可见性,水印嵌入没有对原始文本数据造成视觉影响。

日期	运单号	发段	发站	发货人名称	发专用线	到站
	请输入			请输入	请输入	
2024-11-21	888	车务段	待王	有限责...	有限责...	大家洼
2024-11-21	889	车务段	待王	有限责...	有限责...	大家洼
2024-11-21	831	车站	宣阳	有限公司		商丘
2024-11-21	847	车务段	兰考	有限公司		金马村
2024-11-21	898	车务段	待王	有限责...	有限责...	蒙阳
2024-11-21	255	车务段	阳城	有限公司	专线营...	沁阳
2024-11-21	256	车务段	阳城	有限公司	专线营...	沁阳
2024-11-21	257	车务段	阳城	有限公司	专线营...	沁阳
2024-11-21	258	车务段	阳城	有限公司	专线营...	沁阳
2024-11-21	259	车务段	阳城	有限公司	专线营...	沁阳

日期	运单号	发段	发站	发货人名称	发专用线	到站
	请输入			请输入	请输入	
2024-11-21	204888	车务段	待王	有限责...	有限责...	大家洼
2024-11-21	204889	车务段	待王	有限责...	有限责...	大家洼
2024-11-21	18931	车站	宣阳	流有限公司		商丘
2024-11-21	27547	车务段	兰考	有限公司		金马村
2024-11-21	204898	车务段	待王	有限责...	有限责...	蒙阳
2024-11-21	168255	车务段	阳城	源有限公司	专线营...	沁阳
2024-11-21	168256	车务段	阳城	源有限公司	专线营...	沁阳
2024-11-21	168257	车务段	阳城	源有限公司	专线营...	沁阳
2024-11-21	168258	车务段	阳城	源有限公司	专线营...	沁阳
2024-11-21	168259	车务段	阳城	源有限公司	专线营...	沁阳

Fig. 9 Comparison on browsers
图9 浏览器上的比较

4.2.3 鲁棒性分析

为验证系统的鲁棒性,对嵌入水印的文本数据进行多种常见攻击,包括随机删除、随机插入、随机替换等操作。通过计算攻击后提取的水印与原始水印的相似度 S 来评估其鲁棒性, S 值越接近1,表示鲁棒性越好。计算公

式为:

$$S(A,B)=\frac{\sum_{i=1}^n I(A_i=B_i)}{n} \tag{2}$$

式中: A 为原始水印字符串; B 为攻击后提取的水印字符串; n 为 A 与 B 之间较短字符串的长度; A_i 为 A 的第 i 个字

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	日期	运单号	发段	发站	到站	到局	到站	品名	中类	品类	中类	品名	票类	车种	车号		
2	20241121	04888	车务段	待王	焦大家洼	济山	山01	0130	煤	块煤	洗块煤	洗块煤	1	C70E			
3	20241121	04889	车务段	待王	焦大家洼	济山	山01	0130	煤	块煤	洗块煤	洗块煤	1	C70E			
4	20241121	18931	车站	宜阳	汝商丘	郑商	商08	0813	矿建材料	石料	石子	石子	3				
5	20241121	27547	车务段	兰考	曹金马村	昆	云10	1041	木材	普通人造板	普通人造板	普通人造板	1	P64GK			
6	20241121	04898	车务段	待王	焦蒙阳	郑国	国01	0140	煤	洗、选煤	末煤	末煤	1	C70EH			
7	20241121	68255	车务段	阳城	阝沁阳	郑山	山01	0140	煤	洗、选煤	末煤	末煤	1	C70E			
8	20241121	68256	车务段	阳城	阝沁阳	郑山	山01	0140	煤	洗、选煤	末煤	末煤	1	C64K			
9	20241121	68257	车务段	阳城	阝沁阳	郑山	山01	0140	煤	洗、选煤	末煤	末煤	1	C70E			
10	20241121	68258	车务段	阳城	阝沁阳	郑山	山01	0140	煤	洗、选煤	末煤	末煤	1	C70			
11	20241121	68259	车务段	阳城	阝沁阳	郑山	山01	0140	煤	洗、选煤	末煤	末煤	1	C70E			
12	20241121	82914	北车站	小宋	阝鹤壁时丰	郑河	鹤01	0110	煤	原煤	混煤	混煤	1	C64K			
13	20241121	82913	北车站	小宋	阝鹤壁时丰	郑河	鹤01	0110	煤	原煤	混煤	混煤	1	C64K			
14	20241121	82915	北车站	小宋	阝鹤壁时丰	郑河	鹤01	0110	煤	原煤	混煤	混煤	1	C70			
15	20241121	09141	车务段	戕庄	阝汝合肥北	上皖	皖01	0110	煤	原煤	混煤	混煤	1	C64H			
16	20241121	09142	车务段	戕庄	阝汝合肥北	上皖	皖01	0110	煤	原煤	混煤	混煤	1	C64K			
17	20241121	09145	车务段	戕庄	阝汝合肥北	上皖	皖01	0110	煤	原煤	混煤	混煤	1	C64K			
18	20241121	09146	车务段	戕庄	阝汝合肥北	上皖	皖01	0110	煤	原煤	混煤	混煤	1	C64K			
19	20241121	04149	车务段	宝丰	阝平商务	郑河	河01	0120	煤	洗精煤	洗精煤	洗精煤	1	C70			
20	20241121	09148	车务段	戕庄	阝汝合肥北	上皖	皖01	0110	煤	原煤	混煤	混煤	1	C64K			

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	日期	运单号	发段	发站	到站	到局	到站	品名	中类	品类	中类	品名	票类	车种	车号	箱数	箱型		
2	20241121	888	车务段	待王	焦大家洼	济山	山01	0130	煤	块煤	洗块煤	洗块煤	1	C70E	1				
3	20241121	889	车务段	待王	焦大家洼	济山	山01	0130	煤	块煤	洗块煤	洗块煤	1	C70E	1				
4	20241121	931	车站	宜阳	汝商丘	郑商	商08	0813	矿建材料	石料	石子	石子	3		1	T20			
5	20241121	547	车务段	兰考	曹金马村	昆	云10	1041	木材	普通人造板	普通人造板	普通人造板	1	P64GK	1				
6	20241121	898	车务段	待王	焦蒙阳	郑国	国01	0140	煤	洗、选煤	末煤	末煤	1	C70EH	1				
7	20241121	255	车务段	阳城	阝沁阳	郑山	山01	0140	煤	洗、选煤	末煤	末煤	1	C70E	1				
8	20241121	256	车务段	阳城	阝沁阳	郑山	山01	0140	煤	洗、选煤	末煤	末煤	1	C64K	1				
9	20241121	257	车务段	阳城	阝沁阳	郑山	山01	0140	煤	洗、选煤	末煤	末煤	1	C70E	1				
10	20241121	258	车务段	阳城	阝沁阳	郑山	山01	0140	煤	洗、选煤	末煤	末煤	1	C70	1				
11	20241121	259	车务段	阳城	阝沁阳	郑山	山01	0140	煤	洗、选煤	末煤	末煤	1	C70E	1				
12	20241121	914	北车站	小宋	阝鹤壁时丰	郑河	鹤01	0110	煤	原煤	混煤	混煤	1	C64K	1				
13	20241121	913	北车站	小宋	阝鹤壁时丰	郑河	鹤01	0110	煤	原煤	混煤	混煤	1	C64K	1				
14	20241121	915	北车站	小宋	阝鹤壁时丰	郑河	鹤01	0110	煤	原煤	混煤	混煤	1	C70	1				
15	20241121	141	车务段	戕庄	阝汝合肥北	上皖	皖01	0110	煤	原煤	混煤	混煤	1	C64H	1				
16	20241121	142	车务段	戕庄	阝汝合肥北	上皖	皖01	0110	煤	原煤	混煤	混煤	1	C64K	1				
17	20241121	145	车务段	戕庄	阝汝合肥北	上皖	皖01	0110	煤	原煤	混煤	混煤	1	C64K	1				
18	20241121	146	车务段	戕庄	阝汝合肥北	上皖	皖01	0110	煤	原煤	混煤	混煤	1	C64K	1				
19	20241121	149	车务段	宝丰	阝平商务	郑河	河01	0120	煤	洗精煤	洗精煤	洗精煤	1	C70	1				
20	20241121	148	车务段	戕庄	阝汝合肥北	上皖	皖01	0110	煤	原煤	混煤	混煤	1	C64K	1				
21	20241121	894	车务段	待王	焦蒙阳	郑国	国01	0140	煤	洗、选煤	洗末煤	洗末煤	1	C70E	1				
22	20241121	491	公司	李珍	阝汾水镇	济汾	汾03	0310	焦炭	焦炭	焦炭	焦炭	1	C64K	1				
23	20241121	492	公司	李珍	阝汾水镇	济汾	汾03	0310	焦炭	焦炭	焦炭	焦炭	1	C64K	1				

Fig. 10 Comparison on office software

图 10 办公软件上的比较

符; B_i 为 B 的第 i 个字符; $I(A_i = B_i)$ 为指示函数,当 $A_i = B_i$ 时为1,否则为0。

表3为不同文本数据在各种攻击下的鲁棒性实验结果。可以看出,在各种常见文本攻击下,本文系统提取的水印值均保持在0.6以上,具有较强的鲁棒性。即使在99%的攻击强度下,系统仍能有效提取部分水印信息。

Table 3 Robustness experiment results			
表 3 鲁棒性实验结果			
数据名称	攻击类型	攻击强度/%	S 值
货运发送货票信息	随机插入	100	1.0
职工人员主数据信息	随机删除	50	1.0
职工人员主数据信息	随机删除	99	0.8
收入文件核对信息	随机替换	50	1.0
收入文件核对信息	随机替换	99	0.6

综上所述,本文系统能够在保持原始文本数据完整性的同时有效嵌入和提取水印信息,为铁路数据的安全管理与溯源提供有力支持。

5 结语

基于铁路行业数据安全保护的迫切需求,本文设计并实现了一套不可见字符水印溯源系统。该系统通过零宽字符水印算法确保了水印的隐蔽性和鲁棒性,可有效抵御常见的数据处理操作和恶意攻击。实验结果表明,该系统性能指标优异,满足了铁路数据处理的实际需求,可被铁路部门、科研机构及相关政府单位应用于数据安全管理中,具有重要的实用价值与现实意义。

然而,该系统在实际应用中仍存在一些局限性,例如当面对超大规模数据处理时,系统可能出现效率瓶颈。此外,在日益复杂的网络环境下,现有防护机制仍有提升空间。未来可进行以下优化:一是引入分布式计算技术,提升系统的处理效率;二是整合先进的加密技术和动态水印策略,增强系统的安全性;三是探索将区块链、人工智能等新兴技术与水印系统相结合,提高数据的可信度和溯源

能力。

参考文献:

- [1] HU C H. Discussion on the innovation of passenger and freight transportation organization of Guangzhou Railway Group based on big data[J]. Comprehensive Transportation, 2017, 39(5): 28-32.
胡朝晖. 基于大数据的广铁集团客货运组织创新探讨[J]. 综合运输, 2017, 39(5): 28-32.
- [2] WANG Z. Research on the governance system of railway big data[J]. Cybersecurity and Data Governance, 2022, 41(11): 30-35.
王喆. 铁路大数据治理体系研究[J]. 网络安全与数据治理, 2022, 41(11): 30-35.
- [3] WEI C S, YANG Y J, LI Q, et al. Risk analysis and countermeasures of railway network data security [J]. Railway Computer Application, 2022, 31(11): 33-36.
魏长水, 杨轶杰, 李琪. 铁路网络数据安全风险分析及应对措施[J]. 铁路计算机应用, 2022, 31(11): 33-36.
- [4] LI F H, LI H, NIU B, et al. Research scope and future development trends of data element circulation and security [J]. Journal of Communications, 2024, 45(5): 1-11.
李凤华, 李晖, 牛霖, 等. 数据要素流通与安全的研究范畴与未来发展趋势[J]. 通信学报, 2024, 45(5): 1-11.
- [5] JI H Z, LI X Y. Railway cyberspace security defense based on key element management and control [J]. Railway Computer Application, 2024, 33(10): 35-39.
季宏志, 李向阳. 基于关键要素管控的铁路网络空间安全防护研究[J]. 铁路计算机应用, 2024, 33(10): 35-39.
- [6] JIANG W B, LIU Z L, XIE S K, et al. Intelligent identification and classification and grading method of railway sensitive data based on hierarchical topic analysis [J]. Railway Computer Application, 2024, 33(10): 7-12.
江文彬, 刘兆霖, 谢仕康, 等. 基于层次化主题分析的铁路敏感数据智能识别与分类分级方法[J]. 铁路计算机应用, 2024, 33(10): 7-12.
- [7] LI Q X, WU Y H, ZHOU W, et al. Railway privacy computing solution for data security circulation [J]. Railway Computer Application, 2024, 33(10): 78-82.
李清欣, 吴艳华, 周雯, 等. 面向数据安全流通的铁路隐私计算方案研究[J]. 铁路计算机应用, 2024, 33(10): 78-82.
- [8] MI L, HUANG M Z, MA Z Q, et al. Construction of railway freight data security system and security strategies [J]. Railway Computer Application, 2023, 32(12): 18-23.
米丽, 黄敏珍, 马志强, 等. 铁路货运数据安全体系构建及安全策略研究[J]. 铁路计算机应用, 2023, 32(12): 18-23.
- [9] LI P. Railway sensitive data circulation security management and control system [J]. Railway Computer Application, 2024, 33(10): 67-72.
李苹. 铁路敏感数据流通安全管控体系研究及应用[J]. 铁路计算机应用, 2024, 33(10): 67-72.
- [10] ZOU X Z, HU Y C, ZHU C Q. Design and realization of digital watermarking system for geographic information based on Linux system [J]. Software Guide, 2023, 22(11): 9-17.
邹秀珍, 胡宇宸, 朱长青. 基于Linux系统的地理信息数字水印系统设计与实现[J]. 软件导刊, 2023, 22(11): 9-17.
- [11] WU G H, GONG L C, YUAN L F, et al. Research progress on Chinese text information hiding [J]. Journal of Communication, 2019, 40(9): 145-156.
吴国华, 龚礼春, 袁理锋, 等. 中文文本信息隐藏研究进展[J]. 通信学报, 2019, 40(9): 145-156.
- [12] LING G W. Text watermarking algorithm and DSP implementation based on error correction coding [D]. Changsha: Central South University of Forestry and Technology, 2012.
凌国巍. 基于纠错编码的文本水印算法及DSP实现[D]. 长沙: 中南林业科技大学, 2012.
- [13] ZHAO W J, GUAN H, HUANG Y, et al. Review of text watermarking techniques [J]. Journal of the Communication University of China (Natural Science Edition), 2020, 27(6): 55-62.
赵卫娟, 关虎, 黄樱, 等. 文本水印技术研究综述[J]. 中国传媒大学学报(自然科学版), 2020, 27(6): 55-62.
- [14] YAO H B, WEI M X, ZHOU J W, et al. Print-scan invariant combined with dual domain text watermark algorithm [J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2018, 46(6): 122-127.
姚寒冰, 魏木鑫, 周俊伟, 等. 打印—扫描不变量结合双域的文本水印算法[J]. 华中科技大学学报(自然科学版), 2018, 46(6): 122-127.
- [15] WANG Y Y, GUAN H, HUANG Y, et al. Text image watermarking scheme on the DWT-DCT joint transform domain [J]. Computer Engineering and Design, 2020, 41(6): 1676-1682.
王玉莹, 关虎, 黄樱, 等. DWT-DCT联合变换域上的文本图像水印方案[J]. 计算机工程与设计, 2020, 41(6): 1676-1682.
- [16] MERAL H M, EMRE S, ERSIN Ü, et al. Syntactic tools for text watermarking [C]//9th Conference on Security, Steganography, and Watermarking of Multimedia Contents, 2007: 1-12.
- [17] LI S M, LI B N, YANG C. Chinese text watermarking algorithm based on chaotic AES and synonym extension [J/OL]. Journal of Jilin University (Engineering Edition), 2024-11-05. <https://doi.org/10.13229/j.cnki.jdxbgxb.20240248>.
李书明, 李冰楠, 杨超. 基于混沌AES和同义词扩展的中文文本水印算法[J/OL]. 吉林大学学报(工学版), 2024-11-05. <https://doi.org/10.13229/j.cnki.jdxbgxb.20240248>.
- [18] ZHANG Z Y, LI Q M, QI Y. Text watermark design based on invisible characters [J]. Journal of Nanjing University of Science and Technology, 2017, 41(4): 405-411.
张震宇, 李千目, 戚湧. 基于不可见字符的文本水印设计[J]. 南京理工大学学报, 2017, 41(4): 405-411.
- [19] YAO Y, LIU S H, WANG H, et al. Robust Chinese text watermarking based on character disturbance deformation and font bank replacement [J]. Journal of Cryptography, 2023, 10(4): 769-785.
姚晔, 刘书辉, 王慧, 等. 基于字符扰动变形和字库替换的鲁棒中文文本水印[J]. 密码学报, 2023, 10(4): 769-785.
- [20] DONG J. XML remote sensing file tracing algorithm based on zero width characters [D]. Langfang: North China Institute of Aerospace Engineering, 2021.
董洁. 基于零宽字符的XML遥感文件追溯算法[D]. 廊坊: 北华航天工业学院, 2021.
- [21] XU J J. Research on text watermarking algorithm based on Hadamard transform [D]. Beijing: Beijing Jiaotong University, 2020.
许佳君. 基于哈达玛变换的文本水印算法研究[D]. 北京: 北京交通大学, 2020.
- [22] REN P. Research on structured data watermarking based on classification [D]. Xi'an: Xi'an University of Electronic Science and Technology, 2021.
任盼. 基于分类的结构化数据水印研究[D]. 西安: 西安电子科技大学, 2021.

(责任编辑: 尹晨茹)