

基于混沌 AES 和同义词扩展的 中文文本水印算法

李书明¹, 李冰楠¹, 杨超^{2,3}

(1. 湖北师范大学 计算机与信息工程学院, 湖北 黄石 435002; 2. 湖北大学 计算机学院, 武汉 430062; 3. 湖北大学 大数据智能分析与行业应用湖北省重点实验室, 武汉 430062)

摘要: 针对目前基于修改式和生成式的文本水印算法普遍存在语义歧义较大、嵌入容量有限及安全性不足的问题, 本文提出了一种基于混沌 AES 和同义词扩展的改进算法。该算法通过设计基于 Sentence-Bert 的上下文搭配机制, 不仅能确保同义词替换后的上下文与原文语义高度相似, 而且能判别同义词篡改后产生的语义歧义; 将异体字作为同义词扩展, 在不引起语义歧义的前提下, 实现更多同义词替换, 从而提升嵌入容量; 利用混沌映射机制, 升级传统 MD5 算法, 构建混沌 AES 算法, 实现对隐匿信息的加密保护, 进而增加其破解难度。研究结果表明: 对比同类算法, 该算法在语义歧义、嵌入容量及安全性方面均有良好表现。

关键词: 文本水印算法; 混沌 MD5; 混沌 AES; 混沌映射机制; 上下文搭配机制; 同义词扩展

中图分类号: TP391; TP181 **文献标志码:** A **文章编号:** 1671-5497(2025)11-3715-12

DOI: 10.13229/j.cnki.jdxbgxb.20240248

Chinese text watermarking algorithm based on chaotic AES and synonym expansion

LI Shu-ming¹, LI Bing-nan¹, YANG Chao^{2,3}

(1. College of Computer and Information Engineering, Hubei Normal University, Huangshi 435002, China; 2. College of Computer, Hubei University, Wuhan 430062, China; 3. Hubei Key Laboratory of Big Data Intelligent Analysis and Industry Applications, Hubei University, Wuhan 430062, China)

Abstract: At present, the text watermarking algorithms based on modification and generation generally have the problems of large semantic ambiguity, limited embedding capacity and insufficient security. In view of the above problems, this paper proposes an improved algorithm based on chaotic AES and synonym expansion. By designing a context matching mechanism based on Sentence-Bert, the algorithm can not only ensure that the context after synonym replacement is highly similar to the semantics of the original text, but also distinguish the semantic ambiguity caused by synonym tampering. The variant characters are extended as synonyms, and more synonyms are replaced without causing semantic ambiguity, thereby improving the embedding capacity. By using chaotic mapping mechanism, the traditional MD5 algorithm is upgraded, and the chaotic AES algorithm is constructed to realize the

收稿日期: 2024-03-11.

基金项目: 湖北省教育厅 2023 年教学改革研究项目 (2023388); 国家自然科学基金项目 (61977021)

作者简介: 李书明 (1968-), 男, 教授, 博士. 研究方向: 密码学与区块链技术. E-mail: 10210008@hbnu.edu.cn

encryption protection of hidden information, thereby increasing the difficulty of cracking. The research results show that compared with similar algorithms, the algorithm has good performance in semantic ambiguity, embedding capacity and security.

Key words: text watermarking algorithm; chaotic MD5; chaotic AES; chaotic mapping mechanism; context matching mechanism; synonym expansion

0 引言

文本水印技术是一种将隐匿信息嵌入文本的信息隐藏技术^[1],被广泛应用于版权保护和文本隐写^[2]等领域。该技术要求在嵌入时不能破坏文本的原始语义,而在遭受攻击时仍能保持原文完整性或可检测性,这使得该研究难度较高且具有挑战。

根据文本信息隐藏的原理,将文本水印算法分为基于修改式和生成式两大类。在修改式领域,费文斌等^[3]采用了上下文搭配度替换同义词的方法,尽管降低了语义歧义,但存在嵌入容量受限和隐匿信息未受保护的问题;Xiang等^[4]基于信息论,利用串表压缩算法对秘密信息进行压缩,虽变相提升了嵌入容量,却未对隐匿信息加密保护;姚晔等^[5]将同一汉字编号为不同扰动变形字符,尽管在语义歧义和嵌入容量上表现良好,但隐匿信息以明文嵌入风险较大;张娜^[6]采用无载体零水印、信息熵和 RSA 三者结合的方式,尽管保证了语义的完整和隐匿信息的安全,但嵌入容量受限。在生成式领域,Zheng等^[7]采用 BERT 掩码语言模型生成词候选替换列表,虽突破了同义词库的限制,但容易产生语义歧义;Yu等^[8]运用 GPT 模型集成多位置选词来优化单位位置选词,避免了定长编码选词不当的问题,提升了文本质量;金家立^[9]将机器翻译、自动选择编码机制和动态选择策略三者相结合,在保证嵌入容量的同时,译文质量也得以提升。上述生成式方法在隐匿信息的安全性上同样考虑不足。

由上述内容可知,基于修改式和生成式的文本水印算法在文本的语义歧义和嵌入容量上普遍表现良好,而在隐匿信息的安全性上考虑不足。因此,为同时兼顾语义歧义、嵌入容量和安全性,本文设计了一种中文文本水印改进算法,通过设计基于 Sentence-Bert^[10]的上下文搭配机制,不仅能确保同义词替换后的上下文与原文语义高度相

似,而且能判别同义词篡改后产生的语义歧义;将异体字作为同义词扩展,在不引起语义歧义的前提下,实现更多同义词替换,从而提升嵌入容量;利用混沌映射机制,升级传统 MD5 算法,构建混沌 AES 算法,实现对隐匿信息的加密保护,进而增加其破解难度。

1 相关工作

1.1 传统 MD5 算法的抗破解能力

近 20 年,MD 结构哈希函数安全性研究取得了显著突破,Wang 等^[11]利用差分路径成功实现对 MD5 的近似碰撞攻击;Klima^[12]使用隧道技术大幅提升了 MD5 碰撞消息的搜索效率,而 Stevens^[13]结合差分路径和隧道技术提出了基于中间相遇的自动构造 MD5 差分路径方法,进一步提升了 MD5 碰撞消息的搜索效率。随着碰撞攻击技术的深入发展,MD5 近似碰撞攻击的复杂度正持续降低。

1.2 传统 AES 算法的抗破解能力

近 10 年,有关 AES 的攻击方法取得一定进展,Ronjom 等^[14]利用 YoYo Tricks 提出了 5 轮 AES 的密钥恢复攻击方法;Bardeh 等^[15]提出了交换攻击方法,降低了 YoYo Tricks 中保持零差分模式的概率,实现了 6 轮 AES-128 区分器;Dunkelman 等^[16]融合 YoYo Tricks 与 Boomerang,提出 Retracing Boomerang 攻击方法,实现了 5 轮 AES 最优攻击,时间复杂度仅为 $2^{16.5}$;Bardeh 等^[17]实现了 7 轮 AES-128 的相关差分攻击。关于 AES 攻击方法的研究仍在持续,需要通过增加密钥复杂度、迭代轮数,打乱 S 盒顺序等方法降低被攻击的风险。

1.3 异形词、异体字

异形词是指普通话书面语中并存并用的同音、同义而书写形式不同的词语,如“成分-成份”“筹划-筹画”。异体字是指与正体字字音相同、字形不同而意义相同的字,它是一种汉字的特殊形

式,如“内-内”“出-出”。由此可知,采用异形词、异体字替换后的文本具备与原文语义歧义较小的特点;同时异形词和异体字的引入能对更多字词编码,使嵌入容量得到提升。

2 中文文本水印改进算法

为同时兼顾语义歧义、嵌入容量与安全性,本文设计了一种基于混沌 AES 和同义词扩展的改进算法,该算法包括嵌入和提取两部分,其核心为混沌映射机制和上下文搭配机制,本节将详细描述嵌入流程、提取流程和两大核心机制。

2.1 嵌入流程和提取流程

2.1.1 嵌入流程

嵌入流程,按编号顺序依次执行,如图 1 所示。第 1~3 步,通过基于混沌映射机制的混沌 MD5 算法将密钥加密为增强密钥。第 3~8 步,采用增强密钥,对隐匿信息的 AES 加密结果进行行列置换、移位代换和位运算,得到十六进制增强密文,增强算法使破解难度增加。第 8~9 步,将十六进制增强密文转码为二进制增强密文。第 9~13 步,将二进制增强密文编码为嵌入文本原始词对应的同义词,通过上下文搭配机制反复尝试,直至同义词替换后上下文不产生语义歧义即停止,同时将异体字作为同义词扩展,能增加嵌入容量。第 13~14 步,若嵌入容量满足所有二进制增强密文

文的嵌入,则水印文本制作成功,否则失败。

嵌入流程算法描述如下:

输入:①对称密钥(密钥),②隐匿信息,③嵌入文本。

输出:嵌入结果。

Begin

(1)步骤 1~3:

采用传统 MD5 算法计算密钥 MD5。

for 密钥 MD5 字符(混沌 MD5 算法)

①以每 4 位相邻的十六进制字符表示为 1 个(0~65 535)输入字节;

②采用混沌映射机制生成映射输出;

③得到映射输出中输入字节的匹配结果;

④将匹配结果表示为 4 位新生成的十六进制字符。

拼接所有新生成的十六进制字符,得到增强密钥。

(2)步骤 3~8:采用传统 AES 和增强密钥加密隐匿信息得到十六进制密文。

for 增强密钥

①十六进制密文通过相邻遍历结果实现行列置换;

②十六进制密文通过遍历结果实现移位代换;

③十六进制密文通过遍历结果实现位(非和异或)运算;

增强算法生效,混沌 AES 加密生成十六进制增强密文。

(3)步骤 8~9:将十六进制增强密文转码为二进制增强密文。

(4)步骤 9~13:

for 二进制增强密文

①通过查询同义词词库,将遍历结果编码为嵌入文本原始词对应的同义词(0 表示原始词,1 表示同义词);

if 上下文搭配机制校验同义词与原始词上下文的相似度 $\geq$ 阈值

then 匹配得到筛选后的同义词;

else 重新执行步骤①,直至成功;

②通过查询异体字字库,将遍历结果编码为嵌入文本原始字对应的异体字(0 表示原始字,1 表示异体字),并将其作为同义词扩展。

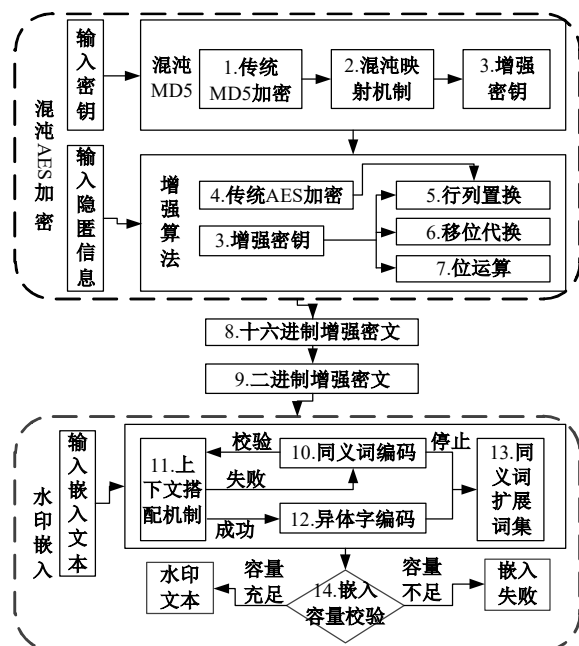


图 1 嵌入流程

Fig. 1 Embedding process

(5)步骤 13~14:

if 同义词扩展嵌入容量满足二进制增强密文的嵌入

then 水印文本制作成功;

else 嵌入容量不足,嵌入失败。

End

2.1.2 提取流程

提取流程,按照编号顺序执行,如图 2 所示。第 1~5 步,从水印文本提取同义词扩展词集并采用异体字和同义词解码,将同义词扩展词集对齐为原始词,通过上下文搭配机制校验水印文本与原始词上下文有无语义歧义,若有,则判别同义词被篡改,若无,则得到二进制增强密文解码结果。第 5~6 步,将二进制增强密文转码为十六进制增强密文。第 7~9 步,通过混沌 MD5 算法将密钥加密为增强密钥。第 9~13 步,采用增强密钥对十六进制增强密文进行增强算法逆向操作,即逆向位运算、逆向移位代换和逆向行列置换,再采用传统 AES 解密得到解密文本。第 14 步,若解密文本与原始隐匿信息一致,则提取成功,否则提示水印文本遭到破坏。

提取流程算法描述如下:

输入:①对称密钥(密钥),②水印文本。

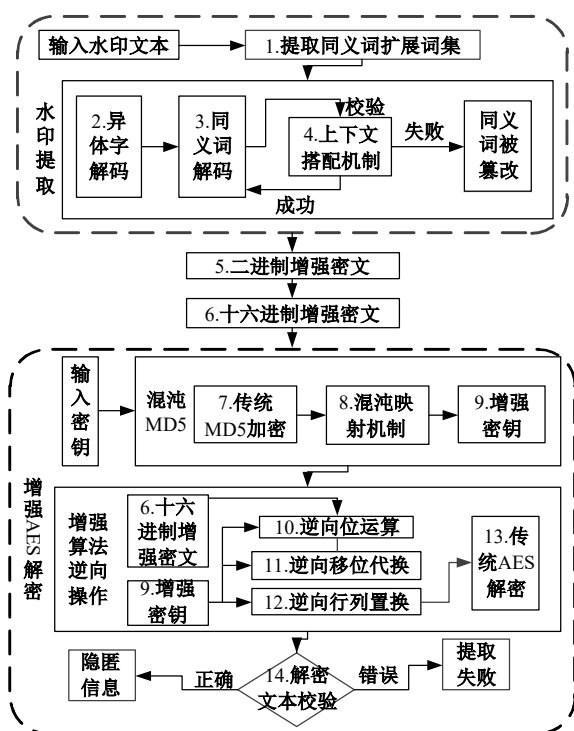


图 2 提取流程

Fig. 2 Extraction process

输出:提取结果。

Begin

(1)步骤 1~5:

for 水印文本

通过查询同义词词库和异体字字库,从遍历结果中提取同义词扩展词集,包括异体字和同义词。

for 同义词扩展词集

①从遍历结果解码异体字,得到二进制密文 1;

②将遍历的同义词对齐为水印文本对应的原始词。

if 上下文搭配机制校验同义词与原始词上下文的相似度 $\geq$ 阈值

then 解码同义词,得到二进制密文 2;

else 同义词被篡改,退出提取流程。

合并二进制密文 1 和密文 2,得到二进制增强密文。

(2)步骤 5~6:将二进制增强密文转码为十六进制增强密文。

(3)步骤 7~9:采用混沌 MD5 加密密钥,得到增强密钥。

(4)步骤 9~13:

for 增强密钥

①十六进制增强密文通过遍历结果实现逆向位(异或或非)运算;

②十六进制增强密文通过遍历结果实现逆向移位代换;

③十六进制增强密文通过相邻遍历结果实现逆向行列置换。

增强算法逆向操作完成,混沌 AES 解密生成十六进制密文。

采用传统 AES 和增强密钥解密十六进制密文得到解密文本。

(5)步骤 14:

if 解密文本与原始隐匿信息一致

then 隐匿信息提取成功;

else 水印文本遭到破坏。

End

2.2 混沌映射机制

图 3 为本文设计的混沌映射机制,共有 3 步。第 1 步,整数 0~65 535 的取值范围是 0x0000~

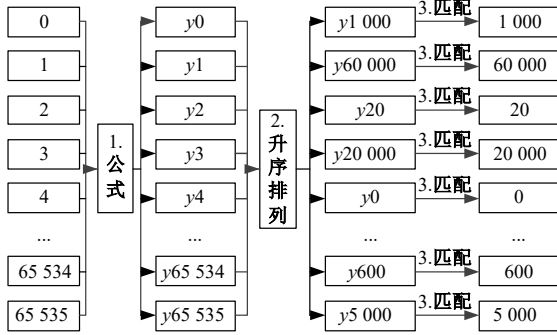


图 3 混沌映射机制

Fig. 3 Chaotic mapping mechanism

0xffff, 可表示 2 个字节或 4 位十六进制字符, 将参数 r 和 2 个字节 (0~65 535) 输入公式 (1) 进行变换; 第 2 步, 将变换后的结果按升序排列; 第 3 步, 在公式 (1) 变换结果中查询排列结果, 将匹配的索引作为输出, 如输入 0 输出 1 000、输入 1 输出 60 000、输入 2 输出 20 等。该机制生成的输出结果有高达 65 535! 种组合, 这使得增强密钥和增强算法生成的十六进制增强密文的取值结果组合较多, 枚举攻击会面临较高的时间成本, 因而破解难度较大, 具有安全性。

$$y = \frac{2 - \cos(-rx)}{2 + \sin(rx)} \quad (1)$$

混沌映射机制算法描述如下:

输入: 整数集合, $X = \{0, 1, 2, \dots, 65\,534, 65\,535\}$; 公式参数 r 。

输出: 整数集合, $Y = \{1\,000, 60\,000, 20, \dots, 600, 5\,000\}$ 且 $Y \neq X$ 。

Begin

① for x in X

将 x 和 r 输入公式 (1), 得到的输出结果加入集合 Y_1 ;

② 将集合 $Y_1 = \{y_0, y_1, y_2, y_3, \dots, y_{254}, y_{255}\}$ 按数值大小升序排列, 得到集合 Y_2 ;

③ for y in 集合 Y_2

将 y 在集合 Y_1 中的索引加入集合 Y ;

输出集合 Y 。

End

2.3 上下文搭配机制

图 4 为本文设计的上下文搭配机制, 旨在确保水印嵌入过程中文本前后语义一致, 不产生歧义, 而在提取过程中能判别同义词被篡改后产生的语义歧义, 从而维护水印文本语义的完整性。

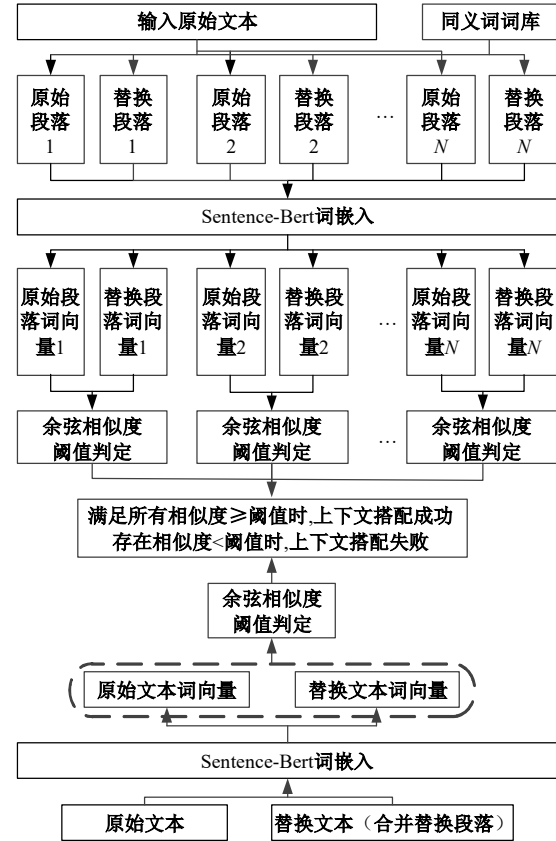


图 4 上下文搭配机制

Fig. 4 Context matching mechanism

在嵌入过程中, 原始文本是嵌入文本, 同义词词库是将嵌入文本中的原始词替换为同义词, 形成替换段落; 在提取过程中, 原始文本是水印文本, 同义词词库是将水印文本中的同义词对齐为原始词, 形成替换段落; 两个流程共用相同的机制。该机制首先将原始文本分成 N 个原始段落, 并尝试使用同义词词库替换原始段落, 形成替换段落; 之后运用 Sentence-Bert 技术校验同义词词库替换前后对应文本和对应段落之间的相似度。当满足所有相似度均不小于阈值条件时, 判定同义词词库替换后与上下文搭配成功; 否则, 当存在相似度小于阈值的情况时, 则视为同义词词库替换后与上下文搭配失败, 需重新尝试。

上下文搭配机制算法描述如下:

输入: ① 原始文本, ② 同义词词库。

输出: 操作状态。

Begin

(1) 将原始文本分成 N 个段落, 即原始段落 1~ N 。

(2) for 同义词词库

替换原始段落 1—N, 形成替换段落 1—N。

(3) for 替换段落 1—N

① 采用 Sentence-Bert 词嵌入技术将原始段落和对应替换段落表示为词向量;

② if 原始段落和对应替换段落词向量的余弦相似度 < 阈值

then 上下文搭配失败, 操作终止;

else 相似度标记加 1, 继续。

(4) 合并替换段落 1—N, 形成与原始文本对应的替换文本。

(5) 采用 Sentence-Bert 词嵌入技术将原始文本和对应替换文本表示为词向量。

(6) if 原始文本和对应替换文本词向量的余弦相似度 < 阈值

then 上下文搭配失败, 操作终止;

else 相似度标记加 1, 继续。

(7) if 相似度标记加 == N+1

提示上下文搭配成功。

End

3 实验与结果分析

3.1 实验数据及实验描述

3.1.1 实验数据及设置

实验数据由嵌入文本、同义词词库和异体字字库 3 部分组成, 其中, 嵌入文本选取“上海科学技术出版社出版的大数据技术与应用系列丛书《教育大数据》, 第七章教育大数据与校园管理应用”的全部内容。同义词词库是在哈工大信息检索研究室同义词词林扩展版的基础上, 增量采集了抖音、知乎、微博、小红书等网络平台同义词, 外加中华人民共和国教育部国家语言文字工作委员会整理与发布的第一批和第二批异形词整理表。异体字字库的数据源于国学大师平台 (<https://www.guoxuedashi.net/zidian/bujian/bjjc.php>), 本实验从该平台整理了与简体中文笔划接近的异体字。此外, 本实验 Sentence-Bert 词嵌入技术的预训练模型选择 distiluse-base-multilingual-cased-v2, 该模型针对中文语义相似度任务, 运用词嵌入技术生成词向量, 以余弦相似度的计算结果表示语义相似度。为保证计算结果的可信度, 将语义相似度的阈值设为 0.95, 认为低于该值的情况会产生语义歧义。

3.1.2 实验描述

(1) 数据预处理

为保证数据的完整性和一致性, 需对嵌入文本进行预处理, 共 2 步。第 1 步, 将嵌入文本中的异体字对齐为原始字; 第 2 步, 将嵌入文本中的同义词对齐为原始词。若对齐前后嵌入文本的语义歧义较小, 则对齐成功。

(2) 实验步骤

为验证本文水印算法的语义歧义、嵌入容量和安全性, 本文采用实验研究法设计了 3 组实验。第 1 组实验采用 2.1.1 节水印算法的嵌入流程, 输入对称密钥、隐匿信息和嵌入文本, 输出水印文本, 通过呈现水印文本的嵌入效果, 并结合基于 Sentence-Bert 的上下文搭配机制校验替换文本与原始文本的语义相似度, 验证算法不产生语义歧义。第 2 组实验采用 2.1.2 节水印算法的提取流程, 通过对比同义词扩展解码的提取结果与仅采用同义词解码的提取结果, 验证算法的嵌入容量得到提升。第 3 组实验由 3 部分组成, 用于验证算法的安全性。第 1 部分, 通过上下文搭配机制校验水印文本是否产生语义歧义, 以此判别同义词是否被篡改; 第 2 部分, 通过混沌映射机制生成的输出结果有高达 65 535! 种组合, 使得枚举攻击时间成本较高, 破解难度较大; 第 3 部分, 使用专业的网站和工具对比“传统 AES 与混沌 AES”及“传统 MD5 与混沌 MD5”加密算法的破解结果, 以此来验证其抗破解的能力。

3.2 语义歧义

3.2.1 水印文本嵌入效果展现抗语义歧义的能力

第 1 组实验, 对称密钥设为“I_love_python”, 隐匿信息设为“©H 大学(2024)计算机学院。保留所有权利。未经本学院书面许可, 任何单位或个人不得以任何方式使用、传播或复制本学院的研究成果。”。为便于人工分析, 选取一段水印文本的嵌入效果, 黑体字词为同义词扩展内容, 下划线字词为水印嵌入前后文本的差异内容, 用于对比语义歧义, 如图 5 所示。其中, 同义词有“可以-能够”“依据-基于”和“比如-例如”等; 异体字有“并-并”“内-内”和“出-出”等。本实验筛选得到的同义词表达含义较为接近, 不易引起歧义; 而异体字之间的笔画差异是细微的, 如果不单独放在一起通常不会引起注意。为验证算法嵌入前后文

>>嵌入前:

1、除此之外,郭水泉认为还可以对学习者的异常行为进行分析,并将异常行为总结为失联学生、晚归学生、舆情异常、学业异常、身心异常等几个方面。

2、失联学生。根据学生在校内的行为数据,与一卡通消费、门禁、上网、借阅等信息,完成与学生行为有关的数据建模。比如,几天无任何刷卡记录,可以认为该学生不在学校,由此判断学生非请假离校,产生失联异常信息,自定义设置预警阈值,并根据规则发出预警信息。

3、晚归与未归分析。门禁通道管理系统是控制和管理人员进出,并准确记录和统计住校学生数据的数字化出入控制系统。门禁通道管理系统与校园的身份识别认证相结合,方便内部人员出入,同时,杜绝外来人员随意进出,通过门禁数据形成当天晚归或未归学生信息,并将信息通过微信端及时推送给相应学生管理人员,及时了解学生异常行为。

>>嵌入后:

1、除此之外,郭水泉以为还能够对学习者的异常行为开展分析,并将异常行为总结为失联学生、晚归学生、舆情异常、学业异常、身心异常等多个方面。

2、失联学生。基于学生在校内的行为数据,与一卡通消费、门禁、上网、借阅等信息,完成与学生行为相关的数据建模。例如,几天无任何刷卡记录,能够认为该学生不在学校,由此判断学生非请假离校,生成失联异常信息,自定义设置预警阈值,并根据规则发出预警信息。

3、晚归与未归分析。门禁通道管理系统是控制和管理人员进出,并正确记录和统计住校学生数据的数字化出入控制系统。门禁通道管理系统与校园的身份识别认证相结合,方便内部人员出入,同时,杜绝外来人员随便进出,借助门禁数据建立当日晚归或未归学生信息,并将信息通过微信端实时推送给对应学生管理人员,及时洞悉学生异常行为。

图 5 水印嵌入前后文本语义歧义效果对比

Fig. 5 Comparison of text semantic ambiguity effects before and after watermark embedding

本是否产生语义歧义,本实验邀请了 20 位自然语言处理和隐私保护研究方向的研究者对水印文本进行评阅,经研究者们的仔细鉴定,一致认为水印文本在理解上与嵌入前保持较高的一致性。由此表明,该算法在保证文本语义不产生歧义的同时,有效实现了水印文本嵌入。

3.2.2 上下文搭配机制具有抗语义歧义的能力

Sentence-Bert 词嵌入技术将文本的语义相似度表示为词向量的余弦相似度,词向量的余弦相似度高,则文本的语义相似度高。同义词不同于异体字,替换后可能产生语义歧义,因此本文采用基于 Sentence-Bert 词嵌入技术的上下文搭配机制来验证替换后的语义歧义。依据该机制,将图 6 同义词替换前的原始文本和替换后的替换文本分别标记为 3 个段落,其中下划线字词为同义词替换前后的差异内容,共 18 处。通过同义词替换生成 3 组“原始段落与替换段落”和 1 组“原始文本与替换文本”,经反复尝试,使这 4 组数据的语义相似度均超过阈值。由图 7 可知,“原始文本与替换文本”和“原始段落与替换段落”之间的语义相似度均大于 0.98,证明同义词替换后的文本和

>>同义词替换前:

1:除此之外,郭水泉认为还可以对学习者的异常行为进行分析,并将异常行为总结为失联学生、晚归学生、舆情异常、学业异常、身心异常等几个方面。

2、失联学生。根据学生在校内的行为数据,与一卡通消费、门禁、上网、借阅等信息,完成与学生行为有关的数据建模。比如,几天无任何刷卡记录,可以认为该学生不在学校,由此判断学生非请假离校,产生失联异常信息,自定义设置预警阈值,并根据规则发出预警信息。

3、晚归与未归分析。门禁通道管理系统是控制和管理人员进出,并准确记录和统计住校学生数据的数字化出入控制系。门禁通道管理系统与校园的身份识别认证相结合,方便内部人员出入,同时,杜绝外来人员随意进出,通过门禁数据形成当天晚归或未归学生信息,并将信息通过微信端及时推送给相应学生管理人员,及时了解学生异常行为。

>>同义词替换后:

1、除此之外,郭水泉以为还能够对学习者的异常行为开展分析,并将异常行为总结为失联学生、晚归学生、舆情异常、学业异常、身心异常等多个方面。

2、失联学生。基于学生在校内的行为数据,与一卡通消费、门禁、上网、借阅等信息,完成与学生行为相关的数据建模。例如,几天无任何刷卡记录,能够认为该学生不在学校,由此判断学生非请假离校,生成失联异常信息,自定义设置预警阈值,并根据规则发出预警信息。

3、晚归与未归分析。门禁通道管理系统是控制和管理人员统。进出,并正确记录和统计住校学生数据的数字化出入控制系统。门禁通道管理系统与校园的身份识别认证相结合,方便内部人员出入,同时,杜绝外来人员随便进出,借助门禁数据建立当日晚归或未归学生信息,并将信息通过微信端实时推送给对应学生管理人员,及时洞悉学生异常行为。

图 6 同义词替换前后文本语义歧义效果对比

Fig. 6 Comparison of text semantic ambiguity effects before and after synonym replacement

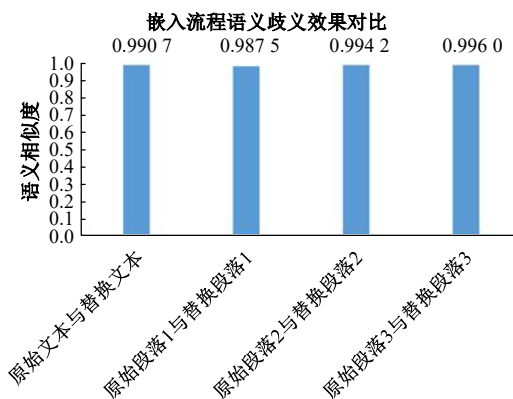


图 7 同义词替换前后文本的语义相似度

Fig. 7 Semantic similarity of text before and after synonym replacement

段落具有抗语义歧义的能力。

3.2.3 同类算法对比分析

文献[3]通过运用早期哈工大信息检索研究室开发的语言技术平台中的语法分析技术,设计了上下文搭配度算法,有效解决了同义词替换带来的语义歧义问题;文献[6]采用无载体零水印的嵌入方法,将编码生成的水印信息传入第三方注册机构。而本文采用的 Sentence-Bert 词嵌入技术属于语义相似度目前研究的热门技术之一,抗

语义歧义的能力会优于文献[3]所采用的早期技术 Word2vec。同时由于上下文搭配机制保证了嵌入前后语义较高的一致性,本文算法嵌入效果会接近文献[6]。因此在抗语义歧义的能力上,文献[6]≈本文>文献[3]。

3.3 嵌入容量

3.3.1 同义词扩展提升嵌入容量的能力

第2组实验采用3.2.1节的一段水印文本,同义词扩展包括同义词和异体字,1表示采用同义词扩展替换方法,0表示未采用同义词扩展替换方法,其提取效果如表1所示。其中同义词提取结果有33组,15组未采用替换方法,如“总结-总结”;而另外18组采用了替换方法,如“可以-能够”。异体字提取结果有20组,10组未采用替换方法,如“外-外”,而另外10组采用了替换方法,如“内-内”。提取结果在表1中的显示先左后右、先上后下。解码同义词得到提取结果111011011010000110010000111101101;同理,解码异体字得到提取结果00011111010001001110。相比仅使用同义词嵌入,异体字的嵌入使嵌入容量提升了 $20/33>60\%$,效果显著,证明该算法在嵌入容量方面表现较好。

3.3.2 同类算法对比分析

文献[3]仅使用同义词替换的编码嵌入方式,嵌入容量明显低于本文同时使用同义词和异体字替换的编码嵌入方式。文献[6]采用主题词编码和分句信息熵编码的嵌入方式,嵌入容量受限于

表1 同义词扩展解码的提取结果

Table 1 Extraction results of synonym expansion decoding

同义词 提取结果	异体字 提取结果
认为-以为:1,可以-能够:1,进行-开展:1	外-外:0,并-并:0
总结-总结:0,几个-多个:1,根据-基于:1	晚-晚:0,业-业:1
完成-完成:0,有关-相关:1,比如-例如:1	内-内:1,任-任:1
几天-几天:0,可以-能够:1,认为-认为:0	值-值:1,并-并:1
不在-不在:0,由此-由此:0,判断-判断:0	出-出:0,晚-晚:1
产生-生成:1,阈值-阈值:1,根据-根据:0	出-出:0,并-并:0
规则-规则:0,准确-正确:1,住校-住校:0	出-出:0,别-别:1
识别-识别:0,结合-结合:0,杜绝-杜绝:0	内-内:0,出一出:0
随意-随意:1,通过-借助:1,形成-建立:1	外-外:1,出一出:1
当天-当日:1,通过-通过:0,及时-实时:1	晚-晚:1,并-并:0
相应-对应:1,及时-及时:0,了解-洞悉:1	

主题词的数量和分句信息熵增加的额外存储开销,致使其表现较差。因此在提升嵌入容量的能力上,本文>文献[3]>文献[6]。

3.4 安全性

3.4.1 同义词篡改判别能力

(1) 本文算法

第3组实验同样采用3.2.1节的一段水印文本,判别前需将水印文本中的同义词对齐为原始词,形成原始词对齐文本。原始词对齐文本和段落分别简称“原始文本”和“原始段落”,同义词篡改后的文本和段落分别简称“篡改文本”和“篡改段落”。由图8可知,尽管下划线字词“不仅如此”为“除此之外”的同义词篡改内容,但其产生的语义歧义不易人为辨别,需使用上下文搭配机制校验“原始文本与篡改文本”或“原始段落与篡改段落”的语义相似度是否有小于阈值的情况,以此判别水印文本的同义词是否被篡改。由图9可知,“原始段落1与篡改段落1”的语义相似度为

>>原始词对齐:

1、除此之外,郭水泉认为还可以对学习者的异常行为进行分析,并将异常行为总结为失联学生、晚归学生、舆情异常、学业异常、身心异常等几个方面。
2、失联学生。根据学生在校内的行为数据,与一卡通消费、门禁、上网、借阅等信息,完成与学生行为有关的数据建模,比如,几天无任何刷卡记录,可以认为该学生不在学校,由此判断学生非请假离校,产生失联异常信息,自定义设置预警阈值,并根据规则发出预警信息。
3、晚归与未归分析。门禁通道管理系统是控制和管理人员进出,并准确记录和统计住校学生数据的数字化出入控制系统。门禁通道管理系统与校园的身份识别认证相结合,方便内部人员出入,同时,杜绝外来人员随意进出,通过门禁数据形成当天晚归或未归学生信息,并将信息通过微信端及时推送给相应学生管理人员,及时了解学生异常行为。

>>同义词篡改后

1、不仅如此,郭水泉以为还能够对学习者的异常行为开展分析,并将异常行为总结为失联学生、晚归学生、舆情异常、学业异常、身心异常等多个方面。
2、失联学生。基于学生在校内的行为数据,与一卡通消费、门禁、上网、借阅等信息,完成与学生行为相关的数据建模,例如,几天无任何刷卡记录,能够认为该学生不在学校,由此判断学生非请假离校,生成失联异常信息,自定义设置预警阈值,并根据规则发出预警信息。
3、晚归与未归分析。门禁通道管理系统是控制和管理人员进出,并正确记录和统计住校学生数据的数字化出入控制系统。门禁通道管理系统与校园的身份识别认证相结合,方便内部人员出入,同时,杜绝外来人员随意进出,借助门禁数据建立当日晚归或未归学生信息,并将信息通过微信端实时推送给对应学生管理人员,及时洞悉学生异常行为。

图8 同义词篡改文本与原始词对齐文本语义歧义效果对比

Fig. 8 Comparison of semantic ambiguity effects between synonym tampering text and original word alignment text

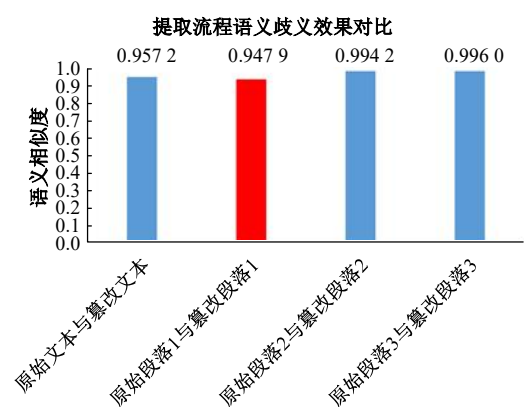


图 9 同义词篡改文本与原始词对齐文本的语义相似度
Fig. 9 Semantic similarity between synonym tampering text and original word alignment text

0.947 9<0.95, 因此判定此段落有同义词被篡改, 验证了同义词篡改的判别能力是有效的。

(2)同类算法对比分析

文献[6]从第三方注册机构获取水印存档信息, 采用词频统计方法计算主题词的相似度, 采用最小误差平方和方法计算分句信息熵的相似度。通过对主题相似度和分句信息熵相似度加权求得总相似度, 若总相似度低于阈值, 则认为与原文相

似度不高, 但该方法局限于词频和语句的统计分析, 未充分考虑上下文的语境关系, 若将主题词以外的同义词篡改, 会出现与原文有较高语义相似度的错误。相比而言, 本文算法对同义词篡改有较好的判别能力。鉴于文献[3]未设计该功能, 在同义词篡改的判别能力上, 本文>文献[6]>文献[3]。

3.4.2 混沌映射机制的抗破解能力

(1)本文算法

参数 r 选择1~1 000, 形成1 000组验证数据, 如表2所示, 采用混沌映射机制将整数(0~65 535)输出为另一组整数。以 $r=1$ 为例, 输入与输出的映射关系为“0→346”“1→1 056”“2→1 766”“3→2 476”...“65 534→60 348”和“65 535→61 058”。结果表明, 当 r 取值1~1 000时, 输入整数(0~65 535), 采用混沌映射机制输出的结果均不同, 这将使本文提出的混沌 MD5 和混沌 AES 输出的加密结果也均不相同。由于 r 可以取全体实数, 使输出结果最大有65 535!种有效组合, 采用混沌 AES 算法对隐匿信息进行加密保护, 其结

表 2 不同 r 取值下混沌映射机制输出结果对比

输入	输出
第1组: ① $r=1$ ② $X=\{0, 1, 2, 3, \cdots, 65\,534, 65\,535\}$	$Y=\{346, 1\,056, 1\,766, 2\,476, \cdots, 60\,348, 61\,058\}$
第2组: ① $r=2$ ② $X=\{0, 1, 2, 3, \cdots, 65\,534, 65\,535\}$	$Y=\{173, 528, 883, 1\,238, \cdots, 30\,174, 30\,529\}$
.....	
第100组: ① $r=100$ ② $X=\{0, 1, 2, 3, \cdots, 65\,534, 65\,535\}$	$Y=\{60\,362, 29\,086, 8\,259, 39\,535, \cdots, 46\,474, 15\,198\}$
第101组: ① $r=101$ ② $X=\{0, 1, 2, 3, \cdots, 65\,534, 65\,535\}$	$Y=\{46\,351, 3\,015, 14\,348, 57\,684, \cdots, 39\,822, 7\,819\}$
.....	
第500组: ① $r=500$ ② $X=\{0, 1, 2, 3, \cdots, 65\,534, 65\,535\}$	$Y=\{22\,493, 60\,010, 45\,424, 7\,907, \cdots, 53\,067, 30\,136\}$
第501组: ① $r=501$ ② $X=\{0, 1, 2, 3, \cdots, 65\,534, 65\,535\}$	$Y=\{27\,026, 56\,562, 14\,333, 39\,719, \cdots, 50\,048, 7\,713\}$
.....	
第999组: ① $r=999$ ② $X=\{0, 1, 2, 3, \cdots, 65\,534, 65\,535\}$	$Y=\{18\,978, 39\,634, 60\,290, 50\,793, \cdots, 37\,826, 17\,170\}$
第1 000组: ① $r=1\,000$ ② $X=\{0, 1, 2, 3, \cdots, 65\,534, 65\,535\}$	$Y=\{60\,229, 30\,005, 22\,712, 52\,936, \cdots, 45\,292, 15\,068\}$

果将难以被破解,安全性得到保障。

(2)同类算法对比分析

文献[3]采用 Logistic 映射将输出在 $[0, 0.5]$ 区间的数据映射为 0,而将 $[0.5, 1]$ 区间的数据映射为 1,通过该方法使预测误差扩展产生的溢出信息映射到由 0 和 1 组成的二进制混沌序列中,但由于 Logistic 映射算法是公开的,导致其仍存在被破解的风险。相比而言,本文混沌映射机制设定的参数 r 使输出结果的有效组合高达 $65\,535!$ 种,破解难度较大。鉴于文献[6]未设计该功能,在混沌算法的抗破解能力上,本文 $>$ 文献[3] $>$ 文献[6]。

3.4.3 加密算法的抗破解能力

(1)本文增强算法

本文采用 CMD5 网站与 MD5Crack 工具攻击 1 000 组传统 MD5 加密密钥和混沌 MD5 增强密钥,其中混沌映射机制参数 r 的 1 000 个取值随机生成。由表 3 可知,使用 CMD5 网站攻击传统 MD5 的加密密钥时,成功破解了 613 次;而攻击混沌 MD5 的增强密钥时,全部破解失败。使用 MD5Crack 工具攻击传统 MD5 的加密密钥时,成功破解次数高达 968 次;但攻击混沌 MD5 的增强密钥时,同样全部破解失败。这表明混沌 MD5 算法的抗破解能力是有效的。

在第(3)节实验的基础上,传统 AES 和本文提出的混沌 AES 采用相同的混沌 MD5 增强密钥加密,分别生成了 1 000 组加密结果。表 4 使用 BurpCrypto 工具尝试破解 1 000 组传统 AES 和混沌 AES 的加密结果。结果显示,对传统 AES 加

表 3 混沌 MD5 与传统 MD5 抗破解能力对比

Table 3 Comparison of anti-cracking ability between chaotic MD5 and traditional MD5

攻击主体	攻击对象	攻击次数	破解次数
CMD5 网站	传统 MD5 加密密钥	1 000	613
	混沌 MD5 增强密钥	1 000	0
MD5Crack 工具	传统 MD5 加密密钥	1 000	968
	混沌 MD5 增强密钥	1 000	0

表 4 混沌 AES 与传统 AES 抗破解能力对比

Table 4 Comparison of anti-cracking ability between chaotic AES and traditional AES

攻击主体	攻击对象	攻击次数	破解次数
BurpCrypto 工具	传统 AES 加密结果	1 000	991
	混沌 AES 加密结果	1 000	0

密结果的破解成功了 991 次,而对混沌 AES 加密结果的破解全部失败。这表明混沌 AES 算法的抗破解能力是有效的。

(2)同类算法对比分析

本文采用混沌 MD5 和混沌 AES 实现对密钥和算法的双增强,有效改善了传统算法抗破解能力不足的问题,使本文增强算法对隐匿信息的加密保护能力优于文献[6]直接采用传统 RSA 的加密方法。鉴于文献[3]未设计该功能,在加密算法的抗破解能力上,本文 $>$ 文献[6] $>$ 文献[3]。

3.5 时间复杂度对比分析

算法的时间复杂度主要由遍历次数和向量维度决定,由于同义词替换的嵌入过程差异不大,本节将分别阐述本文算法、文献[3]方法和文献[6]方法在时间复杂度的相同部分和差异部分。相同部分为通用流程,假设嵌入文本总遍历次数为 L ,同义词有效嵌入遍历次数为 V ,同义词词库的平均访问次数为 M ,则该部分时间复杂度为 $O(VM+L-V)$,简称为 $O(N_1)$ 。差异部分能体现不同算法时间复杂度的差异,相比文献[3]方法,本文算法的同义词词库规模略大,词向量的维度更高,假设词向量维度为 D ,替换同义词数量为 P ,平均重试次数为 Q ,则本文算法差异部分的时间复杂度为 $O(P_1Q_1D_1)$,仅略高于文献[3]方法(代表领域平均水平)的时间复杂度 $O(P_2Q_2D_2)$ 的正常水平。文献[6]采用无载体零水印的方法,将水印信息存于第三方,这将使遍历范围由局部转为全局,假设全局遍历次数为 X ,在局部遍历的基础上,可将时间复杂度增加为 $O(XP_3Q_3D_3)$,因而文献[6]算法的时间复杂度远高于文献[3]方法和本文算法。因此,在时间复杂度上,文献[6] $>$ 本文 $>$ 文献[3]。

3.6 同类算法综合对比

按照 3.2~3.5 节的分析结果,由综合对比可知,本文算法相比文献[6]在抗语义歧义上没有优势,但在嵌入容量上,文献[6]能力较弱;在同义词篡改判别上,文献[3]未设计该功能;在隐匿信息的保护上,文献[3]和文献[6]没有同时结合混沌算法和加密算法,其抗破解能力不足。反观本文算法在嵌入容量、同义词篡改判别、混沌算法抗破解和加密算法抗破解能力的表现上,均优于文献[3]和文献[6],如表 5 所示。由此证明,本文算法

表 5 算法能力综合对比

Table 5 Comprehensive comparison of algorithm capabilities

算法性能	文献[3]	文献[6]	本文
抗语义歧义	中	高	较高
嵌入容量	中	较低	高
同义词篡改判别	不支持	较高	高
混沌算法抗破解	中	不支持	较高
加密算法抗破解	不支持	较高	高
时间复杂度	中	高	略高于中等

能同时兼顾语义歧义、嵌入容量和安全性。

4 结束语

当前文本水印算法在语义歧义、嵌入容量与安全性三者间难以兼顾,为实现这三者的均衡提升,本文设计了一种基于混沌 AES 和同义词扩展的改进中文文本水印算法,通过设计基于 Sentence-Bert 的上下文搭配机制,不仅能确保同义词替换后的上下文与原文语义高度相似,而且能判别同义词篡改后产生的语义歧义;将异体字作为同义词扩展,在不引起语义歧义的前提下,实现更多同义词替换,从而提升嵌入容量;利用混沌映射机制,升级传统 MD5 算法,构建混沌 AES 算法,实现对隐匿信息的加密保护,进而增加其破解难度。实验结果从呈现水印文本的嵌入效果和上下文搭配机制验证了算法具有抗语义歧义的能力;通过对比同义词扩展解码的提取结果与仅采用同义词解码的提取结果,验证了算法的嵌入容量得到提升;通过上下文搭配机制校验水印文本存在语义相似度低于阈值的情况,以此判别同义词被篡改;通过混沌映射机制生成的输出结果有高达 65 535!种组合,使得枚举攻击时间成本较高,破解难度较大;使用专业的网站和工具对比分析“传统 AES 与混沌 AES”及“传统 MD5 与混沌 MD5”加密算法的破解结果验证了其抗破解的能力。在时间复杂度趋于正常水平的前提下,相比文献[3]和文献[6]的方法,本文算法在语义歧义、嵌入容量和安全性上均有较好的表现,再次体现本文算法的优势。

然而,当对水印文本实施同义词或异体字的增删攻击时,本文算法尚无法实现无损修复,甚至可能会降低水印提取结果的准确性,这将是后续

研究亟待解决的问题。

参考文献:

[1] 朱强强. 基于汉字笔画微变形的文本隐写方法研究[D]. 杭州: 杭州电子科技大学网络与信息安全学院, 2024.

Zhu Qiang-qiang. Research on text steganography based on chinese character stroke fine-tuning[D]. Hangzhou: College of Network and Information Security, Hangzhou Dianzi University, 2024.

[2] 黄瑶, 潘丽丽, 熊思宇, 等. 基于生成对抗网络与多头注意力的文本隐写术[J]. 计算机工程与科学, 2023, 45(10): 1789-1796.

Huang Yao, Pan Li-li, Xiong Si-yu, et al. Text steganography based on generative adversarial networks and multi-head attention[J]. Journal of Computer Engineering & Science, 2023, 45(10): 1789-1796.

[3] 费文斌, 唐向宏, 王静, 等. 基于预测误差扩展的可逆文本水印算法[J]. 中文信息学报, 2015, 29(1): 133-138.

Fei Wen-bin, Tang Xiang-hong, Wang Jing, et al. Reversible text watermarking algorithm based on prediction error expansion[J]. Journal of Chinese Information Processing, 2015, 29(1): 133-138.

[4] Xiang L Y, Wu W S, Li X, et al. A linguistic steganography based on word indexing compression and candidate selection[J]. Multimedia Tools and Applications, 2018, 77(21): 28969-28989.

[5] 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.

Lao Ye, Liu Shu-hui, Wang Hui, et al. Robust chinese text watermarking method based on chinese character glyph perturbation and font replacing[J]. Journal of Cryptologic Research, 2023, 10(4): 769-785.

[6] 张娜, 张琨, 张先国, 等. 基于主题词与信息熵编码的文本零水印算法[J]. 计算机与数字工程, 2021, 49(8): 1612-1618.

Zhang Na, Zhang Kun, Zhang Xian-guo, et al. Text zero-watermarking algorithm based on keywords and information entropy encoding[J]. Journal of Computer & Digital Engineering, 2021, 49(8): 1612-1618.

[7] Zheng X Y, Wu H Z. Autoregressive linguistic steganography based on BERT and consistency coding[J]. Security and Communication Networks, 2022:

- 1-11.
- [8] Yu L, Lu Y L, Yan X H, et al. MTS-Stega: linguistic steganography based on multi-time-step[J]. *Entropy*, 2022, 24(5): 585.
- [9] 金家立. 基于自动选择编码方式的文本信息隐藏技术研究[D]. 沈阳: 沈阳工业大学信息科学与工程学院, 2023.
- Jin Jia-li. Research on text information hiding technology based on automatic selection encoding[D]. Shenyang: School of Information Science and Engineering, Shenyang University of Technology, 2023.
- [10] Reimers N, Gurevych I. Sentence-bert: sentence embeddings using siamese bert-networks[J]. *Arxiv Preprint*, 2019, 8: 190810084.
- [11] Wang X Y, Feng D G, Lai X J, et al. Collisions for hash functions MD4, MD5, HAVAL-128 and RIPEMD[J/OL]. *IACR Cryptology eprint Archive*, [2004-08-17]. <https://eprint.iacr.org/2004/199.pdf>
- [12] Klima V. Tunnels in hash functions: MD5 collisions within a minute[J/OL]. *IACR Cryptol. ePrint Arch*, [2006-04-17]. <https://eprint.iacr.org/2006/105.pdf>
- [13] Stevens M. Attacks on hash functions and applications[D]. Leiden: Mathematical institute faculty, science Leiden University, 2012.
- [14] Rønjom S, Bardeh N G, Hellesest T. Yoyo tricks with AES[C] // *Advances in Cryptology-ASIACRYPT 2017: 23rd International Conference on the Theory and Applications of Cryptology and Information Security*, Hong Kong, China, 2017: 217-243.
- [15] Bardeh N G, Rønjom S. The exchange attack: how to distinguish six rounds of AES with chosen plaintexts[C] // *International Conference on the Theory and Application of Cryptology and Information Security*, Kobe, Japan, 2019: 347-370.
- [16] Dunkelman O, Keller N, Ronen E, et al. The retracing boomerang attack[C] // *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Zagreb, Croatia, 2020: 280-309.
- [17] Bardeh N G, Rijmen V. New key recovery attack on reduced-round AES[J]. *Cryptology ePrint Archive*, 2022(2): 43-62.