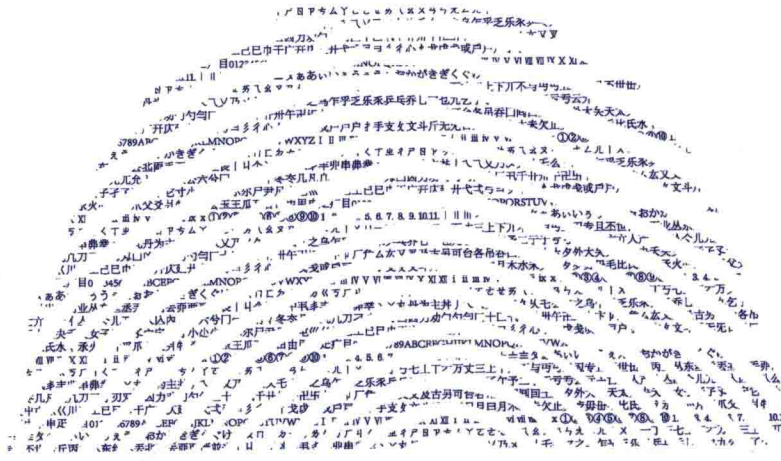

Theory and Method of Text Digital Watermarking

文本数字水印

理论与方法

周新民 著

西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

湖南省移动电子商务协同创新中心
移动商务智能湖南省重点实验室
湖南商学院大数据与互联网创新研

文本数字水印
理论与方法

Theory and Method of Text Digital Watermarking

周新民 著

西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

图书在版编目（CIP）数据

文本数字水印理论与方法 / 周新民著 . ——西安 : 西安交通大学出版社，2016.8

ISBN 978-7-5605-8995-4

Ⅰ. ①文… Ⅱ. ①周… Ⅲ. ①电子计算机－密码术－研究 Ⅳ. ①TP309.7

中国版本图书馆 CIP 数据核字（2016）第 214567 号

书　　名　文本数字水印理论与方法
著　　者　周新民
责任编辑　魏　杰　贺彦峰

出版发行　西安交通大学出版社
（西安市兴庆南路 10 号　邮政编码 710049）
网　　址　http://www.xjtupress.com
电　　话　（029）82668357　82667874（发行中心）
（029）82668315（总编办）
传　　真　（029）82668280
印　　刷　虎彩印艺股份有限公司

开　　本　880mm×1230mm　1/32　**印张** 5.75　**字数** 114 千字
版次印次　2016 年 8 月第 1 版　2016 年 8 月第 1 次印刷
书　　号　ISBN 978-7-5605-8995-4/TP · 623
定　　价　58.00 元

前　言

目前，随着网络的普及和多媒体技术的快速发展，数字作品的信息安全问题已成为一个研究的热点。 数字水印技术作为信息隐藏的一个重要分支，是一种保障数字作品信息安全的重要手段。 它能克服密码技术中加密的内容在解密之后就没有有效的手段来保证其不被非法拷贝、再次传播、恶意篡改的问题，同时也弥补了数字签名不能在原始数据中一次性嵌入大量信息的弱点，克服了数字标签容易被修改和剔除的缺陷，突破了数字指纹仅能给出有关出版权破坏者信息的局限。 因此，数字水印技术成为当前多媒体信息安全研究领域发展较快的热门技术，已经受到国际学术界的高度关注，是解决知识产权保护和信息安全问题极具潜力的多学科交叉技术。

本书应用汉字数学表达式理论，对文本水印算法展开了一系列的研究。 由于人们很多重要的思想以及其他有价值的信息都是以文本的方式存储并在网上传输，如果不对文本采取任何

保护措施，文本的内容很容易被非法编辑和修改。本书提出了一种基于汉字结构知识的鲁棒性公开文本水印算法，为文本的版权保护提供了有效的解决方案。在文本认证方面，本书旨在通过采用文本水印技术，主要从嵌入策略、密钥管理和水印结构三个方面，分别提出了三种不同的基于文本认证的数字水印算法，以解决网络传输过程中文本的认证问题。该研究对电子商务、电子政务和国家安全等多个领域具有重要的学术意义和实用价值。本书取得的主要成果如下：

第一，从版权保护的角度，提出了一种基于汉字结构知识的鲁棒性公开文本水印算法。该方法通过汉字数学表达式，获取汉字的结构类型和笔画数，利用汉字的结构类型将整个文档分成两块，在各块中由汉字笔画数和水印比特位共同确定水印加载的位置，通过混沌加密和海明校验编码使水印生成一种具有鲁棒性的水印结构，然后通过设置字体下划线将水印按对称性的嵌入策略分别嵌入两个文本块中。抗攻击时，能准确判断水印比特位受破坏的具体位置，并能通过块校验和海明校验将破坏的水印比特位进行适当恢复；同时，提取水印时既不需要原始文本，也不需要原始水印，且提取的水印信息是有意义的版权信息，是一种公开水印。

第二，从嵌入策略的角度，提出了一种基于文本内容认证的零水印方案。这种方案通过将认证信息嵌入在中文文档被构造的二进制模式中，而不是在载体文本本身之中，因此可以有效地解决由于水印嵌入导致的文本可感知质量的下降和水印不

可见性与鲁棒性之间固有的冲突。 由于水印嵌入并不需要对宿主文档做任何修改，因此该方案同时适应于长短文本文档。 同时，因为该方案设计的目的是对文本文档进行内容认证，所以认证过程不需要第三方参与。 该方案采用了混沌加密算法，由于混沌系统初始条件的敏感性，即使初始值存在 10^{-6} 的差别，也将导致完全不同的混沌序列值。 所以，如果攻击者不知道有关混沌加密算法密钥的任何消息，是不可能伪造一个能够通过认证的文档的。

第三，从密钥管理的角度，提出了一种基于中文文本内容认证的半脆弱文本水印方案。 该方案通过汉字结构知识，将载体文档分成不同的两个文本块，然后通过计算不同文本块中汉字的频率和笔画数获取两个不同的密钥，并使用这两个不同的密钥对载体文档的 Hash 函数值进行混沌加密， 生成两个不同的水印，分别嵌入到两个文本块中。 该方案密钥生成方式独特，密钥的生成与文本的内容密切相关，为文本的内容认证提供了安全可靠的保证。

第四，从水印结构的角度，提出了一种基于多重认证的文本水印算法。 该方法在两个文本块中实现了多个水印的一次性嵌入，成功地解决了单个水印难以满足的文本多重认证问题。对文本内容的消息认证码、标题水印和作者水印进行了混沌加密，加强了水印的安全控制能力；对有意义的标题和作者信息进行了汉明校验编码，有效增强了水印差错控制的能力。 该方案通过对文本的内容进行认证，可以判断文本的真实性，即文

本是否被篡改；通过对文本的标题与作者信息进行认证，可以判断文本的来源是否可靠。 在嵌入水印的设计上，采用了有意义的水印信息，使得认证过程更加直观化、具体化。 同时，该方案成功实现了篡改的局部定位和一定程度上的篡改恢复。

本书是教育部人文社会科学研究规划基金项目（课题编号：12YJAZH216）的研究成果。 由于作者水平有限，书中难免出现各种疏漏和不当之处，欢迎大家批评指正。

作者

2016 年 6 月于长沙

目　录

第1章　绪　论 …… (1)

1.1　研究背景 …… (1)

1.2　研究意义 …… (4)

1.3　文本水印研究综述 …… (7)

1.4　主要贡献 …… (21)

1.5　组织结构 …… (22)

第2章　汉字数学表达式理论 …… (25)

2.1　汉字数学表达式基本知识 …… (25)

2.2　汉字结构知识的获取 …… (30)

2.3　小　结 …… (33)

第3章　文本数字水印安全性理论 …… (34)

3.1　引　言 …… (34)

3.2　文本水印安全性研究 …… (36)

3.3 文本水印安全性相关问题 …… (44)
3.4 安全性对策 …… (51)
3.5 小 结 …… (52)

第4章 数字版权保护水印方法与模型研究 …… (53)
4.1 数字版权保护的安全性问题 …… (53)
4.2 版权保护水印生成 …… (56)
4.3 版权保护水印嵌入 …… (58)
4.4 版权保护水印检测 …… (62)
4.5 基于数字版权保护的鲁棒性文本水印算法 …… (64)
4.6 网络教育资源数字版权保护模型建立 …… (71)
4.7 小 结 …… (77)

第5章 基于内容认证的文本零水印算法 …… (78)
5.1 引 言 …… (78)
5.2 文本零水印方案 …… (79)
5.3 实验结果与分析 …… (86)
5.4 小 结 …… (88)

第6章 基于内容认证的半脆弱文本水印算法 …… (90)
6.1 引 言 …… (90)
6.2 半脆弱文本水印方案 …… (92)
6.3 实验结果与分析 …… (97)
6.4 小 结 …… (102)

第7章 基于多重认证的文本水印算法 …… (103)
7.1 引 言 …… (103)
7.2 多重认证文本水印方案 …… (105)
7.3 实验结果与分析 …… (116)
7.4 小 结 …… (122)

第8章 文本水印攻击模型与性能评价 …… (124)
8.1 引 言 …… (124)
8.2 文本水印攻击模型 …… (126)
8.3 文本水印性能分析 …… (129)
8.4 文本水印性能评价 …… (134)
8.5 小 结 …… (145)

第9章 结论与展望 …… (146)
9.1 结 论 …… (146)
9.2 进一步工作的方向 …… (150)

参考文献 …… (153)

后 记 …… (173)

第1章 绪论

1.1 研究背景

信息安全是计算机科学与技术的一个重要学科领域。随着电子商务及电子政务的发展，党政机关、企事业单位、民间团体、国防、国家安全等部门将有大量的信息，特别是一些重要文件、秘密信息、军事机密、商业机密、电子合同等，通过网上传输。研究这些重要文件、秘密信息、军事机密、商业机密、电子合同等的安全存储、安全可信传输等安全保障技术已被列入《国家中长期科学和技术发展规划纲要》（2006—2020 年）的重点优先主题。目前，随着计算机技术与网络通信技术以及信息产业的高速发展，接入因特网的个人和单位主机数量快速增长，尤其是计算机在政府、国防、金融、公安和商业等部门的广泛应用，使得社会对计算机的依赖作用越来越大。由于多媒体资源的数字化，数字作品的盗版问题已经越来越严重，使得数字水印技术成为目前一个十分重要的研

究热点，并引起了信息安全领域、计算机技术领域、通信领域和信息处理领域等方面的专家和学者的广泛关注。

我国政府对数字水印等信息安全技术的研究非常重视。2000年1月15日至16日，国家863计划智能计算机专家组会同中科院自动化所模式识别国家重点实验室和北京邮电大学信息安全中心在北京西郊宾馆主持召开了“数字水印学术研讨会”。这次研讨会引起了国家自然科学基金委员会、信息产业部、国家信息安全测评认证中心、国家图书馆及中国社会科学院等多家单位的高度重视。召开数字水印研讨会在我国尚属首次，研讨会的成功召开，极大地坚定了科研工作者的信心，明确了工作方向。这次会议对于促进数字水印技术的前瞻性理论研究与实际应用具有十分重要的作用，对于加强我国科技人员进行信息安全方面的学术交流也有不可低估的作用。2001年，教育部首次批准在武汉大学开设信息安全本科专业，许多高校和科研院所相继设立了信息安全方向的具有硕士和博士学位授予权的学科点。2003年9月，中央《关于加强信息安全保障工作的意见》的27号文件，已经把信息安全工作提升到保护公众利益和维护国家安全以及保障与促进信息化发展的高度。2004年1月，国务院召开全国信息安全保障工作会议，再次特别强调加强信息安全工作的重要性。2007年12月6日至7日，首期信息安全国家标准宣传贯彻培训会议在福州举办。

21世纪是信息时代，在社会信息化的进程中，信息已成为社会发展的重要战略资源，信息安全在信息社会中扮演着极为重要的角色。据相关数据统计，我国2007年数字内容产业的市场规模达到了1 570.1亿元，增长率超过40%，显示出极大的发展潜力。接下来的3年中，数字内容产业继续保持高增长态势。数字内容产业是一个新兴的经济产业，已经创造出惊人的经济、社会价值，成为

21 世纪知识经济产业的核心。世界上一些国家和地区都已把数字内容产业作为新兴战略性产业，并加以关注和研究，积极推动该产业的发展。我国相关政府部门也给予了数字内容产业越来越高的战略定位。数字内容产业作为信息化社会的重要组成部分，已被我国《2006—2020 年国家信息化发展战略》确定为重要发展方向之一。

多媒体数据的数字化为多媒体信息的存取提供了极大的便利，同时也极大地提高了信息表达的效率和准确性。随着因特网的日益普及，多媒体信息的交流已达到了前所未有的深度和广度，其发布形式也愈加丰富了。人们如今可以通过因特网发布自己的作品、重要信息和进行网络贸易等，但是随之出现的问题也十分严重，如作品侵权更加容易，篡改也更加方便。各种侵权问题已成为困扰数字内容产业发展的一个重大问题。根据 2008 年美国政策创新研究所(IPI)发布的最新报告，仅仅数字产业中的盗版问题每年便给美国经济带来高达 580 亿美元的损失。报告作者史蒂芬·斯维克(Stephen E. Siwek)表示："随着政策制定者将注意力转向如何提升美国在全球市场的竞争力，盗版问题无疑应优先得到解决。"因此如何既充分利用因特网的便利，又有效地保护知识产权，已受到人们的高度重视。盗版者对数字多媒体产品的非法操作行为，通常包括以下三种情况：

(1)非法访问，即未经版权所有者的允许从网站中非法复制或翻印数字产品。

(2)故意篡改，即盗版者恶意修改数字产品以破坏或嵌入特征并进行重新发送，从而使原始产品的版权信息丢失。

(3)版权破坏，即盗版者收到数字产品后未经版权所有者的允许将其转卖。

为了保护国家的政治利益和经济利益，各国政府都十分重视信息和网络安全，信息安全已成为一个世纪性、全球性的研究课题。随着理论研究的进行和相关软件的不断推出，国际上一些大公司正在致力于以保护多媒体知识产权为目的的数字水印技术的标准和实用化研究。这标志着一门新兴的交叉学科——数字水印的正式诞生。数字水印技术的研究就是在这种应用要求下迅速发展起来的。如今数字水印技术作为认证、隐蔽通信和知识产权保护等的主要手段，正得到广泛的研究与应用。

本书正是基于目前国家和社会对文本信息安全的较高要求，在文本数字水印的安全性理论与文本认证算法方面开展了深入的研究，其相关研究成果可以为文本的信息安全保护实际应用问题提供必要的理论基础和技术支持。

1.2 研究意义

数字水印是一门充满活力的交叉学科，它涉及的研究领域包括通信、密码学、信息理论、多媒体处理、模式识别和人工智能等。“数字水印”这个术语在1988年第一次被Komatsu和Tominaga使用。随着数字网络通信的飞速发展，信息的发布和传输实现了“数字化”和“网络化”。自从1993年11月因特网上出现了Marc Andreessen的Mosaic网页浏览器，因特网对用户变得友好起来，人们很快便开始乐于从因特网上下载多媒体信息。对数字多媒体而言，因特网成了最出色的分发系统，因为它不但便宜，而且不需要仓库存储，又能实现实时发送。因此，数字多媒体很容易借助因特网被复制、处理、传播和公开。由此引发的数字多媒体信息版

权保护问题和传输的安全问题，已引起了国际学术界、企业界以及政府有关部门的广泛关注。 由于人们对多媒体内容版权保护意识的提高，1995 年前后，关于数字水印技术的研究开始猛增。 内容保护求助的第一种技术是加密。 加密是保护数字内容最常用的一种方法。 密码学是发展得最好的一门科学。 不幸的是，加密在数据传输过程中可以对数字内容进行保护，但一旦数据被解密，就不能提供进一步的保护措施。 因此，迫切需要一种新的信息安全技术，可以使其成为这些技术的一种替代或补充。 数字水印技术具有满足这一需要的潜能，因为它在宿主数据中嵌入的信息在常规操作的情况下很难被去除。

在多媒体水印技术中，文本数字水印理论与关键技术的研究尤其重要。 随着电子商务及电子政务的快速发展，党政机关、企事业单位的一些重要文件均以文本的形式借助因特网快速传输。 如果不对这些文本采取任何保护措施，有关的非法操作很可能会涉及国家安全、法庭举证等方面的问题。 因此，研究如何保证这些文本信息的安全问题是关乎个人、集体甚至国家发展与安危的大事。 文本数字水印技术通过对载体作品嵌入和提取特定的水印信息，为解决上述问题提供了一个很好的技术方案。 目前，国外对文本的研究主要以英文为主，而对中文文本数字水印的研究较少。 同时，由于中文文本数字水印的研究难度较大，国内对该方向的研究报道也相对较少，因此有很多理论与实际问题值得深入研究。

本书主要针对文本数字水印的安全性问题，在文本数字水印安全性理论研究的基础上，充分利用汉字自身的特点，对文本认证进行了深入的研究。 该研究对文本信息安全的发展和应用有着重要的意义，对解决目前因特网上比较脆弱的安全机制是一个有益的尝试。 其研究成果可广泛应用于电子商务、电子政务、国防、国家

安全、数字图书馆等文本安全保护领域。其理论意义和实际应用价值具体如下：

（1）文本数字水印安全性理论的研究，为文本水印系统的安全性提供理论支撑。文本数字水印的研究是一个涉及密码学、中文信息处理、计算机网络、视觉科学、通信、信息安全等的跨学科交叉研究，目前在理论上是一个很新的研究方向，国外已有研究方法比较单一，国内对该方向的研究报道不多，有很多理论问题值得深入研究。目前，与图像、视频、音频数字水印方法相比，文本数字水印所用的算法截然不同。当前文本数字水印的研究中存在的主要安全问题是抗攻击性不强、鲁棒性较差、隐蔽性不好等。而且文本数字水印理论本身十分复杂，人们还远未真正掌握文本水印安全性的实质，因此还需要专门对其安全理论进行广泛深入的研究，努力探求安全性较好的文本水印的嵌入新途径、新方法。本书将针对文本数字水印的安全性问题，研究文本数字水印模型及其相关理论，提供文本水印系统设计的基础理论，为文本数字水印的进一步研究和应用发展，提供一定的理论支撑。

（2）文本数字水印版权保护算法、认证算法的提出，将丰富和发展数字水印理论。本书充分考虑汉字的特点，从水印嵌入方式、嵌入策略和文本数字水印攻击方法与对策等方面研究高效率的、安全性较好的文本水印认证算法。针对格式化中文文本中数字水印抗攻击性不强、鲁棒性较差、透明性不太好的问题，本书提出了一种基于数字版权保护的鲁棒性文本数字水印算法，该研究对解决目前因特网上比较脆弱的安全机制是一个有益的尝试，其研究成果可广泛应用于电子商务、电子政务、国防、国家安全、数字图书馆等版权保护领域。实际应用中的文本数字水印系统对安全性提出了较高的要求。因此，建立满足安全性需求的文本水印模型对文本水印系统的设计、管理和运行是至关重要的。

（3）为实际应用领域中文本信息安全问题提供一定的理论、方法与技术。本书从实际应用需求出发，通过对目前的文本水印算法进行分析、归纳和总结，提炼出共同存在的一个安全性问题，以中文信息处理与数字水印相关理论为基础，展开有关安全性文本数字水印认证算法的研究，从而推进文本水印安全性理论的研究和应用发展，使之更好地服务于社会。

综上所述，本书的研究具有十分重要的理论意义和实际应用价值。文本数字水印技术作为一门新兴的多学科交叉的应用技术，它的发展在一定程度上也带动这些相关领域的进一步发展。因此，无论从理论角度还是应用角度来说，开展对文本数字水印技术的研究，不但具有重要的学术意义，而且还具有极为重要的实际应用价值。本书的研究将为国家安全、经济建设和社会发展所需要的网络信息安全保障提供一定的科学依据；为解决对国民经济发展起关键作用和可能推动国民经济发展的科学技术问题、社会信息化急需解决的网络信息安全问题提供一定的技术支持。

1.3 文本水印研究综述

1.3.1 数字水印研究与发展

1954 年，出现了第一个与数字水印方法类似的技术实例。当时，美国 Muzac 公司的 Emil Hembrooke 申请了一项名为“Identification of Sound and Like Signals”的专利。该专利描述了一种将标识码隐蔽地嵌入到音乐中而证明其所有权的方法。这是迄今为止所知道的最早的电子水印（Electronic Watermarking）技术。直到 20 世纪 90 年代初期，数字水印术语才真正流行起来。早期，关于数字水印的