

学号 22151214548

西 安 电 子 科 技 大 学  
专业学位硕士学位论文开题报告表

论文题目: 基于随机内容插入的可信存证关键技术研究

姓 名 孙恒康

学位类别 专业学位

领 域 电子信息

校内导师 李风华

校外导师 郭云川

学 院 广州研究院

开题日期 2024 年 1 月 11 日

西安电子科技大学研究生院制

## 西安电子科技大学硕士学位论文开题报告要求

一、硕士研究生必须在第三学期末之前进行学位论文开题报告。

二、硕士学位论文的开题报告会由各学院自行组织，硕士研究生必须如实、如期在本学科（领域）或相关学科（领域）范围内公开举行开题报告会，严禁伪造和抄袭开题报告。

三、开题报告结论分为两种：1. 通过，按专家意见修改后进行学位论文撰写工作；2. 不通过，重新开题。

四、在学位论文开题规定的时间期限内休学的硕士研究生，开题的时间期限相应顺延。

五、开题必须在规定时间内通过研究生学位管理系统申请，本表由系统自动生成，用 A3 纸张正反套印。

六、表格填写要求：正文字体宋体，字号小四，行间距固定值 20 磅。

一、论文概况

实习单位名称	中国科学院信息工程研究所
实习岗位	开发工程师
实习实践模式	校外实习
计划实习时间	2023 年 7 月 11 日至 2025 年 7 月 11 日
论文类型	技术论文
选题来源	国家重点研发计划课题
中文摘要	
	<p>随着信息技术的发展，人们在网络上的产生的一些数据或操作记录需要被存证，内部的存证系统安全性不高，数据有被伪造的风险，所以我们需要采取措施防止存证信息被伪造。</p>
	<p>如果存证对象不存储在本地而是存储在一个安全的中心存证系统，虽然安全性增加，但是会增加网络负担而且有泄密风险。所以需要在不传输被存证对象的情况下对本地存证信息进行安全的存证保存。所以选择的存证对象是被存证信息的散列值。</p>
	<p>相比于直接存证对象的散列值，本文提出了针对不同模态的存证信息的随机内容的生成算法和将生成的随机内容随机插入到原始信息的随机插入规则算法。将随机内容按照插入规则插入到原始信息中，然后对插入后的信息进行散列值的生成。由于散列值是本地存证和中心存证系统合作生成的，因此原始信息被伪造的情况理论上不可能，可以有效避免有意伪造存证信息的情况发生。</p>
	<p>关键字：可信存证、随机内容生成、随机插入规则、多模态</p>

## 二、选题依据

### （一）选题意义

随着信息技术的飞速发展，人们在网络上产生的大量数据和操作记录需要进行审计和存证，以确保信息的完整性、可追溯性和安全性。然而，不同组织内部的防护条件差异巨大，有些防护能力较弱的组织无法保证内部的存证信息的安全，存证信息和审计信息有被修改和伪造的可能，因此当需要内部存证信息的时候，信息的可靠性就会大打折扣。

但是如果将需要存证的信息直接传输存放在可信的第三方也会带来诸多问题：当需要存证的信息涉及到商业机密或者个人的隐私数据的时候，将数据存储到第三方会带来数据泄露的风险；而且当需要存证的信息数据量巨大时候，网络传输会影响存证速度，拖慢业务的运行速度。所以常见的处理方法是将需要存证的数据生成认证校验码，只需要将认证校验码传输给第三方的存证系统，这样就能利用避免将信息保存在第三方的面临的数据泄露风险和效率低下的问题。

在单纯对原始存证信息生成认证校验码并向第三方机构发送认证校验码的基础上，还可以为校验码的生成引入新的随机性，即第三方存证中心根据用户提供的存证信息模态信息及其他数据信息（如文件名，文件大小，标题信息）生成和原始信息相同模态的唯一随机内容并将随机内容随机插入原始存证信息中。最后生成插入后的认证校验码。本文针对这一个业务流程中如何生成多模态随机信息和随机插入规则设计了适用于文本模态，结构化数据和图片数据的随机内容生成算法和插入规则算法，提高了存证信息的可信度。

本文提出了针对不同模态的存证信息生成相同模态的随机内容生成算法和将生成的随机内容随机嵌入到原始信息的插入算法。通过将随机内容随机插入到原始信息中然后在对插入后的信息进行认证校验码的生成。由于随机内容生成算法和随机内容插入算法是根据不同的信息生成的，具唯一性，认证校验码具有机密性、完整性和不可否认性的特性，因此原始信息被伪造且伪造信息可以生成相同认证校验码理论上不可能，可以有效避免有意伪造存证信息的情况发生。

### （二）国内外研究现状

根据要实现的算法：多模态随机内容生成算法和多模态随机内容插入算法。首先要解决随机性的问题。目标是不同的存证文件要生成对应模态的完全随机的唯一的内容。所以要充分了如何产生随机性和唯一性，同时兼顾生成内容的模态，图片和文本文件和结构化文件不同，不是文本文件，所以需要如何在图片文件中插入随机内容，这一部分借鉴隐写术对图片文件的修改方式。

## 1. 伪随机生成器

伪随机数发生器（PRNG）在许多应用中是有用的，例如网络通信中的密码系统<sup>[1]</sup>，并且模拟中的重要方法之一是蒙特卡罗<sup>[2]</sup>，其需要产生非常高质量的伪随机序列和数值积分计算，这不是常规方法解决的<sup>[3-5]</sup>。关于对安全的威胁<sup>[3-16]</sup>，PRNG 的统计质量变得比以前更重要。例如，一台超级计算机可能每秒产生 109 个随机数，而密码算法需要 1016 个随机数才能在一个非常重要的通信中创建一个安全通道，因此，所产生的序列中的小相关性或其他弱点很容易导致几个网络层中的关键泄漏。这些分布应该根据它们的商业应用如“正态分布”、“指数分布”、“泊松分布”等来准备，只考虑均匀分布数的生成。更详细地说，集中讨论均匀分布在区间  $(0 \dots 1)$  上的真实的数列。产生伪随机数的基本点是这些产生器是确定性的，因为数字计算机不能产生真正的随机数。因此，需要提出统计测试，并且 PRNG 在被发布用于通信网络中的安全使用之前应该通过一些重要的统计测试。使用混沌映射作为可靠的 PRNG 的概述。

利用组合混沌映射产生随机数是改善统计特性的最佳方法之一。例如，在 2006 年，Wang 等人提出了一种基于 z-logistic 映射的伪随机数生成器<sup>[4]</sup>。2007 年，Ergun 和 Ozogur 提出了非自治混沌电子电路的新随机比特序列<sup>[5]</sup>。然后，Hu 等人提出了一种通过计算机鼠标运动的真随机数生成器<sup>[6]</sup>。2009 年，Patidar 等人设计了一种基于两个混沌 Logistic 映射的随机比特发生器，该混沌 Logistic 映射通过比较两个混沌 Logistic 映射的输出而产生<sup>[7]</sup>。最近在 B. Fechner 和 A. Osterloh 提出了一个元级真随机数生成器<sup>[8]</sup>。分布的均匀性是随机序列统计检验的主要问题。这意味着在生成随机或伪随机比特序列的所有点上，零或一的概率与九的概率一样多<sup>[9]</sup>。

如何挑选合适的伪随机生成器来生成随机内容和随机规则是研究的重点。

## 2. 加密散列

加密散列函数是一种单向数学函数，它将任意长度的输入消息转换为固定长度的唯一散列值，使得给定散列值时无法计算输入消息。产生的哈希值也称为消息摘要或校验和，用作输入消息的“签名”。本质上，一个特定的消息将生成一个唯一的消息摘要；它只能由该消息生成。这在需要验证某些消息或个人的情况下特别有用。例如，当用户在 Facebook 上注册在线帐户时，所提供密码的哈希值存储在 Facebook 的数据库中，当用户稍后尝试登录时，输入密码的哈希值将与存储的哈希值进行比较。如果两个哈希值相同，则授予用户访问权限。SHA-2（Secure Hash Algorithm 2）是由美国国家安全局（NSA）设计的一组加密哈希函数。SHA-2 包含了其前身 SHA-1 的重大变化。SHA-2 家族由六个哈希值为 224、256、384 或 512 位的哈希函数组成：SHA-224、SHA-256、SHA-384、SHA-512、

SHA-512/224、SHA-512/256 [9]。在该算法中，SHA-256 的概念用于生成随机模式；使用该随机模式，消息比特分散在覆盖介质中。

### 3. 关于图片的随机内容插入

密码学是在不安全的信道上安全通信所采用的各种技术的艺术和研究。加密是将普通消息（称为明文或明文）转换为基于密钥的加密消息，并将其加密为看似垃圾的文本（称为密码文本）的过程。解密是这个过程的逆过程，在这个过程中，基于密钥将这些无法理解的文本解扰回原始消息。隐写术是一种将秘密消息（或文件）隐藏在一个封面介质中的艺术，该封面介质可以是无害的，看起来无害的多媒体文件，甚至是 IP 数据包，而没有任何可察觉的变化。秘密消息被称为有效载荷，并且使用密钥均匀地分散在覆盖介质的字节上。由此产生的，修改后的封面媒体被称为隐写图。密钥是从隐写图中解码隐藏信息所必需的。

根据所使用的覆盖介质，存在各种类型的隐写术-例如，图像隐写术、音频隐写术、文本隐写术等。还存在各种技术来实现每种类型的隐写术-例如，为了实现文本隐写术，可以使用行移编码、字移编码、特征编码等<sup>[10]</sup>。隐写术的另一种形式是在常用系统中创建隐蔽通道，例如选择性地将文件系统中的某些文件分段，其中连续文件块之间的差异将表示一个字节的的信息<sup>[11]</sup>。隐写系统背后的基本思想是将数据嵌入到根本不期望数据存在的地方，例如，在硬盘微控制器的固件中<sup>[12]</sup>。LSB (Least Significant Bit, 最低有效位) 编码是图像、音频和视频隐写术中最常用和最简单的技术之一。

何数字数据都以字节存储，每个字节包含 8 位。具有最低数值（位权重= 1）的位被称为最低有效位（LSB）-简单地说，最右边的位-在图 1 中突出显示。第二最低有效位（LSB+1）是具有第二最低数值的位（位权重= 2）。类似地，最高有效位（MSB）是具有最高数值（位权重= 128）的位，其是最左边的位。

图像、音频和视频文件具有大量的细节，并包含合理的冗余量。对这种类型的文件的数据字节进行非常轻微的更改不会导致人类感官系统可感知的差异。例如，在 24 字节的彩色图像系统中，每个像素可以获得 1600 万个以上的值。假设存在红色分量值为 100、绿色分量值为 200 且蓝色分量值为 150 的像素。将该特定像素的每个分量值递增 1，将使整个图像发生非常轻微的变化-人眼无法区分。在 LSB 编码中，需要根据要隐藏的消息的位值来改变覆盖介质的数据字节的最低有效位。因此，覆盖介质的第一字节的 LSB 应该与要隐藏的消息的第一位（MSB）相同；覆盖介质的第二字节的 LSB 与要隐藏的消息的第二位（第二 MSB）相同，覆盖介质的第三字节的 LSB 与消息的第三位（第三 MSB）相同，等等。现在，要解码隐藏的消息，只需读取封面媒体每个字节的最低有效位；将

8 位放在一起形成消息的单个字节。重复此过程以获取整个隐藏消息。

可以将隐写技术对图片信息的处理方式和到随机信息生成和随机规则插入算法相结合，在图片信息中插入生成的随机信息。

#### 参考文献：

- [1] P. L. Ecuyer and R. Panneton, "Fast Random Number Generators Based on Linear Recurrences Modulo 2: Overview and Comparison," Proceedings of the Winter Simulation Conference, IEEE Press, Springer, New York, 2005, pp.110-119. doi:10.1007/978-1-4419-1576-4
- [2] C. Robert and G. Casella, "Introducing Monte Carlo Methods with R," Springer Textbook, New York, 2010.
- [3] B. Jun and P. Kocher, "The Intel Random Number Generator," White Paper Prepared for Intel Corporation, California, April 1999, pp. 1-8.
- [4] L. Wang, F.-P. Wang and Z.-J. Wang, "Novel Chaos Based Pseudo-Random Number Generator," Acta Physica Sinica, Vol. 55, 2006, pp. 3964-3968.
- [5] S. Ergun and S. Ozoguz, "Truly Random Number Generators Based on a Non-Autonomous Chaotic Oscillator," AEU-International Journal of Electronics & Communications, Vol.61, No.4, 2007, pp.235-242. doi:10.1016/j.aeue.2006.05.006
- [6] Y. Hu, X. Liao, K.-W. Wong and Q. Zhou, "A True Random Number Generator Based on Mouse Movement and Chaotic Cryptography," Chaos Solitons and Fractals, Vol. 40, No. 5, 2009, pp. 2286-2293. doi:10.1016/j.chaos.2007.10.022
- [7] V. Patidar, K. K. Sud and N. K. Pareek, "A Pseudo Random Bit Generator Based on Chaotic Logistic Map and its Statistical Testing," Journal of Informatical, Vol. 1, No. 1-3, 2009, pp. 441-452.
- [8] B. Fechner 和 A. Osterloh, "A Meta-Level True Random Number Generator," International Journal of Critical Computer-Based Systems, Vol. 1, No. 1-3, 2010, pp.267-279. doi: 10.1504/IJCCBS.2010.031719
- [9] I. Shparlinski, "On the Uniformity of Distribution of the Decryption Exponent in Fixed Encryption Exponent RSA," Journal of Computation Theory and Mathematics, Vol. 92, No. 3, 2004, pp.143-147.
- [10] Wang, H & Wang, S, "Cyber warfare: Steganography vs. Steganalysis", Communications of the ACM, 47:10, October 2004
- [11] Anderson, R.J. & Petitcolas, F.A.P., "On the limits of steganography", IEEE Journal of selected Areas in Communications, May 1998
- [12] Marvel, L.M., Boncellet Jr., C.G. & Retter, C. "Spread Spectrum Steganography", IEEE Transactions on image processing, 8:08, 1999

### 三、研究方案

#### (一) 研究目标

本文将提出一种针对不同模态的随机内容生成算法和对生成随机内容的插入算法，支持包括文本文件，结构化文件，图片文件。可以做到对不同模态的不同的文件或相同模态的不同文件生成唯一的和原始存证信息相同模态的随机内容：并随机插入到原始存证信息的随机位置：头部、尾部、或者文件任意范围内。这样使得伪造生成具有相同语义且相同类型的散列值理论上不可能，可以有效避免有意伪造存证对象的情况发生，随机内容和随机插入规则也可以作为侵权的证据。在算法的基础上完成多模态随机内容生成和随机插入系统，包含客户端和服务端，作为对算法的检验和具体应用。

#### (二) 研究内容

本部分内容要体现出学位论文的整体设想及构架。

##### 1. 研究内容：

(1) 不同模态的随机内容生成算法：针对 3 种不同的模态：文本文件，结构化文件，图片文件，根据用户传入的文件模态类型和文件其他相关信息：文件名称，大小，标题信息等，生成和原始文件模态相同语义相近的随机内容。

(2) 对 1 中生成的随机内容的随机插入算法：针对 3 种不同的模态：文本文件，结构化文件，图片文件将对应随机内容按照随机的插入规则：包括但不限于文件头部、尾部、数据任意节点，而且使原文件的语义不会发生明显变化。

(3) 多模态随机内容生成和随机插入系统的设计与实现：包括一个客户端和服务端，基于上述研究内容 1 和 2 提出的算法，设计实现多模态随机内容生成和随机插入系统，验证所提出算法的可行性、可用性和高效性。并且要确保系统服务端能够正确处理大量用户的随机生成和插入规则请求，安全保存用户的认证及校验码和随机信息和随机规则。客户端能够正确向服务器传输文件特征，按照服务端返回的随机内容和随机插入规则完成对原始数据的改造，还要保证服务端和客户端的安全通信。

##### 2. 论文的整体设计及架构：

(1) 绪论。说明研究背景和意义、国内外研究现状、本文研究内容、课题工作和本文章节安排。

(2) 背景知识。说明基于多模态的随机内容生成关知识、高效生成算法、随机插入规则算法相关知识等。

(3) 多模态的随机内容生成算法、随机插入算法的设计。说明算法设计概要和算法性能分析。

(4) 处理用户请求的随机内容生成和插入系统的系统设计。说明系统设计



概要和系统详细设计。

（5）实现与测试。说明系统开发与运行环境、系统实现、系统功能测试、系统性能测试。

（6）总结和展望。说明本文的工作总结，并对接下来的工作进行展望。

### （三）拟解决的关键问题

#### （1）不同模态的随机内容生成

由于不同模态的数据有不同的数据结构。需要在在用户有限的特征信息下（文件名称，大小，标题信息等）输出唯一的随机内容，是算法设计中的难点和重点。

#### （2）不同模态的随机插入规则

不同的数据有不同的数据格式和数据大小，根据用户提供的有限的特征信息（文件名称，大小，标题信息等）。生成一个让原始数据高效可用的插入规则，是算法设计的重点。

#### （3）减少对原文件语义的改变

随机内容生成算法和随机规则插入算法需要配合，使得生成的随机内容在按照插入规则插入后尽量少的减少原始数据语义的改变。如何在插入随机内容后使得文本文件，结构化文件和图片文件做到语义尽量少的减少使研究问题。

#### （4）随机内容生成和随机插入规则生成系统处理如何高效处理大量请求

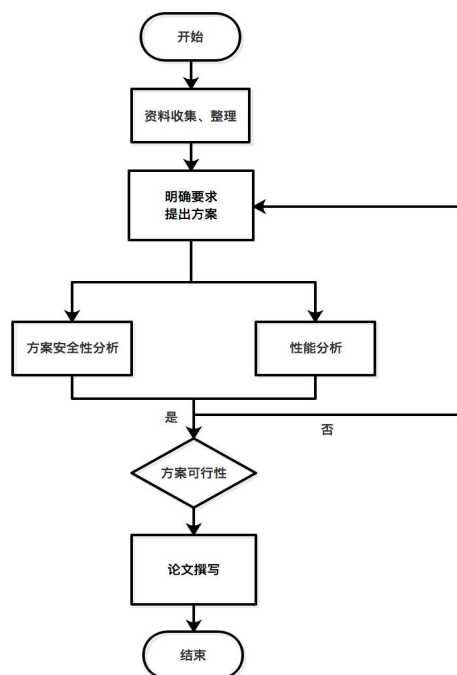
由于网络数据流量具有突发性的特点，面对数量庞大的用户群的突发请求，系统如何做好网络请求的处理，高效生成随机内容和插入规则并返回给用户结果，需要解决如何设计良好的网络分流策略和请求接口。

### （四）拟采取的研究方法、技术路线、实验方案及可行性研究

#### 1. 研究方法

积极跟踪国内外针多模态随机内容生成、随机规则插入最新研究成果，汲取新的思想，发现其中的不足并对其进行改进。进行趋势分析，从实际应用需求出发，找出需要解决的关键技术问题，理论结合实践，将提出的方案进行仿真，并与现有成果进行比较，以证明新提出方案的安全性和高效性。

#### 2. 技术路线



### 3. 实验方案

本文首先提出多模态的随机内容生成和随机插入算法，然后针对特定的模态随机生成和插入中出现的问题：不同模态数据要生成随机的相同模态的随机数据，并且数据数据的插入规则要使得原始数据对随机数据的插入高效准确，且要保证数据的生成内容和插入规则完全随机。最后要使用可处理高并发请求的系统服务端和可以发送存证求情并可以在原始存证信息中插入随机内容的服务端来验证算法的生成效果和插入效果。用大量数据进行实验验证算法的性能和准确确定。

### 4. 可行性研究

通过分析了所研究课题的研究背景和国内外研究现状，以研究内容、技术路线和前期工作准备为基础，我认为在导师的帮助下，通过不断的努力，最终完成本课题的研究目标是可行的。

#### （五）研究计划及预期取得的研究成果

研究计划要具体，要明确指出每一个时间段的学位论文进展情况及预期取得的研究成果。

起 止 时 间	工 作 内 容
2023.09.01 — 2024.01.02	收集相关资料，了解研究背景及国内外相关研究现状，完成开题报告。
2024.01.03 — 2024.03.31	学习不同模态的数据结构和相关的算

	法为之后的研究打下基础。
2024.04.01 — 2024.06.30	系统学习针对不同模态的数据生成和插入算法及算法的应用，并初步完成算法设计。
2024.07.01 — 2024.09.30	继续阅读相关文献，及时跟踪最新研究成果，并改进和完善方案的设计。
2024.10.01 — 2024.12.31	针对设计的认证方案，进行理论和方针上的验证，如果出现问题，及时查找和修正。
2025.01.01 — 2025.02.28	对新提出的方案进行安全性分析和性能分析，并在 Centos 环境进行实现。
2025.03.01 — 2025.04.30	总结毕设研究内容，完善代码，根据论文撰写要求，完成毕设论文。准备答辩。

#### 四、研究基础

##### （一）已具备的实验条件和研究工作积累

- （1）查阅针对不同模态的随机内容生成，随机插入实际应用相关论文、期刊等，基本了解了一些方法；
- （2）从知网上了解了一些随机内容生成算法和不同模态的内容插入算法；
- （3）熟练掌握 c++，qt，数据库编程技术，可以将系统实现落地。

##### （二）已取得的科研成果

联动处置与控制系统 V1.0（排名第 3，学生第一完成人）

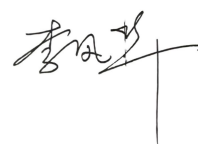
## 五、指导教师意见

（重点对硕士生的开题情况以及是否同意开题予以说明。）

对选题的研究内容进行了比较充分的调研，研究内容和路线可行，同意开题。

校内导师签名：

2024 年 1 月 9 日



（重点对硕士生的开题情况以及是否同意开题予以说明。）

对选题的研究内容进行了比较充分的调研，研究内容和路线可行，同意开题。

校外导师签名：

2024 年 1 月 9 日



## 六、开题报告记录

（着重记录专家对选题报告提出的问题及修改意见和建议。）

对背景的介绍有些繁琐，主要讲解自己的工作。

七、开题报告评语及结论

一级指标	二级指标	评价意见必填
论文选题	1. 选题具有重要的理论意义或实际意义，是直接面向工程或具有探索性的应用课题； 2. 国内外研究现状综合全面反映该领域的最新研究成果，归纳总结正确。	<input type="checkbox"/> 优秀 <input checked="" type="checkbox"/> 良好 <input type="checkbox"/> 合格 <input type="checkbox"/> 不合格
研究方案	1. 研究目标明确； 2. 整体设想及构架科学合理； 3. 研究或设计方法科学合理，关键技术有难度； 4. 预期取得的研究成果具有实用性和新技术应用价值，可产生一定的社会效益和经济效益。	<input type="checkbox"/> 优秀 <input checked="" type="checkbox"/> 良好 <input type="checkbox"/> 合格 <input type="checkbox"/> 不合格
研究基础	具备了较好的实验条件和较为深厚的研究工作积累。	<input type="checkbox"/> 优秀 <input checked="" type="checkbox"/> 良好 <input type="checkbox"/> 合格 <input type="checkbox"/> 不合格
进度安排	时间安排充裕、合理。	<input type="checkbox"/> 优秀 <input checked="" type="checkbox"/> 良好 <input type="checkbox"/> 合格 <input type="checkbox"/> 不合格
<p><b>开题报告评语及结论</b></p> <p>（开题报告结论分为两种：1. 通过，按专家意见修改后进行学位论文撰写工作；2. 不通过，重新开题。）</p> <div><div>组长签名： 苏晓丹</div><div>成员签名： 付玉龙 赵兴文 刘 樵 斗 崑</div><div>2024 年 1 月 11   日</div></div>		

注：填写评价意见时，请在相应评价意见前的“□”中打“√”。