**一、基本情况**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **姓名** | **性别** | **学生类型** | **导师** | **指导老师** |
| **孙恒康** | **男** | **专硕（客座）** | **李凤华** | **郭守坤** |
| **入学年份** | **硕转博时间** | **计划毕业时间** | **本科学校** | **本科专业** |
| **2022** | **—** | **2025** | **西安邮电大学** | **信息对抗** |
| **毕设题目：基于随机内容插入的可信存证关键技术研究** | | | | |

**二、科研来源**

（1）国家重点研发计划2.6项目 “隐私数据的个人权益保障研究”。

**三、研究基础**

（1）基础语言：C/C++，Qt。

（2）系统开发经验：

参与2.6项目课题5子课题1的“异常操作汇聚存储系统”的研发工作，基于Epoll+事件驱动模型实现I/O多路复用网络通信，支持动态调整线程池和任务队列管理，支持FTP、HTTP标准协议支持文件推送。基于Qt研发管理系统前台UI，可视化展示汇聚系统的异常操作日志数据。

参与了1.2项目课题2子课题3的“联动处置与控制系统”UI前台页面的优化与修改的工作。

**四、研究目标**

针对不同模态的存证对象，设计算法生成随机内容和随机内容的插入规则，采用两级协同约束的可信存证架构，由本地存证根据中心存证提供的随机内容和随机内容的插入规则，按规则操作存证对象内容后，生成存证对象的随机认证校验码并上报中心存证，防止存证对象内容被恶意伪造和篡改。

**五、拟解决的问题**

解决存证对象内容被伪造和篡改的问题。

**六、研究内容**



**研究点一：随机内容生成算法**

针对图像、音频、文本不同模态的文件，根据上报存证对象的关键信息集合（文件名称、标题、大小、模态等），针对不同的模态设计随机内容的生成算法，生成的随机内容配合随机内容插入规则使不改变原始存证对象的模态。

对于图像模态，可生成文本信息、图像信息、二进制数据等，转换成二进制数据作为随机内容；对于音频模态，可以生成文本信息、图片信息、音频信息等，转换成二进制数据作为随机内容或是压缩后的音频；对于文本模态，可生成文本信息转换成二进制数据（通过如ASCII编码、Unicode编码等）作为随机内容。

**研究点二：随机内容的插入规则生成算法**

对于研究点一中涉及不同模态的文件，梳理图像、音频、文本等不同模态文件的插入位置对文件模态和内容的影响，结合生成的随机内容，分别设计随机内容的插入规则算法，实现按照随机内容和插入规则操作存证对象生成随机认证校验码，不影响存证对象的模态和内容的展示效果。

对于图像模态，可生成文本信息、图像信息、二进制数据等，将二进制随机内容以最低有效位（LSB）替换、频域嵌入等插入规则；对于音频模态，将二进制随机内容采用时域最低有效位（LSB）或者频域调整的插入规则；对于文本模态，将二进制数据采用格式、不可见编码、特定位置插入规则。

**原型验证系统**

基于上述研究点一和二提出的算法，设计实现原型验证系统（包括两层存证系统），系统验证提出算法的可行性、可用性和有效性。系统采用I/O多路复用、线程池、连接池能技术保证能够正确接收大量用户建立的连接，并正常处理大量的业务的请求响应。中心存证可根据本地存证上报的不同存证对象的关键信息集，生成随机内容和插入规则。本地存证根据随机内容和插入规则完成对原始存证对象的操作，并上报随机认证校验码。原型验证系统可验证本地存证是否按照随机内容及插入规则进行操作生成的随机认证校验码，检查存证对象是否被伪造。

**七、预期产出及验收方式**

（1）软件著作权2项。

（2）支撑2.6项目课题5子课题系统的研发、测试和验收等工作。

**八、成果应用描述**

支撑2.6项目结题验收。

**九、科研条件**

（1）可复用组里目前提供的服务器。

**十、工作计划**

2024年12月前，完成研究点相关内容的调研，完成2.6项目系统功能的研发与测试工作。

2025年3月前，完成毕业论文的撰写。

2025年5月，具备答辩所有条件，完成毕业答辩。

1. **附件**

**论文阅读清单：**

[1] J. Qian, H. Du, J. Hou, L. Chen, T. Jung and X. -Y. Li, "Speech Sanitizer: Speech Content Desensitization and Voice Anonymization," in IEEE Transactions on Dependable and Secure Computing, vol. 18, no. 6, pp. 2631-2642, 1 Nov.-Dec. 2021, doi: 10.1109/TDSC.2019.2960239.

[2] Li Y, Huang Y. End-to-End Speech Hash Retrieval Algorithm based on Speech Content and Pre-training[J]. Frontiers in Computing and Intelligent Systems, 2024, 8(2): 22-28.