密文重复数据删除机制的频率分析攻击

任彦璟

April 2, 2019

现有频率分析攻击在加密重复数据删除中效果不佳，
如何有针对性的提高攻击效果？

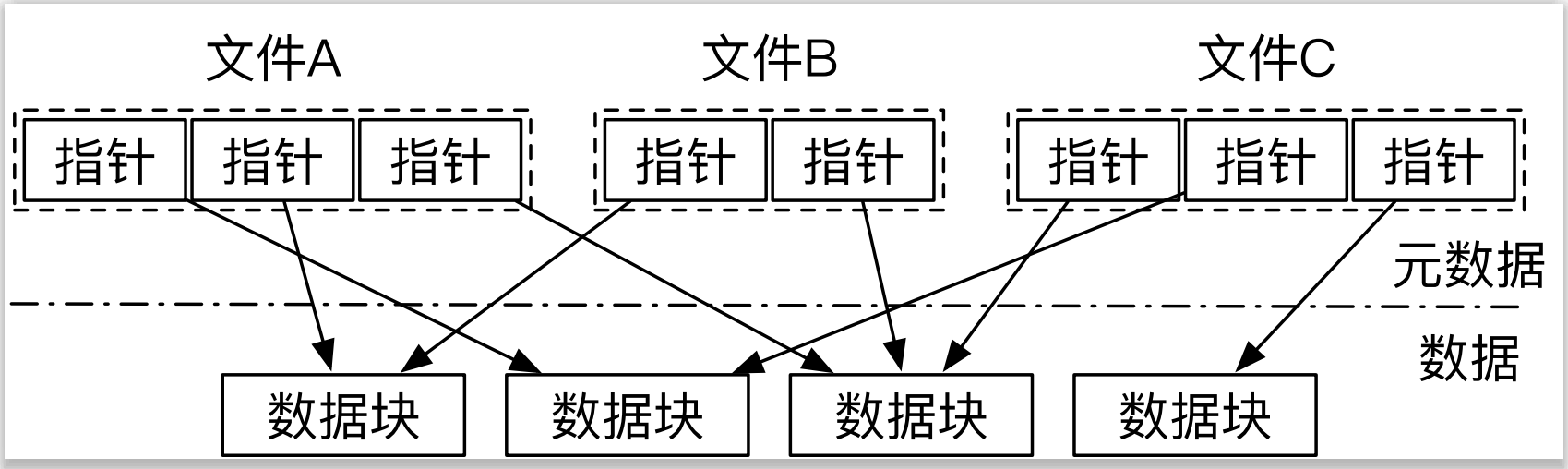
目录

1. 对选题的分析
2. 国内外研究现状
3. 主要研究内容
4. 攻击方案的设计
5. 研究结果与分析
6. 总结建议与参考文献

对选题的分析

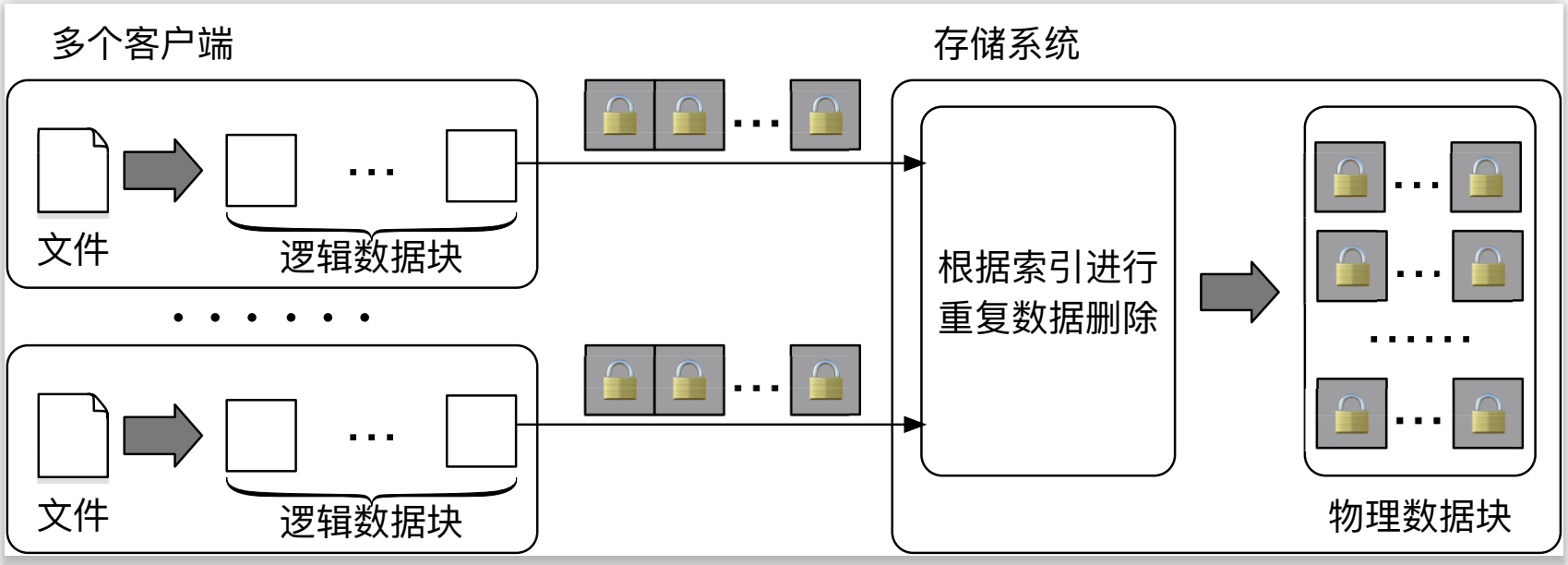
- 背景知识
- 两大核心问题
- 研究的意义
- 研究的目的

背景知识-重复数据删除



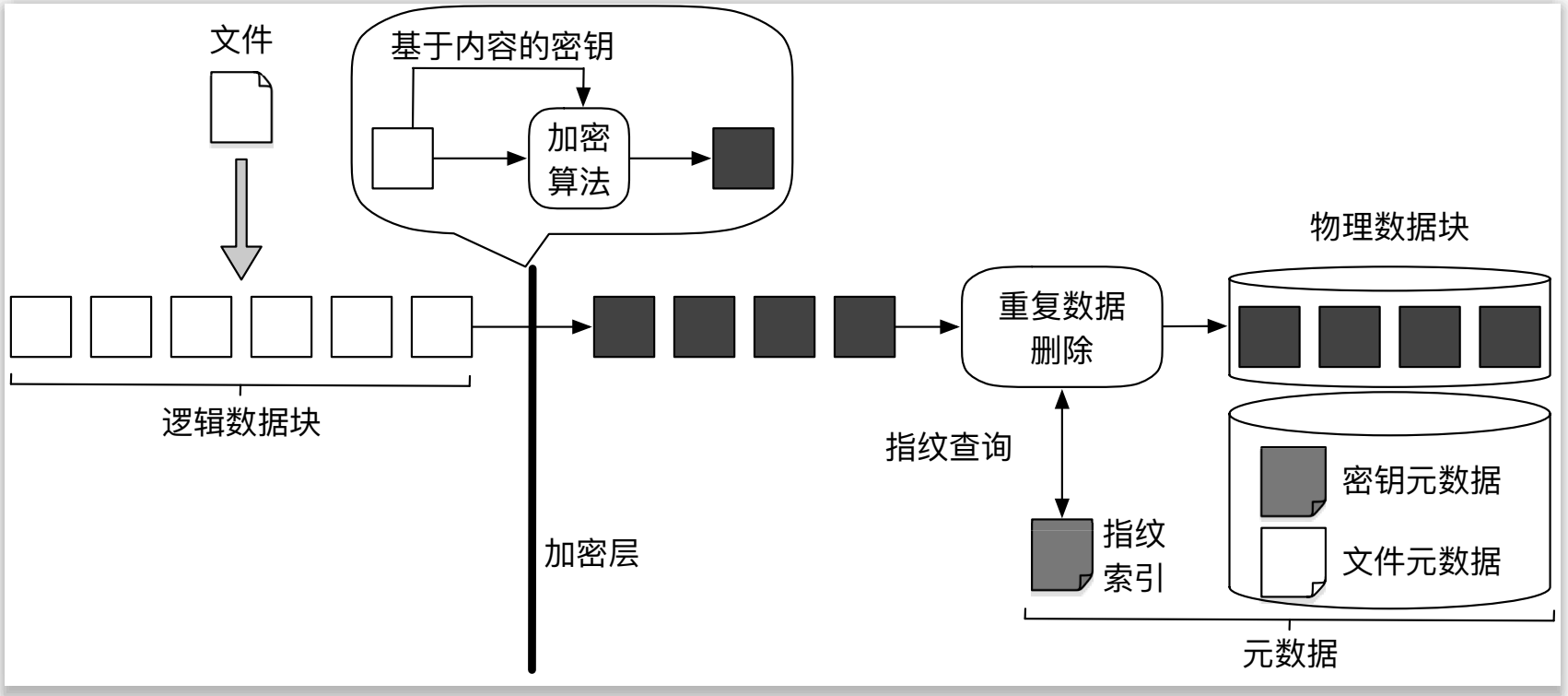
重复数据删除系统的存储方式

背景知识-重复数据删除



重复数据删除系统的运作流程

背景知识-密文重复数据删除



密文重复数据删除的运作流程

背景知识-频率分析攻击

频率分析是一种针对确定性加密的密码分析技术，被应用于破解匿名查询日志、破坏关键词隐私、重构密文数据库记录等实际攻击。

背景知识-MLE加密

消息锁定加密确立了加密重复数据删除的密码学基础：基于数据内容产生密钥，从而将相同明文加密为相同密文。

两大核心问题

数据块频率泄漏问题：加密重复数据删除广泛应用MLE，导致数据块的频率信息泄漏。

针对数据块的频率分析攻击，与现有攻击目标（查询日志条目、关键词、数据库记录等）相比，数据块数量极其庞大（呈千万级），并且大量数据块具有相同频率，致使当前的频率分析攻击算法难以适用。

研究的意义

填补频率分析攻击研究空白。

对理解加密重复数据删除的实际安全性，并降低其在非适合场景下的误用风险具有重要作用。

研究的目的

在理论上：构造针对加密重复数据删除的频率分析攻击，揭示实践中的安全隐患。

在技术上：以理论研究为支撑，设计并实现针对加密重复数据删除的频率分析攻击工具，并在真实系统中进行理论验证和攻击效果测试。

国内外研究现状

- 加密重复数据删除
- 频率分析攻击
- 密文重复数据删除的其他攻击

加密重复数据删除

密码学理论基础：消息锁定加密（MLE）

- 收敛加密（CE）使用明文的哈希值作为MLE密钥，并基于密文哈希值计算指纹，以识别重复数据。
- 哈希收敛加密（HCE）与CE具有相同的MLE密钥产生规则，但基于明文哈希值计算指纹。
- 随机收敛加密（RCE）使用随机密钥加密以产生非确定的密文，同时基于明文哈希值来进行重复检查。
- 收敛扩散（CD）使用明文哈希值作为秘密共享的输入种子，提高了密文存储的可靠性。

频率分析攻击

针对确定性加密的密码分析技术。

面向加密重复数据删除，已有工作提出了基于数据块局部性（chunk locality）的频率分析攻击。

密文重复数据删除的其他攻击

加密重复数据删除可能遭受边信道攻击、副本伪造攻击、基于数据块长度的攻击等威胁，但这些攻击可通过所有权证明、守卫解密(guarded decryption)、固定长度分块等措施进行防御。

本研究的频率分析攻击超出了现有保护措施的防御范畴。

对选题的分析

- 三个主要研究内容.
- 研究的技术路线.

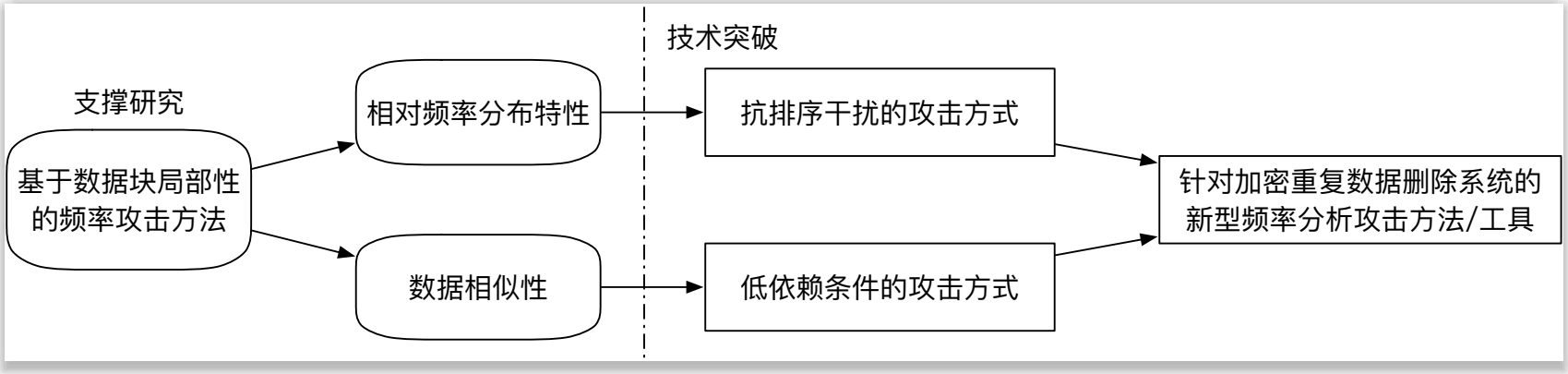
三个主要研究内容

研究基于数据特征的新型频率分析攻击技术，提高传统频率分析的攻击效果。

分别从抵抗频率排序干扰和降低攻击发生条件两方面改进攻击技术。

实现针对真实系统的频率分析攻击原型，并分析该攻击对各类数据安全性的影响。

研究的技术路线

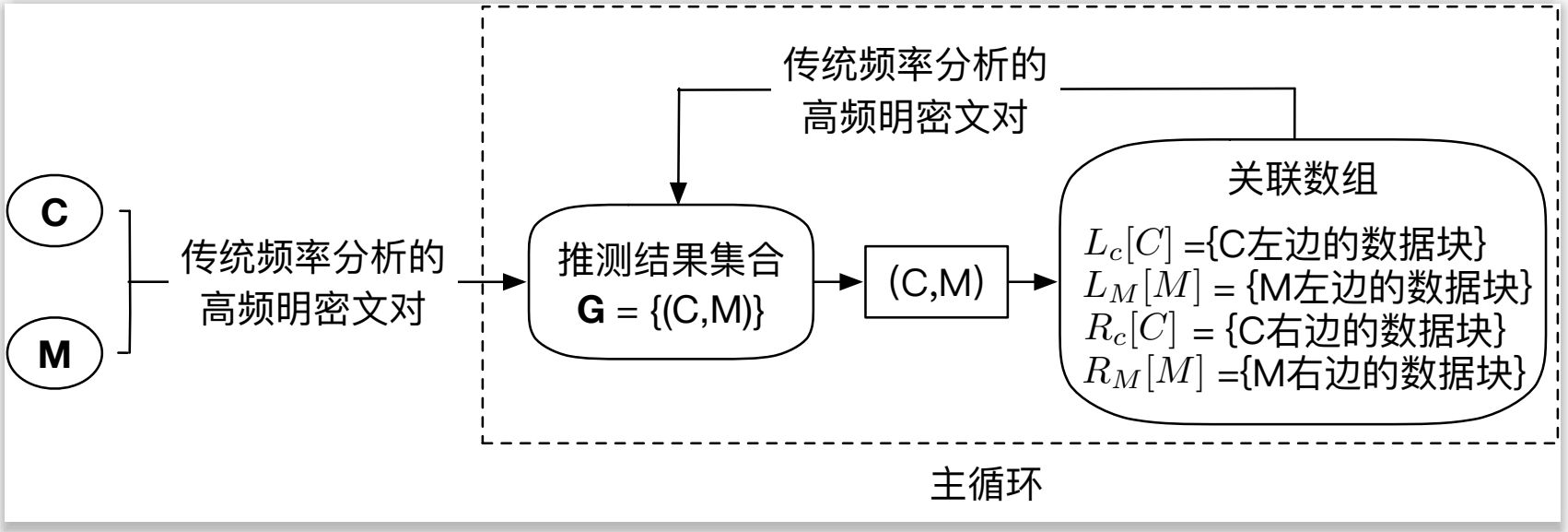


本研究的技术路线

攻击方案的设计

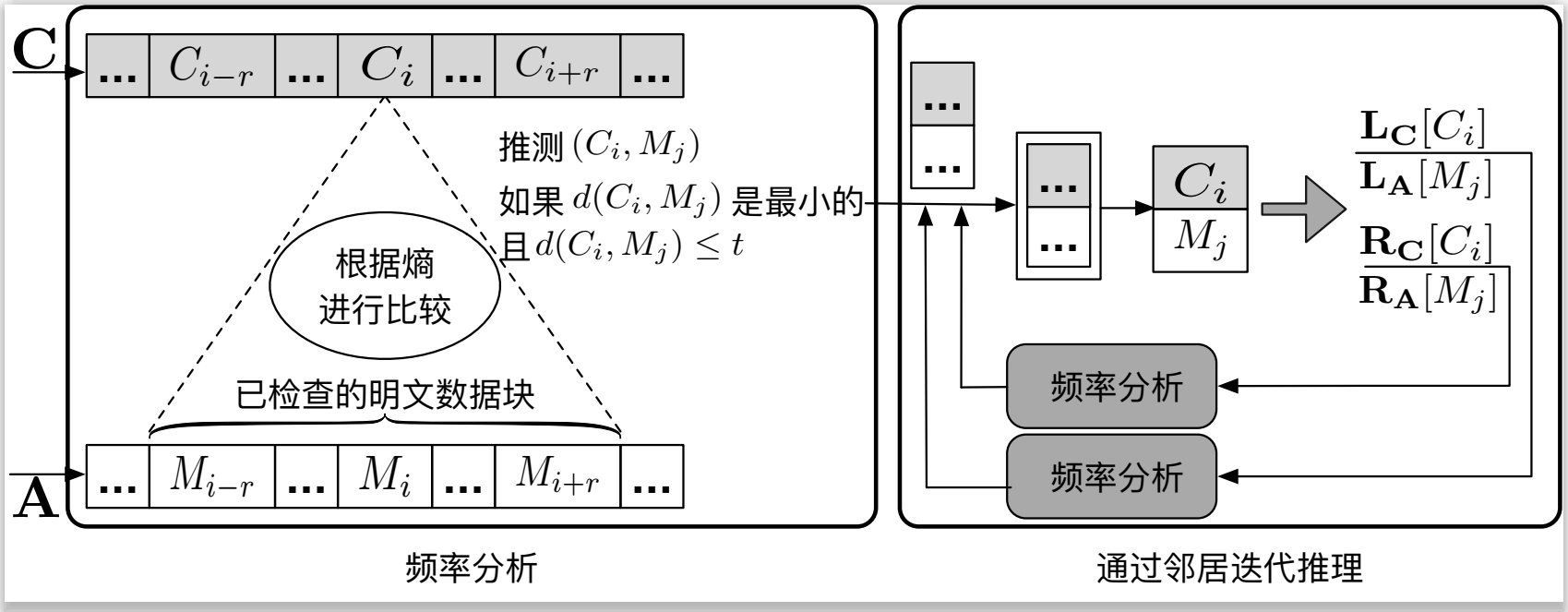
- 已有工作：基于数据块局部性的攻击
- 本课题研究内容：基于分布的攻击
- 本课题研究内容：基于聚类的攻击

基于数据块局部性的攻击



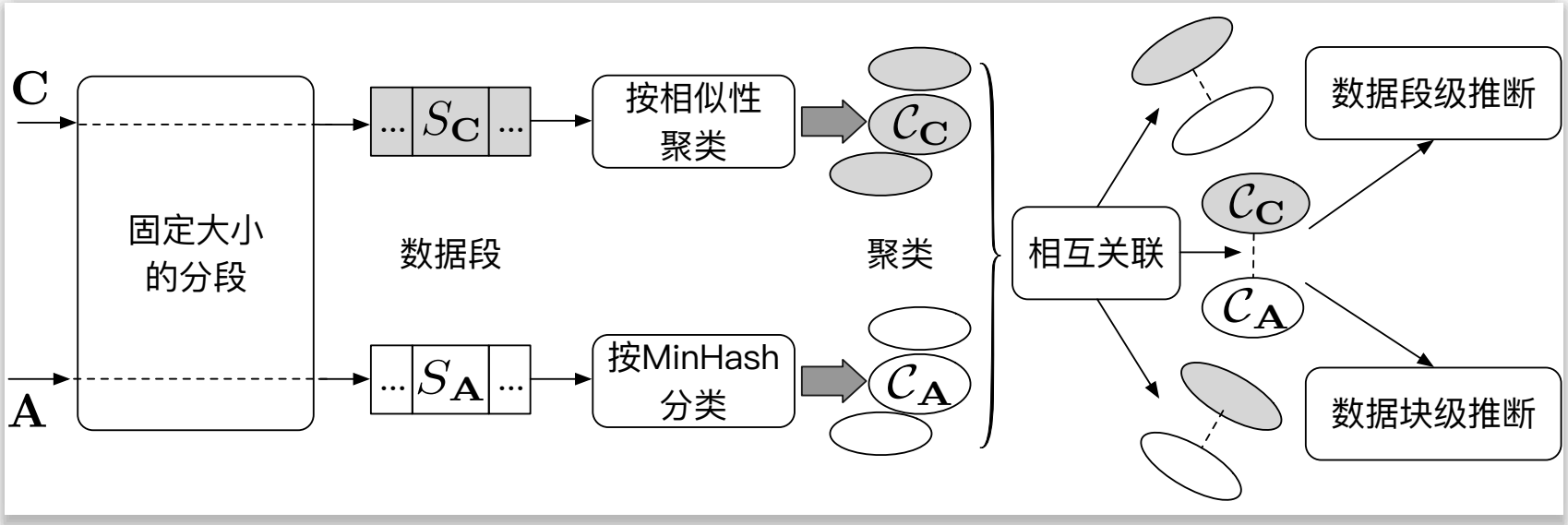
已有工作提出的基于数据块局部性的攻击方案

基于分布的攻击



设计的基于分布的攻击的攻击方案

基于聚类的攻击

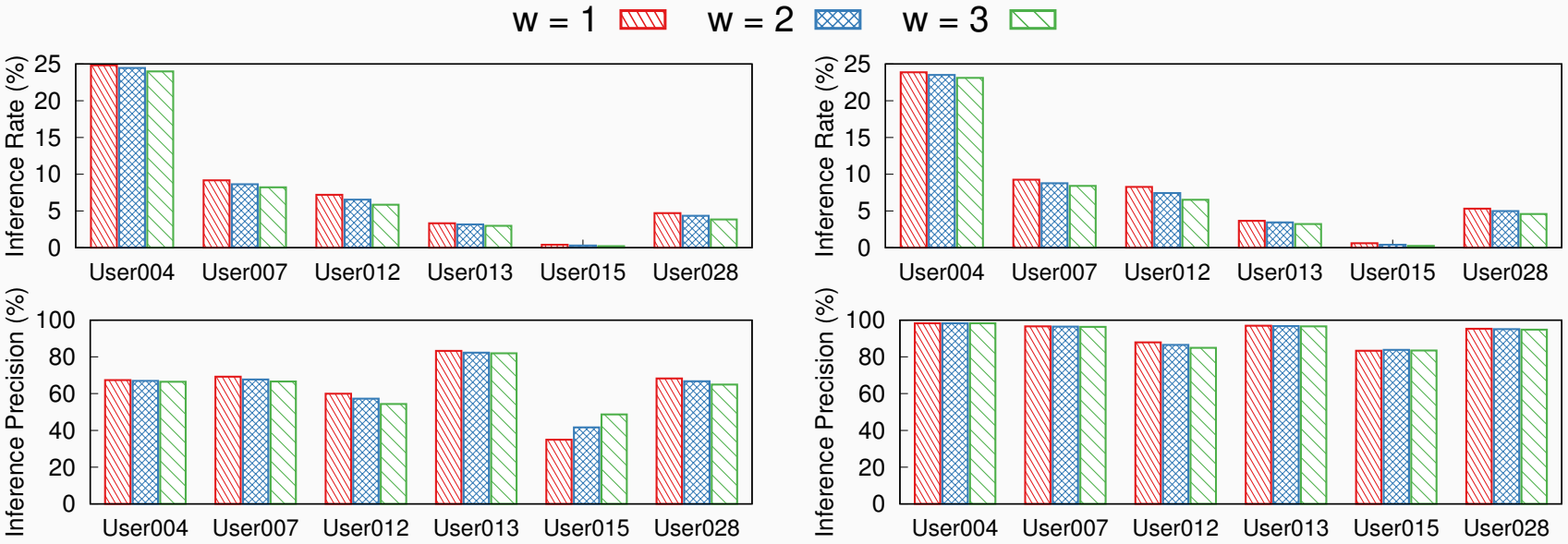


设计的基于聚类的攻击方案

研究结果与分析

- 基于分布的攻击-部分结果
- 基于聚类的攻击-部分结果

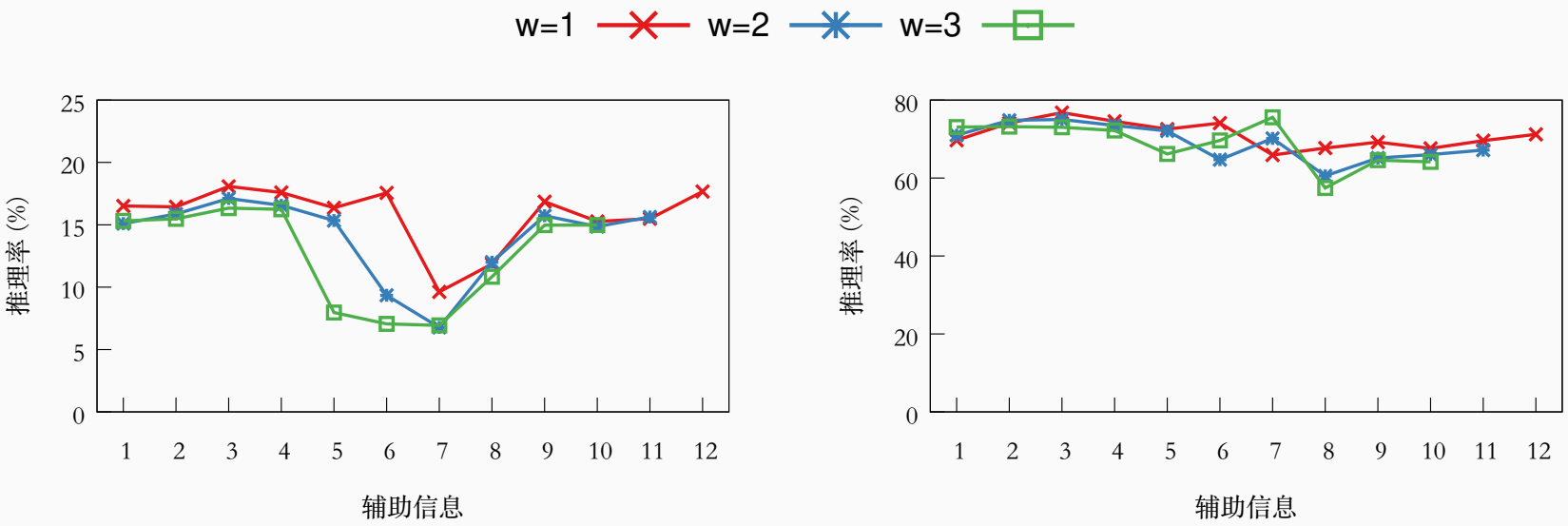
基于分布的攻击结果



设计的基于分布的攻击方案在FSL数据集中的攻击效果

（左无数据块大小信息辅助，右有数据块大小信息辅助）

基于聚类的攻击结果



设计的基于聚类的攻击方案在VM数据集中的攻击效果。

总结建议与参考文献

- 关于研究的总结
- 针对本文提出的攻击的建议
- 参考文献

关于研究的总结

加密重复数据删除应用确定性加密，并由此泄漏了明文的频率。研究重新审视了频率分析引起的安全漏洞，并证明加密重复数据删除更容易受到推理攻击。

研究提出了两种新的频率分析攻击方法，它们在攻击者所具有的条件的不同假设下都能实现高推理率和高推理精度。

关于研究的总结

利用三个真实世界的数据集来验证评估这两种攻击方法，提出关于其性质的各种新观察，并进一步分析它们如何带来实际性的损害。

研究还讨论了加密重复数据删除应对频率分析攻击的可能的对策及其相应的优缺点，以建议从业者安全地实现和部署加密重复数据删除存储系统。

针对本文提出的攻击的建议

- **防止频率泄漏：** MinHash加密、加入冗余数据块。
- **防止顺序泄漏：** 添加扰动。
- **防止大小泄漏：** 数据填充、固定大小分块。

参考文献

- M. Bellare, S. Keelveedhi, T. Ristenpart. Message-locked encryption and secure deduplication[C]. Annual International Conference on the Theory and Applications of Cryptographic Techniques, 2013, 296-312
- M. Naveed, S. Kamara, C. V. Wright. Inference attacks on property-preserving encrypted databases[C]. Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 2015, 644-655
- J. Li, C. Qin, P. P. Lee, et al. Rekeying for encrypted deduplication storage[C]. Dependable Systems and Networks (DSN), 2016 46th Annual IEEE/IFIP International Conference on, 2016, 618-629
- ...