Dynamic Searchable Symmetric Encryption 阅读笔记

李忆诺

2024 年 3 月 12 日

1 基本信息

1.1 论文来源

Kamara, S.; Papamanthou, C.; Roeder, T.: Dynamic searchable symmetric encryption. pp. 965–976. In: Proceedings of the 2012 ACM Conference on Computer and Communications Security, (2012).

1.2 概述

本文基于倒排索引数据结构的 SSE-1 结构，构建了一种动态可搜索对称加密方案，其安全性是 CKA2-secure 的，在 random oracle 模型中是安全的。

2 论文要点

2.1 背景

在现有的 SSE 方案中，虽然有几种 CKA2-secure 的 SSE 方案，但其局限在于：

- 时间复杂度高，搜索耗时较长；
- 加密索引太大；
- 不是明确动态的。

因此，本文提出第一个 SSE 方案来改进上述所有问题。

2.2 价值

本文提出了第一个 SSE 方案，其满足：次线性搜索时间、抵御自适应选择关键字攻击的安全性、紧凑索引以及高效添加和删除文件的能力。

本文的主要贡献有：

- 提出了动态 SSE 的形式化安全定义。特别是，本文的定义抓住了 SSE 安全的一个强有力的概念，即针对选择关键字攻击的适应性安全 (CKA2)；

- 构造了第一个 SSE 方案，它是动态的，CKA2-secure 的，并且实现了最优搜索时间，除此之外，本文的构造在 random oracle 模型中是安全的；
- 描述了基于倒排索引方法的 SSE 方案的第一个实现和评估。实现表明，这种类型的 SSE 方案非常有效；
- 对本文的方案进行了性能评估，显示为可搜索云存储系统增加机密性的增量成本。

2.3 问题陈述

本文提出的方案是基于倒排索引数据结构的 SSE-1 结构的扩展，旨在解决两个问题：

- SSE-1 只能抵抗非自适应选择关键字攻击 (CKA1)，这意味着它只能为批量执行搜索的客户端提供安全性；
- SSE-1 不是显式动态的，即它只能支持使用通用和低效技术的动态操作。

2.4 方法

2.4.1 SSE-1 的构建

加密

假设需要加密的文件集合为 f，该方案为每个关键字 $w \in W$ 构造一个列表 L_W，每个列表 L_W 由 f_W 个节点组成，某一节点 $N_i =< id, addr_s(N_{i+1}) >$，其中 id 是包含关键字 W 的唯一文件标识符，$addr_s(N)$ 表示在范围 A_s 中的节点 N 的位置。

对于每个关键字 w，将指向 L_w 头部的指针添加到搜索键 $F_{K_1}(w)$ 下的搜索表 T_s 中，然后在生成的密钥 $G_{K_2}(w)$ 下使用私钥加密方案 SKE 对每个列表进行加密。

搜索

客户端发送 $F_{K_1}(w)$ 和 $G_{K_2}(w)$，服务器根据 $F_{K_1}(w)$ 恢复指向 L_w 头部的指针，并使用 $G_{K_2}(w)$ 来解密列表并恢复包含 w 的文件的标识符。

2.4.2 SSE-1 的优化

文件删除

添加一个额外的（加密的）数据结构 A_d 作为删除数组，服务器可以通过客户端提供的令牌进行查询，以恢复指向被删除文件对应节点的指针。

指针修改

使用同态加密方案对存储在节点中的指针进行加密，通过向服务器提供适当值的加密，它就可以修改指针，而不需要解密节点。本文使用标准私钥加密方案，该方案包括将消息与伪随机函数输出进行异或。

这种构造还具有不提交的优点，可用来实现 CKA2-secure

内存管理

跟踪 A_s 中哪些位置是空闲的，以便使用空闲列表来添加新节点。

2.5 结果

2.5.1 实验环境

- CPU: Intel Xeon CPU 2.26 GHz (L5520) 单线程
- 方案实现：Microsoft Cryptography API: Next Generation (CNG) 上使用 c++ 实现

为了将加密成本从系统成本中分离出来，本文构建了一个测试框架，该框架在一组文件上执行加密计算，但不会通过网络传输这些文件，也不会产生从磁盘存储和检索索引信息的成本；所有操作都在内存中执行。除此之外，本文还忽略了为文件生成纯文本索引的成本。

数据集：Enron emails、Microsoft Office documents、media files

Micro-benchmarks 结果：无论从搜索中返回的文件数量如何，搜索令牌生成需要恒定的时间 (平均 35µs)。结果表明，客户端的搜索和文件增删是高效实用的，即使是对于常见的单词，或包含许多独特单词的文件。

Full performance 结果：对于大型数据集，SSE 索引生成性能与文件/词对的数量呈线性关系，则建立了初始索引后，后续的添加和删除操作效率很高，动态效果好。

2.5.2 安全性

本文提供了一个表述和比较 SSE 方案泄露的框架，基于该框架，将本方案分别与静态的 SSE-1 方案与动态的 vSDHJ10 方案进行比较。结果表明，本方案在 random oracle 模型中对于泄露情况是 CKA2-secure 的，但本方案泄露的信息比 SSE-1 和 vSDHJ10 更多。

3 评论

3.1 局限性

本文设计的 SSE 方案泄露了更多信息，并且在其各种操作泄露的信息之间存在相关性，攻击者可以依赖于信息之间的相关性以及大量统计数据进行攻击。

3.2 扩展阅读

SSE-1 方案：

M. Chase and S. Kamara. Structured encryption and controlled disclosure. In Proc. Int. Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT), pages 577–594, 2010.

CAK2-secure 的 SSE 方案

R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky. Searchable symmetric encryption: Improved definitions and efficient constructions. In Proc. ACM Conference on Computer and Communications Security (CCS), pages 79–88, 2006.

动态 SSE 方案

P. van Liesdonk, S. Sedghi, J. Doumen, P. H. Hartel, and W. Jonker. Computationally efficient searchable symmetric encryption. In Proc. Workshop on Secure Data Management (SDM), pages 87–100, 2010.

3.3 启示

通过阅读该文献，我了解到动态 SSE 的形式化安全定义，从 CAK1-secure 到 CAK2-secure，也学习到实际的 SSE 方案应满足的四个特性：次线性搜索时间、抵御自适应选择关键字攻击的安全性、紧凑索引以及高效添加和删除文件的能力。除此之外，我学习到该方案所描述的动态可搜索加密构造，以及如何证明其安全性和性能。

在阅读该文献的过程中，我经常被许多从未了解过的 SSE 方案构造所困扰，这深刻说明目前的我对于 SSE 方案及其各种优化改进还不够熟悉，需要在后续的学习过程中，补充更多的相关文献，多进行论文调研工作，来弥补现阶段的不足。