Poster: Normal Assumption and Less Injection Lead to Higher Recovery: File Injection Attack Evolves Step by Step

Ruizhong Du
Hebei University
durz@hbu.edu.cn

Zhendong Zhang*
Hebei University
zhangzd52@outlook.com

Mingyue Li*
Hebei University
limingyue@hbu.edu.cn

Chunfu Jia
Nankai University
cfjia@nankai.edu.cn

***Abstract*—Searchable Symmetric Encryption (SSE) enables queries over encrypted databases while concealing both data contents and query keywords. Although information leakage in SSE schemes is widely recognized as inevitable, the increasing strength of SSE defenses has, to some extent, compelled attackers to adopt attacks with weaker assumptions and more relaxed condition requirements. We propose two file injection attacks, KVA and KFA. They do not rely on the stronger sub-pattern (rsp) of the volume pattern and incur a low injection volume. They are realized through a novel grouping strategy and a n-ary injected document construction method, respectively. The former trades an increased number of injected documents for a substantial reduction in injection volume, while the latter enables KFA to achieve a high recovery accuracy and inherently provides robustness against the threshold countermeasure.**

I. Introduction

Since the introduction of Searchable Symmetric Encryption (SSE), numerous attacks against SSE systems have been proposed [1], [2]. File injection attacks constitute an active attack class, where adversaries inject crafted documents into the database and infer encrypted queries by observing changes in query responses before and after injection. Prior studies aim to improve recovery accuracy while reducing injection cost under acceptable adversarial assumptions.

File injection attacks primarily exploit volume pattern leakage, including the response length pattern (rlp), response size pattern (rsp), and response total size pattern (rtp). Assuming access to rsp is relatively strong and less practical, as rsp is easier for SSE schemes to conceal, and attacks relying on rsp [1] are therefore limited. Our work targets high-accuracy, low-injection file injection attacks without relying on rsp.

We observe two implicit design constraints in prior work on injected document construction. First, existing attacks typically enforce rlp and rtp effects simultaneously within a single injection step, which significantly restricts design flexibility and increases injection volume. Second, most attacks adopt a binary encoding scheme for keyword assignment, which underutilizes the available design space. We show that higher-radix encodings can substantially improve both recovery accuracy and robustness against defenses.

We propose KVA, a new file injection attack that introduces a grouping strategy to partition the keyword space and independently construct rlp- and rtp-oriented injected document sets, thereby decoupling their effects. This design enables an exponential reduction in injection volume while improving recovery accuracy. Building on KVA, we further propose KFA, which revisits keyword encoding from a volume-distribution perspective and employs an n-ary injected document construction scheme, achieving higher accuracy and inherent robustness against the threshold countermeasure.

II. K-Fold volume attack

We propose KVA, which applies a grouping strategy to decouple rlp injection from rtp injection, thereby achieving an exponential reduction in injection size. The core principle of it involves: (1) partitioning the leaked keyword set uniformly into K subsets, denoted as $k_t = \{w_{t,1}, w_{t,2}, \ldots, w_{t,|W|/k}\}$, where $|W|$ denotes the number of leaked keywords; (2) designing two complementary injected file sets F_α and F_β to optimize injections. Specifically, F_α ensures distinct injection sizes for any two keywords within the same subset while maintaining identical injection sizes for keywords with matching indices across subsets. Conversely, F_β enforces uniform injection sizes for all keywords but requires differing injection lengths for keywords sharing the same index. This method achieves a flexible trade-off between rlp/rtp variation and recovery accuracy by adjusting the grouping parameter K. Even with a relatively large K value (e.g., K = 10), it only incurs a minor loss in recovery accuracy.

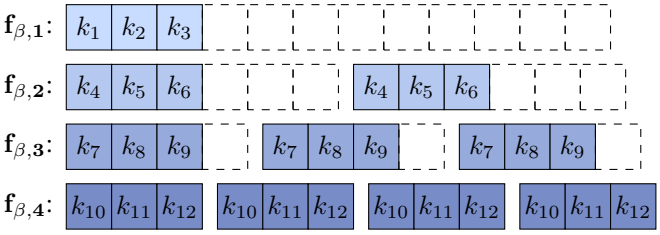


Fig. 1. Illustration of β-injection

Let $f_{\alpha,i}$ be an injected document in the α section. Convert the subscripts of each keyword to binary, and if a subscript n satisfies that the ith bit of its binary is 1, then the injected document $f_{\alpha,i}$ will contain the keywords $\{w_{1,n}, w_{2,n}, \ldots, w_{k,n}\}$.Each keyword appears in only one injection document in the α part, and keywords with the

same subscripts in different subsets are included in the same injection document. The β part, illustrated in Figure 1 with $|W|$ = 12 and $K = 4$, requires the attacker to construct a set of $K = 4$ injected documents $f_{\beta,1}, f_{\beta,2}, f_{\beta,3}, f_{\beta,4}$, each document contains all keywords from a corresponding subset. The injection of β part will not make a difference in the difference in rtp between different keywords.

In the recovery phase of KVA, since the injection phase induces volume characteristics identical to those produced by conventional injection algorithms, the specialized construction of injected documents does not affect KVA's ability to employ standard recovery procedures. KVA compares the volume differences of each query before and after injection with the expected keyword-specific volume changes by design, and recovers a query as the corresponding keyword only if it simultaneously matches the expected rlp and rtp variations, in which case it is added to the recovery result set.

III. Keyword formatted attack

We propose KFA, which aims to further explore the limited yet underutilized design space within the largely fixed framework of injected document construction in existing file injection attacks, in order to achieve improved attack performance. The design of KFA is motivated by the following two observations:

- During injected document construction, the traversal order of keywords directly determines their assigned injection volumes. By optimizing this ordering, an attacker can prioritize keyword recovery and improve overall recovery accuracy.
- While the binary construction method adopted by existing attacks is simple and effective, this does not imply that higher-radix constructions are unnecessary. In fact, higher-radix designs provide advantages in terms of recovery accuracy and robustness against defensive countermeasures.

k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9		Injected files			
1	2	0	1	2	0	1	2	0	→	$f_{1,1}$	k_1, k_4, k_7	$f_{1,2}$	k_2, k_5, k_8
0	0	1	1	1	2	2	2	0	→	$f_{2,1}$	k_3, k_4, k_5	$f_{2,2}$	k_6, k_7, k_8
0	0	0	0	0	0	0	0	1	→	$f_{3,1}$	k_9	$f_{3,2}$	-

Fig. 2. Injection file construction in KFA with n = 3

Corresponding to the first observation, we propose an optimization strategy as follows. During the construction phase, the attacker first estimates the rtp of each keyword from prior data and sorts all keywords in ascending order of their volume, yielding an ordered keyword set. In the binary injection process, keywords with smaller volumes are assigned lower injection costs, whereas keywords with larger volumes are assigned higher injection costs. This strategy assigns higher recovery priority to high-rtp keywords in the recovery phase, thereby improving overall recovery accuracy. Its generality enables broad applicability to volumetric injection attacks, improving efficiency and informing future research.

Corresponding to the second observation, KFA extends the injected document construction method to an n-ary scheme. Specifically, each keyword is assigned a fixed-length n-ary identifier of length k, where k satisfies $n^k \geq |W|$. The attacker constructs k groups of injected documents, each corresponding to one digit of the identifier. Each group contains n-1 documents labeled from 1 to n-1, and the i-th document in a group contains all keywords whose corresponding digit equals i.

By increasing n, KFA significantly reduces the size of individual injected documents, allowing them to fall below the threshold enforced by SSE systems. As a result, the attack no longer relies on additional threshold-evasion techniques. Meanwhile, at the cost of a modest increase in overall injection volume, KFA effectively improves recovery accuracy.

IV. Evaluation

To simulate realistic attack settings, we evaluate our attacks on three real-world datasets. We use the Enron email corpus from 2000-2002, which contains 30,109 emails, and the Lucene mailing list corpus from 2001-2020, comprising 113,201 emails. In addition, we adopt a Wikipedia dump archived on March 1, 2022, which includes 100,000 keywords and 6,358,670 documents. On all three datasets, KVA and KFA significantly outperform BVA and BVMA [2], which also do not rely on the rsp pattern, while achieving performance comparable to the Decoding attack that incurs several orders of magnitude higher injection volume.

V. Conclusion

Our work presents two novel file injection attacks, KVA and KFA, against Searchable Symmetric Encryption. By decoupling injection patterns and employing an n-ary construction, they achieve high query recovery accuracy without relying on strong leakage assumptions and with low injection volume. Evaluations on real-world datasets confirm their superior performance over prior attacks. This demonstrates that even under relaxed conditions, SSE systems remain vulnerable to carefully crafted injection strategies, underscoring the need for stronger defenses in searchable encryption.

Acknowledgment

This paper was supported by the Science ResearchProject ofHebeiEducation Department(QN2024204); the Interdisciplinary Research Program of Hebei University (DXK202501); the Key Research and Development Program of Hebei Province of China (22340701D), and Beijing-Tianjin-Hebei Basic Research Collaborative Special Project (F2024201070).

References

[1] L. Zhang, J. Wang, J. Wu, Y. Wang, and S.-F. Sun, "Violin: Powerful volumetric injection attack against searchable encryption with optimal injection size," *IEEE Transactions on Dependable and Secure Computing*, vol. 22, pp. 4103–4115, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:276474872

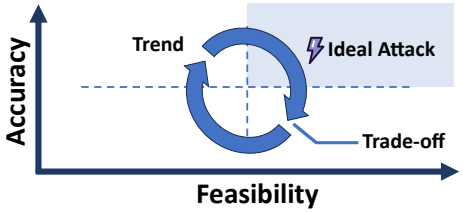
[2] X. Zhang, W. Wang, P. Xu, L. T. Yang, and K. Liang, "High recovery with fewer injections: Practical binary volumetric injection attacks against dynamic searchable encryption," in *USENIX Security Symposium*, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:256827774

Poster: Normal Assumption and Less Injection Lead to Higher Recovery: File Injection Attack Evolves Step by Step

Ruizhong Du[1], Zhendong Zhang*[1], Mingyue Li*[1], Chunfu Jia[2]
[1]Hebei University, Baoding, China, [2]Nankai University, Tianjin, China

Introduction and Key Observing

- The increasing strength of SSE defenses has, to some extent, compelled attackers to adopt attacks with weaker assumptions and more relaxed condition requirements.



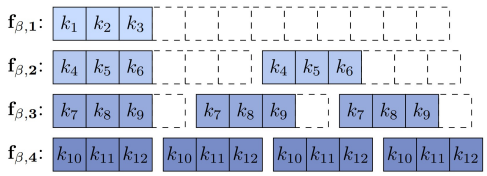
- **We observe two implicit design constraints in prior work on injected document construction:**
  - Existing attacks typically enforce rlp and rtp effects simultaneously within a single injection step, which significantly restricts design flexibility and increases injection volume.
  - Most attacks adopt a binary encoding scheme for keyword assignment, which underutilizes the available design space. We show that higher-radix encodings can substantially improve both recovery accuracy and robustness against defenses.

Powerful File Injection Attacks with Low Assumption

K-Fold Volume Attack

- **(1) Partitioning the leaked keyword set uniformly into K subsets, denoted as $k_t = \{w_{t,1}, w_{t,2}, \ldots, w_{t,|w|/k}\}$.**
- **(2) Designing two complementary injected file sets F_α and F_β to optimize injections.**

- F_α ensures distinct injection sizes for any two keywords within the same subset while maintaining identical injection sizes for keywords with matching indices across subsets.
- F_β enforces uniform injection sizes for all keywords but requires differing injection lengths for keywords sharing the same index.



- This method achieves a flexible trade-off between rlp/rtp variation and recovery accuracy by adjusting the grouping parameter K. Even with a relatively large K value (e.g., K = 10), it only incurs a minor loss in recovery accuracy.

K-Fold Formatted Attack

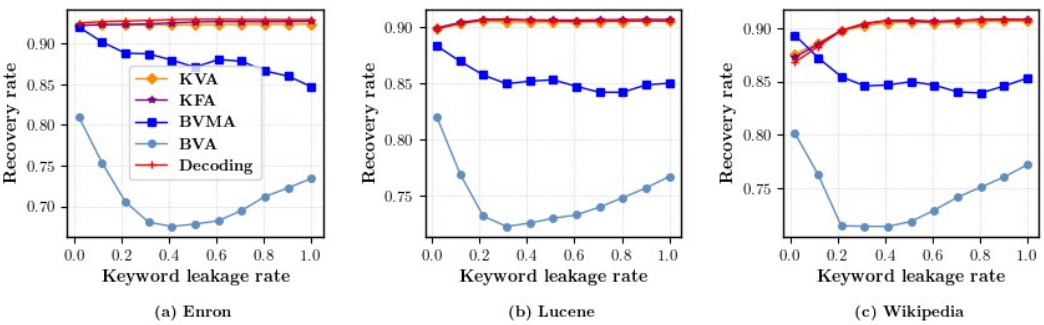
- **During the construction phase, the attacker**:
  - Estimates the rtp of each keyword from prior data.
  - Sorts all keywords in ascending order of their volume, yielding an ordered keyword set.
- In the binary injection process, keywords with smaller volumes are assigned lower injection costs, whereas keywords with larger volumes are assigned higher injection costs.

- KFA extends injected document construction to an n-ary scheme by assigning each keyword a fixed-length n-ary identifier of length k, where $n^k \geq |W|$. The attacker builds k document groups, each corresponding to one identifier digit, and each group contains n-1 documents labeled 1 to n-1, where the i-th document includes all keywords whose digit equals i.

k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9		Injected files			
1	2	0	1	2	0	1	2	0	→	$f_{1,1}$	k_1, k_4, k_7	$f_{1,2}$	k_2, k_5, k_8
0	0	1	1	1	2	2	2	0	→	$f_{2,1}$	k_3, k_4, k_5	$f_{2,2}$	k_6, k_7, k_8
0	0	0	0	0	0	0	0	1	→	$f_{3,1}$	k_9	$f_{3,2}$	-

Evaluation and Conclusion

- To simulate realistic attack settings, we evaluate our attacks on three real-world datasets, including Enron, Lucene and Wikipedia. Across all datasets, KVA and KFA significantly outperform BVA and BVMA, which also do not rely on the rsp pattern, while achieving performance comparable to the Decoding attack despite requiring several orders of magnitude less injection volume.



Conclusion

- Our work presents two novel file injection attacks, KVA and KFA, against Searchable Symmetric Encryption. By decoupling injection patterns and employing an n-ary construction, they achieve high query recovery accuracy without relying on strong leakage assumptions and with low injection volume. Evaluations on real-world datasets confirm their superior performance over prior attacks. This demonstrates that even under relaxed conditions, SSE systems remain vulnerable to carefully crafted injection strategies, underscoring the need for stronger defenses in searchable encryption.