Estimating Confidentiality of Encrypted Search Systems: A Probabilistic Framework

Author Names

January 24, 2026

Abstract

Encrypted search systems enable privacy-preserving queries over confidential data stored on untrusted servers. However, the confidentiality of such systems may be compromised through frequency analysis attacks. We develop a probabilistic framework for quantifying the confidentiality of encrypted search systems based on the sampling distribution of a confidentiality statistic. Using the Bootstrap method with normal approximation, we efficiently estimate the risk of confidentiality breach, enabling proactive countermeasures. Our analysis shows how entropy of the query distribution relates to the adversary's expected accuracy, providing theoretical grounding for resilience engineering approaches to encrypted search security.

1 Introduction

With the advent of *cloud computing*, it is tempting to store our confidential data on remote (untrusted) systems like a cloud storage provider. However, a system administrator may be able to compromise the confidentiality of our data which threatens to prevent further adoption of cloud computing and electronic information retrieval in general if the threat cannot be mitigatedSubashini and Kavitha [2011], Zissis and Lekkas [2012], Claycomb and Nicoll [2012].

The primary challenge is a trade-off problem between confidentiality and usability of the data stored on remote untrusted systems. *Encrypted Search* attempts to resolve this trade-off problem.

Definition 1.1. *Encrypted Search* allows authorized search agents to investigate presence of specific search terms in a confidential target data set, such as a database of encrypted documentsBoneh et al. [2004], Jin Li et al. [2010], Cao et al. [2014], Sun et al. [2013], Kamara and Lauter [2010], while the contents, especially the meaning of the target data set and search terms, are hidden from any unauthorized personnel, including the system administrators of a cloud server.

Essentially, *Encrypted Search* enables *oblivous search*. For instance, a user may search a confidential database stored on an untrusted remote system without other parties being able to determine what the user searched for. We denote any untrusted party that has full access to the untrusted remote system (where the confidential data is stored) as an adversary.[1]

Despite the potential of *Encrypted Search*, *perfect* confidentiality is not theoretically possible. There are many ways confidentiality may be compromised. In this paper, we consider an adversary whose primary objective is to comprehend the confidential information needs of the search agents by analyzing their history of *Encrypted Search* queries.

A simple measure of confidentiality is given by the proportion of queries the adversary is able to comprehend. We consider an adversary that employs a known-plaintext attack. However, since

[1] A system administrator being a typical example.

the confidentiality is a function of the history of queries, different histories will result in different levels of confidentiality. We apply the Bootstrap method to estimate the sampling distribution of the confidentiality. The sampling distribution provides the probabilistic framework to resolve security-related questions such as "what is the probability that the confidentiality is less than 70%?"

The rest of this paper is organized as follows. Section 2 reviews existing work in *Encrypted Search*. There has not been much work that quantitatively analyzes the conditions for information leaks by frequency attacks, such as the number of encrypted words an adversary needs to observe for a certain accuracy and how likely it happens. Section 5 introduces the *moving average bootstrap* (MAB) method, an efficient estimator of the achievable accuracy by the adversary using frequency attacks. Section 7 presents our performance evaluations on the MAB method. Section 9 summarizes our contributions and planned future work, followed by the selected references.

2 Related work

Boneh proposed a method to enable untrusted systems the ability to perform searches over encrypted e-mail messages, called public-key encryption with keyword search (PEKS) Boneh et al. [2004]. Boneh designed PEKS in such a way that e-mail messages are encrypted by the public key of an e-mail receiver, while a third party, such as an e-mail server, to perform search for a particular word (e.g., "urgent") in each encrypted message without all the raw contents in the encrypted e-mail exposed to the third party. The core of this method is trapdoors, which are a hash value of a given word in e-mails. Each e-mail receiver creates trapdoors, one for each target word and trapdoors are included in each encrypted e-mail message for searches on the encrypted e-mail messages.

Li extended this concept to allow untrusted systems to perform encrypted searches that allow approximate matching by enumerating multiple trapdoors, one for each expected deviationJin Li et al. [2010], Cao et al. [2014], Sun et al. [2013], Kamara and Lauter [2010] proposed to apply encrypted search to enhancing security in cloud computing.

Despite the potentials in the encrypted search schemes, risk of information leaks through guessing the searched words has been identifiedByun et al. [2006], Yau et al. [2008], Jeong et al. [2009]. It has been demonstratedByun et al. [2006], Yau et al. [2008], Jeong et al. [2009] that anyone who has access to encrypted data possibly map them to their plain text counterparts.

Use of secure communication channels (e.g., SSL) will be effective in hiding the trapdoors in queries submitted by legitimate users from external adversaries, but use of secure communication channels still can not prevent frequency attacks from internal adversaries, such as malicious administrators, assuming that they can intercept trapdoors within a local host computer, by installing illegal capturing tool or by tampering executables.

Despite the threat from frequency attacks, there has not been much work that delves into quantified analyses on the conditions for when such information leaks exceed a tolerable risk level under various conditions. Rivain proposed a multivariate Gaussian random variable method to estimate the success rate in discovering secret keys under side-channel attacksRivain [2009]. Rivain proposed use of "confidence" for evaluating the effectiveness in side-channel attacksThillard et al. [2013]. Rivain and Thillard both proposed a solution against correlation attacks, but not against frequency attacks. Correlation attacks are different from frequency attacks in that the adversary discovers the encryption keys to deduce the plaintext in the former, while the latter induces the plaintext directly from the observed trapdoors without discovering their encryption keys.

3 ***Encrypted Search*** model

An information retrieval process begins when a *search agent* submits a *query* to an information system, where a query represents an *information need*. In response, the information system returns a set of relevant objects, such as *documents*, that satisfy the information need.

An *Encrypted Search* system may support many different kinds of queries, but we make the following simplifying assumption.

Assumption 3.1. The query model is a *sequence-of-words*.

The *adversary* is given by the following definition.

Definition 3.2. The adversary is an untrusted agent that is able to observe the sequence of queries and corresponding search results submitted by authorized search agent. The objective of the *Encrypted Search* system is to prevent the adversary from being able to comprehend the meaning of the queries or the underlying plaintext data.

A query submitted to an *Encrypted Search* system should not be comprehensible to the adversary.

Definition 3.3. A *hidden query* represents a confidential *information need* of an authorized search agent that is suppose to be incomprehensible to the adversary.

The primary means by which *Encrypted Search* is enabled is by the use of cryptographic *trapdoors* as given by the following definition.

Definition 3.4 (Trapdoor)**.** Search agents map *plaintext* search keys to some cryptographic hash, denoted trapdoors.

A trapdoor for a *plaintext* search key is necessary to allow an *untrusted Encrypted Search* system to look for the key in a corresponding confidential data set.

Assumption 3.5. The *Encrypted Search* system uses a substitution cipher in which each search key in a *plaintext* query is mapped to a unique trapdoor signature. The substitution cipher is denoted by

$$\mathrm{h} \colon \mathbb{X} \mapsto \mathbb{Y}\,, \tag{1}$$

where $\mathbb{X}$ is the set of *plaintext* search keys and $\mathbb{Y}$ is the set of *trapdoors*.

The most straightforward substitution cipher is a *simple substitution cipher* where each *atomic* plaintext search key maps to a single trapdoor as illustrated by Algorithm 1.

Algorithm 1: Simple substitution cipher

Input: $\vec{x}$ is a *plaintext* query.
Output: $\vec{y}$ is the corresponding *hidden query*.
function HiddenQueryGenerator($\vec{x}$)
  $\vec{y} \leftarrow \{\mathrm{h}(x) \colon x \in \vec{x}\}$;
  return $\vec{y}$;

Given a plaintext key $x \in \mathbb{X}$, $\mathrm{h}(x)$ is a random variable whose support is a subset of the trapdoors in $\mathbb{Y}$. Given any plaintext keys $x, x' \in \mathbb{X}$, $x \neq x'$, the supports of $\mathrm{h}(x)$ and $\mathrm{h}(x')$ are disjoint. This makes it possible to *undo* the substitution cipher by some function denoted by

$$\mathrm{g}^* \colon \mathbb{Y} \mapsto \mathbb{X} \tag{2}$$

such that

$$x = \mathrm{g}^*(\mathrm{h}(x))$$

for every $x \in \mathbb{X}$. Thus, given a trapdoor $y \in \mathbb{Y}$, the corresponding plaintext key is given uniquely by $\mathrm{g}^*(y) \in \mathbb{X}$. If h is a *simple substitution cipher* where each plaintext key maps to a single trapdoor, then h is a function and g^* is its inverse denoted by h^{-1}.

Definition 3.6. A *hidden query* time series of size p is a sequence of p hidden queries given by

$$(\vec{y_1}, \ldots, \vec{y_p}) \ , \tag{3}$$

where $\vec{y_j}$ is given by

$$\vec{y_j} = \mathrm{HiddenQueryGenerator}(\vec{x_j}) \tag{4}$$

for $j = 1, \ldots, p$ and $\vec{x_1}, \ldots, \vec{x_p}$ is a time series of n *plaintext* queries submitted by authorized search agents.

Assumption 3.7. The adversary may only observe the *hidden query* time series to estimate the *plaintext query* time series.

We denote the p components of the j-thtrapdoor $\vec{y_j}$ by

$$y_{j1}, \ldots, y_{j j_p} \ ,$$

and thus given a *hidden query* time series

$$(\vec{y_1}, \vec{y_2}, \ldots, \vec{y_p}) \ , \tag{5}$$

we may represent it by the time series given by

$$\left(y_{1\,1}, \ldots, y_{1\,j_1}, q, y_{2\,1}, \ldots, y_{2\,j_2}, q, \ldots, y_{p\,1}, \ldots, y_{p\,j_p}, q\right) \ , \tag{6}$$

where q denotes the *end-of-vector* token.

We denote a time series of such trapdoors by the following definition.

Definition 3.8. A time series of n trapdoors is denoted by

$$\vec{\tau_n} = (y_1, \ldots, y_n) \ , \tag{7}$$

where

$$y_j = \mathrm{h}(x_j) \tag{8}$$

for $j = 1, \ldots, n$ and $(x_1, \ldots, x_n)$ is the corresponding *plaintext* time series.

3.1 Probabilistic model

The two primary sources of information are given by the (unobserved) time series of plaintext which induces the (observable) time series of trapdoors. Other potential sources of information are ignored, such as the time a *hidden query* is submitted.[2]

Since the time series of *plaintext* is uncertain, we model it as a sequence of random variables.

[2]The time series τ_n is just a logical time with the only constraint being that y_j was submitted before or at the same time as $y_{j+1}, \ldots, y_n$.

Definition 3.9. The j-thrandom *plaintext* search key, denoted by $\mathbf{X_j}$, in a time series of size n has a conditional probability given by

$$\Pr[\mathbf{X_j} = x_j \mid \mathbf{X_1} = x_1, \ldots, \mathbf{X_{j-1}} = x_{j-1}] \tag{9}$$

for $j = 1, \ldots, n$ and one of the *keys* is special and denotes *end-of-query*.

The time series of trapdoors is a function of the *plaintext* time series.

Definition 3.10. The uncertain j-thtrapdoor is a random variable given by

$$\mathbf{Y_j} = \mathrm{h}(\mathbf{X_j}) \ , \tag{10}$$

where the *end-of-query* key is not remapped by the substitution cipher h.

4 Threat model: *known-plaintext attack*

In this threat model, the adversary is interested in estimating the *plaintext* time series. However, the adversary is only able to observe the *trapdoor* time series. Thus, the adversary's objective is to infer the *plaintext* from the *trapdoors* using frequency analysis attacks,[3] in particular a *known-plaintext attack*.

In a known-plaintext attack, the adversary the objective of the adversary is to learn how to *undo* the substitution cipher h as given by g^*.

Assumption 4.1. The mapping function g^* is not known to the adversary.

Assumption 4.2. The adversary is able to observe a time series of n trapdoors, i.e., a particular τ_n.

A maximum likelihood estimator of g^* is given by

$$\begin{aligned}\hat{\mathrm{g}} = \arg\max_{\mathrm{g} \in G} \Big\{ & \Pr[\mathbf{X_1} = \mathrm{g}(y_1)] \times \\ & \prod_{i=2}^{n} \Pr[\mathbf{X_i} = \mathrm{g}(y_i) \mid \mathbf{X_{i-1}} = \mathrm{g}(y_{i-1}), \ldots, \mathbf{X_1} = \mathrm{g}(y_1)] \Big\} \ ,\end{aligned} \tag{11}$$

where G is the set of all possible mapping functions from the set of trapdoors $\mathbb{Y}$ to the set of plaintext keys $\mathbb{X}$.

If two plaintext keys $x, x' \in \mathbb{X}, x \neq x'$, may be exchanged without changing the probability distribution of the time series, they are *indistinguishable* and the mapping function g^* necessarily has multiple maximum likelihood estimates (even after observing an infinite time series).[4] However, if some of the random variables are not exchangeable, then the *adversary* may learn *something* about the plaintext by observing the time series of trapdoors.

The greater the uniformity of the *true* distribution, the less accurate the maximum likelihood estimator of g^* is. At the limit of maximum uniformity, where every pair is exchangeable, the adversary can learn nothing about the plaintext by observing the time series. Natural languages have a high degree of non-uniformity and so the primary concern of the adversary is the divergence between the *true* distribution and the *known-plaintext* distribution.

[3] Also known as spectral analysis attacksRaymond [2001].

[4] The maximum likelihood estimator of the mapping function is not *consistent*.

Assumption 4.3. The optimal adversary knows the *true* plaintext distribution $\mathbf{X_1}, \ldots, \mathbf{X_n}$ (or a *known-plaintext* distribution that has a sufficiently small divergence from the *true* distribution).

The *known-plaintext* distribution may be used to solve an approximation to Eq. (11) as given by the following definition.

Definition 4.4. In a *known-plaintext attack*, the adversary substitutes the unknown true distribution with the known-plaintext distribution and solves Eq. (11) under this substituted distribution.

Sub-optimal adversaries A *suboptimal* adversary may have any of the following problems:

1. The distribution of the known-plaintext diverges from the true distribution to the extent that the maximum likelihood estimator is inconsistent. All things else being equal, the less divergence between the *true* distribution and the *known-plaintext* distribution, the better the estimator.

2. The space of mapping functions G may be too large or complex. Note that for a simple substitution cipher, the space of mapping functions G has $k!$ possible mapping functions, where k is the cardinality of the set of plaintext keys $\mathbb{X}$, but a solution to the maximum likelihood estimator may be found in logarithmic time.[5]

3. A simplified probabilistic language model is employed to simplify the problem of finding the maximum likelihood estimate, and thus some of the information in the time series is discarded.

According to Piantadosi, the marginal distribution of words in most documents (and queries) follow a Zipf distributionPiantadosi [2014], where the most frequent word occurs approximately proportional to k times as often as the k-thmost frequently occurring word.

If an adversary ignores correlations in the time series by modeling each time step as an independent and identically distributed random variable, then Eq. (11) simplifies to the trivially solvable

$$\hat{\mathrm{g}} = \arg\max_{\mathrm{g} \in G} \left\{ \prod_{i=1}^{n} \Pr[\mathbf{X} = \mathrm{g}(y_i)] \right\} \tag{12}$$

$$= \arg\max_{\mathrm{g} \in G} \left\{ \sum_{i=1}^{n} \log \Pr[\mathbf{X} = \mathrm{g}(y_i)] \right\} \tag{13}$$

where

$$\Pr[\mathbf{X} = x] = \frac{1}{n} \sum_{i=1}^{n} \Pr[\mathbf{X_i} = x]\,.$$

. If the true distribution is an independent and identically distributed time series, the adversary is optimal if a solution to Eq. (12) can be found.

5 Confidentiality statistic

We are interested in measuring the degree of confidentiality as given by the following definition.

[5] In this case, the desired mapping function $\hat{\mathrm{g}}$ maps the j-thmost frequently occurring *trapdoor* in the *trapdoor* time series to the j-thmost probable *plaintext* key under the *true* distribution.

Definition 5.1 (Confidentiality statistic)**.** Given a time series of n trapdoors

$$\vec{\tau_n} = (y_1, y_2 \ldots, y_t, \ldots, y_n) \ , \tag{14}$$

the confidentiality at time step t, $1 \leq t \leq n$, is given by

$$C_t = 1 - p_t \tag{15}$$

where p_t is the fraction of trapdoors in the first t time steps that the adversary successfully maps to *plaintext*. That is,

$$p_t = \frac{\delta}{t} \, , \tag{16}$$

where

$$\delta = \sum_{i=1}^{t} [\mathrm{g}^*(y_i) = \hat{\mathrm{g}}(y_i)] \ . \tag{17}$$

The following example illustrates the *confidentiality statistic.*

Example 5.2. Suppose the adversary is able to correctly map the set of trapdoors given by

$$\{a, b, c\} \tag{18}$$

to *plaintext* in a time series of 8 trapdoors given by

$$\vec{\tau_8} = (a, c, d, b, e, d, b, d) \ . \tag{19}$$

The adversary correctly maps $\delta = 3$ of the first $t = 4$ trapdoors in the time series. Thus, the confidentiality at time step 4 is given by

$$C_4 = 1 - \frac{3}{4} = 0.25 \, . \tag{20}$$

The adversary correctly maps $\delta = 4$ trapdoors in the total time series. Thus, the confidentiality at time step 8 is given by

$$C_8 = 1 - \frac{4}{8} = 0.5 \, . \tag{21}$$

According to this statistical measure, the confidentiality increased from time step 4 and 8, i.e., the adversary *comprehends* a smaller fraction of the time series at the later time step.

The confidentiality statistic is expected to converge to some asymptotic limit, i.e., as $t \to \infty$, the confidentiality $C_t \to c$, $0 \leq c \leq 1$. If the adversary employs a *known-plaintext attack* where the distribution of the *known-plaintext* is equivalent to the *true* distribution, then $c = 0$, i.e., the adversary eventually comprehends the entire time series.

5.1 Sampling distribution of *confidentiality statistic*

The *confidentiality statistic* is a function of a *random time series* $(\mathbf{Y_1}, \ldots, \mathbf{Y_n})$. Thus, it has a *sampling distribution.*

Definition 5.3. The sampling distribution of C_t is denoted by $\mathbf{C}_t$ for $t = 1, \ldots, n$ for a random time series of n trapdoors.

The sampling distribution quantifies everything there is to know about the statistic. For instance, the *sampling distribution* may be used to make claims like "there is a 1% chance the adversary comprehends 70% of the time series at time step t."

Generally, the sampling distribution is not known and thus must be estimated. If we estimate the generative model of the time series of trapdoors, we may use the Bootstrap methodKusharya [2012] to estimate the sampling distributions.

Several generative models may be used to produce synthetic time series for Bootstrap resampling:

1. *Marginal (unigram) model*: Each trapdoor is drawn independently from the marginal distribution $\Pr[\mathbf{Y} = y]$. This model is computationally simple but ignores sequential dependencies.

2. *Bigram model*: The probability of a trapdoor depends on the immediately preceding trapdoor, i.e., $\Pr[\mathbf{Y_j} = y \mid \mathbf{Y_{j-1}} = y']$. This captures first-order sequential patterns in query behavior.

3. *Higher-order Markov models*: The trapdoor distribution depends on a window of the last q trapdoors. These models capture longer-range dependencies at the cost of increased parameter estimation.

4. *Interpolation techniques*: For smaller sample sizes, smoothing methods interpolate between lower-order and higher-order models to mitigate data sparsity, e.g., Katz backoff or Kneser-Ney smoothing.

The choice of generative model affects the fidelity of the Bootstrap estimate. More sophisticated models better capture the true dynamics of query sequences but require more data for reliable parameter estimation.

In the Bootstrap method, we "resample" from the time series and compute the confidentiality C_t of the resample. If we do this m times, we generate a sample of m confidentiality statistics

$$C_t^{(1)}, \ldots, C_t^{(m)} \tag{22}$$

for $t = 1, \ldots, n$.

Given this sample, we may compute any statistic that is a function of the sample. For instance, the *expected* value of the confidentiality statistic at time step t, $\mathbb{E}[\mathbf{C}_t]$, may be estimated by the sample mean

$$\bar{C}_t = \frac{1}{t} \sum_{i=1}^{m} C_t^{(i)} \,. \tag{23}$$

Another estimator of the *expected* confidentiality is given by a *moving average* like *Exponential smoothing*. However, the Bootstrap sampling distribution makes it possible to compute many other statistics of interest.

The variance of the confidentiality statistic at time step t, $\mathrm{Var}[\mathbf{C}_t]$, is another parameter of potential interest[6] and may be estimated by the sample variance

$$s_{m-1}^2 = \frac{1}{m-1} \sum_{i=1}^{m} \left(C_t^{(i)} - \bar{C}_t \right)^2 \,. \tag{24}$$

If the variance is high at a time step t, the *expected* confidentiality at time step t is not very indicative of the confidentiality of any particular time series at time step t.

By the large sample approximation, the sampling distribution of C_t for $t = 1, \ldots, n$ is approximately normal as given by the following theorem.

[6] The fluctuations demonstrated by Fig. 1 indicate high variance.

Theorem 5.4. *The sampling distribution of C_t converges in distribution to the normal distribution with a mean $\bar{C}_t$ and a variance s^2_{m-1}, denoted by*

$$\mathbf{C}_t \xrightarrow{d} \mathcal{N}\left(\bar{C}_t, s^2_{m-1}\right) . \tag{25}$$

Proof. The confidentiality statistic given by Eq. (15) is a linear function of an average. Therefore, by the Central Limit Theorem, the sampling distribution of C_t converges in distribution to a normal distribution with a mean given by the sample mean and a variance given by the sample variance. □

6 Mapping *entropy* to *confidentiality*

The adversary described in Section 4 may efficiently compromise the confidentiality of a time series of trapdoors if a *simple substitution cipher* is employed as described in Section 3. However, the described adversary is not particularly sophisticated. For instance, a more sophisticated adversary incorporates the search patterns of specific search agents into the probability model described in Section 3.1.

The adversaries we worry about the most are probably more clever than us. Thus, it may be asking too much to simulate them so that a *reliable* confidentiality statistic can be produced. Matters are further complicated if a simple substitution cipher is not used, e.g., a homophonic encryption scheme is used to flatten the distribution of trapdoors. In this case, the confidentiality is expected to improve, but it may be difficult to quantify to what extent.

We may be able to construct a *lower-bound* on confidentiality that is a function of the *entropy*. The entropy of a random time series of t trapdoors is given by

$$\mathrm{H}(\mathbf{Y_1}, \ldots, \mathbf{Y_t}) \text{ bits} . \tag{26}$$

If the random time series is independently distributed, the entropy simplifies to $\mathrm{H}(\mathbf{Y_1}) + \cdots + \mathrm{H}(\mathbf{Y_t})$ and if it is also identically distributed is simplifies to $t\,\mathrm{H}(\mathbf{Y_1})$. Consider the following cases:

1. An *optimal* adversary is expected learn *nothing* about the mapping from *trapdoors* to *plaintext keys* by observing a uniformly distributed time series.[7] A uniformly distributed time series of over a support set of m unique trapdoor signatures has $\log_2 m$ bits/trapdoor of entropy.

2. An *optimal* adversary is expected to learn *everything* about the mapping from *trapdoors* to *plaintext keys* by observing a degenerate time series, which has zero entropy.

By Items 1 and 2, the entropy is bounded by

$$0 \leq \mathrm{H}(\mathbf{Y_1}, \ldots, \mathbf{Y_t}) \leq t \log_2 m . \tag{27}$$

We use these insights to construct an *information again* measure given by the following definition.

Definition 6.1. The *mean information gain* of a random time series $\mathbf{Y_1}, \ldots, \mathbf{Y_t}$ is defined to be the difference between the *maximum entropy* and the actual entropy as given by

$$\mu(t) = t \log_2 m - \mathrm{H}\left(\mathbf{Y_1}, \ldots, \mathbf{Y_t}\right) , \tag{28}$$

which is a real number between 0 and $t \log_2 m$.

[7]The optimal adversary *randomly* chooses a mapping.

If the random time series is independently and identically distributed, then the *mean information gain* is given by

$$\mu(t) = t\left(\log_2 m - \mathrm{H}\left(\mathbf{Y_1}\right)\right) \text{ bits}\,. \tag{29}$$

The rate of change of the *mean information gain* is given by the following theorem.

Theorem 6.2. *The rate of change of the* mean information gain *at time* t *is given by*

$$\lambda(t) = \log_2 m - \mathrm{H}\left(\mathbf{Y_t} \mid \mathbf{Y_{t-1}}, \ldots, \mathbf{Y_1}\right) \text{ bits/trapdoor}\,, \tag{30}$$

which is a real number between 0 *and* $\log_2 m$*.*

Proof. The rate of change at time t is the difference between the *mean information gain* at time steps t and $t-1$, which is given by

$$\begin{aligned}\lambda(t) &= \mu(t) - \mu(t-1) &(31)\\ &= \log_2 m - \mathrm{H}\left(\mathbf{Y_t}, \ldots, \mathbf{Y_1}\right) + \mathrm{H}\left(\mathbf{Y_{t-1}}, \ldots, \mathbf{Y_1}\right)\,. &(32)\end{aligned}$$

The joint entropy $\mathrm{H}(\mathbf{Y_1}, \ldots, \mathbf{Y_t})$ may be rewriten as

$$\mathrm{H}(\mathbf{Y_t} \mid \mathbf{Y_{t-1}}, \cdots, \mathbf{Y_1}) + \mathrm{H}(\mathbf{Y_{t-1}}, \cdots, \mathbf{Y_1})\,. \tag{33}$$

Performing this substitution results in the equivalent equality given by

$$\begin{aligned}\lambda(t) &= \log_2 m - \mathrm{H}\left(\mathbf{Y_t} \mid \mathbf{Y_{t-1}}, \ldots, \mathbf{Y_1}\right)\\ &\qquad + \mathrm{H}\left(\mathbf{Y_{t-1}}, \ldots, \mathbf{Y_1}\right) - \mathrm{H}\left(\mathbf{Y_{t-1}}, \ldots, \mathbf{Y_1}\right) &(34)\\ &= \log_2 m - \mathrm{H}\left(\mathbf{Y_t} \mid \mathbf{Y_{t-1}}, \ldots, \mathbf{Y_1}\right)\,. &(35)\end{aligned}$$

□

If the random time series is independently and identically distributed, then the rate of change is a constant given by

$$\lambda = \log_2 m - \mathrm{H}\left(\mathbf{Y_1}\right) \text{ bits/trapdoor}\,. \tag{36}$$

We may rewrite $\mu(t)$ in terms of the *rate* of the mean information gain as given by

$$\mu(t) = \sum_{j=1}^{t} \lambda(j)\,. \tag{37}$$

For the *uniformly distributed* time series and the degenerate time series, $\lambda(t) = 0$ and $\lambda(t) = \log_2 m$ respectively for all time steps t.

We make the following conjecture about the *mean information gain.*

Conjecture 6.3. *The* mean information gain $\mu(t)$ *quantifies the amount of information the* optimal *adversary is able to extract from observing* $\mathbf{Y_1}, \ldots, \mathbf{Y_t}$ *for the purpose of mapping* trapdoors *to* plaintext keys.

An upper-bound on the expected *accuracy of the* optimal adversary *at time* t *is an* unknown *function*

$$\mathrm{r}\colon \{1, 2, \ldots\} \mapsto (0, 1] \tag{38}$$

that is a function of $\mu(t)$ *and has the following constraints:*

1. $0 < \mathrm{r}(t) \leq 1$ *for* $t \geq 1$. *The accuracy is between* 0 *and* 1. *However,* $\mathrm{r}(t)$ *is an* expectation*, and the optimal adversary has a chance at correctly mapping trapdoors to plaintext even if the random time series has the maximum entropy, thus it is always greater than* 0.

2. $\mathrm{r}(t+1) \geq \mathrm{r}(t)$. *It is a monotonically increasing function since seeing more of the time series is not* expected *to decrease the optimal adversary's accuracy.*[8]

3. *If* $\lambda(t) = 0$*, then* $\mathrm{r}(t+1) - \mathrm{r}(t) = 0$. *If no information is gained from observing a time step* t*, then the optimal adversary is not expected to improve accuracy at time* t.

4. $\lim_{t \to \infty} \mathrm{r}(t) = c, 0 < c \leq 1$. *This is entailed by the other constraints. If the adversary knows the* true *distribution, where the distribution is not uniformly distributed, and the maximum likelihood equation has a unique solution, then* $c = 1$.

Plausible candidates of r take on *sigmoid*-like curves. Initially, r is near its lower-limit (typically near 0) and as t increases, r begins to slowly increase. Given an appropriate mapping from trapdoors to plaintext, the empirical distribution of the mapped trapdoors starts to resemble the unknown true distribution. At some point, the empirical distribution has nearly zero divergence from the true distribution, and thus the adversary achieves maximum accuracy.

6.1 Estimating entropy

Since the probabilistic model for the random time series may not be known, we may estimate the entropy.

Postulate 6.4 (Optimal compressor)**.** The *entropy* of a random time series is equivalent to the *expected* bit length output by an optimal *lossless* compressor given the time series as input as given by

$$\mathrm{H}\left(\mathbf{Y_1}, \ldots, \mathbf{Y_t}\right) = \mathbb{E}\left[\mathrm{BL}\left(\mathrm{Compress}^*\left(\mathbf{Y_1}\mathbf{Y_2}\cdots\mathbf{Y_t}\right)\right)\right], \tag{39}$$

where Compress* is a lossless optimal compressor of the sequence and $\mathrm{BL}(x)$ is the bit length of x.

Thus, we may estimate the entropy as given by the following definition.

Definition 6.5. Given a time series of t trapdoors,

$$\tau_t = (y_1, \ldots, y_t)\ , \tag{40}$$

an estimator of the entropy is given by

$$\hat{\mathrm{H}} = \mathrm{BL}\left(\mathrm{Compress}(y_1 y_2 \cdots y_t)\right)\ , \tag{41}$$

where Compress is a near-optimal compressor of the time series.

The *entropy* is an *expectation*, and is therefore a constant. However, an optimal compressor as a function of $\mathbf{Y_1}, \ldots, \mathbf{Y_2}$ outputs a bit string with a *random* bit length whose *expectation* is given by the entropy. Thus, it has a *sampling distribution.*

[8] Unless more severe countermeasures to throw off the adversary are employed at time $t+1$ than during previous time steps. However, we assume that time $t+1$ is not a special point in time, and that later time steps do not employ more aggressive counter-measures than earlier time steps.

For the purpose of matching the *trapdoors* to *plaintext*, assuming we have the *true* distribution, the most accurate mapping occurs when the empirical distribution of τ_t has zero divergence from the *true* distribution. The empirical distribution converges in distribution to the *true* distribution, so as $t \to \infty$, $p_t \to 1$.

The adversary given by Eq. (12) is *optimal* if the time series τ_t is drawn from a *unigram* language model using a *simple substitution cipher*.

One possibility is to do a *curve fit* of r to the mean confidentiality with respect to time step t. Alternatively, we may use the relationship between $\mathrm{r}(t)$ and the information gain $\mu(t)$ to establish a mapping. Since $\mathrm{r}(t)$ maps time to accuracy through $\mu(t)$, we proceed as follows: first, compute $\mu(t)$ for the observed time series; second, measure the empirical confidentiality at the corresponding information level; third, map this confidentiality to the expected accuracy $\mathrm{r}(t)$. If this mapping is consistent across different time series and system configurations, it provides a principled way to estimate accuracy bounds from entropy measurements alone.

7 Case study: Zipf distribution

Consider a random time series $\mathbf{X_1}, \ldots, \mathbf{X_n}$. If $\mathbf{X_i}$ for $i = 1, \ldots, n$ follow a Zipf distribution, then its rank is a random variable given by

$$\mathbf{K} = \mathtt{Rank}\left(\mathbf{X_i}\right) \tag{42}$$

such that

$$\mathbf{K} \sim \mathrm{Zipf}(s, N)\,, \tag{43}$$

where N is the number of unique *plaintext* words and s characterizes the degree of *uniformity* of the Zipf distribution.

Definition 7.1. The probability mass function of $\mathbf{K}$ is given by

$$p_{\mathbf{K}}(k \mid s, N) = k^{-s} H_{N,s}^{-1}\,. \tag{44}$$

where $H_{N,s}$ is the *generalized harmonic number* given by

$$H_{n,s} = \sum_{k=1}^{n} k^{-s}\,. \tag{45}$$

By Eq. (42), the probability mass function of $\mathbf{X_i}$ is given by

$$p_{\mathbf{X_1}}(x) = p_{\mathbf{K}}\big(\mathtt{Rank}(x) \mid s, N\big)\,. \tag{46}$$

for $i = 1, \ldots, n$. Similiarly, since a *simple substitution cipher* is being used, the probability mass of $\mathbf{Y_j}$ is given by

$$p_{\mathbf{Y_1}}(y) = p_{\mathbf{X_1}}\big(\mathrm{h}^{\text{-}1}(y)\big) \tag{47}$$

for $j = 1, \ldots, n$.

The entropy of the Zipf distribution is given by the following theorem.

Theorem 7.2. *The entropy of the Zipf distribution with parameters s and N is given by*

$$\mathrm{H}_1(N, s) = H_{N,s}^{-1} \sum_{k=1}^{N} k^{-s} \left(s \log_2 k + \log_2 H_{N,s}\right)\,. \tag{48}$$

Proof.

$$\mathrm{H}_1(N,s) = -\sum_{k=1}^{N} p_{\mathbf{K}}(k \mid s,N) \log_2 p_{\mathbf{K}}(k \mid s,N) \tag{49}$$

$$= -\sum_{k=1}^{N} k^{-s} H_{N,s}^{-1} \log_2 \left(k^{-s} H_{N,s}^{-1} \right) \tag{50}$$

$$= H_{N,s}^{-1} \sum_{k=1}^{N} k^{-s} \left(s \log_2 k + \log_2 H_{N,s} \right) . \tag{51}$$

□

Two limiting cases are given by the following corollaries.

Corollary 7.3. *The maximum entropy results when the Zipf distribution has a parameter value $s = 0$ and is given by*

$$\mathrm{H}_1(s=0,N) = \log_2 N \, . \tag{52}$$

Proof.

$$\mathrm{H}_1(0,N) = H_{N,0}^{-1} \sum_{k=1}^{N} k^0 \left(0 \log_2 k + \log_2 H_{N,0} \right) \tag{53}$$

$$= N^{-1} \sum_{k=1}^{N} \log_2 N \tag{54}$$

$$= \log_2 N \, . \tag{55}$$

□

Corollary 7.4. *The minimum entropy results when the Zipf distribution has a parameter value $s \to \infty$ and is given by*

$$\lim_{s\to\infty} \mathrm{H}_1(s,N) = 0 \, . \tag{56}$$

Proof.

$$\lim_{s\to\infty} \mathrm{H}_1(s,N) = H_{N,\infty} \sum_{k=1}^{N} k^{-\infty} \left(\log_2 k + \log_2 H_{N,\infty} \right) \tag{57}$$

$$= 0 \sum_{k=1}^{N} 0 \left(\log_2 k + \log_2 0 \right) \tag{58}$$

$$= \sum_{k=1}^{N} 0 \log_2 0 \, . \tag{59}$$

The limit

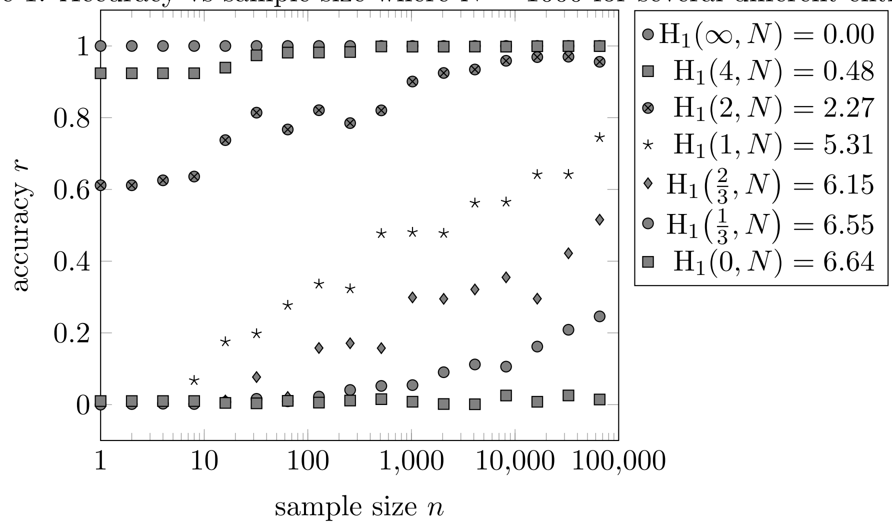
$$\lim_{a\to 0} a \log_2 a = 0 \, , \tag{60}$$

thus

$$\lim_{s\to\infty} \mathrm{H}_1(s,N) = 0 \, . \tag{61}$$

□

Figure 1: Accuracy vs sample size where $N = 1000$ for several different entropies



In Fig. 1, we map the *accuracy* of the adversary with respect to sample size for various entropy levels. The greater the entropy, the less accurate the mapping is expected to be. At one extreme, we have an entropy of 0 (minimum entropy) in which 100% of the traffic is successfully mapped after viewing a sample of size 1 and at the other extreme we have an entropy of 6.64 (maximum entropy) where the accuracy is given by pure random chance and is not correlated with sample size.

8 Application: resilience engineering

Definition 8.1 (Resilience engineering)**.** Here is the definition.

From a resilience engineering perspective, we are interested in the probability that the adversary has compromised the sample of t trapdoors as given by

$$\Pr[\mathbf{C}_t \geq \alpha]\,, \tag{62}$$

where α is the minimum acceptable level of confidentiality. If the probability that this minimum level is relatively low (e.g., less than 95%), the trapdoor signatures could be reassigned to reestablish confidentiality.

As $t \to \infty$, Eq. (62) goes to 0. The minimum sample size the adversary may observe without exceeding some specified level of risk is given by the following definition.

Definition 8.2. The maximum number of trapdoors the adversary may observe with an acceptable level of risk of successfully compromising the confidentiality of the system is given by

$$t^* = \arg\min_t \Pr[\mathbf{C}_t > \alpha] > \beta\,, \tag{63}$$

where

- α is the minimum level of confidentiality the *Encrypted Search* system seeks to maintain and
- β is an unacceptable level of risk (probability) that the minimum level of confidentiality is not met.

Given a set of Bootstrap resample of m confidentiality statistics

$$\mathbb{K} = \left\{ C_t^{(1)}, \ldots, C_t^{(m)} \right\} ,$$

we may estimate Eq. (62) in two ways. The most straightforward way is the proportion of the sample that is greater than α as given by the statistic

$$\Pr[\mathbf{C}_t \geq \alpha] \approx \frac{|\mathbb{A}|}{m} , \tag{64}$$

where

$$\mathbb{A} = \{C \in \mathbb{K} \colon C > \alpha\} . \tag{65}$$

However, by Theorem 5.4, $\mathbf{C}_t$ converges in distribution to a normal distribution. Thus, by the large sample approximation,

$$\Pr[\mathbf{C}_t \geq \alpha] \approx 1 - \phi\left(\frac{\alpha - \bar{C}_t}{s_{t-1}}\right) , \tag{66}$$

where ϕ is the cumulative distribution function of the standard normal, $\bar{C}_t$ is the sample mean, and s_{t-1} is the sample standard deviation. Substituting Eq. (66) into Eq. (63) and simplifying results in a statistic of t^* given by

$$\widehat{t^*} = \arg\min_t \phi\left(\frac{\alpha - \bar{C}_t}{s_{t-1}}\right) < 1 - \beta . \tag{67}$$

9 Conclusions and future work

The primary contributions in this paper are two-folds. First, there has not been much work for studying how safe encrypted searches are against frequency attacks, which can be measured by a large number of attackers for long period of time, possibly infinitely long. We provide studies on the resilience of encrypted searches against frequency attacks from the view point of resilience engineering approach to enhance security on encrypted searches. Resilience engineering is a new way of enhancing safety by precisely estimating the level of possible threats to a system and feeding them back to adjusting or re-designing the system to maintain the acceptable level of safetyAssociation.

Our second contribution is development of a new method, Moving Average Bootstrap (MAB) method, which efficiently and accurately calculates the estimator for the minimum number of encrypted words (N^*) an adversary needs to achieve a given accuracy level (p^*) with a certain level of confidence as soon as a relatively small number of samples (n) (i.e., encrypted words) are submitted by legitimate users. Thus, the MAB method will let the defenders calculate the estimator at an early stage without waiting for a large number of queries submitted by legitimate users. Especially from the view point of "tractability", calculating the estimator using, not to mention an infinitely large number of encrypted words, a large number of encrypted words takes time (waiting for a large number of encrypted words to be submitted) and huge storage (storage space to hold the submitted encrypted words) is required.

Our proposed MAB method calculated the estimated number of encrypted search queries an adversary needs to observe (N^*) for achieving a given accuracy level, $p^* = 0.30$, at the confidence level of 95% using only 5% of the actual observations (250/5000) (Figure 5 (c)). Assuming that the increase in the time an adversary needs to achieve a certain p^* is proportional to the ratio in the increase of the number of the encrypted words observed by an adversary (n) for a large number of encrypted words, the MAB method would allow a defender to estimate N^* in 5% of time (without waiting for legitimate users to issue a large number of encrypted words). We are currently performing analyses using higher p^* (0.55 through 0.80) for different levels of confidence (90 to 98%) for observing how they affect the performance of MAB method and for observing if there is any pathological case for MAB method.

References

Resilience Engineering Association. About resilience engineering. URL `http://www.resilience-engineering-association.org`.

Dan Boneh, Giovanni Di Crescenzo, Rafail Ostrovsky, and Giuseppe Persiano. Public key encryption with keyword search. pages 506–522, 2004.

Jin Wook Byun, Hyun Suk Rhee, Hyun-A Park, and Dong Hoon Lee. Off-line keyword guessing attacks on recent keyword search schemes over encrypted data. *Secure Data Management, Lecture Notes in Computer Science*, 4165:75–83, 2006.

Ning Cao, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou. Privacy-preserving multi-keyword ranked search over encrypted cloud data. *IEEE Transactions on Parallel and Distributed Systems*, 25(1):222–233, January 2014.

William R. Claycomb and Alex Nicoll. Insider threats to cloud computing: Directions for new research challenges. pages 388–394, 2012.

Ik Rae Jeong, Jeong Ok Kwon, Dowon Hong, and Dong Hoon Lee. Constructing peks schemes secure against keyword guessing attacks is possible? *Computer Communications Express*, 32(2): 394–396, 2009.

Qian Wang Jin Li, Cong Wanga, Ning Cao, Kui Ren, and Wenjing Lou. Fuzzy keyword search over encrypted data in cloud computing. pages 441–445, 2010.

Seny Kamara and Kristin Lauter. Cryptographic cloud storage. *Financial Cryptography and Data Security, Lecture Notes in Computer Science*, 6054:136–149, 2010.

Debashis Kusharya. Bootstrap methods and their application. *Technometrics*, 42(2):216–217, March 2012.

Steven T. Piantadosi. Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, 21(5):1112–1130, October 2014.

Jean-François Raymond. Traffic analysis: Protocols, attacks, design issues, and open problems. *Designing Privacy Enhancing Technologies, Lecture Notes in Computer Science*, 2009:10–29, March 2001.

Matthieu Rivain. On the exact success rate of side channel analysis in the gaussian model. *Selected Areas in Cryptography*, 5381:165–183, 2009.

S. Subashini and V. Kavitha. A survey on security issues in service delivery models of cloud computing. *Journal of Network and Computer Applications*, 34(10):1–11, 2011.

Wenhai Sun, Bing Wang, Ning Cao, Ming Li, Wenjing Lou, Thomas Y. Hou, and Hui Li. Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking. pages 77–82, May 2013.

Adrian Thillard, Emmanuel Prouff, and Thomas Roche. Success through confidence: Evaluating the effectiveness of a side-channel attack. *Cryptographic Hardware and Embedded Systems*, 8086: 21–36, 2013.

Wei-Chuen Yau, Swee-Huay Heng, and Bok-Min Goi. Off-line keyword guessing attacks on recent public key encryption with keyword search schemes. pages 100–105, 2008.

Dimitrios Zissis and Dimitrios Lekkas. Addressing cloud computing security issues. *Future Generation Computer Systems*, 28(3):589–592, 2012.