Computer Security Symposium 2014
22 - 24 October 2014

機能安全の概念に基づく秘密情報の分割保管に関する一考察

披田野 清良† 清本 晋作† 三宅 優†

†KDDI 研究所
356-8502 埼玉県ふじみ野市大原 2-1-15
se-hidano@kddilabs.jp

あらまし インターネット上に保管されているパスワードや暗証番号などの秘密情報が漏洩する事例が多発している．それらの多くは，人的ミスや機器の故障，あるいは故意による不正などの潜在的な脆弱性に起因するものであり，根本的な原因を取り除くことは容易ではない．そこで，本稿では，上記のような最悪なケースが起きた場合においてもシステムが正しく動作することを目指す機能安全の概念に着目する．そして，システムから一部の秘密情報が漏洩した際に，残りのデータへの影響を最小限に抑える秘密情報の分割保管の方法を提案する．

A Study on Partitioning Secret Data Based on Concept of Functional Safety

Seira Hidano† Shinsaku Kiyomoto† Yutaka Miyake†

†KDDI R&D Laboratories, Inc.
2-1-15 Ohara, Fujimino-shi, Saitama, 356-8502 JAPAN
se-hidano@kddilabs.jp

Abstract Incidents that secret data such as passwords and PINs stored online are leaked frequently occurred. Since many of them result from potential vulnerabilities like human errors, bugs or intentional misconduct, it is not easy to get rid of underlying causes. In this paper, we thus focus on the concept of functional safety whose goal is that systems work correctly if the worst case like above incidents should occur. We propose a partitioning method to store secret data more securely, which has little or no influence on the remaining data even though a part of secret data are leaked from a system.

1 はじめに

近年，パスワードや暗証番号などの秘密情報が漏洩する事例が多発しており，漏洩情報を用いたシステムの不正利用に関する被害が多数報告されている．しかしながら，それらの事例の多くは，人的ミスや機器の故障，あるいは故意による不正などの潜在的な脆弱性に起因するものであり，根本的な原因を取り除くことは容易ではない．このため，今後も情報漏洩のリスクをゼロにすることはきわめて難しいと考えられる．そこで，本稿では，最悪なケースが起きた場合においてもシステムが正しく動作することを目指す機能安全の概念に着目する．すなわち，情報漏洩は上記理由により現実的に起こりうることを仮定し，情報漏洩によるシステムの安全性への影響を最小限に抑える方法を確立することにより本問題の解決を図る．特に，本稿では，現在多くのサービスがクラウド上で展開されて

いることから，情報漏洩のケースとして以下の場合を想定する．

クラウド上のシステムでは複数のサーバを仮想的に統合して利用しているため，一部のストレージに故障やバグが生じることや，異なる管理者のデータセンターのサーバを統合している場合には一部の管理者が不正を働くなど，部分的にインシデントが発生する可能性が高い．それらのインシデントにより脆弱性が顕在化した場合，当該システムの一部のユーザの秘密情報が大量に漏洩する可能性がある．情報漏洩が発覚した直後，対象ユーザは早期に情報を変更すると考えられるが，それ以外のユーザに関しては，複数のサービスで同一の情報を使用していることなどに起因する変更の煩わしさから，変更に消極的であることが鑑みられる．秘密情報として用いられるパスワードや生体情報などの認証情報，あるいは物理的なソースなどの多くは，各値が一様に生起しない非一様な情報である [1]．それらの非一様性を普遍的にモデル化することは必ずしも容易ではないが，一部のユーザの秘密情報を知ることができれば，当該システムの他のユーザの秘密情報の偏りを推測できる可能性がある．秘密情報の非一様性を利用した場合，総当り的に攻撃するよりも効率的に攻撃を遂行できる．このため，本稿では，情報漏洩のケースとしてシステムから一部のユーザの秘密情報が大量に漏洩する場合を想定し，残りのユーザの秘密情報への影響を最小限に抑える方法について議論する．

また，セキュリティ分野における機能安全の概念に基づく従来研究としては，サイドチャネル攻撃などにより秘密情報について何らかの情報が漏洩することを想定した漏洩耐性暗号に関する試みがあげられる [2, 3]．それらの議論では，上記の暗号プリミティブが情報漏洩時においても安全であることを立証するために，漏洩関数 f により定式化された秘密情報 X に関する漏洩情報 $f(X)$ を用いて情報漏洩時の安全性をモデル化している．しかしながら，それらのモデルではある特定のユーザの X の部分的な情報のみが漏洩することを想定しており，たとえば，$f(X)$ は X の一部のビットであるなど，本稿で想定するようなシステムから大量の秘密情報が漏洩した状況下での安全性を保障するものではない．

以上より，本稿では，まず，システムから一部のユーザの秘密情報が漏洩した際に考えられる，秘密情報の非一様性を考慮した攻撃モデルを明らかにし，機能安全の概念に基づく新たな安全性モデルを構築する．そして，残りのユーザの秘密情報への影響を最小限に抑える方法として，上記の安全性モデルの下，秘密情報のデータベース分割に関する新たな方法を提案する．また，本稿の残りの部分にて，シミュレーション結果を交えて提案手法の有効性を示すとともに，安全性に関する他の従来指標の観点から見ても本提案手法はシステムの安全性を向上する上で有効な手法であることを示す．

2 準備

次章以降での議論の際に使用する 2 次の Renyi エントロピーおよび確率変数の距離の定義について述べる．

2 次の Renyi エントロピー　離散集合 $\mathcal{X}$ 上の確率変数 X の確率関数を $p(x)$ とする．X の 2 次の Renyi エントロピー $H_2(X)$ は次式で定義される [4]．

$$H_2(X) = -\log \sum_{x\in\mathcal{X}} p(x)^2 \tag{1}$$

確率変数の距離　$\mathcal{X}$ 上の確率変数 X，Y の確率関数をそれぞれ $p(x)$，$q(x)$ とする．X と Y の距離は次式で定義される．

$$D(X,Y) = \left[\sum_{x\in\mathcal{X}} d_x(X,Y)^2\right]^{\frac{1}{2}} \tag{2}$$

ただし，$d_x(X,Y) = p(x) - q(x)$ とする．ここで，$|\mathcal{X}| = 2^n$，$H_2(X) = r$ とする．$q(x)$ が一様分布のとき，X と Y の距離 $E(X)$ は次式で表せる．

$$E(X) = \left(2^{-r} - 2^{-n}\right)^{\frac{1}{2}} \tag{3}$$

3 安全性モデル

機能安全の概念の下，秘密情報漏洩時の安全性モデルを構築する．まず，システムから一部のユーザの秘密情報が漏洩した際に考えられる攻撃モデルとして，Distribution Guessing Attack（DGA）を示す．次いで，DGA に対する安全性として DGA 安全を定義し，DGA 安全は確率変数の距離に帰着できることを示す．

3.1 攻撃モデル

あるシステムにおいて一部のユーザの秘密情報が漏洩した場合，攻撃者はそれらの情報を利用することにより，残りのユーザの秘密情報の分布を推測できる可能性がある．本節では，その推測された分布を用いた攻撃モデルの 1 つとして，Distribution Guessing Attack（DGA）を示す．

秘密情報が漏洩したユーザの集合を $\mathcal{U}'$，$\mathcal{U}'$ の各ユーザの秘密情報データの集合を $\mathcal{W}'$ とする．また，攻撃対象のユーザの集合を $\mathcal{U}$，$\mathcal{U}$ の各ユーザの秘密情報データの集合を $\mathcal{W}$ とする．$\mathcal{W}$ の各データはその取りうる値の集合 $\mathcal{X}$ 上の確率関数 $p(x)$ にしたがって生起したものとする．ただし，$u \neq u', u \in \mathcal{U}, u' \in \mathcal{U}'$ とする．DGA の攻撃手順を攻撃者と挑戦者によるゲーム形式で示す．

1. 攻撃者は任意の分布推測アルゴリズム $\mathcal{K}$ を用いて取得した $\mathcal{W}'$ から確率関数 $p(x)$ を推測する．このとき，推測された分布を $q(x)$ とする．
2. 挑戦者はユーザ $u \in U$ をランダムに選択し，対応する秘密情報 $w \in \mathcal{W}$ を抽出する．
3. 攻撃者は $q(x)$ にしたがって秘密情報 $x \in_q \mathcal{X}$ を選択する．
4. x と w が一致すれば，攻撃者の勝ちとする．

3.2 安全性の定義

攻撃対象ユーザの秘密情報の分布 $p(x)$，攻撃者が推測した分布 $q(x)$ より，DGA の平均攻撃成功確率 P_{DGA} は次式で表せる．

$$P_{DGA} = \sum_{x \in \mathcal{X}} p(x)q(x) \tag{4}$$

仮に，システムから何の情報も漏洩していなければ，攻撃者は $p(x)$ に関して何の情報も持たないため，$\mathcal{X}$ 上の一様分布にしたがって DGA を実行する他ない．$|\mathcal{X}| = 2^n$ とすれば，そのときの攻撃成功確率は 2^{-n} で表せる．したがって，一部の秘密情報が漏洩した場合であっても，$P_{DGA} \leq 2^{-n}$ であれば，情報が漏洩していないときと同様に安全であるといえる．本稿ではこのときの安全性を DGA 安全と定義する．

ここで，X，Y をそれぞれ $p(x)$，$q(x)$ を確率関数として持つ $\mathcal{X}$ 上の確率変数とし，$H_2(X) = r$，$H_2(Y) = s$ とする．$r \leq s$ のとき，次の不等式が成立する．

$$P_{DGA} \leq 2^{-r} - \frac{D(X,Y)^2}{2} \tag{5}$$

Proof. $r \leq s$ より，

$$\sum_{x \in \mathcal{X}} p(x)^2 \leq \sum_{x \in \mathcal{X}} q(x)^2 \tag{6}$$

$$\sum_{x \in \mathcal{X}} p(x)^2 \leq \sum_{x \in \mathcal{X}} (p(x) - d_x(X,Y))^2 \tag{7}$$

$$\sum_{x \in \mathcal{X}} p(x) d_x(X,Y) \geq \frac{\sum_{x \in \mathcal{X}} d_x(X,Y)^2}{2}. \tag{8}$$

したがって，

$$P_{DGA} = \sum_{x \in \mathcal{X}} p(x)(p(x) - d_x(X,Y)) \tag{9}$$

$$\leq \sum_{x \in \mathcal{X}} p(x)^2 - \frac{\sum_{x \in \mathcal{X}} d_x(X,Y)^2}{2}. \tag{10}$$

□

式 (3) および式 (5) より，$P_{DGA} \leq 2^{-n}$ のとき，次の不等式が成立する．

$$D(X,Y)^2 \geq 2E(X)^2 \tag{11}$$

したがって，$r \leq s$ かつ $D(X,Y)^2 \geq 2E(X)^2$ を達成できれば，上記の定義と同様にそのシステムは DGA 安全であるといえる．

4 提案手法

システムに保管されている一部の秘密情報が漏洩するケースとして，クラウド化された 2 つのストレージのうち 1 つのストレージのデータがすべて漏洩してしまう場合を考える．秘密情報は 2 つのデータベース DB_1，DB_2 に分割され，異なるストレージに保管されているものとする．一方のデータベースが漏洩した場合，攻撃者は 3.1 節のモデルにしたがって，他方のデータベースに対して DGA を実行することが想定される．本章では，どちらのデータベースが漏洩した場合においても，DGA に対して耐性のある秘密情報の分割保管の方法を提案する．

4.1 問題設定

分割方法について議論するにあたり，攻撃者およびシステムの条件について考える．本稿では，以下の 3 つの条件を設定する．

1. 攻撃者は，分布推測アルゴリズム $\mathcal{K}$ としてヒストグラムを用いる．
2. DB1 および DB2 のどちらのデータベースが漏洩しても DGA への耐性は同様である．
3. いずれのデータベースも漏洩していない場合，DB1 および DB2 は同様に十分に安全である．

$p_1(x)$ を DB1 に含まれるデータのヒストグラム，$p_2(x)$ を DB2 に含まれるデータのヒストグラムとする．条件 1 より，攻撃者の推測分布 $q(x)$ は，DB2 が漏洩した場合は $p_2(x)$，DB1 が漏洩した場合は $p_1(x)$ で表せる．X_1，X_2 をそれぞれ $p_1(x)$，$p_2(x)$ にしたがう確率変数とする．ここで，条件 2 を考慮して，$r = H_2(X_1) = H_2(X_2)$ とすると，式 (5) の等号が成立するため，どちらのデータベースが漏洩した場合においても DGA の平均攻撃成功確率 P_{DGA} は次式で表せる．

$$P_{DGA} = 2^{-r} - \frac{D(X_1, X_2)^2}{2} \tag{12}$$

本稿では条件 1 により攻撃者は分布推測アルゴリズムとしてヒストグラムを用いることを仮定しているが，実際にはより有効なアルゴリズムを使用する可能性も考えられる．攻撃者が攻撃対象のデータベースの分布 $p(x)$ と同一の分布を推測できた場合，$D(X_1, X_2) = 0$ となり，DGA の平均攻撃成功確率は 2^{-r} で表せる．したがって，r の値は十分に大きくする必要がある．ただし，$H_2(X_1) = H_2(X_2)$ を仮定しているため，DB1 および DB2 にランダムにデータを振り分けたときに $H_2(X_1)$，$H_2(X_2)$ はそれぞれ最大となる．したがって，r は次式を満たすように設定する．

$$r = H_2(X_1) = H_2(X_2) = H(X) \tag{13}$$

ただし，X は DB1 および DB2 の総データの分布にしたがう確率変数とする．

また，条件 3 を考慮して，攻撃者が $p(x)$ に関して何の情報も持たないときの DGA への耐性を高くするために，次式を設定する．

$$|\mathcal{X}_1| = |\mathcal{X}_2| = |\mathcal{X}| \tag{14}$$

ただし，$\mathcal{X}_1$，$\mathcal{X}_2$ はそれぞれ DB1 および DB2 のデータの取りうる値の集合とし，$\mathcal{X}$ は DB1 および DB2 の総データの取りうる値の集合とする．

以上より，DGA への耐性を向上させるためには，式 (13)，式 (14) の条件の下，式 (12) の $D(X_1, X_2)$ を大きくすればよいことがわかる．しかしながら，データの個数が増加するにつれて，分割の組み合わせ総数は指数的に増加するため，上記の問題を総当たり的に解くのは現時的に容易ではない．このため，4.2 節において，上記問題を効率的に解くアルゴリズムを示す．

4.2 分割アルゴリズム

$c_1(x)$ および $c_2(x)$ をそれぞれ DB1 または DB2 に含まれる値 $x \in \mathcal{X}$ を取るデータの個数とする．このとき，DB1 および DB2 の総データの中で値 x を取るものの個数は $c(x) = c_1(x) + c_2(x)$ で与えられる．本節では，4.1 節で設定した問題を次の問題に帰着させる．

目的関数

$$\text{maximize} \quad \sum_{x \in \mathcal{X}} [c_1(x) - c_2(x)]^2 \tag{15}$$

制約条件

$$\sum_{x\in\mathcal{X}} c_1(x)^2 = \sum_{x\in\mathcal{X}} c_2(x)^2 = \sum_{x\in\mathcal{X}} \frac{c(x)^2}{4} \quad (16)$$

$$\sum_{x\in\mathcal{X}} c_1(x) = \sum_{x\in\mathcal{X}} c_2(x) = \sum_{x\in\mathcal{X}} \frac{c(x)}{2} \quad (17)$$

$$1 \le c_1(x), 1 \le c_2(x) \quad (18)$$

上記問題を解くアルゴリズムを以下に示す．

初期設定 N 個の秘密情報データを同一サイズのデータベース DB1 および DB2 に分割する．ただし，すべての $x \in \mathcal{X}$ において $c_1(x) = c_2(x)$ を満たすように分割する．次いで，$\Delta = \{\delta(x) = c(x) - 2\}_{x\in\mathcal{X}}$ を作成する．

Step 1 最も大きい $\delta(x) \in \Delta$ と対応する x を選択し，$[\delta(x) - |c_1(x) - c_2(x)|]/2$ 個のデータを DB2 から DB1 に移動する．

Step 2 $[\delta(x) - |c_1(x) - c_2(x)|]/2$ 個のデータを式 (16) を満たすように DB1 から DB2 に移動する．ただし，この移動はグリーディ法により行う [5]．

Step 3 $\delta(x) = \perp$ として Δ を更新する．ただし，Step 2 において，条件を満たすデータの移動ができなかった場合には，Step 1 および Step 2 を行う前の状態に戻し，$\delta(x) = \delta(x) - 2$ として Δ を更新する．

Step 4 Step 1 から Step 3 までの動作を Δ のすべての $\delta(x)$ が 0 以下もしくは $\perp$ になるまで繰り返す．ただし，本ステップは以下の条件にしたがって行う．

- Step 1 において，$c_2(x) > c_1(x)$ が成立する場合，データは DB1 から DB2 に移動する．この場合，Step 2 では，DB2 から DB1 に移動する．
- Step 1 において，$|c_1(x) - c_2(x)| \ge \delta(x)$ が成立する場合，当該 x に関するデータの移動は行わずに，$\delta(x) = \perp$ として Δ を更新し，次の x のプロセスに移行する．
- Step 2 では，$\delta(x) \neq \perp$ の x のデータのみ移動する．

グリーディ法 Step 2 で用いるグリーディ法について解説する．値 x を取るデータを DB1 から DB2 に e 個移動させたときの $c_2(x)^2 - c_1(x)^2$ の変化量は次式で表せる．

$$\begin{aligned} & [(c_2(x)+e)^2 - (c_1(x)-e)^2] \\ & -[c_2(x)^2 - c_1(x)^2] \quad (19) \\ = \ & 2\cdot[c_2(x) + c_1(x)]\cdot e \quad (20) \\ = \ & 2\cdot c(x)\cdot e \quad (21) \end{aligned}$$

したがって，$c(x)$ が大きい x のデータを移動させた方が少ないデータの移動で $|\sum_{x\in\mathcal{X}} c_1(x)^2 - \sum_{x\in\mathcal{X}} c_2(x)^2|$ を大きく変化させられることが分かる．このため，Step 2 のグリーディ法では，$c(x)$ が最も大きい x のデータから順に移動する．

計算量 本アルゴリズムにおいてグリーディ法の計算量は $N/2$ を超えることはない．また，Step 3 の計算量は Δ の更新回数に比例するため，最悪の場合でも高々 $N/2$ である．したがって，本分割アルゴリズムの最悪時の計算量は $\mathcal{O}(N^2)$ となる．

5 評価実験

4.2 節の分割アルゴリズムについて実際に異なる 3 つのデータセットを生成し評価実験を行った．秘密情報は 32 個の異なる値を取るものとし，1000 個のデータを生成した．ただし，それらのデータは標準正規分布にしたがって生起させており，$[-3\sigma, 3\sigma]$ の区間を均等に 32 分割し，それぞれの区間に秘密情報の各値を割り当てている．

本実験で得られた結果を表 1 に示す．表 1 には，それぞれのデータセットについて，式 (4) より算出した P_{DGA}，攻撃者が DB2 から DB1 の分布を完全に推測できたときの DB1 に対する平均攻撃成功確率 2^{-r}，DB1 と DB2 の分布間の距離 $D(X_1, X_2)$ を 2 乗した値，DB1 の分布とすべての生起確率が 1/32 の分布の距離 $E(X_1)$ を 2 乗して 2 倍した値を記載している．どのデータセットにおいても P_{DGA} は 2^{-r} よりも小さくなっており，提案手法の DGA に対する有

表 1: 実験結果

Index	1	2	3
P_{DGA}	0.0129	0.0298	0.0463
2^{-r}	0.0992	0.0776	0.0614
$D(X_1, X_2)^2$	0.169	0.0956	0.0278
$2E(X_1)^2$	0.136	0.0927	0.0603

効性が見て取れる．また，$D(X_1, X_2)^2$ に注目すると，その値が大きい程，式 (12) が示すように，P_{DGA} と 2^{-r} の差分を大きくし，DGA への耐性を高められることが確認できる．ここで，データセット 1 の P_{DGA} に着目すると，その値は，攻撃者が一様分布にしたがって DGA を試みた場合の攻撃成功確率 $1/32 = 0.03125$ よりも小さく，DGA 安全を達成している．このとき，$D(X_1, X_2)^2 \geq 2E(X_1)$ が成立しており，3.2 節で示したように DGA 安全は分布間の距離に帰着できることが確認できる．データセット 2 についても同様に DGA 安全を達成している．しかしながら，データセット 3 に関しては DGA を達成していない．ここで，データセット 1 およびデータセット 3 のヒストグラムをそれぞれ図 1，2 に示す．横軸は秘密情報の取りうる値の参照番号，縦軸はその頻度を示す．図 1 では，ほぼすべての参照番号において DB1 と DB2 の間で出現頻度がまったく異なっている．一方，図 2 では，ほぼ同一の頻度で出現している参照番号が半数を占めている．表 1 の 2^{-r} の値に着目すると，データセット 1 よりもデータセット 3 の方が確率が低い．これは，データセット 1 よりもデータセット 3 の方が 2 次の Renyi エントロピーが大きいことを意味する．2 次の Renyi エントロピーが大きい場合，頻度のばらつきが少なくなるため，式 (21) より，データ 1 個あたりの DB 間の移動における 2 次の Renyi エントロピーの変化量のばらつきも少なくなる．この場合，式 (16) の条件を満たすために必要な微調整可能なデータが少なくなり，分割アルゴリズムの Step 2 を実現できない可能性が高くなる．このため，表 1 が示すように，2 次の Renyi エントロピーが大きいデータセットの順に，$D(X_1, X_2)^2$ が小さくなったと考えられる．

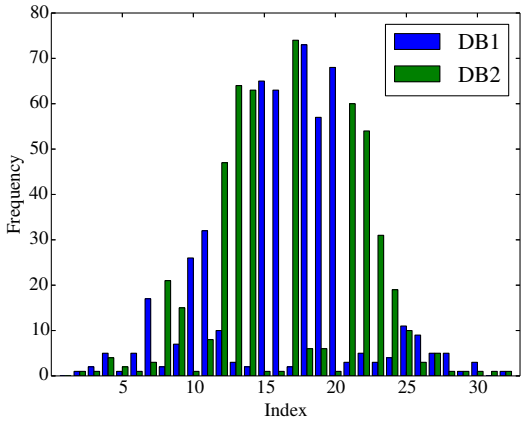

図 1: データセット 1 の分布

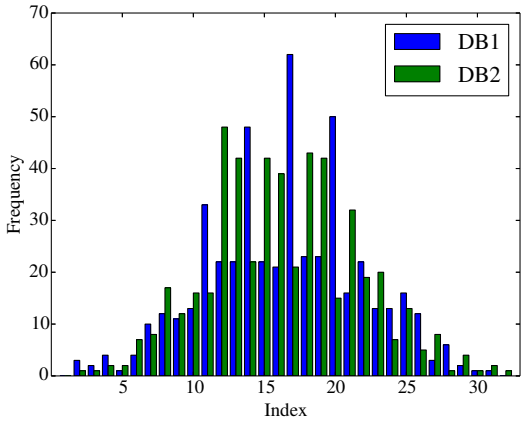

図 2: データセット 3 の分布

ただし，パスワードや生体情報など現実的に利用されている秘密情報の多くは 2 次の Renyi エントロピーがあまり大きくないことが知られている [6]．このため，提案手法を現実的な秘密情報に適用することを考えれば，本手法は DGA への耐性を高める上で非常に有効な手法であるといえる．

6 従来指標による評価

秘密情報の非一様性を利用した攻撃モデルに関する代表的な従来指標としては，推測エントロピーと最小エントロピーがあげられる [7]．本章では，4 章の提案手法の有効性を推測エント

表 2: 推測エントロピー

Index	1	2	3
$G(X_1)$	19.1	14.7	10.7
$\min G(X_1)$	5.78	7.11	8.41

ロピーおよび最小エントロピーの観点から 5 章の実験結果に基づき考察する．

6.1 推測エントロピー

推測エントロピーに関する安全性モデルでは，ある特定のユーザに対し，秘密情報の分布 $p(x)$ を用いて，最も確率の高い x から順に攻撃を試みる攻撃モデルを想定している．推測エントロピーは，攻撃が成功するまでの試行回数 i の期待値として定義される．ここで，3.1 節での議論と同様に，攻撃者が $p(x)$ ではなく取得した一部の秘密情報データから推測した分布 $q(x)$ を用いて上記の攻撃を試みることを想定した場合，$G(X)$ は次式で表せる．

$$G(X) = \sum_{x_i \in \mathcal{X}} i \cdot p(x_i) \tag{22}$$

ただし，$q(x_1) \geq \ldots \geq q(x_i) \geq \ldots \geq q(x_{2^n})$ とする．

5 章の実験結果より算出した推測エントロピーを表 2 に示す．ただし，DB2 のデータが漏洩した場合を想定している．表 2 には，$q(x) = p(x)$ のときの推測エントロピー $\min G(X_1)$ を併記している．$G(X_1)$ と $\min G(X_1)$ の値を比較すると，どのデータセットにおいても上記攻撃への耐性が向上していることがわかる．ここで，図 1 および図 2 の分布を DB2 の頻度の高い順に並び替えたものを図 3，図 4 に示す．推測エントロピーは，式 (22) および Rearrangement inequality より，$p(x)$ と $q(x)$ の確率のランクが異なる程より大きな値を取るといえる．図 3 に着目すると，DB2 と DB1 ではランクがまったく異なることがわかる．一方，図 4 では，一部の頻度の高いものをのぞきほとんどランクが変わっていない．このため，データセット 3 よりもデータセット 1 の方が推測エントロピーの値

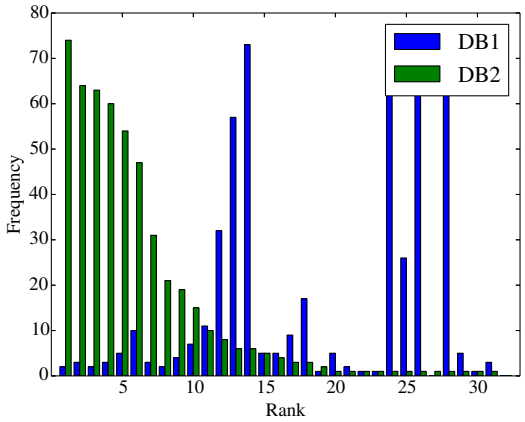

図 3: データセット 1 の分布（ランク順）

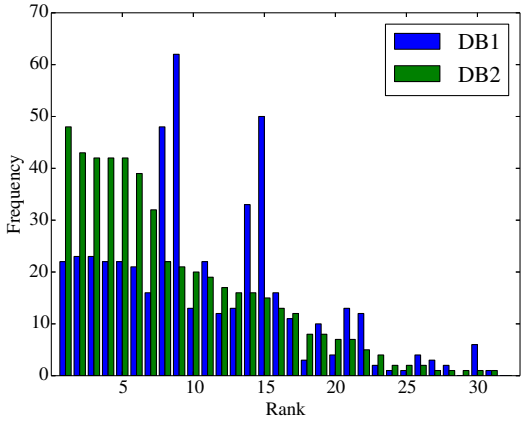

図 4: データセット 3 の分布（ランク順）

が大きく増加したと考えられる．したがって，推測エントロピーの観点から見ても，提案手法は非一様性の強い現実的な秘密情報に適用する上で有効な手法であるといえる．

6.2 最小エントロピー

最小エントロピーに関する安全性モデルでは，任意のユーザに対して，$p(x)$ の中で最も高い確率の x のみを用いて攻撃を試みる攻撃モデルを想定している．6.1 節と同様に，攻撃者は $q(x)$ を用いて上記の攻撃を試みることを想定した場合，最小エントロピー $H_\infty(X)$ は次式で表せる．

$$H_\infty(X) = -\log p(\arg\max_{x \in \mathcal{X}} q(x)) \tag{23}$$

表 3: 最小エントロピー

Index	1	2	3
$H(X_1)_\infty$	7.94	8.95	4.48
$\min H(X_1)_\infty$	2.75	2.76	2.99

5 章の実験結果より算出した最小エントロピーを表 3 に示す．ただし，表 2 と同様に DB2 のデータが漏洩した場合を想定している．表 3 には，$q(x) = p(x)$ のときの最小エントロピー $\min H_\infty(X_1)$ を併記している．推測エントロピーの場合と同様に，どのデータセットにおいても $H_\infty(X_1)$ は $\min H_\infty(X_1)$ の値を上回り安全性が向上していることがわかる．4.2 節の分割アルゴリズムでは，Δ を用いて頻度の差が最も大きくなる x から順にデータを移動させている．このため，一方のデータベースでは最も高い確率を示す x であっても，他方のデータベースではその確率が小さくなるのは明らかである．したがって，提案手法は最小エントロピーの面から見ても有効な手法であるといえる．

7 おわりに

本稿では，まず，機能安全の概念に基づく安全性モデルとして，漏洩した一部のユーザの秘密情報を用いた Distribution Guessing Attack（DGA）を示し，新たな安全性の概念として DGA 安全を定義した．次いで，DGA に対して耐性のあるシステムを構築するための秘密情報のデータベース分割に関する新たな方法を提案した．そして，本提案手法は，特に非一様性の強い秘密情報に適用する上で有効であり，その場合 DGA 安全を達成できる可能性があることをシミュレーション結果を交えて定量的に示した．また，それらのシミュレーション結果から，本提案手法は従来指標の推測エントロピーや最小エントロピーの観点からも有効な手法であることを明らかにした．

今後は，まず，攻撃者が分布推測アルゴリズムとしてヒストグラム以外のより高度な推測アルゴリズムを用いた場合の攻撃モデルを考え，本提案手法の有効性を再検討する．また，本稿では，秘密情報を 2 つのデータベースに分割して保管する場合のみを想定しているが，実際のクラウドではデータを 2 つ以上の複数のストレージに分割する場合が一般的であると考えられる．このため，今後は本方式を拡張し，機能安全の概念の下，秘密情報を複数のデータベースに分割する方法を確立する．

参考文献

[1] Dodis, Y. and Yu, Y.: Overcoming weak expectations, *Proceedings of the 10th Theory of Cryptography Conference (TCC 2013)*, pp. 1–22 (2013).

[2] Adi Akavia, Shafi Goldwasser, V. V.: Simultaneous Hardcore Bits and Cryptography against Memory Attacks, *the 6th Theory of Cryptography (TCC 09), LNCS*, Vol. 5444, pp. 474–495 (2009).

[3] Duc, A., Dziembowski, S. and Faust, S.: Unifying leakage models: from probing attacks to noisy leakage, *Advances in Cryptology - EUROCRYPT 2014, LNCS*, Vol. 8441, pp. 423–440 (2014).

[4] Renyi, A.: On measures of entropy and information, *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 547–561 (1960).

[5] Aho, A. V., Hopcroft, J. E. and Ullman, J. D.: *Data Structures and Algorithms*, Addison-Wesley (1983).

[6] Hidano, S., Ohki, T. and Takahashi, K.: Evaluation of security for protected template in biometric cryptosystem using fuzzy commitment scheme, *IPSJ Journal*, Vol. 54, No. 11, pp. 2383–2391 (2013).

[7] Burr, W. E., Dodson, D. F. and Polk, W. T.: Electronic Authentication Guideline, *NIST Special Publication 800-63 Version 1.0.2* (2006).