Journal of Information Assurance and Security
ISSN 1554-1010 Volume 6 (2011) pp.089-097
© MIR Labs, www.mirlabs.net/jias/index.html

Joint and Conditional Guesswork: Definitions and Implications

Reine Lundin and Stefan Lindskog

Department of Computer Science
Karlstad University
Sweden
{*reine.lundin* | *stefan.lindskog*} @*kau.se*

Abstract: **The need for computer security in today's open computer networks is now undisputed. More and more effort is being spent on security-enhancing methods and techniques. Despite this, there is still a lack of good methods for quantitatively assessing security. New metrics that provide a more exact description of security are therefore desirable. To address this we present an in-depth investigation of the probabilistic measure guesswork, which gives the average number of guesses in an optimal brute force attack. The paper extends the definition of guesswork by defining joint and conditional guesswork. It is proved that joining increases guesswork, while conditioning reduces it. This implies that the joint guesswork is always at least equal to the marginal guesswork and that the conditional guesswork is always at most equal to the marginal guesswork. The paper also provides a description of relations and similarities between guesswork and entropy.**

Keywords: Computer security, security measures, entropy, guesswork, joint guesswork, conditional guesswork.

I. Introduction

Computer security is traditionally defined as an umbrella concept that consists of three attributes: confidentiality (or secrecy), integrity and availability. These are often collectively referred to as the "CIA" [3]. Confidentiality implies prevention of unauthorized disclosure of information, while integrity means preventing unauthorized modification of information. Availability is the prevention of unauthorized withholding of information or resources.

Today, neither security nor its attributes, with the possible exception of availability [2], are easily measurable [10, 11, 25]. The security attributes on which to make measurements are in many cases even not defined or agreed on [13]. Furthermore, when security attributes have actually been defined and agreed on, as in the common criteria [3], the measures are often qualitative [5], i.e., based on experience, and do not carry sufficient information about its values to allow formal analysis. Hence, new ways of measuring security are therefore needed.

Two proposed quantitative confidentiality measures are entropy [23] and guesswork [20, 21]. When trying to break an encrypted message, entropy measures the average number of guesses in an optimal binary search attack, while guesswork measures the average number of guesses in an optimal linear search attack. An in-depth investigation of guesswork based on the findings in [15, 16] Is presented in this paper. The paper extends the definition of guesswork by defining joint and conditional guesswork. It is proved that joining increases guesswork, which implies that the joint guesswork is always at least equal to the marginal guesswork, and that conditioning reduces guesswork, which implies that the conditional guesswork is always at most equal to the marginal guesswork. The relationship and similarity between the extended definitions of guesswork and entropy, using the extended definitions of joint and conditional entropy [4], are briefly described. Unlike in the case of entropy, it is shown that guesswork does not possess the chain rule property.

Beyond entropy and guesswork, other means of measuring security have been proposed. An attempt to quantify security using game theory is described in [22]. A game theoretical method is also used in [17]. Quantifying operational security using state transition diagrams to model attacks and system restoration have been proposed in [12, 18, 24]. In [7], work on developing quantitative metrics for network security monitoring and evaluation is presented.

The remainder of the paper is organized as follows. Section II briefly treats measure theory, and Section III presents the measure entropy. Guesswork and its extending definitions, joint and conditional guesswork, are defined in Section IV. The section also introduces the concept of permutations that are necessary for extending the definition of guesswork. Using the concept of majorization,Section V proves that joining increases guesswork and that conditioning reduces guesswork. This section also shows that guesswork does not possess the chain rule property as does entropy. Finally, Section VI ends the paper with concluding remarks and future work.

II. A Note on Measure Theory

In measure theory [8], also called measurement theory, a measure is defined as a method or process for producing a value of an attribute or a characteristic of an entity that is then put on a scale. The values might be divided into qualitative or quantitative values. Qualitative values have a direct realization by means of a natural language description such as small, medium, and large, while quantitative values have an indirect realization by means of numbers, such as 5 or 42. Hence, for quantitative values to be understandable, more information is needed, which is added through units, for example, 5 meters or 42 degrees Celsius.

A. *Scales*

A scale is a set of values that corresponds to the range of the measure. Essentially, five major types of scales exist [8]:

- Nominal
- Ordinal
- Interval
- Ratio
- Absolute

The first two scales, nominal and ordinal, use qualitative values and the last three, interval, ratio and absolute, use quantitative values. The difference between the different scales lies in how much information they carry about the values. For nominal scales, entities can only be characterized into groups, while ordinal scales also include an order of the groups. For example, the color scale is nominal, while a scale consisting of the elements small, medium, and large is ordinal. In addition to the properties of ordinal scales, interval scales preserve differences between entities. In addition to the properties of interval scales, ratio scales also preserve ratios between entities. Thus, for interval scales, we can use addition and subtraction and for ratio scales we can also use multiplication and division. For example, the Celsius scale is an interval scale, since the difference between an arbitrary interval of one degree on the scale is preserved. However, the Celsius scale is not a ratio scale since it makes no sense to say that 20 degrees of Celsius is twice as hot as 10 degrees of Celsius. An example of a ratio scale is the Kelvin scale, since the measurement now starts at zero. Finally, absolute scales are unique, meaning that there is only one scale that can be used when measuring. Absolute scales allow all sorts of arithmetic analysis; an example is the counting scale.

B. *Set Theory*

In mathematics, measure theory [9] is defined from set theory or more exactly from the concept of σ-algebras. A σ-algebra $\Sigma(\mathcal{X})$ over a set $\mathcal{X}$ is a collection or family of subsets of $\mathcal{X}$ that is closed under complements and countable unions. That is, $\Sigma(\mathcal{X})$ is a subset of the power set $P(\mathcal{X})$ with the following two properties:

1. If $A \in \Sigma(\mathcal{X}) \Rightarrow A^C \in \Sigma(\mathcal{X})$
2. If $A_1, A_2, \ldots \in \Sigma(\mathcal{X}) \Rightarrow \cup_{i=1}^{\infty} A_i \in \Sigma(\mathcal{X})$

The sets in $\Sigma(\mathcal{X})$ can be seen as the possible states of an attribute. For example, if $\mathcal{X} = \{1, 2, 3\}$ then $\Sigma(\mathcal{X}) = \{\emptyset, \{1\}, \{2, 3\}\{1, 2, 3\}\}$ is a σ-algebra, and the corresponding attribute then has four states. The value of the states will depend on the measure, and from a σ-algebra a measure μ is defined as a mapping

$$\mu : \Sigma(\mathcal{X}) \rightarrow [0, \infty] \tag{1}$$

with the following two properties:

1. The empty set has a value of zero
$$\mu(\emptyset) = 0 \tag{2}$$
2. Countable additivity, if $\{A_i\}_1^{\infty}$ is a sequence of disjoint sets then
$$\mu\left(\cup_{i=1}^{\infty} A_i\right) = \cup_{i=1}^{\infty} \mu(A_i) \tag{3}$$

For example, if $\mathcal{X} = \mathbb{R}$ is the real line, then the set of all open intervals (a, b) is a σ-algebra and $\mu((a, b)) = b - a$ is a measure, i.e. the ordinary distance measure, which is referred to as the Borel measure.

III. Entropy

This section gives the formal definition of entropy and its generalizations in terms of joint and conditional entropy. The section also discusses some properties of entropy and states the chain rule of entropy.

A. *Marginal Entropy*

Entropy is usually, and somewhat carelessly, called a measure of uncertainty. However, when entropy was first defined in 1944 by Shannon [23] it was defined as the average amount of information of a (discrete) random variable X. Another interpretation of entropy is that it gives the average number of guesses of a discrete random variable X in an optimal binary search attack [15]. Before stating the formal definition of entropy, some fundamental probability terminology is needed to mathematically express the concept of discrete random variables. Let $\mathcal{X} = \{x_1, \ldots, x_n\}$ be a finite sample space. The mapping

$$\mathcal{P} : \mathcal{X} \rightarrow \mathbb{R} \tag{4}$$

is called a probability distribution if

$$\sum_{i=1}^{n} \mathcal{P}(x_i) = 1 \tag{5}$$

A $\mathcal{X}$-valued random variable X with probability distribution $\mathcal{P}$ is a variable that attains values $x_i \in \mathcal{X}$ with probability $\mathcal{P}(X = x_i)$. For some contexts it is necessary to point out that the random variable $\mathcal{X}$ is connected with probability distribution $\mathcal{P}$. This is written $\mathcal{X}_{\mathcal{P}}$. Furthermore, to shorten the notation, $\mathcal{P}(X = x_i) = \mathcal{P}_i$. Hence, the joint probability distribution of a pair of random variables (X_0, X_1) is written $\mathcal{P}_{ij} = p(X_0 = x_i, X_1 = x_j)$, and the conditional probability distribution of X_1 given X_0 is written $\mathcal{P}_{j|i} = p(X_1 = x_j | X_0 = x_i)$. Note that $\mathcal{P}_{ij} = \mathcal{P}_i \mathcal{P}_{j|i}$.

Definition 1. *The entropy $H(X)$ of a random variable X with probability distribution $\mathcal{P}_i$ is defined as*

$$H(X) = -\sum_{i=1}^{n} \mathcal{P}_i \log_2 \mathcal{P}_i \tag{6}$$

In both computer science and information theory the base of the logarithm is taken to be two, measured in bits, and in mathematics and physics the base of the logarithm is taken to be e, measured in nats. The minimum value of entropy is zero, obtained for the deterministic probability distribution, and the maximum value of entropy is obtained for the uniform probability distribution $\mathcal{U}$, with $H(X_{\mathcal{U}}) = \log_2 n$ [4].

B. Joint Entropy

Definition 1 can be extended to joint entropy [4]. The joint entropy $H(X_0, X_1)$ gives the entropy of a pair of random variables (X_0, X_1) with a joint probability distribution $\mathcal{P}_{ij}$.
Definition 2. *The joint entropy $H(X_0, X_1)$ of a pair of random variables (X_0, X_1) with joint probability distribution $\mathcal{P}_{ij}$ is defined as*

$$H(X_0, X_1) = -\sum_{i=1}^{n}\sum_{j=1}^{n} \mathcal{P}_{ij} \log_2 \mathcal{P}_{ij} \tag{7}$$

C. Conditional Entropy

Definition 1 can also be extended to conditional entropy. The conditional entropy $H(X_1|X_0)$, or equivocation, gives the remaining entropy of a random variable X_1 given another random variable X_0 with conditional probability distributions $\mathcal{P}_{j|i}$. Conditional entropy is defined as the average value of the entropies of the conditional distributions, averaged over the conditioning random variable.
Definition 3. *The conditional entropy $H(X_1|X_0)$ of a random variable X_1 given another random variable X_0 with conditional probability distributions $\mathcal{P}_{j|i}$ is defined as*

$$\begin{aligned} H(X_1|X_0) &= \sum_{i=1}^{n} \mathcal{P}_i H(X_1|X_0 = x_i) \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{n} \mathcal{P}_{ij} \log_2 \mathcal{P}_{j|i} \end{aligned} \tag{8}$$

D. Properties of Entropy

In [4], it is shown that entropy has two important properties expressed through the following inequality

$$H(X_0|X_1) \leq H(X_0) \leq H(X_0, X_1) \tag{9}$$

The first property, $H(X_0|X_1) \leq H(X_0)$, states that conditioning reduces entropy. Hence, by gaining information, the average number of guesses needed to find the value of a random variable decreases. The second property, $H(X_0) \leq H(X_0, X_1)$, states that joining increases entropy. Hence, by adding information, the average number of guesses needed to find the value will increase. Furthermore, the marginal, joint and conditional entropies are related by the chain rule

$$H(X_0, X_1) = H(X_0) + H(X_1|X_0) \tag{10}$$

This rule states that the joint entropy is equal to the marginal entropy plus the corresponding conditional entropy. Hence, by using a binary search attack, the average number of guesses needed to find the value of a pair of random variables is equal to the average number of guesses needed to find the value of a marginal random variable plus the average number of guesses needed to find the value of the other marginal random variable, conditioned with the first chosen marginal random variable. By generalizing Definition 2 and 3 to n random variables [4] the chain rule generalizes to

$$H(X_0, \ldots, X_{n-1}) = H(X_0) + \sum_{i=1}^{n-1} H(X_i|X_0, \ldots, X_{i-1}) \tag{11}$$

IV. Guesswork

The formal definition of guesswork and its generalizations in terms of joint and conditional guesswork is defined in this section. The section also introduces the concept of permutations that are necessary for extending the definition of guesswork.

A. Marginal Guesswork

Guesswork is a measure that gives the average number of guesses of a random variable X in an optimal linear search attack [15], which is also referred to as an optimal brute force attack. In such an attack, the attacker is assumed to have complete knowledge of the probability distribution $\mathcal{P}_i$ of X. Thus, before starting the guessing process, the attacker can arrange all values of $\mathcal{X}$ in a non-increasing probability order

$$\mathcal{P}_1 \geq \mathcal{P}_2 \geq \ldots \geq \mathcal{P}_n \tag{12}$$

From (12), the measure guesswork is defined as follows.
Definition 4. *The guesswork $W(X)$ of a random variable X with probability distribution $\mathcal{P}_i$ that is ordered according to (12) is defined as*

$$W(X) = \sum_{i=1}^{n} i\mathcal{P}_i \tag{13}$$

Similar to entropy, the minimum value of guesswork is one, obtained for the deterministic probability distribution, and the maximum value of guesswork is obtained for the uniform probability distribution $\mathcal{U}$, with $W(X_{\mathcal{U}}) = \frac{n+1}{2}$ [21].
The last term in the sum of guesswork, see Definition 4, is weighted with n. This is not completely correct, however,since the last guess in the guessing process decides for the last two values of the random variable [15]. That is, if the answer to the last question is “yes” then the correct value is $X = x_{n-1}$, and the search terminates. If on the other hand, the answer is “no” the correct value is $X = x_n$, and the search terminates. Thus, the last term in the sum of guesswork should be weighted with $n-1$, and the redefined guesswork becomes

$$W'(X) = W(X) - p_n \tag{14}$$

Since all properties of the original guesswork shown $W(X)$ also hold for the redefined guesswork $W'(X)$, as will be proved in Proposition 2, $W(X)$ will for simplicity be used in this paper instead of $W'(X)$.

B. Permutations

To describe the ordering of the probability distribution according to (12) in a more formal way, the concept of permutation [6] is used. A permutation σ is a mapping

$$\sigma : S \to S \tag{15}$$

of a non-empty set S that is bijective. Hence, permutations of a set are all invertible mappings from the set onto itself. The set of all permutations on S together with composition as the operation forms the symmetric group Θ_S. If $S = \{1, \dots, n\}$, which is denoted Θ_n. An arbitrary element $\sigma \in \Theta_4$, which is a permutation on the numbers $\{1, 2, 3, 4\}$, can be written as

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ \sigma(1) & \sigma(2) & \sigma(3) & \sigma(4) \end{pmatrix} \tag{16}$$

The mapping $\sigma(i)$ gives how all values of the set $\{1, 2, 3, 4\}$ are permuted. Note that $\sigma(i)$ must be one-to-one and onto to be invertible and, hence, a permutation.

An unordered probability distribution $\mathcal{P}_i$ can be ordered in a non-increasing probability order, as in (12), by using a permutation $\sigma(i)$ that maps the index of the largest value to one, the index of the second largest value to two, and so on until the index of the smallest value is mapped to n. Thus,

$$\sigma\mathcal{P}_i = (\mathcal{P}_{\sigma^{-1}(1)}, \mathcal{P}_{\sigma^{-1}(2)}, \dots, \mathcal{P}_{\sigma^{-1}(n)}) \tag{17}$$

where

$$\mathcal{P}_{\sigma^{-1}(1)} \geq \mathcal{P}_{\sigma^{-1}(2)} \geq \dots \geq \mathcal{P}_{\sigma^{-1}(n)} \tag{18}$$

Note that if $\sigma(i) = k$, then

$$k\mathcal{P}_{\sigma^{-1}(k)} = \sigma(i)\mathcal{P}_i \tag{19}$$

Thus, instead of using the terms in (18) to calculate the sum in Definition 4, the same result is achieved if the probability distribution $\mathcal{P}_i$ is left unordered and the i values are instead changed to $\sigma(i)$. Hence, Definition 4 can be redefined as follows.

Definition 5. *The guesswork $W(X)$ of a random variable X with probability distribution $\mathcal{P}_i$ that is ordered in a non-increasing probability order by the permutation $\sigma(i)$ is defined as*

$$\begin{aligned} W(X) &= \sum_{i=1}^{n} i\mathcal{P}_{\sigma^{-1}(i)} \\ &= \sum_{i=1}^{n} \sigma(i)\mathcal{P}_i \end{aligned} \tag{20}$$

C. Joint Guesswork

Definition 5 can be extended to joint guesswork. The joint guesswork $W(X_0, X_1)$ gives the guesswork of a pair of random variables (X_0, X_1) with a joint probability distribution $\mathcal{P}_{ij}$.

Definition 6. *The joint guesswork $W(X_0, X_1)$ of a pair of random variables (X_0, X_1) with joint probability distribution $\mathcal{P}_{ij}$ that is ordered in a non-increasing probability order by the permutation $\pi(i, j)$ is defined as*

$$W(X_0, X_1) = \sum_{i=1}^{n}\sum_{j=1}^{n} \pi(i, j)\mathcal{P}_{ij} \tag{21}$$

By using the inverse permutation $\pi^{-1}(k) = (i, j)$, the ordered joint probability distribution of $\mathcal{P}_{ij}$ becomes

$$\pi\mathcal{P}_{ij} = (\mathcal{P}_{\pi^{-1}(1)}, \mathcal{P}_{\pi^{-1}(2)}, \dots, \mathcal{P}_{\pi^{-1}(n^2)}) \tag{22}$$

and the joint guesswork can hence be rewritten as

$$W(X_0, X_1) = \sum_{i=1}^{n^2} i\mathcal{P}_{\pi^{-1}(i)} \tag{23}$$

For example, using the joint probability distribution given in Table 1, the marginal guesswork of X_0 becomes

$$\begin{aligned} W(X_0) &= \sigma(1)\mathcal{P}_1 + \sigma(2)\mathcal{P}_2 \\ &= 2 * 0.4 + 1 * 0.6 \\ &= 1.4 \end{aligned} \tag{24}$$

and the joint guesswork

$$\begin{aligned} W(X_0, X_1) &= \pi(1, 1)\mathcal{P}_{11} + \pi(2, 1)\mathcal{P}_{21} \\ &\quad + \pi(1, 2)\mathcal{P}_{12} + \pi(2, 2)\mathcal{P}_{22} \\ &= 2 * 0.3 + 3 * 0.2 + 4 * 0.1 + 1 * 0.4 \\ &= 2.0 \end{aligned} \tag{25}$$

Table 1: The joint probability distribution used to calculate the marginal guesswork, $W(X_0)$, the conditional guesswork, $W(X_1|X_0)$, and the joint guesswork, $W(X_0, X_1)$.

$\mathcal{P}_{ij}$	$X_1 = x_1$	$X_1 = x_2$
$X_0 = x_1$	0.3	0.1
$X_0 = x_2$	0.2	0.4

D. Conditional Guesswork

Definition 5 can also be extended to conditional guesswork. The conditional guesswork $W(X_1|X_0)$ gives the remaining guesswork of the random variable X_1 given the random variable X_0 with a conditional probability distribution $\mathcal{P}_{j|i}$. Conditional guesswork is defined as the average value of the guessworks of the conditional distributions, averaged over the conditioning random variable.

Definition 7. *The conditional guesswork $W(X_1, X_0)$ of a random variable X_1 given another random variable X_0 with conditional probability distributions $\mathcal{P}_{j|i}$ that are ordered in a non-increasing probability order by the permutations $\rho_i(j)$ is defined as*

$$\begin{aligned} W(X_1|X_0) &= \sum_{i=1}^{n} \mathcal{P}_i W(X_1|X_0 = x_i) \\ &= \sum_{i=1}^{n} \mathcal{P}_i \sum_{j=1}^{n} \rho_i(j)\mathcal{P}_{j|i} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \rho_i(j)\mathcal{P}_{ij} \end{aligned} \tag{26}$$

Note that the permutations $\rho_i(j)$, in addition to ordering the conditional probability distributions $\mathcal{P}_{j|i}$, also order the joint probability distribution $\mathcal{P}_{ij}$ with respect to j. Hence, by using the inverse permutation $\rho_i^{-1}(k) = (i, j)$, the ordered joint probability distribution of $\mathcal{P}_{ij}$ with respect to j becomes

$$\rho_i \mathcal{P}_{ij} = (\mathcal{P}_{\rho_i^{-1}(1)}, \mathcal{P}_{\rho_i^{-1}(2)}, \ldots, \mathcal{P}_{\rho_i^{-1}(n)}) \tag{27}$$

From this, the joint guesswork can be rewritten as

$$W(X_1|X_0) = \sum_{i=1}^{n} \sum_{j=1}^{n} j \mathcal{P}_{\rho_i^{-1}(j))} \tag{28}$$

For example, using the joint probability distribution given in Table 1, the conditional guesswork becomes

$$\begin{aligned} W(X_1|X_0) &= \rho_1(1)\mathcal{P}_{11} + \rho_1(2)\mathcal{P}_{12} \\ &\quad + \rho_2(1)\mathcal{P}_{21} + \rho_2(2)\mathcal{P}_{22} \\ &= 1 * 0.3 + 2 * 0.1 + 2 * 0.2 + 1 * 0.4 \\ &= 1.3 \end{aligned}$$

In this example, $W(X_0|X_1) = W(X_1|X_0)$. In most cases, however, $W(X_1|X_0) \neq W(X_0|X_1)$.

V. Properties of Guesswork

The relationship between the marginal, joint and conditional guesswork is Investigated in this section by using the concept of majorizations [21]. In [1, 19], the logarithmic guesswork rate

$$W(\mathbb{X}) = \lim_{n \to \infty} \frac{log_a W(X_1, \ldots, X_n)}{n} \tag{29}$$

was investigated for the independent case and for Markov chains, respectively. The behavior in the finite case is still an open research issue.

A. Majorization

Ordered vectors are used in the theory of majorization. Furthermore, since discrete probability distributions can be represented as a vector, the probability distributions will be used instead.

Definition 8. *Let $\mathcal{P}_i$ and $\mathcal{Q}_i$ be two probability distributions ordered in a non-increasing probability order. Then $\mathcal{P}_i$ is majorized by $\mathcal{Q}_i$ if $\forall k : 1 \leq k < n$*

$$\sum_{i=1}^{k} \mathcal{P}_i \leq \sum_{i=1}^{k} \mathcal{Q}_i \tag{30}$$

and

$$\sum_{i=1}^{n} \mathcal{P}_i = \sum_{i=1}^{n} \mathcal{Q}_i \tag{31}$$

Following the notation of Pliam [21], $\mathcal{P}_i \preccurlyeq \mathcal{Q}_i$ is used when $\mathcal{P}_i$ is majorized by $\mathcal{Q}_i$. Further, Pliam used the concept of majorization to prove the following proposition.

Proposition 1. *If $\mathcal{P}_i \preccurlyeq \mathcal{Q}_i$ then*

$$W(X_{\mathcal{P}}) \geq W(X_{\mathcal{Q}}) \tag{32}$$

This proposition can be extended to also hold for $W'(X)$.

Proposition 2. *If $\mathcal{P}_i \preccurlyeq \mathcal{Q}_i$ then*

$$W'(X_{\mathcal{P}}) \geq W'(X_{\mathcal{Q}}) \tag{33}$$

Proof. Let

$$\mathcal{P}'_i = \begin{cases} \mathcal{P}_i & 1 \leq i \leq n-2 \\ \mathcal{P}_{n-1} + \mathcal{P}_n & i = n-1 \end{cases} \tag{34}$$

and

$$\mathcal{Q}'_i = \begin{cases} \mathcal{Q}_i & 1 \leq i \leq n-2 \\ \mathcal{Q}_{n-1} + \mathcal{Q}_n & i = n-1 \end{cases} \tag{35}$$

For the two probability distributions $\mathcal{P}'_i$ and $\mathcal{Q}'_i$, (30) holds $\forall k : 1 \leq k \leq n-2$, and (31) holds for $k = n-1$. Hence, $\mathcal{P}'_i \preccurlyeq \mathcal{Q}'_i$, and by using Proposition 1

$$W(X_{\mathcal{P}'}) \geq W(X_{\mathcal{Q}'}) \tag{36}$$

However, since $W(X_{\mathcal{P}'}) = W'(X_{\mathcal{P}})$ and $W(X_{\mathcal{Q}'}) = W'(X_{\mathcal{Q}})$, the proposition holds. □

Hence, using majorizations, all properties of $W(X)$ shown will therefore also hold for $W'(X)$.

B. Guesswork and Joint Guesswork

A theorem stating that the joint guesswork is always at least equal to the marginal guesswork will now be proved. This then proves that joining increases guesswork.

Theorem 1. *Let X_0 and X_1 be two random variables. Then,*

$$W(X_0) \leq W(X_0, X_1) \tag{37}$$

Proof. From Definition 5, $W(X_0)$ has the ordered probability distribution

$$\sigma \mathcal{P}_i = (\mathcal{P}_{\sigma^{-1}(1)}, \mathcal{P}_{\sigma^{-1}(2)}, \ldots, \mathcal{P}_{\sigma^{-1}(n)}) \tag{38}$$

and from Definition 6, $W(X_0, X_1)$ has the ordered probability distribution

$$\pi \mathcal{P}_{ij} = (\mathcal{P}_{\pi^{-1}(1)}, \mathcal{P}_{\pi^{-1}(2)}, \ldots, \mathcal{P}_{\pi^{-1}(n^2)}) \tag{39}$$

For the two probability distributions to be of the same length, $\sigma \mathcal{P}_i$ is extended with $n^2 - n$ zeros at the end. Furthermore, since $\mathcal{P}_i = \sum_{j=1}^{n} \mathcal{P}_{ij}$, each of the non-zero components in $\sigma \mathcal{P}_i$ is the sum of n components in $\pi \mathcal{P}_{ij}$.

The largest component of $\pi \mathcal{P}_{ij}$ can either be contained in or not contained in the largest component of $\sigma \mathcal{P}_i$. In the first case, obviously, $\mathcal{P}_{\pi^{-1}(1)} \leq \mathcal{P}_{\sigma^{-1}(1)}$. In the second case, the same must hold in order not to contradict that $\mathcal{P}_{\sigma^{-1}(1)}$ is the largest component of $\sigma \mathcal{P}_i$.

Using the same reasoning as above, if the k first components of $\pi \mathcal{P}_{ij}$ are contained in the k first components of $\sigma \mathcal{P}_i$, the sum of the k first components of $\pi \mathcal{P}_{ij}$ is smaller than the sum of the k first components of $\sigma \mathcal{P}_i$. If a set of the k first components of $\pi \mathcal{P}_{ij}$ is not contained in the k first components of $\sigma \mathcal{P}_i$, however, then each of the components in the set must be at least as small as $\mathcal{P}_{\sigma^{-1}(k)}$ in order not to contradict that $\mathcal{P}_{\sigma^{-1}(k)}$ is the kth largest component in $\sigma\mathcal{P}_i$. Thus, $\forall k, 1 \leq k < n^2$

$$\sum_{i=1}^{k} \mathcal{P}_{\pi^{-1}(i)} \leq \sum_{i=1}^{k} \mathcal{P}_{\sigma^{-1}(i)} \tag{40}$$

and

$$\sum_{i=1}^{n^2} \mathcal{P}_{\pi^{-1}(i)} = \sum_{i=1}^{n^2} \mathcal{P}_{\sigma^{-1}(i)} \tag{41}$$

Hence, $\pi\mathcal{P}_{ij} \preccurlyeq \sigma\mathcal{P}_i$, and by Proposition 1 $W(X_0) \leq W(X_0, X_1)$. □

As an illustration, the joint guesswork for the joint probability distribution provided in Table 2 is plotted in Figure 1.

Table 2: The joint probability for Figure 1.

$\mathcal{P}_{ij}$	$X_1 = x_1$	$X_1 = x_2$
$X_0 = x_1$	a	$0.6 - a$
$X_0 = x_2$	b	$0.4 - b$

For this distribution $W(X_0, X_1) \geq W(X_0)$ and $W(X_0) = 1.4$ are independent of the values of a and b. Furthermore, there are four points that set one of the elements in each row to zero in the joint distribution, $(a, b) = (0, 0)$, $(a, b) = (0.6, 0)$, $(a, b) = (0, 0.4)$, and $(a, b) = (0.6, 0.4)$. Hence, for these points, $W(X_0, X_1) = W(X_0)$. The local maximums that occur on the surface correspond to when elements of the columns have $a = b$ or when elements of the rows have $a = 0.3$ or $b = 0.2$. Thus, a uniform conditional probability distribution then exists for the row or column. The global maximum, $W(X_0, X_1) = 2.3$, is achieved at point $(a, b) = (0.3, 0.2)$, which is when the two rows have a uniform probability distribution. This is as close to the real uniform probability distribution as this probability distribution gets.

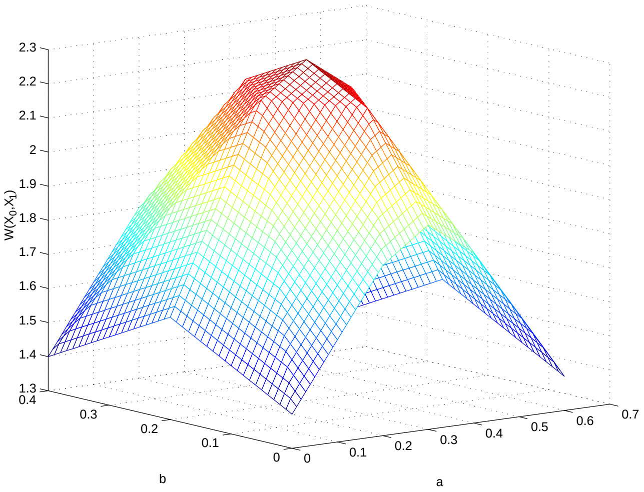

Figure. 1: Joint guesswork $W(X_0, X_1)$, for the joint distribution provided in Table 2. For this distribution, $W(X_0) = 1.4$.

Figure 2 plots the range of the joint guesswork $W(X_0, X_1)$ versus the marginal guesswork $W(X_0)$ using the joint probability distribution given in Table 3. That is, the range of $W(X_0, X_1)$ is plotted for different values of $\mathcal{P}_1$ and $\mathcal{P}_2$ giving $W(X_0)$. By assumption, $\mathcal{P}_1 \geq \mathcal{P}_2$, hence, the possible values of the probabilities, are bounded by the inequality $0 \leq \mathcal{P}_2 \leq 0.5 \leq \mathcal{P}_1 \leq 1$. The change in values of the probabilities were chosen in steps of 0.01. Note that, by setting $\mathcal{P}_1 = 0.6$, and $\mathcal{P}_2 = 0.4$, Table 3 is transformed to Table 2, giving $W(X_0) = 1.4$, and $1.4 \leq W(X_0, X_1) \leq 2.3$.

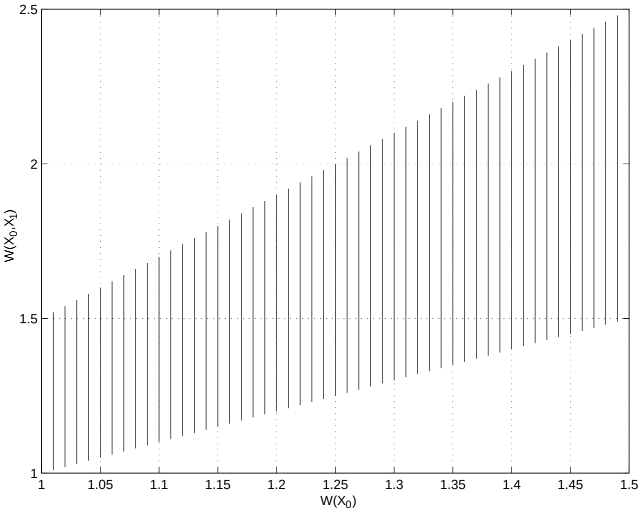

Figure. 2: The range of the joint guesswork $W(X_0, X_1)$ versus the marginal guesswork $W(X_0)$ for the joint distribution provided in Table 3.

Table 3: The joint probability distribution for Figure 2.

$\mathcal{P}_{ij}$	$X_1 = x_1$	$X_1 = x_1$
$X_0 = x_1$	a	$p(x_1) - a$
$X_0 = x_2$	b	$p(x_2) - b$

In Figure 2, the maximum value, $W(X_0, X_1) = 2.5$, is achieved when both the random variables have a uniform probability distribution. The minimum value, $W(X_0, X_1) = 1$, is achieved when both random variables are known. Furthermore, $W(X_0, X_1)$ has four bounds.

- The lower bound, X_1, is known. Hence, $W(X_0, X_1) = W(X_0)$.
- The upper bound, X_1, has a uniform probability distribution.
- The left bound, X_0, is known. Hence, $W(X_0, X_1) = W(X_1)$.
- The right bound, X_0, has a uniform probability distribution.

C. *Guesswork and Conditional Guesswork*

A theorem stating that the conditional guesswork is always at most equal to the marginal guesswork will next be proved. Hence, this proves that conditioning reduces guesswork.

Theorem 2. *Let X_0 and X_1 be two random variables. Then,*

$$W(X_0|X_1) \leq W(X_0) \tag{42}$$

Proof. From Definition 5, $W(X_0)$ has the ordered probability distribution

$$\sigma\mathcal{P}_i = (\mathcal{P}_{\sigma^{-1}(1)}, \mathcal{P}_{\sigma^{-1}(2)}, \dots, \mathcal{P}_{\sigma^{-1}(n)}) \tag{43}$$

And, from Definition 26, $W(X_0|X_1)$ has for each j the ordered probability distribution

$$\rho_j\mathcal{P}_{ij} = (\mathcal{P}_{\rho_j^{-1}(1)}, \mathcal{P}_{\rho_j^{-1}(2)} \dots, \mathcal{P}_{\rho_j^{-1}(n)}) \tag{44}$$

Now, by summing over j, an ordered probability distribution is created.

$$\mathcal{P}'_i = \sum_{j=1}^{n} \rho_j \mathcal{P}_{ij} \tag{45}$$

The distribution is ordered since the first component is the sum of the n largest probabilities, having $\rho_j(i) = 1$, the second component is the sum of the n second largest probabilities, having $\rho_j(i) = 2$, and so on. Thus, $\forall k, 1 \leq k < n$

$$\sum_{i=1}^{k} \mathcal{P}_{\sigma^{-1}(i)} \leq \sum_{i=1}^{k} \mathcal{P}'_i \tag{46}$$

and

$$\sum_{i=1}^{n} \mathcal{P}_{\sigma^{-1}(i)} = \sum_{i=1}^{n} \mathcal{P}'_i \tag{47}$$

Hence, $\sigma\mathcal{P}_i \preccurlyeq \mathcal{P}'_i$, and by Proposition 1, $W(X_0) \geq W(X_0|X_1)$. □

If the random variables are independent, then $\rho_j(i)$ are equal for each j in permuting the random variable X_0. Furthermore, $\rho_j(i) = \sigma(i)$, hence

$$\begin{aligned}\mathcal{P}'_i &= \sum_{j=1}^{n} \rho_j \mathcal{P}_{ij} \\ &= \sum_{j=1}^{n} \sigma \mathcal{P}_{ij} \\ &= \sigma\mathcal{P}_i\end{aligned} \tag{48}$$

Thus, $W(X_0|X_1) = W(X_0)$ if the random variables are independent.

The conditional guesswork using the probability distribution provided in Table 2 is plotted in Figure 3. In the figure, $W(X_0|X_1) \leq W(X_0)$, since $W(X_0)$ is a plane with height 1.4. The plateau of the surface of $W(X_0|X_1)$ is due to the fact that no reordering is made in the distribution before calculating the conditional guesswork. Hence, under those conditions, the sum of each column is unchanged giving no change in the value of the conditional guesswork. Furthermore, the two points $(a, b) = (0.6, 0)$ and $(a, b) = (0, 0.4)$ set the elements to zero on the diagonals, which means that there is only one value to guess on for each row. Hence, for these two points, $W(X_0|X_1) = 1$.

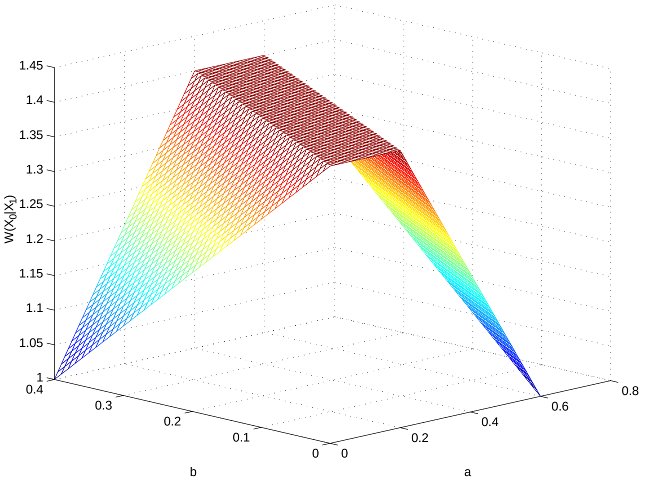

Figure. 3: Conditional guesswork, $W(X_0|X_1)$, for the joint distribution is provided in Table 2. For this distribution, $W(X_0) = 1.4$.

D. The Chain Rule

For entropy, the chain rule (10) relates the marginal, joint and conditional entropies. However, as will now be shown by a counterexample using the uniform probability distribution, guesswork does not possess the chain rule.

The marginal guesswork for the uniform probability distribution $\mathcal{U}_i = \frac{1}{n}$ is

$$W(X_0) = \frac{n+1}{2} \tag{49}$$

In the same way, the joint guesswork for the uniform probability distribution $\mathcal{U}_{ij} = \frac{1}{n^2}$ is

$$W(X_0, X_1) = \frac{n^2+1}{2} \tag{50}$$

Note that this distribution has n^2 elements instead of n. Finally, the conditional guesswork for the uniform probability distribution $\mathcal{U}_{ij}$ is

$$\begin{aligned}W(X_1|X_0) &= n\left(1\frac{1}{n^2} + 2\frac{1}{n^2} + \dots + n\frac{1}{n^2}\right) \\ &= \frac{n+1}{2}\end{aligned} \tag{51}$$

Hence, $W(X_0) + W(X_1|X_0) = n + 1$.

In Figure 4, the joint guesswork, the sum of the marginal guesswork and the conditional guesswork for the uniform probability distribution are plotted as a function of n. The values of n should be integers. However, for the illustration, the graphs are made continuous.

In the figure, $W(X_0, X_1) \leq W(X_0) + W(X_1|X_0)$ when $n \leq 1 + \sqrt{2}$, and $W(X_0, X_1) \geq W(X_0) + W(X_1|X_0)$ when $n \geq 1 + \sqrt{2}$. Hence, while $n \leq 1 + \sqrt{2}$ it is easier to guess at the two joint random variables, and while $n \geq 1 + \sqrt{2}$ it is easier to split the guesses into two parts. For entropy, we have for the uniform distribution, $H(X_0, X_1) = log_2(n^2)$, and $H(X_0) + H(X_1|X_0) = log_2(n) + log_2(n) = log_2(n^2)$.

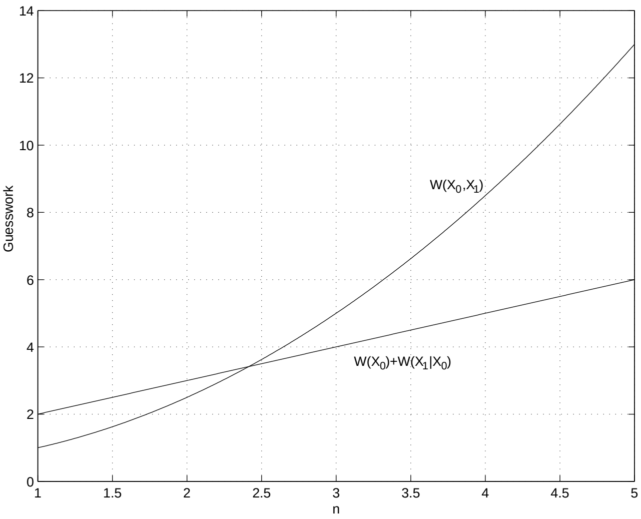

Figure. 4: The joint guesswork, the sum of the marginal guesswork and the conditional guesswork for the uniform probability distribution are plotted as a function of n.

VI. Conclusions and Future Work

This paper has defined joint and conditional guesswork. We have also shown that the following inequality holds

$$W(X_0|X_1) \leq W(X_0) \leq W(X_0, X_1) \tag{52}$$

The first part of the inequality states that conditioning reduces guesswork. Hence, if information is gained, then the average number of guesses needed to find the value of a random variable decreases. The second part of the inequality states that joining increases guesswork. Hence, the average number of guesses needed to find the values of joint random variables is always at least equal to the average number of guesses needed to find the values of the marginal random variables. Thus, guesswork possesses the same two properties as entropy, i.e. joint entropy is always at least equal to the marginal entropy, and conditioning reduces entropy. In contrast to entropy, it has been shown that guesswork does not possess the chain rule property by using the uniform probability distribution.

The goal of our future work is to further investigate $W(X_0)$, $W(X_1|X_0)$ and $W(X_0, X_1)$ with the aim of relating them in an expression similar to the chain rule for entropy. Such a finding will provide a better understanding of guesswork and is thus the natural next step in building up a theory of guesswork. The security implication of the generic selective encryption scheme presented in [14] will also be investigated using the results presented in this paper. In particular, languages of different orders as well as the size and distribution of encryption units should be investigated in order to understand the full impact of selective encryption.

Acknowledgment

The work was carried out within the Compare Business Innovation Centre phase 2 (C-BIC 2) project, funded partly by the European Regional Development Fund (ERDF).

References

[1] E. Arikan. An inequality on guessing and its application to sequential decoding. *IEEE Transactions on Information Theory*, 42(1):99–105, 1996.

[2] A. Avižienis, J.-C. Laprie, B. Randell, and C. E Landwehr. Basic concepts and taxonomy of dependability and secure computing. *IEEE Transactions on Dependable and Secure Computing*, 1(1):11–33, January-March 2004.

[3] Common Criteria Implementation Board. Common criteria for information technology security evaluation, version 3.1. http://www.commoncriteriaportal.org/, September 2006.

[4] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, NY, USA, 1991.

[5] W. de Bruijn, M. R. Spruit, and M. van den Heuvel. Identifying the cost of security. *Journal of Information Assurance and Security*, 5(1):074–083, 2010.

[6] J. R. Durbin. *Modern Algebra an Introduction*. John Wiley & Sons, Inc., New York, NY, USA, 1992.

[7] F. El-Hassan, A. Matrawy, N. Seddigh, and B. Nandy. An experimental approach to network monitoring using quantitative security metrics. *Journal of Information Assurance and Security*, 6(1):048–062, 2011.

[8] N. E. Fenton and S. L. Pfleeger. *Software Metrics: A Rigorous & Practical Approach*. PWS Publishing, 2nd edition, 1997.

[9] G. B. Folland. *Real Analysis, Modern Techniques and Their Applications*. John Wiley & Sons, New York, NY, USA, 1999.

[10] D. S. Herrmann. *Complete Guide to Security and Privacy Metrics*. Auerbach Publications, 2007.

[11] A. Jaquith. *Security Metrics: Replacing Fear, Uncertainty, and Doubt*. Addison-Wesley, 2007.

[12] S. Jha, O. Sheyner, and J. Wing. Two formal analyses of attack graphs. In *Proceedings of the 15th IEEE Computer Security Foundations Workshop (CSFW 2002)*, page 49, Cape Breton, Nova Scotia, Canada, June 24–26, 2002.

[13] S. Lindskog and E. Jonsson. Adding security to QoS architectures. In R. Burnett, A. Brunstrom, and A. G. Nilsson, editors, *Perspectives on Multimedia: Communication, Media and Information Technology*, chapter 8, pages 145–158. John Wiley & Sons, West Sussex, England, 2003.

[14] S. Lindskog, R. Lundin, and A. Brunstrom. Middleware support for tunable encryption. In *Proceedings of the 5th International Workshop on Wireless Information Systems (WIS 2006)*, May 23, 2006.

[15] R. Lundin, T. Holleboom, and S. Lindskog. On the relationship between confidentiality measures: Entropy and guesswork. In *Proceedings of the 5th International Workshop on Security in Information Systems (WOSIS 2007)*, June 12–13, 2007.

[16] R. Lundin, S. Lindskog, A. Brunstrom, and S. Fischer-Hübner. Using guesswork as a measure for confidentiality of selectively encrypted messages. In D. Gollmann, F. Massacci, and A. Yautsiukhin, editors, *Quality of Protection: Security Measurements and Metrics*, volume 23, pages 173–184. Springer, NY, USA, 2006.

[17] K.-w. Lye and J. Wing. Game strategies in network security. In *Proceedings of Foundations of Computer Security*, Copenhagen, Denmark, July 25–26, 2002.

[18] B. B. Madan, K. Goševa-Popstojanova, K. Vaidyanathan, and K. S. Tivedi. A method for modeling and quantifying the security attributes of intrusion tolerant systems. *Performance Evaluation*, 56(1–4):167–186, March 2004.

[19] D. Malone and W. G. Sullivan. Guesswork and entropy. *IEEE Transactions on Information Theory*, 20(3):525–526, 2004.

[20] J. Massey. Guessing and entropy. In *Proceedings of the 1994 IEEE International Symposium on Information Theory*, page 204, Trondheim, Norway, 1994.

[21] J. O. Pliam. *Ciphers and their Products: Group Theory in Private Key Cryptography*. PhD thesis, University of Minnesota, MN, USA, 1999.

[22] K. Sallhammar and S. J. Knapskog. Using game theory in stochastic models for quantifying security. In *Proceedings of the Ninth Nordic Workshop on Secure IT Systems (NordSec 2004)*, Espoo, Finland, November 4–5, 2004.

[23] C. E. Shannon. *Claude Elwood Shannon: Collected Papers*. IEEE Press, Piscataway, NJ, USA, 1993.

[24] F. Stevens, T. Courtney, S. Singh, A. Agbaria, J. F. Meyer, W. H. Sanders, and P. Pal. Model-based validation of an intrusion-tolerant information system. In *Proceedings of the 23rd IEEE International Symposium on Reliable Distributed Systems (SRDS 2004)*, pages 184–194, Florianpolis, Brazil, October 18–20, 2004.

[25] V. Verendel. Quantified security is a weak hypothesis. In *New Security Paradigms Workshop (NSPW)*, pages 227–233, Oxford, UK, September 8-11, 2009.

Author biographies

Reine Lundin received his Licentiate degree in Computer Science from Karlstad University, Sweden, in 2007. He also received a Master's Degree in Physics and a Master's degree in Mathematics from Karlstad University in 1999 and 2003, respectively. He joined the Department of Computer Science at Karlstad University in 2000, where he is currently working as a lecturer. His research focus is quantitative security metrics and tunable security services. He has authored/coauthored 15 book chapters and conference papers.

Stefan Lindskog received his Licentiate and PhD degrees in Computer Engineering from Chalmers University of Technology, Göteborg, Sweden in 2000 and 2005, respectively. In 2008, he received the Docent degree in Computer Science at Karlstad University, Sweden. He joined the Department of Computer Science at Karlstad University in 1990, where he is currently a full professor. His research focus is the design of tunable and adaptable security services and security and performance analysis of security services and protocols. He has authored/coauthored one textbook, eight book chapters, and over 45 journal and conference papers.