Journal of Information Assurance and Security.
ISSN 1554-1010 Volume 8 (2013) pp. 086-096

On Entropy of Selectively Encrypted Bitmap Images using Information Neighborhoods

Reine Lundin and Stefan Lindskog

Department of Computer Science
Karlstad University
Sweden
Reine.Lundin@kau.se

Abstract:
Selective encryption is a concept in which the main goal is to reduce computational cost while providing confidentiality by encrypting only chosen parts of the information to be protected. Previous work on selective encryption has mainly been aimed towards multimedia applications in order to reduce the overhead induced by encryption while still making the information perceptually secure to a desired protection level. This was accomplished by utilizing the fact that different parts of the information have different impacts on our perception senses, i.e., eyes and ears. How computationally secure the information is when using selective encryption has however only briefly been mentioned or rudimentarily analyzed. In this paper, we therefore investigate the security implications of selective encryption by generalizing the work on entropy of selectively encrypted strings to several dimensions and applying it to bitmap images. The generalization is done by constructing information neighborhoods from the order of languages concept and cellular automata theory to capture and model information dependencies in several dimensions.

Keywords: computer security, security measures, selective encryption, entropy, confidentiality, bitmap images, cellular automata.

I. Introduction

When information is protected, a computational cost is imposed on the computing environment. For small computing devices with restricted resources, such as mobile phones or sensor network devices, the security mechanisms that are used can put a significant extra burden on performance and/or energy consumption. The concept of selective encryption may be used to reduce computational cost when providing confidentiality, which provides confidentiality by encrypting only chosen parts of the information. Selective encryption can also be used to trade confidentiality against computational cost by altering the distribution of encrypted parts. The authors in [7] presented a generic selective encryption model. In the proposed model, the information, I, is divided into n equally sized parts, I_i, $0 \leq i < n$. Hence,

$$I = \bigm|_{i=0}^{n-1} I_i \tag{1}$$

where '|' is the concatenation operator. From a bit vector, $\mathbf{b}$, that controls which parts of the information to encrypt, the selectively encrypted information, $E(I)$, is constructed as follows:

$$E(I) = \bigm|_{i=0}^{n-1} \begin{cases} I_i & \text{if } b_{i \bmod |\mathbf{b}|} = 0 \\ E(I_i) & \text{if } b_{i \bmod |\mathbf{b}|} = 1 \end{cases} \tag{2}$$

Figure 1 illustrates selectively encrypted information consisting of five encrypted parts (in gray) and four unencrypted parts (in white).

Figure. 1: A selectively encrypted information consisting of five encrypted parts (in gray) and four unencrypted parts (in white).

Previous work on selective encryption has mainly been aimed towards multimedia applications with a short information lifetime, such as TV or radio broadcasts of events, in order to reduce the overhead induced by encryption while still making the information perceptually secure. This was accomplished by utilizing the fact that different parts of the information have different impacts on our perception senses, i.e., our eyes and ears. The perception impact has also been used in lossy compression of information, like MP3 or JPEG compression, to keep more sensitive parts of the information more intact and thereby not destroying the perception beyond a desirable threshold. However, the security implications of how computationally secure the information is when using selective encryption have only briefly been mentioned or rudimentarily analyzed, e.g., in [17].
When measuring how computationally secure selectively encrypted information is, it must be taken into consideration that the unencrypted parts may, due to information dependencies, leak information to an attacker about the encrypted parts. This extra information must be considered and included in the used security measures, such as entropy [21] or guesswork [12–14, 16]. The entropy $H(X)$ is the classical

security measure of uncertainty that was originally defined by Shannon in 1944. Shannon defined the entropy of a random variable X with probability distribution p_i as follows:

$$H(X) = -\sum_i p_i \log_2 p_i \tag{3}$$

Thus, entropy gives the average amount of information of a discrete random variable. Entropy can also be seen as a measure giving the average number of guesses in an optimal binary search attack [9]. Guesswork $W(X)$, on the other hand, is a measure that gives the average number of guesses needed to find the value of a random variable X in an optimal brute force search attack, i.e., a linear search attack. When performing an optimal brute force attack, the attacker is assumed to have complete knowledge of the probability distribution p_i of X. Hence, before the guessing process starts, the attacker arranges the distribution in a non-increasing probability order:

$$p_1 \geq p_2 \geq \ldots \geq p_n \tag{4}$$

From this, guesswork is formally defined as follows:

$$W(X) = \sum_i i p_i \tag{5}$$

In this paper, the entropy up to the second-order of selectively encrypted bitmap images will be investigated by generalizing the work on entropy of selectively encrypted strings [10, 11] through information neighborhoods that capture and model information dependencies in several dimensions. The used information neighborhoods are constructed from a generalization of Shannon's work on the order of languages together with the neighborhood concept from cellular automata theory [8], using the L^1 metric [4]. Moreover, the construction of information neighborhoods is generic, hence they can be used for any application or idea where it is desired to capture and investigate information dependencies.

The remainder of the paper is organized as follows. Section II, describes the structure of bitmap images and presents previous work on bitplane encryption of bitmap images. Information neighborhoods that allow for information dependencies in several dimension are constructed in Section III. Section IV investigates the entropy of selectively encrypted bitmap images, and Section V concludes the paper.

II. Bitmap Images

In this section, the structure of bitmap images is briefly described together with a presentation of previous work on selectively encrypted bitmap images.

A. Bitmap Structure

An uncompressed $m \times n$ bitmap image can be seen as an m by n matrix $\mathcal{I}$ with pixels in the entries. A pixel $\mathcal{I}_{xy}$ is the smallest perceptive information unit or point in the image, and its position in the image is given by the Cartesian (x, y) coordinate. Moreover, the colors of the pixels are normally represented by natural numbers and the amount of colors is given by the number of bits per pixel, which is referred to as the color depth, c. Hence,

$$0 \leq \mathcal{I}_{xy} < 2^c \tag{6}$$

with the often used convention that white is represented by $\mathcal{I}_{xy} = 0$ and black is represented by $\mathcal{I}_{xy} = 2^c - 1$. Common values of c are 1, 4, 8, 16, 24, 32, 48 or 64 bits per pixel [3], with values less than 8 bits used for grayscale images. Thus, if $c = 1$, the color palette will contain only black and white and, if $c = 8$, the color palette will contain the 256 grayscale colors. For higher values of c the color palette contains other colors as well. In this paper, an 8 bitplanes 512×512 pixels bitmap version of the famous Lena image is used; see Figure 2 for an illustration.

Figure. 2: The famous Lena image.

To address unique bits in the image, a third coordinate z is needed. Hence, $\mathcal{I}_{xyz}$, $1 \leq z \leq c$, gives bit z in the pixel at position (x, y). To shorten the notations, by setting $\mathbf{p} = (x, y, z)$ a bit in the image can be written $\mathcal{I}_{\mathbf{p}}$. Moreover, the subset of bits $\mathcal{I}_z = \mathcal{I}_{**z}$, corresponding to all bits at position z of the image, is referred to as bitplane z of the image. Only encryption of whole bitplanes, $E(\mathcal{I}_z)$, will be further investigated in this paper. Figure 3 illustrates the eight bitplanes of an image. The points in the bitplanes correspond to

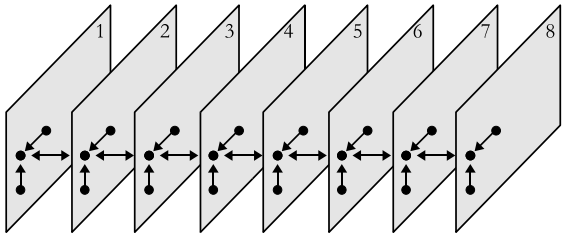

Figure. 3: Eight bitplanes of an image. The points in the bitplanes correspond to single bits, and all bits with equal (x, y) coordinates constitute a pixel.

single bits, and all bits with equal coordinates in the bitplanes constitute a pixel. Note the arrows that are intended to show dependencies between bits, both within the bitplanes and between the bitplanes. Dependencies are further discussed in Section III.

B. Previous Work on Bitplane Encryption

To reduce computational cost while still protecting image perception during transmissions in mobile environments, a selective bitplane encryption was proposed in [17]. In the paper, the authors started by cumulatively encrypting the most significant bitplanes of an eight bitplanes 512×512 pixels bitmap version of the Lena image, see Figure 2. The authors claimed that encryption of only the most significant bitplane gives a low level of perceptive protection, encryption of the two most significant bitplanes gives enough protection if a high degradation of the image is sufficient, and encryption of the four most significant bitplanes gives a high level of perceptive protection. Figure 4 illustrates the cumulative encryption of the most significant bitplanes in the Lena image, starting with encryption of the most significant bitplane in the leftmost subfigure and ending with encryption of the four most significant bitplanes in the rightmost subfigure. The

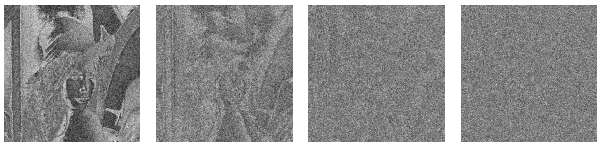

Figure. 4: Cumulative encryption of the most significant bitplanes in the Lena image, starting with encryption of the most significant bitplane in the leftmost subfigure and ending with encryption of the four most significant bitplanes in the rightmost subfigure.

result were not only verified by visual inspections, but also from the assessment of a replacement attack and a reconstruction attack. In the assessment attack encrypted bitplanes were replaced by zero-bit bitplanes, and in the reconstruction attack used that adjacent pixels are dependent and tends to be identical.

In contrast to [17], the authors of [2] instead started by cumulatively encrypting the least significant bitplanes of the Lena image. The reason for starting in the opposite direction is that the less significant bitplanes are harder to perform a plaintext attack on since they are more independent of each other and more random in their internal bit-pattern structure. The cumulative encryption of the least significant bitplanes in the Lena image is illustrated in Figure 5, starting with encryption of the least significant bitplane in the leftmost upper subfigure and ending with encryption of all the eight bitplanes in the rightmost lower subfigure. Note that at least six bitplanes are needed to be encrypted before the perception of the image becomes somewhat degraded.

Besides bitmap image protection, selective encryption has also been studied for MPEG data [22], H.264/AVC video streams [20], JPEG2000 images [1, 15] and a wireless video camera [6]. Moreover, a perception-based selective encryption scheme for telephone data compressed with the ITU-T G.729 8 kb/s speech encoding standard was presented in [19]. Selective encryption for the G.729 speech encoding standard has also been studied in [24].

Figure. 5: Cumulative encryption of least significant bitplanes in the Lena image, starting with encryption of the least significant bitplane in the leftmost upper subfigure and ending with encryption of all the eight bitplanes in the rightmost lower subfigure.

III. Information Neighborhoods

In this section, information neighborhoods that captures and models information dependencies in several dimensions are constructed. This is done by generalizing Shannon's work on order of languages and using the neighborhood concept from cellular automata theory.

A. Generalizing Order of Languages

In [21], contiguous sequences of symbols in a language, called n-grams, were used to derive the probabilities of the symbols in order to approximate texts in the corresponding language. The approximation was carried out to different orders as follows:

- In the zero-order approximation, $\omega = 0$, symbols are independent and uniformly distributed.
- In the first-order approximation, $\omega = 1$, symbols are independent with a distribution according to the 1-grams in the language.
- In the second-order approximation, $\omega = 2$, symbols are dependent on one preceding symbol with a distribution according to the 2-grams in the language.
- In the n-order approximation, $\omega = n$, symbols are dependent on $n - 1$ preceding symbols with a distribution according to the n-grams in the language.

The order gives the size of the ω-grams used in the approximation and thus determines the set of depending symbols that the approximation uses when calculating the probabilities of the symbols. In this paper, the set of depending symbols will be referred to as the neighborhood, based on the neighborhood concept in cellular automata [8]. Moreover, symbols in a language are not only depending on preceding symbols as in the above described approximation model, but also on succeeding symbols. Thus, the overall or total neighborhood can be divided into the preceding neighborhood and the succeeding neighborhood.

The representation state of the information is usually considered one-dimensional when transferring or storing it. However, for better comprehension and higher abstraction, the

representation state might be converted/transformed to another representation state. For instance, the information of a bitmap image can be represented in two or three dimensions. Two dimensions if the pixel state is considered, coordinates of the pixels, and three dimensions if the bit state is considered, coordinates of the pixels and color depth. The conversion between two representation states, possibly without changing the dimension, is called lossless if it is invertible, otherwise it is called lossy. In a lossy conversion, like MPEG audio and video [23, 25], information is lost and the interpretation of it becomes more or less degraded depending on how forgiving the perception environment is. Furthermore, encryption converts information lossless to a representation state that cannot be interpreted by others than those who have a secret that can be used to convert the information back to the original and interpretable representation state.

When the information has a multi-dimensional representation state, such as for bitmap images, the order concept in the language approximation model needs to be generalized. This can be done by using an order vector, $\boldsymbol{\omega}$, where each element gives the order in the corresponding dimension. For instance, $\boldsymbol{\omega} = (2, 1)$ indicates that the order is 2 in the first dimension and 1 in the second dimension.

B. Basic Neighborhoods

Since a ω-gram is a contiguous sequence of ω symbols, it can be represented as an ordered one-dimensional finite integer lattice $[a, b]$ of size ω. In such a structure, by referring to symbols as points and assuming that a step only reaches adjacent points in the lattice, the distance between two symbols is given by the number of steps when walking on the lattice between the two corresponding points. Hence, the nearest symbols is a walk of one step away, the second nearest symbol is a walk of two steps away, and so on. Generalizing this to several dimensions naturally leads to the ordered n-dimensional finite integer lattice I^n with the L^1 metric [4] for calculating distances between points. The L^1 metric between two arbitrary n-dimensionally points, $\mathbf{p}$ and $\mathbf{q}$, is given by

$$||\mathbf{p} - \mathbf{q}||_1 = \sum_{i=1}^{n} |p_i - q_i| \tag{7}$$

and it is often referred to as the Manhattan or Taxicab metric [5]. In one dimension the L^1 distance between two points $\mathbf{p} = (a)$ and $\mathbf{q} = (b)$ becomes $||\mathbf{p} - \mathbf{q}||_1 = |a - b|$. Note that by not restricting the steps in the walk on the integer lattice to adjacent points a lot of other metrics like the Euclidean L^2 metric could be used.

In the above described language approximation model, the point considered is always located at the right boundary point of the ω-gram or lattice. Hence, $\mathbf{p} = (b)$, giving the neighborhood

$$\mathcal{N}_{\omega}^{\omega-1}(b) = [a, \ldots, b-1] \tag{8}$$

consisting of $\omega - 1$ preceding points and zero succeeding points. However, symbols in a language not only depend on preceding symbols but also on succeeding symbols. Thus, if the considered point is instead arbitrary located in the lattice, $\mathbf{p} = (x)$, then

$$\mathcal{N}_{\omega}^{x-a}(x) = [a, \ldots, x-1] \cup [x+1, \ldots, b] \tag{9}$$

consisting of $x - a$ preceding points and $b - x$ succeeding points. To shorten the lattice notation, the neighborhoods will in the following be represented by the tuple

$$\mathcal{N}_{\omega}^{i}(x) = (i, \omega - i - 1) \tag{10}$$

$0 \leq i < \omega$, where the first element gives the number of preceding points and the second element gives the number of succeeding points in the neighborhood. Note that there are ω different neighborhoods, i.e., one neighborhood for each position of the point in the lattice. For instance, if $\omega = 3$, the family of all possible neighborhoods is $\mathcal{N}_3(x) = \{(0, 2), (1, 1), (2, 0)\}$. More generally, the family of all possible neighborhoods of order ω in one dimension is given by

$$\mathcal{N}_{\omega}(x) = \{\mathcal{N}_{\omega}^{i}(x) \,|\, 0 \leq i < \omega\} \tag{11}$$

Figure 6 illustrates the different neighborhoods of $\mathcal{N}_1(x)$, $\mathcal{N}_2(x)$ and $\mathcal{N}_3(x)$. A black square represents the point under consideration and the white squares represent preceding or succeeding points in the neighborhoods.

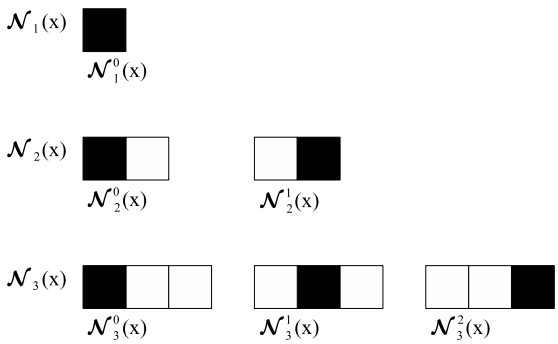

Figure. 6: The different neighborhoods of $\mathcal{N}_1(x)$, $\mathcal{N}_2(x)$ and $\mathcal{N}_3(x)$. A black square represents the point under consideration and the white squares represent preceding or succeeding points in the neighborhoods.

From the families of one-dimensional neighborhoods, also referred to as basic neighborhoods, new families of basic neighborhoods in higher dimensions can be constructed as follows

$$\begin{aligned}\mathcal{N}_{\boldsymbol{\omega}}(\mathbf{p}) &= \overset{n}{\underset{i=1}{\times}} \mathcal{N}_{\omega_i}(p_i) \\ &= \{\mathcal{N}_{\boldsymbol{\omega}}^{\mathbf{i}}(\mathbf{p}) \,|\, \mathbf{0} \leq \mathbf{i} < \boldsymbol{\omega}\}\end{aligned} \tag{12}$$

where $\mathbf{i} = (i_1, \ldots, i_n)$ and

$$\begin{aligned}\mathcal{N}_{\boldsymbol{\omega}}^{\mathbf{i}}(\mathbf{p}) &= \overset{n}{\underset{j=1}{\times}} \mathcal{N}_{\omega_j}^{i_j}(p_j) \\ &= (\mathbf{i}, \boldsymbol{\omega} - \mathbf{i} - \mathbf{1})\end{aligned} \tag{13}$$

Figure 7 illustrates the construction of the two-dimensional family $\mathcal{N}_{2,2}(x, y) = \mathcal{N}_2(x) \times \mathcal{N}_2(y)$. Since $\mathcal{N}_2(x) =$

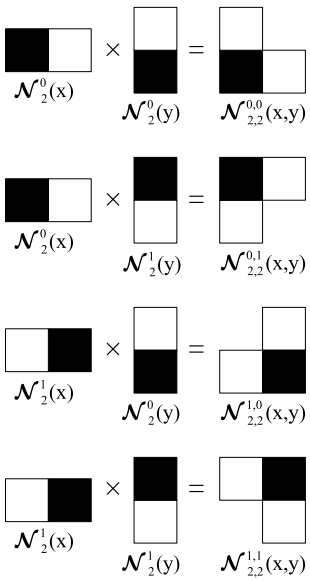

Figure. 7: Construction of the four basic neighborhoods in the two-dimensional family $\mathcal{N}_{2,2}(x,y) = \mathcal{N}_2(x) \times \mathcal{N}_2(y)$.

$\{\mathcal{N}_2^0(x), \mathcal{N}_2^1(x)\}$, the resulting $\mathcal{N}_{2,2}(x,y)$ family will consist of the four axis neighborhoods

$$\begin{aligned}\mathcal{N}_{2,2}^{i_1,i_2}(x,y) &= \mathcal{N}_2^{i_1}(x) \times \mathcal{N}_2^{i_2}(y) \\ &= ((i_1,i_2),(1-i_1,1-i_2)) \end{aligned} \quad (14)$$

Note that the elements in $\boldsymbol{\omega}$ need not be equal; hence, families such as $\mathcal{N}_2(x) \times \mathcal{N}_1(y)$ can also be constructed. Moreover, Figure 8 illustrates all nine states of $\mathcal{N}_{3,3}(x,y) = \mathcal{N}_3(x) \times \mathcal{N}_3(y)$. Note that the $\mathcal{N}_{3,3}^{1,1}(x,y)$ basic neighbor-

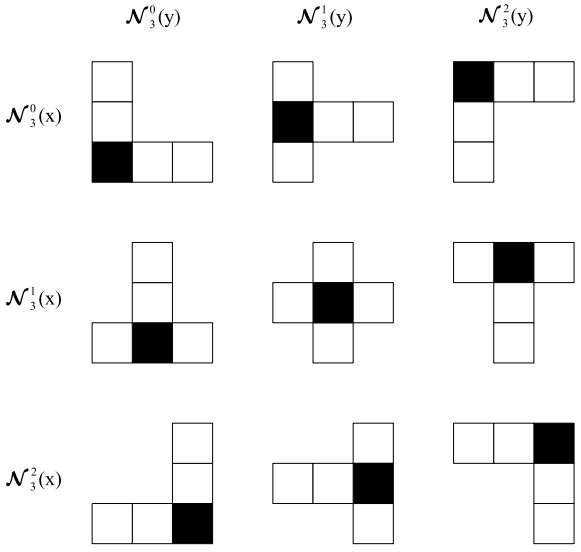

Figure. 8: The nine basic neighborhoods of $\mathcal{N}_{3,3}(x,y) = \mathcal{N}_3(x) \times \mathcal{N}_3(y)$.

hood is equal to the von Neumann neighborhood [8] of range one in the L_1 metric, which is also equal to the concept of 4-connected pixels in computer images. Construction of basic neighborhoods in higher dimensions follows, according to (13), the same structure as in the two-dimensional example.

C. Overall Neighborhoods

The basic neighborhoods contain only points that are located on the axes. However, points located outside the axes but within a specific distance determined by $\boldsymbol{\omega}$ can also be considered to be depending points. Since a circle in the L_1 metric has the shape of a convex polytope, a neighborhood will in the rest of this paper be constructed by joining the outermost points in the corresponding one-dimensional basic neighborhoods, thereby creating a convex polytope acting as the overall neighborhood. An n-polytope is a set or geometric object in n dimensions with flat sides. Usually a 2-polytope is referred to as a polygon and a 3-polytope as a polyhedron. Figure 9 shows the polygon, 2-polytope, for the basic neighborhood $\mathcal{N}_{5,5}^{3,3}(x,y)$ is shown. Note the points

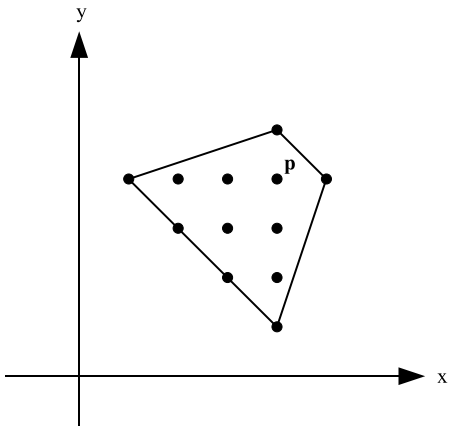

Figure. 9: The polygon, 2-polytope, for the basic neighborhood $\mathcal{N}_{5,5}^{3,3}(x,y)$.

in the overall neighborhood that are not included in the basic neighborhoods. Moreover, if the basic neighborhoods are symmetrically located around the considered point, the overall neighborhoods are actually circles in the L_1 metric.
To mathematically describe the overall neighborhoods the concept of convex hull [18] can be used. The convex hull, Conv($\mathcal{S}$), of a set $\mathcal{S}$ of points is the intersection of all convex sets containing $\mathcal{S}$. Thus, Conv($\mathcal{S}$) forms the smallest convex polytope that contains $\mathcal{S}$. In Figure 9 the overall neighborhood of $\mathbf{p}$ can be written as $\mathcal{D}_{5,5}^{3,3}(x,y) = \text{Conv}(\mathcal{N}_{5,5}^{3,3}(x,y)) \setminus (x,y)$, and generally the overall neighborhood can be written as

$$\mathcal{D}_{\boldsymbol{\omega}}^{\mathbf{i}}(\mathbf{p}) = \text{Conv}(\mathcal{N}_{\boldsymbol{\omega}}^{\mathbf{i}}(\mathbf{p})) \setminus \mathbf{p} \quad (15)$$

where $\setminus$ is the setminus operation. Note that $\mathcal{D}_{\mathbf{0}}(\mathbf{p}) = \mathcal{D}_{\mathbf{1}}(\mathbf{p}) = \emptyset$, hence, no neighborhoods exists in the zero- and first order approximations. In one-dimension the overall neighborhoods are equal to the basic neighborhoods, $\mathcal{D}_\omega(x) = \mathcal{N}_\omega(x)$. Moreover, if the basic neighborhoods are symmetrically located around the considered point, then the

overall neighborhood can also be written as

$$\mathcal{D}^{\mathbf{i}}_{\boldsymbol{\omega}}(\mathbf{p}) = \{\mathbf{q} \neq \mathbf{p} \,|\, ||\mathbf{p} - \mathbf{q}||_1 < \omega\} \tag{16}$$

In two dimensions these neighborhoods are equal to the von Neumann neighborhoods of range one in the L_1 metric. Figure 10 illustrates the symmetrical neighborhood $\mathcal{D}^{2,2}_{5,5}(x, y)$.

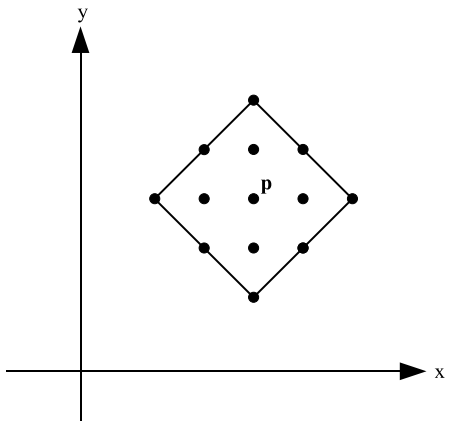

Figure. 10: The overall neighborhood $\mathcal{D}^{2,2}_{5,5}(x, y)$.

IV. Entropy of Selectively Encrypted Bitmap Images using Information Neighborhoods

This section uses the concept of information neighborhoods introduced in the previous section and results from [11] to investigate the entropy of selectively encrypted bitmap images. Only whole bitplanes of the image are assumed to be encrypted, and a bit $E(\mathcal{I}_\mathbf{p})$ will be associated with a random variable X as follows

$$E(\mathcal{I}_\mathbf{p}) = \begin{cases} X_\mathbf{p} = \mathcal{I}_\mathbf{p} & \text{if } b_z = 0 \\ X_\mathbf{p} & \text{if } b_z = 1 \end{cases} \tag{17}$$

Hence, the random variables have a sample space $\mathcal{X} = \{0, 1\}$ and model the behavior of the bits in $E(\mathcal{I})$ by being known if the corresponding bits are unencrypted and being unknown if the corresponding bits are encrypted. Moreover, to simplify the entropy calculations only states having $\omega_i = \omega$ or zero will be considered. Thus, all considered dimensions will have the same family of basic neighborhoods.

A. Zero-order

In the zero-order case, $\boldsymbol{\omega} = 0$, the random variables are independent of each other and the probability distribution on $\mathcal{X}$ is uniform. Hence, bits inside bitplanes and bits between bitplanes are independent of each other. This implies that no information about the image is known. Thus, the entropy of $E(\mathcal{I}_z)$ attains its maximum value of

$$\begin{aligned} H_\mathbf{0}(E(\mathcal{I}_z)) &= \sum_{\mathbf{p}|z} H_\mathbf{0}(X_\mathbf{p}) \\ &= mn \end{aligned} \tag{18}$$

The last step in (18) comes from the observation that $H_\mathbf{0}(X_\mathbf{p}) = log_2(2) = 1$ for the uniform distribution. For the whole image, the entropy in the zero-order case becomes

$$\begin{aligned} H_\mathbf{0}(E(\mathcal{I})) &= \sum_{\mathbf{p}} b_z H_\mathbf{0}(X_\mathbf{p}) \\ &= mn \sum_z b_z \end{aligned} \tag{19}$$

where $\sum_z b_z$ gives the number of encrypted bitplanes in the image. Note that if all bitplanes are encrypted, $\sum_z b_z = c$, then $H_\mathbf{0}(E(\mathcal{I})) = cmn$. For the Lena image, which has 512×512 pixels, $H_\mathbf{0}(E(\mathcal{I}_z)) = 512^2$ and $H_\mathbf{0}(E(\mathcal{I})) = 512^2 \sum_z b_z$.

B. First-order

In the first-order case, $\boldsymbol{\omega} = \mathbf{1}$, the random variables are still independent of each other. However, the distribution on $\mathcal{X}$ is now equal to the distribution of the bits in the bitplane under consideration. Note that there will be c distributions, one for each bitplane. For an encrypted bitplane $E(\mathcal{I}_z)$ the entropy becomes

$$\begin{aligned} H_\mathbf{1}(E(\mathcal{I}_z)) &= \sum_{\mathbf{p}|z} H_\mathbf{1}(X_\mathbf{p}) \\ &= mnH(X_{\mathbf{p}|z}) \end{aligned} \tag{20}$$

where the last step in (20) comes from the fact that all bits in bitplane z have equal entropy $H(X_{\mathbf{p}|z})$. However, $H(X_{\mathbf{p}|z})$ changes for different bitplanes through the value of z. For the whole image, the entropy in the first-order case becomes

$$\begin{aligned} H_\mathbf{1}(E(\mathcal{I})) &= \sum_{\mathbf{p}} b_z H_\mathbf{1}(X_\mathbf{p}) \\ &= mn \sum_z b_z H(X_{\mathbf{p}|z}) \end{aligned} \tag{21}$$

Figure 11 shows the 1-gram bit distribution $p_\mathbf{1}(\mathcal{I}_{\mathbf{p}|z})$ of each bitplane of the Lena image. Note that bitplane five to eight

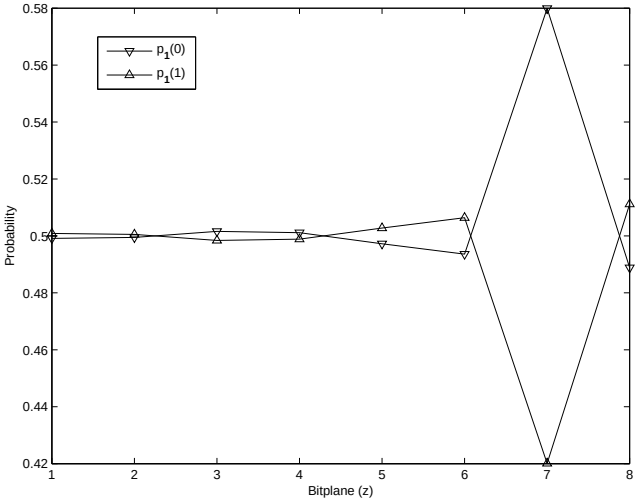

Figure. 11: The 1-gram bit distribution $p_\mathbf{1}(\mathcal{I}_{\mathbf{p}|z})$ of the Lena image.

deviates more from the uniform distribution. This is as expected, since more significant bitplanes are supposed to capture more of the image structure. In Figure 12, the entropies $H_{\mathbf{0}}(E(\mathcal{I}_z))$ and $H_{\mathbf{1}}(E(\mathcal{I}_z))$ are shown for each bitplane of the Lena image. Note that the entropy decreases slowly for more significant bitplanes except at the negative spike at bitplane seven, and that $H_{\mathbf{1}}(E(\mathcal{I}_z) \lesssim H_{\mathbf{0}}(E(\mathcal{I}_z)) = 512^2$.

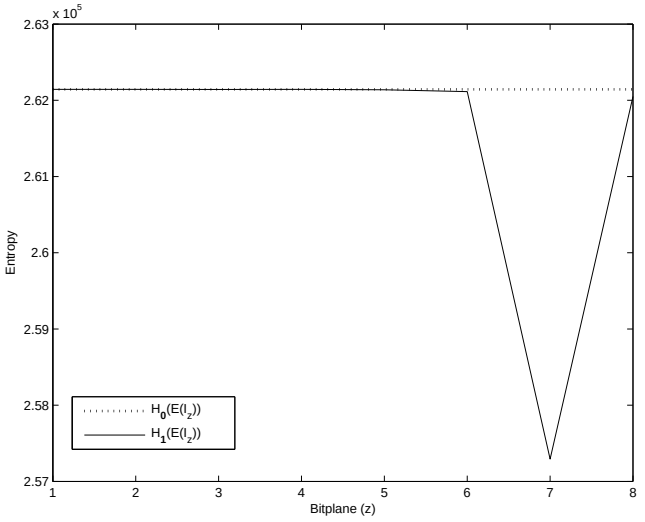

Figure. 12: The entropies $H_{\mathbf{0}}(E(\mathcal{I}_z))$ and $H_{\mathbf{1}}(E(\mathcal{I}_z))$ of the Lena image.

C. Second-order

In the second-order case, $\omega_i = 2$ or 0, the random variables are dependent for each dimension on at most one succeeding or preceding random variable. Hence, for a bitmap image, the neighborhoods will at most contain three random variables, two inside the corresponding bitplane, the x and y coordinates, and one from an adjacent bitplane, the z coordinate. See Figure 3 for an illustration of this. Moreover, by using the results in [11], the second-order entropy of an encrypted bitplane when considering dependence of one preceding symbol in the x-axis becomes

$$H^{1,0,0}_{2,0,0}(E(\mathcal{I}_z)) = \sum_{\mathbf{p}|z} \prod_{x'<x} p(X_{x'}|X_{x'-1}) H(X_x|X_{x-1}) \tag{22}$$

From the previously defined neighborhood notation, see Section III, the entropy of an encrypted bitplane given in (22) can, to allow for other combinations of dependencies in $\boldsymbol{\omega}$, be generalized as follows

$$\begin{aligned} H^{\mathbf{i}}_{\boldsymbol{\omega}}(E(\mathcal{I}_z)) &= \sum_{\mathbf{p}|z} \prod_{\mathbf{p}' \in \mathcal{R}^{\mathbf{i}}_{\boldsymbol{\omega}}(\mathbf{p})} p(X_{\mathbf{p}'}|X_{\mathcal{D}^{\mathbf{i}}_{\boldsymbol{\omega}}(\mathbf{p}')}) H(X_{\mathbf{p}}|X_{\mathcal{D}^{\mathbf{i}}_{\boldsymbol{\omega}}(\mathbf{p})}) \\ &= \sum_{\mathbf{p}|z} p^{\mathbf{i}}_{\boldsymbol{\omega}}(X_{\mathcal{R}^{\mathbf{i}}_{\boldsymbol{\omega}}(\mathbf{p})}) H^{\mathbf{i}}_{\boldsymbol{\omega}}(X_{\mathbf{p}}) \end{aligned} \tag{23}$$

The region $\mathcal{R}^{\mathbf{i}}_{\boldsymbol{\omega}}(\mathbf{p})$ used in the product of (23) is a connected encrypted rectangular subset of the image $\mathcal{I}$, and it is generated by $\mathbf{p}$ and the last encrypted points in the direction given by the indexes. For instance, if the neighborhood considered is $\mathcal{D}^{1,1,0}_{2,2,0}(\mathbf{p})$, then the generating points of $\mathcal{R}^{\mathbf{i}}_{\boldsymbol{\omega}}(\mathbf{p})$ are $\mathbf{p}$ and the left lower point of the bitplane under consideration. How to mathematically express $\mathcal{R}^{\mathbf{i}}_{\boldsymbol{\omega}}(\mathbf{p})$ will be an issue of future research. Finally, the entropy for the whole image in the second-order case becomes

$$H^{\mathbf{i}}_{\boldsymbol{\omega}}(E(\mathcal{I})) = \sum_{\mathbf{p}} b_z p^{\mathbf{i}}_{\boldsymbol{\omega}}(X_{\mathcal{R}^{\mathbf{i}}_{\boldsymbol{\omega}}(\mathbf{p})}) H^{\mathbf{i}}_{\boldsymbol{\omega}}(X_{\mathbf{p}}) \tag{24}$$

1) One-dimensional Dependence

In an one-dimensional dependence of the information, $\omega_i = 2$ for exactly one index value. Considering the x-axis, Figure 13 shows the conditional 2-gram probability distributions, $p^{1,0,0}_{2,0,0}(\mathcal{I}_{\mathbf{p}|z})$, of each bitplane of the image. Note again,

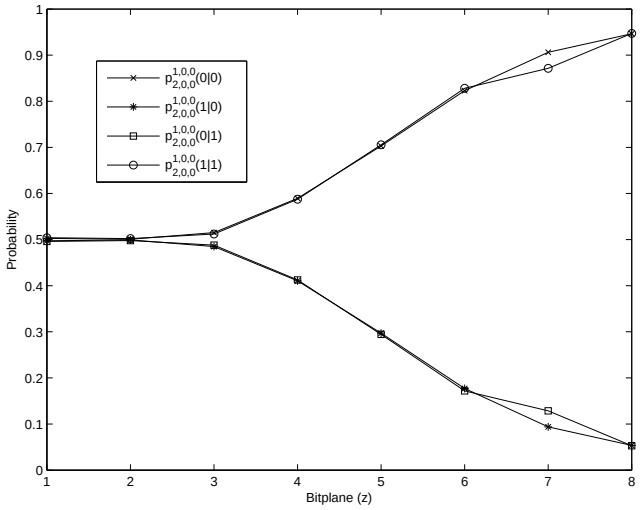

Figure. 13: The conditional 2-gram probability distributions $p^{1,0,0}_{2,0,0}(\mathcal{I}_{\mathbf{p}|z})$ of the Lena image.

that there is a larger deviation from the uniform distribution for more significant bitplanes. The probability distributions $p_{0,2,0}(\mathcal{I}_{\mathbf{p}|z})$ have almost the same shape. The corresponding entropies $H_{2,0,0}(E(\mathcal{I}_z))$ and $H_{0,2,0}(E(\mathcal{I}_z))$ of the Lena image are shown in Figure 14. Note again that the entropy decreases for more significant bitplanes, and that the entropies are almost identical in each case regarding $\mathbf{i}$. The reduction in entropy between bitplane one and eight is 69.9% in the $H_{2,0,0}(E(\mathcal{I}_z))$ case and 75.0% in the $H_{0,2,0}(E(\mathcal{I}_z))$ case. Moreover, Figure 15 shows the conditional 2-gram probability distributions $p^{0,0,1}_{0,0,2}(\mathcal{I}_{\mathbf{p}|z})$ between the bitplanes of the Lena image. Once again the considering distribution deviates more from uniformness for more significant bitplanes. The corresponding entropies $H_{0,0,2}(E(\mathcal{I}_z)|b_{\mathcal{D}_2(z)})$, of the Lena image are shown in Figure 16 in the case when only one adjacent bitplane is considered. If $b_{\mathcal{D}_2(z)} = 0$, the adjacent bitplane is unencrypted. The reduction in entropy between bitplane one and seven, $\mathbf{i} = (0,0,0)$, is 20.1%, and between bitplane two and eight, $\mathbf{i} = (0,0,1)$, is 18.3%. Note that the entropy decreases more when information within bitplanes is used compared to when information between bitplanes is used.

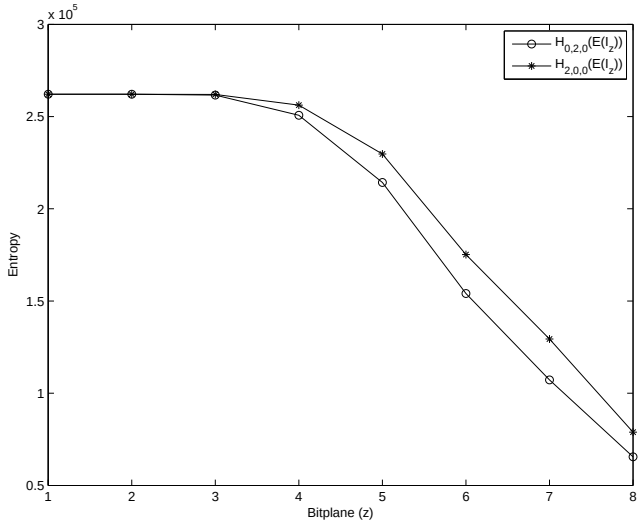

Figure. 14: Entropies $H_{2,0,0}(E(\mathcal{I}_z))$ and $H_{0,2,0}(E(\mathcal{I}_z))$ of the Lena image.

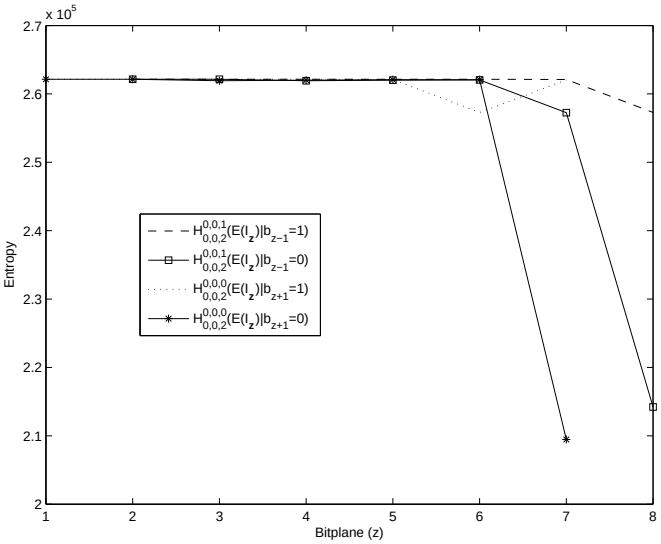

Figure. 16: Entropies $H_{0,0,2}(E(\mathcal{I}_z)|b_{\mathcal{D}_2(z)})$ of the Lena image.

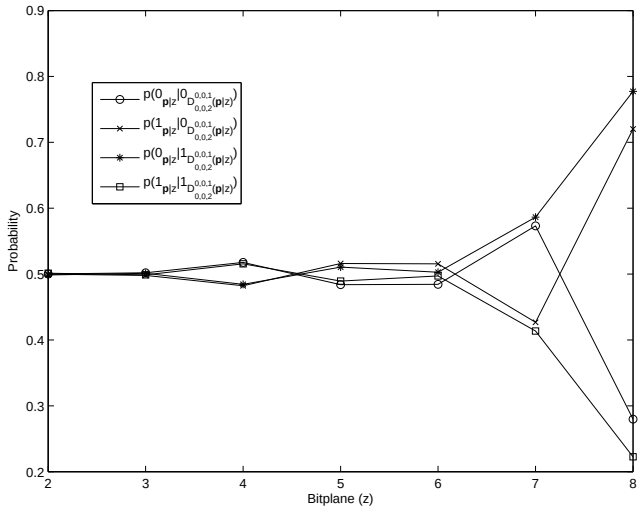

Figure. 15: The conditional 2-gram distributions between the bitplanes $p^{0,0,1}_{0,0,2}(\mathcal{I}_{\mathbf{p}|z})$ of the Lena image.

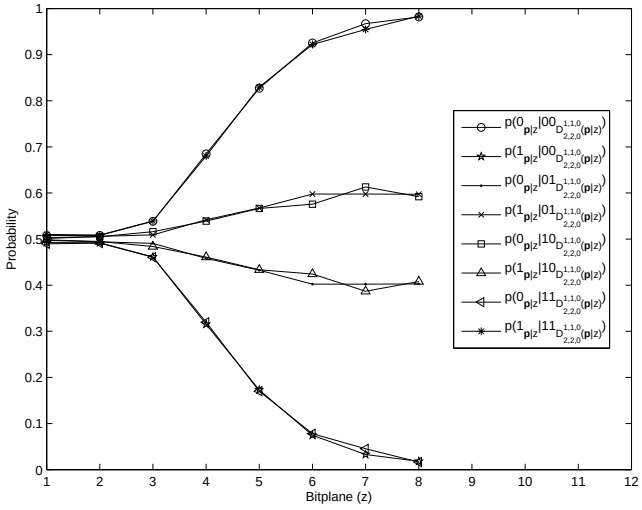

Figure. 17: The conditional 2-gram distributions $p^{1,1,0}_{2,2,0}(\mathcal{I}_{\mathbf{p}|z})$ of the Lena image.

2) Two-dimensional Dependence

In a two-dimensional dependence of the information, $\omega_i = 2$ for exactly two index values. The bitplanes of the image is one example of such two-dimensional dependence where the x and y index is considered. Figure 17 shows the conditional 2-gram distributions $p^{1,1,0}_{2,2,0}(\mathcal{I}_{\mathbf{p}|z})$ of each bitplane of the Lena image. Moreover, in this case the two-dimensional product in (23) is harder to calculate than the corresponding one-dimensional product. The two-dimensional product will also be a focus of future research. However, the product can be calculated by assuming a steady state. Using this, the entropies $H^{1,1,0}_{2,2,0}(E(\mathcal{I}_z))$ and $H^{1,0,0}_{2,2,0}(E(\mathcal{I}_z))$ of the Lena image are calculated and shown in Figure 18. The reduction in entropy between bitplane one and eight is 82.9% when $\mathbf{i} = (1, 1, 0)$ and 80.6% when $\mathbf{i} = (1, 0, 0)$. The other two entropies that are not plotted are $H^{0,0,0}_{2,2,0}(E(\mathcal{I}_z)) \approx H^{1,1,0}_{2,2,0}(E(\mathcal{I}_z))$ and $H^{0,1,0}_{2,2,0}(E(\mathcal{I}_z)) \approx H^{1,0,0}_{2,2,0}(E(\mathcal{I}_z))$. Other two-dimensional entropies of the bitplanes of the Lena image are $H_{2,0,2}(E(\mathcal{I}_z)|b_{\mathcal{D}_2(z)})$ and $H_{0,2,2}(E(\mathcal{I}_z)|b_{\mathcal{D}_2(z)})$. If $b_{\mathcal{D}_2(z)} = 0$ then $\mathcal{R}(\mathbf{p})$ in (23) will be one dimensional in the x or y direction. If instead $b_{\mathcal{D}_2(z)} = 1$ then $\mathcal{R}(\mathbf{p})$ will still be two dimensional, but only consisting of two adjacent rows of the bitplanes. In Figure 19, the entropies $H^{1,0,1}_{2,0,2}(E(\mathcal{I}_z)|b_{z-1})$ and $H^{0,1,1}_{0,2,2}(E(\mathcal{I}_z)|b_{z-1})$ of the Lena image is shown. If $b_{z-1} = 0$ the entropy reduction are 84.5% and 89.5%, respectively, and if $b_{z-1} = 1$ it is 66.4% and 70.7%, respectively.

3) Three-dimensional Dependence

In a three-dimensional dependence of the information, $\boldsymbol{\omega} = \mathbf{2}$, the product in (23) is even harder to calculate. Figure 20 shows the conditional 2-gram probability distributions $p^{\mathbf{1}}_{\mathbf{2}}(\mathcal{I}_{\mathbf{p}|z})$ of the Lena image. Due to the large amount of plots

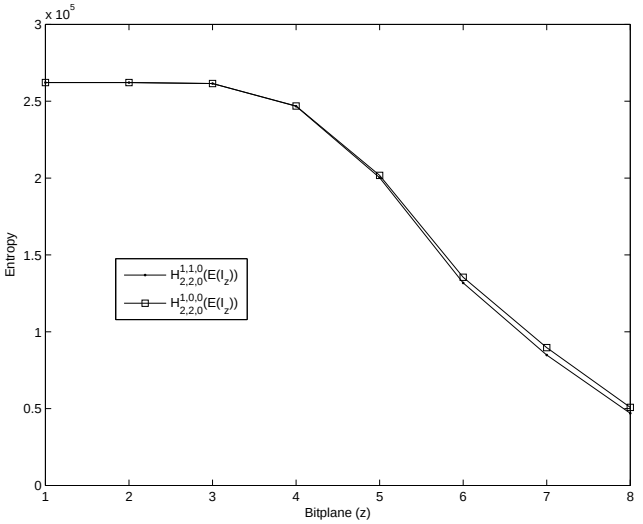

Figure. 18: The entropies $H^{1,1,0}_{2,2,0}(E(\mathcal{I}_z))$ and $H^{1,0,0}_{2,2,0}(E(\mathcal{I}_z))$ of the Lena image.

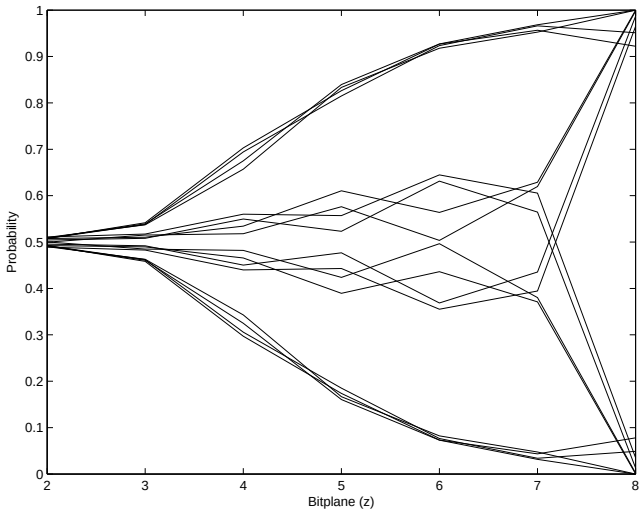

Figure. 20: The conditional 2-gram distributions $p^{\mathbf{1}}_{\mathbf{2}}(\mathcal{I}_{\mathbf{p}|z})$ of the Lena image.

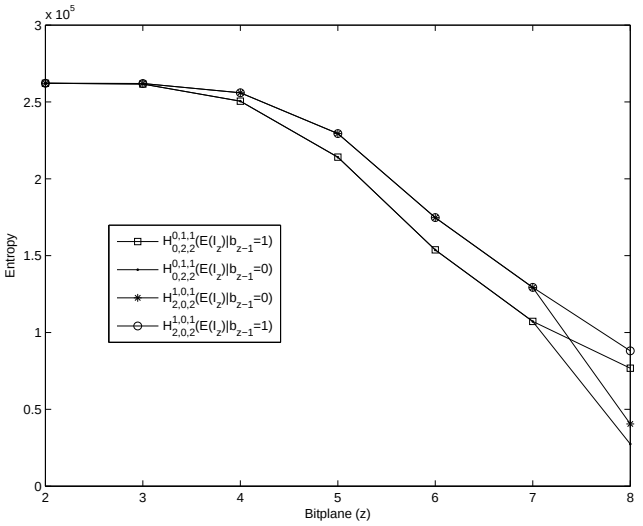

Figure. 19: The entropies $H^{1,0,1}_{2,0,2}(E(\mathcal{I}_z)|b_{z-1})$ and $H^{0,1,1}_{0,2,2}(E(\mathcal{I}_z)|b_{z-1})$ of the Lena image.

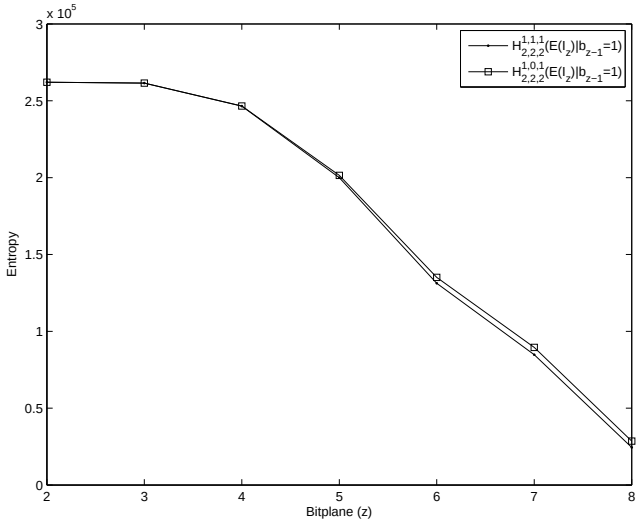

Figure. 21: The entropies $H^{1,1,1}_{2,2,2}(E(\mathcal{I}_z)|b_{z-1} = 1)$ and $H^{1,0,1}_{2,2,2}(E(\mathcal{I}_z)|b_{z-1} = 1)$ of the Lena image.

in the graph, a legend is not inserted. However, bits of equal values tend to cluster, and this property increases for higher bitplanes. The highest probabilities, which occurs with value one, are $p^{\mathbf{1}}_{\mathbf{2}}(1|110)$ and $p^{\mathbf{1}}_{\mathbf{2}}(0|001)$. Moreover, by assuming a steady state, the entropies $H^{1,1,1}_{2,2,2}(E(\mathcal{I}_z)|b_{z-1} = 1)$ and $H^{1,0,1}_{2,2,2}(E(\mathcal{I}_z)|b_{z-1} = 1)$ of the Lena image are calculated and shown in Figure 21. The entropy reduction between bitplane two and eight is 90.7% when $\mathbf{i} = (1, 1, 1)$ and 89.1% when $\mathbf{i} = (1, 0, 1)$. Moreover, state $\mathbf{i} = (0, 0, 1)$ has an entropy almost equal to that of state $\mathbf{i} = (1, 1, 1)$, and state $\mathbf{i} = (0, 1, 1)$ as that of state $\mathbf{i} = (1, 0, 1)$.

V. Conclusions and Future Work

This paper investigated the entropy until the second-order of selectively encrypted bitmap images. To capture information dependencies in several dimensions when performing the entropy calculations, information neighborhoods were constructed by extending Shannon's work on the order of languages together with ideas from the neighborhood concept in cellular automata theory. As expected, the entropy in the Lena bitmap image seems to decrease with more significant bitplanes being encrypted and when larger information neighborhoods are used.

To further investigate the entropy of selectively encrypted bitmap images, the total entropy, not only for single bitplanes, and higher order information neighborhoods, will be considered and applied to other bitmap images as well. Another issue would be to investigate and correlate entropy and the perceptive signal to noise ratio measure, which was used in [2] to investigate the perception of selectively encrypted bitmap images. The product in (23) over the region $\mathcal{R}^{\mathbf{i}}_{\boldsymbol{\omega}}(\mathbf{p})$ will be a further focus of future research.

Acknowledgment

This paper is a revised and extended version of "An Investigation of Entropy of Selectively Encrypted Bitmap Images" by R. Lundin and S. Lindskog, which appeared in Proceedings of the Fourth International Conference on Computational Aspects of Social Networks (CASoN 2012), pages 238–243, So Carlos, Brazil, November 21–23, 2012. © 2012 IEEE.

Part of the work has been carried out within the Compare Business Innovation Centre phase 3 (C-BIC 3) project, funded partly by European Regional Development Fund (ERDF).

References

[1] Z. Brahimi, H. Bessalah, A. Tarabet, and M. K. Khelladi. Selective encryption techniques of JPEG2000 codestream for medical images transmission. *Transactions on Circuits and Systems*, 7(7):718–727, July 2008.

[2] M. Van Droogenbroeck and R. Benedett. Techniques for a selective encryption of uncompressed and compressed images. In *Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS'02)*, pages 90–97, Ghent, Belgium, September 9–11, 2002.

[3] J. D. Foley. *Computer Graphics: Principles and Practice*. Systems Programming Series. Addison-Wesley, Reading, MA, USA, 1996.

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

[5] P. K. Ghosh and K. Deguchi. *Mathematics of Shape Description: A Morphological Approach to Image Processing and Computer Graphics*. John Wiley & Sons, New York, NY, USA, 2009.

[6] J. Goodman and A. P. Chandrakasan. Low power scalable encryption for wireless systems. *Wireless Networks*, 4(1):55–70, 1998.

[7] 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)*, pages 35–46, Paphos, Cyprus, May 23, 2006.

[8] W. M. Luckett. *Cellular Automata for Dynamic S-boxes in Cryptography*. University of Louisville, 2007.

[9] 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)*, pages 135–144, Funchal, Madeira, Portugal, June 12–13, 2007.

[10] R. Lundin and S. Lindskog. Security implications of selective encryption. In *Proceedings of the 6th International Workshop on Security Measurements and Metrics (MetriSec 2010)*, Bolzano, Italy, September 15, 2010.

[11] R. Lundin and S. Lindskog. Entropy of selectively encrypted strings. In *Proceedings of the 5th International Workshop in Information Security Theory and Practice (WISTP 11)*, pages 234–243, Heraklion, Crete, Greece, June 1-3, 2011.

[12] R. Lundin and S. Lindskog. Joint and conditional guesswork: Definitions and implications. *Journal of Information Assurance and Security (JIAS)*, 6(2):89–97, 2011.

[13] R. Lundin and S. Lindskog. Changes in guesswork over time in multi-processor attacks. *Journal of Information Assurance and Security (JIAS)*, 7(4):241–251, 2012.

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

[15] A. Massoudi, F. Lefèbvre, C. De Vleeschouwer, and F.-O. Devaux. Secure and low cost selective encryption for JPEG2000. In *Tenth IEEE International Symposium on Multimedia (ISM 2008)*, pages 31–38, Berkeley, CA, USA, December 15–17, 2008.

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

[17] M. Podesser, H. P. Schmidt, and A. Uhl. Selective bitplane encryption for secure transmission of image data in mobile environments. In *Proceedings of the 5th IEEE Nordic Signal Processing Symposium (NORSIG'02)*, Tromsø/Trondheim, Norway, October 4–6, 2002.

[18] R. T. Rockafellar. *Convex Analysis*. Princeton Mathematical Series. Princeton University Press, Chichester, West Sussex, UK, 1997.

[19] A. Servetti and J. C. De Martin. Perception-based selective encryption of G.729 speech. In *Proceedings of the 2002 IEEE Internatinal Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 621–624, Orlando, Florida, USA, May 13–17, 2002.

[20] Z. Shahid, M. Chaumont, and W. Puech. Fast protection of H.264/AVC by selective encryption of CABAC for I & P frames. In *Proceedings of the 17th European Signal Processing Conference (EUSIPCO 2009)*, pages 2201–2205, Glasgow, Scotland, August 24–28, 2009.

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

[22] G. A. Spanos and T. B. Maples. Performance study of a selective encryption scheme for security of networked, real-time video. In *Proceedings of the 4th International Conference on Computer Communications and Networks (ICCCN'95)*, pages 72–78, Las Vegas, Nevada, USA, September 1995.

[23] W. Stallings. *High-speed networks and internets: performance and quality of service*. Prentice Hall, Upper Saddle River, NJ, USA, 2002.

[24] Z. Su, J. Jiang, S. Lian, G. Zhang, and D. Hu. Hierarchical selective encryption for G.729 speech based on bit sensitivity. *Journal of Internet Technology*, 10(5):599–608, 2010.

[25] A. S. Tanenbaum. *Computer networks*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2003.

Author Biographies

Reine Lundin received his Licentiate and PhD degrees in Computer Science from Karlstad University, Sweden, in 2007 and 2012, respectively. In 1999 and 2003 he also received the Master's Degrees of Physics and Mathematics, respectively, from Karlstad University. He joined the Department of Computer Science at Karlstad University in 2000, where he currently works as an associate professor. His research focus is quantitative security metrics and tunable security services. He has authored/coauthored over 25 book chapters and journal 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 50 journal and conference papers.