Study of imperfect keys to characterise the security of opticalencryption

Lingfei Zhang, Thomas J. Naughton

Department of Computer Science, Maynooth University–National University of Ireland Maynooth,Maynooth, County Kildare, Ireland

Abstract

In conventional symmetric encryption, it is common for the encryption/decryption key to be reused for multiple plaintexts. This gives rise to the concept of a known-plaintext attack. In optical image encryption systems, such as double random phase encoding (DRPE), this is also the case; if one knows a plaintext-ciphertext pair, one can carry out a known-plaintext attack more efficiently than a brute-force attack, using heuristics based on phase retrieval or simulated annealing. However, we demonstrate that it is likely that an attacker will find an imperfect decryption key using such heuristics. Such an imperfect key will work for the known plaintext-ciphertext pair, but not an arbitrary unseen plaintext-ciphertext pair encrypted using the original key. In this paper, we illustrate the problem and attempt to characterise the increase in security it affords optical encryption.

Keywords: Optical information processing, Optical image processing, Optical image encryption

1 Introduction

Optical encryption has received much attention in recent years; the reason can be primarily attributed to some of its distinct advantages over conventional digital electronic hardware and software encryption. Double random phase encoding (DRPE), proposed by Réfrégier and Javidi in 1995 [Refregier and Javidi, 1995], is one of the most studied and extended technologies in optical encryption to date. A security concern about optical encryption was first reported by Carnicer et al. in 2005, where a chosen-ciphertext attack (CCA) was introduced to find the exact decryption key [Carnicer et al., 2005]. Subsequently, Peng et al. proposed a chosen-plaintext attack (CPA) to extract the exact key [Peng et al., 2006a], as well as a proposal that the original key could be obtained by solving a linear system of equations from Frauel et al. [Frauel et al., 2007].In a more practical circumstance, if an attacker only has one plaintext-ciphertext pair, a phase-retrieval algorithm [Peng et al., 2006b] or an heuristic algorithm [Gopinathan et al., 2006] can obtain an approximation of the decryption key. A multiplicity of known pairs could be used to significantly reduce the error in the output image [Situ et al., 2007]. To respond to the above attacks, some DRPE-based security enhancement approaches have introduced additional parameters in the Fresnel domain [Situ and Zhang, 2004] and in the fractional Fourier domain [Unnikrishnan et al., 2000], and have added an extra amplitude mask directly behind the Fourier domain mask [Cheng et al., 2008] as extra keys to force attackers to find improvements from their side. However, the two phase masks are still the main concern of a number of optimized attacks [Kumar et al., 2012, Wang and Zhao, 2012, Zhang et al., 2013, Wang et al., 2015, Li and Shi, 2016].

DRPE is a symmetric encryption algorithm, which means the encryption and decryption steps share the same key. All symmetric encryption keys can only be shared over a secure channel. (It is different from asymmetric key encryption system where the public key used for encryption can be openly shared.) It is inconvenient to apply new key to each subsequent image in symmetric optical encryption because the size of the key is relatively large (routinely hundreds of times that in conventional cryptography). One possible solution is to apply modes of operation to optical encryption [Naughton et al., 2008].

 ISBN 978-0-9934207-2-6

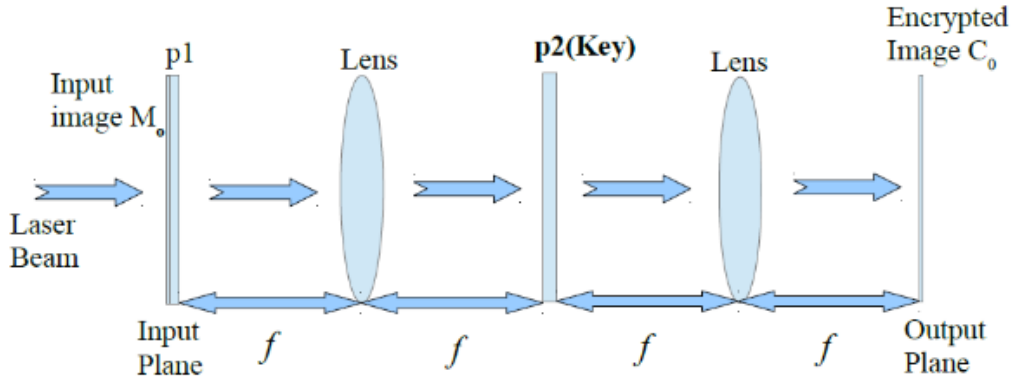

Figure 1: An illustration of symmetric DRPE, showing the location of the two phase-encoded image masks (p_1 and p_2) that constitute the encryption key (and the decryption key). Image mask $p_1$is located immediately after the input image (which effects a pointwise multiplication between the input and p_1). Image mask $p_2$is located in the Fourier plane, where it will be pointwise multiplied by the Fourier transform of the product of the input and p_1. A second Fourier transform returns the encrypted image from the spatial frequency domain to the space domain.

2 Keyspace analysis

The numerical implementation of the encryption process in DRPE (see illustration in Fig.1) can be described as

$$\Psi(x,y) = \mathcal{F}\{\mathcal{F}[f(x,y)\, p_1(x,y)]\, p_2(u,v)\}, \quad (1)$$

where the $p_1(x, y)$ and $p_2(u, v)$ are two statistically independent phase masks representing the encryption key of the system, and $\mathcal{F}$ is a Fourier transform. Each phase key is of the form $\exp(im(x,y))$, where m is a discrete phase mask with height M pixels and width N pixels, and with values randomly taken from the range $[0,2\pi)$. The relevant keyspace has size $K=Q^{2^{(M\times N)}}$, where Q is the number of quantization levels in the phase distribution. A study of the keyspace has been proposed by Monaghan et al. [Monaghan et al., 2007], where a small 3×3 pixels mask with 4 quantisation levels was used. They showed that $4^{3\times3}$ possible keys in the keyspace have to be examined, in the worst case, to decrypt a known plaintext-ciphertext pair. In this discussion, they selected a tolerable decryption error threshold. Multiple keys in the keyspace were found which could decrypt the known ciphertext with an error lower than or equal to the threshold, exactly Q of which (as is well known) were equivalent to the correct key. These Q perfect keys differ from each other only by a constant additive phase $2\pi/Q$.

We determined which of the keys would decrypt subsequently unseen plaintext-ciphertext pairs encrypted using the same original key. We summarise our results using one of the grayscale and one of the binary plaintext images from our experiment. The size for each image continues the use of 3×3 pixels as chosen by Monaghan et al. [Monaghan et al., 2007] and we choose to have 8 quantisation levels in the encryption key. Normalized root mean square (NRMS) error was used to determine the quality of the decrypted output, calculated as

$$E_{\mathrm{NMRS}} = \left\{\frac{\sum_x\sum_y |I_{\mathrm{d}}(x,y)-(I(x,y)|^2}{\sum_x\sum_y |I(x,y)|^2}\right\}^{1/2}, \quad (2)$$

Where I_{d} denotes the intensity of the decryption output and I is the expected intensity. A NRMS error threshold of 0.1 was chosen to decide whether decryption was successful or not. The first mask is not required in the decryption process and therefore the second mask can be regarded as the only decryption key in this system. There are $8^{3\times3}=1.3\times10^9$ possible keys in this keyspace.

Our specific experimental platform runs on a Dell Optiplex 780 desktop PC with an Intel Core™2 Duo E7500 CPU and 4 GB of RAM, running Python 3.5.2 in Linux. For this experiment with binary-valued plaintext

 ISBN 978-0-9934207-2-6

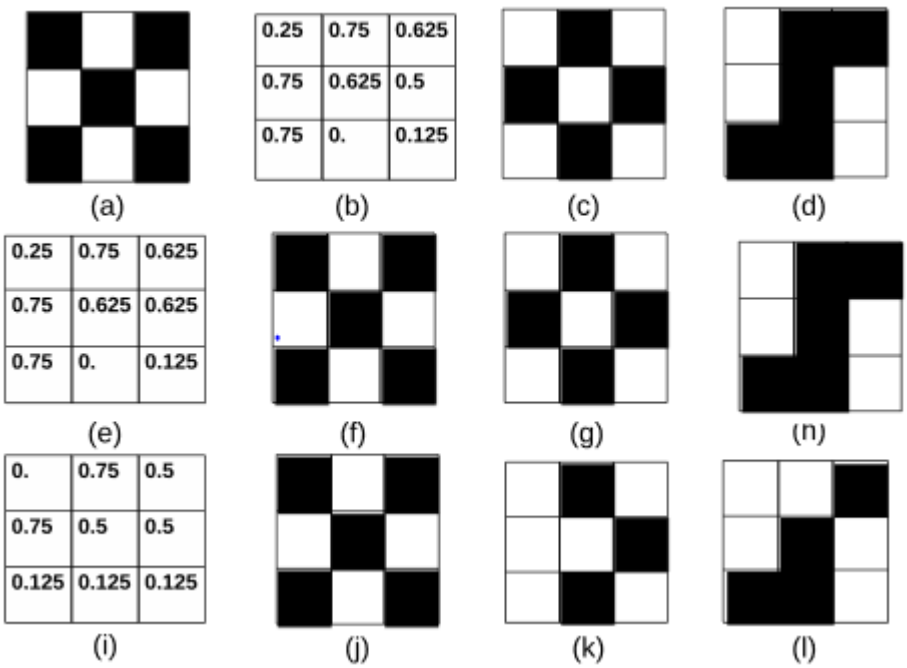

Figure 2: Binary image experiment (all subimages are explained in detail in the main text).

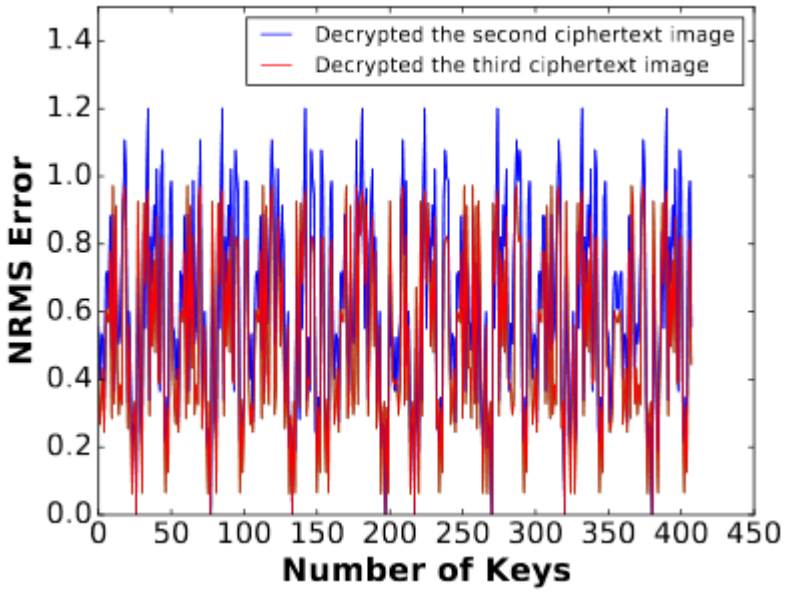

Figure 3: Binary image experiment: for all 408 keys that decrypted the known pair with NRMS error of 0.1 or less, this figure shows how well they decrypted the two unseen images.

 ISBN 978-0-9934207-2-6

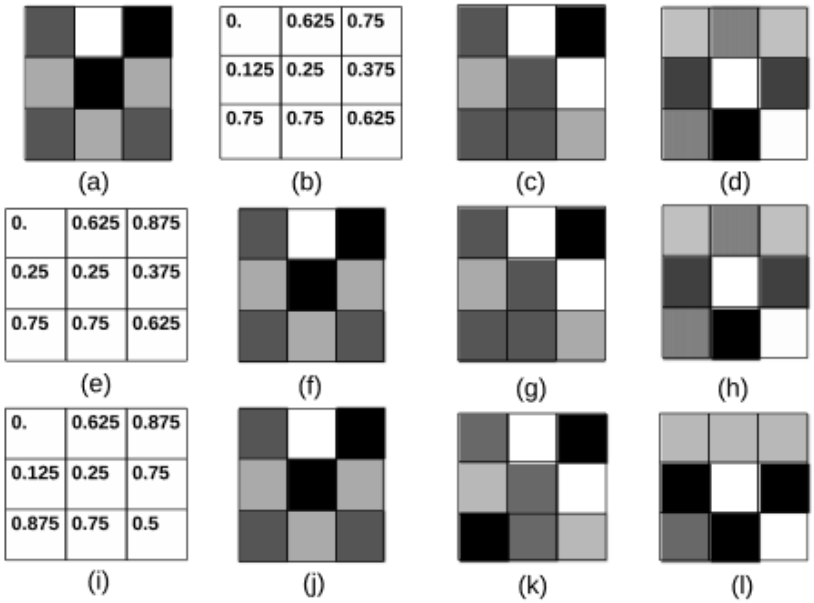

Figure 4: Greyscale image experiment (all subimages have the same meaning as those in Fig. 2).

images, it took approximately 10 hours to investigate all 1.3×10^9 possible keys. Figures 2(a) and (b) show the plaintext part of the known pair and the second random phase mask, respectively, the latter being the phase distribution before being multiplied by 2π. Figures 2(c)-(d) are the second and third binary-valued plaintext images, for which their encrypted versions only will be known to the attacker. Of the 1.3×10^9 possible keys, 408 keys decrypted the encrypted version of the known plaintext image in Fig. 2(a) with a NRMS error of 0.1 or less. Then each of these keys were used to decode encrypted versions of Figs. 2(c) and (d) that were encrypted with the same key Fig. 2(b). The corresponding NRMS errors are plotted in Fig.3, which shows that in general they yield much higher errors with unseen images than the 0.1 error yielded with the known pair. For this second stage, a NRMS error threshold of 0.2 was introduced to discriminate correct decryption (following Monaghan et al. [Monaghan et al., 2007]), followed by the application of a threshold of 0.5 to determine if the binary pixel is white (1) or black (0). From the 408 keys that decrypted the known pair with error up to 0.1, only 24 of them could correctly decrypt both unseen encrypted images, including the Q (Q=8) perfect keys. Another 16 keys produced one correct decryption, with the remainder resulting in errors consistently over 0.2.

Example decryption keys and corresponding decrypted outputs are shown in Figs. 2(e)-(l). Fig. 2(e) is a decryption key with one incorrect pixel – it decrypts encrypted versions of the above three plaintext images with NRMS errors of 0.1, 0.13 and 0.17, respectively, and Figs. 2(f)-(h) are the corresponding outputs. Fig. 2(i) is a decryption key with half of the pixels incorrect – it decrypts the same images with NRMS errors of 0.1, 0.23 and 0.26, respectively, and Figs. 2(j)-(l) are the corresponding outputs. It can seen that although Fig. 2(i) decrypts the known pair with low error, it decrypts the unseen images with higher error.

The experiment was repeated for grayscale 3×3 pixel images. The results are shown in Figs. 4 and 5 and each subimage has the same explanation as those in Figs. 2 and 3. In this test, 120 keys were able to decrypt the known pair with NRMS error of 0.1 or less. Of these, only 32 could successfully decrypt both subsequent unseen images (using our choice of a threshold of 0.3 being reasonable based on visual inspection). As with the binary image case, only a minority of keys that successfully decrypt the known pair can successfully decrypt the unseen images. The key in Fig. 4(e) decrypts encrypted versions of the three plaintext images with NRMS errors of 0.1, 0.21 and 0.22, respectively. The key in Fig. 4(i) decrypts encrypted versions of the three plaintext images with NRMS errors of 0.1, 0.33 and 0.35, respectively. It can seen that although Fig. 4(i) decrypts the known pair with low error, it decrypts the unseen images with higher error.

We define "imperfect keys" as those keys that decrypt the known pair successfully, but do not consistently decrypt unseen images successfully. Of keys that decrypt the known pair, these imperfect keys are in the majority. They disrupt the job of the attacker and their presence increases the security of optical encryption

 ISBN 978-0-9934207-2-6

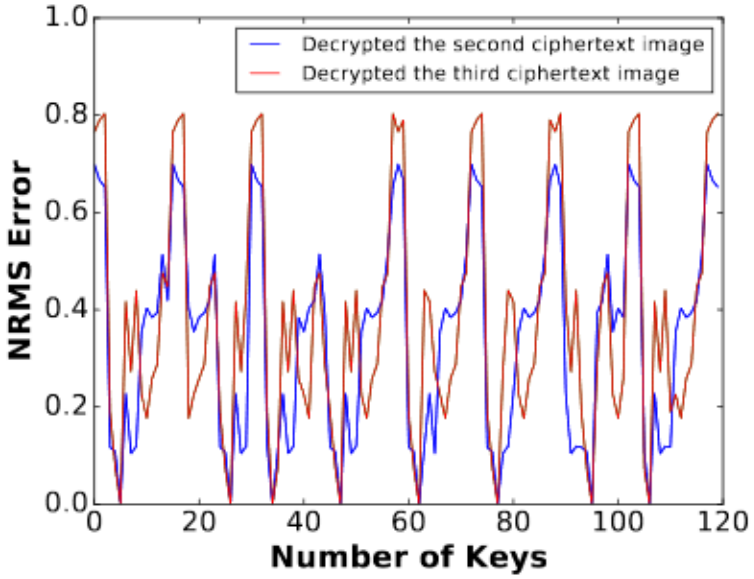

Figure 5: Greyscale image experiment: for all 120 keys that decrypted the known pair with NRMS error of 0.1or less, this figure shows how well they decrypted the two unseen images.

because an attacker cannot know from one known plaintext pair if they have deduced an approximation of aperfect key or a (relatively useless) imperfect key.

3 Investigation of large keyspaces

The examined keyspaces in the previous section were of small sized phase keys. In practice, the plaintext image would have at least two orders of magnitude more pixels. Relatively, the cryptanalysis of the relevant keyspace becomes computationally difficult. In order to prove the existence of the imperfect key in the large size of the keyspace, such as 64×64 pixels, we intentionally selected the keys found using a simulated annealing algorithm. This heuristic approach designed to the DRPE has been proposed by Gopinathan et al. [Gopinathan et al., 2006]. The prerequisite of the SA algorithm is one known plaintext-ciphertext pair and more ciphertexts all encrypted with the same key, that is well fitted with our analysis. In that paper, the keys found in the SA algorithm based on a binary image pair have been examined to decrypt the second unseen image, the decrypting NRMS errors for 32×32 pixels and 64×64 pixels plaintexts were both close to an average of 0.4. Some examples are shown in Fig. 6(a)-(d). Moreover, we have complemented the test adopting grayscale image, the errors to decrypt the second unseen image found that sometime reached to the NRMS of 0.8, which is consistent with the worst cases shown in the Fig. 5. The threshold for the known pair remains the NRMS of 0.1.

As considered binary image is more immune to the noise than grayscale image, the peak error of 0.4 that still provides a tolerable visibility, referring to the Fig. 6(d). Referring to the binary plaintext image can not support details as many as the grayscale image, it is not an ideal choice for hiding text or graph, we do not further discuss it. On the other side, when the noise in the grayscale output exceeds NRMS of 0.8 the attacker can not recover any useful information. It is worth to mention the collision problem in DRPE that also arouse high noise in result, as shown in Fig. 6(j). Collision is commonly inevitable phenomenon in a linear system, such as DRPE. More details of the collision problem can be referred to [Situ et al., 2008]. That suggests to use grayscale image rather than binary image in DRPE to resist the known-plaintext attack.

4 Regions in the keyspace

There is no doubt that all correct but imperfect keys can be presented as adding a small amount of noise to the perfect keys. For example, the key in Fig. 2(e) and Fig. 4(e). That implies a small region around the perfect

 ISBN 978-0-9934207-2-6

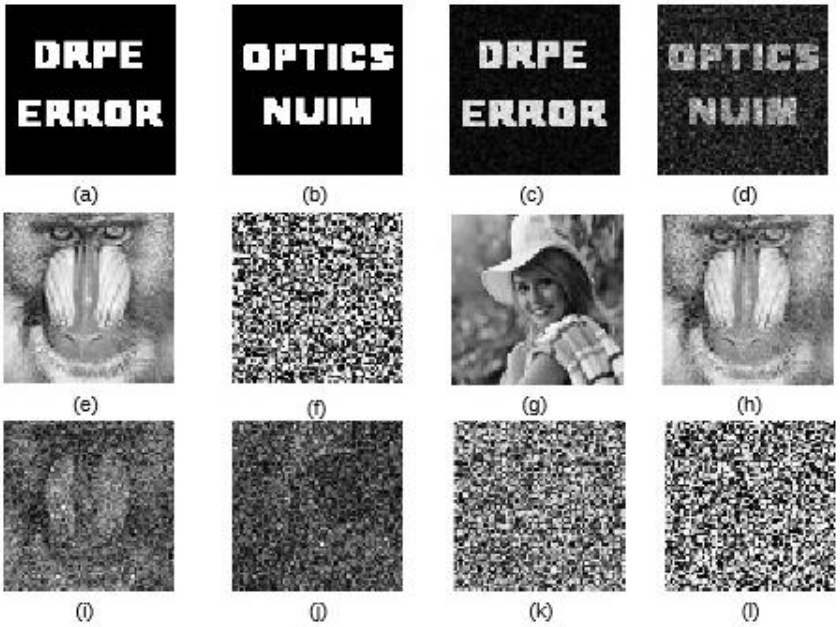

Figure 6: Original and decrypted 64×64 pixel images using the SA algorithm. (a) and (b) are the plaintext part of the known binary pair and a second pair, respectively; (c) and (d) are the decrypted versions using the SA algorithm yielding NRMS errors of 0.1 and 0.4, respectively; (e) is the input part of the known grayscale pair; (f) is the encryption key with 256 Q-levels; (g) is the second grayscale plaintext; (h) is the decrypted version of the original input image, with NRMS error of 0.09; (i) and (j) are decrypted versions of the second unseen image; (k) is the imperfect key found using the SA algorithm; (l) is a highly approximated version of the key in (f).

key in which contains the correct keys (decryption error in 0.1) in keyspace. The theory of region is previously introduced in [Situ et al., 2010]. We also believe that there is the region of imperfect keys in the keyspace.

Table.1 illustrates the increased amount of noise added to a imperfect and an approximate key to explore the region of each that guarantees an average of NRMS of 0.1, the experimental keys were found using the SA algorithm and produced by adding a slight noise to the original encryption key, shown in Fig. 6(k) and (l),respectively, the two keys decrypts the known pair yielding identically NRMS of 0.09. The known plaintext was chosen from the Fig. 6(e), and the same encryption key (f) was reused as well. In this trial, the additional noise was presented as a matrix with the equal size of the keys, and to be randomly produced based on the normal(Gaussian) distribution algorithm which simulates equivalent possibility for all pixels of the trial keys to receive a random phase error. We used one phase-level in phase distribution ($2\pi/Q$) as the unit of the adding noise, Q=256 in the encryption key. The mean parameter of the normal distribution algorithm in this trial was fixed as 0, and its standard deviation(STD) was initialized as 1.0 that indicates noise added to pixels are mainly centralized at one or two phase-levels at the beginning. For the same level of STD, noise was randomly generated for 10 thousand times and each simultaneously added to both experimental keys to generate new keys. If the average error to decrypt the known pair using the new keys is under NRMS of 0.1, the STD of the algorithm increases 0.1 to perform higher amount of noise added to both keys.

Table 1: The table shows the result of additive normal (Gaussian) noise with a standard deviation (STD) of 1.4. Each column shows an average of ten thousand simulations, the mean and standard deviation of decryption errors are listed, followed by the maximum and the minimum error in the decrypted output. The last two columns show the NRMS errors of using the newly found keys to decrypt the second unseen image.

Key category	STD(noise)	Mean(NRMS)	STD(NRMS)	Max(min)	Mean(NRMS)_{2nd}	Max(min)_{2nd}
Imperfect	1.4	0.101	0.001	0.103(0.099)	0.823	0.827(0.823)
Approx	1.4	0.102	0.001	0.106(0.099)	0.094	0.101(0.094)

 ISBN 978-0-9934207-2-6

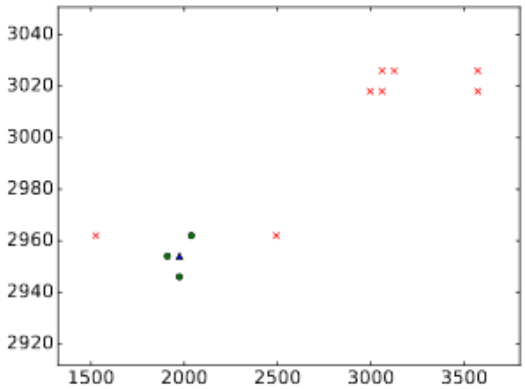

Figure 7: A portion of the keyspace (explained in the text). The total size of the keyspace is $(3\times3)^8$.

The two trial keys showed almost identically capability to accept noise to estimate more keys, that yields errors in a close range (see in Table. 1), the STD was both stopped at 1.4. Furthermore, the new keys derived from the imperfect key are exactly consistent style, the evidence is shown in the last column of Table.1. It is clear that how many of the correct regions in the keyspace that is equivalent to the number of perfect keys in the keyspace. We presume that the size of correct region to be larger than any of the imperfect regions in the keyspace, because the perfect key would provide zero error in decryption but no imperfect could do, that implies the perfect key would undertake higher noise to still satisfy the error threshold.

5 Classification of the keyspace

In this context, the keyspace of the DRPE can be classified as,

1. Incorrect keys, this kind of keys occupied the main part of the keyspace, which is unable to decrypt the known pair yielding NRMS errors in a preset threshold.
2. Perfect keys, they provide zero noise in the output intensity, the exact number is equivalent to the Q-levels in the encryption key.
3. Approximations of perfect keys, the high approximation of the perfect keys that decrypt the known pair with a tolerable noise, these keys can be expected to decrypt all ciphertext with consistently low errors.
4. Imperfect keys, the keys can ideally decrypt the known image pair but which unable to decrypt all ciphertexts within a reasonable error range.

Figure 7 is an illustration of a significant part of the keyspace. The exact data is originated from a previous trial (see in Fig. 4). Each of the possible keys has a unique index according to the sequence of it being examined, the corresponding coordinate is calculated and drawn on a 2d map. We use several remarks to represent different keys, such as, the triangle means the perfect key, the cross stands for the imperfect keys and the circle denotes the approximate key. Fig. 4 typically reflects the characteristic of the entire keyspace, also strongly supports our analysis of the keyspace, the perfect keys (triangle) appear always closely attached with a few approximate keys (circle), the evidence of the correct regions. Theoretically, there is only one correct region for each of the perfect keys. Besides, we notice that one imperfect key separately located at the right and left side of the correct region, that mean the flexible choice of different error threshold would inappropriately determine some keys actually close to the perfect key. Meanwhile, the imperfect keys in Fig. 7 display as clusters or lines, that is regarded as the imperfect region. Fig. 7 can be mapped into the keyspace by continuously adding a constant phase of1/Q.

6 Conclusion

In optical encryption, many previous studies have considered a known-plaintext attack. A commonality among these studies has been them seeking a highly approximated decryption key using an efficient heuristic algorithm

 ISBN 978-0-9934207-2-6

rather than an exhaustive attack. In this paper, we show that these attacks are not robust, and that while the key found will decrypt the known encryption/decryption pair, it is not likely to decrypt unseen images encrypted with the same key. This implies that optical encryption may not be as susceptable to plaintext attacks as previously reported.

Acknowledgements. This publication has emanated from research conducted with the financial support of an Irish Research Council (IRC) Postgraduate Scholarship and of Science Foundation Ireland (SFI) under grant no. 13/CDA/2224.

References

[Carnicer et al., 2005] Carnicer, A., Montes-Usategui, M., Arcos, S., and Juvells, I. (2005). Vulnerability to chosen-cyphertext attacks of optical encryption schemes based on double random phase keys. *Opt. Lett.*, 30(13):1644–1646.

[Cheng et al., 2008] Cheng, X. C., Cai, L. Z., Wang, Y. R., Meng, X. F., Zhang, H., Xu, X. F., Shen, X. X., and Dong, G. Y. (2008). Security enhancement of double-random phase encryption by amplitude modulation. *Opt. Lett*., 33(14):1575–1577.

[Frauel et al., 2007] Frauel, Y., Castro, A., Naughton, T. J., and Javidi, B. (2007). Resistance of the double random phase encryption against various attacks. *Opt. Express*, 15(16):10253–10265.

[Gopinathan et al., 2006] Gopinathan, U., Monaghan, D. S., Naughton, T. J., and Sheridan, J. T. (2006). A known-plaintext heuristic attack on the fourier plane encryption algorithm. *Opt. Express*, 14(8):3181–3186.

[Kumar et al., 2012] Kumar, P., Kumar, A., Joseph, J., and Singh, K. (2012). Vulnerability of the security enhanced double random phase-amplitude encryption scheme to point spread function attack. *Optics and Lasers in Engineering*, 50(9):1196 – 1201.

[Li and Shi, 2016] Li, T. and Shi, Y. (2016). Vulnerability of impulse attack-free four random phase mask cryptosystems to chosen-plaintext attack. *Journal of Optics*, 18(3):035702.

[Monaghan et al., 2007] Monaghan, D. S., Gopinathan, U., Naughton, T. J., and Sheridan, J. T. (2007). Key-space analysis of double random phase encryption technique. *Appl. Opt*., 46(26):6641–6647.

[Naughton et al., 2008] Naughton, T. J., Hennelly, B. M., and Dowling, T. (2008). Introducing secure modes of operation for optical encryption. *J. Opt. Soc*. Am. A, 25(10):2608–2617.

[Peng et al., 2006a] Peng, X., Wei, H., and Zhang, P. (2006a). Chosen-plaintext attack on lensless double-random phase encoding in the fresnel domain. *Opt. Lett.*, 31(22):3261–3263.

[Peng et al., 2006b] Peng, X., Zhang, P., Wei, H., and Yu, B. (2006b). Known-plaintext attack on optical encryption based on double random phase keys. *Opt. Lett.*, 31(8):1044–1046.

[Refregier and Javidi, 1995] Refregier, P. and Javidi, B. (1995). Optical image encryption based on input plane and fourier planerandom encoding. *Opt. Lett*., 20(7):767–769.

[Situ et al., 2007] Situ, G., Gopinathan, U., Monaghan, D. S., and Sheridan, J. T. (2007). Cryptanalysis of optical security systems with significant output images. *Appl. Opt*., 46(22):5257–5262.

[Situ et al., 2008] Situ, G., Monaghan, D. S., Naughton, T. J., Sheridan, J. T., Pedrini, G., and Osten, W. (2008). Collision in double random phase encoding. *Optics Communications*, 281(20):5122 – 5125.

[Situ et al., 2010] Situ, G., Pedrini, G., and Osten, W. (2010). Strategy for cryptanalysis of optical encryption in the fresnel domain. *Appl. Opt*., 49(3):457–462.

[Situ and Zhang, 2004] Situ, G. and Zhang, J. (2004). Double random-phase encoding in the fresnel domain. *Opt. Lett.*, 29(14):1584–1586.

[Unnikrishnan et al., 2000] Unnikrishnan, G., Joseph, J., and Singh, K. (2000). Optical encryption by double-random phase encoding in the fractional fourier domain. *Opt. Lett.*, 25(12):887–889.

[Wang and Zhao, 2012] Wang, X. and Zhao, D. (2012). A special attack on the asymmetric cryptosystem based on phase-truncated fourier transforms. *Optics Communications*, 285(6):1078 – 1081.

[Wang et al., 2015] Wang, Y., Quan, C., and Tay, C. J. (2015). Improved method of attack on an asymmetric cryptosystem based on phase-truncated fourier transform. *Appl. Opt.*, 54(22):6874–6881.

[Zhang et al., 2013] Zhang, Y., Xiao, D., Wen, W., and Liu, H. (2013). Vulnerability to chosen-plaintext attack of a general optical encryption model with the architecture of scrambling-then-double random phase encoding. *Opt. Lett.*, 38(21):4506–4509.