The coin flipping selector for selective encryption

Richard Ostertág*

Department of Computer Science,
Faculty of Mathematics, Physics and Informatics,
Comenius University,
Mlynská dolina, 842 48 Bratislava, Slovak Republic
ostertag@dcs.fmph.uniba.sk
http://www.dcs.fmph.uniba.sk

Abstract. *Some applications require high-speed encryption even at the expense of reduced security. With a fixed secure, but slow cryptographic algorithm, there still is an appealing possibility for encryption speedup by encrypting only some portion of data. In this paper we analyze the ciphertext security obtained this way. We show that it is not possible to exclude from encryption even a small constant fraction of data without significantly compromising security.*

1 Motivation, assumptions, goals

Volume of data is nowadays bigger than ever. Multimedia are a typical example. Fast real-time on-demand encryption of multiple multimedia streams requires specialized powerful hardware.

It is sometimes not possible (or economical) to use powerful enough hardware solution. Then we can replace the encryption algorithm with a faster – although maybe less secure one. Another possibility is to use selective encryption with the original secure algorithm. In this case we encrypt only some fraction of plaintext. Let p denote the fraction of encrypted plaintext. The parameter p ranges between 0 (no encryption) and 1 (full encryption) and is used to control the balance between the encryption speedup and the security.

For example, selective encryption is used for online encryption of MPEG video [1]. In this case, the knowledge of the internal data structure is exploited in order to encrypt only DC coefficients and sign bits of motion vectors. Similar techniques are also used for pictures [2]. For overview of selective encryption methods see [3]. Security of these algorithms is not formally proved.

We formally analyze security of selective encryption in this paper. As we are interested in a general case, we make no assumptions on the internal data structure or on statistical properties of the plaintext.

We originally hoped that it could be possible to selectively encrypt portion of plaintext while maintaining reasonable security. However, we show that this does not work. Since we prove a negative result, it is only better if assumptions are more disadvantageous for the attacker than in practical usage:

1. *One-time pad is used as the encrypting algorithm.* One-time pad is the first and only encryption algorithm for which there is a proof of perfect secrecy if the key is truly random, never reused, and kept secret. We choose this cipher to abstract from eventual weaknesses of the actual cipher which can be exploited by attacker. Theoretical results obtained this way can be used in practice as upper bounds for security of any other selected encryption algorithm.
2. *Attacker can manage no more than ciphertext-only attack.* The attacker is assumed to have access only to a ciphertext and full description of selective encryption algorithm. This means that the attacker knows the enciphering algorithm and also the method of bit selection for enciphering.
3. *Attack is peformed using brute force.* Key space is searched from the most probable key to the least probable key omitting impossible keys to minimize the attacker's work. We assume that the selection algorithm chooses bits for encrypting independently from plaintext content (besides its length). In general it cannot be expected that a better attack is possible. However in actual situation specific properties[1] of plaintext can lead to a more efficient attack.
4. *Attack complexity measure is defined as a fraction of key space that attacker has to search in average to find the key.* Attacker tries every possible key until he finds one that deciphers to the desired plaintext. We ignore the complexity of verifying whether deciphered plaintext is the original one. For selective encryption with $p = 1$ (one-time pad), the expected attack complexity is $1/2$. For selective encryption with $p = 0$ expected complexity is 0. We

* Supported by VEGA grant No. 1/3106/06.

[1] E.g. high redundancy of plaintext poses an even greater risk for selective encryption then for full text encryption.

consider every cipher for which attack complexity approaches 0 as plaintext length goes to $+\infty$ insecure.

We assume that encrypting p percent of plaintext bits with selective encryption reduces sender's work to p percent omitting overhead necessary for selecting those bits. In this situation we will be satisfied with (and accept this as reasonable degradation of security) reduction of attack complexity from $1/2$ to $p/2$, because this means that attacker's work is in average also reduced to p percent but no more.

2 Selectors

In this paper we will assume that plaintext is a bit sequence – sequence of zeros and ones. Let $n > 0$ denote plaintext sequence length. If we want to selectively encrypt $p \in (0, 1)$ percent of plaintext, then we have to choose $k = np$ bits of plaintext for encryption. In [4] we analyzed different ways of bit selection for selective encryption. We introduced the notion of selector – algorithm which performs selection of the k bits for encryption based on n and p. The output of selector on input n and p is a bit sequence of length n with $k = np$ ones – indicating positions of bits chosen for encryption. Selective encryption algorithm proceeds in the following way:

1. The selector selects $k = np$ bits for encryption.
2. Encrypt only selected bits with one-time pad[2].

As it can be seen our model was limited to selections which have exactly $k = np$ bits selected. In [4] we proved that among the analyzed selectors only fully random selection of exactly p percents of bits provides reasonable security for $p \geq 1/2$. In this place it is necessary to mention that in [4] we measured the attack complexity by the number of possible plaintexts[3].

Because we are interested in the values of $p < 1/2$ we relax the assumption that exactly $k = np$ bits have to be selected, and we only require that in average k bits have to be selected. This relaxation allows for using a selector which for every bit flips a biased coin – one falls with probability p, zero with probability $1 - p$. Lets call this selector coin flipping selector. We hope that this step allow us to go with p below $1/2$ because it introduce more uncertainty to attacker as all plaintext are now possible. For that reason we have to also change our attack complexity measure and we choose one mentioned in previous section.

[2] Xor them with truly random noise.

[3] For example, let $p = 1/2$. For a random bit selector there are 2^{n-1} possible plaintexts for every ciphertext. If the selector do not use randomness and deterministically selects every even bit, there are only $2^{n/2}$ possible plaintexts.

3 The coin flipping selector analysis

In the rest of the paper we will show the behavior of the attack complexity for the coin flipping selector for large messages (we will assume that n goes to infinity).

3.1 Average fraction of key space equation

Firstly we need to determine the probability of the key of length n with exactly k ones on fixedly chosen positions if in the selective encryption the coin flipping selector is used. Let denote this probability as $\mathrm{PK}(n, k, p)$, where p is probability of encrypting.

Theorem 1.

$$\mathrm{PK}(n, k, p) = \left(\frac{p}{2}\right)^k \left(1 - \frac{p}{2}\right)^{n-k}$$

Proof.

$$\mathrm{PK}(n, k, p) = \sum_{i=0}^{n-k} \binom{n-k}{i} p^{k+i} (1-p)^{(n-k)-i} \frac{1}{2^{k+i}},$$

because we can get the key with exactly k ones on fixedly chosen positions from any selection with exactly $k + i$ ones with k ones on those fixedly chosen positions and i ones arbitrarily chosen from remaining $n - k$ positions. Also one-time pad has to select for those k positions bit 1 and for remaining i positions bit 0 (thus we get $2^{-(k+i)}$). We can simplify the last equation as follows:

$$\left(\frac{p}{2}\right)^k \sum_{i=0}^{n-k} \binom{n-k}{i} \left(\frac{p}{2}\right)^i (1-p)^{(n-k)-i} =$$

$$= \left(\frac{p}{2}\right)^k \left(\frac{p}{2} + (1-p)\right)^{n-k} = \left(\frac{p}{2}\right)^k \left(1 - \frac{p}{2}\right)^{n-k}.$$

□

Derivative of the function $\mathrm{PK}(n, k, p)$ with respect to k is:

$$\mathrm{PK}(n, k, p) \ln\left(\frac{p}{2-p}\right).$$

Since for all studied $n > 0$ and $0 < p < 1$ expression $\ln(\frac{p}{2-p})$ is negative and $\mathrm{PK}(n, k, p)$ is positive we know, that $\mathrm{PK}(n, k, p)$ is strictly decreasing function with respect to $k \in \langle 0, n \rangle$. Thus effective attacker will start searching key space from the most probable 0^n key to the least probable 1^n key in direction of increasing number of ones in the key. Sort all 2^n keys in this order[4] in an array with indexes from 1 to 2^n. Then denote $L(n, k)$ index of first key of length n with k

[4] Ordering of keys with equal number of ones is irrelevant since all have the same probability. It can be arbitrary but fixed.

ones and $U(n,k)$ will denote index of the last key of length n with k ones. It can easily be seen that:

$$L(n,k) = 1 + \sum_{i=0}^{k-1} \binom{n}{i},\ U(n,k) = \sum_{i=0}^{k} \binom{n}{i}.$$

Theorem 2. *Let $I(n,p)$ be a position in the above mentioned array where attacker finds the key in average case. Then $I(n,p)$ equals to:*

$$\frac{1}{2} + \frac{1}{2}\left(\frac{2-p}{p}\right)^n \sum_{k=0}^{n}\left[\left(\frac{p}{2-p}\right)^k \binom{n}{k} \sum_{i=0}^{k}\binom{n+1}{i}\right].$$

Proof. Let $\Pr(n,p,i)$ be probability that attacker finds key in position i. Then:

$$I(n,p) = \sum_{i=1}^{2^n} i \Pr(n,p,i).$$

Since $\Pr(n,p,i)$ is constant for all i between $L(n,k)$ and $U(n,k)$ we can write:

$$I(n,p) = \sum_{k=0}^{n}\left[\sum_{i=L(n,k)}^{U(n,k)} i\,\mathrm{PK}(n,k,p)\right]$$

Since $\mathrm{PK}(n,k,p)$ does not depend on i we can move it in front of inner sum. The inner sum then reduces to:

$$\sum_{i=L(n,k)}^{U(n,k)} i = [U(n,k) - L(n,k) + 1]\frac{L(n,k) + U(n,k)}{2}$$

Thus equation for $I(n,p)$ changes to:

$$\sum_{k=0}^{n} \mathrm{PK}(n,k,p)\frac{1}{2}\binom{n}{k}\underbrace{\left[1 - \binom{n}{k} + 2\sum_{i=0}^{k}\binom{n}{i}\right]}_{\text{Mark this term as } x(n,k).}.$$

It is obvious that $x(n,0) = 2$ and $x(n,k+1) - x(n,k) = \binom{n+1}{k+1}$. So $x(n,k) = 1 + \sum_{i=0}^{k}\binom{n+1}{i}$. After substituting $x(n,k)$ and $\mathrm{PK}(n,k,p)$ we can write $I(n,p)$ as:

$$\frac{1}{2}\sum_{k=0}^{n}\left(\frac{p}{2}\right)^k\left(1-\frac{p}{2}\right)^{n-k}\binom{n}{k}\left[1 + \sum_{i=0}^{k}\binom{n+1}{i}\right].$$

Then after factoring out $\left(1-\frac{p}{2}\right)^n$ we get:

$$\frac{1}{2}\left(1-\frac{p}{2}\right)^n \sum_{k=0}^{n}\left(\frac{p}{2-p}\right)^k\binom{n}{k}\left[1 + \sum_{i=0}^{k}\binom{n+1}{i}\right].$$

By expanding summand and using binomial theorem for $\left(\frac{p}{2-p}+1\right)^n$ we get:

$$\frac{1}{2}\left(1-\frac{p}{2}\right)^n\left(\frac{2}{2-p}\right)^n +$$

$$+\frac{1}{2}\left(1-\frac{p}{2}\right)^n \sum_{k=0}^{n}\left(\frac{p}{2-p}\right)^k\binom{n}{k}\sum_{i=0}^{k}\binom{n+1}{i} =$$

$$= \frac{1}{2} + \frac{1}{2}\left(\frac{2-p}{2}\right)^n \sum_{k=0}^{n}\left(\frac{p}{2-p}\right)^k\binom{n}{k}\sum_{i=0}^{k}\binom{n+1}{i}.$$

□

Let us denote average fraction of key space which attacker has to search before he finds the key as $F(n,p)$. Now, when we have $I(n,p)$, equation for F is obvious:

$$F(n,p) = \frac{\frac{1}{2} + \frac{1}{2}\left(\frac{2-p}{2}\right)^n \sum_{k=0}^{n}\left(\frac{p}{2-p}\right)^k\binom{n}{k}\sum_{i=0}^{k}\binom{n+1}{i}}{2^n+1}.$$

Although we have assumed that $p < 1$ we can verify, that $F(n,1) = 1/2$ as expected. We can not use $F(n,0)$ because $\Pr(n,0,k)$ is not a valid probability distribution over keys of length n.

3.2 Asymptotics

Based on Figure 1 we will now try to show that for all $p < 1$ holds $\lim_{n\to\infty} F(n,p) = 0$. This will be unwelcome result. It means that even if we encrypt nearly the entire plaintext up to some small fraction, this small fraction is still sufficient to reduce attack complexity to a negligible fraction compared to full text encryption.

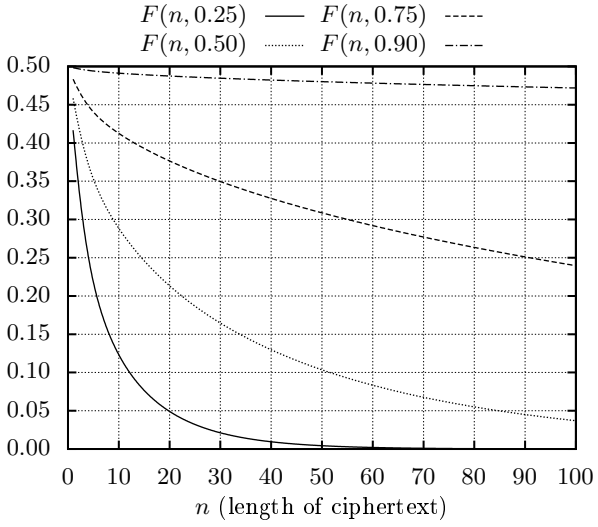

Fig. 1. This graph indicates that $\lim_{n\to\infty} F(n,p) = 0$.

Since we want to prove that the limit goes to zero, it is possible to simplify the proof by realizing that

$F(n,p) \geq 0$ and show that some simpler upper bound $f_0(n,p)+f_1(n,p)+f_2(n,p) \geq F(n,p)$ goes to zero too. We choose $f_i(n,p)$ as follows

$$f_0(n,p) = \frac{1}{2^{n+1}},$$
$$f_1(n,p) = \frac{1}{2^{n+1}}\left(\frac{2-p}{2}\right)^n \sum_{k=0}^{\alpha}\left(\frac{p}{2-p}\right)^k \binom{n}{k} S_k^{n+1},$$
$$f_2(n,p) = \frac{1}{2^{n+1}}\left(\frac{2-p}{2}\right)^n \sum_{k=\alpha}^{n}\left(\frac{p}{2-p}\right)^k \binom{n}{k} S_k^{n+1},$$

where S_k^{n+1} denotes $\sum_{i=0}^{k}\binom{n+1}{i}$ and α is $\frac{n}{2}-\frac{1}{2}n^{\frac{5}{8}}$. To prove the main limit it is sufficient to show that for all $i \in \{0,1,2\}$ $\lim_{n\to\infty} f_i(n,p)$ equals to zero. For $i=0$ it is trivial so we move to $i=1$. In the proof we will use following lemma.

Lemma 1 (for proof see [5]). *Let $\varphi(n)$ be any function satisfying $\lim_{n\to\infty}\varphi(n)=\infty$. Then*

$$\lim_{n\to\infty}\frac{\sum_{k=0}^{n/2-\varphi(n)\sqrt{n}}\binom{n}{k}}{2^n} = 0.$$

Theorem 3. *Let $p<1$. Then $\lim_{n\to\infty} f_1(n,p)=0$.*

Proof. Firstly we replace $f_1(n,p)$ with even simpler upper bound:

$$f_1(n,p) \leq \frac{1}{2^{n+1}}\left(\frac{2-p}{2}\right)^n S_\alpha^{n+1}\sum_{k=0}^{\alpha}\left(\frac{p}{2-p}\right)^k\binom{n}{k} \leq$$
$$\leq \frac{1}{2^{n+1}}\left(\frac{2-p}{2}\right)^n S_\alpha^{n+1}\sum_{k=0}^{n}\left(\frac{p}{2-p}\right)^k\binom{n}{k} =$$
$$= \frac{1}{2^{n+1}}\left(\frac{2-p}{2}\right)^n S_\alpha^{n+1}\left(\frac{2}{2-p}\right)^n = \frac{S_\alpha^{n+1}}{2^{n+1}}$$

Now we can set $\varphi(n) = \frac{(n-1)^{\frac{5}{8}}+1}{2\sqrt{n}}$ and use lemma 1:

$$\lim_{n\to\infty}\frac{S_\alpha^{n+1}}{2^{n+1}} = \lim_{n\to\infty}\frac{\sum_{i=0}^{\frac{n}{2}-\frac{1}{2}n^{\frac{5}{8}}}\binom{n+1}{i}}{2^{n+1}} = 0.$$

Becase $f_1(n,p)\geq 0$ the theorem is proved. □

In the proof for $i=2$ we will utilize another two lemmas.

Lemma 2 (for proof see [6] equation 9.98). *Let $|k| \leq \frac{1}{2}n^{\frac{5}{8}}$. Then binomial coefficient around center for $n\to\infty$ can be aproximated as follows:*

$$\binom{n}{\frac{n}{2}-k} = \frac{2^n}{\sqrt{\frac{\pi}{2}n}}e^{-2\frac{k^2}{n}}\left(1+O\left(n^{-\frac{1}{8}}\right)\right)$$

Lemma 3. *Let $a_k = \left(\frac{p}{2-p}\right)^k\binom{n}{k}$ for $k\in\{0,1,\ldots,n\}$. Then the following inequality holds:*

$$\forall p\in(0,1)\ \exists m\ \forall n>m\ \forall k\geq\frac{n}{2}-\frac{1}{2}n^{\frac{5}{8}} : a_k > a_{k+1}.$$

Proof. We rewrite theorem inequality as a fraction. So we get $\forall p\in(0,1)\ \exists m\ \forall n>m\ \forall k\geq\frac{n}{2}-\frac{1}{2}n^{\frac{5}{8}}$:

$$\frac{a_{k+1}}{a_k} < 1 \Leftrightarrow \frac{p}{2-p}\frac{n-k}{k+1} < 1 \Leftrightarrow \frac{n-k}{k+1} < \frac{2-p}{p}.$$

Because $\frac{n-k}{k+1} < \frac{n-k}{k}$ it is sufficient to find an arbitrary m that $\forall n>m\ \forall k\geq\frac{n}{2}-\frac{1}{2}n^{\frac{5}{8}}$:

$$\frac{n-k}{k} < \frac{2-p}{p} \Leftrightarrow n < k + k\frac{2-p}{p} \Leftrightarrow n < k\frac{2}{p}$$

By solving this inequality we get that it holds for every $k > \frac{p}{2}n$. Because we start with $k=\frac{n}{2}-\frac{1}{2}n^{\frac{5}{8}}$, we will now solve for which n holds:

$$\frac{n}{2}-\frac{1}{2}n^{\frac{5}{8}} > \frac{p}{2}n \Leftrightarrow 1-\frac{1}{n^{\frac{3}{8}}} > p \Leftrightarrow n > \left(\frac{1}{1-p}\right)^{\frac{8}{3}}.$$

Now we have showed that for all $p\in(0,1)$, if we set m to $\left(\frac{1}{1-p}\right)^{\frac{8}{3}}$, then

$$\forall n>m\ \forall k\geq\frac{n}{2}-\frac{1}{2}n^{\frac{5}{8}} : a_k > a_{k+1}.$$

□

Now we are able to proof the last theorem on the limit of function f_2.

Theorem 4. *Let $p<1$. Then $\lim_{n\to\infty} f_2(n,p)=0$.*

Proof. Again we replace $f_2(n,p)$ with even simpler upper bound:

$$\frac{1}{2^{n+1}}\left(\frac{2-p}{2}\right)^n\sum_{k=\alpha}^{n}\left(\frac{p}{2-p}\right)^k\binom{n}{k}\underbrace{S_k^{n+1}}_{\leq 2^{n+1}} \leq$$
$$\leq \left(\frac{2-p}{2}\right)^n\sum_{k=\alpha}^{n}\underbrace{\left(\frac{p}{2-p}\right)^k\binom{n}{k}}_{a_k}$$

Now we use lemma 3 and the fact that we are interested in limit for $n\to\infty$. Thus for large enough n we can upper bound the last equation by

$$\left(\frac{2-p}{2}\right)^n (n-\alpha+1)\,a_\alpha =$$
$$= \left(\frac{2-p}{2}\right)^n (n-\alpha+1)\left(\frac{p}{2-p}\right)^\alpha\binom{n}{\alpha} \leq$$

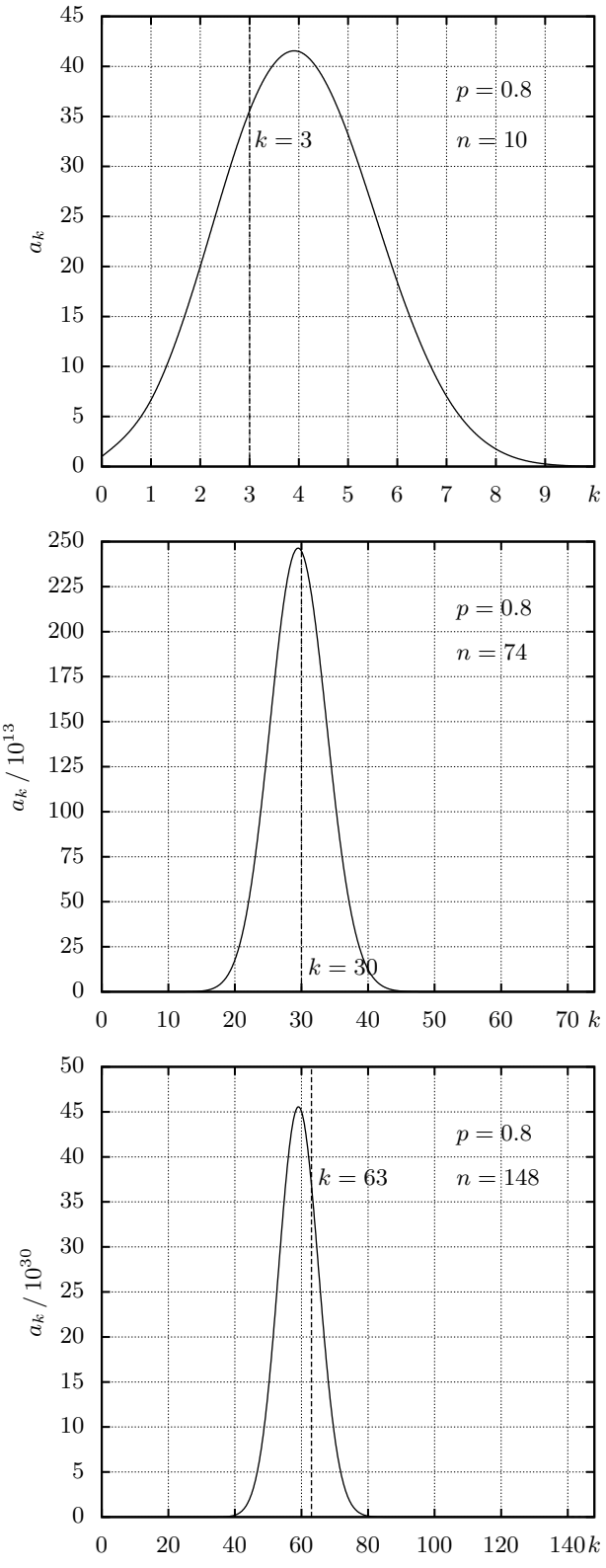

Fig. 2. These graphs illustrate how value of a_k starts to decrease from $k = \frac{n}{2} - \frac{1}{2}n^{\frac{5}{8}}$ if large enough n is used ($n > m$). For $p = 0.8$ based on lemma 3 we get that m approximately equals to 73.1.

$$\leq \left(\frac{2-p}{2}\right)^n n \left(\frac{p}{2-p}\right)^{\alpha} \binom{n}{\alpha} =$$

$$= \left(\frac{2-p}{2}\right)^n \left(\frac{p}{2-p}\right)^{\frac{n}{2}} \left(\frac{2-p}{p}\right)^{\frac{1}{2}n^{\frac{5}{8}}} n\binom{n}{\alpha} =$$

$$= \frac{[(2-p)p]^{\frac{n}{2}}}{2^n} \left(\frac{2-p}{p}\right)^{\frac{1}{2}n^{\frac{5}{8}}} n\binom{n}{\frac{n}{2} - \frac{1}{2}n^{\frac{5}{8}}}.$$

In the sequel we get rid of binomial coefficient by applying lemma 2. We omit the $1 + O(n^{-\frac{1}{8}})$ factor from following equations to save space.

$$\frac{[(2-p)p]^{\frac{n}{2}}}{2^n} \left(\frac{2-p}{p}\right)^{\frac{1}{2}n^{\frac{5}{8}}} n \frac{2^n}{\sqrt{\frac{\pi}{2}n}} e^{-2\frac{\left(\frac{1}{2}n^{\frac{5}{8}}\right)^2}{n}} =$$

$$= \sqrt{\frac{2}{\pi}}[(2-p)p]^{\frac{n}{2}} \left(\frac{2-p}{p}\right)^{\frac{1}{2}n^{\frac{5}{8}}} \sqrt{n}\, e^{-\frac{1}{2}\sqrt[4]{n}}$$

We want to prove now that the last equation goes to zero as n approaches ∞. We will do so by showing that logarithm of the equation goes to $-\infty$. We also omit constant factor $\sqrt{2/\pi}$ as it is irrelevant in this context.

$$\frac{n}{2}\ln(2-p)p + \frac{1}{2}n^{\frac{5}{8}} \ln\left(\frac{2-p}{p}\right) + \frac{1}{2}\ln n - \frac{1}{2}\sqrt[4]{n} + O\left(n^{-\frac{1}{8}}\right)$$

Since $\frac{n}{2}\ln(2-p)p$ is most influencing summand as n goes to infinity and $\ln(2-p)p < 0$ we have proved that the equation goes to $-\infty$. If we again omit the $1 + O(n^{-\frac{1}{8}})$ factor from the right-hand side of inequality we get for large enough n that

$$0 \leq f_2(n,p) \leq \sqrt{\frac{2}{\pi}}[(2-p)p]^{\frac{n}{2}} \left(\frac{2-p}{p}\right)^{\frac{1}{2}n^{\frac{5}{8}}} \sqrt{n}\, e^{-\frac{1}{2}\sqrt[4]{n}}.$$

As we have proved that the right-hand side goes to zero as n goes to infinity, we are done. □

4 Conclusion

In this paper we have showed that even the coin flipping selector tremendously decreases the security of selective encryption for any $p < 1$. In other words it means that even if we encrypt nearly the entire plaintext up to some small fraction, this small fraction is still enough to reduce attack complexity to negligible fraction compared to full text encryption. The same result holds for random bit selector from [4] if the attacker and the attack complexity from this paper is assumed. As a conclusion we can say, that every studied selector significantly degrades security even if the encrypted fraction is closed to 1 for large enough messages.

References

1. Shi, C., Bhargava, B.: A fast mpeg video encryption algorithm. In: Proceedings of the sixth ACM international conference on Multimedia, ACM Press (1998) 81–88
2. Droogenbroeck, M.V., Benedett, R.: Techniques for a selective encryption of uncompressed and compressed images. In: Proceedings of ACIVS 2002 (Advanced Concepts for Intelligent Vision Systems), Ghent, Belgium (2002) 90 – 97
3. Liu, X., Eskicioglu, A.M.: Selective encryption of multimedia content in distribution networks: Challenges and new directions. ASTED International Conference on Communications, Internet and Information Technology (CIIT 2003) (2003)
4. Ostertág, R., Košinár, P.: Analýza selektorov pre selektívne šifrovanie. In: ITAT 2006: Information Technologies-Applications and Theory, Seňa: PONT (2006) 131–137
5. Olejár, D., Stanek, M.: On cryptographic properties of random Boolean functions. J.UCS: Journal of Universal Computer Science **4** (1998) 705–718
6. Graham, R.L., Knuth, D.E., Patashnik, O.: Concrete Mathematics: A Foundation for Computer Science. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1994)