DML Discrete Mathematics Letters
www.dmlett.com

Discrete Math. Lett. **5** (2021) 63–67
DOI: 10.47443/dml.2021.0019

Research Article

Some results on an encryption method using subset-sums of pseudo-recursive sequences

Bence Bakos*, Máté Pálfy

Institute of Mathematics, Eötvös Loránd University, H-1117 Pázmány st. 1/c, Budapest, Hungary

(Received: 27 February 2021. Received in revised form: 24 March 2021. Accepted: 25 March 2021. Published online: 29 March 2021.)

Abstract

Let $A_\eta := \{\lfloor 2^n\eta\rfloor : n = 0, 1, 2, \dots\}$, where $1 \le \eta < 2$. In the recent paper [*Discrete Math. Lett.* **4** (2020) 31–36], the authors examined the subset-sums of the set A_η and then presented an encryption algorithm. The basics of the latter part is that the encoded message is a public natural number N and everyone is allowed to query for the set $(A_\eta \dot{+} A_\eta) \cap [1, N]$ until they find an element of it. Using the results about $P(A_\eta)$, the set of subset-sums of A_η, it was shown that with the secret key γ the message can be decoded in logarithmic time, but for an eavesdropper without the key it takes on average more than N/log^2N time. In this article, we show that the eavesdropper essentially can not figure out the codeword from the found element of S, even if she got that element from the query sequence in a relatively short time.

Keywords: subset-sums; Cantor's representation of integers; encoding a codeword.

2020 Mathematics Subject Classification: 11B30, 11B75.

1. Introduction

The set $\{\lfloor 2^n\eta\rfloor \mid n = 0, 1, \cdots\}$, where $1 \le \eta < 2$, and its related subset-sums were investigated by Erdős and Graham in [4]. Hegyvári also examined some properties of these subset-sums in the articles [2, 3]. Based on these works, in the article [1], an encryption algorithm was examined using the aforementioned set and its subset-sums. In this article, we further investigate that encryption process. We show that if an eavesdropper catches the message, she can not be certain what the codeword was.

For any $\eta \in [1, 2)$, let $a_n := \lfloor 2^n\eta\rfloor$ and $A_\eta := \{a_n \mid n = 0, 1, 2, \cdots\}$. We associate to A_η its so-called subset-sums:

$$P(A_\eta) := \left\{\sum_{i=0}^{\infty} \varepsilon_i a_i \;:\; a_i \in A_\eta;\; \varepsilon_i \in \{0, 1\} \text{ for all } i;\; \sum_i \varepsilon_i < \infty\right\}.$$

It was proven in the article [1] that every element of $P(A_\eta)$ has a unique representation (see Proposition 3.2 in [1]). It was also shown in [1] (see Proposition 3.3 in [1]) that the biggest gap of the set $P(A_\eta) \cap [1, a_k]$ is equal to $\sum_{j=1}^{k} \eta_j$, where $\eta = 1.\eta_1\eta_2\cdots$ in binary form. Using these results, the authors of [1] also presented an encryption algorithm to send a message. The coding process is given as follows:
Let c_n be a binary codeword with n digits: $c_n = \eta_1\eta_2\cdots\eta_n$, where each $\eta_i \in \{0, 1\}$ for $i = 1, 2, \cdots, n$. We associate an $\eta \in [1, 2)$ for the codeword c_n: $\eta = 1.\eta_1\eta_2\cdots\eta_n\cdots$ (after η_n we can extend arbitrarily, only one assumption is important, that we prefer the expression where $\eta_i = 0$ holds infinitely many times). The message is sent by Alice to Bob in the form of $N \in \mathbb{N}$. This N is public in the sense that an eavesdropper (Eve) can catch this message. Alice and Bob on the other hand has a secret key $0 < \gamma < 1$, which is available only for them.

Denote by $B \dot{+} C$ the restricted sum for the sets $B, C \subseteq \mathbb{N}$, i.e.,

$$B \dot{+} C := \{b + c \mid b \in B,\;\; c \in C \text{ and } b \ne c\}.$$

Towards the decoding process, let S be the following set:

$$S := (A_\eta \dot{+} A_\eta) \cap [1, N].$$

For the decoding process, let us define a query function $f : [1, N] \mapsto \{0, 1\}$ such that $f(x) = 0$ if $x \notin S$ and $f(x) = 1$ if $x \in S$. Everyone can query an $(x_0, x_0 + 1, \cdots x_0 + L)$ sequence of integers such that $(f(x_0), f(x_0 + 1), \cdots, f(x_0 + L))$=$(0, 0, \cdots, 0, 1)$. So we can query x_0 and if it is not in S we can query $x_0 + 1$ and so on until we find an element of S. The length of the query sequence is $L_\eta^{x_0}(N) := L$. For simplicity, let us denote by x_L the number $x_0 + L$.

In the 4th section of the article [1], the following two results were presented about this process.

*Corresponding author (bakosbence237@gmail.com).

Theorem 1.1. *If Alice sends the message N for which it holds that*

$$\gamma N \in \left[\sum_{i=0}^{n-1} a_i + 1, a_n\right),$$

then Bob can get the message by asking a query sequence starting by an appropriately chosen x_0 such that

$$L_\eta^{x_0}(N) \leq \log_2 N + 2.$$

Theorem 1.2. *The expected length of the query sequence of an eavesdropper Eve, who chooses the start of the query sequence uniformly at random in $[1, N]$ is*

$$\mathbb{E}(X) \geq \frac{cN}{\log_2^2 N},$$

($c > 0$ absolute).

2. Decomposition lemmas

As we already mentioned, our purpose is to show that Eve can not really recognize the codeword even if she finds an $x_L \in S$ in a relatively short time. We are going to restrict the length of a query sequence by N^β, $0 < \beta < 1$. This is not going to cause problems, because this restriction only makes things harder for Eve, while Bob will still have enough query questions to find the codeword in a short time (see Theorem 1.1). We will return to the right choice of β later.

Assume now that Eve finds an element $x_L \in S$ after querying at most N^β. What she can do is try to decompose x_L into the sum of b and b', where $b, b' \in A_\zeta$ (for some $1 \leq \zeta < 2$) and she hopes that eventually $\zeta = \eta$. This is equivalent to the task of finding b and b', $b + b' = x_L$, $b > b'$ where in binary form the first few digits of b forms exactly b', since both of them are equal to $\lfloor 2^s \zeta \rfloor$ for some $s \in \mathbb{N}$.

Now, we are going to look at this task in a third equivalent form. We search for the pairs (b, k) that satisfy:

$$x_L = b + \left\lfloor \frac{b}{2^k} \right\rfloor \tag{1}$$

where $1 < b < x_L$ and $1 \leq k \leq \lfloor \log_2 x_L \rfloor + 1$. Notice that k determines b, since x_L is fixed. Now, our goal is to guarantee that for many k such b exists. For this, lets analyze (1) for a fixed k.

Lemma 2.1. *For fixed $x_L, k \in \mathbb{N}$, there is a solution for (1) if and only if $x_L \not\equiv -1 \mod (2^k + 1)$.*

Proof. First lets reshape (1):

$$x_L = b + \left\lfloor \frac{b}{2^k} \right\rfloor \iff b + \frac{b}{2^k} - 1 < x_L \text{ and } x_L \leq b + \frac{b}{2^k} \iff$$

$$\iff b\frac{2^k+1}{2^k} < x_L + 1 \text{ and } x_L \leq b\frac{2^k+1}{2^k} \iff \frac{2^k}{2^k+1}x_L \leq b < \frac{2^k}{2^k+1}(x_L+1).$$

To guarantee a solution to (1), we need to prove the existence of an integer greater than

$$\frac{2^k x_L}{2^k+1},$$

but strictly less than

$$\frac{2^k x_L + 2^k}{2^k+1}.$$

We claim that there is such an integer if and only if $x_L \not\equiv -1 \mod (2^k+1)$. Lets look at $2^k x_L$ in the following form:

$$2^k x_L = p_k(2^k+1) + t_k,$$

where $0 \leq t_k \leq 2^k$, $t_k \in \mathbb{Z}$. We need an integer

$$b' \in \left[\frac{2^k x_L}{2^k+1}, \frac{2^k x_L + 2^k}{2^k+1}\right) = \left[p_k + \frac{t_k}{2^k+1}, p_k + \frac{t_k + 2^k}{2^k+1}\right).$$

This is equivalent to the existence of an integer

$$b'' \in \left[\frac{t_k}{2^k+1}, \frac{t_k+2^k}{2^k+1}\right).$$

If $t_k = 0$ then $0 \in \left[\frac{t_k}{2^k+1}, \frac{t_k+2^k}{2^k+1}\right)$, and if $t_k > 1$ then $1 \in \left[\frac{t_k}{2^k+1}, \frac{t_k+2^k}{2^k+1}\right)$. It follows that there is such an integer if and only if

$$t_k \neq 1 \iff 2^k x_L \not\equiv 1 \mod (2^k+1) \iff x_L \not\equiv -1 \mod (2^k+1).$$

□

Our purpose now is to show the existence of many k, where this incongruence holds. The following lemma will ensure this.

Lemma 2.2. *Let k be an integer such that $k \geq \frac{1}{2}\log_2 x_L$. In this case, there is an integer solution for* (1) *either for k or for $k+1$.*

Proof. For contradiction, suppose that both of the following linear congruences hold:

$$x_L \equiv -1 \mod (2^k+1) \text{ and } x_L \equiv -1 \mod (2^{k+1}+1).$$

On the other hand, 2^k+1 and $2^{k+1}+1$ are relative primes, so by the Chinese remainder theorem any solution to the above congruent system has a unique solution under $\mod (2^k+1)(2^{k+1}+1)$. Hence, the smallest positive integer that satisfies both of the above congruences simultaneously is $(2^k+1)(2^{k+1}+1)-1$. Thus $x_L \geq (2^k+1)(2^{k+1}+1)-1$. From this, a trivial estimate shows that $x_L > 2^{2k}$, contradicting the assumptions of the theorem.

□

3. Result on security against an eavesdropper

Now, we want to underestimate the number of b's which Eve considers viable. From the previous lemma, we know that Eve can decompose x_L in many ways to satisfy (1). But, to delude Eve it is not enough to just have b's which satisfy (1). Remember that Eve wants b's which can be the starting sequence of the codeword and she can check if this hypothesis of her is compatible with the queries or not. What do we mean by this?

Fix a $b \in \mathbb{N}$ such that for some k: $x_L = b + \lfloor \frac{b}{2^k} \rfloor$. Considering Eve's point of view, she thinks that she queried from a set $A_\zeta \dotplus A_\zeta$, where $\zeta \in [1,2)$ and she hopes that $\zeta = \eta$. To have this theory, she has to assume that $a_j = \lfloor 2^j \zeta \rfloor = b$, where $j = \lfloor \log_2 b \rfloor$. In this case, the first elements of the set A_ζ would be:

$$1 = \left\lfloor \frac{b}{2^{\lfloor \log_2 b \rfloor + 1}} \right\rfloor, \cdots \left\lfloor \frac{b}{2} \right\rfloor, b.$$

Denote by C_b the set that has exactly the same elements as above:

$$C_b := \left\{ \left\lfloor \frac{b}{2^{\lfloor \log_2 b \rfloor + 1}} \right\rfloor, \cdots \left\lfloor \frac{b}{2} \right\rfloor, b \right\}.$$

Eve can exclude certain b's by looking up if during her query sequence she queried for some $x \in A_\zeta \dotplus A_\zeta$, $x < x_L$. She can do it because if she did indeed query such an x, then she would have got $f(x) = 1$. But, in this case she would have stopped querying there and would have got x as an element of S instead of x_L.

Before we go further, we prove a simple statement that ensures that Eve actually knows $(A_\zeta \dotplus A_\zeta) \cap [1, x_L]$. Or in other words, the exclusion method relies only on b and not on the elements of A_ζ, larger than b (which are unknown to Eve).

Claim. $(A_\zeta \dotplus A_\zeta) \cap [1, x_L] = (C_b \dotplus C_b) \cap [1, x_L]$.

Proof. First, we recall that we assume $b = \lfloor 2^j \zeta \rfloor = a_j$. It is obvious that $(A_\zeta \dotplus A_\zeta) \cap [1, x_L] \supseteq (C_b \dotplus C_b) \cap [1, x_L]$. For the other direction, take an arbitrary $x \in (A_\zeta \dotplus A_\zeta) \cap [1, x_L]$. Notice that it is enough to prove that $x = a_l + a_m$ holds for some $l, m \leq j, l \neq m$. We know by assumption that $x_L = a_j + a_k$ for some $k < j$. Note that for every $n \in \mathbb{N}$: $2a_n = 2\lfloor 2^n \zeta \rfloor \leq \lfloor 2^{n+1} \zeta \rfloor = a_{n+1}$. Assume for contradiction that $x = a_l + a_m$ and $l > j$, in this case we would get $x_L = a_j + a_k < 2a_j \leq a_{j+1} \leq a_l < a_l + a_m = x$. □

We are going to say that a b is *acceptable* for Eve if it is a solution of (1) for some k and it can not be excluded by Eve with the previous method (checking if she queried for any $x < x_L, x \in A_\zeta \dotplus A_\zeta$, which is equivalent, by the above claim, to checking if she queried for any $x < x_L, x \in C_b \dotplus C_b$). We want to emphasize here, that acceptability now relies only on b (and not on the elements of A_ζ which are larger than b). Lets denote the number of acceptable solution for Eve by μ (for a fixed x_L).

We want Eve to have a lot of acceptable b's so we try to deny the ability of exclusion from her. For this purpose, from now on we restrict the length of the query sequence by N^β. How does this affect Eve's exclusion method? First notice that in order to exclude a certain b it is enough for her to check if she queried $\hat{x} = \max\{x \in C_b \dotplus C_b \mid x < x_L\}$, the largest element of $C_b \dotplus C_b$ smaller than x_L. So if we can ensure that $N^\beta \leq x_L - \hat{x}$, then she can not exclude this b. We will use this observation in the next theorem.

As we mentioned earlier, we are proving that Eve can not figure out the codeword uniquely. We need to specify what that means exactly. So far we have investigated only solutions for (1). If Eve finds a k and a corresponding b as a solution of (1), then she assumes $b = \lfloor 2^s \zeta \rfloor$ for some $s \in \mathbb{N},\ 1 \le \zeta < 2$ and hopes that $\zeta = \eta$. The first $n+1$ digits of ζ is equivalent to an n digit codeword ($1.\zeta_1\zeta_2\cdots\zeta_n\cdots \sim d = \zeta_1\zeta_2\cdots\zeta_n$), so she has to get these $n+1$ digits. Since the length of b is at most the length of x_L, Eve only knows at most $\lfloor \log_2 x_L \rfloor + 1$ digits of ζ. So she can choose the values of the remaining $h = n+1-(\lfloor \log_2 x_L \rfloor + 1) = n - \lfloor \log_2 x_L \rfloor$ digits arbitrarily, leaving her with at least 2^h different ζ, which have the property, that their first $\lfloor \log_2 b \rfloor + 1$ digits form the number b. And this is only for one acceptable b. (We know that at least one always exists, which is the one corresponding to η, which is determined by the real codeword.)

So if x_L is too small, then this leaves Eve with a lot of unknown information about the codeword. The second part of the next theorem is about the case when she finds an x_L where $n - \lfloor \log_2 x_L \rfloor$ is small (i.e. x_L is large). In this case there are many acceptable ways to decompose x_L. So she has many b's to start the codeword with, leaving her in inconsistency, even though she has low level of freedom to extend them into possible codewords.

We are going to call an integer $d = \zeta_1\zeta_2\cdots\zeta_n$ (where each $\zeta_i \in \{0,1\}$ for $i = 1, 2, \cdots, n$) possible codeword *viable* if there is an acceptable b which is the prefix of d in binary form. For a fixed x_L denote the number of viable codewords by ν. Obviously $\mu \le \nu$, since every acceptable b can be extended (possibly in many ways) into a viable d.

In the upcoming theorem we will consider two cases, $x_L \ge N^\alpha$ or $x_L \le N^\alpha$, where $0 < \alpha < 1$, and we will give a lower bound on μ and ν respectively. We will give some more restriction on α later and we will also discuss its relationship to other variables N, γ, β.

Theorem 3.1. *(i). If $x_L < N^\alpha$ then*

$$\nu \ge \frac{\gamma N^{1-\alpha}}{2}.$$

(ii). If $x_L \ge N^\alpha$ then

$$\mu \ge \frac{\left(\alpha - \beta - \frac{2}{\log_2 N} - \frac{1}{2}\right)\alpha \log_2 N - 3}{2}.$$

Proof. (i). In this case, even if Eve finds the correct b for x_L she misses some bits of the codeword. To be more precise, if the length of the codeword is n we have the following inequalities:

$$\text{missing bits} = n - \lfloor \log_2 x_L \rfloor \ge \lfloor \log_2 \gamma N \rfloor - \lfloor \log_2 x_L \rfloor \ge \log_2 \gamma - 1 + \log_2 N - \log_2 x_L \ge \log_2 \gamma + (1-\alpha)\log_2 N - 1.$$

Thus,

$$\mu \ge 2^{\text{missing bits}} \ge \frac{\gamma N^{1-\alpha}}{2}.$$

(ii). We are going to prove the existence of acceptable b's belonging to certain values of $k \in [\frac{1}{2}\log_2 x_L, (\alpha - \beta - \frac{2}{\log_2 N})\log_2 x_L]$. From Lemma 2.2, we know that if $k \ge \frac{1}{2}\log_2 x_L$, then for either k and $k+1$ there is a b solution for (1). But, it also needs to be acceptable. Because of the argument above we only have to ensure $N^\beta \le x_L - \hat{x}$ for this.

Since $x_L = b + \lfloor \frac{b}{2^k} \rfloor$, $\hat{x}$ is either $b + \left\lfloor \frac{b}{2^{k+1}} \right\rfloor$ or $\left\lfloor \frac{b}{2} \right\rfloor + \left\lfloor \frac{b}{4} \right\rfloor$ and therefore

$$\left\lfloor \frac{b}{2^k} \right\rfloor - \left\lfloor \frac{b}{2^{k+1}} \right\rfloor \le \min\left(\left\lfloor \frac{b}{2^k} \right\rfloor - \left\lfloor \frac{b}{2^{k+1}} \right\rfloor, b + 1 - \left\lfloor \frac{b}{2} \right\rfloor - \left\lfloor \frac{b}{4} \right\rfloor \right) \le x_L - \hat{x}.$$

So, we are left to see if the following inequality holds:

$$\left\lfloor \frac{b}{2^k} \right\rfloor - \left\lfloor \frac{b}{2^{k+1}} \right\rfloor \ge N^\beta.$$

Starting from the left side we get:

$$\left\lfloor \frac{b}{2^k} \right\rfloor - \left\lfloor \frac{b}{2^{k+1}} \right\rfloor \ge \frac{b}{2^k} - 1 - \frac{b}{2^{k+1}} \ge \frac{x_L}{2^{k+2}} \ge \frac{N^\alpha}{2^{k+2}}.$$

So, we need $\frac{N^\alpha}{2^{k+2}} \ge N^\beta$. Since we are now considering $k \le (\alpha - \beta - \frac{2}{\log_2 N})\log_2 x_L$, we get the following:

$$k \le (\alpha - \beta - \frac{2}{\log_2 N})\log_2 x_L \le (\alpha - \beta - \frac{2}{\log_2 N})\log_2 N = \log_2 N(\alpha - \beta) - 2.$$

If we look at the left and right side and reorder the inequality we can easily get the desired inequality. Now, from this and from Lemma 2.2 we can say that for every k in the interval $[\frac{1}{2}\log_2 x_L, (\alpha - \beta - \frac{2}{\log_2 N})\log_2 x_L]$ either k or $k+1$ gives an acceptable b for Eve. So, by pairing the integers of interval we get

$$\mu \ge \left\lfloor \frac{\left\lfloor (\alpha - \beta - \frac{2}{\log_2 N})\log_2 x_L \right\rfloor - \left\lceil \frac{1}{2}\log_2 x_L \right\rceil + 1}{2} \right\rfloor \ge \left\lfloor \frac{(\alpha - \beta - \frac{2}{\log_2 N})\log_2 x_L - \frac{1}{2}\log_2 x_L - 1}{2} \right\rfloor \ge$$

$$\ge \frac{(\alpha - \beta - \frac{2}{\log_2 N} - \frac{1}{2})\log_2 x_L - 3}{2} \ge \frac{(\alpha - \beta - \frac{2}{\log_2 N} - \frac{1}{2})\alpha \log_2 N - 3}{2}.$$

□

4. Concluding remarks

It is worth to take a look at what the results of Theorem 3.1 really mean. In both parts of this theorem, we get more acceptable solutions if N is large. We can guarantee this part, but after a certain point an absurdly large N is not very practical. So, we want to optimize the parameters α and β as well.

The parameter β is best to be as small as possible (but can not be too small because Bob has to be able to find the codeword). Theorem 1.2 tells us that the lower bound $\log_2 N + 1 < N^\beta$ is required to give Bob the ability to get the codeword with the secret key γ, regardless of the restriction on the length of a query sequence. So, we can choose for example $\beta = \log_N(2\log N)$, which tends to 0, as $N \to \infty$.

In the case of α, it is a bit more complicated than that β. We can say that α is a double-edged weapon in the following sense: According to part (i), α is better to be small, while according to part (ii) it is better to be large (close to 1).

We gather here the acquired bounds on N with respect to the parameters. Here, we consider the case when we only want to ensure more than one viable codeword. From Theorem 3.1, we get that if we want $\nu > 1$ in part (i) and $\mu > 1$ in part (ii), then N has to satisfy the following lower bounds: (i) $\to N > (\frac{2}{\gamma})^{\frac{1}{1-\alpha}}$ and (ii) $\to N > 2^{\frac{7}{\alpha(\alpha-\beta-1/2)}}$.

Acknowledgment

The authors are supported by the European Union, co-financed by the European Social Fund (EFOP-3.6.3-VEKOP-16-2017-00002).

References

[1] B. Bakos, N. Hegyvári, M. Pálfy, X. H. Yan, On subset sums of pseudo-recursive sequences, *Discrete Math. Lett.* **4** (2020) 31–36.
[2] N. Hegyvári, Some remarks on a problem of Erdős and Graham, *Acta Math. Hungar.* **53** (1989) 149–154.
[3] N. Hegyvári, Subset sums in $\mathbb{N}^2$, *Combin. Probab. Comput.* **5** (1996) 393–402.
[4] P. Erdős, R. L. Graham, Old and new problems and results in combinatorial number theory: van der Waerden's theorem and related topics, *Enseign. Math.* **25** (1979) 325–344.