Opowieść o Bogumile, który zły klucz wyrobił

Jak wiemy, każdą nieujemną liczbę całkowitą można *jednoznacznie* zapisać w systemie pozycyjnym o podstawie 2 (systemie dwójkowym). Oznacza to, że dla dowolnej liczby naturalnej n oraz dowolnej nieujemnej liczby całkowitej m mniejszej niż $2^n - 1$ istnieje *dokładnie jeden* ciąg $(b_0, \ldots, b_{n-1})$ liczb ze zbioru $\{0, 1\}$, dla którego

$$m = b_0 2^0 + b_1 2^1 + \ldots + b_{n-1} 2^{n-1}.$$

Jeżeli na przykład $n = 5$ i $m = 22$, to mamy

$$22 = 0 \cdot 2^0 + 1 \cdot 2^1 + 1 \cdot 2^2 + 0 \cdot 2^3 + 1 \cdot 2^4,$$

co odpowiada ciągowi $(0, 1, 1, 0, 1)$. Zwróćmy uwagę, że kolejność cyfr dwójkowych w tym ciągu jest zapisana w odwrotnej kolejności niż w przypadku standardowego zapisu pozycyjnego $22 = (10110)_2$.

Co się stanie, jeśli zamiast kolejnych potęg liczby 2 weźmiemy coś innego? Niech na przykład

$$u_0 = 1, \quad u_1 = 3, \quad u_2 = 7, \quad u_3 = 10.$$

Jakie nieujemne liczby całkowite można przestawić w postaci

$$b_0 u_0 + b_1 u_1 + b_2 u_2 + b_3 u_3, \tag{$\triangle$}$$

przy czym współczynniki b_i są nadal wybierane ze zbioru $\{0, 1\}$? Oczywiście nie uda nam się uzyskać liczby większej niż 21, która odpowiada ciągowi $(b_0, b_1, b_2, b_3) = (1, 1, 1, 1)$, ale podobnie będzie z liczbą 2 — nie istnieje układ zer i jedynek, dla którego wyrażenie ($\triangle$) przyjmie wartość 2. Wynika to oczywiście z tego, że wartość u_1 jest za duża. Co jednak bardziej niepokojące, tej samej liczbie mogą odpowiadać różne ciągi: wartość ($\triangle$) dla ciągu $(0, 1, 1, 0)$ jest taka sama (równa 10), jak dla ciągu $(0, 0, 0, 1)$. Czy można to jakoś naprawić? Tak — wystarczy za u_3 przyjąć, na przykład, 12. Jak się łatwo w tej sytuacji przekonać, każdemu układowi (b_0, b_1, b_2, b_3) odpowiada inna liczba. Oczywiście mamy jakieś „dziury" (liczby bez reprezentacji), ale nie będziemy się tym przejmować, bo w głównej mierze zależy nam na *jednoznaczności*.

Ciągi gwarantujące jednoznaczne przedstawienie

Rozważmy teraz sytuację ogólniejszą. Niech $(u_0, \ldots, u_{n-1})$ będzie rosnącym ciągiem liczb naturalnych. Jak się przekonaliśmy wcześniej, dla dwóch różnych ciągów zero-jedynkowych $(b_0, \ldots, b_{n-1})$ oraz $(b'_0, \ldots, b'_{n-1})$ może zachodzić równość

$$b_0 u_0 + \ldots + b_{n-1} u_{n-1} = b'_0 u_0 + \ldots + b'_{n-1} u_{n-1}. \tag{$\diamond$}$$

Jak się jednak okazuje, jeżeli tylko ciąg $(u_0, \ldots, u_{n-1})$ rośnie wystarczająco szybko, to do takiej sytuacji nie może dojść.

Fakt. *Jeżeli dla każdego $k \in \{1, \ldots, n-2\}$ zachodzi nierówność*

$$u_0 + \ldots + u_k < u_{k+1}, \tag{$\square$}$$

to dla dowolnego ciągu zero-jedynkowego $(b_0, \ldots, b_{n-1})$ wyrażenie

$$b_0 u_0 + \ldots + b_{n-1} u_{n-1}$$

przyjmuje inną wartość.

Fakt ten możemy również sformułować następująco: jeżeli ciąg $(u_0, \ldots, u_{n-1})$ spełnia warunek $(\square)$, to z równości $(\diamond)$ wynika, że $(b_0, \ldots, b_{n-1}) = (b'_0, \ldots, n'_{n-1})$.

Opis kryptosystemu

Alicja chce wysłać do Bogumiła wiadomość. Wiedzą, że Cecylia będzie podsłuchiwała ich komunikację, więc umawiają się, że wykorzystają następującą metodę szyfrowania. Bogumił wybiera, znany tylko sobie, ciąg liczb naturalnych

$$(u_0, \ldots, u_{n-1})$$

spełniający warunek $(\square)$. Ponadto, ciągle w tajemnicy przed innymi, ustala liczbę całkowitą N, która jest większa niż suma wszystkich elementów jego ciągu, to znaczy

$$u_0 + \ldots + u_{n-1} < N.$$

Następnie wybiera liczbę naturalną Q mniejszą niż N i względnie pierwszą z N (czyli $\mathrm{NWD}(Q, N) = 1$). Tak przygotowany, wyznacza on ciąg $(a_0, \ldots, a_{n-1})$, którego elementy dane są wzorem

$$a_i = u_i Q \bmod N. \qquad (\heartsuit)$$

Kluczem prywatnym Bogumiła jest

$$\text{ciąg } (u_0, \ldots, u_{n-1}) \qquad \text{oraz} \qquad \text{liczby } N \text{ i } Q.$$

Te wielkości są znane tylko jemu. Kluczem publicznym Bogumiła, znanym wszystkim, jest

$$\text{ciąg } (a_0, \ldots, a_{n-1}).$$

Alicja, chcąc zaszyfrować wiadomość $m = (b_0, \ldots, b_{n-1})$ składającą się z n bitów, wylicza sumę

$$\hat{m} = b_0 a_0 + \ldots + b_{n-1} a_{n-1} \qquad (\#)$$

i przesyła ją do Bogumiła. Bogumił, który odebrał $\hat{m}$, wyznacza

$$M = (\hat{m} \cdot Q^{-1}) \bmod N. \qquad (\lightning)$$

Liczba Q^{-1} jest tutaj odwrotnością Q modulo N, czyli najmniejszą dodatnią liczbą całkowitą k spełniającą kongruencję $kQ \equiv 1 \pmod N$. Bogumił zna Q^{-1} (może tę wartość wyliczyć), ponieważ zna Q i N, które są liczbami względnie pierwszymi. Zauważmy, że

$$M = \left[(b_0 a_0 + \cdots + b_{n-1} a_{n-1}) Q^{-1}\right] \bmod N \equiv b_0 a_0 Q^{-1} + \ldots + b_{n-1} a_{n-1} Q^{-1} \pmod N.$$

Ponieważ, z definicji a_i, mamy $a_i Q^{-1} \equiv u_i \pmod N$, to

$$M \equiv b_0 u_0 + \ldots + b_{n-1} u_{n-1} \pmod N.$$

Ponieważ $M < N$ (M jest resztą z dzielenia przez N) oraz

$$b_0 u_0 + \ldots + b_{n-1} u_{n-1} < u_0 + \ldots + u_{n-1} < N,$$

to $M = b_0 u_0 + \ldots + b_{n-1} u_{n-1}$. Na mocy faktu sformułowanego wcześniej, istnieje dokładnie jeden ciąg zero-jedynkowy $(b_0, \ldots, b_{n-1})$, dla którego taka równość zachodzi. Bogumił, znając wyrazy ciągu $(u_0, \ldots, u_{n-1})$, może łatwo (jak?) wyznaczyć odpowiednie wartości $(b_0, \ldots, b_{n-1})$, co jest równoważne z odczytaniem m.

Przykład. Bogumił wybrał ciąg

$$u_0 = 3, \qquad u_1 = 5, \qquad u_2 = 10, \qquad u_3 = 20, \qquad u_4 = 45$$

oraz przyjął

$$N = 115, \qquad Q = 34.$$

Bogumił wykorzystał algorytm Euklidesa i wyznaczył Q^{-1}, które w tym przypadku jest równe 44. Ponadto, ze wzoru $a_i = u_i Q \bmod N$ obliczył

$$a_0 = 102, \qquad a_1 = 55, \qquad a_2 = 110, \qquad a_3 = 105, \qquad a_4 = 35.$$

Alicja, chcąc zaszyfrować wiadomość $m = (0, 1, 0, 1, 1)$ oblicza

$$\hat{m} = a_1 + a_3 + a_4 = 195$$

i przesyła tę liczbę do Bogumiła. Teraz Bogumił odtwarza wiadomość, obliczając

$$M = \hat{m} \cdot Q^{-1} \bmod N = (195 \cdot 44) \bmod 115 = 70,$$

i rozwiązując, ze względu na zero-jedynkowe ciągi $(b_0, \ldots, b_{n-1})$, równanie $M = b_0 u_0 + \ldots + b_4 u_4$, które przyjmuje postać

$$70 = b_0 \cdot 3 + b_1 \cdot 5 + b_2 \cdot 10 + b_3 \cdot 20 + b_4 \cdot 45.$$

Ponieważ ciąg $(u_0, \ldots, u_4)$ spełnia założenie ($\square$), to Bogumił ma pewność, że istnieje tylko jeden ciąg $(b_0, \ldots, b_4)$ spełniający ostatnią równość. Ponieważ

$$70 = 0 \cdot 3 + 1 \cdot 5 + 0 \cdot 10 + 1 \cdot 20 + 1 \cdot 45,$$

to odczytana wiadomość ma postać $(0, 1, 0, 1, 1)$, zgodnie z zamierzeniami Alicji.

Zadanie

Napisać program, który dla ustalonego klucza publicznego Bogumiła, czyli ciągu

$$(a_0, \ldots, a_{n-1})$$

skonstruowanego według wzoru ($\heartsuit$) (na podstawie nieznanych nam $(b_0, \ldots, b_{n-1})$, N i Q), oraz zaszyfrowanej przez Alicję wiadomości $\hat{m}$, pomoże Cecylii odczytać wiadomość m. Wiadomość m może składać się z kilku części, ale każdą z nich Alicja zaszyfruje przy pomocy tego samego klucza publicznego — zobacz opis poniżej.

W powyższej postaci, bez dodatkowych informacji, zadanie to jest stosunkowo trudne dla dużych n. Cecylia wie jednak, że Bogumił nie był wystarczająco ostrożny przy konstrukcji klucza prywatnego i wybrał małe liczby u_0 i u_1. Dokładniej, zachodzą nierówności

$$u_0 < u_1 \leqslant U,$$

gdzie U jest pewną z góry ustaloną liczbą (będzie podana w pliku wejściowym). Jeżeli wartość U nie jest duża, to w znacznym stopniu zmniejsza bezpieczeństwo systemu.

Wejście

Dane do zadania są przygotowane w jednym pliku tekstowym, który ma następującą postać:

- W pierwszym wierszu znajduje się jedna liczba naturalna n, będąca długością klucza publicznego. Możemy założyć, że n jest wielokrotnością liczby 8.
- W drugim wierszu podana jest wartość ograniczenia U ($u_1 \leqslant U$).

- W kolejnych n wierszach zapisane są (po jednej w wierszu) liczby a_i.
- Następnie, w wierszu o numerze $n+3$, znajduje się jedna liczba naturalna k równa liczbie części, na które Alicja podzieliła swoją wiadomość.
- W każdym z kolejnych k wierszy zapisana jest zaszyfrowana k-ta część wiadomości.

Zakładamy dodatkowo, że

$$N < 2^{70}.$$

Oryginalna (niezaszyfrowana) wiadomość Alicji jest $k \cdot n/8$-literowym tekstem, którego każdy znak Alicja zamieniła na liczbę całkowitą od 0 do 255 odpowiadającą jego kodowi ASCII. Ponieważ każda taka liczba ma 8-bitową reprezentację dwójkową, to wiadomość Alicji składa się k ciągów o n bitach i każdy taki ciąg n bitów odpowiada $n/8$ znakom.

Wyjście

W jedynym wierszu wyjścia powinna znaleźć się odszyfrowana wiadomość.

Przykład 1

Plik wejściowy:

```
8       // długość klucza n
20      // ograniczenie górne u_1
275     // a_0
9       // a_1
18      // a_2
169     // a_3
214     // a_4
20      // a_5
31      // a_6
337     // a_7
1       // liczba części k
346     // zaszyfrowana wiadomość m
```

Oczekiwany wynik:

```
A
```

Wyjaśnienie: klucz publiczny $(a_0, \ldots, a_7)$ został wygenerowany z klucza prywatnego

$$(u_0, \ldots, u_7) = (1, 3, 6, 11, 26, 52, 101, 203), \qquad N = 408, \qquad Q = 275.$$

Ponieważ ciąg $(u_0, \ldots, u_7)$ ma długość 8, więc za jego pomocą można jednorazowo zaszyfrować 8 bitów. Alicja zaszyfrowała $k = 1$ wiadomości, więc jest to dokładnie 8 bitów $(b_0, \ldots, b_7)$ (jeden znak w kodzie ASCII), które po zaszyfrowaniu zgodnie ze wzorem (#) jest liczbą 346. Odwrotnością $Q = 275$ modulo $N = 408$ jest $Q^{-1} = 227$, to (zob. (ϟ))

$$M = 346 \cdot 227 \bmod 408 = 206.$$

Ponieważ

$$206 = 0 \cdot 1 + 1 \cdot 3 + 0 \cdot 6 + 0 \cdot 11 + 0 \cdot 26 + 0 \cdot 52 + 0 \cdot 101 + 1 \cdot 203,$$

to wiadomością Alicji był ciąg bitów

$$(0, 1, 0, 0, 0, 0, 0, 1).$$

Liczba $(01000001)_2$ jest równa 65, a literą odpowiadającą liczbie 65 w kodzie ASCII jest `A`, więc oryginalna wiadomość brzmi właśnie `A`.

Przykład 2

Plik wejściowy:

```
16       // n
20       // u_1 <= 20 = U
120833   // a_0
241638
112
241792
242002
121617
243248
3304
248190
13216
26418
52850
105700
332191
60385
241589   // a_15
1        // k
980498   // zaszyfrowana wiadomość
```

Oczekiwany wynik:

```
md
```

Wyjaśnienie: klucz publiczny $(a_0, \ldots, a_{15})$ został wygenerowany z klucza prywatnego

$$(7, 8, 24, 41, 86, 175, 353, 708, 1412, 2832, 5661, 11325, 22650, 45298, 90597, 181198)$$

dla $N = 362401$ oraz $Q = 120805$. Ponieważ $n = 16$, to jednorazowo można zaszyfrować 16 bitów, czyli dwa 8-bitowe znaki ASCII. Jednocześnie $k = 1$, więc cała wiadomość składa się z dwóch znaków. Po zaszyfrowaniu ma ona postać $\hat{m} = 980498$. Ponieważ $Q^{-1} = 25886$, to $M = 54792$. Rozwiązaniem równania

$$54792 = b_0 u_0 + \ldots + b_{15} u_{15}$$

jest ciąg `01101101 01100100`, który odpowiada literom `md`.

Przykład 3

Plik wejściowy:

```
16      // n
20      // U
157576  // a_0
157581
25
78833
157666
78978
79168
79553
1540
160646
163726
12310
103403
206811
177263
39394   // a_15
2       // k
885969  // pierwsza część wiadomości
938653  // druga część wiadomości
```

Oczekiwany wynik:

```
Test
```

Wyjaśnienie: klucz publiczny $(a_0, \ldots, a_{15})$ został wygenerowany z klucza prywatnego

$$(2, 5, 15, 28, 56, 115, 229, 460, 924, 1844, 3692, 7386, 14770, 29543, 59086, 118180)$$

dla $N = 236359$ oraz $Q = 78788$. Ponieważ $n = 16$, to jednorazowo można zaszyfrować 16 bitów, czyli dwa 8-bitowe znaki ASCII. Jednocześnie $k = 2$, więc cała wiadomość składa się z czterech znaków, które Alicja podzieliła na dwie części po dwa znaki. Każdą z tych części zaszyfrowała oddzielnie. Po zaszyfrowaniu pierwsza część jest liczbą 885969, a druga liczbą 980498. Ciąg bitów (po odszyfrowaniu) odpowiadający pierwszej części to `01010100 01100101` (znaki `Te`), a drugiej to `01110011 01110100` (znaki `st`).