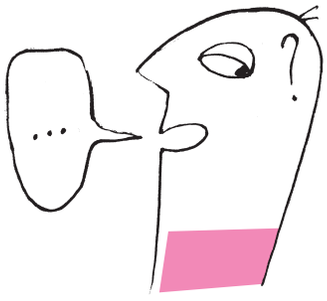
A jednak się da (VII),

czyli saga kryptologiczna w odcinkach. Tym razem: nie wiem, ale powiem.

Łukasz RAJKOWSKI

Rozwiązanie zadania M 1604.
Rozważmy usadzenie przy okrągłym stole spełniające opisany w zadaniu warunek. Kolejne pary uczestników siedzących obok siebie przy stole połączmy na zmianę odcinkami czerwonymi i niebieskimi. Ponadto każdego połączmy zielonym odcinkiem z tym znajomym, obok którego nie siedzi. Wystarczy teraz zauważyć, że każdy z trzech użytych kolorów wyznacza inny sposób zakwaterowania.



Wyobraźmy sobie następującą wakacyjną historię miłosną, która miała prawo się zdarzyć przed nastaniem ery wszechobecnych mediów społecznościowych.

Bogumił ma problem sercowy. Na dyskotece z okazji ostatniego dnia kolonii poznał wspaniałą dziewczynę, Aldonę. Niestety, pechowym zbiegiem okoliczności, rodzice zabrali Aldonę do domu zanim Bogumił zdążył uzyskać od niej numer telefonu. Biedny Bogumił życia bez Aldony sobie nie wyobraża, więc kombinuje jak może, aby ów numer posiąść i bez końca do niej wydzwaniać z wyznaniami miłości. Szczęśliwie, Bogumił ma kolegów, którzy w sprawach sercowych są dużo bardziej biegli od niego. Podobno Dobromir podczas kolonii zapisał w zeszycie numery telefonów wszystkich dziewczyn z wyjazdu. Sęk w tym, że Dobromir to straszna papla, a Bogumiłowi nie uśmiecha się, by wszyscy jego koledzy z podwórka wiedzieli, co czuje do Aldony, bo głąby i tak tego nie zrozumieją – Dobromir nie może zatem dowiedzieć się, że chodzi o Aldonę. Na domiar złego, Dobromir wyjechał do Szwajcarii, choć szczęśliwie akurat on zostawił Bogumiłowi swój numer telefonu (jednak połączenie sporo kosztuje). Co powinien zrobić Bogumił, aby odzyskać kontakt z Aldoną i zapłacić jak najmniejszy rachunek za połączenie z Dobromirem?

To, czego potrzeba Bogumiłowi do szczęścia, to protokół *prywatnego uzyskiwania informacji* (jest to próba tłumaczenia angielskiego terminu *private information retrieval*), czyli sposób na odpytywanie bazy danych (Dobromira) tak, aby baza danych nie wiedziała, o co ją pytamy, i mimo to dostarczyła odpowiedzi (tytułowe *nie wiem, ale powiem*). Oczywiście Bogumił mógłby poprosić o przesłanie *całej* bazy danych (czyli o przedyktowanie zawartości całego zeszytu), ale zależy mu na tym, by zminimalizować ilość przesyłanej informacji. Zaprawieni w lekturze kącika AJSD Czytelnicy nie powinni mieć trudności z przetłumaczeniem tego problemu na bardziej matematyczny język: Dobromir zna ciąg bitów $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. Bogumił chce poznać wartość x_i dla wybranego przez siebie $i \leqslant n$, jednak nie chce zdradzić Dobromirowi wartości i. Zakładamy, że każda informacja wysłana do i od Dobromira kosztuje, a Bogumił chce uczynić ten koszt jak najmniejszym. Wydaje się, że nie może on zrobić niczego istotnie lepszego od poproszenia Dobromira o cały ciąg $\boldsymbol{x}$... Cóż, a jednak się da!

Czytelnik może słusznie skojarzyć przedstawiony tu problem z protokołem transferu utajnionego, opisanym w Δ^2_{19}. Tutaj jednak nie wymagamy, aby jedyną informacją uzyskaną przez Bogumiła była wartość x_i.

W całym artykule: sumowanie należy rozumieć „modulo 2”, czyli np. $1+1+0+1=1$.

Kolegów dwóch (lub więcej)

Na początku uprośćmy sytuację i załóżmy, że Bogumił ma jeszcze jednego kolegę, Eustachego, który zna $\boldsymbol{x}$, ponadto nie zna się on z Dobromirem. Wówczas Bogumił mógłby zastosować następującą egzotyczną (i, jak się zaraz okaże, niezbyt mądrą) procedurę: losuje ciąg binarny $\boldsymbol{a} = (a_1, \ldots, a_n)$ i prosi Dobromira o przesłanie $x_{\boldsymbol{a}} := \sum_{k=1}^{n} x_k a_k$, a Eustachego o $x_{\boldsymbol{a}'}$, gdzie $\boldsymbol{a}' = (a'_1, \ldots, a'_n)$ to ciąg $\boldsymbol{a}$, w którym wartość i-tego bitu została zmieniona (tzn. $a_i + a'_i = 1$ oraz $a_k = a'_k$ dla $k \neq i$). Okazuje się, że wówczas $x_{\boldsymbol{a}} + x_{\boldsymbol{a}'} = x_i$, gdyż

$$x_{\boldsymbol{a}} + x_{\boldsymbol{a}'} = \sum_{k=1}^{n} x_k(a_k + a'_k) = \sum_{k \neq i} 2x_k a_k + x_i(a_i + a'_i) = x_i.$$

W tej sytuacji, znając $x_{\boldsymbol{a}}$ oraz $x_{\boldsymbol{a}'}$, Bogumił jest w stanie odtworzyć x_i. Z drugiej strony, ciąg $\boldsymbol{a}$ jest losowy, więc nie dostarcza Dobromirowi żadnej informacji o i. Ponadto z punktu widzenia Eustachego ciąg $\boldsymbol{a}'$ jest losowy (co wymaga być może krótkiej chwili zastanowienia), więc on też nie uzyskuje żadnej informacji o i. Zauważmy, że gdyby Dobromir i Eustachy połączyli siły, to mogliby łatwo odtworzyć i – jest to jedyny indeks, na którym różnią się ciągi $\boldsymbol{a}$ i $\boldsymbol{a}'$. Zakładamy jednak, że nie komunikują się oni ze sobą, zatem Bogumił nie musi się przejmować taką ewentualnością.

Przykład
Jeśli $\boldsymbol{x} = (1,1,0,1,1,1,1,0,1)$, $i = 3$
oraz $\boldsymbol{a} = (0,1,0,0,1,1,0,1,0)$,
to $\boldsymbol{a}' = (0,1,1,0,1,1,0,1,0)$,
i $x_{\boldsymbol{a}} = 1$, $x_{\boldsymbol{a}'} = 1$.
Koszt komunikacji to 20.

Jak już wspomnieliśmy, powyższy sposób – choć wykorzystuje pewne chytre obserwacje – jest dla Bogumiła bezwartościowy. Przypomnijmy bowiem, że przesył informacji w obie strony jest kosztowny, a w zaprezentowanym protokole Bogumił wysyła dwa ciągi długości n, czyli aż $2n$ bitów. Już (dwa razy) lepiej byłoby poprosić Dobromira o przesłanie całego ciągu $\boldsymbol{x}$!

Rzecz jasna, nie zaprezentowaliśmy powyższej idei tylko po to, by stwierdzić jej bezużyteczność. Okazuje się, że można ją tak zmodyfikować, żeby jednak coś na niej zyskać. W tym celu potrzebujemy kolejnych dwóch kolegów z dostępem do $\boldsymbol{x}$, którzy się ze sobą nie komunikują. Umieśćmy wartości $\boldsymbol{x}$ w kwadratowej tablicy $\boldsymbol{X} = [\overline{x}_{k,l}]_{k,l\leqslant m}$ rozmiaru $m \times m$, gdzie $m = \lceil\sqrt{n}\rceil$. Powiedzmy, że $x_i = \overline{x}_{\alpha,\beta}$. Tym razem Bogumił losuje dwa binarne ciągi $\boldsymbol{a}$ i $\boldsymbol{b}$ długości m i wysyła je Dobromirowi. Dobromir oblicza $\overline{x}_{\boldsymbol{a},\boldsymbol{b}} := \sum_{k,l\leqslant m} \overline{x}_{k,l} a_k b_l$. Podobnie jak poprzednio niech $\boldsymbol{a}'$ powstaje z $\boldsymbol{a}$ przez zmianę α-tego bitu, a $\boldsymbol{b}'$ powstaje z $\boldsymbol{b}$ przez zmianę β-tego bitu. Do kolejnych znajomych Bogumił wysyła pary ciągów bitów $(\boldsymbol{a},\boldsymbol{b})$, $(\boldsymbol{a}',\boldsymbol{b})$, $(\boldsymbol{a},\boldsymbol{b}')$ i $(\boldsymbol{a}',\boldsymbol{b}')$, otrzymując od nich $\overline{x}_{\boldsymbol{a},\boldsymbol{b}}$, $\overline{x}_{\boldsymbol{a}',\boldsymbol{b}}$, $\overline{x}_{\boldsymbol{a},\boldsymbol{b}'}$ oraz $\overline{x}_{\boldsymbol{a}',\boldsymbol{b}'}$. Wówczas w sumie $\overline{x}_{\boldsymbol{a},\boldsymbol{b}} + \overline{x}_{\boldsymbol{a}',\boldsymbol{b}} + \overline{x}_{\boldsymbol{a},\boldsymbol{b}'} + \overline{x}_{\boldsymbol{a}',\boldsymbol{b}'}$

- dla $k \neq \alpha, l \neq \beta$ składnik $\overline{x}_{k,l} a_k b_l$ liczony jest cztery razy (występuje w każdej składowej sumie), więc nie jest liczony wcale,
- dla $l \neq \beta$ składnik $\overline{x}_{\alpha,l} a_\alpha b_l$ liczony jest dwa razy, tak samo jak $\overline{x}_{\alpha,l} a'_\alpha b_l$; obu nie musimy zatem liczyć,
- dla $k \neq \alpha$ składniki $\overline{x}_{k,\beta} a_k b_\beta$ i $\overline{x}_{k,\beta} a_k b'_\beta$ liczone są dwa razy, więc podobnie jak w poprzednim punkcie nie wpływają na wynik,
- wśród liczb $a_\alpha b_\beta$, $a'_\alpha b_\beta$, $a_\alpha b'_\beta$, $a'_\alpha b'_\beta$ są trzy zera i jedynka, a zatem suma składników $\overline{x}_{\alpha,\beta} a_\alpha b_\beta$, $\overline{x}_{\alpha,\beta} a'_\alpha b_\beta$, $\overline{x}_{\alpha,\beta} a_\alpha b'_\beta$, $\overline{x}_{\alpha,\beta} a'_\alpha b'_\beta$ daje $\overline{x}_{\alpha,\beta}$.

Powyższe rozważania dowodzą, że $\overline{x}_{\boldsymbol{a},\boldsymbol{b}} + \overline{x}_{\boldsymbol{a}',\boldsymbol{b}} + \overline{x}_{\boldsymbol{a},\boldsymbol{b}'} + \overline{x}_{\boldsymbol{a}',\boldsymbol{b}'} = x_{\alpha,\beta}$. Bogumił uzyskuje zatem szukaną wartość $\overline{x}_{\alpha,\beta} = x_i$, a jego koledzy nie dowiadują się niczego o α, β (co uzasadniamy tak jak poprzednio). Tym razem wykorzystujemy cztery bazy danych, ale za to rozmiar naszej komunikacji to $4(2\sqrt{n}+1)$. W analogiczny sposób, poprzez zapisanie $\boldsymbol{x}$ w d-wymiarowej tablicy i wykorzystanie 2^d baz danych, możemy dokonać prywatnego uzyskania informacji przy komunikacji rozmiaru $2^d(d\sqrt[d]{n}+1)$. Widzimy zatem, że możemy w ten sposób uzyskać dowolnie mały wykładnik przy n, jednak kosztem wykładniczo rosnącej liczby potrzebnych, nieskomunikowanych baz danych.

Kontynuacja przykładu:

$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$, $(\alpha,\beta) = (1,3)$,
$\boldsymbol{a} = (0,0,1)$, $\boldsymbol{b} = (0,1,1)$, wówczas
$\boldsymbol{a}' = (1,0,1)$, $\boldsymbol{b}' = (0,1,0)$, oraz
$\overline{x}_{\boldsymbol{a},\boldsymbol{b}} = 1$, $\overline{x}_{\boldsymbol{a}',\boldsymbol{b}} = 0$, $\overline{x}_{\boldsymbol{a},\boldsymbol{b}'} = 0$, $\overline{x}_{\boldsymbol{a}',\boldsymbol{b}'} = 1$.

Co prawda, koszt komunikacji to 28 (czyli zwiększył się), ale już dla $n \geqslant 25$ byłby on mniejszy niż przy wykorzystaniu dwóch baz danych.

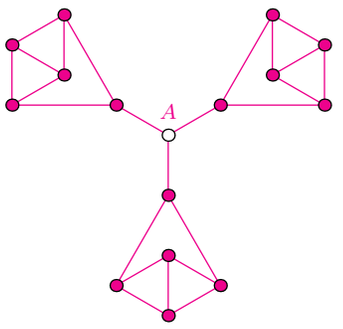
. . . a może jednak wystarczy jeden?

Co, jeśli nie jesteśmy w tak wygodnej sytuacji i Bogumił jest skazany wyłącznie na Dobromira? Czy możemy istotnie zmniejszyć koszt transmisji, mając do czynienia tylko z jedną bazą danych? Cóż, gdyby było inaczej, temat niekoniecznie zasługiwałby na prezentację w naszym kąciku. Okazuje się, że nawet mając do dyspozycji jeden egzemplarz bazy danych, możemy dokonać „prywatnego zapytania" kosztem rzędu rozmiaru bazy danych podniesionego do dowolnie małej potęgi. Po raz pierwszy taki magiczny sposób zaproponowali Eyal Kuszilewić i Rafail Ostrowski w 1997 roku. Aby dokonać jego prezentacji, musimy przedstawić małe przypomnienie z teorii liczb (patrz również artykuł o *commicie* w Δ_{18}^{11}).

Niech n będzie liczbą naturalną. *Kwadratem z dzielenia przez n* nazwiemy taką liczbę naturalną r, że istnieje x, dla którego $r = x^2 \pmod n$. Dla przykładu, 23 jest kwadratem z dzielenia przez 7, gdyż $23 = 3^2 \pmod 7$, ale 3 już nim nie jest. Oznaczmy przez $\mathcal{Q}_n$ zbiór kwadratów z dzielenia przez n, które są względnie pierwsze z n. Dla $n = pq$, gdzie p i q są różnymi liczbami pierwszymi, można nietrudno pokazać, że $\mathcal{Q}_n = \mathcal{Q}_p \cap \mathcal{Q}_q$. Niech $\mathcal{T}_n$ będzie zbiorem liczb względnie pierwszych z n, które nie należą ani do $\mathcal{Q}_p$ ani do $\mathcal{Q}_q$. Można pokazać, że $\mathcal{Q}_n$ i $\mathcal{T}_n$ „zachowują się" jak liczby dodatnie i ujemne, to znaczy: pomnożenie dwóch liczb z $\mathcal{Q}_n$ lub dwóch liczb $\mathcal{T}_n$ daje liczbę z $\mathcal{Q}_n$, a pomnożenie liczby z $\mathcal{Q}_n$ i liczby z $\mathcal{T}_n$ daje liczbę z $\mathcal{T}_n$. Ponadto, dla dużych wartości p, q stwierdzenie, czy liczba z $\mathcal{Q}_n \cup \mathcal{T}_n$ należy do $\mathcal{Q}_n$, jest trudne obliczeniowo, jeśli znamy tylko wartość n (a bardzo łatwe, jeśli znamy p i q).

Rozwiązanie zadania M 1605.
Odpowiedź: Nie wynika.

Na poniższym obrazku zilustrowana jest przykładowa sieć znajomości (uczestnikom konferencji odpowiadają punkty, a znajomościom – odcinki), dla której stosowne zakwaterowanie nie istnieje.



Rzeczywiście, niezależnie od sposobu zakwaterowania osoby A, pozostali uczestnicy dzielą się na dwie grupy 5-osobowe i jedną 4-osobową o tej własności, że każdy ma niezakwaterowanych dotąd znajomych tylko w obrębie danej grupy. Żadnej z grup 5-osobowych nie da się zakwaterować w dwuosobowych pokojach.

Wykorzystując przedstawioną teorię, możemy zaproponować następujący protokół prywatnego uzyskiwania informacji dla jednej bazy danych (Dobromira):

(a) Bogumił i Dobromir umawiają się na reprezentację ciągu $\boldsymbol{x}$ w postaci tablicy $\boldsymbol{X} = [\overline{x}_{k,l}]_{k\leqslant s, l\leqslant t}$. Załóżmy, że $x_i = \overline{x}_{\alpha,\beta}$;

(b) Bogumił wybiera duże liczby pierwsze p, q (których reprezentacja dwójkowa ma K bitów) i oblicza $n = pq$, po czym wybiera losowo $y_1, \ldots, y_s \leqslant n$ w taki sposób, że $y_\alpha \in \mathcal{O}_n$ oraz $y_k \in \mathcal{Q}_n$ dla $k \neq \alpha$. Następnie przekazuje n oraz wszystkie liczby $y_1, \ldots, y_s$ Dobromirowi. Zauważmy, że zgodnie z naszą uwagą Dobromir (nieznający p, q) dla żadnego $k \leqslant s$ nie jest w stanie stwierdzić, czy $y_k \in \mathcal{Q}_n$, nie dowie się zatem niczego o α;

(c) Dla każdego $r \leqslant t$ Dobromir oblicza $z_r = \prod_{k=1}^{s} y_k^{1+\overline{x}_{k,r}}$ modulo n. Jest to iloczyn wszystkich wysłanych liczb y_k, przy czym niektóre – te, którym w r-tej kolumnie odpowiada jedynka – mnożone są „w kwadracie”. Ponieważ tylko y_α nie jest kwadratem modulo n, więc z_r jest kwadratem tylko wtedy, gdy y_α jest mnożone „w kwadracie”, czyli gdy $x_{\alpha,r} = 1$;

(d) Bogumił sprawdza, czy z_β jest kwadratem (może to uczynić, gdyż zna p, q). Jeśli tak, to $x_{\alpha,\beta}$ wynosi 1, w przeciwnym przypadku 0.

Przedstawiona komunikacja zajmuje $Ks + Kt$ bitów, czyli w ten sposób, biorąc $s = t \approx \sqrt{n}$, możemy już osiągnąć komunikację rozmiaru $2K\sqrt{n}$. A można jeszcze lepiej! Zauważmy, że spośród skonstruowanych przez Dobromira liczb $z_1, \ldots, z_t$ Bogumiła interesuje tylko z_β, przy czym nie chce on, by Dobromir poznał β. Toż to brzmi dokładnie jak wyjściowy problem, więc rzecz pachnie rekurencją na kilometr! Bogumił może zastosować ten sam protokół dla ciągu $z_1, \ldots, z_t$, aby poznać z_β. Wówczas rozmiar komunikacji jest rzędu $Ks + K \cdot 2K\sqrt{t}$; optymalizując ze względu na s, t, pod warunkiem $st = n$, dostajemy koszt $3K^{5/3}\sqrt[3]{n}$. W ten rekurencyjny sposób możemy dowolnie zbijać wykładnik przy n (odwrotnie proporcjonalnie do głębokości rekurencji); niestety, kosztem puchnącego (z grubsza liniowo wraz z głębokością rekurencji) wykładnika przy K. Cóż, odwołując się do klasyka, nie udało nam się przyrządzić zupełnie darmowego obiadu, mamy jednak nadzieję, że Czytelnicy i tak docenią (podobnie jak Bogumił) chytrość i elegancję przedstawionych protokołów.

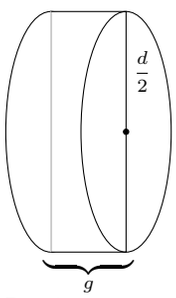
Kto ma rację?

*Jarosław GÓRNICKI**

* Wydział Matematyki i Fizyki Stosowanej, Politechnika Rzeszowska

Skończył się mecz – najważniejsze wydarzenie tygodnia. Po burzliwej wymianie zdań na jego temat trzej przyjaciele: Długi, Gruby i Ludek wracali do domu. Nagle Ludek zapytał o zadanie z matematyki, które było na jutro. Długi i Gruby stanęli jak zaczarowani. Zapomnieli o zadaniu. W necie na chwilę się zagotowało! Nastała cisza przerywana wiadomościami przychodzącymi na komórki. Nikt z klasy jeszcze zadania nie zrobił. Zadanie było krótkie:

Jak gruba powinna być moneta, aby szansa, że wyląduje ona na krawędzi, wynosiła $\frac{1}{3}$?



Rys. 1

Wszyscy zgodzili się przyjąć uproszczenie, że moneta jest jednorodnym, symetrycznym walcem. Gruby, który początkowo zbladł i spocił się, nieśmiało zgłosił pomysł. Posmarujemy deskę miodem (by rzucana moneta nie odbijała się i nie toczyła), sklejając pięciogroszówki stworzymy kilka wariantów „grubych” monet i na podstawie eksperymentu wybierzemy odpowiedź.

Długi uznał, że zadanie jest łatwe i szkoda miodu. To, na której „stronie” wyląduje moneta, jest proporcjonalne do pola powierzchni poszczególnych „stron”. Zatem warunki zadania będą spełnione, gdy pole powierzchni bocznej walca będzie równe polu podstawy. Obliczył (rys. 1):

$$2 \cdot \pi \cdot \frac{d}{2} \cdot g = \pi \cdot \left(\frac{d}{2}\right)^2, \quad \text{zatem} \quad 2g = \frac{d}{2}, \quad \text{skąd} \quad g = \frac{1}{4}d.$$