Materiały do wykładów z analizy algorytmów

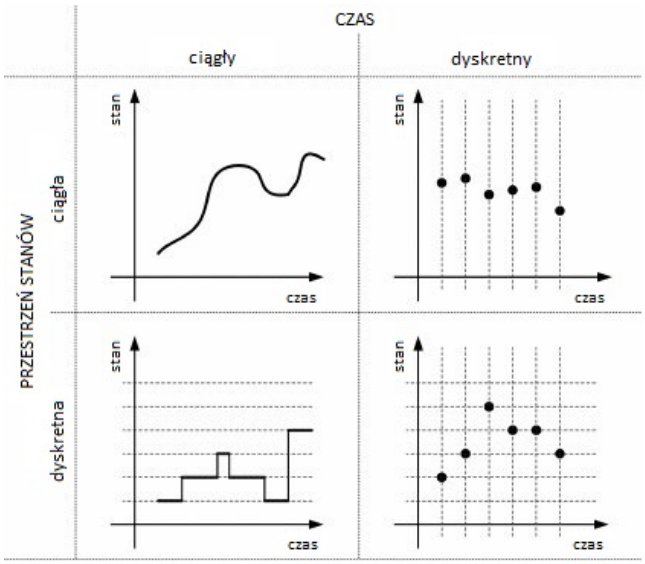
Procesy losowe w analizie systemów kryptowalut

Jakub Lemiesz

1 Procesy stochastyczne

Proces stochastyczny (proces losowy) to rodzina zmiennych losowych $\mathbf{X} = \{X_t : t \in T\}$, gdzie indeks t najczęściej reprezentuje czas. Często przyjmuje się $T = \mathbb{N}_{\geq 0}$ lub $T = \mathbb{R}_{\geq 0}$. Mówimy wtedy odpowiednio o procesie z czasem dyskretnym lub z czasem ciągłym[1]. O procesie stochastycznym możemy wówczas myśleć jak o modelowaniu zmiennej losowej, której wartość zmienia się w czasie. W projektowaniu i analizie algorytmów rozważa się zarówno procesy z czasem dyskretnym (np. PageRank, kryptografia) jaki i ciągłym (np. teoria kolejek, handel algorytmiczny).

Typowo wszystkie zmienne X_t przyjmują wartości z tej samej przestrzeni nazywanej przestrzenią stanów. W zależności od tego czy przestrzeń stanów odpowiada zbiorowi przeliczalnemu czy nieprzeliczalnemu, proces $\mathbf{X}$ nazywamy odpowiednio procesem z dyskretną lub ciągłą przestrzenią stanów.



Niekiedy przyjmuje się, że proces ma dodatkowo własność braku pamięci, czyli, że jego przyszła trajektoria zależy jedynie od aktualnego stanu, a nie od wcześniej przyjmowanych stanów. W zależności od tego czy mamy do czynienia z czasem dyskretnym czy ciągłym mówimy wówczas odpowiednio o łańcuchach lub procesach Markowa. Dalej pokażemy w jaki sposób tego rodzaju proces z czasem ciągłym może być wykorzystany do modelowania i analizy działania systemu kryptowalut opartego na technologii blockchain (np. Bitcoin).

[1] Dla procesów z czasem ciągłym zamiast X_t stosuje się również zapis $X(t)$.

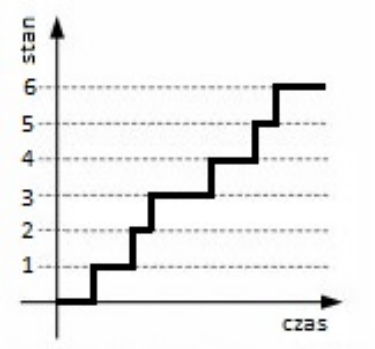
2 Proces Poissona

Jednym z najprostszych procesów stochastycznych z czasem ciągłym jest proces Poissona

$$\{N(t), t \in \mathbb{R}_{\geq 0}\} \ ,$$

gdzie $N(t)$ opisuje liczbę pewnego rodzaju zdarzeń, które zaszły w przedziale czasowym $[0, t]$. Jest to zatem proces zliczający z czasem ciągłym określony na dyskretnej przestrzeni stanów.

Rysunek 1: Przykładowa realizacja procesu Poissona.



Formalnie proces Poissona bywa definiowany na różne sposoby. Można pokazać, że następujące definicje są równoważne (dowód w [2], rozdział 8.4).

Definicja 1 *Procesem Poissona z parametrem intensywności λ nazywamy proces $\{N(t), t \in \mathbb{R}_{\geq 0}\}$ dla którego $N(0) = 0$ oraz który*

1. *jest jednorodny w czasie:*
   dla dowolnych $t, s > 0$ zmienne $N(t+s) - N(s)$ oraz $N(t) - N(0)$ mają ten sam rozkład,
2. *ma niezależne przyrosty:*
   dla $[t_1, t_2] \cap [t_3, t_4] = \emptyset$ zmienne $N(t_2) - N(t_1)$ i $N(t_4) - N(t_3)$ są niezależne
3. $\lim_{t \to 0} \frac{\Pr(N(t)=1)}{t} = \lambda$
   prawdopodobieństwo pojedynczego zdarzenia w krótkim przedziale czasu t dąży do λt,
4. $\lim_{t \to 0} \frac{\Pr(N(t)\geq 2)}{t} = 0$
   prawdopodobieństwo więcej niż jednego zdarzenia w krótkim przedziale czasu t dąży do zera.

Definicja 2 *Procesem Poissona z parametrem intensywności λ nazywamy proces $\{N(t), t \in \mathbb{R}_{\geq 0}\}$ dla którego $N(0) = 0$ oraz który*

1. *ma niezależne przyrosty,*
2. *liczba zdarzeń w dowolnym przedziale długości t ma rozkład Poissona z parametrem λt:*

$$N(t+s) - N(s) \sim Poisson(\lambda t) \ .$$

Warto w tym miejscu przypomnieć sobie podstawowe informacje dotyczące rozkładu Poissona. W szczególności warto zauważyć, że dla zmiennej $X \sim Poisson(\lambda)$ mamy:

$$\mathbb{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda} \ , \ \ \mathbb{E}(X) = \lambda \ , \ \ \mathbb{V}\mathrm{ar}(X) = \lambda.$$

Definicja 3 *Niech $T_1, T_2, T_3, \ldots$ będzie ciągiem niezależnych zmiennych losowych o rozkładzie wykładniczym z takim samym parametrem λ:*

$$(\forall i \in \mathbb{N})\,(T_i \sim Exp(\lambda))\ .$$

Wówczas definiujemy proces Poissona z parametrem intensywności λ jako

$$N(t) = \begin{cases} 0, & T_1 > t, \\ \max\{n : T_1 + \cdots + T_n \leq t\}, & T_1 \leq t. \end{cases}$$

W powyższej definicji wartość zmiennej T_i możemy interpretować jako czas, który upływa między $(i-1)$-szym a i-tym zdarzeniem w procesie Poissona.

W kolejnych sekcjach przedstawimy pokrótce idea działania systemów kryptowalutowych opartych na technologii blockchain. Do procesu Poissona wrócimy w sekcji 6, w której wprowadzimy model matematyczny opisujący proces wydobywania bloków.

3 Kryptowaluty

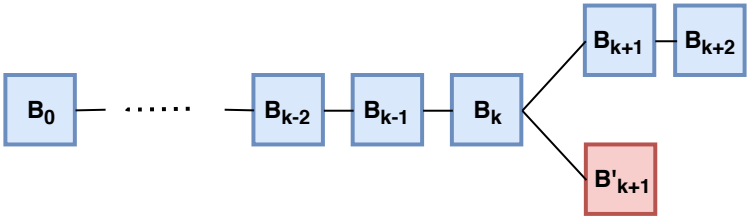
Typowo systemy kryptowalut opierają się na cyfrowo podpisanych deklaracjach, że właściciel pewnych monet (ang. coin) zgadza się przekazać je innemu użytkownikowi systemu, np. sprzedawcy. Właściciel monet zazwyczaj oczekuje w zamian dostarczenia wybranych produktów. Niebezpieczeństwo wiąże się z faktem, że po otrzymaniu produktów właściciel monet może próbować podważyć swoją wcześniejszą deklarację, np. poprzez wystawienie innej deklaracji dotyczącej tych samych monet (podwójne wydatkowanie). Zatem sprzedawca powinien wstrzymać się z dostarczeniem produktów dopóki nie jest pewien, że monety stały się jego niepodważalną własnością. Atak podwójnego wydatkowania (ang. double spending attack) jest próbą przekonania sprzedawcy, że transakcja doszła do skutku, a następnie przekonania innych użytkowników, że tak nie jest.

W celu utrudnienia ataku podwójnego wydatkowania systemy kryptowalut (np. Bitcoin [3]) powszechnie wykorzystują koncepcje dowodu wykonania pracy (ang. proof-of-work) . Do potwierdzenia każdej transakcji wymagany jest pewien koszt obliczeniowy polegający np. na obliczeniu dużej liczby haszy. W praktyce zazwyczaj potwierdza się nie pojedynczą transakcję, ale zbiór wielu transakcji zwany blokiem. Poprzez łączenie bloków w łańcuchy (ang. blockchain) zapewnia się, że koszt adwersarza usiłującego zmodyfikować wybraną transakcję nieustannie rośnie, gdyż wymaga stworzenia na nowo wszystkich bloków, które zostały przyłączone do bloku zawierającego tę transakcję. Modyfikacja transakcji może się jednak powieść jeśli w dyspozycji adwersarza są odpowiednio duże zasoby obliczeniowe i będzie on w stanie stworzyć alternatywny i wiarygodny łańcuch bloków.

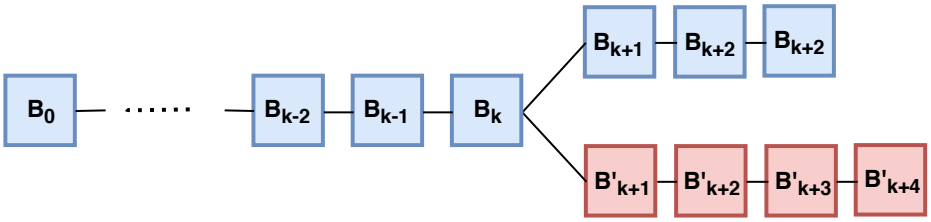
4 Wybór łańcucha bloków

Każdy blok zawiera referencję jednoznacznie identyfikującą jego rodzica, czyli pewien blok, który pojawił się wcześniej. Referencją może być np. hasz tego wcześniejszego bloku. Wyjątkiem jest jedynie pierwszy blok w historii (ang. genesis block). Bloki tworzą więc drzewo, którego korzeniem jest ów początkowy blok. Gałęzią drzewa nazywamy ścieżkę od bloku będącego liściem drzewa do bloku początkowego. Każda taka gałąź reprezentuje jedną wersję historii wykonanych transakcji i musi być wewnętrznie spójna (np. nie może zawierać dwóch wykluczających się transakcji). Różne gałęzie nie muszą być jednak ze sobą spójne.

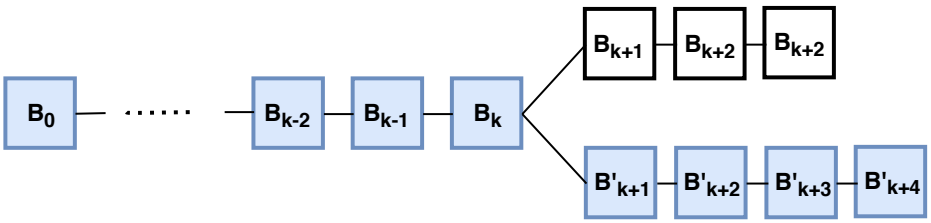
(a) Stan łańcucha przed rozpoczęciem ataku.



(b) Adwersarz rozpoczyna atak ,,double-spend'' budując swoją gałąź.
Uczciwe węzły pracują nad rozszerzeniem najdłuższej gałęzi łańcucha.



(c) Gałąź adwersarz staje się dłuższa od gałęzi, na której bazują uczciwe węzły.



(d) Uczciwe węzły porzucają gałąź, na której dotychczas bazowały.

Rysunek 2: Kroki ataku „double-spending”.

Z punktu widzenia użytkowników systemu kryptowalutowego pożądane jest by ustalić jedną spójna wersję historii wykonanych transakcji. Typowo ustala się, że to najdłuższa gałąź, której stworzenie wymagało największych kosztów obliczeniowych, jest obowiązującą wersją historii. Jeśli wiele gałęzi ma tę samą maksymalną długość, użytkownicy mogą bazować na każdej z nich (ang. fork), ale kiedy w końcu zostanie wygenerowany nowy blok, wydłużając jedną z tych gałęzi, impas zostanie przełamany.

Mówimy, że transakcja ma n potwierdzeń, jeśli

1. należy do pewnego bloku B, który jest częścią obowiązującej, najdłuższej gałęzi, oraz
2. na drodze od bloku B do liścia najdłuższej gałęzi znajduję się n bloków.

Zakłada się, że transakcja z wystarczającą liczbą potwierdzeń jest zabezpieczona przed atakiem podwójnego wydatkowania. W kolejnych sekcjach analizujemy to założenie. Autorem pierwszej analizy, zawierającej pewne uproszczenia, jest Satoshi Nakamoto, pomysłodawca systemu Bitcoin [3]. Druga analiza przeprowadzona przez Grunspana et al. ([1], 2017) jest bardziej precyzyjna.

5 Atak „double-spending”

Atak „double-spending” składa się z następujących kroków:

1. Adwersarz wysyła do systemu deklarację wykonania transakcji t przekazania monet sprzedawcy. Blok B zawierający transakcję t zostaje dołączony do najdłuższej gałęzi.
2. Adwersarz potajemnie buduje alternatywną gałąź bazującą na ostatnim bloku przed dołączeniem bloku B. W nowej gałęzi umieszcza transakcję t', w której nie przekazuje monet sprzedawcy, ale innemu użytkownikowi.
3. Adwersarz czeka, aż transakcja t otrzyma w aktualnie obowiązującej gałęzi odpowiednią liczbę potwierdzeń i sprzedawca wyśle zakupiony produkt.
4. Adwersarz rozbudowuje tajną gałąź do czasu, aż będzie ona dłuższa od aktualnie obowiązującej gałęzi i kiedy to nastąpi publikuje tajną gałąź (zobacz Rysunek 2). Ponieważ gałąź ta jest w tym momencie najdłuższa, inni użytkownicy systemu uznają ją za gałąź obowiązującą. Transakcja t w powszechnym mniemaniu zostaje zastąpioną transakcją t'.

Przyjmijmy, że sprzedawca przed wysłaniem produktu czeka na pojawienie się n potwierdzeń transakcji t. Załóżmy, że w chwili, gdy pojawi się n-te potwierdzenie adwersarz zdążył wygenerować m bloków. Jeśli $m > n$ to atak się powiódł. Jeśli $m \leq n$ interesujące wydaje się pytanie, jakie są szanse, że adwersarzowi, który nie ustaje w rozbudowie tajnej gałęzi, uda się w kiedyś uzyskać gałąź dłuższą niż obowiązująca (która również cały czas się wydłuża).

6 Model

Przyjmijmy, że bloki są generowane w następujący sposób:

1. Wylosuj wartość $nonce$ taką, że $0 \leq nonce \leq 2^{32} - 1$.
2. Skonkatenuj $nonce$ z haszem poprzedniego bloku h_{prev} oraz listą transakcji do umieszczenia w bloku aktualnym:
$$h = H(\, nonce \,||\, h_{prev} \,||\, t_1 \,||\, t_2 \,||\, \ldots) \,.$$
3. Sprawdź, czy hasz h traktowany jako liczba rzeczywista z przedziału $[0, 1)$ jest mniejszy od ustalonego progu δ (innymi słowy, czy ciąg bitów h ma odpowiednio dużą liczbę zer wiodących). Jeśli tak, udało się wygenerować nowy blok. Jeśli nie, powtórz całą procedurę. Wartość δ decyduje o trudności problemu i czasie potrzebnym do jego rozwiązania.

Zakładamy, że całkowita moc obliczeniowa użytkowników systemu jest stała, przy czym cześć p jest w dyspozycji uczciwych użytkowników, a część q w dyspozycji adwersarza oraz $p + q = 1$. Przyjmujemy, że wartość progu δ jest dobrana tak, by średni czas generowania bloku wynosił λ (np. w systemie Bitcoin $\lambda \approx 10\ minut$).

Niech zmienna losowa T oznacza czas potrzebny do wygenerowania nowego bloku. Zauważmy, że rozkład prawdopodobieństwa z którego pochodzi zmienna T posiada własność braku pamięci, gdyż próby wyliczenia odpowiedniego hasza dla kolejnych losowych wartości $nonce$ są od siebie niezależne. Przyjmijmy, że T jest zmienną losową o rozkładzie ciągłym i przypomnijmy, że wówczas własność braku pamięci jest zdefiniowana w następujący sposób.

Definicja 4 *Niech zmienna losowa X o rozkładzie ciągłym przyjmuje wartości z zakresu $[0, \infty)$. Mówimy, że rozkład zmiennej X ma własność braku pamięci jeśli zachodzi warunek:*

$$(\forall x_1, x_2 > 0)\,(\mathbb{P}[X > x_1 + x_2 | X > x_2] = \mathbb{P}[X > x_1])\ .$$

Lemat 1 *Jedynym rozkładem spełniającym warunki powyższej definicji jest rozkład wykładniczy.*

Dowód. Dowód lematu pozostawiamy jako ćwiczenie. Wskazówka: zdefiniuj $G(x) = \mathbb{P}[X > x]$ i pokaż, że z definicji własności braku pamięci musi zachodzić warunek $G(x) = G(1)^x$. Następnie pokaż, że jedyną ciągłą funkcją, która spełnia ten warunek dla dowolnego $x > 0$ jest $G(x) = e^{-\lambda x}$ dla $\lambda > 0$. ■

Z przyjętych założeń możemy zatem wydedukować, że czas potrzebny do wygenerowania bloku ma rozkład wykładniczy $T \sim Exp(\alpha)$ opisany funkcja gęstości

$$f_T(t) = \alpha e^{-\alpha t}\ ,$$

gdzie $\alpha > 0$ determinuje średni czas wydobycia bloku $\mathbb{E}[T] = 1/\alpha$.

Niech $T_k \sim Exp(\alpha)$ oznacza czas wykorzystany na wydobycie k-tego bloku. Zauważmy, że $T_1, T_2, T_3, \ldots$ są niezależnymi zmiennymi losowymi. Niech S_n oznacza łączny czas potrzebny na wydobycie n bloków. Zauważmy, że zmienna S_n ma rozkład gamma z parametrami (n, α).

Lemat 2 *Niech $T_1, T_2, T_3, \ldots, T_n$ będą niezależnymi zmiennymi losowymi oraz niech $T_k \sim Exp(\alpha)$ dla $k \in \{1, 2, \ldots n\}$. Pokaż, że że zmienna*

$$S_n = \sum_{k=1}^{n} T_k$$

ma rozkład gamma z parametrami (n, α).

Dowód. Zobacz ćwiczenie 24 i 25. ■

Dystrybuanta rozkładu gamma z parametrami (n, α) dla $n \in \mathbb{N}_{\geq 1}$ zadana jest wzorem:

$$F_{S_n}(t) = 1 - e^{-\alpha t} \sum_{k=0}^{n-1} \frac{(\alpha t)^k}{k!}\ .$$

Rozważmy ciąg zmiennych $S_1, S_2, S_3, \ldots$ i zdefiniujmy proces stochastyczny $N(t)$ określający liczbę wydobytych bloków do chwili t. Ustalmy $S_0 = 0$, wówczas:

$$N(t) = max\{n \geq 0 : S_n \leq t\}\ .$$

Warunek $N(t) = n$ jest równoważny koniunkcji warunków $S_n \leq t$ oraz $S_{n+1} > t$. Mamy zatem

$$\mathbb{P}[N(t) = n] = F_{S_n}(t) - F_{S_{n+1}}(t) = \frac{(\alpha t)^n}{n!} e^{-\alpha t}\ . \tag{1}$$

Zauważmy, że dla ustalonego t mamy $N(t) \sim Poisson(\alpha n)$. Ponadto czas pomiędzy wydobyciem kolejnych bloków ma rozkład wykładniczy z parametrem α. Zatem $\{N(t), t \geq 0\}$ musi być procesem Poissona o intensywności α (zobacz sekcje 2).

7 Wyścig z adwersarzem

Rozważmy wyścig pomiędzy adwersarzem a uczciwymi użytkownikami, zachodzący w trakcie ataku „double-spending" (zobacz sekcje 5). Przypiszmy zmienne losowe T i S_n oraz proces $N(t)$ zdefiniowane w poprzedniej sekcji do procesu wydobywania bloków przez uczciwych użytkowników. Analogiczne, niech zmienne T' i S'_n oraz proces $N'(t)$ będą związane z wydobywaniem bloków przez adwersarza. Zmienne T oraz T' są niezależne od siebie. Przyjmijmy, że

$$T \sim Exp(\alpha) \quad \text{oraz} \quad T' \sim Exp(\alpha')\ ,$$

gdzie α i α' opisują średnią częstotliwość z jaką wydobywane są nowe bloki i odzwierciedlają moc obliczeniową, którą dysponują uczciwe węzły i adwersarz.

Można pokazać, że zmienna losowa $M = \min(T, T')$, opisująca czas wydobycia bloku biorąc pod uwagę zarówno bloki wydobywane przez uczciwych użytkowników i przez adwersarza, ma rozkład wykładniczy z parametrem $\alpha + \alpha'$ (ćwiczenie). Można również pokazać, że prawdopodobieństwo wydobycie kolejnego bloku przez adwersarza wyraża się wzorem (ćwiczenie):

$$\mathbb{P}[T' < T] = \frac{\alpha'}{\alpha + \alpha'} \, .$$

Zatem prawdopodobieństwo q wydobycia kolejnego bloku przez adwersarza i prawdopodobieństwo p wydobycia kolejnego bloku przez uczciwych użytkowników są związane z tempem wydobywania bloków:

$$q = \frac{\alpha'}{\alpha + \alpha'} \, ,$$
$$p = 1 - q = \frac{\alpha}{\alpha + \alpha'} \, .$$

Dla przykładu, w systemie Bitcoin trudność zadania znalezienia dowodu wykonania pracy jest dostosowywana tak, aby średni czas wydobycia nowego bloku $\tau = \frac{1}{\alpha + \alpha'}$ wynosił około 10 minut. Wówczas oczekiwany czas wydobycia nowego bloku wynosi odpowiednio dla uczciwych użytkowników i dla adwersarza

$$\mathbb{E}[T] = \frac{1}{\alpha} = \frac{\tau}{p} \, ,$$
$$\mathbb{E}[T'] = \frac{1}{\alpha'} = \frac{\tau}{q} \, .$$

W dalszej części będziemy zakładać, że $0 < q < \frac{1}{2}$, czyli, że adwersarz ma do dyspozycji mniej niż połowę całkowitej mocy obliczeniowej dostępnej w systemie. Rozważmy sytuację, w której adwersarzowi brakuje n bloków do zrównania się z aktualnie obowiązującą gałęzią.

Lemat 3 *Niech q_n będzie prawdopodobieństwem zajścia zdarzenia E_n, które odpowiada zrównaniu się aktualnie najdłuższej gałęzi łańcucha i gałęzi adwersarza, w sytuacji kiedy adwersarzowi do zrównania się brakuje n bloków. Wówczas*

$$\mathbb{P}[E_n] = q_n = \left(\frac{q}{p}\right)^n \, .$$

Dowód. Dowód pozostawiamy jako ćwiczenie. Analiza jest podobna do analizy klasycznego problemu ruiny gracza (zobacz np. [2]). ■

8 Analiza Nakamoto

W tej sekcji zaprezentujemy analizę ataku double-spending przedstawioną przez Nakamoto w znanej pracy [3] po raz pierwszy opisującej system Bitcoin. Analiza ta zawiera pewne uproszczenia, to których odniesiemy się w dalszej części.

W systemie Bitcoin powszechnie przyjmuje się, że po nadbudowaniu bloku zawierającego daną transakcję użytkownika przez $n = 6$ bloków, użytkownik może uważać transakcję za potwierdzoną. Na potrzeby analizy przyjmijmy, że wartość powyższego parametru bezpieczeństwa n jest dowolna. Rozważmy moment, w którym uczciwi użytkownicy nadbudowali wyjściowy blok n nowymi blokami i przyjmijmy, że adwersarz wydobył do tego momentu pewną liczbę bloków k. Jeżeli $k \geq n$, to adwersarzowi udało się dogonić łańcuch budowany przez uczciwe węzły i ma duże szansę przeprowadzenia ataku Jeżeli $k < n$, to z Lematu 3 wynika, że prawdopodobieństwo dogonienia przez adwersarza łańcucha budowanego przez uczciwych użytkowników wynosi $(\frac{q}{p})^{n-k}$. Rozważmy prawdopodobieństwo $P(n)$, że adwersarzowi uda się zrównać z łańcuchem budowanym przez uczciwych użytkowników w momencie, gdy nadbudowali oni rozważany blok n blokami lub później. Zauważmy, że możemy przedstawić $P(n)$ jako:

$$P(n) = \mathbb{P}[N'(S_n) > n] + \sum_{k=0}^{n} \mathbb{P}[N'(S_n) = k] \cdot q_{n-k} \, .$$

W dalszej części analizy Nakamoto przyjmuje, że faktyczny czas wydobywania bloków równy wartości oczekiwanej czasu wydobycia tych bloków, co istotnym uproszczeniem. Uproszczenie to pozwala na przybliżenie wartości $N'(S_n)$ przez $N'(\mathbb{E}[S_n])$, gdzie:

$$\mathbb{E}[S_n] = n\mathbb{E}[T] = n\frac{\tau}{p} \, .$$

Ze wzoru (1) wynika, że zmienna $N'(\mathbb{E}[S_n])$ ma rozkład Poissona z parametrem λ, gdzie:

$$\lambda = \alpha' \mathbb{E}[S_n] = \alpha' n \frac{\tau}{p} = n \frac{\alpha'}{\alpha + \alpha'} \frac{1}{p} = n\frac{q}{p} \, ,$$

czyli

$$\mathbb{P}[N'(\mathbb{E}[S_n]) = k] = e^{-\lambda} \frac{\lambda^k}{k!} \, .$$

Zauważmy, że $q_0 = 1$. Możemy zatem napisać:

$$P(n) \approx \mathbb{P}[N'(\mathbb{E}[S_n]) > n] + \sum_{k=0}^{n} \mathbb{P}[N'(\mathbb{E}[S_n]) = k] \cdot q_{n-k}$$

$$= 1 - \sum_{k=0}^{n} \mathbb{P}[N'(\mathbb{E}[S_n]) = k] + \sum_{k=0}^{n} \mathbb{P}[N'(\mathbb{E}[S_n]) = k] \cdot q_{n-k}$$

$$= 1 - \sum_{k=0}^{n-1} \mathbb{P}[N'(\mathbb{E}[S_n]) = k](1 - q_{n-k})$$

$$= 1 - \sum_{k=0}^{n-1} e^{-\lambda} \frac{\lambda^k}{k!}(1 - q_{n-k}) \, .$$

9 Analiza Grunspan'a

Analiza Nakamoto nie jest precyzyjna, gdyż opiera się na założeniu, iż $N'(S_n) = N'(\mathbb{E}[S_n])$. Poniżej przedstawiamy analizę ataku double-spending, którą Grunspan et al. zaprezentowali w pracy [1]. Stosujemy oznaczenia z wcześniejszych sekcji.

Twierdzenie 1 *Niech X_n będzie zmienną losową oznaczającą liczbę bloków wydobytych przez adwersarza w chwili gdy uczciwi użytkownicy nadbudowali rozważany blok n blokami. Wówczas zmienna X_n ma rozkład ujemny dwumianowy z parametrami n oraz p*

$$\mathbb{P}[X_n = k] = p^n q^k \binom{k+n-1}{k}.$$

Dowód. Niech $k \geq 0$. Zauważmy, że zmienne N' oraz S_n są niezależne. Mamy zatem:

$$\begin{aligned}
\mathbb{P}[X_n = k] &= \int_0^{+\infty} \mathbb{P}[N'(t) = k] \cdot f_{S_n}(t) dt \\
&= \int_0^{+\infty} \frac{(\alpha' t)^k}{k!} e^{-\alpha' t} \cdot \frac{\alpha^n}{(n-1)!} t^{n-1} e^{-\alpha t} dt \\
&= \frac{\alpha^n (\alpha')^k}{(n-1)!k!} \cdot \int_0^{+\infty} t^{k+n-1} e^{-t(\alpha'+\alpha)} dt \\
&= \frac{p^n q^k}{(\tau)^{n+k}(n-1)!k!} \cdot \int_0^{+\infty} t^{k+n-1} e^{-\frac{t}{\tau}} dt \left|\begin{matrix} \frac{t}{\tau}=z \\ \\ \frac{dt}{\tau}=dz \end{matrix}\right| \\
&= \frac{p^n q^k}{(\tau)^{n+k}(n-1)!k!} \cdot \int_0^{+\infty} (\tau)^{k+n-1} z^{k+n-1} e^{-z} \tau dz \\
&= \frac{p^n q^k}{(\tau)^{n+k}(n-1)!k!} \cdot (\tau)^{k+n} \cdot \int_0^{+\infty} z^{k+n-1} e^{-z} dz \\
&= \frac{p^n q^k}{(n-1)!k!} \cdot \Gamma(k+n) dz \\
&= \frac{p^n q^k}{(n-1)!k!} \cdot (k+n-1)! \\
&= p^n q^k \binom{k+n-1}{k}.
\end{aligned}$$

W powyższym wyprowadzeniu skorzystaliśmy, ze definicji funkcji gamma:

$$\Gamma(x) = \int_0^{+\infty} z^{x-1} e^{-z} dz,$$

a także z faktu, że dla $n \in \mathbb{N}_{\geq 1}$ mamy

$$\Gamma(n) = (n-1)!\,.$$

■

Nawiązując do uproszczenia dokonanego w analizie Nakamoto zauważmy, że w istocie rozkład zmiennej X_n zbiega do rozkładu Poissona.

Twierdzenie 2 *Niech $q \to 0$ i $l_n = n \cdot \frac{q}{p} \to \lambda$ dla $n \to \infty$. Wówczas:*

$$\mathbb{P}[X_n = k] \to \frac{\lambda^k}{k!} e^{-\lambda}\,.$$

Dowód.

$$\mathbb{P}[X_n = k] = \frac{n^n}{(n+l_n)^n} \cdot \frac{l_n^k}{(n+l_n)^k} \cdot \frac{(k+n-1)!}{(n-1)!k!} = \frac{l_n^k}{k!} \cdot \frac{1}{(1+\frac{l_n}{n})^n} \cdot \frac{n(n+1)\dots(n+k-1)}{(n+l_n)^k} .$$

Po skorzystaniu z faktu, że $(1+\frac{l_n}{n})^n \to e^\lambda$ otrzymujemy wynik. ■

W oparciu o Twierdzenie 1 możemy precyzyjnie wyznaczyć prawdopodobieństwo ataku ataku double spending.

Twierdzenie 3 *Niech $P(n)$ oznacza prawdopodobienstwem zdarzenia, że adwersarz zbuduje łańcuch równy lub dłuższy niż ten budowany przez uczciwych użytkowników w momencie gdy nadbudowali oni rozważany blok n blokami lub później. Wówczas:*

$$P(n) = 1 - \sum_{k=0}^{n-1} (p^n q^k - q^n p^k) \binom{k+n-1}{k} .$$

Dowód. Dowód twierdzenia przebiega w dużym stopniu analogicznie dowodu Nakamoto:

$$P(n) = \sum_{k>n} \mathbb{P}\left[X_n = k\right] + \sum_{k=0}^{n} q_{n-k} \mathbb{P}\left[X_n = k\right]$$

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

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

$$= 1 - \sum_{k=0}^{n} p^n q^k \binom{k+n-1}{k} (1 - q^{n-k} p^{k-n})$$

$$= 1 - \sum_{k=0}^{n} (p^n q^k - q^k q^{n-k} p^n p^{k-n}) \binom{k+n-1}{k}$$

$$= 1 - \sum_{k=0}^{n} (p^n q^k - q^n p^k) \binom{k+n-1}{k}$$

$$= 1 - \sum_{k=0}^{n-1} (p^n q^k - q^n p^k) \binom{k+n-1}{k} .$$

■

Zauważmy, że dystrybuantę zmiennej o rozkładzie ujemnym dwumianowym można wyrazić za pomocą unormowanej funkcji beta:

$$\mathbb{P}\left[X_n \leq k\right] = \sum_{l=0}^{k} p^n q^l \binom{l+n-1}{l} = 1 - I_p(k+1, n) .$$

Zapisując formułę na $P(n)$ z Twierdzenia 3 za pomocą unormowanej funkcji beta Grunspanowi udało się również w zwięzły sposób opisać asymptotyczne zachowanie funkcji $P(n)$.

Lemat 4 *Dla $s = 4pq < 1$ oraz $n \to \infty$ mamy*

$$P(n) = I_s\left(n, \frac{1}{2}\right) \sim \frac{s^n}{\sqrt{\pi(1-s)n}} .$$

Dowód lematu 4 znajduje się w pracy Grunspana [1]. Zauważmy, że w pracy Nakamoto pojawił się postulował, iż prawdopodobieństwo udanego ataku $P(n)$ w tempie wykładniczym maleje do 0 wraz ze wzrostem wartości parametru n, czyli wymaganej liczby potwierdzeń. Prezentowane rozumowanie nie było jednak w pełni formalne.

10 Lista ćwiczeń

1 — Pokaż, że jedynym rozkładem spełniającym warunki Definicji 4 jest rozkład wykładniczy.

2 — Rozważmy dwie niezależne zmienne o rozkładach wykładniczych: $T \sim Exp(\alpha)$ oraz $T' \sim Exp(\alpha')$. Pokaż, że zmienna $\min(T, T')$ ma rozkład wykładniczy z parametrem $\alpha + \alpha'$. Uogólnij ten wynik na minimum $n > 2$ niezależnych zmiennych losowych o o rozkładzie wykładniczym.

3 — Pokaż, że dla zmiennych losowych z poprzedniego zadania mamy:

$$\mathbb{P}[T' < T] = \frac{\alpha'}{\alpha + \alpha'} \, .$$

4 — Udowodnij Lemat 3. Wskazówka: rozważ klasyczny problem ruiny gracza w sytuacji gdy gracz ma nieograniczone środki.

5 — Przeanalizuj dowód Lematu 4 przedstawiony w pracy [1].

Literatura

[1] C. Grunspan and R. Perez-Marco. Double Spend Races. *International Journal of Theoretical and Applied Finance (IJTAF)*, 21(08):1–32, December 2018.

[2] M. Mitzenmacher and E. Upfal. *Probability and computing - randomized algorithms and probabilistic analysis*, pages 212–218. Cambridge University Press, 2005.

[3] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Information Processing Letters*, 5(1):15–17, 2008.