

POP  
Projekt: Sieć optyczna 1  
Dokumentacja wstępna

Bartosz Latosek

Szymon Skarzyński

Styczeń 2023

## Spis treści

<b>1</b>	<b>Wstępny opis problemu</b>	<b>2</b>
<b>2</b>	<b>Użyte technologie</b>	<b>3</b>
<b>3</b>	<b>Sformułowanie sposobu rozwiązania</b>	<b>3</b>
3.1	Ogólne sformułowanie . . . . .	3
3.2	Zarys algorytmu ewolucyjnego . . . . .	4
<b>4</b>	<b>Realizacja</b>	<b>5</b>
<b>5</b>	<b>Operacje genetyczne</b>	<b>5</b>
5.1	Selekcja . . . . .	5
5.2	Krzyżowanie . . . . .	5
5.3	Mutacja . . . . .	5
<b>6</b>	<b>Eksperymenty</b>	<b>6</b>
6.1	Prawdopodobieństwo mutacji genu . . . . .	6
6.2	Prawdopodobieństwo krzyżowania . . . . .	6
6.3	Liczebność populacji . . . . .	7
6.4	Liczba epok . . . . .	7
<b>7</b>	<b>Podsumowanie</b>	<b>7</b>
<b>8</b>	<b>Opis wyników</b>	<b>8</b>

# 1 Wstępny opis problemu

Celem zadania jest minimalizacja kosztu realizacji sieci optycznej za pomocą algorytmu ewolucyjnego przy spełnieniu wszystkich zapotrzebowań, jednocześnie nie przekraczając pojemności włókna światłowodowego.

Sieć optyczna rozumiana jest w tym przypadku jako zbiór węzłów (miast) pomiędzy którymi zdefiniowane są ścieżki o ustalonej przepustowości. Zapotrzebowanie jest wyrażone w GB/s i określa minimalny przepływ, który musi zostać zapewniony między dwoma miastami. Dla  $n$  miast mamy  $n(n-1)/2$  zapotrzebowań. Zapotrzebowanie pomiędzy miastem A i B a miastem B i A są wyrażone osobno, co sprawia, że mamy do czynienia z grafem skierowanym. Oznacza to, że mamy osobną pojemność światłowodów dla połączenia A- $\rightarrow$ B i B- $\rightarrow$ A. Zapotrzebowania również wyrażone są osobno dla połączenia A- $\rightarrow$ B i B- $\rightarrow$ A. Wszystko to wynika z otrzymanego zbioru danych, który założeniem różni się od tego przewidywanego w sprawozdaniu początkowym.

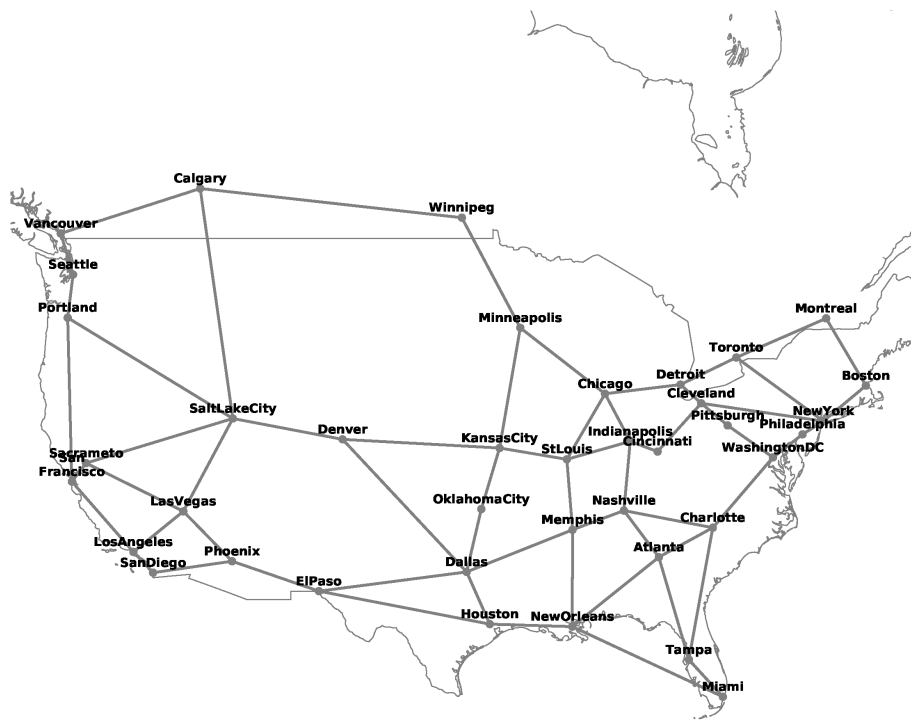
Krawędzie między miastami to światłowód o określonej liczbie par włókien - w rzeczywistości jedno z włókien w parze jest odpowiedzialne za nadawanie, a drugie za odbiór. Każda taka para włókien to nic innego jak po prostu szkło - medium transmisyjne, w którym przesyłane jest światło.

W rozpatrywanym problemie miasta połączone są za pomocą światłowodów o jednej parze włókien (odbior i nadawanie), w każdym takim włóknie możemy wydzielić do 1008 (Co również wynika ze zbioru testowego i jest różne od początkowo założonej wartości 96) 'slotów' komunikacyjnych odpowiadających różnym długościom fali świetlnej. Wynika z tego, że, pojedyncza krawędź grafu może realizować maksymalnie 1008 zapotrzebowań.

W celu obsługi przesyłania i odbierania danych musimy na początku i na końcu wybranej ścieżki zainstalować kartę transpondera o ustalonej pojemności (100 GB/s, 200 GB/s lub 400 GB/s), przy czym każda para kart zajmuje nam jeden 'slot' (długość fali). Załóżmy, że zapotrzebowanie pomiędzy miastem A i B wynosi 400 GB/s. W takim wypadku możemy je zaspokoić instalując na końcach ścieżki 4 najtańsze karty - 100 GB/s, zajmując przy tym 4 sloty, 2 karty - 200 GB/s, zajmując 2 sloty albo jedną, najdroższą kartę - 400 GB/s zajmując przy tym tylko jeden slot.

Problem optymalizacji sprowadza się do takiego doboru kart na końcach ścieżek, aby każde zapotrzebowanie zostało spełnione, przy minimalnym koszcie i nie przekraczając przy tym maksymalnej pojemności pojedynczego włókna - 1008 długości fali ( $\lambda$ ).

W rzeczywistości, jeżeli zapotrzebowanie dotyczy oddalonych od siebie miast, to może być konieczne umieszczenie dodatkowych transponderów na ścieżce w celu wzmocnienia sygnału, ale na potrzeby zadania zakładamy, że sygnał nie musi być wzmacniany.



Schemat amerykańskiej sieci światłowodowej  
Źródło: DOI:10.3390/app10196840

## 2 Użyte technologie

Użyjemy języka Python. Użyte zostały biblioteki:

- **geopy** - do obliczania odległości między miastami
- **matplotlib** - do szkicowania wykresów
- **numpy** - do operacji matematycznych
- **xml** - do parsowania zbioru danych

. Jako, że zadanie dotyczy stricte optymalizacji kosztów, rezygnujemy też z jakiegokolwiek formy graficznej przedstawienia rozwiązania.

## 3 Sformułowanie sposobu rozwiązania

### 3.1 Ogólne sformułowanie

W naszym modelu zostaną przyjęte następujące zbiory:

- $N$  - Zbiór węzłów sieci (Każdy węzeł utożsamiany jest z jednym miastem)
- $E$  - Zbiór krawędzi (połączeń światłowodowych pomiędzy miastami)
- $T$  - Zbiór rodzajów kart transponderów (Pojemności: 100, 200 i 400)

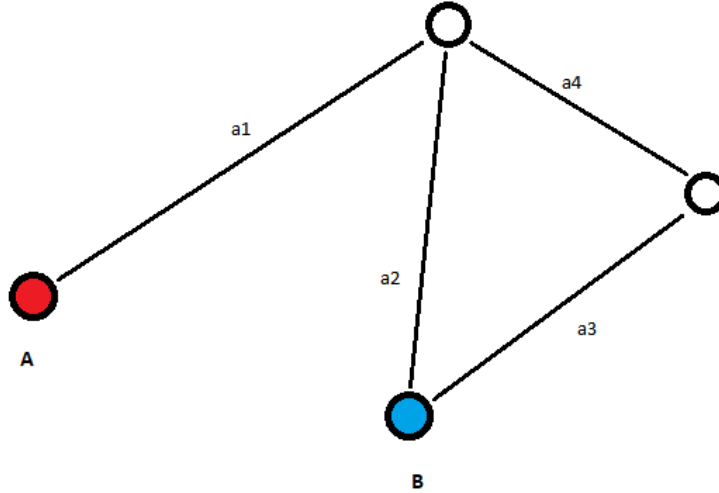
Zadanie polega na dobraniu ilości każdej z kart ze zbioru  $T$  dla każdego węzła osobno tak, aby każde zapotrzebowanie w sieci zostało spełnione przy minimalnym koszcie. Koszt w tym przypadku zdefiniowany jest jako suma pojemności użytych kart transponderów.

### 3.2 Zarys algorytmu ewolucyjnego

Na sam początek należy zastanowić się, czym będzie pojedynczy osobnik populacji. Doszliśmy do wniosku, że osobnikiem będzie zbiór par z których każda składa się z dwóch wektorów:

- reprezentującego dobrane karty transponderów
- reprezentującego ścieżkę (zbiór krawędzi)

Każda taka para, częściowo lub w pełni realizuje jedno zapotrzebowanie.



I tak np. dla powyższego rysunku poglądowego - interesuje nas zapotrzebowanie pomiędzy węzłem A i B. Załóżmy, że wynosi ono 500G (oznaczenie G utożsamiamy z GB/s). I tak, realizacja zapotrzebowania może nastąpić w całości po ścieżce  $[a_1, a_2]$ , lub częściowo po ścieżce  $[a_1, a_2]$  i częściowo po ścieżce  $[a_1, a_4, a_3]$ . Dla pierwszego wariantu PRZYKŁADOWY osobnik będzie postaci:

$$T = [1, 0, 1], E = [a_1, a_2]$$

, gdzie T oznacza zbiór użytych kart transponderów dla odpowiadającej ścieżki. Taki zapis oznacza, że całe zapotrzebowanie będzie realizowane przez ścieżkę  $[a_1, a_2]$  i zostaną do tego użyte karty transponderów: 1 x 100G, 0 x 200G i 1 x 400G, co w sumie daje 500G. To samo zapotrzebowanie może zostać zrealizowane w postaci:

$$T_1 = [1, 1, 0], E_1 = [a_1, a_2]$$

$$T_2 = [2, 0, 0], E_2 = [a_1, a_4, a_3]$$

Powyższe oznacza, że  $1 \times 100G + 2 \times 200G = 300G$  zapotrzebowania będzie przesyłane ścieżką  $[a_1, a_2]$ , a pozostałe  $2 \times 100G = 200G$ , ścieżką  $[a_1, a_4, a_3]$ . Osobnikiem populacji będzie n takich zbiorów realizacji zapotrzebowania, gdzie n oznacza ilość ścieżek zapotrzebowania. Mutacja osobnika będzie polegała na zmianie użytej liczby transponderów danego typu oraz sposobu realizacji zapotrzebowania - ścieżek. Minimalizowany będzie koszt zużytych kart transponderów wraz. Poprawność osobnika będzie sprawdzana za pomocą wyodrębnionej funkcji walidacyjnej. W fazie selekcji algorytmu, wykorzystamy selekcję turniejową, chociaż nie wykluczamy testowania też innych metod.

## 4 Realizacja

W celu weryfikowania poprawności rozwiązania oraz stworzenia abstrakcji pozwalającej nam na modelowanie rozwiązania, zdecydowaliśmy się na stworzenie oddzielnej klasy - *model* spełniającej właśnie te wymagania. Na pojedynczego osobnika populacji przypada jeden model, który opisuje zajętość światłowodów przy uwzględnieniu realizacji zapotrzebowań dla danego osobnika. Zdecydowaliśmy się minimalizować jedynie koszt transponderów - nie interesuje nas zajętość  $\lambda$  o ile nie przekracza ona dopuszczalnej wartości. W związku z czym generowanie populacji początkowej, a później operacje krzyżowania i mutacji opierają się na operacjach związanych z dobraniem odpowiednich kart transponderów spełniających zapotrzebowanie.

Generowanie osobników polega na losowym dobraniu transponderów tak, aby suma ich pojemności spełniała zapotrzebowanie pomiędzy daną parą miast. Następnie, przy użyciu modelu generowana jest (za pomocą algorytmu Dijkstry) najkrótsza ścieżka, która ma jeszcze miejsce na  $\lambda$ dy. W związku z tym, że dostępność  $\lambda$ dy może być mniejsza, równa bądź większa niż zajętość, której potrzebujemy do realizacji zapotrzebowania, mogą wystąpić 2 przypadki:

1. Generowana jest jedna ścieżka spełniająca rozwiązanie - jest ona optymalna (najkrótsze połączenie między miastami)
2. Generowane jest wiele ścieżek, przez co zapotrzebowanie realizowane jest przez sumę transponderów umieszczonych na ich końcach

Z powyższej zależności wynika format genomu osobnika jako mapę zapotrzebowanie - realizacja, gdzie realizacja jest reprezentowana jako  $n$  par (ścieżka - użyte transpondery), których suma realizuje zapotrzebowanie.

Podczas eksperymentów zdecydowaliśmy się też na dodanie ilości zajętych  $\lambda$ dy do funkcji celu, ale był to błąd ponieważ algorytm w podstawowej formie stara się minimalizować koszt transponderów - przy arbitralnie dobranych kosztach oznacza to używanie jak największej liczby transponderów 100GB, a w formie wzbogaconej o długości  $\lambda$ dy chcemy minimalizować zajętość pasm co faworyzuje używanie transponderów 400GB. Cele są rozbieżne, przez co ostatecznie zdecydowaliśmy się usunąć zajętość pasm z funkcji celu.

Funkcja celu w obecnej postaci uwzględnia jedynie koszty użytych kart transponderów i jest wyrażona jako suma wszystkich użytych transponderów dla danego osobnika (a więc dla wszystkich zapotrzebowań).

## 5 Operacje genetyczne

### 5.1 Selekcja

Stosujemy selekcja turniejową z zachowaniem elity reprezentowanej jako najlepszy osobnik poprzedniej populacji. Osobnik ten pomijany jest w procesie krzyżowania i mutacji.

### 5.2 Krzyżowanie

Krzyżowanie występuje ze zdefiniowanym prawdopodobieństwem, w przypadku wystąpienia - dla osobnika dobierany jest losowy osobnik z populacji a następnie ich genom jest krzyżowany wymiennie - realizacja zapotrzebowania jest pobierana losowo z rodzica nr 1 lub 2. Następnie dla nowego zestawu transponderów generowana jest ścieżka spełniająca warunki zadania.

### 5.3 Mutacja

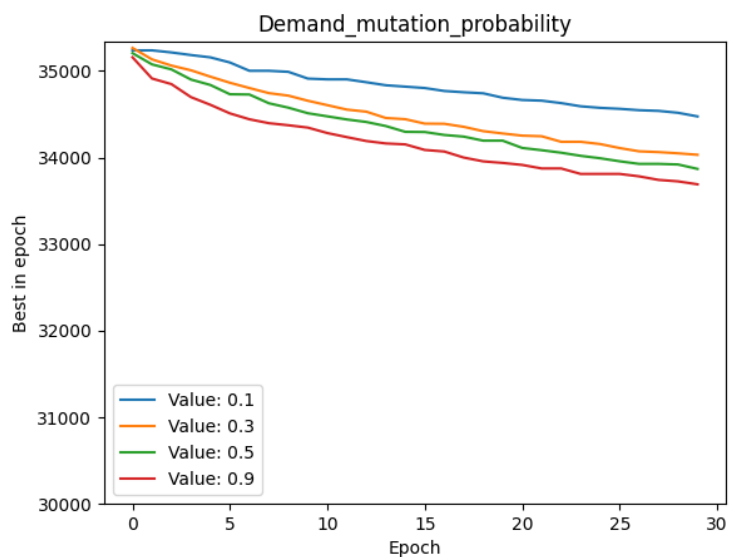
Mutacja polega na losowym przekształceniu zbioru transponderów ścieżki (np. 1x200 w 2x100) oraz zastąpieniu tej ścieżki nową, wygenerowaną dla przekształconego zestawu transponderów.

## 6 Eksperymenty

Zostaną przeprowadzone eksperymenty wizualizowane za pomocą wykresu funkcji celu od czasu, dla podanych ustalonych wartości poszczególnych hiperparametrów. Eksperymenty zostaną przeprowadzone dla następujących domyślnych hiperparametrów.

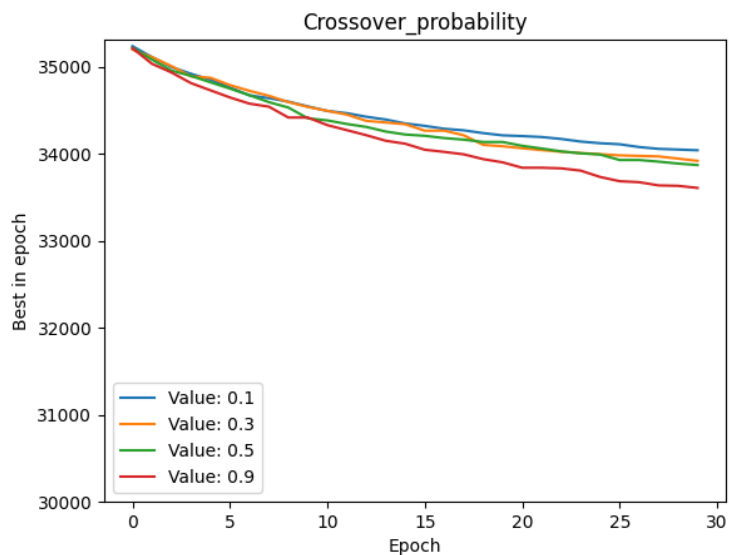
- Liczba epok - 30
- Liczebność populacji - 20
- Prawdopodobieństwo krzyżowania - 0.5
- Prawdopodobieństwo mutacji genu - 0.5
- Prawdopodobieństwo mutacji osobnika - 1 (w celu zachowania różnorodności populacji)

### 6.1 Prawdopodobieństwo mutacji genu



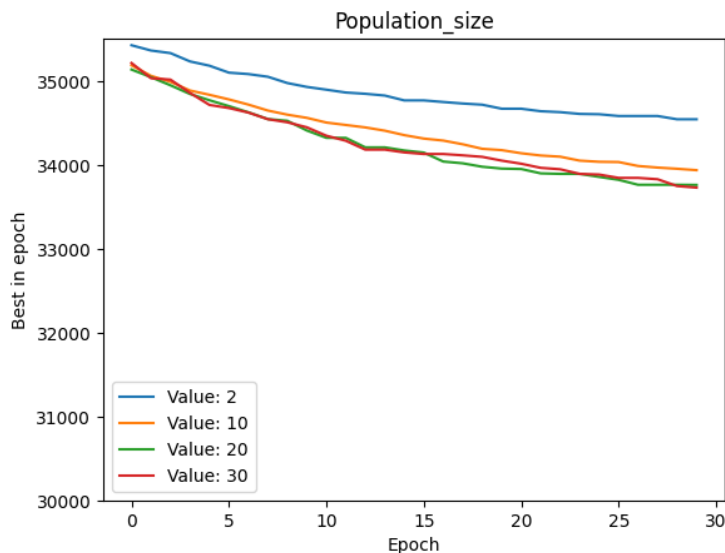
Zwiększanie prawdopodobieństwa mutacji genu (realizacji pojedynczego zapotrzebowania sieci), znacząco poprawiało wyniki. Funkcja celu szybciej zmniejszała wartość, a zwłaszcza w początkowych epokach, w późniejszej fazie, różnice wartości kosztu dla różnych wartości wydają się być dość stałe. Każde zwiększenie prawdopodobieństwa było korzystne i ta tendencja się nie zmieniła nawet przy osiągnięciu wartości 0.9.

### 6.2 Prawdopodobieństwo krzyżowania



Rezultaty zwiększania wartości prawdopodobieństwa krzyżowania były podobne jak w przypadku mutacji, lecz różnice dla różnych wartości nie były aż tak silne. Efekty zmiany tego parametru nie były też widoczne tak bardzo w początkowych epokach, a uwidoczniły się bardziej dopiero po czasie. W przeciwieństwie do poprzedniego wykresu (prawdopodobieństwa mutacji), różnice wartości kosztu dla różnych wartości stają się coraz bardziej oddalone nawet w późniejszych epokach. Jest to związane z tym, że krzyżowanie uśrednia populację a więc wspomaga eksploatację rozwiązań, przez co znajduje lepsze wartości funkcji celu.

### 6.3 Liczebność populacji



Zwiększenie liczebności populacji znacznie poprawia uzyskiwane rozwiązania. Widoczna jest duża różnica pomiędzy liczebnością 2 oraz 10, aczkolwiek wraz z dalszym zwiększaniem różnice nie były już tak duże. Dla liczebności 20 oraz 30 uzyskujemy bardzo podobne wykresy. Dalsze zwiększanie liczebności nie poprawia już rozwiązań, natomiast naturalnie spowalnia działanie algorytmu. Charakter zmiany wartości kosztu najlepszego osobnika jest bardzo podobny dla każdej z testowanych rozmiarów populacji, jednak dla większych wartości algorytm trochę szybciej polepsza rozwiązanie.

### 6.4 Liczba epok

Na powyższych wykresach widać naturalną zależność między liczbą epok, a wartościami funkcji celu. Im dłużej działa algorytm tym lepsze rozwiązania znajduje, co jest naturalnym zjawiskiem i nie wymaga specjalnego testowania.

## 7 Podsumowanie

Działanie algorytmu jest uzależnione od wielu czynników, zarówno hiperparametrów jak i metodyki, tzn. użytej selekcji czy mutacji. Dobranie optymalnych parametrów jest zadaniem trudnym i wymaga inżynierskiego podejścia do problemu - testowania różnych kombinacji parametrów i obserwowania zachowania algorytmu.

## 8 Opis wyników

Przykładowe wyniki działania algorytmu są przedstawione w pliku *output.csv*. Dane mają następującą formę:

1. ID ZAPOTRZEBOWANIA
2. ŚCIEŻKA
3. UŻYTE KARTY 100GB
4. UŻYTE KARTY 200GB
5. UŻYTE KARTY 400GB

Dla danego zapotrzebowania może występować kilka ścieżek symbolizujących rozbić realizacji tego zapotrzebowania.