



Zastosowanie metod uczenia maszynowego w sieciach optycznych

26.05.2022

Krzysztof Walkowiak
Politechnika Wroclawska
Wydział Informatyki i Telekomunikacji



Wprowadzenie



Podziękowania

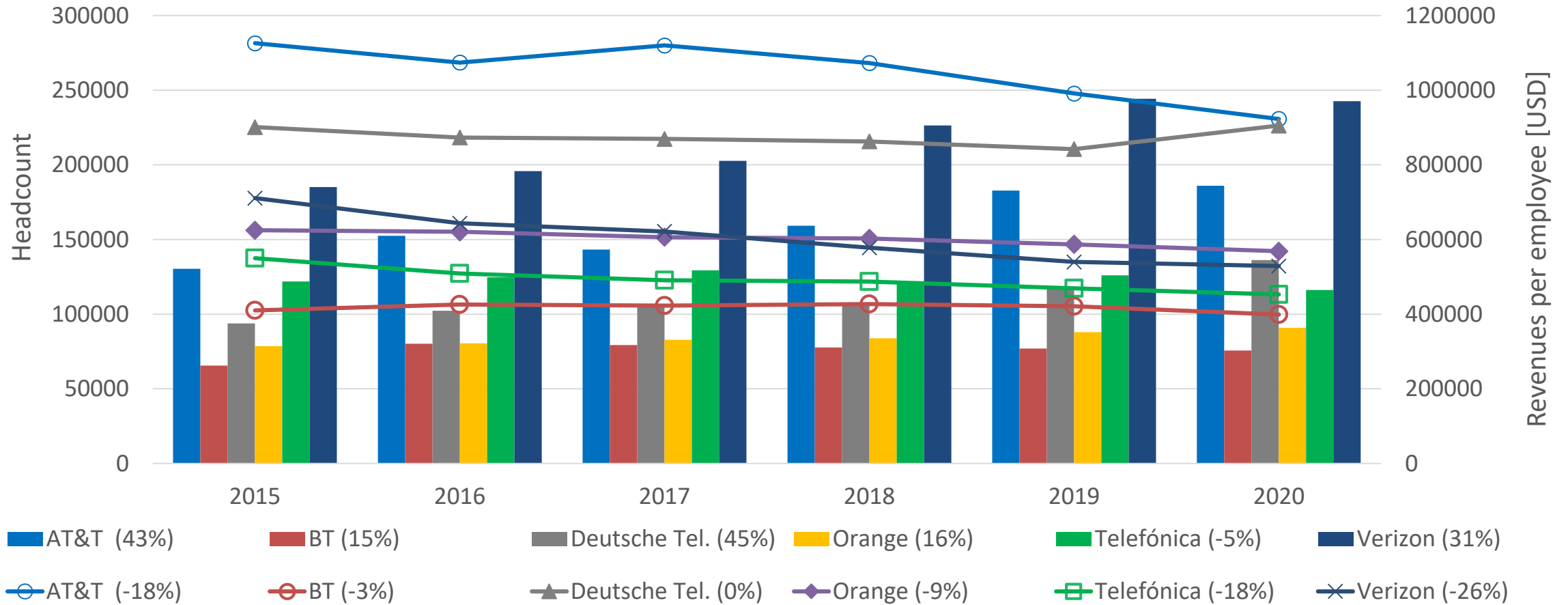
- Projekt **NCN OPUS** „Optymalizacja kognitywnych sieci optycznych”, 2018-2022, nr 2017/27/B/ST7/00888
- Projekt **NCN OPUS** „Zaawansowane metody optymalizacji wielowarstwowych sieci uwzględniających aplikacje”, 2020-2024, nr 2019/35/B/ST7/04272
- **Dr hab. inż. Mirosław Klinkowski**, Instytut Łączności
- **Dr inż. Róża Goścień**, Politechnika Wroclawska
- **Dr inż. Piotr Lechowicz**, Politechnika Wroclawska
- **Mgr inż. Aleksandra Knapińska**, Politechnika Wroclawska
- **Mgr inż. Daniel Szostak**, Politechnika Wroclawska
- **Mgr inż. Adam Włodarczyk**, Politechnika Wroclawska
- **Dr inż. Paweł Ksieniewicz**, Politechnika Wroclawska
- **Mgr inż. Weronika Węgier**, Politechnika Wroclawska
- **Mgr inż. Szymon Wojciechowski**, Politechnika Wroclawska



Motywacja dla zastosowań metod ML w sieciach optycznych

- **Wzrost popularności metod ML** w wielu obszarach nauki i przemysłu
- **Wzrost ilości zbieranych danych** w sieciach optycznych
 - Telemetria
 - Analityka danych
- **Wzrost efektywności metod ML**
 - Rozwój nowych metod ML, szczególnie metod głębokiego uczenia DL (ang. *deep learning*)
 - Wzrost mocy obliczeniowej
 - Dedykowane układy scalone dla metod ML, DL
- Rozwój **nowych metod zarządzania sieciami** teleinformatycznymi
- **Możliwości rozwiązania problemów**, które były **trudne/nieemożliwe** do rozwiązania stosując **dotychczasowe metody**

Telekomy – zatrudnienie i przychody

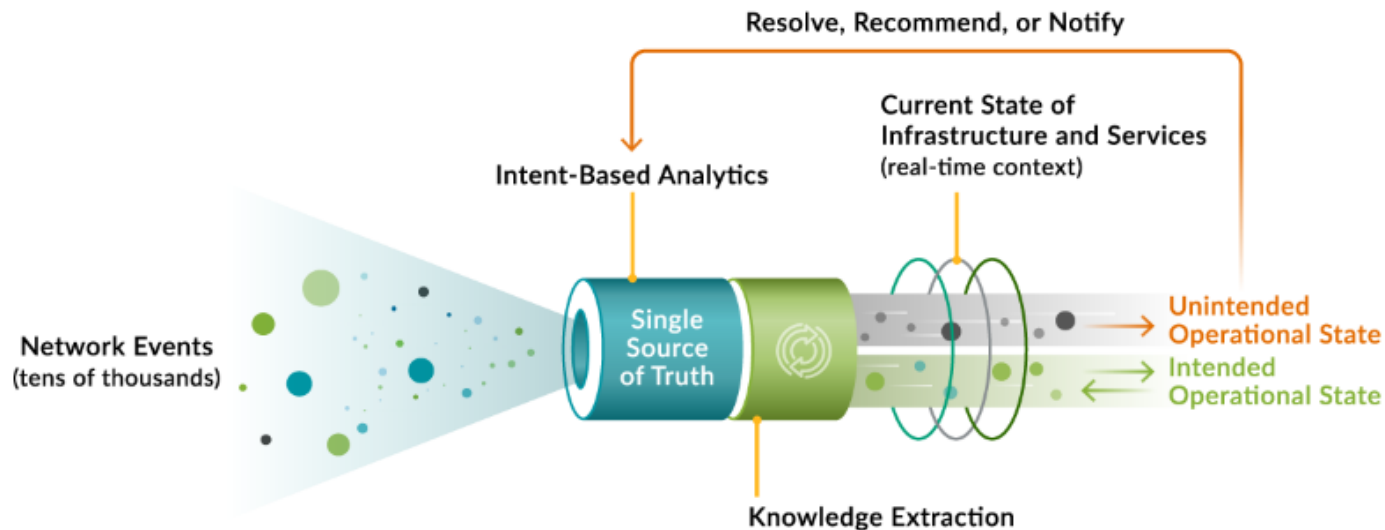


Source: <https://www.lightreading.com/aiautomation/big-20-telcos-have-cut-230k-jobs-12-of-total-in-five-years/d/d-id/769923>

Ewolucja metod zarządzania sieciami teleinformatycznymi

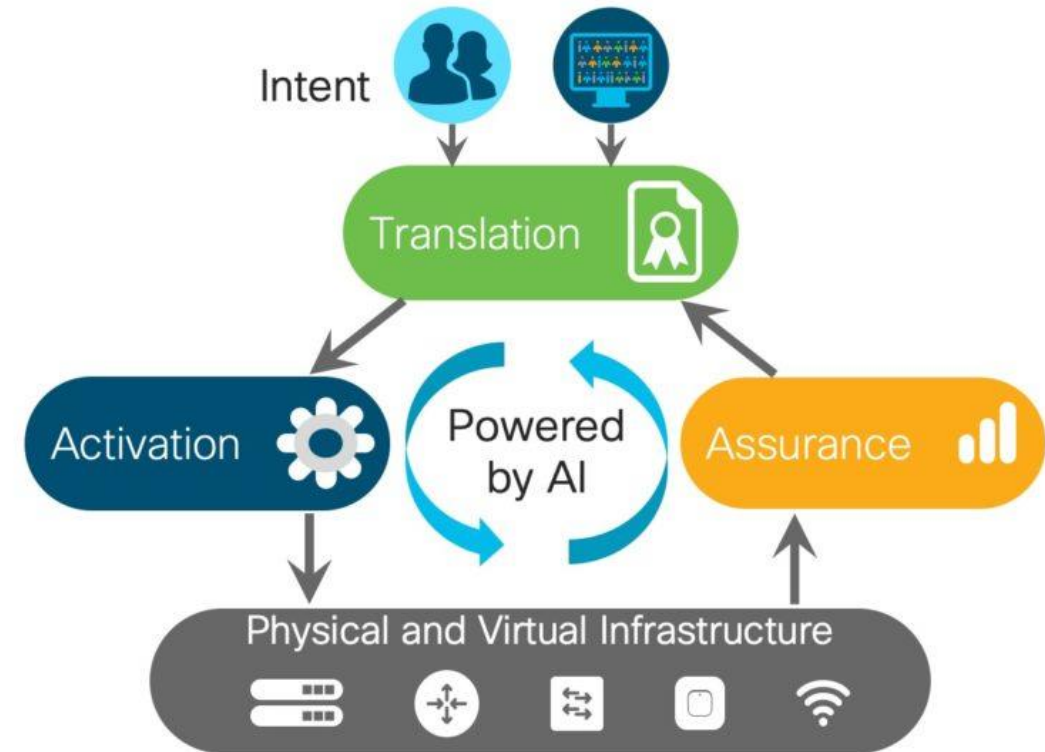
- Podejście „klasyczne”
- SDN (ang. *Software-defined networking*)
- Network automation, zero touch network
- **Intent-based networking**

Network Analytics at its Core



Źródło: <https://www.juniper.net/>

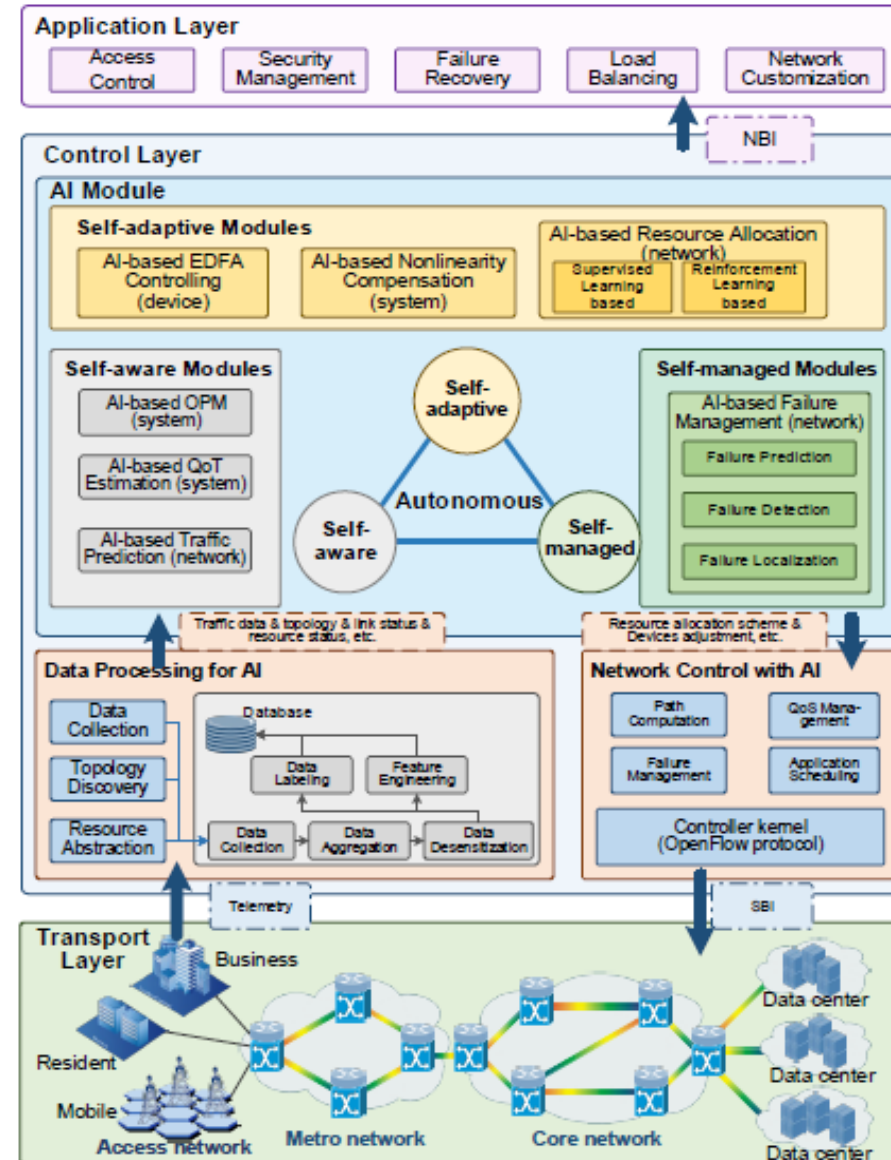
AI in Intent-Based Networking



Źródło: <https://www.cisco.com/>

IEEE Future Networks

- IEEE Future Directions stworzyła grupę **IEEE 5G Initiative** w grudniu 2016
- Aby lepiej odzwierciedlić wkład IEEE w obecne i przyszłe generacje sieci, inicjatywę zmieniono na **IEEE Future Networks** w sierpniu 2018
- Organizacja m.in. publikuje mapy drogowe rozwoju różnych obszarów sieci teleinformatycznych <https://futurenetworks.ieee.org/roadmap>, jeden z obszarów to „**Artificial Intelligence and Machine Learning**”





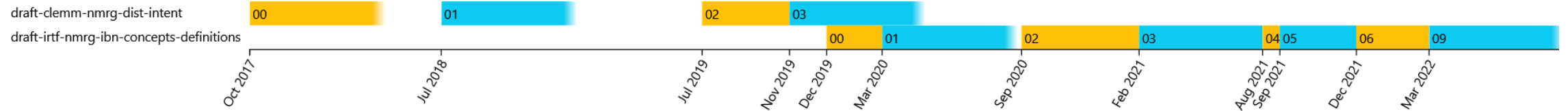
Intent-Based Networking według IETF

Datatracker Groups ▾ Documents ▾ Meetings ▾ Other ▾ User ▾ Sign in

Status [IRSG evaluation record](#) [IESG evaluation record](#) [IESG writeups](#) [Email expansions](#) [History](#)

Versions:

- [00](#)
- [01](#)
- [02](#)
- [03](#)
- [04](#)
- [05](#)
- [06](#)
- [07](#)
- [08](#)
- [09](#)



Document	Type	Active Internet-Draft (nmrg RG)
	Authors	Alexander Clemm ✉, Laurent Ciavaglia ✉, Lisandro Zambenedetti Granville ✉, Jeff Tantsura ✉
	Last updated	2022-04-12 (Latest revision 2022-03-24)
	Replaces	draft-clemm-nmrg-dist-intent
	Stream	Internet Research Task Force (IRTF)
	Intended RFC status	Informational
	Formats	plain text html xml htmlized pdfized bibtex
Stream	IRTF state	In IRSG Poll (Due date 2022-04-26 00:00 PDT)
	Consensus boilerplate	Yes
	Document shepherd	Jérôme François ✉
	Shepherd write-up	Show Last changed 2021-12-21



ETSI (European Telecommunications Standards Institute)

- <https://www.etsi.org/technologies/experiential-networked-intelligence>
- ETSI GS ENI 005 V2.1.1 (2021-12) Experiential Networked Intelligence (ENI). System Architecture:
 - Definitions
- ETSI GR ENI 010 V1.1.1 (2021-03) Experiential Networked Intelligence (ENI). Evaluation of categories for AI application to Networks
- ETSI GR ENI 008 V2.1.1 (2021-03). Experiential Networked Intelligence (ENI). InTent Aware Network Autonomicity (ITANA)



Obszary zastosowań metod ML w sieciach optycznych

- **Analiza i dobór parametrów fizycznych elementów sieci**
 - Badanie jakości transmisji QoT (ang. *Quality of Transmission*)
 - Szacowanie zasięgu transmisji
 - Dobór modulacji
- **Analiza stanu sieci** (ang. *network health*)
 - Na podstawie danych historycznych, predykcja które elementy sieci ulegną awarii, co pozwala na wcześniejszą interwencję
- **Wsparcie dla metod optymalizacji sieci optycznych**
 - Efektywniejszy przegląd przestrzeni rozwiązań
 - Dobór parametrów algorytmu
- **Przewidywanie ruchu i obciążenia sieci**
 - Możliwość zastosowania uzyskanej wiedzy w decyzjach dotyczących routingu i alokacji zasobów
- **Bezpieczeństwo sieci**



Główne wyzwania (i szanse) w stosowaniu ML w sieciach optycznych

- **Brak danych lub trudny dostęp do danych**
 - Operatorzy telekomunikacyjni posiadają takie dane, ale bardzo niechętnie dzielą się tymi danymi z naukowcami
- Dostępne **dane są często niezbalansowane**
 - Niezbędna jest dokładana analiza posiadanych danych i ewentualnie stosowanie odpowiednich metod ML
- **Nie zawsze metody ML działają dobrze**
 - Niezbędne jest przeprowadzenie dokładnych badań oraz odpowiednie dobranie metod referencyjnych
- Potrzeba **dostosowania „klasycznych” metod ML** do rozważanych problemów
- **Tradycyjnie miary jakości działania metod ML nie zawsze pasują** do rozważanych problemów
 - Potrzebne jest opracowanie nowych/dodatkowych miar jakości
- **Badania eksperymentalne** w zakresie ML są **stosunkowo łatwe** dzięki dostępnym bibliotekom, np. Python, R



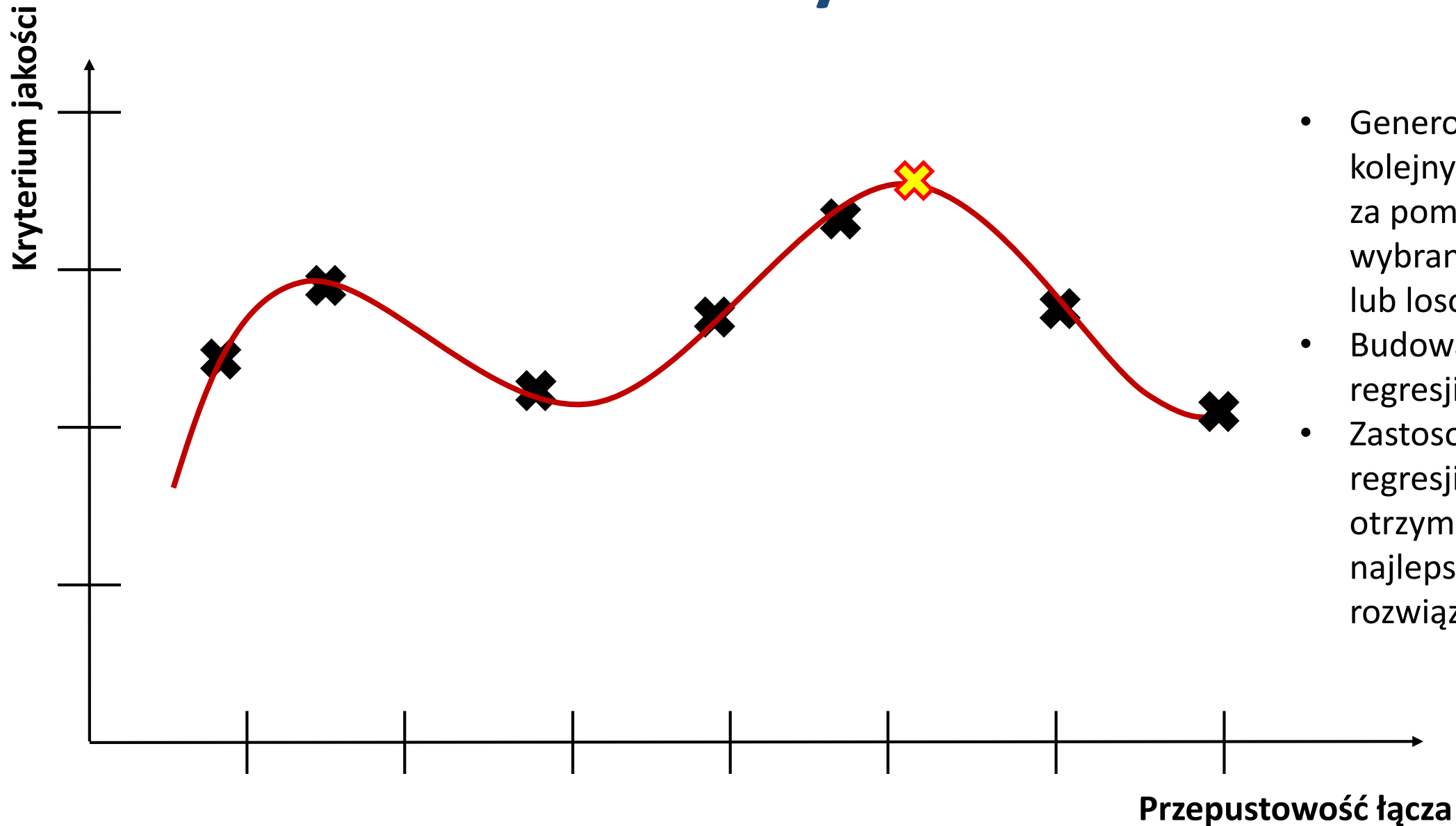
ML jako wsparcie dla metod optymalizacji sieci optycznych



Zastosowanie metod ML do wyznaczania rozwiązań problemów optymalizacyjnych

- W sytuacji kiedy do otrzymania **jednego rozwiązania** danego problemu potrzebny jest na **długi czas obliczeń** trudno stosować zaawansowany algorytm, który przegląda przestrzeń rozwiązań analizując bardzo dużo różnych rozwiązań (np. algorytm ewolucyjny)
- Mając stosunkowo **niewielki zbiór uzyskanych rozwiązań** z innych algorytmów budowany jest **model regresji**:
 - Cechy to wartości wejściowe danego problemu (np. przepustowości łącza w sieci)
 - Zmienna objaśniana to wartość rozwiązania (np. koszt sieci) uzyskanego dla ustalonych wartości wejściowych
 - Model regresji dopasowuje wybraną funkcję/model
 - Otrzymany model regresji jest następnie użyty w celu znalezienia jak najlepszych rozwiązań (np. metodą przeszukiwania typu grid search)

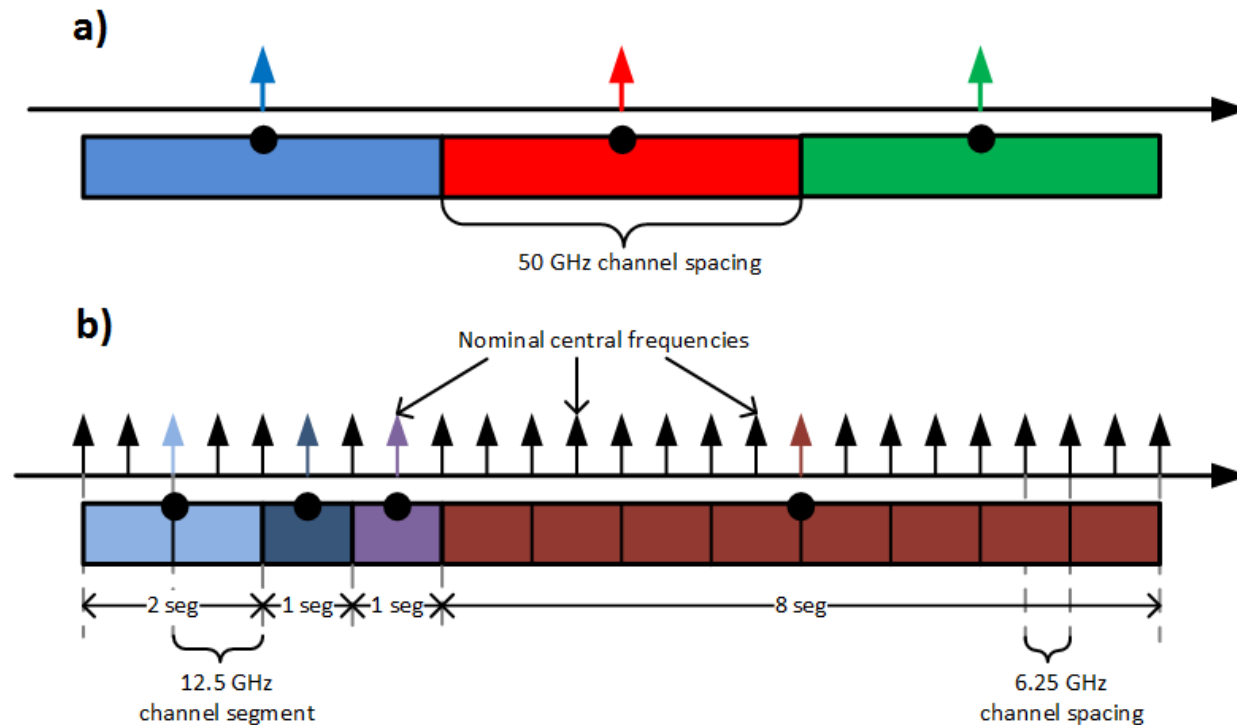
Przykład



- Generowanie kolejnych rozwiązań za pomocą wybranego algorytmu lub losowo
- Budowa modelu regresji
- Zastosowanie modelu regresji do otrzymania jak najlepszego rozwiązania

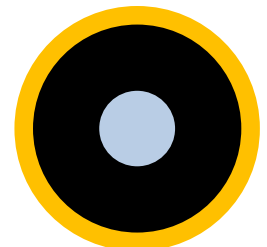
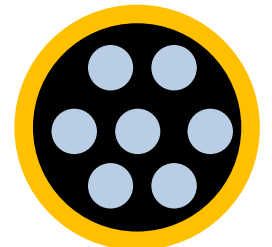
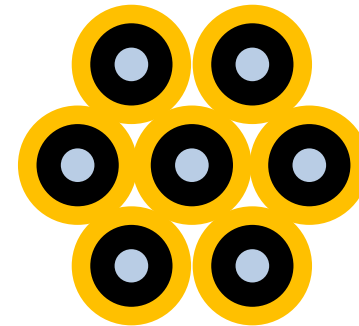
Spectrally-Spatially Flexible Optical Networks (SS-FON)

Sieci **SS-FON** to połączenie koncepcji elastycznych sieci optycznych **EON** (ang. *Elastic Optical Network*) i sieci **SDM** (ang. *Space Division Multiplexing*)



Rodzaje sieci SDM:

- **Fiber bundle** – wiele równoległych włókien światłowodów
- **Multicore fiber** – wiele rdzeni w jednym włóknie światłowodów
- **Multimode fiber** – wiele modów w jednym włóknie



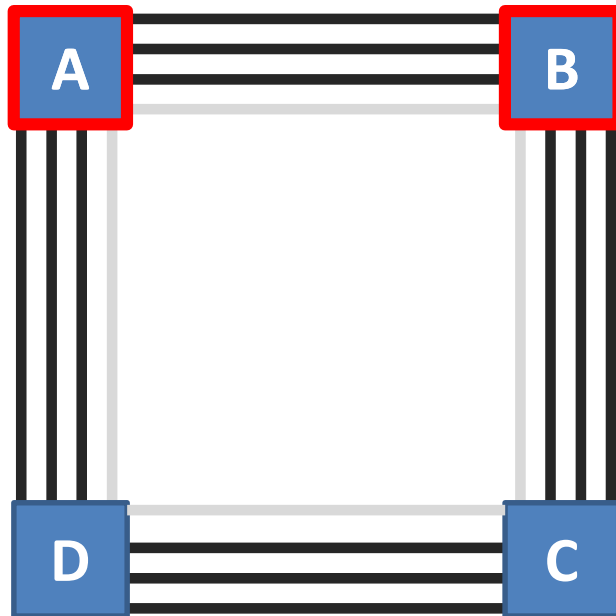
Wymiarowanie sieci SS-FON

- [Sz. Wojciechowski, R. Goścień, P. Ksieniewicz, K. Walkowiak, *Hybrid regression model for link dimensioning in spectrally-spatially flexible optical networks*, IEEE Access. 2022]
- Główny cel to **zwiększenie wolumenu danych**, który można przesłać w sieci SS-FON
- **Rozbudowa sieci SS-FON poprzez dodanie nowych zasobów** (np. instalacja nowych światłowodów oraz urządzeń) oznacza **dodatkowe koszty i zajmuje stosunkowo dużo czasu**
- **Operatorzy sieci preferują metody**, które wykorzystują **obecną infrastrukturę** (brak dodatkowych kosztów)
- Proponowana metoda pozwala na wymiarowanie łączy w sieci SS-FON **poprzez zmianę liczby wykorzystywanych włókien** w niektórych łączach sieci wykorzystując tzw. **ciemne włókna**, co oznacza **brak dodatkowych kosztów**, ponieważ wykorzystywane są istniejące zasoby
- Zmiana wymiarowania może być dokonana stosunkowo szybko, np. stosując podejście **AoD (ang. *architecture on demand*)**
- Procedura wymiarowania może być powtarzana **okresowo lub uruchamiana zgodnie ze zmianami monitorowanych parametrów wydajności sieci**

Model sieci SS-FON - założenia

- Każde łącze SS-FON ma dostępnych **kilka włókien** zagregowanych w wiązkę (**fiber bundle**)
- Włókna mogą być:
 - **aktywne**, tj. mogą być obecnie używane do alokacji ścieżek optycznych
 - **nieaktywne (ciemne)**, nie można ich obecnie używać do alokacji ścieżek optycznych, ale są gotowe do wykorzystania po ich aktywacji
- Jako punkt wyjścia przyjmujemy, że SS-FON jest zwymiarowany **jednolicie**, tj. **każde łącze $e \in E$ ma taką samą liczbę aktywnych włókien** oznaczonych jako K (tj. $active(e) = K$)
- Każdy węzeł $v \in V$ jest wyposażony w OXC (optyczne połączenie krosowe), które ma **ograniczoną liczbę portów wejścia/wyjścia** określoną jako $port(v) = (K \cdot deg(v) + SP)$, gdzie
 - $K \cdot deg(v)$ to porty niezbędne do obsługi K aktywnych włókien w każdym łączy podłączonym do v
 - SP oznacza zapasowe porty pozostawione przez operatora w celu ułatwienia zmiany rozmiaru łączy, te zapasowe porty mogą być użyte do zwiększenia liczby włókien używanych na niektórych łączy

Przykład



- Dla każdego węzła $v=\{A,B,C,D\}$ $port(v)=6$
- Dla każdego łącza e $active(e)=3$
- Dla każdego łącza e $limit(e)=4$

Algorytm

- Nowe włókno w łączu (A,B)
 - Węzły A i B przekraczają limit na portów
 - Łącza (A,D) i (B,C) wyłączają 1 włókno
 - Nowe włókno w łączu (C,D)
- **Wszystkie ograniczenia są spełnione**



Metody rozwiązania problemu wymiarowania SS-FON

- **Algorytm heurystyczny:**
 - Na podstawie wybranej metryki łącza, wybierane są łącza do zwiększenia liczby włókien
 - Możliwe metryki: TopA (Topology Aware), RouA (Routing Aware), TraA (Traffic Aware), ModA (Modulation Aware)
 - Jeżeli po zwiększeniu liczby włókien przekroczony został limit wykorzystanych portów w węźle, należy zmniejszyć liczbę włókien dla tego węzła
- **Model matematyczny ILP (Integer Linear Programming):**
 - Kryterium zdefiniowane jako miara obciążenia sieci (łączy sieciowych)
 - Ograniczenia zgodne z założeniami modelu
 - Rozwiązanie za pomocą metody CPLEX
- **Algorytm losowy:**
 - Liczba włókien na każdym łączu jest losowana według ustalonego zakresu
 - Zachowane są wszystkie ograniczenia modelu

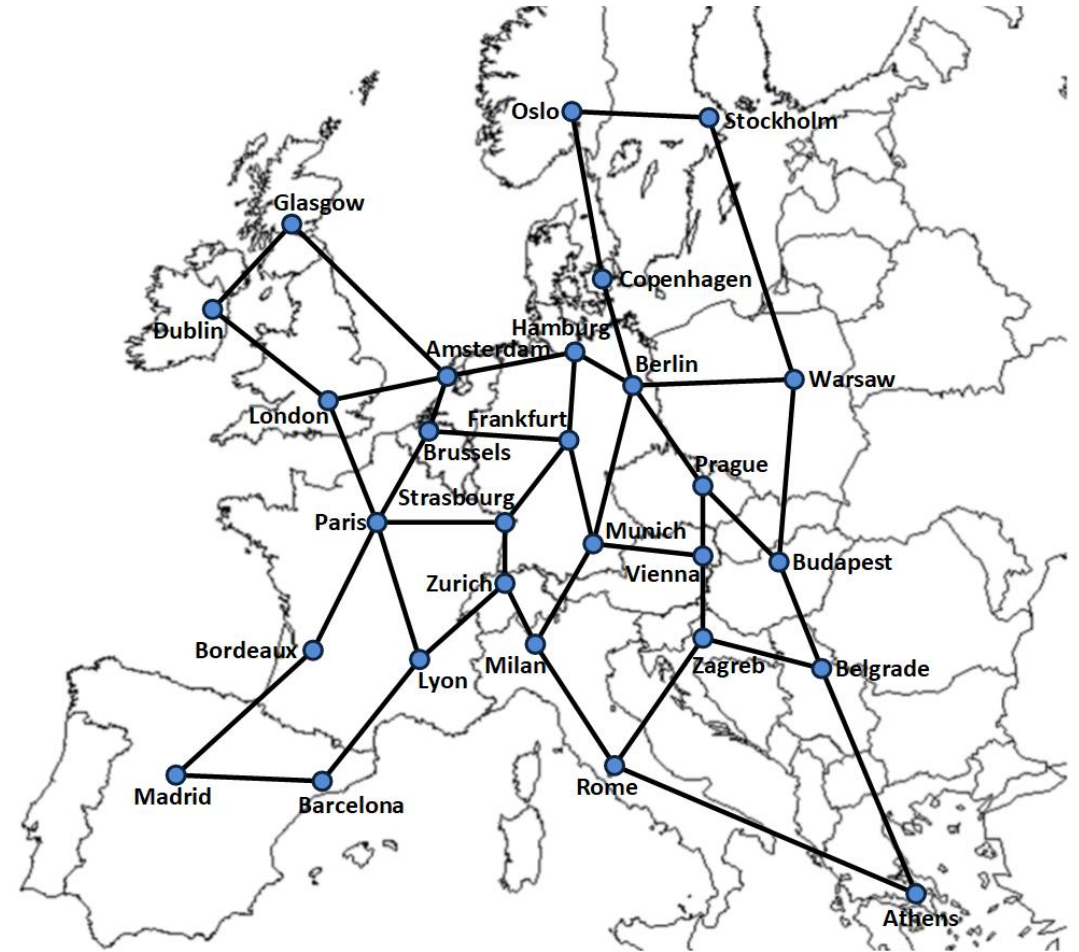
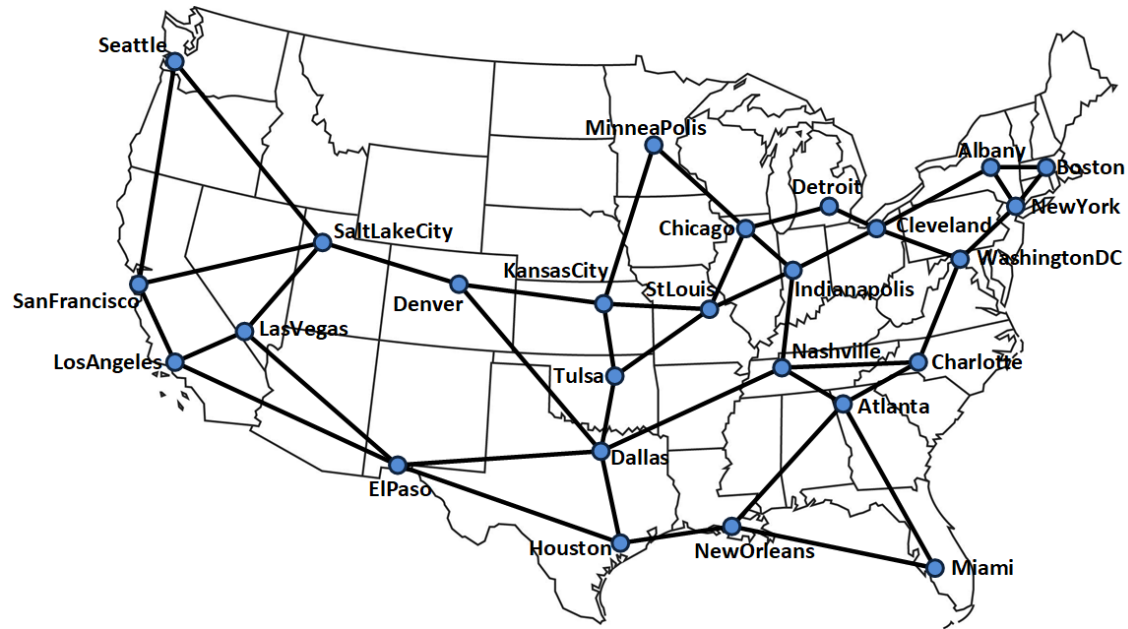


Jakość otrzymanych rozwiązań

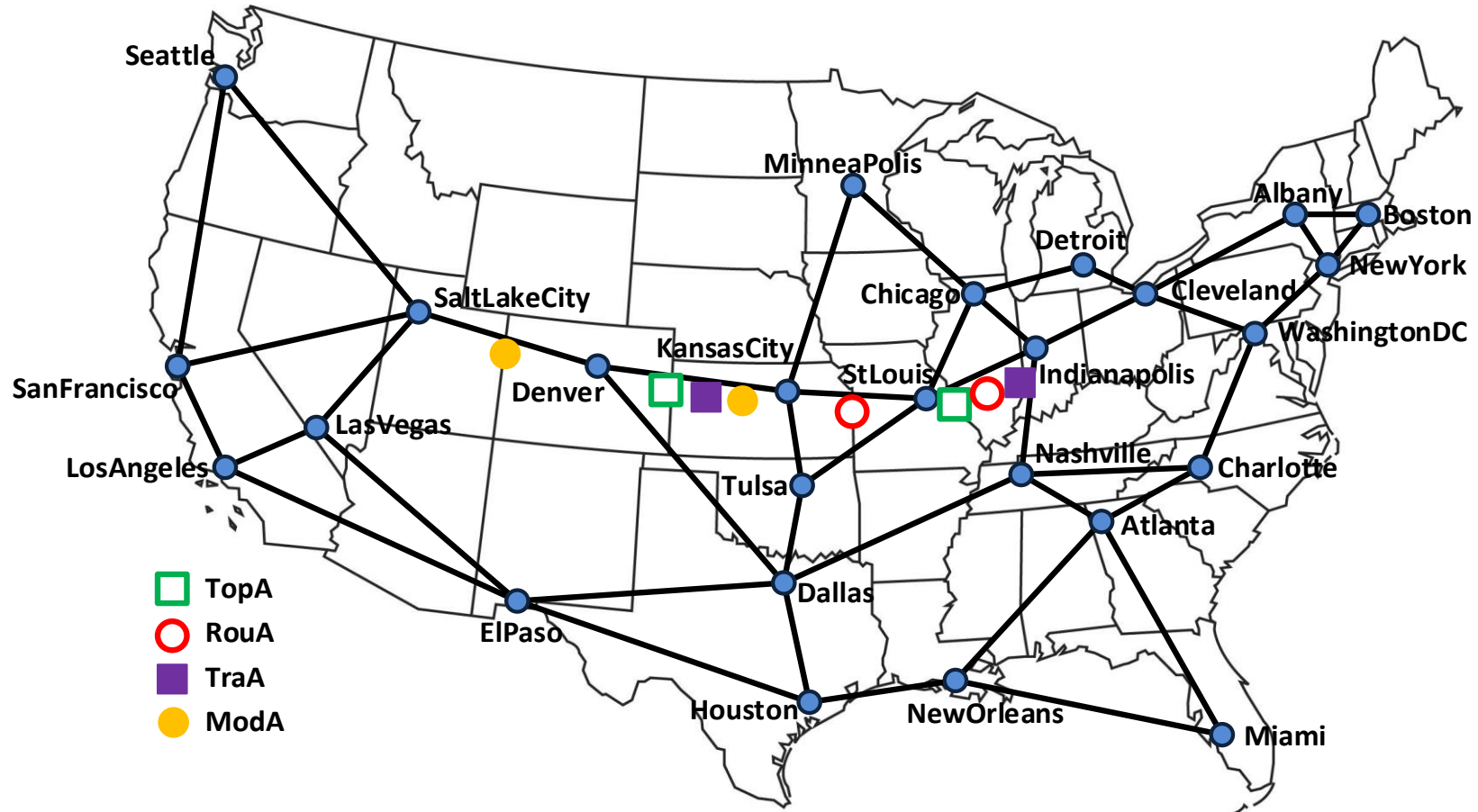
- **Przestrzeń możliwych rozwiązań jest b. duża**, np. dla sieci z 82 łączami, zakładając dopuszczalną liczbę włókien w zakresie 4-8, **liczba wszystkich kombinacji to $5^{82} = 10^{57}$**
- Miarą **jakości rozwiązania** (czyli liczby włókien przypisanych do każdego łącza) jest **oszacowanie sumarycznego ruchu sieciowego**, który można umieścić w sieci zwymiarowanej według uzyskanego rozwiązania
- Do oceny każdego uzyskanego rozwiązania niezbędna jest symulacja **routingu dynamicznego** według wybranego wzorca ruchu **aż do osiągnięcia 1% blokady ruchu**. Jest to czasochłonny zadanie, zajmuje dużo czasu (rzędu minut) dla jednego rozwiązania



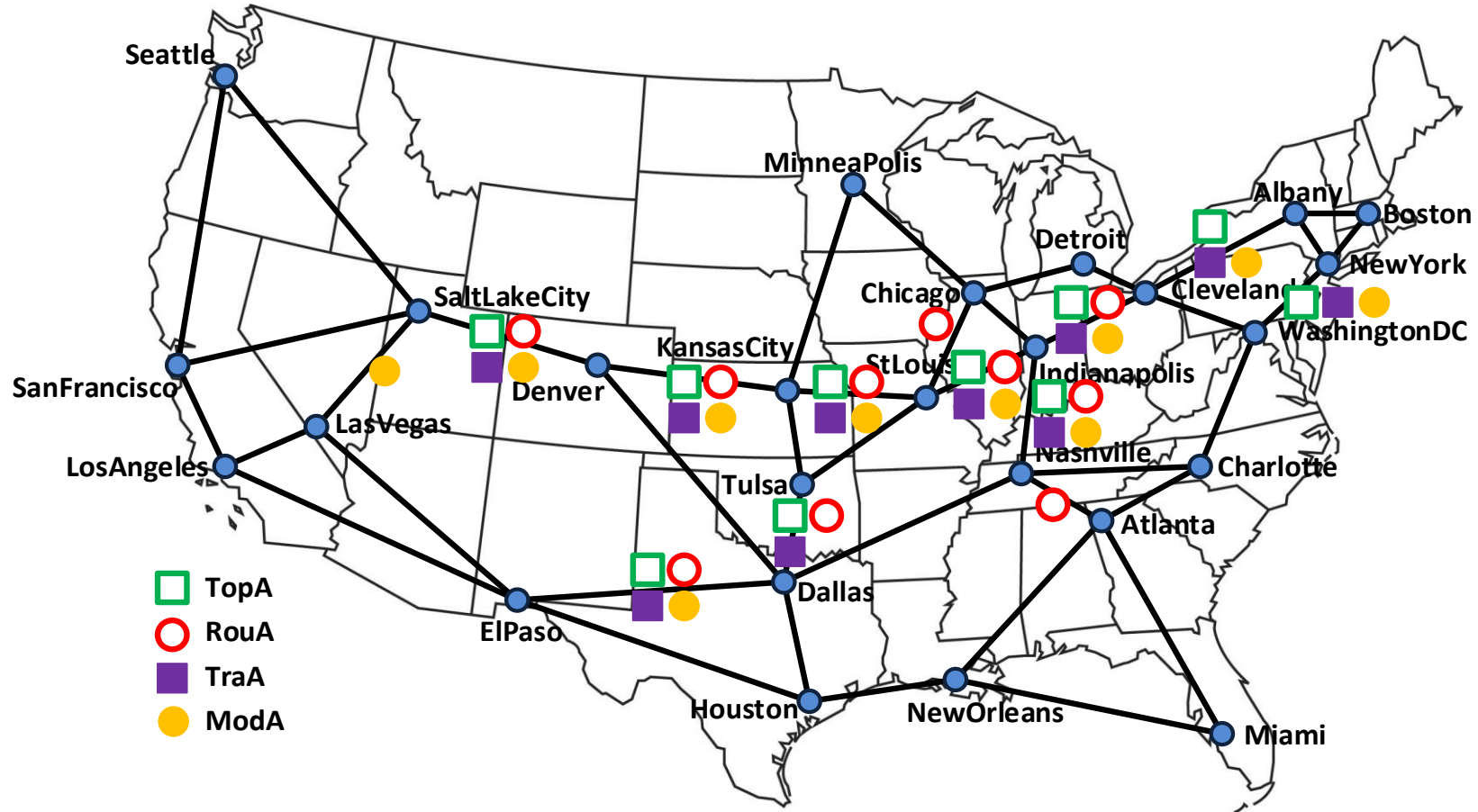
Badane topologie



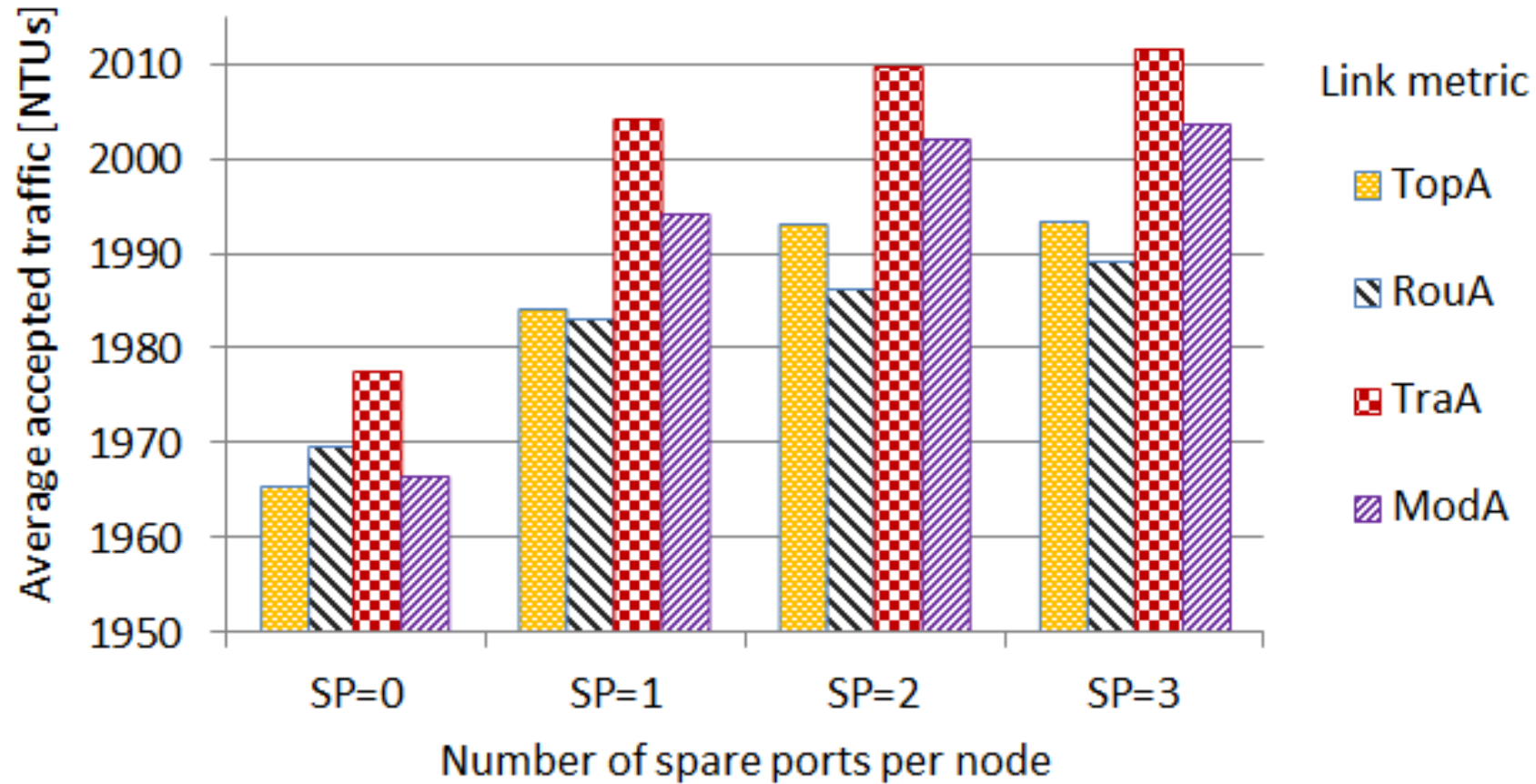
Algorytm heurystyczny, topologia US26, 4 łącza wymiarowane



Algorytm heurystyczny, topologia US26, 20 łączy wymiarowanych

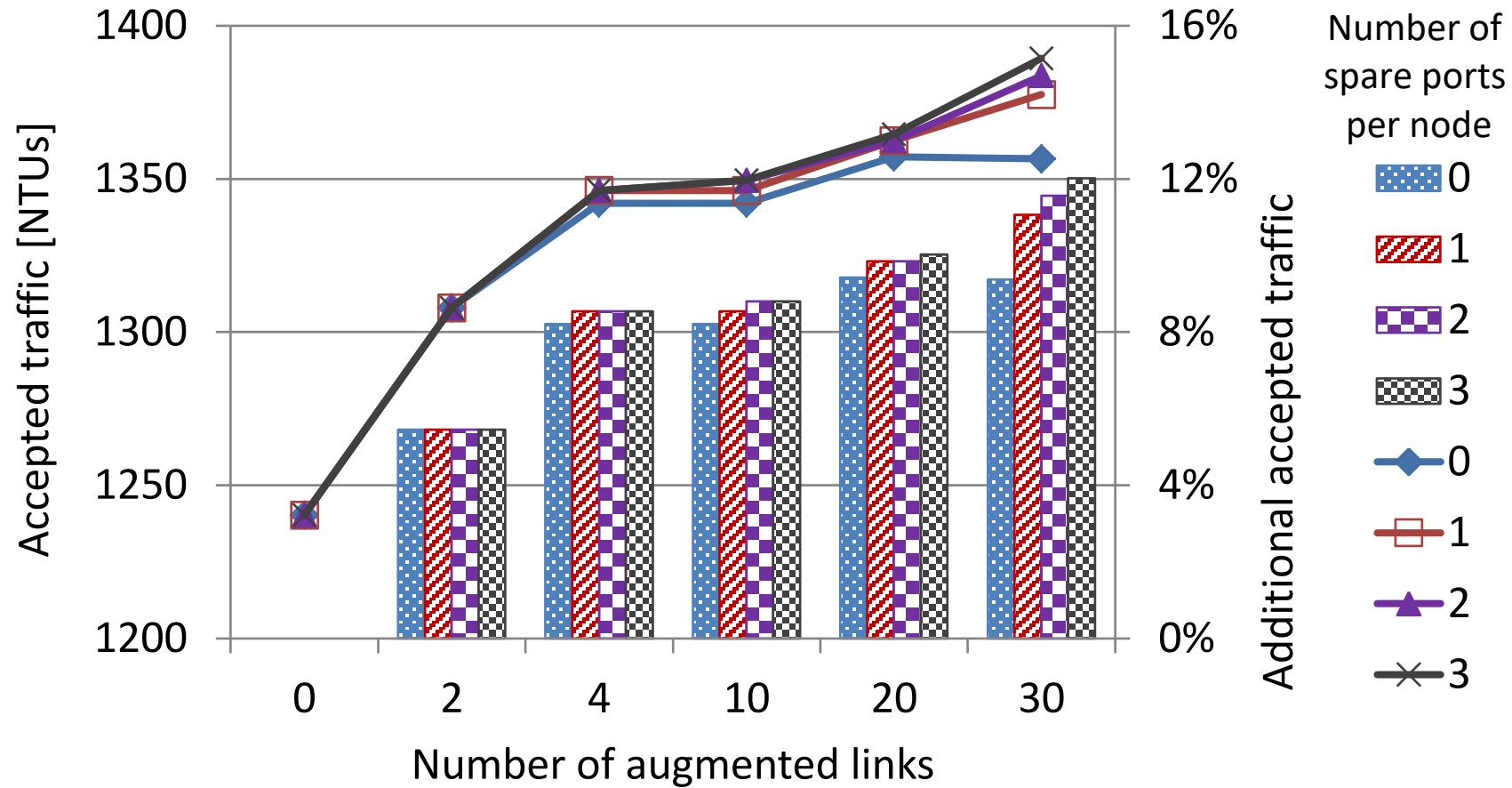


Jakość działania dla różnych metryk łączy, topologia US26

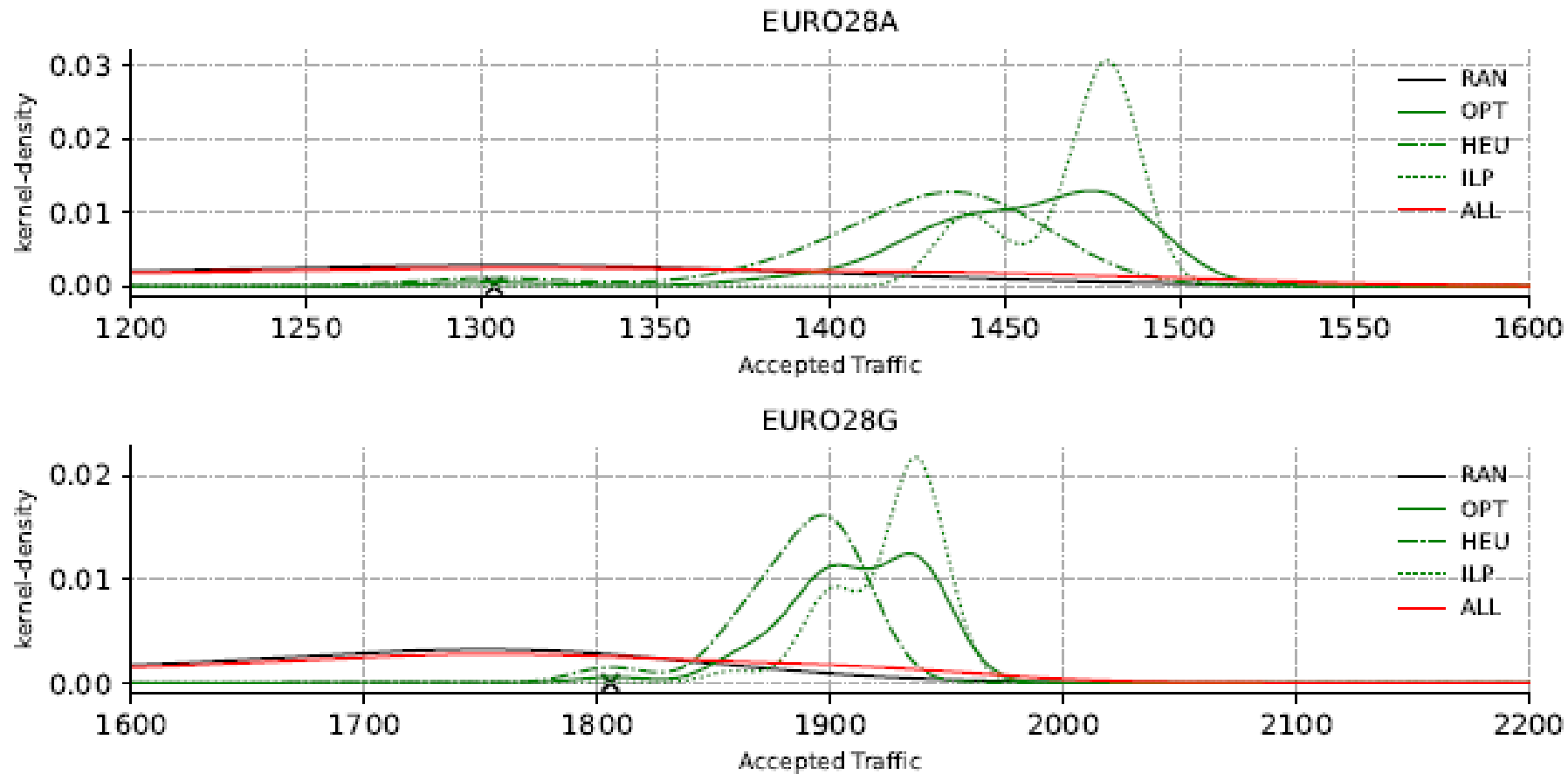




Ruch w sieci (accepted traffic) sieć US26

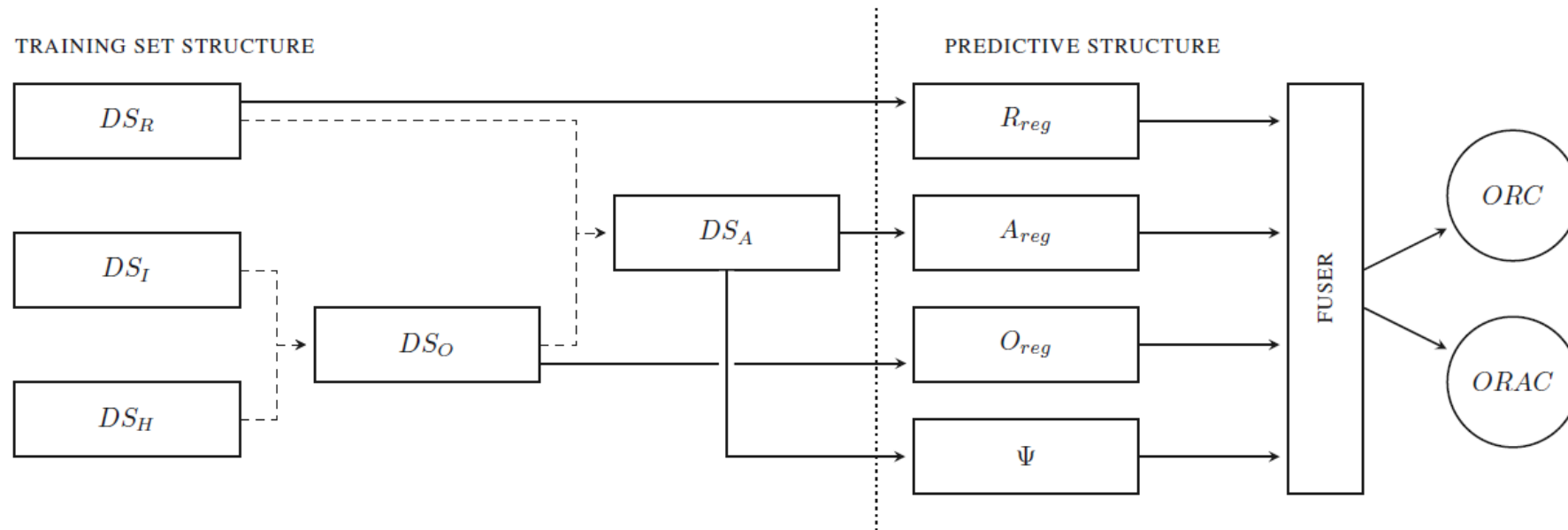


Rozkład wyników algorytmów



- RAN – losowe rozwiązania (300)
- OPT – rozwiązania algorytmów (48)
- HEU – rozwiązania algorytmu heurystycznego (16)
- ILP – rozwiązania modelu ILP (32)
- ALL – wszystkie rozwiązania (348)

Architektura regresora hybrydowego



- A_{reg} - zbudowany na wszystkich wygenerowanych rozwiązaniach (348 przykładów)
- R_{reg} - zbudowany wyłącznie na losowo wygenerowanych rozwiązaniach (300 przykładów)
- O_{reg} - zbudowany wyłącznie na rozwiązaniach z algorytmu heurystycznego i ILP (48 przykładów)
- Zastosowano regresor Multi-Layer Perceptron (MLP)

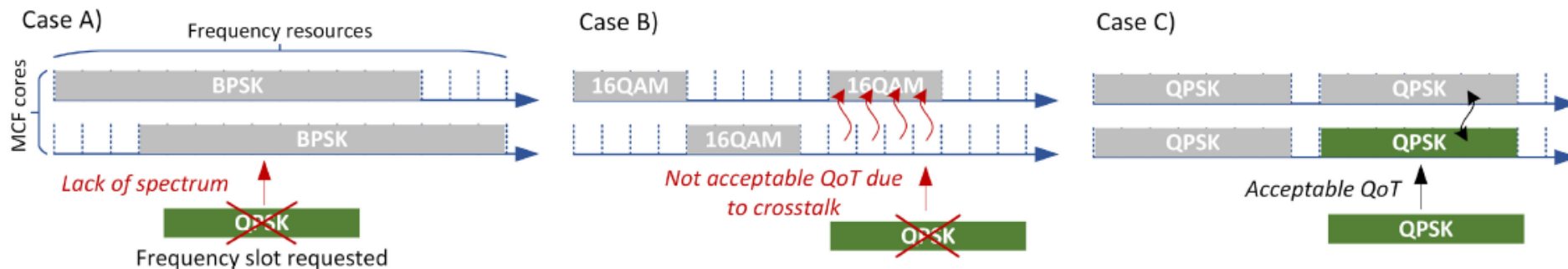
Wyniki – sumaryczny ruch w sieci

LISTED SOLUTIONS				SUPERVISED MODELS				
BASE	<i>DS_R</i>	<i>DS_H</i>	<i>DS_I</i>	HYBRID		SINGLE		
	1	2	3	ARC	AROC	A-REG	R-REG	O-REG
				4	5	6	7	8
EURO28 A								
1304	1284	1426	1461	1484	1486	1482	1481	1482
	—	1	1, 2	1, 2, 3	1, 2, 3	1, 2	1, 2	1, 2
EURO28 G								
1806	1748	1889	1918	1937	1936	1919	1924	1927
	—	1	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
US26 A								
1240	1176	1350	1362	1366	1360	1357	1335	1360
	—	1	1	1, 2	1	1	1	1
US26 G								
1919	1842	1987	2006	2000	2006	1998	1995	2002
	—	1	1	1, 2	1, 2	1	1	1

- D_{SR} – losowe rozwiązania
- D_H – algorytm heurystyczny
- D_H – model ILP
- ARC – model hybrydowy
- AROC – model hybrydowy
- A-REG – model pojedynczy, wszystkie rozwiązania
- R-REG – model pojedynczy, rozwiązania losowe
- O-REG – model pojedynczy, rozwiązania algorytmu heurystycznego i ILP

Dynamiczny routing w sieciach SS-FON z uwzględnieniem przesłuchów (crosstalk)

- [M. Klinkowski, P. Ksieniewicz, M. Jaworski, G. Zalewski, K. Walkowiak, *Machine learning assisted optimization of dynamic crosstalk-aware spectrally-spatially flexible optical networks*, Journal of Lightwave Technology 2020]
- [P. Ksieniewicz, R. Goścień, M. Klinkowski, K. Walkowiak, *Pattern Recognition Model to Aid the Optimization of Dynamic Spectrally-Spatially Flexible Optical Networks*, ICCS 2020]
- Rozważana jest sieć SS-FON stosująca **światłowody wielordzeniowe typu MCF** (ang. *multi-core fibers*), w których występują przesłuchy (ang. *crosstalk*) między poszczególnymi rdzeniami wpływające na jakość transmisji QoT (ang. *quality of transmission*)
- Charakter przesłuchów oraz **dynamicznie zmieniająca się sytuacja w sieci** (ruch dynamiczny) uniemożliwia oszacowanie wpływu QoT, **dlatego zastosowano w tym celu ML**





Dynamiczny ruting w sieciach SS-FON z uwzględnieniem przesłuchów (crosstalk)

- Dostępne są następujące **formaty modulacji**: BPSK, QPSK, 8QAM, and 16QAM,
- Dla każdego formatu modulacji określany jest **parametr MLD** (Modulation format Distance Limit), który jest używany do wyboru zastosowanego **formatu modulacji adekwatnie do długości wybranej ścieżki routingu**
- Dla wybranych wartości MDL symulowany jest ruch dynamiczny w celu oszacowania **wartości blokady ruchu BBP** (ang. *bandwidth blocking probability*) **uwzględniając przesłuchy**

Step 1. Generate at random 1000 feasible MDLs samples for modulation formats QPSK, 8-QAM, and 16-QAM

$$MDL_1 = [MDL_1^{QPSK}, MDL_1^{8-QAM}, MDL_1^{16-QAM}]$$

$$MDL_2 = [MDL_2^{QPSK}, MDL_2^{8-QAM}, MDL_2^{16-QAM}]$$

...

$$MDL_n = [MDL_n^{QPSK}, MDL_n^{8-QAM}, MDL_n^{16-QAM}]$$

Step 2. Run XT-aware dynamic routing algorithm to simulate the SS-FON performance in terms of BBP for each sample MDL_i

$$MDL_1 \rightarrow BBP_1$$

$$MDL_2 \rightarrow BBP_2$$

...

$$MDL_n \rightarrow BBP_n$$

Step 3. Run regression algorithm to predict the expected BBP value and find the best MDL combination using the grid search method

$$[MDL_1, BBP_1]$$

$$[MDL_2, BBP_2] \rightarrow MDL_{BEST}$$

...

$$[MDL_n, BBP_n]$$

Step 4. Run XT-aware dynamic routing algorithm with MF selection according to MDL_{BEST} to assess the SS-FON performance in terms of BBP

$$MDL_{BEST} \rightarrow BBP$$

Badania

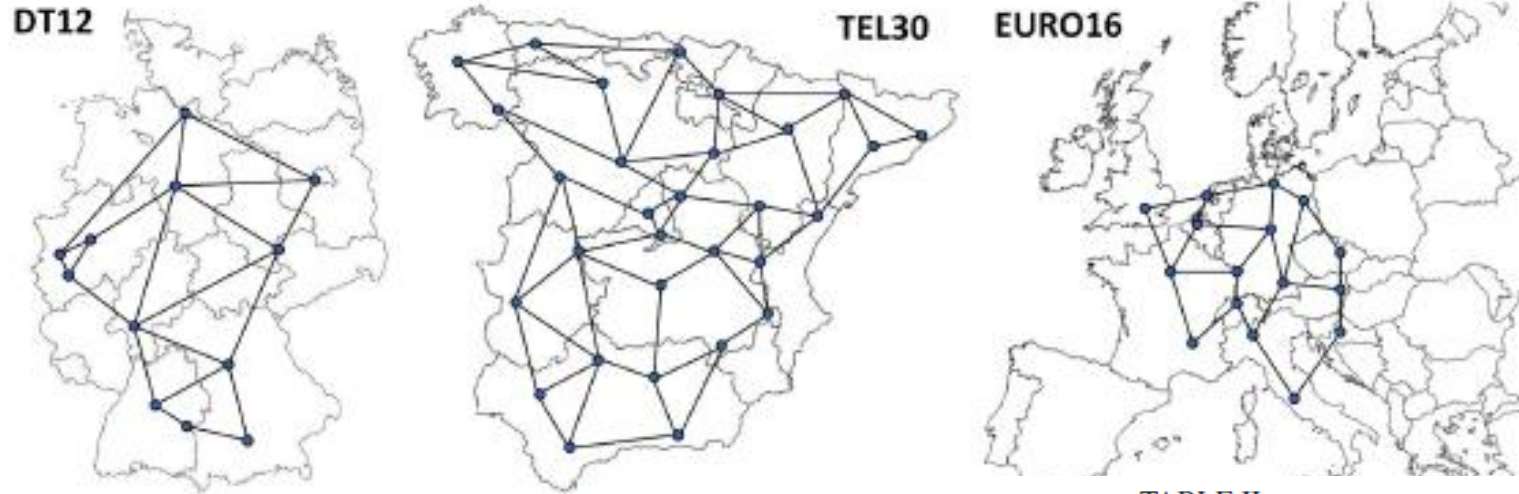


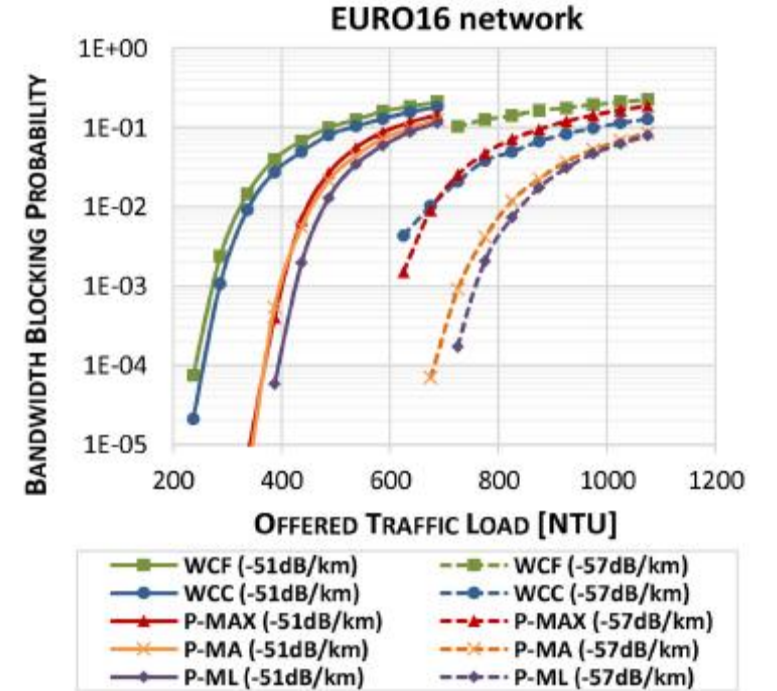
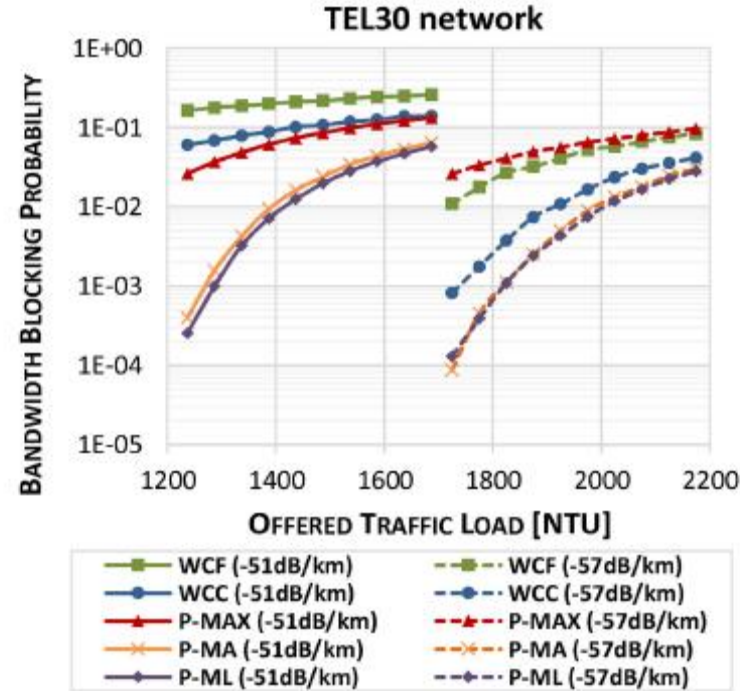
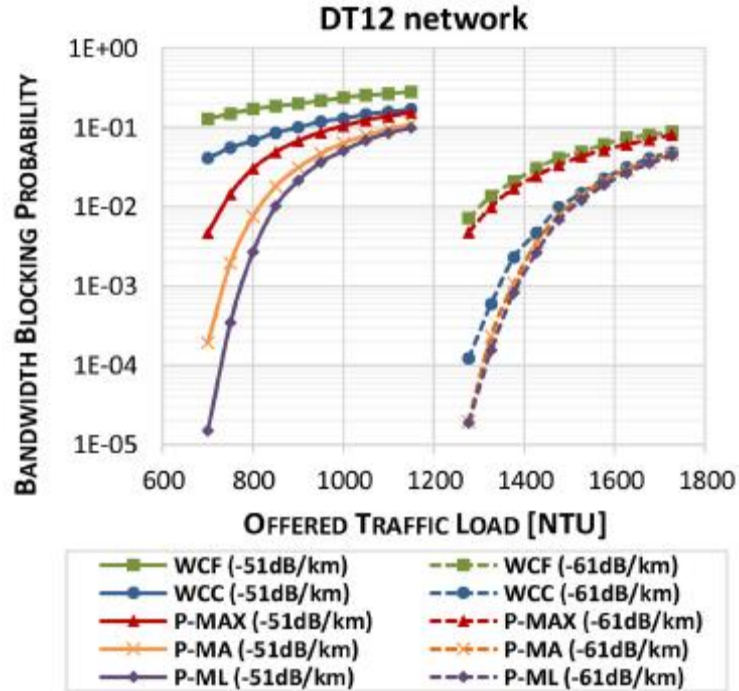
TABLE I
TRANSMISSION SYSTEM PARAMETERS

Attenuation coefficient (α) [dB/km]	0.21
Chromatic dispersion (CD) [ps/nm/km]	16.7
Nonlinear index (n_2) [W^2/m^{-1}]	2.3×10^{-20}
Effective area (A_{eff}) [μm^2]	80
EDFA noise figure (NF) [dB]	5
Required SNR for BPSK [dB]	6.8
Required SNR for QPSK [dB]	9.8
Required SNR for 8QAM [dB]	14.3
Required SNR for 16QAM [dB]	16.5

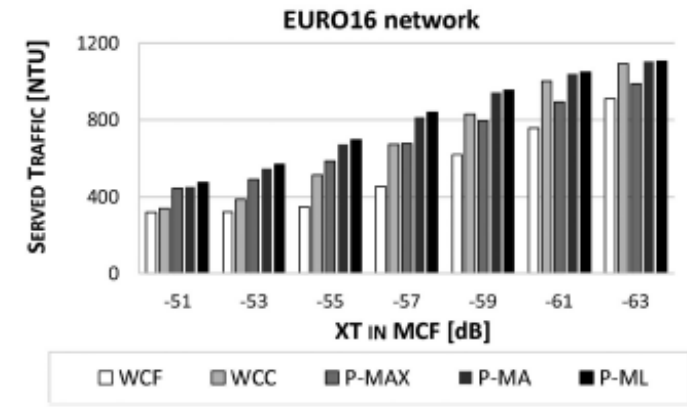
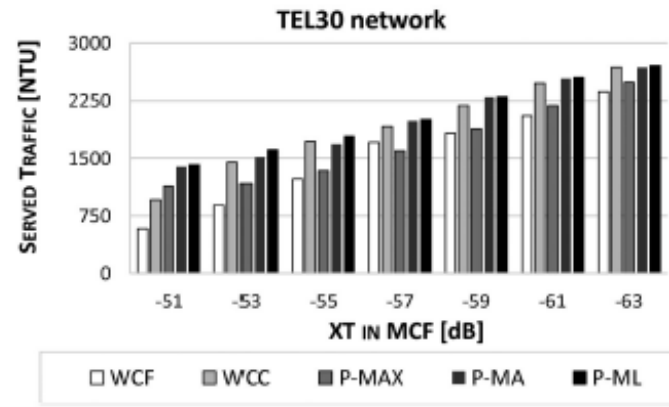
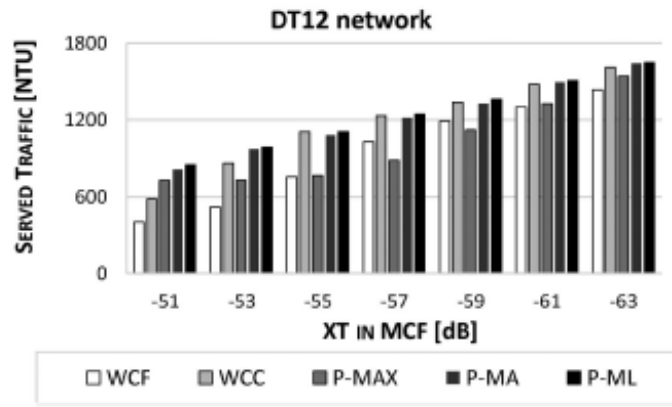
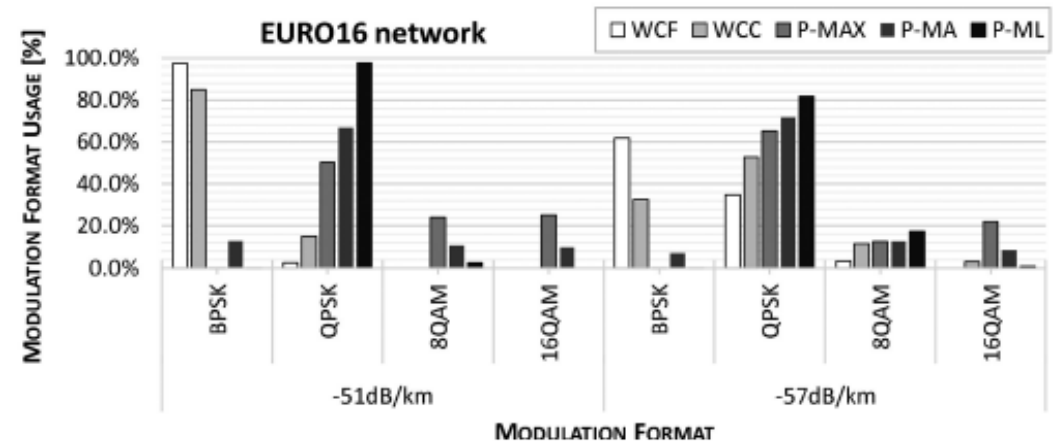
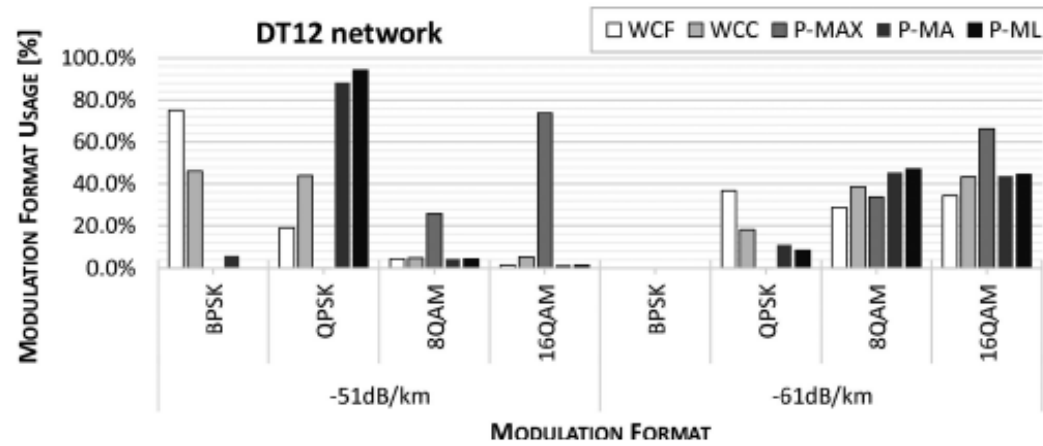
TABLE II
WORST-CASE MDLS (IN KILOMETERS) IN THE MULTI-CORE FIBER
CONSIDERED FOR SELECTED VALUES OF XT

MCF core	XT^{REF} [dB/km]	BPSK	QPSK	8QAM	16QAM
Central core	-51	639	319	113	68
	-57	2039	1020	360	216
	-61	3645	1829	646	387
Outer cores	-51	1180	590	208	125
	-57	3225	1612	569	341
	-61	4961	2481	876	525
All cores	none	7706	3853	1360	816

Wyniki



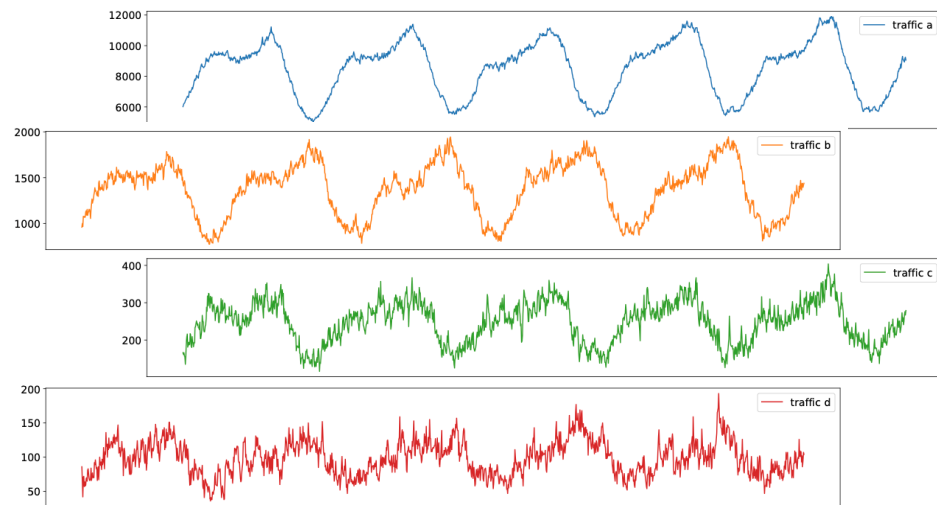
Wyniki



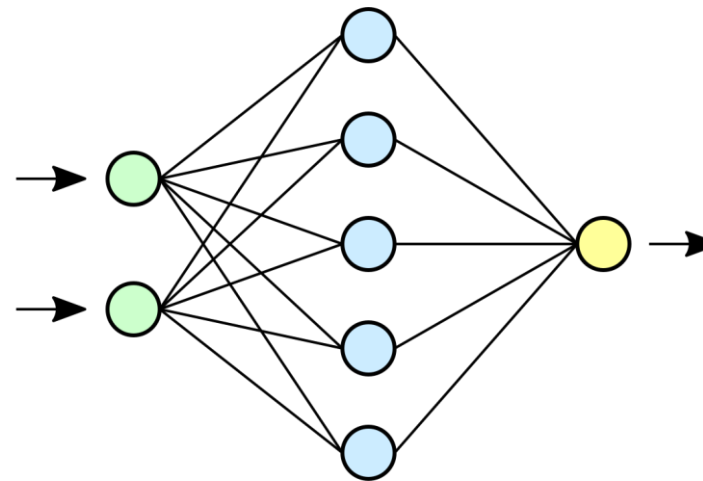


Pre dykcja ruchu w sieciach optycznych

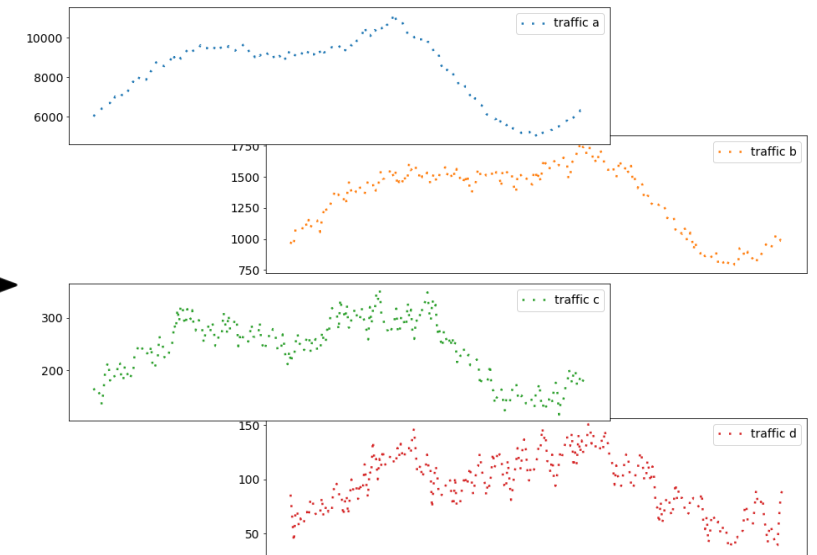
Predykcja ruchu w sieci



Dane historyczne

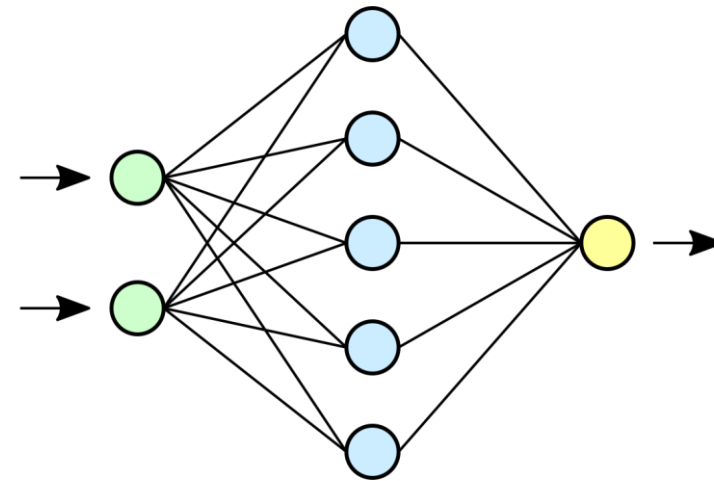
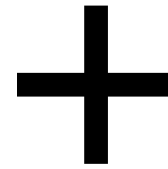
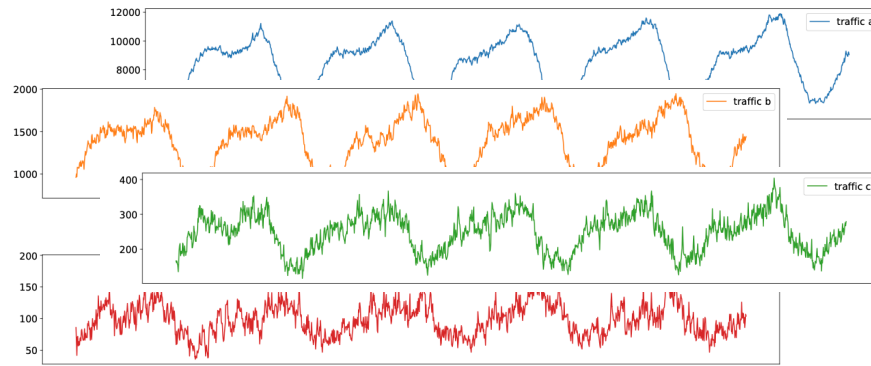


Model predykcji

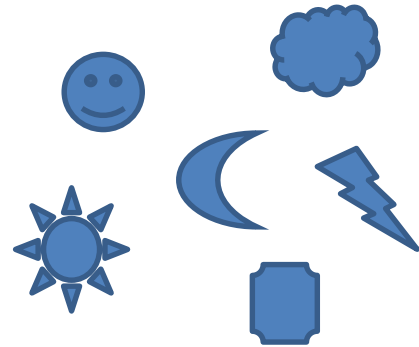


Wynik predykcji

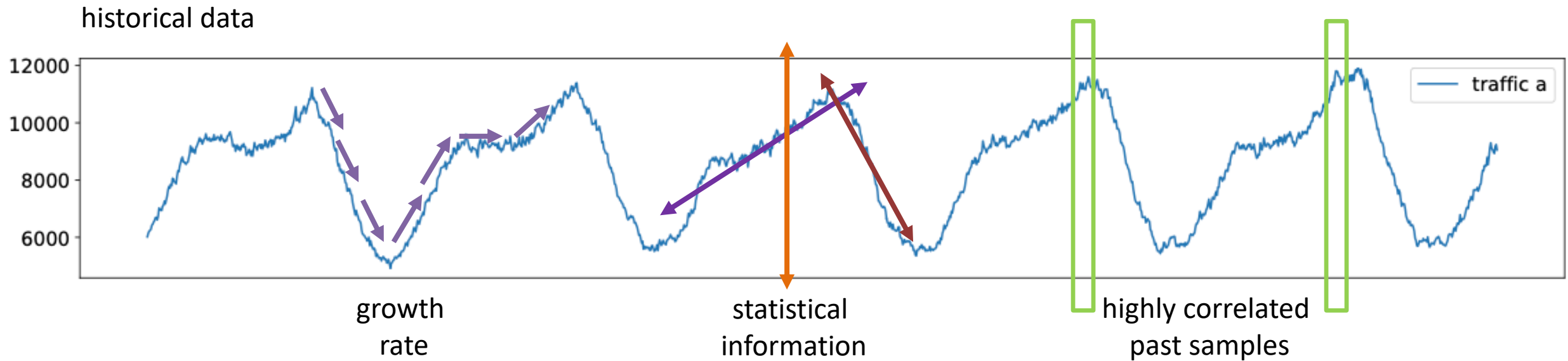
Dodatkowe cechy



Dodatkowe
cechy



Tworzenie dodatkowych cech



... and other possibilities!

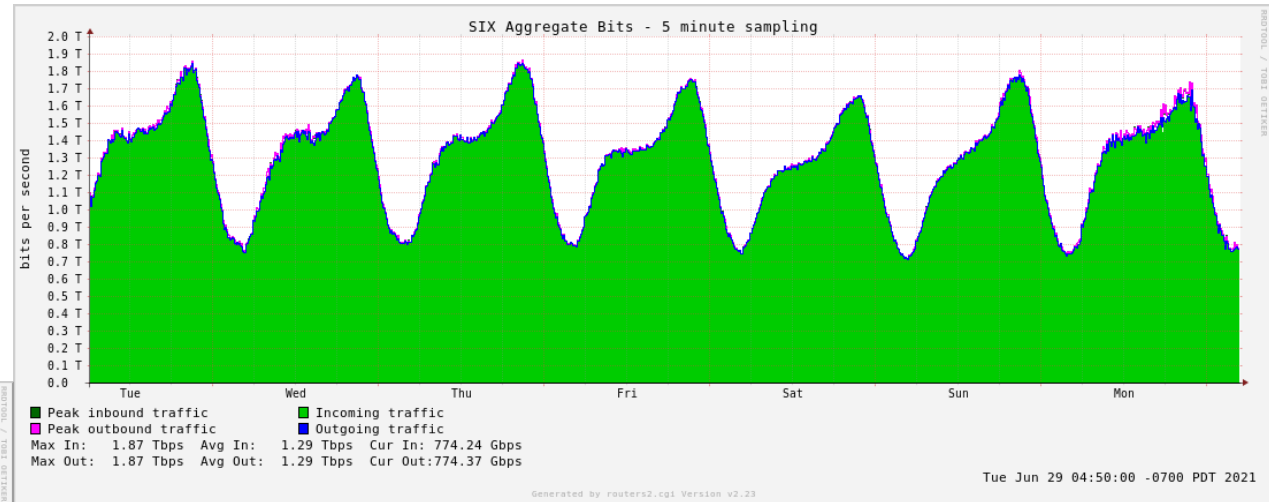
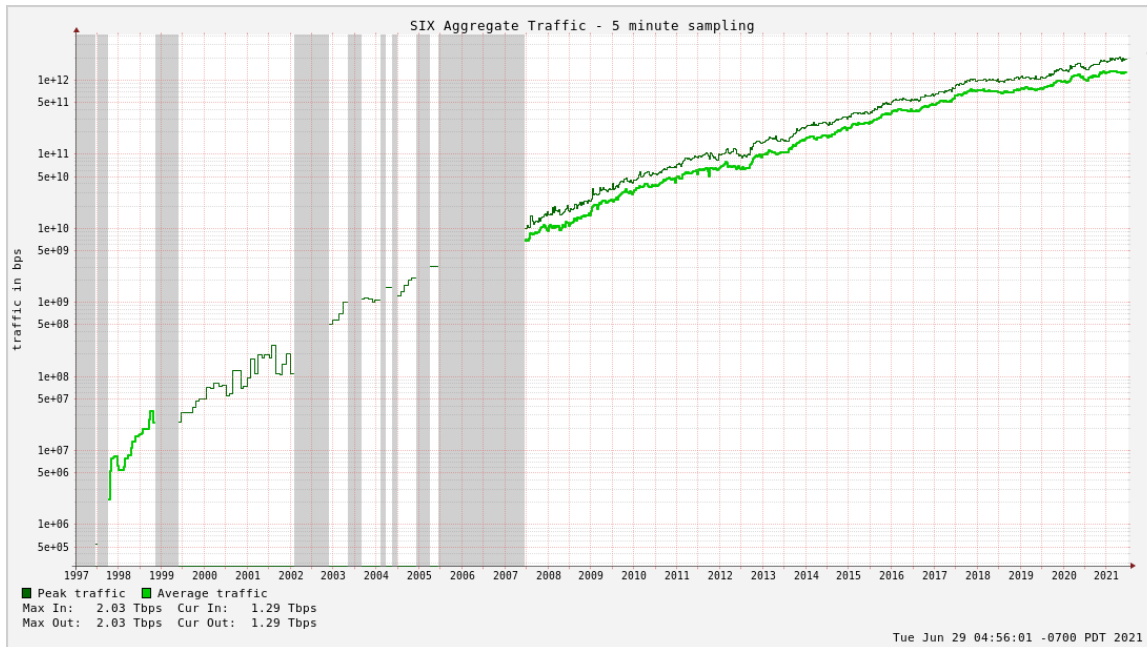
Predykcja ruchu dla różnych rodzajów aplikacji/usług

- [A. Knapińska, P. Lechowicz, K. Walkowiak, *Machine-Learning Based Prediction of Multiple Types of Network Traffic*, ICCS 2021]
- [A. Knapińska, P. Lechowicz, W. Węgier, K. Walkowiak, *Long-term prediction of multiple types of time-varying network traffic using chunk-based ensemble learning*, submitted to Applied Soft Computing]
- Predykcja ruchu z podziałem na **różne typy ruchu** reprezentujące **różne aplikacje lub usługi**
- Możliwość **wykorzystania danych o innych typach ruchu** w celu poprawienia jakości predykcji (zakładając korelację)
- Dodatkowa możliwość zastosowania **komitetów (ang. *ensemble*) klasyfikatorów** oraz metod związanych ze **strumieniami danych (ang. *data stream*)**
- Dwa rodzaje predykcji typu:
 - **Short-term** (tylko następny przedział czasowy, np. 5 minut)
 - **Long-term** (dłuższy okres, np. 24h)
- **Wynik predykcji** może być wykorzystany w procesie w **optymalizacji sieci**



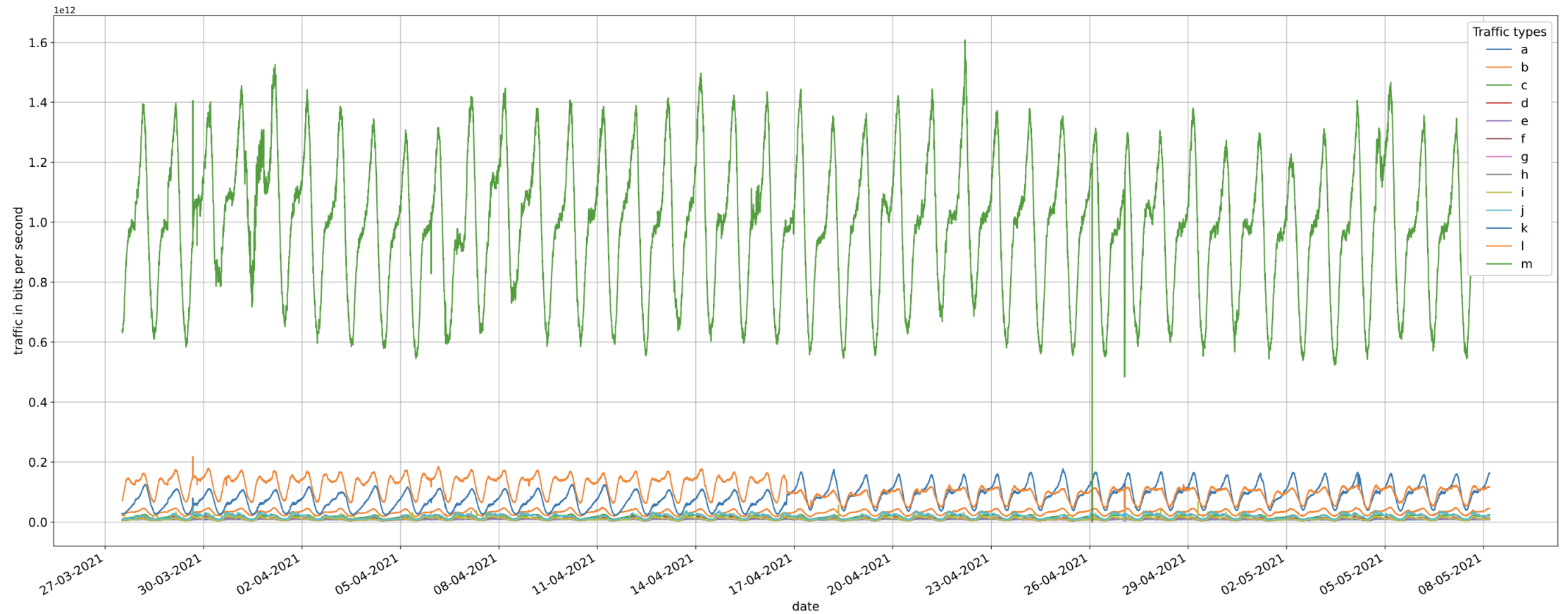
Dane

Seattle Internet Exchange Point
<https://www.seattleix.net/>

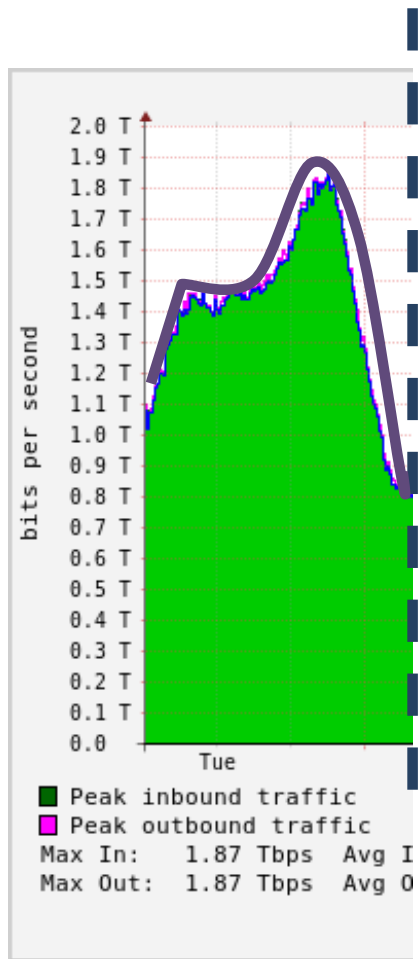




Różne rodzaje ruchu (różne wielkości ramek)

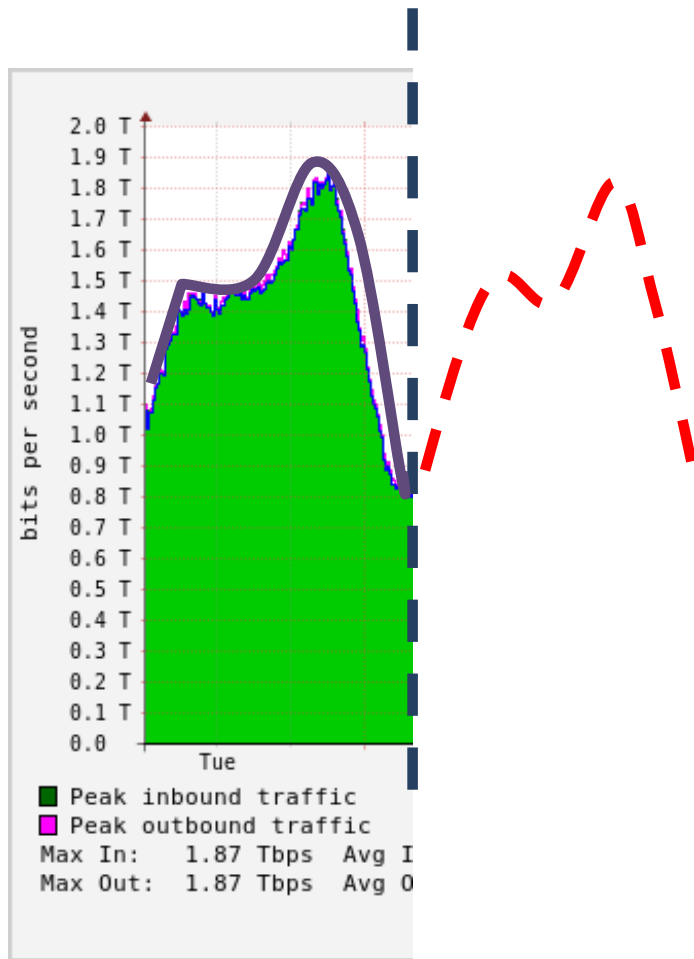


Chunk-based ensemble learning



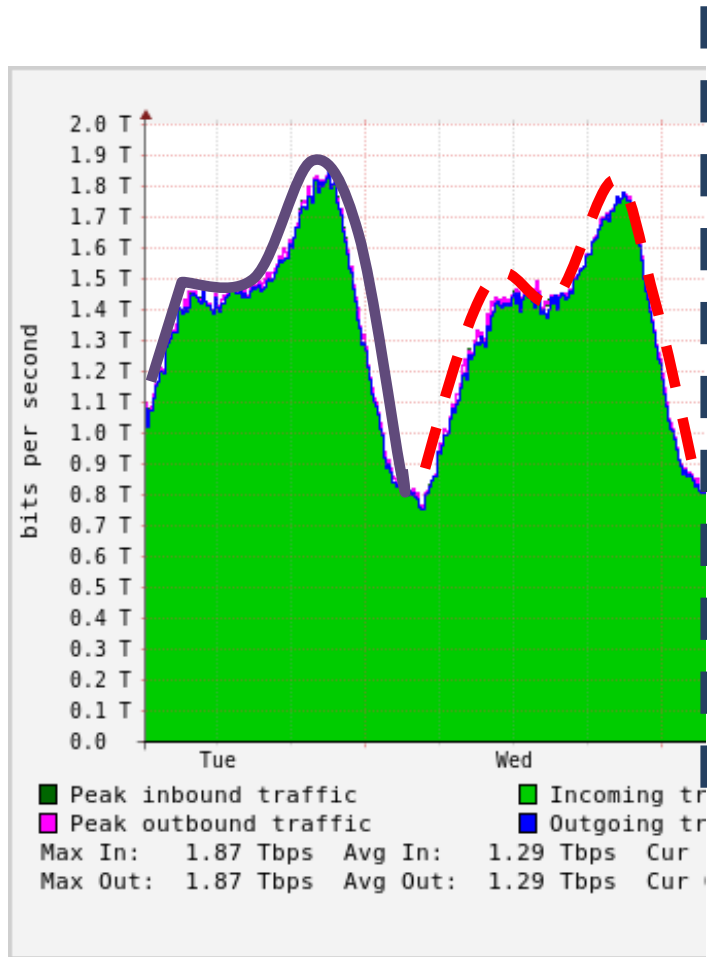
Trenowanie pierwszego klasyfikatora (członka komitetu)

Chunk-based ensemble learning



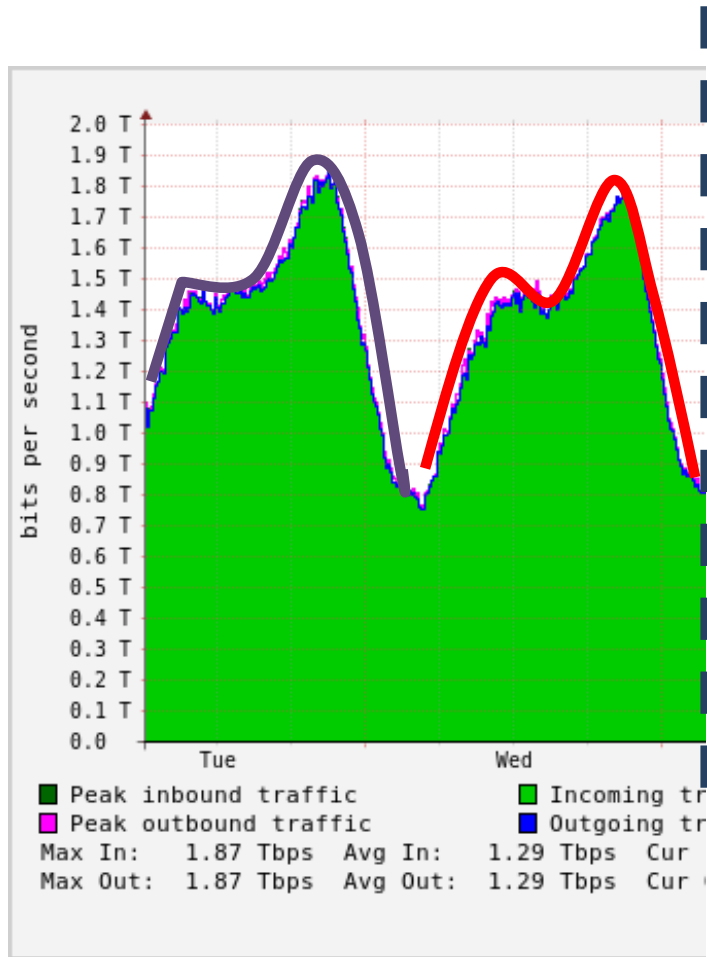
Pierwsza predykcja

Chunk-based ensemble learning



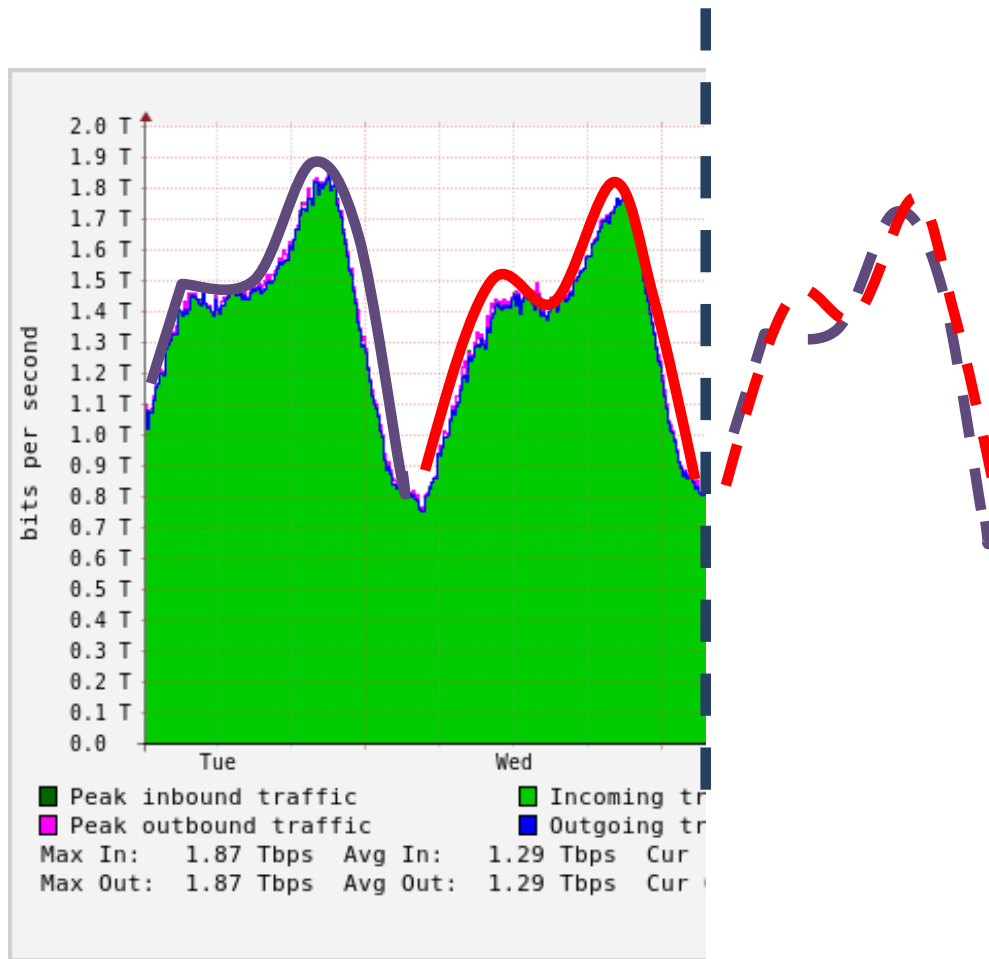
Test-then-train (obliczana jest jakość predykcji (błąd))

Chunk-based ensemble learning



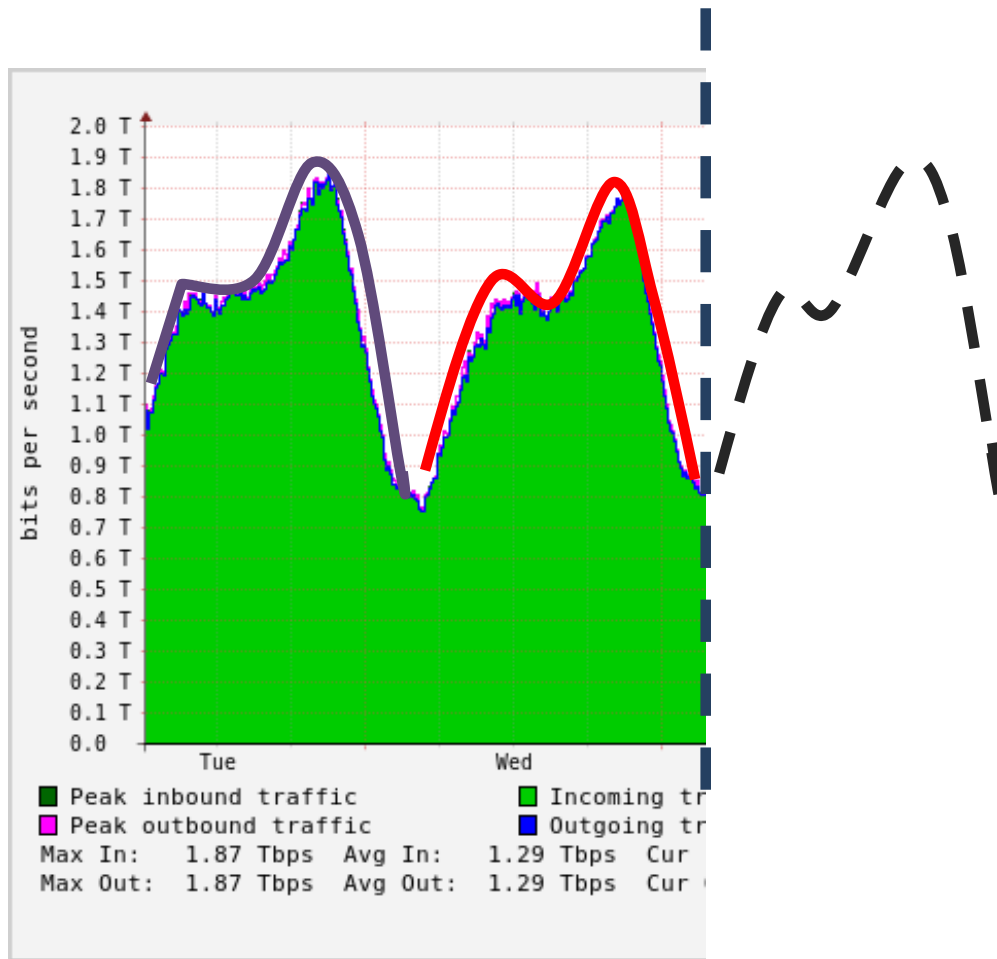
Test-then-train (trenowany jest kolejny klasyfikator w komitecie)

Chunk-based ensemble learning



Każdy członek komitetu wykonuje predykcję

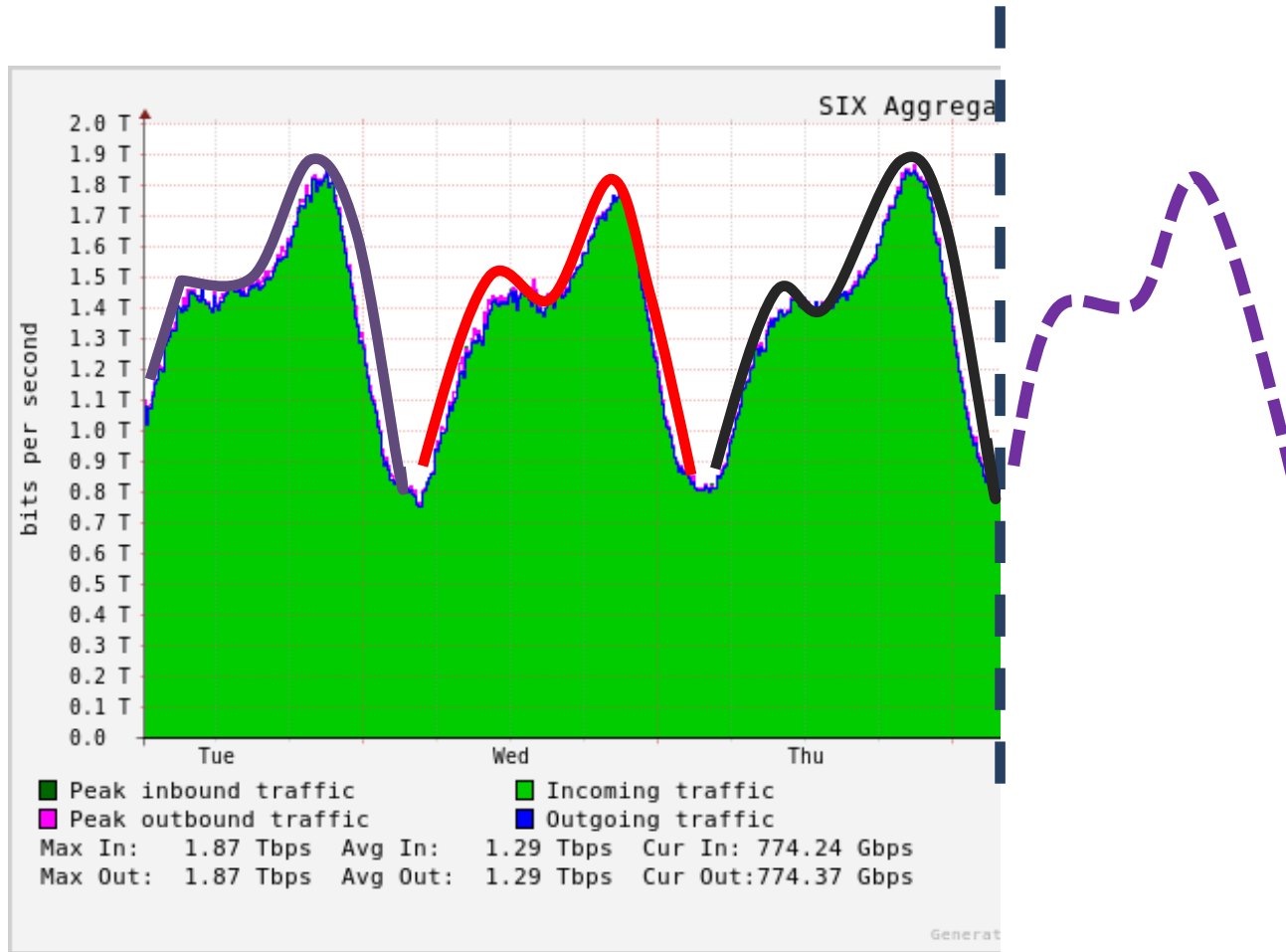
Chunk-based ensemble learning



Wynik predykcji jest liczony jako średnia
(możliwe inne metody)

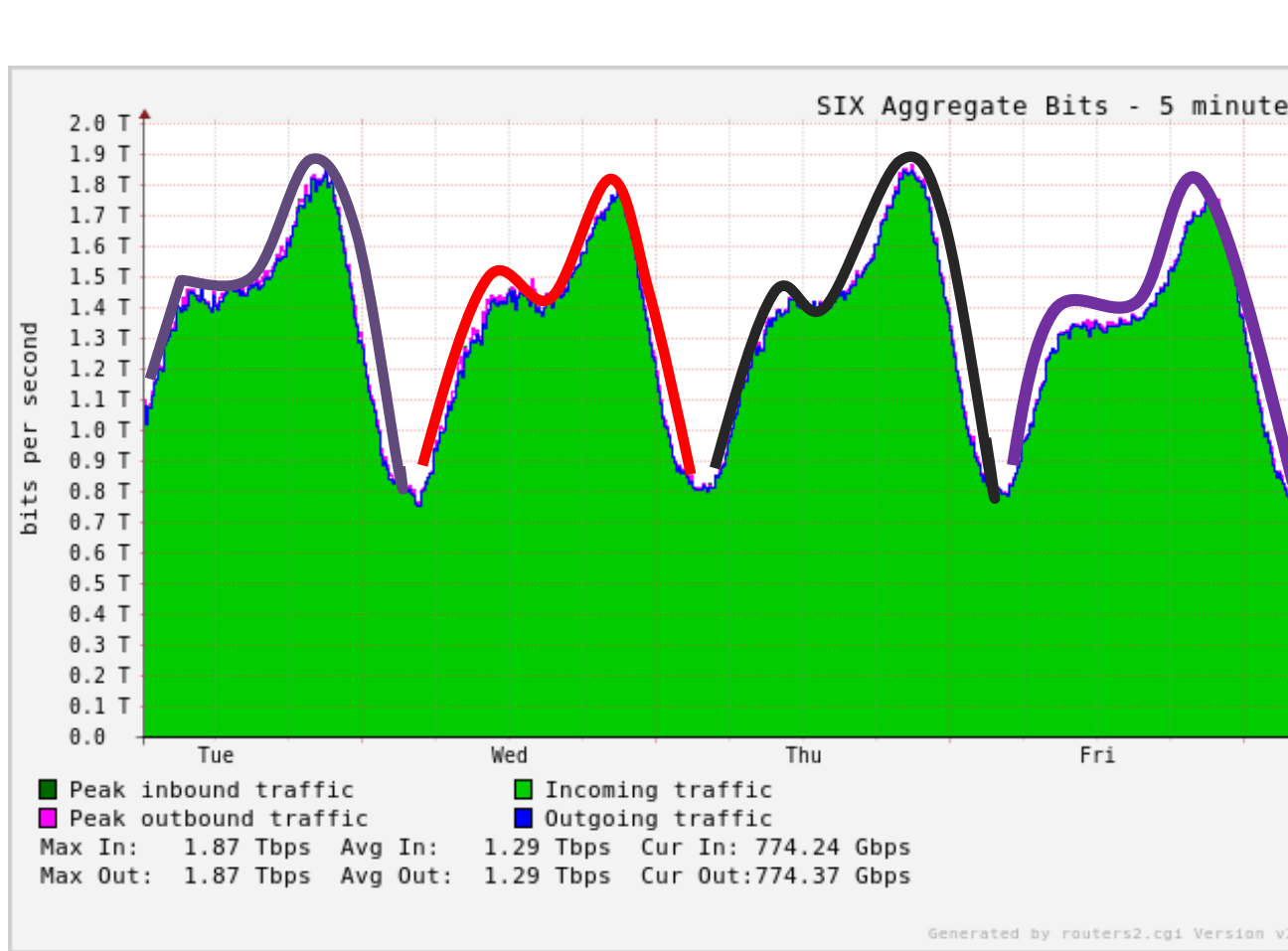


Chunk-based ensemble learning



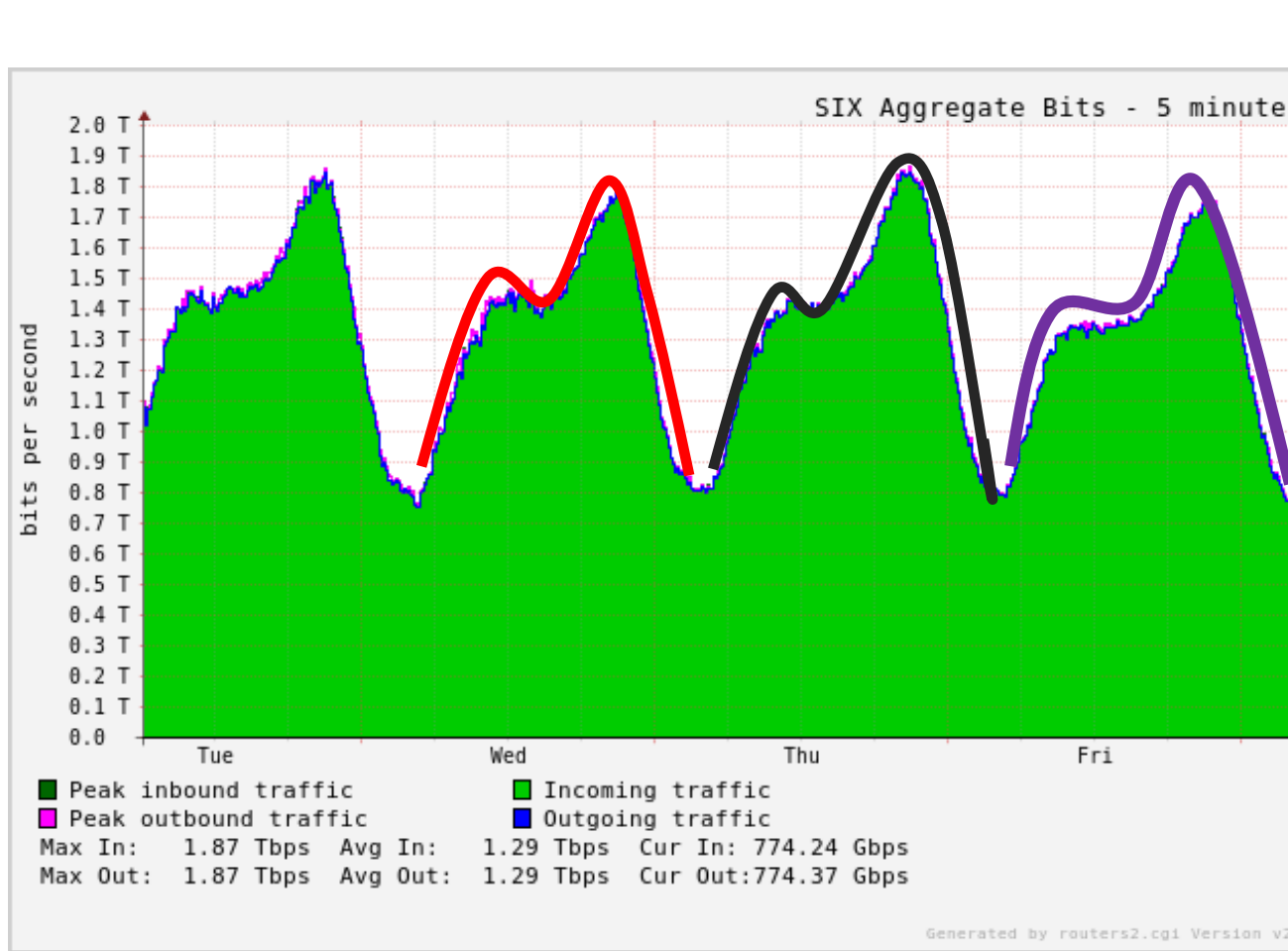


Chunk-based ensemble learning



Za dużo klasyfikatorów (członków komitetu)

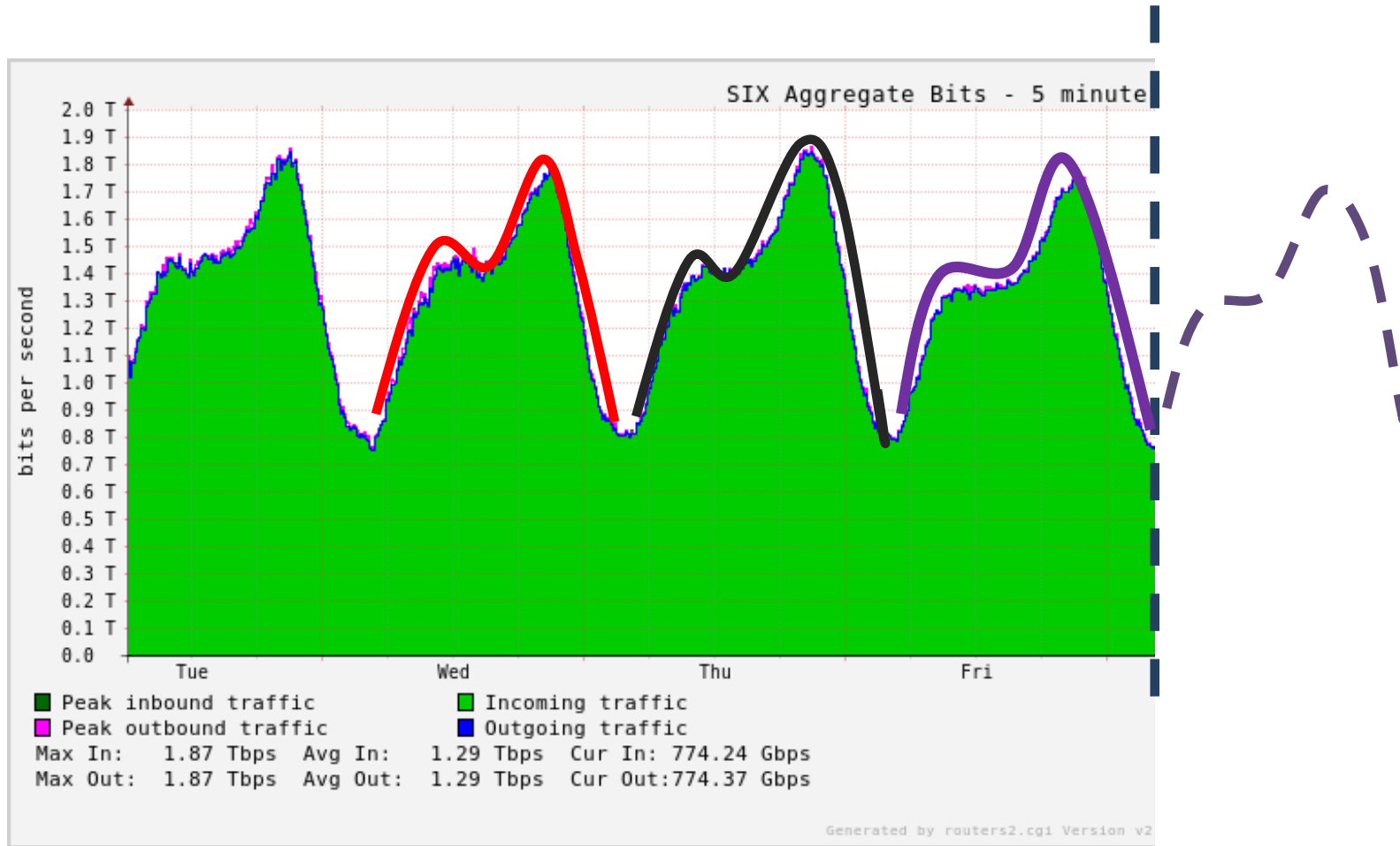
Chunk-based ensemble learning



Usuwanie jednego klasyfikatora z komitetu
(wiele możliwych metod)

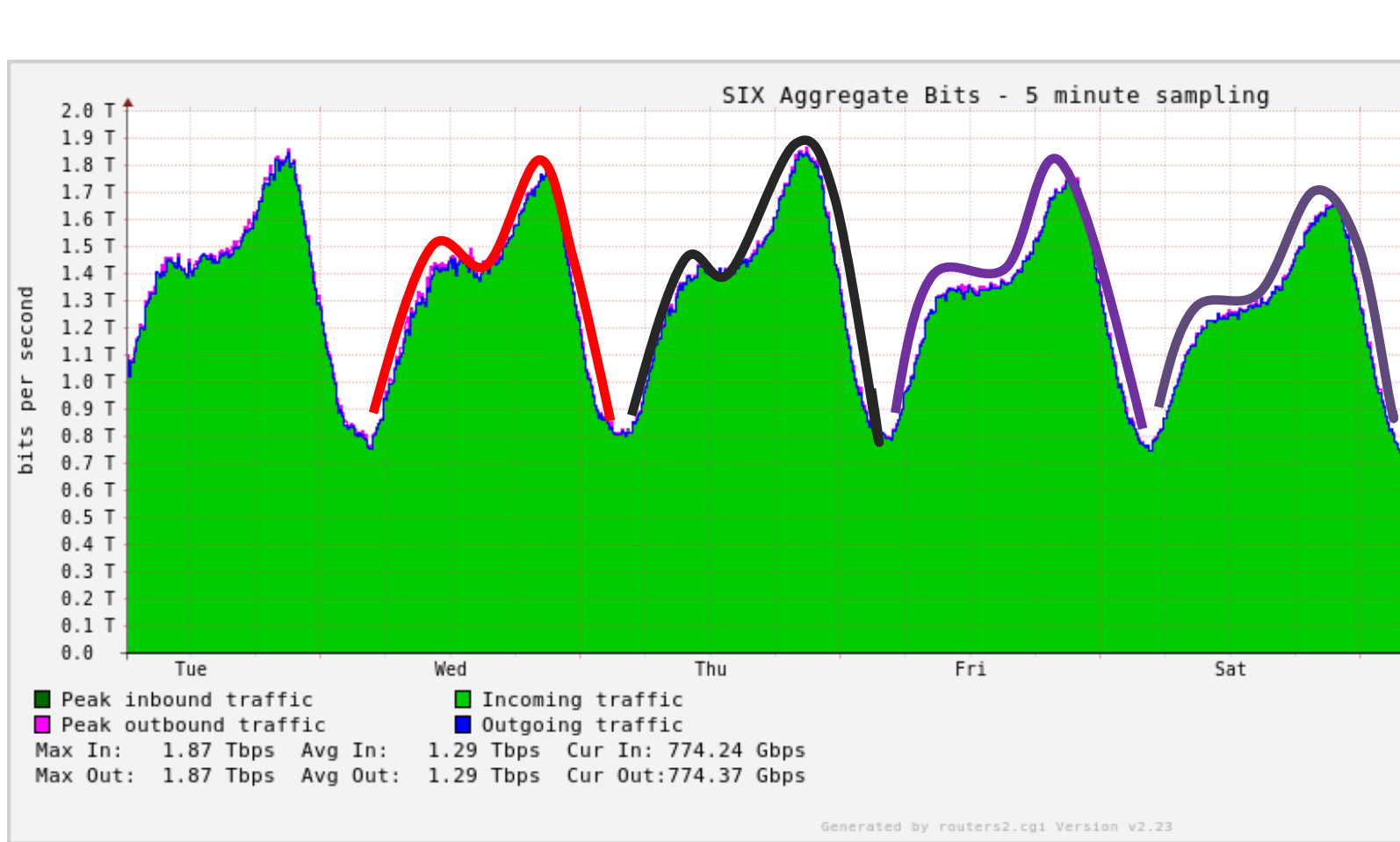


Chunk-based ensemble learning



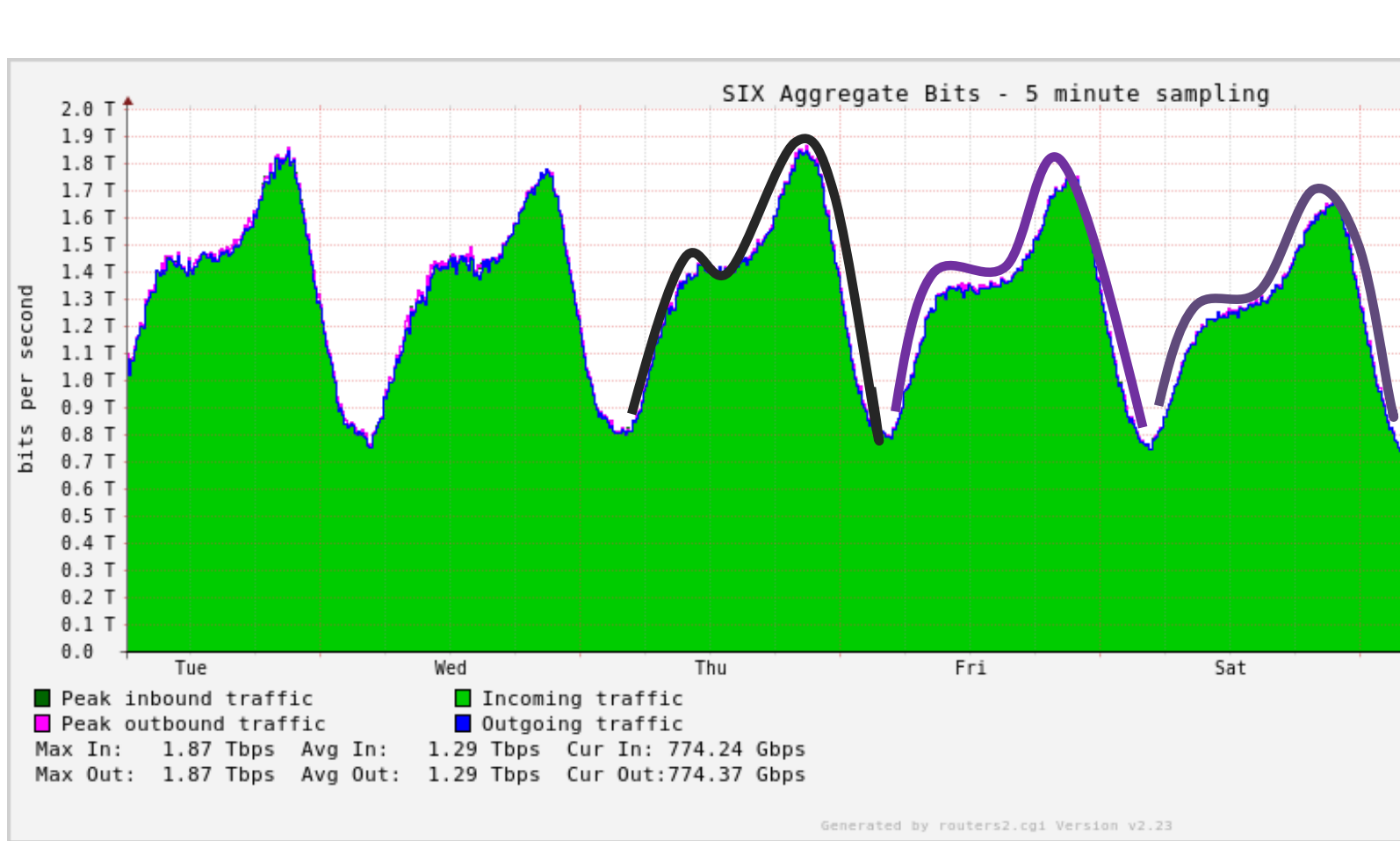


Chunk-based ensemble learning

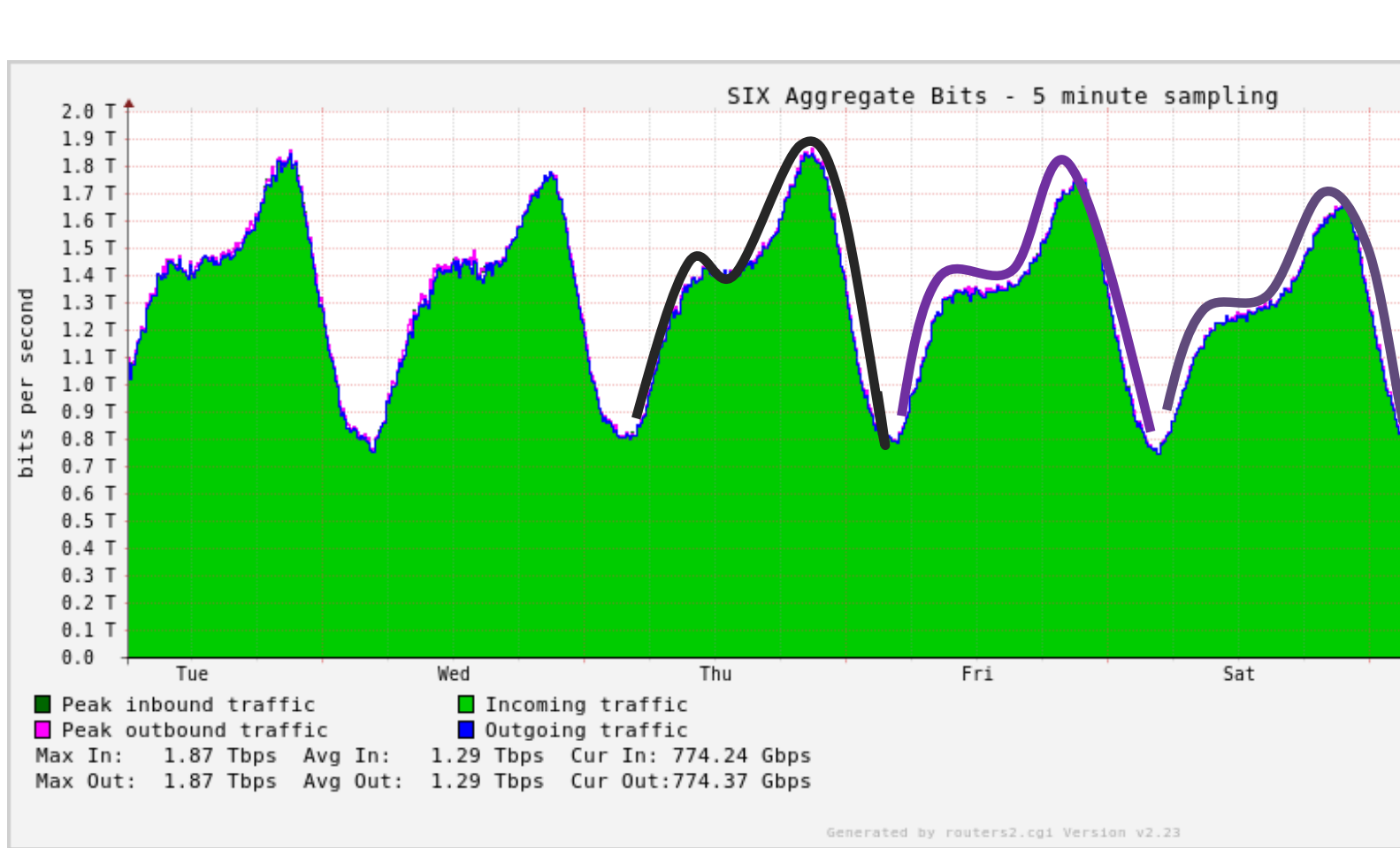




Chunk-based ensemble learning

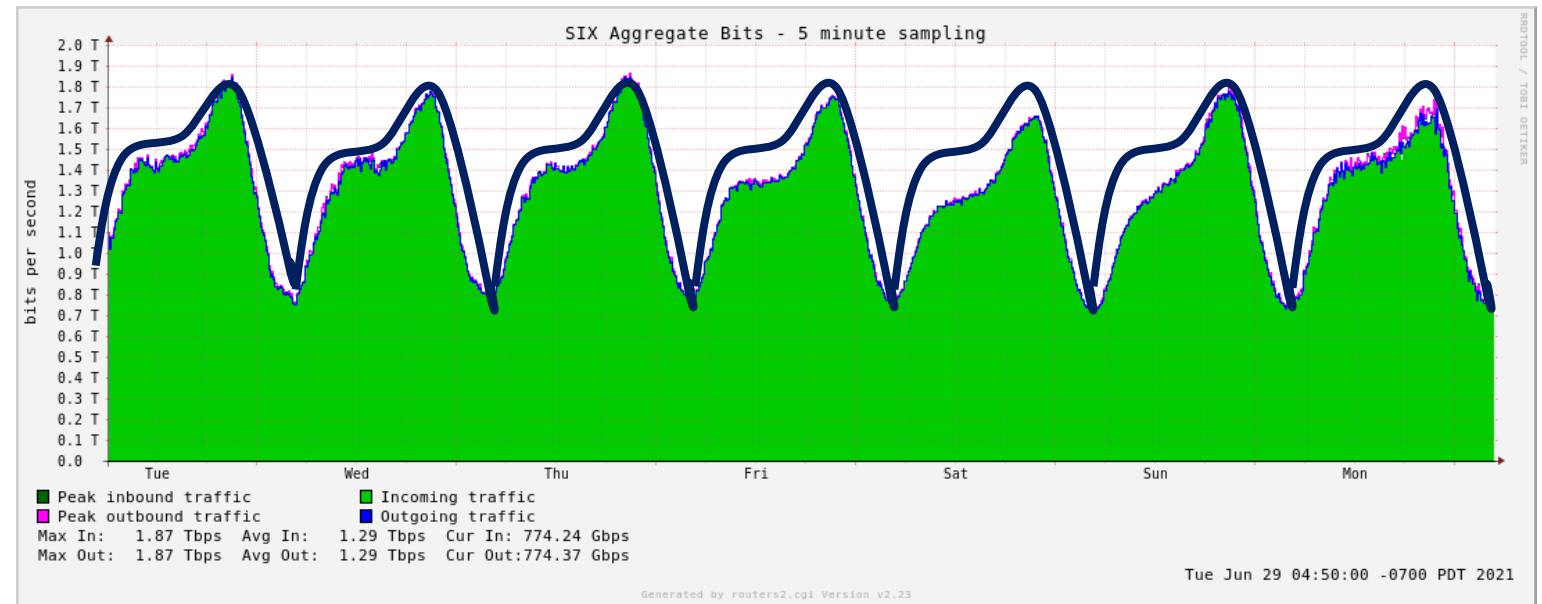


Chunk-based ensemble learning



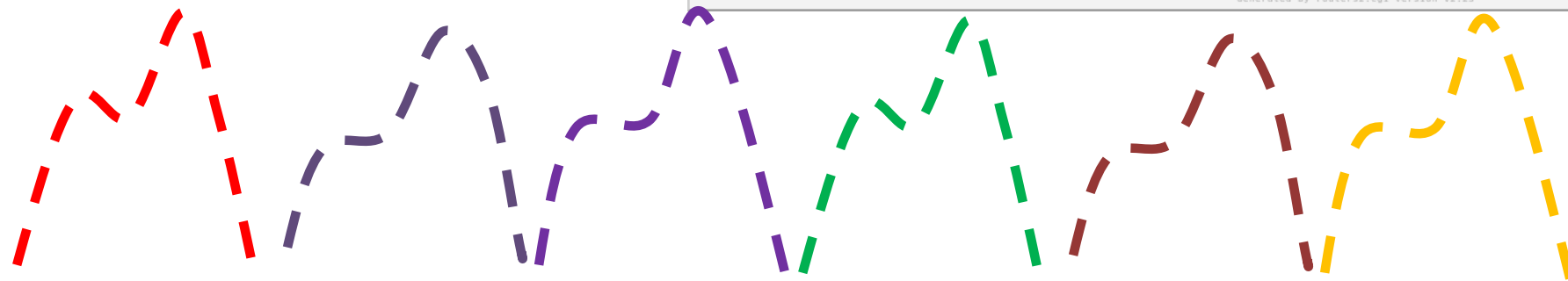
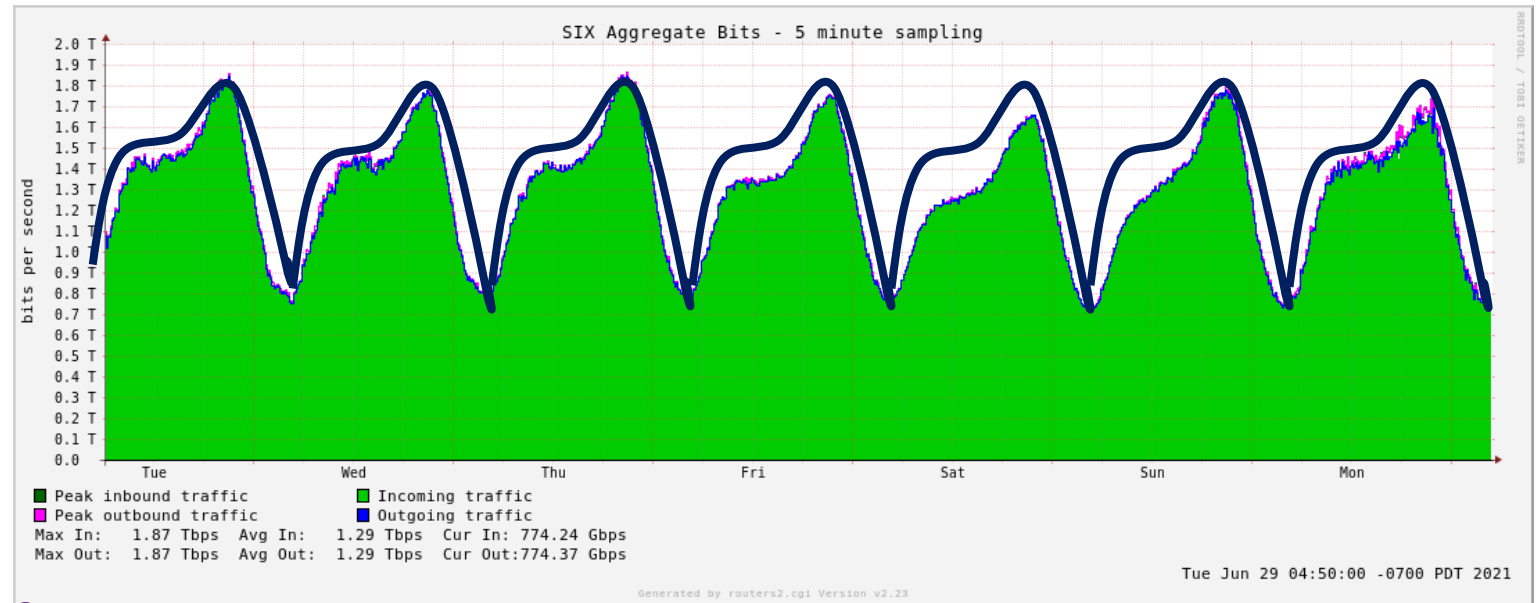


Reference static approach



Reference static approach

Model jest trenowany na danych z 60 dni i następnie użyty do predykcji, bez uwzględniania danych strumieniowych



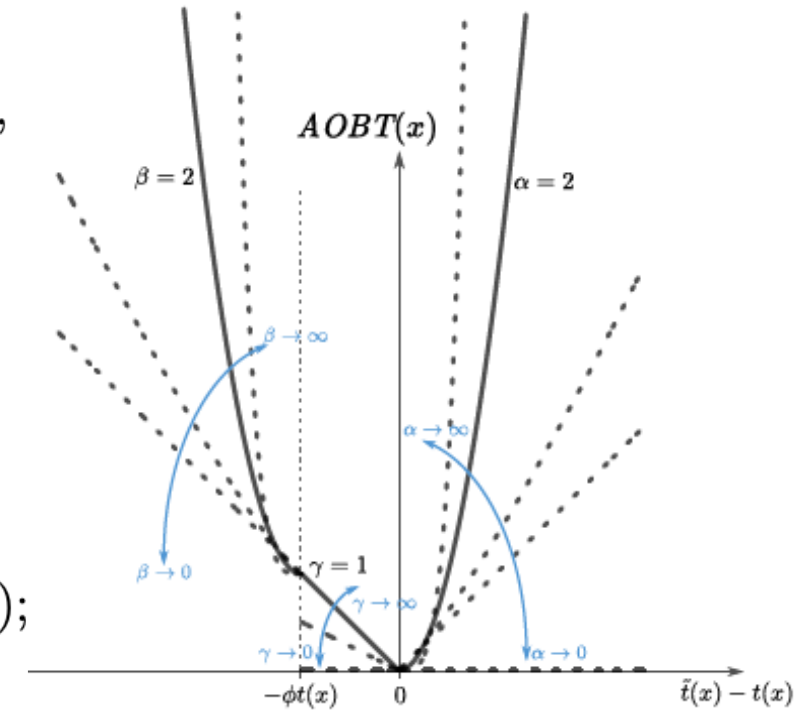
Nowa funkcja kryterialna

AOBT (Allocation outside blocking threshold)

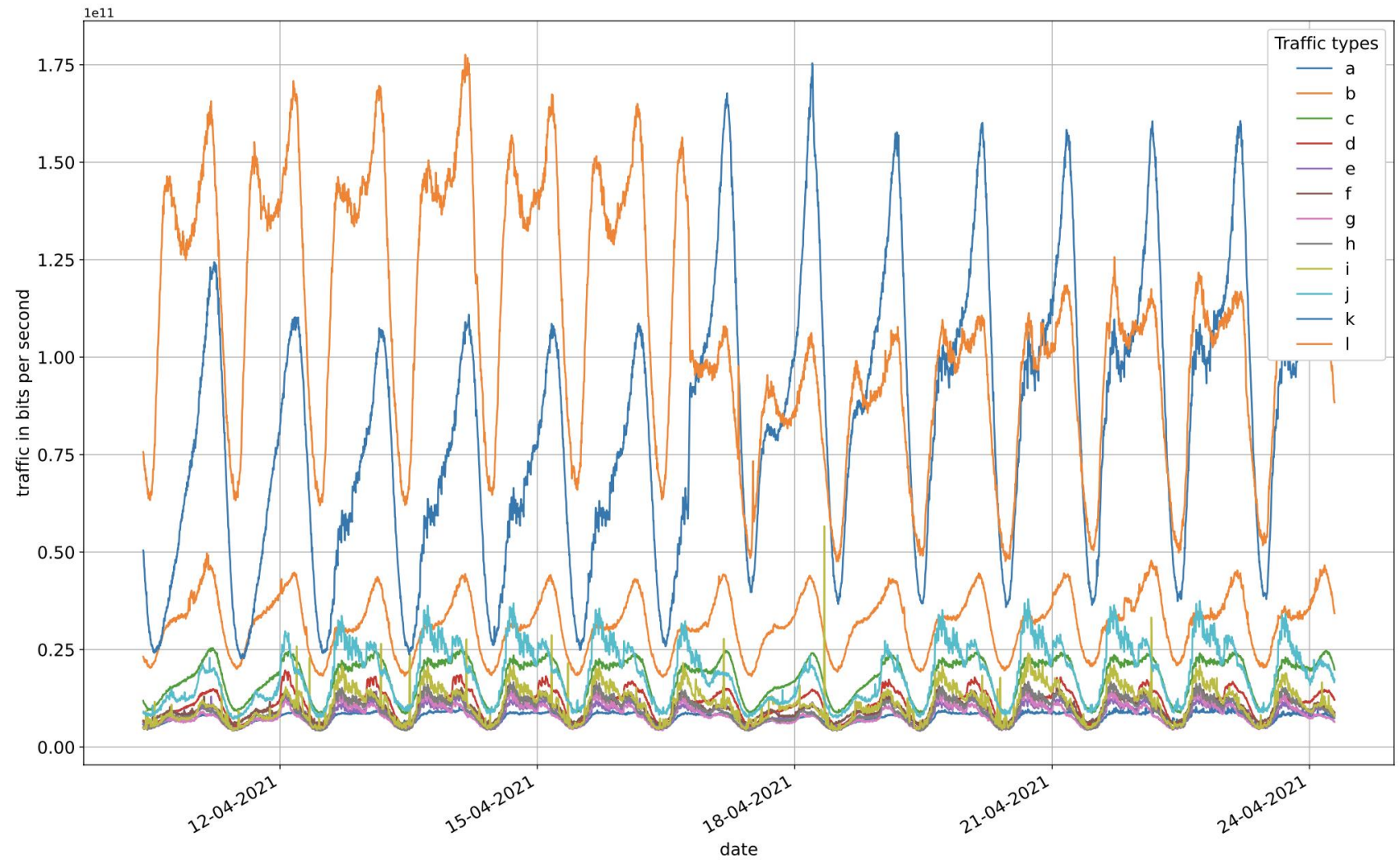
$$AOBT(x) = \begin{cases} [\tilde{t}(x) - t(x)]^\alpha & , t(x) < \tilde{t}(x) \\ [(1 - \phi)t(x) - \tilde{t}(x)]^\beta + \gamma\phi t(x) & , \tilde{t}(x) < (1 - \phi)t(x) \\ \gamma[t(x) - \tilde{t}(x)] & , (1 - \phi)t(x) \leq \tilde{t}(x) \leq t(x) \end{cases} ,$$

where:

- x – a moment in time when traffic is sampled/predicted;
- $t(x)$ – real traffic;
- $\tilde{t}(x)$ – predicted traffic;
- α – slope parameter for overestimation;
- β – slope parameter for underestimation outside blocking threshold (OBT);
- γ – slope parameter for underestimation within the accepted threshold;
- ϕ – accepted blocking threshold.

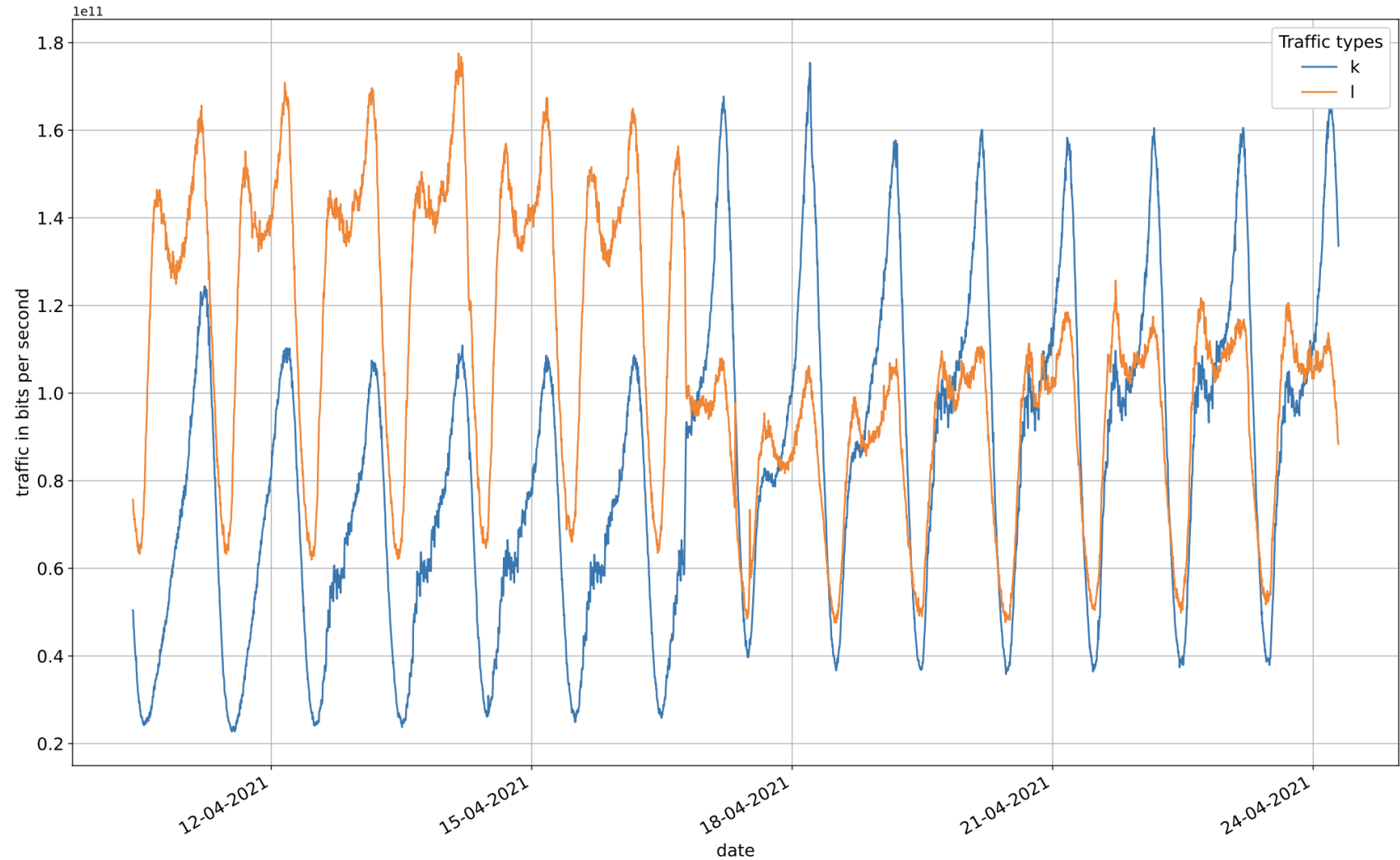


Concept drifts





Concept drifts



Wyniki

 Table 3: RMPSE for model *multi* for various traffic types, methods and estimators; the winner in each column highlighted

method	base estim.	traffic type												
		a	b	c	d	e	f	g	h	i	j	k	l	m
stream	LR	0,1333	0,0979	0,1044	0,1166	0,2168	0,1282	0,1539	0,1586	6,0132	0,1959	0,1207	0,2047	0,1277
	kNN	0,1438	0,1005	0,1483	0,1464	0,2512	0,1913	0,2391	0,2381	4,6241	0,3027	0,1900	0,2582	0,1269
static	LR	0,1839	0,1101	0,1406	0,1320	0,2495	0,2494	0,2418	0,2429	4,6379	0,2702	0,1363	0,2839	0,1613
	kNN	0,2591	0,1560	0,1763	0,1700	0,2909	0,2604	0,2609	0,2626	4,0117	0,3152	0,2293	0,4037	0,1667
	MLP	0,2223	0,1083	0,1304	0,1396	0,2695	0,2025	0,2332	0,2118	4,9902	0,2543	0,1299	0,2728	0,1542

 Table 4: AOBT with $\phi = 1\%$ for model *multi* for various traffic types, methods and estimators; the winner in each column highlighted

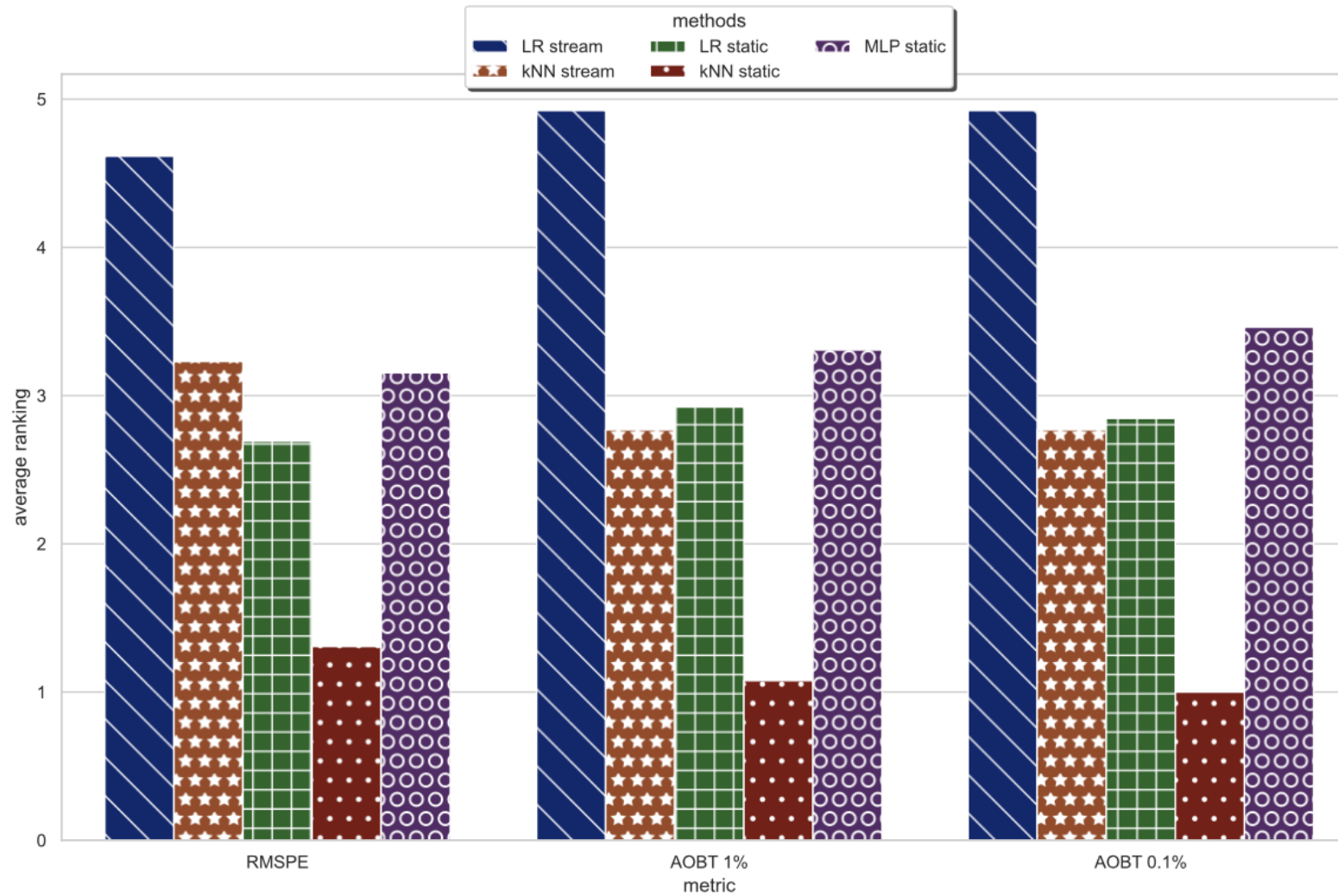
method	base estim.	traffic type												
		a	b	c	d	e	f	g	h	i	j	k	l	m
stream	LR	4,8963	4,4212	1,4660	0,7904	0,7017	0,7924	0,5963	0,8216	4,6462	0,5328	3,7422	3,6643	3,3682
	kNN	5,8065	4,5743	3,4237	1,2982	1,1426	2,1136	1,9839	2,7123	7,6454	1,6925	9,0587	6,5480	3,0884
static	LR	9,7333	5,4180	2,8830	0,9075	0,9330	4,6916	1,8458	2,5605	8,0481	1,2266	4,7735	4,3958	4,6613
	kNN	20,0955	26,2448	4,7814	2,0547	1,8052	6,0891	3,5463	5,1153	13,9112	2,6383	31,2641	22,2525	4,8604
	MLP	13,2472	5,0363	2,0467	0,9718	0,9488	2,3188	1,5166	1,6179	6,4268	0,9793	4,1312	3,9506	4,8657

$\underbrace{\hspace{10em}}_{\times 10^{17}} \quad \underbrace{\hspace{10em}}_{\times 10^{18}} \quad \underbrace{\hspace{10em}}_{\times 10^{19}} \quad \underbrace{\hspace{10em}}_{\times 10^{21}}$

 Table 5: Normalized AOBT with $\phi = 1\%$ for model *multi* for various traffic types, methods and estimators; the winner in each column highlighted

method	base estim.	traffic type												
		a	b	c	d	e	f	g	h	i	j	k	l	m
stream	LR	0,2436	0,1685	0,3066	0,3847	0,3887	0,1301	0,1681	0,1606	0,3340	0,2019	0,1197	0,1647	0,6922
	kNN	0,2889	0,1743	0,7160	0,6318	0,6330	0,3471	0,5594	0,5302	0,5496	0,6415	0,2897	0,2943	0,6347
static	LR	0,4844	0,2064	0,6030	0,4417	0,5168	0,7705	0,5205	0,5006	0,5785	0,4649	0,1527	0,1975	0,9580
	kNN	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9989
	MLP	0,6592	0,1919	0,4281	0,4730	0,5256	0,3808	0,4276	0,3163	0,4620	0,3712	0,1321	0,1775	1,0000

Wyniki - model *multi*



Porównanie modeli multi i single

Model comparison under RMSPE metric in different traffic types and base estimators, the winner in each column highlighted

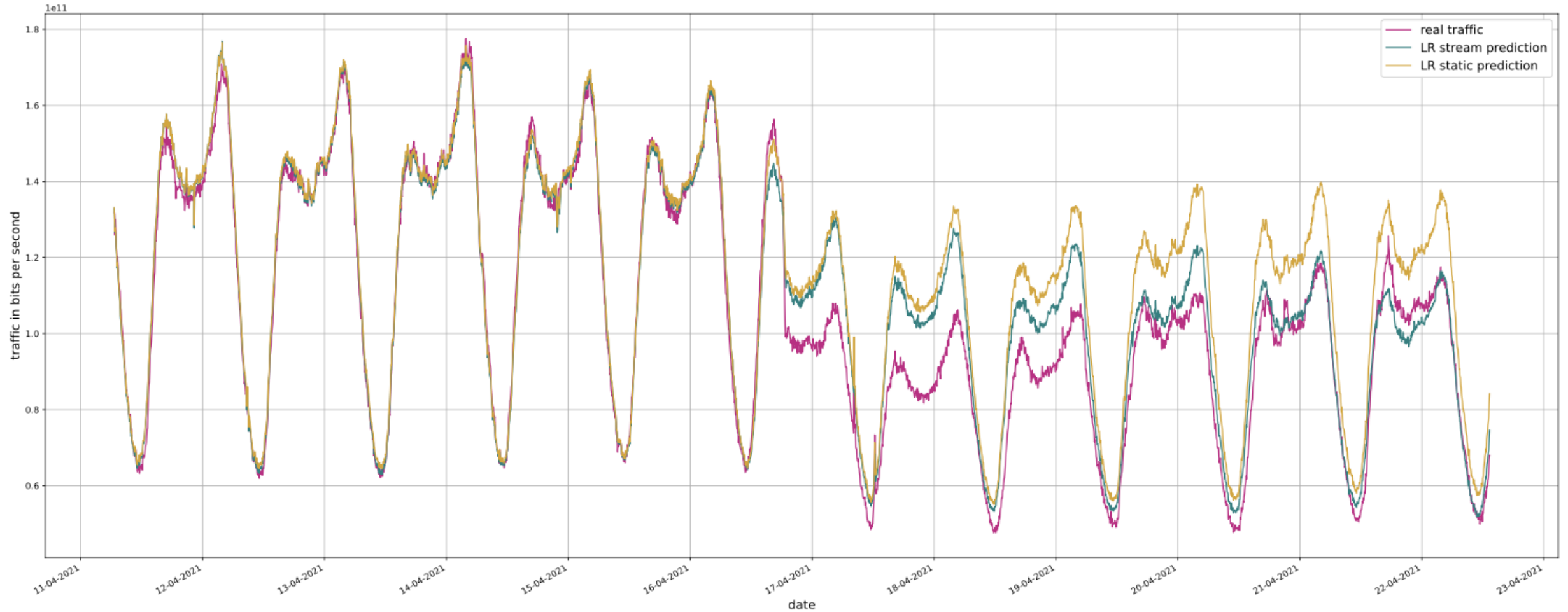
model	base estimator	traffic type												
		a	b	c	d	e	f	g	h	i	j	k	l	m
multi	LR	0,1333	0,0979	0,1044	0,1166	0,2168	0,1282	0,1539	0,1586	6,0132	0,1959	0,1207	0,2047	0,1277
	kNN	0,1438	0,1005	0,1483	0,1464	0,2512	0,1913	0,2391	0,2381	4,6241	0,3027	0,1900	0,2582	0,1269
single	LR	0,1351	0,0961	0,1022	0,1119	0,2352	0,1332	0,1613	0,1628	4,3492	0,1928	0,1219	0,2013	0,1273
	kNN	0,1320	0,0949	0,1059	0,1128	0,2229	0,1333	0,1628	0,1661	4,2198	0,2006	0,1237	0,2035	0,1271

Model comparison under AOBT with $\phi = 1\%$ metric in different traffic types and base estimators, the winner in each column highlighted

model	base estimator	traffic type												
		a	b	c	d	e	f	g	h	i	j	k	l	m
multi	LR	0,8432	0,9665	0,4282	0,6089	0,6141	0,3749	0,3006	0,3029	0,6077	0,3148	0,4131	0,5596	1,0000
	kNN	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9169
single	LR	0,9884	0,8791	0,4111	0,5113	0,6019	0,3750	0,2932	0,3050	0,6921	0,3016	0,4403	0,5239	0,9967
	kNN	0,8303	0,8156	0,4398	0,5288	0,5879	0,4006	0,3488	0,3557	0,6522	0,3681	0,4132	0,5289	0,9252



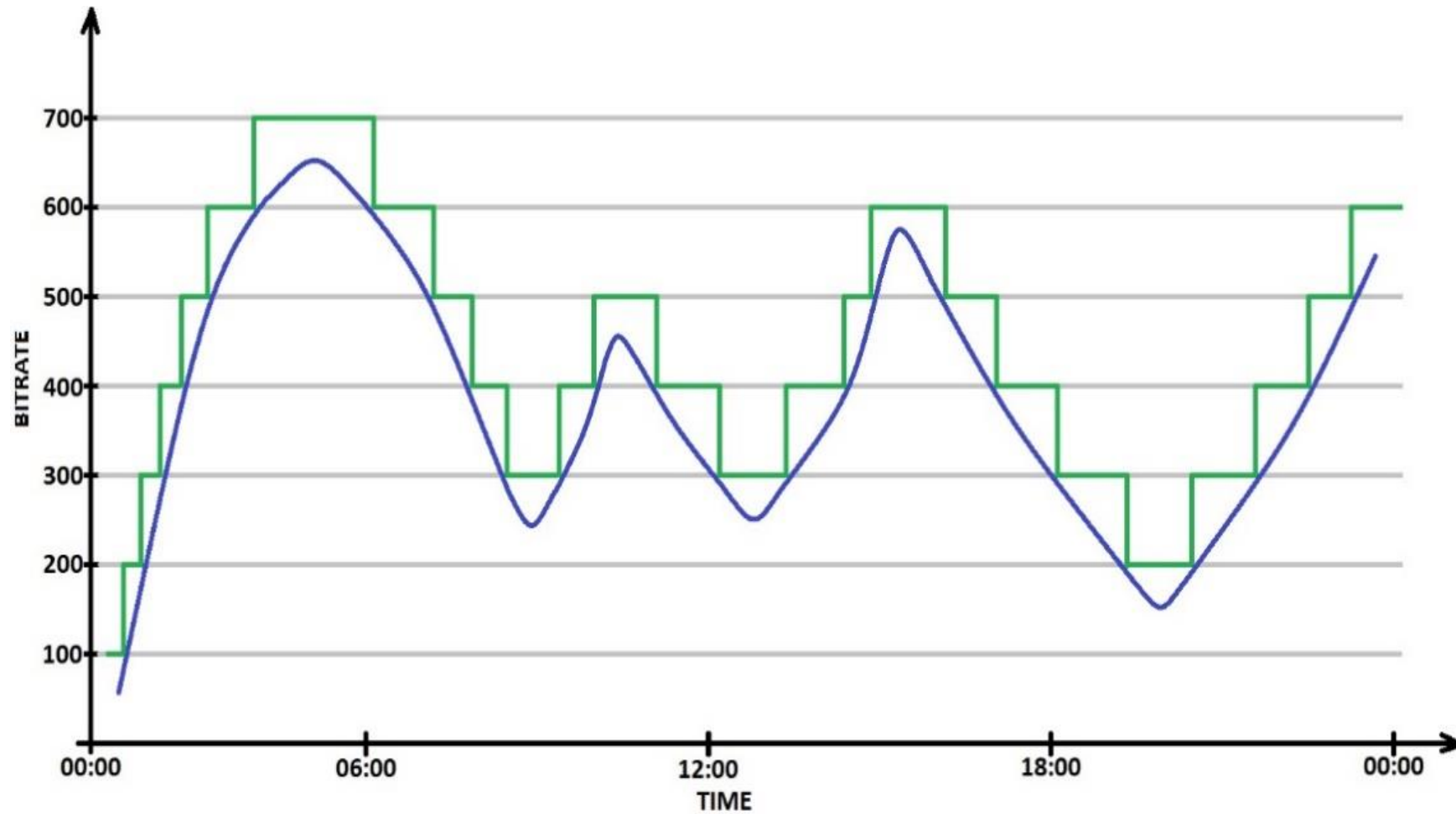
Adaptacja modelu do concept drift



Predykcja ruchu o ustalonych poziomach przepustowości

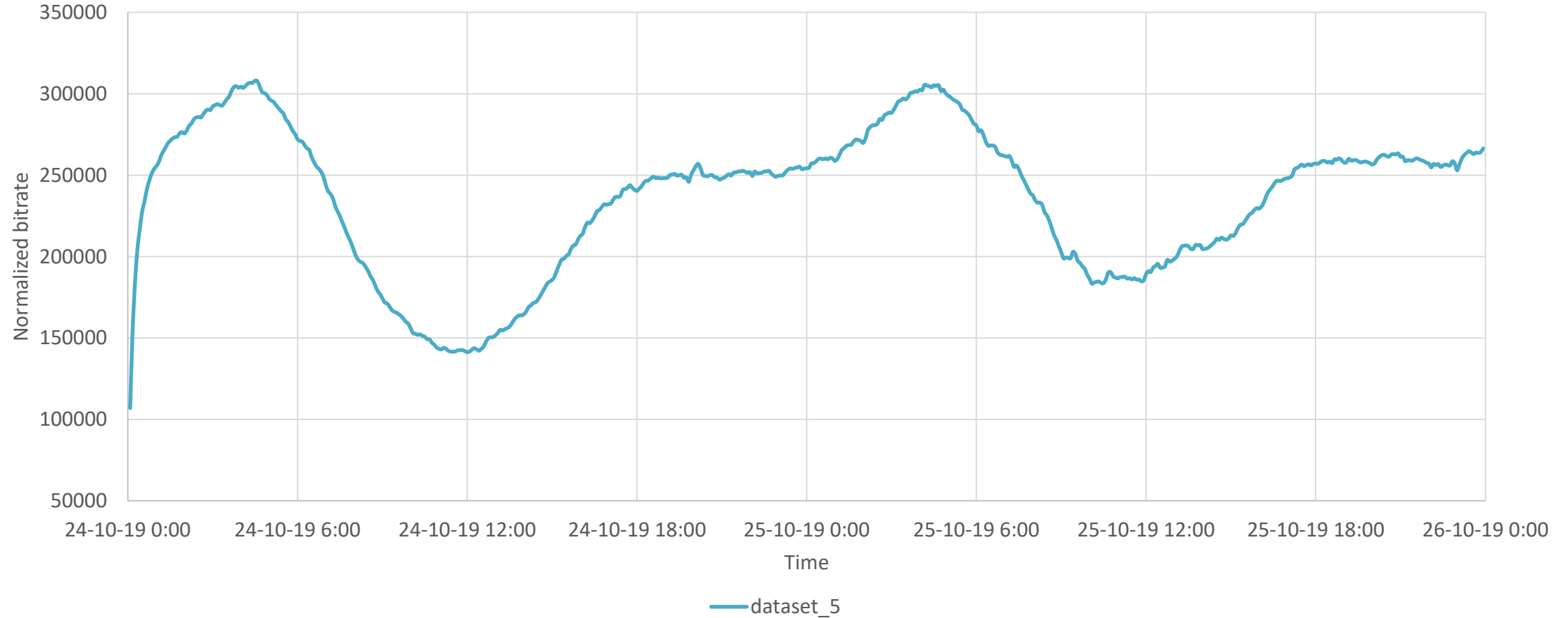
- [D Szostak, A Włodarczyk, K Walkowiak, Machine Learning Classification and Regression Approaches for Optical Network Traffic Prediction, Electronics 2021]
- [D. Szostak, Machine Learning Ensemble Methods for Optical Network Traffic Prediction, CISIS 2021]
- [D. Szostak, K. Walkowiak, A. Włodarczyk, Short-term traffic forecasting in optical network using linear discriminant analysis machine learning classifier, ICTON 2020]
- Predykcja dotyczy określenia **poziomu przepustowości** zgodnie z granulacją stosowaną w sieci – w sieciach optycznych przepustowość zestawianych połączeń (ścieżek optycznych) mają określone poziomy przepustowości wynikające z:
 - **Parametrów** stosowanych **transponderów optycznych**, w tym stosowanego formatu modulacji (np. BPSK 50 Gb/s, QPSK, 100 Gb/s, 8-QAM 150 Gb/s, 16-QAM 200 Gb/s)
 - Granulacji przepustowości w technologiach Ethernet oraz OTN
- **Na podstawie ruchu z Seattle** Internet Exchange stworzono **wzorce ruchu uwzględniające fluktuacje** wynikające ze stosowania różnych usług (na podstawie raportów Cisco)

Poziomy przepustowości



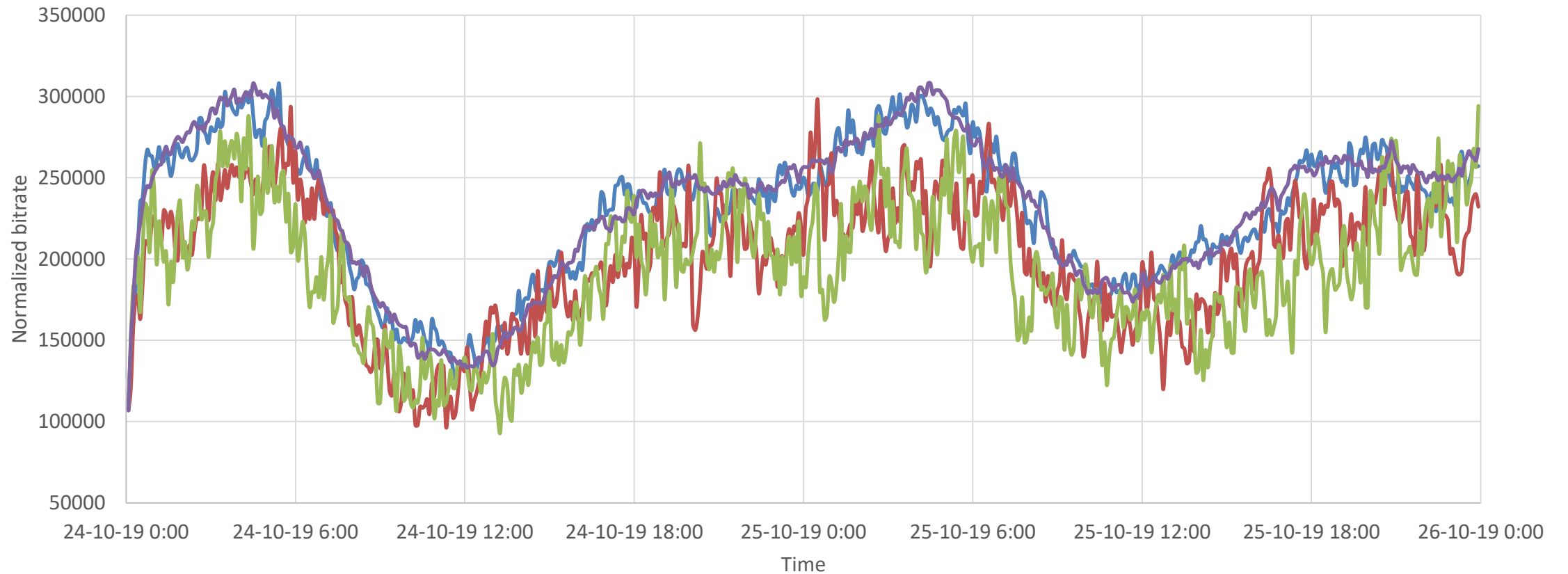


Dataset – wzorcowy ruch Seattle





Dataset – dane dla poszczególnych par węzłów

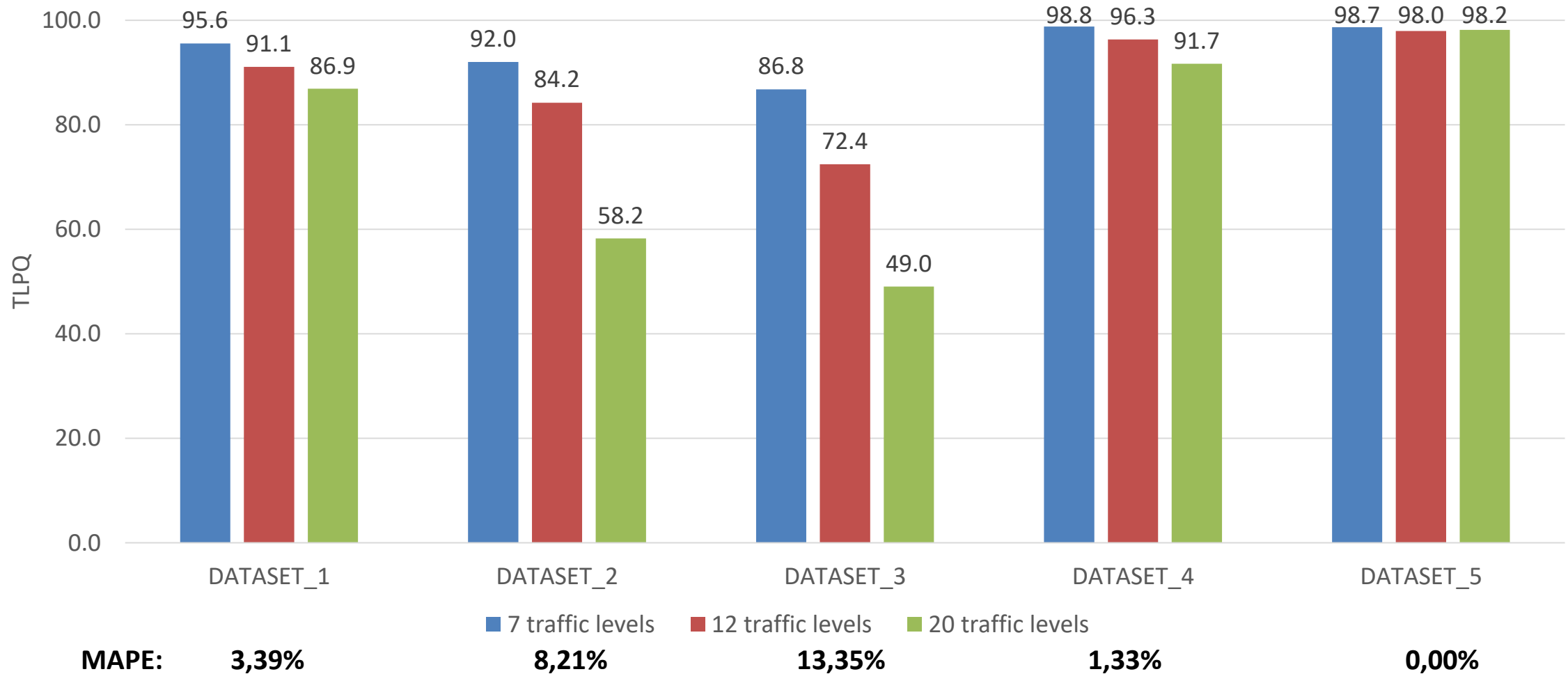


— dataset_1 — dataset_2 — dataset_3 — dataset_4

MAPE: 3,39% 8,21% 13,35% 1,33%



Wyniki





Dziękuję za uwagę 😊

krzysztof.walkowiak@pwr.edu.pl