

POLITECHNIKA BYDGOSKA IM. JANA I JĘDRZEJA ŚNIADECKICH

Politechnika Bydgoska
im. Jana i Jędrzeja Śniadeckich w Bydgoszczy
Wydział Telekomunikacji Informatyki i
Elektrotechniki

Streszczenie rozprawy doktorskiej

Mgr inż. Adam Flizikowski

*Algorytmy sterowania przyjmowaniem zgłoszeń w
sieciach bezprzewodowych przyszłości*

*(Admission control algorithms for future wireless
networks)*

Promotor
Prof. dr hab. inż. Tadeusz Wysocki

Bydgoszcz 2023

STRESZCZENIE

Dla osiągnięcia celu pracy, sformułowano 5 problemów badawczych, które rozwiązano drogą eksperymentu cyfrowego, symulacyjnego, fizycznego z uwzględnieniem warunków rzeczywistych oraz laboratoryjnych, a także uwzględniając warunki idealizowane (np. za pomocą stworzonego stanowiska do emulacji warunków sieciowych w sieciach 4G/5G).

Wszystkie problemy były skierowane na weryfikację funkcji jakości zmiennych systemu komunikacji bezprzewodowej, we wskazanych wyżej warunkach.

Problemy rozwiązano, uzyskując liczne wyniki dotyczące charakterystyk jakościowych tj.: grade of service (GoS), quality of service (QoS), quality of experience (QoE) – uwzględniając metryki takie, jak prawdopodobieństwo blokady, prawdopodobieństwo przedwczesnego zakończenia połączenia, stopień wykorzystania pasma.

Sformułowano wnioski poznawcze dotyczące algorytmów sterowania jakością w kierunku optymalizacji i) jakości sterowania w sieciach bezprzewodowych, ii) jakości doświadczanej przez użytkowników takich sieci, iii) warunków użytkowania algorytmów w różnych konfiguracjach sieci (np. 4G, 5G, WiFi), stosowania algorytmów sterowania przyjmowaniem zgłoszeń, algorytmów kontroli przeciążeń oraz algorytmów sterowania ruchem – powyższe w odniesieniu do zmiennych takich jak m.in.: próg ochronny zasobów radiowych, intensywność napływu zgłoszeń, różne scenariusze jakości sygnału, typy przenoszonego ruchu. Osiągnięto sformułowane cele, wskazano również możliwości doskonalenia jakości systemów telekomunikacji w warunkach mobilnych.

Stworzone zostały innowacyjne algorytmy, a także propozycje ram/procedur badawczych, usprawniające proces ciągłego poprawiania jakości, wg własnego pomysłu, algorytmy te zostały przebadane w odpowiednich warunkach. A zaproponowane rozwiązania mają charakter użyteczny i praktyczny – co zostało potwierdzone na etapie egzemplifikacji i praktykalizacji.

Słowa kluczowe: sterowanie przyjmowaniem zgłoszeń, kontrola natłoku, multi-RAT, algorytmy

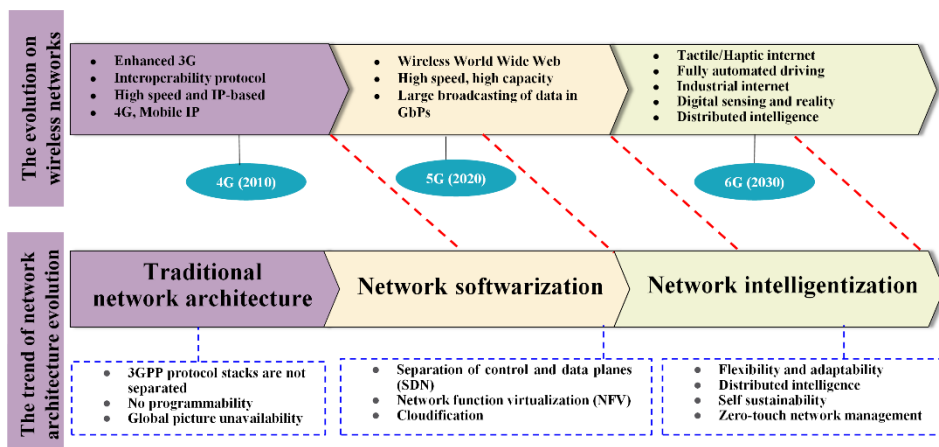
SPIS TREŚCI

Streszczenie	2
1 Wprowadzenie	4
1.1 Sterowanie przyjmowaniem zgłoszeń w sieciach przyszłości	4
2 Cel rozprawy	5
3 Problemy badawcze, teza rozprawy	6
4 Metodyka badań	7
4.1 Rodzaje badań	7
4.2 Zmienne wejściowe	9
5 Modele rozwiązania problemów	13
6 Plan i realizacja badań	15
7 Wyniki procedur działania badawczego	17
8 Podsumowanie osiągnięć i wnioski	24
8.1 Sformułowanie wniosków	26
9 Podsumowanie	27
9.1 Wnioski o charakterze poznawczym	27
9.2 Wnioski praktyczne	29
9.2.1 Synteza kontrolerów wideo	30
9.2.2 Wnioski z badań nad predykcją zużycia procesora w serwerach edge/EMDC	30
9.3 Rekomendacje do dalszych badań	31

1 WPROWADZENIE

1.1 Sterowanie przyjmowaniem zgłoszeń w sieciach przyszłości

W chwili obecnej akademicy, instytuty naukowe i dostawcy rozwiązań z obszaru sieci bezprzewodowych zaangażowani są w prace nad opracowaniem technologii dla sieci B5G oraz 6G. Choć prace standaryzacyjne nad 6G są w stosunkowo wczesnym stadium, a pierwsze specyfikacje spodziewane są w okolicach 2028. Jednak kilka ostatnich lat to bardzo dynamiczne działania naukowe finansowane przez KE np: 6G-BRICKS, 6G-SANDBOX, DAEMON, HEXA-X ale także prace skupiające się na różnych technikach określających, czym 6G ma być. Można tu wymienić choćby takie techniki jak: wielodostęp (NOMA, RSMA), paradygmat alokacji zasobów radiowych (cell-free) i zasobów obliczeniowych (schedulersy zasobów obliczeniowych na platformie Kubernetes), platformy obliczeniowe (MEC), masowe systemy antenowe mMIMO, matryce antenowe wzmacniające sygnały radiowe (RIS), itp. Oczywiście nie można pominąć kluczowej roli algorytmów uczenia maszynowego AI/ML, zapewniających wsparcie na różnych poziomach przetwarzania sygnałów w stosie radiowym (ale także na serwerach edge-cloud).



Rysunek 1 Ewolucja architektury sieci pomiędzy generacjami sieci bezprzewodowych (4G, 5G, 6G)

Kluczowe jednak jest, że z jednej strony nowe aplikacje i usługi (XR, Tactile internet, CCAM, TSN) napędzają oczekiwania wobec technologii, kierując współczesne sieci w stronę sieci wysoko-wydajnych, tj. sieci o wysokich wartościach parametrów jakościowych (za IMT-2020 i niedługo IMT-2030). Z drugiej strony ruch generowany przez użytkowników obecnych aplikacji (głównie wideo), rośnie w tempie wykładniczym – a spodziewany wzrost pojemności, który pozwoli nadążyć za tak szybkim wzrostem określany jest na

około 20-krotny. Niektórzy autorzy określają te zmiany mianem „zalewu danych” (ang. data tsunami). Coraz większą popularnością cieszy się strumieniowanie mediów w kierunku „w górę”, co stanowi przeciwwagę do dotychczas asymetrycznego rozplywu ruchu w sieci Internet, zwłaszcza w świetle rosnącej popularności samochodów (pojazdów) autonomicznych. Autonomia na obecnym etapie rozwoju wymaga wsparcia człowieka (przykładem są projekty KE H2020 Vital-5G, 5GBlueprint), w postaci operatora, który zdalnie powinien być w stanie przejść kontrolę nad pojazdem gdy zawiedzie automatyka.

Sterowanie przyjmowaniem zgłoszeń stanowi kluczową funkcję, która przesądza o efektywnym wykorzystaniu zasobów operatora, która determinuje stabilność zachowania sieci jako takiej a przede wszystkim ma wpływ na realizację polityk optymalnego przydziału zasobów, umożliwiających spełnienie celów strategicznych operatora. Podobnie funkcja przeciwdziałania natłokowi (ang. *congestion control*) stanowi naturalne uzupełnienie roli sterowania przyjmowaniem zgłoszeń. Obie funkcje stanowią kompletny mechanizm, którego stosowalność w dziedzinie czasu, plasuje się na poziomie od sekund do minut, czyli dla okresu trwania sesji. Zasoby w sieciach najłatwiej jest kontrolować zachowując podział na klasy obsługi (np. zgodnie z modelem Diffserv) realizowane np. przez profile usługowe QCI, albo wspomagane mechanizmem grupowania podobnych rodzajów aplikacji w tzw. plastry sieciowe. Natomiast trudno wyobrazić sobie przyszłe sieci radiowe bez odpowiednich mechanizmów szeregowania zasobów radiowych (ang. *scheduler*). Ta ostatnia funkcja stanowi naturalne połączenie z mechanizmami admission/congestion control, jednak nie jest ona przedmiotem niniejszej pracy. Warto jednak zauważyć, że wachlarz topologii sieci radiowych (dostępowych) jest szeroki a w związku z 20-krotnym wymaganym wzrostem pojemności sieci, jednoczesne użycie większej liczby technologii radiowych (ang. *multi-RAT*) stanowi wartościową alternatywę dla zwiększania liczby zasobów radiowych w sieciach bezprzewodowych.

2 CEL ROZPRAWY

Jako cele rozprawy przyjęto: analizę, ocenę stanu wiedzy i praktyki zagadnienia, badania wybranych instalacji, ich charakterystyk jakości efektywności i nieszkodliwości produktu i procesu transmisji oraz rozwój elementów, relacji systemu sterowania przyjmowaniem zgłoszeń w przyszłych sieciach bezprzewodowych.

Szczegółowo, można przyjąć, że celem jest: Analiza, Badania i Inteligentny Rozwój (w oparciu o wiedzę i innowacje) oraz ocena jakości algorytmów sterowania dla przyjmowania i przetwarzania żądań (przeciążenia) w przyszłych sieciach bezprzewodowych.

3 PROBLEMY BADAWCZE, TEZA ROZPRAWY

Aby osiągnąć główne i dodatkowe cele pracy, sformułowano problemy badawcze:

Problem 1: Czy możliwe jest wprowadzenie merytorycznych modeli przydatnych w sterowaniu przyjmowaniem zgłoszeń (i sterowaniu przeciążeniami), w oparciu o istniejący stan wiedzy o tym aspekcie w sieciach bezprzewodowych, oceniany z uwzględnieniem normy jakości (różnica między modelem a rzeczywistością, zgodnie z definicją normy $Q=||x-y||$), które umożliwią: i) ocenę stanu, ii) wskazanie kierunków rozwoju, iii) opracowanie alternatywnych rozwiązań, iv) przeprowadzenie badań (z oszacowaniem/oceną) oraz v) zaproponowanie nowych modeli kontroli (trendów) w kierunku przejścia z 4G na 5G i kolejne generacje;

Problem 2: Czy algorytmy sterowania przyjmowaniem zgłoszeń oparte na deklaracjach 4G (DBAC) i pomiarach (MBAC) mogą być ponownie wykorzystane z niezbędnymi modernizacjami, aby były odpowiednie dla sieci 5G / i poza nimi, aby jeszcze bardziej zoptymalizować warunki techniczne, liczbę użytkowników i zmniejszyć prawdopodobieństwo (P_b , P_d) w stosunku do istniejącego SOTA?

Aby osiągnąć cel i rozwiązać dwa wyżej wymienione problemy badawcze, postanowiono dodatkowo rozwiązać dodatkowe zadania badawcze/twórcze:

Zadanie 1 (rozdział 5): Jak zdefiniować kompletne, zrównoważone, kompleksowe ramy oceny dla innowacyjnych i zmodernizowanych algorytmów kontroli przeciążenia (w 4G/5G), aby umożliwić odpowiednie optymalizacje dostosowujące przesyłany strumień danych do podstawowej przepustowości sieci (np. zdalne monitorowanie samochodów autonomicznych) w celu poprawy możliwości niezawodnych testów mobilnych w laboratoriach, opartych na danych z badań terenowych.

Zadanie 2 (rozdział 6): W jakim stopniu w quasi-optymalizacji parametrów i algorytmów sterowania przyjmowaniem zgłoszeń i przeciążeń, można wykorzystać połączenia jej z inteligentną, opartą na QoE kontrolą ruchu w sieciach z wieloma technologiami radiowymi (ang. RAT), przy użyciu nowatorskich architektur kontrolerów RAN w płaszczyźnie sterowania? Gdzie różne sieci uwzględniają agregację sieci dostępowych, w tym w szczególności opartych na OFDMA i CSMA.

Zadanie 3 (rozdział 7): W jakim stopniu wprowadzenie nowatorskich platform obliczeniowych edge, takich jak Micro Edge Data Center (EMDC), w połączeniu z algorytmami przewidywania obciążenia i umieszczania zwirtualizowanych funkcji sieciowych pozwala poprawić efektywność i przenoszony ruch użytkowników, w powiązaniu z mechanizmami sztucznej inteligencji / ML do

realizacji celów automatycznego skalowania CU-UP w zdezagregowanych plastrach sieciowych 5G opartych na ORAN.

Zadanie 4 (Rozdział 8): Czy obecnie znane algorytmy kontroli dostępu oparte na pomiarach (MBAC), wspierane przez techniki AI / ML uczenia ze wzmocnieniem, powiązane z sekwencyjnym podejmowaniem decyzji w oparciu o MDP, z niezbędnymi modernizacjami, mogą być wykorzystane do ulepszenia rozwiązań CAC w systemach 4G / 5G oraz w jakim stopniu ułatwiają ramowe podejście do wspierania ewolucji przyszłych sieci (np. przez uwzględnienie dodatkowych kryteriów/celów optymalizacji)?

Zadanie 5 (rozdział 7): Jak zaprojektować i wdrożyć skuteczne warunki, ramy optymalizacji, które wspierałyby definicję zrównoważonych, opartych na uczeniu się algorytmów RRM CAC/CC w sieciach 5G/B5G do użytku w centrach danych typu edge-micro?

TEZA ROZPRAWY

Teza pracy została sformułowana w sposób następujący:

Istniejące warunki techniczne w sieciach otwartych RAN 5G/beyond-5G można poznać, opisać, rozwinąć, poprawić, np. poprzez modyfikację algorytmów zarządzających sterowaniem przyjmowania zgłoszeń i kontroli natłoku, zarówno na podstawie deklaracji, jak i ii) opartych pomiarami – a oczekiwanym rezultatem będzie poprawa charakterystyk eksploatacyjnych, w tym wydajności (jakości) sieci w odniesieniu do efektywnego wykorzystania radia, zasobów obliczeniowych i sieciowych.

4 METODYKA BADAŃ

W celu udowodnienia postawionej tezy, osiągnięcia celów i rozwiązania problemów badawczych autor zaplanował zestaw badań o charakterze komplementarnym: a) badania w układzie rzeczywistym tj. eksperymentalne (w terenie, w laboratorium H2020 ORCA, 5G ORAN testbed), b) badania symulacyjne (Matlab, NS2) oraz c) badania teoretyczne (model statystyczny strat/opóźnień).

4.1 Rodzaje badań

Badania eksperymentalne zostały zrealizowane u dwóch polskich operatorów (INEA, Choszczno) ale także w dedykowanym laboratorium w Niemczech (zdalny dostęp do laboratorium projektu H2020 Orca). Pierwsza grupa badań została wykorzystana do zebrania danych potrzebnych do opracowania modelu statystycznego dla kanału mobilnego (rozdział 4 pracy) ale także do przygotowania modelu kanału opartego o uczenie maszynowe (w rozdziale 8). Druga grupa badań eksperymentalnych została przeprowadzona w laboratorium

TU Drezno, wyposażonym w nadajniki/odbiorniki radio programowalnego (SDR) firmy National Instruments dla technologii WiFi oraz 4G/LTE. Ponadto badania z wykorzystaniem komercyjnej sieci polskiego dostawcy 5G (IS-Wireless) zostały wykorzystane do zebrania danych na potrzeby opracowania modelu predykcyjnego, wspomagającego podejmowanie decyzji na potrzeby skalowania wybranych elementów zwirtualizowanej, zdezagregowanej sieci O-RAN 5G.

Badania symulacyjne przeprowadzone zostały z wykorzystaniem symulatorów Matlab (5G Vienna SLS) oraz NS2 (tutaj autor stworzył własne rozszerzenie do algorytmów CAC). W obu środowiskach autor zaimplementował zmodernizowane przez siebie algorytmy CAC. Środowisko NS2 zostało uzupełnione o rozszerzenie do symulatora Matlab jako „środowiska uruchomieniowego” dla algorytmów uczących. Autor zaimplementował znane algorytmy uczące oparte na teorii Markova (MDP) oraz uczeniu ze wzmocnieniem (Q-Learning, SARSA).

Badania teoretyczne realizowane były w oparciu o oprogramowanie MS Excel oraz oryginalną procedurę identyfikacji statystycznej, wykorzystującą zestaw testów statystycznych (m.in. Manna-Whitney’a, T-Studenta). Dodatkowo autor stworzył narzędzie w języku VBasic, pozwalające na stworzenie modelu zmiennej losowej strat i opóźnień dla mobilnego kanału radiowego. Narzędzie zostało sparametryzowane i dostosowane na podstawie danych z pomiarów terenowych przedstawionych w rozdziale 4.

Opracowane **modele statystyczne** zmian kanału radiowego, zostały wykorzystane wewnątrz emulatora zaprojektowanego i zrealizowanego przez autora, do odwzorowania kanału radiowego dla zebranych pomiarów na dwa sposoby: i) w sposób szczegółowy (ang. *trace-based*) albo ii) w oparciu o stworzony model statystyczny opóźnień i strat. Dodatkowo w celu walidacji od-końca-do-końca (E2E), stworzonego modelu oraz emulatora jako całości (narzędzie oparte na systemie Linux, stworzonej aplikacji oraz skryptach mc) wykorzystane zostały dostępne w przestrzeni publicznej, metryki jakościowe QoE typu „no-reference”, opracowane przez naukowców z wydziału telekomunikacji AGH.

Dla wsparcia badań eksperymentalnych autor zaplanował, zaprojektował, zrealizował i porównał z narzędziami komercyjnymi (IXIA) własnego autorstwa **sensor pomiarowy** oparty na mikro-komputerze RaspberryPi, wyposażonym w odbiornik GPS dla właściwej podstawy czasu w pomiarach terenowych. To działanie było niezbędne ze względu na brak dostępu do przenośnych urządzeń pomiarowych, zapewniających odpowiednią dokładność pomiaru ale umożliwiającą pomiary w terenie.

Jeśli chodzi o wykorzystane algorytmy sterowania przyjmowaniem zgłoszeń – autor skoncentrował się na **dwóch grupach algorytmów**: opartych na

deklaracjach (DBAC) oraz opartych na pomiarach (MBAC). Pierwsze głównie w celu porównania wydajności (jakości) oferowanej przez algorytmy autora oraz drugie jako główny punkt ciężkości. Algorytmy MBAC zostały opracowane jako modernizacja istniejących w literaturze algorytmów pomiarowych MBAC, wykorzystujących średnią krocząca liczbę symboli wykorzystywanych przez połączenie pomiędzy symbolami (slotami) czasowo-częstotliwościowymi (OFDM). Algorytmy te zostały rozszerzone o mechanizmy kompensacji zmian modulacji (AMC) oraz napływ zgęstek zgłoszeń do systemu, głównie z perspektywy uwzględnienia trendu dogęszczania przyszłych sieci. Proste wersje algorytmów średniej kroczącej zostały wykorzystane jako punkt odniesienia (EMAC, nscARAC). Także w podejściu opartym o uczenie maszynowe wykorzystano algorytmy uczenia ze wzmocnieniem, ze względu na ich wysoką elastyczność dopasowania do danych z pomiaru zgłoszeń w systemie (definiowania funkcji nagrody).

ZAŁOŻENIA

W pracy nie zastosowano teorii kolejek do pomiaru teoretycznej pojemności systemu radiowego, ze względu na ograniczenie stosowalności tego typu podejść do systemu w stanie ustalonym. Natomiast pomiary pojemności systemu bezprzewodowego zrealizowano za pomocą narzędzi symulacyjnych (NS2, Matlab) i odpowiednio spreparowanych scenariuszy pomiarowych. Ponadto autor korzystał z szeroko stosowanych w literaturze wzorów na obliczanie pojemności systemów bezprzewodowych.

Aby opracować innowacje przedstawione w niniejszej pracy autor wykonał analizę literatury i zamieścił w pracy odwołania do ponad 300 pozycji literaturowych.

4.2 Zmienne wejściowe

System przesyłowy jest zbiorem warunków technicznych W_j , odpowiednich elementów E_1, E_2, \dots, E_m , relacji między tymi elementami R_1, R_2, \dots, R_n , które są funkcjami Θ oraz t (tj. czas pracy i dynamika procesu), które mają na celu realizację zbioru postulowanych stanów $SP_i=H$, charakterystyk (4-1), jako wielkości wyjściowe, od których zależy ocena ogólnej jakości, wydajność, nieszkodliwość transmisji danych, tj. np. od jednolitości, wydajności, wysokiej sprawności, rozsądnego zużycia energii i charakteru, jednostkowego zużycia energii itp.

$$SP_i = f(Wt_j) \quad (4-1)$$

gdzie:

SP_i - i-ty postulowany stan transmisji w sieciach bezprzewodowych,

Wt_j - j-te warunki techniczne opisane koncepcją rozwiązania dla postulowanego stanu,

W celu właściwej identyfikacji modelowanego systemu autor wykorzystał znane

z literatury podejście do zastosowania czterech podsystemów operacyjnych, realizujących jasno określone funkcje:

1. podsystem procesu (SP), fenomenalne funkcjonowanie, tj. realizacja przez sieć (system) wyznaczonego celu (przyjmowanie zgłoszeń, transmisja danych, rozwiązywanie zatorów, ...);
2. podsystem sterowania (SC), wysokowydajne zarządzanie i koordynacja budowy podsystemów (procesowych, informacyjnych i logistycznych) - dla niskoenergetycznej realizacji celów poszczególnych elementów, relacji i zintegrowanych funkcji kontrolnych zapewniających jakość przesyłu;
3. podsystem informacyjny (SJ), samoorganizacja, przetwarzanie i dystrybucja strumieni informacji w ramach systemu, spoza systemu zgodnie z potrzebami i celami działania innych struktur specjalnych (podsystemów);
4. podsystem logistyczny KL, niezawodne utrzymanie i dostawa (wdrożenie) innych podsystemów (np. MEC, LBO), systemy specjalne (i własny), dla niezawodnej realizacji wielu celów operacyjnych.

Powyższe podsystemy powinny być zawsze rozpatrywane całościowo i wzajemnie powiązane podczas analizy systemu, ponieważ łącznie definiują wymiary systemu w sposób kompleksowy.

Metoda analizy, oceny i wskazywania kierunków zrównoważonego rozwoju jakości systemu transmisyjnego w sieciach bezprzewodowych (QbP) została oparta na istniejącej wiedzy na temat wykorzystania potencjału technologii, środowiska, nowoczesnych technologii, urządzeń i procesów oraz cech architektonicznych systemu transmisji, ukonstytuowanych w zakresie projektowania, sterowania, przetwarzania, eksploatacji i rozwoju, a dodatkowo na standardach algorytmów wspomaganych sygnałami sterującymi $\bar{S}(\theta, t)$:

$$\begin{aligned}
 E_1 &= E_1(\bar{W}, \theta, t) \\
 E_2 &= E_2(\bar{W}, \theta, t) \\
 &\dots\dots\dots \\
 E_m &= E_m(\bar{W}, \theta, t) \\
 R_1 &= R_1(W, s, \theta, t) \\
 R_2 &= R_2(W, s, \theta, t) \\
 &\dots\dots\dots \\
 R_n &= R_n(W, s, \theta, t)
 \end{aligned}
 \tag{4-2}$$

W ramach modelowania systemu wejściowo-wyjściowego zastosowano poniższy model. Przy takich założeniach ogólny model transformacji rozwojowej (4-3) w zintegrowanym środowisku telekomunikacyjnym, pozwalający na analizę, koncepcję twórczą, badania i ewaluację, przedstawia się następująco:

$$L(H, E, R, \Theta, t) = P(s, z, \Theta, t - t_0) \quad (4-3)$$

where:

L, P – reprezentuje lewą i prawą stronę równania, gdzie $L(*)$ reprezentuje charakterystykę systemu, podczas gdy $P(*)$ jest zbiorem interwencji twórczych, które działają na zidentyfikowany system w celu zapewnienia spełnienia celów jakości.

\bar{H} – charakterystyka wydajności jako wartości wyjściowe, np. jakość systemu przesyłowego w sieciach bezprzewodowych (QbP/GoS), (energia, energia, emisje, odpady), ekonomiczna i energetyczna efektywność procesów, nieszkodliwość oddziaływania urządzeń i procesów przesyłowych), szkodliwość dla życia i zdrowia ludzkiego, np. w kontekście cyklu życia urządzeń przesyłowych

\bar{E} – elementy transformacji warunków technicznych transmisji w sieciach bezprzewodowych (QbP), mocy i środowiska (nośniki, techniki transmisji, sieci przesyłowe, urządzenia, instalacje, odpady, emisje)

\bar{R} – relacje, łączenie elementów systemu przesyłowego (przyjmowanie zamówień, ruch, transformacja, transmisja, przetwarzanie, magazynowanie, przechowywanie, zużycie, rozwój, akumulacja, serwis itp.), twórczy wpływ na obiekty przesyłowe (optymalizacja, modernizacja, innowacja; użytkowanie, serwisowanie, naprawy, dostawa, utylizacja - złomowanie; recykling materiałów antenowych, ...) Relacje te mogą podlegać ewolucji w czasie.

\bar{W} – reprezentuje zakłócenia między różnymi użytkownikami

Θ – dynamika analizowanego systemu

$t - t_0$ – czas,

\bar{s} – zmienna reprezentująca kontrolowanie, regulowanie, kompensację, monitorowanie, tworzenie, transformacyjną destrukcję,

\bar{z} – zakłócenia technologiczne, społeczne, środowiskowe, przypadkowe i inne.

Lewa strona modelu, równania (4-3), opisuje właściwości danych, wiedzy, procesu transmisji, jego właściwości fizyczne, właściwe dla danej klasy działań. Właściwości te zależą od charakterystyki elementów E_1, E_2, \dots, E_m , połączeń między tymi elementami R_1, R_2, \dots, R_n i są funkcjami Θ i t , które reprezentują odpowiednio dynamikę procesu i czasu. Niewiadome, to elementy zestawu cech H , jako wielkości wyjściowe, od których zależy szczegółowa ocena jakości, oraz ogólna wydajność, nieszkodliwość - niejednorodność, nieskuteczność, zmienna wydajność, nieuzasadnione zużycie energii, specyficzne zużycie energii itp.

Warunki techniczne koncepcji, architektury i parametrów sieci 4G/5G/6G (Wt_j), będą rozpatrywane przy użyciu wag a, b i c, określających: rolę algorytmów CAC

($a \cdot A$), poziom wsparcia rozwiązaniami AI/ML ($b \cdot O$) oraz poziom niedokładności ($c \cdot N$):

$$SP_i = a \cdot A + b \cdot O + c \cdot N \quad (4-4)$$

gdzie:

A – algorytmy sterowania umożliwiające osiągnięcie stanów postulowanych jakości w sieciach bezprzewodowych (QbP),

O – kontrybucja zastosowania elementów uczenia maszynowego do osiągniętego i -tego stanu postulowanego w sieciach bezprzewodowych (QbP),

N – wpływ niedokładności elementów składowych systemu transmisyjnego na osiągnięcie i -tego stanu postulowanego sieci bezprzewodowej (QbP),

a -, b -, c - waga określająca poziom wpływu poszczególnych składowych stanu postulowanego (0-100)% determinujących ostatecznie osiągnięcie i -tego stanu postulowanego jakości w sieci bezprzewodowej (QbP).

Zaproponowany dla tej pracy matematyczny model systemu transmisji w sieciach bezprzewodowych to:

$$H_u(Q_{prod.}, e_{proc.}, N_{prod.}, N_{proc.}, t_0 - t_1) = f(ZST) \quad (4-5)$$

gdzie:

H_u – zmiana charakterystyk użytkowych, w funkcji zmiennych systemu przesyłowego (ZST)

$Q_{prod.}$ – jakość produktu systemu transmisji w sieciach bezprzewodowych (QbP),

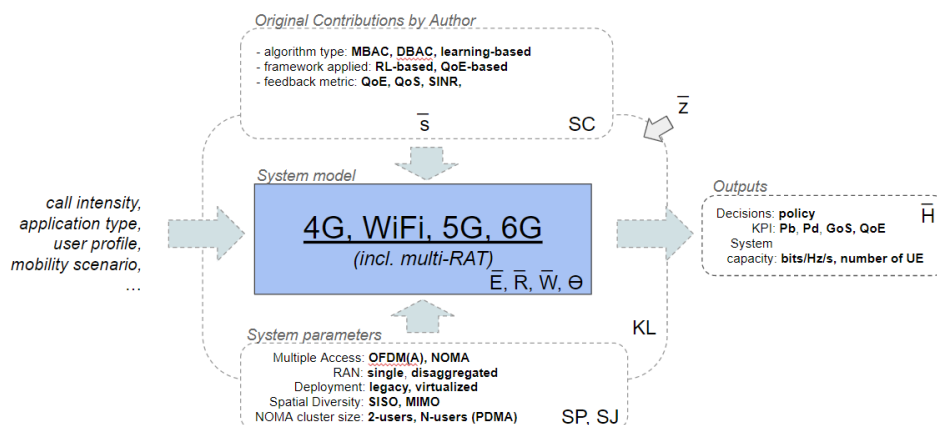
$e_{proc.}$ – sprawność systemu przesyłowego w sieciach bezprzewodowych,

$N_{prod.}$ – nieszkodliwość produktu składającego się na system transmisji bezprzewodowej,

$N_{proc.}$ – nieszkodliwość procesu transmisji w sieciach bezprzewodowych,

$t_0 - t_1$ - czas realizacji transmisji.

Na podstawie wyżej wymienionych zmiennych definiujących zidentyfikowany model systemu (4G/5G/B5G), wszystkie elementy zostały zebrane i przedstawione na diagramie Rysunek 2. Przerywane prostokąty w środkowej części reprezentują cztery podsystemy wprowadzone wcześniej w sekcji (SP, SC, SJ, KL), a zmienne \bar{E} , \bar{R} , \bar{W} , $\bar{\theta}$ reprezentują charakterystykę systemu, a docelowe wskaźniki jakości (tj. zmienne wyjściowe) są identyfikowane jako \bar{H} .



Rysunek 2 Reprezentacja modelowanego systemu bezprzewodowego z uwzględnieniem zmiennych wejściowych i wyjściowych

Ukierunkowane rozwiązanie w zakresie algorytmów sterowania i rozwiązań dla 4G/5G/B5G, które ma cechy zintegrowanego, zaawansowanego, dostosowanego do potrzeb bezprzewodowego systemu transmisji danych, można scharakteryzować przez:

- wysoko wydajny transfer danych
- wysoce wydajną płaszczyznę sterowania
- samoorganizujący się informacyjny system sterowania (np. płynność zużycia zasobów RAN w odpowiedzi na zmieniające się warunki otoczenia)
- wysoce niezawodne zarządzanie i eksploatację sieci (np. poprzez dążenie do optymalnych wartości postulowanej jakości H_U , zwłaszcza gdy zmieniają się warunki brzegowe systemu).

W ten sposób możliwa jest istotna poprawa postulowanych stanów/parametrów w zakresie jakości.

5 MODELE ROZWIĄZANIA PROBLEMÓW

Z perspektywy **rozwiązania problemów badawczych (Problem1 oraz Problem2)** w niniejszej pracy zastosowano modelowanie pojemności systemów radiowych z wykorzystaniem wzorów Shannona dla systemów OFDM. Na etapie modelowania wykorzystano własne rozszerzenia symulatora i emulatora w rozdziale 5, dla potrzeb uwzględnienia informacji o narzucie sygnalizacyjnym różnych klas obsługi. Zastosowanie modelowania narzutu pozwala bardziej realistycznie oceniać dostępne zasoby poszczególnych klas. Zastosowanie tego podejścia jest zasadne dla wyznaczenia wartości maksymalnych pojemności badanych systemów. Ponadto zastosowano znane z literatury modele kanału (Rayleigh, AWGN) ale także modelowanie kanału w oparciu o lokalizację

użytkownika modelowaną za pomocą znanego rozkładu zmiany położenia terminala bezprzewodowego w przestrzeni 2D - tj. Leavy-Walk. Dla właściwego pomiaru wydajności algorytmów sterowania przyjmowaniem zgłoszeń, w zakresie parametrów wydajnościowych (H_u), zastosowano modele ruchu VoIP o stałej przepływności. Aby zapewnić wymaganą stopę błędów przyjęto stałą poziom kodowej stopy błędów (CWER) na poziomie 1% oraz mechanizm AMC. Zaproponowane przez autora algorytmy sterowania przyjmowaniem zgłoszeń (CAC) oraz poziomem przeciążeń (CC) zostały zamodelowane w symulatorach (NS2, Matlab).

Aby zaadresować **problem zdefiniowany w zadaniu 1**, wiarygodnego odwzorowania zachowań sieci na podstawie pomiarów terenowych, zastosowano podejście oparte na pomiarach z zastosowaniem własnego oprzyrządowania zapewniającego odpowiednio niski poziom błędów pomiarowych w zakresie pomiaru parametrów QoS oraz dokładności synchronizacji zegara pomiarowego. Model rozwiązania problemu identyfikacji systemowej zrealizowano w oparciu o oryginalny projekt emulatora sieciowego, opartego na modelach statystycznych rozkładu strat i opóźnień z pomiarów terenowych. Dodatkowo wykorzystany został w pętli pomiar narzutu sygnalizacyjnego z narzędzia symulacyjnego (NS2). Dane wejściowe do emulatora pochodzą z zaproponowanej oryginalnej metodyki generowania rozkładów straty pakietów i opóźnień, a model statystyczny został zweryfikowany i potwierdzony poprzez nieparametryczne testy Mann-Whitney oraz rozkład T-Studenta. Dodatkowo zastosowano podejście pozwalające na weryfikację wpływu emulatora na możliwość wiarygodnej oceny stanów postulowanych (SP_j), wykorzystując w pętli sprzężenia wskaźniki jakości postrzeganej dla ruchu wideo „bez modelu odniesienia” (ang. *no-reference QoE*).

Aby zrealizować zadanie 2 tj. optymalne strategie przełączania ruchu użytkowników pomiędzy sieciami WLAN a LTE (tzw. tryb multi-RAT), zostały zamodelowane w oparciu o eksperyment z użyciem radia programowalnego oraz model pomiaru postrzeganej jakości strumienia wideo usługi IPTV w kierunku w dół, oparty na pomiarze parametrów jakościowych QoS pakietów w sieci. W modelu systemowym zastosowano podejście, w którym model jakości był wyznaczany na bieżąco poprzez (a) odczyt wartości charakterystyk jakościowych QoS po stronie terminala końcowego (UE) i przeliczanie ich na parametry modelu jakości postrzeganej QoE oraz (b) alternatywnie za pomocą pomiaru poziomu sygnału (SINR). Ustawienia optymalne modelu QoE zostały dostosowane do potrzeb niniejszej pracy na podstawie eksperymentów. Wykorzystano tutaj podejście do sterowania jakością w pętli zamkniętej, wykorzystującej kontroler sieci RAN w postaci tzw. elementu funkcjonalnego near-RT RIC, zgodnego z ORAN. Dodatkowo przyjęty model zakłada możliwość sterowania doбором modulacji i kodowania w oparciu o sprzężenie zwrotne z informacją o stanie kanału (CQI) od użytkownika.

Aby znaleźć rozwiązanie dla **problemu wiarygodnego przewidywania**

zapotrzebowania na zasoby obliczeniowe (i radiowe) określonego w zadaniu 3, autor w rozdziale 7 zaproponował autorskie podejście ramowe do optymalizacji sposobu wdrażania (i skalowania) komponentów sieci RAN na infrastrukturze obliczeniowej, w oparciu o algorytmy AI/ML oraz schemat blokowy. Na potrzeby predykcji wykorzystano i poddano optymalizacji ustawień modele ARIMA, LSTM oraz N-Beats i liniowe modele regresji, aby dopasować model wykorzystania zasobów obliczeniowych (CPU, RAM) do bieżącego poziomu interakcji użytkowników sieci 5G z systemem. Ponadto w rozdziale 7 opisana, jako **rozwiązanie zadania 5**, została propozycja procedury ramowej, wspomagającej systemowe podejście do planowania i optymalizacji ustawień sieci z perspektywy postawionego zadania inteligentnej quasi-optymalizacji.

Do **rozwiązania problemu znajdowania optymalnych polityk dla sieci 4G/5G/B5G, w oparciu o uczenie maszynowe (tj. Zadanie 4)**, wykorzystano model sekwencyjnego podejmowania decyzji za pomocą procesów decyzyjnych Markova (MDP oraz SMDP). Rozwiązanie problemu optymalnego podejmowania decyzji wykorzystuje równanie Bellmana oraz algorytmy uczenia ze wzmocnieniem (SARSA, Q-Learning). Ponadto zaproponowane zostało podejście do przyspieszania i generalizacji procesu uczenia poprzez wykorzystanie sieci neuronowych do wsparcia procesu.

6 PLAN I REALIZACJA BADAŃ

W celu umożliwienia kompleksowego spojrzenia na planowane prace badawcze z perspektywy modelu systemu bezprzewodowego, przedstawionego dla tej pracy na rysunku (Rysunek 2), podsumowanie przewidywanych ustawień różnych podsystemów (SP, SC, SJ, KL) przedstawiono w Tabeli 1.

Autor zaplanował eksperymenty uwzględniające plan badawczy, oparty na następujących wymiarach: algorytm CAC, schemat kontroli przeciążeń, różne typy ruchu, różne kierunki transmisji (DL, UL), rodzaj stosowanego sprzężenia zwrotnego, a także wiele technologii radiowych (RAT). W tabeli podane zostało odwzorowanie prac badawczych na rozdział pracy. Z kolei bieżący rozdział, jak wcześniej wspomniano, związany jest z rozwiązaniem postawionych w pracy problemów (1,2) oraz wypełnieniu zadań (1-5).

W rozdziałach 4-8 autor koncentruje się na rozwiązaniach na poziomie pojedynczej komórki do realizacji funkcji CAC, tj. nie zakłada się w pracy wykorzystania mechanizmów do wymiany informacji pomiędzy komórkami w trybie każdy-z-każdym, choć proponowane rozwiązania powinny stosować się także do sieci stosujących wirtualizację i centralizację funkcji stosu radiowego i sterowania. Realizacja badań przebiegała w następującej kolejności:

- Badania eksperymentalne (rozdziały: 4, 5, 6) realizowane były w sieciach operatorskich w Polsce
- Badanie symulacyjne (rozdziały: 4, 8) wykonane zostały w oparciu o popularne platformy NS2 oraz Matlab, a autor stworzył oryginalne

rozszerzenia kodu wspierające modelowanie problemów sterowania przyjmowaniem zgłoszeń, a zwłaszcza algorytmy CAC, będące przedmiotem rozwiązania problemów badawczych.

- Badania teoretyczne (rozdział: 5) zostały przeprowadzone w oparciu o modele statystyczne, ale także w oparciu o teorię związaną z modelowaniem pojemności sieci bezprzewodowych.

Tabela 1 Przegląd planu badawczego

Chapter	Topics covered					
	Admission	Congestion	Traffic	Direction	Feedback	RAT
Chapter 1 (DBAC, MBAC)	declaration and measurement based	Yes	VoIP	DL	QoS, GoS	4G, 5G
Chapter 2 (E2E congestion)	No	adaptive traffic + control	real-time video	UL	QoE metrics	4G
Chapter 3 (multimedia)	No	traffic steering	non-priority traffic	DL	QoE metric X, SINR	4G, WiFi
Chapter 4 (workload prediction)	Yes (computing level)	Yes (computing level)	YT, FTP	DL/UL	CPU utilization	disaggregated virtual 5G RAN
Chapter 5 (AI/ML based CAC)	Yes (RL based)	No	CBR (UGS, BE)	DL	QoS	4G, 5G

Autor wykonał **badania eksperymentalne** na początku procesu badawczego, aby odnieść się w następujących badaniach do rzeczywistych wyników, które zostały zebrane w pomiarach terenowych. Rolą badań eksperymentalnych było poznanie i zrozumienie warunków panujących w sieci jako stan zastany dla pojedynczych połączeń. Szczególnie interesujące dla autora było praktyczne doświadczenie faktycznej (mierzalnej) zmienności jakości postrzeganej w trakcie wstępnych badań, i tak:

- Badania eksperymentalne (rozdziały 4,5) w sieci INEA/Choszczno oraz w lokalnej stacji bazowej PBS, realizowane były z wykorzystaniem faktycznych modemów oraz generatora i analizatora ruchu pakietowego (rozdział 4) oraz po uprzedniej synchronizacji zegarów w terminalach testowych
- Badania eksperymentalne (rozdział 6) z zakresu sieci multi-RAT przeprowadzone zostały z wykorzystaniem zdalnego laboratorium, gdzie dostępna była kompletna instalacja (tj. warstwy wyższe stosu oraz kompletna warstwa fizyczna), wykorzystująca sprzęt National Instruments (urządzenia SDR) oraz sieci WLAN/LTE.
- Badania eksperymentalne (rozdział 7) w sieci 5G ORAN przeprowadzono w laboratorium z użyciem rzeczywistej sieci, celem pozyskania danych o poziomie wykorzystania pamięci, procesora, dysku na maszynie, na której uruchomione zostało zvirtualizowane środowisko

protokołów radiowych stosu 5G. Instalacja wykorzystywała środowisko Kubernetes, jako platformę wirtualizacyjną. Wykorzystano standardowe telefony obsługujące standard 3GPP Rel.15 w zakresie 5G dla częstotliwości 3,5GHz.

Badania symulacyjne w rozdziale 4 oraz 8 miały na celu i) zbudowanie wiedzy o przedmiocie badań, zwłaszcza w sytuacji większej liczby terminali i połączeń, ale także ii) analizę i zrozumienie wpływu poszczególnych parametrów wejściowych na parametry wyjściowe (jakościowe) w modelowanej sieci. W początkowym etapie badań symulacyjnych, autor pozyskał lepsze rozumienie technologii i głównych relacji (R) pomiędzy podstawowymi elementami systemu radiowej sieci dostępowej. W efekcie realizacji tej grupy badań powstały docelowe rozszerzenia do symulatorów NS2 oraz Matlab, które autor opisał w artykułach. Ta grupa badań związana była z implementacją wstępnych wersji modernizowanych algorytmów, a następnie wersji docelowych.

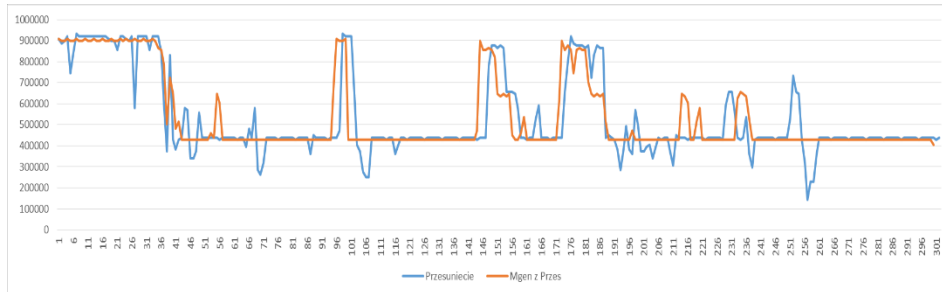
W ramach **badania teoretycznych** autor skoncentrował się w rozdziale 5 na stworzeniu modelu statystycznego rozkładu opóźnień i strat dla pomiarów zebranych w sieciach INEA/Choszczno/PBS. Szczególnie istotne było pozyskanie modelu opóźnień, w związku planem zastosowania tych wyników do budowy wersji emulatora wg oryginalnego pomysłu. Ponadto modele opóźnienia nie są wspierane w modelach L2S dostępnych w literaturze.

7 WYNIKI PROCEDUR DZIAŁANIA BADAWCZEGO

Zaplanowane i przeprowadzone przez autora **badania doświadczalne**, posłużyły głównie do zebrania danych odnośnie jakości połączeń dla różnych strumieni ruchu. Ponadto możliwe było zebranie prawdopodobieństw przejść pomiędzy stanami oznaczającymi różne modulacje (np. Tabela 31 oraz Rysunek 61 w pracy). Udało się wykazać zależność pomiędzy rosnącą przepływnością a malejącymi opóźnieniami transmisyjnymi dla różnych prędkości poruszania terminala, co pokazuje, że w przypadku większej liczby pakietów/sekundę (tj. dla przepływności 1024Kb/s) większa liczba pakietów była przesyłana w trakcie jednej ramki TDD (Rysunek 63 w pracy). Potencjalnie także mechanizmy oszczędzania energii po stronie nadajnika, mają swój wkład w obserwowane zachowanie.

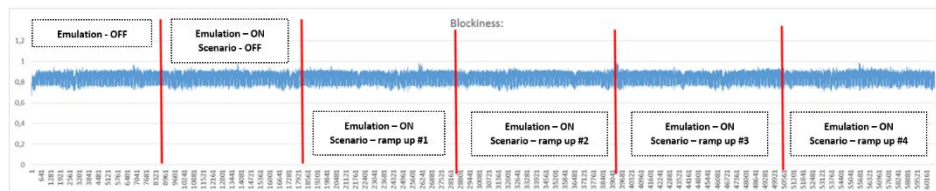
W zakresie modeli teoretycznych zaproponowano model statystyczny (procedura na Rysunku 69 w pracy) oparty o pomiary eksperymentalne, służący do symulacji poziomu opóźnień i strat w sieciach mobilnych. Model ten został opracowany w oparciu o oryginalną procedurę generowania rozkładu strat/opóźnień pakietów, na podstawie pomiarów. Testy statystyczne (T-Studenta, Manna-Whitney'a) potwierdziły wiarygodność modelu. Stworzony model został zintegrowany z emulatorem kanału radiowego własnego projektu, i pozytywnie zwalidowany – zestawiono pomiary wartości oryginalnych (zmierzonych w terenie) oraz wielkości wynikające z zastosowanych modeli strat/opóźnień wpiętych do emulatora. Poniższe wykresy przedstawiają

porównanie pomiędzy tymi przypadkami dla i) chwilowej zmienności przepływności kanału odtwarzanej w emulatorze, ii) różnych metryk QoE które były wykorzystywane do porównania jakości wideo dla sytuacji z zastosowaniem emulatora i rzeczywistej sieci (Rysunek 3, Rysunek 4, Rysunek 5).

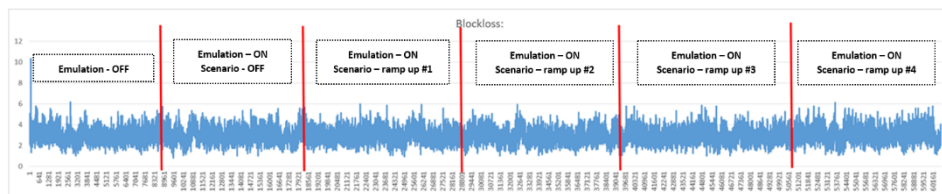


Rysunek 3 Porównanie wartości przepływności z pomiaru (pomarańczowa linia) oraz z emulatora (linia niebieska)

Linie niebieska i pomarańczowa na wykresie reprezentują przepływność chwilową (bit/s) są praktycznie tożsame co do kształtu, występuje jedynie przesunięcie kilku pakietów. Oznacza to, że z perspektywy efektu końcowego po stronie odbioru przesyłanego strumienia wideo jakość postrzegana będzie wiernie odwzorowywała warunki rzeczywiste.



Rysunek 4 Blockiness metrics comparison between test cases

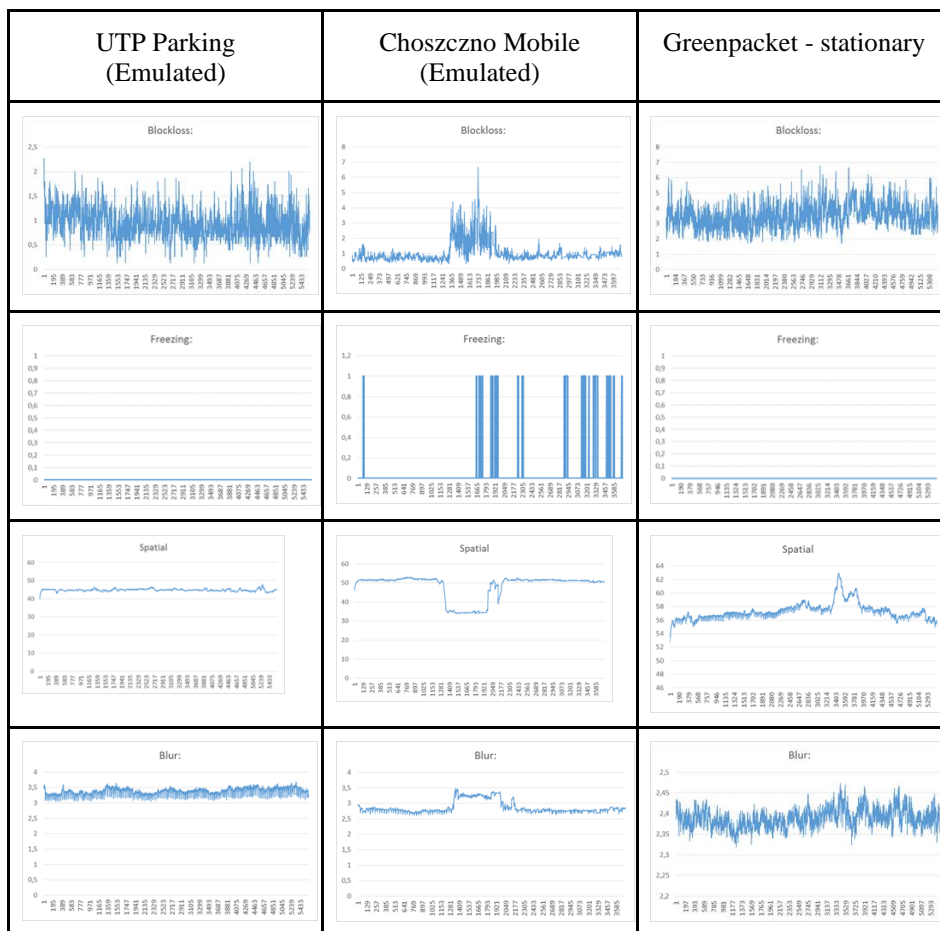


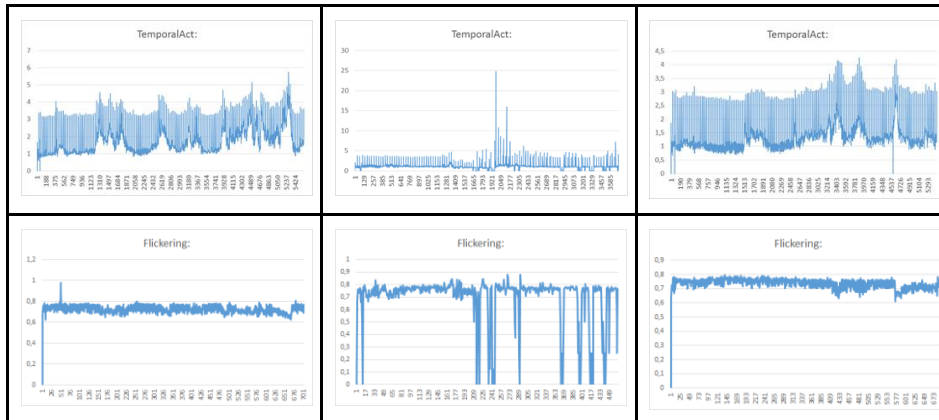
Rysunek 5 Blockloss metrics comparison between test cases

Jak można zauważyć na rysunkach (Rysunek 4, Rysunek 5) porównanie pomiędzy wykresami podczas przełączania ruchu w testach pomiędzy emulatorem a realnymi danymi z pomiaru (tj. testy rzeczywiste), pokazuje że poziomy wartości metryk jakościowych QoE (blockiness, blockloss) są

praktycznie niezmienione pomiędzy poszczególnymi etapami. Oznacza to, że także ze strony użytkownika końcowego nie ma widocznej różnicy w odbiorze i dekodowaniu strumienia wideo, w przypadku wykorzystania zaproponowanego emulatora kanału radiowego. Ponadto w tabeli (Tabela 2) pokazano zestawienie pomiarów dla zebranych metryk porównując pomiar za pomocą modemu (Greenpacket) oraz emulacji (UTP Parking, Choszczno Mobilne). Pomiary dla dwóch skrajnych kolumn reprezentują identyczne warunki radiowe. Jak widać zastosowanie emulatora zachowuje statystyki zmienności parametrów na poziomie QoE. Przesyłany ruch to 1Mb/s w kierunku „w górę” a emulator uwzględnia jednocześnie modelowanie opóźnień i przepływności zgodnie z rozkładem.

Tabela 2 Porównanie pomiarów metryk QoE dla różnych scenariuszy (a/b – emulator, c – rzeczywisty pomiar)





Wyniki badań z użyciem symulacji w środowisku NS2 i Matlab, użyte do oceny jakości algorytmów przyjmowania zgłoszeń i rezerwacji symboli, zostały przedstawione poniżej. Zastosowany mechanizm rezerwacji zasobów na poziomie symboli/slotów OFDM, umożliwia zapewnienie dodatkowej kontroli zasobów w powiązaniu z mechanizmami sterowania przyjmowaniem zgłoszeń. Jest on odpowiednikiem szeregowania zasobów w trybie „z pamięcią”, gdzie dedykowana część zasobów może być zarezerwowana dynamicznie jeśli algorytm decyzyjny aktywuje odpowiednią politykę.

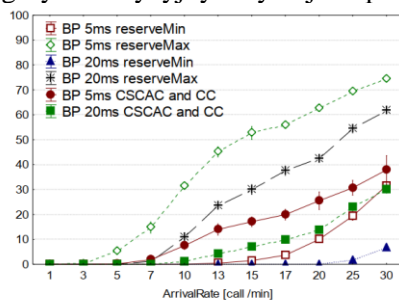


Fig. 1: Average blocking probability for frames 5 ms and 20ms

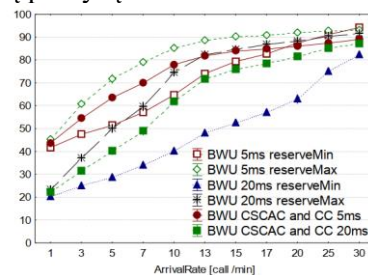


Fig. 4: Bandwidth Utilization (as perceived by AC algorithm) for frames 5 ms and 20 ms

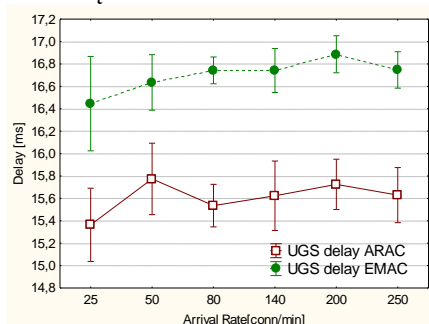
Rysunek 6 Średnie p-stwo blokady dla ramek 5 ms i 20 ms

Rysunek 7 Wykorzystanie pasma (postrzegane przez algorytmy AC) dla ramek 5 ms i 20 ms

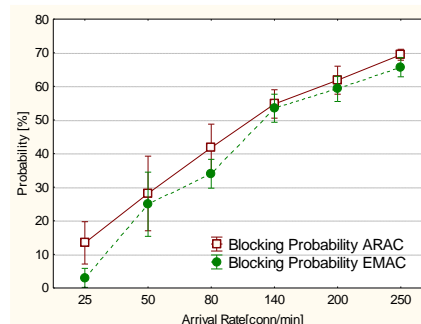
Powyższe rysunki (Rysunek 6 i Rysunek 7) pokazują że niezależnie od konfiguracji warstwy fizycznej (ramka OFDM) prawdopodobieństwo blokady może być kontrolowane poprzez zastosowanie schematów rezerwacji zasobów radiowych (tzw. schematy SRS w pracy). Zastosowanie rezerwacji maksymalnej liczby symboli (tj. dla najniższej wartościowej modulacji) prowadzi do zmniejszenia korzyści z multipleksacji statystycznej strumieni użytkowników, natomiast włączenie mechanizmu sterowania przeciążeniami, może istotnie obniżyć efektywne p-stwo blokady nowych połączeń. Zaletą rezerwacji zasobów

ewidentnie będą zmniejszone opóźnienia transmisyjne.

Rezultaty porównania algorytmu autora (ARAC) z algorytmem bazowym EMAC pokazują, że pozwala on osiągać niższe opóźnienia w transmisji głosowej (VoIP) dla szerokiej gamy intensywności napływu zgłoszeń. Na skutek wbudowanego mechanizmu kompensacji połączeń niedawno rozpoczętych/zakończonych, ale także kompensacji zmiany modulacji (AMC) mechanizm może oferować wyższe prawdopodobieństwo odrzucenia połączeń przenoszonych. Jest to konsekwencją zabezpieczenia połączeń realizowanych już obecnie w komórce przed napływem ruchu z sąsiednich komórek.

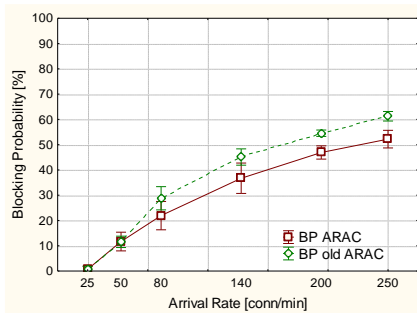


Rysunek 8 Opóźnienia w klasie obsługi UGS/GBR dla ruchu o stałej przepływności VoIP dla algorytmów ARAC i EMAC

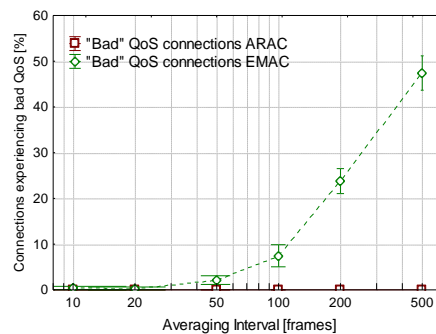


Rysunek 9 Prawdopodobieństwo blokady nowego połączenia dla algorytmów ARAC i EMAC

Prace nad modernizacją algorytmu ARAC doprowadziły do obniżenia prawdopodobieństwa blokady, które w badanych scenariuszach testowych było niższe od rozwiązania oryginalnego nscARAC dla rosnącej intensywności napływu zgłoszeń. Także jeśli chodzi o porównanie z bazowym mechanizmem ruchomego okna (EMAC) rozwiązanie przedstawione w pracy (ARAC) zapewnia gwarancje jakości niezależnie od długości okresu uśredniania w oknie obserwacji (10-500 ramek OFDM).

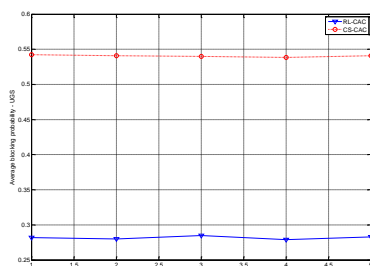


Prawdopodobieństwo blokady połączeń dla algorytmów ARAC i nscARAC (scenariusz 2b)

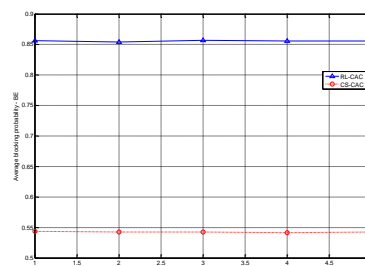


Procent połączeń, dla których nie zapewniono wymaganego poziomu QoS - ARAC vs EMAC

Z kolei algorytm CAC wykorzystujący uczenie ze wzmocnieniem, zaproponowany w rozdziale 8, umożliwia optymalizację prawdopodobieństwa blokady zgodnie z funkcją nagrody określoną dla algorytmu. Uzyskane prawdopodobieństwa blokady i odrzucenia połączenia zostały pokazane na rysunkach Rysunek 10 i Rysunek 11. Jak widać wartości prawdopodobieństw wyznaczone przez algorytm oparty o model MDP i uczenie ze wzmocnieniem dla klasy obsługi priorytetowej (UGS/GBR) osiągają lepsze wartości niż tożsame prawdopodobieństwa dla algorytmu CS-CAC. Poprawa prawdopodobieństw blokady ma miejsce kosztem pogorszenia prawdopodobieństw odrzucenia dla klasy best-effort (BE). Ma to miejsce dlatego że algorytm poszukując optymalnych wartości decyzji o przyjęciu bądź odrzuceniu połączeń kieruje się maksymalizacją długofalowych korzyści, które są motywowane wartością nagrody w danym stanie.



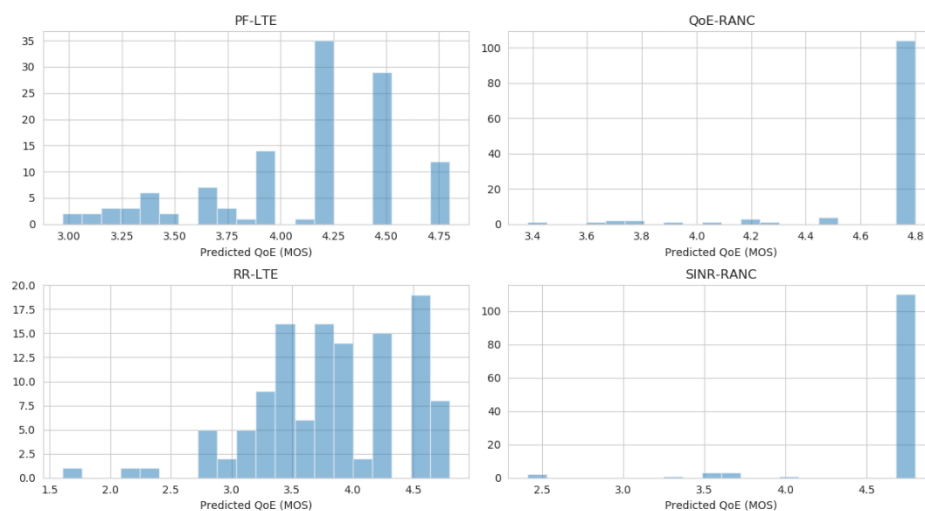
Rysunek 10 Średnie p-stwo blokady dla klasy UGS/GBR (algorytm RL-CAC i CS-CAC)



Rysunek 11 Średnie p-stwo blokady dla klasy BE (algorytm RL-CAC i CS-CAC)

Wyniki badań nad poprawą jakości połączeń wideo z zastosowaniem zaproponowanych przez autora (rozdział 6) algorytmów przełączania ruchu wideo (IPTV) pomiędzy sieciami WiFi oraz 4G, zostały przedstawione na

wykresie poniżej.



Rysunek 12 Histogram wartości MOS dla rozwiązania autora (prawa kolumna) oraz dla rozwiązania multi-RAT

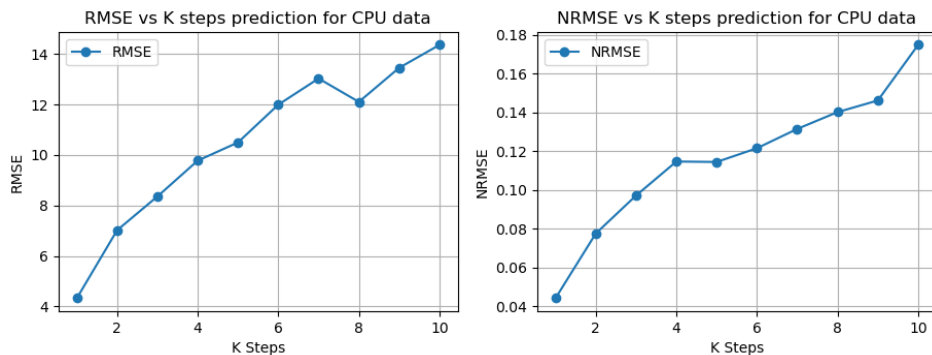
Na rysunku (Rysunek 12) widoczna jest przewaga inteligentnego przełączania ruchu użytkownika w sytuacji, gdy aktywne są zaproponowane algorytmy adaptujące jakość QoE-RANC oraz SINR-RANC na podstawie pomiaru jakości postrzeganej oraz poziomu sygnału. Wykresy pokazują, że transmisja będzie znacząco poprawiona tam, gdzie istnieje możliwość inteligentnego sterowania przełączaniem ruchu pomiędzy siecią mobilną (4G/5G), a siecią WiFi w paśmie nielicencjonowanym. Możliwości te będą tym wyższe jeśli sieć WiFi będzie wykorzystywać mechanizm podziału czasowo-częstotliwościowego OFDMA (np. WiFi6).

Autor w **rozdziale 7** zaproponował ramowe podejście do quasi-optimalizacji mechanizmów sterowania alokacją zasobów w sieciach mobilnych ORAN, z zastosowaniem algorytmów uczenia maszynowego. Zaproponowany model dla predykcji obciążenia procesora (CPU) przez oprogramowanie sieciowe 5G RAN w zastosowaniu z przenośnym centrum danych (EMDC), wykorzystuje sieci neuronowe LSTM w procesie wspomaganie decyzji agenta CAC dla sieci 5G ORAN. Najpierw przeprowadzono zbieranie danych dla zużycia CPU w oparciu o scenariusze testowe. Następnie wytrenowano modele LSTM, różniące się architekturą i liczbą warstw. Wyniki walidacji predykcji dla różnych danych wejściowych i różnych modeli (liniowej i nieliniowej) pokazano poniżej w Tabeli 3. Uzyskane poziomy błęd średniokwadratowy (RMSE) są znacząco niższe dla modeli LSTM, w porównaniu z modelami ARIMA, N-Beats, oraz innymi popularnymi regresorami liniowymi.

Tabela 3 Porównanie wartości błędu (RMSE) dla różnych technik predykcji zużycia CPU

Model	RMSE	Model	RMSE
LSTM1	9	Random Forest	19
LSTM2	4.6	Decision Tree	20
XGBOOST	18	Linear Regression	16
Bagging	19		

Ponadto autor przeprowadził analizę wpływu długości kroku predykcji na poziom błędu RMSE stworzonych modeli predykcji obciążenia CPU przez stos sieciowy 5G vRAN. Wyniki zostały przedstawione poniżej. Jak widać błąd dla predykcji wprzód z krokiem $k=10$ jest nadal niższy niż błąd RMSE dla dowolnej techniki regresji liniowej.



Rysunek 13 Błąd średniokwadratowy predykcji wykorzystania procesora dla różnej długości kroku predykcji $k=\{1-10\}$.

Praktyczne znaczenie powyższego wyniku należy rozumieć jako dodatkowy czas na wykonanie czynności niezbędnych po stronie np. mechanizmu xApp, żeby uruchomić skalowanie poziome płaszczyzny użytkownika w module CU-UP do drugiego serwera edge (EMDC).

8 PODSUMOWANIE OSIĄGNIĘĆ I WNIOSKI

W tym rozdziale podsumowane zostały wyniki w nawiązaniu do celów przedstawionych w rozdziale 3. Analiza wiąże się tu bezpośrednio z postulowanym stanem (SP_i), na który ukierunkowane są rozwiązania opracowane w ramach tej pracy kluczowe pytania badawcze hipotezy o odbiorze sieci wydajności i kontroli zatorów, która jest osiągalna dzięki modernizacjom zaproponowanym przez autora. Poniżej w Tabela 4 autor zebrał osiągnięcia z

rozdziałów z wkładem własnym (4-8). Tabela identyfikuje cele modelowania i związane z nimi osiągnięcia, dostarczone w wyniku prac badawczych, które doprowadziły do uzyskania odpowiedzi na pytania badawcze i potwierdzenia tezy.

Tabela 4 Zestawienie podejść analitycznych w rozdziałach 4-8

Rozdział	Cel modelowania	Osiągnięcie
4,8	Analiza decyzji agentów CAC, dotyczących jakości w sieciach bezprzewodowych 4G/5G/B5G w formie oceny metryki Grade of Service (P_B, P_D, BW_{util}).	Wykorzystanym narzędziem jest symulacja sieci 4G/5G (4) oraz uczenie się przez wzmacnianie jako przybliżenie funkcji (8) w celu zapewnienia zmodernizowanych algorytmów kontroli dostępu, również z systemami rezerwacji symboli, dla przyszłych sieci bezprzewodowych
5	Procedura analizy statystycznej w celu dopasowania rozkładu opóźnień i strat pakietów, na podstawie danych z eksperymentów (pomiar opóźnień i strat w sieciach różnych operatorów)	Wynikiem pracy jest: i) statystyczny model opóźnień i strat pakietów dla kanału radiowego zaimplementowany w formie makro w programie XLS ii) autorskie rozwiązanie do emulacji sieci 4G/5G w związku z symulacją narzutów sygnalizacyjnych dla sieci 4G/5G
6	Testowanie jakości obserwowanej przez użytkownika (QoE) w oparciu o model wykorzystujący obiektywne metryki wynikające z pomiaru sieci (tj. przepustowość, straty, opóźnienia, jitter) w celu wsparcia decyzji o przełączaniu strumieni ruchu między sieciami 4G / WiFi	Analiza statystyczna polega na analizie chwilowego i statystycznego zachowania wartości wydajności QoE postrzeganej przez użytkownika, ale opartej na obiektywnych pomiarach sieciowych w przypadku scenariusza multi-RAT (6, rys. 16, 17)
7	Analiza szeregów czasowych w zakresie metryk wykorzystania mocy obliczeniowej (w postaci metryk CPU) zużywanej przez moduły oprogramowania sieci 5G ORAN, tj. CU, DU, RU. Celem analizy jest przewidywanie przyszłych wartości obciążenia w celu wsparcia decyzji dotyczących skalowania obliczeń CU. To ostatnie ma potencjał bezpośredniego oddziaływania na poprawę pojemności przyszłych sieci (zwłaszcza w zakresie	This chapter provides various prediction techniques to improve the admission capabilities with regards to scaling modern 5G/beyond 5G workloads among the edge servers, when local capacity is drained. For model quality metric the following metrics are used: RMSE, accuracy of prediction (for k-steps prediction). W tym rozdziale przedstawiono różne techniki przewidywania mające na celu poprawę możliwości przyjmowania zgłoszeń, w

	dopasowania obciążeń do dostępnej pojemności centrów obliczeniowych (np. EMDC).	odniesieniu do skalowania nowoczesnych obciążeń 5G/B5G pomiędzy serwerami na brzegu sieci, gdy pojemność lokalna jest niewystarczająca. Do oceny jakości modelu zastosowano typowe metryki: RMSE, dokładność przewidywania (dla przewidywania k-kroków). Metody aproksymacji nieliniowej: LSTM Metody korelacyjne: ARIMA, N Beats Inne metody regresji: XGBoost, Bagging, Random Forrest, Decision Tree, Linear Regression
--	---	---

8.1 Sformułowanie wniosków

Jak widać w powyższych rozdziałach, autor z powodzeniem poradził sobie ze zdefiniowanymi problemami badawczymi (Problem 1-2) oraz zadaniami pomocniczymi (Zadanie 1-5). Należy stwierdzić, że zaplanowane cele zostały osiągnięte. Autor dostarczył różne modele i algorytmy, które umożliwiły odpowiedź na postawione problemy i zadania badawcze, a tym samym przyczyniły się do zaadresowania wyzwań dla sterowania przyjmowaniem zgłoszeń i przeciążeń w sieciach przyszłości (Tabela 43 w pracy). Algorytm ARAC zapewnia konfigurowalne rozwiązanie, które może zaspokoić potrzeby operatorów, zwłaszcza w obecnie zachodzącej ewolucji sieci (5G/6G). Ponadto, w połączeniu ze schematami rezerwacji symboli, może wspierać rezerwację zasobów radiowych (ang. *persistent scheduling*), jak również służyć do kontroli przeciążeń. Dzięki zaproponowanym w algorytmie ARAC modernizacjom, można zarządzać dynamiką zużycia zasobów charakteryzującą przyszłe sieci.

W celu zwiększenia elastyczności definiowania celów sterowania w oparciu o uczenie maszynowe, zastosowanie uczącego agenta MDP jest wygodnym sposobem optymalizacji. Szczególnie interesująca jest elastyczność definiowania celów algorytmu i działań, które może on uwzględniać. Skalowanie zasobów zdezagregowanych sieci, wykorzystujących sieci 5G/B5G ORAN, jest w tej zdaniem autora szczególnie interesujące.

Ponadto rozwiązania multi-RAT, jak wskazano w rozdziale 6, zapewniają ważne wsparcie dla poprawy jakości strumieni wideo i odciążania sieci mobilnych, zwłaszcza w połączeniu z zastosowaną pętlą sprzężenia zwrotnego (QoE lub SINR), w kierunku kontrolera RAN, który może analizować stan takiego sprzężenia zwrotnego i na tej podstawie może dostosować wybór podstawowych technologii RAT. Takie podejście może być odpowiednie dla

dowolnych technologii RAT, które mogą podlegać inteligentnemu przełączaniu (np. łączność z wykorzystaniem 4G oraz 5G, agregacja nośnych), co może zapewnić więcej opcji przenoszenia ruchu dla operatorów.

Zaproponowano ponadto oryginalną, budżetową konfigurację emulatora laboratoryjnego (rozdział 5), który umożliwia ocenę algorytmów kontroli przeciążeń (i kontroli dostępu), wykorzystując opracowany przez autora model statystyczny. W kompleksowym zestawie testów wykazano, że taka struktura może stanowić przydatne narzędzie, zwłaszcza przy projektowaniu algorytmów sterowania dostarczaniem wideo w kierunku „w górę”. Jest to szczególnie atrakcyjne w dzisiejszych czasach, gdy na horyzoncie pojawia się coraz więcej rozwiązań typu proof-of-concept dla zdalnej kontroli pojazdów. Nawet jeśli nie do wspomagania na etapie jazdy z dużymi prędkościami, ze względu na problemy z zasięgiem sieci, zdecydowanie sprawdza się w przypadku zdalnej pomocy, gdy pojazd nie jedzie autonomicznie.

Dostarczone zostało kompletne rozwiązanie predykcji obciążeń procesora opracowane, ocenione i zaimplementowane w postaci skryptów języka Python (opartych na modelach wielu różnych predyktorów – liniowych i nieliniowych). Rozwiązanie to jest szczególnie interesujące ze względu na skuteczne wsparcie przyszłego skalowania sieci, którego znaczenie będzie rosło w związku z przewidywanym wzrostem roli technologii brzegowych (np. EMDC), co zostało zaprezentowane w poprzednich rozdziałach. Predykcja umożliwi wydłużenie czasu na realizację niezbędnych działań wymaganych do faktycznego przeprowadzenia skalowania, pomiędzy węzłami na brzegu sieci (np. EMDC).

9 PODSUMOWANIE

Na podstawie przeprowadzonej analizy, badań własnych sformułowano problemy a w celu udowodnienia tezy sformułowano wnioski z postępowania poznawczego, praktycznego i dalszego.

9.1 Wnioski o charakterze poznawczym

Dzięki pomyślnemu rozwiązaniu problemów badawczych wskazanych w rozdziale 1 (Problem1, Problem2) autor osiągnął główne cele pracy. To z kolei umożliwiło zmierzenie się z główną hipotezą pracy zaprezentowanej w niniejszej rozprawie.

Poprzez ukierunkowanie się na Problem1, możliwe było zrealizowanie drugiego celu głównego, dzięki ocenie stanu istniejących dotychczas algorytmów CAC, zidentyfikowaniu kierunków ich rozwoju, zaproponowaniu unowocześnionych rozwiązań (algorytm ARAC, rama do oceny QoE, predykcja wykorzystania procesora za pomocą sieci LSTM dopasowana do zwirtualizowanych 5G/B5G ORAN sieci, agent RL-CAC uczący się optymalnych polityk CAC online za pomocą algorytmów Q-Learning i SARSA wspartych przez zastosowanie sieci neuronowych (ANN). Rolą sieci neuronowych jest skrócenie czasu uczenia, a także inteligentne i oparte na QOE

przełączanie ruchu wideo między sieciami dostępowymi z licencją komórkową i nielicencjonowanymi (opartymi na Wi-Fi) (rozdział 6). Proponowane rozwiązania składają się na portfolio rozwiązań, które są nie tylko przydatne w przejściu z obecnych sieci 4G na 5G, ale zdają się być wystarczająco solidne, aby zapewnić ciągłą ewolucję w kierunku sieci 6G w najbliższej przyszłości. Cel, który za tym stoi, był możliwy dzięki zaproponowaniu zmodernizowanych algorytmów CAC (mDHCAC, ARAC), które nie tylko poprawiają jakość w istniejących sieciach 4G/5G, ale są odpowiednie do radzenia sobie z ich ewolucją w przyszłości – zwłaszcza biorąc pod uwagę trendy sieciowe i) zagęszczenie, ii) scentralizowana kontrola i kooperacyjne RRM, iii) zwiększone wykorzystanie i rola rozwiązań AI/ML, zwłaszcza w płaszczyźnie sterowania.

Realizacja **Zadania 1**, o którym mowa w rozdziale 1, została osiągnięta poprzez wykazanie, że możliwe, wartościowe i przyszłościowe jest dostarczenie modeli (model statystyczny strat i opóźnień pakietów w bezprzewodowym kanale mobilnym), które można wdrożyć w standardowo wyposażonym laboratorium - oparte na środowisku, aby poradzić sobie z modelowaniem łączy bezprzewodowych o wysokiej wierności i projektować kontrolery wideo oparte na QoE/QoS (i obsługiwane przez lokalne informacje zwrotne radiowe). Wynik ten ma szczególne znaczenie dla przypadków użycia z pojazdami zdalnie sterowanymi. Realizując **Zadanie 2** – autor zdefiniował ramy skutecznego działania mechanizmu inteligentnego przelewania ruchu przedstawionego w rozdziale 6, który może być bardzo przydatnym narzędziem dla operatorów, szczególnie w przypadku, gdy obecnie na konkurencyjnym rynku pojawiają się inicjatywy zachęcające abonentów do stania się również operatorami lokalnych sieci (Wi-Fi, 5G) – np. inicjatywa firmy Helium. Proponowany algorytm (Rysunek 82) może wspierać zwykłe sieci w codziennej pracy, ale może być również odpowiednim rozwiązaniem dla lepszego dostarczania przepustowości na potrzeby zarządzania kryzysowego, gdzie dostarczanie przepustowości ad-hoc (np. agregacja pasma, dual-connectivity, multi-RAT) może być obiecującym rozwiązaniem, prowadzącym do lepszej kontroli QoE przy wsparciu inteligentnego przełączania pomiędzy technologiami radiowymi przez algorytmy oparte na QoE/SINR w kontrolerze RAN (np. RIC).

Proponowane rozwiązanie jest w rzeczywistości odpowiednie również dla sieci wykorzystujących mechanizm „dual-connectivity”, jak również agregację nośnych, ponieważ mają one wspólne cechy z kontekstem zastosowania multi-RAT. Głównymi różnicami jest tutaj rodzaj widma oraz interfejsy i sposób aktywacji dodatkowych zasobów. Jeśli chodzi o ujednolicenie podejścia, wdrożenie algorytmów QoE-RANC i SINR-RANC - jako xApps może w dużym stopniu wesprzeć ten cel. Badania nad efektywnym algorytmem predykcyjnym, mającym na celu realizację celów **Zadania 3** w zakresie skalowania zdezagregowanych sieci RAN teraz i w przyszłości, wykazały, że odpowiednio dostrojony mechanizm LSTM zastosowany do przewidywania obciążenia procesora, jest w stanie zapewnić dodatkowy czas na wykonanie czynności OAM, po stronie operatora, aby przygotować udostępnianie większej liczby

instancji wybranych protokołów (CU-UP w tej pracy) na żądanie i dobrze dopasowane do istniejących narzędzi orkiestracyjnych (np. Kubernetes). Połączenie proponowanej predykcji szeregów czasowych z rozmieszczeniem obciążenia w oparciu o modele MDP/RL przedstawione w rozdziale 8, dodatkowo przyczyniłoby się do realizacji celów optymalizacji, poprzez proste rozszerzenie definicji stanu w algorytmie uczącym (ang. *state*), o działania (ang. *actions*) związane ze skalowaniem CU-UP. W ten sposób dalsza konfiguracja i synergia algorytmów uczących z uwzględnieniem różnych funkcji płaszczyzny kontrolnej (przewidywanie obciążenia, rozmieszczenie obciążenia), ale ujednoczone w ramach wspólnego algorytmu RL-CAC, może przynieść dodatkowe korzyści klientom i operatorom. Takie podejście naturalnie współistnieje z zarządzaniem siecią opartym na zasadach (i intencjach), dostępnym za pośrednictwem już istniejących interfejsów, m.in. A1 (ORAN).

Osiągnięcie celów wskazanych w **Zadaniu5** poprzez dostarczenie metodologii umożliwiającej powyższe i przyszłe modernizacje algorytmów RRM płaszczyzny kontrolnej (niezależnie od tego, czy są one związane z optymalizacją, modernizacją czy innowacją) zostało zrealizowane poprzez wprowadzenie ram dla procedur przewidywania i rozmieszczania obciążeń opartych na technikach uczenia maszynowego. Ta struktura dobrze pasuje do obecnych i przyszłych architektur platform przetwarzania na brzegu sieci. Takie platformy charakteryzują się wspólnymi mianownikami: natury opartej na danych, działania opartej na uczeniu się, podejścia międzywarstwowego (w celu bezkonfliktowych kierunków optymalizacji) itp. Rysunek 95 w pracy, przedstawia proponowaną ewolucję sposobu, w jaki takie mechanizmy powinny być projektowane i współistnieć - należy rozważyć zależności między nimi. Niemniej jednak należy pamiętać i obecnie możliwość osiągnięcia wysokiej efektywności energetycznej i minimalizacja zużycia energii jest bardzo istotna żeby zapewnić samowystarczalność energetyczną sieci.

9.2 Wnioski praktyczne

Rozwiązania dostarczone przez autora zapewniają skonsolidowaną strukturę projektowania i optymalizacji wdrażania, która może być interesującym zestawem modeli i narzędzi dla przyszłych operatorów sieci na małą skalę, w przypadku których procesy obliczeniowe wymagane przez stos RAN 5G/B5G w konkurują o zasoby obliczeniowe aplikacji uruchomionych na brzegu sieci. Dynamiczna skalowalność zasobów (na początek CU-UP) daje operatorom dodatkową dźwignię dla bardziej zrównoważonych wdrożeń sieci. Niedrogie i sprawdzone w terenie rozwiązania laboratoryjne wspierające wydajne projektowanie algorytmów sterowania dostarczaniem wideo w kierunku „w górę”, będą interesującym rozszerzeniem przyszłych zastosowań, w których m.in. drony można instruować sondami QoS/QoE w celu regularnego skanowania obszaru pokrycia w celu dostrojenia algorytmów dla dynamicznych (i potencjalnie również uczących się) kontrolerów wideo, które mają na celu poprawę jakości zdalnego sterowania i m.in. ocena sytuacji kryzysowej.

9.2.1 Synteza kontrolerów wideo

Wykorzystanie projektu emulatora wspomagającego projektowanie kontrolerów wideo, dla różnych zastosowań, będzie mogło być przeprowadzone za pomocą rozwiązań do modelowania kanału. Autor potwierdził to zastosowanie głównie z perspektywy scenariuszy bezpieczeństwa (dla których ruch przesyłany jest w górę sieci, a celem jakościowym jest skuteczna możliwość realizacji zadań wymagających rozpoznawania przez operatora niebezpiecznych sytuacji w strumieniu wideo). Badania doświadczalne pokazały, że zastosowanie transkodera modyfikującego strumień wideo w locie może być procesem mocno konsumującym zasoby obliczeniowe, w porównaniu do czystego przesyłania strumienia wideo który nie podlega transkodowaniu. W efekcie stosowanie transkodera do bieżącej adaptacji strumienia wideo może paradoksalnie doprowadzić do pogorszenia jakości, jeśli nie dobrano odpowiednio mocy obliczeniowych do potrzeb transkodera. Pomiar metryk QoE pokazał, że zastosowanie protokołu UDP zwiększa poziom metryki Blockloss, podczas gdy zastosowanie TCP zwiększa ryzyko zacięć obrazu (głównie w związku z retransmisjami). Należy podkreślić, że dynamiczna zmiana przepływności strumienia wideo wymaga powiązanej z nim zmiany rozdzielczości obrazu, aby zachować wysokie wartości metryk jakości QoE. Straty pakietów i jitter mogą prowadzić do kaskadowych błędów w strumieniu wideo MPEG (np. H.264), dlatego opcją jest zastosowanie alternatywnego standardu np. MJPEG. Należy ponadto pamiętać, że sam pomiar QoE jest procesem złożonym obliczeniowo i niezapewnienie wystarczających mocy obliczeniowych dla maszyny, na której takie metryki są wyliczane, może prowadzić do niekontrolowanego pogorszenia metryk QoE.

9.2.2 Wnioski z badań nad predykcją zużycia procesora w serwerach edge/EMDC

W rekomendacjach ORAN odnośnie stosowania mechanizmów AI/ML, modele uczące (ML) muszą zostać wytrenowane i zweryfikowane przed wdrożeniem. Dlatego modele LSTM wprowadzone w rozdziale 7 powinny być trenowane na odpowiedniej do tego maszynie. Taka maszyna może znajdować się w pobliżu RIC lub po stronie komponentu Service Management and Orchestration (SMO), ale poza samym węzłem near-RT RIC lub non-RT RIC. Ponieważ wytrenowanie głębokiego modelu sieci neuronowej wymaga czasu, bardziej realistyczne może być wdrożenie modeli w węzle SMO (poza NRT RIC). Dane (takie jak statystyki i pomiary obciążenia pracą) z serwerów brzegowych powinny być gromadzone i wykorzystywane do uczenia początkowego modelu. W pełni przetestowany i dobrze zweryfikowany model może być następnie wysyłany do platformy near-RT RIC, która działa jako „host wnioskowania i uczenia maszynowego”. W tym miejscu przeprowadzana będzie predykcja. Kontroler non-RT RIC może wysłać gotowe modele do near-RT RIC przez interfejs A1, w celu wymuszenia odpowiednich działań. Egzekwowanie

polityk można zrealizować za pomocą odpowiedniego zestawu aplikacji sterujących xApp w near-RT RIC. Na przykład, przy przewidywanym dużym obciążeniu dla określonego serwera brzegowego, xApp wykorzystywany do sterowania ruchem może odpowiadać za przeniesienie ruchu użytkowników (albo odciążenia instancji RAN w zakresie CU-UP na serwery o niższym obciążeniu) z wyprzedzeniem, aby uniknąć ewentualnego przeciążenia. Wydajność modelu jest monitorowana i ponownie poddawana trenowaniu, po wykryciu degradacji.

Zaproponowano architekturę wykorzystania modeli prognozowania obciążenia opartych na LSTM, biorąc pod uwagę trendy przemysłowe związane z wykorzystaniem orkiestratorów chmurowych, takich jak Google Kubernetes, wraz z architekturami oprogramowania wspomagającego AI/ML dla serwerów na brzegu sieci. Takie rozwiązanie zostało wprowadzone do komercyjnego systemu 5G ORAN, ponieważ zostało zidentyfikowane przez dostawcę sieci, jako niezbędny czynnik umożliwiający wysoce adaptacyjną sieć.

9.3 Rekomendacje do dalszych badań

Autor pragnie zwrócić uwagę, że pomimo rozwiązania założonych problemów badawczych (i związanych z nimi zadań) – przedstawionych w poprzednich podrozdziałach i rozdziałach (4-8) – przeprowadzone prace badawcze przyniosły wiele inspiracji do dalszych badań. Poniżej wymieniono najważniejsze z nich:

- Problem badawczy poruszony w rozdziale 7 został zrealizowany do poziomu zbadania optymalnego modelu prognozowania zużycia procesora i jako taki właściwie odpowiada na postawione pytania badawcze i zadanie. Jednak kolejnym krokiem może być i) zbudowanie otwartego zbioru danych z bardziej kompleksowym zestawem scenariuszy pomiarowych, w celu dostarczenia społeczności wiarygodnych danych, umożliwiających projektowania i oceny modeli. Według najlepszej wiedzy autora dostępność takich modeli jest bardzo ograniczona, ii) ponadto projekt agenta określającego położenie instancji kodu RAN (ang. *workload placement*), będzie musiał budować na i rozszerzać model przewidywania obciążenia zaproponowany w pracy, aby zapewnić kompletne rozwiązanie. Autor wniósł już swój wkład w powyższą propozycję skalowania CU-UP opartego na predykcjach, w kierunku implementacji proof-of-concept w rzeczywistej sieci 5G ORAN.
- Algorytmy CAC opisane w rozdziałach 4 i 8 powinny być dalej badane szczególnie z perspektywy sieci bezkomórkowych, jak również sieci NOMA. Te dwie techniki, zgodnie z analizą stanu techniki (rozdział 2) i najlepszym zrozumieniem autora, są ważnymi kierunkami badań ze względu na oczekiwany wzrost wdrożeń sieci małych komórek w nadchodzących latach. Proces już się rozpoczął, a operatorzy wkrótce będą obserwować intensywny wzrost małych komórek w ich pobliżu –

mogą to być podejścia oparte na crowdsourcingu (napędzane przez społeczność) lub wdrożenia napędzane m.in. graczy OTT na konkurencyjnym rynku (np. Google). Tutaj szczególnie interesujący jest ogólny model sterowania przyjmowaniem zgłoszeń, który uwzględnia współistnienie wielu technologii radiowych (np. multi-RAT, NOMA). Ponieważ dostępne raporty przewidują, iż zwiększenie dostępnego widma, w paśmie środkowym, ale także dostępnym w pasmach mmWave, będzie w stanie odciążać tylko część ruchu (powiedzmy 20-45%).

- Powyżej wspomniane przyszłe rozszerzenia muszą uwzględniać:
 - Prywatne, prywatno-publiczne konfiguracje sieci komórkowej
 - Współpracę między sieciami komórkowymi i bezkomórkowymi
 - Różne rodzaje usług (tj. w podziale na branże, albo plastry sieciowe)
 - Współdzielenie pojemności w kontekście architektury neutral-host.

Jako przyszłe prace autor planuje także zająć się kontrolerami wideo w powiązaniu z protokołami wielościżkowymi (ang. *multi-path*), które uczą się na podstawie doświadczenia (tj. bazują na historii) lub potrafią dostosować się do lokalizacji (również na podstawie wcześniej zaobserwowanych parametrów sieci w danej lokalizacji).