



УНИВЕРЗИТЕТ „СВ. КИРИЛ И МЕТОДИЈ“ ВО СКОПЈЕ

ФАКУЛТЕТ ЗА ЕЛЕКТРОТЕХНИКА И ИНФОРМАЦИСКИ
ТЕХНОЛОГИИ

Институт за телекомуникации



м-р Марија Паликрушев

**ОПТИМАЛНА РАСПРЕДЕЛБА НА РЕСУРСИ ВО
ИНТЕЛИГЕНТНИ БЕЗЖИЧНИ КОМУНИКАЦИИ
ОД СЛЕДНАТА ГЕНЕРАЦИЈА**

Докторски труд

Скопје, 2026

Кандидат: **м-р Марија Паликрушев**
магистер по електротехника и информациски технологии
број на индекс: 1/2020

Ментор: **проф. д-р Зоран Хаци-Велков**
Универзитет „Св. Кирил и Методиј“ во Скопје
Факултет за електротехника и информациски технологии

Комисија:

Претседател: **проф. д-р Владимир Атанасовски**
Универзитет „Св. Кирил и Методиј“ во Скопје
Факултет за електротехника и информациски технологии

Ментор: **проф. д-р Зоран Хаци-Велков**
Универзитет „Св. Кирил и Методиј“ во Скопје
Факултет за електротехника и информациски технологии

Член: **проф. д-р Џорџ Карајанидис**
Универзитет Аристотел, Солун
Факултет за електротехника и компјутерско инженерство

Член: **проф. д-р Славче Пејоски**
Универзитет „Св. Кирил и Методиј“ во Скопје
Факултет за електротехника и информациски технологии

Член: **проф. д-р Валентин Раковиќ**
Универзитет „Св. Кирил и Методиј“ во Скопје
Факултет за електротехника и информациски технологии

Датум на одбрана: 10.09.2026

Научна област: Електротехника и информациски технологии

Лектор: **Јулијана Јосифовска**

Апстракт

ОПТИМАЛНА РАСПРЕДЕЛБА НА РЕСУРСИ ВО ИНТЕЛИГЕНТНИ БЕЗЖИЧНИ КОМУНИКАЦИИ ОД СЛЕДНАТА ГЕНЕРАЦИЈА

Овој докторски труд истражува методи за ефикасна распределба на ресурси во интелигентни безжични системи од следната генерација, со фокус на три меѓусебно поврзани аспекти: интелигентно управување со радиооколината и безжичниот канал, пресметување на работ на мрежата со федеративно учење и семантички комуникации. Ефикасната распределба на ресурсите во вакви системи бара развој на математички модели и алгоритми насочени кон намалување на комуникациското доцнење, подобрување на енергетската ефикасност и зголемување на адаптивноста во динамични безжични услови.

Во првиот дел се разгледува интелигентното управување со радиооколината преку реконфигурабилни интелигентни површини и напредни релејни системи, кои се користат за контролирано обликување на условите за простирање на сигналот. Преку заедничка оптимизација на параметрите на овие елементи се постигнува подобрување на квалитетот на комуникацискиот канал, покриеноста и искористувањето на расположливите радиоресурси. На тој начин, безжичниот канал се третира како активен елемент во оптимизацијата на системските перформанси, наместо како фиксен надворешен фактор.

Вториот дел се фокусира на енергетски одржливо федеративно учење во безжични мрежи преку интеграција на безжичен пренос на енергија за поддршка на тренирањето и комуникацијата. Развиените модели ја опфаќаат меѓузависноста помеѓу достапната енергија, времето за тренирање, квалитетот на каналот и точноста на учењето, адресирајќи го клучниот предизвик: ефикасно учење без прекумерна потрошувачка на ресурси.

Во третиот дел се разгледуваат семантичките комуникации како надградба на класичниот бит-ориентиран пренос на информација. Наместо исклучиво да се максимизира бројот на успешно пренесени битови, се воведуваат перформансни метрики за оценување на веродостојноста со која се пренесува значењето на информацијата во услови на случаен фединг во безжичниот канал. Дополнително, се развиваат хибридни бит-семантички модели кои овозможуваат динамичка адаптација помеѓу конвенционален и семантички режим на пренос, во зависност од состојбата на каналот и расположливите системски ресурси.

Управувањето со безжичниот канал, пресметувањето на работ на мрежата, енергетската одржливост и семантичкиот пренос на информацијата се темелни столбови на идните 6G мрежи. Во тој контекст, главниот научен придонес на дисертацијата се состои во формулирањето на оригинални оптимизациски проблеми и во изведувањето на нивни аналитички решенија. Врз основа на добиените решенија се развиваат алгоритми со ниска пресметковна комплексност, кои овозможуваат ефикасна онлајн распределба на ресурсите во интелигентни безжични системи од следната генерација.

Abstract

OPTIMAL RESOURCE ALLOCATION IN NEXT-GENERATION INTELLIGENT WIRELESS COMMUNICATIONS

This doctoral dissertation investigates methods for efficient resource allocation in next-generation intelligent wireless systems, focusing on three interconnected aspects: intelligent radio environment and wireless channel management, mobile edge computing with federated learning, and semantic communications. Efficient resource allocation in such systems requires development of mathematical models and algorithms aimed at reducing communication latency, improving energy efficiency, and enhancing adaptability under dynamic wireless conditions.

The first part of the dissertation examines intelligent radio environment management through reconfigurable intelligent surfaces and advanced relay systems, which are employed for shaping of signal propagation conditions. Through joint optimization of system parameters, improvements in link quality, coverage, and utilization of available radio resources are achieved. In this manner, the wireless channel is treated as an active element in the optimization of system performance, rather than as a fixed external factor.

In the second part, the dissertation focuses on energy-sustainable federated learning in wireless networks with energy-constrained wireless devices. To this end, the integration of wireless power transfer is considered, which supports model training in wireless devices and periodic transmission of updated parameters to the central aggregator. The developed models include the interdependence between available energy, training time, channel quality, and federated learning accuracy, addressing one of the key challenges in distributed intelligence: how to ensure efficient learning without excessive consumption of energy and communication resources.

The third part examines semantic communications as an enhancement to classical bit-oriented information transmission. Instead of exclusively maximizing the number of successfully transmitted bits, performance metrics are introduced for evaluating the fidelity with which the meaning of information is conveyed under random fading conditions in the wireless channel. Additionally, hybrid bit-semantic models are developed that enable dynamic adaptation between conventional and semantic transmission modes, depending on channel state and available system resources.

Wireless channel management, mobile edge computing, energy sustainability, and semantic information transmission constitute fundamental pillars of future 6G networks. In this context, the main scientific contribution of the dissertation consists in the formulation of original optimization problems in these aspects and in the derivation of their analytical solutions. Based on the obtained solutions, algorithms with low computational complexity are developed, enabling efficient online resource allocation in next-generation intelligent wireless systems.

На мојот син Касиан.

Касиан, сакам еден ден да знаеш дека твоето доаѓање на овој свет не ме оддалечи од моите соништа. Напротив, ми даде нова сила, нова смисла и уште поголема храброст да ги остварам.

Благодарност

Во јапонската култура постои концептот *мисоги* – предизвик кој изгледа поголем од нашите можности, но токму затоа нè повикува да ги надминеме сопствените граници. Целта на мисоги не е само да се достигне посакуваниот исход, туку низ самиот процес човек да се промени, да стане поиздржлив, похрабар и помудар. Денес, кога ќе погледнам наназад, сфаќам дека докторските студии беа моето животно мисоги. Не само затоа што претставуваа интелектуален предизвик, туку затоа што ме научија на истрајност, трпение и вера во сопствените способности. Оваа дисертација е резултат на долгогодишен труд, но уште повеќе е доказ дека ниту едно големо достигнување не е исклучиво лично. Зад секоја цел што изгледа како индивидуален успех стојат луѓе кои со својата поддршка, љубов и верба ни помагаат да стигнеме таму каде што сами никогаш не би можеле. Затоа, оваа благодарност им припаѓа токму ним.

Најнапред, сакам да му изразам искрена благодарност на мојот академски родител и ментор, проф. д-р Зоран Хаџи-Велков кој ми го пренесе ентузијазмот за научноистражувачката работа и ме научи дека вистинската наука започнува со љубопитност, а се гради со дисциплина, критичко размислување и истрајност. Професоре, Ви благодарам за довербата, поддршката и сите можности за професионален и личен развој што ми ги овозможивте низ овие години.

Неизмерно сум благодарна на моите родители, Сузана и Тони, кои отсекогаш веруваа во мене, дури и во моментите кога јас самата се сомневав. Нивната љубов, пожртвуваност и безрезервна поддршка ми овозможија да сонувам храбро и да истраам на патот кон тие соништа. Посебна благодарност му должам на мојот татко, кој несебично ја презеде една од најважните улоги во текот на ова патување – со огромна љубов и посветеност се грижеше за Касиан во периодот кога се вратив на работа после породилното отсуство и ги довршив докторските студии. Благодарност му должам и на мојот брат Бојан, на Мони и Зафир кои секогаш беа мојата сигурност и потсетник дека семејството е најсилната поддршка.

Посебна благодарност му упатувам на мојот сопруг Јордан, кој беше мој најголем сојузник во текот на целото ова патување. Јордан, ти благодарам за трпението, разбирањето, охрабрувањето и за сите моменти кога ја носеше тежината заедно со мене. На нашиот син Касиан му благодарам што, без да биде свесен за тоа, секој ден ме потсетуваше зошто вреди да се истрае. Неговата насмевка беше најсилната мотивација да продолжам напред.

Им благодарам и на моите колеги и пријатели со кои ги споделувавме предизвиците, неизвесностите и малите победи на академскиот живот. Вашата поддршка, советите, разговорите и чувството дека не чекарам сама низ овој процес беа од непроценлива вредност. Томче, Славче, Валентин, Матеа, Емилија Ч., Емилија К., Горан, Бодан, Александра, Наташа, Филип, Стефан, Илија, Ви благодарам! На Јована и Фелина, моите две најдобри другарки

од детството, им благодарам што низ сите овие години останаа дел од мојот живот и моето сигурно место, без оглед на тоа каде ме одведе животот.

Особено место во оваа благодарност им припаѓа на моите три Елени: Елена Џебр, Елена Бубало и Елена Бошковска. Едната ме научи дека грижата за менталното здравје не е слабост, туку предуслов за личен раст и вистинска сила. Другата ме научи дека силното тело гради и силен ум, и дека истрајноста се тренира секој ден, чекор по чекор. А третата, мојот најдобар пријател и сродна душа, никогаш не ми дозволи да заборавам која сум навистина и беше со мене во најтешките моменти, но и во најголемата радост.

И конечно, сакам да ѝ се заблагодарам на личноста со која го започнав и го завршив ова мисоги – на самата себе. За сите денови кога беше полесно да се откажам отколку да продолжам. За истрајноста, дисциплината и храброста повторно да станам и по секој неуспех. Оваа дисертација не е само доказ за научно достигнување; таа е сведоштво дека најголемите победи не ги добиваме кога сме најсилни, туку кога одбираме да не се предадеме.

На сите вас – благодарам што ми помогнавте да го остварам моето животно мисоги.

Публикации во релевантни научни списанија

- [1] A. Bazrafkan, **M. Poposka**, Z. Hadzi-Velkov, P. Popovski and N. Zlatanov, "Performance Comparison Between a Simple Full-Duplex Multi-Antenna Relay and a Passive Reflecting Intelligent Surface," in *IEEE Transactions on Wireless Communications*, vol. 22, no. 8, pp. 5461-5472, Aug. 2023 (**IF=10.7**)
- [2] **M. Poposka**, B. Jovanovski, V. Rakovic, D. Denkovski and Z. Hadzi-Velkov, "Resource Allocation of NOMA Communication Systems for Federated Learning," in *IEEE Communications Letters*, vol. 27, no. 8, pp. 2108-2112, Aug. 2023 (**IF=4.4**)
- [3] **M. Poposka**, S. Pejoski, V. Rakovic, D. Denkovski, H. Gjoreski and Z. Hadzi-Velkov, "Delay Minimization of Federated Learning Over Wireless Powered Communication Networks," in *IEEE Communications Letters*, vol. 28, no. 1, pp. 108-112, Jan. 2024 (**IF=4.4**)
- [4] **M. Poposka**, H. A. Suraweera, G. K. Karagiannidis and Z. Hadzi-Velkov, "Semantic Wireless Networks With Minimal Energy Consumption," in *IEEE Communications Letters*, vol. 28, no. 8, pp. 1894-1898, Aug. 2024 (**IF=4.4**)
- [5] Z. Hadzi-Velkov, **M. Poposka**, N. G. Evgenidis, H. A. Suraweera and G. K. Karagiannidis, "Throughput Maximization for Wireless-Powered Hybrid Bit-Semantic Communications," in *IEEE Wireless Communications Letters*, vol. 14, no. 2, pp. 445-449, Feb. 2025 (**IF=5.5**)
- [6] S. Pejoski, **M. Poposka** and Z. Hadzi-Velkov, "Optimized Scheduling Transmissions for Wireless Powered Federated Learning Networks," in *IEEE Communications Letters*, vol. 29, no. 3, pp. 640-644, March 2025 (**IF=4.4**)
- [7] M. Marinova, **M. Poposka**, Z. Hadzi-Velkov and V. Rakovic, "Optimal Cut Layer Bounds for Split Learning," in *IEEE Communications Letters*, vol. 29, no. 4, pp. 749-753, April 2025 (**IF=4.4**)

Останати публикации

- [1] **M. Poposka**, V. Rakovic, D. Denkovski, H. Gjoreski and Z. Hadzi-Velkov, “Design Optimization of RF Energy Harvesting Networks for Federated Learning,” *2024 7th International Balkan Conference on Communications and Networking (BalkanCom)*, Ljubljana, Slovenia, 2024, pp. 58-62
- [2] **M. Poposka**, Z. Hadzi-Velkov and T. Shuminoski, “Proportional Fairness in Wireless Powered Mobile Edge Computing Networks,” *2023 30th International Conference on Systems, Signals and Image Processing (IWSSIP)*, Ohrid, North Macedonia, 2023, pp. 1-5
- [3] **M. Poposka**, Z. Hadzi-Velkov and T. Shuminovski, “Design Optimization of Wireless Powered Mobile Edge Computing Systems,” *2022 30th Telecommunications Forum (TELFOR)*, Belgrade, Serbia, 2022, pp. 1-4
- [4] **M. Poposka** and Z. Hadzi-Velkov, “Edge Computing: System Overview and Fusion with Wireless Power Transfer,” *2021 International Balkan Conference on Communications and Networking (BalkanCom)*, Novi Sad, Serbia, 2021, pp. 51-55
- [5] **M. Poposka** and Z. Hadzi-Velkov, “Binary vs partial offloading in wireless powered mobile edge computing systems with fairness guarantees,” *ITU Journal on Future and Evolving Technologies*, 2022

Изјава

Изјавувам дека докторскиот труд го изработив самостојно, дека уредно ги цитирам сите користени извори и литература и дека трудот не е користен во рамките на други универзитетски студии или за стекнување на друго звање.

м-р Марија Паликрушев, _____

Содржина

Апстракт

Abstract

Благодарност

Публикации во релевантни научни списанија

Останати публикации

Изјава

Акроними vi

1	Вовед	1
1.1	Безжични комуникации од следната генерација	1
1.2	Предмет на истражувањето	3
1.3	Преглед на досегашни научни достигнувања	4
1.4	Научни придонеси на дисертацијата	8
1.5	Структура на дисертацијата	9

I Интелигентно управување со радиооколината во безжични комуникации **11**

2	Реконфигурабилни интелигентни површини	12
2.1	Технологии за управување со безжичниот каналот	12
2.2	Оптимизација на безжични системи со реконфигурабилни интелигентни површини	16
2.3	Споредбена анализа со повеќеантенски релејни системи	23

II Пресметување и учење на работ на мрежата во безжични комуникации **30**

3	Пресметување на работ на мрежата	31
3.1	Механизми за пресметување на работ на мрежата	32

3.2	Интеграција со безжичен пренос на енергија	34
3.3	Оптимизација на ресурси според max-min критериум	36
3.4	Оптимизација на ресурси според критериум на пропорционална правичност	45
4	Безжично федеративно учење	49
4.1	Концепт на федеративно учење	50
4.2	Оптимизирање на ресурси со неортогонален повеќекратен пристап	53
4.3	Оптимизирање на ресурси со временски повеќекратен пристап	58
III	Семантички комуникации	65
5	Безжичен семантички пренос на информации	66
5.1	Теориски основи на семантичките комуникации	67
5.2	Системски модели за пренос на семантичка информација	69
5.3	Перформансни метрики за семантички пренос на информација	72
5.4	Оптимизација на ресурси во безжични семантички мрежи	73
5.5	Оптимизација на ресурси во хибридни бит-семантички комуникации	80
	Заклучоци и придонеси	87
	Дел I: Интелигентно управување со радиооколината во безжични комуникации	87
	Дел II: Пресметување и учење на работ на мрежата во безжични комуникации	88
	Дел III: Семантички комуникации	90
	Насоки за идни истражувања	91
A	Основи на конвексна оптимизација	92
B	Докази	96
B.1	Доказ на Теорема 2.1	96
B.2	Доказ на Тврдење 2.1	99
B.3	Доказ на Последица 2.1	102
B.4	Доказ на Теорема 3.1	103
B.5	Доказ на Теорема 4.1	105
B.6	Доказ на Теорема 4.2	108
B.7	Доказ на Теорема 5.1	109
B.8	Доказ на Теорема 5.2	110
B.9	Доказ на Теорема 5.3	111
	Користена литература	114

Листа на слики

1.1	Квантитативна споредба на техничките барања на 5G и 6G преку осум клучни показатели на перформанси [1].	2
2.1	Споредба помеѓу конвенционалните радиооколин и паметните радиооколин.	12
2.2	Архитектура на RIS-систем [2].	13
2.3	Принцип на работа на RIS.	14
2.4	Канален модел на простирање на сигнал во случај на безжични комуникации потпомогнати од RIS.	15
2.5	Модел на комуникациски систем потпомогнат од двонасочно реле.	18
2.6	Модел на комуникациски систем потпомогнат од рефлектирачка површина (RIS).	18
2.7	Информациска брзина во функција од бројот на антенски елементи K во случај на присуство на LoS-компонента.	25
2.8	Информациска брзина во функција од бројот на антенски елементи K во случај на Рејлиев фединг.	26
2.9	Информациска брзина во функција од растојанието d_D	27
2.10	Енергетска ефикасност во функција од бројот на антенски/рефлектирачки елементи во присуство на LoS-компонента.	29
3.1	Трослојна MEC-архитектура.	32
3.2	Системски модел на безжично напојуван MEC-систем со K корисници.	37
3.3	Влијание на растојанието r на пресметковната рата (лево) и на бројот на претоварени бити (десно).	42
3.4	Влијание на моќноста на БС, P_0 , на бројот на претоварени бити.	43
3.5	Влијание на растојанието r_2 на пресметковната рата.	44
3.6	Влијание на предавателната моќност на БС, P_0 , на пресметковната рата.	44
3.7	Влијание на растојанието r_2 на индексот на правичност.	45
3.8	Влијание на растојанието r_2 на пресметковната рата.	47
3.9	Влијание на растојанието r_2 на индексот на правичност.	48
4.1	Шема за дистрибуирано машинско учење со користење на GD.	51
4.2	Системски модел на NOMA-базиран безжичен комуникациски систем за федеративно учење.	53

4.3	Влијание на максималната расположлива енергија $E_{\max}$ на вкупното доцнење T_0	57
4.4	Влијание на тежинските коефициенти b_0 на вкупното доцнење T_0	57
4.5	Системски модел и структура на TDMA-рамка за една рунда на тренирање.	59
4.6	Оптимален број на локални епохи (N_l) и глобални рунди на тренирање (N_g) наспроти предавателната моќност на БС (P_0).	63
4.7	Вкупно време на конвергенција во системот, T , наспроти предавателната моќност на базната станица P_0	64
4.8	Вкупно време на конвергенција во системот, T , наспроти комуникацискиот пропусен опсег B	64
5.1	Споредба на модели на комуникациски системи: (а) класична архитектура на Шенонов комуникациски систем, (б) архитектура на систем за семантичка комуникација.	68
5.2	Архитектура на семантички комуникациски систем базиран на длабоко учење.	70
5.3	Базична архитектура на трансформер.	71
5.4	Системски модел на предложената семантичка комуникациска мрежа.	74
5.5	Влијание на прагот на семантичка информациска брзина врз потрошувачката на енергија.	79
5.6	Влијание на растојанието до базната станица врз потрошувачката на енергија.	80
5.7	Блок-шема на безжично напојуван бит-семантички комуникациски систем.	81
5.8	Семантичка искористеност наспроти просечен однос сигнал-шум.	85
5.9	Максимален системски проток наспроти просечен однос сигнал-шум.	85
A.1	Геометриска репрезентација на конвексна функција.	93

Листа на табели

2.1	Системски параметри за нумеричка анализа за споредба на комуникациски системи потпомогнати од RIS и релеј.	24
3.1	Работни параметри за предложениот безжично напојуван MEC-систем. . .	42
4.1	Системски параметри за безжичен комуникациски систем за федеративно машинско учење базиран на не-ортогонален повеќекратен пристап.	56
4.2	Системски параметри за предложената комуникациска шема.	63
5.1	Работни параметри на безжичната мрежа за семантички комуникации. . . .	78

Акроними

Акроним	Англиски назив	Македонски превод
5G	Fifth-Generation of mobile technologies	Петта генерација мобилни мрежи
6G	Sixth-Generation of mobile technologies	Шеста генерација мобилни мрежи
RIS	Reflective Intelligent Surfaces	Рефлектирачки интелигентни површини
MEC	Mobile Edge Computing	Пресметка на работ на мобилната мрежа
WPT	Wireless Power Transfer	Безжичен пренос на енергија
ФУ	Federated Learning	Федеративно учење
FD	Full-duplex	Двонасочна комуникација
HD	Half-duplex	Еднонасочна комуникација
NOMA	Non-orthogonal Multiple Access	Истовремен повеќекратен пристап
SNR	Signal-to-Noise Ratio	Однос сигнал-шум
FPGA	Field-Programmable Gate Array	Програмабилно логичко поле
IoT	Internet of Things	Интернет на нештата
TDMA	Time Division Multiple Access	Повеќекратен пристап со искористување на временскиот ресурс
SemCom	Semantic Communications	Семантички комуникации
BitCom	Bit-based Communication	Бит-базирана (традиционална) комуникација
SRE	Smart Radio Environments	Паметни радиооколини
MIMO	Multiple-Input Multiple-Output	Повеќе влезови – повеќе излези
CSI	Channel State Information	Информации за состојбата на каналот
BC	Base Station	Базна станица
EM	Electromagnetic	Електромагнетен
LoS	Line of Sight	Директна линија на видливост
РФ	Radio Frequency	Радиофреквенција
ЕЕ	Energy Efficiency	Енергетска ефикасност
CC	Cloud Computing	Пресметка во облак
PC	Edge Servers	„Рабни“ сервери
WP-MEC	Wireless Powered MEC	Безжично напојуван MEC
CPU	Central Processing Unit	Централна процесорска единица
ЕН	Energy Harvesting	„Жетва“ на енергија

Акроним	Англиски назив	Македонски превод
EHUs	Energy Harvesting Units	Единици за собирање енергија
DL	Downlink	Комуникациска врска од базна станица до корисник
UL	Uplink	Комуникациска врска од корисник до базна станица
ML	Machine Learning	Машинско учење
GD	Gradient Descent	„Спуштање“ по градиент
SGD	Stochastic Gradient Descent	Стохастичко „спуштање“ по градиент
FedAvg	Federated Averaging	Федеративно усреднување
TR	Training Rounds	Тренирачки рунди
DNN	Deep Neural Networks	Длабоки невронски мрежи
JSCC	Joint Source-Channel Coding	Заедничко изворно и каналско кодирање
BERT	Bidirectional Encoder Representations from Transformers	BERT модел
S-R	Semantic Rate	Семантичка информациска брзина
S-SE	Semantic Spectral Efficiency	Семантичка спектрална ефикасност
suts	Semantic Units	Семантички информациски единици
EHIT	Energy Harvesting Information Transmitter	EHIT предавател
EBIR	Energy Broadcasting Information Receiver	EBIR приемник

Глава 1

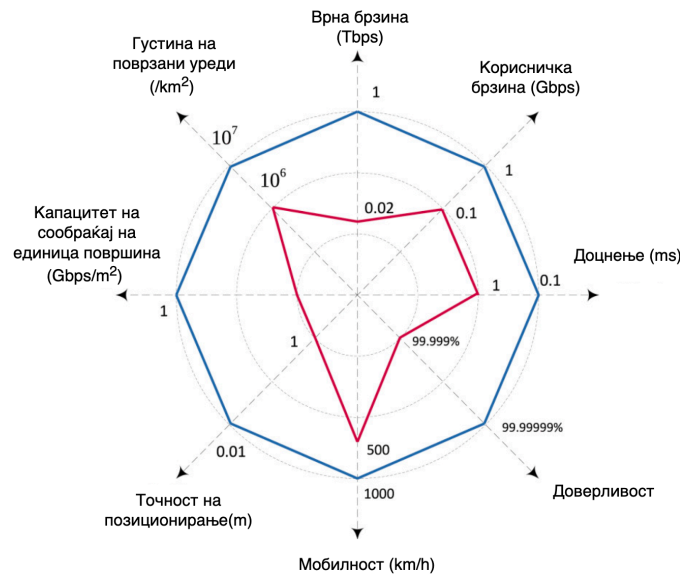
Вовед

1.1 Безжични комуникации од следната генерација

Поради масовната употреба на безжични и мобилни комуникациски уреди, како што се паметните телефони, безжичните сензори и преносните компјутери, безжичните комуникациски системи се во постојан и интензивен развој. Овој развој се огледа во континуираното појавување на нови генерации на безжични комуникациски технологии (4G, 5G, 6G и понатаму), кои нудат значително подобрени перформанси, повисока енергетска ефикасност и поголема интелигенција на комуникациските алгоритми [3]. Следната генерација на безжични и мобилни комуникациски системи треба да бидат проектирани да одговорат на потребата од значително повисоки информациски брзини, за кои се предвидува годишна стапка на раст од над 50% секоја година [1]. Овој тренд не е одреден само од експоненцијалното зголемување на бројот на интелигентни поврзани уреди, туку и од појавата на нови сервиси и апликации кои бараат оптимизација на мрежната ефикасност.

Петтата генерација (5G) на мобилни мрежи воспостави цврста основа за дигитализација на општеството преку трите дефинирани столба: подобрен мобилен широкопојасен интернет, ултрасигурни комуникации со ниско доцнење и масивни комуникации помеѓу машини. Сепак, новите апликации и технологии наметнуваат потреба од уште посоефицицирана мрежна архитектура. Еволуцијата кон шестата генерација (6G) не претставува само зголемување на капацитетот, туку суштинска промена кон концептот на „поврзана интелигенција“ (анг. connected intelligence) [4]. Оваа транзиција е мотивирана од појавата на нови апликации кои побаруваат многу ресурси, како што се холографските комуникации, проширената реалност и автономните системи од големи размери, кои бараат перформанси што ги надминуваат теоретските граници на 5G мрежите [5]. Дополнителен двигател е императивот за континуирана оптимизација на мрежната ефикасност, со особен фокус на пресметковната, енергетската, спектралната и оперативната ефикасност. Оттука, истражувањата се насочени кон подобрување на перформансите и можностите на идните комуникациски мрежи, при што клучна улога имаат технологии како вештачката интелигенција, терахерцниот фреквен-тен опсег и семантичките комуникации. Се предвидува дека 6G ќе работи во терахерцниот (THz) фреквен-тен опсег, кој обезбедува значително поголема расположлива ширина на опсег и овозможува постигнување на битски протоци од редот на терабити во секунда и латентност под милисекунда. Истовремено, се очекува вештачката интелигенција да биде интегрирана како составен дел од мрежната архитектура [5].

За прецизна квантитативна диференцијација помеѓу целите на 5G и 6G-системите, неопходно е да се анализираат најважните показатели на перформанси претставени на Слика 1.1.



Слика 1.1: Квантитативна споредба на техничките барања на 5G и 6G преку осум клучни показатели на перформанси [1].

Слика 1.1 ги прикажува клучните показатели на перформанси кои квантитативно ги разликуваат целите на 5G и 6G-системите. Како што е илустрирано, предвидената густина на поврзани уреди надминува 10^7 уреди/ km^2 ; доверливоста треба да биде над 99.9999%, додека доцнењето треба да се редуцира под 0.1 ms [6]; врвниот битски проток е проектиран на најмалку 1 Терабит/секунда, додека корисничкиот битски проток треба да надминува 1 Гигабит/s. Енергетската ефикасност се очекува да биде за два до три реда на величина повисока во споредба со 5G-системите.

Во овој контекст, 6G-системите не само што ќе пренесуваат податоци, туку ќе служат како интелигентна платформа која ја обединува комуникацијата, пресметката и сензориката. За да се постигнат овие напредни можности, современите истражувања во телекомуникациите се потпираат на иновативни, ефикасни и софистицирани решенија во сите слоеви на системот, од прилагодување на физичките карактеристики на радиоканалот кон барањата на системот, преку пренесување на обработката на сигналите кон работ на мрежата (интелигентни комуникации на работ на мрежата и федеративно учење), па сè до најапстрактното ниво на комуникација, односно толкување и пренос на значењето на информацијата (семантички комуникации). Основната цел на ваквите софистицирани комуникациски системи е оптимално искористување и распределба на достапните ресурси, како што се предавателната моќност и пропусниот опсег. Токму овие меѓусебно поврзани домени се предмет на истражување во оваа дисертација, со централен фокус на сеопфатна оптимизација на распределбата на ресурси во интелигентни безжични мрежи од следната генерација.

На фундаменталното физичко ниво, безжичната пропагандиска средина, која традиционално се сметала за пасивен медиум, денес станува подложна на активно контролирање. Реф-

лектирачките интелигентни површини (анг. reflective intelligent surfaces) (RIS) нудат суштествена промена во начинот на комуницирање преку радиосигнали, овозможувајќи контролирано обликување на радиобрановите преку употреба на специјализирана диелектрична површина составена од пасивни елементи [7]. Ова својство овозможува создавање поволни патеки за простирање на сигналот, со што се постигнува засилување на сигналите и/или прецизно поништување на интерференцијата, што е особено важно за 6G-системите кои работат на милиметарски бранови и THz-фреквенции.

Надоврзувајќи се на подобруениот физички канал, следниот чекор во еволуцијата на комуникациските системи е воведување на машинска интелигенција поблиску до корисниците, на работ на радиомрежата, со цел зголемување на можностите и флексибилноста на комуникациските системи. Пресметувањето на работ на мобилната мрежа (анг. mobile edge computing) (MEC) претставува иновативен пристап кој ги приближува пресметковните способности до мобилните уреди, со што се намалува доцнењето и се зголемува ефикасноста на системот [8]. Федеративното учење (ФУ) претставува посебен облик на дистрибуирана обработка на податоци, каде што процесот на машинско учење се спроведува во соработка меѓу корисниците, со гарантирање на приватноста на нивните локални податоци [9].

Со воведување на интелигентната рабна инфраструктура и на реконфигурабилниот физички канал, современите истражувања во безжичните комуникации се насочуваат кон нови комуникациски концепти, како што е семантички пренос на информацијата, наместо традиционален пренос на битови. Наместо пренос на синтаксички точни битови, семантичката комуникација има за цел пренос на самото значење и контекст на информацијата, што претставува ново ниво на интелигентна комуникација [10].

Во рамките на овој докторски труд ќе бидат адресирани овие фундаментални научно-истражувачки предизвици, со цел развој на иновативни и сеопфатни механизми и шеми за распределба на ресурси и оптимизација на перформансите на интелигентни безжични комуникациски системи од следната генерација.

1.2 Предмет на истражувањето

Истражувачкиот проблем на дисертацијата произлегува од потребата за софистициран дизајн и оптимизација на безжични комуникациски системи од следната генерација, кои, меѓу другото, се засноваат и на примена на вештачка интелигенција. Овие системи, кои функционираат со ограничени енергетски ресурси и во високо динамични услови на канали, бараат иновативни и сеопфатни пристапи за оптимизација на безжичните комуникациски мрежи на повеќе слоеви, од физичко ниво, па сè до апликативно ниво. Во овој контекст, клучен предизвик е здружена оптимизација на комуникациските и на пресметковните ресурси.

Овој комплексен истражувачки проблем во дисертацијата е адресиран преку три фундаментални столба на истражување. Првиот столб е фокусиран на физичкото ниво на комуникација, каде предмет на истражување е активната контрола на безжичниот канал преку рефлектирачки интелигентни површини. Истражувањето опфаќа детална компаративна анализа на RIS со конвенционалните и едноставни релејни системи. Главната цел е да се утврди

релативната супериорност на RIS во однос на перформансите, комплексноста и изводливоста, притоа земајќи го предвид сложениот хардверски дизајн и предизвиците поврзани со интерференцијата.

Вториот столб се однесува на воведување интелигенцијата на работ на мрежата, каде предмет на истражување е оптималното управување со ресурсите во комуникациски мрежи со „рабно пресметување“ и федеративно учење. Овој дел од истражувањето е насочен кон воспоставување на рамнотежа помеѓу енергетските, пресметковните и комуникациските ресурси, како и оптимизација на параметрите на самите машински модели вградени во телекомуникацискиот систем. Посебен акцент е ставен на минимизација на доцнењето и потрошувачката на енергија преку балансирање на енергијата собрана во фазата на безжичен пренос на енергија и потрошената моќност за локално тренирање. Дополнително, се истражува интеграцијата на не-ортогонален повеќекратен пристап (NOMA) како механизам за скалирање на процесите за ФУ и ефикасно справување со интерференцијата во густе безжични средини.

Третиот столб е истражување на апликативното и семантичкото ниво на комуникација, односно оптималниот дизајн на хибридни системи за семантичка комуникација. Предизвикот овде се состои во постигнување на супериорна енергетска ефикасност во споредба со традиционалните бит-базирани комуникации, преку развој на нови шеми за распределба на ресурси. Крајната цел на овој столб е развој на хибриден би-модален комуникациски систем кој динамички се префрла помеѓу традиционален и семантички режим, со цел максимизирање на вкупниот проток и обезбедување високи перформанси во широк опсег на однос сигнал-шум.

1.3 Преглед на досегашни научни достигнувања

Во продолжение е даден преглед на трите клучни истражувачки области од безжични комуникации кои се директно релевантни за оваа дисертација: (1) комуникации со посредство на реконфигурабилни интелигентни површини; (2) комуникациски системи со пресметки на работ на мрежата и со федеративно учење; и (3) семантички комуникации.

Безжичниот радиоканал вообичаено се смета за пасивен и неконтролиран медиум. Сепак, појавата на електромагнетни метаматеријали доведе до значителен напредок, овозможувајќи активна контрола на простирањето на радиобрановите во различни средини – отворени, урбани или затворени [11]. Современата комуникациска технологија дава можност да се реализира ваква контрола со употреба на површини од т.н. метаматеријали распоредени во средината каде се простираат радиосигналите. Овие метаматеријали најчесто се диелектрични површини со низи од полупроводнички елементи, на пример диоди. Диодите се управуват со едноставни процесорски единици, на пример FPGA-контролери [12]. RIS-површината овозможува контролирано одбивање на радиобрановите кои упаѓаат на неа во посакуваната насока, како и потиснување на несакани сигнали од други извори. Со тоа се поставува основата за создавање програмабилен безжичен канал [13–15]. Предностите на RIS се повеќекратни и ги позиционираат како една од клучните технологии за 6G [16]. Потрошувачката

на енергија при нивната работа е минимална, поради што придонесуваат за значително зголемување на енергетската ефикасност на целокупниот телекомуникациски систем. Со други зборови, капацитетот на радиоканалот подржан од RIS-површина значително се зголемува за сметка на минимално дополнителна потрошувачка на моќност која ја троши оваа површина. Со интеграција на RIS-технологијата, безжичните канали во природната средина стануваат контролирани и прилагодливи, при што објекти како згради или внатрешни сидови, кои традиционално го ослабуваат или нарушуваат преносот на сигналот, можат да се искористат за подобрување и оптимизација на комуникацијата [15]. Ова отвора широк спектар на примени, вклучувајќи:

- Подобрување на квалитетот на сервис: Фокусирањето на зракот преку рефлексивна RIS може да го зголеми односот сигнал-шум и да го подобри квалитетот на сервис, истовремено ублажувајќи ја меѓукорисничката интерференција [17].
- Безжичен пренос на енергија: Фокусирањето на електромагнетната енергија кон одредени Internet of Things (IoT) уреди овозможува поефикасен безжичен пренос на енергија и продолжен век на нивното работење [18].
- Безбедност на физичкиот слој: RIS може да се користи за заобиколување на уреди за прислушување или за блокирање на неовластени корисници преку расејување на зракот или алгоритми за поништување (на пр., zero forcing) [19].

RIS-површините се слични по функционалност со конвенционалните релејни системи како што се системите „засили-и-препрати“ и „декодирај-и-препрати“ [20], но нивните фундаментални принципи, предизвици и апликации се разликуваат. RIS функционираат како електромагнетни огледала, рефлектирајќи ги сигналите, наместо да ги ретрансмитираат, како што е случај со активните предаватели во релејни системи. Иако RIS покажуваат јасни предности во однос на еднонасочните релеи, особено во однос на енергетската ефикасност и пасивната природа [21], клучно е да се изврши компаративна анализа со практични двонасочни релејни системи.

Во насоката на еволуцијата кон интелигентни безжични комуникации, се наметнува потребата од распоредување на пресметковни способности поблиску до корисничките уреди, односно на работ на мобилната мрежа. Традиционалниот пристап на пресметување во облак (анг. cloud computing) претставува ефикасно решение за фиксни комуникациски мрежи, каде пресметковните и комуникациските ресурси може да се оптимизираат независно. Сепак, во контекст на безжичните радиомрежи, овој пристап се покажува како несоодветен, бидејќи ги занемарува специфичните ограничувања на радиоканалот. Имено, пренесувањето на сировите податоци до оддалечен централен уред во реално време претставува пречка за апликации кои бараат ниско доцнење, како што се автономните возила, проширената реалност и индустриската автоматизација. Пресметката на работ на мобилната мрежа (МЕС) го решава овој проблем преку децентрализирана обработка на податоци во непосредна близина на крајните корисници [22], овозможувајќи здружена оптимизација на пресметковните и радиоресурсите. На овој начин, не само што се намалува доцнењето и се подобрува пропусниот опсег, туку

корисниците добиваат можност да извршуваат пресметковно интензивни задачи, како тренирање на машински модели, директно на нивните уреди со ограничени ресурси, со помош на блиската MEC-инфраструктура. Ова ја позиционира MEC-архитектурата како клучна технологија за 5G и 6G-системите. И покрај овие предности, напојувањето на корисничките уреди останува предизвик, што доведе до интеграција на технологијата на безжичен пренос на енергија (анг. wireless power transfer) (WPT) со MEC, формирајќи безжично напојувани MEC-системи [23]. Првичните истражувања во оваа област, како оние на You et al. [24] и Bi & Zhang [25] ги поставија темелите за анализа на енергетската ефикасност и компромисите помеѓу собирањето енергија, локалното пресметување и претоварот на пресметковни задачи до MEC-серверите. Со еволуцијата кон повеќекориснички сценарија, Wang et al. [26] го проширија безжично напојуваниот MEC-систем користејќи сукцесивен повеќекратен пристап (анг. time division multiple access) (TDMA), додека Zhou & Hu [27] воведоа пресметковна рата како целна функција. Клучен придонес е преминот кон max-min оптимизација за правична распределба на ресурси, ублажувајќи го двојниот ефект близу-далеку, при што употребата на стратегијата на делумен претовар и NOMA покажуваат супериорни перформанси во однос на стратегијата на бинарен претовар и TDMA [27, 28].

Важен сегмент на пресметките на работ на мрежата е федеративното учење [9], кое овозможува дистрибуција на машинската интелигенција кон крајните безжични комуникациски уреди. Наместо централизирано тренирање на моделите во облакот, овој пристап ги трансформира милионите распространети безжични уреди, од паметни телефони и IoT-сензори до возила и паметни часовници, во активни учесници во процесот на тренирање на сложени машински модели. Ова е особено значајно во контекст на тренирање на големи јазични модели, каде огромните пресметковни и податочни барања можат да се распределат низ огромниот број на достапни безжични уреди, наместо да се потпираат исклучиво на централизирана инфраструктура. Овој принцип овозможува заедничко тренирање на такви модели, при што се гарантира зачувување на приватноста на податоците на крајните корисници, бидејќи сировите податоци никогаш не го напуштаат уредот, туку само ажурираните параметри на моделот се споделуваат со централниот сервер.

Иако ФУ е елегантно решение за проблемот со приватноста, неговата примена во безжични мрежи со голем број корисници открива нов критичен предизвик: комуникациско тесно грло. Итеративната природа на ФУ, која бара повторливи комуникациски рунди, создава значителен товар врз безжичната мрежа, особено во насока од корисниците кон базната станица, што влијае на доцнењето, потрошувачката на енергија и пропусниот опсег. Овие предизвици отвораат простор за истражувања насочени кон комуникациска ефикасност и оптимизација на ресурси во безжично федеративно учење. Еден важен правец е дизајнот на енергетски ефикасни безжични системи за федеративно учење. Бидејќи и локалното пресметување и преносот на параметри кон базната станица трошат значителна енергија, особено под услови на ниска латентност, неопходна е заедничка оптимизација. Авторите во [29] предлагаат оптимизациска рамка која ја минимизира вкупната потрошувачка на енергија во безжичен систем за ФУ преку заедничка оптимизација на фреквенциите на процесорите на уредите, пропусниот опсег, предавателната моќност и точноста на локалниот модел. Нивниот алго-

ритам покажува заштеди на енергија до 59.5% во споредба со конвенционалните методи. Дополнително, резултатите укажуваат на фундаментална размена меѓу точноста на моделот и енергетската ефикасност: поголема точност бара повеќе итерации и, со тоа, зголемено енергетско трошење.

Освен оптимизацијата на потрошувачката на моќност, значајни истражувања се насочени и кон примена на напредни протоколи за повеќекориснички пристап во безжични ФУ мрежи. Во таа насока, во [30] беше предложена адаптирана верзија на NOMA, кој го комбинира не-ортогоналниот повеќекориснички пристап со федеративното учење за намалување на вкупното доцнење во една комуникациска рунда. Во овој случај, уредите паралелно ги завршуваат локалните пресметки, по што истовремено ги пренесуваат параметрите на моделот до централниот сервер, што значително ја подобрува ефикасноста во споредба со традиционалните решенија базирани на временски ортогонален повеќекориснички пристап, како што е TDMA. Дополнително, разгледани се две шеми за детекција при сукцесивно поништување на интерференцијата, и тоа со фиксен и со променлив редослед на декодирање, при што за обете сценарија се врши заедничка оптимизација на комуникациските и пресметковните ресурси со цел минимизација на доцнењето. Резултатите потврдуваат дека NOMA значајно го намалува доцнењето во однос на TDMA, што го позиционира како ветувачко решение за интеграција на федеративното учење во идните 6G мрежи.

Одржливиот развој на федеративното учење во пракса е тесно поврзан со начинот на кој ќе се надминат ограничувањата поврзани со напојувањето на уредите. Анализата и дизајнот на системи кои интегрираат федеративно учење со безжичен пренос на енергија претставуваат релативно ново истражувачко поле, кое досега е обработено само во неколку трудови. Во [31], авторите го разгледуваат компромисот помеѓу брзината на конвергенција на моделот за ФУ и акумулираната енергија од страна на корисничките уреди, при оптимизирани големини на податочните множества и фреквенциите на процесорите на уредите. Сепак, комуникациските ресурси, како што се предавателна моќност и пропусен опсег, не се земени предвид во оптимизацијата. Во [32], се минимизира загубата при тренирање на ФУ преку оптимизација на распоредот по кој корисниците ги испраќаат своите модели, во систем напојуван од обновливи извори (соларна или ветерната енергија), но не се разгледува компромисот помеѓу времетраењето на фазите за „жетва“ на енергија и комуникација. Трудите [33] и [30] се насочени кон оптимизирање на распределбата на ресурси во безжично напојувани системи за ФУ базирани на NOMA, но резултираат со неконвексни оптимизациски проблеми кои бараат пресметковно сложени алгоритми.

Иако претходно опишаните системи за федеративно учење овозможуваат поефикасно искористување на пресметковните ресурси, тие сè уште се потпираат на традиционалниот Шенонов модел на комуникација, каде целта е точен пренос на битови. Сепак, во системи со строго ограничени ресурси, фундаменталното прашање не е „дали сите битови се примени точно?“, туку „дали примачот го разбрал значењето на пораката?“. Токму ова прашање го отвора третиот истражувачки столб – семантичките комуникации. Семантичките комуникации (анг. semantic communications) (SemCom) претставуваат интелигентен начин на пренос на значењето на информацијата која произлегува од изворот на податоци на страната на пре-

давателот. Наместо да се гарантира дека примачот ќе ги добие сите битови идентични со оние што ги испратил испраќачот, како во класичната, бит-базирана Шенонова комуникација (анг. bit-based communication) (BitCom), целта на семантичката комуникација е примачот да ја разбере семантиката (значењето) или намерата на пораката. Така, наместо засебно кодирање и пренос на секоја поединечна порака, системот овозможува размена на пораки во кои се пренесува значењето или целта на комуникацијата. Ова се постигнува со примена на вештачка интелигенција и модели на машинско учење на двете страни на комуникацискиот систем (испраќач и примач), кои „учат“ заеднички јазик и контекст. На тој начин, системот може значително да ја намали потребната предавателна моќност и доцнење, бидејќи се пренесува само она што е релевантно за значењето, а не целосниот синтаксички опис на пораката. Овој вид комуникација претставува еден од клучните концепти во развојот на 6G и идните генерации интелигентни комуникациски системи [34, 35].

Пионерски придонес во оваа област претставува DeepSC-рамката [36], која имплементира кодери и декодери базирани на трансформери за заедничко семантичко и канално кодирање [37]. Оваа рамка не само што ја овозможи дефиницијата на семантички перформансни метрики, како што се семантичка информациска брзина и семантичка сличност [38], туку овозможи и нивната примена за оптимизација на распределба на ресурси во семантички мрежи [39, 40]. Понапредните истражувања се фокусираат на развојот на хибридни комуникациски системи кои ги интегрираат предностите на SemCom и BitCom за постигнување на оптимални перформанси низ целиот спектар на SNR-вредности [41]. Во овие системи, корисниците можат динамично да избираат помеѓу семантички и бит-базирани режими, овозможувајќи долгорочно оптимизирање на достижните семантички/комуникациски брзини.

Врз основа на идентификуваните отворени истражувачки прашања во постоечката литература, главна цел на оваа дисертација е да се развијат, дизајнираат и анализираат иновативни и сеопфатни механизми, шеми и алгоритми за оптимална распределба на ресурси во интелигентни безжични комуникациски системи од следната генерација, кои вклучуваат и техники на машинско учење.

1.4 Научни придонеси на дисертацијата

Научната рамка на оваа дисертација е изградена врз три фундаментални хипотези, кои ги насочуваат истражувањата во трите клучни истражувачки столба на дисертацијата. Овие хипотези се поставени со цел да ги адресираат идентификуваните предизвици во безжичните комуникации од следната генерација:

Прва хипотеза: интелигентната контрола на физичкиот безжичен канал преку реконфигурабилни интелигентни површини може да понуди супериорни перформанси во однос на конвенционалните релејни системи, особено во практични сценарија со ограничени ресурси.

Втора хипотеза: истовремена и сеопфатна оптимизација на комуникациските, пресметковните и енергетските ресурси значително ги подобрува перформансите на системите со рабна интелигенција, особено во споредба при раздвоена и поединечна оптимизација на три-

те вида ресурси.

Трета хипотеза: семантичките комуникации постигнуваат повисока енергетска и спектрална ефикасност, односно бараат помал опсег за пренос во услови на низок однос сигнал-шум, благодарение на преносот на значење наместо поединечни битови.

Се воочува дека секоја од поставените хипотези директно соодветствува на една од трите столба на истражувањето. Преку нивна проверка и верификација, дисертацијата остварува повеќекратен научен придонес кој ги унапредува постоечките теоретски и практични сознанија во областа на интелигентните безжични мрежи. Имено, првата хипотеза е поддржана преку првиот клучен елемент на научниот придонес, кој се однесува на развојот и дизајн на иновативен повеќеантенски релеен систем. Предложениот повеќеантенски релеен систем работи во режим на истовремен прием и праќање и е базиран на 2-битни аналогни фазни поместувачи и пасивно поништување на интерференцијата. Предложениот систем е спореден со RIS во однос на постижните податочни брзини на двата системи. Резултатите покажуваат дека предложениот релеен систем постигнува значително поголеми податочни брзини, особено при мал број на антенски елементи.

Втората хипотеза е потврдена преку вториот елемент на научниот придонес, кој се фокусира на интелигенцијата на работ на мрежата. Развиена е оптимизациска шема за распределба на ресурси во безжични системи со федеративно учење, која овозможува минимизирање на времетраењето на комуникациските и пресметковните фази при ограничени енергетски и пресметковни капацитети на уредите. Предложениот алгоритам овозможува ефикасна распределба на ресурси во повеќекориснички систем, при што е постигнато значително намалување на доцнењето и подобрување на вкупната енергетска ефикасност.

Конечно, третата хипотеза е поддржана преку третиот елемент на научниот придонес, кои се однесуваат на доменот на семантичките комуникации. Прво, формирана е аналитичка рамка за оптимална распределба на ресурси која овозможува минимизација на потрошувачката на енергија при гарантирано ниво на семантичка точност. Второ, предложен е хибриден комуникациски модел кој интегрира традиционален и семантички режим на пренос. Овие придонеси заедно докажуваат дека преносот на значењето овозможува адаптивна и робусна комуникација која го максимизира протокот и спектралната ефикасност, особено во услови на неповолни карактеристики на каналот. Свкупно, дисертацијата нуди заокружена научна целина која ги решава критичните проблеми на распределба на ресурси во интелигентните безжични комуникации од следната генерација.

1.5 Структура на дисертацијата

Докторската дисертација е организирана на следниов начин. Во глава 1 (Вовед) се воведуваат клучните концепти на безжичните комуникации од следната генерација, се дефинира предметот на истражувањето, се дава преглед на досегашните научни достигнувања во трите истражувачки области опфатени со дисертацијата, и се презентираат научните придонеси и хипотези на трудот.

Дел I е насловен „Интелигентно управување со радиооколината во безжични комуни-

кации“ и се состои од глава 2, во која се разгледуваат реконфигурабилните интелигентни површини (RIS) и нивната споредба со предложен повеќеантенски релеен систем. Се презентираат технологиите за управување со безжичниот канал, се изведуваат аналитички изрази за постижните информациски брзини и енергетската ефикасност на двата система, и се врши нивна детална компаративна анализа.

Дел II, насловен „Пресметување и учење на работ на мрежата во безжични комуникации“, се состои од две глави. Во глава 3 се обработува пресметувањето на работ на мрежата (МЕС) во интеграција со безжичен пренос на енергија, при што се предлагаат шеми за распределба на ресурси според max-min критериум и критериум на пропорционална правичност. Во глава 4 се разгледува безжичното федеративно учење, со развој на оптимизациски рамки базирани на неортогонален (NOMA) и временски (TDMA) повеќекратен пристап.

Дел III, насловен „Семантички комуникации“, се состои од глава 5, во која се воведуваат теориските основи на семантичките комуникации, се дефинираат системски модели и перформансни метрики, и се предлагаат оптимизациски шеми за безжични семантички мрежи, вклучувајќи и хибриден бит-семантички комуникациски систем.

Дисертацијата завршува со поглавјето „Заклучоци и придонеси“, каде се сумираат клучните научни резултати и се даваат насоки за идни истражувања. додаток А ги презентира математичките основи на конвексната оптимизација користени низ трудот, додека додаток Б ги содржи деталните докази на сите теореми, тврдења и последици претставени во дисертацијата.

Дел I

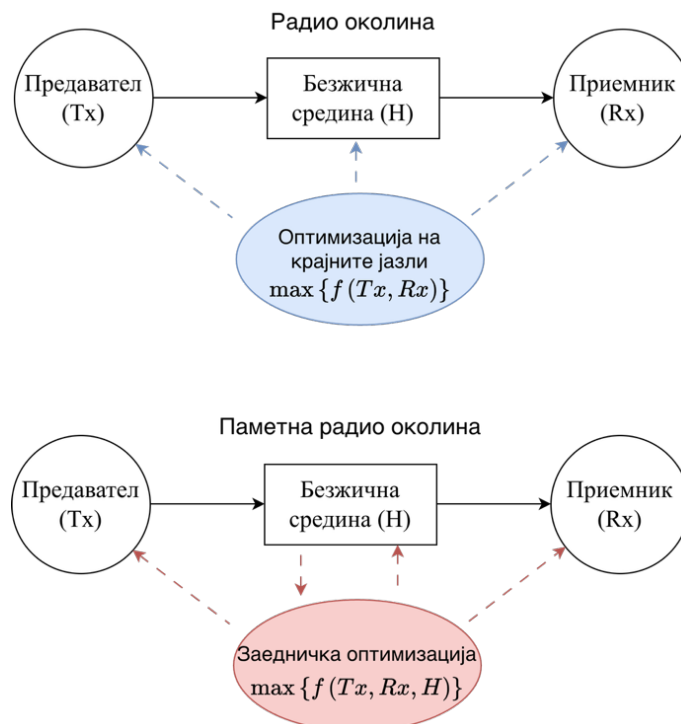
Интелигентно управување со радиооколината во безжични комуникации

Глава 2

Реконфигурабилни интелигентни површини

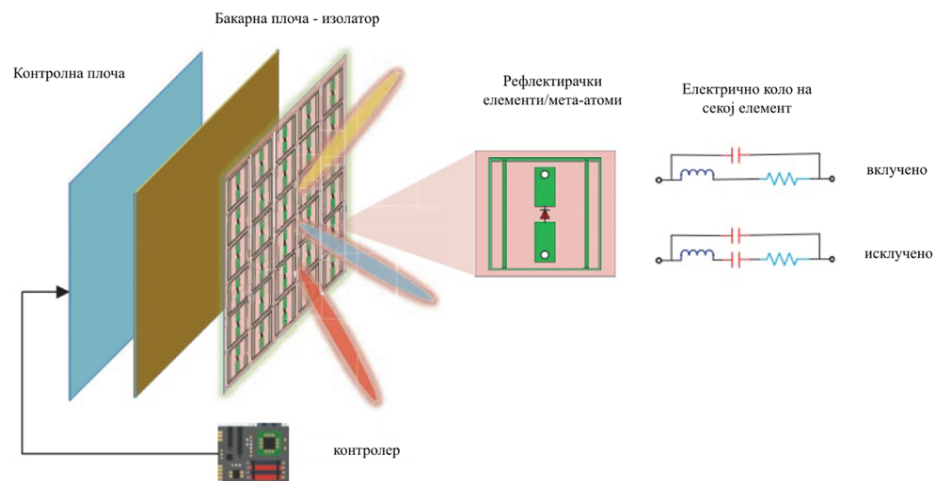
2.1 Технологии за управување со безжичниот каналот

Историски, во безжичните комуникации, пропагацискиот канал формиран од физичките објекти во околината како згради, сидови и терен, се третира како извор на деградација на перформансите, со оглед на неговиот негативен ефект врз простирањето на сигналите. Досега, овие негативни ефекти можат да се компензираат исклучиво преку софистицирани шеми на предавателот и приемникот. Меѓутоа, со појавата на технологијата на рефлектирачки интелигентни површини се овозможува „интервенција“ во самиот радиоканал. Во овие т.н. паметни радиооколинати, условите на пропагација активно се обликуваат со цел оптимизација на безжичната комуникација [42].



Слика 2.1: Споредба помеѓу конвенционалните радиооколинати и паметните радиооколинати.

Со оглед на тоа што RIS имаат способност да го обликуваат брановиот фронт на радио-брановите, безжичната околина теоретски станува подложна на адаптација согласно системските барања, како што е прикажано на Слика 2.1. Авторите во [42] го дефинираат концептот паметна радиооколина како интелигентен реконфигурабилен радиомедиум кој активно учествува во преносот и процесирањето на информациите, со што ја зголемува ефикасноста на размената на податоци помеѓу предавателот и приемникот. Во поново време, RIS [2] се етаблира како едно од најсоодветните решенија за реализација на концептот на паметни радиооколина преку модификација на претходно неконтролираниот пропагациски медиум. RIS претставува матрица од едноставни и евтини пасивни антенски елементи (рефлектирачки единици), погодни за инсталација врз различни објекти како фасади на згради и внатрешни сидови. Енергетската потрошувачка на RIS е занемарлива во споредба со ригорозните енергетски барања на РФ низите кај масивните антенски полиња со повеќе предавателни и приемни антени (анг. multiple-input multiple-output) (MIMO). Способноста на RIS за реконфигурација на медиумот се заснова на специфичните физички својства на метаматеријалите од кои се изградени нивните структурни блокови. Овие структурни блокови, често дефинирани како ќелии или метаатоми, се интегрирани во рамна површина. Секој рефлектирачки елемент е проектиран да врши контролирана реконструкција на карактеристиките на упадните електромагнетни бранови, вклучувајќи ги амплитудата, фазата и поларизацијата. Следствено, RIS воведува дополнителен степен на слобода во оптимизацијата на радиомедиумот. Врз основа на достапните информации за состојбата на каналот (анг. channel state information) (CSI) во базната станица (БС) може да се има целосна контрола врз реконфигурацијата на секој елемент преку наменски контролен канал и контролер.

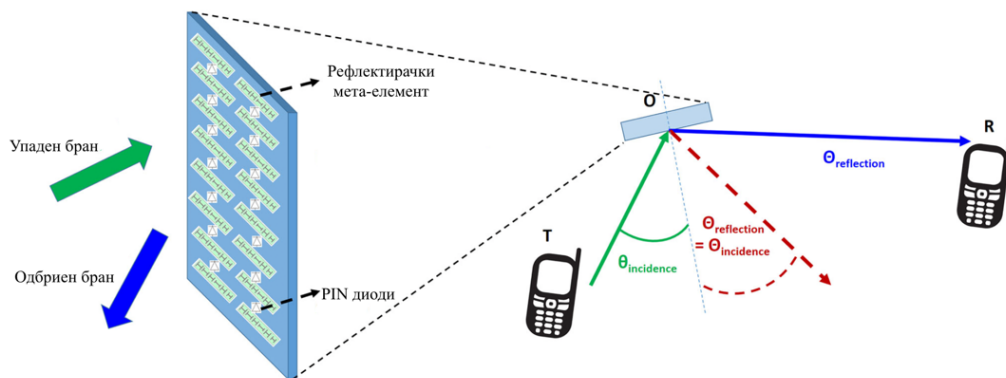


Слика 2.2: Архитектура на RIS-систем [2].

Типичната архитектура на RIS, прикажана на Слика 2.2, се состои од три слоеви и контролер. Надворешниот слој содржи мноштво метаелементи нанесени врз диелектрична подлога и овој слој служи за директна интеракција со упадните електромагнетни сигнали. Меѓуслојот е изведен од бакарна плоча со функција на изолатор за спречување на дисипација на енергијата. Внатрешниот слој содржи контролна логичка плоча задолжена за конфигурација на фазното поместување и амплитудата на рефлексija кај секој засебен елемент. Рекон-

фигурацијата на секој метаелемент се остварува преку вградена PIN-диода, чија состојба „вклучено“ или „исклучено“ се контролира со промена на напонот на поларизација. На овој начин, секој елемент воведува дискретна фазна разлика во рефлектираниот сигнал. Доколку е потребна и контрола на амплитудата на рефлексивната, во колото се вклучува и променлив отпорник кој овозможува континуирано прилагодување на амплитудата во опсегот $[0, 1]$, преку контролирана дисипација на дел од енергијата на упадниот сигнал. Целокупното управување со овие елементи се координира преку контролерот, кој ги прима информациите за состојбата на каналот од базната станица преку наменски безжичен контролен канал и врз основа на нив динамички ги пресметува и применува оптималните поставки за секој метаелемент.

Функционалностите на RIS-технологијата се многукратни. Поради способноста за модулација на упадните радиобранови, овие површини се класифицираат на активни и пасивни. Со нивна помош, примените сигнали можат да се комбинираат конструктивно или деструктивно [43, 44]. Ова овозможува, на пример, да се подобри јачината на саканиот примен сигнал на локацијата на корисникот, да се создаде виртуелна линија на видливост околу дадена физичка препрека, или да се поништи несакана интерференција.



Слика 2.3: Принцип на работа на RIS.

Принципот на работа на RIS е претставен на Слика 2.3. При контакт на упадниот бран со површината, тој се разложува на рефлектирана и пропуштена (рефрактивна) компонента. Овие компоненти се реконфигурираат преку специфично активирање на метаелементите, што се реализира преку PIN-диоди. RIS врши обликување на брановите со примена на генерализираните Снелови закони за рефлексивна и рефрактивна, со што аголот на рефлексивна станува параметар кој може да се контролира.

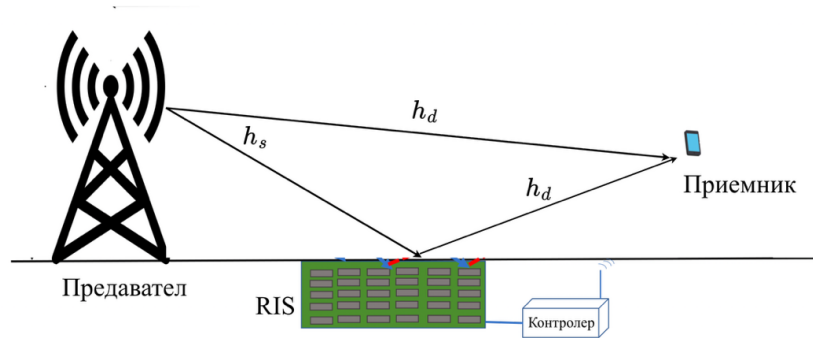
Основната конфигурација на комуникациски системи со посредство на RIS вклучува предавател, приемник и RIS-површина позиционирана меѓу нив, составена од K рефлектирачки елементи, како што е прикажано на Слика 2.4. За разлика од традиционалните релејни системи кои активно детектираат, процесираат и повторно емитуваат сигнали, RIS-елементите работат на пасивен начин, само модулирајќи ја фазата на рефлектираниот сигнал без додавање на дополнителна енергија. Вкупниот канал меѓу предавателот и приемникот во присуство на RIS може да се декомпонира на три основни компоненти. Првата компонента е директниот канал $\mathbf{g} \in \mathbb{C}$ кој претставува комплексен скаларен коефициент за директната Line-of-Sight

(LoS) или Non-Line-of-Sight (NLoS) врска меѓу предавателот и приемникот. Втората компонента е каналот од предавателот до RIS, означен како $\mathbf{h}_s \in \mathbb{C}^{K \times 1}$, кој претставува комплексен вектор со димензии $K \times 1$, каде K е бројот на RIS-елементи. Третата компонента е каналот од RIS до приемникот, означен како $\mathbf{h}_D \in \mathbb{C}^{K \times 1}$, исто така комплексен вектор со димензии $K \times 1$. Овие три компоненти заедно го дефинираат целосниот каналниот модел на безжичен комуникациски систем кој работи со посредство на RIS. Примениот сигнал во тој случај може да се изрази како

$$\mathbf{y} = (\mathbf{g} + \mathbf{h}_s \Phi \mathbf{h}_D) x + \mathbf{n}, \quad (2.1)$$

каде x е предавателниот сигнал, $\mathbf{n}$ е адитивен бел Гаусов шум, а $\Phi = \text{diag}(\beta_1 e^{j\phi_1}, \beta_2 e^{j\phi_2}, \dots, \beta_K e^{j\phi_K})$ е $K \times K$ матрица на фазните и амплитудните коефициенти на RIS. Притоа, $\beta_k \in [0, 1]$ го претставува амплитудниот коефициент, додека $\phi_k \in [0, 2\pi)$ го претставува фазното поместување на k -тиот рефлектирачки елемент. Во идеални услови, амплитудниот коефициент е максимален ($\beta_k = 1$), што значи дека елементот целосно ја рефлектира енергијата на падниот сигнал.

Фундаменталната карактеристика на RIS-каналот која го разликува од традиционалните релејни системи е неговата каскадна природа. Додека во активните релејни системи каналот меѓу изворот и дестинацијата може да се третира како два независни канални сегменти поврзани преку активен репетитор, во RIS-системите каналот е производ од два канала. Конкретно, еквивалентниот RIS-канал може да се запише како $\mathbf{H}_{\text{RIS}} = \mathbf{h}_s \Phi \mathbf{h}_D$. Најзначајната последица на каскадната структура е двојното слабеење радо простирање. Во традиционалните директни комуникациски врски, моќноста на примениот сигнал опаѓа пропорционално на $d^{-\alpha}$, каде d е оддалеченоста меѓу предавателот и приемникот, а α е експонентот на слабеење кој типично изнесува помеѓу 2 и 4 во зависност од условите на пропација. Меѓутоа, во RIS-системите, сигналот мора да помине низ две последователни делници: прво од предавателот до RIS (оддалеченост d_s), а потоа од RIS до приемникот (оддалеченост d_d). Како резултат на тоа, моќноста на примениот сигнал е пропорционална со $(d_s \cdot d_d)^{-\alpha}$. Ова двојно слабеење е директна последица на пасивната природа на RIS и претставува еден од главните предизвици при дизајнирањето на комуникациски системи со посредство на RIS.



Слика 2.4: Канален модел на простирање на сигнал во случај на безжични комуникации потпомогнати од RIS.

2.2 Оптимизација на безжични системи со реконфигурабилни интелигентни површини

Безжичниот пренос во милиметарскиот (30–100 GHz) и субмилиметарскиот опсег (над 100 GHz) типично бара постоење на директна линија на видливост помеѓу предавателот и приемникот, или пак присуство на посреднички уред со LoS до двата краја. Во претходната секција беше покажано дека RIS претставува ветувачко решение за овој проблем, благодарение на способноста за активно обликување на радиобрановите. Сепак, за да се процени реалниот потенцијал на RIS, неопходна е неговата споредба со традиционалните релејни системи, кои исто така служат како посреднички уреди во вакви сценарија. Во таа смисла, RIS наликува на двонасочно реле од тип „засили-и-препрати“ (анг. amplify-and-forward full-duplex relay) со големо антенско поле, но со клучна разлика – наместо активно засилување и повторно пренесување на сигналот, RIS пасивно го рефлектира и насочува упадниот електромагнетен бран кон саканата насока.

Неодамнешниот наплив на истражувачки трудови за комуникации потпомогнати од RIS може да се припише на одредени предности на оваа технологија во однос на традиционалните релеи [45]. Иако резултатите од литературата покажуваат дека RIS ги надминува еднонасочните релеи [46] во однос на постигнатите перформанси, сè уште не е покажано дали тоа важи и за практично двонасочно реле со повеќе антени. Идеалното двонасочно реле кое нема сопствена интерференција е непрактично за реална имплементација поради два клучни проблема. Прво, би барало посебна радиофреквентна низа (анг. RF chain), односно целосен сет на електронски компоненти за предавање и примање на сигналот, вклучувајќи засилувачи и аналого-дигитални конвертори за секој антенски елемент поединечно, што резултира со огромна хардверска комплексност и висока цена. Второ, бидејќи релето истовремено предава и прима сигнал на иста фреквенција, сопствениот предаден сигнал се враќа назад во приемникот и создава интерференција, чие поништување бара дополнителен сложен хардвер.

Повеќеантенските полиња претставуваат клучна технологија која може значително да ја зголеми спектралната ефикасност на комуникациските системи со милиметарски бранови [47]. Едно од најпрактичните и најизводливи решенија за изградба на повеќеантенски полиња кое се карактеризира со ниска потрошувачка на енергија и ниска комплексност е техниката на аналого формирање на сноп [48]. Аналогното формирање на сноп може да се имплементира така што секој антенски елемент би се опреми со аналоген фазен поместувач. Овој поместувач може да ја помести фазата на примениот сигнал за саканата фаза, со што се овозможува повеќеантенскиот систем да изврши пренос/прием со формирање на сноп кон/од саканата насока во просторот. Постојат многу различни технологии за изградба на фазни поместувачи и ова е многу активно поле на истражување, при што постојано се воведуваат напредни технологии за фазно поместување. Една таква технологија беше предложена во [49], каде авторите предлагаат едноставен 2-битен фазен поместувач изграден со PIN-диоды. Бидејќи и фазните поместувачи кај двонасочното реле и рефлектирачките елементи кај RIS се изградени врз основа на истата технологија, т.е. PIN-диоды [50], нивните

трошоци за изградба по елемент се споредливи. Ова значи дека од економски аспект, двонасочното реле со 2-битни фазни поместувачи не претставува значително поскапо решение од RIS, со оглед на тоа дека основната градбена единица е иста кај двата системи.

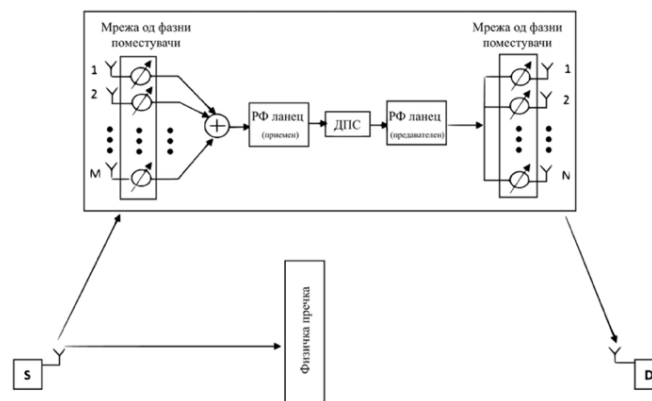
Клучниот предизвик при двонасочна комуникација е сопствената интерференција – појава при која сигналот што го предава релето истовремено се прима од неговиот сопствен приемник, предизвикувајќи деградација на корисниот примен сигнал. Во литературата се познати три пристапи за потиснување на оваа интерференција, кои се разликуваат по комплексност, цена и ефикасност. Првиот пристап е пасивното потиснување, кое се заснова исклучиво на физичка изолација помеѓу предавателното и приемното антенско поле, на пример, преку физичко растојание или употреба на апсорпциски материјали помеѓу предавателните и приемните антени. Иако овој пристап е наједноставен и најевтин за имплементација, тој обезбедува ограничено ниво на потиснување на интерференцијата. Вториот пристап е аналогното потиснување, кое се реализира во радиофреквентниот домен пред аналого-дигиталната конверзија на сигналот. Ова се постигнува со генерирање на копија на предадениот сигнал и нејзино одземање од примениот сигнал во аналоген домен. Иако поефикасен од пасивното потиснување, овој пристап бара дополнителни аналогни кола и прецизна синхронизација. Третиот пристап е дигиталното потиснување на интерференција, кое се применува во дигиталниот домен после аналого-дигиталната конверзија. Овој пристап овозможува највисоко ниво на потиснување на интерференцијата, но бара сложени дигитални алгоритми и значително поголема пресметковна моќ.

Предложеното двонасочно реле се потпира исклучиво на пасивно потиснување преку физичка изолација помеѓу одвоените антенски полиња за пренос и прием, без употреба на аналогни или дигитални кола за потиснување на интерференцијата. Иако ова резултира со поограничено потиснување на сопствената интерференција во споредба со понапредните пристапи, предноста лежи во значително поедноставен дизајн и пониски трошоци за имплементација.

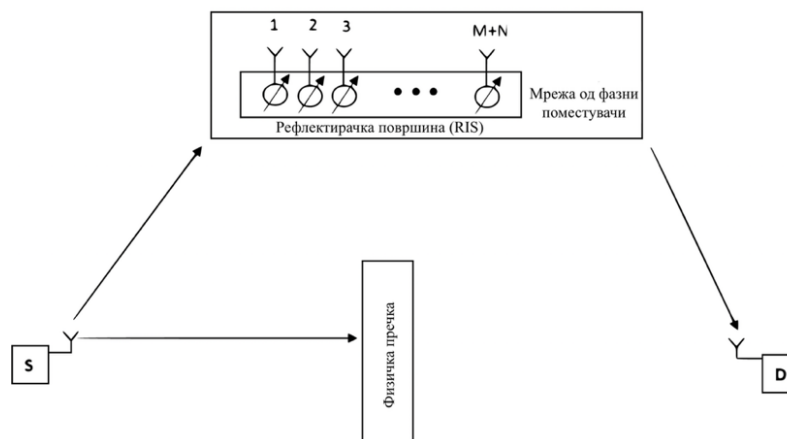
Неодамна, предложени се нови антенски технологии базирани на програмабилни метаматеријали, како што се реконфигурабилните холографски површини [51] и реконфигурабилните рефлектирачки полиња [52], кои овозможуваат прецизно обликување на радиобрановите со ниска хардверска комплексност и ниска потрошувачка на енергија. Слично на пасивните RIS, овие типови антени се базираат на програмабилни метаматеријали, па затоа доколку дизајнираме реле користејќи ги овие антенски технологии, би имале споредлива големина и потрошувачка на енергија како RIS. Природното прашање во овој случај е: дали RIS е подобар од едноставно двонасочно реле? Целта на оваа глава е да се одговори токму на ова прашање. Најпрво, ќе предложиме и истражиме едноставно двонасочно реле со една РФ низа и повеќе антени, изграден со b -битни аналогни фазни поместувачи и пасивно поништување на себе-интерференцијата. Потоа, ќе ја изведеме постигнатата брзина на пренос на податоци за систем кој се состои од предавател на информација (извор), предложениот двонасочно реле и приемник (дестинација). Конечно, ќе ги споредиме постигнатите брзини на пренос на податоци на предложениот двонасочно реле со истиот систем, но кога еднонасочниот релеј е заменет со идеален пасивен RIS. Дополнително, ќе ја анализираме енергет-

ската ефикасност на двата системи, земајќи ги предвид реконфигурабилните холографски површини како имплементација.

Системскиот модел кој го разгледуваме се состои од предавател на информација (извор) со една антена (S), приемник на информација (дестинација) со една антена (D), и посреднички елемент – RIS (Слика 2.5) или двонасочно реле (Слика 2.6). Со цел да се овозможи фер споредба, се претпоставува дека релето и рефлектиракката површина користат ист број на антенски елементи. Дополнително, се претпоставува дека двата системи користат идентична вкупна предавателна моќност, означена со P_T . Во случајот на RIS-системот, вкупната моќност P_T е целосно искористена од изворот, додека во системот со двонасочно реле вкупната моќност P_T се дели помеѓу изворот, кој користи моќност P_S , и релејот, кој користи моќност P_R , така што важи $P_S + P_R = P_T$. Претпоставуваме дека изворот и дестинацијата не можат директно да комуницираат поради физички пречки, додека посредничкиот уред има непречен LoS до изворот и до дестинацијата. Како резултат на тоа, комуникацијата помеѓу изворот и дестинацијата мора да се одвива преку посредничкиот уред, односно со посредство на релејот или RIS.



Слика 2.5: Модел на комуникациски систем потпомогнат од двонасочно реле.



Слика 2.6: Модел на комуникациски систем потпомогнат од рефлектирачка површина (RIS).

Предложеното двонасочно реле „декодирај-и-препрати“ се состои од повеќеантенски

приемник (M антени), РФ низа и повеќеантенски предавател (N антени), каде $M + N = K$, а K е вкупниот број на антенски елементи на двонасочниот релеј. Приемните и предавателните антенски полиња се изолирани само пасивно, што значи дека не се користи активно потиснување на сопствената интерференција. Како што може да се види од структурата на предложеното двонасочно реле, илустрирана на Слика 2.5, приемната страна на релејот се состои од M антени, секоја поврзана со b -битен аналоген фазен поместувач што може да ја помести фазата на примениот сигнал на дадениот антенски елемент за фаза од следното множество:

$$\mathcal{P} = \left\{ 0, \frac{2\pi}{2^b}, 2 \times \frac{2\pi}{2^b}, 3 \times \frac{2\pi}{2^b}, \dots, (2^b - 1) \times \frac{2\pi}{2^b} \right\}, \quad (2.2)$$

каде $b = 1, 2, 3, \dots$. Потоа, фазно поместените сигнали од секоја приемна антена се комбинираат преку аналоген собирач и се испраќаат до приемна РФ низа. Оваа приемна РФ низа има за задача да генерира дигитална репрезентација на влезниот сигнал во основен опсег, кој потоа се испраќа до единица за дигитално процесирање на сигнали (ДПС) за понатамошна обработка во дигиталниот домен. На предавателната страна на предложениот двонасочно реле, како што е илустрирано на Слика 2.5, дигитален сигнал се испраќа до предавателна РФ низа, кој генерира сигнал во опсег, кој потоа се праќа на N предавателни антени. За да се задржи концизен приказ на основните принципи и резултати, сите математички изведувања поврзани со пресметката на постижната брзина на пренос на двонасочното реле се дадени во додаток Б.1. Во продолжение ќе ги дадеме само конечните формулирани теореми и тврдења.

Теорема 2.1. *Кога векторите на фазно поместување $\mathbf{u}$ и $\mathbf{v}$ се дадени со:*

$$\mathbf{u} = \left[e^{-j\hat{\phi}_{S,1}}, e^{-j\hat{\phi}_{S,2}}, \dots, e^{-j\hat{\phi}_{S,M}} \right]^T, \quad (2.3)$$

$$\mathbf{v} = \left[e^{-j\hat{\phi}_{D,1}}, e^{-j\hat{\phi}_{D,2}}, \dots, e^{-j\hat{\phi}_{D,N}} \right]^T, \quad (2.4)$$

каде $\hat{\phi}_{S,m}$ и $\hat{\phi}_{D,n}$ се b -битни квантизирани верзии на фазите на каналите $\phi_{S,m}$ и $\phi_{D,n}$, соодветно, и се дадени со:

$$\hat{\phi}_{S,m} = \frac{2m\pi}{2^b} \text{ ако } \phi_{S,m} \in \left[\frac{2m\pi}{2^b}, \frac{2(m+1)\pi}{2^b} \right), \quad (2.5)$$

$$\hat{\phi}_{D,n} = \frac{2m\pi}{2^b} \text{ ако } \phi_{D,n} \in \left[\frac{2m\pi}{2^b}, \frac{2(m+1)\pi}{2^b} \right), \quad \text{за } m \in \{0, 1, \dots, 2^b - 1\}, \quad (2.6)$$

тогаш постижната брзина на пренос на предложениот двонасочно релеен систем е дадена со:

$$R = \min \left\{ \log_2 \left(1 + \frac{P_S \Omega_S (1 + (M-1)Q)}{N_0 + \frac{\eta}{M} P_R} \right), \log_2 \left(1 + \frac{P_R \Omega_D (1 + (N-1)Q)}{N_0} \right) \right\}, \quad (2.7)$$

каде

$$Q = \left(\frac{2^b}{\pi} \sin \left(\frac{\pi}{2^b} \right) \right)^2. \quad (2.8)$$

Изведувањето и доказот на Теорема 2.1 се дадени во додаток Б.1.

Очигледно е дека брзината на пренос во (2.7) може да се максимизира преку оптимизација на вредностите за M , N , P_S и P_R , со оглед на ограничувањата $M + N = K$ и $P_R + P_S = P_T$. Па така, максималната достижна брзина во разгледуваниот релеен систем се определува со решавање на следниот оптимизациски проблем

$$\begin{aligned} & \underset{P_R, N}{\text{Maximize}} && R \\ & \text{под услов:} && M + N = K \\ & && P_S + P_R = P_T, \end{aligned} \quad (2.9)$$

каде целната функција R се определува според (2.7) и P_T е вкупната предавателна моќност поделена помеѓу изворот и релето. Притоа, за определување на решението во затворена форма се потпираме на следново тврдење.

Тврдење 2.1. *Го предлагаме следново субоптимално решение на (2.9):*

$$M^* = \frac{2K}{3}, \quad (2.10)$$

$$N^* = \frac{K}{3}, \quad (2.11)$$

$$P_R^* = \frac{-\alpha + \sqrt{48P_T K N_0 \eta \Omega_S \Omega_D + \alpha^2}}{6\eta \Omega_D}, \quad (2.12)$$

$$P_S^* = P_T - P_R^*, \quad (2.13)$$

каде

$$\alpha = 2K N_0 \Omega_D + 4K N_0 \Omega_S. \quad (2.14)$$

Доказот на Тврдење 2.1 е даден во додаток Б.2.

Комуникацискиот систем потпомогнат од рефлектирачка површина е прикажан на Слика 2.6. Рефлектирачката површина се состои од K рефлектирачки елементи и има вкупна површина B . Следејќи ги [53–55], RIS може да се моделира како рамнинско антенско поле, каде што секој антенски елемент има површина со големина A , така што $B = KA$ и $A \leq (\lambda/4)^2$. Фазното поместување на секој рефлектирачки елемент може да се прилагоди така што упадниот сигнал се рефлектира и се насочува (анг. beamforming) кон дестинацијата. Со цел да се овозможат најдобрите можни перформанси на RIS, претпоставуваме дека секој фазен поместувач кај RIS може да се постави на која било вредност во опсегот $[0, 2\pi)$, односно, за разлика од предложениот двонасочно реле, кај RIS нема фазна квантизација, што ги максимизира неговите перформанси.

Слично како кај релејот, претпоставуваме дека каналот помеѓу изворот S и k -тиот рефлектирачки елемент на RIS е означен со $h_{S,k}$, а каналот помеѓу k -тиот рефлектирачки елемент на RIS и дестинацијата, D , е означен со $h_{D,k}$. Каналите $h_{S,k}$ и $h_{D,k}$, за $k = 1, 2, \dots, K$,

се дадени со:

$$h_{S,k} = \sqrt{\Omega_{S,k}} e^{j\phi_{S,k}}, \quad (2.15)$$

$$h_{D,k} = \sqrt{\Omega_{D,k}} e^{j\phi_{D,k}}, \quad (2.16)$$

каде $\Omega_{S,k}$ и $\phi_{S,k}$ ја означуваат моќноста и фазата на каналот помеѓу изворот и k -тиот рефлектирачки елемент на RIS, додека $\Omega_{D,k}$ и $\phi_{D,k}$ ја означуваат моќноста и фазата на каналот помеѓу k -тиот рефлектирачки елемент на RIS и дестинацијата. Слично на моделот на релејниот систем, секој канал се моделира како статичен, при што $\Omega_{S,k}$, $\Omega_{D,k}$, $\phi_{D,k}$ и $\phi_{S,k}$ се фиксни за секое k . Понатаму, повторно слично на релејниот систем, претпоставуваме дека изворот и дестинацијата се наоѓаат во регионот на далечно поле на RIS, како што е дефинирано во [53], и затоа соодветните моќности на каналите се дадени со:

$$\Omega_{S,k} = \Omega_S, \forall k \quad (2.17)$$

$$\Omega_{D,k} = \Omega_D, \forall k. \quad (2.18)$$

Во продолжение ќе ја изведеме брзината на пренос на информации за RIS-системот. Нека P_T ја означува предавателната моќност на изворот. Под претпоставка на статични LoS канали од изворот до RIS и од RIS до дестинацијата, постижната брзина на пренос помеѓу изворот и дестинацијата, R_{RIS} , е дадена со [53]:

$$R_{RIS} = \log_2 \left(1 + \frac{K^2 P_T \Omega_S \Omega_D}{N_0} \right). \quad (2.19)$$

Изразите за брзините на пренос кај релејниот систем и RIS-системот дадени со (2.7) и (2.19), соодветно, важат за статичен канал во присуство на LoS-компонента. Меѓутоа, заради комплетност, ги изведуваме и брзините на пренос за канали со Рејлиев фединг.

Последица 2.1. При Рејлиев фединг, постижната брзина на пренос на предложениот комуникациски систем потпомогнат од релеј, R^{Ra} , е дадена со:

$$R^{Ra} = \min \{ \log_2 (1 + \gamma_S^{Ra}), \log_2 (1 + \gamma_D^{Ra}) \}, \quad (2.20)$$

каде

$$\gamma_S^{Ra} = \frac{P_S \Omega_S (1 + (M-1) \frac{\pi}{4} Q)}{N_0 + \frac{\eta}{M} P_R}, \quad (2.21)$$

$$\gamma_D^{Ra} = \frac{P_R \Omega_D (1 + (N-1) \frac{\pi}{4} Q)}{N_0}. \quad (2.22)$$

Брзината добиена во (2.20) може да се максимизира со примена на субоптималните решенија од Тврдење 2.1. Доказот на Последица 2.1 е даден во додаток Б.3.

Последица 2.2. При Рејлиев фединг, постижната брзина на пренос на предложениот ко-

муникациски систем потпомогнат од рефлектирачка површина, R_{RIS}^{Ra} , е дадена со [56]:

$$R_{RIS}^{Ra} = \log_2 \left(1 + \frac{P_T \Omega_S \Omega_D K}{N_0} \times \left(1 + (K-1) \left(\frac{\pi}{4} \right)^2 \right) \right). \quad (2.23)$$

Од (2.19) и (2.23) можеме да забележиме дека Рејлиевит фединг го намалува односот сигнал-шум за фактор од $\pi/4$ споредено со случајот на статичен канал. Доколку наместо во двонасочен режим, релејот работи во еднонасочен режим и ги користи сите свои K антени само за пренос или само за прием во различни временски интервали, достижната брзина на пренос на податоци на соодветниот систем потпомогнат од еднонасочен релеј е дадена со [57]:

$$R_{HD} = \frac{\log_2(1 + \gamma_S^{HD}) \log_2(1 + \gamma_D^{HD})}{\log_2(1 + \gamma_S^{HD}) + \log_2(1 + \gamma_D^{HD})}, \quad (2.24)$$

каде што

$$\gamma_S^{HD} = \frac{P_T \Omega_S (1 + (K-1)Q)}{N_0}, \quad (2.25)$$

$$\gamma_D^{HD} = \frac{P_T \Omega_D (1 + (K-1)Q)}{N_0}. \quad (2.26)$$

Во овој дел ќе ја изведеме енергетската ефикасност на комуникацискиот систем потпомогнат од двонасочно реле, дефинирана како:

$$EE = \frac{R}{P_{total}}, \quad (2.27)$$

каде што R е достижната брзина на пренос, а P_{total} е вкупната потрошувачка на моќност на разгледуваниот систем. Најпрво, ќе ја дефинираме енергетската ефикасност на комуникацискиот систем потпомогнат од RIS. Вкупната потрошувачка на моќност на ваквиот систем е изведена во [58–60]:

$$P_{total} = \xi P_S + p_c^{(S)} + K p_0^{(RIS)} + p_c^{(RIS)} + p_c^{(D)}, \quad (2.28)$$

каде што ξ е ефикасноста на предавателниот засилувач на моќност, P_S е предавателната моќност на изворот (која во случајот со RIS е $P_S = P_T$), K е бројот на рефлектирачки елементи на RIS, $p_0^{(RIS)}$ е потрошувачката на моќност на секој поединечен елемент на RIS, додека $p_c^{(S)}$, $p_c^{(RIS)}$ и $p_c^{(D)}$ се статички потрошувачки на моќност на хардверот на изворот, RIS-површината и дестинацијата, соодветно. Сега, со вметнување на (2.28) и R_{RIS} од (2.19) на местото на R во (2.27), ја добиваме енергетската ефикасност на RIS-системот.

Аналогно на изведувањето на (2.28), вкупната потрошувачка на моќност на комуникацискиот систем потпомогнат од двонасочно реле се добива како:

$$\begin{aligned} P_{total} &= \xi P_S + p_c^{(S)} + M p_0^{(relay)} + p_c^{(relay)} + N p_0^{(relay)} + \xi P_R + p_c^{(D)} \\ &= \xi (P_S + P_R) + p_c^{(S)} + (M + N) p_0^{(relay)} + p_c^{(relay)} + p_c^{(D)}, \end{aligned} \quad (2.29)$$

каде што P_R е предавателната моќност на релејот, $p_c^{(relay)}$ е хардверската статичка потрошу-

вачка на моќност на релето, а $p_0^{(relay)}$ е потрошувачката на моќност на секој антенски елемент кај релето, додека M и N се бројот на приемни и предавателни антени. Имајќи предвид дека кај релето важи $P_S + P_R = P_T$ и $K = M + N$, вкупната потрошувачка на моќност на комуникацискиот систем потпомогнат од двонасочно реле се добива како:

$$P_{total} = \xi P_T + p_c^{(S)} + K p_0^{(relay)} + p_c^{(relay)} + p_c^{(D)}. \quad (2.30)$$

Доколку наместо двонасочно реле се користи еднонасочен релеј, тогаш $p_c^{(S)}$ во (2.30) треба да се замени со $p_c^{(S)}/2$, бидејќи изворот е активен само во половина од времето.

2.3 Споредбена анализа со повеќеантенски релејни системи

Со цел да направиме фер споредба помеѓу комуникацискиот систем потпомогнат од RIS и комуникацискиот систем потпомогнат од релеј, ја усвојуваме структурата на рамнинско антенско поле разгледувана во [53] и притоа претпоставуваме дека површината на секој предавателен/приемен антенски елемент кај релето и на секој рефлектирачки елемент кај RIS, означена со A , го задоволува условот $A \leq (\lambda/4)^2$, каде што λ е брановата должина на носителот на сигналот. Нека d_S го означува растојанието помеѓу центарот на RIS/релејот и изворот, а d_D го означува растојанието помеѓу центарот на RIS/релејот и дестинацијата. Според [53], претпоставката за далечно поле (far-field) е оправдана доколку се исполнети условите $\sqrt{KA} \leq 3d_S$ и $\sqrt{KA} \leq 3d_D$. Во овој случај, $\Omega_{S,k}$ и $\Omega_{D,k}$ се апроксимираат соодветно со [53] како:

$$\Omega_{S,k} \approx \frac{A \cos(\alpha_S)}{4\pi d_S^2} = \Omega_S, \forall k, \quad (2.31)$$

$$\Omega_{D,k} \approx \frac{A \cos(\alpha_D)}{4\pi d_D^2} = \Omega_D, \forall k. \quad (2.32)$$

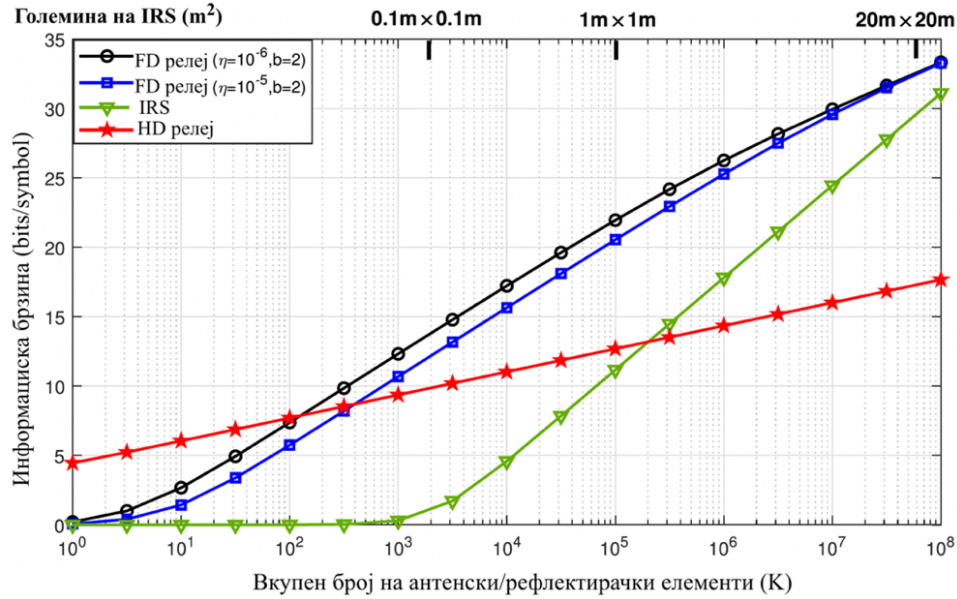
Во (2.31) и (2.32), α_S е аголот помеѓу LoS-патекаата извор-RIS/релеј и нормалата на површината на RIS/реле, а α_D е аголот помеѓу LoS-патекаата дестинација-RIS/реле и нормалата на површината на RIS/релејот.

Табела 2.1: Системски параметри за нумеричка анализа за споредба на комуникациски системи потпомогнати од RIS и релеј.

Параметар	Ознака	Вредност
Бранова должина на носителот (30 GHz)	λ	0.01 m
Површина на антенски/рефлектирачки елемент	A	$(\lambda/4)^2$
Растојание помеѓу изворот и релејот/RIS	d_S	30 m
Растојание помеѓу релејот/RIS и дестинацијата	d_D	30 m
Агол на LoS-патека на изворот	α_S	$\pi/6$
Агол на LoS-патека на дестинацијата	α_D	$-\pi/6$
Вкупна предавателна моќност	P_T	1 W
Моќност на термален шум	N_0	10^{-12} W
Број на битови за фазна квантизација	b	2 бита
Множество на фазни поместувања кај релејот	$\mathcal{P}$	$\{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$
Потрошувачка на моќност по елемент кај релејот	$p_0^{(relay)}$	0.5 mW
Потрошувачка на моќност по елемент кај RIS	$p_0^{(ris)}$	0.33 mW
Статичка хардверска потрошувачка (сите јазли)	$p_c^{(X)}$	1/3 W
Ефикасност на засилувачот на моќност	ξ	1

На Слика 2.7 е претставена постижната брзина на пренос на информации во зависност од вкупниот број на антенски/рефлектирачки елементи (K), во сценарио со доминантна LoS-компонента помеѓу сите јазли, а системските параметри се дадени во Табела 2.1. Овој график директно ги споредува трите различни технологии: предложениот двонасочно релеен систем (FD реле), пасивната интелигентна рефлектирачка површина (IRS) и конвенционалниот еднонасочен релеен систем (HD реле). Анализата и заклучоците од графикот се дадени во продолжение.

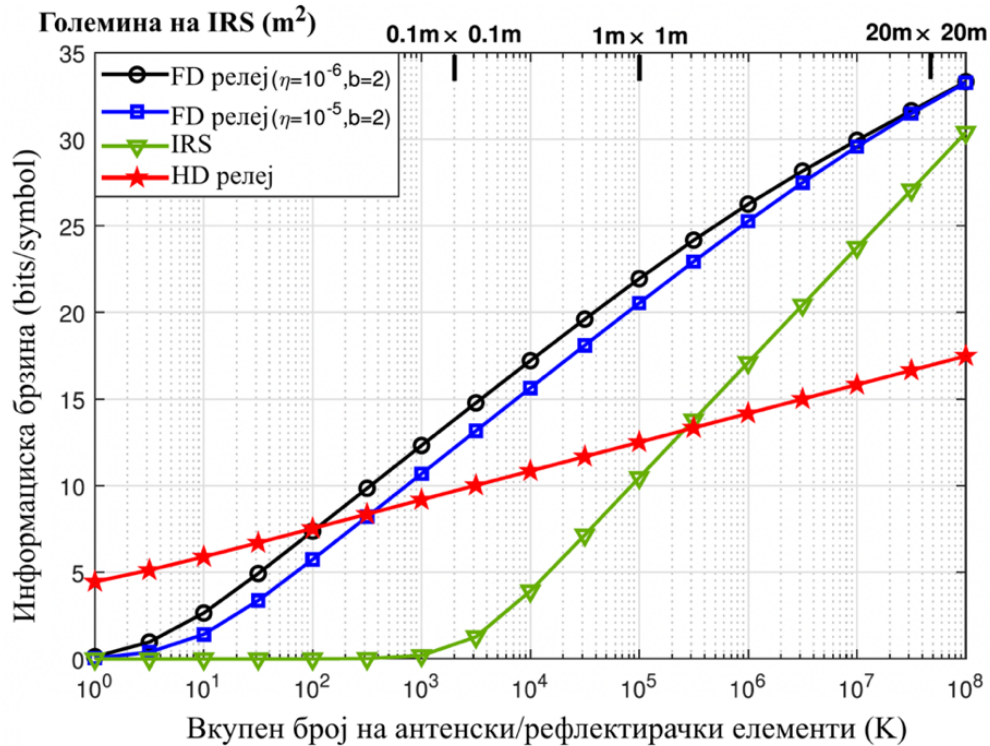
Најочигледниот заклучок од Слика 2.7 е дека предложениот двонасочно релеен систем со една РФ низа и само 2-битна фазна квантизација значително ги надминува перформансите на идеалната пасивна рефлектирачка површина за широк дијапазон на вредности на K . Иако брзината на пренос кај RIS теоретски расте со фактор $\log_2(K^2)$, таа сепак страда од т.н. „закон на двојно опаѓање на моќноста“ (анг. double path-loss). Ова значи дека сигналот слабува со производот од квадратите на растојанијата ($d_S^2 \cdot d_D^2$). Од друга страна, двонасочното реле користи механизам на „декодирај-и-препрати“, кој ефективно го прекинува двојното слабеење и овозможува прочистување на сигналот од примениот шум пред тој да биде повторно пренесен.



Слика 2.7: Информациска брзина во функција од бројот на антенски елементи K во случај на присуство на LoS-компонента.

На графикот се прикажани две криви за двонасочниот релеј за различни вредности на коефициентот на пасивна изолација: $\eta = 10^{-5}$ (50 dB изолација) и $\eta = 10^{-6}$ (60 dB изолација). Се забележува дека дури и при построгото сценарио ($\eta = 10^{-5}$), каде што сопствената интерференција е поголема, релето сè уште доминира над RIS. Разликата помеѓу двете криви за релејот е минимална кај помали вредности на K , што укажува на тоа дека предложената архитектура со една РФ низа е изводлива и оправдана во однос на сопствената интерференција. Горната x-оска на Слика 2.7 дава физичка перспектива на големината на површините. За RIS со димензии од $1m \times 1m$ (што соодветствува на $K \approx 10^5$ елементи на фреквенција од 30 GHz), двонасочниот релеј постигнува брзина на пренос која е речиси двојно поголема од онаа на RIS (околу 24 bits/symbol наспроти 13 bits/symbol кај RIS). Овој резултат е од исклучително значење за практична примена, бидејќи покажува дека во реални услови, едноставно двонасочно реле со една РФ низа е далеку поефикасно од масивна RIS-површина. Ваквото однесување се забележува за големини на RIS до $20m \times 20m$. Интересен феномен се забележува кога K достигнува екстремно високи вредности ($K > 10^8$, што одговара на површина поголема од $20m \times 20m$). Во овој регион, перформансите на RIS и FD-релејот почнуваат да се изедначуваат. Причината за ова е што при толку големи димензии, дестинацијата веќе не се наоѓа во далечното поле (far-field), туку во блиското поле (near-field) на површината. Во блиското поле, површината станува толку голема што таа практично го опкружува изворот или дестинацијата, со што загубите при простирање стануваат занемарливи. Сепак, изградбата на површини со вакви димензии е непрактична за повеќето комуникациски сценарија, што повторно ја нагласува предноста на релејот во оперативниот опсег. Иако еднонасочниот релеј користи сличен механизам за справување со шумот како двонасочното реле, неговата брзина е двојно помала од онаа на двонасочното реле. На Слика 2.7 јасно се гледа дека еднонасочното реле е супериорно во однос на RIS само за многу мали вредности на K , по што RIS го престигнува поради својата способност да работи во двонасочен режим (иако

пасивно). Предложеното двонасочно реле ги комбинира предностите на двете технологии: двонасочна комуникација како RIS и способноста за регенерирање на сигналот на релеен систем.

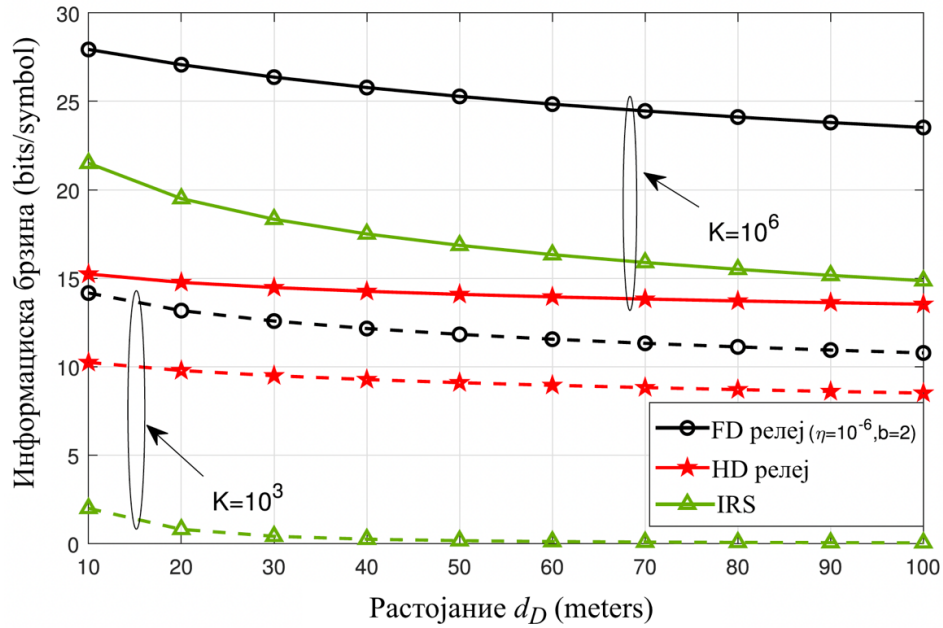


Слика 2.8: Информациска брзина во функција од бројот на антенски елементи K во случај на Рејлиев фединг.

Слика 2.8 ги прикажува остварливите брзини на пренос под истите претпоставки како Слика 2.7, но во сценарио каде нема доминантна LoS-компонента и простирањето на сигналот е моделирано според Рејлиев фединг на канал. Најзначајниот резултат на Слика 2.8 е дека добивката во перформансите на двонасочно реле во однос на рефлектирачката површина е уште поголема во споредба со LoS-сценариото. Иако федингот негативно влијае на трите системи, падот на перформансите кај RIS е многу поизразен. Оваа појава е теоретски оправдана и може да се анализира со споредба на (2.7) и (2.20), како и споредба на (2.19) со (2.23). Имено, во канали со Рејлиев фединг, просечниот однос сигнал-шум кај релето се намалува за фактор од $\pi/4$, додека кај RIS-системот, бидејќи сигналот поминува низ два независни канали со фединг (извор-RIS и RIS-дестинација), вкупното намалување на SNR е пропорционално на квадратот на овој фактор, односно $(\pi/4)^2$. Ова двојно влијание на федингот кај RIS дополнително ја нагласува неговата инфериорност во средини каде нема директна LoS-компонента. Дополнително, резултатите претставени на Слика 2.8 потврдуваат дека способноста на релето да го регенерира сигналот е критична предност во фединг канали. Со декодирање на примениот сигнал, релето спречува акумулација на ефектите од федингот од двете делници на каналот. Спротивно на тоа, RIS само пасивно го рефлектира сигналот, дозволувајќи федингот од двете патеки кумулативно да го деградира примениот сигнал кај дестинацијата. Оваа анализа укажува дека во реални урбани средини, каде што

LoS-патекаата често е блокирана и доминира федингот, предложеното двонасочно реле со едноставна 2-битна квантизација и една РФ низа е далеку посигурно и поефикасно решение од пасивната RIS-технологија.

Додека претходните слики го анализираа влијанието на бројот на елементи, Слика 2.9 го прикажува однесувањето на системите при промена на геометриската конфигурација, по-конкретно при зголемување на растојанието помеѓу посредничкиот уред (реле/RIS) и дестинацијата (d_D), додека растојанието од изворот до посредничкиот уред е фиксирано на $d_S = 25m$.



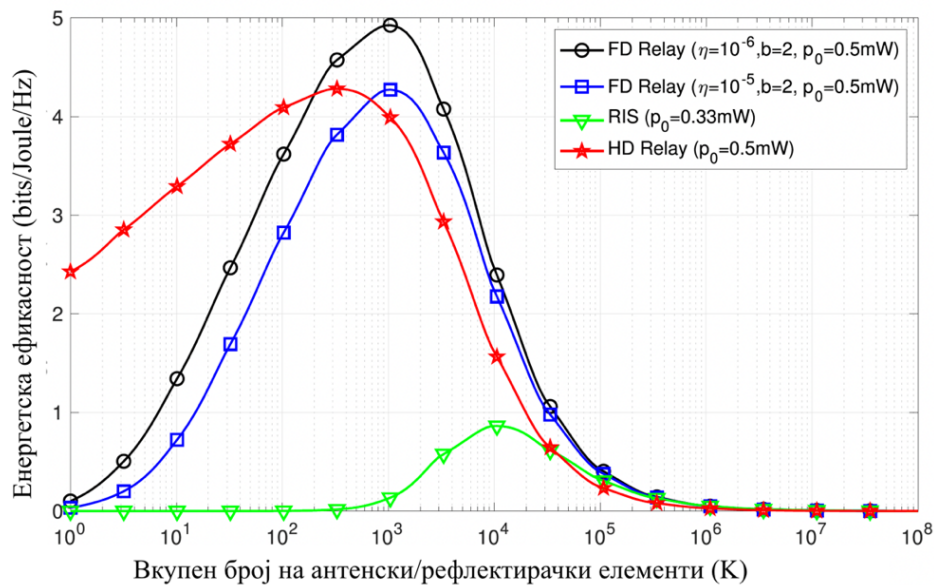
Слика 2.9: Информатиска брзина во функција од растојанието d_D .

На Слика 2.9 јасно се гледа дека трите разгледувани системи покажуваат намалување на брзината на пренос на информации со зголемување на растојанието d_D , што е во согласност со основните закони за простирање на електромагнетните бранови во слободен простор. Меѓутоа, стапката на опаѓање е драматично различна помеѓу пасивната рефлектирачка површина и предложеното двонасочно реле. Кај RIS-системот, примениот сигнал кај дестинацијата е под влијание на производот од слабеењето на двете делници ($1/d_S^2 \cdot 1/d_D^2$). Ова предизвикува брз пад на перформансите со секој изминат метар. Кај двонасочното реле, благодарение на декодирањето на сигналот, влијанието на растојанието d_D е ограничено само на втората делница (реле-дестинација), со што системот е многу поотпорен на поголеми растојанија. На графикот се прикажани две групи на криви кои нудат важен увид. Првото сценарио е со мал број елементи ($K = 10^3$) и тука предноста на двонасочното реле е најизразена. Пасивната рефлектирачка површина со мал број елементи не може да компензира за двојното слабеење на патеката, што резултира со многу ниски брзини на пренос (под 5 bits/symbol) како што растојанието се приближува кон 100 m. Во исто време, двонасочното реле одржува стабилни и значително повисоки перформанси. Второто сценарио е со голем број елементи ($K = 10^6$). Иако зголемениот број елементи кај RIS ја подобрува брзината на пренос (пора-

ди големата добивка од насочувањето), предложеното двонасочно реле со ист број елементи сепак одржува константна предност. Ова покажува дека „добивката од масивност“ кај RIS не може целосно да го надомести отсуството на активна обработка на сигналот што ја нуди релето. Дополнително, Слика 2.9 открива уште еден важен аспект: двонасочното реле конзистентно го надминува еднонасочното реле на сите растојанија. Иако и двата системи користат „декодирање-и-препраќање“, HD-релето губи 50% од своите потенцијални перформанси поради временската ортогоналност. Предложената архитектура на двонасочно реле со една РФ низа успешно ја надминува оваа загуба, докажувајќи дека сопствената интерференција (дури и при пасивна изолација) има помало влијание од губењето на спектралната ефикасност кај еднонасочните релејни системи. Анализата на Слика 2.9 укажува дека RIS-системите се екстремно зависни од прецизно позиционирање и се најефикасни само кога се наоѓаат многу блиску до изворот или дестинацијата. Спротивно на тоа, предложеното двонасочно реле нуди многу поголема флексибилност во планирањето на мрежата, бидејќи може да обезбеди висок пропусен опсег дури и кога е поставено на поголеми растојанија помеѓу комуникациските јазли (извор – дестинација).

Слика 2.10 ја прикажува енергетската ефикасност (EE) на разгледуваните системи, изразена во bits/Joule/Hz , во зависност од вкупниот број на елементи (K). За оваа споредба, релејниот систем е моделиран со користење на реконфигурабилни холографски површини за приемната и предавателната страна, што овозможува фер споредба со пасивната RIS-технологија во поглед на хардверската комплексност и потрошувачката на енергија. Во моделот за енергетска ефикасност се земени предвид три клучни компоненти на потрошувачката на моќност: (i) Статичка моќност: потрошувачката на електричните кола која е константна за сите системи; (ii) Моќност по елемент (p_0): потрошувачката на секој поединечен фазен поместувач – тука релејот е во блага негатива (0.5 mW наспроти 0.33 mW кај RIS), но оваа разлика е занемарлива во споредба со придобивките во брзината на пренос; и (iii) Предавателна моќност (P_T): и двата системи користат идентична вкупна моќност од 1 W . Меѓутоа, релето ја користи оваа моќност многу поефикасно преку процесот на прочистување на шумот кај механизмот „декодирање-и-препраќање“. Енергетската ефикасност е дефинирана како сооднос помеѓу постижната брзина на пренос на информации (R) и вкупната потрошена моќност (P_{total}), како што беше дефинирано во (2.27). Клучниот заклучок од Слика 2.10 е дека предложениот двонасочно релеен систем конзистентно покажува повисока енергетска ефикасност од рефлектирачката површина, за сите разгледувани вредности на K . Иако на прв поглед се чини дека пасивната рефлектирачка површина треба да биде енергетски поефикасна бидејќи нема активни РФ низи за засилување и обработка на сигналот, резултатите го покажуваат спротивното. Причината лежи во фактот што информациската брзина (R) кај релето расте многу побрзо отколку неговата потрошувачка на моќност во споредба со RIS. Исто така, на графикот се забележува специфичен тренд – за мали вредности на K , енергетската ефикасност расте со зголемување на бројот на елементи, бидејќи добивката во брзината на пренос е доминантна. Меѓутоа, како што K станува многу голем, енергетската ефикасност почнува да опаѓа кај двата системи. Ова се случува затоа што вкупната потрошувачка на моќност на антенската површина ($K \cdot p_0$) почнува да доминира над вкуп-

ната предавателна моќност, со што се намалува вкупната ефикасност на системот. Дури и во овој регион, релето ја задржува предноста над рефлектирачката површина.



Слика 2.10: Енергетска ефикасност во функција од бројот на антенски/рефлектирачки елементи во присуство на LoS-компонента.

Со користење на напредни материјали и аналогна обработка на сигналот, релето може да понуди супериорни комуникациски капацитети со помалку потрошена енергија по пренесен бит информација во споредба со RIS. Анализата на Слика 2.10 е силен аргумент во прилог на релејните системи за идните 6G мрежи, каде што енергетската одржливост е од големо значење. Таа покажува дека најефикасниот систем не е оној кој троши најмалку енергија, туку оној кој најпаметно ја користи енергијата за да постигне максимален пренос на податоци.

Дел II

Пресметување и учење на работ на мрежата во безжични комуникации

Глава 3

Пресметување на работ на мрежата

Сеприсутноста на огромен број мобилни уреди, сензори и машини во новата ера на IoT води кон нови предизвици и побарувања од мобилни мрежи [61]. Бидејќи IoT-уредите имаат ограничувања во однос на пресметковната моќ, капацитетот на батеријата и брзината, не се во можност да опслужат апликации кои побаруваат поголеми мемориски или пресметковни ресурси [62]. Традиционално, овој проблем се адресираше со пресметка во „облак“ [63]. Со претовар (анг. offloading) на пресметковни задачи и податоци до „облакот“, апликациите кои побаруваат многу ресурси, може да се извршуваат на моќни сервери, наместо кај крајните корисници [64]. Сепак, заради големото физичко растојание од мобилните корисници до серверот каде се врши пресметка во „облак“, оваа технологија не е изводливо решение за голем број на временски критични апликации кои побаруваат пресметки во реално време [8].

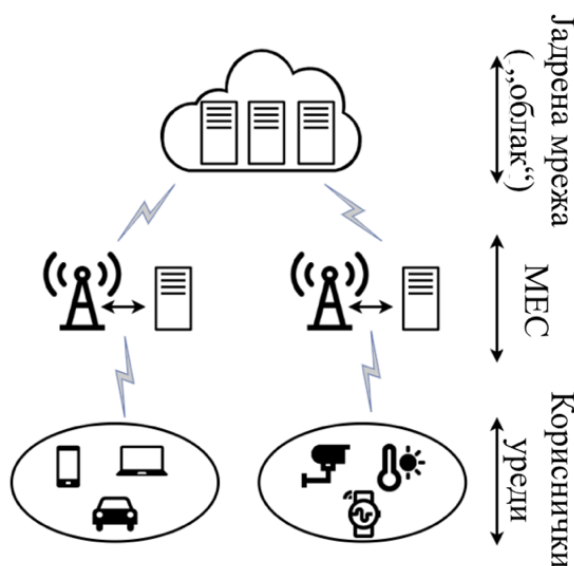
Концептот на пресметка на работ за прв пат беше предложен од страна на европскиот телекомуникациски стандардизациски институт ETSI во 2014 година, дефиниран како нова комуникациска рамка која обезбедува пресметковни способности во делот на радиоприс-тапната мрежа, во непосредна близина на мобилните корисници [22]. Имено, МЕС нуди можност за складирање на податоци и обработка на истите на „работ“ на мобилната мрежа, но во рамки на радио-пристапниот дел на мрежата. Податоците се праќаат на обработка до т.н. „рабни“ сервери, кои може, но и не мора да бидат колоцирани со базната станица. Ова овозможува децентрализирана обработка на податоци и обезбедува подобрени услуги како што се значително намалено доцнење и зголемен пропусен опсег [65–68]. Со поставување на МЕС-серверите блиску до крајните корисници, се избегнуваат високите пропагационски доцнења својствени за пресметката во „облак“, која се одвива на големи географски растојанија. Ова ја прави МЕС-технологијата клучна за временски-критични апликации и клучна технологија за 5G/6G-системи [69, 70].

МЕС-мрежите се карактеризираат со трислојна архитектура (Слика 3.1): (1) Слој на кориснички уреди, (2) МЕС слој и (3) Слој на јадрена мрежа („облак“). Најнискиот слој вклучува мобилни, IoT-уреди и сензори поврзани со јадрената мрежа преку базни станици. МЕС-слојот, составен од базни станици со интегрирани МЕС-сервери, ја оптимизира радио-пристапната мрежа, пренесувајќи „интелигенција“ и пресметковни способности до базната станица. Оваа децентрализирана архитектура, за разлика од целосно централизираната „пресметка во облак“, ја подобрува ефикасноста и овозможува поврзување на уредите со најблискиот МЕС-сервер. Последниот слој, јадрената мрежа/„облак“, ги поврзува сите базни станици и МЕС-сервери со пошироката мрежа. Бидејќи главната задача на МЕС-слојот

е да обезбеди складирање и обработка на податоци на „работ“ на пристапната мрежа, перформансите на МЕС-мрежата силно зависат од нејзината соодветна архитектура [69].

Иако пресметката на „работ“ како технологија еволуираше од пресметката во „облак“, истата нуди неколку клучни предности:

- **Близина:** со МЕС, пресметковните ресурси и услуги се приближуваат до физичката локација на корисникот, намалувајќи го доцнењето при пристап до услугите и овозможувајќи нови можности како што се анализа на корисничкото однесување во реално време [68].
- **Мало доцнење:** целокупното доцнење во мобилна мрежа се состои од пропагациско, процесирачко и доцнење поради брзина на пренос на податоците. Со лоцирање на МЕС-сервери блиску до корисниците (десетици метри до еден километар), пропагациското доцнење е драстично намалено, за разлика од стотиците километри кај пресметка во „облак“. Ова ја прави МЕС-технологијата клучна за временски-критични апликации како Тактилен интернет и IoT [69].
- **Заштеда на енергија:** еден од главните предизвици кај IoT-уредите е оптимизацијата на потрошувачката на енергија поради ограничениот капацитет на батеријата. Со „претовар“ на пресметковни задачи, МЕС го намалува оптоварувањето на уредот, со што го продолжува животниот век на батеријата и ја намалува потрошувачката на моќност [66].



Слика 3.1: Трослојна МЕС-архитектура.

3.1 Механизми за пресметување на работ на мрежата

Благодарение на можноста за „претовар“ на пресметковни задачи во МЕС-системите, потрошувачката на енергија кај крајните уреди е намалена, а процесот на обработка на пода-

тоците е забрзан, што го намалува доцнењето предизвикано од процесирање на податоците. Главниот предизвик во врска со „претоварот“ на пресметките е да се одлучи дали „претоварот“ воопшто треба да се направи, но и колку и кој дел од податоците треба да се пратат на МЕС-серверот. Со оглед на ова, ние разликуваме три различни модели на „претовар“ на пресметковни задачи:

1. Локално процесирање (анг. local computing), кога не се врши „претовар“, т.е. целата обработка се врши локално на страна на корисничкиот уред и овој модел е применлив кога „претоварот“ на податоци до МЕС-серверот не е изводливо или не се исплати;
2. Целосен „претовар“ (анг. fully offloading) на пресметковни задачи, кога пресметковната задача е релативно едноставна или високо интегрирана и мора да биде обработена во целост, а со тоа и „претоварена“ во целост до МЕС-серверот; и
3. Делумен „претовар“ (анг. partial offloading), кога пресметковната задача е поделена на два дела, од кои еден дел се обработува локално кај корисничкиот уред, а остатокот се праќа до МЕС-серверот, каде се обработува.

Локалното процесирање и целосниот „претовар“ во литературата може да се најдат под едно име, бинарен „претовар“ (анг. binary offloading). Изборот на соодветен модел за „претовар“ е ограничен од многу фактори, вклучувајќи ги пресметковните способности на крајните корисници и МЕС-сервери, квалитетот на радиоканалот, типот на апликација и барањата на корисникот.

За да може да се спроведе методолошко академско истражување поврзано со МЕС, најпрво треба да се најде соодветен начин за моделирање на пресметковните задачи. Треба да се напомене дека пресметковните задачи можат да бидат под влијание на различни параметри како што се големината на задачата, интензитетот на пресметување, доцнењето, искористеноста на пропусниот опсег итн., и затоа развивањето на прецизни модели за пресметковни задачи е исклучително сложен процес. Со цел правилно да се опишат својствата поврзани со потрошувачката на енергија и доцнењето како најбитни параметри за МЕС-системите, во оваа дисертација усвојуваме математички обработлив модел на пресметковна задача, кој исто така е широко користен и во постоечката МЕС-литература [71].

Една дадена пресметковна задача може целосно да се карактеризира со следната подредена тројка од параметри $[I; L; O]$. Овде, I ја означува големината на влезните податоци на пресметковната задача изразена во бити, L е количината на потребни пресметковни ресурси за пресметување на 1 бит од влезните податоци (т.е. бројот на потребни циклуси на централната процесорска единица (анг. central processing unit) (CPU)), познато и како пресметковно оптоварување/интензитет, а $O \in (0, 1)$ е односот помеѓу големината на излезните и влезните податоци на задачата, што значи дека пресметувањето на I бити влезни податоци ќе генерира OI бити излезни податоци за конкретниот кориснички уред. Овој модел на пресметковна задача не само што ги опфаќа суштинските својства на мобилните апликации поврзани со пресметковните и комуникациските барања, туку овозможува и математичка обработливост, како што ќе биде прикажано во следните глави.

Потрошувачката на енергија на еден пресметковен сервер/процесор е заеднички определена од употребата на централната процесорска единица, меморијата достапна за складирање на податоци, работната меморија, мрежните интерфејси, итн. Бидејќи придонесот на CPU е доминантен помеѓу овие фактори, тој е главниот фокус кој нашироко се користи во постоечката релевантна литература. Што се однесува до моќноста на CPU, таа се состои од динамичка моќност, моќност на куса врска и моќност на истекување, при што динамичката моќност доминира, а останатите компоненти се занемарливи во споредба со динамичката моќност [72]. Како резултат на тоа, ја земаме предвид само динамичката моќност, означена како P_{com} , која е пропорционална на производот од $V^2 f$ под претпоставка на низок напон на CPU, каде што V е напонот на колото во волти (V) и f е фреквенцијата на часовникот на CPU во циклуси/секунда [73]. Понатаму, во [74] е забележано дека фреквенцијата на часовникот на CPU-чиповите на пресметковниот сервер/процесор, f , е приближно линеарно пропорционална на напонот V . Со други зборови, P_{com} треба да биде линеарно пропорционална на f^3 , и со тоа може да се запише како $P_{com} = \alpha f^3$, каде што α е коефициентот на ефективна капацитивност кој зависи од архитектурата на чипот на пресметковниот сервер/процесор. Следствено, единечната потрошувачка на енергија на пресметковниот сервер/процесор за извршување на еден CPU-циклус може да се претстави како:

$$E_{com} = P_{com} t_{com} = \alpha f^3 \times \left(\frac{1}{f}\right) = \alpha f^2, \quad (3.1)$$

каде $t_{com} = \frac{1}{f}$ е времето за да се изврши еден CPU-циклус [72]. Врз основа на претходно презентираниот пресметковен модел, т.е. $[I; L; O]$ со I бита влезни податоци, каде за пресметување на секој бит се потребни L CPU-циклуси, потрошувачката на енергија за завршување на оваа пресметковна задача може да се пресмета како:

$$E_{task} = IL E_{com} = \alpha IL f^2. \quad (3.2)$$

Соодветно, времето потребно за пресметка, односно завршување на $[I; L; O]$ со извршување на IL CPU-циклуси може да се изрази како:

$$t_{task} = \frac{IL}{f}. \quad (3.3)$$

За ефикасно користење на енергијата за пресметување, пресметковните сервери/процесори можат да ја искористат техниката на динамичко скалирање на напон и фреквенција (анг. dynamic voltage and frequency scaling) (DVFS). На овој начин, енергијата потрошена за пресметување може адаптивно да се контролира преку прилагодување на нивната CPU-фреквенција за секој CPU-циклус [75].

3.2 Интеграција со безжичен пренос на енергија

Иако МЕС има многу предности како што наведовме во претходниот дел, целосното искористување на моќните пресметковни ресурси на работ на мрежата сè уште се соочува со

неколку предизвици. Меѓу нив, ограниченото и недоволно напојување со енергија претставува едно од главните ограничувања кај кориснички уреди кои се напојуваат од батерии. Затоа е соодветно да се искористи технологијата на безжичен пренос на моќност (анг. wireless power transfer) (WPT), која е позната како ветувачко решение за обезбедување на практично и одржливо снабдување со енергија за безжичните уреди [23].

Имено, корисничките уреди (мобилните уреди, сензори, IoT-уреди) кои се опремени со електрични кола за акумулација на енергија (анг. energy harvesting) (EH) и акумулираат РФ енергија од пристапни точки или БС ги именуваме како EH единици (анг. energy harvesting units) (EHUs). Да претпоставиме дека предавателната моќ на БС е означена со P_0 , а ефективното канално засилување помеѓу БС и конкретниот EHU е h . Следствено, акумулираната енергија за овој EHU во текот на еден временски слот T може да се пресмета како [76]:

$$E_{har} = \eta \tau_0 P_0 \Omega_i, \quad (3.4)$$

каде што е усвоен линеарниот модел за акумулирање на енергија, бидејќи претпоставуваме дека влезната РФ моќност на корисничките уреди е во рамките на линеарниот режим на исправувачот, а η е ефикасноста на енергетска конверзија на корисничкиот уред.

Интеграцијата на безжичниот пренос на енергија со пресметката на работ на мрежата создава услови за динамична и возбудлива истражувачка област, фокусирана на создавање самоодржливи пресметковни системи за уреди со ограничени енергетски ресурси. Суштинскиот предизвик во WP-MEC лежи во дизајнот и балансот меѓу преносот на енергија во насока од БС до EHUs (анг. downlink) (DL) и претоварот на пресметковни задачи во насока од EHUs кон БС (анг. uplink) (UL). Ефективната распределба на ресурси е од суштинско значење за искористување на целосниот потенцијал на овој пристап.

Првичниот бран на истражувања во оваа област се фокусираше на воспоставување на фундаменталните принципи и основните компромиси. Еден од пионерските трудови е оној на You et al. [24] кој, иако е во контекст на пресметување во облак, ги постави темелите за анализа на енергетската ефикасност во безжично напојувани мрежи каде пресметката се прави во облак. Тие го воведуваат моделот со три ентитети (извор на енергија, корисник и пристапна точка) и го анализираат компромисот помеѓу локалното пресметување и претоварот, покажувајќи дека постои оптимална политика за претовар која ја минимизира потрошувачката на енергија. Надоврзувајќи се на овие основи, во трудот [25] за прв пат се дефинира архитектура на WP-MEC. Авторите предлагаат систем каде еден корисник, напојуван од пристапна точка (БС), мора да донесе бинарна одлука за претовар. Во овој труд е формализиран протоколот „акумулирај-па-пренеси“ (анг. harvest-then-transmit) и поставена е јасна оптимизациска цел: максимизирање на пресметковната рата.

Природната прогресија е проширувањето кон повеќекориснички сценарија, што воведува нови предизвици како повеќекратен пристап, интерференција и правична распределба на ресурси. Wang et al. [26] го проширија WP-MEC концептот на повеќекорисничко сценарио користејќи TDMA. Во овој труд е развиена рамка за дизајн на MEC-WPT системи со заедничка оптимизација на енергетското зрачење (анг. beamforming), претоварот и локалната пресметка, со цел да се минимизира вкупната потрошувачка на енергија на корисниците,

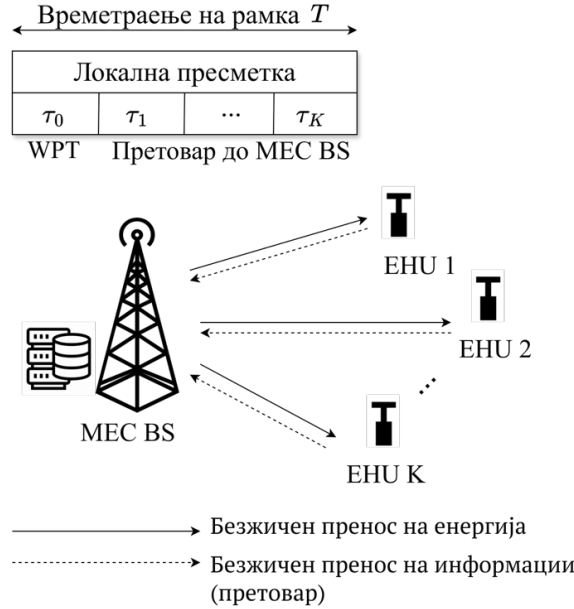
а притоа задоволувајќи ги индивидуалните ограничувања за доцнење на истите. Авторите во [27] воведоа нова целна функција, а тоа е пресметковната ефикасност, дефинирана како вкупниот број на обработени бити поделен со вкупната потрошена енергија за обработка (бити по цул). Клучен придонес во овој труд е промената на целта на оптимизацијата од едноставна максимизација на пресметковната ефикасност кон макс-мин оптимизација, што воведува правична распределба на ресурси. Ова е критично важно во WP-MEC, особено заради појавата на двојниот ефект близу-далеку (анг. double near-far effect). Имено, РФ енергијата емитувана од БС е изложена на слабеење на патеката низ каналот од БС до корисникот кој собира енергија, ENU, како и на слабеење на патеката на комуникацискиот сигнал пренесен преку повратниот канал. Дополнително, резултатите во трудот покажуваат дека делумниот претовар нуди супериорни перформанси во однос на бинарниот претовар, како и дека NOMA секогаш постигнува предност во однос на ортогоналниот, претставен преку TDMA. Во [28] авторите се фокусираат директно на компромисот помеѓу енергетската ефикасност и доцнењето во повеќекориснички WP-MEC систем, анализирајќи и споредувајќи различни шеми за повеќекратен пристап (TDMA и OFDMA).

3.3 Оптимизација на ресурси според max-min критериум

Во продолжение ќе бидат презентирани шеми за распределба на ресурси кои гарантираат правичност во безжично напојуван MEC-систем, каде корисниците применуваат или бинарен или делумен претовар на пресметковните задачи. Целта е да се максимизира минималната пресметковна рата (број на обработени бити) за сите корисници, преку заедничка оптимизација на времетраењето на фазата за безжичен пренос на енергија, времетраењето на фазата безжичен пренос на информација, како и на пресметковните брзини. Разгледуваме безжично напојуван MEC-систем кои се состои од една базна станица, БС, и K корисници кои акумулираат енергија, ENUs. БС е поврзана на стабилно и постојано напојување и покрај за комуникација се користи и како предавател на РФ енергија до корисниците. Секој корисник пак, е опремен со електрично коло за акумулирање енергија и батерија на полнење каде се складира собраната енергија, која подоцна се користи за напојување на неговите функции (пресметка и комуникација). Сите ENUs, како и БС, се опремени со по една антена. Базната станица ги интегрира следниве функционалности: (1) комуникациски предавател/приемник – прима информации од корисниците во uplink насока и пренесува информации до корисниците во downlink насока; (2) РФ предавател – зрачи РФ енергија до корисниците; и (3) MEC-сервер – ги обработува/пресметува претоварените податоци од корисниците.

За да се избегне интерференција меѓу корисниците, користиме ортогонален повеќекориснички пристап со поделба на временските ресурси, поточно времето е поделено на TDMA-рамки со времетраење T . Секоја TDMA-рамка се состои од фаза за пренос/акумулирање на енергија (WPT-фаза) со времетраење τ_0 , проследена со последователни фази за пренос на информации со времетраење $\tau_1, \tau_2, \dots, \tau_K$, за секој корисник соодветно, како што е прика-

жано на Слика 3.2. За време на WPT-фазата, базната станица зрачи РФ сигнали со моќност P_0 која е наменета за безжично напојување на сите корисници истовремено. За време на i -тата фаза на пренос на информации, i -тиот корисник врши претовар на своите пресметковни задачи/информации до БС, а истите податоци треба да се обработат во МЕС-серверот.



Слика 3.2: Системски модел на безжично напојуван МЕС-систем со K корисници.

Времетраењето на сите фази во склоп на TDMA-рамката треба да задоволи следниов израз:

$$\tau_0 + \tau_1 + \tau_2 + \dots + \tau_K = T. \quad (3.5)$$

Ако некои од корисниците (на пр. j -тиот корисник) вршат само локално пресметување, соодветно, времетраењето на фазата за пренос на информации за тој корисник ќе биде нула ($\tau_j = 0$), додека локалното пресметување се врши во текот на целото времетраење на TDMA-рамката. Во WPT-фазата, БС зрачи РФ енергија, при што секој корисник акумулира енергија, а количината на акумулирана енергија зависи од каналот меѓу БС и дадениот корисник. Така, енергијата собрана од i -тиот корисник за време на WPT-фазата е [76]:

$$E_i^{EH} = \eta \tau_0 P_0 \Omega_i, \quad (3.6)$$

каде P_0 е предавателната моќност на БС, $\eta \in (0, 1)$ е коефициентот на ефикасност на енергетска конверзија на ЕН-колото, а $h_i = \Omega_i$ е коефициентот на каналот меѓу БС и i -тиот корисник. Ако претпоставиме дека безжичниот канал h_i е статичен, тогаш неговата вредност зависи само од слабеењето заради пропагација и може да се одреди во зависност од растојанието помеѓу БС и самиот корисник, r_i . Претпоставуваме експонент на слабеење со вредност 3 и слабеење од 30 dB на референтно растојание од 1 метар. Тогаш засилувањето на каналот се пресметува според:

$$h_i = \Omega_i = \frac{10^{-3}}{r_i^3}. \quad (3.7)$$

Разгледуваме два модели за претовар на пресметковни задачи до МЕС-серверот: (1) делумен претовар и (2) бинарен претовар.

Режим на локално пресметување (режим-0). Кога се работи во режим на локално пресметување, пресметувањето се врши во текот на целото времетраење на рамката T . Нека параметарот f_i ја означува брзината на пресметка на процесорот на i -тиот корисник, изразена во број на CPU-циклуси во единица време. Нека L го означува пресметковниот интензитет, односно бројот на CPU-циклуси потребни за обработка на еден бит податоци, а α го означува коефициентот на енергетска ефикасност на процесорот на корисникот. Во тој случај, бројот на локално пресметани битови од i -тата корисничка станица ($1 \leq i \leq K$) согласно (3.3) се пресметува како:

$$d_{L,i} = \frac{T f_i}{L}. \quad (3.8)$$

Со тоа, потрошувачката на енергија за локално пресметување кај i -тиот корисник согласно (3.2) се пресметува како:

$$E_{L,i} = \alpha T f_i^3. \quad (3.9)$$

Режим на претовар (режим-1). Кога работи во режим-1, корисникот може да ја пренесе својата пресметковна задача на МЕС-серверот на БС. Нека P_i ја означува излезната моќ на i -от корисник. Бројот на бити што може да се претоварат кон МЕС БС одговара на капацитетот на безжичниот канал помеѓу соодветниот корисник и БС:

$$d_{O,i} = B \tau_i \log_2 \left(1 + \frac{P_i \Omega_i}{N_0} \right), \quad (3.10)$$

каде B е пропусниот опсег на комуникацискиот канал помеѓу корисникот и БС. Со тоа, потрошената енергија за претовар на податоци од корисникот до БС се пресметува според изразот:

$$E_{O,i} = \tau_i P_i. \quad (3.11)$$

Бидејќи корисниците применуваат стратегија „акумулирај-па-пренеси“ [77], корисникот за време на една TDMA-рамка ја троши енергијата која ја складира за време на WPT-фазата од таа рамка. Следствено, според законот за зачувување на енергијата, i -тиот корисник мора да го задоволи следново неравенство:

$$\alpha T f_i^3 + \tau_i P_i \leq \eta \tau_0 P_0 \Omega_i. \quad (3.12)$$

Стратегиите за распределба на ресурси за МЕС-системи обично се дизајнирани да ја максимизираат вкупната пресметковна рата во системот, што води до неправична распределба на ресурси, бидејќи оние EНUs кои се поблиску до БС добиваат несразмерно голем удел од системските ресурси. За да се справиме со овој проблем, предлагаме правична распределба на ресурси според макс-мин критериумот [78]. Затоа, за разгледуваниот системски модел, го предлагаме следниов оптимизациски проблем кој го максимизира минималниот број на

пресметани битови:

$$\text{Maximize}_{\tau_0, \tau_i, P_i, f_i} \min_{1 \leq i \leq K} \left\{ \frac{T f_i}{L} + B \tau_i \log_2 \left(1 + \frac{P_i \Omega_i}{N_0} \right) \right\} \quad (3.13)$$

$$\text{под услов: } C1 : \sum_{i=1}^K \tau_i + \tau_0 = T \quad (3.14)$$

$$C2 : \alpha T f_i^3 + \tau_i P_i \leq \eta \tau_0 P_0 \Omega_i, \forall i \quad (3.15)$$

$$C3 : \tau_i \geq 0, \tau_0 \geq 0, P_i \geq 0, f_i \geq 0, \forall i \quad (3.16)$$

каде C1 се однесува на вкупното времетраење на сите фази во една TDMA-рамка, C2 се однесува на законот за зачувување на енергијата, додека ограничувањето C3 природно бара сите оптимизациски променливи да имаат ненегативна вредност. Оптимизацискиот проблем не е конвексен, бидејќи во ограничувањето C2 имаме производ на две оптимизациски променливи. Сепак, со едноставна математичка трансформација може да се доведе до конвексен и лесно може да се реши нумерички со користење на различни софтверски алатки, на пример, CVX.

Оптимизацискиот проблем (3.13) може да се реши дури и аналитички во идеален случај кога сите кориснички станици се на исто растојание од БС, односно се поставени на кружница околу БС, каде $\Omega_1 = \Omega_2 = \dots = \Omega_K$, $\tau_1 = \tau_2 = \dots = \tau_K$, $f_1 = f_2 = \dots = f_K$ и $P_1 = P_2 = \dots = P_K$. Решението на оптимизацискиот проблем под овие услови е дадено во следната теорема.

Теорема 3.1. *Оптималната предавателна моќност на корисникот, P_1^* , се пресметува како:*

$$P_1^* = \frac{1}{\Omega} \left[\frac{K \eta P_0 \Omega^2 - N_0}{W \left(\frac{K \eta P_0 \Omega^2 - N_0}{N_0 e} \right)} - N_0 \right], \quad (3.17)$$

каде $W(\cdot)$ е Ламбертовата W -функција [79]¹. Оптималната брзина на локално пресметување f_1^* е:

$$f_1^* = \left(\frac{\ln 2 (N_0 + P_1^* \Omega)}{3 B \Omega \alpha L} \right)^{1/2}. \quad (3.18)$$

Оптималното времетраење на фазата за претовар, означено со τ_1^* , е:

$$\tau_1^* = T \frac{\eta P_0 \Omega - \alpha (f_1^*)^3}{K \eta P_0 \Omega + P_1^*}. \quad (3.19)$$

Оптималното времетраење на WPT-фазата, τ_0^* , е:

$$\tau_0^* = T - K \tau_1^*. \quad (3.20)$$

Доказот на Теорема 3.1 и деталите за тоа како се добиени (3.17)–(3.20) се дадени во дода-

¹ Ламбертовата W -функција е инверзна функција на функцијата $f(w) = w e^w$. За кој било комплексен број z , таа е дефинирана како решение на трансцендентната равенка $W(z) e^{W(z)} = z$. За реални вредности на аргументот $x \geq -1/e$, функцијата е повеќезначна и се состои од две реални гранки: главна гранка $W_0(x)$ за $W(x) \geq -1$ и долна гранка $W_{-1}(x)$ за $W(x) \leq -1$.

ток Б.4.

Во случај на бинарен претовар, оптимизацискиот проблем (3.13) се прилагодува за да се дизајнира соодветна шема за распределба на ресурси. За таа цел, ќе ја користиме следнава индикаторска променлива:

$$I_i = \begin{cases} 0, & i\text{-тиот ЕНУ е во режим-0} \\ 1, & i\text{-тиот ЕНУ е во режим-1} \end{cases}. \quad (3.21)$$

Така, може да го дефинираме следниов max-min оптимизациски проблем:

$$\text{Maximize}_{\tau_0, \tau_i, P_i, f_i, I_i} \quad \min_{1 \leq i \leq K} \left\{ (1 - I_i) \frac{T f_i}{L} + I_i B \tau_i \log_2 \left(1 + \frac{P_i \Omega_i}{N_0} \right) \right\} \quad (3.22)$$

$$\text{под услов: } C1 : \sum_{i=1}^K \tau_i + \tau_0 = T \quad (3.23)$$

$$C2 : \alpha T f_i^3 + \tau_i P_i \leq \eta \tau_0 P_0 \Omega_i, \forall i \quad (3.24)$$

$$C3 : I_i \in \{0, 1\} \quad (3.25)$$

$$C4 : \tau_0 \geq 0, \tau_i \geq 0, P_i \geq 0, f_i \geq 0, \forall i \quad (3.26)$$

Оптимизацискиот проблем (3.22) не е конвексен, бидејќи целната функција и ограничувањата содржат производи од две оптимизациски променливи. Сепак, истиот може да се реши за која било можна комбинација на индикаторски променливи $\{I_i\}_{i=1}^K$, каде има вкупно 2^K можни комбинации. За секој даден сет на индикаторски променливи, (3.22) е конвексен оптимизациски проблем и може да се реши нумерички. Потоа, од сите 2^K можни решенија, оптималното решение е она што ја максимизира целната функција.

Ако претпоставиме симетричен случај, каде сите корисници се поставени на кружница со радиус r , тогаш поради симетријата сите корисници или ќе растовараат или ќе пресметуваат локално ($I_1 = I_2 = \dots = I_K$), ќе пресметаат ист број бити, во исто време за пренос на податоците ($\tau_1 = \tau_2 = \dots = \tau_K$) со иста брзина на пресметување ($f_1 = f_2 = \dots = f_K$), и иста предавателна моќност ($P_1 = P_2 = \dots = P_K$). Бидејќи индикаторската променлива I_1 може да има две вредности, го претставуваме проблемот со два одделни случаи.

Случај 1: $I_1 = 1$ – сите корисници вршат претовар на пресметковни задачи кон БС. Оптимизацискиот проблем ја добива формата:

$$\text{Maximize}_{\tau_0, \tau_1} \quad B \tau_1 \log_2 \left(1 + \frac{A \tau_0}{\tau_1} \right) \quad (3.27)$$

$$\text{под услов: } C1 : K \tau_1 + \tau_0 = T \quad (3.28)$$

$$C2 : \tau_0 \geq 0, \tau_1 \geq 0, \quad (3.29)$$

каде помошната променлива е $A = \frac{\eta P_0 \Omega_1^2}{N_0}$.

Теорема 3.2. Оптималното време за претовар на информации, τ_1^* , се наоѓа како решение

на следнава трансцендентна равенка:

$$\frac{\tau_1^* + A(T - K\tau_1^*)}{AT} \ln \left(1 + A \frac{T - K\tau_1^*}{\tau_1^*} \right) = 0. \quad (3.30)$$

Оптималното време за акумулирање на енергија, τ_0^* , е дадено со:

$$\tau_0^* = T - K\tau_1^*. \quad (3.31)$$

Случај 2: $I_1 = 0$ – сите корисници вршат локално пресметување ($\tau_0 = T$). Оптимизациониот проблем го добива следниов облик:

$$\text{Maximize}_{f_1} \quad \frac{Tf_1}{L} \quad (3.32)$$

$$\text{под услов: } C1 : \alpha f_1^3 \leq \eta P_0 \Omega_1 \quad (3.33)$$

$$C2 : f_1 \geq 0 \quad (3.34)$$

Во овој случај, оптималната брзина на локално пресметување, f_1^* , е дадена со:

$$f_1^* = \left(\frac{\eta P_0 \Omega_1}{\alpha} \right)^{1/3}. \quad (3.35)$$

Равенките (3.30), (3.31) и (3.35) се добиени следејќи слични чекори како во додаток Б.4.

Во продолжение, ги претставуваме и анализираме перформансите на двете предложени шеми за распределба на ресурси, едната базирана на делумен претовар, а другата на бинарен претовар. За да ги оцениме перформансите, ги користиме следниве две значајни перформансни метрики:

1. Вкупна пресметковна рата во системот. Оваа метрика го претставува вкупниот број на обработени бити од сите корисници во системот во единица време (бити/секунда):

$$R = \sum_{i=1}^K R_i, \quad (3.36)$$

каде пресметковната рата на i -тиот корисник, R_i , е дефинирана како:

$$R_i = \frac{f_i}{L} + \frac{B}{T} \tau_i \log_2 \left(1 + \frac{P_i \Omega_i}{N_0} \right). \quad (3.37)$$

2. Индекс на правичност на Џејн (анг. Jain's Fairness Index). Оваа метрика, дефинирана во [80], го квантифицира нивото на правичност при распределба на ресурси меѓу корисници. Вредност од 1 означува совршена праведност, додека вредностите поблиску до 0 означуваат поголема неправичност:

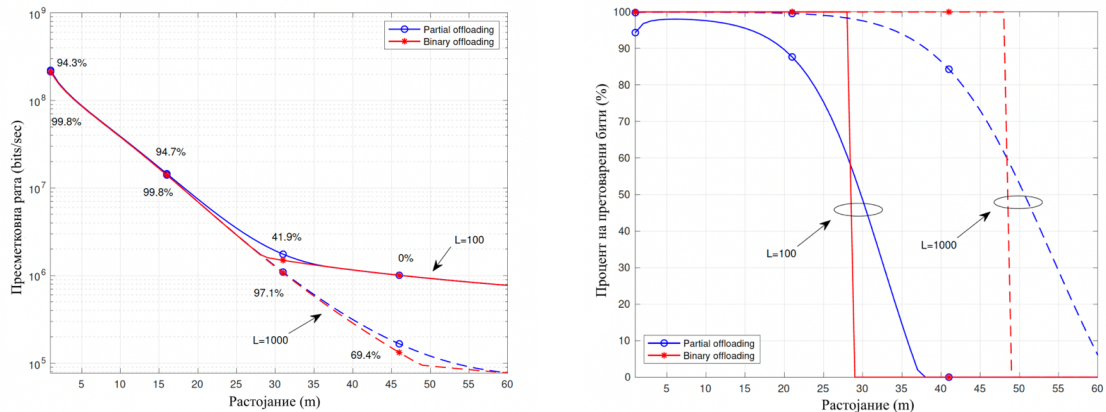
$$J = \frac{\left(\sum_{i=1}^K R_i \right)^2}{K \sum_{i=1}^K R_i^2}. \quad (3.38)$$

Табела 3.1: Работни параметри за предложениот безжично напојуван МЕС-систем.

Параметар	Симбол	Вредност
Коефициент на ефикасност на колото за акумулирање енергија	η	1
Коефициент на енергетска ефикасност на процесорот на корисникот	α	10^{-28} Js
Времетраење на една TDMA-рамка	T	1 s
Моќност на термички шум на влезот	N_0	10^{-12} W/Hz
Пропусен опсег на каналот	B	10 MHz
Број на кориснички станици (EHUs)	K	10
Предавателна моќност на базната станица	P_0	10 W

Прво, го анализираме идеалниот, симетричен случај каде сите корисници се наоѓаат на исто растојание од БС, т.е. на кружница со радиус r , а останатите системски параметри се дадени во Табела 3.1.

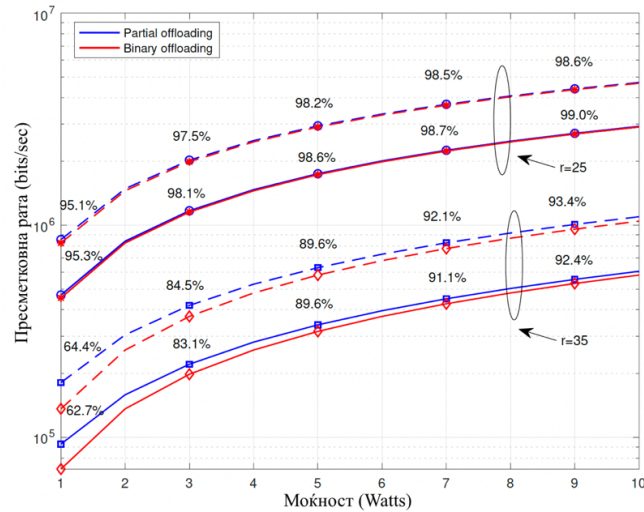
Слика 3.3 ја претставува вкупната пресметковна рата во системот, додека на Слика 3.3 е даден процентот на претоварени бити до МЕС-БС, во функција од растојанието r (0–60 m), за две вредности на пресметковното оптоварување, $L = 100$ CPU-циклуси/бит (полна линија) и $L = 1000$ CPU-циклуси/бит (испрекината линија). За помали вредности на r , оптималната стратегија е податоците да се растовараат на МЕС-серверот, бидејќи локалното пресметување троши повеќе енергија во споредба со енергијата која се троши за безжичен пренос на кратки растојанија (до околу 30 m).

**Слика 3.3:** Влијание на растојанието r на пресметковната рата (лево) и на бројот на претоварени бити (десно).

За мали вредности на r , пресметковното оптоварување L има само мал ефект врз вкупната пресметковна рата, бидејќи во овој регион се врши претовар на сите пресметковни задачи, па локалниот пресметковен интензитет станува ирелевантен. Меѓутоа, како растојанието r се зголемува, ефектот на пресметковниот интензитет станува сè позначаен – за $L = 100$ доцнежниот пораст доведува до значително повисоки пресметковни рати во споредба со $L = 1000$. Во случај на бинарен претовар, постои јасен праг на растојание над кој корисни-

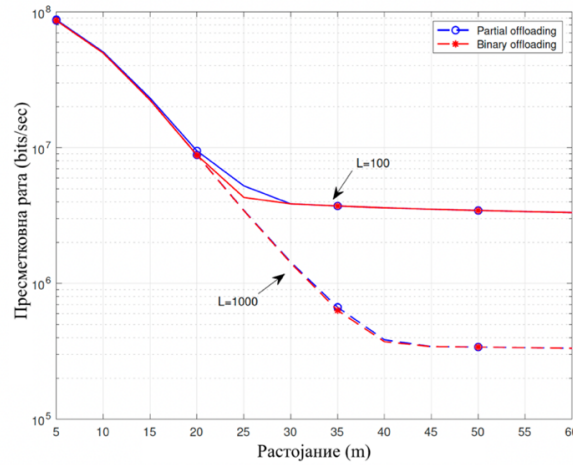
ците целосно се префрлаат на локално пресметување ($r \approx 30$ m за $L = 100$ и $r \approx 50$ m за $L = 1000$), видлив како остро „колено“ на кривите. Шемата за делумен претовар се однесува слично, но без нагол премин – процентот на претоварени битови постепено се намалува од 100% на 0%, а токму во овој транзициски интервал делумниот претовар ги надминува перформансите на бинарниот претовар.

Слика 3.4 ја прикажува зависноста на вкупната пресметковна рата од предавателната моќност на БС, P_0 , за различни радиуси ($r_1 = 25$ m и $r_1 = 35$ m) и различен број на ЕНУs ($K = 10$, полна линија, и $K = 20$, испрекината линија), при $L = 1000$. Бројот над секој маркер го означува процентот на растоварени битови кај стратегијата за делумен претовар. Зголемувањето на P_0 ја зголемува акумулираната енергија кај корисниците (согласно (3.6)), што директно ги зголемува и пресметковната рата и процентот на претоварени бити, со опаѓачки маргинални придобивки при повисоки P_0 .



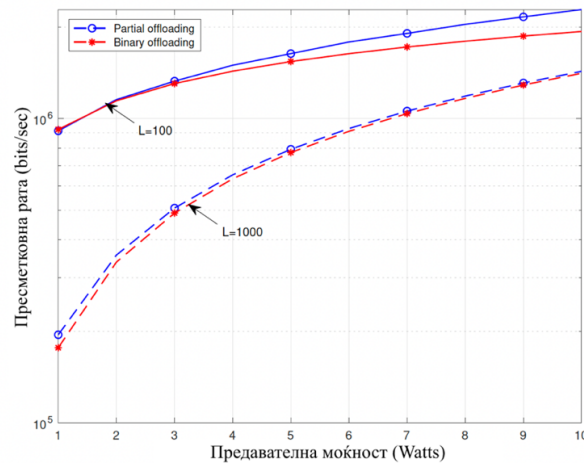
Слика 3.4: Влијание на моќноста на БС, P_0 , на бројот на претоварени бити.

За да го анализираме степенот на правичност, разгледуваме асиметрично сценарио: половина од корисничките станици ($K/2$) се на фиксно растојание $r_1 = 5$ m, додека другата половина се на растојание r_2 , кое варира. Слика 3.5 ја прикажува вкупната пресметковна рата како функција од r_2 (5–60 m). Вкупната пресметковна рата опаѓа со зголемување на r_2 – директна последица на макс-мин критериумот, кој ги приоритизира најслабите корисници (оние подалеку од БС), доделувајќи им повеќе ресурси, што го намалува вкупниот системски капацитет.



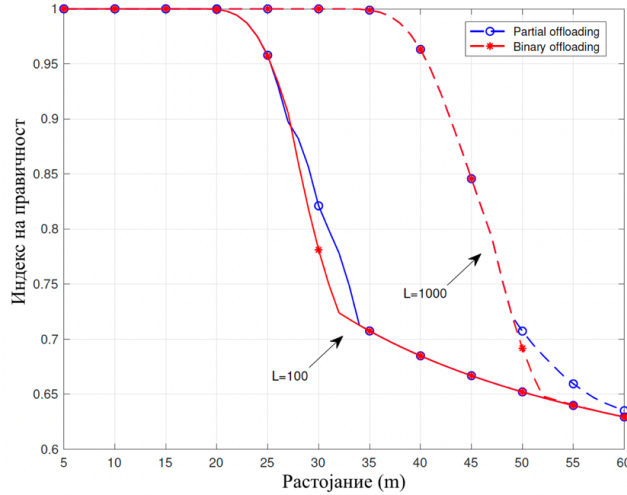
Слика 3.5: Влијание на растојанието r_2 на пресметковната рата.

Слика 3.6 ја прикажува вкупната пресметковна рата наспроти P_0 за фиксни растојанија $r_1 = 15$ m и $r_2 = 30$ m. Кога $L = 100$, шемата за делумен претовар има подобри перформанси од шемата за бинарен претовар за поголеми предавателни моќности, како резултат на способноста да се собере повеќе енергија.



Слика 3.6: Влијание на предавателната моќност на БС, P_0 , на пресметковната рата.

Слика 3.7 го илустрира индексот на правичност на Џејн во функција од растојанието r_2 , кога $r_1 = 15$ m. Индексот на правичност се намалува со зголемување на r_2 , бидејќи минималната пресметковна рата се постигнува кај корисниците на поголемо растојание, додека пресметковната рата на поблиските корисници е значително поголема. И двете шеми постигнуваат максимална вредност за индексот на правичност ($J = 1$) кога радиусот r_2 е помал од одредена гранична вредност, бидејќи тогаш сите корисници вршат само претовар на податоците.



Слика 3.7: Влијание на растојанието r_2 на индексот на правичност.

3.4 Оптимизација на ресурси според критериум на пропорционална правичност

Во продолжение ќе биде презентирана шема за распределба на ресурси за WP-MEC систем која се заснова на критериумот на пропорционална правичност [81]. Овој критериум се базира на компромис – се максимизира вкупниот проток во системот, додека истовремено им се дозволува на сите корисници барем минимално ниво на сервис. Тоа се постигнува преку воведување логаритамска функција на самата целна функција.

Системскиот модел кој го користиме одговара на системскиот модел презентирани во section 3.3. Во случајот на пропорционална правичност, оптимизацискиот проблем се сведува на максимизирање на збирот на логаритмите од пресметковните рати:

$$\text{Maximize}_{\tau_0, \tau_i, P_i, f_i} \sum_{i=1}^K \log \left(\frac{Tf_i}{L} + \frac{B}{T} \tau_i \log_2 \left(1 + \frac{P_i \Omega_i}{N_0} \right) \right) \quad (3.39)$$

$$\text{под услов: } C1 : \sum_{i=1}^K \tau_i + \tau_0 = T \quad (3.40)$$

$$C2 : \alpha T f_i^3 + \tau_i P_i \leq \eta \tau_0 P_0 \Omega_i, \forall i \quad (3.41)$$

$$C3 : \tau_i \geq 0, \tau_0 \geq 0, P_i \geq 0, f_i \geq 0, \forall i \quad (3.42)$$

Заради множење на двете оптимизациски променливи во C2 ($\tau_i P_i$), предложениот оптимизациски проблем не е конвексен. Сепак, со воведување на смената $e_i = P_i \tau_i$, оптимизацискиот проблем даден во (3.39) се трансформира во конвексен проблем. За да го презентираме аналитичкото решение, дефинираме помошна функција:

$$f(p_i^*) = \left(\frac{(p_i^* \Omega_i + N_0) \ln 2}{3B\Omega_i \alpha L} \right)^{1/2}, \quad (3.43)$$

каде p_i^* се наоѓа како корен на следнава трансцендентна равенка:

$$\frac{\eta P_0 \Omega_i^2 + p_i^* \Omega_i}{N_0 + p_i^* \Omega_i} = \log \left(1 + \frac{p_i^* \Omega_i}{N_0} \right). \quad (3.44)$$

Теорема 3.3. *Оптималната предавателна моќност на i -тиот корисник е дадена преку:*

$$P_i^* = \begin{cases} p_i^*, & \text{ако } \alpha T [f(p_i^*)]^3 \leq \eta P_0 \tau_0^* \Omega_i \\ 0, & \text{во спротивно} \end{cases}. \quad (3.45)$$

Оптималната CPU-брзина на i -тиот EHU е:

$$f_i^* = \begin{cases} f(p_i^*), & \text{ако } \alpha T [f(p_i^*)]^3 \leq \eta P_0 \tau_0^* \Omega_i \\ \left(\frac{\eta P_0 \Omega_i}{\alpha} \right)^{1/3}, & \text{во спротивно} \end{cases}. \quad (3.46)$$

Оптималното времетраење на фазата на претовар на i -тиот EHU е:

$$\tau_i^* = \begin{cases} \frac{\eta P_0 \tau_0^* \Omega_i - T [f(p_i^*)]^3}{p_i^*}, & \text{ако } \alpha T [f(p_i^*)]^3 \leq \eta P_0 \tau_0^* \Omega_i \\ 0, & \text{во спротивно} \end{cases}. \quad (3.47)$$

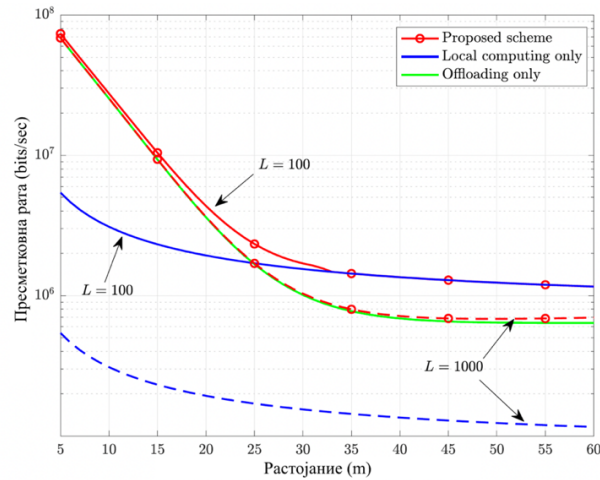
Оптималното времетраење на фазата на пренос/акумулирање на енергија е дадено со:

$$\tau_0 = T \frac{1 + \alpha \sum_{i \in \mathcal{D}} \frac{[f(p_i^*)]^3}{p_i^*}}{1 + \eta P_0 \sum_{i \in \mathcal{D}} \frac{\Omega_i}{p_i^*}}, \quad (3.48)$$

каде $\mathcal{D} = \{i \mid 1 \leq i \leq K, \alpha T [f(p_i^*)]^3 \leq \eta P_0 \tau_0^* \Omega_i\}$. Доказот на оптималното решение е аналоген на доказот на Теорема 3.1, прикажан во додаток Б.4.

Ова решение во затворена форма овозможува негова имплементација во онлајн алгоритам за определување на системските параметри во WP-MEC систем, под гаранција на пропорционална правичност.

Во продолжение, ги анализираме перформансите на предложената шема за распределба на ресурси базирана на критериумот за пропорционална правичност, споредувајќи ги со две референтни (анг. benchmark) шеми: (i) Шема 1 („само локално пресметување“) и (ii) Шема 2 („само претовар“). Перформансите ги оценуваме преку истите две метрики, вкупната пресметковна рата (R) и индексот на правичност на Џејн (J), дефинирани со (3.36) и (3.38). Системските параметри се идентични на оние дадени во Табела 3.1.

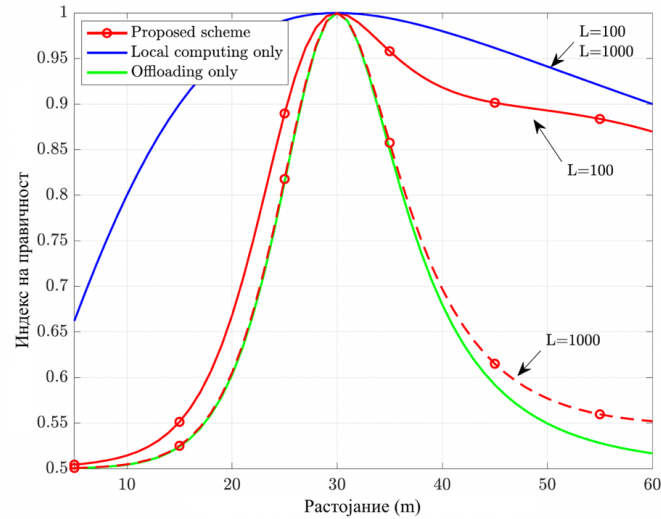


Слика 3.8: Влијание на растојанието r_2 на пресметковната рата.

Слика 3.8 го прикажува влијанието на растојанието на подалечната група корисници, r_2 , врз вкупната системска пресметковна рата, при фиксно растојание на поблиската група ($r_1 = 30$ m) и фиксна предавателна моќ на БС, $P_0 = 10$ W, за две вредности на пресметковното оптоварување ($L = 100$, полна линија, и $L = 1000$, испрекинатата линија). Кога $L = 100$, однесувањето на предложената шема може да се подели на три различни сегменти:

1. Регион на целосен претовар ($r_2 < 15$ m): предложената шема се совпаѓа со референтната шема 2.
2. Регион на флексибилен делумен претовар ($15 \text{ m} < r_2 < 35 \text{ m}$): вкупната пресметковна рата на предложената шема ги надминува двете референтни шеми, бидејќи системот има корист од флексибилноста да се избере оптималниот режим на претовар за секоја група корисници.
3. Регион на доминантно локално пресметување ($r_2 > 35$ m): предложената шема се совпаѓа со референтната шема 1.

Кога пресметковното оптоварување е многу високо ($L = 1000$), локалното процесирање станува исклучително енергетски неефикасно, па предложената шема се совпаѓа со кривите на референтната шема 1 речиси во целиот опсег на вредности на r_2 .



Слика 3.9: Влијание на растојанието r_2 на индексот на правичност.

Слика 3.9 го илустрира индексот на правичност на системот наспроти растојанието r_2 , при истите системски поставки. Индексот на правичност се зголемува од 0.5 до 1 (совршена праведност) како што r_2 се приближува до r_1 ($r_1 = r_2 = 30$ m одговара на случајот кога сите корисници се поставени на иста кружница). За $r_2 > r_1$, индексот на правичност опаѓа со зголемување на r_2 , бидејќи разликата во перформансите помеѓу двете групи корисници расте.

Високото ниво на правичност кога корисниците вршат само локална пресметка се должи на единечното слабеење на патеката на РФ сигналот наменет за акумулирање енергија (т.е. „конвенционалниот“ ефект на близу-далеку). Од друга страна, претоварените податоци се изложени и на слабеење на патеката за време на процесот на собирање енергија, и на слабеење на патеката за време на процесот на претовар (т.е. „двоен“ ефект на близу-далеку), поради што претоварот е инхерентно помалку правичен од локалното процесирање. Затоа, како што системот се префрла кон локално пресметување (при поголеми r_2 и помали L), нивото на правичност се подобрува.

Глава 4

Безжично федеративно учење

Откако ја воспоставивме инфраструктурата и методологијата за ефикасна пресметка на „работ на мрежата“ и претовар на пресметковни задачи, се поставува прашањето: кои се напредните апликации што оваа инфраструктура може да ги овозможи? Одговорот лежи во „интелигенција на работ“, нов принцип кој се наметнува како клучно решение во безжичните комуникации од следната генерација. Огромните количини на податоци што се генерираат на работ на мрежата од страна на милијарди IoT-уреди претставуваат непроценлив ресурс за тренирање на софистицирани модели на машинско учење (МУ).

Меѓутоа, примената на традиционалните методи на МУ на „работ“ на мрежата наидува на фундаментален парадокс. Доколку сите податоци од корисничките уреди се испратат до централен сервер за тренирање, повторно се соочуваме со истите проблеми што МЕС има за цел да ги реши: високо комуникациско оптоварување, зголемена латентност и, најважно, сериозни ризици по приватноста и безбедноста на податоците. Токму со овие предизвици се соочува федеративното учење, како еден засебен вид на дистрибуирано машинско учење. Федеративното учење, првично предложено од Google [82, 83], претставува револуционерен пристап кон дистрибуирано машинско учење кој ја става приватноста на податоците на прво место. Наместо да се централизираат податоците кои ги генерираат корисниците, моделот за учење се носи до самите корисниците. Процесот на ФУ е итеративен и се состои од следниве чекори:

1. **Дистрибуција на моделот:** централниот сервер (агрегатор) го дистрибуира тековниот глобален модел до избрана група на кориснички уреди.
2. **Локално тренирање:** Секој уред го тренира добиениот модел користејќи ги сопствените, локални и приватни податоци, што резултира со локално ажурирани параметри на моделот.
3. **Комуникација на параметрите:** уредите ги испраќаат до централниот сервер само ажурираните параметри на моделот, а не и необработените податоци. Ова е клучниот чекор што ја гарантира приватноста.
4. **Агрегација и ажурирање:** централниот сервер ги агрегира (на пр., ги усреднува) параметрите добиени од сите уреди за да формира нов, подобрен глобален модел.

Овој циклус се повторува во повеќе рунди, при што глобалниот модел постепено се подобрува и конвергира кон оптимално решение, без притоа да се компромитира приватноста

на корисничките податоци. Иако ФУ е елегантно решение за проблемот со приватноста, неговата примена во безжични мрежи со голем број корисници открива нов критичен предизвик: комуникациското тесно грло. Итеративната природа на ФУ, која бара повторливи комуникациски рунди, создава значителен товар врз безжичната мрежа, особено во uplink насоката. Овој товар главно се манифестира на три параметри: (i) **Доцнење** – вкупното време за тренирање на моделот не зависи само од брзината на локалните пресметки, туку и од времето потребно за сите уреди да ги пренесат своите ажурирани параметри; (ii) **Потрошувачка на енергија** – за IoT-уредите кои работат на батерија или се напојуваат преку WPT, енергијата потрошена за безжичен пренос на параметрите на моделот е значаен трошок; и (iii) **Пропусен опсег** – истовремениот пренос на параметри на модели од голем број уреди може да доведе до загушување на мрежата.

4.1 Концепт на федеративно учење

Процесот на машинско учење (МУ) се заснова на способноста на компјутерите да тренираат модели од податочно множество со цел да прават предвидувања и да извлекуваат заклучоци. Во овој процес, централниот проблем е да се пронајде оптималниот сет на параметри на модел, w^* , кој ја минимизира функцијата на загуба $f(h(\mathbf{x}, w), y)$, дефинирана за дадена хипотетичка функција $h(\cdot)$ и податочно множество $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$. Ова се формулира како оптимизациски проблем:

$$\arg \min_{\mathbf{w}} \frac{1}{N} \sum_{i=1}^N f(h(\mathbf{x}_i, \mathbf{w}), y_i). \quad (4.1)$$

За решавање на овој проблем, **методот на спуштање по градиент** (анг. Gradient Descent, GD) е еден од најупотребуваните итеративни алгоритми. GD ги ажурира параметрите на моделот $\mathbf{w}_k$ во секоја итерација k во насока спротивна на градиентот на функцијата на загуба:

$$\mathbf{w}_k = \mathbf{w}_{k-1} - \alpha \nabla_{\mathbf{w}} f(\mathbf{w}_{k-1}), \quad (4.2)$$

каде α е чекорот на учење. За големи податочни множества, често се користи стохастичко спуштање по градиент (анг. Stochastic Gradient Descent, SGD), каде градиентот се пресметува врз основа на мало подмножество од податоците. Квалитетот на тренираниот модел може да се изрази преку релативната грешка при локалното тренирање, η , која ја мери близината на локално тренираниот модел до неговата оптимална вредност.

Имајќи ја предвид дистрибуираната природа на безжичните системи, дистрибуираното машинско учење е од клучно значење во различни безжични мрежни околина [84]. Во рамките на дистрибуирано машинско учење, M уреди во свездеста мрежна топологија се поврзани со централен сервер преку безжичен комуникациски канал. Секој уред $j \in \{1, 2, \dots, M\}$ чува локално подмножество на податоци $\mathcal{D}_j := \{(\mathbf{x}_i^j, y_i^j)\}_{i=1}^{N_j}$, каде $\mathcal{D} = \cup_{j=1}^M \mathcal{D}_j$ ги означува сите податоци достапни во системот. Следствено, на секој уред j соодветствува локална

функција на загуба дефинирана како:

$$f^j(\mathbf{w}) := \frac{1}{N_j} \sum_{i=1}^{N_j} f\left(h\left(\mathbf{x}_i^j, \mathbf{w}\right), y_i^j\right), \quad j = 1, \dots, M. \quad (4.3)$$

Целта на алгоритмот за дистрибуирано учење е уредите колаборативно да го минимизираат конечниот збир на нивните локални функции на загуба и да го добијат оптималното решение $\mathbf{w}^*$, дефинирано како:

$$\mathbf{w}^* \in \arg \min_{\mathbf{w}} \frac{1}{M} \sum_{j=1}^M f^j(\mathbf{w}). \quad (4.4)$$

Да претпоставиме дека за минимизирање на функцијата на трошок се користи GD-алгоритмот, каде M уреди се поврзани со централен сервер во ѕвезда топологија преку безжичен комуникациски канал, како што е прикажано на Слика 4.1. Во секоја итерација $k = 1, \dots, K$ на GD, серверот го испраќа глобалниот параметар $\mathbf{w}_{k-1}$ до сите уреди. Потоа, секој уред j го пресметува својот локален градиент:

$$\nabla_{\mathbf{w}} f^j(\mathbf{w}_{k-1}) = \frac{1}{N_j} \sum_{i=1}^{N_j} \nabla_{\mathbf{w}} f\left(h\left(\mathbf{x}_i^j, \mathbf{w}_{k-1}\right), y_i^j\right), \quad j = 1, \dots, M, \quad (4.5)$$

и го испраќа до централниот сервер преку безжичниот канал. Кога сите уреди ќе го завршат преносот на своите локални градиенти, серверот го ажурира глобалниот параметар:

$$\mathbf{w}_k = \mathbf{w}_{k-1} - \frac{\alpha}{M} \sum_{j=1}^M \nabla_{\mathbf{w}} f^j(\mathbf{w}_{k-1}), \quad k = 1, \dots, K, \quad (4.6)$$

и го испраќа до сите корисници. Овој итеративен процес продолжува сè додека не се постигне претходно дефиниран праг на конвергенција ϵ :

$$\left\| \frac{1}{M} \sum_{j=1}^M \nabla_{\mathbf{w}} f^j(\mathbf{w}^*) \right\|^2 \leq \epsilon. \quad (4.7)$$



Слика 4.1: Шема за дистрибуирано машинско учење со користење на GD.

Во овој контекст, конвергенцијата на моделот е одредена од бројот на локални итерации (епохи) на секој уред (N_l) и бројот на глобални рунди на тренирање (N_g). Квалитетот на локално тренираниот модел по N_l локални епохи се изразува преку релативната грешка η . За дадена η , бројот на локални епохи е $N_l(\eta) = v_1 \log(1/\eta)$. Слично, бројот на глобални рунди за да се постигне целната глобална точност ϵ_0 е $N_g(\epsilon_0, \eta) = v_2(\epsilon_0)/(1 - \eta)$. Параметрите v_1 и v_2 ги опишуваат својствата на конвергенција на конкретниот FL-модел. Овие метрики ќе бидат клучни за оптимизација на системското доцнење во следните делови, овозможувајќи динамично балансирање меѓу локалното тренирање и глобалната комуникација.

Современиот развој на безжичните комуникациски системи, особено во контекст на петтата и шестата генерација мрежи, наметнува интеграција на децентрализирани интелигентни решенија. Како што истакнуваат авторите во [85], овој пристап не само што ја задржува приватноста на податоците, туку и значително го намалува сообраќајниот товар врз безжичната мрежа. Сепак, имплементацијата на ФУ во реални безжични средини се соочува со сериозни предизвици како што се несигурноста на безжичните канали, ограничувањата на пропусниот опсег, како и ограничувања на енергетските и пресметковни капацитети на мобилните уреди.

Еден од клучните аспекти разгледувани во литературата е минимизирањето на вкупното доцнење, односно вкупното времетраење на комуникациските рунди. Во [86], авторите теоретски ја покажуваат зависноста помеѓу вкупното доцнење во системите за ФУ и точноста на моделот на учење, нагласувајќи го критичниот компромис помеѓу локалното време за пресметување на уредот и времето потребно за безжичен пренос на параметрите. Слично на ова, во [87] е предложена сеопфатна рамка која истовремено ги оптимизира предавателната моќност на уредите, изборот на корисници и распределбата на ресурсните блокови со цел да се минимизира функцијата на загуби на моделот за ФУ.

За дополнително подобрување на ефикасноста на спектарот и поддршка на масивен број уреди, NOMA станува доминантна тема во најновите истражувања. Во трудовите [88, 89] авторите воведуваат т.н. „пресметај-па-пренеси“ протокол (анг. Compute-then-Transmit NOMA) кој овозможува симултан пренос на локалните модели преку исти ресурсни блокови, користејќи сукцесивно поништување на интерференцијата. Ова овозможува драматично намалување на доцнењето во споредба со традиционалните ортогонални шеми како TDMA. Надополнувајќи го ова, авторите во [90] го проучуваат федеративното учење во NOMA-базирани МЕС-системи и користат теорија на графови за решавање на проблемите со распредување на корисниците.

Интелигентното управување со корисниците и конвергенцијата на моделите исто така се адресираат преку напредни техники за проценка на локалните ажурирања. Авторите во [91], во нивниот труд во кој вршат минимизирање на времето на конвергенција, предлагаат користење на вештачки невронски мрежи на самата базна станица за предвидување на параметрите на оние корисници кои не се во можност активно да учествуваат во преносот, со што се намалува вкупното време на конвергенција за над 50%. Ваквиот пристап е во целосна согласност со визијата на авторите на [92] за 6G, како мрежа на „поврзана интелигенција“.

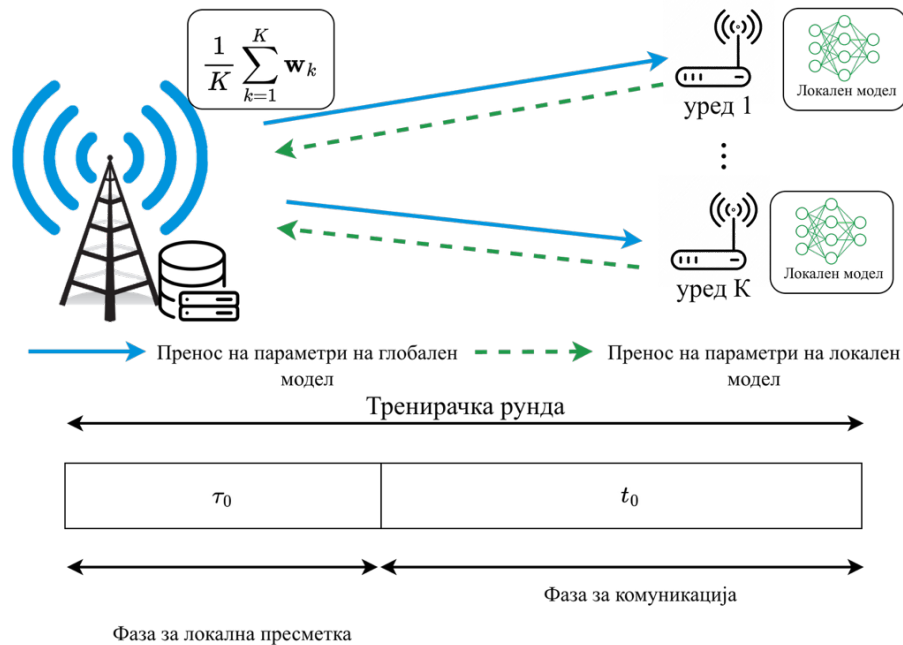
Конечно, сигурноста и приватноста во безжичните системи за ФУ добиваат нова димен-

зија преку концептот на диференцијална приватност. Liu и Simeone во својот труд [93] ја претставуваат иновативната теза за *privacy for free*, каде докажуваат дека природниот шум во безжичните канали може директно да се искористи како механизам за маскирање на податоците при не кодиран пренос на градиенти на локалните модели.

4.2 Оптимизирање на ресурси со неортогонален повеќекратен пристап

Разгледуваме безжичен комуникациски систем за федеративно учење кој се состои од една базна станица и K безжични уреди (анг. wireless devices) (WDs), како што е прикажано на Слика 4.2. БС има централен сервер кој го одржува глобалниот модел за учење, додека секој безжичен уред го одржува својот локален модел за учење базиран на негово сопствено податочно множество. БС е опремена со една предавателна/приемна антена, а секој безжичен уред исто така е опремен со по една антена. Базната станица ги агрегира тежинските коефициенти на локалните модели од WDs и применува стратегија на федеративно усреднување за ажурирање на глобалниот модел [94].

Процесот на тренирање на глобалниот и локалните модели се извршува во многу тренинг рунди (TR), кои се состојат од фаза за локална пресметка и фаза за комуникација. Претпоставуваме дека k -от безжичен уред има локално множество на податоци со големина D_k примероци. Во продолжение детално ќе ги разгледаме овие фази.



Слика 4.2: Системски модел на NOMA-базиран безжичен комуникациски систем за федеративно учење.

Во фазата за локална пресметка, k -от безжичен уред го тренира својот локален модел низ I_k локални епохи. Нека фреквенцијата на процесорот на k -от безжичен уред е f_k

(CPU-циклуси/секунда), а потребниот пресметковен интензитет по примерок е L_k (CPU-циклуси/примерок). Тогаш $a_k = I_k L_k D_k$ е вкупниот број на циклуси потребни за k -от безжичен уред да ја изврши пресметката на локалниот модел во рамки на една ТР. Претпоставуваме дека $a_k = a_0, \forall k$, а времетраењето на фазата на локална пресметка за k -от безжичен уред е $\tau_k = a_0 / f_k$. Количеството енергија што се троши за таа фаза е $E_k^{LC} = \alpha a_0 f_k^2$. Бидејќи сите безжични уреди го започнуваат преносот истовремено, времетраењето на фазата на локална пресметка τ_0 се пресметува како:

$$\tau_0 = \max_{1 \leq k \leq K} \left\{ \frac{a_0}{f_k} \right\}. \quad (4.8)$$

Фазата за комуникација се реализира со истовремен пренос од страна на сите уреди преку користење на NOMA. Ако безжичните уреди се индексирани во растечки редослед на нивните канални засилувања, $|h_1| \leq |h_2| \leq \dots \leq |h_K|$, БС ја декодира информацијата користејќи ја методата на sukcesивно поништување на интерференцијата, во опаѓачки редослед на индексите. Постигнатата информациска брзина на k -от безжичен уред е дадена со:

$$R_k = B \log_2 \left(\frac{|h_k|^2 p_k}{\sum_{j=1}^{k-1} |h_j|^2 p_j + N_0} \right), \quad (4.9)$$

каде B е комуникацискиот опсег на uplink каналот, N_0 е термалниот шум во приемникот на БС и p_k е предавателната моќност на k -от безжичен уред. Претпоставуваме дека големината на пакетот кој ги пренесува тежинските коефициенти на локалниот модел е иста за сите корисници, $b_k = b_0, \forall k$, и бидејќи користиме NOMA, наметнуваме $t_k = t_0, \forall k$. Времетраењето на фазата за комуникација се пресметува според [95]:

$$t_0 = \frac{kb_0}{B \log_2 \left(1 + \sum_{i=1}^k x_i p_i \right)}, \quad 1 \leq k \leq K, \quad (4.10)$$

каде $x_i = |h_i|^2 / N_0$.

Дефинираме оптимизациски проблем со цел да се минимизира вкупното доцнење, T_0 , во тек на една ТР, дефинирано како збир на доцнењето во фазата на локална пресметка (4.8) и доцнењето во фазата на комуникација (4.10):

$$\text{Minimize}_{t_0, p_k, f_k} \quad t_0 + \max_{1 \leq k \leq K} \left\{ \frac{a_0}{f_k} \right\} \quad (4.11)$$

$$\text{под услов:} \quad C1 : \frac{t_0}{k} B \log_2 \left(1 + \sum_{i=1}^k x_i p_i \right) = b_0, \forall k \quad (4.12)$$

$$C2 : p_k \leq P_{\max}, \forall k \quad (4.13)$$

$$C3 : \alpha a_0 f_k^2 + p_k t_0 \leq E_{\max}, \forall k \quad (4.14)$$

$$C4 : f_k \leq f_{\max}, \forall k \quad (4.15)$$

каде C1 се однесува на (4.10), C2 се однесува на максималната предавателна моќност $P_{\max}$, C3 на максималната расположлива енергија $E_{\max}$, а C4 на максималната фреквенција на

процесорот $f_{\max}$. Оптимизацискиот проблем може да се трансформира во конвексен (трансформацијата е претставена во додаток Б.5), а аналитичкото решение е дадено во следната теорема.

Теорема 4.1. *Оптималната вредност на предавателната моќност на k -от безжичен уред е дадена со:*

$$p_k^* = \frac{e^{kz^*} - e^{(k-1)z^*}}{x_k}, \quad (4.16)$$

каде z^* се пресметува според:

$$z^* = \min \{z_{01}, z_{02}, \dots, z_{0K}, g_{01}, g_{02}, \dots, g_{0K}\}, \quad (4.17)$$

каде z_{0k} се определува како решение на следнава трансцендентна равенка:

$$e^{kz_{0k}} - e^{(k-1)z_{0k}} = P_{\max} x_k, \quad (4.18)$$

додека g_{0k} се определува заедно со фреквенцијата на процесорот f_{0k} , како решение на следниот систем трансцендентни равенки:

$$1 - g_{0k}(k-1) - e^{g_{0k}} + kg_{0k}e^{g_{0k}} = 2\alpha a_0(f_{0k})^3 x_k e^{-g_{0k}(k-1)}, \quad (4.19)$$

$$\frac{e^{kg_{0k}} - e^{(k-1)g_{0k}}}{g_{0k}} = \frac{x_k (E_{\max} - \alpha a_0(f_{0k})^2)}{c_0}. \quad (4.20)$$

Оптималната фреквенција на процесорот на k -тиот безжичен уред е дадена со:

$$f_k^* = \begin{cases} f_{0k}, & f_{0k} < f_{\max} \\ f_{\max}, & f_{0k} \geq f_{\max} \end{cases}. \quad (4.21)$$

Минималното системско доцнење, односно минималното времетраење на TP , е дадено со:

$$T_0^* = \frac{c_0}{z^*} + \max \left\{ \frac{a_0}{f_k^*} \right\}, \quad (4.22)$$

каде $c_0 = b_0 \ln(2)/B$. Доказот е даден во додаток Б.5.

За решавање на конвексни оптимизациски проблеми како во додаток Б.5, обично се користи методот на внатрешна точка¹. Овој метод ги решава конвексните оптимизациски проблеми во t_1 итерационски чекори, каде $10 < t_1 < 100$ [96]. Секоја итерација бара $\max\{S_1^3, S_1^2 S_2\}$ операции, каде S_1 е вкупниот број на променливи, а S_2 е вкупниот број на ограничувања. Бидејќи за предложениот проблем $S_1 = 2K + 2$ и $S_2 = 4K$, комплексноста за решавање на проблемот е $\mathcal{O}(t_1 K^3)$. Од друга страна, системот од трансцендентни равенки (4.19)–(4.20)

¹Методите на внатрешна точка (анг. Interior-Point Methods) претставуваат супериорна класа на итеративни алгоритми наменети за решавање на проблеми на линеарно и нелинеарно конвексно програмирање. За разлика од конвенционалниот Симплекс метод, кој го пребарува оптималното решение движејќи се по рабовите (темињата) на полиедарот на дефиниционата област, методите на внатрешна точка ја пребаруваат оптималноста движејќи се директно низ внатрешноста на дефиниционата област.

обично се решава со користење на Њутновиот метод [97], во t_2 чекори од ред на величина десет, со комплексност $\mathcal{O}(t_2 K)$. Споредено со директното нумеричко решение, Теорема 4.1 нуди многу поефикасно решение, а оваа предност станува сè поизразена со зголемување на бројот на корисници K .

Табела 4.1: Системски параметри за безжичен комуникациски систем за федеративно машинско учење базиран на не-ортогонален повеќекратен пристап.

Параметар	Симбол	Вредност
Број на безжични уреди	K	20
Растојание на уредите од БС	d_k	$50k$ m (за $1 \leq k \leq 20$)
Големина на локално множество податоци	D_k	1000 примероци
Пресметковен интензитет по примерок	L_k	1000 циклуси/примерок
Број на локални тренинг итерации	I_k	100
Вкупен број на CPU-циклуси по уред	a_k (a_0)	10^8
Коефициент на пресметковна ефикасност	α	10^{-28}
Моќност на термален шум во БС	N_0	10^{-14} W
Фреквенциски опсег на каналот	B	10 MHz
Експонент на слабеење	n	3
Модел на слабеење	Ω_k	$10^{-3} d_k^{-3}$

За споредба на перформансите на предложената шема, воведовме референтна шема базирана на TDMA. Распределбата на ресурси во референтната TDMA-шема се добива како решение на оптимизациски проблем соодветен на (4.11), но со целна функција која одговара на TDMA-протоколот:

$$\underset{t_k, p_k, f_k}{\text{Minimize}} \quad \sum_{k=1}^K t_k + \max_{1 \leq k \leq K} \left\{ \frac{a_0}{f_k} \right\} \quad (4.23)$$

$$\text{под услов: } C1 : t_k B \log_2(1 + x_k p_k) = b_0, \forall k \quad (4.24)$$

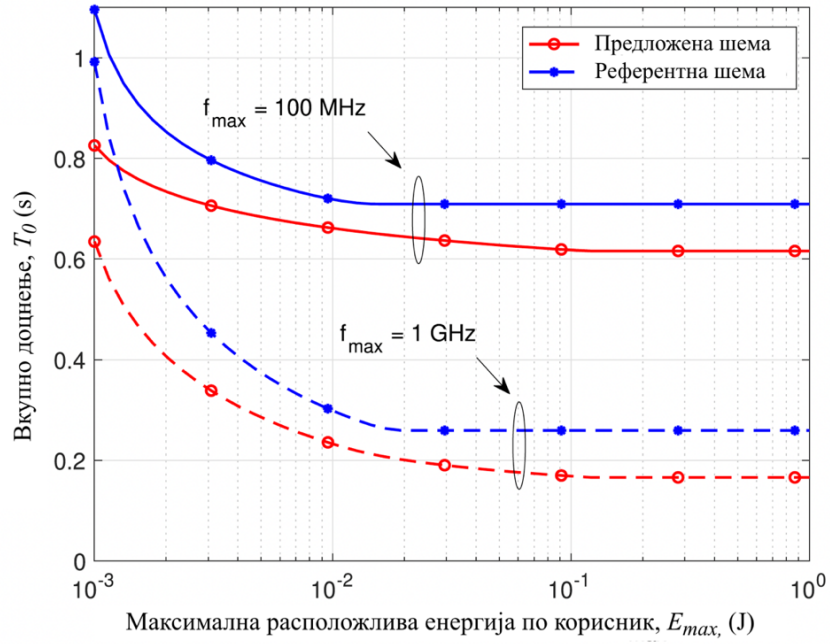
$$C2 : p_k \leq P_{\max}, \forall k \quad (4.25)$$

$$C3 : \alpha a_0 f_k^2 + p_k t_k \leq E_{\max}, \forall k \quad (4.26)$$

$$C4 : f_k \leq f_{\max}, \forall k \quad (4.27)$$

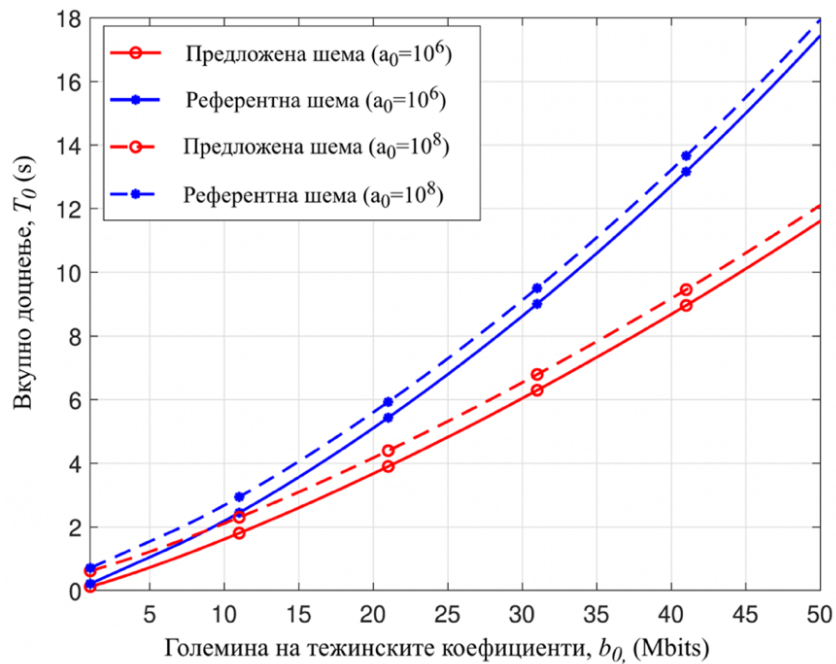
Времетраењето на фазата за локална пресметка, τ_0 , останува непроменето, додека фазата на комуникација се состои од K последователни трансмисии со времетраење t_k .

Слика 4.3 ја претставува зависноста на вкупното системско доцнење T_0 од максималната расположлива енергија $E_{\max}$, за дадено $P_{\max} = 1$ W и $b_0 = 1$ Mbit, за две вредности на $f_{\max}$ ($f_{\max} = 100$ MHz, полна линија, и $f_{\max} = 1$ GHz, испрекинатата линија). Со зголемување на $E_{\max}$, секој безжичен уред може да испраќа со поголема моќност и/или со поголема CPU-фреквенција, но зголемувањето над одреден праг не придонесува понатаму, бидејќи уредите се лимитирани од $P_{\max}$ и $f_{\max}$. Сатурацијата кај предложената NOMA-базирана шема се јавува за поголеми вредности на $E_{\max}$ во споредба со референтната TDMA-шема.



Слика 4.3: Влијание на максималната расположлива енергија $E_{\max}$ на вкупното доцнење T_0 .

Слика 4.4 го прикажува вкупното системско доцнење T_0 во функција од големината на тежинските коефициенти b_0 , за $a_0 = 10^6$ и $a_0 = 10^8$. Доцнењето се зголемува со зголемување на b_0 , но NOMA-базираната предложена шема постигнува помало доцнење споредено со TDMA-базираната референтна шема, што значи дека предложената шема би постигнала поголема точност при тренирање на локалниот модел за фиксно вкупно време [98].



Слика 4.4: Влијание на тежинските коефициенти b_0 на вкупното доцнење T_0 .

4.3 Оптимизирање на ресурси со временски повеќекратен пристап

За да се постигне визијата на идните 6G мрежи, каде безжичните уреди ќе бидат не само интелигентни, туку и енергетски самоодржливи, неопходна е интеграција со технологијата за безжичен пренос на енергија. Оваа синергија помеѓу ФУ и WPT овозможува креирање на системи каде уредите можат континуирано да учествуваат во колаборативно учење без потреба од замена на батерии. Сепак, оваа интеграција го усложнува проблемот на распределба на ресурси – сега, покрај комуникациските и пресметковните ресурси, мора да се управува и со трет, клучен ресурс: времето за собирање енергија.

Нашата централна хипотеза е дека вистинска минимизација на вкупното време за тренирање може да се постигне само преку заедничка, повеќеслојна оптимизација која ги опфаќа сите клучни параметри од три различни домени: (1) комуникациски ресурси (времетраењето на фазите за собирање енергија и пренос на податоци, како и предавателните моќности); (2) пресметковни ресурси (фреквенциите на процесорите); и (3) параметри на учењето (прагот на грешка при локалното тренирање η).

Разгледуваме систем за безжично напојувано федеративно учење кој се состои од една базна станица (БС) и K уреди кои акумулираат РФ енергија (EHUs). БС функционира и како извор на енергија за безжично напојување, и како параметарски сервер за агрегација на ФУ. Секој од K -те EHUs поседува сопствено, приватно податочно множество D_k и е опремен со електрично коло за акумулирање РФ енергија.

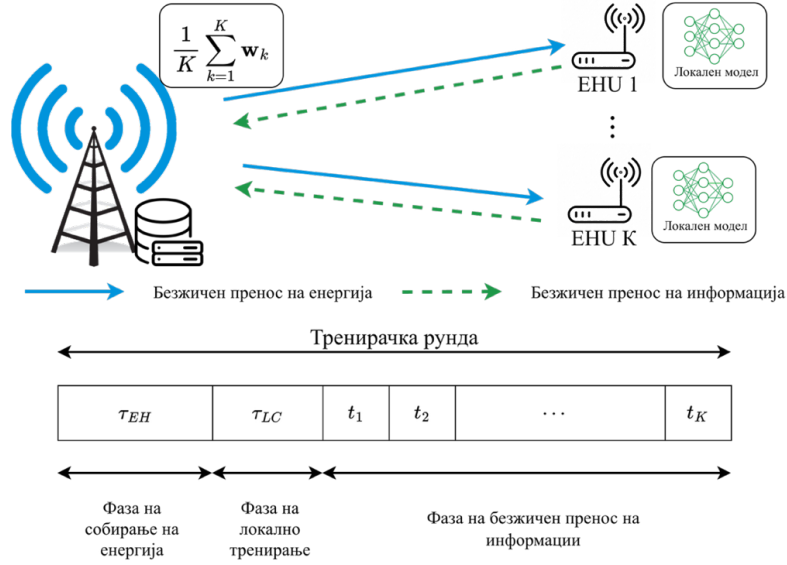
Безжичниот канал помеѓу БС и k -от EHU се претпоставува дека е реципрочен и се карактеризира со квази-статично рамно слабеење, така што засилувањето на каналот, h_k , останува константно во текот на една рунда на тренирање. Како што е илустрирано на Слика 4.5, секоја рунда на тренирање е структурирана во три последователни, непреклопувачки фази, кои се реализираат преку TDMA-протокол: фаза на собирање енергија, фаза на локално тренирање и фаза на пренос на информации.

Параметарот η ($0 \leq \eta \leq 1$) е релативната грешка при локалното тренирање која секој EHU треба да ја исполни, согласно изразот $l_k(\mathbf{w}_k^{(n)}) - l_k(\mathbf{w}_k^*) \leq \eta(l_k(\mathbf{w}_k^{(0)}) - l_k(\mathbf{w}_k^*))$, каде $\mathbf{w}_k^{(n)}$ ги означува параметрите на локалниот модел во n -та локална итерација, а $\mathbf{w}_k^*$ е оптималната вредност. По фазата на локално тренирање, БС ги усреднува локалните параметри, $\mathbf{w} = \frac{1}{K} \sum_{k=1}^K \mathbf{w}_k$ [9], и постапката се повторува во повеќе тренирачки рунди сè додека глобалниот модел не конвергира до саканата точност $1 - \epsilon_0$.

Првата фаза во секоја рунда има времетраење τ_{EH} . Во текот на оваа фаза, БС зрачи РФ сигнал со фиксна преносна моќност P_0 . Вкупната количина на енергија собрана и складирана од страна на k -от EHU во текот на оваа фаза е дадена со:

$$E_k^{EH} = \zeta P_0 \tau_{EH} h_k, \quad (4.28)$$

каде $\zeta \in (0, 1]$ е ефикасноста на конверзија на РФ во DC енергија.



Слика 4.5: Системски модел и структура на TDMA-рамка за една рунда на тренирање.

Бројот на локални епохи, N_l , потребен за секој ЕНУ да ја постигне саканата вредност на η се одредува со [98, 99]:

$$N_l(\eta) = v_1 \log(1/\eta), \quad (4.29)$$

каде v_1 е позитивна константа која зависи од својствата на локалната функција на загуба. Нека $a_k(\eta) = L_k D_k N_l(\eta)$ го означува вкупниот број на CPU-циклуси потребни за k -от ЕНУ. Претпоставуваме дека $L_k = L_0$ и $D_k = D_0$, па $a_k(\eta) = a_0(\eta) = L_0 D_0 v_1 \log(1/\eta)$, $\forall k$. Времетраењето на фазата на локално тренирање е:

$$\tau_{LC} = \max_{1 \leq k \leq K} \left\{ \frac{a_0(\eta)}{f_k} \right\}. \quad (4.30)$$

Енергијата потребна за k -от ЕНУ за таа фаза е:

$$E_k^{LC} = \alpha a_0(\eta) f_k^2, \quad (4.31)$$

каде α е коефициентот на енергетска ефикасност на процесорот на уредот [31].

По фазата на локално тренирање, k -от ЕНУ ги пренесува своите параметри до БС во времетраење t_k . Нека b_0 ја означува големината на параметрите на локалниот модел, еднаква за сите ЕНУs. Капацитетот на каналот мора да го задоволи условот:

$$B \log_2 \left(1 + \frac{p_k h_k}{B N_0} \right) \geq \frac{b_0}{t_k}, \quad \forall k, \quad (4.32)$$

каде p_k е преносната моќност на k -от ЕНУ. Енергијата потрошена за пренос е:

$$E_k^{WC} = p_k t_k. \quad (4.33)$$

Бројот на глобални рунди, N_g , потребен за да се постигнат саканите вредности на ϵ_0 и η ,

се одредува со [29, 99]:

$$N_g(\epsilon_0, \eta) = \frac{v_2(\epsilon_0)}{1 - \eta}, \quad (4.34)$$

каде $v_2(\epsilon_0)$ е позитивна константа која зависи од вредноста на ϵ_0 и својствата на функциите на загуба.

Целта на истражувањето во оваа подглава е да се минимизира вкупното времетраење на тренирање на глобалниот модел за ФУ, T , дефинирано како производ на N_g и времетраењето на една поединечна рунда, $T_r = \tau_{EH} + \tau_{LC} + \sum_{k=1}^K t_k$:

$$T = \frac{v_2}{1 - \eta} \left(\tau_{EH} + \tau_{LC} + \sum_{k=1}^K t_k \right). \quad (4.35)$$

Целта е минимизација на T преку заедничка оптимизација на η , τ_{EH} , τ_{LC} , t_k , f_k , и p_k :

$$\begin{array}{ll} \text{Minimize} & T \\ \eta, \tau_{EH}, \tau_{LC}, f_k, t_k, p_k & \end{array} \quad (4.36)$$

$$\text{под услов: } C1 : \alpha a_0(\eta) f_k^2 + p_k t_k \leq \xi P_0 \tau_{EH} h_k, \forall k \quad (4.37)$$

$$C2 : t_k B \log_2 \left(1 + \frac{p_k h_k}{B N_0} \right) \geq b_0, \forall k \quad (4.38)$$

$$C3 : f_k \leq f_{\max}, \forall k \quad (4.39)$$

$$C4 : \frac{a_0(\eta)}{f_k} \leq \tau_{LC}, \forall k \quad (4.40)$$

$$C5 : 0 \leq \eta \leq 1 \quad (4.41)$$

каде C1 е ограничувањето на енергетската каузалност, C2 го гарантира сигурниот пренос, C3 ја наметнува максималната CPU-фреквенција, C4 ја обезбедува синхронизацијата на завршувањето на локалното тренирање, а C5 ја дефинира валидната област на η .

Оптимизацискиот проблем не е конвексен, поради производните членови во целната функција и ограничувањата, како и сложената зависност од η . За да се справиме со ова, применуваме стратегија на декомпозиција и алтернативна оптимизација [99], делејќи го проблемот на два подпроблеми: (1) распределба на комуникациски и пресметковни ресурси за фиксно η , и (2) оптимизација на параметарот на учење η , користејќи ги оптималните вредности од првиот подпроблем.

За фиксно η , членот $v_2/(1-\eta)$ станува константа, па минимизацијата на T е еквивалентна на минимизација на времетраењето на една рунда. Со воведување на заменската променлива

$e_k = p_k t_k$, оптимизациониот проблем (4.36) за фиксно η се трансформира во:

$$T_{\min}(\eta) = \underset{\tau_{EH}, \tau_{LC}, f_k, t_k, e_k}{\text{Minimize}} \quad T \quad (4.42)$$

$$\text{под услов: } C1' : \alpha a_0(\eta) f_k^2 + e_k \leq \xi P_0 \tau_{EH} h_k, \forall k \quad (4.43)$$

$$C2' : t_k \log \left(1 + \frac{e_k h_k}{t_k B N_0} \right) \geq c_0, \forall k \quad (4.44)$$

$$C3' : f_k \leq f_{\max}, \forall k \quad (4.45)$$

$$C4' : \frac{a_0(\eta)}{f_k} \leq \tau_{LC}, \forall k \quad (4.46)$$

каде $c_0 = b_0 \log(2)/B$. За дадено, фиксно η , оптимизациониот проблем (4.42) е конвексен и неговото решение е дадено со следната теорема.

Теорема 4.2. *Оптималната предавателна моќност на k -от EHU, p_k^* , се добива како:*

$$p_k^* = \frac{B N_0}{h_k} \left[-1 - g_1(\tau_{EH}^*, f_0^*) W_{-1} \left(-\frac{1}{g_1(\tau_{EH}^*, f_0^*)} e^{-\frac{1}{g_1(\tau_{EH}^*, f_0^*)}} \right) \right], \quad (4.47)$$

каде $W_{-1}(z)$ е втората гранка на Ламбертовата W -функција, реална за $-1/e \leq z < 0$. Помошната функција g_1 е дефинирана како:

$$g_1(\tau_{EH}^*, f_0^*) = \frac{h_k (\xi P_0 \tau_{EH}^* - \alpha a_0(f_0^*)^2)}{N_0 b_0 \log(2)}, \quad (4.48)$$

каде f_0^* е оптималната фреквенција на CPU на EHUs, иста за сите K корисници:

$$f_k^* = f_0^* = \begin{cases} f_c^*, & f_c^* < f_{\max}, \forall k. \\ f_{\max}, & f_c^* \geq f_{\max} \end{cases} \quad (4.49)$$

Оптималната фреквенција f_c^* и оптималното времетраење на ЕН-фазата, τ_{EH}^* , се наоѓаат преку решавање на следниов систем трансцендентни равенки:

$$\sum_{k=1}^K h_k g_2(\tau_{EH}^*, f_c^*) = \frac{1}{\xi P_0}, \quad (4.50)$$

$$\sum_{k=1}^K g_2(\tau_{EH}^*, f_c^*) = \frac{1}{2\alpha(f_c^*)^3}, \quad (4.51)$$

каде помошната функција g_2 се дефинира како:

$$g_2(\tau_{EH}^*, f_c^*) = \frac{\frac{h_k}{B N_0 + p_k^* h_k}}{\log \left(1 + \frac{p_k^* h_k}{B N_0} \right) - \frac{p_k^* h_k}{B N_0 + p_k^* h_k}}. \quad (4.52)$$

Откако ќе се пресметаат p_k^* , f_0^* и τ_{EH}^* , оптималното времетраење на фазата на локална

пресметка се одредува преку:

$$\tau_{LC}^* = \frac{a_0(\eta)}{f_0^*}, \quad (4.53)$$

додека оптималното времетраење на фазата за трансмисија на k -от EHU е:

$$t_k^* = \frac{b_0}{B \log_2 \left(1 + \frac{p_k^* h_k}{BN_0} \right)}. \quad (4.54)$$

Доказот на Теорема 4.2 е даден во додаток Б.6.

Резултатот од Теорема 4.2 има важна практична импликација: дури и ако клиентите имаат различни максимални пресметковни капацитети, оптималната стратегија наложува сите тие да работат на иста, заедничка CPU-фреквенција f_0^* , што значително ја поедноставува имплементацијата на системот.

Со замена на оптималните вредности од Теорема 4.2 во (4.35), го изразуваме минималното вкупно време за тренирање како функција само од η :

$$T_{\min}(\eta) = \frac{\alpha - \beta \log(\eta)}{1 - \eta}, \quad (4.55)$$

каде $\alpha = v_2(\tau_{EH}^* + \sum_{k=1}^K t_k^*)$ и $\beta = v_1 v_2 L_0 D_0 / f_0^*$. Функцијата (4.55) е конвексна во однос на η во интервалот $(0, 1)$ и затоа има единствен глобален минимум, η^* , кој се наоѓа со изедначување на изводот $dT_{\min}(\eta)/d\eta$ со нула, што резултира со следнава трансцендентна равенка:

$$\eta^* (\alpha + \beta - \beta \log(\eta^*)) = \beta. \quad (4.56)$$

Оптималната вредност на η^* се одредува нумерички, на пример со Њутновиот метод, кој бара приближно 20 итерации за точност од шест децимални места [100]. Подпроблемот (4.42) може да се реши нумерички со: (а) методот на внатрешна точка (на пр. со CVX [101]), или (б) Теорема 4.2.

Ако се користи методот на внатрешна точка, подпроблемот (4.42) типично се решава во i_2 чекори ($10 < i_2 < 100$), секој со комплексност $\max\{N_1^3, N_1^2 N_2\}$, каде $N_1 = 3K + 2$ и $N_2 = 4K$. Вкупната комплексност за решавање на (4.36) преку алтернативна оптимизација е $\mathcal{O}(i_1 i_2 K^3)$. Со користење на Теорема 4.2, решавањето бара нумеричко решавање на систем од две равенки (4.50)–(4.51), во i_3 чекори ($10 < i_3 < 100$), што резултира со вкупна комплексност $\mathcal{O}(i_1 i_3)$ – независна од бројот на клиенти K , што го прави решението високо скалабилно.

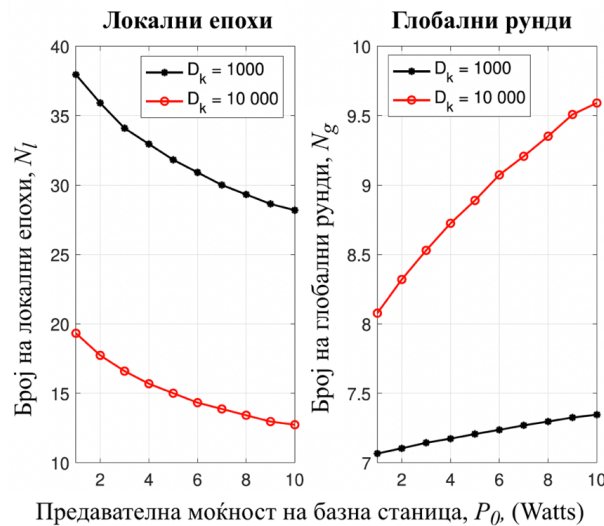
За целите на евалуацијата, разгледуваме безжично напојувана комуникациска мрежа (WPCN) со една БС и $K = 10$ клиенти EHUs, дистрибуирани на растојанија $d_k = 2k$ m (за $1 \leq k \leq K$). Слабењето на каналот е дадено со $h_k = 10^{-3} d_k^{-3}$. Останатите системски параметри се дадени во Табела 4.2.

Табела 4.2: Системски параметри за предложената комуникациска шема.

Параметар	Симбол	Вредност
Спектрална густина на моќност на термички шум	N_0	-174 dBm/Hz
Пресметковен интензитет	L_0	1000 циклуси/примерок
Коефициент на пресметковна ефикасност	α	10^{-28}
Големина на локалното податочно множество	D_0	1000 примероци
Параметри на моделот на ФУ	v_1, v_2	$v_1 = 10, v_2 = 1$
Големина на пренесениот локален модел	b_0	1 Mbit
Комуникациски пропусен опсег	B	10 MHz
Максимална CPU-фреквенција	$f_{\max}$	1 GHz

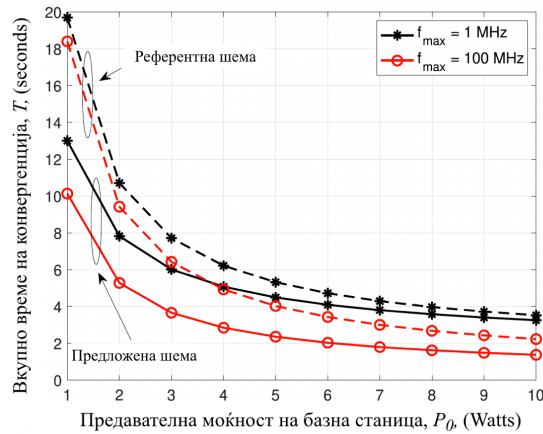
За да ја нагласиме важноста на оптимизацијата на параметарот η , перформансите на предложената шема (полни линии) ги споредуваме со референтна шема (испрекинати линии) која ги оптимизира комуникациските и пресметковните параметри, но не и параметарот на учење – за референтната шема η е фиксиран на 0.5 [29].

Слика 4.6 го прикажува оптималниот број на локални епохи (N_l) и оптималниот број на глобални рунди (N_g) како функција од предавателната моќност на БС (P_0). Се забележува инверзен однос помеѓу N_l и N_g : кога БС зрачи со мала моќност, енергијата е најкритичниот ресурс, па алгоритмот избира мала вредност за η^* (темелно локално тренирање, малку глобални рунди). Кога БС зрачи со голема моќност, енергијата престанува да биде ограничувачки фактор, па алгоритмот избира поголема вредност за η^* (побрзо, поплатко локално тренирање, повеќе глобални рунди), со цел да се искористи брзината на комуникацијата.

**Слика 4.6:** Оптимален број на локални епохи (N_l) и глобални рунди на тренирање (N_g) наспроти предавателната моќност на БС (P_0).

Овој резултат ја демонстрира интелигенцијата на предложената шема: наместо да користи фиксна стратегија, алгоритмот динамички го адаптира балансот помеѓу локалните

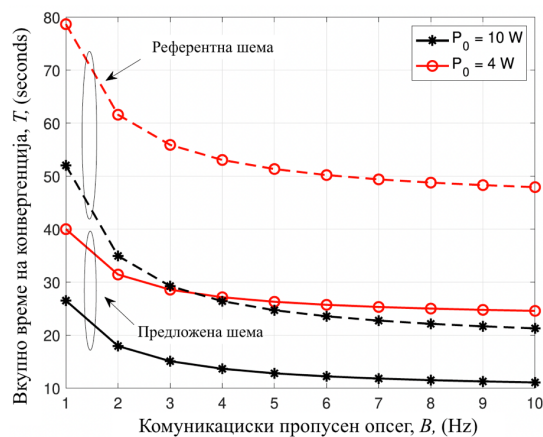
пресметки и глобалната комуникација врз основа на достапноста на клучниот ресурс – енергијата.



Слика 4.7: Вкупно време на конвергенција во системот, T , наспроти предавателната моќност на базната станица P_0 .

Слика 4.7 го прикажува односот помеѓу P_0 и минимизираното вкупно време на конвергенција, за две вредности на $f_{\max}$. Забележуваме јасен тренд на намалување на вкупното доцнење како што се зголемува P_0 , поради побрзото локално тренирање (повисоки f_k) и побрзот пренос (повисоки p_k). Предложената шема постигнува конзистентно и значително помало доцнење во споредба со референтната шема, бидејќи динамички го оптимизира параметарот на учење η . Системите со поголема максимална CPU-фреквенција постигнуваат помало вкупно доцнење, што покажува дека придобивките во времето од побрзите пресметки ги надминуваат дополнителните енергетски трошоци.

Слика 4.8 го прикажува односот помеѓу комуникацискиот пропусен опсег (B) и минимизираното време на конвергенција. Доцнењето се намалува како што се зголемува пропусниот опсег, но трендот не е линеарен, туку асимптотски – при многу големи вредности на B , времетраењето на комуникациската фаза станува занемарливо мало во споредба со времетраењето на фазата на собирање енергија и локалното тренирање, така што кривите се израмнуваат кон константна вредност, ограничена од WPT и пресметковните капацитети на системот.



Слика 4.8: Вкупно време на конвергенција во системот, T , наспроти комуникацискиот пропусен опсег B .

Дел III

Семантички комуникации

Глава 5

Безжичен семантички пренос на информации

Експоненцијалниот раст на интелигентни апликации кои бараат интензивна обработка на податоци ги доведува традиционалните комуникациски системи, изградени врз теоријата на Шенон, до нивните теоретски граници. Овие системи се дизајнирани за совршен пренос на бити на податоци, притоа останувајќи несвесни за нивното значење или за целта на комуникациската задача. Оваа глава претставува сеопфатен преглед на семантичките комуникации, револуционерен концепт кој го префрла фокусот од точност на ниво на бит кон зачувување и интерпретација на значењето. Со пренесување само на суштинските семантички информации потребни за дадена задача, овој пристап ветува значителен исчекор во ефикасноста и одржливоста на системите, особено за идните 6G мрежи.

Повеќе од седумдесет години, теоријата на информации на Шенон претставува теоретски темел на дигиталните комуникации [102]. Нејзиниот фокус на техничкиот проблем, односно на сигурен пренос на бит, беше движечка сила зад технолошката еволуција од 1G до 5G мрежите. Сепак, овој пристап денес се соочува со фундаментални ограничувања, со оглед на тоа што се приближуваме до теоретската граница на Шенон, заради експоненцијалниот раст на податоци генерирани од апликации базирани на вештачка интелигенција, како што се IoT, автономните системи и Метаверзумот [103].

Решението лежи во преиспитување на основните прашања поставени од Вивер, кој ја претставува комуникацијата на три нивоа: технички проблем, семантички проблем и проблем на ефективност [103]. Семантичките комуникации претставуваат суштинска промена во пристапот на комуницирање, од чисто техничко ниво (Ниво А) кон прифаќање на значењето (Ниво Б) и ефективноста (Ниво Ц). Целта повеќе не е совршено и без грешка да се реконструираат изворните податоци (бити), туку да се овозможи на приемникот успешно да го протолкува наменетото значење или да постигне одредена цел. Ова овозможува значителна компресија на податоците со отфрлање на информациите кои се семантички ирелевантни за задачата, ветувајќи огромни придобивки во ефикасноста, доцнењето и одржливоста за 6G мрежите [34].

Иако терминот „семантички комуникации“ често се користи широко, овој термин може дополнително да се систематизира во три различни категории: семантички-ориентирани комуникации, целно-ориентирани комуникации и семантички-свесни комуникации [35].

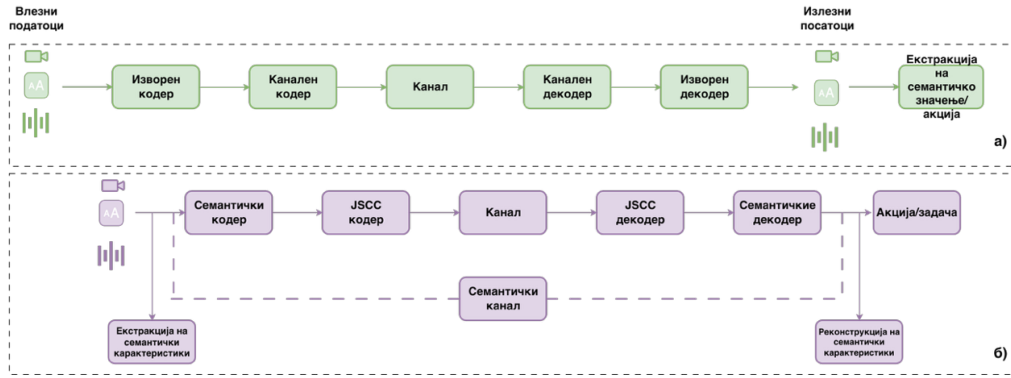
- **Семантички-ориентирани комуникации**, кои одговараат на Нивото Б од теоретската рамка на Вивер, се стремат кон прецизна реконструкција на суштинското значење на информацијата. Овој пристап е фундаментален за апликации како метаверзум и проширената реалност, каде примарната цел не е пренос на необработени податоци, туку пренос на концизна семантичка репрезентација. Перформансите на системот се проценуваат според степенот на семантичка точност (верност) на крајната реконструкција.
- **Целно-ориентираните комуникации** ја издигнуваат целта на ефективност (ниво Ц), каде успехот се мери со перформансите на последователна задача – критичен пристап во случајот на автономни системи и возила, каде крајната метрика за перформанси е успехот на задачата (на пр. избегнување судир), а не верноста на податоците од сензорот.
- **Семантички-свесни комуникации** претставуваат еволуција на системско ниво, опишувајќи мрежна инфраструктура која е свесна за содржината и информациите што ги пренесува, овозможувајќи оптимизација на ресурсите врз основа на семантичката важност, наместо едноставна максимизација на брзината на податоци.

5.1 Теориски основи на семантичките комуникации

Концептуалната промена од комуникација базирана на бити кон семантичка комуникација бара нова теоретска рамка. Во својот епохален труд од 1948 година, Шенон воспостави математичка рамка за комуникација која го апстрахира проблемот во математички модел со користење на теорија на веројатност [102]. Шеноновиот комуникациски модел се состои од извор на информации кој генерира порака, предавател кој ја кодира пораката во сигнал, канал кој го пренесува сигналот (притоа деградирајќи го со шум), приемник кој го декодира ошумениот сигнал, и дестинација која ја прима пораката, како што е прикажано на Слика 5.1.

Основата на теоријата на информации е квантификацијата на информацијата. Ентропијата на дискретна случајна променлива X со функција на веројатност $p(x)$ е дефинирана како:

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log_2 p(x). \quad (5.1)$$



Слика 5.1: Споредба на модели на комуникациски системи: (а) класична архитектура на Шенонов комуникациски систем, (б) архитектура на систем за семантичка комуникација.

Ентропијата ја мери просечната неизвесност за можните исходи на случајната променлива, изразено во бити. За споредба на две распределби на веројатност, $p(x)$ и $q(x)$, се користи Кулбак-Лајблеровата (KL) дивергенција:

$$D_{KL}(p(x)||q(x)) = \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{p(x)}{q(x)}. \quad (5.2)$$

KL-дивергенцијата е мерка за неефикасноста на претпоставката дека распределбата е q кога вистинската распределба е p . Здружената информација $I(X; Y)$ помеѓу две случајни променливи X и Y ја мери информацијата што Y ја дава за X , и може да се дефинира како KL-дивергенција помеѓу здружената распределба $p(x, y)$ и производот на поединечните распределби $p(x)p(y)$:

$$I(X; Y) = D_{KL}(p(x, y)||p(x)p(y)) = \sum_{x \in \mathcal{X}} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}. \quad (5.3)$$

Со проширување на изразот во (5.3), се добива поповообичаената дефиниција базирана на ентропија, според која здружената информација претставува намалување на неизвесноста за променливата X што се добива како резултат на познавањето на променливата Y :

$$I(X; Y) = H(X) - H(X|Y). \quad (5.4)$$

Фундаменталното ограничување за сигурна комуникација преку канал со шум е дефинирано преку капацитетот на канал, C , дефиниран како максималната вредност на здружената информација меѓу влезот X и излезот Y на каналот:

$$C = \max_{p(x)} I(X; Y). \quad (5.5)$$

Според Првата Шенонова теорема, за која било комуникациска брзина R помала од капацитетот C , постои шема за кодирање и декодирање со која може да се постигне произволно ниска веројатност за грешка; за брзина R поголема од C , сигурен пренос е невозможен.

Најдиректниот мост помеѓу класичната и семантичката комуникација го обезбедува те-

оријата на рата-дисторзија, која ја квантифицира минималната рата R (во бити по симбол) потребна за пренос на X така што тој може да се реконструира како Y со просечна дисторзија што не надминува одреден праг D^* :

$$R(D^*) = \min_{D < D^*} I(X; Y), \quad (5.6)$$

каде $D = \sum_{x,y} p(x)p(y|x)d(x,y)$ е функцијата на дисторзија, а $d(x,y)$ е метрика за дисторзија ($d(x,y) = 0$ кога $x = y$). Теоријата на рата-дисторзија може да се адаптира за семантички комуникациски систем:

$$R(D_s, D_a) = \min I(Z; \hat{X}, \hat{Z}), \quad (5.7)$$

каде D_s е семантичката дисторзија помеѓу изворот X и воспоставената информација $\hat{X}$ во приемникот, а D_a е дисторзијата помеѓу семантичката репрезентација Z и примената семантичка репрезентација $\hat{Z}$.

Информациското тесно грло е пристап за наоѓање на оптималниот компромис помеѓу компресијата и точноста, даден преку следниов оптимизациски проблем:

$$\min_{p(Z|X)} I(X; Z) - \beta I(V; Z), \quad (5.8)$$

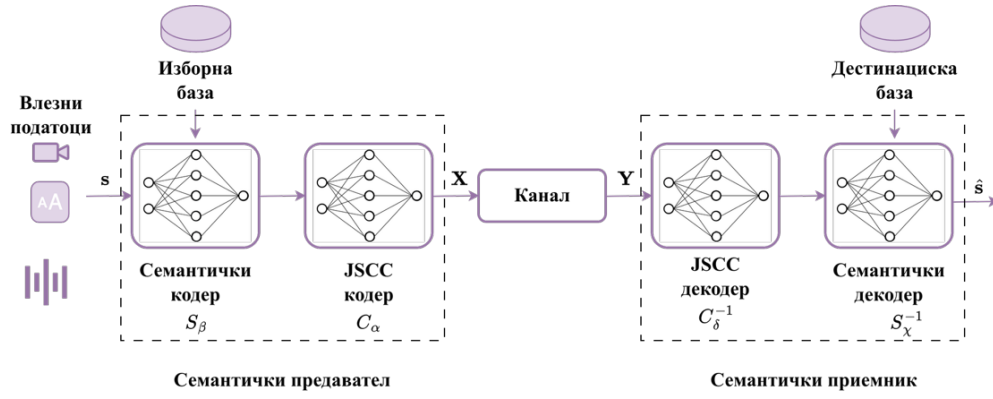
каде V е посакуваната семантичка репрезентација, а β е множител кој го балансира компромисот. Ова може да се прошири преку следнава функција на загуба:

$$\mathcal{L} = \underbrace{I(Z; X)}_{\text{Компресија}} - \underbrace{(1 - \alpha)I(Z; \hat{Z})}_{\text{Здружиена информација}} + \underbrace{\beta D_{KL}(p(x), p(\hat{z}))}_{\text{Инференца}}, \quad (5.9)$$

каде α и β се тежински параметри за прилагодување на компонентите на здружената информација и инференца.

5.2 Системски модели за пренос на семантичка информација

Транзицијата на семантичката комуникација од теоретски концепт во инженерска дисциплина бара развој на конкретни архитектури и методологии, кои се речиси исклучиво овозможени од длабокото учење. Длабоките невронски мрежи (ДНМ) се супериорни во учењето на сложени, хиерархиски репрезентации од огромни количини на неструктурирани податоци, што ги прави совршено погодни за улогата на семантички кодери и декодери. Фундаменталното отстапување од класичниот модел на Шенон е преминот кон мрежна архитектура од крај до крај базирана на длабоко учење, која ги поместува ригидните граници помеѓу кодирањето на изворот, кодирањето на каналот и модулацијата, и ја рedefинира комуникациската верига како еден, заеднички оптимизиран автоенкодер [36], како што е илустрирано на Слика 5.2.



Слика 5.2: Архитектура на семантички комуникациски систем базиран на длабоко учење.

Во овој пристап, класичното одвојување на кодирањето на изворот и кодирањето на каналот е напуштено. Семантичкиот предавател е унифицирана невронска мрежа која извршува две тесно интегрирани функции. Прво, семантички кодир, параметризиран со β и означен како $S_\beta(\cdot)$, ги обработува необработените изворни податоци s за да го извлече нивното суштинско значење (семантичките атрибути). Овој процес го дестилира високо-димензионалниот влез во компактен, латентен вектор, практичен пандан на „тесното грло“ од принципот на ИВ.

Класичниот пристап на теоремата на Шенон за одвојување на кодирањето на изворот и каналот е фундаментално несоодветен за семантичка комуникација, бидејќи при фазата на репрезентација на ниво на бит се отфрла богатата, континуирана структура на семантичките карактеристики. Модерниот пристап ги надминува овие ограничувања со усвојување на целовита методологија од крај до крај позната како Длабоко заедничко кодирање на изворканал (анг. Deep Joint Source-Channel Coding) (Deep-JSCC) [37]. Издвоениот семантички латентен вектор веднаш се проследува до JSCC-кодир, параметризиран со α и означен како $C_\alpha(\cdot)$, кој учи да ја мапира семантичката репрезентација директно во множество на комплексни симболи $\mathbf{x}$, погодни за пренос преку физичкиот канал:

$$\mathbf{x} = C_\alpha(S_\beta(s)). \quad (5.10)$$

Клучниот увид е дека и семантичкиот кодир и JSCC-кодирот се тренираат заеднички, што му овозможува на системот да научи оптимално мапирање, свесно и за семантичката структура на изворот и за физичките својства на каналот.

Откако ќе помине низ канал со шум, примениот сигнал е $\mathbf{y} = h\mathbf{x} + \mathbf{n}$, каде h е коефициентот на каналот, а $\mathbf{n}$ е адитивен шум. Семантичкиот приемник ги извршува инверзните операции: JSCC-декодирот, параметризиран со δ и означен $C_\delta^{-1}(\cdot)$, ги зема ошумените симболи $\mathbf{y}$ и се обидува да го реконструира латентниот семантички вектор, кој потоа се проследува до семантичкиот декодер, параметризиран со χ и означен $S_\chi^{-1}(\cdot)$:

$$\hat{\mathbf{s}} = S_\chi^{-1}(C_\delta^{-1}(\mathbf{y})). \quad (5.11)$$

Целиот систем се тренира како една единствена целина, водена од хибридна функција на загуба која балансира семантичка верност и брзина на пренос. За текстуални податоци, семантичката грешка типично се минимизира со загуба од крос-ентропија (CE), која за реченица со должина L се формулира како:

$$\mathcal{L}_{CE}(\mathbf{s}, \hat{\mathbf{s}}) = - \sum_{l=1}^L [q(w_l) \log(p(w_l)) + (1 - q(w_l)) \log(1 - p(w_l))], \quad (5.12)$$

каде $q(w_l)$ е вистинската веројатност, а $p(w_l)$ е предвидената веројатност за l -от збор. Секундарна цел е да се максимизира здружената информација $I(x; y)$ помеѓу пренесените и примените симболи, преку минимизирање на член на загуба изведен од дуалната репрезентација на KL-дивергенцијата:

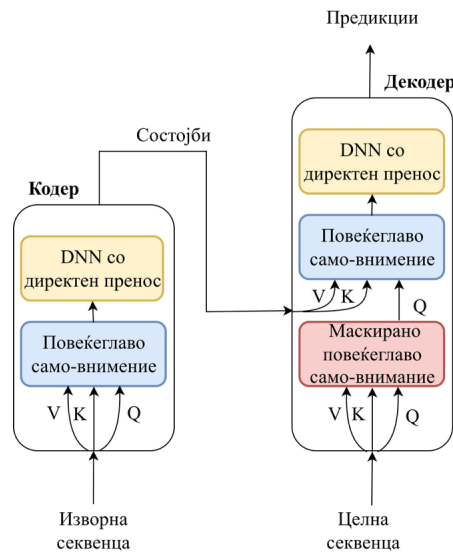
$$\mathcal{L}_{MI}(\mathbf{x}, \mathbf{y}; T) = \mathbb{E}_{p(x,y)} [f_T] - \log \left(\mathbb{E}_{p(x)p(y)} [e^{f_T}] \right), \quad (5.13)$$

каде f_T е невронска мрежа тренирана да разликува примероци од заедничката распределба $p(x, y)$ и производот на маргиналните распределби $p(x)p(y)$. Крајната, вкупна функција на загуба е:

$$\mathcal{L}_{total} = \mathcal{L}_{CE} - \lambda \mathcal{L}_{MI}, \quad (5.14)$$

каде λ е хиперпараметар кој го балансира компромисот помеѓу семантичката точност и брзината на пренос.

Кога станува збор за текст, целта е да се пренесе значењето на реченицата, а не нејзиниот точен зборовен состав. Доминантните модели се базираат на трансформери (анг. transformers), предложени во епохалниот труд „Attention Is All You Need“ [104]. Клучната иновација е механизмот на само-влијание (анг. self-attention), кој оперира на три научени векторски репрезентации за секој влезен токен: Прашање (Query – Q), Клуч (Key – K), и Вредност (Value – V). Архитектурата се состои од кодер и декодер (Слика 5.3).



Слика 5.3: Базична архитектура на трансформер.

Целта на кодерот е да генерира богата, контекстуално-свесна репрезентација на влезната секвенца, чиј финален излез се вектори за Клуч (K) и Вредност (V) кои го енкапсулираат семантичкото значење на целата изворна секвенца. Декодирот ја генерира излезната секвенца елемент по елемент, преку два механизми на внимание – маскиран слој за само-внимание врз сопствениот излез, и слој кој го формира векторот за Прашање (Q) од претходниот слој, додека K и V доаѓаат директно од излезот на кодирот.

Семантичкиот кодир при пренос на текстуални податоци често е базиран на претходно трениран трансформер модел како BERT (Bidirectional Encoder Representations from Transformers) [105], кој ја анализира целата реченица истовремено, наместо збор по збор, со што прецизно го одредува значењето на секој збор врз основа на сите околни зборови. Моделот DeepSC, развиен во [36], е најистакнатиот пример за текстуални податоци – користи авто-енкодер базиран на трансформер за заедничко учење на семантичка компресија и кодирање на канал, докажувајќи супериорни перформанси во однос на традиционалните методи во режими со низок SNR.

5.3 Перформансни метрики за семантички пренос на информација

Класичните комуникациски системи, проектирани за синтаксичка верност, долго време се потпираа на добро воспоставени метрики, како Bit-Error Rate (BER), Symbol-Error Rate (SER) и Peak Signal-to-Noise Ratio (PSNR). Меѓутоа, суштинската премиса на семантичките комуникации – давање приоритет на значењето и ефективността пред буквалната интерпретација – ги прави овие традиционални метрики недоволни, а понекогаш и погрешни. Систем кој совршено го пренесува значењето на реченица со замена на синоним може да биде огромен успех од семантичка перспектива, а сепак би имал лоша вредност на BER.

Доминантниот современ пристап за квантификација на семантичка блискост е да се користат големи, претходно тренирани модели на длабоко учење. За текстуални информации, најсовремената метрика е семантичка сличност (анг. semantic similarity), која ја квантифицира семантичката блискост на две реченици со споредување на нивните векторски вградувања [38]. Нека s е оригиналната изворна реченица, а $\hat{s}$ е реченицата реконструирана на приемникот. Семантичката сличност, ξ , помеѓу нив се дефинира како косинусна сличност на нивните соодветни векторски вградувања:

$$\xi(s, \hat{s}) = \frac{B(s) \cdot B(\hat{s})^T}{\|B(s)\|_2 \cdot \|B(\hat{s})\|_2}, \quad (5.15)$$

каде $B(\cdot)$ е функцијата за вградување на претходно трениран модел како BERT. Метриката ξ се движи од 0 до 1, каде вредност 1 означува совршена семантичка еквивалентност.

За да се овозможи распределба на ресурси во семантички-свесни мрежи, неодамнешните истражувања ги проширија класичните концепти на предавателна информациска брзина и спектрална ефикасност во семантичкиот домен [38]:

- **Семантичка единица (sut):** основна единица на семантичка информација, аналогна на „бит“.
- **Семантичка информациска брзина** (анг. semantic rate) (S-R), Γ : ратата на ефективно пренесени семантички информации во секунда (suts/s). Ако очекуваната количина на семантичка информација по реченица е I (во suts), а очекуваниот број на зборови по реченица е L , и системот користи просечно k_n семантички симболи за да го претстави секој збор, тогаш при пропусен опсег W , ефективната семантичка информациска брзина е:

$$\Gamma_{n,m} = \frac{WI}{k_n L} \xi_{n,m}. \quad (5.16)$$

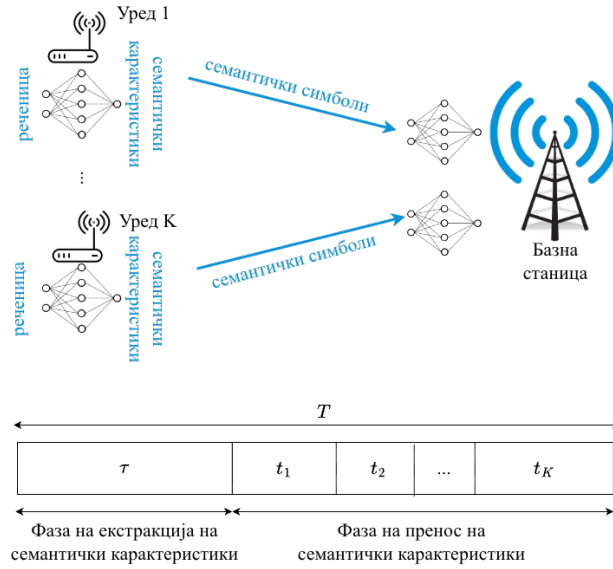
- **Семантичка спектрална ефикасност** (анг. semantic spectral efficiency) (S-SE), Φ : ратата со која семантичките информации можат успешно да се пренесат по единица пропусен опсег (suts/s/Hz):

$$\Phi_{n,m} = \frac{I}{k_n L} \xi_{n,m}. \quad (5.17)$$

5.4 Оптимизација на ресурси во безжични семантички мрежи

Имајќи ја предвид големата пресметковна комплексност за екстракција на семантички атрибути, како и чувствителноста на пренос на истите во услови на безжичен канал, од суштинско значење е да се дефинираат механизми кои истовремено ќе ја оптимизираат енергијата за пресметка и пренос. Врз основа на рамката DeepSC, трудот [36] воведува релевантни метрики за перформанси, како семантичка информациска брзина и семантичка сличност, кои можат да се користат за распределба на ресурси во семантички-свесни мрежи. Користејќи ги овие метрики, [38] ја максимизира семантичката рата за повеќекориснички систем, додека [106] го карактеризира регионот на семантичка наспроти комуникациска рата на хетероген систем. Семантичките безжични мрежи се изучувани и во контекст на енергетската ефикасност во [40], меѓутоа предложеното решение е субоптимално, бидејќи проблемот на екстракција на семантички атрибути се разгледува независно од проблемот на распределба на ресурси, а федингот во каналот е занемарен.

Разгледуваме безжична мрежа наменета за семантички комуникации, која се состои од една заедничка базна станица (БС) и K безжични уреди/корисници. Секој безжичен уред е опремен со семантички кодер кој е спарен со соодветен семантички декодер лоциран на БС, користејќи техники базирани на трансформери, како моделот DeepSC [36]. Системското време е поделено на епохи со времетраење T . Како што е илустрирано на Слика 5.4, секоја епоха е поделена на две последователни, ортогонални фази: фаза на екстракција на семантички атрибути (анг. feature extraction) (FE) и фаза на пренос на семантички атрибути (анг. feature transmission) (FT).



Слика 5.4: Системски модел на предложената семантичка комуникациска мрежа.

Во текот на ФЕ-фазата, k -от безжичен уред ($1 \leq k \leq K$) добива реченица со должина L зборови на својот влез, а неговиот кодер извлекува семантички атрибути со просечна должина s_k семантички симболи по збор. Нека a_k го означува бројот на CPU-циклуси потребни за k -от уред да генерира вектори на семантички атрибути со просечна должина Ls_k по реченица, а f_k фреквенцијата на процесорот. Претпоставуваме дека пресметковниот товар е ист за сите корисници, $a_k = a_0, \forall k$. Времетраењето на ФЕ-фазата за k -от уред е a_0/f_k , а нејзиното вкупно времетраење, τ , се одредува како:

$$\tau = \max_{1 \leq k \leq K} \left\{ \frac{a_0}{f_k} \right\}. \quad (5.18)$$

Енергијата потребна за k -от уред да го генерира својот вектор на семантички атрибути е:

$$E_k^{FE} = \alpha a_0 f_k^2, \quad (5.19)$$

каде α е коефициентот на енергетска ефикасност. Во текот на FT-фазата, корисниците ги пренесуваат своите вектори на карактеристики до БС преку TDMA, каде на секој корисник му се доделува ексклузивен временски слот t_k . Нека g_k го означува засилувањето на каналот, p_k предавателната моќност, B пропусниот опсег, а N_0 спектралната густина на моќност на термичкиот шум. Енергијата за пренос е:

$$E_k^{FT} = p_k t_k, \quad (5.20)$$

а односот сигнал-шум на примениот сигнал е:

$$\gamma_k = \frac{p_k g_k}{B N_0}. \quad (5.21)$$

За да ја квантифицираме семантичката точност помеѓу кодерот на k -от уред и декодерот

на БС, ја усвојуваме метриката семантичка сличност, ξ_k , дефинирана со (5.15). Според [106], оваа функција може да се апроксимира со генерализирана логистичка функција:

$$\xi_k = \xi_k(s_k, \gamma_k) = A_{s_k,1} + \frac{A_{s_k,2} - A_{s_k,1}}{1 + e^{-(C_{s_k,1}\gamma_k + C_{s_k,2})}}, \quad (5.22)$$

каде параметрите $A_{s_k,1}, A_{s_k,2}, C_{s_k,1}, C_{s_k,2}$ ја дефинираат функцијата и се претпоставуваат константни за сите корисници (A_1, A_2, C_1, C_2). Врз основа на семантичката сличност $\xi(\gamma_k)$, ја дефинираме семантичката рата на k -от уред:

$$R_k = \frac{IB}{s_k L} \xi(\gamma_k) = \frac{IB}{s_k L} \xi\left(\frac{p_k g_k}{BN_0}\right), \quad (5.23)$$

каде I е просечната количина на семантички информации содржана во реченица (во suts).

Нашата цел е да ја минимизираме вкупната енергија потрошена од сите K корисници, збир од енергијата за екстракција (E_k^{FE}) и енергијата за пренос (E_k^{FT}). Прво ја анализираме поедноставената, но фундаментална ситуација каде каналите се статични, $g_k = \Omega_k$:

$$\text{Minimize}_{\tau, t_k, f_k, p_k} \sum_{k=1}^K (p_k t_k + \alpha a_0 f_k^2) \quad (5.24)$$

$$\text{под услов: } C1 : t_k \frac{IB}{s_k L} \xi\left(\frac{p_k g_k}{BN_0}\right) \geq R_{th} T, \forall k \quad (5.25)$$

$$C2 : \xi\left(\frac{p_k g_k}{BN_0}\right) \geq \xi_{th}, \forall k \quad (5.26)$$

$$C3 : \tau + \sum_{k=1}^K t_k = T \quad (5.27)$$

$$C4 : p_k \leq P_{\max}, \forall k \quad (5.28)$$

$$C5 : f_k \leq f_{\max}, \forall k \quad (5.29)$$

$$C6 : \frac{a_0}{f_k} \leq \tau, \forall k \quad (5.30)$$

каде C1 е ограничување на квалитетот на услугата (QoS) во однос на семантичката рата, C2 поставува минимален праг на семантичка сличност ξ_{th} , C3 го дефинира вкупното времетраење на една рунда, C4 и C5 се практични ограничувања за максимална моќност и фреквенција, а C6 гарантира дека FE-фазата е доволно долга за најбавниот корисник. Оптимизацискиот проблем (5.24) е неконвексен, но со математички манипулации може да се реши аналитички. Воведуваме помошни променливи $r_k = R_{th} s_k L / (BI)$ и $\sigma_N^2 = BN_0$.

Теорема 5.1. *За постоење на оптимално решение на (5.24), мора да бидат задоволени следниве услови за изводливост:*

$$\frac{a_0}{T f_{\max}} + \frac{1}{\xi_{th}} \sum_{k=1}^K r_k \leq 1 \quad (5.31)$$

и

$$\xi\left(\frac{P_{\max} \Omega_k}{\sigma_N^2}\right) \geq \xi_{th}. \quad (5.32)$$

Кога (5.31) и (5.32) се задоволени, глобално оптималното решение е дадено со: оптимално времетраење на FE-фазата,

$$\tau^* = T \left(1 - \frac{1}{\xi_{th}} \sum_{k=1}^K r_k \right), \quad (5.33)$$

оптимални предавателни моќности,

$$p_k^* = \frac{\sigma_N^2}{C_1 \Omega_k} \left(\log \left(\frac{\xi_{th} - A_1}{A_2 - \xi_{th}} \right) - C_2 \right), \forall k, \quad (5.34)$$

оптимални времетраења на FT-фазите,

$$t_k^* = \frac{r_k}{\xi_{th}} T, \forall k, \quad (5.35)$$

и оптимални CPU-фреквенции,

$$f_k^* = \frac{a_0}{\tau^*}. \quad (5.36)$$

Ако условот (5.31) е задоволен како строга неравенка, тогаш $f_k^* < f_{\max}$; ако е задоволен со строга еднаквост, тогаш $f_k^* = f_{\max}$. Деталниот доказ е даден во додаток Б.7.

Решението во затворена форма нуди неколку важни увиди: сите корисници работат на иста фреквенција на процесорот, $f_k^* = a_0/\tau^*$, што значително ја поедноставува имплементацијата; сите корисници пренесуваат со моќност доволна само за да се постигне минимално дозволена семантичка сличност; а времето за пренос t_k^* е директно пропорционално на „семантичкиот товар“ r_k на секој корисник.

Сега го прошируваме истражувањето за да го опфатиме пореалното сценарио каде безжичниот канал е изложен на случаен фединг, при што засилувањето на каналот g_k повеќе не е детерминистичка вредност, туку случајна променлива, што воведува можност за прекини во комуникацијата (outage). Кога каналот g_k е во состојба на „длабок фединг“, ограничувањето C2 не може да се исполни, дури и при максимална моќност. Поради ова, го рedefинираме ограничувањето C1 со користење на математичко очекување – наместо да бараме семантичката рата да биде задоволена во секој момент, бараме просечната семантичка информациска брзина да го задоволи прагот R_{th} :

$$\frac{t_k B I}{s_k L} \mathbb{E} \left[\xi \left(\frac{p_k g_k}{B N_0} \right) \mathbb{I} \left(\xi \left(\frac{p_k g_k}{B N_0} \right) \geq \xi_{th} \right) \right] \geq \bar{R}_{th} T, \quad (5.37)$$

каде $\mathbb{I}(\cdot)$ е индикаторска функција:

$$\mathbb{I}(\xi \geq \xi_{th}) = \begin{cases} 1, & \xi \geq \xi_{th} \\ 0, & \xi < \xi_{th} \end{cases}. \quad (5.38)$$

Очекуваната вредност во (5.37) може аналитички да се определи со интегрирање преку фун-

кцијата на густина на веројатност на случајната променлива $X_k = p_k g_k / (BN_0)$:

$$\mathbb{E} [\xi(x_k) \mathbb{I}(\xi(x_k) \geq \xi_{th})] = \int_{x_{th}}^{\infty} \xi(x) f_{X_k}(x) dx, \quad (5.39)$$

каде x_{th} го задоволува $\xi(x_{th}) = \xi_{th}$. За висока семантичка сличност ($\xi_{th} > 0.7$), $\xi(x)$ е блиску до A_2 , па (5.39) е ограничено од горе:

$$\mathbb{E} [\xi(x_k) \mathbb{I}(\xi(x_k) \geq \xi_{th})] \leq A_2 F_{X_k}(x_{th}). \quad (5.40)$$

За Рејлиев фединг, PDF на X_k е експоненцијална со средна вредност $p_k \Omega_k / (BN_0)$, каде $\Omega_k = \mathbb{E}[g_k]$, што дава:

$$\mathbb{E} [\xi(x_k) \mathbb{I}(\xi(x_k) \geq \xi_{th})] \leq A_2 e^{-\frac{x_{th} BN_0}{p_k \Omega_k}}. \quad (5.41)$$

Користејќи ја оваа горна граница, го формулираме следниов проблем за минимизација на енергијата во фединг канали:

$$\text{Minimize}_{\tau, t_k, f_k, p_k} \sum_{k=1}^K (p_k t_k + \alpha a_0 f_k^2) \quad (5.42)$$

$$\text{под услов: } \overline{C1} : t_k \frac{BI}{s_k L} A_2 e^{-\frac{x_{th} BN_0}{p_k \Omega_k}} \geq \overline{R}_{th} T, \forall k \quad (5.43)$$

$$C3-C6 \text{ (исто како во (5.27)–(5.30))}$$

Слично на статичниот случај, овој неконвексен проблем може да се трансформира и да се добие решение во затворена форма. Нека $\bar{r}_k = \overline{R}_{th} s_k L / (BI)$ и $x_{th} = \xi^{-1}(\xi_{th})$.

Теорема 5.2. *За постоење на оптимално решение на (5.42), мора да биде задоволен следниов услов за изводливост:*

$$\frac{a_0}{T f_{\max}} + \frac{1}{A_2} \sum_{k=1}^K \bar{r}_k e^{\frac{x_{th} \sigma_N^2}{P_{\max} \Omega_k}} \leq 1. \quad (5.44)$$

Времетраењето на FE-фазата, τ^ , се наоѓа како решение на трансцендентна равенка:*

$$\frac{1}{A_2} \sum_{k=1}^K \bar{r}_k T e^{\frac{x_{th} \sigma_N^2}{p_k(\tau^*) \Omega_k}} = T - \tau^*, \quad (5.45)$$

каде $p_k(\tau)$ е дадено со:

$$p_k(\tau) = \frac{x_{th} \sigma_N^2}{2 \Omega_k} \left(1 + \sqrt{1 + \frac{8 K \alpha a_0^3 \Omega_k}{\tau^3 x_{th}}} \right). \quad (5.46)$$

За дадено τ^* , оптималната предавателна моќност е:

$$p_k^* = \min \{p_k(\tau^*), P_{\max}\}, \forall k, \quad (5.47)$$

оптималното времетраење на FT-фазата:

$$t_k^* = \frac{\bar{r}_k T}{A_2} e^{\frac{x_{th} \sigma_N^2}{p_k^* \Omega_k}}, \forall k, \quad (5.48)$$

и оптималната CPU-фреквенција:

$$f_k^* = \frac{a_0}{\tau^*}, \forall k. \quad (5.49)$$

Деталниот доказ на Теорема 5.2 е даден во додаток Б.8.

Решението за фединг канали е значително покомплексно, но нуди важни увиди – за разлика од статичниот случај, оптималната предавателна моќност $p_k(\tau)$ зависи од времетраењето на FE-фазата, што покажува дека во фединг канали постои длабок компромис помеѓу ресурсите за пресметка и комуникација.

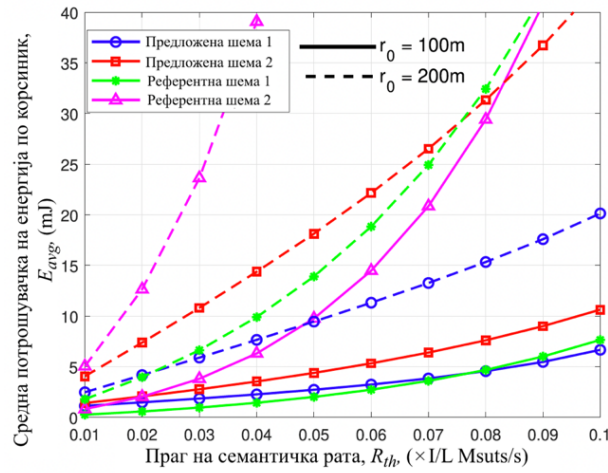
Табела 5.1: Работни параметри на безжичната мрежа за семантички комуникации.

Параметар	Симбол	Вредност
Број на безжични корисници	K	10
Пропусен опсег	B	10 MHz
Времетраење на епоха	T	1 s
Спектрална густина на моќност на шум	N_0	-140 dBm/Hz
Референтно слабеење (на 1 метар)	ρ	-30 dB
Модел на слабеење на каналот	Ω_k	$\rho \cdot r_k^{-2.7}$
Коефициент на пресметковна ефикасност	α	10^{-28}
Вкупен пресметковен товар	a_0	0.2×10^9 циклуси
Максимална предавателна моќност	$P_{\max}$	1 W
Максимална CPU-фреквенција	$f_{\max}$	3 GHz
Праг на семантичка сличност	ξ_{th}	0.7
Број на семантички симболи по збор	$s_k = s_0$	4
Параметри на логистичка функција	A_1, A_2, C_1, C_2	0.37, 0.98, 0.2525, -0.7985

За споредба, воведуваме две референтни шеми за традиционална, бит-базирана безжична мрежа, каде пресметковната енергија за обработка на податоци се занемарува ($\tau = 0$), со цел минимизирање само на енергијата за пренос, $\sum_{k=1}^K p_k t_k$:

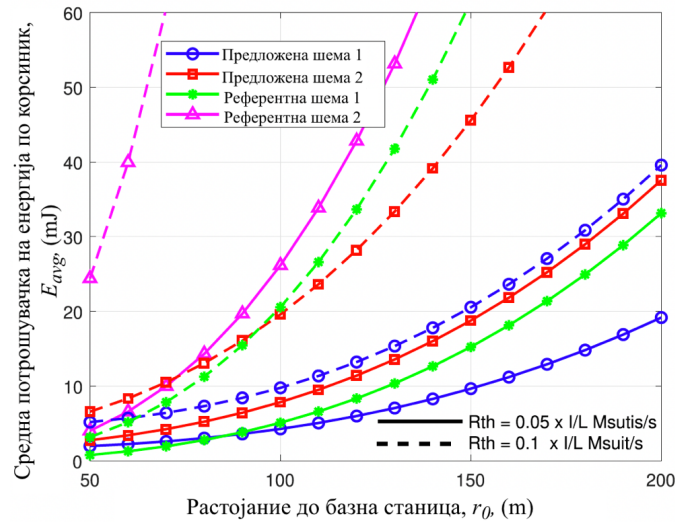
- **Референтна шема 1 (статичен канал):** ги оптимизира (p_k, t_k) да го задоволат условот $t_k B \log_2(1 + p_k \Omega_k / (B N_0)) \geq R_{c,th} T$, каде $R_{c,th} = (\mu_0 L / I) R_{th}$ е еквивалентниот праг за конвенционални комуникации, а $\mu_0 = 32$ е просечниот број на бити по збор [38].
- **Референтна шема 2 (фединг канал):** ги оптимизира (p_k, t_k, R_k) да го задоволат условот $t_k B R_k \cdot \Pr\{\log_2(1 + p_k g_k / (B N_0)) > R_k\} \geq \bar{R}_{c,th} T$.

Перформансите ги споредуваме според средната потрошувачка на енергија по корисник: $E_{avg} = (1/K) \sum_{k=1}^K (p_k t_k + \alpha a_0 f_k^2)$ за предложените шеми, и $E_{c,avg} = (1/K) \sum_{k=1}^K p_k t_k$ за референтните шеми.



Слика 5.5: Влијание на прагот на семантичка информациска брзина врз потрошувачката на енергија.

Слика 5.5 го прикажува клучниот однос помеѓу бараниот квалитет на услуга (R_{th}) и просечната потрошувачка на енергија по корисник (E_{avg}), за корисници распределени униформно на кружница со радиус r_0 ($r_0 = 100$ m и $r_0 = 200$ m). За сите четири шеми, просечната потрошувачка на енергија расте со зголемување на R_{th} , бидејќи за пренос на поголема количина информации е потребна поголема моќност или подолго време. Шемите кои оперираат во фединг канали конзистентно трошат повеќе енергија од своите еквиваленти во статични канали, поради загубата на енергија при прекини во комуникацијата. Најважниот заклучок е демонстрацијата на огромната енергетска ефикасност на предложените шеми при повисоки барања за квалитет на услуга – потрошувачката на енергија кај референтните шеми расте експоненцијално со R_{th} , додека предложените семантички шеми покажуваат многу побавен, речиси линеарен раст, бидејќи пренесуваат само компактни семантички атрибути наместо огромен волумен на не-семантички податоци.



Слика 5.6: Влијание на растојанието до базната станица врз потрошувачката на енергија.

Слика 5.6 го истражува односот помеѓу јачината на каналот, претставена преку растојанието r_0 , и просечната потрошувачка на енергија по корисник, за два прага на семантичка информациска брзина. Во режимот на висок SNR (мал r_0), референтните конвенционални шеми покажуваат малку подобри перформанси, бидејќи дополнителната енергија за пресметка при извлекување на семантички атрибути станува доминантен фактор. Меѓутоа, како растојанието се зголемува и влегуваме во режим на низок SNR, потрошувачката на енергија кај референтните шеми нагло и експоненцијално расте, додека предложените семантички шеми покажуваат многу поблаг раст. Овој график јасно го илустрира практичниот домен во кој семантичката комуникација е најсупериорна: комуникација на долги растојанија или во средини со слаб канал.

5.5 Оптимизација на ресурси во хибридни бит-семантички комуникации

Во претходниот дел, ја истраживме семантичката комуникација како средство за постигнување супериорна енергетска ефикасност. Сепак, добро познат факт од литературата е дека перформансите на SemCom и BitCom се комплементарни: семантичката комуникација е супериорна во режими со низок однос сигнал-шум (SNR), додека конвенционалната бит-базирана комуникација е поефикасна при висок SNR. Оваа дихотомија отвора важно истражувачко прашање: дали е можно да се дизајнира хибриден систем кој динамички се префрла помеѓу двата режима за да ги искористи предностите на секој од нив низ целиот спектар на услови на каналот?

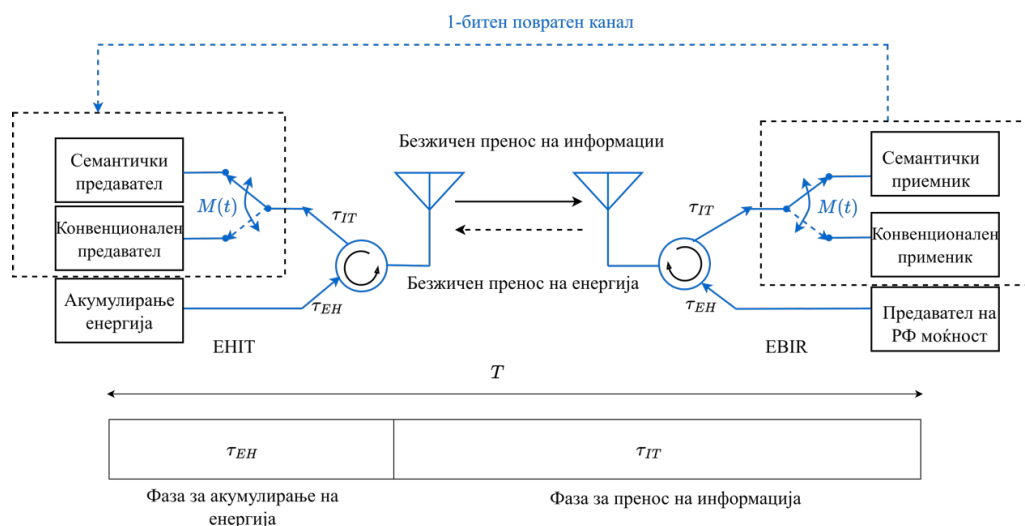
Овој дел го адресира токму тој предизвик. Развиваме и анализираме нов модел за безжично напојуван, би-модален комуникациски систем кој динамички се префрла помеѓу конвенционална и семантичка комуникација. Клучната иновација е воведувањето на полиса за из-

бор на режим која, врз основа на моменталната состојба на каналот, одлучува кој комуникациски режим (BitCom или SemCom) ќе се користи. Оваа одлука се пренесува до предавателот преку еднобитен повратен канал, додека времетраењата на фазите за собирање енергија и пренос на информации остануваат фиксни.

Бидејќи SemCom е релативно нова истражувачка област, литературата нуди мал број трудови кои се занимаваат со хибридни бит-семантички комуникации [41, 106–108]. Во [41], се минимизира доцнењето на точка-точка хибриден бит-семантички систем со повеќе носители, каде секој носител е фиксно одреден да користи или BitCom или SemCom. Во [106], се изучува регионот на семантичка-наспроти-конвенционална рата на хетероген систем со NOMA. Во [107], оваа анализа се проширува со споредба на NOMA, OMA и semi-NOMA системи во фединг канали. Во [108], развиен е систем каде секој корисник може да избира помеѓу BitCom и SemCom-режими за максимизирање на достижната рата.

Клучен аспект на SemCom е што остварливите семантички рати зависат од семантичката сличност, која мора да надмине одреден праг за ефективна SemCom. Слично на BitCom [109], присуството на фединг и недостатокот на CSI на предавателот понекогаш може да резултира со прекин на комуникациите. Бидејќи SemCom покажува значителни предности пред BitCom при низок SNR, тој може да се испреплетува со технологијата за безжично напојувани комуникации (WPC) за да се подобри енергетската одржливост на системите кои типично работат при низок SNR поради двојниот ефект близу-далеку [110].

Дизајнираме безжично напојуван би-модален (бит-семантички) комуникациски систем со 1-битен повратен канал, кој се состои од би-модален предавател на информации кој собира енергија (анг. energy harvesting information transmitter) (EHIT) и би-модален приемник на информации интегриран со РФ предавател, наречен информациски приемник кој емитува енергија (анг. energy broadcasting information receiver) (EBIR).



Слика 5.7: Блок-шема на безжично напојуван бит-семантички комуникациски систем.

Разгледуваме точка-точка систем со еден EHIT и еден EBIR (Слика 5.7), кој го користи протоколот „собери-па-пренеси“. Системското време t е поделено на рамки со времетраење

$T = \tau_{EH}(t) + \tau_{IT}(t)$. Безжичниот канал помеѓу ЕНІТ и ЕВІР е реципрочен и изложен на квази-статичен фединг; моменталното засилување во рамка t е $h_1(t)$, со просечна вредност $\Omega_1 = \mathbb{E}[h_1(t)]$. Клучна претпоставка е асиметријата во познавањето на CSI – ЕВІР совршено го познава каналот, додека ЕНІТ нема достапно CSI. Врз основа на $h_1(t)$, ЕВІР одлучува дали системот ќе биде тивок (во длабок фединг) или активен, а ако е активен, работи или во BitCom или во SemCom-режим.

Ако системот е активен, во текот на ЕН-фазата ЕВІР зрачи РФ енергија до ЕНІТ со фиксна моќност P_0 . Количината на собрана енергија се одредува со:

$$E_{EH}(t) = \eta P_0 \tau_{EH}(t) h_1(t), \quad (5.50)$$

каде η е ефикасноста на конверзија на енергија. Оваа енергија се троши во ИТ-фазата, кога ЕНІТ пренесува информации со излезна моќност $p_1(t)$. Односот сигнал-шум на ЕВІР е:

$$\gamma(t) = p_1(t) h_1(t) / \sigma_N^2, \quad (5.51)$$

каде σ_N^2 е моќноста на термичкиот шум пред ЕВІР, а $E_{IT}(t) = p_1(t) \tau_{IT}(t)$.

SemCom-режим. ЕНІТ користи семантички предавател, а ЕВІР семантички приемник, претходно тренирани за различни услови на каналот. Нека ЕНІТ прима реченица од L зборови на својот влез, со s_1 семантички симболи по збор, а I ја означува просечната количина на семантички информации по реченица. Достижната семантичка информациска брзина е:

$$R_S = \frac{IB}{s_1 L} \frac{\tau_{IT,S}}{T} \xi(\gamma), \quad (5.52)$$

каде $\tau_{IT,S}/T$ е делот од времето искористен за пренос во SemCom-режим. За дадена вредност на каналот h_1 , R_S може да се максимизира со:

$$\begin{array}{ll} \text{Maximize} & R_S \\ \tau_{EH,S}, \tau_{IT,S}, p_{1,S} & \end{array} \quad (5.53)$$

$$\text{под услов: } C1 : \tau_{EH,S} + \tau_{IT,S} = T \quad (5.54)$$

$$C2 : p_{1,S} \tau_{IT,S} \leq \eta \tau_{EH,S} P_0 h_1 \quad (5.55)$$

$$C3 : \xi \left(\frac{p_{1,S} h_1}{\sigma_N^2} \right) \geq \xi_{th} \quad (5.56)$$

Теорема 5.3. Нека a_0 го означува односот сигнал-шум дефиниран како:

$$a_0 = \frac{\eta P_0 h_1^2}{\sigma_N^2}. \quad (5.57)$$

Нека μ_0 е коренот на равенката

$$\xi(\mu_0) = \xi_{th}, \quad \text{односно} \quad \mu_0 = \xi^{-1}(\xi_{th}). \quad (5.58)$$

Нека γ_0 е коренот на следнава трансцендентна равенка:

$$\xi(\gamma_0) = \xi'(\gamma_0) (\gamma_0 + a_0), \quad (5.59)$$

каде $\xi'(\cdot)$ е првиот извод на $\xi(\cdot)$. Решението на (5.59) постои само ако a_0 надминува одреден праг a_{th} ; во спротивно, $\gamma_0 \rightarrow \infty$. Оптималното времетраење на ИТ-фазата е:

$$\tau_{IT,S}^* = \frac{a_0 T}{a_0 + \max(\mu_0, \gamma_0)}, \quad (5.60)$$

оптималното времетраење на ЕН-фазата е:

$$\tau_{EH,S}^* = T - \tau_{IT,S}^*, \quad (5.61)$$

а оптималната предавателна моќност е:

$$p_{1,S}^* = \frac{\eta P_0 h_1 \tau_{EH,S}^*}{\tau_{IT,S}^*}. \quad (5.62)$$

Оптималната остварлива рата е дадена со:

$$R_S^* = \begin{cases} \frac{IB}{s_1 L} \frac{a_0 \xi(\gamma_0)}{a_0 + \gamma_0}, & \text{ако } \gamma_0 > \mu_0 \\ \frac{IB}{s_1 L} \frac{a_0 \xi_{th}}{a_0 + \gamma_0}, & \text{ако } \gamma_0 \leq \mu_0 \end{cases}. \quad (5.63)$$

Доказот на Теорема 5.3 е даден во додаток Б.9.

BitCom-режим. ЕНІТ користи конвенционален предавател, а ЕВІР конвенционален приемник, со строга еднаквост помеѓу влезовите и излезите. Достижната рата, трансформирана во еквивалентна остварлива семантичка информациска брзина според [38], е:

$$R_B = \frac{IB}{z_1 L} \frac{\tau_{IT,B}}{T} \log_2(1 + \gamma), \quad (5.64)$$

каде z_1 е просечниот број пренесени бити по збор. Слично, R_B се максимизира со:

$$\underset{\tau_{EH,B}, \tau_{IT,B}, p_{1,B}}{\text{Maximize}} \quad R_B \quad (5.65)$$

$$\text{под услов: } C1 : \tau_{EH,B} + \tau_{IT,B} = T \quad (5.66)$$

$$C2 : p_{1,B} \tau_{IT,B} \leq \eta \tau_{EH,B} P_0 h_1 \quad (5.67)$$

Врз основа на фундаменталното решение предложено во [110], оптималното времетраење на ИТ-фазата е:

$$\tau_{IT,B}^* = \frac{a_0 T}{a_0 - 1} \left(1 + \frac{1}{W_0 \left(\frac{a_0 - 1}{e} \right)} \right)^{-1}, \quad (5.68)$$

каде $W_0(\cdot)$ е главната гранка на Ламбертовата W -функција, а $\tau_{EH,B}^* = T - \tau_{IT,B}^*$ и $p_{1,B}^* = \eta P_0 h_1 \tau_{EH,B}^* / \tau_{IT,B}^*$. Максимизираната еквивалентна рата е:

$$R_B^* = \frac{IB}{z_1 L} \frac{1}{\log(2)} \frac{a_0}{a_0 - 1} W_0 \left(\frac{a_0 - 1}{e} \right). \quad (5.69)$$

Во пореалистично сценарио каде системот оперира во канал со случаен фединг и EBIR прецизно го следи $a_0(t)$, EBIR може да имплементира полиса за избор на режим:

$$M(t) = \begin{cases} \text{SemCom}, & \text{ако } R_S^*(t) > R_B^*(t) \\ \text{BitCom}, & \text{ако } R_S^*(t) \leq R_B^*(t) \end{cases}, \quad (5.70)$$

по што EBIR го информира EHIT преку 1-битен повратен канал. Целта е да се одреди минималната заедничка остварлива семантичка (еквивалентна) рата, R_0 , со цел да се максимизира системскиот проток дефиниран како $\rho_0 = R_0 \cdot \Pr\{\max(R_S^*, R_B^*) \geq R_0\}$ [109]:

$$\rho_0^* = \underset{R_0 \geq 0}{\text{maximize}} R_0 \cdot \Pr\{\max(R_S^*, R_B^*) \geq R_0\}. \quad (5.71)$$

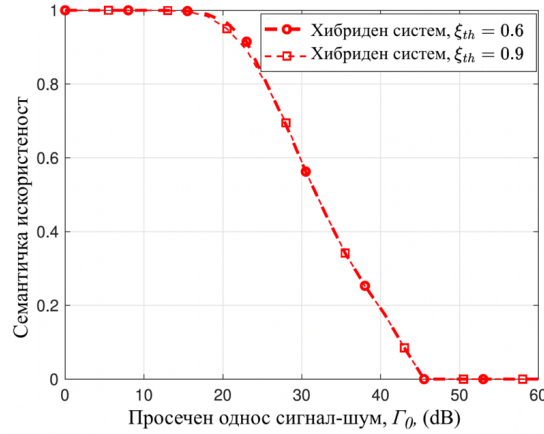
Оваа веројатност се одредува преку Монте Карло симулации, користејќи ги индикаторските функции $I_1(t)$ (BitCom доминира и го надминува R_0) и $I_2(t)$ (SemCom доминира и го надминува R_0). Кога каналот $h_1(t)$ е во длабок фединг, $I_1(t) = I_2(t) = 0$, што резултира со испад и за двата режима.

Пред почетокот на комуникациската сесија, се врши следнава еднакратна конфигурација: (1) конвенционалниот пар кодер/декодер избира кодна книга способна да ја постигне оптималната рата R_0^* ; (2) семантичкиот пар кодер/декодер е претходно трениран еднаш за даден опсег на SNR-вредности; (3) бидејќи R_S^* и R_B^* се монотонно растечки функции од a_0 , се определуваат фиксните времетраења на EH и IT-фазите за секој режим. Овие времетраења остануваат фиксни во текот на целата сесија, гарантирајќи проток ρ_0 .

За анализа на перформансите, претпоставуваме Рејлиев фединг канал, пропусен опсег $B = 10$ MHz, конвенционален пренос со $z_1 = 32$ бити/збор, семантички пренос со $s_1 = 5$ семантички симболи/збор, и параметри на логистичката функција $A_1 = 0.37$, $A_2 = 0.98$, $C_1 = 0.2525$, $C_2 = -0.7985$ [106]. Перформансите на хибридниот систем се споредуваат со два референтни моно-модални системи:

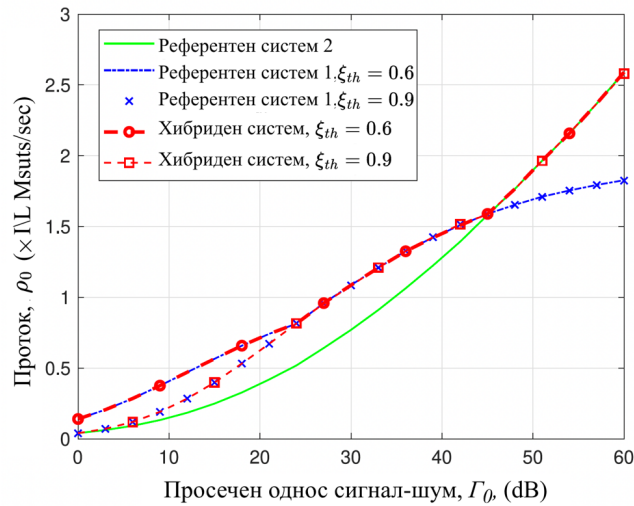
$$\rho_{0,S}^* = \underset{R_{0,S} \geq 0}{\text{maximize}} R_{0,S} \cdot \Pr\{R_S^* \geq R_{0,S}\}, \quad (5.72)$$

$$\rho_{0,B}^* = \underset{R_{0,B} \geq 0}{\text{maximize}} R_{0,B} \cdot \Pr\{R_B^* \geq R_{0,B}\}. \quad (5.73)$$



Слика 5.8: Семантичка искористеност наспроти просечен однос сигнал-шум.

Слика 5.8 ја прикажува семантичката искористеност на предложениот хибриден систем, дефинирана како дел од времето кога системот работи во SemCom-режим, во однос на вкупното време кога не е во испад. Во режимот на низок SNR, системот постигнува максимална семантичка искористеност, што потврдува дека хибридниот систем работи речиси исклучиво во SemCom-режим – фундаменталната предност на SemCom, способноста да зачува значење дури и во услови на силен шум. Како што се зголемува просечниот SNR, Γ_0 , семантичката искористеност постепено се намалува до нула на одреден праг Γ_{th} , што покажува дека системот се префрла на BitCom кога каналот станува доволно добар.



Слика 5.9: Максимален системски проток наспроти просечен однос сигнал-шум.

Слика 5.9 го прикажува максималниот системски проток, ρ_0 , за хибридниот, референтниот SemCom и референтниот BitCom-систем. Клучниот заклучок е дека предложената политика за избор на режим (5.70) гарантира дека хибридниот систем го постигнува протокот на супериорниот моно-модален референтен систем за дадениот SNR: при низок и умерен SNR ($\Gamma_0 < \Gamma_{th}$), максималниот проток на хибридниот систем совршено се совпаѓа со оној на моно-модалниот SemCom-систем, во согласност со резултатите во [36]; при висок SNR

($\Gamma_0 > \Gamma_{th}$), максималниот проток се совпаѓа со моно-модалниот BitCom-систем. Прагот на семантичка сличност ξ_{th} влијае на протокот при низок SNR, а овој ефект се намалува со зголемување на SNR. Резултатите јасно покажуваат дека едноставната, 1-битна политика за избор на режим му овозможува на хибридниот систем беспрекорно да ја следи горната ан-велопа на перформансите на двата специјализирани системи – тој не е компромис помеѓу двата, туку го постигнува најдоброто од двете шеми.

Заклучоци и придонеси

Оваа докторска дисертација систематски ги истражи и значајно придонесе кон напредните стратегии за распределба на ресурси во интелигентните безжични комуникациски системи од следната генерација. Со развивање на рамки за распределба на ресурси кои ги комбинираат реконфигурабилната контрола на безжичниот медиум, рабното пресметување, федеративното учење и семантичките комуникации, ова истражување понуди сеопфатни решенија за фундаменталните предизвици поврзани со распределба на мрежните перформанси и енергетската ефикасност во високо динамични безжични средини со ограничени ресурси.

Сите научни придонеси изнесени во оваа дисертација се поддржани со математички докази кои се систематски презентирани во додаток Б. За секоја предложена оптимизациска шема и аналитичко решение, развиени се детални математички докази кои ја потврдуваат точноста на изведените изрази и оптималноста на предложените алгоритми. Овие докази опфаќаат неколку клучни аспекти: прво, тие ја гарантираат математичката точност на трансформациите и апроксимациите користени при изведувањето на решенија во затворена форма; второ, тие ја докажуваат конвергенцијата на итеративните алгоритми кон оптимални решенија; и трето, тие ја потврдуваат ефикасноста на алгоритмите во смисла на пресметковната комплексност и бројот на итерации потребни за конвергенција.

Дел I: Интелигентно управување со радиооколината во безжични комуникации

Истражувањето започна со анализа на интелигентниот физички слој, каде реконфигурабилните интелигентни површини и напредните релејни системи овозможуваат активно обликување на радиооколината. Во рамките на овој дел, дисертацијата реализира *развој и детален дизајн на иновативен повеќеантенски релеен систем кој работи во режим на истовремен прием и праќање, базиран на 2-битни аналогни фазни поместувачи и пасивно поништување на интерференцијата*. Овој придонес е од суштинско значење бидејќи ги надминува ограничувањата на идеалните реконфигурабилни интелигентни површини кои бараат совршена фазна контрола и претпоставуваат отсуство на хардверски ограничувања.

Посебен акцент беше ставен на компаративна анализа помеѓу идеални пасивни реконфигурабилни интелигентни површини и предложениот едноставен повеќеантенско двонасочно реле. Оваа детална споредба откри дека предложениот релеен систем, кој вклучува 2-битна фазна квантизација и пасивно поништување на интерференција, *конзистентно ги надминува идеалните пасивни рефлектирачки површини во однос на остварливите информациски брзини и енергетската ефикасност*. Овие резултати се особено изразени во сценарија со Реј-

лиев фединг, каде рефлектирачките површини значително губат на ефикасност поради статистичката природа на каналот, додека релејниот систем со активно засилување ја задржува својата супериорност. Квантитативните резултати покажаа дека при мал број на антенски елементи, релејниот систем постигнува значително поголеми податочни брзини, што е од клучно значење за практична имплементација каде комплексноста и цената мора да бидат ограничени.

Значајниот придонес од овој дел лежи во *изведувањето на затворени аналитички решенија за перформансите на овие системи, овозможувајќи ефикасна онлајн имплементација со мала пресметковна комплексност*. Аналитичките изрази за информациските брзини и енергетската ефикасност, изведени низ детална математичка анализа, обезбедуваат директен увид во системските компромиси помеѓу бројот на антенски елементи, предавателната моќност, растојанието и карактеристиките на каналот. Дополнително, *развиената рамка за енергетска ефикасност*, која го зема предвид целокупниот енергетски буџет, покажа дека предложениот релеен систем постигнува супериорна енергетска ефикасност во споредба со идеалните реконфигурабилни површини, особено при присуство на директна компонента на линија на видливост. Овие наоди сугерираат дека предложените релејни системи претставуваат поефикасно и поизводливо решение за 6G-паметни радиооколин, обезбедувајќи практична алтернатива на теоретски идеалните пасивни рефлектирачки површини.

Дел II: Пресметување и учење на работ на мрежата во безжични комуникации

Во доменот на дистрибуирана рабна интелигенција, дисертацијата пионерски разви напредни оптимизациски рамки за безжично напојувани системи за пресметување на работ на мрежата. Овој дел од истражувањето адресира една од најкритичните предизвици во современите безжични мрежи: *како да се постигне енергетска одржливост за уредите со ограничени ресурси кои истовремено мора да извршуваат интензивни пресметковни задачи*. Посебен фокус беше ставен на постигнување енергетска автономија преку интеграција со безжичен пренос на енергија, што претставува клучен предуслов за одржлива работа на уредите со ограничени ресурси во контекстот на рабното пресметување.

Дисертацијата развива *иновативна повеќеслојна оптимизациска шема која заеднички управува со комуникациските ресурси (пропусен опсег, предавателна моќност), пресметковните фреквенции на локалните уреди, и ресурсите за собирање на енергија*. Оваа сеопфатна рамка за оптимизација надминува традиционалните пристапи кои третираат секој вид ресурс одделно, и ги признава комплексните меѓузависности помеѓу енергетското напојување, комуникацијата и пресметувањето. Резултатите од нумеричките симулации покажаа дека заедничката оптимизација доведува до значителни подобрувања во споредба со изолираната оптимизација на поединечни ресурси.

Во рамките на овој домен, дисертацијата понуди два значајни придонеси. Првиот е *развој на оптимизациска рамка базирана на max-min критериум која ја максимизира минимал-*

ната пресметковна рата меѓу сите корисници, обезбедувајќи правичност во системот со безжично напојувани уреди. Развиениот алгоритам ефикасно ја адресира неконвексноста на оптимизацискиот проблем преку серија трансформации, доведувајќи до ефикасно решение кое може да се имплементира во реално време.

Вториот придонес е *развој на оптимизациска шема базирана на критериум на пропорционална правичност која ја максимизира сумата од логаритмите на пресметковните рати на сите корисници*. Овој пристап обезбедува балансиран компромис помеѓу целокупната системска ефикасност и правичноста кон поединечните корисници. Резултатите покажаа дека оваа шема постигнува повисок индекс на правичност во однос на алтернативните пристапи, додека истовремено одржува конкурентни вкупни пресметковни рати. Понатаму, дисертацијата развива *алгоритми за оптимална распределба на режимите на претовар (бинарен и делумен)*, кои динамички одлучуваат колку од податоците ќе се обработуваат локално на самиот уред и колку ќе се пренесат на базната станица за обработка.

Дисертацијата прави значаен чекор напред во областа на безжично федеративно учење со развој на сеопфатни оптимизациски рамки кои ги адресираат специфичните предизвици на дистрибуираното машинско учење во безжични средини со ограничени ресурси. Централниот придонес во овој домен е *развојот на иновативна оптимизациска рамка за системи со федеративно учење која заеднички оптимизира комуникациски параметри (распределба на каналски ресурси и предавателна моќност), пресметковни параметри (фреквенција на процесорот), и параметри на машинското учење (број на локални епохи и глобални рунди на тренирање)* за минимизирање на вкупното време на конвергенција при гарантирана целна точност на моделот.

Традиционалните пристапи кон федеративното учење обично третираат одделно комуникациски и пресметковни аспекти, што резултира со субоптимални перформанси. Напротив, *развиената повеќеслојна оптимизација во оваа дисертација го потенцира фактот дека изборот на параметрите за машинско учење (како бројот на локални епохи) директно влијае на комуникацискиот товар, додека достапните комуникациски и пресметковни ресурси диктираат што е остварливо во рамките на алгоритмот за учење*.

Дисертацијата понуди два комплементарни пристапи за ефикасна распределба на ресурси во федеративно учење. Првиот е *развој на оптимизациска шема базирана на неортогонален повеќекратен пристап (NOMA) која овозможува истовремен пренос на ажурирањата на моделите од повеќе уреди, значително намалувајќи го комуникациското доцнење*. Развиениот алгоритам интелегентно ја распределува предавателната моќност меѓу уредите со цел да се минимизира доцнењето за најспориот уред. Нумеричките резултати покажаа дека оваа шема постигнува значително намалување на вкупното доцнење во споредба со ортогоналните методи за пристап, особено кога бројот на уреди е голем.

Вториот пристап претставува *развој на оптимизациска рамка базирана на временски повеќекратен пристап (TDMA) која заеднички оптимизира временска распределба, предавателна моќност, локални пресметковни фреквенции, и параметри на алгоритмот за учење за постигнување минимално време на конвергенција*. Клучниот технички придонес овде е развојот на ефикасен двослоен алгоритам кој ја разделува сложената неконвексна оптими-

зација на два потпроблеми. Особено значаен резултат е дека *постои нетривијална и силна меѓузависност помеѓу бројот на локални епохи на тренирање и достапните комуникациски ресурси* – кога комуникацискиот пропусен опсег е ограничен, оптималната стратегија е да се зголеми бројот на локални епохи, додека при доволни комуникациски ресурси, пократките локални циклуси со почести глобални ажурирања се поефикасни.

Клучен придонес во овој дел е исто така *интеграцијата на безжичен пренос на енергија во рамката за федеративно учење, овозможувајќи одржлива и самодоволна работа на уредите со ограничени батерии*. Резултатите покажаа дека оптималната распределба на времето помеѓу фазите на собирање енергија, локално тренирање и комуникација води до значителни подобрувања во енергетската ефикасност и намалување на вкупното време на конвергенција.

Дел III: Семантички комуникации

Во областа на семантичките комуникации, оваа дисертација дава суштински придонес со претставување софистицирани стратегии за оптимизација на ресурси кои го ставаат на преден план преносот на значење наместо сирови бити. Фундаменталниот придонес во овој дел е развој на *оптимизациска рамка за минимизација на вкупната потрошувачка на енергија при гарантирано ниво на семантичка сличност помеѓу изворната и примената информација*. Оваа рамка заеднички ја оптимизира предавателната моќност, распределбата на пропусниот опсег, и димензиите на семантичките репрезентации. Резултатите од нумеричките експерименти покажаа дека *семантичките комуникации постигнуваат значителна енергетска ефикасност, особено во режими со низок однос сигнал-шум*, каде традиционалните бит-базирани комуникации бараат прекумерна енергија за одржување на прифатливо ниво на грешки.

Препознавајќи ги комплементарните предности на семантичките и традиционалните бит-базирани комуникации, дисертацијата предлага *иновативен хибриден бит-семантички комуникациски систем кој динамички се префрла помеѓу двата режима на пренос врз основа на моменталните услови на каналот*. Развиениот алгоритам за интелигентен избор на режим користи едноставна 1-битна политика базирана на праговна вредност на односот сигнал-шум, што овозможува ефикасна имплементација со минимална пресметковна сложеност.

Дополнително, дисертацијата развива *сеопфатна оптимизациска рамка за хибридниот систем која заеднички го оптимизира изборот на режим, распределбата на моќноста, распределбата на пропусниот опсег и димензиите на семантичките репрезентации со цел максимизација на вкупниот системски проток*. Нумеричките резултати покажаа дека хибридниот систем постигнува перформанси кои ја следат горната анvelopa на двата специјализирани моно-модални системи, што значи дека тој постигнува најдобрите можни перформанси во целиот спектар на канални услови. Посебно значаен е резултатот што *прагот на семантичка сличност има значително влијание врз протокот при низок однос сигнал-шум, но ова влијание се намалува со подобрување на каналните услови*.

Насоки за идни истражувања

Гледајќи кон иднината, истражувањата би можеле да се прошират во неколку ветувачки и практично применливи насоки. Еден од клучните правци е понатамошен развој на децентрализираните алгоритми за ФУ, специјално оптимизирани за ресурсно-ограничени уреди (TinyML), со длабока интеграција на SemCom и RIS. Ова би вклучило дизајнирање на шеми за ФУ кои овозможуваат ефикасно тренирање и ажурирање на моделите директно на самиот раб на мрежата. RIS би играле клучна улога во динамичното подобрување на безжичните комуникациски врски помеѓу овие TinyML-уреди и централните агрегатори, додека SemCom би овозможил ултраефикасен пренос на ажурирањата на модели, значително намалувајќи го доцнењето и зголемувајќи ја енергетската ефикасност во многу големи, децентрализираните мрежи.

Понатаму, од суштинско значење е проширувањето на апликациите на овие интелигентни системи во специфични 6G-сценарија, со особен фокус на практична имплементација во виртуелна реалност (VR) и проширена реалност (AR). Во овие контексти, доцнењето е критично за корисничкото искуство, а нашите решенија би можеле да обезбедат ниско доцнење и висока пропусност. Ова исто така вклучува развој на механизми за дигитални близнаци (анг. digital twins) и напредни индустриски IoT-системи, каде преносот на семантички информации и рабното учење може да ја оптимизираат работата на комплексни физички системи во реално време, предвидувајќи грешки и подобрувајќи ја ефикасноста.

Друга важна област е понатамошното истражување на сигурноста и приватноста. Ова подразбира развој на напредни криптографски техники кои работат на семантичко ниво, со цел да се заштити самото значење на информациите, наместо само битовите. Имплементацијата на диференцијална приватност е исто така клучна за заштита на податоците од сензитивни информации кои може да се извлечат од семантичките репрезентации или од ажурирањата на моделите за ФУ, особено кога се работи со лични или критични податоци.

На крајот, и покрај длабоката теоретска анализа и симулации, развојот на експериментални платформи и прототипови со хардверски имплементации е клучен за валидација на предложените алгоритми во реални безжични средини. Ова би можело да вклучи користење на FPGA или SDR (анг. Software-Defined Radio) платформи за имплементација на RIS-контролери, MEC-сервери и семантички кодери/декодери. Ваквите експерименти ќе овозможат да се потврдат теоретските предвидувања, да се идентификуваат и решат практичните предизвици и компромиси кои не можат целосно да бидат опфатени со симулации, со што ќе се зацврстат основите за реализација на целосно интелигентни, самоодржливи и безбедни 6G-комуникациски мрежи.

Додаток А

Основи на конвексна оптимизација

Овој додаток нуди детален преглед на теоријата на конвексна оптимизација користена во оваа дисертација. Конвексната оптимизација претставува специјална класа на проблеми на математичка оптимизација каде целните функции и функциите на ограничувања исполнуваат специфични својства кои овозможуваат ефикасно изнаоѓање на глобално оптимално решение.

Проблемот на математичка оптимизација ја има следнава стандардна форма:

$$\begin{aligned} & \underset{\mathbf{x}}{\text{Minimize}} && f_0(\mathbf{x}) \\ & \text{под услов:} && f_i(\mathbf{x}) \leq b_i, \quad i = 1, \dots, m. \end{aligned} \tag{A.1}$$

Тука, векторот $\mathbf{x} = (x_1, \dots, x_n)$ е оптимизациска променлива, функцијата $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$ е целната функција, а функциите $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ се функции на ограничувања. Векторот $\mathbf{x}^*$ е оптимален ако за секој изводлив вектор $\mathbf{z}$ важи $f_0(\mathbf{z}) \geq f_0(\mathbf{x}^*)$. Кај нелинеарното програмирање, овие функции можат да бидат произволни, што често води до проблеми со повеќе локални минимуми. Сепак, кај конвексната оптимизација, се гарантира дека секој локален оптимум е воедно и глобален. Дополнително, основниот оптимизациски проблем дефиниран во (A.1) може да го прошириме со додавање на ограничувања изразени со еднаквост:

$$\begin{aligned} & \underset{\mathbf{x}}{\text{Minimize}} && f_0(\mathbf{x}) \\ & \text{под услов:} && f_i(\mathbf{x}) \leq b_i, \quad i = 1, \dots, m, \\ & && h_i(\mathbf{x}) = 0, \quad i = 1, \dots, p. \end{aligned} \tag{A.2}$$

Сега, можеме да дефинираме конвексен оптимизациски проблем во неговата стандардна форма:

$$\begin{aligned} & \underset{\mathbf{x} \in \mathcal{D}}{\text{Minimize}} && f_0(\mathbf{x}) \\ & \text{под услов:} && f_i(\mathbf{x}) \leq b_i, \quad i = 1, \dots, m, \\ & && h_i(\mathbf{x}) = b_i, \quad i = 1, \dots, p, \end{aligned} \tag{A.3}$$

каде $\mathcal{D}$ е конвексно множество, $f_0, f_1, \dots, f_m$ се конвексни функции, додека $h_i(\mathbf{x}) = \mathbf{a}_i^T \mathbf{x} + b_i$ се афини функции. За да можеме да ја разбереме формата на овој оптимизациски проблем,

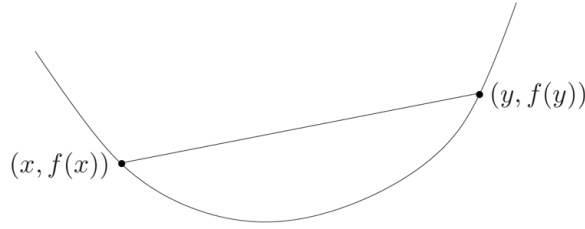
ќе дадеме неколку основни дефиниции во продолжение.

Дефиниција А.1 (Конвексно множество): Множеството C е конвексно ако за секои $x_1, x_2 \in C$ и за секое $0 \leq \theta \leq 1$, важи $\theta x_1 + (1 - \theta)x_2 \in C$.

Дефиниција А.2 (Конвексна функција): Функцијата $f : \mathbb{R}^d \rightarrow \mathbb{R}$ е конвексна функција ако: (1) доменот на функцијата $\text{dom}(f)$ е конвексно множество, и (2) за секои x и $y \in \text{dom}(f)$ и за секое $0 \leq \theta \leq 1$ важи

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y). \quad (\text{A.4})$$

Геометриски, ова неравенство значи дека тетивата меѓу било кои две точки $(x, f(x))$ и $(y, f(y))$ лежи над самиот график на f , како што е прикажано на Слика А.1. Функцијата f е строго конвексна ако важи строго неравенство во (А.4). За функцијата f велиме дека е конкавна ако $-f$ е конвексна, и строго конкавна ако $-f$ е строго конвексна.



Слика А.1: Геометриска репрезентација на конвексна функција.

Покрај оваа дефиниција, постојат уште две дефиниции кои ги карактеризираат конвексните функции.

Дефиниција А.3 (Диференцијабилност): Ако f е диференцијабилна, таа е конвексна ако и само ако:

$$f(y) \geq f(x) + \nabla f(x)^T(y - x), \quad \forall x, y \in \text{dom}(f). \quad (\text{A.5})$$

Ова значи дека линеарната апроксимација во која било точка е секогаш долна граница на функцијата.

Дефиниција А.4 (Двојна диференцијабилност): Ако f е двапати диференцијабилна, таа е конвексна ако и само ако нејзиниот Хесијан е позитивно полуdefинитен:

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in \text{dom}(f). \quad (\text{A.6})$$

Во областа на оптимизацијата, поимот „мазност“ (анг. smoothness) има специфично значење. Функцијата f е β -мазна ако нејзиниот градиент е Липшиц-континуиран (анг. Lipschitz continuous) со параметар β , односно за кои било $x, y \in \text{dom}(f)$ важи:

$$\|\nabla f(x) - \nabla f(y)\| \leq \beta \|x - y\|. \quad (\text{A.7})$$

Постојат неколку клучни импликации од својството на мазност:

1. Ако функцијата f е β -мазна, тогаш функцијата $g(x) = \frac{\beta}{2}\|x\|^2 - f(x)$ е конвексна.

Вообичаено, не би се очекувало дека $-f(x)$ е конвексна функција (освен во случај кога f е афина функција).

- Друга значајна импликација на мазноста е дека таа наметнува квадратно ограничување одозгора на функцијата f . Имено, ако f е β -мазна, тогаш за сите x, y важи:

$$f(y) \leq f(x) + \nabla f(x)^T(y - x) + \frac{\beta}{2} \|y - x\|^2. \quad (\text{A.8})$$

За да се интерпретира ова својство, се фиксира одредена точка x : додека конвексните функции секогаш лежат над нивните тангенти, мазните конвексни функции секогаш лежат под параболата која поминува низ точката $(x, f(x))$.

- Конечно, доколку функцијата f е двапати диференцијабилна, својството на β -мазност е еквивалентно на условот дека нејзиниот Хесијан е ограничен одозгора:

$$\nabla^2 f(x) \preceq \beta I_d. \quad (\text{A.9})$$

Иако не постои општа аналитичка формула за решавање на конвексните оптимизациски проблеми, тие можат ефикасно да се решат со користење на методите на внатрешна точка. Овие методи постигнуваат висока точност во мал број итерации (обично помеѓу 10 и 100), независно од големината на проблемот. Како што истакнува Бојд [91], предизвикот не лежи во самото решавање, туку во „уметноста“ на препознавање и формулирање на практичните проблеми како конвексни.

Лагранжовата дуалност овозможува решавање на примарниот проблем преку неговиот дуален пандан, што често овозможува полесно изведување на решенија во затворена форма. Основната идеја кај Лагранжовата дуалност е земање предвид на ограничувањата во проблемот (A.3) преку проширување на целната функција со тежинска сума од функциите на ограничувањата. Лагранжијанот $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ поврзан со проблемот (A.3) го дефинираме како:

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i (a_i^T x - b_i), \quad (\text{A.10})$$

со дефинициона област $\text{dom } \mathcal{L} = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$. Променливата λ_i ја дефинираме како Лагранжов множител поврзан со i -тото ограничување во вид на неравенство $f_i(x) \leq 0$; аналогно, ν_i го дефинираме како Лагранжов множител поврзан со i -тото ограничување во вид на равенство $h_i(x) = 0$. Векторите λ и ν се нарекуваат дуални променливи или вектори на Лагранжови множители поврзани со проблемот (A.3).

Лагранжовата дуална функција (или само дуална функција) $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ ја дефинираме како минималната вредност на Лагранжијанот по x , за $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$:

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right). \quad (\text{A.11})$$

Кога Лагранжијанот е неограничен оддолу по x , дуалната функција ја зема вредноста $-\infty$. Бидејќи дуалната функција претставува pointwise infimum на фамилија афини функции по (λ, ν) , таа е конкавна, дури и во случаи кога изворниот проблем (A.3) не е конвексен.

Додаток Б

Докази

Б.1 Доказ на Теорема 2.1

Нека каналот помеѓу изворот и приемната страна на двонасочниот релеј е претставен со векторот

$$\mathbf{h}_S = [h_{S,1}, h_{S,2}, \dots, h_{S,M}]^T, \quad (\text{Б.1})$$

каде $h_{S,m}$ го означува каналот помеѓу изворот и m -тата приемна антена на релејот. Нека каналот помеѓу предавателната страна на двонасочниот релеј и дестинацијата е претставен со векторот

$$\mathbf{h}_D = [h_{D,1}, h_{D,2}, \dots, h_{D,N}]^T, \quad (\text{Б.2})$$

каде $h_{D,n}$ го означува каналот помеѓу n -тата предавателна антена на релејот и дестинацијата. Каналите $h_{S,m}$ и $h_{D,n}$ се моделираат како:

$$h_{S,m} = \sqrt{\Omega_{S,m}} e^{j\phi_{S,m}}, \quad (\text{Б.3})$$

$$h_{D,n} = \sqrt{\Omega_{D,n}} e^{j\phi_{D,n}}, \quad (\text{Б.4})$$

каде $\Omega_{S,m}$ и $\phi_{S,m}$ ги означуваат каналното засилување и фазата на каналот помеѓу изворот и m -тата приемна антена на релејот, додека $\Omega_{D,n}$ и $\phi_{D,n}$ ги означуваат каналното засилување и фазата на каналот помеѓу n -тата предавателна антена на релејот и дестинацијата. Во опсегот на милиметарски бранови, каналот се состои од една многу силна директна (LoS) компонента и неколку многу слаби рефлектирани (non-LoS) компоненти. Затоа, секој канал може да се моделира како статичен канал со фиксни $\Omega_{S,m}$, $\Omega_{D,n}$, $\phi_{D,n}$ и $\phi_{S,m}$. Претпоставуваме дека изворот и дестинацијата се наоѓаат во далечното поле на релејот, како што е дефинирано во [111], и затоа соодветните добивки на моќност на каналот се дадени со:

$$\Omega_{S,m} = \Omega_S, \forall m, \quad (\text{Б.5})$$

$$\Omega_{D,n} = \Omega_D, \forall n. \quad (\text{Б.6})$$

Бидејќи двонасочниот релеј истовремено предава и прима сигнал на иста фреквенција, сигналот што го испраќа предавателната антена на релејот се „враќа“ назад во неговиот сопствен приемник, создавајќи сопствена интерференција помеѓу приемната и предавателната антенска низа на релејот. Нека $\mathbf{G} = [g_{mn}]_{M \times N}$ ја означува матрицата на каналот на

сопствена интерференција, каде g_{mn} е каналот на сопствена интерференција помеѓу n -тата предавателна антена и m -тата приемна антена на релејот. Како што е прикажано во [112], најлошото сценарио во однос на достижната брзина на пренос е кога $g_{mn}, \forall m, \forall n$ се независни и идентично распределени комплексни Гаусови случајни променливи со нулта средина:

$$g_{mn} \sim \mathcal{CN}(0, \sigma_I^2), \quad (\text{Б.7})$$

каде σ_I^2 е моќноста на каналот на сопствена интерференција. Под претпоставка дека предавателната моќност на релејот, P_R , е еднакво распределена помеѓу N -те предавателни антени, просечната примена моќност кај релејот како резултат на сопствената интерференција, P_I , е дадена со:

$$P_I = \frac{P_R}{N} \mathbb{E} \left\{ \left| \sum_{m=1}^M \sum_{n=1}^N g_{mn} \right|^2 \right\} = \frac{P_R}{N} MN \sigma_I^2 = MP_R \sigma_I^2. \quad (\text{Б.8})$$

Поради законот за зачувување на енергијата, мора да важи $P_I \leq P_R$. Со комбинирање на (Б.8) со овој услов се добива:

$$MP_R \sigma_I^2 \leq P_R \implies M \sigma_I^2 \leq 1 \implies \sigma_I^2 \leq \frac{1}{M}. \quad (\text{Б.9})$$

Претпоставуваме дека е вметната пасивна изолација помеѓу предавателната и приемната антенска низа на релејот со коефициент на пасивна изолација $\eta \leq 1$. Во овој случај, (Б.9) се трансформира во $\sigma_I^2 \leq \eta/M$. Типичните вредности за коефициентот на пасивна изолација варираат помеѓу 10^{-6} и 10^{-5} [113]. Во остатокот од анализата поставуваме дека σ_I^2 е точно на границата, односно:

$$\sigma_I^2 = \frac{\eta}{M}, \quad (\text{Б.10})$$

што води кон најниската можна остварлива брзина на пренос на предложениот двонасочно релеен систем.

Нека P_S ја означува предавателната моќност на изворот, а P_R предавателната моќност на релејот, каде $P_S + P_R = P_T$. Нека x_S го означува сигналот пренесен од изворот, така што $\mathbb{E}\{|x_S|^2\} = P_S$. Нека $\bar{\mathbf{y}}_R = [\bar{y}_{R,1}, \bar{y}_{R,2}, \dots, \bar{y}_{R,M}]^T$ го означува векторот на примениот сигнал кај релејот пред да се примени какво било фазно поместување, составен од: (i) сигналот од изворот преку каналот $\mathbf{h}_S$; (ii) векторот на шум кај релејот, $\mathbf{w}_R$; и (iii) векторот на сопствена интерференција.

Нека x_R го означува информацискиот сигнал на релејот, $\mathbb{E}\{|x_R|^2\} = P_R$, а $\mathbf{v} = [v_1, v_2, \dots, v_N]^T$ векторот на фазно поместување на предавателната антенска низа, каде $v_n \in \mathcal{P}$. Векторот на сопствена интерференција кај релејот е $x_R \mathbf{G} \mathbf{v} / \sqrt{N}$, а примениот вектор е:

$$\bar{\mathbf{y}}_R = x_S \mathbf{h}_S + \frac{1}{\sqrt{N}} x_R \mathbf{G} \mathbf{v} + \mathbf{w}_R, \quad (\text{Б.11})$$

каде $\mathbf{w}_R = [w_{R,1}, \dots, w_{R,M}]^T$, со $w_{R,m} \sim \mathcal{CN}(0, N_0)$. Нека $\mathbf{u} = [u_1, \dots, u_M]^T$, $u_m \in \mathcal{P}$, го означува векторот на фазно поместување на приемната антенска низа. Примениот сигнал кај

релејот е:

$$y_R = \mathbf{u}^T \bar{\mathbf{y}}_R = x_S \mathbf{u}^T \mathbf{h}_S + \frac{1}{\sqrt{N}} x_R \mathbf{u}^T \mathbf{G} \mathbf{v} + \mathbf{u}^T \mathbf{w}_R. \quad (\text{Б.12})$$

Примениот сигнал кај дестинацијата е:

$$y_D = \frac{1}{\sqrt{N}} x_R \mathbf{v}^T \mathbf{h}_D + w_D, \quad (\text{Б.13})$$

каде $w_D \sim \mathcal{CN}(0, N_0)$.

Од (Б.12), односот сигнал-интерференција-и-шум (SINR) на приемната страна на релејот, γ_S , е:

$$\gamma_S = \frac{\mathbb{E}_{x_S} \{ |x_S \mathbf{u}^T \mathbf{h}_S|^2 \}}{\mathbb{E}_{\mathbf{w}_R} \{ |\mathbf{u}^T \mathbf{w}_R|^2 \} + \mathbb{E}_{x_R, \mathbf{G}} \left\{ \left| \frac{1}{\sqrt{N}} x_R \mathbf{u}^T \mathbf{G} \mathbf{v} \right|^2 \right\}}. \quad (\text{Б.14})$$

Броителот е $\mathbb{E}_{x_S} \{ |x_S \mathbf{u}^T \mathbf{h}_S|^2 \} = P_S |\mathbf{u}^T \mathbf{h}_S|^2$. За голем број антени M , изразот $|\mathbf{u}^T \mathbf{h}_S|^2$ може да се запише како:

$$\begin{aligned} |\mathbf{u}^T \mathbf{h}_S|^2 &= \left| \sum_{m=1}^M \sqrt{\Omega_S} e^{j\bar{\phi}_{S,m}} \right|^2 \\ &= M\Omega_S + M(M-1)\Omega_S \times \sum_{m=1}^M \sum_{j=1, j \neq m}^M \frac{e^{j\bar{\phi}_{S,m}} e^{-j\bar{\phi}_{S,j}}}{M(M-1)} \\ &= M\Omega_S + M(M-1)\Omega_S \times \mathbb{E}_m \{ e^{j\bar{\phi}_{S,m}} \} \mathbb{E}_j \{ e^{-j\bar{\phi}_{S,j}} \} \\ &\stackrel{(b)}{=} M\Omega_S + M(M-1)\Omega_S \left(\frac{2^b}{\pi} \sin \left(\frac{\pi}{2^b} \right) \right)^2, \end{aligned} \quad (\text{Б.15})$$

каде $\bar{\phi}_{S,m} = \phi_{S,m} - \hat{\phi}_{S,m}$, а чекорот (b) произлегува од фактот дека $\bar{\phi}_{S,m}$ и $\bar{\phi}_{S,j}$ имаат униформни распределби во интервалот $[-\pi/2^b, \pi/2^b]$. Просечната моќност на сопствената интерференција е:

$$\mathbb{E}_{x_R, \mathbf{G}} \left\{ \left| \frac{1}{\sqrt{N}} x_R \mathbf{u}^T \mathbf{G} \mathbf{v} \right|^2 \right\} \stackrel{(c)}{=} \frac{P_R}{N} \mathbb{E}_{\mathbf{G}} \{ |\mathbf{u}^T \mathbf{G} \mathbf{v}|^2 \}, \quad (\text{Б.16})$$

каде (c) произлегува од $\mathbb{E}_{x_R} \{ |x_R|^2 \} = P_R$. Очекувањето $\mathbb{E}_{\mathbf{G}} \{ |\mathbf{u}^T \mathbf{G} \mathbf{v}|^2 \}$ се добива како:

$$\begin{aligned} \mathbb{E}_{\mathbf{G}} \{ |\mathbf{u}^T \mathbf{G} \mathbf{v}|^2 \} &= \sum_{m=1}^M \sum_{n=1}^N \mathbb{E}_{\mathbf{G}} \{ |g_{mn}|^2 \} \\ &\quad + \sum_{m=1}^M \sum_{n=1}^N \sum_{i \neq m, i=1}^M \sum_{j \neq n, j=1}^N \mathbb{E}_{\mathbf{G}} \{ g_{nm} \} \mathbb{E}_{\mathbf{G}} \{ g_{ij} \} \times e^{-j\hat{\phi}_{S,m} \hat{\phi}_{S,i}} e^{-j\hat{\phi}_{D,n} \hat{\phi}_{D,j}} \\ &= \sum_{m=1}^M \sum_{n=1}^N \mathbb{E}_{\mathbf{G}} \{ |g_{mn}|^2 \} = MN\sigma_I^2 \stackrel{(d)}{=} N\eta, \end{aligned} \quad (\text{Б.17})$$

каде (d) важи кога е исполнет условот за најлошо сценарио, односно $\sigma_I^2 = \eta/M$. Со вметну-

вање на (Б.17) во (Б.16) добиваме:

$$\mathbb{E}_{x_R, \mathbf{G}} \left\{ \left| \frac{1}{\sqrt{N}} x_R \mathbf{u}^T \mathbf{G} \mathbf{v} \right|^2 \right\} = P_R \eta. \quad (\text{Б.18})$$

Имајќи ги предвид (Б.18), (Б.15) и моќноста на шумот $\mathbb{E}_{\mathbf{w}_R} \{ |\mathbf{u}^T \mathbf{w}_R|^2 \} = M N_0$, изразот за γ_S станува:

$$\gamma_S = \frac{P_S \left(\Omega_S + (M-1) \Omega_S \left(\frac{2^b}{\pi} \sin \left(\frac{\pi}{2^b} \right) \right)^2 \right)}{N_0 + \frac{\eta}{M} P_R}. \quad (\text{Б.19})$$

Од (Б.13), односот сигнал-шум кај дестинацијата, γ_D , е:

$$\gamma_D = \frac{\mathbb{E}_{x_R} \left\{ \left| \frac{1}{\sqrt{N}} x_R \mathbf{v}^T \mathbf{h}_D \right|^2 \right\}}{\mathbb{E}_{w_D} \{ w_D^2 \}} \stackrel{(e)}{=} \frac{P_R |\mathbf{v}^T \mathbf{h}_D|^2}{N N_0}, \quad (\text{Б.20})$$

каде (e) произлегува од $\mathbb{E}\{|x_R|^2\} = P_R$. Изразот $|\mathbf{v}^T \mathbf{h}_D|^2$ се добива на ист начин како во (Б.15), со замена на N со M и Ω_D со Ω_S :

$$\gamma_D = \frac{P_R \left(\Omega_D + (N-1) \Omega_D \left(\frac{2^b}{\pi} \sin \left(\frac{\pi}{2^b} \right) \right)^2 \right)}{N_0}. \quad (\text{Б.21})$$

Конечно, постижната брзина на пренос е дадена со:

$$R = \min \{ \log_2(1 + \gamma_S), \log_2(1 + \gamma_D) \}, \quad (\text{Б.22})$$

каде γ_S и γ_D се дадени во (Б.19) и (Б.21), соодветно, што директно води до финалниот израз во Теорема 2.1. ■

Б.2 Доказ на Тврдење 2.1

Според [57], со цел да се максимизира брзината на пренос кај релејот, потребно е да биде исполнет следниот услов:

$$\frac{P_S \Omega_S (1 + (M-1)Q)}{N_0 + \frac{\eta}{M} P_R} = \frac{P_R \Omega_D (1 + (N-1)Q)}{N_0}. \quad (\text{Б.23})$$

Бидејќи M и N се големи броеви, ги усвојуваме апроксимациите $1 + (M-1)Q \approx MQ$ и $1 + (N-1)Q \approx NQ$, кои во тие услови важат со голема точност. Оттука, (Б.23) се трансформира во:

$$\frac{P_S \Omega_S M Q}{N_0 + \frac{\eta P_R}{M}} = \frac{P_R \Omega_D N Q}{N_0}. \quad (\text{Б.24})$$

Со означување на $a = \Omega_S / \Omega_D$ и $\eta_0 = \eta / N_0$, од (Б.24) се добива:

$$\eta_0 N P_R^2 + M N P_R - a M^2 P_S = 0. \quad (\text{Б.25})$$

За да го максимизираме односот сигнал-шум-и-интерференција кај релејот, го формираме следниов оптимизациски проблем:

$$\begin{aligned}
 & \underset{P_S, P_R, M, N}{\text{Maximize}} && P_R N \\
 \text{под услов:} &&& C1 : \eta_0 N P_R^2 + M N P_R - a M^2 P_S = 0 \\
 &&& C2 : P_R + P_S = P_T \\
 &&& C3 : M + N = K
 \end{aligned} \tag{Б.26}$$

Го користиме методот на Лагранжови множители за решавање на (Б.26). Лагранжовата функција е:

$$\mathcal{L} = P_R N + \lambda_1 (\eta_0 N P_R^2 + M N P_R - a M^2 P_S) + \lambda_2 (P_R + P_S - P_T) + \lambda_3 (M + N - K). \tag{Б.27}$$

Ги наоѓаме изводите на $\mathcal{L}$ во однос на P_R , P_S , M и N и ги изедначуваме со нула:

$$\frac{\partial \mathcal{L}}{\partial P_R} = N + 2\lambda_1 N \eta_0 P_R + \lambda_1 M N + \lambda_2 = 0, \tag{Б.28}$$

$$\frac{\partial \mathcal{L}}{\partial N} = P_R + \lambda_1 P_R^2 \eta_0 + M P_R \lambda_1 + \lambda_3 = 0, \tag{Б.29}$$

$$\frac{\partial \mathcal{L}}{\partial M} = \lambda_1 N P_R - 2\lambda_1 a M P_S + \lambda_3 = 0, \tag{Б.30}$$

$$\frac{\partial \mathcal{L}}{\partial P_S} = -\lambda_1 a M^2 + \lambda_2 = 0. \tag{Б.31}$$

Нека $a_{11} = N(2\eta_0 P_R + M)$, $a_{21} = P_R^2 \eta_0 + M P_R$, $a_{31} = N P_R - 2a M P_S$ и $a_{41} = -a M^2$. Равенките (Б.28)–(Б.31) можат да се запишат во матрична форма:

$$\begin{bmatrix} a_{11} & 1 & 0 \\ a_{21} & 0 & 1 \\ a_{31} & 0 & 1 \\ a_{41} & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} -N \\ -P_R \\ 0 \\ 0 \end{bmatrix}. \tag{Б.32}$$

Системот (Б.32) се состои од четири равенки со три непознати. Бидејќи претпоставуваме дека постои единствено решение, четирите равенки мора да бидат линеарно зависни, па детерминантата на системот мора да биде нула, што бара:

$$(a_{11} - a_{41})P_R - (a_{21} - a_{31})N = 0, \tag{Б.33}$$

односно:

$$(a_{11} - a_{41})P_R = (a_{21} - a_{31})N. \tag{Б.34}$$

Со замена на вредностите за a_{11} , a_{21} , a_{31} и a_{41} во (Б.34), се формира равенката:

$$\eta_0 N P_R^2 + (a M^2 + N^2 + 2a M N)P_R - 2a M N P_T = 0. \tag{Б.35}$$

Од друга страна, со замена на $P_S = P_T - P_R$ во условот $C1$ од (Б.26) добиваме:

$$\eta_0 N P_R^2 + (aM^2 + MN)P_R - aM^2 P_T = 0. \quad (\text{Б.36})$$

Со дефинирање на $\alpha = \eta_0 N$, $b_1 = aM^2 + N^2 + 2aMN$, $c_1 = 2aMNP_T$, $b_2 = aM^2 + MN$ и $c_2 = aM^2 P_T$, равенките (Б.35) и (Б.36) се запишуваат како:

$$\alpha P_R^2 + b_1 P_R - c_1 = 0, \quad (\text{Б.37})$$

$$\alpha P_R^2 + b_2 P_R - c_2 = 0. \quad (\text{Б.38})$$

Корените на (Б.37) и (Б.38) се:

$$P_{R,1} = \frac{-b_1 + \sqrt{4\alpha c_1}}{2\alpha}, \quad (\text{Б.39})$$

$$P_{R,2} = \frac{-b_2 + \sqrt{4\alpha c_2}}{2\alpha}. \quad (\text{Б.40})$$

Под претпоставка дека $b_1^2 \ll 4\alpha c_1$ и $b_2^2 \ll 4\alpha c_2$ (поради постоењето на η_0 во α), (Б.39) и (Б.40) се сведуваат на $P_{R,1} = \sqrt{4\alpha c_1}/(2\alpha)$ и $P_{R,2} = \sqrt{4\alpha c_2}/(2\alpha)$. Бидејќи $P_{R,1}$ и $P_{R,2}$ треба да бидат еднакви, добиваме $c_1 = c_2$, што резултира со:

$$M = 2N. \quad (\text{Б.41})$$

Земајќи го предвид условот $C3$ во (Б.26), од (Б.41) го добиваме оптималното M , означено со M^* :

$$M^* = \frac{2K}{3}. \quad (\text{Б.42})$$

Со замена на (Б.42) во изразот за $P_{R,1}$, ја добиваме оптималната вредност за P_R :

$$P_R^* = \sqrt{\frac{4K\Omega_S N_0 P_T}{3\Omega_D \eta}}. \quad (\text{Б.43})$$

Од условите $C2$ и $C3$, ги добиваме оптималните вредности за N и P_S :

$$N^* = K - M^* = \frac{K}{3}, \quad (\text{Б.44})$$

$$P_S^* = P_T - P_R^*. \quad (\text{Б.45})$$

Оваа апроксимација е валидна кога $b_1^2 \ll 4\alpha c_1$ и $b_2^2 \ll 4\alpha c_2$. Еден проблем со оптималниот резултат добиен за P_R во (Б.43) е тоа што кога K достигнува многу големи вредности, P_R^* станува поголемо од P_T , што е физички невозможно. За да го решиме овој проблем, ги користиме веќе изведените вредности $M^* = 2K/3$ и $N^* = K/3$ како почетна точка за максимизирање на брзината на пренос. Со замена на M^* и N^* во (Б.23), се доаѓа до условот:

$$\frac{P_S \Omega_S \left(1 + \left(\frac{2K}{3} - 1\right) Q\right)}{N_0 + \frac{\eta}{M} P_R} = \frac{P_R \Omega_D \left(1 + \left(\frac{K}{3} - 1\right) Q\right)}{N_0}. \quad (\text{Б.46})$$

Со воведување на апроксимациите $1 + (2K/3 - 1)Q \approx 2KQ/3$ и $1 + (K/3 - 1)Q \approx KQ/3$, кои стануваат сè попречици со зголемување на K , се добива:

$$\frac{P_S \Omega_S \left(\frac{2K}{3}Q\right)}{N_0 + \frac{\eta}{M}P_R} = \frac{P_R \Omega_D \left(\frac{K}{3}Q\right)}{N_0}. \quad (\text{Б.47})$$

Со замена на P_S со $P_T - P_R$ и решавање на квадратната равенка која произлегува од еднаквоста во (Б.47) во однос на P_R , ги добиваме вредностите за P_R^* и P_S^* дадени во равенките (2.12) и (2.13) од Тврдење 2.1. ■

Б.3 Доказ на Последица 2.1

Нека фединг каналот помеѓу изворот S и m -тата приемна антена на релејот биде означен со $h_{S,m}$, за кој се претпоставува дека е комплексна Гаусова случајна променлива со нулта средина и варијанса $\Omega_{S,m}$, односно $h_{S,m} \sim \mathcal{CN}(0, \Omega_{S,m})$. Нека фединг каналот помеѓу n -тата предавателна антена на релето и дестинацијата D биде означен со $h_{D,n} \sim \mathcal{CN}(0, \Omega_{D,n})$. Оттука, фединг каналите можат да се изразат како:

$$h_{S,m} = \alpha_m e^{j\phi_{S,m}}, \quad h_{D,n} = \beta_n e^{j\phi_{D,n}}, \quad (\text{Б.48})$$

каде α_m и β_n се амплитудите на каналите, кои се Рејлиеви случајни променливи со густини:

$$f_{\alpha_m}(x) = \frac{2x}{\Omega_{S,m}} \exp\left(-\frac{x^2}{\Omega_{S,m}}\right), \quad x \geq 0, \quad (\text{Б.49})$$

$$f_{\beta_n}(x) = \frac{2x}{\Omega_{D,n}} \exp\left(-\frac{x^2}{\Omega_{D,n}}\right), \quad x \geq 0. \quad (\text{Б.50})$$

Оттука, $\mathbb{E}\{\alpha_m^2\} = \Omega_{S,m}$, $\mathbb{E}\{\alpha_m\} = \sqrt{\Omega_{S,m}\pi/4}$, и слично за β_n . Поради претпоставката за далечно поле, $\Omega_{S,m} = \Omega_S, \forall m$ и $\Omega_{D,n} = \Omega_D, \forall n$. Од (Б.12), SINR кај релејот, γ_S^{Ra} , е:

$$\gamma_S^{Ra} \stackrel{(f)}{=} \frac{P_S \mathbb{E}_{\mathbf{h}_S} \{|\mathbf{u}^T \mathbf{h}_S|^2\}}{\mathbb{E}_{\mathbf{w}_R} \{|\mathbf{u}^T \mathbf{w}_R|^2\} + P_R \mathbb{E}_{\mathbf{G}} \left\{ \left| \frac{1}{\sqrt{N}} \mathbf{u}^T \mathbf{G} \mathbf{v} \right|^2 \right\}}, \quad (\text{Б.51})$$

каде (f) произлегува од $\mathbb{E}_{x_s} \{|x_s|^2\} = P_S$ и $\mathbb{E}_{x_R} \{|x_R|^2\} = P_R$. Очекувањето во броителот се пресметува како:

$$\mathbb{E} \{|\mathbf{u}^T \mathbf{h}_S|^2\} = \mathbb{E} \left\{ \left| \sum_{m=1}^M \alpha_m e^{j\phi_{S,m}} e^{-j\hat{\phi}_{S,m}} \right|^2 \right\} = \mathbb{E} \left\{ \left| \sum_{m=1}^M \alpha_m e^{j\bar{\phi}_{S,m}} \right|^2 \right\}, \quad (\text{Б.52})$$

што дополнително се поедноставува како:

$$\begin{aligned}
 \mathbb{E} \{ |\mathbf{u}^T \mathbf{h}_S|^2 \} &= \mathbb{E} \left\{ \sum_{m=1}^M \alpha_m^2 \right\} + \mathbb{E} \left\{ \sum_{m=1}^M \sum_{j=1, j \neq m}^M \alpha_m \alpha_j e^{j\bar{\phi}_{S,m}} e^{-j\bar{\phi}_{S,j}} \right\} \\
 &= \sum_{m=1}^M \mathbb{E} \{ \alpha_m^2 \} + \sum_{m=1}^M \sum_{j=1, j \neq m}^M \mathbb{E} \{ \alpha_m \} \mathbb{E} \{ \alpha_j \} \mathbb{E} \{ e^{j\bar{\phi}_{S,m}} \} \mathbb{E} \{ e^{-j\bar{\phi}_{S,j}} \} \\
 &= M\Omega_S \left(1 + (M-1) \frac{\pi}{4} Q \right), \tag{Б.53}
 \end{aligned}$$

каде $\bar{\phi}_{S,m} = \phi_{S,m} - \hat{\phi}_{S,m}$ е грешката (шум) од квантизација воведена од b -битните фазни поместувачи, при што важи $\mathbb{E} \{ e^{j\bar{\phi}_{S,m}} \} = \mathbb{E} \{ e^{-j\bar{\phi}_{S,m}} \} = \frac{2^b}{\pi} \sin \left(\frac{\pi}{2^b} \right) = \sqrt{Q}$. Оттука, γ_S^{Ra} е даден со:

$$\gamma_S^{Ra} = \frac{P_S \Omega_S \left(1 + (M-1) \frac{\pi}{4} Q \right)}{N_0 + \frac{\eta}{M} P_R}. \tag{Б.54}$$

Слично, примениот SINR кај дестинацијата, γ_D^{Ra} , е:

$$\gamma_D^{Ra} = \frac{P_R \mathbb{E}_{h_D} \left\{ \left| \frac{1}{\sqrt{N}} \mathbf{v}^T \mathbf{h}_D \right|^2 \right\}}{\mathbb{E}_{w_D} \{ w_D^2 \}}, \tag{Б.55}$$

каде единствениот член кој е различен од (Б.20) е математичкото очекување во броителот. По слични изведувања како во (Б.18), го добиваме:

$$\gamma_D^{Ra} = \frac{P_R \Omega_D \left(1 + (N-1) \frac{\pi}{4} Q \right)}{N_0}. \tag{Б.56}$$

Од (Б.54) и (Б.56) заклучуваме дека R^{Ra} е дадена со изразот во (2.20), со што доказот на Последица 2.1 е завршен. ■

Б.4 Доказ на Теорема 3.1

Со воведување на заменска променлива $U_k = P_k \tau_k$ во оптимизацискиот проблем (3.13), го добиваме следниов оптимизациски проблем:

$$\text{Maximize}_{\tau_0, \tau_i, U_i, f_i} \quad \min_{1 \leq i \leq K} \left\{ \frac{T f_i}{L} + B \tau_i \log_2 \left(1 + \frac{U_i \Omega_i}{\tau_i N_0} \right) \right\} \tag{Б.57}$$

$$\text{под услов:} \quad C1 : \sum_{i=1}^K \tau_i + \tau_0 = T \tag{Б.58}$$

$$C2 : \alpha T f_i^3 + U_i \leq \eta \tau_0 P_0 \Omega_i, \forall i \tag{Б.59}$$

$$C3 : \tau_i \geq 0, \tau_0 \geq 0, P_i \geq 0, f_i \geq 0, \forall i \tag{Б.60}$$

Ако претпоставиме сценарио со симетрични корисници, каде сите ЕНУ-станции се поставени на кружница со радиус r , оптимизацискиот проблем (Б.57) се поедноставува како:

$$\underset{\tau_0, \tau_1, U_1, f_1}{\text{Maximize}} \quad \frac{Tf_1}{L} + B\tau_1 \log_2 \left(1 + \frac{U_1\Omega_1}{\tau_1 N_0} \right) \quad (\text{Б.61})$$

$$\text{под услов:} \quad C1 : K\tau_1 + \tau_0 = T \quad (\text{Б.62})$$

$$C2 : \alpha T f_1^3 + U_1 \leq \eta \tau_0 P_0 \Omega_1 \quad (\text{Б.63})$$

$$C3 : \tau_1 \geq 0, \tau_0 \geq 0, P_1 \geq 0, f_1 \geq 0 \quad (\text{Б.64})$$

За да го решиме овој проблем, ја формираме Лагранжовата функција $\mathcal{L}(\tau_0, \tau_1, U_1, f_1)$:

$$\begin{aligned} \mathcal{L}(\tau_0, \tau_1, U_1, f_1) = & \frac{Tf_1}{L} + B\tau_1 \log_2 \left(1 + \frac{U_1\Omega_1}{\tau_1 N_0} \right) - \lambda_1 (K\tau_1 + \tau_0 - T) \\ & - \lambda_2 (\alpha T f_1^3 + U_1 - \eta \tau_0 P_0 \Omega_1), \end{aligned} \quad (\text{Б.65})$$

каде λ_1 и λ_2 се соодветните Лагранжови множители. Земајќи го парцијалниот извод по τ_0, τ_1, U_1 и f_1 и изедначувајќи ги со нула, добиваме:

$$\lambda_1 = \lambda_2 \eta P_0 \Omega_1, \quad (\text{Б.66})$$

$$K\lambda_1 = B \log_2 \left(1 + \frac{U_1\Omega_1}{\tau_1 N_0} \right) - B \frac{U_1\Omega_1}{\ln 2 (\tau_1 N_0 + U_1\Omega_1)}, \quad (\text{Б.67})$$

$$\lambda_2 = B\tau_1 \frac{\Omega_1}{\ln 2 (\tau_1 N_0 + U_1\Omega_1)}, \quad (\text{Б.68})$$

$$f_1 = \left(\frac{1}{3\lambda_2 \alpha L} \right)^{1/2}, \quad (\text{Б.69})$$

$$K\tau_1 + \tau_0 = T, \quad (\text{Б.70})$$

$$\alpha T f_1^3 + U_1 = \eta \tau_0 P_0 \Omega_1. \quad (\text{Б.71})$$

Со комбинирање на (Б.66), (Б.67) и (Б.68), го добиваме изразот:

$$\frac{K\eta P_0 \Omega_1^2}{\ln 2 (N_0 + P_1^* \Omega_1)} + \frac{P_1^* \Omega_1}{\ln 2 (N_0 + P_1^* \Omega_1)} - \log_2 \left(1 + \frac{P_1^* \Omega_1}{N_0} \right) = 0. \quad (\text{Б.72})$$

Изразот (3.17) се добива со примена на дефиницијата на Ламбертовата W -функција [79] врз (Б.72). Со вметнување на оптималната предавателна моќност, P_1^* , пресметана во (3.17), во (Б.68), можеме да ја пресметаме оптималната вредност на Лагранжовиот множител λ_2 и да ја замениме во (Б.69), од каде го изведуваме изразот (3.18). Потоа, со користење на оптималните вредности P_1^* и f_1^* и со комбинирање на (Б.70) и (Б.71), стигнуваме до изразот (3.20). ■

Б.5 Доказ на Теорема 4.1

Воведувајќи нова оптимизациска променлива T_0 , операторот $\min$ во целната функција на (4.11) се конвертира во дополнително ограничување, $C5$, што го трансформира (4.11) во:

$$\underset{t_0, p_k, f_k}{\text{Minimize}} \quad T_0 \quad (\text{Б.73})$$

$$\text{под услов:} \quad C1 : \frac{t_0}{k} B \log_2 \left(1 + \sum_{i=1}^k x_i p_i \right) = b_0, \forall k \quad (\text{Б.74})$$

$$C2 : p_k \leq P_{\max}, \forall k \quad (\text{Б.75})$$

$$C3 : \alpha a_0 f_k^2 + p_k t_0 \leq E_{\max}, \forall k \quad (\text{Б.76})$$

$$C4 : f_k \leq f_{\max}, \forall k \quad (\text{Б.77})$$

$$C5 : t_0 + \frac{a_0}{f_k} \leq T_0, \forall k \quad (\text{Б.78})$$

Го обработуваме ограничувањето $C1$ со негово препишување за два последователни индекси, $k-1$ и k :

$$1 + \sum_{i=1}^k p_i x_i = e^{z(k-1)}, \quad (\text{Б.79})$$

$$1 + \sum_{i=1}^{k-1} p_i x_i + p_k x_k = e^{zk}, \quad (\text{Б.80})$$

каде $z = c_0/t_0$, а $c_0 = b_0 \ln(2)/B$. Одземајќи го (Б.79) од (Б.80) добиваме:

$$p_k = \frac{e^{zk} - e^{z(k-1)}}{x_k}. \quad (\text{Б.81})$$

Со вметнување на (Б.81) во $C2$ и $C3$, (Б.73) се трансформира во конвексен оптимизациски проблем:

$$\underset{T_0, z, f_k}{\text{Minimize}} \quad T_0 \quad (\text{Б.82})$$

$$\text{под услов:} \quad C2' : \frac{e^{zk} - e^{z(k-1)}}{x_k} \leq P_{\max}, \forall k \quad (\text{Б.83})$$

$$C3' : \alpha a_0 f_k^2 + \frac{c_0}{x_k} \frac{e^{zk} - e^{z(k-1)}}{z} \leq E_{\max}, \forall k \quad (\text{Б.84})$$

$$C4' : f_k \leq f_{\max}, \forall k \quad (\text{Б.85})$$

$$C5' : \frac{c_0}{z} + \frac{a_0}{f_k} \leq T_0, \forall k \quad (\text{Б.86})$$

Функциите $(e^{zk} - e^{z(k-1)})$ и $(e^{zk} - e^{z(k-1)})/z$ во $C2'$ и $C3'$ се конвексни, што ја потврдува конвексноста на (Б.82). Решението се добива со спојување на решенијата на два специјални случаи: кога $E_{\max} \rightarrow \infty$ (случај 1) и кога $P_{\max} \rightarrow \infty$ (случај 2).

Специјален случај 1: $E_{\max} \rightarrow \infty$. Во овој случај, (Б.82) се редуцира на:

$$\underset{z}{\text{Minimize}} \quad \frac{c_0}{z} \quad (\text{Б.87})$$

$$\text{под услов: } C2' : \frac{e^{zk} - e^{z(k-1)}}{x_k} \leq P_{\max}, \forall k \quad (\text{Б.88})$$

За секое k , левата страна на $C2'$ е монотонно растечка во однос на z , па $C2'$ е еквивалентно на $z \leq z_{0k}, \forall k$, каде z_{0k} е решение на (4.18). Проблемот (Б.87) има решение ако $z \leq z^*$, каде:

$$z^* = \min\{z_{01}, z_{02}, \dots, z_{0K}\}. \quad (\text{Б.89})$$

Ограничувањето $C2'$ е задоволено со строго равенство само за една безжична станица, додека останатите $K - 1$ станици го задоволуваат со строги неравенства. Земајќи го предвид (Б.89), оптималната предавателна моќност на безжичните уреди се одредува како (4.16).

Специјален случај 2: $P_{\max} \rightarrow \infty$. Со воведување на смената $t_0 = c_0/g$, (Б.82) се трансформира во:

$$\underset{T_0, g, f_k}{\text{Minimize}} \quad T_0 \quad (\text{Б.90})$$

$$\text{под услов: } C3' : \alpha a_0 f_k^2 + \frac{c_0}{x_k} \frac{e^{gk} - e^{g(k-1)}}{g} \leq E_{\max}, \forall k \quad (\text{Б.91})$$

$$C4' : f_k \leq f_{\max}, \forall k \quad (\text{Б.92})$$

$$C5' : \frac{c_0}{g} + \frac{a_0}{f_k} \leq T_0, \forall k \quad (\text{Б.93})$$

За $k \geq 1$ и $g \geq 0$, функцијата $(e^{gk} - e^{g(k-1)})/g$ е поголема од 1 и монотонно расте во однос на g и k . Проблемот (Б.90) евозможен само ако $x_k E_{\max}/c_0 > 1, \forall k$. Нека k^* го означува индексот на безжичната станица која го задоволува $C2'$ со строго равенство (а со тоа и $C4'$ и $C5'$), додека кај останатите станици ограничувањата се задоволени со неравенства. Лагранжовата функција за овој проблем со само 3 ограничувања во форма на равенство е:

$$\begin{aligned} \mathcal{L} = T_0 + \lambda \left(\frac{e^{gk^*} - e^{g(k^*-1)}}{g} - \frac{x_{k^*}(E_{\max} - \alpha a_0 f_{k^*}^2)}{c_0} \right) \\ + \gamma \left(\frac{a_0}{f_{k^*}} + \frac{c_0}{g} - T_0 \right) + \mu (f_{k^*} - f_{\max}), \end{aligned} \quad (\text{Б.94})$$

каде λ, γ, μ се позитивни Лагранжови множители. Изедначувајќи ги изводите на $\mathcal{L}$ по T_0, g и f_{k^*} со нула:

$$\frac{d\mathcal{L}}{dT_0} = 1 - \gamma = 0, \quad (\text{Б.95})$$

$$\frac{d\mathcal{L}}{dg} = -\frac{\gamma c_0}{g^2} + \lambda \left(\frac{k^* e^{k^* g} - (k^* - 1) e^{(k^* - 1)g}}{g} - \frac{e^{k^* g} - e^{(k^* - 1)g}}{g^2} \right) = 0, \quad (\text{Б.96})$$

$$\frac{d\mathcal{L}}{df_{k^*}} = -\frac{\gamma a_0}{f_{k^*}^2} + \lambda \frac{2\alpha a_0 f_{k^*} x_{k^*}}{c_0} + \mu = 0. \quad (\text{Б.97})$$

Од (Б.95), $\gamma = 1$, што потврдува дека $C3'$ за k^* е задоволено со строго равенство. Втората равенка се трансформира во:

$$\lambda = \frac{c_0 e^{g(1-k^*)}}{1 + g(1 - k^*) - e^g + k^* g e^g} \geq 0, \quad (\text{Б.98})$$

позитивно за секое $g > 0$ и $k \geq 0$, со што се потврдува дека $C2'$ за k^* е задоволено со строго равенство. Со вметнување на (Б.98) во (Б.97):

$$\mu = \frac{a_0}{f_{k^*}^2} - \frac{2\alpha a_0 f_{k^*} x_{k^*} e^{g(1-k^*)}}{1 + g(1 - k^*) - e^g + k^* g e^g} \geq 0. \quad (\text{Б.99})$$

Сценарио 1: Ако $f_{k^*} < f_{\max}$, тогаш $\mu = 0$. Со изедначување на (Б.99) со нула, и поставување на $C2'$ со равенство, се добива системот равенки (4.19)–(4.20). Решението (g_{0k^*}, f_{0k^*}) е бараното оптимално решение ако $f_{0k^*} < f_{\max}$; во спротивно се применува Сценарио 2.

Сценарио 2: Ако $f_{k^*} = f_{\max}$, тогаш $\mu > 0$. Оптималната вредност на g се одредува директно од (4.20), под услов:

$$2\alpha f_{\max}^3 x_{k^*} e^{g_{0k^*}(1-k^*)} < 1 + g_{0k^*}(1 - k^*) - e^{g_{0k^*}} + k^* g_{0k^*} e^{g_{0k^*}}. \quad (\text{Б.100})$$

Ако е задоволен, оптималното решение е $(g_{0k^*}, f_{\max})$. Конечно, g^* се определува како:

$$g^* = \min\{g_{01}, g_{02}, \dots, g_{0K}\}. \quad (\text{Б.101})$$

Бидејќи или $C2'$ или $C3'$ можат да бидат активни во (4.11), општото решение во Теорема 4.1 се добива со спојување на (Б.89) и (Б.101), што го дава z^* дефиниран во (4.17). ■

Б.6 Доказ на Теорема 4.2

Лагранжовата функција на конвексниот проблем (4.42) е дадена со:

$$\begin{aligned} \mathcal{L} = & \left(\tau_{EH} + \tau_{LC} + \sum_{k=1}^K t_k \right) \\ & + \sum_{k=1}^K \lambda_k (e_k + \alpha a_0(\eta) f_k^2 - \xi P_0 \tau_{EH} h_k) \\ & + \sum_{k=1}^K \mu_k \left(c_0 - t_k \log \left(1 + \frac{e_k h_k}{t_k B N_0} \right) \right) \\ & + \sum_{k=1}^K \gamma_k (f_k - f_{\max}) + \sum_{k=1}^K \beta_k \left(\frac{a_0(\eta)}{f_k} - \tau_{LC} \right), \end{aligned} \quad (\text{Б.102})$$

каде $\lambda_k, \mu_k, \gamma_k$ и β_k се ненегативни Лагранжови множители поврзани со ограничувањата $C1', C2', C3'$ и $C4'$ во (4.42), соодветно. Ги поставуваме изводите на $\mathcal{L}$ во однос на $\tau_{LC}, \tau_{EH}, f_k, t_k$ и e_k на нула:

$$\frac{\partial \mathcal{L}}{\partial \tau_{LC}} = 1 - \sum_{k=1}^K \beta_k = 0, \quad (\text{Б.103})$$

$$\frac{\partial \mathcal{L}}{\partial \tau_{EH}} = 1 - \xi P_0 \sum_{k=1}^K \lambda_k h_k = 0, \quad (\text{Б.104})$$

$$\frac{\partial \mathcal{L}}{\partial f_k} = 2\alpha a_0(\eta) f_k \lambda_k + \gamma_k - \beta_k \frac{a_0(\eta)}{f_k^2} = 0, \quad (\text{Б.105})$$

$$\frac{\partial \mathcal{L}}{\partial t_k} = 1 - \mu_k \left[\log \left(1 + \frac{e_k h_k}{t_k B N_0} \right) - \frac{e_k h_k}{t_k B N_0 + e_k h_k} \right] = 0, \quad (\text{Б.106})$$

$$\frac{\partial \mathcal{L}}{\partial e_k} = \lambda_k - \mu_k \frac{t_k B N_0 h_k}{t_k B N_0 + e_k h_k} \frac{1}{t_k B N_0} = 0. \quad (\text{Б.107})$$

Да претпоставиме дека множителите λ_k, μ_k и β_k се строго позитивни $\forall k$. Во овој случај, ограничувањата $C1', C2'$ и $C4'$ во (4.42) мора да бидат задоволени со строги равенства.

Случај 1: $\gamma_k = 0, \forall k$. Во овој случај, $f_k = f_0 < f_{\max}, \forall k$. Комбинирајќи ги $C1'$ и $C2'$, и користејќи ја смената $e_k = p_k t_k$, добиваме:

$$\frac{c_0 p_k}{\log \left(1 + \frac{p_k h_k}{B N_0} \right)} + \alpha a_0(\eta) f_0^2 = \xi P_0 \tau_{EH} h_k. \quad (\text{Б.108})$$

Равенката (Б.108) има општа форма $p_k = b \cdot \log(1 + a p_k)$, каде $ab > 1$, чие решение во затворена форма е $p_k = -a^{-1} - b \cdot W_{-1}(-(ab)^{-1} \cdot \exp(-1/ab))$, каде $W_{-1}(\cdot)$ ја означува втората гранка на Ламбертовата W -функција. Оттука, решението на (Б.108) е дадено со (4.47). Оптималната вредност на τ_{LC} се добива од $C4'$, што резултира со (4.53), додека оптималната вредност на t_k се добива од $C2'$, што резултира со (4.54). Понатаму, со вметнување на (4.47) во (Б.107) стигнуваме до изразот (4.52). Врз основа на (4.52), равенката (Б.104) се трансфор-

мира во:

$$\sum_{k=1}^K h_k g_2(\tau_{EH}, f_0) = \frac{1}{\xi P_0}, \quad (\text{Б.109})$$

каде $g_2(\tau_{EH}, f_0)$ е дефинирана со (4.52). Вметнувајќи ја (Б.109) во (Б.102), и користејќи ги (4.52) и (4.47), се добива:

$$\sum_{k=1}^K g_2(\tau_{EH}, f_0) = \frac{1}{2\alpha f_0^3}. \quad (\text{Б.110})$$

Равенките (Б.109) и (Б.110) се всушност системот од две равенки дадени со (4.50)–(4.51). Нивното решение (τ_{EH}^*, f_0^*) е оптималното доколку $f_0^* < f_{\max}$; во спротивно, $f_0^* = f_{\max}$, согласно следниот случај.

Случај 2: $\gamma_k > 0, \forall k$. Во овој случај, $f_k = f_0 = f_{\max}, \forall k$. Оптималната вредност на τ_{EH} се одредува со било која од двете равенки од системот (4.50)–(4.51). ■

Б.7 Доказ на Теорема 5.1

Може да се покаже дека оптимизацискиот проблем (5.24) има единствено оптимално решение. Прво, за да се минимизира целната функција, оптималните вредности на f_k мора да бидат што е можно помали, но не помали од заедничката вредност a_0/τ која го задоволува ограничувањето $C6$. Бидејќи f_k не се појавува во останатите ограничувања, $C6$ е задоволено со строго равенство, што ја дава равенката (5.36). Понатаму, да претпоставиме дека, покрај $C6$, ограничувањата $C1, C2$ и $C3$ се исто така задоволени со строги равенства, што ги дава решенијата (5.33), (5.34) и (5.35), додека $C4$ и $C5$ се задоволени со строги неравенства. Условот за изводливост (5.31) го добиваме од (5.33) и условот $\tau \geq a_0/f_{\max}$.

Го користиме методот на Лагранжови множители за да се потврди дека точката $(\tau^*, p_k^*, t_k^*, f_k^*)$, определена со (5.33)–(5.36), е бараниот глобален минимум на (5.24). Претпоставуваме дека $b_k = \sigma_N^2/\Omega_k$, па Лагранжијанот на (5.24) е:

$$\begin{aligned} \mathcal{L}_1 = & \sum_{k=1}^K \left(p_k t_k + \frac{\alpha a_0^3}{\tau^2} \right) - \sum_{k=1}^K \lambda_k \left(t_k \xi \left(\frac{p_k}{b_k} \right) - r_k T \right) \\ & - \sum_{k=1}^K \mu_k \left(\xi \left(\frac{p_k}{b_k} \right) - \xi_{th} \right) - \lambda_0 \left(\tau + \sum_{k=1}^K t_k - T \right), \end{aligned} \quad (\text{Б.111})$$

каде λ_k, μ_k и λ_0 се Лагранжовите множители поврзани со ограничувањата $C1, C2$ и $C3$, соодветно, кои мора да бидат различни од нула ако овие ограничувања се задоволени со

равенство. Ги поставуваме првите изводи на (Б.111) по p_k, t_k и τ на нула:

$$\frac{d\mathcal{L}_1}{dp_k} = t_k - \frac{\lambda_k t_k}{b_k} \xi' \left(\frac{p_k}{b_k} \right) - \frac{\mu_k}{b_k} \xi' \left(\frac{p_k}{b_k} \right) = 0, \forall k, \quad (\text{Б.112})$$

$$\frac{d\mathcal{L}_1}{dt_k} = p_k - \lambda_0 - \lambda_k \xi \left(\frac{p_k}{b_k} \right) = 0, \forall k, \quad (\text{Б.113})$$

$$\frac{d\mathcal{L}_1}{d\tau} = -\frac{2K\alpha a_0^3}{\tau^3} - \lambda_0 = 0, \quad (\text{Б.114})$$

каде првиот извод на ξ во предложената оптимална точка го задоволува условот $\xi'(p_k/b_k) = C_1(A_1 - \xi_{th})(A_2 - \xi_{th})/(A_1 - A_2)$. Од (Б.114) добиваме $\lambda_0 = -2K\alpha a_0^3/\tau^3$, различно од нула за секоја позитивна вредност на τ , со што се потврдува претпоставката дека $C3$ е задоволено со строго равенство. Од (Б.113) добиваме $\lambda_k = \xi_{th}^{-1}(p_k + 2K\alpha a_0^3/\tau^3)$, ненулно за позитивни τ, p_k , со што се потврдува дека $C1$ е исполнето со строго равенство. Од (Б.112) добиваме $\mu_k = t_k[b_k/\xi'(p_k/b_k) - \xi_{th}^{-1}(p_k + 2K\alpha a_0^3/\tau^3)]$, различно од нула за позитивни τ, p_k, t_k , потврдувајќи дека $C2$ е исполнето со строго равенство. Всушност, p_k мора да биде што е можно помало, но не помало од вредноста што ја гарантира минималната семантичка сличност ξ_{th} наметната од $C2$; бидејќи функцијата $\xi(p_k/b_k)$ расте со p_k , оптималната предавателна моќност го задоволува $C2$ со равенство. Ограничувањето $C4$ е релевантно за изводливоста на решението, прикажано преку (5.32). ■

Б.8 Доказ на Теорема 5.2

Иако (5.42) е неконвексен оптимизациски проблем, тој сепак има единствено оптимално решение. Прво, применувајќи слична анализа како во доказот на Теорема 5.1, заклучуваме дека ограничувањето $C6$ е задоволено со строго равенство, што ја дава равенката (5.49).

Случај А. Да претпоставиме дека, покрај $C6$, ограничувањата $\bar{C}1$ и $C3$ се исто така задоволени со еднаквост, додека $C4$ и $C5$ се задоволени со строги неравенства. Методот на Лагранжови множители се користи за одредување на локалните екстреми на (5.42), за кои се покажува дека претставуваат единствен (глобален) минимум. Поставувајќи $b_k = \sigma_N^2/\Omega_k$, Лагранжијанот на (5.42) е:

$$\mathcal{L}_2 = \sum_{k=1}^K \left(p_k t_k + \frac{\alpha a_0^3}{\tau^2} \right) - \sum_{k=1}^K \lambda_k \left(t_k \exp \left(-\frac{x_{th} b_k}{p_k} \right) - \frac{\bar{r}_k T}{A_2} \right) - \lambda_0 \left(\tau + \sum_{k=1}^K t_k - T \right), \quad (\text{Б.115})$$

каде λ_k и λ_0 се Лагранжовите множители поврзани со $\overline{C1}$ и $C3$, соодветно. Ги поставуваме првите изводи на (Б.115) по p_k, t_k и τ на нула:

$$\frac{d\mathcal{L}_2}{dp_k} = t_k \left(1 - \lambda_k \exp \left(-\frac{x_{th} b_k}{p_k} \right) \frac{x_{th} b_k}{p_k^2} \right) = 0, \forall k, \quad (\text{Б.116})$$

$$\frac{d\mathcal{L}_2}{dt_k} = p_k - \lambda_0 - \lambda_k \exp \left(-\frac{x_{th} b_k}{p_k} \right) = 0, \forall k, \quad (\text{Б.117})$$

$$\frac{d\mathcal{L}_2}{d\tau} = -\frac{2K\alpha a_0^3}{\tau^3} - \lambda_0 = 0. \quad (\text{Б.118})$$

Од (Б.116), $\lambda_k = (p_k^2/(x_{th} b_k)) \exp(x_{th} b_k/p_k)$, различно од нула за секој p_k , со што се потврдува дека $\overline{C1}$ е исполнето со строго равенство. Од (Б.118), $\lambda_0 = -2K\alpha a_0^3/\tau^3$, различно од нула за секое $\tau > 0$, потврдувајќи дека $C3$ е исполнето со строго равенство. Вметнувајќи ги овие изрази за λ_k и λ_0 во (Б.117), се добива квадратна равенка по p_k :

$$p_k^2 - x_{th} b_k p_k - \frac{2K\alpha a_0^3 x_{th} b_k}{\tau^3} = 0, \quad (\text{Б.119})$$

која има единствено позитивно решение дадено со изразот (5.46). Комбинирањето на $\overline{C1}$ и $C3$ со (5.46) го дава изразот (5.45). Ако $T > (1/A_2) \sum \bar{r}_k T$, тогаш (5.45) има единствено решение τ^* . Со вметнување на τ^* во (5.46) се добива оптималната моќност $p_k^*(\tau^*)$, додека оптималното времетраење на пренос е $t_k^* = (1/A_2) \bar{r}_k T \exp(x_{th} b_k/p_k^*(\tau^*))$.

Случај Б. Да претпоставиме дека, покрај $\overline{C1}$, $C3$ и $C6$, и ограничувањето $C4$ е задоволено со строго равенство, додека $C5$ е задоволено со строго неравенство. Лагранжијанот $\mathcal{L}_2$ е повторно даден со (Б.115), но со p_k заменето со $P_{\max}$; повторното решавање води до изразите (Б.117) и (Б.118). За да се покаже дека стационарната точка е навистина глобален минимум, се докажува дека целната функција е конвексна по τ и p_k .

Случај В. Ако сите ограничувања во (5.42) се задоволени со строги равенства, сетот на изводливи решенија е единствена точка: $p_k^* = P_{\max}$, $f_k^* = f_{\max}$, $t_k^* = (1/A_2) \bar{r}_k T \exp(x_{th} b_k/P_{\max})$ и $\tau^* = a_0/f_{\max}$. Решение постои ако T е точно еднакво на $T_{\min} = \tau^* + \sum t_k^*$. Оттука, $T \geq T_{\min}$ претставува услов за изводливост (5.44), бидејќи е построг од условот во Случај А. ■

Б.9 Доказ на Теорема 5.3

Иако (5.53) е неконвексен оптимизациски проблем, тој сепак има единствено оптимално решение кое може да се изведе со примена на методот на Лагранжови множители во следниве два случаи.

Случај А. Да претпоставиме дека ограничувањата $C1$ и $C2$ се задоволени со еднаквост, додека $C3$ е задоволено со неравенство. Во овој случај, $C3$ е еквивалентно изразено како:

$$\gamma = \frac{p_1 h_1}{\sigma_N^2} > \mu_0, \quad (\text{Б.120})$$

каде μ_0 е определено со (5.58). Лагранжијанот на соодветниот оптимизациски проблем е:

$$\mathcal{L}_1 = \tau_{IT} \xi \left(\frac{p_1 h_1}{\sigma_N^2} \right) - \lambda_1 (\tau_{EH} + \tau_{IT} - T) - \lambda_2 (p_1 \tau_{IT} - \eta P_0 h_1 \tau_{EH}), \quad (\text{Б.121})$$

каде λ_1 и λ_2 се Лагранжовите множители поврзани со $C1$ и $C2$, соодветно. Ги поставуваме првите изводи на (Б.121) по p_1 , τ_{EH} и τ_{IT} на нула:

$$\frac{d\mathcal{L}_1}{dp_1} = \tau_{IT} \xi' \left(\frac{p_1 h_1}{\sigma_N^2} \right) \frac{h_1}{\sigma_N^2} - \lambda_2 \tau_{IT} = 0, \quad (\text{Б.122})$$

$$\frac{d\mathcal{L}_1}{d\tau_{EH}} = -\lambda_1 + \lambda_2 \eta P_0 h_1 = 0, \quad (\text{Б.123})$$

$$\frac{d\mathcal{L}_1}{d\tau_{IT}} = \xi \left(\frac{p_1 h_1}{\sigma_N^2} \right) - \lambda_1 - \lambda_2 p_1 = 0. \quad (\text{Б.124})$$

Од (Б.123), $\lambda_1 = \lambda_2 \eta P_0 h_1$, што по вметнување во (Б.124) дава:

$$\lambda_2 = \frac{1}{p_1 + \eta P_0 h_1} \xi \left(\frac{p_1 h_1}{\sigma_N^2} \right). \quad (\text{Б.125})$$

Заклучуваме дека λ_2 и соодветниот λ_1 имаат вредности различни од нула за секоја позитивна вредност на p_1 , со што се потврдуваат претпоставките дека $C1$ и $C2$ се исполнети со строги равенства, од каде произлегуваат (5.60) и (5.61). Понатаму, изразот (5.59) се добива со внесување на (Б.125) во (Б.122). Доколку постои решение γ_0 за (5.59), тоа мора да го задоволува условот $\gamma_0 > \mu_0$, со што се потврдува (Б.120). Во спротивно, се применува следниот случај.

Случај Б. Да претпоставиме дека, покрај $C1$ и $C2$, и ограничувањето $C3$ е задоволено со строго равенство, т.е. $\gamma^* = \mu_0$, што дава $p_1^* = \mu_0 \sigma_N^2 / h_1$. Лагранжијанот на соодветниот оптимизациски проблем е:

$$\mathcal{L}_2 = \mathcal{L}_1 + \lambda_3 \left(p_1 - \frac{\mu_0 \sigma_N^2}{h_1} \right), \quad (\text{Б.126})$$

каде $\mathcal{L}_1$ е даден со (Б.121), а λ_3 е Лагранжовиот множител поврзан со $C3$, различен од нула доколку $C3$ е исполнето со равенство. Првите изводи по τ_{EH} и τ_{IT} изедначени на нула соодветствуваат на (Б.123) и (Б.124), додека изводот по p_1 е:

$$\frac{d\mathcal{L}_2}{dp_1} = \frac{d\mathcal{L}_1}{dp_1} + \lambda_3 = 0. \quad (\text{Б.127})$$

Вметнувањето на (Б.125) во (Б.127) дава:

$$\lambda_3 = \frac{\tau_{IT}^* h_1}{\sigma_N^2} \frac{\xi(\mu_0) - \xi'(\mu_0)(\mu_0 + a_0)}{\mu_0 + a_0}. \quad (\text{Б.128})$$

Бидејќи (5.59) не е задоволено, λ_3 дадено со (Б.128) не е нула, со што се потврдува претпоставката дека $C3$ е задоволено со равенство. ■

Користена литература

- [1] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, “The road towards 6G: A comprehensive survey,” *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.
- [2] Q. Wu and R. Zhang, “Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network,” *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, 2020.
- [3] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, “6G Wireless Communication Systems: Applications, Requirements, Technologies, Challenges, and Research Directions,” *IEEE Open Journal of the Communications Society*, vol. 1, pp. 957–975, 2020.
- [4] W. Saad, M. Bennis, and M. Chen, “A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems,” *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, 2020.
- [5] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, “What should 6G be?,” *Nat. Electron.*, vol. 3, no. 1, pp. 20–29, 2020.
- [6] S. Nayak and R. Patgiri, “6G Communication: Envisioning the Key Issues and Challenges,” *EAI Endorsed Transactions on Internet of Things*, vol. 6, no. 24, p. e1, 2020.
- [7] Y. Liu *et al.*, “Reconfigurable Intelligent Surfaces: Principles and Opportunities,” *IEEE Communications Surveys and Tutorials*, vol. 23, no. 3, pp. 1546–1577, 2021.
- [8] S. Barbarossa, S. Sardellitti, and P. D. Lorenzo, “Communicating while computing: Distributed mobile cloud computing over 5G heterogeneous networks,” in *IEEE Signal Process. Mag.*, vol. 31, pp. 45–55, 2014.
- [9] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data.” <https://arxiv.org/pdf/1602.05629>, 2017. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017.
- [10] D. Gunduz *et al.*, “Beyond Transmitting Bits: Context, Semantics, and Task-Oriented Communications,” *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5–41, 2023.
- [11] G. C. Alexandropoulos, G. Lerosey, M. Debbah, and M. Fink, “Reconfigurable Intelligent Surfaces and Metamaterials: The Potential of Wave Propagation Control for 6G Wireless Communications.” <https://arxiv.org/pdf/2006.11136>, 2020.

- [12] S. V. Hum and J. Perruisseau-Carrier, "Reconfigurable reflectarrays and array lenses for dynamic antenna beam control: A review," *IEEE Trans. Antennas Propag.*, vol. 62, no. 1, pp. 183–198, 2014.
- [13] E. Basar *et al.*, "Reconfigurable Intelligent Surfaces for 6G: Emerging Hardware Architectures, Applications, and Open Challenges," *IEEE Vehicular Technology Magazine*, vol. 19, no. 3, pp. 27–47, 2024.
- [14] Q. Wu and R. Zhang, "Intelligent Reflecting Surface Enhanced Wireless Network via Joint Active and Passive Beamforming," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 11, pp. 5394–5409, 2019.
- [15] M. D. Renzo *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: an idea whose time has come," *EURASIP J. Wirel. Commun. Netw.*, vol. 2019, no. 1, pp. 1–20, 2019.
- [16] W. Saad, M. Bennis, and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, 2020.
- [17] T. Jiang and W. Yu, "Interference Nulling Using Reconfigurable Intelligent Surface," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 5, pp. 1392–1406, 2022.
- [18] A. Khalili, S. Zargari, Q. Wu, D. W. K. Ng, and R. Zhang, "Multi-Objective Resource Allocation for IRS-Aided SWIPT," *IEEE Wireless Communications Letters*, vol. 10, no. 6, pp. 1324–1328, 2021.
- [19] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and Secure Wireless Communications via Intelligent Reflecting Surfaces," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2637–2652, 2020.
- [20] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems," in *IEEE Transactions on Signal Processing*, vol. 57, pp. 4837–4851, 2009.
- [21] E. Bjornson, O. Ozdogan, and E. G. Larsson, "Intelligent Reflecting Surface Versus Decode-and-Forward: How Large Surfaces are Needed to Beat Relaying?," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 244–248, 2020.
- [22] M. Patel *et al.*, "Mobile-Edge Computing Contributing Organizations and Authors.."
- [23] S. Bi, C. K. Ho, and R. Zhang, "Wireless powered communication: Opportunities and challenges," *IEEE Communications Magazine*, vol. 53, no. 4, pp. 117–125, 2015.
- [24] C. You, K. Huang, and H. Chae, "Energy efficient mobile cloud computing powered by wireless energy transfer," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1757–1771, 2016.

- [25] S. Bi and Y. J. Zhang, "Computation Rate Maximization for Wireless Powered Mobile-Edge Computing with Binary Computation Offloading," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 6, pp. 4177–4190, 2018.
- [26] F. Wang, J. Xu, X. Wang, and S. Cui, "Joint Offloading and Computing Optimization in Wireless Powered Mobile-Edge Computing Systems," <http://arxiv.org/pdf/1702.00606>, vol. PP, 2017.
- [27] F. Zhou and R. Q. Hu, "Computation Efficiency Maximization in Wireless-Powered Mobile Edge Computing Networks," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 5, pp. 3170–3184, 2020.
- [28] S. Mao, S. Leng, S. Maharjan, and Y. Zhang, "Energy efficiency and delay tradeoff for wireless powered mobile-edge computing systems with multi-access schemes," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 3, pp. 1855–1867, 2020.
- [29] Z. Yang, M. Chen, W. Saad, C. S. Hong, and M. Shikh-Bahaei, "Energy Efficient Federated Learning over Wireless Communication Networks," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 3, pp. 1935–1949, 2021.
- [30] P. S. Bouzinis, P. D. Diamantoulakis, and G. K. Karagiannidis, "Wireless Federated Learning (WFL) for 6G Networks-Part II: The Compute-Then-Transmit NOMA Paradigm," *IEEE Communications Letters*, vol. 26, no. 1, pp. 8–12, 2022.
- [31] Q. Zeng, Y. Du, and K. Huang, "Wirelessly Powered Federated Edge Learning: Optimal Tradeoffs between Convergence and Power Transfer," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 1, pp. 680–695, 2022.
- [32] R. Hamdi, M. Chen, A. B. Said, M. Qaraqe, and H. V. Poor, "Federated Learning over Energy Harvesting Wireless Networks," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 92–103, 2022.
- [33] Y. Wu, Y. Song, T. Wang, L. Qian, and T. Q. S. Quek, "Non-Orthogonal Multiple Access Assisted Federated Learning via Wireless Power Transfer: A Cost-Efficient Approach," *IEEE Transactions on Communications*, vol. 70, no. 4, pp. 2853–2869, 2022.
- [34] T. M. Getu, G. Kaddoum, and M. Bennis, "A Survey on Goal-Oriented Semantic Communication: Techniques, Challenges, and Future Directions," *IEEE Access*, vol. 12, pp. 51223–51274, 2024.
- [35] W. Yang *et al.*, "Semantic Communications for Future Internet: Fundamentals, Applications, and Challenges," *IEEE Communications Surveys and Tutorials*, vol. 25, no. 1, pp. 213–250, 2023.
- [36] H. Xie, Z. Qin, G. Y. Li, and B. H. Juang, "Deep Learning Enabled Semantic Communication Systems," in *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663–2675, 2021.

- [37] J. Xu, T. Y. Tung, B. Ai, W. Chen, Y. Sun, and D. Gunduz, "Deep Joint Source-Channel Coding for Semantic Communications," *IEEE Communications Magazine*, vol. 61, no. 11, pp. 42–48, 2023.
- [38] L. Yan, Z. Qin, R. Zhang, Y. Li, and G. Y. Li, "Resource Allocation for Text Semantic Communications," *IEEE Wireless Communications Letters*, vol. 11, no. 7, pp. 1394–1398, 2022.
- [39] L. Wang, W. Wu, F. Zhou, Z. Yang, Z. Qin, and Q. Wu, "Adaptive Resource Allocation for Semantic Communication Networks," *IEEE Transactions on Communications*, vol. 72, no. 11, pp. 6900–6916, 2024.
- [40] Z. Yang, M. Chen, Z. Zhang, and C. Huang, "Energy Efficient Semantic Communication over Wireless Networks with Rate Splitting," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1484–1495, 2023.
- [41] N. G. Evgenidis *et al.*, "Hybrid Semantic-Shannon Communications," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 9, pp. 10926–10940, 2024.
- [42] M. D. Renzo *et al.*, "Smart Radio Environments Empowered by AI Reconfigurable Meta-Surfaces: An Idea Whose Time Has Come," *EURASIP J. Wirel. Commun. Netw.*, vol. 2019, no. 1, 2019.
- [43] V. S. Asadchy, M. Albooyeh, S. N. Tsvetkova, A. Díaz-Rubio, Y. Ra'di, and S. A. Tretyakov, "Perfect control of reflection and refraction using spatially dispersive metasurfaces," *Phys. Rev. B*, vol. 94, no. 7, 2016.
- [44] A. Díaz-Rubio, V. S. Asadchy, A. Elsakka, and S. A. Tretyakov, "From the generalized reflection law to the realization of perfect anomalous reflectors," *Sci. Adv.*, vol. 3, no. 8, 2016.
- [45] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A New Wireless Communication Paradigm through Software-Controlled Metasurfaces," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 162–169, 2018.
- [46] M. D. Renzo *et al.*, "Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 798–807, 2020.
- [47] E. Bjornson, O. Ozdogan, and E. G. Larsson, "Intelligent Reflecting Surface Versus Decode-and-Forward: How Large Surfaces are Needed to Beat Relaying?," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 244–248, 2020.
- [48] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.

- [49] W. Roh *et al.*, “Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 106–113, 2014.
- [50] Q. Wu and R. Zhang, “Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network,” *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, 2019.
- [51] L. Yin, P. Yang, Y. Gan, F. Yang, S. Yang, and Z. Nie, “A low cost, low in-band RCS microstrip phased-array antenna with integrated 2-bit phase shifter,” *IEEE Trans. Antennas Propag.*, vol. 69, no. 8, pp. 4517–4526, 2021.
- [52] R. Deng *et al.*, “Reconfigurable Holographic Surfaces for Future Wireless Communications,” *IEEE Wirel. Commun.*, vol. 28, no. 6, pp. 126–131, 2021.
- [53] M. H. Dahri, M. H. Jamaluddin, M. I. Abbasi, and M. R. Kamarudin, “A Review of Wideband Reflectarray Antennas for 5G Communication Systems,” *IEEE Access*, vol. 5, pp. 17803–17815, 2017.
- [54] E. Bjornson and L. Sanguinetti, “Power Scaling Laws and Near-Field Behaviors of Massive MIMO and Intelligent Reflecting Surfaces,” *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1306–1324, 2020.
- [55] N. Zlatanov, E. Sippel, V. Jamali, and R. Schober, “Capacity of the Gaussian two-hop full-duplex relay channel with residual self-interference,” *IEEE Transactions on Communications*, vol. 65, no. 3, pp. 1005–1021, 2017.
- [56] E. Everett, A. Sahai, and A. Sabharwal, “Passive self-interference suppression for full-duplex infrastructure nodes,” *IEEE Trans. Wirel. Commun.*, vol. 13, no. 2, pp. 680–694, 2014.
- [57] E. Bjornson and L. Sanguinetti, “Power Scaling Laws and Near-Field Behaviors of Massive MIMO and Intelligent Reflecting Surfaces,” *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1306–1324, 2020.
- [58] O. Ozdogan, E. Bjornson, and E. G. Larsson, “Intelligent Reflecting Surfaces: Physics, Propagation, and Pathloss Modeling,” *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 581–585, 2020.
- [59] S. W. Ellingson, “Path Loss in Reconfigurable Intelligent Surface-Enabled Channels,” in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, vol. 2021-September, pp. 829–835, 2021.
- [60] Q. Wu and R. Zhang, “Beamforming Optimization for Wireless Network Aided by Intelligent Reflecting Surface with Discrete Phase Shifts,” *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1838–1851, 2020.

- [61] N. Zlatanov, R. Schober, and P. Popovski, "Buffer-aided relaying with adaptive link selection," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 8, pp. 1530–1542, 2013.
- [62] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019.
- [63] M. Forouzanmehr, S. Akhlaghi, A. Khalili, and Q. Wu, "Energy Efficiency Maximization for IRS-Assisted Uplink Systems: Joint Resource Allocation and Beamforming Design," *IEEE Communications Letters*, vol. 25, no. 12, pp. 3932–3936, 2021.
- [64] S. Zargari, A. Khalili, and R. Zhang, "Energy Efficiency Maximization via Joint Active and Passive Beamforming Design for Multiuser MISO IRS-Aided SWIPT," *IEEE Wireless Communications Letters*, vol. 10, no. 3, pp. 557–561, 2021.
- [65] M. R. Palattella *et al.*, "Internet of Things in the 5G Era: Enablers, Architecture, and Business Models," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 510–527, 2016.
- [66] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile Edge Computing: A Survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 450–465, 2018.
- [67] M. Bahrami, "Cloud computing for emerging mobile cloud apps," in *Proceedings - 2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, MobileCloud 2015*, pp. 4–5, 2015.
- [68] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-optimal mobile cloud computing under stochastic wireless channel," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 9, pp. 4569–4581, 2013.
- [69] D. Sabella, A. Vaillant, P. Kuure, U. Rauschenbach, and F. Giust, "Mobile-Edge Computing Architecture: The role of MEC in the Internet of Things," *IEEE Consumer Electronics Magazine*, vol. 5, no. 4, pp. 84–91, 2016.
- [70] B. Shi, J. Yang, Z. Huang, and P. Hui, "Offloading Guidelines for Augmented Reality Applications on Wearable Devices," 2015.
- [71] S. Nunna *et al.*, "Enabling Real-Time Context-Aware Collaboration through 5G and Mobile Edge Computing," in *Proceedings - 12th International Conference on Information Technology: New Generations, ITNG 2015*, pp. 601–605, 2015.
- [72] S. Josilo and G. Dan, "Selfish Decentralized Computation Offloading for Mobile Cloud Computing in Dense Wireless Networks," *IEEE Trans. Mob. Comput.*, vol. 18, no. 1, pp. 207–220, 2019.
- [73] W. Yu *et al.*, "A Survey on the Edge Computing for the Internet of Things," *IEEE Access*, vol. 6, pp. 6900–6919, 2017.

- [74] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Latency Minimization for Intelligent Reflecting Surface Aided Mobile Edge Computing," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2666–2682, 2020.
- [75] X. Wang *et al.*, "Wireless Powered Mobile Edge Computing Networks: A Survey," *ACM Comput. Surv.*, vol. 55, p. Article, 2023.
- [76] X. Zhang and S. Debroy, "Resource Management in Mobile Edge Computing: A Comprehensive Survey," *ACM Comput. Surv.*, vol. 55, no. 13 s, 2023.
- [77] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.
- [78] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.
- [79] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient Multi-User Computation Offloading for Mobile-Edge Cloud Computing," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2795–2808, 2016.
- [80] T. Burd and R. Brodersen, "Processor Design for Portable Systems," in *Journal of VLSI Signal Processing*, vol. 13, 1996.
- [81] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-optimal mobile cloud computing under stochastic wireless channel," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 9, pp. 4569–4581, 2013.
- [82] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: Architecture design and rate-energy tradeoff," *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4754–4767, 2013.
- [83] H. Ju and R. Zhang, "Throughput maximization in wireless powered communication networks," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 1, pp. 418–428, 2014.
- [84] B. Radunović and J. Y. L. Boudec, "A unified framework for max-min and min-max fairness with applications," *IEEE/ACM Transactions on Networking*, vol. 15, no. 5, pp. 1073–1083, 2007.
- [85] R. Corless, G. Gonnet, D. Hare, D. Jeffrey, and D. Knuth, "On the Lambert W Function," *Adv. Comput. Math.*, vol. 5, pp. 329–359, 1996.
- [86] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure Of Fairness And Discrimination For Resource Allocation In Shared Computer Systems." <https://arxiv.org/pdf/cs/9809099>, 1998.

- [87] F. Kelly, "Charging and Rate Control for Elastic Traffic," *European Transactions on Telecommunications*, vol. 8, no. 1, pp. 33–37, 1997.
- [88] K. Bonawitz *et al.*, "Towards Federated Learning at Scale: System Design." <https://arxiv.org/pdf/1902.01046>, 2019.
- [89] H. Hellström *et al.*, "Wireless for Machine Learning: A Survey," in *Foundations and Trends® in Signal Processing*, vol. 15, pp. 290–399, 2022.
- [90] S. Ruder, "An overview of gradient descent optimization algorithms." <https://arxiv.org/pdf/1609.04747>, 2016.
- [91] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [92] F. Sattler, S. Wiedemann, K. R. Muller, and W. Samek, "Robust and Communication-Efficient Federated Learning from Non-i.i.d. Data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3400–3413, 2020.
- [93] S. Hu, X. Chen, W. Ni, E. Hossain, and X. Wang, "Distributed machine learning for wireless communication networks: Techniques, architectures, and applications," *IEEE Communications Surveys and Tutorials*, vol. 23, no. 3, pp. 1458–1493, 2021.
- [94] X. Li, W. Yang, Z. Zhang, K. Huang, and S. Wang, "On the Convergence of FedAvg on Non-IID Data." <https://arxiv.org/pdf/1907.02189>, 2020. 8th International Conference on Learning Representations, ICLR 2020.
- [95] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated Learning for Wireless Communications: Motivation, Opportunities, and Challenges," *IEEE Communications Magazine*, vol. 58, no. 6, pp. 46–51, 2020.
- [96] Z. Yang *et al.*, "Delay Minimization for Federated Learning Over Wireless Communication Networks." <http://arxiv.org/abs/2007.03462>, 2020.
- [97] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A Joint Learning and Communications Framework for Federated Learning over Wireless Networks," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 1, pp. 269–283, 2021.
- [98] P. S. Bouzinis, P. D. Diamantoulakis, and G. K. Karagiannidis, "Wireless Federated Learning (WFL) for 6G NetworksPart I: Research Challenges and Future Trends," *IEEE Communications Letters*, vol. 26, no. 1, pp. 3–7, 2022.
- [99] P. S. Bouzinis, P. D. Diamantoulakis, and G. K. Karagiannidis, "Wireless Federated Learning (WFL) for 6G Networks-Part II: The Compute-Then-Transmit NOMA Paradigm," *IEEE Communications Letters*, vol. 26, no. 1, pp. 8–12, 2022.
- [100] X. Ma, H. Sun, and R. Q. Hu, "Scheduling Policy and Power Allocation for Federated Learning in NOMA Based MEC," in *Proceedings - IEEE Global Communications Conference, GLOBECOM*, 2020.

- [101] M. Chen, H. V. Poor, W. Saad, and S. Cui, "Convergence Time Minimization of Federated Learning over Wireless Networks," in *IEEE International Conference on Communications*, vol. 2020-June, 2020.
- [102] Y. Liu, X. Yuan, Z. Xiong, J. Kang, X. Wang, and D. Niyato, "Federated learning for 6G communications: Challenges, methods, and future directions," *China Communications*, vol. 17, no. 9, pp. 105–118, 2020.
- [103] D. Liu and O. Simeone, "Privacy for Free: Wireless Federated Learning via Uncoded Transmission with Adaptive Power Control," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 170–185, 2021.
- [104] F. Fang, Y. Xu, Z. Ding, C. Shen, M. Peng, and G. K. Karagiannidis, "Optimal Resource Allocation for Delay Minimization in NOMA-MEC Networks," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7867–7881, 2020.
- [105] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [106] W. H. Press, S. a Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes 3rd Edition: The Art of Scientific Computing." https://books.google.com/books/about/Numerical_Recipes_3rd_Edition.html?hl=fr&id=1aA0dzK3FegC, 2007. Sample page from NUMERICAL RECIPES IN C.
- [107] K. E. Atkinson, "An introduction to numerical analysis." <https://www.wiley.com/en-us/An+Introduction+to+Numerical+Analysis%2C+2nd+Edition-p-9780471624899>, 1989.
- [108] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [109] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [110] C. E. Shannon and W. Weaver, "The Theory of Mathematical Communication." <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>, 1964. Bell System Technical Journal.
- [111] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." <https://arxiv.org/pdf/1810.04805>, 2019. NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference.
- [112] X. Mu and Y. Liu, "Exploiting Semantic Communication for Non-Orthogonal Multiple Access," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2563–2576, 2023.
- [113] Z. Weng and Z. Qin, "Semantic Communication Systems for Speech Transmission," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2434–2444, 2021.