

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta elektrotechnická
Katedra telekomunikační techniky



Chování mobilního terminálu v čase
Behavior of Mobile Terminal in Time

Bakalářská práce

Studijní program: Elektronika a komunikace

Vedoucí práce: Ing. Bešťák Robert Ph.D.

Filip Krupka

Praha 2020

Čestné prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

Datum: 22. 5. 2020

Podpis autora práce



ZADÁNÍ BAKALÁŘSKÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Krupka** Jméno: **Filip** Osobní číslo: **474243**
Fakulta/ústav: **Fakulta elektrotechnická**
Zadávající katedra/ústav: **Katedra telekomunikační techniky**
Studijní program: **Elektronika a komunikace**

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

Chování mobilního terminálu v čase

Název bakalářské práce anglicky:

Behavior of Mobile Terminal in Time

Pokyny pro vypracování:

Na základě analýzy signalizačních dat v mobilní síti stanovte vzor chování mobilních terminálů při využívání síťových prostředků (technologie, frekvence) v průběhu dne, týdne, měsíce.

Seznam doporučené literatury:

- [1] Marr, B.: Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance, Wiley, 2015. ISBN: 978-1-118-96583-2.
[2] Walke, B.H.: Mobile Radio Networks: Networking, Protocols and Traffic Performance, 2nd Edition, Wiley, 2002. ISBN: 978-0-471-49902-2.
[3] Okabe, A.; Boots, B.; Sugihara, K; Chiu, S.N.: Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, 2nd Edition, Wiley, 2000. ISBN: 978-0-471-98635-5.

Jméno a pracoviště vedoucí(ho) bakalářské práce:

Ing. Robert Bešťák, Ph.D., katedra telekomunikační techniky FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) bakalářské práce:

Datum zadání bakalářské práce: **07.01.2020** Termín odevzdání bakalářské práce: **22.05.2020**

Platnost zadání bakalářské práce: **30.09.2021**

Ing. Robert Bešťák, Ph.D.
podpis vedoucí(ho) práce

podpis vedoucí(ho) ústavu/katedry

prof. Mgr. Petr Páta, Ph.D.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Student bere na vědomí, že je povinen vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

Datum převzetí zadání

Podpis studenta

Anotace

Tato bakalářská práce s názvem „Chování mobilního terminálu v čase“ se zabývá, jak již z názvu vyplývá, stanovením modelu chování mobilních terminálů v rámci celulární rádiové sítě. V práci je mimo jiné popsáno také fungování a struktura této komunikační sítě a jednotlivé technologie, které se v rámci této sítě šíří. Především se práce věnuje predikci možného chování mobilních terminálů. Tato predikce byla stanovena na základě provedených analýz na signalizačních datech z celulární rádiové sítě.

Klíčová slova

Celulární rádiová síť, mobilní terminál, signalizační data, buňka, technologie, predikce chování

Summary

This bachelor's thesis entitled "Behavior of a mobile terminal in time" is focused, as the name implies, on determining the model of behavior of mobile terminals within a cellular radio network. The work describes, among other things, the functioning and structure of this communication network and the individual technologies that are spreading over this network. Primarily, the work is focused on determining the prediction of possible behavior of mobile terminals. This prediction was determined on the basis of analyzes performed on signaling data from the cellular radio network.

Index terms

Cellular radio network, mobile terminal, signaling data, cell, technology, prediction of behavior

Poděkování

Rád bych poděkoval Bc. Jiřímu Hylmarovi, MBA, za poskytnutá signalizační data k provedení potřebných analýz. Nadále bych rád poděkoval vedoucímu práce Ing. Robertu Bešťákovi Ph.D. za věcné rady při zpracování práce.

Obsah

1	Úvod.....	16
2	Velké datové objemy.....	17
3	Celulární radiová síť.....	19
3.1	Mobilní terminál	19
3.2	Buňkový systém.....	19
3.3	Technologie	21
3.4	Sektorizace.....	22
3.5	Procedura přepojení	23
3.6	Velikost buněk	24
3.7	Přístupové metody	24
4	Zpracování a datové sady.....	26
4.1	Použité funkce.....	26
4.2	Datové sady.....	27
5	Analýzy a výstupy.....	30
5.1	Počet buněk jednotlivých technologií.....	30
5.2	Rozložení záznamů uživatelů mezi jednotlivými sítěmi	32
5.2.1	Použití jedné technologií	33
5.2.2	Použití dvou technologií.....	36
5.2.3	Použití tří technologií	41
5.3	Identifikace uživatelských skupin.....	46
5.3.1	Rozdělení podle počtu záznamů	46
5.3.2	Rozdělení podle průměrného času připojení na buňce.....	49
5.3.3	Rozdělení podle počtu signalizačních zpráv vyměněných si s BTS	52
5.3.4	Rozložení záznamů o různé délce připojení.....	55
6	Závěr.....	56
	Seznam použitých informačních zdrojů	57
	Přílohy A	59

Seznam obrázků

Obrázek 1. - Buňkový systém.	20
Obrázek 2. - Překryv buněk.....	20
Obrázek 3. - Překryv technologií.....	20
Obrázek 4. - Sektorizace.....	23
Obrázek 5. - Přístupové metody.	25
Obrázek 6. - Signalizační data.....	27
Obrázek 7. - Počet mobilních terminálů používajících jednu technologii v ranním časovém okně.	35
Obrázek 8. - Počet mobilních terminálů používajících jednu technologii v poledním časovém okně.	35
Obrázek 9. - Počet mobilních terminálů používajících jednu technologii ve večerním časovém okně.	35
Obrázek 10. - Počet mobilních terminálů používajících jednu technologii za celý den.	36
Obrázek 11. - Počet mobilních terminálů používajících dvě technologie v ranním časovém okně.	39
Obrázek 12. - Počet mobilních terminálů používajících dvě technologie v poledním časovém okně.	40
Obrázek 13. - Počet mobilních terminálů používajících dvě technologie ve večerním časovém okně.	40
Obrázek 14. - Počet mobilních terminálů používajících dvě technologie za celý den... ..	40
Obrázek 15. - Počet mobilních terminálů používajících tři technologie v ranním časovém okně.	44
Obrázek 16. - Počet mobilních terminálů používajících tři technologie v poledním časovém okně.	45
Obrázek 17. - Počet mobilních terminálů používajících tři technologie ve večerním časovém okně.	45
Obrázek 18. – Počet mobilních terminálů používajících tři technologie za celý den. ...	45
Obrázek 19. - Počet mobilních terminálů v závislosti na počtu záznamů.	48
Obrázek 20. - Počet mobilních terminálů v závislosti na počtu záznamů – celkový záběr	49
Obrázek 21. - Počet mobilních terminálů v závislosti na průměrném čase připojení na buňkách.....	51
Obrázek 22. - Počet mobilních terminálů v závislosti na počtu signalizačních zpráv vyměněných s BTS.....	54
Obrázek 23. - Počet mobilních terminálů v závislosti na počtu signalizačních zpráv vyměněných s BTS - celkový záběr.	54
Obrázek 24. - Počet záznamů v závislosti na délce času připojení na buňce.....	55

Seznam tabulek

Tabulka 1. - Charakteristiky datových sad.	29
Tabulka 2. - Použití datových sad v jednotlivých kapitolách.	29
Tabulka 3. - Charakteristika datové sady.	30
Tabulka 4. - Počet buněk každé technologie.	32
Tabulka 5. - Počet mobilních terminálů v časových oknech s platnými údaji ID buněk.	33
Tabulka 6. - Rozložení uživatelů mezi technologiemi při rozdělení do časových oken.	34
Tabulka 7. - Procentuální zastoupení mobilních terminálů používajících pouze jedné technologie.	34
Tabulka 8. - Počet mobilních terminálů v časových oknech používajících více než jednu technologie.	37
Tabulka 9. - Rozložení uživatelů mezi kombinacemi dvou technologií při rozdělení do časových oken.	38
Tabulka 10. - Procentuální zastoupení mobilních terminálů používajících pouze daných kombinací technologií.	39
Tabulka 11. - Počet mobilních terminálů v časových oknech používajících více než dvě technologie.	41
Tabulka 12. - Rozložení uživatelů mezi kombinacemi tří technologií při rozdělení do časových oken.	43
Tabulka 13. - Procentuální zastoupení mobilních terminálů používajících pouze daných kombinací tří technologií.	44
Tabulka 14. - Charakteristiky původní datové sady.	46
Tabulka 15. - Počet mobilních terminálů v jednotlivých skupinách podle počtu záznamů.	48
Tabulka 16. - Počet mobilních terminálů v jednotlivých skupinách podle průměrného času připojení.	51
Tabulka 17. - Počet mobilních terminálů v jednotlivých skupinách podle počtu signalizačních zpráv.	53

Seznam použitých funkcí

tibble() - funkce vytvoří datovou tabulku

filter() - funkce ponechá pouze záznamy v tabulce odpovídající zadané podmínce

sum() - funkce sečte všechny numerické proměnné ve vektoru

count() - funkce vygeneruje seznam obsahující počet unikátních proměnných či počet unikátních kombinací proměnných v tabulce

mean() - funkce vygeneruje průměr numerických hodnot ve vektoru

nrow() - funkce spočte počet řádků tabulky

substr() - funkce vyjme či nahradí určenou část řetězce

distinct() - funkce ponechá v tabulce pouze jeden záznam pro každou unikátní proměnnou v zadané skupině proměnných

split() - funkce rozdělí tabulku na dílčí tabulky podle určené skupiny proměnných, následně vytvoří list obsahující tyto dílčí tabulky

lapply() - funkce uplatní určenou funkci na každý objekt v listu

sapply() - funkce uplatní určenou funkci na každý objekt v listu, disponuje možností dodatkového určení datové struktury výstupu

as.numeric() – funkce zobrazí proměnnou jako číslo

as.character() – funkce zobrazí proměnnou jako charakter

as.logical() - funkce zobrazí proměnnou jako log. proměnnou *TRUE/FALSE*

do.call() – volací funkce, pomocí této funkce je možné aplikovat na objekt libovolnou zadanou funkci

rbind() – funkce spojí datové tabulky do jedné

round() – funkce zaokrouhlí numerickou hodnotu na specifikovaný počet číslic

function() – funkce umožňuje vytvořit novou funkci

is.data.frame() – ověřovací funkce, pokud je objekt zadán do funkce ve formátu datové tabulky, vygeneruje na výstup log. proměnnou *TRUE*, v opačném případě *FALSE*

colnames() – funkce umožňuje pojmenovat či přejmenovat sloupce v tabulce

plot() – funkce vygeneruje graf, možnost mnoha definicí vzhledu a obsahu grafu

barplot() – funkce vygeneruje sloupcový graf, možnost mnoha definicí vzhledu a obsahu grafu

Seznam zkratek

DNA	<i>Deoxyribonucleic acid</i>
NMT	<i>Nordic Mobile Telephone</i>
AMPS	<i>Advanced Mobile Telephone System</i>
TACS	<i>Total Access Communication System</i>
GSM	<i>Global System for Mobile Communication</i>
GPRS	<i>General Packet Radio Service</i>
EDGE	<i>Enhanced Data rates for Global Evolution</i>
UMTS	<i>Universal Mobile Telecommunication System</i>
HSPA	<i>High Speed Packet Access</i>
LTE	<i>Long Term Evolution</i>
LTE-A	<i>Long Term Evolution – Advanced</i>
2G	2. generace
3G	3. generace
4G	4. generace
TDMA	<i>Time Division Multiple Access</i>
FDMA	<i>Frequency Division Multiple Access</i>
CDMA	<i>Code Division Multiple Access</i>
WCDMA	<i>Wideband Code Division Multiple Access</i>
OFDMA	<i>Orthogonal Frequency Division Multiple Access</i>
SC-FDMA	<i>Single Carrier Orthogonal Frequency Division Multiple Access</i>
FDD	<i>Frequency Division Duplex</i>
TDD	<i>Time Division Duplex</i>
BTS	<i>Base Transceiver Station</i>
BSC	<i>Base Station Controller</i>
SIM	<i>Subscriber Identity Module</i>
IMSI	<i>International Mobile Subscriber Identity</i>
TMSI	<i>Temporary Mobile Subscriber Identity</i>
LAI	<i>Location Area Identity</i>
MCC	<i>Mobile Country Code</i>
MNC	<i>Mobile Network Code</i>
LAC	<i>Location Area Code</i>

1 Úvod

Celulární rádiová síť představuje velmi široký pojem s mnoha odvětvími, který je atraktivní tématem jak pro studium, tak pro podrobné zkoumání. Interakce s tímto komunikačním prostředkem se stala běžnou rutinou většiny lidí v moderním světě. Uživatelé skrze mobilní terminál mohou využívat možností tohoto komunikačního prostředku, a to zejména k pracovním účelům, zábavě, či ke komunikaci s druhými. Mobilní terminál, se stal součástí každodenní výbavy velké části obyvatelstva. Každý tuto komunikační síť využívá, ale pouze malá hrstka uživatelů skutečně ví, jak taková to síť funguje. Počty mobilních terminálů každým rokem stoupají, a škála typů uživatelů se rozšiřuje. Je ale možné za pomoci velkých objemů dat z této sítě odhadnout, o jaké uživatele se jedná, či nějakým způsobem i předvídat jejich chování? I takovými tématům se věnuje tato práce.

Předním cílem této práce je stanovení modelu chování mobilních terminálů v rámci celulární rádiové sítě v souvislosti s využitím technologií a jednotlivých frekvenčních pásem. V kapitole „Velké datové objemy“ je teoreticky popsán způsob, jakým je možné pracovat s velkými datovými objemy. Dále je zde uvedeno, co nám analýza velkých datových objemů přináší a jaké jsou výhody a nevýhody práce s nimi. Následující kapitola „Celulární rádiová síť“ se věnuje struktuře a chování celulární rádiové sítě. Je zde popsán jak vývoj, tak i fungování jednotlivých technologií provozovaných za pomoci této komunikační sítě. Kapitola „Zpracování a datové sady“ uvádí signalizační data, která jsou klíčová pro stanovení vzoru chování mobilních terminálů. Další část kapitoly se soustředí na způsob, jakým byla data zpracována, a jaké technické prostředky byly za tímto účelem použity. V závěrečné kapitole s názvem „Analýzy a výstupy“ jsou nejprve popsány jednotlivé analýzy společně s postupem provedení jednotlivých úkonů. Následně jsou zaznamenány výsledky analýz a jejich grafické zobrazení, které poukazuje na různé vzory chování, jež jsou hlavním předmětem zkoumání v této bakalářské práci.

2 Velké datové objemy

Velké datové objemy, kterým se v oboru říká „*Big Data*“, představují pojem, objevující se posledních 20 let v mnoha odvětví lidské interakce a velmi rychle se stává nedílnou součástí života každého člověka. „*Big Data*“ představuje oblast, kde objemy dat přesahují technické možnosti nesespecializovaných zařízení a přinášejí do dnes nepředvídatelné poznatky. Každý člověk na planetě dnes zanechává alespoň minimální datovou stopu. Ať už se jedná o záznam o narození, údaje o místě pobytu, či připojení mobilním terminálem k vysílači. Podle studie [10] Martina Hilberta na Annenberg School of Communication and Journalism se objem digitálních dat existujících na planetě Zemi téměř zdvojnásobuje každé 3 roky [1]. Každý občan je obklopen převratnými objemy dat každý den a poznatky a inovace, které každému přinášejí „*Big Data*“ jsou denně využívána, aniž by o tom mnoho občanů vědělo. V budoucnu se dopad a uplatnění analýzy velkých objemů dat bude v lidských životech projevovat každým rokem více a více, o tom není pochyb. Ale čím to, že analýza a zkoumání „*Big Data*“ přišla až v tomto tisíciletí? Počítače a výpočetní technika přeci jen existují již mnoho desítek let. Je to způsobeno přirozeným pokrokem v technice, který je při pohledu zpět znatelný. Dříve neexistovaly superpočítače schopné analyzovat miliony a miliardy vzorků dat, které dnešní superpočítače zvládnou zpracovat v řádech sekund či minut. Než vědci roku 2003 poprvé dekodovali lidský genom (*Deoxyribonucleic acid*, DNA), trvalo jim sekvenování třech miliard párů bází lidského DNA deset let intenzivní práce. O deset let později, dokázala takovou to délku DNA sekvenovat jediná laboratoř za den [1]. Je možné si představit, že schopnosti techniky v roce 2020 jsou o to markantnější.

Práce s velkými objemy dat má mnoho nástrah, a je nepopíratelně technicky i výkonově náročnější, než analýza běžných objemů dat. Má však i mnoho výhod, jako je například benevolence k nečistým datům. Pokud by se v malém objemu dat vyskytovaly chybné údaje, byl by to problém. V malých objemech dat jsou chyby ve výsledných analýzách znatelné, a z tohoto důvodu se lpí na čistotu dat. S velkými objemy dat není nutné řešit, zda je každý bit v objemu dat správný a odpovídá přesně měřené veličině. Ve velkém objemu dat se tyto nedokonalosti znatelně neprojeví. Jak říká citace z knihy [1] „*Raději více horších dat než méně lepších*“. Z tohoto důvodu není vždy nutné lpět na čistotě dat a lze pracovat i s datovými celky, které v menší míře obsahují závadná data. Tento fakt taktéž výstižně definuje citace z knihy [1] „*If you have too much data, the 'good enough' is good enough*“, neboli pokud máme velké množství dat, tak ač data nejsou naprosto čistá, tak to pro analýzu stačí. Jednoduše řečeno, pokud by byla testována sada o padesáti či tisíci vzorcích, tak je nutné vyžadovat vysokou přesnost. V takových to objemech dat se nedokonalosti projeví. V případě testování objemů dat o milionech vzorků už je důležitější množství dat než přesnost vzorků.

Jedním ze základních znaků velkých datových objemů je nemožnost tato data snadnou cestou přesouvat, či jinak duplikovat. Je proto potřeba tato data analyzovat, či jinak zpracovávat z míst, kde jsou umístěna. Zároveň je k analýze velkých datových objemů

zapotřebí jisté technické zázemí. Jelikož zpracování miliónů záznamů nelze provádět na běžném výpočetním zařízení, je často zapotřebí hodnotných technických prostředků. Tento fakt vystihuje popis „*Big Data*“ z knihy [3], který říká „*Big data is high-volume, high-velocity and high-variety information asstes that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.*“, neboli „*Big Data*“ je velkoobjemový, vysokorychlostní a velmi rozmanitý prostředek vyžadující nákladově efektivní a inovativní formy zpracování informací pro lepší přehled a rozhodování.

V této práci se analýza velkých datových objemů ukazuje především v podobě analýzy signalizačních dat z celulární rádiové sítě. (celulární rádiová síť detailněji viz kapitola 3). Tato rozsáhlá data umožňují zkoumání provozu této komunikační sítě pokrývající celou Českou republiku. Dávají tudíž možnost zkoumání velmi širokého spektra mobilních terminálů a buněk v celulární rádiové síti (viz kapitola 5). Struktura těchto signalizačních dat, a jak bylo s těmito velkými objemy dat zacházeno, je popsáno v kapitole 4.

3 Celulární radiová síť

Celulární radiová síť je telekomunikační síť rozdělena do vzájemně překrývajících se buněk. Od toho název celulární (latinské „*celula*“ = buňka) [5]. Tato síť obsahuje mnoho složek, a ty důležité pro pochopení a orientaci v rozebíraném tématu budou zmíněny v této kapitole.

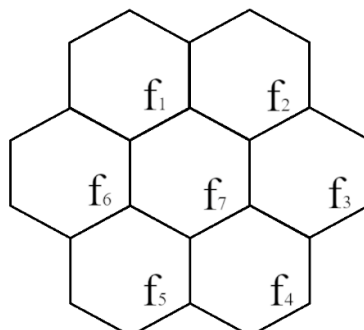
3.1 Mobilní terminál

Mobilní terminál je zařízení určeno k připojení se k celulární rádiové síti. K této síti se mobilní terminál připojí za pomoci základnové stanice (*Base Transceiver Station*, BTS) obsluhující dané geografické prostředí, ve kterém se daný mobilní terminál nachází. Pro komunikaci s celulární rádiovou sítí je nutnost přítomnosti účastnické identifikační karty (*Subscriber Identity Module*, SIM) [11] v mobilním terminálu. Karta SIM identifikuje mobilní terminál unikátním číslem přiděleným mobilním operátorem (*International Mobile Subscriber Identity*, IMSI), kterým lze nadále označovat uživatele [12]. Dále jsou kartou SIM přidělovány mobilnímu terminálu dočasné identifikační označení, jako například dočasný identifikátor mobilního terminálu (*Temporary Mobile Subscriber Identity*, TMSI), který je nejčastěji používán při komunikaci mezi mobilním terminálem a základnovou stanicí [13]. Dalším dočasným identifikátorem je identifikátor aktuální polohy mobilního terminálu (*Location Area Identity*, LAI), který identifikuje oblast, ve které se mobilní terminál nachází [9][8]. Tento identifikátor sestává z identifikačních kódů, kterými jsou identifikační kód země (*Mobile Country Code*, MCC), identifikační kód sítě (*Mobile Network Code*, MNC) a identifikační kód konkrétní oblasti (*Location Area Code*, LAC). LAI se nejčastěji používá ke sledování mobilního terminálu v síti, a napomáhá při přepojení mezi buňkami [9][4] (přepojení detailněji viz kapitola 3.5).

3.2 Buňkový systém

Komunikaci v celulární rádiové síti zajišťují základnové stanice, které jsou hlavním mozkiem každé buňky. Buňky představují geografické oblasti, které jsou pokryty rádiovým signálem od daných základnových stanic. Geografické prostředí je tedy pomyslně rozděleno do pravidelných útvarů, které představují buňky. Každá buňka disponuje vysílačem a přijímačem obsluhujícím danou buňku. Základnové stanice jsou poté řízeny základnovými jednotkami (*Base Station Controller*, BSC) [4], které zejména základnovým stanicím přidělují komunikační kanály. V případě sítí 2.generace (2G) komunikuje každá buňka na jemně rozdílné frekvenci, a to z důvodu zamezení rušení mezi buňkami [5] (generace sítí detailněji viz kapitola 3.3). Stejně frekvence mohou používat pouze buňky, které spolu přímo nesousedí. Z tohoto důvodu se začal využívat princip svazků, kdy každý svazek představuje uskupení buněk, kde každá buňka komunikuje na jiném frekvenčním kanálu. Tento model svazku se s dodržením

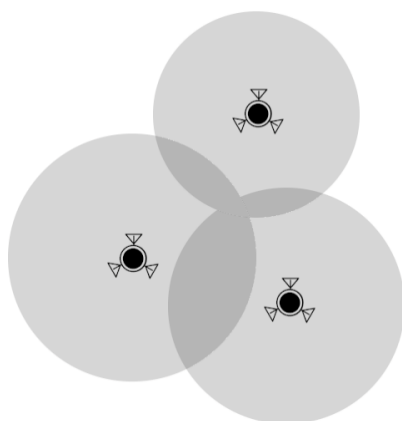
interferenčních podmínek může nadále opakovat o umísťovat vedle sebe. Při dodržení interferenčních podmínek se nikdy neobjeví dvě buňky stejné frekvence vedle sebe a nedojde tak k nežádoucímu rušení [7]. Na obrázku č. 1 je zobrazeno rozložení buněk ve svazku, kdy je zřetelné, že každá buňka využívá jiné frekvence.



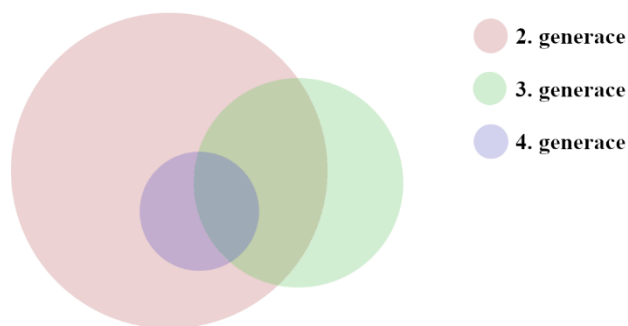
Obrázek 1. - Buňkový systém.

V případě sítí 3. generace (3G) a 4. generace (4G), komunikují buňky na stejných frekvencích, liší se však v použití vícenásobného přístupu (detailněji viz kapitola 3.7). Všechny buňky v sítích 3. generace mohou bez vzájemného rušení používat totožná frekvenční pásma díky metodě širokopásmového vícenásobného přístupu s kódovým dělením (*Wideband Code Division Multiple Access, WCDMA*). Buňky sítí 4. generace jako metodu vícenásobného přístupu využívají přístupové metody ortogonálního frekvenčního dělení (*Orthogonal Frequency Division Multiple Access, OFDMA*) a ortogonálního frekvenčního dělení na jedné nosné (*Single Carrier Orthogonal Frequency Division Multiple Access, SC-FDMA*) [4].

Buňky jsou často představovány jak šestiúhelníky, či jiné pravidelné uskupení, která jsou skládána vedle sebe, tak jak to je zobrazeno na obrázku č. 1. Takové tvary mají však pouze v ideálním případě. Většinou se formují do nepravidelných tvarů závislých na charakteru povrchu. Není poté možné, aby byly buňky poskládány vedle sebe, aniž by se nevytvářela místa bez pokrytí signálem. Z tohoto důvodu se buňky překrývají. Příklad překrývání je zobrazen na obrázku č. 2. Zároveň se často na mnoha místech vyskytuje více buněk o různých technologiích i velikostech. Ku příkladu se mohou překrývat buňky technologií 2. generace, 3. generace a 4. generace, tak jako tomu je na obrázku č. 3 (technologie detailněji viz kapitola 3.3).



Obrázek 2. - Překryv buněk.



Obrázek 3. - Překryv technologií.

3.3 Technologie

Vývoj technologií využívajících jakýchsi typů celulárních rádiových sítí sahají až do 70. let minulého století. V té době se začaly vyvíjet technologie 1. generace, a poprvé byly spuštěny do provozu na přelomu 70. a 80. let, kdy jako první mobilní analogový systém byl roku 1979 v Norsku a Švédsku spuštěn NMT (*Nordic Mobile Telephone*) [14]. Tyto technologie šířily analogový signál a umožňovaly komunikaci v podobě hovorů. Mezi technologie 1. generace patří také například AMPS (*Advanced Mobile Telephone System*) či TACS (*Total Access Communication System*). Od roku 1992 byly do provozu spuštěny technologie 2. generace, které již šířily digitální signál, a krom hovorů obstarávaly taktéž datovou výměnu. Hlavními zástupci této generace jsou technologie GSM (*Global System for Mobile Communication*), GPRS (*General Packet Radio Service*) či EDGE (*Enhanced Data rates for Global Evolution*). Od roku 2004 byly do provozu postupně spuštěny technologie 3. generace. Technologie 3. generace přinášely především mnohem rychlejší přenos dat a větší kapacitu. Hlavními zástupci této generace jsou UMTS (*Universal Mobile Telecommunication System*), HSPA (*High Speed Packet Access*) či technologie LTE (*Long Term Evolution*), často označovaná jako technologie 3,9. generace, která byla počáteční verzí budoucí technologie LTE-A (*Long Term Evolution – Advanced*), považující za technologii 4. generace. Tato technologie již přinesla přenosové rychlosti blízké se až 1 Gb/s, a byla postupně uváděna do provozu od roku 2010 [4]. Ve zbytku práce je pod označením LTE vždy myšlena technologie LTE-A.

Technologie jsou tedy děleny podle generací, při čemž v současné době jsou za pomoci celulární rádiové sítě šířeny pouze sítě technologií 2. generace, 3. generace a 4. generace. Provoz technologií 1. generace byl ukončen okolo roku 2006 [14].

Nejvíce vyskytujícími technologiemi v České republice, které jsou taktéž obsaženy v sadě signalizačních dat, se kterou se pracuje v praktické části této práce, jsou technologie GSM, UMTS a LTE. Technologii LTE se nadále dělí na technologie LTE 800, LTE 1800, LTE 2100 a LTE 2600, a to podle frekvenčních pásem, na kterých jsou technologie používány. Tato frekvenční pásma jsou poté přízviskem jejich označení.

GSM v České republice používají sítě na frekvencích okolo 900 MHz, 1800 MHz a 1900 MHz. Jsou vždy vytvořeny frekvenční pásma o šířce 25 MHz jak pro komunikaci od mobilního terminálu k BTS, tak od BTS k mobilnímu terminálu. Tyto dva kanály mají poté odstup 45 MHz. V případě GSM na frekvenci 900 MHz se jedná o pásmo 890 MHz až 915 MHz pro komunikaci od mobilního terminálu k BTS, a pásmo 935 MHz až 960 MHz pro komunikaci od BTS k mobilnímu terminálu. Pro komunikaci více mobilních terminálů najednou jsou poté použity přístupové metody kmitočtového dělení (*Frequency Division Multiple Access, FDMA*) a časového dělení (*Time Division Multiple Access, TDMA*) (viz kapitola 3.7) [4].

Technologie UMTS se obecně v Evropských zemích provozuje na frekvencích okolo 2100 MHz. Opět se zde dělí pásma na komunikaci od mobilního terminálu k BTS a opačně. Záleží však ještě na tom, zda UMTS pracuje v časovém duplexu (*Time Division*

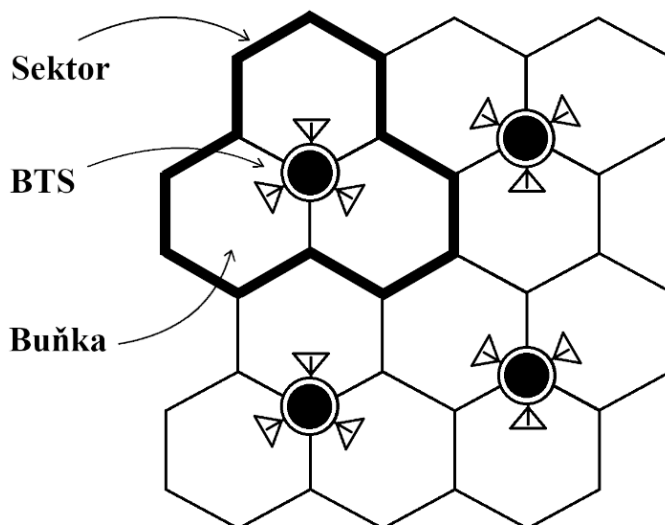
Duplex, TDD) či ve frekvenčním duplexu (*Frequency Division Duplex*, FDD). Pro UMTS pracující v TDD je vyhrazeno pásmo od 1900 MHz do 1920 MHz pro komunikaci od mobilního terminálu k BTS, a pásmo od 2010 MHz do 2025 MHz pro komunikaci v opačném směru. Pro UMTS pracující v FDD jsou vyhrazena širší frekvenční pásma, a to z důvodu, že FDD by mělo být používáno přednostně před TDD. Těmito pásmy jsou pásmo od 1920 MHz do 1980 MHz pro komunikaci od mobilního terminálu k BTS, a pásmo od 2110 MHz do 2170 MHz pro komunikaci od BTS k mobilnímu terminálu. Pro komunikaci více mobilních terminálů najednou je poté v technologii UMTS použito přístupové metody WCDMA (detailněji viz kapitola 3.7) [4].

Technologie LTE používají sítě na frekvencích, které jsou přízviskem k jejich jménům, a tudíž na frekvencích okolo 800 MHz, 1800 MHz, 2100 MHz a 2600 MHz. LTE, stejně jako UMTS, podporuje provoz za použití FDD a TDD. V případě LTE je komunikace rozdělena na typy rámců, Typ 1 pro FDD, a Typ 2 pro TDD. Pro oba typy je komunikace rozdělena na subpásma, při čemž pro Typ 1 jsou určena subpásma pro komunikaci od mobilního terminálu k BTS, a subpásma pro komunikaci v opačném směru, načež pro Typ 2 probíhá komunikace v obou směrech ve stejných subpásmech. Jedno subpásmo je pro tuto technologii široké 180 kHz, při čemž každé subpásmo disponuje subnosnými frekvencemi, které jsou od sebe vzdáleny 15 kHz. V každém subpásmu je tedy sdruženo 12 subnosných frekvencí. Pro velkokapacitní komunikaci je poté v technologii LTE použito přístupové metody OFDMA pro komunikaci od BTS k mobilnímu terminálu, a SC-FDMA pro komunikaci od mobilního terminálu k BTS. Obě tyto metody kombinují metody frekvenčního dělení a časového dělení (detailněji viz kapitola 3.7) [4].

3.4 Sektorizace

Základnové stanice mohou pro každou technologii disponovat jednou, všesměrovou anténou, nebo může disponovat více, směrovými anténami. V případě jedné, všesměrové antény, se nachází základnová stanice ve středu buňky pro danou technologii. Takového to rozdělení se však v dnešní době již moc nevyužívá. V případě, že základnová stanice disponuje více směrovými vysílači pro jednu technologii, vytváří více buněk pro danou technologii a samotná základnová stanice se poté nachází na společné hranici těchto buněk. Tento shluk buněk jedné technologie od jedné BTS se nazývá sektor, a tomuto procesu se říká sektorizace [7]. Ukázka sektorizace je zobrazena na obrázku č. 4. Základnová stanice takto za pomoci sektorizace může vytvořit až 6 buněk jedné technologie, po většinou se ale vytváří 3 či 4 buňky. Sektor nikdy neshlukuje buňky různých technologií, ale každá BTS může obstarávat více sektorů pro více technologií. Obecně platí, že v případě GSM, buňky od jedné BTS tvoří sektor, při čemž každá buňka funguje na jemně rozdílné frekvenci pro dodržení interferenčních podmínek [4].

V případě UMTS a LTE tvoří sektory buňky těchto technologií na jedné BTS, které disponují stejnou nosnou frekvencí.



Obrázek 4. - Sektorizace.

3.5 Procedura přepojení

V případě, že se mobilní terminál pohybuje, a dostane se do míst, kde již připojení k aktuální buňce není výhodné, elektronicky změní frekvenci a připojí se k buňce, která terminálu poskytuje lepší připojení. Přepojení mezi buňkami probíhá bez přerušení a nazývá se „handover“.

Přepojení se dělí na dva druhy:

- Meziuňkové přepojení – Tento typ přepojení nastává při přepojení mezi buňkami různých technologií nebo při přepojení mezi buňkami technologie GSM. Pokud takovéto přepojení nastane, uplatní se tzv. „tvrdý handover“, kdy se mobilní terminál musí první odpojit od první buňka, a až poté připojit na novou buňku. Dochází tedy k prodlevě, kdy mobilní terminál zůstává na malou chvíli bez připojení k síti. Tato prodleva je však v řádu desítek až stovek milisekund, tudíž pro běžného uživatele takřka nepozorovatelné. K „tvrdému handoveru“ dochází z důvodu, že mobilní terminál není schopný být připojený na dvou buňkách o rozdílné frekvenci zároveň.
- Vnitrobuňkové přepojení – Tento typ přepojení nastává při přepojení mezi buňkami technologie UMTS nebo mezi buňkami některé technologie LTE. V takovém případě je možné, aby byl mobilní terminál připojen na obě buňky, jak ke staré, tak i k nové, zároveň. Může tedy dojít k tzv. „měkkému handoveru“, kdy se mobilní terminál první připojí k nové buňce a až poté se odpojí od již nepotřebné buňky. Mobilní terminál může být připojen k obou buňkám zároveň z důvodu, že buňky technologií UMTS a LTE fungují na stejných frekvencích [4].

3.6 Velikost buněk

Buňky se dělí podle rozlohy geografické oblasti, kterou pokrývají signálem. Záleží však na mnoha faktorech. Jedna buňka se může rozkládat až 35 km od základnové stanice. V tomto případě však velmi záleží na profilu povrchu oblasti a také použité technologii. Na tak velké vzdálenosti se musí jednat o ploché, otevřené a nezastavěné oblasti. Ve městech a zastavěných oblastech buňky běžně dosahují poloměru okolo 1 km, na venkově okolo 10 km.

Nejpoužívanější označení pro rozlohu buněk jsou takovato:

- Makrobuňky - Rozloha takových to buněk dosahuje velikosti o průměru několika kilometrů. Tyto buňky jsou určeny pro nezastavěné, řídké osídlené oblasti, venkov, kde by výstavba více stanic BTS nebyla výhodná. Primárně jsou určeny pro rychleji se přemísťující mobilní terminály.
- Mikrobuňky - Takové to buňky dosahují rozlohy o průměru až 1 km. Jsou určeny především pro běžnou i hustou zástavbu, či oblasti s proměnným průřezem povrchu. Jsou určeny především pro běžně se pohybující mobilní terminály.
- Pikobuňky - Buňky velikosti piko dosahují rozlohy o průměru maximálně několika desítek metrů. Jsou určeny pro použití ve větších komerčních prostorách či v prostorách určených pro cestování, jako jsou letiště či vlaková nádraží, kde se sdružuje velké množství mobilních terminálů [4].

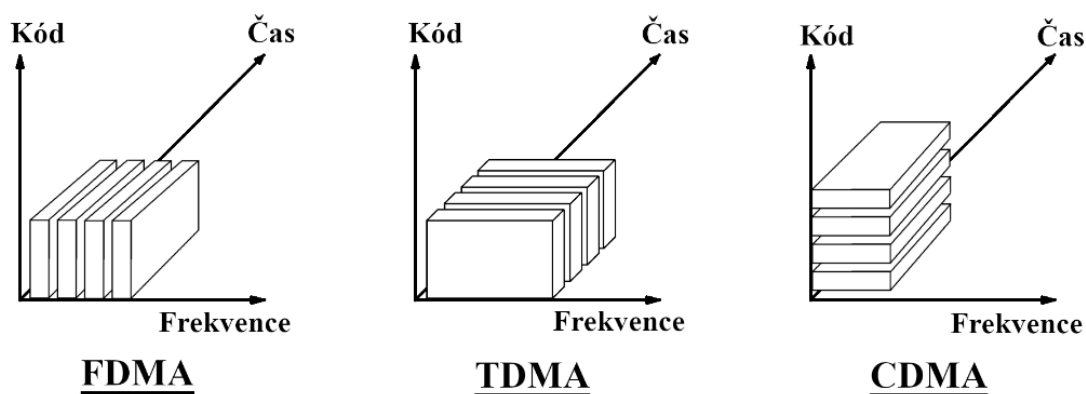
3.7 Přístupové metody

Aby bylo možné zajistit komunikaci více mobilních terminálů se základnovou stanicí v jednom časovém okamžiku, byly vytvořeny přístupové metody, které umožňují tzv. vícenásobný přístup k síti.

K docílení vícenásobného přístupu existují následující metody (přístupové metody jsou graficky znázorněny na obrázku č. 5):

- Kmitočtové dělení (*Frequency Division Multiple Access, FDMA*) - Kmitočtové dělení rozdělí frekvenční pásmo na menší subpásma, kterým jsou následovně přiřazeny komunikační kanály s připojením uživatelů. Tato metoda je zejména použita v technologii GSM, kde se obě frekvenční pásma, jak pro komunikaci od mobilního terminálu, tak i v opačném směru, rozdělí na 124 subpásem o šířce 200kHz.
- Časové dělení (*Time Division Multiple Access, TDMA*) - Časové dělení rozdělí frekvenční kanál do časových slotů, při čemž každý uživatel má přidělen jeden časový slot a komunikace probíhá postupně. Použit je jeden nosný kmitočet pro více uživatelů. Tato metoda je taktéž použita v technologii GSM, kde se po rozdělení frekvenčního pásma na subpásma za pomoci FDMA vytvoří v každém subpásmu 8 časových slotů.

- Kódové dělení (*Code Division Multiple Access*, CDMA) - Kódové dělení kóduje signály vysílané od mobilního terminálu podle daného kódovacího předpisu, kdy je pro každý kanál použit odlišný kódovací předpis. To je poté klíčem pro rozlišení kanálů na přijímací straně. Signály se díky tomu mohou vysílat na stejné frekvenci a nejsou časově závislé. Těto metody je zejména použito v technologii UMTS, pro kterou se však jedná o metodu širokopásmového kódového dělení (WCDMA), kdy je využito širšího pásma, čímž je docíleno vyšší kapacity kanálu a vyšší přenosové rychlosti než v případě jednoduchého dělení CDMA.
- Ortogonální frekvenční dělení (*Orthogonal Frequency Division Multiple Access*, OFDMA) - Jedná se o kombinaci kmitočtového a časového dělení. Kanál je v tomto případě rozdělen jak v časové oblasti do časový slot, tak v kmitočtové oblasti do subnosných frekvencí. Mobilnímu terminálu je poté přiřazen časový slot a několik subnosných frekvencí. Těto technologie využívají zejména technologie LTE, kdy se při komunikaci od BTS k mobilnímu terminálu využívá metody OFDMA, a při komunikaci od mobilního terminálu k BTS se využívá technologie SC-FDMA, kdy je využito pro komunikaci vždy pouze jedné nosné frekvence [4][6].



Obrázek 5. - Přístupové metody.

4 Zpracování a datové sady

Zpracování a analýza dat byla prováděna přes vzdálený přístup ve vývojovém prostředí RStudio na serveru, umístěném na Technické univerzitě Ostrava (VŠB). Veškeré procesy probíhaly na tomto serveru, a to z důvody velmi výkonného technického zázemí, kterým Technická univerzita v Ostravě disponuje. Veškerá data, se kterými bylo v této práci zacházeno, byly taktéž umístěny na tomto serveru v Ostravě. K serveru se přistupovalo za pomoci přímého datového kanálu zvaného tzv. VPN (*Virtual Private Network*). Výhodou této formy připojení je tvorba samostatného datového kanálu pro přímé spojení dvou koncových stanic a snížení rizik úniku dat či jiných problémů tradičnějších druhů připojení [17]. Veškeré analýzy bylo uskutečněny na platformě RStudio za použití programovacího jazyky R.

4.1 Použité funkce

Při analýze a zpracování dat za použití programovacího jazyka R bylo využito již vytvořených funkcí ze základní knihovny jazyka R, kterou je knihovna *base*, a jedná se o funkce *sapply()*, *lapply()*, *nrow()*, *colnames()*, *mean()*, *sum()*, *round()*, *substr()*, *is.data.frame()*, *as.numeric()*, *as.character()*, *as.logical()*, *do.call()*, *rbind()* a *function()*. Taktéž byly použity funkce z datové knihovny *tidyverse*, kterou vytvořil Hadley Wickham a jeho tým, a která nabízí velmi účinné nástroje při zpracování dat [16]. Jedná se tak o funkce *tibble()*, *filter()*, *distinct()* a *count()*. Grafy byly následně vygenerovány za pomoci funkcí ze základní knihovny *graphics*, a to funkcemi *plot()* a *barplot()*.

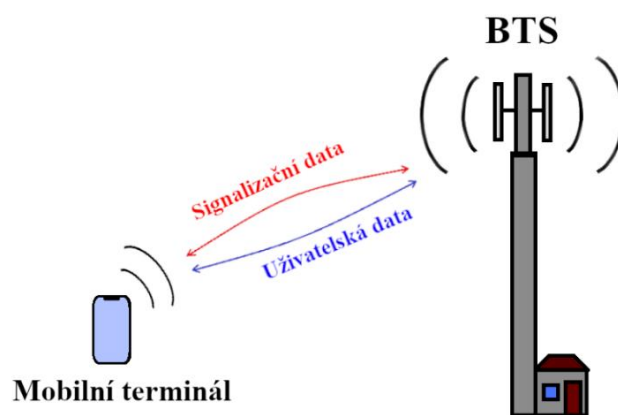
Funkce *mean()* a *sum()* jsou čistě matematické funkce, které pracují s numerickými hodnotami ve vektoru. *Mean()* vygeneruje průměrnou hodnotu hodnot ve vektoru, *sum()* součet hodnot ve vektoru. Funkce *round()* poté zaokrouhlí numerickou hodnotu na určený počet číselných míst. Funkce začínající na „as.“ zobrazují proměnnou v určeném formátu. Funkce *as.numeric()* zobrazuje proměnnou jako numerickou hodnotu, funkce *as.logical()* zobrazuje proměnnou jako log. proměnnou *TRUE/FALSE*, a funkce *as.character()* zobrazuje proměnnou jako řetězec. Funkce *substr()* poté může vyjmout či nahradit určenou část řetězce. Další funkce pracují s datovou strukturou tabulek. Samotná funkce *tibble()* vytvoří datovou tabulku se zadanými hodnotami. Funkce *nrow()* může poté spočítat počet řádků, kolik tabulka obsahuje. Funkce *count()* vygeneruje seznam obsahující počet unikátních proměnných, či počet unikátních kombinací proměnných, které se v tabulce nachází. Pomocí funkce *filter()* je možné uskutečnit filtraci datové tabulky, při čemž v tabulce ponechá pouze hodnoty odpovídající zadané podmínce pro definovaný sloupec. Funkce *distinct()* zase v datové tabulce ponechá pouze jeden záznam pro každou unikátní proměnnou v definovaném sloupci. Funkce *colnames()* umožňuje pojmenovat či přejmenovat jednotlivé sloupce v datové tabulce. *Is.data.frame()* je čistě ověřovací funkcí, která vygeneruje na výstup *TRUE/FALSE* v závislosti na tom, zda se jedná o datovou tabulku, či nikoli. Funkce *split()*

umožňuje rozdělit datovou tabulku na dílčí tabulky podle určené skupiny proměnných, neboli podle proměnných v jednom sloupci, a vytvořit datový list, který obsahuje tyto dílčí tabulky. Každá dílčí tabulka v listu bude tedy poté obsahovat záznamy pro jednu unikátní proměnnou z definovaného sloupce. Tato funkce je velmi důležitým nástrojem v několika analýzách zobrazených v kapitole 5. Funkce *lapply()* a *sapply()* poté umožňují práci se všemi tabulkami či kterýmikoli objekty v listu najednou. Na každý objekt v listu uplatní funkci, která je ve funkci definována jako vnitřní funkce. *Sapply()* poté nabízí dodatekové možnosti, jako je například určení struktury výstupu funkce. Za pomoci funkce *do.call()* a *rbind()* je následně možné zavolat výsledek těchto funkcí, který může taktéž mít strukturu listu, a dílčí tabulky spojit do jedné tabulky. A nakonec funkce *function()*, která je velmi důležitou funkcí, a to z důvodu, že umožňuje vytvoření vlastních funkcí. Grafy pro znázornění vygenerovaných výsledků byly vytvořeny za pomoci funkcí *plot()* a *barplot()*. *Plot()* vytvoří běžný graf, kde je možné určit, zda se hodnoty v grafu kreslí jako body, či jsou pospojovány liniemi. *Barplot()* vytvoří graf, kde výsledné hodnoty zastupují sloupce.

Funkce byly v průběhu práce různými způsoby modifikovány k vygenerování potřebných výstupů, nebo byly využity při tvorbě vlastních funkcí. Použití funkcí je popsáno u jednotlivých analýz viz kapitola 5.

4.2 Datové sady

V průběhu této práce se pracovalo se signalizačními daty z rádiového rozhraní celulární rádiové sítě obstaranými od jednoho operátora. Signalizační data vycházejí z účastnické signalizace, která představuje komunikaci mezi mobilním terminálem a BTS. Tato komunikace přenáší informace, které jsou nezbytné ke správnému řízení a funkci celulární rádiové sítě [15]. Důležitým poznatkem je, že signalizační a uživatelská data jsou posílána mezi mobilním terminálem a BTS separátně. Výskyt signalizace je demonstrován na obrázku č. 6.



Obrázek 6. - Signalizační data.

Signalizační data obsahují informace, se kterými pracují orgány celulární rádiové sítě, zejména základnová jednotka BSC, a zajišťuje terminálům nejvýhodnější připojení. Jedná se například o informace udávající polohu mobilních terminálů či informace o připojení mobilních terminálů k jednotlivým buňkám.

Datová sada se signalizačními daty od operátora byla pro tuto práci zjednodušena, a obsahovala chronologicky seřazené záznamy s těmito údaji:

- Identifikační řetězec mobilního terminálu (identifikátor IMSI)
- Identifikační řetězec buňky (ID buňky, identifikátor buňky)
- Čas strávený mobilním terminálem připojením na buňce
- Počet signalizačních zpráv vyměněných se základnovou stanicí

Identifikátor IMSI je řetězec, který nebylo možné v celém průběhu práce přiřadit k existujícímu mobilnímu terminálu, či fyzické osobě.

Identifikační řetězec buňky sestává z částí označující BTS, která buňku spravuje, technologii, kterou buňka šíří, a směrový vysílač dané buňky. Část unikátně označující BTS tvoří také označení geografické oblasti, ve které se BTS nachází. Toto označení oblasti představují první tři znaky řetězce („PAR“ – Pardubice). Kombinací s následujícími třemi znaky je utvořen unikátní identifikátor BTS („PARab1“). Označení technologie je v identifikačním řetězci buňky obstaráno sedmým znakem řetězce, který vždy tvoří číslice („5“ – LTE). Z tohoto znaku lze taktéž odečíst velikost buňky („5“ – mikro). Následující znak ID buňky přidává v případě UMTS či LTE označení nosné frekvence („1“ – LTE 800). Poslední znak řetězce označuje směrový vysílač („C“). Identifikační řetězec buňky může poté vypadat například takto:

„PARab151C“

Datová sada pro celý den byla rozdělena do tří časových oken. Prvním časovým oknem bylo ranní časové okno od 00:00 hodin do 05:00 hodin. Druhé, polední časové okno pokrývalo rozmezí od 05:00 hodin do 19:00 hodin. Třetí večerní časové okno obsahovalo záznamy od 19:00 hodin do 24:00 hodin.

Signalizační data pochází ze záznamu operátora z data 15.1.2020, čímž byla středa. Středa představuje průměrný den, který běžně není ovlivněn náhlými událostmi či velkými změnami v chování uživatelů mobilních terminálů, jako tomu může být u víkendových dní, či v pondělí nebo pátek. Záznam ze středečního dne by měl tudíž dobře reflektovat běžné chování mobilních terminálů, a z toho důvodu byl vybrán pro testování v této práci.

Datová sada obsahuje při součtu všech časových oken skoro 94,5 milionů záznamů, kde se vyskytuje přes 5 milionů jedinečných mobilních terminálů a více jak 57 tisíc jedinečných buněk. Takto objemná datová sada bude využita pouze při analýze s výsledkem rozdělení uživatelů do skupin podle chování v kapitole 5.4. Ve všech ostatních analýzách v kapitole 5 bude analyzována méně objemná datová sada, vytvořena ovzorkováním původní datové sady. Tato méně objemná datová sada obsahuje 100 tisíc záznamů pro každé časové okno, tudíž 300 tisíc záznamů pro celý den, kde se vyskytují záznamy pro více jak 54 tisíc jedinečných mobilních terminálů s informacemi o 39 tisíci jedinečných buňkách. Zmenšená datová sada byla ovzorkována tak, aby reflektovala skutečné chování mobilních terminálů v celulární rádiové síti i s poskytnutím menšího objemu dat. Použití méně objemné datové sady bylo zvoleno z důvodu snížení

potřebného početního výkonu a zkrácení doby výpočtu při analýzách. Srovnání datových sad je znázorněno v tabulce č. 1. Tabulka č. 2 udává, v jakých kapitolách byla použita která datová sada. Původní datová sada je v tabulkách označena oranžovou barvou, ovzorkovaná datová sada je označena barvou modrou.

Tabulka 1. - Charakteristiky datových sad.

Datové sady pro celý den	Počet záznamů	Počet mobilních terminálů	Počet buněk
Původní sada	94,5 milionů	5 milionů	57 tisíc
Ovzorkovaná sada	300 tisíc	54 tisíc	39 tisíc

Tabulka 2. - Použití datových sad v jednotlivých kapitolách.

Kapitoly, ve kterých byly analyzovány datové sady	Analyzované datové sady
5.1 Počet buněk jednotlivých technologií	Ovzorkovaná sada
5.2 Rozložení záznamů uživatelů mezi jednotlivými sítěmi	Ovzorkovaná sada
5.3 Identifikace uživatelských skupin	Původní sada

5 Analýzy a výstupy

Tato kapitola obsahuje analýzy prováděné na datových sadách signalizačních dat, které představují praktickou část této bakalářské práce. Výsledkem analýz je jakási predikce chování mobilních terminálů v souvislosti s použitím jednotlivých technologií, frekvencí, či na základě podobných vzorů chování vůči celulární rádiové síti.

5.1 Počet buněk jednotlivých technologií

Celulární rádiová síť je od základu zkoumána jako síť tvořená buňkami či sektory. Je proto více než užitečné vědět, jaké buňky či sektory tuto komplexní síť tvoří. Buňky se liší použitou technologií, rozlohou, či azimutem směrového vysílače. Sektory jsou poté tvořeny těmito buňkami.

Na začátku testování bylo provedeno za pomoci funkcí *distinct()*, *nrow()*, *substr()* a *filter()* vygenerování celkového počtu identifikátorů IMSI, záznamů, buněk, BTS a sektorů přítomných v testované datové sadě za celý den, tudíž za všechny tři časová okna dohromady. Použití funkcí je znázorněno v Příloze A, bod (1). Výsledek této analýzy demonstruje tabulka č. 3.

Tabulka 3. - Charakteristika datové sady.

Počet IMSI	Počet záznamů	Počet BTS	Počet buněk	Počet sektorů
54 202	299 997	8 672	39 108	19 033

Následně se přešlo k určení počtu buněk jednotlivých technologií. Základním kamenem této analýzy bylo upravení datové sady na vhodnou datovou strukturu. Jelikož je vhodné tuto analýzu provádět na záznamy z každé BTS zvlášť, byla datová sada rozdělena na dílčí datové tabulky za pomoci funkce *split()*. V tomto případě se rozdělila datová sada podle části identifikátoru buňky, získané za pomoci funkce *substr()*, označující BTS. Výstupem je poté list s dílčími datovými tabulkami obsahujícími záznamy z jednotlivých BTS. Použití funkce je uvedeno v Příloze A, bod (2).

V následujícím kroku bylo potřebné z tabulek obsažených v listu vygenerovat informaci, jaké technologie jsou provozovány na daných BTS, a kolik buněk pro každou přítomnou technologii BTS spravuje. K tomuto účelu bylo potřeba použít funkce, která operuje se všemi datovými tabulkami v listu najednou, a tudíž funkce *lapply()*. Jako vnitřní funkce byla použita vlastní funkce vytvořena z funkcí *function()*, *nrow()*, *substr()*, *distinct()*, *filter()* a *tibble()*. Následně byl výsledek zavolán funkcí *do.call()* s použitím vnitřní funkce *rbind()*. Celé použití je zaznamenáno v Příloze A, bod (3). Výsledkem je seznam všech BTS, při čemž ke každé BTS je zaznamenáno, kolik buněk pro každou technologii BTS spravuje. Z tohoto seznamu byly následně vygenerovány hodnoty počtu buněk jednotlivých technologií (tento výsledek by bylo možné získat i jednodušším způsobem,

takto je ale možné zpětně zjistit, kolik buněk jednotlivých technologií vybrané BTS spravují). Pro vygenerování tohoto výsledku, neboli tabulky s počty buněk jednotlivých technologií, bylo použito funkcí *tibble()* a *sum()*. Při použití funkce *sum()* byl jako vektor pro sečtení určen vždy sloupec zastupující jednu technologii. Použití funkcí pro vytvoření tabulky obsahující počty buněk jednotlivých technologií je uvedeno v Příloze A, bod (4).

Totožným způsobem byla vygenerována tabulka pro počet sektorů jednotlivých technologií, byly pouze upraveny podmínky při použití funkce *filter()* při tvorbě seznamu BTS. Seznam poté nebude udávat počet buněk každé technologie, kterou BTS spravuje, ale bude udávat počet sektorů každé technologie, které BTS spravuje. Tabulka č. 4 zobrazuje výsledek pro obě tyto analýzy, tudíž jak počet buněk každé technologie, tak i počet sektorů každé technologie. Buňky jsou v tabulce označeny modrou barvou, sektory jsou označeny barvou šedou.

Statistika nám odhaluje, že největší počet buněk dohromady tvoří síť LTE, které při součtu buněk pro všechna frekvenční pásma šíří celkem 17 356 buněk. To tvoří 44,4 % z celkového počtu buněk v datové sadě. Lze tedy říci, že okolo 44 % buněk v celé infrastruktuře celulární rádiové sítě v České republice zastupují buňky technologie LTE. Co se týče jednotlivých frekvencí technologie LTE, je nejvíce buněk této technologie provozováno technologií LTE 800, neboli technologií LTE na frekvencích okolo 800 MHz. Technologie LTE 800 v datové sadě obsahuje 8 386 buněk, což pokrývá skoro polovinu všech buněk technologie LTE, a svým počtem stále převyšuje počet buněk technologie UMTS. Druhou nejpoužívanější technologií technologie LTE je LTE 1800, neboli technologie LTE používající frekvence okolo 1800 MHz. Tato technologie v datové sadě obsahuje 4 984 buněk, a představuje tak skoro 29 % všech buněk technologie LTE v datové sadě. Skoro 22 % poté z celkového počtu buněk technologie LTE zastupuje technologie LTE 2100, a pouze okolo 1 % zastupuje technologie LTE 2600. S počtem 13 556 buněk je v datové sadě zastoupena technologie GSM. To tvoří více jak 34,6 % všech buněk v datové sadě. Lze taktéž tedy říci, že okolo 34 % buněk v celé infrastruktuře celulární rádiové sítě na území České republiky je zastoupeno buňkami technologie GSM. Nejméně buněk v datové sadě, co se týče technologie, je zastoupeno buňkami technologie UMTS, a to počtem 8 196 buněk. To v datové sadě představuje okolo 20 % z celkového počtu buněk. V infrastruktuře celulární rádiové sítě v České republice by tedy technologie UMTS měla zastupovat okolo pětiny buněk.

Testem na začátku této analýzy bylo specifikováno, že v datové sadě je obsaženo přesně 39 108 buněk. Výsledky v tabulce č. 4 lze tedy jednoduše ověřit součtem buněk všech technologií. Součet všech těchto buněk udává opět počet 39 108 buněk (viz tabulka č. 4).

Počty sektorů jednotlivých technologií podle očekávání odpovídají přibližně polovině počtu buněk. Je tomu tak z důvodu, že sektory v celulární rádiové síti velmi často shlukují 3 nebo 4 buňky. Jelikož v této datové sadě nejsou všechny buňky vyskytující se v infrastruktuře celulární rádiové sítě, průměrný počet buněk v sektoru se sníží, a tudíž i rozdíl mezi celkovým počtem buněk a celkovým počtem sektorů na technologii se zmenší. Výsledek analýzy lze opět ověřit sečtením počtu sektorů každé technologie

a porovnáním s celkovým počtem sektoru vygenerovaným v úvodní analýze. Po sečtení všech sektorů pro každou technologii vyjde opět počet 19 033 sektorů (viz tabulka č. 4).

Tabulka 4. - Počet buněk každé technologie.

	GSM	UMTS	LTE800	LTE1800	LTE2100	LTE2600	Součet
Počet buněk	13 556	8 196	8 386	4 984	3 788	198	39 108
Počet sektorů	6 367	3 985	4 392	2 484	1 710	95	19 033

5.2 Rozložení záznamů uživatelů mezi jednotlivými sítěmi

Tato kapitola popisuje použití technologií mobilními terminály, a to ať už se jedná o použití jednotlivých technologií, či o používání kombinací dvou či tří technologií.

Stěžejní pro analýzy v této kapitole je vytvoření seznamu všech IMSI, při čemž ke každému IMSI jsou přiřazeny informace, ke kterým sítím byl mobilní terminál v průběhu daného časového okna připojen. Toho bylo dosaženo za pomoci rozdělení analyzované datové sady na list datových tabulek podle identifikátoru IMSI mobilních terminálů. K tomu bylo použito funkce *split()*. Následně se použilo funkce *lapply()* se strukturovanou vnitřní funkcí, která byla tvořena z funkcí *filter()*, *nrow()*, *is.data.frame()*, *substr()*, *distinct()*, *as.logical()*, *as.character()* a *tibble()*. Výsledek této funkce, který představoval seznam všech IMSI s doplňkovými informacemi o použitých technologiích, byl následně zavolán pomocí funkce *do.call()* s použitím vnitřní funkce *rbind()*. Výsledný seznam s identifikátory IMSI disponuje pro každou IMSI informacemi ve formě logických proměnných *TRUE/FALSE*, které označují, jaké technologie v průběhu testovaného časového okna mobilní terminál s danou IMSI použil. Kompletní použití funkcí je zaznamenáno v Příloze A, bod (5).

Následné analýzy v této kapitole, byly provedeny na tento vytvořený seznam. Ve všech těchto analýzách bylo k dosažení výsledku využito filtrace, početních i logických funkcí, a to ať už se jednalo o analýzu počtu uživatelů jedné, dvou, či tří technologií.

Jelikož se v analýzách stanovují počty terminálů, které používají jednotlivé technologie, bylo před tím ještě potřeba zjistit, kolik mobilních terminálů každé časové okno obsahuje. Jsou však počítány pouze mobilní terminály, u nichž je v signalizačních datech udán minimálně v jednom záznamu platný identifikátor buňky, ke které se terminál připojil. Záznamy bez platného identifikátoru buňky totiž nejsou v těchto následujících analýzách možné zkoumat. Toto je podstatný detail, jelikož v datové sadě se vyskytuje velké množství terminálů, u kterých nejsou v záznamech udány identifikátory buněk, na které se připojily.

Pro zjištění tohoto celkového počtu identifikátorů IMSI se jako první na všechna časová okna použije funkce *filter()*, kterou se odeberou záznamy, které neobsahují identifikátor buňky. Poté se na jednotlivá časová okna použijí funkce *distinct()* a *nrow()*. Tím se docílí výsledku počtu mobilních terminálů v jednotlivých časových oknech. Tato statistika je zapsána v tabulce č. 5.

Tabulka 5. - Počet mobilních terminálů v časových oknech s platnými údaji ID buněk.

Časová okna	Ranní okno	Polední okno	Večerní okno	Celý den
Počet IMSI	27 493	11 375	21 756	39 790

Je pozorovatelné, že v poledním i večerním časovém okně je méně mobilních terminálů. To je způsobeno tím, že v základní datové sadě obsahovala tato časová okna více záznamů než ranní okno, tudíž po ovzorkování na stejný počet záznamů musí obsahovat menší počet terminálů.

5.2.1 Použití jedné technologie

Počet mobilních terminálů s identifikátory IMSI, používajících jednu technologii, byl vygenerován metodou logických funkcí uplatněných na seznam identifikátorů IMSI. V případě, že se hledal počet mobilních terminálů, které byly připojeny k některé technologii, použilo se jednoduché kombinace funkcí *filter()* a *nrow()*. Funkcí *filter()* se specifikovalo, že se log. proměnná pro technologii musí rovnat *TRUE*, a poté se za pomoci funkce *nrow()* spočítal počet identifikátorů IMSI, které danou podmínku splňovaly. Použití je znázorněno v Příloze A, bod (6). V případě, že se hledal počet mobilních terminálů, které byly připojeny pouze k jedné technologii v průběhu testovaného časového okna, bylo použito taktéž funkcí *filter()* a *nrow()*. Tentokrát se funkcí *filter()* nespecifikovalo pouze, že se log. proměnná pro danou technologii musí rovnat *TRUE*, ale taktéž, že se log. proměnné pro všechny ostatní technologie musí rovnat *FALSE*. Použití je znázorněno v Příloze A, bod (7). Takovou to metodou byl vygenerován výsledný počet mobilních terminálů, které v každém časovém okně použily některou z daných technologií, a počet mobilních terminálů, které použily pouze jednu z těchto technologií. Přehledný výsledek analýzy pro všechna časová okna je zaznamenán v tabulce č. 6. Tyto výsledky byly také pro vizuální srovnání časových oken graficky znázorněny za pomoci funkce *barplot()*, a jsou zobrazeny na obrázcích č. 7 až 10. V tabulce je pod označením „N“ zobrazen počet mobilních terminálů, které v konkrétním časovém okně použily danou technologii, a pod označením „Np“ je zobrazen počet mobilních terminálů, které v konkrétním časovém okně použily pouze dané technologie. V tabulce č. 7 je poté zobrazeno procentuální zastoupení mobilních terminálů, které použily pouze jedné dané technologie z celkového počtu mobilních terminálů, které použily dané technologii. Žlutou barvou jsou v tabulkách pro každou skupinu znázorněny technologie, které dosahují nejvyšších hodnot, a oranžovou technologie dosahující nejnižších hodnoty.

Mobilní terminály používající pouze jednu technologii často odpovídají terminálům, které jsou součástí různých nepohyblivých účelových systémů, či terminálům, které jsou používány uživateli zůstávajícími na jednom místě. Taktéž odpovídají terminálům, které disponují technickou podporou pouze jedné technologie. To se týká zejména starších zařízení, které disponují pouze podporou technologie GSM.

Z tabulky č. 6 lze vyčíst, že nejvíce mobilních terminálů používá právě technologii GSM. V tabulce č. 7 je poté znázorněno, že pro technologii GSM přibližně 80 % mobilních

terminálů, používající tuto technologii, nepoužívá žádnou jinou technologii. To je největší procentuální část ze všech technologií. Druhou nejvíce používanou technologií je technologie LTE. Technologii podle datové sady dohromady pro všechna frekvenční pásma za celý den použilo 17 214 mobilních terminálů. Využití technologie LTE se v tabulkách č. 6 i 7 dělí na jednotlivá frekvenční pásma. V tabulce č. 7 je poté tedy zobrazeno procentuální zastoupení mobilních terminálů, které nepoužily jiné technologie ani jiného frekvenčního pásma. Z této tabulky je zřetelné, že tato procenta jsou, oproti například GSM, velmi malá. Nejméně mobilních terminálů použilo technologie UMTS, jak je však z tabulky č. 6 zřetelné, nejedná se o nijak nízké číslo. Podobně jako u technologie GSM, i u UMTS je velké procento mobilních terminálů, které použily pouze této technologie. Konkrétně se jedná o přibližně 61 % mobilních terminálů z celkového počtu mobilních terminálů, které použily technologie UMTS.

Při porovnání výsledků časových oken v tabulce č. 7, lze upozorovat, že procentuální zastoupení mobilních terminálů, které používají pouze jedné technologie, je v poledním i večerním časovém okně značně menší, než v ranním časovém okně. To je nejspíše způsobeno vyšší aktivitou a pohybem uživatelů mobilních terminálů v těchto časových oknech

Na závěr je dobré se zaměřit na součet mobilních terminálů, které použily pouze jedné technologie za celý den, který je v tabulce č. 6 vyznačen zelenou barvou. Tento součet tvoří 30 857 mobilních terminálů. Takovýto počet tvoří z celkového počtu 39 790 mobilních terminálů závratných 77 %. Tudiž 77 % procent všech mobilních terminálů v datové sadě za den použilo pouze jedné technologie.

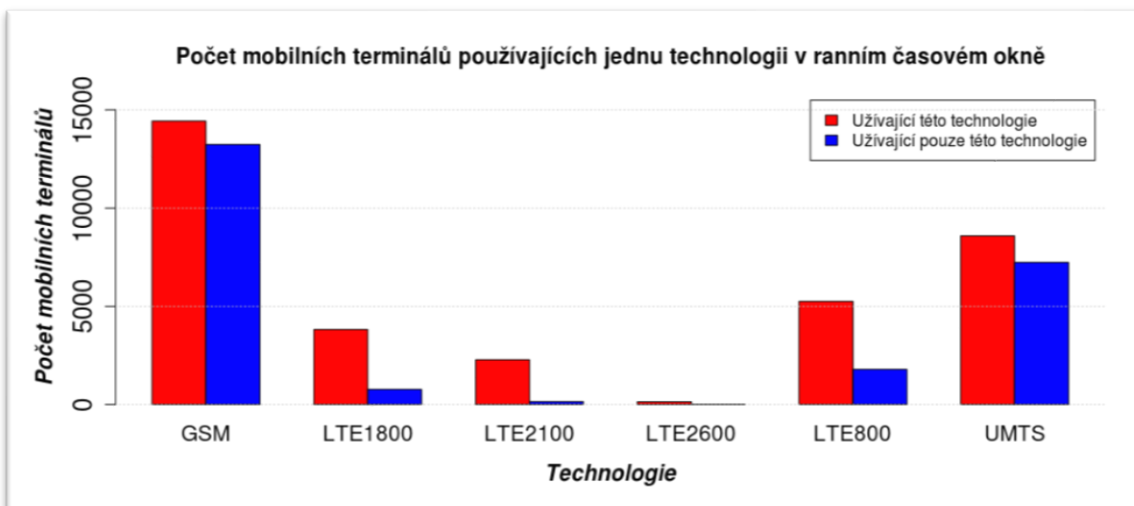
Tabulka 6. - Rozložení uživatelů mezi technologiemi při rozdělení do časových oken.

Technologie	Ranní okno		Polední okno		Večerní okno		Celý den	
	N	Np	N	Np	N	Np	N	Np
GSM	14 427	13 236	8 376	6 087	13 784	11 408	26 049	20 729
UMTS	8 592	7 239	4 243	2 032	8 101	5 917	13 512	8 251
LTE800	5 255	1 789	1 954	83	2 641	279	6 893	1 291
LTE1800	3 821	766	1 821	23	2 426	210	5 684	494
LTE2100	2 277	145	1 528	5	2 003	79	4 198	88
LTE2600	133	4	221	0	157	1	439	4
Součet	34 505	23 179	18 143	8 230	29 112	17 894	56 775	30 857

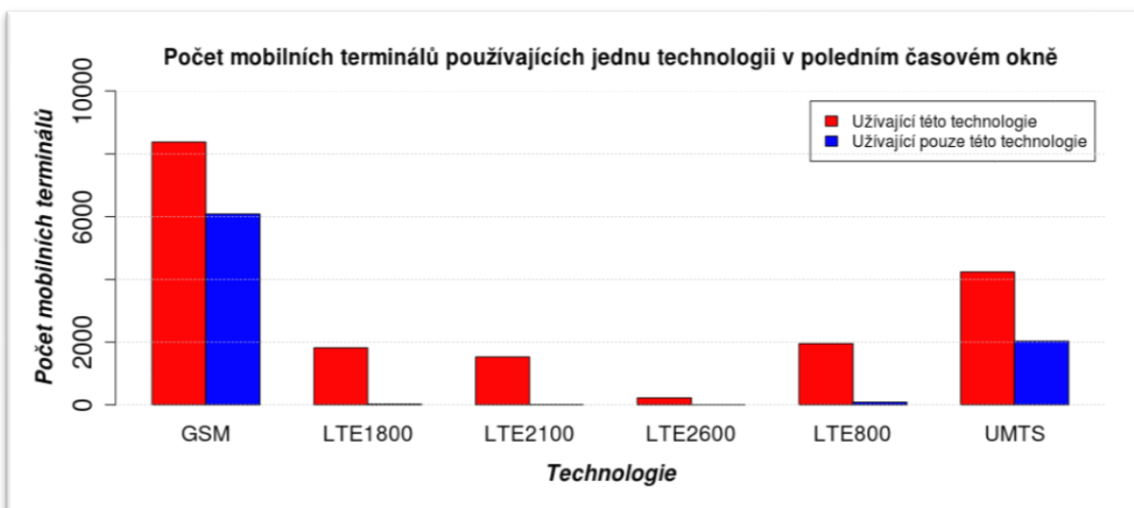
- N – počet mobilních terminálů používajících danou technologii
- Np – počet mobilních terminálů používajících pouze danou technologii

Tabulka 7. - Procentuální zastoupení mobilních terminálů používajících pouze jedné technologie.

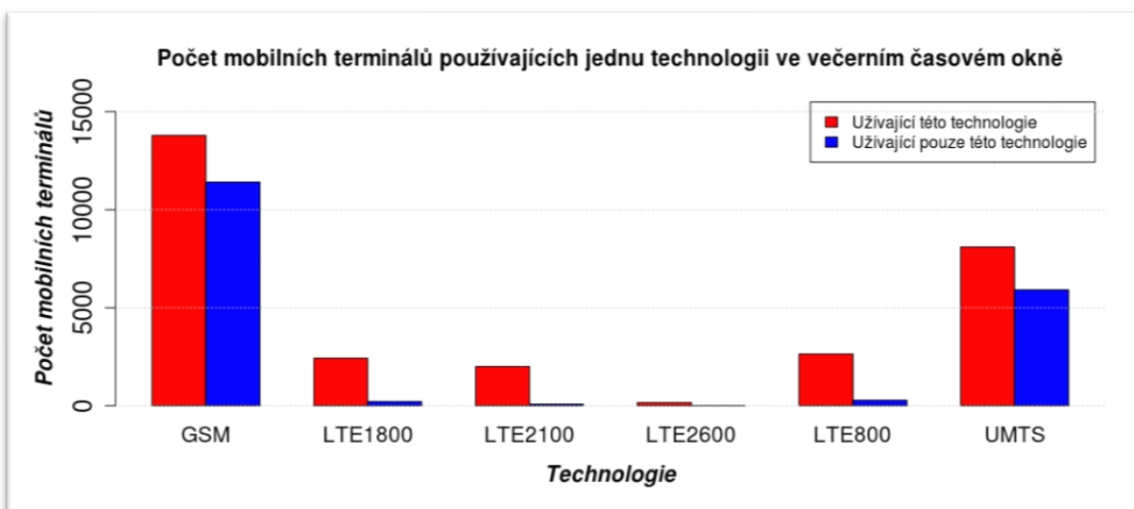
Technologie	Ranní okno	Polední okno	Večerní okno	Celý den
GSM	91,7 %	72,7 %	82,8 %	79,6 %
UMTS	84,3 %	47,9 %	73,0 %	61,1 %
LTE800	34,0 %	4,2 %	10,6 %	18,7 %
LTE1800	20,0 %	1,3 %	8,7 %	8,7 %
LTE2100	6,4 %	0,3 %	3,9 %	2,1 %
LTE2600	3,0 %	0,0 %	0,6 %	0,9 %



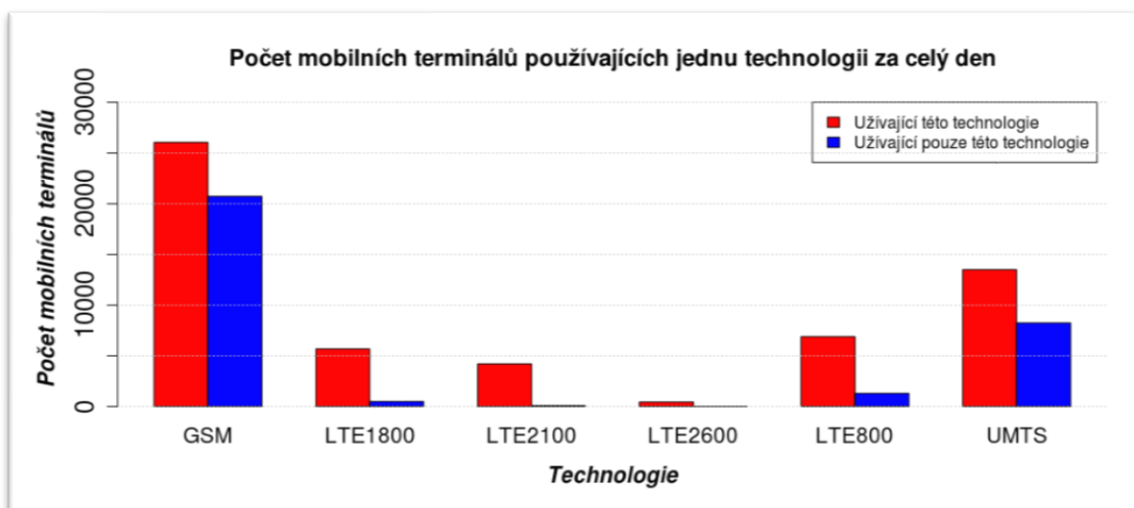
Obrázek 7. - Počet mobilních terminálů používajících jednu technologii v ranním časovém okně.



Obrázek 8. - Počet mobilních terminálů používajících jednu technologii v poledním časovém okně.



Obrázek 9. - Počet mobilních terminálů používajících jednu technologii ve večerním časovém okně.



Obrázek 10. - Počet mobilních terminálů používajících jednu technologii za celý den.

5.2.2 Použití dvou technologií

Tato statistika poukazuje na pravděpodobnost kombinace dvou technologií. Počet mobilních terminálů s identifikátorem IMSI, které v analyzovaných časových oknech použily kombinaci dvou technologií, byl vygenerován pomocí stejných funkcí, jako výsledek předchozí analýzy. Aneb v případě, že se hledal počet mobilních terminálů, které v analyzovaném časovém okně použily kombinace některých dvou technologií, použilo se funkce *nrow()* a *filter()*, při čemž ve funkci *filter()* bylo specifikováno, že log. proměnná pro dané dvě technologie musí být rovna *TRUE*. Poté se spočítal počet identifikátorů IMSI, které tyto podmínky splňovaly. V případě, že se hledal počet mobilních terminálů, které v analyzovaném časovém okně použily pouze těchto dvou technologií, a žádné jiné, specifikovalo se ve funkci *filter()*, že se log. proměnná pro tyto dvě technologie musí rovnat *TRUE*, a však všechny ostatní log. proměnné pro jiné technologie se musí rovnat *FALSE*. Takto byl vygenerován výsledný počet mobilních terminálů, které v každém časovém okně použily některé z možných kombinací technologií, a počet mobilních terminálů, které použily pouze některou z daných kombinací. Přehledný výsledek analýzy pro všechna časová okna je zaznamenán v tabulce č. 9. Tyto výsledky byly také pro vizuální srovnání časových oken graficky znázorněny za pomoci funkce *barplot()*, a jsou zobrazeny na obrázcích č. 11 až 14. V tabulce č. 10 je poté zobrazeno procentuální zastoupení mobilních terminálů, které použily pouze daných dvou technologií, z celkového počtu mobilních terminálů, které použily tyto dvě technologie. V tabulkách a grafech jsou jednotlivé technologie označeny zkratkami, kde „G“ náleží GSM, „U“ náleží UMTS, a technologie LTE jsou vždy označeny písmenem „L“ a nenulovými číslicemi v hodnotě frekvenčního pásma, tudíž například „L21“ náleží LTE 2100. Znakem „+“ jsou vždy technologie přiřazeny k sobě. Tudíž například „G+L8“ označuje kombinaci technologií GSM a LTE 800. Taktéž je opět použito označení „N“, jež označuje počet mobilních terminálů, které použily danou kombinaci technologií, a „Np“, jež označuje počet mobilních terminálů, které již nepoužily jiné než této kombinace. Žlutou barvou jsou v tabulkách v každé skupině

pro všechna časová okna znázorněny kombinace, které dosahují nejvyšších hodnot, a oranžovou kombinace dosahující nejnižších hodnot. Pokud by některé kombinace měly shodné hodnoty, a jednalo by se o kombinace hodné znázornění, může se toto znázornění objevit u více kombinací v jednotlivých skupinách.

Lze odvodit, že mobilní terminály, které použily různé kombinace technologií, jsou zástupci skupiny mobilních terminálů, které v předchozí analýze nepoužily pouze jedné technologie. Tudíž pro každé časové okno počet mobilních terminálů, které použily více než jednu technologii, tvoří vždy rozdíl mezi celkovým počtem terminálů v jednotlivých časových oknech (viz tabulka č. 5), a celkovým počtem terminálů užívajících pouze jednu technologii v jednotlivých časových oknech (viz tabulka č. 6). Počet takových to terminálů pro každé časové okno je zaznamenán v tabulce č. 8.

Tabulka 8. - Počet mobilních terminálů v časových oknech používajících více než jednu technologii.

Časová okna	Ranní okno	Polední okno	Večerní okno	Celý den
Počet IMSI	4 314	3 145	3 862	8 933

Z tabulky č. 9 i 10 si lze opět všimnout markantního rozdílu mezi ranním časovým oknem a ostatními časovými okny. Tento rozdíl je stejně jako v předchozí analýze způsoben vyšší aktivitou uživatelů mobilních terminálů v poledním a večerním časovém okně. Mnoho uživatelů se v době ranního časového okna spíše zdržuje na jednom místě a nevykazuje vysokou aktivitu. Nejvíce mobilních terminálů za celý den použilo kombinaci technologií LTE 800 a LTE 1800, kterou celkem využilo 4 810 mobilních terminálů. Ze všech mobilních terminálů používajících těchto dvou technologií za celý den přibližně 16,5 % mobilních terminálů nepoužilo již jiné technologie. Druhou nejpočetnější skupinou v počtu terminálů podle tabulky č. 9 je skupina mobilních terminálů používající technologie GSM a UMTS. Za celý den jich v datové sadě bylo 4 309. Pro tuto skupinu mobilních terminálů je velmi vysoké procento terminálů, které nepoužilo žádné jiné technologie krom těchto dvou. Konkrétně se jedná o více jak 67 %. Je však důležité upozornit na fakt, že v ranním okně tato kombinace technologií velmi zaostávala za jinými kombinacemi, které ve statistice za celý den mají menší počet mobilních terminálů využívajících těchto technologií. Takovými skupinami, které následují v počtu mobilních terminálů, jsou skupiny používající kombinace LTE 800 s LTE 2100 a kombinace LTE 1800 a LTE 2100. Tyto skupiny již disponují velmi nízkým počtem terminálů, které nepoužily jiné technologie. Jen zřídka se technologie kombinují s technologií LTE 2600. Tomu odpovídá i nejméně používaná kombinace, kterou podle výsledků za celý den je kombinace technologií UMTS a LTE 2600. Tyto dvě technologie použilo za den pouhých 207 mobilních terminálů. Nízký počet terminálů využil i všech dalších kombinací technologie LTE 2600, zejména kombinaci s GSM použilo pouze o 14 terminálů více než v případě UMTS, a tedy 221. To je samozřejmě výsledkem komplexního nízkého používání technologie LTE 2600, a také nízkým počtem buněk, které tato technologie šíří.

Při srovnání grafů na obrázcích č. 11 až 14 lze zpozorovat, že se technologie více kombinují v poledním či večerním časovém okně, při čemž v ranním okně dochází především ke kombinaci technologií technologie LTE. Při pohledu na grafy pro ranní a polední časové okno lze detekovat, jak již bylo zmíněno, markantní nárůst použití kombinace GSM a UMTS.

Lze tedy konstatovat, že uživatelé častěji mění technologie v průběhu poledního či večerního okna. Dominantními kombinacemi jsou kombinace technologií LTE, avšak u těchto kombinací dvou technologií velmi často mobilní terminál použije i jakékoli třetí technologii. Toto se netýká kombinace technologií GSM s UMTS, kdy při použití této kombinace je pravděpodobnost, že nedojde k použití již žádné další technologie, více jak 67 %. Nejméně mobilních terminálů poté využívá kombinací s technologií LTE 2600.

Na závěr je opět dobré se zaměřit na součet mobilních terminálů, které za celý den použily pouze dvě technologie. Součet terminálů je v tabulce č. 9 vyznačen zelenou barvou. Tento součet odpovídá počtu 4 534 mobilních terminálů, což z celkového počtu téměř 40 tisíc terminálů odpovídá přibližně 11 %.

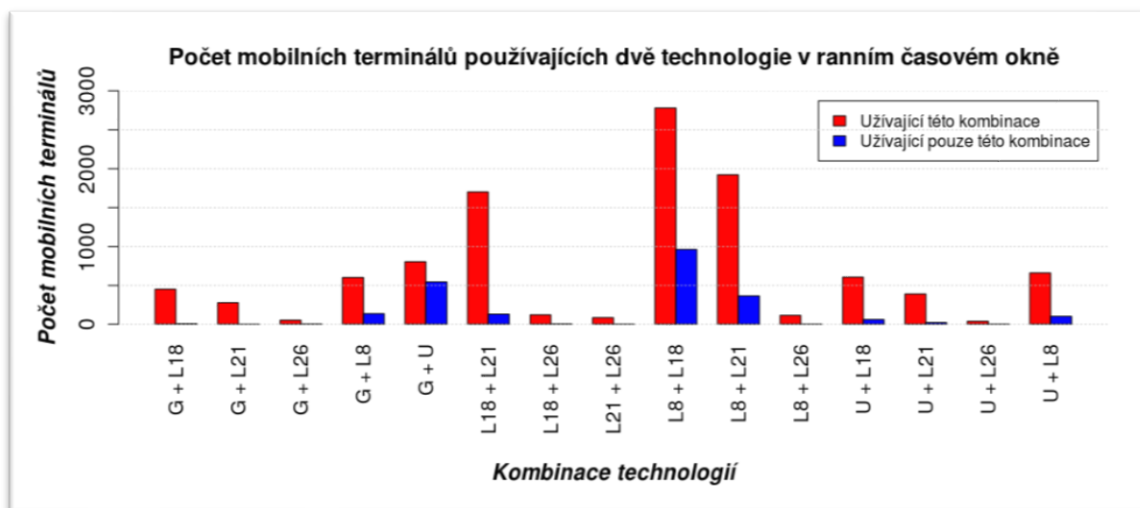
Tabulka 9. - Rozložení uživatelů mezi kombinacemi dvou technologií při rozdělení do časových oken.

Kombinace	Ranní okno		Polední okno		Večerní okno		Celý den	
	N	Np	N	Np	N	Np	N	Np
G + U	805	545	1 912	1 161	1 773	1 286	4 309	2 900
G + L8	602	136	1 080	38	994	103	2 275	199
G + L18	451	7	1 009	6	886	16	2 039	21
G + L21	277	0	882	1	762	5	1 687	3
G + L26	52	3	104	0	55	0	221	3
U + L8	661	101	963	12	801	9	2 095	99
U + L18	607	60	944	27	777	28	2 036	82
U + L21	390	22	811	11	665	9	1 651	33
U + L26	38	1	107	0	56	2	207	2
L8 + L18	2 780	963	1 703	114	2 020	205	4 810	794
L8 + L21	1 922	365	1 455	30	1 775	99	3 823	282
L8 + L26	114	1	216	0	144	3	414	2
L18 + L21	1 701	130	1 424	13	1 715	62	3 666	111
L18 + L26	121	4	214	0	150	2	418	2
L21 + L26	84	1	202	0	136	0	370	1
Součet	10 605	2 339	13 026	1 413	12 709	1 829	30 021	4 534

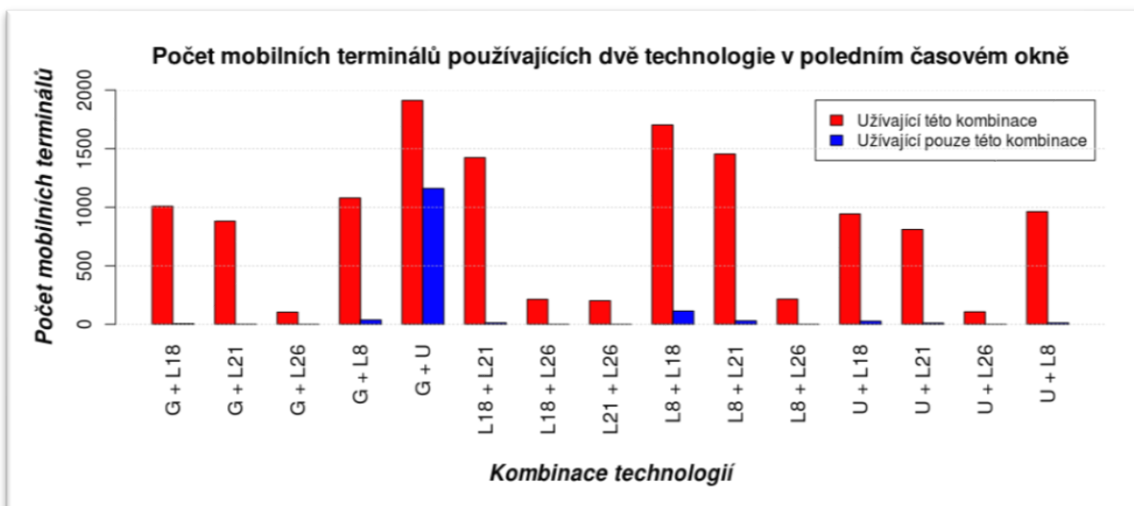
- N – počet mobilních terminálů používajících danou kombinaci technologií
- Np – počet mobilních terminálů používajících pouze danou kombinaci technologií

Tabulka 10. - Procentuální zastoupení mobilních terminálů používajících pouze daných kombinací technologií.

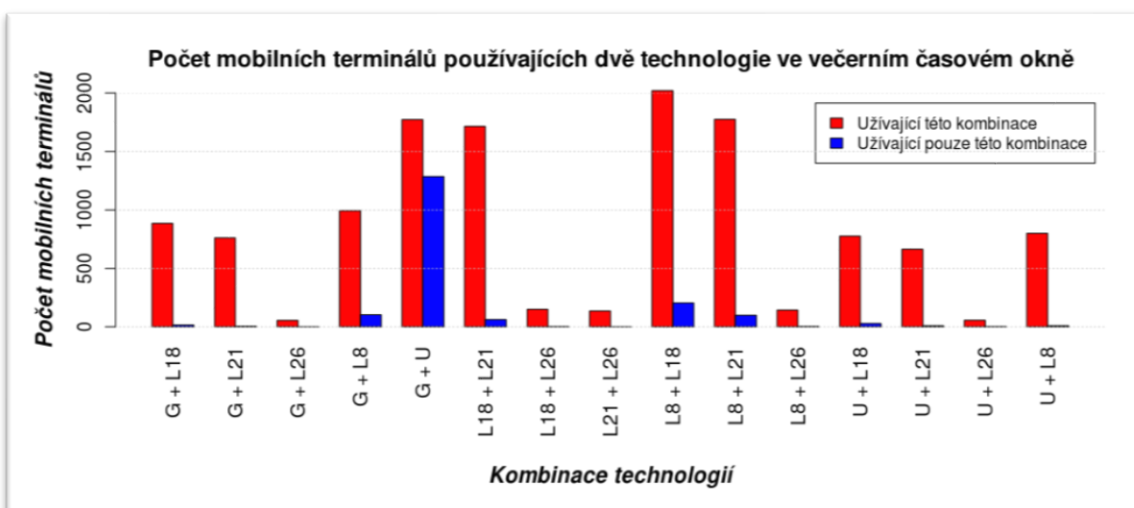
Kombinace	Ranní okno	Polední okno	Večerní okno	Celý den
G + U	67,7 %	60,7 %	72,5 %	67,3 %
G + L8	22,6 %	3,5 %	10,4 %	8,7 %
G + L18	1,6 %	0,6 %	1,8 %	1,0 %
G + L21	0,0 %	0,1 %	0,7 %	0,2 %
G + L26	5,8 %	0,0 %	0,0 %	1,4 %
U + L8	15,3 %	1,2 %	1,1 %	4,7 %
U + L18	9,9 %	2,9 %	3,6 %	4,0 %
U + L21	5,6 %	1,4 %	1,4 %	2,0 %
U + L26	2,6 %	0,0 %	3,6 %	1,0 %
L8 + L18	34,6 %	6,7 %	10,1 %	16,5 %
L8 + L21	19,0 %	2,1 %	5,6 %	7,4 %
L8 + L26	0,9 %	0,0 %	2,1 %	0,5 %
L18 + L21	7,6 %	0,9 %	3,6 %	3,0 %
L18 + L26	3,3 %	0,0 %	1,3 %	0,5 %
L21 + L26	1,2 %	0,0 %	0,0 %	0,3 %



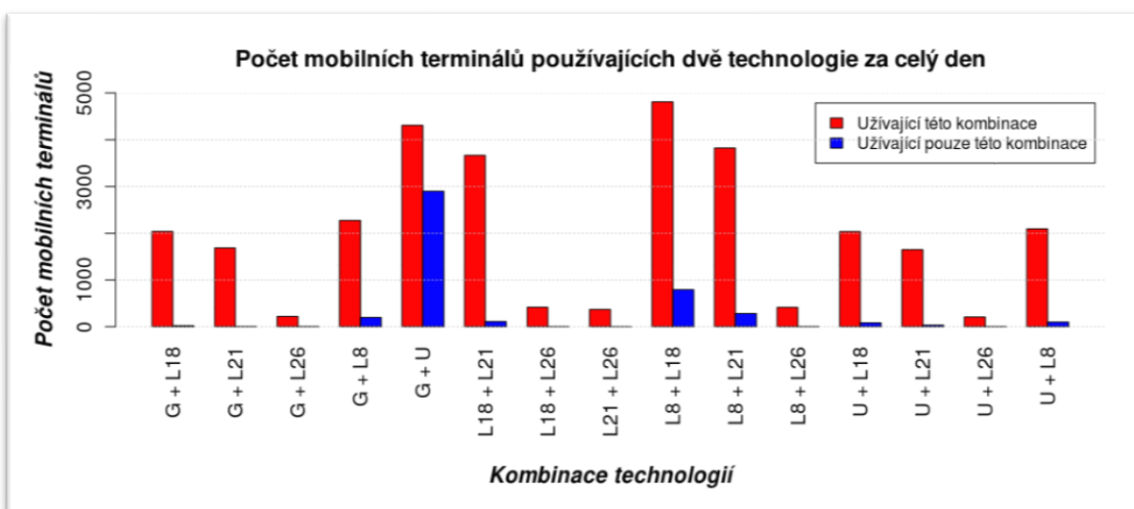
Obrázek 11. - Počet mobilních terminálů používajících dvě technologie v ranním časovém okně.



Obrázek 12. - Počet mobilních terminálů používajících dvě technologie v poledním časovém okně.



Obrázek 13. - Počet mobilních terminálů používajících dvě technologie ve večerním časovém okně.



Obrázek 14. - Počet mobilních terminálů používajících dvě technologie za celý den.

5.2.3 Použití tří technologií

Tato statistika poukazuje na pravděpodobnost kombinace tří technologií. Počet mobilních terminálů s identifikátorem IMSI, které v analyzovaných časových oknech použily kombinaci tří technologií, byl vygenerován pomocí stejných funkcí, jako výsledky předchozích analýz. Tudíž v případě, že se hledal počet mobilních terminálů, které v daném časovém okně použily kombinace jakýchkoli tří technologií, použilo se funkce *nrow()* a *filter()*, při čemž ve funkci *filter()* bylo specifikováno, že log. proměnná pro dané tři technologie musí být rovna *TRUE*. Poté se za pomoci funkce *nrow()* spočítal počet mobilních terminálů, které tyto podmínky splňovaly. V případě, že se hledal počet mobilních terminálů, které v daném časovém okně použily pouze těchto tří technologií, a žádné jiné, specifikovalo se ve funkci *filter()*, že se log. proměnná pro tyto tři technologie musí rovnat *TRUE*, a však všechny ostatní log. proměnné pro jiné technologie se musí rovnat *FALSE*. Takto byl vygenerován výsledný počet mobilních terminálů, které v každém časovém okně použily některé z možných kombinací tří technologií, a počet mobilních terminálů, které použily pouze těchto tří technologií. Výsledek analýzy pro všechna časová okna je zobrazen v tabulce č. 12. Tyto výsledky byly také pro vizuální srovnání časových oken graficky znázorněny za pomoci funkce *barplot()*, a jsou zobrazeny na obrázcích č. 15 až 18. V tabulce č. 13 je následně zobrazeno procentuální zastoupení mobilních terminálů, které použily pouze kombinace daných tří technologií, z celkového počtu terminálů, které použily těchto tří technologií. V tabulkách a grafech jsou opět jednotlivé technologie označeny zkratkami, stejně jako tomu bylo při minulé analýze, a tudíž že „G“ náleží GSM, „U“ náleží UMTS, a technologiím LTE vždy náleží písmeno „L“ a nenulové číslice v hodnotě frekvenčního pásma, tudíž například „L18“ náleží LTE 1800. Znakem „+“ jsou vždy technologie přiřazeny k sobě. A tedy například „G+U+L8“ označuje kombinaci technologií GSM, UMTS a LTE 800. Taktéž je opět v tabulkách použito označení „N“, které označuje počet mobilních terminálů, jenž použily danou kombinaci technologií, a „Np“, jež označuje počet mobilních terminálů, které již nepoužily jiné než této kombinace. Žlutou barvou jsou v tabulkách pro každou skupinu znázorněny kombinace, které dosahují nejvyšších hodnot, a oranžovou kombinace dosahující nejnižších hodnot. Pokud by některé kombinace měly shodné hodnoty, a jednalo by se o kombinace, které by byly hodné znázornění jednou z těchto barev, může se toto znázornění objevit u více kombinací v jednotlivých skupinách.

Lze opět odvodit, že mobilní terminály, které použily různé kombinace tří technologií, nejsou zástupci skupin, které v předchozích analýzách použily pouze jedné či dvou technologií. Počet těchto terminálů lze tedy pro každé časové okno vypočítat jako rozdíl mezi celkovým počtem terminálů, které používají více než jednu technologii (viz tabulka č.8), a celkovým počtem terminálů, které používají maximálně dvou technologii (viz tabulka č. 9). Počet takových to terminálů pro každé časové okno je zaznamenán v tabulce č. 11.

Tabulka 11. - Počet mobilních terminálů v časových oknech používajících více než dvě technologie.

Časová okna	Ranní okno	Polední okno	Večerní okno	Celý den
Počet IMSI	1 975	1 732	2 033	4 399

V tabulce č. 12 lze pozorovat dominanci v počtu mobilních terminálů používajících kombinaci technologií LTE 800, LTE 1800 a LTE 2100, kdy za celý den tuto kombinaci použilo 3 425 mobilních terminálů. Tato dominance je obzvláště viditelná na výsledcích pro ranní a večerní časové okno, kdy kombinace jiných technologií dosahují maximálně polovičního počtu mobilních terminálů, co dosahuje tato kombinace. Tento fakt je velmi dobře zřetelný na obrázcích č.15 a 17, kde zejména v ranním okně jinou kombinaci nepoužívá ani třetina počtu terminálů, které používají tuto dominantní kombinaci. Z tabulky č. 13 je poté možné vyčíst, že procentuální podíl počtu mobilních terminálů, které za celý den použily pouze tuto kombinaci technologií, z celkového počtu terminálů používajících tuto kombinaci, je roven 38,5 %. Je nutné však upozornit, že v ranním časovém okně tento procentuální podíl mobilních terminálů tvořil skoro 70 %. Následně, při pohledu na výsledky za celý den, vysoký počet mobilních terminálů také použil kombinaci technologií GSM, LTE 800 a LTE 1800, a také technologií UMTS, LTE 800 a LTE 1800. Tyto kombinace shodně použilo za celý den lehce pod 2 000 mobilních terminálů, a procentuální podíl mobilních terminálů, které použily pouze těchto technologií, se drží okolo 10 %. Nejnižší počet terminálů, jak je v tabulce č. 12 zřetelné, využil kombinace technologií GSM, UMTS a LTE 2600. Za celý den tuto kombinaci použilo pouze 149 terminálů, při čemž žádný terminál nepoužil pouze těchto tří technologií. Při celkovém pohledu je vidět, že počty terminálů používající i ostatní kombinace s technologií LTE 2600 opět velmi zaostávají za počty terminálů používající jiné technologie.

Při srovnání grafů na obrázcích č. 15 až 18 je zřetelné, že v poledním časovém okně je běžnější použití více různých technologií. To je pravděpodobně opět výsledkem vyšší aktivity a pohybu uživatelů mobilních terminálů v poledním časovém okně. Tato aktivita může vyplývat například z dojíždění do zaměstnání, či v případě mladších uživatelů dojíždění do školy, nebo ze samotného vykonávání pracovní činnosti.

Lze tedy říci, že v poledním časovém okně jsou uživatelé mobilních terminálů více aktivní, a z tohoto důvodu se používá více různých technologií. Zároveň lze konstatovat, že kombinace technologií LTE 800, LTE 1800 a LTE 2100 je co se týče kombinací o třech technologiích používána bezkonkurenčně největším počtem mobilních terminálů. Z druhé strany nejméně mobilních terminálů využívá kombinací s technologií LTE 2600.

Tradičně je dobré na závěr zmínit součet všech mobilních terminálů používajících přesně tři technologie za celý den, který je v tabulce č. 12 vyznačen zelenou barvou. Tento součet tvoří 1 954 mobilních terminálů, a z celkového počtu téměř 40 tisíc terminálů odpovídá přibližně 5 %.

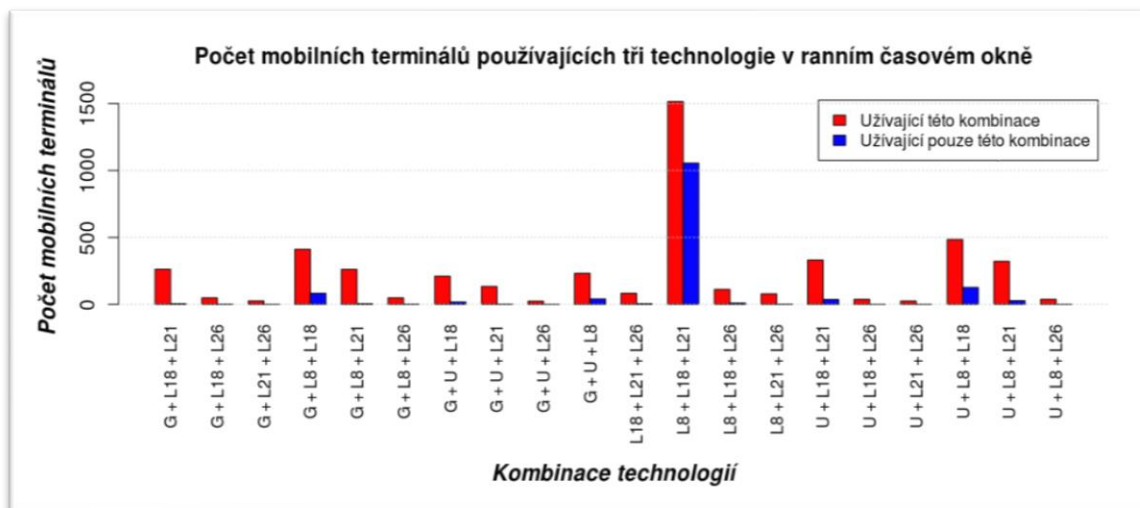
Tabulka 12. - Rozložení uživatelů mezi kombinacemi tří technologií při rozdělení do časových oken.

Kombinace	Ranní okno		Polední okno		Večerní okno		Celý den	
	N	Np	N	Np	N	Np	N	Np
G + U + L8	232	40	714	35	440	34	1310	91
G + U + L18	210	17	681	11	420	16	1260	31
G + U + L21	133	1	612	6	363	1	1083	8
G + U + L26	23	0	78	0	33	0	149	0
G + L8 + L18	411	82	968	60	796	82	1 908	167
G + L8 + L21	261	5	852	10	699	29	1 599	26
G + L8 + L26	48	1	101	0	53	0	213	1
G + L18 + L21	262	5	836	4	695	27	1 600	18
G + L18 + L26	48	1	101	0	55	1	214	1
G + L21 + L26	25	0	93	0	48	0	181	0
U + L8 + L18	484	126	874	54	693	69	1 814	190
U + L8 + L21	321	27	763	10	614	33	1 502	38
U + L8 + L26	37	0	104	2	51	0	198	2
U + L18 + L21	331	36	754	12	589	9	1 520	47
U + L18 + L26	37	0	101	0	53	0	198	1
U + L21 + L26	24	0	92	0	49	1	174	1
L8 + L18 + L21	1 515	1 055	1 374	312	1 582	631	3 425	1 318
L8 + L18 + L26	111	9	209	1	141	5	404	8
L8 + L21 + L26	78	1	198	1	129	0	359	1
L18 + L21 + L26	82	5	197	2	135	4	363	5
Součet	4673	1411	9702	520	7638	942	19474	1954

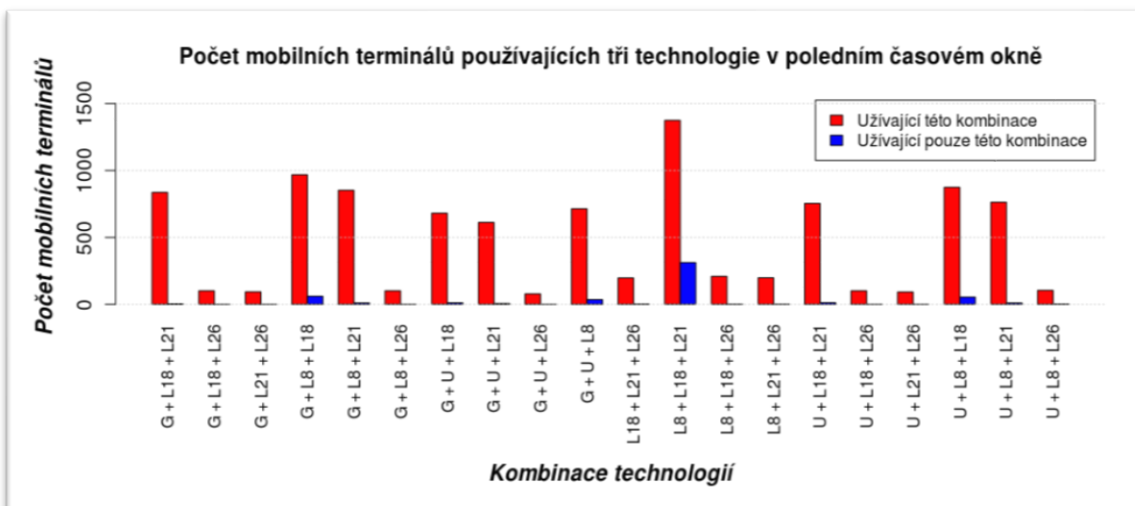
- N – počet mobilních terminálů používajících danou kombinaci technologií
- Np – počet mobilních terminálů používajících pouze danou kombinaci technologií

Tabulka 13. - Procentuální zastoupení mobilních terminálů používajících pouze daných kombinací tří technologií.

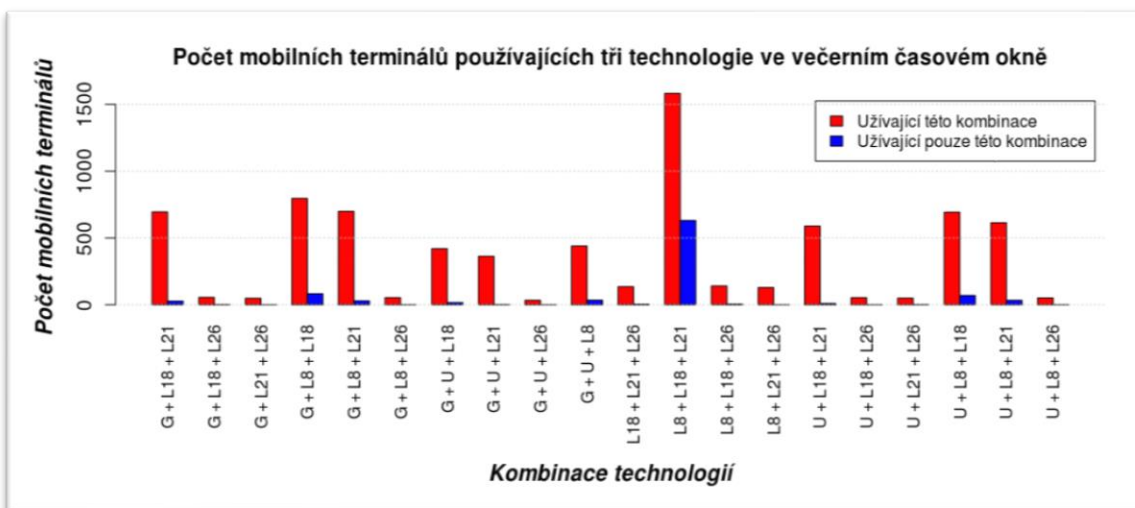
Kombinace	Ranní okno	Polední okno	Večerní okno	Celý den
G + U + L8	17,2 %	4,9 %	7,7 %	6,9 %
G + U + L18	8,1 %	1,6 %	3,8 %	2,5 %
G + U + L21	0,8 %	1,0 %	0,3 %	0,7 %
G + U + L26	0,0 %	0,0 %	0,0 %	0,0 %
G + L8 + L18	20,0 %	6,2 %	10,3 %	8,8 %
G + L8 + L21	1,9 %	1,2 %	4,1 %	1,6 %
G + L8 + L26	2,1 %	0,0 %	0,0 %	0,5 %
G + L18 + L21	1,9 %	0,5 %	3,9 %	1,1 %
G + L18 + L26	2,1 %	0,0 %	1,8 %	0,5 %
G + L21 + L26	0,0 %	0,0 %	0,0 %	0,0 %
U + L8 + L18	26,0 %	6,2 %	10,0 %	10,5 %
U + L8 + L21	8,4 %	1,3 %	5,4 %	2,5 %
U + L8 + L26	0,0 %	1,9 %	0,0 %	1,0 %
U + L18 + L21	10,9 %	1,6 %	1,5 %	3,1 %
U + L18 + L26	0,0 %	0,0 %	0,0 %	0,5 %
U + L21 + L26	0,0 %	0,0 %	2,0 %	0,6 %
L8 + L18 + L21	69,6 %	22,7 %	39,9 %	38,5 %
L8 + L18 + L26	8,1 %	0,5 %	3,5 %	2,0 %
L8 + L21 + L26	1,3 %	0,5 %	0,0 %	0,3 %
L18 + L21 + L26	6,1 %	1,0 %	3,0 %	1,4 %



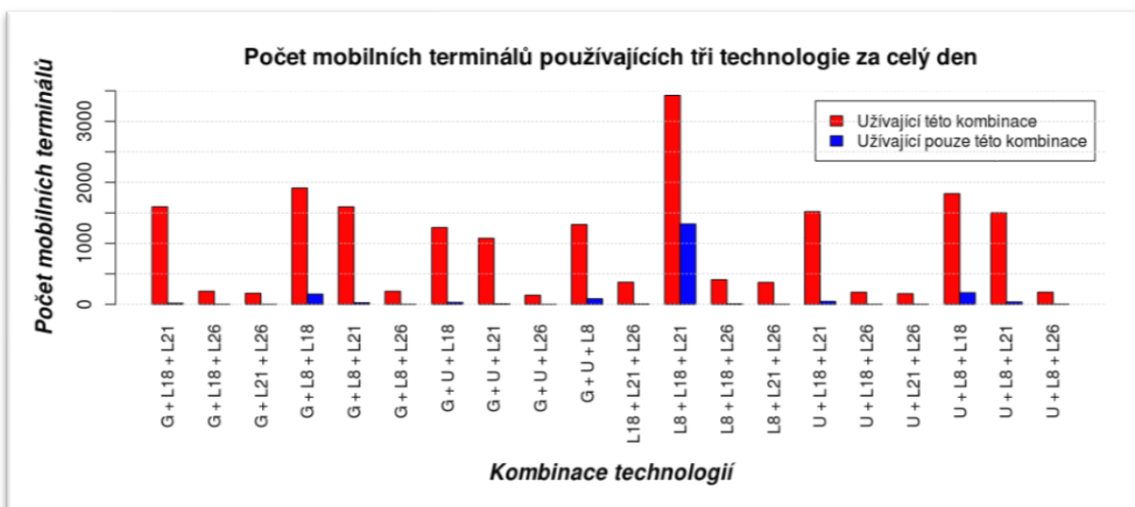
Obrázek 15. - Počet mobilních terminálů používajících tři technologie v ranním časovém okně.



Obrázek 16. - Počet mobilních terminálů používajících tři technologie v poledním časovém okně.



Obrázek 17. - Počet mobilních terminálů používajících tři technologie ve večerním časovém okně.



Obrázek 18. – Počet mobilních terminálů používajících tři technologie za celý den.

5.3 Identifikace uživatelských skupin

Mobilní terminály je možné dělit do skupin podle podobných vzorů chování. Vzory chování lze nalézt v mnoha charakteristikách. V této kapitole se ukáže rozdělení mobilních terminálů do skupin podle celkového počtu záznamů za den, podle průměrné doby připojení na buňkách, a podle celkového počtu signalizačních zpráv vyměněných si se základnovými stanicemi. Skupiny mohou podle chování odpovídat běžným uživatelům mobilních terminálů, systémům nastaveným pro určité účely, či i porouchaným mobilním terminálům. Skupiny jsou vytvořeny na základě předpokladů vázajících se k daným typům chování. V této kapitole je použita původní datová sada pro celý den obsahující 94 milionů záznamů ze všech sítí a obsahující více jak 5 milionů mobilních terminálů. Přesné počty jsou zaznamenány v tabulce č. 14.

Tabulka 14. - Charakteristiky původní datové sady.

Charakteristiky	Počet záznamů	Počet identifikátorů IMSI
Původní datová sada	94 399 454	5 006 847

Z důvodu, že se tato datová sada pro celý den skládá ze tří časových oken, byly tímto výsledky analýzy jemně ovlivněny. Zejména z tohoto důvodu každý mobilní terminál, co se vyskytl ve všech časových oknech, zaznamenal minimálně 3 záznamy. To má následně vliv na velikost průměrného času připojení na buňkách, kdy mobilní terminál, obsažený ve všech třech časových oknech, může dosáhnou maximálního průměrného času připojení na buňkách rovného osmi hodinám.

5.3.1 Rozdělení podle počtu záznamů

První dělení mobilních terminálů do skupiny je uskutečněno na základě počtu záznamů, které jsou v datové sadě obsaženy s informacemi o připojení daných terminálů na libovolnou buňku. Toto je velmi důležité rozdělení, jelikož reflektuje, jak moc se mobilní terminály pohybují. Každý záznam totiž obecně představuje novou buňku, na kterou se terminál připojil. Tudíž čím více záznamů bylo pro daný terminál vytvořeno, tím více se daný terminál pohyboval. Lze takto určit, jak aktivní daní uživatelé mobilních terminálů jsou, a či zda se vůbec jedná o uživatele z řad populace, a ne o účelové systémy, kterých součástí mobilní terminál může být. Toto rozdělení bylo uskutečněno na základě určitých předpokladů získaných ze zkoumání signalizačních dat, a taktéž na základě rad z praxe, které byly získány od osob znalých v tématu.

Rozdělení mobilních terminálů do skupin podle počtu záznamů bylo uskutečněno za pomoci funkcí *count()* a *filter()*. Funkce *count()* v tomto případě spočítala, kolikrát se v datové sadě vyskytovaly jednotlivé identifikátory IMSI, neboli kolik záznamů který mobilní terminál za celý den zaznamenal. Výsledkem je poté datová tabulku obsahující první sloupec se všemi unikátními identifikátory IMSI vyskytujícími se v datové sadě, a druhý sloupec udávající kolik záznamů mobilní terminál s daným identifikátorem za celý den zaznamenal. Následně bylo uskutečněno rozdělení do skupin na základě určených předpokladů, a to za pomoci funkce *filter()*, při čemž se určila spodní

a horní hranice počtu záznamů pro dané skupiny. Použití funkcí pro vytvoření skupiny mobilních terminálů, které za den zaznamenaly až 10 záznamů, je poznamenáno v Příloze A, bod (8). Nakonec byl za pomoci další funkce *count()* spočten počet mobilních terminálů s totožným počtem záznamů, a tato statistika počtu mobilních terminálů ku počtu záznamů byla poté vygenerována do grafu za pomoci funkce *plot()*, a je znázorněna na obrázcích č. 19 a 20.

Mobilní terminály byly rozděleny do těchto skupin podle počtu záznamů (počet mobilních terminálů v jednotlivých skupinách je zaznamenán v tabulce č. 15):

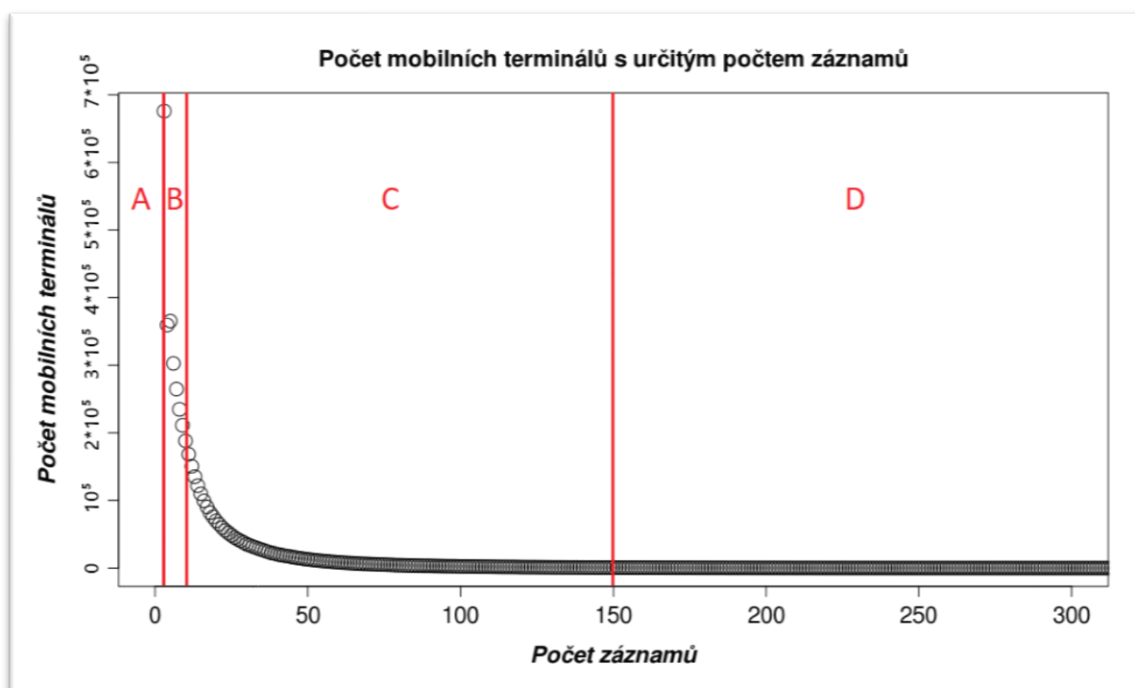
- **A** - Skupina obsahující mobilní terminály, které za celý den zaznamenaly pouze 3 záznamy. Jak již bylo zmíněno, jelikož datová sada je rozdělena do tří časových oken, které obsahují záznamy zvlášť, každý terminál, vyskytující se ve všech těchto časových oknech, dosáhl minimálně tři záznamů. Může se tedy jednat o mobilní terminály, které byly ve všech třech časových oknech připojeny na jedné buňce, ve výsledku však zaznamenaly tři záznamy. Tato skupina bude tedy nejspíše obsahovat mobilní terminály, které se nepohybují, a fungují z jednoho místa. U této skupiny je význačné mnohem větší využití technologie GSM, než je tomu u ostatních skupiny. Zároveň mobilní terminály v této skupině disponují velmi dlouhými časy připojení na buňkách. Tato skupina bude tedy pravděpodobně obsahovat mnoho terminálů, které jsou součástí různých účelových systémů, které mohou fungovat nepřetržitě. Může se jednat například o telefonní linky a podobná zařízení. V této skupině je obsaženo více jak 675 tisíc mobilních terminálů.
- **B** - Skupina obsahující mobilní terminály, které za celý den zaznamenaly 4 až 10 záznamů. Tato skupina zastupuje mobilní terminály, které zaznamenaly velmi nízký počet záznamů, což znamená, že se za celý den pravděpodobně připojily pouze na pár buněk. Bude se tedy jednat o ne příliš přemísťující se terminály. Uživatelé těchto terminálů mohou být běžní uživatelé z řad populace, co převážně zůstávají na jednom místě. Taktéž se může jednat o různé účelové systémy, jakž tomu je ve skupině A. V této skupině nadále převládá použití technologie GSM. Tato skupina zastupuje více jak 1,9 milionu mobilních terminálů.
- **C** - Skupina obsahující mobilní terminály, které za celý den zaznamenaly 11 až 150 záznamů. Tato skupina by ve větší míře měla odpovídat mobilním terminálům používaným běžnými uživateli z řad populace. Zaznamenání 11 až 100 záznamů za den, neboli připojení se k takovému to počtu buněk, odpovídá běžnému pohybu obyvatel. Mobilní terminály, které zaznamenaly od 100 do 150 záznamů, již odpovídají uživatelům, kteří se více pohybují, či více střídají technologie. Například takto může uživatel z řad populace střídát připojení mezi buňkami technologií LTE a GSM, a to v závislosti na tom, zda právě využívá hovorových služeb či datových služeb. U mobilních terminálů v této skupině převažuje připojení k celulární rádiové síti za pomoci technologií LTE. V této skupině je obsaženo skoro 2,4 milionů mobilních terminálů.

- **D** - Skupina obsahující mobilní terminály, které za celý den zaznamenaly více jak 150 záznamů. Tato skupina mobilních terminálů se vyznačuje velkým počtem buněk, na které se jednotlivé terminály připojily. Bude se tedy jednat o vysoce pohyblivé terminály. Může se například jednat o terminály používané uživateli z řad populace, kteří více cestují, a překonávají například autem či jiným prostředkem delší vzdálenosti. Může se taktéž jednat o uživatele, jejichž práce obnáší časté přesuny, jak tomu může být například u řidičů autobusů či strojvedoucích vlaků. V této skupině dominuje použití technologií LTE. Tato skupina obsahuje skoro 33 tisíc mobilních terminálů.

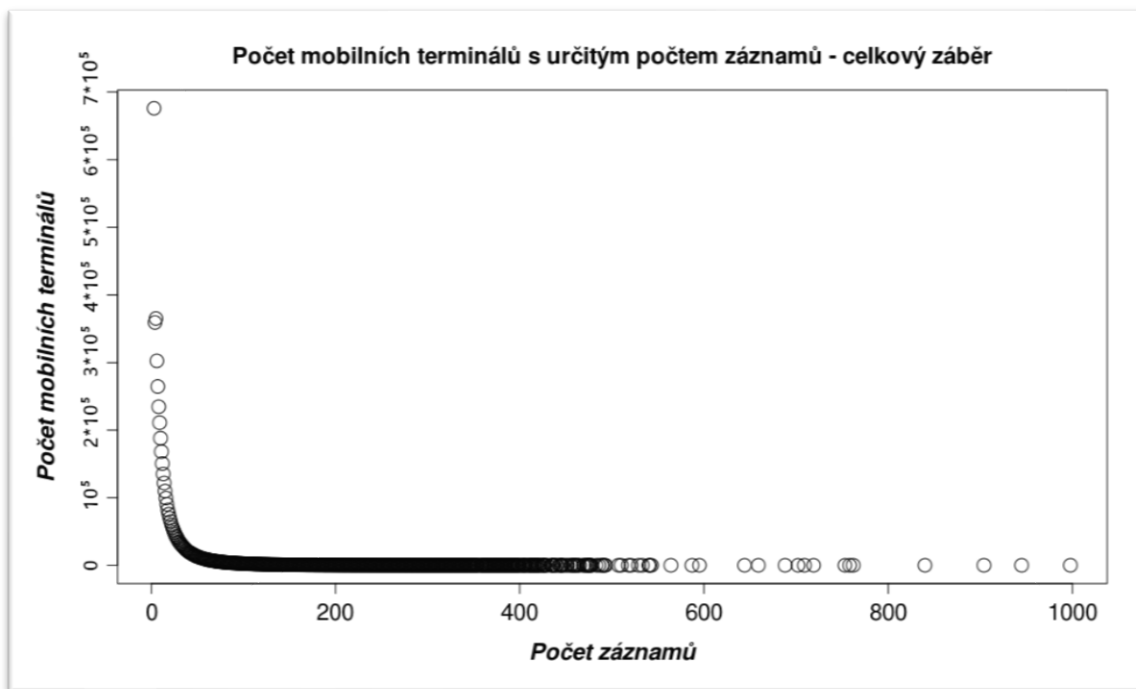
Tabulka 15. - Počet mobilních terminálů v jednotlivých skupinách podle počtu záznamů.

Skupiny	A	B	C	D	Celkem
Počet IMSI	675 827	1 925 594	2 372 438	32 988	5 006 847

Graf na obrázku č. 19 zobrazuje rozložení počtu mobilních terminálů ku počtu záznamů, neboli ku počtu buněk, které terminál navštívil. Na grafu je taktéž zobrazeno rozdělení do jednotlivých skupin. V grafu na obrázku č. 20 je zobrazena tato statistika pro celé spektrum počtu záznamů. Na grafu 19. je viditelný markantní rozdíl mezi mobilními terminály, které zaznamenaly 3 záznamy za celý den (terminály ze skupiny A), a mobilními terminály, které zaznamenaly 4 záznamy za celý den. Tento rozdíl tvoří více jak 300 tisíc terminálů. Rozdíl je takto markantní z důvodu opravdu velkého množství terminálů ve skupině A odpovídajících pouze jedné hodnotě počtu záznamů. Počet mobilních terminálů se čtyřmi či pěti záznamy je poté velmi podobný. Nadále se počty mobilních terminálů pozvolnou křivkou snižují s rostoucím počtem záznamů.



Obrázek 19. - Počet mobilních terminálů v závislosti na počtu záznamů.



Obrázek 20. - Počet mobilních terminálů v závislosti na počtu záznamů – celkový záběr

5.3.2 Rozdělení podle průměrného času připojení na buňce

Rozdělení mobilních terminálů do skupin podle průměrného času připojení na buňce je velmi účinný nástroj k rozpoznání, zda je mobilní terminál užíván běžnými uživateli z řad populace, či zda se jedná o součásti účelových systémů.

Pro tuto analýzu bylo opět velmi důležité, stejně jak u analýz v kapitole 5.2, rozdělit datovou sadu na jednotlivé datové tabulky pro jednotlivé mobilní terminály. Toho se docílilo za pomoci funkce *split()*, při čemž jako dělicí faktor byl použit identifikátor IMSI zastupující mobilní terminály. Vytvořil se tak list obsahující tyto tabulky. V následujícím kroku bylo potřebné z daných tabulek vygenerovat informaci o průměrném čase stráveném připojením k buňkám. K tomuto kroku byla použita funkce *sapply()* s vnitřní funkcí *mean()*. Funkce *sapply()* funguje při použití v této analýze stejně jako funkce *lapply()*, byl však u ní určen výsledný formát výstupu jako vektor. Vnitřní funkcí *mean()* byl generován pro každý mobilní terminál průměrný čas připojení na buňkách. Bylo ale potřeba průměrný čas připojení vždy přiřadit k danému identifikátoru IMSI. Bylo proto navíc ještě použito funkce *tibble()* a další funkce *sapply()*. Druhá funkce *sapply()* současně s prvním generovala k průměrnému času také identifikátor IMSI daného mobilního terminálu, a následně se pomocí funkce *tibble()* tvořila tabulka obsahující tyto dvě položky. Kompletní použití funkcí je zaznamenáno v Příloze A, bod (9). Nakonec byl za pomoci další funkce *count()* spočten počet mobilních terminálů s totožným průměrným časem připojení, a tato statistika počtu mobilních terminálů ku průměrnému času připojení byla poté vygenerována do grafu za pomoci funkce *plot()*, a je znázorněna na obrázku č. 21.

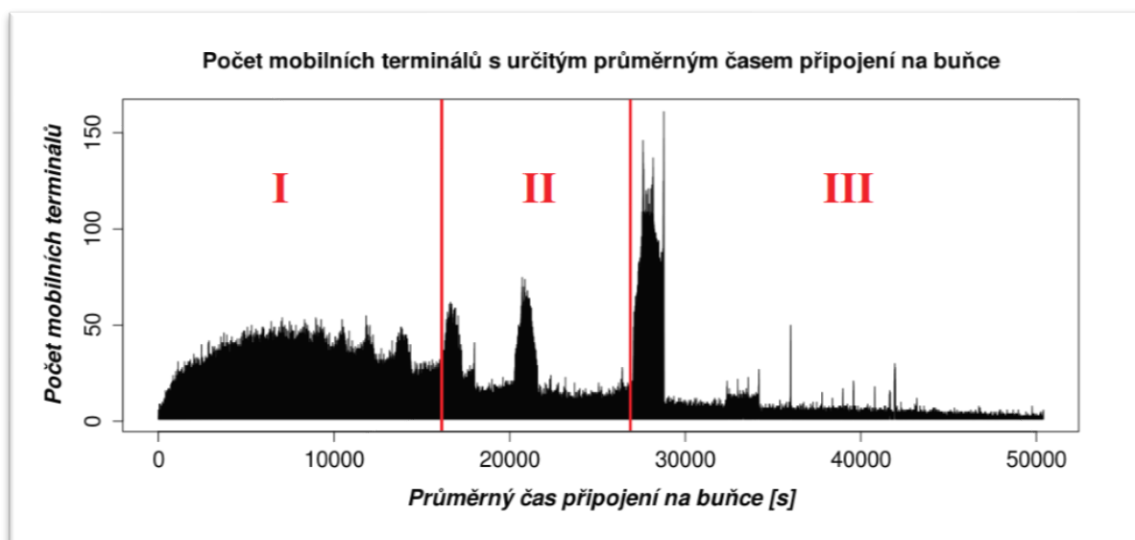
Mobilní terminály byly nakonec rozděleny do tří skupin za pomoci funkce *filter()* určující horní a spodní hranici velikosti průměrného času připojení mobilních terminálů na buňkách. Rozdělení bylo provedeno na základě řady předpokladů, které byly získány při zkoumání datové sady. Počet mobilních terminálů v jednotlivých skupinách je poté zaznamenán v tabulce č. 16. Skupiny jsou následující:

- **I** – Skupina obsahující mobilní terminály, které za celý den zaznamenaly průměrný čas připojení na buňkách v rozmezí od 0 s až po 17 000 s. Mobilní terminály v této skupině jsou používány ve velké míře běžnými uživateli z řad populace. Tyto terminály se za den mohou připojit k mnoha buňkám, v některých strávit delší čas, v některých třeba jen vteřiny při přesunu. Terminály s průměrným časem menším než 1000 s odpovídají zejména uživatelům, kteří se často přesouvají a zaznamenají mnoho záznamů za den. Na druhou stranu terminály s vyšším průměrným časem v této skupině odpovídají více uživatelům, kteří se připojili pouze na pár buněk, ve kterých se zdrželi delší dobu. V této skupině převládá použití technologií LTE, a také GSM. Tato skupina obsahuje velkou část z celkového počtu mobilních terminálů, a to více jak 4,1 milionů terminálů.
- **II** – Skupina obsahující mobilní terminály, které za celý den zaznamenaly průměrný čas připojení na buňce v rozmezí od 17 001 s až po 28 000 s. Mobilní terminály v této skupině disponují velmi vysokým průměrným časem připojení na buňkách. To odpovídá terminálům, které za den zaznamenaly pouze malý počet záznamů s vysokými časy připojení. Může se už jednat jak o účelové systémy, které není nutné za provozu přemísťovat, tak o běžné uživatele, kteří mobilní terminál používají pro časově delší činnosti, aniž by se při tom přesouvali. Co se týče použití technologií, převládá technologie GSM. Tato skupina zastupuje 618 tisíc mobilních terminálů.
- **III** - Skupina obsahující mobilní terminály, které za celý den zaznamenaly průměrný čas připojení na buňkách roven a větší 28 001 s. Jedná se o mobilní terminály, které mají průměrný čas o velikosti okolo 8 hodin a vyšší. Při úvaze, že mobilní terminály, které byly obsaženy ve všech časových oknech, zaznamenaly minimálně 3 záznamy, při čemž pouze v poledním časovém okně je možné dosáhnout delšího času než 5 hodin, bylo dospěno k závěru, že se jedná o mobilní terminály, které jsou pravděpodobně součástí nepohyblivých systémů, a které jsou připojeny k buňkám a fungují nepřetržitě. Může se taktéž jednat o mobilní terminály obsažené pouze v poledním časovém okně, které zaznamenaly průměrný čas delší než 8 hodin. To však nejspíš tvoří pouze minoritu této skupiny. V této skupině taktéž převládá použití technologie GSM. Celkem je v této skupině obsaženo přes 266 tisíc mobilních terminálů z datové sady.

Tabulka 16. - Počet mobilních terminálů v jednotlivých skupinách podle průměrného času připojení

Skupiny	I	II	III	Celkem
Počet IMSI	4 122 138	618 055	266 654	5 006 847

Graf na obrázku č. 21 zobrazuje početní rozložení mobilních terminálů v závislosti na jejich průměrném čase připojení k buňkám. V grafu je vidět několik skokových nárůstů na určitých hodnotách průměrného času. Největší nárůst je pozorovatelný na hodnotách okolo 28 000 s až 29 000 s. To odpovídá mobilním terminálům s průměrným časem okolo osmi hodin. To, jak již bylo řečeno, odpovídá zejména mobilním terminálům, které jsou obsaženy ve všech třech časových oknech, a které jsou připojeny a fungují nepřetržitě. Mobilní terminály, které zaznamenaly průměrný čas připojení delší než 8 hodin, by měly odpovídat terminálům, které nebyly obsaženy ve všech časových oknech. V případě, že byly obsaženy pouze v poledním časovém okně, a byly připojeny nepřetržitě celé časové okno, mohly dosáhnout až průměrného času 50 000 s. Další dva velké nárůsty v počtu mobilních terminálů jsou pro průměrné časy připojení pohybující se okolo 18 000 s a 21 000 s. To odpovídá přibližně 5 a 6 hodinám. Jedná se nejspíše opět o mobilní terminály připojené takřka nepřetržitě, které v tomto případě už z nějakých důvodů za den zaznamenaly alespoň 4 nebo 5 záznamů. Největší celkový počet mobilních terminálů je obsažen v rozmezí od 1000 s do 17 000 s, kdy se na mnoha hodnotách pohybuje až 50 mobilních terminálů. Je očividné, že toto rozmezí reflektuje mnohem větší škálu různých výsledků, které odpovídá různorodosti uživatelů z řad populace.



Obrázek 21. - Počet mobilních terminálů v závislosti na průměrném čase připojení na buňkách.

5.3.3 Rozdělení podle počtu signalizačních zpráv vyměněných si s BTS

Třetím dělením je dělení mobilních terminálů do skupin podle počtu signalizačních zpráv, které si terminály za celý den vyměnily se všemi BTS, ke kterým byly připojeny. Toto rozdělení do skupin indikuje, jací uživatelé používají dane mobilní terminály, či zda se vůbec jedná o správně funkční terminály.

Rozdělení mobilních terminálů do skupin je uskutečněno skoro totožně, jako v předchozí analýze. Namísto počítání s časem stráveným na buňce, se však při této analýze počítalo s počty signalizačních zpráv vyměněných se všemi BTS, ke kterým byl terminál připojen, a namísto průměrování se sčítalo. Tudíž ve vnitřní funkci funkce *supply()* se na místo funkce *mean()* použilo funkce *sum()*, a jako vektor, který se sčítal, se určil sloupec zastupující počty zpráv vyměněných s BTS. Takto se vygeneroval seznam všech identifikátorů IMSI, při čemž ke každému identifikátoru byl přiřazen celkový počet zpráv, které si za celý den mobilní terminál vyměnil se všemi BTS.

Následně byl za pomoci funkce *filter()* seznam na základě rad osob znalých v tématu rozdělen do skupin podle počtu těchto signalizačních zpráv. Počty terminálů spadající do jednotlivých skupin jsou zaznamenány v tabulce č. 17. Rozdělení do skupin je poté také vyobrazeno na grafu na obrázku č. 22. Na obrázku č. 23 je poté zobrazen graf celého rozložení mobilních terminálů vůči počtu vyměněných signalizačních zpráv. Tyto grafy byly stejně jako v minulých analýzách vygenerovány za pomoci funkce *count()* a *plot()*. Skupiny, do kterých byly mobilní terminály rozděleny, jsou následující:

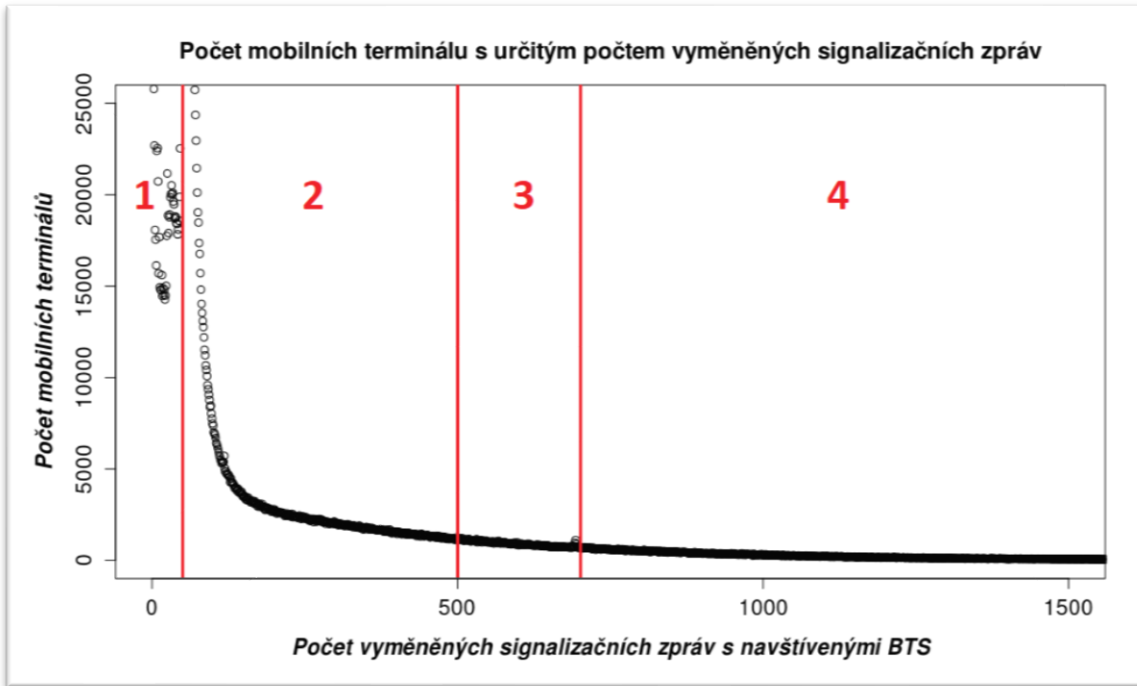
- **1** - Skupina obsahující mobilní terminály, které si za celý den vyměnily se všemi BTS, k jejichž buňkám byly připojeny, pouze 1 až 47 signalizačních zpráv. Tato skupina bude nejspíše obsahovat mobilní terminály, které buďto fungují z jednoho místa, nebo se ne příliš často pohybují, a z tohoto důvodu navštívily velmi malý počet buněk. V této skupině převládá použití technologie GSM. Terminály v této skupině budou pravděpodobně buďto součástí různých účelových systému, zejména na bázi využívání hovorových služeb, jako jsou například telefonní linky, nebo budou užívány uživateli, kteří jsou při využívání telekomunikací méně aktivní. Tato skupina obsahuje přes 976 tisíc mobilních terminálů.
- **2** – Skupina obsahující mobilní terminály, které si za celý den vyměnily se všemi BTS, k jejichž buňkám byly připojeny, 48 až 500 signalizačních zpráv. Tato skupina by ve větší míře měla odpovídat mobilním terminálům používaným běžnými uživateli z řad populace. U mobilních terminálů v této skupině převažuje připojení k celulární rádiové síti za pomoci technologií LTE. To odpovídá zejména vyšším počtům signalizačních zpráv vyměněných s BTS, jelikož technologie LTE vyžaduje výměnu vyššího počtu signalizačních zpráv než ostatní technologie zmíněné v této práci. Do této skupiny spadá velká část mobilních terminálů, a to více jak 3,6 milionu.

- **3** - Skupina obsahující mobilní terminály, které si za celý den vyměnily se všemi BTS, k jejichž buňkám byly připojeny, 501 až 700 signalizačních zpráv. Do této skupiny taktéž spadají mobilní terminály používané uživateli z řad populace. Tentokrát se již jedná o velmi aktivní uživatele trávící hodně času připojením a využívající zejména datových služeb. Jedná se tedy zejména o mladší uživatele. V této skupině je více jak 60 % záznamů vytvořeno na sítích LTE. Tato skupina obsahuje něco málo přes 180 tisíc mobilních terminálů.
- **4** - Skupina obsahující mobilní terminály, které si za celý den vyměnily se všemi BTS, k jejichž buňkám byly připojeny, více jak 700 signalizačních zpráv. Do této skupiny již spadají mobilní terminály, u kterých množství signalizačních zpráv vyměněných se základnovými stanicemi se pohybuje na velmi vysokých počtech, že se až jeví jako podezřelé. Zároveň více jak 90 % záznamů vytvořených terminály v této skupině je vytvořeno za připojení na síť LTE. Jedná se často o mobilní terminály, které jsou součástí porouchaného systému, či samy mohou být porouchané. Taktéž se může jednat o terminály, které jsou chybně nastaveny, a s BTS si nerozumí. Takové to terminály poté mohou zaznamenat velmi vysoký počet signalizačních zpráv pouze tvořených například z požadavků o připojení k buňce, které k BTS neustále posílají. V této skupině je obsaženo více jak 243 tisíc terminálů.

Tabulka 17. - Počet mobilních terminálů v jednotlivých skupinách podle počtu signalizačních zpráv.

Skupiny	1	2	3	4	Celkem
Počet IMSI	976 606	3 605 842	180 617	243 782	5 006 847

Graf na obrázku č. 22 zobrazuje rozložení počtu mobilních terminálů ku počtu vyměněných signalizačních zpráv se základnovými stanicemi. Na grafu je zároveň zobrazeno rozdělení terminálů do jednotlivých skupin. V grafu na obrázku č. 23 je poté zobrazena tato statistika pro celé spektrum jak počtu mobilních terminálů, tak počtu vyměněných signalizačních zpráv. Na grafu na obrázku č. 22 je pozorovatelná klesající křivka počtu mobilních terminálů ku počtu vyměněných signalizačních zpráv s BTS. Tedy s rostoucím počtem vyměněných zpráv klesá počet terminálů. Z této křivky pouze vystupují terminály, které dosahují celkového počtu signalizačních zpráv vyměněných s BTS menší než 48. Tyto terminály spadají do skupiny **1**, a jedná se zejména o terminály, které navštívily velmi malý počet buněk. Z důvodu, že každá technologie vyžaduje výměnu různého počtu signalizačních zpráv, je rozložení mobilních terminálů, které navštívily nízký počet buněk, v grafy nestálé.



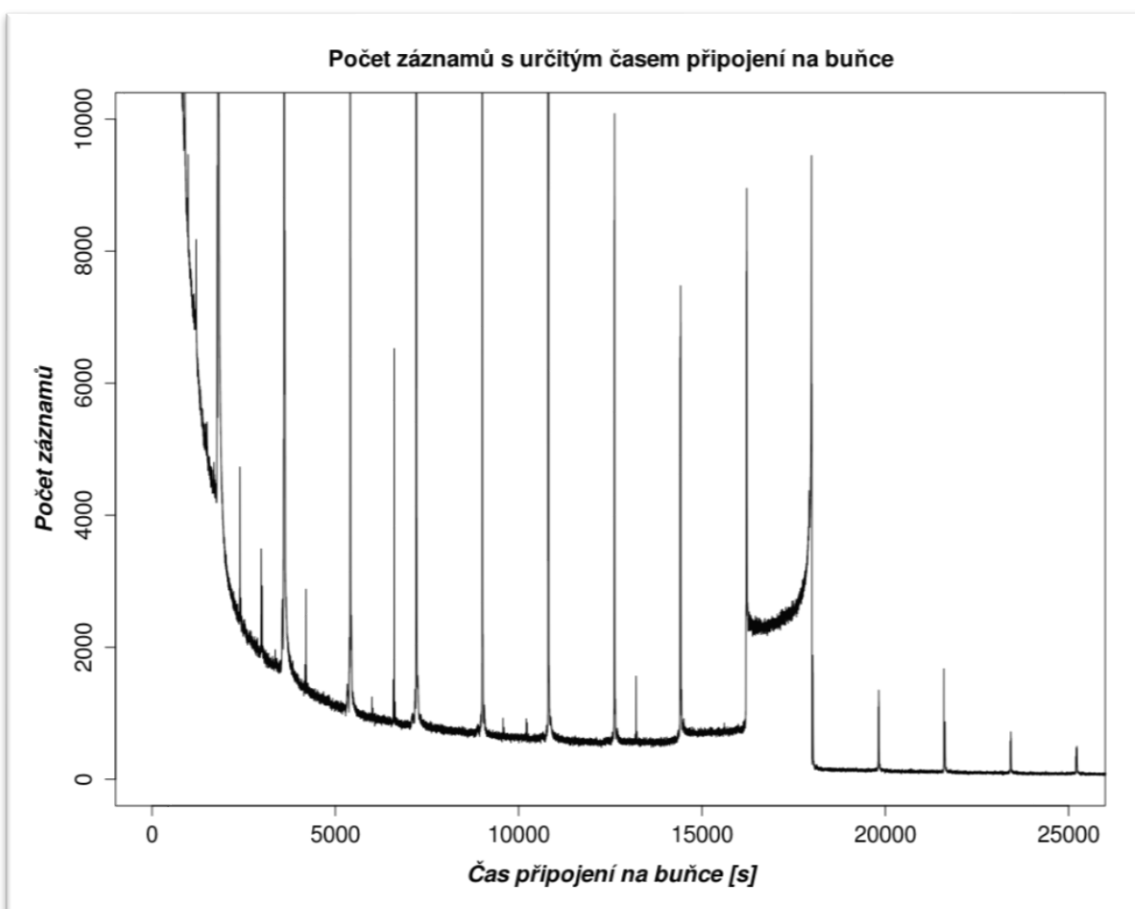
Obrázek 22. - Počet mobilních terminálů v závislosti na počtu signalačních zpráv vyměněných s BTS.



Obrázek 23. - Počet mobilních terminálů v závislosti na počtu signalačních zpráv vyměněných s BTS - celkový záběr.

5.3.4 Rozložení záznamů o různé délce připojení

V této kapitole se zobrazí rozložení záznamů v závislosti na délce připojení mobilních terminálů na buňkách. V tomto případě se nebude dělit vzorek na skupiny mobilních terminálů, zejména jelikož nejde o charakteristiky mobilních terminálů, ale záznamů, které byly jimi vytvořeny. Rozložení bylo uskutečněno za pomoci funkce `count()`. Tato funkce byla použita na datovou sadu a spočetla počet záznamů, které zaznamenaly stejný čas připojení. Následně byla tato statistika za pomoci funkce `plot()` vygenerována do grafu, který je možné vidět na obrázku č. 24. Na grafu je zřetelný pokles počtu záznamů s délkou připojení. To je předvídatelný průběh, jelikož se vyprodukuje více záznamů při pohybu mezi buňkami, či při kratších zastávkách na buňkách, než při delším připojení na buňkách. Nadále si lze všimnout nárůstu záznamů o délce okolo 17000 s, což odpovídá přibližně pěti hodinám. To svědčí o velkém počtu záznamů, které dosáhly takové to délky připojení, a můžou například odpovídat záznamům o připojení uživatelů, co stráví okolo pět hodin třeba v práci či ve škole. Jedná se však také o záznamy z mobilních terminálů, které byly v ranním či večerním okně připojeny nepřetržitě na jedné buňce celé časové okno, a tudíž jejich záznam je dlouhý přibližně 5 hodin. Na grafu je taktéž pozorovatelný značný nárůst každých 1800 s. Tento nárůst je pravděpodobně výsledkem určitého nastavení sítě.



Obrázek 24. - Počet záznamů v závislosti na délce času připojení na buňce.

6 Závěr

V této bakalářské práci byl na úvod objasněn jev, kterému mnoho lidí říká „*Big Data*“. Následně byly probrány základy fungování celulární rádiové sítě, zejména struktura buňkového systému, sektorizace, vývoj a fungování jednotlivých technologií či funkce přístupových metod. Po shrnutí, jakým způsobem probíhala práce na signalizačních datech, a jakou strukturu daná poskytnutá data měla, se přešlo k praktické části bakalářské práce, kterou představovala řada analýz zkoumající chování mobilních terminálů a strukturu celulární rádiové sítě.

První sada analýz představovala samotné zkoumání obsahu datové sady, a výsledkem bylo vygenerování údajů o počtu buněk, BTS, sektorů a samotných mobilních terminálů v datové sadě.

Druhá sada analýz byla zaměřena na využívání jednotlivých technologií. Přinesla poznatky, které technologie jsou nejvíce využívány mobilními terminály, a které technologie se nejčastěji používají samostatně neboli, aniž by se mobilní terminál připojil k jiné technologii. Následně zobrazila, jaké technologie se často kombinují, a též naopak, ke kterým technologiím se jeden mobilní terminál ne příliš často připojí v jediném dni. Výsledky analýz pro kombinace technologií byly vygenerovány pro kombinace dvou a tří technologií. Veškeré výsledky byly taktéž znázorněny na grafech.

Závěrečná sada analýz byla zaměřena na rozdělení mobilních terminálů do skupin podle podobného vzoru chování, při čemž jako určující charakter chování byl celkový počet záznamů za den, průměrný čas připojení na buňkách za den a celkový počet signalizačních zpráv vyměněných se základnovými stanicemi za den. Mobilní terminály byly vždy rozděleny do skupin podle stanovených kritérií, a pro každou skupinu bylo udáno, jaké uživatele mobilních terminálů tato skupina může představovat. Následně bylo rozložení počtu terminálů ku danému charakteru chování znázorněno na grafu. Nakonec bylo znázorněno rozložení veškerých záznamů vyprodukovaných v celulární rádiové síti v závislosti na délce připojení.

Prvotním záměrem této práce bylo stanovit vzory chování mobilních terminálů v průběhu dne, týdne a měsíce. Z důvodu velké rozsáhlosti tématu bylo po konzultaci s vedoucím práce téma zpracováno pouze pro dobu jednoho dne.

Seznam použitých informačních zdrojů

- [1] MAYER-SCHÖNBERGER, Viktor a Kenneth CUKIER. *Big Data*. Brno: Computer Press, 2014. ISBN 978-80-251-4119-9.
- [2] MARR, Bernard. *Big data: using smart big data, analytics and metrics to make better decisions and improve performance*. [Hoboken], New Jersey: Wiley, 2015. ISBN 9781118965832.
- [3] LOSHIN, David. *Big data analytics: from strategic planning to enterprise integration with tools, techniques, NoSQL, and graph*. Amsterdam: Elsevier, Morgan Kaufmann, 2013. ISBN 0124173195.
- [4] BEČVÁŘ, Zdeněk, Pavel MACH a Ivan PRAVDA. *Mobilní sítě*. V Praze: České vysoké učení technické, [2013]. ISBN 978-80-01-05305-8.
- [5] Příspěvatelé Wikipedie, *Celulární radiová síť* [online], Wikipedie: Otevřená encyklopedie, c2020, Datum poslední revize 17. 02. 2020, [cit. 2020-4-20]. Dostupné z: https://cs.wikipedia.org/wiki/Celulární_radiová_síť
- [6] HANUS, Stanislav. *Rádiové a mobilní komunikace*. Brno: FEKT VUT v Brně, 2002.
- [7] UNZEITIG, Lubomír. *Model pokrytí území buňkové sítě*. Brno, 2008. Diplomová práce. VUT v Brně. Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky.
- [8] TRÁVNÍČEK, Tomáš. *Monitoring GSM sítě*. Brno, 2008. Diplomová práce. VUT v Brně. Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací.
- [9] LAI - Telecom ABC. *Telecom ABC - Telecommunications and Internet dictionary* [online]. Copyright © 2005 Telecom ABC. All Rights Reserved. [cit. 2020-4-27]. Dostupné z: <http://www.telecomabc.com/l/lai.html>
- [10] The World's Technological Capacity to Store, Communicate, and Compute Information. *MartinHilbert.net - Research, videos, news, teaching & fun* [online]. [cit. 2020-4-22]. Dostupné z: <https://www.martinhilbert.net/worldinfocapacity-html/>
- [11] Příspěvatelé Wikipedie, *SIM karta* [online], Wikipedie: Otevřená encyklopedie, c2019, Datum poslední revize 16. 09. 2019, [cit. 2020-4-27]. Dostupné z: https://cs.wikipedia.org/wiki/SIM_karta
- [12] Příspěvatelé Wikipedie, *International Mobile Subscriber Identity* [online], Wikipedie: Otevřená encyklopedie, c2019, Datum poslední revize 11. 07. 2019, [cit. 2020-4-27]. Dostupné z: https://cs.wikipedia.org/wiki/International_Mobile_Subscriber_Identity
- [13] Příspěvatelé Wikipedie, *Mobility management* [online], Wikipedie: Otevřená encyklopedie, c2020, Datum poslední revize 06. 04. 2020, [cit. 2020-4-27]. Dostupné z: https://en.wikipedia.org/wiki/Mobility_management

- [14] Příspěvatelé Wikipedie, *Nordic Mobile Telephone* [online], Wikipedie: Otevřená encyklopedie, c2019, Datum poslední revize 13. 06. 2019, [cit. 2020-4-27]. Dostupné z: https://cs.wikipedia.org/wiki/Nordic_Mobile_Telephone
- [15] JURÁNEK, Michal. *Emulátor Registru uživatelů mobilní sítě* [online]. Brno, 2014 [cit. 2020-05-18]. Dostupné z: <http://hdl.handle.net/11012/53270>. Diplomová práce. Vysoké učení technické v Brně. Fakulta informačních technologií. Ústav informačních systémů. Vedoucí práce František Ščuglík.
- [16] Příspěvatelé Wikipedie, *Tidyverse* [online], Wikipedie: Otevřená encyklopedie, c2020, Datum poslední revize 20. 04. 2020, [cit. 2020-5-18]. Dostupné z: <https://en.wikipedia.org/wiki/Tidyverse>
- [17] Příspěvatelé Wikipedie, *Virtuální privátní síť* [online], Wikipedie: Otevřená encyklopedie, c2019, Datum poslední revize 24. 10. 2019, [cit. 2020-5-18]. Dostupné z: https://cs.wikipedia.org/wiki/Virtu%C3%A1ln%C3%AD_priv%C3%A1tn%C3%AD_s%C3%AD%C5%A5

Přílohy A

Zde jsou zapsány jednotlivé části kódu, které byly použity při analýzách k vygenerování různých typů výsledků, a demonstrují, jakým způsobem byly analýzy provedeny při využití programovacího jazyka R, a funkcí zmíněných v kapitole 4. Tyto části kódu jsou zapsány postupně v bodech, na které je v textu odkazováno.

- (1)

```
x <- filter(x, cells != "N")
NIMSI <- nrow(distinct(x, users))
Nrows <- nrow(x)
Ncells <- nrow(distinct(x, cells))
NBTS <- nrow(distinct(x, substr(cells, 1, 6)))
xG <- filter(x, substr(cells, 7, 7) < 3)
xU <- filter(x, substr(cells, 7, 7) == 3 | substr(cells, 7, 7) == 4)
xL <- filter(x, substr(cells, 7, 7) > 4)
substr(xU$cells, 7, 7) <- "_"
substr(xL$cells, 7, 7) <- "_"
Nsectors <- nrow(distinct(xG, substr(cells, 1, 6))) + nrow(distinct(xU,
substr(cells, 1, 8))) + nrow(distinct(xL, substr(cells, 1, 8)))
```

, při čemž x představuje datovou sadu, na kterou jsou funkce použita.
- (2)

```
BTS <- split(x, substr(x$cells, 1, 6))
```

, při čemž x představuje datovou sadu, na kterou je funkce použita.
- (3)

```
out <- lapply(BTS, function(x){
name <- substr(x[1,2], 1, 6)
G <- nrow(distinct(filter(x, substr(cells,7,7) == "1" | substr(cells,7,7) == "2"),
cells))
U <- nrow(distinct(filter(x, substr(cells,7,7) == "3" | substr(cells,7,7) == "4"),
cells))
LTE800 <- nrow(distinct(filter(x, substr(cells,7,8) == "51" | substr(cells,7,8)
== "61"), cells))
LTE1800 <- nrow(distinct(filter(x, substr(cells,7,8) == "52" | substr(cells,7,8)
== "62"), cells))
LTE2100 <- nrow(distinct(filter(x, substr(cells,7,8) == "53" | substr(cells,7,8)
== "63"), cells))
LTE2600 <- nrow(distinct(filter(x, substr(cells,7,8) == "54" | substr(cells,7,8)
== "64"), cells))
tibble(BTS = name,
GSM = G,
UMTS = U,
LTE800 = LTE800,
LTE1800 = LTE1800,
LTE2100 = LTE2100,
```

```
LTE2600 = LTE2600)
```

```
})
```

```
out <- do.call(rbind, out)
```

, při čemž *BTS* představuje list obsahující datové tabulky s informacemi o základnových stanicích.

```
(4) out2 <- tibble( GSM = sum(out$GSM),
```

```
UMTS = sum(out$UMTS),
```

```
LTE800 = sum(out$LTE800),
```

```
LTE1800 = sum(out$LTE1800),
```

```
LTE2100 = sum(out$LTE2100),
```

```
LTE2600 = sum(out$LTE2600))
```

, při čemž *out* představuje seznam BTS s údaji počtu buněk jednotlivých technologií.

```
(5) cutted <- split(x, x$users)
```

```
insidefunc <- function(x, long, technum1, technum2){
```

```
out <- filter(x, substr(cells, 7, long) == technum1 | substr(cells, 7, long) ==  
technum2)
```

```
out <- !(is.data.frame(out) && nrow(out)==0)
```

```
return(out)}
```

```
out <- lapply(cutted, function(x){
```

```
x <- distinct(x, substr(cells,7,8), .keep_all = TRUE)
```

```
colnames(x)[2] <- "cells"
```

```
users = as.character(x[1,1])
```

```
GSM = as.logical(insidefunc(x, 7, "1", "2"))
```

```
UMTS = as.logical(insidefunc(x, 7, "3", "4"))
```

```
LTE800 = as.logical(insidefunc(x, 8, "51", "61"))
```

```
LTE1800 = as.logical(insidefunc(x, 8, "52", "62"))
```

```
LTE2100 = as.logical(insidefunc(x, 8, "53", "63"))
```

```
LTE2600 = as.logical(insidefunc(x, 8, "54", "64"))
```

```
tibble(users = users,
```

```
GSM = GSM,
```

```
UMTS = U,
```

```
LTE800 = LTE800,
```

```
LTE1800 = LTE1800,
```

```
LTE2100 = LTE2100,
```

```
LTE2600 = LTE2600)
```

```
})
```

```
technologytab <- do.call(rbind, out)
```

, při čemž *x* představuje analyzovanou datovou sadu.

```
(6) nIMSI_technology <- c(GSM = nrow(filter(technologytab, GSM == TRUE)),
```

```
UMTS = nrow(filter(technologytab, UMTS == TRUE)),
```

```
LTE800 = nrow(filter(technologytab, LTE800 == TRUE)),
LTE1800 = nrow(filter(technologytab, LTE1800 == TRUE)),
LTE2100 = nrow(filter(technologytab, LTE2100 == TRUE)),
LTE2600 = nrow(filter(technologytab, LTE2600 == TRUE)))
```

, při čemž *technologytab* je seznam mobilních terminálů s údaji, na buňky kterých technologií se mobilní terminál připojil.

```
(7) nIMSI_onlyonetechnology <- c(GSM = nrow(filter(technologytab, GSM ==
TRUE & UMTS == FALSE & LTE800 == FALSE & LTE1800 == FALSE &
LTE2100 == FALSE & LTE2600 == FALSE)),
UMTS = nrow(filter(technologytab, GSM == FALSE & UMTS == TRUE &
LTE800 == FALSE & LTE1800 == FALSE & LTE2100 == FALSE &
LTE2600 == FALSE)),
LTE800 = nrow(filter(technologytab, GSM == FALSE & UMTS == FALSE &
LTE800 == TRUE & LTE1800 == FALSE & LTE2100 == FALSE & LTE2600
== FALSE)),
LTE1800 = nrow(filter(technologytab, GSM == FALSE & UMTS == FALSE &
LTE800 == FALSE & LTE1800 == TRUE & LTE2100 == FALSE & LTE2600
== FALSE)),
LTE2100 = nrow(filter(technologytab, GSM == FALSE & UMTS == FALSE &
LTE800 == FALSE & LTE1800 == FALSE & LTE2100 == TRUE & LTE2600
== FALSE)),
LTE2600 = nrow(filter(technologytab, GSM == FALSE & UMTS == FALSE &
LTE800 == FALSE & LTE1800 == FALSE & LTE2100 == FALSE &
LTE2600 == TRUE)))
```

, při čemž *technologytab* je seznam mobilních terminálů s údaji, na buňky kterých technologií se mobilní terminál připojil.

```
(8) Users_0_10 <- count(x, users)%>%
filter(n>=0)%>%
filter(n<=10)
```

, při čemž *x* představuje datovou sadu, která poskytuje záznamy uživatelů, a *n* představuje počet záznamů pro uživatele.

```
(9) list <- split(x, x$users)
sapply(list, function(x) mean(x$time), simplify = "array")
tibble( users = as.character(sapply(list, '[', 1, 1, simplify = "array")),
average_time = as.numeric(sapply(list, function(x) mean(x$time), simplify =
"array")))
```

, při čemž *x* představuje datovou sadu, na kterou je funkce použita.