

METODIKA SPRACOVANIA GEOLOKAČNÝCH ÚDAJOV MOBILNEJ SIETE**Príloha č.1: Opis predmetu zákazky****Návrh spracovania geolokačných údajov mobilnej siete**

Cieľom predkladanej metodiky je poskytnúť postup pre spracovanie geolokačných údajov z mobilnej siete na účely využitia v dopravnom modelovaní. Sledovaným javom je rozmiestnenie a mobilita užívateľov mobilnej siete (SIM kariet) na území Slovenska. Cieľom je odhadnúť počty pravidelne sa pohybujúcich osôb. Vzhľadom na citlivosť údajov a detailnosť ich spracovania bude agregácia do cieľových zón (obce SR) spracovaná na infraštruktúre mobilného operátora podľa nižšie uvedenej metodiky. Metodika je rozdelená do 6 krokov.

1. Príprava vstupných údajov

Pri príprave vstupných údajov je potrebné použitie takých SIM kariet, ktoré v čo najväčšej miere zodpovedajú fyzickým osobám (obyvateľom). V praxi ide teda o selekciu len tých užívateľov mobilnej siete, ktorí majú aktivovanú hlasovú službu, prípadne podľa typu zariadenia (vylúčia sa tak SIM karty v iných zariadeniach, ako napr. tabletoch, autách, alarmoch a pod.)

Typ lokalizačných údajov: CDR údaje + signalizačné údaje (location update)

Typ SIM: len SIM karty s hlasovou službou

Východisková priestorová jednotka: vyžarovacie polygóny BTS

Cieľová priestorová jednotka: katastrálne územia obcí SR (2927 obcí)

Sledované územie: SR

Dĺžka pozorovania: 28 dní (20 pracovných dní a 8 víkendových dní)

Obdobie pozorovania: 30.9.2024 - 27.10.2024

Tab. 1 Štruktúra vstupných údajov o aktivite SIM kariet v mobilnej sieti

MSISDN	Time_stamp	Id_BTS
ID_SIM_0001	2024-10-01 17:20:13	BTS-001

Tab. 2 Priestorová geometria vyžarovacích polygónov BTS

Id_BTS	Geometria vyžarovacieho polygónu (service area)
BTS-001	17,0574; 48,1556 ; 17,0566, 48,1776 ;...

Vstupné údaje je potrebné uchovať na backendoch mobilného operátora po celú dobu riešenia zadania, tak aby bolo možné optimalizovať metodiku spracovania týchto údajov. Je pravdepodobné, že bude potrebné počas analýzy vykonať korekcie v algoritmoch s cieľom zabezpečiť validné časové a priestorové agregáty. Presnosť výsledkov bude priebežne hodnotená porovnaním s referenčnými údajmi z národného cenzu (SODB 2021) a pomocou niekoľkých korelačných mier.

METODIKA SPRACOVANIA GEOLOKAČNÝCH ÚDAJOV MOBILNEJ SIETE

2. Konštrukcia modelu pokrytia

Model pokrytia má za úlohu priradiť každej BTS bunke podiel, ktorým pokrýva jednotlivé obce Slovenska. Určujúcim kritériom je objem budov v prieniku s vyžarovacím polygónom danej BTS. Vstupné údaje:

1. Priestorový model vyžarovacích polygónov
2. Bodová vrstva objemu budov v katastrálnych územiach obcí SR (2890 obcí). Body sú rozmiestnené vo vrcholoch štvorcov s rozstupmi 25 m. V atribútovej tabuľke je o každom bode údaj o objeme budov a o príslušnosti ku jednej z 2890 obcí SR. Bodovú vrstvu objemu budov dodá objednávateľ.

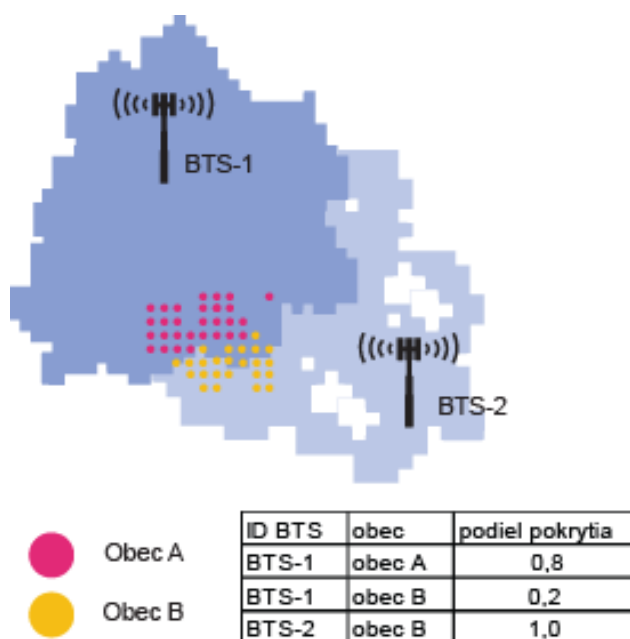
Priestorovým prienikom týchto dvoch vrstiev bude nasledujúca tabuľka:

Tab. 3. Podiel pokrytia buniek mobilnej siete v katastrálnych územiach obcí

ID BTS	obec ID	názov obce	podiel pokrytého objemu budov
BTS-001	508233	Stupava	0,8
BTS-001	507831	Borinka	0,2
BTS-002	508233	Stupava	1
BTS-003	508055	Lozorno	0,98
BTS-003	508233	Stupava	0,02

Kontrola výpočtu: Súčet podielu jednotlivých BTS = 1 (resp. 100%)

Pozn.: Nemusí platiť v prípade buniek mobilnej siete, ktoré obsluhujú cezhraničné územie.



Obr. 1. Schéma modelu pokrytia. Body reprezentujú objem budov vo štvorci 25x25 metrov.

METODIKA SPRACOVANIA GEOLOKAČNÝCH ÚDAJOV MOBILNEJ SIETE

3. Priradenie časových slotov

Základnou časovou jednotkou analýzy sú nasledujúce intervaly, ktoré reflektujú typický denný režim populácie. Ide o 100 slotov pre pracovné dni (5 slotov x 20 dní), 16 slotov pre soboty (4 sloty pre 4 dni) a 12 slotov pre nedele (3 sloty pre 4 dni), spolu 128 slotov.

Tab. 4. Časové intervaly spracovania údajov z mobilnej siete

	pondelok	utorok	streda	štvrtok	piatok	sobota	nedeľa
0:00							
1:00							
2:00	P1	P1	P1	P1	P1	S1	N1
3:00							
4:00							
5:00							
6:00							
7:00	P2	P2	P2	P2	P2	S2	N2
8:00							
9:00							
10:00							
11:00	P3	P3	P3	P3	P3	S3	N3
12:00							
13:00							
14:00							
15:00	P4	P4	P4	P4	P4	S4	
16:00							
17:00							
18:00							
19:00							
20:00	P5	P5	P5	P5	P5		
21:00							
22:00							
23:00							

4. Agregácia eventov pre individuálne SIM (MSISDN)

V ďalšom kroku je potrebné určiť tú bunku mobilnej siete, ktorá má v danom časovom slotе dominantné postavenie v individuálnych záznamoch užívateľa mobilnej siete (MSISDN). Principiálne existujú dva spôsoby:

- 1) *Event count*: pre každú SIM kartu (MSISDN) sa sčíta počet eventov (CDR+signalizačných) na jednotlivých bunkách mobilnej siete za každý časový slot pozorovania
- 2) *Time spent*: pre každú SIM kartu (MSISDN) sa sčíta dĺžka zotrvania na jednotlivých bunkách mobilnej siete za každý časový slot pozorovania.

Preferencia je pre druhý spôsob, ak by to však bolo pre spracovateľa údajov náročné, akceptovateľný je aj prvý spôsob.

METODIKA SPRACOVANIA GEOLOKAČNÝCH ÚDAJOV MOBILNEJ SIETE

Tab. 5 a tab. 6. Zachytávajú transformáciu údajov pre jednotlivé MSISDN

Tab. 5. Vstupné údaje z mobilnej siete.

MSISDN	Time_stamp	Id_BTS
ID_SIM_0001	2024-10-01 17:20:13	BTS-001
ID_SIM_0001	2024-10-01 17:25:11	BTS-001
ID_SIM_0001	2024-10-01 17:26:05	BTS-001

Tab. 6. Agregované údaje podľa časové slotu a počtu eventov / dĺžky zotrvania.

MSISDN	Time_slot	Id_BTS	Event_count	Time_spent
ID_SIM_0001	2024-10-01 P2	BTS-001	10	00:38:52
ID_SIM_0001	2024-10-01 P2	BTS-002	3	00:49:15

5. Identifikovanie mobility

V ďalšom kroku sa budú sledovať presuny SIM kariet na infraštruktúre mobilnej siete medzi zvolenými intervalmi (spolu 128 intervalov). Pre každý časový slot pozorovania sa určí BTS, kde mala daná SIM karta najväčší počet eventov (**primárna BTS**).

Tab. 7. Určenie primárnej BTS pre každý časový slot

MSISDN	Time_slot	Event_count (alebo Time spent)	BTS sa najväčším počtom eventov v danom časovom slotu (primárna BTS)
ID_SIM_0001	2024-10-01 P1	15 (02:08:52)	BTS-001
ID_SIM_0001	2024-10-01 P1	2 (00:15:11)	BTS-002
ID_SIM_0001	2024-10-01 P2	25 (04:01:15)	BTS-003
ID_SIM_0001	2024-10-01 P2	5 (00:15:12)	BTS-003
ID_SIM_0001	2024-10-01 P2	14 (01:25:30)	BTS-001

Následne sa agreguje počet SIM kariet pre každú po sebe nasledujúcu dvojicu časových slotov (napr. 2024-10-01 P1 – 2024-10-01 P2) a usporiadanú dvojicu BTS (východisková primárna BTS - cieľová primárna BTS). Výsledkom tejto agregácie bude nasledujúca tabuľka 8, ktorá zaznamenáva všetky vektory mobility medzi primárnymi BTS.

METODIKA SPRACOVANIA GEOLOKAČNÝCH ÚDAJOV MOBILNEJ SIETE

Tab. 8. Agregácia individuálnych tokov SIM kariet do matice tokov.

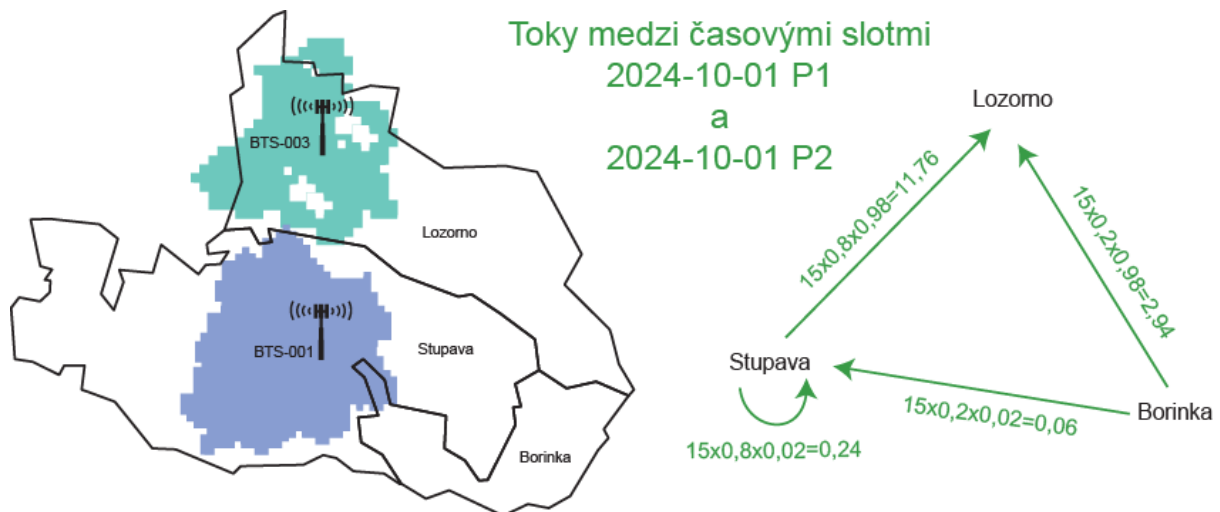
Dátum a časový slot (start)	Dátum a časový slot (end)	Východisková bunka	Cieľová bunka	Počet SIM
2024-10-01 P1	2024-10-01 P2	BTS-001	BTS-003	15
2024-10-01 P1	2024-10-01 P2	BTS-003	BTS-001	10

6. Agregácia tokov do úrovne obcí

Pri agregácii vektorov mobility do úrovne obcí budeme potrebovať tabuľku s podielom pokrytia (tab. č. 3) a maticu tokov medzi BTS v jednotlivých časových slotoch (tab. 8). Každý tok sa prepočíta (prenásobí) podielom pokrytia vo východiskovej aj cieľovej obci. Ak teda východisková bunka BTS zasahuje do dvoch obcí v pomere 80/20 a cieľová bunka do dvoch obcí v pomere 98/2, tak tok 15 SIM kariet medzi danými BTS sa prerozdelení do 4 tokov s veľkosťou, ktorú zachytáva tabuľka 9.

Tab. 9. Pravdepodobnostný model prerozdelenia tokov medzi usporiadanými dvojicami primárnych BTS

Dátum a časový slot (start)	Dátum a časový slot (end)	Východisková BTS	Cieľová BTS	Východisková obec ID	Východisková obec	Cieľová obec ID	Cieľová obec	Veľkosť toku
2024-10-01 P1	2024-10-01 P2	BTS-001	BTS-003	508233	Stupava	508055	Lozorno	$15 \times 0,8 \times 0,98 = 11,76$
2024-10-01 P1	2024-10-01 P2	BTS-001	BTS-003	507831	Borinka	508233	Stupava	$15 \times 0,2 \times 0,02 = 0,06$
2024-10-01 P1	2024-10-01 P2	BTS-001	BTS-003	508233	Stupava	508233	Stupava	$15 \times 0,8 \times 0,02 = 0,24$
2024-10-01 P1	2024-10-01 P2	BTS-001	BTS-003	507831	Borinka	508055	Lozorno	$15 \times 0,2 \times 0,98 = 2,94$



Obr. 2. Tok o veľkosti 15 SIM kariet sa prerozdelení podľa modelu pokrytia medzi 4 usporiadané dvojice obcí (zelená farba). Ako zachytáva obrázok, v prípade, keď východisková primárna

METODIKA SPRACOVANIA GEOLOKAČNÝCH ÚDAJOV MOBILNEJ SIETE

BTS a cieľová primárna BTS zasahujú do tej istej obce, je potrebné počítať aj s vnútorným tokom, kedy je východisková obec totožná s cieľovou obcou.

Toky medzi lokalitami pravidelnej lokalizácie budú následne agregované do úrovne obcí za každú po sebe nasledujúcu dvojicu časových slotov (127 dvojíc). Hodnoty je potrebné nezaokrúhľovať na celé čísla a ponechať v tvare na 2 desatinné miesta.

7. Štruktúra výstupných údajov

Výstupné údaje vo formáte CSV a s očakávanou veľkosťou niekoľko miliónov riadkov (127 slotov x max 80-tisíc medziobecných tokov = 10 mil. riadkov) obcí budú mať nasledovnú štruktúru:

Tab. 10. Štruktúra výstupných údajov

časový slot (start)	časový slot (end)	Východisková obec ID (muni_A)	Cieľová obec ID (muni_B)	Veľkosť toku (flow)
2024-09-30 P1	2024-09-30 P2	508233	508055	50,25
2024-09-30 P1	2024-09-30 P2	508233	507831	1,25*
...
2024-10-27 N2	2024-10-27 N3	528595	508055	15,4

*Tok, ktorý sa anonymizuje a bude mať hodnotu 1.

Extrapolácia

Extrapoláciu údajov na celú populáciu **nepožadujeme**.

Anonymizácia

Vzhľadom na citlivosť individuálnych záznamov a rešpektovanie platných právnych predpisov o ochrane súkromia by údaje boli anonymizované a agregované na infraštruktúre mobilného operátora. Aj po agregácii údajov sa však prirodzene medzi usporiadanými dvojicami priestorových jednotiek (obcí) vyskytnú toky, ktoré budú mať malý počet alokovaných SIM kariet. Malé toky **väčšie ako 0 a menšie ako 5** budú označené jednotne číslom "1". Pre redukciu veľkosti tabuľky je tiež vhodné **neuvádzať nulové toky**.