WSN Dataset Generator

Afonso Oliveira, Teresa Vazão
INESC-ID / Instituto Superior Técnico / Universidade de Lisboa

I. Overview

The generated dataset is composed by traffic and signal quality information. Figure 1 illustrates the generation process with the interaction between the different modules. The generation process uses as a base Simulation of Urban MObility (SUMO) [1], a mobility simulator that generates traces of pedestrian and vehicular movement on a given map.

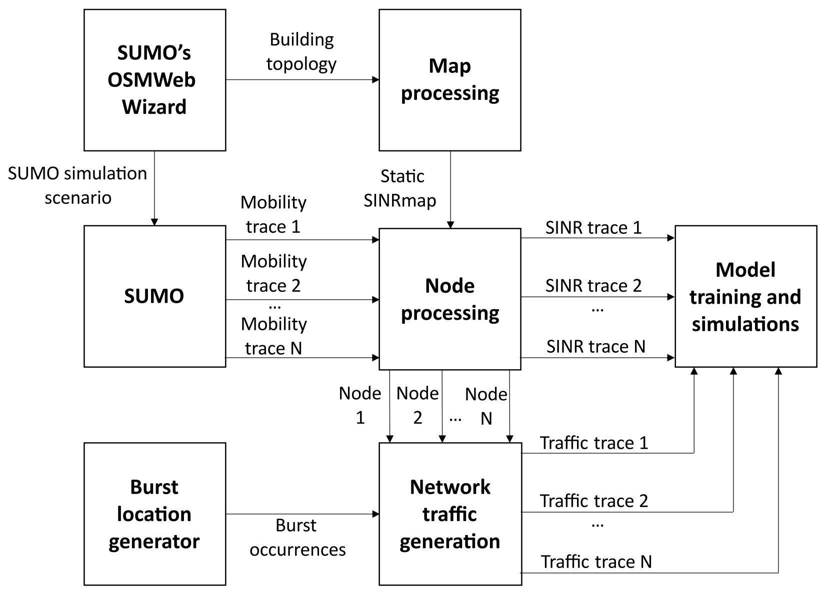

Fig. 1. SINR and traffic synthetic trace generation

SUMO includes a scenario generator that extracts maps from Open Street Maps (OSM). Also, it generates a scenario with pedestrian and vehicle placement along with desired itineraries so that it can be used by the SUMO simulator to generate mobility traces. The information extracted from OSM also includes the building topology, which is used by the **Map processing** module to calculate which areas are in Line of Sight (LoS) of the best cell it can be connected to as well as the pathloss and shadow fading of the signal. This information is then processed along with the mobility traces by the **Node processing** module to generate the final Signal-to-Noise Ratio (SNR) traces.

The traffic generation of each node is based on its characteristics (vehicle, passenger, pedestrian, static) and the burst events that affect some traffic types. The traffic generation is done for a set of defined traffic types. The following sections provide a detailed description of the used modules.

II. SUMO and SUMO's OSM Web Wizard

SUMO is a mobility simulator that simulates the movement of pedestrians and vehicles with predetermined itineraries, which may be done following a main route or backup ones. To generate activity and maps, one possibility is to use a tool named OSM Web Wizard provided by SUMO that extracts map information from OSM and generates vehicle activity based on it.

This work uses OSM Web Wizard to generate activity for the SUMO simulations and extract the information on the map. Using map information the area topology in terms of buildings (*Building topology*) is given to the **Map processing** module so that LoS and non-LoS areas can be calculated. The activity generated for the SUMO simulation is fed to SUMO through the form of a *SUMO simulation scenario* which contains the pedestrians and vehicles to be simulated along with their respective itineraries.

With the scenario as input, SUMO generates *mobility traces* of vehicles and pedestrians which are exported in SUMO's Floating car Data (FCD) file format that includes information such as position, speed, and direction angle. The traces are fed to the **Node processing** module for the generation of simulation nodes and calculation of fast fading models.

III. Map processing

The map processing module decides on the base station locations. Each base station includes 3 cells 120º apart. From each cell the map processing module calculates the pathloss to other relevant map points. The pathloss calculation depends on whether the target point is or not in LoS with the cell, which is decided based in the building topology provided by SUMO's Web Wizard.

The Map processing module exports a map for each cell with the SNR received at different points 10m apart from each other. The SNR maps are also compiled into one static SNR map containing the cell that gives the better signal quality to each specific point. The **Node processing** module uses this information to choose the cell it calculates the SNR relative to at each timestamp depending on the device position. The following sections detail how the Map processing module calculates the received SNR values.

A. Cell placement

The cells are placed on the map around a chosen location. In this work seven base station with three-sectored cells were created as represented in Figure 2. The Intersite Distance (ISD) of 500m was chosen in accordance with the Urban Macro scenario from a 3GPP report on channel models [2].

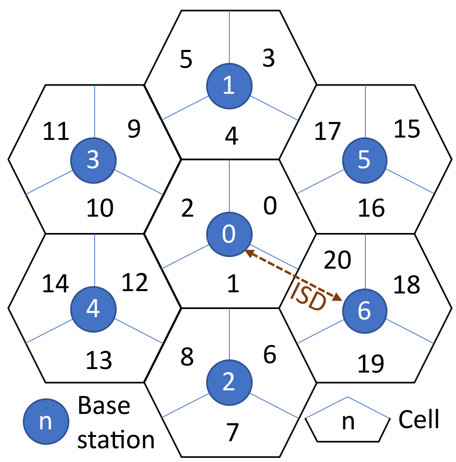

Fig. 2. Cell placement.

B. Static SNR calculation

For calculating the SNR to and from each device, equation (1) is used. For each transmitting (tx) or receiving (rx) node, P_{tx} represents the transmission power, G the directional antenna gain, NF the noise factor, PLF represents the pathloss and fading, TN the thermal noise, and BW the carrier bandwidth. The parameters for this calculation are taken from the Configuration A of the Urban Macro (UMa) environment in the 3GPP report [2]. The used parameters and calculations are summarized in Table III, which is complemented by Table II that detail the equations for the pathloss calculation.

$$\begin{aligned} SNR =& P_{tx} + G_{tx}(\theta, \phi) + G_{rx}(\theta, \phi) \\ & - PL - SF - NF_{rx} - TN - 10\log(BW) \end{aligned} \tag{1}$$

As each base station contains three cells, the antenna gain for each cell is directional. The equations for antenna gain are extracted from the same 3GPP report [2] and presented in Table III, where θ represents the vertical angle, and ϕ the horizontal angle between the direction of the cell source and the line between the UE position and the cell source being considered.

The use of multiple antennas for transmitting and receiving is not considered. The variation characteristics provided by position and speed are considered to be the most important for this work and multi-antenna consideration was discarded to generate the dataset in a feasible time frame.

The pathloss and slow fading are static per position and calculated in the **Map processing** module. The fast fading component is then added in the **Node processing** module, since it varies with time and depends on node specific characteristics such as speed.

The calculations for the pathloss and slow fading component are done from each cell source to separate points in the map 10m meters apart and forming a grid. Points farther then 500m from the cell source are not considered. Each point can be in LoS with the cell or without LoS (nLoS) depending if a direct line between the cell source and the point being considered

TABLE I
SNR CALCULATION PARAMETERS [2].

Parameter	Value
P_{cell} (dBm)	46
$G_{cell}^{\max}$ (dBi)	8
$G_{cell}(\theta, 0)$ (dBi)	$\min\left\{12\left[(\theta - 90°)/65°\right]^2, 30\right\}$
$G_{cell}(0, \phi)$ (dBi)	$\min\left[12\,(\phi/65°)^2, 30\right]$
$G_{cell}(\theta, \phi)$ (dBi)	$G_{cell}^{\max} - \min\left[G_{cell}(\theta, 0) + G_{cell}(0, \phi), 30\right]$
NF_{cell} (dB)	5
P_{UE} (dBm)	23
$G_{UE}(\theta, \phi)$ (dBi)	0
NF_{UE} (dB)	7
TN (dBm/Hz)	-174
BW (Hz)	20×10^6

is obstructed by a building taken from the OSM map. A special case for points without line of sight is when they are inside buildings (IN). In this case, the propagation characteristics are different and the pathloss and slow fading vary accordingly.

The used parameters and equations for pathloss and slow fading are taken from the mentioned 3GPP report [2] following the Configuration A of the Urban Macro (UMa) environment. Table II lists the equations used for the different levels of obstruction: LoS, $nLoS$, and IN. The values of d_{3D}, d_{2D}, and d_{2D-in} respectively represent the distance between the cell and the point being considered, its horizontal projection, and the component of that horizontal projection inside a building. In the IN obstruction level, the $PL(intersection\_edge)$ refers to the path loss where the line between the cell source and the point being considered. Slow fading (SF) is generated as a random value taken from a normal distribution which is added to the the pathloss value.

The values of h_{BS} and h_{UE} represent the height of the Base Stations (cell source) and the UEs (point being considered), respectively. f_c represents the carrier frequency. Finally, d_{BP} is a breaking point distance where the calculation of the pathloss in LOS condition changes. It is calculated according to a formula where c represents the speed of light. These parameters are also listed in Table III.

TABLE II
PATHLOSS AND SHADOW FADING CALCULATIONS [2].

Obst.	PL (dB)	SF (dB)
LoS	$\begin{cases} 28 + 22\log(d_{3D}) + 20\log(f_c) & \text{if } d_{2D} < d_{BP} \\ 28 + 40\log(d_{3D}) + 20\log(f_c) & \\ -9\log\left(d_{BP}^2 + (h_{BS} - h_{UE})^2\right) & \text{if } d_{2D} \geq d_{BP} \end{cases}$	$\mathcal{N}(0, 6)$
nLoS	$\max(PL_{LoS}, 13.54{+}39.08\log(d_{3D}){+}20\log(f_c){-}0.6(h_{UE} - 1.5))$	$\mathcal{N}(0, 6)$
IN	$PL(intersection\_edge) + 20 + 0.5d_{2D-in}$	$\mathcal{N}(0, 7)$

TABLE III
PATHLOSS CALCULATION PARAMETERS [2].

Parameter	Value
f_c (Hz)	4×10^9
h_{BS} (m)	25
h_{UE} (m)	1.5
d_{BP} (m)	$4(h_{BS} - 1)(h_{UE} - 1)f_c/c$

C. Output static maps

The **Map processing** module exports a map for each cell containing the calculated static SNR values at each point. A global map is also generated containing the best static SNR value for each point along with the associated cell. Figure 3 illustrates this global map in the form of a heatmap with the values of the best SNR value for each point.

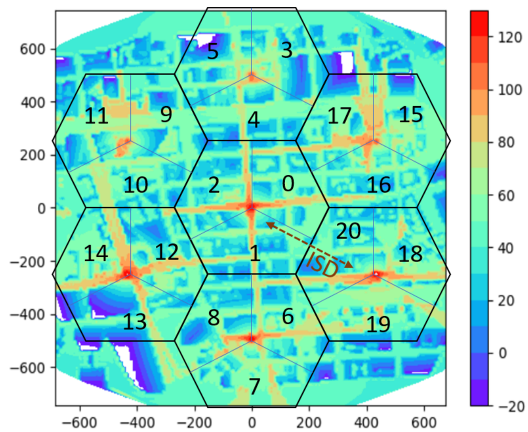

Fig. 3. SNR heatmap with generate cells. SNR in dBs.

IV. Node processing

The **Node processing** module generates the UEs based in the SUMO traces. It also generates static nodes, which do not exist in the SUMO simulation. Following that, the module calculates the SNR of each UE over time using by adding calculated fast fading traces to the static SNR according to the UE position at each time.

A. UE generation

The UEs are generated based on the SUMO traces. The generated SUMO traces include three types of nodes: pedestrian, and vehicle. Each pedestrian is associated with a UE. Vehicles are in itself a UE, for eV2X applications, and include as much UEs as passengers inside. The number of passengers is taken from a uniform distribution. Additionally static UEs are dispersed along the map in a random manner following a uniform distribution for the values of both horizontal axis within the bounds of the map generated by the **Map generation** module. The positions are re-generated so that a configured percentage of the static UEs is inside buildings and the remaining outside. In total there are four UE node types: *pedestrian*, *vehicle*, *passenger*, and *static*. Table IV lists the parameters for the UE generations.

TABLE IV
UE generation parameters.

Parameter	Value/Distribution
Number of static UEs	5000
Static UE location	continuous uniform within map bounds
Static UEs inside (%)	80
Passengers per vehicle	discrete uniform $(1, 4)$

B. SNR traces

One SNR trace is created for each UE. The SNR values vary over time depending on the UE position and the impact of the modeled fast fading. For simplification due to the amount of samples generated, multi-path fading is simulated using a simple single antenna configuration. For that the Method of Exact Doppler Spread (MEDS) [3] is used to simulate the signal envelope under a Rician channel. From each movement trace one or more channels are generated since the model simulates static scenarios, for instance with the same speed and obstruction level. The movement trace is divided into different channels depending on the variability of the speed, direction or obstruction level of the UE.

1) Fast fading model: The fast fading model is done by generating a Rician channel. Equation (2) describes a deterministic model for generating the envelope of a generic Rice process. To generate a Rice process with the MEDS the definitions in Table V are applied to the Rice process generation [3]. The value of N as well as the sampling period T_s, which defines the time between two consecutive t values in equation (2) are custom configurations. The values of f_{max}, ρ, σ_0, θ_ρ, f_ρ are parameters that is specific of a channel and depends on UE characteristics. The used values are presented in Table VI.

$$\xi(t) = \left\| \sum_{n=1}^{N} \left[C_{1,n} \cos\left(2\pi f_{1,n} t + \theta_{1,n}\right) \right] + \rho \cos\left(2\pi f_\rho t + \theta_p\right) + j \left\{ \sum_{n=1}^{N} \left[C_{2,n} \cos\left(2\pi f_{2,n} t + \theta_{1,n}\right) \right] + \rho \cos\left(2\pi f_\rho t + \theta_p\right) \right\} \right\|, \quad \forall t \in \{n T_s | n \in \mathbb{N}\} \tag{2}$$

TABLE V
RICE PROCESS PARAMETERS USING MEDS [3].

Parameter	Value
$C_{i,n}$	$\sigma_0 \sqrt{2/N}, \forall_{i\in\{0,1\}, n\in\{1,2...N\}}$
$f_{i,n}$ (Hz)	$f_{max} \sin\left[(n-1/2)\ \pi/(2N)\right], \forall_{i\in\{0,1\}, n\in\{1,2...N\}}$
$\theta_{i,n}$ (rad)	$\mathcal{U}(0, 2\pi), \forall_{i\in\{0,1\}, n\in\{1,2...N\}}$
f_{max} (Hz)	$v f_c / c$

TABLE VI
MEDS PARAMETERS FOR EACH CHANNEL.

Parameter		Value
N		200
T_s (s)		1×10^{-3}
ρ	LoS	$\sqrt{K/(K+1)}$
	nLoS	0
σ_0	LoS	$\sqrt{1/[2(K+1)]}$
	nLoS	$\sqrt{1/2}$
f_ρ (Hz)		$f_{max} \cos\Phi$
θ_ρ (rad)		$-2\pi d_{3D} f_c / c$
K		$10^{K_{dB}/10}, K_{dB} \sim N(\mu_k, \sigma_k^2)$
μ_k		9
σ_k		3.5

2) Channels per movement trace: Four different states of an UE have an impact on the Rician channel being generated. They are the UE's speed (v), the horizontal angle between the movement and direction to the cell source (Φ), the obstruction level, and the distance between the cell source and the UE (d_{3D}). The generation of a new channel is triggered when the obstruction level or the connected cell changes or when the UE's speed or direction varies by 10%.

3) Channel specific parameters: The parameters specific for each channel are listed in Table VI. The values of ρ, σ_0, depend on the obstruction level. Note that the obstruction level referred as inside (IN) is also in nLoS. The values for generating K are taken from the 3GPP report [2] and refer to the Urban Macro scenario.

4) UE inside vehicles: As they are inside a vehicle, additional pathloss applies to passenger UEs. This additional obstruction to the signal is also taken from the mentioned 3GPP report [2]. The value to be added is randomly generated per passenger and follows the normal distribution $\mathcal{N}(9, 5)$.

V. TRAFFIC GENERATION

A. Traffic types

The network traffic for the synthetic dataset is generated based on a set of traffic classes that aim to emulate different applications expected to be part of a 5G mobile wireless network. These classes are: Machine Type Communications (MTC) sensors, offline video, vehicle based information, VoIP, and Web. Their generation procedure is detailed in Table VII.

1) MTC sensors: MTC sensors are emulated using an exponentially distributed inter packet time and a truncated normal distribution for the packet size. In a survey on traffic models by Navarro-Ortiz *et al.* this was identified as recurrent traffic model appointed by different entities such as **3GPP!** (**3GPP!**).

2) Offline video: Offline video is simulated using a different approach than a simple Constant Bitrate (CBR) transmission. Through observation of traces from Youtube sessions, we noticed chunks of the video are loaded depending on the playback offset position. To simulate offline video traces from the network generated by Youtube videos using an uninterrupted playback was used.

TABLE VII
TRAFFIC GENERATION PARAMETERS. TIME METRICS IN SECONDS, DISTANCE IN METERS AND INFORMATION IN BYTES.

Type	Component	Distribution	Parameters / Values	Direction	Slice	Comment
V2I (CAM)	Inter packet arrival (s)	Truncated normal	$\mu = 0.4$; $\sigma = 1$; min = 0.1; max = 1	Uplink	URLLC	Base in CAM message report [4]
	i^{th} packet size (B)	Truncated normal	$\mu = 400$; $\sigma = 200$; min = 200; max = 500			
V2I (DENM)	Inter burst occurrence time (s)	Exponential	$1/\lambda = 60$	Uplink	URLLC	Geo-located burst occurrences trigger messages. Packet size from DENM message.
	Burst position (m)	Uniform	Not applicable			
	Burst radius (m)	Fixed	50			
	Packet size (B)	Fixed	320			
Offline Video	Inter packet arrival (s)	Lognormal	Fitted to type	Downlink	eMBB	Types correspond to real traces. Possible types: music video, sports compilation, mostly static tutorial, cooking, reaction.
	Clip duration (s)	Normal	$\mu = dur(type)$; $\sigma = 0.3 * dur(type)$			
	i^{th} Packet size (B)	Normal	$\mu = size(type, i)$;[1] $\sigma = 0.3 * size(type, i)$			
	Inter session time (s)	Exponential	$1/\lambda = 3600$			
	Session duration (s)	Exponential	$1/\lambda = 1200$			
Web	Inter session time (s)	Exponential	$1/\lambda = 600$	Downlink	eMBB	Each web session consists in a given number of pages. Each page contains a given number of objects [5].
	Pages per session	Normal	$\mu = 3.86$; $\sigma = 2.465$;			
	Page size (B)	Truncated log-normal	$\mu = 8.37$; $\sigma = 1.37$; min = 100; max = $20 * 10^6$			
	Number of objects	Truncated Pareto	$\alpha = 1.11$; $k = 2$; min = 0; max = 53			
	Object size (B)	Truncated log-normal	$\mu = 2.36$; $\sigma = 6.17$; min = 50; max = $20 * 10^6$			
	Parsing time (s)	Exponential	$1/\lambda = 0.13$			
	Reading time (s)	Exponential	$1/\lambda = 3$			
VoIP	Inter call arrival (s)	Exponential	$1/\lambda = 600$	Both	eMBB	When active a 40 byte packet is sent every 20 ms, otherwise a 16 byte packet is sent every 160 ms [5].
	Call duration (s)	Exponential	$1/\lambda = 180$			
	Voice activity	Markov Chain w/ sampling T	$p(on\|off) = 0.01$; $p_{(}off\|on) = 0.01$; $T = 0.02$			
MTC	Inter packet arrival (s)	Exponential	$1/\lambda = 234$	Uplink	MTC	Negative size: new random generation.
	i^{th} packet size (B)	Normal	$\mu = 200$; $\sigma = 60$			

A set of videos ranging from sports, to music, or simple popular reaction videos was collected and set as possible templates. From these templates, the inter packet arrival is fitted to a lognormal distribution using Python's lybrary *scipy*, the duration of a video clip is taken from a normal distribution based on the template clip duration, and each i^{th} packet length is also taken from a normal distribution based in the template's i^{th} packet length. The last 2/3 of the packet length list are repeated to accommodate for generated video traces that have a higher duration than the template duration. Only the last values are repeated since it was observed that the first packets had considerably higher lengths to provide a faster initial load of the video clip. Table VIII lists the used templates along with their Maximum Likelihood Estimates (MLEs) of the parameters in the lognormal distribution fit if the Inter Packet Arrival (IPA) and clip duration.

TABLE VIII
IPA DISTRIBUTION PARAMETERS AND DURATION FOR EACH TEMPLATE

Template	IPA - parameter's MLE			Duration (minutes)
	Shape	Location	Scale	
News	4.84×10^{-3}	3.28×10^{2}	3.31×10^{2}	21:21.701
Cooking	1.61	3.48×10^{-3}	1.79	07:37.381
Music	2.40	2.84×10^{-3}	8.72×10^{-1}	04:00.501
Static	3.41×10^{-1}	-4.45	8.15	16:38.661
Reaction	1.51×10^{-1}	-1.92×10^{1}	2.37×10^{1}	12:38.401
Sports	4.70×10^{-1}	-4.21	8.75	10:07.041

3) Vehicle based information: Vehicle based communication portraits V2I communication. The traffic generation is based in periodic and emergency traffic. Periodic traffic was generated from random distributions configured based in a report of Cooperative Awareness Message (CAM) messages by Car 2 Car Communication Consortium (C2C-CC) [4]. For emergency traffic, random emergency areas are generated that force nodes within that area to trigger a Decentralized Environmental

TABLE IX
TRAFFIC DIRECTION AND SLICES.

Traffic	Direction	Slice
V2I (CAM)	Uplink	URLLC
V2I (DENM)	Uplink	
Offline Video	Downlink	eMBB
VoIP	Both	
Web	Downlink	
MTC	Uplink	MTC

TABLE X
TRAFFIC TYPE ALLOCATED TO EACH NODE TYPE

Source type	Node type	Traffic
Person	Passenger, Pedestrian, 60% Static	20% Idle, 20% Offline Video, 20% VoIP, 40% Web
Vehicle	Vehicle	V2I (CAM+DENM)
Machine	40% Static	MTC

Notification Message (DENM) packet. These areas are generated using an uniform distribution for a center with a fixed radius. The time in between the occurrence of bursts is exponentially distributed.

4) VoIP: The traffic model for VoIP is a model with silence suppression according to a 3GPP guidelines [5]. The model follows a conversation on-off switch to dictate when a person is talking (and traffic is flowing normally) or not. In addition to the 3GPP model, the inter-call and duration times are sampled from exponential distributions.

5) Web: Web traffic also follows a traffic model from 3GPP guidelines [5], where a user requests web pages that are processed and require their specific objects. After a given reading time from the last web page request, a new web page is requested.

B. Traffic sources

Traffic is distributed to the different node types according to a defined traffic source type depending who or what is operating the UE: a *person* using general broadband services, a *vehicle* sending CAM and DENM messages, and a *machine* sending sensor information uplink. The mapping between source types to node types and the equivalent traffic types used is listed in Table X.

C. Traffic scenarios

To study the solution under different traffic scenarios different traffic behaviors were created in the traffic of the eMBB slice. Besides the *default* generation in Table VII, different parameterization is specified to originate an *occasional* and *constant* behavior. The parameters used for each behavior are listed in Table XI.

TABLE XI
TRAFFIC BEHAVIORS.

Traffic	Behavior	Component	Value
Offline Video	Default	Inter session time (t)	$1/\lambda = 1200$
		Session duration (t)	$1/\lambda = 1200$
	Occasional	Inter session time (t)	$1/\lambda = 2400$
		Session duration (t)	$1/\lambda = 600$
	Constant	Inter session time (t)	$1/\lambda = 0$
		Session duration (t)	$1/\lambda = 1200$
VoIP	Default	Inter call arrival (t)	$1/\lambda = 600$
		Call duration (t)	$1/\lambda = 180$
	Occasional	Inter call arrival (t)	$1/\lambda = 1200$
		Call duration (t)	$1/\lambda = 60$
	Constant	Inter call arrival (t)	$1/\lambda = 10$
		Call duration (t)	$1/\lambda = 180$
Web	Default	Inter session time (t)	$1/\lambda = 600$
	Occasional	Inter session time (t)	$1/\lambda = 1200$
	Constant	Inter session time (t)	$1/\lambda = 10$

Using the defined traffic behaviors three scenarios are created: *standard*, which exclusively uses the default behavior, *sporadic*, which exclusively uses the occasional behavior, and *undistributed* that uses a combination of behaviors to provide a scenario where a smaller number of UEs is responsible for more traffic. The scenarios are specified in Table XII.

[1]For index i values higher than the type's number of packets the index circulates over the last 2/3 of the packets.

TABLE XII
TRAFFIC BEHAVIOR USED IN EACH SCENARIO

Scenario	Traffic behaviors
Standard	100% Default
Sporadic	100% Occasional
Undistributed	25% Idle, 25% Occasional, 50% Constant

VI. OUTPUT

The output of the dataset returns three main types of traces: SNR traces, traffic traces and aggregated traces. The first are traces by node with the SNR value over sample periods. The second group is composed by traces that have one entry for each time a traffic chunk is generated with the specific timestamp. Finally, a third dataset combines both by averaging the signal quality levels and summing the generated traffic for each node and slice through different time intervals.

A. SNR traces

The SNR traces contain measures of the SNR value over sample periods of 1 ms. They are composed by the columns:

- **Time**: the timestamp of the SNR measure;
- **UE**: the UE with the SNR measure;
- **X**: X position according to the SUMO map;
- **Y**: Y position according to the SUMO map;
- **SNR**: calculated SNR at instant **Time**;
- **MCS**: the Modulation Coding Scheme (MCS) used according to the SNR;
- **Cell**: the cell which the UE is attached to;

The choice of MCS, according to a node's signal quality uses the same process that is used by the NS-3 network simulator [6]. The process consists in converting the signal quality of a UE u, SNR_u, to the corresponding spectral efficiency, η_u, according to the Shannon theorem [7], as described in equation 3. To better adapt the result to a real environment, there is the coefficient $\Gamma = 1.5$, dependent on the target Bit Error Rate (BER) = 10^-5. The MCS is taken from a table from the **3GPP!** standard [**?**], where the MCS with the highest spectral efficiency lower than η_u is selected.

$$\begin{aligned} \eta_u &= \log_2\left(1 + \frac{SNR_u}{\Gamma}\right) \\ \Gamma &= -\frac{\ln(5 \cdot BER)}{1.5} \end{aligned} \tag{3}$$

B. Traffic traces

The traffic traces are composed traces recorded at the timestamps when a new traffic packet is triggered. They are composed by:

- **Time**: Timestamp of the triggered traffic packet;
- **UE**: the traffic source or destination;
- **Direction**: uplink or downlink;
- **Bytes**: length of the packet in bytes;
- **Types**: The types of traffic used by the UE;
- **Slice**: slice of the traffic.

C. Aggregated traces

Since the SNR traces have a fine time period of 1 ms, and to have a workable dataset for some applications, a new aggregated dataset is generated containing aggregated information of both the SNR and traffic traces within time windows of 10 seconds. This trace is composed by the columns:

- **Time**: time of the beginning of the time window;
- **UE**: the UE the dataset entry is relative to;
- **Cell**: The cell the UE is attached to. If one UE has been in two cells over one time window, there is a UE entry for each cell;
- **Slice**: the slice of the UE the dataset entry is relative to;
- **Bytes**: the sum of packet lengths within the time window;
- **SNR**: the average SNR within the time window;
- **MCS**: MCS calculated with the average SNR as input;

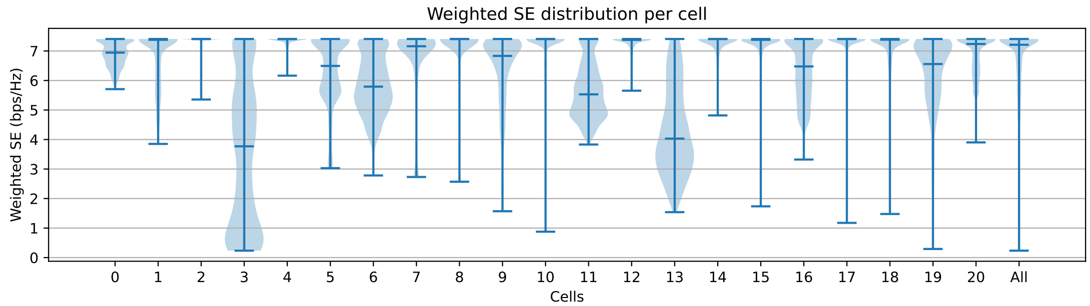

Fig. 4. Weighted spectral efficiency distribution per cell in the downlink of the standard scenario

- **SE**: spectral efficiency;
- **RBs**: the number of Resource Blocks (RBs) required to schedule the summed packet lengths;
- **UE Ratio**: UE participation within the slice based in the used RBs;

The equation for the UE ratio of a given UE u (r_u) is presented in Equation (4). In the equation RBS_u represents the used RBs by UE u, and $U_{c,s}$ is the set of all UEs from the slice s attached to the cell c.

$$r_u = \frac{RBS_u}{\sum_{n \in U_{c,s}} RBS_n} \tag{4}$$

D. Distribution visualization

Figure 4 represents violin plots of the distribution of the traffic weighted average of the spectral efficiency (SE), or simply weighted SE, for the standard scenario in the downlink direction. For each cell, the edge marks represent the minimum and maximum value and the middle mark the median value. As can be observed some cells represent different distributions. For instance, while cell 4 has a very predictable behavior, cells 3 and 13 have a spread distribution of the weighted SE, making it distribution theoretically more difficult.

To understand the impact of the different scenarios we calculate the cost of contracting the same resources as the ones required in the last monitoring period without over-provisioning. The value $\alpha = 500$ was used and the bitrate target set to $1Mbps$. The monitoring period is the same as the update period, which is 60 s.

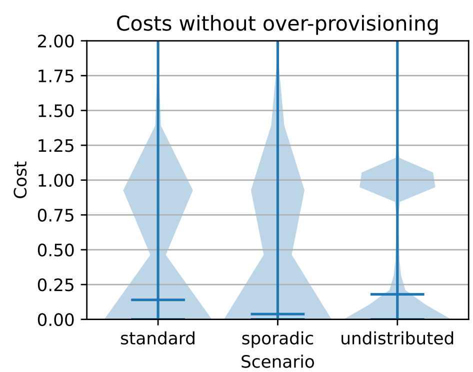

Fig. 5. Costs without over-provisioning in the downlink

The result is in Figure 5 that portraits the different scenarios in the downlink direction, which is where the difference in behaviors is more accentuated due to the affected traffic types. In the undistributed scenario the over-provisioning happens less frequently, since the obtained values are either close to 0 (low over-provisioning) or 1 (contract violations), which means the value for the next window is almost always similar to the previous. Since less nodes are responsible for the existing traffic, the weighted SE does not vary much. The sporadic scenario ends up having lower costs than the standard.

Figure 6 shows the same costs without over-provisioning for the different slice, direction pairs. Note that each direction, Downlink (DL) or Uplink (UL), has its of resource pool and for that reason in this work they are treated almost as different slices. The URLLC slice, which is the vehicular application, also has close values since the vehicles move outside where the connection quality is better. It should be noted that higher MCS schemes have a lower spectral efficiency difference between them. It seems from the plot of the eMBB slice in the uplink direction has higher costs with over-provisioning.

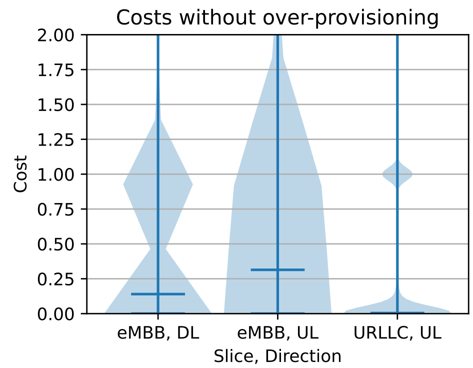

Fig. 6. Costs without over-provisioning in the standard scenario.

References

[1] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. Microscopic traffic simulation using sumo. In *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE, 2018.

[2] 3GPP. Study on channel model for frequencies from 0.5 to 100 GHz. Technical Report (TR) 38.901, 3rd Generation Partnership Project (3GPP), 10 2019. Version 16.0.0.

[3] M. Patzold, U. Killat, F. Laue, and Yingchun Li. On the statistical properties of deterministic simulation models for mobile fading channels. *IEEE Transactions on Vehicular Technology*, 47(1):254–269, 1998.

[4] Vincent Martinez and Friedbert Berens. Survey on CAM statistics. Technical Report TR2052, Car 2 Car Communication Consortium (C2c-CC), 12 2018. Version 1.0.1.

[5] TSG-RAN1. LTE physical layer framework for performance verification. Meeting R1-070674, 3rd Generation Partnership Project (3GPP), 02 2007. 3GPP TSG-RAN1#48.

[6] Giuseppe Piro, Nicola Baldo, and Marco Miozzo. An lte module for the ns-3 network simulator. *Proceedings of the 4th International ICST Conference on Simulation Tools and Techniques: March 2011*, 01 2011.

[7] Proakis. *Digital Communications 5th Edition*. McGraw Hill, 2007.