Slice-aware energy saving algorithm for 5G networks based on simplicial homology

Aymeric de Javel, J.S. Gomez, P. Martins, J.L. Rougier, P. Nivaggioli

Slice-aware energy saving algorithm for 5G networks based on simplicial homology.

A. de Javel, J.S. Gomez, P. Martins, J.L. Rougier, P. Nivaggioli

***Abstract*—Network slicing is an important feature introduced by 5G. It enables an operator to deploy multiple independent end-to-end virtual networks, called network slices, on top of a common physical infrastructure. Network slices require the allocation of resources at different levels of the system, among which power budget is of upmost importance at RAN level. This work proposes a new power allocation algorithm for 5G network slicing. Network slices are introduced as a set of five parameters that reflect slices capacity, latency and reliability requirements. Simplicial homology is used to analyse network coverage. Reliability is then defined in terms of a multiple network connectivity criteria. The performances of the proposed algorithm are evaluated by simulations for a scenario with three different network slices: eMBB, uRLLC and mMTC. This method can be used to determine the required power budget at RAN level for different 5G network slices deployment scenarios.**

***Index Terms*—network optimization, simplicial homology, network slicing, 5G system behavior.**

I. Introduction

5G addresses a wide range of business use cases (called verticals) ranging from industry 4.0 to smart cities and augmented reality. Each usage has specific and opposite network requirements such as delay, throughput, reliability and supported devices density that are difficult to satisfy on top a common physical infrastructure. Network slicing has been introduced to address this critical issue. The network slice concept is defined to support the service quality required by each vertical on top a common physical infrastructure. Network slicing grants also isolation between the resources and the flows associated to different verticals. In this context, physical resources, located at different network levels, have to be allocated to the different slices.

Many solutions have been proposed in the literature to address these different issues, mainly from a datacenter perspective. The resources allocated are CPU, memory and bandwidth. Allocating those resources can be done by using classic linear optimization algorithms like in [1] and [2]. However, the previous methods do not take into account Radio Access Network (RAN) specific constraints where power budget is of the upmost importance. This work investigates the relation between power optimization in RAN and slices requirements fulfillment. The capacity (throughput and devices density) and the reliability criteria have been introduced as the main slices requirements. Reliability is modelled as a multiple connectivity constraint. For that matter, simplicial homology is used for studying network coverage and connectivity. It has already been successfully applied for energy saving methods, like in [3]. Given that coverage constraints are hardly linear by nature, classic optimization methods cannot be used to solve the power allocation problem. Specific heuristics have to be developed. This work provides a near optimal heuristic for 5G slice-aware systems based on simulated annealing. It can be used to determine the amount of resources required to deploy a new 5G network slice on a given physical network.

The first part of this paper details the system model where 5G flexible numerology is used to fit slices latency and throughput requirements, the second one describes simplicial homology, the third one explains the methods used to reduce power budget and the last part presents our simulations results and analysis.

A. de Javel, J.S. Gomez, P. Martins and J.L. Rougier are with the Institut Polytechnique de Paris (IP Paris), Télécom Paris, LTCI (Paris, France), contact e-mail: aymeric.dejavel@telecom-paris.fr

P. Nivaggioli is with Cisco (Paris, France).

II. System model

A. Cellular network

We consider a cellular network in which cells are represented by circles. Circles center are base stations and circles radius are cell's coverage zone.

The slice s capacity κ_s (in bits per seconds) and Signal over Interference plus Noise Ratio (SINR) η_s are given by the following equations. W_s is the slice allocated bandwidth, $P_r^i(r)$ is the received power from cell i, I the interference and N_s the noise.

$$\kappa_s = W_s \cdot \log_2(1+\eta_s), \quad \eta_s = \frac{P_r^i(r)}{I+N_s}.$$

The received power (in Watts) i at a distance r from an emitter which emission power is P_e^i is given by equation 1 where r_0 is the reference distance, γ is the path-loss exponent and λ is the signal wavelength.

$$P_r^i(r) = P_e^i \cdot \left(\frac{r_0}{r}\right)^{\gamma} \cdot \left(\frac{\lambda}{4\pi r_0}\right)^2. \tag{1}$$

Interference I on cell i can be written as the sum over all the cells c inside C (excluding cell i) of the received power at distance d_{ic} (given by figure 1).

$$I = \sum_{c \in C \setminus \{i\}} P_e^c \cdot \left(\frac{r_0}{d_{ic}}\right)^{\gamma} \cdot \left(\frac{\lambda}{4\pi r_0}\right)^2.$$

The noise N_s is modelled as a random variable following a normal distribution of parameter $k.T.W_s$ where $k.T = 4.0039 \times 10^{-20} W/Hz$

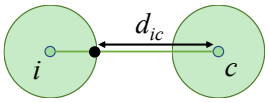


Fig. 1: Representation of distance d_{ic}.

B. Flexible numerology and network slicing

One main improvement of 5G mobile networks is network flexibility that leads to network slicing.

1) Bandwidth parts: In 5G New Radio (NR), the numerology (i.e. the radio frame structure) can be tuned as standardized in [4], meaning that subcarrier spacing and slot time duration are flexible. In 4G LTE, the subcarrier spacing is fixed to 15 kHz and slot time duration (14 OFDM symbols) is 1 ms. Flexible numerology gives the operator the ability to change the Subcarrier Spacing (SCS) that can be 15 kHz, 30 kHz, 60 kHz in FR1 bands (bands above 7.125 GHz, used in the first 5G deployments) and 120 kHz or 240 kHz in FR2 bands. Each time the SCS is multiplied by two, the number of slots in a subframe is multiplied by two. More information about 5G radio frame structure can be found in [5].

For FR1 frequency bands, a 15 kHz SCS might be chosen by default as frequency selectivity is narrowed. One principal requirement that could lead to the use of high SCS is latency. Indeed, as 5G NR should be mainly deployed in Time Division Duplexing (TDD) mode, and as 5G NR TDD schedules downlink and uplink resources at the slot level (inside a slot, there can be downlink and uplink OFDM symbols), the shorter the slot is, the finer the DL/UL switching is. Using high SCS (and thus a small slot duration) can then fit with short delay requirements.

As different numerologies cannot be mixed on a single bandwidth, the operator will have to split the bandwidth in different parts (which correspond to the frequency dimension of a 5G resource grid) and each part will operate with a single SCS. Inside those parts, Bandwidth Parts (BWP) can be defined and allocated to each UE or group of UE.

One BWP is given by two parameters. The first one is the BWP SCS, and the second one in the number of Resource Blocks (RB) that compose this BWP. In 5G NR, one resource block is a frequency indication only, and is defined as being a set of 12 contiguous subcarriers. So the BWP width W_{BWP} is equal to $k \cdot 12 \cdot \omega$ where k and ω are respectively the number of RBs and the SCS. Figure 2 represents 5G NR flexible numerology and BWPs.

5G NR flexible numerology enables an operator to split its bandwidth in different BWPs. All the BWPs have a different set of parameters that are BWP width (number of RBs) and BWP SCS. BWP width is related to the capacity, as described in Section II-A and BWP SCS is related to the communication latency. Those two parameters can be tuned by the operator depending on the requirements of the UEs that will connect to the network through the allocated BWP. It makes sense to consider that operators will allocate a single BWP to each slice.

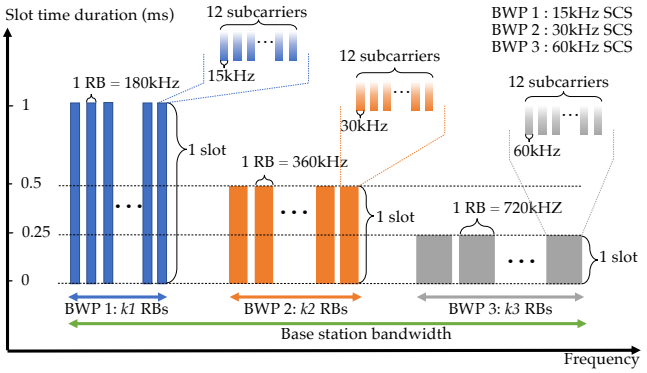


Fig. 2: 5G NR flexible numerology and BWPs.

2) Network slicing: Consequently, a slice is allocated a BWP depending on its capacity and latency requirements. We define slices as being a set of parameters that represent their capacity and latency requirements.

For capacity requirements, we define three parameters:

- UE density: This parameter represents the density of UEs that might connect to a base station. This parameter is the number of UEs per surface unit.
- Activity : This represents the time proportion a UE might be in connected mode.
- Per UE throughput: This represents the capacity required by a single UE while in connected mode.

To take into account latency requirements, we define one parameter:

- Latency importance: This parameter represents the latency requirement level, and is represented by 0,1 or 2. This is used to compute BWP SCS which will be equal to 15, 30 or 60 kHz respectively.

Finally, as some network slices might require ultra reliable communications, we define a parameter to take this aspect into account:

- Connectivity: This represents the number of antennas to which each UE should be connected simultaneously. As handover procedure can take a lot of time compared to the latency requirements, network reliability can be improved by increasing the number of antennas to which each UE is connected to, avoiding handovers communication overhead. This parameter is a key reason for using simplicial homology, which enables us to compute coverage quality including multiple network connectivity.

In the evaluation framework described in Section V, we define three typical 5G network slices:

- Slice 1 (eMBB): Legacy mobile communications for making calls, surfing on the web and video streaming.
- Slice 2 (URLLC): Industry 4.0 for connecting critical actuators and sensors.
- Slice 3 (mMTC): Smart cities that collect data from ultra dense sensors.

III. Simplicial Homology

The need of a method for analysing network coverage and computing network connectivity led us to use simplicial homology, which is a mathematical tool from algebraic topology. We model a cellular network as being a set of cells that are represented by a vertex (the base station) and a radius r. We consider that the intersection between two vertices is not empty if the coverage zone of the corresponding cells intersect each other as depicted in Figure 3.

The first step for studying a cellular network is to get its simplicial complex representation. A simplicial complex is a combination of vertices (i.e. a combination of base stations). The number of elements in the combination gives the dimension of the simplicial complex. A k dimension simplicial complex is called a k-simplex. Thus, a vertex is a 0-simplex, an edge is a 1-simplex, a triangle a 2-simplex, a tetrahedron a 3-simplex etc.

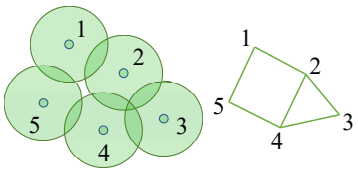


Fig. 3: Cell deployment example and corresponding two-by-two cells intersections. Points represent base stations and circles represent cells coverage.

There are two main kinds of simplicial complexes for network topology representation: the Cěch complex and the Rips complex. The real representation of a network is given by the Cěch complex but its computation complexity is high compared to the Rips complex. It has been shown by [6] that between 0% and 11% of the time, the Rips complex captures k-simplexes that do not exist compared to the Cěch complex. Regarding the error rate and the complexity gain, the Rips complex provides a good engineering approximation.

In the example shown in Figure 3, using Rips complexes, the 0-simplexes are the vertices $\{1, 2, 3, 4, 5\}$, 1-simplexes are $\{[1, 2], [1, 5], [2, 3], [2, 4], [3, 4], [4, 5]\}$ and there is one 2-simplex: $\{[2, 3, 4]\}$.

The method described in [7] is used to derive the Betti numbers that represent network coverage holes. Betti numbers (noted β_0, β_1,...,β_k) are the dimension of each homology group and they can be used to compute the number of k-dimension holes in the network. Thus, β_0 represents the number of related components in the network, β_1 represents the number of coverage holes, β_2 represents the number of zones where there is no 2-connectivity etc. Betti numbers are computed from simplicial complexes using the rank theorem.

In the example shown in Figure 3, β_0 and β_1 are equal to one as there is one related component and one coverage hole.

IV. Energy saving algorithm

A. Sub-optimal heuristic

In this Section, we provide an algorithm for optimizing cells emission power while respecting slices requirements and keeping a good coverage quality. As the problem is not convex (a power increase might imply new interference and thus decrease coverage quality), we use simulated annealing method which is a probabilistic metaheuristic that gives a computationally tractable optimal solution. It has already shown good results in different similar projects such as [3] and [8].

Algorithm 1 gives an overview of the simulated annealing algorithm. Cells are initialized with an emission power which is an algorithm input. At each algorithm iteration, a cell is randomly chosen and we try either to reduce its emission power or either to increase it. An emission power decrease can only be accepted if there is no lack of k-connectivity in the ending state. A power increase will be accepted with a probability given by a defined temperature which will be cooled along iterations. The parameters of this algorithm are K (number of temperature steps), L (temperature step length), α (cooling factor), t_0 (initial temperature) and Δp (emission power variation). The statement $if\ H_l > H$ has to be understood as: *if one new Betti number from* 0 *to* D *(*β_0*,...,*β_D*) is greater than initial Betti numbers*. D is the maximum connectivity dimension we check and is equal to the maximum slices connectivity requirement (defined in Section II-B2). C and $P(c)$ are respectively the set of cells and the emission power of cell c.

Algorithm 1 Simulated annealing

```
allocateBWPs()
R = computeRadius(C)
H = computeConnectivity(C, R)
for k ∈ [0, K] do
  t_k = t_0.α^k
  p = e^(−Δp / t_k)
  for l ∈ L do
    c = Rand(C)
    s = Rand({−1, 1})
    if s = −1 then
      P_e^c = P_e^c − Δp
      R = computeRadius(C)
      H_l = computeConnectivity(C, R)
      if H_l > H then
        P_e^c = P_e^c + Δp
      end if
    else if s = 1 then
      v = Rand([0, 1])
      if v < p then
        P_e^c = P_e^c + Δp
      end if
    end if
  end for
end for
```

The different functions used in Algorithm 1 are explained below:

1) BWPs allocation (function allocateBWPs): BWP_s SCS ω_s is allocated depending on slice latency importance. BWP_s

number of RBs (NRB_s) is allocated proportionally to the total available bandwidth W_T with the following equation. c_{us}, t_{ps} and d_s are respectively the per UE capacity, activity and density for slice s.

$$NRB_s = \left\lfloor \frac{1}{12.\omega_s} \cdot \frac{c_{us}.t_{ps}.d_s}{\sum_{x \in S} c_{ux}.t_{px}.d_x} \cdot W_T \right\rfloor .$$

2) Calculating cells radii (function computeRadius): For each cell, we initialize the current radius at r_0 which is the minimal cell radius. Then we compute the maximum radius of the cell using a minimum reception threshold and equation (1) and which corresponds to the maximum distance from the base station for which a UE could decode radio signal. Beyond this distance, no UE could connect the base station so this distance is the maximum cell radius. The circle defined by the base station and the maximum cell radius defines the maximum coverage zone. We create sub-circles inside the maximum coverage zone and study the cell capacity for each sub-circle. The number of sub-circles to be tested inside the maximum coverage zone is an input parameter. As we observed that the cell capacity decreases exponentially over distance, we create a geometric series to increase exponentially the sub-circles radius so that capacity over sub-circles is expected to decrease linearly. For each sub-circle, we compute the received power depending on the cell emission power following Equation (1). Interference are computed looping over all other cells and computing the received power from this cell (for detailed explanations, see Section II-A). Next step is to loop over each slice and for each slice we compute the required capacity (given by slices parameters, see Section II-B2) and the offered capacity (described in Section II-A) using W_{BWP_s} and undergone SINR. If, for each slice, the offered capacity is greater than the required capacity, the sub-circle is able to serve the different slices, so current radius is set to the current sub-circle radius and the algorithm continues. Otherwise, the sub-circle cannot serve different slices and the algorithm is stopped. Cell radius is the last validated sub-circle radius.

3) Computing connectivity: Function *computeConnectivity* computes simplicial complexes based on the topology and return the derived Betti numbers as explained in Section III.

V. Simulations and results

We have investigated four different scenarios. The first three scenarios are based on single slice simulations. eMBB and mMTC simulations have been done for a connectivity requirement equal to one and uRLLC simulations for a connectivity requirement equal to three. The last scenario integrates all the slices defined previously with a connectivity requirement equal to three. Each slice will use a set of antennas randomly deployed inside a finite square area according to a Poisson Point Process (PPP). Table I gives the deployment parameters of the cells used in the different scenarios. Table II describes the radio parameters of cells associated with different slices. Ten thousand simulations are performed for each scenario to compare power budget, spatial density and transmission energy efficiency. Power budget is defined as the total power consumed by cells of a given deployment. The code and parameters used for these simulations are available online (https://github.com/adejavel/5GSliceAwareEnergySaving).

TABLE I: Deployment parameters

Slice	Vertical	Square width (m)	Cells density (cells/km^2)
Slice 1	eMBB	500	50
Slice 2	URLLC	300	500
Slice 3	mMTC	500	30
All slices	All	300	500

TABLE II: Simulations parameters

Slice	Initial power	Δp	Band	Bandwidth	γ
Slice 1	40 W	0.5 W	3.5GHz	400 MHz	3
Slice 2	0.5 W	5 mW	26GHz	500 MHz	6
Slice 3	1 W	10 mW	700 MHz	10 MHz	3
All slices	10 W	0.1 W	3.5 GHz	400 MHz	4

Table III gives the different slices parameters used in the four simulation scenarios. D is the UE density, A is the UE activity factor, T is the required per UE throughput, LI is the slice latency importance, C is the connectivity requirement and ST is the total throughput per squared kilometer. Slice 1 capacity requirements are low compared to what have been announced with 5G as massive MIMO and beam-forming are not taken into account. For the single slice scenarios, simulations have been performed in their specific frequency bands. In the last scenario, the three slices share the same frequency band, in a context where bandwidth is limited.

The different results obtained during the simulations are shown in Table IV. Mean power budget is expressed per squared kilometers and spectral efficiency represents the power budget per transmitted bit. Figures 4a, 4b, 4c and 4d are histograms representing the overall distribution of antennas emission power (expressed in Watts) for all the simulation cases. It can be observed that all the histograms have the same shape: an initial peak followed by a second one around the median.

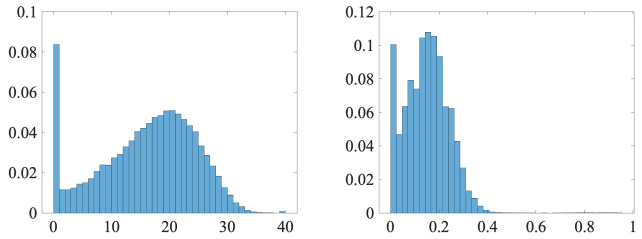
The initial peak is due to the clustering inherent to PPP. This process often creates clusters of close cells. Inside these

TABLE III: Slices parameters

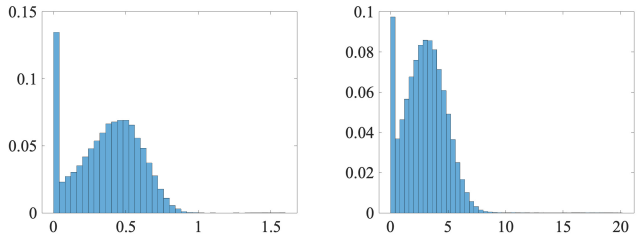
Slice	D (UE/km^2)	A	T (b/s)	LI	C	ST (b)
Slice 1	10^2	0.1	$100 \cdot 10^6$	0	1	10^9
Slice 2	$2 \cdot 10^4$	0.05	10^3	2	3	10^6
Slice 3	10^5	0.05	10^3	0	1	$5 \cdot 10^6$

TABLE IV: Simulations results

Slice	Power budget (W/km^2)	Energy efficiency (W/b)
Slice 1	777	$7.77 \cdot 10^{-7}$
Slice 2	70	$7 \cdot 10^{-5}$
Slice 3	10.8	$2.16 \cdot 10^{-6}$
All slices	1425	$1.41 \cdot 10^{-6}$



(a) Slice 1: eMBB scenario. (b) Slice 2: URLLC scenario.



(c) Slice 3: mMTC scenario. (d) All slices scenario.

Fig. 4: Simulations results (X axis is the emission power in Watts and Y axis is the proportion of cells).

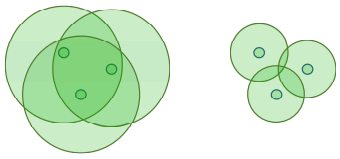


Fig. 5: PPP cells cluster creation illustration.

clusters, base stations can reduce their emission power to the minimum required power, without having any impact on the number of coverage holes and connectivity. This is illustrated in Figure 5 where Betti numbers are equals in both the left and right deployments for different emission power.

Cells density are of the same order of magnitude for both eMBB and mMTC. eMBB slice capacity requirements are much bigger than those of mMTC. Foremost, mMTC propagation conditions are significantly better that those available in frequency bands allocated to eMBB. These two reasons explain why mMTC power budget is much smaller than eMBB. One can deduce that slice capacity and cells density have an important impact on power budget but a marginal one on transmission energy efficiency. At the opposite, for uRLLC slice, which requires high reliability, it can be observed that power budget is between mMTC and eMBB whereas transmission energy efficiency is much higher than the two other slices. It can be deduced that, while deploying slices independently, reliability constraints imply a loss of energy efficiency. On the other side, the last scenario simulation results show that running the three slices on a same physical network imply a huge increase of the power budget while keeping a reasonable energy efficiency. The huge power budget can be explained by cells density, which is high, that implies a lot of interference. On the other side, high cells density also improves network capacity and explains the gain in energy efficiency. Thus, on a power budget point of view, it is worse to have all the slices running on the same physical infrastructure. In the same time, this technique brings a gain in energy efficiency compared to uRLLC slice (which has the highest reliability constraint).

VI. Conclusion

In this paper, a solution for determining per-slice power budget has been proposed. It takes into account slices capacity, latency and reliability constraints. Simplicial homology was introduced as a way to model reliability requirement. Simulated annealing method was used to reduce RAN power budget. Extensive simulations has been realized for three different types of slices : eMBB, uRLLC and mMTC. Simulations have shown that reliability constraints imply a huge loss of energy efficiency and that deploying slices independently preserves power budget but can decrease energy efficiency for applications requiring high reliability.

As future research, we plan to further optimize simulated annealing parameters, introduce new power reduction algorithms and methods to include radio technologies such as massive MIMO and beam-forming in our network optimization. We also plan to investigate slicing impact on RAN networks using millimeter waves. Finally, latency constraints will be explicitly integrated in the optimization algorithm.

VII. Acknowledgements

The work presented in this article benefited from the support of NewNet@Paris, Cisco's Chair "Networks for the Future" at Télécom Paris (https://newnet.telecom-paristech.fr). Any opinions, findings or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of partners of the Chair.

References

[1] D. G. Cattrysse and L. N. Van Wassenhove, "A survey of algorithms for the generalized assignment problem," Tech. Rep., 1990.

[2] M. Leconte, G. S. Paschos, P. Mertikopoulos, and U. C. Kozat, "A resource allocation framework for network slicing," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2018, pp. 2177–2185.

[3] N. Le, P. Martins, L. Decreusefond, and A. Vergne, "Simplicial homology based energy saving algorithms for wireless networks," in *IEEE International Conference on Communication Workshop*, 2015, pp. 166–172.

[4] 3GPP, "Technical Specification Group Radio Access Network; NR; Physical channels and modulation," 3GPP, TS 38.211, 04 2020. [Online]. Available: http://www.3gpp.org/dynareport/38211.htm

[5] M. Kottkamp, R. und Schwarz, A. Pandey, D. Raddino, A. Roessler, and R. Stuhlfauth, *5G New Radio: Fundamentals, Procedures, Testing Aspects*. Rohde & Schwarz GmbH & Company KG, 2019.

[6] F. Yan, P. Martins, and L. Decreusefond, "Accuracy of homology based coverage hole detection for wireless sensor networks on sphere," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3583–3595, 2014.

[7] A. Vergne, L. Decreusefond, and P. Martins, "Simplicial homology for future cellular networks," *IEEE Transactions on Mobile Computing*, vol. 14, no. 8, pp. 1712–1725, 2015.

[8] M. Lanza, A. L. Gutiérrez, J. R. Pérez, J. Morgade, M. Domingo, L. Valle, P. Angueira, and J. Basterrechea, "Coverage optimization and power reduction in sfn using simulated annealing," *IEEE Transactions on Broadcasting*, vol. 60, no. 3, pp. 474–485, 2014.