This is a postprint version of the following published document:


DOI: 10.1109/mcom.2018.1700960

© 2018, IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Design and Analysis of 5G Scenarios with simmer: An R Package for Fast DES Prototyping

Iñaki Ucar, José Alberto Hernández, Pablo Serrano, Arturo Azcorra

Abstract—Simulation frameworks are important tools for the analysis and design of communication networks and protocols, but they can result extremely costly and/or complex (for the case of very specialized tools), or too naive and lacking proper features and support (for the case of ad-hoc tools). In this paper, we present an analysis of three 5G scenarios using simmer, a recent R package for discrete-event simulation that sits between the above two paradigms. As our results show, it provides a simple yet very powerful syntax, supporting the efficient simulation of relatively complex scenarios at a low implementation cost.

Index Terms—5G, simmer, network simulation, QoS, small cells, C-RAN, PON, IoT.

I. INTRODUCTION

Simulation frameworks are undoubtedly one of the most important tools for the analysis and design of communication networks and protocols. Their applications are numerous, including the performance evaluation of existing or novel proposals, dimensioning of resources and capacity planning, or the validation of theoretical analyses, which are based on simplifying assumptions whose impact is to be assessed.

In fact, simulation frameworks also make a number of simplifying assumptions, typically about components of the considered system that are not directly related to the performance variable of interest, to reduce complexity so the development of the scenario is easier and numerical figures are obtained faster. This “complexity” axis goes from very specialized, large simulation tools such as NS-3, OMNeT++, OPNET, and ad hoc simulation tools, consisting on hundreds of lines of code, typically used to validate a very specific part of the network or a given mathematical analysis. The latter are often developed over general-purpose languages such as C/C++ or Python, over numerical frameworks such as Matlab, or over some framework for discrete-event simulation.

On the one hand, the complexity of specialized tools (as their cost, if applicable) preclude their use for short-to-medium research projects, as the learning curve is typically steep plus they are difficult to extend, which is mandatory to test a novel functionality. On the other hand, the development of ad-hoc tools also require some investment of time and resources, lack a proper validation of their functionality and, furthermore, there is no code maintenance once the project is finished, for the few cases in which the code is made publicly available.

This article provides a quick overview of simmer and its key features. The analyses of the three considered 5G-
related scenarios showcase the multiple benefits of different approaches and the versatility of simmer to easily implement their key features.

II. AN INTRODUCTION TO SIMMER

Simmer\(^1\) [1] is a discrete event simulation (DES) library for the R Project\(^2\), the open source programming language for statistical computing that has been receiving increased attention, primarily due to its widespread adoption for data science, analytics and statistical research.

By developing simmer for R, it can benefit from this growing ecosystem. Note that simmer does not aim at substituting NS-3 or OMNeT++, which are the de facto standards for open-source network simulations. Instead, simmer is designed as a general-purpose DES framework with a human-friendly syntax and a very gentle learning curve. It can be used to complement other field-specific simulators as a rapid prototyping tool that enable insightful analyses of different designs. As we will illustrate in the next section, with simmer it is simple to simulate relatively complex scenarios, with the added benefit of the availability of many convenient data analysis and representation libraries, thanks to the use of R.

The R application programming interface (API) exposed by simmer revolves around the concept of trajectory, which defines the "path" in the simulation for entities of the same type. A trajectory is a recipe for the arrivals attached to it, an ordered set of actions (or verbs) chained together with the pipe operator (\%\%\%), whose behaviour is similar to the \%\%\% pipe operator in Julia for the Python and Julia languages respectively [1]. The key for this speed is its underlying simulation core, which is written in C++. Furthermore, and perhaps more importantly, simmer implements automatic monitoring capabilities: every event is accounted for by default, both for arrivals (starting and ending times, activity time, ending condition, resources traversed) and resources (server and queue status), and all this information can be easily retrieved in standard R data frames for further processing of results (lines 13-14 of the clerk example).

III. MODELING 5G SCENARIOS

In what follows, we will model and analyze three representative 5G scenarios using simmer. The source code for the use cases presented here, including configuration (definition of constants and parameters), simulation and analysis of results, is available online,\(^3\) while some summary statistics of the simulations performed and their complexity are provided below.

A. Crosshauling of FH and BH Traffic

This scenario is motivated by the Cloud Radio Access Network (C-RAN) paradigm [3], where the mobile base station functionality is split into simple Remote Radio Heads (RRH), spread across the deployment and connected by fiber to centralized (and possibly virtualized) Base Band Units (BBU), at the operators’ premises. C-RAN is an architectural shift, aiming at providing CAPEX and OPEX savings while supporting better interference reduction and improved performance via Coordinated Multi-Point (CoMP).

In this C-RAN paradigm, fronthaul (FH) traffic from the RRH has stringent delay requirements, while backhaul (BH) traffic from the BBU has mild delay requirements. In a general topology, such as the one illustrated in Fig. 1a ("scenario no. 1") [4], XPFEs (Crosshaul Packet Forwarding Element) will forward both types of traffic, and therefore introducing service differentiation might improve the ability to fulfill the delivery guarantees of FH traffic, which is in-line with the IEEE 802.1CM Time-Sensitive Networking for Front haul standard under development.\(^6\) To quantify the benefits of this differentiation, we use simmer to simulate the scenario, this way supporting for example a design decision about the best scheme to deploy.

---

\(^1\)Available for download on the Comprehensive R Archive Network (CRAN), at https://CRAN.R-project.org/package=simmer

\(^2\)https://www.R-project.org/

\(^3\)See the “Articles” section at http://r-simmer.org.

\(^4\)See http://www.ieee802.org/1/pages/802.1cm.html, Draft 0.6.

---

1: library(simmer)
2: cust <- trajectory("customer") <- %>
3: seize("clerk", amount=1) <- %>
4: timeout(function() rexp(1, 1)) <- %>
5: release("clerk", amount=1) <- %>
6: env <- simmer("rack") <- %>
7: add_resource("clerk", capacity=1, queue.size=Inf) <- %>
8: add_generator("cust", cust, function() rexp(1, 1)) <- %>
9: run(until=1000)
10: arrivals <- get_non.arrivals(env)
11: resource <- get_non.resources(env)
12: Given that both the time at the clerk and the time between customers are exponential random variables and infinite queue length, this example corresponds, in Kendall’s notation, to an M/M/1 queue. It serves to illustrate the two main elements of simmer: the trajectory object and the simmer environment (or simulation environment).

The customer trajectory (line 3) defines the behaviour of a generic customer: seize a clerk, spend some time, and release it. The env simulation environment (line 8) is then defined as one clerk with infinite queue size and a generator of customers, each one following the trajectory defined above. Based on this syntax, the flexibility is provided through a rich set of activities (more than 30) that can be appended to trajectories, which support: changing arrivals’ properties (attributes, priority, batches), different interactions with the resources (select, seize, release, change their properties), and the generators (activate, deactivate, change their properties), and even the definition of branches (simple, depending on a condition, or parallel) and loops. Finally, some support to asynchronous programming is also provided (subscription to signals and registration of handlers). Not only simmer provides a powerful yet simple syntax, but it is also fast, for example, faster than equivalent frameworks such as SimPy and SimJulia for the Python and Julia languages respectively [1].
We assume that XPFEs are at 75% load and operate at 40 Gb/s line rate. We assume 50% of this load corresponds to FH traffic, and the other 50% corresponds to BH traffic. Packet arrivals follow a Poisson process (although any other arrival process may be used in the simulation, for instance, bursty, self-similar); regarding packet sizes, FH packets are assumed to be CPRI (Common Public Radio Interface) Basic Frames of 80 byte length (i.e. CPRI option 4, see [5]), while BH packets follow the classical AMS-IX (Amsterdam Internet Exchange) trimodal function, namely 7 out of 12 packets are short (40 bytes), 4 out of 12 are medium size (576 bytes) and 1 out of 12 packets is long (1500 bytes).²

We consider three different policies for service differentiation at the N XPFEs:

1) No service differentiation.
2) Strict Priority (SP) is given to FH over BH, but without preemption.
3) SP is given to FH over BH with preemption.

For each of these policies, we first simulate a one-XPFE scenario with the traffic characteristics described above, and compute the queueing delay for FH and BH traffic. The results are depicted in Fig. 2a, where we use box-plots to illustrate the 5, 25, 50, 75, 95-th percentiles.

As the figure shows, the first strategy (‘without SP’) results in both traffic experiencing the same queueing delay. The second strategy (‘with SP’), that is, service differentiation without preemption, results in a much improved service for FH traffic, as FH packets only have to wait for other FH packets ahead and, sometimes, one BH packet being served (i.e., the residual service time). The third strategy (‘with SP & preemption’) implements a preemption strategy, discarding BH traffic from the transmission line if a FH packet arrives which, as the figure shows, does significantly improve delay performance (the 95-th percentile drastically decreases), which was caused by the long transmission times of long BH frames (1500 bytes). For this scenario, we conclude that a preemption strategy reduces queueing delay to the minimum, with very high delivery guarantees.

We next analyze scenarios where the FH has to transverse N XPFEs in tandem, each one also serving BH traffic, as illustrated in Fig. 1a. We depict in Fig. 2b the queueing delays of FH traffic for N = 1, 2, 5, under the three considered policies (the N = 1 case corresponds to the same results as in the previous experiments). As shown, FH queueing delay accumulates after traversing multiple XPFEs for the first and second policies, both in terms of median and percentiles, hence jitter too. Only the SP strategy with preemption keeps both delay and jitter values extremely low, since FH packets have only to wait if the server is occupied by other FH packets. In contrast, the SP strategy without preemption shows accumulated packet delay and packet delay variability after traversing multiple XPFEs.

Implementation Details: The simulation implements a single trajectory for the FH traffic, fh_traffic, that seizes the N XPFEs sequentially. Additionally, a list of N trajectories called bh_traffic, one for each XPFE, models the interfering BH traffic. The XPFEs are defined as resources with capacity=1 and infinite queue length. A generator of FH traffic plus N generators of BH traffic are attached to their respective trajectories. All the elements are encapsulated into a function and a number of constants are parametrized, namely, the number of XPFEs, FH traffic’s priority and whether the XPFEs should be preemptive or not. As can be seen, all

the cases, defined as a data frame (one case per row), are easily parallelized using standard R tools (i.e., the R-core parallel package). All the monitoring information automatically collected by simmer for the arrivals (BH and FH packets) can be retrieved as a data frame using the `get_mon_arrivals()` method. This enables further analysis and visualization in only a few lines of code using the `dplyr` and `ggplot2` packages.

B. Mobile Traffic Backhauling with FTTx

We next consider the case of a residential area with a Fiber-To-The-Premises (FTTx) infrastructure, that is, an Optical Distribution Network (ODN), composed of the Optical Line Terminal (OLT), splitters, and the Optical Network Unit (ONU) at the users’ premises. As Fig. 1b illustrates (“scenario no. 2”), we assume that an operator is planning to deploy an antenna, carrying the mobile traffic over the ODN, and is considering two implementation options:

1) Deployment of a Small Cell, reducing the amount and requirements of the generated traffic. In both cases, we analyze the upstream channel of a Time-Division Multiplexed Passive Optical Network (TDM-PON) providing broadband access to the residential users and the mobile users. We assume for simplicity the case of ITU-T G.984 Gigabit PON (it would be straightforward to extend the results to other TDM-based PONs, such as XG-PON, XGS-PON, EPON, 10G-EPON), with a total capacity of 1.25 Gb/s in the upstream, and where each ONU generates 20 Mb/s of upstream traffic.

1) Small Cell: Here we assume that the small cell generates 150 Mb/s peak traffic, and shares the ODN with 31 residential users, which corresponds to an average total load of 61.6%.

We assume bursty arrivals following a compound Poisson model, where both the bursts and the burst length are Poisson-distributed, with a mean of 20 packets. As in the previous use case, packet sizes are randomly chosen following the AMS-IX trimodal distribution, yielding an average burst size of 6407 Bytes. In this case, we assume a Dynamic Bandwidth Allocation (DBA) algorithm similar to the IPACT protocol [6], where each ONU requests at the end of its transmission window resources for the next cycle, while the OLT receives such requests and decides when and for how long each ONU may transmit its data in the next cycle granting transmission windows in a round-robin fashion. We assume a 1 s guard-time between transmission windows of consecutive ONUs.

2) RRH Backhauling: Next, we assume an RRH generating FH traffic, following the MAC-PHY (Media Access Control, physical layer) functional split, that is, one OFDM (Orthogonal Frequency Division Multiplexing) symbol is sampled, quantized (approx. 6000 B) and transmitted to the BBU every 66.67 µs, this resulting in an approximate rate of 720 Mb/s. This FH traffic shares the ODN with now 7 residential ONUs, this resulting in an approx. total load of 68.8%. To fulfil the tight delivery requirements of FH traffic, we assume that the DBA algorithm guarantees periodic TDM reservations for the FH traffic, and then the ONUs use a similar algorithm as before to share the rest of the available bandwidth.

We depict in Fig. 3 (left) the results corresponding to the uplink delay in Case 1, that is, one Small Cell and 31 ONUs, for both types of traffic and the four QoS policies. Like in the previous case, we focus on the queueing delay, and represent the different percentiles with box-and-whisker plots.

---

The figure shows queueing delay experienced by BH and FH traffic under different QoS policies: a) BH/FH queueing delay comparison for a single XPFE; b) accumulated FH delay for several XPFEs. The whiskers represent the 5th and 95th percentiles.

We assume bursty arrivals following a compound Poisson model, where both the bursts and the burst length are Poisson-distributed, with a mean of 20 packets. As in the previous use case, packet sizes are randomly chosen following the AMS-IX trimodal distribution, yielding an average burst size of 6407 Bytes. In this case, we assume a Dynamic Bandwidth Allocation (DBA) algorithm similar to the IPACT protocol [6], where each ONU requests at the end of its transmission window resources for the next cycle, while the OLT receives such requests and decides when and for how long each ONU may transmit its data in the next cycle granting transmission windows in a round-robin fashion. We assume a 1 s guard-time between transmission windows of consecutive ONUs.

We consider four different quality-of-service (QoS) policies for the DBA, these being defined by the maximum supported request per cycle per ONU (the cellular traffic is always granted their requests in each case). If an ONU requests more that this maximum transmission window (namely, 1500 B, 3000 B, 6000 B or infinite, which means no limit), the OLT will only grant this maximum, and therefore the rest of the traffic will queue at the ONU.

2) RRH Backhauling: Next, we assume an RRH generating FH traffic, following the MAC-PHY (Media Access Control, physical layer) functional split, that is, one OFDM (Orthogonal Frequency Division Multiplexing) symbol is sampled, quantized (approx. 6000 B) and transmitted to the BBU every 66.67 µs, this resulting in an approximate rate of 720 Mb/s. This FH traffic shares the ODN with now 7 residential ONUs, this resulting in an approx. total load of 68.8%. To fulfil the tight delivery requirements of FH traffic, we assume that the DBA algorithm guarantees periodic TDM reservations for the FH traffic, and then the ONUs use a similar algorithm as before to share the rest of the available bandwidth.

We depict in Fig. 3 (left) the results corresponding to the uplink delay in Case 1, that is, one Small Cell and 31 ONUs, for both types of traffic and the four QoS policies. Like in the previous case, we focus on the queueing delay, and represent the different percentiles with box-and-whisker plots.

---

8https://CRAN.R-project.org/package=dplyr
9https://CRAN.R-project.org/package=ggplot2
As the figure shows, when requests are not limited ('Inf') the traffic from the ONUs and the Small Cell experience the same delay, as expected. However, the moment the DBA algorithm enforces a limit to ONUs requests, differences appear: the Small Cell delays are smaller and less dispersive, while the ONU results are longer and show more variability. It can be seen that with a 1500 B limit, the delivery guarantees for the Small Cell traffic are very tight (e.g., 95-th percentile around 250 µs), which is achieved at the expense of significant delays for the ONUs.

Next, we depict in Fig. 3 (right) the results for the uplink delay in Case 2, that is, a RRH and 7 ONUs. We do not depict here the delay results corresponding to the FH traffic, as in this case the reservation mechanism guarantees its delivery with zero queuing delay. As the figure show, the delay performance obtained by users is slightly better than in the previous case, although less users can be accommodated due to traffic demands.

**Implementation Details:** This scenario serves to illustrate a different strategy to code a simulation with **AirSim**. Instead of a trajectory attached to an unlimited generator of arrivals, the OLT is defined here as a trajectory with a single worker in an infinite loop by using the `rollback()` activity. This OLT executes the DBA logic encapsulated into the `set_next_window()` function, and timely activates the ONUs, defined as resources, in a round-robin fashion. This is achieved by modifying the capacity of a given ONU resource to meet the number of packets allocated for the next transmission window.

As in the previous use case, a number of constants are parameterized (scenario, number of ONUs and TDM reservation limit) and all the combinations are parallelized with the R-core parallel package. Finally, the monitoring statistics for the arrivals are retrieved, and a similar analysis is made using **dplyr** and **ggplot** in very few lines of code.

### C. Energy Efficiency for Massive IoT

Finally, we consider the case of a massive Internet-of-Things (mIoT) scenario, a use case for Long Term Evolution (LTE) and next-generation 5G networks, as defined by the Third Generation Partnership Project (3GPP) in [7]. As Fig. 1c (top) illustrates, we consider a single LTE macrocell in a dense urban area. The buildings in the cell area are populated with N smart meters (for electricity, gas and water), and each meter operates independently as a Narrowband IoT (NB-IoT) device.

The devices’ behaviour is modelled following the diagram depicted in Fig. 1c (bottom), which is a simplified version of the Markov chain model developed in [8] (Fig. 5). A device may stay in **RC Idle** (‘Off’) and awakes with some periodicity to upload its reading. This communication phase encompasses a contention-based random access (RA) procedure, with a backoff time randomly chosen between (0, W) time slots and up to m retransmissions. If the connection request fails, the reading is dropped and the device returns to the ‘Off’ state. If the connection is successful, we assume that the device implements the Control Plane Cellular IoT (CP) optimization [8], so that the data is transmitted over the RRC Connection request phase using the Non Access Stratum (NAS) level. Then, the device has to wait (‘Inactive’) until the connection is released, and eventually returns to the ‘Off’ state.

The goal of this use case is to study the effect of synchronization across IoT devices (for instance, due to a power outage) in the energy consumption. As in [9], we assume that a device provides its readings as often as every hour, and the cases of N = 5,10, 30-10^6 devices in one cell are considered.

In order to study different levels of synchronization, each node implements an additional backoff window prior to the RA procedure. Furthermore, we selected m = 9 and W = 20; the rest of the parameters (power consumption, timings, message sizes...) can be found in [8] (Table 1).
Fig. 4 shows the results of the simulation for one day. It depicts the energy consumed per reading considering a uniform backoff window between 0 and 5 (highly synchronized), 10, 30 and 60 seconds (non-synchronized). As the number of devices and the level of synchronization grow, the random-access opportunities (RAOs) per second grow as well producing more and more collisions. These collisions cause retries and a noticeable impact in the energy consumption (up to 12% more energy per reading). Therefore, this use case shows the paramount importance of randomizing node activation in mIoT scenarios in order to avoid RAO peaks and a premature battery drain.

**Implementation Details:** This scenario requires a single meter trajectory implementing the logic of each IoT device in an infinite loop, and N workers are attached to it at \( t = 0 \). Each device registers itself for a given signal (“reading”) and waits in sleep mode until a new reading is requested, which is triggered by a secondary trajectory (trigger). As soon as a new reading is signalled, the RA procedure starts by randomly selecting one of the 54 preambles available, which are defined as resources. The process of seizing a preamble encompasses two sub-trajectories:

- If there are no collisions, the preamble is successfully seized and the post seize sub-trajectory is executed, which transmits a reading.
- If there is collision, rejection occurs, and the reject sub-trajectory is executed, which performs the RA back-off (for a random number of slots) and restarts the RA procedure (for a maximum of \( m \) retries).

Both sub-trajectories set the appropriate power levels \( P \) for the appropriate amount of time. In this case, these power levels throughout the simulation time are retrieved with the `get_non_attributes()` method. Again, the energy is concisely computed and represented using `dplyc` and `ggplot2` packages.

**D. Wrap up**

Thanks to these scenarios, we have demonstrated the usability and suitability of `simmer` for fast prototyping of three different 5G scenarios. The code developed highlights some of the characteristics that make `simmer` attractive for researchers and practitioners in communications research:

- A novel and intuitive trajectory-based approach that simplifies the simulation of large networks of queues, including those with feedback.
- Flexible resources, with dynamic capacity and queue size, priority queueing and preemption.
- Flexible generators of arrivals that can draw interarrival times from any theoretical or empirical distribution via a function call.
- Asynchronous programming features and monitoring capabilities, which helps the researcher focus into the model design.

Table I summarizes the main simulation statistics for each scenario (simulation time is computed with a machine equipped with an Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz x4 (32 cores) and 64 GB of RAM. Debian GNU/Linux 8, R 3.3.2 and `simmer` 3.6). These numbers attest that `simmer` can be used to simulate relatively complex scenarios with very few lines of code (under 100 lines in all cases). Furthermore, the automatic monitoring capabilities embedded in `simmer`, and the integration with the R language, enable sophisticated analyses and visualizations just with a few more lines. It is likewise remarkable the ease with which multiple scenarios, with different parameters, can be simulated concurrently thanks to base R functions. Thus, exploring a large number of combinations of parameter values is not only straightforward, but also as fast as the slowest thread given enough number of CPU cores available.

**IV. CONCLUSIONS**

In this paper, we have introduced the use of the `simmer` R package for the simulation of communication scenarios. We have illustrated its simple yet powerful syntax, and have demonstrated its ease of use and functionality with the analysis of three 5G-inspired scenarios, corresponding to radio, access and metro deployments. The results obtained, which can be easily computed thanks to the powerful capabilities of R, help taking design decisions related to hardware choices, traffic prioritization or access scheme configuration. Because of these, we believe `simmer` is a powerful tool to validate analytical studies, or to complement the use of more complex and costly simulation frameworks.

**ACKNOWLEDGEMENTS**

This article has been partially supported by the 5G-City project (TEC2016-76795-C6-3-R) and the TEXEO project (TEC2016-80339-R), both funded by the Spanish Ministry of Economy and Competitiveness.

**REFERENCES**


**Biographies**

**Iñaki Ucar** (inaki.ucar@uc3m.es) received his M.Sc.Eng. in Telecommunications Engineering and M.Sc. in Communications from Universidad Pública de Navarra (UPNA) in 2011 and 2013 respectively, and his M.Sc. in Telematics Engineering from Universidad Carlos III de Madrid (UC3M) in 2014. Currently, he holds the position of Teaching Assistant and pursues his Ph.D. in the Department of Telematics Engineering of UC3M. His work focuses on energy efficiency of wireless networks.

**José Alberto Hernández** (jahgutie@it.uc3m.es) completed his five-year degree in Telecommunications Engineering at UC3M in 2002, and his Ph.D. degree in Computer Science at Loughborough University, Leicester, United Kingdom, in 2005. He has been a senior lecturer in the Department of Telematics Engineering since 2010, where he combines teaching and research in the areas of optical WDM networks, multi-generation access networks, metro Ethernet, energy efficiency, and hybrid optical-wireless technologies. He has published more than 75 articles in both journals and conference proceedings on these topics. He is a co-author of the book *Probabilistic Models for Computer Networks: Tools and Solved Problems*.

**Pablo Serrano** (pablo@it.uc3m.es) received his degree in Telecommunications Engineering and his Ph.D. from Universidad Carlos III de Madrid (UC3M) in 2002 and 2006, respectively. He has been with the Telematics Department of UC3M since 2002, where he currently holds the position of Associate Professor. He has over 80 scientific papers in peer-reviewed international journal and conferences. He has served as guest editor for Computer Networks, and on the TPC of a number of conferences and workshops including IEEE INFOCOM, IEEE WoWMoM and IEEE Globecom.

**Arturo Ancorra** (ancorra@it.uc3m.es) received his M.Sc. degree in Telecommunications Engineering from the Universidad Politécnica de Madrid in 1986 and his Ph.D. from the same university in 1989. In 1993, he obtained an M.B.A. from the Instituto de Empresa. He has a double appointment as Full Professor (with chair) at the Telematics Engineering Department of Universidad Carlos III of Madrid and as Director of IMDEA Networks. Prof. Ancorra has coordinated the CONTENT and E-NEXT European Networks of Excellence, has served as a Program Committee Member in numerous international conferences and has published over 100 scientific papers in books, international journals and conferences.