

Efficient global optimization of multi-parameter network problems on wireless testbeds

Michael T. Mehari *, Eli De Poorter, Ivo Couckuyt, Dirk Deschrijver, Jono Vanhie-Van Gerwen, Daan Pareit, Tom Dhaene, Ingrid Moerman

*Ghent University - iMinds, Department of Information Technology (INTEC)
Gaston Crommenlaan 8 (Bus 201), B-9050 Ghent, Belgium
michael.mehari@intec.ugent.be
Tel: +32 (0)93 313800 Mob: +32 484 418362*

Abstract

A large amount of research focuses on experimentally optimizing the performance of wireless solutions. Finding the optimal performance settings typically requires investigating all possible combinations of design parameters, while the number of required experiments increases exponentially for each considered design parameter. The aim of this paper is to analyze the applicability of global optimization techniques to reduce the optimization time of wireless experimentation. In particular, the paper applies the Efficient Global Optimization (EGO) algorithm implemented in the SURrogate MOdeling (SUMO) toolbox inside a wireless testbed. Moreover, to cope with the unpredictable nature of wireless testbeds, the paper applies an experiment outlier detection which monitors outside interference and verifies the validity of conducted experiments. The proposed techniques are implemented and evaluated in a wireless testbed using a realistic wireless conferencing scenario. The performance gain and experimentation time of a SUMO optimized experiment is compared against an exhaustively searched experiment. In our proof of concept, it is shown that the proposed SUMO optimizer reaches 99.79% of the global optimum performance while requiring 8.67 times less experiments compared to the exhaustive search experiment.

Keywords: wireless experimentation, optimization, testbeds, surrogate modeling, outlier detection

1. Introduction

Wireless networks are utilized in many application domains. For example, if a home user is wirelessly connected, he can move around with his laptop or mobile device, while staying connected to his peers. Wireless sensor networks can be used in applications as diverse as early-warning systems for forest fire and home automation. Body area networks attached to a patient for health-monitoring purposes make the patient-doctor interaction more productive. These wireless innovations trigger the wireless research community to continuously introduce and validate novel wireless concepts. Such research problems often have several design parameters that can be changed. For example, Wi-Fi networks have parameters that can be

tweaked at the physical layer (e.g. transmit power, channel, modulation), MAC layer (e.g. inter frame spacing, contention window), network layer (e.g. routing protocol, mobility, topology) and application layer (e.g. throughput, server configurations). Optimizing all or a subset of these parameters (a.k.a. multi-parameter optimization) in order to find the *optimum operating point* is time consuming since the design space grows exponentially for every investigated design parameter.

Often, these wireless networks are optimized using wireless network simulations. These simulators generate a number of interference and traffic patterns, create a propagation model of the wireless medium, execute the optimization algorithms and analyze a set of performance metrics. However, wireless network simulators also have a number of disadvantages. Results can be very different when executing identical experiments on multiple wireless network simulators. In [1], the accuracy of Opnet, ns-2, and GloMoSim simulators indicate significant differences when evaluating a single protocol problem. Another limitation of a wireless network simulator is its incapability to accurately model the underlying wireless transmission properties such as channel characteristics and antenna diversity. It is also very hard to model the hardware's imperfections and dissimilarities between devices of the same type [2], which often have a considerable impact on the overall network performance.

As a result, experimentally driven research is necessary to complement simulations [2]. Measurements and performance evaluations on a real-life testbed are gaining more attention as they account for hardware imperfections and dissimilarities. However, wireless testbeds also have limitations. They require more set-up overhead compared to their simulator counterparts before, during and after experimentation. Typical examples are resource management, turning on radio interfaces, message orchestration and output post processing. For example, when using the Orbit Management Framework (OMF) for experimentation control, an experiment having N wireless devices adds an average delay of $5.17 \cdot N$ ms on a single message orchestration [3]. In addition, experiments on real-life testbeds can not be artificially speed up, which is possible when using simulations. In order to mitigate the time overhead, efficient optimization algorithms can be used that are best fitted to wireless testbeds. Two of their most widely used approaches are selective sampling of the design space and sensitivity analysis on the design parameters. In this paper, we investigate the selective sampling approach of Efficient Global Optimization (EGO) [4] implemented in the SURrogate MOdeling (SUMO) toolbox [5]. EGO uses Kriging approximations to find optimal operation point(s) of a complex problem while minimizing the number of experiments needed. This way, the overall experimentation time is kept to a minimum [6]. In a nutshell, this paper examines the strengths of the SUMO optimizer by applying it to a network problem in a wireless testbed having multiple design parameters.

This paper presents the following novel contributions.

- Integration of the SUMO toolbox in a wireless testbed.
- Definition of a wireless conferencing scenario which involves multiple design parameters and performance objectives.

- A simple mechanism for detecting outliers during Wi-Fi experiments.
- Repeatability analysis of Wi-Fi experiments.
- Sensitivity analysis of global optimization to the choice of the initial sample experiments.
- A generic stopping criteria that can be used in a variety of optimization problems.

The remainder of this paper is organized as follows. Section 2 explores the related work on multi-parameter optimization in wireless networks. The principles of SUMO optimization and modifications to the SUMO toolbox are explained in section 3. In section 4, the SUMO optimizer is experimentally validated by optimizing a wireless conference network problem. The results of the experiment optimization process are presented and analyzed in section 5. Finally Section 6 proposes future work and Section 7 concludes the paper.

2. Related work

Solutions of wireless network problems often involve multi-objective optimizers in order to optimize multiple design parameters. In literature, a wide range of multi-objective optimization algorithms exist. The effectiveness of such algorithms greatly depends on the methodology behind their implementation as measured by time, processing power, memory and performance. During the optimization process, optimizers carefully investigate two aspects. These are exploration and exploitation [7]. Exploration refers to the phase in which an optimizer understands the dynamics of a problem by selecting as few random sample points as possible. These random sample points have to be selected carefully in order not to waste valuable experimentation time. On the other hand, the exploitation phase locates local optimums starting from the explored design space. If the problem has been explored very well, the exploitation phase guarantees to locate global optimums. Therefore, the question of predicting global optimums in a short period of time creates the *exploration vs exploitation trade off* [8] which all multi-objective optimizers target.

Exhaustive search approaches evaluate all operating points of a solution to select optimum settings from the design space. A generic numerical calculation approach using MATLAB is presented in [9]. This algorithm exhaustively searches the design space and determines the optimum point to give the highest performance objective.

Genetic Algorithms (GA) [10] are heuristic algorithms that mimic the process of natural selection. Starting from an initial population (that consists of so-called chromosomes), new generations are produced, which hopefully contain better (i.e. fitter) chromosomes than the previous generation. The optimization process selects new offsprings according to a fitness function and the evolutionary iterations continue until a predefined stopping criterion is met.

A Particle Swarm Optimization (PSO) [11] algorithm optimizes a problem by exchanging information with neighboring particles such that a single particle with given position and velocity parameters searches an optimum setting. PSO works based on a mathematical

Algorithm	Problem definition	Design parameters	Performance objectives	validation method	Refer.
Numerical calculation	Tuning of physical layer parameters in Wireless Sensor Network	Node hop distance, Transmit energy, Modulation schemes	Energy per Successful received Bit↓	simulation	[9]
GA	Maximizing sensing converge of wireless sensor network	Sensor positions	Relocation energy↓	simulation	[10]
PSO	Wireless Sensor Network deployment, Node localization, Node clustering and Data aggregation	Node positions, Transmit power, Sensor configuration	Quality of Service↑ Network lifetime↑ Localization error↓ Transmit power↓ Reliability↑	simulation	[11]
DE	Radio Frequency Identifier network planning	Position, Angle, Transmit power	Coverage↑ Interference↓ Cost↓	simulation	[12]
SA	Cognitive Radio system optimization	Transmit power, Modulation type	Power consumption↓ Bit Error Rate↓ Throughput↑	simulation	[13]

Table 1: Design parameters, performance objective, validation method of different optimization algorithms, applied to a variety of complex wireless network problems

formula optimizing a population of solutions (i.e. particles). Finally the optimization process stops when the improvement is below a given limit.

Differential Evolution (DE) [12], similar to GA, starts from a given population and a fixed number of randomly initialized vectors. In every iteration, a newer generation is produced by randomly combining the vectors in order to create a mutation. The newer generation mixed with the target vector is evaluated against an objective function and the selector decides whether or not it should be accepted to compose the next generation.

Simulated Annealing (SA) [13] algorithm is based on the analogous principle of freezing liquid when forming a crystalline structure such that with sufficient time the structure acquires a minimum energy state. In each iteration step, the newly generated point is checked against the current point based on a probability distribution scale proportional to the problem’s analogous temperature. Such points are accepted when the total objective function decreases and the iteration continue until the stopping criteria are met.

Table 1 compares the different multi-parameter optimization algorithms that have been applied to wireless network problems. All multi-parameter optimization algorithms applied on the wireless network problems made use of simulation as a validation method which has several disadvantages, as outlined in the introductory section. On the other hand, this paper investigates the SUMO toolbox to evaluate its suitability for wireless network optimization. The SUMO optimization toolbox is often used in electromagnetic [14] and aerodynamic [15] optimization problems. Even though we are validating the SUMO toolbox in a wireless testbed for the first time, previous comparisons on multi-objective optimizers [14] [15] favours the SUMO variants which our preference is based upon. Therefore this paper goes beyond the state-of-the-art by (i) evaluating the suitability of the SUMO optimizer for wireless

problems and (ii) evaluating the feasibility of multi-objective optimization algorithms in real-life experimentation rather than simulation.

3. SUMO

3.1. Optimizer principles

The SUMO optimizer is an efficient implementation of the well-known Expected Improvement (EI) criterion, popularized by Jones et al. in [6]. In this work, the optimizer is applied to accelerate the optimization of wireless network problems, because the execution of multiple experiments on a wireless testbed is often a time-consuming procedure. A typical optimization problem comprises a set of N network parameters, denoted by a vector $\mathbf{x}=\{x_i\}$ for $i = 1, \dots, N$, which may consist of either discrete numerical or continuous variables. These parameters need to be tuned within the bounds of a pre-specified parameter range of interest $[L_i, B_i]$ in order to reach an optimal network performance. The aim of the overall procedure is to optimize a given objective function $f(\cdot)$ (for example, to maximize the Quality of Service/Quality of Experience (QoS/QoE), or to minimize a certain cost such as energy consumption, etc.). The optimization algorithm starts from a well-chosen initial experimental design, and a global (but only locally accurate) Kriging surrogate model of the objective function is computed. Such Kriging models are part of a broader class of approximation methods, called the Gaussian Processes (GP), and have some interesting properties that can be exploited by the optimizer. Whereas the standard approximation methods predict only a single function value, GP methods can predict the uncertainty of a function value as the realization of a normally distributed random variable $Y(\mathbf{x}) \sim N(\mu(\mathbf{x}), \sigma^2(\mathbf{x}))$, where $\mu(\mathbf{x})$ represents the predicted value for $f(\mathbf{x})$ and $\sigma^2(\mathbf{x})$ the prediction variance at an arbitrary point \mathbf{x} in the parameter space. Based on this random variable $Y(\mathbf{x})$, different statistical criteria (such as the Probability of Improvement (PoI) or EI) can be computed to quantify how interesting a new point in the design space is. In this work, we adopt the EI criterion which simultaneously balances exploration and exploitation [8] of the parameter space. This corresponds to the improvement that is expected to occur when compared to the optimum value obtained so far (i.e., f_{min} or f_{max}). By picking additional points with the highest EI value in the parameter space, the optimization process is directed towards a configuration with optimal performance. For example, in the case of a minimization problem, it can be written in the form of an integral as in [14] where $\varphi(\cdot)$ represents the probability density function of a random variable and $I(x)$ is the improvement function.

$$E[I(x)] = \int_{-\infty}^{f_{min}} I(x)\varphi(Y(x))dY$$

where the improvement function is defined as

$$I(x) = \max(f_{min} - Y(x), 0)$$

$E[I(x)]$ corresponds to the improvement that is expected to occur when compared to the

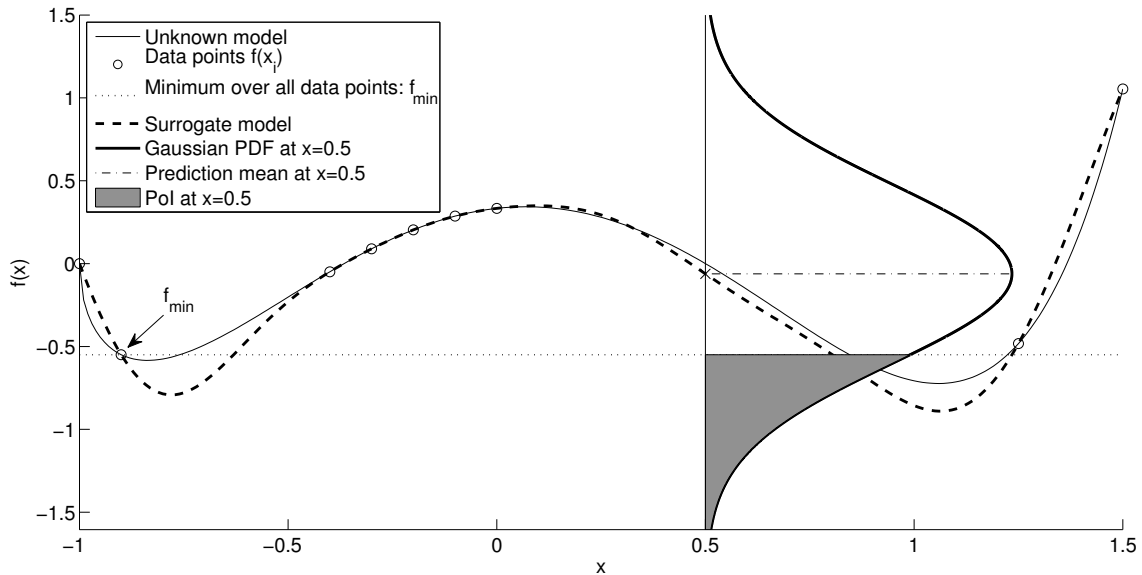


Figure 1: Graphical illustration of a Kriging model and the expected improvement criterion. A surrogate model (dashed line) is constructed based on a set of data points (circles). For each point the surrogate model predicts a Gaussian probability density function (PDF). An example of such a PDF is drawn at $x = 0.5$. The volume of the shaded area is the probability of improvement (PoI) and the first moment of this area is the expected improvement

optimal value of the objective. A graphical illustration of this criterion is shown in Figure 1.

A more detailed explanation can be found in Section II-B of [14]. Note that this EI criterion can also be expressed and evaluated in a closed-form, and it is optimized over the parameter space. The selection of new points corresponds to the execution of new experiments on the testbed and the outcome of these results is used to update the Kriging surrogate model. The process of performing experiments and subsequently updating the model to optimize the objective function is iterated until a stopping criterion is met. Typically, Kriging and the EI criterion are used to solve continuous optimization problems though it can be easily applied to discrete optimization problems too. The optimization of the EI criterion for discrete problems can simply be done by traditional discrete optimizers (such as a discrete pattern search or the discrete version of the CMA-ES algorithm [15]). However, regarding the dimensionality of the problem in this paper it is chosen to evaluate the EI criterion over the complete (and limited) set of discrete possibilities and the best point is chosen. The discrete variables of this work are ordinal of nature, which means that standard Kriging can be directly applied (the standard continuous correlation functions can be used). Though, nominal or categorical variables can also be easily handled by Kriging by using an appropriate function to describe the correlation between the discrete data points [16].

In short, the SUMO optimizer is an effective approach to optimize network performance on a real-life testbed. The experimental results confirm its effectiveness and robustness.

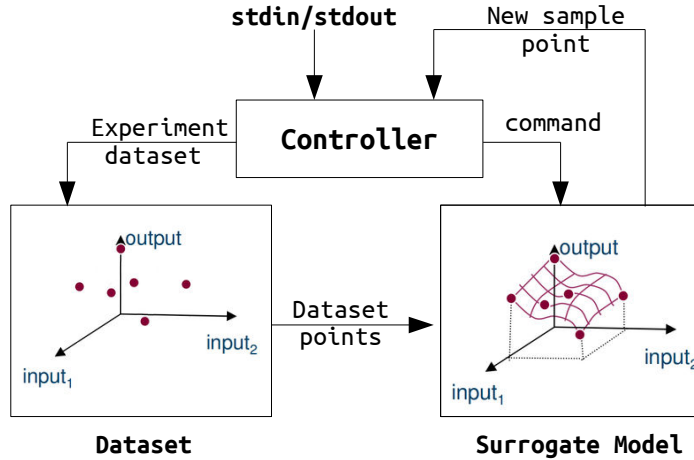


Figure 2: Overview of generic SUMO toolbox

3.2. Toolbox modification

Out of the box, the SUMO toolbox is used as a complete multi-parameter optimizer. It has a controller unit sitting at the highest level and manages the optimization process. Figure 2 describes the SUMO toolbox highlighting the control and optimization functions.

The controller manages the optimization process starting from a given initial dataset (i.e. initial sample points + outputs) and generates a surrogate model. The surrogate model approximates the dataset over the continuous design space range and is used by the optimizer instead of the dataset. Next, the controller predicts the next design space element from the constructed surrogate model at locations where the expected improvement is the largest, with the aim of further meeting the optimization’s objective. The optimization process iterates until stopping conditions are met.

Normally, the SUMO toolbox works as a complete optimization solution when used out of the box. However, in the context of wireless testbeds, the SUMO toolbox has to be controlled by the underlying testbed management framework. This means the controller part of SUMO toolbox has to be replaced by the testbed’s management framework and Figure 3 shows the modification and integration work.

This testbed management framework performs similar tasks as the original SUMO controller except for the addition of a number of tasks like experimentation on the wireless testbed, storing the dataset on a separate file, and reading the experiment configuration from a file.

4. Experimental validation

This section verifies, by integrating theoretical solutions from the previous section, the use of SUMO optimization toolbox in a wireless conference network problem using a wireless testbed. First we give a description of the wireless testbed where experimental validation is carried out. Next the experiment scenario and the optimization processes are presented.

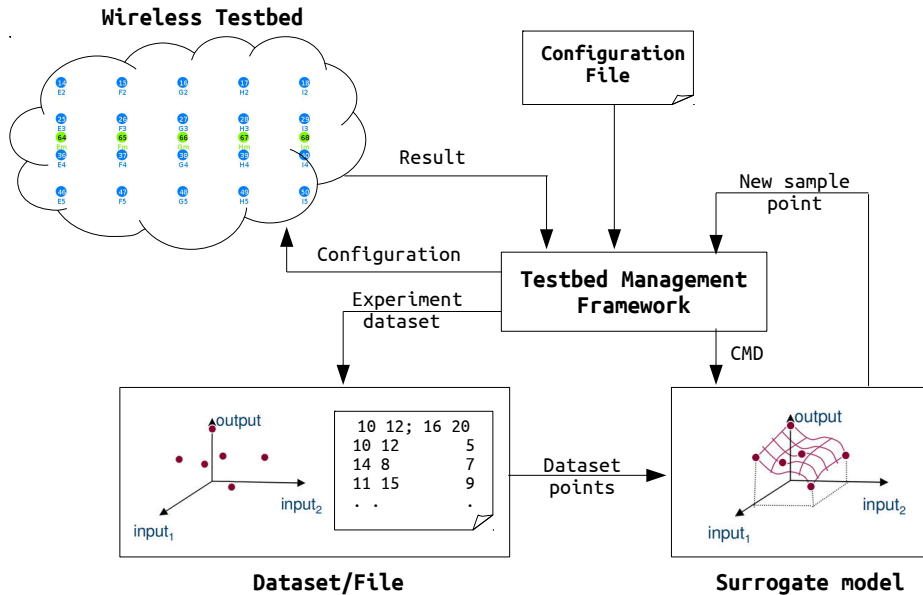


Figure 3: Integration of the modified SUMO toolbox in the wireless testbed

Finally, we look in more detail at the performance objectives and discuss how conflicting performance objectives can be combined into a single objective.

4.1. Wireless testbed

The wireless *iMinds w-iLab.t* testbed, located at Zwijnaarde (Ghent, Belgium) [17], is shown in Figure 4. The testbed, equipped with heterogeneous wireless devices, is mainly used for wireless experimentation. It has 60 nodes each consisting of an embedded Zotac PC having two Wi-Fi interfaces, a sensor node, a Bluetooth dongle and a wired control interface connected to the testbed management framework. Furthermore, the testbed is equipped with advanced spectrum sensing devices. These include Universal Software Radio Platform (USRP), IMEC Sensing Engines, and Wireless open Access Research Platform (WARP) boards. The wireless testbed is also equipped with mobile nodes mounted on Roomba robots allowing mobility experiments.

4.2. Experiment scenario

The SUMO optimizer is validated by experimentally optimizing a wireless conferencing scenario. Figure 5 shows the wireless conferencing scenario that comprises a wireless speaker broadcasting a speaker’s voice over the air and multiple wireless microphones receiving the audio at the listener end. This type of wireless network is used in a multi-lingual conference room where the speaker’s voice is translated into different languages and multiplexed into a single stream. Next, the stream is broadcasted to all listeners and each listener selects their preferred language.

Often, the speaker’s audio quality is reduced by external interference and the surrounding environment is impacted by external interference. Thus, the main objective of the wireless

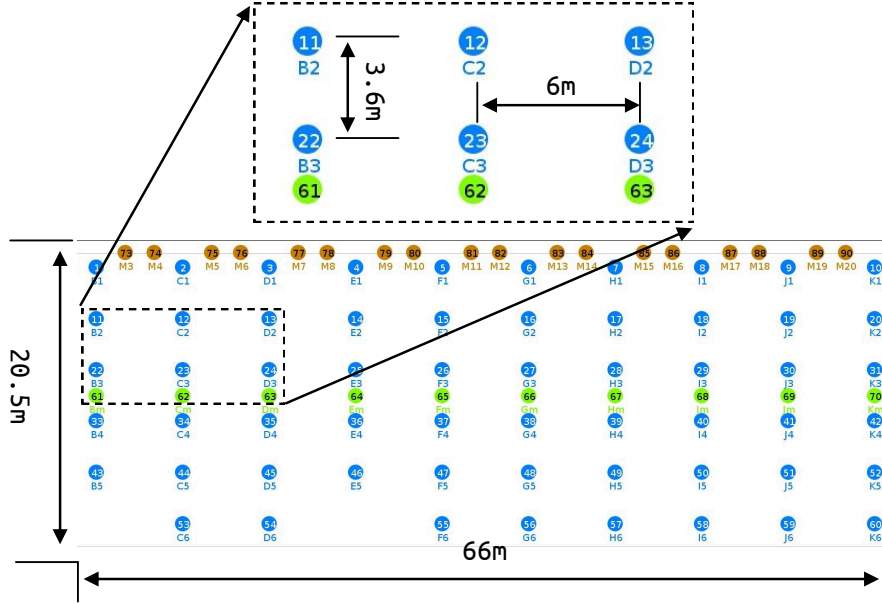


Figure 4: Top view of iMinds w-iLab.t wireless testbed

conferencing scenario is to improve the received audio quality while keeping the transmission exposure at a minimum. To this end, the conferencing operator has the possibility to adapt the speaker’s channel and transmission power parameters.

The experiment is composed of 1 interferer creating background interference and a System Under Test (SUT) having 1 speaker and 8 listeners. The speaker broadcasts a 10s audio stream, obtained from ITU-T Test Signals for Telecommunication Systems [18], and each listener calculates the average audio quality within the time frame. The audio stream is encoded using A-Law encoding format at 64 Kbps bitrate. Moreover, the interferer transmits a 10 Mbps continuous UDP stream on dual channels (i.e. 1 and 13) generated using the iperf [19] application. The speaker, listeners and interference generator are shown in Figure 5.

On the left hand side of Figure 5, the realistic wireless conferencing scenario is shown, where as on the right hand side, the experimentation scenario is mapped on the iMinds w-iLab.t testbed. All listener nodes (i.e. 38, 39, 40, 48, 50, 56, 57, and 58) are associated to the speaker access point (i.e. node 47). Background interference is created by the access point (i.e. node 49) using two separate Wi-Fi cards. The Wi-Fi card and driver used for this experiment are ”Atheros Sparklan WPEA-110N/E/11n mini PCI 2T2R” and ”Atheros ath9k” respectively. The SUMO algorithms run on a dedicated PC that can communicate with all nodes of the experiment.

So far, we have looked one scenario which is a speaker broadcasting 1 language stream to 8 listeners. We also considered two more scenarios to analyze the effect of an intense wireless medium usage and a varied network topology that has on the SUMO optimization. For the intense wireless medium case (i.e. scenario II), we increased the number of language

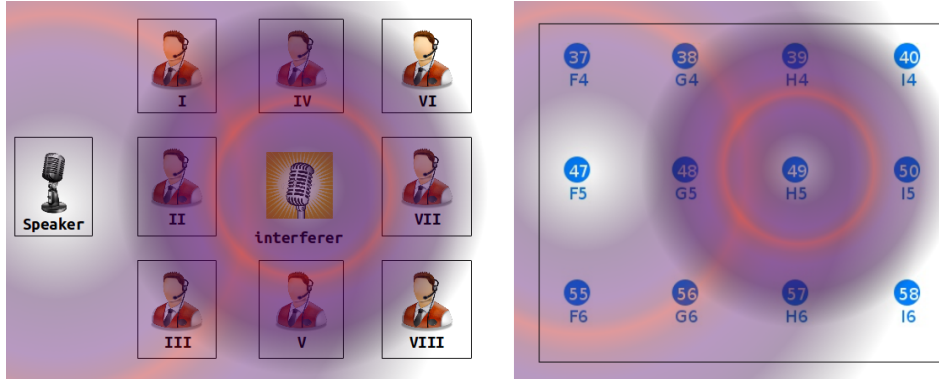


Figure 5: Left: wireless conferencing scenario consisting of 8 listeners, 1 speaker, and 1 interferer. Right: mapping of the conferencing scenario to the testbed nodes. The transmission range of the speaker and interferer is indicated.

streams from 1 to 8 and for the varied network topology case (i.e. scenario III), we increased the number of listeners from 8 to 16 using 1 language stream. The physical data rate used in all cases is 1 Mbps using the Direct Sequence Spread Spectrum (DSSS) 802.11 modulation technique.

4.3. Optimization process

The optimization process is illustrated step by step in Figure 6. At (1) the controller is given a list of settings of the first experiments that needs to be configured on the wireless testbed. (2) Experiments are deployed on the wireless testbed using the requested settings, thus resulting in an initial sample set. (3) At the end of each experiment, the controller retrieves the evaluation criteria of the experiment. For the conferencing scenario, the evaluation criteria are the audio quality and exposure performances from all listeners. (4) An objective function is created by processing the evaluation criteria (see Section 4.4). (5) When the SUMO optimizer receives a sufficiently large dataset, it generates a surrogate model. (6) The next sample point with highest expected improvement is predicted. (7) The controller starts the next optimization experiment using the new design parameters. (8) Again, the evaluation criteria are retrieved and (9) the objective function is calculated for the new design parameters. (10) Based on the current dataset, extended by one record, the surrogate model is updated and (11) a new sample is predicted. The optimization process continues until stopping conditions are met.

4.4. Performance objectives

Dual objectives are applied in the wireless conference network problem. The first objective is maximizing the received audio quality which is measured using the Mean Opinion Score (MOS). MOS is a subjective audio quality measure represented on a 1 to 5 scale (i.e. 1 being the worst quality and 5 being the best quality). To calculate the MOS score, the experiment described in Section 5 uses the ITU-T Perceptual Evaluation of Speech Quality (PESQ) P.862 standard. It calculates the PESQ score out of packet loss, jitter and latency network parameters and maps it onto a MOS scale [20].

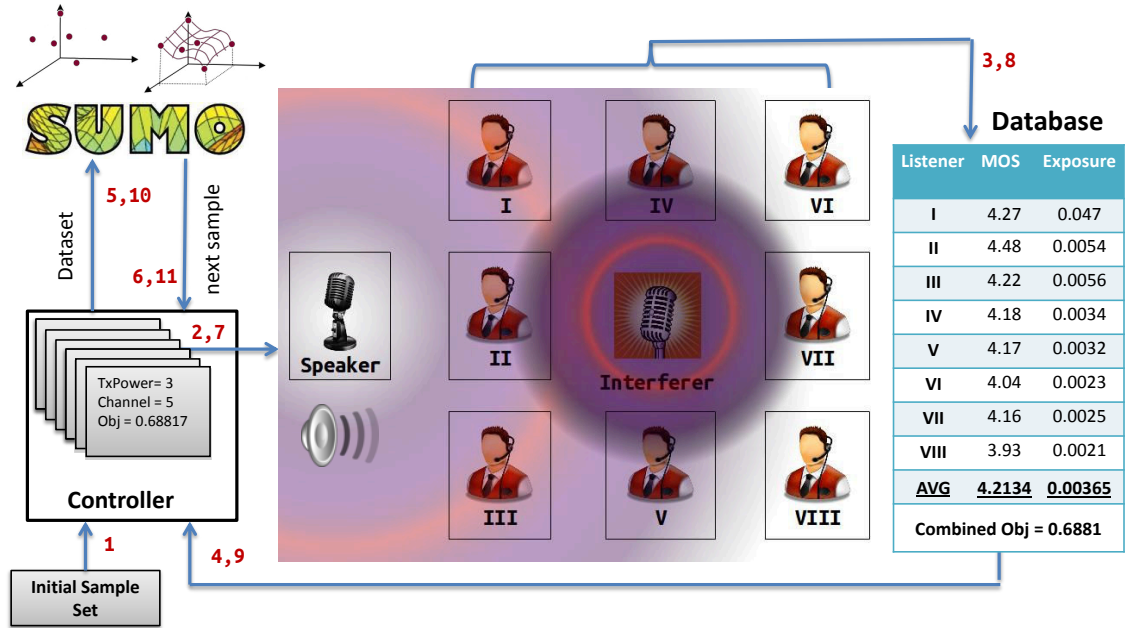


Figure 6: The process of SUMO optimization in the wireless conference network problem. The different sequential steps are numbered from 1 to 11.

The second objective is minimizing transmission exposure. In [21] an in depth calculation of transmission exposure is presented. The exposure at a certain location is a combined measure of received power and transmit frequency. Transmission exposure is an important evaluation metric related to potential health issues, leading the regulatory bodies to set limits on maximum allowable radiation levels.

As maximizing the combined objective is the goal, the weight of performance metrics needs to be defined depending on the problem type. For example, a person who wants to install a wireless conferencing system in urban areas applies tighter exposure requirement than in rural areas. We would also apply high audio quality requirement in parliament auditoriums compared to office meeting rooms. However in our case, the aim is to validate the SUMO toolbox and we combined both metrics first by normalizing them to a $[0 \ 1]$ scale, followed by subtracting the two metrics into a $[-1 \ 1]$ scale and finally renormalizing the combined metric back to a $[0 \ 1]$ scale. To this end, the exhaustive search experiment is used to reference the normalization. Moreover, the combined metric is evaluated in each listener nodes and a representative combined metric is calculated first by averaging the combined metrics and next by selecting the one closest to the average.

5. Result and discussion

This section will analyze the viability and efficiency of using the SUMO global optimization technique for wireless experimentation. A methodology for detecting the validity of experiments by detecting outliers is described in section 5.1. The overhead of performing



Figure 7: PRE and POST experiment monitoring.

an exhaustive search is given in section 5.2, the outcome of which is used as a reference for experiment comparison. Experiment repeatability is discussed in section 5.3. The sensitivity of experiments to the choice of the initial sample size is discussed in section 5.4. Potential stopping criteria are analyzed in section 5.5. Finally, the SUMO optimized experiment is compared against the exhaustive search model in Section 5.6.

5.1. Experiment outlier detection

Each experiment iteration has a chance of returning invalid measurement data. An experiment conducted at a certain time can show a different result when repeated at a later time. The most recurring reasons for this anomalous behavior are node malfunctioning and external interference. Since node malfunctions can be detected by most experimental testbeds, this section only focuses on methods to observe and handle experiment outliers due to external interference. This however does not mean interference generated inside SUT since it is part of the experiment.

In any wireless experiment, external interference (i.e. from an external device not participating within the experiment) competes for the scarce wireless medium and as such affects the behavior of the SUT. Ideally, experiments are performed in a controlled environment where unwanted external interference is blocked. However, this cannot be easily achieved, as shielding an experimental environment requires a great deal of money. As a result, many experimental facilities are currently installed in readily-accessible environments that serve multiple purposes, such as office buildings. For these areas, an alternative approach is to measure the effect of interference on an experiment. This approach does not block the interference but gives an interference score for each experiment conducted. Depending on the score, the experimenter decides whether or not to discard the experiment. One way of measuring external interference is by doing correlation measures on the experiment outcome [22]. Such measures correlate the output of identical experiment runs and discard those having lower correlation scores. The disadvantage of this method is that at least three experiment runs are required before getting a tangible result. A second approach for experiment outlier detection, shown in the Figure 7, is by doing PRE and POST experiment monitoring [23].

In this approach, the interference level is measured before the experiment starts (PRE) and only if the environment is clean that the experiment execution is triggered. After the experiment ends, the interference level is measured again (POST) to estimate a possible interference on the executed experiment. The main idea of this approach is that an experiment has a higher chance of being interfered by external interference if the environment is not clean during the PRE or the POST experiment.

The above two approaches give indirect indication of experiment interference levels. In contrast, a direct approach sniffs the wireless medium for external interference during experimentation. The level of difficulty imposed varies with the types of interference considered

[24]. As the Industrial Science and Medical (ISM) band is a non-licensed band, a number of technologies may coexist together and the impact of external interference from such devices is difficult to characterize. As such, all these technologies need to be considered, thereby requiring a multi-layer (i.e. feature and energy detection) and multi-technology (i.e. Wi-Fi, Zigbee, Bluetooth, . . .) distributed sensing solution [25].

Implementation of a distributed sensing solution was under way by the time of writing this paper. Hence we revert to a different approach by taking advantage of the pseudo-shielded nature of the iMinds w-iLab.t wireless testbed. The wireless environment is pseudo-shielded mounted on top of a clean room and experiments are guaranteed with clean environment from external (outside the wireless testbed) interference. Therefore, the PRE and POST experiment monitoring along with the pseudo-shielded wireless environment are used for the experiment outlier detection.

5.2. Exhaustive search model

In this section, we describe a reference experiment that was performed to generate an exhaustive search model of the wireless conference network problem. Neither SUMO nor any optimization algorithm is used to generate the model. The exhaustive search model evaluates all possible combinations of settings and will be used as a reference model for comparing SUMO optimization experiments. In total, 260 experiments (i.e. 13 Channels \times 20 Transmit Power) were executed during which an interference is created continuously on dual channels (i.e. 1 and 13). We start by making a complete analysis using the first scenario and later present the models of the remaining two scenarios (see Section 4.2).

Figure 8 shows the outcomes of the exhaustive search model for the first scenario. The exposure model of Figure 8(a) only considers the exposure from the speaker but not from background interference, since the goal of the SUT is to reduce its own exposure. Moreover, exposure depends on the distance of the speaker and the average exposure over all listener nodes is calculated at the medium exposure point (Section 4.4). For the first scenario, the range of exposure values at the medium location is shown on the color bar of Figure 8(a). The exposure objective degrades with increased transmission power independent of the used channel. In contrast, the audio quality objective increases with increased transmission power and the influence of interference can be noted on multiple channels. There is an area on the non-interfered channels (i.e. 6 to 8) where adequate performance is observed also for lower transmit Power (i.e. 1dBm to 6dBm). This area is of interest because it represents a region where exposure is low. On the other hand, the worst performance from the audio quality model is shown between channels 2 to 4, 10 to 12 and transmit power 1 dBm to 7 dBm. Interestingly, this region is not located on channels where background interference is applied on but on the neighboring channels. This is due to the fact that the speaker and interferer nodes apply CSMA/CA medium access on identical channels but to a limited scale on neighboring channels which results in degraded performance [25].

The combined objective model from Figure 8(c) is a combination of the exposure model and the audio quality model from Figure 8(a) and 8(b) respectively. As expected, the non-interfered regions with low transmission power have the highest values for the combined objective function. Figure 8(d) shows a different view of Figure 8(c) where the combined

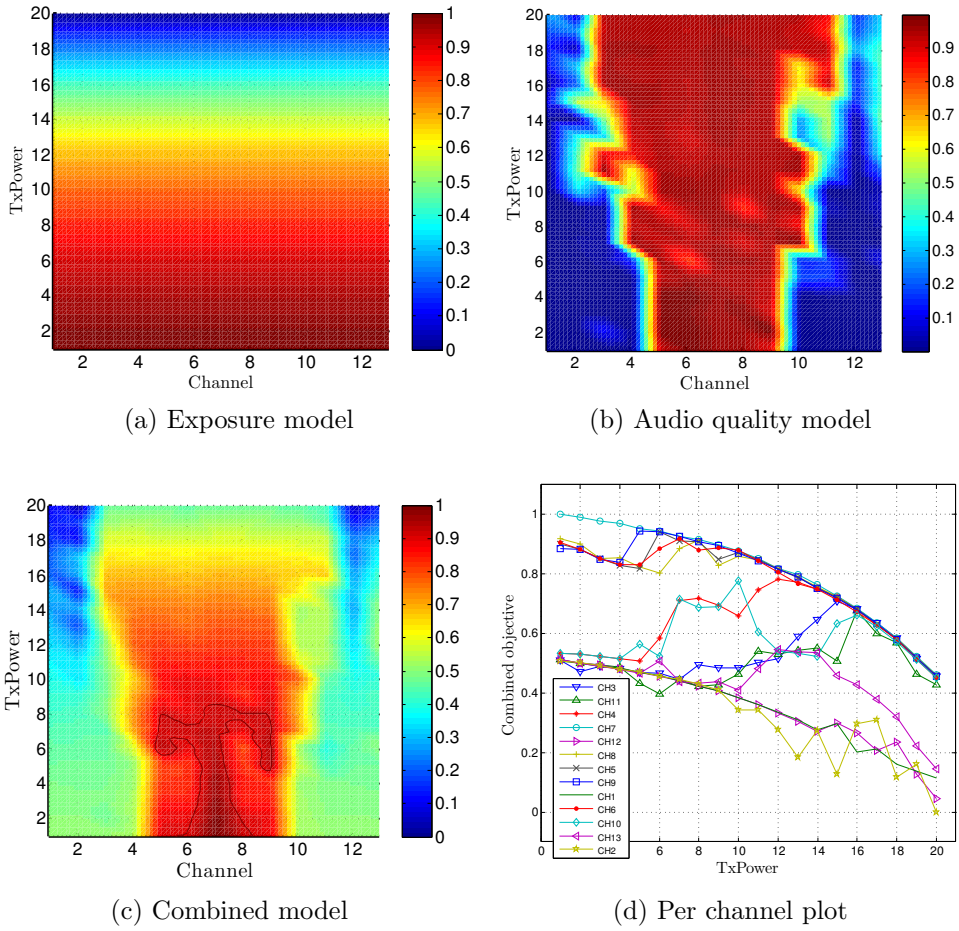


Figure 8: Exhaustive search model. Background interference at channels 1 and 13.

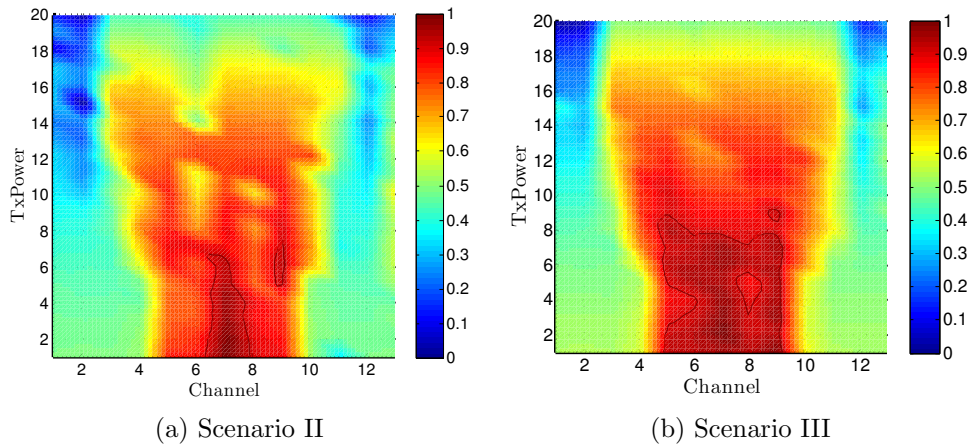


Figure 9: Exhaustive search model. Area inside the black contour is the Optimum region.

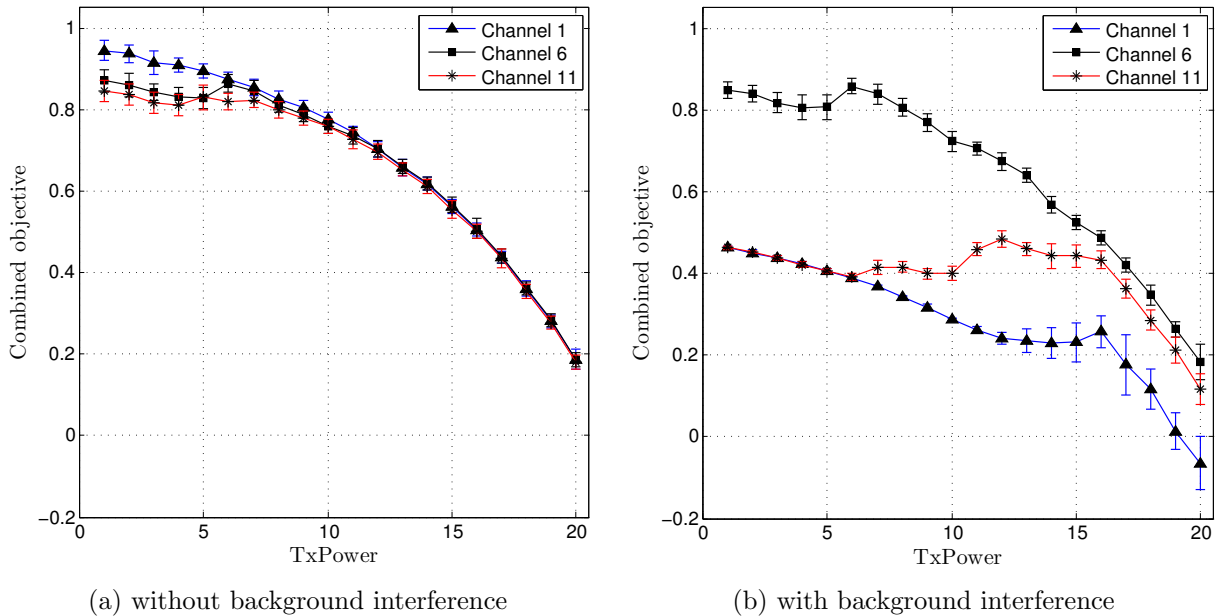


Figure 10: Repeatability test at Wi-Fi channels 1, 6, and 11

objective model is plotted for different transmission channels. Color bar shows the strength of the objectives in their respective limits where red indicates the best performance and blue indicates the worst performance.

Figure 9 shows the exhaustive models of scenarios II and III. Scenario II used the wireless medium intensely bringing a lot of agitation to the system and thus a smaller optimum region compared to scenarios I and III. On the other hand, scenario III used the same wireless medium as to scenario I but increased the number of listeners which stabilizes the average performance over the listener nodes and a larger optimum region as a result.

5.3. Experiment repeatability

A basic criterion for comparing wireless experiments is the requirement that experiments are repeatable. Identical experiments conducted in different time frames should show similar performance. One way of checking repeatability is by calculating the Standard Deviation (STD) of identical experiments and compare it with a threshold. Using scenario II (see Section 4.2), two sets of experiments were performed each on three different channels (i.e. 1, 6, and 11) and 35 repeated experiment runs were conducted. The first set considered a clean environment without background interference whereas the second set considered background interference at channels 1 and 13. Once again, a 10 second audio stream is transmitted by the speaker at different transmission power levels and listeners calculate the combined objective.

Figure 10 shows the experiment outcomes, using error bars to show the STD of the objective function on top of the average value. The lower the STD on the error bar, the higher the experiment repeatability. From all tests, the experiment of Figure 10(b) at

channel 1 shows the worst repeatability with a decreasing trend as we increase the transmit power. A physical layer investigation between the speaker and interferer packets reveals that at lower transmit power, the interferer does not see packets coming from the speaker thus jamming the environment constantly assuming it is clean. Whereas for the speaker, it applies the CSMA/CA medium access method and it avoids the medium for most of the time or collides with the interferer in case it transmits. This increases the number of lost packets and eventually to a very low audio quality for the listener nodes. Because of the loss in audio quality and the same exposure at fixed power levels, the combined objective at a listener node does not show variation with repeated experiments. However, when we increase the transmit power, the interferer feels the presence of the speaker and it starts applying the CSMA/CA medium access method before transmitting its packets. This lets the speaker to transmit without being interfered and the listeners to receive a higher audio quality. Since the medium is now shared by the two transmitters, the audio quality starts fluctuating depending on the time share the speaker has possessed during the experiment. This creates the difference in the combined objective and reduces repeatability between repeated experiments. On the other hand, experiments on channels 1 and 11 from Figure 10(b) show similar repeatability trend except on a reduced scale at channel 11. On channel 11, the interferer overlaps part of the spectrum with the speaker. Thus the same principle applies to the experiment on channel 1 and shows decreasing repeatability as transmit power increases. However, due to the fact that CSMA/CA is used on a limited scale on neighboring channels [25], repeatability at channel 11 is slightly better than at channel 1. On the other hand, the experiment of Figure 10(b) at channel 6 has minimal background interference, hence the experiment behaves similar to the case without interference.

So far we have analyzed the repeatability test with and without interference. But the question we need to answer should be, is the wireless environment repeatable? Since any wireless environment can not be 100% repeatable, we have to leave a certain margin depending on the problem type. This margin depends on the dynamics of the wireless environment but addressing it requires state of the art ideas and tools. Instead we revert to a different approach by comparing repeatability at its worst condition (i.e. with a highest co-channel interference on SUT) and at its best condition (i.e. with no interference on SUT). This comes down to comparing the repeatability variation of Figure 10(a) and 10(b) and the best case variation (i.e. $STD = 0.0301$) is smaller than the worst case variation (i.e. $STD = 0.0747$) and that proves the repeatability test.

5.4. Initial sample size sensitivity

As explained in Section 4.3, a surrogate model predicts the next experiment input parameters with a highest expected output performance. However, the initial model requires a set of initial sample points from the design space and performance outputs. This section investigates how many initial samples are required before a usable surrogate model can be created.

The initial sample points for any problem have to be selected carefully such that the optimization process quickly converges to the optimum. If the number of initial sample points is large, the optimizer spends too much time during exploration work. On the other

hand, considering few initial sample points leads to the risk of missing global optimums and thus exploiting local optimums instead. One way to address the trade-off between exploration and exploitation during optimization is by selecting an appropriate initial sample size. Usually this depends on the complexity of a problem’s global model. The more complex a problem’s global model is, the larger the initial sample size needed to have good surrogate model approximation and vice versa. It was indicated in [26] that extreme points of a surface can be used to measure the complexity of a problem. These are the minimums, maximums and saddle points of a problem’s global model. Moreover, it is also indicated that by setting the initial sample size to the number of extreme points, an optimizer has a higher chance to arrive at the global optimum in short amount of time. This assumption only works if the problem’s extreme points are known beforehand. Most of the time this is not the case as we generally optimize unknown problems. Moreover, initial sample size selection depends on the problem type [26]. For our specific problem, setting the initial sample size to 8 points is found a good choice. The 8 initial sample points together with the corner points which the SUMO optimizer adds, sums up to 12 initial points in total.

In the following sections, we will each time analyze four different sampling methods to pick the 12 initial sample points from the design space. These are

- Latin Hypercube Sampling (LHS) [27], which is a stratified sampling method that selects sample points evenly along the design space while ensuring proportional representation of design variables.
- Orthogonal sampling, which divides the design space into a number of sub-spaces and LHS is applied in each sub-space.
- Random sampling, which selects points randomly over the design space.
- Hammersley Sequence Sampling (HSS) [28], which is a low-discrepancy quasi-random sampling method providing better uniformity properties and uniform distribution of points in space.

5.5. Stopping criteria

The main goal of the global optimization is to reduce the number of required experiments. This section investigates the effect that different stopping criteria have on the problem’s optimum value. In this paper, we look at two stopping criteria named Fixed Iteration (FI) and Objective Function Improvement (OFI). With the FI stopping criterion, a fixed number of iterations are conducted and the optimum value from the output is selected. On the other hand, the OFI stopping criterion looks at the relative difference in performance and stops the iteration when the STD of the top sorted N iterations falls below a certain threshold. The idea behind the OFI stopping criterion is that the sorted objective function of a list of experiments ideally approaches a flat curve as the number of experiments increases.

To perform sensitivity analysis, experiments are conducted using the SUMO toolbox until the stopping criterion is met. Using scenario I (Section 4.2), a plot of normalized combined objectives as a function of iteration count is shown on Figure 11. Among the different design

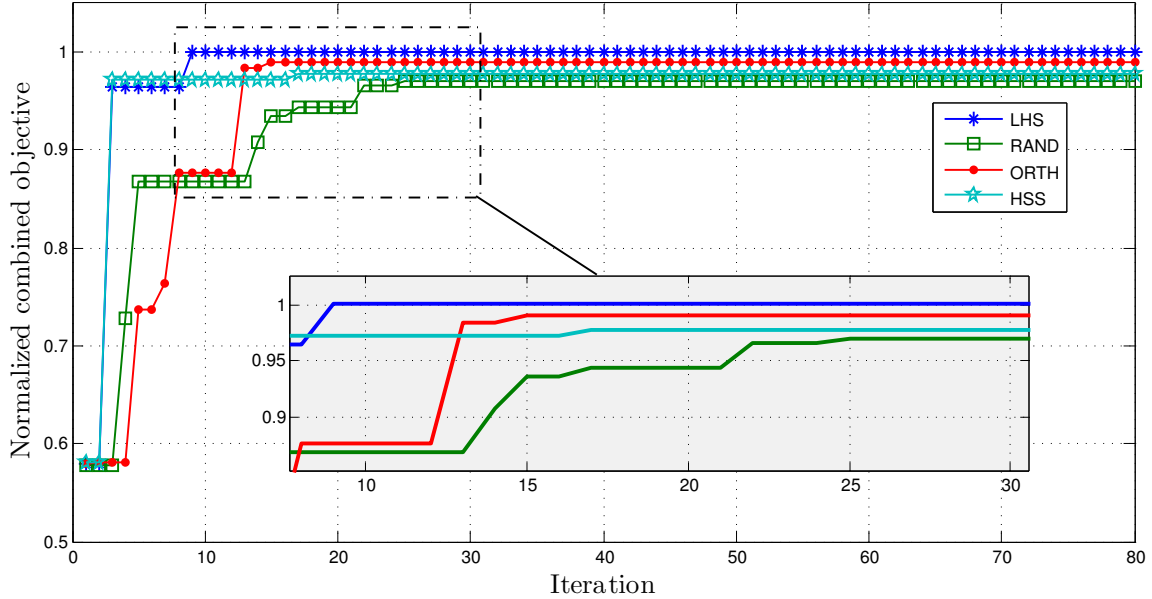


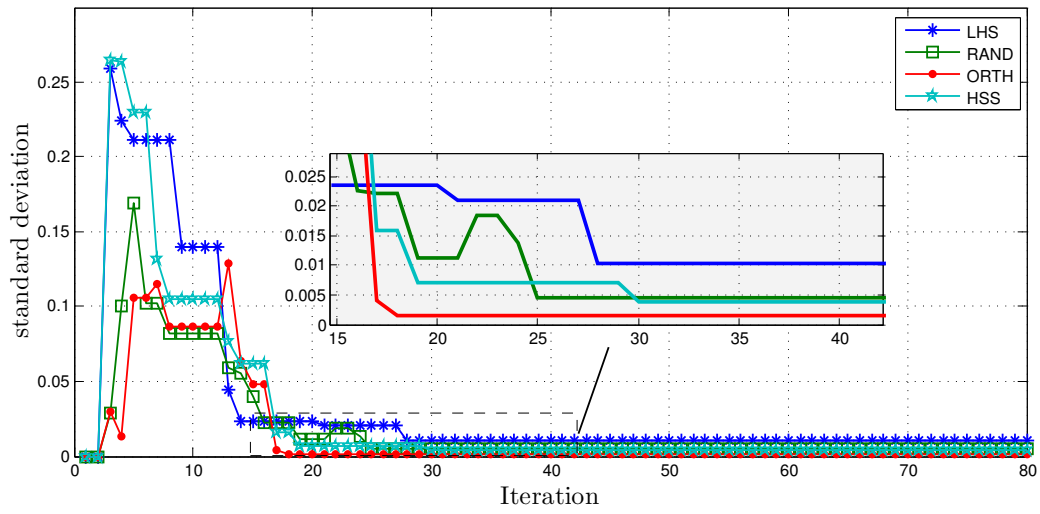
Figure 11: Scenario I normalized combined objective as a function of experiment iteration

space sampling methods applied, the LHS method reaches the Global Maximum Combined Objective (GMCO) first (after iteration 9) whereas the Random sampling method arrives last (iteration 25). This, however, does not mean LHS is preferable for all problem types but for the current problem it approximates the global model better than any other sampling method. On the other hand, we see the plots of the three experiments (i.e. ORTH, HSS, RAND) does not reach 1 and this is due to the small repeatability variation we have seen on the Figure 10(b) at channel 6.

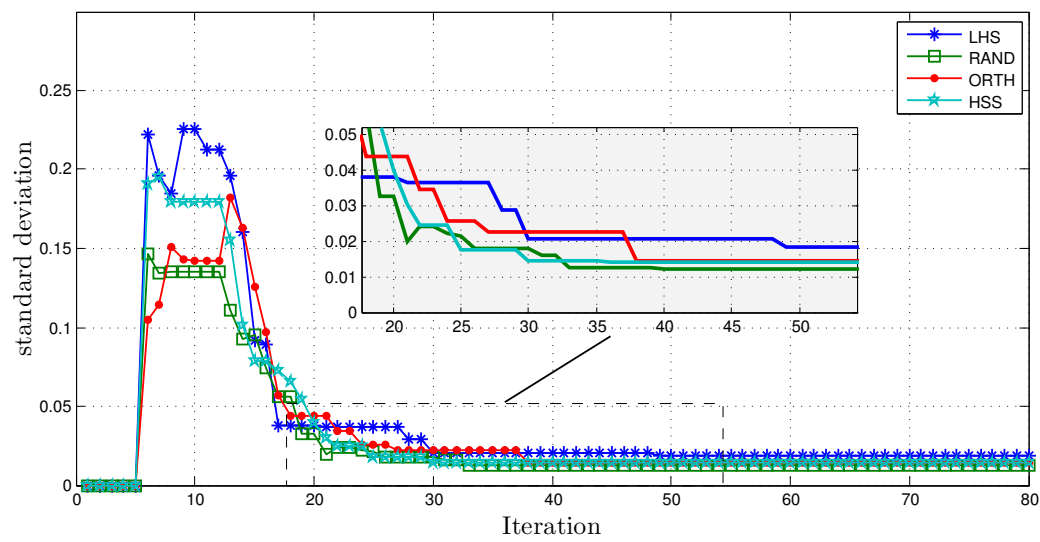
FI stopping criterion sets one parameter which is the number of iterations an experiment needs to execute. It is clear from Figure 11 that it is difficult to draw a conclusion about the number of iterations since the iteration count of the different sampling methods to reach the GMCO is highly variable. As such, the FI stopping criterion is mainly useful for time-constrained testbeds where experimenters can only reserve resources for a limited time. It can be used as an upper limit in case all other stopping criteria fail to satisfy.

On the other hand, the OFI stopping criterion, as stated previously, considers the relative performance difference between experiments and stops the iteration when the STD of the top sorted N iterations falls below a given threshold. The OFI stopping criterion has two parameters to set. These are the STandard Deviation WIDTH (STD-WIDTH) which sets the number of objective performance values in the STD calculation and the STandard Deviation THreshold (STD-THLD) which is used as a lower limit for the stopping criterion. Figure 12 shows the STD curve of scenario I (section 4.2) as a function of iteration count for STD-WIDTH 3, 6 and 10. These numbers are wide enough to show the behavior and variation of different STD curves. Calculation starts after the iteration count reaches STD-WIDTH.

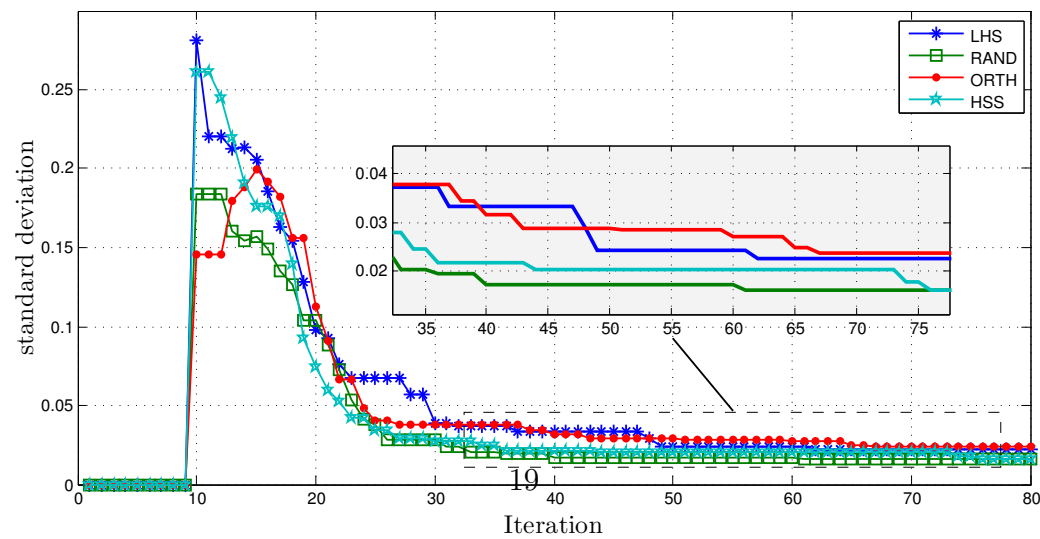
As stated previously, the output of the plots for each STD width approaches a flat curve when the optimization reaches the optimum. On the other hand, the randomness of the



(a) STD-WIDTH=3

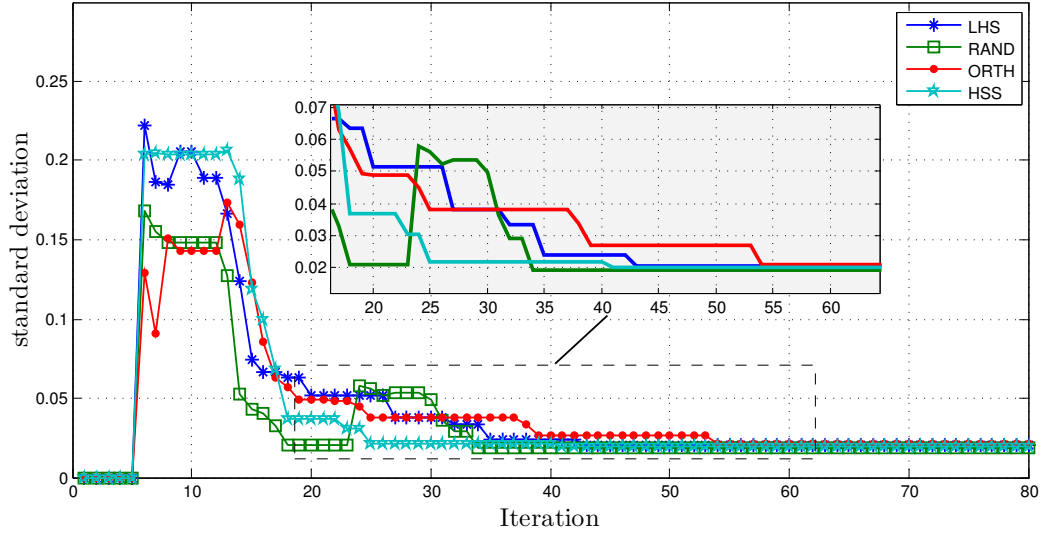


(b) STD-WIDTH=6

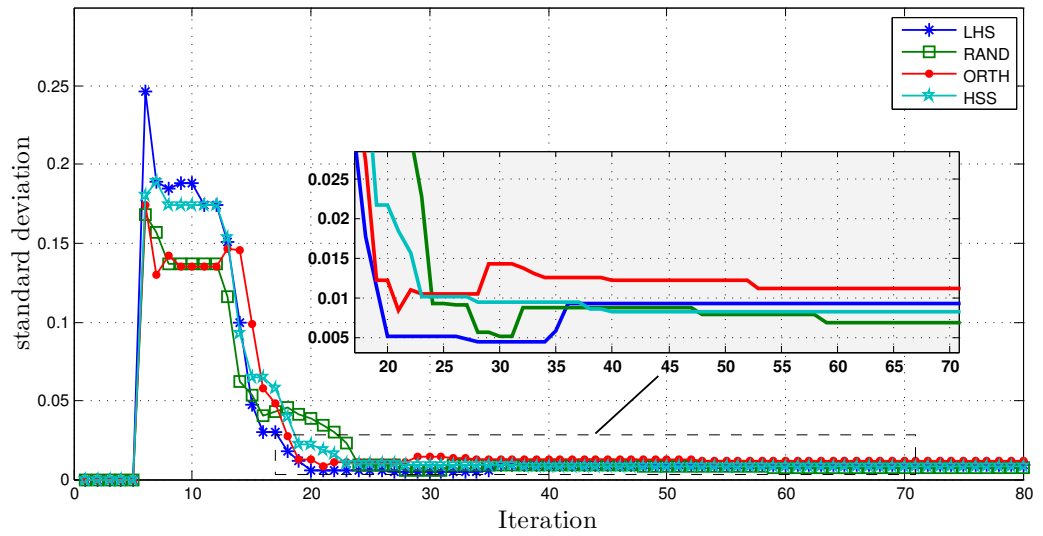


(c) STD-WIDTH=10

Figure 12: Scenario I standard deviation as a function of experiment iteration



(a) STD-WIDTH=6



(b) STD-WIDTH=6

Figure 13: Scenario II and III standard deviation as a function of experiment iteration

curves gradually decreases as the STD-WIDTH increases. This also increases the settling time until the lowest STD value is reached. For example looking the LHS experiment from Figure 12, the settling times for the three STD-WIDTH parameters 3, 6 and 10 are 28, 49 and 62 iterations respectively. Also note the benefit of the SUMO optimization with a sharp declining curve after the 12 initial experiments. As the optimization continues, the STD curve starts decreasing and converges to a stable value.

There are two things we want the STD curve to achieve. First, we want the curve to reach a stable value as fast as possible. This depends on the size of the optimum region in the problem’s global model. The optimum regions of all scenarios are shown as a black contour on top of the exhaustive search model (i.e. Figures 8(c), 9(a) and 9(b)). The larger this area, the sooner the optimization locates the optimum and the STD curve converges to a stable value and vice versa (see next section). However, the size of a problem’s optimum region is not known beforehand and a good value of STD-WIDTH, in such cases, is half the elements of the initial sample size. In our case, STD-WIDTH will be 6 (i.e. $12/2 = 6$).

Second, we want the curve to reach a very small stable value. Again, this value never approaches to zero as the wireless medium shows a small repeatability variation. Since the data points used in the STD calculation after it gets stable are inside the optimum region, STD-THLD can assume the maximum STD where repeated experiments show over the optimum region. Again, the optimum region is not known before hand and the work around is to perform repeatability tests without background interference (i.e. Figure 10(a)) and select the maximum value from the list (i.e. scenario I = 0.02418, II = 0.02865 and III = 0.02067).

The STD performance for scenarios II and III are shown in Figure 13.

5.6. Performance comparison

Now the stopping criteria and initial sample size are selected and experiment repeatability is verified. Next, we compare the SUMO approach to the traditional experimentation that exhaustively searches all parameters. For the comparison, we have defined the parameters of the OFI stopping criterion to the following: STD-WIDTH = 6, STD-THLD1 = 0.02418, STD-THLD2 = 0.02865 and STD-THLD3 = 0.02067. Table 2 shows the performance metrics of each conducted experiment for the three scenarios when these parameters are applied. The four different sampling methods from Section 5.4 are also included. The required *number of iterations*, before the stopping conditions are met, are shown for each of the sampling methods. The *Duration Gain* metric calculates the rate by which SUMO experiment duration is reduced compared to the exhaustive search experiment that took 260 experiments. The *Performance Gain* metric evaluates how close the optimum solution of the SUMO experiment is to the GMCO value.

When comparing all scenarios from Table 2, LHS is found the best sampling method and RAND is the worst sampling method in terms of performance gain. The LHS sampling method almost all the time achieved the GMCO value of the problem but for the RAND sampling method, depending on the scenario, it shows a wide performance variation. On the other hand, RAND sampling method converges the quickest in the first two scenarios but at the expense of a lower performance gain (scenario I = 89.15% and II = 78.76%). This

Sampling Method	No. of iterations	Duration Gain	Performance Gain
LHS	30	260/30=8.67	3.9398/3.9480=99.79%
RAND	21	260/21=12.38	3.5198/3.9480=89.15%
ORTH	27	260/27=9.63	3.8663/3.9480=97.93%
HSS	25	260/25=10.4	3.7708/3.9480=95.51%

(a) Scenario I

Sampling Method	No. of iterations	Duration Gain	Performance Gain
LHS	35	260/35=7.42	3.9139/3.9192=99.86%
RAND	18	260/18=14.44	3.0869/3.9192=78.76%
ORTH	39	260/39=6.67	3.9102/3.9192=99.77%
HSS	25	260/25=10.4	3.7749/3.9192=96.32%

(b) Scenario II

Sampling Method	No. of iterations	Duration Gain	Performance Gain
LHS	18	260/18=14.44	3.5419/3.6803=96.24%
RAND	24	260/24=10.834	3.4565/3.6803=93.92%
ORTH	19	260/19=13.68	3.3069/3.6803=89.85%
HSS	21	260/21=12.38	3.4926/3.6803=94.89%

(c) Scenario III

Table 2: Duration Gain and Performance Gain of SUMO optimized experiments using 4 sampling methods

is because of poor initial samples are explored and it leads to a local optimum instead of a global optimum. Looking the RAND curve of Figure 13(a) in particular reveals this finding by having bumps along the curve (iteration 2334), had we continue the optimization. The other finding discerned from Table 2 is that the number of iterations generally decreases with an increase in the problem’s optimum region. Scenario III, having the largest optimum region, has the smallest number of iterations followed by scenario I and II. On the other hand, when looking at the different sampling method performances, LHS and ORTH show similar performance in both metrics and in all scenarios. This is because both are using latin hypercube sampling and so does their similarity on performance. In addition, both LHS and ORTH are the most sensitive sampling methods to a change in the optimum region.

6. Future work

As was explained in the introductory section, selective sampling of the design space and sensitivity analysis on design parameters are the two most widely used approaches for reducing experimentation time overhead. Selective sampling is used in this paper and sensitivity analysis is planned as a future work.

We also planned to test additional types of multi-parameter optimizers on different network problems using a wireless testbed. Similar to the aerodynamic problem where the

SUMO variant, CMO, outperforms all other optimizers [29], we want to test the performance of different multi-parameter algorithms. We also want to test the performance of combined optimization on wireless network problems. For example, by using SUMO and GA working on specific tasks of the optimization process.

In addition, future work will focus on network optimization in dynamic environments. SUMO optimizer works in a static environment and it builds the surrogate model along the optimization process. However if the environment is changing from time to time, the SUMO optimizer fails to incorporate this. This work will be a vital part of cognitive wireless networks.

Finally, the experiment outlier detection used in this paper is not full-proof and we want to extend this by incorporating a multi-layer (i.e. feature and energy detection) and multi-technology (i.e. Wi-Fi, Zigbee, Bluetooth, ...) interference estimators. In addition, other criteria for checking the validity of experiments will be investigated.

7. Conclusion

This paper investigated the feasibility of the SUMO optimizer when used in experimental optimization of wireless solutions. In particular, a wireless conferencing scenario is considered. This paper also described the integration work of the SUMO optimizer in the iMinds w-iLab.t wireless testbed.

To compare the efficiency of SUMO optimized experiments, an exhaustively searched experiment is first conducted which leads to an accurate model of the problem to be optimized. However experiment repeatability needs to be guaranteed before comparison. To this end, identical experiments both with and without interference are conducted to validate this property. Moreover, experiments might become invalid due to external interference and an experiment outlier detection is applied to check validity of each conducted experiment.

SUMO is a powerful optimizer but a number of configurable parameters affect its efficiency. The sensitivity to initial sample size and the effect of stopping criteria are investigated in this paper. The initial sample size sensitivity exploits the exploration and exploitation balance of an optimization problem such that with few initial samples, an optimizer locates the optimum in a short period of time. Next, the Fixed Iteration (FI) and Objective Function Improvement (OFI) stopping criteria are considered. The FI stopping criterion is found not suited for our problem as it shows a wide variation in iteration count to reach the Global Maximum Combined Objective (GMCO). On the other hand, the OFI stopping criterion is well suited since it considers a relative difference in combined objective performance. Four sampling methods (Latin Hypercube Sampling, Random sampling, Orthogonal sampling and Hammersley Sequence Sampling) were combined with the SUMO toolbox to optimize the experiment until the OFI stopping criteria is met. The experiment is grouped into three scenarios. Scenario I realized the simplest wireless conferencing system where a speaker streamed one language to 8 listener nodes. Scenario II utilized the wireless spectrum intensely by increasing the language count from 1 to 8. And scenario III varied the network topology by increasing the listeners count from 8 to 16. In our proof of concept, the SUMO

experiment with LHS sampling method arrived the GMCO value at least 7.42 times faster than the exhaustive search experiments in all scenarios.

Acknowledgment

The research leading to these results has received funding from the European Union's Seventh Framework Programme FP7/2007-2013 under Grant agreement No. 258301 (CREW Project www.crew-project.eu) and 318273 (LEXNET Project www.lexnet-project.eu).

Eli De Poorter, Dirk Deschrijver and Ivo Couckuyt are postdoctoral research fellows of the Research Foundation Flanders (FWO-Vlaanderen). This work was also supported by the Interuniversity Attraction Poles Programme BESTCOM initiated by the Belgian Science Policy Office.

References

- [1] D. Cavin, Y. Sasson, A. Schiper, On the accuracy of manet simulators, in: Proceedings of the Second ACM International Workshop on Principles of Mobile Computing, POMC '02, ACM, New York, NY, USA, 2002, pp. 38–43.
- [2] K. Tan, D. Wu, A. Chan, P. Mohapatra, Comparing simulation tools and experimental testbeds for wireless mesh networks, in: World of Wireless Mobile and Multimedia Networks (WoWMoM), 2010 IEEE International Symposium on a, pp. 1–9.
- [3] T. Rakotoarivelo, G. Jourjon, M. Ott, Technical Report: Designing and Orchestrating Reproducible Experiments on Federated Networking Testbeds, Technical Report, NICTA, Sydney, Australia, 2012.
- [4] D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene, K. Crombecq, A surrogate modeling and adaptive sampling toolbox for computer based design, *J. Mach. Learn. Res.* 11 (2010) 2051–2055.
- [5] D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene, K. Crombecq, A surrogate modeling and adaptive sampling toolbox for computer based design, *J. Mach. Learn. Res.* 11 (2010) 2051–2055.
- [6] D. Jones, M. Schonlau, W. Welch, Efficient global optimization of expensive black-box functions, *Journal of Global Optimization* 13 (1998) 455–492.
- [7] C. E. S. ANIL K. GUPTA, KEN G. SMITH, The interplay between exploration and exploitation, *Academy of Management Journal* 49 (2006) 693–706.
- [8] G. Hawe, J. Sykulski, Balancing exploration exploitation using kriging surrogate models in electromagnetic design optimization, in: Electromagnetic Field Computation, 2006 12th Biennial IEEE Conference on, pp. 229–229.
- [9] M. Holland, T. Wang, B. Tavli, A. Seyedi, W. Heinzelman, Optimizing physical-layer parameters for wireless sensor networks, *ACM Trans. Sen. Netw.* 7 (2011) 28:1–28:20.
- [10] Y. Qu, S. Georgakopoulos, Relocation of wireless sensor network nodes using a genetic algorithm, in: Wireless and Microwave Technology Conference (WAMICON), 2011 IEEE 12th Annual, pp. 1–5.
- [11] R. Kulkarni, G. Venayagamoorthy, Particle swarm optimization in wireless-sensor networks: A brief survey, *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 41 (2011) 262–267.
- [12] X. Gao, Y. Gao, Tdma grouping based rfid network planning using hybrid differential evolution algorithm., in: F. L. Wang, H. Deng, Y. G. 0001, J. Lei (Eds.), AICI (2), volume 6320 of *Lecture Notes in Computer Science*, Springer, 2010, pp. 106–113.
- [13] K. Kaur, M. Rattan, M. S. Patterh, Optimization of cognitive radio system using simulated annealing, *Wirel. Pers. Commun.* 71 (2013) 1283–1296.
- [14] I. Couckuyt, F. Declercq, T. Dhaene, H. Rogier, L. Knockaert, Surrogate-based infill optimization applied to electromagnetic problems, *INTERNATIONAL JOURNAL OF RF AND MICROWAVE COMPUTER-AIDED ENGINEERING* 20 (2010) 492–501.

- [15] N. Hansen, A. Ostermeier, Completely derandomized self-adaptation in evolution strategies, *Evol. Comput.* 9 (2001) 159–195.
- [16] Q. Zhou, P. Z. G. Qian, S. Zhou, A simple approach to emulation for computer models with qualitative and quantitative factors, *Technometrics* 53 (2011) 266–273.
- [17] S. Bouckaert, P. Becue, B. Vermeulen, B. Jooris, I. Moerman, P. Demeester, Federating wired and wireless test facilities through emulab and omf: The ilab.t use case, in: T. Korakis, M. Zink, M. Ott (Eds.), *Testbeds and Research Infrastructure. Development of Networks and Communities*, volume 44 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, Springer Berlin Heidelberg, 2012, pp. 305–320.
- [18] I. telecommunication, Itu-t test signals for telecommunication systems, <http://www.itu.int/net/itu-t/sigdb/genaudio/AudioForm-g.aspx?val=1000050,????> Last accessed Feb 27, 2014.
- [19] NLANR/DAST, iperf: Tcp and udp bandwidth performance measurement tool, <https://code.google.com/p/iperf/>, 2009. Last accessed Feb 27, 2014.
- [20] N. K. Base, How is mos calculated in pingplotter pro?, <http://www.nessoft.com/kb/50>, 2005. Last accessed Feb 27, 2014.
- [21] D. Plets, W. Joseph, K. Vanhecke, L. Martens, Exposure optimization in indoor wireless networks by heuristic network planning, *Progress In Electromagnetics Research* 139 (2013) 445–478.
- [22] S. Keranidis, W. Liu, M. Mehari, P. Becue, S. Bouckaert, I. Moerman, T. Korakis, I. Koutsopoulos, L. Tassiulas, Concrete: A benchmarking framework to control and classify repeatable testbed experiments, in: *FIRE Engineering Workshop*, Abstracts.
- [23] S. Bouckaert, M. Mehari, W. Liu, I. Moerman, P. V. Wesemael, D. Finn, M. Chwalisz, J. Hauer, M. Doering, N. Michailow, D. Depierre, C. Heller, M. Smolnikar, Z. Padrah, M. Vucnik, Methodology for performance evaluation, www.crew-project.eu/sites/default/files/CREW_D4.2.IBBT_R.PU_2012-09-30_final.pdf, 2010. Section 3.3.
- [24] Wikipedia, Electromagnetic interference at 2.4 ghz, http://en.wikipedia.org/wiki/Electromagnetic_interference_at_2.4_GHz, 2008. Last accessed Feb 27, 2014.
- [25] W. Liu, S. Keranidis, M. Mehari, J. Vanhie-Van Gerwen, S. Bouckaert, O. Yaron, I. Moerman, Various detection techniques and platforms for monitoring interference condition in a wireless testbed, in: L. Fbrega, P. Vil, D. Careglio, D. Papadimitriou (Eds.), *Measurement Methodology and Tools*, volume 7586 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2013, pp. 43–60.
- [26] Z. Zhou, Two-phase IMSE-optimal Latin hypercube design for computer experiments, Ph.D. thesis, University of Southern California, 2006.
- [27] F. A. C. Viana, Things you wanted to know about the latin hypercube design and were afraid to ask, 10th World Congress on Structural and Multidisciplinary Optimization (2013).
- [28] W.-S. L. Tien-Tsin Wong, P.-A. Heng, Sampling with hammersley and halton points, *Journal of Graphics Tools* 2 (1997) 9–24.
- [29] P. H. Reisenthel, D. J. Lesieutre, Statistical benchmarking of surrogate-based and other optimization methods constrained by fixed computational budget, in: 51st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference 18th, 51st Structural Dynamics, and Materials Conference.