CONCRETE: A benchmarking framework to CONtrol and Classify REpeatable Testbed Experiments

Stratos Keranidis[⊕], Wei Liu[⋆], Michael Mehari[⋆], Pieter Becue[⋆], Stefan Bouckaert[⋆], Ingrid Moerman[⋆], Thanasis Korakis[⊕], Iordanis Koutsopoulos[⊕], and Leandros Tassiulas[⊕]

[+]Department of Computer Engineering and Telecommunications, University of Thessaly, Greece
[○]Centre for Research and Technology, Hellas
[⋆]Ghent University - IBBT, Gent, Belgium

Abstract. The global research community has been recently directed towards implementation approaches and evaluation through experimentation in networking testbeds. A unique problem that arises during experimentation in such environments is that several different factors may impact the monitored performance of networks under consideration. More specifically, the major factor that significantly impacts performance is the effect of interference that is generated either from testbed nodes that simultaneously transmit or from collocated commercial devices belonging to external networks. Moreover, as the level of interference significantly varies over time, the monitored performance may present high variation among several executions of the same experiment. In order to ensure stable experimental conditions, specific tools have to be developed that would be able to detect whether the gathered results have been affected by external factors. In this work, we present a novel benchmarking framework, developed within the collaboration of CREW and OpenLab EU FP7 projects. The developed framework provides for proper evaluation of experimentally gathered results and thus enables the experimenter to arrive at solid conclusions regarding the validity of the conducted experiments.

1 INTRODUCTION

The tremendous growth of 802.11 WLANs, has resulted in the congestion of the limited unlicensed spectrum, especially in densely populated urban areas. In order to perform research on such complex networks, the global research community has been directed towards implementation approaches and evaluation through experimentation in real world network scale and settings. To this aim, several networking testbeds have been recently deployed that are composed of both prototype, as well as off-the-self commercial devices.

Testbed experimentation introduces several issues that arise due to several different factors that may impact the monitored performance of networks under consideration. Interference generated by testbed nodes or by neighboring external networks may significantly impact monitored performance. In order to ensure validity and stability of experimental conditions, specific monitoring tools should be running in parallel with the experiment execution to ensure that the gathered results have not been affected by external factors.

In this work, we present a novel benchmarking framework, developed within the collaboration of CREW [1] and OpenLab [2] EU FP7 projects. The developed framework is based on the performance comparison of multiple executions of the experiment under consideration. We illustrate the usage of the proposed framework in section 2, while in section 3 we point out conclusions and directions for future work.

Fig. 1: Visualization support

2 The CONCRETE Framework

The developed framework enables the experimenter to run several runs of the specified experiment and analyze the corresponding results. The main functionalities that are currently supported, are:

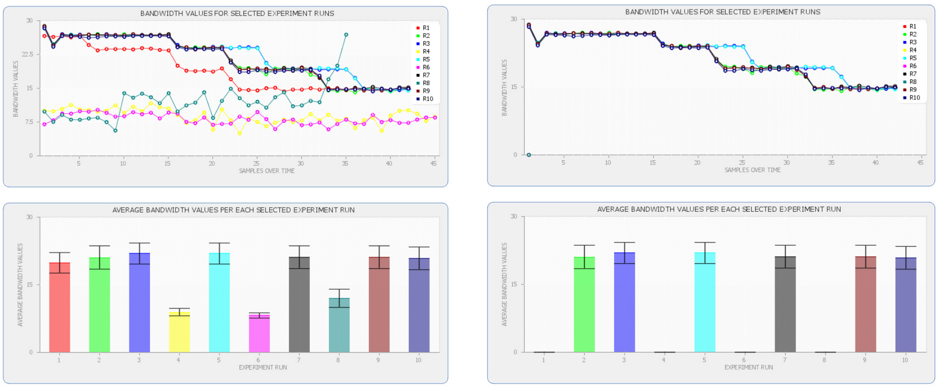
- scheduling the execution of several runs for the same experiment
- visualization of prevailing channel conditions before each run and moreover visualization of the performance achieved in each run
- visualization of performance over all executed rounds
- estimation of correlation among the different runs, as the benchmarking score that describes the stability of each run, accompanied by an automatic mechanism that selects the most stable runs, based on the calculated score
- comparison of performance obtained over all executed rounds in comparison with the performance achieved only in the subset of selected rounds.

Below, each different functionality is briefly detailed and is also accompanied by some representative screenshots.

As for the first step, the experimenter uses a custom Graphical User Interface (GUI) to provide all input required to describe the experiment to be executed. The framework proceeds with the execution of multiple experiment runs and also instruments the measurement gathering and storage procedures, through the use of the OMF (cOn- trol and Management Framework) [3].

As for the second step, the framework accesses data stored during the execution and generates screenshots that illustrate both performance in each different round as well as interference conditions, before each discrete run. The interference monitoring is based on energy detection applied through distributed spectrum sensing, by utilizing USRP devices. Based on the use of different specified metrics (packet loss, ping delay, RSSI levels), we can get different plots for each specified experiment. In the left part of Fig. 1, we can see a representative plot of throughput measurements over 3 rounds, while on the right part, we can see a representation of the channel conditions over time, in terms of signal strength in dBm values.

In the next step, the tool constructs a joint performance representation plot over all runs, which also presents average and standard deviation values for each different run. Fig. 2(a) illustrates a representative screenshot.



(a) Performance plot over all executed rounds (b) Performance plot over stable rounds

Fig. 2: Representative screenshots

Apart from providing simple statistics that result in an overall comparison over the various executions, an automated process for detecting whether the experiment is affected by external factors has also been implemented. The effect of external factors, such as interference, constitute the temporal correlation between each consecutive experimental run. Thus, the degree of correlation between multiple executions of the same experiment may vary according to the temporal variation of factors that impact the testing environment. Based on the estimated degree of temporal correlation among the multiple executions of the same experiment, the tool is able to provide a list with the runs that are considered stable. Fig. 2(b) illustrates a representative screenshot, where only performance of stable rounds is plotted.

Based on the final set of selected rounds, a final average performance and standard deviation is calculated across the subset of selected rounds. Series of different experiments conducted in both w-ilab.t [4] and NITOS [5] testbeds have verified the great performance difference between cases where all runs are taken into consideration in comparison solely with stable runs.

3 CONCLUSIONS AND FUTURE WORK

CONCRETE framework provides for proper evaluation of results obtained during testbed experiments, by excluding experiment runs that have been affected by external factors. The consideration of Cross-Correlation calculations as the experiment validity score, the distributed energy detection before each run and the use of the OMF framework to instrument experiment execution are the main building blocks of the developed framework. In future, we plan on advancing the interference estimation procedure, by incorporating measurements gathered during the experiment execution by nodes set in monitor mode and also by applying sophisticated feature detection techniques.

References

1. *"Cognitive Radio Experimentation World (CREW) project"*, http://www.crew- project.eu/.
2. *"OpenLab project"*, http://www.ict-openlab.eu/.
3. T. Rakotoarivelo, M. Ott, G. Jourjon, and I. Seskar. OMF: A Control and Management Framework for Networking Testbeds. *SIGOPS Oper. Syst. Rev.*, 43(4):54 –59, 2010.
4. *"w-ilab.t Wireless Testbed"*, http://www.crew-project.eu/wilabt.
5. *"NITOS Wireless Testbed"*, http://nitlab.inf.uth.gr/NITlab/index.php/testbed.