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*

University of Thessaly and CERTH, Greece*
Ghent University - iMinds, Gent, Belgium

Introduction

- The Problem:
  - During experimentation in networking testbeds several different factors may impact the monitored performance of networks under consideration.
  - As a result high variation exists among several executions of the same experiment.

- The Need:

  Stable experimental conditions have to be guaranteed, in order to arrive at solid conclusions.

- Our Solution:

  The novel CONRETE benchmarking framework that provides for evaluation of experimental stability.

Outline

- Related Projects
- Basic Experimental Scenario
- Interfering Factors
- Building Blocks
- CONCRETE Benchmarking Framework
- Insights and Future work

Related Projects

- CREW
  - Establishes an open federated test platform, which facilitates experimentally-driven research on advanced spectrum sensing, cognitive radio and cognitive networking strategies in view of horizontal and vertical spectrum sharing in licensed and unlicensed bands.

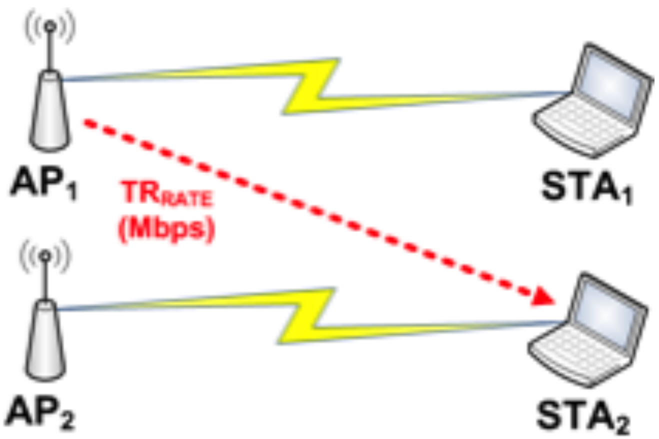
- OPENLAB
  - Delivers control and experimental plane middleware to facilitate early use of testbeds and exploiting proven technologies, developed in the OneLab and Panlab initiatives.

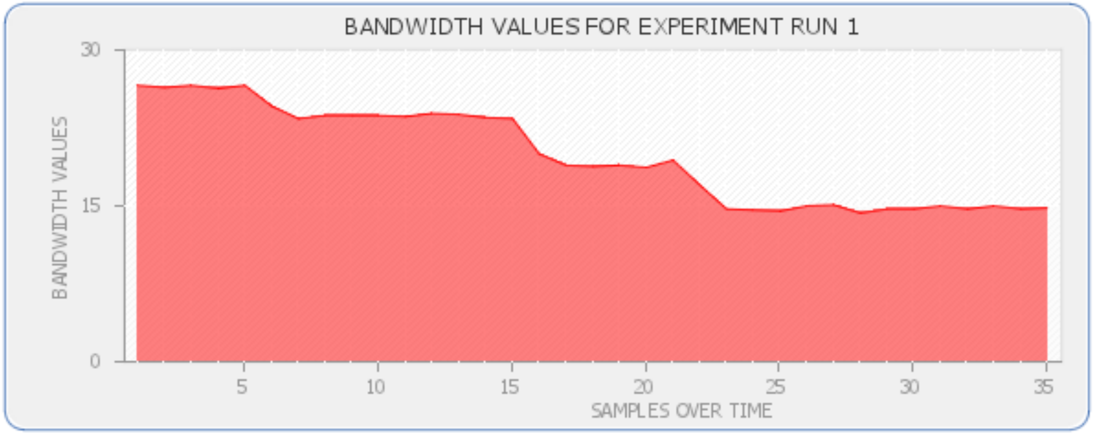
- OPENLAB – CREW Collaboration
  - In order to improve the reproducibility of wireless experiments, OpenLab is interested to augment the OpenLab facilities with the CREW spectrum sensing benchmarking scenarios.

Outline

- Related Projects
- Basic Experimental Scenario
- Interfering Factors
- Building Blocks
- CONCRETE Benchmarking Framework
- Insights and Future work

Basic Experimental Scenario





- 2 pairs of nodes contending for channel use.

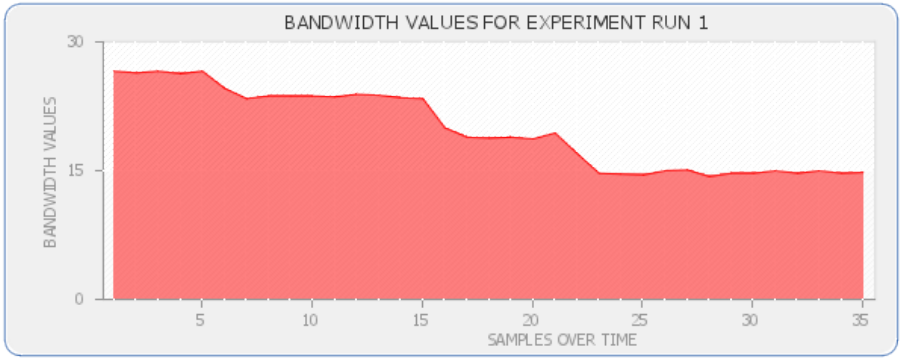
- **AP2 - > STA2**: saturated traffic conditions
- **AP1 -> STA1**: varying traffic rate (**TR_{RATE}**) conditions

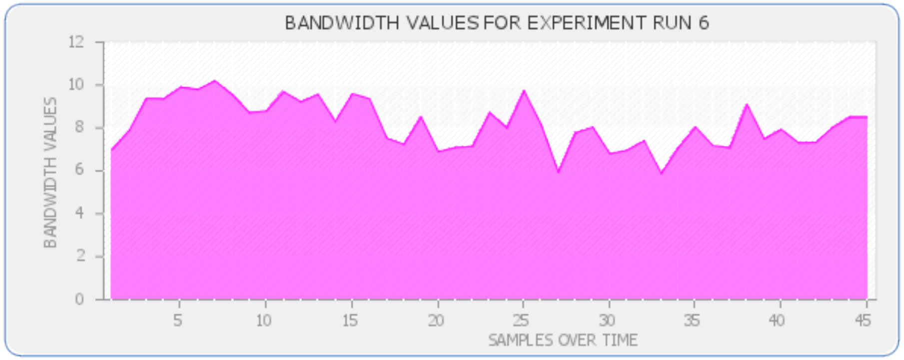
- We monitor the throughput performance of the **AP2-STA2** pair

Outline

- Related Projects
- Basic Experimental Scenario
- Interfering Factors
- Building Blocks
- CONCRETE Benchmarking Framework
- Insights and Future work

Interfering Factors (1/5)



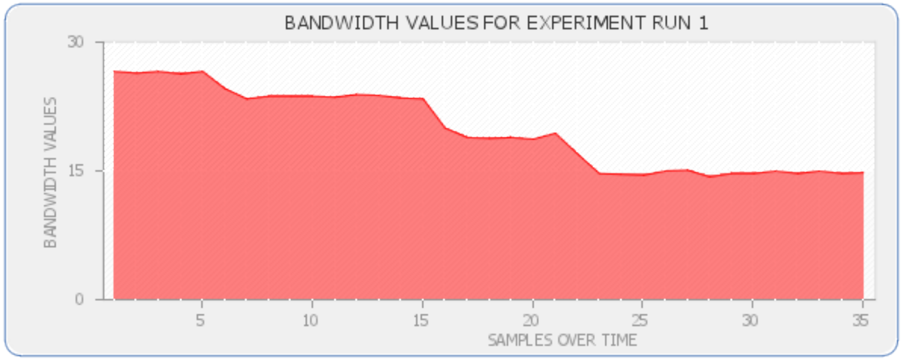


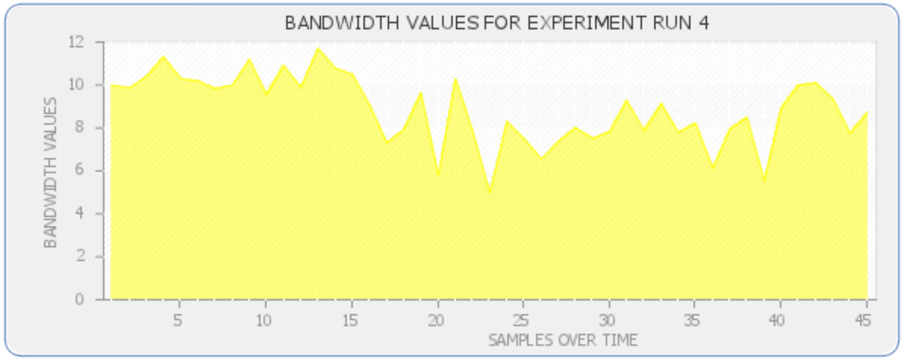
Specific executions of the same experiment may present different performance, due to:

Internal Interference

generated by testbed nodes, operated by other experimenters, which simultaneously transmit on the same or overlapping frequencies.

Interfering Factors (2/5)



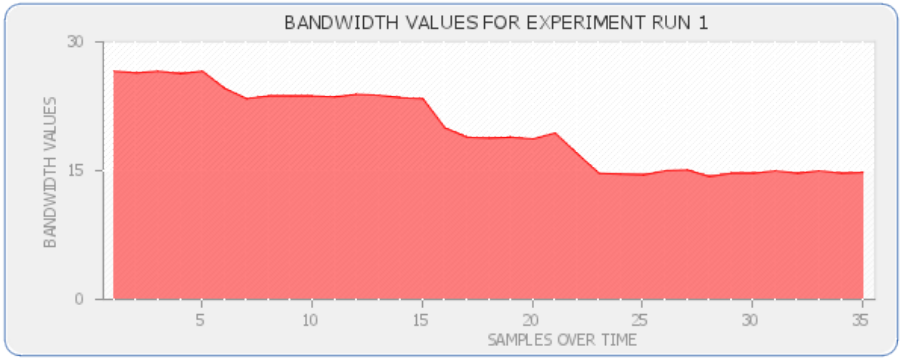


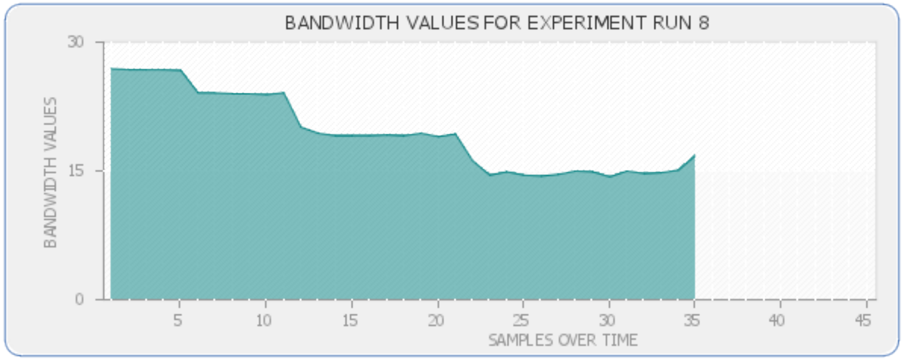
Specific executions of the same experiment may present different performance, due to:

External Interference

generated by collocated commercial devices belonging to external networks, which simultaneously transmit on the same or overlapping frequencies.

Interfering Factors (3/5)

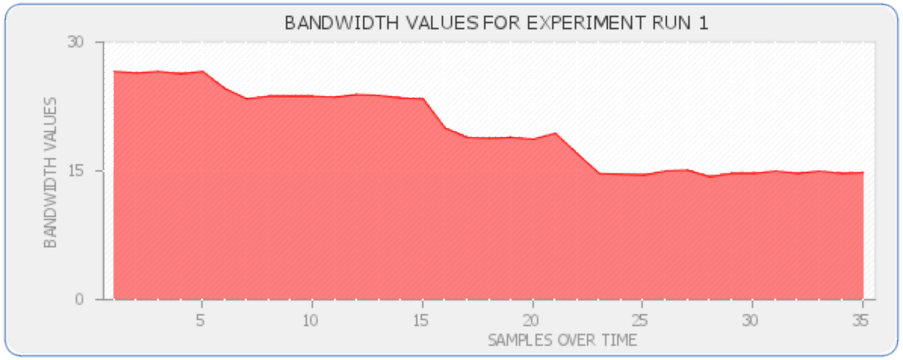


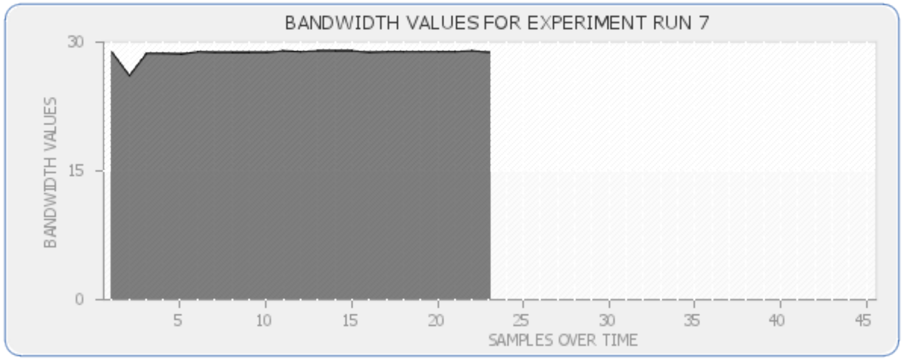


Specific executions of the same experiment may present different performance, due to various factors, such as:

stopping of normal execution due to hardware or software failure

Interfering Factors (4/5)



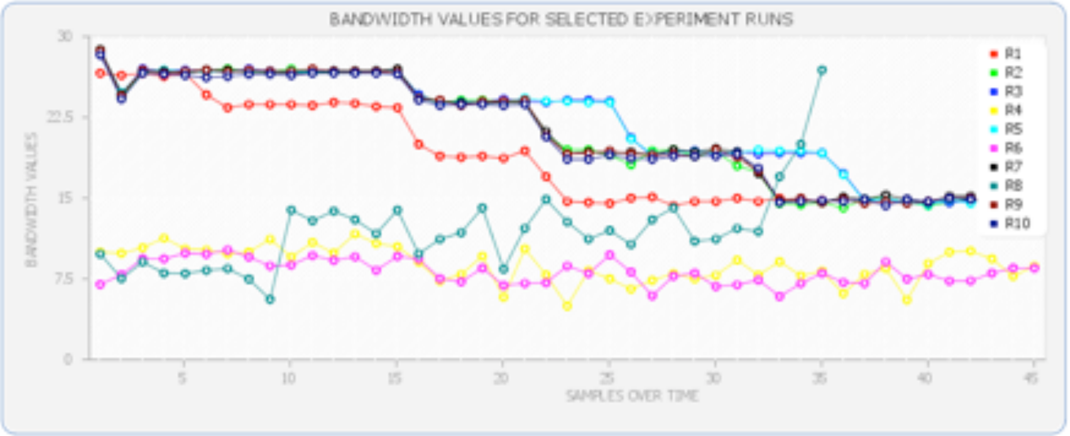


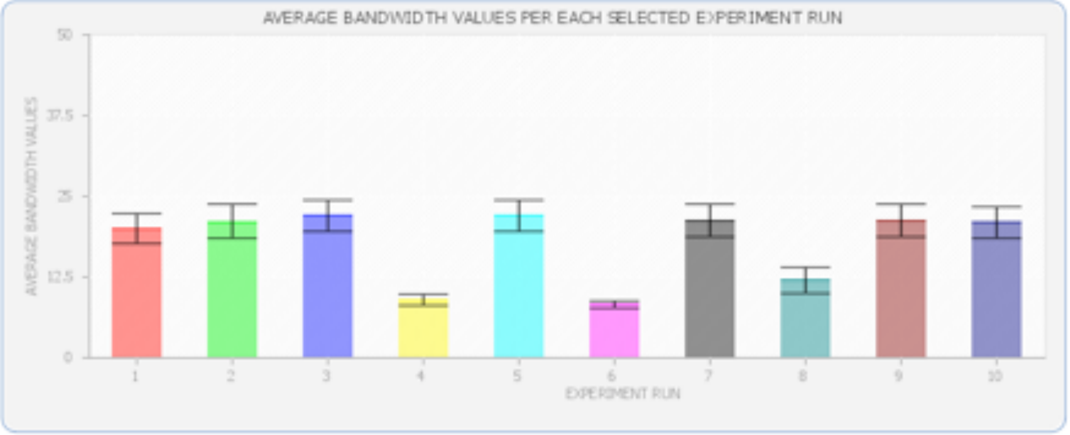
Specific executions of the same experiment may present different performance, due to various factors, such as:

Different node positioning
(etc. mobile nodes behind obstacles)

Interfering Factors (5/5)

The Result





Outline

- Related Projects
- Basic Experimental Scenario
- Interfering Factors
- Building Blocks
- CONCRETE Benchmarking Framework
- Insights and Future work

Building Blocks

Advanced Spectrum
Sensing Techniques

OMF Control and
Management Framework

CONCRETE

iMinds w-ilab.t
Cognitive Testbed

Long experience with
instrumentation of
testbed experiments

Building Blocks – Correlation

$$\rho_{X,Y} = corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

$$r = \frac{1}{(n-1)} \sum \frac{(X-\mu_X)(Y-\mu_Y)}{\sigma_X \sigma_Y}$$

- The well known measure of dependence is Pearson's correlation, which indicates the extent to which two random variables covary.
- The μx and μy represents the mean of the data set X and Y respectively.
- The σx and σy represents the standard deviation of the data set X and Y respectively

Outline

- Related Projects
- Basic Experimental Scenario
- Interfering Factors
- Building Blocks
- CONCRETE Benchmarking Framework
- Insights and Future work

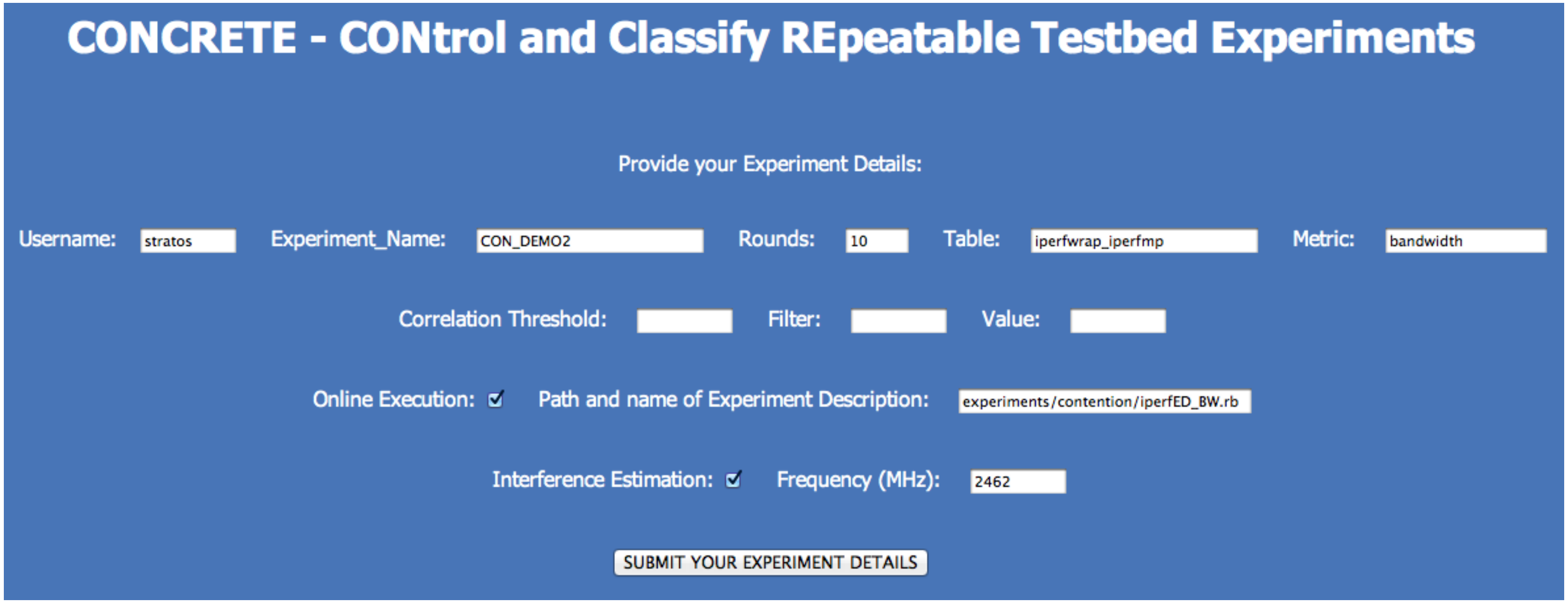
CONCRETE Benchmarking Framework
CONtrol and Classify REpeatable Testbed Experiments

The 6 main functionalities that are currently supported, are:

1. Scheduling the execution of several runs for the same experiment
2. Visualization of prevailing Channel Conditions before each run and moreover visualization of the Performance achieved in each run
3. Estimation of Correlation among the different runs, in order to provide an appropriate benchmarking score that describes the stability of each run
4. Calculation of average performance and st. deviation values for each run
5. Automatic mechanism that selects the most stable runs, based on their correlation score
6. Calculation of performance over all executed rounds in comparison with the performance achieved only in the subset of selected rounds.

CONCRETE Benchmarking Framework (1/6)

1. Scheduling the execution of several runs for the same experiment



CONCRETE Benchmarking Framework (2/6)

2. Visualization of Channel Conditions before each run and moreover visualization of the Performance achieved in each run

CONCRETE Benchmarking Framework (3/6)

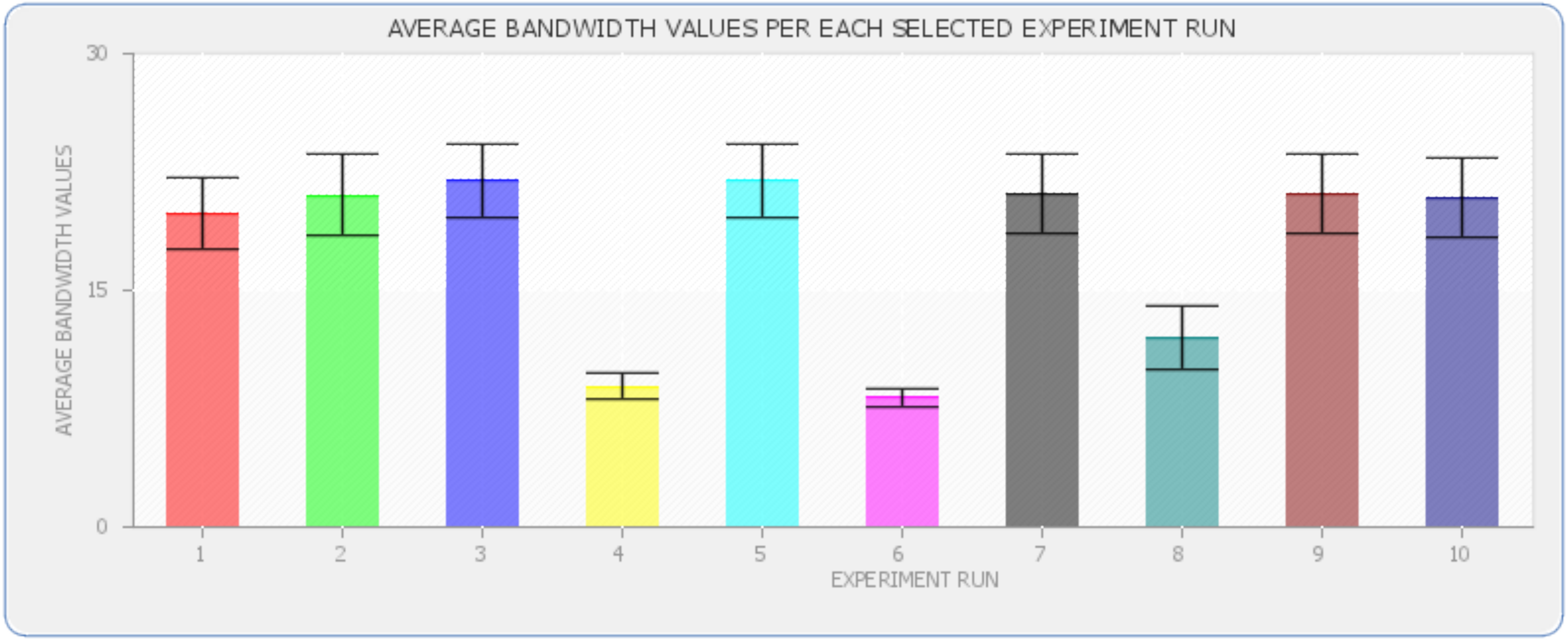
3. Estimation of Correlation among the different runs

CORRELATION MATRIX

ID	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
R1	1	0.9076	0.8738	0.7706	0.8802	0.5732	0.9127	-0.5042	0.9151	0.9178
R2	0.9076	1	0.9191	0.6533	0.9214	0.5852	0.9966	-0.6115	0.9973	0.9959
R3	0.8738	0.9191	1	0.6086	0.9987	0.6736	0.9127	-0.5077	0.9189	0.9131
R4	0.7706	0.6533	0.6086	1	0.6214	0.4738	0.6661	-0.2388	0.6673	0.6799
R5	0.8802	0.9214	0.9987	0.6214	1	0.6591	0.9148	-0.5007	0.9212	0.9154
R6	0.5732	0.5852	0.6736	0.4738	0.6591	1	0.5808	-0.308	0.5888	0.5869
R7	0.9127	0.9966	0.9127	0.6661	0.9148	0.5808	1	-0.6226	0.9987	0.9982
R8	-0.5042	-0.6115	-0.5077	-0.2388	-0.5007	-0.308	-0.6226	1	-0.6177	-0.6071
R9	0.9151	0.9973	0.9189	0.6673	0.9212	0.5888	0.9987	-0.6177	1	0.9986
R10	0.9178	0.9959	0.9131	0.6799	0.9154	0.5869	0.9982	-0.6071	0.9986	1

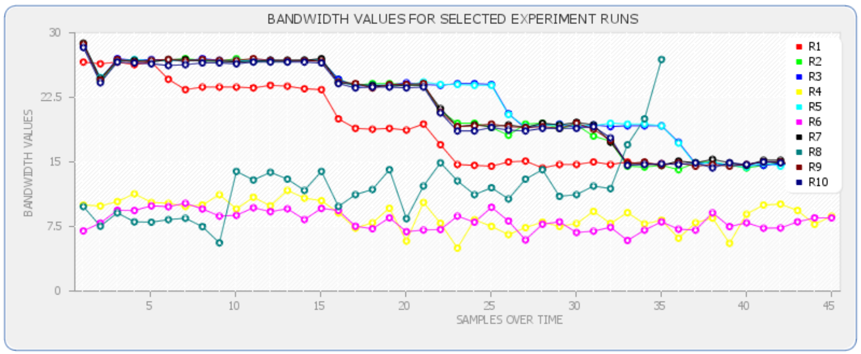
CONCRETE Benchmarking Framework (4/6)

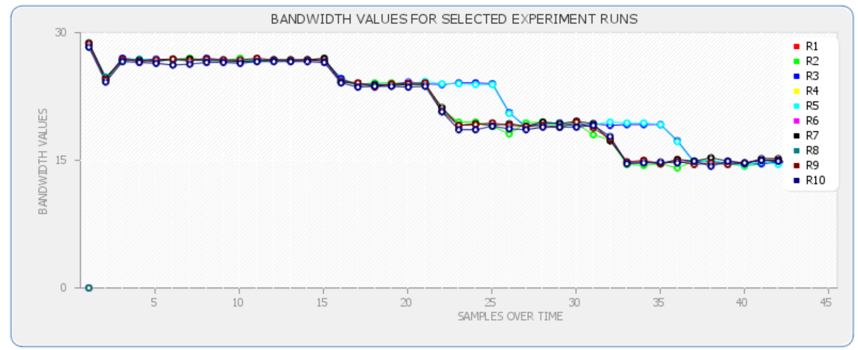
4. Calculation of AVG performance and ST. DEV. for each run

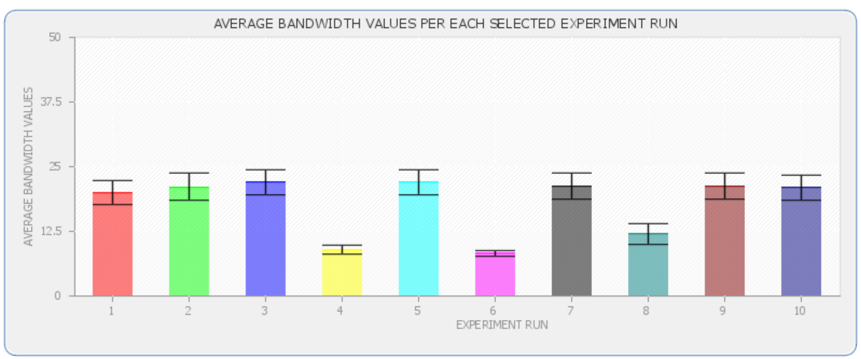


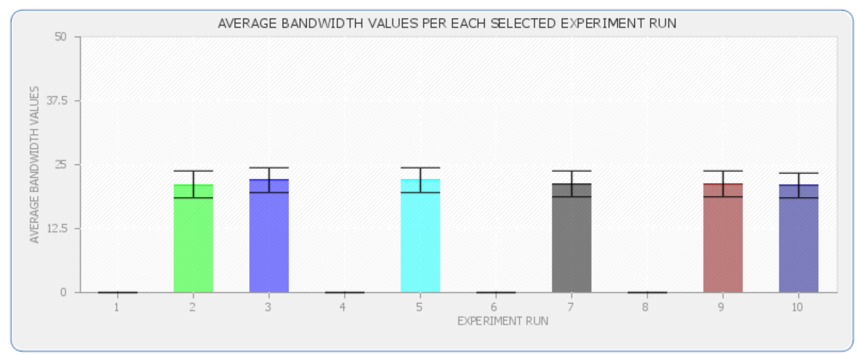
CONCRETE Benchmarking Framework (5/6)

5. Automatic mechanism that selects the most stable runs, based on their correlation score









CONCRETE Benchmarking Framework (6/6)

6. Calculation of performance over all executed rounds in comparison with the performance achieved only in the subset of selected rounds.

AVERAGE AMONG ALL SUCCESFULL RUNS: 17.656495238095
STDEV AMONG ALL SUCCESFULL RUNS: 4.0677154655928

CALCULATE BASED ON SET OF SELECTED ROUNDS

YOU SELECTED 6 Round(s): 2 3 5 7 9 10

AVERAGE AMONG SELECTED RUNS: 21.301111111111
STDEV AMONG SELECTED RUNS: 4.9083900551418

CORRELATION MATRIX

ID	R2	R3	R5	R7	R9	R10
R2	1	0.9426	0.9422	0.9977	0.998	0.9971
R3	0.9426	1	0.9991	0.9415	0.9444	0.9412
R5	0.9422	0.9991	1	0.9409	0.9442	0.9408
R7	0.9977	0.9415	0.9409	1	0.9987	0.9984
R9	0.998	0.9444	0.9442	0.9987	1	0.9988
R10	0.9971	0.9412	0.9408	0.9984	0.9988	1

MINIMUM CORRELATION AMONG ALL SUCCESFULL RUNS: -0.6226
MINIMUM CORRELATION AMONG SELECTED RUNS: 0.9408

Outline

- Related Projects
- Basic Experimental Scenario
- Interfering Factors
- Building Blocks
- CONCRETE Benchmarking Framework
- Insights and Future work

Insights and Future Work

- Experimental Insights:
  - Due to the high variation of wireless channel conditions there is a clear need for environment monitoring mechanisms
  - that aid in arriving at CONCRETE conclusions.

- Future Work:
  - Enable channel monitoring during the experiment execution through Wi-Fi Monitor nodes.
  - Implement Feature detection mechanism to enable detection of transmissions generated by devices using heterogeneous technologies
  - Examine performance under various experiments and metrics (energy etc.) and propose possible enhancements

Thank You!