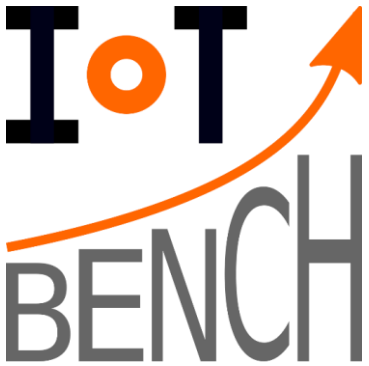
IoTBench - Past, present, and future of a community-driven benchmarking initiative



Romain Jacob and Markus Schuß

on behalf of IoTBench

CPS-IoTBench Workshop

April 15, 2019

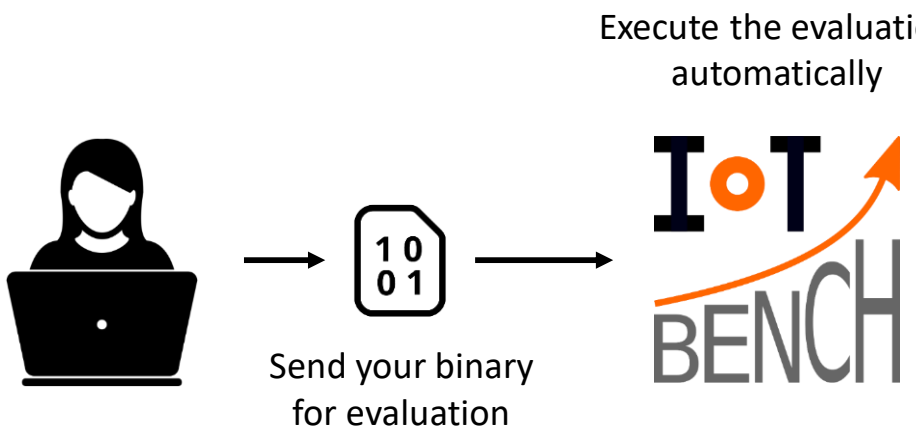
How does your algorithm compare to the state-of-the-art?
method
< insert your research object here >

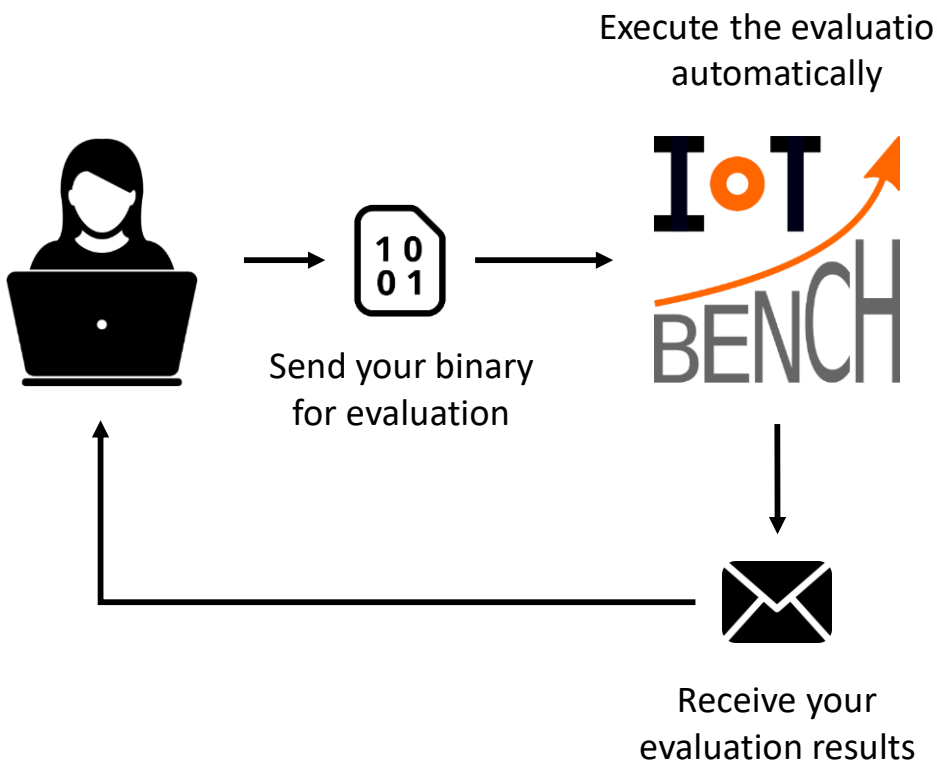
How does your algorithm compare to the state-of-the-art?
method
< insert your research object here >
low-power wireless protocol

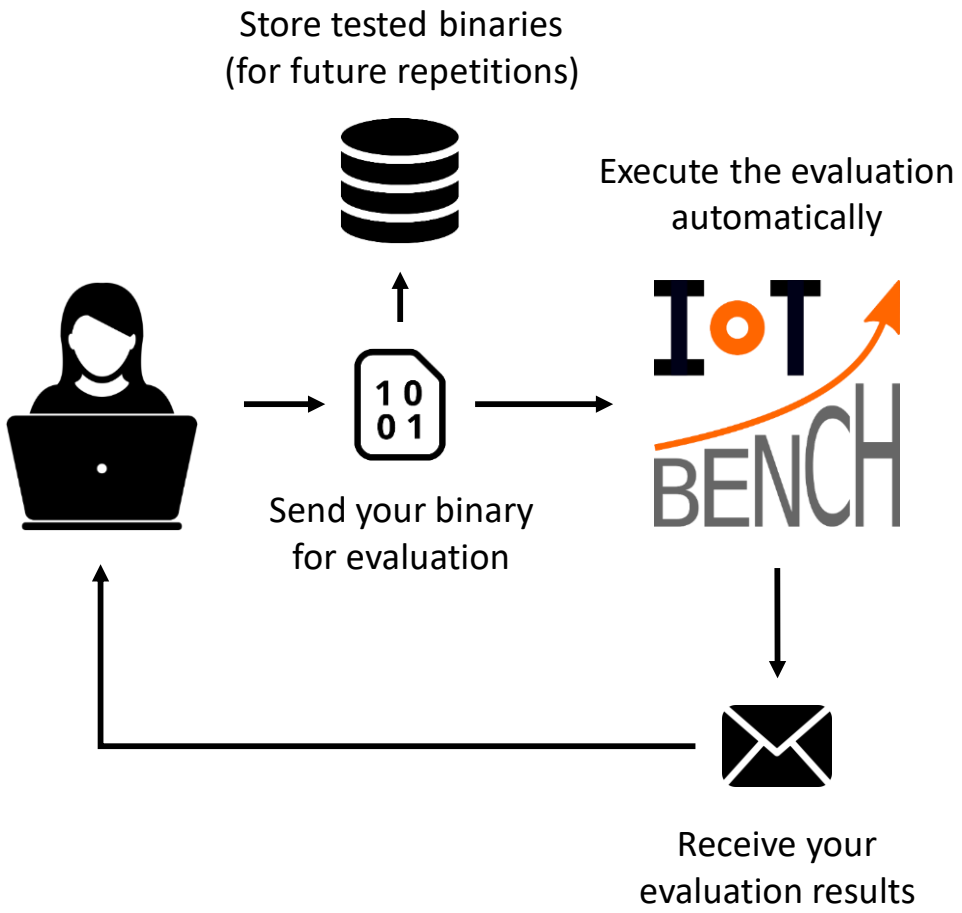
“ We need a benchmark for IoT networking. ”

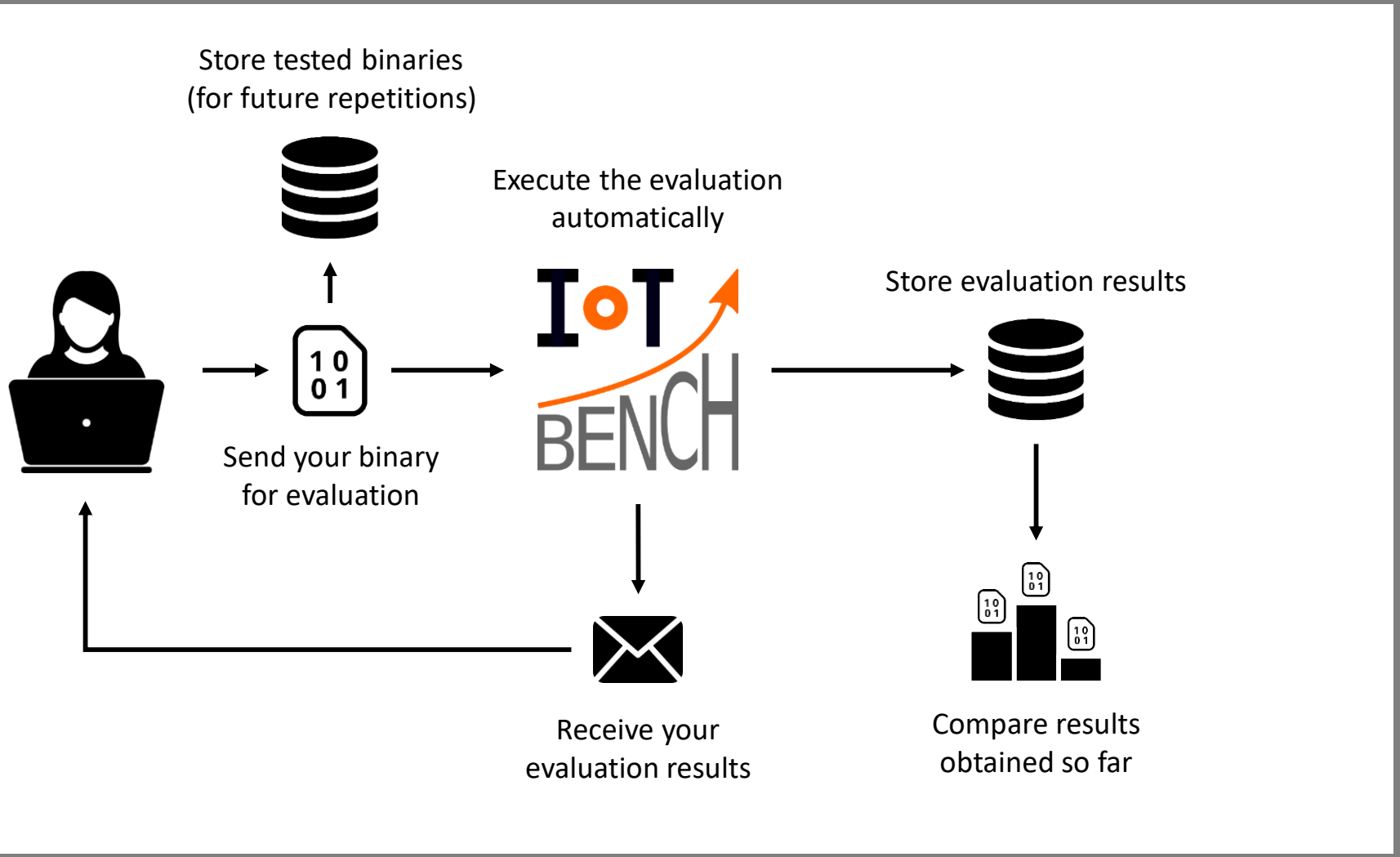
The Vision

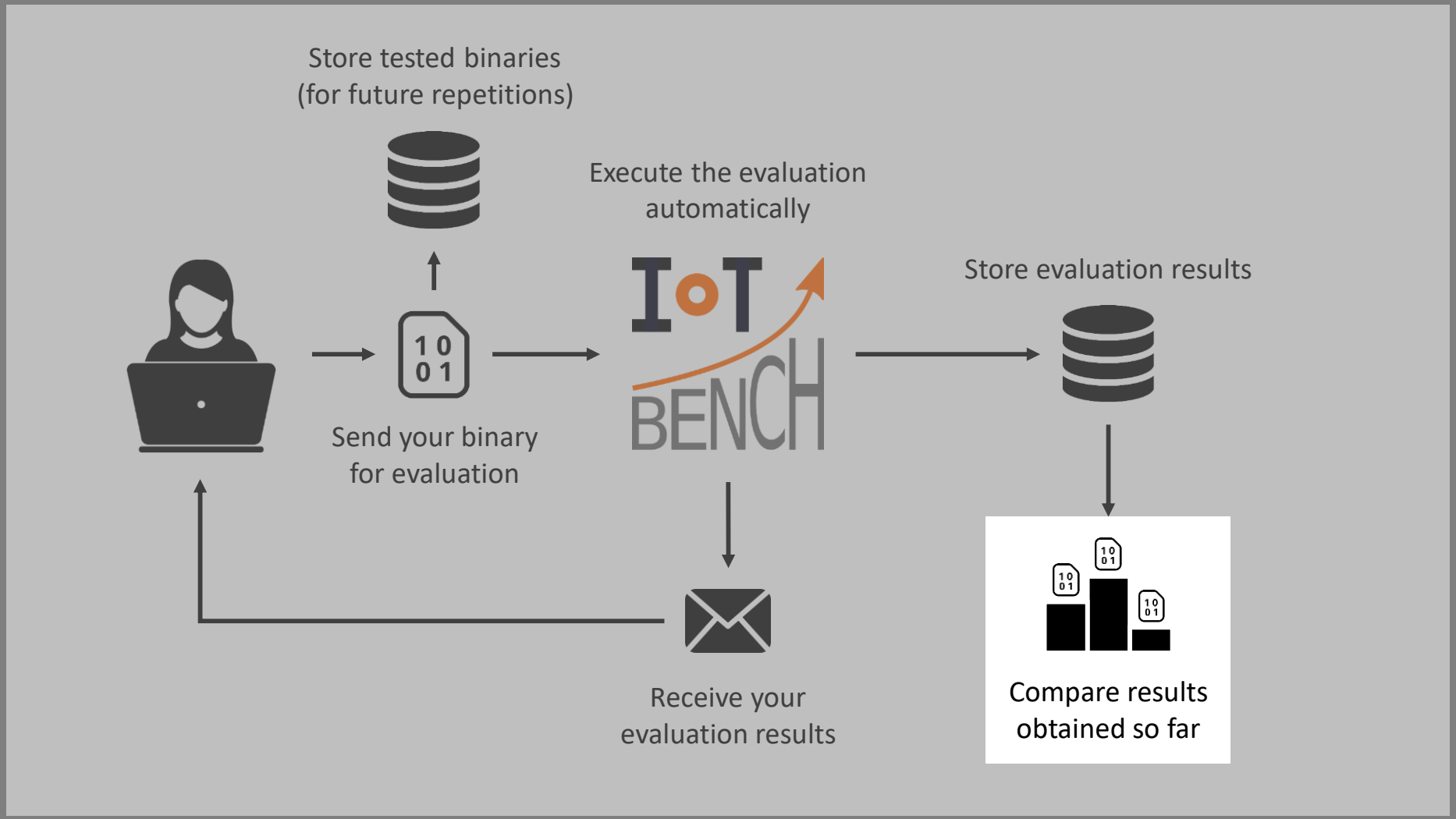
" Okay, this protocol is really cool.
Let's see how well it perform... "







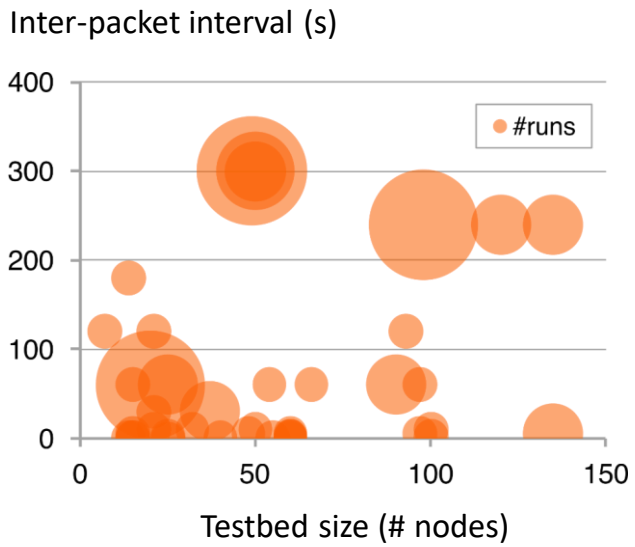




The Problem(s)

Why is it so difficult to compare (low-power) wireless protocols?

Many different test settings



Periodic data collection only

Why is it so difficult to compare (low-power) wireless protocols?

Many different test settings

Experiments are not reproducible
not comparable

Only 16,5% wireless networking papers provide enough info to only attempt to reproduce the results [1]

[1] G. Z. Papadopoulos et al., "Performance evaluation methods in ad hoc and wireless sensor networks: a literature study," IEEE Communications Magazine, Jan. 2016.

Why is it so difficult to compare (low-power) wireless protocols?

Many different test settings

Experiments are not reproducible
not comparable

No reference results available

How to address these challenges?

Many different test settings

Experiments are not reproducible
not comparable

No reference results available

How to address these challenges?

Many different test settings

Experiments are not reproducible
not comparable

No reference results available

Formalized test configurations

How to address these challenges?

Many different test settings

Experiments are not reproducible
not comparable

No reference results available

Formalized test configurations

Experimental methodology

Definition of repeatability

Comparison methodology

How to address these challenges?

Many different test settings

Experiments are not reproducible
not comparable

No reference results available

Formalized test configurations

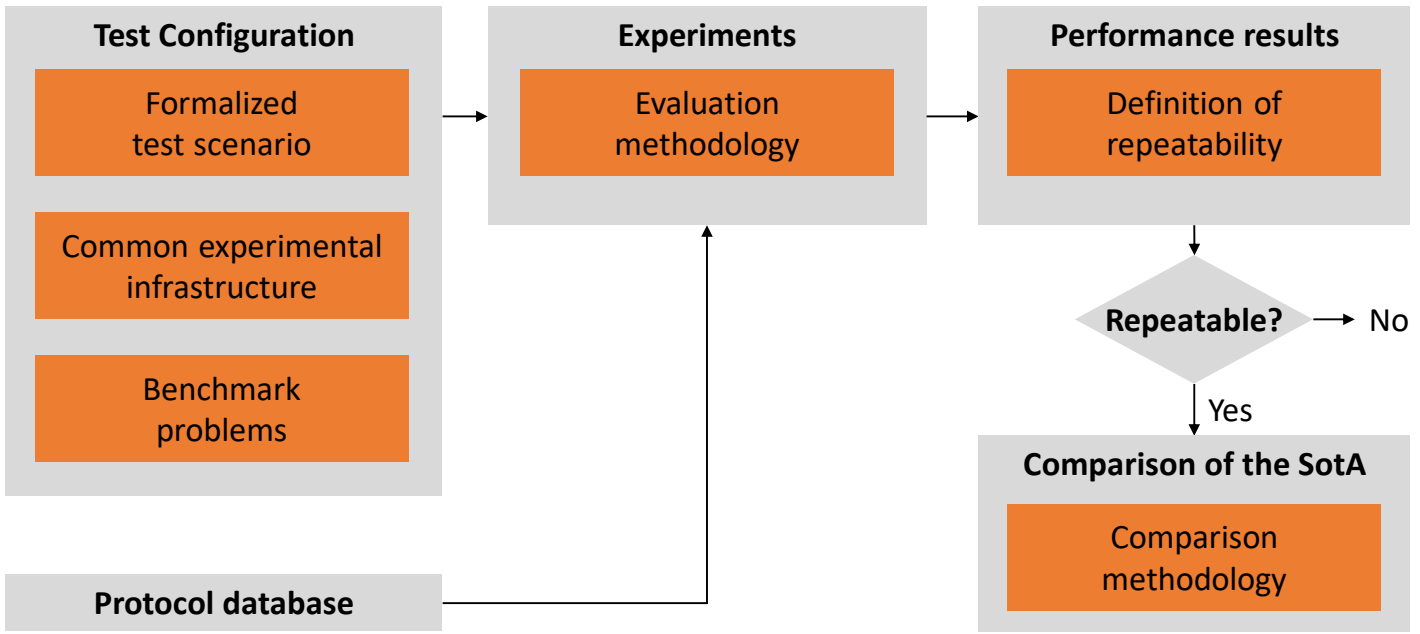
Experimental methodology

Definition of repeatability

Comparison methodology

Benchmark problems

Common experimental infrastructure



Getting there

The EWSN Dependability Competition

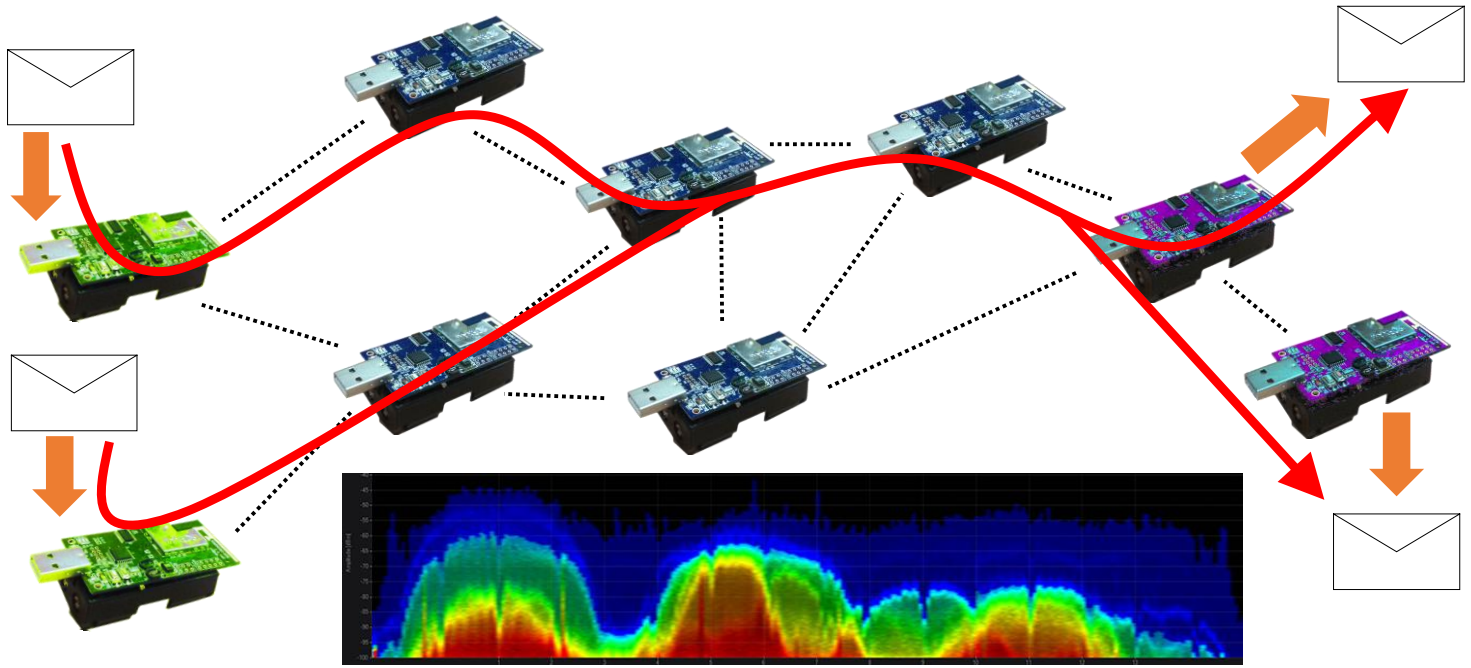
EWSN Dependability Competition

Goal: quantitatively compare the performance of low-power wireless systems

... and many others!

EWSN Dependability Competition

Step 1: define a common test scenario



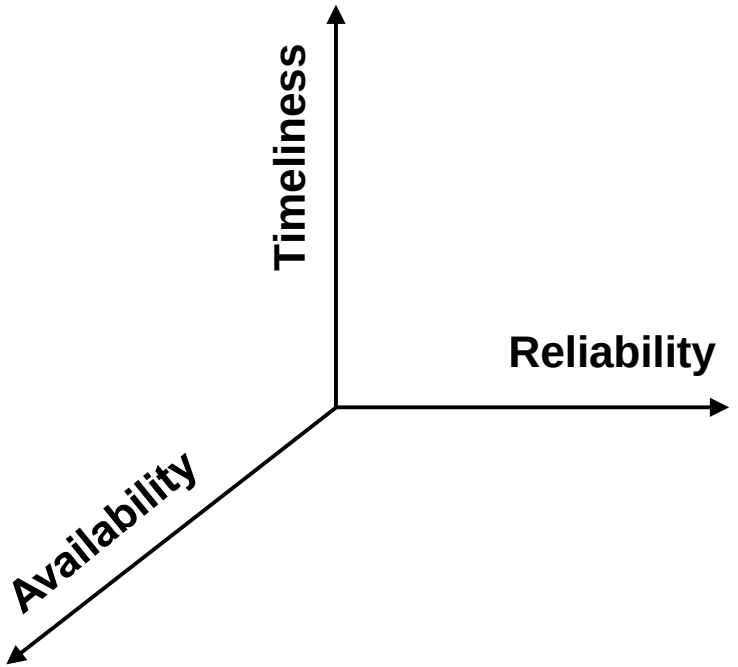
EWSN Dependability Competition

Step 2: define performance metrics to enable comparison

Do solutions allow a reliable, timely, and energy-efficient communication?

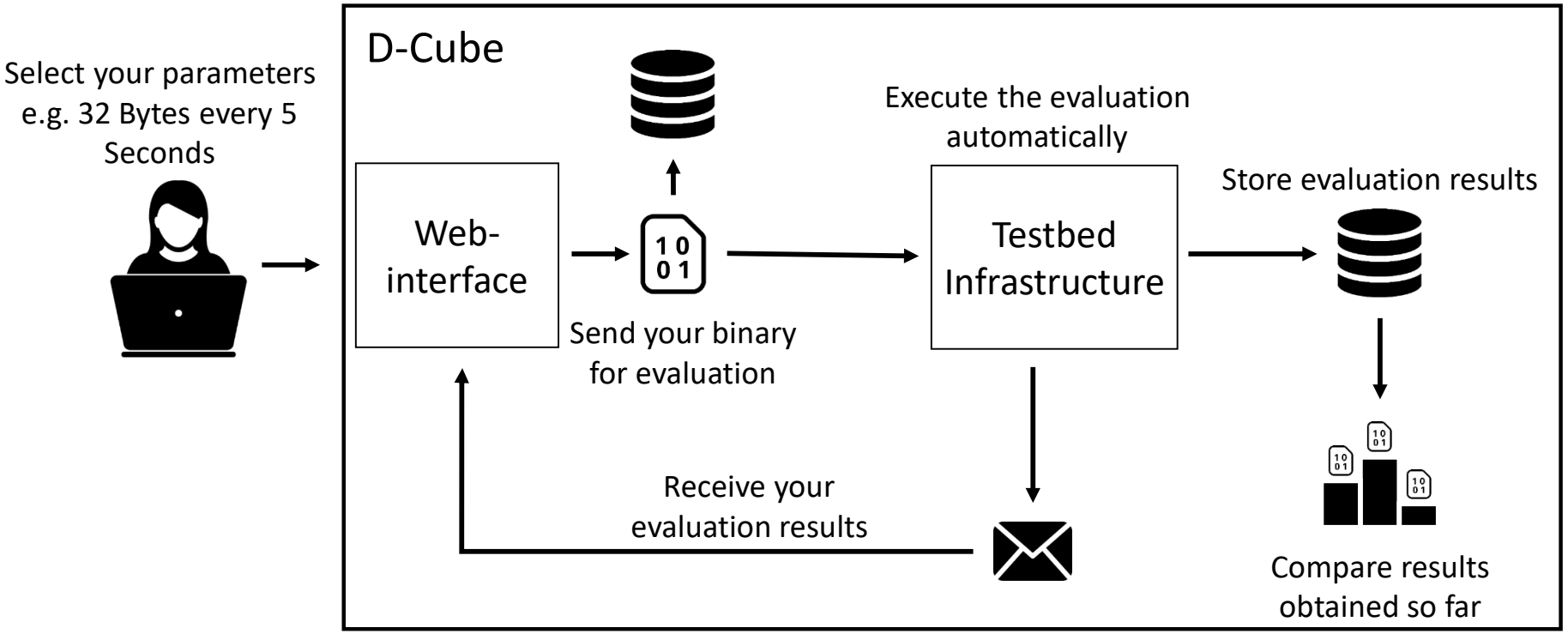
Three evaluation metrics

- Number of messages delivered correctly
- End-to-end latency
- Total power consumption of all nodes



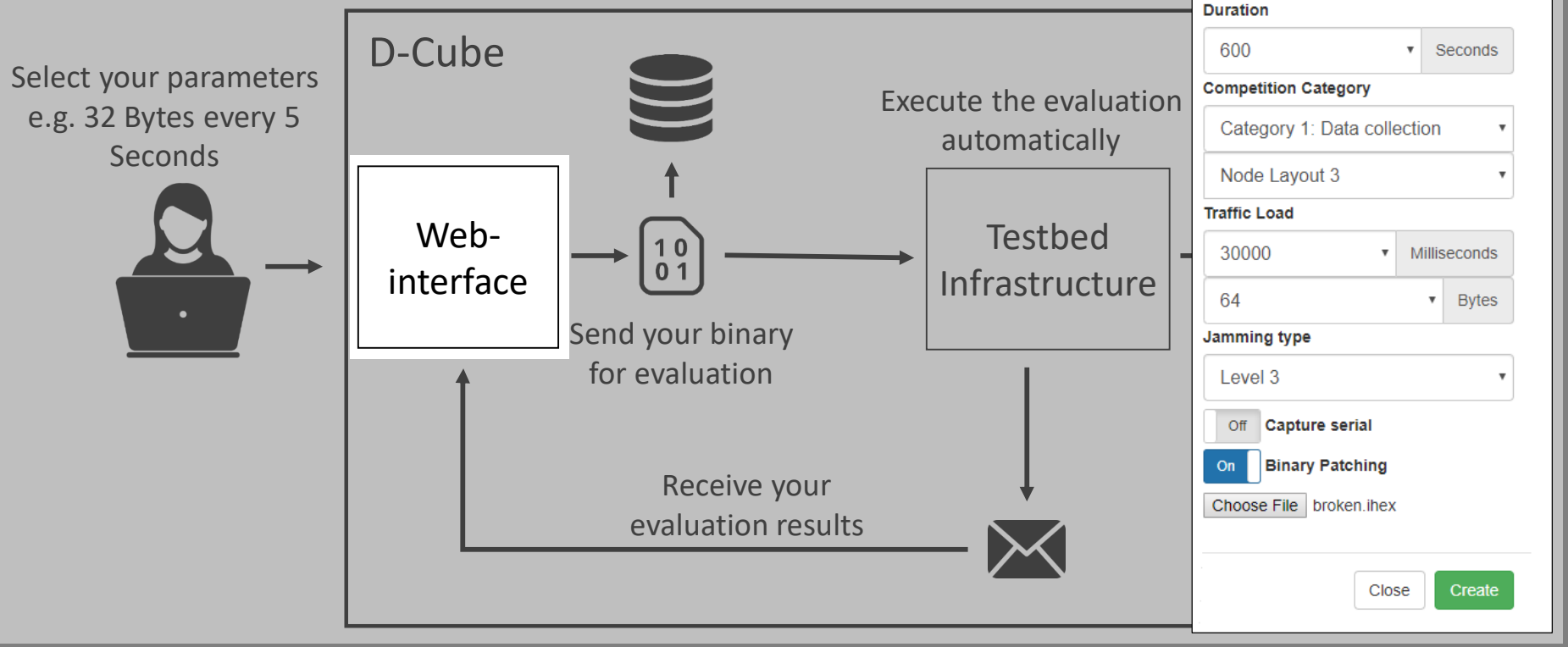
EWSN Dependability Competition

Step 3: define the benchmark problem(s)



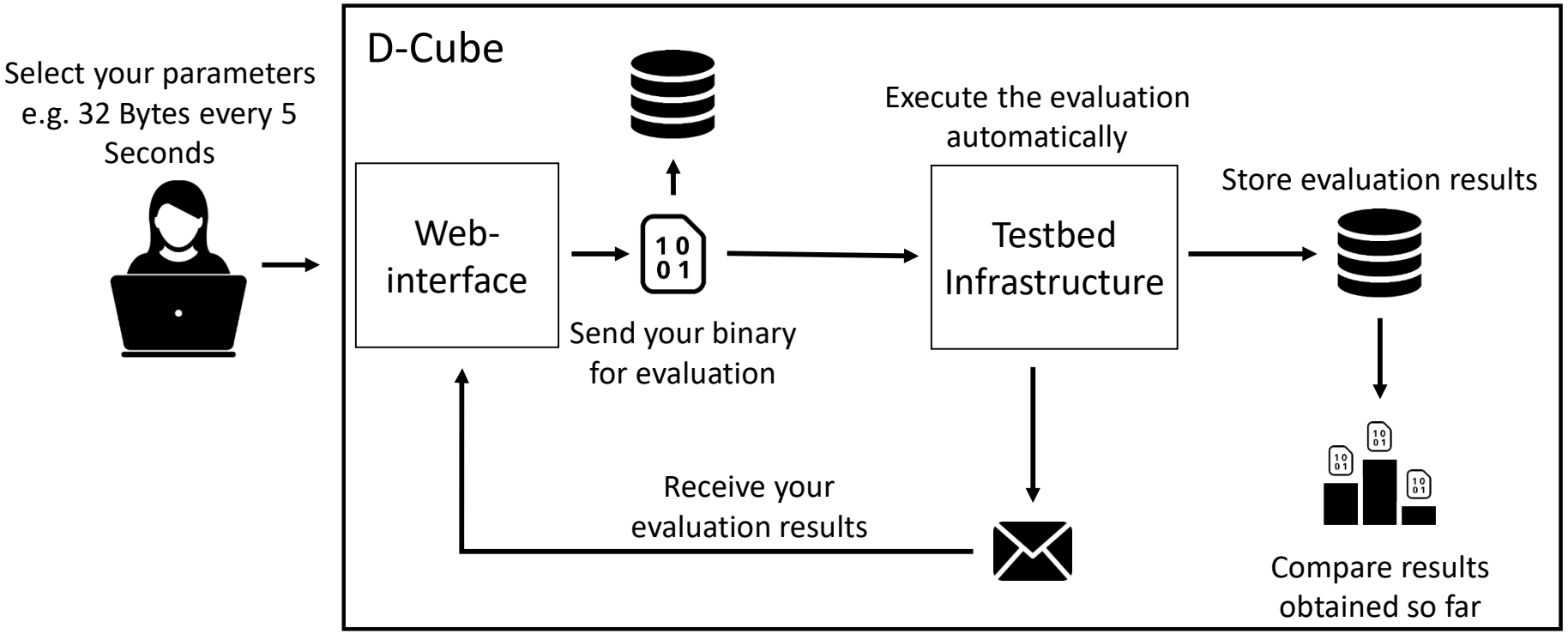
EWSN Dependability Competition

Step 3: define the benchmark problem(s)



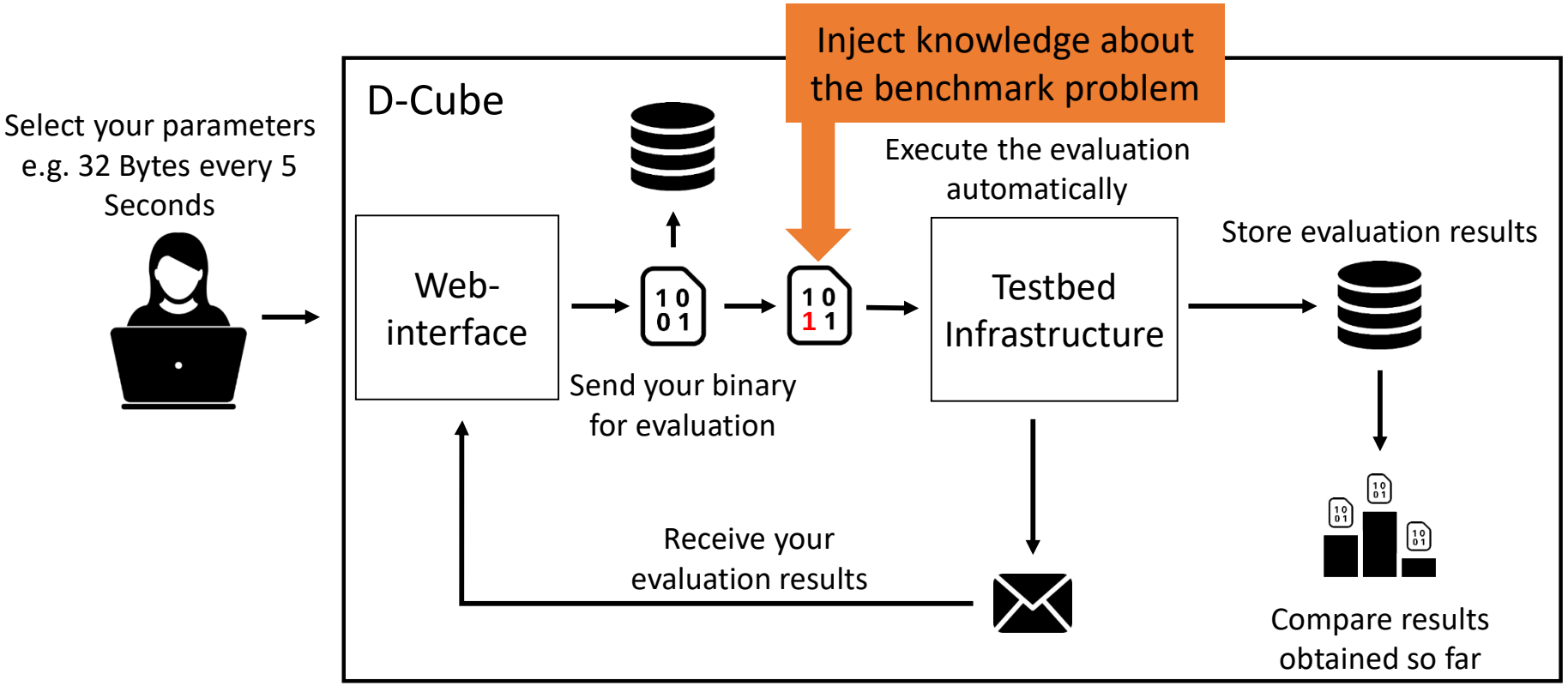
EWSN Dependability Competition

Step 3: define the benchmark problem(s)



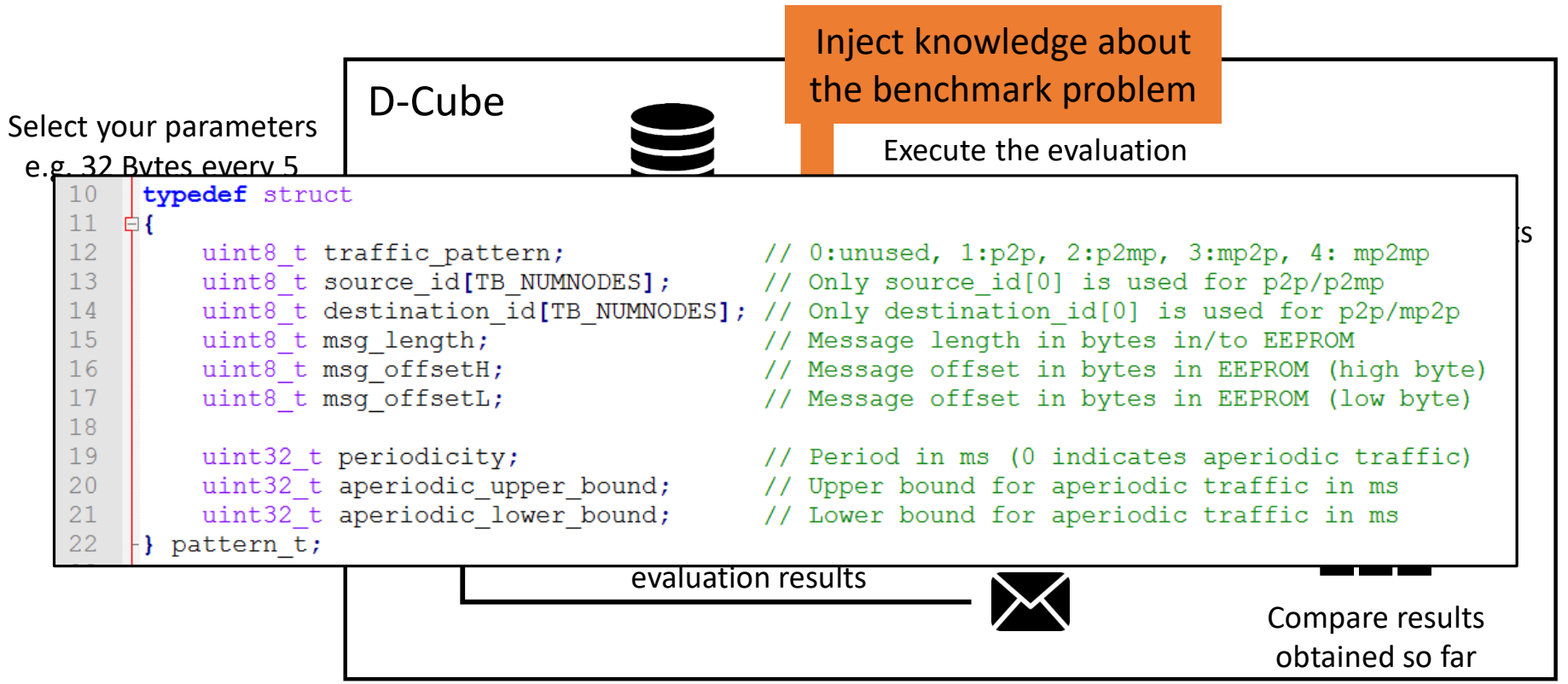
EWSN Dependability Competition

Step 3: define the benchmark problem(s)



EWSN Dependability Competition

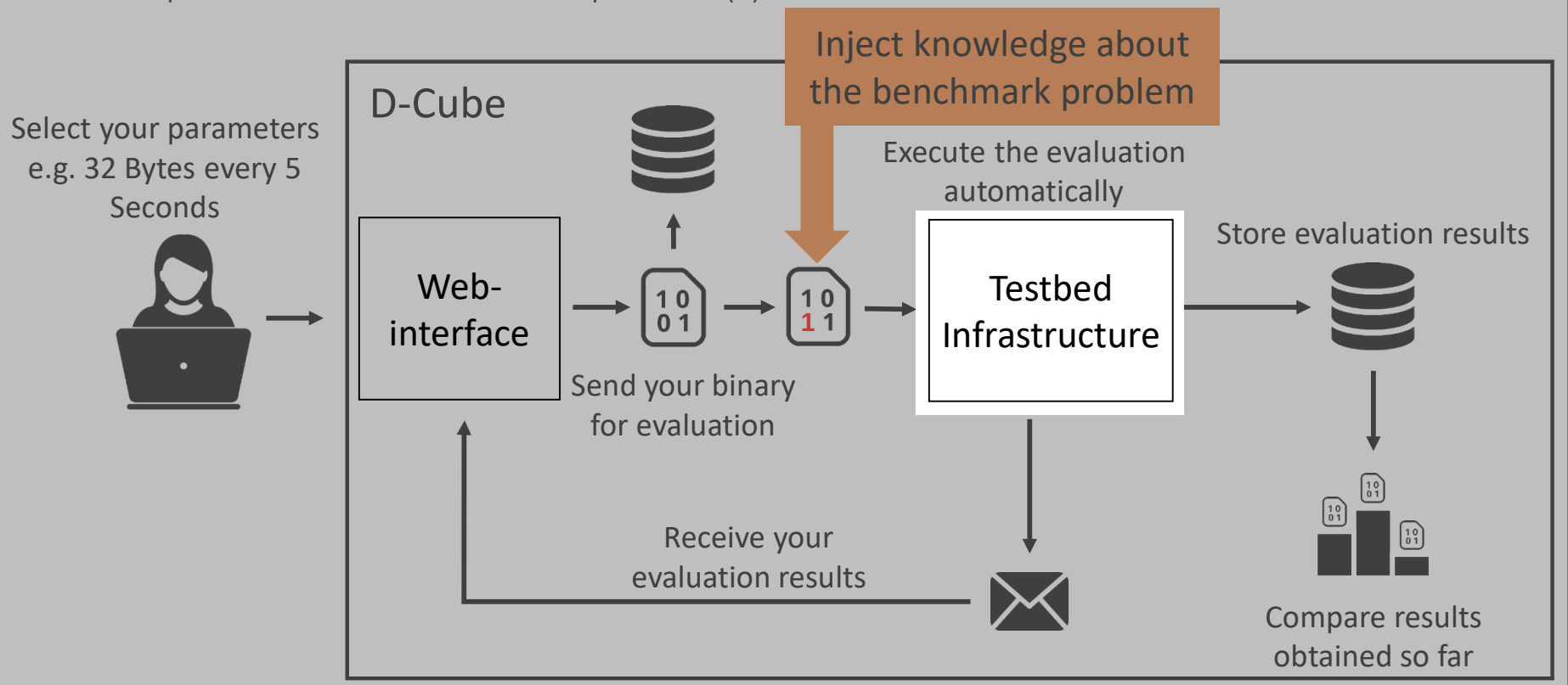
Step 3: define the benchmark problem(s)




```
10  typedef struct
11  {
12      uint8_t traffic_pattern;              // 0:unused, 1:p2p, 2:p2mp, 3:mp2p, 4: mp2mp
13      uint8_t source_id[TB_NUMNODES];       // Only source_id[0] is used for p2p/p2mp
14      uint8_t destination_id[TB_NUMNODES]; // Only destination_id[0] is used for p2p/mp2p
15      uint8_t msg_length;                   // Message length in bytes in/to EEPROM
16      uint8_t msg_offsetH;                  // Message offset in bytes in EEPROM (high byte)
17      uint8_t msg_offsetL;                  // Message offset in bytes in EEPROM (low byte)
18
19      uint32_t periodicity;                 // Period in ms (0 indicates aperiodic traffic)
20      uint32_t aperiodic_upper_bound;       // Upper bound for aperiodic traffic in ms
21      uint32_t aperiodic_lower_bound;       // Lower bound for aperiodic traffic in ms
22  } pattern_t;
```

EWSN Dependability Competition

Step 3: define the benchmark problem(s)



D-Cube

Low-cost Benchmarking infrastructure

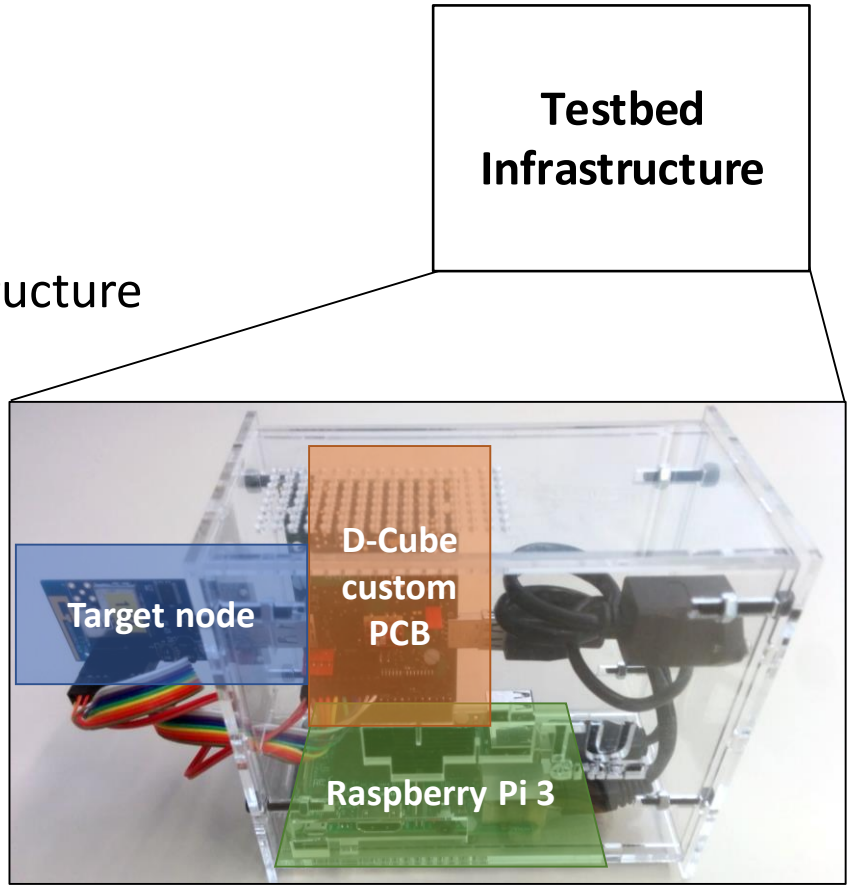
The physical component of the infrastructure

Build on top of off-the-shelf hardware

Raspberry Pi + **open-source** addon PCB

Focus on easy deployment

~ 100€ for the box, excluding the node



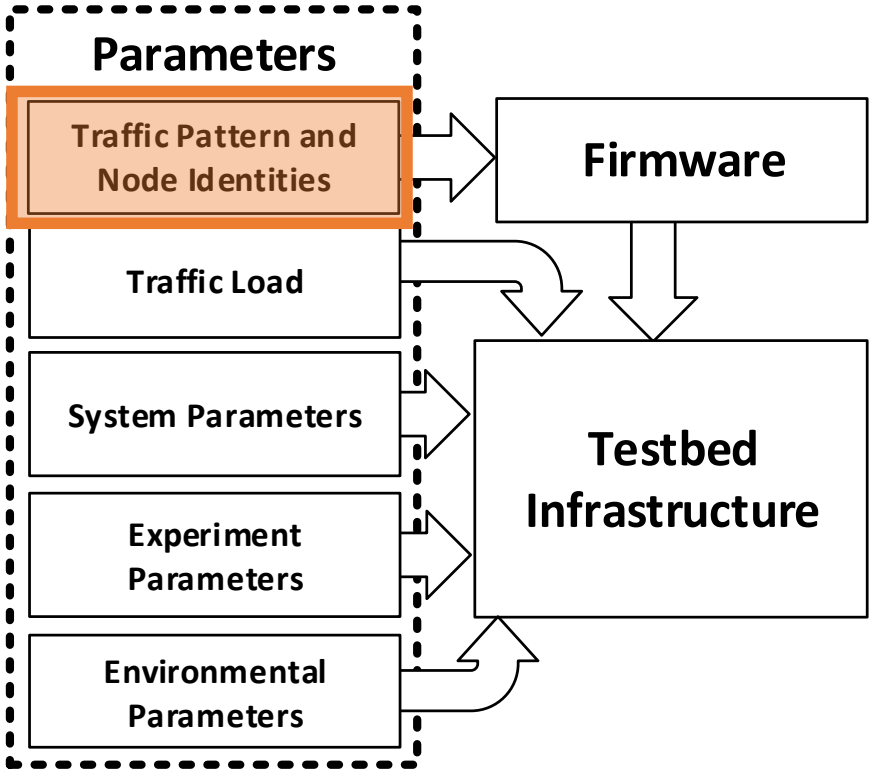
D-Cube

Low-cost Benchmarking infrastructure

Parameters can be modified on a per experiment basis (fully automated)

Supports the generation of stimuli

- Unix-style single purpose applications



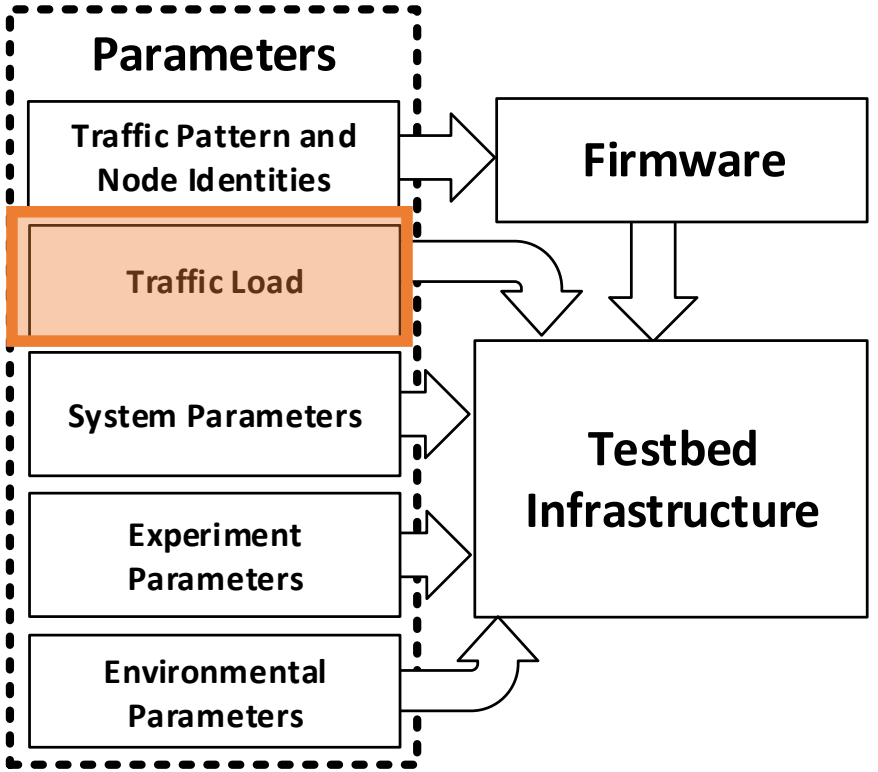
D-Cube

Low-cost Benchmarking infrastructure

Parameters can be modified on a per experiment basis (fully automated)

Supports the generation of stimuli

- Unix-style single purpose applications



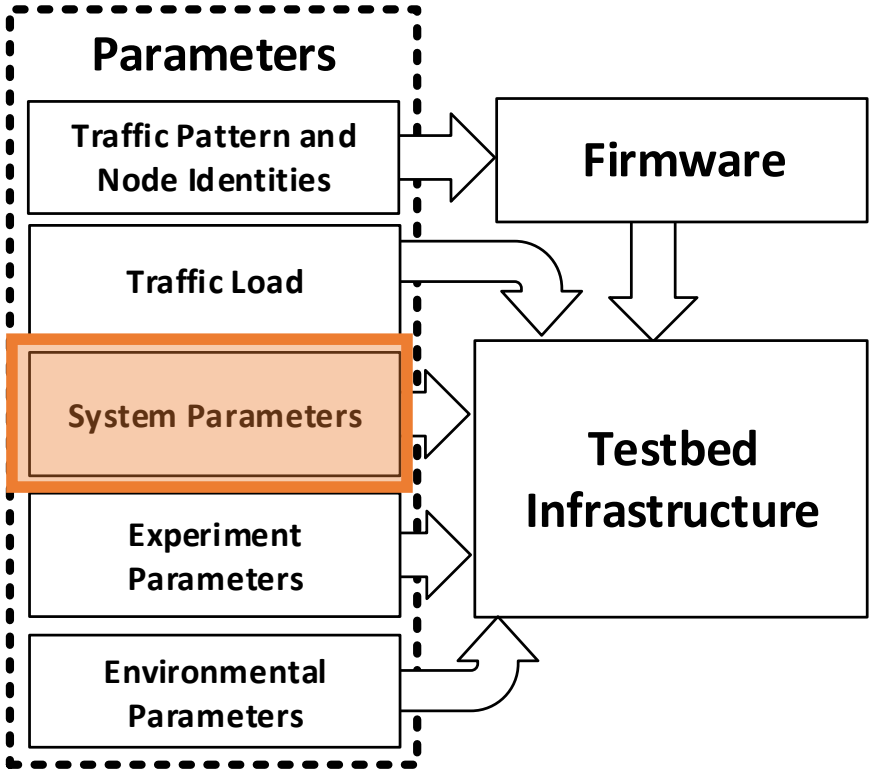
D-Cube

Low-cost Benchmarking infrastructure

Parameters can be modified on a per experiment basis (fully automated)

Supports the generation of stimuli

- Unix-style single purpose applications



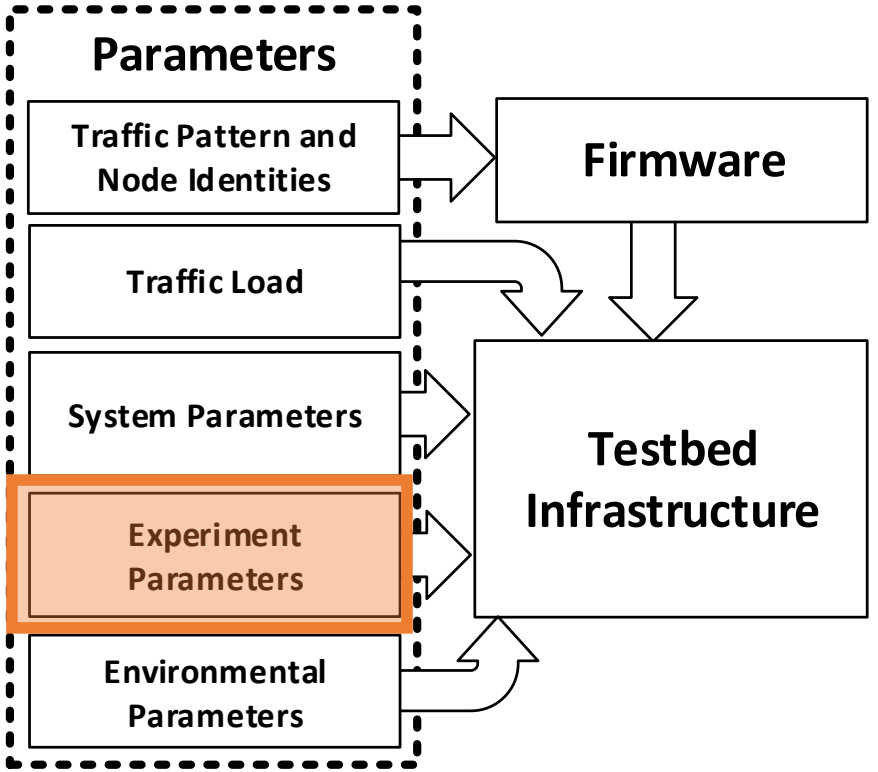
D-Cube

Low-cost Benchmarking infrastructure

Parameters can be modified on a per experiment basis (fully automated)

Supports the generation of stimuli

- Unix-style single purpose applications



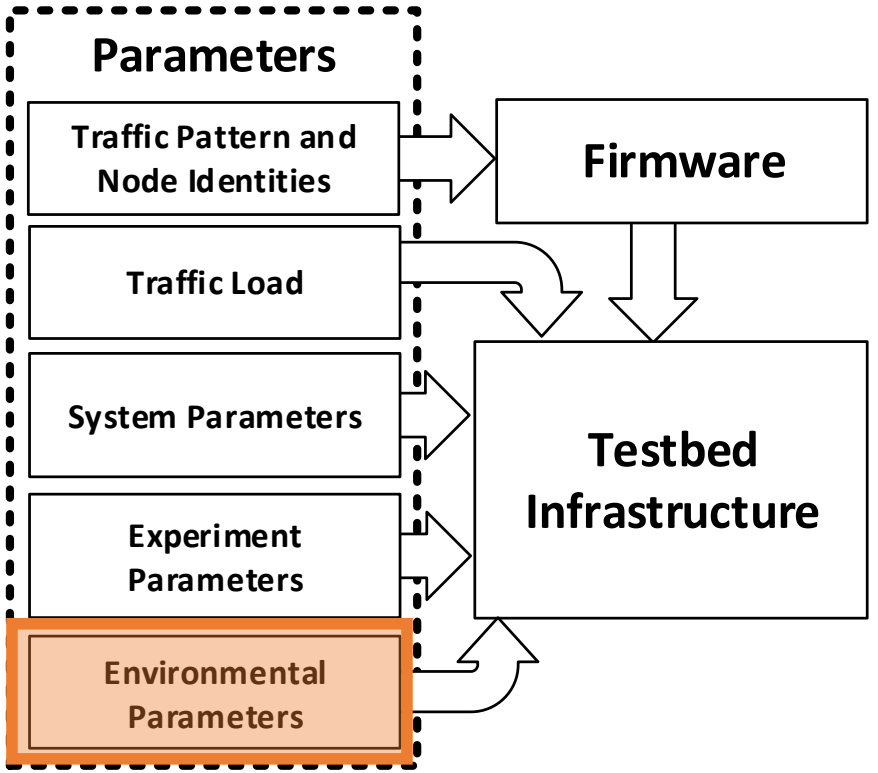
D-Cube

Low-cost Benchmarking infrastructure

Parameters can be modified on a per experiment basis (fully automated)

Supports the generation of stimuli

- Unix-style single purpose applications



JamLab-NG

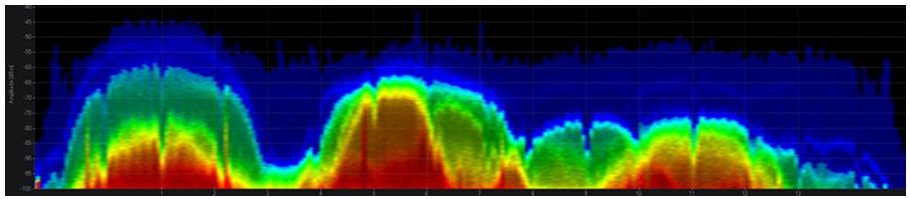
Repeatable interference generation

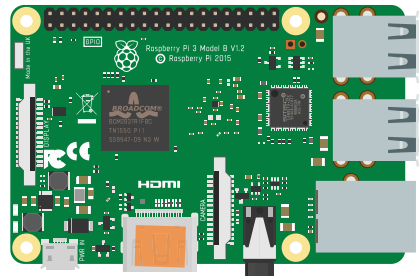
To compare results, running experiments in an “office” is insufficient

Raspberry Pi used in D-Cube comes with a build-in Wi-Fi card

Firmware is modified to generate interference on the Wi-Fi card itself

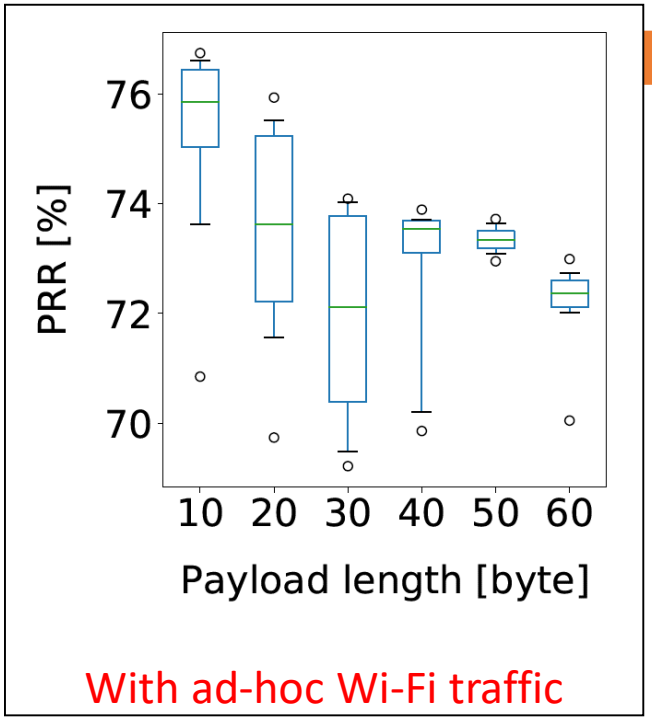
Complete control would require a testbed devoid of any RF interference



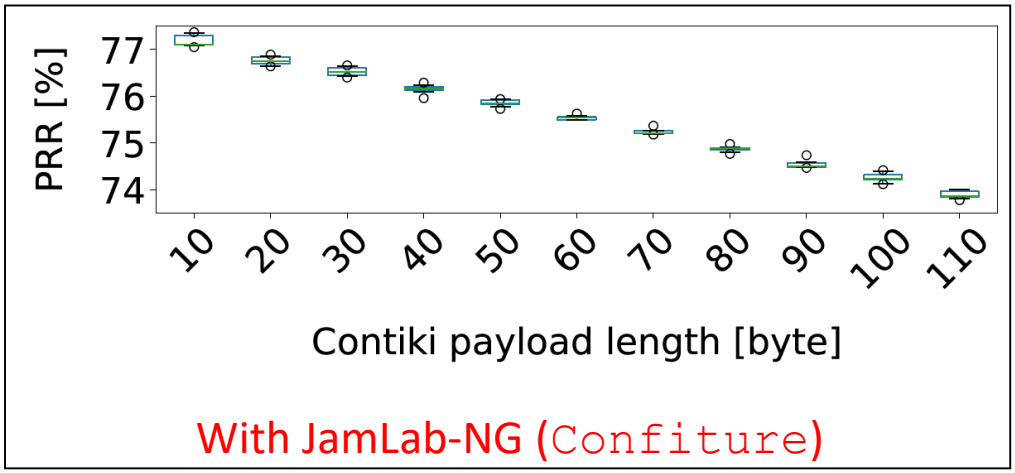


JamLab-NG

Repeatable interference generation



With ad-hoc Wi-Fi traffic



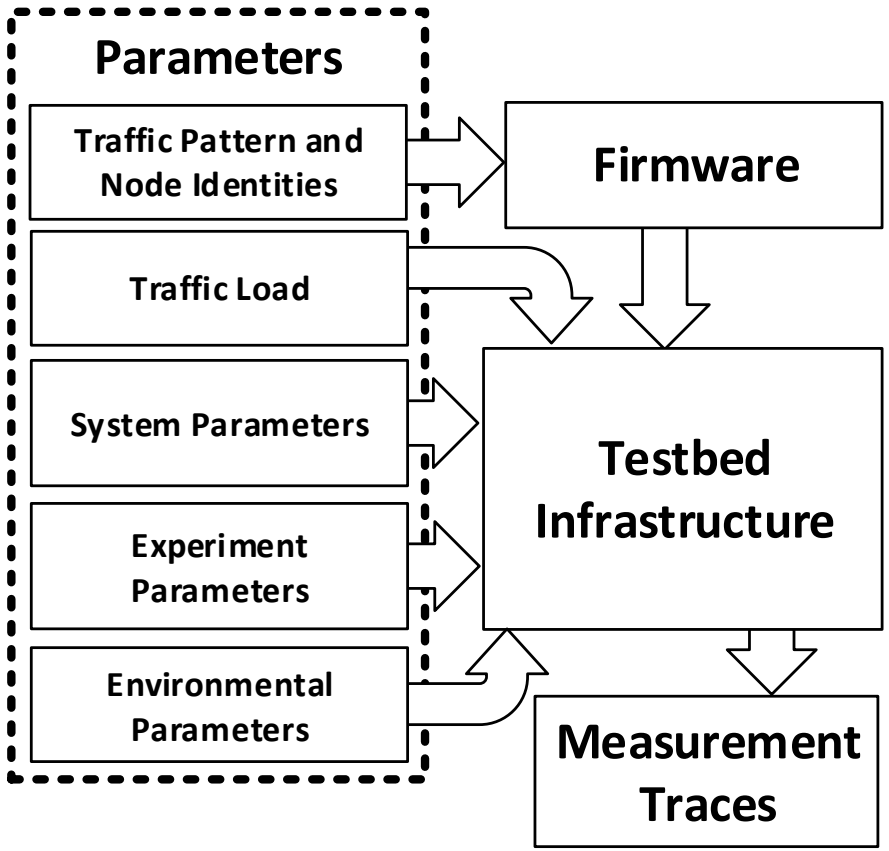
With JamLab-NG (`Confiture`)

D-Cube

Low-cost Benchmarking infrastructure

Testbed performs the measurements

- Does not affect the target node
- Real-time monitoring
- Energy, I/O with timestamp, node communication

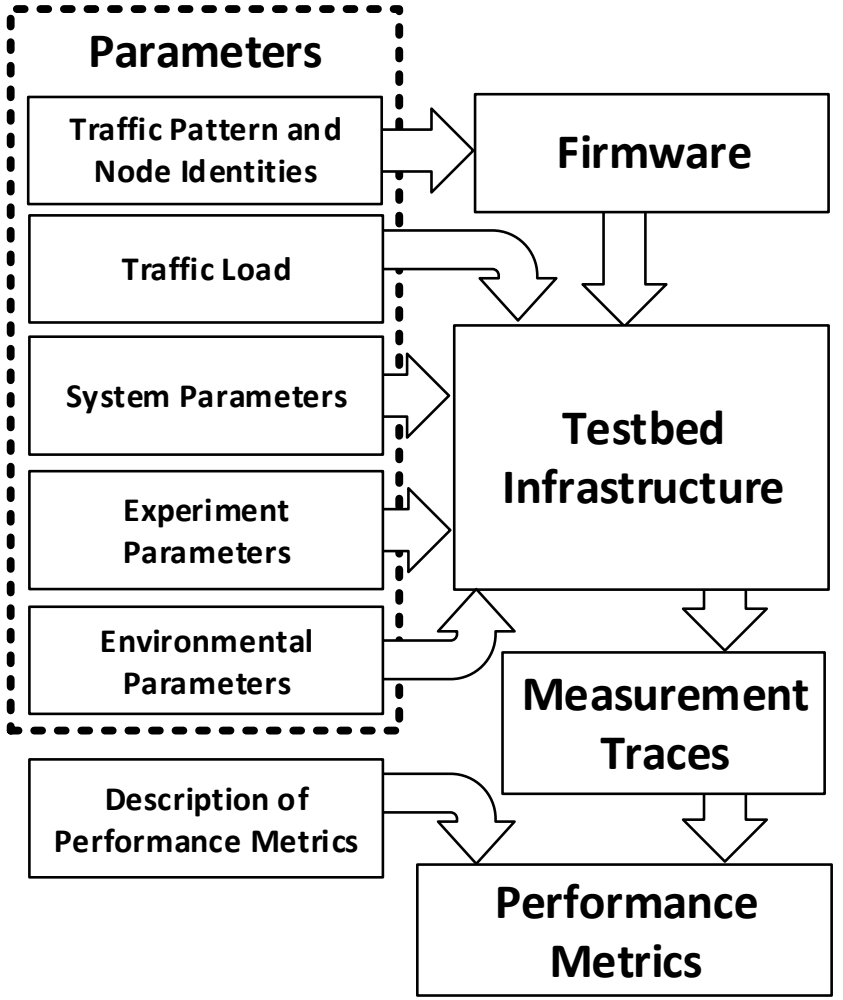


D-Cube

Low-cost Benchmarking infrastructure

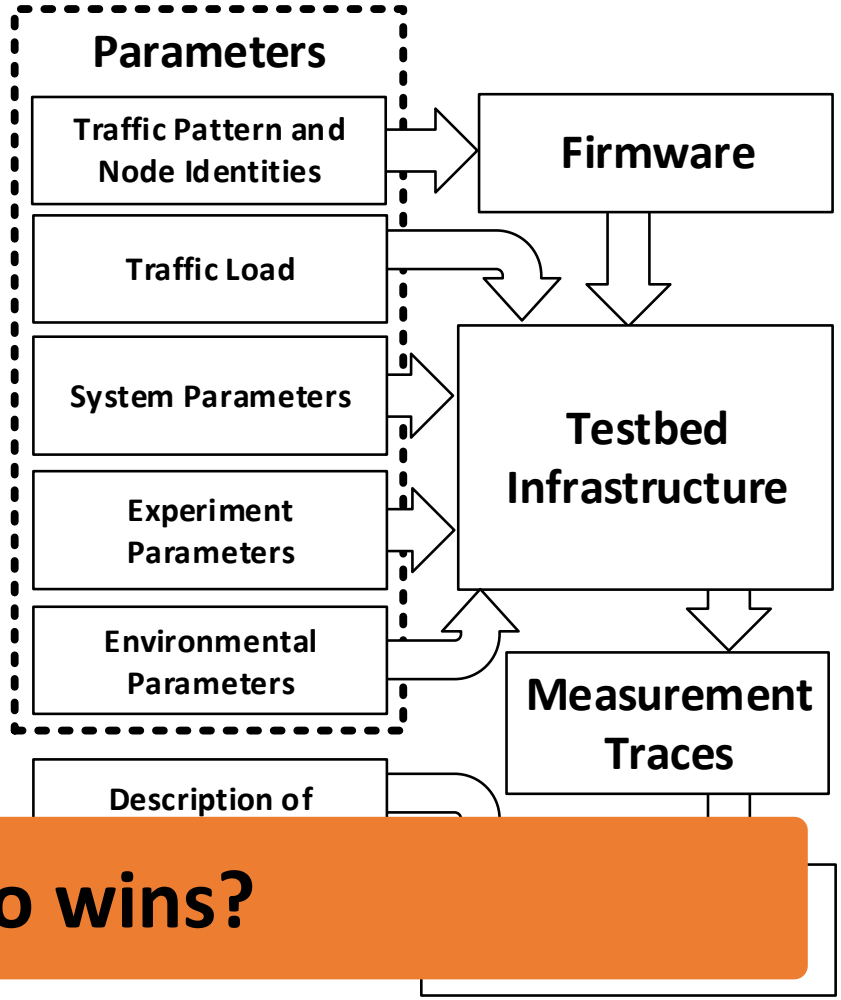
Computation of Performance metrics

- Reduction of dimensionality
- Are computed after completion
- Using python + pandas for easy analysis
- Used for comparison of results



D-Cube

Low-cost Benchmarking infrastructure



Computation of Performance metrics

- Reduction of dimensionality
- Are computed after completion
- Using python + pandas for easy analysis
- Used for comparison of results

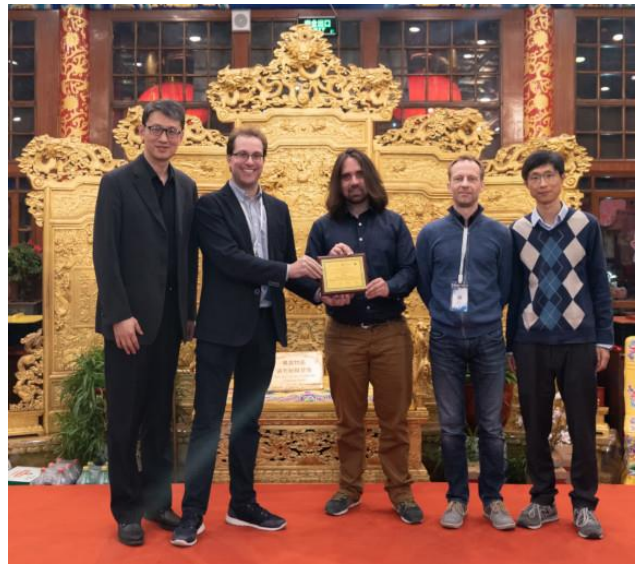
… And who wins?

EWSN Dependability Competition

Who wins the competition?

2019 Data Collection
DeCoT+ (Academia)

2019 Dissemination
BigBangBus (Industry)

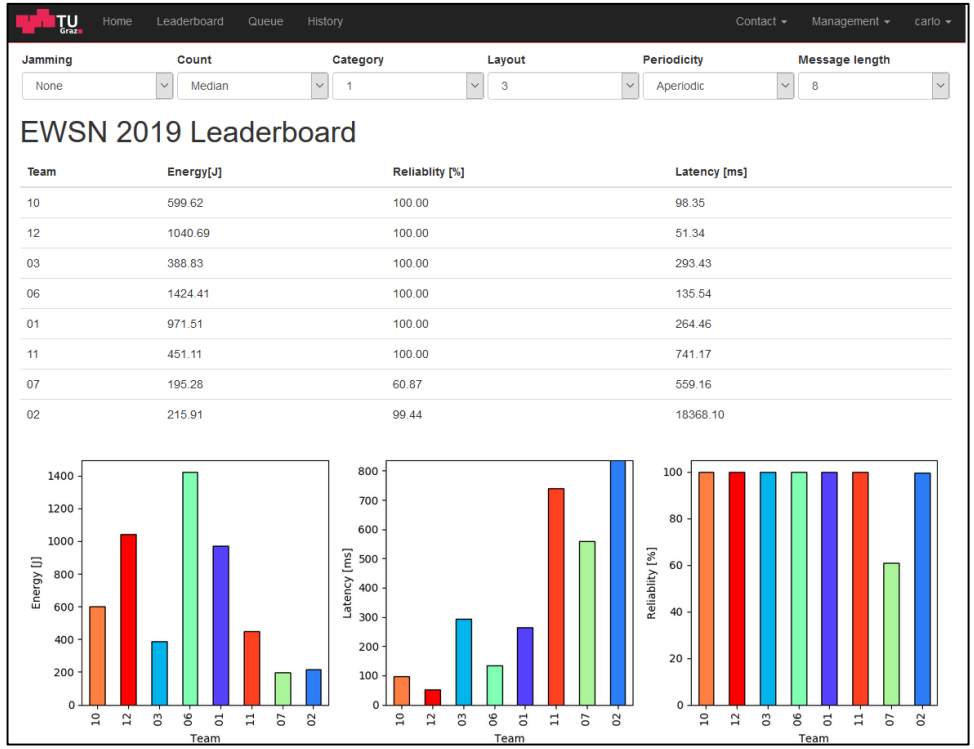


EWSN Dependability Competition

Comparison of performance metrics

Leaderboard

- Public version during the preparation
- Final version after evaluation phase
- Comparison for a single combination of parameters

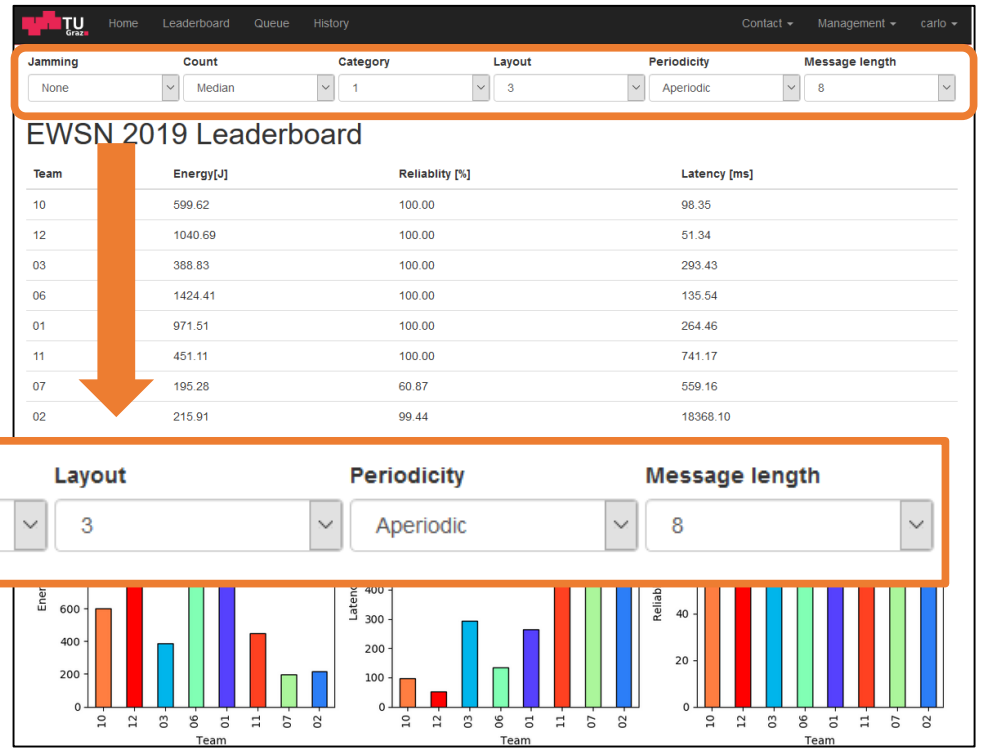


EWSN Dependability Competition

Comparison of performance metrics

Leaderboard

- Public version during the preparation
- Final version after evaluation phase
- Comparison for a single combination of parameters

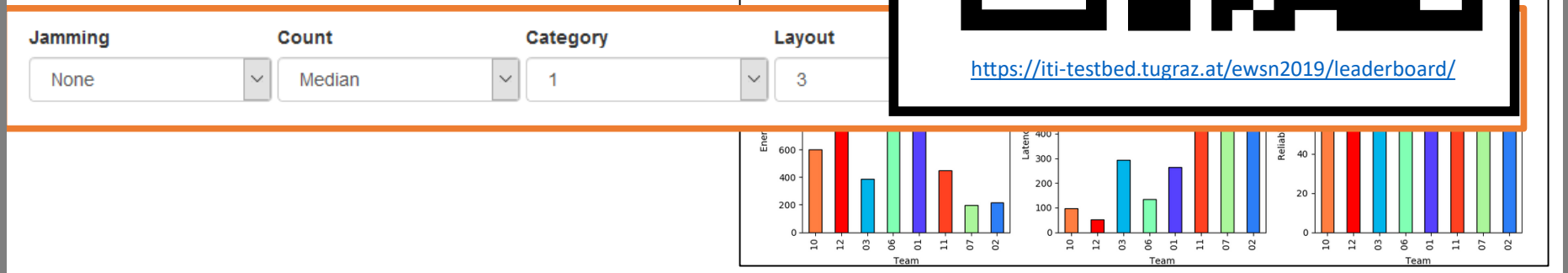


EWSN Dependability Comp

Comparison of performance metrics

Leaderboard

- Public version during the preparation
- Final version after evaluation phase
- Comparison for a single combination of parameters



https://iti-testbed.tugraz.at/ewsn2019/leaderboard/

EWSN Dependability Competition

Comparison of performance metrics

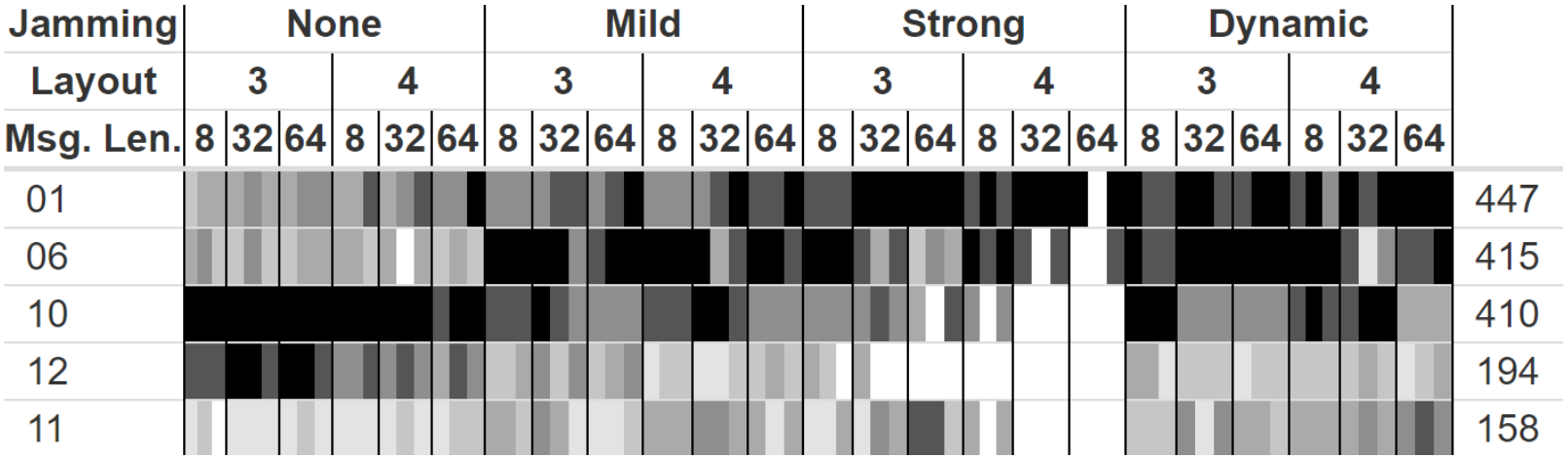
Heatmap

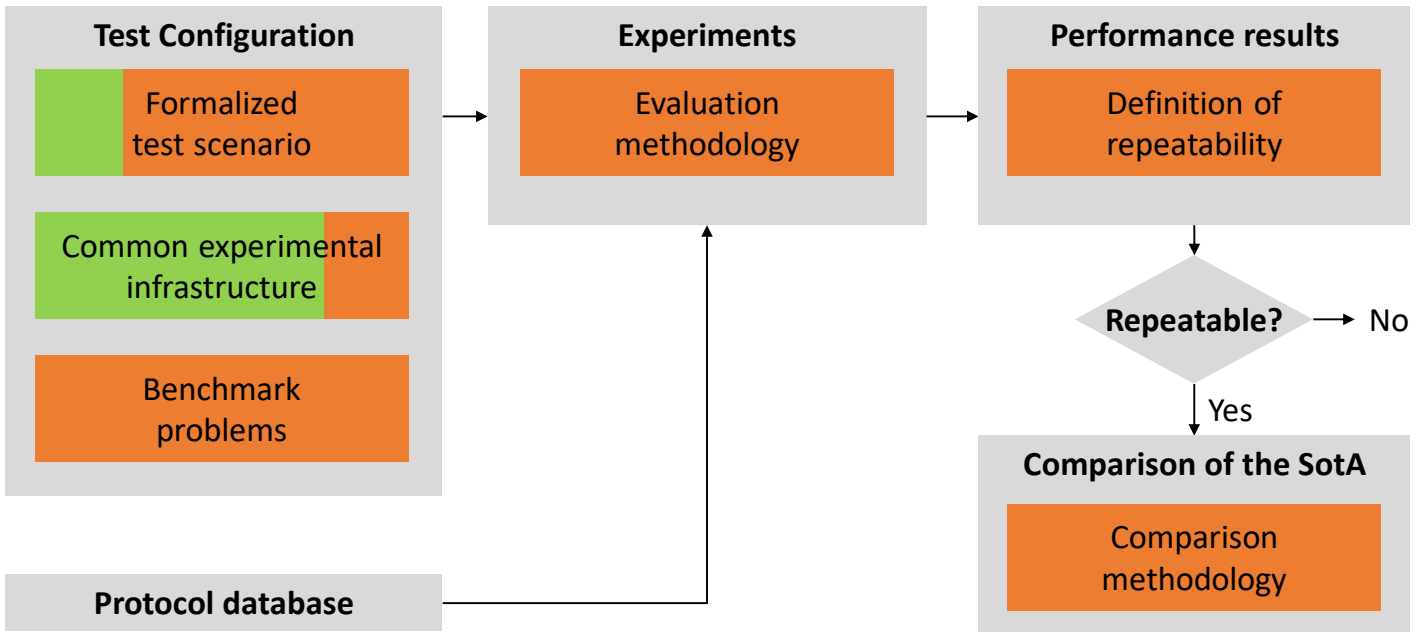
Comparing the results of individual Benchmark Problems

Insight into overall performance

Legend 9 6 4 3 2 1

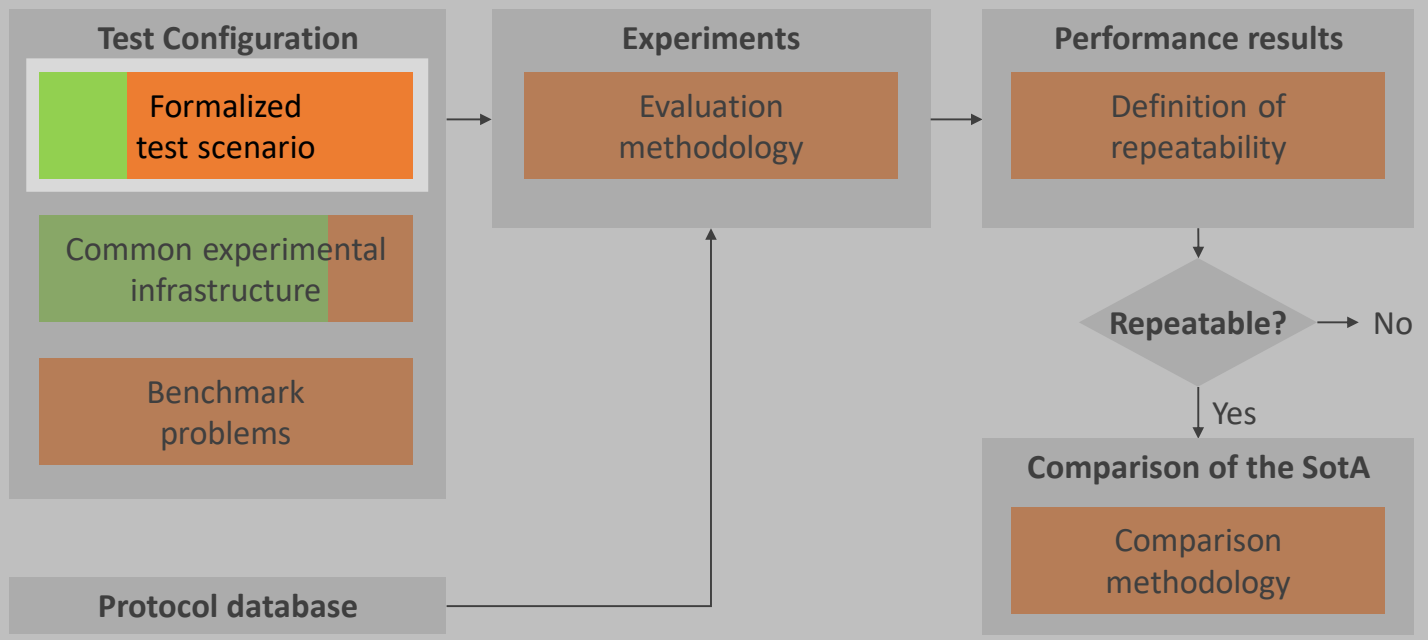
From left to right, for each message length: aperiodic, periodic 5s, periodic 30s





Looking ahead

On-going and future work



Test configuration = Test scenario + Test environment

Test Scenario	Traffic type Period Payload Number of sources …
Test Environment	Number of nodes Platform Frequency band …

tiny.cc/TestConfig

The description framework serves to describe profiles and benchmark problems

		Profile
Test Scenario	Traffic type	Terminating
	Period	Short
	Payload	Medium
	Number of sources	High
	...	...
Test Environment	Number of nodes	20
	Platform	TelosB
	Frequency band	2.4 GHz
	...	...

The description framework serves to describe profiles and benchmark problems

		Profile
Test Scenario	Traffic type	Terminating
	Period	Short
	Payload	Medium
	Number of sources	High
	...	...
Test Environment	Number of nodes	20
	Platform	TelosB
	Frequency band	2.4 GHz
	...	...

Describe your own setup

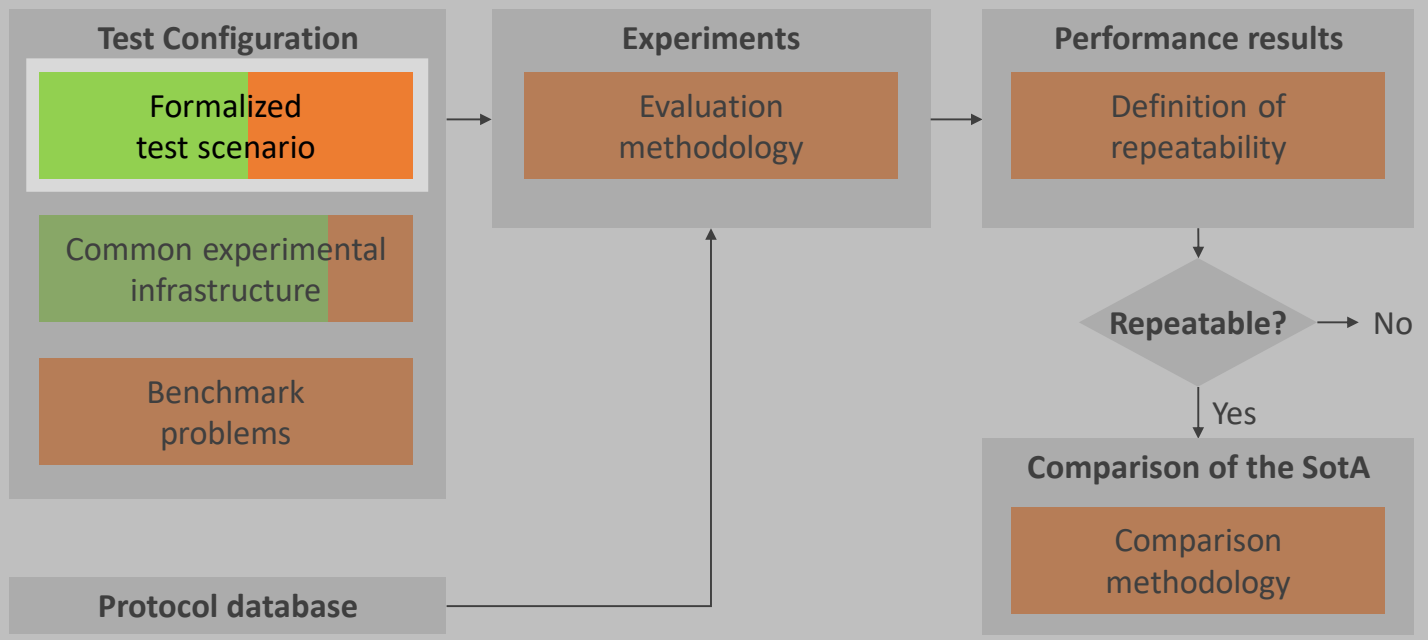
The description framework serves to describe profiles and benchmark problems

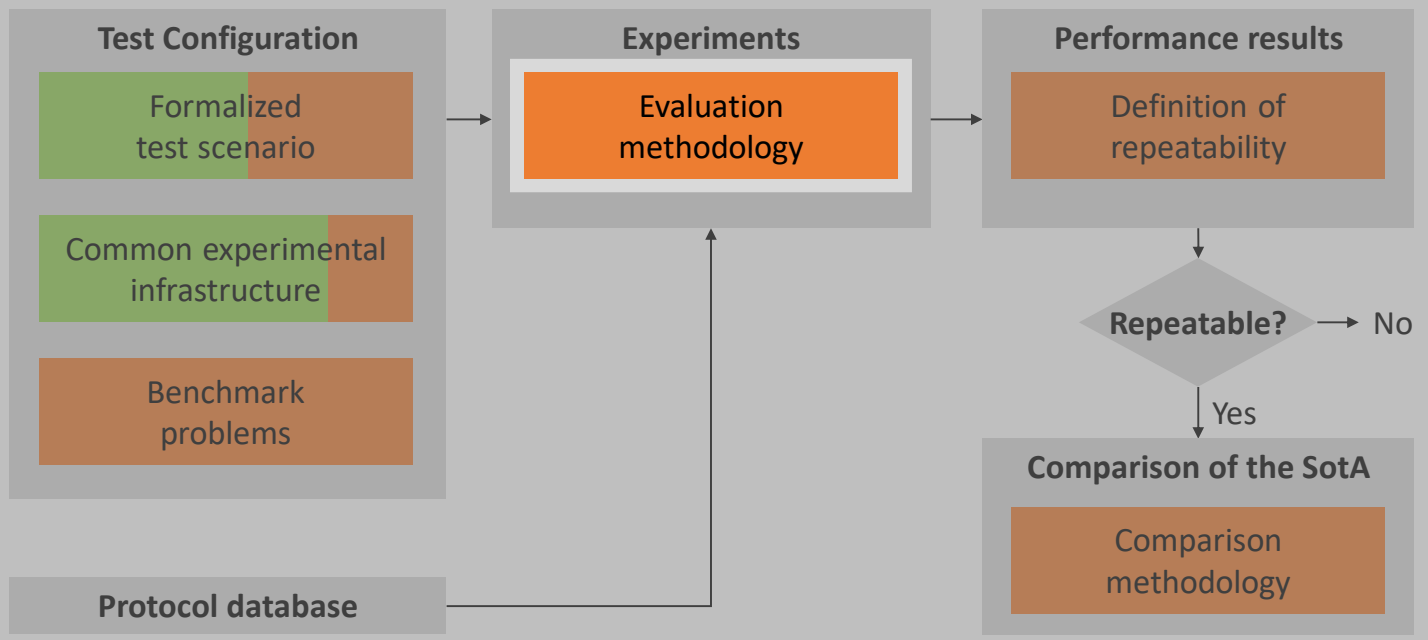
		Profile	Benchmark pbm
Test Scenario	Traffic type	Terminating	40 s
	Period	Short	100 ms
	Payload	Medium	16 B
	Number of sources	High	19
	...	...	...
Test Environment	Number of nodes	20	Graz_Layout3
	Platform	TelosB	
	Frequency band	2.4 GHz	
	...	...	

The description framework serves to describe profiles and benchmark problems

		Profile	Benchmark pbm
Test Scenario	Traffic type	Terminating	40 s
	Period	Short	100 ms
	Payload	Medium	16 B
	Number of sources	High	19
	...	...	...
Test Environment	Number of nodes	20	Graz_Layout3
	Platform	TelosB	
	Frequency band	2.4 GHz	
	...	...	

Exact setup (public)



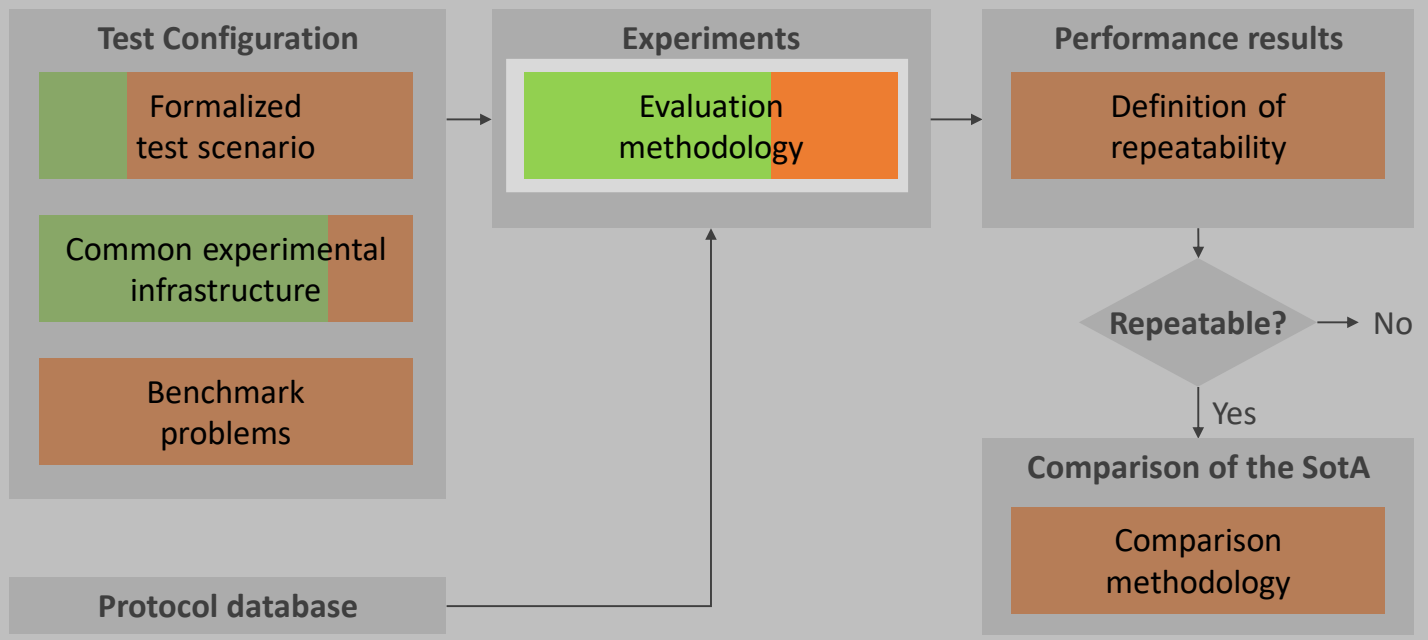


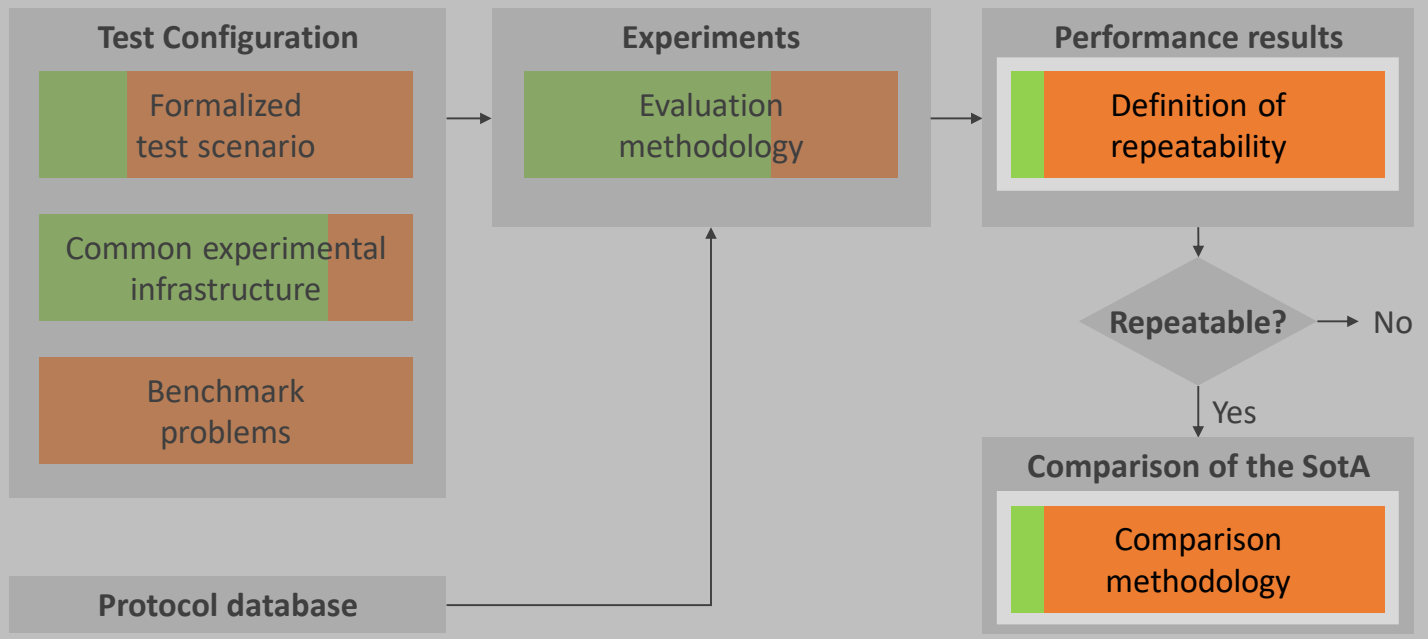
Towards a Methodology for Experimental Evaluation in Low-Power Wireless Networking

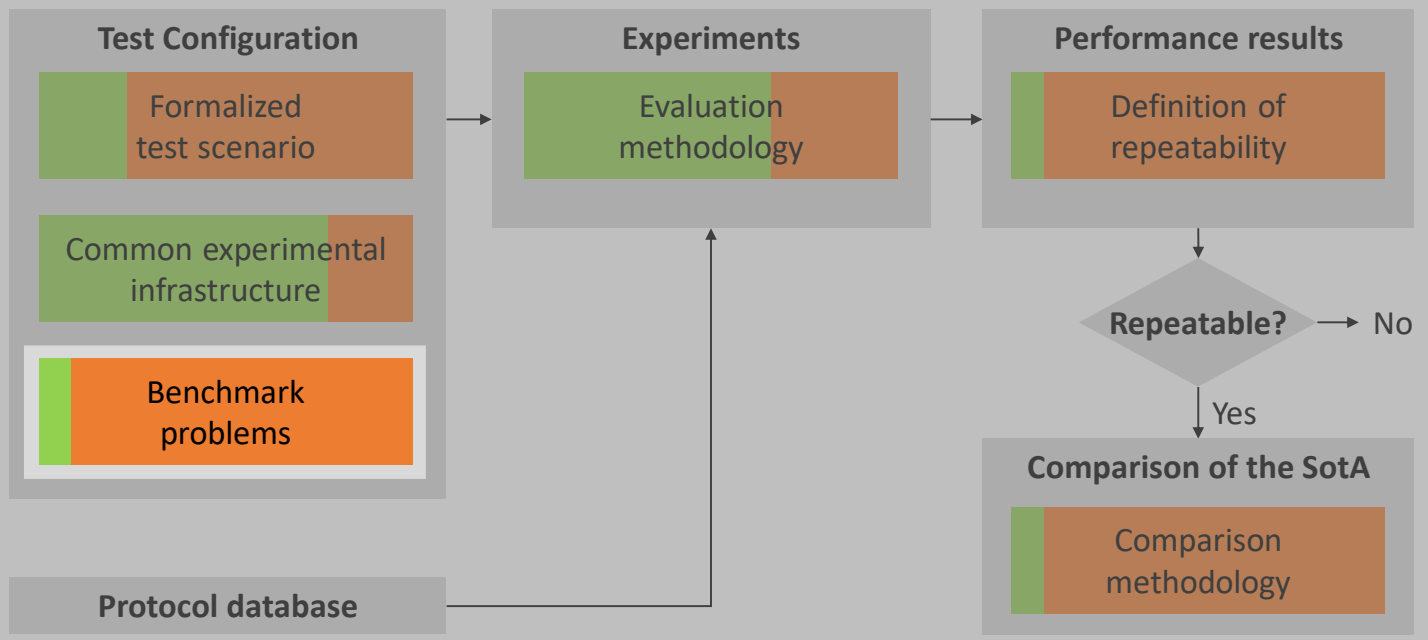
Romain Jacob
Usman Raza
Lothar Thiele
Carlo Alberto Boano
Marco Zimmerling

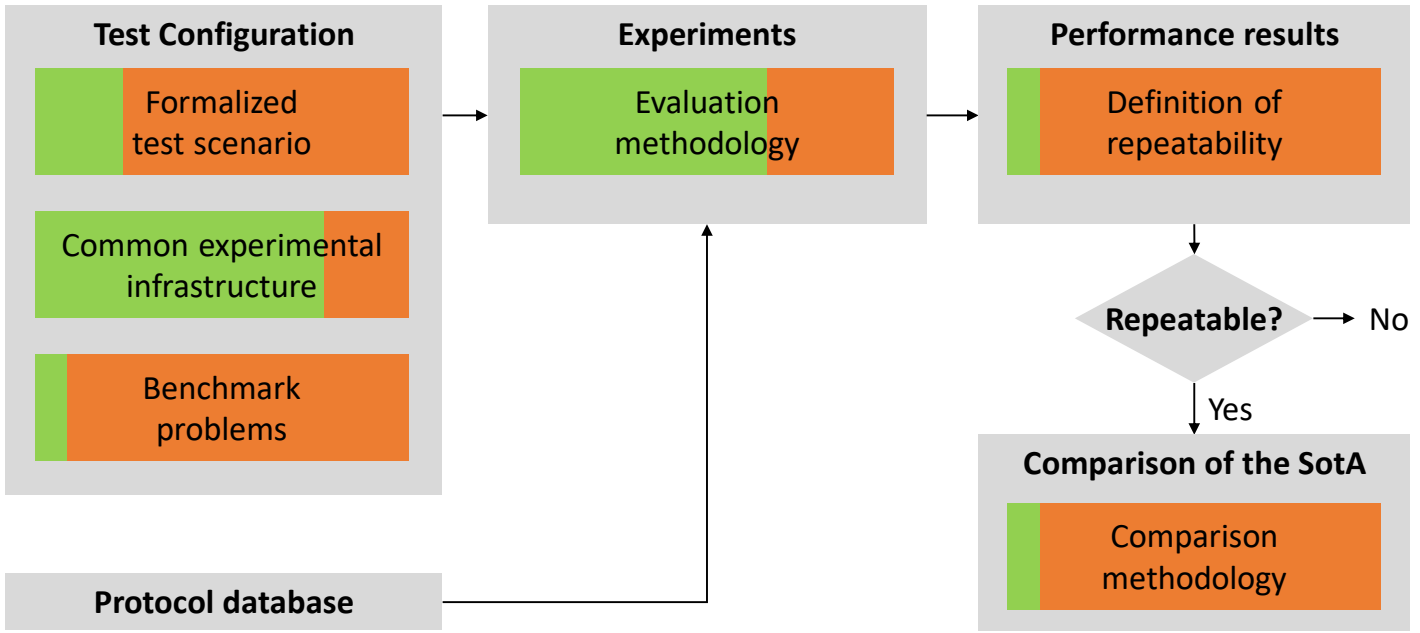
ETH zürich
TOSHIBA
TU Graz
TECHNISCHE UNIVERSITÄT DRESDEN

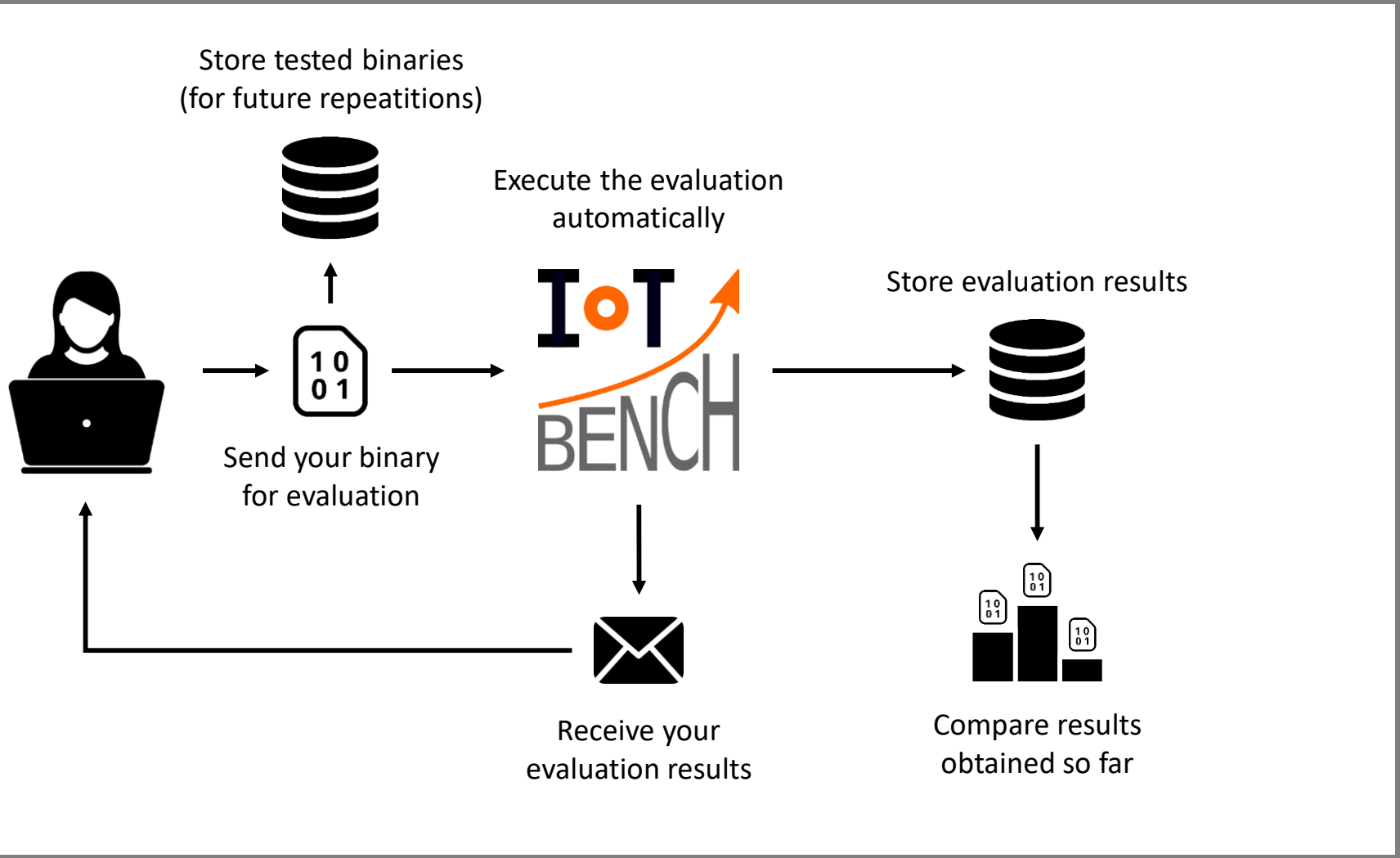
Includes material from Hanspeter Schmid and Alex Huber











We need

Standardized metrics for the evaluations

Standardized test scenarios

Central repository

Common interface to the different testbeds

Test environments

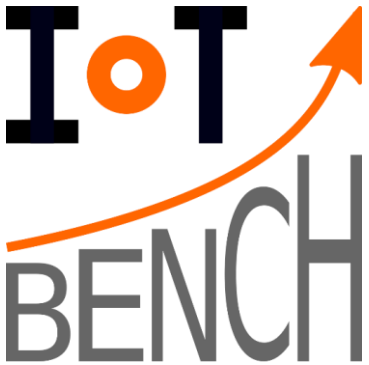
Benchmarks

Scenarios description
Protocols binary
Evaluation results

Testbeds
Simulators

We need
You!

IoTBench - Past, present, and future of a community-driven benchmarking initiative



Join us and
Get involved!

www.iotbench.ethz.ch

@iot_bench

Icons from thenounproject.com