IoT Bench: Towards a Benchmark for Low-power Wireless Networking
There is a lack of confidence in the results despite tremendous efforts put in the evaluation.

- **2016**
  - **Data Prediction + Synchronous Transmissions = Ultra-low Power Wireless Sensor Networks**

- **2017**
  - **Five-Nines Reliable Downward Routing in RPL**

- **2017**
  - **RedFixHop with Channel Hopping: Reliable Ultra-Low-Latency Network Flooding**
... because there is still no standard way to evaluate low-power wireless protocols

Huge variety of settings

Inter-packet interval (s)

Experimental setup
2010 to 2015
Periodic data collection only
... because there is still no standard way to evaluate low-power wireless protocols

Huge variety of settings

Hard to compare against references

- Sources not available
- Experiment not reproducible
... because there is still no standard way to evaluate low-power wireless protocols

Huge variety of settings

Hard to compare against references

**Heterogeneous comparisons**

- Protocol only
- Protocol + Platform
A common benchmark is desperately needed

The benchmark should...

be **reproducible** across different settings

be **simple** and minimalistic

allow **consistent** comparisons

suggest the **appropriate experiments** to perform
A common benchmark is desperately needed

Researchers want to evaluate new protocols
A common benchmark is desperately needed ... by everyone!

Researchers evaluate new protocols

Companies want to showcase the advantage of their products
A common benchmark is desperately needed
... by everyone!

Researchers evaluate new protocols

Companies showcase the advantage of their products

Customers want to objectively compare protocol strengths
The community is getting together to tackle this challenge

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aug. 2016</td>
<td>Poster at SenSys</td>
</tr>
<tr>
<td>May 2017</td>
<td>1&lt;sup&gt;st&lt;/sup&gt; Meeting in Milan</td>
</tr>
<tr>
<td>Oct. 2017</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Meeting in Stockholm</td>
</tr>
<tr>
<td>Feb. 2018</td>
<td>Poster and presentation at EWSN</td>
</tr>
<tr>
<td>Today</td>
<td>CPSBench Workshop at CPSWeek</td>
</tr>
</tbody>
</table>
Benchmarking is fundamentally a tradeoff between generality and accuracy.
Benchmarking is fundamentally a tradeoff between generality and accuracy.
Benchmarking is fundamentally a tradeoff between generality and accuracy.

![Graph showing the tradeoff between generality and accuracy]

- **Step 1**: Common Conceptual Framework
- **Perfect Comparison**
  - Same experimental conditions
  - High accuracy
  - Poor usability

Accuracy

Generality
Step 1  Define a common conceptual framework

Parameters  ———  Input  Controlled by the user

*e.g.*  number of nodes
traffic load ...
Step 1 Define a common conceptual framework

Parameters

Metrics

Output

Determined by the protocol behavior

\textit{e.g.}\ packet reception rate, energy consumption

Observed

Characterize the environment

\textit{e.g.}\ level of external interference
Step 1  Define a common conceptual framework

Parameters
- System: Topology, link quality, mobility...
- Traffic: Traffic load and pattern
- Experiment: Number of runs, acceptable deviation...
- Environment: Interference, people’s presence...
- Performance: Delivery, latency, energy...

Metrics
Step 1  Define a common conceptual framework

<table>
<thead>
<tr>
<th>Parameters</th>
<th>&quot;assignment of concrete values to input parameters, and a precise definition of observed and output metrics to be measured&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metrics</td>
<td>e.g. Low-rate data collection</td>
</tr>
<tr>
<td></td>
<td>▪ Input parameters #nodes: 100, #sources: 99, #destinations: 1, traffic load: from 0.1 msg/min to 1 msg/min.</td>
</tr>
<tr>
<td></td>
<td>▪ Observed metrics link qualities and external interference.</td>
</tr>
<tr>
<td></td>
<td>▪ Output metrics packet delivery rate, end-to-end packet latency, and average energy consumption.</td>
</tr>
</tbody>
</table>
Benchmarking is fundamentally a tradeoff between generality and accuracy.
Benchmarking is fundamentally a tradeoff between generality and accuracy.

- **Generality**
  - Step 1: Common Conceptual Framework
  - Step 2: Common Experimental Infrastructure

- **Target of IoT Bench**
  - Perfect Comparison
    - Same experimental conditions
    - High accuracy
    - Poor usability

- **Accuracy**
Step 2  Setup a common experimental infrastructure

Key Idea

Separate networking code from experiment scenario

1. Node runs networking code
2. Testbed/Simulator runs the experiment
e.g. using GPIO/serial to instruct nodes and measure the profile’s metrics

similar to the 2018 EWSN Competition
### Step 2

Setup a common experimental infrastructure

<table>
<thead>
<tr>
<th>Key Idea</th>
<th>Separate networking code from experiment scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benefits</strong></td>
<td>Fully automated&lt;br&gt;Rules out mis-interpretation of profiles</td>
</tr>
<tr>
<td><strong>Drawbacks</strong></td>
<td>More complex&lt;br&gt;Stricter&lt;br&gt;More infrastructure maintenance</td>
</tr>
</tbody>
</table>
A lot of remains to be done!

Define useful profiles

Improve the testbeds

Improve the simulators
To make IoTBench a reality, we have...

A team

Communication channels

A plan

and

20 years of experience in wireless protocols
What we don’t have, *yet*

- Industry support and partners
- Experience about benchmarking
- Nor Time!

Need to tackle this problem
- quickly
- efficiently
IoTBench: Towards a Benchmark for Low-power Wireless Networking

Stay tuned!
and get involved 😊

www.iotbench.ethz.ch
@iot_bench

Icons from thenounproject.com