

Towards a Benchmark for Low-power Wireless

EWSN'18, February 15
Madrid

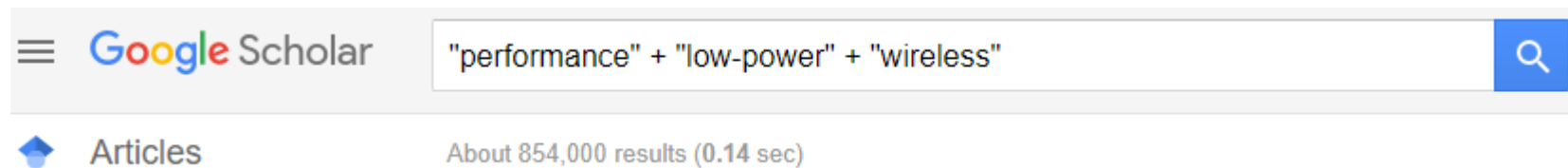


@iot_bench

Simon Duquennoy
on behalf of the IoT Benchmarking consortium

What is this about?

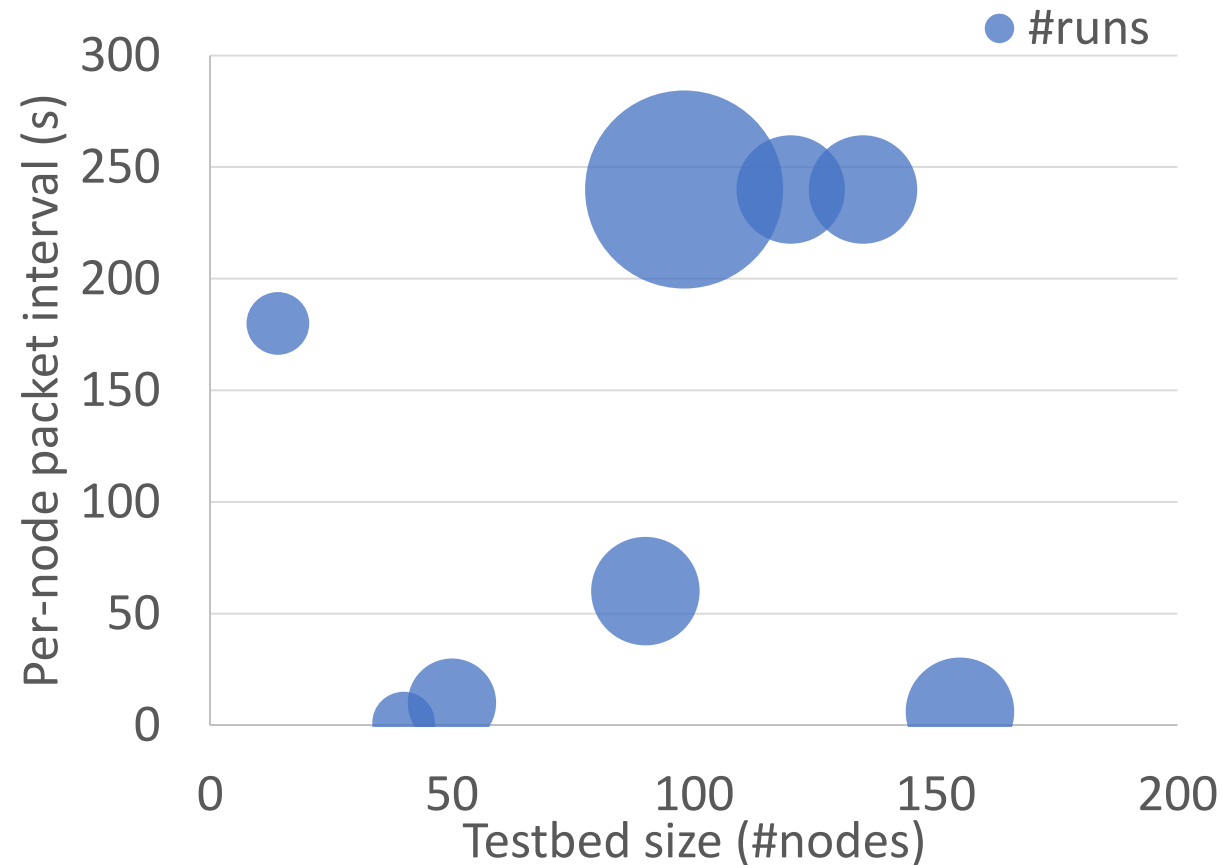
- Almost two decades of WSN / IoT wireless research
- ... and yet no standard way to evaluate!



- Either we are done
 - > switch to next topic
- ... or we believe more research is needed
 - > then we need a benchmark!

A little case study

- Manually gathered stats from papers
 - 2004 – 2016
 - Only periodic data collection here



What we mean with a benchmark

- Challenges in evaluation of low-power wireless protocols
 - Variety of settings: deployment, metrics, application scenario..
 - Comparing against reference implementations is hard
 - Comparison: protocols vs. protocols+platform
- Benchmark: a set of tools and practices for performance evaluation
 - Enables fair comparison
 - Enables repeatability (to a certain extent..)
 - *(as a complement to custom evaluations)*

History

- 2016, June
 - Small group (Olaf, Carlo, Marco and I) discuss the idea of a benchmark
- 2016, August
 - Poster submitted to SenSys (11 unique affiliations)
 - Drafts goals and challenges
- 2017, February
 - Ad-hoc meeting at EWSN, Uppsala
 - Group expands
- 2017, May
 - Plenary meeting in Milan
 - Group expands
- 2017, October
 - Plenary meeting in Stockholm
 - Group expands
- 2017, December-onwards
 - Bi-monthly telcos, face-to-face when needed

Poster Abstract: A Benchmark for Low-power Wireless Networking

Simon Duquennoy^{1,2}, Olaf Landsiedel³, Carlo Alberto Boano⁴, Marco Zimmerling⁵, Jan Beutel⁶, Mun Choon Chan⁷, Omprakash Gnawali⁸, Mobashir Mohammad⁹, Luca Mottola^{9,2}, Lothar Thiele⁶, Xavier Vilajosana¹⁰, Thiemo Voigt^{11,2}, Thomas Watteyne¹

¹Inria, France ²SICS Swedish ICT, Sweden ³Chalmers University of Technology, Sweden
⁴Graz University of Technology, Austria ⁵TU Dresden, Germany ⁶ETH Zurich, Switzerland
⁷NUS, Singapore ⁸University of Houston, USA ⁹Politecnico di Milano, Italy
¹⁰Open University of Catalonia, Spain ¹¹Uppsala University, Sweden

ABSTRACT

Experimental research in low-power wireless networking lacks a reference benchmark. While other communities such as databases or machine learning have standardized benchmarks, our community still uses ad-hoc setups for its experiments and struggles to provide a fair comparison between communication protocols. Reasons for this include the diversity of network scenarios and the stochastic nature of wireless experiments. Leveraging on the excellent testbeds and tools that have been built to support experimental validation, we make the case for a reference benchmark to promote a fair comparison and reproducibility of results. This abstract describes early design elements and a benchmarking methodology with the goal to gather feedback from the community rather than propose a definite solution.

1. INTRODUCTION

The low-power wireless community has put a lot of effort

2. GOALS

With the benchmark, we aim to provide researchers with a standardized way to evaluate their protocols and compare the results in a fair manner to the state of the art. To this end, the benchmarking suite should allow to run protocols on testbeds in a reproducible manner under different settings (*e.g.*, traffic load, traffic pattern, controlled interference) and measure key metrics including end-to-end packet reliability, latency, and radio duty cycle. The defined scenarios, settings, and testbeds should be maintained as a continuous community process, allowing the benchmark to evolve and adapt to technology and application trends.

The benchmarking suite should execute the selected protocols and scenarios automatically on behalf of the user. Ideally, users simply upload their firmware, select a benchmark scenario, and hit the *play* button. The testbed then takes care of running the corresponding experiment and collecting the results. The experimental setup and settings of each scenario in the benchmarking suite should be configurable.

The current team

- Simon Duquennoy (RISE SICS)
- Olaf Landsiedel (Chalmers)
- Carlo Boano (TU Graz)
- Marco Zimmerling (TU Dresden)
- Omprakash Gnawali (Univ. Houston)
- Mobashir Mohammad (NUS)
- Mun Choon Chan (NUS)
- Lothar Thiele (ETHZ)
- Jan Beutel (ETHZ)
- Luca Mottola (Politecnico di Milano and RISE SICS)
- Thiemo Voigt (Uppsala Univ and RISE SICS)
- Thomas Watteyne (Inria and Linear Technology)
- Xavier Vilajosana (Open University of Catalonia)
- Gian Pietro Picco (University of Trento)
- Anna Förster (Universität Bremen)
- Idrees Zaman (Universität Bremen)
- Ramona Marfievici (Cork Institute of Technology)
- Koen Langendoen (TU Delft)
- Marco Zuniga (TU Delft)
- Kay Römer (TU Graz)
- Hyung-Sin Kim (UC Berkeley)
- JeongGil Ko (Ajou University, Seoul)
- Jeongyeup Paek (Chung-Ang University, Seoul)

- Europe-heavy (and many here today)..- .. but some US and Asia 😊
- You're welcome to join!!

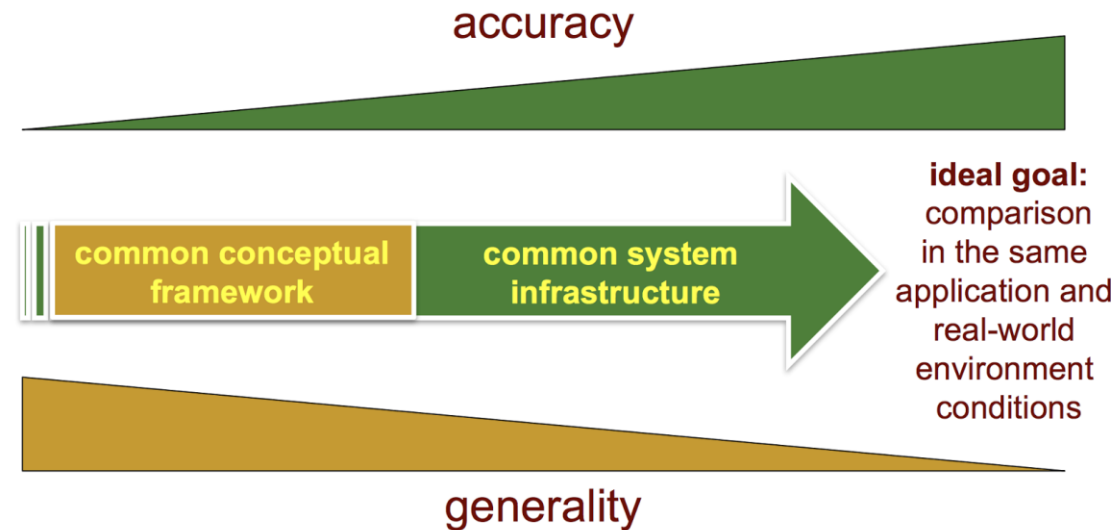


What we do

- Discuss what a benchmark might be
 - Talk with other communities (e.g. Robotics, DB)
 - Talk with IoT companies (5 were invited in Stockholm meeting)
 - Tackle research challenges (e.g. around reproducibility)
 - Define an initial benchmark
- Try to coordinate as a community
 - Sync with EWSN Dependability Competition
 - SenSys'16 poster
 - CPSBench: a workshop of CPSWeek'18
 - Dagstuhl?
 - Research grants..

Design space

- The accuracy – generality tradeoff



- Currently looking into two approaches
 1. Specification only (most general)
 2. Standardized testing architecture (most accurate)

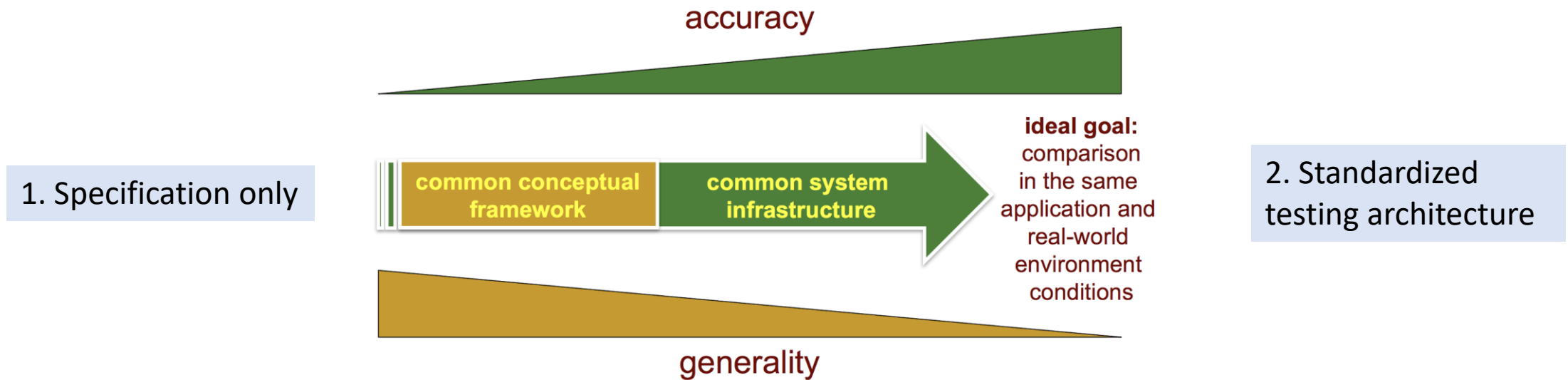
Approach #1: Specification only

- Metrics
 - Can be *input*, *observed*, or *output*
 - Ex. *input*: traffic load & pattern
 - Ex. *observed*: wireless noise
 - Ex. *output*: delivery, latency, energy
- Profiles
 - Assignment of concrete values to **input**
 - And interpretation of **observed** and **output**
 - Ex: data collection
 - Input: nodes: 100, #sources: 99, #destinations: 1, traffic load: from .1 msg/min to 1 msg/min, ...
 - Observe: link qualities, external interference, ...
 - Output: delivery, latency, ...

Approach #2: Standardized Testing Architecture

- Builds on the same concepts as #1: metrics and profiles
- Key idea: separate networking code from experiment scenario
 1. Node runs networking code
 2. Testbed runs experiment: e.g. uses GPIO/serial to instruct nodes and measure perf.
(similar to the Competition, but standardized/portable to any testbed)
- Benefit
 - Fully automated
 - Rules out mis-interpretation of profiles, etc
- Drawbacks
 - More complex and strict
 - More infrastructure maintenance

Design space – stepping back



- There are many more plausible design points..
- Which one should we focus on first?
- How to foster adoption?
 - Balance with comparability/repeatability

Conclusion



@iot_bench

- Thanks!
- Is this useful?
- Are we on the right track?
- Other ideas?
- Come and see our poster!
- Next, Ramona about the CPSBench workshop 😊

We are also organizing a workshop: CPSBench



- 1st Workshop on Benchmarking Cyber-Physical Networks and Systems
- Co-located with CPSWEEK (IPSN, RTAS, ...) in Porto on April 10, 2018
- Looks at benchmarking challenge from a CPS perspective: control, embedded, real-time, robotics, communication, machine learning, ...

<https://cpsbench2018.ethz.ch/>

CPSBench: Organizers and PC members

General Chair



Omprakash
Gnawali

TPC Co-Chairs



Sebastian
Trimpe



Marco
Zimmerling

Web Chair



Romain
Jacob

Publicity Chair



Ramona
Marfievici

Publication Chair



Anna
Förster

Tarek Abdelzaher (UIUC, USA)

José Araújo (Ericsson Research, Sweden)

Carlo Alberto Boano (TU Graz, Austria)

Berk Calli (Yale University, USA)

Patricia Derler (National Instruments, USA)

Simon Duquennoy (RISE SICS, Sweden)

Karl Henrik Johansson (KTH, Sweden)

Matteo Matteucci (Politecnico di Milano, Italy)

Ian Mitchell (University of British Columbia, Canada)

Luca Mottola (Politecnico di Milano, Italy)

Lothar Thiele (ETH Zurich, Switzerland)

Klaus Wehrle (RWTH Aachen University, Germany)

CPSBench: Aims and scope

- Bring together researchers and practitioners from the different CPS sub-communities to
 - Learn about each other's challenges and evaluation methodologies
 - Reflect on success stories (e.g., existing benchmarks in a particular community) or failures in using standard evaluation criteria
 - Debate future research agendas to jointly define the performance metrics and benchmark scenarios that matter from a system's perspective

Consider attending and drop us a line if you want to get involved!