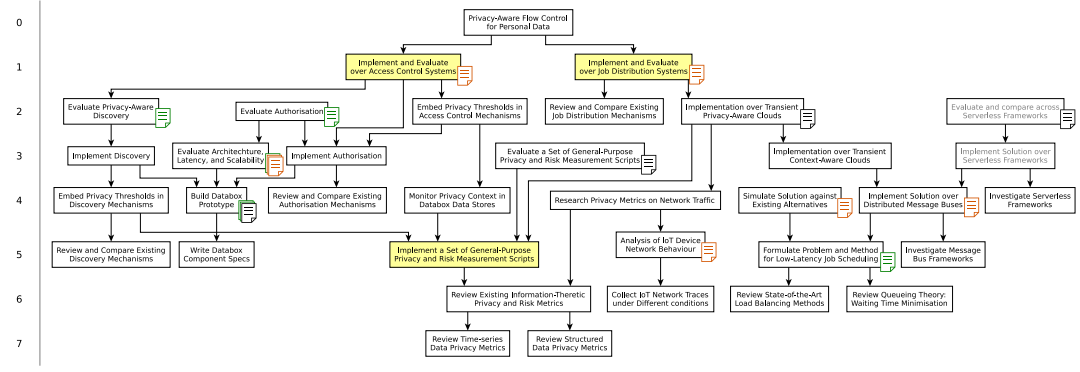
Privacy-Aware Load Balancing for Distributed Computation

Yousef Amar
QMUL Supervisor: Gareth Tyson
UniGe Supervisor: Lucio Marcenaro
Starting Date: 2016-01-16

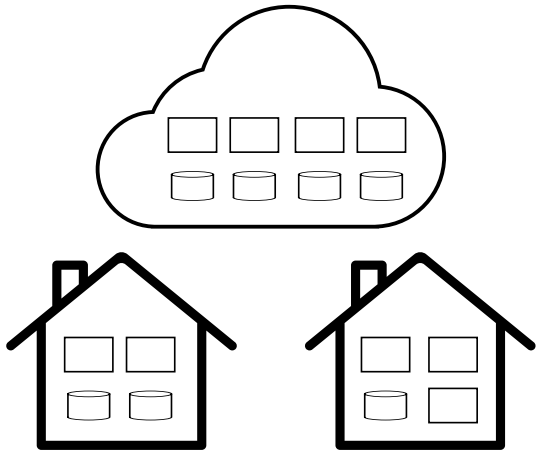
Previous Work

- Privacy-Aware Access Control
- Load Balancing on the Network Edge

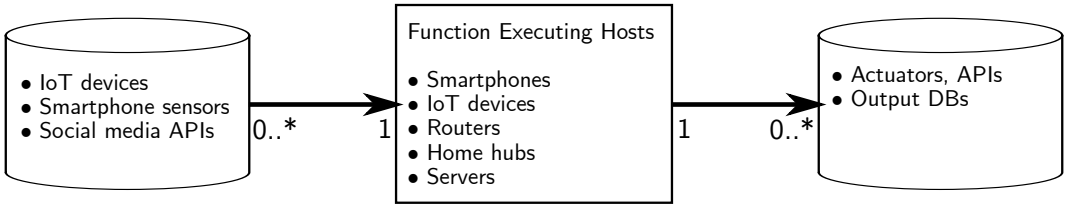


Research Context

- Data sources at the edge and in the cloud
  - IoT devices
  - Smart phones
  - Social media
- Computation shifting towards the edge
  - Untapped processing power
  - Near sources, low latency
  - An increase in privacy



Research Context



Research Problem

- Which host should execute a given function?
- Minimising factors such as
  - Latency
  - Information disclosure
  - Cost (e.g. power consumption, bandwidth usage)
  - Device capabilities
  - Other
- Tradeoff between these factors
- A system for tracking these factors and controlling job distribution

Related Work

- SOTA Load Balancing applicable to this context:
  - Beamer (NSDI 2018)
  - Maglev (NSDI 2016)
- Lots of research in privacy and information disclosure
- *None* combining the load balancing with privacy

Requirements and Constraints

- A system that distributes computation in a way that minimises response time, within the bounds of privacy constraints imposed by the user
- Overhead must be small enough for execution on edge and home IoT devices
- System must be versatile enough to run on a range of edge devices

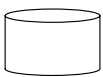
Design

Architechture — Symbols

Generic Function

Privacy-adding Function

Privacy-measuring Function

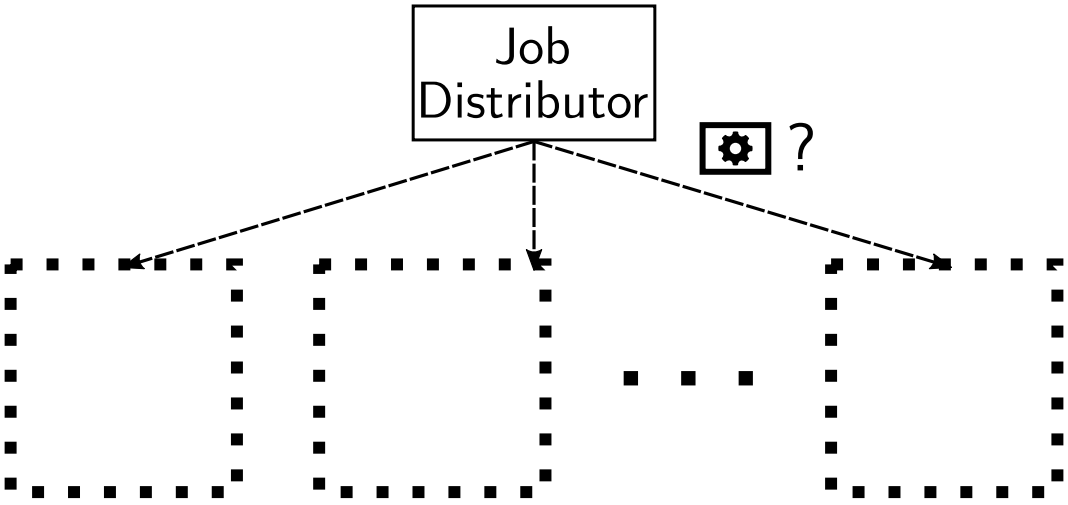


Data source

Private data source

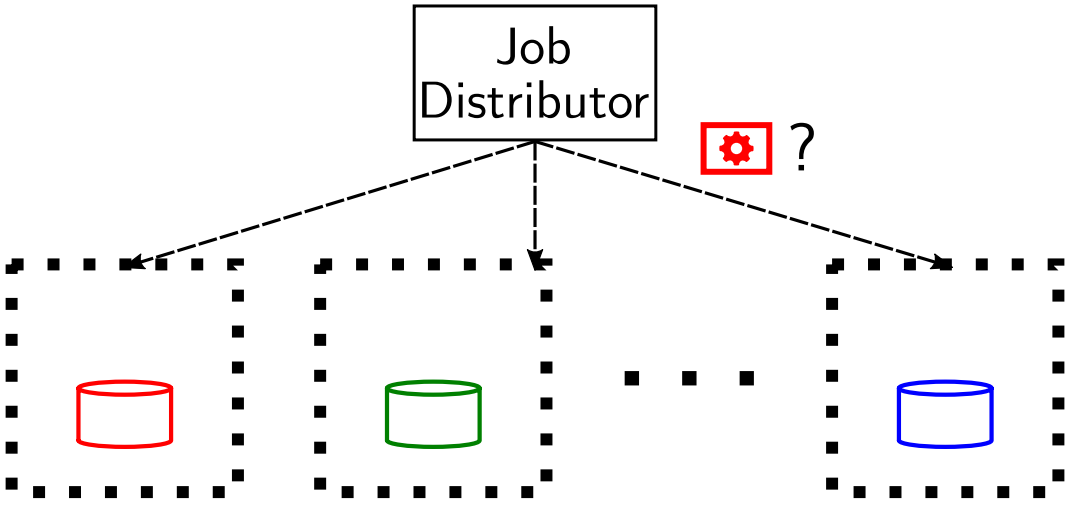
Method

No sources, no cross-host comms



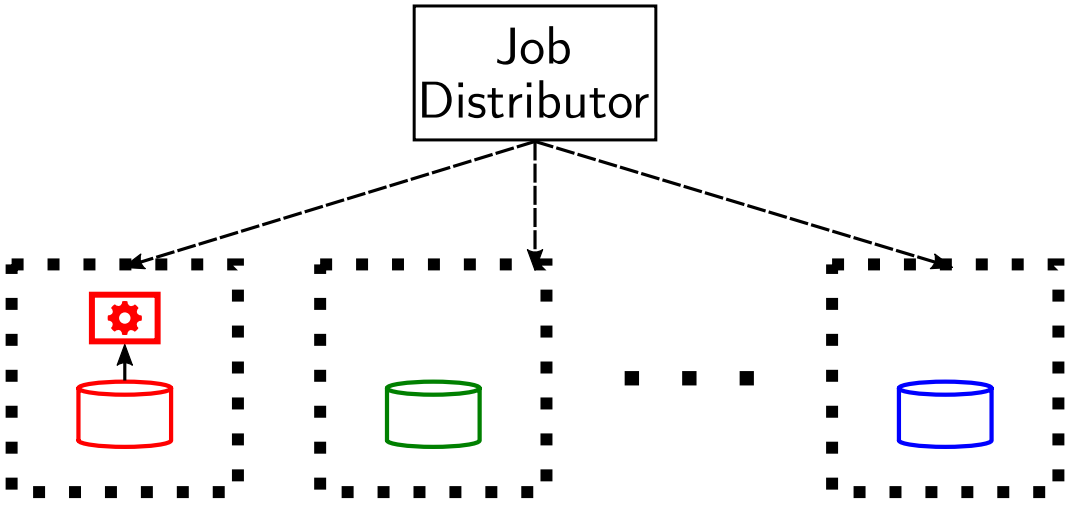
Method

Single-source, no cross-host comms, no linkage



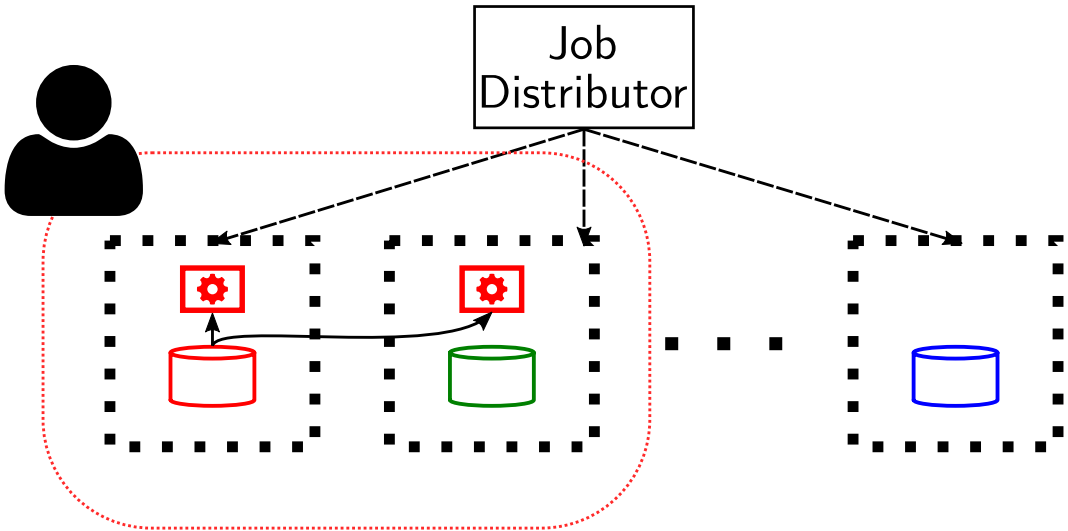
Method

Single-source, no cross-host comms, no linkage



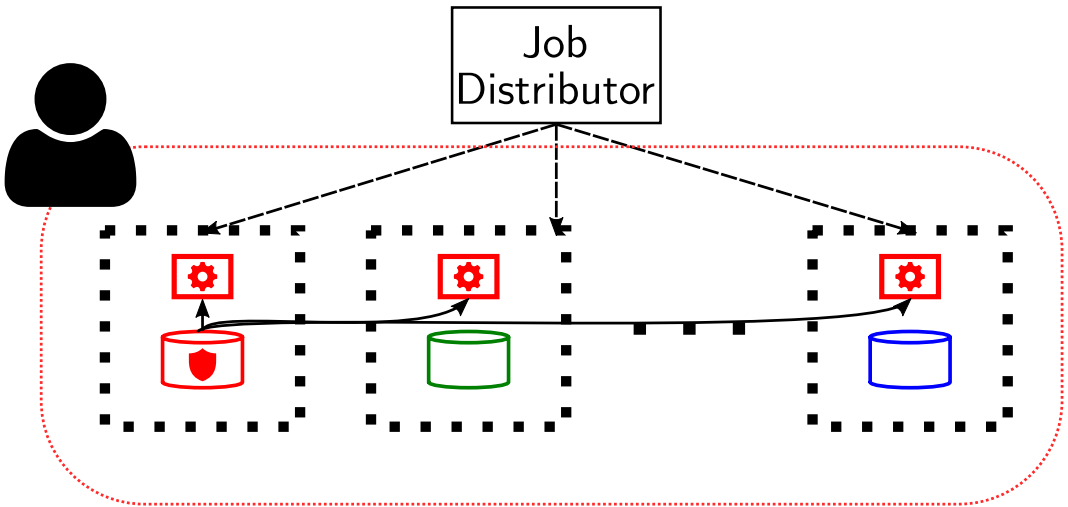
Method

Single-source, cross-source comms, host whitelist



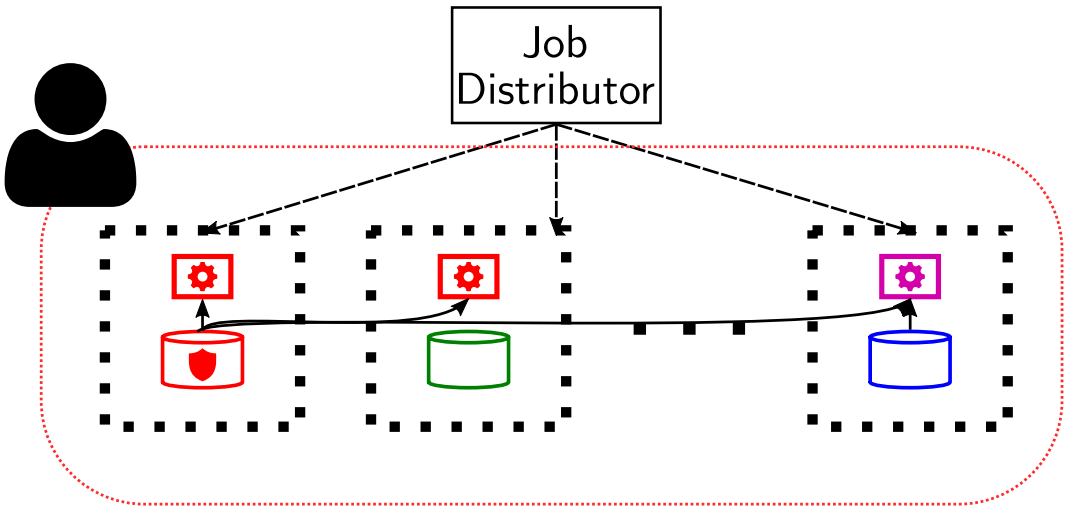
Method

Private sources allow larger whitelists



Method

Multi-source, cross-host sources

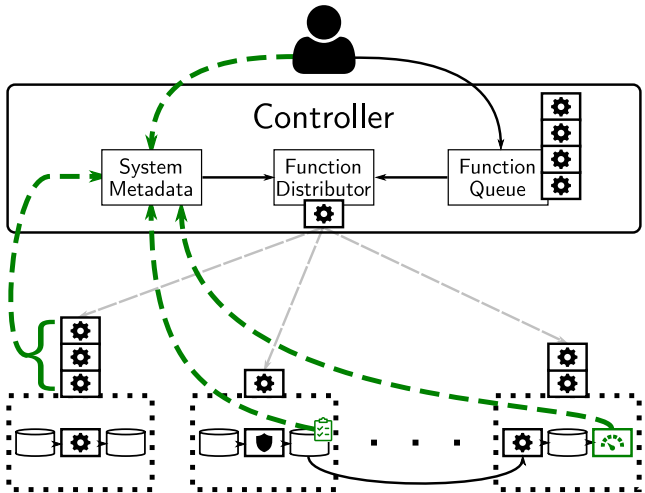


Method
Whitelist Generation

- User, community, and/or app provider list inferences for individual and combinations of data source access
- System lists risks vs utility, and user decides which risks are acceptable for the utlity they require
- System selects a subset of hosts that would satisfy these constraints
- From the system-side, classes of jobs simply have a subset of hosts available to execute on
- Problem reduces to whitelist generation on top of load balancing

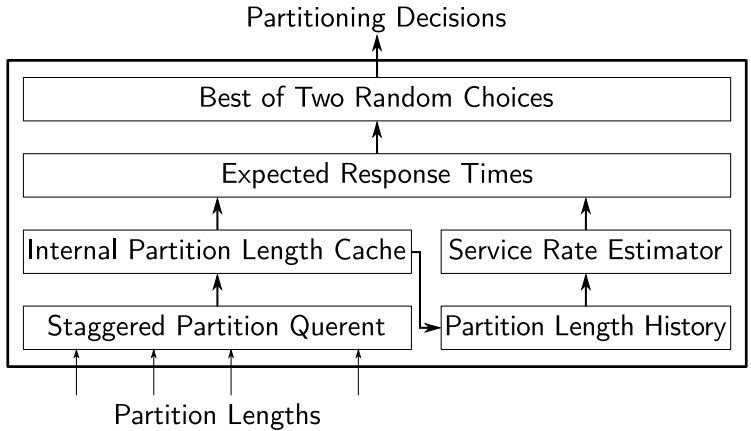
Method

Architechture – Job, data, and metadata pipelines



Method

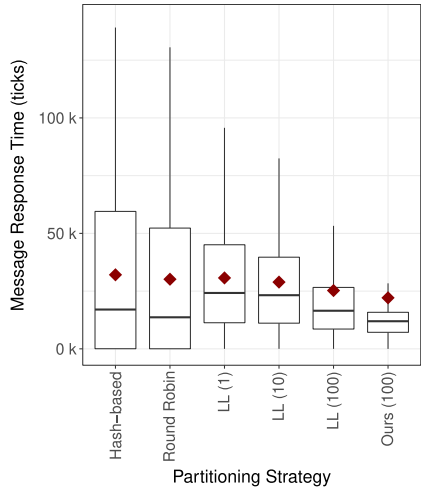
Load Balancing – Optimising for Performance



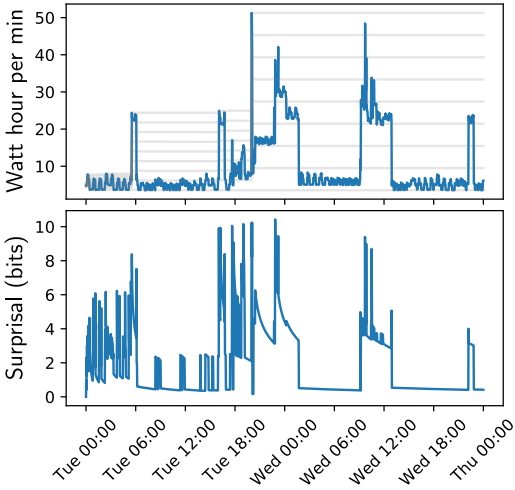
Method

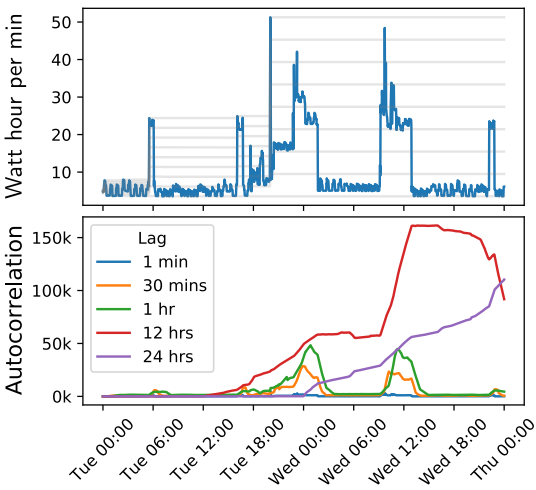
Load Balancing – Optimising for Performance

Message response times with different partitioning strategies as Tukey Box-Whisker plots. Whiskers show 1.5× inter-quartile range above (respectively below) the third (respectively first) quartile, whereas red diamonds show the mean values.



Privacy Heuristics





Past Evaluation

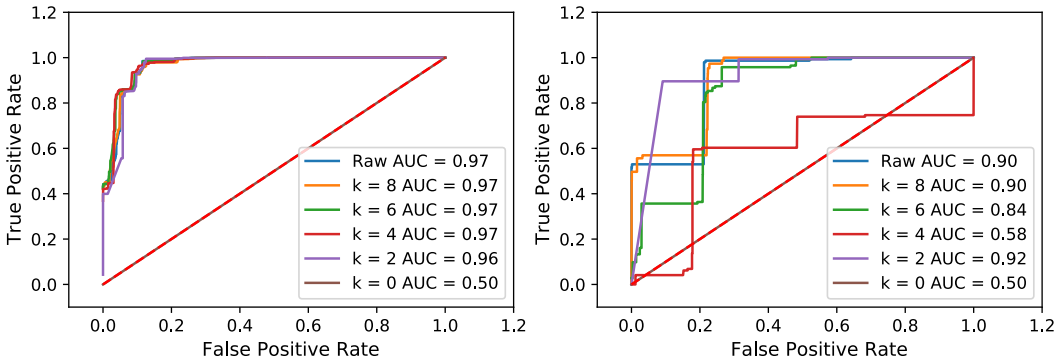


Figure: Receiver Operating Characteristic (ROC) curves for washer-dryer (utility; left) and microwave (attack; right)

Past Evaluation

- Gains in privacy
- Without impacting utility
- Negligible latency overhead
- Showing that the overhead on IoT devices is justifiable

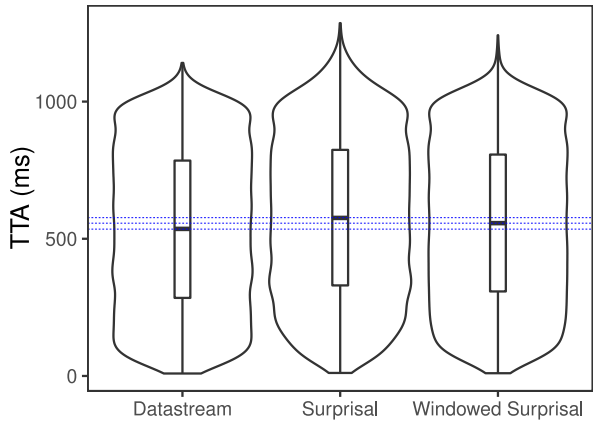


Figure: Distributions of time to availability under different conditions

Theis ToC

- Chapter 1: Privacy Measures
- Chapter 2: Privacy-Aware Access Control
- Chapter 3: Edge Load Balancing
- Chapter 4: Privacy-Aware Edge Load Balancing

Publications

- 2018 – Providing Occupancy as a Service with Databox *In Proceedings of the 1st ACM International Workshop on Smart Cities and Fog Computing* (pp. 29-34) ACM
- 2018 – An Information-Theoretic Approach to Time-Series Data Privacy *In Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems* (p. 3) ACM
- 2018 – Building accountability into the Internet of Things: the IoT Databox model *Journal of Reliable Intelligent Environments* (pp. 1-17) Springer
- 2017 – Balanced Message Distribution in Distributed Message Handling Systems *US Patent (serial number: 15/794440)*
- 2017 – Route-based authorization and discovery for personal data *In the 11th EuroSys Doctoral Workshop*
- 2016 – Personal Data Management with the Databox: What's Inside the Box? *In Proceedings of the 2016 ACM Workshop on Cloud-Assisted Networking* (pp. 49-54) ACM
- 2016 – Privacy-aware infrastructure for managing personal data *In Proceedings of the 2016 ACM SIGCOMM Conference* (pp. 571-572) ACM
- 2015 – Incremental dense multi-modal 3d scene reconstruction. *In Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on* (pp. 908-915) IEEE

Conference and Workshop Attendance

- W-P2DS 2018 (presenting)
- EuroDW 2018 (presenting)
- Databox Annual Symposium (presenting)
- EuroDW 2017
- EuroSys 2017
- SIGCOMM 2016
- NGN-MSN Coseners 2016
- UK HDAN Research Roadmap Workshop

Awards and Scientific Recognition

- EuroSys 2018 Conference Grant
- EuroSys 2018 Shadow PC Grant
- EuroSys 2017 Grant
- IEEE Circuit Building Competition Winner
- SIGCOMM 2016 Grant
- `yousefamar.com#awards`

Miscellaneous

- amar.io/skills-points-record.png
- Involvement in Databox Project databoxproject.uk
- Internship at Nokia Bell Labs Stuttgart
- Time at UniGe and areas of overlap

Thank you for your attention!