E2E-Loader: A Framework to Support Performance Testing of Web Applications

E. Battista, S. Di Martino, S. Di Meglio, F. Scippacercola, and **L. L. L. Starace**

Università degli Studi di Napoli Federico II, Italy

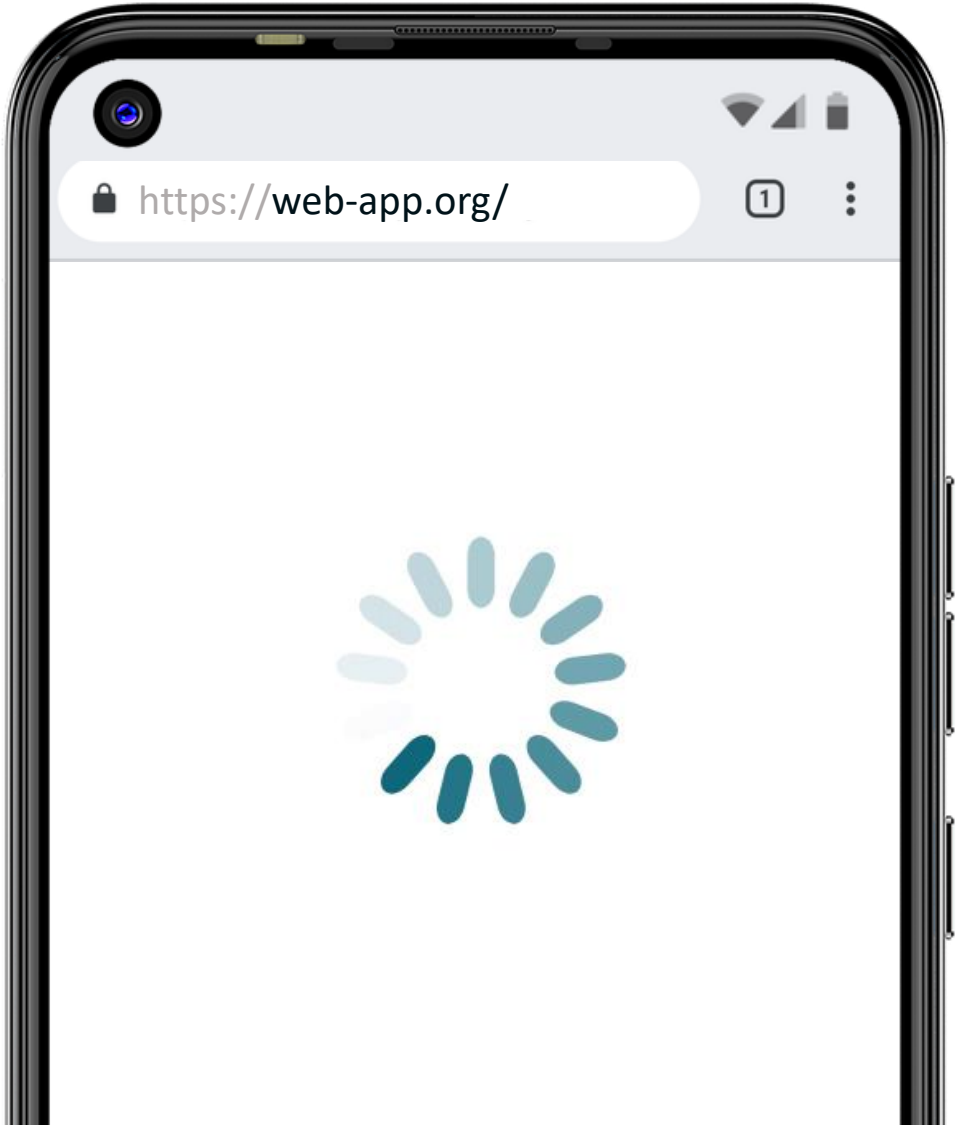
luigiliberolucio.starace@unina.it

https://luistar.github.io/

16th International Conference on Software Testing (ICST 2023), Dublin, Ireland

Ensuring the Quality of Web Applications

- Functional Testing is crucial
- But does the web app behave as expected also **under load**?
  - Quality issues related to performance aspects
  - Functional issues might only manifest under certain load conditions



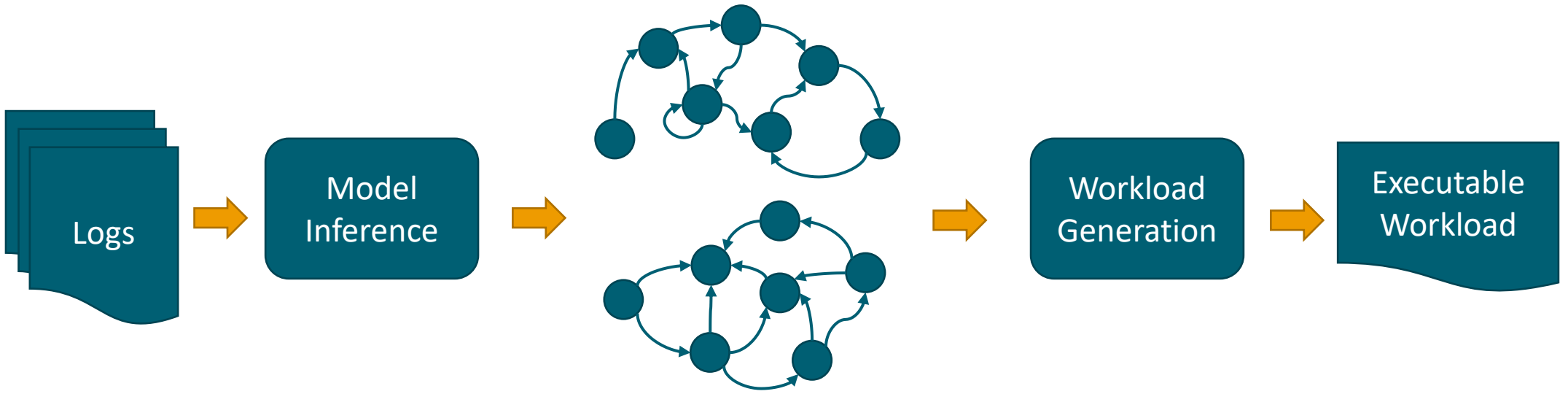
Performance Testing

- **Goal**: Uncover load-related issues.
- **How**: Generate workloads for the System and monitor its behaviour
- For Web Apps, a workload is a specific sequence of web requests
  - Made from a number of concurrent users
  - Each performing its own use cases

- How do we generate workloads?
  - Manually (does not scale very well)
  - Automatically, via driver-based executable scripts

Related Works

- Most related works leverage **log analysis** [1]



Model of user behaviours

[1] Parrott, Chester, and Doris Carver. "Lodestone: A Streaming Approach to Behavior Modeling and Load Testing." *2020 3rd International Conference on Data Intelligence and Security (ICDIS)*. IEEE, 2020.

Limitations

Logs needed

- Can’t be applied before the system (or new features) are deployed

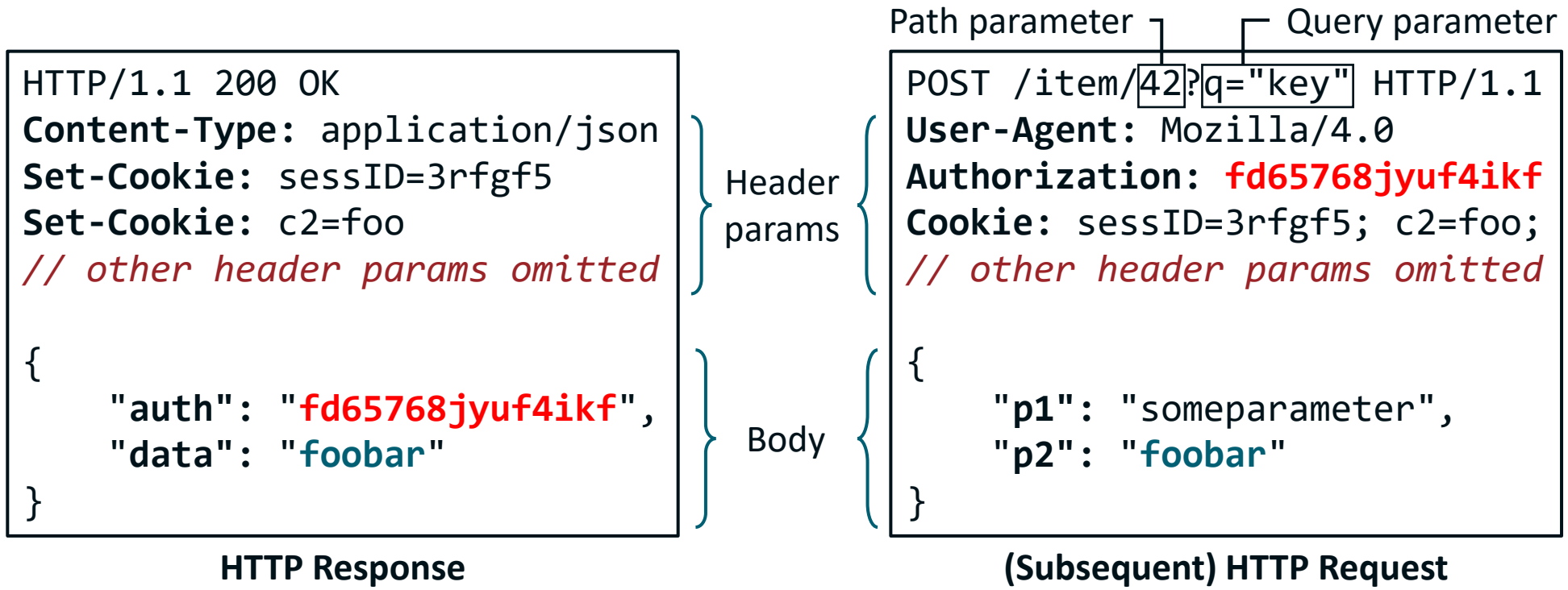
Little support for emerging protocols (e.g.: WebSocket)

- More manual work needed

Little support for managing data dependencies between responses/requests

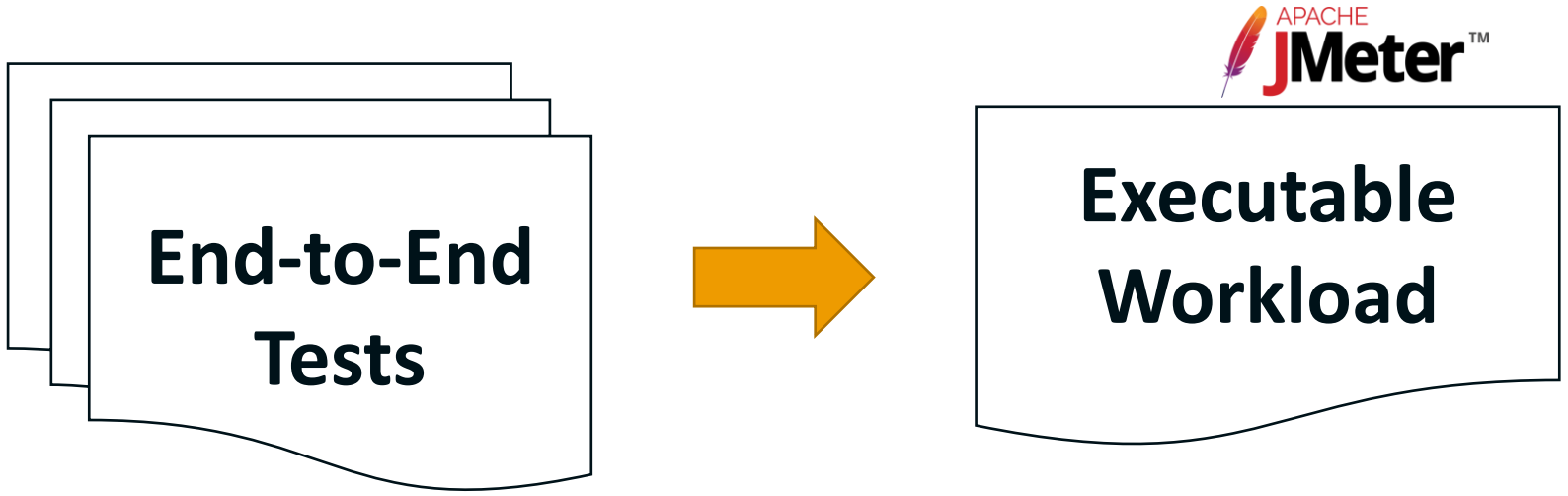
- More manual work needed

Data Dependencies

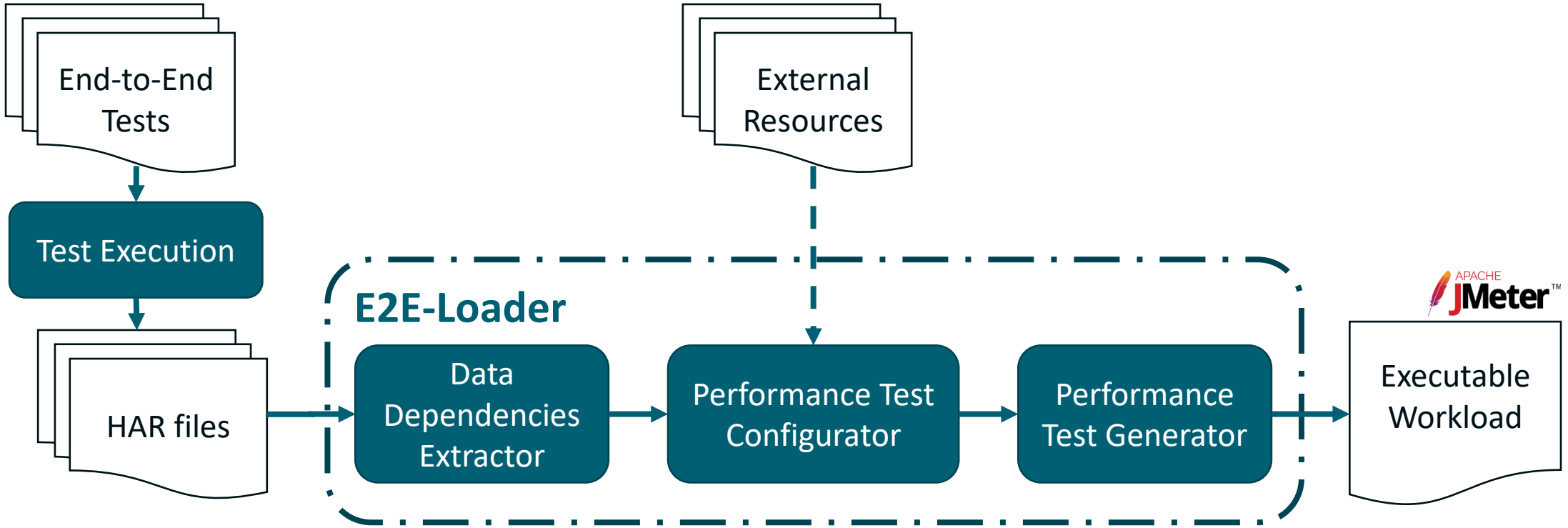


Proposed Approach: E2E-Loader

Key idea: Leverage **existing E2E functional tests**, representing meaningful user behaviours, to support the generation of **executable workloads**



Proposed Approach: E2E-Loader



Managing Data Dependencies

- Two steps process:
  1. Get candidate data dependencies by matching params with the same value
  2. Refine results by considering also parameter names
     - Alternatives sorted by Levenshtein distance
- Performance Test Configurator GUI allows practitioners to
  - Confirm/discard/fix automatically detected data dependencies
  - Add new dependencies that were not detected
  - Make the tests parametric, leveraging external resources (CSV files)

Empirical Evaluation: Industrial Case Study

RQ: Are workloads generated with E2E-Loader comparable to those manually-generated by practitioners using state of the art tools?

Web App Under Test

Modern app, 50 Microservices, Orchestrated with Kubernetes

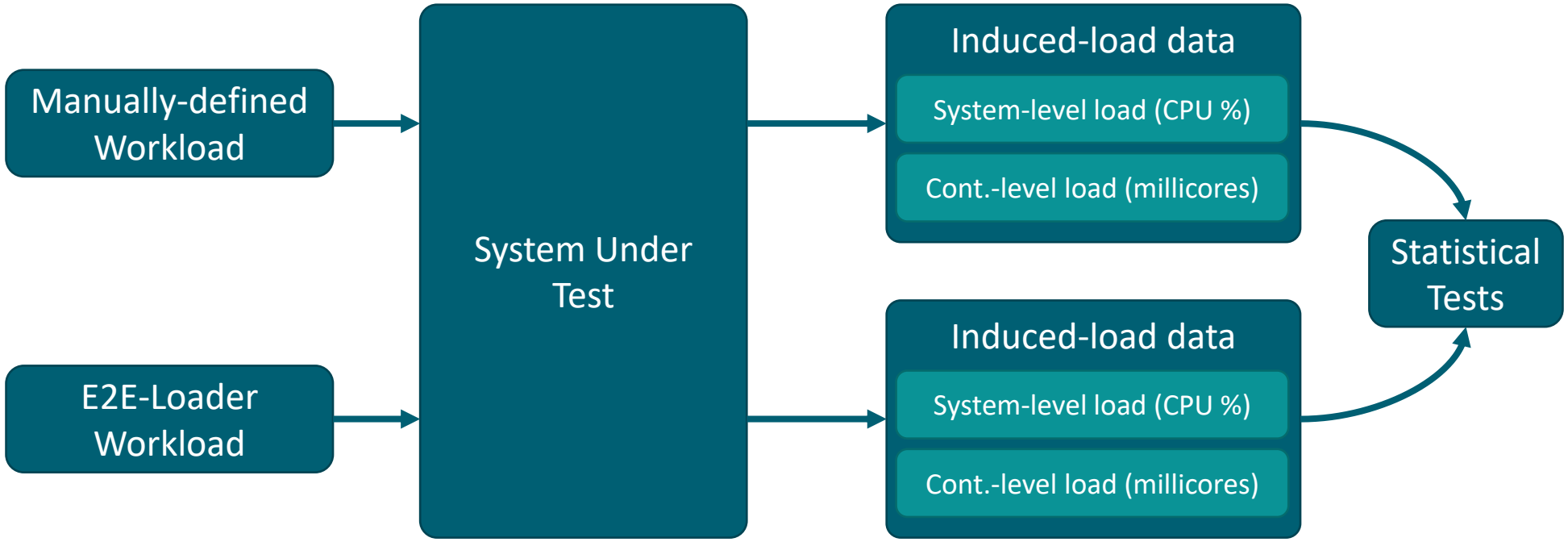
5 Workloads

Defined Manually By Practitioners using JMeter

Existing E2E Functional Tests

Cypress-based

Empirical Evaluation: Procedure



Statistical Tests

- We compare the load induced by the manually defined workload and the one induced by the workload obtained using E2E-Loader
- The induced loads are basically sequences of CPU levels over time
- Wilcoxon Signed Rank Test is used [1]
- H0: The load induced by manually-defined workloads has the same distribution as the one induced by E2E-Loader workloads
- If statistically significant differences exist, we measure the effect size using Cohen's δ

[1] Chen et al., "An experience report of generating load tests using log-recovered workloads at varying granularities of user behaviour," in 2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE).

Empirical Evaluation: Results

- Workloads obtained with E2E-Loader induce similar loads w.r.t. those manually obtained with JMeter
  - In all cases, at CPU level
  - In 93% of the cases, at millicore (container) level

	Workloads				
	1	**2**	**3**	**4**	**5**
p-value	0,54	0,66	0,15	0,07	<0,05
Eff. size	-	-	-	-	small

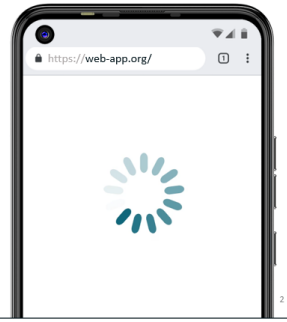
Future Works

- Compare against log-based approaches
- Quantify improvements in productivity
- Replicate evaluation on a broader set of subject systems

E2E-Loader: A Framework to Support Performance Testing of Web Applications

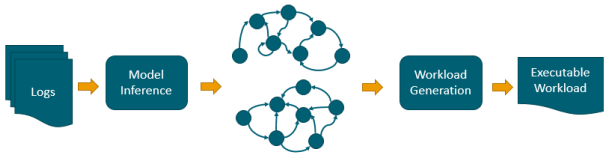
Ensuring the Quality of Web Applications

- Functional Testing is crucial
- But does the web app behave as expected also **under load**?
  - Quality issues related to performance aspects
  - Functional issues might only manifest under certain load conditions



Related Works

- Most related works leverage **log analysis** [1]



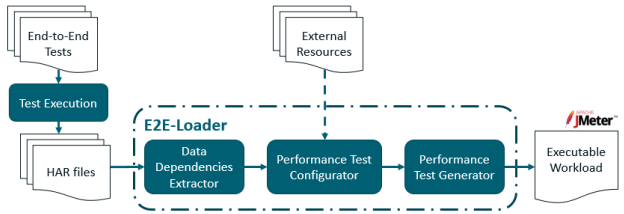
Model of user behaviours

[1] Parrott, Chester, and Doris Carver. "Lodestone: A Streaming Approach to Behavior Modeling and Load Testing." *2020 3rd International Conference on Data Intelligence and Security (ICDIS)*. IEEE, 2020.

Limitations

Logs needed	Little support for emerging protocols (e.g.: WebSocket)	Little support for managing data dependencies between responses/requests
• Can't be applied before the system (or new features) are deployed	• More manual work needed	• More manual work needed

Proposed Approach: E2E-Loader



Empirical Evaluation: Industrial Case Study

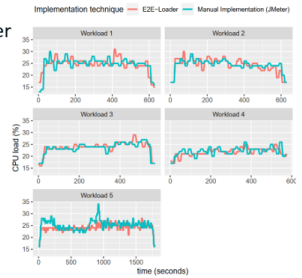
RQ: Are workloads generated with E2E-Loader comparable to those manually-generated by practitioners using state of the art tools?

Web App Under Test	5 Workloads	Existing E2E Functional Tests
Modern app, 50 Microservices, Orchestrated with Kubernetes	Defined Manually By Practitioners using JMeter	Cypress-based

Empirical Evaluation: Results

- Workloads obtained with E2E-Loader induce similar loads w.r.t. those manually obtained with JMeter
  - In all cases, at CPU level
  - In 93% of the cases, at millicore (container) level

	Workloads				
	1	2	3	4	5
p-value	0,54	0,66	0,15	0,07	<0,05
Eff. size	-	-	-	-	small



Luigi Libero Lucio Starace
luigiliberolucio.starace@unina.it