Introduction to Artifact Evaluation

Neea Rusch
Augusta University

3 April 2024

What is an Artifact?

An artifact is a supplement that extends beyond a scientific paper and supports the claims or results of that paper.

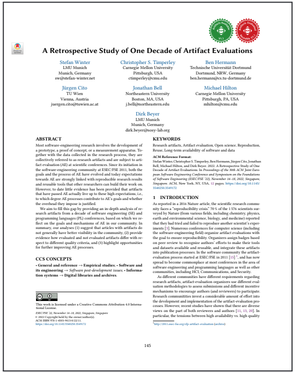
Artifact may contain: software, mechanized proofs, test suites, data sets, benchmarks, video of a difficult/impossible-to-share system in use, hardware, or any other artifact described in a paper.

An artifact captures a point-in-time matching the paper – it should be packaged for long-term preservation to facilitate future research.

Artifact Evaluation Motivations

- Encourage and support authors to provide supplements to papers
- Help future researcher to build on and compare with previous work
- Validate claims and results presented in a paper
- Reward authors who put in effort to create useful artifacts
- Recognize the effort to release usable software systems

Papers with artifacts are recognized with badges.



Historical Background

Insufficient respect paid to the artifacts that back papers.

Areas so centered on software, models, and specifications should want to evaluate artifacts as part of the paper review process.

Not examining artifacts enables everything from mere sloppiness to, in extreme cases, dishonesty.

More subtly, it also imposes a subtle penalty on people who take the trouble to vigorously implement and test their ideas.

---

https://artifact-eval.org/motivation.html

Historical Background

In 2011, Andreas Zeller, the program chair for ESEC/FSE, decided to institute a committee to address this problem. Andreas asked Carlo Ghezzi and Shriram Krishnamurthi to run this process.

Shriram had long wanted to create such a committee and call it the "Program Committee" (ha, ha). However, not only is that name taken, we also wanted to be open-minded to all sorts of artifacts that are not programs […]. We therefore called this the ***Artifact Evaluation Committee (AEC)****.*

---

https://artifact-eval.org/motivation.html

Artifact Evaluation Today

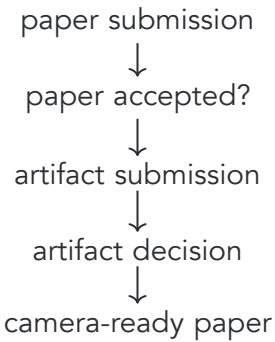
Software Engineering & Programming Languages
ICSE, ESEC/FSE, ASE, ECOOP, ISSTA, OOPSLA, POPL, PLDI, ICFP, SAS, ESOP, TACAS, CAV, TSE, TOSEM, EMSE, TOPLAS…

Systems research
SOSP, USENIX ATC, EuroSys, FAST, OSDI, SC…

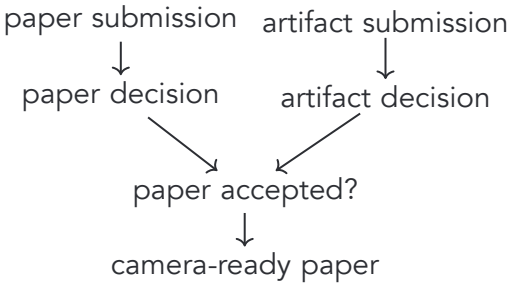
Security research
ACSAC, CHES, CCS, NDSS, USENIX Security, SysTEX, WOOT…

High-Level Workflow

Common



Alternative



Artifact Evaluation Process

Time	Step	Responsible party
	Artifact submission	Authors
2–5 days	Bidding	AEC members
	Artifacts assigned (usually 2–3)	AEC chairs
1–2 weeks	**Phase 1:** Kick the tires	AEC members
1–2 weeks	Author responses, possible fixes	Authors
2–4 weeks	**Phase 2:** Full review	AEC members
3–7 days	Discussion and badging decisions	AEC members
	Decisions announced	AEC chairs

Expect an artifact to take on average 8h to review completely.

Badges

Artifacts are evaluated against badging criteria, in up to three categories (available, evaluated, results validated). The current commonly applied criteria is ACM Artifact Review and Badging policy v1.1[1].

[1] https://www.acm.org/publications/policies/artifact-review-and-badging-current

Badges: Available

Available Author-created artifacts relevant to the paper have been placed on a publicly accessible *archival repository*[2]. A DOI or link to this repository along with a unique identifier for the object is provided.

[2]Archival repositories: Zenodo (zenodo.org), Figshare (figshare.com), Software Heritage (softwareheritage.org), Dagstuhl Artifacts Series (DARTS),…

Badges: Artifacts Evaluated

Two levels are distinguished, only one of which should be applied.

Functional The artifacts associated with the research are found to be *documented, consistent, complete, exercisable*, and include appropriate evidence of verification and validation.

Reusable The artifacts have all the qualities of Functional level, but, in addition, they are very carefully documented and well-structured to the extent that reuse and repurposing is facilitated.

Badges: Results Validated

The main results of the paper have been successfully obtained by a person or team other than the author.

Reproduced The main results of the paper have been obtained in a subsequent study by a person or team other than the authors, *using, in part, artifacts provided by the author.*

Replicated The main results of the paper have been independently obtained in a subsequent study by a person or team other than the authors, *without the use of author-supplied artifacts.*

Artifact Evaluators

AEC members are usually senior graduate students, postdocs, or recent PhD graduates.

Among researchers, experienced graduate students are often in the best position to handle the diversity of systems expectations that the AEC will encounter.

Graduate students represent the future of the community, so involving them in the AEC process early will help push this process forward.

---

https://pldi24.sigplan.org/track/pldi-2024-pldi-research-artifacts

Benefits of Participating in AECs

- Early experience in peer review process, learn how to write reviews
- Early access to cutting-edge works at top conferences
- Gain exposure to new research topics
- Develop intuition for what a top-conference publication requires
- Learn the artifact process and improve quality of own artifacts
- Start recognizing researchers, research trends, etc.
- Service experience for your CV

General Artifact Preparation Tips

For all artifacts:

- Make artifact claims explicit in the artifact readme
- Try prepare a push-button/single command evaluation
- The artifact should support regenerating experimental claims[3]
- Remember a license

[3] If paper contains tables or figures of measurements, include scripts to regenerate them.

General Artifact Preparation Tips

For artifacts requiring high computational resources:

- Prepare a partial (small/short) evaluation + full evaluation
- Try include critical functionality in the partial evaluation
- Give time estimates of all long latency tasks
- Provide static files with full result details for manual inspection

Generally: need evidence to show the artifact produces claimed results.
Sometimes, video or alternative approaches are helpful to demonstrate this.

General Artifact Preparation Tips

Artifact should be useful long-term (10 years+):

- Provide a container or VM that captures the expected environment
- Detail software dependencies including versions
- Make artifact self-contained: avoid external references that may change or get deleted