SAS2013 Artifact Submission Experience Report

Manuel Fähndrich and Francesco Logozzo

July 9, 2013

CFP and Submission

This year, for the first time, SAS invited the submission of virtual machine artifacts in support of submitted papers. The call for papers worded the invitation as follows:

> New this year, we are encouraging authors to submit a virtual machine image containing any artifacts and evaluations presented in the paper. The goal of the artifact submissions is to strengthen our field's scientific approach to evaluations and reproducibility of results. The virtual machines will be archived on a permanent Static Analysis Symposium website to provide a record of past experiments and tools, allowing future research to better evaluate and contrast existing work.

> Artifact submission is optional. Details on what to submit and how will be forthcoming.

> The submitted artifacts will be used by the program committee as a secondary evaluation criteria whose sole purpose is to find additional positive arguments for the paper's acceptance. Submissions without artifacts are welcome and will not be penalized.

We sent instructions on how to submit artifacts after the final paper deadline to all authors who submitted a paper. The deadline for artifact submission was one week after the final paper submission. The instructions for artifact submissions were as follows:

> If you plan to submit a VM, those are the steps to follow.

> 1. Prepare a virtual machine using some widely available software, running on most platforms. Examples are VirtualBox, Hyper-V, and VMware.

> 2. The virtual machine should contain:
>   - A self-contained prototype, mentioned in the paper, *in binary form* (ex., myanalyzer.exe)
>   - The benchmarks used in the paper (ex. difficultTest.c)

- A text file explaining how to reproduce the results in the paper (e.g., run "myanalyzer.exe /analyze:difficulTest.c /iterations:3")

3. Put the VM in some location reachable from us (e.g., your website, skydrive, dropbox, etc.)

4. Send the URL with the VM to (our email address). The email subject should mention the SAS paper number (e.g., VM for SAS submission 1234)

5. Expect for an acknowledgment from us (to be on the safe side and be sure your email was not mistakenly captured by the spam filter)

The motivation to submit artifacts as complete VMs was to provide flexibility to the authors in terms of operating system etc, as well as to ensures reviewers and future users of the archive would be able to examine the artifacts without hardware/OS dependencies.

The authors seemed positive about the submission process and made the extra effort of submitting VM images, with instructions on how to use them. Upon submission, we made sure that, at a minimum, we were able to start the VM—that was not always the case at the 1st attempt. During the submission process, some authors had trouble preparing VMs and were in email contact with the program chairs. Some authors were able to resolve issues and ended up submitting VMs, others didn't due to technical or time constraints. During this process, we reassured authors that not submitting an artifact was not penalizing the submission. We ended up receiving 22 usable artifacts out of the 56 submitted papers, i.e., 40% of submissions included an artifact.

All but one artifact was submitted using VirtualBox (`www.virtualbox.org`), an open virtual machine environment available on all current platforms. VirtualBox was also the easiest environment to use for starting the VMs. We would thus encourage future submission in VirtualBox.

Evaluation

During the review process, we encouraged the reviewers to look at the submitted VMs and suggested some criteria to consider during the VM exploration, such as:

- Did you run the VM?
- Did you run the experiments?
- How much time did you spend playing with the VM/experiments?
- How do the experiments support the paper?
- Is it clear what the experiments measure/produce ?
- Can the experiment be changed and run?
- Did you play with alternative ways to run it? (new test problems, small variations)

- - Other (positive) observations
- - Anything you want to tell us about the VM experiment

Some reviewers used the artifacts to enrich their review. As PC chairs, we did make sure that in reviews and discussions, artifact evaluation was not used to negatively influence the evaluation, as we promised in the call-for-papers. This generally was not an issue at all.

The existence of artifacts gave more confidence about the experimental results of a submission, enabled the reviewers to answer some questions about the papers, and in the case of papers leaning towards rejection provided another mechanism to save the paper. Here are a few excerpts from reviews of accepted papers:

> "I appreciate that the authors have submitted a VM. It was easy to rerun the experiment. Although the result is not identical to what is reported in the paper (aes, sha1 and lzss show better figures than the paper while srcode shows worse), it generally supports the paper.
>
> Also, I was able to solve some questions (mentioned in the minor comments below) regarding the Coq code. Thanks!"

Another reviewer said:

> "As regards the provided Virtual Machine, it was easy to install and execute. I used it to analyze the very few examples provided with the virtual machine, as well as some other examples (mainly to draw conclusions on the limitations)."

The VMs also were used during the discussion phase:

> "In terms of practice, this paper was clearly the strongest in my pile. This may be because the authors have built up a very strong tool chain with lots of things that others cannot match right now. But, that is not a negative—especially since they did submit a VM (even if not a publicly available tool)."

Out of the 23 accepted papers, 11 had associated VMs, i.e., 48%, thus a slightly higher percentage than the percentage of artifacts among the total submissions. We are pleased with that outcome as it lends support to the idea that artifacts may help paper acceptance, but should not penalize the paper.

Archival

We gave authors of accepted papers a chance to revise the VMs submitted for archival, as well as the option to opt-out of the archival completely. Only VMs of accepted papers were archived. These are accessible at `http://staticanalysis.org/` as a scientific record of the state of the art at this point in time and they will hopefully serve as a comparison base for future research.

A big shout of thanks goes to Manuel Hermenegildo, who set up the static-analysis.org website to host the SAS 2013 artifacts.

Conclusion

Among the different virtual machines submitted, we found The SAS'2013 VM evaluation was less structured than other recent artifact submission experiments [1] (FSE, ECOOP, and OOPSLA). We viewed the VMs and their evaluation more in the light of how PCs use author-responses to reviews. Author-responses can be considered or ignored by PC members as seen fit. As a result, we did get a more superficial and sparse evaluation of the VMs than the experiments conducted at FSE, ECOOP, and OOPSLA, where a separate artifact evaluation committee reviewed the artifacts. In [1], the evaluation of the paper and the artifact was considered completely separate, with a chinese wall between the reviewers of each side. Thus, in their approach, artifacts cannot influence a paper's acceptance. In contrast, we wanted the artifacts to be part of the paper submission rather than their own separate submission. Another difference is that we are archiving artifacts in order to keep a record of the state of the art in our field.

Overall, the artifact submission and evaluation for SAS'2013 was successful and we were happy with the outcome. Clearly, there are many ways artifact evaluation for software conferences can be improved in future experiments. The motivation for authors to submit artifacts in our view is sufficient given that it has the ability to increase the chance of acceptance of the paper. We did not need any prizes to entice 40% of authors to submit an artifact.

On the other hand, the evaluation of the artifacts in our view could be improved vastly. Since artifacts should influence the PC, we see it necessary that the PC is involved in the artifact evaluation. But it isn't clear how to encourage the PC members to take the time to evaluate the artifact, as they are already burdened with paper reviewing. One possible approach might be to assign each paper to 3 PC members for the normal paper evaluation and to 1 additional PC member for artifact evaluation. The artifact PC member should read the paper as well, but writes an artifact evaluation, rather than a paper evaluation. The other PC members are free to play with the artifact as well, but they can also ask the artifact PC member to try to anwer certain questions they have about the artifact.

Manuel Fahndrich and Francesco Logozzo

References

[1] Shriram Krishnamurthi, "Artifact Evaluation for Software Conferences", `http://cs.brown.edu/~sk/Memos/Conference-Artifact-Evaluation/`