
[My Submissions](#)[pap241 Review Details](#)

Submission: P*: A Model of Pilot-Abstractions

Contributors: Jha, Luckow, Mantha, Merzky, Santcroos, Weidner

Summary of reviews of pap241s2

Reviewer

[Reviewer 1](#)[Reviewer 2](#)[Reviewer 3](#)[Reviewer 4](#)

Committee Comments [jump](#)

Review of pap241s2 by Reviewer 1

[top](#)

Summary and High Level Discussion

Indicate the topic of the paper and summarize the importance of the contribution. List the major points of the paper and your evaluation of them including problem and motivation, related work, main idea(s), algorithms and techniques, experiments and experimental method, results, etc. **If the paper is an SOP submission, indicate the secondary area of contribution (if any) from among the remaining seven areas** (Algorithms; Applications; Architectures and Networks; Clouds and Grids; Performance, Energy, and Dependability; Programming Systems; Storage, Visualization, and Analytics). **Summarize the importance and novelty of the contribution towards the recording and dissemination of the state-of-the-art in the practice of HPC.** List the major experiences and observations of the paper and your evaluation of their strengths and weaknesses.

The paper first introduces a common abstraction model (P*) for Pilot-Jobs. Then it show that P* can map various existing Pilot-Jobs production frameworks. The last part of the article describes the Pilot-API, a Python-based API which adheres to P* and which provides a common unified API to existing Pilot-jobs frameworks. This API enables concurrent usage of multiple Pilot-Jobs frameworks, providing some form of interoperability between these frameworks.

Comments for Rebuttal

Use this section to clearly Indicate the one or two major issues that you would like the author(s) to address in the rebuttal. This is prose that the authors will see and use in providing their rebuttal which is limited to 750 words. Please provide clear and specific text to guide the response by the author(s).

Section V-B: the text references figure elements using a terminology which differs from the one used for figure axis labels: C1, C2, HTC, ... while the figure axis are labelled BigJob/XSEDE, DIANE/EGI, XSEDE:Trestles/FG, etc. This is confusing for the reader. Please provide a common naming convention.

Detailed Comments for Authors

Provide detailed comments on the presentation, organization, grammar, typographical errors, clarity problems with tables and graphs, or detailed technical points. This is prose that the authors will see and it helps to be specific here.

- A short definition of "Pilot-Jobs" in the abstract would help those readers which do not read the whole paper

- sentences a bit difficult to understand: II-C: "The PM which is responsible for mapping DUs to SUs and for conducting decision regarding resource assignments." missing verb? III-D: "... the different PJ frame- works have different usage modalities mainly cause by the fact that each PJ framework has evolved..."

Summary and High Level Discussion

Indicate the topic of the paper and summarize the importance of the contribution. List the major points of the paper and your evaluation of them including problem and motivation, related work, main idea(s), algorithms and techniques, experiments and experimental method, results, etc. **If the paper is an SOP submission, indicate the secondary area of contribution (if any) from among the remaining seven areas** (Algorithms; Applications; Architectures and Networks; Clouds and Grids; Performance, Energy, and Dependability; Programming Systems; Storage, Visualization, and Analytics). **Summarize the importance and novelty of the contribution towards the recording and dissemination of the state-of-the-art in the practice of HPC.** List the major experiences and observations of the paper and your evaluation of their strengths and weaknesses.

The paper describes a generalization of the pilot frameworks (as exemplified by the Condor glide-in). The authors describe the models, map various existing frameworks to the model, propose an interoperable API for it, and demonstrate that it can be used to provide interoperability between important infrastructures running pilot jobs. They also attempt to generalize the pilot job framework to data management.

Comments for Rebuttal

Use this section to clearly indicate the one or two major issues that you would like the author(s) to address in the rebuttal. This is prose that the authors will see and use in providing their rebuttal which is limited to 750 words. Please provide clear and specific text to guide the response by the author(s).

The strength of the paper is that it attacks an important problem: there is a proliferation of pilot frameworks, each working in essentially the same way, so providing a unifying model and underpinnings of interoperability can potentially simplify moving between infrastructures using different implementations.

A significant weakness of the paper is that the P* model is not explained well. The authors see the pilot jobs as primarily facilitating mapping tasks onto heterogeneous resource pools. In practice though, pilot frameworks seem to be popular primarily because they offer some control over resource provisioning in a job/batch scheduler dominated world (i.e., a pilot glide-in is used to obtain a resource lease onto which an application can then submit jobs, further, infrastructures like MyCluster have been used to configure the resource). This resource versus job provisioning aspect has not been articulated enough; for example, the SUs are in practice resource slots. Later in the paper, there is talk of "the process of mapping a SU to resources via a Pilot" and "potential" multiple layers of scheduling (there are always at least two in a pilot framework). A clearer model and explanation here would significantly improve the paper. Also, in the evaluation, it would be more interesting to see more fine-grained performance results to get a better feel for the potential performance issues of pilot frameworks. Is the API itself introducing any overhead? What are the trade-offs for the user for choosing to submit to two several different pilot-based systems instead of one?

Detailed Comments for Authors

Provide detailed comments on the presentation, organization, grammar, typographical errors, clarity problems with tables and graphs, or detailed technical points. This is prose that the authors will see and it helps to be specific here.

- "SU can consist" instead of "SU can consists" - III.B "It utilizes" instead of "it utilizes" - V "Instead" instead of "In stead"

Summary and High Level Discussion

Indicate the topic of the paper and summarize the importance of the contribution. List the major points of the paper and your evaluation of them including problem and motivation, related work, main idea(s), algorithms and techniques, experiments and experimental method, results, etc. **If the paper is an SOP submission, indicate the secondary area of contribution (if any) from among the remaining seven areas** (Algorithms; Applications; Architectures and Networks; Clouds and Grids; Performance, Energy, and Dependability; Programming Systems; Storage, Visualization, and Analytics). **Summarize the importance and novelty of the contribution towards the recording and dissemination of the state-of-the-art in the practice of HPC.** List the major experiences and observations of the paper and your evaluation of their strengths and weaknesses.

This paper defines the general concept of a Pilot Job as the unit for independently parallel tasks and looks at three implementations Condor: DIANA and Bigjob. Extensive experimentation is done. The concept is generalized to one involving CPU and data use -- the Pilot-Data PD concept. It is not discussed but I think this "Bag of Tasks" concept.

Comments for Rebuttal

Use this section to clearly indicate the one or two major issues that you would like the author(s) to address in the rebuttal.

This is prose that the authors will see and use in providing their rebuttal which is limited to 750 words. Please provide clear and specific text to guide the response by the author(s).

What are research deductions from work.

What do experiments show?

Where do clouds fit in?

Detailed Comments for Authors

Provide detailed comments on the presentation, organization, grammar, typographical errors, clarity problems with tables and graphs, or detailed technical points. This is prose that the authors will see and it helps to be specific here.

This is a solid paper but neither very exciting (as basic ideas are old) and not very insightful. The paper presents results on several infrastructures but says we shouldn't take detailed results seriously as impacted by quirks of data placement and presumably details of chip set in nodes. More interesting should have been comparison between different frameworks but here paper says there is essentially no difference as job runtimes large compared to set up time. Several important frameworks like Swift and Nimrod are not discussed. Further most futures in this area of independent jobs is likely to be in clouds where "appliances" "roles" (Azure) or just machine images are the pilots. These already explore data extension with prototypes like the Venus-C "Generalized Worker" and compSS framework built at Barcelona. OCCI and CDMI underlie this. To be accepted this paper must have a nontrivial deduction and discuss more modern technology.

Review of pap241s2 by Reviewer 4

[top](#)

Summary and High Level Discussion

*Indicate the topic of the paper and summarize the importance of the contribution. List the major points of the paper and your evaluation of them including problem and motivation, related work, main idea(s), algorithms and techniques, experiments and experimental method, results, etc. **If the paper is an SOP submission, indicate the secondary area of contribution (if any) from among the remaining seven areas** (Algorithms; Applications; Architectures and Networks; Clouds and Grids; Performance, Energy, and Dependability; Programming Systems; Storage, Visualization, and Analytics). **Summarize the importance and novelty of the contribution towards the recording and dissemination of the state-of-the-art in the practice of HPC.** List the major experiences and observations of the paper and your evaluation of their strengths and weaknesses.*

This paper proposes a model of pilot-abstractions; an abstraction of pilot-jobs and pilot-data, which provides the ability to utilize a placeholder job as a container for a dynamically determined set of compute tasks such as Condor-G/Glide-in, Swift and DIANE. It shows pilot API for the proposed abstraction, and implements it for Condor-G/Glide-in, DIANE, and authors' BigJob. The evaluation results show low abstraction overhead and interoperability between different systems.

Comments for Rebuttal

Use this section to clearly indicate the one or two major issues that you would like the author(s) to address in the rebuttal. This is prose that the authors will see and use in providing their rebuttal which is limited to 750 words. Please provide clear and specific text to guide the response by the author(s).

Based system like Condor-G/Glide-in has a scheduling feature by, for example, match-making. How can it be expressed by the proposed pilot job framework? Relation of existing systems that provide the same functionality as the proposed framework is not clear.

Detailed Comments for Authors

Provide detailed comments on the presentation, organization, grammar, typographical errors, clarity problems with tables and graphs, or detailed technical points. This is prose that the authors will see and it helps to be specific here.

The explanation of figure 8 is not complete. It is not described what each bar means.

Committee Comments to Authors:

[top](#)

None

Authors' Rebuttal of Reviews:

[top](#)

[submit rebuttal](#)

