Pipelined Reasoning of Verifiers Enabling Robust Systems (PROVERS)

HR001123S0020

Frequently Asked Questions (FAQ)

As of May 4, 2023

Q63: If the Phase 2 platform is classified, will TA1 teams need to obtain appropriate security clearances and/or be read into those programs?

A63: A TA1 performer that is paired with a TA2 performer, where the Phase 2 platform is classified, will need to have appropriate security clearances already in place. This would be confirmed in advance of any award by DARPA.

Q62: In previous discussions, there was mention of a DARPA-provided System Under Test. Will there be such a system?

A62: DARPA will review the set of proposals submitted and decide whether a DARPA-provided SUT is needed; if so, DARPA will provide one.

Q61: The BAA talks of a wide range of properties and tooling, from relatively simple properties that can currently be automated at scale (e.g., the absence of certain runtime errors), to rich properties that need much more sophisticated proof techniques and tools. Are these of equal interest? Some approaches provide a “gradual on-ramp” from one to the other---is that a priority?

A61: Richer properties, that cannot routinely be automated at scale, requiring more sophisticated proof techniques and tools are of greater interest Although, as suggested, it is imagined that in some settings a “gradual on-ramp” will be appropriate and valued.

Q60: How will TA2 performers “put forward updates/change requests to exercise maintenance/repair capabilities of TA1 throughout each of the phases”? In particular, will the changes be planned in advance? Will the upcoming changes be “surprises” for the TA1 performers working with the case study? Will they be surprises for the developers working on the TA2 team? Do they have to be pre-agreed with the TA3 or QE/EC teams?

A60: TA1 and TA2 performer teams will be fully collaborative, through a collaborating proposal submitted through the BAA process or upon selective pairing by the PROVERS PM. The updates/change requests are not meant to be “surprises” for the TA1 or TA2 performers. As well they do not have to be pre-agreed with the TA3 or QE/EC teams. Updates/change requests do need to be framed within the TA1/TA2 joint, or TA2 independent proposal, sufficient for the PROVERS PM to have a sense that the performer has a strong plan for carrying out the updates/change requests for their system throughout each of the phases of the PROVERS program.

Q59: The BAA mentions that collaboration between TA1 and TA2 members will be unconstrained in phase 1, but constrained in later phases. What will that constrained collaboration look like, especially in a combined TA1+2 proposal?

A59: The constraining relates to the demonstration of TA1 capabilities by formal methods experts. Within phase 1 exercises formal methods experts will be allowed to participate fully, in an unlimited capacity, in the demonstration of TA1 capabilities. In phase 3 exercises traditional developers, without formal methods expertise, will be expected to carry out the demonstration of TA1 capabilities.

Q58: The quantitative evaluation and metrics are important to the program, but what they mean is not yet clear to us. What is a “class of system properties”? And the percentage of lines of proof will vary hugely depending on exactly what the changes introduced by TA2 are, and the nature of the proof system; how can that be made meaningful?

A58: A “class of system properties” can be viewed as a classification of properties. For instance, properties might be grouped as correctness, information flow, isolation, etc. – classes of properties. Regarding percentage of lines of proof requiring manual change, it is made meaningful as it relates to automation enabled within TA1.1 and TA1.2 to reduce the need for manual intervention when updates/changes are made, intervention requiring scarce expert resources.

Q57: Is the TA1/TA2 pairing 1-1? The BAA allows the plural "SUTs" for a TA2 effort; can several TA2 partners and their respective SUTs be part of the same TA2 team or TA1+2 team? For instance, a TA1 effort may want to target several TA2 partnerships at the time of proposal, fostering synergies between them.

A57: The TA1/TA2 pairings are anticipated to be one-to-one.

Q56: Is hardware verification in scope for TA2, or as part of a mixed software/hardware verification challenge? SUTs candidates may have both a software and a hardware part.

A56: Yes.

Q55: For the TA2 SUT(s), do they need to currently exist and be functional? Or can they be "under development" or "to be developed"? - Formal Methods may be more efficiently deployed when they are part of the development from day 1, rather than when applied to a legacy code base / system. Are we evaluating the use of formal methods throughout the design lifecycle from requirements validation to system maintenance?

A55: TA2 SUT(s) need to be able to be sufficiently developed, at a minimum “under development”, to meet TA2 SUT requirements. A “to be developed” SUT would not meet these requirements.

Q54: What is the rough scale of the expected SUT in terms of complexity (lines of code, number of components, data throughput, etc.)?

A54: Preference will be given to SUT(s) that require the analysis of a composition of components over the course of the PROVERS program, as opposed to those that require analysis of only a single component.

Q53: What is the level of expertise of the target users? Bachelors/Masters/PhD with/without training in FM?

A53: Target users, or traditional developers/engineers, are likely to hold bachelor’s/master’s degrees without formal methods training.

Q52: Does the program / DARPA / the FFRDC have anticipated usability metrics in mind?

A52: Yes, although these will be more fully developed by the FFRDC in concert with the PROVERS PM. Usability metrics, for example, may focus on such concepts as friction, how much does the technique itself require effort from the programmer in order to adopt and continue to use the technique – inertial friction and drag.

Q51: Is it okay to prove properties for newly written (or rewritten) code as opposed to legacy code?

A51: Yes.

Q50: How important is including CPS aspect to the proposal?

A50: This would be valued but not required.

Q49: Will the teams be coming up with spec and code changes throughout the program or will others (e.g., the FRDC government team) be doing so?

A49: The TA2 performers will have responsibility to provide modifications/updates of the SUT for use within each of the three phases.

Q48: Are different TA1 teams encouraged, allowed, discouraged, or not allowed to collaborate?

A48: TA1 performers are encouraged to collaborate.

Q47: Would it make sense to do a TA1.1-only proposal, or would it be better to form a collaboration with another group that could provide matching TA1.2 and TA1.3 capabilities?

A47: A TA1 proposal submission must cover the three TA1 subareas, TA1.1, TA1.2, and TA1.3.

Q46: Is it in-scope for TA1.1 to verify combined hardware + software systems (i.e., SoC implementation together with firmware/software), like the dishwasher example mentioned in Sandia's slides?

A46: Yes.

Q45: Can you comment on the color of money. 6.1, 6.2, etc.

A45: PROVERS phase 1 is 6.1, with phases 2 and 3 being 6.2.

Q44: Do you envision the technologies developed under PROVERS to be applied more to newly constructed systems (clean slate assured design and systems engineering) or on existing systems that may or may have al already developed corpus of proofs and assurance cases?

A44: Either is within scope.

Q43: In terms of the QE/EC team, can you elaborate on the broad range of verification capability and the associated number of classes of system properties proven metric? Will the focus be more on functional or non- functional (e.g., usability, cybersecurity) properties?

A43: Current tools are limited in the range of properties, functional and quality, supported. Therefore, tools should support a diversity of system properties and qualities. The capability is to provide for a broad range of properties, with the measure to be the number of classes of system properties proven by phase.

Q42: You identified adoptability and ability to adopt (in terms of resources) as hinderances to current widespread use and adoption of formal methods in the DoD systems engineering space. The desire for PROVERS to push the envelope on both of those seems to highlight the need for human-systems expertise, as does the Sandia presentation. What technical areas do you see that as integral to? Workflow integration?

A42: TA1 Proof Engineering, and its three sub-areas scalable automation, workflow integration, and continuous feedback, is integral to adoptability of formal methods.

Q41: Can you explain the connection (or the whole process) between a requirements change, the consequent design change and any system modeling updates that necessitates, and the proof repairs/formal verification updates that would need to happen?

A41: One might imagine a change to a system requirement will necessitate a subsequent update to the design and system model. As well such a change may cause existing proofs to break, necessitating repair of the proof.

Q40: How does a requirements change relate to the metric on the percentage of lines of proof requiring manual change after a software update/during maintenance?

A40: TA2 will put forward updates/change requests to exercise maintenance/repair capabilities of TA1 throughout each of the phases. As a basis of confidence mentioned during the PROVERS Proposers Day for TA1.2, Amazon's success in the continuous formal verification of s2n, the open-source TLS implementation used in numerous Amazon services, was described. Within this AWS effort, and over the course of a year, AWS demonstrated that changes to code resulted in automatic corrections to the corresponding proof scripts 953/956 times: only 3 times did they have to call in proof engineers. Workflow integration of this repair technology enabled performance enhancements with little involvement of proof engineers. Hence the goal is for automation enabled within TA1.1 and TA1.2 to reduce the need for manual intervention when updates/changes requests are made.

Q39: Can you say more about the various assessments that are planned for the program? We see references to cyber evaluations (4 total?), impact of requirements changes (8?), and phase evaluations (4?).

A39: The TA3 performer will provide for state-of-the-art security assessments, with assessment results supporting the quantitative evaluation, and to inform end-of-phase FFRDC evaluations. TA2 will provide for SUT change requests (i.e., requirement changes) that will notionally occur twice in phase 1, four times in phase 2, and twice in phase 3. SUT change requests are to be put forward to exercise and access TA1 developed maintenance/repair capabilities.

Q38: What is the anticipated timeline of how long TA3 would have for each of the baseline (Phase 1 and 2) and verified (end of each Phase) assessments?

A38: For the PROVERS program it is anticipated that the TA3 performer would have 2 – 4 weeks based on complexity of the use-case to perform each assessment, however specific timelines for assessment will be formally communicated at the PROVERS kickoff meeting.

Q37: Will the TA3 red team restrict their attention to cybersecurity vulnerabilities and findings, or will they also look at the requirements that TA1/TA2 teams are verifying?

A37: The TA3 performer will focus attention on security assessment of the SUT as described within the BAA, as a baseline assessment and at the conclusion of each phase.

Q36: Will the FFRDC evaluations be the ones to evaluate the human-related metrics that are gathered by TA1.3 instrumentation, or will that be evaluated by the TA1 teams?

A36: The FFRDC will evaluate human-related metrics to assess usability of capabilities developed within the program. The data that the FFRDC is evaluating is generated by TA2 teams using TA1 capabilities instrumented for data collection. Guidance will be provided by the FFRDC on what minimally should be collected per metric.

Q35: Will combined TA1/TA2 teams be required to provide platform support and expertise to other performer teams?

A35: TA2 platform expertise will be provided to TA3 and QE/EC performers as appropriate.

Q34: What are the expectations for initial versions of TA2 SUTs? Should they have a set of verified requirements at the start of the program, or will that develop over the course of phase 1?

A34: An ideal TA2 SUT will have a rich set of security and safety requirements, with many requirements verified. As well the TA2 performer is to provide SUT digital development artifacts as outlined within the BAA.

Q33: Who will be acting as IRB for any human subject research? AFRL, Sandia, or performers?

A33: The Government awarding agent has an assigned DOD HSR IRB which would be used within the program.

Q32: Is there any relation of PROVERS to SoSITE and IDAS?

A32: There is no intentional relationship between PROVERS to SoSITE and IDAS. Although the PROVERS PM also currently serves as the IDAS PM.

Q31: TA1.3 talks about collecting usage statistics and use that to improve performance/acceptance of the system we build. Can the system also solicit explicit feedback from the user, or is this considered too disruptive?

A31: TA1 proposals could propose soliciting explicit feedback of users of TA1 developed capabilities.

Q30: Solicitation implies that TA1s will be paired with TA2s. if a TA1 proposer does not explicitly propose a TA2 partner, how will it be ensured that the proof-engineering ideas of TA1 and the platform proposal of TA2 performers are compatible?

A30: This will be the responsibility of the PROVERS PM team.

Q29: Will the continuous development environment, including the CI/CD pipelines and the choice of formal method tools, will be determined by the target platform (i.e., at TA2 discretion, perhaps the tools that they already use for the target development), or by the TA1s (potentially causing TA2s to significantly augment their development process)?

A29: The goal should be to incorporate TA1 proof engineering tools into TA2 systems engineering workflows for development.

Q28 What is the scope of artifacts that they will curate? Is this only for the purposes of evaluation or are the TA1/TA2 teams expected to use this curation service?

A28: QE/EC will curate program artifacts such as: SUTs and related artifacts, capabilities and evidence developed, and assessments. TA1/TA2 teams will provide artifacts to the QE/EC performer for curation.

Q27: Can you describe the OTA process? Specifically, should we submit to the BAA or wait for an OTA? How long would it take the OTA to be created? Will (can) the OTA be specific technical areas from the

BAA or will the OTA be broader to cover multiple TAs from the BAA? -Will DARPA fund the BAA and OTA or one or the other?

> A27: An OTA is not a separate solicitation method, but one of many award mechanisms available under the BAA. A proposer may suggest the appropriate award type (e.g., fixed price or cost reimbursement contract, cooperative agreement, etc.) they expect in response to their proposal's selection for an award, but the Government reserves the right to determine the final award type. An OTA may be an appropriate award mechanism for non-traditional proposers looking for innovative, commercial-like contractual arrangements. To understand the flexibility and options associated with Other Transactions, consult http://www.darpa.mil/work-with-us/contract-management#OtherTransactions.

Q26: Brad mentioned Amazon s2n when talking about TA1.2 Workflow Integration and Facebook Infer when talking about TA1.3 Continuous Feedback. Do these two subareas (TA1.2 and TA2.3) focus primarily on engineering and integration efforts? In other words, do you foresee an all-academic team addressing these two subareas or do you recommend an all-academic team to work with industrial partners to address these two subareas?

> A26: Mention of Amazon s2n and Facebook Infer related accomplishments were only meant to convey a basis of confidence for success in these two TA1 sub-areas.

Q25: The BAA mentions that TA2 phase-1 platforms will be unclassified, and later in phases 2&3 may shift to classified. Since TA1 teams will be collaborating with TA2 teams, does it mean that TA1 teams would need security clearance starting from phase 2, or would an unclassified TA1 team for all three phases be okay?

> A25: TA1/TA2 pairings, made by the DARPA PM after the Scientific Review Official's selections, will ensure TA1 performer capabilities, workflows, and accesses (if needed) are a strong fit for proposed TA2 workflows, use case artifacts, and any necessary accesses. Therefore, if a TA1 performer desires only to perform unclassified work, if selected, the performer would be paired with a TA2 performer where the proposed SUT remains unclassified throughout the entire PROVERS program.

Q24: If a team focuses on TA1, do they need to address all of TA1.1, TA1.2 and TA1.3, or could they choose to only address a subset of those, for example TA1.1 and TA1.3 only?

> A24: A TA1 proposal submission must cover the three TA1 subareas, TA1.1, TA1.2, and TA1.3.

Q23: I'm trying to understand why the Sandia Human Interaction lab is relevant to this project. Are human interfaces and human computer interaction for theorem provers in scope of the project? What about verification of human-in-the-loop (as in a pilot or a driver), which is a very different problem?

> A23: A goal of the PROVERS program is to enable traditional software developers to incrementally produce and maintain high-assurance national security systems. To ensure that TA1.1 and 1.2 PROVERS capabilities continually adapt to feedback from the developers who use them, capabilities developed within the program will be instrumented to collect usage data. This usage data will be used to continuously improve performance and user acceptance within the program.

Q22: Are there any plans to discuss this program at upcoming ITEA workshops? I think there are a number of potential transition partners in attendance that would be very interested in the work to be done on this program.

A22: There are no current plans to discuss the program at upcoming workshops, however as appropriate such workshops would be considered.

Q21: Given that a combined TA1/TA2 bid is a single proposal: Is it also a single abstract? Or can a single bidder provide a separate abstract for TA1 and TA2?

A21: A combined TA1/TA2 proposal should submit a single abstract. The abstract is a concise version of the proposal comprising a maximum of five (5) pages.

Q20: Can references in the abstract be included in an appendix that does not impact page count?

A20: The abstract is a concise version of the proposal comprising a maximum of five (5) pages including all figures, tables, and charts. The required cover sheet, and optional submission letter, table of contents, or appendices are not included in the page count.

Q19: Can you provide some more insight into what you expect from a successful TA1 proposal with respect to the notion of proof. Some of the examples cited in the BAA involve powerful interactive theorem provers that are sound and are used to state and prove arbitrary properties, but most of the examples involve static analysis methods that are neither sound nor complete, can only check a limited set of properties and don't require any proof engineering.

A19: An ideal TA1 proposal would provide for a range of formal methods capabilities to include proof assistants as well as other static and dynamic analysis methods.

Q18: Can you provide some insight/examples into the kinds of processes/pipelines that are in scope for TA1?

A18: There are numerous “pipelines”, specific “pipelines” referenced in the PROVERS Proposers Day presentation by way of example were Travis CI, a continuous integration service, and MuseDev, a DevOps-native code analysis platform.

Q17: The languages, development cycles, processes, IDEs, etc. used are numerous. For example, are Python/Java/C/Javascript/ADA/C/Rust/Go of interest? What about languages based on interactive theorem provers?

A17: The PROVERS program is not focused on specific languages, however it is clear that there may be certain advantages in the choice of a programming language, such as languages that are more amenable to formal approaches. Similarly, the PROVERS program is not restricting its focus on specific proof tools.

Q16: For budgeting purposes, how many TA2 teams should the TA3 team expect the program to have (basically - how many systems will TA3 team should plan to evaluate in each assessment)? Do you have any guidance on the anticipated complexity of the TA2 systems?

A16: For planning purpose, assume three TA2 teams/systems and address in the cost volume any pricing assumptions if there are more or less than three. Regarding TA2 system complexity, we envision working with individual system components then composition of a few system components.

Q15: The BAA seems to indicate that TA2 will provide TA3 with access to the digital artifacts, but not any hardware or simulation capability - does that mean that TA3 is solely responsible for any simulation that could be needed to perform dynamic evaluation?)

A15: Any hardware or simulation capability made available in TA2 will be available to TA3.

Q14: The schedule section of BAA specifies two assessments in Phase 2, but the TA3 section specifies only one - does that mean TA3 would not need to assess the verified systems for the Phase 2 intermediate assessment?

A14: That's correct, the phase 2 intermediate assessment will not require a TA3 assessment.

Q13: Will the classes of system security properties be standardized across all performers or are performers expected to proposer their own security properties?

A13: The properties depend on the proposed SUT so all system properties are proposed, not standardized. Proposals should highlight these property classes

Q12: The GAO report referenced in the Proposers Day presentation mentions that "if it is not in the contract, do not expect to get it." Does DARPA anticipate tackling any of the challenges related to this statement from the report?

A12: PROVERS is focused on making formal methods more accessible, not developing contracting language.

Q11: If you propose to both Ta1 and ta2 do we need to file a conflict-of-interest statement?

A11: Proposers can submit combination TA1 and TA2 proposals and be selected for award of both. This does not create a conflict of interest (COI). It's on the proposer to confirm whether any other COI exists. If the performer identifies there is a COI, the performer should take the action to submit a COI mitigation plan with the proposal.

Q10: Can program management provide some kind of "matchmaking" help for having a T1 team and a T2 team write a joint proposal? Or do you wish to maintain separation between T1 and T2 teams? How would you recommend we seek out such proposal partners?

A10: A hyperlink to the PROVERS teaming website was distributed to all registered Proposers Day attendees to assist with teaming.

Q9: Is the FFRDC QE&EC a role specific to Sandia, or other DoE labs such as National Renewable Energy Laboratory (NREL) can contribute to this role as well?

A9: Sandia is the only FFRDC on the on the QE/EC team. Contribution to Sandia QE/EC work would be negotiated with Sandia.

Q8: Can FFRDCs contribute to two roles (e.g., QE&EC and TA role)?

A8: There will a lead FFRDC on the QE/EC team. That FFRDC will be ineligible as a performer at any level under the BAA. Other eligible FFRDCs may propose to the TAs in accordance with the BAA requirements.

Q7: For TA2, would systems that are CUI or sensitive but not classified be allowed in the first Phase?

A7: There are no classified systems involved in phase 1. The phase 1 SUTs must be open, no CUI.

Q6: Will proposers be able to submit CUI and/or classified material as part of their response?

A6: Yes, the process is outlined in the BAA Section IV.B.3.b.2 and IV.B.3.b.3.

Q5: Will TA2 be responsible for developing threat models for their systems or just properties they are interested in formally verifying/proving?

A5: The inclusion of threat models would strengthen a TA2 proposal.

Q4: What are the boundaries between TA1.2 and TA2 in integration of TA1 techniques into the software development tool chain? For example, is TA1.2 responsible for integrating their techniques into an IDE that TA2 then uses, or is TA2 responsible for actually integrating TA1 capabilities into their DevSecOps toolchains/processes?

A4: TA1 and the TA2 pairing will necessitate collaborative workflow toolchains.

Q3: Are there any restrictions on partnering with foreign-owned companies or including foreign nationals on the team?

A3: There are no restrictions on eligible foreign entities, but refer to the BAA for foreign-owned company and FN eligibility requirements.

Q2: What is the anticipated start date of the program?

A2: Late fall of 2023

Q1: To what degree is public TA1 work a priority in phases 2 and 3?

A1: Continued public work is important to promoting FM adoption and meeting program objectives and will continue throughout phases 2 and 3.