Crowd Sourced Formal Verification (CSFV)
Questions & Answers
As of February 7, 2012

Q42: What is the responsibility of the Technical Area 3 performer with respect to the deliverables received from the Technical Area 2 performers? Specifically, the BAA states that "Technical Area 2 will provide beta-quality software to the Technical Area 3 performer at months 12 and 30" and that "Technical Area 2 will provide final software at months 15 and 33." What will the Technical Area 3 performer need to do with the optimizing compiler received from the Technical Area 2 performer?

A42: The Technical Area 3 performer will need to integrate the optimizing compiler into their system (i.e., ensure the compiler will work on program benchmarks).

▲ ▲ ▲ Latest Q/A ▲ ▲ ▲

Q41: Is there any restriction on, or prejudice against, subcontracting to a foreign company?

A41: Per Section III.A of the BAA, all responsible sources capable of satisfying the Government's needs may submit a proposal that shall be considered by DARPA. Non-U.S. organizations and/or individuals may participate to the extent that such participants comply with any necessary nondisclosure agreements, security regulations, export control laws, and other governing statutes applicable under the circumstances.

Q41: Will the hosting and setup of the database during the Technical Area 1 development and Beta testing periods be covered by the Technical Area 3 team? Is it correct to assume that the associated database will be furnished by the government via TA3?

A41: Yes, hosting and setup of the database is a Technical Area 3 responsibility. We expect that the Technical Area 3 team will interact directly with all Technical Area 1 performers to make the database and interfaces available as needed.

Q40: How many games may we describe in Technical Area #1?

A40: There is no restriction on the number or type of games a proposer may offer to develop. Technical Area 1 proposals should clearly indicate how many games of each type could be developed for the proposed level of effort, and clearly delineate the cost for each game.

Q39: How interested is DARPA in proposals that develop games to train programmers to effectively use existing verification tools? ACL2s is an attempt to make

theorem proving accessible to the masses. Gamifying ACL2s could be a big step in that direction. Is this of interest?

A39: Training software engineers is not a goal of CSFV.

Q38: Is there any more publicly available information on the PipeJam game? For example, given the short video, the game seems to be reducible to SAT, which would indicate that there is no reason to have humans play the game since machines will be much more effective players. Presumably DARPA does not want games that can be solved by current verification technology.

A38: PipeJam attaches to Prof. Michael Ernst's pluggable type-checkers. Please see his ICSE 2011 paper, "Building and Using Pluggable Type-Checkers" for more details. In particular, PipeJam is capable of generating the annotations that would otherwise require human analysis.

CSFV is not defined by PipeJam. It is merely an example to demonstrate the feasibility of mapping formal verification problems into games.

Q37: The answer given in Q6 is somewhat confusing. For example, it seems that a better metric is one based on the number of conjectures that are resolved: either fully proven or falsified with a concrete counterexample. For example, if it is desired to determine if a buffer overrun in possible at some program point, then the conjecture is that buffer overrun is not possible. One game may be able to prove that no buffer overrun is possible using 100 steps and the other game after 1,000,000 steps might still not be done. Why is the number of proof steps a useful metric?

DARPA may want to consider how various competitions compare tools, such as those involved with the SAT competition. The metrics used to compare solvers involved weighing problems: the fewer solvers that solve the problem, the more problems are worth. Also, portfolio solvers typically perform best on problems and wind up winning competitions. It may be interesting to explore that option once the games are up and running, since there may be several complementary approaches.

A37: For the purpose of CSFV, a proof step should be considered the quantum of human input that gets a verification tool unstuck, whether that is one or more literal proof step(s) (in the terms of an interactive theorem prover) or solving an instance of SAT or SMT for a model checker, etc. Getting unstuck can happen in either direction: the proof of a lemma, or a counterexample.

The goal of CSFV, making formal verification practical at scale, is very different than the goals of the SAT competition. Therefore, different metrics apply.

Q36: Are there any restrictions on the kind of games that are of interest for this BAA? For example, what about games that are text-based or that show code? Are such games within the scope of this BAA or do you want 2d and 3d graphics based games? What about games that expect players to have a certain skill set (say a skill set that 1M people have, which is a few orders of magnitude larger that the 1K estimated FV experts) to even start playing?

A36: Refer to Q27, Q30 and Q34. All kinds of games are within the scope of CSFV, however, players will not have access to source code of programs being verified (see slide 28 [slide 15 of Drew Dean's talk] from Proposers' Day titled "Plenty of Code Available to Verify..."). Pending IRB approval, the proposer should choose their audience to maximize program goals, whether that is a targeted or general audience.

Q35: It is thought that some of the output from Technical Area 1 would be configuration optimizations which could then be used by existing static analysis engines or by a new technology. If DARPA is seeking a new compiler backend that would self correct insecure code, based on the output from Technical Area 1, that is a completely different task. The last paragraph in Technical Area 2 implies the self correcting compiler option. Can you please clarify?

A35: In Technical Area 2, DARPA is interested in compilers that generate more optimal code than existing compilers. It is not seeking development of compilers that change the meaning of a software program.

Q34: What are the targeted platforms and communities? Are Facebook, iOS, or Android games being created?

A34: Any platform with Internet connectivity can be targeted.

Q33: If the game makes use of standard parser and compiler tools, does the game need to make suggestions on how to improve these tools?

A33: No.

Q32: Is the game concerned solely with the 2011 CWE/SANS Top 25 Most Dangerous Software Errors? What if the game only addresses some of the 25 software errors but not all? Is it still applicable for an award?

A32: The 2011 CWE/SANS list is a starting point for the 25 properties to be proven of each benchmark. Not all properties will apply to every software program. Other properties to be verified will be supplied so that each benchmark verifies 25 properties. Please refer to Q15.

Q31: Can a player select a random sample of code to evaluate or will the game only have a fixed set of examples?

A31: The TA3 performer will be responsible for choosing which game instances are provided to game players.

Q30: Can a user / player ever view and /or edit the code as part of the game?

A30: No.

Q29: Are performers allowed to require special commercially available peripherals that the players must obtain on their own to play the games? For example, a player may be required to have a Kinnect and Xbox or a camera, etc.

A29: Yes. However, the proposer should consider the impact of their decision with regard to potential audience size.

Q28: Do the games, as they will be deployed on the public internet, need to follow any internationalization / localization requirements?

A28: There are no requirements in CSFV for internationalization or localization. The proposer should consider the impact of its decision with regard to potential audience size.

Q27: Are there any geographical, demographic, background, occupational, age, etc. restrictions when targeting the audience for the games? While performers would make the games available to all, are performers allowed to target a specific audience for their game designs (e.g., children, college-level undergraduates, retired seniors with a technical background, etc.)?

A27: One must consider Institutional Review Board (IRB) issues regarding human use and the targeting of specific demographics. Please see the IRB requirements in the BAA. Pending IRB approval, the proposers should choose their audiences to maximize program goals, whether they are a targeted or general audience.

Q26: Are players allowed to interact with user interface elements from the code as part of the game? For example, part of the code might create a visualization, menu, or input textbox. Are players allowed to see these elements, click on them, or type something?

A26: No. Please see slide 28 of the Proposers' Day briefing (slide 15 of Drew Dean's talk) titled "Plenty of Code Available to Verify ..." regarding approaches to verifying proprietary software. Game players will not have access to the program they are verifying.

Q25: The three benchmark codebases, Hadoop, BIND, and Linux kernel, are usually used by software developers with limited (e.g., command line, network, or API-level) or no actual user interface interactions (as opposed to, for example, a GUI-based OS layer, or a mobile app). Is there a preference for verifying these kinds of codebases, or should performers also consider codebases that involve end-user interfaces?

A25: Those are the program benchmarks, and will be used for evaluating progress. CSFV seeks to make formal verification of software cost effective at scale, without limitation to a specific type of software.

Q24: Java and C are used for the benchmarks. Are performers restricted to these specific languages when targeting the games and tools for this project? Regarding Java, are languages compiled to Java byte code / JVM within scope? (e.g., Groovy, JRuby, Jython, Rhino/Javascript, etc)? Are performers allowed to focus on one specific language (e.g., just Java)?

A24: Please see the answer to Question 20 for background. Performers may focus on a single, specific language. Performers will need to describe how their solutions for languages other than C and Java will support meeting program goals, including benchmarks. This statement also applies to languages other than Java that are compiled to JVM bytecode. Availability of open source software to be verified in languages other than C/C++/C# and Java is a key concern that should be addressed by proposals seeking to verify software in such languages.

Q23: Is there an approved list of formal verification tools to be consulted?

A23: No, each performer picks the formal verification tool(s) of its choice, with modifications as needed to support CSFV. CSFV is not building new verification tools, as specified in the BAA.

Q22: Page 10 of the BAA states that the beta quality software from TA1 is due to TA3 in months 10 and 28. However, the schedule chart on page 10 seems to show that the beta quality software due in months 12 and 30. Which one is correct?

A22: At month 10, the CSFV performers need to have their systems ready for integration — this could be considered a private beta. At month 12, a 3 month public beta cycle begins, with public launch at month 15.

Q21: What is the difference between initial closing and final closing of the BAA?

A21: Proposals must be received by DARPA by the initial closing in order to be considered during the initial evaluation phase. Proposals received after the initial closing will not be reviewed until the review/selection process is complete for the proposals received by the initial closing. However, while proposals may be received and reviewed until the final closing date (BAA expiration), proposers are warned that the likelihood of funding is greatly reduced for proposals submitted after the initial closing date.

Q20: Are optimizing (contemporary) interpreters in scope for TA2, or just compilers?

A20: The goals of TA2 are first, to test the hypothesis that additional information generated from formal verification can help a program run more optimally (along some dimension), and second, provide additional incentives (besides security benefits) for developers to work with the CSFV program.

The scope of TA2 should be understood in relation to the above stated goals. While interpreter vs. compiler is not per se the important distinction, the second goal requires a language with a significant user base. Most common languages (e.g., C, C++, C#, Java, JavaScript) have (just-in-time) compilers as their execution environment, with the exception of AWK, Perl, etc., which are most often I/O bound in any event, and would not assist in meeting the first goal. (Note that Facebook has produced a PHP compiler.) Non-traditional choices should supply a concise motivation for how they will meet the stated goals.

Q19: Is the "crowd-sourcing" aspect of CSFV part of TA1 or TA3?

A19: Generating an audience for CSFV is the responsibility of both TA1 and TA3. Additionally, DARPA will provide assistance by issuing press releases and the use of social media.

Q18: Does DARPA recognize that acquiring artists, musicians and graphic designers is a likely element of a successful proposal?

A18: It is the proposer's task to generate the best possible proposal with the most compelling game. Obviously, there are going to be creative resources needed beyond software developers.

Q17: Is DARPA open to proposals for a game idea that is larger, more technically ambitious and risky but which takes longer than 10 to 14 months?

A17: That would be very difficult to accomplish in this program, because the games need to integrate into an entire system and each game does not entirely stand alone. Proposed schedules should not only be consistent with the maturity of the proposed approach (including any risk reduction), they should also fit within the overall program schedule.

Q16: Which TA is responsible for recruiting players?

A16: None of the TAs are specifically responsible for recruiting game players. DARPA will have press releases and other media interactions to raise awareness of the games at the appropriate points in time.

Q15: Would a TA1 proposal that uses completely different formal verification tools for the different top 25 vulnerability classes, perhaps partially unifying at the game level, be

in scope or is there expectation that each TA1 team would have a single unified formal verification framework?

A15: Yes, that would be in scope. It is not expected that a single TA1 team will necessarily cover all 25 vulnerability classes.

Q14: For TA1 and TA2, do you prefer teams that would handle both C and Java or should teams focus on one programming language?

A14: There is no preference.

Q13: What degree of system behavior is expected as a result of Crowd Sourced Formal Verification? Is the intent only to reduce the number of bugs?

A13: CSFV is focused on formal verification at the source code level. Issues such as correctness of compilers, runtime systems, et cetera, are out of scope of this program. The program is not interested in finding bugs. It is focused on proving correctness for SANS/CWE 25 properties.

Q12: How are compelling or engaging games going to be measured in the general audience or specific target markets?

A12: The primary goal of this program is to increase the amount of formal verification work being performed. Any game that drives this is meeting the program goals. Obviously, general audience is a key aspect.

Q11: What do you expect the budget breakdown to be by technical area?

A11: The budget breakdown has not been predetermined and will depend on the quality of the proposals received. .

Q10: Will PipeJam, including source code, be made available to proposers?

A10: No.

Q9: Can TA2 teams develop games?

A9: Game development is out of scope for TA2. TA2 is strictly about building the optimizing compiler back end.

Q8: Is the development of robots a requirement for TA1? Is that in the scope of the program or should only APIs for the robots be developed?

A8: TA1 performers will just create the API. The robots will be developed by third party game players. Robot development is not an aspect of the program.

Q7: Regarding privacy preserving analytics, should proposals include any analytics on play trace data (whether web or not), or will privacy constraints severely limit the data collected?

> **A7:** Privacy constraints are important and considerations will have to be made as to whether or not the trace data is personally identifiable. Proposals may propose analytics that include such data, however the data must not contain any personally identifiable information.

Q6: Does the program seek evaluations of playability attractiveness to players of the game (or parts of the game) beyond simple popularity and play duration counts?

> **A6:** The key metric for this program is how many proof steps or formal verification steps are accomplished through the playing of the games. Other metrics beyond that would be potentially beneficial to have but are not a core component of CSFV.

Q5: Is there any flexibility towards changing the gaming concept after the proposal?

> **A5:** Changes to the gaming concept between proposal and source selection is not allowed. After contract award, any possible changes would have to be negotiated with the contracting agent.

Q4: Is there a research component included in the work that Technical Area 3 is performing?

> **A4:** There is interest in TA3 proposals which include researching new privacy-preserving analytics. This is not a mandatory requirement but it is something of particular interest to the program. This work will be based on the possibility of obtaining data needed to analyze the usage and popularity of the various games in ways that are better than currently available commercial products.

Q3: Is it possible to be a member of a Technical Area(TA) 3 team and also perform on Technical Area 1, or does participation in a TA3 team rule out being awarded to TA1?

> **A3:** Being a member of a team chosen to perform in TA3 does preclude you from participating in any other technical area in the program. However, you are free to send as many proposals to as many technical areas as you see fit.

Q2: When will the slides, video and transcript from the webcast be released?

> **A2:** Proposers' Day materials will be posted to http://www.darpa.mil/Opportunities/Solicitations/I2O_Solicitations.aspx in the near to immediate future.

Q1: Are there any preferences or limitations in terms of the procurement of Information Technology (IT) equipment in this program?

> **A1:** Procurement of general IT equipment is discouraged in this program. If computing equipment beyond the general needs of PCs are needed, performers should consider utilizing existing cloud computing services instead of proposing to purchase IT equipment.