DARPA-BAA-13-11
VET – Vetting Commodity IT Software and Firmware
Frequently Asked Questions

January 25, 2013

Q28. The BAA notes that we may propose to replace the software and firmware shipped by commodity IT device manufacturers with Open Source software and firmware with similar functionality if we desire, to make experimentation easier. Some commodity IT devices are based on FPGAs that ship with bitstreams pre-configured by the manufacturer. Rather than work with these proprietary bitstreams, can we replace the manufacturer's bitstreams with our own? Can we work on an FPGA development board rather than a finished commercial product that contains an FPGA?

A28. Use of FPGA development boards is in scope provided that you can argue convincingly that the development board is essentially equivalent to the hardware found in some commodity IT device of interest. Use of a bitstream not found in actual interesting commodity IT devices is in scope, provided that you can make a convincing argument that it is similar enough to an actual commodity IT device bitstream to ensure that the analysis techniques developed on it will apply to actual commodity IT devices.

▲ ▲ ▲ Latest Q/A ▲ ▲ ▲

Q27. For TA1, may offerors assume that the firmware and software has been inventoried from the commodity IT device and exported off-board? That is, may offerors begin creating the TA1 output checklist with the firmware and (binary) software in hand?

A27. The mission for TA1 is to start with some sample devices and end with a prioritized list of components to examine and problems to rule out. You will have to propose some kind of coherent story that leads from that start to that end. It seems likely that that story would contain a step that involves figuring out what software and firmware components are on the device that deserve to be considered in the later analysis. However, it's up to you to decide whether or not that step is an interesting research problem, and whether it will become a focus of your work or not be fleshed out much beyond just the story.

Q26. There are many techniques in the anti-malware industry for identifying new malware based on how similar it is to previously-witnessed examples of malware in terms of its structure or behavior. For example, there are traditional signature-based schemes that compute a hash over part or all of a binary and compare the result to the hashes of binaries previously determined to be malicious. There are heuristic schemes that run software in an instrumented emulator and look for behavior that is similar to that exhibited by software previously determined to be malicious. There are clustering

schemes that group similar binaries together based on a variety of features and identify new malware by its closeness to old malware. The common thread among these schemes is their reliance on similarity. Are these similarity-based approaches in scope for VET TA2?

A26. No. In order to make the VET scenario a reality, TA2 will need to provide human analysts with tools to help them rapidly gain some understanding of what a particular piece of software or firmware does or does not do, or what flaws it does or does not contain. Furthermore, these tools will need to provide useful results even in cases where the software or firmware contains malice that is not simply a slight variation on malice that has been seen before, - malice that may instead be entirely unique. Consequently these similarity-based approaches are out of scope for VET TA2.

Q25. Rather than examining software and firmware for backdoors and other hidden malicious functionality, are solutions that instead propose to replace the software on commodity IT devices with new software constructed for correctness from the ground up in scope for TA2? Are solutions that instead propose to harden the software on commodity IT devices with new features that detect misbehavior or sandbox misbehavior in scope for TA2?

A25. No.

Q24. Can we propose work that is a superset of what is called for in the BAA? If alternate approaches like sandboxing and runtime monitoring are out of scope, what could possibly fit into the Alternate Approaches TA7?

A24. DARPA will consider proposals that meet all the requirements of one or more of the VET Technical Areas, supports the overall VET scenario, and also extends or adds detail to the VET story in a way that makes sense. Proposers may specify additional tasks as separately priced options if desired, but this is not required.

Proposers are welcome to send questions on scope to VET@darpa.mil. As a general rule, proposers should consider the program vision described in the Proposer's Day talk: a fully-automated diagnostic device backed directly or indirectly by a checklist of broad classes of hidden malicious functionality that non-specialist technicians can connect to commodity IT devices and await some simple indicator of whether or not the device is safe to deploy. Approaches that do not fit into TA1, TA2, or TA3 yet still support this vision or part of this vision are likely in scope. Other approaches are likely out of scope.

Sandboxing is a valid way of remediating flaws, but remediation is not part of the VET program vision. Runtime monitoring is a valid way of detecting exploitation as it happens, but detecting exploitation as it happens is not part of the VET program vision. Consequently both of these approaches are out of scope for VET. Proposers are reminded that other DARPA BAAs are open, including the I2O Office-Wide, DARPA-BAA-12-29.

Q23. What about False Positives?

A23. DARPA will consider both fully-automated and semi-automated approaches in TA1 and TA2. In semi-automated scenarios, a human analyst might make good use of a tool that produces an acceptably low number of false positives. However, in TA3 DARPA will consider only fully-automated approaches. Since genuinely malicious functionality will presumably be exceedingly rare in deployment, an extremely low False Positive (False Alarm) rate is essential in TA3 to avoid the base-rate fallacy.

Q22. Can we work on development boards?

Q22. Yes, you may begin on development boards but your plan must describe how you will apply your techniques to real devices as the VET program proceeds.

Q21. What about the very likely scenario that every COTS device we analyze in sufficient depth turns out to have many vulnerabilities? Do we revert to paper, pencil and analog phones? Do we wait for clean slate secure software?

A21. The VET program seeks to develop a new capability – the capability to demonstrate that the software and firmware on board commodity IT devices is (or is not) free of specific broad classes of backdoors and other hidden malicious functionality. Remediating problems once they have been found is not the focus of VET. However, if VET succeeds in providing this new capability then it's not hard to imagine remediation scenarios that it might enable. For example, if VET enabled analysts to give a numeric score to the software and firmware on devices ("passed 13521 out of 14213 diagnostics"), then policymakers might be able to choose a numeric threshold for which devices were worthy of deployment. Alternately, if VET enabled analysts to enumerate attack scenarios which imperfect devices might permit, then administrators might deploy external mechanisms to block those scenarios – blocking certain kinds of network traffic, mandating certain manual precautions, and so on.

Q20. Why isn't code that "lies to us" counted in the bad behavior TA2 should look for?

A20. TA1 focuses on producing a prioritized checklist of components to examine and problems to rule out. TA2 focuses on examining those components and ruling those problems out. You could imagine a deployment scenario in which these TA1 and TA2 tasks are handled by experts working on samples of a device in the laboratory, with the goal of approving a particular set of software/firmware for deployment. T3 focuses on examining devices at scale in the field. This examination may re-answer the same questions as the TA2 analysis, or it may attempt to confirm that the device under examination is running a set of software/firmware approved earlier in the laboratory. Either way, the analysis must be fully-automated, and should avoid requiring analysts to

disassemble the device under examination whenever possible. Doing so may require that the analysts run their diagnostic computations on the device under examination itself. This is an inherently dangerous act, as the adversary may have rigged the software and firmware on the device to cause the diagnostic to lie – that is, to report a state of perfect health when the device is in fact compromised. TA3 must produce tools and techniques for running trustworthy diagnostic computations on untrustworthy devices. TA2 and TA3 focus on different problems.

Q19. What risks to the government raised by proposing GPR instead of unlimited rights should proposer's address in their proposals?

A19. Offerors may choose to propose Government Purpose Rights (GPR) in cases where they propose to conduct their research and development with a mixture of Government and their own internal research and development funding as defined by DFARS Part 227. Cases where a performer develops a viable technology with GPR but fails to commercialize it are of less benefit to the Government. DARPA encourages offerors proposing less than unlimited rights to give their commercialization plans some thought at proposal time.

Q18. Can proposals cover multiple Technical Areas? Are there any limitations on which ones? What kind of teaming arrangements are you looking for? Who is responsible for making sure the output of TA1 can be consumed by TA2? How is the integration of TA1 and TA2 evaluated, and what weight is given to integration during evaluation?

A18. Proposers may address one or more TAs in a single proposal. However, proposals addressing multiple TAs may address only combinations of TAs 1, 2, and 3. Proposers wishing to address combinations of other TAs should submit multiple separate proposals, each addressing a single TA, and should not expect DARPA to fund more than one of these proposals. The decision as to which proposal to consider for award is at the discretion of the Government.

Note that covering multiple technical areas in a single proposal is not mandatory – DARPA will consider proposals that address only a single technical area. DARPA expects TA 1, 2, 3, and 7 performers to conduct their research fairly independently through the first 3 years of the program. DARPA expects integration to form demonstrable prototypes will become a focus towards the end of the program. DARPA does not require proposers to form teams covering all of TAs 1, 2, and 3 from the start of the program. DARPA will consider proposals for efforts focusing on single TAs as well as multiple TAs.

The competitive engagements will evaluate TA1, TA2, and TA3 performers individually. They will not evaluate prototypes that integrate work from multiple performers.

DARPA expects the integrator to work out the details of integrating work from multiple performers. DARPA expects the aim of this integration will be to produce demonstrable

prototypes near the end of the program, not to support the periodic competitive engagements.

Q17. Is there any rough guidance on overall program size? Size per TA? Number of likely awards per TA? How large is the total program funding?

A17. The total funding available for award will be listed in the BAA. DARPA expects proposers to propose novel approaches, and to propose a level of effort based on what is needed to explore and demonstrate the viability of those approaches. Proposers may propose some tasks as separately priced and fundable options, if they desire.

Q16. How long will we have to respond after the issuance of the BAA?

A16. 45 days is typical. The precise deadline will be indicated in the BAA.

Q15. Is the vetting of FPGA bitstreams in scope for this program?

A15. Yes.

Q14. Are there any limitations or assumptions on the scope for TA4? What are the limits on hidden malicious functionality? Is VET interested in malicious functionality in software that relies on specific behavior of the underlying hardware? Are covert channels in scope? Will research on the TAs be serialized so that TA2 does not begin until TA1 has produced checklists, and TA3 does not begin until TA2 has produced diagnostics?

A14. For programmatic reasons, VET will assume non-malicious hardware and microcode. We will assume hardware and microcode can contain the usual errata and quirks that might be of use to an adversary, but not features deliberately included for any purpose other than to support malicious activity. Malicious functionality in software that depends on the benign semantics of the hardware is in scope, including examples that depend on errata. Malicious functionality in software that depends on malicious functionality in hardware or hardware features that exist for no other reason than to support malicious software is out of scope.

DARPA expects that the traditional broad classes of flaws will be among those considered by TA4:

1. Generic flaws common in programs written in unsafe languages, such as buffer overflows, integer overflows, and use after free errors
2. Missing authorization checks, or improperly implemented checks such as time-of-check-time-of-use errors, or the provision of potentially dangerous undocumented commands in general
3. Missing or improper use of encryption or checks on hashes or certificates
4. Failure to sanitize input or confusion between code and data

TA4 performers will be free to explore other kinds of hidden malicious functionality, as well. For example, covert channels are also in scope.

Note that all of the backdoors and hidden malicious functionality described in the Proposer's Day mobile phone example could plausibly be explained either as valid administrative commands, or as accidental, rather than intentional flaws. The VET program emphasizes the search for this kind of ambiguously malicious/benign functionality rather than the search for more complex malware, rootkits, or implants whose presence would inarguably indicate malicious intent.

Research on the TAs will not be serialized – all TAs will proceed in parallel. TA2 performers can begin with the traditional broad classes of flaws listed above.

Q13. What is the classification level of TA4? Is it necessary to have a facilities clearance?

A13. The VET program is entirely unclassified. You do not need a facilities clearance.

Q12. Is Firm Fixed Price a valid contract type for this effort?

A12. Proposers can request any contract type, but the contracting officer shall have sole discretion to select an award instrument type and to negotiate all instrument provisions with selectees.

Q11. Do proposers need to have a DCAA auditing system?

A11. No. DARPA encourages proposals from small businesses that have innovative ideas even if they do not yet have an acceptable accounting system as defined by DFARS Clause 252.242-7006.

Q10. Does VET have an interest in vetting applications on smartphones?

A10. Yes, preferably ones implemented in native code rather than "safe" languages like Java. The system software on smartphones is also of interest, of course.

Q9. Will the proposer's day slides be available online?

A9. The slides are available on the teaming website (https://www.schafertmd.com/darpa/i2o/vet/teaming), as well at http://www.darpa.mil/Opportunities/Solicitations/I2O_Solicitations.aspx under VET.

Q8. Are solutions focusing on particular platforms (such as iOS or ARM or Linux) acceptable if they are generalizable to other platforms?

A8. Yes.

Q7. Are solutions that require human analyst decision-making for some scenarios in scope?

A7. Semi-automated solutions are in scope for TA1 and TA2. TA3 solutions must be fully-automated.

Q6. To what extent are approaches that require the physical disassembly of devices in scope?

A6. Solutions that involve physical disassembly are in scope without limitation for TA1 and TA2. TA3 solutions should avoid disassembly to the greatest extent possible.

Q5. What kind of teaming will there be between performers, both within the same TA and across TAs?

A5. DARPA will require performers to sign an Associate Contractor Agreement clause for the portions of the contracts requiring joint participation in the accomplishment of program requirements. This Agreement will ensure appropriate coordination and integration of work by performers, ensure complete compatibility between equipment, data, and services, and prevent unnecessary duplication of efforts and maximize commonality.

TA1, 2, 3, and 7 performers should expect to cooperate most closely with the integrator, experimentation lead, and adversarial challenge teams.

Q4. Will there be early discussion and coordination between the TA 1, 2, 3, and 7 performers and the adversarial teams in order to focus efforts?

A4. Yes.

Q3. Are you expecting proposers to focus on specific hardware/firmware, or will the hardware/firmware targets be specified after award? Will the VET program assume that DRM/software protection on the device is not an issue? That is, can we assume access to the binary or firmware?

A3. TA2 proposers, TA3 proposers, and TA7 proposers addressing the same problem as TA2 or TA3 must all identify the specific makes and models of the commodity IT devices with which they intend to experiment. Furthermore, proposers must choose devices with software and firmware that can be modified both by themselves and by TA4 performers in order to support experimentation. Strong proposals will specify devices that ship with large amounts of software and firmware and that interface directly with Wide- or Local-Area Networks. Some examples of these devices include mobile phones, network routers, printers, and computer workstations. DARPA encourages proposers specifying devices that do not typically interface directly with these kinds of networks to include a

convincing argument that this lack of connectivity does not preclude interesting malicious activity. Performers should consider both technical and legal limitations on their ability to experiment on particular sets of software and firmware before making their choice. Proposers may propose to replace the software and firmware shipped by the manufacturer with Open Source software and firmware with similar functionality if desired.

Proposers addressing TA4 must identify a set of commodity IT devices to which they are capable of applying their technical approach. TA4 proposers must choose sets that contain at least some devices with software and firmware that they and the TA2, and TA3 performers can modify in order to support experimentation. During contract negotiation, DARPA expects to require proposers of selected TA4 AC proposals to settle on specific makes and models of commodity IT devices that match those chosen by the TA2 and TA3 performers.

Q2. To what extent should proposers assume source is available? Would an approach to TA1 based on analyzing design documentation be in scope?

A2. Demonstrating the absence of these classes of flaws is a difficult challenge. Targeting software and firmware shipped on commodity IT devices makes this challenge even more difficult. Analysts must examine this software and firmware as delivered by the manufacturer, and cannot require any cooperation from the manufacturer at all, including the provision of documentation, specifications, or source.

Approaches based on analyzing design documentation are out of scope.

TA2 and TA3 proposers can propose to target Open Source software and firmware, but their analyses must not rely on source. TA4 proposers can use source in the production of their challenge programs, if it is available.

Q1. Is it in scope to consider attacks that involve multiple devices working in concert? Is it required?

A1. TA1 proposers may propose to consider scenarios involving multiple devices or a single device or both.