Cyber Grand Challenge

Frequently Asked Questions (FAQ)

August 3, 2016

Defense Advanced Research Projects Agency
Information Innovation Office
675 North Randolph Street
Arlington, VA 22203-2114

CYBER

GRAND_CHALLENGE

Document Change Summary

Section	Description	Date
	Initial Publication	Nov 7, 2013
Q1	East Coast Competitor Day dates added	Nov 14, 2013
Q1	Modified to include West Coast Competitor Day information and removed CGC website URL	Nov 27, 2013
Q10	Added - What type of security vulnerabilities will CGC address?	Nov 27, 2013
Q11 – Q33	Added	Dec 17, 2013
Q34 – Q56	Added	Dec 24, 2013
Q57 – Q59	Added – Update to scoring methods and initial CGC environment API.	Mar 10, 2014
Q26 – Q27	Obsoleted entries replaced by entries 58 and 59.	Mar 10, 2014
Q60 – Q64	Added	Jul 24, 2014
Q65 – Q73	Added	Aug 29, 2014
Q74	Added	Oct 21, 2014
Q75 – Q77	Scored event update 1.	Nov 14, 2014
Q78 – Q79	Tiebreaker public comment update 1.	Feb 13, 2015
Q80 – Q92	Scored event update 2.	Feb 13, 2015
Q93 – Q102	Added	Mar 18, 2015
Q103 – Q106	Added	Apr 10, 2015
Q107 – Q111	Added	May 6, 2015
Q112	Added	May 12, 2015
Q112	Revised	May 22, 2015
Q113 – Q119	Added	May 22, 2015
Q120 – Q124	Added	May 28, 2015
Q125 – Q126	Added	May 29, 2015
Q127 – Q130	Added	Jun 2, 2015
Q131	Added	Sep 1, 2015
Q132 – Q144	Added	Oct 20, 2015
Q145 - Q156	Added	Dec 10, 2015
Q157	Added	Jan 27, 2016
Q158	Added	Mar 16, 2016
Q159 – Q161	Added	Mar 16, 2016
Q162 – Q163	Added	Mar 25, 2016
Q164	Added	May 6, 2016
Q165 – Q170	Added	May 19, 2016
Q171	Added	May 27, 2016
Q172 - Q174	Added	Jun 3, 2016
Q175	Added	Jun 8, 2016

Q176	Added	Jun 10, 2016
Q177 – Q178	Added	Jun 23, 2016
Q179	Added	Jun 28, 2016
Q180	Added	Jul 12, 2016
Q181	Added	Jul 15, 2016
Q182	Added	Jul 21, 2016
Q183 - Q184	Added	Jul 22, 2016
Q185	Publish CFE TeamPhrases	Aug 3, 2016
Q186	Added	Aug 3, 2016

Q186: Effective August 3, 2016 at 11:00pm PT, this FAQ is closed. How will new FAQ entries be published?

A186: All new FAQ entries will be published in the CGC Event FAQ.

[1] https://github.com/CyberGrandChallenge/Event-FAQ

Q185: What were the competitor team TeamPhrases used to contribute to the calculation of the master seed?

A185: The TeamPhrases solicited from finalists and used according to A176 of the FAQ are published in the below JSON:

```
[
    {
        "phrase":
"defa5b1925203b76ee19bb1102e620754fb655b11b52399da226354630e1f18b61f439b8cb2d520de9958
9c68fdc5312ab6b229879f7bda06d285cba98a961b7fe63ba3e9b96de11254196e2a73dab6099058af816a
747a1182868b868e58eda8206bc33aba51964c4ef77aa4378d5665b66db8b18ae4eb6ed99e560b89ce2467
b4bfff16dea49d6dcb88101392f91e0ca4c4ed672e30cc52b3f7c45a0a8d39c7a4b41b83a0f7e00b50c5ce
123c38645a7c495b53d32df5b8b57dfb3a0933bce478930cd6b4692e8a57b0c335997c6e86a99114a0ca5c
0751118ffe7d989b298d15e5f3df7cd9546290bedb7d79d87f91abafbb4a953078cac4aa53fd10caca1",
        "name": "CodeJitsu"
    },
    {
        "phrase": "Remember: Innovation can occur anywhere -- If we knew what we were
doing, it would not be called research",
        "name": "CSDS"
    },
    {
        "phrase": "Deep Red [team photo: (-(-_(-_-)_-)-)] is excited to participate in
the Cyber Grand Challenge.  They do not battle with o==]::::::::::>'s, but with 1's
and 0's for treasure and eternal glory.  Their specially chosen TeamPhrase shall
surely secure their victory.",
        "name": "DeepRed"
    },
    {
        "phrase": "           ;.\n :         .                                   ..\n
 .NNN.   cco                                  XMM      k.   .\n .MMM.  .OK0
                                  kMM     l:;   l0Wo\n .,;'.,MMM.   . .
                                  kMW     xkl.  ,MMo\n '0WMNKNMNMMM:   dKK
    :dMMMMMXO:      ,xXMMMWKk,    0MM.  ;XKWW  :0MMN0x..\n cMMM'
   lMMMMMX..KMX...lMMMMc.'cWM0   .KMMXl,llWMM:   WMM ;0NM0:' ..'MMK'.\n NMMc
   d0XMMMx.  ;MMd .XMMW0,         XMMk     :MMW0. NMMNMMNx.  .,..MMK\n MMMc   c:oXM0.
   .MMl  0:dxMMMWXx:..xOMMMKKXNWMMMWW0'.KMMXMMWO0'     WMK\n MMM0c   .lXMK    .MM;
       .;dMMMWl.'MMNW0          .KMXl XW0MO.  . MM0\n ,NMMd   .XMMW.   'MM'  oWM;
   .NMMMx  OMNWN.   :MMXl. XMMd  0WMMN   'MMX\n kOWMMWXWMXWMM.   ;MN    cMMXxdNMMMl.
   xMMMKkd0MMKc   XMN'   .KMWK.  kMWKxo\n : l;Ko'.    .    .;c      .cxxdOO
       .:odkxol'     ,.       . .    cddl;..\n :kO.    .      .\n .",
        "name": "Disekt"
    },
    {
        "phrase": "\n              AAAAA\n          AAAAAAAAAAAAAAA\n
       AAAAAAAAAAAAAAAAAAA\n     AAAAAAAAAAAAAAAAAAAAAAA\n    AAAAA  AAAAAAAAAAA
  AAAAA\n   AAAAAAA  AAAAAAAAA  AAAAAAA\n AAAAAAAAA  AAAAAAA  AAAAAAAAA\n  AAAAAAAAAA
         AAAAAAAAAAA\n  AAAAAAAAAAA  AAA  AAAAAAAAAAAA\n  AAAAAAAAAAA  A  AAAAAAAAAAA\n
    AAAAAAAAAAA   AAAAAAAAAAA\n     AAAAAAAAAAA AAAAAAAAAAA\n
      AAAAAAAAAAAAAAAAAAAAA\n         AAAAAAAAAAAAAAA\n              AAAAA\n",
        "name": "ForAllSecure"
    },
    {
        "phrase": "Shell we play a game?",
        "name": "Shellphish"
    },
    {
        "phrase": "GrammaTech and UVA bring you Xandra, Defeneder of Humanity.",
        "name": "TECHx"
    }
]
```

Q184: We noticed the TI API [1] publishes NNN values "between 0 and 100, reflecting percentage" for poll feedback by multiplying the availability score specified in the CFE scoring document [A58] by 100. Does this convention hold for scores as well (i.e. a score of 4 becomes 400 in the TI API)?

A184: Yes

[1] https://github.com/CyberGrandChallenge/cgc-releasedocumentation/blob/master/ti-api-spec.txt

Q183: In the Finalist Event Information document, you state "the possibility exists that competitor HPC access will not be restored prior to CFE computation". Can we specify a powerup schedule for our blades in the event that our CRS must be powered up cold for CFE?

A183: No.

Q182: Can the same CSID appear in two different competitions?

A182: No protection mechanism exists that would prevent this.

Q181: We'd like to optimize for identical binaries; how many of our competitors will field Challenge Sets containing identical Challenge Binaries during CFE?

A181: Unknown.

Q180: What is the tiebreaker algorithm for CFE?

A180: In the highly unlikely event of a tied score in CFE, DARPA/CGC will use the following tiebreaking algorithm on the scores of teams compliant with the CGC Rules

For CFE, a Tied Block is any group of scores in adjacent places for which 1> the scores are equal and 2> the block occludes a prize place (1st, 2nd, or 3rd). For instance, three teams tied for 2nd, 3rd, and 4th place would constitute a single Tied Block, occluding 2nd and 3rd prize places.

For every Tied Block at the conclusion of CFE computation, the tiebreaking process is as follows:

a) Discard one scored round's worth of points from every CRS in a Tied Block from the end of the CFE computation record

b) Remove any CRS from the Tied Block whose score is no longer tied with the Block

c) If any Tied Blocks remain, repeat the tiebreaker at a)

For example, at the conclusion of CFE computation, 1st place and 2nd place systems have the same score, constituting a Tied Block. 3rd and 4th place systems are also tied, constituting a second Tied Block. The tiebreaker is run and the scores from the final round are removed. 1st place and 2nd place are no longer tied, however 3rd place and 4th place remain tied. For the purposes of separating 3rd and 4th only, a second round is removed. This process repeats until 3rd and 4th place are separated.

In the event that the above algorithm can not separate a Tied Block over the full CFE computation duration, DARPA will issue a ranking by expert judgment.

Q179: Will the CFE Event Plan include the number of rounds each CS will be fielded?

A179: Yes; DARPA/CGC will never Plan to field a CS for less than 10 rounds.

Q178: We were reading through the CGC Rules and noted that Area of Excellence #5 says "mitigate security flaws [using] a network security device". Do we really have to deal with all the problems of a network security device? We uploaded some IDS signatures that DoS the IDS device and noticed that max'ing out the load on the device caused performance impact to every Challenge Set on our defended host. Could you perhaps give us a virtualized network security device for every Challenge Set?

A178: No. The CGC Rules specify that "the CRS confronts the CFE network from the vantage point of a real world network defender" using a "network security device" and a "defended host". These challenges: load, resource contention, and network concurrency are real world constraints that have always been part of the Capture the Flag tradition and were guaranteed by the CGC Rules since the inception of the Challenge. Competitors are assured that no late-breaking changes will be made to the Rules that would create a less realistic Challenge.

Q177: What is the official OS kernel of CFE?

A177: No such kernel has been specified, exists, or can exist. DECREE has been implemented as a layer over an open source operating system that continues to receive updates and security patches from a global community of researchers on a timeline that cannot be predicted by DARPA/CGC. DARPA/CGC reserves the ability to continue to secure all parts of the CFE architecture including the kernel <u>up to and during CFE against any threat to the integrity of CFE</u>, to include security weaknesses, emergent threats, execution divergence attacks, data and memory attacks, hardware/software attacks utilizing Rowhammer, memory deduplication, etc. DARPA/CGC has committed to the DECREE ABI and the TI API; see also A137, A174. Any competitor approach whose success is dictated by hardware, firmware, or kernel errata may fail in a real world exercise, to include deployment to millions of computers or the CFE infrastructure.

Q176: How will the CFE framework generate random numbers during the event?

A176: A seed value will initialize pseudorandom number generators used in CFE, to include the PRNGs used to create the flag page, polling schedule, etc. This seed value will be arrived at through a calculation conducted in public view, described in this entry.

The name of this seed value is master seed, hereafter MS.

Each team is asked to provide to DARPA/CGC, by midnight EDT on June 17, 2016, a phrase that will contribute to the calculation of MS. The calculation is explained below, where:

- H is defined as the binary digest output of SHA384
- HEX is an ASCII transform in which every byte of input is converted into a 2-character hexadecimal representation
- ASCIINUM is an ASCII transform in which the decimal form of a number is printed in ASCII
- XOR is defined as bit-wise exclusive or
- the comma character (",") is defined as concatenation.

The calculation is as follows:

```
MS = H(TeamName1,TeamPhrase1) XOR
     H(TeamName2,TeamPhrase2) XOR
     H(TeamName3,TeamPhrase3) XOR
     H(TeamName4,TeamPhrase4) XOR
     H(TeamName5,TeamPhrase5) XOR
     H(TeamName6,TeamPhrase6) XOR
     H(TeamName7,TeamPhrase7) XOR
     H(TeamNameDARPA,TeamPhraseDARPA) ;
```

DARPA will commit to TeamPhraseDARPA on June 10, 2016 by choosing a random number *r* and publishing HEX(H(H("DARPA",TeamPhraseDARPA),ASCIINUM(*r*))) in this entry. DARPA will publish all competitor team TeamPhrases ahead of the CFE in the CGC FAQ. DARPA will publish TeamPhraseDARPA and *r* after the CFE in the CGC FAQ.

Teams may communicate their TeamPhrase to DARPA/CGC via:

Email: cybergrandchallenge@darpa.mil (S/MIME auth supported per CQE instructions).

Post: DARPA/I2O
Attn: Cyber Grand Challenge
675 North Randolph Street
Arlington, VA 22203-2114

A TeamPhrase may be of any length. Teams are encouraged to choose a TeamPhrase that can be expressed in ASCII and will survive government review for public posting.

Any TeamPhrase not received by midnight EDT on June 17, 2016 will be set to the NULL string.

HEX(H(H("DARPA",TeamPhraseDARPA),ASCIINUM(*r*))) is:

```
1773ec260768b2d17bd1cdd6fc54a19e619ee0cc58edcab0f19ef2a32128
79670ede7121914ed1a55eac4565a0bf88ac
```

Q175: Is DARPA/CGC accepting suggestions for its cryptographic commitment strategy for CFE?

A175: The CFE commitment mechanisms have been designed over two years of Challenge execution and cryptography review is complete. No suggestions will be accepted.

Q174: Will the scoring algorithm described in A58, accessed on the CFE infrastructure per A13 of the Trials FAQ [1], and provided as an oracle to competitors per A137 be modified prior to CFE?

A174: No.

[1] https://github.com/CyberGrandChallenge/Event-FAQ/blob/master/event_faq.md

Q173: Will our CRS know in advance on which round a Challenge Set will be removed?

A173: No.

Q172: Will DARPA/CGC provide a mapping that identifies each opponent CRS during CFE?

A172: No.

Q171: What is the autonomy policy for CFE?

A171: Following consideration of responses to the period of public comment, careful observation of Trials, and continued observation of Sparring Partner, DARPA/CGC has determined that the following autonomy policy will govern CFE:

> All rounds of CFE will be computed autonomously.

See also Section 4 of the CGC Rules.

Q170: In CFE, in every round, will every team be scored on the same corpus of Challenge Sets?

A170: Yes.

Q169: What is the duration of the computation of CFE?

A169: CFE computation will begin on the morning of August 4th, 2016, and occur over roughly ten hours of elapsed time, subject to constraints described in A168. CFE computation is projected to finish during the CFE live event [1]. Successful computation of CFE requires a minimum of 40 scored rounds.

[1] Details on the CFE live event are distributed on www.cybergrandchallenge.com

Q168: What is the fixed round schedule for CFE?

A168: CFE will not be computed using a fixed round schedule. CFE will be computed using an event plan. The event plan will be consulted if any risk to event execution is encountered, such as:

- Natural disaster
- Coolant leak
- Power failure [1]
- Major software fault generated by adversarially formed machine inputs (DECREE kernel panic, et al).
- Thermal overload
- Threat to storage of the event record (infrastructure storage failure, primary or secondary)
- Other infrastructure hardware failure (suspected or confirmed)
- Other HPC hardware failure (suspected or confirmed)
- Previously undiagnosed infrastructure software fault

DARPA/CGC will publish a cryptographic commitment to this event plan and distribute the plaintext of this event plan after CFE.

[1] https://lasvegassun.com/news/2016/may/11/power-outage-stops-play-for-40-minutes-at-vegas-st/

Q167: We've prototyped up some code that sends all our cell phones a text message when Sparring Partner matches occur. Does DARPA/CGC object to semi-autonomous participation in Sparring Partner matches?

A167: No.

Q166: We're running a "Chaos Monkey" [1] full time on our CRS that does things like randomly corrupt memory, shut down nodes, corrupt the disk, and terminate processes- getting the CRS ready for a fully autonomous event. Unfortunately, our Chaos Monkey is causing us to perform sub-optimally in our matches against Sparring Partner. Will imperfect play versus Sparring Partner count against our CFE score in any way?

A166: No.

[1] http://techblog.netflix.com/2012/07/chaos-monkey-released-into-wild.html

Q165: We're concerned that our competitors may analyze our replacement binaries (RBs) through the consensus evaluation process and co-opt portions of our RB, making all of our CBs/RBs a CRS-formed input emitted onto the CGC network and potentially destined for a competitor's CB. Could this happen?

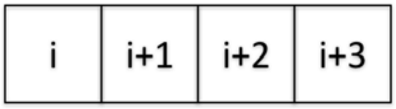
A165: Yes.

Q164: Once removed, will a Challenge Set ever be re-introduced in CFE?

A164: No.

Q163: What happens with a newly introduced Challenge Set (CS)?

A163: In A157 *normal play* was described. For any given Challenge Set M, its initial rounds are covered in this entry as follows:



i	i+1	i+2	i+3

When M is first introduced in round *i:*

> *i*: M is fielded. The first round M is polled.
> *i:* The first round in which CRS formed inputs for M may be submitted.
> *i + 1*: The first round PoV modules that seek to prove vulnerability in M may be thrown.
> Replacement Challenge Binaries and IDS rules files for M received per A161 prior to round i+2:
>
> > *i+2:* available for consensus evaluation.
> > *i+3:* fielded.

The fielding of Replacement Challenge Binaries and IDS rules files incurs service down time in a manner similar to A157.

Q162: How many Challenge Sets will be active at once during CFE?

A162: A maximum of 30 Challenge Sets will be active during any given round.

Q161: If we upload a bunch of IDS rules files for the same CS, which one do you use?

A161: The infrastructure will field the last completed upload per round. This applies to Replacement Challenge Binaries as well.

Q160: Will a CS round score be zero per A157 if all CBs and the IDS rules file for that CS are identical between rounds?

A160: No.

Q159: Is a CRS required to re-submit identical Replacement Challenge Binaries, IDS rules files or PoV modules every round a CS is in play?

A159: No.

Q158: What is DARPA's process to answer competitor questions?

A158: DARPA/CGC is a competition with substantial prizes at stake; this introduces unique requirements on traditional software Q&A processes:

- *Never provide competitive advantage to a particular competitor.* All answers to all questions must be made available to all competitors simultaneously; this ensures that no competitor has access to more information about the competition than any other competitor.

- *Never provide competitive disadvantage to a particular competitor.* If a competitor asks DARPA/CGC "our secret sauce is **ketchup**, please predict the performance of **ketchup** in CGC", DARPA/CGC cannot use the question as written as revealing the text of the question would reveal the secret sauce and create competitive disadvantage.

- *DARPA/CGC must not restate publicly available information*. When a question has already been answered publicly, restating the answer provides the possibility of alternate interpretations that harm the integrity of the contest. For this reason, DARPA/CGC must always repeat existing public answers when they exist.

- *DARPA/CGC must not distribute uncertainty.* When a software bug occurs DARPA/CGC must independently root-cause and authoritatively fix prior to responding, as collaborating with any competitor during the debugging process would provide that competitor with additional insight and, as a result, competitive advantage.

- *DARPA/CGC must work to minimize FAQ rewriting.* Competitors depend on accurate FAQ entries; revoking an entry and rewriting it can cause competitors to back up and change development course. For this reason, many FAQ entries cannot be released until they are backed by repeated testing.

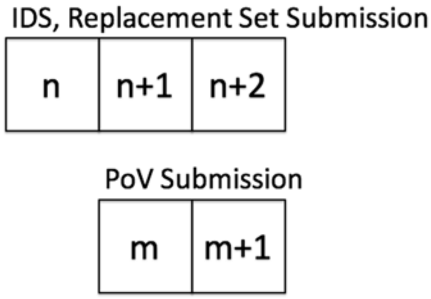
- *DARPA/CGC must observe an approval workflow consistent with public release of information by a government agency.* All publicly posted answers flow through an approval process before release.

CGC has been constructed as a technology competition that is globally accessible, fair to all competitor parties, high-integrity, and as transparent as possible within these constraints.

Q157: How long are rounds and how do they work?

A157: Each round of CFE play will be followed by a short, unscored break. The duration of these breaks will be a minimum of 30 seconds with no maximum duration. Each round of play will last at least 240 seconds. During a break, no game actions will be initiated by CFE infrastructure.

A CRS formed input whose submission completes during an unscored break will be submitted to the round directly following the break.



In response to feedback received during the period of public comment, DARPA/CGC has formulated a round structure to enable consensus evaluation. For any Challenge Set, IDS rules files and Replacement Challenge Binaries submitted during a round *n* are scheduled as follows during *normal play**:

n: CRS formed input submitted.
n + 1: DARPA/CGC inspects CRS-formed inputs.
n + 1: Service is down to pollers and PoVs (round score is zero).
n + 1: IDS rule file and Replacement Challenge Binaries available to opponents via consensus evaluation.
n + 2: IDS rule file and Replacement Challenge Binaries are fielded.

PoV modules submitted during a round *m* are scheduled as follows:
m: CRS formed input submitted.
m + 1: PoV is thrown as specified.

**normal play* does not include the initial and final rounds that a Challenge Set is fielded during CFE; CS introduction and removal will be covered separately.

Q156: Our CRS had trouble with nonces, challenge-response and concurrency in CQE Challenge Sets; we are worried that this will be an obstacle for our CRS in CFE. Will CFE Challenge Sets include these features?

A156: Yes. Challenge Set authors have been instructed to provide increased challenge set difficulty for the CFE phase of the competition. The guidelines provided to the Challenge Set authors can be found at https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/walk-throughs/submitting-a-cb.md

Q155: Will access to the DARPA cloud nodes be maintained 24x7 in the months prior to CFE?

A155: No. DARPA/CGC will schedule maintenance windows whenever possible. DARPA/CGC observes no fixed mapping between teams and hardware and reserves the ability to re-provision within the DARPA cloud at any time.

Q154: What should I do if my DARPA cloud nodes fail during development prior to CFE?

A154: DARPA cloud issues can be reported to cybergrandchallenge@darpa.mil. In the case of a software fault DARPA/CGC may automatically re-provision the node. In the case of a hardware fault DARPA may pursue remedy, replacement, and/or re-provisioning.

Q153: Are the DARPA cloud node disks backed up?

A153: No. Competitors are encouraged to follow cloud computing best practices and remain prepared to deploy their CRS on a freshly re-provisioned cloud. Automated installation and devops best practices may assist in the event of CFE disaster recovery.

Q152: What guarantees does DARPA/CGC make against HPC node failure?

A152: DARPA/CGC cannot eliminate the possibility of node failure. During development prior to CFE, replacement nodes will be installed per commercial warranty as possible. During CFE, failed nodes will not be replaced. CRS design should take place informed by these facts.

Q151: What runtime limits are placed on Network Appliance rules?

A151: Each Network Appliance instance will utilize one rules file. For each instance, the Network Appliance maintains two analysis ring buffers, one for incoming bytes read from each side of the connection. These ring buffers are capped at 100x1024 bytes.

Q150: What runtime limits are placed on CFE PoV modules?

A150: The competition framework enforces a hard limit on PoV modules of 15 seconds of wall time and limits physical memory usage to 64x1024^2 bytes per PoV instantiation (e.g. a PoV "throw").

Q149: What runtime limits are placed on CFE Challenge Binaries?

A149: The competition framework enforces a hard limit on Challenge Binaries of 15 seconds of wall time and 3x1024^3 bytes of virtual address space. See also, A58 and A119.

Q148: What are the CFE upload limits?

A148: The CFE competition framework Team Interface API [1] will accept input no larger than 10x1024^2 bytes for PoV modules, 10x1024^2 bytes for Network Appliance rules files and 50x1024^2 bytes for Replacement Challenge Binaries.

See also: Q58, Q77, Q119, and Q132.

[1] https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/ti-api-spec.txt

Q147: Will the value of MAX_THROWS in the Team Interface API specification be hardcoded for CFE?

A147: Yes: 10. See also [1].

[1] https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/ti-api-spec.txt

Q146: Does the CFE framework facilitate saving state between PoV throws?

A146: No.

Q145: How will the Network Appliance impact the Performance component of the Availability score?

A145: Network appliance ruleset execution time will be measured relative to the execution time of a reference empty ruleset; these measured increases in execution time will be combined with existing CB execution time measurements to determine the execution time portion of the Performance score. See also A58.

Q144: In A58, Retained Functionality is specified in terms of network test cases; are these network test cases "service polls"?

A144: Yes.

Q143: Can a service poll fail by timing out?

A143: Yes.

Q142: Why is a facility provided to query the location of the type 2 PoV memory range and PoV length in cfe-pov-markup-spec.txt?

A142: The DECREE API allows competitions hosted using DECREE to specify the start address of the type 2 PoV memory location and the required size of a PoV through requests to the competition framework. This flexibility will assist in porting the DECREE framework to future architectures. This facility allows researchers utilizing DECREE flexibility in designing current and future competitions and research. During the Cyber Grand Challenge Final Event, the Type 2 PoV memory range and PoV length will be consistent with the "submitting-a-cb.md" walkthrough. See also A29.

Q141: Will teams have access to a scoring oracle during CFE that can be utilized prior to fielding RBs?

A141: Per A4, A17, and A137: No.

Q140: Is it possible for a fielded replacement binary to receive a network input that has not transited the network appliance?

A140: No

Q139: Can fielding a poorly performing network appliance rule negatively affect availability score?

A139: Yes.

Q138: During CQE, a secondary scoring document was released which provided degradation curves and formulae. Will a similar document be released for CFE?

A138: Additional information about CFE scoring will be released through the FAQ.

Q137: What additional information about the impact of network appliance rules, execution time, etc. on the Availability score will be released?

A137: DARPA/CGC intends to allow competitors to answer these questions independently by providing access to a competition framework oracle via the DARPA cloud. At the time of this writing, construction of the competition framework continues. This access will be granted prior to CFE Trials and revoked prior to CFE.

Q136: Which Challenge Sets were excluded from computing CQE ranking in A130?

A136: The exclusion proof decrypts to:

```
Common-name,CSID
YAN01_00007,fd0e1101
YAN01_00009,c9967603
YAN01_00010,31502e01
YAN01_00011,77f39101
YAN01_00012,9d97ef01
```

The decryption can be verified by taking the encrypted base64 blob from A130 and decrypting it with the key in the below script:

```
PASS="The best way to predict the future is to invent
it.qR59RBQiIbMnuFwJFJpNUHMusu6RjyrGxvDI17v9egiGRloXME9
bGjwBjF32bc2qdNqLVwTz3CNuogh1XcWapd5IGQ1VEE4B9/Mt9TAKN
0S1P97XZj4WFCH+KAhMzTR3AkYM/mTbtzXfhDzSOEgl7G+7T1IE0uY
aYYYuEwm4uL4="

A130_PROOF="U2FsdGVkX1/21aa4u0ZhcTCAyGr8oX7aCimSvKfGG3
HFvjy5sCWi7Dh0dY9mb2nizlldmCCeB7g2RbUv3R0PkvFVp+EeTeWj
8MpTiBCP0cVS3/uXy2rWIZtkFpOYBXIqYdybwpawzlRshJ8BeA2/WC
WBeupEhCWFjJT7XigmLyEZqIujQ/rsWJTLOTfWFWC0"

echo "$A130_PROOF" | \
base64 --decode | \
openssl aes-256-cbc -d -pass "pass:$PASS"
```

Q135: During a review of the CQE corpus, our team discovered that some replacement binaries released by our competitors disrupt the function of our program analysis software. We're concerned about this effect during a more autonomous competition (CFE). The CGC Event Participation Agreement explicitly allows this effect from CRS-formed inputs; will DARPA/CGC provide additional protections against this effect?

A135: No.

Q134: The requirement for full autonomy brings with it a slew of engineering decisions, including watchdog timers, redundant components, parallel execution, and high assurance programming. Does DARPA/CGC have a recommended path towards full autonomy?

A134: No. See also A4 and A17.

Q133: Which CRS-formed inputs named in the CGC Event Participation Agreement can be identified in the current DECREE architecture?

A133: At the present time:

- Replacement binaries
- Data produced by replacement binaries
- Network appliance rules
- Traffic modified by network appliance rules
- Proofs of vulnerabilities
- Data produced by proofs of vulnerabilities

Q132: Is a CRS permitted to submit povml files as a Proof of Vulnerability to the competition framework during CFE?

A132: No; see also section 2.2 of the CRS Team Interface API spec [1] that describes how to submit a PoV during CFE.

[1] https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/ti-api-spec.txt

Q131: **What is the duration and autonomy policy of the CFE?**

A131: In A58, DARPA/CGC indicated that CFE points would accrue "per Challenge Binary per round". DARPA/CGC is opening a period of public comment that will close on **September 25th, 2015**. This period of public comment encourages feedback on the appropriate timing of the CFE consistent with the goals of CGC: the ability to demonstrate, if feasible, real-time proactive and reactive mitigation of novel security faults on machine-scale timelines measured in seconds. Feedback topics include:

> 1> Duration of CFE Rounds, in seconds
>
> 2> Total number of Rounds in CFE
>
> Note: Feedback for 1> and 2> should be paired to indicate a total event duration.
>
> 3> Release and/or expiration schedule for Challenge Sets (CS) in CFE.
>
> 4> Policy on Human Interference (hands on debugging). A partial list of theorized policies on CFE debugging follows:
>
>> 4a> Teams could be allowed to debug their CRS at any time during CFE. During debugging, a CRS would be disconnected from the framework and unable to accumulate points during debugging. After debugging completed, any Challenge Set (CS) previously revealed to the debugging team would be fruit of the poisoned tree and unavailable for scoring to that team for the remaining duration of the event. For example, if team alpha chose to debug in the first second when a single CS was fielded, then all points for that CS and the ability to field that CS would be unavailable to alpha for the duration of the event; a team beta that chose to debug in the final minute with every CS fielded would be unable to score any further points after debugging.
>>
>> 4b> A number of scheduled debugging breaks could be announced prior to the event, dividing CFE into Stages. For each Stage, all previous CS's would be wiped and fresh CS's would be provided.

4c> CFE could be totally autonomous; catastrophic errors could not be corrected.

4d> Competitors would be permitted to fully interact with a CRS for an initial period of time at the onset of CFE to ensure the correct operation of the CRS. After this time the CRS would continue autonomously as in 4c.

Competitors are encouraged to provide comments to cybergrandchallenge@darpa.mil by **September 25th, 2015.**

Q130: Which Challenge Sets will be excluded from computing CQE ranking?

A130: See FAQ A117. The proof of excluded CS selection is:

```
U2FsdGVkX1/21aa4u0ZhcTCAyGr8oX7aCimSvKfGG3HFvjy5sCWi7Dh0dY9m
b2nizlldmCCeB7g2RbUv3R0PkvFVp+EeTeWj8MpTiBCP0cVS3/uXy2rWIZtk
FpOYBXIqYdybwpawzlRshJ8BeA2/WCWBeupEhCWFjJT7XigmLyEZqIujQ/rs
WJTLOTfWFWC0
```

Q129: An empty write in a PoV XML passes poll-validate but causes an exception in cb-replay; will this be fixed?

A129: A fix will be released in a future version of DECREE. During CQE, competitors are encouraged to refer to FAQ entry A111 with respect to debugging their CRS during the event. Prior to CQE, competitors are encouraged to ensure their CRS does not generate a PoV with an empty write statement similar to the below:

```
<write><data></data></write>
```

Q128: What is the official Twitter feed for the CGC Qualification Event?

A128: https://twitter.com/DARPA_CGC_CQE

Q127: How do I confirm that I am running the latest version of DECREE?

A127: In the last DECREE release email, a test [1] was issued to check the vagrant version. If your current system responds with the output cited below [1] [2], there is no need to update. If your system responds with different output, DARPA/CGC advises that you update.

[1] The expected output of vagrant ssh of an upgraded VM will be: Linux cgc-linux-packer 3.13.2-cgc #1 SMP Mon Apr 13 18:33:57 UTC 2015 i686
[2] vagrant@cgc-linux-packer $ cat /etc/decree_version cqe_development-vm-61

Q126: Where is the CQE challenge bundle?

A126: The Cyber Grand Challenge CQE bundle is available! Please download and verify at your earliest possible convenience prior to CQE. Download instructions are available [1]. Please continue to monitor the Event FAQ [2] and the CQE Twitter Feed [3] throughout CQE.

Package Name: cgc_qualifier_event.ar.gz.enc
Package MD5: 0ec6a5708aea0b0d7c5a9f1c37924423
Package SHA256: 8274d4086ee1039e7901b3cc7221fa89e2f923dc5ad39924965f2ff5b110449d

[1] https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/walk-throughs/cgc-qualifier-event-api.md
[2] https://github.com/CyberGrandChallenge/Event-FAQ/blob/master/event_faq.md
[3] https://twitter.com/darpa_cgc_cqe

Q125: Do all hardware components of the CRS need to be physically present at a CQE site visit?

A125: No.

Q124: Were all submissions to Scored Event 2 scored?

A124: No; some submissions to Scored Event 2 were not scored. All **final** submissions were scored and ranked. Per A67 a best effort was made to collect and score multiple submissions. During SE-2, new submissions were collected every 10-20 seconds; in the case of multiple submissions from the same team to the same challenge set within that time window, the latest submission was scored and the rest were not scored. Five teams were affected.

Q123: Is mem_use(CB) the maximum RSS size for the CB process?

A123: No; mem_use(CB) is calculated congruent with the formula presented in the CB scoring walkthrough [1].

[1] https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/walk-throughs/scoring-cbs.md

Q122: Is exec_time(CB) the user CPU time + system CPU time for the CB process? Just user time? Or something else, e.g. wall clock? If wall-clock, what interval?

A122: exec_time(CB) does not include system time per A113; exec_time is user time measured with a sensor as described in A113.

Q121: Are mem_use(CB) and exec_time(CB) computed from the values we see in ru_maxrss, ru_minflt and (ru_utime + ru_stime), respectively, in the rusage structure returned by a wait3/wait4 system call from a parent?

A121: The rusage data structure returned by wait3/wait4 contains ru_maxrss and ru_minflt that are used to compute mem_use(CB), as described in CB scoring walkthrough [1]. DARPA/CGC does not derive timing measurements from the rusage data structure per A113.

[1] https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/walk-throughs/scoring-cbs.md

Q120: How are the measurements for all the service polls against a CB aggregated into a single mem_use(CB) or exec_time(CB) component? Is mem_use(CB) the highest maximum memory usage across any service poll/input, or is it an average? Likewise, is the exec_time(CB) the average across service polls?

A120: These calculations are performed in a manner congruent with the scoring CBs walkthrough [1].

[1] https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/walk-throughs/scoring-cbs.md

Q119: File size and memory usage restrictions imposed on Replacement Binaries appear very severe and potentially not representative of real-world constraints; in the case of memory usage restrictions, per-page granularity imposes a heavy overhead burden due to relatively small size of Challenge Binaries. Can these limits be adjusted and memory measurement granularity increased?

A119: Cyber Grand Challenge does not impose performance penalties by measuring the increase in performance between a competitor submitted RB and the original CB. Performance penalties are assessed based on the difference between a successfully defended (patched) CB created by a human expert and a competitor submitted RB [1]. Thus a defense is proven to exist and competitor defenses are measured against known successful defenses. From its inception, the CGC has set the ambitious goal of measuring the skill of automated systems against the abilities of human experts. While some defensive competitions have imposed absolute performance limitations of 5% or less [2], CGC imposed a graceful degradation curve and opened this curve to public comment in early 2014. To provide continuity to all competitors this mechanism will not be modified after the period of public comment per A100. On the topic of granularity of memory measurement, the CGC adopted the use of *pages*, a MMU construct; traditionally human experts have skillfully optimized the use of pages.

[1] https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/CQE%20Scoring.pdf
[2] http://www.microsoft.com/security/bluehatprize/rules.aspx

Q118: When handling multiple CBs in a CS, is a crash in any or all of the CBs considered proof of a vulnerability?

A118: Yes, a crash in any or all CBs in a CS that meets the criteria described in A28 is considered to have proven a vulnerability.

Q117: Will every Challenge Set in the CQE Corpus be scored as part of the CQE event?

A117: No. There will be over 100 Challenge Sets in the CQE Corpus; a small number of these are purely diagnostic and will be excluded from computing CQE ranking.

For each of these *excluded CSs:*

- PoVs against these CSs will not be part of CQE Consensus Evaluation
- CS Scores will not be used for any part of CQE competitor ranking
- CS Scores will not be considered for any part of section 3.1.3 of the CGC Rules

To prove that excluded CSs were chosen in advance, DARPA/CGC will provide a cryptographically verifiable proof of this selection prior to CQE. After CQE, a selection document matching this proof will be released. *Competitors are advised to field a CRS that attempts to solve every CS in CQE.*

Q116: What is the maximum size of the uncompressed CQE Challenge Bundle?

A116: Size will not exceed $5000x1024^2$ bytes.

Q115: Why were there no vulnerabilities that lacked a reference PoV in the ranking of Scored Event 2?

A115: DARPA/CGC has refined its procedures following Scored Event 1. Note that due to the potential conflict of interest involved in making CB changes while simultaneously handling de-anonymized scoring material and PoVs, the DARPA/CGC scoring team cannot determine flaw uniqueness or formulate CB repairs per Attack 1 in A110. See also A9 and A97.

Q114: The public specification for cb-replay [1] indicates that the read tag accepts a timeout tag. This does not appear to be implemented. Will the cb-replay implementation be adjusted to match the spec before CQE?

A114: No -- the mismatch will be maintained through CQE for continuity purposes.

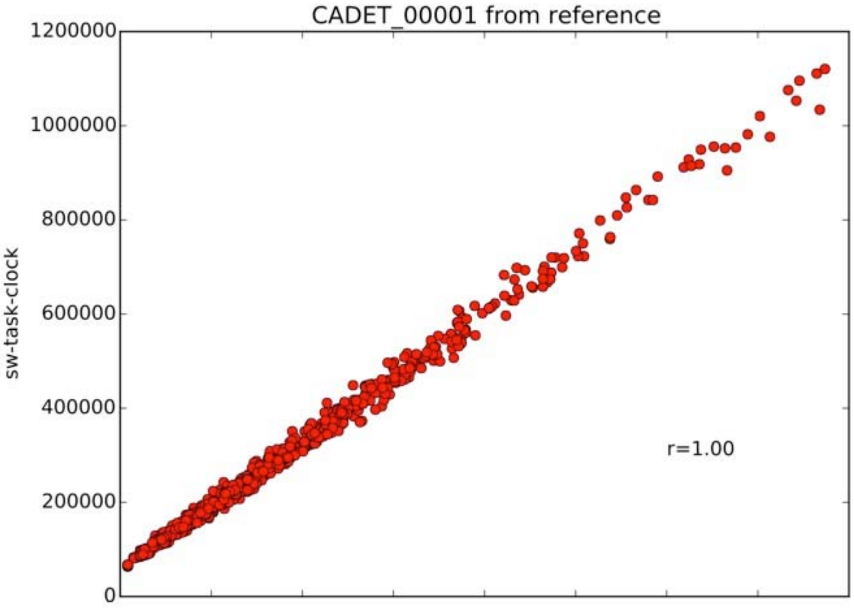
[1] https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/replay.dtd

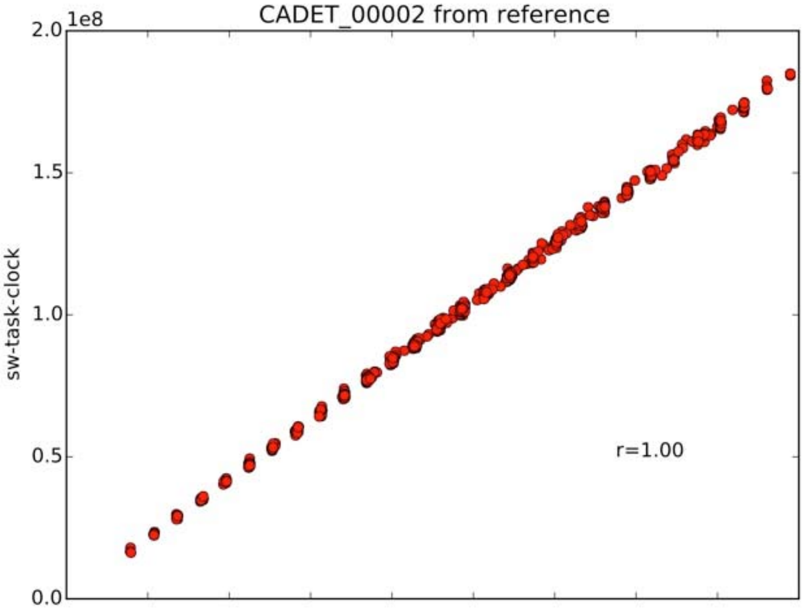
Q113: Following up A99 in the FAQ, is additional information available about the correlation between operating system timing estimates and DARPA/CGC sensor timing following SE2?

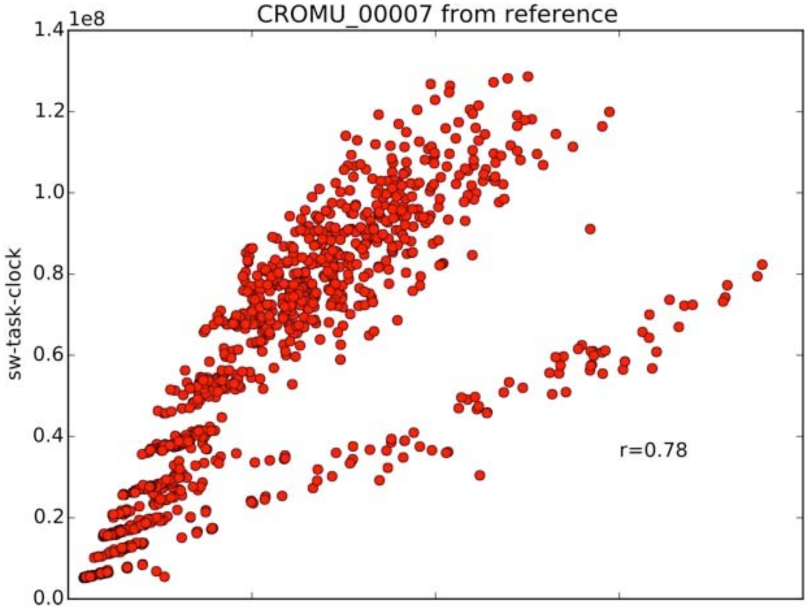
A113: The DARPA/CGC team's timing sensor was designed to measure process execution time with unprecedented accuracy and precision. This sensor stops measuring time immediately during the system call transition from user to kernel, and resumes measuring time immediately during the transition from kernel to user. Unfortunately, the task clock provided by Linux is not as rigidly accurate during this transition and can measure some kernel execution as user execution, causing divergence between the DARPA/CGC sensor and estimates based on the system clock (using the improved estimation facility in DECREE described in A99). This divergence increases based on the number and type of system calls made during execution. The random() call, in particular, is a major source of divergence. The graphs provided show the correlation results of the DARPA/CGC sensor and the system clock, with divergence increasing when large numbers of system calls are made. DARPA/CGC expects that for most Challenge Set execution (unburdened by a preponderance of system calls) the correlation between the task clock estimate and the high accuracy DARPA/CGC sensor should be high. Graphs of this correlation over samples from the service polls for SE-2 Challenge Sets are provided below. Please note the divergent graph for YAN01_00006, a unique CB whose execution is predominantly system calls.

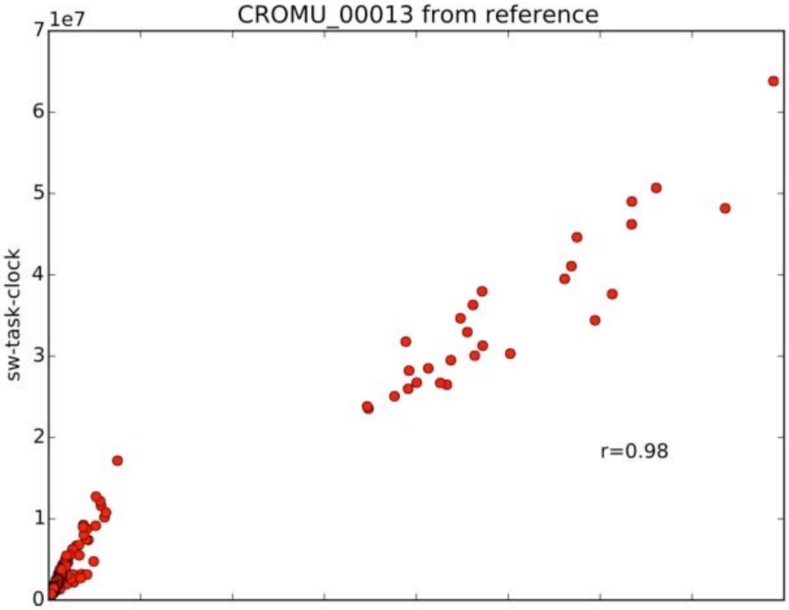
The following plots show linear correlation; an r value [1] of 1.00 denotes maximal positive linear correlation (optimal). The x-axis is generated by the output of the DARPA/CGC sensor; the y-axis is generated by the OS-based estimates issued by DECREE in A99. Each dot in the scattergram is a single service poll.

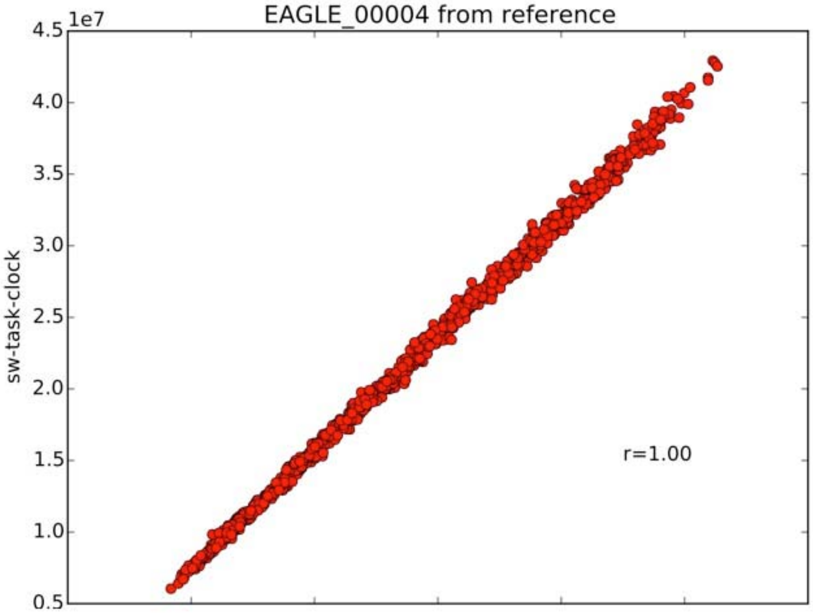
[1] http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient

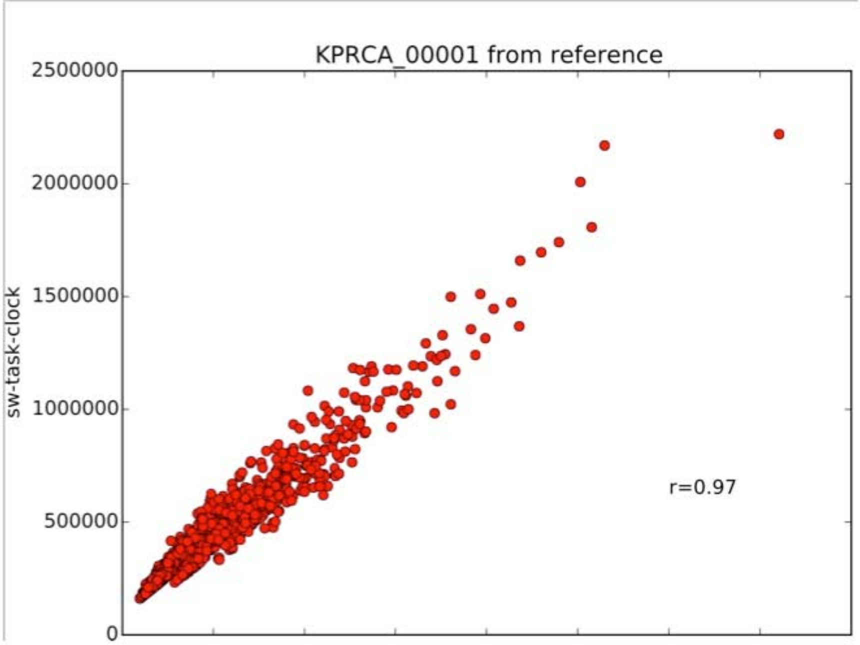

CADET_00001 from reference
sw-task-clock
1200000
1000000
800000
600000
400000
200000
0
r=1.00

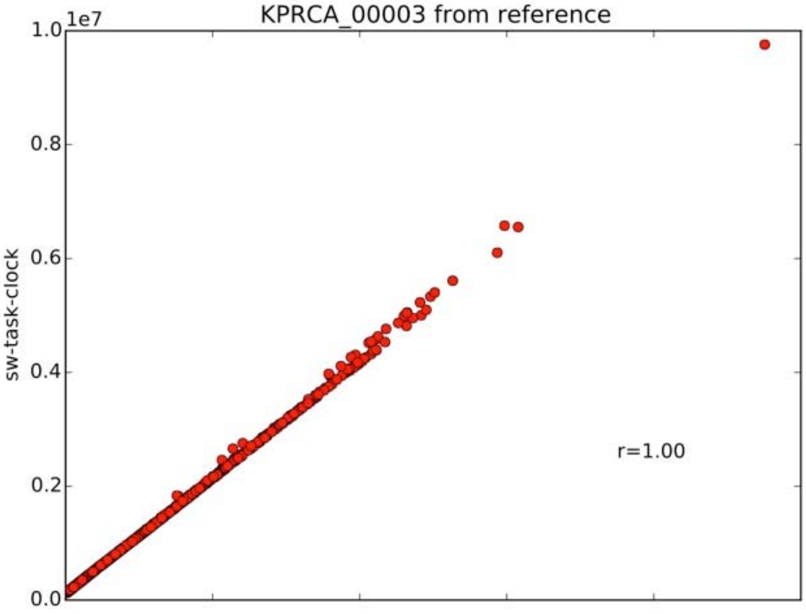

CADET_00002 from reference
1e8
sw-task-clock
2.0
1.5
1.0
0.5
0.0
r=1.00

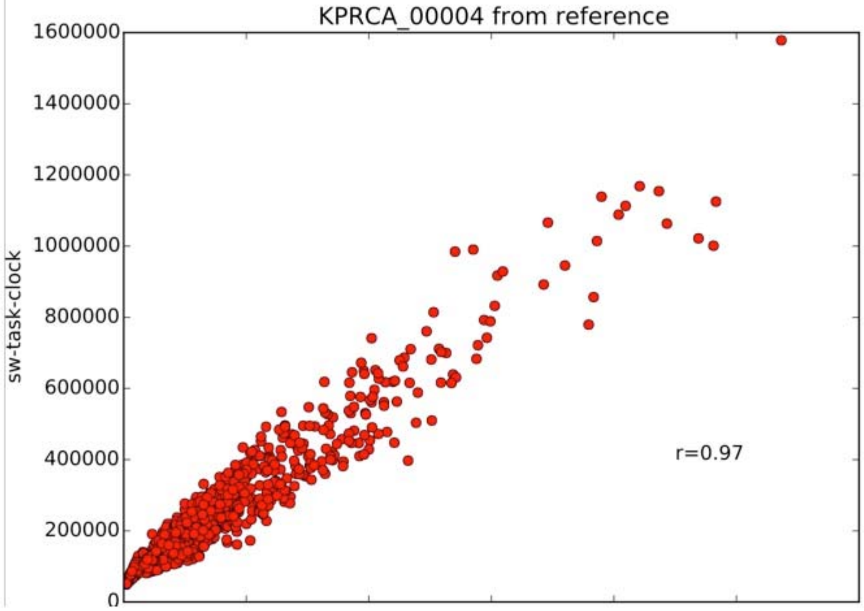

CROMU_00007 from reference
1e8
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
sw-task-clock
r=0.78

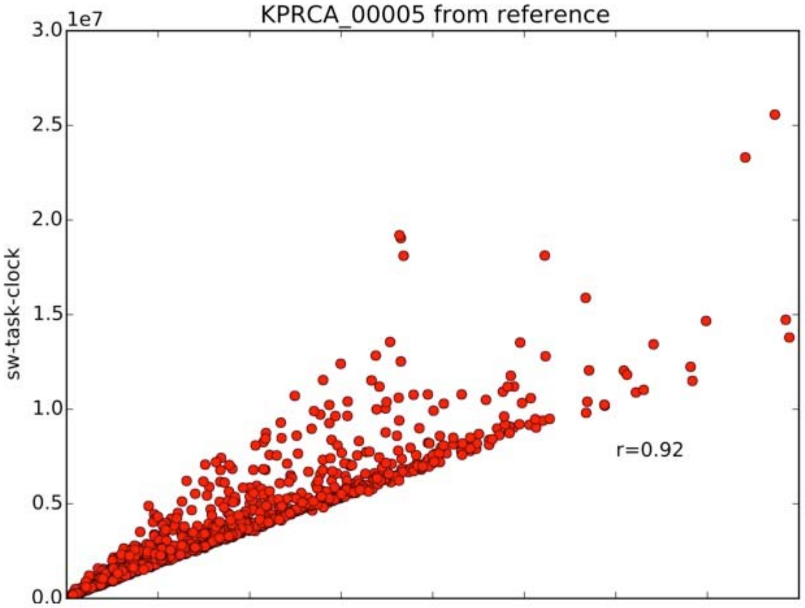

CROMU_00013 from reference
1e7
7
6
5
4
3
2
1
0
sw-task-clock
r=0.98

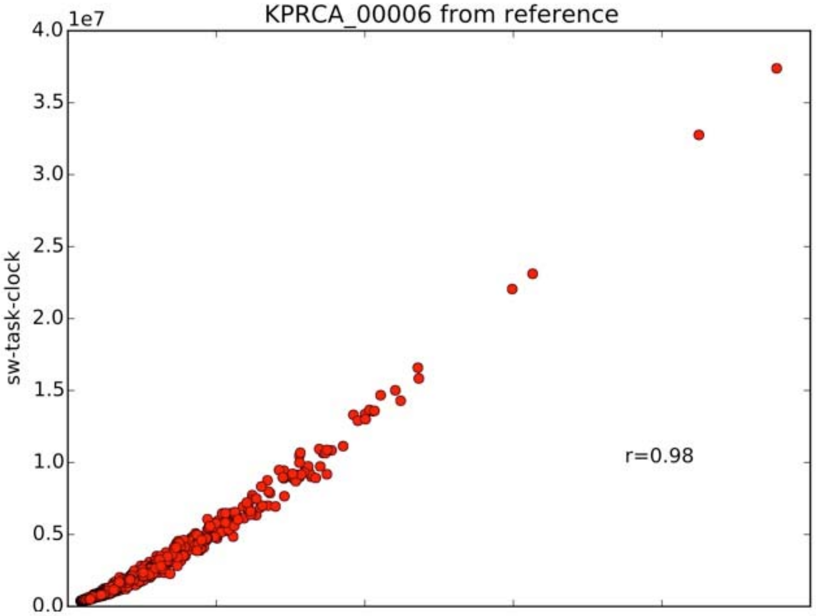

EAGLE_00004 from reference
1e7
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
sw-task-clock
r=1.00

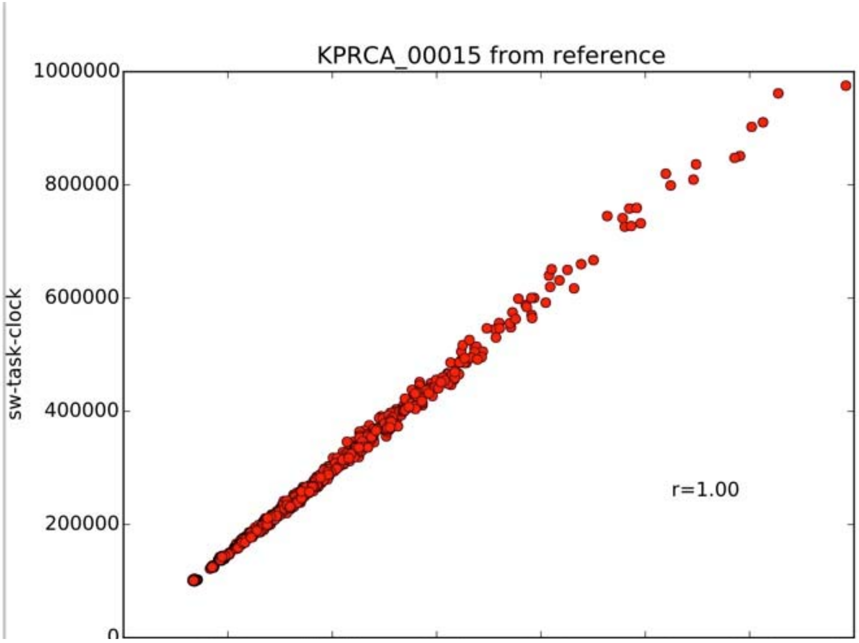

KPRCA_00001 from reference
2500000
2000000
1500000
1000000
500000
0
sw-task-clock
r=0.97

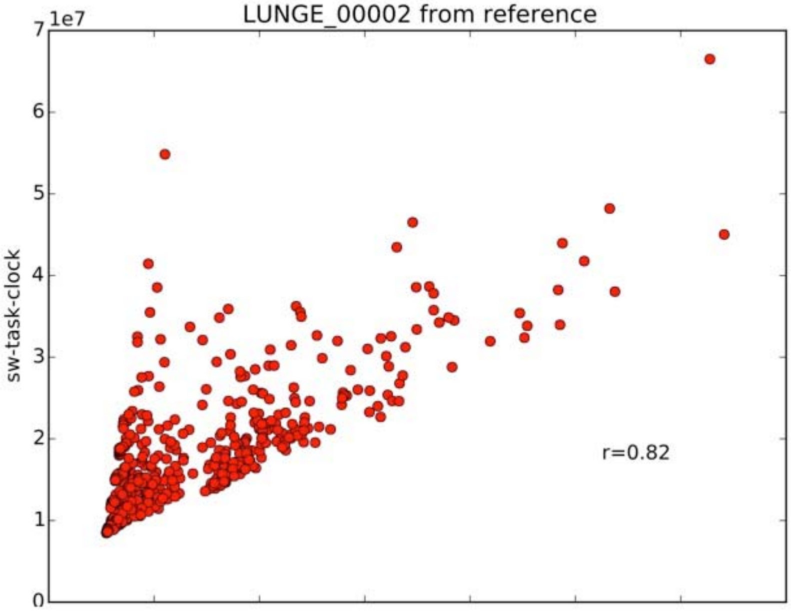

KPRCA_00003 from reference
1e7
1.0
0.8
0.6
0.4
0.2
0.0
sw-task-clock
r=1.00

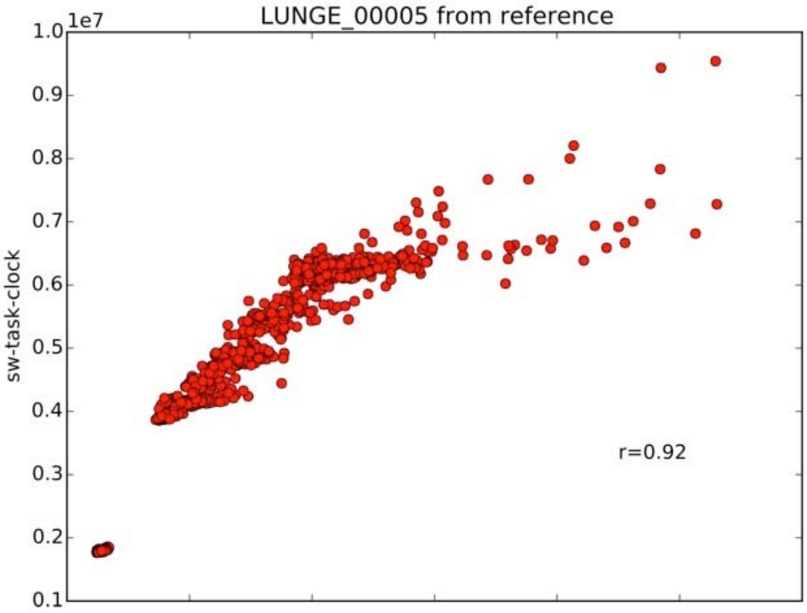

KPRCA_00004 from reference
1600000
1400000
1200000
1000000
800000
600000
400000
200000
0
sw-task-clock
r=0.97

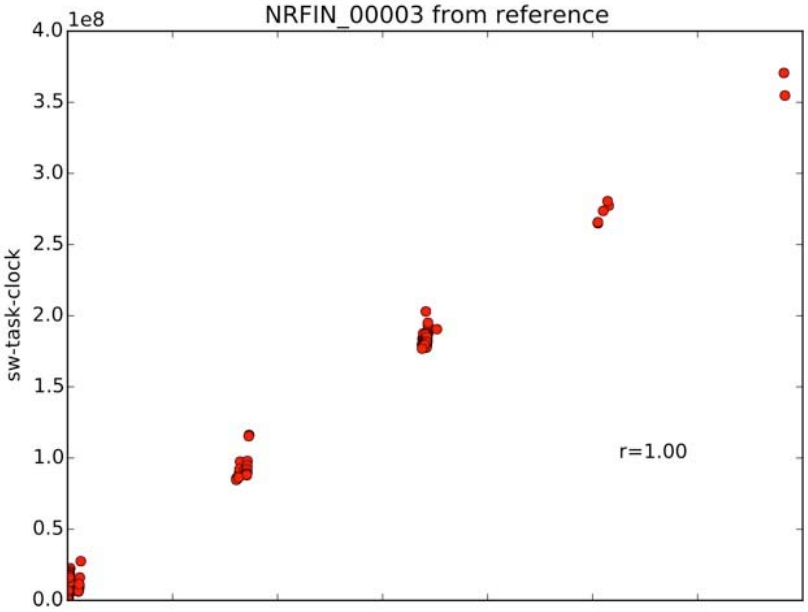

KPRCA_00005 from reference
1e7
3.0
2.5
2.0
1.5
1.0
0.5
0.0
sw-task-clock
r=0.92

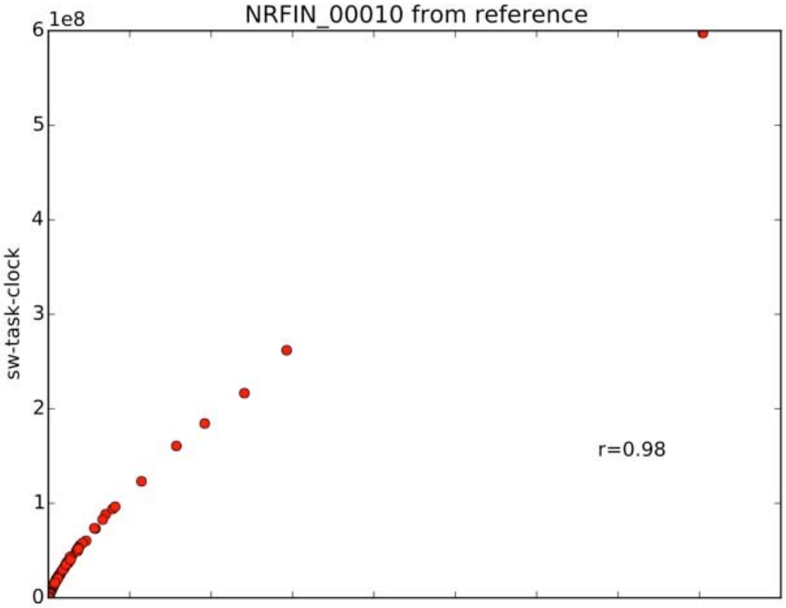

KPRCA_00006 from reference
1e7
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
sw-task-clock
r=0.98

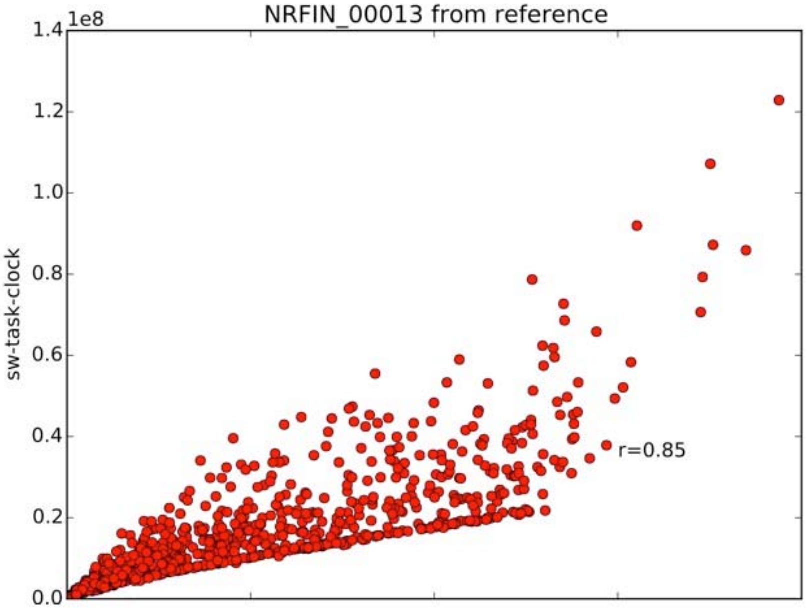

KPRCA_00015 from reference
sw-task-clock
1000000
800000
600000
400000
200000
0
r=1.00

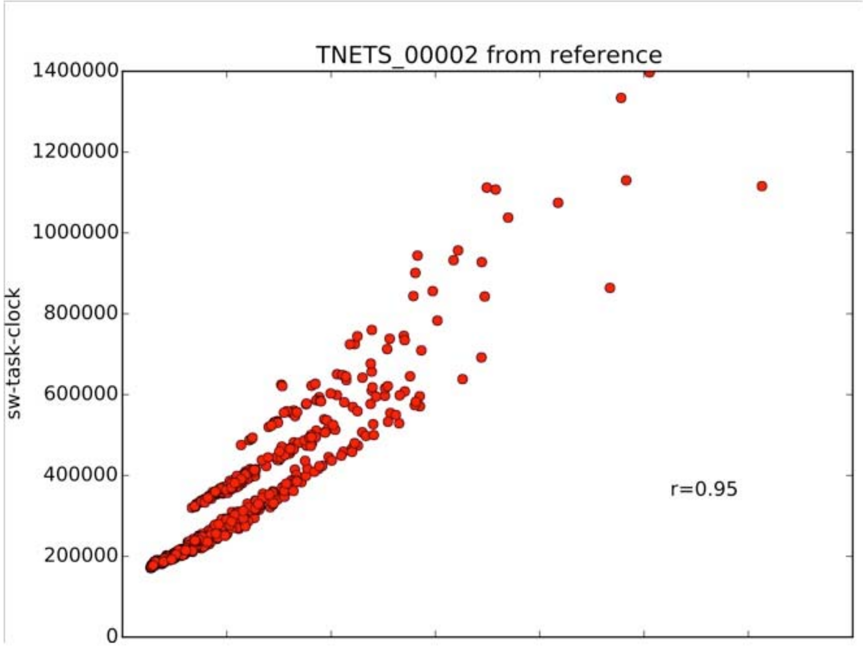

LUNGE_00002 from reference
1e7
sw-task-clock
7
6
5
4
3
2
1
0
r=0.82

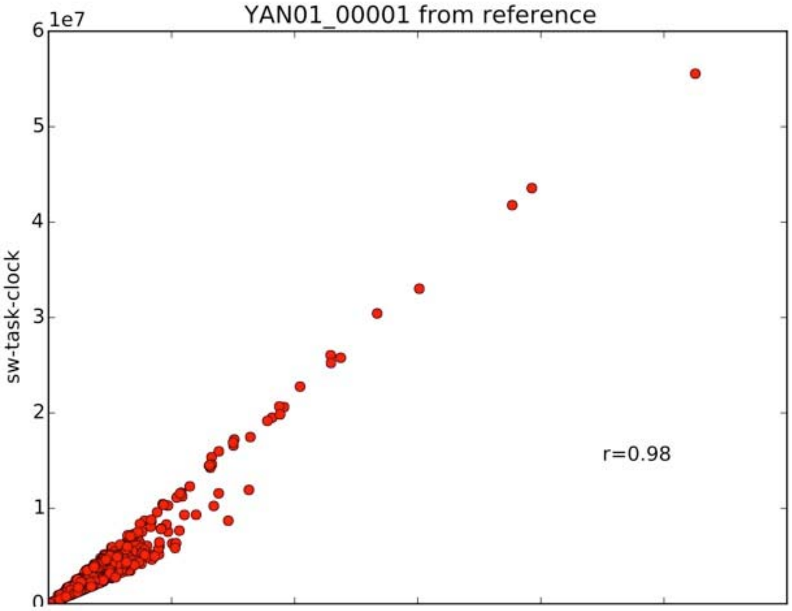

LUNGE_00005 from reference
1e7
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
sw-task-clock
r=0.92

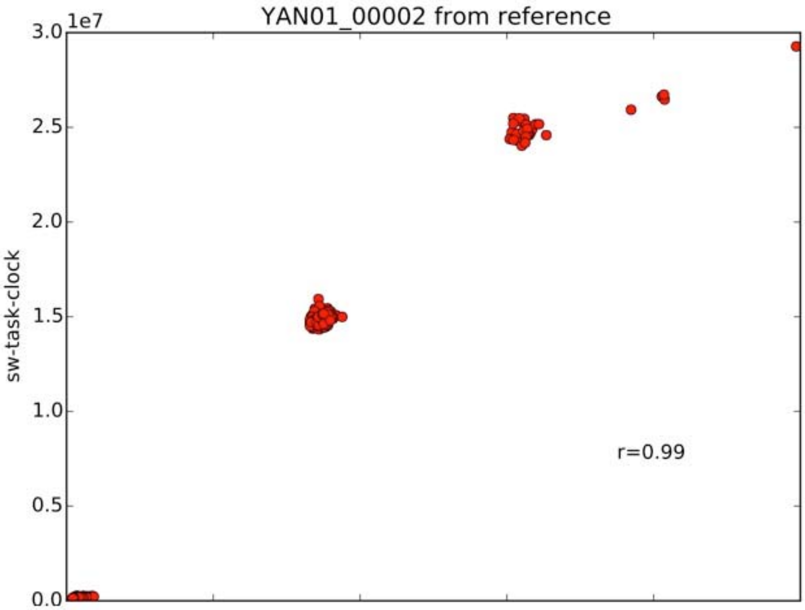

NRFIN_00003 from reference
1e8
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
sw-task-clock
r=1.00

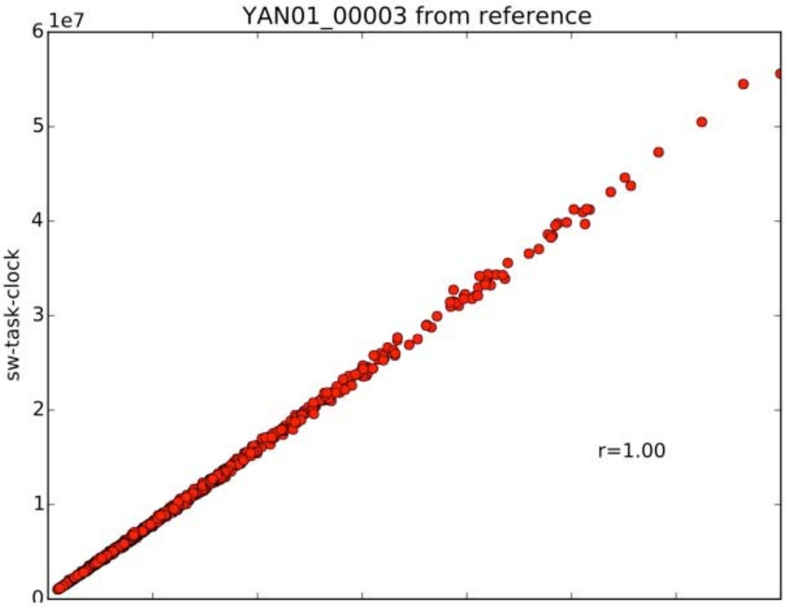

NRFIN_00010 from reference
1e8
6
5
4
3
2
1
0
sw-task-clock
r=0.98

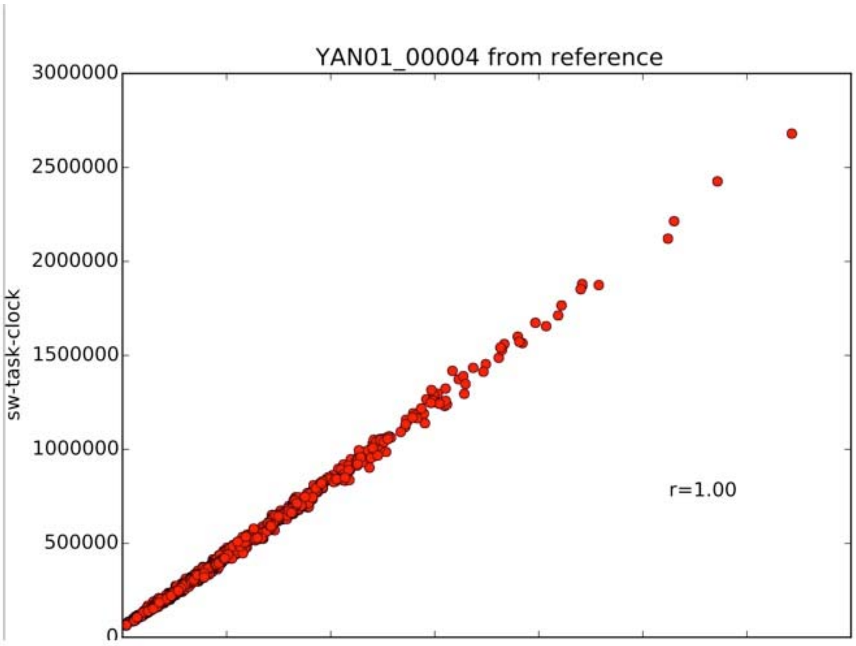

NRFIN_00013 from reference
1e8
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
sw-task-clock
r=0.85

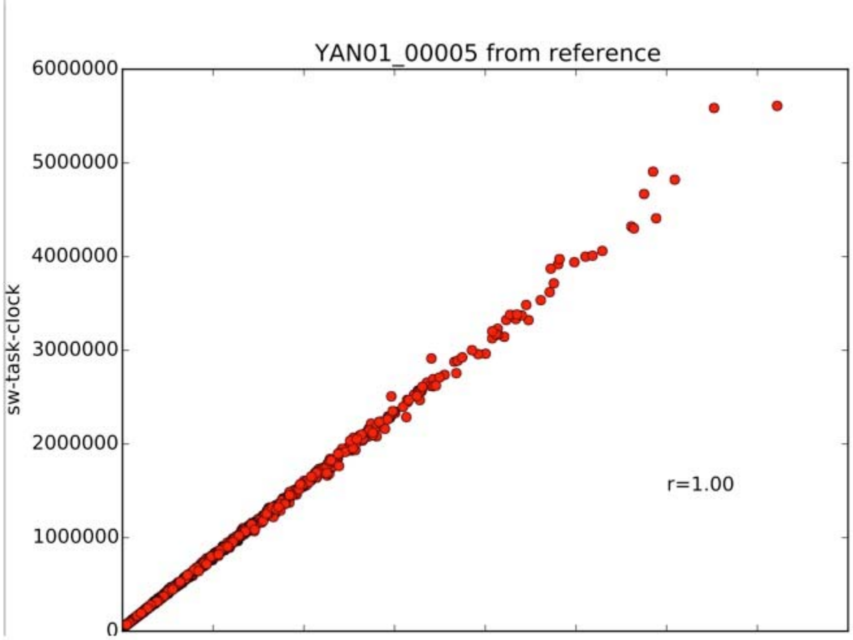

TNETS_00002 from reference
1400000
1200000
1000000
800000
600000
400000
200000
0
sw-task-clock
r=0.95

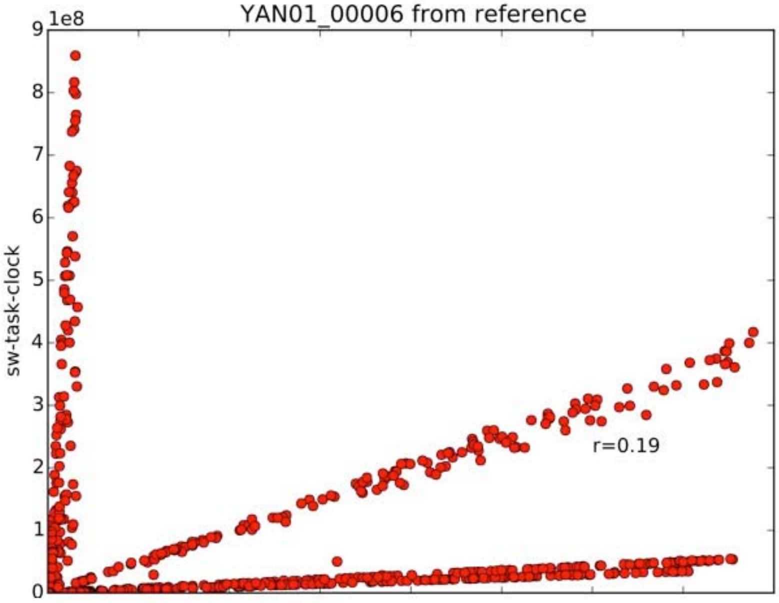

YAN01_00001 from reference
1e7
6
5
4
3
2
1
0
sw-task-clock
r=0.98

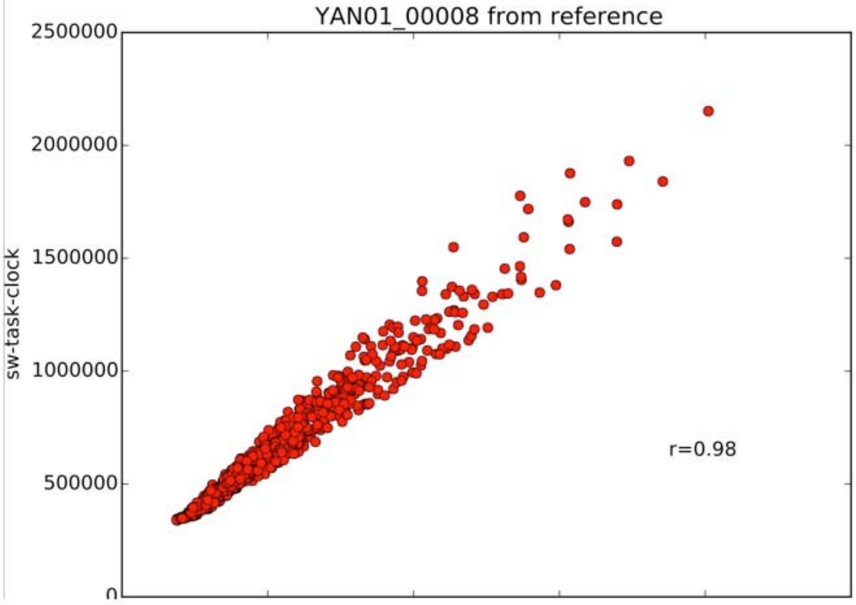

YAN01_00002 from reference
1e7
3.0
2.5
2.0
1.5
1.0
0.5
0.0
sw-task-clock
r=0.99


YAN01_00003 from reference
1e7
6
5
4
3
2
1
0
sw-task-clock
r=1.00


YAN01_00004 from reference
sw-task-clock
3000000
2500000
2000000
1500000
1000000
500000
0
r=1.00


YAN01_00005 from reference
sw-task-clock
6000000
5000000
4000000
3000000
2000000
1000000
0
r=1.00





Q112: What is the methodology to find my submissions in the released SE-2 scoring data?

A112: *Revised in FAQ v16*: Replacement Binary scores and PoV scores have been released as separate .csv files at cgc.darpa.mil; per the bundled instructions the SHA256 of each RB and PoV is used to calculate the **RB and PoV identifiers** used in these .csv files. Please note that these identifiers are not provided by the *submission* *verification* script [1]; these identifiers should be calculated separately using the *solution package verification* script [2].

[1] https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/scripts/cqe_submission_verification.py
[2] https://github.com/CyberGrandChallenge/cgc-release documentation/blob/master/scripts/cqe_verify_solution_package.py

Q111: During the CQE, what kind of human involvement is allowed? Although we have read Section 4 (Full Automation Requirement) of the CGC Rules, it remains unclear what "human assistance in cyber reasoning processes" means, exactly.

On a scale of what we imagine to be most reasonable to the least reasonable, here are a few examples about which we are curious (of course, specific examples such as splitting up pcaps are things we will handle as we have thought of them, these are just examples we have encountered to help motivate our questions):
**** power or hardware failure and we manually reboot servers;***
**** kernel panic or OOM and we manually reboot servers;***
**** launching VMs/scripts (such as for fuzzing or submitting binaries) by hand at the start of the CQE;***
**** large pcap causes us to manually split it up before feeding it to our CRS;***
**** trivial bug (such as misnamed python variable) in our CRS, so we fix the bug and restart the CRS;***
**** binary uses behavior we didn't account for (such as writing to STDIN) and we patch our system to account for this;***
**** bug in our CRS causes us to miss the faulting address for a crash, so we fix the bug and rerun that portion of the CRS;***
**** crash is not reproducible, so we manually fix it up***
**** CRS cannot find crash, so we manually do it***

The first three we currently deem as "in scope". The last two we currently deem "out of scope". The four in the middle we are very unsure of.

A111: DARPA/CGC believes that the issue of human interference in cyber reasoning processes is clear-cut and can easily be differentiated by any of our expert competitors in this field. The DARPA Competition Agreement governs human interference in cyber reasoning functions: any function that interacts with Challenge Sets and makes decisions about how to formulate Replacement Binaries and PoV's. A change to a reasoning function will change the outcome of a system's performance in any Area of Excellence [1].

When a human performs a cyber reasoning process on the contents of the Challenge Bundle, the resulting knowledge is ***fruit of the poisonous tree****. After gaining such knowledge from any source, a human may not modify the CRS.*

Below, this test is applied to the list of examples provided in the question:

1> Action: power or hardware failure and we manually reboot servers

- Human did not possess tainted knowledge. No integrity issue.

2> Action: Kernel panic or OOM and we manually reboot servers

- Human did not possess tainted knowledge. No integrity issue.

3> Action: Launching VMs/scripts (such as for fuzzing or submitting binaries) by hand at the start of the CQE

- Human did not possess tainted knowledge. No integrity issue.

4> Action: Manually un-packaging Challenge Bundle and feeding contents into CRS.

- Human did not possess tainted knowledge. No integrity issue.

5> Action: Manually packaging results from the CRS for submission to DARPA.

- Human did not possess tainted knowledge. No integrity issue.

6> Action: Large pcap causes us to manually split it up before feeding it to our CRS

- Human did not possess tainted knowledge. No integrity issue.

7> Action: Trivial bug (such as misnamed python variable) in our CRS, so we fix the bug and restart the CRS

- This example describes an unanticipated system bug that occurred due to automated processing of the Challenge Bundle:

  - If the CRS simply crashes on input and the bug can be fixed through CRS code inspection, no tainted knowledge was involved; no integrity violation.
  - If a human examines the Challenge Bundle and performs a cyber reasoning process to arrive at the knowledge used to patch the bug, an integrity violation has occurred*.* The knowledge used to patch the bug derived from human examination of the Challenge Bundle, and is therefore *fruit of the poisonous tree*.

  For example, if a CRS contains a flaw that is designed to crash on the first CB processed, it is not an integrity violation to use the crash dump to identify this flaw as long as inspection of the crash dump does not yield tainted cyber reasoning knowledge gained from the Challenge Bundle. However were a team to examine the Challenge Bundle in an attempt to "optimally fix" this flaw, perhaps by inserting new PoV seed cases, an integrity violation would have occurred.

8> Action: Binary uses behavior we didn't account for (such as writing to STDIN) and we patch our system to account for this

- Challenge Sets whose specification is unanticipated by teams may cause faults and should be handled per case 7, above.

9> Action: Bug in our CRS causes us to miss the faulting address for a crash, so we fix the bug and rerun that portion of the CRS

- In this hypothetical scenario:

  a. A bug exists in the dynamic analysis/concrete input portion of a CRS.
  b. This bug causes a false negative in which the CRS does not identify a crash.
  c. A human uses a debugger to identify the false negative.
  d. Human uses bug report from 9.c to fix the CRS bug and restarts CRS.

  An integrity violation occurred in step 9.c when a human performed a cyber reasoning process on the Challenge Bundle, and then used fruit of this poisonous tree to identify a flaw and formulate a fix in the CRS.

10> Action: Crash is not reproducible, so we manually fix it up

- In this hypothetical scenario:

  a. A human examines the output of a CRS in a debugger or test harness by testing the CRS-generated PoV with the CB provided in the Challenge Bundle.
  b. A human determines, through iterative testing and analysis using the data components in 10.a, a higher reliability formulation for the PoV.
  c. Using the knowledge from 10.b, a human reprograms the CRS to emit a newly formulated PoV.

  Due to the use of the CB provided by the Challenge Bundle in 10.a, the change made in 10.c is fruit of the poisonous tree; this is an integrity violation.

11> Action: CRS cannot find crash, so we manually do it

- Human PoV generation requires cyber reasoning; the creation of PoVs is an Area of Excellence [1]; integrity violation.

[1] Areas of Excellence are described in the CGC Rules, Section 1.3

Q110: Some CBs in SE1 contained vulnerabilities that were not proven by the reference PoV set. A CRS that patches a flaw proven by a reference PoV can consistently increase the Reference component of its Security score. On the other hand, a CRS that patches a flaw not tested by the reference PoV set can only improve the Consensus component of its Security score. This inconsistency appears to lead to a competition that may not consistently distinguish effective patching. Why can't the DARPA scoring team directly add CRS-discovered PoVs to the list of reference PoVs for the purpose of Security scoring?

A110: In order to mitigate Attack 3, below.

Q109: Why does the scoring system described in A58 only allow for one PoV per CB?

A109: In order to mitigate Attack 1 and Attack 2 below.

Q108: Why wasn't a PoV weighted by the number of RBs it proves vulnerable?

A108: In order to mitigate Attack 2 below.

Q107: Why wasn't an RB's Security score weighted by the total number of PoVs it defends against?

A107: In order to mitigate Attack 1 & Attack 4 below.

CQE Threat Modeling

Below is a list of mitigated attacks against the integrity of the CGC competition. DARPA has made every effort to increase competitor confidence in competition results by mitigating attacks against competition integrity. CQE scoring is the result of extensive threat modeling. A subset of this threat modeling exercise is included in the list of mitigated attacks below that are provided as a reference point for answers about specific scoring design decisions:

> Attack 1: PoV Stuffing
>
> This attack requires a scoring system in which a CRS may submit and receive credit for multiple PoVs that evaluate a single CB. To mount this attack, a CRS uses knowledge of a flaw to generate a great number of PoVs that prove a single flaw in a CB. For example, in the case of a simple buffer overflow, a CRS could submit thousands of PoVs that overflow the same buffer with different contents. This attack cannot be easily distinguished from a CRS that has accomplished the more difficult feat of generating multiple PoVs that prove multiple flaws. Attempts to determine when a PoV proves a "unique flaw" requires subjective judgments on behalf of the DARPA/CGC scoring team; see A9.

Attack 2: Sock Puppet Crash Dummies

In this attack, a rogue competitor, in violation of the DARPA competition agreement, stands up multiple "sock puppet" teams. A set of these sock puppet teams intentionally introduce "keyed" flaws in their replacement binaries (RBs) that require unique knowledge (a "key") to prove. In violation of the DARPA competition agreement, the sock puppets share this knowledge with the rogue competitor. This rogue competitor provides consensus evaluation PoVs for the "keyed" flaws, thereby improving its Evaluation score without actually finding legitimate software flaws. Many subtle variations of this simple attack exist.

Attack 3: Reference Collusion

In this attack, a rogue competitor, in violation of the DARPA competition agreement, stands up multiple "sock puppet" teams. In violation of the DARPA competition agreement, the sock puppets collude with the rogue competitor, only introducing PoVs that prove flaws known to be patched by the rogue competitor. This attack allows Security scoring to be dominated by flaws chosen by the rogue competitor.

Attack 4: Consensus Collusion

In this attack, a rogue competitor, in violation of the DARPA competition agreement, stands up multiple "sock puppet" teams. Each of these sock puppets introduces a valid set of PoVs containing a "key". In violation of the CGC Rules, the sock puppets collude with the rogue competitor, sharing the "key" value. The rogue competitor is able to artificially improve its Security score without actually mitigating a flaw by submitting RBs that block any input containing the "key". Many subtle variations of this simple attack exist.

Q106: Will additional Challenge Binary specifications be released that exceed the specificity of the existing documentation [1]?

A106: No. Cyber Reasoning Systems should reason about software per the CGC Rules [2].

[1] https://github.com/CyberGrandChallenge/cgc-release-documentation
[2] https://cgc.darpa.mil/documents.aspx

Q105: How will DARPA/CGC prove that all PoVs scored during CQE could cause unlimited execution of arbitrary code?

A105: DARPA/CGC does not assert or intend to prove this. Software vulnerabilities can lead to a variety of effects ranging from denial of service to code execution; DARPA/CGC will accept Proof of such Vulnerabilities in SE2 and CQE per A28.

Q104: We're unsure that our approach will work on the DARPA/CGC scoring apparatus. Can you guarantee the success of our approach?

A104: No. DARPA/CGC recommends that adventurous or risky use of hardware or software features be evaluated thoroughly during Scored Events. Commercial security software is often required to operate in uncertain computing environments and relies on hardware interrogation (such as CPUID) and software interrogation (such as O/S version numbers) to proceed.

Q103: Did DARPA/CGC change the way memory usage is quantified or measured in DECREE?

A103: No. Memory usage measurement has remained consistent with the formula reviewed during public comment, published in A58 and detailed in the DECREE documentation [1]. In an attempt to assist competitors with estimating this quantity, DARPA/CGC built and continues to update a walkthrough to assist competitors in making their own measurements. One Linux quirk that could cause measurement problems for competitors is that child processes are assigned a maximum RSS value inherited from their parent. This causes estimation problems when a large parent process launches a small child process. In order to allow competitors to make higher-fidelity estimates, this quirk was disabled for CGC processes in DECREE [2]. DARPA/CGC will continue to work to create the most high fidelity implementation of the scoring methodology published in A58 as possible and provide assistance to competitors to make accurate measurements.

[1] https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/CQE%20Scoring.pdf
[2] https://github.com/CyberGrandChallenge/linux-source-3.13.2-cgc/blob/master/fs/binfmt_cgc.c#L300

Q102: In SE1, the released reference security scores are all either 0.0 or 1.0. Has this been confirmed?

A102: Yes.

Q101: While traditionally Capture the Flag contests and DARPA Challenges have never given competitors access to the scoring system and associated sensors prior to competition, could Cyber Grand Challenge become the first to allow open source access to the scoring system?

A101: DARPA/CGC will not allow competitor access to the scoring measurement code and its associated sensors due to the current lack of technology capable of making strong integrity promises about executable code subject to adversarial introspection. DARPA/CGC will continue to use an open, competitor-reviewed scoring algorithm described in this FAQ (A58/A59), allow periodic competitor access to the scoring system through the scheduled Scored Events, and provide measurement assistance to competitors through DECREE releases.

Q100: Can the scoring methodology be periodically updated similar to the DECREE releases?

A100: No. All competitors have had equal access and ability to contribute to the scoring algorithm during its period of public comment. Changing the scoring system mid-competition is disruptive to the overall competitor pool and will degrade competitor confidence. Competitors should be assured that scoring fundamentals will not be altered.

Q99: We're using the operating system to measure performance and the OS seems to miss significant increases or decreases in performance timing in our replacement CBs. What does DARPA suggest?

A99: DARPA/CGC achieved repeatable, high precision performance measurements using uniform hardware and hardware-supported measurements. This measurement apparatus builds on the work of Levinthal [1-2], Du et al [3-4], Weaver et al [5-6], and Zapanuks et al [7].

To assist competitors in achieving timing measurements that more closely track the DARPA/CGC measurement apparatus, DECREE has been updated to provide higher precision timing approximations to the competitors than the previous 10ms resolution clock. Please see https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/walk-throughs/scoring-cbs.md

DARPA/CGC performed 100 testing runs of the Scored Event 1 challenge bundle using both the DARPA performance timing apparatus and the task.clock based tool released in DECREE (above). The results of this experiment show that for a given challenge binary, ExecTimeOverhead possesses a mean difference from DARPA's scoring calculation of 0.5% with a 95% confidence interval of (0.441%, 0.576%). This experiment was performed on a Haswell microarchitecture with 25MB of cache and a 2.3Ghz base clock.

[1] Levinthal, D. Performance Analysis Guide for Intel Core i7 Processor and Intel Xeon 5500 processors 2009
[2] Levinthal, D. Cycle Accounting Analysis on Intel Core2 Processors 2009
[3] Du, J.; Sehrawat, N. & Zwaenepoel, W. Performance Profiling of Virtual Machines, SIGPLAN Not., ACM, 2011, 46, 3-14
[4] Du, J.; Sehrawat, N. & Zwaenepoel, W. Performance Profiling in a Virtualized Environment Proc. HotCloud 2010, USENIX, 2010
[5] Weaver, V. M.; Terpstra, D. & Moore, S. Non-Determinism and Overcount on Modern Hardware Performance Counter Implementations International Symposium on Performance Analysis of Systems and Software, 2013, 215-224

[6] Weaver, V. M. & McKee, S. A. Can Hardware Performance Counters be Trusted? International Symposium on Workload Characterization, 2008, 141-150
[7] Zaparanuks, D.; Jovic, M. & Hauswirth, M. Accuracy of Performance Counter Measurements IEEE Symposium on Performance Analysis of Systems and Software, 2009, 23-32

Q98: We're concerned that DARPA's performance measurements are not repeatable. Are they repeatable?

A98: Yes. In a process known as continuous integration testing, the SE1/SE2/CQE scoring process and its associated unit tests are continually re-tested and compared against hand-confirmed performance measurements. This testing corpus will grow to include all scored events. These unit tests include all competitor submissions to date. See A99.

Q97: Per the limited IP rights described in A81 and the CGC Rules is there a conflict of interest issue with the CQE technical paper; specifically, will these technical papers be used to manipulate the competition to defeat my described approach?

A97: No. DARPA/CGC was created with the intention of bringing the game of Capture the Flag to automated systems. CTF contests as a whole are designed to be objective tests of skill - a level playing field in which competitors from around the globe are tested solely on their ability to solve difficult reverse engineering problems. DARPA/CGC is a CTF competition in both letter and spirit.
Specifically:

- DARPA/CGC will never take any action to intentionally demote or promote any competitor or set of competitors.
- Construction of the challenge has been undertaken in an effort to mirror real world challenges.
- Per the IP rights specified in the CGC Rules and A81, neither the CQE nor the CFE technical reports will be shared with Challenge Binary author performer teams.

Q96: Regarding the CQE technical paper, I assume you want an overview of the tools, approaches, game theory and program analysis, but not a full design document and discussion of every algorithm and heuristic used. Correct?

A96: Yes

Q95: What happens when DARPA confirms acceptance of our CQE technical paper?

A95: Per section 3.1.3 of the CGC rules, CQE technical paper acceptance is required in order to advance as a finalist and receive CQE prizes in accordance with section 3.1.4 of the Rules.

Q94: In our testing, some of our PoVs don't seem to work reliably against some CBs. What should we do?

A94: DARPA has developed a high-reliability PoV harness for scoring SE1/SE2/CQE that may mitigate this issue somewhat during event scoring. In CFE, PoV reliability will be the responsibility of competitors.

Q93: Section 4 of the CGC Rules requests notification of potential threats to the integrity of the CGC competition. We are aware of a methodology to cause a total mapping of Linux process memory that does not cause Linux to report the memory as allocated or consumed. Will this threaten the integrity of the CGC competition? <METHODOLOGY REDACTED>

A93: No. Please note the following:

1> DECREE is not Linux
2> DECREE is a specification
3> A reference implementation of DECREE atop Linux has been released on github
4> DECREE and its associated reference implementations do not support the reported shenanigans
5> DARPA appreciates competitor support in conducting a high integrity competition

Q92: The reference security scores for replacement binaries for CB e7cd3901 in Scored Event 1 appeared to be incorrect. Can DARPA confirm those scores?

A92: An error in the SE1 scoring system incorrectly scored the reference PoV for e7cd3901 as failing to prove vulnerability. This resulted in inflated reference security scores for several replacement CBs. Corrected scores have now been posted at cgc.darpa.mil. DARPA would like to thank the FuzzBomb Team for reporting and helping diagnose this issue. The SE2 scoring system will incorporate all lessons learned from SE1. Competitor participation and feedback continues to be invaluable on the road to CQE.

Q91: Will the test_event buckets be maintained for team testing during the events?

A91: Yes, the test_event buckets will exist before and during scored events. Competitors are advised these buckets are a temporary space and files may be deleted by DARPA for resource consumption reasons at any time.

Q90: Will libcgc implement all LLVM intrinsics?

A90: No. Prior to CQE, DARPA will release the version of libcgc to be used during CQE.

Q89: Will the source code to the new challenges included in the scored events be released?

A89: Yes, source code will be released prior to the next event.

Q88: The Security:Reference score released in the SE1 scoring appears to have a maximum value of 1. Why is this?

A88: Per A59, the Reference component of the Security score has a maximum value of 1.

Q87: Will DARPA provide the total scores for the top ranked teams in Scored Event 1?

A87: No. See Q/A68.

Q86: Scored Event 1 included previously released/examined Challenge Binaries. How will these Challenge Binaries be treated in future CGC events?

A86:
Definitions:

- Previously released & examined Challenge Binaries will be referred to as Old CBs
- Challenge Binaries released for the first time during a competition event will be referred to as New CBs.

- Scored Event 2: A mix of Old CBs and New CBs. Only New CBs will be considered for ranking purposes.
- CQE: Scoring using New CBs only.
- CFE Trials: Old CBs and New CBs may be used.
- CFE: Scoring using New CBs only.

Please note that CQE & CFE are the only CGC events that involve prize authority.

Q85: Should teams expect packet loss in the sample packet captures released as part of the event challenge bundle?

A85: Yes. DARPA is evaluating the packet loss rate for Scored Event 2.

Q84: Were some PoV submissions not scored in Scored Event 1?

A84: No. All submitted PoVs were scored. In a few cases the CGC team had to fix the XML DTD path of submitted PoVs or add missing <xml> tags. Please run the script found at https://github.com/CyberGrandChallenge/Event-FAQ/blob/master/cqe-verify-solution-package.py to calculate the hashes used in the scoring release. In future events, the CGC team will not manually correct malformed submissions.

Q83: I'm trying to reproduce results on DECREE using mixed components from an old release and a new release. How should I proceed?

A83: DECREE and its associated test suites do not maintain backwards compatibility. Competitors should use the latest VM and tool suite release available.

Q82: Are challenge binary creators allowed to use inline assembler?

A82: The guidelines provided to CB authors can be found at https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/walk-throughs/submitting-a-cb.md

Q81: What is the distribution of the CQE Technical Paper required for teams to advance to CFE?

A81: This question is answered in detail in the CGC Rules, Section 5, Intellectual Property; technical papers are private communications to DARPA. DARPA will observe all limitations detailed in Section 5. DARPA will not allow access to technical papers by any party not listed in Section 5. Competitors wishing to use email encryption when sending technical papers may request a signed email from cybergrandchallenge@darpa.mil; the S/MIME public key associated with this email has the fingerprint 1E 48 B8 A5 EF C6 82 F9 2C 86 7B B6 D0 1A 34 55.

Q80: Will the CBs settle on a standard libc replacement that we can rely upon for analysis?

A80: No. See instructions provided to CB authors: https://github.com/CyberGrandChallenge/cgc-release-documentation/blob/master/walk-throughs/submitting-a-cb.md

Scoring Tie-Breaker comments (Answers Round1, Q78-79):

> ***Q79: Could DARPA use the time of last submission as a simple tiebreaker?***
>
> A79: DARPA does not wish to penalize international competitors who may: reside in locations with increased transit time to the CQE servers have access to slow, lower bandwidth connections reside in an inconvenient time zone relative to the start of CQE.

> ***Q78: Could the tiebreaker algorithm first remove the Proof-of-Vulnerability score to incentivize purely defensive solutions?***
>
> A78: The SE1 tiebreaker algorithm increases the importance of the Defensive score in the first four of its five computation rounds.

Q77: May I submit a PoV of arbitrary size?

A77: No. PoVs are limited to no greater than $10x1024^2$ bytes.

Q76: What constitutes a test case when calculating the retained functionality score in CQE?

A76: A test case is a single service poll within a single TCP connection. No TCP connection will contain more than one service poll. Individual service polls will either pass or fail. Replacement Challenge binaries will be tested by at least 1000 individual service polls. These results will be used to calculate retained functionality per A59.

Q75: How will DARPA communicate with the teams during the Scored Events and CQE?

A75: Competitors can email questions to cybergrandchallenge@darpa.mil during the events. An event FAQ will be updated as needed on github at https://github.com/CyberGrandChallenge/Event-FAQ in order to disseminate information to all teams per A70. The event FAQ will be incorporated into this document after the completion of each event.

Q74: I am a foreign national who is eligible to participate per the CGC Rules. I have created a US-based LLC with a US-based Registered Agent to serve as the Entrant for my CGC team; this LLC is also eligible to participate per the CGC Rules. Is this approach compliant with the CGC Rules?

A74: Yes.

Q73: What happens when a connection is made to a DECREE service?

A73: *inetd-style.* A new instance is created to handle the new connection. This new instance is torn down after the connection terminates.

Q72: What types of connections will be made during CQE scoring?

A72: Multiple connections will be made from Service Pollers. Multiple connections will also be made from Proof of Vulnerability modules. Service Polls and PoV modules will never share connections.

Q71: What types of connections will be made during CFE scoring?

A71: Multiple connections will be made from service pollers. Multiple connections will also be made from logic built by competitors. Service polls and competitor logic will never share connections.

Q70: What other access to Cyber Grand Challenge is available to competitors outside of the cybergrandchallenge@darpa.mil email box and the FAQ responses?

A70: In the interests of conducting a fair and equitable global competition, access to challenge information is made available electronically to all competitors. All competitors whether next door or across the globe, may submit questions through the mailbox, and responses will be communicated through this FAQ.

Q69: Are CFE finalists required to bring hardware to compete in CFE?

A69: No. Finalists will have the option of either:

1. Bringing a competition system to CFE in accordance with A31, *or*
2. Competing in CFE on a DARPA-provided compute cloud instance after having accepted the DARPA Cloud Agreement.

Each DARPA-provided compute cloud instance will be on the order of hundreds of x86-64 cores.

Further details regarding the Cloud Agreement and system specifications will be released at a later date.

Q68: What information will be released to competitors after Scored Event #1?

A68: Please note that information release after Scored Events will be entirely different from the post-CQE information release addressed in A25. After Scored Event #1, the following information will be released publicly:

- The names of the seven top-scoring teams in rank order.
- A list of SHA-256 hashes for submitted Challenge Binaries and their associated scores and corresponding reference CB name.
- A list of SHA-256 hashes for PoVs and their associated scores and corresponding reference CB name.

Please note, these released hash lists will not correlate scored submissions to teams. Competitors will be required to calculate SHA-256 hashes of their submitted inputs in order to determine their scores.

Q67: How will ranking occur in Scored Event #1?

A67: Multiple submissions may be scored; hash list information on multiple submissions will be available via the hash list format (A68). Ranks will be determined using the score assigned to each team's final submission.

Q66: What will CQE Challenge Bundle contain?

A66: At the beginning of CQE, competitors will gain access to CQE Challenge Bundle (bundle will contain a collection of Reference CBs, as well as some pcap recordings of some service poll interactions between Service Pollers and these Reference CBs). These service poll interaction samples, where present, are not guaranteed to be complete.

Q65: What will Scored Event Challenge Bundles contain?

A65: Scored events are intended to provide technical preparation for CQE; therefore the Scored Event Bundles will mirror the format of the CQE Challenge Bundle to the greatest extent possible. Competitors should note that the CQE Bundle will be much larger than the Scored Event Bundles. These Scored Event Bundles may also re-use previously released CBs.

Q64: What is DECREE?

A64: DECREE is an open-source extension built atop the Linux operating system. Constructed from the ground up as a platform for operating small, isolated software test samples that are incompatible with any other software in the world—DECREE aims to provide a safe research and experimentation environment for the Cyber Grand Challenge.

DECREE binaries and source are available:
http://repo.cybergrandchallenge.com/
http://github.com/cybergrandchallenge/

Q63: How should issues in DECREE be reported?

A63: Email cybergrandchallenge@darpa.mil

Q62: Will all advanced application defenses that prevent arbitrary code from running increase the security score in CQE?

A62: No. CGC scoring does not require arbitrary code execution, therefore mechanisms which frustrate arbitrary code execution will not necessarily prevent scoring events. In CQE, competitors have the opportunity to mitigate denial of service flaws. See also Q4.

Q61: Will the Reference Patched CB perform differently than the Original CB?

A61: A diverse group of software authors are building a large corpus of CBs for CGC incorporating many classes of vulnerabilities. These CB authors are required to provide a single Reference Patched CB that passes the same functionality test suite as the Original CB and is not susceptible to any of the reference PoVs.

Q60: How does the Inter Process Communication (IPC) work in Challenge Binaries (CBs)?

A60: DECREE precludes communication via shared memory, network, or persistent storage between different CBs as well as different connections serviced by the same CB.

In order to offer a rich CB portfolio with broad CWE coverage including concurrency issues, DARPA allows for the use of a CGC IPC mechanism within a single CB, which works as follows. Each CB may be composed of multiple binaries running in distinct processes. The CGC competition framework will launch all of the binaries associated with the challenge. Each of these processes will be pre-connected with file descriptors to communicate with the others via receive() and transmit() system calls (see Figure 1) in a manner determined by the CB author (see Figure 2). Example IPC CBs will be provided prior to the first CQE Scored Event.

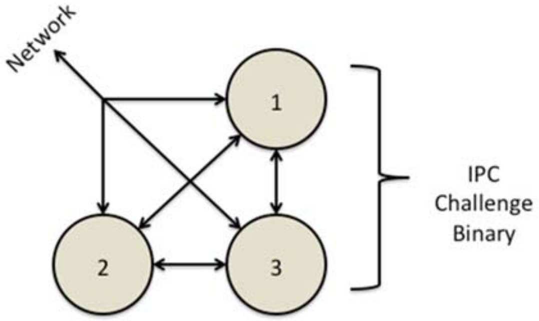


Figure 1: File descriptor connections

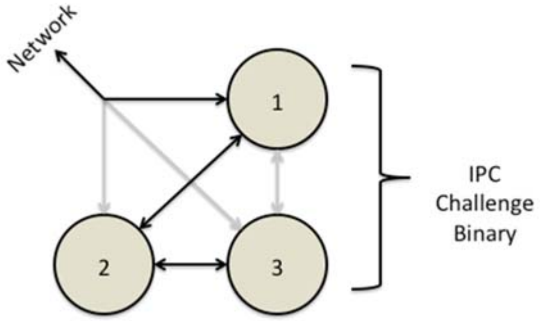


Figure 2: Example use of IPC inter-connections

Q59: What is the scoring method for CQE?

A59: DARPA held a period of public comment for the CFE and CQE scoring methods, and feedback from the CGC community was received and reviewed by DARPA. Updated scoring methods for CQE and CFE were released on Monday, March 10, 2014. These scoring methods are consistent with the Cyber Grand Challenge Rules. See also Q/A 26, 27, and 30.

CQE scoring is the product of three assessed quantities: Availability score, Security score, and Evaluation score. These scores map to the Areas of Excellence (AoE) located in the Cyber Grand Challenge Rules as follows:

Availability: Area of Excellence 4
Security: Area of Excellence 2
Evaluation: Area of Excellence 3

During CQE, the CBs distributed by DARPA shall be referred to as "reference CBs", not to be confused with replacement CBs submitted by competitors. PoVs used by DARPA for scoring will be referred to as "reference PoVs". PoVs submitted by competitor systems to DARPA during CQE will be referred to as "submitted PoVs".

DARPA will score one replacement CB and one submitted PoV per reference Challenge Binary. CQE Scores will be assessed per reference Challenge Binary ("CB score"). Each CB score will be calculated as follows:

Availability * **Security** * **Evaluation**

Availability:
This quantity shall vary as a multi-step function between 0 and 1, with 1 being a perfect score. **Performance** and **Retained Functionality** will be measured, with **Availability** being set to the minimum of these quantities.

- **Performance** of a submitted CB will decrement based on the greatest measured increase in system resource utilization. These measured increases include file size, execution time, and memory usage. For each of these measured quantities, a maximum acceptable increase is indicated below, after which **Performance** begins to decrement:
    - File size: +40%
    - Execution time: +10%
    - Memory usage: +10%

- **Retained Functionality** will be the percentage of test cases the replacement CB passes.

Competitors are advised that use of a multi-step function imposes a faster-than-linear **Availability** dropoff based on damage to **Performance** or **Retained Functionality**.

Security:
This quantity shall be determined using two quantities, **Reference** and **Consensus:**

- **Reference:** The number of reference PoVs which do not prove vulnerability in the replacement CB, divided by the number of reference PoVs

- **Consensus:** This quantity will be set to 0 or 1:
    - 0: Any submitted PoV proved vulnerability in the replacement CB
    - 1: No submitted PoV proved vulnerability in the replacement CB

If **Reference** is zero, **Security** will be set to zero.
If **Reference** is nonzero, **Security** will be calculated as follows:
1 + (**Reference** + **Consensus**)/2
Evaluation:
This quantity will be set to 1 or 2:
- 1: The PoV emitted by this CRS did not prove vulnerability in the reference CB
- 2: The PoV emitted by this CRS did prove vulnerability in the reference CB

A CRS's total score at the end of CQE shall be the sum of that CRS's CB scores.

Q58: What is the scoring method for CFE?

A58: DARPA held a period of public comment for the CFE and CQE scoring methods, and feedback from the CGC community was received and reviewed by DARPA. Updated scoring methods for CQE and CFE were released on Monday, March 10, 2014. These scoring methods are consistent with the Cyber Grand Challenge Rules. See also Q/A 26, 27, and 30.

CFE scoring is the product of three assessed quantities: Availability score, Security score, and Evaluation score. These scores map to the Areas of Excellence (AoE) located in the Cyber Grand Challenge Rules as follows:
Availability: Area of Excellence 4
Security: AoE 2 and 5
Evaluation: Area of Excellence 3

CFE scoring will take place over many rounds. Scores will be assessed per Challenge Binary per round ("CB round score"); it follows that the sum of the CB round scores for a single CRS will be the total score for that system. Each CB round score will be calculated as follows:

Availability * **Security** * **Evaluation**

Availability:
This quantity shall vary as a multi-step function between 0 and 1, with 1 being a perfect score. **Performance** and **Retained Functionality** will be measured, with **Availability** being set to the minimum of these quantities.

- **Performance** of a submitted CB will decrement based on the greatest measured increase in system resource utilization. These measured increases include file size, execution time, and memory usage. For each of these measured quantities, a maximum acceptable increase is indicated below, after which **Performance** begins to decrement:

- File size: +20%
- Execution time: +5%
- Memory usage: +5%

- **Retained Functionality** will be the percentage of network test cases the replacement CB passes.
  Competitors are advised that use of a multi-step function imposes a faster-than-linear **Availability** dropoff based on damage to **Performance** or **Retained Functionality**.

Security:
This quantity will be set to 1 or 2:
- 1: At least one PoV from a competitor proved vulnerability in this CB during this round
- 2: No PoVs from competitors proved vulnerability in this CB during this round

Evaluation:
This quantity will vary between 1 and 2:
- 1: No PoVs emitted by this CRS proved vulnerabilities in this CB service on competitor systems
- 1+(x/(N-1)) PoVs emitted by this CRS for this CB service proved vulnerability in x competitor systems, where N is the number of CRSs participating in CFE.

A CRS's total score at the end of CFE shall be the sum of that CRS's CB round scores.

Q57: What is the API to the CGC environment?

A57: The following C language function prototypes are provided:

```
void _terminate(unsigned int status);
int allocate(size_t length, int prot, void **addr);
int deallocate(void *addr, size_t length);
int fdwait(int nfds, fd_set *readfds, fd_set *writefds,
        struct timeval *timeout, int *readyfds);
int random(void *buf, size_t count, size_t *rnd_bytes);
int receive(int fd, void *buf, size_t count, size_t
        *rx_bytes);
int transmit(int fd, const void *buf, size_t count, size_t
        *tx_bytes);
```

These function prototypes are notional and may be improved due to feedback prior to CGC kickoff.

Q56: Can foreign nationals participate in this challenge?

A56: This question is addressed in the CGC Rules Section 2 and Section 6. Foreign nationals may participate in Cyber Grand Challenge within a team which conforms to the CGC Rules.

Q55: DARPA-BAA-14-05 mentions DARPA-BAA-14-03, which describes the architecture framework. Where is DARPA-BAA-14-03?

A55: DARPA anticipates DARPA-BAA-14-03 to be published in the near future.

Q54: Does DARPA have a complete government team or are there opportunities for CGC support in development, judging, operating, etc.?

A54: DARPA anticipates a second BAA with other opportunities within this challenge.

Q53:   Can foreign teams apply for the funding also or can teams have foreign members?

A53: Review the eligibility section of DARPA-BAA-14-05 (3.1.4) and the Rules (2.1).

Q52: Is this 6.1 or 6.2 money?

A52: DARPA anticipates 6.2. funds for awards under DARPA-BAA-14-05 and DARPA-BAA-14-03.

Q51: Does fundamental versus non-fundamental affect desirability?

A51: See DARPA-BAA-14-05 section 2.2.

Q50: Are there any restrictions on foreign subcontractors? If so, what are the restrictions?

A50: See section 3.1.3 of DARPA-BAA-14-05.

Q49: Will the proposal evaluations favor small business, or is it a level playing field based on merit?

A49: See section 5 of DARPA-BAA-14-05. All proposals are evaluated on the same criteria.

Q48: Are the deliverables and payment percentages in DARPA-BAA-14-05 fixed, or can we propose alternatives?

A48: They are notional, not fixed. You can propose alternatives.

Q47: Can you clarify the length of the periods of performance for the base and option periods?

A47: Under DARPA-BAA-14-05, each period of performance is 12 months. The schedule in DARPA-BAA-14-05 is notional. Plan for all activities to take place within two 12 month phases.

Q46: Is it possible to combine with another group after the CQE?

A46: Yes.

Q45: Can an organization have two teams, one for Open track and one for Proposal track?

A45: This is excluded in the Rules. Teams are intended to be wholly separate.

Q44: If I submit a proposal to the Competition BAA (DARPA-BAA-14-05) and do not get selected, can I submit to the Architecture BAA (DARPA-BAA-14-03)?

A44: There's nothing to prevent you from submitting to both, but you cannot be selected for award under both. In the event that a proposer submits an otherwise selectable proposal to both DARPA-BAA-14-05 and DARPA-BAA-14-03, the decision as to which proposal to consider for award is at the discretion of the Government.

Q43: Must we deliver a working spreadsheet as part of the proposal for DARPA-BAA-14-05 or is that just DARPA's preference? You said it would be "helpful" versus "required?"

A43: Per section 4.2.1.2 of DARPA-BAA-14-05, the cost proposal should include a spreadsheet file (.xls or equivalent format) that provides formula traceability among all components of the cost proposal. The spreadsheet file must be included as a separate component of the full proposal package.

Q42: Can we talk to the Contracting Officer before a proposal is submitted?

A42: Reference Section 7 of DARPA-BAA-14-05, questions should be submitted to CGC-CompetitorBAA@darpa.mil.

Q41: Are there two BAA's anticipated for this program, the Architecture BAA (DARPA-BAA-14-03) and the Competition BAA (DARPA-BAA-14-05)?

A41: Yes.

Q40: What is the eligibility for using an OT for prototypes (845)?

A40: See DARPA's contract management website (http://www.darpa.mil/Opportunities/Contract_Management/Other_Transactions_and_Technology_Investment_Agreements.aspx) for information regarding OT for Prototype awards.

Q39: Is the electronic submittal system similar to T-FIMS?

A39: Yes.

Q38: Could the amounts of the project be larger if an entity supplied a cost share beyond the $750k?

A38: Yes.

Q37: With regard to Section 4.2.1.2.3 of DARPA-BAA-14-05, where are government rates and Defense Contract Audit Agency (DCAA) rates defined?

A37: FAR Part 42 discusses procedures for establishing forward pricing rates. Information is also available on the Defense Contract Management Agency's (DCMA) Website http://guidebook.dcma.mil/41/. You do not have to have DCMA approved rates to propose and receive an award under DARPA-BAA-14-05. Section 4.2.1.2.3 requires a proposer to justify its proposed direct labor rates and provides several examples of how that can be accomplished.

Q36: With regard to Section 4.2.1.1.1 of DARPA-BAA-14-05, where are the types of businesses described?

A36: Business sizes are defined by the Small Business Administration (http://www.sba.gov/content/table-small-business-size-standards). A definition of HBCU and Minority Institutions can be found in DFARS 252.226-7000 (http://www.acq.osd.mil/dpap/dars/dfars/html/current/252226.htm#252.226-7000).

Q35: Is there a limit to the number of teams awarded or total amount of grants?

A35: No grants will be awarded under DARPA-BAA-14-05, only Firm-Fixed-Price Procurement Contracts and Other Transactions. Under DARPA-BAA-14-05, DARPA anticipates multiple awards of $750,000 per phase of a two-phase effort; however, per the BAA, the number/amount of awards will depend on the quality of the proposals received and the availability of funds.

Q34: Will accepted proposals become public?

A34. DARPA will not publish awarded proposals under DARPA-BAA-14-05. Per section 4.2.2 of the BAA, DARPA treats proposals as source selection information (see FAR 2.101 and 3.104) and protects them as such, using secure handling and destruction procedures.

Q33: During CFE, how will a CRS monitor and modify traffic to a networked host?

A33:

<u>Monitor:</u>
During CFE, each competitor CRS will receive a read-only stream of all Competitor CRS network traffic directed toward its network host over the CFE network.

<u>Modify:</u>
Competitor systems will be provided with access to a DARPA-managed network appliance within the competition framework which will allow for traffic modification between the CFE network and the network host defended by the CRS. The managed appliance will use filters provided by the CRS through the Competition Framework API. Filters need not modify traffic; depending on their formulation they may modify traffic, alert, or take no action. Like Challenge Binaries, filters provided by a CRS will be distributed to all competitor systems for purposes of consensus evaluation (Shannon's Maxim). Alerts generated on the managed appliance by fielded filters will be communicated back to the CRS through the Competition Framework API. These CRS-provided filters will conform to a subset of the open source Snort rule syntax; this subset will be published by DARPA prior to kickoff of Cyber Grand Challenge.

Q32: How will CRS hardware interface with the CQE event?

A32: CQE is an Internet-accessible event that will not require a CRS to be physically present at an event location. Section 3.1 of the CGC Rules provides information on the CQE Corpus and workflow. Competitors should note the requirements of the post-CQE Site Visit from the CGC Rules, Section 3.1.3.2.

Q31: How will CRS hardware interface with the competition framework during CFE?

A31: A CRS used in CFE must fit entirely in a single, standard 19" 42U rack. CFE will take place at a physical location to be specified by DARPA. Competitor systems must be physically present at the CFE Compute Location in order to interface directly with the competition framework. Further details about available power, cooling, and infrastructure will be published as Cyber Grand Challenge progresses.

Q30: When will CQE and CFE scoring be finalized?

A30: DARPA is currently holding a period of public comment for the CFE and CQE scoring methods. Feedback should be addressed to cybergrandchallenge@darpa.mil. Scoring will be open for a period of public comment concluding on Friday, February 14, 2014. A final scoring algorithm for CQE and CFE will be released on Monday, March 10, 2014. Like the current proposed scoring algorithms, the final scoring methods will be consistent with the Cyber Grand Challenge Rules. See also Q/A 26 and 27.

Q29: What constitutes a Proof of Vulnerability in CFE?

A29: Proof of Vulnerability (PoV) may be demonstrated during CFE by interacting with the Competition Framework via two different methods.

Method 1: Registers

To prove that a remote CB is vulnerable, a Method 1 PoV shall redirect execution to an instruction pointer value negotiated with the Competition Framework API. Execution must fault at this value. At time of fault, one general purpose register must also be set to a second value negotiated with the Competition Framework API. This redirection will create an effect detectable by instrumentation operated on the competition framework.

Method 2: Memory

To prove that a remote CB is vulnerable, a Method 2 PoV will retrieve a value present in the memory of a competitor CB. This value must be retrieved from a memory range specified by the Competition Framework API, must be of a size specified by the Competition Framework API, and must be submitted through the Competition Framework API. This submittal will create an effect detectable by instrumentation operated on the competition framework.

Q28: What constitutes a Proof of Vulnerability in CQE?

A28: A CQE Proof of Vulnerability will cause a CB to crash in a manner consistent with memory corruption flaws with security impact, e.g., a segmentation fault or illegal instruction fault.

Q27: What is the scoring method for CQE?

A27: ~~The below scoring method candidate will be open for public comment until Friday, February 14, 2014.~~

~~CQE scoring is the product of three assessed quantities: Availability score, Security Score, and Evaluation score. These scores map to the Areas of Excellence (AoE) located in the Cyber Grand Challenge Rules as follows:~~

~~Availability: Area of Excellence 4~~
~~Security: Area of Excellence 2~~
~~Evaluation: Area of Excellence 3~~

~~During CQE, the CBs distributed by DARPA shall be referred to as “reference CBs”, not to be confused with replacement CBs submitted by competitors. PoVs used by DARPA for scoring will be referred to as “reference PoVs”. PoVs submitted by competitor systems to DARPA during CQE will be referred to as “submitted PoVs”.~~

~~CQE Scores will be assessed per Challenge Binary (“CB score”). Each CB score will be calculated as follows:~~

~~Availability * Security * Evaluation~~

~~**Availability**:~~
~~This quantity shall vary as a multi-step function between 0 and 1, with 1 being a perfect score. Performance and retained functionality will be measured, with Availability being set to the minimum of these quantities. Competitors are advised that slowing down the function of a replacement CB will result in a faster-than-linear Availability score dropoff.~~

~~**Security**:~~
~~This quantity will be calculated as follows: 1+ (**Reference + Consensus**)/2~~
- ~~**Reference:** The number of reference PoVs which do not prove vulnerability in the replacement CB, divided by the number of reference PoVs~~
- ~~**Consensus:**~~

  ~~This quantity will be set to 0 or 1:~~
  - ~~0: Any submitted PoV proved vulnerability in the replacement CB~~
  - ~~1: No submitted PoV proved vulnerability in the replacement CB~~

~~**Evaluation**:~~
~~This quantity will be set to 1 or 2:~~
- ~~1: A PoV emitted by this CRS did not prove vulnerability in the reference CB~~
- ~~2: A PoV emitted by this CRS did prove vulnerability in the reference CB~~

~~A CRS’s total score at the end of CQE shall be the sum of that CRS’s CB scores.~~

Q26: What is the scoring method for CFE?

~~This scoring method candidate will be open for public comment until Friday, February 14, 2014.~~

~~A26: CFE scoring is the product of three assessed quantities: Availability score, Security Score, and Evaluation score. These scores map to the Areas of Excellence (AoE) located in the Cyber Grand Challenge Rules as follows:~~

~~Availability: Area of Excellence 4~~
~~Security: AoE 2,5~~
~~Evaluation: Area of Excellence 3~~

~~CFE scoring will take place over many rounds. Scores will be assessed per Challenge Binary per round ("CB round score"); it follows that the sum of the CB round scores for a single CRS will be the total score for that system. Each CB round score will be calculated as follows:~~

~~Availability * Security * Evaluation~~

~~Availability~~~~:~~
~~This quantity shall vary as a multi-step function between 0 and 1, with 1 being a perfect score. Performance and retained functionality will be measured, with Availability being set to the minimum of these quantities. Competitors are advised that slowing down the function of a replacement CB will result in a faster-than-linear Availability score dropoff.~~

~~Security~~~~:~~
~~This quantity will be set to 1 or 2:~~

- ~~1: At least one PoV from a competitor proved vulnerability in this CB during this round~~
- ~~2: No PoVs from competitors proved vulnerability in this CB during this round~~

~~Evaluation~~~~:~~
~~This quantity will vary between 1 and 2:~~

- ~~1: No PoVs emitted by this CRS proved vulnerabilities in this CB service on competitor systems~~
- ~~1+(x/(N-1)) PoVs emitted by this CRS for this CB service proved vulnerability in x competitor systems, where N is the number of CRSs participating in CFE.~~

~~A CRS's total score at the end of CFE shall be the sum of that CRS's CB round scores.~~

Q25: What will be publicly released Post-CQE?

A25: DARPA intends to release the following items post-CQE:

- Reference CBs (initial Corpus distributed for CQE)
- PoVs, including both reference PoVs and PoVs gathered during the CQE

- Replacement CBs from the CQE, including reference patched CBs
- PCAP of traffic used during CQE evaluation
- Reference service pollers for each CB
- Reference CB source code
- A detailed list of scores for each CB for each finalist
- Team rankings (including Open Track and Proposal Track)

DARPA may modify this list of intended deliverables at its sole discretion.

Q24: What information about challenge binaries will be provided ahead of time (e.g., sample input and response; interaction protocol, API for service, etc.)?

A24: DARPA will provide an interface document detailing the methods CBs will use to interface with their execution environment.

Q23: What will we know about challenge network configuration (e.g., address ranges) before the final event?

A23: The CFE network topology will be known prior to CFE. In addition, competitors will have the opportunity to test technology interoperability during CFE Trials.

Q22: Will the execution environment be provided to the teams?

A22: A sample environment will be provided prior to the program commencing (proposal track awards have been finalized and open track teams have been registered/accepted) in the form of a virtual machine.

Q21: Will sample inputs be provided with some of the challenge binaries in the CQE corpus?

A21: Yes.

Q20: Can secure replacement CBs be submitted by a CRS throughout CFE?

A20: Yes.

Q19: What is the impact of submitting a replacement CB?

A19: The submission of secure replacements may be rate limited by the Competition Framework API, and fielding a replacement CB may impact service availability.

Q18: Are there networking constraints on patching? Reaching out to remote servers? May CBs communicate with the CRS while executing on the network host?

A18: During CFE, Challenge Binaries will not have the ability to initiate connections.

Q17: During CFE, for network defense, will existing tools for scanning and defending (TCP/UDP/NMAP, wireshark, snort, etc.) work, or must we develop new tools? Do you expect the teams to develop program analysis tools themselves or use off-the-shelf ones?

A17: DARPA will not dictate what automated approaches are acceptable within a CRS.

Q16: During CFE, what information (data sources) will our CRS have access to? Specifically will our CRS have access to crash logs, core dumps, and full network traffic feed?

A16: During CFE, a CRS will have access to a read only network tap. During CFE, a CRS will have the ability to request some CB status information through the Competition Framework API. Data sources automatically generated by a CRS internally will not be dictated by DARPA.

Q15: During CFE, how many networked hosts will competitors be responsible for monitoring/protecting?

A15: One.

Q14:   During CFE, will competitors have access to the network host?

A14: A CRS will have the ability to query the Competition Framework API for some CB status information. A CRS will have the ability to field replacement CBs through the Competition Framework API.

Q13: During CFE, will you be issuing new binaries to teams after competition start, or will you give all binaries to teams before start?

A13: During CFE, a CRS will be notified that a CB is available through the Competition Framework API.

Q12: What programming languages will CBs be written in?

A12: The C family of languages.

Q11: Does the U.S. Government assert any intellectual property rights to CRS source code developed by open track competitors?

A11: No.

Q10: What type of security vulnerabilities will CGC address?

A10: CGC Challenge Binaries shall contain traditional memory corruption flaws. A subset of relevant flaw types drawn from the MITRE Common Weakness Enumeration entries as found on http://cwe.mitre.org/ follows; teams are encouraged to make use of this list as a starting point, not a reference.

CWE-120: Buffer Copy without Checking Size of Input ('Classic Buffer Overflow')
CWE-121: Stack-based Buffer Overflow
CWE-122: Heap-based Buffer Overflow
CWE-123: Write-what-where Condition
CWE-124: Buffer Underwrite ('Buffer Underflow')
CWE-128: Wrap-around Error
CWE-129: Improper Validation of Array Index
CWE-130: Improper Handling of Length Parameter Inconsistency
CWE-131: Incorrect Calculation of Buffer Size
CWE-134: Uncontrolled Format String
CWE-135: Incorrect Calculation of Multi-Byte String Length
CWE-147: Improper Neutralization of Input Terminators
CWE-158: Improper Neutralization of Null Byte or NUL Character
CWE-170: Improper Null Termination
CWE-190: Integer Overflow or Wraparound
CWE-191: Integer Underflow (Wrap or Wraparound)
CWE-193: Off-by-one Error
CWE-194: Unexpected Sign Extension
CWE-195: Signed to Unsigned Conversion Error
CWE-196: Unsigned to Signed Conversion Error
CWE-401: Improper Release of Memory Before Removing Last Reference
CWE-409: Improper Handling of Highly Compressed Data (Data Amplification)
CWE-415: Double Free
CWE-416: Use After Free
CWE-457: Use of Uninitialized Variable
CWE-466: Return of pointer value outside of expected range
CWE-467: Use of sizeof() on a Pointer Type
CWE-468: Incorrect Pointer Scaling
CWE-469: Use of Pointer Subtraction to Determine Size
CWE-763: Release of Invalid Pointer or Reference
CWE-786: Access of Memory Location Before Start of Buffer
CWE-787: Out-of-bounds Write
CWE-788: Access of Memory Location After End of Buffer
CWE-805: Buffer Access with Incorrect Length Value

CWE-806: Buffer Access Using Size of Source Buffer
CWE-822: Untrusted Pointer Dereference
CWE-823: Use of Out-of-range Pointer Offset
CWE-824: Access of Uninitialized Pointer
CWE-825: Expired Pointer Dereference

Q9: What constitutes a software flaw in Cyber Grand Challenge?

A9: DARPA CGC will not provide a formal definition of a software flaw; this question lies outside the scope of the challenge. The CGC will operate in the tradition of existing cyber competitions: a flaw is proven when an input delivered from the network to a flawed software program (CB) creates an effect detectable by instrumentation operated by the competition framework. CGC Challenge Binaries will contain memory corruption flaws representative of flaws categorized by the MITRE CWE[1], however, Competitor Systems may prove any software flaw they discover through automated reasoning. A list of representative CWE categories will be released prior to the kickoff of Cyber Grand Challenge.

Q8: What platform will CGC run on?

A8: CGC Challenge Binaries (CBs) will be incompatible with any known OS architecture. CBs will run in an environment custom built for the competition. Knowledge of the operating system will not be in scope for the competition; rather, CGC requires a competition system to reason about the function of compiled binaries receiving inputs from the network. CBs will not conform to any currently known application layer protocols. CB protocol knowledge must be generated automatically by competition systems during CGC events through a process of automated reasoning about software. These constraints will ensure that all knowledge in use by competition systems during CGC events is generated via automatic processes.

Q7: What CPU architecture will CGC run on?

A7: For the purpose of maximizing accessibility and participation: Intel x86, 32-bit.

Q6: What compiler will be used to build the binaries?

A6: CGC will distribute a reference compiler toolchain prior to challenge kickoff. However, challenge binaries may be produced by any compiler including the reference compiler.

[1] http://cwe.mitre.org/

Q5: During the final event, what happens when my Competition System fields a new Challenge Binary?

A5: During CFE, in order to enact defenses, a CRS may choose to replace a CB with a newly secured version. To field a replacement CB, a CRS must submit the replacement through an automated API operated by the competition framework. The competition framework will deploy the replacement binary on behalf of the CRS to its networked host. Additionally, the competition framework will make a copy of the replacement CB available to all competitor systems for the purposes of consensus evaluation (Shannon's Maxim). Once deployed, replacement CBs will be required to function as self-contained replacements without custom dependencies, libraries, etc.

Q4: I'm interested in advanced application defenses. Will these be part of CGC?

A4: During CFE, systems will have the ability to deploy network defenses as well as application defenses. To deploy application defenses, competition systems may analyze CBs and field secure replacements. Due to the competitive nature of CGC, DARPA expects that competitors will field many approaches of varying type, advancement, and efficacy.

Q3: What limitations are imposed on replacement CBs during CFE?

A3: During CFE, the competition framework will monitor the availability and correct function of each CB. If a CRS deploys replacement CBs that degrade CB function by impacting performance, correctness of CB responses, or the ability to service network requests, a negative impact on scoring is expected. Similar constraints will be imposed on replacement CBs during CQE scoring.

Q2: In the CGC Rules, Area of Excellence 2 specifies Autonomous Patching. Does this mean a Cyber Reasoning System (CRS) is required to isolate and remove flaws, or may a CRS field any secure replacement Challenge Binary (CB)?

A2: During the CGC Qualification Event (CQE) and Final Event (CFE), CBs will be evaluated based on availability, correct function, and the mitigation of flaws, as described in the CGC Rules and this FAQ. No specific requirements are imposed on the formulation method for secure replacement CBs.

Q1: Are you planning an Industry Day for competitors?

A1: Two Competitor Day sessions are planned, one on the East Coast, and one on the West Coast.

> - The East Coast Competitor Days are currently scheduled for December 3 and 4, 2013 at the DARPA Conference Center, 675 North Randolph Street,

Arlington, VA 22203. Note: the second day will be a repeat of the first day to accommodate registered attendees. Availability is on a first-come-first-served basis. All registrations will be for the December 3 session until capacity is reached; at that point, registrations will be for the December 4 session. Please visit http://www.sa-meetings.com/darpacgccompetitorday for more information and to register.

- The West Coast Competitor Day is currently scheduled for December 9, 2013 at the Westin St. Francis, 335 Powell St, San Francisco, CA. Availability is on a first-come-first-served basis. Please visit http://www.sa-meetings.com/darpacgccompetitordaywest for more information and to register.