DARPA Cyber Grand Challenge: CB Author Ranking

Defense Advanced Research Projects Agency
Information Innovation Office

September 10, 2014

Introduction

Several independent teams will be creating Challenge Binaries (CBs) for use in DARPA's Cyber Grand Challenge. To incentivize production of a broad range of CBs that will enable a thorough understanding and comparison of Cyber Reasoning Systems (CRSs) participating in CGC, an algorithm has been developed to score and rank CB author performance, as detailed in this document.

After completion of Cyber Grand Challenge Qualifier Event (CQE), each Challenge Set (CS) will be scored on its ability to differentiate CRSs participating in CQE uniquely. Each CS will be scored individually; an authoring team's score will be the sum of scores for CSs created by this team.

CS Score

Each CS will receive a score based on its ability to differentiate performance of CRSs; while there are many ways to differentiate performance, the CS score focuses on ability to differentiate CRSs on their ability to generate Proofs of Vulnerability (PoVs) and provide Replacement CBs that are invulnerable to Reference PoVs. This is reflected in CRS's **Evaluation Score** and **Security Score**, respectively, as computed during CQE scoring and defined in [2].

To compute a CS score, we first collect data on pairwise differentiation between CRS that this CS provides. Then, the values are adjusted based on "uniqueness" of the differentiation, determined by how many other CSs also gave different scores to this pair of CRSs.

The final CS score is the sum of component scores, as detailed below.

Raw Differentiation Matrix

A CRS x CRS Raw Differentiation Matrix (RDM), representing this CS's ability to differentiate pairs of competitors, will be computed for each CS. The values in the RDM will be filled in as follows (N_{CRS} is the number of CRS participating in CQE):

1. Initialization:

$$\forall i, j \in \{0, N_{CRS}\}\ RDM(CS)_{i,j} = 0$$

2. Evaluation Score (upper triangle):

$$\forall i \in \{0, N_{CRS}\}\ \forall j \in \{i, N_{CRS}\}$$
$$RDM(CS)_{i,j} = (EvaluationScore(CS, CRS_i) \neq EvaluationScore(CS, CRS_j))$$

3. Security Score (lower triangle):

$$\forall i \in \{0, N_{CRS}\}\ \forall j \in \{0, i\}$$
$$RDM(CS)_{i,j} = (SecurityScore(CS, CRS_i) \neq SecurityScore(CS, CRS_j))$$

For example, consider a CS (CS_1) that induced the following scores in the 3 participating CRSs:

CRS	Evaluation Score	Security Score
CRS_1	1	1.5
CRS_2	1	0.0
CRS_3	2	1.5

Then the RDM for CS_1 will look as follows (note that the diagonal is always 0):

$$RDM(CS_1)_{i,j} = \begin{array}{c} \\ CRS_1 \\ CRS_2 \\ CRS_3 \end{array} \begin{array}{c} \begin{array}{ccc} CRS_1 & CRS_2 & CRS_3 \end{array} \\ \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \end{array}$$

Unique Differentiation Matrix

To compute unique differentiation, a matrix showing total differentiation counts over all CSs will be computed by adding together the individual CS RDMs.

$$TDM_{i,j} = \sum_{CS} RDM(CS)_{i,j}$$

A Uniqueness-adjusted Differentiation Matrix (UDM) will be computed for each CS by dividing the entries in the RDM by the corresponding Total Differentiation Matrix (TDM) entries (ignoring zeros).

$$UDM(CS)i, j = \begin{cases} \frac{RDM(CS)_{i,j}}{TDM_{i,j}} & \text{if } TDM_{i,j} \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

For example, if the TDM is:

$$TDM_{i,j} = \begin{array}{c} \\ CRS_1 \\ CRS_2 \\ CRS_3 \end{array} \begin{array}{c} \begin{array}{ccc} CRS_1 & CRS_2 & CRS_3 \end{array} \\ \begin{pmatrix} 0 & 2 & 10 \\ 5 & 0 & 2 \\ 9 & 1 & 0 \end{pmatrix} \end{array}$$

then the UDM for CS_1 in the example above will be:

$$UDM_{i,j} = \begin{array}{c} \\ CRS_1 \\ CRS_2 \\ CRS_3 \end{array} \begin{array}{c} \begin{array}{ccc} CRS_1 & CRS_2 & CRS_3 \end{array} \\ \begin{pmatrix} 0 & 0 & 0.1 \\ 0.2 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix} \end{array}$$

Final CS Differentiation Score

The final CS *unique differentiation score* will be computed as the sum of the elements of the corresponding UDM.

$$CSScore(CS) = \sum_{i,j} UDM(CS)_{i,j}$$

In the running example, CS_1 will receive a score of

$$CSScore(CS_1) = \sum_{i,j} UDM(CS_1)_{i,j} = 0.1 + 0.2 + 0.5 + 1 = 1.8$$

To mitigate an unlikely event of a tie, DARPA will release a tie-breaker algorithm before CQE.

References

[1] Cyber Grand Challenge Rules, `https://cgc.darpa.mil/documents.aspx`

[2] Cyber Grand Challenge: CQE Scoring Document, `https://cgc.darpa.mil/documents.aspx`

Glossary

AoE Area of Excellence.

CB Challenge Binary.

CGC Cyber Grand Challenge.

Challenge Binary A vulnerable network service that accepts remote network connections, composed of one or more communicating binaries.

Challenge Set A Challenge Binary with associated Reference PoVs, Service Polls, and Reference Patched CB.

CQE Cyber Grand Challenge Qualifier Event.

CRS Cyber Reasoning System.

CS Challenge Set.

Cyber Reasoning System Unmanned systems that autonomously reason about novel program flaws, prove the existence of flaws in networked applications, and formulate effective defenses.

PoV Proof of Vulnerability.

Proof of Vulnerability An input that activates and proves the existence of a hidden flaw in a CB.

RDM Raw Differentiation Matrix.

Reference Patched CB DARPA's solution for a CB.

Reference PoV PoV supplied by CB author.

Replacement CB Solution for a CB supplied by a team's CRS.

Submitted PoV PoV supplied by a team's CRS.

TDM Total Differentiation Matrix.

UDM Uniqueness-adjusted Differentiation Matrix.