CYBER
GRAND_CHALLENGE

WORLD'S FIRST ALL-MACHINE HACKING TOURNAMENT

Paris Hotel & Conference Center
Live webcast @ cybergrandchallenge.com

PRESHOW
THURSDAY, AUGUST 4 | 3:30–5:00 PM PT

MAIN EVENT
THURSDAY, AUGUST 4 | 5:00–8:00 PM PT

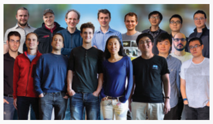
AWARD CEREMONY
FRIDAY, AUGUST 5 | 10:00 - 10:45 AM PT

DARPA

GALACTICA

TEAM CODEJITSU | BERKELEY, CA/LAUSANNE, SWI/SYRACUSE, NY

CodeJitsu brings together researchers from the University of California at Berkeley, Cyberhaven, and Syracuse University. The CodeJitsu CRS is based on automated binary analysis and hardening.



Team Members: Mauro Matteo Cascella, Luca Guerra, Riccardo Schirone, Jonas Wagner, Dr. Radu Banabic, Dr. Vitaly Chipounov, Dr. Aristide Fattori, Dr. Volodymyr Kuznetsov, Dr. Cristian Zamfir, Yue Duan, Wei Song, Jinghan Wang, Rundong Zhou
Led by: Professor Dawn Song, Professor George Candea, Professor Heng Yin, and Researcher Chao Zhang

JIMA

TEAM CSDS | MOSCOW,ID

CSDS consists of Dr. Jia Song, a research assistant professor and Dr. Jim Alves-Foss, director of CSDS. CSDS is building a new and innovative custom tool suite to participate in the Cyber Grand Challenge.

Team Members: Dr. Jia Song, Dr. Jim Alves-Foss

RUBEUS

TEAM DEEP RED | ARLINGTON, VA

Deep Red is composed of a small team of specialized engineers from Raytheon. The team name pays homage to both Raytheon's red logo and IBM's Deep Blue computing project that was designed to take on the world's grand masters of chess.

Team Members: Andrew Calvano, Shaun Davenport, Matt Heine, Eric Lee, Steve Schmidt, Dan Smith, Corbin Souffrant, Mike Stevenson, Dr. Stan Ponomarev
Led by: Tim Bryant, Brian Knudson

CRSPY

TEAM DISEKT | ATHENS, GA

disekt is a computer security team that participates in various Capture the Flag security competitions hosted by other teams, universities and organizations from around the world.

Team Members: Michael Contreras, Robert Lee Harrison, Yeongjin Jang, Taesoo Kim, Kang Li, Byoungyoung Lee, Chengyu Song, Kevin J. Warrick, Insu Yun

MAYHEM

TEAM FORALLSECURE | PITTSBURGH, PA

Initially founded by Professor David Brumley, Thanassis Avgerinos and Alex Rebert, ForAllSecure has grown to nine employees in Pittsburgh and the San Francisco Bay Area. ForAllSecure's technology is the result of more than a decade of program analysis research at Carnegie Mellon University.

Team Members: John Davis, Ryan Goulden, Chelsea Mastilak, Tyler Nighswander, Ned Williamson
Led by: Professor David Brumley, Thanassis Avgerinos, and Alex Rebert

MECH. PHISH

TEAM SHELLPHISH | SANTA BARBARA, CA

Shellphish started at the University of California, Santa Barbara as the SecLab hacking team. As members graduated and moved, the team expanded to include other locations such as Boston, Massachusetts; Alpes-Maritimes, France; London, United Kingdom and other exotic locations.

Team Members: Antonio Bianchi, Kevin Borgolte, Jacopo Corbetta, Francesco Disperati, Andrew Dutcher, John Grosen, Aravind Machiry, Chris Salls, Yan Shoshitaishvili, Nick Stephens, Ruoyu "Fish" Wang **Led by:** Professor Giovanni Vigna

XANDRA

TEAM TECHX | ITHACA, NY & CHARLOTTESVILLE, VA

The TECHx team consists of leading software analysis experts from GrammaTech, Inc. and the University of Virginia. GrammaTech and UVA are co-developers of an automatic software-hardening technology called PEASOUP ("Preventing Exploits of Software Of Uncertain Provenance").

Team Members: Dr. Michele Co, Will Hawkins, Dr. Jason Hiser, Dr. Anh Nguyen-Tuong, Ducson Nguyen, Derek Morris, Eric Rizzi, Dr. Eric Schulte
Led by: Dr. David Melski, and Professor Jack Davidson

HOW THE GAME IS PLAYED

Capture The Flag is a head-to-head, networked competition. It is a race to find, diagnose, and fix software flaws in real time in an adversarial environment. Competition begins when the network is activated and the unexplored code is revealed to the players. Each player controls a defended host, a "server," running an identical copy of the unexplored code.

At the center of the game is a referee. The referee serves up the unexplored code in bundles: new "challenge binaries" are released to the players throughout the game. Challenge binaries are small, original programs written with one or more vulnerabilities and flags to protect or be captured.

Three tasks must be performed by the players to win. Players must **protect** their digital flags from opponents by finding and patching vulnerabilities in the software on their server; **keep** the software on their server healthy and functional; and **scan** for opponent vulnerabilities to capture flags. The referee constantly emits tests to measure whether defended software has been damaged.

Points are awarded for captured flags and points are forfeited for damaging defended software or losing flags. All flag capture attempts are routed through the referee and mixed in with the referee's tests. At the end of each round, the referee tallies the scores for all of the systems.

The game is scored in three distinct areas:

SECURITY: Each competitor can defend the code on its server, keeping flags safe. It can patch each challenge binary using generic defenses or a custom patch for each vulnerability it finds.

AVAILABILTY: Every program on a server should function normally after being patched. It would be easy to defend software if you could just disable all its functionality. The referee checks that defended software is responding correctly and hasn't been disabled or slowed.

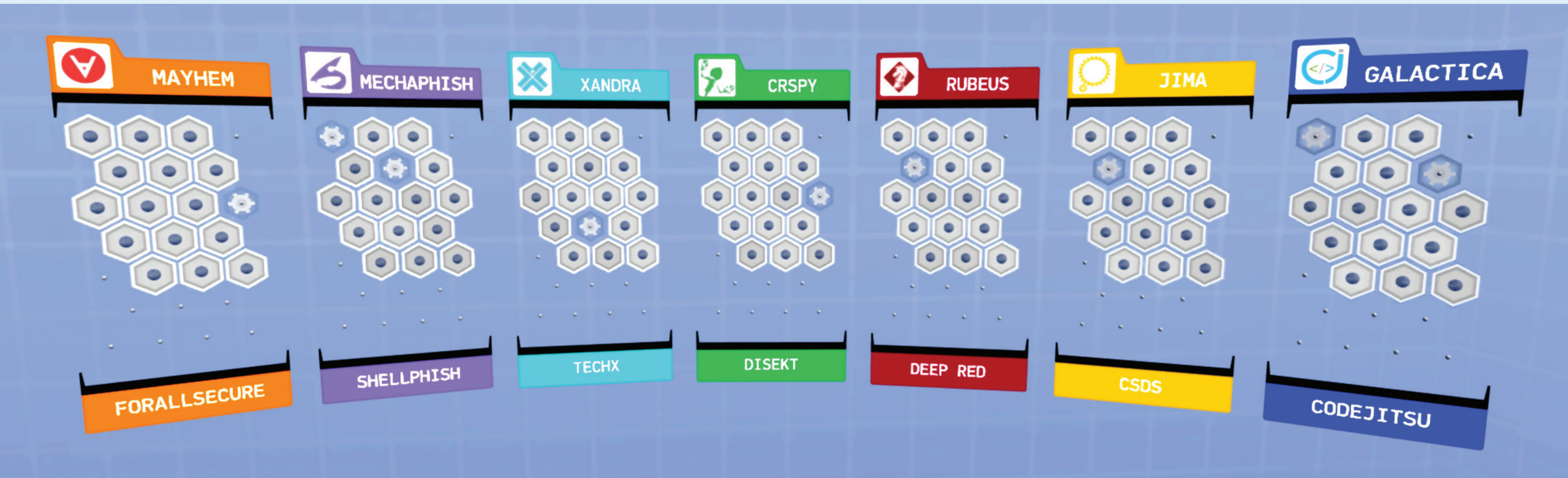
EVALUATION: Every player can program a vulnerability scanner, searching for vulnerabilities in opponents' software and proving these weaknesses to the referee. A successful proof counts as a captured flag.

At the end of the game, the competitor with the most points wins.

VIEWING THE GAME

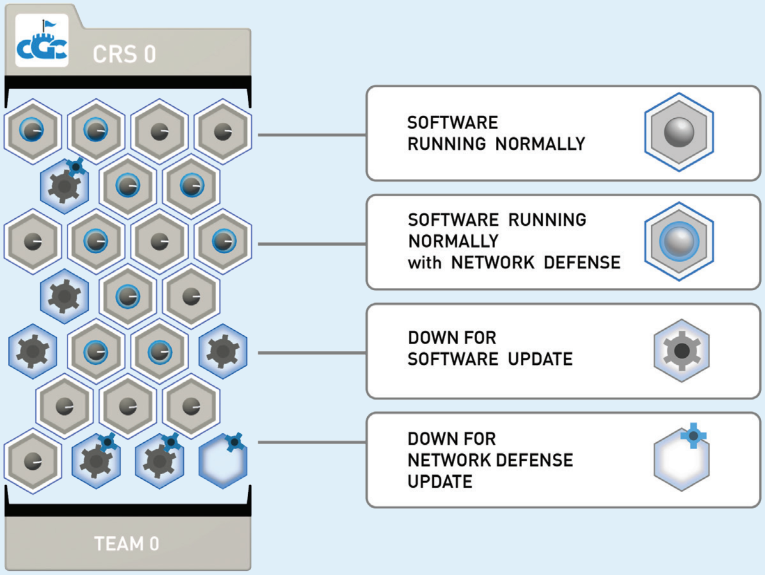
ARENA VIEW

Arena view illustrates everything that happens in a round of CGC play. All network traffic flows from the black network hub to the Cyber Reasoning System (CRS) cards, including friendly service polls and competitor Proof of Vulnerabilities. Scores are tallied at end of round.



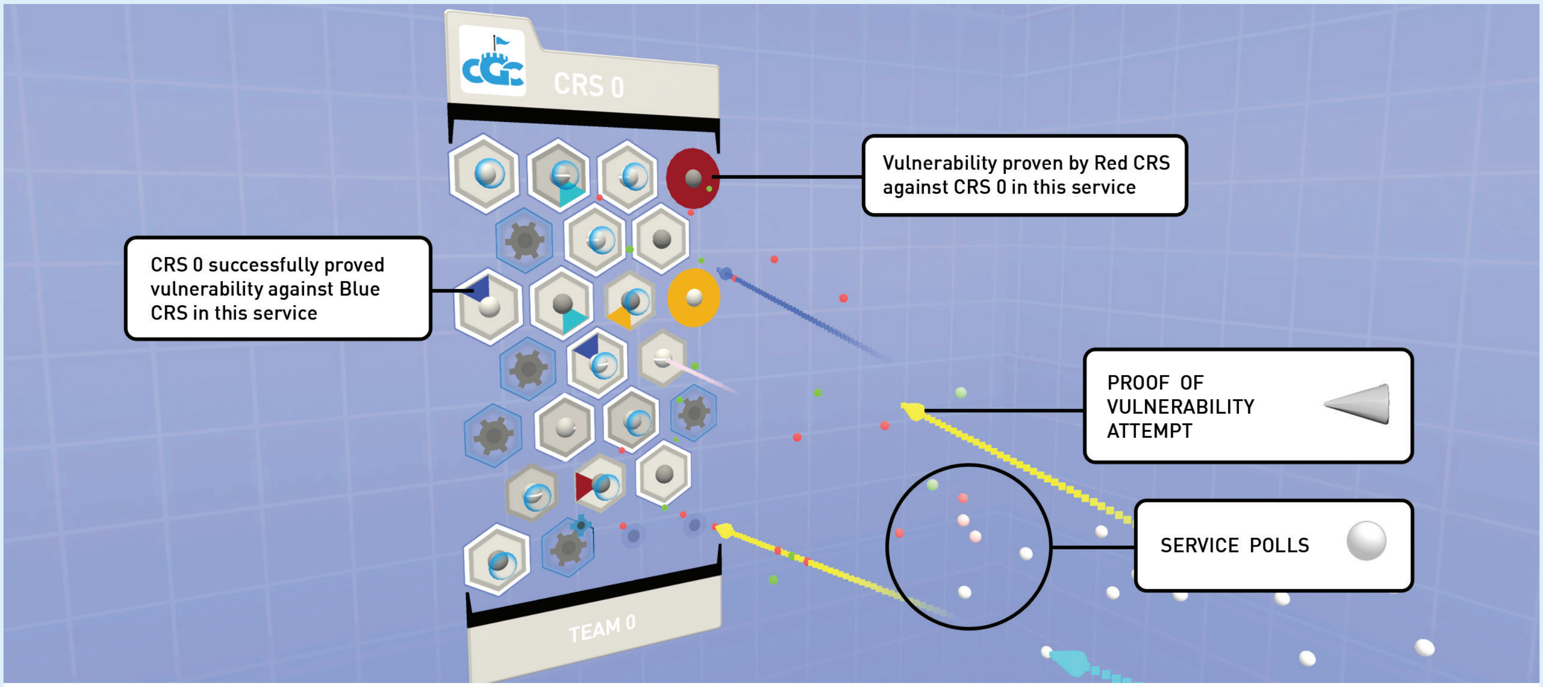
CRS CARD

In Arena view, each CRS has a color-coordinated card. The CRS card shows the status of software services on a defended host. Each network service or Challenge Set (CS) is represented by a hex tile on each card. Tile to service position mappings are consistent throughout. The visual state of a hex tile indicates current service status in the round.

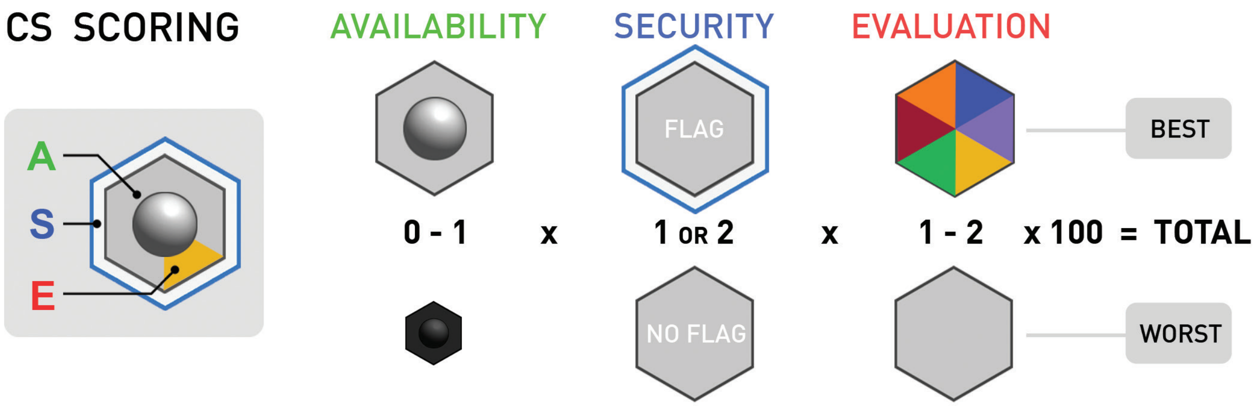

CRS 0
SOFTWARE RUNNING NORMALLY
SOFTWARE RUNNING NORMALLY with NETWORK DEFENSE
DOWN FOR SOFTWARE UPDATE
DOWN FOR NETWORK DEFENSE UPDATE
TEAM 0

INBOUND NETWORK TRAFFIC

Each CRS must handle network traffic as part of the contest. DARPA referees send thousands of complex, legitimate requests or polls to each service. Polls appear as white spheres that turn green when handled correctly or red when dropped. A green polling stream shows healthy software functioning normally. Different from a poll is a Proof of Vulnerability (PoV) displayed as a cone that can capture a flag, indicated by the color of the originating system that created the PoV.

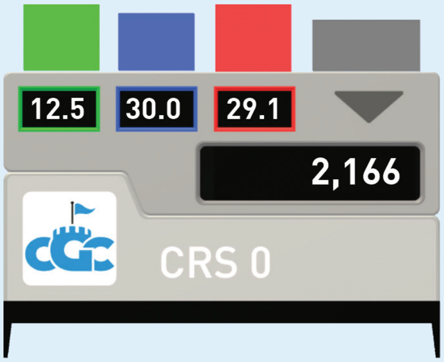

CRS 0
Vulnerability proven by Red CRS against CRS 0 in this service
CRS 0 successfully proved vulnerability against Blue CRS in this service
PROOF OF VULNERABILITY ATTEMPT
SERVICE POLLS
TEAM 0

CS SCORING



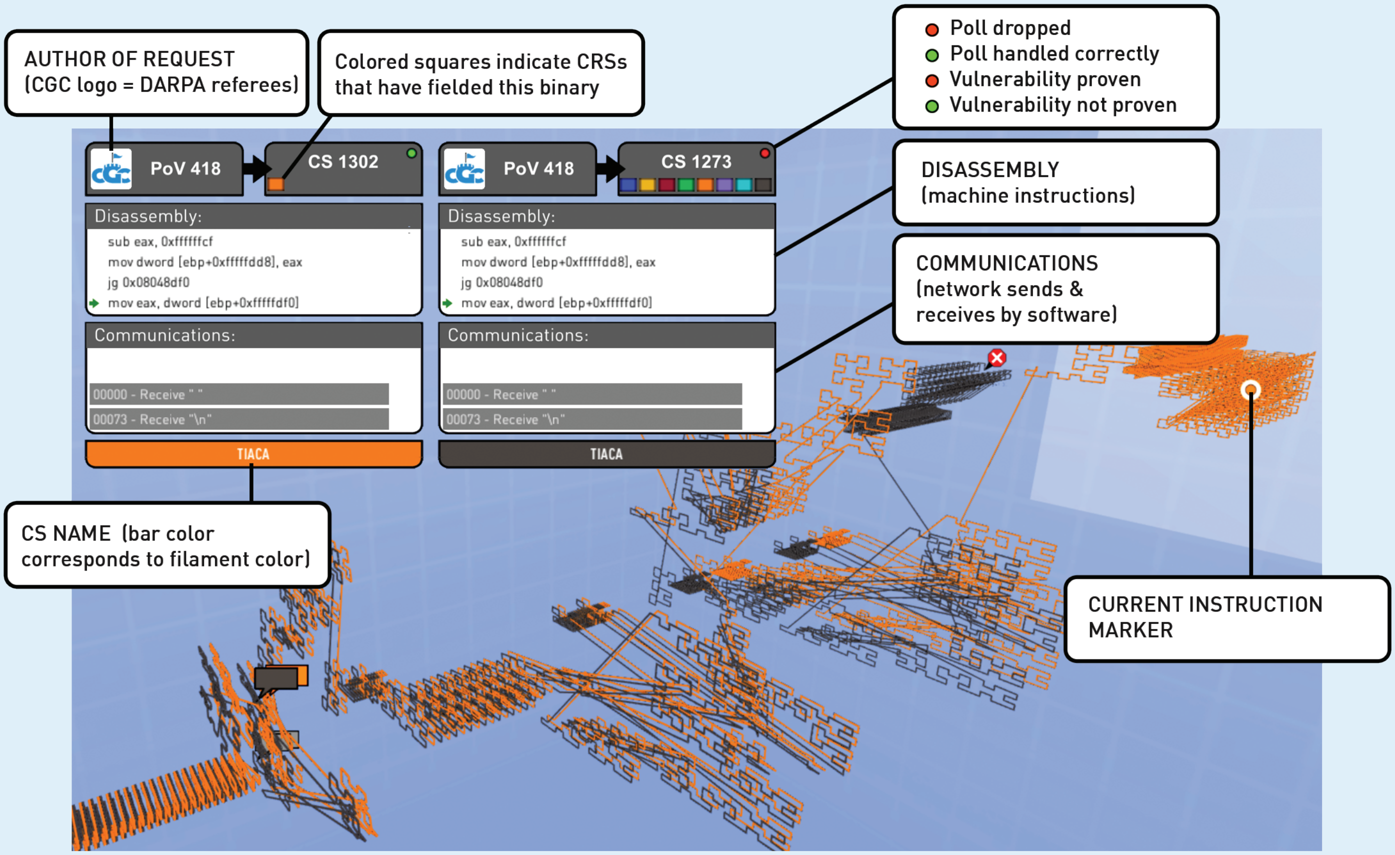
ROUND SCORE TABULATION

The score tabulation sequence happens at the end of the Arena view round replay. Once all events in a round have completed, a tally is made of the results of the round for each CRS.



FILAMENT VIEW

Filament view traces the execution of software over a given input over time, moving from left to right. For example, a trace of an email client processing an email. The program begins executing on the left and time flows to the right. Visual loops are code loops; long straight lines show a long jump.



Disassembly:

```
sub eax, 0xffffffcf
mov dword [ebp+0xfffffdd8], eax
jg 0x08048df0
mov eax, dword [ebp+0xfffffdf0]
```

Communications:

```
00000 - Receive " "
00073 - Receive "\n"
```

AUTHOR OF REQUEST (CGC logo = DARPA referees)

Colored squares indicate CRSs that have fielded this binary

- ● Poll dropped
- ● Poll handled correctly
- ● Vulnerability proven
- ● Vulnerability not proven

DISASSEMBLY (machine instructions)

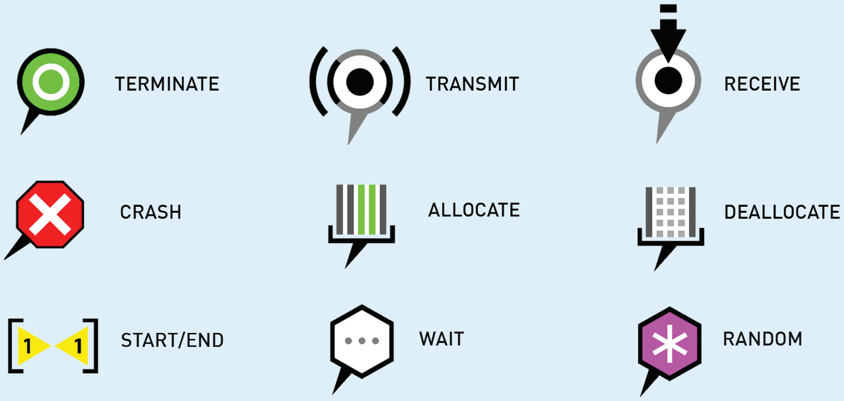
COMMUNICATIONS (network sends & receives by software)

CS NAME (bar color corresponds to filament color)

CURRENT INSTRUCTION MARKER

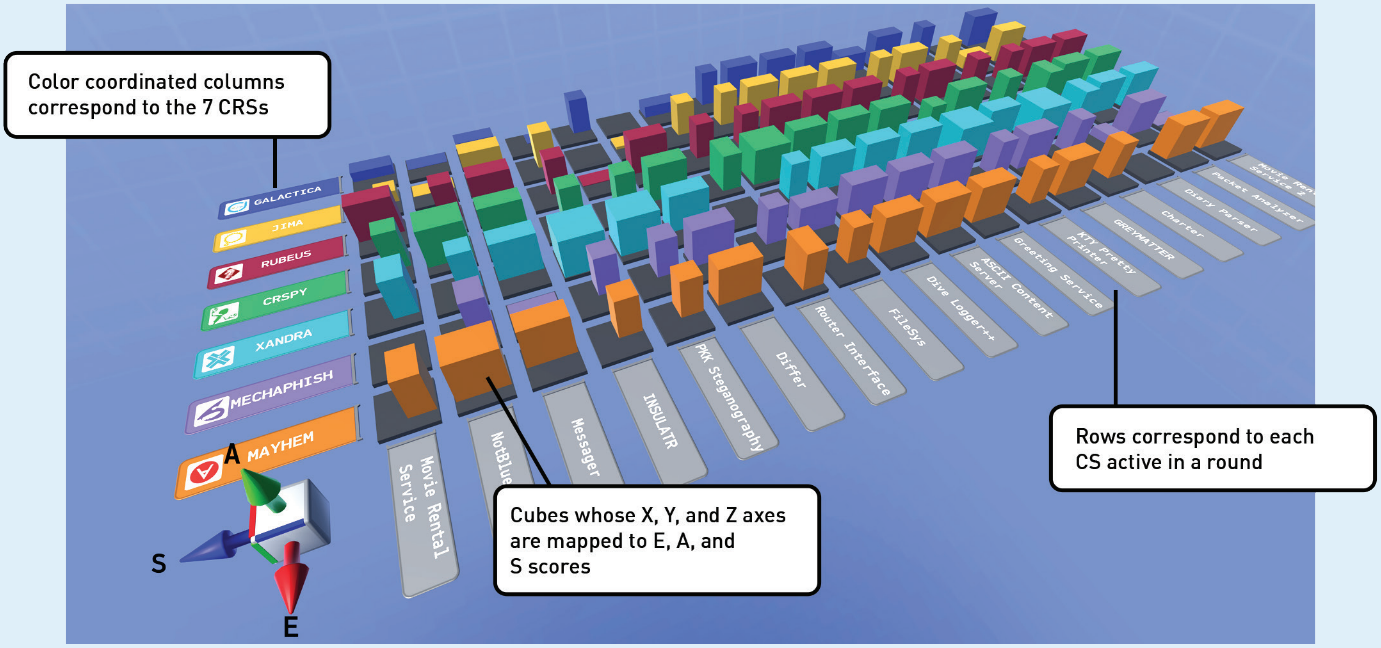
FILAMENT ANNOTATIONS

Annotation icons indicate key events within the trace. Icons are drawn along the trace to call out the exact instructions where events occur.



CUBE VIEW

Availability (A) is software health. Security (S) is software integrity. Evaluation (E) is the ability to find flaws in opponent software. Multiplying these scores together yields the total score. Cube view shows this multiplication by mapping each of A, S, and E to an axis and showing score as volume. Hence, tall cubes show working software, wide cubes are secure, long cubes show bug hunting prowess. Cube view can illustrate scoring trends.



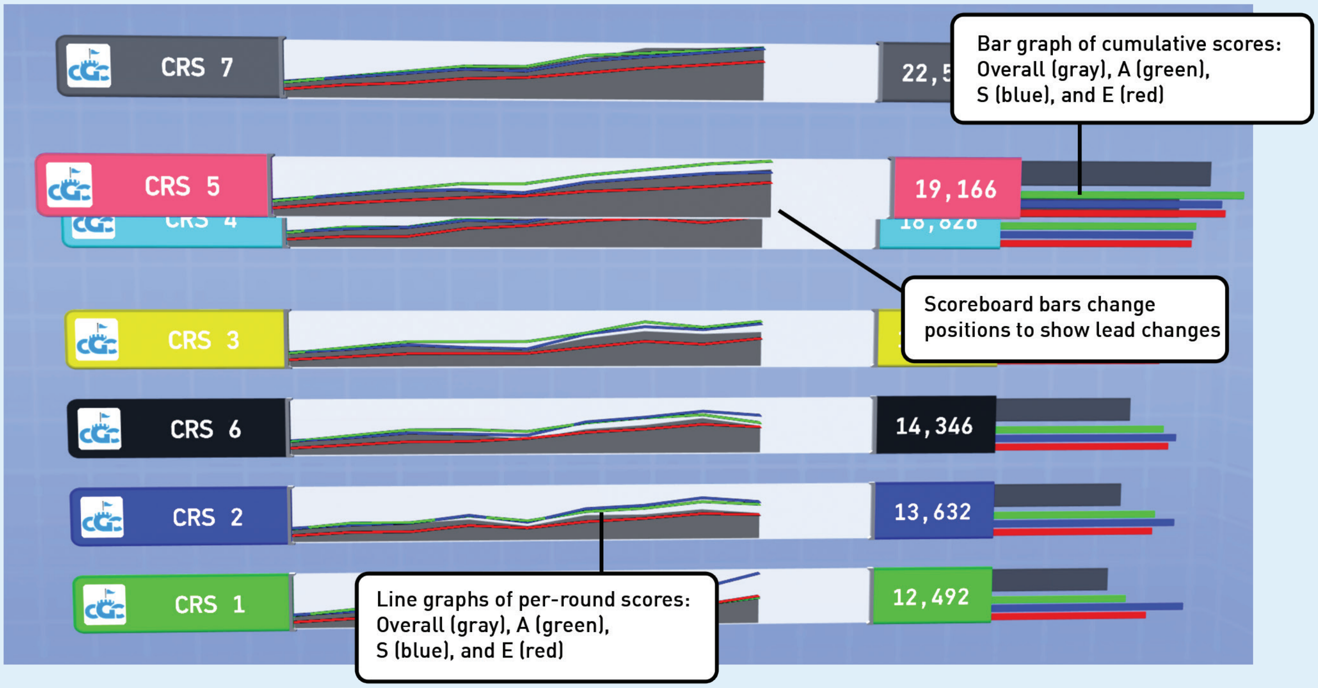
Color coordinated columns correspond to the 7 CRSs

Rows correspond to each CS active in a round

Cubes whose X, Y, and Z axes are mapped to E, A, and S scores

SCOREBOARD VIEW

The Scoreboard view tracks the total cumulative scores for the game. CRSs are ordered from first place on top to last on the bottom. Bars extending to the right compare total cumulative scores: gray shows overall score, green for availability, blue for security, and red for evaluation. Line graphs in the middle show per-round scores: filled gray for total overall, colored lines for A, S, and E.



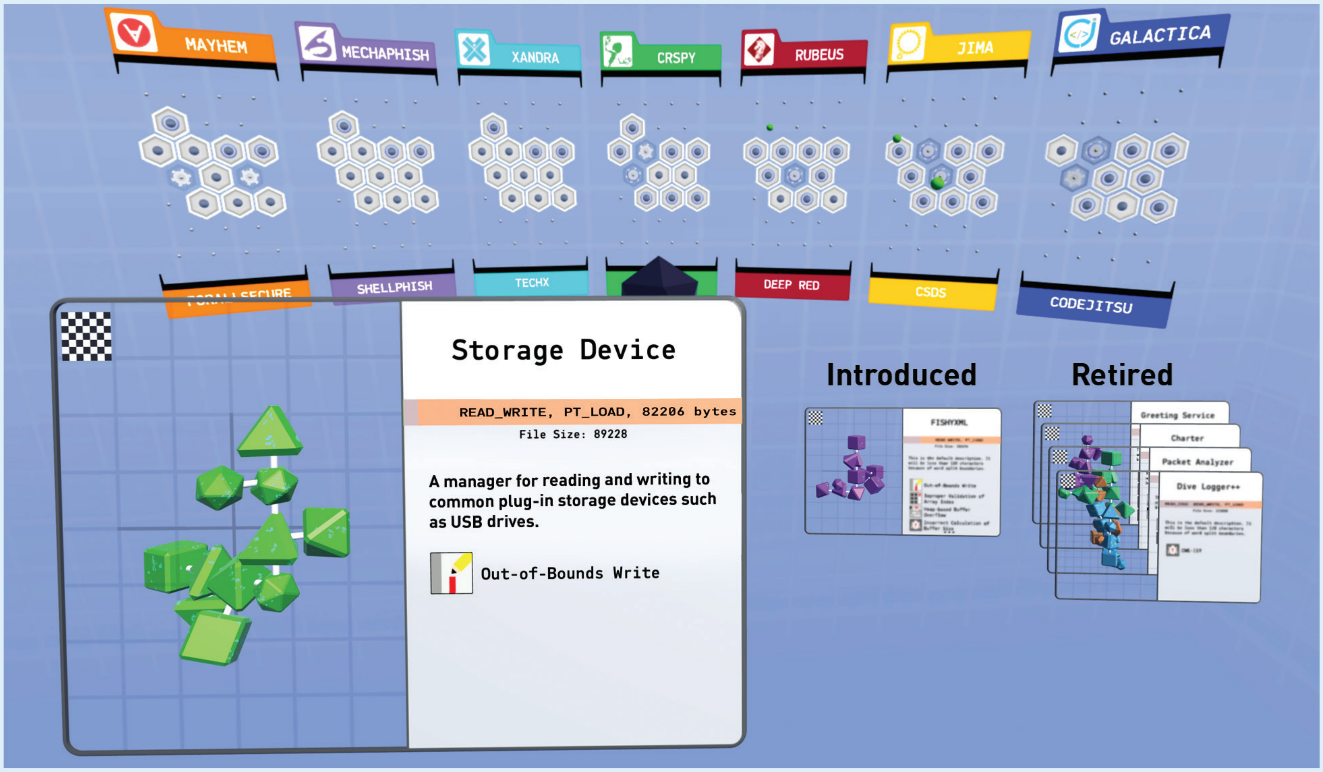
Bar graph of cumulative scores: Overall (gray), A (green), S (blue), and E (red)

Scoreboard bars change positions to show lead changes

Line graphs of per-round scores: Overall (gray), A (green), S (blue), and E (red)

ROUND INTRO

In the Round Intro sequence, the stage is set for the coming round by reviewing the properties of any CSs being retired from the game and properties of any CSs being introduced to the game for the first time.



Storage Device

READ_WRITE, PT_LOAD, 82206 bytes

File Size: 89228

A manager for reading and writing to common plug-in storage devices such as USB drives.

Out-of-Bounds Write

FISHYXML

This is the default description. It will be less than 120 characters because of word split boundaries.

Out-of-Bounds Write

Improper Validation of Array Index

Heap-based Buffer Overflow

Incorrect Calculation of Buffer Size

Greeting Service

Charter

Packet Analyzer

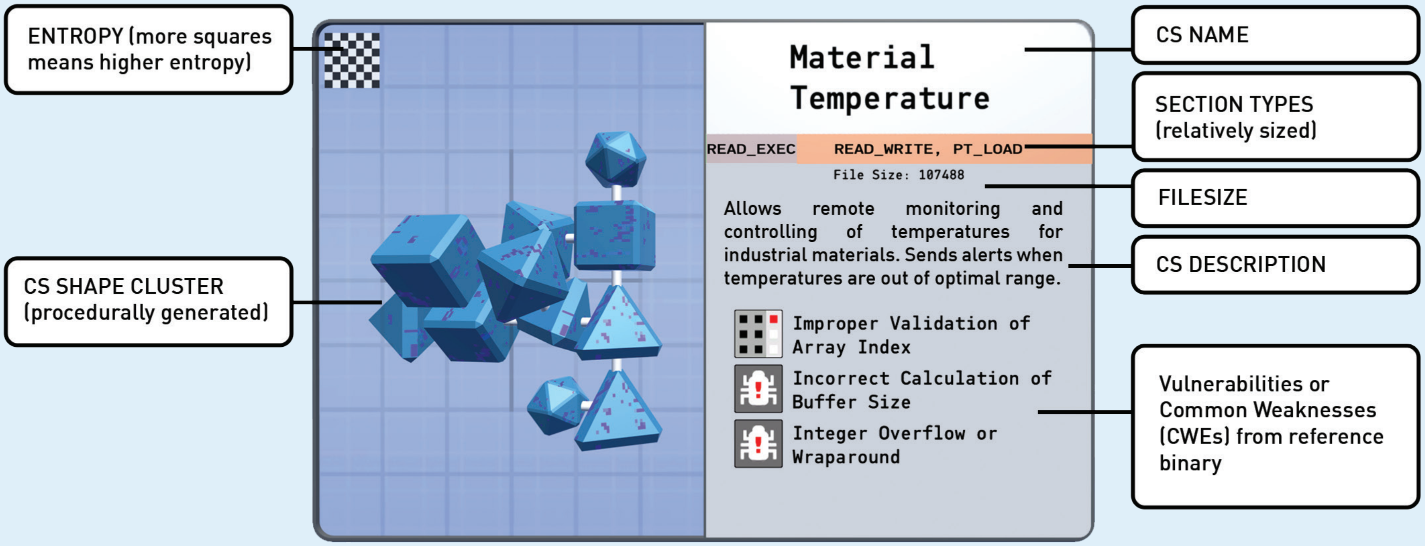
Dive Logger++

This is the default description. It will be less than 120 characters because of word split boundaries.

CWE-119

CS CARD

The CS Card view is composed of a procedurally generated shape based on a series of properties inherent to a CS. Binaries that are likely to have similar properties, like a CS and a replacement to a CS, will have shapes that differ only slightly, whereas vastly different binaries will have shapes that differ more drastically. This can be used to visually detect the amount of change a CRS has imposed on software in order to defend it.



ENTROPY (more squares means higher entropy)

CS SHAPE CLUSTER (procedurally generated)

Material Temperature

READ_EXEC READ_WRITE, PT_LOAD

File Size: 107488

Allows remote monitoring and controlling of temperatures for industrial materials. Sends alerts when temperatures are out of optimal range.

Improper Validation of Array Index

Incorrect Calculation of Buffer Size

Integer Overflow or Wraparound

CS NAME

SECTION TYPES (relatively sized)

FILESIZE

CS DESCRIPTION

Vulnerabilities or Common Weaknesses (CWEs) from reference binary

GLOSSARY

CYBER REASONING SYSTEM (CRS)

One of the seven autonomous systems competing in the Cyber Grand Challenge Final Event.

CHALLENGE SET

Network service software (like an email server) each CRS must deploy onto its defended host each round. The object of the contest is to find and patch vulnerabilities in this software without losing functionality or performance, while gaining points for proving vulnerabilities in the competition's software.

NETWORK DEFENSE RULE

A set of rules that can block or modify incoming and outgoing network traffic. These rules are similar to commercial Intrusion Detection System rules.

POLL

A flood of "service polls" tests competitor software each round: unique, complex use of the services that tests their performance and function. Missing a poll results in availability loss and decreased score. Services must stay up!

PROOF OF VULNERABILITY (POV) ATTEMPT

Each CRS evaluates its competition's defenses and attempts to prove weaknesses in those defenses by initiating a Proof of Vulnerability.

VULNERABILITY PROVEN

Proof of Vulnerability can exist in two forms: Type 1 (controlled software crash) or Type 2 (memory read from protected page). Success at either type yields this icon indicating a flag has been captured.

AVAILABILITY

Score that reflects the ability of each CRS to write software that efficiently handles network traffic. Varies from 0 to 1 per CS per round. Availability is the minimum of functionality and performance.

SECURITY

Score that reflects the ability of each CRS to maintain software security. Is either 1 or 2 per CS per round.

EVALUATION

Score that reflects the ability of each CRS to find and prove the existence of vulnerabilities. Varies from 1 to 2 per CS per round.

FILAMENT / TRACE

A visual trace of data flowing through the code of a network service, with time moving from left to right.

COMMON WEAKNESS ENUMERATION (CWE)

A reference index of frequently encountered software vulnerability categories.

DOWNLOAD PLATFORM AND LIVE RESULTS

Test your mettle by downloading the CGC Virtual Machine and, once the air gap is certified, the challenges from **https://repo.cybergrandchallenge.com/CFE**. The live results of CFE, including competitor submissions, can be downloaded during DEF CON at **https://defcon.cybergrandchallenge.com/CGC**.

ABOUT CYBER GRAND CHALLENGE

The ultimate test of wits in computer security occurs through open competition on the global Capture the Flag (CTF) tournament circuit. In CTF contests, experts reverse-engineer software, probe its weaknesses, search for deeply hidden flaws and create securely patched replacements. What if a purpose-built computer could compete against the CTF circuit's greatest experts? The DARPA Cyber Grand Challenge is the world's first all-computer Capture the Flag tournament co-located with DEF CON, in which automated systems will take the first steps towards a defensible, connected future. The DARPA Cyber Grand Challenge seeks to create public proof that it's possible to automate the cyber defense process with machines that can discover, confirm and fix software flaws in real time.

ABOUT DARPA

For more than 50 years, the Defense Advanced Research Projects Agency has held to a singular and enduring mission: to make pivotal investments in breakthrough technologies for national security. In close collaboration with our Defense R&D partner agencies, DARPA engages top-tier public and private innovators—including academics, companies large and small, and colleagues across the federal government—to deliver on that mission, transforming revolutionary concepts and even seeming impossibilities into practical capabilities. For additional information, please visit **www.darpa.mil**.

#DARPACGC

CYBERGRANDCHALLENGE.COM