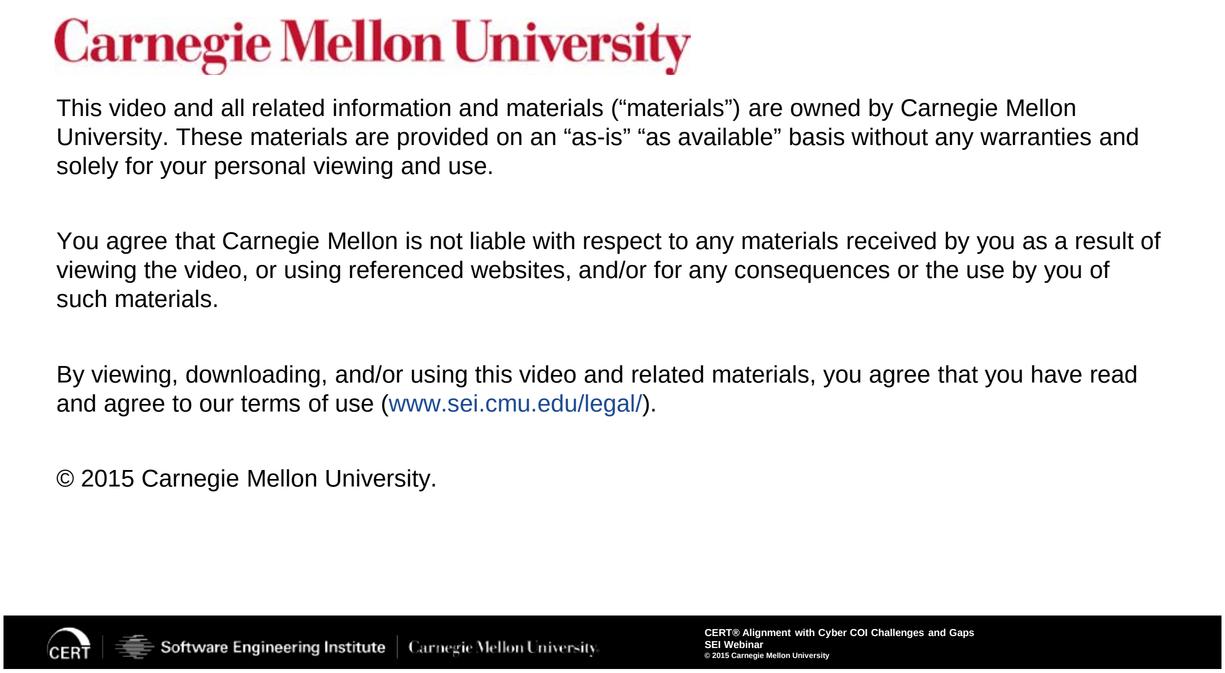
Generalized Automated Cyber-Readiness Evaluator (ACE) – Rotem Guttmann

Table of Contents

Carnegie Mellon University Notice



Carnegie Mellon University

This video and all related information and materials (“materials”) are owned by Carnegie Mellon University. These materials are provided on an “as-is” “as available” basis without any warranties and solely for your personal viewing and use.

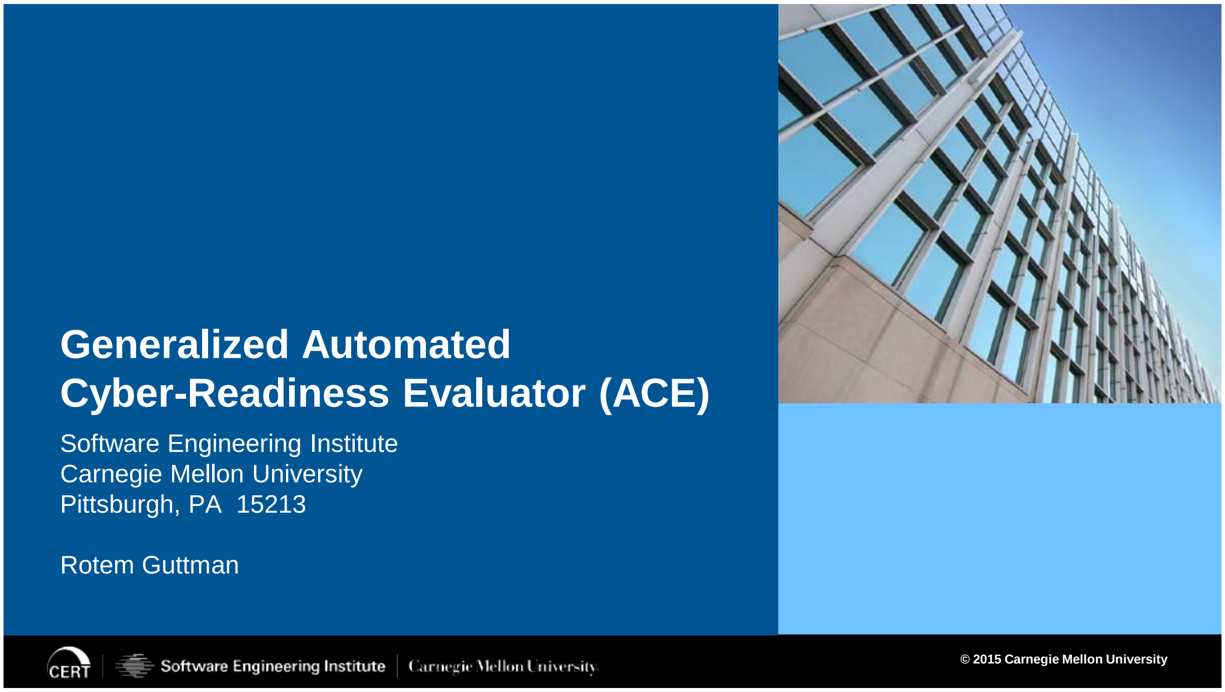
You agree that Carnegie Mellon is not liable with respect to any materials received by you as a result of viewing the video, or using referenced websites, and/or for any consequences or the use by you of such materials.

By viewing, downloading, and/or using this video and related materials, you agree that you have read and agree to our terms of use (www.sei.cmu.edu/legal/).

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar

**001 ~~~

Generalized Automated Cyber-Readiness Evaluator (ACE)



Generalized Automated Cyber-Readiness Evaluator (ACE)

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Rotem Guttman

CERT
Software Engineering Institute
Carnegie Mellon University

**041 Announcer: Our next topic

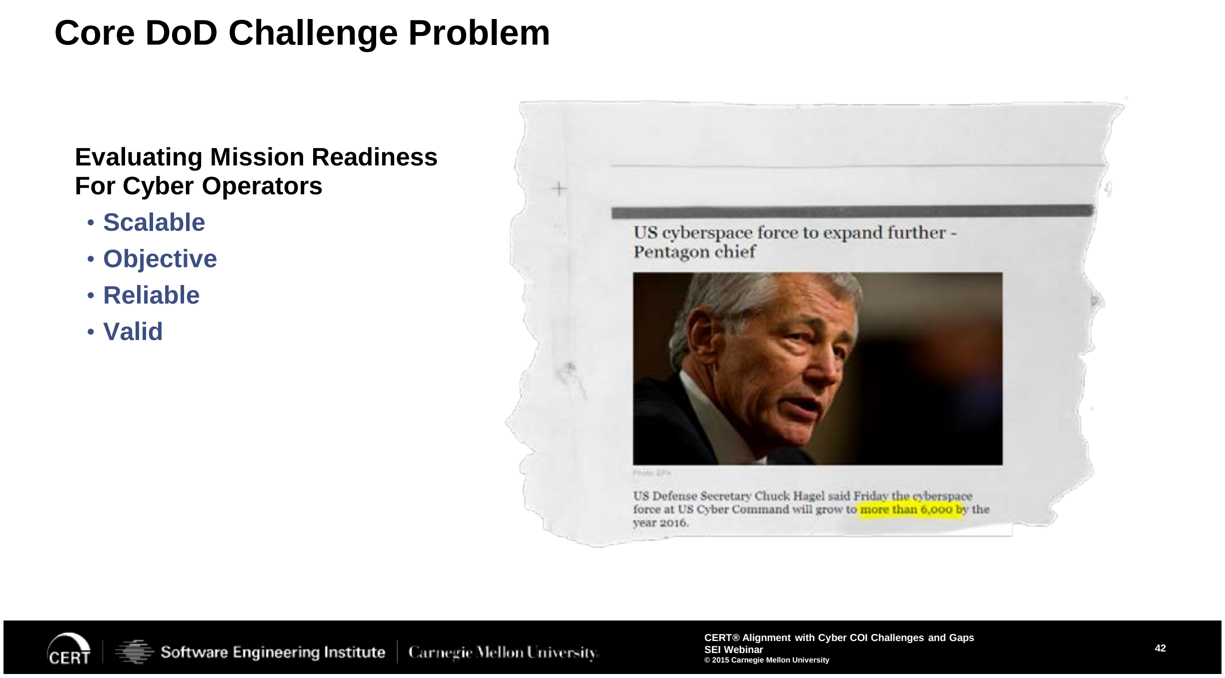
is "Generized Automated Cyber-Readiness Evaluator, or ACE," by Rotem Guttman.

Rotem is a cybersecurity researcher focused on developing new and engaging methods to improve training for cybersecurity personnel. Guttman is the PI for Automated Cyber-Readiness Evaluator research project and developing a system to the Automated Evaluation Cyber Personnel.

Now I'd like to turn it over to Rotem Guttman. Rotem? All yours.

Presenter: Thank you. Okay. So I am the PI for the Generalized Automated Cyber-Readiness Evaluator.

Core DoD Challenge Problem



Core DoD Challenge Problem

Evaluating Mission Readiness For Cyber Operators

- Scalable
- Objective
- Reliable
- Valid

US cyberspace force to expand further - Pentagon chief

Photo: EPA

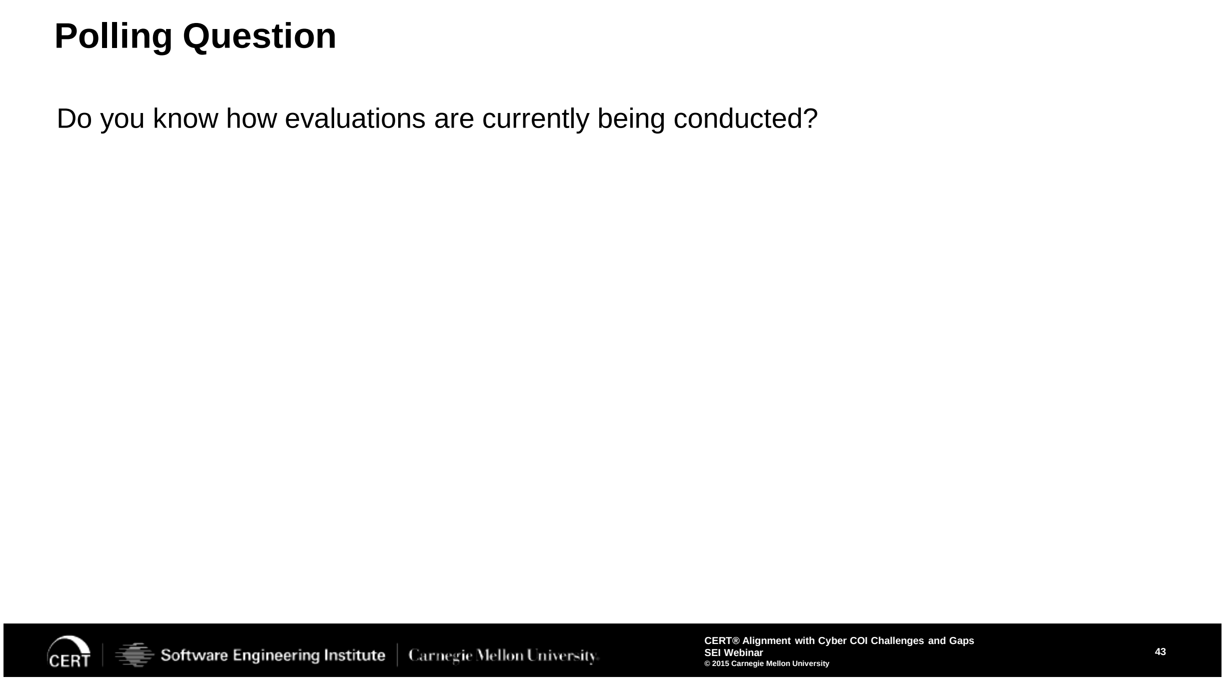
US Defense Secretary Chuck Hagel said Friday the cyberspace force at US Cyber Command will grow to more than 6,000 by the year 2016.

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar
© 2015 Carnegie Mellon University

42

**042 The key problem that we're really looking at with the system is the fact that DoD needs to staff up to a rather sizable amount of cyber operators, over 6,000 by 2016. And the problem is all of these personnel need to be evaluated as mission ready. And in order to do that we have to do that in a manner that's first and foremost scalable. But beyond that it also has to be objective. Whether the evaluation is done at one location or another, it has to be to the same criteria or it's useless to the commander to know what level they're actually ready for. It has to be reliable. It has to be valid. It has to be something that we can go back and certify that we can have confidence in the evaluation.

Polling Question



Polling Question

Do you know how evaluations are currently being conducted?

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar
© 2015 Carnegie Mellon University

43

**043 So the--well, pass that to you.

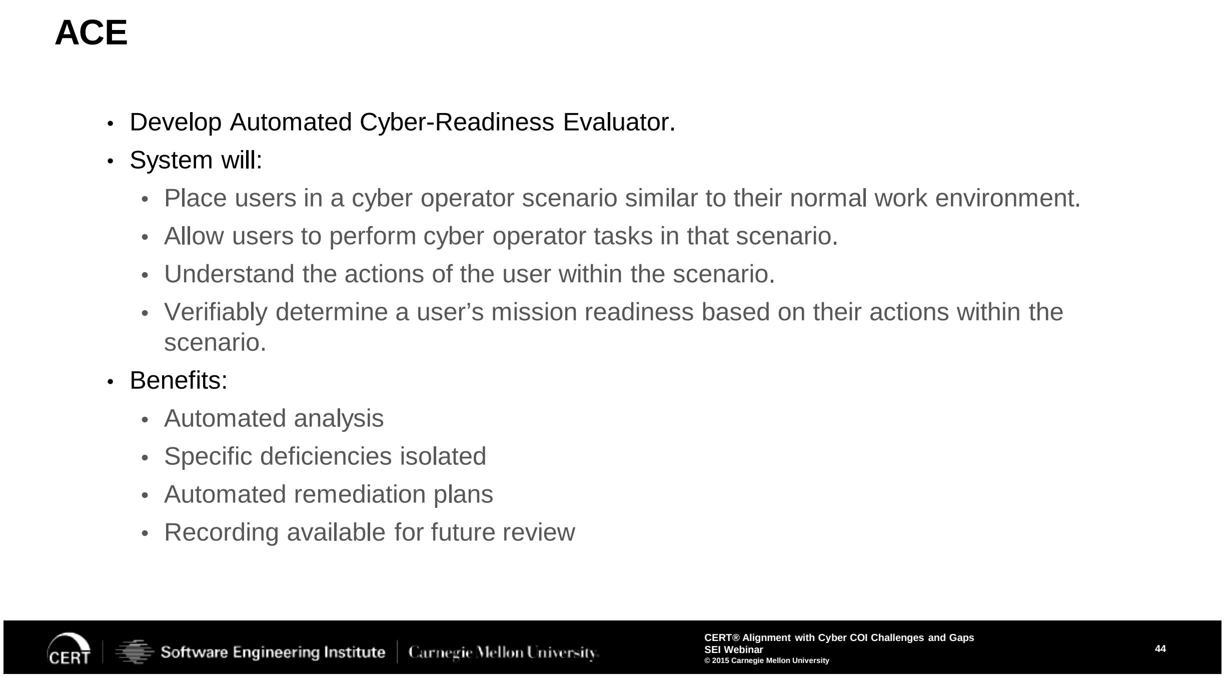
Announcer: Yeah. So we're going to kick it off with a polling question. And that'll be on your screen now. Just wanting to know, do you know many evaluations that are currently being conducted? So there's no right answer, but what we want you to--there is a right answer, but since it's a polling question we need to you to type the answer into the Q&A section for your guess. So there's not a multiple choice option appearing in the pop-up, so type in your answer.

Presenter: There's more than one right answer.

Announcer: So good. So type in your answer into the Q&A field and we'll let Rotem move on, and then we'll come back to some of the responses once you're ready.

Presenter: Okay.

ACE



ACE

- Develop Automated Cyber-Readiness Evaluator.
- System will:
  - Place users in a cyber operator scenario similar to their normal work environment.
  - Allow users to perform cyber operator tasks in that scenario.
  - Understand the actions of the user within the scenario.
  - Verifiably determine a user's mission readiness based on their actions within the scenario.
- Benefits:
  - Automated analysis
  - Specific deficiencies isolated
  - Automated remediation plans
  - Recording available for future review

CERT | Software Engineering Institute | Carnegie Mellon University

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar
© 2015 Carnegie Mellon University

44

**044 Well, have they had a chance to put it? Because I was going to talk a little about--

Announcer: Ah, let me see here.

Presenter: --what's being done right now.

Announcer: Let's go to Q&A. Let's hit Refresh.

Presenter: So in the meantime, basically what we're proposing is to address some of the issues with the way that evaluations are being done right now. The methods that are done right now don't meet the criteria that I outlined earlier. And so because of that we want to evaluate users in a more natural system, in a more natural method. So the system that we're designing places the cyber operator in their normal work environment. So basically we have them sit down and do their job and when we ask them to do their job we put them in a scenario. So we basically say, "Here's what's going on. Here's your mission. Complete the mission." More importantly, the system can then understand what they do in that scenario. What they do in the environment. And based on their actions, actually make a determination of their mission readiness.

Now, the benefits of this are rather broad. First of all, the automated analysis. By automating the analysis we don't have to have personnel to do the evaluation, dedicated personnel to sit and do the evaluations. Have we had--

Announcer: We have 100. We have 50. We have 10,000. So the number just all over the board.

Presenter: Oh, no. I'm sorry. Maybe those--

Announcer: Oh, from the poll. Yeah.

Presenter: Yeah. The polling question was how the evaluations are being done, not how many.

Announcer: Oh, I'm sorry. Yeah. I read it wrong, so that is my fault.

Presenter: Oh, that's quite all right. I'll just give away the answer.

Announcer: Okay.

Presenter: So first of all, as far as the number goes, there are thousands being done.

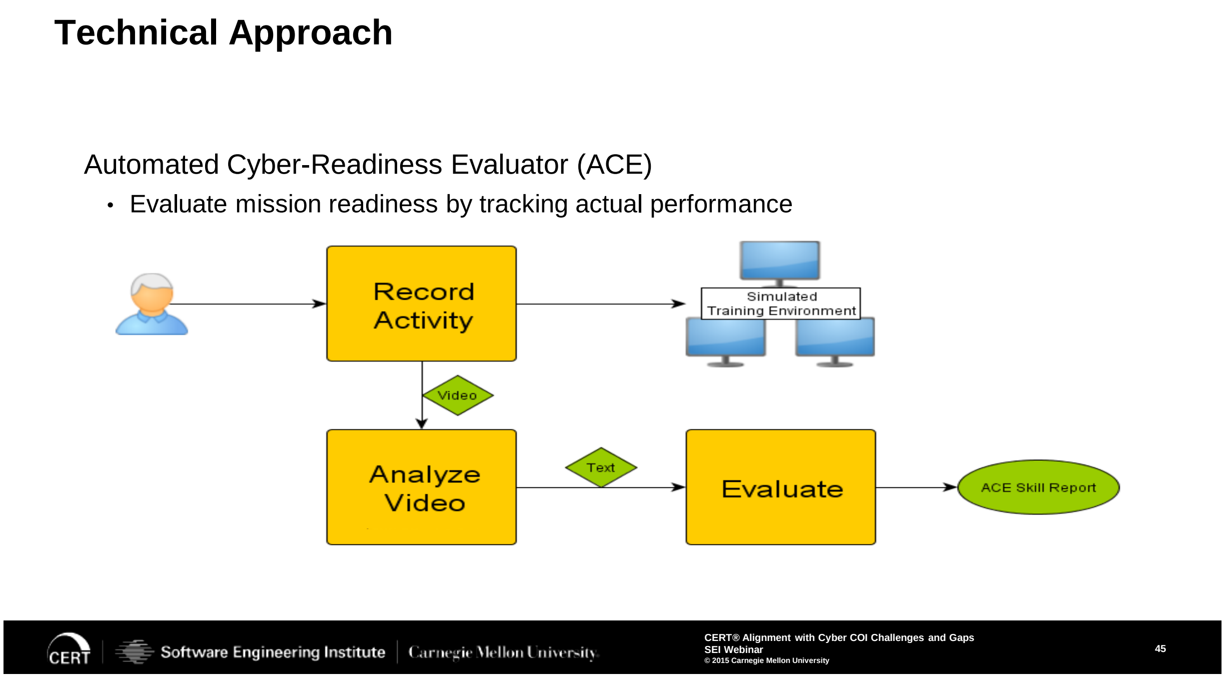
Announcer: Okay.

Presenter: Right now. Across all of the cyber operator roles. But the methods vary. Unfortunately, a lot of groups are actually relying on checklist solutions, where basically the criteria are a checklist of their mission readiness criteria. For example, knowledge of file carving tools is one of the criteria. So the evaluator would actually stand behind the evaluee, look over his shoulder tell him, "Carve out all the files on this machine, all the PDF files on this machine." Watch them do it, and then put a check mark in the box. Now, the problem with that is that it's not uniform. If he struggles, if he kind of gets it done halfway, whether that gets marked or not is really up to the evaluator. Beyond that, you're prompting him to do it. The fact that he knows when told, "Go carve out the files," doesn't mean that as part of his job he knows when to carve out files or why he's carving them out. It doesn't get to the deeper requirement of can he actually fulfill his tasks? Can he actually do his job?

So back to the benefits. The automated analysis relieves us of the need for that actual evaluator to stand over their shoulder of those one-on-one evaluations. Another problem is actually because of that need, because we're growing so quickly in the amount of cyber operators, certain groups have actually been unable to even do the checklist approach and have fallen back on self-assessments. Where actually personnel are just getting that list and filling out for themselves what their skill levels are on various capabilities.

Additionally we can isolate specific deficiencies. So if we have this large document that specifies all of the criteria for mission readiness for a particular job role we can actually specify, "Okay. He's actually mostly mission ready except for here's the three skills that he's lacking in, might need additional training in." And because we can identify that we can generate automated remediation. Remediation plans. We can say, "Okay. These are the three skills that he's lacking in, so here's some course material that we have, additional training that we have," that we can then auto enroll him in. And beyond that, having the data available is extremely valuable because we can go back and review it in the future. We can take a look, you know, a year or two years, three years down the line and actually take a look. "Okay. Here's your top performers. Let's go look back at when they first started working, when they first finished their training. What traits did your top performers have that perhaps the other people didn't? How can we train towards those skills as well?"

Technical Approach



Technical Approach

Automated Cyber-Readiness Evaluator (ACE)

- Evaluate mission readiness by tracking actual performance

Record Activity

Simulated Training Environment

Video

Analyze Video

Text

Evaluate

ACE Skill Report

CERT | Software Engineering Institute | Carnegie Mellon University

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar

45

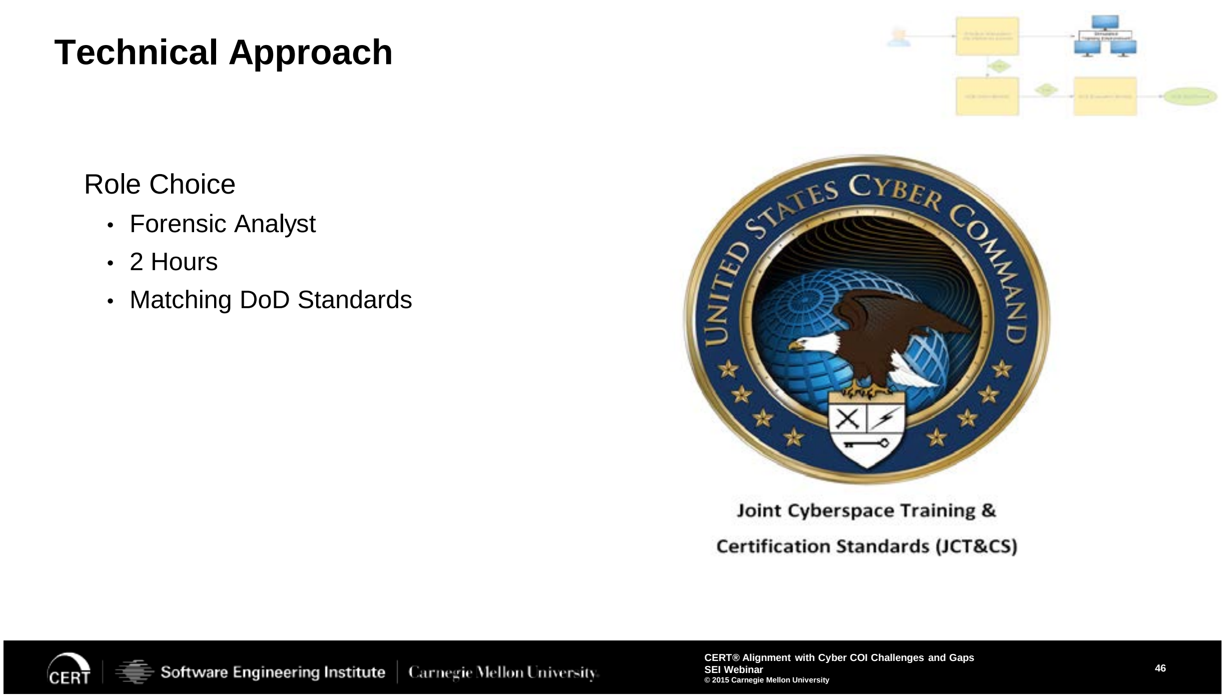
**045 So this is all well and good. Something that we already have the system, but how do you actually instrument such a system? So ACE is designed to put users in a real environment. So the first thing that we need is a real environment for them to operate in. Thankfully we have developed here at CERT the STEPfwd platform. And the STEPfwd platform, if your users aren't already familiar with it, it allows us to push out full training environments. Now, this could be anything from the learning management system where we have thousands of hours of classroom content, lectures, materials, where they can actually sit and get training on these topics. In addition to hands-on labs where they can sit down and actually practice doing particular skills. And this is the material that they would get auto enrolled in if there are deficiencies found.

But the key aspect of the STEPfwd platform that we're leveraging in ACE is actually the exercise capability. So we have exercise and simulation capability in there where we can actually push out anything from one individual workstation to entire corporate networks with thousands of machines and thousands of users that we can push out to people anywhere in the world through the browser. I'll go into a little more detail about that later, but what we did is we leveraged that capability in order to take the environment that we developed and push it out to our evaluees.

So the next thing that we needed after that is the ability to record their capabilities. So by recording their capabilities we can then capture their behavior. Now, this recording takes the form of an actual video stream of everything that they do in the environment. That video output then gets transferred to our video analysis engine. This is known as ACE Vision. ACE Vision can actually take that video and parse out what's going on on the screen and determine, "What are they doing right now?" Beyond that, the output from that system, now text, gets passed into the ACE Eval system. Ace Eval is a two-layer system where the first layer basically takes the raw detections from the video system of, "Okay. He did this thing. He clicked on that icon. He typed this thing." And then pulls that text in and translates it into, "Okay. Now, here's what he's doing. Here's the action that he's taking right now."

The second layer of that system then takes that log of actions, activities, and maps those activities to the knowledge, skills and abilities that they have or have not displayed. The end result of that second layer is the ACE Skill Report. So the ACE Skill Report is the final result, what we actually want to give to the commanding officer or to the evaluee himself. That says, "Okay. Here's where you have met the criteria for mission readiness," or, "Here's where you're deficient." So in order to do this, the first thing we needed is an actual scenario.

Technical Approach



Technical Approach

Role Choice

- Forensic Analyst
- 2 Hours
- Matching DoD Standards

Joint Cyberspace Training & Certification Standards (JCT&CS)

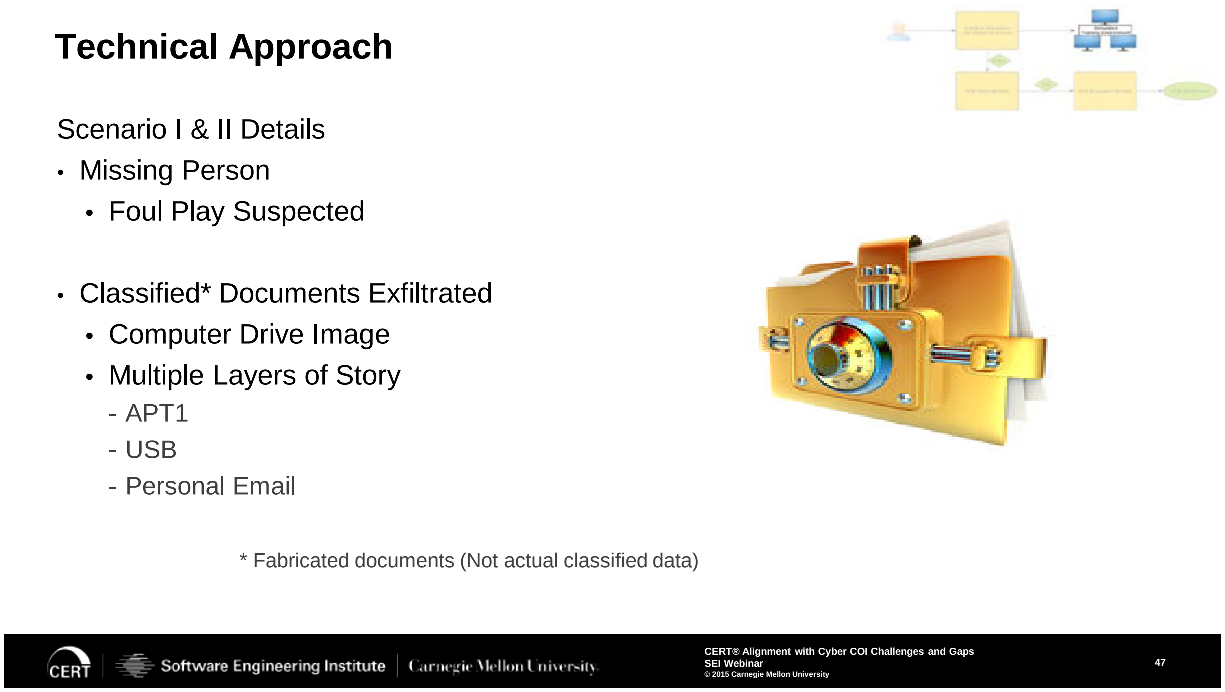
CERT | Software Engineering Institute | Carnegie Mellon University

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar
© 2015 Carnegie Mellon University

46

**046 In order to do that we need to know, "Okay. Which job role are we going to address first?" We chose the role of a forensic analyst, simply because he can be reasonably expected to do his job, to do an initial analysis within two hours, when given a capture. And we had a good standard to work from. We used the JCT&CS standard material for the forensic analyst role. Additionally, the forensic analyst role lets us minimize our development footprint. A forensic analyst can be reasonably expected to do his job on an analysis workstation. We wanted to get to the actual research problem of doing this analysis without having to build a large amount of infrastructure. For example, in incident responder role you would have to build out the full network, let them respond to the incident, jump to different machines throughout their evaluation. Now, the platform, I want to be clear, has the capability to do that, but we want it to minimize the footprint for the purpose of the research.

Technical Approach



Technical Approach

Scenario I & II Details

- Missing Person
  - Foul Play Suspected
- Classified* Documents Exfiltrated
  - Computer Drive Image
  - Multiple Layers of Story
    - APT1
    - USB
    - Personal Email

* Fabricated documents (Not actual classified data)

CERT | Software Engineering Institute | Carnegie Mellon University

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar
© 2015 Carnegie Mellon University

47

**047 Now, as I mentioned, now that we have the role, we can develop the scenario. So we've actually developed two scenarios. The first scenario is a missing persons case. Essentially there's a person of interest that we wanted to question. We went to go to their residence. Their residence was trashed, there's evidence of foul play. We manage to recover a damaged laptop. Here's the hard drive from it. Get us any actionable intel you can as fast as possible. You know, we want to speak to this person. We need to figure out where they are. And so it's on the forensic analyst to actually look at that drive and get the relevant information back to us in a timely manner so agents can actually go out and speak to this person, apprehend them if possible.

The second case is a somewhat larger case where there are classified documents. And I say classified. These are fabricated documents that take place of classified documents, just to be clear. So these documents, we know that they're in enemy hands. We know that they've been found to have already leaked. We want to know, "How did they get there?" And because of the nature of the documents, we know that they were created at a particular defense contractor. That what we then do is give forensic capture of these machines from the defense contractor to the forensic analysts and they have to actually determine, "Okay. How did these files get from point A to point B?" And what's nice about this story is there's actually multiple layers to the story.

So there's indications of possible activity by APT1. Indications of hardware being connected to these machines that probably shouldn't have been connected to them. Indications of accessing personal e-mail and networks that probably should not be accessed from this type of machine. So there's a lot going on there for the analyst to sift through, and a skilled analyst can actually look, "Okay. Where are the things that matter and where are the things that are just distracting me from the job at hand?"

Technical Approach

Technical Approach

Data Capture Capability

- Background Data Collection
- Restricted to Environment
- Scalable

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar
© 2015 Carnegie Mellon University

48

**048 Next we had to actually have the capability of doing that data capture.

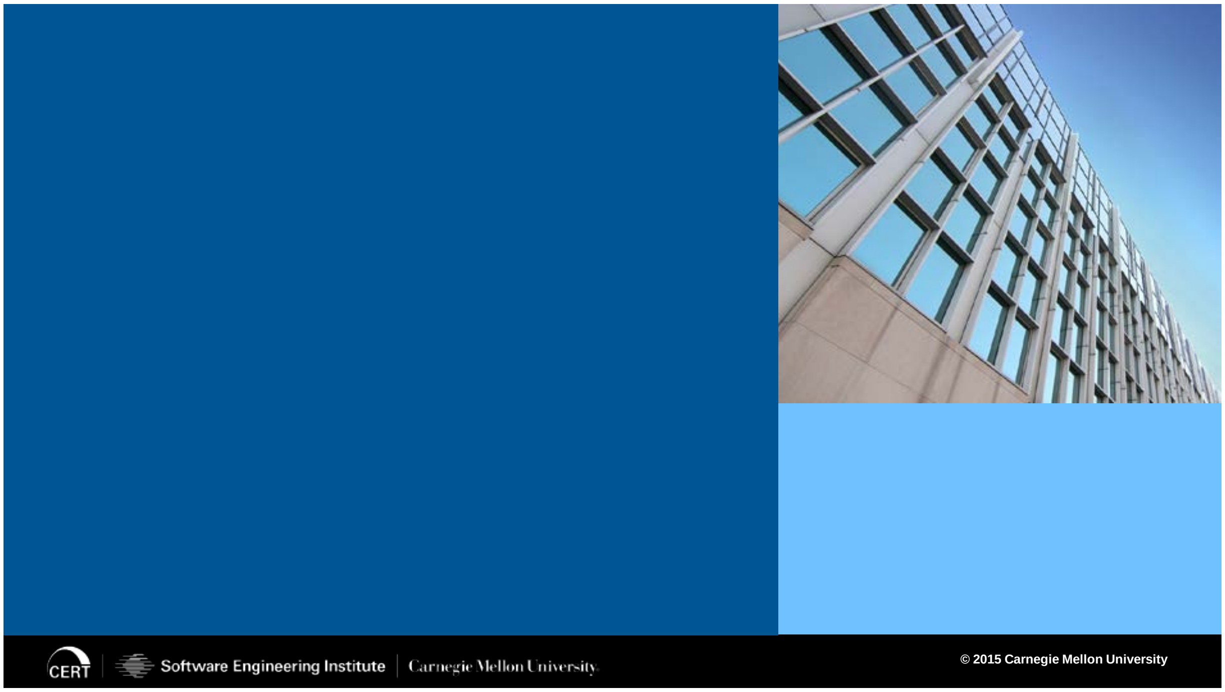
Getting that video stream.
So that was a capability that was added to the STEPfwd platform. And I just want to take a moment to talk about that capability itself. So this is a data capture capability that happens in the background. So we can actually set up the accounts for people that are going in for evaluations and mark them, "Okay. This is an evaluation exercise. This is an evaluation scenario. And this person is the person being evaluated." And then whenever they log in to that particular simulation, everything that they do will get recorded. That being said, the recording is being done server side. We're only recording the video stream that we're pushing out to

them. So by doing that, we restrict the recording to the environment. That way we prevent a lot of possible captures of data that we don't want. We don't want to see what's on their desktop. We don't want to see what's in other tabs of their browser or whatever they might be doing on their machine. This is good both for us and for them. It's good for us in terms of IRB review and meeting the criteria that we need to to respect their privacy, but it's good for them as well to know that their privacy is being respected and they don't have to worry about us accidentally collecting data about them specifically.

Additionally, we've disabled all of the other channels that could go in. You can't even cut and paste into the environment simply because we didn't want them to accidentally paste in something that was in their clipboard from before. So we've put a lot of attention there to making sure that our data capture is only the data that, of their behavior, in our environment.

Beyond that, the way that we're doing the data capture also makes it much more scalable. Because of how we're doing it, we can have 20, 30, 40 users all on one blade, and we can actually separate this out into shards. We could have a large number of blades dedicated to doing these captures, so it's scalable to a DoD scale of problem.

Actual data is needed

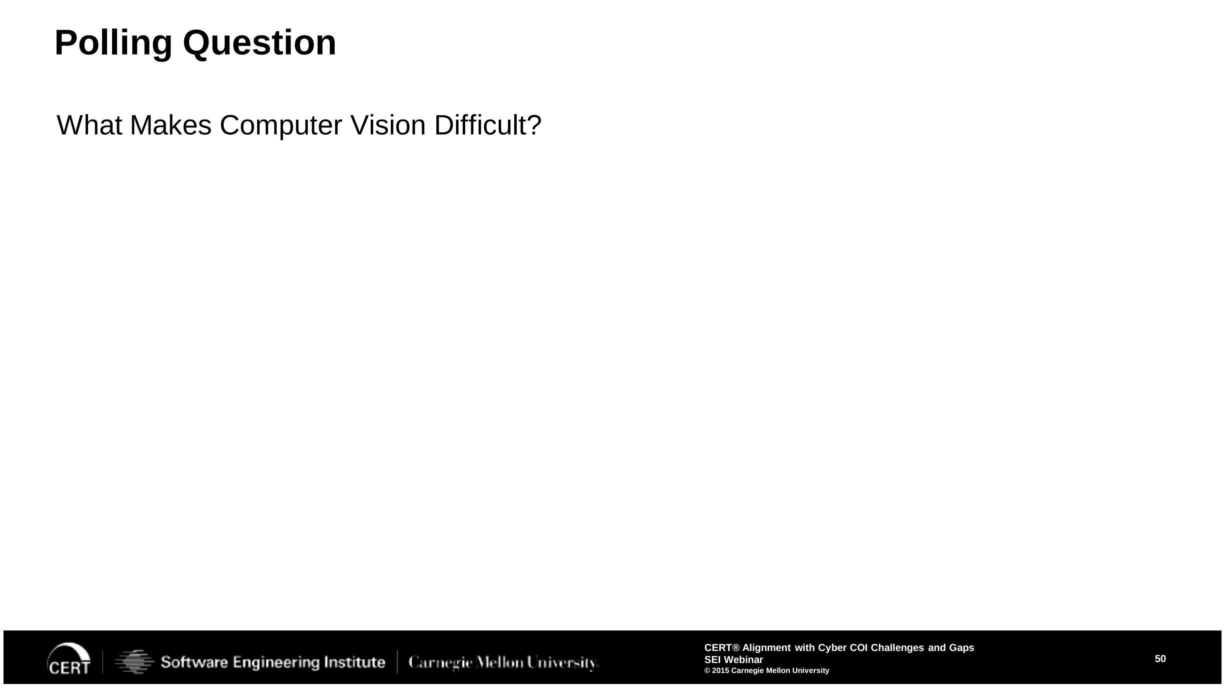


**049 The next thing we needed is actual data. Now that we have the capability, we need to actually record the behavior of forensic analysts. So what we want here is really a broad spectrum of capability. We want to be able to look at these forensic analysts, both ones that are, you know, top of their field, know what they're doing. Ones that are just out of training, may or may not be ready. And even people that aren't necessarily, wouldn't be mission ready. We want to see people that would fail this analysis that might be lacking some or all of the skills that are required. So in order to address those needs, what we did is we actually had CERT staff do captures, and the CERT staff generally gave us the higher end of the capabilities. These are people that are familiar with the software, these tools. Some of them even wrote these tools.

So in order to give us a broader variety, we also did captures with CMU graduate students. We actually integrated with, there's a forensics program here on campus we integrated with their capstone class where they actually do some of these scenarios as part of their classroom material and they can volunteer to have their data also captured so that we can use it. This gave us a better variety of skill level. So you have your A students who would probably be mission ready today. And you have your students that might not be getting an A in the class that have, you know, more or less capabilities in this realm.

Additionally, we had DoD personnel that we integrated into this. These are live personnel that are, you know, in these roles right now or coming into these roles now that we had do these captures as well. As well as volunteers from the National Cyber Forensic Training Alliance that were willing to come in and actually participate in our evaluations. We're also looking at integrating with CMU's information security office as part of their continuing education plan.

Polling Question



Polling Question

What Makes Computer Vision Difficult?

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar
© 2015 Carnegie Mellon University

50

**050 Announcer: Okay. So next. Another polling question. It's going to be on your screen now. And again, just feel free to type in your response to the Question and Answer section. And that question is, "What makes computer vision difficult?" So I'm assuming there's going to be a diversity of answers here, so--

Presenter: Oh yeah. I expect that there's a lot of good answers.

Announcer: We'll give them-- so again, there's no multiple choice here. Just type that right into your Q&A box in your console.

And while we're waiting for that, there's a couple questions streaming in throughout the day about, asking about CEU credit for the event. We are offering CEUs for attending the event, to get the certificate of attendance for the event. Just send an e-mail to info@sei.cmu.edu, and we'll be sure to get your certificate out there to you.

And let's see what we get here. I don't know. People are being shy. Nothing's coming in yet. But...

Presenter: Okay.

Announcer: Hold on. We got some in.

Presenter: Oh, there we go. They showed up.

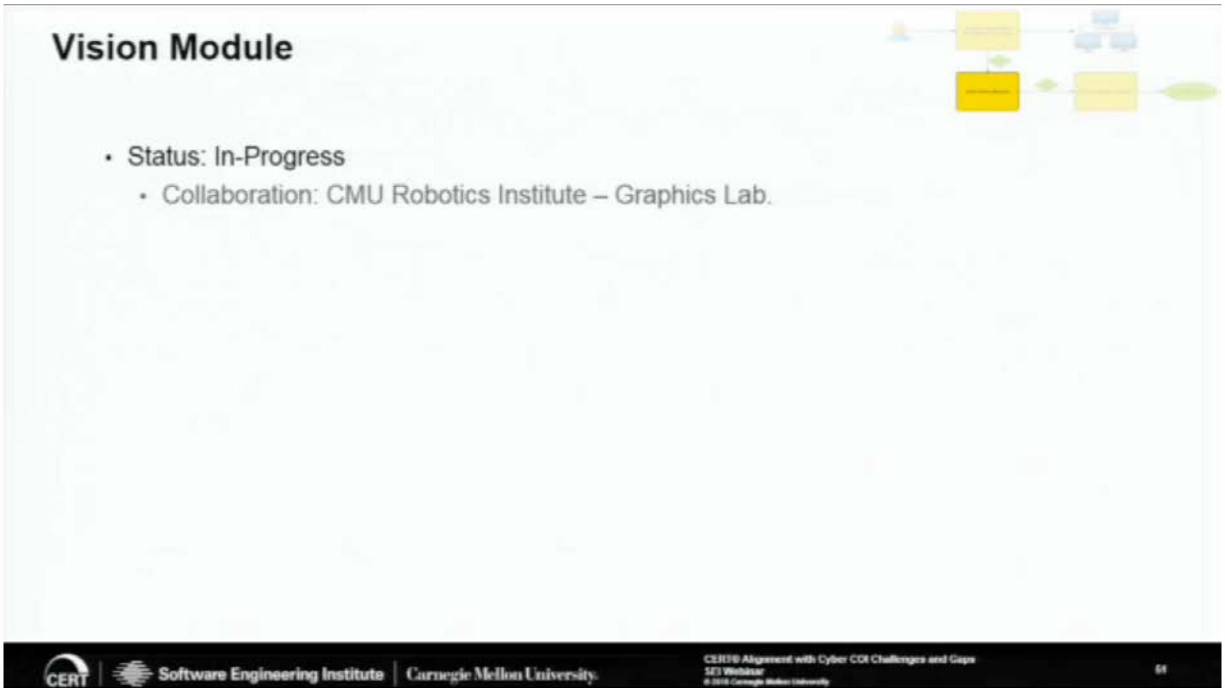
Announcer: Bright white background. How about that? Too much data. Another. What else? Full context and only the necessary context?

Presenter: Yeah. So a lot of the answers are, I mean, all of that's true.

Announcer: Yeah. Lighting, reflection, complexity. Another one.

Presenter: Oh, there. That's a great answer. So actually I'll address some of that. What we wanted to do is actually look at these systems. So take these video captures in an automated fashion. Determine, you know, "What's going on here?" And that's a really hard question. We know that's a really hard question. I mean, since the '80s people have been looking at, "How do we do computer vision? How do we have a computer know what it's looking at?" And so we didn't want--why did we think that we could address this question in a manner that's relevant? Well, the reason is that those things that you listed, you know, occlusions, lighting, perspective, shadows, you know. All of these things that make computer vision hard go away in our environment. So basically we cheated. We've created an environment that's optimized for doing this analysis.

Vision Module



Vision Module

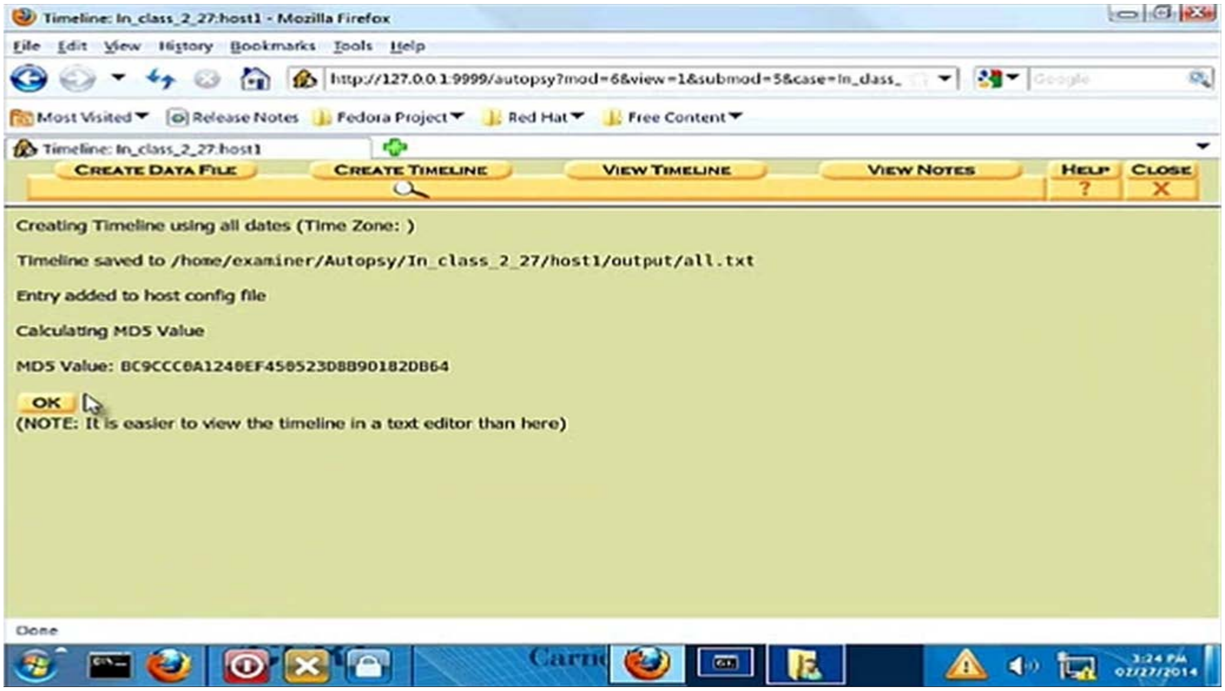
- Status: In-Progress
  - Collaboration: CMU Robotics Institute – Graphics Lab.

CERT | Software Engineering Institute | Carnegie Mellon University

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar

**100 So we're currently in progress developing the computer vision module. I'll give some data in a little bit about where we are on that. But I wanted to give an idea of how it actually works, how we're taking advantage of the way that we've instrumented the environment.

Vision Module

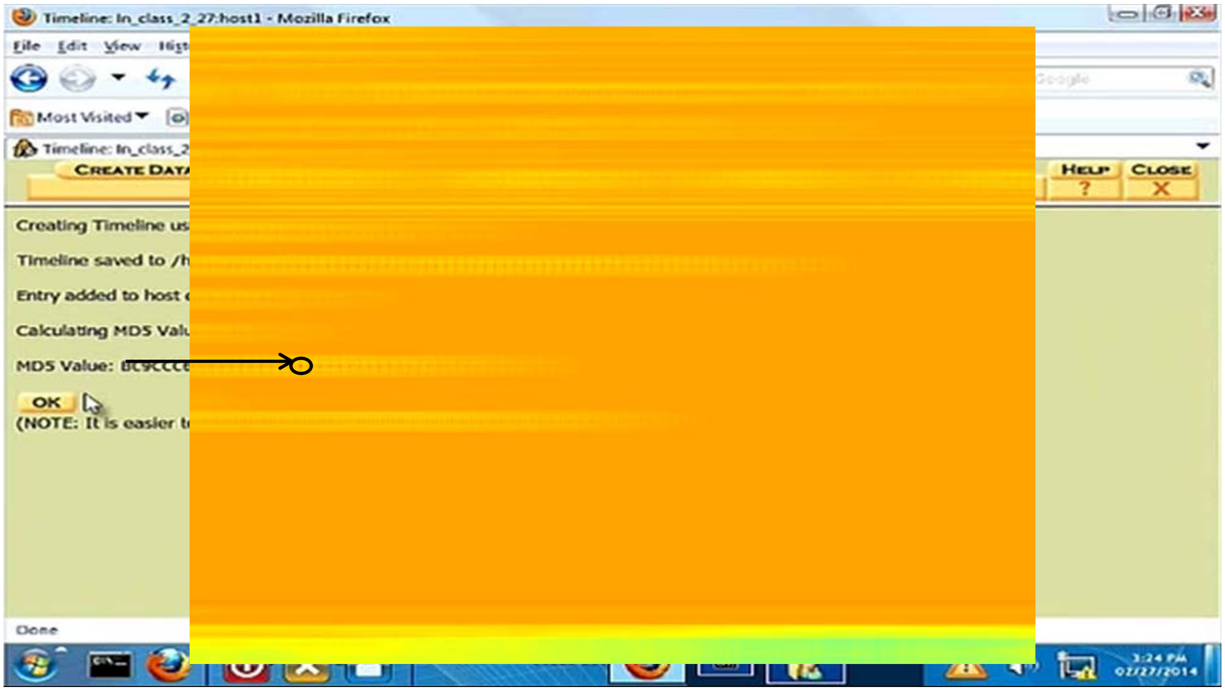

Timeline: In_class_2_27:host1 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://127.0.0.1:9999/autopsy?mod=6&view=1&submod=5&case=In_class_
Google
Most Visited Release Notes Fedora Project Red Hat Free Content
Timeline: In_class_2_27:host1
CREATE DATA FILE
CREATE TIMELINE
VIEW TIMELINE
VIEW NOTES
HELP
?
CLOSE
X
Done
3:24 PM
02/27/2014

```
Creating Timeline using all dates (Time Zone: )

Timeline saved to /home/examiner/Autopsy/In_class_2_27/host1/output/all.txt

Entry added to host config file

Calculating MD5 Value

MD5 Value: BC9CCC0A1240EF450523D8B90182DB64

OK
(NOTE: It is easier to view the timeline in a text editor than here)
```

**051 So here's a still from actually that same video capture that you saw just a couple of slides ago. In the still you can see that the forensic analyst has loaded the environment. This is the autopsy web interface. And so the nice thing is, the way that a lot of the graphical elements in this data capture are rendered right now, they're going to be rendered exactly the same way. We actually control, you know, the screen resolution, the fonts, the colors. The way that that graphic is rendered this time and the next thousand times is going to be pixel by pixel, exactly the same, because we're actually capturing the video stream right off the graphics that we're pushing out to them. So we're getting it at full quality, no distortions, and we're able to leverage that to create actually template images of the elements that we want to track.

So to give you one example, one of the things that's actually in the requirements is "verifies the state of the evidence." So they're given an MD5 value for the evidence when it was captured and have to recreate that value and verify it. So as you can see on screen, he's done that. How does our system know that he's done that? There's a myriad of ways to do this. He could've done MD5 sum, MD5 deep or generated it in any number of forensics tools. So as you see on the screen here, he has generated the MD5 value here.

Vision Module



Creating Timeline us

Timeline saved to /h

Entry added to host

Calculating MD5 Valu

MD5 Value: bc9ccce

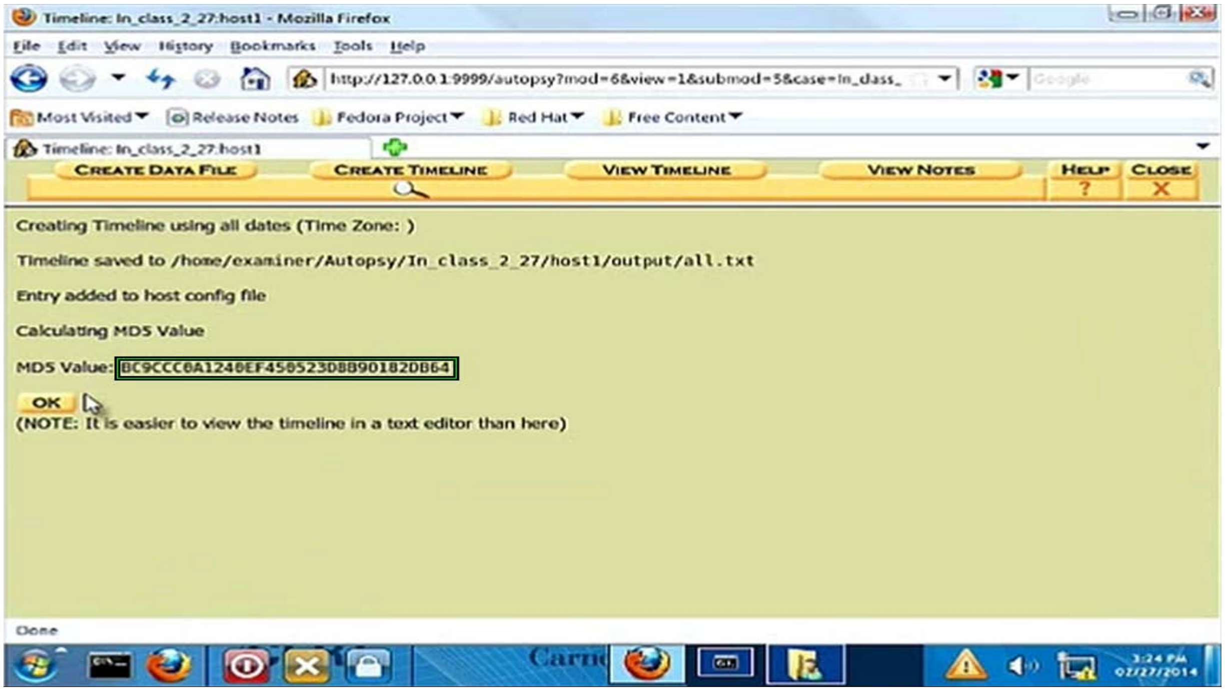
OK

(NOTE: It is easier t

**052 And so what you're seeing now is the vision system as it parses this particular frame, this is a, and I want to be clear here, this is a representation of the memory structure that's created in memory. We don't actually create this image for each frame. This is just to give us a visualization.

So this memory structure gives us the confidence that it's found this template at any particular location. If you look closely, most of it gives you a nice, smooth background, you know, relatively low confidence. But right at this location there's one pixel where you get a bright red dot. Where you have a much, much higher confidence, because we have had a good match to the template of the MD5 value. That location maps to this position on screen. So now you can see we've had a positive detection and now the system knows, "Okay.

Vision Module


Timeline: In_class_2_27:host1 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://127.0.0.1:9999/autopsy?mod=6&view=1&submod=5&case=In_class_
Most Visited
Release Notes
Fedora Project
Red Hat
Free Content
Timeline: In_class_2_27:host1
CREATE DATA FILE
CREATE TIMELINE
VIEW TIMELINE
VIEW NOTES
HELP
CLOSE
OK
Done
3:24 PM
02/27/2014

Creating Timeline using all dates (Time Zone:)

Timeline saved to /home/examiner/Autopsy/In_class_2_27/host1/output/all.txt

Entry added to host config file

Calculating MD5 Value

MD5 Value: BC9CCC0A1240EF450523D8890182DB64

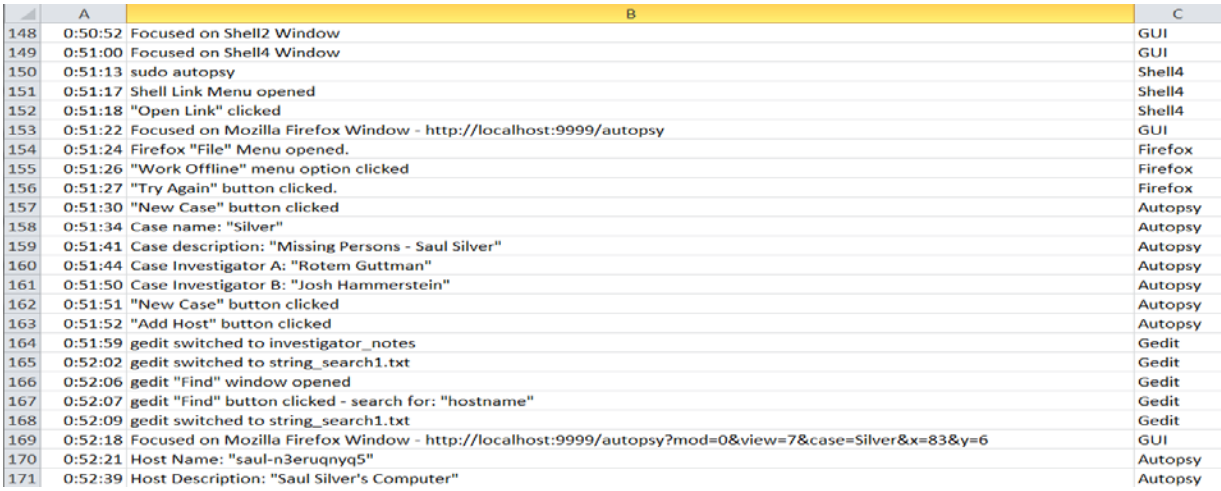
(NOTE: It is easier to view the timeline in a text editor than here)

**053 We've detected that MD5 value." Now, what we're doing is 15 times per second, so at 15 frames per second we're actually grabbing stills from the video stream and analyzing them. So now what the system's going to look for in the vision system, it's going to detect that there is this MD5 value and add it to that output stream.

Now in reality, this is just one of hundreds, in some cases for the second scenario, thousands of templates that we've created for that scenario. Now, those templates then get taken and it does the same analysis for a large variety of them at once. The way that it's doing it at once is we use a series of PCA filters in order to determine which templates actually exist in the sample.

So moving forward, this is
the type of output that you'll get
from the first layer of the evaluation
system.

Spreadsheet

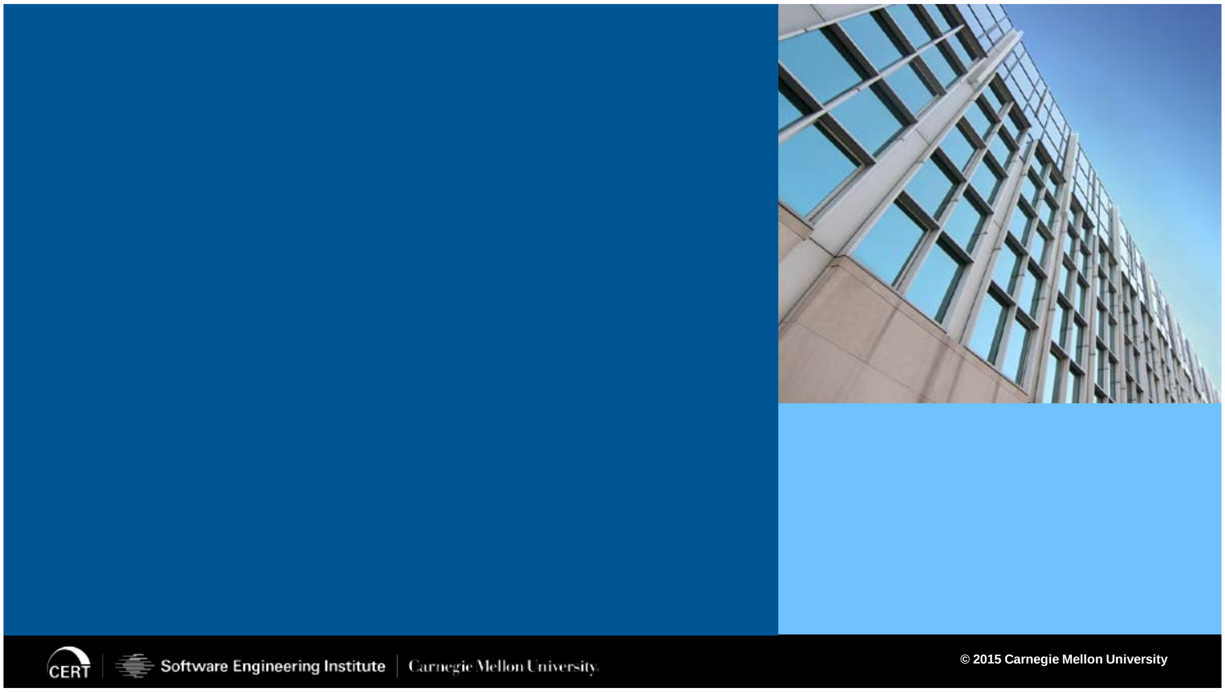


	A	B	C
148	0:50:52	Focused on Shell2 Window	GUI
149	0:51:00	Focused on Shell4 Window	GUI
150	0:51:13	sudo autopsy	Shell4
151	0:51:17	Shell Link Menu opened	Shell4
152	0:51:18	"Open Link" clicked	Shell4
153	0:51:22	Focused on Mozilla Firefox Window - http://localhost:9999/autopsy	GUI
154	0:51:24	Firefox "File" Menu opened.	Firefox
155	0:51:26	"Work Offline" menu option clicked	Firefox
156	0:51:27	"Try Again" button clicked.	Firefox
157	0:51:30	"New Case" button clicked	Autopsy
158	0:51:34	Case name: "Silver"	Autopsy
159	0:51:41	Case description: "Missing Persons - Saul Silver"	Autopsy
160	0:51:44	Case Investigator A: "Rotem Guttman"	Autopsy
161	0:51:50	Case Investigator B: "Josh Hammerstein"	Autopsy
162	0:51:51	"New Case" button clicked	Autopsy
163	0:51:52	"Add Host" button clicked	Autopsy
164	0:51:59	gedit switched to investigator_notes	Gedit
165	0:52:02	gedit switched to string_search1.txt	Gedit
166	0:52:06	gedit "Find" window opened	Gedit
167	0:52:07	gedit "Find" button clicked - search for: "hostname"	Gedit
168	0:52:09	gedit switched to string_search1.txt	Gedit
169	0:52:18	Focused on Mozilla Firefox Window - http://localhost:9999/autopsy?mod=0&view=7&case=Silver&x=83&y=6	GUI
170	0:52:21	Host Name: "saul-n3eruqnyq5"	Autopsy
171	0:52:39	Host Description: "Saul Silver's Computer"	Autopsy

**054 So it takes those raw
detections, treats them as essentially
the, each detection is essentially the
letters in alphabet, essentially. And it
takes those and determines, okay,
which letters appear?

What is that in
these particular frames or in these
groups of frames? What activity does
that map to? So it's either a
switched focus to a shell window or
clicked on a link or ran this
application, et cetera. And then
taking that output it matches it with a
timestamp for when that appeared in
the stream.

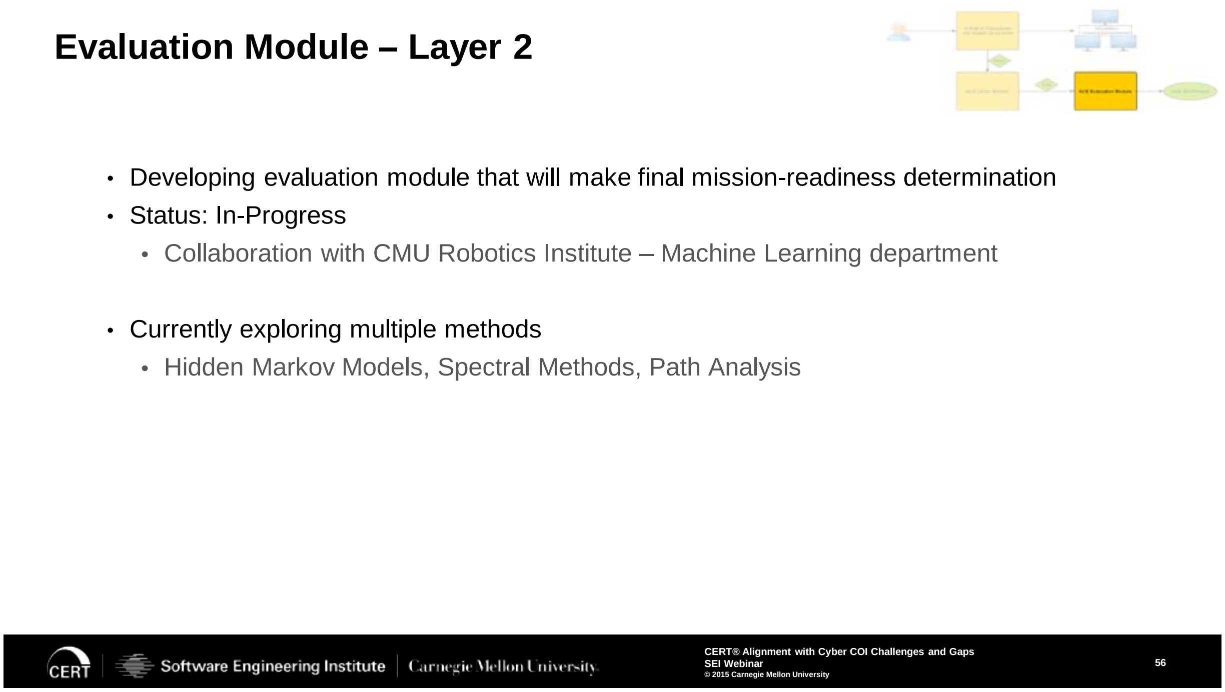
Actual data capture video


CERT
Software Engineering Institute
Carnegie Mellon University
© 2015 Carnegie Mellon University

**055 What you see here now is an actual video. Now, again, full disclosure. We don't actually generate this video for each output. I've just created this in order to give you a visualization. On the top of the screen, everywhere above that subtitle bar, you can see this is an actual data capture. This is exactly the video that we're capturing. What I've done is I've placed on the bottom there the actual output from that first layer of the vision system, what that type of output would look like at that time, and I've just matched the timestamps of that output to the subtitles.

And so you can see they're actually going through and clicking, going, doing the analysis, and the system is tracking what behavior they're doing at that time.

Evaluation Module – Layer 2



Evaluation Module – Layer 2

- Developing evaluation module that will make final mission-readiness determination
- Status: In-Progress
  - Collaboration with CMU Robotics Institute – Machine Learning department
- Currently exploring multiple methods
  - Hidden Markov Models, Spectral Methods, Path Analysis

CERT | Software Engineering Institute | Carnegie Mellon University

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar
© 2015 Carnegie Mellon University

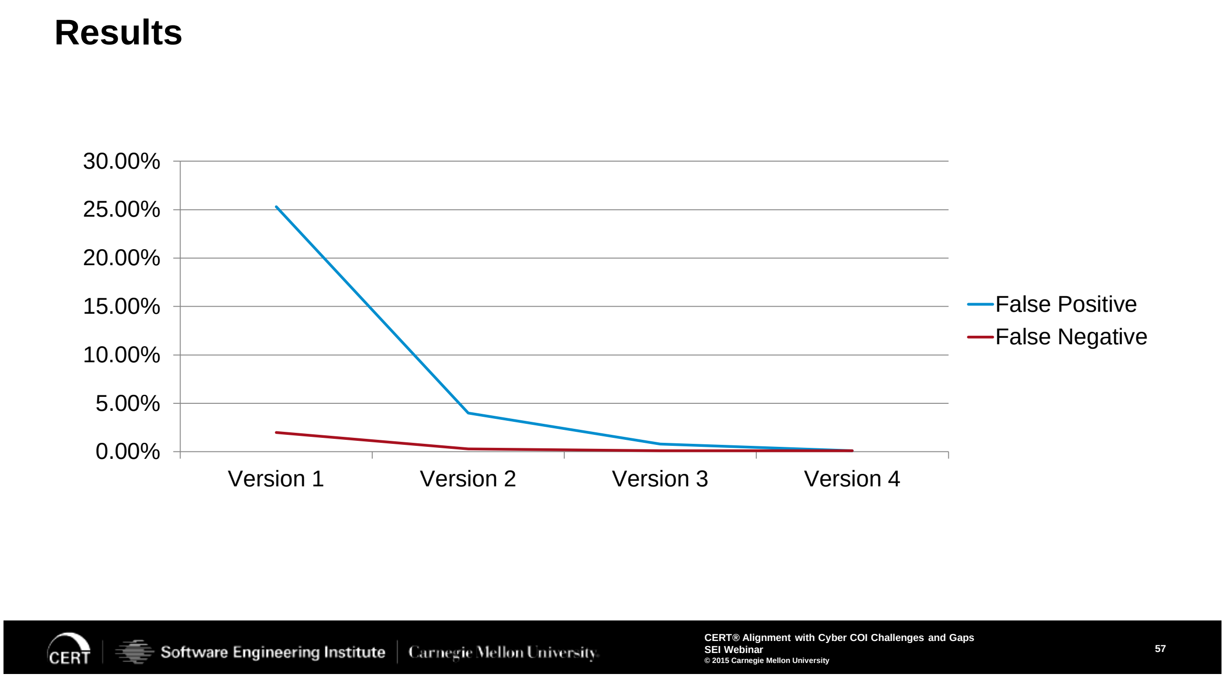
56

**056 Finally, that, layer two of that evaluation module, will take those detections, those activities that they've done, and map them to, "Okay. What knowledges, skills and abilities do these represent?" Again, this development is ongoing, but we've had some good responses from this actually. What we've found is depending on what the knowledge, skill or ability in question is, the correct answer there is the right way to do the analysis is different. Some things, like, for the example I gave before, actually, of knowledge of file carving tools, if you really want to get the deeper knowledge and see, "Okay. Does he really understand how and when to use file carving tools?" what you want to see in the output stream is first any one of a couple of different indicators that might give him a hint, "Hey, it would be worth to go and carve out the files on the system. There's indications here of a PDF file that you might want to look for." And maybe he goes, looks through the file system to attempt to recover it. Can't recover it. And then goes and carves out the PDF files and manages to actually get it from slack space. Great. I see all indications of all that behavior in the stream. Now I know that he's, that he has, actual knowledge of file carving tools. Not just how to do it, but when to do it and why he's doing it. Because he's gone, seen indicators and done it.

On the other hand, that sort of analysis where I'm looking, you know, creating HMMs for detecting that is somewhat heavy handed, unnecessary if I want to look at, verifies--to take another example that I mentioned--verifies the state of evidence. There all I really care about is did he or did he not generate the MD5 value again, and if he did, did he put it in his examiner notes? That's basically all I care about in that case. So for that, we can take a much simpler approach for that type. So what we're doing at the moment is looking at having actually different analysis methods for the different types of knowledges, skills, abilities that we're attempting to detect.

So to give you a little bit of a idea for the results, where are we right now in this work?

Results



Results

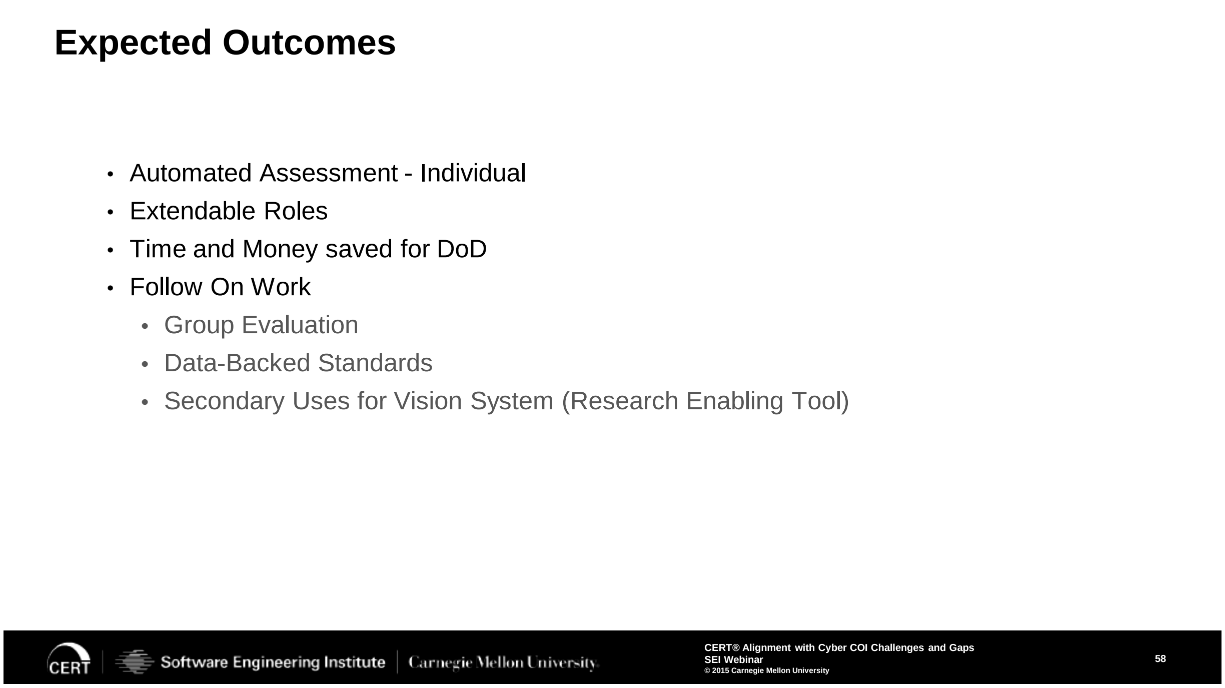
CERT | Software Engineering Institute | Carnegie Mellon University

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar
© 2015 Carnegie Mellon University

57

**057 Well, on the vision system, to begin with, we've had spectacular results. This was where we had really rapid improvement as we took into account essentially, as I mentioned before, the optimizations that we've made, knowing that we don't have to deal with occlusions, we don't have to deal with shadows. And so actually the numbers for version four there are actually our worst-case scenario. We were actually, weren't, able to find a false negative in the data stream in a two-hour capture that we manually reviewed to generate this data set. We actually weren't able to find a negative. So we put one in just to give us some detection. So it's--we're reasonably confident that we're doing a really good job of doing the vision system detections for these environments.

Expected Outcomes



Expected Outcomes

- Automated Assessment - Individual
- Extendable Roles
- Time and Money saved for DoD
- Follow On Work
  - Group Evaluation
  - Data-Backed Standards
  - Secondary Uses for Vision System (Research Enabling Tool)

**058 As far as expected outcomes go for the further development of the system, obviously automated assessment of individual cyber operators and the roles, in the forensic analyst role, that's already being developed. I'm very confident that we're going to have that as an outcome.

Additionally, extendable roles. What we want to do after we've proven the methodology of the system is extend this to all cyber operator roles. So whether you're an instant responder, infrastructure support, whatever the case may be, there's no reason that we can't do the evaluation the exact same way. It's just a matter of training the vision system for any additional tools that you're using that haven't already been used by other work roles.

In addition to that, what we're going to want to do is--sorry. In addition to that, the overhead for actually adding additional roles goes down. As there's more overlap between tools that have been used in other roles, we don't have to retrain the vision system for those roles. All we have to do is for whatever's specific to your role, just train that small amount of tools. Plenty of roles use a text editor. Plenty of roles use a browser. So we've already trained that in for the forensic analyst role. So now when you have somebody in another role that uses those tools, we don't have to retrain the system. All we have to do is retrain the, is extend the training, of the evaluation system and say, "Okay. What are we looking for now as far as knowledges, skills and abilities for this new role?"

Obviously doing it in this manner saves a lot of time and money for our DoD customers. To be able to do that analysis in automated method where they don't have to assign personnel to it. You can basically take the new guy, tell him, "Okay. I've signed you up for this evaluation. Go do your evaluation." He completes it. Couple of hours later you get an e-mail with a report of whether he's mission ready or not. And if he's not, here's the additional training we've already signed him up for.

Follow on work. We'd like to do group evaluation. I don't see any reason why we can't apply these methods for entire teams. Cyber protection teams or otherwise. As well as taking actually a longer-term view. After we've had the system running for a year or two, let's go back and rewrite the standard by looking at the actual performance. Let's look at your top performers and then use the things that really differentiate them from your bottom performers. To actually be, "Okay. Here's the standard for what actually we should be looking at." And we can do that in a data-backed way.

Additionally, all of the different elements that I've talked about, the ACE Vision, ACE Eval, they've been designed to be modular so that we can actually use them separately. Use them as a research enabling tool. So for example, taking the ACE Vision system and saying, "Okay. Now that we can do detections of behaviors of somebody in a restricted environment, let's build another environment." You know, if we want to do insider threat modeling or if we want to do whatever the case may be. We can use that to advance other research by allowing them to capture data sets that they wouldn't have access to otherwise, or at least wouldn't have access to without expending considerable resources.

Q&A

FloCon 2016

Daytona Beach, FL

January 11–14, 2016

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar
© 2015 Carnegie Mellon University

59

**059 Announcer: Okay. Before we jump into our Question and Answer with Rotem, I wanted to remind everybody about the Files tab that's located on your console. You can walk away with a PDF copy of each of the presentations from today, along with other work from CERT on cybersecurity, in the cybersecurity area. Also, upon exiting today's webinar, make sure you fill out that Survey tab as your feedback. It's greatly appreciated.

So let's go to the first question from Ronald asking, "As a systems engineer, I am wondering if there is a collection of typical cybersecurity requirements that might appear in a systems spec or other system-leveled documents so that testing against such requirements can be performed."

Presenter: So I'm not actually familiar with a document that--they may exist. I personally am not familiar with them--out of the material that we've gone through for the roles that we've looked at. But the key thing for this environment is if that document exists we can build to it. So the way that we build our evaluations is we actually start with the document, the evaluation of what we're looking at, and we work backwards from there. So basically we look at the documents, say, "Okay. What are the skills, the knowledges that they need to display?" And then we create a separate document of, "What would show those skills? What would exercise them? What would we be looking for?" And then we'd build a scenario based on that, to build a scenario that forces them to leverage those knowledges, skills and abilities. And so by doing it that way we can actually create that if that document exists for, you know, what would be systems engineering or really any cyber role.

Announcer: Okay. Next one from Shannon asking, "What have you seen in the data, in the data captures you perform, that you did not expect?"

Presenter: Okay. So in the data captures, we honestly, we never cease to be surprised with the data that we can capture. One thing is we anonymize the data, so I don't know who is, a particular capture comes from. But I have seen captures where people have surprised me both with the extremely high level of skill that they've had, where they've managed to skip entire portions of content because they just got it in a trivial manner and we've had to actually make the scenario harder to complete in order to force them to exercise all the individual steps. Or cases where people, you know, didn't have even minimal skills, you know, be able to mount an image. You know, it's something that we just assumed, you know, anybody that would be doing this analysis would be able to. And then all of a sudden we find one capture where somebody spends two hours reading through the manpage for mount, never quite gets it right, you know. And so it's all a matter of being able to have the system deal with all of these cases.

So one nice thing that we've added to the system is actually it has the capability of returning confidence measures on all of its detections. And so when you get this scenario where somebody does something completely new, completely unexpected, the system will actually kick that back to a human evaluator, basically say, like, "Something happened here and I don't know what." You know, "I'm just getting low confidences on everything." And in a two-hour capture, it's not going to do that for the two-hour capture. It'll actually do that for the specific time period that we're talking about here. You know, that the new material is in. Which allows our human evaluator to keep their workload very low. They can just go in and say, "Oh, okay. Well, that's neat, that's new. Now let's add that knowledge to the evaluation system," or to the vision system, if it's some new tool that we've put in. And then we can add that training now that that knowledge has been trained into it. It now knows that for the next time that we get somebody in. If somebody does that trick or that thing next time, great, it'll already detect it, it'll already map it appropriately.

Announcer: Great. Next from Joel asking, "Is there an effort to standardize the toolkit for cyber warriors across the services?" So I know you mentioned STEPfwd. So...

Presenter: Well, STEPfwd, STEPfwd addresses the issues of training, you know. So whatever the toolkit is, we can train to it in STEPfwd. That being said, I believe that there is an effort to standardize that toolkit. I am not personally involved in that effort. I can't speak to that.

Announcer: Okay. Next one. Let's go to Robert asking, "Do you have any experience in linking data from multiple sources? Locations."

Presenter: Oh. So--

Announcer: "And how do you correlate the data?"

Presenter: Okay. So I'm actually really glad that that question was asked, because I forgot to mention something important.

Announcer: Okay.

Presenter: We have a couple of additional sensors on the environment in addition to the vision system. So we don't only rely on the vision system. But the vision system is the primary data source that we use. So to give you an example, that forensic workstation that we have our participants, in they actually have a keylogger on there, for example, that's running throughout the analysis, which generates a text file. We use that actually to shore up the OCR in the vision system. So basically what happens is we take the keylogger with a higher degree of confidence than we take the character detections in the vision system. But so what happens is the vision system will do the detection and, you know, you might get a command, an entire long string that they've run, and then it might say, like, the vision system thinks they ran it with dash D but the keylogger says they ran it with dash C. So it'll trust the keylogger and it'll shore up, it'll fix, that detection if necessary. Pardon me.

What we use in order to make that determination, the way that we pull that data together and associate it, is actually timestamps. So we have very good reliability as far as the timing of these, because we're logging all of them simultaneously. And so that's actually what we use in order to match up the timestamp from the keylogger data with the detections in the vision system. As I mentioned, we're doing it at 15 hertz, so we have a relatively high confidence with exactly where in the stream we are for any given detection.

Announcer: Okay. So one more question in Q here for Rotem. So if you have a question, feel free to type it in now. From John, asking, "Would the ACE system be usable for other roles, other than the cybersecurity-related positions that you've highlighted for us?"

Presenter: Yeah. So ACE is usable for a broad variety of roles, but I don't want to overstate it. I mean, there's certain roles that require a creative element where there's no necessarily absolutely correct answer, right way of doing things, to give you, you know. I wouldn't want to use ACE to evaluate somebody's capabilities in web development, for example. You know, there's a lot of ways to write code that could be correct and I'd rather use a system that's going to evaluate their output, you know, evaluate their code if I'm going to do that. Whereas if I'm looking at somebody like an incident responder, where I know what they need to do. They need to go to, you know, detect here's the intrusion. I know that they need to find, you know, this machine, they need to remediate this infection, whatever the case may be. Or I can actually say, "Okay." Here's the actions that correspond to doing their job correctly." Those types of roles are roles where I think ACE would be appropriate and could have really good results for it.

That being said, you're absolutely right. They don't necessarily need to even be cyber roles. If I have somebody that's, you know, in a non-cyber role but has activities that they need to perform on a computer, be it in healthcare or, you know, they need to login, you know, billing or whatever the--HR or whatever the case may be. If they have to perform tasks in a digital environment that we can replicate and step forward, then there's no reason that we can't do that.

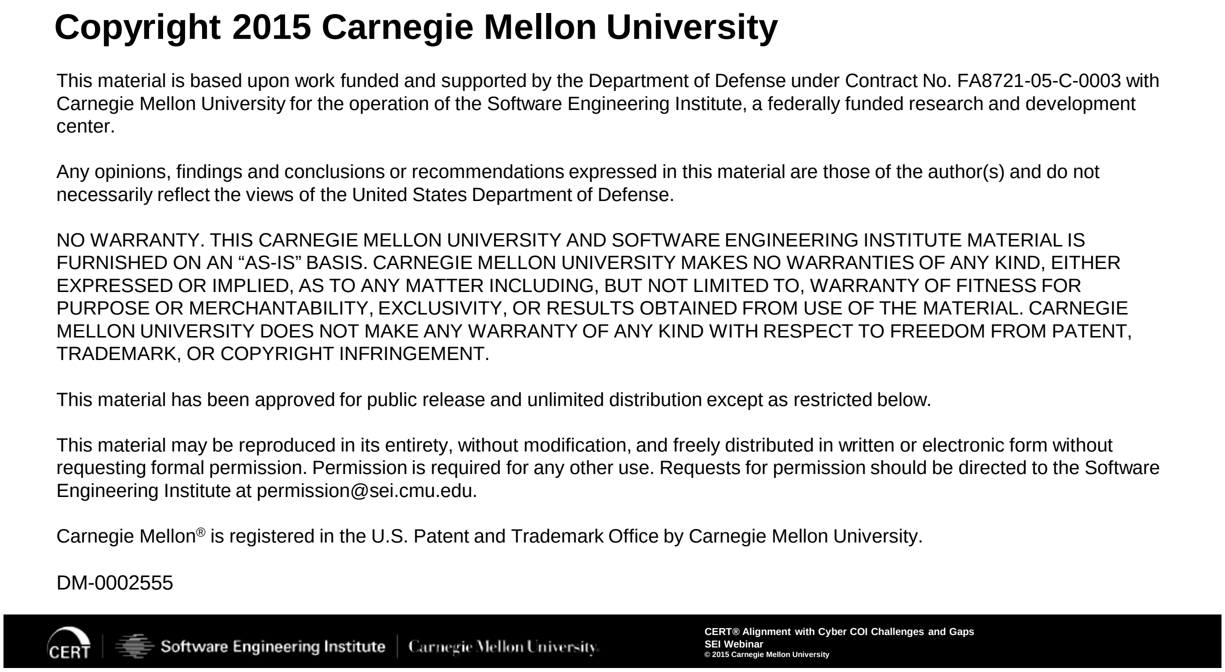
One question that hasn't been asked that I've been asked a lot of times that I just want to make sure to note is, "Why are we doing vision?" Because as I mentioned, we have a keylogger on there. Why don't we just instrument the environment? And one of the key things that using vision gives us is it gives us the ability without having to re-instrument the environment to move to any number of systems. The way that I like to put it, the way I've always explained it, is every application that we've written for decades now has the same API and we didn't notice it. And that API is two eyes, two hands. You know, we all use it, we all interact with the machine.

Well, having the two hands, that's easy. I can send keystrokes and mouse movements and, you know, I've always had that capability. It's the two eyes that we're missing. And so by adding that vision system capability, we can go in and do an evaluation on basically any system without re-instrumenting anything at all. And the example I always like to give when people say, is like, "Why don't you instrument the system?" Well, that's great when I'm on a Windows workstation. It's fine when I'm on a Linux box. What happens when I just jumped into console access on a Cisco router, for example? Now what am I going to do in order to track what they're doing? Well, I could instrument that as well, but now that's, again, I have to go engineer that. What happens when I go to a Juniper router now, you know, or any number of other manufacturers? Well, if I'm doing it with the vision system it's still, it's the same thing. It's just going to look different. I just have to train the vision system with those new templates. What does this system look like? And there it's the same process no matter what the system is. It's a process that isn't a lot of effort for us to do in order to train the system for that, because the methodology already exists.

Announcer: Excellent. Rotem, excellent presentation. Folks, that's about--we're going to wrap up this presentation here now so we can get a five-minute break to prepare for our next presentation, which will start promptly at 3:05. And that talk is going to be, "Using DidFail to Analyze Flow of Sensitive Information in Sets of Android Apps," by Lori Flynn and Will Klieber.

So we'll see you back at 3:05. Thank you.

Copyright 2015 Carnegie Mellon University



Copyright 2015 Carnegie Mellon University

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Defense.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN “AS-IS” BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

DM-0002555

CERT® Alignment with Cyber COI Challenges and Gaps
SEI Webinar
© 2015 Carnegie Mellon University