

Department of Defense Bloggers Roundtable With Dan Kaufman, Director of the Information Innovation Office of the Defense Advanced Research Projects Agency (DARPA); Norm Whitaker, Deputy Director of the Information Innovation Office of the Defense Advanced Research Projects Agency (DARPA) Subject: DARPA's Shredder Challenge Time: 2:00 p.m. EST Date: Monday, November 21, 2011

Copyright (c) 2011 by Federal News Service, Inc., Ste. 500 1000 Vermont Avenue, NW, Washington, DC 20005, USA. Federal News Service is a private firm not affiliated with the federal government. No portion of this transcript may be copied, sold or retransmitted without the written authority of Federal News Service, Inc. Copyright is not claimed as to any part of the original work prepared by a United States government officer or employee as a part of that person's official duties. For information on subscribing to the FNS Internet Service, please visit <http://www.fednews.com> or call (202) 347-1400

PETTY OFFICER WILLIAM SELBY (Office of the Secretary of Defense for Public Affairs): Roger that. I'd like to welcome you all to the Department of Defense's Bloggers Roundtable for Tuesday -- I'm sorry -- for Monday, November 21st, 2011. My name is Petty Officer William Selby with the Office of the Secretary of Defense Public Affairs, and I'll be monitoring the call today.

Today we are honored to have as our guests Mr. Dan Kaufman and Dr. Norm Whitaker, the director and deputy director of the Information Innovation Office of the Defense Advanced Research Projects Agency, DARPA. They will be speaking on DARPA'S Shredder Challenge.

And a note to the bloggers on the line, please remember to clearly state your name and organization in advance of your question. Please respect our guests' time and keep your questions succinct and to the point. And if you are not asking a question, we ask that you place your phone on mute.

With that, Mr. Kaufman and Dr. Whitaker, if you have opening statements you can go ahead with those now.

DAN KAUFMAN: Yeah. This is Dan Kaufman. I'll start real quick and then I'll kick it over to Norm. I just -- I assume you guys are all sort of plugged in and know what the Shredder Challenge is, but basically the idea was to take five documents, shred them in different ways. And we were curious whether people could put it back together,

particularly looking at it inside a sort of a -- in a military- relevant way, meaning, you know, can you do it fast enough and cheap enough and, you know, to be within sort of a mission time frame.

And probably the biggest news, at least for us, is we found out just today that the fourth puzzle has been solved, so there -- one remains. To give you some of the details about the Challenge, I'm going to pass it over to Norm.

NORM WHITAKER: Hi. This is Norm Whitaker. So as Dan said, the Shredder Challenge was started back in October. It's been actually a real well -- (inaudible) -- and it's exciting to continue with the announcement of the fourth puzzle being solved.

DARPA, as you may know, is the -- is the single joint R&D organization for the Department of Defense, and our job is to -- (electronic interruption) -- is to (prevent strategic problems ?) basically. And so we looked at the Shredder Challenge, this idea of potential vulnerabilities having to do with shredded documents as something we just wanted to look into, to understand, you know, what was vulnerable.

We as an agency have a long history of engaging in crowd-sourcing-type events. The DARPA -- the Grand Challenges and the DARPA Urban Challenge and, as another example, the Red and Blue Challenge that we conducted a couple of years ago were all examples of times when we reached out to a wide variety of people in the public to see, you know, who out there has a great idea for solving a difficult problem. So this is no different than that, and we've been utterly pleased with the results so far.

To take you into the details of the event, as Dan pointed out, we are looking at just shredding a relatively small group of documents. So this is not a trash bag full of shreds. This is something that takes between, say, two and 20 pages to lay the shreds out, which is what we've done for people, and we put it up on our website for downloading. We've invited a team, including computer scientists, people who are good at jigsaw puzzles and anybody else who is just curious to take it on, and so far we've had more than 8,000 registered teams try and do this.

As you might imagine, the five challenges we put forth are graded, in the sense that the first one is the easiest and the fifth one is the hardest. And so we saw almost immediate returns on the first puzzle, within a few days. Some real clever jigsaw puzzle fans, I think,

are the ones who stepped up. And we've seen quite a few people able to solve for the one.

As we get into the more difficult puzzles, we're seeing a combination of techniques in use. The (hand ?) solvers, the jigsaw puzzle solvers, have had a lot more difficulty with, you know, with thousands and thousands of pieces. And we're seeing automated approaches -- people who are using computers to divide up the images (into ?) shreds and align them and suggest (actually ?) some of the pieces that a human doesn't take a look at, and verifies or basically trains the system as you go to become better and better at solving problems.

We're also seeing a -- at least two crowd-source efforts. These are, if you can imagine, a computer program where hundreds, say, of people are simultaneously trying to solve each puzzle by putting pieces together. And that looks like -- of course we had the real potential to -- (look back ?).

So of the top five, the top five teams on the leader board currently, three of them are automated approaches or semi-automated, with a human involvement. One of them, as far as we know, is a by-hand approach and the fifth one is the crowd-source approach.

DARPA is very pleased with what we've seen from this. We've got about 13 days to go before the event is over, so it looks like it's going to be a horse race up to the end among the -- sort of the three competing approaches to solving the Shredder Challenge.

We are -- we can take any of your questions now.

PETTY OFFICER SELBY: Roger that, sir. And two other people had joined while we were on the call.

Q: Yeah. This is Richard Walker. I came in slightly late and I'm wondering who was just speaking, if you could identify yourself.

MR. WHITAKER: This is Norm Whitaker from DARPA.

PETTY OFFICER SELBY: And before that, that was Mr. Kaufman, Mr. Dan Kaufman of the DARPA.

Q: And Mr. Whitaker's title is what?

MR. WHITAKER: I'm the deputy of the Information Innovation Office.

Q: OK. Great, thank you.

PETTY OFFICER SELBY: I can send all that over to you too, Richard.

Q: Excellent. PETTY OFFICER SELBY: And just for your -- so you know, you're seventh on the line, and we're going to take questions in the order that people came on line. And if you're not asking a question, we ask that you please place your phone on mute.

Chuck, you were first to be on the line, so you can go with your questions.

Q: Good afternoon, gentlemen, and thank you for taking our call. Chuck Simmins from America's North Shore Journal.

What obligation do the people participating in this challenge have to share their methods with DARPA? And is there any potential for the winner or some of the other challengers to possibly get additional research funding for their method in the future?

MR. WHITAKER: This is Norm Whitaker. I'll provide the answer to that.

The participants, when they submit an answer, also submit a -- basically a picture of the page that they reconstructed so that, you

know, they are in fact using the data, not finding it on line or something like that. And they are providing us with a short review of how they solved it.

So the obligation -- there is an obligation to tell us how you did it. We've found that in practice, people are happy to talk about it. Usually they've done something very, very clever and share freely.

Now, we as an agency do fund lots and lots of research projects. We don't have a solicitation open for something like this that we -- if there's a great idea out there, we'd be happy to hear about it. You know, there's always a -- we're always looking for another great idea.

Q: Thank you.

PETTY OFFICER SELBY: And Katie (Drummond ?). Katie, are you still there? We'll go on to Tim then. Tim, are you available still? Jared (sp).

Q: I guess I'm wondering, is the point of this more to sort of scare it out and identify, you know, ineffective shredding techniques that are being used by us today, or is it -- is this more of an offensive capability that we're trying to develop that might let us decode messages that have been shredded by U.S. adversaries?

MR. WHITAKER: This is a sort of a first crack. You know, we're an R&D organization, and what we do doesn't always directly relate to a product or it's not always a direct analysis of a capability. So this first look is trying to understand what's involved with a relatively -- I want to say easy -- unshredding task. Because there's not that many pages involved and it's -- they're handwritten documents that actually make it tougher.

So, you know, on the surface we are just getting our feet wet, if you will, in this problem, see where this goes in the future. We think, though, that this is the type of study of, you know, information security issues having to do with the war fighter is important and (essentially ?) help us understand what might be possible in the future.

Q: OK, but again, is -- what does DARPA hope to have at the end of this project, once you're all done? What's the outcome we're driving toward?

MR. WHITAKER: DARPA in general gets a lot of things out of these kinds of challenges. We identify really, really clever people. We identify new approaches, and there'll be software developed, but won't be developed through us. DARPA won't take any rights away to the software that's developed. But we'll have identified the people who know how to solve problems like this.

Q: OK. Thanks.

PETTY OFFICER SELBY: And Neil Savage?

Q: Hi. Neil Savage from Communications of the ACM. I'm wondering what is it about this problem that confounds computers, that people can bring to it? And are you hoping to find a way to teach a computer to be better at this sort of thing, based on some of the things you might learn?

MR. WHITAKER: The -- once you get into this problem you realize that there's probably a lot of different domains and there are different clues that one might use to solve it. For example, if each shred contains multiple letters, that's a different problem than if -- shreds containing only a piece of a letter.

So in addition, any kind of information that you can pull off a page, like if there's a gradient along the page or that perhaps the paper was different at one end than it was at the other, or there was a - - (inaudible) -- (formulating ?) as the copy was made before it was shredded, those kinds of things can provide real -- are huge clues for getting to a solution. So actually, understanding which of those components, if any of those, are comparable to other (traits that people -- are used ?) are the most important is part of the goal here.

If you know exactly what was the most important (in any given case ?), then probably you could start developing code and get the problem solved.

Q: But is it the human visual ability as compared to what computers can make of images, or what are the -- what are the human capabilities here that elude computers?

MR. WHITAKER: I think that's a really interesting question that we don't necessarily know the answer to. We may know something more than we know now at the end of the challenge, but that certainly is just sort of a potential intellectual question that's being asked.

MR. KAUFMAN: One thing -- you know, one thing I'd add that humans tend to -- seem to do better than computers, although it's obviously an area of research -- (inaudible) -- contextual understanding, right? And so rather than having to -- you know, machines are very good at doing things very, very rapidly (but ?) exploring the entire design space -- (inaudible) -- and, you know, notions of intuition and notions of being able to pick up context that humans are particularly strong at. And the question is where does that lie, right? Obviously some things machines do better and so we're seeing interesting things, right?

A machine will move faster than an individual human, but now we're seeing ways of using crowds of people. So as Norm says, and I think he's exactly right, this is exactly why we hold these challenges, so that we can sort of understand what's going on and where those lines lay.

Q: And just -- the other part of (this ?) is, so were you hoping that what you learn from this, aside from the specific challenge of trying to recreate shredded documents, is something you could perhaps use as a basis for developing new computer algorithms that can address whatever it is that the humans are addressing?

MR. WHITAKER: You know, both visually, but in general, the problem of piecing together lots and lots of pieces of information that are -- you're not really sure how they're related, and trying to find order in that, that's a real generic and interesting problem that we have to solve all the time. So there's a visual problem, as you're pointing out. And we do hope just to learn things.

But we also are looking just more generally, piecing together this patchwork quilt of pieces, that it's hard to see the connection between those till you put them up next to each other. Q: OK. Thank you.

PETTY OFFICER SELBY: And Corinne?

Q: Hi. Corinne Reilly, from the Virginian-Pilot. I just want to make sure I understand this right. So it's not five documents but five challenges, and there's a varying number of documents in each challenge?

MR. WHITAKER: That's correct. Q: OK. What's the range? I don't know if you can say.

MR. WHITAKER: I think, shredded, the puzzle number five is about 20 pages. So that doesn't mean it's a 20-page document, but when you lay out all the little shreds, whoever solves that will see that's how big it is.

Q: OK.

MR. WHITAKER: The first -- document is pretty simple.

Q: OK. And at this point is it too late for a new group to join and start competing, or if they solve the final challenge, could they still win at this point?

MR. WHITAKER: Somebody could start right now and win. They could win the whole thing, but they've probably got a really great computer program or a huge crowd of friends to help them.

MR. KAUFMAN: (Inaudible) -- (pretty clear ?) to help you. So the fact that somebody solved puzzle one doesn't mean that you can't get points for solving puzzle one.

Q: OK, so you could log on now and start from the beginning. You'd just obviously have less time. OK, that's it.

PETTY OFFICER SELBY: And Richard.

Q: Richard Walker, from AOL Government.

It seems as if your ultimate objective is a fully automated, agile technique or system for reconstructing shredded documents in the field and doing it very quickly. Is that -- is that correct? Does that sum up your objectives?

MR. WHITAKER: If we could get that, that would be really great.

Q: Right. So what have you learned from the -- from the manual technique so far? Because it seems as if, you know, they're awfully time-consuming. And what -- you know, what purposes have they served toward your ultimate objective of automation? MR. WHITAKER: I think the manual technique -- (inaudible) -- a baseline where, you know, sometimes manual's actually faster. In fact, the first -- (inaudible) -- were solved that way.

Q: Oh, really?

MR. WHITAKER: And you don't always know what you're getting when you -- if you were to just recover a bunch of shreds, you don't know what the nature of the document is. Particularly the sort of a lower (boundary ?) of how long it would take to get a complex -- (inaudible) -- put together.

Q: OK. Thank you.

PETTY OFFICER SELBY: And was there anybody on the line -- we've been through the first line of questions. Is there anybody that didn't get to ask a question?

Q: Hi. Jennifer -- (inaudible) -- from CNN. I dialed in late, so I wasn't part of this group.

But I had a question about the demographics of the participants. And I apologize if you've already -- (inaudible) -- at the beginning. But do you have any sort of way of telling, you know, where people are from, (and if it's ?) global, is this, you know, people from the United States, ages, anything like that?

MR. WHITAKER: So we don't collect that information directly. But from looking at email addresses and so on we do know that we get participants from all over the place, from all over the world. And we also -- we sort of can tell if we're getting lots of university participants. There's a lot young people with great ideas who are out there who have participated in this in one form or another. So we're really excited by that.

Q: Great. Thank you.

PETTY OFFICER SELBY: And back around to you, Chuck.

Q: I've got nothing further to ask.

PETTY OFFICER SELBY: Katie, are you still on the line? Tim? Jared (sp)?

Q: I guess the only thing I'd ask is, have you at this point been able to identify specific shredding techniques that are more easy to -- that are more easy to decipher than others? Is that -- is that part of the difficulty in the steps between one and five, or do they vary throughout them?

MR. WHITAKER: So there are -- you can Wikipedia and see that there are multiple levels of shredders. I think we go up to a level four, and of course the more completely you shred, the harder it is to put it together. That's not a surprise. I think that the difference in solutions, though, is going to be where the excitement is.

I know there's a specific team -- (I think ?) UCSD. You might follow them up. They have a -- the crowd sourcing that they're doing is -- they just started a few days ago. But things are moving so fast that -- (inaudible) -- they have an interesting story to tell.

Q: All right. Thanks.

PETTY OFFICER SELBY: Thank you. And Neil?

Q: Yeah. Could you say what other sorts of things you're trying to approach with this technique of crowd-sourcing? Particularly, you know, it's a widely used term and I've seen it for people trying to come up with advertising campaigns and such, but to solve the types of problems that, you know, you'd like to solve but that you can't really with computers, are you doing a lot of these? And then what are some of the other examples of them?

MR. KAUFMAN: Yeah. So -- this is Dan. Let me ask you sort of two things. One, we've done -- one we're in the process of doing and one we just started, right?

Q: OK.

MR. KAUFMAN: So probably the one we did out of DARPA is something called Foldit. Are you familiar with Foldit?

Q: Yes, I am. I've talked to --

MR. KAUFMAN: OK. So Foldit came out of DARPA. I was the program manager -- (inaudible) -- and so that obviously is doing -- (approaching design fold ?). We're now taking that and we've released a new program called Engage where we're looking at that, using it for STEM education so we find better, smarter ways to tutor our children. Because we all talk about this need to, say, improve education.

And then if you want to get to the really obscure and slightly odd but interesting, we're interested in formal methods actually being able to prove that code is formally correct, which is not something you would ever think of as amenable to games or to crowds. But we've got some early research that indicates you can do that. So that -- (inaudible) -- crowd is formidable.

You have to -- as always, it won't -- it's not a panacea. It won't solve all problems for all people, but we are continually sort of impressed, pleased and interested in the breadth of problems that can be formulated in a way that's accessible to a crowd. Q: I don't want to take up a lot of the conference's time, but is there a way I can find out more about that last one you mentioned?

MR. KAUFMAN: Yeah. The formal DARPA --

MR. WHITAKER: There's a press release that's about to go out.

MR. KAUFMAN: Yeah. We're going to put a press release out, I think in a few days, and then it will -- it'll actually key you to the -- (inaudible) -- office where, you know, where we call people to come and do it, and there should be a fair amount of detail in there for you.

Q: OK.

MR. WHITAKER: One thing that we've learned from crowd sourcing is that the people out there are incredibly smart and -- incredibly capable -- (inaudible) -- are able to -- (audio break).

PETTY OFFICER SELBY: Sir, are you still there?

MR. WHITAKER: Yes.

PETTY OFFICER SELBY: OK.

Q: Hello?

MR. WHITAKER: Yes, hello?

PETTY OFFICER SELBY: OK. (Laughs.)

Q: OK, I guess I finished that line of questioning.
(Inaudible)

PETTY OFFICER SELBY: OK. Corinne? (No audible response.)
OK, Richard, are you still there?

Q: Yes, I am.

How long has DARPA been using crowd sourcing as a way of seeking innovation? MR. WHITAKER: So this actually began with authority that was granted by Congress, and it started back in 1999 -- (inaudible) -- used in 2003. So the Congress and the White House have both been extremely supportive of this approach to trying to get more and more people involved in solving some of the hard problems.

Q: It just seems like it's really taking off in government now. Other agencies are using it.

MR. : Other agencies are using it, that's right. It's been very, very effective.

Q: Thank you.

PETTY OFFICER SELBY: And Jen?

Q: I'm good, thank you.

PETTY OFFICER SELBY: OK. Did anybody else have any more questions?

Q: Can I try just one more?

PETTY OFFICER SELBY: Yes. Who's this?

Q: Neil Savage.

PETTY OFFICER SELBY: Oh, yeah. Sure, Neil. Go ahead.

Q: Just wanted to ask, are there any sort of best practices you're starting to find out about in designing this sort of crowd-sourcing problems with what you need to take into account to make this work?

MR. WHITAKER: So I would say that the -- I think the agency has learned a lot, but it's basically the things that we would learn from a commercial industry who have lots of experience working with crowds and understanding what motivates people.

We found in general that the people in the United States were extremely supportive of projects that support the war fighter, and we have lots of people who watch what we do and sign up strictly for that purpose because they want to be involved in things that, you know, advance technology that helps protect people on the battlefield. So as

an agency and as a department we've been extremely lucky in all the support that we get.

Our challenges, we try to design them, you know, logically. We make sure that the incentives are clear and that the rules are easy for people to understand. We generally look for things that are easy to judge so that instead of a beauty contest where we try to rate on some sort of unquantified grounds, we look for things that are quantifiable. So in this case we have a six-point system that we use for the puzzle, so that everybody knows exactly how many points they've got depending on what they're able to achieve. So I think that that's -- sort of sums up our approach.

MR. KAUFMAN: Yeah. I'd say there are a couple things that -- (inaudible) -- is you want something transparent. When you're with the government I think it's really important to make it transparent and incredibly fair to people so that they understand what it is. And then we do work hard to make sure that it's successful.

And part of it is not just about solving the problem, but it's getting people generally interested in science and technology and getting people excited and understand that they can be a part of it. So we try to craft things that even though, as in this case, where -- (word inaudible) -- is quite difficult, we try to open it up to do --

And then the last thing is -- and it's difficult, but it's important -- is to not describe the answer. So there's a temptation often to think, oh, well, this is a smart way to do it, and you over-constrain the problem. And one thing we've learned about the crowd and the creativity, particularly of the U.S. population, is we don't -- (inaudible) -- stating what the problem is clearly and sitting back and allowing people to solve it in ways that frankly we would never have thought of them in solving it. And so the most important thing to do is get out of the way a little bit and let them do what they do well.

Q: OK.

PETTY OFFICER SELBY: Is that good for you, Richard?

Q: Was that Dan just speaking?

MR. KAUFMAN: Yeah.

Q: Just want to be clear on that. When do you expect to issue the press release? Will it be by the end of this week, or by Wednesday before the holiday?

MR. WHITAKER: We think we're going to get it out tomorrow.

Q: Oh, actually? OK.

MR. WHITAKER: Tomorrow on the DARPA website.

Q: Very good. Thank you.

PETTY OFFICER SELBY: Thank you -- sorry about that.
(Inaudible) -- did you have a follow-up? Well, I guess he's dropped off the line.

So with that, I'd like to thank everybody that's has joined today and that participated for your questions and for your comments. Thank you, Mr. Kaufman and Dr. Whitaker.

Today's program will be online on DODlive.mil and where we'll have a transcript of the call as well as an audio file and a short story. Once again, I'd like to thank everybody for your time.

And sir, Mr. Kaufman and Dr. Whitaker, if you have any closing comments, you can go ahead with those.

MR. KAUFMAN: I just -- just thanks and appreciate you guys covering this stuff. I think the more the word gets out, sort of the more responses we get, and it's just fun to watch. So thanks.

MR. WHITAKER: It's www.shredderchallenge.com. Thank you.

Q: Thank you very much.

Q: When do you expect the transcript to be --

PETTY OFFICER SELBY: That can be anywhere within three hours to 24 hours, but if you check on DODlive.mil and click under the blogger link, you'll see the link to this roundtable.

Q: Excellent. Thank you very much.

PETTY OFFICER SELBY: And I can send you a -- I'll send you an update if we get it earlier than that.

Q: Super. Thank you.

PETTY OFFICER SELBY: Thank you, everybody, on the call. (Inaudible) -- today's call. Please feel free to disconnect at this time.

Q/MR. : OK. Thank you very much. Appreciate it. END.