December 11, 2018 Community Telecon

Interface I1 - Endpoint Data Collection

Presented By - Charles Schmidt (MITRE)

Charles Schmidt started the teleconference by asking the composition of the audience.

* Who creates products that are part of an enterprise security solution (tools, etc.)?
	+ David Ries (jOVAL), Bill Munyan (CIS), Josh Lubbell (NIST), Andreas Steffen (strongSwan), Jarrett Lu (Oracle), Jack Vander Pol (SPAWAR)
* Who wants security automation to be part of your enterprise solution?
	+ Debra Gentile-Granat (???), Jessica Fitzgerald-McKay (NSA), Tyler ??? (???), Jerome Athias (???), Jarrett Lu (Oracle), Josh Lubbell (NIST), Bill Munyan (CIS)
* Who is an application/OS vendor that would like to use security automation?
	+ Jarrett Lu (Oracle)

Charles Schmidt explained how this talk was focused on data collection from an endpoint and transporting this data to a Posture Collection Server in a standardized way where it can be stored and analyzed. In this proposal, endpoints can self-report and the Posture Collection Server provides some orchestration capabilities. The Posture Collection Server is the point through which endpoints do that.

Charles Schmidt explained the main reasons behind the SCAP v2 vision for endpoint collection.

* Scale collection on an endpoint. Lots of collectors that can leverage the same underlying infrastructure.
* Scale analytics in an enterprise without overloading endpoints with redundant agents.
* Single point of management for endpoint collection (i.e. Posture Collection Server). This avoids each tool having to contact each endpoint for data. Useful when considering across all endpoint types.
* Support timely awareness of endpoint state.

Charles Schmidt asked if people had questions or concerns about this vision statement.

* Tyler ??? - I would prefer an agentless solution.
* Charles Schmidt - To the extent that you can get endpoint information from an agentless solution, that would be preferable. However, until endpoints are exposing native capabilities, that will be difficult. We would like agentless solutions to be supported.
* Blair Heiserman - I would argue that an agent-based solution will return more data.
* Charles Schmidt - I agree with both Tyler ??? and Blair Heiserman and both would be in scope.
* David Solin - This is a very ambitious vision. From my own experience, working at a large software company that does a lot of M&A, we would have never been able to get all of our different agents to interoperate. So, it seems like it will be very hard to get clients and servers from different vendors to interoperate with each other. For all of our product integrations, we did them at the database level with Extract, Transform, Load (ETLs). It was the only way to do it financially, politically, etc.
* Charles Schmidt - Can you expand on that? Is the data normalized?
* David Solin - The data would live in each tool natively and then we would have a single tool that could access the data. Or, there would be ETLs between the individual databases. We had this vision that we wanted to supply our customers with all of our products. But, in reality, customers would only have some products, so everything had to be possible, but, optional. We wouldn’t have been able to do that, and this is just one organization. I also find companies like Microsoft build multiple things because they are so large and at a component level, they can’t see the reuse possibilities. So, my concern is the idea that we are going to collect data from all the endpoints using a single interface, I think it is something that will be very difficult.
* Jessica Fitzgerald-McKay - I think what you are saying is you are needing an endpoint architecture that is pluggable and scalable? Or, is it not a technical issue?
* David Solin - One problem is people assumed data centers and systems were always on. Others did other things. They were just different.
* Jarrett Lu - To achieve some level of interoperability for SCAP, you would need a mandatory set of things that an endpoint would support and there would be extensions and an endpoint would just say I don’t support that. There would be a set of things an endpoint must support to be SCAP v2 compliant and the collector has to be aware that I may or may not support that type and be prepared for that.
* David Solin - Another possibility is to have an SCAP gateway and you can query this service/gateway and say you want information about a particular endpoint and how fresh it needs to be and the product could decide how it is going to put it together for you.
* Charles Schmidt - You are absolutely right. This is ambitious. It would be nice if there was a single command-and-control framework, but, I don’t think we are going to see that, but, I guess the question that I got for David Solin and others, is there a degree to which we can work towards that vision maybe with some key datatypes? The whitepaper focused on software inventory. Is there an opportunity to standardize around that for now and then increment over time? Would that be worthwhile?
* David Solin - Any small thing that you can get adopted by everyone is a big thing. For example, if every endpoint was able to understand the asset identification schema and then you could query it and get that data that would be enormous. Maybe in version three, we could get endpoints to speak this language we got. Then, you would have adoption and interoperability solved and pushing down to the endpoint would be more conceivable.
* Charles Schmidt - You are saying it would be good to get common ground on defining a common structure and semantic for this information.
* David Solin - I think so, but to go a little further. In the architecture diagram you have there, is an endpoint between the endpoint and Posture Collection Server? It might come from a different vendor. I think that is a bit much to ask for when there are other interfaces on the server side.
* Charles Schmidt - If there was an agent on the endpoint and the Posture Collection Server then it is likely that the connection between these two are to be a single vendor and we should focus more of our work on the other interfaces.
* David Solin - Yes, that’s my professional opinion.
* Blair Heiserman - We need a well-defined data format because most tools can pull back the same information, but, everyone formats it differently. Then, it doesn’t matter which tool you pull it from as long as you have that format. CCE was big in doing that for us and giving us a standard way to talk about the configuration setting. Then, the vendor of the platform or application is they can write it using their own capability. Everyone just translates, and you might not get back what you need, but, if you give them a single format of what you want, they can do whatever they want as long as they give you it in the correct format.
* Charles Schmidt - So, if we had a way to express the data on these endpoints and normalize this understanding then that would be a huge win and we don’t have to do all the translations between products.
* Blair Heiserman - Right now, everyone talks about CVE. Some of the mechanics differ, but, pretty much the same things. If you do an installed software list, you can say it must include product name, organization, product hash, etc. Everyone is going to do it differently, but, if they give you that data, you can be in a good spot. In some cases where you have commonality on the platform, you could talk about password policy compliance rather than having a CCE for each version of Windows because luckily Microsoft has kept it the same across multiple versions of Windows. Do you meet the password requirements yes or no?
* Charles Schmidt - To product vendors, to what extent can we say we are comparing apples-to-apples if we are looking at normalized structures, but, disparate mechanisms by which the data is collected? Is this problematic for tool vendors? Back when writing OVAL definitions, if checking the registry versus WMI, they didn’t always reconcile. Even if talking about different data and letting vendors go about data collection independently, is that a concern?
* Blair Heiserman - I think the collection is really the first step. There are now many ways for that interface to work, but, what comes to me listening to this conversation is the Posture Collection Server is a posture distribution server and there is still a place to do rationalizing of data and the data needs to be correlated to get insight across all the data. I think there is still value even if not all of I1 is standardized.
* David Solin - With validation programs, I think we have a good handle on if products are collecting the correct data. In order to be an SCAP validated product, you need to run all this content. I think the program does a good job ensuring all the vendors are processing the content in similar ways.
* Charles Schmidt - So, if we could standardize around the data we care about and then expand on the validation programs to ensure we are collecting the right data, that would be sufficient to make it possible?
* David Solin - I think it would.
* Blair Heiserman - I would hate to throw in the normalization of comparing like-to-like that kind of defines your configuration. Comparing the registry and WMI, one is the correct comparison, but, how you pull the WMI data may vary (PowerShell script, custom tool, etc.) and there is one correct way to collect data and query the system.
* Charles Schmidt - Any other comments or thoughts on this discussion.
* Tim Jones - I would like to see this as a standard that multiple tools would support.
* Jessica Fitzgerald-McKay - Part of the vision is that tools can be swapped out and communicating amongst themselves without there being an administrator in the loop. Is this something we want to achieve with SCAP?
* Blair Heiserman - Yes. Some form of being able to pull data from a tool or a central location point. I’m also fine with a tool dropping a file on an endpoint that I can query and read. I want to be able to provide an SCAP file to a tool and know the data I get back in a good format. How it does that doesn’t matter so much. If people want to provide more access, it becomes more a selling point for the tool itself.
* Charles Schmidt - So, your key value is the exposure of information with APIs (standardized or not) and standardized structures of representation. Those are the keys to tool interoperability you are seeing.
* Blair Heiserman - Right now, I am getting data from different tool sets and have to make it so it works with other tool sets (e.g., CMDB, other efforts, enterprise visibility management, etc.).
* Charles Schmidt - When I say a standard structure, I mean syntax and semantic. This goes to David Solin’s point that a validation program would ensure some common ground truth and understanding of the meaning and structure of data.
* Blair Heiserman - Agreed.
* Charles Schmidt - So, if I am following correctly, the normalization of an infrastructure between an endpoint and Posture Collection Server, people are not seeing a lot of value, but, people are seeing a lot of value in a common way to describe endpoint data as well as access to the data through the Posture Collection Server and having a shared understanding.
* Tim Jones - Are you envisioning something that looks like OpenC2 with SCAP data as part of the content we would like to share?
* Jessica Fitzgerald-McKay - I would like to see SCAP results inform how OpenC2 actions are carried out.
* David Solin - We naturally need validation from end user types that this would deliver the value they are seeking from an SCAP 2.0.

# Question: End users - How many different vendors are you using in your security space (order of magnitude)?

* Blair Heiserman - I am currently running 5 tools. It’s too many with massive performance applications.
* Josh Lubbell - +1.
* Brady Alleman - > 3 vendors.
* Tyler ??? - +1.
* Blair Heiserman - Yes, more than that in the enterprise, but, less on the enterprise itself.
* Charles Schmidt - More than a dozen?
* Jessica Fitzgerald-McKay - +1.
* David Kennel - +1.
* Charles Schmidt – David Kennel, how big is your enterprise in number of endpoints?
* David Kennel - 40k+ devices, 20K+ endpoints.
* Charles Schmidt - For everyone, why don’t you use more vendors in your solution?
* Blair Heiserman - Most products overlap significantly. Because they overlap, they tend to cause performance conflicts.
* Tyler ??? - Maintenance responsibility (of agents).
* David Kennel - TCO (total cost of ownership).
* Jessica Fitzgerald-McKay - The number of tools we would need is prohibitive with cost and just maintenance. It also creates gaps because you need to open ports, etc.
* Josh Lubbell - Since my use is research/experimental, I favor free or open source.
* Jessica Fitzgerald-McKay - Agentless is more of a marketing term as there is something running on there.
* Charles Schmidt - I would take this to mean we are getting all the data from our tools. I would also take this to say if we could create a situation where the TCO was not so high then this would be useful.
* David Kennel - Acquisition is easy. Paying the mortgage is hard.
* Blair Heiserman - Cost and expertise to really learn and understand the results of the products. Managing exceptions is always interesting.

# Question: End users, how often are you in a situation where you wish you could gather additional information, but, you cannot get it?

* Charles Schmidt - Experience weekly or more often?
* Jessica Fitzgerald-McKay - +1.
* Charles Schmidt - Few times a year?
* Blair Heiserman - It’s the one degree off questions that kill me. I can gather the information, but, have to modify the content to gather the information.
* David Kennel - We collect as close to real-time as possible.
* Jessica Fitzgerald-McKay - Yes, or the tool that does the analytic I need does not get the right data for the questions we have.
* Charles Schmidt - Interesting there isn’t a lot of data they can’t get, but, people say they would like more tools. Could someone help me understand the combination of those answers?
* John Field - I just want to add to the question. I think I have the same cognitive confusion. We can get everything we need, but, want more tools. I can get the information, but, then want to get more posture. We want to develop more useful tools, do we think the data is sufficient today? I certainly don’t feel that way.
* Tim Jones - Shadow IT scratches this itch, but, does not fit into the larger solution.
* Blair Heiserman - What I don’t have is good data reporting tools. I know I have the data, but, no easy way to pivot to answer the questions.
* David Kennel - Visibility and analytics could both be improved.
* John Field - +1.
* Jessica Fitzgerald-McKay - Visibility and avoiding vendor lock-in.

# Question: To vendors, SCAP v2 envisions a federated content creation model. This means *Vendor A* might perform collection based on instructions from *Vendor B.* What assurances are necessary for this to be acceptabile if you are *Vendor A*?

* *<no concerns were raised>*

# Question: To vendors, SCAP v2 envisions a collect-once/use-many system. This means *Vendor A* might perform a collection of data used by *Vendor B*. What assurances are necessary for this to be acceptable if you are *Vendor B*?

* *<no concerns were raised>*
* Charles Schmidt - Can I get a vendor to say that you aren’t concerned with that?
* David Solin - Our concerns are very narrow, but, we are listening.
* David Ries - I can throw out some. It really depends on how you are combining data. There are things like Splunk where it gets all this data. It doesn’t vouch for the data, but, it makes the promise to be able to analyze the data. I can’t imagine they would have a problem because that is the value of their product. There are other products that go get the data and show it to you. But, if it doesn’t align with their value and contract proposition, it won‘t work for them.
* Tyler ??? - Software that tracks traffic through a kernel. I don’t know if that data would be useful to an agent vendor.
* Charles Schmidt - Would it be fair to say data consumers (e.g., Splunk) might have one answer whereas other vendors, who are making assertions about the enterprise rather than the data, might have significantly more concerns?
* David Ries - If you are designed to get the data, analyze, and report that data then it would get tricky because those tools now have to put some caveats about their assertions and they do interesting things (aging, how data is collected, etc.). There may be more ways in how the information goes in that.
* Charles Schmidt - So, you are saying additional, contextual information would be necessary?
* David Ries - Potentially.
* David Kennel - Not a vendor, but, it scares me. We’ve seen significant false positive/false negative issues with some tools.
* David Kennel - Data quality is very important in this context.
* Blair Heiserman - I want visibility into what is being tested (or a certification that someone validated that what was being collected was done properly).
* Blair Heiserman - I prefer to be able to do an independent validation of content. Once, I can validate the content aligns with reality, then I have trust for the vendors tooling regardless of how they gather it.
* David Kennel - +1 to ^.

# Question: Is this question of endpoint interactions (query and event-driven) overcome by discussions earlier in this presentation?

* Blair Heiserman - Both interactions have value.
* Blair Heiserman - Also, periodic gathering of state with reporting when the state changes.
* Charles Schmidt - If we are not standardizing the endpoint and Posture Collection Server protocol, can we even say the data must be queried or event-driven? Can we reconcile with this as we try to be more event-driven reporting?
* David Solin - Not in the world I live in. They are different products, different customers, different requirements, different release cycles. Trying to stick them in the same agent is really hard. It would have to be a big pot of gold at the end of that rainbow to get something generic that works for everyone in all situations. It’s fanciful and a question you might ask in a classroom. Kind of seems like a non-starter.
* Charles Schmidt - We aren’t trying to do this for everything necessarily, but, rather the most critical elements.
* Jessica Fitzgerald-McKay - You would have a hard time getting all these capabilities into one agent. How is that different than all the agents on the same system?
* David Solin - I am not sure we are talking about the same thing when we are talking about an agent.
* Jessica Fitzgerald-McKay - Agree. Also, Charles Schmidt tell me where I am misunderstanding. If you have a Posture Collection Server, even when the agents are from that product, and you are saying all the agents from different vendors are on an endpoint. I think they aren’t compatible?
* David Solin - There is what is possible and what is downstream. The direction in which the water is going to flow. I think that they are not totally incompatible, but, I think it is going to be hard to drive. When I had products that work in radically different ways, no one at the company was going to sign up to make everything interoperate because customers were always willing to pay more for other things. And these are products that barely talk client-server in the same version rather than just a hotfix off from one another. I just find this hard to have a one-size fits all protocol to not have issues.
* Jessica Fitzgerald-McKay - One message that was covered in the last teleconference is we aren’t thinking about one protocol for all endpoint types. We are just trying to get to the place where we are saying there might be many options for that standard and agents and servers don’t talk to each other using home-grown shims. We should talk about this more. Specifically, what we think of I1 and what capabilities we care about and what is out of scope.
* David Solin - WMI is query and event-driven so it’s in the realm of possibilities. Event-driven is also polled-query. If you can limit the vocabulary, you can do this.
* Charles Schmidt - You said your products were radically different and normalizing them wouldn’t be easy. Can you expand on this?
* David Solin - Limited R&D value or technical?
* Charles Schmidt - Technical.
* David Solin - Written in different languages. Different communication capabilities, etc.
* Charles Schmidt - Are there any other comments or questions?
* David Ries - I wanted to see if I can throw out an idea. I am an SCAP v1 wonky person. We have a basket of benefits query/event-driven, collect-once/use-many, single agent that does things. All seem good, but, there are lots of ways to do these things. This is an architectural approach and there was some resistance to that. Maybe there is another approach where we could make incremental changes to SCAP v1, but, then vendors will deliver these capabilities to you. One benefit to this is inventory. Drives software, vulnerability, etc. doesn’t change as frequently as overall endpoint state. Right now, in SCAP/OVAL content, you have a huge piece of OVAL content with lots of queries and it’s hard to do on an event-driven basis and hard to deliver benefits. But, if there was a change to OVAL definitions rather than having criteria to and/or together you had to define inventory components to the definition SWID, etc. here are the components, vulnerability, application version, etc. If those had to be separated out and labeled as inventory definitions. Then, you could have a much bigger definition and would enable tools to process that content and SWID inventory data that had been collected by other agents on an event-driven basis etc.
* Charles Schmidt - What you are saying is up-level the expressions. Instead of saying seeing if bit 3, 5, 7 are set and check this file hash, you can now say check for this software name X, version 1.2.3. Did I capture that?
* David Ries - You could still have all that. But, if you had those checks, some of them are trying to check if a certain version of the operating system, then a check for whatever you are actually doing. But now, checking for an operating system could be in its own file with SWID and CPE. I think that would support the inventory event-driven use case in a way that collects all endpoints (SWID, OVAL, etc.) and stored centrally so you could evaluate entire SCAP content.
* Blair Heiserman - One thing that I find interesting is the inventory piece and whitelisting vulnerability or IoT thing. Especially, when gathering this data is massive. I like SWID tagging, but, I am still waiting for a malicious SWID tag here with people dropping them in a known location.
* Charles Schmidt - SWID tags are going to be a whole other conversation, but, the idea of using available evidence and segmenting and simplifying the inventory part of the process. I also hear that one and whether SWID tags or OVAL definitions are good ground truth for that.
* Blair Heiserman - I think it depends what you are trying to do. SWID tags could be a quick light-touch run. If I need some assurance that the SWID tag is correct, you need to look at something deeper level (executables, hashes, DLLs, etc.) which falls on the vulnerability side of things. Also, whitelisting would be good too. I think the concept that has been talked about is all of these tools are collecting the same data and all attempting to hash every file on your system to look for malicious content, it can be a huge performance hit.
* Charles Schmidt - To summarize, this is not a binary answer. There are cases where you want something quick-and-easy and others where you need more authoritative/validated checks.
* Blair Heiserman - Correct.

# NEA

Charles Schmidt gave quick overview of NEA and the three components (transport, routing, application layer). Charles Schmidt also explained how NEA was interesting because it provides a common infrastructure for different collectors which are extensible and plug-and-playable.

* Blair Heiserman - How does this fit into the general industry transition to MDM?
* David Solin - What if one app needs steaming and another needs pub-sub? Products are not as cleanly built as you might think in reality. The quickest and easiest path to implementing this stuff will supersede anything you might get by implementing this.
* Charles Schmidt - it’s not just a matter of different data, but, it’s more operational requirements that can’t be abstracted away in this.
* David Solin - Yes, this stuff could be abstracted away, but, you can’t start from scratch with products. Someone suggesting we do this with R&D money will be replaced with someone who wants to spend it on ways to get more revenue.
* Jerome Athias - Do you mean machine2machine?
* Blair Heiserman - Mobile device management (InTune, Apple policies).
* David Solin - With this, we have just hidden everything, but, not added value.
* Blair Heiserman - I am okay with accepting a vendor’s implementation approach as long as I can validate it. It can be done with native tooling or something like Common Criteria. Then, I am satisfied with how it is completed and the technical specifics. With this, I am more interested in sticking with vendors’ native capabilities. While I appreciate standards-based approaches, I would like to move away from vendors always translating standards capabilities. I want to have the assurance that I get the same data from the vendor that I would get if I queried it manually.
* Charles Schmidt - I am hearing vendors have ways they engage with endpoints to collect the information they need. What I am hearing is not a lot of people are asking for a unification of the methods by which those collections occur and that in fact it might be prohibitive to do so, but, if there could be a normalization of how information is exposed and a way to get a level of assurance to the quality of that data. The semantics providing that ground truth, they see a lot more value for their enterprise and tools.
* Blair Heiserman - I have requested these standards-based approaches and getting vendor tools that take these standards and provide answers through proprietary mechanisms. So, I am waving the white flag and saying I don’t care how it is collected as long as it is validated.
* Jessica Fitzgerald-McKay - I still see value in this standardized interface because we still have the problem of running too many agents on an endpoint to collect all of the information we need. Also, I am not certain without an interface to endpoints, we will be able to achieve our goals for interoperable SCAP tools.
* Blair Heiserman - I agree a standard interface would be nice because if five vendor products want to do file hashing, they could all put the data in one file. But, right now, they all try to do it independently and that means there is too much going on, so I use one too. If they did this, I could use more than one tool and not cause performance problems on endpoints.