20190717 Endpoint Data Collection Telecon
Attendees: Chris Argenio (MITRE), Jose Luis Flores (IKERLAN), Bob Masucci (MITRE), Jess Fitzgerald-McKay (NSA), Adam Montville (CIS), Bill Munyan (CIS), Joe Sain (MITRE), Charles Schmidt (MITRE), David Solin (Joval), Andreas Steffan (strongSwan)

This is the first of the quick sync-up calls with each of the SCAP v2 to see where things stand and what the next steps are for each group. The first item to be discussed is the subgroup charter and scoping document, which was drafted by MITRE.

Thoughts on the charter:
AM – My question: what does "support event-driven" mean. Does that mean "support" but not require? Does this mean that event-driven data collection is supported and that other methods of data collection are supported as well?
JFM – It is important that a set of critical information be gathered in an event-driven way. A useful discussion item would be to enumerate the pieces of information that ought to be collected in an event-driven way.
CMS – This portion of the document was intentionally open-ended. Event-driven information collection is great when it's possible and useful.
AM – I agree that event-driven collection is great when possible and useful, and that it may be preferred in some situations. We need to recognize that other methods of data collection will be necessary as well.
DS – I look at this as a web service for collection. I read that there was a way to subscribe for the data rather than poll. The subscribing party could indicate what data is of interest.
AS – I've been collecting data using IETF standards for the past 10 years. I would like to see the information collected in an event-driven way if possible.
JFM – I'm interested in what can be reasonably be gathered. David is talking about a device that perhaps advertises what it can push?
DS – Yes. You would have the ability to query for all kinds of information. Some of the info would be delivered on demand (pull model), but some could be available by subscribing for alerts.
JFM – So we agree that not everything can be event-driven but some subset of the info can be event-driven, which will benefit the security of the system.
DS - There should be a way for the endpoint to describe what can be collected in what way.
AS – The concept of subscriptions is a good idea. Critical events should be delivered in real time. Other events can be accessible by polling.
JFM – It feels like this discussion has expanded on what is in the charter. Perhaps we should generalize the second bullet, which describes "characterizing the endpoint." We need to be able to allow for variations in capabilities such as the concept of developing a subscription capability in which the endpoint can describe its capabilities to the collection service.
AM – Agreed. We are looking for a way to understand what data an endpoint supports. We will need to find that info - maybe that happens at enrollment, maybe it happens periodically – but it seems necessary.
DS – I had read "characterization" – if a device is registering into the ecosystem, it would self-describe the data it can provide, although that is an implementation that fits the requirement rather than a requirement.
JFM – It sounds like an initial work item would be to clarify what we mean by "characterization." What are the options and what implementations do we prefer?
DS – Yes.

AM – If a system or endpoint is going to self-describe the data it can provide, do we need to consider how we describe what that data is (Format/version, etc.). It might be too early for that conversation. The system can provide software inventory, but it should also provide what language it speaks. Are we going to need an enumeration of languages that are used? Is it too early for this discussion, as it is starting to get into solution space?
JFM – Agreed. It is something that we are going to have to do eventually, but it is very focused on a solution. We should capture this for future work items.
JFM – I would like to hear some discussion on the initial work statement in the last paragraph of the charter.
AM – Would part of the "overall architecture" be the identification of things like interfaces, languages, processing, order of activities, etc.? I've been looking at this problem from an interface/ information perspective more than anything.
JFM – The big picture architecture for SCAP v2 is very hand-wavy over the endpoint data collection. In previous conversations we discussed existing capabilities that we might want to leverage for endpoint data collection. Would be good to understand how those collection mechanisms work. This would allow us to drill down into each of those potentially useful existing technologies, but also generalize them. We could then be more specific and explicit description of how endpoint data collection works and be more expressive of all the different technologies we want to leverage.
AM – Regarding the sequence diagrams – in SACM at one point we had a vulnerability assessment graph that laid out a workflow – is that what is envisioned for the sequence diagrams?
JFM – I'd like to see that kind workflow. The bigger SCAP v2 project has a set of use cases. We could use them to develop those workflows; how do we support each of those use cases?
AM – Does overall architecture also include component identification, where these are the things that need to talk to each other?
JFM – Yes.
AM – The SACM architecture draft is currently looking at the workflows, the components of the workflows, the the interfaces that those components need to support, and the data that is going to travel over those interfaces.
JFM – This gets to the second item: prioritization. Is there something to tackle first? Would this be the core of the work to bring to the September meeting?
AM – A lot what we are talking about seems similar to what we have been tackling in the SACM working group. Is this intended to be separate from that work, parallel to that work, or the same?
JFM – I see SACM as an input to this process. Dave Waltermire and I have performed a certain amount of work in SACM that we have included in the SCAP v2 documents as an option for endpoint data collection. Your SACM Architecture should be considered. I felt in earlier conversations that there were collection methods beyond what has been described in SACM that the group was interested in pursuing. I think David Solin expressed that there are other non-standardized methods of data collection. Can you elaborate?
DS – I was thinking that we might consider the possibility of having an infrastructure node that acts as a broker on behalf of the endpoints that want to live on the network, rather than imposing new requirements on them. Right now, there are many means of collecting data from different endpoints. You can run an agent on some endpoints; some endpoints support SNMP; some support SSH. What we could do, instead of saying that every device has to speak this specific protocol and has to provide this set of information, we could allow third parties to create a "SACM compatible" node for your existing CISCO infrastructure, for example. Otherwise it will take a long time to achieve that ecosystem.
CMS – It sounds like you are describing a translation node that could take the native collection/reporting capabilities for an infrastructure and provide in a SACM-compatible way.
DS – Yes.

JFM – So it sounds like you are interested in focusing on the data sent between the 3rd party node and the posture collection service; what is that information and what does that node need to translate on our behalf?
DS – Yes. my proposal is to move the interface from the endpoint to some other device. It won't necessarily be at the endpoint. There will be a SACM-compliant way that you can query information for all the devices.
CMS – I will note that it is not incompatible with having that ability on the device.
DS – Absolutely. Something reports on one device (self) or many.
AM – That is compatible with what we are trying to put forward in SACM.
JFM – Ideally, I would like a standardized way to talk to a device type or system type. In the interim, though, since this is not how are networks are configured, having a translator is very useful, as long as we don't disregard the ability to communicate directly with the device as well.
DS – I don't think we will ever get to that ideal state, but as long as we don't require it, we can support it.
BM – I agree. Having the interface would be great, and that interface would be transparent to whoever is invoking it. It could be a broker speaking for the endpoints or the endpoint itself.
JFM – Sounds like we have an idea of a direction that we want to go and what the focus should be.
JFM –What would be our initial product look like and what would early milestones be? Do we have thoughts on something that we could draft up and have ready for the next SCAP workshop in September?
CMS – What are the things that we are working toward, and what are the near-term milestones in that path?
DS – We need more than we have. We should decide what that will be and work backward from there. I'd say a draft specification, but not sure what that would entail.
JFM – A draft specification would show some forward progress. Some ideas:

- A list of existing capabilities we want to capture
- description of the broker/translator node
- ideas on what the connection between the broker node and the posture collection service would look like
- what information we would like to see collected in an event-driven way
- what is out of scope for event-driven collections
- what is needed for subscriptions?
- how endpoints announce their capabilities

DS – I remember seeing lots of diagrams – were those parts of a document?
JFM – The architecture diagrams could be found on the NIST site.
DS – We should revise those diagrams.
AM – Could the SACM Architecture be an input to this? Changes to that draft that would benefit this charter. Since Bill and I have the pen on the draft, we could update that draft.
JFM – I think it would be good to share the SACM with this group if you don’t mind.
DS – Should we do as an update to the transitioning to SCAP v2 document, or should we develop a new document that is based on it?
JFM – Doc owned by NIST so none of us have the ability to modify it. We could take the initial diagrams, use them as a starting point, and suggest ways to elaborate on them. We could use that elaboration to capture existing capabilities that we would like to leverage. We could add in the 3rd party node concept. We could say that we are creating a more fleshed architecture that deals with endpoint data collection.
CMS – Adam – I wanted to go back to your comment about the interaction between the SACM draft and the EDC charter. Could you elaborate on the opportunities there between the two documents?

Conceptually, they are very closely aligned, but in terms of actions that could be taken, do you have something in mind?
AM – My opinion is that everything discussed on today's call fits into SACM Architecture. I would say that what we want in that architecture draft are also the things we want here (additional workflows, use cases, asset management, vulnerability management, configuration management, etc.). What I was trying to get at is, should we continue the SACM architecture work and have that be a product for this group to leverage. This could save the group from starting from scratch. We have something written down; we've done hackathons that experimented with implementations of the SACM architecture. We have made some reasonable progress. It would be great if we had more people to work on this as a concerted effort. Is this group the place to do that?
CMS – The SACM architecture is an IETF document, so I'm not sure how difficult it would be to evolve that architecture outside the auspices of IETF. I will have to read the SACM document in its entirety, but my recollection is that it includes a lot that this group is not focused on. Can we work on that document in a way that remains focused on the objectives of this team?
AM – You are pointing out that there is an intersection between this group and the SACM group. My ultimate goal here is to support an ecosystem of cybersecurity-related tools. It is why SACM exists. I would say that there is an intersection but not a whole set match. I do want to see an ecosystem of tools where this would plug into the ecosystem and they would work together. This is one important piece, data collection, that supports all other workflows. Maybe it is the case if we looked at the architecture and ensure it is satisfying what we need here. If it doesn't, comment to the list and we can try to push through SACM. Then we have the connection between the data collection work here and the SACM work. Maybe we could eventually move this work to SACM.
JFM – As a participant in both efforts, I would dislike seeing the SACM Architecture stray too far afield from what we are doing here. But I also agree that the SACM Architecture addresses an additional set of use cases beyond what we care about in this group. Let's treat the SACM Architecture as one capability we want to support. It is a feed into this effort. I agree, if we find places where what we want to do requires a change to the SACM architecture, we should bring to the list. If that does come up, I will be happy to take the action on behalf of the group. Charles and Andreas are also participants in both efforts; we can monitor progress of both and keep them from separating too far.
AM – Makes sense. Regarding the SACM Architecture, it really is two architectures:

1. How we get collection done.
2. Supporting the "downstream architecture."

To me the collection feeds into the SACM architecture directly – being able to reuse that data is the intent.
JFM – Agree. Let's look as an existing capability and go forward from there.
DS – My fear is that there is a lot of work to catch up on. I started attending SACM meetings 5 years ago and it is still ongoing.
JFM – I hear you. If we treat the SACM architecture as an input, like other inputs, I don't think we need to review everything. Those of us in SACM can bring any commentary back to the list.
AM – You wouldn't need to catch up on 5 years of work.
JFM – Seems like the group's interest is on an initial product focusing on architectural over information or data model concepts.
AM – Agree, but I think that the data model comes closely after that.
DS – Agree
BM – Agree
AS – Agree. To me the SACM architecture is fine. I would like them to go forward with implementing things.

JFM – That is something we can talk about between now and September I would like to set some milestones between now and September 16. Should we have another call? Is there a draft document that we can begin? What would our early milestones be?
JFM – Perhaps, Adam, Bill, and I can get together in Montreal and discuss it there. If we could put an outline together and share it to the list, that would be a good start.
JFM –What does this group prefer? Do we prefer more email list discussions, or do we want additional telecons?
CMS – Telecon.
DS – Agreed.
JFM – OK, we will send out a poll for meeting times for the week of August 5. Bill, Adam, and I will meet in Montreal, and we will share our progress. We will meet again on the week of the 5th.

Action Items:

- Set up a telecon for the week of August 5
- Charles/Andreas/others will track the SACM architecture and note any concerns about compatibility with the needs of this workgroup.
- Adam/Bill/Jessica meet at the IETF meeting and put together an outline of milestones to share with the list.