vendor id / vid keyword to give rulesets unique sid ranges

Since more and more rulesets are added to the rule index, risk of sid space collisions increase. It's been suggested that 'gid' should be used to separate them, but the gid field already has other uses in the real world.

So the idea is to add a 'vid' keyword to keep rulesets apart. The rule index should probably also be the master of the vid id's.

Related issues:
- Related to Task #2685: SuriCon 2018 brainstorm
- Related to Task #4097: Suricon 2020 brainstorm
- Related to Documentation #3303: Add a documentation about the used sid and gi...

History
#1 - 12/21/2018 11:36 AM - Victor Julien
- Related to Task #2685: SuriCon 2018 brainstorm added

#2 - 05/23/2019 09:44 PM - Andreas Herz
- Assignee set to OISF Dev
- Target version set to TBD

#3 - 11/19/2020 04:29 PM - Victor Julien
Could be as a string to avoid vid collisions.

#4 - 11/19/2020 04:52 PM - Jason Ish
- Related to Task #4097: Suricon 2020 brainstorm added

#5 - 11/23/2020 12:25 PM - Victor Julien
Several things were discussed at the 2020 brainstorm:
- vid in rulesets would still require some kind of central registry
- vid in output would mean all post-processing would need updates
- could we just add it to 'metadata'? No solution for sid collisions
- Suricata-Update could also auto assign a vid, so sid collisions wouldn't happen. Doesn't solve the output issue
- suggestion was made to move away from sid/gid in favor of some kind of uuid. Rev would have to stay.

In conclusion: no decision. Needs more thought.

#6 - 11/23/2020 04:35 PM - Jason Ish
Based on the discussion in the Brainstorm Slack, no one was really in support of a new "vid" field. A new UUID field seemed to have a little more support.

But there seemed to be more support for using more of the existing fields to generate a unique identifier. At the most basic, this could be left to post-processing. Suricata could be modified to not enforce a unique sid, and as it already logs rule metadata, output processors could derive uniqueness from the sid, and some metadata like a "vendor_id" or "rule_source" in the metadata. However, existing processing tools that depend on uniqueness of the SID may fail in this case, or reporting tools that aggregate on sid will provide bad reports.

Another idea is to pre-process rules to provide a unique SID. Suricata-Update could dynamically create SID ranges for each rule_source. Then update the metadata in the rule with enough data to map back to the original rule (rule source, and original sid). This ensures all rules provided to Suricata via Suricata-Update would have a unique SID, and post-processing tools for the most part would not need to change. Tools that link back to the original sources of the SID would need to change though, as the SID may not be the same. This should be done as needed, not every ruleset
should have its SIDs remapped.

This pre-processing approach would also work for generated IOCs, where I hear they just start at some low number, or use a date.

Related tickets:
- #4113 -- intel index: register sid ranges per rule source
- #3303 -- Add a documentation about the used sid and gid ranges

#7 - 11/23/2020 04:36 PM - Jason Ish
- Related to Documentation #3303: Add a documentation about the used sid and gid ranges added