Summary of the Workshop

On

Cyber Incident Data and Analysis Repository

National Protection and Programs Directorate/Department of Homeland Security

April 19-20, 2016

Background

In 2015, The Department of Homeland Security's (DHS) National Protection and Programs Directorate (NPPD) facilitated a series of discussions among insurers, chief information security officers (CISOs), and other cybersecurity professionals on the concept of a trusted cyber incident data and analysis repository (CIDAR). The repository aims to enhance national cyber resilience through enhanced voluntary sharing of cyber incident data. Over the course of several months, the Cyber Incident Data and Analysis Working Group (CIDAWG) identified cyber incident data categories that notionally could form the basis of a future repository development effort and be used by contributing companies, insurers, and cybersecurity researchers to perform trend and other analyses. Such repository-supported analyses, conducted in strict accordance with all applicable legal and privacy requirements, could help both private and public sector organizations better assess cyber risks, identify effective controls, and improve their cyber risk management practices. The CIDAR could also foster the development of new cybersecurity insurance policies that "reward" businesses for adopting and enforcing best practices.

On April 19-20, 2016, NPPD hosted a workshop to discuss the value and the feasibility of a CIDAR. The workshop built on the work the CIDAWG has accomplished thus far and focused on the execution of the repository. The goals of this workshop included:

1. Share the findings of the CIDAWG on the:
    a. Value proposition of a CIDAR;
    b. Cyber incident data points that could be shared into a repository to support needed analysis; and
    c. Perceived challenges to sharing data into the repository and overcoming those challenges.
2. Validate the feasibility of and soliciting support for a CIDAR from the broad cybersecurity community. Receive input on how cyber incident data points shared into a CIDAR should be prioritized, operationalized and automated and how the repository should be executed.
3. Receive input on voluntary information sharing approaches, models and best practices that could inform any future repository implementation.

These findings can be found on: Cyber Incident Data and Analysis Working Group White Papers.

Summary of the Plenary - Day 1, April 19, 2016

- Background and Overview of the CIDAWG, its Key Findings and Conclusions; Tom Finan, Chief Strategy Officer, Ark Network Security Solutions and Cynthia Wright, MITRE, supporting DHS/NPPD.

  CIDAR Background_04192016

- Welcome Remarks; Dr. Phyllis Schneck, Deputy Under Secretary for Cybersecurity and Communications, NPPD

- Panel; Sharing of Sensitive Data
  - **Description:** Panelists discussed how their organizations collect, share and anonymize data as well as lessons learned in metric selection and other decisions made when establishing and managing a data repository**.**
    - **Moderator:** Moderator: Dr. Sandor Boyson, Supply Chain Management Center, Robert H. Smith School of Business, University of Maryland College Park;
    - **Panelists**:
      - Rick Lacafta, Financial Services - Information Sharing and Analysis Center (FS-ISAC):
      - Randy L. McGuire, Aviation Safety Information Analysis and Sharing/Federal Aviation Administration:
      - Tom Millar, US-CERT:
      - Robert Frost, Cybersecurity Information Sharing Partnership (CiSP), UK-CERT:
      - Harold Booth, National Vulnerability Database, National Institute of Standards and Technology
  - **Discussion Summary:**
    - The panelists agreed that a CIDAR would require:
      - a clear focus to differentiate it from other information sharing systems;
      - the establishment of reporting thresholds to minimize burdensome reporting of relatively minor incidents;
      - a clear alignment of participation incentives with business interests; and
      - automation of data to the greatest extent possible for both scalability and ease of participation.
  - **Panelists provided the below answers to the following questions:**

- <u>What are the major concerns of your various stakeholders? What did you need to do in order to obtain their cooperation?</u>
    - non-disclosure and terms of use agreements;
    - traffic light protocol
    - anonymization
    - structured operating procedures/rules
    - two factor authentication
    - digital authentication
    - end-to-end-encryption
    - outreach, education and clear explanation of the value of sharing
    - ensuring the validity, currency and relevance of data
    - transparency and strong governance of the information sharing structure
    - when participation reached a certain mass, segmentation of participants into smaller communities or finding other forms of authentication and reliability of the data helped maintain trust
- <u>What would need to happen to make the CIDAR work?</u>
    - over commitment in setting up the repository information sharing environment
    - continuous reviews and engagement with contributors to ensure the system is working as planned and to revise as needed
    - Clear incentive structure for every participant
    - "crawl-walk-run" approach; make sure the effort is scalable, and automate as much as possible

Summary of the Breakout Sessions – Day 1 and 2

- **Description**: NPPD facilitated two breakout sessions where workshop participants evaluated each cyber incident data category the CIDAWG has developed and published in December 2015. In order to give all workshop attendees an opportunity to discuss each data point in some detail, the facilitators divided them into groups that discussed data category subsets (General Incident Data, Organizational Practices/Maturity, Incident Response and Recovery, and Consequences and Impacts) in rotation. In addition to observations about the data categories and their notional input mechanisms, participants were asked to comment on:
    1. What cyber incident metrics do organizations already track; what additional data points they should be tracking for the purposes of the repository; and, what would be the additional cost of tracking those new data points?

2. What data is the easiest to obtain?
3. How should organizations collect the data?
4. How should the data be automated and operationalized?

Because of the CIDAWG's disparate membership, many of its discussions during the development and refinement of the proposed 16 data categories revolved around arriving at a common understanding of terms. In an effort to level-set workshop participants from an even broader array of industries, the NPPD team provided an overview of each data category as it was conceived by the CIDAWG, and an explanation of how the CIDAWG members envisioned them being used. In brief, the 16 proposed data categories offered for consideration and discussion included:

1. **Type of Incident** - High-level descriptor or "tag" (e.g., "Ransomware") to differentiate the incident for ease of reference, leaving the capture of specific technical details about the incident to other data categories.

2. **Severity of Incident** - The relative scale or scope of an incident within the context of the incident contributor's industry and circumstances. .

3. **Use of Information Security Standards and Best Practices** - The cyber risk management practices, procedures, and standards that an organization had in place at the time of an incident and/or attack.

4. **Timeline** - The date of detection of a cyber incident and the date of effective control.

5. **Apparent Goals** - The assets apparently targeted, implying their financial, reputational, and operational value to an attacker.

6. **Contributing Causes** - People, process, and/or technology failures contributing or otherwise relevant to an incident and/or attack.

7. **Security Control Decay** - A set of circumstances where a security control, although present, did not operate effectively enough to withstand an incident and/or attack.

8. **Assets Compromised/Affected** - The points in a network and/or business where an incident and/or attack took place.

9. **Type of Impact(s)** - The specific effects of an incident and/or attack on all affected parties.

10. **Incident Detection Techniques** - The techniques used to identify an incident and/or attack, and their effectiveness.

11. **Incident Response Playbook** - The tactics, techniques, and procedures (TTPs) used to respond to an incident and/or attack and to bring it to a close, and their effectiveness.
12. **Internal Skill Sufficiency** - Availability and sufficiency of an organization's skills and capacity to quickly address and resolve incidents and/or attacks.
13. **Mitigation/Prevention Measures** - Actions taken to stop incidents and/or attacks and to prevent similar future occurrences.
14. **Costs** - Financial and other quantifiable costs incurred as a result of an incident and/or attack.
15. **Vendor Incident Support** - Vendor behavior during the assessment and resolution of a cyber incident and/or attack.
16. **Related Events** - Related activities that provide incident and/or attack context.

General observations on the 16 data categories:

- Based on insurance carrier experience, ideally, there should be fewer than 10 categories and 10 data points in each category.
- Eliminate overlaps of similar data categories.
- Each data category should be analytically independent of the others to the greatest degree possible so that lack of data in one area doesn't hinder analysis in another.
- Based on prioritizations within the break-out groups, participants recommended the following changes:
    - Delete Data Category #5: "Apparent Goals." It is highly speculative, and "Assets Affected" effectively captures the data participants considered most useful.
    - Delete #7 "Security Control Failure." It can be effectively covered under #6 "Contributing Causes."
    - Combine #10 "Incident Detection Techniques and #11 "Incident Response Playbook" into a single category called "Incident Detection & Response." Alternatively, consider deleting both since other venues exist for sharing playbooks. As currently written, these categories do not provide sufficient detail to deliver insight into what detection and response approaches are effective. Moreover, #13 "Mitigation and Prevention" addresses long-term response techniques (though participants suggest tying this category to effectiveness as well).

- Change “Mitigation Prevention Measures” to “Recovery”
- Delete #15: “Vendor Incident Support.” Participants felt the category is not useful if it’s anonymous and is legally problematic (and may break the anonymity of the contributor) if vendors are named.
- Delete #16: “Relevant Events.” Attendees found it too speculative with regard to causality, and feared that data could be aggregated to support class action suits.

If these recommendations are adopted by the CIDAWG, the number of data categories will drop from 16 to either 10 or 11 (depending on whether the incident detection/playbook response categories are combined or deleted). The initial priorities for CISOs include: ***Incident Type, Assets Affected, and Recovery***; with the addition of ***Costs and Type of Impact*** (including non-financial) for insurers. The impact of a cyber incident may require some time to emerge, and may not be as readily available.

<u>General Incident Information</u>

1. **Type of Incident** - High-level descriptor or “tag” (e.g., “Ransomware”) to differentiate the incident for ease of reference, leaving the capture of specific technical details about the incident to other data categories. **COMMENTS:**
    a. A common, limited, and consistent taxonomy will be necessary to eliminate overlapping ways of characterizing an incident and. - For instance, a successful phishing attack could be categorized as phishing, human error, or failure in the email security gateway. Similarly, an ICS attack might be typified simply as malware.
    b. The incident data must be understandable throughout the organization including the C-suite.
    c. Check-box categorization is a challenge since new attack modes are constantly appearing - repository would have to be easily updated in order to remain current.
    d. Suggested frameworks that could be used to characterize cyber incidents include: PrEP and VERIS. One industry representative suggested that a combination of terms from those and others might work best to capture the unique benefits the CIDAR is envisioned to provide.

2. **Apparent Goals** - The assets apparently targeted, implying their financial, reputational, and operational value to an attacker. **COMMENTS:**
    a. Its intent could be realized by assessing the full scope of an incident (Data Category #9: Type of Impacts)
    b. Although it does call for a certain amount of speculation, it is important to know who might be targeting you and why. The difference between organized crime and insider threat has significant implications for cyber defense, particularly for on-line companies.

3. **Assets Compromised/Affected** - The points in a network and/or business where an incident and/or attack took place. **COMMENTS**:
    a. Difficult to capture the array of assets in a meaningful way due to the sheer variety in network architectures. Recommendations:
        i. Classify by layers according to the standard 7-layer network model—application layer, hardware layer, etc.- Remove "databases" from data point list.
        ii. With regard to data classification, input fields could focus on the type of server—e.g., file, web, or e-mail—as a way of characterizing the assets.
        iii. Capture where the data sits—locally, on third party servers, or in a private, hybrid, or public cloud.
        iv. Approaches will need to account for the moving target represented by the internet of things, where smart devices are ubiquitous.
            1. Insurance industry would need to increase flexibility to accommodate the constant changes in technology.
            2. The basis of insurance underwriting is to look at trends in an effort to identify exposure. It may be impossible to account for every asset in assessing risk, but capturing this data can help companies and insurers exercise due diligence in looking at third-party vendors.
        v. Need detailed list of compromises

4. **Related Events** - Related activities that provide incident and/or attack context. **COMMENTS**:
    a. Recommend deleting - too speculative, and calls for conjecture that could have legal or other repercussions.

<u>Consequences and Impacts</u>

5. **Severity of Incident** - The relative scale or scope of an incident within the context of the incident contributor's industry and circumstances. **COMMENTS:**
    a. Difficult to quantify.
    b. Financial loss severity scales can be useful for insurers when correlated with other information. Within the insurance industry, OCTAVE and FAIR are cyber models developed to quantify the amount of risk and the cost impact of cyber events.
    c. Financial measures of severity may not be meaningful in themselves, as the severity of a cyber incident could be measured in anything from market delay to loss of life.
    d. Severity should focus solely on the number of records affected, since "Cost" is captured in another data category, but in the case of as ICS incident, it is the system itself, not particular records, that is affected.
    e. Medical systems and devices would require a different measure than number of records as well.
    f. Use subjective assessment of what constituted "major" versus "minor" within the context of the affected industry or process,

g. Use existing legal standards (restitution) to establish comparative levels of severity.

6. **Type of Impact(s)** - The specific effects of an incident and/or attack on all affected parties. **COMMENTS**:
    a. The data points within this category do not realistically capture the impacts of a cyber incident on ICS. Too record focused – need to refine the categories of impacts.
    b. Need to simplify data collection for ease of use; avoid duplication (ex: employee information is the same as PII).
    c. Need to find a way to capture the secondary and tertiary impacts of an incident.
    d. Need to correlate the data points within this data category to make sure the right set of data points are being collected. – Consider: What are the insurance companies looking for in order to determine types and levels of coverage? What are the analyses needs of different industry sectors when it comes to impact assessment?
    e. Use the FISMA reporting categories—confidentiality, integrity, and availability—along with categories of system effects captured in the NIST 800.61 guidance.
        i. Include other aspects of aggregate exposure, such as injury/death, environmental harm, production loss, property damage, market share losses, etc.

7. **Costs** - Financial and other quantifiable costs incurred as a result of an incident and/or attack. **COMMENTS**:
    a. Important data category for insurers to support underwriting and determine insurance parameters.
    b. CISOs also need this kind of data to help making informed decisions about cyber investments.
    c. Need to find the right detail of cost categories to make them meaningful but easy to report on.

8. **Contributing Cause(s)** - People, process, and/or technology failures contributing or otherwise relevant to an incident. **COMMENTS**:
    a. Because this data point was intended to encompass human, process, and/or technology failures contributing or otherwise relevant to an incident, the Security Control Decay data category should be captured here. With the exception of human error instances such as successful phishing attacks or misconfigured devices, failure of existing security controls was of particular interest in order to track technologies that are becoming obsolete.
    b. The list of possible control failures could be kept manageable by using the NIST 800-53 control families or similar established classification systems.

<u>Organizational Practices and Maturity</u>

9. **Use of Information Security Standards and Best Practices** - The cyber risk management practices, procedures, and standards that an organization had in place at the time of an incident and/or attack. **COMMENTS**:
    a. Both insurers and CISOs raised concerns about the inherent difficulty in assessing the effectiveness of framework implementation without requiring some outside assessment.
        i. Different parts of a company could be working under different security frameworks, or the same framework could be implemented differently in different organizations.
        ii. A relatively weak framework implemented well could perform better than a prestigious framework implemented poorly.

10. **Security Control Decay** - A set of circumstances where a security control, although present, did not operate effectively enough to withstand an incident and/or attack. **COMMENTS**:
    a. Participants viewed this data category as being adequately covered by the Contributing Causes data category and recommend deleting it.

11. **Incident Response Playbook** - The actions, methods, procedures, and tools used to respond to an incident and to bring it to a close, and their effectiveness. **COMMENTS**:
    a. There needs to be a way to sync up response actions with what actually helped reduce the cost, duration, or impact of the incident. - This data is only important in the context of what happened as a result of using incident response TTP.
    b. This data category would add little analytical value to the CIDAR, and the CIDAWG-proposed data input options were ineffective to capture this data.
    c. Combine with data category “Incident Detection Techniques”

12. **Internal Skills Sufficiency** - Availability and sufficiency of an organization's skills and capacity to quickly address and resolve incidents.
    a. It might be more useful to ask which skills were accessed and whether they were internal, external, both, or neither.

<u>Incident Response and Recovery</u>

13. **Timeline** - The date of detection of a cyber incident and the date of effective control. **COMMENTS**:
    a. Using the time between detection and effective control (eschewing specific dates as identifiable data) was essentially a proxy for either “severity,” which is covered in data category #2.
    b. Focus on the first event that allowed the perpetrator access, along with the time intervals between initial access and further lateral moves within the system that provided access to more critical network services.

14. **Incident Detection Techniques** - The techniques used to identify an incident, and their effectiveness. **COMMENTS:**
    a. Discussion on this data point suggested that incident detection is largely captured in the combination of timeline and incident response data categories. Beyond identifying who (e.g., external security provider, FBI, internal cyber defenders, etc.) discovered the breach, which may or may not be indicative of the security competence of the breached enterprise team, this data point appears to offer little stand-alone value.

15. **Mitigation/Prevention Measures** - Actions taken to stop incidents and to prevent similar future occurrences. **COMMENTS**:
    a. This data category should form the core of a CIDAR prototype.
        i. it can be used to find out what peer companies are investing in
        ii. it can help identify specific software and component vulnerabilities if the specifics of the system patched (such as the CVE number) are captured.
    b. Need to capture whether the "mitigating" controls actually worked.

16. **Vendor Incident Support** - Vendor behavior during the assessment and resolution of a cyber incident. **COMMENTS**:
    a. Delete data category - given the anonymity provisions that must be integral to a CIDAR, any information sufficient to help identify what vendors should be avoided would also suffice to identify them—exposing the contributor to legal action—and possibly the contributing company itself.

<u>General Recommendations for a CIDAR</u>

- Greater clarity on whom the CIDAR is intended to benefit and how to incentivize companies to participate, as well as to shape the types and specifics of the data requested.

- A common taxonomy for the CIDAR is a must. Ideally, it should be based on one or more already-recognized standards—VERIS, the PrEP framework, and NIST 800-53 were mentioned by name in various contexts—applicability may vary by category. Explanatory material, preferably formal training, should also be provided because decision-makers in a possible contributor company may not be familiar with the cyber lexicon, but will need to understand what types of information they are contributing.

- A CIDAR pilot should start with basic, useful, and easy-to-acquire data categories in order to gain market acceptance. Over the longer term, data input must be practical—and if possible automated. Companies that experience thousands of "incidents" a week are not going to hire 15 extra people just to do voluntary data reporting.

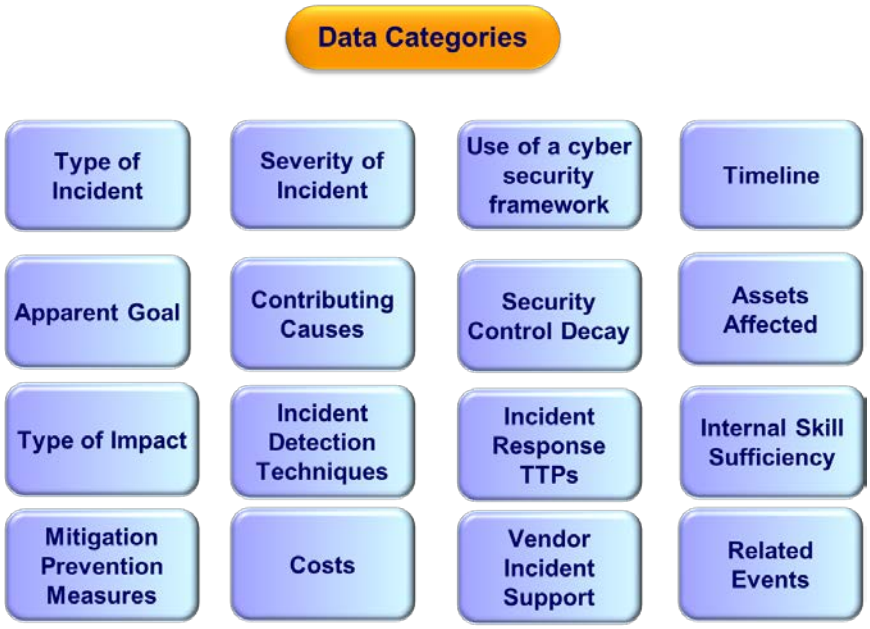
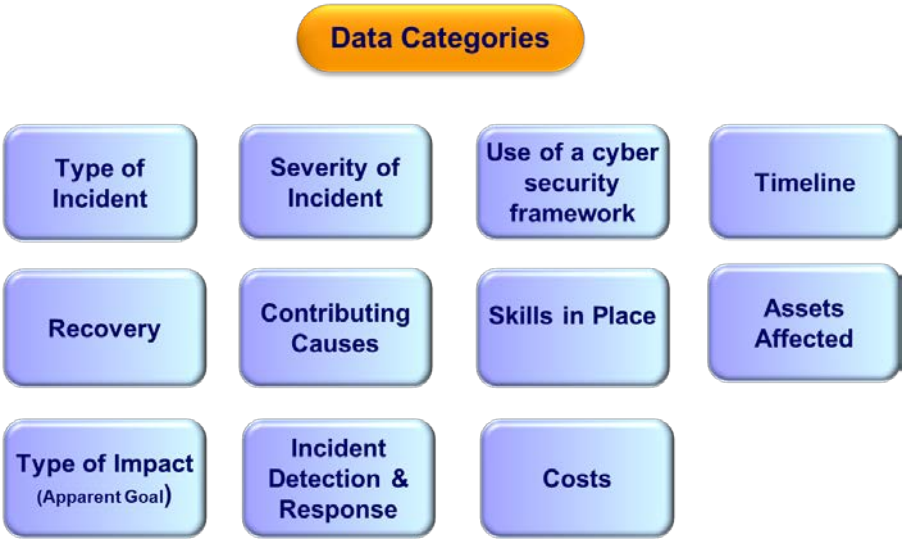
- During most cyber incidents, there will be a lot of unknowns. The CIDAR must be able to function with incomplete data, and the data it collects needs to be simple (the proposed list/checkbox system is good; freeform entry fields are problematic) in order to get the consistency that will support predictive analysis.
- Incident data will evolve over time—particularly costs, impacts, and the mechanics of the attack. A CIDAR must be designed to allow updates by the original contributor.
- Providing participating companies with a unique identifier can help with practicality by allowing much of the relatively static contextual data to auto-populate. Such an identifier should remain separate from the incident identifiers used in the externally accessible database in order to protect the anonymity of the contributor.

Conclusion and Way Forward

Most of the CIDAR workshop attendees appeared to find the forum informative and useful. In general, there was little resistance to the idea that a CIDAR as envisioned by the CIDAWG could be a valuable addition to the cybersecurity risk management toolkits of both insurers and enterprise owners. The DHS facilitators thanked participants for their thoughtful and insightful inputs, which will meaningfully contribute to the national cyber resiliency discussion. Attendees are encouraged to remain engaged in the CIDAR development discussion including, if interested, volunteering to participate in future CIDAWG efforts. Finally, DHS NPPD wishes to call attendees' attention to the material already published on the CIDAR, particularly the Value Proposition and Overcoming Perceived Obstacles to Incident Data Sharing whitepapers, available at: http://www.dhs.gov/cybersecurity-insurance.

Appendix A: Cyber Incident Data Categories

1. Cyber Incident Data Categories (initial)



2. Cyber Incident Data Categories (result of feedback from workshop participants)