Carnegie Mellon University
Software Engineering Institute

# Cyber Intelligence Tradecraft Report

## The State of Cyber Intelligence Practices in the United States

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.



DM19-0447

# Cyber Intelligence Tradecraft Report

The State of Cyber Intelligence Practices in the United States

By
Jared Ettinger

**Contributors**
Hollen Barmer
Jennifer Kane
Heather Evans
Erica Brandon
Ritwik Gupta
Daniel DeCapria
Andrew Mellinger

**Implementation Guides**
*Artificial Intelligence and Cyber Intelligence*
April Galyardt, Ritwik Gupta, Dan DeCapria, Eliezer Kanal, Jared Ettinger

*The Internet of Things and Cyber Intelligence*
Dan Klinedinst, Deana Shick, Jared Ettinger

*Public Cyber Threat Frameworks and Cyber Intelligence*
Samuel Perl, Geoffrey Dobson, Geoffrey Sanders, Daniel Costa, Lawrence Rogers, Jared Ettinger

**Design**
David Biber, Alexandrea Van Deusen, Todd Loizes

# Contents



## Implementation Guides

# Executive Summary

## *Strengthening Cyber Intelligence*

Intelligence dates to ancient times when early civilizations used it to protect their assets and gain an advantage over their adversaries. Although the ways we perform the work of intelligence have changed, it remains as critical as ever. And this can be no truer than in the cyber domain. In performing cyber intelligence, we collect, compare, analyze, and disseminate information about threats and threat actors seeking to disrupt the cyber ecosystem,[1] one of our most critical assets. Through cyber intelligence, we know ourselves and our adversaries better. And with that knowledge, we can proactively take steps to better understand risks, protect against threats, and seize opportunities.

In 2013, the Software Engineering Institute (SEI) at Carnegie Mellon University conducted a study on behalf of the U.S. Office of the Director of National Intelligence to understand the state of cyber intelligence practices at organizations throughout the country. We conducted a similar study in 2018, and this report details our most recent findings.

We built on outcomes from the 2013 study to develop foundational concepts that drive the 2018 study. First, we define cyber intelligence as acquiring, processing, analyzing, and disseminating information that identifies, tracks, and predicts threats, risks, and opportunities in the cyber domain to offer courses of action that enhance decision making. Second, we propose a framework for cyber intelligence; based on the intelligence cycle, its components provide for Environmental Context, Data Gathering, Threat Analysis, Strategic Analysis, and Reporting and Feedback.

During the 2018 study, we interviewed 32 organizations representing a variety of sectors to understand their best practices and biggest challenges in cyber intelligence. During conversations guided by questions designed to elicit descriptive answers, we noted organizations' successes and struggles and how they approached each component of the Cyber Intelligence Framework. We also provided an informal assessment of how well each organization was performing for certain factors within each component. We aggregated and analyzed these answers, grouping what participants told us into themes. This report moves through the Cyber Intelligence Framework, detailing our findings for each component. Three companion implementation guides provide practical advice about artificial intelligence and cyber intelligence, the internet of things and cyber intelligence, and cyber threat frameworks.

---

1 https://www.dhs.gov/sites/default/files/publications/DHS-Cybersecurity-Strategy_1.pdf

There are a number of areas where organizations can take action to improve their cyber intelligence practices. They include differentiating between cyber intelligence and cybersecurity, establishing repeatable workflows, breaking down silos that fragment data and expertise, enabling leadership to understand and become more engaged in cyber intelligence, establishing consistent intelligence requirement and data validation processes, and harnessing the power of emerging technologies.

Since 2013, the practice of cyber intelligence has gotten stronger. Yet it is not strong enough. In the coming years, data and compute power will continue to increase, and artificial intelligence will enable us to make sense of threats while also making threats themselves more complex. Organizations of any size can learn from and apply the best practices and performance improvement suggestions outlined in this report. Together we can achieve higher levels of performance in understanding our environment, gathering and analyzing data, and creating intelligence for decision makers.

# Cyber Intelligence Study Report

*Introduction*

## ABOUT THIS REPORT: IMPROVING THE PRACTICE OF CYBER INTELLIGENCE

This report details the findings of a study the Software Engineering Institute (SEI) at Carnegie Mellon University conducted at the request of the United States Office of the Director of National Intelligence (ODNI). Our mission was simple: understand how organizations across sectors conduct the work of cyber intelligence and share our findings.

In this report, we describe the practices of organizations that are performing well and the areas where many organizations struggle, and we identify the models, frameworks, and innovative technologies driving cyber intelligence today. We believe this report can provide a starting point to enable organizations across the country to adopt best practices, work together to fix common challenges, and reduce the risk of cyber threats to the broader cyber community.

## WHO SHOULD READ THIS REPORT?

We have designed this report to be informative for anyone concerned with cyber threats. The following readers will find this report useful:

- Organizational Decision Makers: understanding where to direct funding and resources
- Cyber Intelligence Team Managers: understanding best practices for your team, including hiring, workflow, and leveraging data
- Cyber Intelligence Analysts: understanding best practices, tools for analysis, and what your peers are doing

Whether your organization has a robust cyber intelligence program or is just getting started, the actionable recommendations provided in each section of this report can serve as guideposts for helping you achieve high performance.

## THE IMPORTANCE OF CYBER INTELLIGENCE

***Cyber intelligence:*** *acquiring, processing, analyzing, and disseminating information that identifies, tracks, and predicts threats, risks, and opportunities inside the cyber domain to offer courses of action that enhance decision making.*

**CYBER INTELLIGENCE DEFINED**

Your organization may protect the confidentiality, integrity, and availability of data and computer systems. Such practices are part of cybersecurity. However, do you know which threat actors have the intent and capability to target your organization now and in the future? Do you track malware campaigns? Do you know which of your technologies are at risk? Do you know how certain attacks would affect your organization? Do you perform supply chain analysis, produce targeting packages for your pen-testing team, or provide assessments on the impact/opportunity of emerging technologies? Are you able to produce threat priority and vulnerability lists or industry threat assessments? Do you know if your organization should open a line of business in a foreign country? Cyber intelligence can provide this insight to protect your organization.

## TERMINOLOGY

In this report, we use the following terms and definitions:

- **Cyber Hygiene:** "Activities such as inventorying hardware and software assets; configuring firewalls and other commercial products; scanning for vulnerabilities; patching systems; and monitoring."[2]
- **Cybersecurity:** Actions or measures taken to ensure a state of inviolability of the confidentiality, integrity, and availability of data and computer systems from hostile acts or influences.[3] The term "cyber hygiene" is sometimes referred to as both cybersecurity and as actions to improve cybersecurity.
- **Cyber Threat Intelligence:** Intelligence analysis on threats in the cyber domain. Cyber intelligence includes cyber threat intelligence, but cyber threat intelligence does not represent all of cyber intelligence.[4]
- **Data:** "A set of values of subjects with respect to qualitative or quantitative variables."[5] "Data can be any character, text, word, number, and, if not put into context, means little or nothing to a human."[6]

**TIP**

**See the Glossary** for more terms and definitions.

2 https://www.nist.gov/blogs/taking-measure/rethinking-cybersecurity-inside-out

3 The definition for cybersecurity created based on analyzing participating organizational responses and from the DHS Lexicon Terms and Definitions Instruction Manual 262-12-001-01 (October 16, 2017) https://www.dhs.gov/sites/default/files/publications/18_0116_MGMT_DHS-Lexicon.pdf

4 A number of organizations expressed confusion over the difference between cyber threat intelligence and cyber intelligence, specifically whether these terms describe the same thing. Many organizations told us that introducing "threat" into this phrase breeds that confusion. Although threats are a large part of the cyber intelligence picture, cyber intelligence also includes analysis of areas like technologies, geopolitics, and opportunities. For these reasons, this report deliberately excludes the term "cyber threat intelligence." We refer to the activities typically associated with cyber threat intelligence as Threat Analysis, a component of the Cyber Intelligence Framework.

5 https://en.wikipedia.org/wiki/Data

6 https://www.computerhope.com/issues/ch001629.htm

- **Information:** "Data formatted in a manner that allows it to be utilized by human beings in some significant way."[7]
- **Intelligence:** "1. The product resulting from the collection, processing, integration, evaluation, analysis, and interpretation of available information concerning foreign nations, hostile or potentially hostile forces or elements, or areas of actual or potential operations. 2. The activities that result in the product. 3. The organizations engaged in such activities."[8]

## CYBER INTELLIGENCE FRAMEWORK

### ENVIRONMENTAL CONTEXT

A deep understanding of your organization, including your organization's entire attack surface; threats, risks, and opportunities targeting your organization and industry; and your organization's internal and external network and operations. Gaining this understanding is a continuous process and influences what data is needed to perform cyber intelligence.

### DATA GATHERING

Through automated and labor-intensive means, data and information is collected from multiple internal and external sources for analysts to analyze to answer organizational intelligence requirements.

### THREAT ANALYSIS

Assessing technical telemetry and non-technical data pertaining to specific threats to your organization and industry to inform cybersecurity operations/actions and Strategic Analysis. Threat Analysis is built on operational and tactical analysis and enhances CSO/CISO and other mid- to senior-level decision making.

### STRATEGIC ANALYSIS

Holistically assessing threats, risks and opportunities to enhance executive decision making pertaining to organization-wide vital interests such as financial health, brand, stature, and reputation.

### REPORTING AND FEEDBACK

Communication between analysts and decision makers, peers, and other intelligence consumers regarding their products and work performance. Reporting and feedback help identify intelligence requirements and intelligence gaps.

### HUMAN-MACHINE TEAMING

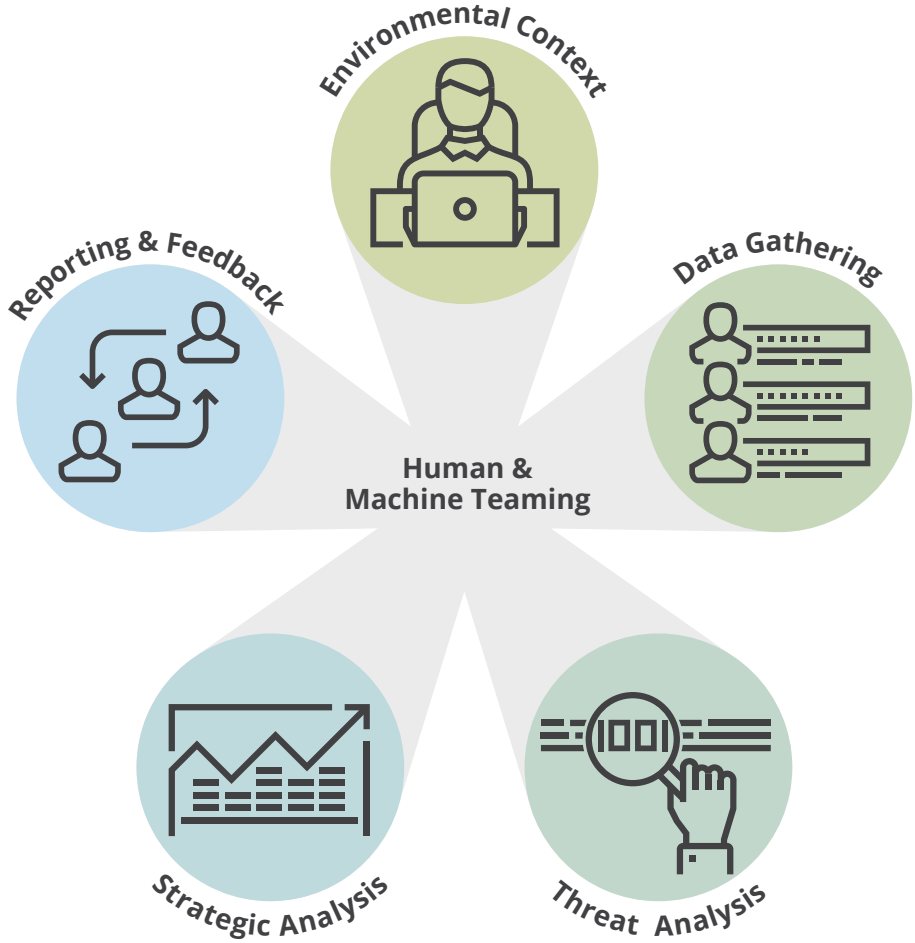
At the center of the cyber intelligence framework, human analysts use their analytical acumen alongside the computational power and speed of machines—computers able to automate processes and, increasingly, to learn through artificial intelligence—to produce timely, actionable, and accurate intelligence, depending on the cyber issue being analyzed.

---

7 ibid.

8 https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf

## CYBER INTELLIGENCE FRAMEWORK




**Cyber Intelligence Framework** Rooted in the U.S. Government's traditional intelligence cycle, the analytical framework above provides a structure for cyber intelligence efforts and forms the basis for the concepts in this study.

## BACKGROUND: 2013 CYBER INTELLIGENCE STUDY

This study is a follow-up to a similar cyber intelligence study the SEI conducted at the request of ODNI in 2013. The ***Cyber Intelligence Tradecraft Project: Summary of Key Findings*** report highlights cyber intelligence best practices and biggest challenges we observed several years ago. We used our 2013 findings as a foundation for the most recent study, and as a baseline to understand changes in cyber intelligence practices over the years. In this report, we point out areas where cyber intelligence practices are improving rapidly and areas where progress has been almost glacial.

## HOW WE CONDUCTED THE STUDY

To understand the state of cyber intelligence practices nationwide, we set out to interview companies and organizations about their cyber intelligence practices. Using our 2013 report as a foundation, we developed updated interview questions rooted in the five components of our 2013 cyber intelligence framework: Environmental Context, Data Gathering, Threat Analysis, Strategic Analysis, and Reporting and Feedback. We asked conversational questions that helped us determine how well organizations were doing in relation to 33 assessment factors.

Our SEI team interviewed 32 U.S. organizations during sessions that ranged from 2-4 hours. We performed both on-site and virtual interviews of small, large, new, and established organizations representing a variety of critical infrastructure sectors: Finance, Health and Public Health, Information Technology, Communications, Food and Agriculture, Commercial Facilities, Government Facilities, Energy, Defense Industrial Base, Transportation, and Academia. We interacted with representatives from these organizations' cyber intelligence and cybersecurity teams and leadership.

After completing all of the interviews, our team benchmarked the data we collected against the 33 assessment factors within the five components of the cyber intelligence framework. We compiled an extensive list of the challenges and best practices interview participants shared (a total of 2,268 items) and grouped them by themes. The resulting themes drive the content of this report.

## HOW WE UNDERSTOOD HIGH PERFORMANCE

Using information from our 2013 study, we developed some baseline criteria for high performance. We refined and adjusted these criteria based on information from interviews we conducted during the current study to define the methodologies, technologies, and processes that constitute high performance in cyber intelligence today. We then scored performance according to the following scale:

| | |
|---|---|
| **High Performing:** | Organization meets all high-performing criteria. |
| **Almost High Performing:** | Organization generally meets all high-performing criteria, except one. |
| **Getting Started/Doing a Few Things:** | Organization generally meets one or two high-performing criteria. |
| **Low Performing:** | Organization meets no high-performing criteria. |
| **Insufficient Information:** | Insufficient information to make an assessment. |

## WHAT HAS CHANGED SINCE THE 2013 STUDY? WHAT HAS STAYED THE SAME?

### THE CYBER INTELLIGENCE FRAMEWORK

We changed some terminology within the Cyber Intelligence Framework. We first introduced the Cyber Intelligence Framework, rooted in the traditional intelligence cycle, in 2013, with the components Environment, Data Gathering, Functional Analysis, Strategic Analysis, and Decision Maker Reporting and Feedback. To reflect terminology we heard from participants, we changed Functional Analysis to Threat Analysis. Because we heard time and again from participants whose reporting and feedback practices involved a variety of individuals, especially at the peer level, we changed Decision Maker Reporting and Feedback to simply Reporting and Feedback.

## TRADITIONAL INTELLIGENCE CYCLE

Our recent research showed some high performing organizations using frameworks that are modeled on the traditional intelligence cycle and that successfully incorporate cutting edge technology into their cyber intelligence programs. These high-performing organizations have long established cyber intelligence programs and foster a complete people, processes, and technologies approach to cyber intelligence. In contrast to our 2013 report, which described the traditional intelligence cycle as limited by its linear format, we now assess the traditional intelligence cycle as an interrelated and non-linear process. The success and failure of one or more steps in the cycle may spawn a rippling effect on the entire cycle. The traditional intelligence cycle is therefore an acceptable way for organizations to approach cyber intelligence; our cyber intelligence framework is ideal because it addresses the intersection and pervasiveness of cyber and technology.

## GAP BETWEEN TECHNICAL AND ANALYTICAL EXPERTISE

A gap remains and is widening between individuals experienced in intelligence analysis and operations and those experienced in information security, computing fundamentals, and artificial intelligence. Some organizations have only technical people on their team with zero to little understanding, background, or training in intelligence analysis. Other organizations that employ individuals experienced in intelligence analysis and information security encounter stark cross-team communication challenges.

## INCREASED ADOPTION OF AUTOMATION AND ARTIFICIAL INTELLIGENCE

Computing hardware and software is changing and improving every day; machines, with their computational power and speed, have the potential to transform cyber intelligence. As organizations create and have access to more data, these organizations are increasingly adopting automation and artificial intelligence. Specifically, many are using machine learning to assist human analysts with understanding their environment, data collection, analysis, and report generation.

*"Knowing yourself is the beginning of all wisdom."*

—Aristotle

# Environmental Context

*Understanding Your Organization Inside and Out*

## INTRODUCTION

A cyber intelligence team should have a deep understanding of its organization's entire attack surface; threats, risks, and opportunities relevant to the organization and industry; and the impact of those threats, risks, and opportunities. Environmental Context refers to this understanding, which requires knowledge of your organization's internal and external network and operations, including services, operating systems, endpoints, mission and culture, processes and policies, business partners, suppliers, geopolitics, emerging technologies, and position in industry relative to competitors. Because your environment is constantly changing, gaining and maintaining this understanding is a continuous process.

### ENVIRONMENTAL CONTEXT ASSESSMENT FACTORS

In evaluating the state of the practice of cyber intelligence in terms of Environmental Context, we considered the following factors:

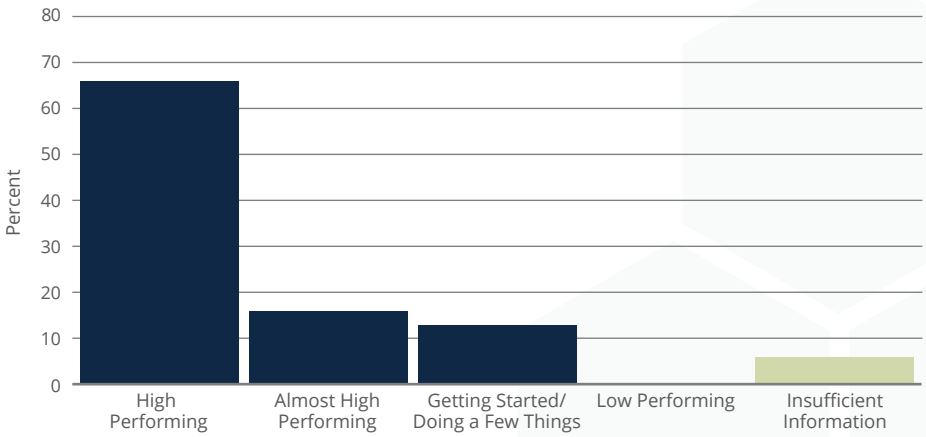
1. Knowing Your Attack Surface
2. Understanding the Difference Between Cyber Intelligence and Cybersecurity
3. Aligning Cyber Intelligence Roles with Your Organization's Needs
4. Having Enough People, Having the Right People
5. Placement of Your Cyber Intel Effort in Your Organization
6. Cyber Intelligence Workflow
7. Threat Prioritization Process
8. Using Past, Present, and Future Data
9. Relationship Between Cyber Intelligence and Insider Threat Teams

## ENVIRONMENTAL CONTEXT FACTOR 1: KNOWING YOUR ATTACK SURFACE

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization holistically understands its people (including relevance and access) and cyber footprint (including infrastructure, internet presence, physical assets and access, and technology). This understanding informs the tactics, techniques and procedures (TTPs) the organization implements to support cybersecurity and cyber intelligence.

## PERFORMANCE SNAPSHOT




**Environmental Context Factor 1 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

### Silos blind

A major challenge we observed across organizations was silos. In some organizations, internal business units have separate, distinct IT systems. These business units may not communicate or share data efficiently because IT systems and technology stacks are completely different. Cultural differences and network fragmentation among internal business units exacerbate the effects of silos.

A related challenge is the inability to actively and continuously monitor third parties due to policy and IT architecture and technology stack differentiations. Without visibility into the activities and services of partners, suppliers, and sub-contractors, cyber intelligence teams cannot know how threat actors—and which threat actors—could exploit vulnerabilities within their attack surface.

### Inability to identify and track important organizational data presents dangers

Many organizations have trouble identifying the location of confidential and intellectual property data, how data moves across the organization, and when and how individuals interact with it. Many study participants expressed frustration over not having a data loss prevention (DLP) tool. These organizations tended to also lack formalized insider threat programs. Although access control lists help to prevent unauthorized access, they cannot, for example, easily detect an insider stealing 40 pages of sensitive information at a slow rate.

**IMPROVE YOUR PERFORMANCE**

- Conduct a crown-jewel exercise to identify critical assets.
- Work with cybersecurity teams to know and monitor the users accessing your network, the data they use, and their computing equipment.
- Promote regular sharing among your Information Technology, Technology Development and Integration, Cyber Intelligence, Program Management, Security Operations Center, and Security Engineering and Asset Security teams. See Environmental Context Factor 5 for more information.
- Hold daily standup meetings, calls, or video conferences.
- Create a physical or virtual fusion center.

## BEST PRACTICES

### Know your critical assets

High-performing cyber intelligence teams demonstrate a keen understanding of their organization's critical assets, from network

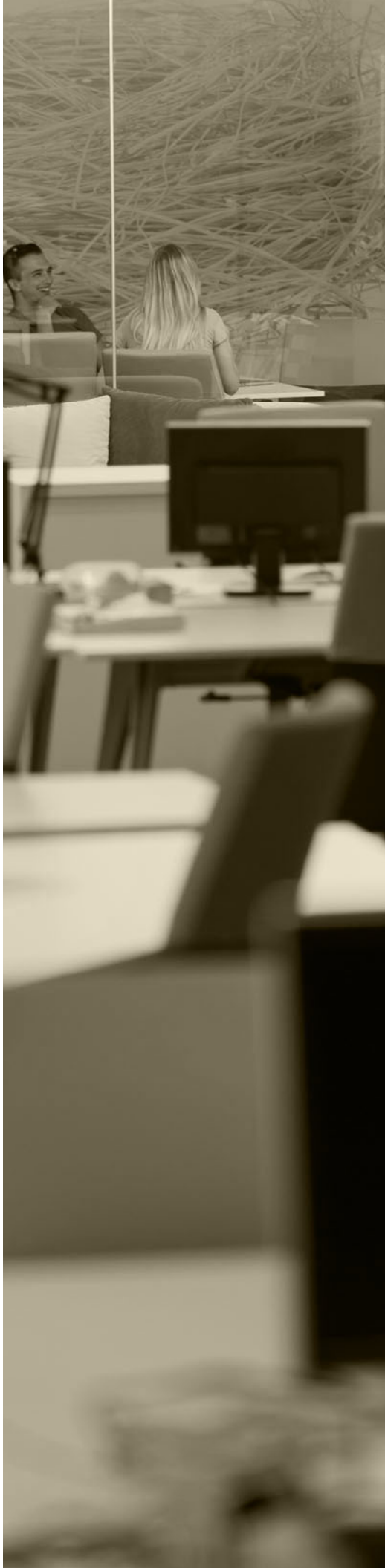
endpoints to patent pending technologies. These teams understand information technology and operational technology assets (such as industrial control and supervisory control and data acquisition systems), infrastructure, and the convergence and associated vulnerabilities between the two. These organizations understand their internet-facing systems, data centers, cloud-based components, network infrastructure, servers, hosts, portals, mobile platforms, and internet of things and other embedded technologies; and they keep track of their hardware and software inventory via a number of commercially available IT asset management and operational technology monitoring solutions.

Conducting a crown-jewel exercise or analysis can help you understand your critical assets, which range from sensitive technologies to data types moving and resting within your organization. During the course of the exercise, you'll identify the assets themselves, their owners, the risk to your organization if they are compromised, and how they interact with other assets. High-performing organizations reported using existing models for crown-jewel exercises[9] or developing their own crown-jewel exercises by meeting and building relationships with colleagues working on critical assets or patent-pending technologies. For organizations just starting out, the crown-jewel exercise can provide a foundation for building a cyber intelligence effort.

Don't forget about people. High-performing cyber intelligence teams know their organization's employees, contractors, executives, and business partners—and how these individuals access the organization's network and data. High-performing organizations use DMZs and internal and external firewalls for instances where their own employees access internet-facing systems. These organizations use DLP, security information and event management (SIEM), and user and entity behavior analytical (UEBA) tools to identify abnormal behavior across users and services such as simple mail transfer protocol (SMTP), file transfer protocol (FTP), Telnet, virtual private network (VPN), webmail, and Remote Desktop, as well as exposures from Wi-Fi hotspots and rogue access points.

**Explore creating a fusion center**

High-performing cyber intelligence teams build strong relationships with cybersecurity teams and across organizational business units. A "fusion center" is a model for bringing together diverse teams to analyze disparate information. Virtual or physical fusion centers



9 NIST IR 8179 Criticality Analysis Process Model: Helping Organizations Decide Which Assets Need to Be Secured First, NIST Special Publication 1800-5 IT Asset Management, and NIST Special Publication 800-171 Protecting Controlled Unclassified Information in Nonfederal Systems and Organizations

facilitate interaction between the cyber intelligence team, cybersecurity team, and other teams such as network defense, vulnerability management, forensics, incident response, penetration testing, and insider threat. In a fusion center, these teams are often physically co-located, and report on their current work and observations in daily standup meetings.

**Foster cross-functional collaboration**
Some high-performing teams meet and collaborate daily with other internal business units such as human resources, governance and compliance, information technology, software development, physical security, and business development and marketing. Formal and informal relationships give the cyber intelligence team a holistic understanding of the organization's environment and future business direction, such as the release of patented technologies, the roll-out of software, and significant mergers or acquisitions. With an understanding of developments in these areas as well as business unit needs and requirements, the cyber intelligence team can provide relevant cyber intelligence reporting these teams and to managers and executives to aid in decision making.

## CREATING A FUSION CENTER

Creating a fusion center takes time, dedication, and resources. There are many ways to create a fusion center; some fusion centers come together organically while others form at the direction of leadership. The implementation and organizational structure of the fusion center should be specific to the organization. On the next page, we provide some examples of how organizations of various sizes and stages of maturity may structure a fusion center, and the teams fusion centers may add as they mature. These examples are based on information from our interviews as well as the SEI technical note *Structuring the Chief Information Security Officer Organization*[10] and specific roles and positions from *NIST-NICE Standard Practice 800-181*.[11]

**Physical or virtual?**
Organizations we interviewed described advantages of physical and virtual fusion centers. Physical fusion centers have the obvious advantage of allowing individuals across teams to literally turn their chairs and talk with their coworkers to develop meaningful relationships based on working together in the same space and cultural environment.

High-performing organizations described two key advantages to virtual fusion centers: attracting and retaining talent, and forcing collaboration. In a job market where it is difficult to hire and keep skilled cyber intelligence team members, a virtual fusion center can both expand options for attracting talent and provide flexibility to aid in retention. When employees can work from anywhere, an organization can hire from everywhere. Team members can live where cost of living is lower and can easily relocate based on family needs or interests. In addition, the very nature of virtual fusion centers makes collaboration a given. Virtual fusion centers support proactive communication with a variety of tools (e.g., Slack, Skype, a shared threat intelligence platform), and team members hold daily and weekly standups.

10 https://resources.sei.cmu.edu/asset_files/TechnicalNote/2015_004_001_446198.pdf

11 https://doi.org/10.6028/NIST.SP.800-181

## EVOLUTION OF A FUSION CENTER

The following chart presents an approach for creating a fusion center. Organizations just starting out should consider creating a fusion center with the "Beginning" components and positions. The numbers shown in the position titles are specific roles and positions from *NIST-NICE Standard Practice 800-181*.

### BEGINNING

**Security Operations**

- Hunt
- Vulnerability
  *Vulnerability Assessment Analysts: PR-VAM-001*
- Host and Network Security Monitoring
- Incident Response
  *Cyber Defense Incident Responder: PR-CIR-001*

**Security Engineering and Asset Security**

- Host and Network Security
- Malware and Forensics Analysis
- Physical Access Control
- Information Asset Security
- Identity and Access Management
- Applications Security
- Security Engineering

### DEVELOPING

**Security Operations**

- Hunt
- Vulnerability
  *Vulnerability Assessment Analysts: PR-VAM-001*
- Host and Network Security Monitoring
- Incident Response
  *Cyber Defense Incident Responder: PR-CIR-001*

**Security Engineering & Asset Security**

- Host and Network Security
- Malware and Forensics Analysis
- Physical Access Control
- Information Asset Security
- Identity and Access Management
- Applications Security
- Security Engineering

**Cyber Intelligence**

- Operational and Tactical
  *Threat/Warning Analysts: AN-TWA-001*
  *Cyber Defense Forensics Analyst: IN-FOR-001*
  *Cyber Defense Analysts: PR-CDA-001*

**Insider Threat**

**Physical Security**

**Program Management**

- Program Management Office
  *Mission Assessment Specialist: AN-ASA-002*
- Governance, Risk and Compliance
  *Cyber Legal Advisor: OV-LGA-001*
  *Privacy Officer / Compliance Manager: OV-LGA-002*
- Internal and External Relationships
  *Partner Integration Planner: CO-OPL-003*
- Business Development and Marketing

### MATURE

**Security Operations**

- Hunt
- Vulnerability
  *Vulnerability Assessment Analysts: PR-VAM-001*
- Host and Network Security Monitoring
- Incident Response
  *Cyber Defense Incident Responder: PR-CIR-001*

**Security Engineering & Asset Security**

- Host and Network Security
- Malware and Forensics Analysis
- Physical Access Control
- Information Asset Security
- Identity and Access Management
- Applications Security
- Security Engineering

**Program Management**

- Program Management Office
  *Mission Assessment Specialist: AN-ASA-002*
- Governance, Risk and Compliance
  *Cyber Legal Advisor: OV-LGA-001*
  *Privacy Officer / Compliance Manager: OV-LGA-002*
- Internal and External Relationships
  *Partner Integration Planner: CO-OPL-003*
- Business Development and Marketing

**Cyber Intelligence**

- Threat Analysis
  *Threat/Warning Analyst: AN-TWA-001*
  *Cyber Defense Forensics Analyst: IN-FOR-001*
  *Cyber Defense Analyst: PR-CDA-001*
- Collection Management
  *Cyber Intelligence Planner: CO-OPL-001*
  *All Source Collection Manager: CO-CLO-001*
  *All Source Collection Requirements Manager: CO-CLO-002*
- Strategic Analysis
  *All Source Analyst: AN-ASA-001*
  *Strategic Analyst*
  *Geopolitical Analyst*
  *Intelligence Analyst*
  *Data Analysts: OTM-DTA-002*

**Insider Threat**

**Physical Security**

**Technology Development & Integration**

- Data Science and Machine Learning
  *Data Analysts: OTM-DTA-002*
  *Machine Learning Engineer*
- Software Application and Development
  *Research and Development Specialist: SP-TRD-001*
  *Software Developer: SP-DEV-001*
- Knowledge Management
  *Knowledge Manager: OM-KMG-001*

Key

**Groups**

Team

*Positions*

**Consider hiring a dedicated physical security analyst**
Study participants told us that physical intelligence is the highest-volume, lowest-yield intelligence available, with countless Internet user comments that could constitute threats to physical assets. The alerting makes an enormous amount of work for analysts, and the subjective nature of potential physical threats makes automated detection difficult. That said, organizations are increasingly concerned about physical threats to their organization and are dedicating resources to provide intelligence about them.

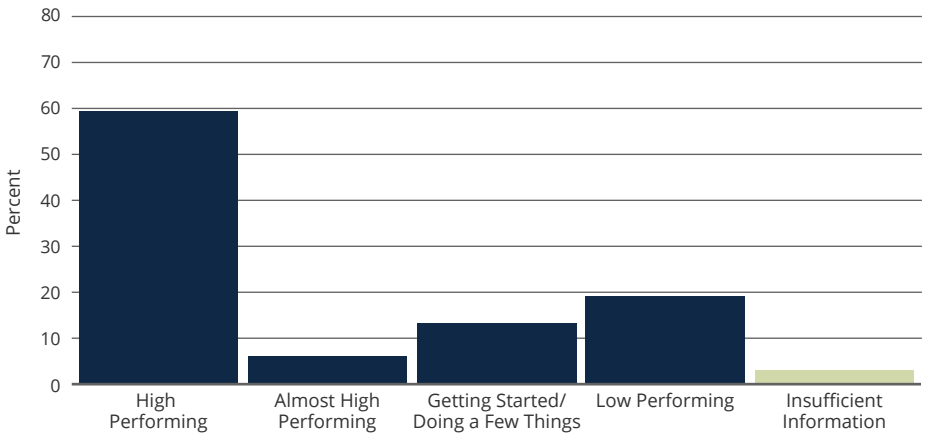
A practice of high-performing organizations is to have a dedicated physical security analyst, sometimes within their fusion center, to provide intelligence on physical threats that could cause harm to the organization's people, operations, and brand. The analyst provides intelligence on threats to the organization's physical locations and partner locations across the globe. Threats can range from malicious cyber actors looking to inflict physical harm, internal foreign country developments (geopolitics), and natural disasters impacting business operations.

## ENVIRONMENTAL CONTEXT FACTOR 2: UNDERSTANDING THE DIFFERENCE BETWEEN CYBER INTELLIGENCE AND CYBERSECURITY

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization establishes and maintains cyber intelligence and cybersecurity as two work functions distinguished in their mission, purpose, roles, and responsibilities. Entities performing these two work functions interact and collaborate proactively to run the organization's cyber efforts.

### PERFORMANCE SNAPSHOT




**Environmental Context Factor 2 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

### Misunderstanding cyber intelligence

While some organizations might receive third-party intelligence daily feed(s), resources constraints mean that they can improve their organizations' security only through cyber hygiene actions. Failing to create a distinct cyber intelligence team puts your organization at increased risk for harm because you are constantly in a reactive position.

### Lack of communication between cybersecurity and cyber intelligence teams

Some cyber intelligence teams explained that communication and collaboration with the organization's other cybersecurity functions is inefficient. In the absence of fusion centers or other collaboration mechanisms, communication may be one-way or may occur only through email and chat, hampering collaboration and cyber intelligence performance.

### Fusion centers that lack cyber intelligence functions

In some organizations, fusion centers resemble operations centers, which consist of cybersecurity teams (vulnerability, incident response, and hunt teams) that typically reside in a security operations center (SOC). These fusion centers do not include cyber intelligence or other teams (physical security, knowledge management, insider threat, technology development teams).

**IMPROVE YOUR PERFORMANCE**

- Get CISO and Board support to create a cyber intelligence team that has a clear mission. Define and document roles and responsibilities that are approved and understood by the entire organization.
- Build relationships with leadership to help promote your team across the organization.
- Build relationships with business unit leaders to get buy-in on the need for a fusion center.
- Exchange ideas with colleagues in cybersecurity, IT, intelligence, technology development, software development, and physical security.

## BEST PRACTICES

### Create a defined cyber intelligence team

High-performing organizations build cyber intelligence teams that have their own mission, purposes, roles, and responsibilities. Mission, purpose, roles, and responsibilities are matured and approved by the Chief Information Security Officer and the board and are documented and accessible to the team and throughout the organization. They are evaluated bi-annually to ensure the team's support to the organization is consistent, meaningful, and lasting.

## TERM CLARITY

### Fusion Center

- Multiple teams of different disciplines
- Located in one physical/virtual location
- Proactively collaborating: information sharing and analysis
- Advances organization-wide decision making for
  - cybersecurity operations
  - preventive and anticipatory actions based on Threat Analysis
  - organizational vital interests based on Strategic Analysis
- Engages entire organization and external partners

### Operations Center

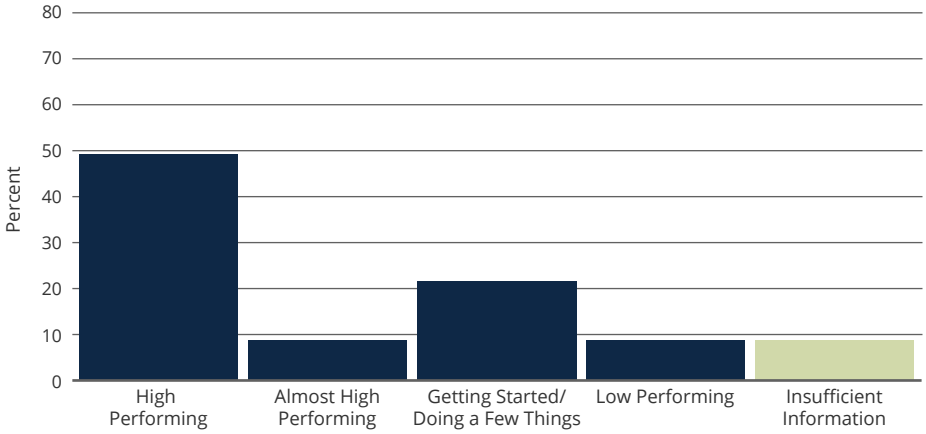
- Multi-disciplined staff
- One or more teams in one physical/virtual locations
- Focused on cybersecurity operations and Threat Analysis (for example, detecting and responding to incidents, maintaining the current status of operations, and tactical and operational analysis of possible threats)
- Often a component within a fusion center

## ENVIRONMENTAL CONTEXT FACTOR 3: ALIGNING CYBER INTELLIGENCE ROLES WITH YOUR ORGANIZATION'S NEEDS

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization distinguishes between cybersecurity and cyber intelligence analysts. The organization clearly defines responsibilities for these individuals that support organizational needs in cybersecurity, cyber intelligence, and business mission needs.

### PERFORMANCE SNAPSHOT




**Environmental Context Factor 3 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**Unclear roles and responsibilities**

Some organizations lack clearly defined and documented roles and responsibilities for their cybersecurity and cyber intelligence teams. These organizations, (mostly smaller organizations) explained that while roles and responsibilities are conceptually understood, formal documentation and clarity regarding how roles and responsibilities align to support the overall organization mission were unclear or not established. The SEI team also met with organizations that, due to resource constraints, have roles and responsibilities strictly dedicated to cybersecurity efforts. These organizations usually have teams that consist of network monitoring analysts, vulnerability analysts, incident response analysts, hunt analysts, and forensic analysts.

## BEST PRACTICES

### Cross functional teams

High-performing organizations distinguish between and have a mix of cybersecurity and cyber intelligence analysts. These organizations clearly document and articulate each team member's role and responsibilities (defined by skill set, domain, or even product line) and map them to organizational needs. Team roles and responsibilities are visible and understood across the organization. Visibility streamlines processes and helps break down silos.

### Regular evaluation

High-performing cyber intelligence teams regularly evaluate (at least every six months) that they have the right personnel performing the right roles to support
the organization.

### Balancing technical skills and responsibilities with analytical expertise

Cyber intelligence teams should strike the right balance of having technical staff working alongside those who possess strong intelligence and geopolitical analysis and experience. Consider two types of analysts:

**Threat analysts** are highly technical; they use technical telemetry (internal/external atomic, behavioral, and computed indicators and artifacts[12]) to provide tactical and operational analysis regarding threats to the organization or industry to advance cybersecurity operations, and inform Strategic Analysis. Roles, responsibilities and skills typically associated with threat analysts are similar to those in NIST SP 800 181 for Cyber Defense Analysts or Threat Warning Analysts—position titles are sometimes used interchangeably.

**Strategic analysts** provide holistic intelligence assessments. These analysts produce intelligence rooted in Threat Analysis considered alongside other information (all-source intelligence) and analytical tradecraft (structured analytical techniques, data science, human-centered design activities). Example assessments relate to strategic threats, threat actors, risks, and opportunities and provide information for decision makers regarding the organization's vital interests. Roles, responsibilities, and skills typically associated with strategic analysts are similar to all-source intelligence analysts, intelligence analysts, threat actor analysts, risk analysts, or country and geopolitical analysts—position titles are sometimes used interchangeably.

**IMPROVE YOUR PERFORMANCE**

- Document team roles and responsibilities and map them to organizational needs.
- Ensure your cyber intelligence team has both strategic analysts (those who are well versed in intelligence, analytical tradecraft, emerging technologies, and geopolitics) and threat analysts (those who are well versed in technical analysis).
- Ensure your cyber intelligence team has access to data scientists and machine learning experts.

12 https://digital-forensics.sans.org/blog/2009/10/14/security-intelligence-attacking-the-kill-chain

### Use of data science and machine learning

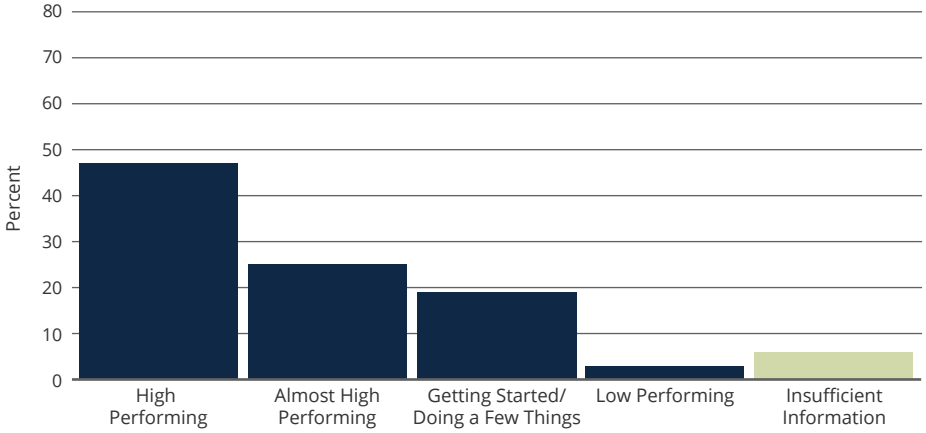
High-performing cyber intelligence teams have access to data scientists and machine learning experts and engineers, as members of their team or as resources they can call on from other parts of the organization. These experts help the team make sense of their data and automate processes and analysis.

## ENVIRONMENTAL CONTEXT FACTOR 4: HAVING ENOUGH PEOPLE, HAVING THE RIGHT PEOPLE

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization has the personnel to support its cyber intelligence needs. The cyber intelligence team has sufficient staff to surge and free time to perform self-initiated research. The organization consistently evaluates personnel needs against cyber intelligence needs to ensure that its cyber intelligence team members have expertise to meet those needs.

### PERFORMANCE SNAPSHOT




**Environmental Context Factor 4 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**"We need more people!"**

Well established and nascent cyber intelligence efforts share the challenge of personnel. Some organizations have a one-person cyber intelligence effort, and others are merely staying afloat in complete reactive mode. Without adequate personnel, teams lack the time and resources to do long-term holistic assessments or self-initiated research, and may not be able to surge to support cybersecurity efforts.

In many organizations that struggle with a lack of personnel, budget is a factor. Other organizations report that leadership does not recognize cyber intelligence as a worthy investment or does not understand the difference between cybersecurity and cyber intelligence.

#### IMPROVE YOUR PERFORMANCE

- Consider NIST SP 800-181 as a resource for building your cyber intelligence team.
- Give your analysts the freedom to explore and perform self-initiated research.

**Difficulties recruiting and retaining cyber intelligence professionals**
Organizations find it difficult to pay enough money to attract the right talent and increase salaries annually at a competitive rate to retain talent. Organizations in the finance sector especially noted the acquisition and retention of talent as a recurring challenge. This difficulty seems to arise in the financial sector because of intense competition among organizations that have robust cyber intelligence programs and can continually outbid one another for talent.

## BEST PRACTICES

**Leaders invest in cyber intelligence**
Organizations with a budget to hire cyber intelligence talent tend to be organizations where leadership values the importance of cyber intelligence.

**A variety of approaches and resources for staffing and surging**
High-performing organizations dedicate resources to surging for both cybersecurity and cyber intelligence efforts using in-house teams and third-party retainers. Some organizations cross-train between teams to provide an internal surge capability. One high-performing organization described training a floating surge force of generalists who can pick up slack anytime anywhere. Another organization is adopting a plan that uses interns to augment its cyber intelligence staff. These interns have cyber intelligence, cybersecurity, and intelligence analysis experience and education. Last, a common practice of high-performing cyber intelligence teams is to have veteran cybersecurity and intelligence analysts train less experienced analysts.

**The right personnel**
In our 2013 report, we noted that high-performing organizations were pairing traditional intelligence analysts with cybersecurity and other technical analysts to ensure analytical tradecraft and Strategic Analysis was formulated into the cyber intelligence team's workflow. This approach is still a best practice. Many organizations are now hiring data scientists and machine learning experts as part of a technology development and integration team. These individuals work with the cyber intelligence team as team members or collaborators; they help derive meaning out of large data lakes and build in-house customizable tools to assist analysts with pattern and prediction analysis.

**Mapping position requirements to NIST/NICE Cybersecurity Workforce Framework**
A practice of high-performing organizations is to map position requirements to National Institute of Standards and Technology (NIST) Special Publication 800-181: *National Initiative for Cybersecurity Education (NICE) Cybersecurity Workforce Framework*[13] categories. Positions and roles highlighted in NIST SP 800-181 are designed to strengthen the cybersecurity posture of an organization.

13 https://nvlpubs.nist.gov/nistpubs/specialpublications/nist.sp.800-181.pdf

### BUILDING A HIGH-PERFORMING TEAM

Although organizations should tailor staffing to their own needs, the following positions—based on NIST/NICE 800-181 and information from study participants—can help organizations achieve high performance.

**Cyber Intelligence Team**
- All Source-Collection Requirements Manager
- All Source-Collection Manager
- All Source Analyst
- Cyber Intelligence Planner
- Multi-Disciplined Language Analyst
- Threat/Warning Analyst
- Threat Analyst
- Strategic Analyst
- Cyber Defense Forensics Analyst
- Geopolitical Analyst

**Cybersecurity Team or Security Operations Team**
- Cyber Defense Incident Responder
- Cyber Defense Analyst

**Technology Development and Integration Team**
- Data Analysts
- Machine Learning Engineer
- Software Developer
- Research and Development Specialist
- Knowledge Manager

**Program Management**
- Mission Assessment Specialist
- Partner Integration Planner
- Privacy Officer
- Cyber Legal Advisor

**Create a culture of innovation**

Organizations that encourage exploration and innovation tend to have high-performing cyber intelligence teams. Proactive self-initiated research, with top-down encouragement and approval, leads cyber intelligence team members to identify new threat actors targeting the organization and to develop new tools and solutions for addressing complex problems. One high-performing cyber intelligence team allows each analyst two research weeks each year to work on a project of their choice. Another high-performing cyber intelligence team requires self-initiated research every day as a scheduled activity.

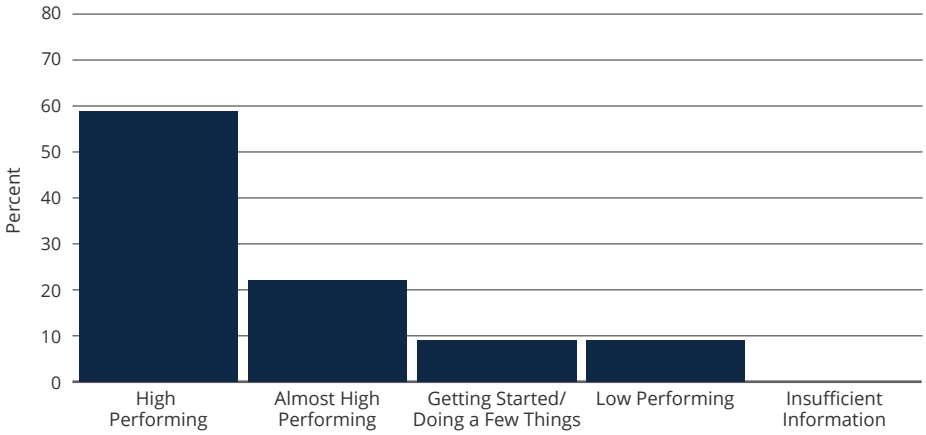
A culture of innovation not only leads to useful tools and solutions, but also gives cyber intelligence team members the chance to be proactive and the freedom to showcase their creative skills and ideas. In this way, retaining great people becomes less of a challenge.

## ENVIRONMENTAL CONTEXT FACTOR 5: PLACEMENT OF YOUR CYBER INTEL EFFORT IN YOUR ORGANIZATION

### WHAT THIS ASSESSMENT FACTOR MEANS

The cyber intelligence team has consistent access to teams and decision makers throughout the organization as well as associated data.

## PERFORMANCE SNAPSHOT




**Environmental Context Factor 5 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

**Aligning cyber intelligence too closely with cybersecurity**

In 2013, we found that the cyber intelligence team's organizational location affected its focus and performance; this finding holds true for organizations today. Cyber intelligence teams should be closely aligned with functions where they can influence strategic decision making (for example, risk management). However, organizations often align cyber intelligence with security operations and network management, relegating their analysts to reactive, technical tasks supporting cybersecurity.

Organizations that struggle in this area commonly take a "cybersecurity plus" approach to cyber intelligence: they may add a cyber intelligence analyst or a budding intelligence effort within or below a cybersecurity team. As a result, the cyber intelligence analyst may end up reporting to a security operations center (SOC) team lead or other manager focused on cybersecurity, which may limit the analyst to a reactive approach.

**Unnecessary bureaucracy**

Organizations we interviewed reported widespread difficulties with layers of management that prevent them from getting intelligence to the right people in a reasonable timeframe, and from getting approvals for new tools or research ideas. For example, one organization reported that its cyber intelligence team analysts report to the team manager, who reports to the lead for physical security, who then reports to the chief information security officer (CISO). The CISO for this organization often tasks the cyber intelligence team directly to circumvent the bureaucracy and get quick answers.

## BEST PRACTICES

**Elevate the CISO position**

A common organizational structure is for the cyber intelligence team to report to the chief security officer (CSO) or CISO,[14] who then reports to the chief information officer (CIO), who then reports to the chief executive officer (CEO), who sits on the board of directors. This structure can perpetuate challenges related to locating cyber intelligence too closely to IT or cybersecurity efforts. High-performing organizations elevate their CISOs, giving them the ability to report directly and frequently to the CEO and board of directors. A growing body of research and reporting describes the advantages of this approach.[15]

Different organizations elevate their CISOs in different ways. For some high-performing organizations, the cyber intelligence team lead (chief of cyber intelligence) has direct, easy, and ongoing formal and informal access to the CSO/CISO. The CSO/CISO has this same level of direct and easy access to the CEO. In other high-performing organizations, the CSO/CISO also sits on the board of directors. In this structure, leadership is very much engaged, and the cyber intelligence team can provide intelligence in a timely and efficient manner to advance organization-wide business decisions.

**Augment your fusion center with an enterprising capability**

Fusion centers, described in Environmental Context Factor 1, help information flow to the right people at the right time; they increase information sharing efficiency, speed the leadership approval process, and ensure everyone is collaborating and on the same page. Some high-performing organizations with fusion centers go a step further, embedding cyber intelligence analysts in organizational lines of business like human resources, legal, business development, public relations, finance, and contracts. These individuals sit with the business units and explain cyber threats to the organization, take specific requests for information, and provide tailored cyber intelligence products to the business unit.

# ENVIRONMENTAL CONTEXT FACTOR 6: CYBER INTELLIGENCE WORKFLOW

## WHAT THIS ASSESSMENT FACTOR MEANS

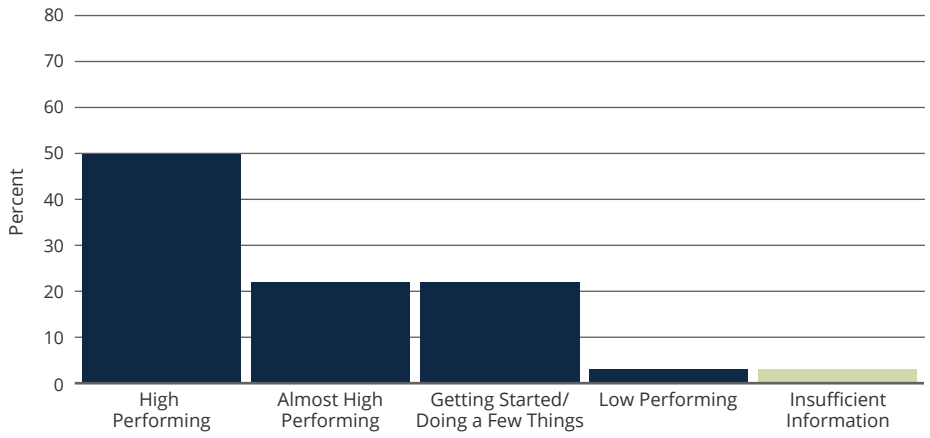
The organization has an established and repeatable process that accounts for environment, data gathering, Threat Analysis, Strategic Analysis, and reporting and feedback components. This process is reviewed and updated regularly.

---

14 Although the CSO and CISO are distinct positions with distinct roles, many organizations use the terms interchangeably in practice. Broadly speaking, the CSO/CISO is responsible for strategically managing and providing risk guidance associated with physical, people, and asset security as well as cybersecurity.

15 https://www.isc2.org/-/media/FAA17021673C4D0387CE9EFD45009EBC.ashx
https://www.fsisac.com/article/fs-isac-unveils-2018-cybersecurity-trends-according-top-financial-cisos
https://er.educause.edu/articles/2018/6/its-time-to-set-cisos-free
https://www.linkedin.com/pulse/cio-report-ciso-why-j-j-guy
http://www.bankinfosecurity.com/blogs/role-reversal-cio-reports-to-ciso-p-1648
https://www.cio.com/article/3247251/cio-role/goals-for-cios-in-2018.html
https://www.forbes.com/sites/forbestechcouncil/2018/01/09/the-evolving-role-of-the-cio-in-2018/#48b459a21c8e

## PERFORMANCE SNAPSHOT




**Environmental Context Factor 6 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

### Conceptual or incomplete cyber intelligence workflows

Many organizations lack a formalized, documented, and repeatable cyber intelligence workflow. Some of these organizations explained that their workflow is largely conceptual and exists in the minds of team members.

A related challenge is incomplete cyber intelligence workflows that most commonly omit Strategic Analysis. Teams in organizations with incomplete workflows often conduct Strategic Analysis only if time is permitted, or if the organization has a distinct separate team of analysts capable of performing that level of analysis. Other organizations have separate workflows for each specific team (incident response team, SOC team, vulnerability management team, forensics team), and these distinct workflows do not join into a single comprehensive cyber intelligence workflow. Still other organizations had reactive workflows that were documented and formalized, yet only for cybersecurity and incident response.

**IMPROVE YOUR PERFORMANCE**

- Incorporate the Cyber Intelligence Framework as a guide to perform cyber intelligence.
- Define and document your workflow to ensure that it is repeatable.

## BEST PRACTICES

### Use the Cyber Intelligence Framework to perform cyber intelligence

High-performing organizations account for all Cyber Intelligence Framework components in workflows that are written down, easy to find, and clearly show how each team contributes. The following list shows practices described by high-performing organizations at every step of the Cyber Intelligence Framework.

## BEST PRACTICES FOR WORKFLOWS THROUGHOUT THE CYBER INTELLIGENCE FRAMEWORK

**Environmental Context—Planning and Direction**

- Understand current organizational exposure to the threat because of vulnerabilities (Risk): People + Cyber Footprint + Physical + Technology
- Conduct crown-jewel exercise for critical asset and sensitive technology identification
- Understand organization's entire internal and external networking infrastructure, including associations with partners and suppliers
- Understand organization's mission, industry, and role within industry
- Identify and align gaps and requirements: intelligence requirements, priority intelligence requirements, and specific intelligence requirements
- Cyber intelligence team creates and manages request-for-information (RFI) process
- Cyber intelligence team **owns** the intelligence requirement process for the **entire** organization

**Data Gathering—Collection, Processing, and Exploitation**

- Collect technical telemetry from internal sources (e.g., SIEM, SOAR, all logs) and external sources (e.g., third-party providers, publicly available information, classified sources) to answer SIRs and PIRs.
- Strategic Analysis: Incorporate Threat Analysis and collect other non-technical information, including geopolitics, business intelligence, human resources data, research and development data, physical security data, and social media.

**Threat Analysis—Analysis and Production**

- Collect technical telemetry from internal sources.
- High-performing organizations have Threat Analysis workflows (or playbooks) to support time-sensitive and action-oriented decisions for network and host monitoring, vulnerability management, and incident response.
- Workflows are defined, documented, repeatable, and scalable
- Indicators of Compromise (IOCs)—atomic, behavioral, and computed[16]—are automatically correlated and matched against internal network and endpoint telemetry activity; automated data enrichment through integrated internal platforms, and external integrations
- Machine or analyst alerts senior analyst or another machine for decision on elevating—A "yes" decision leads to triggering an automated workflow within security information and event management/threat intelligence platform (SIEM/TIP) playbook integrations or security orchestration and automation response (SOAR), or Jira solution
- Lead analyst(s) assigned adds context (additional current and historical data) creating tactical analysis to answer what/where/when/how questions regarding threats, attacks, incidents, vulnerabilities, or other unusual network activity for the purpose of generating human and machine mitigating actions.
- Depending on event and time constraints, fusion center analysts perform operational analysis, adding context to existing tactical analysis (threat actors, campaigns) to start to answer the who and why behind threats
- Enhance mid- to senior-level leadership decisions regarding non-immediate but near-term (weekly–quarterly) business process and operational decisions.

16 https://digital-forensics.sans.org/blog/2009/10/14/security-intelligence-attacking-the-kill-chain

**Strategic Analysis**

- Fuse Threat Analysis with other external and non-traditional data sources
- Depending on data collected, work with data science team to identify any larger trends or anomalies in data collected
- Provide analytical assessments based on threat actor potential, organizational exposure, and organizational impact of threat
- Analyze current and future technologies and geopolitics that may positively/negatively impact the organization and industry
- Perform structured analytical techniques as needed
- Enhance executive leader decision making pertaining to organization-wide financial health, brand, stature, and reputation

**Reporting and Feedback—Dissemination and Integration, Reporting and Feedback/ Evaluation**

- Produce written and verbal reports and briefings (weekly, monthly, quarterly, semi-annually, annually) per leadership and organization-wide requests on topics. Explain threats to organization in **risk to business** based scenarios.
- Evaluate workflow processes quarterly—what can be streamlined, what can be updated, what can be automated?
- Create quarterly metrics of intelligence products produced and activity disrupted
- Create informal and formal mechanism for feedback (web portal, email address to team, surveys)
- Create quarterly metrics of feedback received on intelligence products through portal-specific comments, likes, views, downloads of reports
- Identify new requirements based on feedback, analyst requirements, and leadership concerns

**Human: Analytical Acumen**

- Apply critical thinking, creativity, and imagination to complex problems
- Understand the allure of "sexy" intelligence, cognitive biases, and logical fallacies
- Perform structured analytical techniques/ human-centered design techniques
- Bring context to information (risk to business/industry, trends, threat actor TTP insights)
- Manage, advance, and evaluate relations with internal and external partners (third-party intelligence providers, subsidiaries
- Evaluate processes, policies and tradecraft to ensure feedback is incorporated to ensure effective and efficient intelligence analysis

**Human-Machine Team**

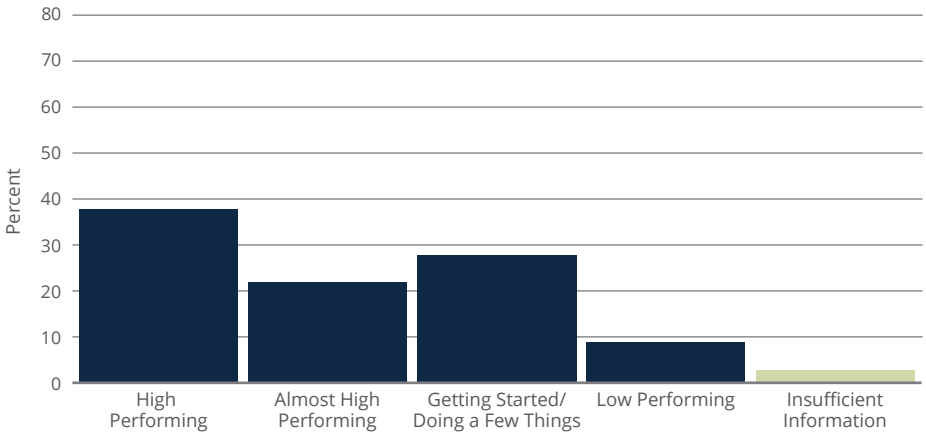
- Real-time status on cyber threats, organizational and international polices, new technologies, organizational developments, business offerings, new patents, new industry developments
- Detect anomalies
- Predict user behavior trends
- Real-time status on network architecture and attack surface
- Automation of manual tasks (parsing emails, attachments, URLs, file detonation, creating incidents, performing historical searches, notifying team members, and sending attachments or indicators through tools like Virus Total or WHOIS.
- Evaluate and score data and data sources on top of automation scoring process
- Generate concise tailored reports and presentations to specific audiences and leadership internal and external of organization

## ENVIRONMENTAL CONTEXT FACTOR 7: PRIORITIZING THREATS

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization uses a repeatable threat prioritization process (such as a matrix or scoring system) that incorporates components of the cyber intelligence workflow to identify and prioritize cyber threats based on threat actor potential, target exposure, and organizational impact. This process is reviewed and updated regularly.

### PERFORMANCE SNAPSHOT




**Environmental Context Factor 7 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**Threat prioritization is ad hoc or narrowly focused**

Teams across sectors often take an ad hoc approach to prioritizing threats, basing their judgments on current relevant news or gut feelings. In some organizations, executive leadership sets the organization's highest level intelligence requirements (sometimes going several years without updating them), and cyber intelligence analysts are left to identify organization gaps and establish appropriate priority intelligence requirements (PIRs) and specific intelligence requirements (SIRs) to collect against executive-level intelligence requirements—with no established process for doing so.

Some organizations also struggle to create a holistic threat prioritization process, meaning that their process fails to consider threat actor potential to target the organization, organizational exposure to the threat, and the impact of the threat on the organization. Additionally, a number of organizations rely solely on paid threat intelligence platforms to automate threat prioritizations, without conducting additional analysis and evaluation to determine if the automated prioritization is actually organizationally relevant. Some organizations do evaluate and review their own threat

#### IMPROVE YOUR PERFORMANCE

- Use public threat frameworks to assist with answering intelligence requirements and for tactical and operational threat prioritization.
- Consider threat actor potential to target the organization, organizational exposure to the threat, and the impact of the threat on the organization to strategically prioritize threats.
- Evaluate strategic threat prioritizations on a quarterly basis.

prioritization process; however, such evaluations occur annually at best. When considering the dynamic and emerging threat landscape, along with rapid industry and technological developments, organizations should holistically evaluate their threat prioritization process and corresponding IRs and PIRs quarterly. SIRs should be evaluated every 60 days.

Threat prioritization requires organizations to understand their environment. This means having a holistic understanding of the attack surface in relation to cyber threats: physical and logical attack surface, critical assets, patent pending technologies, executive-level intelligence requirements (IRs), industry developments, geopolitics, and knowledge gaps. Using that information, organizations establish PIRs and then lower-level, technical SIRs. The next step is to collect information to answer the IRs, PIRs, and SIRs. With the information collected as part of the Data Gathering component of the Cyber Intelligence Framework, organizations use human-machine teams to perform Threat Analysis or Strategic Analysis to create actionable intelligence for leadership. See Data Gathering Factor 1 for more information about the intelligence requirement process.

## BEST PRACTICES

**Use public threat frameworks**

High-performing organizations use public cyber threat frameworks to support intelligence analysis and communicate threat prioritizations. Our Public Threat Framework Implementation Guide describes how to use these frameworks and incorporate them into your cyber intelligence effort. Specifically, some teams have their Threat Analysis, threat/warning, and cyber defense analysts map technical internal and external telemetry (atomic, behavioral, and computed indicators) to the MITRE ATT&CK Framework[17] to track changes in threat actor behavior (TTPs) over time. This process assists with answering tactical and technical SIRs and for informing threat prioritizations. When it comes to briefing and writing for senior leadership and the board of directors, some organizations switch to the Lockheed Martin Cyber Kill Chain[18] to communicate attack stages. We also met with organizations that use the Diamond Model[19] to conduct analysis when leadership is primarily interested in attribution. Last, the ODNI Cyber Threat Framework[20] enables analysts to translate technical activities (what, when, where, and how—Threat Analysis) and strategic (who and why) analysis into common attributes and a common vocabulary or lexicon, which facilitates external organizational communication and collaboration. The ODNI CTF overlaps with other frameworks to create a common language to simplify metrics, reporting, and situational awareness.

**Prioritize threats based on threat actor potential, target exposure, and organizational impact**

High-performing organizations tend to consider a variety of factors when prioritizing threats. These considerations commonly fall into the three categories we described in our 2013 Cyber Threat Prioritization Implementation Guide:

- Threat Actor Potential to Execute the Threat (Capability + Intent)
- Organizational Exposure to the Threat because of Potential Vulnerabilities (People + Cyber Footprint + Physical + Technology)
- Organizational Impact of the Threat (Operational Costs+ Strategic Interests)

17 https://attack.mitre.org

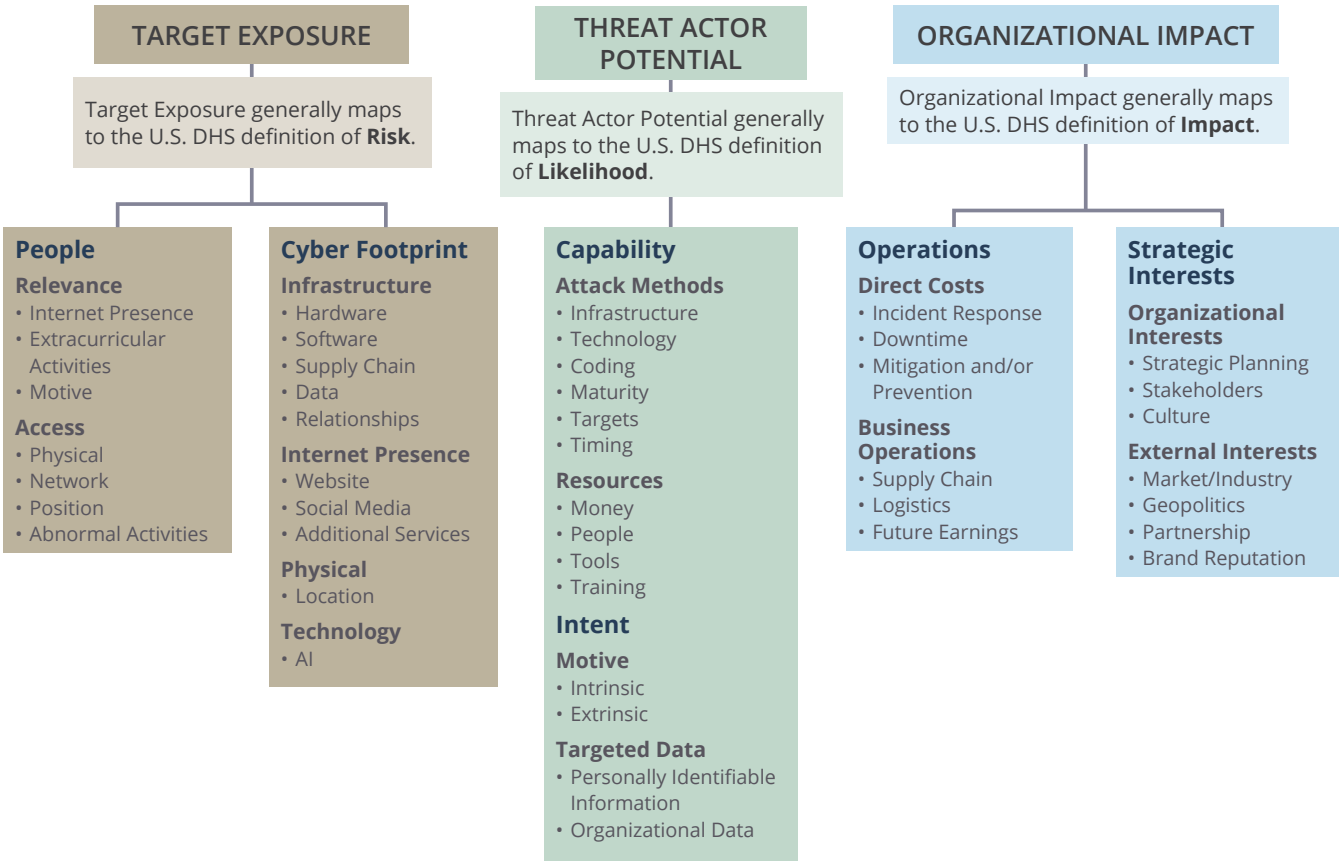
18 https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

19 https://apps.dtic.mil/docs/citations/ADA586960

20 https://www.dni.gov/index.php/cyber-threat-framework

## THREAT = RISK + LIKELIHOOD + IMPACT

## PRIORITIZING THREATS FOR MANAGEMENT[21]




**Prioritizing Threats for Management** High-performing organizations tend to consider a variety of factors when prioritizing threats. These considerations commonly fall into the three categories shown above.

### Use a tiered model to prioritize threats

Since 2013, high-performing cyber intelligence teams have reported using tiered models to prioritize threats. These models can be homegrown or based on existing tools. Tiered models enable teams to be more agile, focusing on the most important threats; such models also provide a framework for communicating strategic threat prioritizations to leadership. The simple scenario and matrix below provide one example of an approach to tiering threats.

21 DHS definitions can be found at https://www.dhs.gov/sites/default/files/publications/18_0116_MGMT_DHS-Lexicon.pdf

## FICTIONAL SCENARIO: THREAT PRIORITIZATION

**Threat Actor VenomSYN Using B00MAI Malware**
**Threat Prioritization Score: Medium**

Bottom Line Up Front: A medium risk exists that VenomSYN will target our organization using B00MAI malware. Threat Actor Potential: VenomSYN sends spear-phishing emails wrapping B00MAI malware in a PDF document. VenomSYN has been targeting organizations in defense and academic sectors, not organizations in our health sector. Target Exposure: VenomSYN may target our employees; however, overall exposure to B00MAI malware is low due to our cyber hygiene policies, two-factor Identity and Access Management practices and algorithmic detection capability based on sandbox testing. Organizational Impact: Organizational impact of this threat is assessed as medium. Should VenomSYN breach our systems, containment would be almost immediate. That said, public awareness of the breach could harm our organization's reputation.

**Likelihood (Threat Actor Potential)**

| | High Intent | Medium Intent | Low |
|---|---|---|---|
| High Capabilities | High Threat Actor Potential to execute threat **(OIS)** | High Threat Actor Potential to execute threat | Medium Threat Actor Potential to execute threat |
| Medium Capabilities | High Threat Actor Potential to execute threat | Medium Threat Actor Potential to execute threat **(OI,OP, OO)** | Low Threat Actor Potential to execute threat |
| Low Capabilities | Medium Threat Actor Potential to execute threat | Low Threat Actor Potential to execute threat | Low Threat Actor Potential to execute threat |

**Risk (Target Exposure to the threat because of potential vulnerabilities: People, Cyber, Physical, Technological (CPT)**

| | High CPT Vulnerabilities | Medium CPT Vulnerabilities | Low CPT Vulnerabilities |
|---|---|---|---|
| High People Vulnerabilities | High Target Exposure to the threat because of vulnerabilities **(OIS)** | High Target Exposure to the threat because of vulnerabilities | Medium Target Exposure to the threat because of vulnerabilities |
| Medium People Vulnerabilities | High Target Exposure to the threat because of vulnerabilities | Medium Target Exposure to the threat because of vulnerabilities **(OI,OP)** | Low Target Exposure to the threat because of vulnerabilities |
| Low People Vulnerabilities | Medium Target Exposure to the threat because of vulnerabilities | Low Target Exposure to the threat because of vulnerabilities | Low Target Exposure to the threat because of vulnerabilities **(OO)** |

**Impact (Organizational Impact of the cyber threat on the Target) = Operational Costs + Strategic Interest Impact**

| | High Strategic Interest Impact | Medium Strategic Interest Impact | Low Strategic Interest Impact |
|---|---|---|---|
| High Operational Costs | High Organizational Impact | High Organizational Impact **(OIS)** | Medium Organizational Impact |
| Medium Operational Costs | High Organizational Impact | Medium Organizational Impact **(OI,OP)** | Low Organizational Impact |
| Low Operational Costs | Medium Organizational Impact **(OO)** | Low Organizational Impact | Low Organizational Impact |

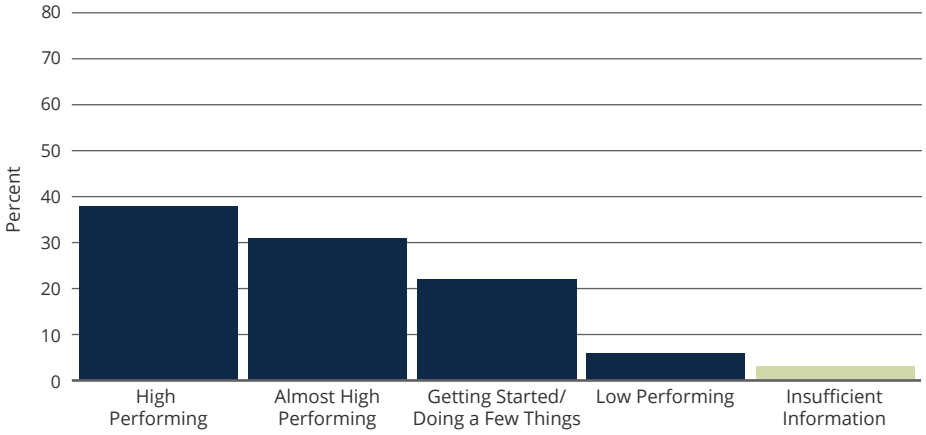
**Scenario Matrix** OI: Our Industry, OO: Our Organization, OP: Our Partners, OIS: Other Industry Sectors

## ENVIRONMENTAL CONTEXT FACTOR 8: USING PAST, PRESENT, AND FUTURE DATA

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization consistently uses past, present, and future data regarding cyber threats to the organization itself, within its industry, and across industries. The organization reviews lessons learned from prior incidents as part of its cyber intelligence efforts. Data includes significant historical data, current data and both self-developed and vendor-based predictions on future threats.

### PERFORMANCE SNAPSHOT




**Environmental Context Factor 8 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**The focus is only on today**

Although organizations widely acknowledge the importance of past data for informing present and future analysis, many struggle to effectively use past data. Besides the common challenge of resource constraints, organizations struggle with the lack of technology to query and manage past data. Some organizations use email to collect and manage all of their data. Other organizations described limitations with portal search functions and difficulties accessing logs. Even when organizations are able to manage and access old data, many lack a formal structure, method, or documented workflow to incorporate this data.

Organizations also struggle with looking toward the future. Many are not using past and present data, along with data about future threats, geopolitics, and technologies to predict future threats, risks or opportunities to the organization and industry. Resource constraints, along with lack of demand—likely due to the reactive approach we observed at many organizations—make predictive analysis difficult.

**IMPROVE YOUR PERFORMANCE**

Create capabilities and resources to leverage past data and intelligence on threat actors, IoCs and adversary behavioral trends to derive present and future adversary intent and capabilities.

### BEST PRACTICES

**Make use of tools and veteran team members**

High-performing organizations use historical reporting on threat actors, IoCs, and adversary behavioral trends to derive present and future adversary intent and capabilities. Many high-performing organizations use past data and trends to support link analysis, perform IoC reconstruction, inform leadership of current events, or show organizational defense capability improvement overtime. For past data, some organizations leverage the cloud to query logs, incidents, and post mortems going as far back as 10 years. Other organizations have built custom graph databases that enable quick and easy searches to help analysts understand past, present, and future data relationships.

High-performing organizations that have longtime employees do a good job of drawing from those team members' knowledge of past threats and events and the organization itself. Although relying solely on knowledge contained in team members' minds is a bad practice, leveraging team member experiences and perspective along with the appropriate tools and processes can increase the effectiveness of your cyber intelligence effort.

## ENVIRONMENTAL CONTEXT FACTOR 9: RELATIONSHIP BETWEEN CYBER INTELLIGENCE AND INSIDER THREAT DETECTION, PREVENTION, AND RESPONSE

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization's cyber intelligence effort has a relationship with its insider threat mitigation effort that supports mutual, proactive information sharing; the teams can access one anothers' databases and people when needed.

### TERMINOLOGY[22]

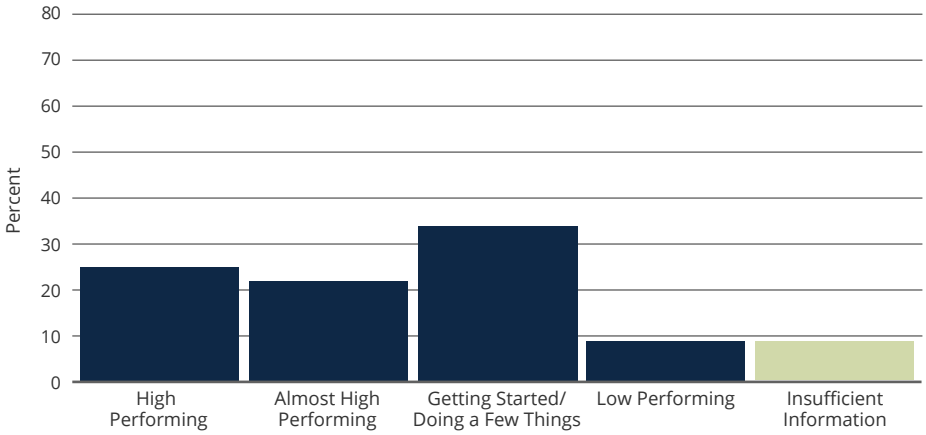
**Insider** – anyone given authorized access to organization assets (people, facilities, information, technology)

**Insider Threat** – the potential for an insider, either maliciously or unintentionally, to act in a way that could negatively affect the organization.

**Insider Incident** – harm realized by an organization due to the actions (or inactions) of an insider.

22 https://resources.sei.cmu.edu/asset_files/TechnicalReport/2019_005_001_540647.pdf

## PERFORMANCE SNAPSHOT




**Environmental Context Factor 9 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

### Absence of a true insider threat effort

Some organizations we interviewed do not have dedicated insider threat programs or teams. For some of these organizations, insider threat detection, prevention, and analysis fall to one person who has other full-time responsibilities within the information technology division or cybersecurity team. Some organizations rely exclusively on technical measures such as standard activity monitoring of databases, access management policies, and Data Loss Prevention (DLP) tools that make up their insider threat program. Other organizations have leadership who view insider threat only as a human error (for example, employees who fall victim to phishing emails); those organizations have not invested in tools like a DLP and instead simply provide training to employees. Still other organizations have not yet built an insider threat team because they are still coming to consensus on what an insider threat actually means to the organization or because they have not yet experienced an insider incident.

### Lack of information sharing between insider threat and cyber intelligence teams

A prevailing challenge for organizations that have insider threat programs is the lack of information sharing between the insider threat team and the cyber intelligence team. Some organizations have no information sharing at all—no passing of indicators, intelligence reports, or insider threat data sources. Some cyber intelligence teams only know if there is an insider threat issue at the organization if the insider threat team reaches out for additional information. Other organizations' cyber intelligence teams pass indicators and intelligence reports to the insider threat team without any reciprocity.

#### IMPROVE YOUR PERFORMANCE

- Create a formal insider threat mitigation program or function that uses a combination of policies, procedures, and technical controls across the organization to protect against malicious and unintentional insider threats.
- Create formal mechanisms to ensure bi-directional and proactive information sharing between the insider threat and cyber intelligence teams.

Lack of information sharing is sometimes due to data sensitivity, law enforcement/company investigations, and privacy concerns; even so, information sharing should not be one-sided.

## BEST PRACTICES

### Create an insider threat effort

The goal of an insider threat program is to prevent insider incidents and detect insider threats to an organization's critical assets without alienating insiders. High-performing organizations have formal insider threat teams, resources, and authorities with policies, procedures and technical controls. High-performing organizations often locate the insider threat program under the CISO/CSO/CRO to ensure appropriate information sharing with all cyber and non-cyber teams (including human resources and physical security) across the organization. Although some organizations embed an insider threat analyst in their fusion center to advance collaboration and communication, most organizations house their insider threat team outside the fusion center.

### Build relationships between insider threat and cyber intelligence teams

Cyber intelligence teams and insider threat teams in high-performing organizations recognize that working together is better for the overall protection of the organization's mission. The teams communicate not only through informal personal relationships, but in regular weekly calls and monthly formal meetings. Furthermore, these teams acknowledge that they are each consumers of the other's intelligence products. For example, the cyber intelligence team can send information to the insider threat team: keywords about organizational critical assets and technologies, TTPs for threat actors, organizational references in third-party intelligence reporting, and algorithms to support DLP and behavioral analytics. The insider threat team uses this information to make DLP and other adjustments to its monitoring and training capabilities. In return, the insider threat team can share case results, feedback on keywords, and RFIs to the cyber intelligence team. For additional information about how to create high-performing insider threat programs, refer to the SEI's *Common Sense Guide to Mitigating Insider Threats, Sixth Edition*.[23]

### Practice defense in depth; consider a DLP system

When it comes to technical controls, most high-performing organizations use a DLP system and conduct topical DLP analysis in combination with user activity monitoring, user behavioral analytics, or user entity behavioral analytic tools.

BEST PRACTICES

**APPLYING MACHINE LEARNING TO ASPECTS OF THE INSIDER THREAT PROBLEM**

A high-performing organization has created a neural network that learns on unstructured data from sensors surrounding the organization's web browsers and proxy sensors (including partners and affiliates). The organization has applied random forest decision trees to predict a probability that a user will head toward a website or category focused on weapons, criminal networks, and other nefarious sites.

23 https://resources.sei.cmu.edu/asset_files/TechnicalReport/2019_005_001_540647.pdf

*“...information consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention.”*

—Herbert A. Simon

# Data Gathering

*Collecting the Right Information*

## INTRODUCTION

When organizations know their environment, they can create the right intelligence requirements for data gathering. Through automated and labor-intensive means, data and information is collected from multiple internal and external sources for analysts to analyze to answer organizational intelligence requirements.

### DATA GATHERING ASSESSMENT FACTORS

In evaluating the state of the practice of cyber intelligence in terms of Data Gathering, we considered the following factors:

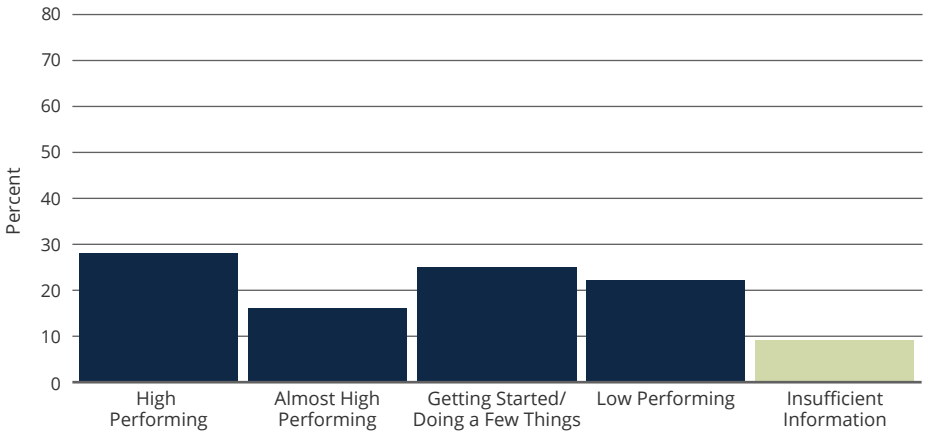
1. Intelligence Requirement Process
2. Intelligence Requirement and Data Source Alignment
3. Organization Information Sharing Process
4. Technology for Data Gathering
5. Data Source Validation

## DATA GATHERING FACTOR 1: INTELLIGENCE REQUIREMENT PROCESS

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization collects data that addresses Threat Analysis and Strategic Analysis needs according to intelligence requirements. The organization has a process to ensure analytical needs are met.

## PERFORMANCE SNAPSHOT




**Data Gathering Factor 1**

**Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

**Lack of organization-wide intelligence requirement process**

Without an intelligence requirement process where all leadership, analyst, and business unit intelligence requirements are understood and approved, an organization may have trouble identifying gaps, overlaps, or duplication of efforts. Some organizations have no mechanism to create, track, and satisfy intelligence requirements. Other organizations are building their cyber intelligence programs and are just beginning to engage leadership and analysts for intelligence requirements. Some organizations have intelligence requirements that address only cybersecurity concerns such as compliance, patch, and vulnerability management issues. Still others have different intelligence requirement processes for different teams across the organization.

**Stale intelligence requirements**

Organizations struggle with outdated requirements that lead to irrelevant data collection or data collection with diminishing analytical returns. Some organizations have high-level intelligence requirements that were established years ago by senior leadership, some of whom are no longer at the organization.

**Difficulties with third-party intelligence providers**

Organizations described a variety of challenges with third-party intelligence providers not meeting the organization's intelligence requirements. One organization explained that intelligence provider feeds do not contain raw data its cyber intelligence team needs for Threat Analysis. Some third-party intelligence providers produce only finished intelligence products and provide access to sales people, when organizations prefer raw data and access to vendor-specific analysts. Similarly, some third-party intelligence providers require an organization to buy an entire intelligence portfolio when they only

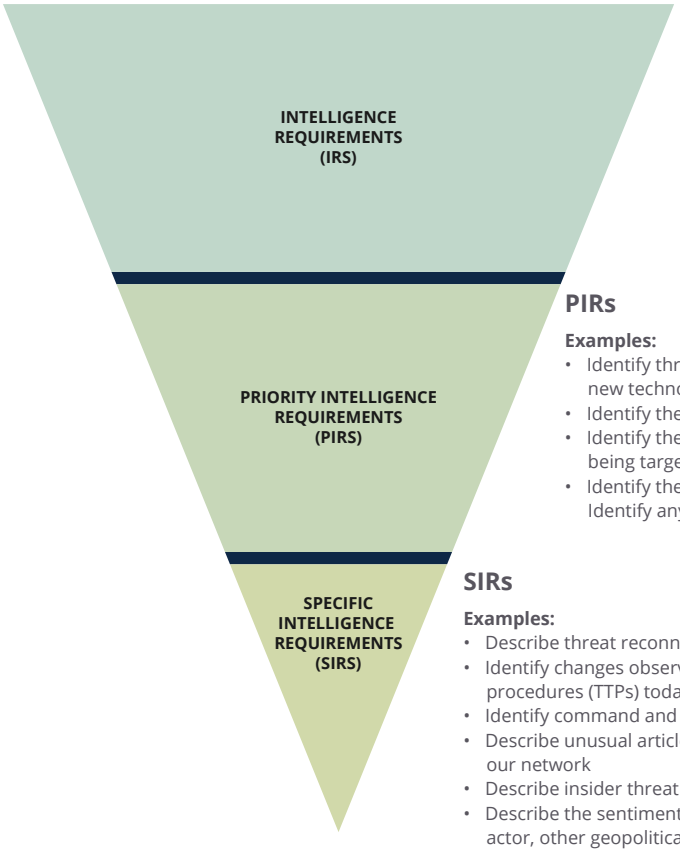
**IMPROVE YOUR PERFORMANCE**

- Create a collection management team to manage the intelligence requirements process.
- Use intelligence requirements, priority intelligence requirements, and specific intelligence requirements.
- Tag organizational specific intelligence requirements to DHS Homeland Security (HSEC) Standing Information Needs (SIN) as appropriate.

need one small aspect of the intelligence provider's service. In a few cases, organizations admitted that they themselves had failed to alert vendors of intelligence requirement changes.

## INTELLIGENCE REQUIREMENTS

Cyber intelligence teams consider intelligence requirements (IRs) alongside environmental context information about their attack surface, critical assets, patent pending technologies, business unit needs, industry developments, geopolitics, and knowledge gaps to develop priority intelligence requirements (PIRs) and then more granular and technical specific intelligence requirements (SIRs).




**IRs**

**Examples:**
- Identify notable threats to the organization
- Identify internal and external cyber threats targeting the organization
- Identify cyber threats targeting related industries
- Identify cyber threats to our partners

**PIRs**

**Examples:**
- Identify threat actors targeting our organization's critical assets or new technologies
- Identify the threat actors' motives
- Identify the person, group, entity or asset in the organization that is being targeted
- Identify the impact of the threat (reputation, revenue, operations). Identify any mitigating controls in place.

**SIRs**

**Examples:**
- Describe threat reconnaissance activity that occurred today
- Identify changes observed in a specific threat actor tactics, techniques, and procedures (TTPs) today
- Identify command and control infrastructure a specific threat actor is using
- Describe unusual articles, indicators, or behavior changes across our network
- Describe insider threat attempts observed today
- Describe the sentiment of a country government towards a specific threat actor, other geopolitical information regarding a specific threat actor

IRs reflect senior leadership and board concerns about threats and risks to the organization's environment, mission, operations, revenue, bottom line, and reputation. They are general in nature and are approved at the highest level of the organization (CEO, president). IRs serve as a baseline and starting point for the organization's collection plan.

PIRs are more detailed and operationally focused and align to IRs. PIRs should be approved by the CEO, vice president, and CSO/CISO, and should be updated at least every six months.

SIRs are operational, tactical, and technical in nature and focus on particular facts, entities, or activities. They also tend to be greater in number than IRs and PIRs and change more frequently based on both the dynamic nature of an organization's environment and the cyber threat landscape. SIRs are created by the cyber intelligence team in collaboration with others in the fusion center and should be approved at the CSO/CISO level. SIRs should be evaluated and audited at least every 60 days.

## BEST PRACTICES

**Create a collection management team to manage intelligence requirements**

A practice of high-performing organizations is having a collection management team responsible for capturing, managing, and evaluating senior-executive-level intelligence requirements, priority intelligence requirements, and specific intelligence requirements. The collection management process has three core aspects: a requirement, the actual data gathering, and analysis of the data to answer the requirement. These responsibilities fall to the collection management team. In other words, the collection management team owns, manages, produces, and evaluates the cyber intelligence requirement process, and assists with the data gathering and vetting processes. The collection management team establishes collection requirements to ensure the data collected comes from a variety of sources and is aligned to answer IRs, PIRs, SIRs and RFIs. The collection management team also ensures that data collected meets present needs and is aligned to support organizational strategic plans and vision. Last, the collection management team develops and tracks the rationale for each data source used and continuously looks for new data sources and technologies to help automate some of these processes.

Based on this best practice and drawing from *Intelligence Community Directive 204, National Intelligence Priorities Framework*,[24] organizations can create an organizational intelligence priorities framework (OIPF). The OIPF informs future planning, budgeting, programming, and allocation of resources to data collection and analysis. The OIPF should be actively managed so that it reflects organization-wide stakeholder priorities, and the entire OIPF should be reviewed quarterly. Organizations should consider imposing expiration dates on intelligence requirements to force reevaluation. To increase visibility, organizations should consider providing access to the OIPF to all departments that may be able to use it. The OIPF should also show how specific collection sources and their source validation status align to intelligence requirements. Advanced organizations could incorporate an OIPF into existing dashboard capabilities, permitting users to drill down through the IRs, PIRs, and SIRs.

### INTELLIGENCE REQUIREMENTS VS. COLLECTION REQUIREMENTS

**Intelligence Requirement:** Request for information about threats, risks, and opportunities for the purpose of protecting and advancing the organization's mission. Answering intelligence requirements requires data collection, analysis and reporting and feedback.

**Collection Requirement:** Request for using specific types of internal and external data sources and/or variety of sources that provide data to help answer IRs, PIRs, and IRs.

24 Intelligence Community Directive 204. National Intelligence Priorities Framework. 2 January 2015 https://www.dni.gov/files/documents/ICD/ICD%20204%20National%20Intelligence%20Priorities%20Framework.pdf

### Track customer needs using standing information needs

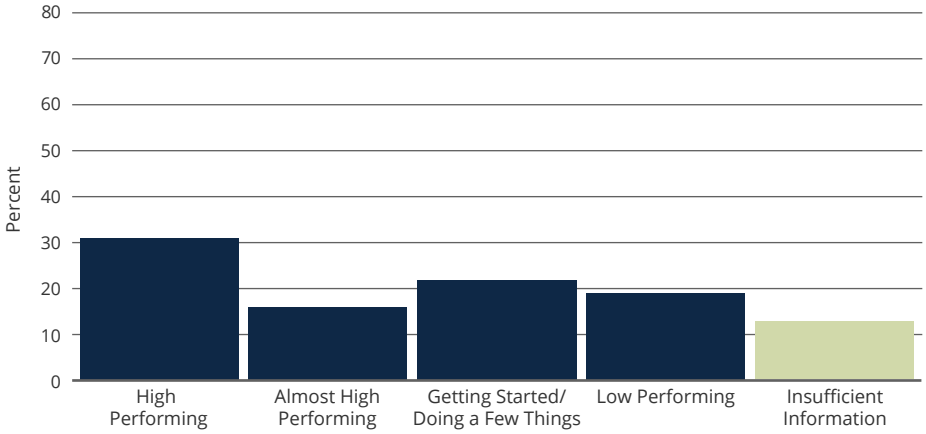
The Department of Homeland Security (DHS) uses Homeland Security (HSEC) Standing Information Needs (SIN) to identify and track customer needs across the department. DHS national fusion centers also establish their own specific SINs to identify, track, and satisfy customer needs within their area of responsibility. National fusion centers and ISACs provide information and intelligence analysis in response to these needs.[25] Some fusion centers and ISACs have created special interest groups to determine customers' intelligence requirements. High-performing organizations also align and tag their own IRs, PIRs, and SIRs to HSEC SINs and fusion center and ISAC-specific IRs. Aligning organizational requirements to national requirements helps guarantee operational relevance and enhances public and private information sharing and trust.

## DATA GATHERING FACTOR 2: INTELLIGENCE REQUIREMENT AND DATA SOURCE ALIGNMENT

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization has a formal repeatable process for aligning data sources to meet intelligence requirements. This process is reviewed and updated regularly.

### PERFORMANCE SNAPSHOT




**Data Gathering Factor 2 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

#### Lack of people leads to lack of process

For some organizations, no formal, repeatable process exists to align data sources to intelligence requirements, often due to resource constraints. Most of these organizations do not have the people and

#### IMPROVE YOUR PERFORMANCE

- Create a system or mechanism to align data sources to intelligence requirements.
- Use both internal and external data sources to support your cyber intelligence effort.
- Continuously evaluate third-party intelligence providers via scoring criteria.

25 https://www.archives.gov/files/isoo/oversight-groups/sltps-pac/national-network-of-fusion-centers-2015.pdf https://www.hsdl.org/?view&did=817528

time to align data sources to particular intelligence requirements and end up following an ad hoc or trial-by-error process.

**Fragmentation and decentralization**
Several organizations explained that no central holistic view exists of all sources used by every analyst across the organization. Instead, each cyber intelligence analyst has their own set of data sources. One organization noted that its SOC has a collection of sources and procedures for aligning sources, while the cyber intelligence team has different sources and procedures. A lack of a central location for sources may result in duplicative efforts or may lead to a collection gap against an IR, PIR, or SIR. Organizations should have a location the entire fusion center can access showing the source, the source's validation, and what is being collected from that source to answer IRs, PIRs, and SIRs. Incorporating this location into any capabilities associated with an OIPF would be beneficial.

## BEST PRACTICES

**Map data sources to intelligence requirements**
High-performing organizations map their data sources to their intelligence requirements. One high-performing organization is currently building an automated capability that aligns existing and new data sources to existing organizational IRs, PIRs, and SIRs.

**Evaluate and communicate with intelligence vendors**
High-performing organizations often use their collection management teams to manage the organization's relationship with its third-party intelligence providers, specifically pertaining to intelligence requirements. The collection management team communicates new requirements, explains the justification and priority behind them, and provides feedback to the third party. For some high-performing organizations, the collection management team collaborates with other members of the cyber intelligence team, (specifically the cyber intelligence analysts) to continuously evaluate third-party intelligence providers via scoring criteria like letter grades. Other high-performing organizations track the third-party provider's performance using month-to-month graphs to show how intelligence provided by the vendor answered intelligence requirements and helped the organization; organizations send that feedback to the vendor to let them know how they are doing.

**Differentiate between third-party intelligence aggregators and intelligence originators**
In evaluating third-party intelligence providers, high-performing organizations identify whether the provider is an intelligence aggregator or an intelligence originator. An intelligence aggregator simply collects and passes intelligence to its customers, while an intelligence originator provides new context to the information, making it actionable and relevant to the customer.

**Use a wide variety of sources**
High-performing organizations emphasized two key ideas regarding data source collection: "any data all the time" and "data finds data." High-performing organizations use a variety of internal and external data sources to support intelligence analysis.

First, internal data sources are typically generated messages (logs) or machine data from organizational hardware and software regarding device usage. There are many types of internal logs: traffic logs, operating system logs, firewall logs, IDS and HIDS logs, IoT logs, cloud logs, and vulnerability management logs, just to name a few. These internal data sources are typically

ingested, viewed, and analyzed in a SIEM, DLP, Intrusion Detection/Intrusion Prevention (IDS/IPS), Endpoint Detection and Response (EDR) Platform, or Security Orchestration Automation and Response (SOAR)—or a Third-Party Threat Intelligence Platform (TIP) that integrates many tools. Internal data sources, however, should not be limited to just machine data and logs. Internal data sources should include logs, tips, and other information from data sharing relationships, service level agreements, and collaboration with other internal business units such as human resources, marketing/sales, research and development, finance, and supply chain management.

External sources are both paid and free third-party intelligence providers or platforms that provide aggregated intelligence and/or additional originated context (actionable and organizationally relevant) about atomic, behavioral, and computed indicators of comprise[26] and associated meta-data analysis (email addresses, IP addresses, user agent strings, etc.) related to vulnerabilities, threat actor groups, threat actor TTPs, threat actor capabilities and motivations, and threat campaigns.

**TIP**

**See Appendix: Popular Cyber Intelligence Resources** for a list of free and paid intelligence vendors and sources that organizations told us they are currently using.

External intelligence vendors may provide information from a collection of sensitive sources, which could include adversary communications in dark/deep/surface web forums, C2 servers, forensic analysis, Virus Total, Shodan, endpoint, and network security data that they have access to from their organizational customers. The Intelligence Community, defense and other government agencies, may also receive indicators and information about threat actors, capabilities and motivations via unclassified and classified sources and means such as signals intelligence (SIGINT), imagery intelligence (IMINT), human intelligence (HUMINT), measurement and signatures intelligence (MASINT), open source intelligence (OSINT), and geospatial intelligence (GEOINT).

High-performing and larger organizations also create their own global/external business information security officer (BISO) collection capability. These organizations train BISOs in intelligence collection and analysis. The BISOs provide country-specific intelligence by gathering information from local sources and conducting analysis on that information. Adding a BISO collection capability increased one organization's overall monthly production by 30 intelligence reports.

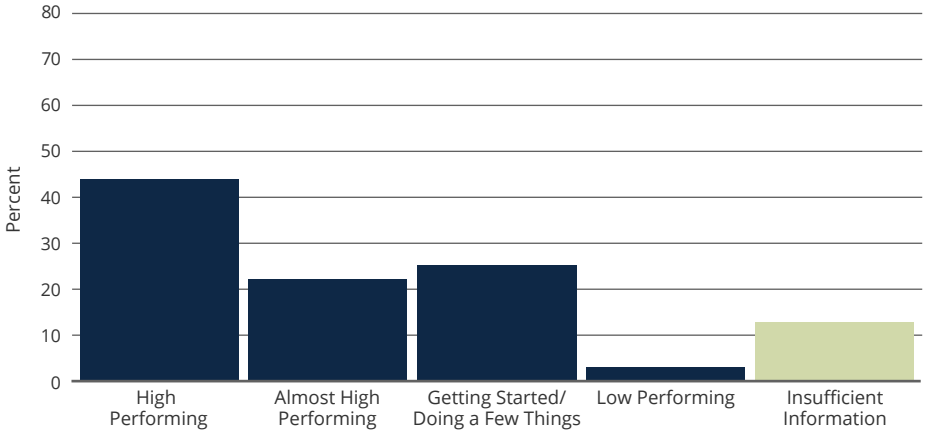
26 https://digital-forensics.sans.org/blog/2009/10/14/security-intelligence-attacking-the-kill-chain

## DATA GATHERING FACTOR 3: ORGANIZATION INFORMATION SHARING PROCESS

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization has formal and informal, bi-directional, and proactive sharing of information and analysis with appropriate internal organizational business units and external partners. The organization assigns staff members to lead information sharing relationships when appropriate. There is a process to review and update the value of information sharing relationships.

### PERFORMANCE SNAPSHOT




**Data Gathering Factor 3 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**Balkanization impedes an organization's internal information sharing**

We described the challenge of organizational silos in Environmental Context Factor 1. Although logical purposes exist for separation between certain business units (for example, data privacy, proprietary and classified information), silos stymie formal and informal information sharing between internal business units. Information that could be used to protect the organization and support its overall mission is not being shared proactively and across the organization.

Organizational policies, organizational structure, and business-specific technology stacks impede bi-directional and proactive sharing of relevant functional and strategic information and analysis. Organizations described a variety of challenges related to silos, including the absence of formal sharing mechanisms and service level agreements with other key business units, communicating cyber intelligence and important threat data with an organization's own overseas business subsidiaries that are unable to provide headquarters with relevant intelligence and threat data, and lack of involvement by legal and HR departments until those departments have a critical need.

**IMPROVE YOUR PERFORMANCE**

- Work with other organizations and across sectors to develop joint publications, create CTFs, and host brown bags on subjects such as best practices and lessons learned.
- Use the collection management team and BISOs to build internal and external relationships.

### Shortcomings in external information sharing

External information sharing—that is, sharing by government, industry, and academia—has improved since our 2013 cyber intelligence study, but challenges remain. Many organizations we met described benefits from the Department of Homeland Security's (DHS) free Automated Indicator Sharing (AIS) capability, which allows organizations to receive and share anonymized cyber threat indicators. The Cyber Information Sharing and Collaboration Program (CISCP) provides Indicator Bulletins and Threat Actor and Malware Analysis reports that organizations can use to support their own analysis. The partnership among industry, academia, government and law enforcement appears to be growing. Some of the organizations we interviewed are deepening their relationships with other government and non-profit organizations such as local FBI field offices, the Intelligence Community, and the National Cyber Forensics Training Alliance (NCFTA).

Shortcomings remain in the quantity, type, and level of information shared. Government organizations conveyed that industry organizations do not share enough cyber intelligence with the government, and companies conveyed that government organizations do not share enough cyber intelligence with industry. Several organizations described challenges with law enforcement in particular: these organizations perceive information sharing to be a one-way street, with industry and academic organizations receiving little or no feedback from law enforcement about how information is ultimately used. These organizations report that the lack of sharing makes them less inclined to share data and intelligence. Several industry organizations expressed minor frustration with DHS AIS and CISCP and FBI Private Industry Notifications (PINs). They described the information as being occasionally negligible, or already known before the reports were released to industry.

Separately, organizations voiced their desire for increased cyber intelligence collaboration and partnership with and among financial organizations and Silicon Valley (specifically the "big five" technology companies[27]).

### Meaningful participation in ISACs

Information sharing and analysis centers (ISACs) face challenges when members do not participate in meaningful ways. Organizations explained that because of privacy and proprietary information sharing concerns, they can often only receive information from ISACs; ISACs then struggle to get insight about the members' missions, environments, vulnerabilities, requirements, threat prioritizations, and internal cyber intelligence products.

**TIP**

To identify your sector's ISAC, visit **nationalisacs.org/member-isacs**.

27 https://en.wikipedia.org/wiki/Big_Four_tech_companies

## BEST PRACTICES

**Share the right information**

High-performing organizations recognize the difference between meaningful information sharing and just information sharing. One high-performing organization shares intelligence with relevant fusion centers and ISACs only when the intelligence is actionable and has received a 51% confidence rating from analysts.

An interesting practice of some high-performing organizations is to share draft cyber intelligence reports and initial analytical judgments with trusted cyber intelligence teams that work for external entities or organizations. Trusted external teams provide comments, analytical recommendations, and other feedback to improve the report. For industry, this practice has the potential to grow into something bigger, such as companies publishing joint reports. Collaborative reporting in industry can emulate National Intelligence Estimates and Intelligence Community Assessments, which serve as the IC's authoritative statements on particular issues.

## INFORMATION SHARING—A CONFIDENCE-BUILDING MEASURE

Organizations and companies have different missions and business goals. Free-market and vigorous competition naturally exists among companies to generate wealth by creating the best cyber intelligence product, invention and innovation. Yet what is more true every day, is that cyber touches everything. And in the open and free internet, a threat to one can quickly become a threat to us all. Are there ways for organizations to continue to be the best they can—create new products, intellectual property and innovations, and work together in new and meaningful ways? Collaboration efforts in the Cybersecurity Tech Accord, the Global Cyber Alliance, and the Public-Private Analytic Exchange Program (AEP)[28] are some good examples of these types of efforts. We offer a few additional ideas that adhere to the general concepts of organizations across government, industry, and academia doing more things together and being more transparent:

**Do things together**

Contact other organizations and companies to create formalized brown bags, town halls, cyber threat frameworks, joint cyber assessments, cross-sector virtual blogs, and chat rooms.

The joint creation of cyber intelligence reports by private sector companies, ISAC members and third-party intelligence providers can increase teaming, collaboration, and transparency, which leads to trust. Moreover, jointly produced reports (with appropriate legal guidance to protect privacy/proprietary information and within Traffic Light Protocol guidelines) could bring greater authority and credibility to assessments on cyber issues. Joint publication conveys the reality that a threat to one is a threat to all. Organizations could also reserve the option to publicly disclose their contribution to the report and include supporting and dissenting views on analytical judgments.

**Be transparent**

Share data (indicators) *and knowledge*. Government, private sector, and academic organizations as well as ISACs, fusion centers, and third-party intelligence providers, can share knowledge about

28 https://cybertechaccord.org ; https://www.globalcyberalliance.org; www.dhs.gov/intelligence-and-analysis-private-sector-engagement

- prior attacks and how your organization handled them (lessons learned)
- new attack surfaces
- using common tools and technology more efficiently
- internal best practices and challenges
- team compositions (roles, talent, responsibilities)
- current strategic threats, campaigns, attribution

**Task collection management team with managing information sharing**

High-performing organizations usually have a collection management team squarely focused on ensuring successful formal and informal sharing of cyber information and intelligence with internal and external partners, including vendors. The collection management team regularly evaluates its relationships, thinking about new and more efficient ways to share and receive information. The collection management team also helps to build, in coordination with the program management office's internal and external relationship team, successful information sharing relationships with other internal organizational business units that fall traditionally outside of a fusion center, such as HR, business intelligence, physical security, legal, marketing, finance, technology development, and corporate leadership.

**Formalize and document information sharing practices**

High-performing organizations often develop cyber intelligence guides and best practices for sharing intelligence with internal business units—and their people understand those guides. Organizations that had fusion centers but were still building a collection management team relied on business information security officers (BISOs) embedded in each organizational business unit to manage the relationship with the greater fusion center. BISOs act as both a liaison and officer for the fusion center by ensuring CISO polices are formulated into the business unit and enhancing intelligence sharing (intelligence requirements, cyber intelligence reports) with the fusion center.

**Foster fusion center culture through engaged leadership**

Fusion centers must be actively managed by leadership. Leaders of high-performing organizations ensure their fusion centers have a culture that inspires innovation, teamwork, hard work, and a sense of mission. Additionally, the leaders of the fusion centers themselves are engaged, providing guidance and decisions in a timely manner. We discuss more on leadership engagement in the Reporting and Feedback section of the report.

**TIP**

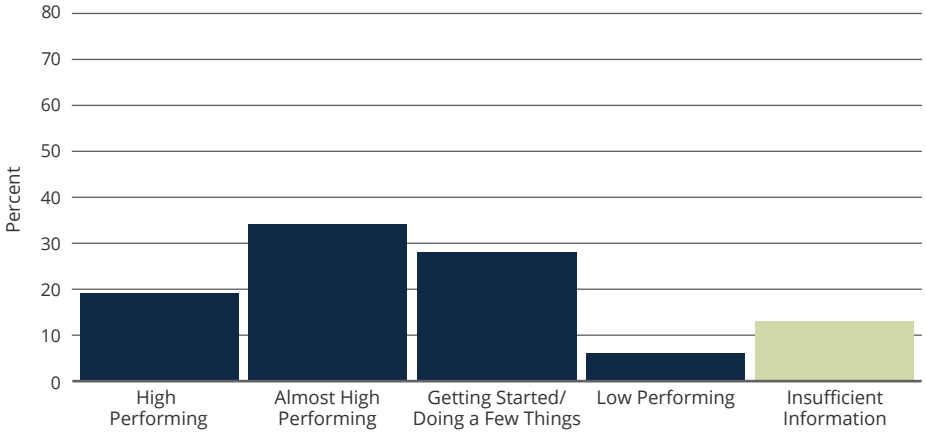
The collection management team is not responsible for managing overall information security and compliance relationships with other organizational internal business units, suppliers, partners, contractors, and stakeholders. The program management office (PMO), a component of the fusion center, should have an internal/external relationship team managing internal and external relationships. This team should also speak to internal business units and other partners and contractors about the value of cyber intelligence. This team should coordinate and work very closely with the cyber intelligence collection management team.

## DATA GATHERING FACTOR 4: TECHNOLOGY FOR DATA GATHERING

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization aligns homegrown and off-the-shelf technology with specific environmental context factors and data gathering requirements to tailor tools that consistently satisfy analytical needs. The organization has a technology review process. The organization uses current and emerging technology such as machine learning and automation as appropriate.

### PERFORMANCE SNAPSHOT




**Data Gathering Factor 4 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**Outdated technologies, resource challenges, and lack of a technological review**

Since our 2013 study, more organizations have turned to technologies like SIEMs, SOAR platforms, and threat detection platforms that apply automation, data science, and behavioral analytics to log threat data to support data gathering, incident response, and Threat Analysis. However, some organizations rely on outdated tools and technologies to support data gathering and data management. These organizations discussed struggles normalizing data and find themselves continuously weeding through false positives. One organization has dedicated a full-time analyst to manually identify and work through daily false positives. Even some large organizations use email as their primary method to collect and manage data, and in other organizations, strategic analysts rely on spreadsheets to track threats and threat actors.

#### IMPROVE YOUR PERFORMANCE

- Write scripts to assist with data ingestion, product dissemination, and phishing responses.
- Adopt SOAR platforms to assist with workflow creation and manual data enrichment tasks.
- Create a technology development and integration team to build customized tools that leverage automation and machine learning for cyber intelligence needs.

Organizations commonly lack resources (a recurring challenge related to people, capability, and funding) to build customized tools to meet specific analytical and data gathering needs that cannot be met by off-the-shelf technology. Some organizations also expressed challenges acquiring funding approval for new technology; others discussed how technology fragmentation within their organization hampers mission and collaboration.

Still other organizations lack a technological review process. These organizations struggle to know if their existing technology is outdated, if it is capable of answering new needs, if new technology is available that could help the organization, or if other divisions across the organization are using same or better technology.

**Data normalization and ingestion still a challenge**

In our 2013 cyber intelligence study, we found that organizations were inundated with data feeds that came in different formats, making data consumption and integration for analysis extremely challenging. Although significant progress has been made with data language and serialization formats and exchange standards such as MITRE's Structured Threat Information Expression (STIX), Trusted Automated Exchange of Indicator Information (TAXII) 2.0, and OpenIOC, challenges remain. Data normalization is a never-ending hurdle for both organizations and vendors. The problem is compounded by the 2.5 quintillion bytes of data generated every day[29] from existing machines and the increasing number of connected devices and learning systems.

Multiple permutations exist for how organizations receive, document and capture (tag and index), and extract structured and unstructured relevant data and metadata resources (ports, domains, IPs and hashes, timestamps) in XML, JSON, free text, and CSV coming from these devices. A data resource from one organization or from one threat intelligence vendor might actually be the same data resource from a different organization or different vendor, even though it is represented by different strings and formats. Today's machines are generally not yet smart enough to recognize the same information formatted in different ways.

29 https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#6c197fa060ba

## CASE STUDY: DATA NORMALIZATION

A high-performing organization dedicated resources to establishing an organic internal cyber intelligence system using big data frameworks and natural language processing to automate the ingestion and normalization of data received from internal and external sources. The system generates a record from each data source that populates field constructs in a context such as the following:

- **UID**: Unique data record (line) identifier, or article reference number
- **TYPE**: Common object category (e.g., Actor, Malware)
- **NAME**: Common object designation or name (e.g., FIN5, Sofacy, Emotet, COOLPANTS)
- **ALIAS**: Familiar name(s) associated with object from all sources
- **GEOGRAPHY**: Geopolitical boundary of actor/group activity (e.g., World, Continent, Country, Region, State, City, Local/Tribe)
- **INTENT**: Explicit (e.g., Criminal, Political, Espionage, Personal reward, Fame, Money, Hacktivism)
- **REQUIREMENT REFERENCE**: Intelligence requirement, priority requirement, and specific intelligence requirement number
- **COMPENSATING CONTROL**: Freeform (from a defined list) security-led operations existing, emergent, or recommended physical or logical risk/threat mitigations
- **ADMIRALTY CODE RATING**: Source Reliability (A-F rating) and Information Content (1-6 rating)")
- **FSEEN**: Date/Time of first seen activity
- **LSEEN**: Date/Time of last seen activity
- **TAG**: Identifier for sorting, searching, and sharing
- **NOTE**: Freeform text field
- **ATTACHMENT**: Object or link extension (actual article/object or referrer)

## BEST PRACTICES

**Form a team to investigate emerging technology**

High-performing organizations have technology development and integration teams comprised of security engineers, developers, data scientists, statisticians, and machine learning experts. The technology development and integration team meets frequently with analysts and leadership and incorporates their input and needs into future technology builds and procurements. The team then builds customized tools for cybersecurity and cyber intelligence purposes and applies automation and machine learning as appropriate.

We met organizations that have created in-house analytical tools that perform like Maltego but are specific to the organization's needs. Another high-performing organization has a team that built a large graph database of all internal and external data it has collected. The graph database is curated and highly-structured and is used for discovery, analysis, and knowledge sharing. The organization's technology development team is currently working on automating tasks within the graph database to hunt for interesting data, connections, and correlations. We also met with an organization that created an in-house,

automated collection management system. One participant shared a piece of wisdom with our team: "anything you have to do more than once, you can script," which frees up time, money, and people to focus on more complicated analytics.

**Use diverse technology to support cyber intelligence**
Most high-performing organizations do not rely exclusively on a single tool or an "all-in-one" solution via integrations into a Threat Intelligence Platform (TIP) or SIEM. Rather, they incorporate homegrown and a variety of free and paid off-the-shelf tools and technologies to support current data gathering and analysis. For instance, a number of high-performing organizations have incorporated the free open-source ELK stack (Elastic Search, Logstash, and Kibana) for data processing/aggregation, search, analysis, and visualization. Other organizations use Hadoop, MongoDB, or cloud-based solutions for data storage and management. For intelligence analysis and visualization, a number of high-preforming organizations use free and paid for tools such as BRO, Kali Linux, Process Monitor, Maltego, Analyst's Notebook, Malware Information Sharing Platform (MISP), Tableau, and Adobe InDesign/Photoshop. Naturally, SIEMs, DLPs, SOAR, and TIPS provide analysis and visualization features in addition to product integrations with some of these same tools.

Technology also enables organizations to share information quickly and efficiently. We met organizations using Slack, SharePoint and their internal SIEM, TIP, or SOAR platform ticketing systems to share event and incident information. Organizations use Microsoft's Yammer tool as both an organizational social networking tool and incident tracker. Information and reports can be shared, posted, and edited in Yammer, and analysts and leadership can provide feedback and "like" reports and comments. In many high-performing organizations, the fusion center—and specifically the cyber intelligence team—maintains a website for sharing and receiving information such as cyber intelligence reports, current working drafts, best practices, new developments, opportunities for feedback, future reports, and RFIs. On the RFI page, the option exists to explain priority of the information need and track the status once it is submitted

**Automation, artificial intelligence, and applied machine learning**
High-performing organizations recognize that automation is no longer simply nice to have; it is a necessity. Since our 2013 study, organizations have built more scripts to assist with data ingestion, product dissemination, and phishing response. Additionally, a number of organizations are using or incorporating SOAR platforms to help automate incident response and data enrichment tasks. SOAR platforms are designed to automatically integrate data from a variety

**TIP**

Since our 2013 study, cyber intelligence tools have become more versatile. Tools that were once single-feature technologies now have a variety of functions. In **Appendix: Most Popular Cyber Intelligence Resources**, we present a list of some of the most popular tools and resources reported by study participants and their uses.

of internal security tools and gather incident data and context into one single location. SOAR platforms can produce both standard and customizable step-by-step playbooks or workflows that automate manual repeatable tasks such as parsing emails, file detonation, creating incidents, notifying team members, and sending attachments or indicators through Virus Total or WHOIS. Our research also shows that high-performing organizations with resources and funding to purchase or apply machine learning will see direct savings in labor, giving analysts time to work on more pressing issues

## IMPLEMENTING MACHINE LEARNING

Organizations we met are implementing machine learning in the following ways:

- **Feeding a neural network normalized data using natural language processing.** Physical, logical, and sociocultural data dimensions are systematically categorized by machines. Data artifact, indicator, and behavior characteristics are equalized and weighted against organization risk and decision-making models. The system ranks risk to prioritize threat matching and initiate predictive pattern recognition beyond human analyst capacity. The system qualifies matches of 100% malicious activity and has the option to monitor, act, or maneuver the threat through artificial intelligence and series of mitigating controls. The system generates summary risk and threat judgment for appropriate consumers (C-Suite to Analyst). The system is currently able to process 1.25 petabytes every day and can search back through data on demand.
- **Using supervised learning to train a model on a dataset of 5,000 articles.** The model generates articles twice a day for the entire team. One analyst is responsible for triaging and drilling down on the most serious and pressing items. The model also gets better every day because the analysts provide new training and feedback data to the model as they work. For example, any report written by the cyber intelligence team is tagged with the same tags they used to label and ingest articles originally. This organization claims that the process has reduced the time required for a particular task from eight hours to one hour.
- **Applying dynamic topic modeling to enhance intelligence analysis.** Dynamic topic modeling is a way to analyze the evolution of (unobserved) topics of a collection of documents over time. The ML application helps them answer the questions: What do we believe will happen in the next year? What topics are we seeing or did not look at in our analysis?
- **Using machine learning to help tackle the inside threat problem.** Specifically, training model(s) to learn how web browsers are susceptible to vulnerabilities and also internal user behavior (all logs, files and artifacts the user interacts with). Using a random forest decision tree algorithm, the model predicts the probability that a user's experience is heading toward a threat vector.

**TIP**

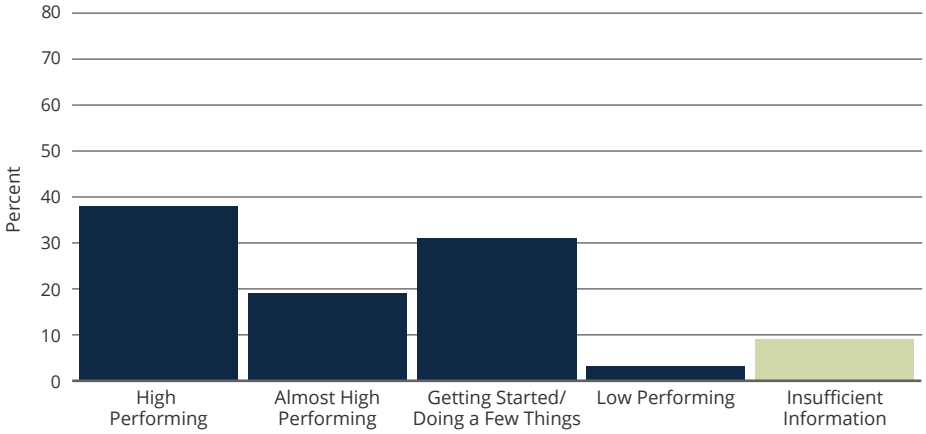
For more information on machine learning, see

## DATA GATHERING FACTOR 5: DATA SOURCE VALIDATION

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization has a repeatable process of validating data through tagging, using multiple sources, and assessing data sources. This process is reviewed and updated regularly.

### PERFORMANCE SNAPSHOT




**Data Gathering Factor 5 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**Lack of a common lexicon**

Organizations use different terminology to describe a source's validation, such as a credibility ranking, confidence ranking, uniqueness ranking, or reliability ranking. Criteria used to justify validation rankings also vary across organizations and industries, with some organizations using only corroboration by other data sources as the justification for validation.

**Lack of processes**

Some organizations have no process for validating data and data sources, while others have processes that lack consistency, formalization, or transparency across the organization. These difficulties are compounded when analysts have their own data sources outside the central location where an organization's data sources are managed and evaluated. Organizations that have instituted a process for validating data sources explained that they might not review all of their sources regularly (at best annually) to determine if the data sources are still credible and reliable and provide relevant data to support the organization's mission. Last, some organizations only validate the data itself and not the data source.

**Reliance on vendors to validate data sources**

Some organizations rely completely on third-party intelligence providers to perform data source validation, often due to lack of resources (people and

#### IMPROVE YOUR PERFORMANCE

- Use the Admiralty Code as a starting point for data source validation.
- Set a 30-day time limit for vetting data sources and ensuring the data they provide aligns with intelligence requirements.

time) to perform their own validation of sources. In these cases, the notion of "trust but verify" becomes simply "trust." Organizations also explained that some third-party intelligence providers apply different types of ratings and scores that pertain only to the credibility of the data, yet there is no rating or scoring regarding the data source itself. Additionally, because some third-party intelligence providers generate scores using their own proprietary algorithms, organizations often have no clear understanding for the reasoning behind a given score.

## BEST PRACTICES

**Evaluate data sources in a repeatable and transparent way that incorporates multiple sources**
High-performing organizations have formal, holistic, transparent, and repeatable processes for evaluating data sources. These organizations receive third-party intelligence from vendors, yet perform additional separate validation. One organization explained that all internal and external data sources are currently manually reviewed, assessed, and classified every 30 days by a qualified analyst and to ensure they are correctly aligned to intelligence requirements. Another organization looks for a minimum of three data sources to corroborate each source's reporting. Some organizations, especially those in law enforcement, validate the data and data sources to the point that there is no uncertainty. There are no confidence levels because "evidence" they gathered must be able to stand up in a court of law.

**Building off the Admiralty Code for source validation**
A number of high-performing organizations and third-party intelligence providers that generate original context use the NATO or Admiralty Code Grading System[30] for conveying source reliability and credibility of information. The Admiralty Code, which provides a binary rating system that considers the reliability of both sources and the information they provide, is a positive step toward a common lexicon or ontology for data source validation. Additionally, the Admiralty Code is an incorporated taxonomy in the Malware Information Sharing Platform[31] (MISP), a free and open source threat sharing platform used by organizations we met.

## EVALUATION OF SOURCE RELIABILITY

| | | |
|---|---|---|
| A | Reliable | No doubt of authenticity, trustworthiness, or competency; has a history of complete reliability |
| B | Usually Reliable | Minor doubt about authenticity, trustworthiness, or competency; has a history of valid information most of the time |
| C | Fairly Reliable | Doubt of authenticity, trustworthiness, or competency but has provided valid information in the past |
| D | Not Usually Reliable | Significant doubt about authenticity, trustworthiness, or competency but has provided valid information in the past |
| E | Unreliable | Lacking in authenticity, trustworthiness, and competency; history of invalid information |
| F | Cannot Be Judged | No bias exists for evaluating the reliability of the source |

30 https://fas.org/irp/doddir/army/fm2-22-3.pdf, https://en.wikipedia.org/wiki/Admiralty_code

31 https://www.misp-project.org/features.html

## EVALUATION OF INFORMATION CONTENT

| | | |
|---|---|---|
| 1 | Confirmed | Confirmed by other independent sources; logical in itself; consistent with other information on the subject |
| 2 | Probably True | Not confirmed; logical in itself; consistent with other information on the subject |
| 3 | Possibly True | Not confirmed; reasonably logical in itself; agrees with some other information on the subject |
| 4 | Doubtfully True | Not confirmed; possible but not logical; no other information on the subject |
| 5 | Improbable | Not confirmed; not logical in itself; contradicted by other information on the subject |
| 6 | Cannot Be Judged | No bias exists for evaluating the validity of the information |

**Toward a Common, Robust Lexicon for Validating Data Sources**

Trusting data and data sources—identifying what is true and not true and having confidence that data is accurate, is reliable, and hasn't been tampered with—will become a more important challenge in coming years. As more organizations turn to machine learning to assist with decision making and prediction analysis, data quality is increasingly important; organizations must be able to validate the data and models used, and explain the process. Additionally, learning models can be vulnerable to poisoning, model inversion, and extraction attacks that could bias or trick a model's output. The potential for attacks like these means that demonstrating and explaining data source validation will require a greater level of detail, vetting capability, and transparency.

The Admiralty Code is a framework that high-performing organizations are using to form a common approach for vetting data sources (Evaluation of Source Reliability) and data (Evaluation of Information Content). It also provides a simple binary lexicon for explaining source reliability and information content. Potential exists to build upon the Admiralty Code to vet and explain a source's authenticity, reliability, and freedom from hostile control.

*“When everything is intelligence—nothing is intelligence.”*

—Wilhelm Agrell
University of Lund, Sweden

# Threat Analysis

*Technical Approach to Inform Cyber Intelligence*

## INTRODUCTION

Threat Analysis is the assessment of technical telemetry and non-technical data pertaining to specific threats to your organization and industry to inform cybersecurity operations/actions and Strategic Analysis. Threat Analysis is built on operational and tactical analysis and enhances CSO/CISO and other mid- to senior-level decision making.

- **Tactical Analysis:** Analysis of specific threats, attacks, incidents, vulnerabilities, or unusual network activity that enhances decision making for network defenders, incident responders, and machines pertaining to cybersecurity and incident response. Information analyzed is usually technical telemetry such as network and endpoint activity, atomic, behavioral, and computed indicators such as malware samples, hash values, domains, IPs, logs, and email header information. Tactical analysis tends to answer specific intelligence requirements and immediate, daily, and weekly what/where/when/how questions about threats.
- **Operational Analysis:** Analysis of specific threats, threat actors, and threat actor campaigns, intentions, and capabilities against an organization and its industry. Operational Analysis answers priority and specific intelligence requirements (PIRs, SIRs[32]) to enhance CSO/CISO and other mid- to senior-level decision-makers' leadership decisions regarding non-immediate but near-term (weekly–quarterly) business process and cybersecurity decisions.

### THREAT ANALYSIS ASSESSMENT FACTORS

In evaluating the state of the practice of cyber intelligence in terms of Threat Analysis, we considered the following factors:

1. Threat Analysis Workflow
2. Timeliness and Accuracy of Threat Analysis
3. Diversity in Technical Disciplines
4. Traits, Core Competencies, and Skills
5. Threat Analysis Tools

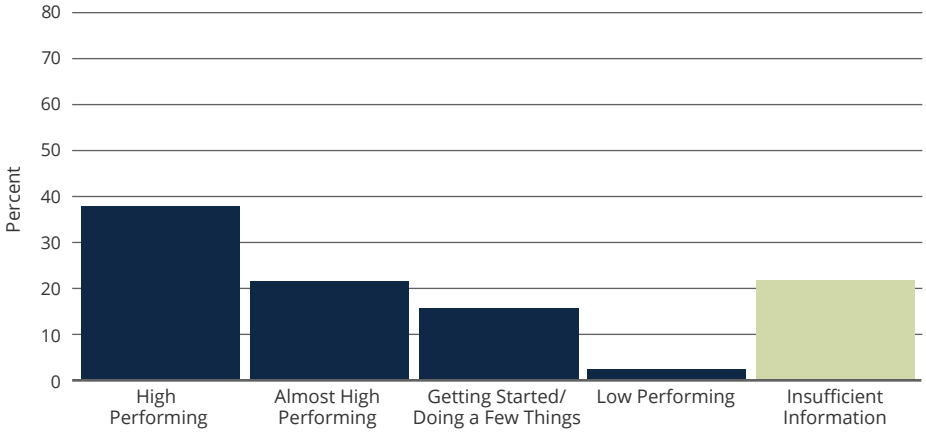
32 https://digital-forensics.sans.org/blog/2009/10/14/security-intelligence-attacking-the-kill-chain

## THREAT ANALYSIS FACTOR 1: THREAT ANALYSIS WORKFLOW

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization has a defined and repeatable Threat Analysis workflow with clear timelines, roles, and responsibilities. The workflow incorporates other Cyber Intelligence Framework components to provide analysis on specific threats to the organization and industry for the purposes of informing cybersecurity operations/ actions and Strategic Analysis.

### PERFORMANCE SNAPSHOT




**Threat Analysis Factor 1 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**No formal Threat Analysis workflow**

Organizations struggle with workflows that are largely conceptual and abstract; for these organizations, no documentation exists for workflow triggers, roles, responsibilities, or timelines to produce Threat Analysis. Although this challenge was more common in smaller organizations, even some larger organizations lacked formally documented and accessible processes and procedures.

We also interviewed organizations that described specific challenges: some lack a ticketing/tracking system to show the status and workflow steps pertaining to an incident. Some organizations that have a Threat Analysis workflow are struggling to integrate their organization's threat prioritizations into the workflow or to get their vendor to understand the organization's threat prioritizations.

**Threat Analysis workflow is the only workflow**

We did meet organizations with defined and documented Threat Analysis workflows supporting cybersecurity and incident response missions. Some organizations, often due to the recurring challenge of resource constraints, only focus on internal technical telemetry

**IMPROVE YOUR PERFORMANCE**

- Create a defined and repeatable Threat Analysis workflow.
- Use public threat frameworks and SOAR technologies to assist with Threat Analysis and workflow creation.

and do not receive or conduct analysis on other technical and non-technical data feeds from internal business units, industry data, or third-party strategic intelligence. Without this information, the cyber intelligence team lacks the insight to produce Strategic Analysis.

### THREAT ANALYSIS GENERAL WORKFLOW:

1. Know your environment
2. Identify and understand gaps and intelligence requirements (IRs, and especially PIRs, SIRs)
3. Collect/normalize internal and external telemetry from data sources
4. Conduct tactical analysis to answer what/where/when/how questions regarding threats, attacks, incidents, vulnerabilities, or other unusual network activity for the purpose of generating human and machine mitigating actions
5. Conduct operational analysis, adding context (threat actors, campaigns) to existing tactical intelligence, starting to answer the who and why behind threats
6. Enhance mid- to senior-level leadership decisions regarding non-immediate but near-term (weekly–quarterly) business process and operational decisions.
7. Leadership provides feedback

### BEST PRACTICES

**Create a Threat Analysis playbook**

High-performing organizations have Threat Analysis playbooks that ensure their workflows are defined, documented, repeatable, and scalable. Roles, responsibilities, and timelines are clearly understood. Many of these organizations also use SOAR and other customized platforms to manage the process.

Threat Analysis workflows for some high-performing organizations start when indicators are automatically correlated and matched against internal network and endpoint telemetry activity in a SIEM. Pre-built alerts notify a junior cyber defense analyst to decide if the alert requires additional analysis. For alerts that require additional analysis, the cyber defense analyst creates a new case within the SIEM, TIP, SOAR Platform, JIRA, or other customized platform with read/write/edit privileges for the entire fusion center.

Threat Analysis workflows in other high-performing organizations operate like a tree diagram, and analysis proceeds when certain thresholds are met or workflow milestones are completed. If a threshold for additional analysis is met, a senior cyber defense analyst or cyber defense incident responder becomes the lead analyst. The lead analyst gathers additional current and historical data with assistance from a team of analysts in the fusion center. These analysts have the option to simultaneously add input to the case at any time.

**Use common frameworks and tools**

Many high-performing organizations are using the MITRE ATT&CK Framework to identify and understand adversarial tactics and techniques that interact with their systems. They also rely on Zeek (formerly Bro) in addition to a SIEM, EDR, or IDS/IPS utility. Zeek assists with searching historical data, malware, and network traffic analysis, and other interesting and important technical data such as user agent strings, protocols, headers, mac addresses, IPs, and certificates. High-performing teams then evaluate collected data, validate the data and data source, and make analytical judgments about the

threat potential to the organization with recommendations for mitigation. Depending on the severity of the threat, the fusion center may immediately take action to stop and remediate the threat and will later report to leadership and other internal business units about the threat and actions taken. Again, Threat Analysis is threat specific and enables mid- to senior-level leaders to make immediate to near-term decisions about cyber hygiene, cybersecurity, and incident response to ensure sustained success of business processes and operations.

**Save time and resources by using security orchestration, automation, and response (SOAR) technologies**

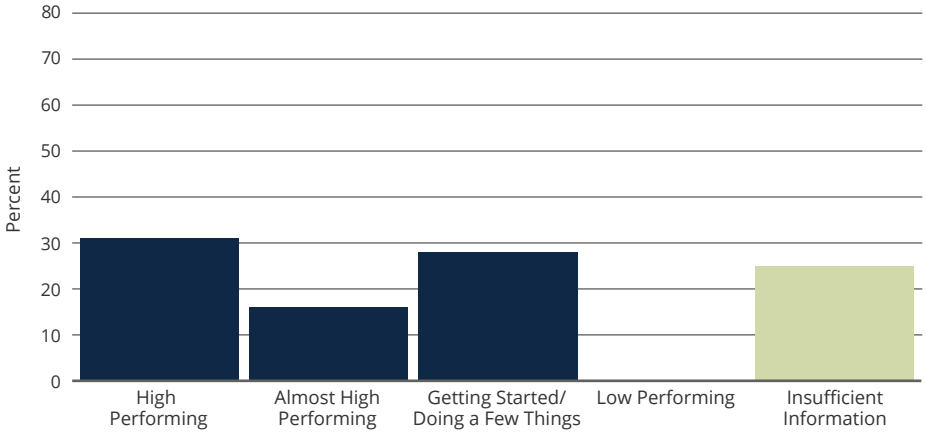
Some high-performing organizations use SOAR solutions to support Threat Analysis. When configured appropriately, SOAR technologies can be a force multiplier for organizations with limited staff and time—especially when analysts are drowning in repetitive manual tasks. SOAR technologies automatically connect and coordinate disparate cybersecurity tools, threat intelligence platforms, and other non-security tools and products into a single dashboard. By connecting these tools—as well as people—a SOAR solution automates data enrichment and the execution of tasks like parsing URLs, file detonation, performing historical searches, and sending attachments or indicators through tools like VirusTotal or WHOIS. This automation saves response time and reduces analyst workload and human error. The SOAR tool also works with an organization's playbook, allowing organizations to create playbooks from templates or to customize a playbook. The playbooks mimic a tree diagram process with scheduled timelines for sequential or multiple tasks.

## THREAT ANALYSIS FACTOR 2: TIMELINESS AND ACCURACY OF THREAT ANALYSIS

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization's cyber intelligence team is capable of producing time-sensitive and multi-source validated functional analysis. The cyber intelligence team provides analytical updates as needed for information sharing and decision making purposes.

### PERFORMANCE SNAPSHOT




**Threat Analysis Factor 2 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

### Inadequate reporting

Many organizations do not produce Threat Analysis reports due to common challenges like lack of resources or lack of process. Others struggle to produce reports in a timely manner: one organization explained that four days is considered a quick turnaround given their entire Threat Analysis workflow, from environmental context to report generation and feedback. Still others produce reports that do not include data source validation language, estimative language, or acknowledgment of intelligence gaps.

**IMPROVE YOUR PERFORMANCE**

- Use a "shot clock" for Threat Analysis reports on particular issues.
- Include data source validation scores, estimative language, and acknowledgment of intelligence gaps in Threat Analysis reports as appropriate.

## BEST PRACTICES

### Create processes to support speed and efficiency

High-performing organizations place a premium on speed and efficiency with formalized processes, plans, and timelines for report generation based on event/incident severity. A high-performing organization described their formalized "shot-clock" process for producing Threat Analysis reports: depending the severity of a case, the team must answer immediate leadership requirements within one hour. Within 24 hours, the team must complete an incident analysis or notification report with added original context/analysis and actionable recommendations for decision makers.

To meet leadership-approved timelines, many high-performing organizations incorporate milestones and timelines into SOAR playbooks to assist with Threat Analysis and incident notification reports. Some organizations also have service level agreements (SLAs) with other internal business units and external partners that dictate timelines for delivery of functional reports.

### Provide specific and actionable reporting

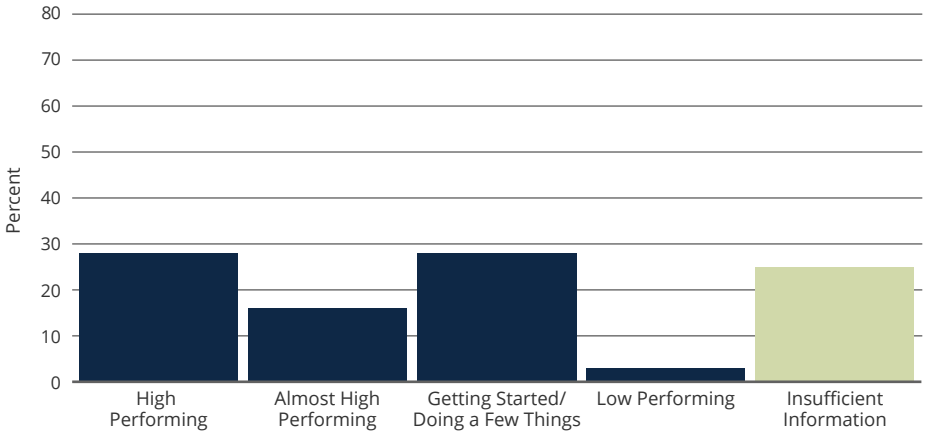
A number of high-performing organizations we met, specifically in the finance, health and public health, and government facilities sectors, produce a variety of Threat Analysis reports such as daily reports, weekly situational reports, vulnerability notification reports, after-action reports, and monthly and bi-monthly technical reports on malware behavior, and network and user-behavior telemetry trends. These reports tend to be actionable/operational in nature and are targeted to fusion center leadership and the CISO.

## THREAT ANALYSIS FACTOR 3: INCORPORATING DIVERSE DISCIPLINES TO CONDUCT THREAT ANALYSIS

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization has a repeatable process and structure to incorporate diverse technical knowledge for Threat Analysis. The organization regularly evaluates that process to ensure it incorporates the technical knowledge and skills to conduct effective and comprehensive Threat Analysis.

### PERFORMANCE SNAPSHOT




**Threat Analysis Factor 3 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**Lack of technical diversity**

Some organizations simply do not have a diversity of skills represented on their teams. Even large organizations may have small teams made up of members with similar technical backgrounds. Other cyber intelligence teams explained that they are unable to get management approval to hire new team members, or that they have no evaluation methodology to ensure the team has the right number of people with the right skill sets.

**Lack of visibility into technical skills**

Many organizations explained that no information about skills is documented. The team simply knows who to go to for any particular technical situation. In small organizations with cyber intelligence teams of 1-3 people, a conceptual process makes sense. For larger teams, the lack of a formal process to incorporate diverse technical skills raises challenges. For example, one team explained that at times they actually do not know who is working on a ticket or issue. For other organizations, the CISO or management simply selects the analyst(s)

**IMPROVE YOUR PERFORMANCE**

Identify, document, and publish a listing of all team members with technical skills to support Threat Analysis.

they think should work on a particular technical issue. This approach leads to single points of failure when the manager or analyst is not available, or if the manager is not aware of all technical skills and experiences existing within the organization.

### BEST PRACTICES

**Know and document your team's skills**

High-performing organizations have teams that have an informal understanding of team member skills as well as formal documentation of team member technical skills and expertise. These organizations have the types of organic relationships we saw in fusion centers, where analysts often know who has what skills based on working closely together. But these organizations also document team member skills and ensure they are visible across the entire team. One high-performing organization has created a matrix listing subject matter experts and their skills sets. This helps the entire organization quickly triage events and assign the right technical analyst as well as identify appropriate peer-review analysts.

**Open lines of communication with support from management**

Many high-performing organizations recognize that creating a process to pull in the right analysts at the right time is largely a management responsibility. This doesn't mean that managers always pick the analyst(s) they want working on a particular issue. Rather, management creates open lines of communication (across the fusion center and the entire organization) that are effectively aligned to ensure that team members with the right skills are pulled in at the right time. While management ensures lines of communication are open, the whole team must participate in proactive communication necessary to incorporate the right people. For example, high-performing organizations often hold weekly sync meetings to educate everyone on current issues and work status. These sync meetings also help everyone know where expertise and transactional memory exists across the team.

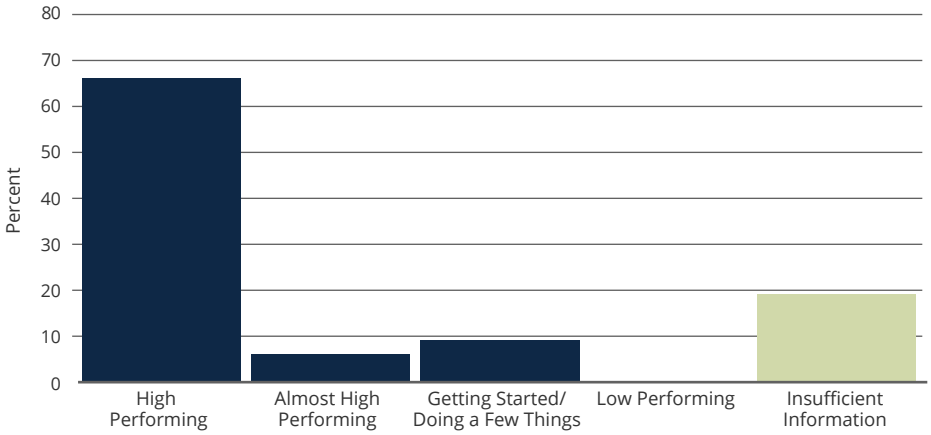
## THREAT ANALYSIS FACTOR 4: TRAITS, CORE COMPETENCIES, AND SKILLS

### WHAT THIS ASSESSMENT FACTOR MEANS

Threat Analysts are deeply skilled in computing fundamentals, cybersecurity, technical exploitation, cyber forensics, data collection and examination, networking, and incident response. They are generally inquisitive, persistent, open-minded critical thinkers and problem solvers. Threat analysts are familiar with intelligence analysis, computer science, and data science. Opportunities for formal and informal training are available and encouraged for team members to keep core competencies and skills fresh.[33]

33 For a list of more specific traits, core competencies and skills, see CITP1 Training and Education White Paper and NIST NICE SP 800-181

## PERFORMANCE SNAPSHOT




**Threat Analysis Factor 4 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

### Small cyber intelligence teams and limited opportunities for training

Some organizations we met have small cyber intelligence teams and rely heavily on third-party intelligence providers. Even when some of these teams collaborate on particular issues, they are unable to cover necessary skills, core competencies, and traits to perform effective Threat Analysis. Such organizations also explained that they struggle with identifying people to hire that are technically proficient in more than one technical discipline. In other words, a candidate may have excellent experience in networks and networking but little experience with malware or programming.

### Technical teams that lack people skills

Some cyber intelligence teams have highly technical people, yet those team members lack communication, collaboration, and self-awareness skills. One organization expressed that it would be beneficial for the team to learn about emotional intelligence.

### No management buy-in for training

Some cyber intelligence teams explained that there isn't much encouragement, funding, and opportunity to attend technical training or conferences.

## BEST PRACTICES

### Build teams with depth and breadth in technical disciplines

High-performing organizations have deep and wide benches across many technical disciplines. From a strictly technical standpoint, high-performing organizations have team members with backgrounds that broadly fit into computing fundamentals, cybersecurity, technical exploitation, data collection and examination, communication and collaboration, and applied artificial intelligence. More specifically,

**IMPROVE YOUR PERFORMANCE**

- Use NIST SP 800 -181 as a hiring guide; look for individuals with subject matter knowledge across many technical disciplines and deep technical expertise in a least one discipline.
- Ensure technical applicants have critical thinking, self-awareness, and communication skills.
- Test applicants by having them provide a work sample addressing a relevant cyber issue.
- Require new employees to complete mandatory introductory training on a particular technical specialization.
- Conduct internal mock threat scenarios where new analysts draft and brief threat assessments.
- Match new employees with senior technical analysts for ongoing mentoring.

we met people skilled in forensics and malware analysis, reverse engineering, intrusion analysis, incident response, network forensics, network and information architecture engineering, operating systems, networking, mobile devices, mobile and web applications, social engineering, operational technologies, vulnerability analysis, cryptography, penetration testing, programming and software development, data science, and machine learning.

High-performing organizations expressed that many team members have deep knowledge and experience with a variety of tools or that they are fast learners. Individuals need to rapidly manipulate tools to generate additional context and provide options and solutions quickly for decision makers.

**Test candidates for technical skills and look for non-technical skills**
High-performing organizations commonly assess skill gaps across their teams. Then, using NIST NICE 800-181 as a guide, they look to hire individuals with a proven record of expertise, aptitude, hands-on tool familiarity, and a deep desire to learn and improve. Many organizations explained that experience carries greater weight than education. They also test applicants with some type of work sample. For example, one organization evaluates applicants based on whether they can choose an important cyber intelligence question and answer it effectively.

**TIP: HIRING**

A common theme when hiring is to shoot for the letter "T" model for technical positions, meaning that employees should have broad subject matter knowledge and experience across many different Threat Analysis disciplines and one area in which they have tremendous technical depth and experience. Better than the "T" model, is Π, where an employee has broad knowledge and experience across many different cyber intelligence disciplines and two areas of technical depth and experience.

Many organizations expressed that while a basic understanding of IT and cybersecurity is important, technical skills can be taught. A major theme throughout our interviews with study participants was the importance of non-technical skills. Organizations across finance, health and public health, government, and the defense industrial base sectors emphasized the importance of a passion to learn, curiosity, open-mindedness, adaptability, critical thinking -specifically problem solving, and the ability to communicate effectively without ego (writing, briefing) technical concepts to different audiences. Additionally, individuals performing Threat Analysis should have familiarity with and understanding of intelligence analysis and structured analytical techniques.

**Create a culture that encourages everyday learning and training**
High-performing organizations recognize that experts want to work for winning and highly capable companies—training their people is good for morale and their bottom line. Organizations we interviewed, specifically in finance, energy, and government facilities, continuously provide a variety of internal and external learning and training opportunities. Examples include mandatory introductory training for new employees in particular technical areas, conducting internal

mock threat scenarios where newer analysts draft and brief threat assessments, and matching new employees with senior technical analysts for ongoing mentoring.

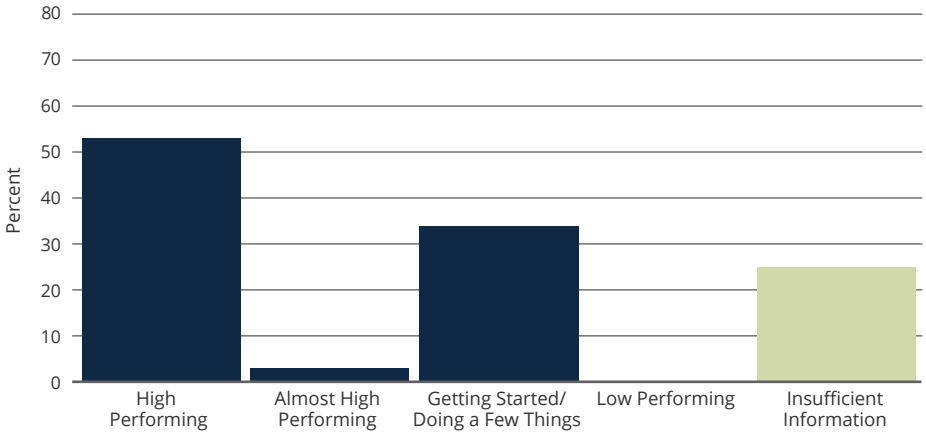
Many high-performing organizations encourage employees to take online technical training and attend conferences, technical exchanges, and free webinars. These organizations typically also have large budgets for training, in some cases more than $8,000 per year per employee. Employees are sent to well-known industry training venues and conferences to build technical skills on topics such as malware and network analysis, forensics, and incident response—and to make connections with other cyber intelligence professionals. Employees receiving technical training or attending conferences brief or teach their team about what they learned when they return. Some organizations additionally set funding aside for outside vendors to visit on-site and train the team on a particular skill or new tool.

## THREAT ANALYSIS FACTOR 5: THREAT ANALYSIS TOOLS

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization has an appropriate combination of homegrown and off-the-shelf technical analysis tools to support Threat Analysis. Tools are appropriately configured to support the organization, are readily available, and are evaluated routinely to ensure they meet organizational needs.

### PERFORMANCE SNAPSHOT




**Threat Analysis Factor 5 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

### Getting the right tools and technology

Some organizations expressed challenges with creating or acquiring technology to support Threat Analysis. For instance, we met with an organization relying primarily on email as its mechanism for data collection, management, and analysis. Additionally, we interviewed organizations expressing the need, yet lacking the purchasing authority, to acquire new and better technology. Some organizations are specifically seeking DLPs, better event correlation and analysis tools, and integration technologies like a SIEM, SOAR, EDR, DLP, or TIP.

We also met organizations that don't have people with the skills/expertise to build customized programs and tools, or write scripts to make internal and external information more useful to their organization's needs. Last, we interviewed organizations that explained they do not have a process/strategy for evaluating their current tools and technologies against future needs to perform Threat Analysis.

**IMPROVE YOUR PERFORMANCE**

- Create a strategy to analyze your current tool and technologies needs to identify current gaps and future needs.
- Use a diverse set of tools (open source, off the shelf, and homegrown) to support Threat Analysis.

## BEST PRACTICES

### Create a strategy for using open-source, free, paid, and customized tools and technologies to support Threat Analysis

A practice of high-performing organizations is creating a Threat Analysis tools and technologies strategy. Such a strategy usually involves regular evaluation of current organizational tools and technologies vs. current needs, identification of tools and technologies that will be built in-house vs. purchased, and identification of tools and technologies needed in the next few years. Routine evaluation of tools and technologies ensures they assist the cyber intelligence team in performing effective Threat Analysis to answer changing SIRs, PIRs, and IRs. A method for evaluation may involve leadership issuing an annual or bi-annual solicitation for tool and technology requirements from the fusion center and other parts of the organization to understand organizational needs before exploring COTS or in-house solutions.

High-performing organizations also take the necessary time to configure and test new tools and technology before launching them on their network.

### Use tools to their full potential

High-performing organizations use tools like IDA Pro, Joe Sandbox, Virus Total Premium, Splunk, and RSA NetWitness. Other tools include Kali Linux, MISP, WHOIS, Cuckoo Sandbox, VirusTotal, OllyDbg, Shodan, Wireshark, Snort, the ELK Stack (Elastic Search, Logstash, and Kibana), and Zeek.

The following are just a few interesting examples of how organizations we interviewed use tools and technologies to support Threat Analysis:

- As a premium customer of VirusTotal Intelligence, the cyber intelligence team creates specific YARA rules looking for indicators important to their organization. When a team member uploads a file to VirusTotal and it meets the team's established criteria, the team is immediately alerted. The team then retrieves the document for additional investigation.
- The cyber intelligence team uses Splunk and Zeek concurrently for analysis and validation. The organization's Zeek clusters provide analytics on network traffic such as top protocols, top talkers, and top ports, acting as an audit on top of Splunk.
- The cyber intelligence team writes scripts to facilitate IOC extraction from .pdf and .doc files, and creates tools to perform secure remote file retrieval. The team is working on creating ML algorithms for use in Splunk to identify anomalous user activity, malware beaconing, and data exfiltration.
- The cyber intelligence team is building in-house malware labs for testing and analysis using open-source tools such as VMware, pestudio, Process Monitor, Process Explorer, Wireshark, and Zeek.
- The organization has created a system where a neural network is fed normalized data (indicators and artifacts) using Natural Language Processing (NLP). The system then searches for matches with 100% malicious activity and has the option of generating risk and threat judgments reports to the appropriate human analysts for additional analysis.

**TIP**

As part of our research, we captured a list of the tools participants are using for Threat Analysis. See **Appendix: Most Popular Cyber Intelligence Resources**.

**TIP**

See the implementation guide "Artificial Intelligence and Cyber Intelligence" to learn more about using machine learning to support Threat Analysis on challenges such as malware attribution, insider Threat Analysis, and identifying, sorting, and prioritizing information.

*"If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle."*

—Sun Tzu
*The Art of War*

# Strategic Analysis

*Understanding the Big Picture*

## INTRODUCTION

Strategic Analysis is the process of conducting holistic analysis on threats and opportunities. Holistically assessing threats is based on analysis of threat actor potential, organizational exposure, and organizational impact of the threat. Strategic Analysis answers "who" and "why" questions related to threats and threat actors.

Strategic Analysis is not only comprehensive, but anticipatory. Strategic Analysis goes beyond Threat Analysis to incorporate analysis regarding emerging technologies and geopolitics that may impact or provide opportunities for the organization now and in the future. It can be actionable, enabling executive leaders to make risk-based decisions pertaining to the organization's financial health, brand, stature, and reputation.

### STRATEGIC ANALYSIS ASSESSMENT FACTORS

In evaluating the state of the practice of cyber intelligence in terms of Strategic Analysis, we considered the following factors:

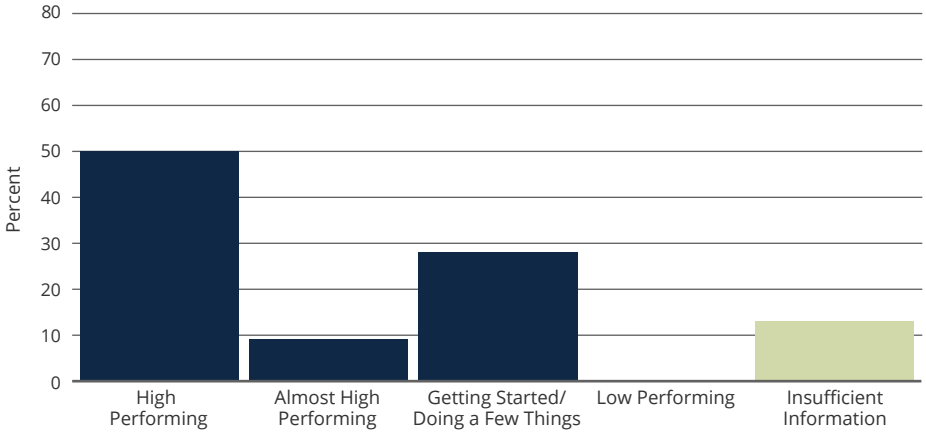
1. Understanding the Difference Between Strategic Analysis and Threat Analysis
2. Strategic Analysis Workflow
3. Diversity Among Strategic Disciplines
4. Traits, Core Competencies and Skills
5. Strategic Analysis Tools
6. Analytical Tradecraft Applied to Cyber Intelligence Analysis

## STRATEGIC ANALYSIS FACTOR 1: UNDERSTANDING THE DIFFERENCE BETWEEN STRATEGIC AND THREAT ANALYSIS

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization distinguishes between Threat Analysis and Strategic Analysis. Collaboration between threat analysts and strategic analysts is proactive and efficient.

## PERFORMANCE SNAPSHOT




**Strategic Analysis Factor 1 Performance Snapshot** The graph on the left shows how participants in the study are performing in this assessment factor.

## COMMON CHALLENGES

### Inability to implement Strategic Analysis

Most cyber intelligence teams we interviewed recognize the importance of performing Strategic Analysis, but many simply aren't doing it. Lack of resources and leadership commitment and understanding lead to cyber intelligence teams that are geared towards Threat Analysis to inform cybersecurity and/or cyber hygiene actions rather than anticipatory Strategic Analysis.

We met organizations without any strategic analysts and with no requisitions to perform that type of work. We also met organizations that have only one person on the entire team creating Strategic Analysis reports—a task too large for any one person, especially in larger organizations. One team explained that all of its leadership has backgrounds in cybersecurity and cyber hygiene and as a result, they do not understand the importance of Strategic Analysis. Most organizations lacking a Strategic Analysis capability tend to rely solely on third-party intelligence providers to provide that type of analysis.

**IMPROVE YOUR PERFORMANCE**

Create a separate and distinct Strategic Analysis team.

### Data silos

Additionally we encountered some strategic analysts discussing challenges accessing cybersecurity data and intelligence from cybersecurity or threat teams. Most of these data silos stem from differentiations in technology stacks, culture, sharing policies or SLAs, and teams being physically separated from one another. One organization explained that while their TIP supports threat actor profiling (good for Strategic Analysis) they face challenges mapping/ tagging data in the TIP to the MITRE ATT&CK framework, which could later be used to support Strategic Analysis.

## BEST PRACTICES

**Create a separate team focused on Strategic Analysis**

High-performing organizations have Strategic Analysis teams with formalized responsibilities, policies, and procedures—and those teams proactively collaborate with cybersecurity and threat teams. A large organization we met dedicated resources and commitment by standing up a 10-person team focused on Strategic Analysis.

## THE VALUE OF STRATEGIC ANALYSIS

In analyzing threat intelligence alongside non-traditional data from departments such as HR, physical security, and legal, strategic analysts develop depth, context, and perspective on particular issues. These analysts understand the circumstances that form the setting for a past, current, or future event, incident, or issue. They use this understanding to create reports and briefings to executive leadership that contain judgments and actionable recommendations, going from technical to non-technical with a risk-based perspective.

Strategic Analysis not only informs leadership about organizational risks; it also informs the more technical threat and cybersecurity teams about holistic current and future threats, risks, and opportunities. Analysis detailing threat actor behavior over time, or specific threat actor capabilities and intent, or even how emerging technologies enable new threats and opportunities gives these more technical teams insight into how to better prepare for and respond to events and incidents. Lastly, Strategic Analysis provides the collection management team with ideas and guidance on new areas for tasking data sources.

**Answer IRs and PIRs**

In high-performing organizations, strategic analysts aim to answer (usually in quarterly/annual reports and briefings) executive leadership-level intelligence requirements and priority intelligence requirements. This level of analysis is typically geared towards assisting executive leadership in making risk-based decisions pertaining to an organization's financial health, brand, stature, and reputation. Analysis can be extremely deep and detailed on a particular topic, and it can also be more broad-based and focused on trends.

**Foster collaboration**

Strategic Analysis provides technical threat analysts with insight on threat actors' motivations and capabilities, and threat and risk trends impacting the organization and industry. Because of the complimentary nature between Threat Analysis and Strategic Analysis, strong collaboration must exist. Most high-performing organizations have fusion centers or one location where all analysts physically sit together to foster that collaboration. However, we interviewed one high-performing organization that purposely locates its strategic analysts outside of the fusion center to prevent these analysts from becoming mired in the tactical and operational intelligence.

**Produce the right reports for your organization**

We interviewed strategic analysts who produce or contribute a variety of reports. A number of these reports focus on future threats and opportunities to the organization, which may help identify new intelligence requirements and research and development areas. Typical reports include

- ranking and tracking threat actor motivations, capabilities, and lifecycles against the organization's

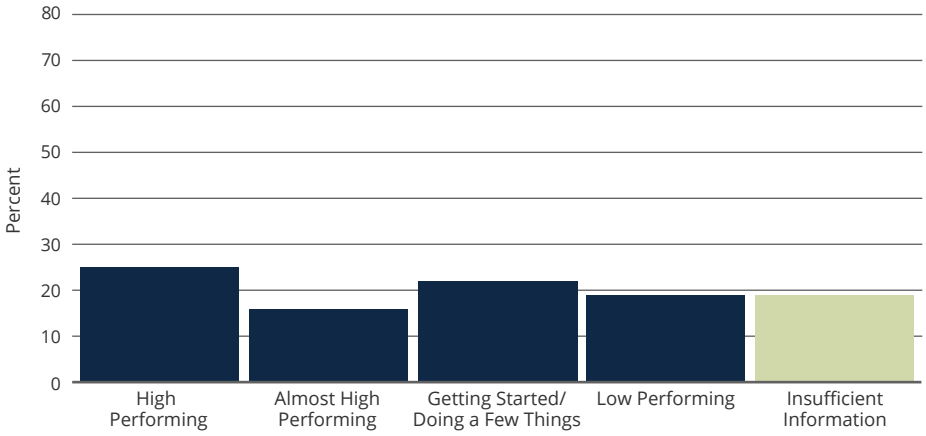
critical assets and technologies at risk
- tracking APTs as a mission and identifying threat actors (down to individual people) and why they are motivated to target the organization, its third parties, and industry
- identifying and mapping threat actors to geographic locations
- impact/opportunity presented by quantum computing, machine learning, 5G, and crypto-currencies
- foreign travel concerns
- opening a business in a foreign country
- where/what the organization should be investing in (technology, other companies)
- company mergers and acquisitions
- supply chain analysis
- how a particular technology may impact a line of business
- potential geopolitical, technological, and economic disruptions to business
- future foreign country forecasts
- assessing organizational emerging technology and how that lines up with company five year plans and threat actor capabilities
- assessing what specific threat incidents mean for the company moving forward
- targeting packages for the pen-testing team

## STRATEGIC ANALYSIS FACTOR 2: STRATEGIC ANALYSIS WORKFLOW

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization has a defined and repeatable Strategic Analysis workflow with clear timelines, roles, and responsibilities. The workflow incorporates other Cyber Intelligence Framework components to create analytical products holistically assessing threats, risks, and opportunities for the organization and industry.

### PERFORMANCE SNAPSHOT




**Strategic Analysis Factor 2 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**Non-existent, ad hoc, or multiple workflows**

Several of the organizations we met told us they did not produce Strategic Analysis reports; these organizations do not have a Strategic Analysis workflow. Other organizations incorporate aspects of a Strategic Analysis workflow; however the workflow is ad hoc, not formalized, and not repeatable. For instance, we met some organizations that produce Strategic Analysis reports—but they have no

established timelines for report production. Methodologies, processes, technologies, and templates used vary across analysts.

## SEPARATING WORKFLOWS: THREAT ANALYSIS AND STRATEGIC ANALYSIS

A recurring theme we noticed during our interviews was that participating organizations had difficulty distinguishing between strategic and non-strategic components and activities—and workflow is one area where we saw this difficulty. The workflows begin and end with the same components. However, unique components within the Threat Analysis workflow are designed to inform cybersecurity operations/actions, while components within the Strategic Analysis workflow involve holistically assessing threats, risks, and opportunities.

**Threat Analysis**

Performed to make immediate to near-term decisions pertaining to cyber hygiene, cybersecurity, and incident response (deny, disrupt, neutralize, deceive, exploit, defeat) to ensure sustained success of business processes and operations. It relies heavily on technical skills and is threat specific.

**General Workflow**

1. Know your environment
2. Identify and understand gaps and intelligence requirements (IRs, and especially PIRs, SIRs)
3. Collect/normalize internal and external telemetry from data sources
4. Conduct tactical analysis to answer what/where/when/how questions regarding threats, attacks, incidents, vulnerabilities, or other unusual network activity for the purpose of generating human and machine mitigating actions
5. Conduct operational analysis, adding context (threat actors, campaigns) to existing tactical intelligence; starting to answer the who and why behind threats
6. Enhance mid- to senior-level leadership decisions regarding non-immediate but near-term (weekly–quarterly) business process and operational decisions.
7. Leadership provides feedback

**Strategic Analysis**

Performed to holistically assess threats, risks, and emerging technologies and geopolitics that may impact/provide opportunities for the organization now and in the future. Informs threat analysts, the collection management team and enhances executive decision-making about organizational strategic issues and opportunities.

**General Workflow**

1. Know your environment
2. Identify and understand gaps and intelligence requirements
3. Fuse Threat Analysis with other external and non-traditional data sources
4. Depending on data collected, work with data science team to identify larger trends or anomalies in data collected
5. Perform structured analytical techniques and human-centered design activities as needed
6. Provide analytical assessments based on threat actor potential, organizational exposure, and organizational impact of threat
7. Analyze current and future technologies and geopolitics that may positively/negatively impact the organization and industry
8. Enhance executive leader decision making by answering IRs and providing intelligence pertaining to organizational strategic risks regarding financial health, brand, stature, and reputation
9. Leadership provides feedback

## BEST PRACTICES

**Attribution matters**

Attribution can be extremely challenging, especially in situations involving hybrid threats—cyber actors from a nation-state using some terror proxy group, cutout, or cartel to conduct the attack. That said, high-performing organizations recognize the importance of knowing your adversary. When organizations know the threat actor(s) intent on targeting them, they study and continuously monitor the threat actor's TTPs. This enables the organization to be anticipatory and take proactive measures against that specific threat actor(s). Working towards attribution (at any level: country, specific people, etc.) enables your cyber intelligence team to work with the collection management team to task new collection against that threat actor, revealing more insight into threat actor modus operandi. Armed with attribution knowledge, cyber intelligence teams can also generate targeting packages mimicking the specific threat actor TTPs to give to the penetration testing team. Last, attribution leads to accountability: high-performing organizations share attribution intelligence with the U.S. government (FBI and DHS) to hold malicious threat actors accountable. Sharing attribution intelligence with the proper government authorities enhances collaboration and trust between government, industry, and academia, and lets threat actors know that there are consequences for their actions.

### IMPROVE YOUR PERFORMANCE

- Create a defined and repeatable Strategic Analysis workflow to answer IRs and PIRs. The workflow should leverage all components of the Cyber Intelligence Framework to support Strategic Analysis on threats, risks, and opportunities.
- Focus on attribution to open new collection tasking against a particular threat actor, to reveal greater insight into threat actor modus operandi, and to assist with target package generation to mimic the specific threat actor for the penetration testing team.

## ATTRIBUTION RESOURCES

The ODNI *Guide to Cyber Attribution*[34] describes how analysts can assess responsibility for a cyber attack. The guide suggests three ways:

1. Point of origin (neighborhood, city, state, country, region)
2. Specific digital device or online persona
3. Individual or organization that directed the activity

The *Cyber Intelligence Tradecraft Project Threat Prioritization Guide*[35] provides categories for collecting and analyzing information on threat actor potential, which could assist with cyber attribution:

- infrastructure
- technology
- coding
- maturity
- targets of interest
- timing ability
- funding
- people
- tools and training
- intrinsic and extrinsic motivations
- targeted data and organizational systems

34 https://www.dni.gov/files/CTIIC/documents/ODNI_A_Guide_to_Cyber_Attribution.pdf

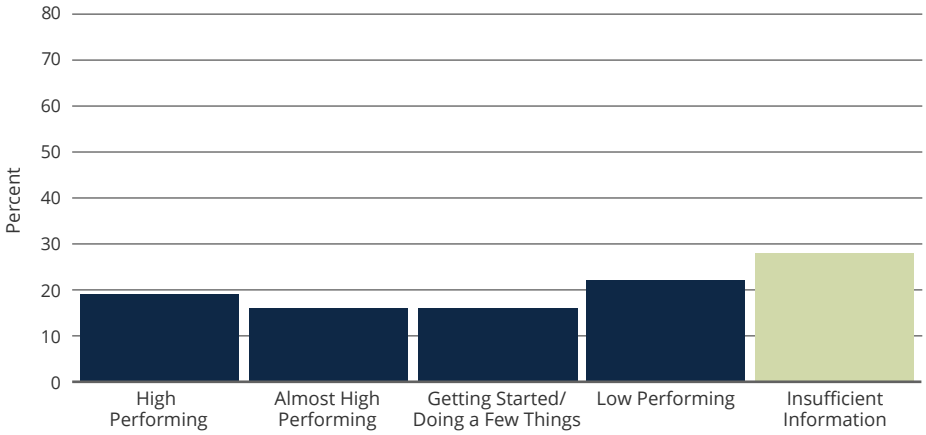
35 https://insights.sei.cmu.edu/sei_blog/2016/02/cyber-intelligence-and-critical-thinking.html

## STRATEGIC ANALYSIS FACTOR 3: INCORPORATING DIVERSE DISCIPLINES TO CONDUCT STRATEGIC ANALYSIS

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization has a process and structure to incorporate diverse disciplines to conduct Strategic Analysis. The organization regularly evaluates the Strategic Analysis process to ensure it incorporates the right knowledge and skills to enhance executive leadership decision making pertaining to organizational vital interests (financial health, brand, stature, and reputation).

### PERFORMANCE SNAPSHOT




**Strategic Analysis Factor 3 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**Resource constraints**

Organizations in all sectors have resource limitations preventing the production of strategic assessments. Organizations simply lack personnel to build a strategic team and as a result are unable to commit time and energy to produce these assessments. On some occasions we met with teams of one to three people responsible for both cybersecurity and cyber intelligence for large—even global—organizations.

**Hiring team members with the same skills**

Additionally, we met with organizations that struggle to diversify skills when hiring. They seem to hire individuals with the same skills and experience, typically those technically competent in cybersecurity, forensics, reverse engineering, intrusion analysis, operating systems, and network and information architecture engineering.

#### IMPROVE YOUR PERFORMANCE

- Create open lines of communication (across the fusion center and the entire organization) to ensure the right group of diverse people is pulled in at the right time.
- Hire data scientists to work with cyber intelligence analysts to identify trends, patterns, and anomalies.
- Regularly evaluate your organization's processes to ensure the right knowledge and skills across the entire organization are brought to bear on a particular issue.

**Lack of process**

Few organizations were high-performing for having a process to incorporate diverse disciplines to conduct Strategic Analysis. Some organizations maintain a process, yet explained it is more ad hoc in nature—nothing is written down explaining whose expertise is needed or good to leverage for particular issues. Analysts who do contribute to these products are typically analysts with the same experience or background.

## BEST PRACTICES

**Build collaboration in**

We met organizations that have entire teams performing Strategic Analysis. These analysts are typically intelligence analysts and geopolitical analysts. Analysts tend to be organized or assigned to threats, threat types, or regions or countries. A practice of high-performing organizations is to ensure there is proactive collaboration between strategic analysts and data scientists. The data scientists build tools for both strategic analysts and threat analysts. They also help with identifying trends and correlations. Indeed, one high-performing organization explained that "you need data scientists to win wars."

Another practice of high-performing organizations is to have a codified process to incorporate diverse disciplines to conduct Strategic Analysis. As noted in Threat Analysis Factor 3: Incorporating Diverse Disciplines to Conduct Threat Analysis, management investment and oversight ensures the right analysts are pulled in at the right time. High-performing organizations also regularly evaluate that process to ensure the right knowledge and skills across the entire organization are brought to bear.

To assist with Strategic Analysis, high-performing cyber intelligence teams bring in people with diverse backgrounds to participate in brainstorming sessions, weekly sync and collaboration meetings, and peer reviews of strategic products. An area of interest for future research might be exploring "SOAR-like" technology for automated data enrichment of data sets within and outside the organization and playbook generation that connects diverse analysts across the organization to contribute to strategic analytical products on holistic threats, risks, and opportunities.

## STRATEGIC ANALYSIS FACTOR 4: TRAITS, CORE COMPETENCIES, AND SKILLS

### WHAT THIS ASSESSMENT FACTOR MEANS

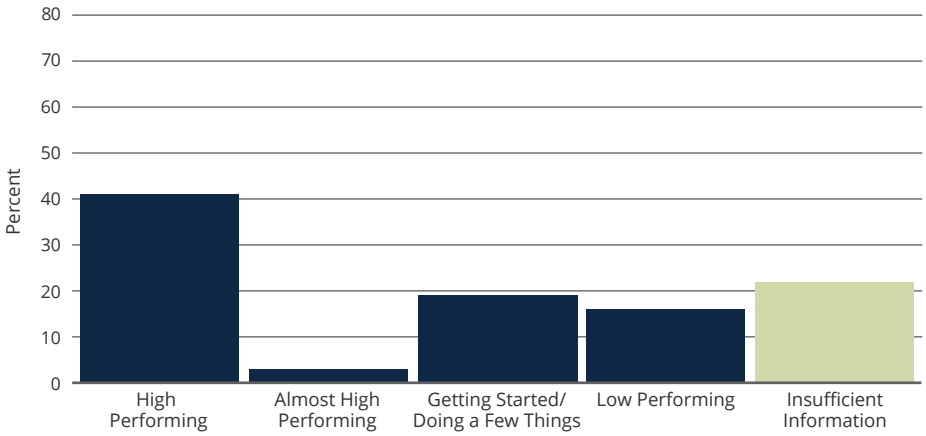
Analysts have the traits, core competencies and skills to perform Strategic Analysis. Many opportunities for formal and informal training are available and encouraged for team members to keep core competencies and skills fresh.

**Traits**
- curiosity
- persistence
- self-motivation
- intellectual independence
- ability to learn quickly
- open mindedness
- adaptability

**Core Competencies**
- critical thinking
- problem solving
- intelligence analysis
- data collection
- communication and collaboration
- knowledge about industry and geopolitics

**Basic Skills**
- computing and cybersecurity fundamentals
- technical exploitation
- computer science and data science

### PERFORMANCE SNAPSHOT




**Strategic Analysis Factor 4 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**Organizations lack personnel and leadership commitment to perform Strategic Analysis**

Some organizations interviewed performing Strategic Analysis face talent and acquisition challenges, tending to lack a bench of analysts to support this level of analysis. Other organizations with Strategic Analysis teams explained that some team members have no intelligence analysis experience or background in analytical techniques or geopolitics. When it came to formal and informal training for Strategic Analysis, organizations we interviewed primarily

**IMPROVE YOUR PERFORMANCE**
- Refer to NIST SP 800-181 as a guide for hiring to perform Strategic Analysis.
- Hire individuals that have experience and can demonstrate strong critical thinking skills. You can always teach and provide on-the job-training for technical skills.

in the finance and government facilities sectors indicated that they do not offer formal training in intelligence analysis, data collection, or human-centered design techniques. Training for these organizations is very much on the job.

**Difference in styles between military and other government agency trained intelligence analysts**

Based on their experience, a few industry organizations explained that hiring former military officers with training in intelligence may not be the best fit for Strategic Analysis. For instance, a team commented that military officers are usually more skilled and interested in operations and not Strategic Analysis and writing. Another team from a large industry organization remarked that officers with a straight military intelligence background (and NO technical experience) tend to see things in pure military terms and perspectives. For Strategic Analysis, these high-performing industry organizations recommend hiring intelligence analysts that have had experience from a "three letter" intelligence agency.

## BEST PRACTICES

**Prioritize critical thinking and other non-technical skills when building your team**

Critical thinking—specifically problem solving—is the skill high-performing organizations cited most frequently when describing their strategic team. Organizations explained that critical thinking skills are needed for identifying patterns, relationships, and sources and for corroborating information. One high-performing organization noted that their strategic analysts need to have the ability to think about problems in non-rigid ways, have a healthy skepticism, be imaginative, see the big picture, and have the foresight to ask broad questions, such as "Do we still need to be doing things this way?" Indeed, other organizations explained to us that they will always hire a candidate with a great analytical mind and a mediocre cyber background, over a candidate who has an extensive cyber background but is not a critical thinker. These organizations emphasized that while it is possible to provide technical training, it is more difficult to teach critical thinking.

A practice of high-performing organizations is to refer to NIST SP 800-181 as a guide for hiring individuals with the right knowledge, skills, and abilities (KSAs) to perform Strategic Analysis. The following NIST NICE SP 800-181 KSAs map to critical thinking and problem solving: S0359, A0035, A0080, A0081, A0070, A0106, A0118, A0122.

Other traits high-performing organizations either hire for or already have on their strategic team include intellectual independence, curiosity, tenacity, strong work ethic, inquisitiveness, the ability to let others poke holes in their analysis, recognizing when they don't know something, a sense of humor, confidence to arrive at judgments without complete information, and strong interpersonal skills and emotional intelligence. Many high-performing organizations also explained that their strategic analysts have a desire and passion to stay current on cyber threats, geopolitics, industry developments (always reading news and blogs), and developments within their own organization.

**Provide professional development to learn technical skills and make connections**

A practice of high-performing organizations is to send their intelligence analysts and other non-technical analysts to industry training venues and conferences to build and in some cases take introductory technical skills courses on topics such as network analysis, forensics, and incident response and to make connections with other professionals. Employees receiving technical training or attending conferences return and brief/teach their team about what they learned.

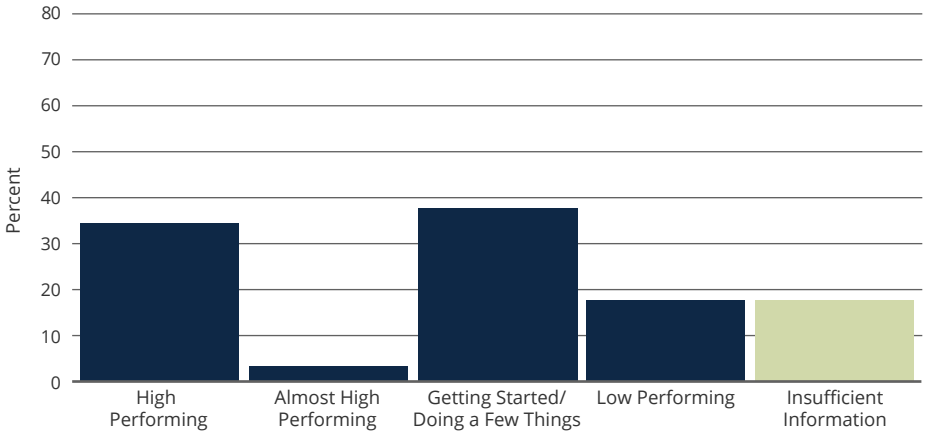
**Communicate clearly with technical and non-technical audiences**
Strategic analysts need sufficient technical acumen to communicate effectively with other technical analysts across the organization. They also need skills to communicate clearly and efficiently with non-technical audiences, specifically executive leadership. Many high-performing organizations, mostly in the finance, communications, food and agriculture, and government facilities sectors stressed how their analysts are strong at presenting at different altitudes depending on the audience. They are really strong at communicating (writing and briefing) the strategic context and risk perspective to executive leadership.

## STRATEGIC ANALYSIS FACTOR 5: STRATEGIC ANALYSIS TOOLS

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization has a combination of homegrown and off-the-shelf (as appropriate) tools to support Strategic Analysis. Tools are appropriately configured, readily available, and evaluated routinely to ensure they meet organizational needs.

### PERFORMANCE SNAPSHOT




**Strategic Analysis Factor 5 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**Over reliance on third-party intelligence provider assessments**
We met organizations that rely solely on third-party intelligence providers to provide strategic analytical assessments on threat actors, industry developments, and geopolitics. These organizations do not have tools and resources to conduct additional analysis incorporating third-party assessments and making them relevant to their specific organization's mission and interests. As noted earlier, we met with organizations where strategic analysts rely on spreadsheets to track threats and threat actors.

### Fragmentation of tools and knowledge

Some organizations we interviewed expressed the need for a single "pane of glass" across their systems that enables analysts to search and conduct analysis at all levels. Other organizations are hoping to purchase or create a knowledge management system that allows strategic analysts to access data and conduct analysis using a system like Palantir. That said, some organizations expressed that tools should not dictate or put strategic analysts in a box in terms of how they perform their jobs. A tool is one instrument assisting in the entire Strategic Analysis process.

## BEST PRACTICES

### Regularly evaluate Strategic Analysis tools

A practice of high-performing organizations is to regularly evaluate Strategic Analysis tools to ensure they meet current and future organizational needs. Evaluation leads to purchasing or building homegrown customized tools to make data and subsequent analysis relevant to the organization's mission. Before incorporating new tools on their network, these organizations ensure the tools are appropriately configured to integrate well with other tools.

### Use a mix of tools

Since Strategic Analysis is grounded in Threat Analysis and other non-traditional data sources, technical tools used for Threat Analysis are certainly useful for Strategic Analysis.

Most high-performing organizations additionally employ a good mix of analytical and visualization tools. Common Strategic Analysis tools used by high-performing organizations we met include ELK Stack, Maltego, MISP, i2 Analyst's Notebook, Palantir, Tableau, and Adobe InDesign and Photoshop. A more detailed list of tools can be found in the appendix **Most Popular Cyber Intelligence Resources**.

**TIP**

**THE PROMISE OF MACHINE LEARNING**

Incorporating machine learning into Strategic Analysis will become prevalent in the future as organizations find more efficient ways to complement Threat Analysis by gathering data from human resources, business intelligence, physical security, legal, marketing, finance, technology development, and corporate leadership and external technical and non-technical data about geopolitics, industry developments, and technology and innovation. Our Artificial Intelligence and Cyber Intelligence Implementation Guide discusses, among other thing, how machine learning can enhance Strategic Analysis on challenges such as identifying attack commonalities and associations between threat actors and events, and predicting possible insider threats or geopolitical events in a country.

**IMPROVE YOUR PERFORMANCE**

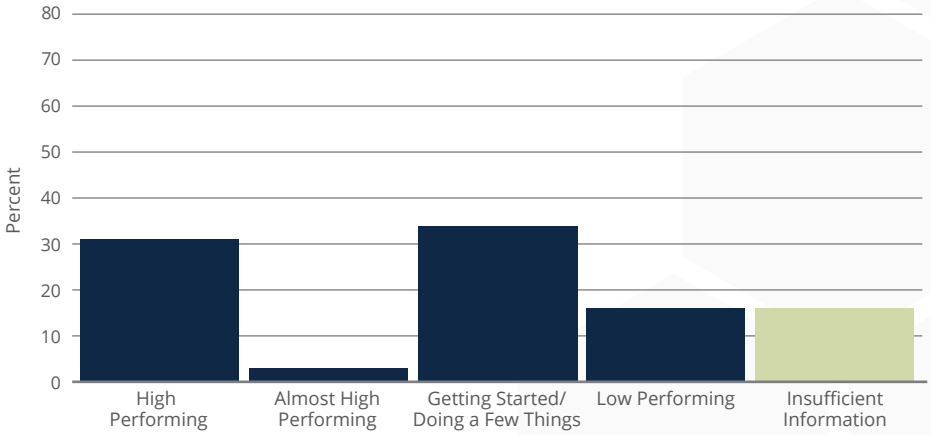
- Regularly evaluate Strategic Analysis tools to ensure they meet current and future organizational needs.
- Before incorporating new tools on your network, ensure they are appropriately configured to integrate well with other tools.
- Use a mix of analytical and visualization tools.

## STRATEGIC ANALYSIS FACTOR 6: ANALYTICAL TRADECRAFT APPLIED TO CYBER INTELLIGENCE ANALYSIS

## WHAT THIS ASSESSMENT FACTOR MEANS

The organization has a repeatable process for incorporating structured analytical techniques into its cyber intelligence analysis. The organization writes cyber intelligence reports that describe the quality of and credibility of sources and data methodologies, use estimative language (expressions of likelihood and confidence), are customer relevant, and incorporate visual information where appropriate. This process is reviewed and updated regularly.

## PERFORMANCE SNAPSHOT




**Strategic Analysis Factor 6 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

**Lack of formalized process for incorporating analytical tradecraft**
Several organizations we interviewed do not apply any analytical tradecraft into their analysis process. We met organizations not incorporating analysis of alternatives via structured analytical techniques or using estimative language (expressions of likelihood and confidence) in intelligence assessment reports to leadership. Organizations also do not include source descriptors and/or source validation, intelligence gaps and uncertainties, and the impact of intelligence gaps and uncertainties on assessments and judgments.

Other organizations explained they lack resources (people and time) to incorporate analytical techniques, yet recognize the importance of analytical tradecraft. Indeed, some organizations explained that their team could have benefited from learning about intelligence analytical standards earlier, as they have worked with others that have written assessments that jumped to conclusions, lacked analytical thought, and were personality driven.

Most organizations we met attempt to incorporate, albeit on an ad hoc basis, analytical tradecraft into workflows, specifically for performing Strategic Analysis. For these organizations, there is no agreed upon policy/procedure for how to incorporate analytical tradecraft into assessments. In other words, there is no formalized process in terms of when and how to include source descriptions and validation, expressions of likelihood, and confidence levels. Additionally, one organization explained that they will only occasionally perform Red Teaming, Analysis of Competing Hypotheses (ACH), or Devil's Advocacy for Strategic Analysis. Another organization talked about how only some analysts (not all) use estimative language and include intelligence gaps and source validation.

### IMPROVE YOUR PERFORMANCE

- Apply, as appropriate, structured analytical techniques on top of and in addition to cyber threat frameworks such as the Lockheed Martin Kill Chain and Diamond Model when performing Strategic Analysis.

**TIP**

**ANALYSIS OF ALTERNATIVES DEFINED**

"The systematic evaluation of different hypotheses to explain events or phenomena, explore near-term outcomes, and imagine possible futures to mitigate surprises and risks." (Intelligence Community Directive 203)

## MAKE INFORMED DECISIONS ABOUT HOW TO APPLY ANALYTICAL TRADECRAFT

It is generally not realistic to apply structured analytical techniques and analytical standards to every threat report. That is simply not feasible or logical when it comes to the speed and demands of mission (such network defense, cyber hygiene and incident response) and other fast-paced (machine and human) generated analysis that leads to immediate and near-term actionable cybersecurity focused recommendations. For example, it doesn't make sense to perform a structured analytical technique or write a report with source validations that suggest patch management or blocking an IP address. Just do the patch and block the IP address first, and then draft a weekly report or perform analysis later addressing the particular event(s), anomalous behavior(s), and mitigation actions taken. Organizations should make informed decisions, pending resources and timing based on threat/event criticality if incorporating analytical tradecraft into Threat Analysis is feasible either before or after mitigation actions are taken. Most reports, at least at the operational and Strategic Analysis level, should include estimative language, source descriptors or source validation, confidence level, and intelligence gaps.

## BEST PRACTICES

### Adopt ICD 203 and structured analytical techniques

A practice of high-performing organizations is to use Intelligence Community Directive 203[36] (ICD 203) as the foundation and guideline for applying analytic standards to their cyber intelligence analysis workflows. Most organizations we interviewed incorporate analytical standards into cyber intelligence analysis workflows, specifically when performing Strategic Analysis. While some processes are not truly formalized in these organizations, they do apply structured analytical techniques on top of and in addition to cyber threat frameworks such as the Lockheed Martin Kill Chain and Diamond Model. Structured analytical techniques are used to help the analyst be mindful of cognitive biases and logical fallacies and not "run on automatic."

Some organizations we met explained that they use these structured analytical techniques: brainstorming/ideation sessions, key assumptions checks, analysis of competing hypotheses, futures analysis, devil's advocacy, red teaming, decision trees, and what-if analysis. We met with cyber intelligence teams that conduct Root Cause Analysis. One organization brought in specialists to help their team perform Root Cause Analysis on a particular event. Other organizations bolster analytic rigor by purposely pairing intelligence analysts with data scientists on threat actor behavior deep dives, emerging threats, and opportunities assessments. Another high-performing organization gives each of its strategic analysts a copy of the CIA's *A Tradecraft Primer: Structured Analytical Techniques for improving Intelligence Analysis*.[37] They also post ICD 203 standards on an internal wiki page.

**TIP**

Organizations should explore human-centered design techniques such as Affinity Clustering, Bull's Eye Diagramming, and Importance/Difficulty Matrixes when evaluating threats, risk and opportunities. See the Luma Institute's *Innovating for People: Handbook of Human-Centered Design Methods.* **luma-institute.com**

**TIP**

Products expressing an analyst's confidence in judgments (Confidence Level) should *not* combine a degree of probability of an event or development (Very Likely) in the same sentence. Make them two sentences.*

**For example, don't write this:**

- "We assess with moderate confidence that cyber espionage malware ABC is linked to Threat Group XYZ and that its spear-phishing emails targeting machine learning experts will almost certainly continue in the near term."

**Write this instead:**

- "We assess with moderate confidence that cyber espionage malware ABC is linked to Threat Group XYZ."
- "Their use of spear-phishing to target machine learning experts in our organization and industry will almost certainly continue in the near term."

*https://fas.org/irp/dni/icd/icd-203.pdf

36 https://www.dni.gov/files/documents/ICD/ICD%20203%20Analytic%20Standards.pdf

37 https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/Tradecraft%20Primer-apr09.pdf

Some strategic analysts incorporate expressions of likelihood and confidence in their intelligence assessments. Strategic analysts in high-performing organizations (defense industrial base, government facilities, information technology, and communications sectors) are doing this and also include a scale or description describing the meaning of likelihood degrees and confidence levels. A number of high-performing organizations are pulling from the ICD 203 expression of likelihood scale and then use *high, medium/moderate,* and *low* to describe confidence levels for assessments or judgments in reports. A practice of high-performing organizations is to also include intelligence gaps, source descriptors/characterization, and source validation in both Threat Analysis and Strategic Analysis reports. Of the organizations we met that are doing this, source validation ratings are usually based on the Admiralty Code.

*"If you can't explain it simply, you don't understand it well enough."*

—Albert Einstein

# Reporting and Feedback

*Communicating with Teams and Decision Makers*

## INTRODUCTION

Reporting and Feedback is the communication of and subsequent feedback to analysts regarding their products and work performance. It identifies intelligence requirements and intelligence gaps.

### REPORTING AND FEEDBACK ASSESSMENT FACTORS

In evaluating the state of the practice of cyber intelligence in terms of Reporting and Feedback, we considered the following factors:

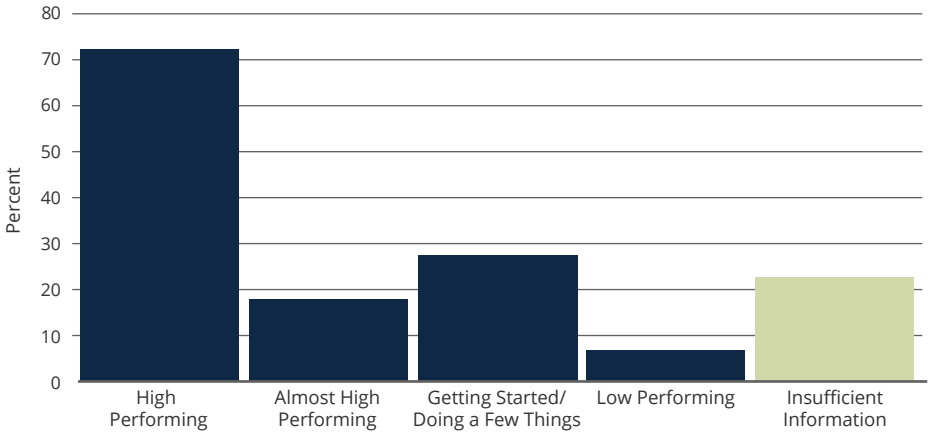
1. Cyber Intelligence Report Types
2. Actionable and Predictive Analysis
3. Leadership Involvement
4. Influence on Decision Making
5. Feedback Mechanisms for Analysts
6. Influence of Feedback on Data Gathering and Analysis
7. Satisfying Intelligence Consumers
8. Capturing Return on Investment

## REPORTING AND FEEDBACK FACTOR 1: CYBER INTELLIGENCE REPORT TYPES

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization applies a strategy and timeline to generate reports from a varied product line. The product line addresses customer needs, is stored, and can be accessed by internal and external partners as appropriate.

### PERFORMANCE SNAPSHOT




**Reporting and Feedback Factor 1 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**Lack of resources and leadership strategy to produce cyber intelligence reports**

We met organizations that do not produce cyber intelligence reports; most of these organizations explained that they do not have enough people and time to generate reports. With such limited resources, these teams can only focus on cybersecurity issues associated with cyber hygiene and incident response.

Several organizations explained that they produce Threat Analysis and Strategic Analysis reports on an ad hoc basis. There is no formalized schedule for report production or timeline for creating different report types. For other organizations, reports are simply event-driven emails. Other teams we met told us that there is no leadership (CISO and up) buy-in, vision, or strategy to create cyber intelligence reports. More specifically, these teams said that there is no strategy for a cyber intelligence product line and that they have received little guidance from leadership on requirements, timelines, and layouts for cyber intelligence reports.

Some organizations discussed challenges pertaining to delays in the review and dissemination process of operational and tactical-level

**IMPROVE YOUR PERFORMANCE**

- Create a strategy for a cyber intelligence product line that includes timelines and layouts for cyber intelligence reports.
- Build a varied cyber intelligence product line that addresses immediate needs, CISO and executive leadership requests, as well as specific internal business units and external customers and partner requests.

**TIP**

Reasons for why there is no leadership buy-in for report production could vary from budget constraints to a lack of understanding about cyber intelligence.

cyber intelligence reports. These organizations talked about workflow issues and the high number of coordinators. Today's reports are being disseminated about yesterday's issues. Indeed, an organization said that 24 hours is ideal for creating and disseminating tactical and operational reporting, but that is rarely achieved (four days is actually considered quick).

### BEST PRACTICES

**Create a variety of reports**

A practice of high-performing organizations is to have a varied cyber intelligence product line. These organizations have threat (operational and tactical) analysis and Strategic Analysis reports that address immediate needs, CISO and executive leadership requests, as well as specific internal business units and external customers and partner requests. SLAs and SOPs hold these organizations to their commitments and ensure that decision makers and other readers know what to expect. We met with organizations that produce reports daily, weekly, monthly, quarterly, and annually. Study participants reported producing a variety of reports/briefings, including

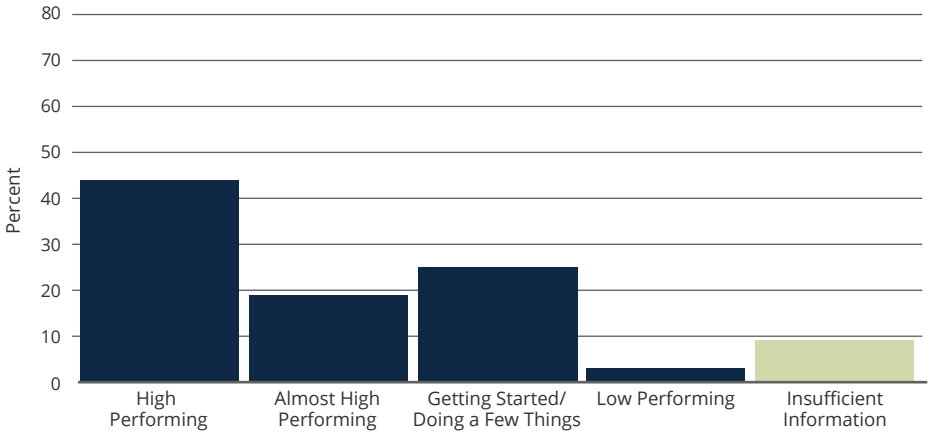
- vulnerability reports
- threat analysis reports
  - threat actors
  - threats to sectors
  - malware analysis
- threat priority lists
- bi-annual and annual threat assessment
- targeting packages for penetration testing team
- vulnerability reports
- technology program threat assessments
- geopolitical events
- industry developments
- patch status reports
- anti-virus reports
- threat news
- executive reports
- future threat analysis reports
- daily sector reports
- tactical reports: articles, indicators, and behavior summary
- incident responses reports
- after action reports
- briefings to CISO/CSO twice a week
- monthly executive council briefings
- bi-annual board briefings

## REPORTING AND FEEDBACK FACTOR 2: ACTIONABLE AND PREDICTIVE ANALYSIS

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization has a mechanism for reporting actionable and predictive analysis when necessary. Cyber intelligence reports include predictive analysis focusing on near- and long-term threats to the organization. Measures for evaluating prediction accuracy are in place.

## PERFORMANCE SNAPSHOT




**Reporting and Feedback Factor 2 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

### Lack of predictive analysis

Predictive analysis is performed and incorporated into longer-term strategic reports pertaining to threats, risks, and opportunities involving organizational vital interests. Many organizations interviewed do not include predictions in their cyber intelligence reports. These organizations generally have zero to little resources (people and time) to support a strategic analytical capability. These organizations were represented in the finance, health and public health, government, academic, and energy sectors. Some organizations also explained they probably could perform predictive analysis; however, they are not collecting the right data to support that type of analysis. Other organizations remarked that they do include predictive analysis in cyber intelligence reports, but it is done inconsistently. Many organizations we met stated that they have no measures in place to evaluate for prediction accuracy.

Some organizations interviewed do not include actionable recommendations in their cyber intelligence reports. Additionally, other organizations explained that actionable recommendations are designed to answer only tactical-level SIRs and are only for cybersecurity operations, mitigations, and cyber hygiene, which usually falls to the SOC. One organization explained that they do not put formal recommendations into strategic reports because the cyber landscape changes so fast and is too dynamic. The organization is concerned about recommending an action that has been overcome by events.

### IMPROVE YOUR PERFORMANCE

- Include actionable recommendations in Key Judgments or Bottom-Line-Up-Front sections of cyber intelligence reports.
- Incorporate analytical predictions into strategic reports pertaining to threats, risks, and opportunities involving organizational vital interests.

## BEST PRACTICES

### Include predictions and actionable recommendations in cyber intelligence reports

A practice of high-performing organizations is to incorporate predictions into strategic reports pertaining to threats, risks, and opportunities involving organizational vital interests. We met with cyber intelligence teams predicting when their own emerging patent-pending technology will become profitable, and how that aligns with the organization's own future business plans. We also interviewed cyber intelligence teams producing future country risk assessments, predicting what and how new technologies will impact the organization' business operations, and how new tools enable the organization to be proactive against threats. A practice of high-performing organizations is to also include predictions into more time-sensitive operationally focused reports about threat actor intentions, capabilities, operations, and campaigns. Some organizations produce reports that predict whether a specific threat actor will target the organization, or malware types that could cause the most damage to the organization. To assist with prediction analysis, another practice of high-performing organizations is for their cyber intelligence team to work closely with data scientists (data scientists are either part of the cyber intelligence team or co-located within the fusion center). These organizations "apply data science to actions on objectives" to determine a risk score associated with a given or proposed action. If time permits, structured analytical techniques such as *Alternative Futures Analysis*[38] or human-centered design techniques such as *What's on Your Radar* and *Creative Matrix* can assist prediction analysis.[39]

Many high-performing organizations include actionable recommendations in all their reports, even quarterly reports describing threat actor TTPs. Others said the benefits of including actionable recommendations depend on the situation and audience. For these organizations, actionable recommendations are mostly used to support cybersecurity and cyber hygiene needs based on data and subsequent analysis collected at the tactical/technical SIR level. These are more immediate–near term actions/mitigations/controls such as blocking IP addresses, implementing network IDS rules, patching vulnerabilities, or searching for specific hashes or strings. As mentioned earlier, a practice of high-performing organizations is to not write a report recommending a particular course of action at the SIR level that is immediate (e.g., block IP address). Rather these organizations take the necessary course of action first to protect the

**TIP**

A report without recommendations can still be useful. These reports add value with their insight and context about threat activity.

38 https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/Tradecraft%20Primer-apr09.pdf

39 Luma Institute. Human-Centered Design Thinking.

organization, and then later write a daily or weekly after-action report describing what actions were taken. Another practice of high-performing organizations is to have daily operations briefings or standups in the fusion center in front of the CSO/CISO. The briefings include proposed actionable recommendations, or actions that have already been taken to protect the organization over the course of the day. Finally, actionable recommendations should be included in a cyber intelligence report's Key Judgments section or in a Bottom Line Up Front.

## DESCRIBING ACTIONABLE RECOMMENDATIONS AT THE TACTICAL, OPERATIONAL, AND STRATEGIC LEVEL

Threat Analysis leads to actionable recommendations at an operational and tactical level in response to threats, threat actors, and campaigns. Actionable recommendation examples at the tactical level could be to patch particular vulnerabilities or disable a particular feature in an application. Actionable recommendations at the operational level follow from internal and external technical telemetry evaluation regarding a specific threat actor. Actionable recommendation examples at the operational level might be updating organizational-wide password rules, segmenting controls systems with a DMZ from the public-facing internet and business networks, incorporating a DLP, creating a honeypot, putting sensitive technology research on separate servers, or engaging the collection management team to task new collection on a specific threat actor.

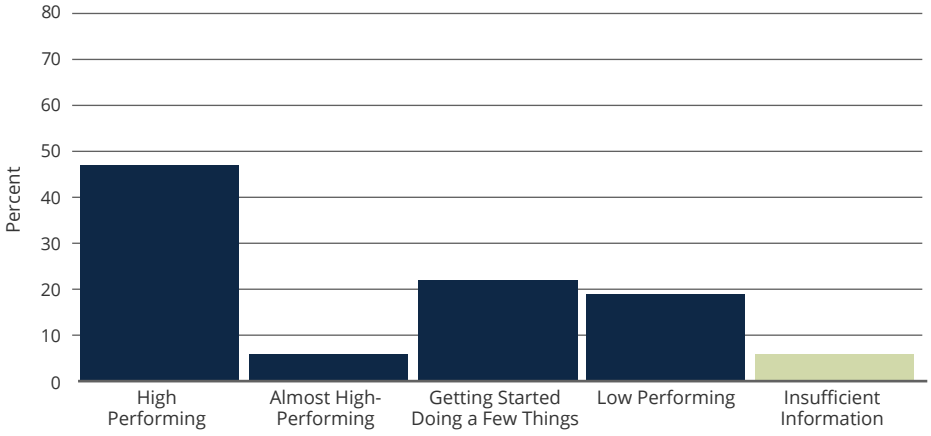
Strategic Analysis can be actionable, yet is based more on analytical judgments, enabling executive leaders to make risk-based decisions pertaining to organizational vital interests. Analysts may recommend strategic actions such as opening an office in one foreign location rather than another, merging with one organization rather than another, using a particular supplier, switching to new a software provider, or investing in new technology.

## REPORTING AND FEEDBACK FACTOR 3: LEADERSHIP INVOLVEMENT

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization's leadership influences the cyber intelligence effort by consistently providing items of interest, suggestions, praise, and format and production timeline requests for functional and strategic analytical products.

### PERFORMANCE SNAPSHOT




**Reporting and Feedback Factor 3 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**Reactionary involvement from leadership**

Some organizations mentioned that leaders get involved only when there is a crisis. Leaders at these organizations take a "no news is good news" approach to cyber intelligence; at best, they may request a briefing during a crisis, ask for follow-up information after an incident, or express appreciation that an incident has not happened. Cyber intelligence teams facing this challenge expressed a desire for leadership to be more active with setting strategy, specifically in risk management and setting PIRs.

#### IMPROVE YOUR PERFORMANCE

- Send strategic reports to your organization's board of directors.
- Create and perform a "road show" to showcase your team's capabilities.

### BEST PRACTICES

**Involve your organization's board of directors**

High-performing organizations frequently have a very involved board of directors that understands the importance, if not the details, of cyber intelligence and cybersecurity. For some organizations, the CISO or CTO sits on the board or has close contact with the board and can be an advocate for the cyber intelligence team. In other cases, the cyber intelligence team sends reports to the board. One team mentioned that they instituted this practice after a high-profile breach.

**Be your own advocate**

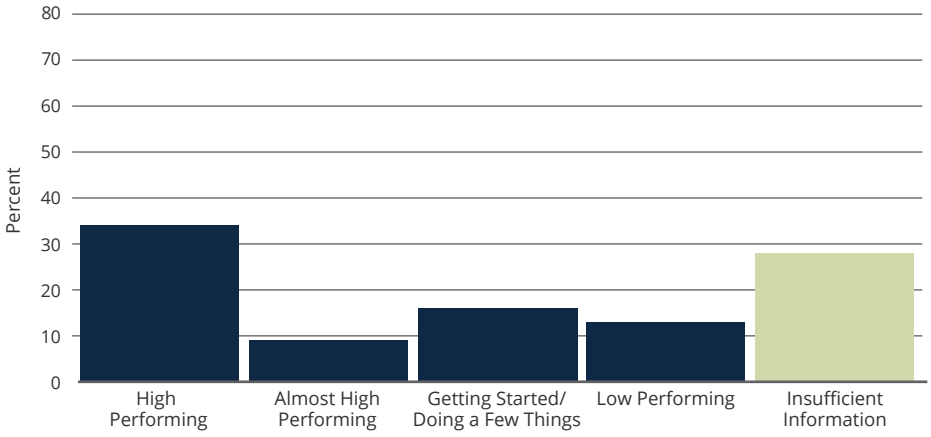
To build relationships with leaders, teams at high-performing organizations take proactive steps to showcase their work. One team developed a road show it performed for departments across the organization to familiarize those groups with their capabilities and successes. The team was initially discouraged and somewhat exhausted by what they described as a huge education process, but in the end, the payoff was worth it. That team has buy-in from senior leadership, who helps them get into hard-to-crack directorates.

## REPORTING AND FEEDBACK FACTOR 4: INFLUENCE ON DECISION MAKING

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization's leadership incorporates cyber intelligence reporting into its decision making on issues relating to its Environment, Data Gathering, Threat Analysis, Strategic Analysis, cybersecurity, and overall risk management and business decisions regarding organizational vital interests.

### PERFORMANCE SNAPSHOT




**Reporting and Feedback Factor 4 Performance Snapshot** The graph on the left shows how participants in the study are performing in this assessment factor.

### COMMON CHALLENGES

**Lack of leadership knowledge about cyber intelligence**

Some organizations explained that their leadership (specifically at the board level) does not understand cyber or the return on investment cyber intelligence brings to the organization. Teams at these organizations commented that they are constantly educating leadership about cyber. They do this to enable leadership to ask the right questions and know what to do with cyber intelligence when it is presented to them. Other teams explained that leadership only uses cyber intelligence when it involves technology purchase decisions.

### IMPROVE YOUR PERFORMANCE

- Senior leadership should champion the cyber intelligence team by referencing the team's reports in speeches and talks across the organization.
- Use the cyber intelligence team as a testing ground for new tools and technologies that could later be adopted and scaled across the entire organization.
- Keep metrics and feedback on leadership, partner, and customer usage and implementation of the cyber intelligence team's recommendations.

One cyber intelligence team supporting a large organization commented that their leadership does not seem to be doing anything on a strategic level with cyber intelligence reports the team produces.

**Lack of access to leadership**

As mentioned in earlier sections, some cyber intelligence teams lack consistent access to the CISO and board. For example, one team has briefed its CISO just three times in the last seven years, and another was briefing the board for the first time in 10 years. Additionally, some organizations within the finance, defense industrial base, and academic sectors explained that they lack information and/or have zero visibility into how leadership actually uses cyber intelligence to enhance decision making. There is no feedback.

## BEST PRACTICES

**Use cyber intelligence to enhance decision making**

Leadership using cyber intelligence to enhance decision making is a practice of high-performing organizations. Some teams we met explained how their leadership—CSO/CISO and up through C-suite executives and the board—is constantly refining how the organization conducts business based the cyber intelligence team's work. Indeed, one organization said that senior executives meet every day to discuss, among other things, cyber issues and the cyber intelligence team's analysis. We met with cyber intelligence teams that explained how their CEO champions the cyber intelligence team by referencing the team's reports in speeches and talks across the organization. Other teams said that leaders of different business units regularly receive cyber intelligence reports.

We also met organizations where the cyber intelligence team is considered such a trusted authority that they are constantly being pulled into internal organization-wide business unit leadership meetings. For instance, one organization is tapping its cyber intelligence team's expertise to help build the organization's insider threat program.

Cyber intelligence teams across the communications, commercial services, government facilities, and financial services sectors explained how leadership leverages their reporting and recommendations pertaining to tool and technology purchases that will better protect the organization. For example, cyber intelligence teams we interviewed have influenced leadership to purchase passive DNS scanning tools and bitcoin wallet analysis tools. Another practice of high-performing organizations is to use the cyber intelligence team as a testing ground for new tools and technologies that could later be adopted and scaled across the entire organization.

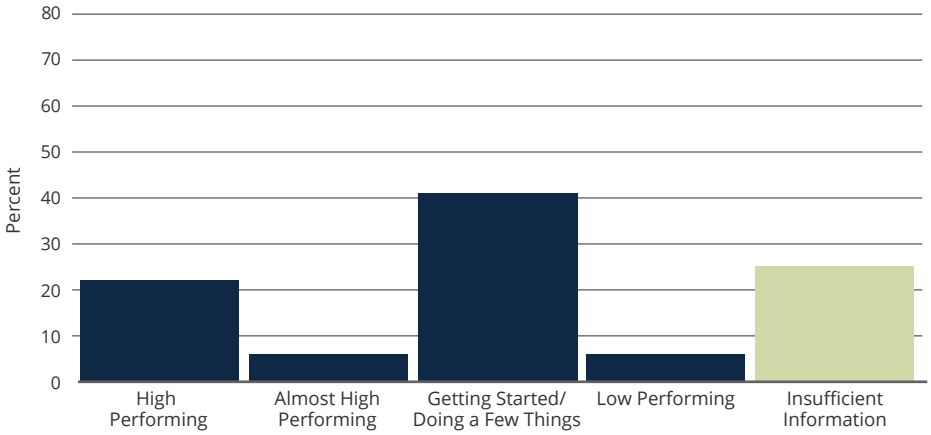
Organizations provided other examples of how cyber intelligence is influencing their leadership's decision making: helping executives, the board, and lawyers understand who/what is and will be the biggest threats to the organization; leadership requiring the organization to review and enhance existing controls; opening offices in foreign locations; re-prioritizing resources and budgets; increasing support to new or existing projects; providing recommendations on vendor purchase options; and acquisition support. Finally, a practice of high-performing organizations is to track and keep metrics and feedback on leadership, partner, and customer usage and implementation of the cyber intelligence team's recommendations.

## REPORTING AND FEEDBACK FACTOR 5: FEEDBACK MECHANISMS FOR THE CYBER INTELLIGENCE TEAM

### WHAT THIS ASSESSMENT FACTOR MEANS

Formal and informal mechanisms are in place for customers, collaborators, and stakeholders to provide feedback to the cyber intelligence team.

### PERFORMANCE SNAPSHOT




**Reporting and Feedback Factor 5 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

### COMMON CHALLENGES

**Lack of feedback mechanisms**

Several organizations we interviewed have no formal mechanisms in place for analysts to receive feedback from leadership, customers, collaborators, and stakeholders before and after a report is published. Most cyber intelligence teams interviewed receive feedback on their intelligence reports informally before and after publication. Informal mechanisms include email, peer-to-peer reviews, conversations, and leadership reviews. Formal mechanisms may range from websites, portals, wikis, surveys, and annual or biannual performance reviews. These cyber intelligence teams explained that they are sometimes unclear if they are meeting leadership, customer, collaborator, and stakeholder expectations. For other organizations, specifically in the finance and energy sectors, email is the primary and often only mechanism for analysts to receive feedback. Still other organizations said that external customers and stakeholders as well as internal business units do not regularly provide feedback on their cyber intelligence reports. This may be due, however, to the fact that the published intelligence reports lack a comment or feedback section. Lastly, some cyber intelligence teams commented that peer-review and coordination processes are too extensive, preventing and holding up timely report publication and dissemination.

#### IMPROVE YOUR PERFORMANCE

- Use a combination of portals, wikis, surveys, email, peer-to-peer conversations, annual reviews, and engagement teams for the cyber intelligence team to receive feedback.
- Append surveys or feedback links to finished cyber intelligence reports.
- Create a system, policy, and culture where rapid feedback to draft reports is the norm so originating analysts can quickly course correct and make necessary adjustments.

## BEST PRACTICES

### Create multiple ways analysts receive feedback before and after report publication

A practice of high-performing organizations is using multiple informal and formal mechanisms to receive feedback. Feedback may be in the format of questions or comments about reports, new requirements, ideas for new sources, and suggestions for analytical and workflow improvements. We met cyber intelligence teams using a combination of portals, wikis, surveys, email, peer-to-peer conversations, annual reviews, and engagement teams to interact with/receive feedback from organizational leadership and other internal and external customers. Having a distinct internal and external relationship engagement team (as noted in Environmental Context Factor 5) that is co-located with the cyber intelligence team as part of an organization's fusion center is a best practice. More specifically, it enables cyber intelligence teams to be readily available for contact by leadership, internal and external customers, collaborators, and stakeholders at any time. In addition to the ongoing daily engagement with internal and external customers, one organization's cyber intelligence team holds bi-annual meetings/conferences with *all* customers and stakeholders together about cyber issues, where they solicit feedback on their performance.

Another practice of high-performing organizations is to append surveys to finished cyber intelligence reports. Surveys inform the cyber intelligence team about what's working, what's not working, and internal and external customers' interest in reports. Other organizations have created a feedback link in every published report. A method that one high-performing organization has adopted is the creation of a pop-up window on the cyber intelligence team's website where readers can enter feedback or ask questions. As noted in Data Gathering Factor 4: Technology for Data Gathering, Microsoft's Yammer tool is used as both an organizational social networking tool and as an incident tracker. Yammer enables the cyber intelligence team to receive feedback from across the organization in a real-time social network-type environment. Employees (to include C-suite executives for instance) have the option to like, share, reply to, praise, and update posts and to create polls.

### Commit to peer reviews

High-performing organizations have rigorous, yet rapid, peer review processes to ensure the timely publication of reports. One organization explained that they have instituted a cultural practice of providing rapid feedback to draft reports so originating analysts can quickly make necessary adjustments. Another organization requires analysts to have two peers—one from inside the cyber intelligence team and

**TIP**

**IMPROVING PEER REVIEW EFFICIENCY**

One suggestion to improve peer review efficiency is a policy where reviewers are allotted a given amount of time to review/edit a draft report before being automatically skipped in the process. Mandatory reviewers are established and cannot be skipped. For example, cyber issues requiring less than 24 hours for a report should naturally and generally have a short list of reviewers. Individuals are automatically alerted about the report, and are only allotted 1 hour (for example) to provide feedback on the report. Feedback options could range from approve, disapprove with suggestions, approve with corrections, etc. Longer review time-frames are report-type dependent. The entire process should be visible and auditable across the fusion center so everyone knows who contributed and provided feedback, and who was automatically skipped. In the future, it is foreseeable that such a system could learn and provide suggestions as to which individuals across an organization should review a draft report based on time sensitivities, people's availability, and team expertise.

one from outside the team—review all reports. These organizations also ensure draft reports are reviewed by supervisors and direct managers before publication and dissemination.

**Don't wait to publish the report**

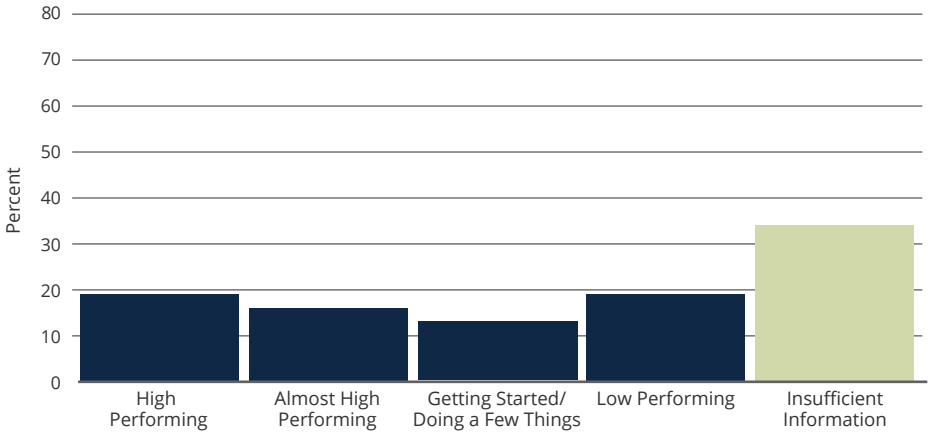
High-performing teams live by the axiom "don't let the perfect be the enemy of the good." Waiting to publish a report or disclose until you have the complete picture tends to diminish operational relevance. In other words, the law of diminished returns comes into effect the longer you wait. It is certainly better to publish and openly note any intelligence gaps or areas where you lack confidence than to wait for the full picture.

## REPORTING AND FEEDBACK FACTOR 6: INFLUENCE OF FEEDBACK ON DATA GATHERING AND ANALYSIS

### WHAT THIS ASSESSMENT FACTOR MEANS

Formal and informal processes ensure that data gathering and analysis efforts are influenced by feedback received from customers, collaborators, and stakeholders.

### PERFORMANCE SNAPSHOT




**Reporting and Feedback Factor 6 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

### Cyber intelligence teams receive little feedback regarding their analysis and data gathering

If feedback mechanisms are not in place for analysts to receive feedback (Reporting and Feedback Assessment Factor 5), then there is no way for feedback to influence the cyber intelligence team's data gathering and analysis efforts. For this particular assessment factor, several teams explained they receive zero to very little feedback from leadership, customers, collaborators, and stakeholders that influence the team's data gathering and analysis efforts. When cyber intelligence teams do not receive feedback (either in the evaluation and feedback step in the traditional intelligence cycle or via continuous feedback implied/encouraged within all components of Cyber Intelligence Framework), the cyber intelligence team's performance suffers. And the organization's ability to better protect itself may also suffer. More specifically, when teams are not receiving new or updated intelligence requirements, the data they are collecting and subsequently performing analysis on may no longer be relevant. New threats and risks emerge every day that could be missed.

That said, some organizations we interviewed explained that feedback from leadership, customers, collaborators, and stakeholders can influence the creation of new requirements, specifically SIRs, at the more technical/tactical level. One cyber intelligence team discussed how leadership and other stakeholders can influence data collection and analysis, but not necessarily the team's workflow.

**IMPROVE YOUR PERFORMANCE**

- Strategically formalize cyber intelligence into the organization's overall business decision calculus from a systems perspective (people, process, and technology).
- Frame cyber intelligence ROI in financial terms.

**TIP**

This Reporting and Feedback Assessment Factor 6 is closely related, but is not the same as Reporting and Feedback Assessment Factor Five, *Feedback Mechanisms for Analysts*. The distinction between these two assessment factors is that Reporting and Feedback Assessment Factor 5 assessed if organizations have mechanisms in place for analysts to receive feedback. Reporting and Feedback Assessment Factor 6 is more concerned with whether feedback analysts receive *actually influences data gathering and analysis efforts.*

## BEST PRACTICES

### Take action based on feedback

Your cyber intelligence team's performance depends on feedback from leadership, customers, collaborators and stakeholders. Many organizations explained that feedback from leadership, customers, collaborators and stakeholders influences the cyber intelligence team's data gathering and analysis efforts.

Organizations discussed how leadership, internal business unit, and external customer feedback enabled the cyber intelligence team to identify new intelligence requirements and subsequent intelligence gaps. New requirements lead to changes in internal data collection strategies, the passing of new requirements to third-party intelligence providers, and subsequent analysis of that data. Because one organization received so many requirements, the organization created a new position for an analyst to be the central point for all requirements—a starting point for a collection management team. We also met with organizations that described how feedback from

leadership enhanced the cyber intelligence team's strategy and workflow. For instance, one high-performing organization explained that leadership's feedback led to an extensive review and update of how all tactical alerts were created and disseminated. Other organizations described how feedback identified gaps that could be filled with better data gathering and analysis tools, leading to new budget requests. Some cyber intelligence teams also said that feedback influenced not just content, but the format and manner in which the intelligence reports are presented to leadership and customers.

**Build trust by being transparent**

Being transparent is a practice of high-performing organizations. In other words, publishing a cyber intelligence report on an important cyber issue that also clearly explains areas where you lack information, have intelligence gaps, or are less confident in judgments is a best practice. Again, 100% solutions are less relevant when 70% solutions are possible. Don't wait to disclose or release your report. Release it and continue to acquire the information you need. Being transparent creates trust with leadership, customers, collaborators, and stakeholders. Trust is the bedrock for receiving meaningful feedback that can influence data gathering, analysis, and overall strategy. Being transparent builds stronger relationships and understanding. With better understanding, cyber intelligence analysts can start predicting questions, and answer them before they are even asked.

**TIP**

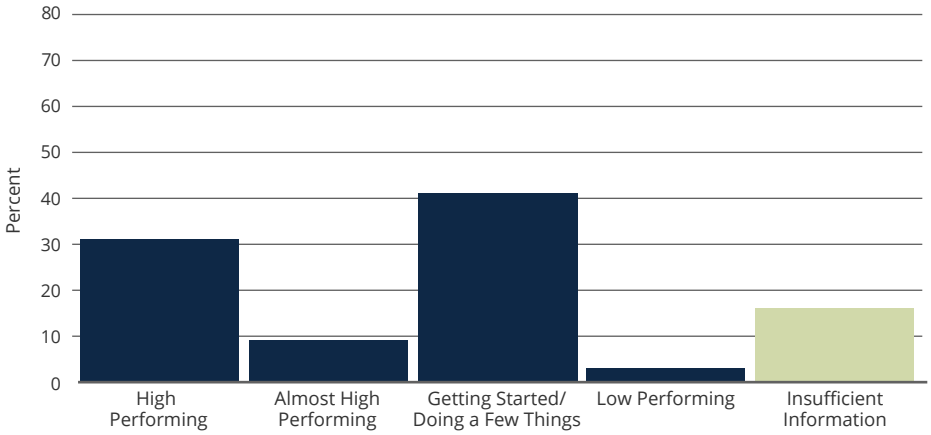
On reports where you have intelligence gaps or lack confidence, say so. Being up front builds trust and understanding.

## REPORTING AND FEEDBACK FACTOR 7: SATISFYING INTELLIGENCE CONSUMERS

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization has formal and informal processes to consistently ascertain whether consumers are satisfied with the cyber intelligence team's performance, specifically the quality, quantity, and timeliness of cyber intelligence reports.

## PERFORMANCE SNAPSHOT




**Reporting and Feedback Factor 7 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

### Cyber intelligence teams struggle to know if they are satisfying consumers

As with cyber intelligence analysts needing feedback to improve data gathering and analysis, organizations should build mechanisms to know if their consumers are satisfied the cyber intelligence team's performance. Consumers can consist of internal and external leadership, collaborators, customers, and stakeholders. Cyber intelligence teams we met struggled with knowing if consumers were satisfied with their performance and the quality, quantity, and timeliness of their products. Several teams we interviewed reported that they are unable to consistently know if their consumers (internal and external) are satisfied with their cyber intelligence reports. These organizations explained that while consumers may occasionally provide feedback, they did not have a formalized and repeatable process established, or tools such as a website, survey, portal, or wiki to ascertain consumer feedback. Indeed, one team talked about how they are constantly trying to figure out how consumers will interpret reports they produced.

Other cyber intelligence teams explained that their organization had yet to create a formal method to track and document that feedback. Some cyber intelligence teams keep metrics on the number of reports produced, yet do not track if/how the reports produced meet consumer requirements. Lastly, it was mentioned again to the SEI team that consumers may not know enough about cyber to know if they are satisfied or not.

**IMPROVE YOUR PERFORMANCE**

Interact with intelligence consumers: build an engagement team, hold brown-bags, attend internal business unit meetings, track consumer satisfaction using tools, and host periodic "cyber intelligence days."

**TIP**

There is a distinction between Reporting and Feedback Assessment Factor 7, *Satisfying Intelligence Consumers* and Reporting and Feedback Assessment Factor 5, *Feedback Mechanisms for Analysts.* Reporting and Feedback Assessment Factor 5 assessed if organizations have mechanisms in place for cyber intelligence analysts to receive feedback. Reporting and Feedback Assessment Factor 7 is more concerned with organizations knowing if their intelligence consumers (internal and external) are satisfied with the cyber intelligence team and its products. There is some overlap, however, between the two assessment factors. For instance, mechanisms to know if consumers are satisfied (surveys, wikis, portals, websites, meetings) may be the same, overlap, or are entirely different than mechanisms created for cyber intelligence analysts to receive feedback on their reports and performance.

### BEST PRACTICES

**Create multiple avenues to ascertain consumer satisfaction**

Creating avenues for your cyber intelligence team to know if consumers are satisfied is a practice of high-performing organizations. They do this because consumer feedback leads to changes (people, process, and technology) that enable your cyber intelligence team to perform at a higher level and meet/exceed consumer demands.

Most high-performing organizations adopt multiple methods to determine consumer satisfaction pertaining to their cyber intelligence team. First, and as noted earlier, a practice of high-performing organizations is to have an internal and external engagement team to make certain consumers are satisfied. In addition to ensuring IRs are met, the engagement team prioritizes consumers and report publication and distribution cycles for the team. For example, executive leadership is likely the highest priority consumer, perhaps followed by specific internal business units or key partners and subsidiaries.

Most organizations we interviewed, though, did not have an engagement team. Some of these organizations shared how giving feedback was a core value embedded in their organization's culture. Some cyber intelligence teams, for instance, discussed how they have daily standups with the CSO/CISO and receive direct and immediate feedback. Other teams explained that their manager briefs the C-suite and board frequently (several times a week) and returns with feedback. Additional methods to determine internal/external consumer satisfaction include holding brown bags, attending other internal business unit meetings, portals, surveys, websites, surveys, blogs, and holding annual or bi-annual "cyber intelligence days" where the team showcases its work and provides opportunities for feedback.

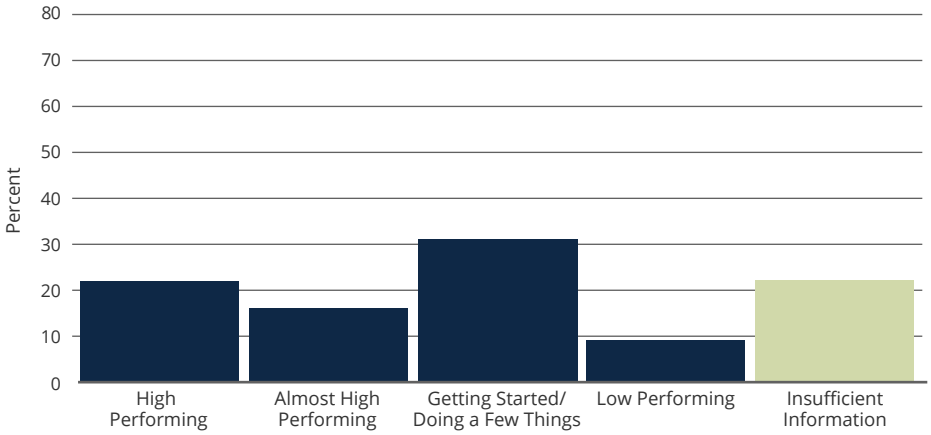
Another practice of high-performing organizations is building metrics to assess/show consumer satisfaction. These metrics are utilized to justify the cyber intelligence team's return on investment for the organization (see Reporting and Feedback Factor 8 for more information about demonstrating ROI). Example metrics organizations include are report production numbers, the number of reports addressing or tagged to executive leadership intelligence requirements, priority intelligence requirements, tools showing how often reports were opened and by whom, and internal and external service level agreements renewed or newly established.

## REPORTING AND FEEDBACK FACTOR 8: CAPTURING RETURN ON INVESTMENT

### WHAT THIS ASSESSMENT FACTOR MEANS

The organization captures return on investment (ROI) for its cyber intelligence efforts. High-performing organizations have a clear definition of what ROI means to them. The organization regularly tracks, monitors, and reports ROI to leadership for its cyber intelligence efforts, tools, personnel, and data feeds. The organization uses ROI information in a strategic fashion to manage current and future cyber intelligence investments.

## PERFORMANCE SNAPSHOT




**Reporting and Feedback Factor 8 Performance Snapshot** The graph on the left shows how study participants are performing in this factor.

## COMMON CHALLENGES

**Cyber intelligence teams struggle to show why they matter**

Cyber is ubiquitous. Yet a common challenge expressed to the SEI team was that not everyone, especially executive-level leadership, is comfortable with cyber. Some cyber intelligence teams discussed difficulties demonstrating their cyber intelligence efforts because educating leadership is a constant endeavor. Other organizations expressed concern that leadership "doesn't care" about cyber, yet wonders why security is so expensive; or that leadership only cares about "celebrity vulnerabilities." Some cyber intelligence teams struggle with demonstrating ROI because their organization has no clear definition about what ROI means. Teams explained how they have no metrics or ways to track ROI. Other organizations highlighted that their challenge was more of an issue of access to leadership. These teams have metrics, yet find it difficult to get in front of leadership. A few teams explained that their leadership doesn't even ask for metrics—at least not on a routine basis. Still, some cyber intelligence teams were of the belief that demonstrating ROI will always be a challenge, similar to how it is for Intelligence Community as a whole. In other words, you don't typically hear about Intelligence Community successes, usually only mistakes or incidents. One organization explained that as long as they don't make the news, they are demonstrating ROI. Last, some teams expressed that leadership views cyber intelligence and more specifically cybersecurity only through the lens of cost avoidance, rather than as an asset that can be both cost avoidance and a return on investment.

## BEST PRACTICES

**Track and showcase metrics for cost avoidance and revenue generation**

High-performing organizations demonstrate ROI by protecting the organization and providing actionable insights to enhance leadership decision making about emerging threats, risks and opportunities pertaining to organizational vital interests. This is possible because the cyber intelligence team's input is strategically formalized into the organization's overall business decision calculus from a systems perspective (people, process, and technology). Additionally, high-performing organizations grasp the concept that cyber intelligence teams demonstrate value beyond just cost avoidance. There can be an income component to cyber intelligence. We list below metrics organizations track and provide to leadership, as well as ways to demonstrate cost avoidance and return on investment for cyber intelligence.

High-performing organizations track the following metrics on a daily, weekly, monthly, and annual basis:

- External reports from other sources confirming your own cyber intelligence team's analysis
- New and repeat internal and external consumers for cyber intelligence products and tools
- New cases/incidents initiated and successfully resolved
- Vulnerabilities identified and fixed
- Phishing pages taken down
- People accessing your website or portal
- Threats identified targeting the organization
- Reports types downloaded
- The number of times reports were downloaded from your website or portal
- Important business decisions and meetings where the cyber intelligence team provided advice and guidance
- Business decisions across the organization that leveraged cyber intelligence products

Teams at high-performing organizations show cost avoidance. For example,

- Develop deep internal and self-generating cyber intelligence expertise, as well as tools and systems. This enables your organization to not be so reliant on hiring outside consultants, typically a cash expense.
- Cyber intelligence influencing leadership to not open a facility in a foreign location saves costs.
- Cyber intelligence passed to cybersecurity teams (SOC, Incident Response, Vulnerability Team, Network Defense) leads to new mitigations and controls that protect the organization.
- Expenses saved after updating networks or patching "ABC" policy.
- Adopting a virtual fusion center or aspects of a virtual fusion center may save on location expenses.
- Showing organizational impact/costs of specific threats targeting industry partners—and if the threat targeted the organization itself
- Showing organizational impact/costs of specific threats targeting the organization itself were not stopped
- Streamlining manual tasks with automation and machine learning may reduce expenses.
- Creating targeting packages for the penetration testing team to use against organizational assets or proprietary technology. This may demonstrate how hard/easy it is to target a specific asset or technology.

- Renegotiating deals with vendors based on vendor performance
- Keeping current with security and compliance regulations

Demonstrating ROI tends to be more challenging than demonstrating cost avoidance. Specifically, ROI implies there is a monetary, specifically income, value attributed to the cyber intelligence team's performance. A practice of high-performing organizations is to first create a financially defined ROI definition that has clear measures and timeframes. A possible and hypothetical example might be:

*With an annual budget of X dollars, the cyber intelligence team over the next year will protect the organization's critical infrastructure and technologies valued at X dollars. The cyber intelligence will aim to generate X dollars in revenue this year. Revenue generation will be accomplished by establishing new internal and external partner agreements, and informing leadership about threats, risks, and opportunities pertaining to the organization's vital interests.*

**Examples of ways to demonstrate or achieve ROI:**

- You have built such an amazing high-performing cyber intelligence team, that as a result, your organization is very appealing to other companies looking to be acquired or merge with a better cyber intelligence performing organization
- Cyber intelligence advancing leadership decision making regarding strategic technology development and procurement
- Embedding BISOs or cyber intelligence analysts in internal business units (business development, physical security, marketing, technology procurement, legal, and HR) to provide tailored cyber intelligence to those units. This may not result in a true cash transaction, yet at a minimum would likely show as an internal business expense for that specific business unit
- Your cyber intelligence team becomes an industry leader in providing cyber intelligence. Other organizational peers are charged annual fees to receive your organization's cyber intelligence products, briefings, or partnership for joint simulations and other related expertise.

Lastly, the manner in which ROI and cost avoidance is communicated to executive leadership is critical. Cyber intelligence teams may track all the metrics they want. However, it either won't matter or will go unnoticed if the cyber intelligence team is unable to communicate metrics in business risk-based terms, ascribing monetary values to events, incidents, and opportunities such as those listed above.

# Conclusion

In this report, we defined cyber intelligence as acquiring, processing, analyzing, and disseminating information that identifies, tracks, and predicts threats, risks, and opportunities in the cyber domain to offer courses of action that enhance decision making. Performing cyber intelligence is about knowing which threat actors have the intent and capability to target your organization and industry; tracking malware campaigns that may disrupt your operations; understanding your supply chain vulnerabilities; and assessing potential mergers and acquisitions, geopolitics, and emerging technologies that may impact your organization.

In 2018, we interviewed 32 organizations representing a variety of sectors to understand their best practices and biggest challenges in cyber intelligence. This study includes a report of our findings as well as three implementation guides, which provide how-to-steps for leveraging machine learning, the Internet of Things, and cyber threat frameworks to support cyber intelligence.

We found a number of best practices, including the following:

**Understanding that cyber intelligence is not cybersecurity.**
Organizations should create a dedicated cyber intelligence team that follows a defined and repeatable cyber intelligence workflow based on these framework components: Environmental Context, Data Gathering, Threat Analysis, Strategic Analysis, and Reporting and Feedback. We learned that having a collaborative, diverse fusion center with strong leadership engagement is a best practice.

**Establishing a fusion center.**
Fusion centers help break down silos and enable quick information sharing and analysis. A mature fusion center may comprise the SOC, security engineering and asset security, cyber intelligence, program management, and technology and development teams.

**Building a collection management team.**
High-performing organizations have collection management teams to identify and track intelligence requirements and work with analysts to validate data and data sources.

**Using emerging technologies.**
We also saw high-performing organizations bring in machine learning engineers and data scientists, and incorporate SOAR technologies to automate manual tasks in the cyber intelligence workflow.

**Ensuring that the cyber intelligence team's analysis is incorporated into leadership decision making processes from tactical to strategic levels.**
Cyber intelligence reports and briefings should be produced on a variety of subjects and according to an agreed upon schedule. A committed and engaged leadership team should provide feedback to the cyber intelligence team and champion their efforts.

We also found a number of challenges:

**Lack of formal workflows.**
We interviewed organizations without formal workflows for producing cyber intelligence. Practices were conceptual and ad hoc.

**Difficulty accessing data.**
Another challenge was that organizations (big and small) expressed difficulty accessing relevant data across their organization, industry, and other sectors.

**Lack of resources.**
We met organizations seeking more people with diverse skills to perform different types of Threat and Strategic Analysis. Additionally, some organizations lack formal intelligence requirement and data validation processes and rely exclusively on third-party intelligence providers. We interviewed cyber intelligence teams using outdated tools and technology for data gathering and analysis.

**Lack of leadership buy-in.**
Last, a good number of cyber intelligence teams expressed the desire for their leadership to have more cyber education, and for leadership to support the team's efforts and provide feedback on its performance.

Looking ahead, we see the promise of emerging technologies. New technologies can provide us with ways to capture large amounts of data and make sense of it. Artificial intelligence using machine learning has the potential to relieve human analysts of the burden of manual tasks and free them to think critically. Human-machine teaming, the center of our Cyber Intelligence Framework, is a key to the future of cyber intelligence.

In conclusion, the state of the practice of cyber intelligence in the United States is strong, but there are many ways we can be stronger. We can work better together, both within our own teams and across organizations—and with the tools and technologies that are already improving the practice of cyber intelligence.

# Appendix: The Future of Cyber and Cyber Intelligence

During our interviews we asked participants questions about the future of cyber and cyber intelligence. The SEI team grouped participants' responses into themes. The most common groupings are shown below.

**Five years from now, what skills, knowledge, and experience do you think will be important to have for cyber intelligence analysts?**

A diversity of skills, knowledge and experience will be needed to become a high-performing cyber intelligence team. Most, if not all the skills, knowledge and experience listed below are already in need. Organizations we interviewed simply explained however, that they will need more of it.

**Technical skills, Knowledge and Experience**

- Computing
  - Networking fundamentals
- Programming and Coding: Python, C++, API programming, REST,
  - Databases: Mongo DB
- Artificial Intelligence, specifically Machine Learning
  - How to build models
  - Data Science
    - Big Data Analytics
  - Automation
    - Scripting
- Experience working on a cyber intelligence team
- Cloud Analysis and engineering
- Mobile
- Embedded Devices
- SOC skills
- Malware Analysis
- Staying Fresh on Tools

**Non-Technical Skills Knowledge and Experience**

- Knowledge about threat actors
- Cross-Domain Intelligence Analysis
  - Critical Thinking
  - Connecting Dots, Link Analysis
- Communication skills but have technical aptitude to learn
  - Integration and communication
  - Interpersonal Skills

    - Emotional Intelligence
- Privacy Analysis
- Criminal Psychology
- Organizational skills
- Research skills
- Social Media Exploitation and Open Source Intelligence Techniques

## What technologies will impact cyber intelligence performance in the next five years? Why and how?

We asked organizations what technologies they believe will be relevant and impact the future of cyber intelligence performance in the next five years. The most common/frequent technologies that were mentioned are listed below. Maybe not so ironically, some technologies listed are also viewed by organizations as the biggest future threats in the following question.

- Artificial Intelligence
    - Will Impact how we respond to attacks
    - Will change how organizations recruit new talent and allocate monetary investments
    - Machine Learning
        - Help analyze bigger data sets that will require more software development
        - Technology that automatically answers intelligence requirements
        - Making risk decisions about other types of telemetry aside from Hashes and IPs
- Automation
- Cloud
    - Presents new challenges and opportunities
    - Cloud becomes operations infrastructure
    - Machine Learning capabilities through the cloud will better alert you to threats
- Unified Digital Landscape
    - Everything Smart (IoT Devices, Phones, Vehicles, Buildings)
- Big Data and Big Data Analytics
    - Changing Data sets and collection sources
    - Ability to process big data, draw connections,
    - Anything that can house big data, manage it, run analytics on it
- Quantum Computing
- Encryption
- Brain–Computer Interfaces

## What are your biggest future threats?

- Technology, and its unintended consequences
  - Artificial Intelligence
    - Adversaries using Artificial Intelligence such as machine learning against us, so it will continue to be an arms race
    - Malware that learns
    - Generative Adversarial Networks
  - Cloud
    - How to secure it and get value out of it at same time
  - Botnets
    - Ransomware at scale
- Data
  - Threat of drowning in data
  - Loss of trust in data
    - Disinformation
      - What is true and not true will be an increasing challenge
- Targets
  - Failing to educate people
    - People are weakest link
  - Unified Digital Landscape
    - Everything Smart (IoT Devices, Phones, Vehicles, Buildings)
      - Not enough security built into IoT devices
    - Machine Learning Models
    - Cloud
      - Huge attack surfaces, largely controlled by small number of big companies
  - Industrial Control Systems
  - Mergers and acquisitions creating larger attack surfaces
  - Vertical pivoting from user networks to operational critical infrastructures and ICS
  - Third-party vendors
  - Supply-chain Threats
  - Social Media Targeting of employees
- Policy Stagnation
  - Laws and sharing of data
  - Intersection of technology and rules (Cyber and GDPR)
    - Laws too slow to keep up with pace of technology
  - Block-Chain decentralization, lack of regulation and monitoring
- Cyber Sovereignty and Internet Balkanization
  - Privacy
    - Leveraging GDPR for advantage
- Encryption
  - Quantum Computing
    - Some algorithms today are non- quantum safe.
  - Not have enough diversity and wider adoption of the same algorithms

    - TLS version 1.3 could make deep packet inspection challenging
    - Threat actors are moving more towards encrypted chats like WeChat, WhatsApp and Telegram to conduct business.
        - Some encrypted chats have their own block chain platform and cryptocurrency
- People
    - Staffing and Retention
        - Not enough people that understand security, intelligence, forensics, and technology
- Threat Actors
    - Understanding the threat actor supply chain
        - Not just one person behind a threat (programmer, buyers, seller)
    - Foreign Nation States/Cyber Criminal Organizations
        - China's cyber strategy
        - Nation State Hacking from North Korea, Iran, Russia and China
        - State Sponsored attacks: More state actors and criminal organizations working together
        - Diffusion/Proliferation of Nation-state capabilities to other nation-states and to individuals
        - Nation-State attacks more sophisticated, incorporating levels of deception, operational security awareness
    - Insiders

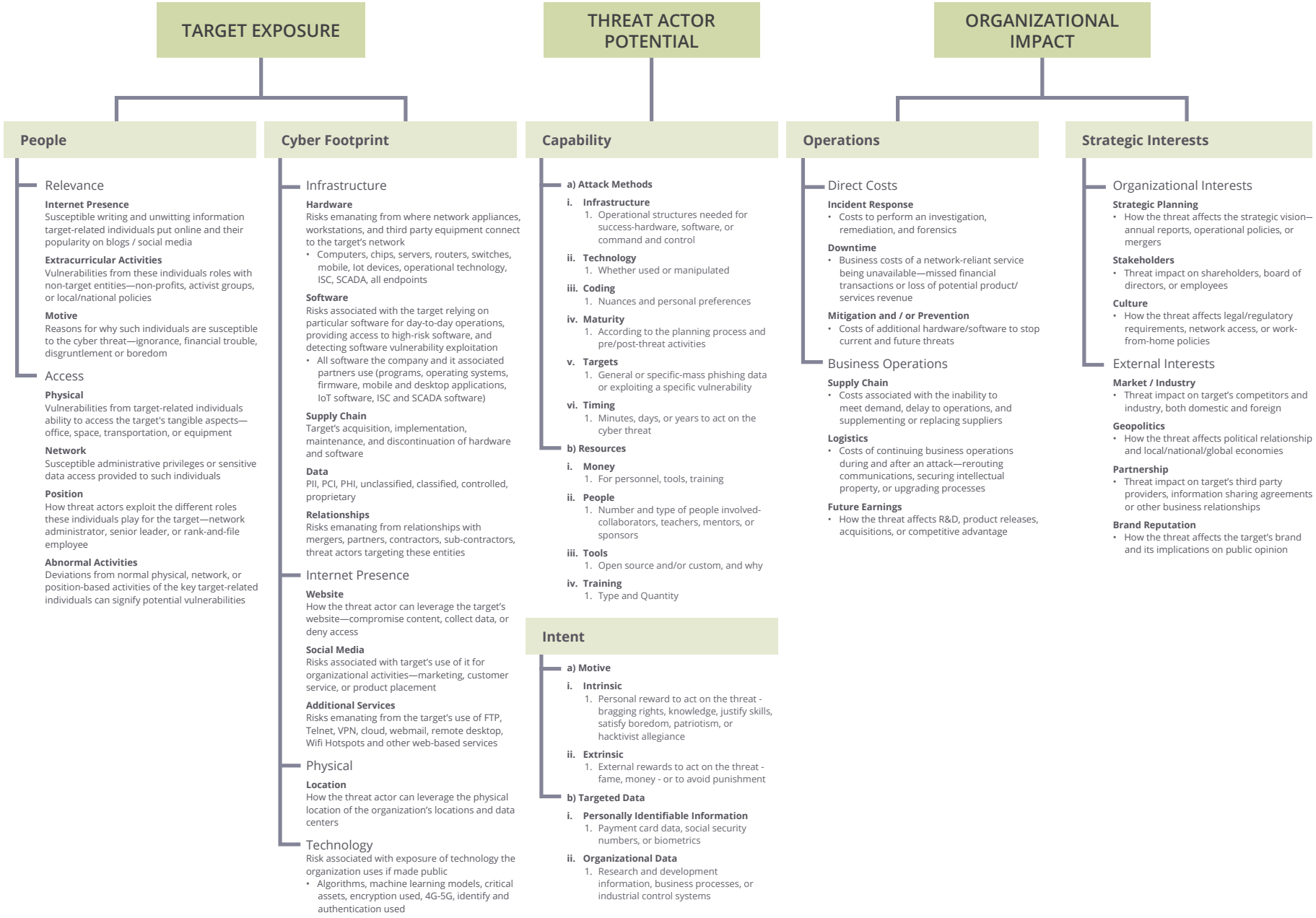
# Appendix: Most Popular Cyber Intelligence Resources[40]

Participants reported using a wide array of tools, sources, and services in their cyber intelligence practices. The following graph shows the most frequently reported resources among participants and their uses. The list includes a mix of free, open-source, and paid resources.

| Data Management | Analysis | Visualization | Third-Party Intelligence | Resource |
|---|---|---|---|---|
| | ■ | ■ | ■ | Anomali |
| | ■ | ■ | ■ | CrowdStrike |
| | | | ■ | DHS - CISCP |
| ■ | | | | Elastic Search |
| | | | ■ | FBI |
| | ■ | ■ | ■ | FireEye |
| | ■ | ■ | ■ | Flashpoint |
| ■ | | | | Hadoop |
| | ■ | ■ | | i2 Analyst's Notebook |
| | | | ■ | IBM X-Force Threat Intelligence |
| | ■ | | | IDA for Malware Analysis |
| | ■ | | | Joe Sandbox |
| | | ■ | | Kibana |
| | ■ | ■ | | Maltego |
| ■ | ■ | ■ | | Malware Information Sharing Platform (MISP) |
| | | | ■ | NCFTA |
| | | | ■ | Proofpoint ET Intelligence |
| | ■ | ■ | ■ | Recorded Future |
| | ■ | ■ | | Splunk |
| | ■ | | | VirusTotal |

40 The resources listed on this page were reported to the SEI by study participants. The SEI does not endorse or recommend any specific commercial product, process, or service.

# Appendix: Prioritizing Threats for Management (full view)



# Glossary

**Analytical Acumen**: Part of the Cyber Intelligence Framework's center of gravity; represents what a human analyst brings to cyber intelligence. Analytical Acumen is and art and science. As an art, no human analyst produces intelligence the same way, and the reason for that is that we all have our own personal instincts, biases, experiences, and a host of other nuances that impact the creativity and imagination that we bring to a cyber issue. As a science, an analyst seek outlets: technology, conceptual frameworks, analytical methodology, information collection methods, to best channel their creativity and imagination (the Art) into intelligence.

**Artificial Intelligence**: Systems that understand the world and independently make smart decisions based on that understanding.[41]

**Atomic Indicators**: "Pieces of data that are indicators of adversary activity on their own. Examples include IP addresses, email addresses, a static string in a Covert Command-and-control (C2) channel, or fully-qualified domain names (FQDN's)."[42]

**Behavioral Indicators**: "Those which combine other indicators—including other behaviors—to form a profile."[43]

**Business Information Security Officers (BISOs):** Used by high-performing organizations to embed in each organizational business unit to manage the relationship with the greater fusion center. BISOs act as both a liaison and officer for the fusion center by ensuring CISO polices are formulated into the business unit and enhancing intelligence sharing (intelligence requirements, cyber intelligence reports) with the fusion center. Global external BISOs may provide external country specific intelligence collection and analysis.

**Capability**: "Means to accomplish a mission, function or objective"[44]

**Computed Indicators**: "…those which are, well, computed. The most common amongst these indicators are hashes of malicious files, but can also include specific data in decoded custom C2 protocols, etc. Your more complicated IDS signatures may fall into this category."[45]

**Cyber Hygiene**: Cybersecurity efforts are sometimes called "cyber hygiene." "Cyber hygiene includes such activities as inventorying hardware and software assets; configuring firewalls and other commercial products; scanning for vulnerabilities; patching systems; and monitoring."[46]

---

41 https://ai.cs.cmu.edu/about

42 https://digital-forensics.sans.org/blog/2009/10/14/security-intelligence-attacking-the-kill-chain

43 https://digital-forensics.sans.org/blog/2009/10/14/security-intelligence-attacking-the-kill-chain

44 https://www.dhs.gov/sites/default/files/publications/18_0116_MGMT_DHS-Lexicon.pdf

45 https://digital-forensics.sans.org/blog/2009/10/14/security-intelligence-attacking-the-kill-chain

46 https://www.nist.gov/blogs/taking-measure/rethinking-cybersecurity-inside-out Ron Ross. November 15, 2016

**Cyber Intelligence**: Acquiring, processing, analyzing and disseminating information that identifies, tracks, and predicts threats, risks, and opportunities in the cyber domain to offer courses of action that enhance decision making.

**Cybersecurity:** Actions or measures taken to ensure a state of inviolability of the confidentiality, integrity, and availability of data and computer systems from hostile acts or influences.[47]

**Cyber Threat Intelligence:** Intelligence analysis on threats in the cyber domain. Cyber intelligence includes cyber threat intelligence, but cyber threat intelligence does not represent all of cyber intelligence.[48]

**Data Gathering**: Through automated and labor-intensive means, data and information is collected from multiple internal and external sources for analysts to analyze to answer organizational intelligence requirements.

**Data Loss Prevention (DLP) Tool/Software:** "Detects potential data breaches/data ex-filtration transmissions and prevents them by monitoring, detecting and blocking sensitive data while in-use (endpoint actions), in-motion (network traffic), and at-rest (data storage)."[49]

**Diamond Model of Intrusion Analysis**: "model establishing the basic atomic element of any intrusion activity, the event, composed of four core features: adversary, infrastructure, capability, and victim. These features are edge-connected representing their underlying relationships and arranged in the shape of a diamond, giving the model its name: the
Diamond Model."[50]

**Environmental Context**: Everything you need to know about your organization internally and externally. Includes understanding organization's entire attack surface; and threats, risks and opportunities targeting your organization and industry, and the impact of those threats, risks and opportunities to your organization and industry. Includes deeply knowing your internal and external network and operations, to include but not limited to: the organizations servers, operating systems, endpoints, data centers, organization's business, its mission and culture, organizational processes and policies, business partners, geopolitics, emerging technologies, and position in industry relative to competitors. Attaining Environmental Context is a continuous process and influences what data is needed to perform cyber intelligence.

**Human-Centered Design**: "Design and management framework that develops solutions to problems by involving the human perspective in all steps of the problem-solving process. Human involvement typically takes place in observing the problem within context, brainstorming, conceptualizing, developing, and implementing the solution."[51]

---

47 The definition for cybersecurity created based on analyzing participating organizational responses and from the DHS Lexicon Terms and Definitions Instruction Manual 262-12-001-01 (October 16, 2017) https://www.dhs.gov/sites/default/files/publications/18_0116_MGMT_DHS-Lexicon.pdf

48 A number of organizations expressed confusion over the difference between cyber threat intelligence and cyber intelligence, specifically whether these terms describe the same thing. Many organizations told us that introducing "threat" into this phrase breeds that confusion. Although threats are a large part of the cyber intelligence picture, cyber intelligence also includes analysis of areas like technologies, geopolitics, and opportunities. For these reasons, this report deliberately excludes the term "cyber threat intelligence." We refer to the activities typically associated with cyber threat intelligence as Threat Analysis, a component of the Cyber Intelligence Framework.

49 https://en.wikipedia.org/wiki/Data_loss_prevention_software

50 https://apps.dtic.mil/docs/citations/ADA586960

51 https://en.wikipedia.org/wiki/Human-centered_design

**Impact**: "Measure of effect or influence of an action, person, or thing on another—extended definition: may occur as either direct or indirect results of an action."[52]

**Intelligence**: "1. The product resulting from the collection, processing, integration, evaluation, analysis, and interpretation of available information concerning foreign nations, hostile or potentially hostile forces or elements, or areas of actual or potential operations. 2. The activities that result in the product. 3. The organizations engaged in such activities."[53]

**Intent**: "Determination to achieve an objective."[54]

**Likelihood: "**Chance of something happening, whether defined, measured or estimated objectively or subjectively, or in terms of general descriptors (such as rare, unlikely, likely, almost certain), frequencies, or probabilities."[55]

**Lockheed Martin Kill Chain**: "The Cyber Kill Chain framework is part of the Intelligence Driven Defense model for the identification and prevention of cyber intrusions activity. The model identifies what the adversaries must complete in order to achieve their objective."[56]

**Machine Learning**: A field at the intersection of Statistics & Computer Science. Fundamentally, it is about learning from data: summarizing patterns, making predictions, and identifying key characteristics of a group of interest, among many other tasks.

**MITRE Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK)**: "a globally-accessible knowledge base of adversary tactics and techniques based on real-world observations. The ATT&CK knowledge base is used as a foundation for the development of specific threat models and methodologies in the private sector, in government, and in the cybersecurity product and service community."[57]

**Office of the Director of National Intelligence, Cyber Threat Framework**: "Developed by the US Government to enable consistent characterization and categorization of cyber threat events, and to identify trends or changes in the activities of cyber adversaries. The Cyber Threat Framework is applicable to anyone who works cyber-related activities, its principle benefit being that it provides a common language for describing and communicating information about cyber threat activity. The framework and its associated lexicon provide a means for consistently describing cyber threat activity in a manner that enables efficient information sharing and cyber Threat Analysis, that is useful to both senior policy/decision makers and detail oriented cyber technicians alike."[58]

---

52 DHS Lexicon Terms and Definitions Instruction Manual 262-12-001-01 (October 16, 2017) https://www.dhs.gov/sites/default/files/publications/18_0116_MGMT_DHS-Lexicon.pdf

53 https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf

54 DHS Lexicon Terms and Definitions Instruction Manual 262-12-001-01 (October 16, 2017) https://www.dhs.gov/sites/default/files/publications/18_0116_MGMT_DHS-Lexicon.pdf

55 https://www.dhs.gov/sites/default/files/publications/18_0116_MGMT_DHS-Lexicon.pdf

56 https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/Gaining_the_Advantage_Cyber_Kill_Chain.pdf

57 https://attack.mitre.org

58 https://www.dni.gov/index.php/cyber-threat-framework

**Operational Analysis:** Analysis of specific threats, threat actors, their campaigns, intentions and capabilities against an organization and its industry. Operational Analysis answers Priority and specific intelligence requirements (PIR, SIR) to enhance CSO/CISO and other mid-to senior-level decision-makers' leadership decisions regarding non-immediate but near-term (weekly–quarterly) business process and cybersecurity decisions.

**Organizational Intelligence Priorities Framework (OIPF):** A framework for creating and managing organizational intelligence requirements (IRs, PIRs, and SIRS) , the data sources aligned to answer those intelligence requirements, and the validation of those data sources. The OIPF informs future planning, budgeting, programming, and allocation of resources to data collection and analysis.

**Reporting and Feedback**: Communication between analysts and decision makers, peers, and other intelligence consumers regarding their products and work performance. Reporting and feedback help identify intelligence requirements and intelligence gaps.

**Risk**: "Potential for an unwanted outcome as determined by its likelihood and the consequences… potential for an adverse outcome assessed as a function of hazard/threats, assets and their vulnerabilities, and consequences."[59]

**Security Orchestration, Automation and Response (SOAR): "**Technologies that enable organizations to collect security data and alerts from different sources."[60]

**Strategic Analysis**: Strategic Analysis is the process of conducting holistic analysis on threats AND opportunities. Holistically assessing threats is based on analysis of threat actor potential, organizational exposure and organizational impact of the threat. One might also perform Strategic Analysis to provide deep clarity on the who and why behind threats and threat actors. Strategic Analysis goes beyond Threat Analysis to incorporate analysis regarding emerging technologies and geopolitics that may impact/provide opportunities for the organization now and in the future. In this light, Strategic Analysis is not only comprehensive, but ANTICIPATORY. It can be actionable, yet is based more on analytical judgments, enabling executive leaders to make risk-based decisions pertaining to organizational wide financial health, brand, stature, and reputation.

**Structured Analytical Techniques**: analytic techniques designed to help individual analysts challenge their analytical arguments and mind-sets. Techniques are grouped by diagnostic, contrarian and imaginative thinking.[61]

---

59 DHS Lexicon Terms and Definitions Instruction Manual 262-12-001-01 (October 16, 2017)

60 https://www.gartner.com/en/documents/3860563

61 https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/Tradecraft%20Primer-apr09.pdf

**Tactical Analysis:** Analysis of specific threats, attacks, incidents, vulnerabilities, or unusual network activity that enhances decision making for network defenders, incident responders, and machines pertaining to cybersecurity and incident response. Information analyzed is usually technical telemetry such as network and endpoint activity, atomic, behavioral, and computed indicators[62] such as: malware samples, hash values, domains, IPs, logs, email header information. Tactical analysis tends to answer specific intelligence requirements (SIRs) and the immediate, daily and weekly what/where/when/how questions about threats.

**Threat**: "Indication of potential harm to life, information, operations, the environment and/or property—extended definition—may be a natural or human-created occurrence and includes capabilities, intentions, and attack methods of adversaries used to exploit circumstances or occurrences with the intent to cause harm."[63]

**Threat Analysis**: Assessing technical telemetry and non-technical data pertaining to specific threats to your organization and industry to inform cybersecurity operations/actions and Strategic Analysis. Threat Analysis is built on operational and tactical analysis and enhances CSO/CISO and other mid- to senior-level decision making.

62 https://digital-forensics.sans.org/blog/2009/10/14/security-intelligence-attacking-the-kill-chain

63 DHS Lexicon Terms and Definitions Instruction Manual 262-12-001-01 (October 16, 2017) https://www.dhs.gov/sites/default/files/publications/18_0116_MGMT_DHS-Lexicon.pdf

Carnegie Mellon University
Software Engineering Institute

# Artificial Intelligence and Cyber Intelligence

## An Implementation Guide

# Artificial Intelligence and Cyber Intelligence: An Implementation Guide

**Authors**
April Galyardt
Ritwik Gupta
Dan DeCapria
Eliezer Kanal
Jared Ettinger

**Design**
David Biber
Alexandrea Van Deusen
Todd Loizes



This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.



DM19-0447

# Table of Contents

# Machine Learning and Cyber Intelligence

## INTRODUCTION

Machine learning (ML) tools and techniques have been demonstrated to effectively improve cyber intelligence workflows across environment, data gathering, threat analysis, strategic analysis, and feedback to decision makers. Simply adding ML capability to existing organizational toolsets, procedures, and workflows will not solve all cyber intelligence challenges. These technologies work in concert with experienced and qualified personnel who know how to understand, integrate, and even improve ML processes in the context of cyber intelligence challenges. Only by combining modern tooling with personnel knowledgeable about its use and procedures can organizations begin to realize the significant benefits ML can provide.

### KEY JUDGMENTS

- Setting up an effective ML–cyber intelligence collaboration will require proper consideration, preparation, and communication.
- Introducing operationally effective ML into the cyber intelligence workflow requires a repeatable, consistent, and well-defined process.
- Prior to using ML, it is essential to walk through the ML checklist to answer relevant questions such as "Does ML help with this?" and "Have we considered the broader context?" Any doubts that arise when completing this checklist highlight gaps in analytical understanding that must be discussed with the cyber intelligence team.
- There are important ethical and data-use dilemmas associated with ML, especially when paired with the world of intelligence. Enumerate, weigh, and address these dilemmas to the fullest extent possible before proceeding with ML capabilities.
- In ML, the biggest performance improvements result from higher-quality data, not more sophisticated algorithms. Expect to spend the majority of your time and effort on data acquisition, curation, and processing.
- A large variety of people are needed to make an effective ML and cyber intelligence effort; namely, talented cyber intelligence analysts, ML scientists, and ML and/or data engineers. While expertise may initially be divided into silos in each member's domain, the team must work together to nurture domain expertise in a cross-functional manner and maintain open lines of communication.
- Be creative and have fun with the data sources you have. The reason to use ML is to tap the hidden knowledge potential within those data sources, so think critically about what new things can be extracted from the data that already exists and how to use it effectively. Don't forget to balance this creativity with operational and engineering considerations.

## DEFINITIONS

**ML** is a field at the intersection of statistics and computer science. Fundamentally, it is about learning from data: summarizing patterns, making predictions, and identifying key characteristics of a group of interest (among many other tasks). The term **artificial intelligence (AI)** has many definitions, but broadly speaking it refers to systems that understand the world and independently make smart decisions based on that understanding.[1] If an AI system can interact with and learn from interactions with the surrounding world, it must be learning from data. To that extent, ML is an integral part AI. Unfortunately, the language around AI and ML is further muddied by the fact that some ML algorithms, particularly neural networks, are often referred to as "AI" by the general public. In this guide, we focus on ML and the practice of learning from data.

Within ML, a **model** refers to a set of equations that could describe the data. When we **train** or **fit** a model, we search over a family of models to find the single model that best fits the data. This **trained model** is often referred to as simply **the model**. Within this context, an **algorithm** is a specific process for fitting the model to data. **Features** or **variables** refer to the different kinds of information recorded in the data; for example, if our data is a set of documents, one feature might be the author's name and another might be the number of words in the document.

The work of designing a model, fitting a model, and extracting information is generally performed by an **analyst**. However, within a cyber intelligence framework, we must disambiguate this work from the work of a cyber intelligence analyst.

We use **ML scientist** to refer to people who carry out the ML analysis and **data engineer** to refer to people who collect and prepare the data for analysis.

## CONDITIONS FOR SUCCESS IN ML

ML is a powerful tool, and it has spurred tremendous leaps in capability and productivity across many areas. However, just as hammers work well with nails but poorly with screws, ML is ideally suited to some tasks and poorly suited to others. This section is designed to help identify problems for which setting up an ML pipeline would justify the investment.

In popular conceptualizations, ML focuses on algorithmic capabilities; for instance, recommender systems in shopping carts ("You may like

**ARE YOU READY TO USE ML FOR CYBER INTELLIGENCE?**

1. Can you state your problem as either:
   a. I would like to use ___ data to predict ___, or
   b. I would like to understand the structure of the features recorded in ___ data?
2. Is it a large-scale problem?
3. Have you already done exploratory analysis?
4. Have you considered the broader context?

1 https://ai.cs.cmu.edu/about

this") or automated labeling of people, places, or objects in images. However, less-visible aspects of ML are just as critical to the process as algorithm choice: data storage, processing, and cleaning; the augmentation of data ("feature engineering"); reporting and visualization; and the development of software and hardware to support the entire ML pipeline
(among others). Many organizations already perform these tasks to support other business needs, and ML is often added to existing analysis pipelines to take advantage of existing tools that perform some of these duties.

To help determine whether your organization is ML ready, we've developed a checklist of necessary capabilities. If you aren't yet performing these data analytic practices, we recommend that you incorporate these items into your analytics pipeline and familiarize yourself with their output and value before adding ML capabilities.

## PROBLEM STRUCTURE

ML algorithms can be broadly divided into two categories: supervised learning algorithms and unsupervised learning algorithms. A supervised learning problem can be framed as "I would like to use some set of data to predict an unknown." For example, you might want to identify the actors responsible for constructing and deploying a particular piece of malware. Statistically, this is a prediction problem and can be reframed as "I would like to use hashes of malware files to predict who made the malware." As another example, cyber intelligence analysts often have more information coming in than they are able to process. The supervised learning problem can be framed as "We might want to use data about what information cyber intelligence analysts have previously found most useful to predict which new information is most important for the analysts to consume." These two examples will use vastly different data, but because they can both be framed as a problem of making a prediction using data, supervised learning algorithms can be applied to these problems.

Unsupervised learning problems can be framed as "I would like to understand the structure of the features recorded in a given set of data." The structure we're looking for could be very different depending on context. One common kind of structure we might look for comprises subgroups and clusters; for example, we might analyze resolved incident tickets collected over the past year by looking for clusters of related tickets. A second kind of structure comprises groups of closely related features. For example, if we are collecting data on insider threat indicators, we might want to examine which features are highly correlated with each other. If we identify 10 to 12 features that are all closely related and effectively measuring the same thing, then we may be able to reduce our data collection burden and only collect the 5 or 6 most useful. Note that in the section below, "Examples of ML for Cyber Intel," we refer to these two problem structures so you can see how they are applied in practice. Later, in the section "The ML Pipeline," we discuss the requirements and conditions under which you can apply both supervised and unsupervised learning.

## SCALE

ML will frequently reap the largest return on investment when it is applied to a task at a large scale. This scale could take many different forms. A common problem of scale is a task that needs to be executed repetitively. For example, satellites collect images faster than humans can label them. However, an ML algorithm can label the petabytes of images collected each day and flag anomalous images for human review. In other situations, an analysis might only be needed once, but the data available is of large scale—more than can be handled by a single person. For example, analysts might have a large amount of information on a particular set of network intrusions. In such a case, ML algorithms could find patterns in the data. This information then increases the capability and speed of the human cyber analyst.

The second way ML can address a problem of scale is to provide greater consistency across repetitions. For example, if we are looking at data and deciding whether or not to open an insider threat investigation, two humans might reasonably disagree. Supplementing the human analyst with information from an ML algorithm can foster greater consistency in decision making.

## EXPLORATORY ANALYSIS

A good rule of thumb is to always run simple analyses first. This includes basic information about data, such as how much data is missing, lists of the most frequently observed data for different datatypes, and data visualizations, such as histograms and time series charts. Statisticians frequently refer to this as exploratory data analysis. You should be able to answer basic questions about your data, such as "How many?", "How often?", and "What kind?" before attempting to apply ML techniques.

There are two reasons for addressing these questions. First, you can gain tremendous insights from your answers. How many of your incident tickets contain the words "technical debt?" How many contain the word "malware?" Simply identifying the "top 10" lists for different types of data frequently uncovers significant trends you may not be aware of. This type of analysis is very straightforward to perform using data analysis tools, and taking advantage of this low-hanging fruit can be a very cost-effective way to make use of existing data.

The second reason for addressing these questions is that the data cannot be put into an ML algorithm until it has already been sufficiently processed. The ability to answer these basic questions indicates that the data is processed enough for use in an ML algorithm. Furthermore, often what you find when conducting the simple exploratory analysis provides insight that will help shape the ML analysis. For example, you might discover that one sensor is, essentially, replicating the information from another sensor. Therefore, only one of those sensors should be used by the ML algorithm. Another common discovery is that, from date X to date Y, a sensor was misfiring and therefore should be omitted for those dates. When you try to apply an ML algorithm without first acquiring this basic understanding of the data, errors will happen.

### BROADER CONTEXT

Every ML analysis takes place within a broader context that includes ethical and legal ramifications. These issues will vary with context, but there are two that every ML analysis will share in common and which should be addressed when deciding whether to implement an ML algorithm: 1) What are the consequences of a data breach? and 2) What are the consequences of an incorrect decision based on the ML algorithm?

The large amounts of data required to make ML efficient also make data breaches more problematic. The Pennsylvania Supreme Court recently ruled that businesses can be held legally responsible for not taking adequate safeguards with sensitive information.[2] We also know that many algorithms, particularly neural networks, "leak" information: if someone has access to any of the decisions or output of an algorithm, they can make strong inferences about the information that was used to train the algorithm. The consequences of such a breach vary from case to case, ranging from an increased risk of identity theft for consumers to national security issues.

ML models are probability based. There is always a level of irreducible error in the output of an ML algorithm. For instance, if we are looking at predicting insider threat, there will always be cases in which the algorithm indicates someone is a threat (but they are not) and cases in which someone is a threat (but the algorithm misses it). This is true in every single application. As you implement an ML model, you must develop the procedures and responses to situate the output within your organization. Is this a case in which the response to the ML output can be automated? Or, do you have a case in which the response must be escalated to a human decision maker?

## USE CASES FOR ML IN CYBER INTELLIGENCE

There are many different types of ML, each best suited to solving a particular set of challenges. To provide a better understanding of how these tools can augment your cyber intelligence analysis, the following section describes a number of use cases demonstrating these capabilities in a variety of common scenarios.

### ENVIRONMENTAL CONTEXT: INSIDER THREAT ANALYSIS

One increasingly common application of ML is to predict which individuals within an organization might represent insider threats. This use case is usually a supervised learning problem: Collect as much relevant behavioral computer activity as possible on users (web browsing, network share access, logon/logoff logs) and use this data to predict the extent to which an individual is, or has the potential to be, an insider threat. This problem could also be framed as an unsupervised learning problem of anomaly detection: Most employees will exhibit relatively consistent usage patterns (e.g., logging on at a consistent time of day). A starkly anomalous usage pattern may be a red flag. The statistical problem is then to identify the anomalies in the data.

2 https://www.jdsupra.com/legalnews/pennsylvania-supreme-court-recognizes-34420/

There are two points to highlight in this use case. First, this application of ML would require proactive coordination between the insider threat team, the cyber intelligence team, and the ML team. Second, models and data collection should be updated regularly: Attack surfaces change constantly, and gaps in coverage are continually relevant.

## DATA GATHERING: IDENTIFYING REDUNDANT INFORMATION

When dealing with large, disparate datasets, there is frequently significant redundancy in the available data. Since this redundancy can substantially increase analyst workload, especially as the scale of the data increases, ML can help identify which information is redundant to save time and storage requirements. Unsupervised learning methods (including clustering and the feature reduction algorithms discussed below), combined with simple comparison metrics, can be used to group the data and flag redundancy. On large datasets, this can result in a significant reduction in data the analyst needs to examine and in greatly reduced data storage needs.

## THREAT ANALYSIS

### Malware Attribution

Given a set of executable files that are known to be malware, we may be interested in identifying the sources of the malware to identify a threat actor. If we have access to a labeled dataset in which different pieces of previously collected malware have been tagged with their source, we can use that data to build a supervised learning model to predict which of our new malware files has come from each source.

### Sorting and Prioritizing Information for Cyber Intelligence Analysts

Not all information has equal importance or priority when running through a cyber intelligence pipeline. Under normal circumstances, it is only after an intelligence analyst has reviewed the information that it can be given a priority rating; however, ML methods may be able to predict these priority ratings, and the historical relevance of similar data, from the task at hand. Assuming that we have access to past data that has already been given priority ratings, we can use supervised learning methods to sort incoming data by priority using a trained model.

## STRATEGIC ANALYSIS: IDENTIFYING COMMONALITIES IN ATTACKS

Strategic analysis is the work of conducting holistic analysis on both threats and opportunities. Holistic assessment of threats is based on analysis of threat actor potential, organizational exposure, and the impact the threat has on an organization. One might also perform strategic analysis to provide deep clarity on the who and why behind threats and threat actors. Strategic analysis goes beyond threat analysis to incorporate analysis of emerging technologies and geopolitics that may impact and/or provide opportunities for the organization now and in the future.

When reviewing data for strategic analysis, analysts often search for associations between actors, events, or activities. We can rephrase this as an unsupervised learning question: “Given a dataset consisting of resolved incident tickets from the past year, can we find clusters of tickets that relate to similar threats or threat actors?” Identifying commonalities would ease the discovery of a modus operandi or positive threat actor identification.

Clustering analysis could be augmented using active learning techniques, which help analysts identify where more data is required to make proper decisions. For example, consider an analyst attempting to identify the threat actor or actors behind a series of seemingly unrelated, discrete threats. In the course of her analysis, she classifies different threats into discrete buckets. This allows her to apply a clustering technique, which can automatically label new data as it comes in by comparing it to existing buckets. Furthermore, by applying active learning to her data, she can understand where her own model is strongest and weakest. When new data comes in and is automatically labeled, she can quickly know the extent to which the new label requires further manual analysis.

### REPORTING: VISUALIZATIONS AND AUTOMATIC REPORT GENERATION

Pairing ML with other automation provides an additional advantage. When an ML analysis is repeated regularly, the reports and graphs based on that analysis can be automatically updated and sent to cyber intelligence analysts or leadership for human review. There are several commercially available tools that streamline automatic report generation even further.

## THE ML PIPELINE

The fuel powering the entire ML process is data. This data must be processed, cleaned, and prepared before being put through the algorithm. The output of the algorithm is some useful result that the analyst can use. However, just as a car isn't very useful without seats, the ML process requires significant external tooling to make the whole thing useful.

In this section, we outline the main steps for performing analysis:

5. requirements definition
6. data input and processing
7. model design, development, and execution
8. reporting and feedback

The overall model is visible in Figure 1. Note that we do not show intelligence requirements (IR) in the figure since they are too high-level for the purposes of ML.

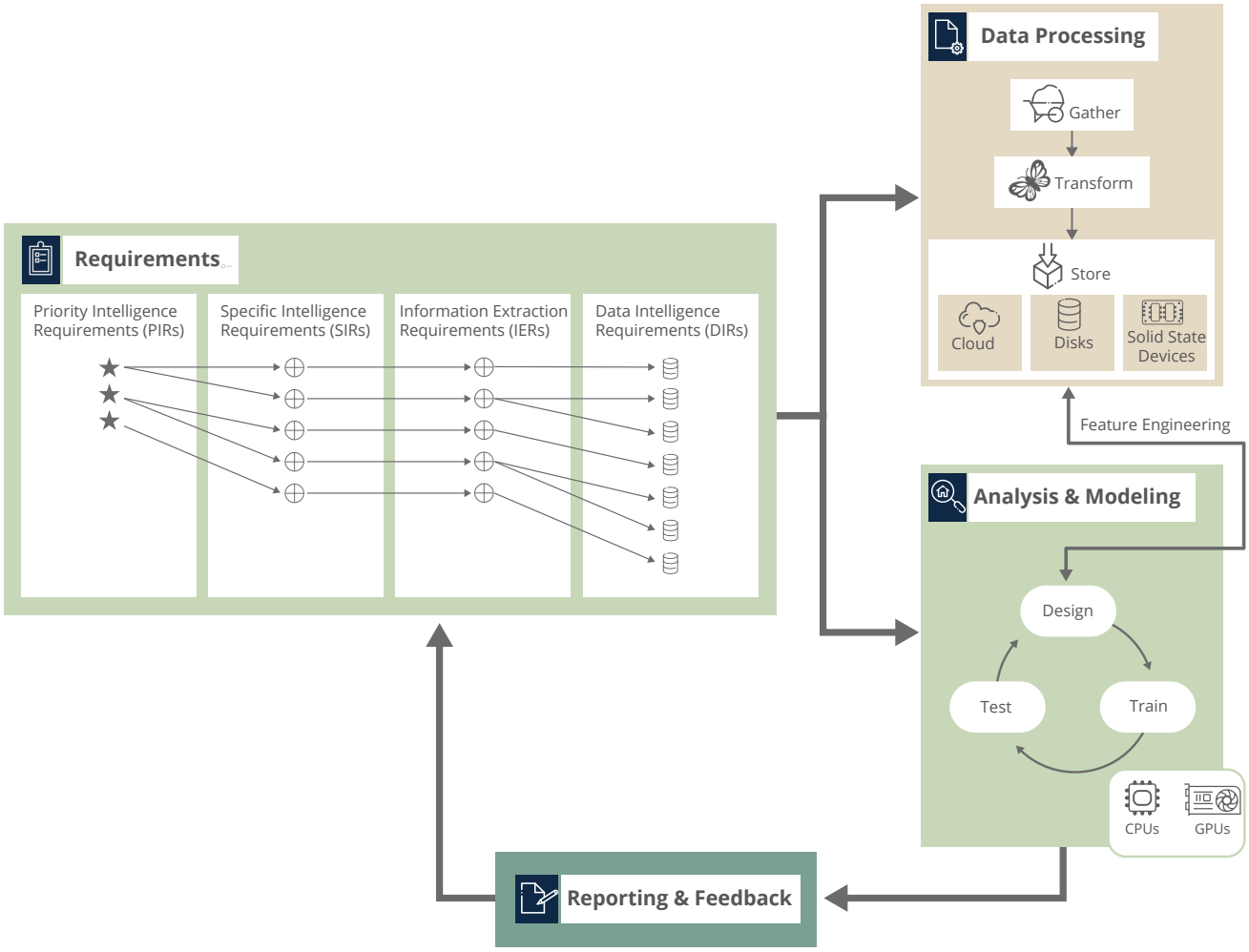



**Figure 1.** The machine learning pipeline for strategic analysis.

## DEFINING REQUIREMENTS

Consistent with current cyber intelligence best practices, an ML planning process should begin with specific intelligence requirements (SIR) that map to priority intelligence requirements (PIR) from your collection management team. These cyber intelligence requirements are your north star. Without an intelligence requirement stated in a clear problem definition, analysts can veer into questions such as "What will happen if I try this fancy new algorithm on this data?" While that kind of exploration can be useful for getting to know the data and trying out whether a new algorithm works as advertised, it probably won't solve the big picture problems for the organization.

A common complaint from many organizations is, "I have more data than I know what to do with." This happens because data was collected without any particular use in mind. Consequently, the organization often does not record information that would be useful to answer any particular question. Not all data is equal, and for most companies enormous stores of inert data provide no incremental value. The planning stage is critical to avoiding this problem.

We suggest creating explicit information extraction requirements (IERs) and data intelligence requirements (DIRs) that map directly to the organization's intelligence priorities held by the collection management team. These requirements will help guide the ML analysis and data collection and provide an explicit mechanism for tracking how data is used. The IERs and DIRs need to be developed in tandem, because the kinds of ML analysis desired will have different data needs and the data available will shape what ML analyses are possible. The IERs need to answer questions such as

- What kind of data science method should we be using?
- What metrics will we use for success?
- Are there any other criteria necessary to make the results useful?

The IERs should be developed by the ML scientist and the cyber intelligence analysts. The consumers of the ML output are there to ensure that the ML scientists understand their needs. The ML scientist must be there to translate those needs into properties of the analysis. When an ML scientist first meets with a client, she will listen and ask a series of questions in order to understand the client's needs. The IERs simply make these needs explicit and directly link them to the organizations' PIRs and SIRs.

To develop the IER, it's helpful to begin by asking questions such as

- What does a minimally sufficient solution look like?
- How good does an ML model have to be to be useful?

The answers very much depend on the use context for a problem. If all conclusions must be strong enough to present as evidence in a court of law, that is a very different threshold than one needed to simply ask analysts to investigate a suspicious anomaly. When we are deciding what criteria our model should optimize and what thresholds it must meet, we must take this context into account.

**BEST PRACTICES**

**SETTING THRESHOLDS**

While a 5% error rate has been a standard threshold in statistics and data science for about a century, it may be a bad threshold for your application. A popular social media platform recently deployed an ML algorithm to detect adult content published on its platform. This is a prediction algorithm designed to detect whether a given image contains particular types of nudity. Like any prediction algorithm, it will make mistakes. However, for this large platform, a 5% false positive rate constitutes approximately 500 million misclassified images and a corresponding number of users unhappy that their images are being blocked for no apparent reason. The default 5% error rate is not low enough for this context. It is worth spending the time and effort to determine what kind of criteria must be met for a solution to be useful for your problem.

DIRs follow from IERs. Once IRs, PIRs, SIRs, and IERs are specified, DIRs address concerns such as

- What data do I need to fulfill these IERs?
- How much data do I need?
- Are there any data collection methodologies that need to be followed?
- What are potential sources of the data that is needed for analysis?

Different analyses require different amounts of data. Simple analyses might require hundreds of data points, while more complex analyses, such as neural networks, require tens of thousands. The DIRs make these kinds of requirements explicit. In the section "Modeling" below, we discuss when labeled data is necessary and why it is imperative that the collection conditions match exactly the conditions in which an ML model will be used. The DIRs specify these requirements.

Defining the collection conditions is an essential part of a documented and repeatable process of requirements generation. It is also important to frequently verify DIRs. As the nature of any cyber intelligence analysis changes, the types, amount, and sources of data also change. Consequently, ensure that DIRs are not carried over from previous tasks simply because previous tasks seem similar.

One last point regarding IERs and DIRs: They should all be designed with specific expiration dates. The questions of interest to an organization will necessarily change as the cyber threat landscape changes, and an organization should adapt its analysis to meet the new questions. Design the requirements with this in mind.

### Gap Analysis

Gap analysis can be useful in defining requirements. This is a formalized process
for answering

- Where are we now?
- Where do we want to be?
- How do we close the gap between here and there?"

Through this gap analysis process, you may discover that the data you currently collect does not actually contain the necessary information to address your problem.
For example

- Where are we now?
  - We analyze netflow data to determine whether our network is under attack.
- Where do we want to be?
  - We want to learn the identity of our attacker.
- Identify the gap.
  - Netflow data is too coarse to identify how the attack is being executed.
- How are we going to close the gap?
  - We need to collect data directly on the computer being attacked and, possibly, examine logs and memory dumps.

This netflow example highlights an important lesson: Just because you have data, that doesn't mean it has the right information to answer your current question. Netflow data is perfectly adequate to provide evidence that an attack is occurring. But it's completely insufficient to answer how the attackers got in, how they are moving through your network, or what they're doing inside. Gap analysis is a useful tool for identifying requirements in general, including which new data needs to be collected.

### QUESTIONS YOU SHOULD BE ABLE TO ANSWER AT THE PLANNING STAGE

- What is the relevant IR/PIR/SIR?
- What does a "minimally sufficient solution" look like?
- Create information extraction requirements (IERs).
  - What kind of analysis is required (e.g., classification, anomaly detection, clustering…)?
  - What metrics will be used to measure success?
  - Are there other criteria that must be met (e.g., specific run time, processing limitations)?
    **Note:** You should not settle on any particular algorithm at this stage, simply identify the needed metrics and criteria
- Create data intelligence requirements (DIRs).
  - What data do I need to answer this question?
  - How much of that data do I need?
  - Are there any specific collection requirements (e.g., random sample)?
  - You might have a couple of data sources in mind, but it's too early to commit to a specific one.

## DATA PROCESSING

Data gathering and data processing are where you should expect to spend the majority of your time and effort. It should be noted that, within the field of ML, the biggest improvements in performance come from a foundation of better, higher-quality data.[3] This is true in cyber intelligence as well. For example, in one study the authors tried seven different algorithms to predict, from three sets of features, whether or not a file was malware. The differences among the algorithms' performances was minimal. The differences among the prediction accuracy for different features were pronounced (Table 1). This example highlights two things that are almost always true about ML: 1) Better data generally makes more of a difference than algorithm choice, and 2) There is generally a different cost associated with different features.

| Feature of Executable File | Ease of Extraction | Prediction Accuracy |
|---|---|---|
| n-grams of bytes | Cheap and easy to extract | 60–80% |
| Opcodes | Requires disassembling the file, medium cost and effort | 85–95% |
| API calls used by executable | High effort and computational time to extract | 90–95% |

**Table 1.** Prediction accuracy in malware detection algorithm study

3 https://arxiv.org/pdf/1808.01201.pdf

The majority of the work in ML takes place in the data preparation stage. Some estimates suggest you'll need as many as five data engineers for each ML scientist. In our experience with customers, this is not unrealistic; the amount of work required to prepare data for ML should not be underestimated.

One of the best practices we observed was to automate data gathering and processing as much as possible. Anything you do more than once should be automated. In fact, one team noted that achieving any sort of scale without automation is impossible. Having an automated data collection and processing system, on the other hand, allows teams to get more and different data, allows for continuous improvement, and results in direct savings in labor costs. Furthermore, the automation provides time for the analysts to work on more pressing issues.

## GATHERING DATA

First and foremost, identify data sources in your DIRs that meet your needs rather than collecting and storing data using an ad-hoc approach. We also recommend tracking the sources and making that information part of how you store the data. When gathering data, it is already common practice among database experts to create an extensive data dictionary describing what each data element is and how it was generated. However, in the context of ML, consider adding the source of each data element to the dictionary. Doing so will not only help you track where data elements come from, reducing technical debt later on as models are updated, but also allow you to assess the usefulness of each data source in the future. This assessment can help you decide whether or not to continue to collect particular data. In addition, an ML algorithm can be adapted to weight each source based on prior knowledge about source quality or the algorithm's assessment of the data's value in making predictions.

It should also be noted that manual data entry is highly error prone, expensive, and often infeasible at scale. Avoid manual entry as much as possible. We reiterate that automation in all aspects of data gathering should be pursued to the extent possible.

## DATA TRANSFORMATION

Once data sources have been obtained, the data must be prepared for storage and subsequent analysis. This typically entails at least three steps: cleaning, storage, and feature engineering. In this section, we will discuss the first two steps. Feature engineering will be discussed in greater detail at the end of the section "Modeling."

Data cleaning entails ensuring the data is of proper quality for future analysis. This work includes tasks such as handling missing data, outliers, correct-but-highly-unlikely values, and similar problems. While each of these tasks can be handled in any number of ways, the key here is consistency. Mishandling these data corner cases can result in the loss of significant data, incorrect conclusions, or missing entire swaths of data for a variety of reasons. Given that data is the fuel that powers all analysis, the intelligence analyst, the ML specialists, and any subject matter experts should all agree on how such corner cases are handled.

Over the past two decades, the technology for storing data has become incredibly sophisticated: It is now possible to store almost any type of data using commercial-off-the-shelf products. Unfortunately, this can make life difficult for the ML practitioner, because it's now often far easier to simply toss

data into a database than it is to prepare it for subsequent analysis. Consequently, storage should be performed with the end goal in mind; in this case, subsequent usage in an ML algorithm. So, when considering a data storage solutions such as relational databases, time-series databases, NoSQL engines, binary blobs, graph entities, and document-based storage, it is critical to consider how the stored data will be consumed. It may be, for instance, that the data should be stored in multiple formats—columnar tables as well as JavaScript Object Notation (JSON)—to enable different types of analytics. Because the nature of the analysis will vary tremendously based on the use case, analysts should ensure that the data storage format is working to their advantage rather than posing a hindrance.

## MODELING

In this section, we will discuss the two broad categories of ML problems (supervised and unsupervised learning) with the goal of enabling leadership and cyber intelligence analysts to work effectively with their ML team. The type of modeling you need is determined by your PIR and/or SIR. It should be specified in the IER, which you established based on your PIR. Categorizing an ML problem this way can help constrain it, and it can help you begin the process of translating an organizational need into a tractable data science problem.

We emphasize that basic exploratory descriptive statistics should always be completed before beginning a more complex ML analysis. Descriptive summaries are some of the most basic tools, but also the most useful. Such summaries include visualizations, averages, standard deviation, correlations, and proportions. Every project will need them, and many problems can be solved with these tools alone. They answer the question, “What does normal look like in this dataset?” For example, knowing how much traffic your network usually carries on Monday morning as opposed to Thursday afternoon is very useful in planning infrastructure and scheduling system updates. This information could also be used to provide a baseline for an ML algorithm against which network traffic anomalies could be detected. Moreover, creating a simple visualization is often the fastest way to reveal errors in data collection; for example, a negative number denoting the number of logons on a particular day would be a clear indicator that the data collection process needs to be checked. Descriptive summaries provide a necessary foundation for interpreting the results of more complicated analyses.

## PREDICTION AND SUPERVISED ML

A supervised learning problem can be framed as "I would like to use _____ data to predict _____." The key requirement for supervised learning is labeled data. For instance, if you want to use the hash of a binary executable file to predict who made the malware, then you need to have the hashes of other binary files that have already been labeled by source. Likewise, if you want to predict which new pieces of information are most important for human cyber intelligence analysts, then you must have data on prior information that the analysts have labeled as useful or not useful.

This labeled data will serve as the "ground truth" for training the ML algorithm.[4] The quality of labels applied to the training data will directly impact the quality of the ML output.[5] For example, let's assume a case in which labels from cyber intelligence analysts on different pieces of information indicate whether the information was useful or not. It is easy to imagine different analysts disagreeing about how to label the same piece of information. Moreover, certain pieces of information might meet some, but not all, of the criteria for earning the label "useful." How should these data points be labeled? All of these small decisions will impact the utility of the results from any ML algorithm applied to the data. Indeed, when we fit a supervised algorithm, the final model is the one that makes the best predictions on a test set of the labeled data, so it is imperative that the labels be meaningful and accurate.

The second consideration in selecting a supervised learning algorithm is to identify what kinds of information you need to be able to get out of your analysis. Do you need to predict whether or not a new source, piece of data, or method of analysis will be useful? Or do you need to infer what features of the information indicate whether it will be useful? The first case could be valuable if you just need to help your intelligence analysts figure out what needs their attention most. The second case is more important if you are evaluating which information services you want to continue to collect. Many methods can be used for either prediction or inference (e.g., logistic regression), but some methods can only be used for prediction (e.g., k-nearest neighbors).

Supervised learning models work under specific assumptions and constraints. A supervised learning model is trained on one set of labeled data and then applied to new data. For this to be successful, the assumptions and conditions that held true during training must also be enforced during deployment. Consider again the malware example: There are large open source repositories of malware files that could be used to train an algorithm. But these repositories have been highly curated and contain files that are easily and distinctively identified as malware. Malware "in the wild" will not be so easy to identify. If you train your algorithm to identify malware from one of the curated repositories, it will only catch the "easy-to-identify" cases.

Moreover, the order for a supervised learning process is always
1. Collect the labeled data.
2. Train a model.
3. Apply the model.

4 Ground truth is simply defined as "the correct answer." Labeled data provides the correct answers for every input.

5 https://ssrn.com/abstract=2477899 or http://dx.doi.org/10.2139/ssrn.2477899

Therefore, the labeled data is always data from the past that we will leverage for the present. When we apply a supervised learning model, we are making a bet that tomorrow will be the same as yesterday. This is often a bet we are willing to make. For example, using the pixels of an image to predict whether or not there is a face in that image is not a task that will evolve rapidly over time. However, tomorrow's malware attack probably won't look very much like one in the past. Immutability over time is simply one example of the general immutability constraint. There can be disastrous results if the application conditions do not sufficiently match the conditions for the collection of training data.

## SUMMARIZING STRUCTURE AND UNSUPERVISED ML

Unsupervised learning methods answer the question, "I would like to understand the structure of the features recorded in ____ data." The most important consideration for using unsupervised ML is "What kind of structure are you looking for?" Different algorithms will find different kinds of structure. The three most common types of structures we might be interested in are clusters, anomalies, and sets of highly correlated variables. Other kinds of structures include lower-dimensional representations, density estimates, and latent variable representations. The IERs based on the PIRs and SIRs should specify the kinds of structures we are interested in for a particular analysis. The reason you must pre-specify the structure you are interested in is twofold. First, any patterns discovered in the data have to be functionally meaningful to the intelligence analyst. Second, many of these methods will impose structure on the data, even when that structure might not actually be present; if you pre-specify the patterns of interest, it is easier to evaluate whether or not the patterns you find are actually real.

In a clustering problem, the question is "Are there meaningful subgroups in this data?" For example, we might be interested in identifying users with similar patterns of usage (e.g., early risers, night owls, and system administrators). We could be looking for groups of incident tickets that all have the same attack patterns. Common methods for extracting clusters include k-means, mixture modeling, and distance-based hierarchical clustering.

Anomaly detection tries to answer the question "Is there anything in this data that doesn't look similar to the rest?" For example, if our computer is infected with malware, it might be sending or receiving unusually large amounts of data; an anomaly detection algorithm might be useful for implementing a system to detect
an infection.

Assume the question "I have a whole pile of variables—are any of them related to each other?" For this case, simple dimension reduction methods that look for sets of highly correlated variables are appropriate. For example, if you are looking at usage statistics on a website, then number of clicks and total time spent on the page are going to be highly related: The more time someone spends on the page, the more links they're likely to click. In general, dimension reduction techniques (called factor analysis in some communities of practice), focus on finding a simpler representation of the data that contains the same information as the original data. One of the most popular techniques for dimension reduction is principal components analysis (PCA), which uses basic linear algebra to find groups of features that are linearly correlated. However, PCA is most appropriate for numerical data. Other methods, such as word2vec or latent Dirichlet allocation, are more appropriate for textual data.

Measuring the success of a supervised learning method is fairly easy: How well do my predictions match the truth? In comparison, measuring the success of an unsupervised learning method is much trickier. If you ask a clustering algorithm to find five clusters in the data, it will find five clusters, but they may not be meaningful (see Figure 2). One of your criteria for success is that the patterns discovered have to be functionally meaningful to an intelligence analyst; this criterion could be measured informally or with a survey, depending on how big the team is.

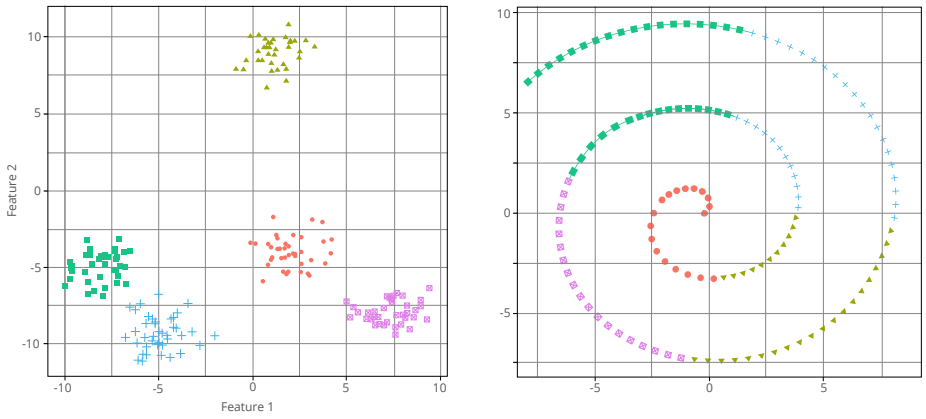



**Figure 2.** When clusters are present (left), a clustering algorithm will find clusters. When a different structure is present, (the spiral is a type of lower dimensional manifold), a clustering algorithm will still try to find clusters.

## FEATURE ENGINEERING

Feature engineering is an integral part of the feedback between data processing and modeling. If we can refine and improve the features we use to train a model, then we can achieve increased operational effectiveness through greater model accuracy or reduced training time. Feature engineering can take two forms. The first is feature selection, which is used to mitigate redundant features (information is already contained in another feature, duplication) and irrelevant features (features contain no lift to an applicable ML task at hand). The second, feature extraction, is used in creating new features by combining original features under a consistent methodology. The PCA technique mentioned above is often used for automated feature extraction.

It is worth emphasizing that manual feature engineering requires more thought and effort but often produces greater rewards. Through close collaboration between the ML team and the cyber intelligence team, it is common to discover that, for example, "We've been using the total count of event X, but we get better results if we use the time that has passed since the last event X." As discussed earlier, better data often provides more improvement than better algorithms. Similarly, leveraging the domain expertise of the cyber intelligence team in feature extraction will produce better
ML results.

### ADVERSARIAL ML

Adversarial ML is not a substantial threat yet, but it will be in the near future. Adversarial ML is still largely the domain of the research community, which is working to document the different ways in which ML systems might be vulnerable; methods for hardening AI systems against adversarial attacks are cutting edge research. The methods of attack have not permeated the script kiddie community yet; however, adversarial examples are already abundant. Most of these examples follow a pattern in which an organization trains and deploys an ML system, then the adversary builds an attack that deceives it. However, they are nonetheless powerful for demonstrating the dangers: a 3D printed turtle[6] classified as a gun and a pair of glasses[7] that causes facial recognition software to believe you are John Malkovich. Another attack type we can expect in the medium-term future is the injection, by an adversary, of malicious examples into your training data.[8]

## DECISION MAKER REPORTING AND FEEDBACK

No matter the ML technique or the area of application, the results must be understandable to the end users of the output (the "consumer"). Even more importantly, these ML consumers must have a mechanism for providing actionable feedback to developers and analysts to ensure that the analysis is not only understandable but also valuable. The most effective way to ensure this conversation can happen is by defining a common language. Given the expected widespread adoption of ML solutions, developers and analysts cannot assume that all consumers will be literate in ML techniques, lingo, and nuances. However, a certain amount of literacy is required to ensure that useful feedback can be provided. Literacy in the following specific concepts should be enforced:

- **The concept of "probabilistic answers" is common in ML.** Many ML algorithms do not answer questions with a "yes" or "no" but rather with a likelihood that a given scenario has occurred. For example, consider an algorithm observing a large amount of network traffic coming from a known bad set of IP addresses. The algorithm may be intelligent enough to recognize the activity as a component of an attack, but may not have enough additional data to further classify the attack, or it may not be trained to recognize this specific type of activity as an attack, or any of a number of similar possibilities. In this scenario, algorithm output may indicate that an attack is occurring with 34% likelihood. While this is not something a person would say, it represents how algorithms process input.

  Additionally, these outputs could be mapped to ICD 203[9] expressions of likelihood. ICD 203 §D.6.e.2.a describes colloquial terminology ranging from "almost no chance" to "almost certainly," mapping likelihoods to seven possible categories. It is critical for executives to understand that these terms are not chosen arbitrarily but correspond to specific likelihoods provided by the algorithm.

- **Algorithm performance depends on two factors: the training set and the currently available data.** Continuing our previous example, assume that our analyst recognizes that this type of traffic

6 https://www.theverge.com/2017/11/2/16597276/google-ai-image-attacks-adversarial-turtle-rifle-3d-printed

7 https://qz.com/823820/carnegie-mellon-made-a-special-pair-of-glasses-that-lets-you-steal-a-digital-identity/

8 https://www.cs.umd.edu/~tomg/projects/poison/

9 https://www.dni.gov/files/documents/ICD/ICD%20203%20Analytic%20Standards.pdf

is perfectly normal. Does this recognition mean that the algorithm is junk and should therefore be ignored? Of course not! The algorithm only knows what it was shown in the past and what it has available to it at the moment. Unfortunately, it is all too common for the consumer to use this data point to dismiss the AI as "junk." Misbehavior in one scenario does not imply misbehavior elsewhere. Users of AI systems should find out what types of data the AI is best equipped to handle and be extra cautious about trusting output when feeding the system data outside its expertise. Similarly, when dealing with the AI's "specialty" data, pay close attention to the output before dismissing a seemingly spurious result: The AI may see a pattern or trend that a human would normally miss.

Note also that some AI systems possess the ability to continuously learn new information. For example, modern spam filters are "preprogrammed" to identify generic spam. As users tag the spam they personally receive, the system learns new types of spam and classifier performance increases. Some cyber intelligence systems possess a similar capability; if this is the case, consumers should be aware that their labels are being included in the system.

- **Appropriate trust is key.** When it comes to AI, trusting the output too much or too little can be problematic. In the earliest uses of AI in aviation, there were crashes because pilots did not trust the AI system. In contrast, we know that there are systematic biases in which AI results deserve less trust. Trusting a system too much may be particularly problematic if an adversary figures out how to craft an attack specifically targeted to avoid detection by the AI—even if the human would have identified the attack without AI assistance, overreliance on an AI system may lead analysts to trust output without validating it. "Trust but verify" is a healthy motto.

Once consumers understand and internalize these concepts, they must then understand how to convey feedback to analysts. "This doesn't make sense" is almost never considered useful feedback. Rather, we recommend that consumers try to make their feedback more actionable, focusing on the levers that analysts can tweak. The following examples demonstrate various types of actionable feedback. In all cases, "you" refers to the consumer.

## ORGANIZATIONAL STRUCTURE FOR ML

There are many concerns related to creating an ML activity within your cyber intelligence organization. It is difficult to understand the team composition, how to collaborate with an ML activity, and how to best support the ML activity with proper policies and infrastructures. In this section, we outline how to organize your cyber intelligence team to achieve success with ML, and we also look at how you can incorporate some classic software engineering principles to ML to ensure high-quality output.

## ORGANIZING AN ML EFFORT WITHIN AN INTEL TEAM

An understanding of the relationships among IRs, PIRs, and SIRs over time, region, and industry is maintained through individual roles and responsibilities. A team that can function effectively at the intersection of cyber, intelligence, and ML must include people with specific backgrounds, skillsets, and traits. Moreover, the team members must have a clear separation of roles and responsibilities while at the same time allowing close collaboration and effective information sharing.

A successful ML–cyber intelligence effort requires three parts:

- **domain expertise**: knowledge of cyber intelligence and other organizational context
- **ML expertise**: understanding of the underlying theory in ML and how to apply it
- **data engineering expertise**: ability to engineer systems that integrate and scale ML and cyber intelligence capabilities

Without these three kinds of expertise, an ML effort within a cyber intelligence team will find it difficult to succeed and scale.

ML scientists, cyber intelligence analysts, and data engineers must all have depth in their respective domains, but they must also be able to understand and, most importantly, communicate across their domains. These are three very large bodies of knowledge, so it is rare to be able to hire an individual with expertise in more than one of these areas. However, within a cross-domain cyber intelligence team, individual members will usually have a primary area of specialty and will develop expertise in a secondary area as they work in the intersection. This can be facilitated with formal training and collaboration sessions, but is often achieved informally via day-to-day interaction and collaboration.

Close collaboration and open communication are critical at all times and can likely be better facilitated within a fusion center, where diverse teams come together to analyze disparate information. There are complex design requirements that exist in each domain of work that require each practitioner to take the restrictions and needs of another domain into consideration at every step of their work. For example, there are fundamental limitations to what a compute resource can accomplish. The cyber intelligence analysts and ML scientists must listen and make adjustments to ensure their solutions do not grossly overestimate the fundamental assumptions that a software engineer is taking for granted. The ML and engineering personnel need to be on, or work closely with, the cyber intelligence team. Since their jobs involve directly modeling and analyzing data collected and tagged by cyber intelligence analysts, it is essential for them to be included in regular information sharing and planning meetings, especially regarding information collection practices or procedures.

## SOFTWARE ENGINEERING FOR ML

While ML processes can be used to effectively monitor, assess, and model environmental events, they are not without their operational concerns. Under the best circumstances, an ML system designed today will need to be monitored and adjusted as time passes.

Stakeholders must understand, at least on a high level, that software engineering for ML looks different from software engineering elsewhere in their organization, because following good software engineering practices will enable you to adapt to changing circumstances. Traditional software engineering attributes, such as functionality, usability, reliability, performance, and supportability (FURPS), still apply to the world of ML. However, their specific implementations will look different for the ML pipeline.

Specifically, ML systems require much more verification at each step of the process than the surrounding software pipeline. The following list breaks this idea down into the specific FURPS components:

- **Functionality**: As the model is training and constantly updating, and as it is being exposed to more data, does it still achieve the task it is being created to solve? The passage of time and the application of training updates does not guarantee functionality. Consequently, verification must be woven into the functionality pipeline.
- **Usability**: Is the model consistent? Does it produce outputs humans can understand and reason about? Does the consistency change as the model gets new training updates?
- **Reliability**: Does the model provide stable outputs? If the model predicts an input as class 1 with high confidence at time t, will it still predict class 1 within the same confidence at time t+1? This must be verified regularly and not taken for granted.
- **Performance**: Does training the model more and more increase the runtime requirements for inference? Does the model become too large to feasibly deploy to production systems? Constant performance checks must be in place.
- **Supportability**: Is it simple to influence the behavior of the model early in the training process? How about late in the training process? How can you update a production model and ensure its veracity? A verification pipeline must be constructed here as well.

There are two places in which the differences between software engineering for ML and traditional software engineering are particularly stark. First, the mental representations a software engineer uses to think about data are fundamentally different from the mental representations an ML scientist uses to think about data and modeling. These differences can cause communication difficulties between the software engineering team and the ML team, so it is critical to have open lines of communication to resolve these issues.

Second, ML creates a tight coupling between the content of the data, the model, and the final use of the information. In fact, the model is created specifically to connect the data to the end use. This is in stark contrast to traditional software engineering, in which tight coupling is forbidden. However, the format of the data storage, the hardware, and the implementation of the algorithm can and should all be loosely coupled. Once again, close collaboration between the ML team and the software engineering team is required to address these issues.

## INFRASTRUCTURE AND HARDWARE

A strong computing infrastructure is necessary to maintain a healthy and capable data science effort. Namely, the availability of computing infrastructure with networked storage and compute (CPU and GPU) capabilities is essential. Many companies will already have infrastructure for this purpose elsewhere in the organization, so partnering with those teams and growing their infrastructure is a strong possibility.

Along with computing infrastructure, quick and reliable data storage is a necessity. Data is processed repeatedly in ML R&D and production; strong data storage capabilities will augment the volume and speed of the ML effort.

With large data environments, there need to be strong measures in place to protect the transfer, usage, and availability of data. All ML and cyber intelligence personnel must follow rules set by the CISO organization. These measures also include data governance and compliance, which can apply to both on-premise and cloud infrastructures. Due to the tight data coupling in the ML domain, guidance must be created in conjunction with the CISO organization to avoid any potential issues.

Carnegie Mellon University
Software Engineering Institute

# The Internet of Things and Cyber Intelligence

## An Implementation Guide

# The Internet of Things and Cyber Intelligence: An Implementation Guide

**Authors**
Dan Klinedinst
Deana Shick
Jared E. Ettinger

**Design**
David Biber
Alexandrea Van Deusen
Todd Loizes



This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.



DM19-0447

Contents

# The Internet of Things and Cyber Intelligence

## INTRODUCTION

The Internet of Things (IoT) can be a valuable source of data for cyber intelligence analysis. Sensors can provide information about the real world that is absent from traditional information systems, and often cost far less than general purpose computers. This implementation guide extends the basic Cyber Intelligence Tradecraft Project (CITP) analytic framework with guidance specific to the IoT. Organizations might find that, as they implement more automation, they have a wealth of new data available to them, both from their own devices and third-party providers.

### DEFINITION OF "IOT"

The IoT is a "network of physical objects that contain embedded technology to communicate and sense or interact with their internal states or the external environment."[1]

Gartner's definition of IoT is fairly concise yet comprehensive. However, for our purposes, we are going to limit IoT to objects that communicate, directly or sporadically, with the Internet. This excludes devices that communicate only via voice, SMS, or data networks that are entirely separate from the global Internet. However, it does include devices that communicate via Internet Protocol (IP) but are intended to be on standalone networks. The reason for this is that any device that can communicate via the Internet is likely to do so at some point, even if it's intended to be on a virtual private network or physically isolated network.

Some categories that are specifically included in our definition include

- industrial control systems (SCADA, Modbus, etc.) that are connected to IP networks
- embedded devices in vehicles
- building automation systems
- sensors that communicate their data via local radio networks (e.g., Zigbee and Bluetooth) to gateways that communicate over the Internet

Our working definition of IoT does not include apps or cloud services with which the devices communicate, although those may be involved in gathering and analyzing data from IoT devices.

1 https://www.gartner.com/it-glossary/internet-of-things/

## SCOPE OF IMPLEMENTATION GUIDE

This implementation guide focuses on the data gathering and analysis components of the cyber intelligence framework. It examines how intelligence from IoT devices can be incorporated into broader cyber intelligence analysis to advance decision making. Organizational risks can be purely cybersecurity, such as an effort to steal data over the Internet, or they can be financial, political, or physical risks that involve cyber and non-cyber components or indicators.

Certain cyber-specific IoT risks could be defined as risks that directly impact your IoT devices and/or sensors or threat actors using IoT systems to pivot to other digital targets. Some potential risks in this category include

- loss of confidentiality of data stored on or collected by IoT devices
- loss of integrity of data generated by IoT devices (that is, can you trust what your sensors are telling you?)
- loss of integrity of actions of actuators (e.g., causing an incorrect action in the physical world)
- loss of availability of IoT devices
- loss of availability of larger cyber-physical systems, such as the inability to use your car because your "smart key" was hacked
- attackers recovering credentials for other systems (IoT devices might have privileges on cloud services, databases, or other IoT systems)
- attackers using IoT to launch a distributed denial of service (DDOS) attack on other systems (e.g., Mirai)
- attackers pivoting through IoT devices or networks to attack other, less exposed networks
- using compromised IoT devices to "jump the air gap" (for example, compromising a device that uses Bluetooth and then using the Bluetooth radio to connect to an air-gapped system)

We define non-cyber-specific risks as those in which threat actors compromise IoT to facilitate another crime, or in which a non-cyber crime could be detected via IoT devices. It is important to note that the boundary between "cyber" and non-cyber" is becoming less clearly delineated as IoT devices become more widely adopted. Hardware, from computers to entire networks, is increasingly becoming virtualized. Additive manufacturing ("3D printing") is making it possible to instantiate virtual designs as physical hardware with minimal logistics. Therefore, non-cyber risks will increasingly have cyber threats and implications.

The number of these scenarios will grow as IoT becomes increasingly integrated into everyday life, but some examples might include

- physical threats, such as disabling cameras to rob a bank or hacking a car to aid in a kidnapping
- fraud, such as bypassing a subway turnstile by hacking the smart card reader
- corporate or nation-state espionage
- terrorism or sabotage
- theft of intellectual property
- insider threats, which cover a wide variety of crimes performed by a trusted person

### PRIVACY

Collecting data from the IoT poses unprecedented privacy risks. Any organization that chooses to collect data, even from its own systems, needs to carefully consider both the legal and ethical implications of that collection. This implementation guide cannot definitively prescribe privacy practices, as the standards and mores regarding privacy vary dramatically between polities, cultures, and industries. Different standards will apply depending on whether your organization is collecting information about employees, customers, users, competitors, or other stakeholders.

The following list presents several resources to help you address privacy concerns as they relate to the gathering of data from IoT devices. However, every organization should consider privacy implications and consult with lawyers and privacy experts before implementing an IoT component in a cyber intelligence program.

- NIST Considerations for Managing Internet of Things (IoT) Cybersecurity and Privacy Risks Workshop[2]
- IoT Privacy and Security in a Connected World[3]
- Privacy Expectations and Preferences in an IoT World[4]

### OUT OF SCOPE

This implementation guide is intended to advance the state of practice of cyber intelligence. This document is not intended to be a guide for securing IoT devices or the ecosystems in which they operate. The NIST Cybersecurity Framework is a good starting point for thinking about managing cybersecurity-related risk. This framework is also not intended to help organizations meet regulatory requirements, such as those in the General Data Protection Regulation (GDPR) or industry-specific regulatory frameworks.

While business intelligence is important to consider when your organization does any holistic cyber intelligence assessment, this implementation guides also does not address gathering business intelligence from IoT systems, although many of the same tools and analytical techniques could be applied.

## APPLYING THE CITP FRAMEWORK TO IOT

### ENVIRONMENTAL CONTEXT

In cyber intelligence, the environmental context is everything you need to know about your organization, both internally and externally. It includes understanding your organization's entire attack surface; the threats, risks, and opportunities facing your organization and industry; and the impact of those threats, risks, and opportunities on your organization and industry.

2 https://www.nist.gov/sites/default/files/documents/2018/06/28/draft-IoT-workshop-pre-read-document.pdf

3 https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127IoTrpt.pdf

4 https://www.ftc.gov/system/files/documents/public_comments/2017/11/00018-141696.pdf

The threat environment is defined by first enumerating risks such as those listed in the section on scope (both cyber and non-cyber) and then identifying which are most likely to impact the organization's mission. The integration of IoT into operations should be accompanied by an integration of IoT risks into an organization's overall risk management program. Additionally, IoT can be investigated as a possible source of information that can mitigate other risks that have been previously identified.

Cyber intelligence teams can use IoT to delineate an organization's attack surface. IoT sensors and telemetry data can help organizations identify and inventory networks and systems that have not previously been managed by corporate IT security. These can include operational technology (OT) assets, such as manufacturing, logistics, and maintenance. Today's organizations often have large deployments of networked equipment that are not managed from a cybersecurity perspective. Cyber intelligence teams should identify where they have current blind spots and look for data sources that can improve visibility. However, adding IoT devices into these shadow networks can also increase risk, as the devices themselves are a potential attack vector.

Externally, organizations will have to consider the IoT threat environment in which they operate. Connected products are going to operate in potentially hostile environments. They will be used in homes, businesses, and "smart" cities, where they will interact with networks and other IoT devices that might be owned or operated by adversaries, competitors, or even criminals. However, these external IoT networks are also potential sources of information—or otherwise useful information—if they share or make available data.

To identify new threats and vulnerabilities posed by IoT, organizations should update threat assessment and threat modeling activities to include IoT. There are numerous threat assessment and threat modeling tools available for these activities. The important thing is to identify possible attack vectors so that you can use intelligence capabilities to monitor for indicators of those attacks. Your threat assessment activities will vary depending on the nature of the threats, but some common tools for network- and software-centric threat modeling include STRIDE, AADL-Security Annex, FAIR, red teaming, and the NIST Risk Management Framework. These tools operate at various levels of technical detail, but all focus on identifying gaps in your knowledge of your environment. These gaps will guide your intelligence gathering activities.

When defining the environment for your cyber intelligence program, you should also consider the business, legal, and cultural environments you operate within. These environments will affect what data you need and what data you can legally gather. In some cases, you might be required to gather certain data (for example, GDPR in Europe and the payment card industry worldwide.) In some countries, the government itself might be a potential adversary, while in others it might put up barriers to your collection or transportation of certain data. The technical infrastructure will also form part of your environment. For example, areas with extensive high-speed Internet and current cellular technologies will differ from those still dependent on 2G and slow or sparse Internet access.

## DATA GATHERING

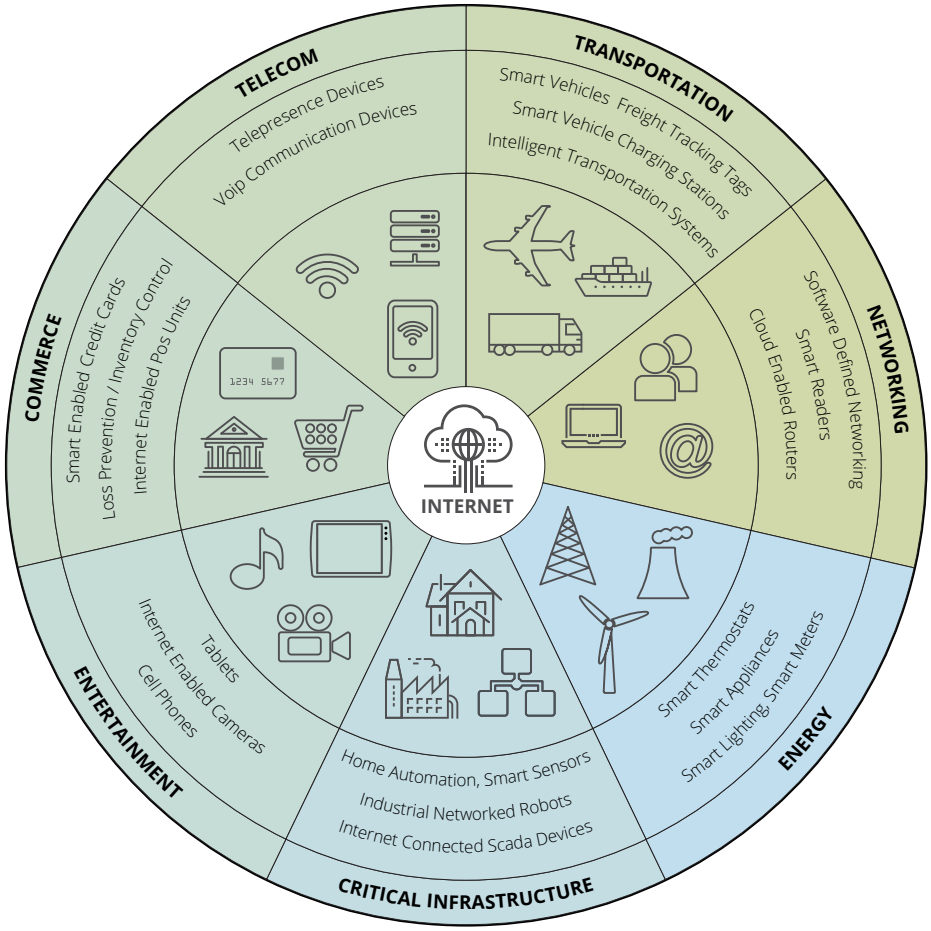



**Figure 1.** The number and type of devices from which data can be gathered is growing.

IoT devices are increasing the complexity of data analysts use to perform essential functions. The sheer amount of data available from these devices is overwhelming due to ease of implementation, the number of devices, and the frequency of readings. Unlike traditional computing devices, the number of IoT devices on a network can easily reach the thousands due to their lower cost and ease of use. Collecting data from these devices could create a data-scaling problem, making it very challenging for analysts to find the needle in the haystack. IoT data comes from a variety of sources: sensors built into the technology, metadata about the device, radio frequency (RF) communications produced by the device, and open source or public applications. Data gathering consists of two stages: first, collecting information from machine-to-machine communications and, second, gathering information from the public Internet or cloud service, where much of the machine-to-machine data is sent to be processed.

The sheer volume of potential data available means that analysts need to determine what data is most likely to help them. They should focus on collecting that data first. Analysts should make this determination by reviewing common requirements fed into their intelligence lifecycle and identifying what new data could help meet those requirements. Their environment, discussed above, will help them decide what IoT-based data is available and useful. Then they will need to set up mechanisms to access the data from internal or external IoT devices.

IoT devices can be implemented in a variety of environments and networks. Sometimes, the same device can be found across different sectors and implemented for different purposes. Several vendors compete in the space that provides security information and event management (SIEM) support for

analysts working with IoT data, and their products provide a collection mechanism for this data. These products sometimes have built-in algorithms to sift data accordingly. However, analysts will sometimes need access to raw data from the sensors. Sensor logging technology must be placed between the information flows to ensure analysts have proper visibility over the organizations and to gather IoT communications for analysis. It probably is not cost effective to record every network packet and store it for very long, but network security monitoring (NSM) systems have very robust mechanisms for distilling large amounts of data into the most important information. Organizations will probably want to focus on longer-term storage of the most relevant information: Netflow (standard records of which device communicated with which other devices), REST transactions (a common way of querying or directing IoT devices), and security events, such as logins or updates to credentials.

A traditional NSM will filter for these events and transfer the most relevant information to a SIEM. Increasingly, both the NSM and SIEM will use various machine learning techniques to identify what data is useful and extract insights from it (see the Machine Learning section below.) However, analysts will still need to have an understanding of the organization's IoT environment to derive useful intelligence and provide it to decision makers.

## THREAT ANALYSIS

After identifying data sources and setting up methods for collecting the data, analysts will commonly do various types of technical analysis. The results of these analyses will feed into more holistic and strategic types of analyses. In the IoT space, different types of technical analyses might be possible, depending on the type of data.

Network traffic analysis examines both the content of the traffic and information about the network traffic. Examining the content of the traffic can include monitoring for attack signatures, monitoring for exfiltration of high-value data, or even conducting sentiment analysis on text communications.

Analysts can also look at metadata about devices, which will vary depending on the device. Some useful data points might include changes of state (on/off), reset or pairing events, power levels, GPS coordinates, version numbers, security events (logins, key changes), etc. For example, the cellular modem in a car might be able to report its GPS coordinates, which could be cross-referenced with electronic logging devices in fleet vehicles, to look for fraud.

Most IoT devices will connect to the Internet directly or indirectly via one or more radio frequency (RF) interfaces: WiFi, cellular, Bluetooth, Zigbee, DSRC, etc. All RF devices provide information by broadcasting administrative commands, usually many times per second. Some of this information—either from the organization's own devices or from others' devices—could be useful. WiFi and Bluetooth identifiers can reveal identity information, and signal strength can be used for location triangulation. There is a large corpus of work on using SIGINT from RF transmissions.

Finally, there is the data that is being collected by the IoT devices themselves. This can be very general environmental information, such as ambient temperature, or much more detailed information, such as LiDAR scans. The data could be telemetry from industrial control systems or vehicles, or audio and/or video from cameras and smart assistants. Any of this data that could prove useful for cyber intelligence needs to be stored, normalized, and categorized. This is a prerequisite for further analysis via machine learning, data analytics, pattern detection, and other techniques.

## AUTOMATION

Automation is the process of telling a computer to run certain algorithms over data and output a result. It is a critical function for any organization pursuing an IoT analytic strategy. Intelligence can be derived from the result, and effort can be focused on more complex issues. Automation can detect issues within collection or analytic mechanisms. It can also detect failures, promote optimization of processes, and drive faster threat detection and situational awareness.

Automation is increasingly important to intelligence efforts because it reduces analyst fatigue in the face of massive amounts of data. The output from these processes answers questions quickly and could allow analysts to perform trend analysis over time. The data used during the analysis process must be gathered and normalized before any sort of automation can begin. A consistent format for the data should be considered based on analyst requirements and capabilities. Automation can provide results for tasks that are reoccurring or cyclical. For example, an automated process could detect malicious beaconing from within a network that contains IoT devices. Since beacons are relatively small and typically communicate over regular intervals, an analyst could automate that detection process instead of doing it by hand. The output of this automated task could be indicative of an infection on the network or an ongoing campaign against an organization. Ideally, organizations should try to automate as many tasks as possible to allocate resources to harder, more complex problems.

## MACHINE LEARNING

Machine learning is covered in depth in the *Machine Learning Implementation Guide* section of this report. However, there are a few areas where machine learning and IoT overlap that are worth discussion.

- The amount of data that needs to be processed might be very large, depending on the IoT devices in use. Examples might include video feeds from cameras, mapping data from vehicles, or environmental data, such as wind, humidity, and temperature. Any of these sources could generate hundreds of gigabytes of data per day. This fact needs to be considered if an organization wants to use machine learning platforms to analyze physical data.
- The type of data being collected might have different sensitivity or greater privacy concerns than traditional network or financial data. Anecdotal evidence suggests people are more interested in privacy if they are being recorded by video cameras or if the words they speak are being analyzed.
- Sensors might be interpreted by machine learning systems, and devices might take action in the physical world based on those interpretations. Self-driving vehicles provide an excellent example. Such interaction with the physical world will change your threat model and risk assessment activities.

- Attack vectors will also change. Attacks have already been demonstrated against both physical sensors (e.g., tricking cameras on a self-driving car) and against machine learning systems (model extraction attacks, data tainting attacks, etc.)
- Using machine learning systems for real-time detection of malicious activity within a network will be different when the network consists partially or entirely of IoT devices. Baseline behaviors, expected states and characteristics, and anomalies will all look different from traditional computing resources.
- It is likely that, in the near future, IoT devices will commonly have machine learning hardware and algorithms built in, rather than always depending on remote processing. This development will change how information can be gathered from IoT devices, since they will be pre-processing data and making decisions on it locally. The device's decision about what information is important to share will not necessarily be the same as a security analyst's.

## CORRELATION AND CORROBORATION

Data correlation allows analysts to uncover relationships between two datasets or variables. Correlation algorithms can associate seemingly disparate events, provide insights into known malicious activity, and allow better situational awareness across an organization.

IoT is novel because it might provide data points outside of traditional network data. IoT sensors can record data about their physical surroundings, such as video, sound, and environmental data. Correlating this data with other security data, such as network accesses, can uncover indicators of hybrid cyber-physical security issues. Furthermore, anomalies in the data being reported can indicate purely cybersecurity issues affecting IoT devices. Outages, abnormal readings, or increased activity might indicate an attacker is probing IoT devices or has already compromised them. This data can be correlated with network and authentication logs to identify nontraditional attack vectors.

## PATTERN RECOGNITION

Pattern recognition is defined as "the automatic discovery of regularities in data through the use of computer algorithms and the use of these regularities to take actions such as classifying the data into different categories."[5]

5 http://cds.cern.ch/record/998831/files/9780387310732_TOC.pdf

For IoT devices, pattern recognition will be applied to two overarching categories of patterns. One is the behavior of the devices themselves (e.g., network traffic). If a device makes the same connection to a remote Internet site every day, it is probably checking for updates or bulk uploading data. If that pattern is known, anomalies can be detected if the device suddenly starts making the connection 20 times per day. This is why identifying the correct data is so important; without it, analysts will not be able to recognize baseline patterns, anomalies, or potentially malicious patterns.

The other main category of pattern detection is patterns in the data observed by the devices. For example, IoT devices such as lights, door locks, and thermometers could reveal patterns about when certain buildings or rooms are in use. When correlated with other data, these patterns can reveal a larger pattern of life for an individual or a "pattern of business" for an organization. Anomalies in these patterns can also be detected, although the tools to do so are not as mature as network-based pattern and anomaly detection.

Pattern recognition and anomaly detection suffer from certain challenges. Cognitive biases might cause analysts (or even machine learning algorithms) to "detect" patterns that aren't there or miss ones that are. Anomaly detection is prone to false positives; that is, anomalies are usually more likely to be non-malicious variations than indicators of threats. Finally, adversaries will try to overcome these detection techniques. A well-known IoT example is the Stuxnet malware: in addition to changing the speeds of nuclear centrifuges, it also changed the speed readings being sent to operators so they wouldn't notice the anomalous speeds.

## STRATEGIC ANALYSIS

Strategic analysis is the process of incorporating technical IoT data into additional data feeds to paint a more clear and holistic picture about threats, risks, and other activity. It is unlikely that data from IoT devices or networks will be sufficient to assess risks and threats, except in the most straightforward attacks. More likely, IoT data will be used to augment other intelligence sources. It's important to remember that the additional data could either support or refute intelligence conclusions. Consider this example: An analyst discovers anomalous network traffic and suspects it might be reconnaissance for an attack. Then the analyst adds IoT network traffic to the dataset and realizes that a recently deployed set of IoT devices is generating the anomalous network traffic, and that this is expected behavior for that device. This fact doesn't preclude the data indicating an attack (a savvy attacker could try to blend into the traffic), but it lowers the probability.

Strategic analysis often has a predictive component to it. While technical analysis can tell you what is or has happened, strategic analysis is geared toward identifying possible future states as well as the likelihood those states will materialize. When it comes to IoT data, these future predictions are likely to be based on extrapolating trends from patterns in existing data and identifying gaps or problems that could be exploited by an adversary.

Below we discuss several techniques for doing strategic analysis and how IoT data might be incorporated. However, we are first going to present some examples of how you might merge IoT data with other data sources to obtain better intelligence:

- Compare IoT network traffic, such as DNS, HTTP and SSL, to similar traffic from traditional computers to identify gaps where security policies are different or not enforced.
- Combine delivery vehicles' telemetry data with inventory lossage to detect theft.
- Integrate WiFi, Bluetooth, etc. into your data loss prevention tools to detect localized theft of intellectual property or other espionage.
- Create baselines of network activity, RF activity, and foot and/or vehicle traffic (via video, for example) to detect anomalies that night be suspicious. This technique could be used in office buildings, retail spaces, transportation hubs, or manufacturing facilities.

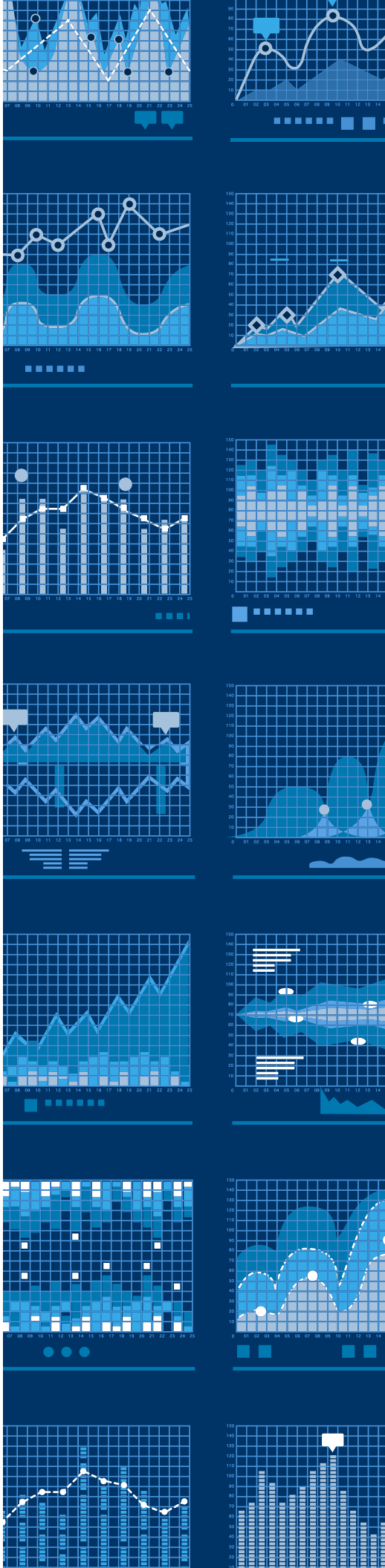
These are just a few examples of how different data sources offer a more complete picture. In the following sections, we'll discover specific tools organizations can use and how to incorporate IoT data into them.

## TREND ANALYSIS

In the paper *Trend Analysis as Pattern Recognition*, Dr. Stephen Millett describes three types of trend analysis: background, signals, and scatters.[6] We can apply all three types to IoT data to support cyber intelligence performance.

Background, or Type 1, analysis is about establishing baselines and looking for deviations or anomalies from those baselines. In IoT data, this could be regular network traffic (for instance, when devices update their configurations or report telemetry data). It could be physical patterns, such as daily fluctuations in room temperatures or noise levels. It could be physical traffic patterns, including vehicles, foot traffic, opening doors, etc. At the threat analysis level, these data points can be used to identify incidents, events, or tactical threats. In strategic analysis, these data points are used to extrapolate future states: the building will be cooler in the evenings, or there will be more foot traffic on weekends. These trends (Dr. Millett refers to them as continuities) can help the analyst decide what filters or anomaly detection to put in place in anticipation of possible incidents.

Signals, or Type 2, trend analysis means looking for specific patterns or changes. Network-based intrusion detection is a classic example of this: most intrusion detection systems include some ability to monitor for signatures. These signatures indicate a certain attack, a known family of malware, or an anticipated error ("Access Denied!") for example. IoT devices might exhibit these behaviors on the network.



6 https://doi.org/10.1177/194675670900100403

There might also be changes in the data recorded by the devices: for instance, a security camera might not expect to see movement in the middle of the night and will alert someone if it does. The important point is that this type of analysis can only detect *known* changes in the baseline trends already established.

Scatters are the third form of trend analysis. They describe previously *unknown* signals; that is, data that does not correspond to the baseline but also is not an anomaly that we predicted or expected. Understanding these signals, categorizing them, and deciding whether they are relevant can be time consuming. Fortunately, machine learning and data science techniques have increasingly sophisticated tools to both find and identify new patterns, trends, and anomalies in data.

## LINK ANALYSIS

In intelligence analysis, link analysis looks at the relationships between various entities. "Links connect people, things, organizations, processes, transactions, interactions, and activities. At the same time, they are reliable conduits of information."[7] In the IoT domain, links are likely to be technological, such as network links, or transactional, such as the data collected and promulgated through IoT networks. It's common when performing cybersecurity red teaming to discover that the assessed organization has many network connections and trust relationships that are undocumented or even unknown. The same is true of adversaries. Malicious code (malware) is linked via code reuse, threat actors are linked by their tradecraft and infrastructure, and insider threats are often linked to outsiders. These are just a small subset of links that can be discovered via analysis of IoT data. There are many tools for performing link analysis, from entry-level OSINT tools, such as Maltego, to high-end analytical tools, such as Palantir. Network modeling tools, such as Red Seal, can help you identify network connections and monitor or block them.

## TARGETING

Data from IoT devices can be used in target selection as well. While the military might use the data for selection of literal targets for attack, there are non-military uses as well. In a Red Team defense scenario (explained in more detail below), you might perform targeting against your own organization to identify and protect your critical assets. In a competitive business environment, you might use IoT data to help choose a new location, select a marketing campaign, or "target" a competitor's customers. In all cases, IoT data can be used to gather information about buildings, locations, people, and patterns of life. For example, imagine your IT environment consists of both desktops and sales kiosks in access-controlled office buildings in public malls. Your IoT devices (cameras, accelerometers) might indicate that people are likely to attempt unauthorized actions on the kiosks (trying many passwords or shaking it). Therefore, perhaps you should allocate more of your security budget to securing the kiosks than the desktops.

One free tool for doing target selection and prioritization is the CARVER methodology. Originally developed by the military for prioritizing how to use scarce resources to attack an enemy's assets, it has been retooled to prioritize technological and/or cyber assets. CARVER simply assigns every asset a score from 1 to 5 in each of 6 categories: Criticality, Accessibility, Recuperability, Vulnerability, Effect, and Recognizability. Those with the highest total score are notionally assigned as the highest priority ("most important") targets.

---

7 Hall, Wayne Michael and Gary Citrenbaum. *Intelligence Analysis: How to Think in Complex Environments: How to Think in Complex Environments*. Praeger Security International, 2010.

## USING IOT INTELLIGENCE TO HELP RED OR PEN TESTING TEAMS

The term "red teaming" has become more popular in the cybersecurity industry, but is most often used in a narrow sense to denote technical testing. We use "red teaming" to describe the general concept of taking an adversarial look at your organization, your possible courses of action, your security controls, etc. This exercise might include technical red teaming, wherein a team of specialists attempts to gain access to your networks and vital data or other assets. However, such testing can be very expensive and time-consuming. It can also be too dangerous to perform against production, safety-critical systems. Therefore, it is often worth performing tabletop red team exercises or performing them in simulation or test environments. This might be especially true of IoT-heavy systems, which might have dangerous effects on the physical world (traffic control systems, hospitals, HVAC, etc.). Even the most cautious red team can inadvertently damage systems that were not built to be resilient to active exploitation.

IoT devices and data can also be used to inform red team activities against other systems as well. They can be used to pivot to other networks, they can be compromised in order to hide the red team's activities (which might include social engineering or physical infiltration), and their data can be used to plan and prioritize testing activities.

## REPORTING AND FEEDBACK

When including IoT data and analysis of IoT devices and networks in intelligence products, there are a few key facts to communicate. Consumers of cyber intelligence might not be accustomed to data from IoT, which can differ from common threat and network intelligence. Consequently, it might be important to do the following:

- Identify any information that came from IoT devices, especially if you suddenly start including information about the real world (telemetry, video, audio, geolocation, etc.).
- Identify your confidence in the data of those devices and the devices themselves.
- Identify how the IoT data supports overall intelligence assessment. That is, explain how IoT data can provide a more complete, nuanced, or reliable assessment.

Also, communicate with stakeholders and leadership about which additional IoT data might support your analyses in the future. There might already be IoT devices in use that you can use or query. Or, you might need to acquire or implement new devices to achieve the visibility you desire.

# IOT CASE STUDY

The following is a fictional scenario, using a fictional organization, to illustrate how IoT can support cyber intelligence.

A2M2 Inc. is a large electronics retailer that sells products both online and in stores. Its products, and products from other manufacturers, are assembled at a facility in Asia. Many products are integrated with a cloud-based service to provide customers with software updates, cloud storage, and other services. A2M2 has strong vertical integration with its logistics providers and supply chain, and runs its

own warehouses and distribution centers. It has been aggressively deploying technology, both in its stores and its backend facilities, to enable market research and just-in-time inventory control.

Recently, A2M2's Fraud department has noticed two anomalies: Its physical inventories don't match the records in its inventory management system, and more of its store-brand products seem to be popping up on auction sites. After some research by its IT department, A2M2 has discovered a possible breach in its inventory management and financial systems. It is concerned that inventory is systematically being stolen, but without being able to trust its compromised IT systems it cannot pinpoint where the theft might be taking place. It's possible that the IT systems were compromised primarily to cover up the theft, but A2M2 cannot identify the original attack vector. The company decides to look into the data available from its many IoT devices to see if it can determine whether systematic theft is taking place and, if so, where and when it's occurring.

In the stores, A2M2 wants to gather more information about people—customers, employees, and suppliers—such as when they are entering and exiting, where they go within the store, and how long they stay. The company upgrades its security cameras to count the number of people who move past. Sensitive to privacy concerns, A2M2 opts against facial recognition of customers but enables it in non-public areas. The company also gathers data from automatic door openers, motion detectors, and Bluetooth Low Energy devices that allow customers with smartphones to view additional data about products via an augmented reality app.

At the stores' shipping and receiving docks, the company gathers information from the telematics units in delivery trucks and the RFID tags used for inventory control. While most of A2M2's products are delivered to the stores from its regional warehouses via its own fleet, some products are delivered by other freight companies. They work with these companies to share data about routes, delivery times, and unloading times.

In the warehouses, the company starts collecting telemetry from the automated packing robots and sorting machines. At the assembly facility in Asia, it similarly instruments the industry control systems and process management tools. Now that A2M2 is collecting much more data about its operations, it needs to deploy tools that will analyze and correlate all that data. Its initial goals are a) to determine whether any of its IoT devices were used as an attack vector to pivot into the financial systems, and b) to get an accurate picture of the amount of inventory that is arriving and leaving the assembly plant, warehouses, and stores.

A2M2 uses a SIEM to aggregate the data pulled from its IoT devices. From there, analysts pivot between data sets, find anomalies, and identify trends in the data over time. Not only does this help A2M2 narrow the cause of this particular breach, the organization utilizes this technology to identify other intrusions, weak spots in its current architecture, and other inefficiencies. Ultimately, the data from IoT informs process improvement throughout the company.

Carnegie Mellon University
Software Engineering Institute

# Public Cyber Threat Frameworks and Cyber Intelligence

## An Implementation Guide

# Public Cyber Threat Frameworks and Cyber Intelligence: An Implementation Guide

**Authors**
Samuel J. Perl
Geoffrey Dobson
Geoffrey Sanders
Daniel Costa
Lawrence Rogers
Jared E. Ettinger

**Design**
David Biber
Alexandrea Van Deusen
Todd Loizes



This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.



DM19-0447

# Contents

# Public Cyber Threat Frameworks and Cyber Intelligence

## INTRODUCTION

Cyber threat frameworks (CTFs) aim to provide a structured representation of the common and typical behaviors exhibited by cyber threats. The main promise of a cyber threat framework is that an intelligence picture can be developed from empirical data and used for conversations with executives and third parties. CTFs aid decision makers because they provide input on both past and predicted adversarial behavior. Different CTFs provide different types of input to decision makers on the behavior of adversaries. Some frameworks provide general stages of probable cyber threat behavior, while others can describe specific techniques occurring in each stage. Both levels of detail prove useful depending on the decision maker's goals, and they also help the cyber intelligence team generate new requirements.

### KEY FINDINGS

1. People, process, and technology are foundational to successful CTF implementation.
2. There is no one-size-fits-all CTF that organizations can "set and forget." The frameworks complement each other and target different levels of detail. The Lockheed Martin Cyber Kill Chain model is a reasonable high-level assumption of adversarial behavior, but depending on the organization and its leadership it may not meet all of the detailed intelligence requirements. In such cases, organizations can use more detailed models, such as the Diamond Model and MITRE's ATT&CK framework, to collect input. Other alternatives are the Office of the Director of National Intelligence Cyber Threat Framework (ODNI CTF), which provides more complexity and intelligence focus than the Kill Chain. If an organization chooses the ODNI CTF, the National Security Agency Central Security Service (NSA/CSS) Technical Cyber Threat Framework can be used to collect more detailed input or to interoperate with input data in MITRE's ATT&CK framework.
3. Organizations that have prerequisites in place, including necessary support and a strong executive champion, should implement a CTF and begin using it to support cyber intelligence analysis.
4. Organizations that have implemented a CTF should consider what part of the information they collect can be shared with other defenders. The only way to systematically increase the cost of attacking in cyberspace is to share our collective defensive progress with each other. Adversaries are rarely specific to one organization, and sharing information about all observed activities may allow hidden connections between events to be discovered—especially if those events are at different organizations.

## PREREQUISITES FOR CYBER THREAT FRAMEWORKS

CTFs depend on the right combination of people, process, and technology for success. Organizations seeking to use CTFs should first consider whether they have high functioning capabilities in the key areas of people, process, and technology. Organizations should assess their own capabilities in these areas and should not assume that they are already sufficient to support CTFs.

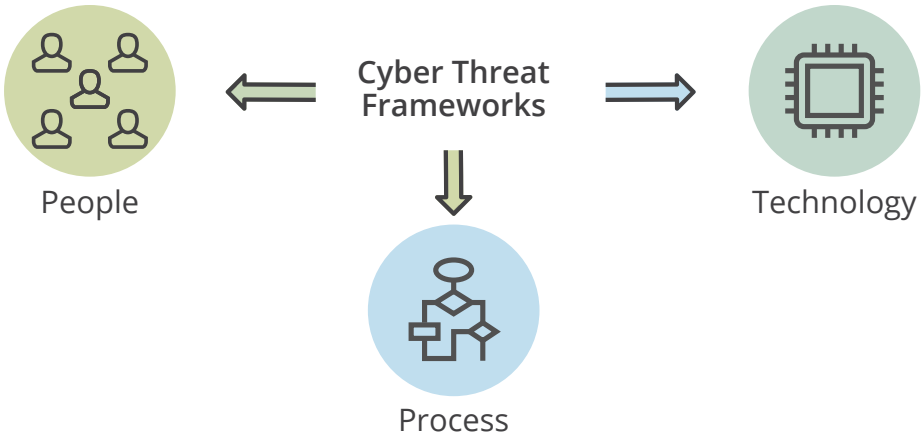



**Figure 1.** Cyber threat frameworks provide shared knowledge for people, process, and technology (the triad of information systems). They do this by embedding assumptions and knowledge about attacker behaviors, and provide analysis 'tooling' to help organizations collect, analyze, and share their data collected during contact with attackers.

**People, process, and technology are foundational to successful CTF implementation**

- *People, a foundational prerequisite.* People are the foundational prerequisite of CTFs. Skilled analysts combine critical thinking, technical expertise, and writing skills with CTFs to produce and convey intelligence to decision makers. Without capable people, organizations will struggle to use a CTF properly.
- *Process, which helps ensure repeatable outcomes.* Having strong processes for information security governance, network and host visibility, threat and indicator management, active defense, incident management, and situational awareness is a strong prerequisite to the successful deployment and use of CTFs in the organization.[1]
- *Technology, a time reducer and force multiplier.* Technology enables CTF implementation through people and process. CTF activities can require a lot of data and analysis, and automation can reduce the time it takes analysts to use a CTF.

1 Threat and indicator management provides formal processes to collect, measure, prioritize, and monitor threats over time and to manage the indicators used to identify and defend against them. Active defense assumes that defensible architecture, automation, and passive defense patterns reduce manual analysis, freeing analysts to actively identify, respond to, and learn from adversaries. Incident management processes codify the details of who within an organization will respond to threats and how they will respond. Situational awareness defines methods of aggregating organizational information and intelligence to provide a standard decision support view.

## CTFS AND THE CYBER INTELLIGENCE FRAMEWORK

One way to conceptualize CTFs is to characterize them relative to the components of the Cyber Intelligence Analytical Framework they align to. Each of the CTFs covered in this report can help complete some aspects of each component.

### ENVIRONMENTAL CONTEXT

CTFs can help organizations model and describe their environments, which can help scope a cyber intelligence function. They can also identify the data needed to perform cyber intelligence. Some CTFs can be used by analysts to help gain a holistic understanding of their organization's attack surface in relation to cyber threats. MITRE ATT&CK's domains and platforms provide a filter that can be used to enumerate the tactics, techniques, and procedures (TTPs) applicable to particular operating systems or platforms (mobile devices, for example). The Kill Chain provides progressive stages, which can be used as a model to help prioritize threats based on organizational impact.

CTFs can help organizations describe and model past incidents, current pertinent information, and potential future threats. They can also provide a common model between aspects of an organization's cyber intelligence functions and its insider threat detection, prevention, and response programs.

### DATA GATHERING

Organizations can use CTFs to help align data sources to meet intelligence requirements. Frameworks that provide a knowledge base of attacker TTPs, such as MITRE ATT&CK and the NSA CSS TCTF, can help organizations identify what data they need to collect and analyze to detect certain threat actor activity. Internal and external information sharing relationships can be facilitated by the CTFs that provide a controlled vocabulary, including the ODNI Cyber Threat Framework, MITRE ATT&CK, and the NSA CSS TCTF. Many homegrown and off-the-shelf tools for facilitating data collection for cyber intelligence analysis—security information and event management (SIEM) tools in particular—make use of CTFs as logical models that help organize and categorize collected data.

All of the CTFs can play a part in helping an organization establish and maintain a repeatable cyber intelligence workflow that can consider past, present, and future data regarding cyber threats

### THREAT ANALYSIS

CTFs can be used to help organizations analyze the technical complexities and characteristics associated with threats, incidents, and events. The ODNI CTF, MITRE ATT&CK, and the NSA CSS TCTF can all be used to help analysts model and describe the what, when, where, and how of activity associated with cyber threats. The consistency, accuracy, and timeliness of the more technical analysis can be aided by using these CTFs. Frameworks with enumerations of techniques, such as MITRE ATT&CK and the NSA CSS TCTF, can help inform what technical disciplines, expertise, and core competencies are needed to produce threat analysis reports. Many threat intelligence platforms also facilitate threat analysis by allowing an analyst to tag or categorize indicators based on the phase of the Kill Chain or MITRE ATT&CK they are associated with.

## STRATEGIC ANALYSIS

Strategic analysis is the process of producing a holistic assessment of threats, risks, and opportunities to enhance executive decision making pertaining to organization-wide vital interests, such as finances health, brand, stature, and reputation. More specifically, one might also perform strategic cyber intelligence analysis to provide deep clarity on the who and why behind threats and threat actors. Strategic analysis goes beyond threat analysis to incorporate analysis about emerging technologies and geopolitics that may impact and/or provide opportunities for the organization, now and in the future. The Diamond Model supports attribution of a particular set of actions to a threat actor, and both MITRE ATT&CK and the NSA CSS TCTF support deducing the intent of a particular action. Higher-level CTFs, such as the ODNI CTF, Diamond Model, and Lockheed Martin Cyber Kill Chain, can all be used as tools that can enhance an organization's ability to repeatably produce strategic analysis.

## REPORTING AND FEEDBACK

The outputs produced by a cyber intelligence analyst can also benefit from the use of CTFs. By providing standardized models and controlled vocabularies, CTFs enable organizations to consistently use a common lexicon with consumers. Consistently using CTFs to describe, organize, and share aspects of cyber intelligence analysis can increase the quality and timeliness of the organization's intelligence reporting. Consumers providing feedback can also indicate which parts of the intelligence were useful and in what context. CTFs that contain a knowledge base of recommended mitigation strategies associated with a particular activity can help organizations produce actionable intelligence. CTFs that model incident progression, such as the Kill Chain, can be used to help measure the impact, severity, or loss associated with specific threat activity. Both types of information can also be used by decision makers to help reduce future exposure to cyber risks.

## REVIEW OF SURVEY AND INTERVIEW RESULTS

The Cyber Intelligence Tradecraft Survey asked respondents whether and how they used CTFs within their cyber intelligence processes. Of the survey respondents, 22 of 31 (71%) claimed to use at least one of the CTFs. The most commonly used CTF among survey respondents was the Kill Chain (17 of 31, or 55%), followed by ATT&CK (14 of 31, or 45%), then the Diamond Model (5 of 31, or 16%), and finally the ODNI CTF (1 of 31, or 3%). Survey respondents listed threat prioritization, internal information sharing, and threat actor attribution as their use cases for CTFs within their cyber intelligence processes.

## CTF COMPARISON

In this study, we are considering, comparing, and contrasting the five most prevalent cyber threat frameworks: The ODNI Cyber Threat Framework, the Kill Chain, the MITRE Pre-ATT&CK/ATT&CK knowledge bases, the Diamond Model, and the National Security Agency's Technical Cyber Threat Framework (NSA TCTF).[2] As we interviewed organizations, contacted experts, reviewed literature, and tested each of the frameworks, some clear differences emerged. To begin, consider the following table depicting each of framework's stated goals:

| ODNI | Kill Chain | Diamond | MITRE ATT&CK | NSA TCTF |
|---|---|---|---|---|
| The Cyber Threat Framework was developed by the U.S. government to enable consistent characterization and categorization of cyber threat events and to identify trends or changes in the activities of cyber adversaries.[3] | Institutionalization of this approach reduces the likelihood of adversary success, informs network defense investment and resource prioritization, and yields relevant metrics of performance and effectiveness.[4] | The model establishes the basic atomic element of any intrusion activity, the event, composed of four core features: adversary, infrastructure, capability, and victim.[5] | ATT&CK is useful for understanding security risk against known adversary behavior, for planning security improvements, and verifying defenses work as expected.[6] | This framework was designed to help NSA characterize and categorize adversary activity by using a common technical lexicon that is system agnostic and closely aligned with industry definitions.[7] |

**Table 1.** A comparison of the stated focus areas of the five CTFs in our analysis.

When analyzing the frameworks' goals, it's clear that the NSA TCTF and the ODNI CTF are similar: they both aim to "characterize and categorize" so that disparate organizations can describe threat activities through a common lexicon. The remaining three frameworks seek to technically understand adversary behavior, but from different angles. The Lockheed Martin Cyber Kill Chain seeks to track adversary movement, the Diamond Model seeks to correlate threat data into cohesive events, and the MITRE ATT&CK model seeks to define and predict specific behaviors. The following table lays out the major differences our analysis uncovered among the cyber threat frameworks.

| CTF | Example Artifacts | Organizational Maturity Level | Target Audience |
|---|---|---|---|
| ODNI | Threat glossary | High | Intel Analyst |
| Kill Chain | Enterprise security plan | Low | Leadership |
| ATT&CK | Sensor signatures | Medium | Engineer |
| Diamond | Machine learning features | High | Researcher |
| NSA TCTF | Adversary activities glossary | Medium | Security Analyst |

**Table 2.** Main differences among cyber threat frameworks.

2 We are considering PRE-ATT&CK and ATT&CK as one framework.

3 https://www.dni.gov/index.php/cyber-threat-framework

4 https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/LM-White-Paper-Intel-Driven-Defense.pdf

5 http://www.activeresponse.org/wp-content/uploads/2013/07/diamond.pdf

6 https://attack.mitre.org/

7 https://apps.nsa.gov/iaarchive/library/reports/nsa-css-technical-cyber-threat-framework-v1.cfm

## COMMUNICATING TO OTHERS USING CYBER THREAT FRAMEWORKS

A key difference among all frameworks is the audience each targets: the ODNI Cyber Threat Framework is primarily intended for intelligence analysts, the Lockheed Martin Kill Cyber Chain is geared towards corporate leadership, the Diamond Model aligns with research queries, the MITRE ATT&CK library informs cyber security engineers, and the NSA Technical Cyber Threat Framework speaks to security analysts. Another key difference lies in the types of artifacts each framework will produce. Both the ODNI CTF and NSA TCTF will primarily produce communication documents that analysts can use to translate the widely varying technical activities into common attributes, thereby making cross-organizational communication much more efficient. Organizations using the Kill Chain can provide leadership decision-making aids by converting low-level intrusion detection and prevention systems into nontechnical and enterprise-level risk-based concerns. The outputs of the MITRE ATT&CK methods, such as sensor signatures and anomalous traffic patterns, exist at a more technical level and generally target an engineering audience. Lastly, the Diamond Model, which targets more research-focused personnel, can be used to correlate adversary intents with infrastructure vulnerabilities in machine learning applications.

To create the word clouds for each of the frameworks below, we collected the words from two articles by authors attempting to use or describe the frameworks. These words clearly illustrate the characterization arrived at by our analysts. That is,  the ODNI CTF is a threat-based framework, the Kill Chain models cyber security in general, the Diamond Model focuses on the adversary, and ATT&CK informs detection activities.[8]

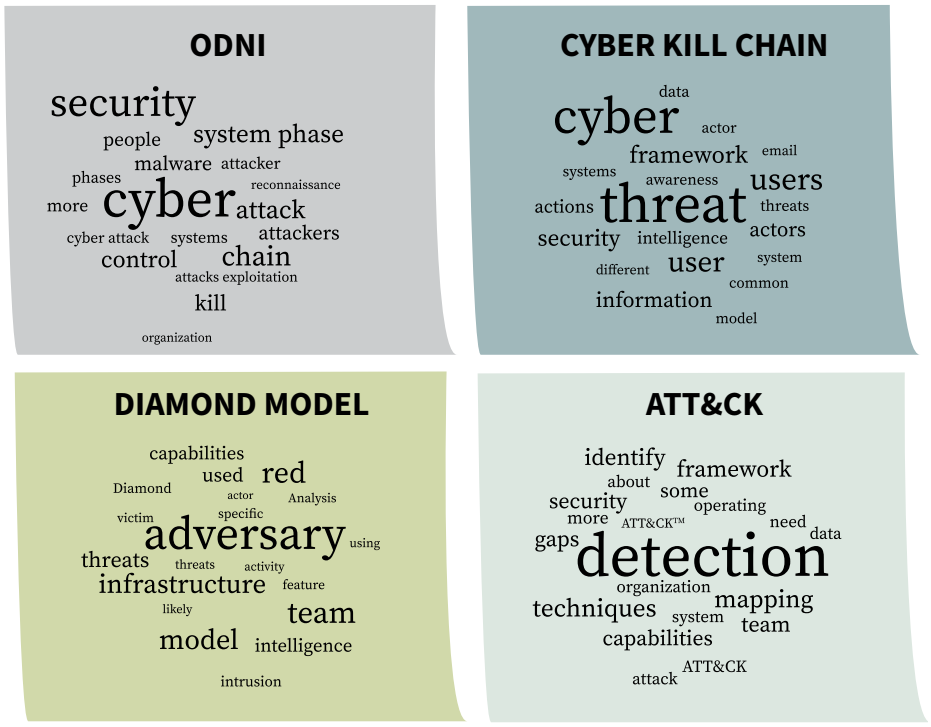



**Table 3.** Word-cloud graphics characterizing each CTF.

8 ODNI CTF: https://www.afcea.org/content/creating-common-language-cybersecurity and https://smallwarsjournal.com/jrnl/art/odni-common-cyber-threat-framework-new-model-improves-understanding-and-communication
Kill Chain: https://www.sans.org/security-awareness-training/blog/applying-security-awareness-cyber-kill-chain and http://techgenix.com/cyber-kill-chain/
Diamond Model: recordedfuture.com/diamond-model-intrusion-analysis/ and sixdub.net/?p=762
ATT&CK: tanium.com/blog/getting-started-with-the-mitre-attack-framework-improving-detection-capabilities/ and tanium.com/blog/getting-started-with-the-mitre-att-and-ck-framework-lessons-learned

## CTF CONSIDERATIONS

Organizations should consider implied assumptions, limitations, and overlaps of CTF models when adopting them as part of a cyber intelligence program. Not accounting for these considerations can negatively impact the success of a highly capable cyber intelligence team.

**INSIDER TIP**

Before implementing a framework, organizations should verify that they

- understand and manage CTF model assumptions
- reduce areas of model overlap
- account for model limitations
- adjust framework implementation plans accordingly

### ASSUMPTIONS

Each framework contains assumptions about organizations attempting to use it. These assumptions are not specifically listed in the frameworks, but they can affect the success of an implementation. The following list presents our opinion of the most important assumptions organizations should be aware of. They are ordered by their importance for determining an organization's successful implementation and usage of a CTF (see Figure 3 below).

**Assumption 1: Stakeholders support an organizational cyber intelligence function.**
Executives from different organizational units, such as business, information technology, and information security, can have a profound impact on framework implementation. Organizational buy-in and positive stakeholder relationships are key components that enable cyber intelligence analysts to execute strategic, operational, and tactical intelligence functions. CTFs require input and feedback from multiple areas of an organization for success. For example, executives provide valuable input to cyber intelligence teams that are trying to understanding the organization's specific threat landscape. Cyber intelligence teams leverage information from business executives to develop focused intelligence on which specific threats may impact the organization. Without this unique organizational knowledge, significant gaps may appear in areas such as intelligence collection or priority intelligence requirements (PIRs). For example, PIRs could be something a line of business might send to a cyber intelligence team to understand how threat could change due to a future sale or acquisition.

**Assumption 2: A specific infrastructure exists to support cyber intelligence.**
Another common assumption of CTFs is that an organization has the infrastructure, such as architecture, devices, logs, and applications, that provides cyber intelligence analysts the ability to collect data, test hypotheses, and produce intelligence. Without appropriate infrastructure, analysts cannot use CTFs to support the cyber intelligence analytic framework. They cannot collect important data

and are thus unable to build an appropriate analytical assessment. Organizations should consider implementing automated indicator collection, reduction, and prioritization technologies to free up analysts and reduce human error.

**Assumption 3: Cybersecurity analysts and cyber intelligence analysts need to consistently work together and support each other to get the most value out of the CTF they are using.**
Cyber intelligence teams develop intelligence used by cybersecurity analysis teams to manage the security infrastructure and conduct security operations (including incident handling). This cooperative relationship is foundational to intelligence production because, as cybersecurity teams consume intelligence and provide feedback, additional data is collected to continue intelligence production. If security operations teams are not mature enough, the benefits of using frameworks may not be realized. For example, if security operations cannot effectively use cyber intelligence to identify threats and new indicators, frameworks cannot be used to develop a larger intelligence picture. Therefore, both security operations and cyber intelligence should work together to address threats by using and producing intelligence as the organization addresses threats.

**Assumption 4: A cyber intelligence function is established and operating.**
CTFs assume that a cyber intelligence team is established and operating, including appropriate funding and stakeholder support.

## LIMITATIONS

All frameworks focus on adversary intrusions and do not account for other operating environment attributes, such as friendly and neutral forces, culture, natural disasters, or the marketplace. These considerations would require framework expansion or new model development.

Organizations should also consider the legal requirements of any security actions. Offensive cyber operations, such as attack and exploitation, are limited to law enforcement or national defense entities. Therefore, if frameworks generate intelligence that falls outside of the traditional organizational defense scope, it may need to be transitioned to the appropriate authorities.

Some frameworks do not address all intelligence levels (strategic, operational, and tactical) and can also introduce bias. For example, the MITRE ATT&ACK framework "can be used to better characterize and

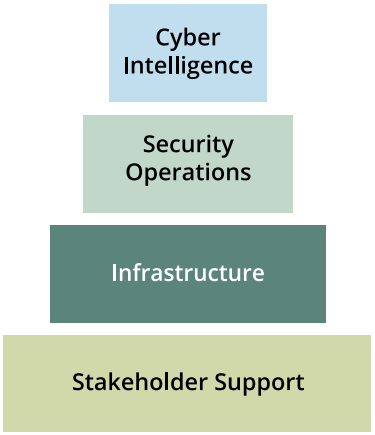



**Figure 3: The relationship and hierarchy of CTF assumptions.**
Cyber threat frameworks assume that organizations have
1. organizational stakeholder support
2. infrastructure for cyber intelligence
3. a mature cybersecurity team
4. a defined cyber intelligence team

These assumptions have direct relationships to one another. For example, implementing cyber intelligence infrastructure depends on stakeholder buy-in for funding, while cyber intelligence couldn't function without a cybersecurity team to consume intelligence and produce data for intelligence production.

describe post-compromise adversary behavior."[9] Therefore, strategic analysis focuses contained in other CTFs (such as attacker intentions, capabilities, and attribution) are not specifically addressed by ATT&CK. Additionally, MITRE ATT&CK is known to introduce specific bias where the model schema does not enforce reporting attributes of aggregate counts.[10]

## OVERLAPS

The evolutionary nature of CTFs has led to overlaps among the models. These overlaps should be reviewed by organizations prior to implementation. For example, the Kill Chain was one of the earliest models to formally define adversary intrusion operations.[11] Since then, other models have adopted, enhanced, and expanded Kill Chain concepts. As the Kill Chain was implemented, additional needs were identified, resulting in new and different models. One example, the Diamond Model of Intrusion Analysis, integrates Kill Chain concepts, but adds granularity, complex relationships, and formal mathematical methods.[12] In contrast, the ODNI Cyber Threat Framework overlaps frameworks for a different purpose.[13] As organizations adopted various CTFs to support cybersecurity operations, it was difficult to label and aggregate the collected information. The ODNI CTF addressed this challenge by creating a common framework language to simplify metrics, reporting, and situational awareness.

CTF overlaps are not always equal in comparison, however. The Kill Chain and Diamond Model both discuss correlation of indicators for intrusions and campaigns. While the Kill Chain paper discusses correlation concepts, the Diamond Model provides discrete attributes, formulas, and graphs for correlation. Another example of overlap difference is the MITRE ATT&CK framework.[14] It focuses on post-compromise sections of the Kill Chain and enumerates attacker TTPs that are not detailed in the Kill Chain or Diamon Model.

Overlaps of CTF models should be compared and contrasted to determine how they specifically contribute to cyber intelligence operations. Organizations should determine their current CTF coverage and which areas of specific CTFs best fit their needs.

## TAKEAWAYS

CTF implementation plans should include steps to address these considerations. We recommend organizations meet the minimum assumptions listed above to successfully implement frameworks. To overcome limitations, we also recommend identifying intelligence gaps or possible bias introduced by a specific framework. If gaps or biases do not impact intelligence or operations, they should be documented and periodically reviewed to verify that their status does not change. Limitations that impact intelligence analysis can be addressed by combining frameworks or extending them to meet organizational needs. If adopting multiple frameworks, organizations should analyze overlaps to determine which frameworks provide the best features for the overlap area. For example,

---

9 MITRE Corporation. "ATT&CK for Enterprise." attack.mitre.org/resources/enterprise-introduction

10 twitter.com/MITREattack/status/1026532833018478593

11 Lockheed Martin. "The Cyber Kill Chain," 2019. lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

12 threatconnect.com/wp-content/uploads/ThreatConnect-The-Diamond-Model-of-Intrusion-Analysis.pdf

13 Office of the Director of National Intelligence. "Cyber Threat Framework Frequently Asked Questions." dni.gov/files/ODNI/documents/features/Cyber_Threat_Framework_Frequently_Asked_Questions_20180718.pdf

14 MITRE Corporation. "ATT&CK for Enterprise." https://attack.mitre.org/resources/enterprise-introduction/

organizations combining the Kill Chain and Diamond Model may determine that the Kill Chain addresses the intrusion lifecycle and courses of action, while the Diamond Model is best suited for intrusion analysis. Whatever circumstances may surround the consideration and implementation of frameworks, they provide structure and formal models to mature and improve threat analysis.

## APPLYING A CYBER THREAT FRAMEWORK TO A REALISTIC SCENARIO

In this section, we apply the frameworks previously described in this guide to a real-world scenario. By laying out a complex attack, step by step, and identifying the underlying layers of each CTF that can be applied, we can gain insights about their practicality. We began by devising a scenario in which a threat actor selected a cyberspace target on a corporate network rich in personally identifiable information (PII) that the attacker believed could be stolen and monetized. In the upper portion of Figure 4, we identify each step taken by the attacker in this scenario. Then, for each step of the attack, we identify the underlying layer of the CTF that can be used to apply best practices for mitigating the threat.

The scenario: A threat actor exploits vulnerabilities in Internet-facing services to gain access to the victim's network. Next, the actor moves laterally to gain access to the servers on which the PII is located. The actor also takes the necessary precautions to hide all actions, where possible, in an attempt to reduce the likelihood of detection. Finally, the threat actor installs the tools needed to maintain continued access and to continue to pilfer additional PII. All the steps of this scenario, from 1 to 18, are presented in Figure 4.

We walked through the scenario, step by step, and identified the part of each CTF that could be applied, from an organizational security perspective, to identifying ongoing attacks. In each case, matching the attacker step with the CTF layer provides insight into the actor's intentions and likely next actions. This matching also provides clues about how deeply the actor has likely penetrated the victim's network. This can be especially helpful in marshalling resources to counter the actor at a technical level while communicating to the appropriate parties the status of the attack and the recommended actions to mitigate the threat.

Figure 4 also shows the attack scenario from the incident handling perspective, juxtaposed against the threat actor timeline. The incident handling team must act on the information it gathers, in the order received, and then make decisions with incomplete information. As shown in Figure 4, the first indicator the team encounters is unusual network traffic. This information, while vital for cyber situational awareness, is actually an artifact from the attacker's step 10. Digging deeper, the next indicators the team discovers are anomalous web server log files, which correspond to attacker step five. Ultimately, organizational leadership must optimize the application of resources to both securing cyber assets and responding to live incidents. The CTFs can be utilized by breaking comprehensive security into manageable parts and then guiding the organization to industry best practices for each part.

| Scenario Step and Title | ODNI | Lockheed Martin Cyber Kill Chain | Mitre Pre-ATT&CK and ATT&CK | NTCTF (NSA) |
|---|---|---|---|---|
| 1 Ideation | Preparation | does not address ideation | Priority Definition Planning: Conduct Cost/Benefit Analysis | Administer |
| 2 General Reconnaissance | | Reconnaissance | Target Selection: Determine Operational Element | Preparation |
| 3 Victim Identified | | | Target Selection: Determine Strategic Element | |
| 4 Victim Infrastructure Reconnaissance | | | Technical Information Gathering | |
| 5 Web Server Vulnerability Discovered | | | Technical Weakness ID | |
| 6 Prepare Web Server Attack | | Weaponization | Build, Test, Stage Capabilities | |
| 7 Attack Web Server | Engagement | Delivery Exploitation Installation | Initial Access: Exploit Public-Facing Application | Engagement |
| 8 Establish Interactive Access to Web Server | | Command and Control | Execution: Command-Line Interface | |
| 9 Clean Up Web Server attack | | Actions on Objectives | Defense Evasion | Presence |
| 10 Establish Continued Access to Web Server | Presence | | Persistence | |
| 11 Victim Network Reconnaissance | | Reconnaissance | Discovery | |
| 12 Next Target Identified | | | | |
| 13 SQL Database Vulnerability | | | | |
| 14 Prepare SQL Injection Attack | Effort/ Consequence | Weaponization | | |
| 15 Attack SQL Database Server | | Delivery Exploitation Installation | Lateral Movement: Exploitation of Remote Services | Effect |
| 16 Harvest PII from SQL Database | | Actions on Objectives | Collection: Data from Local System | |
| 17 Exfiltrate Harvested PII | | | Exfiltration | |
| 18 Clean Up Harvesting | | | Defense Evasion | |

| Incident Handling Timeline | Adversary Timeline | Indicator |
|---|---|---|
| 1 | 10 | Unusual Network Traffic |
| 2 | 5 | Web Server Log Files |
| 3 | 11,12 | Historical Flow Data |
| 4 | 15, 16 | SQL Server Log Files |
| 5 | 17 | Web Server Log Files |

**Figure 4.** Network intrusion scenario in which an attacker attempts to steal personally identifiable information.

## KEY TAKEAWAYS

1. When applied to an intrusion, ideation (defined as the formation of an idea or concept) describes the threat actor's motivation for an attack. It is unlikely that an actor arbitrarily selects a victim and the date and time of attack. One important goal during incident response is to understand why the attacker specifically selected the victim. Of the cyber threat frameworks that were analyzed, the Lockheed Martin Cyber Kill Chain does not address this step.

2. Incident handlers and system administrators can be become distracted from the task of identifying and stopping a threat actor by focusing on remediating the technology exploited by that actor. As the juxtaposition of the timelines shows, this distraction pointed their focus in the wrong direction with respect to attempting to defeat the actor.

3. Most traditional cybersecurity defense practices assume that attackers are not targeting the organization specifically. While this is still true for many attacks, more organizations than ever before are being targeted for specific reasons.[15] The CTFs provide analysts and managers a range of alternative defensive considerations for predicting, defending against, and possibly preventing future adversary behaviors.

15 Targeting U.S. Technologies. A Trend Analysis of Cleared Industry Reporting. By The US Defense Security Service Coordinated with: AFOSI, MCIA, and NCSC (9/7/2017) https://www.dss.mil/Portals/69/documents/ci/2017_CI_Trends_Report.pdf [accessed March 13, 2019]

## CONCLUSION

Capability in cyber talent, processes, and technology is a prerequisite for the efficient use of CTFs. Different CTFs appeal to different audiences. Some are intended to help communicate with other cyber intelligence and security analysts while others target nontechnical audiences. Frameworks that have been around longer, such as the Lockheed Martin Cyber Kill Chain, are used the most, but use of some newcomers, such as the Diamond Model and MITRE ATT&CK, is growing quickly. Organizations should consider the implied assumptions, limitations, and overlaps of each CTF when adopting it as part of a cyber intelligence program—not accounting for these considerations can negatively impact the success of a highly-capable cyber intelligence team. Cybersecurity is a focus and limitation of all frameworks and should be considered before implementation. The evolutionary nature of CTFs has led to overlaps between the models. Organizations may often use multiple models for a more complete analysis. Use of one or more CTFs can help provide input to every part of the cyber intelligence analytical framework.

## About the SEI

The Software Engineering Institute is a federally funded research and development center (FFRDC) that works with defense and government organizations, industry, and academia to advance the state of sthe art in software engineering and cybersecurity to benefit public interest. Part of Carnegie Mellon University, the SEI is a national resource in pioneering emerging technologies, cybersecurity, software acquisition, and software lifecycle assurance.

## Contact Us

CARNEGIE MELLON UNIVERSITY
SOFTWARE ENGINEERING INSTITUTE
4500 FIFTH AVENUE; PITTSBURGH, PA 15213-2612

sei.cmu.edu
412.268.5800 | 888.201.4479
info@sei.cmu.edu