National Cybersecurity Protection System Cloud Interface Reference Architecture

Volume One: General Guidance

May 14, 2021

Version 1.4

Cybersecurity and Infrastructure Security Agency
Cybersecurity Division

Capability Delivery Subdivision
NCPS Program Management Office

Revision/Change Record

Version	Date	Revision/Change Description	Section/Pages Affected
Version 1.0	12/12/2019	Initial Release Version	All
Version 1.1	4/17/2020	Response to Comments and Feedback	Added new Section 3; moved old Section 3 to Section 4 and revised content; updated Sections 1.2, 1.4, 1.5, 2.2, Appendix A; added Figures 3-9; minor graphic and text revisions throughout.
Version 1.2	7/24/2020	Response to Comments and Feedback	Added Appendix C; updated Executive Summary and Conclusion; minor graphic and text revisions throughout.
Version 1.3	10/16/2020	Alignment with Volume Two	Added new Section 3; moved old Section 3 to new Section 4.
Version 1.4	5/14/2021	Additional alignment with Volume Two	Restructured Volumes One and Two based on feedback from stakeholders; incorporated data from pilots to update options for Telemetry Type and Data Transformation Attributes; added new Section 5, in part from Volume Two draft; moved old Section 5 to new Section 6; added Appendixes D-I, in part from Volume Two draft; added Figure 13; minor graphic revisions throughout.

EXECUTIVE SUMMARY

The National Cybersecurity Protection System (NCPS) Program is evolving to ensure that security information about cloud-based traffic can be captured and analyzed and Cybersecurity and Infrastructure Security Agency (CISA) analysts can continue to provide situational awareness and support to the agencies. To support this goal, CISA is developing a cloud-based architecture to collect and analyze agency cloud security data. This reference architecture explains how agencies can interact with that system. It includes background about how the cloud impacts NCPS, discusses what security information needs to be captured in the cloud and how it can be captured, and provides reporting patterns to explain how that information can be sent to CISA.

The *NCPS Cloud Interface Reference Architecture* (NCIRA) will be released as two individual volumes. This first volume provides an overview of changes to NCPS to accommodate the collection of relevant data from agencies' cloud environments and provides general reporting patterns for sending cloud telemetry to CISA. The second volume provides an index of common reporting patterns and considerations for how agencies can send cloud-specific data to the NCPS cloud-based architecture. Individual cloud service providers (CSPs) can use Volumes One and Two to offer guidance on vendor solutions that align with these reporting patterns.

Reference Architecture is being released to Federal Civilian Agencies in advance of a production system to accomplish the following:

- Notify agencies about changes in the NCPS Program and give them time to plan.
- Solicit feedback from agencies so that a final version of this reference architecture provides desired content and meets the needs of agencies.
- Gather requirements from agencies to ensure the cloud-based NCPS architecture can support agency use cases.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND DEFINITIONS

Abbreviation	Definition
AI	Artificial Intelligence
API	Application Programming Interface
AWS	Amazon Web Services
C2	Command & Control
CASB	Cloud Access Security Broker
CEF	Common Event Format
CIO	Chief Information Officer
CISA	Cybersecurity and Infrastructure Security Agency
CISO	Chief Information Security Officer
CLAW	Cloud Log Aggregation Warehouse
CSP	Cloud Service Provider
DDOS	Distributed Denial of Service
DGA	Domain Generation Algorithms
DHS	Department of Homeland Security
DNS	Domain Name System
FedCIRC	Federal Computer Incident Response Capability
FIPS	Federal Information Processing Standards
FISMA	Federal Information Security Management Act
FOUO	For Official Use Only
GCP	Google Cloud Platform
GMT	Greenwich Mean Time
HIDS	Host-Based Intrusion Detection System
HR	Human Resources
HTTP	Hypertext Transfer Protocol
IaaS	Infrastructure as a Service
IAM	Identity and Access Management
ICRF	International Celestial Reference Frame
IDS	Intrusion Detection System
IERS	International Earth Rotation and Reference Systems
IOC	Indicators of Compromise
IP	Internet Protocol
IPFIX	Internet Protocol Flow Information Export
IPS	Intrusion Prevention System
IT	Information Technology
JSON	JavaScript Object Notation
LEEF	Log Event Extended Format
ML	Machine Learning
MOU	Memorandum of Understanding
MTIPS	Managed Trusted Internet Protocol Services
NAT	Network Address Translation
NCIRA	NCPS Cloud Interface Reference Architecture

NCPS	National Cybersecurity Protection System
NIST	National Institute of Standards and Technology
NOC	Network Operations Center
OMB	Office of Management and Budget
PaaS	Platform as a Service
PIA	Privacy Impact Assessments
PII	Personally Identifiable Information
PKI	Public Key Infrastructure
SaaS	Software as a Service
SECaaS	Security as a Service
SIEM	Security Information and Event Management
SOAR	Security Orchestration, Automation and Response
SOC	Security Operations Center
TIC	Trusted Internet Connections
TLS	Transport Layer Security
URL	Uniform Resource Locator
UT	Universal Time
UTC	Coordinated Universal Time
VM	Virtual Machine
VPC	AWS Virtual Private Cloud
VPN	Virtual Private Network

1 INTRODUCTION

Federal civilian departments and agencies[1] are required to meet the requirements of the National Cybersecurity Protection System (NCPS).[2] In general, this means that the Cybersecurity and Infrastructure Security Agency[3] (CISA) monitors the flow of agency network traffic and network flow logs are forwarded to CISA. CISA analysts use this data for 24/7 situational awareness, analysis, and incident response. Traditionally, network flow data has been collected by NCPS sensors located at Trusted Internet Connections (TIC) and Managed Trusted Internet Protocol Services (MTIPS) gateways, which capture security information as traffic passes between the agency and the Internet. As agencies move their information technology (IT) infrastructure to the cloud, some of their network traffic no longer traverses traditional NCPS sensors, and security information about that traffic is no longer captured by NCPS.

The NCPS Program is evolving to ensure that security information about cloud-based traffic can be captured and analyzed and CISA analysts can continue to provide situational awareness and support to the agencies. To support this goal, CISA is deploying a cloud-based architecture, the Cloud Log Aggregation Warehouse (CLAW), to collect and analyze agency cloud security data. This document, the *NCPS Cloud Interface Reference Architecture* (NCIRA) explains how agencies can provide cloud-generated security information to CLAW.

1.1 Document Organization

This document is structured to facilitate readability and ease of use by agencies. *NCPS Cloud Interface Reference Architecture: Volume One* consists of seven sections and nine appendices.

- Section 1 provides a document overview, assumptions, and constraints.
- Section 2 presents an overview of NCPS, describes how agency adoption of cloud computing impacts the program and introduces the NCPS cloud telemetry cycle.
- Section 3 describes cloud telemetry uses and related considerations for agencies.
- Section 4 expands on the NCPS cloud telemetry cycle by introducing a staged approach to cloud sensing, agency processing, and reporting to CISA.
- Section 5 describes cloud telemetry reporting pattern characteristics and their implications.
- Section 6 describes the cloud-based architecture that CISA is developing to collect and process NCPS-relevant data from cloud deployments of federal civilian agencies.
- Section 7 offers summary information.
- Appendix A discusses different types of cloud telemetry logs.
- Appendix B explores the various locations at which network flow information can be collected.
- Appendix C presents the implementation workflow for deploying NCPS in the cloud.
- Appendix D provides in-depth analysis of the Cloud Telemetry Timeliness characteristic.

[1] For the purposes of this document, the term "agency" will hereinafter be used to refer to all Federal Civilian Executive Branch departments and agencies.
[2] https://www.dhs.gov/cisa/national-cybersecurity-protection-system-ncps
[3] This document discusses programs (e.g., NCPS) that predate the creation of CISA. When discussing these programs, the term "CISA" refers to both the current agency and the predecessors that previously managed those programs.

- Appendix E provides in-depth analysis of the Cloud Telemetry Timing Coordination characteristic.
- Appendix F provides in-depth analysis of the Cloud Telemetry Provenance characteristic.
- Appendix G provides in-depth analysis of the Reporting Connection Administration characteristic.
- Appendix H provides in-depth analysis of the Cloud Telemetry Sharing Cost characteristic.
- Appendix I provides in-depth analysis of the Agency Data Retention and Use Constraints characteristic.

NCPS Cloud Interface Reference Architecture: Volume Two is a companion document that provides a catalog of the most common reporting patterns. These volumes can be used together to inform agency implementers on best practices and considerations for different deployment scenarios.

1.2 Purpose

A reference architecture is an authoritative source of information about a specific subject area that guides and constrains the instantiations of multiple architectures and solutions. The purpose of this reference architecture is to explain what information agencies need to capture in the cloud for NCPS, how that information can be captured, and how it can be sent to CISA. This reference architecture is divided into two volumes.

1. Volume One of the *NCPS Cloud Interface Reference Architecture* provides general guidance for agencies reporting cloud telemetry to CISA. The information provided includes the introduction of general reporting patterns. The discussion in Volume One is vendor-agnostic and not specific to any particular CSP.
2. Volume Two of the *NCPS Cloud Interface Reference Architecture* contains a catalog of reporting patterns for how agencies can participate in NCPS in the cloud under different cloud service models. The catalog includes individual reporting patterns (typical of an agency using a single CSP) as well as complex reporting patterns (illustrating how an agency can use several cloud service models and providers and send cloud security data to NCPS in the cloud).

1.3 Audience

This document is designed primarily for the federal civilian agencies, contractors, and vendors that are required to comply with the NCPS Program. This document can also be leveraged by stakeholders ranging from policy, acquisition, technical, and cybersecurity personnel to agency information technology leadership (e.g., Chief Information Officers (CIOs) and/or Chief Information Security Officers (CISOs)). Non-federal organizations may also derive value from this document as programs, strategies, and approaches are considered to address cloud security needs.

1.4 Assumptions

The following assumptions were used in the development of this reference architecture.

1. CISA will expand NCPS to include cloud data sources (rather than develop a new program to accommodate this new deployment model).
2. CISA will operate its own security telemetry collection infrastructure.

3. Agencies will continue to seek CISA assistance in securing their operations and data by participating in NCPS.
4. Cloud computing products and services will continue to evolve and expand and their adoption by Federal Civilian Executive Branch agencies will increase.
5. Federal cybersecurity policy will permit agency security data hosted on cloud services to be accessed directly by CISA (rather than through agency on-premise infrastructure).
6. Agencies are expanding the use of encryption for all types of data and encryption is expected to become increasingly common in the future.
7. CISA's initial telemetry requirements can be satisfied without payload decryption.

1.5 Constraints

The following constraints were used in the development of this reference architecture:

1. Agencies remain as data owners for all cloud telemetry and are merely sharing a copy of that data with CISA.
2. CISA makes efforts to reduce costs to agencies for sending cloud telemetry to CISA. However, agencies may still incur financial expense to fully participate in NCPS in the cloud. This occurs most naturally when an agency operates within one cloud service provider (CSP) and CISA's collection infrastructure resides in another.
3. CISA and agencies will have a written and signed memorandum of understanding (MOU) which governs the information sharing and handling relationship between both parties.
4. CISA information collection and use shall comply with public privacy impact assessments (PIA) for the NCPS Program.
5. Richness of telemetry shared with CISA is bound by the agency's encryption policy. If the agency does not perform encryption "break and inspect" functions, the agency and CISA will both be unable to observe traffic payload details.

1.6 FCISA Preferences

The CISA preferences listed throughout this document are starting points for the current state of cloud telemetry sharing maturity. These preferences will be revisited as implementation and testing results from the various cloud pilots are collected and examined. The preferences may require adjustments to coincide with technology advances and may eventually be codified into requirements.

2 BACKGROUND

NCPS is an integrated system-of-systems that delivers a range of capabilities, including intrusion detection, analytics, intrusion prevention, and information sharing capabilities that defend the civilian federal government's information technology infrastructure from cyber threats. The NCPS capabilities, operationally known as EINSTEIN, are one of several tools and capabilities that assist in federal network defense.

NCPS sensors are integrated into TIC access points. As such, agencies have traditionally been able to fulfill NCPS requirements simply by complying with the TIC Program. However, in 2019, Office of Management and Budget (OMB) issued an updated TIC policy, OMB Memorandum M-19-26[4], which does not require TIC access points to be embedded in all TIC use cases. Many of these new TIC use cases describe cloud services. In these use cases, network traffic between an agency and a CSP does not pass through an NCPS sensor.

As agencies and CISA adopt cloud environments and conform to the new TIC use cases, they will continue to share telemetry and security insights. This document provides guidance on how agencies can share telemetry with CISA and fulfill the requirements of NCPS when both agencies and CISA are operating in cloud environments. It furthers the NCPS objective to support "cyber" information sharing between CISA and federal agencies in order to enable a shared situational awareness between CISA and federal civilian networks. Under this platform, CISA and agencies gain increased security visibility and enhance existing incident response capabilities needed to tackle modern cyber threats on U.S. networks.

2.1 NCPS Overview

Traditionally, TIC access points (either MTIPS gateways[5] or agency-managed TIC Access Points[6]) contain EINSTEIN[7] sensors, so when an agency participates in the TIC Program, they also automatically utilize the capabilities of the NCPS Program. EINSTEIN 1 (E1) monitors the flow of network traffic (i.e., network flow records) to and from a Federal Civilian Executive Branch agency's on-premise networks. EINSTEIN 2 (E2) is an intrusion detection service that identifies potentially malicious network activity in Federal Government network traffic based on specific known signatures.[8]

Under the traditional (on-premise) TIC model, both E1 and E2 are deployed and screen all network traffic that is routed from an agency through TICs, MTIPS, and the EINSTEIN 3 Accelerated (E^3A) NEST[9] locations. For E1 and E2, the agency's telemetry, in the form of network traffic logs, are forwarded to CISA, and CISA analysts use these data for 24/7 situational awareness, analysis, and incident response. Hence, participation in TIC and ensuring all agency traffic from "inside" networks to "outside" systems traverses a TIC access point is all that is required to be in full compliance with NCPS demands for E1 and E2. The top data flow arrow in Figure 1 shows this traditional E1 and E2 telemetry pattern.

[4] https://www.whitehouse.gov/wp-content/uploads/2019/09/M-19-26.pdf
[5] https://www.dhs.gov/cisa/managed-trusted-internet-protocol-services
[6] https://www.dhs.gov/cisa/trusted-internet-connections
[7] https://www.dhs.gov/einstein
[8] https://www.dhs.gov/cisa/national-cybersecurity-protection-system-ncps
[9] https://www.gao.gov/assets/680/674829.pdf (page 48)

Figure 1: Current On-Premise Telemetry Configuration

Security insights are concrete intelligence data formulated for the timely identification and prevention of imminent cyber threats. Security insights may include security rules provided in a rule-based language (e.g., Snort[10] rules, Yara[11] rules, etc.), attack signatures (e.g., malware hash, malicious macros, etc.), and indicators (e.g., blacklisted IPs, email header indicators, etc.). Security insights are furnished by CISA and delivered to agencies to enable them to mitigate and counter cyber-attacks. Security insights may trigger internal processes and incident response within the agency's network to enact needed security reinforcements. Under the NCPS Program, security insights can also be provided in the form of a CISA security alert to an agency concerning detected suspicious activity on the agency's network. This CISA alert may include a mitigation recommendation from CISA analysts, which will trigger an agency workflow to remediate the security threat. The bottom data flow arrow in Figure 1 shows the flow of security insights from CISA to a TIC access point.

2.2 How Agency Cloud Adoption Impacts NCPS

As part of their IT modernization efforts, many agencies are utilizing commercial cloud products and adopting cloud email, collaboration, and software tools. Many agencies are using multiple CSPs in order to meet their mission needs and are utilizing all three cloud service models: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).[12] When an agency creates a tenancy within a CSP, traffic between that CSP and the agency may no longer pass through a TIC access point or an NCPS sensor.

Figure 2 shows the relationship between an agency's CSP tenancy and CISA. In this diagram, an agency still has some of its network traffic traversing the traditional TIC access point, but network traffic to or within one or more CSPs does not pass through the TIC access point. The top data flow paths show the traditional flow of E1/E2 telemetry from the agency to CISA and the flow of security insight from CISA to the agency. The bottom data flow paths show the new data flows between the agency, the CSP, and CISA. Reporting patterns for data flows and telemetry collection and sharing are discussed in Section 3 with more details and specific use cases provided in *NCPS Cloud Interface Reference Architecture: Volume Two*.

[10] https://www.snort.org
[11] https://yara.readthedocs.io/en/latest/
[12] Email as a Service (EaaS) is a sub-type of SaaS.

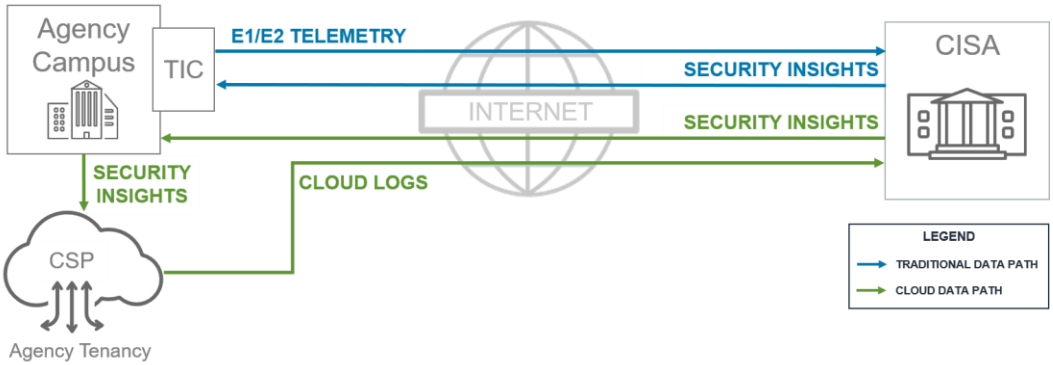

Figure 2: On-Premise and Cloud Telemetry Configuration

Because there are a wide range of CSPs and tenant-controlled security tools, there will be new data formats for telemetry (other than traditional network flows) and potential new formats for security insights for NCPS in the cloud. Data formats are discussed in Section 4.1.1 with more details.

NCPS Cloud Telemetry Cycle

In order to fully realize the collection of cloud data to fulfill NCPS requirements, there is a need to define the NCPS cloud telemetry cycle, as shown in Figure 3. Each of the entities in the cycle have unique roles and responsibilities.

- CISA sends intelligence and requirements to agencies (as depicted by the blue arrow).
- An agency is responsible for protecting its data, both on-premise and in the cloud, and the agency leverages intelligence and requirements to set configurations and indicators of compromise (IOCs) in its cloud instances (as depicted by the red arrow).
- CSP monitoring and policy enforcement agents generate logs and send them to CISA as cloud telemetry (as depicted by the black arrow).
- CISA uses the cloud logs to inform situational awareness and threat discovery, resulting in new intelligence sent to agencies (as depicted by the blue arrow).

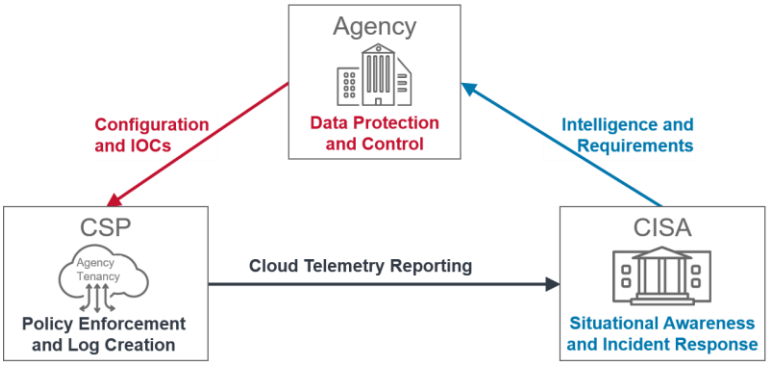

Figure 3: NCPS Cloud Telemetry Cycle

Benefits of Sharing Cloud Security Data With CISA

There are several benefits associated with sharing cloud security data with CISA.

1. Expanding NCPS to include agency cloud data provides CISA with the ability to gain situational awareness of threats and threat actors across the .gov domain, including on federal agencies' cloud communications. As a result, CISA can proactively respond to and mitigate cloud-based attacks against federal networks.
2. The inclusion of agency cloud telemetry extends CISA's security visibility and protection perimeter to include cloud-hosted software interactions and third-party services. This increased visibility informs and enhances incident response capabilities and federal cloud security posture. All agencies and CISA benefit from that extended visibility.
3. Additional cloud telemetry provides CISA with the ability to aggregate and correlate threat data generated and consumed in the cloud to aid in the timely discovery of security vulnerabilities and attack campaigns facing federal network cloud infrastructure.
4. Data gathered from the cloud network flow and cloud security logs provide CISA with additional intelligence and information to predict the changing security landscapes of both on-premise and cloud infrastructure, as well as to accurately plan, execute, and manage security countermeasures on the federal scale.
5. NCPS in the cloud provides a centralized model for log aggregation and analysis of a broad data set from federal cloud deployments, which result in a greater risk reduction for individual agencies as well as better availability of indicators of compromise to Federal Government information resources.

NCPS Roles, Responsibilities, and Cloud Operations

Transitioning to the cloud introduces new roles, actors, and procedures (e.g., an autonomous CSP, absence of TIC, third-party cloud monitoring tools, etc.) and the existing system for NCPS security insights transmission needs to be adapted. Specifically, in existing NCPS on-premise deployments, security insights in E2 are forwarded from CISA to the TIC access point (as shown in Figure 1). However, when agencies utilize CSPs there is the introduction of a new telemetry exchange. E2 security insights continue to be transmitted from CISA to the TIC access points (as shown in the Figure 2 blue data flow), but agencies also need to "pull" E2 security insights from CISA and transmit those security insights to their agency tenant protections hosted by CSPs (as shown in the green data flow path in Figure 2).

CISA's cloud presence for collecting and analyzing NCPS information is called CLAW. It is based on a functional, module-based architecture, hosted in multiple clouds, and ingests, stores, and analyzes cloud security logs and EINSTEIN sensor data from multiple agencies using commercial CSP services. It is geared towards enabling secure and efficient methods to process cloud data in a manner that offers CISA a similar level of situational awareness provided by current EINSTEIN on-premise deployments.

Figure 4 shows a more detailed analysis of the shifting relationships for NCPS implementation in the cloud. Roles that must be implemented or coordinated by more than one party are shown within the shared space of the overlapping ovals, with the participants identified. Traditional NCPS was almost entirely implemented by CISA, with the agency only playing a role in provisioning a network tap for CISA observation and use. This two-party interaction is shown with roles labeled "Traditional NCPS (TIC)." For cloud telemetry, the agency, CISA, and CSPs each have a responsibility to enable functionality. For information on the roles and responsibilities for implementing NCPS in the cloud, refer to Appendix C.

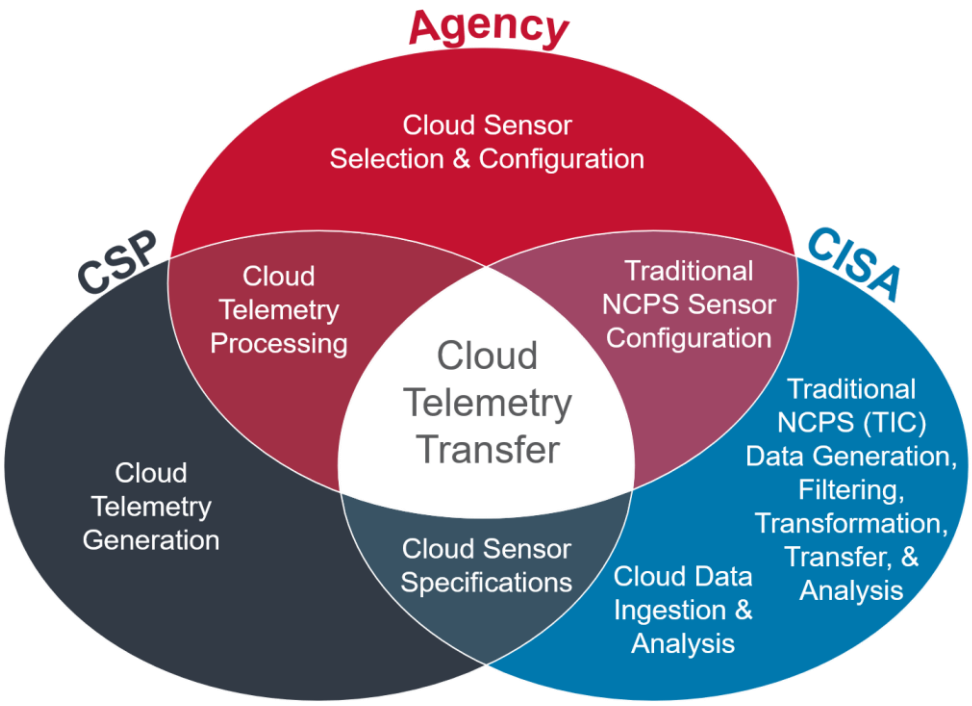

Figure 4: NCPS Roles and Responsibilities

3 CLOUD TELEMETRY USE AND REPORTING

As NCPS evolves to accommodate cloud services, agencies will be required to implement reporting patterns and maintain telemetry sharing with CISA. This section provides more details on cloud telemetry and reporting.

3.1 Cloud Telemetry Uses

Cloud service providers offer multiple mechanisms for their customers to gain insights into how their services are utilized. These insights support billing, system health monitoring, licensing compliance, auditing, security monitoring, user experience, and other functions. As cloud customers become more familiar with cloud services and how they are leveraged, CSPs also provide new and enhanced visibility into their services.

Cloud consumers seek to optimize different attributes of their cloud services. As a result, some visibility offered by the CSPs will be underutilized by some consumers and fully utilized by others. Agencies must therefore collect and analyze appropriate artifacts and observables from the full set of available CSP telemetry. CISA also has visibility preferences based on their stated missions and those preferences often align with the agency's visibility preferences. The relationship between these CSP telemetry sets is shown in Figure 5.

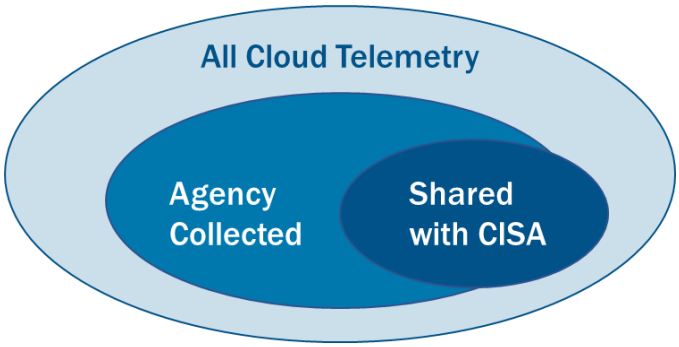

Figure 5: Cloud Telemetry Sets

CSPs offer multiple mechanisms to access this telemetry, with web consoles, APIs, log streaming, and management tool integration being the most prolific options. The cloud telemetry data obtained by each of these access methods may be different, so agencies may need to consider multiple mechanisms to ensure full visibility coverage.

3.2 Agency Cloud Telemetry Usage

For many agencies operating in the cloud, security information directed toward CISA is one of potentially many outputs derived from an integrated analysis solution. For agencies that have undertaken a significant investment in their cloud operations, it is typical to integrate security information with other operational information.

Generally, the integrated approach combines security-relevant logging information collected from multiple components within the CSP (and possibly across CSPs) responsible for cloud sensing of various cloud telemetry such as network flow, access/authentication logs, etc. This information can directly support reporting pattern Stage A (cloud sensing activities). Such information is often combined with operational telemetry, such as billing, performance, availability, and compliance data. Other

sources for analysis input include threat intelligence (TI) provided either by the CSP through its own relationships and subscriptions or directly by the tenant (e.g., CISA-provided security insights). In addition to these sources, an agency may incorporate information from other CSPs within which it operates (or coordinates), along with additional reporting from other infrastructure (e.g., in hybrid cloud cases this might involve on-premise sensing from vulnerability scans, active directory, firewall/intrusion detection system (IDS), application logs, endpoint/server logs, and security configuration management sources).

Aggregation of myriad data streams is generally accomplished with a centralized log aggregator and filtering system. The agency can apply artificial intelligence (AI) and machine learning (ML) techniques for heuristic-based anomaly detection, threat and advanced persistent threat detection, and risk and compliance assessment analysis. For reporting pattern Stage B, centralized aggregation and filtering is a function shared by CISA and the agency integrated analysis pipeline.

Outputs of the integrated analysis services support multiple uses. For presentation purposes, data visualization and dashboards can help agency analysts make sense of the information collected. For response functions, cloud-native security information and event management (SIEM) and automated response tools can act upon the integrated analysis results. Finally, reporting outputs can include reporting to CISA (as reporting pattern Stage C functions) as well as other external reporting destinations.

Controlling access is facilitated using features such as a CSP's identity and access control functions and policy implementation framework, which also provide useful logging information. These provide a baseline to help ensure the data reported is reliable. All major CSPs provide sophisticated identity and access management (IAM) systems and configuration management/verification systems. Once this is established, cloud-native SIEM, cloud access security broker (CASB), and (possibly) security orchestration, automation, and response (SOAR) capabilities can also have their telemetry integrated.

To illustrate this concept, Figure 6 shows how cloud telemetry reporting to CISA (denoted in red) fits within an agency's integrated analysis solution.

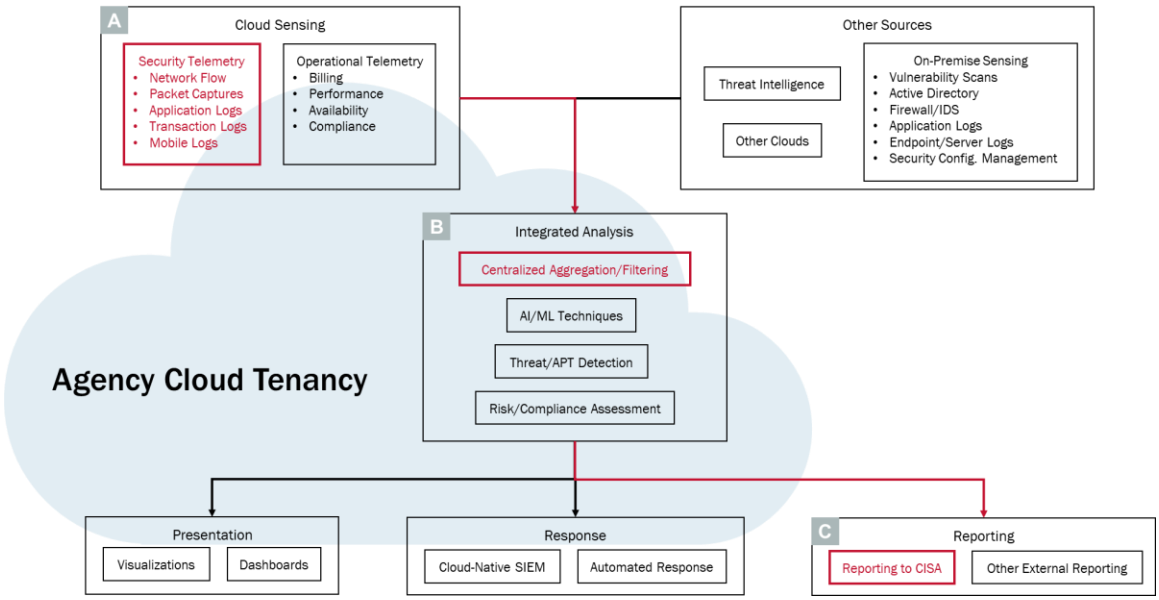

Figure 6: Agency Integrated Telemetry Solution Architecture

4 AGENCY REPORTING PATTERNS

The NCPS cloud telemetry cycle was introduced in Figure 3 to depict the relationship between an agency, CISA, an agency's authorized CSPs, and the information passed between parties. In this section, "cloud telemetry reporting" from the CSP to CISA (the black arrow from CSP to CISA in Figure 3) will be further developed into general reporting patterns that will be used to describe unique reporting instances and possible vendor solutions in Volume Two. Figure 7 further delineates the reporting of telemetry from agency cloud resources to CISA as taking place in three stages.

- **Stage A**: Cloud Sensing; Generates the telemetry.
- **Stage B**: Agency Processing; Prepares the telemetry for communication.
- **Stage C**: Reporting to CISA; Includes the transmission of information and transition from the agency to CISA infrastructure and control.

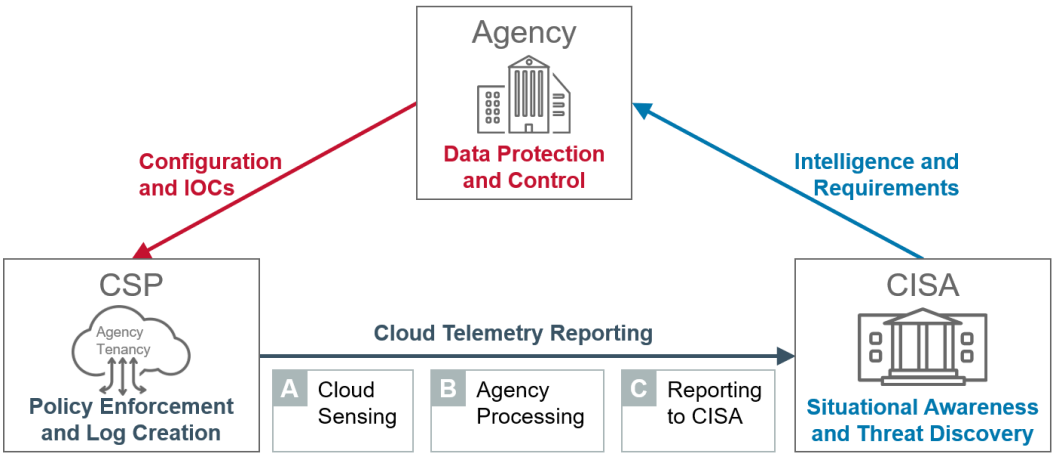

Figure 7: NCPS Cloud Telemetry Cycle Reporting Detail

Within each of the Reporting Pattern Stages there are attributes that capture the functions that take place within the stage (as shown in Figure 8). Each attribute describes a specific processing element that requires the agency to select from one or more options. Within the Cloud Sensing stage (Stage A), the two attributes are "Sensor Positioning" and "Telemetry Types," which describe where and what cloud telemetry is generated. Within the Data Processing stage (Stage B), there are four attributes that describe how the cloud telemetry may be processed prior to reporting to CISA: "Data Filtering," "Data Enrichment," "Data Aggregation," and "Data Transformation." Within the Reporting to CISA stage (Stage C), there are two attributes that are used to describe how the data will be transferred to and received by CISA: "Data Transfer" and "CLAW Distribution." Details for activities within the stage as well as options for these attributes are presented in the following sections.

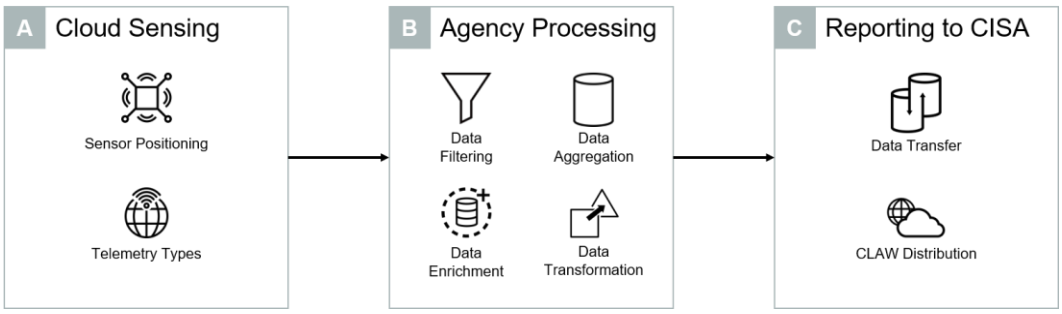

Figure 8: Agency Reporting Pattern Stages

Tools Available to Agencies

A wide range of tools are available to agencies for generating telemetry and for performing filtering, aggregation, and/or transformation, with differing functionality and costs. Options can be classified as cloud-native, agency-provided, or third-party.

Cloud Native

Cloud-native tools are provided by the CSP as services or configuration options. They are likely to offer a similar degree of trust, scalability, and interoperability as other cloud components from the same CSP. Given their native cloud support, they are also typically easier to configure and deploy. For these reasons they may be preferable to other options. However, configuration and customization of these generic capabilities will be required to align with the reporting pattern used by each agency. Services are typically priced on a tiered pay-for-what-you-use model. If data crosses CSP/region boundaries and/or is stored on agency resources, agencies also incur the associated costs.

Examples of CSP-provided capabilities include Amazon Web Services (AWS) Virtual Private Cloud (VPC) Flow Logs, Azure Network Security Groups Flow Logging, and Google Cloud Platform (GCP) VPC Flow Logs for generating network flow telemetry; AWS CloudWatch with Athena, Azure Monitor, GCP Cloud Logging with BigQuery for filtering; AWS Elasticsearch, Azure Monitor and Event Hubs, and GCP Pub/Sub for data aggregation; AWS Lambda, Azure Functions Consumption Plan, and GCP Cloud Functions for data transformation; and more sophisticated pipelines such as AWS Glue, Azure Data Factory, and GCP Cloud Data Fusion and Data Flow, which may include support for filtering, aggregation, and transformation.

Agency

Agency tools are capabilities that are developed by the agency and provided within the cloud (e.g., in virtual machines or "serverless"). Agencies using virtual machines (VMs) under their own control will be responsible for the proper scaling and load balancing between the VM instances; they will also be responsible for the associated operational costs. Software licensing costs may also apply.

Third Party

Third-party tools are provided by an external entity in the form of cloud SaaS services, cloud-based virtual machines (deployed by agencies), and/or remotely accessible web services. Agencies must interface with third-party services (e.g., through application programming interface (API) calls or publish/subscribe channels) to retrieve raw or processed data; using such services external to the agency and CLAW CSPs will entail a dependency on the provider for continuous operations and security of the service. It may also involve additional procurement and purchasing arrangements, including vendor

vetting. Software licensing costs are typical when using virtual appliances available from a CSP's marketplace (although they also come with support, upgrade, and training offerings).

4.1 Stage A: Cloud Sensing

Cloud data creation is the first step of the reporting pattern (as shown in Figure 6). The success of the NCPS Program is directly impacted by the type of data or logs made available for analysis. Different types of security logs are available for different cloud service models (IaaS, PaaS, or SaaS) and different CSPs. The selection of the cloud log types used as E1/E2-equivalent cloud telemetry will impact whether CISA is able to attain efficient and high-fidelity threat correlation and may affect the processing performance and costs incurred by the agencies providing the telemetry.

Network flow logs will initially be considered as the primary source of data to satisfy NCPS in the cloud visibility objectives. Later, additional types of cloud logs may be considered as a data source. Appendix A elaborates on network flow and other log types.

CISA and the participating agency will determine the specific collection location for network flow records based on the agency's requirements. In all cases, the following guiding principles help scope the generation of network flow records to be sent to CISA and those retained by the agency.

1. CISA is primarily interested in the network flow records that describe agency interactions with systems or components beyond the agency visibility, control, and administration (as opposed to interactions between internal agency components).
2. Network flow record collection is enabled for all data sensitivity designations of agency information hosted in the cloud (i.e., low and high sensitivity data will both require observation).
3. When agencies have more robust information collection needs for their internal purposes, as well as for post-collection processing or filtering of logs, network flow records can be used to align data sharing with agency and CISA MOU requirements.

NCPS Cloud Interface Reference Architecture: Volume Two details specific reporting patterns.

In the Cloud Sensing stage (Stage A), agencies configure one or more telemetry sources to send raw data to the Agency Processing stage (Stage B), or, in the case of no processing, directly to the Reporting to CISA stage (Stage C).

4.1.1 Attributes and Options

As listed in Table 1, the two attributes for consideration in the Cloud Sensing stage are Sensor Positioning and Telemetry Types.

Table 1: Cloud Sensing Options

Stage A – Cloud Sensing	
Attribute	**Options**
Sensor Positioning	Gateway
	Subnet
	Interface
	Service
	Application
Telemetry Types	Network Flow Logs
	Access/Auth Logs
	IDS/IPS Logs
	API Activity Logs
	DNS Logs
	VPN Logs
	Firewall Logs

Sensor Positioning: Options are based on where the telemetry is generated. Sensors for network flow logs may be placed at the gateway, subnet, or interface level. Other logs may be generated on application servers or with the various CSP services used by the agency. Appendix B discusses potential network flow data collection locations in more detail.

- **Gateway**: Network sensors are placed at the gateway between an agency's cloud tenancy and the internet[13], allowing monitoring to and from all agency cloud resources. When network address translation (NAT) is used, the agency must ensure that the records report public IP addresses. A suitable example would be an agency hosting a public website – and no private resources – on its cloud tenancy. NOTE: Traffic at the gateway location may include agency "private/internal" sources not typically monitored by NCPS. Processing may be required to exclude these records prior to sharing.
- **Subnet**: Network sensors are placed at individual subnets within an agency's cloud tenancy, allowing monitoring to and from agency cloud resources in each subnet. Private and public data flows can be separated so that only the latter is shared with CISA. A suitable example would be a cloud tenancy cohosting a public website and internal HR applications on different subnets. Only the subnet with the public website is provisioned to share records with CISA.
- **Interface**: Network sensors are placed at the individual network interfaces used by cloud virtual machines, allowing monitoring to and from each configured interface. Private and public flows can be separated, but with finer granularity than the subnet option and potentially greater insights for event correlation and analysis. A suitable example would be a server cohosting a public website and internal human resource (HR) applications on different interfaces. Only the interface with the public website is provisioned to share records with CISA.
- **Service**: Telemetry is generated from CSP services that provide key functions, such as load balancing, network/application firewalls, DNS, identity/authentication, key management, and more. In this way, services can double as sensors; however, they differ in their ability to be used as a data feed. Some can be configured to periodically deliver telemetry to agency cloud storage

[13] As well as other networks peered with the agency's cloud tenancy.

resources or to the CSP's monitoring service. Others may only make telemetry available through API calls or a manual export process.

- **Application**: Telemetry is generated from application servers, such as web servers and mail servers. Similar to services, application servers double as sensors, generating logs. They also vary in the level of visibility offered and log access mechanisms.

CISA Preference

CISA prefers that agencies place network flow sensors at each public subnet. This is the simplest way to ensure coverage of all public data flows while excluding many private data flows.

Telemetry Types: Options are based on what kind of telemetry is generated and include network flow logs, access logs, IDS/IPS logs, and API activity logs, DNS logs, VPN logs, and firewall logs. Appendix A describes each type in greater detail.

- **Network Flow Logs**: Network flow logs provide basic information about the data flows to and from agency publicly accessible cloud resources.
- **Access/Authentication Logs**: Access and authentication logs provide information on attempted and actual system/application access, with greater detail provided when privileged accounts are concerned.
- **IDS/IPS Logs**: Intrusion Detection System and Intrusion Prevention System (IDS/IPS) logs provide distilled information regarding suspected malicious activity. Other sources of distilled information (e.g., SIEM logs) also fall into this category.
- **API Activity Logs**: API activity logs record any operation employing cloud APIs, which covers nearly all activities which change some state within a cloud deployment.
- **DNS Logs**: Domain Name System (DNS) query/response logs can provide information regarding C&C (Command and Control) servers, DGA (Domain Generation Algorithms), etc. in addition to legitimate activity.
- **VPN Logs**: VPN logs provide the history of remote accesses (e.g., time, user, assigned address) to a VPN server, and may also include security posture of remote client system and user.
- **Firewall Logs**: Firewall logs provide information on network packets allowed and denied for forwarding and can be used to detect both attacks and benign misconfigurations.

CISA Preference

CISA prefers that agencies generate and share network flow logs as a first step and will coordinate with agencies on additional telemetry types.

4.1.2 Caveats and Considerations

Agencies can use the following caveats and considerations for evaluating options in the Cloud Sensing stage.

- **Processing Requirements**: The telemetry at this stage is raw and unfiltered and agencies must provision the storage, network, and compute resources necessary to process the full volume of the data they share; agencies should consider these requirements when choosing what sources to share and how to configure them.
- **Output Formatting**: Many sources can be configured with settings affecting the formatting and fields of the telemetry they provide. Agencies can use these settings to eliminate some work that would otherwise be done in the Agency Processing stage but should do so cautiously. For example, fields excluded from the original telemetry cannot be recovered later, whereas they would still be available to the agency if they were filtered out during the processing stage.
- **Visibility**: Agencies should provide as much visibility as possible about "public" interactions between their cloud systems and external networks and should minimize sharing on "private" interactions between internal components. Agencies should also consider whether each additional source they share increases visibility or is merely redundant.
- **Break-and-Inspect:**[14] Sensors that generate telemetry by inspecting traffic payloads should be able to "break-and-inspect" encrypted traffic. They should be positioned to minimize the associated risks (embedding a certificate authority, decrypting potentially sensitive traffic, etc.) and to ensure that break-and-inspect is not performed redundantly.
- **Encryption**: Telemetry data should be protected in transit. This includes when data from cloud-native and third-party sources is transferred to the agency and when data is transferred to later stages. Parameters for key length, key rotation, and cipher suites should be restricted to those that provide sufficient protection; specific details are given in Volume Two. Telemetry data should also be protected at rest (e.g., when it resides in cloud storage or on an agency-provisioned sensor).
- **Source Costs**: To the extent that agencies are sharing telemetry that they already generate and use for themselves, the costs of generating telemetry to share with CISA are minimal; however, an agency may need to add additional sensors. CSP services providing cloud-native telemetry are typically provided at little to no cost,[15] whereas agency-provided telemetry typically involves the cost of operating the sensors (and potentially the cost of licenses) and third-party telemetry typically involves subscription costs.

4.2 Stage B: Agency Processing

In the Agency Processing stage (as shown in Figure 8), data collected from the cloud are filtered, enriched, aggregated, and transformed into appropriate data formats that can be ingested by CLAW. The simplest option is where no processing takes place: essentially when raw log data is copied or moved directly from agency cloud sources to the reporting stage via a push or pull operation. More sophisticated processing may involve aggregation, enrichment, filtering, and data format transformation (called "data wrangling" or "data munging"). An agency may have multiple sensors and data streams distributed across multiple cloud subscriptions, regional cloud instances, or even CSPs. Combining these together implies some method for ordering and interleaving (e.g., by time) and filtering out certain

[14] This consideration does not apply to network flow logs and may be ignored by agencies which are in the early processes of sharing telemetry with CISA. It is, however, applicable as a general best practice.
[15] For telemetry generated by one CSP and delivered to another (i.e., the agency is using multiple CSPs), traffic egress and ingress costs also apply. See Section 3.2.1, Data Aggregation.

information, such as internal data transfers or logs containing sensitive information not required by CISA for threat analysis. Filtering of data may also be necessary to reduce the volume of information delivered to CLAW.

Data wrangling may be performed by an agency or by another party chosen by the agency for this task. Many CSPs offer such cloud-based capabilities either in their native ingestion pipelines and services or using third-party capabilities from within their respective marketplaces. Note that a very broad set of implementation options are available for the processing stage; the details vary among different CSPs and continue to change as their offerings improve.

4.2.1 Attributes and Options

As listed in Table 2, the four attributes for consideration in the Agency Processing stage are Data Filtering, Data Enrichment, Data Aggregation, and Data Transformation.

Table 2: Agency Processing Options

Stage B – Agency Processing	
Attribute	**Options**
Data Filtering	None
	Removal
	Sanitization
	Obfuscation
Data Enrichment	None
	Derived
	Agency-Defined
Data Aggregation	None
	Multi-Account
	Multi-Region
	Multi-Provider
Data Transformation	None (Native Forms Align)
	IPFIX
	JSON
	Parquet
	CEF
	Syslog
	LEEF
	CISA Coordinated

Data Filtering: Agencies providing data to CLAW will be required to filter logs to only provide the material the agency wishes to be analyzed by NCPS, both to satisfy privacy requirements and to improve the efficiency of analysis. Data selectors may include sensitivity markings (e.g., FOUO), network flow information (e.g., IP addresses, port numbers, protocols), or other information (e.g., domain names, user or system credentials, time of activity). Records containing unwanted information may be handled through a variety of mechanisms, including removal, sanitization, and obfuscation.

- **None**: Unfiltered logs are sent to CISA. This option is only appropriate when the agency is confident that the raw logs will not contain any information that they do not wish to share.
- **Removal**: Records containing unwanted information are discarded. As they otherwise contain information of interest, this option is only appropriate when the agency is confident that discarding these records will not result in a gap in visibility.
- **Sanitization**: Records containing unwanted information are sanitized, such that the information of interest is retained, and the undesirable information is completely erased. This can range from systematically removing a field from every record to blanking fields in individual records on a case-by-case basis.
- **Obfuscation**: Records containing unwanted information are obfuscated, such that the information of interest is retained. The undesirable information undergoes a transformation that preserves its usefulness for analytics while making it impossible to derive the original values. For example, real names may be substituted with a number, permitting the agency to recover the original content using a lookup table.

CISA Preference

CISA prefers that agencies sanitize records containing unwanted information. Agencies may also perform obfuscation; however, CISA analytics will not be dependent on data that agencies would only share in an obfuscated form.

Data Enrichment: As opposed to data filtering, in which agencies subtract unwanted information, with data enrichment agencies add desirable information to the records they share with CISA. Enrichment, if performed, may consist of either derived data and/or agency-defined data.

- **None**: Unenriched logs are sent to CISA. This option is acceptable when the logs already contain all the data fields that are expected by CISA.
- **Derived**: Agencies use existing information within records to derive and insert additional information of interest. Derivation can be used to provide required fields. For example, an agency has a cloud telemetry feed that provides a destination IP address/URL but omits the destination port; the agency derives the missing port based on the service offered at the IP address/URL.
- **Agency-Defined**: Agencies supplement records with additional contextual information that would otherwise only be known to the agency. Examples include identifying endpoints as either client or server, distinguishing between administrator, user, and guest entities, and mapping IP addresses to names of subnets. Agencies should coordinate with CISA when providing this additional information so that CISA is able to make the best use of it.

CISA Preference

Agencies may choose to perform any or no data enrichment, as long as CISA receives records with all the desired fields.

Data Aggregation: Agencies may wish to aggregate multiple sensor data sources. They may possess sources in multiple CSPs, multiple regions/cloud types, and/or multiple accounts/tenancies within the same CSP. Likewise, there are multiple instances of CLAW in different locations acting as targets. As moving data between CSPs or regions is comparatively expensive versus remaining "local," aggregation functions will likely be located based on the relative position of the sources and CLAW(s). Aggregation options include none, multi-account, multi-region, and multi-provider; more details are provided in Volume Two.

- **None**: Each telemetry stream is sent separately to CISA; alternatively, the agency only has a single telemetry stream.
- **Multi-Account**: Cloud logs from multiple accounts or tenancies within the same CSP are aggregated.
- **Multi-Region**: Cloud logs from multiple regions within the same CSP are aggregated.
- **Multi-Provider**: Cloud logs from more than one CSP are aggregated.

CISA Preference

Agencies may choose to use data aggregation at any level – account, region, and/or provider – as long as CISA receives the desired log fidelity and delivery is timely.

Data Transformation: Agencies will need to transform their data into a format known to CLAW; target formats include the native log format, Internet Protocol Flow Information Export (IPFIX), and other CISA-coordinated formats. Additional details and guidelines, including supported formats and expected fields, may be found in Volume Two.

- **None (Native Forms Align)**: The native log format generated by the sensors is already known to CLAW.
- **IPFIX**: The native log format generated by the sensors is unknown to CLAW and logs are transformed into IPFIX.
- **JSON**: The native log format generated by the sensors is unknown to CLAW and logs are transformed into JSON (JavaScript Object Notation).
- **Parquet**: The native log format generated by the sensors is unknown to CLAW and logs are transformed into Parquet.[16]
- **CEF**: The native log format generated by the sensors is unknown to CLAW and logs are transformed into CEF (Common Event Format).
- **LEEF**: The native log format generated by the sensors is unknown to CLAW and logs are transformed into LEEF (Log Event Extended Format).
- **Syslog**: The native log format generated by the sensors is unknown to CLAW and logs are transformed into Syslog.

[16] A format where data is organized by column/field instead of row/record, allowing for significant compression savings when large numbers of records are included; suitable when telemetry volume is high.

- **CISA Coordinated**: Agency-managed, third-party, and/or proprietary formats may be provided to CLAW if they provide a mechanism for CLAW ingestion processing and use. Approval for such log formats will be given on a case-by-case basis as coordinated with CISA.

CISA Preference

CISA prefers no data transformation (assuming format is compatible with CLAW).

4.2.2 Caveats and Considerations

Agencies can use the following caveats and considerations for evaluating options in the Agency Processing stage:

- **Content**: The selection of which data to provide to CLAW and NCPS may be driven by several factors, including privacy, data rates, network or storage costs, and formats. Every agency is responsible to provide all the required log data with appropriate security controls at a rate that does not overwhelm CLAW's ingestion system.[17] The data must also contain enough detail that NCPS analysts and analytic processes can produce useful threat analyses and alerting. CLAW guidelines for expected fields within various log formats and log categories will be further discussed in this document as well as in Volume Two and other guidance provided by CISA.
- **Aggregating Across CSPs**: Data aggregated from several sources that span CSP, region, or account may require careful account access control configurations and may incur additional communication costs. Security protections and usage tracking (billing) also tend to be tailored for use within a single CSP, so spanning providers may imply the need to create a set of compatible contracts and procurement processes across multiple vendors.
- **Combining Streams**: Several log streams may be combined into a smaller number to reduce the volume of data ingested by CLAW. This is an important factor, given CLAW's task of handling the volume of data created by all participating Federal Civilian Executive Branch departments and agencies. When combining data streams, data is generally ordered by some field. Most commonly for network metadata, this is a timestamp provided by sensors during collection. Issues regarding timing include time synchronization, precision, and accuracy of the timestamp. Different sensors, if not synchronized in a uniform manner, or with insufficient accuracy, will likely result in an unwanted stream of interleaved data record order, making subsequent analysis more difficult. An insufficient precision may result in the erroneous appearance of multiple simultaneous events. This can also frustrate subsequent processing and analysis.
- **Formats**: Data format conversion may be required if sensors produce logs in formats that are unknown to CLAW. Formats such as NetFlow and IPFIX are similar and may have relatively straightforward transformations. Logs from other components (e.g., IDSs, web proxies) tend to be of a more proprietary nature and more complicated data transformations may be necessary. Complicated data transformations may consume significant processing time, thereby increasing the overall end-to-end ingestion processing time.

[17] In addition, given that multiple agencies may start providing log data to CLAW simultaneously, a throttling and/or load shedding mechanism between the agencies and CLAW for the log feeds may be required (but is not yet specified). Without such a mechanism, some ingestion data loss may be unavoidable.

- **Data Rates**: Although it can use cloud scaling to handle additional load, CLAW (and the environment in which it resides) ultimately has a limit to the rate at which it can ingest data. This may be limited by one or more bottlenecks in processing, networking, or storage. In performing data wrangling involving the aggregation of multiple flows, assuming de-duplication has already been performed, some method for adjusting the incoming flow rate to CLAW may be required. Options include flow control back to the sensor sources (e.g., employing the underlying network protocol flow control), buffering data for a limited period of time (if the sensor data is bursty rather than persistently exceeding CLAW ingestion rate), or sampling the incoming data to reduce its rate. Care must be taken in sampling, as periodic sampling can under- or over-emphasize certain periodic phenomena.
- **Encryption**: Data arriving for processing is likely to be encrypted. In sophisticated processing scenarios (e.g., that involve de-duplication or merging by timestamp) the processing must have access to the contents of the sensor data streams supplied. Consequently, the processing agents must have access to keys to decrypt incoming data. Likewise, the outgoing (to CLAW) data or connections are also encrypted. The keys or certificates used to support this encryption must also be of sufficient strength and maintained appropriately in order to ensure the security of the most sensitive data handled between all the sensors and CLAW.

4.3 Stage C: Reporting to CISA

Once cloud telemetry has been sourced and processed in the Cloud Sensing and Agency Processing stages (respectively), the results are reported to CISA. The Reporting to CISA stage (as shown in Figure 8) consists of the data transfer of cloud telemetry to one or more CISA CLAW repositories. This represents a transition from data handling protections being handled by the agency to being handled primarily by CISA. As stated above, agencies retain authoritative data ownership and are only relaying a copy of their security telemetry to CISA for use in situational awareness and incident response.

4.3.1 Attributes and Options

As listed in Table 3, the two attributes for consideration in the Reporting to CISA stage are Data Transfer and CLAW Distribution.

Table 3: Reporting to CISA Options

Stage C – Reporting to CISA	
Attribute	**Options**
Data Transfer	Agency Push
	CLAW Pull
CLAW Distribution	Single Region
	Multi-Region
	Multi-Cloud

Data Transfer: Data transfer involves the mechanism by which agency cloud data is transferred to CISA after the data is collected and processed. Based on the party initiating the communication request, data transfer to CLAW can be classified as an agency push or a CLAW pull.

- **Agency Push:** The agency initiates a data transfer from their infrastructure (either at the Cloud Sensing or from Agency Processing stage) to CLAW. CISA hosts the receiving end (CLAW) in a listening fashion and issues credentials for agency use. The agency utilizes these credentials to authenticate, establishes a secure transfer means, then transfers the telemetry to CLAW. The agency may use the same credentials to transfer more than one data type to CLAW. In other words, CLAW will host a unique repository for each protected entity and agencies may populate their repository with multiple data types comingled in the same data store.
- **CLAW Pull:** The agency establishes a repository of interesting telemetry with reachability from CISA systems, issues access credentials for CISA use, and then listens for CISA pull requests. Agencies must negotiate polling interval, buffer duration during connection disruptions, link capacity, multiplexing, error correction, and other technical details with CISA. Although some of these details also apply with push option, they are more relevant here, as the agency is no longer in direct control of the data transfer.

CISA Preference

CISA prefers the agency push option to enable agencies to more fully execute their role as data owner and more tightly control the volume, rate, and content of shared telemetry. CISA hosts the receiving end in a listening fashion and issues credentials for agency use.

CLAW Distribution: To reduce cost (both latency and monetary), CLAW infrastructure is hosted in multiple locations. The intent is to position CLAW infrastructure in such a way that agencies can transfer their cloud data to a "local" repository that is in the same CSP and region.

- **Single Region:** Agency data transfers are to a single CLAW location. This option is especially low-cost if the agency telemetry is hosted within the same CSP infrastructure (such as AWS GovCloud West) as a CLAW instantiation (this may not always be possible).
- **Multi-Region:** Agency data transfers occur to more than one regionally-located CLAW within the same CSP infrastructure, such as AWS GovCloud East and AWS GovCloud West.
- **Multi-Cloud:** Agency data transfers occur to more than one CLAW hosted on more than one CSP infrastructure (such as AWS GovCloud and Microsoft Azure).

CISA Preference

CISA prefers an agency to use the CLAW distribution option that matches its cloud deployment in order to reduce cost (both latency and monetary).

4.3.2 Caveats and Considerations

Agencies can use the following caveats and considerations for evaluating options in the reporting to CISA stage:

- **Encryption**: The mechanisms employed to provide protection of data in transit must be mutually agreed upon by the agency and CISA. Parameters such as accepted encryption ciphers, key lengths, key lifetimes, authentication factors, key/credential distribution, and others must be established.
- **Initiation Costs**: Typically, the party that initiates the data transfer incurs additional costs. However, regardless of the initiating party, data is always outbound from the agency; if the data leaves the CSP, the agency incurs outbound data transfer costs, which can be greater than inbound costs.
- **Transfer Frequency**: The frequency or timeliness of the transfer can be based on time differentials (e.g., polling every five minutes) or based on log size (e.g., after every 20MB) of accumulated new data. In addition to the batching mechanism, transfers may also be triggered by noteworthy events (such as an unusually high volume of traffic).
- **High Availability/Durability**: CLAW infrastructure will provide high availability and data durability at each instantiation, ensuring resilient service offerings and increasing agency confidence.
- **"Local" Transfers**: Agency telemetry sharing with multi-region and multi-cloud CLAW infrastructure can reduce agency data transfer costs. This is due to multiple "local" data transfers potentially being more cost efficient than a single transfer to a "remote" location. However, this may increase technical complexity and administrative overhead.
- **Finite Deployments**: While attempts will be made to host CLAW resources as close to agency tenancies as possible, there will only be a finite quantity of CLAW instantiations. These locations may not fully accommodate the breadth of agency service locations. This will require the transfer of agency cloud telemetry to a "nearby" CLAW that may not be co-hosted on the same CSP.
- **Data Retention**: After confirmed transfer to CISA of telemetry copies, agencies must determine the duration to retain transferred data prior to deletion or removal.
- **Least Privilege**: CLAW must only be able to obtain data intended to be shared with CISA. Access permissions granted to the principal that pushes data to CLAW, or that pulls data from the agency on behalf of CLAW, should follow the principle of least privilege.

5 REPORTING PATTERN-LEVEL CHARACTERISTICS

As NCPS evolves to accommodate cloud services, agencies will be required to implement reporting patterns and maintain telemetry sharing with CISA. In the previous section, the three stage reporting pattern concept was introduced and discussed a series of attributes and options that agencies needed to select in each stage of a reporting pattern. This section introduces six high-level characteristics that apply to reporting patterns as a whole (including all three of their stages) that agencies need to evaluate when selecting which reporting pattern(s) to employ. These characteristics will be negotiated between agencies, CSPs, and CISA during CLAW integration activities. These six reporting pattern-level characteristics are as follows:

1. **Cloud Telemetry Timeliness**: The duration between cloud telemetry creation and presentation of that information to CISA analysts (to accommodate response within cyber-relevant time).
2. **Cloud Telemetry Timing Coordination**: Telemetry timestamp labelling mechanism in use (to enable CISA processing and proper record sequencing).
3. **Cloud Telemetry Provenance**: Telemetry source attribution and labelling (which may be complicated by agency aggregation and processing).
4. **Reporting Connection Administration**: Data transfer initiation, maintenance, and retirement execution.
5. **Cloud Telemetry Sharing Cost**: Potential expenses incurred for cloud sensing, agency processing, and reporting to CISA (based on attribute options selected).
6. **Agency Data Retention and Use Constraints:** Any additional data handling, retention, and use constraints, as captured in agency and CISA Memorandums of Understanding (MOUs).

Each of these characteristics has different implications that agencies will need to weigh when selecting a reporting pattern. The first three characteristics (cloud telemetry timeliness, cloud telemetry timing coordination, and cloud telemetry provenance) involve nuanced technical implications, whereas the next three characteristics (reporting connection administration, cloud telemetry sharing costs, and agency data retention and use constraints) are largely administrative. Additional details for each characteristic are presented in Appendices D-I (respectively).

5.1 Cloud Telemetry Timeliness

Different CSPs have different timeframes for log delivery. While typical values range from between a few minutes to fifteen minutes of event occurrence, agencies must confirm the timeliness of a CSP's log delivery through discussions with the CSP and their own testing. In most cases, the service documentation does not include guarantees regarding the timeliness of log delivery. While one CSP might claim that "events are delivered within five minutes of occurrence," another might claim that "events are delivered in real-time," and another might only provide hints in its documentation. Even within a CSP's offerings, more common/popular services are likely to have better documentation around timeliness than other services.

Some generalizations about timeliness may be made based on log type. Logs concerning point-in-time events (e.g., Application Programming Interface (API) activity logs) can be delivered quickly, whereas those concerning continuous events (e.g., network flow logs or application metrics detailing resource usage) have some interval that must transpire before the event is recorded and delivered. In the latter case, tenants may be given some control over the interval, with the caveat that shorter intervals may incur greater costs than the default/free interval.

CISA's goal is to detect, investigate, and respond to *any* threat before it has time to evolve and progress. Although CISA acknowledges that an agency has limited control over the timeliness of a CSP's delivery of raw logs, once the logs are received from the CSP, it is the agency that largely determines how long it takes to process the logs and deliver them to CLAW. Agencies should ensure that the time between raw logs release to the agency tenant from the CSP and the delivery of the processed logs to CLAW is within thirty minutes.

Additional cloud telemetry timeliness details can be found in Appendix D.

CISA Preference

When agency processing is performed, CISA expects the time between receiving raw logs from the CSP and the delivery of processed logs to CLAW to not exceed thirty minutes.

This CISA preference is a starting point for the current state of NCIRA architecture maturity. It is probable that this preference will require adjustment as hardware and software technology evolves and advances.

5.2 Cloud Telemetry Timing Coordination

Cloud telemetry includes indications of when observable activities took place, captured as timestamps that are typically applied by the system generating the telemetry. In order to ensure proper telemetry sequencing for event recreation and analysis when combining multiple telemetry sources, the clock of each telemetry source should be synchronized off of an agreed upon standard source. Time synchronization is the coordination of the system and services clocks (servers, workstations, network devices, services, etc.). Modern network infrastructures may have multiple links, network tiers, or data centers between the point where telemetry data is captured and the point where analysis is performed. The insertion of a standardized timestamp is a common method for preserving the telemetry generation times. This method is widely used in the industry, but the implementation specifics (timestamp accuracy, format, etc.) vary based on application.

One of the key requirements for accuracy when performing any kind of analytics is understanding precisely when an event that generated an observable record took place. Telemetry timestamping is essential for data analysis in modern networks (network troubleshooting, application performance tracking, security or threat analysis, and legal compliance). Any time-specific analysis performed is dependent upon the accuracy and precision of the timestamps of data being analyzed.

In the simplest case, the source (CSP) and the destination (CLAW) both influence timing synchronization, and discrepancies may occur between the systems. The cloud telemetry logs are timestamped when the log entries are generated. The logs are available to agency processing tools, where the original telemetry timestamps can be examined (but must not be altered). When the logs are pushed to CLAW, the originally generated log timestamps are therefore retained. The cloud telemetry timestamp format must be coordinated between agencies and CISA to ensure compatibility and accurate processing. CISA prefers that timestamps are in Coordinated Universal Time (UTC); if timestamps must

be generated in a different time zone, agencies and CISA must coordinate how the time zone – or the offset to UTC – is communicated within the telemetry.

Additional cloud telemetry timing coordination details can be found in Appendix E.

CISA Preference

When feasible, cloud-native telemetry timestamp format, precision, and accuracy should be preserved by agency processing to ensure accurate processing and use by CLAW systems and analysts. Timestamps in UTC are preferred.

This CISA preference is in accordance with the USNO Metrological and legal traceability of time signals II Official Time in the United States standard.

5.3 Cloud Telemetry Provenance

Data provenance of an information object refers to the process of tracing and recording the object's origin and history. Generally, provenance will include author identification, modification times, and some degree of information about activities performed that have affected the object's content or handling, as well as a method to ensure integrity of the history and object itself. Provenance information can help analysts and systems trace telemetry sources, track changes in data richness over time, identify the scope of untrustworthy data (if a telemetry source is misbehaving for whatever reason), and track updates as new telemetry versions become available, as well as other purposes. Provenance of cloud telemetry should be conveyed by agencies to CISA at sharing initiation and on an ongoing basis.

When an agency inhabits multiple tenancies and reports information to CISA in a push form, log information may be aggregated. Assuming a common log type and format across each data source in each tenant, the agency can aggregate the sources either by interleaving or combining them in some other fashion (e.g., data from one tenancy might precede that from another). In this case, provenance claims are likely to be made by the agency as an author of the combined data. Although multiple streams may arrive at the agency labeled and integrity-protected, the process of interleaving would create a new stream that itself requires provenance metadata. The agency would be responsible for asserting that it provided the aggregation of the multiple streams, and constituent streams may retain sufficient provenance information to be checked end-to-end by CISA when no agency filtration is performed.

The agency is also an author of log information when data filtering and/or enrichment is performed. Provenance claims in this context are (at least) three-fold: (1) the origin of the information from the CSP service, (2) the origin of the information used in performing the enrichment, and (3) the resulting stream provided to CISA by the agency. Agency processing should be arranged to convey (as new provenance claims) the nature of the modifications (e.g., enrichment) performed, the type of information removed, and the processing mechanisms (e.g., software artifacts) used in performing the processing.

Additional cloud telemetry provenance details can be found in Appendix F.

CISA Preference

Provenance of cloud telemetry should be conveyed by agencies to CISA at sharing initiation and on an ongoing basis.

This CISA preference requires telemetry provenance to provide a historical record of data, its origin, and generated evidence to support forensic activities. Proper forensics requires telemetry provenance for the entire data transfer session.

5.4 Reporting Connection Administration

Reporting connection administration consists of three primary lifecycle phases: connection initiation, connection maintenance, and connection retirement. Each phase entails unique considerations and objectives, as described below.

Connection Initiation

The first phase of the reporting connection administration lifecycle is connection initiation. Data transfers can be initiated by the agency, the CSP, or CLAW. The agency and CSP transfers are pushes and the CLAW transfer is a pull coordinated with the agency or the CSP. Determining the purpose, applicability, and responsibilities of the persons involved in creating, maintaining, and operating the reporting pattern systems should take place during initiation. Negotiation of protections of data in transit (encryption parameter selection), which party will initiate the machine-to-machine connection, what credentials to use, and what connectivity reachability changes are required (e.g., firewall permit lists for agency and CLAW sources and destination IP addresses) will also take place during this phase. Key management and distribution have multiple functions and scope depending on the service models utilized and reporting pattern selected.

Connection Maintenance

The second phase of the reporting connection administration lifecycle is connection maintenance. Monitoring the health and timeliness of the data transfers is a vital part of both system operation and cybersecurity. The health of the data transfers should be verified to ensure completeness and integrity. A mechanism for monitoring the timeliness of the data transfers should be established to ensure the occurrence of expected transfers. The trigger for data transfer intervals should be negotiated between the agency and CISA and could be determined by elapsed time or size of the accumulated data. Determining if there are visibility gaps, identifying the causes of these gaps, and instituting remedies is essential to the continued secure functioning of the system. Additional data transfer considerations should include the documentation of conditions that warrant emergency termination of telemetry sharing and the methods for accomplishing that termination, as well as the performance of data transfers in alignment with the “least function” principle (permitting only connectivity for those services required for the transfer).

To remedy visibility gaps when they are discovered, agencies should retain access to data after initial transfer to CLAW for a minimum period of seven days and a recommended period of thirty days. This allows for CISA to request re-delivery of said data if necessary. This does not override more stringent retention requirements based on the nature of the data (see Section 5.6).

Connection Retirement

The last phase of the reporting connection administration lifecycle is connection retirement. During decommissioning (when a reporting pattern implementation is no longer in use by an agency), the following activities may need to take place: any relabeling or retirement of agency shared data (for history), any sanitation of data processing equipment, closing any firewall permit-list entries, removing any special routing details, turning off data transfer connection health triggers, and removing queuing resource pools. This can assist in preserving the accuracy and security of agency and CLAW systems.

Additional reporting connection administration details can be found in Appendix G.

CISA Preference

Reporting connection administration details should be established prior to data transfer initiation to ensure each party understands its roles and expectations are met. Agencies should retain data delivered to CLAW for at least seven days, with thirty days being recommended, to facilitate re-delivery when necessary.

This CISA preference encourages agencies to negotiate documentation that formalizes the details of data transfer mechanisms. The indicated data retention windows are initial attempts at establishing window lengths that will allow for recovery in the event of data loss during a transfer. Window length adjustment may be necessary as the systems and environments grow and evolve.

5.5 Cloud Telemetry Sharing Cost

The process of sharing telemetry presents several areas where costs can be incurred and where steps may be taken to reduce such costs. CISA is taking steps to ensure that there are regional CLAW locations that can be used by the agencies to minimize overall reporting costs to the agencies.

Unless telemetry is delivered directly from the CSP to CLAW, agencies will likely incur costs at all stages of a reporting pattern. The following discussion assumes that agencies implement their reporting in the cloud; however, they may also use on-premise capabilities (and incur the on-premise costs of doing so).

For Cloud Sensing (Stage A), the primary cost is from the CSP generating the telemetry. Often, this is done free of charge,[18] though higher resolution data may be available for a fee. Regardless, agencies likely incur costs if the raw data is delivered to a CSP's storage service, and if the same CSP is the source of the data, this may be the only option. Receiving the data in a compressed format, if supported, can mitigate this cost.

For Agency Processing (Stage B), agencies may implement processing using IaaS, PaaS, and/or SaaS capabilities.[19] Compute is the primary cost driver for IaaS, and decisions about data filtering,

[18] For example, an IaaS tenant pays for network transfers and can enable network flow logs free of charge, or a SaaS tenant pays a monthly subscription where application log export is a basic and included feature.

[19] Due to the potentially sensitive nature of the telemetry, consider deploying PaaS-developed or SaaS-provided capabilities in an IaaS tenancy. Some IaaS CSPs have a PaaS offering that integrates with their other services, and most have SaaS offerings through their third-party marketplace.

enrichment, aggregation, and transformation all impact theoretical compute cost. Actual cost also depends on the agency's ability to right-size telemetry processing resources and identify appropriate places to use containerized or serverless services. Outside of compute costs, agencies also likely pay to store intermediate data, as well as egress traffic, depending on the type of aggregation performed.

For Delivery to CLAW (Stage C), egress traffic is the primary cost driver (when data is not delivered to a regional CLAW. This cost can be transferred to CISA in some (but not all) cases.[20] Storage costs may apply for the CLAW Pull option, and, in general, to implement data retention requirements. Data compression can mitigate costs. Another way for agencies to mitigate costs is by implementing a lifecycle plan in which data is migrated to progressively cheaper storage (and eventually deleted) as the need for low-latency, high-throughput access diminishes over time. Compute costs may apply for the Agency Push option if compute resources are used to perform the data push.

With the agency integrated analysis process approach (see Section 3.2 for more details), much of these costs are sunk costs. In order to avoid imposing an undue cost burden on agencies, CISA provides a variety of reporting patterns and options and sets balanced requirements for other factors (such as timeliness, provenance, data retention).

Additional cloud telemetry sharing cost details can be found in Appendix H.

CISA Preference

To minimize new costs, CISA encourages agencies to take advantage of their existing integrated analysis process when implementing their chosen reporting pattern and to utilize their regional CLAW when possible and appropriate.

This CISA preference recognizes the potential for increased impact the NCPS program may have upon agency budgets. Reapplication of existing infrastructure, systems, and processes may guide the selection of the reporting pattern and pilot project that best meets the agency budget and needs.

5.6 Agency Data Retention and Use Constraints

Cloud telemetry may include various levels of detail about the systems that are monitored. As agencies adopt new cloud services and select different reporting patterns, their ability to filter or obfuscate this telemetry for potentially sensitive information may be limited. In addition, visibility may evolve as the cloud services undergo changes, potentially increasing the details shared with CISA in established connections. While rare for telemetry data, an agency may have unique demands that necessitate special handling and use of their telemetry data by CISA. These may include regulatory,[21] privacy,[22] or agency

[20] The self-service mechanism some CSPs provide is a "requester pays" flag on storage containers, where the data requester is billed instead of the data owner for costs associated with the request (including data transfer).
[21] These include things like Law Enforcement Sensitive, Health Insurance Portability and Accountability Act, etc.
[22] Examples include: mobile device location may be sensitive to welfare case workers, persistent connections to external parties may disclose business relationships which are not publicly known, telemedicine delivery may disclose the IP addresses (or even login names) of patients or healthcare providers, citizen users of public-facing agency portals can disclose which IP address ranges (and locations) are most reliant on those government services, etc.

risk determinations. These data handling constraints should be communicated to CISA early in the reporting pattern selection process to ensure mutual understanding and to guarantee that proper protections can be established. CISA will ensure proper alignment with agency constraints or negotiate alternative mechanisms for preserving CISA mission objectives within those parameters prior to any telemetry data transfer. Special handling constraints may impact data retention windows, permitted data transfer and storage encryption standards, data tagging and labels, personnel qualifications for analysts, minimum technical and/or administrative controls, etc.

Additional agency data retention and use constraint details can be found in Appendix I.

CISA Preference

Agency cloud telemetry selected for sharing with CISA should adhere to Official Use Only guidelines. Agency special data retention and use constraints should be communicated to CISA prior to establishing any telemetry sharing.

This CISA preference addresses the sensitivity of the telemetry generated when agency applications and databases are relocated to the cloud. It sets a minimum sensitivity level and calls for formal documentation to specify the required protections to safeguard the shared telemetry data.

6 CISA CLOUD DATA AGGREGATION

CISA seeks to improve performance, reduce costs, and enhance threat discovery and incident responsiveness for agencies. CLAW is designed to support these goals by supporting agency adoption of cloud technologies. This section explains CLAW in more detail.

6.1 Cloud Log Aggregation Warehouse Overview

CLAW is a CISA-deployed architecture for the collection and aggregation of NCPS data from agencies using commercial CSP services. While agency NCPS data is currently aggregated on-premise at CISA, CLAW is deployed in the cloud to aggregate agency security logs that originate in the cloud. CLAW presents a functional, module-based architecture to ingest, store, and analyze security logs and sensor data from agencies. It is geared towards enabling secure and efficient methods to process cloud data in a manner that offers CISA a similar level of situational awareness provided by current EINSTEIN on-premise deployments.

6.1.1 CLAW Distribution

The CLAW architecture supports log aggregation at multiple locations (optimized for performance, cost, and efficiency) utilizing centralized threat discovery with distributed analytics. Figure 9 shows how agencies in different cloud regions can each transfer their data to CLAW without requiring each to push their data to a single region.

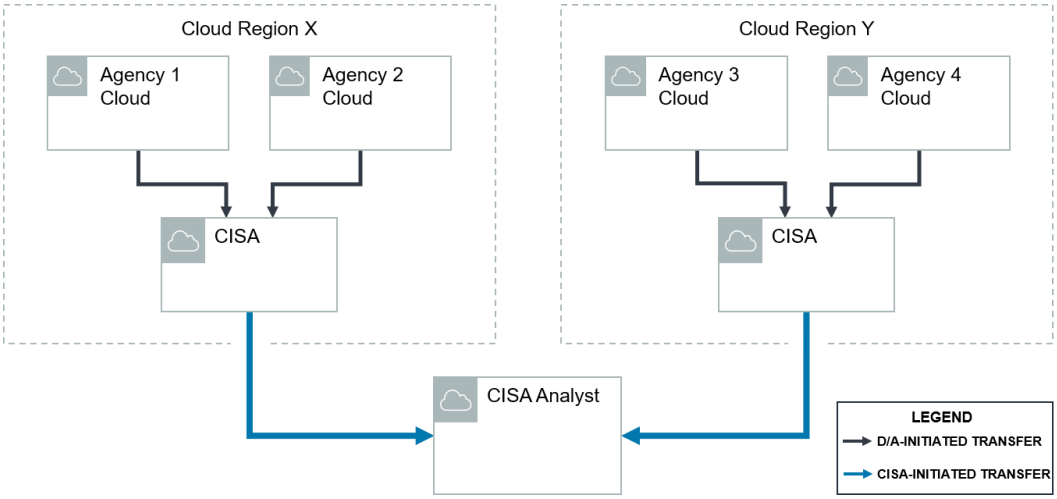

Figure 9: Responsibility for Transferring Security Data (Agency vs. CISA)

Agencies can transfer their data to the closest CISA CLAW location based on their CSP and region, reducing data transfer costs, technical complexity, and transmission latency. The CLAW architecture supports collocating CLAW aggregation points with agency tenants on major CSPs. Any additional data aggregation or consolidation required will occur within CISA's purview. Further details about issues and caveats related to CLAW distribution can be found in Section 4.3.2.

6.1.2 CISA Analysis of Agency Data

Using CLAW, CISA provides the environment and tools to correlate and discover threats from application and network data that has been shared by agencies. Current analytics approaches involve

signature-based (pattern recognition) and non-signature-based (heuristic and statistical) analytics for identification of IOCs and for identification of anomalous activities. Analysis will also bring in enrichment data to enhance the analysis results.

Figure 10 shows how cloud data from individual agency cloud tenancies is collected and analyzed at CISA cloud sites while preserving agency data isolation. Each agency's data is separated to prevent data comingling and corruption (using means such as independent data indexes and data stores). Analysis results obtained will subsequently be sent to CISA analysts for processing and assimilation (i.e., for threat detection and correlation, and for synthesis of security indicators).

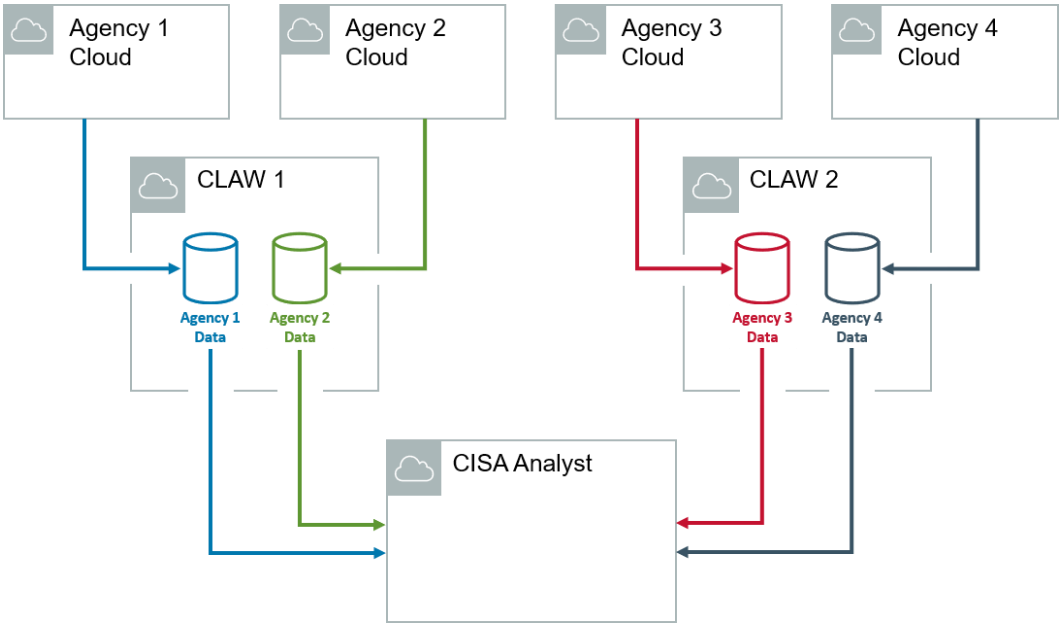

Figure 10: Agency Log Ingestion (Autonomy Preserved with Log Isolation)

Analysis will be coordinated from a central location with standardized tools. Those centralized tools will be able to interact with the sensor data distributed across CLAW locations. The data will be ingested and processed at a "local" CLAW location relevant to their CSP and region; in other words, the agency data will not be backhauled to a central repository. This will provide analysts with global situational awareness without requiring a corresponding centralized data store or requiring multiple copies of the same tools at each of the distributed data stores.

The data will be protected to ensure confidentiality and integrity using encryption for both data in transit and at rest that is compliant with Federal Information Processing Standard (FIPS) Publication 140-2[23]. In addition, data retention compliance requirements and data recall capability for long-term forensic discovery will be met until data is destroyed or removed. CISA sustainment operations (Network Operations Center/Security Operations Center) will have oversight of the CLAW.

[23]Refer to Appendix I Encryption Requirements: FIPS 140-3 for future FIPS 140 requirements.

7 CONCLUSION

As agencies move more of their applications and services to CSPs, the NCPS Program is evolving to ensure that security information about cloud-based traffic can be captured and analyzed and that CISA analysts can continue to provide situational awareness and support to the agencies. This document introduces general reporting patterns for how cloud logs will be collected and transferred to CLAW. Appendix C describes the implementation workflow for how agencies, CSPs, and CISA will partner to deploy NCPS in the cloud. The companion document (*NCPS Cloud Interface Reference Architecture: Volume Two*) provides a catalog of reporting patterns and options that match common agency cloud use cases. Together, these two documents provide guidance for how an agency can adapt their cloud environments to allow for security data to be sent to NCPS. Individual CSPs (e.g., AWS, Microsoft, etc.) can use these documents to provide vendor solutions for agencies and CISA to utilize.

APPENDIX A: CLOUD TELEMETRY TYPES

Network Flow Logs

IP network flow logs describe the communication that takes place between endpoints and enables modern network management and security. Network flow log protocols[24] specify how information about such communication is to be generated and formatted. Participating network devices implement these protocols by generating, compiling, and organizing network flow records as traffic traverses them. When collected and sent to CISA, network flow logs enable CISA to have situational awareness of an agency's cloud network activities.

Cloud activities deployed in multiple service models (IaaS and PaaS specifically) may generate different types of network flow logs. For example, when a node in the cloud initiates communication with another node, an intermediate sensor device with network flow record generation capability creates a flow record for that communication. Subsequent packets with the same network flow attributes update previously created flow records, which are continuously monitored and updated until the communication concludes. Flow records are then sent to a collector, where data logs are stored and further analyzed.

Access/Authentication Logs

Access and authentication logs contain a history of transactions, including success/failures of delivery of credentials for users, groups, and roles. Accesses which involve two-factor authentication and/or privileged accounts may be logged with additional/more detailed information. These logs are useful for detecting unexpected account activity or privilege escalations during or after attacks and may help to determine "blast radius" of successful attacks and active probing campaigns.

Access and authentication logs appear in a wide variety of contexts and are not mutually exclusive from the other telemetry types discussed below. API activity logs describe access/authentication to a CSP's API by cloud principals, while VPN logs describe access/authentication to a network by remote users. IDS/IPS logs include instances of access/authentication suspicious enough to warrant an alert. A firewall may generate multiple telemetry streams – one stream containing firewall logs proper and another describing access/authentication to its management functions.

IDS/IPS Logs

IDS/IPS logs contain output from assets that detect known and suspect malicious activity and report distilled results. These may come from monitoring network traffic (NIDS – network IDS or NTA – network traffic analysis), from monitoring VM/container activity (HIDS – host IDS), or from analyzing other log inputs (SIEM – security incident and event management). These logs can be used to detect anomalies/known malicious patterns in input data sources, focus attention on understanding attackers' current or past activities, and possibly orchestrate responses automatically.

IDS/IPS solutions may be provided as a CSP-native service, as a service from a SECaaS (security-as-a-service) vendor, or as a VM/container appliance in IaaS environments. IDS/IPS systems typically have access to databases of known current malicious patterns, with log records identifying the detected

[24] IPFIX is a prominent example and is related to the E1 and E1E formats.

pattern along with other contextual information. The ability to detect cloud/CSP-specific patterns, and to include cloud/CSP-specific context in logs, depends on the type of solution used and its configuration.

API Activity Logs

API activity logs consist of activities employing cloud APIs, which covers nearly all activities with any lasting change of state within a cloud deployment. They reveal actions that misuse/re-configure existing resources, create unauthorized resources, or perform actions in an unauthorized fashion (e.g., by wrong owner, wrong time, utilizing inappropriate access). These logs can be used to understand attacker's current and past activities outside (and to a lesser extent, inside) VM instances at a fine level of detail.

API activity logging is available in all cloud service models at minimal to no cost. (When applicable, anonymous access to public tenant resources may also be logged, but at a higher cost.) It forms the basis of many CSP-native and third-party security monitoring services while also being a common starting point for tenants' own security analytics.

DNS Logs

DNS logs are the history of DNS queries and responses derived from either network traffic or recursive server app logs. While not specifically mentioned by ATT&CK, they may provide very useful information regarding Command & Control (C2) servers and Domain Generation Algorithms (DGA) used by malware. Additionally, they also help to better understand attacker activities related to lateral movement (e.g., network enumeration) and data exfiltration (e.g., sensitive data in DNS requests).

DNS logs are more likely to be available in the PaaS and especially IaaS service models, where tenants can exert greater control over the DNS server used by internal resources. They may take the form of logs from a CSP-native DNS service or application logs from a DNS server VM controlled by the tenant.

VPN Logs

VPN logs contains the history of VPN accesses from remote users (e.g., time, user, assigned address), and may also include security posture of remote client system and user. They can be used to detect attempted unauthorized access, attempted (and successful) exploits on the VPN server, and to improve understanding of which users have authorized access as a function of time.

VPN logs are available in the IaaS service model, where client-based VPN gateways can be provisioned to allow remote users to connect to a tenant's cloud network. They may take the form of logs from a CSP-native VPN service or application logs from a VPN appliance VM controlled by the tenant. Tenants may also provision IPsec-based gateways to connect with on-premises networks, but the associated logs are different, focusing on the establishment and teardown of tunnels between sites.

Firewall Logs

Firewall logs indicate which packets/flows have been allowed or denied forwarding. They help to determine if a current attack is underway or has occurred in the past – including large traffic-based attacks such as DDoS – and are also helpful to detect benign misconfigurations. Firewall logs can serve as network flow logs but can also differ as in the case of application firewalls.

APPENDIX B: FLOW RECORD COLLECTION LOCATION

The flow record collection guidelines discussed in this document apply differently depending on where in the agency's cloud system the demarcation point occurs. While CISA provided guidance on demarcation points, found in Section 4.1, it is the agency's responsibility to identify the relevant demarcation points within their cloud environments. Three typical deployment locations for network flow generation in IaaS deployments are shown in Figure 11. In addition, this appendix enumerates some conditions under which each is suitable.

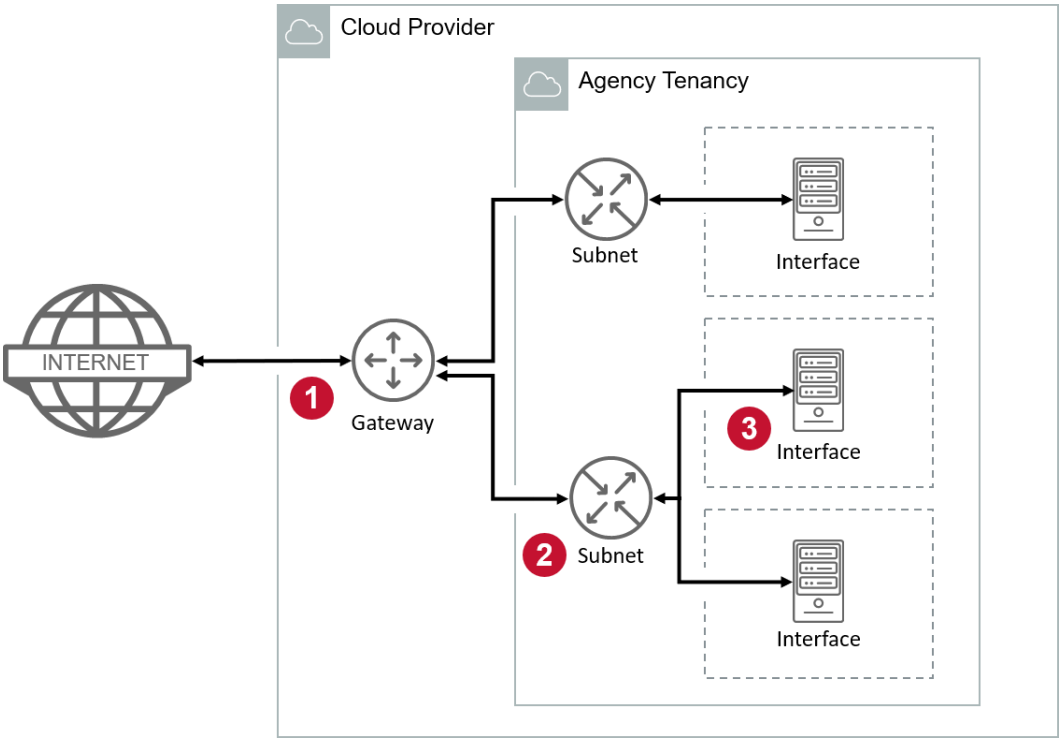

Figure 11: Network Flow Log Generation Positions for IaaS

There are at least three potential collection locations for agency tenancy network flow records. Each of these collection locations has unique visibility scope and detail.

❶ Internet Gateway

The first potential collection location is the gateway at the internet to agency cloud tenancy interface(s). Collection of network flow records at the gateway allows monitoring of all traffic to and from all agency cloud resources. When NAT is used, the agency must ensure that the records gathered report the public IP addresses. The traffic monitored at this location may include agency "private/internal" sources not typically monitored by the NCPS sensors. The records gathered here may require processing to exclude those records prior to being sent to CISA. An example would be a publicly accessible web site hosting publicly available information, where the agency is not cohosting any additional resources on the same cloud tenancy.

❷ Subnet

The second potential collection location(s) are the subnet(s) utilized by the agency tenancy. Collection of network flow records at the subnet level allows for monitoring of all traffic to and from cloud server(s) on each individual subnet. The “private/internal” and “public” data flows can be separated, thereby constraining the sharing of data flow information with CISA to only the “public” resources and reducing post-collection processing requirements. An example would be a publicly accessible web site hosting publicly available information and internal human resources applications in the cloud with “public” and “private/internal” data flows (respectively) destined for resources on independent subnets. The subnet with the publicly available information is provisioned to share network flow records with CISA.

❸ Interface

The third potential collection location(s) are the interface(s) utilized by the agency tenancy to provide access to cloud virtual machines. Collection of network flow records at the interface level allows for monitoring of all traffic to and from the individual interfaces on each of the cloud server(s) that has been properly configured to provide this capability. The “private/internal” and “public” data flows are separated by the individual virtual interfaces. This type of collection location provides the finest granularity for the network flow records, minimizes the post-collection processing requirements, and permits greater insights for event correlation and analysis. An example would be a publicly accessible web site hosting publicly available information in the cloud with the public data flows destined for resources on independent interfaces. The interfaces with the publicly available information are provisioned to share network flow records with CISA.

APPENDIX C: NCPS IN THE CLOUD IMPLEMENTATION WORKFLOW

In order to implement NCPS in the cloud, CISA, agencies, and CSPs will need to take separate, coordinated actions. Figure 12 shows a workflow sequence of the actions that must come together in order to successfully implement NCPS in the cloud. The subsections in this section give details about the actions so that agencies can plan accordingly.

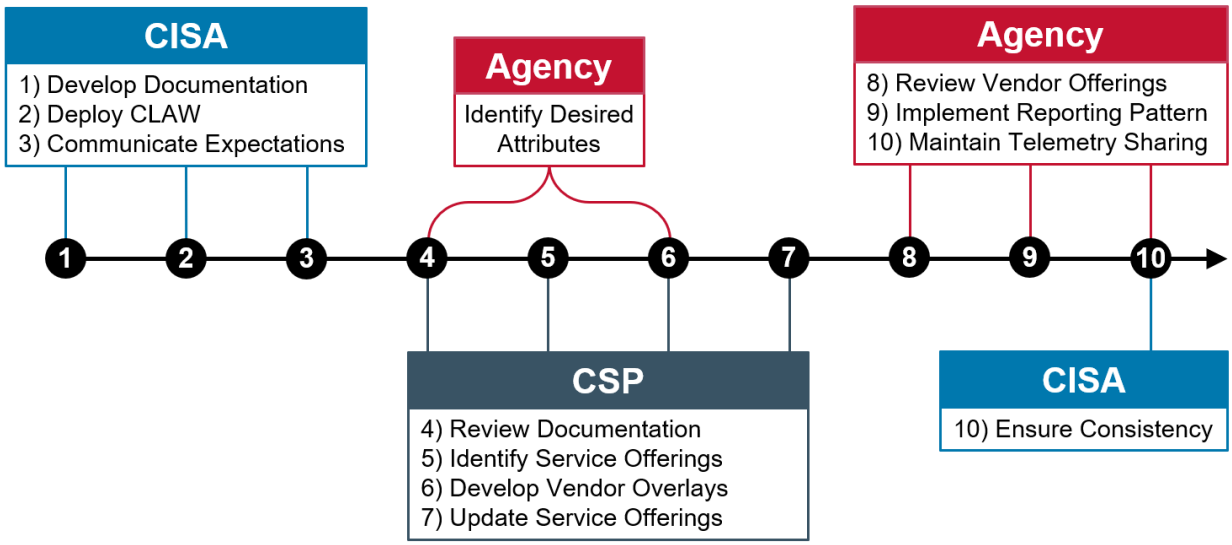

Figure 12: Implementation Workflow for NCPS in the Cloud

CISA

As NCPS evolves to accommodate cloud services, CISA will have multiple roles and responsibilities in order to implement NCPS in the cloud. CISA's roles and responsibilities are as follows.

- **Develop Documentation**: CISA will enumerate generic reporting patterns and components within Volumes One and Two of this reference architecture.
- **Deploy CLAW**: CISA will deploy CLAW across a number of CSPs and regions, giving agencies options for where to deliver cloud telemetry.
- **Communicate Expectations**: CISA will communicate reporting expectations to CSPs and agencies in a number of ways including released documentation, outreach activities and one-on-one interactions between an agency and CISA.
- **Ensure Consistency**: CISA will work to ensure the consistency of agency cloud telemetry inputs. This will be a continuous improvement process.

CSP

As NCPS evolves to accommodate cloud services, CSPs will have multiple roles and responsibilities in order to implement NCPS in the cloud. CSP roles and responsibilities are as follows.

- **Review Documentation**: CSP will review the cloud telemetry reporting pattern documentation in Volumes One and Two of this reference architecture.
- **Identify Service Offerings**: CSP will identify which service offerings they provide that could satisfy cloud telemetry reporting pattern options.
- **Develop Vendor Overlays**: CSP will author and publish agency guidance for utilizing service offerings in alignment with NCPS cloud telemetry reporting patterns.
- **Update Service Offerings**: If desired, the CSP vendor may modify their product offerings to align with NCPS in the cloud.

Agency

As NCPS evolves to accommodate cloud services, agencies will have multiple roles and responsibilities in order to implement NCPS in the cloud. The agency roles and responsibilities are as follows.

- **Identify Desired Attributes**: Agency identifies its desired options for each of the attributes for reporting cloud telemetry in a reporting pattern that matches its use case.
- **Review Vendor Offerings**: Agency will review vendor reporting pattern documentation and identify cloud service offerings that can be used to satisfy options.
- **Implement Reporting Pattern**: Agency will select, configure, and verify vendor or agency-created services to instantiate reporting patterns for sharing cloud telemetry with CISA.
- **Maintain Telemetry Sharing**: Agency will maintain telemetry sharing with CISA.

APPENDIX D: CLOUD TELEMETRY TIMELINESS

Different CSPs have different timeframes for log delivery. While typical values range from between a few minutes to fifteen minutes of event occurrence, agencies should confirm the timeliness of a CSP's log delivery through discussions with the CSP and their own testing. In most cases, the service documentation does not include guarantees regarding the timeliness of log delivery. While one CSP might claim that "events are delivered within five minutes of occurrence," another might claim that "events are delivered in real-time," and another might only provide hints in its documentation. Even within a CSP's offerings, more common/popular services are likely to have better documentation around timeliness than other services.

Some generalizations about timeliness may be made based on log type. Logs concerning point-in-time events (e.g., transaction logs for auditing Application Programming Interface (API) calls) can be delivered quickly, whereas those concerning continuous events (e.g., network flow logs or application metrics detailing resource usage) have some interval that must transpire before the event is recorded and delivered. In the latter case, tenants may be given some control over the interval, with the caveat that shorter intervals may incur greater costs than the default/free interval.

CSPs may tailor their log delivery based on the destination. For example, a CSP may do hourly batching to its general-purpose storage destination, within-minutes delivery to its big data service, and real-time streaming to its publish/subscribe service.[25] If agencies perform processing before delivery to CLAW, they must recognize when this behavior is present in a CSP and provide a receiving destination that allows timely receipt of raw logs. If logs are pushed directly to CLAW, then CISA will provide an appropriate receiving destination for the CSP service.

Completeness

A complement to log timeliness is log *completeness*; data that never arrives is not timely at all. CSPs may not provide complete logs for several reasons, some intentional and some not.

- The methodology for generating data is based on sampling (e.g., only a sample of network packets being used for network flow logs, values for application metrics only being sampled at certain intervals, etc.).
- Events occur faster than the CSP can log them. This is more likely for data plane events (e.g., HTTP GET requests to a public storage container) than for management plane events (e.g., API calls that change the configuration of the storage container).
- Misconfiguration results in some events being seen but not logged (e.g., different audit settings based on roles) or sensors being placed such that events are not seen at all (e.g., network sensor placement relative to a gateway/firewall).
- Log comprehensiveness is related to log completeness; for some services, tenants may change the default settings to allow for more detailed or less detailed reporting.

[25] https://en.wikipedia.org/wiki/Publish%E2%80%93subscribe_pattern

Cyber-Relevant Timeframe

To be timely, data must be received in a cyber-relevant timeframe. However, what is considered a cyber-relevant timeframe varies depending on threats. Recent open-source reporting[26] has measured the "breakout time" of several well-known threat actors, where breakout time is defined as the time from initial compromise to the start of lateral movement (including steps such as local network reconnaissance and privilege escalation on the compromised host). Effective action within this time can stop an attack in its early stages. While most threat actors had an average breakout time of over two hours, Russian actors were found to have an average breakout time of under twenty minutes. This is significantly faster than what many organizations are prepared to handle.

Delivery

CISA's goal is to detect, investigate, and respond to any threat before it has time to evolve and progress. Although CISA acknowledges that an agency has limited control over the timeliness of a CSP's delivery of raw logs, once the logs are received from the CSP, it is the agency that largely determines how long it takes to process the logs and deliver them to CLAW. Agencies should ensure that the time between raw logs release to the agency tenant from the CSP and the delivery of the processed logs to CLAW is within thirty minutes.

CISA Preference

When agency processing is performed, CISA expects the time between receiving raw logs from the CSP and the delivery of processed logs to CLAW to not exceed thirty minutes.

This CISA preference is a starting point for the current state of NCIRA architecture maturity. It is probable that this preference will require adjustment as hardware and software technology evolves and advances.

[26] https://www.crowdstrike.com/resources/reports/2019-crowdstrike-global-threat-report/.

APPENDIX E: CLOUD TELEMETRY TIMING COORDINATION

All systems that consist of multiple servers and clients (such as in cloud service delivery models) require timing synchronization. Timing synchronization is an important issue that must have its own considerations based on the level of accuracy specified by individual agencies. This concept applies to CLAW. The passage of data from CSP to agency to CLAW, through multiple servers and clients, dictates the need for time synchronization, as depicted in Figure 13, below.

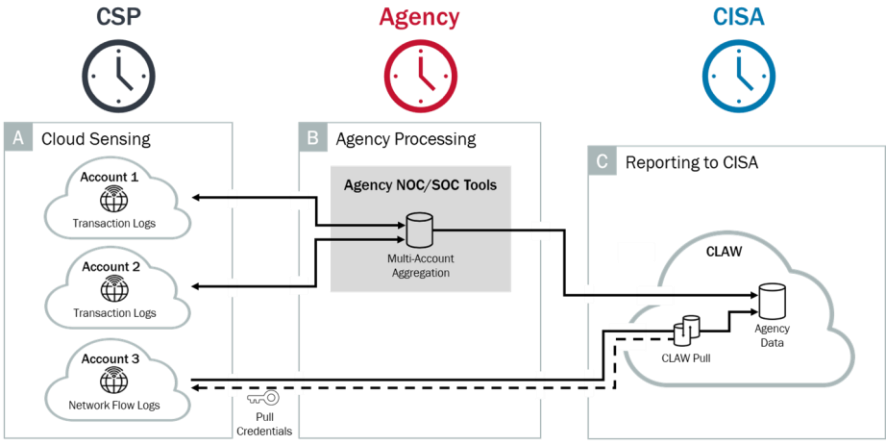

Figure 13: Typical Organizations Involved in a CLAW Reporting Transaction

Standard Terms

As indicated above, synchronization is desired across the various geographic regions (on the CSP side) that participate in the generation of the cloud logs stored in CLAW. In addition, it is important to ensure that the systems of the individual agencies use a standard timestamp (precision, format, etc.). This includes the various processing stages that are involved in the manipulation or filtering of the logs before they are stored in CLAW. To achieve this, this document defines several standard terms that are used in the process of synchronization.

System Time

The current time and date used by computer systems to supply applications with accurate time. Computer systems base their system time on the current time in relation to UTC (Coordinated Universal Time)[27] and each time zone is designated as an offset ahead or behind by a specific number of hours.

Authoritative Time Source

A single source synchronized to UTC by which events can be time-stamped and correlated is required for any network. An authoritative time source is critical to support essential operational and analytical cybersecurity functions and processes with an accuracy determined by the functions conducted at the local site.

Time Synchronization

Time synchronization is the coordination of the system and services clocks (servers, workstations, network devices, network services, etc.). Time synchronization with the authoritative time source is

[27] https://www.nist.gov/publications/metrological-and-legal-traceability-time-signals.

critical to support essential operational and analytical cybersecurity functions and processes with an accuracy determined by the functions conducted at the local site.

Timestamping

Telemetry timestamping is essential for data analysis in modern networks (network troubleshooting, application performance tracking, security or threat analysis, and legal compliance). Any time-specific analysis performed is dependent upon the accuracy and precision of the timestamps of data being analyzed. One of the key requirements for drawing valid conclusions from any kind of analytics is understanding precisely when a packet was captured. Modern network infrastructures may have multiple links, network tiers, or data centers between the point the data is captured and the point where the analysis is performed. The insertion of a standardized timestamp (precision, format, etc.) is a common method for preserving the data capture times. This method is widely used in the industry, but the implementation specifics vary based on the application.

Coordinated Universal Time

UTC is a standard universal time system that does not depend on the local calendars or the geographic location of the affected systems. It is highly accurate and can be used as a standard timing source for the purposes of synchronization and other tasks that require the knowledge of time with a high level of accuracy.

UTC is the primary time standard by which the world regulates clocks and time. UTC is also referred to as Greenwich Mean Time (GMT), Universal Time (UT), Zulu, or Z Time.[28]

From a scientific perspective, time is a hard subject to regulate. Science (and society) measures time with respect to the International Celestial Reference Frame (ICRF), which is computed using long baseline interferometry of distant quasars, GPS satellite orbits, and laser ranging of the moon (the local moon of the planet Earth). Irregularities in Earth's rate of rotation cause UTC to drift regularly from the time with respect to the ICRF. To address this clock drift, the International Earth Rotation and Reference Systems (IERS) occasionally introduces an extra second into UTC to keep it within 0.9 seconds of real time. This is also known as a leap second.

Leap seconds are known to cause application errors, as many applications are not prepared to deal with a 61st second. This can be a concern for developers and systems administrators. This can also introduce issues with various servers across multiple geographic regions unless it is taken into consideration and accounted for in timing calculations. In some cases, timing sources smooth out leap seconds over a given period (commonly called "leap smearing"), which makes it easy for applications to deal with leap seconds. In all cases (and for the purposes of CLAW reporting), it is important to know how various CSPs deal with this matter and to take that into consideration whenever timing synchronization is addressed.

Timing Synchronization

In the simplest case, the source (CSP) and the destination (CLAW) both influence timing synchronization, and discrepancies may occur between the systems. The cloud telemetry logs are

[28] There are subtle differences between these items. GMT is equivalent to UT1, which along with UTC are variants of Universal Time that differ by fractions of a second. Zulu/Z Time is the time zone with no offset from UTC.

timestamped when the log entries are generated. The logs are available to agency processing tools, where the original telemetry timestamps can be examined (but must not be altered). When the logs are pushed to CLAW, the originally generated log timestamps are therefore retained. The cloud telemetry timestamp format must be coordinated between agencies and CISA to ensure compatibility and accurate processing. CISA prefers that timestamps are in Coordinated Universal Time; if timestamps must be generated in a different time zone, agencies and CISA must coordinate how the time zone – or the offset to UTC – is communicated within the telemetry.

CISA Preference

When feasible, cloud-native telemetry timestamp format, precision, and accuracy should be preserved by agency processing to ensure accurate processing and use by CLAW systems and analysts. Timestamps in UTC are preferred.

This CISA preference is in accordance with the USNO Metrological and legal traceability of time signals II Official Time in the United States standard.

APPENDIX F: CLOUD TELEMETRY PROVENANCE

Provenance claims are assertions about the origin, authorship, and modification history of a particular object, accompanied with a verifier such as a cryptographically secure signature. Advanced forms involve a complete history of all modifications to the object along with an authenticated trail of modification. Provenance in the cloud is generally more useful than for private data stores. This is due to the ease with which such information is distributed and incorporated into other data products. Generally, provenance will include author identification, modification times, and some degree of information about activities performed that have affected the object's content or handling.

Provenance can be applied to the major services many applications use, including object storage, databases, and messaging services. The consistency of the provenance may be a system variable. For example, log file hashes can be conveyed as digital signatures on each file or collected into a signed digest file delivered periodically. Finer-grain considerations might include read-after-write synchronization of underlying objects, i.e., whether a read immediately following a write returns the updated object or an older version.

The concept of provenance can also be applied to software artifacts, especially those involved in manipulating or drawing conclusions from important or sensitive data sets themselves. Likewise, the computing environments (e.g., containers) in which software artifacts execute are important components in the overall provenance picture.

Degree of Processing

The degree of provenance claims processing relates closely to the degree of agency processing of the telemetry, which ranges from "pass-through" to "authored." In the pass-through case, sensor data may simply be forwarded from agency sensors to CISA, whereas at the "authored" end of the spectrum, sensor data may be interpreted, edited, summarized, transformed, otherwise manipulated, or even replaced before it is reported to CISA. In this latter case, and in many intermediate cases, the agency itself can be considered the author of the data (or at least one of the contributing authors) as opposed to merely a pipeline for sensor data.

Agencies that provide multiple telemetry streams to CLAW may thus provide different provenance claims for each, such as when streams undergo differing levels of processing or simply contain different types of data.

Integrity Checking Mechanisms

For pass-through cases, most CSPs provide provenance claims regarding which sensors provided logging information, and the connection from CSP to agency to CISA uses an encrypted and integrity-protected channel (e.g., Transport Layer Security). CSPs also often provide periodic checksums or hashes on log files. These integrity checking mechanisms may be invoked to provide an end-to-end assessment as to the veracity of the CSP-provided log data. Where the CSP or service has access to higher level information, such as the individual user or account responsible for an action, this may also be included in the provenance claims.

Agency Authoring

When an agency inhabits multiple tenancies and reports information to CISA in a push form, log information may be aggregated. Assuming a common log type and format across each data source in each tenant, the agency can aggregate the sources either by interleaving or combining them in some other fashion (e.g., data from one tenancy might precede that from another). In this case, provenance claims are likely to be made by the agency as an author of the combined data. Although multiple streams may arrive at the agency labeled and integrity-protected, the process of interleaving would create a new stream that itself requires provenance metadata. The agency would be responsible for asserting that it provided the aggregation of the multiple streams, and constituent streams may retain sufficient provenance information to be checked end-to-end by CISA when no agency filtration is performed.

The agency is also an author of log information when data filtering and/or enrichment is performed. Provenance claims in this context are (at least) three-fold: (1) the origin of the information from the CSP service, (2) the origin of the information used in performing the enrichment, and (3) the resulting stream provided to CISA by the agency. Agency processing should be arranged to convey (as new provenance claims) the nature of the modifications (e.g., enrichment) performed, the type of information removed, and the processing mechanisms (e.g., software artifacts) used in performing the processing. Along with data aggregation and data transformation, agencies have a large degree of freedom; an indicator of the agreement between the agency and CISA demonstrating how the stream provided to CISA is sufficient for NCPS operations should be included or referenced from the provenance claims.

Temporal Ordering

In many cases, a cloud tenant will have multiple sensors, services, and analytics running simultaneously to achieve multiple objectives, such as security, reliability, and performance. Consequently, an individual scenario may involve provenance from different types of systems and from multiple geographic locations. In these cases a timestamp or transaction identifier is commonly used to provide temporal ordering and correlation, but note that time should be of sufficient precision and accuracy to make such log aggregation possible. See Appendix E for more details.

CISA Preference

Provenance of cloud telemetry should be conveyed by agencies to CISA at sharing initiation and on an ongoing basis.

This CISA preference requires telemetry provenance to provide a historical record of data, its origin, and generated evidence to support forensic activities. Proper forensics requires telemetry provenance for the entire data transfer session.

APPENDIX G: REPORTING CONNECTION ADMINISTRATION

Reporting Pattern Generation

The system administrators for the agency infrastructure, the agency cloud tenancies, and CLAW must establish, configure, and maintain the credentials for the human and digital entities necessary to the creation and functioning of the reporting connections.

Data Transfer Overview

Data transfers can be initiated by the agency, the CSP, or CLAW. The agency and CSP transfers are pushes and the CLAW transfer is a pull coordinated with the agency or the CSP.

Key management has three functions: confidentiality, integrity, and source authentication. The scope of key management has several aspects: generation, storage, distribution, recovery, and destruction. The availability and implementation of key management services is dependent upon the individual CSP and is also dependent upon the reporting pattern and services model the agency chooses to implement. The key exchange between the involved entities is essential to providing security for the data being transferred. Protection of the required keys must be established and monitored for unauthorized disclosure.

Agencies must also determine if all data being transferred requires encryption, and which encryption algorithms are supported by all involved entities. The reporting patterns that involve a push from the CSP to CLAW require the CSP to initiate the credential exchange. For the reporting patterns that are performed via a CLAW pull, the agency configures the credentials.

NIST References

There are several National Institute of Standards and Technology (NIST) documents that provide useful guidance on this characteristic:

- NIST.SP.800-53r5[29] is a reference for determining the purpose, applicability, and responsibilities of the persons involved in creating, maintaining, and operating the reporting pattern systems.
- NIST.SP.800-57 Part 1,[30] Part 2,[31] and Part 3[32] are references to provide recommendations for key management.
- NIST.SP.800-60 Vol. 1[33] and Vol. 2[34] provide guidance for mapping types of information and information systems to security categories.
- NIST.SP.800-152[35] provides guidance for establishing a federal profile for key management.
- NIST.SP.800-175A[36] provides guidelines for the use of cryptographic standards.

[29] https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final.
[30] https://csrc.nist.gov/publications/detail/sp/800-57-part-1/rev-5/final.
[31] https://csrc.nist.gov/publications/detail/sp/800-57-part-2/rev-1/final.
[32] https://csrc.nist.gov/publications/detail/sp/800-57-part-3/rev-1/final.
[33] https://csrc.nist.gov/publications/detail/sp/800-60/vol-1-rev-1/final.
[34] https://csrc.nist.gov/publications/detail/sp/800-60/vol-2-rev-1/final.
[35] https://csrc.nist.gov/publications/detail/sp/800-152/final.
[36] https://csrc.nist.gov/publications/detail/sp/800-175a/final.

Data Transfer Monitoring

Monitoring of the health and timeliness of the data transfers is a vital part of the system operation and security. The health of the data transfers should be verified to ensure completeness and integrity (see Appendix C). A mechanism for monitoring the timeliness of the data transfers should be established to ensure the occurrence of expected transfers. The trigger for data transfer intervals should be negotiated between the agency and CISA and could be determined by elapsed time or size of the accumulated data. Determining if there are visibility gaps, the causes of these gaps, and instituting remedies is essential to the continued secure functioning of the system. Some potential causes of visibility gaps are connectivity failures, data transfer failures, API changes, etc. Alerts for unsuccessful completion should be generated and forwarded to the appropriate entities for remediation. It is optimal if any unsuccessful data transfer occurs that it is addressed before the agreed upon data retention window has elapsed (refer to Appendix F for details).

Monitoring the production systems for security vulnerabilities should be performed on a regularly scheduled basis. Additional vulnerability testing and remediation should be performed when notifications of newly discovered vulnerabilities are transmitted to the agency. An ideal situation would be that both the agency and CISA would monitor the completeness, integrity, and timeliness of data transfers for conformance to the agency and CISA agreement. Additional data transfer considerations should include the documentation of conditions that warrant emergency termination of telemetry sharing and the methods for accomplishing the termination,[37] performance of data transfers in alignment with the "least function" principle[38] (permitting only those services required for the transfer), and assurance that shared trust of the CLAW instantiations for the individual agencies is configured to prevent attackers achieving cross-agency compromise.[39]

When a communications session is completed the associated network connection must be terminated.[40] Network disconnect applies to internal and external networks. Terminating network connections associated with specific communications sessions includes de-allocating TCP/IP address or port pairs at the operating system level and de-allocating the networking assignments at the application level if multiple application sessions are using a single operating system-level network connection. Periods of inactivity may be established by organizations and include time periods by type of network access or for specific network accesses.

[37] Per NIST.SP.800-53r5 AC-12.
[38] Per NIST.SP.800-53r5 AC-6.
[39] Per NIST.SP.800-53r5 AC-4 and AC-20.
[40] Per NIST.SP.800-53r5 SC-10.

CISA Preference

Reporting connection administration details should be established prior to data transfer initiation to ensure each party understands its roles and expectations are met. Agencies should retain data delivered to CLAW for at least seven days, with thirty days being recommended, to facilitate re-delivery when necessary.

This CISA preference encourages agencies to negotiate documentation that formalizes the details of data transfer mechanisms. The indicated data retention windows are initial attempts at establishing window lengths that will allow for recovery in the event of data loss during a transfer. Window length adjustment may be necessary as the systems and environments grow and evolve.

APPENDIX H: CLOUD TELEMETRY SHARING COST

For many of the major IaaS CSPs, one consequence of the pay-for-use model is that each service has its own pricing structure. Combined with their large and growing service portfolios, tenants find it difficult to select the most cost-effective solution for a given problem, balance cost with other considerations, or grasp the full range of factors affecting their monthly bill. Tenants should take full advantage of billing and cost explorer tools provided by the CSP to understand current costs and estimate future costs; combined with strict resource tagging, these tools can help tenants achieve high traceability for costs. At the same time, tenants will find it useful to frame costs in terms of fundamental cost drivers, which are not always evident when using the billing and cost explorer tools.

Types of Costs

Most costs in the CSP's monthly bill fall under one of three fundamental cost drivers: compute, storage, and network.[41] Common compute costs include the hourly rate of a VM instance and the number of invocations to a serverless function. Common storage costs include the data stored in a general-purpose storage service (volume, storage class, replication, etc.) and the disk volumes that back VM instances. Common network costs include a general data transfer cost and additional costs for resources such as public/static IP addresses, peering connections, and VPN gateways; egress traffic from a CSP region is the dominant network cost as ingress traffic is free in most cases.

IaaS

IaaS CSPs commonly offer managed services that build upon basic services, ranging from managed databases to serverless functions to PaaS-like services; their pricing breakdown can be framed in terms of the above cost drivers. For example, a managed database service might charge for reads and writes (compute), the volume of data plus any snapshots (storage), and returning query results and exporting data (network); the costs are analogous to what a tenant would incur operating their own solution, but with a premium for the convenience of a managed service.

SaaS

While SaaS CSPs must pay for the compute, storage, and network infrastructure to run their applications, they present tenants with a different set of cost drivers based on the benefit the application provides rather than the resources it consumes. Common factors include the number of supported users, the availability of advanced features, and the speed of technical support. Subscription-based pricing is the prevalent model – typically with several tiers to choose from – although examples of pay-for-use exist (e.g., a payments application that charges per processed transaction). While most SaaS applications are delivered over the web, some can be deployed as appliances in an IaaS environment, where the tenant pays both subscription costs to the SaaS CSP as well as pay-for-use costs to the IaaS CSP.

PaaS

The pricing structure for PaaS varies depending on the CSP. The pricing may be closer to IaaS, as in an IaaS CSP that offers a PaaS service at minimal cost other than the compute, storage, and network resources consumed by the hosted application (which the CSP bills as usual). Alternatively, the pricing

[41] https://docs.aws.amazon.com/whitepapers/latest/how-aws-pricing-works/key-principles.html.

may be closer to SaaS, as in a dedicated PaaS CSP that offers to host the application on one of several performance tiers, each with progressively higher monthly subscription costs.

Staffing

In addition to the costs billed by the CSP, staffing is another significant cost driver. While IaaS offers more flexibility than PaaS, which in turn offers more flexibility than SaaS, time and money is required to exploit that flexibility. Tenants pay for the people that design, develop, test, deploy, monitor, and maintain a system and may even need to hire new staff or train existing personnel. With SaaS, some of these activities are less demanding (compared to PaaS or IaaS) or not necessary.

CISA Preference

To minimize new costs, CISA encourages agencies to take advantage of their existing integrated analysis process when implementing their chosen reporting pattern and to utilize their regional CLAW when possible and appropriate.

This CISA preference recognizes the potential for increased impact the NCPS program may have upon agency budgets. Reapplication of existing infrastructure, systems, and processes may guide the selection of the reporting pattern and pilot project that best meets the agency budget and needs.

APPENDIX I: AGENCY DATA RETENTION AND USE CONSTRAINTS

Telemetry data (including network flow logs) have the potential to rise to the level of sensitive information protected under various regulatory, privacy, and/or risk frameworks. As examples, telemetry may reveal:

- Mobile device location sensitive to welfare case workers
- Persistent connections to external parties indicating business relationships not publicly known
- IP addresses or even login names of patients or healthcare providers in telemedicine
- IP address ranges (and locations) of citizen-users reliant on other government services

Regulatory, Privacy, and Risk Determinations

At minimum, telemetry sharing should adhere to Official Use Only (OUO) guidelines and comply with the Federal Information Security Management Act (FISMA). The use of CSPs certified through the Federal Risk and Authorization Management Program (FedRAMP) can assist with FISMA compliance, as both are based on the NIST.SP.800-53 collection of security and privacy controls. The handling of agency data by CLAW and the standard mechanisms for sharing data with CLAW are consistent with these requirements.

Beyond that, telemetry may be subject to further privacy laws, such as the US Privacy Law and the Health Insurance Portability and Accountability Act (HIPAA). Designations other than OUO may also apply depending on the agency, such as Law Enforcement Sensitive (LES), Sensitive Security Information (SSI), and Critical Infrastructure Information (CII). Agencies must determine whether telemetry falls under one or more of these categories and communicate such to CISA. The agency, CISA, and, if needed, the relevant CSP(s), will then negotiate special handling constraints including data retention windows, permitted data transfer and storage encryption standards, data tagging and labels, personnel qualifications for analysts, minimum technical and/or administrative controls, etc.

Major regulation passed in recent years, such as General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), and similar state legislation do not normally apply to federal agencies. However, they may serve as the basis for future legislation enacted at the federal level.

Encryption Requirements: FIPS 140-3

The Federal Information Processing Standard (FIPS) 140-3, *Security Requirements for Cryptographic Modules*,[42] is a standard that is applicable to all federal agencies that use cryptographic-based security systems to protect sensitive information in computer and telecommunication systems (including voice systems).[43] This standard is developed and maintained by NIST. FIPS 140-3, Security Requirements for Cryptographic Modules, went into effect September 22, 2019. FIPS 140-2 modules can remain active for 5 years after validation or until September 21, 2026, but support and validation work for FIPS 140-3 modules must be in place by September 2020.It is a common requirement for data handling in

[42] https://csrc.nist.gov/publications/detail/fips/140/3/final; https://csrc.nist.gov/Projects/fips-140-3-transition-effort.
[43] As defined in Section 5131 of the Information Technology Management Reform Act of 1996, Public Law 104-106 and the Federal Information Security Management Act of 2002, Public Law 107-347.

government applications. To achieve FedRAMP certification (which a CSP must have to work with federal agencies), a CSP must use FIPS-validated cryptography.

For the telemetry data in transit from Cloud Sensing (Stage A) to Agency Processing (Stage B), agencies may decide to satisfy the requirements of FIPS 140-3 or a later version (whichever is available). However, the agencies may also decide to use another method of ensuring secure data communication, such as public key infrastructure (PKI) certificate, Transport Layer Security (TLS) 1.3, or comparable. Likewise, in the communication from Agency Processing (Stage B) to the CLAW (Stage C), similar security requirements are in place. However, once the data is transmitted to the CLAW, CISA will ensure adherence to FIPS 140-3.

CISA Preference

Agency cloud telemetry selected for sharing with CISA should adhere to Official Use Only guidelines. Agency special data retention and use constraints should be communicated to CISA prior to establishing any telemetry sharing.

This CISA preference addresses the sensitivity of the telemetry generated when agency applications and databases are relocated to the cloud. It sets a minimum sensitivity level and calls for formal documentation to specify the required protections to safeguard the shared telemetry data.