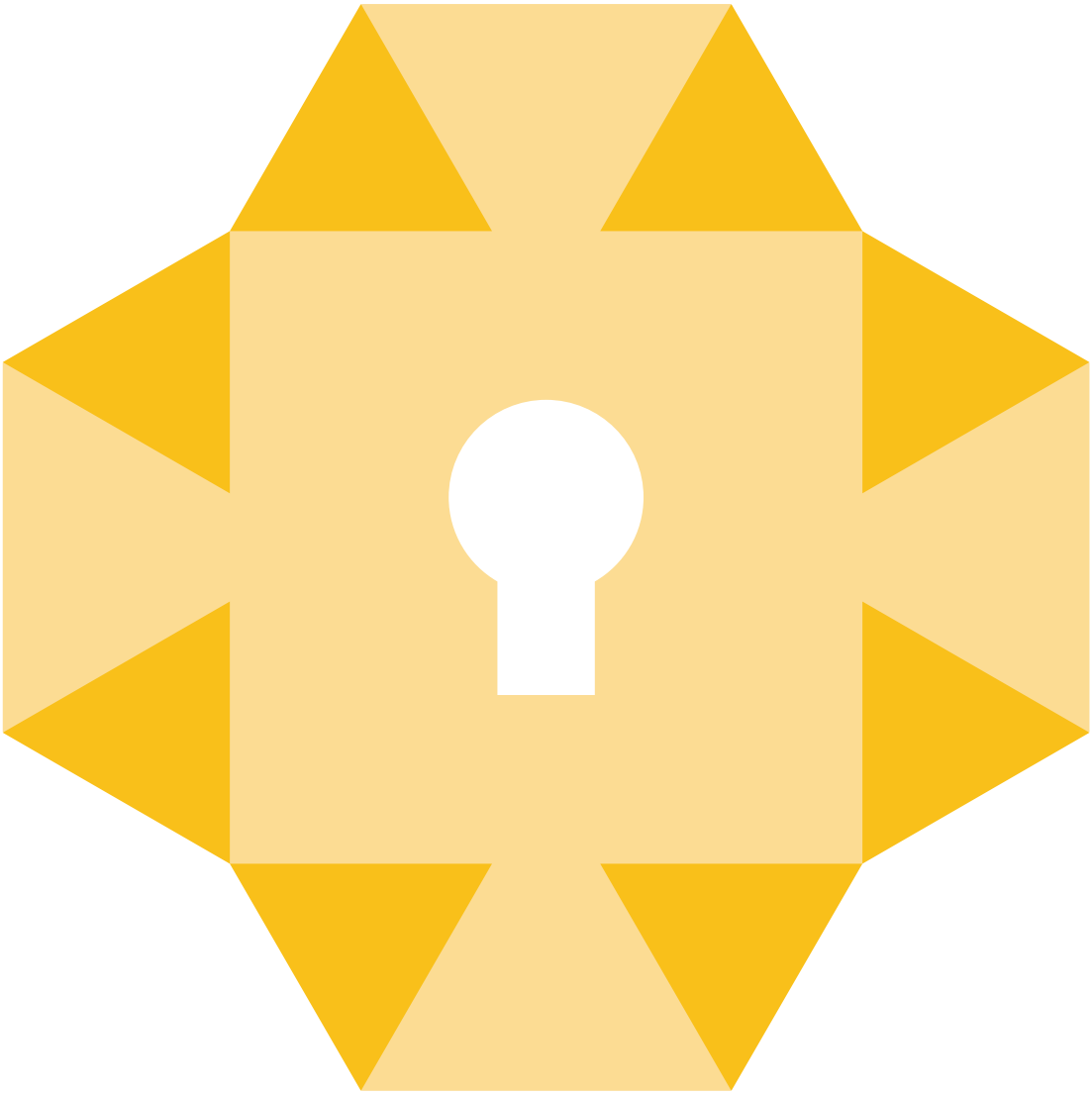

Integrated Cyber Defense Exchange

WHITE PAPER

Symantec™

Contents

Introduction 3

What is ICDx? 3

ICDx Instance 3

Architecture 3

Service Architecture 4

Schema 4

Collectors. 5

Archives 5

Queries, Search Conditions, and Filters 6

Forwarders 6

API Gateway and Actions. 7

ICDx Chaining 7

Extensibility 8

ICDx Installation 8

SaaS Support, and Cloud and On-Premises Deployment. 8

ICDx Console 8

Console Search and Filter Operations 8

Configuration 9

Settings 10

Security 10

Performance and Sizing 12

Introduction

Integrated Cyber Defense (ICD) is a Symantec strategic platform comprising a broad range of on-premises and SaaS security products, services, and partner ecosystems.

The size and diversity of the platform's products require an integration and interoperability framework that does not constrain the individual applications. That framework is the Integrated Cyber Defense Exchange (ICDx).

What is ICDx?

ICDx standardizes the interfaces to the ICD platform and its ecosystem. It solves the problem of point-to-point integration complexity via an event and API gateway that bridges internal and external applications. ICDx presents customers and technology partners with a single point of integration that provides:[1]

- Centralized data collection, normalization, and archiving for all integrated products
- Centralized data filtering, and data forwarding to customer SOC tools and data platforms
- Threat intelligence ingestion and publication
- A message bus and API for information exchange

Typical use cases for ICDx include:[2]

1. Collecting, normalizing, and archiving events across multiple applications, and forwarding as a single event feed
2. Sending different subsets of collected data to various destinations
3. Adding Symantec or supported third-party products without the need for downstream rules or analytics reconfiguration
4. Filtering PII or other GDPR-sensitive attributes
5. Supporting orchestration and automation products that invoke product actions
6. Communicating among applications via messaging

Applications and services *integrate* with ICDx and *interoperate* with each other or with external systems. This avoids combinations of point-to-point API calls among systems. ICDx bridges to application APIs, network ports and protocols, and databases, and does not impose changes on existing products and services.

ICDx has a simple console for operational statistics, data and schema browsing, collector and forwarder configuration, threat intelligence feeds, and action adaptor configuration. Products are not required to plug into a single administrative console or database.

Compressed, indexed archives are used to search, store, and forward collected data and can be used for long-term data retention.

ICDx Instance

Each instance of ICDx is a stand-alone deployment, packaged as installable software.[3] Multiple ICDx instances can be chained together via their message buses or over HTTP/S, and clustered and federated for common administration.[4]

Each ICDx node includes the following:

1. A message bus and programming interface
2. A standard, extensible attribute dictionary and event schema, and a standard language for command and control actions
3. A built-in set of collectors that receives or retrieves information from many data sources and normalizes to standard attributes and schema
4. A built-in set of forwarders that filters and sends information to multiple external destinations and message buses
5. A built-in set of action adaptors to translate standard commands to external APIs
6. An HTML web GUI and REST API

Architecture

ICDx is modular and quite simple. It uses the Extensible Protection and Management Platform (EPMP)—a Symantec service architecture and set of services for building single- or multi-tenant applications. Collectors and *forwarders* handle data sources and destinations, and *adaptors* support *actions* on applications.

Sources and destinations can be external or internal.

Examples of external sources include:

- Databases
- Web services
- Messaging systems
- Product-specific APIs

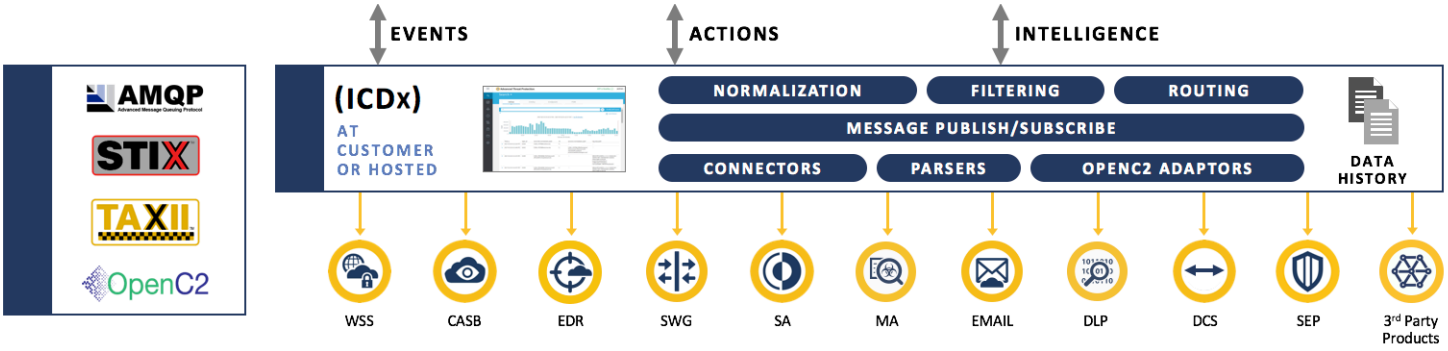

Figure 1: ICDx High-Level Architecture

Examples of external destinations include:

- SIEMs
- Service desks
- Data lakes
- Streaming analytics systems

ICDx bridges to external sources and destinations via built-in connectors for each class of source and destination.

Internal communication is provided by a built-in AMQP forwarder that publishes data to an exchange, and a built-in AMQP collector that consumes from a queue bound to an exchange. Message bus exchanges and queues can span multiple ICDx instances or be limited to a single instance. Applications can integrate with the ICDx message bus via AMQP network protocol bindings (SDKs), or the ICDx HTTP API. Figure 1 illustrates the high-level ICDx architecture.

Service Architecture

ICDx is a native EPMP application whose services are integrated via a message bus control plane. The message bus is also used as a data plane for internal collectors and forwarders.

The REST API and identity-related functionality is implemented via the EPMP Rest Request Router (R3) service, and the EPMP Identity service. Authorization is performed via OAUTH2 tokens generated by the Identity service. The event pipeline, or Data Exchange (DX) service, is an enhanced EPMP Event service with additional plug-ins. An Archive service is used to store and forward collected data, or to provide optional long-term data retention. All ICDx console operations are performed via the REST API. Refer to EPMP documentation for EPMP architecture and services details.

Schema

Perhaps the most important ICDx aspect is its standardization of the rich information produced by the ICD platform products. The Integrated Cyber Defense Schema is organized into sets of attributes, objects, event types, and categories.

Fields extracted from product logs and telemetry are mapped to a standard dictionary of attribute names, and their values are normalized to standard data types, defined ranges, and enumerations. Attributes that model common entities such as files, processes, emails, and network connections are combined into reusable objects. Scalar attributes and objects are then combined into event types that impart meaning to the behaviors that produced the events. Finally, event types are classified into high-level categories such as Security, System Activity, and Compliance.

While there are comprehensive information models for management and applications—such as the Distributed Management Task Force (DMTF) Common Information Model—or threat intelligence, such as the Structured Threat Intelligence Exchange (STIX) standard, there is no such equivalent for event information.

Threat intelligence is well modeled via STIX, and ICDx supports STIX and TAXII transport.[5] The Integrated Cyber Defense Schema complements STIX when matching indicators of compromise with actual observables in the environment modeled by the ICDx Schema and collected as events within the system. The standardization of events across products makes integration much easier.

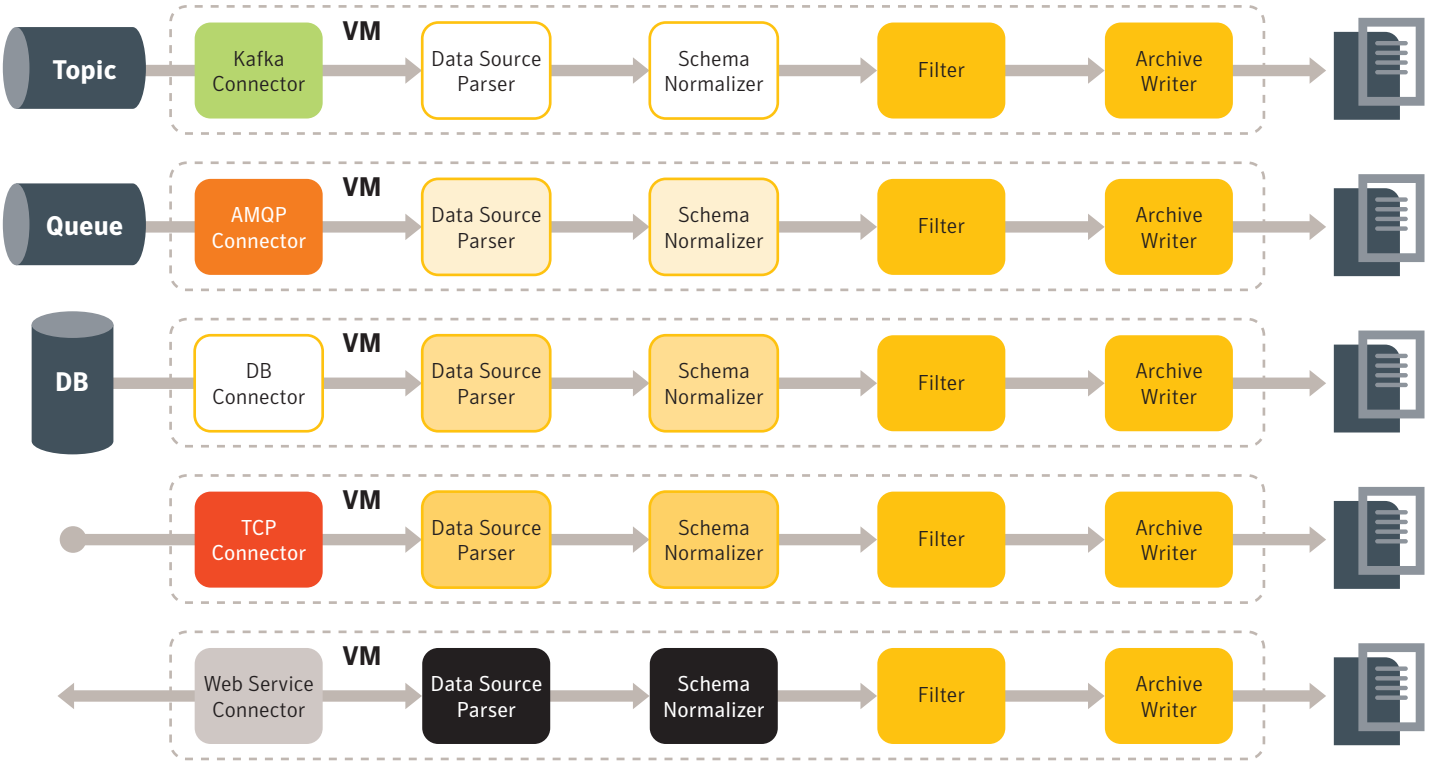

Figure 2: Collector Architecture

Collectors

ICDx ingests events by connecting to a variety of data sources, pulling or capturing data, and then parsing, normalizing, and storing the information. The ICDx component that performs this function is a *collector*. For performance reasons, built-in collectors bypass the message bus and stream directly to a store-and-forward archive subsystem.

Collectors can be either data source-specific or generic. Data source-specific collectors parse and normalize the specific data streams into the Integrated Cyber Defense Schema, while generic collectors expect the data to be prepared externally before ingestion. Data source-specific collectors are built-in for many Symantec products, such as Symantec Endpoint Protection, Secure Web Gateway, and more. Generic collectors are intended for add-on products external to ICDx.

Each collector can optionally retain the raw data prior to parsing and normalization in a raw_data attribute. Any parsed fields that are not normalized are retained in a product_data attribute. These attributes are handled as any other schema attribute; however, their values are not tokenized and are string types.

There can be multiple instances of each collector, corresponding to each data source being ingested.

Collectors have the following subcomponents:

- Data source connector
- Parser
- Normalizer
- Filter
- Archive writer

Each connector type handles the protocol to connect to the data source. A generic collector has a pass-through parser and normalizer, while data source-specific collectors have a priori knowledge of the form and meaning of the data source. Each instance of a collector runs in its own process. Figure 2 illustrates the ICDx collector architecture.

Archives

Collector instances stream data to archives. An archive is a time-series set of files containing the normalized data, along with a corresponding set of index and statistics files. The archive files are compressed and allow for quick time-span access. The event data is then accessed via the index files, which maintain offsets into the compressed data files.

An archive file set is open while data is ingested and stored. After a certain time period elapses, by default 2 hours, or a maximum data size is reached, by default 512MB, the archive file set is closed and a new file set is created to store the next time period. Each archive has lifecycle management settings. For example, data can be purged based on retention time, free disk space, etc.

ICDx allows an archive configuration to point to any Linux-mounted volume. Data will then be archived to that volume. By default, a local volume is configured, but a volume mounted on NAS or SAN storage is supported for longer-term retention policies, high availability, and disaster recovery (for example, using storage system snapshots).

Each collector can have a dedicated archive or it can use a default archive. Collector-specific archives perform better and are more easily managed. For example, higher volumes of data or long retention times may benefit from archives mounted on larger data volumes, while data that is only stored and forwarded may not require its own archive and can be purged shortly after forwarding.

If a collector is subsequently deleted, the archive is preserved and is accessible for query and forwarding, even though no new data will be captured.

Queries, Search Conditions, and Filters

In general terms, a query combines a time span, a search condition, and a filter condition. Within ICDx, queries are used in the Search view, Top N view, collector configurations, and forwarder configurations. A query created in a Search view may be saved for reuse in collector and forwarder configurations.

Queries leverage the power of the standardized Integrated Cyber Defense Schema to refine a data stream. Since data is normalized into the standard schema across data sources, queries can operate against well-known fields independent of a particular source.

Searches retrieve data based on indexed fields of events in an archive; filters discard events or extract attributes after data have been searched and retrieved. Filters do not require fields to be indexed.[6] In addition, filters can include or exclude individual attributes from events.

Filters can be used when only certain attributes of interest are to be forwarded to an external system, such as when sending events to a public cloud storage that requires personal identifiable information is removed to meet privacy and compliance requirements.

Forwarders

After ICDx captures and streams data into an archive via a collector, all or part of the data may be routed to other systems via forwarders. A forwarder queries the archives and routes desired data to a destination using its associated protocol. Each forwarder is its own communication channel running in its own process. Selecting particular archives can improve the efficiency of the query. Forwarders provide the ability to route different sets of data to multiple destinations simultaneously.

For example, all Category = Security events with Severity > Minor can be forwarded to ServiceNow for ticket creation. Category = Security AND Category = System Activity can be forwarded to Splunk for rule processing, and separately to Elasticsearch for cluster analysis, while Product = DLP OR Product = SEP is forwarded to Symantec Information Centric Analytics (ICA) for User Behavioral Analytics (UBA).

Forwarders have the following subcomponents:

- Archive selection/reader
- Archive search
- Filter
- Connector

As with collectors, forwarders are configured as instances of a class of forwarders. There are forwarders for popular event consumers such as Elasticsearch and Splunk, message queues such as RabbitMQ and Kafka, cloud services such as ServiceNow, AWS S3, and generic output to JSON files.[6]

An updating marker selects the time span and a search condition selects event data using indexes across the archives. A filter then refines the data before handing off to the connector, which communicates to the forwarder destination.Figure 3, on the next page, illustrates the ICDx forwarder architecture.

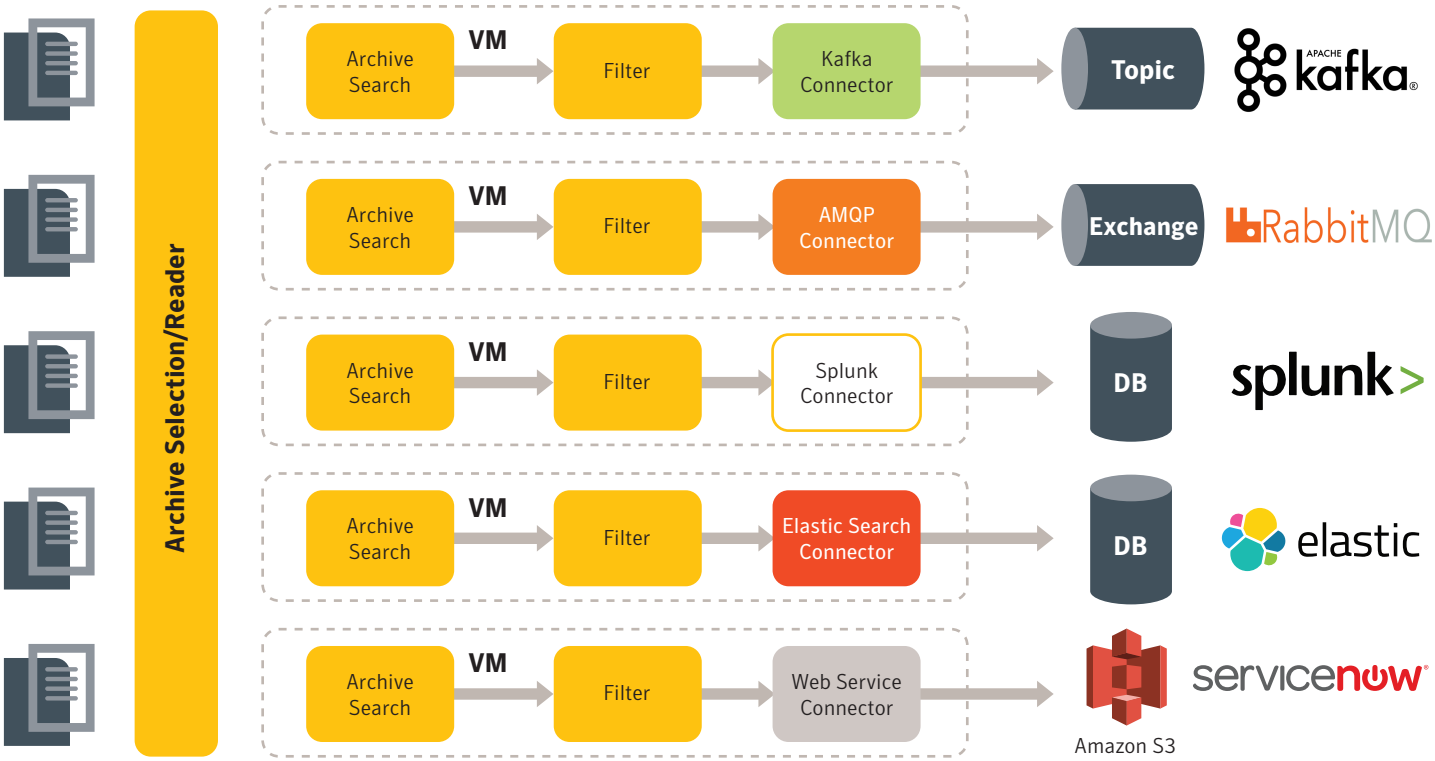

Figure 3: Forwarder Architecture

API Gateway and Actions

ICDx abstracts individual API calls to underlying products using the API Gateway to invoke actions. Consider the API Gateway as the inverse of the data collectors. While collectors aggregate data from different sources into the common Integrated Cyber Defense Schema, actions fan out from the API Gateway to one or more product-specific APIs that have the same or similar semantics.

ICDx supports the draft version of the OpenC2 standard, in which API calls are standardized by action to different targets using actuators. Physically, ICDx uses adaptors to represent OpenC2 actuators. Specific adaptors are implemented per product, which translate standard calls to product-specific calls. Product termination points are categorized by standard OpenC2 targets. Using this pattern, a common action call can be routed to product termination points and invoked in the context of each product. ICDx can configure actions to be invoked on multiple targets simultaneously.

Actions supported by ICDx are allow, deny, contain, query, and remediate. Targets include device, file, process, URL, directory, email message, registry key, and others.

One use case is to match event data, including historical archived event data, with IOCs from one or more threat intelligence sources. This can be done by applications written on ICDx via the ICDx ecosystem of security partners.

ICDx Chaining

ICDx instances may be connected together or chained. As with ordinary forwarders, selected information from the archives of one instance may be copied to another instance. For large enterprise environments with many products and large data volumes, or for complex network topologies, ICDx chaining may provide better performance and load distribution, as well as data segregation.

For example, Symantec Secure Web Gateway (SWG) events may be stored at one instance in a dedicated archive, while forwarding only a portion of those events to another ICDx instance in another site—from there, a forwarder can include that stream with other data sent to Elasticsearch.

Extensibility

ICDx is an open system that delivers built-in functionality for Symantec and selected third-party products. It is a foundation piece for the Symantec Technology Integration Partner Program (TIPP). Extending ICDx to support TIPP products or customer-specific applications is most easily accomplished by integrating into the ICDx message bus using the AMQP forwarder and collector support.

To date, the types of products integrated via the message bus forwarder are SIEM and TIP products. Third-party product data sources can be similarly integrated via the message bus collector.

Examples of Java- and Python-based collectors and forwarders are included with the ICDx installation package.[7]

ICDx Installation

ICDx is written in Java. The ICDx installer includes the following open-source dependencies:

- Java Runtime Environment OpenJDK 8
- NGINX
- RabbitMQ (including Erlang runtime)

These are not prerequisites; the installer retrieves and installs these dependencies, configures and starts them, and starts the ICDx services. Optionally, pre-installed components that meet a minimum criterion may be used.

SaaS Support, and Cloud and On-Premises Deployment

ICDx instances support both SaaS applications as well as on-premises applications. ICDx instances may be hosted by the customer in a public or private cloud, or in the customer's data center. Connectivity to data sources and destinations is required.

For certain use cases, a hosted multitenant deployment is planned. These use cases depend on the applications being integrated and the customer environment. An all-SaaS environment would not normally require a customer-hosted node. A hybrid environment, however, requires a customer node to access applications within the firewall or within the customer cloud boundary.

ICDx Console

After ICDx is installed, the console is available via a web browser to configure the overall system of collectors, forwarders, queries, etc. The ICDx console is primarily an administration tool but it also can search and display data from archives. This capability is not intended to serve as a primary dashboard or search tool, but rather to ensure the data being collected and forwarded is what is desired.

Typically, a user will use the search capabilities to set up queries for forwarders, and to monitor collector activity. The purpose of ICDx event functions is to get the right data to the right system, and as such functions as data middleware. The console is fully API driven; every console function is available as an API.

The following functions are provided:

- Dashboard for important metrics
- Search view for query, display, and event details of collector and archive data
- Top N view for counts and details of events by different attributes
- Configuration for setting up collectors, forwarders, threat intelligence and actions
- Settings for Active Directory, API keys, and default archive management

Console Search and Filter Operations

Both the Search view and the Top N view can search and filter collected data across archives and display the events in the console.

The Search view can search across selected archives associated with specific collectors, or across all collectors within time spans, and display the resulting counts that match the search criteria in a time-series histogram. Clicking a bar in the histogram displays the individual events and enables the user to display the columns for the attributes of each event. In turn, each event can be inspected by right-clicking and viewing all the populated fields for that event; alternatively, quick filters can be chosen based on the particular value of a column (for example, Severity > Minor or IP = 192.168.205.221).

The query can then be saved and reused in this view or in a collector or forwarder configuration.

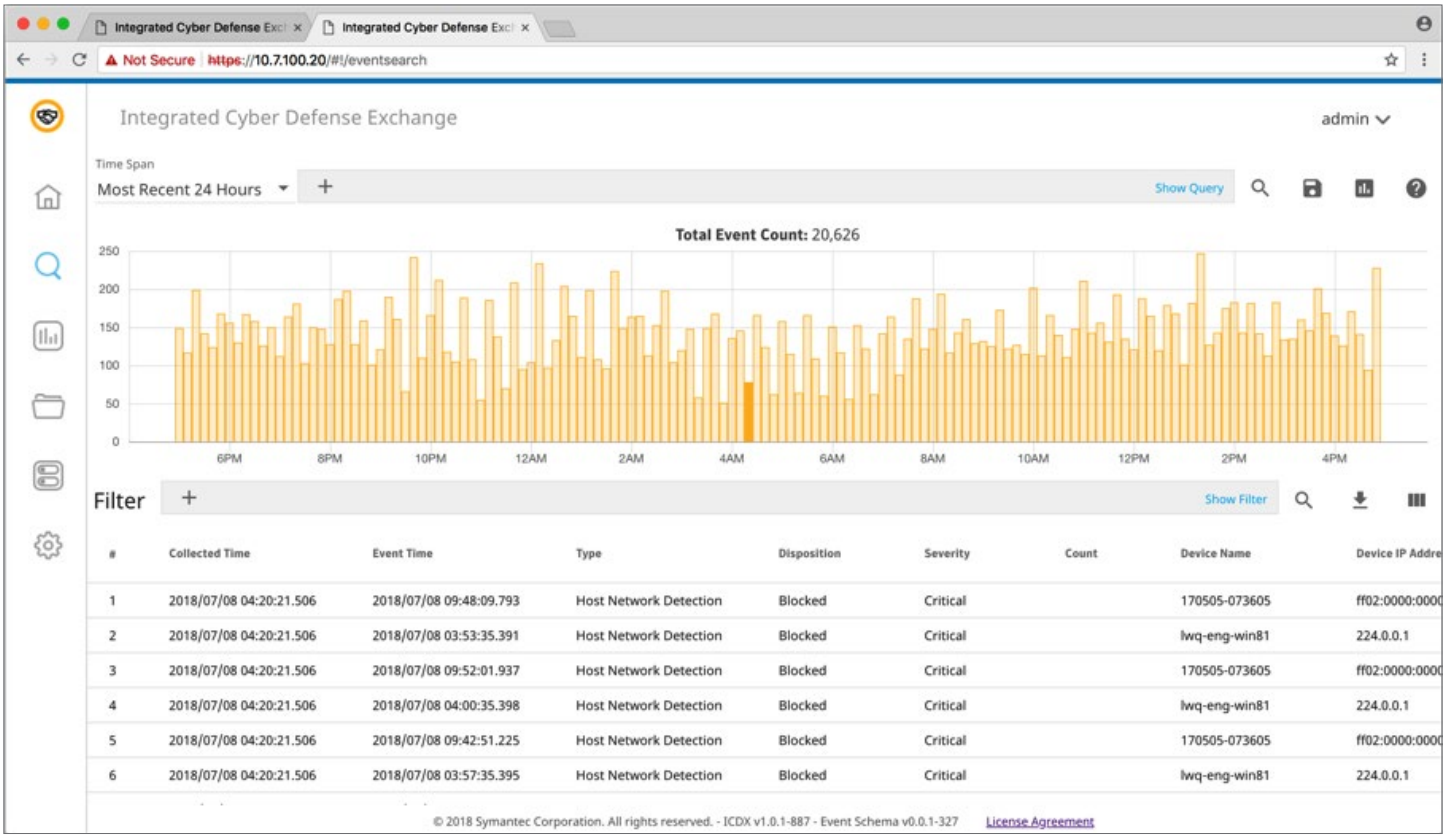

Figure 4: Console Search View

The Top N view aggregates events by indexed field across selected archives associated with specific collectors, or across all collectors within time spans, and displays the top counts in a histogram with similar drill-down and filtering options to the Search view. For example, the top five events can be counted over the last 12 hours by event type, product name, or severity.

Configuration

The heart of the ICDx console is the configuration of collectors and forwarders. Each class of collector or forwarder is shown in its respective tab of the configuration view. Within each class, one or more instances of a collector or forwarder is set up with simple dialog boxes where connection information is configured.

For example, a Symantec Endpoint Protection (SEP) collector minimally requires the database IP address or name, and the credentials needed for reading the database tables. More advanced options allow for filtering the collected data, storing data in separate archives, polling intervals and batch sizes when querying the database, etc. Configurations for Secure Web Gateway require setting the IP address and port that will receive the access log data, and configurations for web services such as Email Security Service require endpoint and API key.

A forwarder configuration for Splunk requires the IP address and port, and Splunk API key with optional index name and source type of the destination. Other collectors and forwarders have similar configurations based on their respective connectors.

RabbitMQ - AMQP exchange and queue are special types of forwarder and collector, respectively. Data may be forwarded (as with any other forwarder) to an exchange on the local RabbitMQ message bus, or a remote message bus, by specifying an exchange name. Data may be collected from an exchange by creating a RabbitMQ queue collector, which binds to the exchange. This is the most general way of exchanging information among ICDx data sources without having a specific class of collector or forwarder.

An example use case for an exchange forwarder is to send selected data to an external system not directly supported by ICDx—for example, a UEBA—that can be fed via a Python program that binds to the exchange and sends data via the specific APIs of those systems.

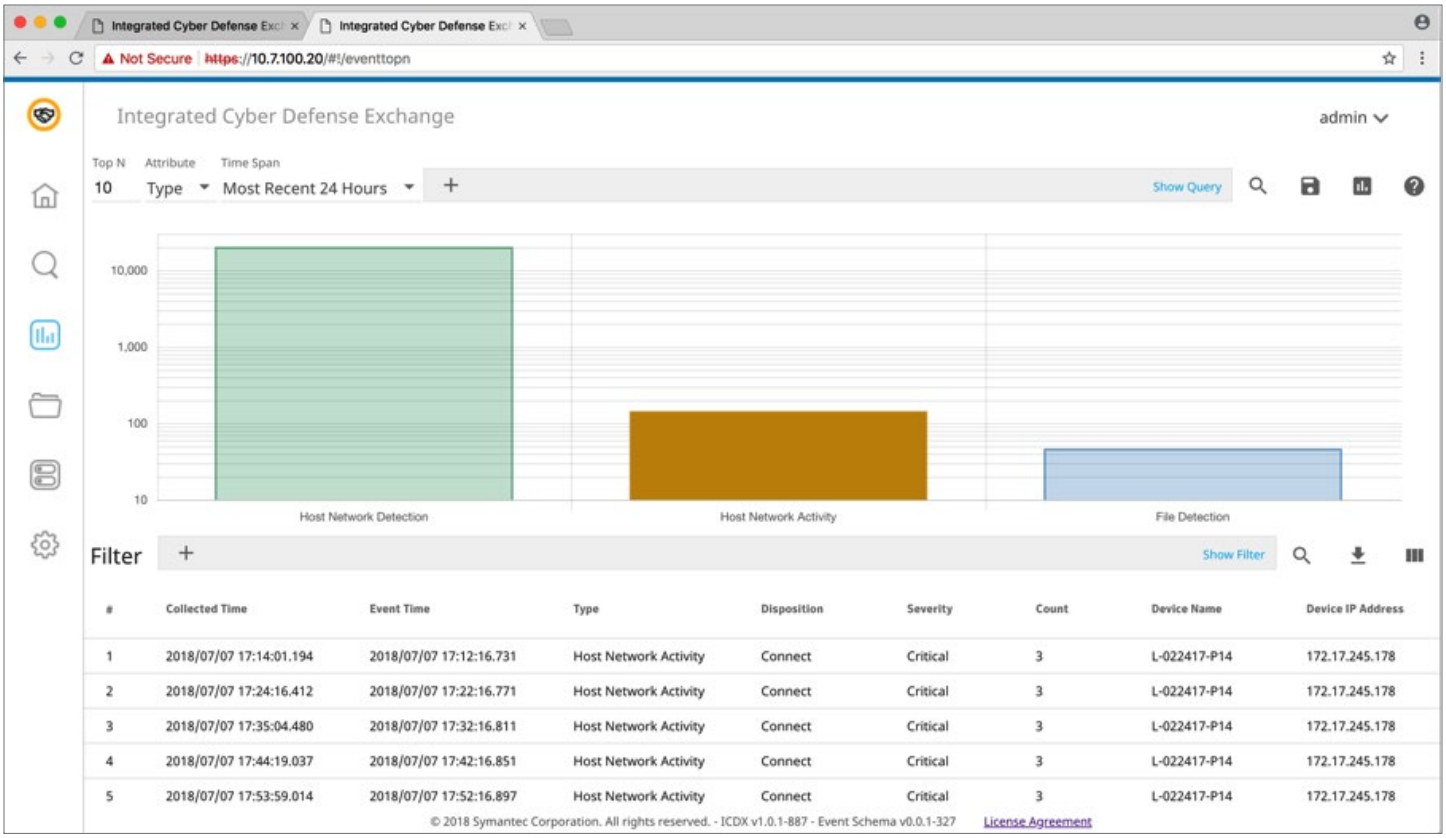

Figure 5: Console Top N View

An example use case for a queue collector is the reverse: An external system publishes data to an exchange and an ICDx queue collector binds to that exchange and archives the data (as with any other collector). That data, if prenormalized into the ICDx Schema, is available to a forwarder, such as the Elasticsearch or Splunk forwarders, in combination with any other data collected.

Settings

Console settings include the following tabs:

- Event archive management for the default archive
- Active Directory settings for enterprise login
- API key generation for HTTP or AMQP external system interface

As can be seen in Figure 7, access to the ICDx console can be restricted to users in a specified Active Directory group.

Security

The ICDx install must be performed as root on the Linux system. The installer creates an ICDx user under which all the ICDx processes run. The ICDx user owns the files and folders created by the ICDx processes. No direct login is permitted as the ICDx user.

In order to access the ICDx files and folders, a user must sudo to create a shell as this user. For example, sudo –E –u –s ICDx to copy archives to another volume and view a configuration file. The data is not encrypted at rest and is protected by OS access control only.

In order to log in to the console, either a built-in user and credential can be used, or an Active Directory credential. ICDx uses OAUTH2 tokens to grant privileges to the console. Login requires membership in a specified AD group for the token to contain the necessary privileges.

Transport security is performed via TLS/SSL. The ICDx install creates a self-signed certificate that can be replaced with a third-party CA signed certificate. Both HTTP/S and AMQP/S are supported.

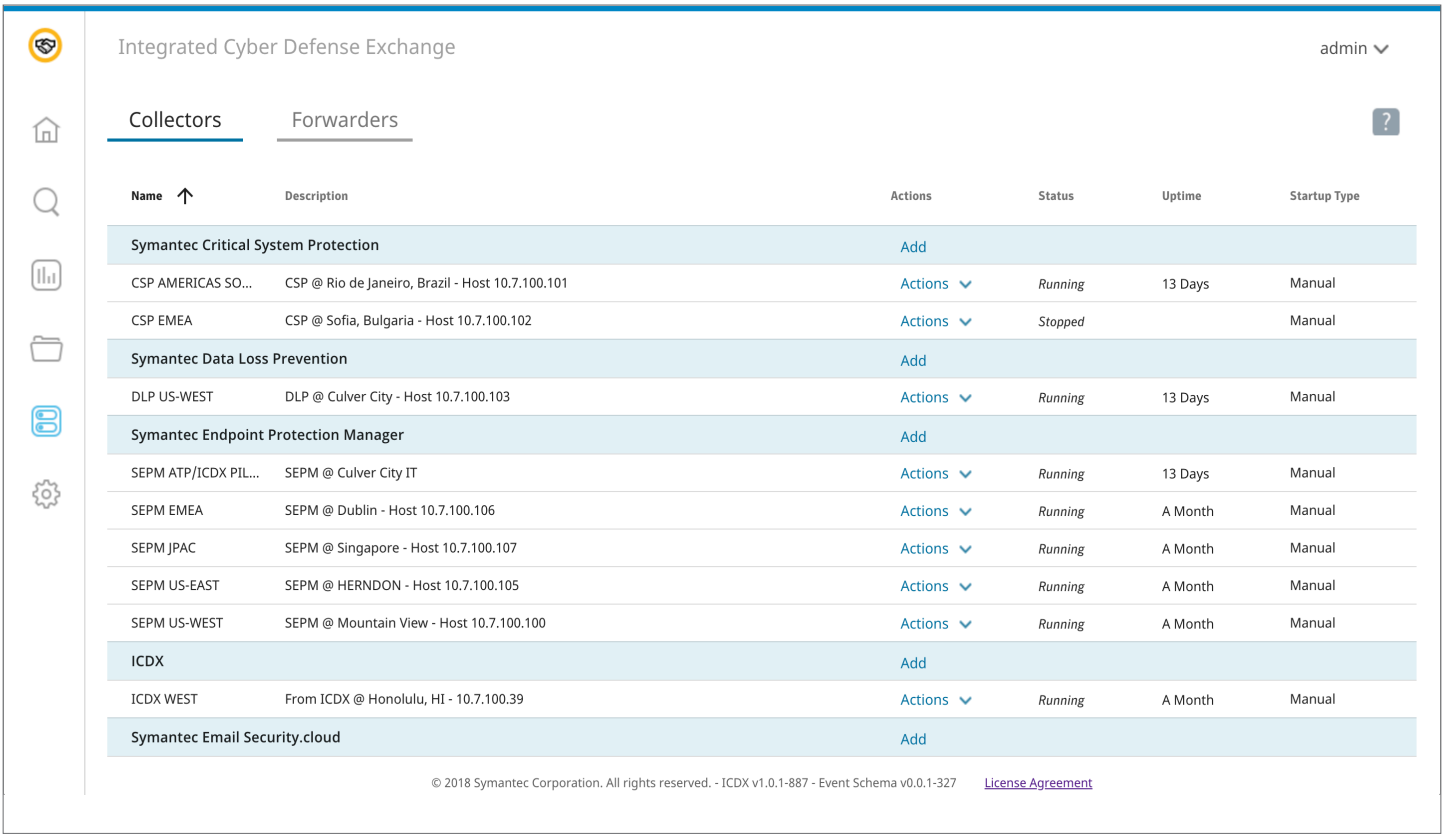

Figure 6: Configuration View

The embedded RabbitMQ broker runs under the ICDx user; its administrative credential can be set at install time. Using this credential and the RabbitMQ portal, access may be restricted to individual exchanges and queues.

The ICDx machine is not hardened by the ICDx installer in any way. It is recommended that only the necessary ports are accessible, OS patches are kept current, and other best practices are maintained.

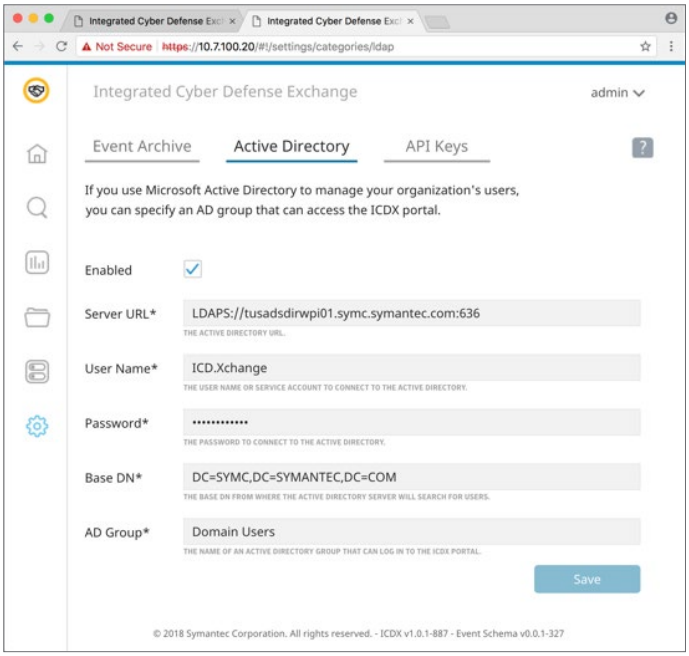

Figure 7: Settings View

Performance and Sizing

There is no single measure for ICDx performance, which will depend on the class of computer it runs on, and the throughput and latency of the collector and forwarder connections, including any collector throttling that is configured. In addition, the archive performance and capacity will depend on the I/O systems for the local or remote archive mount points.

In addition, the number of collectors and forwarders running will affect overall performance of a single ICDx instance.

To get an estimate of a particular collector running in isolation, consider this: An ICDx instance running on a typical four-core virtual machine with 8GB of memory and local storage can stream many tens of thousands of events per second to an archive.

Overall performance and throughput can be increased by adding more collectors, using dedicated archives, and chaining multiple ICDx instances. Load balancing can be achieved by chaining together multiple ICDx instances such that separate instances segregate the collection and archiving load, while subsets of data are forwarded to a central instance or separately forwarded to external destinations such as a customer-managed data lake.

Disaster recovery should be handled using reliable media for archives, and regularly backing up configuration files.

[1]See endnotes for current ICDx capabilities.
[2]Version 1 of ICDx is focused on event schema, event collection, and forwarding. A subsequent release will include actions and threat intelligence.
[3]The first release of ICDx supports Linux via a software installer (a single tarball file) that installs and starts up in about two minutes. A future deployment model will include containers.
[4]Federation is not available in the first release of ICDx.
[5]Threat intelligence is not supported in the Controlled Availability release of ICDx.
[6]In Version 1 of ICDx, indexes are predefined on important common fields such as log time, severity, event type, and other attributes.
[7]Version 1 of ICDx supports Elasticsearch, Splunk, ServiceNow, RabbitMQ, and JSON files.

About Symantec

Symantec Corporation (NASDAQ: SYMC), the world's leading cyber security company, helps organizations, governments and people secure their most important data wherever it lives. Organizations across the world look to Symantec for strategic, integrated solutions to defend against sophisticated attacks across endpoints, cloud and infrastructure. Likewise, a global community of more than 50 million people and families rely on Symantec's Norton and LifeLock product suites to protect their digital lives at home and across their devices. Symantec operates one of the world's largest civilian cyber intelligence networks, allowing it to see and protect against the most advanced threats. For additional information, please visit www.symantec.com, subscribe to our blogs, or connect with us on Facebook, Twitter, and LinkedIn.

Symantec

350 Ellis St., Mountain View, CA 94043 USA | +1 (650) 527 8000 | 1 (800) 721 3934 | www.symantec.com

19D219599_wp_ICDx_EN