FACULTAD DE INGENIERIA

UNIVERSIDAD DE LA REPUBLICA

Advanced Threats Information Sharing and Collaboration

Autor:
Julio Saráchaga

Supervisor:
Dr. Gustavo Betarte

Contraparte del cliente:
Ing. Fernanda Molina

Supervisor alterno:
Ing. Marcelo Rodríguez

Brief Article

Julio Saráchaga

Glosario

- Botnet: Es una colección de computadoras comprometidas (llamadas computadoras zombie), instaladas generalmente mediante gusanos, troyanos y backdoors, controladas remotamente, usualmente con fines maliciosos, en especial para denegaciones de servicio.
- CAPEC: Es una lista pública desarrollada por la comunidad de patrones de ataque comunes así como un esquema comprensivo junto con una clasificación.
- CSIRT: Es una organización responsable de recibir reportes de incidentes de seguridad, analizarlos y responder a ellos. [CERT/CC]
- CVE: El CVE (Common Vulnerabilities and Exposures) es un diccionario de comunes (por ejemplo los identificadores CVE) para publicar información conocida de vulnerabilidades de seguridad. El CCE (Common Configuration Enumeration) provee identificadores para problemas de configuración de seguridad. Los identificadores hacen más fácil compartir información a través de diferentes bases de datos y herramientas de seguridad.
- Cyber Observables:
- Cyber Threat information: Es cualquier información representable en STIX. Esto incluye, pero no esta limitado a, observables, indicadores, incidentes, TTPs (Tactics, Techniques y Procedures), exploit targets, campaigns, threat actors y courses of action.
- TAXII Data Feed: Es una colección de cyber threat information estructurada expresable en uno o mas documentos STIX que pueden ser intercambiados utilizando TAXII. Cada TAXII Data Feed *debe* tener un nombre que lo identifica de forma única entre el resto de los feeds de un productor dado. Cada elemento de un TAXII data feed debe ser

etiquetado con un timestamp y puede tener otras etiquetas a discreción del productor.

- IDS: Intrusion Detection System. Sistema que detecta intrusiones no deseadas en una red o equipo, que no pueden ser detectadas por un firewall convencional.
- MAEC: Es un lenguaje estandarizado para codificar y comunicar información de malware basandose en atributos como comportamiento, patrones de ataque, etc.
- Mensae TAXII: Un bloque de información que es pasado de una entidad a la otra. Un mensaje TAXII representa un pedido o una respuesta.
- Intercambio de mensajes TAXII: Una secuencia definida de mensajes TAXII intercambiados entre dos entidades.
- Servicio TAXII: Son funcionalidades albergadas por algunas entidades y que es accedido o invocado usando uno o mas TAXII Message Exchange.
- TAXII Capability: Una actividad de alto nivel soportado por TAXII por medio del uso de uno o mas servicios TAXII.

1. Introducción

Históricamente, se ha visto un aumento en la sofisticación, velocidad, impacto y cantidad de los ataques informáticos. Por ello es necesario que las estrategias de defensa se adapten a los nuevos actores y ataques. Para responder a esto, las organizaciones han tenido que recurrir al intercambio de información de amenazas con el fin de tener un mejor panorama de las actividades de sus adversarios y así ayudar a administrar sus recursos de forma de obtener el mejor resultado posible de sus defensas.

Las aproximaciones tradicionales de seguridad tienen como objetivo entender y registrar vulnerabilidades, debilidades y configuraciones necesarias pero insuficientes. Para defenderse, las organizaciones tienen estrategias basadas en alertas que buscan bloquear los ataques y arreglar las vulnerabilidades. Si bien dichas estrategias pueden ser efectivas contra algunas amenazas no logran detener ataques avanzados o proveer información sobre las actividades de un atacante luego de que ingresó a la red. Una estrategia más adecuada es basarse en **cyber kill-chain**, en esta estrategia se busca descomponer las fases de un ataque con la finalidad de obtener una mejor comprensión del ataque y el atacante, así como mejorar las posibilidades de defensa.

[scale=0.55]./images/killChain.png

Figura 1: Cyber kill-chain [2]

Como se ve en la figura 1, los primeros pasos de esta estrategia representan una oportunidad para detectar y mitigar las amenazas de forma proactiva antes de que al adversario realice un acceso no autorizado en los sistemas de la organización. En los pasos posteriores es donde se realiza la detección, respuesta y aseguramiento de los activos más importantes. Al entender al adversario, se puede tener una mejor oportunidad para descubrir sus intenciones y responder al ataque. Entender la amenaza permite la toma de mejores decisiones, dando prioridad a los recursos y así tener una ventaja ante el atacante. El resultado de defensas en base a inteligencia es una mejor opción que las respuestas a los ataques dado que estos ajustan sus operaciones basándose en el éxito o falla de sus intentos. En un modelo como el de la figura 1 los intentos del adversario pueden ser reconocidos, logrando que los defensores tengan la posibilidad de ajustar sus tácticas para que al adversario le sea más difícil alcanzar sus objetivos.

Compartir información con socios y comunidades de confianza le permite a las organizaciones tener un conjunto de datos relevante para lograr una

identificación precisa de la amenaza. Por medio de este intercambio, cada organización puede entender mejor las amenazas, no solo de forma abstracta sino que también con evidencias específicas que indiquen la presencia del atacante. Las tareas de cyber inteligencia permiten anticipar y mitigar las amenazas antes de que sean difíciles de encontrar y erradicar utilizando los métodos tradicionales de detección y respuesta. Con la información recolectada los analistas pueden agrupar patrones de actividades similares, atribuir actividades a ciertos actores, identificar e implementar estrategias para mitigar ataques de forma rápida y anticiparse al lanzamiento de ataques similares en el futuro. Para aprovechar de forma adecuada los beneficios de la cyber inteligencia, las organizaciones deben compartir la información recolectada con socios de su confianza.

Por medio del análisis del comportamiento de los adversarios en distintos objetivos y en un período de tiempo adecuado, los defensores son capaces de identificar un conjunto importante de indicadores, tácticas, técnicas y procedimientos. De esta forma se obtiene información de los objetivos y las estrategias lo cual permite al defensor predecir el comportamiento del ataque y generar defensas dinámicas. Dada la forma y complejidad con la cual evolucionan las amenazas, la velocidad con la cual ocurren los eventos y la basta cantidad de datos que se deberían intercambiar, es necesario establecer una forma automática para ayudar a los analistas y tomadores de decisión a llevar a cabo acciones defensivas.

Existen múltiples métodos para el intercambio de información y estos juegan un rol importante en los tipos, volúmenes y naturaleza de la información compartida con la comunidad. Algunos de estos medios de intercambio limitan el tipo de contenido que es compartido mientras que otros promueven ciertos tipos de intercambio. Los procesos para compartir información son manuales, llevan mucho tiempo, son repetitivos y en muchos casos requieren que las organizaciones reescriban o traduzcan la información a una amplia variedad de formatos. Otro problema que se presenta cuando se quiere intercambiar información es la utilización de tecnologías y/o formatos propietarios, esto provoca la necesidad de desarrollar scripts y módulos para permitir compartir información por fuera de las comunidades. Para aquellas comunidades con algún grado de automatización, sus modelos son generalmente bajos en prestaciones y usan soluciones propietarias, comerciales o adaptadas a su comunidad.

La mayoría de dichos métodos no permiten el consumo de información de amenazas de forma automática, esto hace que rutinariamente las organizaciones deban tomar dicha información y sintetizarla en sus bases de datos locales. Si bien los métodos utilizados en la actualidad han ayudado a me-

jorar las capacidades defensivas de numerosas organizaciones, éstas no han logrado explotar su máximo potencial.

La automatización requiere de información de calidad, esto no puede ser logrado adecuadamente con los distintos productos y sistemas de hoy en día. Para lograr un nivel adecuado de automatización es necesario contar con un estándar el cual tenga representaciones estructuradas de información, de ésta forma se puede lograr un mejor aprovechamiento de los datos sin saber de antemano quien los provee. Dicha información debe ser legible por un humano y parseable por una máquina. Estos requerimientos tienen varias justificaciones, primero que nada, un analista podría realizar un análisis que es inapropiado para ser automatizado o que sea focalizado en tomas de decisiones por parte de personas. También podría ser de interés que un analista tenga conocimiento de la situación actual, de la fidelidad de las fuentes o de los métodos utilizados para producir la información. Por lo dicho anteriormente, es necesario la existencia de un estándar con una representación estructurada de la información, siendo ésta expresiva, flexible, extensible, automatizable y legible. Además se deben contar con medios para permitir el intercambio seguro y confiable de información entre distintas organizaciones.

Si bien existen varios esfuerzos para crear herramientas con las características presentadas, la realidad es que ninguno de ellos ha logrado su cometido. El objetivo de dicha herramienta debería ser:

- Permitir compartir información de forma más rápida y precisa.
- Reducir el análisis humano y liberar a los recursos humanos para realizar trabajo de análisis más valioso.
- Permitir que las amenazas más conocidas sean analizadas por computadoras.
- Permitir que se comparta de forma automática un gran número de datos, siendo estos datos complejos. Esto debería permitir una defensa activa.
- Proteger la información intercambiada.
- Permitir que se agregue información a las bases locales con discreción y limitando el número de analistas que acceden a la información. Dicha información debe tener datos de contexto.
- Permitir la colaboración de analistas de distintas organizaciones en los incidentes.

2. Gestión de incidentes

2.1. Herramientas

2.1.1. Que es RTIR

RTIR es un sistema de manejo de incidentes diseñado para ser utilizado por los equipos de seguridad de sistemas. A sido creado en conjunto con equipos de CERT y CSIRT para manejar el creciente número de incidentes reportados. Presenta la ventaja de ser opensource, contener una API completa y una comunidad de usuarios grande y experta. Además es simple de integrar con otras herramientas existentes. Está implementado por medio de módulos PERL y las herramientas que provee RT. Se puede pensar en RTIR como una extensión de RT para ser utilizada por CERT's y CSIRT's.

Existen algunas alternativas a RTIR como lo son AIRT (Application for Incident Response Teams) cuya última versión data de Julio de 2009. AIRT es una aplicación web desarrollada para los equipos de respuesta a incidentes. Busca proveer facilidad ante los reportes de incidentes de seguridad así como un seguimiento simple de estos. Este sistema no cuenta con una comunidad comparable a la de RTIR así como con documentación tan extensa como la de RTIR.

Otra de las opciones existentes es OTRS (Open Technology Real Services), así como los anteriores también es open soruce. Presenta las ventajas de tener una comunidad más numerosa que AIRT y que el código esta siendo desarrollado continuamente. Así como RTIR esta desarrollado por medio de Perl y permite conectarse a varias bases de datos. Presenta documentación extensa para implementadores.

3. Representación

3.1. Modelos de datos

3.1.1. IDMEF

Intrusion Detection Message Exchange Format (IDMEF) fue especificado como un protocolo experimental que no especifica ningún estándar. El propósito de IDMEF es definir formatos de datos y procedimientos de intercambio para compartir información de interés con sistemas de detección de intrusos y sistemas de respuesta con los sistemas de administración que deben interactuar con estos. El RFC de IDMEF describe el modelo de datos para representar información tomada de los sistemas de intrusión. Se busca

que dicho formato pueda ser utilizado por los Intrusion Detection Systems (IDSs) para reportar alertas sobre eventos que parezcan sospechosos.

Por medio de la firma digital de XML se busca proveer integridad, autenticidad de los mensajes y de los servicios. La responsabilidad para la integridad y autenticación de los mensajes es responsabilidad del protocolo de comunicación y no del formato de mensajes. La inclusión de firmas digitales en mensajes IDMEF debería ser realizada en casos en los que éstas deban ser archivadas para su uso posterior o cuando los mensajes IDMEF son intercambiados sobre protocolos poco seguros.

3.1.1.1 Modelo de datos de IDMEF

El modelo de datos de IDMEF es una representación orientada a alertas enviadas a los administradores desde los sistemas de intrusión. El modelo de datos presenta varios problemas referentes a la representación de los datos:

- Generalmente, la información de las alertas es heterogénea. Algunas alertas son definidas con poca información, otras proveen información más detallada. Ejemplos de alertas con poca información son aquellas que solo presentan datos de origen, destino, nombre u hora, estos datos pueden set extendidos por alertas detalladas en las cuales se presenta información de usuarios, procesos o puertos de servicios. Es necesario que el modelo de datos que represente dicha información sea flexible para adaptarse a las distintas necesidades. Un modelo de datos orientado a objetos es extensible por medio de agregación y sub clases, de esta forma una implementación que no entienda estas extensiones podrá seguir entendiendo el subconjunto de información que está definido en el modelo. Estas dos formas de extender el modelo permiten que se mantenga su consistencia.

- Los IDSs son diferentes, algunos de estos detectan ataques analizando el tráfico en la red, otros utilizando los logs de sistemas operativos o aplicaciones que auditan información. Las alertas generadas para un mismo ataque pero por medio de diferentes herramientas, no necesariamente contienen los mismos datos. El modelo de datos define clases que soportan las diferencias en las fuentes de los datos. En particular, las nociones de fuente y objetivo para la alerta son representadas por una combinación de nodo, procesos, servicio y clases de usuario.

- Dependiendo del tipo de red o sistema operativo utilizado, los ataques serán observados y reportados de manera diferente lo cual lleva a que el modelo de datos deba adaptarse a dichas diferencias.

- Los instrumentos comerciales persiguen objetivos diferentes. Por varias razones, estos desean entregar más o menos información sobre ciertos tipos de ataques. El modelo debe permitir la flexibilidad necesaria preservando su integridad.

El diseño del modelo de datos busca proveer una representación estándar de las alertas de manera que la información no sea ambigua y que se describa la relación entre alertas simples y complejas.

El objetivo de dicho modelo es proveer una representación estándar de la información analizada por un sistema de intrusión cuando es detectado un evento inusual. Estas alertas pueden ser simples o complejas dependiendo de las capacidades de la herramienta que las creo.

Los nuevos objetos son introducidos para referenciar contenido adicional. Esto es importante debido a que la tarea de clasificar y nombrar vulnerabilidades es difícil y sumamente subjetiva. El modelo de datos no debe ser ambiguo, por lo que mientras que se permite que los análisis sean más o menos precisos entre ellos, no se debe permitir que estos produzcan información contradictoria en dos alertas que describen el mismo evento. De todas formas, siempre es posible insertar toda la información útil de un evento en campos de extensión de la alerta en lugar de en los campos a los que estos pertenecen, sin embargo, dichas prácticas reducen la interoperabilidad y deberían ser evitadas en lo posible.

3.1.2. IODEF

Incident Object Description Exchange Format (IODEF) define una representación de datos que provee un framework para el intercambio de información entre CSIRTs, dicho tipo de información de seguridad es comúnmente intercambiado entre este tipo de organizaciones. IODEF provee una representación en XML para transportar información de incidentes entre pares en distintos dominios administrativos pero que tienen las responsabilidades operacionales de remediar o analizar y establecer advertencias en un dominio definido. El modelo de datos provisto por IODEF codifica la información referente a hosts, redes y servicios corriendo en estos sistemas; metodología de ataques y evidencia forense asociada; impacto de la actividad realizada y aproximaciones a los trabajos realizados.

IODEF debería ser compatible con IDMEF y tener la capacidad de incluir mensajes IDMEF. Los actores principales en IODEF son los CSIRTs y no los IDSs como ocurre en IDMEF. Se puede ver a IODEF como una interfaz orientada a ser utilizada por personas, por ello un mensaje IODEF

es parseable por una máquina y a su vez leíble por humanos. Los objetos IODEF tienen un tiempo de vida mayor a IDMEF, en éste último los mensajes son utilizados una sola vez. Los mensajes IODEF consideran información referente al manejo de incidentes, almacenamiento de dicha información o estadísticas y análisis.

El objetivo primordial de IODEF es mejorar las capacidades operacionales de los CSIRTs. La adopción por parte de la comunidad provee una habilidad mejorada para resolver los incidentes y transmitir información de contexto simplificando la colaboración e intercambio de datos. El formato estructurado provisto por IODEF permite:

- Incrementar la automatización en el procesamiento de los datos.
- Bajar el esfuerzo necesario para normalizar datos similares de diferentes fuentes.
- Un formato común con el cual construir herramientas interoperables para el manejo de incidentes y análisis subsecuente, específicamente cuando los datos provienen de dominios distintos.

La coordinación entre CSIRTs no es un problema estrictamente técnico. La confianza, los procedimientos y las leyes son consideraciones que pueden impedir que las organizaciones intercambien información. IODEF no busca evadir dichas consideraciones, sin embargo, las implementaciones operacionales de IODEF deben considerar este contexto.

3.1.2.1 Modelo de datos de IODEF

A la hora de diseñar IODEF se realizaron ciertas consideraciones de diseño

- El modelo de datos sirve como formato de transporte, lo que lleva a que su representación específica no sea óptima para almacenamiento en disco, ser archivado por plazos largos o procesamiento en memoria.
- Por medio de la implementación no se busca establecer un consenso respecto a la definición de un incidente, en su lugar se trata de dar un entendimiento amplio que sea lo suficientemente flexible para abarcar la mayoría de las operaciones.
- Describir un incidente para todas las definiciones requeriría un modelo de datos extremadamente complejo. Por ello, IODEF solo busca

dar un marco para transmitir información de incidentes comúnmente intercambiada. Se asegura un mecanismo que pueda ser extendido para soportar información de la organización así como técnicas para referenciar información mantenida por fuera del modelo de datos.

- El dominio de análisis de seguridad no está totalmente estandarizado y debe basarse en descripciones textuales. IODEF busca conseguir un balance entre el contenido libre y el procesamiento automático de la información de incidentes.
- IODEF es una de las representaciones que han sido estandarizadas. El modelo de datos de IDMEF influenció el diseño de IODEF.

El modelo de datos de IODEF no introduce problemas de seguridad. Este solo define una representación simple para información de incidentes. Como los datos codificados por IODEF pueden ser considerados sensibles por las partes que los intercambian se deben tomar precauciones para asegurar la confidencialidad durante el intercambio y el subsecuente procesamiento. El primero debe ser resguardado por un formato de mensaje, pero luego se deben tener consideraciones de seguridad en el sistema que procesa los datos, los almacena y archiva la documentación y la información derivada de éstos.

El formato de mensajes subyacente y el protocolo utilizado para intercambiar datos provee una garantía de confidencialidad, integridad y autenticidad. El uso de protocolos de seguridad estandarizados es recomendado. Por ejemplo IODEF/RID.

3.1.2.2 Trabajo de MILE

Managed Incident Lightweight Exchange (MILE) es un grupo de trabajo del IETF que enfoca sus esfuerzos en dos áreas. La primera es el formato de datos y extensiones para representar incidentes y datos de indicadores con IODEF. La segunda área es el protocolo RID.

Respecto a IODEF se busca revisar el RFC para incorporar mejoras y extensiones basándose en la experiencia que ha sido obtenida de su utilización. Se busca poder extender IODEF para soportar extensiones especificas necesarias por la industria y permitir utilizar contenido especifico. Además MILE busca proveer guías en la implementación y uso de IODEF para ayudar a los implementadores en el desarrollo de sistemas.

Respecto a RID se busca definir una aproximación orientada a los recursos que permita a los CSIRTs ser más dinámicos y ágiles a la hora de

colaborar. También se busca proveer guías en la implementación y uso de RID. RID podría requerir modificaciones para agregar información de políticas u otros cambios. Con el incremento en la cantidad de implementaciones RID, podría ser necesaria una revisión de su RFC.

3.1.3. STIX

STIX es un lenguaje para la especificación, captura, representación y comunicación de información de seguridad. Lo realiza de manera estructurada para soportar un mejor manejo de la información así como para permitir procesos automatizados.

Se busca que la información sea representada de forma estructurada, a su vez ésta debe ser expresiva, flexible, extensible, automatizable y legible. Esto se presenta como un desafío ya que la información intercambiada debe ser estructurada para que pueda ser parseada por una máquina pero no se debe perder la posibilidad de que un analista la pueda leer y entender. Que sea entendible por una persona es necesario para que no se pierda el juicio y control que ésta pueda tener.

Al convertirse STIX en un estándar para la comunidad se obtendrá información de mejor calidad la que será mejor aprovechada y de la cual no se sabrá la fuente de los datos de antemano.

3.1.3.1 Aproximaciones de la actualidad

La información que se intercambia y utiliza actualmente es atómica, inconsistente y muy limitada en sofisticación y expresividad. Además, el uso de estructuras estandarizadas se enfoca en una porción del problema y la integración no se realiza de forma adecuada entre si o carece de la flexibilidad para hacerlo. Generalmente, las actividades de intercambio de indicadores son entre humanos, siendo los indicadores desestructurados o semi-estructurados e intercambiados por medio de portales web o encriptados y enviados vía email. Recientemente se ha visto el surgimiento de transferencias entre máquinas, éstas intercambian conjuntos de indicadores simples de modelos de ataques bien conocidos.

A diferencia de IODEF, STIX provee una lista de elementos para construir indicadores de compromiso y se integra con CAPEC, MAEC o CVE. Además posee soporte para tácticas, técnicas y procedimientos del adversario y tareas realizadas por la organización. IODEF fue pensado para compartir información de incidentes y no indicadores de compromiso. STIX permite el intercambio de indicadores referentes a amenazas así como información

del contexto en el que éstas se dan. Juntos, permiten tener un conocimiento más rico de las intenciones, capacidades, motivaciones y actividades de un adversario y de esta forma defenderse de él.

STIX busca extender los indicadores para permitir el manejo e intercambio de estos de forma más expresiva y con un espectro más amplio de información.

Actualmente, el intercambio y manejo de información automática es visto típicamente en líneas de productos, servicios ofrecidos o soluciones específicas de una comunidad. STIX busca permitir el intercambio de información de forma comprensiva, rica, de alta fidelidad entre organizaciones, comunidades, productos y servicios ofrecidos.

STIX provee una arquitectura unificada con un conjunto amplio de información para permitir una solución práctica que permita la implementación de distintos casos de uso. El conjunto de información incluye:

- Cyber Observables
- Indicadores
- Incidentes
- Tácticas, técnicas y procedimientos de los adversarios
- Objetivos de exploits
- Cursos de acción
- Campañas de ataques
- Actores de ataques

3.1.3.2 Casos de uso

Ha sido desarrollado para soportar varios casos de uso involucrados en el manejo de amenazas de seguridad. A continuación se dan descripciones de esos casos de uso.

- Análisis de Cyber Amenazas: Un analista de seguridad estudia información estructurada y no estructurada referente a actividades de una amenaza procedentes de varias fuentes. El analista busca entender la naturaleza de las amenazas más relevantes, identificarlas, y representarlas para poder expresar y actualizar el conocimiento relevante de la amenaza. La información de la amenaza incluye acciones realizadas, comportamientos, capacidades, quienes lo realizaron, etc. De este

conocimiento y representación el analista puede especificar patrones de la amenaza, sugerir acciones a ser realizadas para responder a sus actividades y/o compartir información con socios de su confianza.

- Especificar patrones de amenazas: Un analista especifica patrones para representar las características de amenazas junto con su contexto y metadatos para interpretar, manejar y aplicar el patrón y sus resultados. Esto puede ser realizado manualmente o con la asistencia de una herramienta automática.

- Manejo de las actividades de respuesta: Los encargados de tomar decisiones y el personal de operaciones trabajan en conjunto para prevenir o detectar actividades que presenten una amenaza y responder a los incidentes detectados que sean realizados por dichas amenazas. Las acciones preventivas pueden mitigar vulnerabilidades, debilidades o malas configuraciones que sean objetivos de exploits. Luego de la detección e investigación de incidentes específicos, acciones reactivas pueden ser realizadas. Se pueden desprender tres sub-casos de uso de lo anterior.

    - Prevención de amenazas: Quienes deben tomar las decisiones evalúan acciones preventivas para las amenazas relevantes que sean identificadas y seleccionan acciones apropiadas para su implementación. El personal de operaciones implementa las acciones seleccionadas por los tomadores de decisión. Dichas medidas preventivas son obtenidas por medio de la interpretación de los indicadores.
    - Detección de amenazas: El personal de operaciones aplica mecanismos (automáticos y manuales) para monitorear y asistir las operaciones con la finalidad de detectar la ocurrencia de amenazas específicas basándose en la evidencia histórica, análisis del contexto actual e interpretación de los indicadores que se presentan. Esta detección se realiza generalmente por medio de patrones de indicadores.
    - Respuesta a incidentes: El personal de operaciones responde a amenazas detectadas, investiga que ha ocurrido o está ocurriendo, trata de identificar y representar la naturaleza de las amenazas y lleva a cabo acciones para mitigar sus efectos, también puede aplicar acciones preventivos. Una vez que los efectos son entendidos, el personal de operaciones puede implementar acciones acordes para prevenir o llevar a cabo tareas correctivos.

- Intercambio de información: Los encargados de tomar las decisiones establecen políticas respecto a que tipo de información será compartida, además toma decisiones respecto a con quienes se comparte y como debería ser manejada basándose en frameworks de confianza de forma de mantener niveles adecuados de consistencia, contexto y control. Esta política es luego implementada para compartir los indicadores de amenazas apropiados y otra información de amenazas.

3.1.3.3 Principios de desarrollo de STIX

En el enfoque dado a STIX se ha buscado implementar un conjunto de principios para su desarrollo con el consenso de la comunidad. Esos principios son los siguientes:

- Expresividad: Con el fin de soportar la diversidad de casos de uso relevantes, STIX apunta a proveer una cobertura expresiva en todos sus casos de uso específicos en lugar de dirigirse específicamente a alguno de ellos.

- Integración en lugar de duplicación: Cuando STIX abarca conceptos de información estructurada para los cuales ya existen representaciones estandarizadas con el consenso adecuado y que se encuentran disponibles, se busca integrar estas representaciones a la arquitectura STIX en lugar de duplicar esta información innecesariamente.

- Flexibilidad: Con el fin de soportar un amplio rango de casos e información variable con varios niveles de fidelidad, se ha diseñado a STIX para ofrecer tanta flexibilidad como sea posible. STIX se adhiere a una política de permitir a los usuarios utilizar cualquier porción de representaciones estándar que sean relevantes para un contexto dado y evita elementos obligatorios siempre que sea posible.

- Extensibilidad: Con la finalidad de soportar un amplio rango de casos de uso con un potencial diferente de representación y para facilitar el perfeccionamiento impulsado por la comunidad, se ha diseñado STIX para construir mecanismos de extensión para usos específicos, para usos localizados, para refinamientos del usuario y evolución y para facilidad de refinamiento y evolución centralizada.

- Automatización: El diseño realizado en STIX busca maximizar la estructura y la consistencia para soportar métodos de procesamiento automático por máquinas.

- Lectura: El diseño de STIX busca estructuras del contenido para que no sea únicamente consumible o procesable por máquinas sino que también sea leíble por humanos. Esto es necesario para claridad y comprensibilidad durante las primeras etapas de desarrollo y adopción.
- Implementación: La implementación inicial de STIX utiliza XML como un mecanismo portátil, estructurado y que se puede encontrar en cualquier parte para la discusión, colaboración y refinamiento entre las comunidades involucradas. Está pensado para el desarrollo en colaboración de un lenguaje estructurado sobre información de amenazas entre expertos de la comunidad. Se ha pensado que el uso del lenguaje sea estimulado y soportado por medio del desarrollo de varias herramientas como APIs. Solo por medio de niveles adecuados de colaboración entre miembros de la comunidad y con la utilización de datos reales se puede llegar a que la solución evolucione.

Cuadro 1: Comparativa de STIX e IODEF

STIX	**IODEF**
Se representan observables que son propiedades o eventos que se dan durante las operaciones de redes o computadores. Se refiere a información sobre un archivo, una key en la registry, un servicio iniciado, etc. Para su representación se utiliza CyBox. Se da información de eventos en el host y la red por medio de Observables. Se puede considerar información como nombres, hashes, tamaños de archivos. Keys que se cambian en la registry, servicios que se ejecutan o request realizadas a un host (ie: request http)	Descripción de eventos particulares del incidente que se dan en un host o red. También se puede tener información de Log de los eventos o acciones significativas realizadas por los involucrados.
Se representa información de indicadores que afectan a la organización así como nuevos datos conocidos durante la respuesta al incidente. Se dan datos sobre lo que busco realizar un atacante, fuente de la información del incidente y tareas realizadas por los analistas.	Se da una descripción textual del incidentes. Se da información de contacto de las partes que participaron del incidente. Se da información de la razón por la cual se envía el documento y de cuales son las posibilidades que tiene el receptor para divulgar información.

Continua en la página siguiente

Cuadro 1 – *Continuación de la página anterior*

STIX	**IODEF**
	Se permite la extensión del modelo para representar datos que no están en éste.
Se da información sobre socios involucrados.	Información de contacto para personas u organizaciones involucradas en el incidente.
La clase Incidents da representaciones de los tiempos de detección reporte, etc del incidentes.	La clases Time dan información sobre el comienzo, fin, detección del incidente.
Se da una representación del modus operandi del adversario. Esto se realiza por medio de TTP, en estas se busca representar comportamientos específicos que muestra el adversario, recursos que éste tiene como herramientas e infraestructura, información de las victimas, objetivos de exploits, fuentes de la información de los TTP. Se utilizan otros estándares como CVE, CAPEC o MAEC para la representación de los ataques o del malware.	Una representación libre de la metodología utilizada por el adversario. Se representan las técnicas utilizadas por el atacante.
Por medio de la clase incidents se da información de bienes afectados y de la naturaleza del incidente.	Repercusiones técnicas y no técnicas del incidente, ie: impacto monetario, de tiempo, etc.
Las campaigns dan referencia a actividades similares que fueron realizadas.	Referencia a actividades similares.
Representación de campañas, esto son adversarios con una intención. Las campañas consisten de las intenciones del adversario, sus TTP utilizadas en la campaña, los incidentes realizados en la campaña, indicadores asociados a la campaña.	
Representaciones de adversarios que tienen ciertas intenciones y han sido observados históricamente. Los adversarios tienen las siguientes características: una representación de identidad, se sospecha de una motivación, una intención de conseguir algo, tienen un historial de TTP, ciertas campaigns son asociadas con un adversario, etc.	

Continua en la página siguiente

Cuadro 1 – *Continuación de la página anterior*

STIX	IODEF
Se representan objetivos de exploits, estos se pueden ver como vulnerabilidades en software, sistemas, configuraciones de red que son objetivos para ser explotados por las TTP de un adversario. Se utilizan CVE, OSVBD (entre otras) para la identificación de vulnerabilidades publicas.	Se da una clasificación de vulnerabilidades conocidas.
Se representan medidas que deberían ser realizadas para prevenir o corregir para evitar ser objetivo de exploits.	Se dan acciones que debería realizar el receptor.
Se representan etiquetas para la información que indican restricciones o datos potencialmente sensibles.	

3.2. Estándares

4. Intercambio

4.1. Protocolos

4.1.1. RID

Real-Time Inter-Network (RID) es un método de comunicación entre redes para facilitar el intercambio de datos de incidentes. A su vez, busca integrar mecanismos existentes de detección, seguimiento, identificación de fuentes y mitigación que aporta una solución al manejo de incidentes. Combinar estas capacidades en un sistema de comunicación permite incrementar el nivel de seguridad en la red. Las políticas para manejar incidentes son recomendadas y pueden ser acordadas por un consorcio utilizando recomendaciones y consideraciones de seguridad.

RID ha sido ampliamente utilizando en comunidades de investigación, pero no ha sido muy adoptado por otros sectores. Fue desarrollado como un mecanismo de comunicación para facilitar la transferencia de información entre distintos proveedores de servicios de Internet para trazar precisa y eficientemente el flujo de paquetes nocivos a lo largo de la red. RID considera la información que necesitan varias implementaciones para seguir paquetes dentro de una red y los requerimientos de los proveedores de servicios de decidir si se permite trazar el recorrido de un paquete o no.

Los datos en RID son representados como documentos XML utilizando IODEF. De esta forma se simplifica la integración con otros aspectos del manejo de incidentes. Además las ventajas de la representación en XML permite conservar la privacidad y tener un buen nivel de aseguramiento. RID busca proveer un método para comunicar la información relevante entre CSIRTs manteniendo la compatibilidad con varios sistemas de rastreo y respuesta.

Los mensajes RID son encapsulados en otros protocolos para el transporte, este procedimiento se define en el RFC 6046. La autenticación, integridad y autorización son el resultado de las capacidades de cada capa y son utilizadas para alcanzar un nivel de aseguramiento adecuado.

Los mensajes RID tienen la intención de ser usados en el manejo coordinado de incidentes para localizar la fuente de un ataque y detener o mitigar sus efectos. Los objetivos de los ataques incluyen redes o sistemas que se vieron comprometidos con ataques como denegaciones de servicio u otro tipo de tráfico malicioso en una red. Por medio de RID los CSIRTs tienen la posibilidad de reportar ataques que estén ocurriendo a otros CSIRTs o pedir información a éstos de la detección de ataques.

Los sistemas comprometidos pueden ser el resultado de otros incidentes de seguridad como worms, troyanos o virus. El manejar incidentes es una tarea difícil para algunas organizaciones debido al tamaño de su red y la cantidad de recursos disponibles. RID provee el framework necesario para realizar la comunicación entre redes involucradas en el manejo, seguimiento y mitigación de incidentes de seguridad. Distintos tipos de mensajes son necesarios para facilitar el manejo de incidentes. Los mensajes que se incluyen son *Report*, *IncidentQuery*, *TraceRequest*, *RequestAuthorization*, *Result* e *InvestigationRequest*. El mensaje Report es utilizado cuando se ingresa un incidente en un sistema RID y no se deben realizar más acciones. Un mensaje Incident Request es utilizado para pedir información de un incidente en particular. Un mensaje Trace Request es utilizando cuando la fuente del tráfico puede estar oculta. En ese caso, cada proveedor de red que reciba uno de estos mensajes enviará uno a la red anterior en el camino del mensaje para obtener la fuente del tráfico. Los mensajes Request Authorization y Result son utilizados para comunicar el estado y resultado de un mensaje Trace Request o Investigation Request. El mensaje Investigation Request solo considera lo sistemas RID en el camino a la fuente del tráfico.

En RID se pueden diseñar topologías que permitan el intercambio de información facilitando la comunicación entre socios. La topología básica para comunicar sistemas RID es la de una comunicación directa. En una organización se pueden establecer y utilizar varias topologías. Una podría

fortalecer las relaciones bilaterales entre socios. Los socios podrían reenviar los mensajes RID entre ellos. Este enfoque permite un rastreo iterativo en donde la fuente es desconocida.

La comunicación entre los sistemas RID debe ser protegida. RID tiene muchas consideraciones de seguridad incluidas en el diseño del protocolo. Al considerar el transporte de mensajes RID, una red fuera de banda, ya sea física o lógica, puede prevenir ataques externos contra las comunicaciones RID. Se deben utilizar conexiones autenticadas y encriptadas entre sistemas RID para proveer confidencialidad, integridad, autenticidad y privacidad de los datos. Las relaciones de confianza son realizadas por socios y establecen relaciones de confianza por medio de infraestructura de PKI.

El protocolo de transporte utilizado debe proveer encriptación para dar un nivel de seguridad e integridad adicional, mientras se provee autenticación por medio de certificados. De todas formas los mensajes RID no utilizan únicamente la seguridad provista por el protocolo de transporte.

La infraestructura PKI provee la base para la autenticación, autorización, encriptación y firmas digitales necesarias para establecer una relación de confianza entre los miembros de una comunidad RID.

Problemas de privacidad pueden ser de preocupación cuando se habla de compartir información y deben ser considerados a la hora de alcanzar la meta de detener o mitigar los efectos de un incidente de seguridad. Para ello hay clases específicas para automatizar las políticas de privacidad.

IODEF define un formato de mensaje, no un protocolo de transporte, esto se realiza para permitir a los CSIRTs intercambiar y almacenar los datos de la forma que más le conviene a cada organización. Sin embargo, RID requiere una especificación de un protocolo de transporte para asegurar la interoperabilidad entre las organizaciones socias. El RFC6046 mencionado anteriormente, especifica el transporte de mensajes RID sobre HTTPS/TLS. En esta especificación, cada servidor RID funciona como servidor y cliente. Todos los sistemas RID deben estar preparados para aceptar conexiones HTTP/TLS de cualquier socio con la finalidad de soportar llamadas a respuestas tardías de pedidos que se realizaron.

4.1.2. TAXII

El objetivo que plantea Trusted Automated eXchange of Indicator Information (TAXII) es extender la capacidad de compartir indicadores, siendo dichos intercambios robustos, seguros y de gran volumen de datos. A su vez, los datos intercambiados deberían ser mas expresivos que en la actualidad.

Con TAXII no se busca crear una comunidad para compartir, sino que se

le da a las organizaciones una herramienta que facilite el intercambio entre ellas. TAXII mejora las deficiencias existentes dando especificaciones abiertas y comunes para transportar los mensajes con información. También se provee un conjunto de capacidades como encriptación, autenticación, direccionamiento, alertas y pedidos entre sistemas.

TAXII es un conjunto de especificaciones técnicas y de documentación para permitir el intercambio de información procesable entre organizaciones. Para realizar dichos intercambios, se definen protocolos y formatos de datos que permiten intercambiar información de forma segura. Ha sido diseñado para permitir la interoperabilidad de diferentes soluciones en lugar de ligarse a una tecnología o producto en particular. Además, se busca incentivar a los proveedores de tecnología a incorporar soporte para las especificaciones de TAXII en sus productos. La información intercambiada ayuda a detectar, prevenir y mitigar amenazas informáticas en tiempo real. Es importante recalcar que no se buscan definir acuerdos para el intercambio de información o gobierno. En su lugar, permite a las organizaciones mejorar el contexto en el que se encuentran respecto a las nuevas amenazas y además compartir la información que ellos elijan con las organizaciones que deseen de forma simple y rápida aprovechando las relaciones y sistemas existentes.

En el desarrollo de TAXII se buscó consenso y participación de la comunidad. TAXII permite el intercambio de información sobre amenazas de forma eficiente y comprensiva por medio de *automatización* y *articulación* de un modelo detallado de información. Para lograr esto, se utiliza una representación estándar de información de amenazas y un framework para soportar el intercambio de datos. El modelo permite el envío y recepción de un conjunto amplio de información de seguridad para soportar las necesidades referentes al intercambio de información. Se le da la libertad a los proveedores de determinar como sus productos producen, consumen o toman ventaja de los flujos de información especificados por TAXII.

[width=150mm]./images/TAXIIArchitecture1.png

Figura 2: Arquitectura de TAXII [2]

TAXII cubre un amplio número de casos de uso, tecnologías, especificaciones e implementaciones. Los casos de uso son desarrollados de manera secuencial, permitiendo un conjunto inicial de casos de usos que permiten el intercambio de información. TAXII utiliza protocolos y especificaciones existentes siempre que es posible, de esta forma se integra con mecanismos de intercambio de información para reducir los costos de implementación y

permitir la adopción rápida por parte de organizaciones ya establecidas y que se encuentran intercambiando información.

Las motivaciones para tener una mejor solución que permita intercambiar información lleva a que en el diseño de TAXII se hayan planteado los siguientes objetivos:

- Permitir el intercambio seguro y rápido de información referente a amenazas entre comunidades de defensores de seguridad.
- Lograr un estándar para permitir compartir indicadores entre otros elementos entre organizaciones.
- Extender el intercambio de indicadores para permitir intercambios seguros, robustos y de gran volumen que tengan una expresividad mayor a la actual.
- Soportar un amplio número de casos de uso y practicas comunes a las comunidades.
- Tomar los estándares existentes que sean adecuados.
- Llegar a una adopción por parte de organizaciones internacionales de estándares.

Para representar la información, TAXII utiliza STIX. STIX es un lenguaje desarrollado por la comunidad para la especificación, captura, representación y comunicación de información de amenazas cibernéticas de forma estandarizada.

Para automatizar el intercambio de información, es necesario especificar como ésta es compartida. Para lograr esto, TAXII define especificaciones técnicas y documentación de soporte. En particular, las especificaciones de TAXII definen un conjunto de capacidades necesarias para el transporte exitoso de mensajes. Los mensajes TAXII llevan datos de amenazas informáticas transformadas a formato STIX. El conjunto completo de los mensajes incluyen mensajes con datos y de control.

TAXII utiliza protocolos y especificaciones existentes siempre que sea posible y los integra con los mecanismos actuales para reducir los costos de implementación y permitir una adopción rápida por parte de las organizaciones ya establecidas y que ya comparten información. TAXII ha sido desarrollado de forma modular para soportar una variedad de mecanismos y formatos de datos para ser intercambiados.

4.1.2.1 Casos de uso

TAXII ha sido desarrollado para soportar casos de uso comunes para el intercambio de información. A continuación se detallan los casos de uso desarrollados.

4.1.2.1.1. Alertas o Advertencias públicas Estas son advertencias al público en general o a varios miembros de CSIRTs, son enviadas a todos los subscriptores. Estas alertas son de una naturaleza muy amplia, por ello no necesitan ser encriptadas o tener autorizaciones especiales. Sin embargo es importante asegurar la autenticidad del emisor. Las entidades u organizaciones deben ser especificadas para identificar la fuente de la alerta.

4.1.2.1.2. Alertas y reportes privados Las alertas privadas son similares a las públicas, la diferencia radica en que la información compartida es sensible y restringida a los socios que comparten datos. Los mecanismos para enviar datos deberían ser similares a los de las alertas públicas. Como las alertas y reportes se suponen sensibles y no para el uso general, es importante que TAXII soporte formas adecuadas de encriptación, autenticación, autorización e identificación de datos. Dependiendo en la naturaleza de las comunicaciones, podría requerirse un manejo explícito de marcas o restricciones.

Las alertas son generalmente cortas, teniendo mensajes estándar con indicadores muy específicos o acciones especificadas. Los reportes tienden a ser mas largos y pueden incluir reportes de incidentes, análisis de malware o amenazas así como otras observaciones.

4.1.2.1.3. Soporte para Consultas Es común que los analistas busquen información entre sus comunidades así como por fuera de ellas. Para esto se soportan dos tipos de mensajes:

- Request For Information (RFI): es un mensaje simple que se espera sea manejado de forma manual y que permite que se pida información a otra organización.
- Repository Search: Para este tipo de consulta, es esperado que una organización ofrezca repositorios en los cuales buscar, los cuales podrían ser compatibles con TAXII o STIX.

4.1.2.1.4. Transferencia Varias organizaciones que transfieren información necesitan en algunas instancias agregar miembros. El nuevo miembro necesita obtener los datos del repositorio de alguna organización. Por ello se desea un caso de uso para realizar un intercambio de un gran volumen de datos.

4.1.2.2 Componentes de TAXII

Como se dijo anteriormente, TAXII es un conjunto de especificaciones técnicas y documentación para el intercambio de información. A continuación se enumeran los distintos componentes de TAXII.

- Especificación de TAXII: Define la especificación de los componentes y provee una guía y requerimientos sobre como dichas especificaciones interactúan en TAXII.
- Especificación de servicios TAXII: Define una serie de servicios que deben ser implementados para ser compatible con TAXII. Describe la información intercambiada en alto nivel y no se limita ningún mecanismo de intercambio en especial.
- Implementación de servicios: Se realiza una implementación de los servicios TAXII para un mecanismo de intercambio. Cada implementación de servicios provee una guía técnica y requerimientos para implementar la especificación de los mecanismos de intercambio.
- Modelo de datos de mensajes: Se define una estructura para los mensajes TAXII, incluyendo cabezales, datos del mensaje, control y mensajes de datos. Los mensajes de datos utilizan STIX como carga de los mensajes TAXII.
- Implementación de los mensajes de datos: Es una implementación del modelo de datos de mensajes, incluyendo la carga STIX. Cada implementación de mensajes define una guía técnica y requerimientos para utilizar un formato de mensajes particular para expresar el modelo de datos de mensaje.

4.1.2.2.1. TAXII Toolkit Es provisto para soportar la adopción de TAXII y asistir en el desarrollo de capacidades compatibles. El toolkit provee una colección de implementaciones de referencia, un conjunto de herramientas y una colección de librerías e interfaces.

TAXII está definido por múltiples especificaciones relacionadas. Esta sección describe las especificaciones definidas en TAXII.

- Especificación de servicios: Provee los requerimientos por los cuales se definen los servicios e intercambios de TAXII. No provee detalles respecto al formato de los datos o como los mensajes TAXII son transportados por la red. Dichos detalles y requerimientos pueden ser encontrados en la especificación de los protocolos de enlace y en la especificación de mensajes de enlace.
- Especificación de protocolos de enlace: Define los requerimientos para transportar mensajes TAXII por la red. Puede haber varias especificaciones creadas para TAXII. Cada especificación define requerimientos para el transporte de mensajes TAXII utilizando protocolos de red y se proveen requerimientos respecto a como los servicios TAXII son soportados por los protocolos de red.
- Especificación de mensajes de enlace: Se definen requerimientos para representar mensajes TAXII en un formato particular. Puede haber múltiples especificaciones para dichos mensajes. Se provee información detallada sobre como las especificaciones definidas en la especificación de servicios son expresadas en los mensajes.

Figura 3: Relación de las especificaciones de TAXII [2]

Para dar flexibilidad en el proceso evolutivo de TAXII, se han separado las especificaciones de servicios, de los protocolos de enlace y de los mensajes de enlace. Lo dicho anteriormente es expresado en la figura 3. Debido a que las organizaciones generalmente tienen restricciones respecto a los protocolos que soportan, TAXII busca no ligarse a un único protocolo que excluya a una parte de la comunidad. Cuando se ve que la comunidad expresa interés en un nuevo protocolo o tipo de mensaje, TAXII puede dar soporte para ellos sin cambiar los componentes centrales.

Dos grupos que usen el mismo protocolo de red y formato de mensajes serán capaces de intercambios de información estructurada de forma automática. Las políticas de intercambio de los participantes pueden limitar estos intercambios si es necesario, pero el uso de servicios compatibles con TAXII asegura que se puede intercambiar cualquier información con los mecanismos definidos por TAXII. Los grupos que usen diferentes protocolos o

formatos de mensajes no serán capaces de comunicarse directamente, pero como están utilizando mensajes y servicios en el núcleo de las comunicaciones de sus comunidades significa que es posible establecer caminos para que ocurra la interacción.

4.1.2.2.2. Especificación de Servicios Esta especificación provee normativas respecto a los servicios, mensajes e intercambios de mensajes en TAXII. No provee detalles respecto a como los mensajes son transportados, dejando eso a la especificación de los protocolos de enlace. Se da información respecto a los datos presentes en los mensajes TAXII y no a como los mensajes son expresados.

Las unidades funcionales de TAXII representan conjuntos discretos de actividades requeridas para soportar TAXII. Una unidad funcional representa algún componente con un rol bien definido en TAXII.

- TAXII Transfer Agent (TTA): Es una unidad funcional conectada a la red que envía o recibe mensajes TAXII. Una TTA interactúa con otras TTAs por medio de la red y maneja los requerimientos de una o más de las especificaciones de los protocolos de enlace. Una TTA provee un mensaje TAXII a un TAXII Message Handler permitiendo que éste último sea independiente del protocolo de red utilizado. De la misma forma, el TTA puede ser independiente del contenido de los mensajes TAXII, dejando el manejo de la información al TAXII Message Handler.
- TAXII Messsage Handler (TMH): Es una unidad funcional que produce y consume mensajes TAXII. EL TMH es responsable de parsear y construir mensajes con el formato especificado en uno o más TAXII Message Binding Specifications. Un TMH interactúa con un TTA, el cual maneja los detalles necesarios para transmitir mensajes por la red. El Backend TAXII interactúa con el TMH para convertir su contenido en mensajes TAXII y llevar a cabo actividades basadas en los mensajes TAXII que son recibidos por el TMH.
- TAXII Backend: Cubre todas las unidades funcionales distintas al TTA y al TMH. Las especificaciones de TAXII no proveen requerimientos sobre como son implementadas las capacidades en un backend más allá de como debe interactuar con el TMH. Las organizaciones o implementadores pueden decidir que capacidades implementar según los servicios TAXII que deseen soportar o según como quieran dar ese soporte.

- Arquitectura TAXII: Cubre los aspectos de las unidades funcionales de la infraestructura de productor o consumidor que provee o utiliza servicios TAXII. Una arquitectura TAXII incluye una TTA, un TMH y un backend TAXII.

[width=150mm]./images/TAXIIArchitecture.png

Figura 4: Unidades funcionales de TAXII [2]

4.1.2.3 Capacidades

TAXII provee capacidades especificas para aquellos que desean compartir información de amenazas cibernéticas. Las capacidades TAXII son el nivel más alto en el cual se pueden expresar las acciones de TAXII. Hay tres capacidades que soporta la actual versión de TAXII, estas son: push messaging, pull messaging y discovery.

En push messaging la información puede ser enviada de un productor a un consumidor. Esto puede reflejar una relación pre-existente entre el productor y el consumidor en la que el consumidor ha pedido que se le envíen datos desde el productor. También puede usarse en caso de que el consumidor desee aceptar contribuciones de cualquier productor, y estos le envíen datos en cualquier momento.

Pull messaging permite a un consumidor requerir información de un productor. Esto no solo le permite al consumidor el control sobre el momento en el que recibe los datos sino que también le permite hacerlo sin tener que aceptar conexiones entrantes. Así como en push messaging, el productor y consumidor pueden tener acuerdos pre-existentes para que el consumidor tenga acceso a los datos del productor. De forma alternativa, un productor puede hacer su información pública de forma que cualquier consumidor pueda obtenerla. La versión actual de pull messaging limita a los consumidores a hacer pedidos por medio de las organizaciones productoras de los datos en lugar de por los datos en si. Toda la información provista por un productor debe estar organizada en grupos llamados "TAXII Data Feeds". Cada elemento en un TAXII data feed es etiquetado utilizando timestamps. El productor tiene total dominio sobre como el contenido se mapea en TAXII data feeds y en el significado de los timestamps. La capacidad de pull messaging está atada a entender el contenido del productor.

Para facilitar las comunicaciones automatizadas, TAXII soporta capacidades para descubrir los servicios específicos que ofrece un servidor o grupo

de servidores, así como los protocolos o mensajes que este servidor ofrece. Esto no quita la necesidad de que personas estén involucradas para establecer acuerdos de cooperación lo cual esta por fuera del objetivo de TAXII. Sin embargo, permite el intercambio de información respecto a las capacidades que un productor soporta y cuales son los mecanismos que utiliza para hacerlo.

4.1.2.4 Servicios TAXII

Los servicios TAXII representan un conjunto de mecanismos necesarios para soportar capacidades TAXII. Una implementación TAXII pudiera implementar alguno, todos o incluso ninguno de los servicios definidos. TAXII define los siguientes servicios:

- Servicio de descubrimiento: Es utilizado para recibir y responder a mensajes que requieren información sobre los servicios ofrecidos.
- Feed Managment Service: Es utilizado para recibir o responder a mensajes utilizados para el manejo de subscripciones a TAXII Data Feed.
- Inbox Service: Es utilizado para recibir información de amenazas cibernéticas por medio de intercambios iniciados por el productor en intervalos dictados por este.
- Poll Service: Es utilizado para recibir y responder a mensajes de pedido a el TAXII Data Feed iniciados por el consumidor.

A continuación se describen los distintos servicios.

4.1.2.4.1. Discovery Service Es un mecanismo para comunicar información referente al uso de servicios TAXII y a su disponibilidad. Para un pedido al servicio, se retorna una lista de los servicios TAXII y como estos pueden ser invocados. Un solo servicio de descubrimiento puede reportar servicios TAXII en diferentes equipos finales o incluso en múltiples organizaciones, los propietarios del servicio pueden definir su alcance a gusto. Un servicio de descubrimiento puede utilizar varios factores para determinar cuales servicios revelar ante una petición, incluyendo, pero no limitado a la entidad del cliente TAXII. El servicio de descubrimiento debe soportar "Discovery Message Exchange".

4.1.2.4.2. Feed Managment Service Es el mecanismo con el cual un consumidor pide información referente a TAXII Data Feeds, pidiendo subscripciones a estos, o modificando las existentes. Este servicio facilita el intercambio de mensajes para manejar las subscripciones. No se entrega contenido de los TAXII Data Feed, en su lugar se envía contenido del TAXII Data Feed al servicio de Inbox de un consumidor en intercambios iniciados por un productor o en respuesta directa a un pedido del consumidor al servicio de poll. Dicho servicio debe implementar soporte para subscription managment exchange. Dicho servicio podría implementar soporte de feed information exchange.

4.1.2.4.3. Inbox service Este servicio es el mecanismo con el cual un consumidor acepta los mensajes en un intercambio iniciado por el productor. Un consumidor puede implementarlo para recibir datos del TAXII Data Feed. El servicio de inbox debe implementar soporte para Data Push Exchange.

4.1.2.4.4. Poll service Es provisto por un productor para permitir pedidos al TAXII Data Feed iniciados por el consumidor. Un consumidor contacta a este servicio explícitamente pidiendo el contenido del TAXII Data Feed. Los productores podrían ofrecer Data Feeds combinando envíos al Inbox service del consumidor o por medio de pedidos al servicio de poll de productor. Una implementación de éste servicio debe dar soporte a Data Poll Exchange.

4.1.2.5 Intercambio de mensajes TAXII

Esta sección describe los mensajes intercambiados que son necesarios para soportar los servicios definidos antes. Estos intercambios solo consideran mensajes TAXII y son independientes a los protocolos sobre los cuales viajan los mensajes. En particular, esos protocolos podrían requerir intercambios de red adicionales antes de transmitir mensajes TAXII o romper un mensaje TAXII en múltiples mensajes del protocolo subyacente que son transmitidos independientemente. Los siguientes diagramas representan conceptualmente la secuencia en la cual los mensajes TAXII son transmitidos y como actúan.

4.1.2.5.1. Data Push Exchange En este intercambio, un mensaje STIX es transmitido desde un cliente a un servidor inbox que está esperando. El mensaje STIX puede ser solicitado o no solicitado. El servidor inbox puede ser capaz de filtrar mensajes según la autenticidad del emisor. Los mensajes

enviados en este intercambio no deberían tener un campo 'In Response to' en su header.

[width=150mm]./images/DataPushExchange.png

Figura 5: Mensajes intercambiados en un Data Push Exchange [2]

El cliente TAXII envía un mensaje STIX al inbox server. El inbox server podría descartar el mensaje o pasar el mensaje STIX junto con cualquier información de la entidad autenticada al Back-end TAXII. El cliente TAXII no recibe respuesta del servidor de inbox y no sabrá si el mensaje ha sido aceptado o descartado por el servidor, aunque el protocolo confiable de la capa inferior puede asegurar que el mensaje fue entregado a la TTA del servidor de inbox. El inbox server no enviará un mensaje de error TAXII si hay algún problema con el mensaje TAXII.

4.1.2.5.2. Discovery Exchange Un cliente TAXII pide información sobre el servicio TAXII ofrecido por un productor. El discovery server del productor responde con una lista de servicios. Si bien el cliente puede ser informado de la existencia de un servicio, este no necesariamente tendrá acceso inmediato al servicio.

[width=150mm]./images/DiscoveryExchange.png

Figura 6: Mensajes intercambiados en un Discovery Exchange [2]

El cliente TAXII envía un pedido de descubrimiento al servidor. Cuando el servidor recibe el pedido puede retornar un mensaje de error o pasar la información al Backend TAXII. Información relevante incluye la identidad autenticada si ésta fue provista. El Backend TAXII podría utilizar esta información junto a su propia política de control de acceso para crear una lista de servicios a ser retornada. Esto podría ser empaquetado en una respuesta de discovery lo cuales podrían ser enviados al cliente TAXII. El cliente TAXII recibe esa respuesta y la pasa la información del servicio a su propio Back-End para ser procesado.

4.2. Feed Information Exchange

En éste intercambio un cliente TAXII pide información sobre fuentes de datos disponibles en un Feed Server. El servidor responde con una lista de fuentes de datos disponibles. Dicha respuesta es realizada por el backend y

en ella se pueden considerar decisiones de control de acceso para realizar la respuesta.

[width=150mm]./images/FeedInformationExchange.png

Figura 7: Mensajes intercambiados en un Feed Information Exchange [2]

El cliente TAXII envía el Feed Information Request al servidor. Cuando el servidor recibe la request podría retornar un mensaje de error o pasar la información relevante al TAXII backend. Entre la información relevante se podría incluir la identidad. El backend podría utilizar esta información junto con sus políticas de control de acceso para crear una lista de fuentes de datos para ser enviadas al cliente. Esta lista es empaquetada en una Feed Information Response. El cliente recibe este mensaje y pasa el Feed a su propio backend para ser procesado.

4.3. Subscription Managment Exchange

En este un cliente intenta establecer, borrar, pausar, resumir o modificar una subscripción a un TAXII Data Feed conocido enviando un mensaje subscription managment request al servidor. El servidor pasa la request al TAXII Back-end el cual determina la respuesta. Esta respuesta es luego enviada al cliente.

[width=150mm]./images/SubscriptionManagmentExchange.png

Figura 8: Mensajes intercambiados en un Subscription Managment Exchange [2]

El cliente TAXII envía una Manage Feed Subscription Request al servidor. Éste podría retornar un mensaje de error TAXII o pasar la información relevante al TAXII backend. La información relevante podría incluir la identidad, parámetros que identifiquen la subscripción a ser modificada o creada, y la acción realizada. El backend TAXII puede usar dicha información junto con sus políticas de control de acceso y las funcionalidades que posea para determinar si la acción está permitida o no. Dependiendo de la respuesta, el servidor podría retornar un mensaje de error TAXII o enviar una respuesta Manage Feed Successful Response.

4.4. Feed Poll Exchange

Es utilizado por un consumidor para pedir contenido de un productor de datos. El TAXII Data Feed content es enviado al consumidor en el mismo intercambio. Esto permite a un consumidor devolver el Data Feed Content en su propia tabla de tiempo y sin necesidad de utilizar un Inbox Server o aceptar conexiones entrantes.

[width=150mm]./images/FeedPollExchange.png

Figura 9: Mensajes intercambiados en un Feed Poll Exchange [2]

El cliente consumidor inicia el intercambio enviando un mensaje Poll Request al servidor de Poll del productor. El servidor puede enviar un mensaje de error inmediatamente o pasar la información relevante al TAXII backend. Información relevante incluye el nombre de la fuente, los parámetros de subscripción, timestamps indicando el intervalo de tiempo de la información que pide el consumidor y la identidad del consumidor. El backend TAXII evalúa ésta información para determinar la respuesta. Las respuesta posibles:

- El pedido de información podría ser denegado. En este caso el servidor de poll crea un error TAXII
- Un conjunto de contenido TAXII Data Feed podría ser provisto. En este caso, el servidor de Poll construirá y enviará un mensaje de Poll Response. Este mensaje indica el intervalo de tiempo que cubre el TAXII Data Feed que es transmitido y los cuerpos de mensajes STIX que forman el contenido del TAXII Data Feed.

En todos los casos, el cliente TAXII recibe el mensaje apropiado y pasa esta información al Backend TAXII para ser procesado.

4.5. Modelos

4.5.1. Comunidades

Actualmente el número de organizaciones que buscan compartir información de amenazas es creciente. Esto lleva a que el número y tipo de comunidades que busca hacerlo también se incremente. En la actualidad se identifican tres tipos de comunidades:

- Pares
- Comerciales

- Gubernamentales

Para que se pueda compartir información entre dichas organizaciones es necesario cierto nivel de confianza dado que compartir información sensible podría exponer a las organizaciones a daños en su reputación, demandas o advertir a un atacante de la investigación que se lleva a cabo haciendo que el trabajo realizado haya sido inútil. Se deben definir medidas para la protección de los datos como restricciones en su manejo, sanitización y el establecimiento de confianza entre las dos organizaciones. Lo mencionado anteriormente es particularmente importante cuando las organizaciones forman parte de varias comunidades que intercambian información, se encuentran casos en los que datos compartidos con una organización no deberían ser compartidos con otra.

Las comunidades entre pares son las más comunes, estas organizaciones o individuos tienen el propósito común de mejorar las defensas colectivas contra adversarios que tienen en común. La información compartida por dichas organizaciones es más especifica que la provista por organizaciones comerciales.

Las comunidades comerciales son anónimas y los miembros poseen algún tipo de acuerdo común, por ejemplo el pago de cuotas para pertenecer a la comunidad. La organización comercial maneja la información de forma centralizada y la distribuye entre los miembros de la organización. El acceso a la información por parte de los socios es rápido teniendo la posibilidad de que dicha información sea más amplia que la compartida por pares y que además no siempre sea aplicable a las necesidades de la organización.

Las comunidades gubernamentales son establecidas y manejadas por el gobierno, son voluntarias u obligatorias e incluyen participantes tanto del gobierno como de la industria privada. En ellas el gobierno controla la información y la distribución de esta. Así como en las comunidades comerciales, la información y los participantes son altamente confidenciales.

4.5.2. Modelos

Se pueden identificar tres modelos para el intercambio de información entre organizaciones:

- Hub and Spoke
- Peer to peer
- Source/subscriber

En el método hub and spoke, la entidad hub controla la recepción y diseminación de los datos, además se encarga de mantener anónimos y proveer un análisis adicional de los datos recolectados para luego diseminarlos entre los participantes. Este modelo es comúnmente visto en comunidades de gobierno o comerciales.

En el modelo peer to peer, los participantes intercambian y reciben información directamente de los otros participantes. La información es compartida entre todos los miembros de la comunidad por igual y la fuente está claramente identificada.

El modelo faltante es el source/subscriber, este modelo es utilizado por las comunidades comerciales que proveen de información. El proveedor de información envía regularmente información a todos los subscriptores y estos podrían eventualmente enviarle información a la fuente. Generalmente, la información se codifica de una manera propietaria y puede faltar información esencial sobre algunos intentos de irrupción. Presenta la ventaja de que se tiene acceso rápido a un conjunto de datos amplio y es útil para organizaciones con recursos limitados.

4.5.3. Implementación de Modelos en TAXII

A continuación se muestra como los servicios TAXII pueden ser utilizados para la implementación de los modelos para el intercambio de información.

4.5.3.1 Source/Suscriber

En este modelo una entidad es la fuente de información y algunos subscriptores tienen acuerdos con dicha entidad para recibir información periódicamente. Se busca que los subscriptores no se conozcan entre si, para ello la entidad fuente realiza acuerdos con cada uno de ellos. En este modelo, la fuente es un productor TAXII mientras que los subscriptores son consumidores.

TAXII soporta este tipo de modelo para el intercambio con el uso de los servicios de Discovery, Feed Management, Inbox y Poll. Una organización que desee subscribirse al TAXII Data Feed de la fuente necesita conocer los servicios TAXII que la fuente ofrece y como contactar con ellos. Si bien esto podría ser realizado por una mecanismo fuera de banda (e.g. publicando la información en otro medio) también podría ser logrado contactando al Discovery Service de la fuente. Desde este punto, el subscriptor podría contactar al servicio de Feed Managment identificado para aprender que fuentes ofrece el productor y que restricciones podría tener su acceso.

Si el contenido del TAXII Data Feed es restringido solo a algunas entidades autorizadas y el productor ha determinado que el subscriptor tiene permitido recibir el contenido, la fuente y el subscriptor necesitan acordar en como el subscriptor se autenticara. Dependiendo en el protocolo que soporta la fuente, esto se puede realizar por medio de una contraseña, un certificado u otro método. Si el contenido de un TAXII Data Feed es abierto y no requiere autenticación, éste paso es innecesario cuando se establecen las subscripciones al TAXII Data Feed.

Luego de la autenticación,el subscriptor puede contactar al Feed Managment Service del productor y pedir subscripciones a sus fuentes. La entidad fuente puede comparar dichos pedidos con su propio entendimiento de lo que el subscriptor puede recibir y permitir o denegar dichos pedidos según corresponda. La fuente puede enviar contenido al subscriptor al Inbox Service de éste en el intervalo apropiado. Alternativamente, el subscriptor podría contactar al Poll service de la fuente para descargar el contenido deseado.

[width=150mm]./images/SourceSuscriberModel.png

Figura 10: Flujo de información en modelo Source/Subscriber [2]

La figura 10 muestra como el modelo Source/Subscriber puede ser soportado por los servicios TAXII. También se ven los mensajes TAXII intercambiados entre la fuente y el subscriptor. Con los intercambios que están por encima de la línea punteada se establece la subscripción. Los intercambios pueden ser realizados repetidamente sin la necesidad de realizar el proceso de subscripción nuevamente.

4.5.3.2 Peer-to-peer

En un modelo Peer-to-peer, los pares de organizaciones entran en un acuerdo mutuo para compartir información entre si. En este modelo, cada Peer puede operar como productor y consumidor. Los socios en este intercambio podrían establecer fuentes utilizando un procedimiento similar al establecido en el modelo Source/Subscriber. Alternativamente, podrían acordar subir o descargar contenido sin ninguna subscripción formal. No tener una subscripción formal permite a un Peer albergar un Inbox Service sin necesidad de un Feed Managment Service.

El modelo Peer-to-Peer tiene dos variantes: acuerdos para el intercambio entre comunidades y acuerdos para el intercambio ad-hoc. En el primero la comunidad constituye acuerdos entre pares en los que todos los miembros acuerdan una única política la cual será respetada por todos. A diferencia de los otros dos modelos que tienen un punto central desde el cual la información es diseminada, todo el intercambio ocurre puntualmente entre dos pares. Si los pares A y B desean recibir información del peer C directamente, ambos necesitaran establecer un acuerdo apropiado con el peer C para que éste les envíe la información que desean.

Alternativamente, el intercambio entre pares puede ser realizado de forma individual con intercambios ad-hoc. Esto podría ocurrir si dos compañías hacen acuerdos individuales para compartir entre ellas. En este caso, los acuerdos sobre que compartir son específicos para las partes. Una sola entidad podría participar en ambas variantes, perteneciendo a una o mas comunidades en las cuales los miembros comparten entre si siguiendo un acuerdo común entre los miembros y a su vez negocian acuerdos individuales con otras entidades. Alguna información recibida por medio de un acuerdo no siempre debería ser compartida con otros pares que no son parte del acuerdo. De esta forma un participante debería hacer un seguimiento de quien fue el proveedor de la información recibida, como se realiza ese seguimiento está por fuera de la especificación de TAXII.

[scale=0.55]./images/PeerToPeerModel.png

Figura 11: Flujo de información en modelo Peer to Peer [2]

La figura 11 muestra como el modelo Peer to peer puede ser realizado por medio de los servicios TAXII. En este diagrama se ve que dos pares se contactan para pedir subscripciones para obtener información. Se asume que ambos pares tienen un Feed Managment Service que es utilizado para manejar todos los pedidos de subscripción.

4.5.3.3 Hub and Spoke

En un modelo Hub and Spoke, la entidad Hub es un consumidor de información que le proveen las entidades Spoke, pero a su vez se comporta como un productor que brinda información a entidades Spoke. Una entidad Spoke podría ser un productor, dando información al Hub, un consumidor que reciba actualizaciones del Hub o ambas. El Hub puede utilizar un Inbox Service para recibir información de cualquiera que desee enviar información de forma voluntaria y/o podría requerir información de ciertas fuentes para guardar la información en una única ubicación. Desde este punto, el Hub puede funcionar como una entidad Source del modelo Source/Subscriber mientras que los Spoke serían Subscribers de dicho modelo. El Hub puede adoptar cualquier política respecto de la información que recibe, desde pasar toda la información automáticamente a solo pasar información de socios reconocidos, o realizar ediciones y análisis antes de reenviar la información.

[scale=0.75]./images/HubAndSpokeModel.png

Figura 12: Flujo de información en modelo Hub and Spoke [2]

La figura 12 muestra como se puede implementar el modelo Hub and Spoke utilizando los servicios provistos por TAXII. En este modelo algunas entidades Spoke podrían ser consumidores, otras productores y en algunos casos ambas. El diagrama muestra los intercambios que podrían ser utilizados por el Spoke que actúa como productor y consumidor. Si este desea actuar de una sola forma solo los intercambios necesarios serían relevantes. Independientemente del rol que tome el Spoke, es necesario que éste conozca los servicios relevantes en el Hub. Esto se realiza utilizando el Discovery Service provisto por el Hub, de todas formas esto podría realizarse con mecanismos fuera de banda.

Referencias

[1] Mitre Corporation. *Active Defense Strategy for Cyber*. Mitre Corporation, 2012.

[2] Mitre Corporation. *Cyber Information-Sharing Models: An Overview*. Mitre Corporation, 2012.

[3] Mitre Corporation. *A New Cyber Defense Playbook*. Mitre Corporation, 2012.

[4] Mitre Corporation. *STIX Whitepaper*. Mitre Corporation, 2012.

[5] M.J. Cloppert E.M. Hutchins and R.M Amin PH.D. *Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains*. Lockhead Martin, 2012.

[6] D. Curry H. Debar and B. Feinstein. *RFC 4765 - IDMEF*. IETF, 2007.

[7] M. Richard J. Connolly, M. Davidson and C. Skorupka. *TAXII Whitepaper*. Mitre Corporation, 2012.

[8] D. Rolsky D. Chamberlain J. Vincent, R. Spier and R. Foley. *RT Essentials*. A. Randal and T. Apandy, 2005.

[9] K. Moriarty and B. Trammell. *RFC 6546 - RID Messages*. IETF, 2010.

[10] J. Meijer R. Danyliw and Y. Demchenko. *RFC 5070 - IODEF*. IETF, 2007.

[11] B. Trammell. *RFC 6546 - Transport of RID messages over HTTP/TLS*. IETF, 2012.