CURSO 2020-2021

MÁSTER EN BUSINESS INTELLIGENCE Y DATA SCIENCE

MODULO:

Ciberseguridad en BI y Data Science

Nombre estudiante: Marc Faravelli Rodríguez

No.Matricula: 2047104

E-mail: marcfaravelli@gmail.com

Preguntas del Caso Práctico

1. ¿Se te ocurre/n alguna/s otra/s regla/s de BI que puedan ayudar a disminuir el fraude en la entidad bancaria?

Las técnicas de detección de anomalías pertenecen a la disciplina de minería de datos y le permiten encontrar eventos raros que difieren significativamente de la mayoría de los datos en un conjunto de datos. La identificación rápida de eventos anómalos, que ocurren raramente o que nunca han ocurrido en el pasado, puede permitir una intervención reactiva y precisa, anticipando la evolución de situaciones nocivas o la pérdida de oportunidades. Estas técnicas se utilizan para detectar comportamientos ilegales en Internet, como ataques de piratas informáticos, estafas de seguros, fraude bancario etc

Los autores de los fraudes tratan de engañar a la contraparte enmascarando el comportamiento fraudulento como un comportamiento perfectamente lícito. En cuanto a las intrusiones informáticas, muchas técnicas se pueden adaptar con variaciones más o menos consistentes a diferentes contextos y por ello es muy difícil identificarlas con precisión y rapidez.

Dada la naturaleza misma de ciertos sistemas de detección de fraude, producen una gran cantidad de falsos positivos y, por lo general, tienen una tasa de detección baja. Un aspecto aún más importante es que las reglas no son adaptables y, por lo tanto, pierden efectividad a medida que los estafadores evolucionan sus estrategias, lo que requiere una actualización manual constante. Por estos motivos, sería oportuno poner el foco en el estudio e implementación de un sistema antifraude que utiliza metodologías y enfoques de aprendizaje automático (machine learning).

2. Dentro de los 3 ejes de la ciberseguridad CIA (Confidencialidad, Integridad y Disponibilidad) y viendo tu sistema de Big Data, no solo como los sistemas operativos, sino también los datos, tanto los datos en crudo como la información de negocio que sale del sistema: ¿Podrías argumentar cómo mitigaríamos el riesgo en cuanto a la ciberseguridad en cada uno de los 3 ejes?

Los conceptos de confidencialidad, integridad y disponibilidad están estrechamente relacionados con el diseño de un sistema de gestión de seguridad de datos que debe considerarse de manera unificada:

1. Confidencialidad

Una estrategia dirigida a la privacidad de la informática debe, en primer lugar, ofrecer confidencialidad, es decir, garantizar que los datos y los recursos se preserven de un posible uso o acceso por parte de terceros no autorizados. La confidencialidad debe garantizarse en todas las etapas de vida de los datos, desde su almacenamiento, durante su uso o su tránsito por una red de conexión.

Existen diversas herramientas que pueden utilizarse para garantizar la confidencialidad de la información: el cifrado de las comunicaciones, los procedimientos de autenticación, la creación de modelos de gobernanza de datos bien definidos y acciones de sensibilización a los usuarios. El concepto de confidencialidad no es único, ya que existen varios elementos que deben ser considerados por la organización en relación a su negocio, por ejemplo el grado de sensibilidad de la información que se procesa y el nivel de criticidad y secretismo que la caracterizan.

2. Integridad

Hablar de integridad significa tener en cuenta diferentes escenarios: evitar cambios no autorizados en la información por parte de los usuarios, pero también garantizar que la información en sí sea identificable y verificable de manera única en todos los contextos en los que se utiliza.

Para garantizar la integridad, es necesario implementar políticas de autenticación claras y monitorear constantemente el acceso y uso efectivo de los recursos, con herramientas capaces de crear registros de auditoría. El control de acceso (por ejemplo, a través de los sistemas de Gestión de Identidad y Acceso), los procedimientos de autenticación, los sistemas de Detección de Intrusos, las restricciones de acceso y, una vez más, la formación de los usuarios representa soluciones útiles para respetar este principio.

3. Disponibilidad

Hacer que un servicio esté disponible significa esencialmente dos cosas: evitar que se produzcan interrupciones del servicio durante el intervalo de tiempo definido y garantizar que los recursos de la infraestructura estén listos para la correcta prestación de lo que se requiere.

Por lo tanto, se deben implementar mecanismos para mantener los niveles de servicio definidos, haciendo uso de herramientas de recuperación ante desastres, respaldo y continuidad del negocio, capaces de limitar los efectos de una posible indisponibilidad del servicio o pérdida de datos. Las contramedidas que se pueden implementar se refieren, por ejemplo, al diseño de infraestructuras de red capaces de garantizar la redundancia de los sistemas y ofrecer los servicios requeridos incluso en caso de falla o accidente, sistemas de firewall capaces de proteger las redes, sistemas de monitoreo de tráfico interno y continuo. Las políticas de Continuidad del Negocio también aseguran la implementación de soluciones capaces de limitar posibles puntos de ataque.

3. A nivel de los sistemas que deberían proteger tu plataforma de Big Data: ¿Debería estar segmentada de la red interna o de lo contrario podría estar integrada dentro de alguna red interna?

Si se desea ejecutar servicios accesibles desde el exterior, como HTTP, correo electrónico, FTP y DNS, se recomienda que estos servicios disponibles públicamente estén segmentados física y / o lógicamente por la red interna. Sin embargo, los hackers motivados siempre pueden encontrar una forma de acceder a la red si los servicios que

han sido violados residen en la misma ruta lógica que el resto de la red. Los servicios accesibles externamente deben residir en lo que el campo de seguridad reconoce como una zona desmilitarizada (DMZ), un segmento lógico de la red, donde el tráfico entrante de Internet solo podrá acceder a esos servicios y no estará habilitado para acceso a la red interna.

4. ¿Cuáles son los Pros y Contras de tenerla segmentada?

Pros

La división de una red en segmentos permite distintos niveles de seguridad para los diferentes segmentos. Un puente no puede ejecutar software de seguridad, por lo que los enlaces entre segmentos protegidos por una mayor seguridad son implementados por otros componentes de hardware. La segmentación de seguridad solo puede implementarse mediante un dispositivo controlado por software, como un enrutador. Cuando se segmenta la red, esta se divide en partes que no se comunican o que están separadas por controles de seguridad. El objetivo es asegurar que cualquier problema (la propagación de malware o el rango de acción de quienes han entrado en una fuga de red) quede confinado al segmento de la red donde ocurrió, evitando infectar a otros.

Contras

Esta solución tiene algunos inconvenientes importantes. Primero, el equipo adicional conlleva costes adicionales; en segundo lugar, la modificación de una infraestructura de red existente es siempre un problema para los administradores de sistemas. El motivo radica en su complejidad arquitectónica, que ciertamente no es insuperable, pero parece muy preocupante para quienes están acostumbrados a las redes "planas". Y es muy probable que en las empresas sean precisamente los que no se ocupan de la seguridad los que más se oponen a la segmentación de la red, como elemento que complica la comunicación de las aplicaciones corporativas entre sí y con los datos.

5. Si decidimos segmentarla: ¿crees necesario instalar un firewall?

Si, es aconsejable instalar al menos un firewall de hecho son una forma eficaz de disuadir a los atacantes casuales. Es un sistema diseñado para evitar el acceso no autorizado a una red privada y viceversa, es decir, evita conexiones peligrosas desde fuera de la red corporativa pero también lo contrario, es decir, puede evitar que, desde dentro de la red corporativa, se visiten páginas webs potencialmente peligrosas. Los firewalls pueden ser de software o hardware, sin embargo, estos últimos brindan un mayor nivel de seguridad y, por lo tanto, se prefieren para redes corporativas donde hay servidores y la seguridad tiene la máxima prioridad.

En última instancia, preguntarse por qué instalar un firewall en nuestra red es como preguntarse por qué deberíamos instalar una puerta de seguridad en nuestro hogar. Hoy, lamentablemente, así como no existe una zona residencial en la que uno pueda sentirse a salvo de visitas desagradables, tampoco hay lugar en Internet donde un atacante no pueda pasar intentando violar nuestro sistema, con el objetivo de robar

información o use nuestras conexiones de datos para llegar a nuevos destinatarios y comenzar un efecto dominó.

6. Si decidimos añadir un IPS (Intrusion Protection System): ¿Por qué ganaríamos en seguridad?

Los IPS se encuentran entre los dispositivos de seguridad más sofisticados que se utilizan en este momento. Inspeccionan los paquetes de red y bloquean los paquetes sospechosos, además de alertar a los administradores sobre los intentos de ataque. Los registros de estos sistemas contienen información valiosa sobre las amenazas de la red con respecto al tipo de ataque, los dispositivos objetivo y más. Es bueno monitorear estos registros y extraer la información que brindan para aumentar la seguridad de su red.

La mayor ventaja es que luego de haber constatado la posibilidad de un ataque, este tipo de sistema no se limita solo a informar al administrador, sino que activa inmediatamente las medidas de seguridad adecuadas. De esta forma, evitan un intervalo de tiempo demasiado largo entre la detección de un intruso y la implementación de acciones para detenerlo.

7. ¿Qué entiendes por Hardening de Servidores?

El desarrollo e implementación de medidas de seguridad, utilizando las mejores prácticas, se conoce como "Hardening" de Servidores o Sistemas. Es un proceso continuo para identificar y comprender los riesgos de seguridad y para implementar las medidas adecuadas para contrarrestarlos. El proceso es dinámico ya que las amenazas y los sistemas a los que se dirigen están en constante evolución. El Hardening del sistema se utiliza para poder implementar acciones que mitiguen las amenazas para cada etapa del ciclo de vida de las amenazas mismas.

Para ser eficaz, requiere mantener actualizados los conocimientos sobre seguridad y todas sus novedades. Implica: estar al tanto de los problemas de software y hardware, incluidos los sistemas operativos, dispositivos móviles, cámaras, codificadores, dispositivos de almacenamiento y dispositivos de red. En última instancia, establecer un punto de contacto para todos los componentes del sistema.

8. Al tener las redes segmentadas y requerir que los flujos de datos pasen por Firewalls para una mayor seguridad: ¿qué problemas pueden conllevar este escenario comentado?

Las redes segmentadas o DMZ que consiste en segmentos LAN aislados (unas "subredes") accesibles desde redes internas y externas, son caracterizadas por el hecho de que los hosts conectados a la DMZ tienen posibilidades limitadas de conectarse a los hosts específicos de la red interna.

En cuanto a los firewalls pueden tener diferentes desventajas:

- Acceso limitado para usuarios: si es una organización grande, la restricción de usuarios puede ser bastante molesta para los usuarios;

- Rendimiento: los firewalls basados en software pueden limitar el rendimiento general de la computadora y ralentizarlo. En general, los firewalls pueden consumir muchos recursos del sistema cuando se ejecutan en segundo plano y obstaculizar el funcionamiento general;

- Indefenso ante ataques de malware: si bien los firewalls pueden bloquear troyanos, de ninguna manera son efectivos contra virus y otro malware. Este malware puede ingresar al sistema camuflado. Entonces, incluso si se tiene un firewall, el ataque puede ocurrir. Si es así, se debe de tener instalada una herramienta anti-malware;

- Complejidad: en general, los firewalls en sistemas pequeños son bastante simples. Sin embargo, este no es el caso de las organizaciones más grandes. Un sistema de firewall completo instalado en una organización más grande requiere personal dedicado independiente para mantenerlo. Por lo tanto, se tendrá que asumir un costo adicional para contratar expertos. Además, administrar un firewall complejo puede resultar costoso en muchos casos.

9. Mírate este enlace y responde brevemente por qué es importante, principalmente, clasificar los datos a nivel de seguridad.

Es evidente que hoy en día los datos están creciendo y se necesita una cierta ratio para administrarlos de manera eficiente y segura. Las empresas tienen la tarea de definir el nivel de sensibilidad de los datos procesados con el fin de implementar controles adecuados y así proteger tanto al negocio como a los clientes.

Los datos (en todas sus diferentes formas) representan un activo valioso y tangible para todas las empresas y, por lo tanto, deben protegerse con el nivel adecuado de protección. En esta perspectiva, un único estándar uniforme en todas las actividades de una empresa no es la solución óptima: es conveniente aplicar diferentes niveles de seguridad en función del tipo (y por tanto del valor) de la información a proteger.

Por lo tanto, el primer paso a dar es crear procesos que sean útiles para clasificar sus datos y determinar su valor: puede, por ejemplo, confiar en el nivel de sensibilidad (basado en el riesgo de pérdida o divulgación) o en la importancia que tiene su uso dentro de las actividades de la empresa.

Hay varios parámetros que se pueden utilizar para clasificar los datos. El criterio más utilizado en el sector privado, por ejemplo, es el valor de la información. Cuanta más información sea valiosa para una empresa, más debe protegerse. Incluso el factor tiempo debe tenerse en cuenta y le permite evaluar la degradación de ciertos datos: la importancia de la información, de hecho, puede disminuir con el tiempo. Otro criterio útil para evaluar la degradación es el ciclo de vida: los datos de una empresa pueden volverse obsoletos por una multitud de razones, desde la actualización con nueva información hasta cambios sustanciales dentro de la empresa. Por lo general, en la

comunicación interempresarial es necesario tener en cuenta toda una serie de variables para optimizar la gestión de datos.

10. Imagínate que tienes que gestionar un incidente relacionado con una fuga de datos de la información ya procesada de utilidad para el negocio. Tu equipo de seguridad ha detectado un Command & Control (telecontrol de un sistema que un hacker realiza desde fuera de la empresa a través de Internet) y tienes el dilema de erradicar cuanto antes el problema o buscar evidencias Forensics para una posible denuncia: ¿Cuál sería tu decisión?

Cuando se detecta un Command & Control hay algunas acciones que se podrían realizar y son:

- Desconectar las máquinas comprometidas de Internet especialmente de la red local.
- Actualizar el antivirus e instalar los parches del sistema operativo.
- Utilizar herramientas anti troyanas.
- Bloquear todas las tarjetas cuyos datos bancarios se almacenaron en el anfitrión comprometido.
- Cambiar todas las contraseñas del host atacado.

11. Respecto a las estrategias de Recuperación del Negocio del banco que va a instaurar un Data Science para mejorar las ventas y teniendo en cuenta que el coste de la estrategia debe ser coherente con los datos a proteger: argumenta si utilizar un COLD, WARM, HOT SITE.

Si la prioridad es reducir costos, el Cold Site es probablemente la mejor opción. Si, por el contrario, el proceso de producción no permite tiempo de inactividad en la recuperación de datos, el uso de este sistema podría ser un problema.

Antes de elegir la Recuperación del Negocio, se debe de considerar su objetivo de tiempo de recuperación (RTO) y los datos del objetivo del punto de recuperación. Si el tiempo de recuperación es corto, un Hot Site es una buena opción porque los sistemas y las configuraciones ya están configurados para requisitos específicos.

Si en cambio, se tiene un RTO extendido durante más tiempo, se puede configurar un Cold Site. Incluso si lleva más tiempo instalarlo, es posible que no tenga ningún efecto en nuestro negocio y, al mismo tiempo, reduzca los costes.

12. Imagínate que tu Data Science lo tienes afinado y empieza a dar información que guía o enfoca las campañas de marketing a públicos objetivos ofreciéndoles productos que encajan a sus necesidades. Se filtra esta información a la competencia. En esta situación: ¿crees que tu empresa estaría más expuesta a un ataque APT? Argumenta la respuesta.

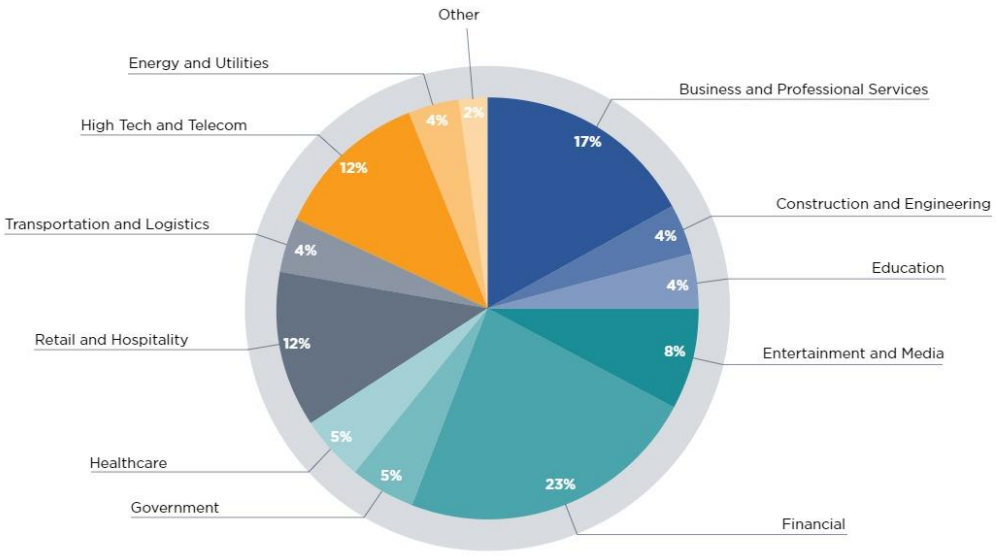
Una APT requiere enormes recursos, tecnologías y tiempos muy largos, por lo que tiene costes extremadamente altos.

Esta consideración es fundamental para comprender qué víctimas se eligen para ataques similares: para que haya un retorno positivo de la inversión, los objetivos de las APT siempre suelen ser de alto perfil, como estados soberanos o grandes corporaciones.

Es importante subrayar que las pequeñas y medianas empresas, sin embargo, no están a salvo de los ataques relacionados con las APT.

Las PYME suelen formar parte de la cadena de suministro de las grandes empresas, pero no comparten sus elevados estándares de seguridad. Esto, a ojos de los actores detrás de las APT, las convierte en una base perfecta desde la que iniciar la infiltración ascendiendo en la cadena de suministro hasta alcanzar el verdadero objetivo estratégico.

Los sectores afectados por las APT son muchos: en la siguiente imagen se puede ver una estadística reciente de FireEye con los porcentajes de industrias afectadas:



Los primeros en la lista son las finanzas y el comercio, seguidos de la hospitalidad, la tecnología y el entretenimiento / medios.

Las entidades bancarias siempre han sido un objetivo principal para los ciberdelincuentes. Sin embargo, los ataques llevados a cabo por estos últimos casi siempre están dirigidos a clientes de instituciones bancarias. A veces los atacantes también pueden apuntar directamente a las entidades financieras, llevando a cabo un

ataque decidido, altamente profesional y bien coordinado; los ciberdelincuentes, además, utilizan numerosos medios y herramientas, con el objetivo preciso de robar la mayor cantidad de dinero posible, hasta llegar a un límite, aparentemente autoimpuesto. Por estas razones, considero que mi empresa tenga que estar preparada con las correspondientes medidas de seguridad ya que tiene probabilidades altas de ser atacada.

13. ¿Por qué deberíamos anonimizar los datos de los clientes?

El valor potencial del anonimato de los datos de los clientes se consideraría como una estrategia para permitir que las personas y la empresa en un sentido amplio disfruten de los beneficios de los "datos abiertos", al tiempo que se mitigan los riesgos para las personas interesadas. Sin embargo, lo difícil es crear conjuntos de datos verdaderamente anónimos mientras se mantiene toda la información subyacente necesaria para realizar la actividad solicitada.

A este propósito hay diferentes técnicas de anonimización que pueden proporcionar garantías de protección de la privacidad y pueden utilizarse para crear procedimientos de anonimización efectivo. Eso sí, solo si su aplicación está correctamente diseñada para lograr la anonimización deseada al tiempo que produce datos útiles. La solución óptima debe decidirse caso por caso, si es posible utilizando una combinación de diferentes técnicas.

14. En el caso que parte de tu solución requiriera servicios SaaS en el Cloud: ¿Qué tienes que decir sobre las medidas de seguridad que adoptarías para interconectar tu red al Cloud?

En el modelo SaaS, la responsabilidad de implementar y mantener la seguridad se comparte entre el proveedor de la nube y el proveedor de servicios. SaaS hereda los problemas derivados de PaaS e IaaS, ya que depende de ellos, incluida la seguridad de los datos y la red. También se debe considerar el análisis de vulnerabilidades y multi-tenancy, lo que conlleva problemas provocados por el intercambio de recursos como el hardware y las propias aplicaciones, arriesgando la pérdida de información.

En los entornos de Software como Servicio (SaaS), las medidas de seguridad y su alcance se formulan a través de contratos. En este modelo, como ya se ha comentado previamente, es principalmente el proveedor quien debe garantizar las medidas de seguridad. Él, por ejemplo, será responsable de proporcionar a los usuarios un sistema de autenticación y un sistema de autorización que incluye un conjunto predefinido de roles, autorizaciones y reglas comerciales.

15. ¿Crees necesario que Cloud Computing puede ser de ayuda para tu Análisis de Datos? Razona la respuesta.

El Cloud Computing puede ofrecer diversas características de servicio que nos pueden resultar útil:

- Infraestructura “ágil”: los servicios de TI, basados en la nube, son cada vez más capaz de apoyar a una fuerza laboral en constante crecimiento, permitiendo a los consumidores acceder a sus recursos dondequiera que estén;

- Respuestas rápidas: la nube nos permite expandirnos fácilmente pudiendo cubrir las necesidades. Ya no es necesario tener que comprar recursos informáticos, ya que la nube nos garantiza variaciones y solicitudes de procesamiento y potencia de cómputo "bajo demanda";

- Menores costes: no hay más costes de mantenimiento excesivos de las infraestructuras de TI. Un servicio de Cloud está disponible "bajo demanda", siguiendo la lógica de "pago por uso", es decir, comprando solo lo que se utiliza. Se reduce el despilfarro de recursos y, en consecuencia, el gasto. Si se requiere más potencia informática, es posible adoptar nuevas aplicaciones sin tener que hacer frente a costosas inversiones.

- Menos fallos del sistema: con la adopción de la computación en la nube, se evitan los problemas de interrupción del servicio. Esto significa que, si un solo nodo falla, todavía tiene la funcionalidad del servicio. Los tiempos de pedido, construcción, instalación y configuración son extremadamente cortos. Además de los costes de mantenimiento y reparación.

- Fácil de usar: su interfaz es bastante simple, que en la mayoría de los casos es web, lo que conduce a una relación fácil entre nuestros usuarios y la propia aplicación en la nube.