3. Protección de Datos

4. Data Leak Prevention

4.1 Introducción

Prevención de la fuga de datos (Data Leak Prevention) y Prevención de la pérdida de datos (Data Loss Prevention) son términos para resumir un enfoque de gestión de la seguridad de la información con muchos años de vida. El DLP se focaliza en analizar y entender los flujos de datos en una organización, la situación de los datos (en movimiento, en uso o almacenados), e implantar las medidas de seguridad necesarias para proteger su confidencialidad.

DLP se entiende mejor pensando en distintos escenarios que afectan a la seguridad de la información de cualquier empresa que en la actualidad necesite proteger activos digitales como diseños, imágenes, planes, documentación sensible, etc. Esta metodología, podría utilizarse para evitar que empleados descontentos se llevaran información sin autorización de una organización imprimiéndola, en un dispositivo USB de almacenamiento (pendrive, disco externo...) o enviándola por Internet (mediante correo electrónico, redes sociales, sistemas de almacenamiento externo como Dropbox…).

Las soluciones de DLP no solo permiten bloquear la impresión, el copiado de datos a dispositivos USB o el envío de datos por Internet, sino que permiten controlar y monitorizar dichos flujos de datos: permitiendo la impresión, copiado o envío a determinados grupos de usuarios y registrando todas las impresiones, copias y envíos en un histórico.

La implantación de un sistema de este tipo no es trivial. Se requiere un trabajo inicial de análisis de los flujos de datos, clasificación de la información, análisis de riesgos y configuración de sistemas. Sin embargo, con el inicio de un proyecto de este tipo, se pueden obtener resultados inmediatos que ayudan a prevenir daños mayores mientras se trabaja en una implantación global de una solución definitiva.

4.2 Soluciones DLP

Cada poco tiempo, los fabricantes de estas soluciones introducen nuevos productos en el mercado. Aunque esto varía de un proveedor a otro, se dan comúnmente tres niveles diferentes para esta solución y se les conoce como datos en reposo, datos en movimiento y datos en los puntos finales.

Entonces, ¿cómo difiere DLP de cualquier otra tecnología de seguridad? Mientras que herramientas tales como cortafuegos e IDS/IPS buscan cualquier cosa que pueda representar una amenaza para una organización, DLP está interesada en identificar datos sensibles. Se busca el contenido que es fundamental para una organización.

Podría parecer que DLP es una solución cuyo único propósito es impedir las violaciones de datos por parte de intrusos. Si bien se puede evitar este tipo de fugas de datos, la mayoría de las veces esta solución se utiliza como un mecanismo para el descubrimiento de procesos que fallan en el curso normal de los negocios. Un ejemplo sería la presencia de datos sensibles en el ordenador portátil de un socio.

Las organizaciones gastan gran cantidad de tiempo y dinero en formación del usuario. Por lo tanto, es de suponer que las fugas de datos como resultado de acciones involuntarias de un usuario deben ser mínimas. Esto no es cierto. Sabemos que es un hecho que la mayoría de las empresas que sufren brotes de malware se deben a este tipo de acciones de los usuarios. Esta tendencia no ha cambiado mucho a pesar de la formación y concienciación del usuario actual.

Aunque aparezca por escrito que las políticas y los procedimientos se siguen correctamente, sólo cuando se aplican controles de detección vemos el panorama completo de en qué medida se están aplicando realmente. Con el fin de tener éxito en la aplicación de una política, el control de detección por sí solo no es suficiente. Si bien el control de detección proporcionaría visibilidad, el control preventivo es una necesidad para reducir las fugas de datos tanto de forma accidental como intencional. DLP es una tecnología que puede ayudar a hacer cumplir estas políticas de manera efectiva.

4.3 Identificación de Datos Sensibles

Del mismo modo que ocurre en la minería de datos, el uso de expresiones regulares es una herramienta útil para encontrar contenido coincidente en DLP. Esto se vuelve aún más preciso cuando se aplica el cruce de datos contra contexto. Por ejemplo, si se observa un empleado de administración revisando la nómina de otra persona, este evento es un comportamiento normal y puede ser ignorado. Sin embargo, si este evento llegara a ocurrir en otro departamento, el DLP debe levantar una alerta y por lo tanto debe ser escalado.

Casi todos hablan de comparar datos estructurados y datos no estructurados. Los datos estructurados son aquellos que existen en formatos definidos, como los números de la seguridad social o de tarjetas de crédito. Los datos no estructurados son aquellos que no se ajustan a un formato definido, que es todo lo demás. Algunos ejemplos de datos no estructurados son los códigos de fuente, archivos multimedia, etc. Por lo que se refiere a los datos estructurados, un formato predefinido simplifica la construcción de la expresión regular. En cuanto a los no estructurados, no nos queda otra usar las “huellas digitales de los datos” debido a su formato complejo. Las huellas digitales se hacen usando una forma de control seguro y se guardan en una base de datos. Esta información puede ser usada para la identificación de dicho contenido sensible en otros lugares. Dependiendo del resultado, se toma la decisión de si requiere o no una razón para levantar una alerta.

Unas expresiones regulares precisas son importantes para minimizar los falsos positivos. Esto, junto con el contexto aumentará en gran medida la capacidad de detección real y nos ahorra tratar con los falsos positivos. De esta forma, podemos invertir más tiempo en los esfuerzos de remediación.

Por ejemplo, echemos un vistazo a esta firma de la herramienta Snort:

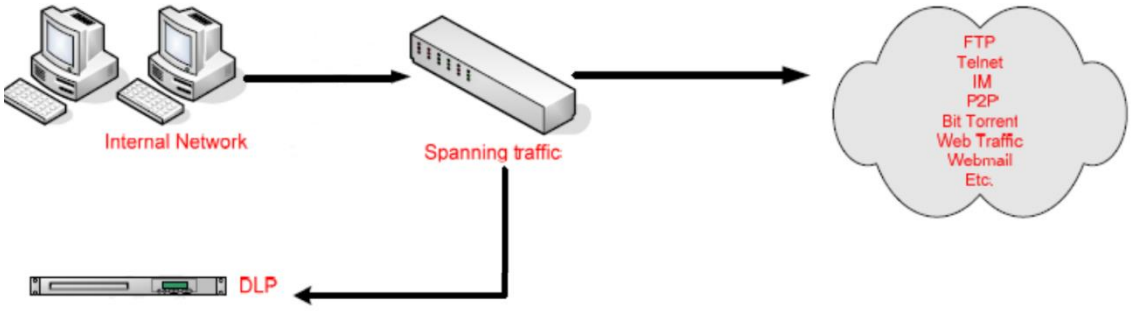
alert tcp $EXTERNAL_NET any -> $SQL_SERVERS 1433 (msg:"SQL SA brute force"; flow:to_server,established; content:"|10|"; depth:1; content:"|00 00|"; depth:2; offset:34; content:"|00 00 00 00|"; depth:4; offset:64; pcre:"/^.{12}(\x00|\x01)\x00\x00(\x70|\x71)/smi"; byte_jump:2,48,little,from_beginning; content:"s|00|a|00|"; within:4; distance:8; nocase; threshold:type threshold, track by_src)

Además de una cadena de comparación usando la expresión regular *pcre:"/^.{12}(\x00|\x01)\x00\x00(\x70|\x71)/smi”*, también podemos definir la profundidad de inspección del paquete donde llevar a cabo esa comprobación.

4.4 Estados de la Información

4.4.1 Datos en movimiento

Esta característica de una solución de DLP se aplica a todos los datos en el cable. Veamos a continuación un ejemplo de colocación de este dispositivo.



Como se muestra arriba, todo el tráfico que de la red interna a través de cualquiera de los canales comunes se reflejará a DLP para la inspección.

Esto proporciona visibilidad en un gran número de violaciones. Por ejemplo, si un archivo sensible fue transferido a través de FTP. FTP, un protocolo que usa texto simple debe ser la primera preocupación en la transmisión de los archivos confidenciales. En segundo lugar, esto lleva a la pregunta de si el archivo nunca debe salir de la empresa. En tercer lugar, tendremos que verificar si las partes involucradas están autorizadas para ver y transmitir datos. La mayor parte de esto se aplica no sólo a FTP, sino cualquier canal de comunicación.

4.4.2 Datos en Reposo

Como su nombre indica, se aplica a todo lo que almacene datos, tales como recursos compartidos, bases de datos, etc. El descubrimiento de datos tiene dos usos. El uso principal de esta función es el descubrimiento de datos sensibles en los repositorios de datos. Esto utiliza la política actual para buscar cualquier información sensible. El escaneo de descubrimiento también se puede utilizar para obtener las huellas dactilares que serán utilizadas en otra parte de la identificación de datos no estructurados.

Como se muestra arriba, este dispositivo puede ser colocado en cualquier lugar de la red con el único requisito de conectividad IP a los objetivos. Los DLP están equipados para crear varias sesiones virtuales para reducir al mínimo la necesidad de varios dispositivos en la red. Cada sesión virtual puede ser configurada para analizar un conjunto de servidores en una red dada.

Esto es ideal para redes más grandes. Mientras que el uso de ancho de banda puede ser una preocupación para tan alto volumen de tráfico, hay soluciones para este problema. Una de las soluciones es la utilización de la función de escaneado progresivo. Una vez que un servidor está totalmente explorado, una exploración gradual sólo buscará cambios desde el último análisis.

Las revelaciones de datos más comunes que se identifican durante esta exploración de descubrimiento son aquellas en las que residen datos críticos en servidores DMZ sin el conocimiento del cliente.

4.4.3 Datos en los Extremos

Los datos en los puntos finales es una solución basada en un agente que se encuentra en las estaciones de trabajo del usuario final y los portátiles haciendo un seguimiento de todos los datos que salen a través de dispositivos extraíbles, como disquetes, CDs, memorias USB, etc. Esto también proporciona auditoría y protección contra los usuarios que imprimen datos clasificados.

Debido a su enfoque basado en agente, en realidad no ha sido una solución que haya gustado mucho entre los clientes. Sin embargo, prevé una gran protección frente a los datos que salen a través de dispositivos extraíbles. La implementación de esta solución es comparable a un IDS basado en host (Intrusion Detection System).

4.5 Elegir una Solución

Hay varios elementos en la elección de un proveedor de soluciones DLP. Vamos a señalar las áreas clave que los clientes deben considerar. Los criterios de evaluación usados son los siguientes:

1. Monitorización vs prevención
2. Gestión centralizada
3. Requisitos de copia de seguridad y almacenamiento
4. Facilidad de integración
5. Presencia en el mercado
6. Personal adicional

4.5.1 Monitorización vs Prevención

Los vendedores utilizan nombres complejos y elegantes para estas dos características. Aunque una mirada más profunda en la solución podría revelar algunas características únicas para cada proveedor, a un nivel más alto vemos que se refieren simplemente a al funcionamiento de DLP en modo de supervisión y en modo de prevención. Por ejemplo, un importante fabricante afirma que una de sus características clave es su capacidad para mover los archivos expuestos. Una buena analogía para la discusión de si la tecnología de protección de contenido debe ejecutarse en el modo de monitorización o en el modo de prevención es la comparación entre los sistemas de detección de intrusos (IDS) y los sistemas de prevención de intrusos (IPS). Cuando IPS fue introducido por primera vez, se pensaba que esta tecnología era capaz de bloquear, si no todos, la mayoría de los ataques y que los falsos positivos, simplemente desaparecerían. Poco sabían los clientes en ese momento de que sólo un puñado de firmas podría ir en modo bloqueo y que era necesario un estudio a fondo del entorno para crear otras reglas y ampliar la capacidad de bloqueo.

La misma regla se aplica a DLP también. La precisión de una firma es muy importante antes de decidirse a poner en cuarentena o bloquear una determinada actividad. Por otra parte, DLP puede requerir hardware y el software adicional con el fin de hacer cumplir la prevención. Para citar un ejemplo, si optamos por bloquear un correo electrónico con datos sensibles, algunos proveedores pueden requerir la integración con un MTA, como IronPort, Sendmail, Proofpoint, etc. Optar por el modo de prevención puede ser muy costoso, especialmente si tenemos la intención de bloquear canales múltiples. Además del costo, la facilidad de integración debe tenerse en cuenta en también. Es importante que todos los objetivos de futuro se incluyan en el ámbito de aplicación.

4.5.2 Gestión Centralizada

La sobrecarga en la gestión es la pesadilla de cualquier organización. La gestión centralizada puede reducir en gran medida esta sobrecarga. Algunas de las características clave a incluir son la creación y aplicación de políticas, informes y filtros de datos.

4.5.3 Requisitos de Backup y Almacenamiento

Cada organización tiene un conjunto de requisitos para el almacenamiento de datos. Aunque la mayoría de los vendedores de DLP se basan en software, hay algunos que son basados en

hardware. El producto llega en un dispositivo de hardware y tiene la capacidad de retener los datos durante periodos de tiempo significativos. Si la política de retención de datos establece que los datos deben conservarse durante seis meses, algunos productos basados en hardware están construidos para soportar terabytes de datos. Esto puede ser una buena solución para las organizaciones con un presupuesto más ajustado.

4.5.4 Facilidad de Integración

Algunos elementos pueden desempeñar un papel significativo en la facilidad de integración. Los vendedores no siempre tienen la solución en la mano para satisfacer las necesidades de un cliente. Hay temas complejos que verán la luz sólo cuando la implementación se lleve a cabo. Una de las cuestiones que podríamos encontrarnos es la función de descubrimiento de datos sin agente. Todas las plataformas de operación que formen parte de la digitalización deben tomarse en consideración. En algunos casos, la función de descubrimiento sin agente puede estar disponible para sistemas basados en Windows, y sin embargo requerir un agente en AIX OS.

Si la política de la compañía establece que dichos agentes no pueden estar en ejecución en los servidores críticos, el despliegue se complica. Muchas veces, esta excepción llevaría a convocar una reunión con la Junta Directiva de Tecnología (TSB) y puede retrasar el proyecto de manera significativa.

4.5.5 Presencia en el Mercado

Este es un factor clave a considerar en la elección de un proveedor. Un vendedor con buena presencia en el mercado ya ha experimentado y se ha ocupado de los problemas de implementación. En segundo lugar, esto puede ayudar con la creación de la política, que es el núcleo de esta tecnología y tiene un impacto directo sobre el flujo de trabajo. Para aquellos que tengan que cumplir con regulaciones del gobierno, hay políticas predefinidas que estas organizaciones pueden utilizar. Si un proveedor en particular ya tratado con organizaciones de salud, si no todos, la mayoría de los requisitos son muy similares en un punto de vista regulatorio y este proveedor en particular puede ser una buena opción para otras organizaciones de salud.

4.5.6 Necesidades de Personal

Cuando los IDS hicieron su primera entrada en la industria de la seguridad, muy pocas organizaciones se dieron cuenta de la necesidad de personal dedicado a eliminar a los falsos positivos y tratar con amenazas reales. En la actualidad, casi todas las organizaciones que han desplegado dispositivos IDS, emplean personal suficiente para cubrir una operación 24/7.

DLP está en sus primeras etapas y hay que estimar la cantidad de trabajo adicional que esto puede crear y la necesidad de personal dedicado. Hemos visto bastantes falsos positivos en el mundo IDS para darnos cuenta de que DLP no es ninguna excepción. Aunque el contenido que se busca es diferente, el mecanismo es el mismo.