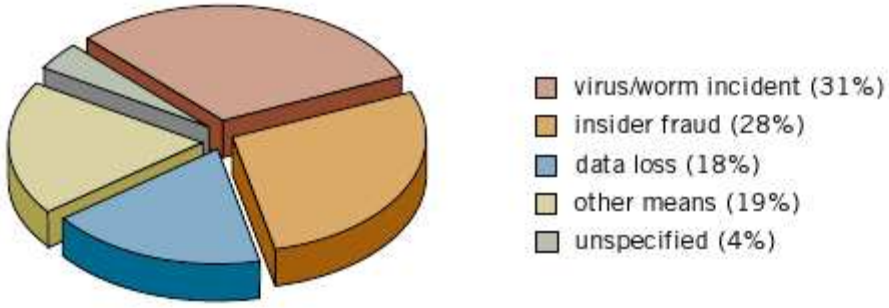
DATA LOSS PREVENTION (DLP)

Hemen her gün yeni bir bilgi hırsızlığı ya da kişisel bilgilerin yersiz ifşa edilmesine tanık oluyoruz. Şirketlerin finansal bilgileri, stratejik hedefleri, planları, bütçe uygulamaları, müşteri portföyü, kamu kurumlarının vatandaşlara dair çok önemli verileri, hatta askeri kurumların gizlilik dereceli bilgilerinin ifşa edilebildiğini, zaman zaman medyaya bile yansıdığını görebiliyoruz.

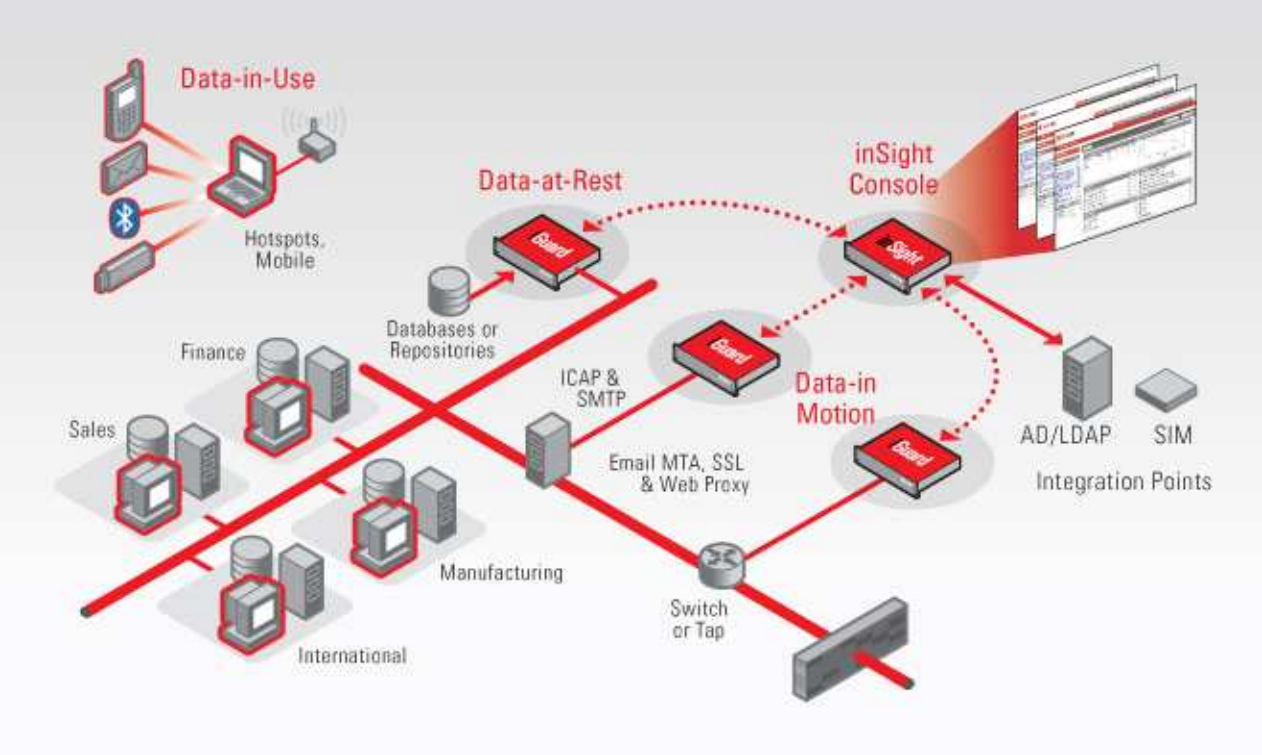
ABD'de 2006 yılının veri kayıp ve kaçak envanteri incelendiğinde en karlı 1000 şirketin(Fortune-1000) yaklaşık %75'inin söz konusu istenmeyen durumlara maruz kaldığını görürüz.Yine aynı listedeki şirketlerin bilgi güvenliğinde sancılı olduğu alanlar incelendiğinde veri kaçağının %18 ile yüksek düzeylerde olduğu görülür.



Şeki.1 Fortune 1000 şirketlerinde bilgi güvenliği zafiyetlerinin dağılımı

Bilgi güvenliğinde Veri Kaçağı (Data Loss) olarak nitelendirilen bu olayın önlenmesi için birbirinden çok farklı özellikler barındıran bir dizi ticari ürün mevcuttur. Bu çözümlerin geneline verilen isim de Veri Kaçağı Önleme (Data Loss Prevention - DLP) çözümleri olmuştur. Çok temel sayılabilecek yöntemlerle bu korumayı sağlamaya çalışan ticari çözümlerin zamanla geliştirilebileceği ve çok daha akıllı sistemlere erişileceği umulmakta. Yazımızda belirgin bir ticari ürünü tanıtmaktan çok bu gibi ürünlerin sağladığı genel güvenlik çözümlerinden bahsedilecektir. Mevcut çözümlerin çalışma prensipleri ve örnek koruma senaryoları aktarılacak, bu çözümlerin yeteneğini artırabilecek veri madenciliği ve iş zekası uygulamalarının muhtemel katkılarını tartışılacaktır.

Veri Kaçağı Önleme (DLP), kullanımda olan(son kullanıcı aksiyonlarındaki), hareketli olan(ağ üzerindeki) ve geri kalan(veri ambarı/tabanındaki) verinin kritiklik derecelerinin saptanması, monitör edilmesi ve korunmasını sağlayan politika, yazılım ve donanımdan oluşan bir çözüm platformudur.



Şekil.2 Genel kabul görmüş DLP mimarisi

Böyle bir çözümün şekillenmesi ve hatta ticari bir ürün haline gelmesinde rol oynayan motivasyonların başında kurumların özel/gizli bilgilerinin kaçağından ötürü yaşadıkları zararlar gelir. Bunun yanında sektörel düzenleyici örgütlerin yaptırımları da şirketleri veri güvenliğini sağlama konusunda mecbur bırakmıştır. Sağlık sektöründe HIPAA, finansta GLBA, BASEL II ve PCI(Payment Card Industry) Standards gibi regülasyon yaptırımları bilgi güvenliği ve dolayısıyla DLP konusunda ciddi adımların atılmasında önemli rol oynar. Hatta bunların tümünü geride bırakabilecek kadar kritik olan, Wall Street borsasında kağıdı işlem gören her kurumun çok sıkı denetimlere maruz kaldığı Sarbanes-Oxley yasaları DLP konusunda en önemli yaptırımdır diyebiliriz.

DLP çözümlerinin çalışma mantığını incelemeden önce veri koruma gereksinimlerini inceleyelim:

Kullanımdaki verinin kritiklik derecelerinin belirlenmesi ve korunması:

Bu aşamada elde bulundurulan veriler çeşitli mantık örgüleri, o kuruma haiz gizlilik kriterleri, iş sahibi düşünceleri gibi belirleyiciler yardımıyla sınıflandırılır ve ölçek ölçek ayrılan bu sınıfların ne aşamada korunması gerektiği belirlenir. Örneğin aynı rakam bilgisi bir yerde çok önemsiz bir sıra numarası olabilirken diğer yanda bir finansal bilgi ya da kişilerin sosyal güvenlik numarası olabilir. Bunları belirleyecek, çeşitli örüntüleri yakalayacak uygulamalar yardımıyla sınıflandırma yapılır.

Veri kaynaklarında saklanan verinin kritiklik derecelerinin belirlenmesi ve korunması:

Göz önünde bulunan veri sınıflandırıldığı gibi dosya sistemi, veri ambarı, dizüstü makineler gibi kalan veri kaynaklarının üzerindeki veri de sınıflandırılarak koruma ekosistemine dahil edilmelidir.

Ağ üzerinde akan verinin tespiti, sınıflandırılması ve korunması:

Aynı şekilde ağ üzerinde akan veri de kontrol edilebilmelidir. Sadece belirli portların/uygulamaların (SMTP-25 gibi) kontrolü değil, tüm ağ akışının izlenmesi sağlanmalıdır.

Sınıflandırmaların ve güvenlik politikalarının periyodik yenilenmesi:

Değişen şartlar ve güvenlik gereksinimleri DLP'de uygulanması gereken politikaların zaman zaman güncellenmesini gerektirir.

Kontrol ve yönetim araçlarının yeterliliği:

DLP politikalarının rahatlıkla implemente edilebileceği bir kullanıcı dostu ortam oluşturulabilmelidir.

Veri kaçağı olaylarının anlık takibi:

Belki de en önemlisi DLP politikalarını ihlal eden bir hareket görüldüğünde, olayın yönetim biriminde anlık olarak görüntülenebilmesidir.

Kurum bazı bilinçlendirme:

Tüm kurum çalışanlarının ve sorumlularının belirlenen DLP politikaları hakkında gerekli eğitim ve bilgilendirme(broşür, el kitabı vb) sahibi olması sağlanmalıdır.

Bu gereksinimleri karşılaması planlanan DLP sistemleri esasta 2 farklı prensibe göre çalışmaktadır:

Ağ Tabanlı DLP Sistemleri:

Ağ geçidi (gateway) tabanlı sistemler olarak da anılan bu sistemler, ağın kritik noktalarına konuşlandırılır. Hazırlanan politikalar uyarınca gerçek zamanlı kontroller uygulayan sistem politika ile çelişen bir veri akışı saptadığı zaman alarm üretmekte ve engelleyebilmektedir. Kolayca kurulabilmeleri, bakım maliyetlerinin düşük olması bu sistemlerin tercih nedenlerinin başında gelir. Akan verinin kontrolü imzalar yardımı ile olur. Veri kaynaklarında durağan olan veri sistem tarafından kritiklik sınıflandırmasından sonra belirli bir etiket ile etiketlenir ya da veriye göre imzası çıkarılır.(dijital imza ile karıştırılmamalıdır) Eldeki etiket ya da imzanın ağda aktığını tespit eden sistem aksiyon alabilmektedir.

Sunucu Tabanlı DLP Sistemleri:

Son kullanıcıya hizmet veren iş istasyonları ya da sunuculara kurulan bu tip sistemler ağ tabanlı sistemlerin yeteneklerinin yanında kurum içi grupların birbiri ile iletişimini de denetleyebilmektedir. Bunun yanında getirdikleri bir diğer açılım sadece ağdaki akışı değil daha faklı şekillerde veri kaçırma ihtimallerini de(usb, taslak olarak bekleyen e-postalar vb.) engelleyebilmeleridir.

DLP YARDIMIYLA ÖNLENEBİLECEK BİRKAÇ BİLGİ KAÇAĞI SENARYOSU

Her şirkette ya da kurumda olabileceği gibi sizin kurumunuzda da kötü niyetli çalışanların olması muhtemeldir. Bir kötü niyetli çalışanın kuruma dair gizli bir bilgiyi dışarı çıkarmayı planladığını varsayalım. Öncelikle mevcut DLP çözümlerinde bu gibi gizli dokümanların sistem tarafından etiketlenmiş ve bir politika gurubuna sokulmuş olması gerekmektedir.

Kötü niyetli çalışanımız bu gizli belgeyi e-posta yoluyla şirket dışına çıkarmak istesin. E-posta gönderme işlemini yapmaya çalıştığı anda ağ geçidine takılan mesaj düşürülecek, kişiye de uyarı mesajı verilecektir.

Bunun üzerine ajanımız e-postanın şirket denetiminde olabileceğini düşünüp farklı yollarla belgeyi dışarı çıkarmayı deneyebilecektir. MSN, Skype, ya da diğer portlarda çalışan bir çok protokol DLP çözümümüze takılacak ve dosyanın çıkışı engellenecektir.

Bu şekilde dokümanı ağ üzerinden çıkaramayacağını anlayan ajanımız dokümanı sistem tarafından tanınmayan, güvensiz bir USB belleğe almak istediğinde masaüstünde yer alacak bir uyarı ile uyarılacak ve dosyayı kopyalaması engellenecektir.

Bu sorunla da yüzleşen kötü niyetli şahsımız bu sefer de dokümanın içerisindeki gizli kısmı başka bir word belgesine aktarıp bu yeni belgeyi USB'ye almayı deneyecektir. Fakat burada da sistem kendisini uyaracak ve işlemi engelleyecektir. Çünkü DLP çözümümüz belgeyi etiketlediği gibi belgenin içeriğini de etiketler.

Bir diğer çalışanımız yazılım şirketimizde yazılan kaynak kodların çok kritik bir kısmını daha ürün çıkmadan rakip firmaya ulaştırmak istesin. (Tabii ki çalışanlarımıza güveniyoruz ve bu sistemleri çalışanlarımızın huzurunu bozmak için kurmuyoruz. Fakat çok önemli bilgilerimizi bu yolla rakiplere kaptırdığımız da bir gerçek.) Kaynak kodların ufak bir parçası bile bizim belirlediğimiz politikalarla çakıştığı sürece sistem tarafından tespit edilir ve önlenir.

Bunlara ilaveten önceden belirlediğimiz örüntülere uyan içerikler de iletişim sırasında ya da harici aygıtlara aktarımda bloklanabilir. Örneğin TCkimlik numaraları 11 haneli sayısal ve son hanesi çifttir. Benzeri şekilde kredi kartı numaraları 16 haneli sayısal örüntülerdir. Buna uyan sayılardan oluşan bir listenin aktarımı tespit edilirse içerik kolaylıkla bloklanabilir.

Artık kötü niyetli bir kullanıcı için tek çıkar yol şirket ağının dışında iken(örneğin dizüstü bilgisayarında dışarıda bir ağdan internete ulaşmışken) verileri kaçırmak istemesidir. DLP sistemleri bu gibi senaryolarda da başarı sağlar. Kurum bünyesinde envantere girmiş tüm bilgisayarlara ajan yazılım kurulabilir ve bu sayede makinadan e-posta vb aktarımlar, usb gibi cihazlara aktarımlar dış ağlarda iken de izlenebilmiş olur.

Tüm bu senaryoları engelleyebilen DLP yazılımları belirli yönleriyle zayıf kalmaktadır. Halihazırdaki çözümler ya öntanımlı kelimeleri tespit eder, engeller; ya öntanımlı örüntüleri (kredi kartı numarası gibi) tespit eder engeller; ya da etiketlenen bir dosyanın tümü ya da bir parçasının kaçırılmasını engeller. Fakat eğer ki bilgi kaçırmak isteyen kişi örneğin kendi ifadelerini kullanarak bir e-posta hazırlar ve dışarı çıkarırsa, sistem bunu kesinlikle tespit edemez. Buna benzer, kullanıcı tarafından tanımlı anahtar kelimeleri kullanmadan dışarı sızdırılan her bilginin kaçağı ne yazık ki önlenememektedir. Bu gibi kaçaklar ancak ve ancak yapay öğrenme özelliğine sahip sistemlerce engellenebilir. Bu noktada veri/metin madenciliği tekniklerinin makine öğrenmesi özellikleri devreye girer.

VERİ MADENCİLİĞİ YÖNTEMLERİNİN DLP'DE SAĞLAYABİLECEĞİ AÇILIMLAR

Veri madenciliği, veritabanı gibi yapısal verilerde olduğu gibi, metin, görüntü, ses gibi yapısal olmayan verilerin de içindeki saklı örüntüleri bulmakta kullanılmaktadır. Bu örüntüler içerik(content) anlamında belirli anahtar "ifade"ler belirleyebilir. İşte bu ifadeler üzerinden içerik saptanması ve içeriğe göre gizli bilginin dışarı çıkarılması engellenmiş olacaktır.

Halihazırda ürünleştirilmiş DLP sistemleri de "ifade"ler üzerinden içerik taraması yapmaktadır. Fakat bu yazılımların ifade olarak anlayacağı kurallar kullanıcı tarafından sisteme önceden tanıtılmak zorundadır. İşte tam bu noktada metin madenciliğinin başlık ya da anahtar olgu belirleme(topic detection, indexing) özelliklerinden yararlanmanın mantıklı olacağı ortadadır. Sisteme gizli olarak tanıtılan dokümanların içeriği akıllı bir metin madenciliği sistemi ile taranır ve önemli olgular/başlıklar otomatik olarak saptanabilirse, kullanıcının kural girme zorunluluğu ortadan kaldırılmış olacaktır. Bu hem uygulamadaki başarıyı artıracak hem de uygulamayı kullanıcı için çok daha kolay hale getirecektir.

Örneğin mevcut sistemlerde gizlilik belirtisi olan ifadeler kullanıcı tarafından belirlenecek ve 25 değişik "ifade" gizliliği belirleyecek ifadeler olarak sisteme girilecektir. Ancak çok büyük organizasyonlarda kişilerin tecrübeye dayalı olarak bu ifadeleri seçmesi ve belirlemesi mümkün olmayabilir. Bazı belirleyici ifadeler sisteme girilmemiş olabilir ya da ifade seti güncellenmemiş olabilir. İşte bu gibi sıkıntıların yaşanmaması için sistem tarafından "ifade"lerin otomatik çıkarsanması metin madenciliği yöntemleri ile oldukça basittir. Anahtar ifade sayısı belki yine 25'te kalacak fakat doküman gizliliğini belirlemede çok daha önemli "ifade"lerin seçilmesi mümkün olacaktır.