AĞ TABANLI VERİ SIZINTISI TESPİTİ VE ÖNLENMESİ ÜZERİNE BİR İNCELEME

***

A REVIEW ON NETWORK BASED DATA LEAKAGE DETECTION AND PREVENTION

Cengiz PAŞAOĞLU[*]
Habibe GÜLER[**]
Masoma JAFARI[***]

DOI: 10.33461/uybisbbd.611768

Öz

Bilgi güvenliği sistemlerinin temel amacı yetkisiz kişilerce gerçekleştirilen veri ihlallerine karşı önlem almaktır. Bu sebeple, kullanımda, hareket halinde veya durağan durumda olan gizli/hassas verilerin sızıntılarının tespiti ve önlenmesinde daha etkili çözümler sunabilen veri sızıntısı önleme sistemleri (DLPS-Data Leakage Prevention System) geliştirilmiştir. Bu çalışma kapsamında, özellikle ağ tabanlı veri sızıntısı tespit (DLP-Data Leakage Prevention) sistemleri üzerinde durularak veri sızıntısı tespitinde kullanılan bağlam tabanlı, içerik tabanlı ve içerik etiketleme yöntemleri detaylı bir şekilde açıklanmıştır. Bunun yanı sıra önleme yöntemleri de incelenmiştir. Son olarak günümüz DLP sistemlerinin yaygın olarak karşılaştığı zorluklardan bahsedilmiştir.

***Anahtar Kelimeler**: Hassas Veri, DLPS, Bağlam Tabanlı Denetim, İçerik Tabanlı Denetim, İçerik Etiketleme, Ağ Tabanlı Veri Sızıntısı Tespiti.*

Abstract

The main purpose of information security systems is to take measures against unauthorized data violations. For this reason, Data Leakage Prevention Systems (DLPS) which can provide more effective solutions in detecting and preventing leakage of confidential data in use, in motion or at rest, have been developed. In this study, context-based, content-based and content tagging methods that are used -especially with network based Data Leakage Prevention (DLP) systems- in data leak detection are explained in detail. In addition, prevention methods are also examined. In conclusion, the challenges encountered by today's DLP systems have been discussed.

***Keywords:** Sensitive Data, DLPS, Context Based Method, Content Based Method, Content Tagging, Network Based Data Leakage Detection.*

[*] Dr., Kişisel Verileri Koruma Kurumu, cengizpasaoglu@kvkk.gov.tr,
ORCID: 0000-0002-4583-5461
[**] Gazi Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, habibeguler@outlook.com,
ORCID: 0000-0003-2607-4801
[***] Gazi Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, masomashams1@gmail.com,
ORCID: 0000-0003-0824-8601

Geliş Tarihi / Received: 27.08.2019
Kabul Tarihi / Accepted: 09.10.2019

1. GİRİŞ

Günümüzde hassas/gizli verilerin ifşası bireyler ve organizasyonlar için giderek büyüyen bir sorun haline gelmeye başlamıştır. Bilgi güvenliği sistemlerinin temel amacı yetkisiz kişilerce gerçekleştirilen veri ihlallerine karşı önlem almaktır. Veri sızıntısı riskine karşı güvenlik duvarları (firewall), sanal özel ağlar (VPN-Visual Private Network), saldırı tespit ve önleme sistemleri (IDS/IPS-Intrusion Detection/Prevention Systems) gibi birçok geleneksel güvenlik çözümü vardır. Ne var ki, bu sistemler veri sızıntılarının tespiti ve önlenmesinde yetersiz kalmaktadır. Hassas/gizli veriler farklı sızıntı kanallarında farklı formlarda olabileceği için bu sistemlerin yetersizliği sebebiyle gerçekleşecek veri sızıntıları hem bireyler hem de organizasyonlar açısından çok ciddi sonuçlara yol açabilir. Bu sebeple, kullanımda, hareket halinde veya durağan durumda olan verilerin sızıntılarının tespiti ve önlenmesinde daha etkili çözümler sunabilen veri sızıntısı önleme sistemleri geliştirilmiştir.

Bu çalışma kapsamında özellikle ağ tabanlı veri sızıntısı tespit sistemleri tarafından kullanılan yöntemler üzerine kapsamlı bir inceleme yapılmıştır. Bu bağlamda makale yedi ana başlık altında ele alınmıştır. Öncelikle veri sızıntısı problemi ele alınarak konu hakkında genel bilgiler verilmiştir. Daha sonra konu ile ilgili literatür taramasına yer verilmiştir. İlerleyen bölümlerde önce veri sızıntısı tespitinde yaygın olarak kullanılan yöntemler üç ana başlık altında irdelenmiş sonrasında ağlarda veri sızıntısı tespiti detaylandırılmıştır. Bunun yanı sıra veri sızıntıları önleme yöntemlerinden de bahsedilmiştir. Son olarak günümüz DLP sistemlerinin yaygın olarak karşılaştığı zorluklardan bahsedilerek sonuç bölümü ile makale sonlandırılmıştır.

2. VERİ SIZINTISI PROBLEMİNE GENEL BAKIŞ

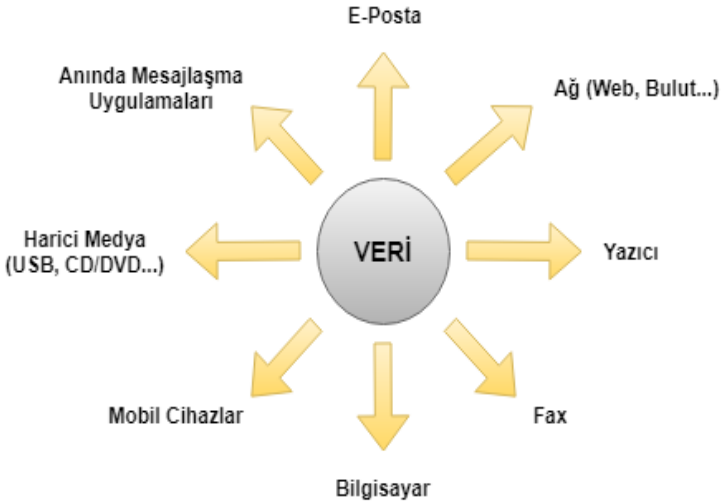
Bilgi piramidinin temelinde yer alan veri kavramı, işlenip anlamlandırılmamış en küçük bilgi (information) parçasına verilen isimdir. Henüz işlenmemiş veriler tek başlarına bir anlam ifade etmezler fakat belirli kriterlerde bir araya gelmeleri sonucu önemli hale gelirler.

Kişisel veri, kimliği belirli veya belirlenebilir gerçek kişiye ilişkin her türlü bilgiyi ifade etmektedir (Kişisel Verilerin Korunması Kanunu, 2016). Genel bir ifade ile hassas veri ise sadece yetkili kişi veya kişiler tarafından erişilebilen ve herhangi bir kanuni dayanak olmadan ifşa edilmeye karşı korunan bilgiler olarak tanımlanabilir. Hassas/özel nitelikli kişisel veriler (ırk, etnik köken, din, mezhep, siyasi düşünce, sağlık vb.), başkaları tarafından öğrenildiği takdirde ilgili kişinin mağdur olabilmesine veya ayrımcılığa maruz kalabilmesine neden olabilecek nitelikteki verilerdir (Kişisel Verilerin Korunması Kanunu, 2016). Öte yandan şirketler, kurum ve kuruluşlardaki hassas veriler ise fikri mülkiyet, finansal bilgiler, hasta bilgileri, kişisel kredi kartı verileri, iş ve işletmeye bağlı diğer bilgilerden oluşmaktadır (Shabtai, Elovici, & Rokach, 2012). Kısaca kurumsal veri ise kurumlar tarafından kullanılan, işletilen veya üretilen verilerdir. Bahsi geçen tüm bu kavramlar doğrultusunda veri sızıntısı, hassas/gizli kurumsal veriler ya da hangi nitelikte olduğuna bakılmaksızın kişisel verilerin ihmal/kaza yahut kasti bir şekilde hukuka aykırı olarak, farklı yollardan yetkisiz kişilerin eline geçmesi olarak tanımlanabilir.

2.1. Veri Sızıntısı Kanalları

Farklı kanallar vasıtasıyla birçok sebepten ötürü veri sızıntıları gerçekleşmektedir. INFOWATCH güvenlik firması tarafından hazırlanan 2017 yılı küresel veri sızıntısı raporuna göre (Global Data Leak Report, 2017) veri sızıntılarının en çok %69,8 ile ağ üzerinden gerçekleştiği görülmektedir. Bunu %13,3 ile e-postalar, %8,2 ile yazdırılan dokümanlar takip etmektedir. Yine aynı rapora göre 2017 yılında gerçekleşen veri sızıntılarının %60,5 gibi büyük bir çoğunluğu içeriden kaynaklı, %39,5'i ise dışarıdan kaynaklı sızıntılardır. Veri sızıntısına bilinçli veya bilinçsiz yollarla %50,3 oranında çalışanların sebep olduğu görülmektedir. %41,7 ile bunu harici saldırganlar takip etmektedir. Tüm bunlara ek olarak sızdırılan verilerin %64,8'i kişisel verilerden, %21,1'i

ödeme ayrıntılarından, %8'i ticari sırlar ve teknik bilgilerden, %6,1'inin ise devlet sırlarından oluştuğu görülmektedir. Şekil 1'de veri sızıntılarının gerçekleştiği bazı kanallar görülmektedir.



Şekil.1. Bazı veri sızıntısı kanalları

2.2. Veri Sızıntılarının Etkileri ve Sonuçları

Hassas veriler, stratejik gereksinimler, iş gizliliği, veri gizliliği, hukuki yükümlülükler, kişisel verilerin gizliliği, Sağlık Sigortası Taşınabilirlik ve Sorumluluk Yasası (HIPAA- Health Insurance Portability and Accountability Act), Gramm–Leach–Bliley Yasası (GLBA- Gramm–Leach–Bliley Act), BASEL II (Uluslararası Bankacılık 2. Basel Uzlaşısı), Sarbanes-Oxley Yasası (Sarbanes-Oxley Act), Veri Güvenliği Standardı (DSS- Data Security Standard) gibi sağlık, finans, bankacılık ve birçok alanda kullanılan verilerin güvenliğini sağlamaya yönelik çıkarılan yasalara, yönetmeliklere ve uluslararası bilgi güvenliği standart ve akreditasyonlarına uyum açısından önemli olup sızıntılara karşı korunması gereken verilerdir (Başak, 2016). Söz konusu verilerin kanuni olmayan yollarla başkaları tarafından ele geçirilmesine veri ihlali denilmektedir. Gerek ülkemizde 2016 yılında çıkarılan Kişisel Verilerin Korunması Kanunu (KVKK), gerekse Avrupa Birliği Parlamentosu tarafından 2016 yılında kabul edilip Mayıs 2018 yılında yürürlüğe giren Genel Veri Koruma Tüzüğü (GDPR-General Data Protection Regulation) kapsamında veri ihlalleri hususu özel maddeler ile düzenlenmiştir.

Hem kötücül yazılımlar, davranışlar, tutumlar veya düşüncelerle hem de istenmeyen bir durum ile içeriden veya dışarıdan hassas verilerin saldırıya maruz kalmaları, bir kurumu veya organizasyonu ciddi bir şekilde zarara uğratabilmektedir (Canbay ve Sağıroğlu, 2016). Saldırılar sonucu oluşabilecek veri sızıntılarının dolayısıyla veri ihlallerinin kurum veya kuruluşlara hem maddi olarak hem de hukuki olarak ciddi olumsuz etkileri olabilmektedir. Sonuç olarak kurumlar itibarlarını ve hatta müşterilerini kaybetmekte bunun sonucunda hisse senetleri düşmekte ve dolaylı olarak daha çok maddi zarara uğramaktadırlar. Bu noktada veri sızıntısı probleminin günümüzde gerek ülkemizdeki gerekse dünyadaki tüm kurum ve kuruluşlar için çok büyük ve en önemli problemlerden biri olduğu açıkça görülmektedir. Ponemon Enstitüsü'nün IBM'in de sponsorluğunda yaptığı çalışma sonucu yayınladığı 2018 Veri İhlalinin Küresel Maliyeti raporuna göre (Cost of a Data Breach Study: Global Overview, 2018) bir veri ihlalinin küresel ortalama toplam maliyetinin 3.86$ milyon dolar olarak hesaplandığı görülmektedir. Ülkeler bazında incelendiğinde ise ABD'nin ortalama 7.91$ milyon dolar ile en yüksek paya sahip olduğu, bunu 5.31$ ve 4.74$ milyon dolar ile sırasıyla Orta Doğu ülkeleri ve Kanada'nın takip ettiği görülmektedir. Söz konusu rapora göre Türkiye 2.16$ milyon dolar ortalama maliyet ile 12. sırada yer almaktadır. Bu açıdan bakıldığında veri ihlalleri sonucu ülkemizin mali kaybının ortalamadan

kısmen daha az olduğu söylenebilir, ancak ülkemiz anılan raporda da vurgulandığı üzere en çok kaybı olan ilk 15 ülke içinde yer almaktadır.

3. LİTERATÜR TARAMASI

Bu bölümde makale kapsamında incelenen ve makalenin hazırlanmasında yol gösterici olan bazı çalışmalar ele alınacaktır. Bu bağlamda ağ tabanlı veri sızıntılarının tespiti ve önlenmesine yönelik çalışma yapan ve yeni yaklaşımlar öneren makaleler incelenmiştir.

İncelenen söz konusu makalelerin birçoğunda veri sızıntısı tespitine yönelik yöntemlerin ele alındığı görülmektedir. Bunlardan (Trieu, Tran, Tran, & Tran, 2017) ve (Alneyadi, Sithirasenan, & Muthukkumarasamy, 2015) de yapılan çalışmalarda hem tespit hem de önleme/engelleme yöntemlerinin beraber kullanıldığı görülmüştür.

Yine bu makalelerde ağdaki paketleri analiz ederken içerik tabanlı analiz tekniği olan istatistiksel analiz tekniğinin yaygın olarak kullanıldığı görülmüştür. (Shu, Zhang, Yao, & Feng, 2016) ve (Canbay, Yazıcı, & Sağıroğlu, 2017) yaptıkları çalışmalarında istatistiksel analiz yönteminin yanında anahtar kelime eşleme yöntemini de kullanarak ağ tabanlı veri sızıntılarının tespitini yapmayı hedeflemektedirler. (Huang, Lu, Li, & Ma, 2018; Katz, Elovici, & Shapira, 2014; Soumya & Smitha, 26 June, 2014) de incelenen diğer makalelerden farklı olarak, bağlam tabanlı yönteme yeni bir yaklaşım getirerek veri sızıntılarının tespitine yönelik çalışmalar önermişlerdir.

İncelenen çalışmalarda önerilen yöntemlerin performans ölçümleri için farklı birçok ölçütün kullanıldığı görülmüştür. Bunlar; ayrım eşik değerinin farklılık gösterdiği durumlarda doğru pozitiflerin yanlış pozitiflere olan kesri ile hesaplanan alıcı işletim karakteristiği (ROC- Receiver Operating Characteristic), doğruluk (ACC- Accuracy), kesinlik (precision), hassasiyet (recall), tespit oranı (DR- Detection Rate), doğru pozitif, yanlış pozitif ve yanlış negatif oranı (TPR-True Positive Rate, FPR- False Positive Rate, FNR- False Negative Rate) olarak söylenebilir. Ek olarak önerilen bu modellerin verinin hangi durumu için uygun oldukları konusuna da incelenen çalışmalarda değinilmiştir. Bu bağlamda durağan veri (DAR- Data At Rest), kullanımda veri (DIU- Data In Use) ve hareket halinde veri (DIM- Data In Motion) olarak veri durumu sınıflandırılabilir. Sonuç olarak bu çalışmalar tarafından kullanılan yöntem, analiz tekniği, verinin hangi durumu için uygun olduğu ve önerdikleri modelin doğruluğunu test etmek üzere kullandıkları performans metriği Tablo 1' de özetlenerek incelenen makalelerin benzerlikleri ve farklılıkları ortaya konmuştur.

Tablo 1: Literatür İncelemesi

Makale	***Yöntem***	**Analiz Tekniği**	**Veri Durumu**	**Performans Ölçütü**
(Trieu vd., 2017)	Tespit/Önleme	İçerik Tabanlı – İstatistiksel Analiz + Makine Öğrenmesi	DIU, DIM, DAR	DR, FNR
(Alneyadi vd., 2015)	Tespit/Engelleme	İçerik Tabanlı – İstatistiksel Analiz	DIM	Precision, Recall
(Shu vd., 2016)	Tespit	İçerik Tabanlı – Anahtar Kelime Eşleme + İstatistiksel Analiz	DIM, DAR	TP, FP
(Canbay vd., 2017)	Tespit	İçerik Tabanlı – Anahtar Kelime Eşleme + İstatistiksel Analiz	DIM	ACC
(Soumya ve Smitha, 2014)	Tespit	Bağlam Tabanlı	DIM	DR
(Katz vd., 2014)	Tespit	Bağlam Tabanlı (CoBAn)	DIM	TPR, FPR
(Huang vd., 2018)	Tespit	Bağlam Tabanlı (AGW)	-	ACC, ROC, Recall
(Liu vd., 2009)	Tespit	İçerik Tabanlı – İstatistiksel Analiz	DIM	ROC
(Gugelmann, Studerus,	Tespit	İçerik Tabanlı	DIM	-

Lenders, & Ager, 2015)				
(Hemalatha, Somasundaram, & Thirugnanam, 2016)	Tespit	İçerik Tabanlı – İstatistiksel Analiz	DIM	-
(Shu, Yao, & Bertino, 2015)	Tespit	İçerik Tabanlı – Bulanık Parmak izi	DIM	FNR, FPR
(Breitinger ve Baggili, 2014)	Tespit	İçerik Tabanlı- Yaklaşık Eşleşme	DIM	DR
(Shapira, Shapira, & Shabtai, 2013)	Tespit	İçerik Tabanlı – Parmak izi (Geliştirilmiş Versiyon)	-	ROC

4. VERİ SIZINTISI TESPİTİ VE ÖNLENMESİ

Veri sızıntısı tespitine yönelik yaklaşımlar kurum dışına çıkarılmaması gereken hassas verilerin korunmasına yönelik bir takım tespit ve önleme teknikleri kullanırlar. Bu tanıma göre S. Alneyadi ve ark. (Alneyadi, Sithirasenan, & Muthukkumarasamy, 2016) yazdıkları inceleme yazısında veri sızıntısı önleme sistemlerini geleneksel yöntemlerden ayıran üç temel özelliği olduğundan bahsetmişlerdir. Bu özellikler;

- DLP sistemleri hassas veya gizli verilerin ve bunları çevreleyen kaynağın içeriğini analiz edebilmektedir,
- DLP sistemleri hareket halinde, kullanımda veya durağan durumda olan gizli veriler için koruma sağlayabilir,
- DLP sistemleri bildirme, denetleme, engelleme, şifreleme gibi birtakım eylemler yoluyla gizli veriyi koruyabilmektedir.

şeklinde sıralanabilir.

4.1. DLP Tanımı

Güvenlik duvarları (firewall), saldırı tespit sistemleri (IDS-Intrusion Detection Systems), saldırı önleme sistemleri (IPS-Intrusion Prevention Systems) ve sanal özel ağlar (VPN-Virtual Private Network) gibi geleneksel sistemlerin veri sızıntılarının tespiti ve önlenmesi konusunda yetersiz kaldığı görülmüştür (Alneyadi vd., 2016). Bu sorunun üstesinden gelmek ancak veri sızıntısı tespiti ve önlenmesine yönelik yeni bir sistemin olması ile mümkündür. Bu noktada veri sızıntısı önleme çözümleri devreye girmektedir. DLP sistemleri ilk kez 2006 yılında geliştirilmeye başlanmıştır, bu sebeple henüz yeni bir sistem olarak kabul görmektedir.

Shabtai ve arkadaşları (Shabtai vd., 2012) DLP çözümünü "gizli bilgilerin yetkisiz erişimi, kullanılması veya iletilmesini önlemek için tasarlanmış bir sistem" olarak adlandırmışlardır. Başka bir çalışmada (Mogull ve Securosis, 2007) ise DLP çözümleri, beklemede, kullanımda veya hareket halindeki verileri derin içerik analizi vasıtasıyla tanımlayan, izleyen ve koruyan merkezi politikalara dayanan ürünler olarak tanımlanmaktadır. DLP çözümleri, hem yetkili olmayan bir kullanıcının gizli verilere erişmesini tespit etme ve önlemede hem de gizli verilerin yanlışlıkla paylaşılmasını engellemek için kullanılmaktadır (Tahboub ve Saleh, 2014) .

4.2. Veri Sızıntısı Tespit Yöntemleri

Veri sızıntısı tespitinde kullanılan yöntemler, incelenen makalelerden (Alneyadi vd., 2016; Mogull ve Securosis, 2007; Shabtai vd., 2012) yola çıkılarak üç ana başlıkta incelenebilir. Bunlardan en yaygın olarak karşımıza çıkan yöntemler ise içerik tabanlı denetim yapan veri sızıntısı

tespit yöntemleridir. Bu yöntemler aynı zamanda ağ tabanlı veri sızıntısı tespitinde derin paket analizinde kullanılmaktadır.

Veri sızıntılarını DLP sistemleri yardımıyla doğru bir şekilde tespit edebilmek için öncelikle verinin durumunu ve verinin hassas nitelikli/gizli veri olup olmadığını belirlemek gereklidir. DLP sistemleri farklı yaklaşımlar kullanarak verinin 3 durumu için de çözüm sunmayı amaçlamaktadır. Bu durumlar şu şekilde tanımlanabilir;

- Durağan veri (Data-at-Rest), diğer bir deyişle kaynakta duran veriler, veri tabanlarında, dosya ve ağ sunucularında, bulut depolama alanlarında, dosya yönetim sistemlerinde, sabit diskler veya hafıza kartları gibi veri depolama cihazlarında depolanan henüz kullanılmayan ve hareket halinde olmayan verilerdir. Bu tip verilerin yetkisiz kişilerce erişimi, çalınması veya değiştirilmesini engellemek için genellikle veri şifreleme veya erişim kontrolü gibi güvenlik önlemleri alınır (Shabtai vd., 2012). Bu önlemlerin alınabilmesi için öncelikle içeriklerin nerede tutulduğunun bilinmesi gerekmektedir. Bu noktada depolanan hassas içeriklerin tespit edilmesinde DLP sistemlerinin içerik keşif özelliği kullanılmaktadır (Securosis, 2010; Shabtai vd., 2012).
- Kullanımda olan veri (Data-in-Use), bu tip veriler herhangi bir kullanıcı veya yazılım tarafından etkileşimde olunan verilerdir. Word, Excel gibi ofis dosyaları, veri tabanı erişimleri, bulut uygulamaları veya mobil uygulamalar tarafından kullanılan veriler bunlara örnek olarak verilebilir. Uç nokta ile ilgili DLP sistemleri bu tip verileri korumak ve kullanıcı bu verilerle etkileşime girdiğinde verileri izlemek için kullanılırlar (Securosis, 2010; Shabtai vd., 2012). Kullanımdaki verileri korumaya yönelik geliştirilen DLP araçları kopyala-yapıştır, ekran görüntüsü alma, USB, CD-DVD veya akıllı telefona aktarma, yazdırma ve fakslama gibi faaliyetleri izleyerek hassas verilerin kurum dışına çıkarılmasını engellemeye çalışmaktadır (Shabtai vd., 2012).
- Hareket halindeki veri (Data-in-Motion), kurum içinde bir yerden bir yere giden veya internet yoluyla kurum dışına giden verilerdir. Hareket halindeki veriler için geliştirilen DLP çözümleri iletişim kanalları vasıtasıyla ağ üzerinden bilinen veya bilinmeyen protokoller kullanılarak gönderilen verilerin organizasyon tarafından önceden belirlenmiş politikalar doğrultusunda incelenerek hassas veri olup olmadığının tespiti için kullanılır (Al-Sanabani, 2016; Shabtai vd., 2012).

Bu kapsamdaki denetim yaklaşımları sonraki kısımlarda ele alınmıştır.

4.2.1. Bağlam Tabanlı Denetim

Bu yaklaşımda sistem temel olarak denetimi yapılacak dosyanın alıcı/gönderici, zaman, boyut, format ve başlık bilgileri gibi meta verileri üzerinde inceleme yapar. Bu meta verilere ek olarak günümüz sistemleri tarafından bağlam-tabanlı denetim kapsamında dosya sahibi ve izinleri, kullanılan şifreleme formatı veya ağ protokolleri, web tabanlı e-posta ve sosyal ağ siteleri gibi özel servisler, site adresleri, masaüstü uygulamaları gibi faktörlerde analiz edilmektedir (Mogull ve Securosis, 2007).

4.2.2. İçerik Tabanlı Denetim

Tipik bir içerik tabanlı DLP sistemi düzenli ifadeler, veri parmak izi, istatistiksel analiz gibi yöntemleri kullanarak veri havuzunda veya hareket halinde olan hassas verileri izleyerek çalışır (Alneyadi vd., 2016). İçerik tabanlı analiz yapılırken veriyi çevreleyen bağlamdan ziyade verinin kendisi üzerinde denetim amaçlanır, bu sebeple içerik tabanlı denetim yaklaşımı bağlam tabanlı denetim yaklaşımına göre daha yaygın ve tercih edilen bir yöntemdir (Securosis, 2010).

Aynı zamanda derin içerik analiz teknikleri olarak da kabul edilen içerik tabanlı denetim tekniklerinden yaygın olarak kullanılanları (Kaur, Gupta, & Singh, 2017; Mogull ve Securosis, 2007; Securosis, 2010) aşağıda olduğu gibi altı başlıkta incelenebilir.

4.2.2.1. Parmak izi oluşturma

En bilinen içerik tabanlı denetim yaklaşımlarından biri olan ve parmak izi çıkarma olarak bilinen metinsel bir özellik oluşturulan bu yöntemde, hassas nitelikli veri içerdiği bilinen dosya veya veri tabanı girdilerinin genellikle özet (hash) fonksiyonları yardımıyla parmak izi çıkarılır, bu bilgiler bir veri tabanında veya incelenecek olan makinede yerel olarak saklanabilir, incelenecek olan içeriklerinde aynı şekilde parmak izi çıkarılır ve daha önceden saklanmış parmak izi bilgileriyle karşılaştırılarak veri sızıntısı olup olmadığını tespit etmek için tam eşleşme aranır (Al-Sanabani, 2016; Shabtai vd., 2012; Shapira vd., 2013).

Y. Shapira ve ark. (Shapira vd., 2013) tarafından yapılan çalışmada parmak izi yönteminin kullanılan özet fonksiyonlarına göre 4 farklı tipte olduğu görülmektedir. Bunlar; klasik, bölgeye duyarlı karma (LSH-Locality Sensitive Hashing) tabanlı, toplama istatistikleri tabanlı, ankraj tabanlı parmak izi oluşturmadır.

Parmak izi oluşturma yöntemi normal koşullarda düşük oranda yanlış alarm verse de bazı kısıtları olabilmektedir. Örneğin, parmak izi çıkarılmış bir dosyada yapılan herhangi küçük bir değişiklik bu dosyadan üretilen özet değerinin farklı olmasına yol açacaktır bunun sonucunda gizli içerikteki bazı karakterler değiştirilerek bu yöntem bypass edilebilir (Shapira vd., 2013). İkinci olarak ise, parmak izi çıkarma işlemi sırasında genellikle bir dokümanın içeriği bütün olarak ele alınır ve bu yaklaşım bazı durumlarda standart içeriğe sahip fakat gizli olmayan dokümanlar için yanlış alarma sebep olabilir. Bu iki problemin üstesinden gelmek için k-atlama-n-gram (Shapira vd., 2013) ve bulanık parmak izi oluşturma (Shu vd., 2015) gibi genişletilmiş teknikler üzerine çalışmaların yapıldığı görülmektedir.

4.2.2.2. Çoklu anahtar kelime eşleme

Bu yaklaşımda, hassas nitelikli/gizli bir dokümanda bulunan karakter katarlarından seçilerek bir hassas kelime listesi oluşturulur. Bu listeler oluşturulurken makine öğrenmesi algoritmaları (V. Gupta, 2013) veya Terim Frekansı – Ters Metin Frekansı (TF/IDF- Term Frequency/Inverse Document Frequency) (Canbay vd., 2017) gibi yöntemlerin kullanıldığı görülmüştür. Gizli dokümanlardan elde edilen bu anahtar kelimeler bir araya toplanarak tek bir veri tabanında saklanmak suretiyle büyük bir sözlük oluşturulur (V. Gupta, 2013). Daha sonra, veri sızıntısı olup olmadığını tespit edebilmek için ağda hareket eden paketlerin analizi sırasında sözlükteki kelimelerin Naive String Match, Knuth- Morris- Pratt, Boyer- Moore (Canbay vd., 2017), Boyer-Moore-Horspool, Boyer-Moore-Horspool-Raita, Rabin Karp, Aho-Corasick, Sun-Manber gibi tekli veya çoklu desen eşleştirme algoritmaları ile paket içerisinde bulunma durumuna bakılmaktadır (Ren, 2013).

4.2.2.3. Tam dosya eşleme

Bu yöntemde video, resim gibi medya dosyaları veya proje ve çizim gibi özel formattaki dosyaların özetleri (hash) çıkarılır, karşılaştırma sırasında ise üretilen özet değeriyle tam eşleşen bir değerin olması beklenir. Bu yaklaşımda bir dosyadaki hassas verilerin kelimesi kelimesine analizinden ziyade dosyanın bir bütün olarak özeti çıkarılır bu açıdan parmak izi oluşturma yaklaşımından farklıdır (Mogull ve Securosis, 2007; Securosis, 2010). Bu yaklaşım her dosya tipine uygulanabilir fakat içerikte yapılan ufak bir değişiklik tüm dosyanın özet değerini değiştireceğinden yeterince etkili bir çözüm sunamaz.

4.2.2.4. Kısmi doküman eşleme

Bu yaklaşımda, korunan içeriğin tam veya kısmi eşleşmesi denetlenir (Pesen, 2015; Securosis, 2010). Önceden belirlenen politikalar doğrultusunda hassas veya gizli içeriğe sahip dokümanın ya tüm içeriği ya da cümle bazlı olarak özeti alınır. Bu yöntemle içeriğin tamamının veya bir kısmının anlık mesajlaşma, forum siteleri, sosyal ağlardaki formlardan herhangi birine kopyalanıp yapıştırılması tespit edilebilmektedir. Bunu yaparken kullanılan en yaygın özet çıkarma şekli

dairesel özetleme metodudur, bu yöntemde önceden belirlenmiş sayıda karakter seçilip özeti alınır ve bu şekilde devam ederek tüm belgenin özeti alınmış olur (Securosis, 2010). Dışarı giden içeriğe de aynı yöntem uygulanarak bulunan özet değerleri karşılaştırılır. Yapılandırılmamış hassas veriler üzerinde, kaynak kodlarda koruma sağlayabilen bir yaklaşımdır. Bu yöntem fazla miktarda içerik olması durumunda düşük performans göstermektedir ayrıca ortak sözcükler ve standart kalıplar bu yöntemin yanlış alarmları üretmesini tetiklemektedir (Pesen, 2015).

4.2.2.5. Düzenli ifade eşleme

Düzenli ifadeler 1951 yılında Kleene (Kleene, 1951) tarafından tanıtılmıştır ve veri sızıntısı önleme sistemlerinde en yaygın kullanılan analiz tekniklerinden bir diğeridir (Alneyadi vd., 2016). Sadece endüstriyel DLP sistemlerinde değil DLP özelliği gösteren tüm araçlarda kullanılan bir tekniktir ve hassas/gizli içeriği belli kurallar dahilinde analiz eder (Securosis, 2010). Literatürde desen eşleştirme veya kural tabanlı eşleştirme (Al-Sanabani, 2016; Alneyadi vd., 2015; K. Gupta ve Kush, 2017; Securosis, 2010; Shabtai vd., 2012) olarak da adlandırılan bu yöntem ile sosyal güvenlik numarası, kredi kartı numarası, TC kimlik numarası, vergi numarası, banka hesap numarası gibi hassas/gizli kurumsal veya kişisel kayıtların tam veya kısmi tespitinde kullanılır (Securosis, 2010). Düzenli ifadelerle analiz yapılırken hassas veya gizli nitelikli ifadelerin tespiti için birtakım desenler kullanılır bu desenler; terim, gerçek anlam ifade eden karakterler veya özel anlam ifade eden (. | * $? +) gibi meta karakterlerden oluşur. Bu yaklaşım kolayca tanımlanmış yapılandırılmış veri parçalarının tespitinde oldukça etkili olmasına rağmen yapılandırılmamış verilerin tespitinde yetersizdir ve yanlış alarm üretir (Hauer, 2015; Securosis, 2010). Bu yöntemin bir diğer özelliği ise bir doküman içeriğinin anlamsal ifadesinden çok şekilsel gösterimi üzerine analiz yapmasıdır. Bu bazı durumlarda sızıntı tespiti için yeterli iken bazı durumlarda yetersiz bir çözümdür.

4.2.2.6. İstatiksel analiz

Bu yöntem bir içeriğin yapısını analiz etmek ve korunan içeriğe benzeyen içeriklerde politikaları ihlal eden kısımları bulmak için makine öğrenmesi, Bayes analizi ve diğer istatistiki yöntemleri kullanmaktadır (Pesen, 2015; Securosis, 2010). İstatistiksel analiz sırasında kullanılan bu teknikler karışık veya belirsiz tipteki veriler üzerinde de etkilidir. Ayrıca bu yöntem kısmi dokuman eşleme yönteminden farklı olarak yapılandırılmamış içeriklerin tespitinde de oldukça başarılıdır. N- gram ve bir dokümanda geçen kelimelerin önemini belirten terim ağırlığı bulma teknikleri (örneğin TF-IDF) temel istatistiksel analiz teknikleridir (Alneyadi vd., 2016). N-gram analiz yöntemi makine öğrenmesi algoritmaları ile birlikte dokümanların hassas veya hassas olmayan şeklinde sınıflandırılmasında DLP sistemlerinde yaygın olarak kullanılır (Alneyadi vd., 2016). İstatistiksel analiz yöntemlerinin en belirgin avantajı yapılandırılmamış veriler üzerinde içerikleri hassas ve hassas olmayan şeklinde sınıflandırabiliyor olmasıdır. Bunun yanında bu sınıflandırmanın en doğru şekilde yapılabilmesi algoritmanın öğrenmesine bağlıdır bu da algoritmanın büyük miktarlarda veriye ihtiyacı olduğu anlamına gelmektedir bu sebeple bu yöntem yanlış pozitif veya yanlış negatif sonuçlar vermeye eğilimlidir (Securosis, 2010).

4.2.3. İçerik Etiketleme

Bu yaklaşımda, hassas veri içeren bir dosyaya içerdiği verinin çok gizli, gizli, özel gibi gizlilik seviyesine göre bir etiket atanır ve atanan etikete dayalı bir politika uygulanır. İçerik diğer uygulamalar tarafından işlendiğinde bile etiketlenmeye devam edecektir (Al-Sanabani, 2016; Shabtai vd., 2012; Shapira vd., 2013). Etiketler farklı yollar kullanılarak atanabilir örneğin; manuel olarak hassas verinin yaratıcısı tarafından, otomatik olarak içerik veya bağlam tabanlı analiz yöntemleri kullanılarak, otomatik olarak belirli bir yerde tutulan tüm dosyalara ve son olarak özel bir uygulama ya da kullanıcı tarafından oluşturulan dosyaların tümüne uygulanabilir. Geleneksel içerik etiketleme çözümleri, hedef iş istasyonunu veya sunucuyu tarayan, veri dağıtım politikasını

ihlal eden depolanmış verileri algılayan aracılar kullanır ve sonrasında da gerekli işlemleri yapar (Matthee, 2016).

4.3. Veri Sızıntısı Önleme Yöntemleri

Bu yaklaşımlar uygun teknik ve yöntemler kullanılarak veri sızıntısı olayları meydana gelmeden önce önlenmesi için kullanılan proaktif yaklaşımlardır. İncelenen makalelerde (Al-Sanabani, 2016; Alneyadi vd., 2016; Shabtai vd., 2012) önleyici yaklaşımlar için dört temel yöntem olduğu görülmüştür. Bunlar erişim kontrolü, şifreleme, devre dışı bırakma ve farkındalık olarak söylenebilir.

4.3.1. Erişim Kontrolü

Erişim kontrolü, bir kullanıcı tarafından belirli bir kaynağın kullanımına izin verme veya reddetme olarak söylenebilir. Bu bağlamda erişim kontrolü, hassas verilere kimin veya kimlerin erişebileceğine dair kuralların belirlenmesi ve uygulanması hususundaki süreçleri kapsamaktadır. Organizasyon tarafından tanımlanan politikaya göre, bir kullanıcının veya çalışanın hassas bilgilere erişim izni yoksa, DLP bu bilgilerin kullanımını kısıtlar, aksi halde erişim kabul edilir. Erişim kontrolü sağlamanın bir yolu, erişim denetimini belgelere otomatik olarak uygulamak için kurumsal dijital haklar yönetimi (EDRM-Enterprise Digital Rights Management) ile entegrasyondur (Shabtai vd., 2012).

4.3.2. Şifreleme

Durumu ve niteliği göz önüne alınarak hangi hassas verinin şifrelenmesi gerektiği ve şifrelenmiş bu hassas verilerin şifresini çözme hususunda kimlerin yetkili olabileceğine dair ilkelerin belirlendiği bir yaklaşımdır. Bu yaklaşımda DLP sistemleri, hassas verilerin yalnızca onaylı kurumsal uygulamalarla şifrelemeye ve şifresinin çözülmesine izin vererek hassas verilerin güvenliğini garantilemeyi amaçlamaktadır.

4.3.3. Devre Dışı Bırakma

Bu yaklaşım hassas verilerin sızmasına neden olabileceği düşünülen işlevleri devre dışı bırakmayı içeren önleyici bir yaklaşımdır. Örneğin, hassas içeriğe kopyalama ve yapıştırma işlemlerinin kısıtlanması, içeriğin taşınabilir depolama birimine kaydedilmesi veya ekran görüntüsünü alma işlevlerinin devre dışı bırakılması gibi önlemlerdir.

4.3.4. Farkındalık

Farkındalık, kullanıcıları ve çalışanları hangi verinin hassas/gizli nitelikte olduğu, kimin neye erişiminin olduğu ve bunu korumak için ne yapılması gerektiğine dair bilgilendiren bir süreçtir (Al-Sanabani, 2016; Shabtai vd., 2012). Bunun için çalışanlara ve kullanıcılara birtakım eğitimler verilerek onların farkındalıklarını artırmak amaçlanmaktadır çünkü güvenlik sistemlerinde her ne kadar en son teknolojiyi kullanarak önlemler alınsa da en zayıf halka olan insan faktörü asla göz ardı edilmemelidir. Çalışanlar organizasyonun güvenlik politikaları hakkında sürekli olarak bilgilendirilmeli ve bu ilkelere uyulması için gereken çalışmalar yapılmalıdır.

5. AĞ TABANLI VERİ SIZINTISI TESPİTİ VE ÖNLENMESİ

Veri sızıntısı olaylarının büyük bir kısmı ağ üzerinden gerçekleşmektedir. Bu sebeple ağ tabanlı DLP sistemleri geliştirilmiştir. Bu sistemler hareket halindeki veriler için koruma sağlamayı hedeflemektedir. Bunu yaparken DLP merkezi sunucusunda bulunan organizasyon tarafından önceden belirlenmiş politikaları ihlal eden içeriğin olup olmadığına bakılır.

Bilgi güvenliğinin sağlanabilmesi hususunda gerekli gizlilik, bütünlük ve erişilebilirlik kriterlerinin korunması kurumsal ağın çıkış noktasına kurumun ihtiyaçlarına uygun olacak şekilde tasarlanmış bir DLP sisteminin konumlandırılması ile mümkündür. Söz konusu DLP sistemleri organizasyonlarda yürütülen işin ve uygulamanın erişilebilirliğini garanti altına alabilmeli ve bunun için HTTP, FTP, IM, Telnet, TCP/IP, SMTP, POP3, IMAP gibi güncel protokoller üzerinden gönderilen paketlerin yalnızca başlık değil ayrıca derin paket analiziyle paketin payload kısımlarına da bakarak olası veri sızıntılarını tespit edebilmelidir (Oğuz ve Cevahir, 2010). Bu makale kapsamında detaylı bir şekilde incelenen bağlam tabanlı, içerik tabanlı ve içerik etiketleme olarak üç temel başlıkta sınıflandırılan veri sızıntısı tespit yöntemleri bugün endüstriyel amaçlı geliştirilen DLP çözümlerinde derin paket analizi sırasında kullanılan yaklaşımlardır. Email, webmail, HTTP/FTP, HTTPS, IM/Messaging, TCP/IP, P2P (Peer-to-Peer) izlenme ve uyarma gerektiren network kanalları olarak söylenebilir (Oğuz ve Cevahir, 2010). Bu açıdan bakıldığında bir organizasyonun dış ağa çıkmadan önce DLP sistemlerinin konumlandırılması Şekil 2'deki gibidir.

Şekil 2. Kurum dış ağı DLP pasif izleme mimarisi (Securosis, 2010).

Bunun yanı sıra e-posta için organizasyon ağında pasif izleme mimarisinden farklı olarak posta sunucusu ile aktarım temsilcisi (mail transfer agent) arasına DLP sistemi yerleştirilmektedir (Securosis, 2010). Ağlarda veri sızıntısı tespiti için endüstriyel DLP çözümleri pasif izleme dışında bloklama/filtreleme için 3 farklı şekilde kullanılabilmektedir. Bunlar; köprü (bridge), tcp reset ve vekil sunucu (proxy) olarak belirtilebilir. Köprü görevi yapan DLP sistemi iki farklı ağ arasına konumlandırılarak ortada derin içerik analizi yapar. Bu yöntemde DLP merkez sunucusunda önceden belirlenmiş politikayı ihlal eden bir içeriğin tespit edilmesi durumunda oturum sonlandırılarak bağlantı kesilir. İkinci yöntem olarak ise trafik izleme sırasında veri sızıntısı tespit edildiğinde ağa TCP reset paketi enjekte edilerek bağlantı sonlandırılır ve hassas nitelikli verilerin kurum dışına çıkması engellenmiş olur. Üçüncü ve en önemlisi kendisine ait vekil sunucusu bulunan DLP sistemleridir. Bu durumda ağ trafiği önce Proxy tarafından yakalanıp analiz için ilgili DLP ürününe gönderilir burada derin paket analizleri gerçekleştirilir. Bu yöntemle paket analizleri çok daha sağlıklı yapılabilmektedir. Ayrıca bu yaklaşımın en büyük avantajı günümüzde yaygın olarak kullanılmaya başlanan akan trafiğin şifreli iletilmesini sağlayan SSL/TLS gibi protokollerle sarmalanmış şifreli paketlerin analizinin yapılabiliyor olmasıdır. "Reverse SSL/TLS" özelliğine sahip bir DLP vekil sunucusu yakaladığı şifreli paketleri deşifre edebilme yetkisine sahiptir bu sayede şifreli akan trafikte bile veri sızıntısı tespiti yapılabilmektedir (Farrell, 2017; Securosis, 2010).

6. VERİ SIZINTISI TESPİTİ VE ÖNLENMESİNDE KARŞILAŞILAN ZORLUKLAR

Günümüzde bilgi güvenliğinin sağlanması adına farklı sistemler organizasyonlar tarafından kullanılmaktadır. Ne var ki artan teknoloji kullanımı beraberinde birçok güvenlik açığını da getirmektedir. Bunlardan en önemlisi de veri sızıntısı problemidir. Veri sızıntılarının tespitinde ve

engellenmesinde oldukça etkili olmasına rağmen DLP sistemleri, diğer güvenlik mekanizmalarında olduğu gibi farklı birçok zorlukla karşı karşıya kalmaktadır. İncelenen makalelerde (Alneyadi vd., 2016; K. Gupta ve Kush, 2017; Kaur vd., 2017) karşılaşılan zorluklar şifreleme, erişim kontrolü, uyarlama/güncelleme ve insan faktörü olarak dört ana başlık altında toplanabilir.

Şifreleme veri güvenliğini sağlamak için yaygın olarak kullanılan bir yaklaşımdır ve özellikle ağ tabanlı DLP sistemlerinde en temel zorluklardan birisi olarak söylenebilir. Bu sistemler ağ üzerinden akan trafiği çeşitli analiz yöntemleri ile paket içerikleriyle orijinal verileri karşılaştırarak veri sızıntılarını tespit etmeye çalışır fakat güçlü şifreleme algoritmalarıyla şifrelenmiş içeriklerin analiz edilmesi, organizasyon dışına ağ yoluyla çıkarılmaya çalışılan içeriğin hassas olup olmadığının sınıflandırılması oldukça zor olmaktadır (Alneyadi vd., 2016).

Bilgi güvenliği alanında, erişim kontrolü, bir sisteme, fiziksel veya sanal kaynaklara erişimi sınırlandırmanın bir yoludur (Kaur vd., 2017). Bir organizasyondaki kullanıcıların veya çalışanların hangi veriye erişebileceğinin öncelik ve izinlerinin önceden belirlenmesi önemlidir. Bu erişim izinlerinin doğru ve net bir şekilde tanımlanmaması durumunda DLP sistemleri erişilen verinin meşru bir kullanıcı tarafından mı yoksa izinsiz bir kullanıcı tarafından mı erişildiğine karar verememektedir (Alneyadi vd., 2016). Bazı DLP sistemleri erişim izin kontrolünü sağlamak için Microsoft tarafından sağlanan bir dizin hizmeti olan Active Directory sistemleri tarafından sağlanan erişim listelerini kullanırlar (Mogull ve Securosis, 2007). Erişim izin listelerinin güncel olması konusu işten ayrılan veya görevi değişen çalışanların veri sızıntılarına sebep olabilme riski açısından DLP sistemleri için önemli bir konudur. Erişim izinleri kontrolü veri sızıntılarının önlenmesinde önemli bir rol oynamaktadır bu bağlamda DLP sistemleri, verileri kazara veya kasti sızıntılardan korurken erişim izinleri dahilinde çalışmasını sürdürebilmelidir.

Kurumlar tarafından işlenen verilere yenileri eklenmekte bazıları ise artık işlenmemektedir yani veriler sürekli bir değişim içindedir, sabit değildir. Bu veriler içinde hassas nitelikli olanların kurum dışına izinsiz çıkarılmaması için DLP sistemleri ile sızıntılar denetim altına alınmaya çalışılmaktadır. Özellikle düzenli ifadeler, veri imzası veya parmak izi oluşturma gibi yöntemlerle incelenen trafik ile orijinal gizli veriyi karşılaştırarak sızıntı tespit etmeye çalışan DLP sistemlerinde politikaların güncel olması büyük önem teşkil etmektedir. Var olan veriler üzerinde yapılan değişiklikler veya yeni verilerin eklenmesi DLP sistemlerinde önceden belirlenmiş kuralların bu değişimlere uyarlanarak güncel tutulması gerekliliğini ortaya çıkarmaktadır. Örneğin, parmak izi çıkarılmış bir veri tabanında bulunan hassas veriler üzerinde değişiklik yapılmış olsun, bu durumda daha önceden üretilen parmak izi ile veri tabanının son durumdaki parmak izi birbiriyle uyuşmayacaktır. Böyle durumlarda DLP sisteminde belirlenen kuralların değiştirilmemesi, anılan veriler trafik üzerinde akarken onların tespit edilememesi anlamına gelir. Sonuç itibariyle veri sızıntısı tespiti ve önlenmesinde DLP sistemlerinin sürekli güncellenmesi ve yeniliklere uyum sağlayabilmesi karşımıza çıkan zorluklardan bir diğeridir.

Son olarak veri sızıntısı tespiti ve önlenmesinde karşılaşılan zorluklardan en önemli ve kontrol edilmesi en güç olanı insan faktörüdür. Birçok psikolojik ve sosyal faktörlerden etkilendiği için insan davranışlarının tahmin edilmesi her zaman zor olmuştur (Alneyadi vd., 2016). Veri sızıntısı tespiti ve önlenmesinde özellikle verilerin gizlilik düzeyini tanımlama, belirli kullanıcılara erişim hakları atama ve bir DLP sisteminin algılama eşiğini kalibre etme gibi karar verme durumlarında insan eylemi gerekmektedir. Bu noktada insanların doğası gereği objektiflikten uzaklaşıp yanlış kararlar vermesi kaçınılmaz bir gerçektir. Bunun yanı sıra organizasyonlar tarafından belirlenen güvenlik politikaları ne kadar katı olursa olsun, çalışanların bu kurallara tam olarak uyacağı hiçbir zaman garanti değildir. Örneğin, gizli verilere erişim yetkisi olmayan bir çalışan yetkili bir çalışanın erişim bilgilerini izinli veya izinsiz bir şekilde kullanarak kuralları ihlal edebilmektedir ve bunun gibi birçok örnek daha verilebilir. Bu bağlamda, insan faktörü olduğu sürece DLP ve diğer güvenlik sistemlerinde her zaman zorluklar olmaya devam edecektir.

7. TARTIŞMA VE SONUÇ

Bu makale çalışmasında, ağ tabanlı veri sızıntılarının tespiti ve önlenmesinde kullanılan yöntemlerin incelenmesi kapsamında veri sızıntısı tespiti hakkında bir literatür taraması yapılmış, veri sızıntısı tespitinde kullanılan yöntemler araştırılmış ve yapılan çalışmalar gözden geçirilmiştir.

Çalışmamız için gerçekleştirdiğimiz literatür incelemesi sonucunda veri sızıntısı tespiti ve önlenmesi konusunda benzerlik ve farklılıklarının görülebilmesi amacıyla çalışmalar (Trieu vd., 2017; Alneyadi vd., 2015; Shu vd., 2016; Canbay vd., 2017; Soumya ve Smitha, 2014; Katz vd., 2014; Huang vd., 2018; Liu vd., 2009; Gugelmann vd., 2015; Hemalatha vd., 2016; Shu vd., 2015; Breitinger ve Baggili, 2014; Shapira vd., 2013) kullanılan yöntem, analiz tekniği, incelenen verinin durumu ve tasarlanan sistemin performansının test edilmesinde kullanılan ölçütler şeklinde kategorize edilmiştir. Tablo 1'de görüleceği üzere veri sızıntısı tespitinde bağlam tabanlı, içerik tabanlı ve içerik etiketleme analiz tekniklerinin kullanıldığı görülmüştür. İncelenen araştırmalar, verilerin en çok ağ kanalıyla sızdırıldığını göstermektedir, bu sebeple literatürde özellikle hareket halindeki verilerin incelenmesi çalışmaların en belirgin benzerliklerinden biri olarak tespit edilmiştir. Yapılan çalışmalarda özellikle veri sızıntısı tespitine odaklanıldığı ve bunun için içerik tabanlı yöntemlerin yaygın olarak tercih edildiği saptanan benzerliklerden bir diğeridir. Bununla birlikte çalışmalarda veri sızıntısı tespitinde yüksek oranda doğru sonuçlar elde edilebilen ve daha iyi biçimsel ve anlamsal analiz yapılabildiği için içerik tabanlı analiz tekniklerinden biri olan istatistiksel analizin tercih edildiği görülmüştür. Bazı çalışmalarda ise belirsiz veri desenine sahip hassas veriler veya veriler üzerindeki modifikasyon ataklarına karşı istatistiksel analizle beraber anahtar kelime eşleme yönteminin kullanıldığı saptanmıştır. Özellikle kasti veri sızıntısı vakalarında ise tespit yapabilmek amacıyla istatistiksel analizin yanında parmak izi eşleme yönteminin tercih edildiği görülmüştür. Veri sızıntısı tespiti ve önlenmesi konusuna yeni bir yaklaşım sunmayı amaçlayan çalışmaların ise analiz tekniği olarak bağlam tabanlı yöntemi tercih etmesi ise en belirgin farklılık olarak karşımıza çıkmaktadır. Ayrıca bağlam tabanlı analizi kullanan çalışmalarda gizli olmayan bir dokümandaki gizli bilginin tespit edilebilmesi de bu çalışmaların diğer bir amacı olarak saptanmıştır.

Veri sızıntılarının önüne geçmek için kullanılan erişim kontrolü, şifreleme, devre dışı bırakma ve farkındalık gibi yaklaşımlar ise literatürde (Al-Sanabani, 2016; Alneyadi vd., 2016; Shabtai vd., 2012) bahsi geçen önleme yöntemleri olarak görülmektedir. Çalışmamız kapsamında incelenen makalelerde veri sızıntısının hem tespiti hem de önlenmesine odaklanan çalışmaların analiz tekniği olarak yine içerik tabanlı yaklaşımı kullandığı görülmüştür. İncelenen çalışmalarda geliştirilen sistemlerin performanslarının test edilmesi konusunda veri sızıntısı tespitinde kullanılan analiz tekniğine uygun olacak şekilde farklı birçok ölçütün kullanıldığı tespit edilmiştir.

Yaptığımız bu çalışmanın hedefi güçlü bir bilgi güvenliği sisteminde bulunması gereken DLP sistemini her yönüyle ele almaya çalışarak bu alanda yapılacak olan çalışmalara ışık tutabilmektir. Günümüz teknolojisi her ne kadar gelişmiş ve güçlü güvenlik sistemleri için katkı sağlıyor olsa da tek başına DLP sistemlerinin yeterli olamayacağı ve DLP sistemleri için hala var olan bir takım engel ve zorlukların olduğu görülmektedir. Özellikle bilgi güvenliğinde en zayıf halka olan insan faktörüne karşı bu sistemlerin de yetersiz kaldığı ortaya çıkmıştır. Yine de uygulanacak protokollerin ve güvenlik sistemlerinin doğru bir şekilde planlanması özellikle kurumsal bilgi güvenliği konusunda riskleri en aza indirmeyi sağlamaktadır. Bununla birlikte çalışanların veri güvenliği konusunda bilgi sahibi olması, bilinç ve farkındalık seviyelerinin verilecek eğitimlerle artırılması ile planlı ve rastgele denetimlere tabi tutulup bilgilerinin ölçülmesi diğer önemli bir faktördür ve yaptığımız çalışmanın buna katkı sağlayacağı değerlendirilmektedir.

Son olarak yaptığımız bu çalışmada veri sızıntısı tespit sistemlerinde kullanılan analiz tekniklerinin verinin durumuna, tipine göre değişkenlik gösterdiği tespit edilmiştir. Sadece bir tekniği kullanan bir sistemin tüm veri sızıntılarının tespitinde yetersiz kaldığı görülmüştür. Bu sebeple gelecekte bu alanda yapılabilecek çalışmalar; birden fazla tespit yönteminin entegre edildiği

veri sızıntısı tespit sistemlerinin geliştirilmesine odaklanmalıdır. Ayrıca odaklanılması gereken bir diğer konu ise verilerin aynı kalmayıp sürekli değişime uğramasıdır. Bu durum DLP sistemlerinde tanımlanan protokollerin sürekli değiştirilip güncellenmesine sebep olmaktadır. Bu konuda bağlam tabanlı, içerik tabanlı ve içerik etiketleme yaklaşımlarında yapılan biçimsel analize ek olarak verileri anlamsal bazda inceleyecek ve sürekli kendi kendini güncel tutabilecek bir yaklaşımın geliştirilmesine ihtiyaç vardır. Böyle bir yaklaşımın geliştirilmesinde makine öğrenmesi ve derin öğrenme algoritmalarının DLP sistemleriyle beraber kullanılmasının yüksek oranda başarı sağlayabileceği öngörülmektedir. Bu kapsamda yapılacak yenilikçi çalışmalar DLP sistemlerinin güvenlik kapasitelerinin gelişmesinde önemli katkı sağlayacaktır.

KAYNAKÇA

Al-Sanabani, H. (2016). *Eklentiler Kullanarak Veri Kaybını Engelleme.* (Yüksek Lisans Tezi), Sakarya Üniversitesi, YÖK Ulusal Tez Merkezi.

Alneyadi, S., Sithirasenan, E., & Muthukkumarasamy, V. (2015). *Detecting data semantic: a data leakage prevention approach.* Paper presented at the Trustcom/BigDataSE/ISPA, 2015 IEEE.

Alneyadi, S., Sithirasenan, E., & Muthukkumarasamy, V. (2016). A survey on data leakage prevention systems. *Journal of Network and Computer Applications, 62*, 137-152.

Başak, C. D. (2016). Veri Sınıflandırılması ve Hassas Verinin Sızdırılması. Retrieved from https://www.platinbilisim.com.tr/TR/Medya/SiberBulten/siber-bulten-agustos-2016

Breitinger, F., & Baggili, I. (2014). File detection on network traffic using approximate matching.

Canbay, Y., & Sağıroğlu, Ş. (2016). Veri Kaçağı Tespitinde Yeni Bir Yaklaşım. *Savunma Bilimleri Dergisi, 15*(1), 149-177. doi:2148-1776

Canbay, Y., Yazici, H., & Sagiroglu, S. (2017). *A Turkish language based data leakage prevention system.* Paper presented at the Digital Forensic and Security (ISDFS), 2017 5th International Symposium on.

Cost of a Data Breach Study: Global Overview. (2018). Retrieved from https://databreachcalculator.mybluemix.net/assets/2018_Global_Cost_of_a_Data_Breach_Report.pdf

Farrell, C. (2017). *Looking Under the Rock: Deployment Strategies for TLS Decryption.* (Master), The SANS Institute. Retrieved from https://www.sans.org/reading-room/whitepapers/dlp/paper/38240

Global Data Leak Report 2017. Retrieved from https://infowatch.com/report2017

Gugelmann, D., Studerus, P., Lenders, V., & Ager, B. (2015). Can content-based data loss prevention solutions prevent data leakage in Web traffic? *IEEE Security & Privacy, 13*(4), 52-59.

Gupta, K., & Kush, A. (2017). A Review on Data Leakage Detection for Secure Communication. *International Journal of Engineering and Advanced Technology (IJEAT), 7*(1).

Gupta, V. (2013). *File detection in network traffic using approximate matching.* Institutt for telematikk.

Hauer, B. (2015). Data and information leakage prevention within the scope of information security. *IEEE Access, 3*, 2554-2565.

Hemalatha.N.C, Somasundaram.R, & Thirugnanam, M. (2016). Privacy Preserving Data Leak Detection in Large Scale Organizations. *International Journal of Future Innovative Science and Engineering Research (IJFISER), 2*(2). doi:2454- 1966

Huang, X., Lu, Y., Li, D., & Ma, M. (2018). A novel mechanism for fast detection of transformed data leakage. *IEEE Access, 6*, 35926-35936.

Katz, G., Elovici, Y., & Shapira, B. (2014). CoBAn: A context based model for data leakage prevention. *Information Sciences, 262*, 137-158.

Kaur, K., Gupta, I., & Singh, A. K. (2017). *A Comparative Evaluation of Data Leakage/Loss prevention Systems (DLPS).* Paper presented at the Proc. 4th Int. Conf. Computer Science & Information Technology (CS & IT-CSCP), Dubai, UAE.

Kleene, S. C. (1951). *Representation of events in nerve nets and finite automata*. Retrieved from

Liu, Y., Corbett, C., Chiang, K., Archibald, R., Mukherjee, B., & Ghosal, D. (2009). *SIDD: A framework for detecting sensitive data exfiltration by an insider attack.* Paper presented at the 2009 42nd Hawaii International Conference on System Sciences.

Matthee, M. H. (2016). *Tagging Data to Prevent Data Leakage (Forming Content Repositories)*. Retrieved from SANS Institute InfoSec Reading Room:

Mogull, R., & Securosis, L. (2007). Understanding and selecting a data loss prevention solution. *Technicalreport, SANS Institute*, 27.

Oğuz, B., & Cevahir, H. K. (2010). BT Yönetiminde Bilgi Sızıntısı ve Ağ Tabanlı Çoklu Protokol Bilgi Sızıntısı Engelleme.

Pesen, M. M. (2015). DLP'de İçerik Analizi Yöntemleri. Retrieved from https://www.sibergah.com/veri-guvenligi/veri-sizintisi-onleme/dlp-de-icerik-analizi-yontemleri/

Ren, L. (2013). DLP Systems: Models, Architecture and Algorithms. Retrieved from https://www.researchgate.net/publication/304080339_DLP_Systems_Models_Architecture_and_Algorithms

Securosis, L. (2010). Understanding and Selecting a Data Loss Prevention Solution. *Securosis, LLC,[Online]. Available: https://securosis. com/assets/library/reports/DLP-Whitepaper. pdf*.

Shabtai, A., Elovici, Y., & Rokach, L. (2012). *A Survey of Data Leakage Detection and Prevention Solutions*. In P. N. Stan Zdonik, Shashi Shekhar (Series Ed.) (pp. 92). doi:10.1007/978-1-4614-2053-8

Shapira, Y., Shapira, B., & Shabtai, A. (2013). Content-based data leakage detection using extended fingerprinting. *arXiv preprint arXiv:1302.2028*.

Shu, X., Yao, D., & Bertino, E. (2015). Privacy-preserving detection of sensitive data exposure. *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, 10*(5), 1092-1103.

Shu, X., Zhang, J., Yao, D. D., & Feng, W.-C. (2016). Fast detection of transformed data leaks.

Soumya, S. R., & Smitha, E. S. (2014). Data Leakage Prevention System By Context Based Keyword Matching And Encrypted Data Detection. *International Journal of Advanced Research in Computer Science Engineering and Information Technology, 3*(1), 375-384.

T.C. Resmi Gazete, Kişisel Verilerin Korunması Kanunu, 6698 C.F.R. (7 Nisan 2016).

Tahboub, R., & Saleh, Y. (2014). *Data leakage/loss prevention systems (DLP).* Paper presented at the Computer Applications and Information Systems (WCCAIS), 2014 World Congress on.

Trieu, L. Q., Tran, T.-N., Tran, M.-K., & Tran, M.-T. (2017). *Document Sensitivity Classification for Data Leakage Prevention with Twitter-Based Document Embedding and Query Expansion.* Paper presented at the 2017 13th International Conference on Computational Intelligence and Security (CIS).