Veritabanı Güvenliğinde Saldırı Tahmini ve Tespiti için Kullanıcıların Sınıflandırılması

Çiğdem Bakır, Veli Hakkoymaz

Özet—**Kullanıcı ve uygulamaların Internet üzerindeki web sitelerini kullanarak bilgi edinmesi, belge ve bilgi paylaşımı, bankacılık ve diğer işletimsel işlemleri gerçekleştirmesi her geçen gün daha da artmaktadır. Ancak son zamanlarda Internet kullanımının yaygınlaşmasıyla birlikte yetkisiz erişim, veri tutarlılığı, veri bütünlüğü ve veri gizliliği gibi birtakım güvenlik problemleri de ortaya çıkmıştır. Bu durum kişisel ve kamusal alanlarda kullanılan bilgilerin korunmasını zorunlu kılmaktadır. Bu çalışmada, kullanıcıların yaptıkları işlem sayısı, IP adresleri, kullandıkları veri miktarı, yaptıkları işlem türü ve üstlendikleri roller gibi kriterler dikkate alınarak kullanıcı saldırılarının tespit edilmesi amacıyla ortak bir model oluşturulmuştur. Böylelikle risk teşkil eden kullanıcı grupları önceden farkedilerek bilgi güvenliğini sağlamak amaçlanmıştır.**

***Anahtar Kelimeler*—Veritabanı Güvenliği, Saldırı Tespit Sistemleri, Log Kayıtları, Risk Analizi, Saldırı Tahmin Sistemleri.**

Abstract— **Nowadays, user and applications are increasingly getting used to sharing documents and information, performing banking and other business operations by using the web sites on the Internet. However, together with the proliferation of the Internet use, some security problems such as unauthorized access, data coherency, data integrity and data confidentiality have resulted in. This necessitates that information used in personal and public fields to be protected. In this study, a common model is formed in order to detect the user attacks by means of the criteria such as the number of operations performed by the users, IP addresses, the amount of data used, types of operations performed to the database and user roles. Thus, in order to provide database security, it is our aim to recognize in advance and categorize the user groups that may pose a risk.**

***Index Terms*—Database Security, Intrusion Detection Systems, Log Records, Risk Analysis, Intrusion Prediction Systems.**

I. GİRİŞ

Teknolojinin hızla gelişmesiyle birlikte hem kişisel amaçlar hem de bankacılık, bilgi paylaşımı, e-ticaret, iletişim gibi birçok alanda Internet vazgeçilmez bir ihtiyaç olmuştur. Ancak Internetle birlikte bilgi sızması, bilginin yetkisiz kişilerce ele geçirilmesi, değiştirilmesi, bilgi gizliliğinin sağlanamaması vb. ortaya çıkan sorunların çözülmesi amacıyla birtakım çalışmalar yapılmıştır[13,14,15,16]. Bu çalışmalar genellikle oluşan saldırı sonrasında yapılacak işlemlerle ilgilidir. Bu çalışma ise, farklı olarak log kayıtları ile olası kullanıcı saldırılarını önceden tahmin etmeyi ve ileride risk oluşturabilecek saldırıları oluşmadan önlemeyi amaçlamıştır. Yani, sistemi tehdit eden kullanıcılar belirlenmiş ve risk analizi yapılmıştır.

Saldırısızlık garantisi olmadığı (bilgi/kaynak paylaşımının vazgeçilmezliği ve sadece iyiliklerle dolu bir dünyada yaşamadığımız) için veritabanı güvenliği problemini ele almak gerekir. Veritabanı sistemlerinde veri/bilgi güvenliği problemi birkaç aşamada ele alınabilir.

Bilgi güvenliği problemi oluşmadan önce: Potansiyel saldırganı (uygulama/kullanıcı) saldırı oluşmadan önce, çeşitli sisteme erişim biçimi ya da işlemler gibi davranışlara bakarak tahmin etme ve sistem yönetici ve sorumlularını bu potansiyel saldırı ihtimaline göre uyanık olmalarını sağlamak amaçlanır.

Bilgi güvenliği problemi oluştuktan sonra: Bilgi güvenliği problemi olmayacak varsayımı ile hiçbir önlem almama ve problemin oluşmasını bekleme durumudur. Ancak problem oluştuktan sonra sistemin durumunu inceleyerek bir saldırı ya da bilgi güvenliği problemi oluştuğunu tespit etmek ve bu sorunu ortadan kaldırmak için yapılacak çalışmalar.

Eğer bir sistem, bilgi güvenliği problemi olmayacağını garanti etmiyor ve saldırı tespit/kurtarma mekanizması sunmuyorsa, sisteme bir saldırı gerçekleşebilir ancak kullanıcılar ne olduğunu ayırt edemez. Bu durumda güvenli olmayan sistem, kullanıcılarca kullanılmaya devam edecek ve istenmeyen durumlar oluşacaktır. Bu, kabul edilemez bir durum gibi gözükse de eğer saldırı olma olasılığı çok düşükse ve sistem kritik bilgi içermiyorsa tolere edilebilir bir durumdur. Ancak tersi durumda, bilgi güvenliğini garanti eden ve saldırı potansiyellerini ortadan kaldıran önlemleri maliyetine bakmaksızın değerlendirmek gerekir. Bu çalışma, bilgi güvenliği problemi oluşmadan önce alınacak önlemler kapsamında değerlendirilmelidir.

Yapılan çalışmada örnek log kayıtları analiz edilerek veritabanına aktarılmıştır. Buradan, kullanıcıların davranışlarını modelleyebilecek çeşitli kıstaslar kullanarak kurallar oluşturulmuştur. Diğer çalışmalardan farklı olarak bilgi sızmalarına karşı bir risk analizi gerçekleştirilmiştir. Ayrıca oluşturulan modele göre *en riskli*, *riskli*, *orta riskli*, *düşük riskli* ve *en düşük riskli* olmak üzere kullanıcılar risk durumuna göre beş grupta kümelenmiştir.

Çiğdem Bakır, Yıldız Teknik Üniversitesi, Elektrik-Elektronik Fakültesi, Bilgisayar Mühendisliği, İstanbul, Türkiye,(e-mail: cigdem@ce.yildiz.edu.tr)

Veli Hakkoymaz, Yıldız Teknik Üniversitesi, Elektrik-Elektronik Fakültesi, Bilgisayar Mühendisliği, İstanbul, Türkiye,(e-mail: veli@ce.yildiz.edu.tr)

Yapılan çalışmanın *ilk aşamasında* düzensiz, dağınık halde bulunan log kayıtları toplanmış ve filtreden geçirilerek gereksiz, üzerinde işlem yapılmayacak alanlar temizlenmiştir[3]. Böylelikle log kayıtları analiz edilerek içerdiği veriler anlamlı bilgi haline getirilmiştir. *İkinci aşamada*, normalize edilmiş veriler veritabanına aktarılarak kullanıcı kullanım ve operasyonlarını takip etmek için çeşitli kurallar oluşturulmuştur. *Üçüncü aşamada*, oluşturulan kuralları kullanarak kullanıcı saldırılarını belirleyebilmek amacıyla, ortak bir model gerçekleştirilmiştir. *Son aşamada* ise oluşturulan modelle hesaplanan risk oranına bağlı olarak kullanıcılar risk durumlarına göre gruplandırılmış ve veritabanına aktarılmıştır. Ek olarak, risk grubu en yüksek düzeyde olan kullanıcılar ve IP adresleri raporlanmıştır.

Makalenin kalan kısmı şu şekilde organize edilmiştir: İlgili ve benzer çalışmalar II. kısımda ele alınacaktır. III. kısımda log kayıtları ve bilgi güvenliği anlatılırken, IV. kısımda yapılan çalışma ve tasarımı anlatılacaktır. Son kısımda (V), çalışmanın katkıları özetlenecek ve gelecek çalışmaların ne yönde olacağı tartışılacaktır.

II. İLGİLİ ÇALIŞMALAR

Bu çalışmada veritabanı güvenliği özellikle sisteme tanımlı olan kullanıcılar tarafından yapılabilecek saldırıların tahmin edilmesi amaçlanmıştır. Sistem kullanıcılarının risk bazlı sınıflandırılmaları için ortak bir model geliştirilerek kullanıcıların risk oranları hesaplanmıştır. Virüs, solucan, Truva atı, casus yazılım gibi kötümcül yazılımlar ya da sahte IP adresi ile sisteme girme, port tarayıcı saldırıları, kullanıcıların yapmış oldukları hatalı işlemler ve yetkisiz kullanıcıların veri üzerinde okuma ve yazma yaparak sisteme zarar vermesi veritabanı güvenliğini olumsuz etkilemekte, bu durum bu tür ölçümlerle tahmin edilmektedir[17,19].

Veritabanı saldırı tespit sistemleri imza tabanlı ve imza tabanlı olmayan sistemler olmak üzere ikiye ayrılır. İmza tabanlı olanlar, daha önceden gerçekleşmiş, bilinen saldırıların sayısal imzalarının kaydedildiği bir veritabanından yararlanırlar[4]. Bu saldırılar genelde virüs, solucan, Truva atı ve casus yazılım gibi kötücül yazılımlar aracılığıyla yapılmaktadır[8]. Bunların dezavantajları, imza bilgilerinin güncel tutulmasının maliyetli olması, yani karşılaşılan saldırıların imzası henüz bilinmediğinden yeni saldırı türlerini tespit edememesidir.

İmza tabanlı olmayan saldırı tespit sistemleri, sistemde gerçekleşen anormal durumların tespitinde kullanılmaktadır. Bunlar kullanıcı erişimi, rol bazlı erişim gibi iç kaynaklı saldırıları tespit etmek içindir. Saldırıların çoğu, bazı kullanıcıların donanım ve yazılım açıklarını kullanarak istedikleri bilgilere ulaşması ile oluşur. Genelde SQL enjeksiyon saldırılarını içerir[4].

Veritabanı saldırılarını önceden tahmin etme ve güvenlik açıklarını saldırı oluşmadan farketmek amacıyla çok etmenli istatiksel bir tahmin sistemi Quickprop sinir ağları geliştirilmiştir[19]. Bu çalışmada gizli katmanlarını hesaplayabilmek için Pearson korelasyon katsayısı kullanılmış ve bir banka verisi üzerinde yetkisi olmayan kullanıcılar belirlenmeye çalışılmıştır. Kısa vadede gerçekleşen anormal ve hatalı kullanıcı davranışları bulunmuştur. Ancak uzun vadede gerçekleşen potansiyel riskli kullanıcılar ve kontrolsüz kullanıcı işlemleri bir günlük log kayıtları kullanıldığından tam olarak belirlenememiştir. Yani, uzun vadede gerçekleştirilecek saldırılar için bir risk analizi içermemektedir.

Hatalı kullanıcı davranışlarını çözebilmek amacıyla yapılan bir başka çalışma da, genetik algoritma kullanımıdır[20]. Genetik algoritma, sinir ağlarına dayalı olarak ağ özelliklerinden çeşitli kurallar oluşturarak elde edilen kurallar ile sınıflandırma yapar. Bu çalışmanın sonuçları diğer çalışmalarla karşılaştırılmalı olarak verilmiştir. Ancak bu çalışma da öncekinde olduğu gibi kısa vadede gerçekleşebilecek saldırılar için bir çözüm önerisi getirmiştir.

Saklı Markov Modeli kullanarak saldırıyı tahmin ve önleme çalışması, diğer bir ilgili çalışmadır[21]. Saklı Markov Modeli, verilen durumlardan yola çıkarak gizli durumları bulmak için gerçekleştirilen bir sınıflandırma algoritmasıdır. Dağıtık veriler birbirleriyle çok büyük ağlar üzerinde haberleşir ve bu sebeple ciddi saldırılara açıktır. Bu çalışmada fuzzy tekniği kullanılarak risk analizi yapılmış, tehlikeli olarak giden paket oranı tespit edilmeye çalışılmıştır. Ayrıca dağıtık çevreler için risk oluşturacak saldırılar belirlenmeye çalışılmıştır.

Rol tabanlı erişim kontrol modeli kullanan saldırı tespit sisteminin ilk aşamasında veritabanı günlüğü incelenir. Geçmiş hareketler ve dahil oldukları rollere göre sınıflandırma modeli (Naive Bayes) oluşturulur. Kullanıcı hareketleri bu modele göre sınıflandırılır. Sonuçta bulunan rol veritabanı günlüğündeki kullanıcılarla karşılaştırılır. Buna göre, eğer bulunan rol ilgili kullanıcıyla tanımlı ise bir rol olarak kabul edilir, değilse alarm verilir. Ancak bu çalışma, sadece kullanıcıların rollerine göre saldırı tespiti yapmakta, kullanıcıların bireysel olarak yapmış oldukları hareketleri göz önüne almamaktadır[13].

Veritabanında kullanıcı erişiminden kaynaklanan saldırı tespit sistemine örnek Detection of Misuse in Database Systems (DEMIDS) gösterilebilir. Bu çalışma, ilişkisel veritabanları için iç kaynaklı saldırıların (hatalı davranışlar) tespit edilmesi içindir. Bu sistem denetleyici, veri işleyici, profil düzenleyici ve algılayıcı olmak üzere dört ana bileşenden oluşur. Denetleyici verileri toplar ve denetim günlüğüne kaydeder. Veri işleyici, verileri istenen yapı ve türlere dönüştürür. Profil düzenleyici, öğrenme ile herbir kullanıcı için bir profil oluşturur. Denetim aşamasında, kullanıcı aktivitelerinin şüpheli olup olmadıkları sık kullanılan kullanıcı profilleri ile karşılaştırılarak hesaplanmıştır. Önceki çalışmadan farklı olarak, sadece sistemdeki kullanıcı hareketleriyle değil, veritabanında olmayan kullanıcı için de bir profil belirlemektedir[14].

Çeşitli veri madenciliği teknikleri kullanan veritabanı saldırı tespit sistemleri veri nesneleri arasındaki okuma ve yazma bağımlılıklarını analiz eder[15]. Öncelikle saldırı içermeyen veritabanı hareketleri (günlüklerde) analiz edilerek okuma ve yazma bağımlılıklarını ifade eden kurallar belirlenir. Bu kurallar sistem için öğrenilen bir modeli oluşturur. Daha sonra gelen yeni hareketlerin, bu kurallara uyup uymadığına bakılarak saldırılar tespit edilir. Bu çalışma gerçek veriler için

yapılmamış olup, büyük sentetik veriler için veritabanı sadece okuma ve yazma hareketleri için yapılmıştır.

Gerçek zamanlı veritabanı sistemlerinde, hareketler için belirli zaman kısıtlamaları (deadline) vardır. Bu veritabanları, değerleri zaman içerisinde değişen ve periyodik olarak güncellenen zaman boyutlu veri nesneleri için tasarlanmıştır. Hareketler belirli zaman sınırları içerisinde tanımlanmıştır. Gerçek zamanlı veritabanı sistemleri için Lee tarafından yapılan bir çalışmada, hareketlerin zaman imzaları kullanılmış ve zaman boyutlu verileri güncelleyen hareketler için ortaya çıkarmıştır[16]. Farklı olarak, gerçek zamanlı sistemlerde sadece yazma hareketi için güvenlik uyarısı verir.

Yapılan çalışmada gerçekleştirilen model ile kullanıcı erişiminden kaynaklanan saldırıların olup olmamasına bakılmaksızın kullanıcılar risk durumlarına göre sınıflandırılmıştır. Sistemdeki tüm kullanıcıların hareketleri dikkate alınmıştır ve riskli kullanıcılar belirlenmeye çalışılmıştır. Böylelikle bir veritabanı saldırı tahmin sistemi geliştirilmeye çalışılmıştır.

III. BİLGİ GÜVENLİĞİ İÇİN LOG KAYITLARI

Bilgi güvenliği, bilgilerin izinsiz olarak yetkisiz kişilerin kullanımından ve değiştirilmesinden korunmasıdır. Bütünlük, gizlilik ve erişebilirlik olmak üzere üç temel unsurdan meydana gelmektedir. Bütünlük bilginin yetkisiz kişilerce yapılan rastgele değişiklerden korunmasını; gizlilik bilginin yetkisiz kişilerin erişimine izin verilmemesini; erişebilirlik ise bilginin yetkili kişiler tarafından ulaşılabilir olmasını ifade eder[9,11]. Bilgi güvenliğiyle sistemi tehdit eden riskler belirlenir. Gizlilik, erişebilirlik ve bütünlük sağlanarak iş sürekliliği artar[10].

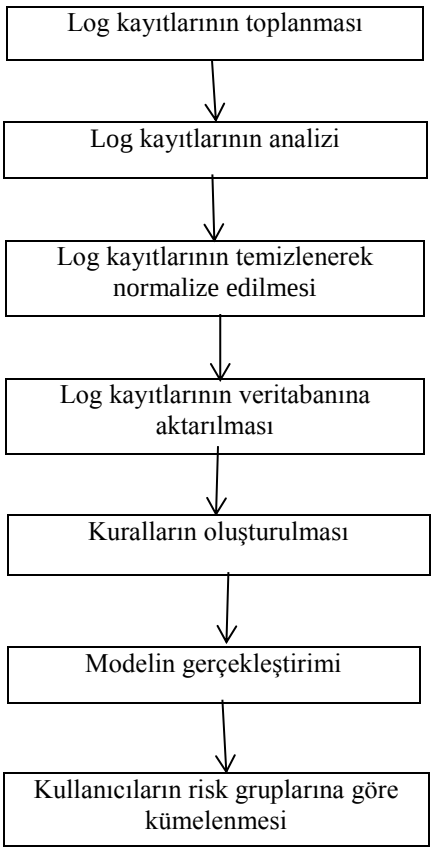
Günümüzde bilgi teknolojilerinin yaygınlaşmasıyla özellikle uygulamaların Internet üzerine taşınması ve Internet ortamında birçok bilginin paylaşılması ve hemen hemen tüm işlemlerin internet üzerinden gerçekleşmesi kötü niyetli veya yetkisiz kişilerin sistem bütünlüğüne zarar vermesine yol açmaktadır[2,12]. Dolayısıyla bir güvenlik sorunu ortaya çıkmaktadır. Veritabanında log yönetimi değişen bilgilerin izlenmesini olanaklı kıldığı gibi, bilgi güvenliğinin de en temel yapısını oluşturur. Bilgi güvenliği ihlallerinin yaşanmaması için kesintisiz bir log yönetimi gerekir. Log yönetimiyle birlikte sistemdeki tüm kullanıcılar, işlemler, işlemlerin zamanı, IP adresleri, başarılı ya da başarısız olma durumu izlenir ve tehlikeli bir durumla karşılaşıldığında log kayıtlarına bakılarak nedeni belirlenir. Bilgi güvenliği ve risklerin ortadan kaldırılması bakımından log yönetimi büyük önem taşır[1].

Log kayıtları incelenerek, veritabanında hatalı kullanım ve saldırılarının tespiti yapılmaktadır. Bunun için veritabanı nesneleri ve bu nesnelere yapılan işlemler değerlendirilir. Özellikle veritabanı sistem nesnelerine erişim işlemleri izlenmelidir. İşlemler ise özellikle sadece okunma amaçlı tasarlanmış bir tabloda değişiklik yapmaya çalışan bir işlem zararlı olarak görülebilir ve tespiti önemlidir.

Log yönetimi farklı kaynaklardan toplanan log verilerinin tek bir merkezde tutulması ile kolayca gerçekleştirilebilir[6]. Log yönetimiyle sadece sistemdeki anormal durumlar belirlenmez. Aynı zamanda ileride oluşabilecek güvenlik sorunlarına karşı da kurumlarda çalışan personeller için farkındalık oluşturur. Cobit ve ISO27001 gibi uluslararası standartlar log yönetimini desteklemektedir[7]. Log kayıtları tek bir merkezde toplanır, depolanır ve istenildiğinde hızlı bir şekilde erişilir. Ayrıca log kayıtları istenildiği zaman, geri getirilme özelliğine sahiptir. Veri bütünlüğü, tutarlılığı ve veri gizliliğinin sağlanması için veriler ile ilgili kullanıcıların yetkileri ve ilişkiler düzeyindeki erişimi belirlenir. OSSIM, Manage Engine Event Log Analyzer, Swatch ve File System Auditor gibi hazır yazılımlar log analizini gerçekleştirmek için kullanılan açık kaynak kodlu ücretsiz yazılımlardır[7].

IV. ÇALIŞMA ADIMLARI VE SİSTEM TASARIMI

Log kayıtlarından veri güvenliğine tehdit oluşturabilecek kullanıcıların belirlenmesi çalışması birden fazla adımı içermektedir. Çalışmanın genel adımları Şekil 1'de gösterilmiştir.



Şekil. 1. Log kayıtları ile risk gruplarını belirleme adımları

A. Log Kayıtlarının Toplanması ve Normalize Edinmesi

Örnek çalışmada, 1 ay içinde sisteme giren 2673 kullanıcı ve 10872 işlem esas alınarak log veritabanı oluşturulmuştur. Ancak toplanan log kayıtları (veri) karmaşık ve düzensiz bir haldedir. Toplanan log kayıtlarını anlamlı hale getirmek ve çalışma ile amaçlanan hedefe ulaşabilmek için çeşitli web madenciliği ve veri madenciliği teknikleri ile veri normalize edilmiş ve gerekli önişlem adımları yapılmıştır[1]. Normalize edilecek veri, log kayıtlarının içerdiği bilgiler, sunucu tarafından toplanan veriler, kullanıcılar ve kullanıcıların

gerçekleştirdiği işlemlerden oluşur. Tüm gereksiz ve kullanılamayacak olan veriler log kayıtlarında temizlendikten sonra benzer işlemi gerçekleştiren kullanıcılar, yaptıkları işlemler, işlem miktarları ve kullanıkları veri miktarları belirlenerek veriler düzenli hale getirilmiştir. Normalizasyon, log dosyasındaki verilerden anlamlı bilgi oluşturabilmek için yapılan veri temizleme, filtreleme, raporlama ve indexleme işidir. Yani, log kayıtlarının istenilen formatta kullanılmasını sağlamak için gereksiz alanlar çıkarılarak normalizasyon işlemi yapılmıştır. Şekil 2'de normalize edilen bir kısım log kayıtlarının biçimi gösterilmiştir.

Results Messages

	ID	Kullanici	Tarih	Girdigi_IP_Adresi	Islem_Turu	Kullandigi_Veri
1	1	admin	2014-09-01 01:00:00.000	195.174.39.01	rwu	60
2	2	kullanici4	2014-09-05 17:44:00.000	217.251.10.63	rw	620
3	3	yazilim1	2014-09-04 12:14:00.000	198.174.39.08	u	16
4	4	kullanici4	2014-09-06 09:08:00.000	215.250.41.19	rw	452
5	5	muhasebe1	2014-09-04 14:12:00.000	195.174.39.06	w	55
6	6	muhasebe2	2014-09-05 15:19:00.000	195.174.39.07	w	15
7	7	kullanici1	2014-09-02 04:53:00.000	192.168.2.125	r	470
8	8	kullanici1	2014-09-04 09:37:00.000	192.168.2.125	r	678
9	9	kullanici2	2014-09-03 15:27:00.000	214.254.120.55	w	189
10	10	kullanici3	2014-09-02 04:55:00.000	198.164.45.15	u	561
11	11	kullanici5	2014-09-01 14:26:00.000	165.101.50.12	rwu	784
12	12	kullanici3	2014-09-07 11:07:00.000	198.164.45.15	u	58
13	13	kullanici5	2014-09-03 00:00:00.000	165.101.50.12	rwu	125
14	14	kullanici6	2014-09-01 19:51:00.000	198.167.54.25	rw	254
15	15	kullanici7	2014-09-04 21:27:00.000	194.164.65.15	wu	541
16	16	kullanici8	2014-09-05 18:16:00.000	65.14.152.36	w	256
17	17	kullanici7	2014-09-06 10:54:00.000	194.164.65.15	wu	15
18	18	yazilim2	2014-09-04 16:04:00.000	198.174.39.12	u	25
19	19	kullanici5	2014-09-02 22:41:00.000	165.101.50.12	rwu	745
20	20	kullanici2	2014-09-06 18:36:00.000	198.164.45.15	u	201

Query executed successfully.

Şekil. 2. Normalize edilmiş log kayıtları

B. *Log Kayıtlarının Veritabanına Aktarılması*

Results Messages

	ID	Kullanici	Islem_Sayisi	Islem_Turu	Kullandigi_Veri	Girdigi_IP_Adresi
1	1	kullanici254	2501	rw	642	198.164.45.16
2	2	kullanici16	2416	r	1245	65.14.152.40
3	3	kullanici541	2364	rw	541	217.125.42.20
4	4	kullanici1264	2350	rwu	2510	165.100.51.20
5	5	kullanici3	2343	u	3406	198.164.45.15
6	6	kullanici2543	2267	w	154	198.160.46.24
7	7	kullanici64	2264	r	184	65.15.150.37
8	8	kullanici27	2258	wu	6217	195.176.40.20
9	9	kullanici184	2243	ru	1567	195.120.41.12
10	10	kullanici1026	2196	ru	2589	198.164.45.17

Query executed successfully.

Şekil. 3. Log Kayıtlarında kullanacak kıstasların veritabanına aktarılması

Şekil 2'de gösterilen normalize edilen kısmi log kayıtları oluşturulacak modeldeki kıstaslarda kullanılan işlem miktarı, işlem türü, kullandığı veri miktarı ve IP adreslerini gösterecek biçimde Şekil 3'de verilmiştir.

C. *Modelin Gerçekleştirilmesi*

Saldırı tahmin ve tespit sistemlerinde, özellik seçimi, modeli oluşturmakta oldukça önemlidir[18]. Bu çalışmada sistem kullanımı, kullanılan veri miktarı, kullanıcı yetkileri ve IP adresi ve sistemi yöneten ve gerçekleştiren kullanıcıların üstlendikleri roller model oluşturmak için temel kriterler olarak belirlenmiştir. Yapılan diğer çalışmalarda tek ya da iki kriter kullanılarak model gerçekleştirimi sağlanmıştır[17,18]. Gerçekleştirilen çalışmada ise birden fazla ve en çok kullanılan kriterler seçilerek sistemin güvenliğini tehdit edebilecek kötü niyetli kullanıcıların risk durumları analiz edilmiştir. Ayrıca sistem kullanıcılarının hepsini kapsayan ortak bir model gerçekleştirilmesi amaçlanmıştır. Sistem kullanımı ve sistemde kalma süresi çeşitli saldırıları ayırt etmeyi sağlar. Kullanıcıya verilen yetkiler, veritabanında meydana gelen hatalı işlemleri ya da yetkisi olmayan kullanıcıların sisteme girip girmediklerini belirlemede ana göstergelerdendir[4]. Kullanıcıların, okuma ve yazma işlemlerine yetkili ya da yetkisiz olup olmadığını yansıtır. Ayrıca port numarası, kaynak ve hedef IP adresi bilgileri SQL komutlarının gruplandırılmasında kullanılır. Veritabanı haberleşmesi farklı port üzerinde gerçekleştirilir. Port numarası, oturumda paketleri düzenlemek, paketlerin hatalı gidip gitmediğini, verinin eksik bir şekilde sunucuya gönderilip gönderilmediğini denetleyebilmek için kullanılabilir. Sahte IP adresi ile sisteme grime ve port tarayıcı saldırılarını tespit amacıyla IP adresi ve veri miktarı model oluşturmak için seçilen kriterlerdendir[17,19].

D. *Kuralların Oluşturulması*

Bizim çalışmamızda, kullanıcıların sistemdeki kullanım ve operasyonlarını dikkate alarak bazı kıstaslar tanımlanmış ve bu kıstaslara göre risk oranını belirleyebilmek için kurallar geliştirilmiştir (bkz. Şekil 3). Risk oranının hesaplanması için öngörülen kurallar aşağıdaki gibi belirlenmiştir:

1) Sistem Kullanımı: Sistemde kalma süresi, sisteme giriş sayısı ve sisteme ne kadar sıklıkla girdiği kullanıcının sistemle ne kadar alakalı olduğunu gösterir. Kullanıcının sistemde kalma miktarı arttıkça sistemle ilgili bilgi alma, bilgiyi değiştirme ya da bilgi ekleme gibi temel hareket (transaction) işlemlerini gerçekleştirme olasılığı artar. Bu sebeple kullanıcının sistem kullanım durumu bu çalışmada kıstas olarak kullanılmıştır. Sisteme girme sıklığına göre aşağıdaki kurallar belirlenmiştir: Sisteme,

- 1-10 kez girenlerin ağırlıkları 0,
- 11-20 kez girenlerin ağırlıkları 1,
- 21-50 kez girenlerin ağırlıkları 2,
- 51-100 kez girenlerin ağırlıkları 3,
- 101-250 kez girenlerin ağırlıkları 4,
- En az 251 kez girenlerin ağırlıkları 5,

olarak belirlenmiştir. Bu kurallar örnek veritabanındaki kullanıcıların kalma miktarına göre büyükten küçüğe doğru

sıralandığında, tüm örnek log kayıtlarının dengeli bir şekilde dağılması amacıyla bu biçimde oluşturulmuştur.

2) *Kullanılan veri miktarı:* Bir kullanıcının sistemde kullandığı veri miktarı kullanıcılar arasında risk analizi yapabilmek için önemlidir. Çünkü kullanılan veri miktarı arttıkça kullanıcılar sistemde daha fazla bilgi edinebilir ve daha fazla bilgiyi kullanarak değiştirme imkanı bulabilir. Sistemde kullanılan veri miktarına göre;

- 0-50 KB arası kullananların ağırlıkları 0,
- 51-100 KB arası kullananların ağırlıkları 1,
- 101-150 KB arası kullananların ağırlıkları 2,
- 151-500 KB arası kullananların ağırlıkları 3,
- En az 501 KB arası kullananların ağırlıkları 4,

olarak belirlenmiştir.

3) *Kullanıcı Yetkileri:* Bir sistemde kullanıcı yetkileri read, write ve update işlemleri için belirlenir[5]. Sistem üzerinde okuma ve yazma yetkisi olmayan bir kullanıcının gizli bilgiler içeren veriyi okuması, veri üzerinde değişiklik yapması ve veri üzerine yazması veri gizliliği, veri bütünlüğü ve veri güvenirliği açısından risk taşır.
Sistemde kullanıcı yetkilerine göre;

- Read işlemi yapanların ağırlıkları 1,
- Write işlemi yapanların ağırlıkları 2,
- Update işlemi yapanların ağırlıkları 3,

olarak belirlenmiştir.

4) *IP Adres Sıklığı:* En çok kullanılan IP adresleri veritabanı güvenliği için riskli olarak kabul edilmiştir. Sistemde en çok kullanılan ilk 100 IP adresi belirlenmiştir.

- Sık kullanılmayan IP adreslerinden bağlananların ağırlıkları 0,
- Sık kullanılan 100 IP adreslerinden bağlananların ağırlıkları 1,

olarak belirlenmiştir.

5) *Sistemi yöneten ve gerçekleyen kullanıcıların üstlendikleri roller:* Sistemi gerçekleştiren kullanıcılar içinde yetkisi dışında başka kullanıcıların hareketlerini (transaction) görüntülemesi veri gizliliği, veri bütünlüğü ve veri güvenirliği açısından risk taşır. Bu kullanıcıların üstlendikleri rollere bakarak bir risk ve güven analizi yapılmaktadır[13]. Sistemi yöneten ve gerçekleştiren kullanıcıların üstlendikleri rollere göre;

- Yapılan işlemleri sadece görüntüleyenlerin ağırlıkları 0,
- Yapılan işlemleri görüntüleme ve yazma yapanların ağırlıkları 1 ,

olarak belirlenmiştir.

E. Kullanıcıların Kümelenmesi

Kurallar oluşturulduktan sonra risk oranının hesaplanabilmesi için eşitlik 1 kullanılmıştır:

$$Risk\ Oranı = \sum_{i=1}^{n} (w_{a,i}, w_{b,i}, w_{c,i}, w_{d,i}, w_{e,i})/T \qquad (1)$$

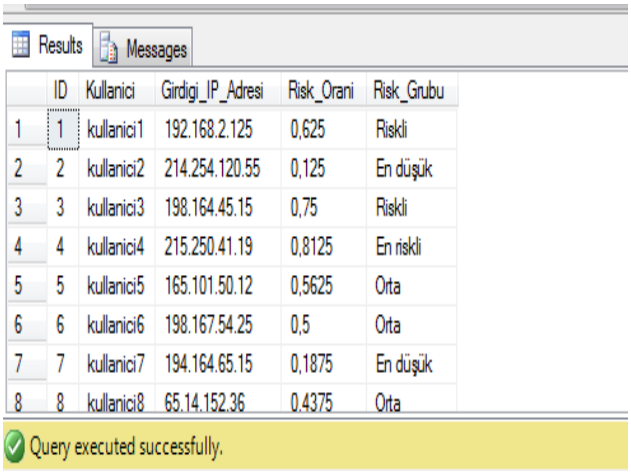
Bu hesaplamada, $w_{a,i}$ **a** özelliğinin (sisteme giriş sayısı), $w_{b,i}$ **b** özelliğinin (kullanılan veri miktarı), $w_{c,i}$ **c** özelliğinin (kullanıcılara verilen yetki), $w_{d,i}$ **d** özelliğinin (IP adresi), $w_{e,i}$ ise **e** özelliğinin (sistemi yöneten ve gerçekleyen kullanıcıların rolleri) ağırlığını gösterir. **n** sistemde bulunan kullanıcı sayısını gösterirken, **T** ise herbir kriter içerisinde en güçlü ağırlığa sahip kuralların oluşturduğu toplam ağırlığı ifade eder. Ancak kullanıcı yetkileri kriteri için bir kullanıcı tüm yetkileri (read, write, update) gerçekleştirebileceğinden dolayı risk oranı hesaplanırken bu kriterin tüm ağırlıkları **T** değişkenine aktarılmıştır. Yukarıdaki formül dikkatle incelendiğinde risk oranının 0-1 arasında değerler alacağı gözlenecektir. Buna göre risk sınıflandırılması yaparsak;

Risk Oranı =
0-0.2 ise *en düşük riskli*
0.21-0.4 ise *düşük riskli*
0.41-0.6 ise *orta riskli*
0.61-0.8 ise *riskli*
0.81-1.0 ise *en riskli*

olmak üzere sistemi tehdit eden 5 grupta kümelenir. Tablo 1 de bazı kullanıcıların geliştirilen modele göre risk oranları ve risk grupları gösterilmektedir. Ayrıca bu bilgiler riskli kullanıcıların bulunmasını kolaylaştırmak için ilgili personel tarafından kullanılmak üzere veritabanında kaydedilmiştir (bkz.Şekil 4).

TABLE I
SISTEMDEKI BAZI KULLANICILARIN RISK ORANLARI VE RISK GRUPLARI

ID	Kullanıcı	Risk Oranı	Risk Grubu
1	kullanici1264	1	En riskli
2	kullanici10	0.75	Riskli
3	kullanici124	0.25	Düşük
4	kullanici26	0.18	En düşük
5	kullanici2350	0.68	Riskli
6	kullanici1557	0.12	En düşük
7	kullanici65	0.5	Orta
8	kullanici546	0.56	Orta
9	kullanici42	0.37	Düşük
10	kullanici9	0.06	En düşük


Results Messages

	ID	Kullanici	Girdigi_IP_Adresi	Risk_Orani	Risk_Grubu
1	1	kullanici1	192.168.2.125	0,625	Riskli
2	2	kullanici2	214.254.120.55	0,125	En düşük
3	3	kullanici3	198.164.45.15	0,75	Riskli
4	4	kullanici4	215.250.41.19	0,8125	En riskli
5	5	kullanici5	165.101.50.12	0,5625	Orta
6	6	kullanici6	198.167.54.25	0,5	Orta
7	7	kullanici7	194.164.65.15	0,1875	En düşük
8	8	kullanici8	65.14.152.36	0,4375	Orta

Query executed successfully.

Şekil. 4. Kullanıcıların risk oranına göre gruplandırılması

V. DEĞERLENDİRME VE SONUÇ

Bu çalışma bir risk analizi yaparak, güvenlik açıklarının azaltılmasına yardımcı olmayı hedeflemektedir. Riskli kullanıcılar belirlenerek güvenlik analizleri yapılmış, sistemi kullanan personele gereken önlemlerin alınması önerilmiştir. Çeşitli kurallar kullanılarak geliştirilen modele göre sistemi tehdit edenlerin risk oranları hesaplanmış ve risk durumları sınıflandırılmıştır. Log kayıtlarına bakılarak olası risk grubu oluşturacak kullanıcı ve IP adresleri belirlenmiş ve bunlar log dosyasının tutulduğu veritabanına aktarılmıştır. Kümeleme işlemi kullanıcıların sisteme girme sıklığına, kullandıkları veri miktarına, kullanıcılara verilen yetkilere ve IP adreslerine göre yapılmıştır.

Bu çalışma özellikle kurumda güvenlik konusunda çalışan personele yol gösterici olacaktır. Sistemi tehdit edebilecek kullanıcılar önceden belirlenerek muhtemel saldırı senaryolarının önlenmesi amaçlanmıştır. İleride yapılacak çalışmalar için, örnek bir çalışma olarak değerlendirilebilir. Devamında, bu çalışmanın yeni kıstaslar kullanarak genişletilmesi hedeflenmektedir.

KAYNAKLAR

[1] E.Sahinaslan, A.Kanturk etc., "Kurumlarda Log Yönetiminin Gerekliliği", *Akademik Bilişim Konferansları,* 2013.

[2] T.Ozseven, M.Dugenci, "Log Analiz: Erişim Kayıt Dosyaları Analiz Yazılımı ve GOP Üniversitesi Uygulaması", *Bilişim Teknolojileri Dergisi*, pp.55-66, 2011.

[3] T.Aye, " Web log cleaning for mining of web usage patterns", 3rd International Conference on volüme 2, pp.490-494, 2011.

[4] Y.Zhang, X.Ye etc, "A practical database intrusion detection system framework", *In Computer and Information Technology,* vol.1, pp.342-347, 2009.

[5] C.Mohan, B.Linday and R.Obermarck, "Transaction management in the R distributed management system", *ACM Transactions on Database Systems,* vol.11, issue 4, pp.378-396, 1986.

[6] I.Ray, K.Belyaev atc., "Secure logging as a service-delagating log management to cloud", *IEEE Journal Systems,* vol.7, issue.2, pp.323-334, 2013.

[7] K.Kent, M.Souppaya, "Guide to Computer Security Log Management, *National Institute of Standarts and Technology,* 2006.

[8] C.Pfleeger, S.Pfleeger, "Security in Computing, *3rd Prentice Hall Professional Technical Reference,* 2002.

[9] J.Andress, "The Basics of Information Security", *Understanding the Fundamentals of InfoSec in Theory and Practice",* 2014.

[10] M.Tekerek, "Bilgi Güvenliği Yönetimi", *KSU Journal of Science Engineering,* pp.132-137, 2008.

[11] E.Yildiz, "Gerçek Zamanlı Bir Saldırı Tespit Sistemi Tasarımı ve Gerçekleştirimi", *Journal of New World Sciences,* vol.5, no 2, pp.143-159, 2010.

[12] M.Ogun ve A.Kaya, "Siber Güvenliğin Milli Güvenlik Açısından Önemi ve Alınabilecek Tedbirler", *Journal of Security Strategies,* issue 18, pp.145-181, 2013.

[13] E.Bertino, E.Terzi etc., "Intrusion detection in RBAC-administered databases", 21st Annual i*n Computer Security Application Conference,* pp.10- 182, 2005.

[14] C.Chung, Y.Gertz etc., "Demids:A missue detection system for database systems", *Integrity and Internal Control in Information Systems,* vol.37, pp.159-178, 2000.

[15] Y.Hu, B.Pand, "A data mining approach for database intrusion detection", *Proceedings of the 2004 ACM symposium on Applied computing,* pp.711-716, 2004.

[16] V.C.Lee, J.A.Stankovic, "Intrusion detection in real-time database systems via time signatures", *Real-Time Technology and Applications Symposium (RTAS),* pp.124-133, 2000.

[17] Bridges S.M, Vaughnn R.B, "Fuzzy Data Mining and Genetic Algorithms Applied to Intrusion Detection", *23rd National Information Systems Security Conference,* 2000.

[18] Jemili F., Zaghdounal M. Etc, "Hybrid Intrusion Detection and Prediction multiAgent System, HIDPAS", *International Journal of Computer Science and Information Security,* vol.5, no.1, 2005.

[19] Ramasubramanian P. And Kannan A., "Multi-Agent based Quickprop Neural Network Short-term Forecasting Framework for Database Intrusion Prediction System", *CiteSeerX,* 2014.

[20] Romasubramanian P., Kannan A., "A genetic-algorithm based neural network short-term farecasting framework for database intrusion prediction system", *Soft Computing,* vol.10, issue 8, pp.699-714, 2006.

[21] Haslum K., Abrater A. Etc, "Disp: A framework for distributed intrusion prediction and prevention using hidden markov models and online fuzzy risk assasment", *3rd International Symposium on Information Assurance and Security,* pp.183-190, 2007.

Çiğdem Bakır Sakarya Üniversite Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümünden 2010 yılında mezun oldu. 2013 yılında Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği Bölümünde Araştırma Görevlisi olarak göreve başladı. 2014 yılında aynı üniversitenin Bilgisayar Mühendisliği Ana Bilim dalında Yüksek Lisansını tamamlayarak Doktora eğitimine başladı. İlgi alanları; veri madenciliği, bilgi güvenliği, biyomedikal görüntü ve işaret işleme konularıdır.

Veli Hakkoymaz lisans derecesini Hacettepe Üniversitesi Bilgisayar Bilimleri Mühendisliği Bölümünden 1987'de, yüksek lisans derecesini University of Pittsburgh (PA) Bilgisayar Bilimlerinden 1992'de, Doktora derecesini CWRU (OH) Bilgisayar Bilimleri Mühendisliğin'de 1997'de tamamladı. 2011'de Doçent ünvanını aldı. İlgi alanları; veritabanı yönetim sistemleri, bilgisayar mimarisi, işletim sistemleri ve dağıtık sistemlerdir. Halen Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği Bölümünde Öğretim Üyesi olarak görevini sürdürmektedir.