Fırat Üniv. Müh. Bil. Dergisi
28 (2), 291-302, 2016

Science and Eng. J of Fırat Univ.
28 (2), 291-302, 2016

Web Sunucu Erişim Kütüklerinden Web Ataklarının Tespitine Yönelik Web Tabanlı Log Analiz Platformu

Muhammet BAYKARA[1], Resul DAŞ[1*], Gürkan TUNA[2]
[1]Fırat Üniversitesi Teknoloji Fakültesi Yazılım Mühendisliği Bölümü, 23119/ELAZIĞ
[2]Trakya Üniversitesi, Teknik Bilimler Meslek Yüksek Okulu, Bilgisayar Programcılığı, EDİRNE
*rdas@firat.edu.tr

(Geliş/Received: 28.06.2016; Kabul/Accepted: 29.09.2016)

Özet

İnternetin her alanda kullanıldığı ve yaşantımıza yön verdiği bir zamanda yaşamaktayız. Doğal olarak bu değişim bilgisayarları ve dijital dünyayı hayatımızın vazgeçilmezleri arasına yerleştirmiş ve sürekli bir ihtiyaç haline getirmiştir. Günümüzde hemen hemen her şey dijital ortamlarda yapılmaktadır. İşte dijital ortamların bu kadar yaygın olması ve günlük işlerin büyük çoğunluğunun buradan gerçekleştirilmesi, yapılan işlemlerin güvenliğini sağlamanın gerekliliğini de büyük oranda arttırmış, bilgilerin yetkisiz kişilerin eline geçmesini önlemek zorlaşmıştır. Bu sebeple bilgi güvenliğinin sağlanmasına yönelik yapılan çalışmalar da artmıştır. Bu çalışmada web sunucularına yönelik olarak gerçekleştirilen kötücül aktivitelerin tespitine yönelik bir log analiz ve takip yazılımı geliştirilmiştir. Web erişim kayıtlarının analiz edilmesi, sistem üzerinde yaşanabilecek olası saldırıların erken fark edilmesinde ve saldırı örüntülerinin tespit edilebilmesinde önemli bir rol oynamaktadır. Geliştirilen log analiz platformu ile web sunucularına yapılan saldırılar istatistiksel olarak rapor edilebilmektedir.

Anahtar Kelimeler: Bilgi güvenliği**,** Saldırı tespit sistemleri, Log analizi, Web sunucu güvenliği.

Web Based Log Analysis Platform for Detection of Web Attacks from Web Server Access Logs

Abstract

Internet is used in all areas and we are living at a time when it gives direction to our lives. Naturally, these changes have placed computers and digital world has become an indispensable part of our lives and a constant need. Nowadays, almost everything is done in the digital environment. Here the digital realization here the majority of the daily work is so common, the need to ensure the security of transactions has also increased to a large extent, the information is difficult to avoid falling into the hands of unauthorized persons. This is done in order to ensure information security reason, the work also increased. In this study, a log analysis and monitoring platform for the detection of malicious activity performed for the web server has been developed. Analysis of Web access logs, in the early detection of possible attacks that may occur on the system and plays an important role in the attack pattern is detected. Made to the web server log analysis platform developed by attacks it can be reported as statistically significant.

Keywords: Information security, Intrusion detection systems, Log analysis, Web server security.

1. Giriş

Bilgi güvenliği, günümüzde dijital ortamlarda artan veri hacmi düşünüldüğünde önemi her geçen gün daha da artan bir unsur haline gelmiştir. Bilgi sistemlerinde güvenliğin sağlanması için çok çeşitli çalışmalar yapılmaktadır [1]. Bilgi güvenliği sistemlerinde insan faktörü en zayıf halkadır. Sistemler üzerinde kullanım yetkisi bulunan kişiler dahi bilinçsiz kullanım sonucu sisteme farkında olmadan zarar verebilmektedir. Bu sebeple bilgi sistemleri hem iç tehditlere karşı hem de dışarıdan gelebilecek olası kötücül amaçlı tehditlere karşı korunmak zorundadır [1, 2].

Bilgi güvenliği sistemlerinin korunması için alınması gereken birçok önlem vardır. Günümüzde özellikle gelişen teknoloji ile birlikte yeni açıklıklar meydana gelebilmektedir. Bu da güvenliğin sağlanabilmesini zorlaştırmakta; bilgi güvenliği yönetim sistemi araçlarının ve güvenlik yönetiminin sürekli olarak güncellenmesini gerektirmektedir. Bilişim sistemlerinde tehditlerin belirlenmesi ve kötücül aktivitelerin

tespit edilebilmesi için kullanılabilecek veri kaynaklarından birisi de log dosyalarıdır. Bilişim sistemlerinde kullanılan birçok cihaz ve sistem üzerinde meydana gelen olayları kayıt altına alırlar. Bu log kaydı, zamana bağlı olarak hangi kullanıcının hangi aktiviteyi yaptığını gösterir. Log kayıtları bir uçağın kara kutusuna benzer bir şekilde işlev gören text dosyalarıdır. Farklı formatlardaki log kayıtları, üzerinde bulundukları sistem ile ilgili birçok bilgiyi tutmaktadır. Bu bilgiler ile saldırı tespiti, adli bilişim süreç analizi, performans analizleri, adli olayların aydınlatılması vb. mümkün olabilmektedir.

Log dosyaları, öneminden dolayı günümüzde artık yasal bir zorunluluk olarak düzenli bir biçimde arşivlenmesi ve korunması zorunlu hale gelmiştir. Bu zorunluluk belirli standartlar ve kanunlar ile desteklenmiştir. 5651 sayılı kanun, FISMA (Federal Information Security Management Act), HIPAA (Health Insurance Portability and Accountability Act), SOX (Sarbanes-Oxley), ISO 17799 gibi ismi sıkça duyulan kanun, düzenleme ve standartlar da log tutma konusunu zorunlu tutmuştur [3].

Günümüzde artık bilişim sistemleri, otomasyon sistemleri, elektronik doküman yönetim sistemleri vb. birçok yazılım web tabanlı olarak geliştirilmektedir. İnternete erişimin mümkün olduğu her yer ve cihazdan ulaşılabilir olan web tabanlı bu sistemler arttıkça güvenlik hususu da önem kazanmaktadır. Web tabanlı bu sistemler, kurumsal anlamda büyük bir öneme sahip olan altyapılardır. Bu gibi sistemlere yapılabilecek olası saldırılar maddi manevi kayıplara yol açabilmektedir.

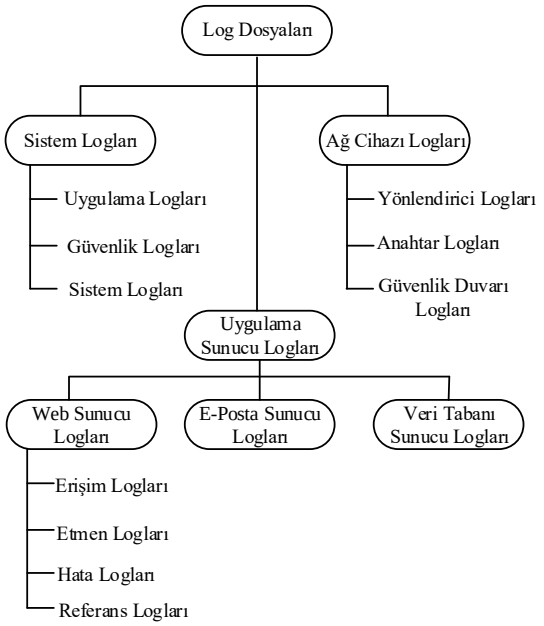
Bu çalışmada web sunucularına yapılabilecek olası saldırıların tespit ve analizine yönelik olarak log analizi ve takip yazılımına sahip web tabanlı bir platform geliştirilmiştir. Sağlıklı gerçekleştirilmiş bir log/izleme altyapısı, yaşanmış saldırıların analizinde veya yaşanacak muhtemel saldırıların erken fark edilmesinde hayati bir role sahiptir. Gerçekleştirilen yazılım ile sistem üzerindeki log kayıtları analiz edilip çeşitli istatistikler çıkarılabilmekte ve saldırı tespiti yapılabilmektedir.

2. Log Dosyaları ve Türleri

Bilişim sistemlerinde birçok sistem üzerinde gerçekleşen olayları log dosyalarında kayıt altına almaktadırlar. Saklanan bu dosyalar türüne göre farklı formatlarda olabilmektedir. Log dosyaları için bir kategorizasyon Şekil 1'de verilmiştir [3].

2.1. Sistem Log Dosyaları

Sistem log dosyaları, işletim sistemleri üzerinde çalışan ve gerçekleşen olayları kaydeden dosyalardır. Microsoft Windows ya da Linux gibi sistemlerde meydana gelen olayları saklayan dosyalardır. Bu kayıtlar, bu tarz sistemlere normal ve anormal girişlerin belirlenmesine yardımcı olmaktadır. İşletim sistemlerinde bu log dosyalarının toplanması, bilişim sistemi yöneticilerinin işletim sistemleri ve uygulamalardaki performans sorunlarını çözmelerinde önemli bir rol oynamaktadır. Sistem loglarındaki bu kayıtlar, sistem üzerinde meydana gelen yetkisiz erişimlerin, tehlikeli aktivitelerin, veri hırsızlığının, virüs ya da solucan gibi çeşitli sebeplerle oluşan kötücül, riskli aktivitelerin tespiti için büyük önem taşımaktadır. Windows işletim sistemlerinde üç çeşit sistem log dosyası vardır. Bunlar uygulama log dosyaları, güvenlik log dosyaları ve sistem log dosyalarıdır [3].



Şekil 1. Log dosya kaynakları [3]

2.1.1. Uygulama Logları

Uygulama logları, işletim sistemleri vb. çatı yazılım sistemleri üstünde çalıştırılan uygulama

ve programların gerçekleştirdikleri olaylar hakkında bilgi içermektedir. Örneğin bir veri tabanı yönetim sistemi programı, uygulama log dosyasına dosya hatalarını kaydedebilir. Program geliştiriciler hangi olayların kaydedileceğine karar verebilmektedir [3, 4].

2.1.2. Güvenlik logları

Güvenlik log dosyası (Security Log), sistem üzerinde oturum yönetimi, dosya yönetimi ve kaynak kullanımı ile ilgili olayları kaydeden log dosyası türüdür. Geçerli geçersiz oturum açma girişimleri, dosya oluşturma, dosya açma, dosya silme vb. kaynak kullanımı olayları güvenlik log dosyalarında tutulmaktadır. Hangi olayların güvenlik log dosyalarına kaydedileceğini belirlemek için ilgili sistem üzerinde yönetici olarak oturum açılmış olması gerekmektedir [3, 5].

2.1.3. Sistem loğları

Sistem logları (System Logs) Windows işletim sistemi bileşenleri tarafından kaydedilen olayları içermektedir. Örneğin, bir sürücü veya sistem bileşeni, işletim sistemi başladığında yüklenemiyorsa bu olay sistem günlüğüne kaydedilmektedir. Sistem bileşenleri tarafından kaydedilen olay türleri Windows tarafından önceden belirlenmektedir [3].

2.2. Uygulama sunucusu log dosyaları

Kişisel bilgisayarlardaki işletim sistemleri gibi sunucular üzerinde bulunan işletim sistemleri de sunucu üzerinde gerçekleşen olaylar ile ilgili kayıtları log dosyalarında tutmaktadır. Kayıtların tutulduğu bu dosyalarda, sunucu üzerinde gerçekleşen erişimler, hatalar, bağlantı hataları, yüklenen dosyalar, çalıştırılan komutlar, erişilen nesneler, giriş kayıtları, veri tabanı aktiviteleri gibi önemli olaylar kaydedilmektedir [3].

2.2.1. Web sunucu loğları

Web sunucu log dosyaları, sunucu platformundan bağımsız düz metin (ASCII) dosyalarıdır. Bu dosyalar, web tarayıcısı üzerinden web sunucusuna erişim sağlayan istemcilerin aktivitelerini düz metin formatında saklamaktadırlar [6, 7]. Saklanan bu bilgilerin analiz edilmesiyle, kullanıcılara ait birçok önemli bilgi elde edilebilmektedir. Kullanıcıların istekte bulunduğu URL adresleri, IP adresleri, işlemlerin gerçekleştirildiği tarih-saat bilgileri, referans başlığı bilgileri, HTTP durum kodları, servis edilen verinin byte olarak miktarı ve kullanıcı etmen dosyaları bu bilgilere örnek olarak verilebilir.

Literatürde genel olarak dört tip sunucu log kaydından söz edilmektedir [3]. Bunlar;

- Erişim Logları (Access Logs)
- Etmen Logları (Agent Logs)
- Hata Logları (Error Logs)
- Referans Logları (Referer Logs)'dır.

Erişim logları:

Web sunucusu tarafından işlenen bütün istekleri kaydeden log dosyalarıdır. Bu dosyalar üzerlerinde bulundukları servisin açılmasından kapanmasına kadar geçen süre içerisinde kullanıcı erişim bilgilerini kaydederler. Sunucu servisi açık olduğu sürece bu dosyalar diskte bulundukları yerden silinemez ya da taşınamazlar. Bu dosyaların silinmesi için sunucu servislerinin kapalı olması gerekmektedir.

Etmen logları:

Etmen logları, sunucu üzerinde sistem yöneticisi tarafından aktif ya da pasif edilebilen dosyalardır. Sistem yöneticisi gerektiğinde sunucu üzerinde bu seçeneği aktif yaparak etmen dosyalarının kaydedilmesini sağlayabilir. Bu dosyaların içerdiği bilgiler birleştirilmiş günlük biçimi tarafından da kaydedilebildiği için sunucu performansını düşürücü bir etken olmaması açısından bu tip log dosyalarının aktif yapılması aksi bir durum olmadıkça tercih edilmez.

Etmen dosyaları, sunucu ile bağlantı kuran istemcilerin web tarayıcı bilgileri ile kullandıkları işletim sistemi bilgilerini tutarlar.

Hata logları:

En önemli log dosyalarından olan hata logları, sistem yöneticilerinin sunucu üzerinde oluşan hataları görmelerini sağlamaktadırlar. Herhangi bir istemci, sunucu üzerinde bulunmayan bir dosyaya erişmek istediğinde kırık link hatası ile karşılaşacaktır. Bu hata sunucu üzerindeki hata log dosyasına kaydedilmektedir.

Referans Logları:

Referans logları, HTTP isteği başlığındaki "referer" alanı bilgilerini kaydetmektedirler. Bu

alanda bulunan doküman adresleri, bir önceki adres de göz önünde bulundurularak referans loglarına kaydedilir.

2.2.2. E-posta sunucu loğları

Bir e-posta sunucusu, SMTP (Simple Mail Transfer Protocol), POP3 (Post Office Protokol 3) ve IMAP (Internet Message Access Protocol) protokollerini kullanarak e-posta alıp gönderme hizmeti vermektedir.

E-posta sunucuları, 25. port üzerinden mesaj gönderip almaktadır. İki e-posta sunucusu aralarında konuşmaya başladıklarında ilk olarak SMTP protokolüne özgü olarak el sıkışma diye tabir edilen safhayı gerçekleştirmektedir. O anda her sunucu hangi kullanıcıdan hangi kullanıcıya e-posta göndereceği ve alınacağı konusunda birbirlerine gerekli bilgileri verirler ve eğer alıcı sunucuda ilgili e-posta hesabı var ise veri akışı başlamaktadır [3]. Bu veri akışı esnasında e-posta sunucuları gelen ve giden e-postalar ile ilgili bilgileri log dosyalarında tutmaktadırlar.

2.2.3. Veri Tabanı Sunucu Logları

Veri tabanı sunucuları, herhangi bir veri tabanı yönetim sisteminin üzerinde bulunduğu sunuculardır. Bu sunucular, web sunucuları gibi işlev görmektedir. Web sunucularında olduğu gibi istemciler ağ ortamından belli kurallar doğrultusunda veri tabanı sunucularına erişim sağlayabilmektedirler. Bu erişimler sonucunda web sunucularında olduğu gibi veri tabanı sunucuları da sunucu üzerinde meydana gelen hatalarla ilgili, başarılı ve başarısız erişimler ile ilgili bilgileri log dosyalarına kaydetmektedirler.

2.3. Ağ Cihazı Log Dosyaları

Ağ cihazları, ağ alt yapısı tarafından gerçekleştirilen işlemler ile ilgili log dosyalarını saklamaktadırlar. Yönlendirici (Router), anahtar (Switch) ve güvenlik duvarı (Firewall) gibi ağ cihazları üzerlerinden geçen ağ trafiği ile ilgili bilgileri kaydetmektedirler. Bu kayıtlar, güvenlik zafiyetlerinin belirlenmesinde ve önlem alınmasında büyük önem taşımaktadır. Cihaz arayüzlerinin durum değişikliği, sistem konfigürasyon değişikliği, erişim listelerine (access list) takılan bağlantılar gibi güvenlik açısından önemli olan bilgilerin kaydı tutulabilmektedir [3].

2.3.1. Yönlendirici Log Dosyaları

Yönlendirici, OSI referans modelinin 3. katmanında çalışan ağ protokollerini destekleyen ve ağları birbirine bağlayan cihazdır. Bir ya da daha fazla ağ arasındaki kesişim noktasıdır. İsminden de belli olduğu üzere, yönlendirme görevinden dolayı iki ağ arasındaki veri iletişiminin doğruluğunu ve iletişimini sağlamaktadırlar.

2.3.2. Anahtar Log Dosyaları

Anahtarlar, ağ üzerindeki bilgisayarların ve diğer ağ aygıtlarının birbirlerine bağlanmasını sağlayan ağ aktif cihazlarındandır. OSI referans modelinin 2. katmanında çalışırlar. Ancak yeni nesil anahtarlarda IP Routing özelliği de olduğundan bu anahtarlar OSI referans modelinin 3. katmanında da çalışmaktadırlar. Diğer ağ cihazları gibi anahtarlar da üzerlerinde gerçekleştirdikleri işlemler için log kayıtları tutmaktadırlar.

2.3.3. Güvenlik Duvarı Log Dosyaları

Güvenlik duvarları bilgisayarlar ve ağlar arasındaki ağ trafiği akışını kontrol eden yazılım veya donanımlardır. Günümüzde güvenlik duvarları artık sadece ağ trafiğini denetlememekte, ayrıca bazı saldırıların tespitini de yapabilmektedir. İnternet ve intranet kullanıcıları güvenlik duvarlarını kullanarak bağlantılarını güvenli hale getirmektedirler. Yönlendirici ve anahtarlarda olduğu gibi donanımsal ya da yazılımsal olarak hizmet sunan güvenlik duvarları da ağ ortamında gerçekleşen olayları log dosyalarına kaydetmektedirler.

3. Saldırı Tespit Sistemleri Açısından Log Dosyaları

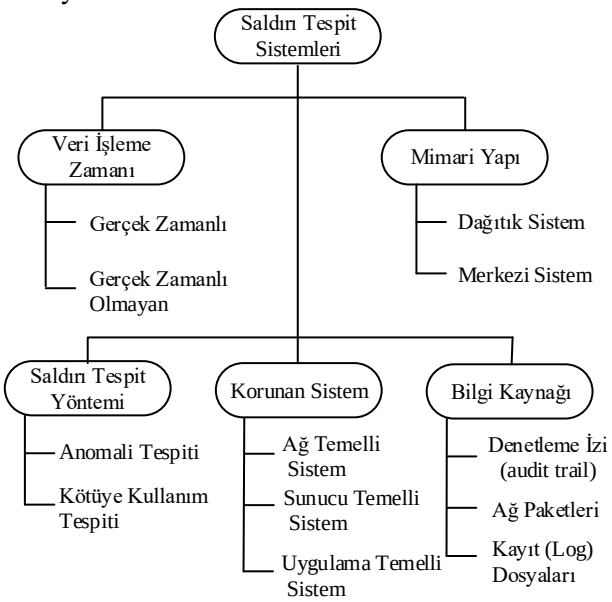
Saldırı tespiti, bilişim sistemlerinde veya ağda meydana gelen olayları izleyip analiz ederek, bilginin gizliliğini, bütünlüğünü ve erişilebilirliğini bozmak ya da sistemin güvenlik mekanizmalarını aşmak için yapılan aktiviteler

olarak tanımlanan saldırı işaretlerini yorumlama sürecidir [9]. En basit tanımıyla, saldırı tespit işlemini yapmak için geliştirilen sistemlere ise "*saldırı tespit sistemleri* (STS)" denir.

STS'lerin sınıflandırılmaları için birçok farklı kriter kullanılabilir. Genel olarak kullanılan kriterler şöylece sıralanabilir:

- Saldırı tespit yöntemi,
- Mimari yapı,
- Korunan sistem türü,
- Veri işleme zamanı,
- Kullanılan bilgi kaynağı.

STS'lerin en çok kullanılan bu sınıflandırma kriterlerine göre sınıflandırılması Şekil 2'de verilmiştir. STS'ler, saldırıların tespitine yönelik olarak bilgi kaynaklarını analiz ve karşılaştırma yapmak için kullanırlar. Saldırı tespiti için temelde kullanılabilecek bilgi kaynakları, bilgisayar veya ağ paketlerinin dinlenmesinden elde edilebildiği gibi, kullanıcıların davranış örüntülerinden de elde edilebilir. STS açısından kullanılabilecek bilgi kaynakları "denetleme izi", "ağ paketleri" ve "uygulama kayıt (log)" dosyalarıdır.



Şekil 2. Saldırı tespit sistemlerinin sınıflandırılması

Denetim (hesap, günlük) izi; bilgi ve iletişim güvenliği için, sistem aktivitelerinin zamana göre sıralanmış şeklidir. Bu yapı, sistem üzerinde gerçekleşen olayların sıralaması bozulduğunda veya bu olaylarda değişiklikler meydana geldiğinde, sistemin yeniden yapılandırılmasında ve buradaki test işlemlerinin gerçekleştirilmesinde kullanılır. STS'ler denetleme izini, daha çok sistemde tanımlı bulunan kullanıcıların veya grupların hareket-davranış örüntülerini çıkarmak için kullanır. Günlük yapılan işler ve bu işlere yönelik olarak sistemdeki yetkiler göz önünde bulundurularak kullanıcı profilleri oluşturulur. Çeşitli analizlerle bu profillerin doğrulukları ispatlanır. STS'ler bu belirlenen profillere aykırı her türlü hareketi saldırı olarak algılar [9].

STS'ler açısından bir başka bilgi kaynağı yapısı olan ağ paketleri; koklayıcılar (sniffers) tarafından ağ trafiğinin dinlenmesiyle elde edilir. Bu günlükler daha çok hizmet aksattırma (DoS) saldırılarını tespit etmekte kullanılır. Ağ paketlerinin dinlenmesiyle elde edilen bilgiler sayesinde, sunucu tabanlı STS'lerden farklı olarak, ağ katmanında gerçekleşen saldırı olaylarını da tespit etmek mümkün olabilmektedir.

Uygulama log dosyaları; uygulama katmanında gerçekleşebilecek saldırıları tespit etmekte kullanılabilir. Bu veri kaynağı, diğer iki kaynaktan daha kolay elde edilebilmektedir ancak saldırı tespit oranı daha düşüktür. Ayrıca gerçek zamanlı bir saldırı tespiti yapılması gerekiyorsa, eş zamanlı olarak uygulama log dosyalarının hem oluşturulup hem analiz edilmesi gibi bir yapı pratikte pek mümkün olamayacağından bu bilgi kaynağı kullanılmaz. Daha çok saldırı ile analiz aşamasının farklı olduğu, gerçek zamanlı olmayan sistemlerde log kayıtlarının analizleri yapılır [9]. Bu çalışmamızda, log kayıtlarını kullanan, gerçek zamanlı olmayan bir saldırı analiz platformu ve istatistiksel çıkarım motoru olarak nitelenebilecek bir öneri sistemi geliştirilmiştir.

4. Materyal ve Metot

Web tabanlı uygulamaların ve web sitelerinin erişim bilgileri sunucu üzerinde bulunan erişim log dosyalarında tutulmaktadır. Oluşturulan her bir log dosyası sunucu tarafından otomatik olarak her gün için ayrı bir log dosyası olarak kaydedilmektedir. Web sitesine ait alt domainler mevcut ise sunucu tarafından her alt domain için ayrı klasörler oluşturularak erişim bilgileri bu klasörlerde tutulur. Ziyaretçilerin her bir erişimi log dosyasına yeni bir satır olarak eklenir. Eklenen her bir satır erişimle ilgili çeşitli bilgiler

tutmaktadır. Tutulan bilgi türleri kullanılan web sunucusuna ve kullanılan log formatına göre farklılık gösterebilir. Ayrıca sunucu üzerinde yapılan ayarlamalara göre tutulacak bilgi türü sayısı artırılabilir veya azaltılabilir. Şekil 2'de Windows Server 2003 işletimi sistemi üzerinde çalışan IIS 6.0 web sunucusunda tutulan log dosyasından örnek bir satır verilmiştir.

Tablo 1. Örnek erişim log satırı

2014-11-20 00:22:31 193.140.180.4 GET /Default.aspx - 80 - 212.154.80.164 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1; +SV1;+GTB6.4) - 200 0 0 67049 428 31

Tablo 1'de verilen web erişim log satırı, ilgili log dosyasındaki bir satıra karşılık gelmekte ve çeşitli alanlardan meydana gelmektedir. Bu alanlar farklı anlamlara gelmektedir. Erişim log dosyalarındaki bu alanlar ve anlamları Tablo 2'de verilmiştir.
Gerçekleştirilen sistem, Tablo 1'de gösterilen log dosyalarındaki satırları, Tablo 2'de gösterilen alanlara göre ayrıştırarak [10] saldırı tespiti ve çeşitli analizler gerçekleştirmektedir. Web erişim log dosyaları üzerinde çalışan sistem genel olarak şu yeteneklere sahiptir:

- Erişim sağlanan dosyaların bilgilerini, ne kadar erişim sağlandığı, kim tarafından erişim sağlandığı bilgisini kaydeder.
- Erişim sağlanan web sayfa ve uygulamalarının verilerini tutar.
- Web uygulamasına erişim sağlayan kullanıcıya ait IP bilgisini kaydeder.
- Web uygulamasına hangi internet tarayıcısından erişildiği bilgisini kaydeder.
- Web uygulamasına nereden referans alıp eriştiği bilgisini kaydeder.
- Web uygulamasına erişim metodunu kaydeder (GET, POST gibi).
- Web uygulamasını, kullanıcı IP numaralarına göre raporlar.
- Uygulama ile ilgili tüm bu bilgileri istatistiksel olarak kayıt altına alır.
- XSS, Brute Force, DoS, DDoS ve injection saldırıları ile ilgili tespit raporları oluşturur ve kaydeder.

Tablo 2. Erişim log alanlarının açıklaması

Alan Adı	Örnek Değer	Açıklama
date	2015-11-20	Aktivitenin meydana geldiği tarih
time	00:22:31	Aktivitenin meydana geldiği saat
c-ip	212.154.80.164	İstekte bulunan kullanıcının IP adresi
s-ip	193.140.180.4	Web sitesinin bulunduğu sunucunun IP adresi
cs-uri-stem	/Default.aspx	İstekte bulunulan web adresi
sc-status	200	İsteğe verilen cevabın durum kodunu içerir
cs(user-agent)	Mozilla/4.0+(compatible; +MSIE+6.0; +Windows+NT+5.1;)	İstemci tarafından kullanılan tarayıcının tipi ve diğer özellikler
cs-referrer	-	Aktif sayfaya hangi kaynaktan gelindiğini gösterir

Web logları üzerinde yapılan analizlerin önemli bir kısmını ön işleme adımı oluşturmaktadır. Log dosyalarında büyük miktarda içerdiği gereksiz bilgi, ön işleme aşamasını zorunlu bir hale getirmektedir. Ön işleme aşaması ile gereksiz bilgilerden temizlenen log dosyalarının yönetimi ve analizi kolaylaşmaktadır.

Ön işleme aşamasında veri temizleme, veri filtreleme ve birleştirme gibi tekniklerden yararlanılmaktadır. Web log dosyalarındaki ön işleme süreci tüm veri madenciliği sürecinin %80'ini kapsamaktadır. Önişleme adımı daha sonraki adımlar olan örüntü keşfi ve örüntü analizinin daha etkili ve kolay yapılmasını sağlamaktadır [11].

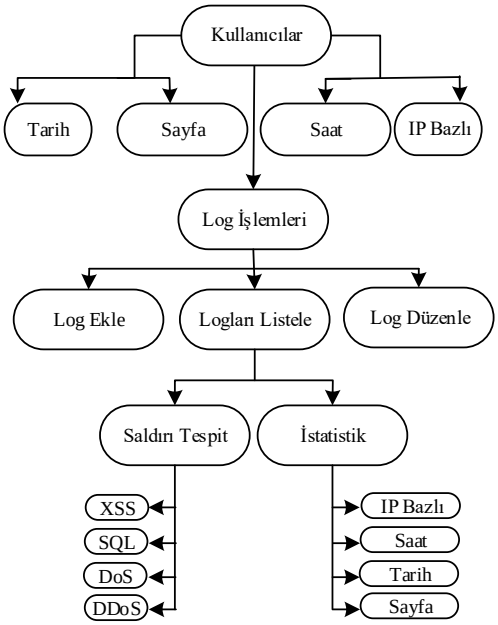
Literatürde yapılan birçok çalışmada veri temizleme algoritmasıyla analiz edilen verideki gereksiz olan kısımlar çıkarılmıştır. Log kayıtları Tablo 2'de verildiği üzere bölünmesi gereken çeşitli alanlar içermektedir. Log dosyasındaki tek bir satırdan alanların bölünerek çıkarılması işlemine "alan ayrıştırma" adı verilmektedir. Burada en çok kullanılan ayırıcı karakterler ',' veya 'boşluk' karakteridir. Aye yaptığı çalışmada bu işlemi gerçekleştirmek için ayrıştırma algoritması oluşturmuştur [12].

Bu çalışmada bu işlemi gerçekleştirmek için Şekil 4'te akış şeması verilen ayrıştırma yöntemi kullanılmıştır. Veri temizleme aşamasında ise analiz edilen log dosyasına özel olarak analiz amacı için önem arz etmeyen nesneler (stil dosyaları, grafik dosyaları, ses dosyaları vb.)

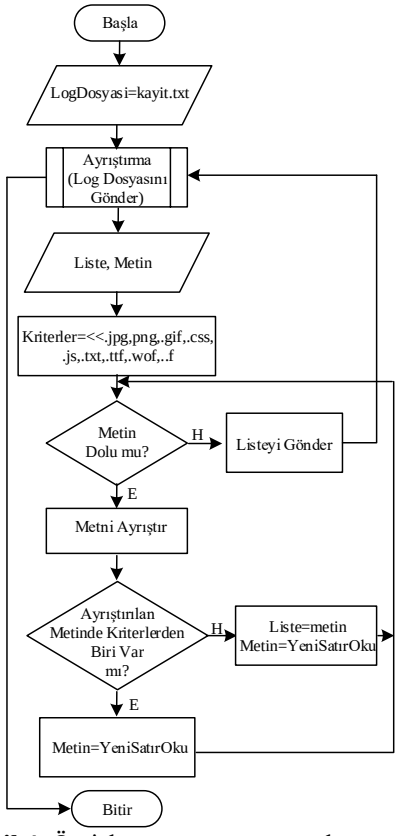
temizlenmektedir [12]. Bununla birlikte bazı özel resim sitelerinde veya haber sayfalarında kullanıcıların ilgisinin analiz edilmesi gibi bir durumda özellikle bu dosyalara odaklanmak da gerekebilmektedir. Burada önemli olan web log madenciliğinin amacının doğru olarak belirlenmesidir [13]. Bu çalışmada resim ve diğer bağlı olan dosyalar olmadan daha hızlı analiz etmek amacıyla .gif, .jpg, .css, .js, .png, javascript uzantılı dosyalar analiz edilmemiştir.

5. Uygulama Sonuçları

Geliştirilen web tabanlı log analizi platformu uygulaması ile web erişim logları üzerinde Şekil 3'te verilen işlemler yapılabilmektedir. Sistem üzerinde kullanım yetkisi bulunan her kullanıcı log dosyalarını analiz etmek üzere sisteme yükleyebilir, log analiz ve dosyalarını listeyebilir, düzenleyebilir. Listelediği log dosyaları üzerinde gerçekleştirdiği saldırı tespit ve istatistiksel analizleri görebilir. Log dosyalarından tespit edilebilecek olan Sql enjeksiyonu, XSS, DoS, DDoS, brute force, kırık link tespiti, RFI/LFI vb. kötücül atakların analiz sonuçlarını ve log dosyasını incelediği web sitesi ya da uygulamasının erişimine yönelik istatistiklere erişebilmektedir.



Şekil 3. Log dosyasının analizi

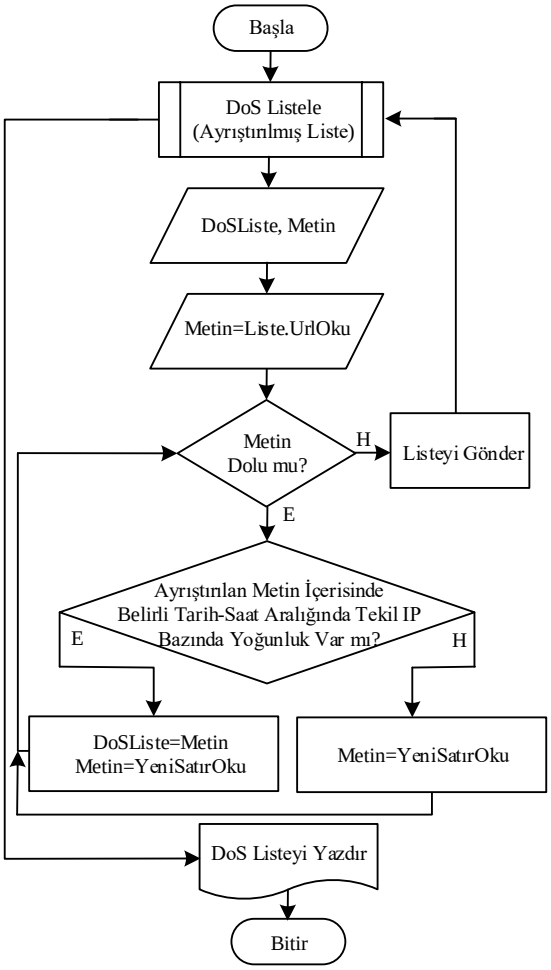


Şekil 4. Ön işleme ve ayrıştırma akış şeması

Şekil 5, DoS türü atakların tespiti için kullanılan akış şemasını göstermektedir. Log dosyası üzerinde yapılan analizler sonucunda, belirli bir zaman aralığında belirlenen eşik değerden fazla bir istek yapılması durumunda ilgili durum bir DoS atağı olarak değerlendirilmekte ve ilgili log satırı DoS atak listesine kaydedilmektedir. DoS atağı, hedef sistemin kaynaklarının tüketimine yönelik TCP/IP protokollerini istismar eden bir saldırı türüdür. Burada web erişim kayıtlarındaki belirli bir IP tarafından sıklıkla tekrar eden istekler değerlendirilmiştir.

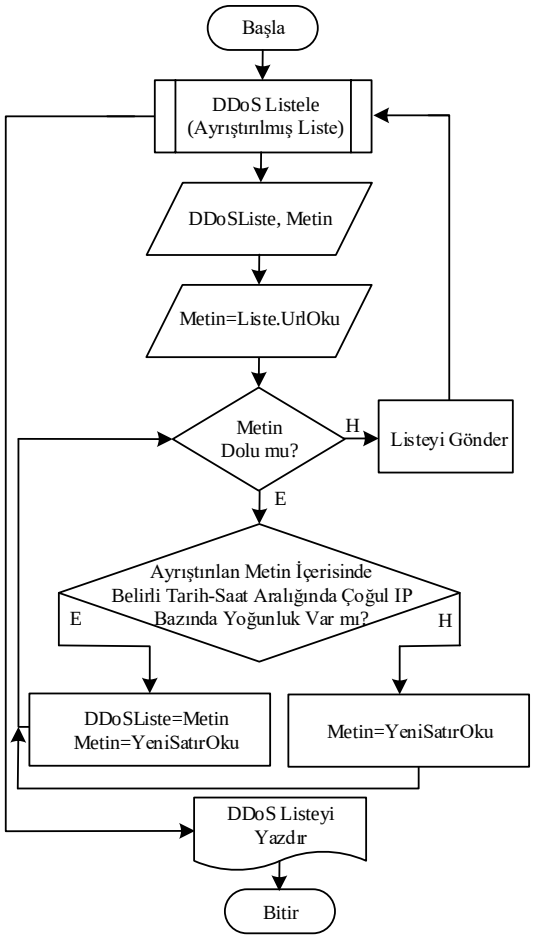
Şekil 6, DDoS türü atakların tespiti için kullanılan akış şemasını göstermektedir. Log dosyası üzerinde yapılan analizler sonucunda, belirli bir zaman aralığında birden fazla IP numarasından, belirlenen eşik değerden fazla bir istek yapılması durumunda ilgili durum bir DDoS atağı olarak değerlendirilmekte ve ilgili log satırı DDoS atak listesine kaydedilmektedir. DDoS atağı, hedef sistemin kaynaklarının tüketimine yönelik TCP/IP protokollerini istismar eden ve

birden fazla saldırganın eş zamanlı olarak gerçekleştirdiği bir saldırı türüdür. Burada web erişim kayıtlarındaki birçok IP tarafından sıklıkla tekrar eden belirli istekler değerlendirilmiştir.



Şekil 5. DoS atakları tespiti akış şeması

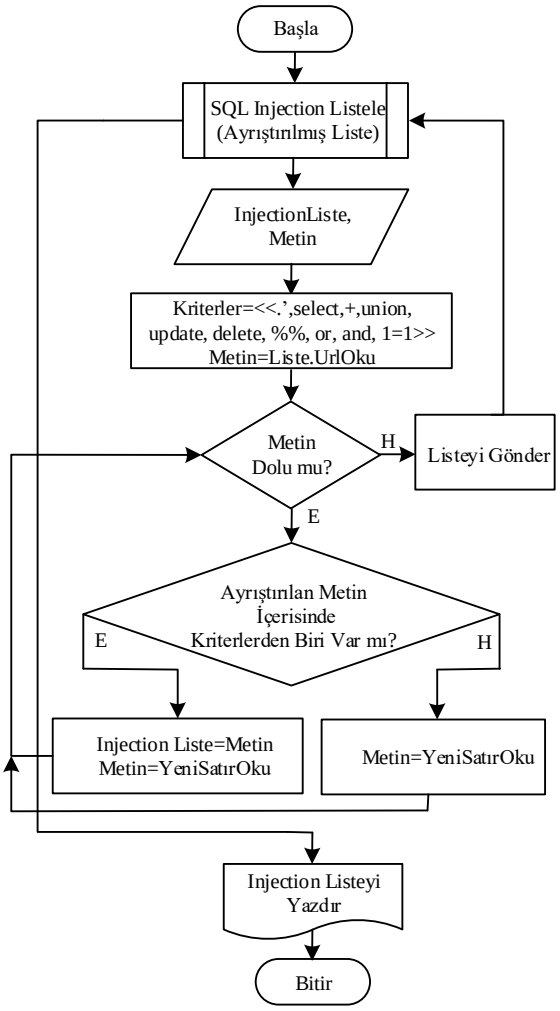
Günümüzde profesyonel olarak hizmet sunan web uygulamalarının neredeyse tamamı veri tabanlarını kullanmaktadır. Online çalışan bu web uygulamaları, veri tabanı ile yapısal bir sorgulama dili olan SQL aracılığıyla haberleşirler. SQL Enjeksiyonu, web uygulamalarından alınan kullanıcı girdileri ile oluşturulan SQL sorgularının manipülasyonu olarak tanımlanmaktadır [2, 9]. Kullanıcıların etkileşimde bulundukları veri tabanlı web uygulamalarında, sorgu veya parametreler kullanılarak veri tabanı tablolarındaki bilgiler belli şartlara göre filtrelenerek uygulama ara yüzüne aktarılır. Aktarılan bu sonuç değerleri, uygulamanın tasarımına göre kullanıcıya veya yöneticiye belli formatlarda sunulur. SQL enjeksiyonu yöntemi tam bu işlemler gerçekleştirilirken yapılır. Saldırgan, web tarayıcı adres çubuğuna veya uygulamada bulunan giriş kontrollerine kötücül kodlar ekleyerek, SQL enjeksiyon saldırısını gerçekleştirir.



Şekil 6. DDoS atakları tespiti akış şeması

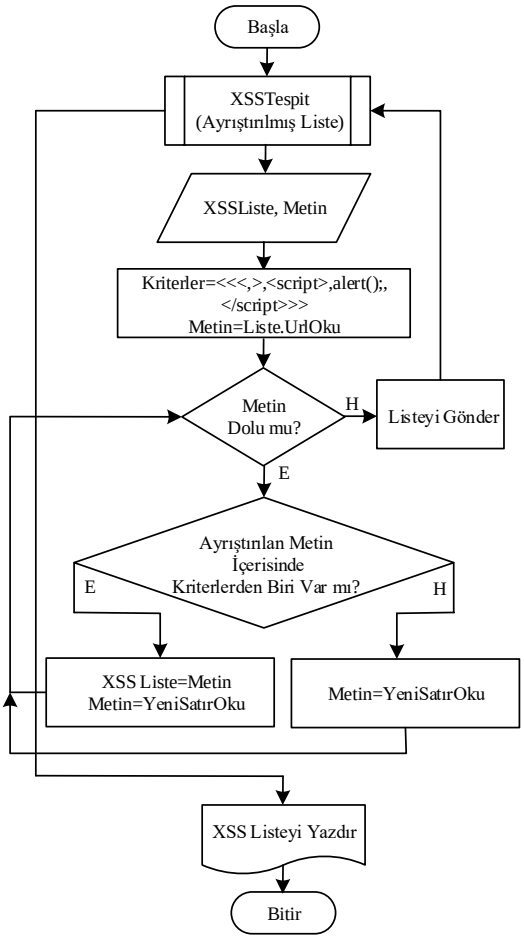
Genel kullanıma açık olmayan ancak bu şekilde elde edilen bilgiler önemli ve gizli olabilir. Saldırgan, sistem ve veri tabanı hakkında elde ettiği bu önemli bilgilerle SQL enjeksiyon senaryosuna farklı boyutlar kazandırarak, veri tabanında bulunan diğer bilgilere ulaşabilir. Sonrasında elde ettiği bilgileri kullanarak, hedefini gerçekleştirir. SQL enjeksiyonu ataklarında kullanılan anahtar kelimeler [2]'den yararlanılarak buradaki yazılımda tanımlanmış ve kullanılmıştır. Kural tabanlı saldırı tespit sistemi mantığında, ayrıştırılan log satırları içerisinde SQL enjeksiyonu anahtar kelimelerinden [2] herhangi birinin bulunması durumunda

enjeksiyon saldırısı tespit edilebilmektedir. Şekil 7'de SQL enjeksiyonu tespitinde kullanılan akış şeması verilmektedir.



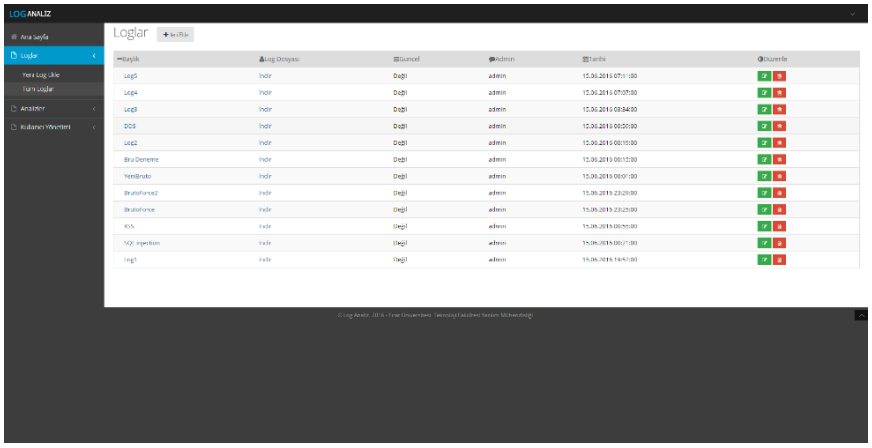
Şekil 7. SQL enjeksiyonu tespiti akış şeması

Çapraz site betikleme (XSS-Cross Site Scripting) saldırıları, web siteleri arasında kod yazma şeklinde gerçekleştirilen saldırılardır [9]. Bu saldırı, sitedeki kontrol yapısı içermeyen a lanlarda betik kodların çalıştırılmasıyla gerçekleştirilir. Bu betik, HTML veya Javascript kodlarından oluşabilir. Genel olarak Javascript kodlarıyla gerçekleştirilir. Ancak web tarayıcılar tarafından desteklenen VBScript, Ajax, ActiveX benzeri diller tarafından yazılan betiklerle de bu saldırılar yapılabilmektedir. Betik kodlar, HTML etiketleri arasında çalıştırıldığından XSS yerine HTML Enjeksiyonu ifadesi de kullanılmıştır [9, 14]. XSS saldırısı ile web siteleri oturum çerezleri ele geçirilebilir, web tarayıcısı istenen adrese yönlendirilebilir, DoS saldırıları yapılabilir. XSS saldırısında kurban, üzerinde XSS açığı bulunan web sitesine normal giriş yaptığında, oturum bilgileri normal olarak makinesine yüklenecektir. Bu sırada saldırgan, bir XSS betiğinin linkini e-mail vb. bir yolla yollamaktadır. Kurban, XSS betiğini çalıştırdığı anda sunucuyla etkileşime geçilip kod çalıştırılacak ve kurbanın oturum bilgileri çalınabilecektir. Şekil 8, XSS saldırılarının tespiti için kullanılan akış şeması gösterilmiştir. Burada [9]'da verilen XSS anahtar terimlerinden yararlanılarak yazılımda tanımlanmış ve kullanılmıştır.

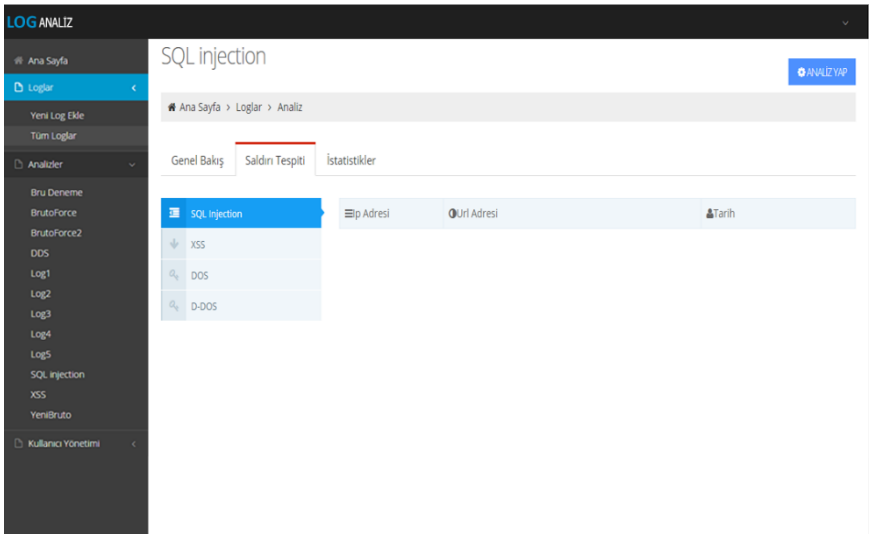


Şekil 8. XSS tespiti akış şeması

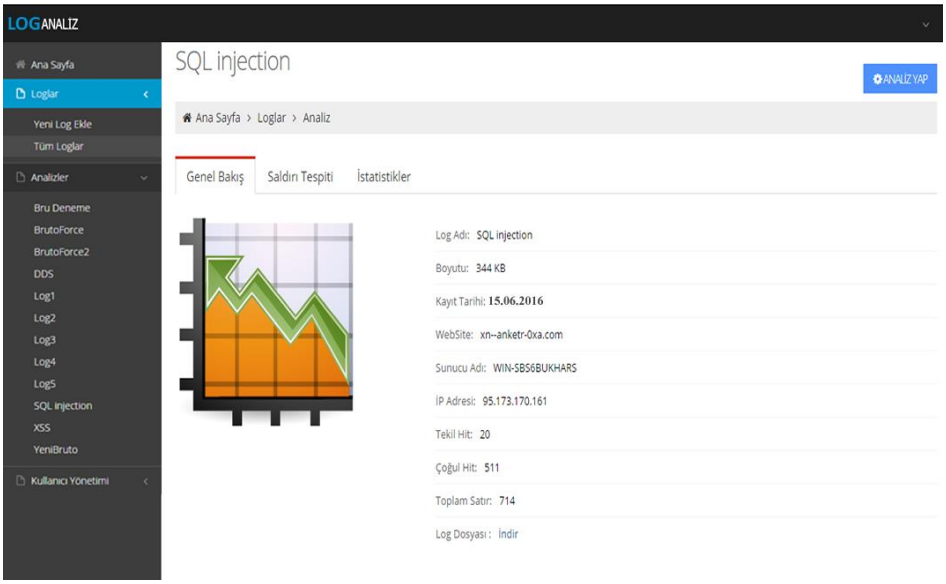
Uygulamaya ait ekran görüntüleri Şekil 9-12'de gösterilmiştir. Şekil 9'da log dosyaları sisteme yüklenmekte, Şekil 10'da analizi yapılan log dosyası saldırı tespiti için değerlendirilmekte ve Şekil 11 ve Şekil 12'de bu sonuçların rapor edilmesi gösterilmektedir.



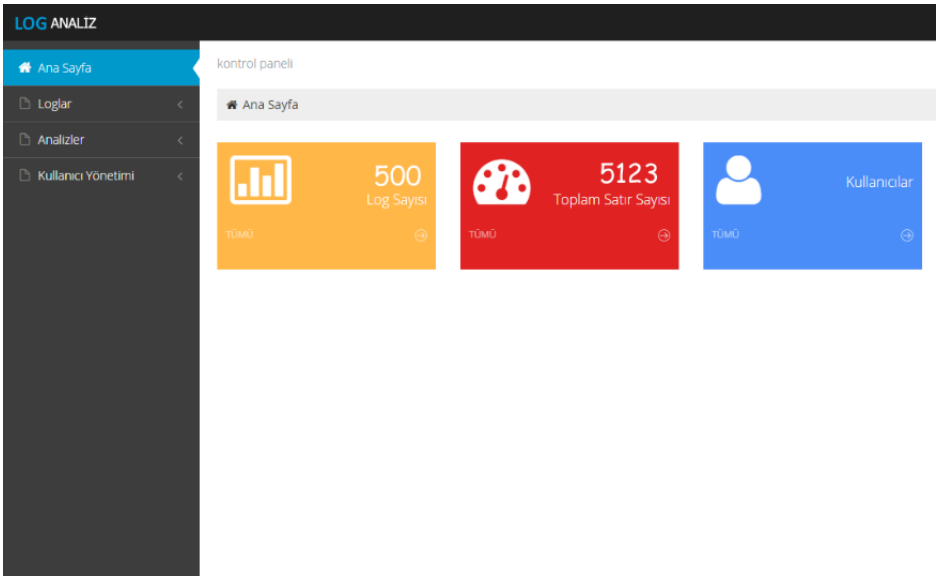
Şekil 9. Log dosyalarının sisteme yüklenmesi



Şekil 10. Log dosyalarından saldırı kayıtlarının tespit edilmesi



Şekil 11. Log dosyalarının saldırı türüne göre analiz edilmesi



Şekil 12. Log dosyalarının saldırı türüne göre analiz edilmesi

6. Sonuç ve Değerlendirme

Bilgi sistemlerinde artık birçok araç ve sistem kendi log kayıtlarını tutmaktadır. Hatta günümüzde CRM ya da ERP sistemlerinin bir çoğu kendi log kaydını tutmak zorundadır. Bu çalışmada bilgi sistemleri ve güvenliği açısından kullanılabilecek olan veri kaynaklarından biri olan log kayıtları üzerinde çeşitli incelemeler yapılmıştır. Bu incelemeler sonucunda log kayıtlarından tespit edilebilecek olan saldırı türleri belirlenmiş ve gerçekleştirilen log analiz platformu yazılımı ile log kayıtları analiz edilmiştir. Bu analizler sonucunda belirtilen saldırılar tespit edilmiş ve bu saldırıların sonuçları sistem yöneticisine rapor edilebilecek formata getirilmiştir. Geliştirilen sistem genel olarak web sunucu yönetim işlemlerini kolaylaştırmaktadır. Gerçekleştirilen sistem online olarak çalışan bir web sayfası üzerinden yüklenecek olan web erişim kayıtlarının analizini yapıp sonuçlarını raporlayacak niteliktedir.

Sunucu üzerinde tutulan erişim kayıtları herhangi bir metin editörü ile açılarak incelendiğinde herhangi bir anlam ifade etmeyen, karmaşık ve düzensiz bir yapıda olduğu görülecektir. Bu veriler web kullanım madenciliği ile analiz edilerek anlamlandırılmaktadır.

Bu çalışma ile log kayıtları ve türleri ile ilgili detaylı bilgi verildikten sonra web sunucu logları üzerinde web madenciliği konusunda yapılabilecekler hakkında literatürdeki çalışmalardan faydalanarak web kullanım madenciliğinin tüm süreçlerini içeren ve ilgili erişim kayıtlarından çeşitli istatistiksel bilgiler çıkaran bir yazılım tasarlanmıştır.

Ön işleme ve veri madenciliği teknikleri uygulandıktan sonra çıkan sonuçlarda herhangi bir anormallik olup olmadığını bulmak için kötüye kullanım tespiti yapılmıştır. Belirlenen kriterler dışındaki istekler tehdit olarak algılanıp tespit edilen saldırının türüne göre; enjeksiyon, hizmet engelleme, kaba kuvvet, dağıtık hizmet engelleme vb. kategorilerde sistem yöneticisine analiz sonuçları sunulmuştur.

7. Bilgilendirme

Bu çalışma, Fırat Üniversitesi, Bilimsel Araştırmalar Projeleri Biriminin (FÜBAP) TEKF.15.04 numaralı projesi ile desteklenmiştir.

8. Kaynaklar

1. Baykara, M., Daş, R., Karadogan, İ., "Bilgi Güvenliği Sistemlerinde Kullanılan Araçların İncelenmesi", 1st International Symposium on Digital Forensics and Security (1. Uluslararası Adli Bilişim ve Güvenlik Sempozyumu)", 231-239, 20-21 Mayıs 2013, Elazığ - Turkey.

2. Demirol, D., Daş, R., Baykara, M., " SQL Enjeksiyon Saldırı Uygulaması ve Güvenlik Önerileri", 1st International Symposium on Digital Forensics and Security (1. Uluslararası Adli Bilişim ve Güvenlik Sempozyumu)", 62-66, 20-21 Mayıs 2013, Elazığ – Turkey

3. Daş, R., Demirol, D., Tuna, G., "A Novel Software Tool for Mining Access Patterns Efficiently from Web User Access Logs", 2nd International Conference on Engineering and Natural Sciences (ICENS), pp.2836-2843. 24-28 Mayıs 2016, Sarajevo (Saraybosna), Bosnia and Herzegovina (Bosna ve Hersek)

4. A. Makanju, A. N. Zincir-Heywood and E. E. Milios, "A Lightweight Algorithm for Message Type Extraction in System Application Logs," in IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 11, pp. 1921-1936, Nov. 2012.

5. R. Vaarandi and M. Pihelgas, "Using Security Logs for Collecting and Reporting Technical Security Metrics," 2014 IEEE Military Communications Conference, Baltimore, MD, 2014, pp. 294-299.

6. Daş, R., Türkoğlu, İ. and Poyraz, M., Web Kayıt Dosyalarından İlginç Örüntülerin Keşfedilmesi, Fırat Üniv. Fen ve Müh. Bil. Dergisi 19 (4), 493-503, 2007.

7. Daş, R., Türkoğlu, İ. and Poyraz, M., Genetik Algoritma Yöntemiyle Internet Erişim Kayıtlarından Bilgi Çıkarılması, SAÜ Fen Bilimleri Enstitüsü Dergisi 10. Cilt, 2.Sayı, s. 67-72, 2006.

8. Daş, R. (2008). Web Kullanıcı Erişim Kütüklerinden Bilgi Çıkarımı, Doktora Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü.

9. Baykara, M. (2016). Bilişim Sistemleri İçin Saldırı Tespit ve Engelleme Yaklaşımlarının Tasarımı ve Gerçekleştirilmesi, Doktora Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü.

10. Özseven, T., Düğenci, M. Web Log Analizi İle Dışarıdan Verilen Kırık Link Tespiti, 15.Türkiye'de İnternet Konferansı, 2010

11. T. Hussain, S. Asghar and N. Masood, "Web usage mining: A survey on preprocessing of web log file," Information and Emerging Technologies (ICIET), 2010 International Conference on, Karachi, 2010, pp. 1-6.

12. T. T. Aye, "Web log cleaning for mining of web usage patterns," 3rd International Conference on Computer Research and Development (ICCRD), Shanghai, 2011, pp. 490-494.

13. M. Shu-yue, L. Wen-cai and W. Shuo, "The Study on the Preprocessing in Web Log Mining," Fourth International Symposium on Knowledge Acquisition and Modeling (KAM), Sanya, 2011, pp. 315-317.

14. Vural, Y., Sağıroğlu, Ş., "Kurumsal Bilgi Güvenliği ve Standartları üzerine bir İnceleme", Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, Cilt 23, No 2, s.507-522, Haziran 2008.