T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

WEB ERİŞİM KÜTÜKLERİNİN TEMİZLENMESİNE YÖNELİK YAZILIM GELİŞTİRME

YÜKSEK LİSANS TEZİ

Doygun DEMİROL
(112131108)

Anabilim Dalı: Elektronik ve Bilgisayar Eğitimi
Programı: Bilgisayar Sistemleri

Tez Danışmanı: Yrd. Doç. Dr. Resul DAŞ (F.Ü.)

2013

ÖNSÖZ

Bu yüksek lisans tez çalışmasında, web kullanım madenciliğine yönelik bir yazılım geliştirilmiştir. Geliştirilen yazılım ile web erişim kayıtları analiz edilmiş ve elde edilen veriler, veri tabanına aktarılmıştır. Veri tabanına aktarılan bu bilgilerle istatistiksel analizler ve saldırı tespiti yapılmıştır. Yapılan istatistiksel analizlerin olumlu sonuç verdiği gözlemlenmiştir.

Yüksek lisans tez çalışmam süresince benden maddi ve manevi desteklerini esirgemeyen aileme ve benden yardımlarını, desteğini, sabrını ve bilgisini esirgemeyen değerli hocam Yrd. Doç. Dr. Resul DAŞ'a teşekkürü bir borç bilirim.

Doygun DEMİROL

İÇİNDEKİLER

Sayfa

ÖZET

YÜKSEK LİSANS TEZİ

WEB ERİŞİM KÜTÜKLERİNİN TEMİZLENMESİNE YÖNELİK YAZILIM GELİŞTİRME

Doygun DEMİROL

Fırat Üniversitesi
Fen Bilimleri Enstitüsü
Elektronik ve Bilgisayar Eğitimi Anabilim Dalı
2013, Sayfa: 84

İnternet ve bilgisayar sistemleri başta olmak üzere birçok sistem, üzerinde gerçekleştirilen işlemleri kayıt altına alırlar. Bu işlem, bir uçağın kara kutusunun gerçekleştirdiği işleve benzemektedir. Kara kutu, uçakta olan bitenle ilgili her aktiviteyi nasıl kayıt altına alıyorsa, bilişim sistemlerinde kullanılan çoğu önemli cihaz da üzerinde gerçekleştirilen işlemlerle ilgili izleri kayıt altına almaktadır. Kayıt altına alınan bu bilgiler web tabanlı saldırıların tespitinde, adli olayların aydınlatılmasında, adli bilişim süreçlerinde, elektronik ticaret sitelerinde kullanıcıların davranışlarına göre ürün sayfalarının yeniden yapılandırılmasında, web uygulamalarının performanslarının arttırılması gibi birçok konuda önemli bilgilerin çıkarılması için kullanılabilmektedir.

Web sunucularında üretilen ve saklanan kütük dosyaları, sunucular üzerinde bulunan web sitelerine ait etkinliklerin takip edilmesine ve bu sitelerin çeşitli yönlerden analiz edilmesine yönelik birçok önemli veriler ihtiva etmektedir. Bu dosyaların boyutu sunucu trafiğinin yoğunluğuna göre farklılık gösterebilmektedir. Bu veriler, web sunucusu üzerindeki web sitelerini ziyaret eden kullanıcıların, web sayfalarına erişirken bıraktıkları örüntülerin yanı sıra, siteye üye olurken kullanıcının web formuna girmiş olduğu bilgileri ve arama motorlarına ait botların web sayfalarında dolaşırken bıraktıkları izlerden oluşmaktadır. Karmaşık, anlamsız ve boyutu büyük olabilen bu dosyaların içeriklerinde bulunan kayıt satırlarından anlamlı verilerin çıkarılabilmesi, web sitelerinin analiz edilmesi web madenciliği ile gerçekleştirilmektedir. Metin tabanlı bu karmaşık verilere, veri

madenciliği yöntemleri uygulanarak web sitesi yöneticilerine, sitenin geliştirilebilmesi ve etkinliğinin yükseltilmesi için birçok yararlı ve anlamlı bilgiler sunulabilmektedir.

Bu tez çalışmasının genel amacı, web sunucuları tarafından farklı biçimlerde saklanan web erişim kütüklerinin temizlenmesine yönelik uygun bir yazılım geliştirmektir. Ayrıca, geliştirilen bu yazılım ile temizlenen erişim kütüklerinden anlamlı bilgilerin elde edilmesine yönelik örüntü keşfi ve örüntü analizi yapılmıştır. Bu kapsamda, önemli istatistiki bilgiler ve saldırı tespiti çalışmaları yapılmıştır.

Anahtar Kelimeler: Web madenciliği, Web kullanım madenciliği, Kütük analizi, Web erişim kütükleri, Veri madenciliği, Bilgi aktarımı.

SUMMARY

Master Thesis

SOFTWARE DEVELOPMENT FOR THE REMOVAL OF THE WEB ACCESS LOGS

Doygun DEMİROL

Firat University
Graduate School of Natural and Applied Sciences
Department of Electronics and Computer Education
2013, Page: 84

Internet and computer systems, should be first to take into consideration, because recording all the operations received and issued. This process is similar to the one performed by the black box of a plane. As the black box of the plane saves all activities of each section, a computing device records all the operations performed by a system. By the help of this System, many web based attacks were discovered and eased the clarification on the e-commerce sites, so using an application like the presented one, will lead to an increased performance of the web application and the Access of the customers will be restricted for the reasons mentioned above.

Produced and stored in the files of the web servers, the sites contain many important data that can be followed and analyzed. The size of these files may vary depending on the density of the traffic of the server. The data may contain a list of the users who visited the web site, as well as patterns of their access on the web pages or the information introduced by them in the search engine. All these data can be complex, meaningless and the contents of the files can be large. So, a significant amount of data can be analysis with web usage mining. Applying the web usage mining phases to log files provides to web site managers meaningful and useful information.

The overall objective of this thesis is to understand how we can clean the access logs from the servers. Excepting this, we can discover how we can Access important patterns

of the information and how to analyze them. In this context, we should analyze important statistical information and intrusion detections.

Keywords: Web mining, Web usage mining, Log analysis, Web access logs, Data mining, Information extraction.

ŞEKİLLER LİSTESİ

Sayfa

TABLOLAR LİSTESİ

Sayfa

KISALTMALAR LİSTESİ

HTTP	**:**	Hypertext Transfer Protocol
CERT	**:**	Computer Emergency Response Team
IIS	**:**	Internet Information Service
SQL	**:**	Structured Query Language
NIST	**:**	National Institute of Standards and Technology
PCI DSS	**:**	Payment Card Industry, Data Security Standards
FISMA	**:**	Federal Information Security Management Act
HIPAA	**:**	Health Insurance Portability and Accountability Act
SOX	**:**	Sarbanes-Oxley
ISO	**:**	International Organization for Standardization
OSI	**:**	Open System Interconnection
XSS	**:**	Cross Site Scripting
REGEX	**:**	Regular Expression
IP	**:**	Internet Protocol
ANSI	**:**	American National Standards Institute
FTP	**:**	File Transfer Protocol
URL	**:**	Uniform Resource Locator
CLF	**:**	Common Log Format
ECLF	**:**	Extended Common Log Format
GMT	**:**	Greenwich Mean Time
SMTP	**:**	Simple Mail Transfer Protocol
MTA	**:**	Mail Transfer Agent
POP3	**:**	Post Office Protocol 3
IMAP	**:**	Internet Message Access Protocol
VTYS	**:**	Veri Tabanı Yönetim Sistemi
DLP	**:**	Data Loss Prevention
YSA	**:**	Yapay Sinir Ağları
OLAP	**:**	On-Line Analytical Processing
ODBC	**:**	Open Database Connectivity

1. GİRİŞ

Web sunucuları tarafından üretilen ve saklanan erişim kütük dosyaları, sunucu üzerinde bulunan web sitelerine ait etkinliklerin takip edilmesine ve bu sitelerin çeşitli yönlerden analiz edilmesine yönelik birçok önemli veri ihtiva etmektedirler. Bu dosyaların boyutları sunucu trafiğinin yoğunluğuna göre farklılık gösterebilmektedir. Bu veriler, web sunucusu üzerindeki web sitelerini ziyaret eden kullanıcıların, web sayfalarına erişirken bıraktıkları örüntülerin yanı sıra siteye üye olurken kullanıcının web formuna girmiş olduğu bilgilerden ve arama motorlarına ait botların web sayfalarında dolaşırken bıraktıkları izlerden oluşmaktadır. Karmaşık, anlamsız ve boyutu büyük olabilen bu dosyaların içeriklerinde bulunan kayıt satırlarından anlamlı verilerin çıkarılabilmesi, web sitelerinin analiz edilmesi web madenciliği ile mümkündür. Metin tabanlı bu karmaşık verilere, veri madenciliği yöntemleri uygulanarak web sitesi yöneticilerine, sitenin geliştirilebilmesi ve etkinliğinin yükseltilmesine yönelik yararlı ve anlamlı bilgiler sunulabilmektedir. Ayrıca, e-ticaret hizmeti veren web sitelerinin kütük kayıtları analiz edilerek, elde edilen kar miktarları ve web sitesinin popülaritesi arttırılabilir. Çevrimiçi eğitim hizmeti veren internet sitelerinde öğrenci erişim izleri analiz edilerek en yararlı ders materyallerinin seçimi, eğitim başarısını pozitif yönde etkileyebilecek kararların alınması gibi farklı ilgi alanlarında kararlar alınabilir.

1.1 Tez Çalışmasının Amacı

Bu tez çalışmasının genel amacı, web sunucuları tarafından farklı biçimlerde saklanan web erişim kütüklerinin temizlenmesine yönelik uygun bir yazılım geliştirmektir. Ayrıca, geliştirilen bu yazılım ile temizlenen erişim kütüklerinden anlamlı bilgilerin elde edilmesine yönelik örüntü keşfi ve örüntü analizi yapılmıştır. Bu kapsamda, önemli istatistiki bilgiler ve saldırı tespiti çalışmaları yapılmıştır.

Erişim kayıtlarından, web madenciliği yöntemleri kullanılarak elde edilen anlamlı bilgiler, web sitelerine erişen kullanıcıların davranışları hakkında bilgi edinilmesini ve elde edilen bu bilgilerle kullanıcı davranışlarının tahmin edilmesine olanak sağlamaktadır. Bunun yanı sıra tasarımcılar için site tasarımının düzenlenmesine, iyileştirilmesine ve performansının arttırılmasına yönelik bilgiler de elde edilebilmektedir. Erişim kayıtlarındaki HTTP durum kodlarıyla, web sitesinde bulunan problemlerin giderilmesi, web saldırını tespit etme gibi önemli bilgiler de çıkarılabilir.

1.2 Literatür Taraması ve Değerlendirilmesi

Bilişim sistemlerinde kullanılan birçok uygulama ve sistem, üzerlerinde yapılan işlemleri sistem yöneticilerini bilgilendirmek amacıyla farklı formatlarda kütük dosyası olarak kaydederler. Kaydedilen bu işlem kayıtlarına kütük denmektedir. Literatürde kütük kayıtları üzerine birçok akademik çalışma yer almaktadır. CERT (Computer Emergency Response Team)'in yayınlamış olduğu çalışmada; kütük, bilgisayar sistemleri üzerinde gerçekleştirilen eylem ve olayların kayıtları olarak tanımlanmıştır. Çalışmada, ayrıca Windows ve Linux Web sunucularında, Oracle ve Microsoft SQL Server 2005 veri tabanı yönetim sistemlerinde kütük tutma ve kütük yönetimi hakkında detaylı bilgilere yer verilmiştir [1]. Grace ve arkadaşları log dosyalarını, sistemde meydana gelen olayların listesini içeren dosyalar olarak tanımlamışlardır. Çalışmada log dosyalarından, log dosyalarının oluşturulmasından, formatlarından, kullanımlarından ve web madenciliği sürecinde kullanılabilecek çeşitli algoritmalar hakkında detaylıca bilgi verilmiştir [2]. NIST (National Institute of Standards and Technology) yayınladığı çalışmada log tanımını, bir kurum ya da kuruluşun sistem ya da ağında meydana gelen olayların kaydı olarak tanımlamıştır. Çalışmada, kuruluşlar için başta kütük yönetimi olmak üzere bilgisayar güvenliğine yönelik bilgiler de verilmiş ve kütük yönetimi altyapısına, planlanmasına, operasyonel süreçlerine ve kütük kayıtlarının analizine dair detaylı bilgilere yer verilmiştir [3]. Şahinaslan ve çalışma arkadaşları çalışmalarında log yönetiminin önemini ve gerekliliğini belirterek, kurumlarda log yönetiminin etkin bir şekilde nasıl sağlanabileceğine dair bilgiler vermişlerdir [4]. Tyagi ve arkadaşları yaptıkları çalışmada web sunucu log kayıtlarını toplamış ve bu kayıtları analiz etmişlerdir. Analiz sonucunda web sitesi erişim kayıtlarının ayrıntılı istatistiklerini çıkarmış ve web kullanım madenciliği tekniklerini kullanarak analiz ettikleri log dosyalarından, web sitesi yöneticisi ve web tasarımcısının kullanabileceği anlamlı bilgileri çıkarmışlardır. Çıkarılan anlamlı bilgilerle web sitesinin etkinliğini ve başarımını arttırmışlardır [5]. Vellingiri ve Pandian web kullanım madenciliğinin ilk aşaması olan ön işlem aşamasında uygulanan veri temizleme ve işlem tanımlama adımları için yeni ve etkili teknikler sunmuşlardır. Bu teknikler sonucunda veri temizleme ve işlem tanımlama aşamaları diğer yöntemlere göre daha hızlı sonuç vermiştir [6]. Pamutha ve diğ. çalışmalarında web madenciliği ön işlem aşaması üzerinde durmuşlardır. Çalışma sonucunda kullanıcı oturumlarına ait benzersiz IP adresleri, benzersiz sayfa erişimleri, kullanıcılara ait oturum süreleri, ziyaret edilen sayfalara ait frekans değerleri gibi istatistiksel sonuçlara ulaşmışlardır [7]. Daş ve diğ.

yaptıkları akademik çalışmalarında e-öğrenme üzerine geliştirilen web tabanlı öğretim materyallerinin web kullanım madenciliği ile analiz edilebilmesi ve değerlendirilebilmesi için uygulama aşamalarını sunmuşlardır. Ayrıca, web tabanlı öğretim materyallerinin web kullanım madenciliği ile analiz edilmesine ilişkin önerilerde bulunmuşlardır [8]. Daş ve arkadaşlarının yaptıkları bir başka çalışmada Fırat Üniversitesi'ne ait kullanıcı erişim kayıtlarını analiz ederek web site yöneticisi ve tasarımcısı için kullanışlı bilgiler bulmuşlardır. Elde edilen bu bilgiler, sistem performansının arttırılmasına yönelik olumlu bilgilerdir [9,10]. Daş ve diğ. yapmış oldukları bir başka çalışmada site içerisindeki bağlantıları kullanarak yol analizi yöntemini yardımıyla web sitesinin etkinliğinin arttırılmasını sağlamışlardır [11]. Daş ve diğ. bir başka çalışmada ise birliktelik kuralını kullanarak web sitesinin etkinliğini arttırmaya yönelik olumlu bilgilere ulaşmışlardır. Benzer bir çalışmalarında ise kullanıcı erişim kayıt dosyasındaki ham veriler düzenleyerek, genetik algoritma yöntemi ile bu verilerden istatistiksel bilgi çıkarımı yapmışlardır. Yapılan çalışma sonucunda, internet kullanıcılarının en fazla kullandığı veri tabanı adres bilgisini tespit etmişlerdir [12, 13]. Shenoy yapmış olduğu tez çalışmasında, vekil sunucularının performansını artırmak için web sayfalarını önbelleğe alma ve web sayfalarını önceden yüklenmesine (pre-fetching) yönelik bir yaklaşım sunmuştur [14]. Salama ve arkadaşları, farklı formatlardaki web erişim kayıtlarından web saldırılarını tespit etmek ve izlemek için bu dosyaların tek bir XML formatına çevrilmesini sunmuşlardır [15]. Sumathi ve diğ. yaptıkları çalışmalarında kullanıcı gezinme örüntülerini kullanarak, kullanıcının mevcut ihtiyaçlarını karşılamaya yönelik uygun öneriler sağlamışlardır. Bir ticari web sitesinden sağladıkları gerçek kullanım verileri üzerinde yaptıkları deney sonrasında elde ettikleri sonuçlar, önerilen sistemin önerisi etkinliğini göstermiştir [16]. Romero ve arkadaşları yapmış oldukları çalışmalarında, Eindhoven Üniversitesi'nden elde ettikleri verileri kullanarak web sitelerini kişiselleştirmişlerdir. Üniversite öğrencilerinin kullanmış oldukları sisteme entegre edilmiş olan bu çalışma sayesinde, öğrenciler için en uygun içerik bağlantı/sayfaları tavsiye edilmiştir [17].

2. KÜTÜK DOSYALARI

2.1 Giriş

Kütük bir sistem üzerinde gerçekleşen bütün aktivitelerin düz metin formatında saklandığı dosyalardır. Web sunucuları, istemcilerden gelen istekleri bir kütük dosyasında saklarlar. İsteklerinin saklandığı bu dosyalarda istemcilere ait olan IP adresleri, erişim tarih zaman bilgileri, erişilen web adresi bilgisi, kullanılan port numarası, durum kodları gibi önemli bilgiler yer almaktadır. Bu önemli bilgiler sistem yöneticileri için çok önemlidir. Önceleri kütük kayıtları, bir yığın düz yazı olarak değerlendirilmekte ve çoğunlukla üstünde bulunduğu sistemle ilgili bir problem olduğu zamanlarda kullanılmaktaydı. Fakat günümüzde bu dosyaların içerdiği bilgiler daha detaylı incelemelerle daha etkin kullanılabilmektedir. Örneğin bir üniversite otomasyonuna, bir elektronik alışveriş sitesine ya da benzeri bir otomasyona yapılan saldırıların tespiti bu dosyaların içerdiği bilgiler sayesinde yapılabilmektedir. Bunun yanı sıra saldırılara senkronize olarak saldırı tespitleri de yapılabilmektedir. Kütük dosyaları sadece saldırı tespitinde veya adli süreçlerde kullanılmamaktadır. Bu dosyalardaki bilgiler, analizler yapılarak, web uygulamalarının tasarımlarını geliştirme, uygulamada kullanılan kodların optimizasyonu, uygulama üzerindeki hata oranının azaltılması, üye sistemi olan uygulamalarda üyelerin daha çok nelerle ilgilendiklerinin bulunması ile üyelere özel reklamların veya önerilerin sunulması gibi birçok alanda kullanılabilmektedir. İnternet bankacılığı hizmetini müşterilerine sunan banka yetkilileri, müşteri davranışları hakkında elde edilebilecek önemli özel bilgilerle ticari kazanımlarına katkı sağlayabilirler. Erişim kayıtları analiz edilerek, bu tarz hizmetlere ait birçok sayısal veriler elde edilebilir. Elde edilen bu sayısal verilerle, kullanıcılarla ilgili farklı bilgilere yönelme, kullanıcı davranışlarıyla ile ilgili tahminlerde bulunma ve benzeri araştırmalar yapma gibi konularda kolaylık sağlayarak, sunulan hizmetin kalitesi daha çok arttırılabilir.

Toplumun her kesiminde internet kullanımının yaygınlaşması, internetin sektörel olarak genişlemesini sağlamıştır. Örnek olarak elektronik haberleşme, bilgi tarama, finans, elektronik ticaret vb. örnekler verilebilir. İnternete bağlı kişisel bilgisayarların kullanımının artması, bu sektörlerin hızla gelişmesini ve bunlara bağlı olarak alt sektörlerin oluşmasını sağlamıştır. Bu sektörlere yapılan yatırımlar, hem alt sektörlerin gelişimini hızlandırmış hem de istihdamı arttırıcı bir etki meydana getirmiştir. Bütün bunlara paralel olarak internet ve web sayfaları üzerindeki verilerin hacmi ve karmaşıklığı da gün geçtikçe hızla

artmıştır. Artan bu verilerin düzenli olarak saklanması ve analiz edilmesi, sistem yöneticileri ve firma sahipleri açısından artık bir gereksinim ve zorunluluk haline gelmiştir [18]. Bu zorunluluk ve gereksinim bazı standartlar ve kanunlar tarafından da desteklenmiştir [19]. Log toplama konusuna kanun ve standartların ne kadar önem verdiğine örnek olarak PCI veri güvenliği standardını verebiliriz. PCI DSS (Payment Card Industry, Data Security Standarts) altı ana madde altında on iki gereksinimden söz etmektedir. Bunlardan biri de log toplama ile alakalıdır. Bunun yanı sıra T.C. 5651 sayılı kanunu, 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.

2.2 Erişim Kütüklerinden Web Tabanlı Saldırıların Tespit Edilmesi

Web tabanlı saldırıların tespit edilmesinde erişim kayıtlarının sağladığı en büyük avantaj, kolayca analiz edilebilir olmasıdır. IIS ve Apache sunucuları varsayılan olarak erişim kayıtlarını kaydederler. Ağ trafiği, erişim kayıtlarına göre daha fazla bilgi içermelerine rağmen ağ trafiği içeriklerini toplamak ve işlemek daha maliyetlidir. Ağ trafiğini dinlemek, tüm ağ paketlerini toplamayı ve ekstra donanım yükü gerektirmektedir. Ağ trafiğinden elde edilen veriler toplandıktan sonra verilerin uygun bir formata çevrilmesi gerekmektedir. Ağ trafiği verileri ancak bu şekilde erişim kayıtlarıyla aynı biçime dönüştürülerek analiz edilebilir [20].

Web uygulamaları OSI referans modelinin 7. katmanı olan uygulama katmanında çalışmaktadırlar. Web uygulamalarına yönelik saldırıları tespit etmek için web saldırı tespit mekanizmalarının uygulama katmanında çalışması gerekmekte ve sunucu üzerindeki trafiği tamamen izlemesi gerekmektedir.

Erişim kayıtları, ağ üzerindeki trafiğe ait bilgilerden sadece bir kısmını içermektedir. Tablo-1, erişim kayıtları ile ağ üzerindeki tüm trafiğin analizi arasındaki önemli farkları göstermektedir.

Web tabanlı saldırıları tespit etmek için iki farklı strateji kullanılmaktadır. Bunlar; kural tabanlı saldırı tespiti ve anomali tabanlı saldırı tespit stratejileridir.

Tablo 1: Erişim kütükleri verileri ile ağ trafiği verilerinin karşılaştırılması [20].

	Avantajlar	Dezavantajlar
Web Erişim Kütükleri	- Erişim kayıtlarındaki veriler kolayca elde edilerek kullanılabilirler.	- Erişim kayıtları, genellikle tüm ağ trafiğinden elde edilecek verilerden sadece belli bir kısmını içermektedir.
Tüm Ağ Trafiği	- Tüm bilgiler analiz edilebilir.	- Ağ trafiğinin içerdiği verilerin ilk önce toplanması gerekmektedir. - Verilerin birleştirilmesi, normalize edilmesi vb. gerekebilir. - Yüksek trafik yükünü toplamak zor alabilir. - Şifreli trafiği çözmek zor olabilir.

2.2.1 Kural Tabanlı Saldırı Tespiti

Kural tabanlı saldırı tespit stratejilerinde kurallar, statik olarak analiz öncesinde tanımlanmış olmalıdır. Bunlar belli karakterlerin tespit edilmesi gibi basit kurallar olabildiği gibi oturum sabitleme saldırıları gibi daha karmaşık kurallar olabilmektedir. Statik kurallar saldırı tespitinden önce bir kez tanımlanır ve saldırı tespit aşaması boyunca aynen kalırlar, değişmezler. Statik kurallar her uygulama için özel olarak hazırlanmış olmalıdır. Statik kuralları belirlemek için en mantıklı yol, önceden bilinen giriş karakterlerinin tanımlanması, sabit uzunluktaki parametrelerin tanımlanması ve üst limitlerin belirlenmesidir. Statik kurallar ayrıca negatif ve pozitif olarak iki ayrı algılama modeline ayrılmaktadır.

Negatif algılama modelindeki (kara liste yaklaşımı) davranış biçimi gelen her isteğin kabul edilmesi kuralına göre işlemektedir. Hangi isteklerin kabul edilmeyeceği model içinde önceden tanımlanmıştır. Model, gelen istekleri kara listedeki verilerle karşılaştırır. Eşleşmeye uyan istekleri kabul etmez, eşleşmeye uymayan istekleri ise kabul eder. Bu modelin uygulanması kolay olmakla beraber güvenlik yaklaşımlarına göre mantıklı değildir. Bu modelin en büyük dezavantajı, güvenliğin sadece belirlenen kurallar

çerçevesinde sağlanabilmesidir. En önemli avantajlarından biri ise hata oranının çok az olmasıdır [20].

Pozitif algılama modelinde ise negatif modeldekinin aksine bir yaklaşım kullanılmaktadır. Bu modelde öncelikle gelen isteklerin hiçbiri varsayılan olarak kabul edilmez. Negatif modelin aksine kabul edilecek isteklerin özellik listesi (beyaz liste) göz önüne alınır ve gelen istekler bu listedeki nesnelerle karşılaştırılır. Eşleşmeye uymayan istekler saldırı olarak kabul edilir. Eşleşmeye uyan istekler ise normal istek olarak tanımlanır ve erişime kabul edilir. Beyaz liste otomatik olarak öğrenme aşamasında ya da manuel olarak yönetici tarafından tanımlanabilmektedir. Güvenlik yaklaşımlarına göre bu algılama modeli, negatif modele göre daha çok kullanılabilmektedir. Güvenlik duvarları genellikle bu modele göre yapılandırılmaktadır [20].

2.2.2 Anomali Tabanlı Saldırı Tespiti

Anomali tabanlı saldırı tespit sistemleri, dinamik kuralları da içermektedir. Adından da anlaşılacağı üzere bu kurallar statik değildir ve manuel olarak tanımlanmamaktadır. Anomali tabanlı saldırı tespit sistemlerinde kurallar bir öğrenme aşaması ile tanımlanmaktadır. Öğrenme aşamasında, normal ağ trafiği esas alınmaktadır. Temel olarak esas alınacak trafiğin saldırılardan arındırılmış ve temiz olması, öğrenme aşamasında büyük önem taşımaktadır. Öğrenme aşamasının amacı, normal bir ağ trafiği akış setinin oluşturulmasıdır. Bu setteki davranışlara uymayan web istekleri saldırı olarak işaretlenmekte ve saldırılar bu şekilde engellenmektedir [20].

2.2.3 Çapraz Site Betik Saldırısı ve Tespiti

Çapraz site betik saldırısı (Cross Site Scripting), genellikle XSS olarak bilinen ve HTML enjeksiyon yönteminin bir alt kümesidir. XSS zararlı bir web uygulaması güvenlik sorunudur. XSS ile saldırgan, kurbanın tarayıcısında betik (script) çalıştırmaya izin verebilir, kullanıcının oturum bilgilerini çalabilir, web sitesine zarar verebilir, siteye zararlı içerikler girebilir ve kötü niyetli yazılım çalıştırılabilir [21].

Basit bir XSS saldırısı <h1> veya <script> gibi HTML etiketlerini içerir. En sık kullanılan örneklerden biri "<script>alert('XSS')</script>" şeklindedir. XSS açıklarını tespit etmenin en basit yolu HTML etiketlerinin incelenmesidir [20]. Tablo 2'de gösterilen

düzenli ifade, (Regular Expression - REGEX) HTML etiketlerini taramak için kullanılabilir. Tablo-3'de, Tablo-2'de gösterilen düzenli ifadenin açıklamaları yer almaktadır.

Tablo 2: XSS saldırısı için düzenli ifade.

```
/((\%3C)|<)((\%2F)|\/)*[az09\%]+((\%3E)|>)/ix
```

Tablo 3: Düzenli ifadenin açıklaması.

((\%3C)|<)	Küçüktür işareti kontrolü veya hex eşdeğeri
((\%2F)|\/)*	Kapalı etiket kontrolü veya hex eşdeğeri
[a-z0-9\%]+	Etiketler içinde alfanümerik ve % karakteri kontrolü veya hex eşdeğeri
((\%3E)|>)	Büyüktür işareti kontrolü veya hex eşdeğeri

Tablo-2'deki düzenli ifadenin sonunda bulunan 'i' ve 'x' karakterleri, eşleştirmede büyük-küçük harf duyarlılığının olmamasını ve boşlukların göz ardı edilmesini sağlamaktadır.

XSS saldırısı için birçok HTML etiketi kullanılabilmektedir. Örneğin, JavaScript, vbscript, applet, meta, xml, img, title, base, iframe gibi HTML etiketleri XSS saldırısında kullanılabilmektedir. Web saldırılarının erişim kayıtlarından tespit edilebilmesi için XSS saldırısının uygulanabileceği tüm HTML etiketleri için bir düzenli ifade tanımlanması gerekmektedir. Örnek olarak Tablo-4'de 'img' etiketi için bir düzenli ifade gösterilmiştir. Düzenli ifadeye ait açıklamalar ise Tablo-5'de sunulmuştur.

Tablo 4: img etiket için tanımlanan düzenli ifade.

```
/((\%3C)|<)((\%69)|i|(\%49))((\%6D)|m|(\%4D))((\%67)|g|(\%47))[^\n]+((\%3E)|>)/I
```

Tablo 5: Düzenli ifadenin açıklanması.

(\%3C)|<)	Küçüktür işareti kontrolü veya hex eşdeğeri
(\%69)|i|(\%49)) ((\%6D)|m|(\%4D)) ((\%67)|g|(\%47)	'img' etiketinin tanımlanması veya büyük küçük harf duyarlılığının olmaması ve hex eşdeğeri
[^\n]+	Mevcut satırda '<img' karakterlerinin kontrolü
(\%3E)|>)	Büyüktür işareti kontrolü veya hex eşdeğeri

Saldırı tespit sisteminin daha sağlıklı bir sonuç vermesi için, XSS saldırısında kullanılabilecek tüm HTML etiketlerine uygun düzenli ifadeler yazılmalıdır. Fakat tüm HTML etiketleri için ayrı ayrı düzenli ifadeler tanımlansa dahi saldırı tespit sisteminin tüm XSS saldırılarını tespit etmesi garanti değildir.

Bu çalışmada, XSS saldırısında kullanılabilecek HTML etiketleri ve JavaScript kodları, bölüm 2.2.1'de anlatıldığı gibi negatif algılama modeli referans alınarak tespit edilmeye çalışılmıştır. Negatif modelde olduğu gibi tehlikeli enjeksiyon kodları bir kara liste olarak tanımlanmış ve erişim kayıtlarındaki kullanıcı istekleri bu kara liste ile karşılaştırılmıştır. Kara listede tutulan verilerle eşleşen satırlar saldırı olarak tanımlanmış ve bu satırlara ait IP adresleri saldırgan olarak tanımlanmıştır.

XSS saldırısına ait örnek erişim kayıtları Tablo-6'da gösterilmiştir.

Tablo 6: XSS saldırısı sonucunda oluşan kütük kaydı örnekleri.

```
1.2.3.4  [12/Mar/2004:22:31:12 0500] "GET /foo.jsp?<SCRIPT>foo</SCRIPT>.jsp HTTP/1.1" 200 578
-------------------------------------------------------------------------------------------
1.2.3.4 [12/Mar/2004:22:37:17 0500] "GET /cgibin/cvslog.cgi?file=<SCRIPT>window.alert</SCRIPT> HTTP/1.1" 403 302
```

2.2.4 SQL Enjeksiyon Saldırısı ve Tespiti

SQL (Structured Query Language), veri tabanlarından veri seçme, silme ve güncelleme gibi işlemleri yapabilmek için kullanılan yapısal bir sorgulama dilidir. Hem ANSI, hem de ISO standardı olan SQL, yaygın olarak kullanılan Oracle, PostgreSQL, MSSQL Server, MySQL, DB2 gibi modern veri tabanı yönetim sistemlerinin temelini oluşturmaktadır [22].

Günümüzde profesyonel olarak hizmet sunan web uygulamalarının tamamına yakını veri tabanlarını kullanmaktadır. Online çalışan bu web uygulamaları, veri tabanı ile yapısal bir sorgulama dili olan SQL aracılığıyla haberleşirler [23]. SQL Enjeksiyon ise, web uygulamalarından alınan kullanıcı girdileri ile oluşturulan SQL sorgularının manipülasyonu olarak tanımlanabilir [24]. 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. 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, saldırı hedefini gerçekleştirir.

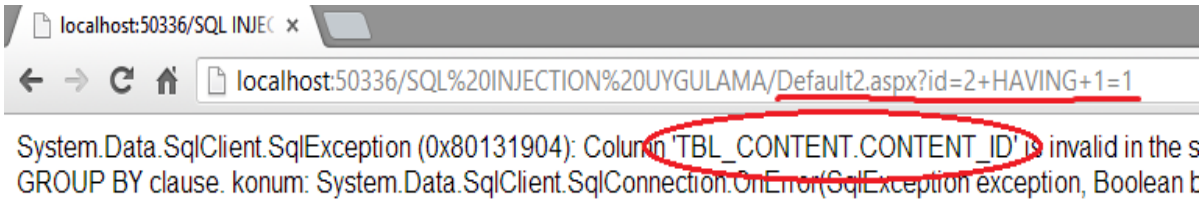
Şekil-1'de kullanıcının gerçek verileri kullanarak yönetici formuna giriş yapması durumunda oluşacak SQL sorgusu örneği ve kullanıcı girişi sonrasında oluşan SQL sorgusu resmedilmiştir.

Şekil 1: Örnek kullanıcı giriş formu.

Şekil 2: SQL enjeksiyon saldırısından örnek bir kesit.

Şekil-2'de saldırgan tarafından yönetici giriş formuna girilen bilgiler gösterilmiştir. Gösterilen SQL sorgusuna göre şifre alanına girilen değerle SQL sorgusuna "true" değer döndürülmesi sağlanmıştır. Saldırgan bu şekilde yönetici paneline erişip amacına ulaşmıştır.

Web formlarından POST metodu ile yapılan istekler kütük dosyalarına kaydedilmeyeceğinden, bu saldırı tekniği erişim kayıtlarından algılanamayacaktır. Fakat SQL enjeksiyonu saldırı tekniği Şekil-3'te görüldüğü gibi tarayıcı adres çubuğundan uygulanacak olursa, erişim kayıtlarından bu tip saldırılar algılanabilmektedir.



Şekil 3: SQL enjeksiyon saldırısından örnek bir kesit.

Şekil-3'de uygulanan teknik sonucunda elde edilecek erişim kaydı Tablo-7'de gösterilmiştir.

Tablo 7: SQL enjeksiyon saldırısı sonrasında oluşan erişim kaydı örneği.

2012-12-09 13:17:34 W3SVC100000 WEB151 1.2.3.4 GET /Default2.aspx?id=2+HAVING+1=1 80 HTTP/1.0 Mozilla/4.0+(compatible;+Synapse)

Tablo-8'de SQL enjeksiyonu tekniği için kullanılabilecek enjeksiyon kodlarına ait örnek düzenli ifadeler yer almaktadır.

Tablo 8: SQL enjeksiyon saldırısı için tanımlanmış düzenli ifadeler.

/(\')|(\%27)|(\\)|(#)|(\%23)/ix
/((\%27)|(\'))(select|union|insert|update|delete|replace|truncate)/ix
/exec(\s|\+)+(s|x)p\w+/ix
/(||%00|system\(|eval\(|`|\\)/i
/\w*((\%27)|(\'))(\s|\+|\%20)*((\%6F)|o|(\%4F))((\%72)|r|(\%52))/ix

XSS saldırı türünde olduğu gibi SQL enjeksiyon saldırısında da kural tabanlı saldırı tespiti yöntemi kullanılmıştır. Saldırıların algılanması için negatif algılama modeline dayalı bir kara liste oluşturulmuştur. Düzenli ifadeler yardımıyla erişim kayıtlarındaki kullanıcı istekleri kara liste ile karşılaştırılmıştır. Karşılaştırma sonucunda eşleşen kayıt satırları SQL enjeksiyon saldırısı olarak tanımlanmış ve bu kayıtlara ait IP adresleri saldırgan olarak tanımlanmıştır.

2.3 Kütük Verisi Kaynakları ve Tipleri

Birçok sistem, bünyesinde gerçekleştirilen işlemleri kütük dosyalarında saklamaktadırlar. Saklanılan bu dosyaların formatları günlük kaynaklarının türüne göre farklılık gösterebilmektedir. Bu tez çalışmasında, log tutulabilen sistemler dört ana başlık altında incelenmiştir.

2.3.1 Sistem Kütükleri

Sistem kütük dosyaları Microsoft Windows ya da Linux gibi sistemlerde gerçekleşen olayları kaydederler. Bu kayıtlar, bu tarz sistemlere normal ve anormal girişlerin belirlenmesine yardımcı olur. Sistem yönetimi için bu çok önemlidir. İşletim sistemlerinde önemli olan bu log dosyalarının toplanması, yöneticilerin işletim sistemi ve uygulamalardaki performans sorunlarını çözmesinde büyük rol oynamaktadır. Ayrıca bu dosyalardaki veriler, sistemde yapılan yetkisiz ve tehlikeli aktivitelerin, veri hırsızlığının,

virüs ya da solucan gibi nedenlerle oluşan tehlikeli aktivitelerin tespiti için de büyük önem taşımaktadırlar. Log dosyalarının toplanması ve güvenliğinin sağlanması, güvenlik sürecinde sistem için hayati önemi taşımaktadır. Windows sistemlerde üç çeşit sistem kütük dosyası vardır. Bunlar uygulama kütük dosyaları, güvenlik kütük dosyaları ve sistem kütük dosyalarıdır.

2.3.1.1 Uygulama Kütükleri

Uygulama kütükleri, işletim sistemleri üstünde çalıştırılan uygulama veya programların gerçekleştirdikleri olayları içerirler. Örnek olarak bir veri tabanı yönetim sistemi programı, uygulama kütüğüne dosya hatasını kaydedebilir. Program geliştiriciler hangi olayların kaydedileceğine karar verebilmektedir [25]. Şekil-4'de Windows işletim sistemine ait uygulama kütüğü örneği yer almaktadır. Kütük kayıtlarında, uygulamalara ait kayıtların düzeyleri, kaydın oluşturulduğu tarih-saat bilgisi, kaydın kaynak bilgisi gibi önemli bilgiler yer almaktadır.

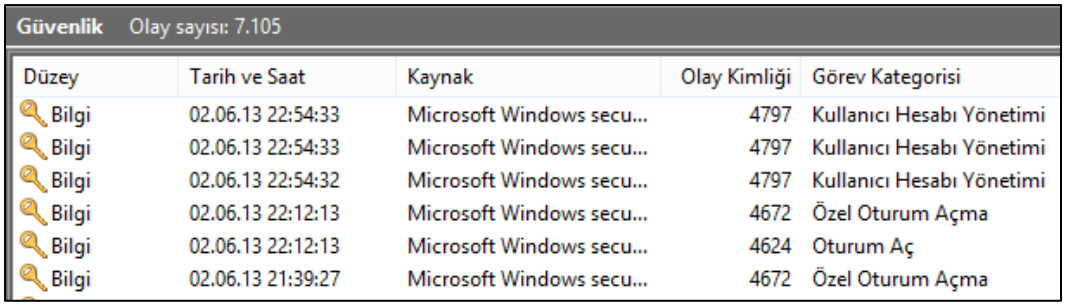
Uygulama Olay sayısı: 6.181

Düzey	Tarih ve Saat	Kaynak	Olay Kimliği	Görev Kategorisi
Bilgi	26.05.13 21:31:12	Security-SPP	1003	Yok
Bilgi	26.05.13 21:31:15	Security-SPP	1003	Yok
Hata	26.05.13 21:31:16	Security-SPP	8198	Yok
Bilgi	26.05.13 21:32:56	Search	1010	Arama Hizmeti
Uyarı	26.05.13 21:32:56	Search	1008	Arama Hizmeti
Bilgi	26.05.13 21:32:56	SecurityCenter	1	Yok
Bilgi	26.05.13 21:32:57	ESENT	105	Genel
Bilgi	26.05.13 21:32:57	ESENT	102	Genel
Bilgi	26.05.13 21:32:57	Search	1004	Arama Hizmeti
Bilgi	26.05.13 21:32:58	ESENT	325	Genel
Bilgi	26.05.13 21:33:08	CAPI2	4112	Yok
Bilgi	26.05.13 21:33:55	Search	1003	Arama Hizmeti
Bilgi	26.05.13 21:33:55	Search	1005	Arama Hizmeti
Bilgi	26.05.13 21:35:35	LoadPerf	1000	Yok
Bilgi	26.05.13 21:36:18	CAPI2	4111	Yok
Bilgi	26.05.13 21:36:18	CAPI2	4109	Yok
Bilgi	26.05.13 21:36:18	CAPI2	4109	Yok
Bilgi	26.05.13 21:36:18	CAPI2	4097	Yok
Uyarı	26.05.13 21:36:39	User Profile Service	1530	Yok

Şekil 4: Windows işletim sistemi örnek uygulama günlüğü kayıtları.

2.3.1.2 Güvenlik Kütükleri

Güvenlik kütüğü (Security Log) geçerli ve geçersiz oturum açma girişimlerinin yanı sıra dosya oluşturma, dosya açma ve dosya silme gibi kaynak kullanımı ile ilgili olayları da kaydetmektedir. Örneğin, oturum açma denetimi etkinleştirildiği zaman, kullanıcı bilgisayarda oturum açmayı her denediğinde güvenlik kütüğüne bir olay kaydedilir. Hangi olayların güvenlik kütüğüne kaydedileceğini belirlemek, olayları etkinleştirmek ve kullanmak için *administrator* grubunun bir üyesi olarak oturum açmış olmak gerekmektedir [25]. Şekil-5'de Windows işletim sistemine ait güvenlik kütüğü örneği yer almaktadır.



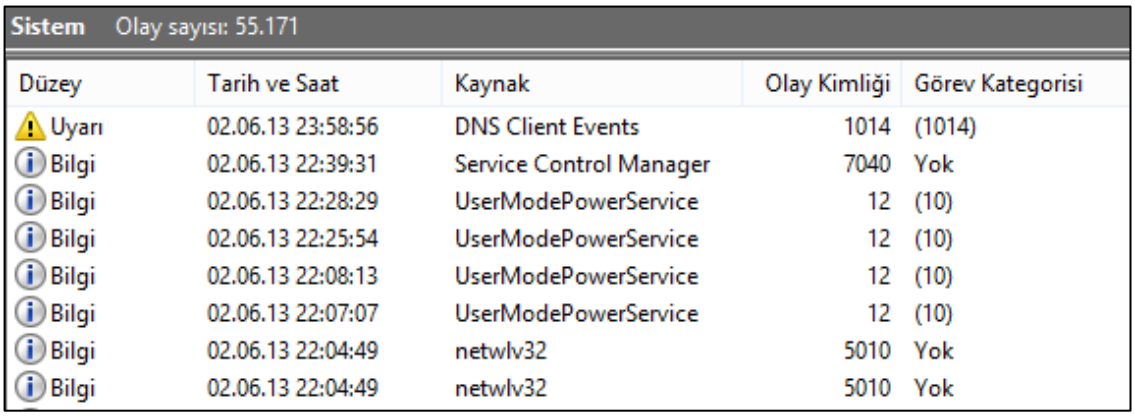
Güvenlik Olay sayısı: 7.105

Düzey	Tarih ve Saat	Kaynak	Olay Kimliği	Görev Kategorisi
Bilgi	02.06.13 22:54:33	Microsoft Windows secu...	4797	Kullanıcı Hesabı Yönetimi
Bilgi	02.06.13 22:54:33	Microsoft Windows secu...	4797	Kullanıcı Hesabı Yönetimi
Bilgi	02.06.13 22:54:32	Microsoft Windows secu...	4797	Kullanıcı Hesabı Yönetimi
Bilgi	02.06.13 22:12:13	Microsoft Windows secu...	4672	Özel Oturum Açma
Bilgi	02.06.13 22:12:13	Microsoft Windows secu...	4624	Oturum Aç
Bilgi	02.06.13 21:39:27	Microsoft Windows secu...	4672	Özel Oturum Açma

Şekil 5: Windows işletim sistemi örnek güvenlik günlüğü kayıtları.

2.3.1.3 Sistem Kütükleri

Sistem kütükleri (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 [25]. Şekil-6'da Windows işletim sistemine ait sistem kütüğü örneği yer almaktadır.



Sistem Olay sayısı: 55.171

Düzey	Tarih ve Saat	Kaynak	Olay Kimliği	Görev Kategorisi
Uyarı	02.06.13 23:58:56	DNS Client Events	1014	(1014)
Bilgi	02.06.13 22:39:31	Service Control Manager	7040	Yok
Bilgi	02.06.13 22:28:29	UserModePowerService	12	(10)
Bilgi	02.06.13 22:25:54	UserModePowerService	12	(10)
Bilgi	02.06.13 22:08:13	UserModePowerService	12	(10)
Bilgi	02.06.13 22:07:07	UserModePowerService	12	(10)
Bilgi	02.06.13 22:04:49	netwlv32	5010	Yok
Bilgi	02.06.13 22:04:49	netwlv32	5010	Yok

Şekil 6: Windows işletim sistemi örnek sistem günlüğü kayıtları.

2.3.2 Uygulama Sunucusu Kütükleri

Kişisel bilgisayarlardaki işletim sistemleri gibi sunucular üzerinde bulunan işletim sistemleri de sunucu üzerinde gerçekleşen olaylar ile ilgili kayıtları kütük dosyalarında tutmaktadırlar. Kayıtların tutulduğu bu dosyalarda, sunucu üzerine gerçekleşen erişimler, hatalar, bağlantı hataları, yüklenen dosyalar, çalıştırılan komutlar, erişilen nesneler, giriş kayıtları, veri tabanı aktiviteleri gibi önemli olaylar kaydedilmektedir. Bu bölümde web sunucusu kütükleri, elektronik posta sunucusu kütükleri, FTP (File Transfer Protocol) sunucusu kütükleri ve veri tabanı sunucusu kütükleri detaylıca incelenip, bu kütüklerin çeşitlerinden, tiplerinden ve formatlarından etraflıca bahsedilecektir.

2.3.2.1 Web Sunucu Kütükleri

Web sunucu günlük dosyaları (Web Server Log Files), sunucu platformundan bağımsız düz metin (ASCII) dosyalarıdır. Bu dosyalar, web tarayıcısı üzerinden web sunucusuna erişim sağlayan istemcilerin aktivitelerini düz metin halinde saklamaktadırlar [26, 27]. Saklanılan bu bilgilerin analiziyle, kullanıcılara ait birçok kullanışlı bilgi elde edilebilir. Kullanıcıların istekte bulunduğu URL adresleri, IP adresleri, kullanıcı isteğinin tamamlandığı ana ait tarih saat bilgileri, referans başlığı (Referer Header) bilgileri, http durum kodları, servis edilen byte miktarları ve kullanıcı etmen (User Agent) bilgileri bu bilgilere örnek olarak verilebilir.

Genel olarak dört tip sunucu günlük dosyası bulunmaktadır. Bunlar;

- Erişim Kütükleri (Access Logs)
- Agent Kütükleri (Agent Logs)
- Hata Kütükleri (Error Logs)
- Referans Kütükleri (Referer Logs)'dir.

İlk iki kütük dosyası tipi standart kütük dosyalarındandır. *Referer* ve *agent* kütük dosyaları ise sunucu üzerinden açık ya da kapalı yapılabilmektedir. Erişim kütükleri bu iki kütük dosyasının taşıdığı bilgileri içerecek şekilde genişletilerek (Extended Format) de kullanılabilmektedir [27].

2.3.2.1.1 Web Erişim Kütükleri

Web sunucusu tarafından işlenen bütün istekleri kaydeden kütüklere erişim kütükleri (Access Logs) denmektedir. Bu kütük dosyalarına *proxy* kütükleri de denmektedir. Bu dosyalar üzerlerinde bulundukları servisin açılmasından kapanmasına kadar geçen süre içerisinde kullanıcı erişim bilgilerini kaydederler. Sunucu servisi açık olduğu sürece bu dosyalar diskte bulundukları konumdan silinemez ya da taşınamazlar. Bu dosyaların silinmesi için sunucu servislerinin kapalı olması gerekmektedir.

Farklı işletim sistemleri üzerinde çalışan web sunucusu erişim kayıtları ile vekil sunucuları üzerindeki yazılımların sakladığı erişim kayıtlarının biçimleri birbirinden farklı olabilir. Örneğin, Linux işletim sistemi üzerinde çalışan bir Apache web sunucusu ile Windows Server işletim sistemi üzerinde çalışan bir IIS (Internet Information Services) web sunucusunun ürettiği erişim kütüklerinin biçimi birbirinden farklıdır [18]. Tablo-9'da bir web sunucusuna ait web erişim kütüğü örneği yer almaktadır. Örnek kayıt satırındaki bilgiler birbirlerinden boşluklarla ayrılmıştır.

Tablo 9: Erişim kütük dosyası örnek satırı.

```
1.2.3.4- - [31/May/2012:08:01:38 +0300] "GET /images/green.gif
   HTTP/1.1" 304 1346"http://www.examplesite.com/index.asp"
   "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)"
```

Web sunucu günlükleri ortak günlük biçimi (Common Log Format), birleştirilmiş günlük biçimi (Combined Log Format) ve çoklu erişim günlüğü (Multiple Access Log) olmak üzere üç ana başlık altında incelenebilirler.

Ortak günlük biçimi (Common Log Format), Apache ve Microsoft IIS sunucu yazılımlarında kullanılabilen bir kütük tipidir. Tablo-10'da örnek bir CLF günlük kaydı satırı gösterilmiştir.

Tablo 10: CLF günlük kaydı örnek satırı.

```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET
/apache_pb.gif HTTP/1.0" 200 2326
```

Tablo-11'de, Tablo-10'da verilen CLF günlük kaydı örnek satırının içerdiği alanların açıklamaları verilmiştir.

Tablo 11: CLF örneği alanlarının açıklanması.

Örnek Veri	Açıklama
127.0.0.1	İstemciye ait IP adresidir.
-	RFC931 veya kimlik tanımlamalarıdır. Özel tanımlamalar yapılmadığı sürece bilgi bulunmaz.
frank	İstekte bulunan kişinin kullandığı kimlik.
10/Oct/2000	Tarih bilgisi.
13:55:36	Saat bilgisi.
-0700	GMT saat bilgisi. (-0700 yerel saatten 7 saat öncesini temsil etmektedir.)
GET	İstemcinin kullanmış olduğu http istek metodu.
/apache_pb.gif	İstemci tarafından gidilen adres bilgisi.
HTTP/1.0	İstemcinin kullandığı http protokolü sürümü.
200	Sunucunun verdiği cevabın durum kodudur.
2326	Sunucudan dönen veri boyutudur.

Birleştirilmiş günlük biçimi (Combined Log Format), CLF tipi ile aynı alan bilgilerini barındırmaktadır. Fakat bu tipte, CLF tipine ek olarak kullanıcı etmen bilgisi(user agent) ve referans (referer) bilgileri de bulunmaktadır [28]. Tablo-12'de birleştirilmiş günlük biçimine örnek bir kayıt satırı verilmiştir.

Tablo 12: Birleştirilmiş günlük biçimi kaydı örnek satırı.

```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif
HTTP/1.0" 200 2326 "http://www.example.com/start.html" "Mozilla/4.08
[en] (Win98; I ;Nav)"
```

Tablo-13'de, Tablo-12'de verilen birleştirilmiş günlük biçimi kaydı örnek satırının içerdiği alanların açıklamaları verilmiştir.

Tablo 13: Birleştirilmiş günlük biçimi örneği alanlarının açıklanması.

Örnek Veri	Açıklama
127.0.0.1	İstemciye ait IP adresidir.
-	RFC931 veya kimlik tanımlamalarıdır. Özel tanımlamalar yapılmadığı sürece bilgi bulunmaz.
frank	İstekte bulunan kişinin kullanıcı kimliği.
10/Oct/2000	Tarih bilgisi.
13:55:36	Saat bilgisi.
-0700	GMT saat bilgisi. (-0700 yerel saatten 7 saat öncesini temsil etmektedir. '–' Yerel saatten öncesi, '+' Yerel saatten sonrası)
GET	İstemcinin kullanmış olduğu http istek metodu.
/apache_pb.gif	İstemci tarafından erişilen adres bilgisidir.
HTTP/1.0	İstemcinin kullandığı HTTP protokolü sürümüdür.
200	Sunucunun verdiği cevabın durum kodudur.
2326	Sunucudan dönen veri boyutudur.
http://www.example.com/start.html	Kullanıcının istekte bulunduğu sayfaya gelmeden önce kullanmış olduğu referans sayfa adresi.
Mozilla/4.08 [en] (Win98; I ;Nav)	İstemci tarafından kullanılan web tarayıcısının ve işletim sisteminin adı ve diğer özellikleri.

Çoklu erişim günlüğü (Multiple Access Log), yapılandırma dosyasında çok sayıda özel günlük yönergesi kullanılarak kolayca oluşturulabilir. Örneğin, Tablo-14'deki yönergelerle üç tane erişim günlüğü oluşturulacaktır. İlki temel CLF bilgisini içerirken diğer ikisi isteğin kaynaklandığı yeri ve tarayıcı kimliğini içerir. Son iki satır ayrıca, referans günlüğü ve kullanıcı etmen yönergelerinin etkilerinin nasıl oluşturulacağını göstermektedir [28].

Tablo 14: Çoklu erişim kütüğü yapılandırması.

```
LogFormat "%h %l %u %t \"%r\" %>s %b" common
    CustomLog logs/access_log common
CustomLog logs/referer_log "%{Referer}i -> %U"
  CustomLog logs/agent_log "%{User-agent}i"
```

2.3.2.1.2 Etmen Kütükleri

Etmen günlükleri (agent logs), sunucu üzerinde sistem yöneticisi tarafından aktif ya da pasif edilebilen kütük dosyalarıdır. Sistem yöneticisi gerektiğinde sunucu üzerinde bu

seçeneği aktif yaparak etmen günlük dosyalarının kaydedilmesini sağlayabilir. Bu kütük dosyalarının içerdiği bilgiler *birleştirilmiş günlük biçimi* (Combined Log Format) tarafından da kaydedilebildiği için ve sunucu performansını düşürücü bir etken olmaması açısından bu tip günlük dosyalarının aktif yapılmaması aksi bir durum olmadıkça tercih edilmeyebilir [11].

Kullanıcı etmen günlük dosyaları, sunucu üzerinde bağlantı kuran istemcilerin web tarayıcı bilgileri ile kullandıkları işletim sistemi bilgilerini tutarlar. Şekil-7'de örnek bir kullanıcı etmen günlük dosyası satır örneği verilmiştir.

```
Mozilla/3.0 (Win 95; 1)
```

Şekil 7: Etmen kütük dosyası satırı örneği.

- ***Tarayıcı (Browser):*** İstemcinin, sunucuya hangi web tarayıcısı ile bağlandığını gösteren alandır. Günümüzde internete bağlanırken birçok tarayıcı kullanılmaktadır. Şekil-7'deki örnekte istemci, Mozilla web tarayıcısını kullanmıştır.

- ***Tarayıcı Versiyonu(Browser Version):*** İstemcinin kullanmış olduğu web tarayıcısının versiyon bilgisini gösteren alandır. Şekil-7'deki örnekte istemci, Mozilla tarayıcısının 3.0 versiyonu kullanmıştır.

- ***İşletim Sistemi (Operating System):*** İstemcinin kullanmış olduğu işletim sistemi ve modeli hakkında bilgi içeren alandır. Şekil-7'deki örnekte istemci, Windows 95 işletim sistemini kullanmıştır.

Agent günlük dosyası, web sitelerinin tasarımı ve geliştirilmesi için gerekli bilgiler içermektedir. Yöneticiler bu bilgilerle kullanıcılara yönelik olarak web sitelerinin tasarımlarını değiştirebilirler. Bu sayede hem web sitesinde görüntülenen içerikler daha düzenli olur hem de kullanıcılar erişmek istedikleri bilgileri daha kolay bir şekilde elde etmiş olurlar.

2.3.2.1.3 Hata Kütükleri

En önemli kütük dosyalarından olan hata kütük dosyaları (Error Log Files), sistem yöneticilerinin sunucu üzerinde oluşan hataları görmelerini sağlamaktadırlar. Örneğin, bir istemci sunucu üzerinde bulunmayan bir dosyaya erişmek istediğinde kırık bağlantı hatası ile karşılaşacaktır. Oluşan bu hata sunucu üzerinde bulunan hata günlük dosyasına kaydedilmektedir. Hata günlüklerine ait bazı hata mesajları ve uyarı seviyeleri Tablo-15'de verilmiştir.

Tablo 15: Hata günlüğünde bulunan mesajların seviye tablosu [1].

Uyarı Seviyesi	Açıklama
emerg	Acil durumlar, sistem kullanılamaz.
alert	Hemen müdahale edilmelidir.
crit	Kritik durum mesajlarıdır.
error	Hata durumu mesajlarıdır.
warn	İkaz durumu mesajlarıdır.
notice	Normal ama önemli durumlardır.
info	Bilgilendirme mesajlarıdır.
debug	Hata ayıklama seviyesi mesajlarıdır.

Tablo 16: Hata günlüğü örneği.

```
[Wed Oct 11 14:32:52 2012] [error] [client 127.0.0.1] client denied by
        server configuration: /export/home/live/ap/htdocs/test
```

Tablo-16'da örnek hata mesajı satırı gösterilmiştir. İlk kısımda mesajın günlüğe işlendiği tarih ve saat bilgisi yer almaktadır. İkinci kısımda Tablo-15'de açıklanan hata seviyesi bulunmaktadır. Üçüncü kısımda istemciye ait IP adresi bilgisi yer almaktadır. Dördüncü kısımda oluşan hataya karşı sunucunun verdiği mesaj yer almaktadır. Son kısımda ise hataya yol açan dosyanın bulunduğu fiziksel yol belirtilmektedir.

Sistem yöneticileri, sistem üzerinde hataya yol açan sorunları tespit edip gerekli önlemleri hata günlük dosyalarından elde ettikleri bilgiler ile yapabilmektedirler. Bu bilgiler sistem üzerinde performans düşüklüğüne neden olan sorunların, hatalı kullanıcı girişlerinin ve hatalı dosya erişimlerinin azaltılması gibi sorunlarda kullanılabilir.

2.3.2.1.4 Referans Kütükleri

Referans kütükleri (Referer Logs), HTTP isteği başlığındaki *referer* alanı bilgilerini kaydetmektedirler. Bu alanda bulunan doküman adresleri, bir önceki adres de göz önünde bulundurularak referans kütüklerine kaydedilir. Tablo-17'de örnek referans kütüğü satırı gösterilmiştir.

Tablo 17: Örnek referans kütüğü satırı [12].

http://ornek.com/forum/index.html -> /resimleri/logo.gif

Tablo-17'de örnek referans satırındaki kayıt iki bölümden oluşmaktadır. Birinci kısımda istemcinin bir sonraki adımda gideceği adrese hangi adres üzerinden ulaştığı yer almaktadır. Diğer kısımda ise kullanıcının son olarak gittiği adres bilgisi yer almaktadır.

Bir istemci, bir web sitesine istekte bulunduğunda, sunucu gelen bu isteği HTTP durum (HTTP Status) kodlarından uygun olanı ile eşleştirmektedir. İstek kaydında bulunan durum kodlarına bakılarak, yapılan isteğin amacı ile ilgili bilgiler çıkarılabilmektedir. HTTP durum kodları üç haneden oluşan sayısal değerlerdir. Bu sayısal değerlerin ilk rakamı yapılan isteğe verilen cevabın kategorisini belirlemektedir. Son iki rakamın ise sınıflandırma değeri yoktur. HTTP durum kodları aşağıdaki gibi sınıflandırılmıştır:

1xx (Geçici Yanıt): Sunucuya istek geldiği zaman, sunucu tarafından yapılan işlemin devam ettiğini belirten geçici bir bilgilendirici yanıtın verildiğini gösteren durum kodlarıdır [18].

2xx (Başarılı): İstemci tarafından yapılan isteğin sunucu tarafından başarılı bir şekilde işlendiğini belirten durum kodlarıdır [18].

3xx (Yeniden Yönlendirme): İstemci tarafından yapılan isteği tamamlamak için ek işlemlerin yapılması gerektiğini belirten durum kodlarıdır. Bu kodlar genellikle yeniden yönlendirme amacıyla kullanılmaktadır [18].

4xx (İstek Hatası): İstemci tarafından yapılan isteğin hatalı olduğunu, engellendiğini ya da yapılan işlemin tamamlanamadığını belirten durum kodlarıdır [18].

5xx (Sunucu Hatası): Bu kodlar, isteğin işleme koyarken sunucunun dahili bir hatayla karşılaştığını gösterirler. Bu hatalar genellikle istemcinin isteğiyle ilgili değil, sunucu ile ilgili hatalardır [18].

2.3.2.2 E-Posta Sunucu Kütükleri

Hem kurumsal hem de gündelik hayatın vazgeçilmezlerinden biri olan e-postalar günümüzde iletişim için en yaygın olarak kullanılan araçlardandır. Bir e-posta sunucusu genellikle SMTP(Simple Mail Transfer Protocol) adı verilen ve 25 numaralı portta çalışan E-Posta Transfer Ajanı (Mail Transfer Agent) aracılığı ile e-posta alıp göndermektedir. MTA, SMTP protokolü üzerinden diğer e-posta sunucuları ile e-posta alış verişi yapabilen ve aldığı mailleri POP3, IMAP veya web desteği ile istemcilerin almasına yardımcı olan e-posta sunucu yazılımlarına verilen genel isimdir.

E-posta sunucularına gelen mesajlar 25. port üzerinden hareket etmektedirler. İki e-posta sunucusu aralarında konuşmaya başladıklarında ilk olarak SMTP protokolüne özgü olarak el sıkışma diye tabir edilen safhayı gerçekleştirmektedirler. 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 [29]. Bu veri akışı esnasında e-posta sunucuları gelen ve giden e-postalar ile ilgili günlük dosyaları tutmaktadırlar. Tablo-18'de örnek bir SMTP e-posta sunucusu günlüğü örneği verilmiştir. Günlük kaydına ait alanların açıklaması Tablo-19'da yer almaktadır.

Tablo 18: E-Posta sunucusu günlüğü örneği.

```
{09/05/2012 14:14:24} [5420] [Debug] 127.0.0.1:50760|25 <- MAIL FROM: <user@test.com>
{09/05/2012 14:14:24} [4904] [Debug] 127.0.0.1:50759|25 <- MAIL FROM: <user@test.com>
```

Tablo 19: SMTP E-Posta sunucusu günlüğü alanlarının açıklanması.

Örnek Veri	Açıklama
09/05/2012	E-Postaya ait tarih bilgisi.
14:14:24	E-Postaya ait saat bilgisi.
5420	E-Posta Kimlik (ID) bilgisi.
127.0.0.1:50760|25	Gönderen kişiye ait IP bilgisi, kullanılan port numarası ve SMTP port numarası.
MAIL FROM: <user@test.com>	E-Posta gönderen kişiye ait e-posta adresi bilgisi.

2.3.2.3 Veri Tabanı Sunucu Kütükleri

Veri tabanı sunucuları (Database Servers), herhangi bir veri tabanı yönetim sisteminin (VTYS) üzerinde bulunduğu sunuculardır. Web sunucuları gibi işlev görmektedir. Genel olarak bir sunucunun hem web sunucusu hem de veri tabanı sunucusu olarak kullanılması mümkündür fakat aşırı istemci talebi sunucu kaynaklarının tüketimini artıracağından ve bu durumun da performansın düşmesine sebep olacağından genellikle tavsiye edilmemektedir.

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 günlük dosyalarını kaydetmektedirler. Kaydedilen bu günlük dosyaları farklı VTYS'lerde farklı formatlarda ve isimlerde olabilmektedir. Tablo-20'de Microsoft SQL Server (MSSQL) tarafından oluşturulan günlük dosyalarında kullanılan hata mesajlarına ait önem dereceleri ve açıklamaları verilmiştir.

Tablo 20: MSSQL sunucusunda oluşan hatalarda kullanılan önem dereceleri [30].

Seviye	Açıklama
0 - 10	Bilgilendirme amaçlı uyarıları içerir gerçek hata mesajlarını içermezler.
11-16	Kullanıcı sorunlarından dolayı oluşturulan mesajlardır ve kullanıcı tarafından tespit edilebilirler.
17	SQL Sunucusunun yapılandırılmış ayarlardan farklı çalıştığını göstermektedir.
18	Ölümcül olmayan iç yazılım hatası
19	Kaynak limitinin aşıldığını belirtir.
20	Mevcut işlemle ilgili sorun olduğunu belirtir.
21	Sunucu üzerindeki tüm süreçleri etkileyen bir sorunla karşılaşıldığını belirtir.
22	Bir tablo yada index'in zarar gördüğünü belirtir. Sorunun çözümü için SQL sunucusunun yeniden başlatılması gerekmektedir.
23	Sunucu üzerinde bulunan veri tabanlarından birisinin hasarlı olduğunu belirtir.
24	Donanım hatası olduğunu belirtir.
25	Donanımsal ya da yazılımsal olarak ölümcül bir sistem hatasının olduğunu belirtir.

Tablo 21: Veri Tabanı sunucusu günlüğü örneği.

```
06/05/2013 22:05:02,Server,Unknown,Server is listening on [ 127.0.0.1 <ipv4> 1434].
```

Tablo-21'de bir VTYS kurulumu sonrasında oluşan günlük kaydı örneği gösterilmiştir. Günlük örneğinde alanlar virgül ile ayrılmıştır. Alanlar ile ilgili açıklamalar Tablo-22'de verilmiştir.

Tablo 22: Veri tabanı sunucusu günlüğünün alanlarının açıklanması.

Örnek Veri	Açıklama
06/05/2013	Sunucu üzerinde yapılan işleme ait tarih bilgisi.
22:05:02	Sunucu üzerinde yapılan işleme ait saat bilgisi.
Server	Kaynak bilgisi.
Unknown	Tablo 5'de açıklanan hata önem derecesi.
Server is listening on [127.0.0.1 <ipv4> 1434].	Mesaj.

2.3.3 Ağ Cihazları Kütükleri

Ağ cihazları (Network Devices), ağ alt yapısı tarafından gerçekleştirilen işlemler ile ilgili günlük dosyalarını saklamaktadırlar. Yönlendiriciler (Routers), anahtarlar (Switches) ve güvenlik duvarları (Firewall) gibi ağ cihazları, üzerlerinden geçen ağ trafiği ile ilgili günlükleri kaydetmektedirler [31]. Bu kayıtlar, güvenlik zafiyetlerinin belirlenmesinde ve önlem alınmasında büyük önem taşımaktadır. Cihaz ara yüzlerinin durum değişiklikleri, sistem konfigürasyon değişiklikleri, erişim listelerine (Access list) takılan bağlantılar gibi güvenlik açısından önemli olan bilgilerin kaydı tutulabilmektedir [32].

2.3.3.1 Yönlendirici Kütükleri

Yönlendirici (Router), OSI modelinin 3.katmanında çalışan ağ protokollerini destekleyen ve ağları birbirine bağlayan cihazlardır. Bir ya da daha fazla ağ arasındaki kesişim noktasıdır. Adından da belli olduğu üzere yöneltici görevinden dolayı iki ağ arasındaki veri iletişiminin doğruluğunu ve iletişimini sağlamaktadırlar.

Tablo 23: Yönlendirici kütük satırı örneği.

```
1|Fri, 15 Feb 2012 16:38:14 Source:192.168.0.2 BLOCK:www.examplesite.com
```

Tablo-23'de bir yönlendiriciye ait günlük dosyası örneği verilmiştir. Bu günlük kaydına ait alan bilgilerinin açıklamaları Tablo-24'de verilmiştir.

Tablo 24: Yönlendirici kütük dosyasının alanlarının açıklanması.

Örnek Veri	Açıklama
15 Feb 2012	Yönlendirici üzerinde gerçekleşen işleme ait tarih bilgisi.
16:38:14	Yönlendirici üzerinde gerçekleşen işleme ait saat bilgisi.
192.168.0.2	Kaynak cihazın IP adresi bilgisi.
BLOCK | ALLOW	Erişimin engellendiğini veya izin verildiğini belirtir.
www.examplesite.com	Engellenen ya da izin verilen IP adresi veya web adresi bilgisini belirtir.

2.3.3.2 Anahtar Kütükleri

Anahtarlar (Switch), ağ üzerindeki bilgisayarların ve diğer ağ aygıtlarının birbirlerine bağlanmasını sağlayan ağ cihazlarındandır. OSI modelinin 2. Katmanında çalışırlar ancak yeni nesil anahtarlarda *IP Routing* özelliği de olduğundan bu anahtarlar OSI modelinin 3. katmanında da çalışmaktadırlar. Diğer ağ cihazları gibi anahtarlar da üzerlerinde gerçekleştirdikleri işlemler için günlük dosyaları tutmaktadırlar.

Tablo 25: Anahtar kütüğü örneği.

```
Date/Time    Mod  Type  SType Dev   Origin          MSGID Source File/Line
-------------------------------------------------------------------------------
15:38:47  7  SYS  REST  NORM  00000 Local           00001 shostart.c:179
03-JUN-1997  00400040         LOCTIME
             Switch startup, version 7.4-00, 12-May-1997, Clock Log  : 15:38:33
             on 03-Jun-1997
```

Tablo-25'de bir anahtar cihazının üretmiş olduğu günlük satırından örnek bir kesit verilmiştir. Örnek günlük doyasında verilen alanlar ile ilgili açıklamalar Tablo-26'da gösterilmiştir.

Tablo 26: Anahtar kütüğü alanlarının açıklanması.

Alan Bilgisi	Açıklama
Date/Time	Anahtar üzerinde yapılan işleme ait tarih ve saat bilgisi.
S	Günlük iletisi önem seviyesi.
Mod	Günlük iletisinin oluşturulduğu modül.
Type	Mesaj tipi.
SType	Alt mesaj tipi.
Dev	Günlük iletisinin oluşturulmasını tetikleyen cihaz.
Origin	Günlük iletisinin yerel bir adresi veya bir IP adresi tarafından oluşup oluşmadığını belirtir.
MSGID	Mesaj kimlik numarası.
Source File/Line	Günlük mesajının oluşturulduğu modül kaynak dosyasının adı ve satır numarası.
Ref	Günlük iletisindeki referans alanının içeriği.
Flags	Günlük iletisindeki bayrak alanının içeriği.
Message	Günlük iletisindeki mesaj alanının içeriği.

2.3.3.3 Güvenlik Duvarları Kütükleri

Güvenlik duvarları bilgisayarlar ve ağlar arasındaki ağ trafiği akışını kontrol eden yazılım veya donanımlardır. Günümüzde güvenlik duvarları artık sadece ağ trafiğini denetlememektedirler. Bunun yanı sıra bazı saldırıların tespitini de yapabilmektedirler. İnternet ve intranet kullanıcıları güvenlik duvarlarını kullanarak bağlantılarını güvenli hale getirmektedirler [33].

Diğer ağ cihazlarında olduğu gibi donanımsal ya da yazılımsal olarak hizmet sunan güvenlik duvarları da ağ ortamında gerçekleşen olayları günlük dosyalarına kaydetmektedirler. Örnek olarak donanımsal bir güvenlik duvarı cihazı, üzerinden geçen ağ trafiği günlüklerini, yönetici olay günlüklerini, istenmeyen kaçak veri günlüklerini, uygulama günlüklerini, anti virüs tarama günlüklerini, web filtreleme günlüklerini, saldırı günlüklerini, ağ paketleri günlüklerini, e-posta filtreleme günlüklerini ve ağ tarama günlüklerini kaydedebilmektedir. Ayrıca elde edilen günlük dosyaları ile ayrıntılı web istatistikleri de çıkarabilmektedir. Günlük dosyalarının analizi ile ağda kullanılan protokollerin yüzdelik dağılımları, bant genişliği tüketimi, istemci ve kaynaklar tarafından kullanılan bant genişliği, sistemde bulunan toplam virüs sayısı, sisteme yapılan toplam saldırı sayısı, anti-spam aktiviteleri, web filtreleme aktiviteleri, ağ aktiviteleri, mail aktiviteleri, anti virüs aktiviteleri gibi analizler de çıkarabilmektedir. Bu aktivitelere ait

saatlik, günlük, haftalık, aylık, yıllık raporlar gibi sistem yöneticilerini yakından ilgilendiren ek raporlar da oluşturabilmektedir.

Tablo-27'de bir güvenlik duvarı cihazının oluşturduğu DLP (Data Loss Prevention) tipi günlük iletisi örneği verilmiştir. Bu örneğe ait alanların açıklanması Tablo-28'de verilmiştir. Tablo-28'de belirtilen kural seviyesi alanına ait açıklamalar Tablo-29'da gösterilmiştir.

Tablo 27: Güvenlik duvarı kütük dosyası örneği [34].

```
policyid=1 identidx=0 serial=73855 src="10.10.10.1" sport=1190
src_port=1190 srcint=internal dst="192.168.1.122" dport=80 dst_port=80
dst_int="wan1" service="https" status="detected" hostname="example.com"
url="/image/trees_pine_forest/" msg="data leak detected(Data Leak
Prevention Rule matched)" rulename="AllHTTP" action="log-only" severity=1
```

Tablo 28: Güvenlik duvarı kütük dosyasının alanlarının açıklanması [34].

Alan Bilgisi	**Açıklama**
policyid=(1)	Güvenlik duvarı politikasının kimlik numarası.
identidx=(0)	Kimlik temelli politika kimlik numarası.
serial=(73855)	Güvenlik duvarı oturum seri numarası.
src=(10.10.10.1)	Kaynak IP adresi bilgisi.
sport=(1190)	Kaynak port numarası bilgisi.
src_port=(1190)	Kaynak port numarası bilgisi.
srcint=(internal)	Kaynak arabirimi adı.
dst=(192.168.1.122)	Hedef IP adresi bildisi.
dport=(80	Hedef port numarası bilgisi.
dst_port=(80)	Hedef port numarası bilgisi.
dst_int=(wan1)	Hedef arabirimi adı.
service=(https)	Oturum veya paket için geçerli olan IP ağ servisi.
status=(detected)	Güvenlik duvarı cihazının gerçekleştirdiği eylem.
hostname=(example.com)	Web sitesinin anasayfa adresi.
url=(/image/trees_pine_forest/)	Kullanıcının görüntülediği web sayfasının URL adresi.
msg=(data leak detected(Data Leak Prevention Rule matched)	Kaydedilen cihaz aktivitesinin açıklama mesajı.
rulename=(All-HTTP)	DLP algılayıcısının içindeki DLP kuralının adı.
action=(log-only)	Kural içinde belirtilen eylem.
severity=(1)	Kural seviyesi.

Tablo 29: Güvenlik duvarı günlük kural seviyelerinin açıklamaları [34].

Seviye	Açıklama
1 – Uyarı (Alert)	Acil müdahalenin gerekli olduğunu belirtir.
2 – Kritik (Critical)	İşlevselliğin etkilendiğini belirtir.
3 – Hata (Error)	Bir hata durumunun olduğunu ve işlevselliğin etkilenebileceğini belirtir.
4 – İkaz (Warning)	İşlevselliğin etkilenebileceğini belirtir.
5 – Bildiri (Notification)	Normal olaylar hakkında bilgi verildiğini belirtir.
6 – Bilgi (Information)	Sistem işlevleri hakkında genel bilgi verildiğini belirtir.

2.4 Kütük Verilerinin Temizlenmesi ve Analiz Süreci

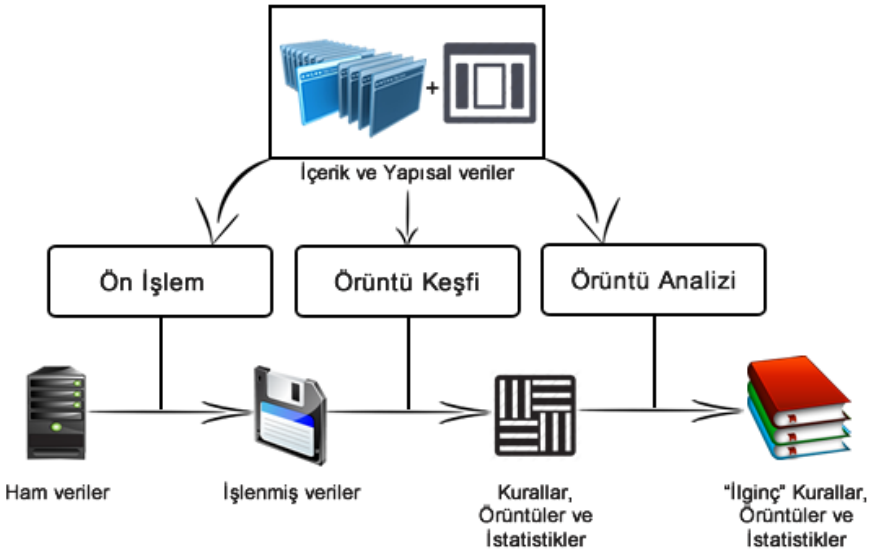
Son yıllarda artan internet kullanımına paralel olarak bilgi kaynaklarının ve servislerinin artması, bu hizmetleri kullanan kullanıcılar ve firmalar için çok önemli bir hale gelmiştir. Bu hizmetleri kullanan kullanıcıların internet davranışlarının incelenmesi özellikle elektronik ortamda pazarlama işi ile ilgilenen firmalar açısından daha da önemlidir. Internet kullanıcılarının işlemlerinden sonra oluşan verilerin web günlüklerine yazılması kurum ve kuruluşların kullanıcı davranışlarını analiz etmesi açısından büyük önem taşımaktadır. Elektronik ortamda ticari web sitesi yöneten web site sahipleri daha fazla satış yapabilmek, web sitesine ilk defa gelen kullanıcıları müşteri yapabilmek için sayfayı ziyaret eden kullanıcılardan daha fazla bilgi edinmeyi hedeflemektedirler. Bilgi amaçlı hizmet veren sitelere sahip olan kurum ve kuruluşlar da web sitelerinin daha aktif kullanılabilmesi ve gelen ziyaretçilerin ulaşmak istedikleri bilgi kaynaklarına en hızlı şekilde ulaşabilmelerini sağlamak için yoğun çaba göstermektedirler [35].

İnternet kullanımı ile ilgili yapılan çoğu istatistik, bir web sayfasını ziyaret eden kullanıcının, ziyaret ettiği sayfadan mümkün olan en kısa sürede ayrılmayı tercih ettiğini göstermektedir. Bu da web sitesi sahiplerinin web sitelerinde kullandıkları bilgilerin ne kadar yararlı olup olmadığının analizinin yapılmasını, neredeyse zorunlu hale getirmektedir. Kullanıcıların, istedikleri bilgileri en hızlı şekilde bulmasını sağlamak ancak web sitesi erişim kayıtlarının analizi ile mümkündür [35]. Bu kayıtların analiz edilmesi web madenciliği teknikleri ile yapılmaktadır. Erişim kayıtlarının temizlenmesi web kullanım madenciliğinin alt basamağıdır.

Web kullanım madenciliği, veri madenciliği tekniklerinin kullanılmasıyla, web verilerinden kullanıcı örüntülerinin keşfedilmesi ve bu verilerin analiz edilerek ilginç örüntülerin ortaya konulmasını sağlamaktadır [12]. Tipik veri madenciliği tekniklerini

uygulamadan önce veri temizleme işlemleri, veri kalitesini arttırmak için çok önemli bir rol oynamaktadır [18].

Web kullanım madenciliği işlemleri üç temel adımdan oluşmaktadır. Bu adımlar Şekil-8'de gösterilmektedir.



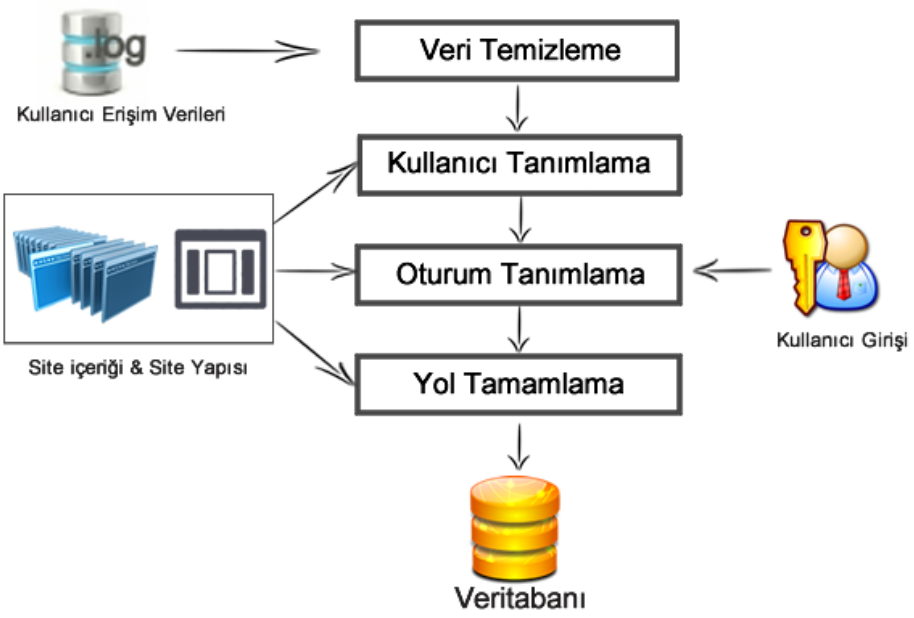
Şekil 8: Web kullanım madenciliğinin genel uygulama adımları.

2.4.1 Ön İşlem

Web kullanım madenciliğinin ilk aşaması olan ön işlem aşamasında, web sunucularından alınan düzensiz biçimdeki kullanıcı erişim kayıtlarının, analiz değeri olmayan ilişkisiz alanlardan arındırılarak belirli bir düzene getirilmesi sağlanmaktadır [12]. Ön işlem aşamasından geçirilen erişim kayıtları, bir sonraki adım olan örüntü keşfi için hazır hale getirilmektedir. Bu adımda yapılan işlemlerle veriler gürültüden temizlenmiş olmaktadır. Bu aşamada, büyük boyuttaki veriler temizlenip belli hesaplama aşamalarından geçirildiği için uzun zaman alabilmektedir.

Çoklu kütük kaynaklarından alınan ham verilerin ortak bir formatta birleştirilip herhangi bir veri tabanına kaydedilmesi büyük önem taşımaktadır. Web kullanım madenciliği sonucunda elde edilen verilerin doğru yorumlanabilmesi için ön işlem aşamanın başarılı ve doğru bir şekilde yapılması gerekmektedir.

Web kullanım madenciliği ön işlem aşamasının genel yapısı ve adımları Şekil-9'da gösterilmektedir.



Şekil 9: Web kullanım madenciliği ön işlem aşamaları.

2.4.1.1 Veri Temizleme

Kullanıcı erişim kayıt dosyasında web kullanım madenciliği açısından birçok gereksiz veri vardır. Veri temizleme adımıyla kütük dosyasında bulunan satırlardan resim, çoklu ortam, script (betik) ve erişilen web sitesinin tasarımında kullanılan stil ve javascript vb. dosyalara olan erişim kayıtlarının silinmesi gerekmektedir. Bunun yanında robot (spider veya bot) adı verilen ve web sitelerinde otomatik tarama yapan web uygulamalarının bırakmış oldukları izler de kütük dosyalarından silinmelidir. Tablo-30'da veri temizleme aşamasından geçirilmemiş (ham) örnek erişim kayıtları gösterilmektedir.

Tablo 30: Örnek ham erişim kayıtları.

	date	time	clp	csMethod	csUriStem	csUriQuery	scStatus
1	2007-03-30 00:00:00.000	09:01:20	65.55.208.93	GET	/robots.txt	NULL	404
2	2007-03-30 00:00:00.000	09:01:21	65.55.208.93	GET	/Default.asp	[95][80004005][Microsoft][ODBC_Microsoft_Access_S...	500
3	2007-03-30 00:00:00.000	10:07:52	88.229.205.159	GET	/STYLES.CSS	NULL	200
4	2007-03-30 00:00:00.000	10:07:52	88.229.205.159	GET	/iletisim.asp	NULL	200
5	2007-03-30 00:00:00.000	10:07:52	88.229.205.159	GET	/strbkgde.gif	NULL	404
6	2007-03-30 00:00:00.000	10:07:52	88.229.205.159	GET	/images/topnav_bg.gif	NULL	200
7	2007-03-30 00:00:00.000	10:07:52	88.229.205.159	GET	/images/1_03.gif	NULL	200
8	2007-03-30 00:00:00.000	10:07:52	88.229.205.159	GET	/images/1_02.gif	NULL	200
9	2007-03-30 00:00:00.000	10:07:52	88.229.205.159	GET	/images/1_04.gif	NULL	200
10	2007-03-30 00:00:00.000	10:07:53	88.229.205.159	GET	/images/1_05.gif	NULL	200

Bir internet kullanıcısı bir web sayfasına istekte bulunduğu zaman o sayfanın içeriğinde bulunan resim, çoklu ortam dosyaları, stil dosyaları gibi web kullanım

madenciliği açısından önemi olmayan dosyalara da istekte bulunmuş olur. Tablo-30'da görüldüğü üzere, 88.229.205.159 IP adresine sahip olan kullanıcı saat 10:07:52 'de "iletisim.asp" sayfasına başarıyla (scStatus=200) giriş yapmıştır. Fakat kullanıcı "iletisim.asp" sayfasına giriş yaptığında o sayfanın içeriğinde bulunan ".css", ".gif" vb. dosyalarına da erişim sağlamış olmaktadır. "iletisim.asp" sayfasına yapılan erişim sonucunda oluşan erişim kayıtları Tablo 30'da görülmektedir. Veri temizleme adımında, Tablo-30'da görülen ve web kullanım madenciliğinde dikkate alınmayan satırlar silinmelidir. Veri temizleme adımından sonra oluşan örnek kayıtlar Tablo-31'deki gibidir.

Tablo 31: Veri temizleme adımından kalan satırlar.

	date	time	clp	csMethod	csUriStem	csUriQuery	scStatus
1	2007-03-30 00:00:00.000	2013-01-01 10:07:52.000	88.229.205.159	GET	/iletisim.asp	NULL	200
2	2007-03-30 00:00:00.000	2013-01-01 14:01:48.000	65.54.188.55	GET	/iletisim.asp	NULL	200
3	2007-03-30 00:00:00.000	2013-01-01 14:01:53.000	65.54.188.55	GET	/yonetim.asp	NULL	200
4	2007-03-30 00:00:00.000	2013-01-01 14:19:11.000	88.245.65.22	GET	/misyon.asp	NULL	200
5	2007-03-30 00:00:00.000	2013-01-01 14:19:29.000	88.245.65.22	GET	/iletisim.asp	NULL	200
6	2007-03-30 00:00:00.000	2013-01-01 16:07:29.000	88.64.151.239	GET	/misyon.asp	NULL	200
7	2007-03-30 00:00:00.000	2013-01-01 16:07:50.000	88.64.151.239	GET	/misyon.asp	NULL	200
8	2007-03-30 00:00:00.000	2013-01-01 16:08:17.000	88.64.151.239	GET	/misyon.asp	NULL	200
9	2007-03-30 00:00:00.000	2013-01-01 16:08:26.000	88.64.151.239	GET	/iletisim.asp	NULL	200
10	2007-03-30 00:00:00.000	2013-01-01 17:06:35.000	88.244.22.1	GET	/misyon.asp	NULL	200

2.4.1.2 Kullanıcı Tanımlama

Bu adımda amaç, erişim kayıtlarında oluşan kayıtların hangi kullanıcılara ait olduğunun tespit edilmesidir. Eğer web sayfasına istekte bulunan kullanıcı web sayfasına kullanıcı kimliği ve şifresiyle giriş yapmışsa kullanıcı tanımlama adımının uygulanması daha kolay olacaktır ve daha az zaman alacaktır. Genelde çoğu web sitesi yayınlamış olduğu içerikleri, kendi üyelerinin dışında anonim kullanıcılar ile de paylaşmaktadır. Anonim kullanıcıların web sitesine erişimleri sonucunda erişim kayıtlarında bir nevi kullanıcı adı olarak kullanabileceğimiz bazı bilgiler kayıt altına alınmaktadır. Bunlar kullanıcıya ait "agent" bilgisi, çerez bilgisi ve IP adresleri gibi bilgilerdir. Kayıtlı kullanıcı olarak giriş yapılmamış web sitelerinde, kullanıcı erişim davranışlarının analiz edilebilmesi için benzersiz kullanıcıların tespit edilmesi gerekmektedir. Proxy sunucularının bulunduğu sistemlerde web sitelerine yapılan istekler sonucu oluşan erişim kayıtlarında IP adresi bilgisi tüm satırlar için aynı olacaktır. Kullanıcı erişim kayıtlarından farklı kullanıcıları ayırt edebilmek için "agent" bilgisi kullanılabilir. Erişim kayıtlarında bulunan "agent"

alanında siteye erişen kullanıcının tarayıcı ve işletim sistemi bilgiler yer almaktadır [18]. Burada, aynı IP adresine sahip kayıtların "agent" bilgileri karşılaştırılarak kullanıcılar ayırt edilebilir. Farklı kullanıcıları tespit etmekte kullanılan bir başka bilgi de çerez (cookie) bilgileridir. Kullanıcıların istekte bulunduğu bir web uygulaması, her kullanıcı için ayrı bir çerez üretmekte ve üretilen bu çerezler kullanıcı bilgisayarında saklanmaktadır. Temizlenmiş erişim kayıtlarının çerez bilgileri karşılaştırılarak farklı kullanıcılar tespit edilebilir. Tablo-32'de kullanıcı tanımlama adımından geçirilmiş örnek erişim kayıtları yer almaktadır.

Tablo 32: Kullanıcı tanımlama adımı sonrasında oluşan kayıtlar.

	cIp	csUserAgent
4	85.106.229.0	Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1)
5	88.226.0.194	Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+.NET+CLR+2.0.50727)
6	88.229.186.66	Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+5.1)
7	88.229.205.159	Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1)
8	88.231.143.36	Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1)
9	88.244.22.1	Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1)
10	88.245.65.22	Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+5.1;+.NET+CLR+1.1.4322;+InfoPath.1)
11	88.64.151.239	Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+5.1;+.NET+CLR+1.0.3705;+.NET+CLR+1.1.4322;+Media+Center+PC+4.0;+.NET+CLR+2.0.50727)

2.4.1.3 Oturum Tanımlama

Oturum, bir kullanıcının bir web sitesine girişinden çıkışına kadar gerçekleştirmiş olduğu aktiviteler topluluğudur. Oturum tanımlama işlemi, bir web sitesine giren tüm kullanıcıların oturumlarının gruplandırılması işlemi olarak tanımlanabilir. Oturum tanımlama adımı için oturum süresi temelli (session-duration-based), sayfada kalma süresi temelli (page-stay-time-based method) ve referans temelli (referrer-basic heuristic method) olmak üzere çeşitli sezgisel yöntemler kullanılabilir [18].

Oturum süresi temelli sezgisel yönteminde, kullanıcının ardı ardına giriş yaptığı iki sayfa arasındaki erişim zamanı farkı temel alınır. Bu konuda birçok çalışma 30 dakikayı temel almıştır. Temel alınan zaman aşımı eşik değerine göre her kullanıcının sayfa geçişlerinin zaman farkı hesaplanmakta ve oturumlar bu aralığa göre belirlenmektedir.

Sayfada kalma süresi temelli sezgisel yöntemde, kullanıcıların sayfada harcadığı toplam süre 10 saniyeyi geçmemelidir. Kullanıcının iki aktivite arası geçirdiği süre 10 saniyeyi geçmediği sürece kullanıcı aktiviteleri aynı oturum içerisinde kabul edilir.

Kullanıcıların bağlanmış olduğu sunucu üzerindeki aktiviteler 10'ar saniyeler şeklinde bölünerek tanımlanan oturumlara göre analiz edilmektedir [18].

Referans temelli sezgisel yöntemde, kullanıcıların oturumlarını belirlemek için erişim kayıtlarında bulunan alanlardan kullanıcının istekte bulunduğu sayfa ile beraber referans (referer) sayfası kullanılmaktadır. Kullanıcının yapmış olduğu ilk istek erişim kayıtlarına kaydedilirken genelde referans adresi boş olarak kaydedilmektedir. Bu yüzden kullanıcıya ait oturum başlangıcı bu satır referans alınarak oluşturulmaktadır [18].

Bu çalışmada oturum tanımlama adımı için kullanılan sezgisel yöntemlerden oturum süresi temelli sezgisel yöntemi kullanılmıştır. Tablo-33'de oturum tamamlama adımından geçirilmiş örnek erişim kayıtları yer almaktadır. Tablodaki ilgili veriler, gizlilik nedeniyle sembolik olarak A, B, C, D harfleriyle isimlendirilmiştir.

Tablo 33: Oturum tanımlama adımından kalan kayıtlar.

	Tarih	Zaman	IpAdresi	DomainAdresi	Referer	OturumNo	KullaniciNo
1	16/05/2008	09:15:35	10.1.1.212	A	NULL	1	1
2	16/05/2008	17:12:04	10.1.2.104	A	NULL	1	2
3	16/05/2008	17:27:17	10.1.2.104	B	A	1	2
4	16/05/2008	17:01:45	10.1.2.105	A	NULL	1	3
5	16/05/2008	05:31:09	10.1.2.106	B	A	1	4
6	16/05/2008	05:31:16	10.1.2.106	C	B	1	4
7	16/05/2008	12:24:45	10.1.2.106	D	C	2	4
8	16/05/2008	12:24:48	10.1.2.106	A	D	2	4
9	16/05/2008	12:24:58	10.1.2.106	C	A	2	4
10	16/05/2008	11:49:36	10.1.2.124	C	NULL	1	5

2.4.1.4 Yol Tamamlama

Erişim kayıtlarında bazen bazı bağlantıların eksik olduğu görülmektedir. Bunun sebebi genelde web tarayıcısının önbelleğinden ya da kullanıcının kullandığı vekil sunucudan (Proxy Server) dolayı bazı sayfaların önbelleğe alınmış olmasıdır. Çoğu kullanıcı internette gezinirken, önceden açmış olduğu sayfaya tekrar bağlanabilmek için kullanmış olduğu tarayıcının "geri" düğmesini kullanır. Bu durumda web tarayıcısı, kullanıcının gezmiş olduğu sayfaları önbellekten alarak tekrar kullanıcıya göstermektedir. Önbellekten kullanıcıya tekrar gösterilen web sayfası için web erişim kayıtlarında yeni bir satır oluşturulmamaktadır. Bu durumun çözülmesi yol tamamlama adımı ile gerçekleştirilmektedir.

2.4.2 Örüntü Keşfi

Web kullanım madenciliğinin ikinci uygulama aşaması olan örüntü keşfi aşaması, ön işlemden geçirilen verilere, veri madenciliği tekniklerinin uygulandığı aşamadır. Bu aşamada ön işlem aşamasından elde edilen düzenli ama anlamsız olan verilerden ilginç ve gerekli olan bilgilerin çıkarımı yapılmaktadır.

Örüntü keşfi aşamasında birçok yöntem ve algoritma kullanılabilmektedir. İstatistiksel analiz, ilişkilendirme kuralları, kümeleme, sınıflandırma, sıralı örüntüler ve bağımlı modelleme gibi teknikler bu yöntem ve algoritmalara örnek olarak verilebilir [11,12].

2.4.2.1 İstatistiksel Analiz

İstatistiksel yöntemler web kullanıcıları hakkında bilgi keşfetmek için en sık kullanılan yöntemlerdir. Bu yöntemde oturum dosyası analiz edilerek, çeşitli istatistiksel analizler (frekans, ortalama, medyan, vb.) elde edilebilir. İstatistiksel analiz sürecinin amacı, bir web sitesi içerisindeki birçok temel bilgiyi elde etmektir [18]. Bu temel bilgilere örnek olarak; kullanıcıların bir web sayfasında gezindikleri sayfaların görüntülenmesi, kırık bağlantıların tespit edilmesi, başarılı ve başarısız isteklerin belirlenmesi, en çok giriş yapılan sayfaların bulunması, en az giriş yapılan sayfaların bulunması, kullanıcı isteklerinin çeşitli parametrelere göre gruplandırılarak gösterilmesi gibi önemli bilgiler gösterilebilir. İstatistiksel bilgiler analiz edilerek sistem performansının arttırılması, sistem güvenliğinin arttırılması ve pazarlama kararlarını destekleyici raporlar çıkartılabilir.

Erişim kayıtlarının analizi sonucunda elde edilen istatistikî bilgilerle web siteleri daha etkin bir şekilde organize edilebilmektedir ve web site performansının iyileştirilmesi için web tasarımcılarına ve site yöneticilerine yardım edilmektedir.

2.4.2.2 Birliktelik kuralı

Bu yöntem genellikle elektronik ticaret uygulamalarında kullanılmaktadır. Bu yöntemin amacı bir küme içerisinde bulunan nesnelerin birbirleriyle olan ilişkilerinin tespit edilmesidir.

Bu yönteme örnek verecek olursak; bir X ürününün satın alınması ile Y veya Z ürününün alınması arasında bir ilişki veya bağlantı olup olmadığının tespit edilmesi şayet bu ürünlerin satın alınmaları arasında bir bağlantı var ise bu bağlantının kuvvet derecesinin

çıkarılması sağlanır. Ortaya çıkan sonuçlarla Y ve Z ürünlerinin alımlarıyla ilgili olarak sistemde değişiklik yapmaktır. Örneğin, bir teknoloji mağazasından dizüstü bilgisayar alan kişilerin aldığı bilgisayarın yanında dizüstü bilgisayar çantası alması ya da dizüstü soğutucusu alması arasında bir bağlantı olduğunu düşünürsek, mağaza yetkililerini bu tip ürünlerin aynı rafta ya da yakın raflarda sergilenmesini sağlaması, birliktelik kuralını uyguladıklarını göstermektedir. Bu işlem bir web sitesinde ürün sayfalarının yapılandırılmasında kullanılabilir [36].

2.4.2.3 Sınıflandırma

Sınıflandırma tekniği, bir veriyi daha önceden tanımlanmış sınıflara dağıtma tekniğidir. Bu teknikte, verilerin özelliklerini en iyi şekilde açıklamak için seçim ve açığa çıkarma uygulamalarına ihtiyaç duyulur. Sınıflandırma; karar ağaçları, bayesian, en yakın komşu ve destek vektör makineleri gibi denetlenen tümevarımsal öğrenim algoritmaları kullanılarak yapılabilir [36].

2.4.2.4 Sıralı Örüntüler

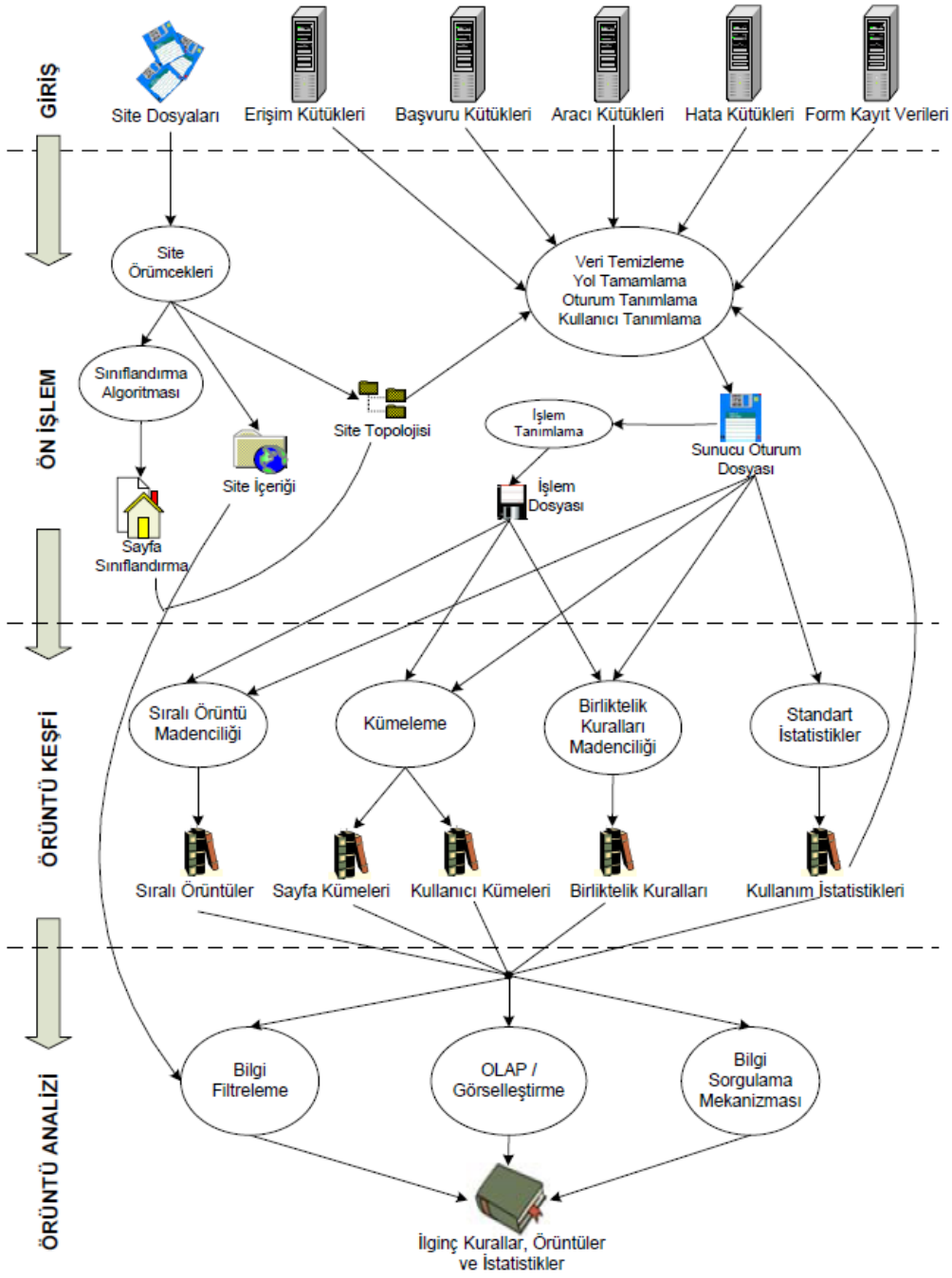
Sıralı örüntü yöntemiyle, kullanıcı oturumları arasında örüntü bulunmaya çalışılır. Bu işlemde, sabit bir zaman aralığı içerisindeki kullanıcı oturumları incelenir ve karşılaştırılır. Karşılaştırılan oturumlardan elde edilen örüntülerle kullanıcıların eğilimleri belirlenir. Sıralı örüntülerin bulunması, elektronik ticaret sunan bir web sitesinde yöneticilere, kullanıcıların gelecekteki eğilimleri tahmin edici bilgiler sunarak, yöneticilerin bu yönde ilan vermesini sağlayabilir.

2.4.2.5 Kümeleme

Kümeleme, verilerin gruplandırılmasıdır. Benzer ya da birbiriyle ilişkili olan veriler aynı kümeyi oluştururken farklı ya da aralarında ilişki olmayan veriler farklı bir kümeyi oluşturmaktadır. Kümelemenin temel hedefi, dağınık bir halde bulunan verileri benzerliklerine göre bir araya getirip sınıflandırarak işlenebilir hale getirmektir. Kümelemede genellikle yapay sinir ağları (YSA) ile istatistiksel metotlardan yararlanılır ve örüntü tanıma, görüntü işleme, benzer ortak arkadaş grupları keşfetme, veri madenciliği, istatistik, biyoloji ve makine öğrenmesi gibi pek çok alanda kullanılır [37].

2.4.3 Örüntü Analizi

Örüntü analizi web kullanım madenciliğinin son aşamasıdır. Örüntü analizinde amaç örüntü keşfinde bulunan ilginç olmayan kural ve desenlerin filtrelenmesidir. Yaygın olarak kullanılan bilgi sorgulama mekanizmaları MSSQL, MySQL gibi veri tabanı uygulamaları ve OLAP(On-Line Analytical Processing) uygulamaları bu aşamada gerçekleştirilmektedir [12]. Web kullanım madenciliğine ait tüm adımlar Şekil-10'da gösterilmiştir.



Şekil 10: Web kullanım madenciliği uygulama adımları [18].

3. WEB KULLANICI ERİŞİM KÜTÜKLERİNİN TEMİZLENMESİNE YÖNELİK YAZILIMIN GELİŞTİRİLMESİ

3.1 Giriş

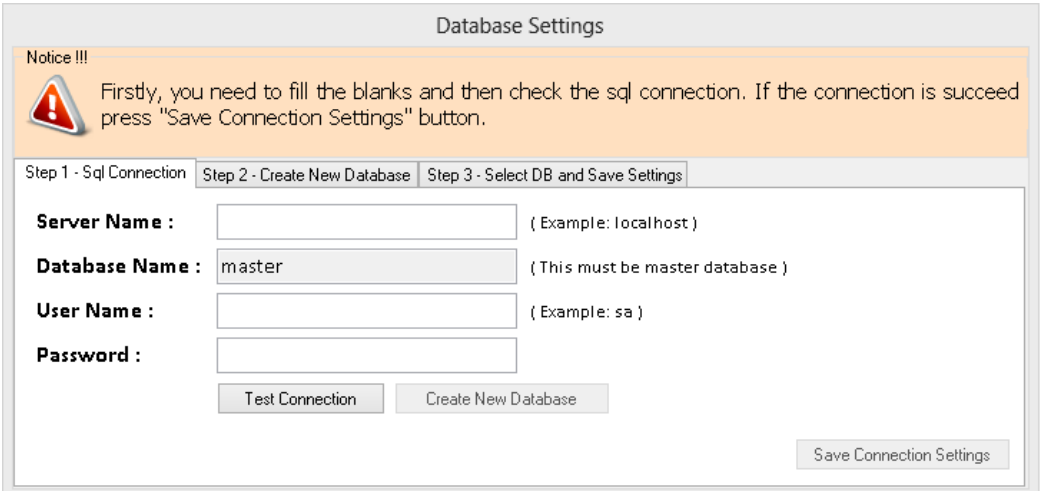
Bu tez çalışmasında web erişim kütüklerinin temizlenmesine yönelik bir yazılım geliştirilmiştir. Geliştirilen bu yazılım ile web sunucularından elde edilen erişim kütük dosyaları, web kullanım madenciliğinin ilk aşaması olan ön işlem aşamasından geçirilerek bir oturum dosyası oluşturulmuştur. Yazılım, Visual C# programlama dilinde yazılmıştır ve Microsoft SQL Server veri tabanını kullanmaktadır. Geliştirilen yazılımla kullanıcı, MSSQL veri tabanı sunucusu üzerinde bulunan veri tabanlarını kullanabilecek veya yeni bir veri tabanı oluşturabilecektir. Oluşturulan veri tabanı içerisinde, web kullanım madenciliğinin ön işlem aşamasındaki adımlara ait veri tabanı tabloları bulunmaktadır. Uygulama içinde kullanılan veri tabanı tabloları, programın işleyişi esnasında oluşturulmaktadır. Bu tablolar, ön işlem aşamasındaki adımlara ait veri temizleme, kullanıcı tanımlama ve oturum tanımlama işlemleri sonucunda elde edilen bilgileri içermektedir. Elde edilen bu bilgilerle, web sitesine ait bazı erişim istatistiklerine ulaşılabilmekte ve web sitesine yapılan saldırı girişimleri tespit edilebilmektedir. Uygulama, veri temizleme adımında elde edilen veri tabanı tablosu verileriyle, web sitelerine yapılan XSS ve SQL enjeksiyon saldırılarını, kural tabanlı saldırı tespiti yöntemine göre analiz etmektedir. Yapılan analiz sonucunda saldırıların yapıldığı tarih-zaman bilgileri, IP adresi bilgileri, saldırı satırlarını içeren kütük dosyasının adı, saldırının yapıldığı anda oluşturulan erişim satırının hangi kütük dosyasında kaçıncı satırda olduğu ve saldırı türü bilgileri uygulama ara yüzünde tablo halinde gösterilmektedir.

Bu çalışmada kütük dosyalarının veri tabanına aktarımı için ücretsiz olan Microsoft Log Parser 2.2 yazılımı kullanılmıştır. Log Parser yazılımı sisteme kurulduğunda, kurulum dizininde iki adet uygulama oluşmaktadır. Bunlar; “LogParser.dll” ve “LogParser.exe” uygulamalarıdır. “LogParser.exe” dosyası, komut satırından parametre gönderilerek çalıştırılabilen bir uygulamadır. “LogParser.dll” dosyası ise ActiveX nesnesidir ve kolaylıkla C# uygulamalarına referans edilebilmektedir. ActiveX nesnesi Log Parser kurulumu esnasında "MS Utility 1.0 Type Library - LogParser Interfaces Collection" anahtar adıyla, sistem kayıt defterine kaydedilmektedir. “LogParser.dll” ActiveX nesnesi VBScript, C#, JavaScript gibi .Net programlama dilleri ile kullanılabilmektedir. İsmi geçen

script dillerinde LogParser nesnesi oluşturabilmek "MSUtil.LogQuery" sınıfı ile mümkündür [38].

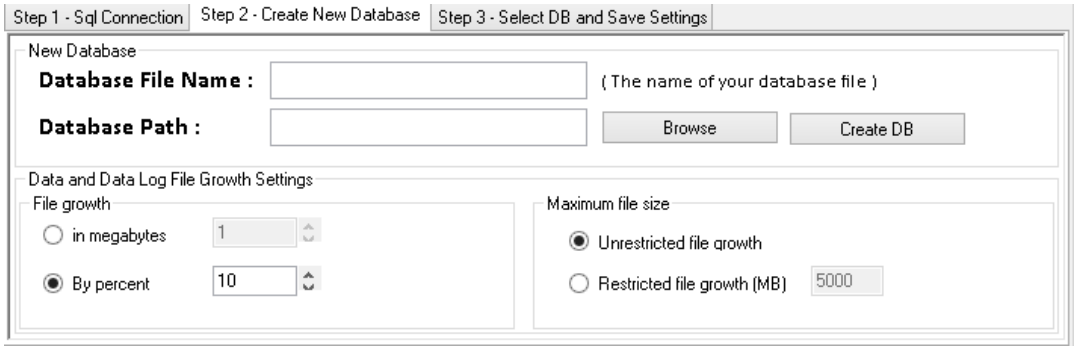
3.2 Geliştirilen Log Cleaning Software (LCS) Yazılımının Özellikleri

Web erişim kütüklerini temizlemek ve analiz etmek amacıyla geliştirilen bu yazılım, iki ana formdan oluşmaktadır. Bu formlardan ilki, yazılımda kullanılacak veri tabanına ait ayarların yapıldığı "Database Settings" formudur. Şekil-11'de veri tabanı ayarlarının gerçekleştirildiği form resmedilmiştir.



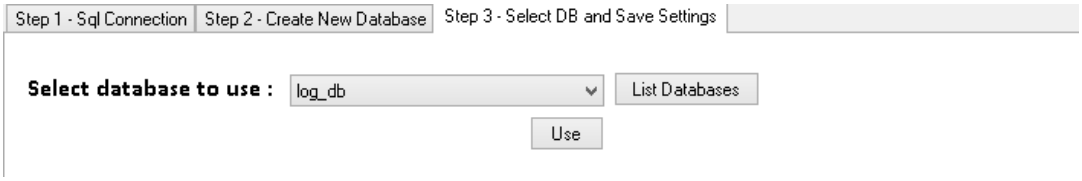
Şekil 11: Veri tabanı ayarları ara yüzü.

Şekil-11'deki form üç kısımdan oluşmaktadır. Kullanıcı, ilk olarak SQL sunucusuna erişmek için ilgili alanları doldurmalı ve girmiş olduğu bilgilerle veri tabanına bağlanıp bağlanamayacağını "Test Connection" düğmesi ile test etmelidir. Test sonucunda, kullanıcının girmiş olduğu SQL sunucu giriş bilgileri doğruysa daha sonraki girişlerde bu bilgilerin otomatik olarak geri getirilmesi için uygulama ara yüzünde "Save Connection Settings" düğmesine tıklanmalıdır. Kullanıcının girmiş olduğu veriler test edildikten sonra yeni veri tabanı oluşturma adımına geçilmelidir. Şekil-12'de veri tabanı oluşturma formu resmedilmiştir.



Şekil 12: Veri tabanı oluşturma ara yüzü.

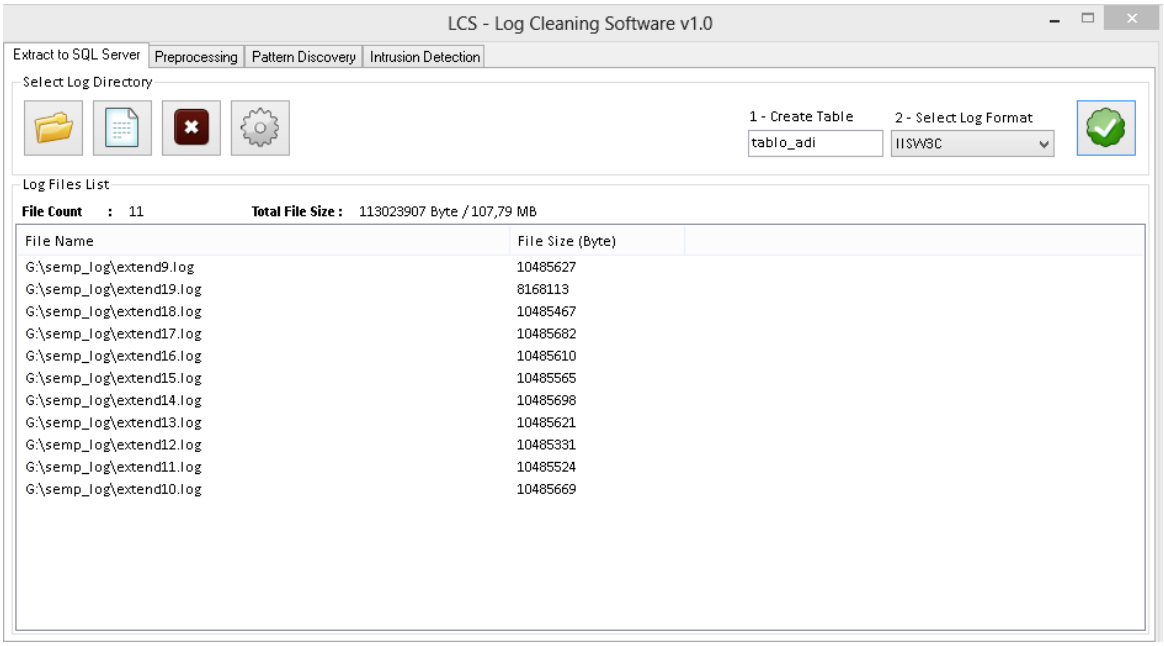
Kullanıcı, Şekil-12'de gösterilen formda, oluşturulacak veri tabanı adını, veri tabanı dosyasının oluşturulacağı fiziksel dosya yolunu ve veri tabanı büyüme ayarlarını gerçekleştirebilmektedir. Bir sonraki adımda kullanıcı, önceden oluşturduğu veri tabanını ya da Şekil-12'de oluşturduğu veri tabanını seçmelidir. Şekil-13'de bu adıma ait ekran görüntüsü yer almaktadır.



Şekil 13: Web madenciliği aşamalarında kullanılacak veri tabanının seçilmesi.

Şekil-13'de, web kullanım madenciliği aşamasında kullanılacak veri tabanının seçimi yapılmalıdır. Web kullanım madenciliği ön işlem aşamalarından geçirilen kütük dosyası verileri, bu formda seçilen veri tabanına aktarılmaktadır. Kullanıcı ilk olarak, "List Databases" düğmesine tıklayarak SQL sunucusunda bulunan veri tabanlarını listelemelidir. Daha sonra kullanılacak veri tabanı seçilerek "Use" düğmesi tıklanmalıdır.

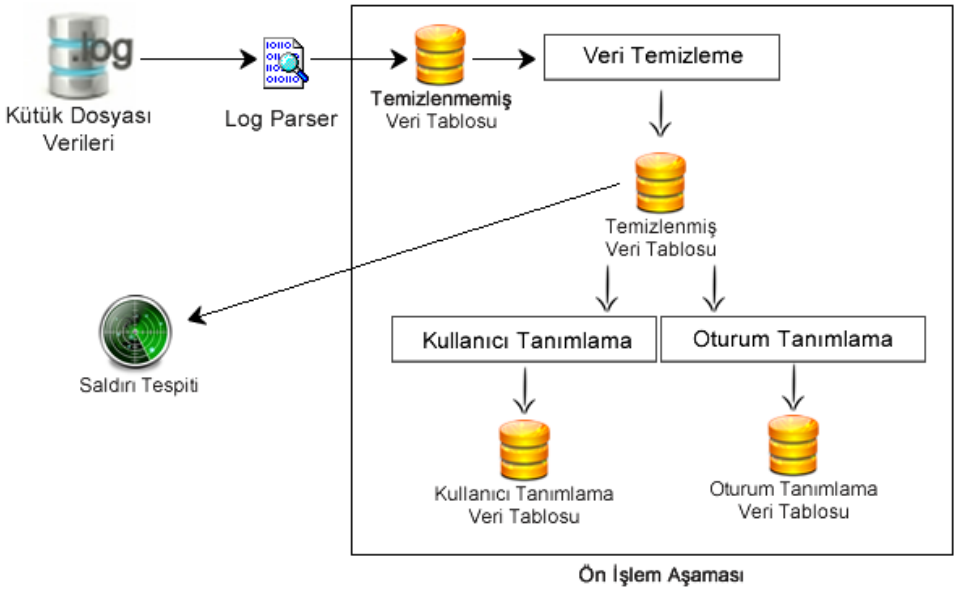
Veri tabanı ayarlarının yapılmasının ardından yazılım, kullanıcıyı web kullanım madenciliği işlemlerinin yapıldığı ana forma yönlendirmektedir. Şekil-14'de bu forma ait ekran görüntüsü yer almaktadır.



Şekil 14: Web kullanım madenciliği işlemleri ve saldırı tespitinin gerçekleştirildiği ana işlem formu.

Kullanıcı, ham erişim verilerinin veri tabanına aktarılması işlemini, aktarılan ham verilerin temizlenmesi işlemini, kullanıcı ve oturum tanımlama işlemlerini Şekil-14'de gösterilen ana formdan yapabilmektedir. Ayrıca bölüm 2.2'de bahsedildiği gibi SQL enjeksiyon ve XSS saldırılarının tespiti de bu ana form üzerinden yapılabilmektedir.

LCS yazılımına ait genel mimari Şekil-15'de resmedilmiştir.



Şekil 15: LCS yazılımına ait genel mimari.

3.2.1 Verilerin SQL Veri Tabanına Aktarılması

Şekil 14'deki formda, analiz edilecek kütük dosyalarının seçilmesi ve veri tabanı tablo adının ayarlanması gibi ayarlar yapılmaktadır. Kullanıcı kütük dosyalarını tek tek seçebileceği gibi klasör bazında bir seçim de yapabilmektedir. Kullanıcı, analiz edilecek kütük dosyalarını seçtikten sonra oluşturulacak veri tabanı tablosunun ismini ilgili alana girmelidir. Ayrıca kullanıcının seçmiş olduğu kütüklere ait kütük formatı da form üzerinden seçilmelidir.

Bu çalışmada Windows sunucularına ait W3C, IIS ve NCSA kütük formatları analiz edilebilmektedir. Aynı türden bilgileri içeren bu formatlar, alan isimlerinin farklı olması ve içerdikleri bilgilerin genişliği bakımından birbirlerinden ayrılmaktadırlar.

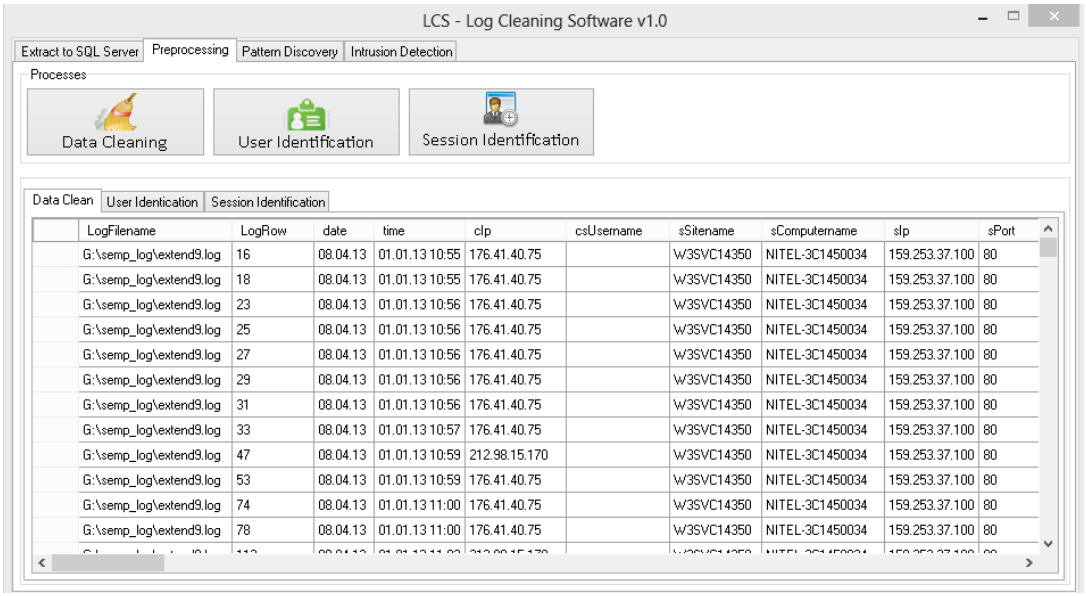
Kütük dosyalarının seçimi, veri tabanı tablo isminin girilmesi ve kütük formatının seçilmesi işlemlerinden sonra kütük dosyasında bulunan bütün erişim satırları, oluşturulan veri tabanında, oluşturulan tabloya aktarılmaktadır. Erişim kütüklerinin veri tabanına aktarılması işleminde, Microsoft Log Parser yazılımına ait "LogParser.dll" ActiveX nesnesi kullanılmıştır. Veri tabanı tablosuna aktarılan erişim kütük satırları, ön işlem aşaması için hazır hale gelmiştir. Veri tabanı tablosuna aktarılan erişim kayıtları, web kullanım madenciliği ön işlem aşamaları olan veri temizleme, kullanıcı tanımlama ve oturum tanımlama adımlarından geçirilmelidir.

3.2.2 Ön İşlem Aşaması

Erişim kütükleri öncelikle Log Parser kullanılarak veri tabanına parçalanmış şekilde aktarılmaktadır. Aktarılan bu veriler, SQL veri tabanında bir tabloda tutulmaktadır. Ham erişim verilerinin tutulduğu bu tablodaki verilerin, web kullanım madenciliği ön işlem aşamasının ilk adımı olan veri temizleme adımından geçirilmesi gerekmektedir. Veri temizleme adımından geçirilip, değersiz satırlardan temizlenmiş olan bu veriler yeni bir veri tabanı tablosuna aktarılmıştır. Uygulanan her adım sonrasında, veri tabanında ilgili adıma ait veri tabanı tablosu oluşturularak, işlenen veriler bu tablolara aktarılmıştır.

Geliştirilen bu yazılım ile web madenciliği ön işlem aşamasının yanı sıra kütük dosyası satırları üzerinde saldırı tespiti yapılmıştır. Temizlenmiş satırlar, XSS ve SQL enjeksiyon saldırılarında kullanılabilecek kötücül sözcüklerle karşılaştırılmış, eşleşen kayıtlar saldırı olarak işaretlenmiş ve bu kayıtlar uygulama ara yüzünde listelenerek resmedilmiştir.

Ön işlem aşamasına ait uygulama ara yüzü Şekil-16'da gösterilmiştir. Uygulamanın bu yüzünde veri temizleme, kullanıcı tanımlama ve oturum tanımlama adımlarına ait sekmeler bulunmaktadır. Bu sekmeler, ilgili adımlara ait kodların çalıştırılmasından sonra veri tabanından sorgulanan bilgilerle doldurulmaktadır.



Şekil 16: Ön işlem aşamasına ait uygulama ara yüzü.

3.2.2.1 Verilerin Temizlenmesi

Erişim kayıtlarındaki ham veriler, web kullanım madenciliği süreci için uygun formatta değildir. Bu verilerin, madencilik sürecinde işlenebilir veriler olabilmesi için birtakım aşamalardan geçirilmesi ve madencilik süreci için kullanılmayan, değeri olmayan ilgisiz kayıt satırlarının temizlenmesi gerekmektedir. Kullanıcılar bir web sitesini ziyaret ettiklerinde sadece istekte bulundukları sayfayı değil, sayfa içerisinde gömülü diğer kaynakları da ziyaret etmiş olmaktadırlar. Sayfa içerisinde bulunan resim dosyaları, çoklu ortam dosyaları, web sayfasının tasarımı ile ilgili bilgileri içeren stil dosyaları, betik dosyaları gibi dosyalar da, erişim sonrasında kütük dosyasında birer satır olarak yer almaktadır. Bu satırların, madencilik süreci gereği temizlenmesi gerekmektedir.

Geliştirilen yazılımla, erişim kütük dosyalarından madencilik süreci içerisinde değeri olmayan bu satırlar, bir SQL sorgusu yardımıyla filtrelenerek yeni bir SQL veri tablosuna aktarılmaktadır. Böylece oluşan yeni veri tablosu, madencilik sürecinde gereksiz verilerden

temizlenmiş olmaktadır. Yazılımın, bu adımıyla ilgili genel yapısı Şekil-17'de gösterilmiştir.

Şekil 17: Ham kütük verilerinin temizlenmesi adımı.

LCS tarafından bir önceki adımda veri tabanına aktarılan ham veriler, SQL sorgusuyla ilgisiz satırlardan temizlenerek, yeni bir veri tabanı tablosuna kaydedilmektedir. Veri temizleme aşamasında elde edilen temiz kayıtlar Tablo-34'de gösterilmektedir.

Tablo 34: Veri temizleme adımından sonra elde edilen temiz erişim satırları.

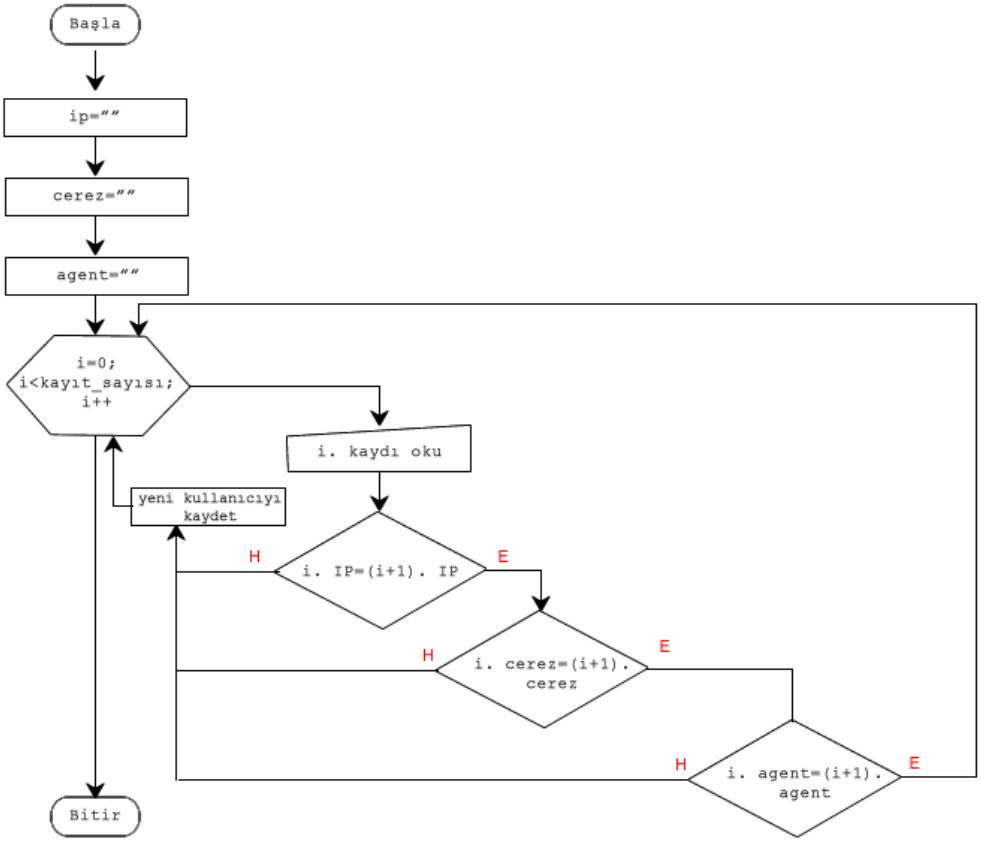
	date	time	clp	csMethod	csUriStem	csUriQuery	scStatus
1	2007-03-30 00:00:00.000	2013-01-01 10:07:52.000	88.229.205.159	GET	/iletisim.asp	NULL	200
2	2007-03-30 00:00:00.000	2013-01-01 14:01:48.000	65.54.188.55	GET	/iletisim.asp	NULL	200
3	2007-03-30 00:00:00.000	2013-01-01 14:01:53.000	65.54.188.55	GET	/yonetim.asp	NULL	200
4	2007-03-30 00:00:00.000	2013-01-01 14:19:11.000	88.245.65.22	GET	/misyon.asp	NULL	200
5	2007-03-30 00:00:00.000	2013-01-01 14:19:29.000	88.245.65.22	GET	/iletisim.asp	NULL	200
6	2007-03-30 00:00:00.000	2013-01-01 16:07:29.000	88.64.151.239	GET	/misyon.asp	NULL	200
7	2007-03-30 00:00:00.000	2013-01-01 16:07:50.000	88.64.151.239	GET	/misyon.asp	NULL	200
8	2007-03-30 00:00:00.000	2013-01-01 16:08:17.000	88.64.151.239	GET	/misyon.asp	NULL	200
9	2007-03-30 00:00:00.000	2013-01-01 16:08:26.000	88.64.151.239	GET	/iletisim.asp	NULL	200
10	2007-03-30 00:00:00.000	2013-01-01 17:06:35.000	88.244.22.1	GET	/misyon.asp	NULL	200

Veri temizleme aşamasından sonra elde edilen temiz erişim satırları, sonraki aşamalar için kullanılmaktadır. Kullanıcı tanımlama ve oturum tanımlama aşamaları için veri temizleme adımından elde edilen veri tabanı tablosundan yararlanılacaktır.

3.2.2.2 Kullanıcıların Tanımlanması

Bu adımda amaç, veri temizleme adımından sonra oluşan verilerin hangi kullanıcılara ait olduğunun tespit edilmesidir. Kullanıcı, web sayfasına kullanıcı adı ve şifresiyle giriş yaptığında kullanıcı tanımlama adımının uygulanması daha kolay olacaktır fakat çoğu web sitesinde yayınlanan içerikler, web sitesi üyelerinin dışında anonim kullanıcılar ile de paylaşıldığından, bu adımda kullanıcıların birbirlerinden ayırt edilmesi gerekmektedir. Bir web sitesine üyelik bilgileri ile giriş yapan kullanıcının erişim bilgileri, erişim kütüklerine

kullanıcı adı bilgisiyle kaydedilmektedir. Fakat bu işlem, anonim kullanıcılar için geçerli olmamaktadır. Anonim kullanıcıların web sitelerine erişimleri sonucunda erişim kayıtlarında, bir nevi kullanıcı adı olarak kullanabilecek bazı bilgiler kayıt altına alınmaktadır. Bunlar kullanıcıya ait kullanıcı etmen bilgisi, çerez bilgisi ve IP adresleri gibi bilgilerdir. Aynı yerel ağ içerisinden dış ağlara tek bir IP adresi üzerinden, birçok kullanıcı erişim yapabilmektedir. İlk bakışta bu kullanıcıların IP adreslerinin aynı olmasından dolayı birbirlerinden ayırt edilmesi imkânsız gibi görülmektedir. Fakat bu kullanıcıların web sitelerine erişimleri sonucunda erişim kayıtlarında aynı IP adresi görülse de kullanıcıların, kullanıcı etmen bilgileri ve çerez bilgileri birbirlerinden farklı olabilmektedir. Bu farklılıktan dolayı kullanıcılar birbirlerinden ayırt edilebilmektedir. Geliştirilen bu yazılımda, kullanıcı etmen bilgileri ve çerez bilgileri göz önüne alınarak kullanıcılar birbirlerinden ayırt edilmiş ve kullanıcı tanımlamaları yapılmıştır. Yazılımın bu adımla ilgili kod parçasının akış diyagramı Şekil-18'de, sözde kodu ise Tablo 35'te gösterilmiştir. Akış diyagramına göre temizlenmiş veri tablosundaki tüm veriler, bir döngü içerisine alınmıştır. Döngü içerisinde öncelikle her kayıta karşılık gelen IP adresi bir değişkene atanmış ve bir sonraki kayıtın IP adresi bilgisi ile karşılaştırılmıştır. Karşılaştırılan IP adresleri aynı değilse sonraki kayıta ait IP, yeni bir kullanıcı olarak SQL veri tablosuna kaydedilmiştir. Fakat IP adresleri aynı ise ilgili kayıtlara ait çerez bilgileri karşılaştırılmıştır. Karşılaştırma sonuçları aynı değilse ilgili kayıta ait IP adresi yeni bir kullanıcı olarak kaydedilmiştir. Fakat ilgili kayıta ait çerez bilgileri aynı ise bu kayıtlara ait kullanıcı etmen bilgilerinin karşılaştırılması gerekmektedir. Bu karşılaştırma sonucunda ilgili kayıtlara ait kullanıcı etmen bilgileri farklı ise ilgili kayıta ait IP adresi yeni kullanıcı olarak veri tablosuna kaydedilmiştir. Fakat karşılaştırma sonucunda kullanıcı etmen bilgileri de aynı ise herhangi bir işlem yapılmadan sonraki kayıtları karşılaştırmak için döngüye devam edilmiştir.



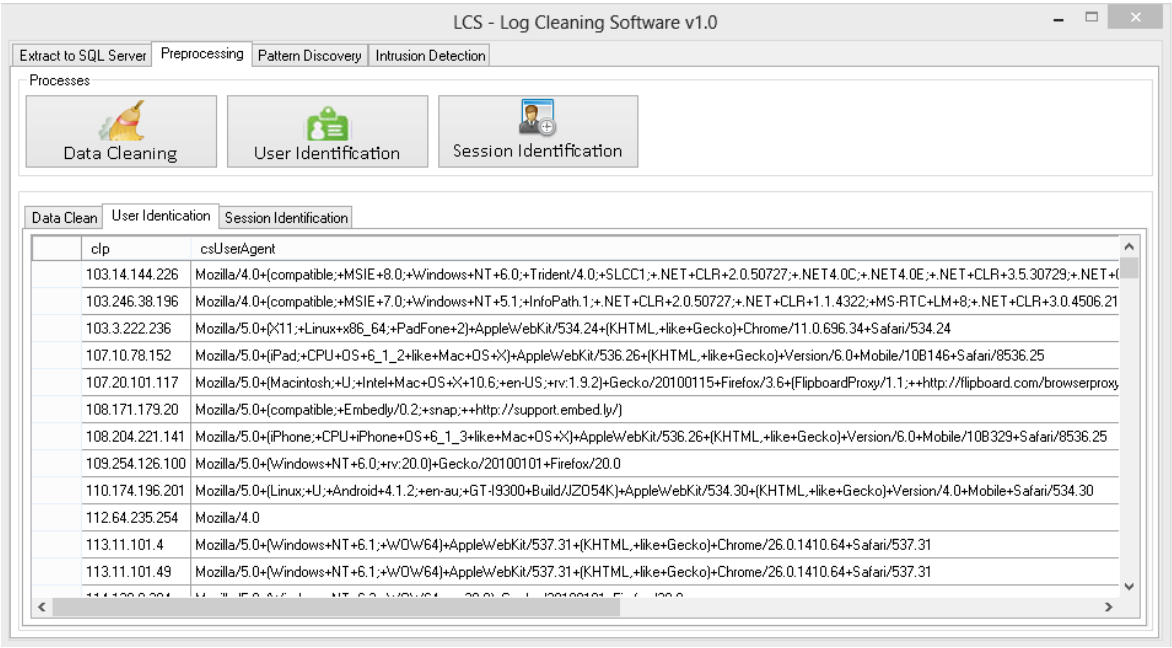
Şekil 18: Kullanıcı tanımlama akış diyagramı.

Tablo 35: Kullanıcı tanımlama adımına ait sözde kod.

1	**Başla**
2	IP, Agent, Cerez değişkenlerini tanımla.
3	**for**(int i=0; i<kayit_sayisi; i++)
4	**if**(IP[i] == IP[i+1])
5	**if**(Cerez[i] == Cerez[i+1])
6	**if**(Agent[i] == Agent[i+1]) return;
7	**else**
8	Yeni_kullanici_kaydet();
9	**else**
	Yeni_kullanici_kaydet();
10	**else** Yeni_kullanici_kaydet();
11	**Bitir**

Temizlenmiş verilerin bulunduğu veri tabanı tablosundan, kullanıcılara ait kullanıcı etmen bilgisi, çerez bilgisi ve IP bilgisi göz önüne alınarak kullanıcılar seçilmiştir. Elde edilen veriler, yeni bir veri tabanı tablosu oluşturularak veri tabanına kaydedilmiştir. Yazılım tarafından gerçekleştirilen bu işlem sonucunda, elde edilen verilerden örnek bir kesit Tablo-36'da gösterilmiştir.

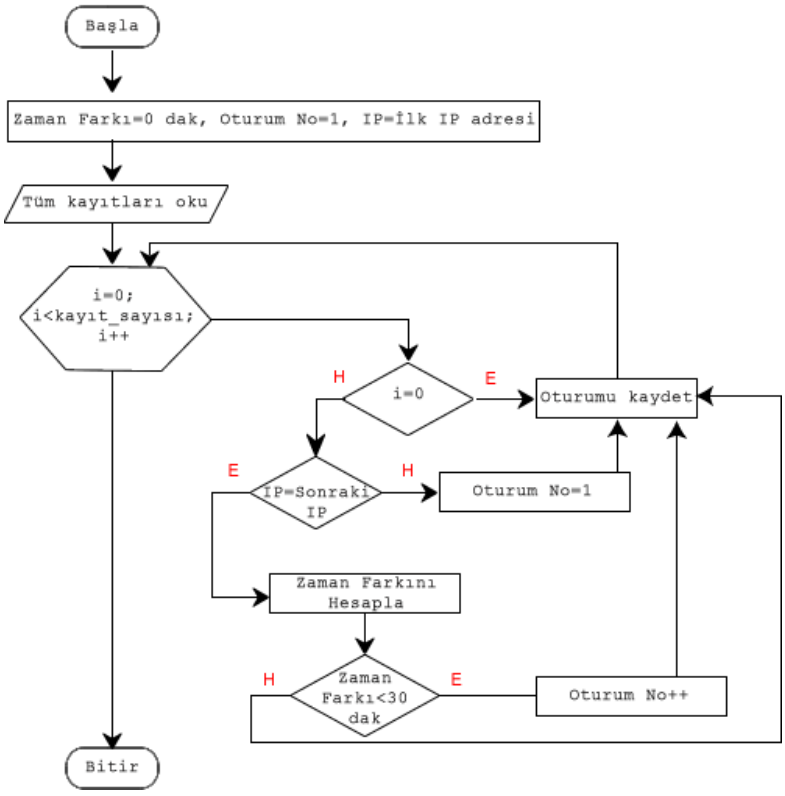
Tablo 36: Kullanıcı tanımlama adımından sonra elde edilen satırlar.



3.2.2.3 Oturumların Tanımlanması

Oturum, bir kullanıcının bir web sitesine girişinden çıkışına kadar web sitesinde gerçekleştirdiği tüm işlemlerin kümesi olarak veya bir web sitesine giren kullanıcıların oturumlarının gruplandırılması olarak tanımlanabilir. Oturum tanımlama adımı için oturum süresi temelli (session-duration-based), sayfada kalma süresi temelli (page-stay-time-based method) ve referans temelli (referrer-basic heuristic method) olmak üzere çeşitli sezgisel yöntemler kullanılabilmektedir [18]. Bu çalışmada, oturum süresi temelli sezgisel yaklaşım kullanılmıştır.

Geliştirilen uygulamada bulunan oturum tanımlama işlemine ait akış diyagramı Şekil-19'da, sözde kodu ise Tablo 37'de gösterilmektedir.



Şekil 19: Oturum tanımlama akış diyagramı.

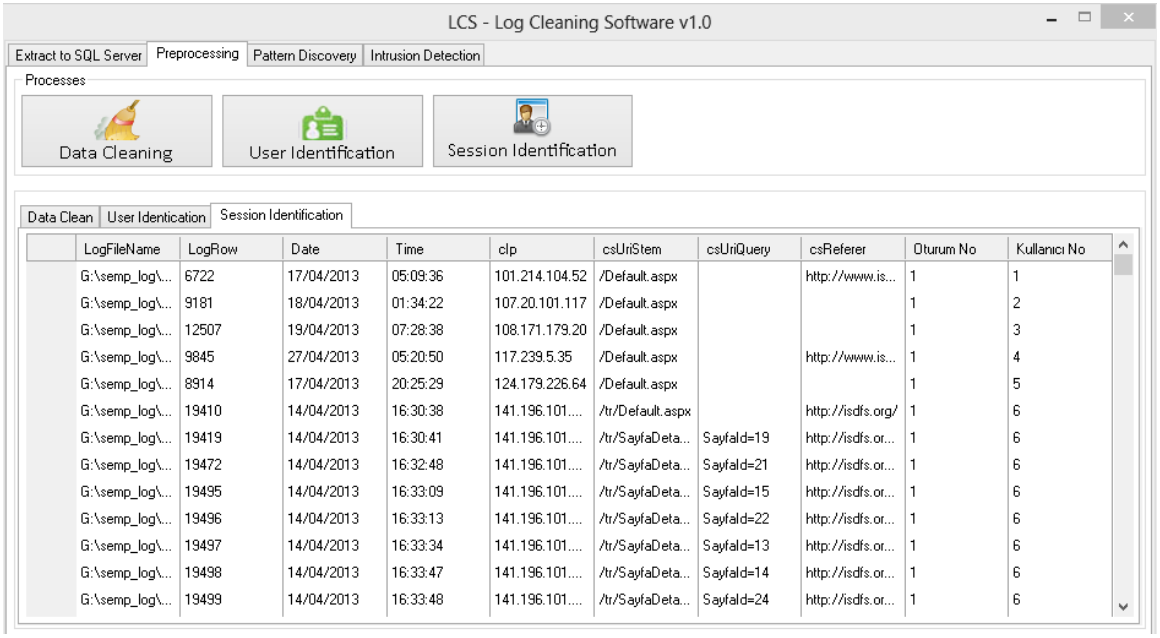
Tablo 37: Oturum tanımlama adımına ait sözde kod.

1	**Başla**
2	Zaman_Farki, Oturum_No=1, IP değişkenlerini tanımla.
3	**for**(int i=0; i<kayit_sayisi; i++)
4	**if**(i=0)
5	i. Oturumu_Kaydet();
6	**else**
7	**if**(IP[i] == IP[i+1])
8	**if**(Zaman_Farki<30 dak) Oturum_No++;
9	**else** i. Oturumu_Kaydet();
10	**else** Oturum_No=1; i. Oturumu_Kaydet();
11	**Bitir**

Oturum tanımlama adımında, temizlenmiş veri tablosundaki verilerden yararlanılmıştır. Şekil-19'daki akış diyagramına göre ilk satırdaki kayıta ait IP bilgisi, bir değişkene atanmıştır. Daha sonra bir döngü yardımıyla tüm satırlar IP değişkeni ile karşılaştırılmıştır. Bu karşılaştırmada IP değişkeni bir sonraki kayıtın IP bilgisi ile aynı değilse oturum numarası 1 olarak okunan kayıtın oturum kaydı gerçekleştirilmiştir. Fakat IP değişkeni, sonraki kayıttaki IP bilgisi ile aynı ise bu iki kayıtın arasındaki zaman farkı hesaplanmıştır. Eğer zaman farkı 30 dakikadan az ise oturum numarası bir arttırılıp oturum kaydedilmiştir. Eğer zaman farkı 30 dakikadan fazla ise oturum numarası 1 olarak okunan kayıtın oturum kaydı gerçekleştirilmiştir.

Tablo-38'de oturum tanımlama adımından geçirilmiş kütük verilerinden sonra oluşan tablo gösterilmiştir.

Tablo 38: Oturum tanımlama adımından sonra elde edilen satırlar.



LogFileName	LogRow	Date	Time	clp	csUriStem	csUriQuery	csReferer	Oturum No	Kullanıcı No
G:\semp_log\...	6722	17/04/2013	05:09:36	101.214.104.52	/Default.aspx		http://www.is...	1	1
G:\semp_log\...	9181	18/04/2013	01:34:22	107.20.101.117	/Default.aspx			1	2
G:\semp_log\...	12507	19/04/2013	07:28:38	108.171.179.20	/Default.aspx			1	3
G:\semp_log\...	9845	27/04/2013	05:20:50	117.239.5.35	/Default.aspx		http://www.is...	1	4
G:\semp_log\...	8914	17/04/2013	20:25:29	124.179.226.64	/Default.aspx			1	5
G:\semp_log\...	19410	14/04/2013	16:30:38	141.196.101....	/tr/Default.aspx		http://isdfs.org/	1	6
G:\semp_log\...	19419	14/04/2013	16:30:41	141.196.101....	/tr/SayfaDeta...	Sayfaİd=19	http://isdfs.or...	1	6
G:\semp_log\...	19472	14/04/2013	16:32:48	141.196.101....	/tr/SayfaDeta...	Sayfaİd=21	http://isdfs.or...	1	6
G:\semp_log\...	19495	14/04/2013	16:33:09	141.196.101....	/tr/SayfaDeta...	Sayfaİd=15	http://isdfs.or...	1	6
G:\semp_log\...	19496	14/04/2013	16:33:13	141.196.101....	/tr/SayfaDeta...	Sayfaİd=22	http://isdfs.or...	1	6
G:\semp_log\...	19497	14/04/2013	16:33:34	141.196.101....	/tr/SayfaDeta...	Sayfaİd=13	http://isdfs.or...	1	6
G:\semp_log\...	19498	14/04/2013	16:33:47	141.196.101....	/tr/SayfaDeta...	Sayfaİd=14	http://isdfs.or...	1	6
G:\semp_log\...	19499	14/04/2013	16:33:48	141.196.101....	/tr/SayfaDeta...	Sayfaİd=24	http://isdfs.or...	1	6

3.2.3 Örüntü Keşfi

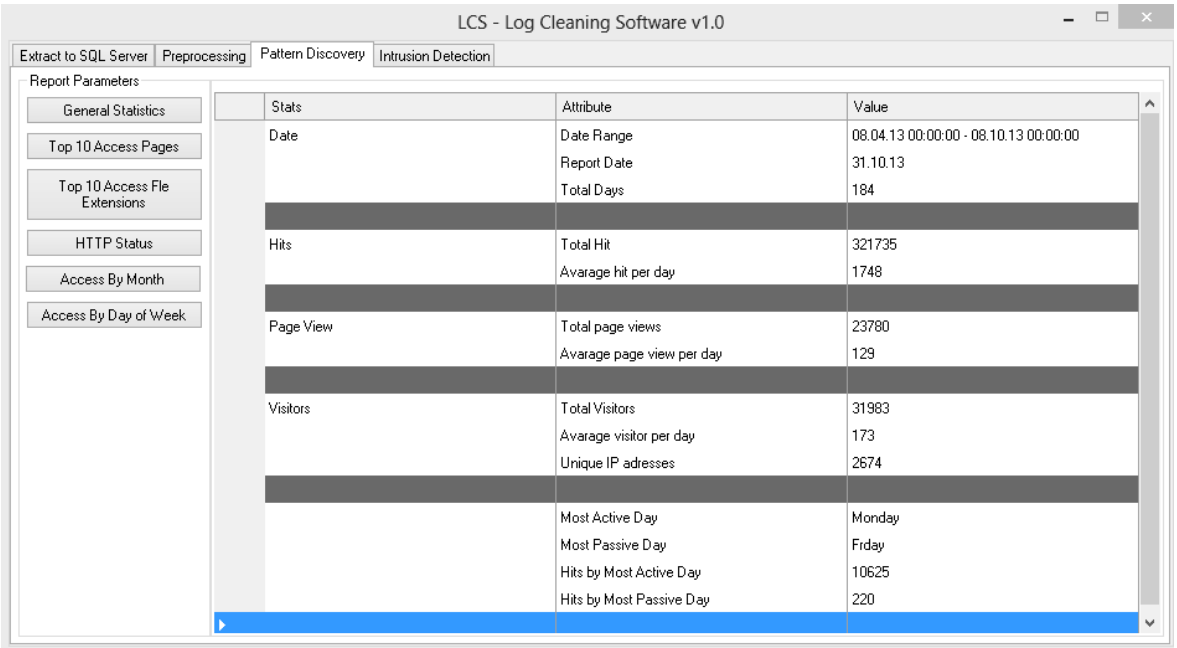
Tez çalışmasının bu bölümünde erişim kütük kayıtları, istatistiksel analiz yöntemi kullanılarak analiz edilmiştir. Analiz sonucunda elde edilen bilgilerle web sitesi kullanıcılarının davranışlarına yönelik önemli verilere ulaşılmıştır. Elde edilen bu verilerle, bir web sitesine ait birçok bilgi çıkarılabilmektedir. Erişim kütük verilerinden çıkarılabilecek bu bilgiler, web sitesinin amacı doğrultusunda değişiklik gösterebilmektedir. Örneğin; elektronik ticaret hizmeti sunan bir web sitesinin, erişim

kayıtları analiz edilerek elde edilen veriler, kullanıcıların ilgi alanlarına göre ilişkilendirilerek, kullanıcıların web sitesi içerisindeki araştırdığı ürünler ile ilgili kampanyalardan haberdar edilebilmesi için kullanılabilir. Eğitim içerikli bir web sitesinin erişim kayıtlarının analiz edilmesi ile web sitesinden eğitim amaçlı yararlanan öğrencilerin davranışları incelenebilir. Bu sayede ders materyallerinin düzenlenmesi, web sitesinde bulunan içeriklere, öğrencilerin daha kolay erişebilmesi için site tasarımının yenilenmesi ya da var olan web site tasarımının geliştirilmesi, web sitesinde bulunan hata sayfalarının ve kırık bağlantıların tekrar yapılandırılarak web sitesinin performansının arttırılması gibi site başarımını ve eğitim kalitesini arttırıcı birçok işlem uygulanabilir.

3.2.3.1 Genel İstatistikler

Geliştirilen LCS yazılımı ile temizlenmiş veri tablosundan elde edilen genel erişim istatistikleri Tablo-39'da gösterilmiştir.

Tablo 39: LCS yazılımından elde edilen genel istatistikler.



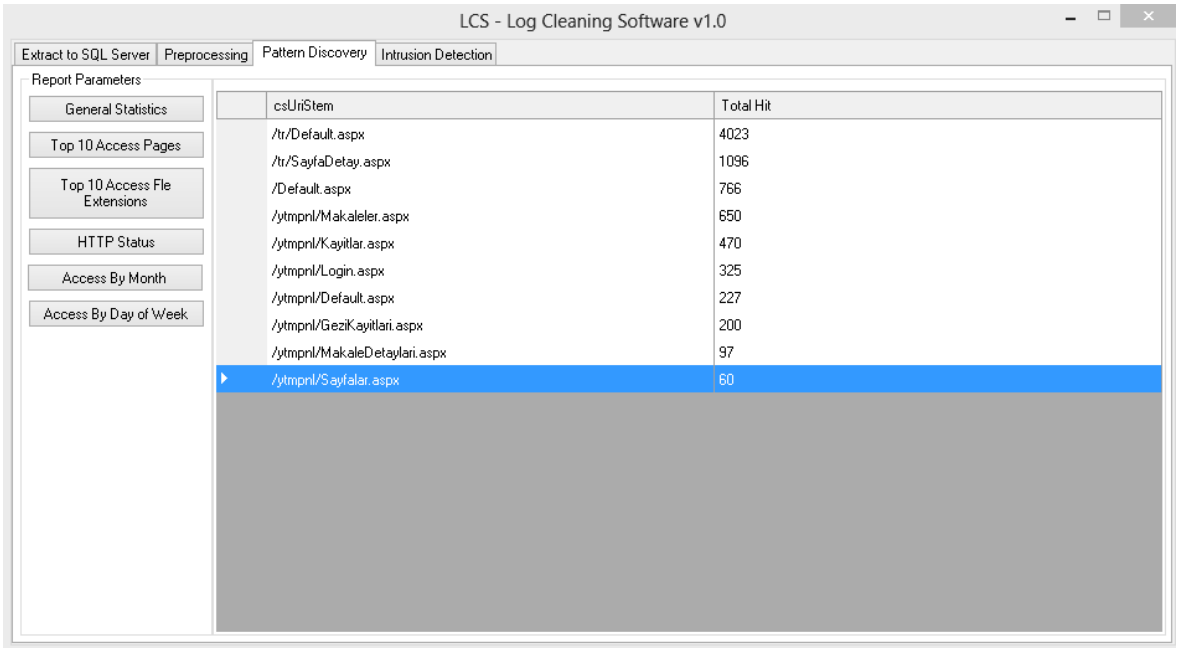
Bu istatistiklerden çıkarılan bilgilerle, web sitesine ait toplam istek sayısı, sayfa görüntüleme sayısı, toplam ziyaretçi sayısı, en çok aktivitenin yapıldığı gün ve bu güne ait toplam istek sayısı, en az aktivitenin yapıldığı gün ve bu güne ait toplam istek sayısılarına ulaşılabilmektedir. Çıkarılan istatistiklere göre web sitesine en çok erişim yapılan günün

Pazartesi, en az erişim yapılan günün ise Cuma günü olduğu gözlemlenmiştir. Bu bilgiler ışığında, web sitesinde yapılacak herhangi bir bakım ya da web sitesi yapılandırma çalışmasının Cuma günü yapılması tercih edilebilir.

3.2.3.2 En Çok Erişilen Sayfalar

Erişim kütükleri analiz edilen web sitesine ait en çok erişilen sayfalar ve erişim sayıları Tablo 40'da gösterilmiştir. Elde edilen bilgilere göre web sitesinin ana sayfası, kullanıcılarca en çok ziyaret edilen sayfa olarak görülmektedir. Bu bilgiye göre kullanıcılar, web sitesini ziyaret ettiklerinde, ulaşmak istedikleri bilgileri genellikle web sitesinin ana sayfasından elde etmiştir. Web sitesinin ana sayfasının sık kullanılıyor olması, web sitesi yöneticilerine bu sayfanın daha etkin kodlanması için fikir verebilir.

Tablo 40: En çok erişilen sayfalar.

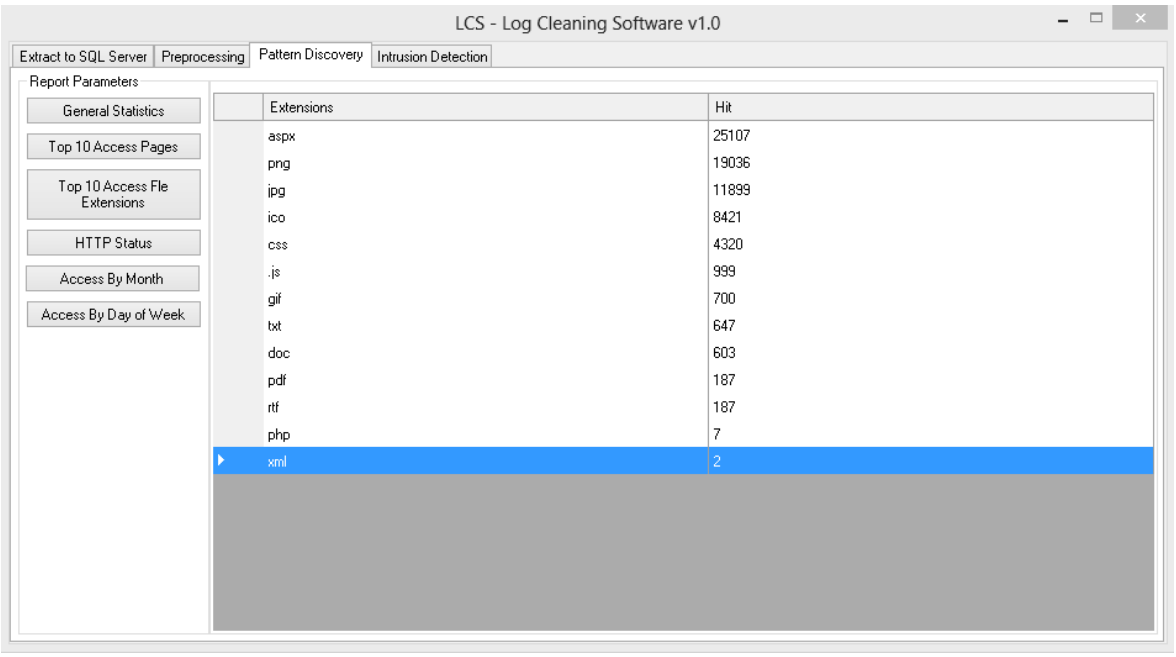


3.2.3.3 En Çok Erişilen Dosya Uzantıları

Erişim kayıtları analiz edilen web sitesinden en çok talep edilen dosya uzantıları Tablo-41'de gösterilmiştir. Elde edilen istatistiklere göre web sitesinin yapıldığı web programlama dilinin dosya uzantısı olan ".aspx" uzantısı en çok talep edilen dosya uzantısıdır. Diğer dosya uzantıları incelendiğinde, resim dosyası uzantılarının da çok talep edildiği görülmüştür. Web sitesi yöneticilerinin bu bilgiyi göz önünde bulundurarak, web

sitesinde kullanılan resim dosyalarının boyutlarının en az olarak ayarlaması, bant genişliğini düşüreceği için kullanıcılar web sitesinde daha hızlı gezinebilirler.

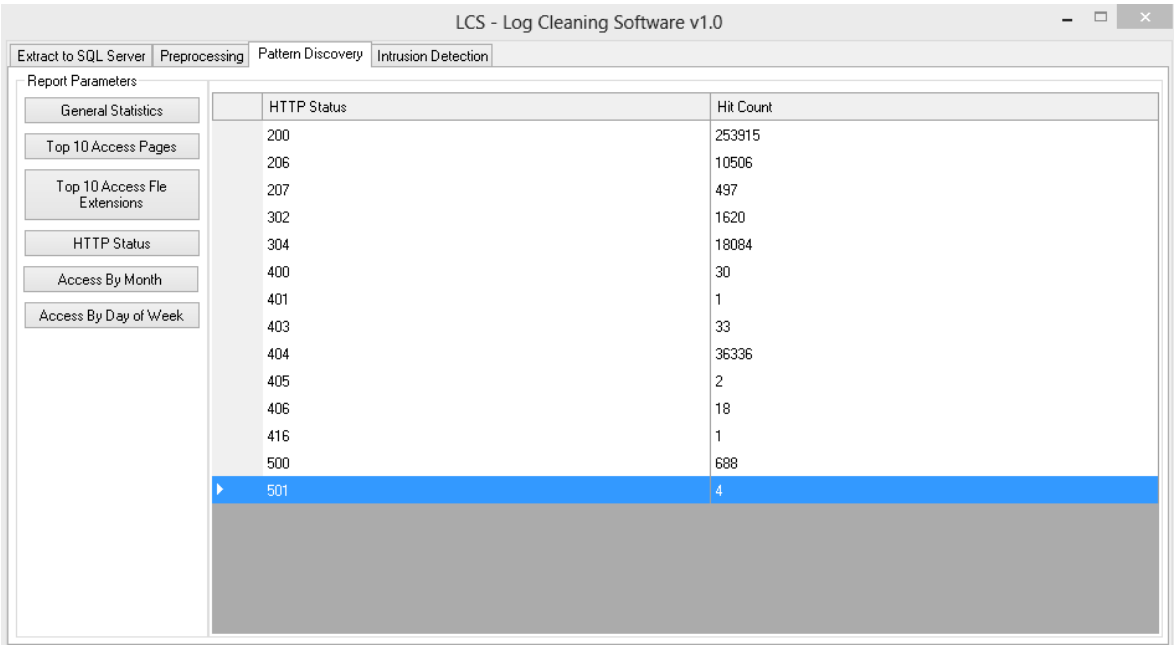
Tablo 41: En çok erişilen dosya uzantıları.



3.2.3.4 Web İsteklerinin HTTP Durum Kodlarına Göre Dağılımları

Kullanıcılar bir web sitesini ziyaret ettiğinde, kullanıcı işleminin sonucuna göre erişim kayıtlarına HTTP durum kodu bilgisi de kaydedilmektedir. Bu durum kodları sayesinde web sitesine yapılan başarılı ve başarısız erişim kayıtlarına ulaşabilmektedir. Tablo-42'de, web sitesine yapılan isteklerin HTTP durum koduna göre dağılımları gösterilmiştir.

Tablo 42: Web sitesine yapılan isteklerin HTTP durum koduna göre dağılımları.

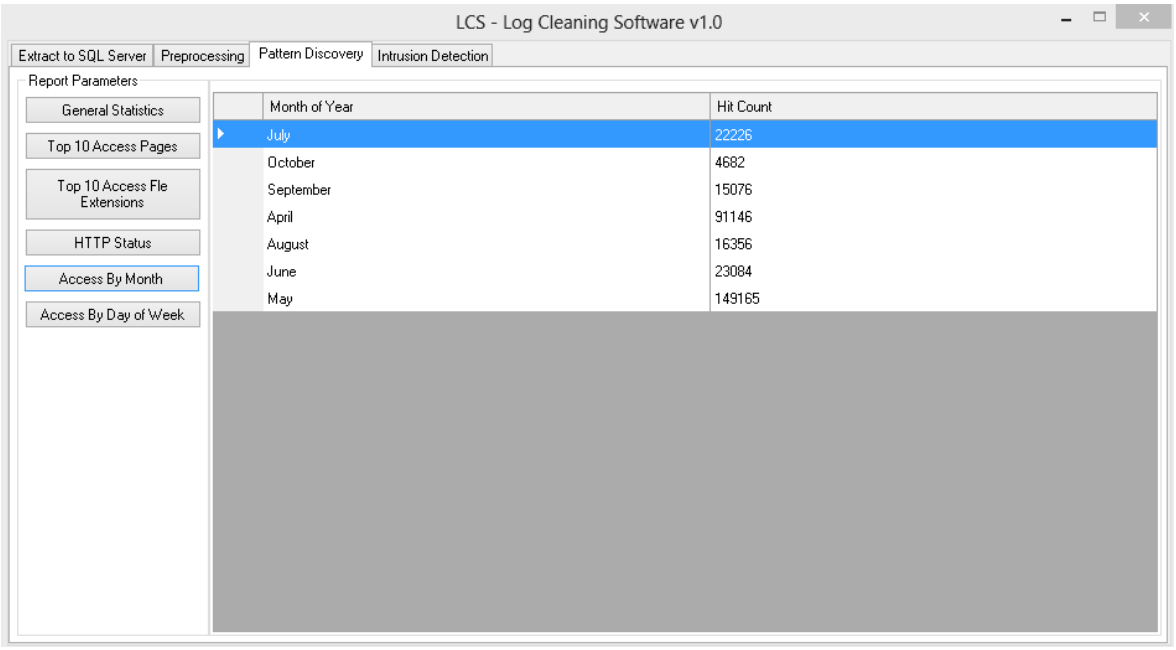


Elde edilen bilgilere göre "404 bulunamadı" istek sayısının fazla olduğu görülmektedir. Bu da web sitesinde kırık linklerin fazla olduğunu göstermektedir. Bu bilgiler ışığında, web sitesi yöneticisinin kırık linkleri tespit edip web sayfasını yeniden yapılandırması, web sitesinin ziyaret potansiyelini ve içerik kalitesini arttırabilir.

3.2.3.5 Web İsteklerinin Aylara Göre Dağılımları

Web sitesinden elde edilen erişim kayıtlarının, aylara göre istek sayısı dağılımları Tablo-43'de gösterilmiştir.

Tablo 43: Web isteklerinin aylara göre dağılımları.

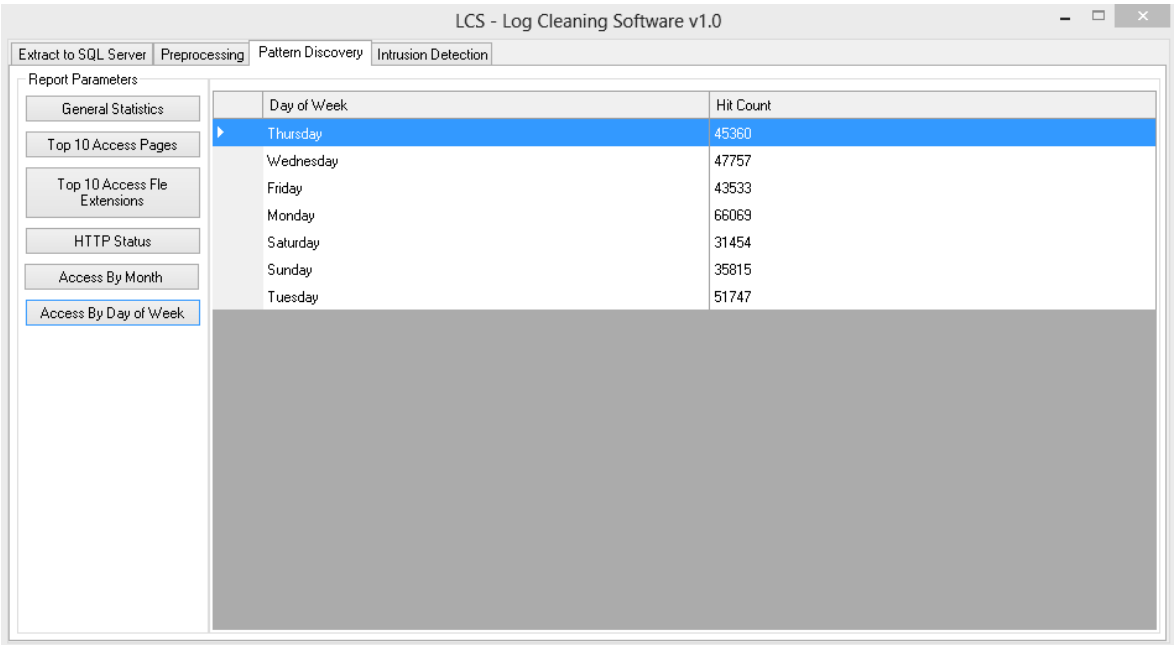


Elde edilen bilgiler incelendiğinde erişim istek sayısının Mayıs ayında, diğer aylara oranla daha fazla olduğu görülmektedir.

3.2.3.6 Web İsteklerinin Haftanın Günlerine Göre Dağılımları

Erişim kayıtlarından elde edilen istatistiksel bilgiler doğrultusunda web sitesine yapılan isteklerin, haftanın günlerine göre dağılımları Tablo-44'de gösterilmiştir.

Tablo 44: Web isteklerinin haftanın günlerine göre dağılımları.



Elde edilen verilere göre web sitesine en çok isteğin yapıldığı gün Pazartesi günüdür. Bu sebeple, web sitesinde herhangi bir bakım çalışması veya yapılandırmanın Pazartesi günü yapılması, diğer günlere göre daha çok kullanıcının web sitesine erişmesinde problem oluşturacaktır.

3.2.4 Saldırı Tespiti Uygulaması

Bu çalışmada gerçekleştirilen yazılım ile web kütük dosyalarında bulunan satırlardan XSS ve SQL enjeksiyon saldırılarının tespiti yapılmıştır. Kötü niyetli kişilerin tarayıcı adres çubuğuna bu saldırıları gerçekleştirmek için yazdığı kodlar, erişim kütük kayıtlarına kaydedilmektedir. Kütük kayıtlarında bulunan bu satırlar, saldırı türüne göre belirlenen ve program üzerinde tanımlanan dizilerle karşılaştırılmış, eşleşen kayıtlar program ara yüzünde listelenmiştir.

3.2.4.1 XSS (Cross Site Scripting)

Tablo-45'de virgüllerle ayrılmış olan sözcükler, erişim kayıtlarıyla karşılaştırılmıştır. Tabloda belirtilen sözcükler, XSS saldırı yöntemi için uygulamada tanımlanan kara listeyi oluşturmaktadır. Bu sözcüklerle eşleşen erişim kayıtları, saldırı olarak işaretlenmekte ve uygulama ekranında listelenmektedir. Ancak saldırı amacı taşımayan erişim kayıtlarından

bazıları, kara listedeki sözcüklerle eşleşebileceğinden uygulama ekranında saldırı amaçlı yapılmış kayıtlarla beraber listelenecektir.

Tablo 45: Filtrelenen XSS sözcükleri.

XSS
javascript, vbscript, expression, applet, meta, xml, blink, style, script, embed, object, iframe, frame, frameset, ilayer, layer, bgsound, title, base, onabort, onactivate, onafterprint, onafterupdate, onsubmit, onunload, alert, onclick

3.2.4.2 SQL Enjeksiyon

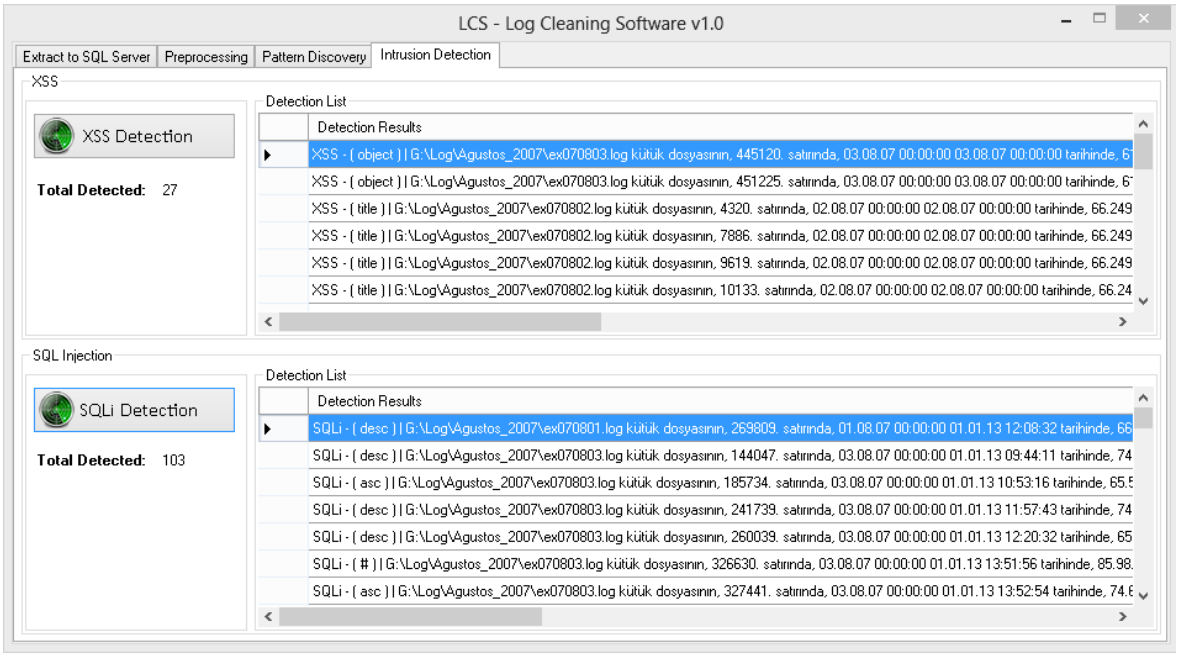
Tablo-46'da virgüllerle ayrılmış olan sözcükler, SQL enjeksiyon saldırı yönteminde kullanılabilecek sözcüklerdir. Bu sözcükler, erişim kayıtlarıyla karşılaştırılmış eşleşen kayıtlar uygulama ara yüzünde listelenerek gösterilmiştir.

Tablo 46: Filtrelenen SQL sözcükleri.

SQL Enjeksiyon
*', \, --, ' --, --';, ' ;, = ', = ;, = --, \x23, \x27, \x3D \x3B', \x3D \x27, \x27\x4F\x52 SELECT *, \x27\x6F\x72 SELECT *, 'or select *, admin'--, ';shutdown--, <>\'%;)(&+, ' or ''=', ' or 'x'='x, \ or \x\=\x, ') or ('x'='x, 0 or 1=1, ' or 0=0 --, \ or 0=0 --, or 0=0 --, ' or 0=0 #, \ or 0=0 #, or 0=0 #, ' or 1=1--, \ or 1=1--, ' or '1'='1'--, \' or 1 --'\, or 1=1--, or%201=1, or%201=1 --, ' or 1=1 or ''=', \ or 1=1 or \\=\, ' or a=a--, \ or \a\=\a, ') or ('a'='a, \) or (\a\=\a, hi\ or \a\=\a, hi\ or 1=1 --, hi' or 1=1 --, hi' or 'a'='a, hi') or ('a'='a, hi\) or (\a\=\a, 'hi' or 'x'='x';, @variable, ,@variable, PRINT, PRINT @@variable, select, insert, procedure, limit, order by, asc, desc, delete, update, distinct, having, truncate, replace, like, handler, bfilename, ' or username like '%, ' or uname like '%, exec xp, exec sp, '; exec master..xp_cmdshell, '; exec xp_regread, t'exec master..xp_cmdshell 'nslookup www.google.com'--, --sp_password, \x27UNION SELECT, ' UNION SELECT, ' UNION ALL SELECT, ' or (EXISTS), ' (select top 1, '||UTL_HTTP.REQUEST, 1;SELECT%20*, to_timestamp_tz, tz_offset, <>"'%;)(&+, '%20or%201=1, %27%20or%201=1, %20$(sleep%2050), %20'sleep%2050', char%4039%41%2b%40SELECT, '%20OR, 'sqlattempt1, (sqlattempt2), \|, %7C, *\|, %2A%7C, *(\|(mail=*)), %2A%28%7C%28mail%3D%2A%29%29, *(\|(objectclass=*)), %2A%28%7C%28objectclass%3D%2A%29%29, %28, %29, %26, !, %21, ' or 1=1 or ''=', ' or ''=', x' or 1=1 or 'x'='y, //, //*, */*, @*, SELECT @@VERSION, SELECT * from v$version;, union, delete, alter, having*

Tablo-47'de, erişim kayıtları analiz edilen web sitesine yapılan SQL enjeksiyon ve XSS saldırı girişimleri gösterilmiştir.

Tablo 47: SQL enjeksiyon ve XSS saldırılarının tespiti.



Tablo-48'de, erişim kayıtları analiz edilen web sitesine yapılan SQL enjeksiyon saldırı girişiminden örnek bir satır gösterilmiştir. Yapılan saldırı tespiti sonucunda, web sitesine yapılan saldırının türü, SQL enjeksiyon için kara listede tanımlanan hangi sözcük ya da karakterin kullanıldığı, saldırı girişiminin hangi kütük dosyasında kaçıncı satırda olduğu, saldırı girişimine ait tarih ve saat bilgisi, kullanıcıya ait IP adresi bilgisi ve saldırı girişiminin hangi sayfa üzerinden yapıldığı gösterilmiştir.

Tablo 48: SQL enjeksiyon saldırıları girişimine ait örnek satır.

SQLi - (') | G:\semp_log\extend10.log kütük dosyasının, 22641. satırında, 22.04.13 15:42:00 tarihinde, 78.186.46.161 IP numaralı kullanıcı /tr/SayfaDetay.aspxSayfaId=38' adresine SQLi saldırısı yapmış olabilir!

Kullanıcıların POST metodu ile yaptığı isteklerde, erişim kayıtlarına web formundan gönderilen sorgular kaydedilmemektedir. Bu yüzden erişim kütük dosyalarından elde edilen bilgiler ile tüm SQL enjeksiyon ve XSS saldırıları tespit edilememektedir. Ayrıca tarayıcı adres çubuğunda saldırı tespiti için filtrelenen sözcüklerin bazıları normal kullanımda kullanılabilecek sözcüklerle eşleşebileceği için saldırı tespitinde oluşan sonuçlar yüzde yüz doğru değildir. Örneğin; kullanıcı "/Default.asp?git=linkler" bağlantısını ziyaret ettiğinde, adres çubuğundaki "linkler" sözcüğü ile XSS kara

listesindeki “link” sözcüğü eşleşeceği için bu isteğe ait erişim kayıt satırı saldırı olarak algılanacaktır.

4. WEB KULLANICI ERİŞİM KÜTÜKLERİNDEN ELDE EDİLEN İSTATİSTİKSEL BİLGİLER

Akademik alanda hizmet sunan bir web sitesinden alınan W3C formatındaki erişim kütük dosyaları, 07 Nisan 2013 ile 08 Ekim 2013 tarihleri arasındaki erişim bilgilerini kapsamaktadır. Toplamda 11 adet, 107 MB boyutunda ve 321.735 satırdan oluşan erişim kütük dosyaları, LCS uygulamasıyla analiz edilmiş ve elde edilen bilgiler MSSQL veri tabanına aktarılmıştır. Bu süreçte, temizleme ve dönüşüm işlemlerine tabi tutulan erişim kütük dosyalarından geriye 23.780 satır kalmıştır. Çalışmanın bu bölümünde, geliştirilen LCS yazılımından elde edilen işlenmiş veriler, ODBC aracılığıyla *Nihuo Web Log Analyzer 4* yazılımına kaynak gösterilmiş ve erişim kayıtları analiz edilerek, detaylı istatistiksel bilgiler elde edilmiştir. İstatistiksel analizler sonucunda kullanıcı erişim kütüklerinden çıkarılan anlamlı bilgiler, web sitesinin başarımının ve performansının arttırılmasında kullanılabilmektedir.

4.1 Genel İstatistikler

Erişim kütük dosyalarından elde edilen web istek istatistikleri, sayfa gösterim istatistikleri, ziyaretçi istatistikleri ve bant genişliği istatistikleri özet olarak Tablo-49'da gösterilmiştir.

Tablo 49: Erişim kayıtlarına ait genel istatistikler.

Genel İstatistikler	
Web İstekleri	
Toplam İstek Sayısı	321.735
Kullanıcı İstek Sayısı	75.499
Örümcek-Bot İstek Sayısı	34.950
Günlük Ortalama İstek Sayısı	1.748
Ziyaret Başına Ortalama İstek Sayısı	10,06
Başarısız İstek Sayısı	37.113
Sayfa Gösterimi	
Toplam Sayfa Gösterimi	60.861
Günlük Ortalama Sayfa Gösterim Sayısı	332
Ziyaret Başına Düşen Ortalama Sayfa Gösterim Sayısı	1,90
Benzersiz Sayfa Gösterimi Sayısı	45.636
Ziyaret	
Toplam Ziyaret Sayısı	31.983
Kullanıcı Ziyaret Sayısı	14.183
Örümcek-Bot Ziyaret Sayısı	9.694
Günlük Ortalama Ziyaret Sayısı	174
Bant Genişliği	
Toplam Bant Genişliği	59,21 GB
Kullanıcı Bant Genişliği	24,46 GB
Örümcek-Bot Bant Genişliği	17,29 GB
Günlük Ortalama Bant Genişliği	331,32 MB
İstek Başına Ortalama Bant Genişliği	192,98 KB
Ziyaret Başına Ortalama Bant Genişliği	1,90 MB

4.2 Web Aktivite İstatistikleri

Bu istatistik sonucunda elde edilen bilgilerle, kullanıcıların hafta içi ve hafta sonunda web sitesine ortalama kaç kere istekte bulunduklarına, web sitesinin en aktif ve en pasif kullanıldığı günler ve bu günlere ait istek sayılarına, web sitesinin en aktif ve en pasif kullanılan saatlerine ve bu saatlere ait istek sayılarına ulaşılabilmektedir. Web aktivite istatistiklerine ait bilgiler Tablo-50'de gösterilmiştir. Ayrıca web sitesine ait kullanıcı erişim bilgileri, günlük ve aylık olarak tablo halinde Tablo-51 ve Tablo-52'de verilmiştir.

Tablo 50: Erişim verilerine ait aktivite istatistikleri.

Web Aktivite İstatistikleri	
Aktivite Toplamları	
Hafta içine Ait Ortalama Ziyaret Sayısı	185
Hafta içine Ait Ortalama İstek Sayısı	1.927
Hafta sonuna Ait Ortalama Ziyaret Sayısı	144
Hafta sonuna Ait Ortalama İstek Sayısı	1.293
En Aktif Gün	Pazartesi
En Pasif Gün	Cumartesi
En Aktif Tarih	Pazartesi, 20 Mayıs, 2013
En Aktif Güne Ait İstek Sayısı	10.625
En Aktif Güne Ait Ziyaret Sayısı	926
En Aktif Güne Ait Bant Genişliği	1,74 GB
En Pasif Gün	Cumartesi, 28 Eylül, 2013
En Pasif Güne Ait İstek Sayısı	220
En Pasif Güne Ait Ziyaret Sayısı	52
En Pasif Güne Ait Bant Genişliği	44,30 MB
Günün En Aktif Saati	14.00 - 14.59
Günün En Pasif Saati	00.00 - 00.59

Tablo 51: Erişim kayıtlarına ait günlük erişim istatistikleri.

Günlere Göre Erişim İstatistikleri				
Günler	**İstekler**	**Sayfa Sayısı**	**Ziyaretler**	**Bant Genişliği**
Pazartesi	35.815	6.772	4.129	6,70 GB
Salı	66.069	12.632	5.572	9,24 GB
Çarşamba	51.747	9.671	4.696	9,16 GB
Perşembe	47.757	8.795	4.693	9,33 GB
Cuma	45.360	8.537	4.753	9,59 GB
Cumartesi	43.533	8.310	4.744	7,99 GB
Pazar	31.454	6.144	3.396	7,21 GB
Toplam	**321.735**	**60.861**	**31.983**	**59,21 GB**

Erişim kayıtlarından elde edilen aktivite istatistikleri Tablo-50'de gösterilmiştir. Tablo-50 ve Tablo-51'deki bilgiler doğrultusunda web sitesi ziyaretçi trafiğinin en yoğun olduğu günün Pazartesi, en az yoğun olduğu günün ise Cuma olduğu görülmektedir. Bu bilgiler ışığında, web sitesi yöneticisinin site içi bakım çalışmalarını, tasarım ya da programlama çalışmalarını ziyaretçi trafiğinin en az yoğun olduğu gün olan Cuma günlerinde yapması önerilebilir.

Tablo 52: Erişim kayıtlarına ait aylık erişim istatistikleri.

Aylara Göre Erişim İstatistikleri				
Aylar	**İstekler**	**Sayfa Sayısı**	**Ziyaretler**	**Bant Genişliği**
Nis 2013	91.146	17.068	6.889	4,01 GB
May 2013	149.165	23.757	11.391	11,79 GB
Haz 2013	23.084	4.954	3.557	12,51 GB
Tem 2013	22.226	6.061	3.584	14,30 GB
Ağu 2013	16.356	4.369	3.157	8,06 GB
Eyl 2013	15.076	3.740	2.660	6,35 GB
Eki 2013	4.682	912	745	2,19 GB
Total	**321.735**	**60.861**	**31.983**	**59,21 GB**

Erişim kayıtlarına ait aylık ziyaretçi istatistiklerinden elde edilen verilere bakıldığında, Mayıs ayından sonra ziyaretçi sayısında ciddi bir düşüş görülmektedir. Bu düşüşün nedeni, web sitesinde bulunan içeriklere bağlı olabileceği gibi herhangi bir organizasyonun son bulmasından da kaynaklanmış olabilir.

4.3 En Çok Erişilen Kaynaklar

Web erişim kütükleri analiz edildiğinde, web sitesinden istekte bulunulan dosya tipleri, en çok erişilen sayfalar, en çok giriş ve çıkış yapılan sayfalar çıkarılabilmektedir. Tablo-53'de en çok erişilen sayfalar listelenmiştir. Tablodaki verilere göre en çok erişilen sayfa, web sitesinin ana sayfasıdır. Tablo-54'de web sitesine en çok giriş yapılan sayfalar gösterilmektedir. Tablo-53 ve Tablo-54'deki bilgiler ışığında, erişim kayıtları analiz edilen bu web sayfasının ana sayfası, reklam amaçlı ya da akademik anlamda önemli haber ve duyuruların konuşlandığı bir sayfa olarak kullanılabilir. Kullanıcıların çoğunun bu web sitesine doğrudan ana sayfadan erişmeleri, web sitesinde yayınlanan reklamların tıklanma sayısını arttırabilir ya da duyuru ve haberlerden kullanıcıların web sitesine ilk erişimlerinde haberdar edilmeleri sağlanabilir. Bu şekilde, web sayfasının amacına hizmet etmesi daha aktif olarak sağlanabilir ve reklam gelirlerinin arttırılması sağlanabilir. Tablo-55'de listelenen bilgiler de bu önerileri desteklemektedir. Kullanıcıların en çok çıkış yaptığı sayfanın ana sayfa olması, kullanıcıların duyuru ve haberleri aktif bir şekilde izlediğini göstermektedir.

Tablo 53: En çok erişilen sayfalar.

No	Sayfalar	Görüntüleme	Ziyaret
1	/	36.132	25.161
2	/tr/	14.217	9.567
3	/symposium_program_22	9.399	7.080
4	/registration_17	4.605	3.159
5	/project_competition_33	4.035	3.444
6	/accepted_papers_40	3.768	3.021
7	/paper_submission_21	3.615	2.427
8	/committees_20	3.519	3.036

Tablo 54: En çok giriş yapılan sayfalar.

En Çok Giriş Yapılan Sayfalar			
No	**Sayfalar**	**Girişler**	**Toplam Girişler (%)**
1	/	19.353	36,91%
2	/symposium_program_22	3.798	7,24%
3	/tr/	4.629	8,83%
4	/project_competition_33	1.797	3,43%
5	/committees_20	1.221	2,33%
6	/registration_17	1.035	1,97%
7	/accepted_papers_40	861	1,64%
8	/ulasim_bilgileri_41	801	1,53%

Tablo 55: En çok çıkış yapılan sayfalar.

En Çok Çıkış Yapılan Sayfalar			
No	**Sayfalar**	**Çıkışlar**	**Toplam Çıkışlar (%)**
1	/	12.729	24,27%
2	/symposium_program_22	3.843	7,33%
3	/project_competition_33	1.554	2,96%
4	/tr/	3.447	6,58%
5	/committees_20	1.287	2,45%
6	/accepted_papers_40	1.167	2,23%
7	/registration_17	1.101	2,10%
8	/keynote_speakers_24	921	1,76%
9	/ulasim_bilgileri_41	852	1,62%

Tablo 56: En çok erişilen dosya uzantıları.

En Çok Erişilen Dosya Uzantıları					
No	**Uzantılar**	**İstek**	**Toplam İstek (%)**	**Bant Genişliği**	**Toplam Bant Genişliği (%)**
1	.png	72.624	22,57%	5,05 GB	8,53%
2	.aspx	63.142	19,63%	3,08 GB	5,20%
3	.jpg	38.613	12,00%	4,07 GB	6,88%
4	.ico	26.590	8,26%	11,67 GB	19,71%
5	.css	25.888	8,05%	6,62 GB	11,17%
6	.js	24.470	7,61%	3,50 GB	5,91%
7	.pdf	13.004	4,04%	823,87 MB	1,36%
8	.txt	11.754	3,65%	9,20 GB	15,54%
9	.gif	5.755	1,79%	3,96 GB	6,69%

Yapılan analiz sonucunda en çok erişim yapılan dosya uzantıları ve istatistiksel bilgileri Tablo-56'da gösterilmektedir. Tablo-56'daki bilgiler ışığında, toplam erişim istekleri içerisinde en çok istekte bulunulan dosya uzantısı .png dosya uzantısıdır. Resim dosyası uzantısı olan .png tipindeki dosyaların, web sitesi trafiğinde çok yer kapladığı görülmektedir. Bu da, web sitesinin tamamen açılma süresini uzatmakta ve web sitesinde bulunan resim dosyalarının, kullanıcı ekranına daha geç geldiğini göstermektedir. Tablo-56'daki bilgilere bakıldığında ikinci olarak en çok erişilen dosya uzantısının .aspx dosya uzantısı olduğu görülmektedir. Bilgi amaçlı ya da dinamik olarak kullanılabilen bu dosyalar, web sitesinde bulunan sayfalardaki bilgilerin güncellenebilir olduğunu ya da yükleme-indirme işlemlerinin aktif olarak yapılabileceğini göstermektedir. Bu tarz işlemlerin yapıldığı web sitelerinin, kötü niyetli kullanıcılarca saldırıya uğrama ihtimalleri fazla olduğundan, erişim bilgileri analiz edilen bu akademik web sitesinin güvenliğinin sağlanması ayrıca bir önem taşımaktadır.

4.4 Kullanıcılara Ait İstatistikler

Erişim kayıtları analiz edilen akademik web sayfasını ziyaret eden kullanıcılara ait isteklerin ülkelere göre dağılımları Tablo-57'de, web tarayıcılarına göre dağılımları Tablo-58'de, kullanılan işletim sistemlerine göre dağılımları Tablo-59'da ve kullanılan mobil aygıtlara göre dağılımları da Tablo-60'da gösterilmiştir.

Tablo 57: Ülkelere göre kullanıcı istatistikleri.

Ülke İstatistikleri				
No	**Ülke**	**İstek**	**Ziyaret**	**Bant Genişliği**
1	Türkiye	252.562	15.432	34,72 GB
2	Amerika	40.870	10.155	16,83 GB
3	Bilinmiyor	2.546	1.650	1,48 GB
4	Çin	2.973	752	1,07 GB
5	İngiltere	3.294	576	953,14 MB
6	Japonya	968	563	214,24 MB
7	Almanya	3.129	453	412,81 MB
8	Fransa	1.212	407	412,00 MB
9	Hollanda	1.277	225	355,91 MB
10	Rusya	442	142	98,61 MB

Yapılan analizler sonucunda, akademik anlamda hizmet sunan web sitesine en çok erişimin Türkiye'den olduğu görülmektedir. Türkiye'yi takiben en çok erişim Amerika'dan yapılmıştır. Yaklaşık 7 aylık erişim kaydı analiz edilen bu web sayfasına yapılan erişimlerin ülke bazında çeşitlilik göstermesi, web sayfasının uluslararası yayın yaptığını da göstermektedir. Bu bilgiler ışığında, web sayfasında bulunan içeriklerin yabancı dillerde de sunulması gerekliliği ortaya çıkmaktadır.

Tablo 58: Web tarayıcılarına göre kullanıcı istatistikleri.

Tarayıcı İstatistikleri					
No	**Tarayıcı**	**İstekler**	**Ziyaretler**	**Sayfa Görüntüleme**	**Bant Genişliği**
1	Mozilla/5.0	109.794	5.775	17.417	5,51 GB
2	Chrome 26.x	22.483	1.275	2.891	1,77 GB
3	Mozilla 2.x	22.468	1.344	3.076	1,31 GB
4	Internet Explorer 9.x	15.320	1.411	1.322	8,17 GB
5	Internet Explorer 8.x	14.889	1.040	1.645	1,46 GB

Tablo-58'deki bilgilere bakıldığında Mozilla ve Chrome web tarayıcılarının kullanıcılar tarafından daha çok kullanıldığı görülmektedir. Bütün web tarayıcıları .css uzantılı stil dosyalarında bulunan stil kodlarını aynı şekilde yorumlamaktadır. Fakat stil kodlamasında form kontrolüne ait özellik değerleri belirtilmeyen bazı form kontrolleri, farklı tarayıcılar tarafından farklı yorumlanabilmektedir. Bu da aynı web sayfasının farklı tarayıcılarda farklı görüntülenmesine sebep olabilmektedir. Web sitesi tasarımcısının bu farklılığı göz önünde bulundurarak, stil kodlamasında değeri belirtilmeyen form kontrol

özelliklerine varsayılan bir değer ataması gerekmektedir. Bu sayede, web sayfasının tüm tarayıcılarda aynı şekilde görüntülenmesi sağlanabilir.

Tablo 59: İşletim sistemlerine göre kullanıcı istatistikleri.

İşletim Sistemleri				
No	**İşletim Sistemi**	**İstekler**	**Ziyaretler**	**Bant Genişliği**
1	Windows 7	180.072	11.303	23,87 GB
2	Windows XP	38.505	3.820	6,49 GB
3	Windows NT	28.147	1.876	4,84 GB
4	Google Android	7.540	553	892,42 MB
5	Mac OS	6.570	859	1,18 GB
6	Linux	5.706	662	726,82 MB
7	Windows Vista	3.884	483	724,89 MB
8	iOS 4	3.392	418	3,11 GB
9	Linux Ubuntu	1.197	94	90,68 MB
10	Windows 2003	768	108	142,80 MB

Tablo-59'da kullanıcıların, web sitesine erişirken kullandıkları işletim sistemleri listelenmektedir. Tablodan elde edilen bilgilere bakıldığında, Microsoft firmasına ait olan Windows işletim sisteminin daha çok kullanıldığı görülmektedir. Bu bilgi birçok bakımdan yorumlanabilir. Örneğin, Windows XP işletim sisteminin, Windows Vista işletim sisteminden daha çok kullanıldığı, Windows 7 işletim sisteminin ise en çok kullanılan işletim sistemi olduğu görülmektedir. Windows XP işletim sisteminin, Windows Vista işletim sisteminden daha eski bir işletim sistemi olmasına rağmen daha çok kullanılması, XP işletim sisteminin kullanıcılara daha rahat bir kullanım sunduğunu ve Windows Vista işletim sisteminin kullanıcılarca çok fazla tercih edilmediğini göstermektedir. Buna rağmen Windows 7 işletim sisteminin XP ve Vista işletim sistemlerinden daha çok kullanılmış olması, Windows 7 işletim sisteminin daha kullanışlı olduğunu göstermektedir.

Tablo 60: Mobil aygıtlara göre kullanıcı istatistikleri.

Mobil Aygıtlar				
No	**Mobil Aygıt**	**İstekler**	**Ziyaretler**	**Bant Genişliği**
1	iPhone	7.445	929	3,51 GB
2	iPad	2.894	351	399,57 MB
3	HTC Droid Incredible	71	69	55,99 MB
4	Samsung SGH-E250	128	62	97,70 MB
5	iPod Touch	49	42	11,38 MB
6	BlackBerry 9800	49	11	43,28 MB
7	Samsung Galaxy S	31	3	1,29 MB
8	Samsung Galaxy Nexus Prime	72	3	2,88 MB
9	Nokia N97	19	3	561,78 KB
10	BlackBerry 9700	2	2	20,75 MB

Tablo-60'da kullanıcıların, web sitesine erişirken kullandıkları mobil aygıtlar gösterilmektedir. Tablodaki bilgilere bakıldığında web sitesine, birçok mobil aygıttan giriş yapıldığı ve bu aygıtlara ait toplam erişim bilgilerinin azımsanmayacak bir değerde olduğu gözlemlenmiştir. Bu nedenle akademik anlamda hizmet sunan bu web sitesinin, mobil aygıtlar için ayrıca bir tasarımının olması gerekliliği görülmektedir.

4.5 HTTP Protokolü Durum Kodlarına Göre İstatistikler

Kullanıcıların web sitelerine yaptıkları istekler, sunucu tarafından durum kodları ile cevaplanmaktadır. Sunucu, kullanıcının yapmış olduğu her isteğin sonucunu, sistem yöneticisine HTTP durum kodlarıyla haberdar eder. Bölüm 2.3.2.1.4'te HTTP durum kodları ile ilgili gerekli bilgiler verilmiştir. Sistem yöneticisi yapılan isteklerin başarılı ve başarısız olduğunu bu durum kodları ile anlamaktadır. Erişim kütüklerinden çıkarılan bu durum kodlarıyla, yapılan isteklerin durumları analiz edilir. Bu çalışmada, erişim kayıtları incelenen web sitesine ait HTTP durum kodları, Tablo 60'da verilmiştir. Elde edilen bilgilere göre hatalı isteklerin toplam sayısı 111.336'dır. Bu değer, web sitesine yapılan toplam istek sayısının %11,5'idir. Bu bilgilere göre bakıldığında hata oranlarında en çok pay 404 hata türüne aittir. Bu hata türüne ait olası nedenler şu şekildedir;

- İstenen dosya yeniden adlandırılmış olabilir.
- İstenen dosya başka bir konuma taşınmış ve/veya silinmiş olabilir.
- İstenen dosya bakım veya diğer bilinmeyen nedenlerden dolayı geçici olarak kullanılamıyor olabilir.
- İstenen dosya sunucuda olmayabilir.

Web sitesi yöneticisinin 404 hata türüne ait olası nedenleri göz önünde bulundurarak, bu problemi çözmesi site etkinliğinin ve başarımının daha da artmasına sebep olabilir.

Tablo 61: HTTP protokolü durum kodlarına göre istatistikler.

Hata İstatistikleri			
No	**Hata Türü**	**İstek**	**İstek(%)**
1	404 – Bulunamadı	36.336	97,91%
2	500 – Dâhili sunucu hatası	688	1,85%
3	403 – Yasak	33	0,09%
4	400 – Geçersiz istek	30	0,08%
5	406 – Kabul edilemez	18	0,05%
6	501 – Uygulanmadı	4	0,01%
7	405 – Yönteme izin yok	2	0,01%
8	401 – Kimlik doğrulama hatası	1	0,00%
	Toplam	**37.113**	**100,00%**

5. SONUÇ VE ÖNERİLER

İnternet kullanımının yaygınlaşması ile web sunucularında saklanan günlük dosyalarının analiz edilmesi de önem kazanmıştır. İnternet kullanımının artması, başta güvenlik olayları olmak üzere performans, adli süreçler ve benzeri birçok olayın anlamlandırılması ve aydınlatılması adına günlük dosyalarının analizini gerekli kılmaktadır. Günlük dosyalarının analizinin yapılması, güvenlik ihlallerinin tespitinin yapılması ve delillerin toplanması, performansın izlenmesi, başarılı ve başarısız erişimlerin tespitinin yapılması gibi birçok konuda bizleri bilgilendirmektedir. Bu bilgiler, günlük dosyalarının sistem yöneticileri açısından ne kadar önemli bir veri kaynağı olduğunu göstermektedir.

Günlük dosyalarının sistem yöneticiler tarafından güvenle saklanıp, düzenli olarak analiz edilmesi, sistem yöneticilerini hem kanuni zorunluluk ve standartlar açısından hem de sistem performansı açısından ilgilendirmektedir. Bu anlamda 5651 sayılı kanun gereğince günlük dosyalarının tutulmasının yanı sıra, bu dosyaların zaman damgalarıyla beraber saklanmasına dikkat ederek profesyonel olarak günlük kaydı analizleri yapılmalıdır. Bilgi güvenliği kapsamında da günlük kayıtlarının toplanması, elde edilmesi, analizi ve raporlanması büyük önem taşımaktadır. Ayrıca kütük dosyaların analiz edilmesi ile web tabanlı saldırıların da tespit edilmesi, güvenlik açısından da kurum kuruluşları yakından ilgilendirmektedir. Bu nedenle bilişim teknolojileri alanında hizmet sunan kurum ve kuruluşların bu alana özel çalışmalar yapması gerekmektedir.

Bu tez çalışmasında bir web sitesine ait erişim kayıtları, geliştirilen yazılım ile web kullanım madenciliği disiplini kapsamında analiz edilmiş ve elde edilen istatistiksel bilgiler tablolar halinde gösterilmiştir. Erişim kayıtlarının analizi sonucunda birçok bilgi elde edilmiştir. Bu bilgiler;

- Site trafiğine ait genel istatistiksel bilgiler.
- Site kullanıcılarına ait aktivite istatistikleri.
- Web sitesine erişen kullanıcıların günlere göre dağılımları.
- Web sitesine erişen kullanıcıların aylara göre dağılımları.
- Kullanıcıların en çok eriştikleri sayfalar.
- Kullanıcıların en çok giriş ve çıkış yaptığı sayfalar.
- Kullanıcıların en çok eriştikleri dosya uzantıları.
- Kullanıcı isteklerinin, ülkelere göre dağılımları.

- Kullanıcı isteklerinin, web tarayıcılarına, işletim sistemlerine ve erişim yapan mobil aygıtlara göre dağılımları.
- HTTP durum koduna göre istatistikler.
- XSS ve SQL enjeksiyon saldırı girişimleridir.

Bu tez çalışmasında, web sitesinden elde edilen erişim kayıtları C# programlama dilinde gerçekleştirilen LCS yazılımı ile web kullanım madenciliği aşamalarından geçirilmiş elde edilen bilgiler MSSQL veri tabanına tablolar halinde aktarılmıştır. Veri tabanında bulunan düzenli bilgilerden önemli istatistiksel bilgiler çıkarılmıştır. Bu bilgiler web sitesi yöneticilerine, web sitesinin geliştirilmesi veya yeniden tasarlanması için önemli bilgiler sunmaktadır. Yapılan analizler sonucunda;

- Web sitesinin ana sayfası diğer sayfalara göre daha çok ziyaret edilmiş, iç kısımlarda kalan diğer sayfalara ait ziyaret sayıları, birbirine yakın olduğu görülmüştür.
- Kullanıcı isteklerinin oluşturduğu HTTP durum kodları incelendiğinde 404 "kaynak, sayfa bulunamadı" hatasının çok olduğu görülmektedir. Bu sonuç doğrultusunda, web sitesine ait sayfalarda kırık linklerin olduğu anlaşılmış ve bu problemin giderilmesi gerektiği anlaşılmıştır. 5XX kodlu isteklerinin az olduğu görülmektedir. Bu sonuç da web sunucusunun genel olarak sorunsuz çalıştığını göstermektedir.
- Kullanıcıların web sitesini en az ve en çok kullandığı günler çıkarılmıştır. Bu bilgi ışığında web sitesinde yapılacak herhangi bir bakım çalışmasının bu günler göz önünde bulundurularak yapılması önerilebilir.
- Kullanıcıların en çok erişim yaptığı dosya uzantılarının ağırlıklı olarak resim dosyaları olduğu görülmüştür. Web sitesinde bulunan resim dosyalarının boyutları, web sitesinin hızlı yüklenmesini geciktiriyorsa bu resimlerin web sitesi yöneticisi tarafından daha dikkatli seçilmesi gerekmektedir. Boyutu büyük slayt resimleri, ikon vb. resim dosyaları, web sitesinin yavaş yüklenmesine yol açmaktadır. Bu durum hem web sitesi performansını negatif olarak etkilemekte hem de limitli internet kullanıcılarının aleyhine olmaktadır.
- Mobil aygıtlar ile web sitesine erişimin yadsınamaz derecede olduğu görülmüştür. Bu sebeple mobil kullanıma uygun bir sayfanın tasarlanmasının gerekli olduğu öngörülmektedir.

KAYNAKLAR

[1] Guidelines for Auditing and Logging, 2008. Computer Emergency Response Team, Security Guidelne, India.

[2] Grace, J., V, Maheswari., D, Nagamalai., 2011. Analysis of Web Logs and Web User in Web Mining, International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.1.

[3] Kent, K. and Souppaya, M., 2006. National Institute of Standarts and Technology. Special Pub. 800-92. *Guide to Computer Security Log Management* , Gaithersburg, USA.

[4] Ş, Ender., K, Arzu., K, Rembiye. ve Ş, Önder, 2009. Kurumlarda Log Yönetiminin Gerekliliği, XI. Akademik Bilişim Konferansı Bildirisi, Harran Üniversitesi, Şanlıurfa, 11-13 Şubat, s. 613-615.

[5] Navin Kumar Tyagi, A. K. Solanki and Manoj Wadhwa, 2010. Analysis of Server Log by Web Usage Mining for Website Improvement, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 8.

[6] J. Vellingiri and C. Pandian, 2011. A Novel Technique for Web Log mining with Better Data Cleaning and Transaction Identification, Journal of Computer Science 7 (5): 683-689, ISSN 1549-3636.

[7] T. Pamutha, S. Chimphlee, C. Kimpan and P. Sanguansat, 2012. Data Preprocessing on Web Server Log Files for Mining Users Access Patterns, International Journal of Research and Reviews in Wireless Communications, Vol. 2, No. 2, ISSN: 2046-6447.

[8] Daş, R. and Türkoğlu, İ., Web Tabanlı Öğretim Materyallerinin Web Kullanım Madenciliği ile Analiz Edilmesi, Fırat Üniv. Mühendislik Bilimleri Dergisi, 22 (1), 111-122, 2010.

[9] Daş, R., Türkoğlu, İ. and Poyraz M. Analyzing Of System Errors For Increasing A Web Server Performance By Using Web Usage Mining, Journal Of Electrical & Electronics Engineering, Istanbul University, Vol. 7, No.8, 379–386, 2007.

[10] Daş, R., Türkoğlu, İ. and Poyraz, M., Bir Web Sitesine Ait Kullanıcı Erişim Kayıtlarının Web Kullanım Madenciliği Yöntemiyle Analizi: Fırat Üniversitesi Örneği, e-Journal of New World Sciences Academy 2008, Volume: 3, Number: 2, ISSN:1306-3111.

[11] Daş, R. and Türkoğlu, İ., Creating Meaningful Data From Web Logs For İmproving The İmpressiveness Of A Website By Using Path Analysis Method, Expert Systems with Applications, Elsevier, 6635–6644, 2009.

[12] 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.

[13] 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.

[14] Akshay Shenoy, 2011. Improving The Performance Of A Proxy Server Using Web Log Mining. *Master Thesis*, San Jose State University, The Department of Computer Science, ABD.
[15] S, E, Salama., M, I, Marie. and L, M, El-Fangary & Y, K, Helmy., 2011. Web Server Logs Preprocessing for Web Intrusion Detection, Canadian Center of Science and Education, Vol. 4, No. 4.
[16] C, P, Sumathi., R, P, Valli. and T, Santhanam., 2010. Automatic Recommendation of Web Pages in Web Usage Mining, International Journal on Computer Science and Engineering, Vol. 02, No. 09, 2010, 3046-3052.
[17] C, Romero., S, Ventura., A, Zafra. and P, de Bra., 2009. Applying Web Usage Mining For Personalizing Hyperlinks In Web-Based Adaptive Educational Systems, Computers & Education, Elsevier, 828–840.
[18] Daş, R., 2008. Web Kullanıcı Erişim Kütüklerinden Bilgi Çıkarımı, Doktora Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ.
[19] Internet: İnternet ortamında yapılan yayınların düzenlenmesi ve bu yayınlar yoluyla işlenen suçlarla mücadele edilmesi hakkında kanun, http://www.tbmm.gov.tr/kanunlar/k5651.html Erişim Tarihi: 09.06.2013.
[20] R, Meyer., 2008. Detecting Attacks On Web Applications From Log Files, Information Security Reading Room, SANS Institute.
[21] Internet: En Kritik 10 Web Uygulaması Güvenlik Zayıflıkları 2007 Güncellemesi, http://csirt.ulakbim.gov.tr/dokumanlar/Ceviri_OWASP.pdf, OWASP, Erişim Tarihi: 10.06.2013.
[22] Vural, Y., "Kurumsal Bilgi Güvenliği ve Sızma Testleri" Yüksek Lisans Tezi, Bilgisayar Mühendisliği, Gazi Üniversitesi, ANKARA, 2007.
[23] Internet:http://www.gokselcuryan.com/haberler/3-teknoloji-yenilikleri/100-SQL-injection-nedir.html, Erişim Tarihi: 20.07.2013.
[24] Anley, C., "Advanced SQL Injection In SQL Server Applications", Next Generation Security Software Publication, Surrey, 2002.
[25] Internet: http://technet.microsoft.com/en-us/library/cc722404(v=WS.10).aspx. Erişim Tarihi: 05.06.2013.
[26] Internet: https://www.owasp.org/images/e/e0/OWASP_Logging_Guide.pdf. Erişim Tarihi: 05.06.2013.
[27] Wahab M.H.A., Mohd M.N.H., Hanafi H.F. and Mohsin M.F.M., 2008. Data Pre-processing on Web Server Logs for Generalized Association Rules Mining Algorithm, Proceedings Of World Academy Of Science, Engineering And Technology Vol.36 ISSN 2070-3740.
[28] Internet: http://httpd.apache.org/docs/1.3/logs.html. Erişim Tarihi: 07.06.2013.
[29] Internet:http://wiki.internet.gen.tr/index.php/E-posta_sunucuları_nasıl_çalışır. Erişim Tarihi: 08.06.2013.
[30] Internet: http://msdn.microsoft.com/en-us/library/ms164086.aspx. Erişim Tarihi: 08.06.2013.
[31] Chuvakin, A., Schmidt, K. and Philips C., 2013. Logging and Log Management, Sygnress, U.S.A.

[32] K, Enis., 2002. Ağ Cihazlarının Güvenliğinin Sağlanma Yöntemleri, Ege Üniversitesi, İzmir.
[33] Karen, S. and Hoffman, P., 2009. National Institute of Standarts and Technology. Special Pub. 800-41. Guidelines on Firewalls and Firewall Policy, Gaithersburg, USA.
[34] Fortinet, Logging and Reporting Handbook. 01-432-112804-20120124, 2012.
[35] Gezer, M., Erol, Ç. and Gülseçen, S., 2007. Bir Web Sayfasının Web Madenciliği İle Analizi, Akademik Bilişim 2007, Dumlupınar Üniversitesi, Kütahya.
[36] H, İsmail., 2007. Veri Madenciliği Algoritmaları Kullanılarak Web Günlük Erişimlerinin Analizi, Başkent Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
[37] Kaya, H., ve Köymen, K., 2008. Veri Madenciliği Kavramı Ve Uygulama Alanları, Doğu Anadolu Bölgesi Araştırmaları.
[38] Internet: http://www.aspnedir.com/Article/DisplayArticle.aspx?ID=610, Erişim Tarihi: 07.08.2013.

ÖZGEÇMİŞ

1986	:	Elazığ'da doğdu.
1998 - 2002	:	Afyon Milli Piyango Anadolu Lisesi'nde ortaöğretimini tamamladı
2002 - 2004	:	Bayburt Anadolu Lisesi'nden mezun oldu.
2005 - 2010	:	Fırat Üniversitesi Teknik Eğitim Fakültesi Bilgisayar Öğretmenliği Bölümü'nden mezun oldu.
2010 - 2011	:	Elazığ/Karakoçan'da ücretli öğretmenlik yaptı.
2012 - …..	:	Fırat Üniversitesi Teknik Eğitim Fakültesi Elektronik ve Bilgisayar Eğitimi Bölümünde Yüksek Lisans öğrenimine başladı ve halen aynı bölümde öğrenimine devam etmektedir.

Doygun DEMİROL