Gazi Üniv. Müh. Mim. Fak. Der.
Cilt 22, No 2, 247-253, 2007

J. Fac. Eng. Arch. Gazi Univ.
Vol 22, No 2, 247-253, 2007

WEB ATAKLARI İÇİN METİN TABANLI ANORMALLİK TESPİTİ (WAMTAT)

Hidayet TAKCI[1], Turker AKYUZ[2], İbrahim SOGUKPINAR[3]
[1, 2, 3] Computer Engineering Dept., Gebze Institute of Technology, 41400, Gebze, Kocaeli
[1]htakci@bilmuh.gyte.edu.tr, [2]takyuz@bilmuh.gyte.edu.tr, [3]ispinar@bilmuh.gyte.edu.tr

(Geliş/Received: 15.12.2005; Kabul/Accepted: 12.12.2006)

ÖZET

Bugünlerde birçok web sitesi kullanıcılarla etkileşim içerisinde olup bu etkileşimde kullanıcılar isteklerini URL içinde gömülü olarak web sunucuya iletirler. URL içerisine giriş verisi olarak zararlı kodun gömülmesi atak yöntemlerinden biridir ve bu tip atakların tespiti için giriş verisi analiz edilebilir. Bu çalışmada, atak tespiti için metin tabanlı bir anormallik tespiti yöntemi önerilmektedir. Önerilen yöntem kullanıcı girişlerinin analizinde giriş verisinin metinsel özelliklerini kullanır. Gerçeklemesi yapılarak deneysel sonuçları bu makalede verilen yöntem web tabanlı atakların anormallik tabanlı tespitinde yeni bir yaklaşımdır.

Anahtar Kelimeler: Anormallik tespiti, metin özellikleri, web atak tespiti.

A TEXT BASED ANOMALY DETECTION FOR WEB ATTACKS

ABSTRACT

Nowadays, there is an interaction between the web sites and users. In this interaction, user requests are sent to web servers in URL strings. Sometimes, harmful code may be embedded into those strings. Harmful code embedding is one of web attacks. User input data may be analyzed for detection of this type of attack. In this study, a text based anomaly detection method has been proposed. Proposed method uses textual properties of input data for analysis. This method that is implemented and given experimental results is particularly a new approach for web based anomaly detection.

Keywords: Anomaly detection, text properties, web attack detection.

1. GİRİŞ (INTRODUCTION)

Web, son yıllarda bilgi paylaşımı için en çok tercih edilen ortamlardan birisi olmuştur. Web sitelerinin sayısı çok hızlı şekilde artmış ve web tabanlı servisler yaşamımızın önemli bir parçası haline gelmiştir. Bununla birlikte, web sitelerine karşı yapılan saldırılar sebebiyle web ortamını kullanmak riskli durumdadır. Bu riskleri ortaya koyan, web ataklarıyla ilgili bir çalışma Foltz tarafından yapılmıştır. İlgili çalışmada 1999 – 2004 yılları arasındaki web atakları raporlanmıştır [1].

Web ile ilgili saldırılar 2001 nisan ayı ile 2002 mart ayı arasında bütün saldırıların yüzde 23'lük kısmını oluşturmuştur [2]. Bu oran her geçen gün daha da artmakta olup bu duruma paralel olarak web güvenliğinin önemi artmış, web güvenliği için bir dizi teknik geliştirilmiştir. Web güvenliği ile ilgili detaylı bir çalışma Rubin ve Geer tarafından yapılmıştır [3].

Web saldırılarını tespit için kullanılabilecek araçlardan birisi saldırı tespit sistemleridir. Saldırı tespit sistemleri, bilişim sistemlerindeki olayları analiz ederek zararlı davranışı tespit ederler. Bu sistemlerin kötüye kullanım ve anormallik tespiti şeklinde iki yaklaşımı bulunur. Kötüye kullanım tespitinde; iyi bilinen atak örüntüleri kullanılarak atak imzaları ile eşleşen olaylar saldırı olarak belirlenir.

Günümüzde birçok kötüye kullanım tespiti tabanlı saldırı tespit ürünü kullanılabilir durumdadır. Bu ürünlerin bir kural veritabanı vardır ve bu veritabanları atak imzaları içerir. İmza veritabanları yeni bir saldırı türü ortaya çıktığında kolayca güncellenebilirler. Bu ürünler bilişim sistemlerindeki nüfuzların yüzde 68'ini tespit etme yeteneğine

sahiptir [4]. Bununla birlikte, kötüye kullanım tespiti tabanlı saldırı tespit sistemleri web saldırıları konusunda yeterli kapasiteye sahip değildir [5, 6]. Kötüye kullanım tespiti yönteminde; sadece bilinen saldırılar tespit edilebilmekte, atak imzaları genelleştirilememekte ve fazla sayıda hatalı alarm meydana gelmektedir.

Anormallik tespiti yöntemi, saldırı tespit yaklaşımlarından bir diğeridir. Bu yaklaşımın kötüye kullanım tespitine göre bazı avantajları bulunmakla birlikte bilinmeyen saldırıların tespiti bu yaklaşımın en önemli avantajıdır. Bu yüzden son zamanlarda birçok anormallik tespiti yöntemi geliştirilmiştir [7, 5]. Anormallik tespiti sistemleri beklenen normal kullanım profillerinden sapma gösteren etkinlikleri anormallik olarak işaretlerler. Anormallik tespitinde, sistemin normal davranışı genellikle istatistiksel yöntemler yardımıyla elde edilir.

Bu çalışmada web atak tespiti için anormallik tabanlı bir yöntem önerilmiştir. Önerilen yöntemde kullanıcı giriş verisi, içerisinde ataklarla ilişkili karakterler veya kelimeler olup olmamasına göre analiz edilir. Yöntem, her bir giriş verisi için bir anormallik skoru hesaplamaktadır. Bu hesaplamada kullanıcı verisine bazı metinsel işlemler uygulanmaktadır. Eğer kullanıcı giriş verisine ait anormallik skoru eşik seviyesinden daha büyükse kullanıcının saldırgan olduğu aksi takdirde normal bir kullanıcı olduğu tahmin edilir. Bu yöntem metinlerin anlamı ile onların içeriği arasında bir ilişki bulunduğu varsayımı üzerine bina edilmiştir.

Önerilen yöntem, özellikle giriş verisi içerisinde özel karakterler veya kelimeler bulunduğu durumlarda daha başarılıdır. Örneğin, bazı ataklar giriş verisi içerisine zararlı kod gömülmesi ile yapılır. Zararlı kod içerisinde (**.. , ; , < , > , |** vs) gibi özel karakterler ve (DELETE, UPDATE, vs.) gibi özel kelimeler bulunabilmektedir. Özel karakterlerden (.) karakteri directory traversal atağında, (;) karakteri komut çalıştırma atağında sıklıkla rastlanan karakterlerdir. Dolayısıyla kullanıcı tarafından web sitesine URL içinde sunulan giriş verisi içerisindeki karakterler ve kelimeler taranarak erişimin bir atakla ilgili olup olmadığı anlaşılabilir. Önerilen yöntem Gebze Yüksek Teknoloji Enstitüsü (GYTE) web sunucusu üzerinde toplanan erişim verileriyle test edilmiştir.

Makalenin organizasyonu şu şekildedir: giriş bölümünün ardından, ikinci bölümde geri plan bilgisi ve web atak tespiti konusundaki çalışmalar sunulmuştur. Üçüncü bölümde önerilen yöntem açıklanmıştır. Dördüncü bölümde ise deneysel sonuçlar ile analizler sunulmuştur. Son bölümde sonuçlar yer almaktadır.

2. WEB ATAK TESPİTİ ve İLGİLİ ÇALIŞMALAR (WEB ATTACK DETECTION and RELATED WORKS)

2.1. WEB ATAKLARI (WEB ATTACKS)

Web uygulamaları; web sunucu üzerinde yer alan, veritabanları veya diğer dinamik içerik ile etkileşim halinde olan betiklerden (scriptt) meydana gelir. Bu uygulamalar, istemci sunumcu etkileşim içindedirler ve web uygulama açıklıklarının önemli bir kısmı zararlı kod girişlerinden meydana gelir. İstemci istekleri sunucu üzerinde işlenir ve bu yüzden sunucular tehdide açıktırlar. Web sunucu yazılımları ve web uygulama programları bazı açıklar içerebildiği için web sunucular saldırganlar için popüler birer hedeftir.

Web atakları aslında http protokolünü kullanan ağ ataklarıdır. Web ataklarını daha iyi anlayabilmek için atak yaşam döngüsüne bakmak faydalı olacaktır. Atak yaşam döngüsü bir atağın başlangıcından başarılı olmasına kadar geçen süreci işaret eder. Bir saldırganın giriş noktası web sunucu yazılımı veya bir web uygulamasıdır. Saldırgan bazı güvenlik servislerini etkisiz hale getirmek için veya servisleri bozmak için bir açıklık arar. Daha sonra açığı kullanarak o açıktan sisteme nüfuz eder [8].

Eğer web uygulamasına yapılan saldırı başarılı olduysa, web sunucu veya web uygulaması üzerinde bir açık vardır. Bütün web sunucu yazılımları açık içerebilir ve üreticiler genellikle açıklıkları ortadan kaldırmak için yama çıkarırlar. Diğer taraftan web uygulama kodu hatalar içerebilir, örneğin, sunucu taraflı betikler (asp, php, jsp, etc.), SQL cümlelerinde veya farklı teknolojilerde kullanılan bazı nesneler hataya sebep olabilir.

2.2. WEB ATAK TESPİTİ ÇALIŞMALARI (WEB ATTACK DETECTION WORKS)

Kruegel ve arkadaşları tarafından yönetilen bir çalışmada anormallik skoru; istek tipi, istek uzunluğu ve yük dağılımı gibi parametrelerden elde edilmektedir [9]. Onların yaklaşımına göre bir isteğin boyutu ortalama istek boyutundan daha büyükse o zaman isteğin atak olma ihtimali yüksektir. Kruegel ve ekibi http ve DNS trafiğini analiz ederek DNS ataklarını tespit için bir prototip geliştirmişlerdir.

Web atak tespiti için bir diğer yöntem karakter dağılımı yöntemidir. Bu yöntem Kruegel ve arkadaşları ile Vigna ve arkadaşlarının çalışmasında anlatılmıştır [5]. O yöntemin varsayımlarından birisi; normal web erişimlerindeki sorgu verisinin normal metinlerdekine benzer karakter dağılımı vereceği, bir atağa ait sorgu verisinin de farklı karakter dağılımı vereceğidir. İlgili çalışmada [5], sorgu sahasındaki özellik isimleri sorgu verisi olarak kullanılmıştır. Anormallik yaklaşımını desteklemek için webstat

isimli bir diğer çalışma Vigna ve arkadaşları tarafından önerilmiştir [10]. Bu yöntemde ataklar, durumların ve geçişlerin bir kompozisyonu olarak sunulmuştur.

Cho ve Cha [6] tarafından yönetilen oturum anormallik tespiti (Session Anomaly Detection - SAD) anormallik tespiti konusunda bir diğer çalışmadır. Onlar kullanıcılar tarafından istenen web sayfası dizilerinin benzer örüntülere sahip olduğunu varsaymışlardır. SAD web erişim günlüklerinden web oturumlarını açığa çıkarmakta, belirli istek sıralarına göre profiller oluşturmakta ve hesaplama yapmaktadır [6]. SAD bayes parametre tahmini (bayesian parameter estimation) tekniğini kullanmaktadır. Bu teknik öncel olasılık dağılımlarına göre olaylar arasındaki benzerliği tahmin etmektedir.

Önerdiğimiz çalışma web atak tespitini metin tabanlı bir yöntem ile yerine getirmektedir. Kullanıcı giriş verisinin tamamı içerik verisi olarak kullanılmakta ve analiz edilmektedir (özellik adları + özellik değerleri, örneğin; http://www.deney.com/default.asp?isim="Bilgisayar" gibi bir URL için giriş verisi olarak veri="**isim** + **Bilgisayar**" bilgisi kullanılmaktadır) Kruegel ve Vigna'nın [5] çalışmasında sadece özellik isimleri kullanılırken (yani giriş verisi olarak onlar değişken adı olan **isim** bilgisini kullanırlar) bizim çalışmamızda özellik değerleri de kullanılmıştır. Bunun sebebi daha uzun metinlerde yöntemimizin daha başarılı sonuçlar vermesidir. Yöntemimiz bir anormallik skoru elde etmekte olup anormallik skorunun parametrelerini giriş verisinin metinsel özellikleri oluşturmaktadır. Elde ettiğimiz skorlara göre 10 ile 30 puan arası puana sahip erişimler **normal** bu aralık haricinde kalan skorlar ise **anormal** kullanımları yani atak durumlarını vermektedir. Sistemin normal erişimler için alt eşiği 10, üst eşiği ise 30'dur.

3. METİN TABANLI ANORMALLİK TESPİTİ (TEXT BASED ANOMALY DETECTION)

Özel karakterler, özel kelimeler, karakter frekansları ve buna benzer metinsel bilgiler metnin içeriği hakkında bilgi verebilirler. Özel karakterler ve özel kelimeler metnin analiz nedenine göre değişir. Örneğin, metin güvenlik maksadıyla analiz ediliyorsa güvenlik için kullanılabilecek karakterler özel karakterlerdir. Web ataklarında sıklıkla kesme işareti (') kullanıldığı için kesme işareti güvenlik çalışmasında bir özel karakterdir. Aynı şekilde bir SQL injection atağında kullanılan DELETE kelimesi bir özel kelimedir. Ataklarda yer alan bu tipten özel karakterler ve özel kelimelerin bir kod içinde yer alması o kodun zararlı kod olma ihtimalini yükseltmektedir. Bunlara ek olarak bir metin ortalamadan daha uzun ise bu da o metnin içeriği hakkındaki kanaati değiştirecektir. Metnin normalden uzun olması web uygulamasını çökertebileceği için metnin uzunluğu da metnin zararlı olma ihtimalini artırır.

Takçı ve Soğukpınar tarafından yapılan bir çalışmada metin içerisinde yer alan karakterlerin sıklık bilgileri ile metin dili bulunabilmiştir [11]. Bu da göstermektedir ki metinle ilgili parametrelerin doğru seçimi sayesinde metnin içeriği hakkında bilgi elde edilebilmektedir. Dil tanıma çalışmasında sadece karakter sıklıkları yeterli olmuştur fakat bir metnin zararlı kod içerip içermediğini anlamak için karakter sıklıklarından daha fazla bilgiye ihtiyaç vardır.

3.1. YÖNTEMİN PARAMETRELERİ (PARAMETERS of THE METHOD)

Bir web erişiminin atak olup olmadığını tespitten önce bir web erişiminde hangi parametrelerin kullanılabileceği konusuna bakmak gerekir. Yapılan deneyler göstermiştir ki web erişimlerinde kullanılan istek tipi ile web erişimlerinin tipi (atak veya normal) arasında bir ilişki vardır. O yüzden istek tipi parametrelerden biri olmalıdır. Normal web erişimlerinde web sunucudan genellikle belirli web doküman tipleri istenmektedir (.htm, .asp vs.) eğer bunlar haricinde bir dosya tipinde istek varsa bu durum anormaldir ve kimi zaman saldırı amaçlıdır. Web sunucu üzerindeki hassas veriyi elde etmeye yöneliktir. O yüzden web olmayan doküman tiplerinin istenmesi de bir parametre olarak hesaba katılabilir. Web ataklarında komut çalıştırma atakları önemli yer tutmakta olup bir web erişiminde komut çalıştırma girişimi olup olmadığını tespit için bazı anahtar kelimeleri taramak bir çözüm olabilir. Örneğin, "cmd.exe" gibi sözcükler web erişim verisinde yer alıyorsa bu erişim bir web atağı erişimidir. Bu sebeple komut çalıştırma atağıyla ilgili kelimeler de parametrelerden biridir. Normal giriş verilerinde pek fazla rastlanmamakla birlikte genellikle web ataklarında yer alan bazı karakterler de web atağını tespitte önemli yer tutmaktadır. Parametrelerden birisi de özel karakterlerdir. En son parametre ise özellikle tampon taşması (buffer overflow) atağını yakalayabilmek için giriş verisinin uzunluğuna bakılmasıdır. Giriş verisinin uzunluğu da böylece parametrelerden biridir.

Yöntemin parametreleri şunlardır;

1. *istek tipi (GET, POST and HEAD)*
2. *web olmayan dokümanlar*
3. *komut çalıştırma durumu*
4. *özel karakterlerin giriş verisi içinde bulunması*
5. *giriş verisi uzunluğu*

İstek tipi(Request Method)
Web ataklarıyla istek tipleri arasında bir ilişki vardır. Her bir istek tip'i için anormallik skoru diğerinden farklıdır. Tablo 1 de verilen istek yöntemleri için anormallik skorları deneysel olarak elde edilmiştir.

Tablo 1. İstek yöntemleri için anormallik skorları (anomaly scores for request methods)

Yöntem	Durum Kodu=200	Anormallik Skoru
GET	72	18
POST	88	12
HEAD	0	100
PROPFIND	0	100
OPTIONS	100	0

Web olmayan dokümanlar (Non web files)
Kimi zaman web sunucudan web formatında olmayan dokümanlarda istenir. Bu durum genellikle atak durumunda meydana gelir. Bu yüzden web olmayan dokümanların kullanıcılar tarafından istenmesi anormallik skoru hesabında önemli bir parametredir. Web olmayan doküman formatları şu gruplardan birinde olabilir;

Arşiv dosyaları (.zip, .tar.gz, vs.)
Yedekleme dosyaları (.bak, vs.)
Başlık dosyaları (.inc, .asa, vs.)
Metin dosyaları (.txt, ör. Readme.txt)

Dolayısıyla eğer giriş verisini tutan URL içerisinde bu dosyalara ait uzantılardan biri veya birkaçı bulunuyorsa o zaman ilgili isteğin web atağı olma ihtimali artmaktadır. Bu gruptaki her bir dosya tipinin puanı eşit kabul edilmiştir. URL isteğinde başlık dosyası istenmesi ile arşiv dosyası istemenin anormallik skoru aynı olarak verilmiştir.

Komut Çalıştırma (Command execution)
Önemli web ataklarından birisi komut çalıştırma atağıdır. Komut çalıştırma atağında saldırgan kişi sistemin açıklarını kullanarak "cmd.exe", "dir" tarzında komutları ve uygulamaları çalıştırır ve daha sonra sistem üzerindeki yetkisini artırır. Dolayısıyla, URL satırında eğer komut çalıştırma atağında kullanılan sözcükler bulunursa bu web erişiminin de anormallik skoru daha yüksek olacaktır.

Giriş verisi içerisinde özel karakterlerin bulunması (special characters in input data)
Web atak tespitinde ataklarla ilişkili karakterlerin (., <, >, |, ; vs.) taranması da önemlidir. Bu karakterler genellikle ataklarda rastlanan özel karakterlerdir. Özel karakterler giriş doğrulama (input validation) ve kod enjeksiyonu (code injection) ataklarında kullanılır. Bu yüzden ataklar ile özel karakterler arasında bir ilişki vardır ve özel karakterler anormallik parametrelerinden birisidir. Eğer URL içerisinde ataklarla ilişkili karakterler bulunursa o isteğin anormallik skoru da artırılır.

Giriş verisi uzunluğu (the length of input data)
Buffer taşması atakları genellikle normalden daha uzun giriş verisinde meydana gelir. Uzun URL satırları URL'in bir atakla ilgili olduğunu işaret eder. Örneğin code red kurdu tampon bellek taşmasına sebep olur ve tampon bellek taşması ataklarıyla ilgili URL satırları normal URL satırlarından genellikle daha uzun olmaktadır. Dolayısıyla ataklarla uzun sorgu satırları arasında bir ilişki vardır ve sorgu satırlarının uzunluğu anormallik parametrelerinden biridir.

3.2. PARAMETRELERİN SUNUMU (PARAMETERS REPRESENTATION)

Bir önceki bölümde bahsedilen parametrelerin tamamı metin tabanlı anormallik tespiti yönteminin özellik kümesini meydana getirir. Her bir erişim, bir erişim vektörü ile sunulur ve her bir parametre ikili değerlerden (0, 1) birini alır. Erişim vektörü ikili değerler tutmaktadır. Özellik setinde 50 adet özellik bulunmakta olup özellik setinin detayları aşağıda verilmiştir.

Tablo 2. Özellik vektörü (attribute vector)

İstek Tipi	Web Olmayan Dokümanlar	Komut Çalıştırma	Özel Karakterler	Giriş Verisi Uzunluğu

Özellik vektörünün detayları aşağıda sırayla verilmiştir.

İstek tip'i detayları (details of request type)
GET POST HEAD PROPFIND OPTIONS

Web olmayan dosyaların detayları (details of non web files)

Arşiv dosyaları			Yedek dosyaları	Başlık dosyaları		Metin dosyaları
.zip	.tar.gz	.rar	.bak	.inc	.asa	.txt

Komut çalıştırma detayları (details of command execution)
dir ls-l .exe .dll … .bat

Özel karakterler(special characters)
.. < > ' | * - { } [ ] %5c %5e ... %2f

Erişim vektörlerinin genişletilmiş özellikleri Tablo 3'de verilmiştir.

Tablo 3. Genişletilmiş özellik vektörü (expanded attribute vector)

Get	Post	…	.zip	.rar	…	.exe	…	<	>	'	…

3.3. ANORMALLİK SKORU HESABI (ANOMALY SCORE CALCULATION)

Anormallik skoru anormallik tabanlı tespit yöntemlerinde önemli bir bilgidir. Eğer anormallik skor hesabının parametreleri iyi şekilde seçilebilirse o zaman sistemin başarı oranı artacaktır. Önerilen yöntemde her bir web erişimi, anormallik parametrelerinin bir vektörüyle sunulur. Bu vektörün parametreleri anormallik skoru hesabında kullanılır. Parametreler atak özellikleriyle ilişkili parametrelerdir. Hesaplanan anormallik skorlarına

göre web erişimleri normal ve anormal şeklinde sınıflanır ve anormal sınıfında yer alanlar atak olarak kabul edilir.

Anormallik skoru (AS) hesabı için;
E genişletilmiş özellik vektörü ve W ağırlık vektörü olmak üzere E ve W şöyle sunulur;

$\vec{E} = (e_1, e_2, ..., e_m)\, , \vec{W} = (w_1, w_2, ..., w_m)$, m burada parametrelerin sayısını vermektedir.

$e_i = \{0,1\}$ ve $w_i = \{0, \ldots, 100\}$ arasında değerler alır. Anormallik Skoru (AS) , her bir web erişimi için (1) ile hesaplanır;

$$AS = \sum_{i=1}^{m} (e_i * w_i) \quad (1)$$

, i={1,…,m} (burada skaler çarpım yapılmaktadır)

Eğer bir web erişiminin AS değeri belirlenen eşik değerinden daha büyük ise o zaman web erişimi anormal olarak etiketlenir. Anormal olarak bulunan bu web erişimi büyük ihtimalle bir ataktır.

3.4. KULLANICI ERİŞİM VERİSİ (USER ACCESS DATA)

Web kullanıcılarına ait erişim verileri web sunucu üzerinde web günlük (log) dosyalarında tutulur. Erişim veya diğer adıyla denetleme verisi genellikle kullanıcı davranışlarının analizinde kullanılır. Kullanıcı erişimlerinin her biri bir günlük kaydı olarak saklanır ve günlük kayıtları bazı sahalardan meydana gelir. Bu sahalardan birisi giriş verisi sahasıdır. Kullanıcının web sunucuya giriş olarak sunduğu veri log kaydındaki sorgu sahasında tutulur. Örneğin bir arama motoru ile "metin işleme" sözcük öbeğini arattığımızda sorgu sahasının değeri "metin işleme" olur. Kullanıcı web sunucudan bazı sonuçlar elde edebilmek için ona veri girişi yaptığında bu veriler kullanıcının sorgu kelimeleri olmaktadır. Görüldüğü gibi kullanıcı web sunucuya veri gönderebilmektedir. Gönderilen veriler kimi zamanda zararlı kod içerebilir. O zaman da sorgu verisi ile atak düzenlenmiş olmaktadır. Kullanıcı erişim verisi saldırganların tespiti için en önemli kaynaktır.

4. DENEYSEL ÇALIŞMA ve SONUÇLAR (EXPERIMENTAL STUDY and RESULTS)

Bu çalışmada web atak tespiti web erişimlerinin metin tabanlı analizi ile yerine getirilmektedir. Çalışmanın en önemli motivasyonu web erişimlerinin metinsel özelliklerinin onları ayırt etmede kullanılabilmesidir. Önerilen yöntemin testi için Gebze Yüksek Teknoloji Enstitüsü web sitesine gelen istek verileri kullanılmıştır. Bu verilerin içerisinde normal ve atak verisi bulunmaktadır. Örnek web sorgu verileri Tablo 4'de sunulmuştur.

Tablo 4. Örnek Web Sorgu Verisi (Sample Web Query Data)

İstek tipi	URL Yolu	Sorgu Verisi
HEAD	/scripts/check.bat/..┴£..┴£..┴£winnt/system32/cmd.exe	/c%20dir%20C:\?/c+dir+c:\
GET	/scripts/..%5c%5c../winnt/system32/cmd.exe	/c+dir
GET	/scripts/..%5c%5c../winnt/system32/cmd.exe	/c+dir
HEAD	/scripts/check.bat/.. └».. └».. └»winnt/system32/cmd.exe	/c%20dir%20C:\?/c+dir+c:\
HEAD	/scripts/check.bat/..┴£..┴£..┴£winnt/system32/cmd.exe	/c%20dir%20C:\?/c+dir+c:\
GET	/scripts/root.exe	/c+dir
HEAD	/adsamples/.. └»../.. └»../.. └»../winnt/system32/cmd.exe	/c+dir?/c+dir+c:\
HEAD	/adsamples/..┴£../..┴£../..┴£../winnt/system32/cmd.exe	/c+dir?/c+dir+c:\
HEAD	/adsamples/check.bat/.. └/.. └/.. └/winnt/system32/cmd.exe	/c+dir?/c+dir+c:\
HEAD	/etc/passwd	/c+dir+c:\
HEAD	/iisadmpwd/achg.htr	/c+dir+c:\
HEAD	/iisadmpwd/..%2f..%2f..%2f..%2f..%2f..%2fwinnt/system32/cmd.exe	/c+dir+c:\
GET	/lisans/yenibolumders.asp	bno=104&dno=BIL%20102&bolumad=Bilgisayar%20M³hendislii
GET	/lisans/yenibolumders.asp	bno=104&dno=BIL%20102&bolumad=Bilgisayar%20M³hendislii
GET	/default.ida	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXcbd3%u7801%u9090

Web Sorgu Verisi = URL Yolu + Sorgu Verisi

Tablo 4'de ham halde bulunan veriler önce ikili veriye dönüştürülür. Dönüşümde, web erişimi içerisindeki özel karakterler, anahtar kelimeler vs. taranır. Eğer web sorgu verisi içerisinde özel karakterlerden veya anahtar kelimelerden biri varsa erişim vektörünün ilgili boyutu TRUE (1) değerini alır. Daha sonra, erişim vektörlerinden bir anormallik skoru elde edilir. Önerilen sistemde her bir parametre erişim vektörünün bir boyutu ile eşleşmektedir.

GET /scripts/root.exe /c+dir

Tablo 5. Erişim Vektörü Örneği - kısaltılmış (Sample Access vector)

get	…	zip	.rar	…	.exe	…	<	'	|	…
1	…	0	0	…	1	…	0	0	0	….

Önerdiğimiz sistem veri toplamadan analize kadar bütün işlemleri web betikleri yardımıyla yerine getirmektedir. Geliştirdiğimiz sistem için veri toplama ve analiz amaçlı fonksiyonlar yazılmış olup bunlar veri analizinin yapılacağı betiklere eklenerek rahatça kullanılabilirler. Kullanım esnasında kullanıcıdan gelen istek verisi işleme alınmadan önce hazırladığımız fonksiyonlardan geçirilerek içerisinde zararlı kodlar olup olmadığı bulunabilmektedir. ASP ile geliştirdiğimiz veri analiz fonksiyonları

Request.Querystring ve Request.Form nesneleri ile alınan bütün verilere uygulanabilir durumdadır. Ayrıca analiz sonrası rapor fonksiyonları yardımıyla da bir veritabanı veya bir metin dosya içine zararlı kodların bulunduğu istekler daha sonra kullanılmak üzere kaydedilir. Dolayısıyla asp betiklerini destekleyen her web sunucuda kolayca kullanılabilir ve o sistemlere kolayca entegre edilebilir.

4.1. SONUÇLAR ve YORUMLAR (RESULTS and INTERPRETATIONS)

GYTE web sitesine gelen istekler değerlendirilip istekler puanlandığında Tablo 6'daki değerler elde edilmiştir. Bu değerlerin minimumu 0 (sıfır) ve maksimumu 80'dir. Sistem için normal durum sıfır değerine yakınlık manasına gelmektedir. Normal ile anormal isteklerin birbirinden ayrıldığı değer ise 30'dur. Sistemimiz anormal durumlara daha yüksek puanlar vermek yoluyla anormal durumları normal durumlardan ayırt etmektedir.

Sistem test edilirken çevrimdışı 56244 adet web erişim verisi kullanılmış olup bu veriler testten önce uzman yardımıyla etiketlenmiştir. Bu verilerden 53051 tanesi normal 3193 tanesini anormaldir. Önerdiğimiz sistem için eşik değeri 30'dur, sistemimize göre anormallik skoru 30 puandan daha fazla olanlar anormal olarak 30 puandan düşük olanlar normal olarak tespit edilmiştir. Etiketi NORMAL olduğu halde sistemimiz tarafından ANORMAL olarak tespit edilen erişimlerin adedi 46'dır. Etiketi NORMAL olan ve sistemimiz tarafından da NORMAL olarak tespit edilenlerin adedi ise 53005'tir. Dolayısıyla NORMAL erişimler için doğru tanıma oranı yüzde 99,9133'dir. ANORMAL erişimler için doğru tanıma oranı ise yüzde 97,9956'dir.

Anormallik tespiti yaparken eşik değerinin ne olması gerektiği önemli bir konudur. Çünkü bu değer tanıma doğruluğunu etkileyen konulardan biridir. Eşik değeri verilirken genellikle uzman görüşünden faydalanılır. Yani sistemin kurulması esnasında o sistem için hangi eşik değerinin normali verdiği bilinerek sabit bir değer verilir. Daha sonra bu değer doğru tanıma oranına bakılarak aşağı yukarı hareket ettirilir. Yinelemeli bir yapıda ve deneysel olarak bu değerin optimumu bulunur. Ayrıca her anormallik tespiti sistemi için eşik değeri farklı olacaktır.

Tabloda görüldüğü üzere yöntemimiz web erişim verilerini yüksek doğrulukla tespit etmektedir. Hatalı tanıma ile ilgili satırlar hatalı pozitif ve hatalı negatif değerleri vermektedir. Dolayısıyla yöntemimizin içinde zararlı kod bulunan web erişimlerini yakalama oranı oldukça yüksektir. Zararlı içeriğin bulunduğu türden web atakları için yöntemimiz başarılıdır. Yöntemin code red için başarı oranı yüzde 100 olup path traversal ve code injection için yüzde 97-99 arasındadır. Diğer tip ataklar için oran %1 oranında düşüktür.

4.2. DİĞER YÖNTEMLERLE KARŞILAŞTIRMA (COMPARISONS WITH OTHER METHODS)

Bu çalışmada önerilen metin tabanlı anormallik tespiti yöntemi web tabanlı atakları analiz için yeni bir yaklaşımdır. Web erişimlerindeki anormallikleri bulmak için web günlük verilerinin analizi konusundaki benzer çalışmalar Kruegel ve Vigna [5] ile Cho ve Cha [6] yapılmıştır. Bu yöntemlerin karşılaştırması Tablo 7'de verilmiştir.

Tablo 6. Anormallik skoruna dayalı doğru tanıma tablosu (accuracy table based on anomaly score)

	Puan	Miktar	Anormal	Normal
Normal	0	5	5	0
Normal	10	52917	23	52894
Normal	20	129	18	111
Anormal	30	67	3	64
Anormal	40	3	3	0
Anormal	50	1208	1208	0
Anormal	60	1468	1468	0
Anormal	70	167	167	0
Anormal	75	7	7	0
Anormal	80	273	273	0
Toplam		56244	3175	53069

	asıl etiket	tespit edilen	tanınma	toplam	oranlar
Hatalı Tanıma	N	AN	46	53051	0,000867
Doğru Tanıma	N	N	53005	53051	0,999133
Hatalı Tanıma	AN	N	64	3193	0,020044
Doğru Tanıma	AN	AN	3129	3193	0,979956

5. SONUÇLAR (CONCLUSIONS)

Bu çalışmada web ataklarının tespiti için giriş verisinin metinsel özellikleri kullanan bir yöntem önerilmektedir. Yöntem atak tespiti için anormallik

tabanlı bir yaklaşım ile çalışır. Her anormallik tabanlı yöntemde olduğu gibi yöntemimizin de bir eşik değeri vardır ve bu eşik değeri deneysel olarak elde edilmiştir. Eşik değeri web erişimlerini normal ve anormal şeklinde iki temel gruba ayrır. Bununla birlikte, eğer eşik değeri hatalı verildiyse o zaman atak tespitindeki hata miktarı artacaktır.

Önerilen yöntemde bütün web erişimleri birer erişim vektörü tarafından sunulur. Erişim vektörleri birinden diğerine farklılıklar göstermektedir. Web ataklarına ait giriş verilerinin metinsel özellikleri genellikle birbirine benzerdir. Benzer erişim vektörleri benzer

Tablo 7. Yöntemimiz ve diğerleri arasında karşılaştırmalar (comparisons between our method and the other methods)

Özellik	SAD	Diğer Yöntemler Krugel &Vigna	WAMTAT
Detaylı Bilgi	Oturum tabanlı, Parameter estimation tekniğini kullanıyor.	Özellik boyu ve özellik karakter dağılımını kullanıyor	Web erişimlerinin metinsel özelliklerini kullanıyor.
Giriş Verisi	Web oturumları	Sunucu taraflı programlara değer göndermek için kullanılan parametrelerin isimleri	Web sorgu verisi (sorgu parametreleri ve değerleri).
Anormallik Skoru Hesabı	Bayesian parameter estimation tarafından hesap edilen alt oturumların maksimum puanları.	x^2-testlerinden idealleştirilmiş karakter dağılımları bulunur.	Web erişimlerinin metinsel özelliklerinden elde edilen parametre vektörü ile ağırlık vektörü çarpımlarından elde edilir.
Test Verisi	Whisker tarafından türetilen web log verileri kullanılmıştır	Google ve Kaliforniya üniversitesi web sunucu günlüklerinden elde edilmiştir.	GYTE Web günlük dosyaları
Tespit verimiliği	Bütün whisker taramalarında 91% . Eğer bazı atak tipleri çıkartılırsa skor 99% ' a yükseliyor.	Karakter dağılımı ve yapısal model broad range ataklarına karşı çok etkili.	Karakter duyarlı ataklarda başarılı.
Eğitim Süresi	Uzun eğitim süresi daha doğru sonuç üretiyor .	Uzun eğitim süresi daha doğru sonuç üretiyor.	Kısa
Sistem sınırlamaları	Sadece sayfa sıraları ve sıklıkları analiz ediliyor..	Web erişim günlüklerinin güvenilirliği Sistemin sınırlamasıdır.	Parametre ağırlıkları doğru seçilmezse doğru tanıma oranı düşüyor.

giriş verisinden elde edilmektedir. Böylece, web erişimlerinin metinsel özelliklerini kullanarak web ataklarını tespit edebiliriz. Bu yöntem özellikle; code red, path traversal ve code injection ataklarında başarılı olmaktadır.

Yöntemimizde bir web erişiminin atak olup olmadığının tespiti için 50 civarında özellik kullanılmakta olup özellik adedinin az olması etkinliği artırmaktadır. Yöntemin özellik kümesinde ataklara özel karakterler ve kelimeler bulunduğu için özellik kümesi küçüktür.

KAYNAKLAR (REFERENCES)

1. Foltz, C. B., Cyberterrorism, **Computer Crime, and Reality, Information Management&Computer Security**, vol 12, no 2, 2004, p.154-166.
2. Security Tracker. Vulnerability statistics April 2001-march 2002. http://www.securitytrcker.com/learn/statistics.html, April 2002.
3. Rubin A. D. and Geer Jr. D. E., "A Survey of Web Security", **IEEE Computer**, Vol. 31, No. 9, September 1998, pp. 34-41.
4. Gordon, L.A., Loeb, M.P., Lucyshyn W. and Richardson R., 2004 CSI/FBI Computer Crime and Security Survey. 2004, available at http://gocsi.com
5. Kruegel C., Vigna G., (2003), Anomaly Detection of Web-Based Attacks, **Proceedings of the 10th ACM Conference on Computer and Communication Security** (CCS '03) ACM Press Washington, DC. pp. 251-261.
6. Cho S., Cha S., SAD:Web Session Anomaly Detection Based on Parameter Estimation, **Computers & Security**, Volume 23, Issue 4, June 2004, pp. 312-319
7. Vigna G. , Valeur F., and Kemmerer R.A., Designing and Implementing A Family of Intrusion Detection Systems, **Proceedings of the European Conference on Software Engineering** (ESEC) Helsinki, Finland September 2003.
8. Alvarez G., Petrovic S., A new taxonomy web attacks suitable for efficient encoding, **Computers & Security**, vol. 22, 2003, pp. 435-449.
9. Kruegel C., Toth, T. and Kirda E.., Service Specific Anomaly Detection for Network Intrusion Detection. **Proceedings of Symposium on Applied Computing**(SAC). ACM Scientific Press, March 2002.
10. Vigna G. , Robertson W. , Kher V. , and Kemmerer R.A. , A Stateful Intrusion Detection System for World-Wide Web Servers, **Proceedings of the Annual Computer Security Applications Conference** (ACSAC) 34-43 Las Vegas, NV December 2003
11. Takci H., Sogukpinar I., Centroid-Based Language Identification Using Letter Feature Set, **Lecture Notes in Computer Science**, Vol. 2945/2004, February 2004, pp. 635-645.