AĞ GÜVENLİĞİ

Bahar 2005

SALDIRI TESPİT SİSTEMLERİNDE İSTATİSTİKSEL ANORMALLİK BELİRLEME KULLANIMI

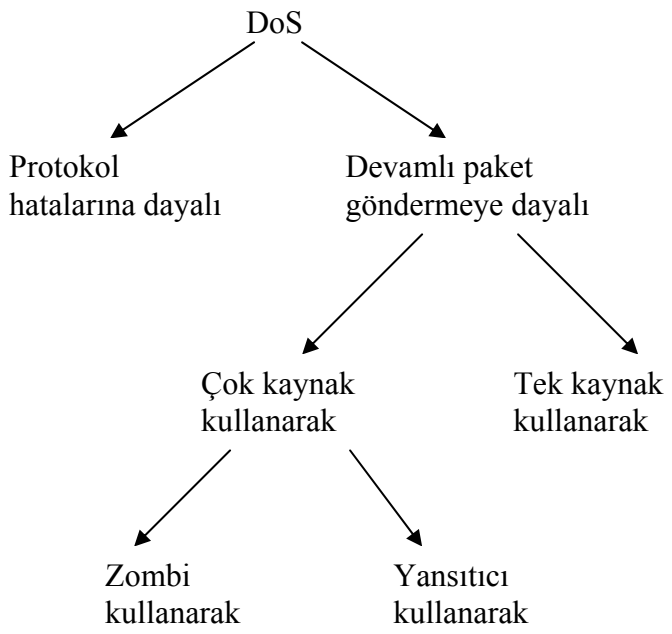
MELİKE EROL, Yük. Müh.
504032511

1. GİRİŞ

Bilgisayar sistemlerinde en genel anlamda sızma ya da saldırı makine başından yapılacak izinsiz erişimlerden başlar ve çok geniş bir spektruma yayılır. Bilgisayar ağları söz konusu olduğunda ise saldırılar sadece bu tip kullanıcı ve erişim temelli saldırılar ile sınırlı kalmaz. Ağ üzerinden yapılan saldırılar günümüzde en sık karşılaşılan problemlerdir. Bu çalışma ağ üzerinden yapılan saldırıları ve onlara karşı geliştirilen saldırı tespit mekanizmalarını incelemeyi hedeflemektedir. Bu tip saldırılar 4 temel kategoride incelenebilirler.

1) **Bilgi Tarama (Probe ya da scan):** Bu saldırılar bir sunucunun ya da herhangi makinanın, geçerli ip adreslerini, aktif portlarını veya işletim sistemini öğrenmek için yapılan saldırılardır. Bilinen saldırılardan bunlara örnek olarak:
   a. ipsweep: belirli bir protu sürekli tarama saldırısı.
   b. Portsweep: bir sunucu üzerindeki hizmetleri bulmak için tüm portları tarar.
   verilebilir.

2) **Hizmet Engelleme (Denial of Service - DoS):** Bu saldırılar genelde TCP/IP protokol yapısındaki açıklardan faydalanarak veya bir sunucuya çok sayıda istek yönelterek onu tıkamaya sebep olan saldırılardır. DoS saldırıları kendi içinde gruplara ayrılır [18].



Şekil 1. DoS saldırı tipleri

Protokol hatalarına dayalı saldırılara örnek olarak ping-of-death (ölümüne ping) saldırısı yani bir tek büyük boyutlu ICMP eko mesajı gönderilmesi saldırısı vardır. Başka bir saldırı, TCP SYN paketinin içersine kaynak ve varış adresi aynı makine olan bir paket gönderilmesiyle olur. Bunlar tek paketle ya da az paketle gerçekleştirilen, protokollerin açıklarını kullanan saldırılardır. Diğer gruptaki saldırılar devamlı istekte bulunulmasına dayanır. Hem sunucu makine hem de ağ meşgul edilir. Örneğin bu, bir sunucuya devamlı bağlantı isteği yapmak olabilir. Saldırı tek bir makine kaynaklı olabileceği gibi ağ üzerinden ele geçirilen birçok makine ile de yapılabilir. Bu tip makinelere zombi denir. Eğer zombi kullanılmıyorsa yansıtıcı da kullanılabilir. Yansıtıcı herhangi bir sunucu

olabilir. Saldırgan kurbanın adresi ile sorgu yapar ve dolayısıyla sunucu farkında olmadan kurbanın adresine cevaplar gönderir. Özel olarak kullanılan saldırlardan bazıları:

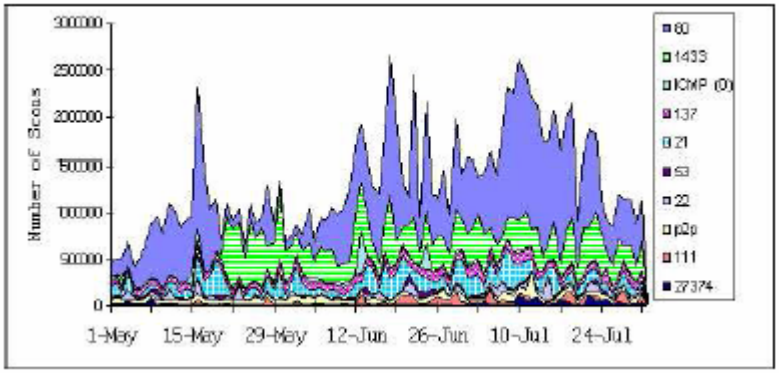
a. Smurf: ICMP mesajlarının broadcast ile tüm ağa dağıtılmasıyla oluşur.
b. Selfping: kullanıcının makinayı sürekli pinglemesiyle gerçekleşir.
c. tcpreset: saldırgan kurbanın kurmaya çalıştığı bağlantılar için kurban adına reset göndererek bağlantısını engeller.
d. mailbomb:saldırgan sunucuya sürekli mail gönderir.

3) **Yönetici Hesabı ile Yerel Oturum Açma (Remote to Local - R2L**): Kullanıcı haklarına sahip olunmadığı durumda misafir ya da başka bir kullanıcı olarak izinsiz erişim yapılmasıdır. Bunlara örnek:
   a. Sshtrojan: unix üzerinde çalşına bir trojan saldırısıdır.
   b. guest: tahmini kolay şifreleri bularak sisteme girilmesidir.

4) **Kullanıcı Hesabının Yönetici Hesabına Yükseltilmesi (User to root - U2R):** Bu tip saldırılarda sisteme girme izni olan fakat yönetici olmayan bir kullanıcının yönetici izni gerektirecek işler yapmaya çalışmasıdır. Örnekleri:
   a. Eject: Solaris üzerinde eject programı ile tampon taşmasına yol açıp, yönetici haklarına sahip olunmasıdır.
   b. Sqlattack: sql veritabanı kurulu Linux makinalarda sunucuya bağlanan kullanıcının belirli komutlarla yönetici hakları ile komut satırı elde etmesidir.

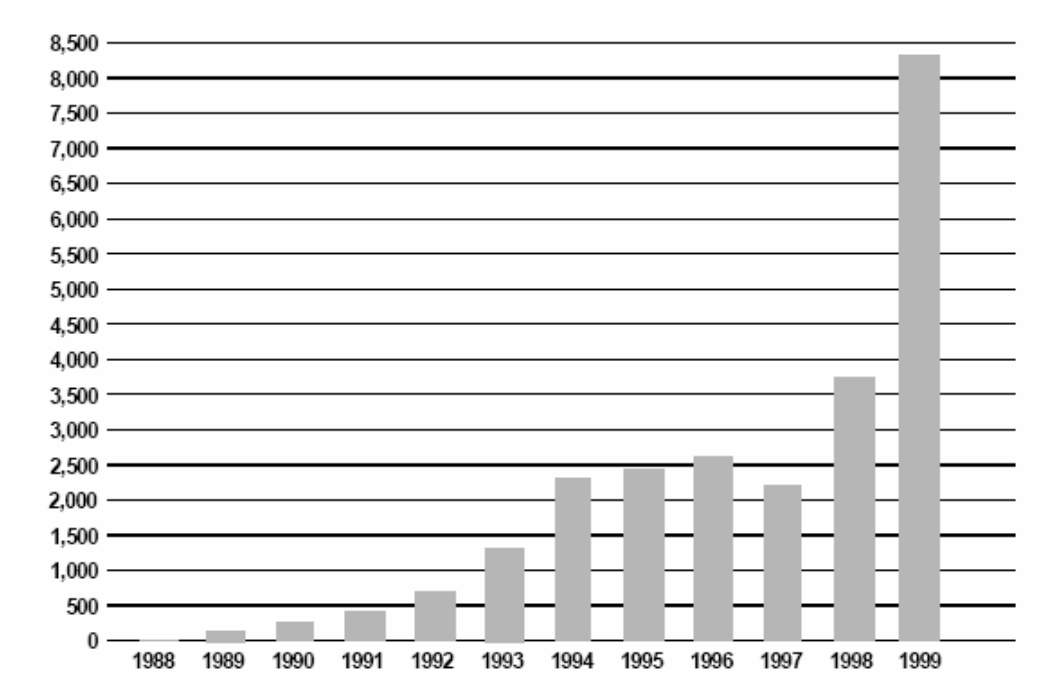
Bu çalışma çerçevesinde tüm bu saldırıları yakalamaya odaklanan saldırı tespit sistemlerinin yapıları, kullandıkları tespit yöntemleri genel olarak tanıtılacaktır. Fakat özelde odaklanılacak konu bir kullanıcının hakkı olmadan sisteme giriş yapmak istemesi veya sisteme girmeye hakkı olup da bazı işlemlere izni olmayan kullanıcının bu izinleri aşmaya çalışması gibi saldırılar olmayacaktır. Genel tanımlar verildikten sonra daha çok ağ üzerinden yapılan saldırı tespitine yönelik çalışmalara odaklanacağız.

Ağ trafiğini esas alarak, burada oluşan anormallikleri gözlemleyerek, saldıryı belirlemek mümkündür. Bu tip modellere ağ aktivitesi modelleri (Network activity models) adı verilir ve ağdaki trafik yoğunluğu üzerinden saldırı tespitine odaklanırlar. Bu durumda tespit edilebilecek saldırılar yukarda sıralanan saldırılardan 3. ve 4. maddelere giren ve en yaygın saldırılardan olan DoS ve tarama (probe) saldırıları olacaktır. Bu tipteki saldırıların ne kadar önemli olduğu 2002 yılında yapılan bir çalışmada gösterilmiştir. Burada internet üzerindeki 1600 ağdan toplanmış verilerin analizi yapılmıştır [1]. İnternet üzerinde hergün 25000 civarında saldırı gerçekleştiği saptanmıştır. Bu çalışmadan alınan aşağıdaki grafikte sadece yoklama (probe/scan) saldırısı göz önüne alındığında Mayıs-Temmuz ayları arasındaki kapı numaralarına karşı düşen saldırı yoğunluğu verilmiştir. Tek bir saldırı tipi için verilen Şekil 2, durumun ne kadar ciddi boyutlarda olduğunu göstermektedir. Bundan başka sadece 1999 yıllında NASA, NATO, Beyaz saray, Pentagon gibi merkezlere yapılan ve başarı sağlanan saldırıların sayısı 10'u geçmektedir [17].



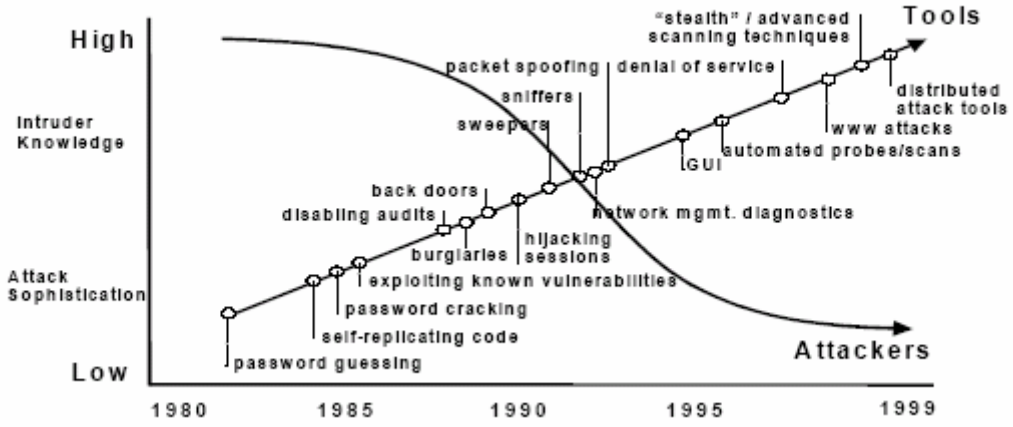
Şekil 2. Mayıs-Temmuz 2002 arasında 10 kapı numarasına yapılan saldırı yoğunluğu [1].

1980'lerde sadece bilgisayar sistemleri uzmanları tarafından yapılan saldırılar artık otomatikleşmiş saldırı araçları ile çok daha hızlı sürede yapılmaktadır. CERT (Computer Emergency Readiness Team) tarafından 2000 yılında yayınlanan bir rapordan alınan değerler Şekil 3'te gösterilmiştir. Saldırı sayılarındaki artışın boyutları daha o senelerden çok ciddi boyutlara ulaşmış ve giderek elektronik ticaret uygulmalarının da artmasıyla fazlaca maddi kayıba yol açmaya başlamıştır.



Şekil 3 CERT'e raporlanan saldırı sayısı [17].

Şekil 4'te ise saldırgan davranışlarındaki değişim ve saldırı tiplerinin yıldan yıla değişimi verilmiştir [17].



Şekil 4 1980-1999 arası saldırı tiplerindeki ve saldırgan profillerindeki değişiklikler [17].

Saldırı tanımını yaptıktan sonra Saldırı Tespit Sistemi (STS) tanımını yapmak gerekir. Saldırı tespitinde kullanılan sistemlere STS denir. Her ne kadar bir STS'nin hedefi DoS, tarama, U2R, R2L saldırılarını yakalamak olsa da bu pratikte pek mümkün olmamıştır. Ticari, devlet tarafından geliştirilen ve araştırmacılar tarafından geliştirilen bir çok STS mevcuttur. Sayıları çok fazla olduğundan çeşitli sınıflandırmalara göre incelemek yerinde olacaktır. STS'ler değişik yaklaşımlar ile sınıflandırılabilirler. Bunlar sistemin gerçek zamanlı olup olmaması, dağıtık mimaride olup olmaması, ağ veya sunucu tabanlı olması ve kullandığı analiz yöntemlerine göre dört adet sınıflandırmayı içerir.

Bu çalışmada ilk iki sınıflandırma üzerinde fazlaca durmayacağız. Kendi gerçekleyeceğimiz basit analiz için gerçek zamanlı ve dağıtık olmayan, ağ tabanlı bir yapı ve anormallik tespit yöntemleri kullanılacağı için bunlar üzerinde daha fazla duracağız. Bu sınıflandırmalara daynarak gruplanan ticari ve akademik STS'lerin kısaca tanıtılmasının ardından iyi bir STS'de olması gereken özellikleri Bölüm 3'te vereceğiz. Saldırılara karşı verilebilecek tepkileri Bölüm 4'te inceledikten sonra Bölüm 5'te birçok çalışmada kullanılan DARPA saldırı tespit test veri setlerini tanıtacağız. Daha sonra 6. Bölümde kendi analizimizde örnek alabileceğimiz yakın tarihli çalışmalar hakkında detaylı bilgi vereceğiz. Burada çalışmamızın yakından ilgili olduğu anormallik tespinde istatistiksel yaklaşımları kullanan makale ve bildirileri inceleyeceğiz. Son olarak da Bölüm 7'de bu ders kapsamında gerçeklemeye çalıştığımız kendi uygulamamız hakkında bilgi vereceğiz.

2. STS'LERİN SINIFLANDIRILMASI

2.1 Veri İşleme Zamanına Göre:

Burada STS'ler "gerçek zamanlı" ve "gerçek zamanlı olmayan" şeklinde ikiye ayrılırlar. Gerçek zamalı olmayan sistemlerde veri önce depolanır, sonra analiz için ilgili STS'ye gönderilir. Gerçek zamanlı sistemlerde veri o anda analiz edilir. Bu tip STS'ler yoğun bilgi akışı olan ağlarda uygulanması zor olan ama aktif olarak cevap üretilmesi gereken durumlarda da tek çözüm olan sistemlerdir. Ticari uygulamaların gerçek zamanlı olması tercih edilir. Biz daha kolay olması açısından biz kendi çalışmamızda depolanmış veriler üzerinde analiz yapacağız.

2.2 Mimari Yapıya Göre:

STS'ler ağ üzerinden birçok noktadan veri toplayıp bir merkezde işleyebilirler. Bu tip mimarilere sahip STS'lere dağıtık STS denir. Bunun karşısında ağ üzerinden tek bir yerden

veri toplayan STS'ler vardır. Bilinen STS'lerin çoğu bu kategoriye girer. Bizim hedeflediğimiz sistemde tek bir yerden veri toplanıp işlenmesi durumunu göz önüne almaktadır. Bunun için İnternet üzerinden aldığımız DARPA veri setlerine ait bir adet ağ dinleyicisi (sniffer) dosyasını kullanacağız (tcpdump.inside[19]).

2.3. Bilgi Kaynaklarına Göre:

STS'ler bilgi kaynaklarına göre 2'ye ayrılılar. Ağ temelli STS'ler ve sunucu-temelli STS'ler. Sunucu temelli STS'lerin özel bir hali olarak uygulama temelli STS'ler de mevcuttur.

Ağ-Temelli STS'ler

Bu STS'lerin temel amacı ağ üzerinden yapılan saldırıları ağ trafiğini gözetleyerek tespit etmektir. Ağ paketlerini yakalayıp bunları analiz ederek saldırı tespiti yaparlar. Son yıllarda geliştirilen ağ temelli STS'ler bizim çalışmamızın da temelini oluşturmaktadır. Bunlardan bir kısmı bölüm 6'da ayrıntılı olarak inceleneceklerdir.

Sunucu-Temelli STS'ler

Sunucu-temelli STS'ler bir bilgisayar sistemi içerisinde toplanan veriler üzerinde çalışırlar. Bu şekilde işletim sistemine yönelen saldırılar için hangi sistem çağrılarının ve hangi kullanıcıların sorumlu olduğu tespit edilebilir. Sunucu-temelli STS'ler iki tür bilgi kaynağı kullanır: işletim sistemi izleme seçenekleri ve sistem günlük dosyaları.

Bunların ağ temelli sistemlere göre bir avantajı: olayların olduğu yerel sunucuyu izleme yetenekleri sayesinde ağ-temelli STS'lerin yakalayamayacağı saldırıları tespit edebilirler. Bu saldırılar çoğu zaman fazla trafik yaratmayan R2L veya U2R saldırılarını içerir.

Uygulama-Temelli STS'ler

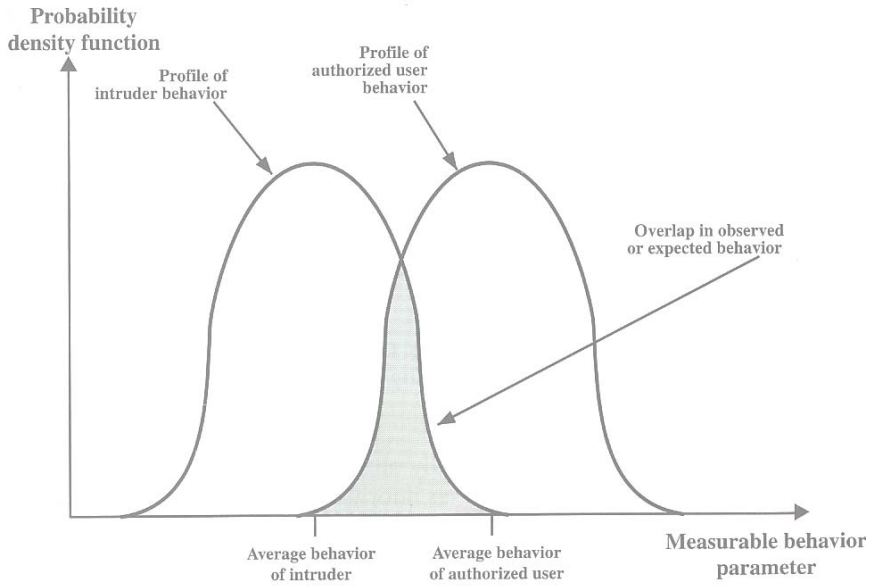
Bu STS'ler bir yazılım uygulaması ile meydana gelen olayları analiz eden sunucu-temelli STS'lerin özel bir türüdür. Uygulama-temelli STS'lerde bilgi kaynağı olarak genellikle uygulamaya ait günlük dosyalar kullanılır. Analiz-motoruna belirli uygulamalar için o uygulamalara has özellikler bildirildiği taktirde, yetkisini aşan kullanıcıların gerçekleştirdiği saldırılar uygulama-temelli STS'lerce tespit edilebilir [2].

2. 4 Saldırı Tespit Yöntemlerine Göre:

STS'ler saldırı tespit yöntemlerine göre sınıflandırılırsa 2 temel yöntem söz konusudur. Bunlardan ilki anormallik tespitine dayanır. Diğeri kötüye kullanım tespiti ya da bazı kaynaklarda tanımlandığı üzere imza-tanıma üzerinden gider. Aşağıda iki yaklaşım da ayrıntılı olarak incelenecektir.

2.4.1. Anormallik Tespiti:

Anormallik (anomaly) normal davranıştan sapma anlamına gelir. Burada normal davranıştan farklılık göstren davranışların saldırı olarak işaretlenmesi söz konusudur. Normal bir sistemde kullanıcı istekleri tahmin edilebilir istatistiksel değerlerle uyuşur. Burada normal davranışın bilinmesi ve modelenmesi esastır. Ancak bundan sonra bir anormallik varsa tespit edilebilir. Normal davranış bir dağılım ile (bkz. Şekil 5) veya dağılımın momentleriyle modellenebilir. Belirli kurallar zinciri ile tanımlanabilir veya başka yöntemler uygulanabilir ama sonuçta bu tanımlı ya da öğrenilen davranıştan sapma anormallik olarak değerlendirilir ve sapmanın şiddetine göre saldırı olarak sınıflandırılabilir. Bu yöntemin avantajı daha önceden tanınmayan saldırıların keşfedilmesi olasılığıdır. Dezavantajı ise yanlış alarmların (false alarm/positive) sayısının yüksek olmasıdır [3]. Yanlış alarm tanımı bir sonraki bölümde ayrıntılı olarak verilecektir fakat kabaca "var olmayan" bir durumu "var" olarak tanımaktır .



Şekil 5 Normal ve saldırı içeren dağılımın teorik ayrışımı [20]

Anormallik tespitinde istatistiksel yöntemler, yapay sinir ağları, veri madenciliği bilgisayar bağışıklık sistemi (computer immunology) gibi birçok yaklaşım uygulanabilir [4]. Bunları kendi içinde sınıflandırmak gerekirse [5]:

a) **Kendi-kendine öğrenen sistemler:** Belli bir veri seti ile eğitilerek normal davranışı öğreniler. Bunlar i) zaman serilerini kullanmayan kural tabanlı sistemler ve açıklayıcı istatistikleri kullanan sistemler olarak ve ii) zaman serilerini kullanan yapay sinir ağları temelli teknikler olarak üçe ayrılırlar.

1. Zaman serisi kullanmayan sistemler:
   a. **Kural Tabalı (Rule Based) Sistemler:** Sistem kendisi trafiği inceleyip kurallar oluşturuyor ve saldırı tespiti sırasında bu kararlara göre davranıyorsa bu sınıfa girer.
   b. **Açıklayıcı İstatistikleri (Descriptive Statistics) Kullanan Sistemler:** Kullanıcı profilinin basit istatistiklerle oluşturulup, buradan uzaklık vektörlerini (distance vector) kullanarak karar alan sistemlerdir.

2. Zaman serisi kullanan sistemler:
a. Yapay Sinir ağları (Artificial Neural Networks – ANN) yaklaşımı: Burada sistem önceden bir zaman serisi ile eğitilir ve çalışmaya başladığında, buradan öğrendikleri ile karar verir.

b) Programlanan Sistemler:
   a. **Açıklayıcı İstatistikleri kullanan sistemler:** Burada yukardakinden farklı olarak profiller daha önceden tanımlanmıştır. Profilleri sınırlayan belli bir eşik değeri, bir istatistiksel büyüklük veya basit bir kural olabilir. Bu yaklaşımı kullanan çalışmaların örnekleri Bölüm 6'da incelenecektir.
   b. **Bilinmeyeni reddet (Default Deny) yaklaşımı:** Tek başına çok sık kullanılan bir yaklaşım olmasa da, sistem davranışındaki durum-geçiş değerlerini kontrol ederek, farklı bir geçiş (transition) görüldüğünde izin vermemek şeklinde bir yaklaşıma sahiptir.

c. **Sinyal İşleme Tekniklerini Kullanan yaklaşım:** Burada sistem parametreleri sinyal işleme teknikleriyle belirlenir ve saldırı içermeyen normal veri setinden çıkartılan eşik değerleri ile anormallik tespiti yapılır. Bu yaklaşım diğerlerine göre yenidir ve çok geniş bir alanı kapsamaktadır. Henüz sınırlı sayıda sinyal işleme tekniği saldırı içeren zaman serilerine uygulanmıştır. Bölüm 6'da bunların örnekleri üzerinde duracağız. Özellikle son yıllarda popüler olan temel bileşen analizi (Principle Component Analysis - PCA) yaklaşımı ile elde edilen özdeğerlerin sistem parameteresi olarak kullanılması bu alandaki ilginç çalışmalardan biridir.

2.4.2. İmza-tanıma Temelli Saldırı Tespiti:

Kötüye kullanım tespiti (Misuse detection) ya da imza-tanımaya dayalı sistemlerde ise her davranışın bir imzası-karakteri vardır. Bunlar daha önce görülen davranış şablonlarıdır. Eğer gözlenlenen davranış daha önceden bilinen bir saldırı imzası ile eşleşiyorsa saldırı olarak sınıflandırılır. Daha önce karşılaşılmadıysa saldırı olarak nitelenmez. Avantajı saldırıyı kesin olarak tanıyabilmesidir. Yani yanlış alarm vermezler fakat yeni bir saldırı gelirse bunu sezemezler. İmza temelli sistemler daha çok ticari sistemlerde kullanılırlar.
İmza-tanımaya dayalı sistmelere ait yaklaşımlar sınıflandırılırsa [5]:

a. **Programlanan Sistemler:** Kendi kendine öğrenen hibrid bir teknik olan Ripper dışında tüm imza temelli teknikler bu sınıfa girerler.
    a. **Durum modellemesi:** Burada tüm davranışlar durumlara karşı düşer. Eğer bir davranış daha önceden tanımlı durumlara ve durum geçişlerine denk düşen hareketler yapıyorsa saldırı olarak tanınır.
    b. **Exper sistemler:** Sızma belirlerme sistemlerin ilkleri kural-tabanlı (rule-based) exper sistemlerdir.
    c. **Örüntü eşleme (Pattern Matching):** Sistemde daha önceden tanımlanmış, olmaması gereken bazı sözcüklerin tanınmasına yardım eder. Esnek değildir fakat basittir. Örneğin "parola dosyasını kopyala" komutu görüldüğünde bunun bir saldırı olduğunu en basit şekilde bu yöntem tespit eder.

Aşağıda Tablo 1'de saldırı tespit sistemlerinin sınıflandırması şeması ve bunlara denk düşen gerçeklenmiş sistemler yer almaktadır. Bu tablo ve sınıflandırma 2000 yılında yapılan bir çalışmada yer aldığından en son gelişmeleri içermemektedir. Biz 6. Bölümde bu seneden sonra önerilen istatistiksel yaklaşımları içeren anormallik tespiti sistemleri inceleyeceğiz.

Tablo 1. Saldırı Tespit Yöntemlerinin ve Araçlarının Sınıflandırılması

Anormallik Tabanlı	Kendi-kendine Öğrenen Sistemler	Zaman serisi kullanmayan	Kural tabanlı	**W&S**
			Açıklayıcı İstatistikler	**IDES** **NIDES** **EMERALD** **Haystack**
		Zaman serisi kullanan	Yapay sinir ağları	**Hyperview**
	Programlanan Sistemler	Açıklayıcı İstatistikler	Basit istatistikler	**MIDAS** **NADIR** **Haystack**
			Basit kural	**NSM**
			Eşik değeri	**ComputerWatch**
		Bilinmeyeni Reddet	Durum Serileri Modeli	**JANUS** **Bro**
İmza Tanıma tabanlı	Programlanan Sistemler	Durum Modellemesi	Durum Geçiş	**USTAT**
			Petri-net	**IDIOT**
		Exper sistem	**NIDES** **EMERALD** **MIDAS** **DIDS**	
		Örüntü eşleme	**NSM**	
		Basit kural	**NADIR** **Bro** **HayStack**	
İmza esinlenmeli	Kendi kendine öğrenen	Otomatik özellik seçme	**Ripper**	

Tablo 1'de bazı sistemlerin birden fazla kategoride yer alması, ilgili yaklaşımların tümünü içerdiğini gösteriyor. Burada amacımız varolan bu sistemlerin incelenmesi olmadığından dolayı sistemler hakkında sadece kısa bilgi vermekle yetineceğiz.

Haystack: Amerikan Hava Kuvvetlerinde kullanılan Unisys 1100/60 ana bilgisayarları için tasarlanmış bir saldırı tespit sistemidir. 1988 yılında geliştirilmiştir.

MIDAS: 1988'de National Computer Security Centre (Ulusal Bilgisayar Güvelik Merkezi) için geliştirilmiştir. Exper sistem tabanlıdır.

IDES (Intrusion Detection Expert System): IDES 1988-1992 yılları arasında üzerinde çalışılan birçok sistemi içerir. Bu arada birçok versiyonu çıkmıştır ve son olarak NIDES (Next-Generation Intrusion Detection Expert System) adını almıştır. Bu çalışmalardan en eskilerinden ve ünlülerinden biri olan Denning'in sistemi [6] aşağıda daha ayrıntılı olarak

W&S (Wisdom and Sense – Zeka ve His): Geliştirilmesine 1984 yılında başlanmasına rağmen ilk makalesi 1989'da çıkmıştır. Zeka kısmı geçmişteki audit verileri inceleyerek normal davranışı oluşturması anlamına gelir. His kısmı ise bunların kural haline getirilip exper sisteme verildikten sonra anormal davranışların yakalanması anlamına gelmektedir.

ComputerWatch: AT&T Bell Laboratuarları tarafından ticari bir ürün olarak 1990 yılında geliştirilmiştir.

NSM (Network Security Monitor): NSM'de IDES gibi revizyonlardan geçmiştir ve geliştirilmesi 1990-94 arasına denk düşer. NSM ağı dinleyerek, ağın kullanımıyla ilgili bir profil geliştirir ve geçerli kullanımı onunla karşılaştırır. Elde edilen veri beklenen bağlantı verisiyle karşılaştırılır ve beklenen aralıkta çıkmayan her veri anormal olarak işaretlenir. Biz çalışmamızda NSM tipinde bir STS gerçeklenmesi üzerinde duracağız.

NADIR: 1991-93 yılları arasında Los Alamos Laboratuarlarında geliştirilmiştir. NADIR kullanıcılar hakkında haftalık istatistikler tutar. Daha sonra bu istatistikleri exper sistem kurallarıyla karşılaştırır.

Hyperview: 1992 yılında geliştirilmiş diğer sistemlerden oldukça farklı bir sistemdir. İki ayrı parçadan oluşur. İlki davranışları izleyen ve sınıflayan bir exper sistem, ikincisi buradan öğrendikleriyle eğitilen yapay sinir ağlarını içeren parçadır.

DIDS: 1992 yıllarında geliştirilmiş dağıtık mimarili sistemleri kapsar. Bu tip sistemlerde ağın değişik noktalarından veriler toplanır ve bir merkezde incelenir.

USTAT: 1993-95 yılları arasında geliştirilmiştir. Durum geçiş analizi yapar. Eğer bir davranış saldırı için tanımlı durum geçişlerini yapıyorsa saldırı olarak sınıflandırılır.

IDIOT: 1994-96 yılları arasında CERIAS (Center for Education and Research in Information Assurance and Security) geliştirilmiştir. Örüntü tanıma için petri-net'lerin kullanımına dayanır.

JANUS: 1996'da Berkley'de geliştirilmiş bir sistemdir.

EMERALD (Event Monitoring Enabling Responses to Anomalous Live Disturbances): 1997-98 yıllarında geliştirilmiş ölçeklenebilir, dağıtık bir STS'dir. Bu konuda çokca referans verilen çalışmalardan birdir.

Bro: 1988'de Vern Paxson tarafından geliştirilmiş bir STS'dir. Gerçek zamanda ağ trafiğini pasif olarak gözlemleyerek saldırı tespiti yapmaktadır. Birçok değişik özelliği kendinde barındırması açısından sıkça söz edilen bir STS'dir.

Ripper: 1999 yılında geliştirilen bu sistem veri madenciliği yaklaşımını kullanır. Ripper DARPA değerlendirmesine katılmış olan sistemlerden bir tanesidir.

Burada adı geçen sistemlerden başkaca birçok STS mevcuttur. Bundan sonraki bölümde IDES tipinden bir STS modeli incelenecektir.

Denning'e ait IDES tipinden STS modeli:

Sızma belirleme dendiğinde konunun temel taşlarından sayılan Denning'in çalışmasından bahsetmek gereklidir [6]. Bu çalışmada exper sistemin karar vermesi için 3 farklı istatistiksel ölçü tanımlanmıştır.

1. Eşik ölçüsü: Kullanıcının belirli bir aralıkta bir işlemi tekrarlamasına izin verilir. Eğer bu değer aşılırsa anormallik olduğu tespit edilir. Örneğin bir sisteme kullanıcı *n* den fazla hatalı giriş (login) denemesi yaparsa kullanıcı hesabı kilitlenebilir.
Burada normal davranış için belirlenen eşik değeri sistemden sisteme çok fazla değişkenlik gösterebileceği için zayıf bir yöntemdir.

2. İstatiksel momentlerin kullanılması: Belirlirli bir davranış için dağılım ve ilk iki momentin bilindiği varsayılırsa ve eğer değerler belirli bir aralıkta beklenen değerlerden dışarı çıkarsa bu davranış anormallik olarak kabul edilir. Eşik değerinde geçerli problemler burada da mevcuttur.

3.Markov modeli kullanılması: Sistem bir durum makinası veya Markov zinciri gibi düşünülür. Yeni bir olay olunca sistem yeni bir duruma geçer. Eğer bir olay oluşunca, düşük olasılıklı durum geçişine neden oluyorsa bu davranış kural dışı olarak kabul edilir. Anormallikler, tek olaya değil bir dizi olaya bağlıdır.

Denning'in kullanıcı hareketlerinden oluşturduğu audit kayıtları şu şekildedir. Örneğin bir dosyayı kopyalamak için Ali kullanıcısı şöyle bir komut yürütsün:

[ali@computer]# COPY GAME.EXE TO <Library> GAME.EXE

Bu komut için oluşturulacak kayıtlar şu şekilde olacaktır.

Kullanıcı	**İşlem**	**Üzerinde Çalışılan Nesne**	**Hata kodu**	**Kaynak Kullanımı**	**Zaman Damgası**
Ali	Execute	<Library>COPY.EXE	0	CPU=00002	11...60
Ali	Read	<Ali>GAME.EXE	0	RECORDS=0	11....61
Ali	Execute	<Library>COPY.EXE	Write-viol	RECORDS=0	11....62

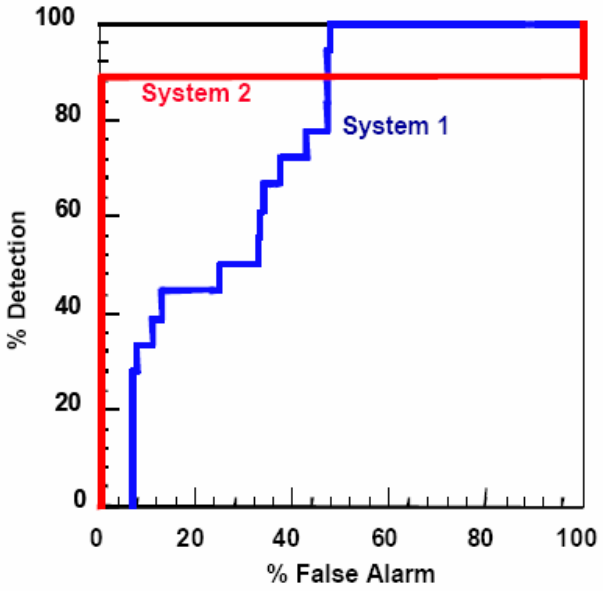
Bu kayıtlardan hangi kullanıcının hangi işlemi ne zaman yaptığı, herhangi bir hataya yol açıp açmadığı ve ne kadarlık kaynak harcayıp, kaç kayıt üzerinde değişiklik yaptığı okunabilmektedir. Bu kayıtlara bakılarak uygulanan yukarda anlatılan 3 istatistiksel ölçüt ile saldırının varlığını tespit edilir.

3. SALDIRI TESPİT SİSTEMLERİNİN BAŞARIMLARININ DEĞERLENDİRMESİ

Bu bölümde bir saldırı tespit sisteminin hangi özelliklere göre başarılı veya başarısız şeklinde değerlendirilebileceği üzerinde tartışacağız. Değerlendirme kriterleri aşağıda verilmiştir [16].

1- Yanlış Alarm Seviyesi

Yanlış alarm, tespit edilmesi planlanan büyüklük için yapılan yanlış değerlendirmeleri içerir. Yanlış değerlendirmeler iki çeşit olabilir. Biri var olan bir değeri kaçırmak diğeri varolmayan bir değeri varmış gibi tespit etmektir. Yanlış alarm ikinicisidir. Genelde STS sistemlerinin başarımını ölçmekte önemli bir parametredir çünkü bir sistemde izin verilen yanlış alarm sayısı ve doğru tespit miktarı birbiriyle ilintilidir. Yanlış alarmlara izin verildikçe doğru tespit oranı artmaktadır. Sistem parametreleri ikisinin de optimum olduğu noktaya ayarlanmalıdır. Burada yapılacak analiz ve başarım hesaplamasında ROC (Receiver Operating Characteristic – Alıcı Karakteristiği) eğrileri kullanılır. ROC adını radar uygulamalarından almıştır. Şekil 6'da iki sistem için ROC eğrileri verilmiştir. Burada örneğin %40'lık bir yanlış alarma izin verilirse sistemin doğru yakalama şansı %50 olur. Eğer %60'lık yanlış alarm kabul edilirse %100'luk bir doğru tespit şansı vardır.



Şekil 6 Örnek ROC eğrileri [16]

2- Doğru tespit seviyesi:
Kaç adet saldırılarının doğru olarak tespit edildiğidir. Yukarıda anlatıldığı gibi doğrudan yanlış alarm ile ilgilidir ve Şekil 6'daki y-eksenini oluşturur. Bir STS'nin var olan saldırıların kaçını yakaladığı önemli bir parametredir. Fakat bunu yaparken yanlış alarm sayısını da çok yükseltmemelidir.

3- Yoğun veri ile başa çıkabilme:
Bu kriter ağ tabanlı sistemler için geçerlidir. Trafik yoğunluğu arttıkça tüm sistemler paket düşürebilir hale gelirler. Bazı STS'ler trafik yoğunluğu çok arttığı zaman saldırı tespitinde tamamen başarısız olabilirler. Özellikle DoS saldırlarında sistemin güvenli olması için bu önemli bir parametredir.

4- Olayları ilintilendirebilme:
Farklı kaynaklardan gelen saldırı bilgilerini birleştirebilme yeteneğidir ve halihazırdaki STS'lerin bu konuda sınırlı yeteneği vardır.

5- Yeni saldırıları tanıyabilme:
Bu sadece anormallik tespitine dayalı sistemler için geçerlidir. İmza temelli sistemler yeni gelen hiçbir saldırıyı tanıyamazlar ve bir çok ticari STS bu tip bir yöntem benimsediği için bu yetenekten zaten yoksundurlar.

6- Saldırıyı tanıyabilme:
Saldırnın varlığını tespit etmek ve saldırının tipini söyleyebilmek iki ayrı kavramdır. STS'lerin ilk odaklandığı konu doğal olarak saldırıların varlığını tespit etmektir. Genelde saldırı tipi, saldırı olduktan sonra ağ yöneticisi tarafından çeşitli kayıtlar incelenerek ortaya çıkartılır. Fakat gene de çok genel sınıflandırmalar yapabilen STS'ler mevcuttur.

7- Saldırının başarılı olup olmadığını belirleme:
Aslında gerçekleştirilen saldırıların hepsi başarı ile sonuçlanmaz yani sisteme zarar vermez. Genelde STS'ler sisteme zarar veren veya vermeyen saldırıları sınıflandırmak konusunda bir çalışmayı barındırmazlar. Fakat bu özellik bir STS'nin sahip olması istenen özelliklerdendir.

4. SALDIRILARA KARŞI ALINACAK ÖNLEMLER:

Saldırı olduktan sonra alınabilecek önlemler saldırının tekrar olmasını engellemeye yarar ve o andaki saldırı için hiçbirşey yapamazlar. Bu önlemler örneğin yazılım açıklarından kaynaklanan saldırıları önlemek için yazılım yaması yapmak olabilir. Ama asıl önemli olan saldırı anında önlem alabilmektir. Saldırılara karşı iki farklı önlem gerçeklenebilir; pasif gözetleme ve erişim engelleme.

a) Pasif Gözetleme: Bu yaklaşımda saldırganın davranışları kaydedilir, saldırıda araya girilmez. Amaç, saldırganın istediği şeyi ve kullandığı yöntemleri öğrenmektir. Burada en büyük sorun, sistemin savunmasız bırakılmasıdır. Saldırgan bu savunmasız durumda diğer sistemlere de saldırabilir.
b) Erişimi Engelleme: Burada amaç saldırganın erişebileceği sistem kaynaklarını sınırlamaktır.

Saldırıya karşı yukardaki önlemler alındıktan sonra, saldırıya karışmış sistemlerin erişimi kaldırılır veya süreçleri sonlandırılır.

STS'leri genel olarak tanıttıktan sonra daha özelde yeni geliştirilen STS'ler ve önerilen yeni yöntemler üzerinde biraz bilgi vermek gerekir. Bu yakın tarihli çalışmalara geçmeden önce bu çalışmaların kullandığı veri setlerinin özelliklerinden bahsetmek yerinde olacaktır. Bundan sonraki bölümde STS'lerin kullanabileceği MIB tabanlı veri setleri ve internetten de bulunabilecek olan DARPA 1998 ve 1999 saldırı tespit değerlendirme veri setleri ile KDD Cup veri setlerini tanıtacağız.

5. STS TESTLERİNDE KULLANILAN VERİ SETLERİ

STS'lerin sınanmasında en zorlu kısım güvenilir veri setleri elde edilmesidir. İnternet ortamından elde edilmiş verilerin çoğu saldırı var ya da yok bilgisi içermez. Herhangi bir ağda bu bilginin edinilmesi için ağ yöneticisinin sistemi gözlüyor olması ve sistem kayıtlarını tutuyor olması gerekir. Bu birçok ISP için çok masraflı ve gereksiz görülen bir iştir. Saldırıları içeren veri setine tümüyle hakim olunmak isteniyorsa, bir benzetim (simulasyon) yapılarak sentetik test verisi yaratılabilir. Fakat bunun da internet trafiğine benzediğini kanıtlamak zordur. Bu konudaki zorlukları kısaca özetlersek:

1- Gerçek veri bulmak zordur.
2- Atak sırasında etiketlenmiş veri seti bulmak daha da zordur.
3- Verileri saklamak masraflıdır.
4- Normal davranışı tanımlamak ve simule etmek zordur (çünkü internet trafiği fazlasıyla dinamiktir.)

Bütün bu zorluklara rağmen STS'lerin test edilmesi için bir veri seti şarttır. Bu durumda 2 değişik veri toplama yöntemi kullanılabilir [7]. Bir yöntem yönlendirici MIB'inde (Management Information Base – Yönetim Veri Tabanı) tutulan, giren-çıkan IP paketi ve bunun gibi bazı sayaç değerlerinin SNMP ile sorgulanarak bir makinaya aktarılmasından ibarettir. Bu yöntem sıkça kullanılmaktadır çünkü maliyeti düşüktür. Fakat veri 5 dakikada bir sorgulandığından ve 5 dakikanın ortalaması sonuç olarak geri döndüğünden dolayı yeterince hassas değildir. Ayrıca ağ yöneticisi ile işbirliği içinde olunması gerekir aksi takdirde her yönlendirici üzerindeki bilgilere erişim hakkı tanımaz. Bir diğer problem bu yöntemle sadece yönlendirici MIB'inde tutulan verilerin elde edilebilir olmasıdır. Yani sayaç değerleri dışında, örneğin gidip-gelen paketlerin içeriği gibi bilgilere kesinlikle ulaşılamaz. Burada paket içeriğinde belirli bir anahtar kelime saldırı yapıldığı anlamına geliyorsa, tespit edilmeden

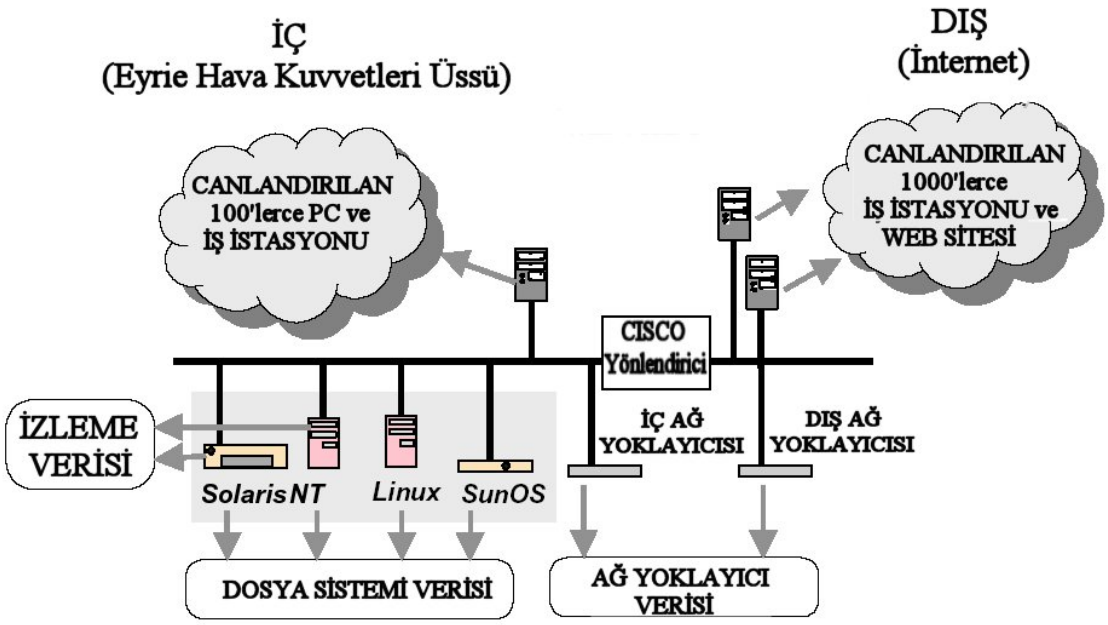
sistemden geçer. Bu durumda zaman serilerini kullanan yaklaşımlar ile analiz yapılabilir ve sadece trafik volumunde değişikliğe yol açan saldırılar tespit edilebilir.

Diğer yöntem ise IP flow monitor'lerinden veya ağ dinleyicisi (sniffer) programlarından veri toplanmasıdır. IP Flow monitorleri genelde pahalıdır. Sadece paket başlıklarına bakıldığından dolayı sınırlı bir bilgi içerir. Yoklayıcı programlar ise en ayrıntılı bilgiyi içerirler ve dolayısıyla çok fazla disk alanına ihtiyaç duyarlar. Ayrıca çok yoğun trafiklerde paketlerin hepsini yakalayamayabilirler.

Simule edilen bir ağdan ağ ve en sık kullanılan veri seti DARPA 1998 ve 1999 veri setleridir. Bu veri setleriyle ilgili bilinen problemler olmasına rağmen kontrollü bir çerçevede üretilip, ortak kullanıma açık olan yegane veri setleridir. Daha yeni tarihli çalışmalara geçmeden önce bu çalışmaların kullandığı DARPA veri setlerinin özelliklerinden bahsetmek yerinde olacaktır. Bundan sonraki bölümde DARPA 1998 ve 1999 saldırı tespit değerlendirme veri seti, bu veri setine getirilen eleştiriler ile yine bu verilerden türetilen KDD Cup'99 veri setlerini tanıtacağız.

5.1 DARPA Veri Seti Detayları:

DARPA'nın sponsorluğunda MIT Lincoln Laboratuarlarında STS'ler için bir karşılaştırma ortamı sunan IDEVAL veri setlerini oluşturulmuştur [8]. Özellikle saldırı tespit sistemlerinin sınanması amacıyla saldırıların hedefi olacak bir ağ ve saldırıları gerçekleştiren başka bir ağ dizayn edilmiştir. İç ağda Amerikan Hava Kuvvetlerindeki bir yerel ağın simülasyonu gerçeklenmiştir. 1999 değerlendirmesinde kullanılan ağ aşağıdaki gibidir.



Şekil 7 DARPA 99 Ağ yapısı[2]

Hava kuvvetlerini temsil eden iç ağ 172.16.0.0/16 IP adres uzayına sahiptir. Bu ağ içerisinde dört "kurban" makine bulunmaktadır. Bunların üzerinde SunOS, Solaris, Linux, ve Windows NT koşmaktadır. (1998 veri setlerinde sadece UNIX makinalar kullanılmıştır.) Şekilde görülen trafik oluşturucular yüzlerce sunucuyu ve çeşitli uygulamaları çalıştıran İnternet kullanıcılarını simüle etmektedir. Protokollerin (HTTP, SMTP, telnet,...) karışımı, trafik yoğunluğunun saatlik değişimleri, 1998'de gerçek Hava Kuvvetleri ağından toplanan trafiğe benzer olacak şekilde tasarlanmıştır. Ağ üzerinden 2 noktadan veri toplanmıştır: dört

kurban makine ile yönlendirici arasındaki iç ağ dinleyicisi ve yönlendirici ile İnternet arasındaki dış ağ dinleyicisi üzerinden [2].

1999 değerlendirmesi iki aşamadan oluşmaktadır. Veri setleri toplam 5 haftalık bir süreyi kapsamaktadır, ilk 3 haftası eğitim (training) verileri, son 2 haftası da test verileridir. 1. hafta ve 3. hafta saldırı içermemektedir. 2. hafta bazı saldırılar içermekte ve bu saldırılar ayrıca bir dosyada işaretlenmiştir. Bu kısım değerlendirmeye alınacak sistemlerin kendi başarımlarını test etmeleri içindir. Geriye kalan 4. ve 5. haftalarda ise 58 değişik saldırı tipinde 200 adet saldırı mevcuttur. Bu saldırı yazılımları internet ortamından ve hacker sitelerinden toplanmış saldırılardır. (1998 veri setinde çok fazla sayıda değişik saldırı bulunmamaktadır).

1999 değerlendirmesine 8 grup 18 IDS ile katılmış ve katılan gruplara:

1) Ağ dinleyicisindan elde dilen dosyalar
    a. İç ağ dinleyicisindan alınan tcpdump dosyası
    b. Dış ağ dinleyicisidan alınan tcpdump dosyası
2) Solaris Temel Güvenlik Modülü (Basic Security Module -BSM) audit dosyaları
3) Windows NT olay log dosyaları
4) Kurban makinalardan her gece alınan güvenlik kayıtları.

sağlanmıştır. Bu dosyalar şu anda İnternet sitesinden de indirilebilir durumdadır [19]. Değerlendirmenin ilk aşamasında katılımcılara üç haftalık veri dağıtılmıştır. İkinci aşamada, katılımcılara sistemlerini test edebilmeleri için iki haftalık test verisi dağıtılmıştır. Bunlar tüm setin dördüncü ve beşinci haftalarını oluşturur. Tüm günler için yukarda belirtilen 5 tip dosya sağlanmıştır. Tahmin edileceği üzere exper sistem tabanlı çalışmalar 2, 3 ve 4'ü kullanırken, anormallik tespiti yapan sistemler 1 numaralı verileri kullanmayı tercih etmişleridir. Veriler 5 hafta için 5'er gunluk olmak üzere toplam 25 gunluk trafik içerir. Verinin toplam boyutu 7,5 GB civarındadır. Hergun sabah 08:00 ve ertesi gün sabah 06:00 arasında 22 saat veri toplanması sonucu oluşturulmuş verilerdir. Gün başına ortalama 41MB veri duşmektedir [8]. Bunlardan protokol istatistikleri TPC (384MB), UDP (26MB) ve ICMP (98MB) olarak en baskın protokoller olmuştur. Aslında bir gunde oluşan trafik internet ortamına kıyasla düşük seviyededir fakat dosya boyutları ve daha sonra bu dosyaların internet ortamından araştırmacılara aktarılacak olması dolayısıyla bu şekilde tutulmuştur. Giriş kısmında belirttiğimiz 4 tip saldırı, Hizmet Engellme (DoS), Bilgi tarama (probe), yönetici hesabı ile yerel oturum açma (R2L), kullanıcı hesabının yönetici hesabına yükseltilmesi (U2R) saldırıları çeşitli saldırı sitelerinden toplanan kodlarla gerçeklenmiş ve sisteme aralıklarla enjekte edilmiştir.

Burada saldırı tespitinde dikkat edilmesi gereken nokta 5 dosyadan hangisinin hangi tip saldırılar için kullanılacağıdır. Sniffer verisi ağ saldırıları yani probe ve DoS için kullanılabilecekken, BSM kayıtları ile tüm saldırılar tesbit edilebilir. Biz çalışmamızda ağ saldırılarını hedef alacağız dolayısıyla ağ dinleyicisi verilerini kullanacağız.

DARPA veri setleri bazı özellikleri dolayısıyla bir sonraki bölümde ayrıntılı olarak inceleyeceğimiz eleştiriler almıştır fakat yine de var olan tek açık veri setidir. Bundan başkaca KDD-99 Cup Veri Setleri [9] olarak da bilinen veri seti mevcuttur. Aslında bu veri seti de DARPA 1998 veri setinden bazı özelliklerin ayıklanmış olduğu ve 5. Uluslararası Knowledge Discovery and Data Mining konferansı çerçevesinde düzenlenen bir yarışmada kullanıma sunulan veri setleridir. Sonuçta yine DARPA verileri ile aynı özelliklere sahiptir.

5.2. DARPA Veri Setine Getirilen Eleştiriler

DARPA veri setleri ile ilgili problemler olduğu bilinmektedir. DARPA verilerini ve dökümantasyonunu derinlemesine inceleyen bir çalışma [10] McHugh tarafından yapılmıştır. 2000 yılında yapılan bu çalışma daha çok 1998 veri setlerinin kritiğini yapmaktadır. Fakat benzer problemlerin 1999 setlerinde de sürdüğü bilinmektedir.

Getirilen eleştirilerden ilki arka plan trafiği denebilecek gerçek bir ağın çalışmasına ilişkin verinin alındığı Hava Kuvvetleri trafiği ile benzetim sonucu ortaya çıkan verinin hangi özelliklerinin benzediğinin Lincoln araştırma ekibi tarafından belirtilmediğidir. Aslında geliştirilecek STS'lerin DARPA veri setlerini temel alması ve buradan alacakları sonuca göre hareket etmeleri ve kendilerini kalibre etmeleri DARPA'nın temel amacı olarak düşünülürse bu verilerin sağlıklı olması yani gerçeğe yakın olması son derece önemlidir.

Bundan başka arka plan trafiği, internette sıkça gözlenen paket fırtınalarını içermemektedir. Yanlış protokol gerçeklemeleri sonucu ortaya çıkan bu davranış anormallik içerir. Saldırı olmayan veri setlerinde de bu paket fırtınaları yer alabilirdi. Ayrıca çok daha önemli bir parametre olarak iletişim yoğunluğunun ne olduğu DARPA setlerinde belirtilmemiştir. McHugh çalışmasında veri hızını incelediğinde bazı gunler için 10Kbps bazı günler için 50Kbps mertebesinde olduğu yani çok farklılık gösterdiği ve gerçek ağlarla kıyaslandığında bunun en az 10- 100 katı bir trafikle karşılaşıldığını söylemektedir.

DARPA setlerinde başka bir yanılgı ise saldırıların nasıl enjekte edildiğidir. Burada da gerçekçi bir yaklaşımdan ziyade rastgele, günde 3-4 saldırı olacak şekilde yerleştirildiği söylenmektedir. Ayrıca en yaygın saldırı probe olmasına rağmen saldırı karışımı aşağıda verildiği şekildedir.

Tablo 2. DARPA veri setindeki saldırların sınıflandırılması

U2R	114
R2L	34
DoS	99
Probe	64

McHugh'ün çalışmasından başka DARPA veri setlerine ait özbenzeşimliliği inceleyen bir başka çalışma 2003 tarihinde yapılmıştır [22] Burada da Hurst parametresinin günün değişik saatlerinde gündüz (H=0.7), gece (H=0.5) farklı değerler aldığı yani özbenzeşimli yapının herzaman korunmadığı söylenmiştir.

Ayrıca bizim yaptığımız çalışmada ihtiyaç duyduğumuz propagasyon gecikmesi ya da en azından TCP bağlantılarına ait ortalama RTT değerleri verilemesi gerekirdi. Zaman serisi oluşturulurken örnekelme frekansının seçilmesi için bu bilgilerin elde edilmesi önemlidir. Aksi takdirde deneme yanılma yoluyla çalışılmak zorundadır.

Tasarlanan STS'ler kendi başarımlarını bu veri setleri ile değerlendireceği için veri setinden kaynaklanan bir hata yanlış yollara sapılmasına yol açacaktır. Dolayısyla bu hatalara dikkat çeken çalışmalar önemlidir.

6. İSTATİSTİKSEL ANORMALLİK TESPİTİ KULLANAN AĞ TEMELLİ STS'LERDE YENİ YAKLAŞIMLAR

Anormallik tespitinde en populer yaklaşımlar zaman serilerini kullanmak veya veri setinden çıkartılan özellikler (feature) ile uzaklık hesabı yapmaktır.

Zaman serisi kullanılacaksa sinyal işleme teknikleri de rahatça kullanılabilme olanağı bulurlar. Verinin herhangi bir zaman serisi olarak elde edilebilmesi uygulama alanını genişletir. Belirli aralıklarla olaylar (gelen paket sayısı, gelen bağlantı sayısı, vs.) zamanda dizilirse, buradan en basit istatistiksel testler (ortalama ve diğer momentler) başlayarak, spektrum kestirimi, özdeğer analizi, dalgacık analizi gibi birçok karmaşık sinyal işleme tekniğine de yol açılmış olur. Özellikle son zamanlarda sinyal işleme ile STS'lerin birleştirilmesi çok sayıda çalışmaya konu olmuştur [11,12]. Bu çalışmalarda genellike MIB verisi kullanılır.

Eğer zaman serileri değil de özellikler kullanılıyorsa bu durumda ayrıksı tespitine (outlier detection) olanak veren örüntü tanımadan bildiğimiz kümeleme algoritmaları kullanım olanağı bulur. Bu bölümde bu iki temel yaklaşımı kullanan yakın zamanlı çalışmaları inceleyip kendi çalışmamızda kullanılabilecek özellikleri gözden geçireceğiz.

İnceleyeceğimiz çalışmalar 3 ayrı tipte veri seti kullanmaktadır ve biz de buna göre sıralayacağız. Öncelikle DARPA veri setlerini kullanan çalışmalara daha sonra KDD Cup verisini kullanan çalışmaya ve son olarak da yönlendirici MIB'lerini kullanmak suretiyle kendi yönetimlerindeki ağlardan veri toplayan ve analiz eden grupların çalışmalarına yer vereceğiz.

DARPA 1998 veri setini kullanarak yapılan bir çalışma [13]dır. Bu çalışma kaynak varış adresleri arasında yapılan bağlantı sayısını kullanarak saldırı tespiti yapmayı öngörür. Bağlantı sayıları bir zaman sersine dönüştürülüp üzerinde basit istatistik testler uygulanır. Bu çalışmada verilere değişik ölçeklerde bakılarak trafiğin gunduz, akşam ve gece rejimlerinde farklı ölçeklerde farklı saldırıların tespitinde kolaylık sağladığı ortaya çıkartılmıştır. Örneğin gündüz saatlerinde 100sn'lik örnekleme frekansı kullanılırsa ve gece saatlerinde 10sn'lik örnekleme frakansı kullanılırsa saldırı yakalamada daha başarılı olunduğu gözlemlenmiştir. Bu rejimlerin "normal" davranışlarının farklı olacağı beklenen bir özelliktir. Uygun ölçeklerde gelen bağlantı sayıları belirli bir dağılıma oturtulmuş ve uygun olmayan veriler saldırı olarak işaretlenmiştir. Dağılım kullanılarak yapılan çalışmalar her ne kadar sağlıklı olmasa da DARPA veri setine zaman serisi olarak bakması açısından önemli bir çalışmadır.

DARPA veri setlerini kullanarak ayrıksı eleman tespiti (outlier detection) yapan bir sistem [14]de tanıtılmıştır. Burada önce eldeki verilere uygun olacak şekilde bazı özellikler seçilmiştir. Bu özelliklere karşı düşen sayısal büyüklükler uyarınca yapılan kümelemede dışarda kalan örnekler ayrıksı olarak nitelenip saldırı olarak işaretlenmişlerdir. Örnegin bu çalışmada kullanılan bazı özellikler: "kaynak_varış_paket_sayısı", "varış_kaynak_paket_sayısı", "son 15saniyede bir kaynağa yapılan bağlantılar" gibi 23 değişik özellik seçilmiş ve bunların sayısal değerleri ile saldırı içermeyen veri setinden normal bir küme oluşturulmuştur. Daha sonra test verileri bu küme ile karşılaştırılıp genel davranıştan ayrılma varsa bunun tespitine çalışılmıştır. Burada en yakın komşu (nearest neighbor), Mahalanobis uzaklığı algoritması gibi örüntü tanımadan bilinen yaklaşımlar uygulanmıştır.

DARPA veri setlerini kullanan bir başka çalışma [4] dur. Burada yukarda anlatılan iki tekniğe ait presiplerin; özellik çıkartma (feature selection) ve sinyal işlemenin birleştirilmesi söz konusudur. Bunları uygulamak için 3 adımlı bir kombine işlem serisi gerçeklenmiştir. Önce ham DARPA veri setinden yani TCP/IP bilgilerinden 3 ayrı set elde ediyor. Bunları elde etmek için tek-aşamalı bağdaştırma (single-level association mining), çok-aşamalı bağdaştırma (multi level association mining) ve özellik çıkartma adlı 3 veri madenciliği algoritması koşturuluyor. Bu bağdaştırma kurallarını anlamak için bir örneğe bakalım. Bir TCP/IP bağlantısı kaydı şu şekilde tanımlanabilir.

K= {Bt, kaynak.ip, kaynak.port, varış.ip, varış.port, servis, bayrak}

Bt bağlatı başlangıç zamanını verir. Daha sonra sırasıyla kaynak IP adresi, kaynak kapı numarası, varış IP adresi, varış kapı numarası yer alır. Servis ftp, smtp gibi uygulama protokolerini içerir. Bayrak ise bir bağlantı için {açık, kapalı, albaştan} gibi durumları ifade eder. Örneğin aşağıdaki kayıt:

(kaynak.ip: 129.174.111.3, kaynak.kapı: 3167, varış.ip: 73.8.19.5, varış.kapı: 23, servis: telnet, zaman: 15:23:12)

sadece saat 15:23:12 de 129.174.111.3 ve 73.8.19.5 arasında telnet bağlantısı olduğunu söylerken, aşağıdaki bağlantı kuralı bunu daha anlaşılır bir halde ifade edebilir.

(kaynak.ip: 129.174.111.3 →varış.ip: 73.8.19.5, varış.kapı: 23, servis: telnet, süre: 15:20:49-15:29:03)

Buna benzer bağdaştırma kuralları ortaya çıkartılıp herbir kayda sayısal bir değer atanıp 3 ayrı matrise yazılır. Bağdaştırma kuralları Denning'in tablolarına benzetilebilir. Herbir kayda değer (skor) atanması konusunda alınacak kararlar da başlı başına bir çalışmadır. Şimdilik sayısal değerlerin oluşturulup matrislerle ifade edildiğini var sayalım. Bu 3 matrisin özilinti (otokerelasyon) matrisleri üzerinde temel bileşen analizi (principal component analysis - PCA) ya da başka adıyla factor analizi yapılıp yüksek boyuttaki veri daha az boyutta ifade ediliyor. Bunun anlamı örneğin 100x100'luk bir matrisin toplam 100 özdeğer ile belirtilmesi ve de bu değerlerin sadece belli başlı olanlarının (mesela ilk 10) kullanılmasıdır. Yani verinin kendisi yerine ondan daha az boyuttaki temel bileşenler ile analiz yapılıyor. Bu bileşenlerin davranışı saldırı içermeyen veri setleri için normal davranışı modellemek üzere çıkartılıyor. Daha sonra test veri için de aynı işlem yapılıp normal değerler arasındaki Mahalanobis uzaklığına bakılıyor. Eğer belli bir eşik değeri aşılmışsa eldeki veri ayrıksı olarak adlandırılıyor ve saldırı olarak damgalanıyor. Burada korelasyon matrisinin temel alınması bağlantı kayıtları arasında korelasyon varsa bunu açığa çıkartmaya yarar. Bu çalışma birçok tekniği bünyesinde barındıran bir çalışma olması açısından buraya alınmıştır. Karmaşık olması dolayısıyla gerçeklenmesi düşünülmemiştir.

Buraya kadar DARPA veri setini kullanan çalışmalara odaklandık. Buradan sonra KDD Cup verisi üzerinde yapılmış sinyal işleme tekniklerini içeren bir çalışmaya bakacağız. KDD'99 verileri kullanarak temel bileşen analizini uygulayan ve bu konuda yapılmış farklı çalışmalardan biri [15]tir. Temel bileşen tabanlı sınıflandırıcı yöntemini önerilmiştir. Temel bileşen analizinin güzel yanının kestirim ya da saldırı tespitinde kullanılacak verinin boyutunu azaltması olduğunu biliyoruz. Temel bileşenler (TB) verinin varyansını ve korelasyonunu yakalarlar. TB'ler büyükten küçüğe sırayla dizildiklerinde ilk TB'ler en yuksek varyansa

sahip olanlardır. Sondaki TB'lerde verideki otokorelasyon bilgisini içerirler. TB'lerin hesabı korelasyon matrisi R den yapılabilir. i. TB y_i şu şekilde elde edilir:

$$y_i = e_i z \text{, } e_i \text{ özvektör ve z:}$$

$$z = (z_1, z_2, ..., z_p)\text{, } z_k = \frac{(x_k - \bar{x}_k)}{\sqrt{s_{kk}}}, k = 1,2,..., p$$

s_{kk} standart sapma olmak üzere. Buradan temel bileşen ve ona ilişkin özdeğerler birlikte bir sınıflandırma ölçütü oluştururlar:

$$\mu = \sum_{i=1}^{p} \frac{y_i^2}{\lambda_i}$$

λ_i'ler özdeğerler olmak üzere, μ değerine göre bir sınıflandırma yapılacaktır. Fakat burada hangi özdeğerler kullanılacak bunu belirlemek önemli bir problemdir. İlk özdeğerler yuksek varysansı temsil ederler. Eğer eğitim veri setinden elde edilen varyans değerini aşan değerler test verisinde varsa bunlar ayrıksı sayılırlar. Fakat küçük TB'lerde korelasyon yapısına duyarlıdırlar. Saldırı sırasında farklı bir korelasyon yapısı söz konusu ise bunlar işe yarayabilir. Dolayısıyla bu çalışmada ilk TB'lerin bir kısmı

$$\sum_{i=1}^{q} \frac{y_i^2}{\lambda_i}$$

ve son TB'lerin bir kısmının kullanılması önerilmiştir:

$$\sum_{i=p-r+1}^{p} \frac{y_i^2}{\lambda i}$$

Burada $\sum_{i=1}^{q} \frac{y_i^2}{\lambda_i} \leq c_1$ ve $\sum_{i=p-r+1}^{p} \frac{y_i^2}{\lambda i} \leq c_2$ için c_1 ve c_2 değerleri eşik değerleridir. Yanlış alarm değerini belirli bir seviyede tutacak şekilde seçilebilirler. Dolayısıyla belirli bir yanlış alarm seviyesi için belirlenen eşik ile özdeğerler ve TB'lerden oluşan sayısal değer karşılaştırılarak bir sınıflandırıcı gerçeklenmiş olur. [15].

Hazır veri setlerini kullanmak yerine kampüs ağından topladığı verilere sinyal işleme yöntemlerini uygulayan bir saldırı tespiti çalışması Wisconsin- Madison Üniversitesinden Barford ve arkadaşlarının yaptığı çalışmalardır [7]. Kampüs ağının dış ağ bağlantısını sağlayan yönlendiricisinden SNMP mesajlarını ve IP flow monitörlerinden alınan bilgileri işlemişlerdir. Buraya kadar anlatılan yöntemlerden farklı olarak verilere dalgacık yöntemi uygulanmıştır. Zaman serisi olarak elde edilen veriler üç ayrı frekans bileşenine: düşük, orta ve yüksek olmak üzere ayrılmışlardır. Burada dalgacık yöntemi kullanılarak veri ayrıştırılmıştır. Verinin kendisinden oluşturulan zaman serisi ve bu filtrelemenin sonundaki, 3 değişik ölçekteki zaman serisine kayan ortalama hesabı (moving average) uygulanıp belirlenen bir eşik değerinden sapmalara saldırı etiketi konmuştur. Bu zaman-frekans incelemesi sonucunda saldırı uzunluğu konusunda da bilgi sahibi olunabilecek bir teknik oluşturulmuş olup, saldırılar uzun-vadeli (long-lived) ve kısa vadeli (short-lived) olarak sınıflandırılabilmiştir.

Kendi topladığı verileri kullanan başka bir çalışma [12] dır. Bu çalışmanın dayandığı temel nokta "anormal bir davranış öncesi keskin–ani bir degisiklik (abrubt change) oluyor olması" ve bu degisiklik noktası yakalanırsa, anormallik olmadan tespit yapılabileceğinin düşünülmesidir. Burada da bir zaman serisi elde edilip örnekleme zamanından büyük olacak aralıklardaki ani değişkliklerin tespiti söz konusudur. Zaman serilerini oluşturmada MIB değişkenlerinden, ip katmanı sayaçları ve yönelendirici arayüz sayaçları ndan, ipIR (ip in received), ipIDe (ip in delivred) ve ipOR (ip out requestst) kullanılıyor. Ani davranışı kestirme başlı başına bir çalışma alanı olduğundan burada detaylarına girmeyeceğiz fakat bu örnek bize sinyal işleme konularının STS'lerde uygulanmasının açacağı yeni ufuklar konusunda ip uçları vermektedir.

7. DARPA VERİ SETLERİNİ KULLANARAK ZAMAN-SERİSİ OLUŞTURULMASI VE BUNLARIN ANALİZİ: Bir Gerçekleme Çalışması

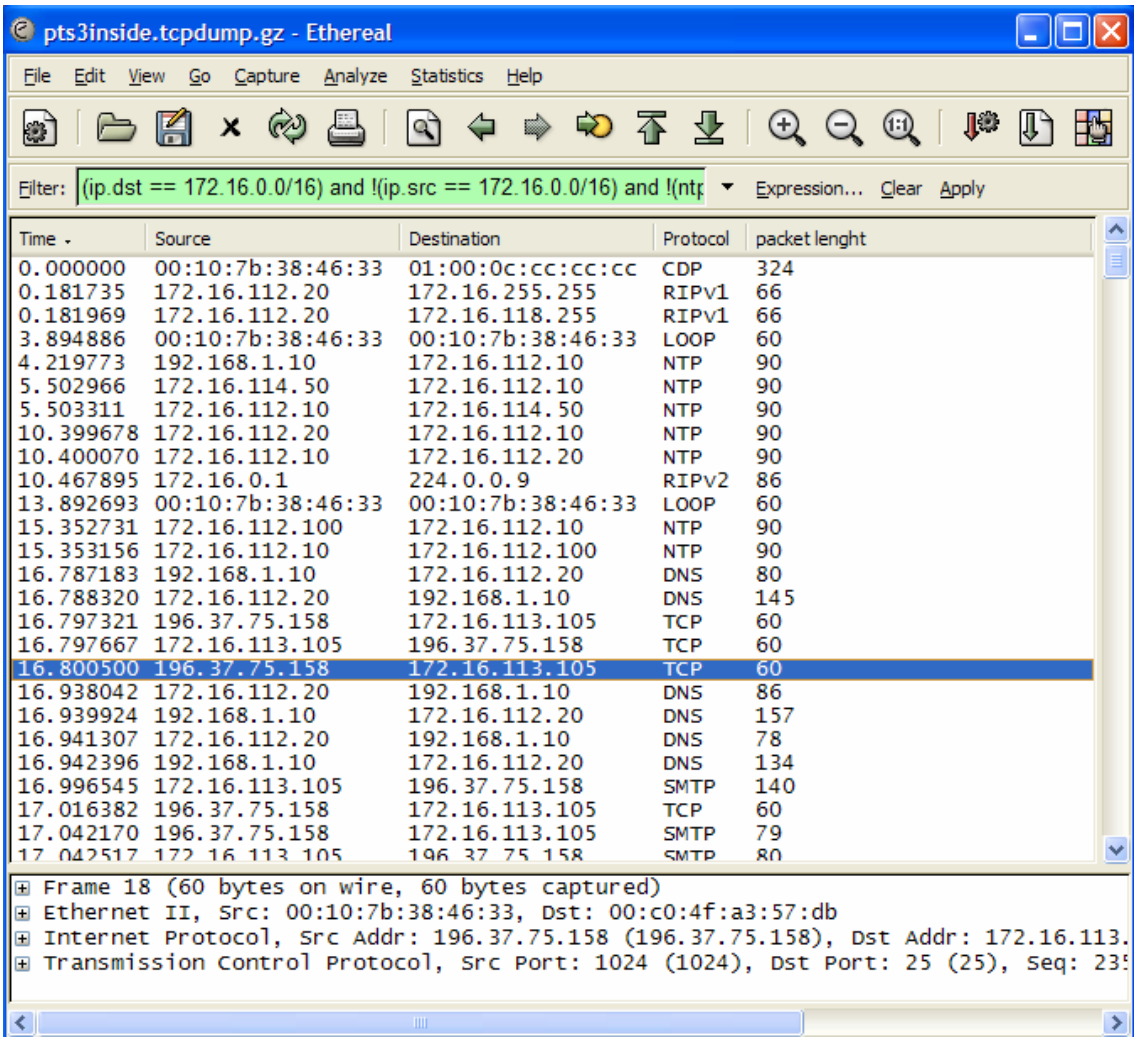
Yapmayı hedeflediğimiz çalışma DARPA veri setlerini kullanarak MIB verilerine benzer IP trafik yoğunluğunu gösteren bir zaman serisi elde etmek ve buradan DoS ve probe saldırılarını tespit etmektir. Aslında yaptığımız çalışmanın hedefi eldeki DARPA 1999 veri setine sinyal işleme tekniklerini uygulamaktır. Bunun için iç ağ dinleyicisi dosyalarından, 5 haftalık verilerden saldırı içermeyen 3. hafta ve saldırı içeren test haftası 5. haftadan ilk güne (pazartesi) ait dosyalar alındı. Burada düşünülen amaç saldırı içermeyen günlerin gürültü, saldırıların sinyal olarak tanımlanıp, gürültüde sinyal tespiti yapılması işlemini gerçeklemektir. Bu bildik bir kestirim (detection-estimation) konusudur.

Ağ dinleyicisi verisi ile zaman serisi oluşturmak ilk elden yapılabilecek kolay bir işlem değildir. Dinleyici veri seti tcpdump veri formatında olup, ağa giren ve çıkan tüm paketleri tüm ayrıntılarıyla tutmaktadır. Öncelikle bu çok fazla disk yeri tutan bir durumdur. Dinleyici verileri ağ üzerinde, çift yönlü gelip-giden tüm paketleri içerdiğinden, bir bağlantının iki yönlü paketleri, yönlendirme paketleri ve diğer tüm gereksiz trafik filtrelenmelidir. Her 2 gün için tcpdump veri seti ethereal aracılığıyla açılmış ve filtrelenmiştir. İç ağdaki yönlendiricilere gelen mesajların makinalara yapılan saldırı ile ilgisi olmadığı düşünülerek ağa giren yönlendirme mesajları rip, loop, cdp, vsvs gibi mesajlar filtrelenerek dışarda bırakılmıştır. Ağ dinleyicisi iki yönlü paketleri yakaladığı için kurban makinalara gelen paketler için varış ip adresi 172.16. xxxx olan paketler olarak süzülmelidir.

Bunun için ilk olarak ele alınan 3. hafta pazartesi (saldırı içermeyen veriler) tcpdump.inside dosyası ethereal de açılmalı ve uygun filtre uygulanmalıdır. Şekil 8'de bugüne ait ilk paketler görülmektedir. Uygulanan filtre:

```
(ip.dst == 172.16.0.0/16) and !(ip.src == 172.16.0.0/16) and !(ntp) and !(rip) and !(loop) and !(arp) and !(nbns)
```

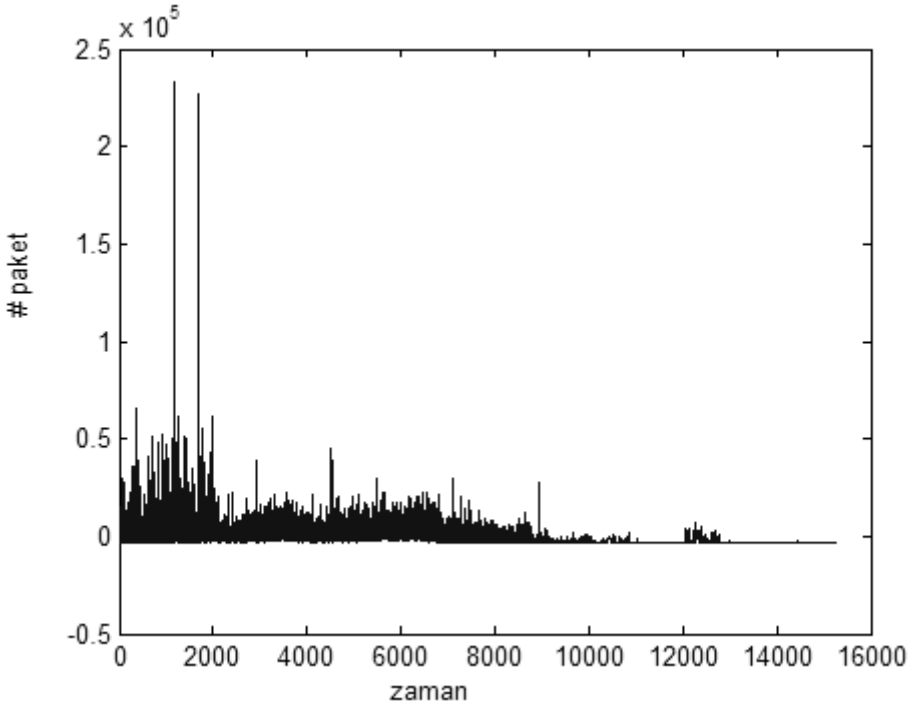
dir.



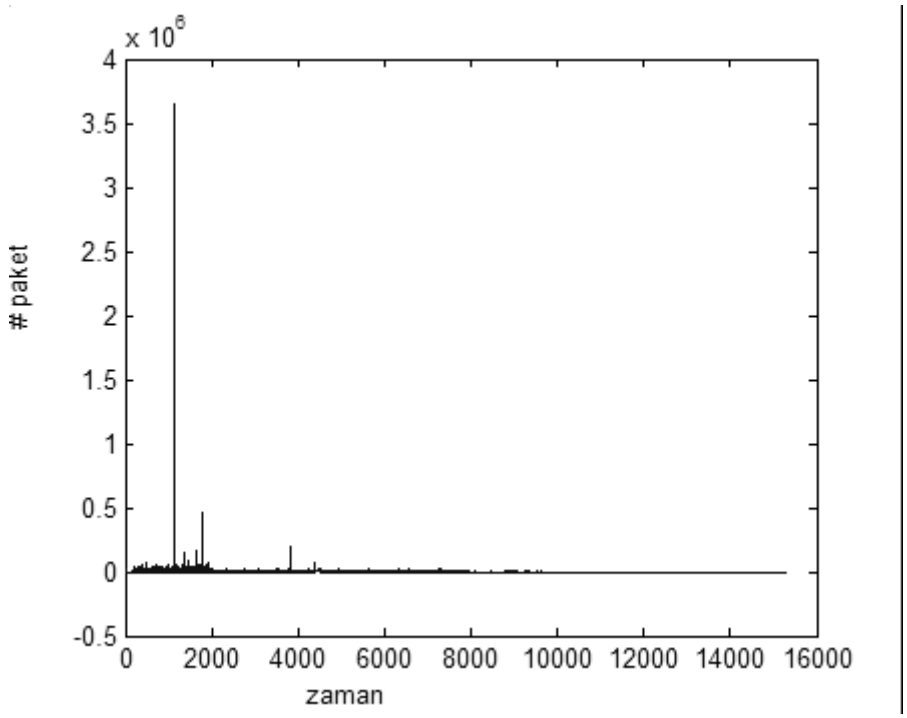
Şekil 8. Ethereal ekran görüntüsü

Nihayi amaç bir zaman-serisi oluşturmak olduğundan ayıklanan veriler zaman damgası ve paket boyu kalacak şekilde 2 sütuna düşürülecektir. Buradan belli zaman aralıkları için kovalarda biriktirilerek örnekleme yapılacaktır. Burdan sonra Ethereal'daki bilgi, programın export özelliği kullanılarak text dosyasına aktarılmıştır. Artık bu dosya herhangi bir programlama dilinde incelenebilir. Biz çalışmamızda C++ kullandık. Yazılan kod ile sadece zaman ve paket boyu bilgisi alınıp 2 sütunlu başka bir dosya oluşturulmuştur. Aslında buradan sonra veriler sinyal işleme kodlarının da kolaylıkla yazılacağı MATLAB ortamına aktarılabilirdi fakat MATLAB'deki matris ve dizi boyları dosyadaki verilerden kısa olduğundan dolayı bir adım daha önişleme yapılmalıdır.

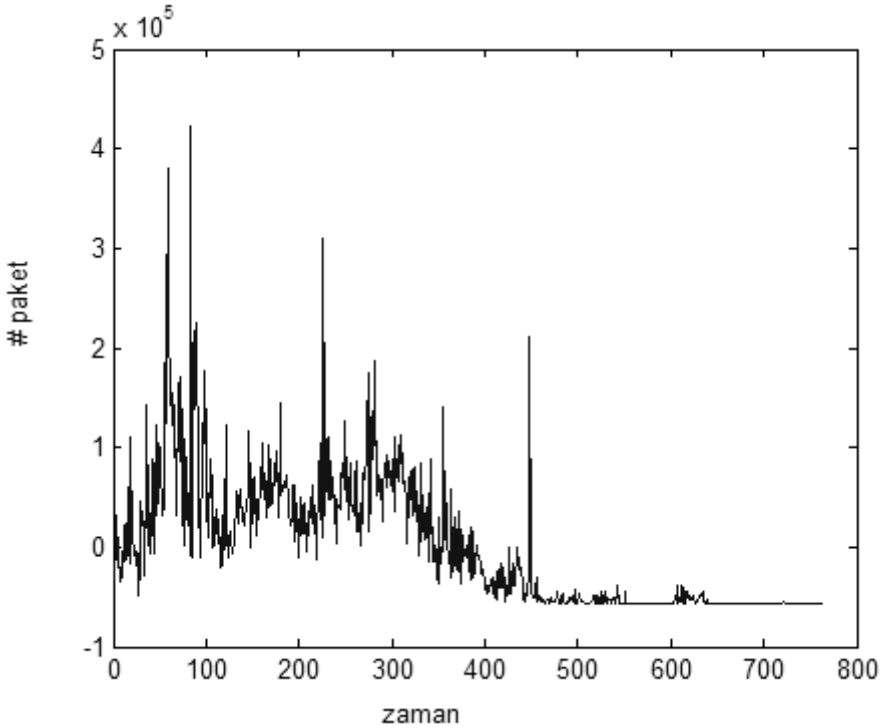
Önişlemenin ikinci adımında yine bir C++ kodu ile her 5 saniyede bir gelen paketler kovalarda toplanmıştır. Bu adımdan sonra MATLAB'e aktarılacak zaman serisi oluşmuş olur. Şekil 9 ve 10'da 3. Hafta (saldırı içermeyen) ve 5. hafta (saldırı içeren) Pazartesi günlerine ait ağda akan byte sayısı görülmektedir. Şekil 11 ve Şekil 12'te yine aynı günler için 100snlik kovalarda toplanan veriler gösterilmiştir.



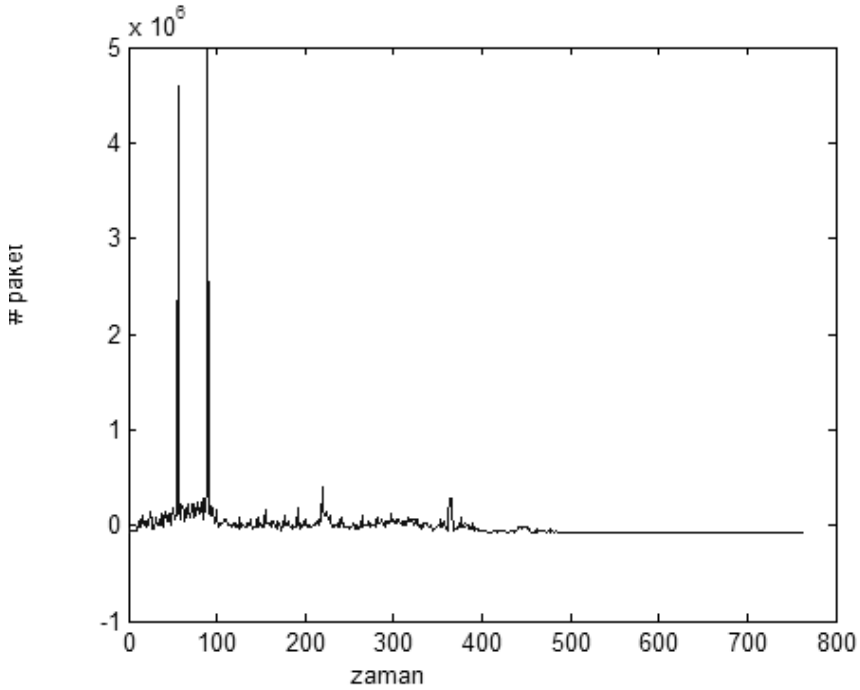
Şekil 9. 3. hafta pazartesi verileri (5sn lik kovalarda toplanmış)



Şekil 10. 5. hafta pazartesi verileri (5sn lik kovalarda toplanmış)



Şekil 11. 3. hafta pazartesi verileri (100sn lik kovalar)



Şekil 12. 5. hafta pazatesi verileri (100sn lik kovalar)

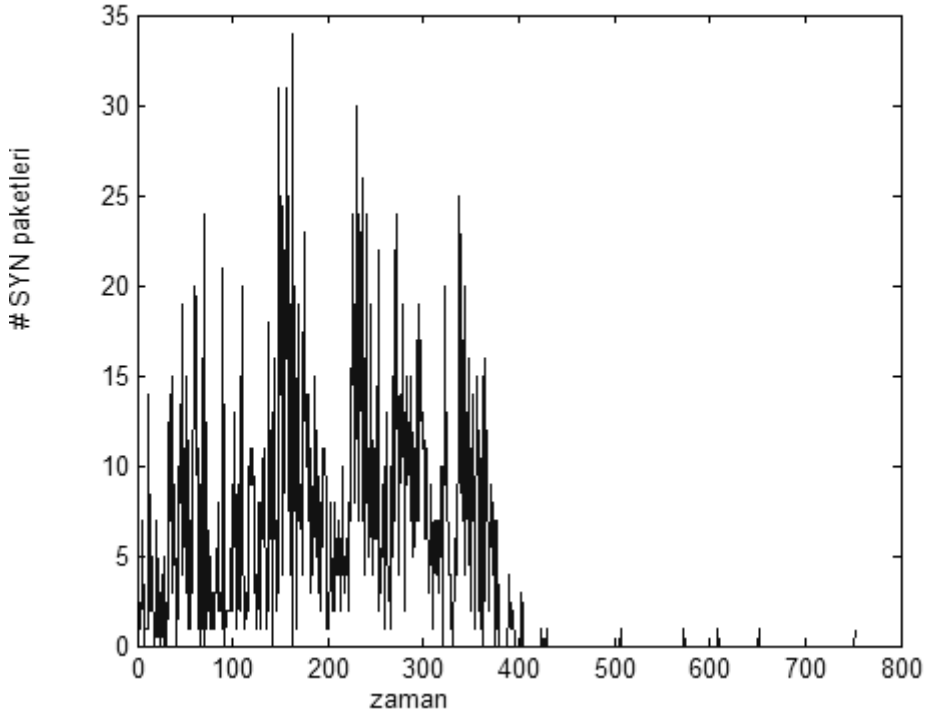
Şekillerden görülen en belirgin özellik normal veri setindeki 2 adet tepenin saldırı setinde daha yüksek değerlerde oluşmasıdır. Aslında burada normal veri setinde bu şekilde tepelerin görülmesi gerçek trafikle kıyaslandığında bir problem olduğunu göstermektedir. Bunun DARPA setiyle ilgili daha önce bahsedilen problemlerden kaynaklandığını düşünmekteyiz. Çünkü Hava kuvvetleri ortamı simule edilirken istatistiksel özellikler göz önüne alınmamış gibi gözükmektedir.

Bütün bu işlemler sonucunda sadırı tespitinde, tcpdump verisinden zaman ve paket boyu bilgisi yani trafik hacmi verisi çıkartıldığı için DARPA veri setinin içerdiği çok az özellik kullanılmış durumdadır. Bu durumu engellemek için farklı bir yaklaşım düşünülmüştür. Bu yaklaşım bağlantı isteklerinden zaman serisi oluşurmak olabilir. Burada DoS saldırıları hedef alınabilir. Bir zaman diliminde bir kaynağa gelen bağlantı sayılarındaki ani değişiklik saldırı habercisi olabilir. Yalnız burada saldırı hakkında ön bilgi olamayacağı için her makinaya ait ayrı ayrı zaman serileri oluşturmak gerekir. DARPA ağında dört adet kurban makina bulunduğundan bunu yapmak çok zor değildir fakat çıkan sonuçlar gerçekçi olmayacaktır. Dolayısıyla biz makine tabanlı olmayan ama yeni bağlantı sayılarını gösteren TCP paketlerinin SYN bayrağına bakacağız. Ağ dinleyicisi paket başlık detaylarını da gösterdiğinden dolayı SYN=set durumuna bakılabilir. Burada gelen SYN paketleri sayılacaktır. Örneğin son 10 saniyede gelen SYN sayılarındaki artış bir belirteç olabilir. Fakat SYN mesajına SYN-ACK gelip gelmediğinin kontrolünün yapılması da gereklidir. Bu bağlantı isteğinin istekte bulunan taraf tarafından onaylandığını gösterir [21]. Bunlar da sayıldığı takdir de normal gelişen bağlantılara ait SYN'ler boş yere sayılmamış olur. Biz çalışmamız da bu kısma dikkat etmeden sadece SYN sayılarındaki artışa eşik değeri uygulayarak bir saldırı tespiti yapmaya çalışacağız.

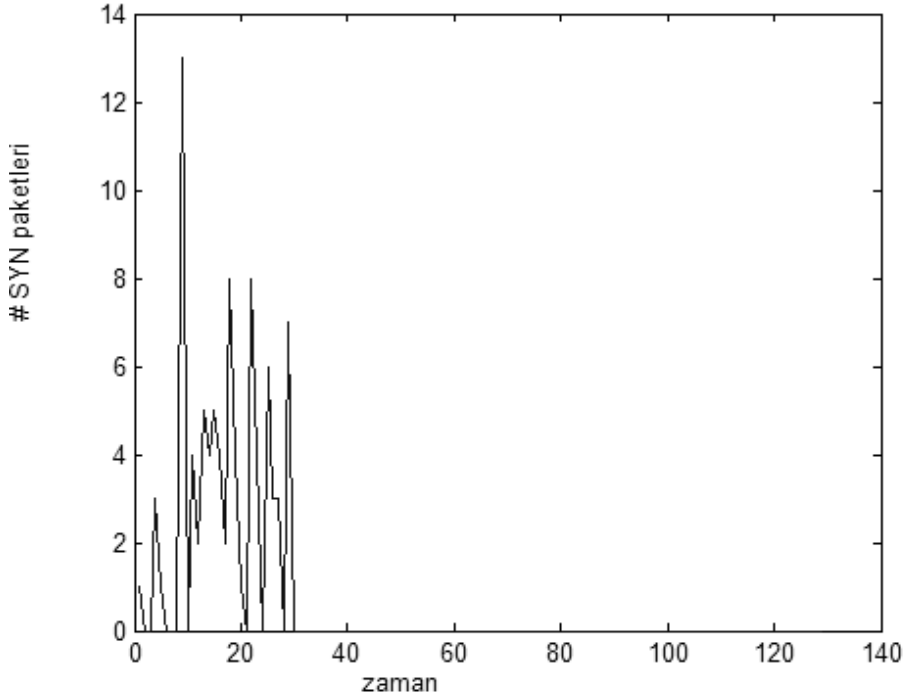
Yine Ethereal ile açtığımız dosyaya sadece TCP bağlantılarını filtreleyecek bir filtre koymak gereklidir:

```
(ip.dst == 172.16.0.0/16) and !(ip.src == 172.16.0.0/16) and tcp
```

Buradan sonra yine bir C++ kodu ile zaman damgaları ve SYN anahtar kelimesini içeren satırlar ayıklanacaktır. Daha sonra 10sn'de kaç adet SYN içeren satır geçtiği sayılıp toplanmıştır. Buradan oluşan sonuç MATLAB'e aktarılacak zaman serisini oluşturulmuştur. Burada dikkat edilmesi gereken nokta seçilen örnekleme süresinin saldırıyı gösterebilecek kadar SYN mesajının toplanmasına yetmesi gerekliliğidir. DARPA setinde propagasyon gecikmesi ya da ortalama RTT süreleri verilmediğinden burada bir hesap yapılamamakta ve denyesel olarak sonuçlar üretilmektedir. Bu sürelerin gerçekçi olup olmadıkları da bilinemektedir ve bizim tarafımızdan veri setine getirilen bir eleştiridir. Şekil 13 ve 14'te bu analizin sonucunda çıkan zaman serileri verilmiştir.



Şekil 13. 3. hafta Pazartesi için SYN paket sayısı (100sn)



Şekil 14. 5. hafta Pazartesi için SYN paket sayısı (100sn)

Şekillerden de görüldüğü gibi SYN paketleri de yeterli bilgi içermemektedir. Ayrıca kendi içersinde de problemleri vardır. Normal veri setinde daha fazla bağlantı bulunması analizde bir hata yapılmış olduğunu düşündürtmektedir. Bunu görmek için ilk aşamada SYN-ACK iklisinin kaç bağlantıda gönderilmediği araştırılmalıdır.

8. SONUÇ

Ağ Güvenliği dersi kapsamında yapılan bu projde ağ üzerinden gerçekleşen saldırılar hakkında genel bir bilgi birikimi oluşması amaçlanmıştır. Literatür taramasıyla oluşturulan bilgi birikimi, kendi çabamızla tasarladığımız basit sistem üzerinde kullanılmıştır. Bu sayede STS çalışmalarının temel aşamaları hakkında bilgi sahibi olunmuştur.

REFERANSLAR

[1] V. Yegneswaran, P. Barford, J. Ullrich, "Internet Intrusions: Global Characteristics and Prevelance", ACM SIGMETRICS, 2003.

[2] M. A. Aydın, "Bilgisayar Ağlarında Saldırı Tespiti için İstatistiksel Yöntem Kullanılması", İTÜ Yüksek Lisans Tezi, 2005.

[3] R.A. Kemmerer, G. Vigna, "Intrusion Detectsion: A Brief History and Overview", IEEE Computer Special Issue on Security and Privacy, 2002.

[4] N. Wu, J. Zhang, "Factor Analysis Based Anomaly Detection", Proc. IEEE Workshop on Information Assurance, 2003.

[5] S. Axelsson, "Intrusion Detection Systems: A Survey and Taxonomy", Technical Report Dept. of Computer Eng., Chalmers University, March, 2000.

[6] D. Denning, "An Intrusion-Detection Model", IEEE Trans on Software Enginering, vol. 13, no. 2, 1987.

[7] P. Barford, J. Kline, D. Plonka, A. Ron, "A Signal Analysis of Network Traffic Anomalies", Proc. of ACM/SIGCOMM, 2002.

[8] R. Lippmann, J. W. Haines, D. J. Fried, J. Korba, K. Das, "Analysis and Results of the 1999 DARPA Off-Line Intrusion Detection Evaluation", MIT Lincoln Laboratory Technical Report TR-1062, 2001.

[9] KDD Cup veri setleri, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[10] J. McHugh, "Testing Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln Laboratory", ACM Trans on Information and System Security, vol. 3, no. 4, 2000.

[11] A. Lakhina, M. Crovella, C. Diot, "Diagnosing Network-Wide Traffic Anomalies", Proc. of ACM SIGCOMM, 2004.

[12] M. Thottan, C. Ji, "Anomaly detection in IP Networks", IEEE Trans on Signal Processing, vol. 51, no. 8, 2003.

[13] J.B.D. Cabrera, B. Ravichandran, R. K. Mehra, "Statistical Traffic Modeling for Network Intrusion Detection", Proc. of International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, 2000.

[14] A. Lazarevic, L. Ersöz, V. Kumar, A. Özgür, J. Srivastava, "A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection", Proc. of SIAM Conf. Data Mining, 2003.

[15] M. L. Shyu, S. C. Chen, K. Sarinnapakorn,L. W. Chang, "A Novel Anomaly Detection Scheme Based on Principal Component Classifier", Proceedings of ICDM Foundation and New Direction of Data Mining workshop, pp 172-179, 2003.

[16] P. Mell, V. Hu, R. Lippmann, J. Haines, M. Zissman, "An Overview of Issues in testing Intrusion Detection Systems", NIST Technical Report, Haziran 2003.

[17] J. Allen, A. Christie, W. Fithen, J. McHugh, J. Pickel, "State of the Practice of Intrusion Detection Technologies", Tech. Rep. CMU, 2000.

[18] A. Hussain, J. Heidemann, C. Papadopoulos, "Distingushing between Singel and Multi-source Attacks Using Signal Processing", Computer Networks, vol. 46, 2004.

[19] DARPA IDEVALweb sitesi, "http://www.ll.mit.edu/IST/ideval/data".

[20] W. Stallings, “Cryptography and Network Security: Principles and Practice”, Prentice Hall, 1998.
[21] V. A. Siris, F. Ppapagalou, “Application of Anomaly Detection Algorithms for Detecting SYN Flooding Attacks”, Globecom, 2004.
[22] W. H. Allen, G. A. Marin, “On the Self-similairty of Synthetic Traffic for Evaluation of Intrusion Detection Systems”, Symp. On the Applications and the Internet, 2003.