ÖZELLİK İNDİRGEMESİ İLE ÖRÜNTÜ SINIFLANDIRMASI KULLANILARAK BİLGİSAYAR SALDIRILARININ TESPİTİ

Müge ÇEVİK[1] **Bülent ÖRENCİK**[2]

[1]Siemens San. Tic. A.Ş. Programlama ve Sistem Mühendisliği Bölümü
Yakacık Caddesi No:111 34870 Kartal, İstanbul
[2]Bilgisayar Mühendisliği Bölümü
Elektrik-Elektronik Fakültesi

İstanbul Teknik Üniversitesi, 80626, Maslak, İstanbul

[1]e-posta: muge.cevik@siemens.com [2] e-posta: orencik@cs.itu.edu.tr

Anahtar sözcükler: Saldırı Tespiti, Örüntü Sınıflandırması, Gözetimli Öğrenme

ABSTRACT

There are two types of intrusion detection systems: Behaviour based and knowledge based. Pattern classification can combine both of them and guides to find the optimum solution. In this paper, the KDD Cup 1999 intrusion detection data have been used to develop an intrusion detection system with pattern classification which is named CLIDS (Cluster based Intrusion Detection System). The proposed system uses FCBF algorithm developed by Lei Yu and Huan Liu to reduce the dimension of the patterns, then it creates an attack database which consists of clusters of attacks and the normal patterns and the radius of these clusters; in the test phase it tries to find if the test pattern anomaly is, or a known attack type is or normal is.

1. GİRİŞ

Bilgisayar saldırılarının gün geçtikçe artması yeni saldırı sistemleri araştırmayı teşvik etmektedir. Temel olarak iki tip saldırı tespit sistemi vardır: Davranış bazlı ve bilgi bazlı. Bilgi bazlı sistemler sadece önceden tanımlanmış saldırıları yakalayabilirler; yeni saldırılara karşı dayanaksızdırlar. Davranış bazlı saldırı tespit sistemleri ise normal davranışları öğrenirler ve bu davranıştan farklı olan davranışları anormal olarak tanımlarlar

Örüntü sınıflandırması hem bilgi bazlı hem davranış bazlı saldırı tespitini bir araya getirerek optimum sonuca ulaşmada yol gösterici olmaktadır. Örüntü sınıflandırmada iki tür yöntem vardır: Öğretimli sınıflandırma, öğretimsiz sınıflandırma. Öğretimli sınıflandırmada belli bir örüntü kümesiyle algoritma çalıştırılır ve algoritma bu kümede önceden belirlenmiş sınıfları ve sınıfların özelliklerini öğrenir. Test örüntüsü algoritmaya verildiğinde bu test verisinin hangi sınıftan olduğu belirlenir. Öğretimsiz sınıflandırmada ise örüntülerin hangi sınıftan oldukları önceden bilinmez. Örüntülerden birbirine yakın özellikte olanlar aynı sınıfta toplanır. Daha sonra bunlar etiketlenir[3].

Bu bildiride önerilen sistemde ACM Special Interest Group on Knowledge Discovery and Data Mining tarafından 1999 yılında yapılan veri madenciliği yarışmasında sunulan saldırı tespit verileri kullanılarak bir örüntü sınıflandırması ile saldırı tespit sistemi gerçeklenmeye çalışılmıştır. Bu saldırı tespit sistemi CLIDS (**Cl**uster based **I**ntrusion **D**etection **S**ystem – Küme tabanlı Saldırı Tespit Sistemi) olarak adlandırılmıştır. Bu sistemde öncelikle bilinen ataklarla eğitim verileri içinde sınıf karakteristikleri çıkarılmaktadır. Bunu yaparken de bilinen sınıflandırma algoritmaları doğrudan kullanılmamıştır. Eğitim verileri, FCBF (Fast Correlation Based Filter) [1] algoritmasıyla ayırt edici özellikleri bulunduktan sonra sınıflandırılmıştır. Bu sınıflandırmada saldırı tespitinde çok önemli rol oynayan sembolik veriler (protokol tipi, hizmet tipi, bayrak tipi vb.) öne çıkarılarak, FCBF tarafından seçilmiş sembolik verilerle etiketlenen sınıflar oluşturulmuştur. Bundan sonra CLIDS'in içindeki algoritma yardımıyla test örüntüleri, "normal – atak " verilerinin oluşturduğu sınıflara göre yapılan karşılaştırmalarla bunların hangi sınıflara yakın oldukları belirlenir. Eğer bir örüntü birden fazla sınıfa yakınsa bu sınıflardan hangisinin seçilmesi gerektiği irdelenir. Hiç bir sınıfa önceden belirlenmiş bir eşikten daha yakın olmayan örüntüler "anormal" olarak etiketlenir. CLIDS gerçek zamanlı çalışmamakla birlikte öğretimli örüntü sınıflandırmasının ve özellik seçiminin saldırı tespitinde kullanılabileceğini kanıtlamakta ve anormal durumları bulmada yeni bir bakış açısı getirmektedir.

2. SALDIRI VERİLERİ VE ÖZELLİK SEÇİMİ ALGORİTMASI

Saldırı verileri, öğretim verilerinde 22 adet (ağ bazlı ve konak bazlı) ve test verilerinde 40 adet (ağ bazlı ve konak bazlı) saldırı tipi içermektedir; ayrıca saldırı

tiplerinin yanı sıra bu verilere normal trafik kayıtları da eklenmiştir. Değerlendirmeye alınan veriler, normal ve saldırı trafiğinin yaratıldığı bir bilgisayar laboratuarında elde edilen tüm ağ paketleri ve konak bazlı kayıtları sembolize eden 41 özellikten oluşan vektörleri içermektedir. Bu vektörler bir metin dosyasında okunabilecek şekilde biçimlendirilmiştir. Her bir özellik sayısal (örneğin içerilen sekizli sayısı) veya sembolik (örneğin bir bağlantının kurulu olması yada olmaması durumu) değerler alabilir. Söz konusu özelliklerin bazıları şunlardır: Bağlantı süresi uzunluğu, protokol tipi (sembolik), varış hizmet tipi (sembolik), kaynak veri uzunluğu, varış veri uzunluğu, TCP/IP bağlantısı kurmanın normal ya da hatalı olma durumu (sembolik), varış ve konak adresinin aynı olma durumu (sembolik), hatalı TCP/IP paket parçası sayısı, acil paket sayısı, sistem dosyalarının yaratılma ve erişilme sayısı, hatalı "login" sayısı, başarılı "login" olup olmama durumu (sembolik), dosya yolunun hatalı verilme sayısı, kök kabuğun ele geçirilip geçirilmemesi (sembolik), kök kullanıcı olmaya girişilip girişilmemesi (sembolik), kök kullanıcı olma sayısı, dosya yaratma sayısı, kabuk komutu sayısı, giriş kontrol dosyalarında yapılan işlem sayısı, ftp oturumundaki komut sayısı, kullanıcının "hot" listeye ait olup olmaması (sembolik), kullanıcının "guest" listeye ait olup olmaması (sembolik), son 2 saniye içinde aynı konağa yapılan bağlantı sayısı, son 2 saniye içinde aynı konağa yapılan bağlantıların "SYN" hata yüzdesi, son 2 saniye içinde aynı konağa yapılan bağlantıların "REJ" hata yüzdesi, son 2 saniye içinde aynı konağa yapılan bağlantılarda aynı hizmet tipi olma yüzdesi, son 2 saniye içinde aynı konağa yapılan bağlantılarda farklı hizmet tipi olma yüzdesi, son 2 saniye içinde aynı hizmet tipini kullanan bağlantı sayısı, son 2 saniye içinde aynı hizmet tipini kullanan bağlantılarda "SYN" hata yüzdesi, son 2 saniye içinde aynı hizmet tipini kullanan bağlantılarda "REJ" hata yüzdesi, son 2 saniye içinde aynı hizmet tipini kullanan bağlantıların farklı varış noktalarının sayısı. Burada sembolik olarak belirtilmemiş olan özellikler sayı tipindedir[4].

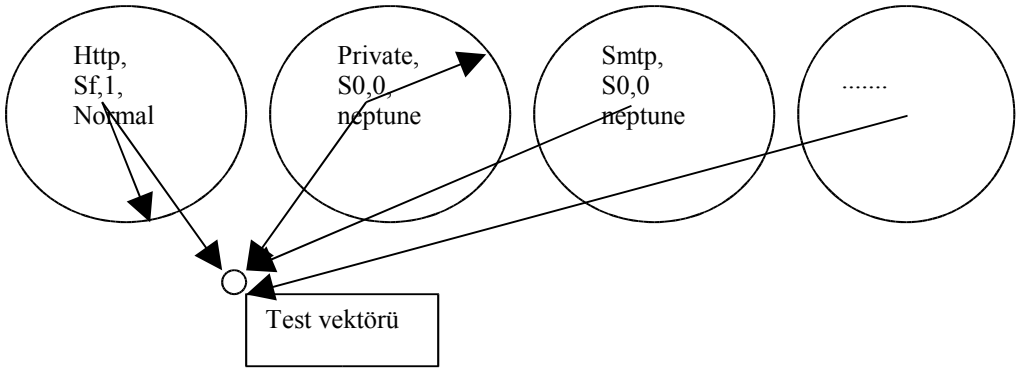
Tüm özellik seçimi algoritmalarında hep aynı yaklaşımın söz konusu olduğu gözlenmektedir: Şöyle ki; eğer bir özellik sınıf kavramıyla ilişkili ise ve sınıf kavramıyla ilişkili bütün diğer özelliklerden bağımsız ise amaca uygundur ve değerlendirmede kullanılabilecek özellikler arasında yer alabilir. Bu çalışmada mevcut özellik seçimi algoritmaları içinde FCBF algoritmasından yararlanılmıştır. Bu algoritma her özellik için değer skalasının önceden verilmesini gerektirmektedir. Bu durum sadece sembolik verileri kullanmayı olanaklı kıldığı için, algoritmaya sembolik olmayan verileri de değerlendirmeye katabilecek şekilde ekleme yapılmıştır. Bunun için sayı tipindeki veriler kullanıcı tarafından verilen bir sayı kadar bölmelere ayrılmış ve test örüntülerindeki özelliklerin o bölmelerden hangisinde olduğuna bakılarak sayısal değerler sembolik değerlere çevrilebilmiştir.

3. ÖNERİLEN SALDIRI TESPİT SİSTEMİ

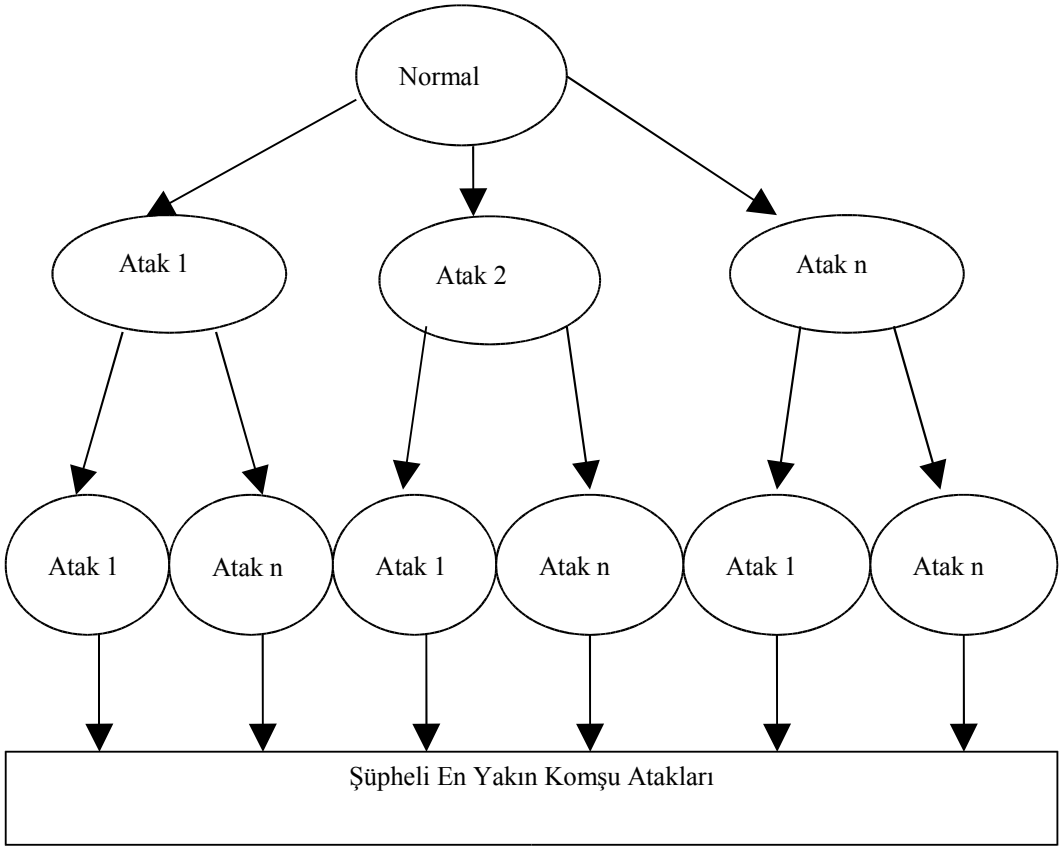
Önerilen saldırı tespit sistemi 2 aşamalı bir çalışma sistemine sahiptir: Öğrenme aşaması ve saldırı tespit aşaması (test aşaması). Öğrenme aşamasında atak ve normal sınıflarının her biri birebir olarak, geliştirilmiş FCBF algoritması kullanılarak karşılaştırılmış, herbiri için belirleyici özellikleri çıkarılmış, bu özelliklerden sembolik olanlar oluşturulan örüntü sınıflarının etiketleri haline getirilmiştir. Bu örüntü sınıflarının her birinin ortalama vektörü bulunmuş, bu vektör ile sınıfa ait olan en uzak vektör arasındaki uzaklık sınıfın yarıçapı olarak adlandırılmıştır. Saldırı tespit aşamasında ise test vektörü ile örüntü sınıflarının her birinin ortalama vektörü arasında Euclid uzaklığı hesaplanmış, buna göre test vektörü en yakın sınıfa atanmıştır. Euclid uzaklığı hesaplanırken seçilmiş ve sembolik olmayan özellikler için ağırlık verme olanağı tanınmıştır.

Şekil 1'de "Neptune" atağı ve "Normal" bağlantı arasındaki örüntü sınıflarının gösterimi ve test vektörünün bunlarla karşılaştırılması yer almaktadır. Burada görülmektedir ki, sınıfları ayırt edici özellikler varış hizmet tipi (http, private, smtp v.b.), TCP/IP bağlantısı kurmanın normal (SF ile gösterilmiştir) ya da hatalı olma durumu (S0 ile gösterilmiştir), başarılı "login" olma (1 ile gösterilmiştir) ya da olmama (0 ile gösterilmiştir) durumudur. Bunlara karşılık olarak birinci örüntü sınıfının etiketi "Http, SF, 1, Normal", ikincinin etiketi "Private, S0, 0, Neptune", üçüncünün etiketi "Smtp, S0, 0 Neptune" v.b. dir. Test vektörü için öncelikle aynı seçilmiş özelliklerden sembolik olanlar birebir karşılaştırılmakta, bunlardan aynı olan birden fazla örüntü sınıfı bulunursa Euclid uzaklığı hesaplamasına geçilmektedir.

Şekil 2'de sistemin genel bakış açısı verilmiştir. Sistem test durumunda da iki aşamada çalışmaktadır. Birinci aşamada test vektörü "Normal" sınıfı ile diğer tüm atak sınıfları arasındaki karşılaştırmadan geçmektedir. Bu karşılaştırma sembolik verilerin eşitliği ve test vektörünün örüntü sınıfı ortalama vektörü arasındaki uzaklığın örüntü sınıfının yarıçapının belli bir sabitle çarpımından küçük olması koşullarını kapsar. İlk aşamada sembolik verileri hiçbir sınıfa uygun bulunmadıysa ya da tüm sınıflara olan uzaklığı sınıfın yarıçapının ilgili sabitle çarpımından büyükse test vektörü "Anormal" olarak tanımlanmaktadır. Test vektörü, "Normal" sınıfına yakın bulunduysa normal olarak tanımlanmakta, herhangi bir ya da birden fazla atak sınıfına yakın bulunduysa ikinci aşamaya geçilmektedir. Burada test vektörü bulunmuş atak sınıflarına ait olduğu düşünülüp diğer atak sınıflarının hepsiyle karşılaştırılmaktadır. Buradan çıkan sonuçlar (atak tipi ya da anormal olma durumu) şüpheli en yakın komşu kümesine atılmaktadır. Şüpheli en yakın komşu kümesinde yapılan oylama sonucu en fazla bulunan atak tipi ya da "anormal" olma durumu test vektörünün tipini belirler.



Şekil 1 Neptune atağı ve normal bağlantı arasındaki örüntü sınıflarının gösterimi ve test vektörünün bunlarla karşılaştırılması (Sınıf ortalama vektörleriyle test vekörü arasındaki Euclid uzaklığının karşılaştırılması)



Şekil 2 Şüpheli En Yakın Komşu Atakları

3.SONUÇ

Bu çalışmada örüntü sınıflandırmasını baz alarak geliştirilmiş bir saldırı tespit sistemi geliştirilmiştir. Bu sistemde iki problem çözülmeye çalışılmıştır: 1) Sistemin test vektörlerini etiketleyebilmesi; 2) Sistemin "Anormal" durumları yakalayabilmesi. KDD Cup 99 yarışmasında kazanan yarışmacının sonuçları şunlardır: % 83.3 DOS (Denial of Service- Hizmeti aksatma), % 97 Probe (Deneme), % 13 U2R (Unathorized access to local super user - Yerel kök

kullanıcı olmak için yetkisiz erişim), %8 R2L (Unauthorized access from a remote machine - uzak bir makineden yetkisiz erişim) atakları. Bu sonuçlar içinde anormal durumları yakalayabilme üzerine bir bilgiye rastlanmamaktadır[5]. Sistemin varsayılan değerleri kullanılarak (yarıçap çarpım sabiti 1.2, sayı tiplerinin bölmeleme sayısı 10, ağırlık verme sabiti 10) elde edilen test sonuçları şunlardır: % 74.9 DOS, % 60.9 Probe, % 24.2 U2R, % 8.4 R2L, %72.5 Normal, % 85.9 Anormal. Elde edilen sonuçlar oldukça ümit vericidir. Özellikle DOS ve Probe yani ağ bazlı atakların tespit oranının sistemin öğrenme veritabanının artırılmasıyla birebir bağlı olduğu görülmüştür. Program veritabanın genişletilmesi, bir veritabanı programının kullanılması ve donanımsal

yönden güçlü bir makinede koşturulması halinde daha iyi sonuçlar alınması beklenmektedir. U2R ve R2L yani konak bazlı saldırılar için KDD Cup 99 verilerinin kullanılmasının zaten uygun olmadığı kanıtlanmıştır.

CLIDS'in sistem parametreleri, kullanıcı arayüzü kullanılarak kolaylıkla değiştirilebilmektedir. Kullanıcının isteğine bağlı olarak "Anormal" durumlar ayrıca herhangi bir atak tipiyle etiketlenerek öğrenme veritabanı geliştirilebilmekte, böylece geri beslemeli bir sistem yaratılabilmektedir.

Gelecekte özellikle ağ bazlı saldırı tiplerinin gerçek zamanlı yakalanması, bunun için program altyapısının uygun hale getirilmesi, CLIDS'in birden fazla birbirinden bağımsız makinede koşturulması halinde elde edilen sonuçların paylaşılması üzerine çalışılabilir. Bütün saldırı tespit sistemleri için geçerli olan, öğrenme veritabanının nasıl genişletileceği konusu da gene çok önemli bir araştırma konusudur.

KAYNAKLAR

[1]**Yu L., Liu H.,** 2003, Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution , Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC

[2]**Debar H., Dacier M., Wespi A.,** 1999, Research Report, A Revised Taxonomy for Intrusion-Detection Systems IBM Reserch Zurich Research Laboratory

[3]**Duda R. O. Hart P. E., Stork D.G.,** 2001, Pattern Classification, Wileyans Sons, Second Edition

[4]http://kdd.ics.uci.edu/databases/kddcup99/task.html

[5]http://www-cse.ucsd.edu/users/elkan/ clresults.html

[6]**Sabhnani M. and Serpen G.**, Why Machine Learning Algorithms Fail in Misuse Detection on KDD Intrusion Detection Data Set, Electrical Engineering and Computer Science Department-The University of Toledo