Akış Çizgesi Tabanlı Sızma Tahmin Yöntemi
(Intrusion Prediction Method Based on Flow Graph)

Öznur Erdur-Sever
GYTE Bilgisayar Mühendisliği Bölümü
oznurerdursever@gmail.com

İbrahim Soğukpınar
GYTE Bilgisayar Mühendisliği Bölümü
ispinar@bilmuh.gyte.edu.tr

Özetçe

Bilgi Teknolojilerinin gelişimi ile birlikte sistem ve sisteme dahil olan varlıkların içinde bulunduğu tehditler artmaktadır. Bu sebeple bilgi güvenliğinin sağlanması daha da önem kazanmaktadır. Günümüzde tasarlanan bilgi sistemlerinin altyapısı genellikle bilgisayar ağlarına dayanmaktadır ve bu da bilgi sistemlerini saldırılara daha açık hale getirmektedir. Ağ güvenliğinin sağlanmasında Sızma Tespit Sistemleri (Intrusion Detection Systems, IDS) önemli bir araçtır. Sızma Tespit Sisteminde uygulanan teknikler ile saldırılar, ancak kısmen ya da tamamen gerçekleştikten sonra saptanabilmektedir; buna bağlı olarak saldırının kontrol altında tutulması ya da durdurulması zor olmaktadır. Bu nedenle yeni geliştirilecek olan IDS sistemlerine saldırıyı tahmin etme özelliği dahil edilmelidir. Bu çalışmada Bilgisayar ağlarına olan sızmaların önceden tahmin edilmesine yönelik bir yöntem önerisinde bulunulmuştur. Önerilen yöntem test edilerek sonuçlar verilmiştir.

Anahtar Sözcükler: Sızma Tahmin Sistemleri, Grafa Dayalı Sızma Tahmini, İçerik Tabanlı İmza Çıkartılması, Hibrit İmza Tanımlaması

Abstract

Along with the improvement of Information Technologies, the threat; that the system and the included entities are in, has been increasing. Therefore ensuring information security gains more importance. The infrastructure of the information systems designed lately, is generally built upon computer networks; which makes information systems more prone to attacks. Intrusion Detection Systems are important tools in providing network security. The attacks in Intrusion Detection Systems (IDS) can only be detected after they occur partially or fully. And because of this, taking attacks under control or ceasing them is difficult. For this reason, attack prediction feature should be included to the new IDS systems to be designed. In this research; a method to predict intrusions through computer networks is suggested. Experimental results has been presented after testing the method.

Key words: Intrusion Prediction Systems, Graph Based Intrusion Prediction, Content Based Signature Generation, Hibrit Signature Definition

1. Giriş

Bilgi Teknolojilerinde tasarlanan sistemler zamanla daha çok iletişime ihtiyaç duymaktadır; bu nedenle bilgisayar ağlarının sistem tasarımında kullanımı artmış ve buna bağlı olarak ağ güvenliği önemli bir konu haline gelmiştir.

Bilgisayar ağlarında; ağdaki veri paketlerinin dinlenmesi, hizmet dışı bırakma saldırıları, parola saldırıları, yazılımdaki zayıflıkların kötüye kullanılması, zararlı kodlar, virüsler, truva atları ve solucanlar gibi unsurlar güvenlik için tehdit olarak örnek verilebilir.

Bilgisayar ağlarında tehditlere karşı ağ güvenliğini sağlamada kullanılan yöntemler özetle şöyle sınıflandırılabilir:

Anti Virüs Yazılımları; bilgisayar virüslerinin sisteme yerleşmesini önlemek ya da sisteme yerleşmiş virüsleri ortadan kaldırmak için kullanılan yazılımlardır.

Güvenlik Duvarları; bilgisayarları ve bilgisayar ağlarını dışarıdan gelebilecek tehditlere karşı filtrelemeye benzer bir yöntemle koruyan sistemlerdir.

Veri Şifreleme; verilerin gönderici ile alıcı haricinde başka kişiler tarafından anlaşılmasını önlemek amacıyla belli kurallar dahilinde değiştirilmesidir.

Sızma Tespit Sistemleri; bilgisayar sistemlerinde gerçekleşen olayları izleyerek güvenlik sorunu olmaları açısından değerlendirmektir. Bir sorunla karşılaştıkları zaman çeşitli alarmlar üretebilirler ya da saldırıyı önlemek için ağın işleyişinde değişiklikler yapabilirler.

Bu yöntemler atakları gerçekleştikten sonra tespit edebilmektedirler ve atakları gerçek zamanlı olarak engellemeyi zorlaştırmaktadırlar. Bu nedenle, sisteme önemli bir hasar verilmeden uygun bir cevap verilebilmesi ve amacın tipini tahmin etme özelliğine sahip proaktif sistemler tasarlanmalıdır. Bu da Sızma Tahmin Sistemi tasarımı ve çalışmalarının yapılmasını gerekli hale getirmiştir. Böylece saldırgan, hedeflediği amacına ulaşmadan tespit edilerek önlem alınması sağlanmış olacaktır.

Bu çalışmada, akış tabanlı olarak sızma tahmini için bir yöntem önerilmiştir. Önerilen yöntem KDDCup99 verileri üzerinde sınanarak elde edilen sonuçlar benzer yöntem önerisinde bulunan iki çalışma ile karşılaştırılmıştır. [14][15]

Makalenin sonraki bölümleri aşağıdaki şekilde organize edilmiştir. Bölüm 2'de Sızma tahmin sistemleri açıklanmıştır. Önerilen yöntem Bölüm 3 de verilmiştir. Deneysel sonuçlar ve karşılaştırmalar Bölüm 4'de verilmiştir. Sonuç ve öneriler son bölümde açıklanmıştır.

2. Sızma Tahmin Sistemleri ve İlgili Çalışmalar

Sızma Tahmin Sistemleri, bilgisayar ağlarında saldırılar henüz gerçekleşmeden uyarı vermeye yönelik tasarlanan sistemlerdir. Sisteme gerçek bir hasar vermeden saldırıyı önceden tahmin etmek için bağlantı kayıtlarını, sistem çağrılarını ya da transfer edilen ağ paketlerini gözlemlemek ve değerlendirmek gibi yöntemler kullanırlar. Ağ trafiği gözlemlenerek veya sistem analiz edilerek elde edilen veriler istatistik, yapay sinir ağları ve veri tabanı sınıflandırma gibi teknikler kullanılarak değerlendirilir ve kullanıcıların amaçları ve planlanan bir saldırı olma ihtimali belirlenmeye çalışılmaktadır.

Li Feng ve arkadaşları yaptıkları çalışmada; bir dizi sistem çağrısının amacının normal erişim ya da saldırı amaçlı olduğunu anlamaya yönelik bir Dinamik Bayes Ağ yaklaşımı geliştirmişlerdir. Bu şekilde sızma tehlikesini önceden tahmin ederek uyarılar verilmesini ya da saldıyı engellemek için önlemler alınmasını sağlamaya çalışmaktadırlar. [1][2]

Kjetil Haslum ve arkadaşları yaptıkları çalışmalarda dağıtık bir Sızma Tahmin ve Önleme sistemi tasarlamışlardır. Bu sistemde ağ üzerindeki değişik noktalardaki Sızma Tespit Sistemlerinden edinilen bilgileri merkezi bir sistemde Saklı Markov Modeli kullanarak değerlendirerek sızmayı tahmin etmektedirler. [3] Daha sonra bu çalışmalarına risk değerlendirme özelliği kazandırarak daha akıllı hale getirmişlerdir ve sistemlerinde performans iyileştirmelerinde bulunmuşlardır. [4][5]

Zhang Zhengdao ve arkadaşları ise çalışmalarında Saklı Yarı-Markov Model kullanarak sistem çağrılarını değerlendirmişler ve ağ üzerinde sızma tahmininde bulunmuşlardır. [6]

Liao Cheng-Bin çalışmasında anahtarlardan geçen ağ paketlerinin özelliklerini çıkartmış ve bunları önceden belirlenmiş bir ağ paketleri özellikleri kümesinde veri madenciliği sınıflandırma yöntemini kullanmıştır. Bu şekilde oluşabilecek saldırıları tahmin etmiştir. [7]

Zhi-tang Li ve arkadaşları çalışmalarında veri madenciliği tekniğini kullanarak gerçekleşen saldırıları derecelendirirken ardından sızma gelme ihtimalini hesaplanmaktadır. [8] Bir başka çalışmalarında ise yine veri madenciliği tekniğini kullanarak gerçekleşmiş olan saldırılardan bir saldırı grafiği oluşturulur ve bu grafiğe göre saldırıyı gerçekleştirenin saldırı planını hesaplamaya ve bir sonraki saldırısı tahmin etmeye çalışılmaktadır. [9]

3. Akış Çizgesi Tabanlı Sızma Tahmin Yöntemi

Bu çalışmada, daha önce polimorfik solucan tespiti için önerilen CCM (Conjunction of Combinational Motifs) metodunun uyarlanması ile bilgisayar sistemlerine sızma tahmininde kullanılacak yeni bir yöntem tasarlanmıştır. [10]

CCM, yönlü kenarlar ve bağımsız düğümlerin birleşimine dayanan yeni bir hibrit polimorfik solucan tespit imza şeması ve graf tabanlı hibrit imza sınıflarının uygulamasıdır.

Düğüm çıkartılması, önerilen CCM yönteminin ilk aşamasıdır. Önerilen düğüm skoru hesaplama yöntemini kullanarak düğümler, güçlü ve zayıf olarak ayrılmaktadır. Güçlü düğümler, CCM imza kümesinin doğrudan elemanı olarak kullanılmaktadır.

Güçlü düğümler, keşfedilmiş polimorfik solucan akış çizgesinden çıkartılır. Kalan akış çizgesi, zayıf düğümleri içerir ve bu zayıf düğümler güçlü kenarları bulmak için analiz edilir.

Sonuç olarak, elde edilen imza kümesi güçlü düğümlerin ve güçlü yönlü kenarların birleşimi olarak tanımlanır ve polimorfik solucanların tespitinde kullanılır.[10]

3.1 İçerik Tabanlı İmza Yapıları

İçerik tabanlı imza yapıları; Düğüm Tabanlı İmzalar, Kenar Tabanlı İmzalar ve Hibrit İmzalar olarak sınıflandırılabilir.

Düğüm Tabanlı İmzalar; Sızma tahmini için düğümlerden faydalanır. Tahmin yönteminde keşfedilmiş düğümlerin tamamı kullanılabileceği gibi keşfedilmiş düğümlerin bir alt kümesi de kullanılabilir.

Kenar Tabanlı İmzalar; bir sızma tahmin mekanizması tanımlamak için kenarlardan faydalanır. Kenar, iki düğümün yönsüz birleşimi ya da iki düğümün yönlü dizilimi olarak tanımlanır. Yöntem, kenar kümesi E'nin kenarlarının tamamını kullanılabileceği gibi kenarlarının bir alt kümesini de kullanılabilir.

Hibrit İmzalar; bir sızma tahmin mekanizması tanımlamak için düğümlerden ve kenarlardan faydalanır.

Önerilen sızma tahmini yönteminde; CCM çalışmasında tanımlanan imza çıkarma metodu uyarlanarak ağ kayıtları analiz edilmekte ve içerik tabanlı hibrit imzalar çıkartılmaktadır.

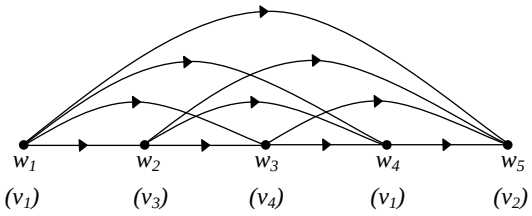
3.2 Genel Tanımlar ve Notasyon

Düğüm kümesi V : V, $n \geq 1$ olmak üzere n düğümden oluşan ve her düğümün şüpheli akış havuzundaki sızma kopyalarının belirli bir bölümünde yer alan bir ortak kayıt katarını temsil ettiği bir küme olsun. $V = \{v_1, v_2, ..., v_n\}$

Kenar kümesi E : E, her kenarın VxV kümesindeki sıralı bir düğüm ikilisini temsil ettiği, n^2 kenardan oluşan bir küme olsun $\forall_{i,j}$ için $e_{ij} = (v_i, v_j)$ ve $1 \leq i, j \leq n$ iken, $E=VxV=\{e_{ij}\}$.

Akış çizgesi X : X, keşfedilmiş düğümler $v_i \in V$ cinsinden bir akış çizgesi olsun. X; $1 \leq i \leq m$ ve $x_i \in V$ olmak üzere m düğümden oluşur. X'te hiçbir $v \in V$ düğümü yoksa $X=\emptyset$'dir. X, bir komşu düğümler listesi, komşu kenarlar kümesi veya yönlü çizge olarak ifade edilebilir.

Sızma akış çizgesi W : W, keşfedilmiş düğümler $v_i \in V$ cinsinden bir sızma çizgesi olsun. W; $1 \leq n \leq p$ olmak üzere p düğümden (w_i) oluşur ve $1 \leq i \leq p$ olmak üzere $\forall_i$ için $w_i \in V$'dir. W, bir komşu düğümler listesi, komşu kenarlar kümesi veya yönlü çizge olarak ifade edilebilir.



Şekil- 1 Sızma akış çizgesi

Düğüm skor fonksiyonu

$$f_{V_{score}} : v_i \rightarrow (V_{score})_i \in R : \ f_{V_{score}} \quad , (1)$$

düğüm kümesi V üzerinde her $v_i \in V$ için düğüm skorunu hesaplayan bir fonksiyon olsun ve V_{score} düğüm skorlarının oluşturduğu küme olsun.

Kenar skor fonksiyonu ,

$$f_{E_{score}} : e_{ij} \rightarrow (E_{score})_{ij} \in R : \ f_{E_{score}} , \ (2)$$

kenar kümesi E üzerinde her $e_{ij} \in E$ için kenar skorunu hesaplayan bir fonksiyon olsun ve E_{score} kenar skorlarının oluşturduğu küme olsun.

Akış skor fonksiyonu,

$$f_{X_{score}} : X \rightarrow (X_{score}) \in R : \ f_{X_{score}} \ , (3)$$

akış çizgesi X üzerinde akış skorunu hesaplayan bir fonksiyon olsun.

İmzaların üretilmesinde iki çeşit havuz kullanılmaktadır. Bunları Normal Akış Havuzu ve Şüpheli Akış Havuzu olarak adlandırılır. Normal Akış Havuzu, sızma içermeyen trafik akışları için tanımlanmaktadır. Sızma içeren akış havuzlar ise Şüpheli Akış Havuzu olarak tanımlanmaktadır.

3.3 Düğümlerin Tanımlanması

Önerilen sızma tahmini yönteminde düğümler, atak akışlarında görülen en uzun ortak bağlantı kayıtları katarıdır.

Düğümler çıkartılırken bilgisayar ağındaki bağlantı kayıtları analiz edilir. Atak akış havuzundaki n atak akışı örneğinden en az K tanesinde yer alan, en az α en fazla β uzunluğundaki atomik alt bağlantı kayıt katarlarından her birisi bir düğüm olarak belirlenir.

3.4 Düğüm Skorlarının Hesaplanması

Düğüm kümesi V ve Sızma akış kümesi I, oluşturulduktan sonra her düğüm $v_i \in V$ için şüpheli veya normal akışta görünme olasılıklarına göre skorlar hesaplanır.

Bu olasılıklar Bayes' kuralı kullanılarak hesaplanır. '$L(v_i) = true$' notasyonu düğümün sızma akışında olma durumunun; '$L(v_i) = false$' notasyonu ise düğümün normal akışta olma durumunun gösterimi için kullanılmaktadır.

Bir düğüm v_i'nin sızma akışında olma olasılığı $P[L(v_i) = true | v_i]$, normal akışta olma olasılığı $P[L(v_i) = false | v_i]$ olur.

Başlangıçta bir düğümün sızma akışı ve normal akış elemanı olması olasılıkları eşit kabul edilmektedir,

$$P[L(v_i) = true | v_i] = P[L(v_i) = false | v_i]$$

Düğüm kümesi V ve Akış kümesi X analiz edilerek her düğüm v_i için sızma akışı ve normal akışta görünme olasılıkları oranı hesaplanır. Bu hesaplama (4)'deki gibi olur.

$$f'_{V_{score}} = \frac{P\left[L(v_i) = true | v_i\right]}{P\left[L(v_i) = false | v_i\right]} \quad ,(4)$$

Elde edilen sayılar çok büyük olacağından bu oranın logaritmik değeri kullanılacaktır. Bu nedenle düğüm skor fonksiyonu $f_{V_{score}}$ (5)'deki denklem ile tanımlanmaktadır:

$$f_{V_{score}} : v_i \rightarrow (V_{score})_i \in R,$$
$$f_{V_{score}}(v_i) = (V_{score})_i = \log\left(\frac{P(v_i | L(v_i) = true)}{P(v_i | L(v_i) = false)}\right) \quad , (5)$$

Sızma akışında daha sık, normal akışta daha az görünen düğümler iki akışta da ortak olanlara göre daha yüksek skorlara sahiptirler. Bu düğümler güçlü düğümler, diğerleri ise zayıf düğümler olarak tanımlanırlar. Güçlü düğümlerin içerisinde görüldükleri akışların sızma akışı olması ihtimali daha fazladır. Bu nedenle, skorları hesaplanan düğümler, Güçlü ve Zayıf olarak iki kümeye ayrılmaktadır.

Güçlü düğümler V_{strong} kümesini, Zayıf düğümler ise V_{weak} kümesini oluşturmaktadır. $P_{cluster\_V}$ prosedürü, f_{Vscore} fonksiyonunu kullanarak kenar skorlarının hesaplanmasını ve k-means clustering algoritmasını kullanarak V_{strong} ve V_{weak} kümelerin oluşturulmasını sağlar.

3.5 Kenar Skorlarının Hesaplanması

V_{strong} kümesinin çıkartılmasından sonra bir sonraki adım E_{strong} yönlü kenar kümesinin belirlenmesidir. E_{strong} sızma akışı grafı *I*'daki güçlü kenarlardan oluşur.

Bir zayıf ve bir güçlü düğüm veya iki güçlü düğüm içeren kenarlar genellikle güçlüdür çünkü güçlü düğümün kendisi şüpheli akış havuzunda daha sık görülürken, normal akış havuzunda daha seyrek görülmektedir. Bu tür güçlü kenarlar göz ardı edilir ve güçlü düğümler imza kümesinde tek başına kullanılır.

İmza kümesindeki güçlü kenarlar yalnızca V_{weak} kümesindeki zayıf düğümleri içerirler. V_{strong} kümesindeki güçlü düğümler sızma akışı grafı *W*'dan silinirler ve akış grafı yönlü ve devirsiz hale getirilir bu yeni akış grafına W_{weak} adı verilir. Bu grafda her düğüm W_i tüm diğer düğümlere W_j, $j>i$ olacak şekilde bağlıdır. W_{weak}, E_{strong} kümesini çıkarmak üzere incelenir.

E_{strong}, W_{weak}'*in* kenar kümesinin alt kümesidir ve imza kümesinin güçlü kenarlarını içerir. W_{weak}'in elamanı olan her bir kenar için düğüm skorlarındaki yöntem kullanılarak sızma ve normal akışlarda görülme olasılıklarına göre kenar skorları hesaplanır.

Kenar skorları; W_{weak} kümesi elemanlarından; içerisinde bulundukları akışların sızma akışı olma ihtimali daha fazla olanları belirlemek için hesaplanır. Hesaplanan skorlara göre sızma akışlarında daha sık görülenler Güçlü kenarlar ve sızma akışlarında çok fazla görülmeyenler Zayıf kenarlardır.

$P_{cluster\_E}$ prosedürü, f_{Escore} fonksiyonunu kullanarak kenar skorlarını hesaplamak ve k-means algoritmasını kullanarak E_{strong} ve E_{weak} kümelerini oluşturmak için tanımlanmıştır.

$$f_{E_{score}} : e_{ij} \rightarrow (E_{score})_{ij} \in R,$$
$$f_{E_{score}}(e_{ij}) = (E_{score})_{ij} = \log\left(\frac{P(e_{ij} | L(e_{ij}) = true)}{P(e_{ij} | L(e_{ij}) = false)}\right) \quad , (6)$$

W_{weak} akışında *e* adet kenar olduğunu varsayarsak ve akış havuzlarındaki akışların ortalama boyutunu *L* byte olarak varsayarsak, kenar skor hesaplama ve kenar kümelerini belirlemek *O(e x L)*'de tamamlanmaktadır. En iyi dürümda tüm bulunan düğümler güçlüdür ve güçlü kenar aramaya ihtiyaç yoktur. "*e*" sıfıra eşit ise çalışma zamanı karmaşıklığı da sıfıra inecektir. En kötü durumda, bulunan tüm n adet kenar zayıf olacaktır ve tüm $e=n^2$ yönlü kenar adayları sızma akışı grafında görünecektir. Bu durumda çalışma zamanı karmaşıklığı $O(n^2 x L)$ olacaktır.

3.6 Çıkarılan İmzaların Sızma Tahmininde Kullanılması

Sızma tahmininde bulunmak için güçlü düğümler ve güçlü kenarlar kullanılarak çıkarılan hibrit imzalar kümesi kullanılır.

Bilgisayar ağlarında görülen bağlantı kayıtları tek tek değerlendirilerek analiz edilmekte olan akış kümesi *Akis* olsun,

t anında yapılan bir bağlantı kaydı $x \in Akis$ olsun, öncelikle bu *x* bağlantı kaydı için analiz edilen atak tipi için Tablo- 2'deki ilgili öznitelikler kullanılarak düğüm sınıflandırılması yapılır.

Düğüm sınıflandırması şöyle yapılır;

Analiz edilmekte olan atak tipi için önceden çıkartılan düğüm kümesi $V_{AtakTipi}$, bu kümenin eleman sayısını veren fonksiyon $s(V_{AtakTipi})$ olsun,

$V_{AtakTipi}$ kümesindeki her bir düğüm v_i için düğümü oluşturan bağlantı kaydı sayısını belirten düğüm uzunluğunu veren fonksiyon $l(v_i)$ olsun,

Eğer analiz edilmekte olan bağlantı kaydı *x* ve öncesinde gelmiş olan ardışık $l(v_i)$-1 bağlantı kaydının oluşturduğu kayıt katarı, v_i'ye eşit ise bu düğüm bu akışta görülmüş demektir.

Analiz edilmekte olan akışta görülen düğümlerin oluşturduğu küme $(Akis \cap V_{AtakTipi})$, bu kümenin eleman sayısını veren fonksiyon $s(Akis \cap V_{AtakTipi})$ olur,

Sızma tahmin değerlendirmesinde; *x* bağlantı kaydı analizinde kullanılan karar kuralı $L(x)$ aşağıdaki gibidir:

$$KararKurali: \begin{cases} L(x) = Sizma, eger \\ s(Akis \cap V_{AtakTipi}) >= s(V_{AtakTipi}) * EsikDeger \\ \\ L(x) = SizmaDegil, eger \\ s(Akis \cap V_{AtakTipi}) < s(V_{AtakTipi}) * EsikDeger \end{cases} , (7)$$

3.7 Yöntemin Uygulanması

Bu kısımda önceki bölümlerde verilen matematiksel ifadelerin sözde kodları verilmiştir.

3.4 Düğüm Skorlarının Hesaplanması başlığı altında anlatılan $P_{cluster\_V}$ prosedürünün sözde kodu aşağıdaki gibidir:

$P_{cluster\_V}$: Düğüm skorlarını hesaplar, Güçlü ve Zayıf düğümler kümelemelerini oluşturur

Girdiler: Düğüm kümesi V, Sızma Akış Havuzu, Normal Akış Havuzu.

Çıktılar: Güçlü düğüm kümesi V_{strong}, Zayıf düğüm kümesi V_{weak}, Düğüm skorları kümesi V_{score}

```
V_score = ∅
V_strong = ∅
V_weak = ∅
S = ∅  //v_i güçlü ise S(v_i)=strong
       //v_i zayıf ise  S(v_i)=weak
c_strong = 1 //V_strong kümesi merkez değeri
c_weak = 0 //V_weak kümesi merkez değeri
for all v_i € V do
  (V_score)_i = f_Vscore(vi)
end for
for all v_i € V do
  if (distance((V_score)_i, c_weak) <
      distance((V_score)_i, c_strong))
    S(v_i) = weak
  else
    S(v_i) = strong
  end if
end for
while (S changed) do
  //zayıf ve güçlü düğümler için
  //yeniden merkez hesaplanır
  for all x € {strong, weak} do
    c_x = centroid{v_j | S(v)=x}
  end for
  for all v_i € V do
    if (distance((V_score)_i, c_weak) <
        distance((V_score)_i, c_strong))
```

```
      S(v_i) = weak
    else
      S(v_i) = strong
    end if
  end for
end while
for all v_i € V do
  if S(v_i) = strong then
    V_strong = V_strong ∪ v_i
  else
    V_weak = V_weak ∪ v_i
  end if
end for
```

3.5 Kenar Skorlarının Hesaplanması başlığı altında anlatılan $P_{cluster\_E}$ prosedürünün sözde kodu aşağıdaki gibidir:

$P_{cluster\_E}$: Kenar skorlarını hesaplar, Güçlü ve Zayıf kenar kümelemelerini oluşturur

Girdiler: Kenar kümesi $E=(V_{weak} x V_{weak})$, Sızma Akış Havuzu, Normal Akış Havuzu.

Çıktılar: Güçlü kenar kümesi E_{strong}, Zayıf kenar kümesi E_{weak}, Kenar skorları kümesi E_{score}

```
E_score = ∅
E_strong = ∅
E_weak = ∅
S = ∅  //e_ij güçlü ise S(e_ij)=strong
       //e_ij zayıf ise  S(e_ij)=weak
c_strong = 1 //E_strong kümesi merkez değeri
c_weak = 0 //E_weak kümesi merkez değeri
for all e_ij € E do
  (E_score)_ij = f_Escore(e_ij)
end for
for all e_ij € E do
  if (distance((E_score)_ij, c_weak) <
      distance((E_score)_ij, c_strong))
    S(e_ij) = weak
  else
    S(e_ij) = strong
  end if
end for
while (S changed) do
  //zayıf ve güçlü kenarlar için
  //yeniden merkez hesaplanır
  for all x € {strong, weak} do
    c_x = centroid{e_ij | S(v)=x}
  end for
  for all e_ij € E do
    if (distance((E_score)_ij, c_weak) <
        distance((E_score)_ij, c_strong))
      S(e_ij) = weak
    else
      S(e_ij) = strong
    end if
  end for
end while
for all e_ij € E do
  if S(e_ij) = strong then
    E_strong = E_strong ∪ e_ij
  else
    E_weak = E_weak ∪ e_ij
  end if
end for
```

4. Deneysel Sonuçlar ve Karşılaştırma

Yöntemin analizinde testler, KDD Cup 99 veri kümesi kullanılarak yapılmıştır. Bu çalışmada KDD Cup veri kümesinin kullanılmasının sebebi; geçmiş ve günümüzde yapılan sızma tespit ve tahmin çalışmalarında kullanılıyor olması ve bu nedenle önerilen yöntemin sonuçlarının diğer çalışmalarla karşılaştırılmasına olanak sağlamasıdır.

KDD Cup 99 veri kümesi; DARPA 98 veri kümesinin bazı özniteliklerinin çıkartılmasıyla (başlangıç tarihi, ip ve port) oluşturulmuştur, ve yaklaşık 4.900.000 tane kayıt içermektedir. Saldırı oranı doğal değildir. Yaklaşık %80 oranında saldırı içermektedir. [11]

KDD Cup 99 veri kümesinde kayıtlar 41 öznitelikten oluşmaktadır. 42. öznitelik ise o kaydın göründüğü akışın tipini belirtir.

Tablo- 1 KDD Cup 99 öznitelikler

1	duration	22	is_guest_login
2	protocol_type	23	count
3	service	24	srv_count
4	flag	25	serror_rate
5	src_bytes	26	srv_serror_rate
6	dst_bytes	27	rerror_rate
7	land	28	srv_rerror_rate
8	wrong_fragment	29	same_srv_rate
9	urgent	30	diff_srv_rate
10	hot	31	srv_diff_host_rate
11	num_failed_logins	32	dst_host_count
12	logged_in	33	dst_host_srv_count
13	num_compromised	34	dst_host_same_srv_ rate
14	root_shell	35	dst_host_diff_srv_rate
15	su_attempted	36	dst_host_same_src_ port_rate
16	num_root	37	dst_host_srv_diff_ host_rate
17	num_file_creations	38	dst_host_serror_rate
18	num_shells	39	dst_host_srv_serror_ rate
19	num_access_files	40	dst_host_rerror_rate
20	num_outbound_cmds	41	dst_host_srv_rerror_ rate
21	is_host_login	42	attack_type

KDD Cup 99 veri kümesindeki saldırı tipleri 4 ana başlıkta gruplanmaktadır:

Hizmet Engelleme (Denial of Service – DoS): Bilgisayar ağlarında; bazı hesaplamalar yapılarak veya kaynakların çok fazla meşgul edilmesi ile sistemin gerçek isteklere cevap veremeyecek hale getirilmesidir.

Yönetici Hesabı ile Yerel Oturum Açma (Remote to local – R2L): Bilgisayar ağlarında; kullanıcı haklarına sahip olunmadığı halde ağa izinsiz erişim yapılmasıdır.

Kullanıcı Hesabının Yönetici Hesabına Yükseltilmesi (User to root – U2R): Bilgisayar ağlarında yönetici izni olmayan bir kullanıcının hakkı olmadığı erişimde ve işlemlerde bulunmasıdır.

Bilgi Tarama (Probe): Bilgisayar ağlarında; bilgi toplamak ya da önceden bilinen açıkları bulmak amaçlı gerçekleştirilen saldırılardır.

4.1 Deneysel Sonuçlar

KDD Cup 99 veri kümesinde, ardışık olarak görülen aynı tipteki kayıt katarları, akışlar olarak tanımlanmaktadır. Bir atak tipi analiz edilirken; analiz edilmekte olan atak tipindeki kayıt dizileri atak akışları, diğer tipte olan kayıt dizileri ise normal akışlar olarak kabul edilmektedir.

Düğümlerin seçilmesinde her atak tipi için seçilen öznitelikler kullanılarak analiz yapılmıştır. Aynı tipteki atak akışlarında ortak görünen minimum *n*, maximum *m* uzunluğundaki ardışık kayıt katarları belirlenmektedir. Bu kayıt katarları düğümler olarak tanımlanır. Düğümlerin belirlenmesinde kullanılan özniteliklerin belirlenmesinde KDD Cup 99 veri kümesindeki saldırı tipleri ve öznitelikler arasındaki bağlantıları inceleyen çalışmalardan faydalanılmıştır. ([11], [12], [13]) Analizde kullanılan öznitelikler Tablo- 2'de verilmiştir.

Tablo- 2 Atak tipleri analizinde kullanılan öznitelikler

Atak Tipi	Öznitelik	Kayıt Sayısı (Öğrenme)	Atak Grubu
Back	10, 13	2203	DoS
Land	7	21	DoS
Neptune	33, 35, 38, 39	107201	DoS
Pod	5	264	DoS
Smurf	3, 23, 28, 26, 41	280790	DoS
Teardrop	5, 8	979	DoS
Ftp write	5, 9, 23	8	R2L
Guess password	3, 4, 35	53	R2L
Imap	3, 26	12	R2L
Multihop	23	7	R2L
Phf	5, 6, 10, 14	4	R2L
Spy	-	2	R2L
Warezclient	10, 5, 1	1020	R2L
Warezmaster	-	20	R2L
Buffer overflow	14	30	U2R
Loadmodule	3, 24, 36	9	U2R
Perl	14, 16	3	U2R
Rootkit	4, 23, 24	10	U2R
Ipsweep	3, 19, 22	1247	Probe
Nmap	4, 5	231	Probe
Portsweep	15, 28	1040	Probe
Satan	22, 23, 27	1589	Probe

Normal	3, 34, 35, 39, 38, 26, 25, 30, 29	97277	-

Analizler yapılırken Sızma Tahmin Sistemi her atak tipi ve normal kayıtları için ayrı ayrı çalıştırılmıştır. Bir akış analiz edilirken alarm üretilirse o atak akışı tahmin edilmiş sayılmaktadır. Tüm veri kümesi analiz edildikten sonra tahmin edilen ve tahmin edilemeyen atak akışları değerlendirilerek aşağıdaki deneysel sonuçlar çıkartılmıştır:

Saldırı Tespit Oranı(Detection Rate): Tahmin edilen saldırıların tüm saldırılara oranıdır.

Yanlış Pozitif Oranı(False Positive Rate): Saldırı olarak tahmin edilen normal kayıtların tüm gerçek normal kayıtlara oranıdır.

Yanlış Negatif Oranı(False Negative Rate): Normal olarak sınıflandırılan saldırı kayıtlarının tüm gerçek saldırı kayıtlarına oranıdır.

Tablo- 3 Akış Çizgesi Tabanlı Sızma Tahmin Sistemi Detaylı Analiz Sonuçları

Atak Tipi	Saldırı Tespit Oranı	Yanlış Pozitif Oranı	Yanlış Negatif Oranı	Kayıt Sayısı (Test)
Back	1	0	1	2203
Land	1	0,0002	0	21
Neptune	0,803	0	0,197	1072002
Pod	0,697	0	0,303	264
Smurf	1	0	0	2807886
Teardrop	0,897	0	0,103	979
Ftp write	0,50	0,391	0,50	8
Guess password	1	0,125	0	53
Imap	1	0,047	0	12
Multihop	1	0,323	0	7
Phf	1	0	0	4
Spy	-	-	-	2
Warezclient	0,613	0,158	0,387	1020
Warezmaster	-	-	-	20
Buffer overflow	0,867	0,129	0,133	30
Loadmodule	0,333	0,126	0,667	9
Perl	1	0,002	0	3
Rootkit	1	0,779	0	10
Ipsweep	0,926	0,114	0,074	1247
Nmap	0,892	0,001	0,108	2316
Portsweep	0,708	0	0,292	10413
Satan	0,995	0	0,004	1589
Normal	0,945	0	0,055	972780

4.2 Karşılaştırma

Atak gruplarının saldırı tespit oranları hesaplanırken ilgili gruptaki atak tiplerinin saldırı tespit oranları atak tipindeki kayıt sayılarına göre ağırlıklı ortalaması alınır.

$$\frac{\sum_{i=AtakTipi}\left(SaldiriTespitOrani_i * KayitSayisi_i\right)}{KayitSayisi_{AtakGrubu}}, (8)$$

Tablo- 4 Akış Çizgesi Tabanlı Sızma Tahmin Sistemi Analiz Sonuçları

Veri Kümesi	Saldırı Tespit Oranı	Yanlış Pozitif Oranı	Yanlış Negatif Oranı
DoS	0,946	~0,000	0,054
R2L	0,626	0,154	0,374
U2R	0,808	0,246	0,192
Probe	0,782	0,009	0,218
Normal	0,945	-	0,055

Önerilen yönteme benzer ve KDD Cup 99 veri kümesi ile test edilmiş iki farklı yöntemin sonuçları aşağıda sunulmuştur:

Tablo- 5 A Neuro-Genetic Based Short-Term Forecasting Framework for Network Intrusion Prediction System Analiz Sonuçları [14]

Veri Kümesi	Saldırı Tespit Oranı	Yanlış Pozitif Oranı	Yanlış Negatif Oranı
DoS	0,968	-	0,032
R2L	0,97	-	0,003
U2R	0,885	-	0,115
Probe	0,6873	-	0,3127
Normal	0,998	0,002	-

Tablo- 6 An Adaptive Automatically Tuning Intrusion Detection System [15]

Veri Kümesi	Saldırı Tespit Oranı	Yanlış Pozitif Oranı	Yanlış Negatif Oranı
DoS	0,992	0,023	0,008
R2L	0,244	0,255	0,756
U2R	0,131	0,891	0,869
Probe	0,883	0,134	0,117
Normal	0,953	0,147	0,047

Yapılan analizlerde önerilen yöntemin Remote to Local (R2L) ve User to Root (U2R) saldırı grubu atakların tahmininde Tablo- 6'daki yöntemden daha iyi sonuç verirken Probe saldırı grubu atakların tahmininde ise Tablo- 5'deki yöntemden daha iyi sonuç verdiği gözlemlenmiştir. Denial of Service (DoS) grubu atakların tahmininde ise iki yönteme yakın sonuçlar vermiştir.

5. Sonuç ve Öneriler

Sızma Tespit Sistemlerinin kullanımının ve buna bağlı olarak da öneminin artması nedeniyle bu konuda oldukça fazla çalışma yapılmaktadır. Gelişen teknoloji kullanılarak bu sistemlere saldırıları tespit etmenin yanında kullanıcının amacını önceden tahmin etmek ve uyarı vermek gibi yeteneklerin kazandırılması zorunlu hale gelmiştir.

Bu çalışmada, bilgisayar ağları saldırılarını önceden tahmin etmek için akış çizgesi tabanlı bir bir yöntem önerisinde bulunulmuştur. Bu yöntemde daha önceden gerçekleşen saldırı kayıtları analiz edilerek çıkarılan hibrit imzalar kullanılarak saldırılar önceden tahmin edilir. Yapılan deneysel sonuçlar karşılaştırılmalı olarak verilmiş ve yöntemin güvenilirliliği gösterilmiştir.

Önerilen yöntem ilk kez gerçekleşecek olan saldırı tiplerini tahmin edememektedir. Bu yaklaşımda, saldırıları önceden tahmin etmek üzere kullanılacak imzaları çıkarmak için, daha önce gerçekleşmiş olan saldırı bilgilerine ihtiyaç duyulmaktadır. İleriki çalışmalarda; sistemin gerçekleşen ataklarla öğrenebilir ve iyileşebilir olması konusunda çalışmalarda bulunulabilir.

Bununla birlikte ağ güvenliğinde sızma tahmin ve harekete geçme süresi oldukça önem taşımaktadır. Dolayısı ile yapılacak çalışmalarda önerilen yöntemin zaman açısından performansı analiz edilebilir ve sonuçları değerlendirilebilir.

Kaynakça

[1] **Feng L., Wang W., Zhu L., Zhang Y.**: Predicting the intrusion intentions by observing system call sequences. Journal of Computers & Security 23, 241–252 (2004)

[2] **Feng L., Wang W., Zhu L., Zhang Y.**: Predicting intrusion goal using dynamic Bayesian network with transfer probability estimation. Journal of Network and Computer Applications 32 721–732 (2009)

[3] **Haslum K., Abraham A., Knapskog S.**: DIPS: A Framework for Distributed Intrusion Prediction and Prevention Using Hidden Markov Models and Online Fuzzy Risk Assessment. IEEE, Third International Symposium on Information Assurance and Security (2007)

[4] **Haslum K., Abraham A., Knapskog S.**: Fuzzy Online Risk Assessment for Distributed Intrusion Prediction and Prevention Systems. IEEE, Tenth International Conference on Computer Modeling and Simulation (2008)

[5] **Haslum K., Moe M.E.G., Knapskog S**: Real-time Intrusion Prevention and Security Analysis of Networks using HMMs. IEEE (2008)

[6] **Zhengdao Z., Zhumiao P., Zhiping Z.**: The study of intrusion prediction based on HsMM. IEEE, Asia-Pacific Services Computing Conference (2008)

[7] **Cheng-Bin L.**: A New Intrusion Prediction Method Based on Feature Extraction. IEEE, Second International Workshop on Computer Science and Engineering (2009)

[8] **Li Z., Lei J., Wang L., Li D.**: A Data Mining Approach to Generating Network Attack Graph for Intrusion Prediction. IEEE, Fourth International Conference on Fuzzy Systems and Knowledge Discovery (2007)

[9] **Li Z., Lei J., Wang L., Li D.**: Assessing Attack Threat by the Probability of Following Attacks. IEEE, International Conference on Networking, Architecture, and Storage (2007)

[10] **Bayoglu B., Soğukpınar I.**: Graph based signature classes for detecting polymorphic worms via content analysis. Elsevier, Computer Networks 56 832–844 (2012)

[11] **Kayacik, H.G., Zincir-Heywood, A.N., Heywood, M.L.** (2006). Selecting Features for Intrusion Detection: A Feature Analysis on KDD 99 Intrusion Detection Datasets.

[12] **Olusola, A.A., Oladele, A.S., Abosede, D.O.**: Analysis of KDD'99 Intrusion Detection Dataset for Selection of Relevance Features. In Proceedings of the World Congress on Engineering and Computer Science (Vol. 1, pp. 20-22) (2010, October).

[13] **Kim, B. J., & Kim, I. K.**: Robust Real-time Intrusion Detection System. International Journal of Information Processing Systems Vol. 1, No. 1 (2005)

[14] **Sindhu, S. S. S., Geetha, S., Marikannan, M., Kannan, A.**: A neuro-genetic based short-term forecasting framework for network intrusion prediction system. International Journal of Automation and Computing, 6(4), 406-414 (2009).

[15] **Yu, Z., Tsai, J. J., & Weigert, T.**: An adaptive automatically tuning intrusion detection system. ACM Transactions on Autonomous and Adaptive Systems (TAAS), 3(3), 10. (2008)

[16] KDD-cup data set, Available at URL http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (2004)