K×KNN: K-MEANS VE K EN YAKIN KOMŞU YÖNTEMLERİ İLE AĞLARDA NÜFUZ TESPİTİ

Sibel Kırmızıgül Çalışkan
Doğan Yayın Holding
Sistem Geliştirme
skirmizigul@dmg.com.tr

İbrahim Soğukpınar
Gebze Yüksek Teknoloji Enstitüsü
Bilgisayar Mühendisliği Bölümü
ispinar@bilmuh.gyte.edu.tr

ABSTRACT

Today, network security is one of the most important study and research topics in information technology systems. Intrusion detection systems which supports network security that is defined as to achieve the continuity of integrity, secrecy and availability of information, is improved by data mining methods for discovering hidden, important, unknown and useful information from databases which are include network traffic information and analyzing network status and satisfies the detection of anomalous in a short time before damaging network security are used widespread.
In this work, a hybrid structure has been proposed using two methods that are k-means and k nearest neighbors together with working on attribute selection, relations between attributes and data preprocessing. Firstly data set has been divided in subsets by the k-means algorithm. Later the k nearest neighbor algorithm has to been run on all subsets having different characteristics. Finally the test result of all data set has been obtained with combining the all subsets' results. When analyzing the results of k-means, KNN and new algorithms, the hybrid structure has been produced better results is seen.

Key words: Intrusion detection, data mining, k-means, k nearest neighbour

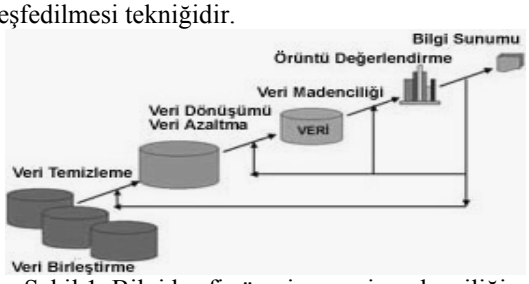
1. GİRİŞ

Günümüzde artan veri sayısı ve çeşitliliği, bilgisayar ve internet kullanımının yaygınlaşması ve bilgi toplumu olma yolundaki adımlar veri madenciliğini daha fazla gündeme getirmiştir. Büyük miktardaki veri içerisinden desenlerin, ilişkilerin, önemli bilgilerin keşfedilmesi tekniği olan veri madenciliği; bütünlük, gizlilik ve erişime açıklık koşullarının sağlanması ile olabilecek ağ güvenliği alanında ve güvenlik duvarları, erişim kontrolleri gibi güvenlik önlemlerinin yerini almaktansa, var olan güvenlik önlemlerini desteklemek için geliştirilen nüfuz tespit sistemlerinde yaygın olarak kullanılmaktadır.
Nüfuz tespiti için yapılan çalışmalarda, kötüye kullanım tespiti ve anormallik tespiti yöntemleri kullanılırken [Leung and Leckie, 2005]; yapay sinir ağları, bayes ağlar ve KNN [Nikulin, 2005] gibi sınıflandırma yöntemleri; bölünmeli, çizge tabanlı ve hiyerarşik demetleme yöntemleri, karar ağaçları ve genetik algoritmalar da mevcuttur [Stein et all, 2005]. Ayrıca farklı yöntemlerin birleştirilmesi ile oluşturulan hibrit yöntemler de kullanılmaktadır.
Bu çalışmada, nüfuz tespiti için kullanılan kümeleme ve sınıflandırma yöntemlerinin eksik ve güçlü yönleri incelenerek; kümelemeyi ve sınıflandırmayı, denetimli ve denetimsiz öğrenimi, k-means ve k en yakın komşu yöntemlerini bir arada kullanan hibrit bir yapı geliştirilmiştir. K-means ve k en yakın komşu yöntemleri ile ayrı ayrı alınan sonuçların daha da iyileştirilmesi amaçlanan çalışmada, tek ve geniş bir küme için belirlenen k ve eşik değerlerin, tüm kümeyi etkilemesi ve hepsi için zorunlu kılınması yerine, karakteristik özelliklerine göre ayrılan her alt küme için ayrı k ve eşik değerler belirlenerek zorunluluk kaldırılmış ve kümelere özgü değerler ile esnek bir yapı oluşturulmuştur. Geliştirilen yeni yöntemle yapılan testlerde, nüfuz tespitinde doğruluk oranının arttığı görülmüştür.

2. VERİ MADENCİLİĞİ, AĞ GÜVENLİĞİ VE NÜFUZ TESPİTİ İLE İLGİLİ ÇALIŞMALAR

2.1. Veri Madenciliği

Doksanlı yılların sonlarına doğru veri tabanı sistemlerinin artan kullanımı ve hacimlerindeki bu olağanüstü artış, organizasyonları elde toplanan bu verilerden nasıl faydalanılabileceği problemi ile karşı karşıya bırakmıştır. Problem çözümü için geliştirilen veri madenciliği, büyük miktardaki veri içerisinden desenlerin, ilişkilerin, değişimlerin, düzensizliklerin ve önceden fark edilmemiş, üstü kapalı, çok net olmayan ancak önemli olan bilgilerin keşfedilmesi tekniğidir.



Şekil 1. Bilgi keşfi süreci ve veri madenciliği

Teoride veri madenciliği bilgi keşfi sürecinin bir parçası olarak kabul görürken pratikte veri madenciliği ve bilgi keşfi eş anlamlı olarak kullanılmaktadır. [Lee and Stolfo, 2000]

2.2. Ağ Güvenliği ve Veri Madenciliği

Ağ güvenliği, bilgilerin güvenilir bir ortamda bozulmadan iletiminin sağlanması, güvenliğe aykırı durumların ve saldırıların tespit edilmesi ve ağ içerisindeki araçların çalışmalarının kontrolü işlemlerinin düzenli olarak gerçekleştirilmesi ile bütünlüğün, gizliliğin ve erişime açıklığın devamlılığının sağlanması olarak tanımlanmaktadır.
Nüfuz tespit sistemleri, internet veya yerel ağdan gelebilecek, ağdaki sistemlere zarar verebilecek, çeşitli paket ve verilerden oluşan saldırıları fark etmek üzere tasarlanmış sistemlerdir. Ağ trafiği (ağ-temelli) ya da bir bilgisayar sistemi içerisinde toplanan veriler (sunucu-temelli) incelenerek saldırılar tespit edilmektedirler. CERT (Computer Emergency Readiness Team) tarafından yıllara göre rapor edilen saldırı sayıları, her geçen yıl artışın ciddi boyutlara ulaştığını göstermektedir. [CERT/CC, 1988]
Saldırılar dört temel kategoride toplanmaktadır:

- DOS: Hizmet engelleme.
- R2L: Yönetici hesabı ile yerel oturum açma.
- U2R: Kullanıcı hesabının yönetici hesabı gibi davranmaya çalışması.
- PROBE: Bilgi tarama.

2.3. Nüfuz Tespiti ile İlgili Çalışmalar

Nüfuz tespit sistemleri, kötüye kullanım(imza tanıma-temelli) tespiti ve anormallik tespiti olmak üzere iki temel yöntemi yaygın olarak kullanmaktadırlar. Kötüye kullanım tespitinde, ağ verisinde incelenen davranış, önceden bilinen bir saldırı ile eşleşiyorsa saldırı olarak sınıflandırılmaktadır. Fakat bu yöntem ile sadece bilinen saldırılar tespit edilebilmektedir. Doğruluğun arttırılması için çok büyük miktarlarda etiketlenmiş veri olmalıdır ve her yeni bulunan saldırı örneği ile küme güncellenmelidir. Diğer taraftan anormallik tespiti, etiketlenmiş veri üzerinde modelleri eğiterek normal davranıştan sapmaları incelemektedir. Sadece normalliği tanıyan ve kategorize eden bu sistem ile normal davranıştan farklılık gösteren davranışlar saldırı olarak algılanmaktadır. Anormallik tespitinin kötüye kullanım tespitine göre avantajı, daha önceden tanınmayan saldırı türlerini keşfedebilme olasılığıdır. [Leung and Leckie, 2005]
Önemli ağ karakteristiklerinin keşfi için veri tabanı sistemleri ve sorgulama dilleri de kullanılmaktadır. Bunun için gerekli ilişkileri ve koşulları içeren protokol tabloları ve ilişkisel veri tabanı ile birleştirilen ağ trafiği verileri içindeki hatalar ve anormallikler tespit edilmektedir.[Zaki and Sobh, 2005]
1980'lerden sonra, amaç fonksiyonu birbirine bağlı basit işlemci ünitelerinden oluşan bir ağ üzerine dağıtan ve kullanılan öğrenme algoritmaları ile veriden üniteler arasındaki bağlantı ağırlıklarını hesaplayan yapay sinir ağları yaygın olarak kullanılmaya başlanmıştır. Öğrenilen fonksiyonların ve kuralların kökten yaprağa doğru inilerek çıkartılmasını ve anlaşılabilecek şekilde sunulmasını sağlayan karar ağaçları ve genetik algoritmalar [Stein et all, 2005] da yapay sinir ağları gibi nüfuz tespiti için kullanılan sınıflandırma yöntemleri arasında yer almaktadırlar. Olasılık yöntemlerini kullanarak örneklerin hangi sınıfa hangi olasılıkla ait olduklarını gösteren bayes sınıflandırıcılar ve bayes ağlar, belirlenen komşu sayısı ve eşik değer ile k en yakın komşu yöntemi [Nikulin, 2005], veriyi ayırmada doğrusal bir sınır kullanan karar destek makinaları ve bulanık küme sınıflandırıcıları da diğer sınıflandırma yöntemlerindendir. Demetleme yöntemleri arasında ise bölünmeli demetleme yapan k-means, çizge tabanlı demetleme ve hiyerarşik demetleme yer almaktadır.
Nüfuz tespit sistemlerinin başarımı, kullandıkları farklı yöntemler ile normal ve anormal verileri birbirinden ayırıken sadece anormallikleri ne oranda tespit ettiği ile ölçülmemektedir. Bunula birlikte gürültülü veri ile çalışabilmesi, düşük yanlış alarm seviyesi, büyük veriler ile başa çıkabilmesi ve olayları ilintilendirebilmesi de nüfuz tespit sistemlerinden beklenenler arasında yer almaktadır. Nüfuz tespitlerinde karşılaşılan en büyük problemlerden birisi olan fazla miktarda normal davranışın saldırı olarak algılanmasının azaltılması amacıyla yapılan çalışmalar arasında saldırı olarak algılanan verilerin sınıflandırma ve demetleme yöntemleri ile ikinci bir işlemden geçirilmeleri de yer almaktadır. [Pietraszek and Tanner, 2005]

3. K-MEANS, K EN YAKIN KOMŞU VE GELİŞTİRİLEN YÖNTEM

3.1. K-Means Yöntemi

K-means yöntemi, kümeleme problemini çözen en basit denetimsiz öğrenme yöntemleri arasında yer alır. Algoritmanın genel mantığı n adet veri nesnesinden oluşan bir veri kümesini(X), giriş parametresi olarak verilen k ($k \leq n$) adet kümeye bölümlemektir. Amaç, gerçekleştirilen bölümleme işlemi sonunda elde edilen kümelerin, küme içi benzerliklerinin maksimum ve kümeler arası benzerliklerinin minimum olmasını sağlamaktır.
Yöntemin performansını k küme sayısı, başlangıç olarak seçilen küme merkezlerinin değerleri ve benzerlik ölçümü kriterleri etkilemektedir.

3.2. K En Yakın Komşu Yöntemi

K En Yakın Komşu yöntemi, sınıflandırma problemini çözen denetimli öğrenme yöntemleri arasında yer alır. Yöntemde; sınıflandırma yapılacak verilerin öğrenme kümesindeki normal davranış verilerine benzerlikleri hesaplanarak; en yakın olduğu düşünülen k verinin ortalamasıyla, belirlenen eşik değere göre sınıflara atamaları yapılır. Önemli olan, her bir sınıfın özelliklerinin önceden net bir şekilde belirlenmiş olmasıdır.
Yöntemin performansını k en yakın komşu sayısı, eşik değer, benzerlik ölçümü ve öğrenme kümesindeki normal davranışların yeterli sayıda olması kriterleri etkilemektedir.

3.3. Geliştirilen Yöntem

K-means ve k en yakın komşu yöntemleri ile ayrı ayrı alınan sonuçların daha da iyileştirilmesi amacıyla yapılan çalışmada, denetimli ve denetimsiz öğrenimi, kümelemeyi ve sınıflandırmayı, k-means ve KNN yöntemlerini bir arada kullanan hibrit bir yapı geliştirilmiştir. KNN ile tek ve geniş bir küme için belirlenen k ve eşik değerin, tüm kümeyi etkileyerek küme içindeki farklı özelliklere sahip normal davranış ve saldırı verileri için zorunlu kılınması, saldırı tesip oranının yüksek tutulmasını ve aynı zamanda yanlış pozitif oranının düşük olmasını zorlaştırmaktadır. Nüfuz tespiti için yüksek tutulan k ve eşik değerler normal davranışların saldırı olarak algılanmasına neden olmaktadır. Farklı türdeki verilerin hepsi için tek bir k ve eşik değer sınırlaması yerine, karakteristik özelliklerine göre ayrılan her alt küme için ayrı k ve eşik değerler belirlenerek zorunluluk kaldırılmış ve kümelere özgü değerler ile esnek bir yapı oluşturulmuştur.

Yöntemin adımları aşağıda gösterilmiştir.

1. Test kümesi, k-means yöntemi ile k alt kümeye bölünür.
   a. Bölme işleminde k küme için ilk küme merkezleri belirlenir. $C = \{c_1, c_2, c_3 \ldots c_k\}$ Bunun için nesneler arasından k adet rasgele nokta seçilebilir ya da tüm nesnelerin ortalaması ile de belirlenebilir.
   b. Test kümesindeki her verinin $X = \{x_1, x_2, x_3 \ldots x_n\}$ seçilen merkez noktalara yakınlığı kosinüs benzerliği ile hesaplanır. Her veri kendine en yakın merkez noktanın olduğu kümeye dahil edilir.

$$sim(x_i, merkez(c_j)) = \frac{x_i \bullet merkez(c_j)}{\| x_i \| \, \| merkez(c_j) \|} \quad (1)$$

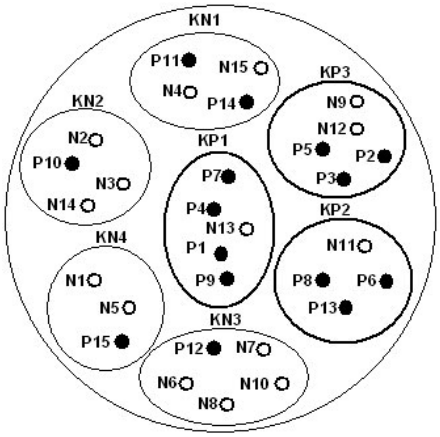
$(i = \{1, 2, 3...n\}, j = \{1, 2, 3...k\})$

   c. Oluşan kümelerin merkez noktaları o kümedeki tüm nesnelerin ortalama değerleri ile değiştirilir.

$$merkez(c_j) = \frac{\sum_{i=1}^{n_j} (x_i)}{n_j} \quad (2)$$

$(x_i \in c_j)$ve $n_j = c_j$ kümesindeki veri sayısı

   d. Merkez noktalar değişmeyene kadar b. ve c. adımları tekrarlanır. (Şekil 2)
   KP kümeleri saldırıları, KN kümeleri normal davranışları içermektedir. K-means yöntemi ile her veri tek bir kümede içinde yer almaktadır ve küme kesişimleri sıfırdır.
   $KN1 \cap KN2 = KN1 \cap KP3 = KN2 \cap KP1 = 0 \ldots$
   Bu nedenle;
   $KN1 \cup KN2 \cup KN3 \cup KN4 \cup KP1 \cup KP2 \cup KP3$ = Test Kümesi



Şekil 2. K-means ile test kümesinin bölünmesi

2. K-means ile test kümesinin bölünmesiyle oluşan farklı karakteristik özellikli alt kümenin her biri için KNN yöntemi uygulanır.
   a. Alt kümelerdeki her verinin $X' = \{x'_1, x'_2, x'_3 \ldots x'_n\}$, öğrenme kümesindeki verilere $D = \{d_1, d_2, d_3 \ldots d_m\}$ yakınlığı kosinüs benzerliği ile hesaplanır.

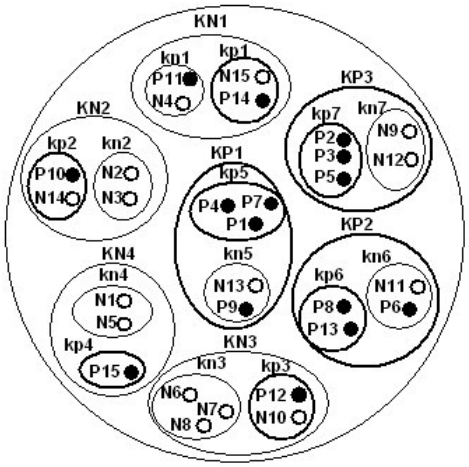
$$sim(x'_i, d_l) = \frac{x'_i \bullet d_l}{\| x'_i \| \, \| d_l \|} \quad (3)$$

$(i = \{1,2,3...n_j\}, l = \{1,2,3...m\})$

   b. Her verinin öğrenme kümesindeki verilere olan yakınlıkları sıralanıp ilk "k" tanesi alınarak ortalamaları hesaplanır.

$$sim\_avg(x'_i) = \frac{\max \sum_{l=1}^{k} sim(x'_i, d_l)}{k} \quad (4)$$

c. Ortalama değerleri, belirlenen eşik değerinden büyük olanlar normal, küçük olanlar ise anormal olarak sınıflandırılır. Tüm alt kümelerin yapıları ile farklı k (k={5,10,15,20}) ve eşik değerler için (ed={0.75,0.80,0.85,0.90,0.95}) vereceği sonuçlar incelenerek her alt kümeye ait en uygun k ve eşik değer seçilir. (Şekil 3)



Şekil 3. KNN ile her alt kümenin sınıflandırılması

3. Alt kümelerden elde edilen doğru pozitif (DP), doğru negatif (DN), yanlış pozitif (YP) ve yanlış negatif (YN) sayıları toplanarak test kümesi için toplam sonuç elde edilmiş olur.

Algoritma Şekil 4'te gösterilmiştir.

```
Başlangıç olarak k kümenin merkezlerini belirle;
Öğrenme kümesi D belirlenir D = {d1, d2, d3, ...., dm);
repeat
    for test kümesi(X) içindeki her veri xi için do
        for k küme (C) içindeki her cj küme için do
            benzerlik hesapla sim(xi, merkez(cj));
            if (sim>max_sim) then
               max_sim = sim;
               küme(xi) = j;
        for k küme (C) içindeki her cj küme için do
            küme merkezlerini yeniden hesapla;
            merkez(cj) = sum(xi)/nj   (xi∈cj);
until verilerin küme atalamaları değişmediği sürece
for k küme (C) içindeki her cj küme için do
   for cj kümesi içindeki her veri x'i için (x'i∈cj) do
       if x'i bilinmeyen bir sistem çağrısı ise then
          x'i = anormal;
       else
          for öğrenme kümesindeki (D) her veri dl için do
              benzerlik hesapla sim(x'i, dl);
              if sim(x'i, dl) = 1 then
                 x'i =normal; exit;
          En büyük k tane sim(x'i, dl) benzerliği bul;
          En yakın k komşu için benzerlik ort. hesapla;
          if sim_avg > eşik değer then   x'i =normal;
          else    x'i =anormal;
```

Şekil 4. K-Means ve KNN algoritması

4. UYGULAMA SONUÇLARI VE ANALİZ

Çalışmada 41 nitelik değerli KDD Cup 1999 veri kümesi kullanılmıştır. KDD Cup 1999 veri kümesi, DARPA98 veri kümesinden birkaç niteliğin çıkartılmasıyla (başlangıç tarihi, ip ve port) oluşturulmuştur. [KDD, 1999] 494.017 veri içeren test kümesi, 2430 normal davranış verisi içeren öğrenme kümesi ve 4.898.430 veri içeren sınama kümesi kullanılmıştır.

Nüfuz tespit sistemlerinde sonucu etkilemeyecek, önemsiz niteliklerin kullanılması işlem zamanını arttırırken, performansın azalmasına da neden olabilmektedir. [Chebrolu et al, 2005] Uygulamada davranış türlerini ayırmada en etkili olan nitelikler, bilgi kazancı yöntemleri ve test kümesi üzerinde alınan sonuçlara göre belirlenmiştir ve 29 nitelik değeri kullanılmıştır. Nitelik azaltma işleminin yanı sıra nitelikler arasındaki ilişkiler de incelenmiştir. Veri aktarım boyutlarında anormallik olup olmadığının incelenebilmesi için kaynağa gelen (src_bytes) ve hedefe gönderilen (dst_bytes) byte miktarlarındaki farkın oranı ve her iki değerin toplamda ne kadar bant genişliği oluşturduğu iki ayrı nitelik değeri olarak ele alınmıştır.

Z-score normalizasyon ile veri dönüşümü yapılarak nitelik değerleri ortalama ve standat sapma değerleri ile belirli aralıklara çekilmişlerdir.

$$\text{yeni nitelik (j)} = \frac{\text{nitelik değeri(j) - ortalama(j)}}{\text{standart sapma(j)}} \quad (5)$$

$$\text{ortalama(j)} = \frac{1}{n}\sum_{i=1}^{n} \text{nitelik(j)} \quad (6)$$

$$\text{standart sapma(j)} = \sqrt{\left(\frac{1}{n\text{-}1}\right)\sum_{i=1}^{n}(\text{nitelik değeri(j)} - \text{ortalama(j)})^2} \quad (7)$$

Veri önişleme sonrasında çalıştırılan üç yöntemin sınama kümesi üzerindeki sonuçları incelendiğinde(Tablo 1) geliştirilen yöntem ile doğru pozitifin ve doğruluk oranının arttığı görülmüştür.

Tablo 1. Sınama kümesi için yöntem karşılaştırma

	K-Means	KNN	Hibrit Yöntem
DP:	3.785.561	3.794.252	3.918.604
DN:	941.083	953.064	947.066
YP:	31.697	19.716	25.714
YN:	140.089	131.398	7.046
Doğruluk:	%96,49	%96,92	%99,33
Hata:	%3,51	%3,08	%0,67
STO:	%96,43	%96,65	%99,82
YPO:	%3,26	%2,03	%2,64
STO/YPO:	%3,38	%2,10	%2,65

Tablo 2. DP, DN, YP, YN değerleri

		Öngörülen Sınıf	
		Saldırı	**Normal**
Doğru Sınıf	**Saldırı**	DP	YN
	Normal	YP	DN

K-means ve KNN yöntemlerinde yüksek oranda görülen yanlış negatif (YN) değeri, geliştirilen yöntemde azalırken, yanlış pozitif oranında (YPO) KNN yöntemine göre %0,6 değerinde bir artış görülmektedir. Doğruluk ve saldırı tespit oranında (STO) %3 değerinde bir artış görülmektedir.

Üç yöntemin saldırı türlerine göre nüfuz tespit oranları incelendiğinde de önerilen yöntemin daha iyi sonuç verdiği görülmüştür.(Tablo 3)

Tablo 3. Davranış türlerinin tespit oranları

Davranış türü	K-Means	KNN	Hibrit Yöntem
DOS	%96,81	%96,99	%99,94
PROBE	%63,59	%47,30	%89,62
R2L	%5,33	%25,49	%76,20
U2R	%1,92	%63,46	%71,15
NORMAL	%96,74	%97,97	%97,36

K-means ve KNN yöntemleri zaman karmaşası açısından karşılaştırıldığında, sonuçları çok hızlı vermesiyle ilk sırayı k-means yöntemi alırken, test kümesindeki her verinin saldırı olup olmadığının tespiti için öğrenme kümesindeki her veri ile benzerliğini kontrol eden KNN yöntemi ikinci sırada yer almaktadır. KNN yönteminde öğrenme kümesinin büyüklüğü ile doğru orantılı olarak zaman karmaşası da değişmektedir. Öğrenme kümesindeki veriler benzerliklerine göre daha küçük alt kümelere bölünebilirler. Böylece test kümesindeki her verinin saldırı tespiti için önce küme ortalamalarıyla benzerlikleri kontrol edilebilir. Benzerlik değerlerine göre sıralanan öğrenme kümeleri sırayla en yakın k komşu hesaplaması için kullanılabilirler.

5. SONUÇ VE ÖNERİLER

Bu çalışmada, Veri madenciliği yöntemlerinden "K-means" ve "K en yakın komşu" yöntemlerinin iyileştirilmesi amacıyla; nüfuz tespiti için kümelemeyi ve sınıflandırmayı, denetimli ve denetimsiz öğrenimi, k-means ve k en yakın komşu yöntemlerini bir arada kullanan hibrit bir yapı geliştirilmiştir.

Farklı boyutlardaki veri gruplarında düşük performans gösterebilen fakat gerçeklemesi kolay ve zaman karmaşası az olan "K- means" ile tek ve geniş bir küme için belirlenen k ve eşik değeri, küme içindeki farklı özelliklere sahip normal davranış ve saldırı verileri için zorunlu kılan ve zaman karmaşası çok olan, fakat k komşu ortalaması aldığı için gürültülü verilerden az etkilenen "k en yakın komşu" yöntemleri bir arada kullanılmıştır. KNN tarafından tüm küme için kullanılan tek bir k ve eşik değer, saldırı tesip oranının yüksek tutulması ve aynı zamanda yanlış pozitif oranının düşük olmasını zorlaştırmaktadır. Tek bir k ve eşik değer sınırlaması yerine, karakteristik özelliklerine göre ayrılan her alt küme için ayrı k ve eşik değerler belirlenerek zorunluluk kaldırılmış ve kümelere özgü değerler ile esnek bir yapı oluşturulmuştur.

Geliştirilen uygulamada en hızlı sonucu veren k-means uygulaması ile test kümesi daha küçük alt kümelere ayrılarak k en yakın komşu yönteminin zaman karmaşası ve bellek gereksinimi azaltılmıştır.

KAYNAKLAR

[1] Leung, K. and Leckie, C., 2005, Unsupervised Anomaly Detection in Network Intrusion Detection Using Clusters, *Proceedings of the Twenty-eighth Australasian conference on Computer Science*, ACM, 38, 333-342

[2] Nikulin, V., 2005, Threshold-Based Clustering with Merging and Regularization in Application to Network Intrusion Detection, *Computational Statistics & Data Analysis*, Elsevier, 51.2, 1184-1196.

[3] Stein, G., Chen, B., Wu, A. S. and Hua, K. A., 2005, Decision Tree Classifier For Network Intrusion Detection With GA-based Feature Selection, ACM, 2, 136 – 141.

[4] Lee, W. and Stolfo, S. J., 2000, A Framework for Constructing Features and Models for Intrusion Detection Systems, ACM, 3.4, 227 – 261

[5] CERT/CC Statistics, 1988-2005, Mellon Software Engineering Institute, CERT Coordination Center, http://www.cert.org/stats/cert_stats.html, [Ziyaret Tarihi: 10 Ocak 2007].

[6] Zaki, M. and Sobh, T.S., 2005, NCDS: Data Mining for Discovering Interesting Network Characteristics, *Information and Software Technology*, Elsevier, 47.3, 189-98.

[7] Pietraszek, T. and Tanner, A., 2005, Data Mining and Machine Learning Towards Reducing False Positives in Intrusion Detection, *Information Security Technical Report*, Elsevier, 10.3, 169-83.

[8] KDD, 1999, The Third International Knowledge Discovery and Data Mining Tools Competition Dataset, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, [Ziyaret Tarihi: 25 Aralık 2007].

[9] Chebrolu, S., Abraham, A. and Thomas, J. P., 2005, Feature Deduction and Ensemble Design of Intrusion Detection Systems, *Computers & Security*, Elsevier, 24, 295-307.