Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı, S. 46-56, Ekim 2019

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue, pp. 46-56, October 2019

Research Article

Yapay Bağışıklık Algoritmaları ile Web Trafik Verilerinde Anomali Tespiti

Emre Dandıl[1*], Kadir İlhan[2]

[1] Bilecik Şeyh Edebali Üniversitesi, Bilgisayar Mühendisliği Bölümü, Bilecik (ORCID: 0000-0001-6559-1399)
[2] Bilecik Şeyh Edebali Üniversitesi, Bilgisayar Mühendisliği Anabilim Dalı, Bilecik (ORCID: 0000-0003-1935-0572)

(Bu yayın HORA kongresinde sözlü olarak sunulmuştur.)

(First received 1 August 2019 and in final form 22 October 2019)

(**DOI:** 10.31590/ejosat.636309)

ATIF/REFERENCE: Dandıl, E. & İlhan, K. (2019). Yapay Bağışıklık Algoritmaları ile Web Trafik Verilerinde Anomali Tespiti. *European Journal of Science and Technology*, (Özel Sayı), 46-56.

Öz

Günümüzde internet dünyasında farklı türdeki tehditler ve saldırılar artarak devam etmekte ve buna paralel olarak alınan güvenlik önlemlerinde de önemli gelişmeler olmaktadır. Ayrıca, çevrimiçi ziyaretçi sayılarının oranı ile zaman serileri şeklinde gösterilen web trafik verilerinde anormal değişikliklerin hızlı ve doğru bir şekilde tespiti ve önlenmesi büyük önem taşımaktadır. Ağ verilerinde anormal trafiklerin tespiti için farklı metodolojiler ve veri sınıflandırılma teknikleri kullanılmaktadır. Bu problem genellikle sinyal pencereleri üzerinde özellik çıkarılarak sınıflandırma yapılarak değerlendirilmektedir. Bu çalışmada, ağ üzerindeki anormal web trafiklerinin tespiti için Yapay Bağışıklık Sistemlerinin Negatif Seçim Algoritmasına (NSA) dayalı bir yöntem önerilmiş ve kullanıcı dostu bir uygulama yazılımı geliştirilmiştir. Web trafiği için Yahoo Webscope S5 verisetinde bulunan gerçek veriler kullanılmış ve pencere kaydırma yöntemi kullanılarak veriler pencerelere ayrılmıştır. Yapılan deneysel çalışmalarda, web trafik verilerinde oluşan anormal trafik verilerinin tespiti, NSA'nın yapısında bulunan aktifleşen detektör sayılarındaki değişimin izlenmesi ile gerçekleştirilmiştir. Çalışmada, geliştirilen uygulama yazılımı üzerinde NSA ile web ağ trafik verilerindeki anormal durumların düşük hata oranlarıyla tespit edildiği görülmüştür.

Keywords: Ağ güvenliği, Web trafik verileri, Anomali tespiti, Yapay bağışıklık sistemleri, Negatif seçim algoritması.

Anomaly Detection in Web Traffic Data using Artificial Immune Algorithms

Abstract

In recent years, different types of threats and attacks continue to increase in the internet world. It is important to detect anomalies in the time series quickly and accurately for web traffic data measured by the number of online visitors. Different methodologies and data classification techniques are used to detect abnormal traffic in network data. This problem is generally evaluated by classifying the signal windows by feature extraction. In this study, a method based on the Negative Selection Algorithm (NSA) of Artificial Immune Systems for the detection of abnormal web traffic on the network is proposed and a user-friendly application software is developed. For web traffic, the real data in the Yahoo Webscope S5 dataset is used and the data is split into windows using the window shift method. In the experimental studies, the detection of abnormal traffic data in the web traffic data is realized by

[*] Sorumlu Yazar: Bilecik Şeyh Edebali Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Bilecik, Türkiye, ORCID: 0000-0001-6559-1399, emre.dandil@bilecik.edu.tr

monitoring the changes in the number of activated detectors in the NSA structure. On the application software developed in the study, it is observed that abnormal conditions in the web traffic data are detected with low error rates with NSA.

Anahtar Kelimeler: Network security, Web traffic data, Anomaly detection, Artificial immune systems, Negative selection algorithm.

1. Giriş

Teknolojinin gelişmesi ile internetin hayatın her alanında yer alması, özel yaşantıları, şirket bilgilerini hatta kurumsal itibarı tehdit edecek boyutlara kadar ulaşması bilgi güvenliğinin ve sürekliliğinin ne kadar önemli boyutlara geldiğinin bir göstergesi niteliğindedir. Artık sadece gerçek ortamda güvenliği sağlamak yetmeyip, sanal ortamın da güvenliğinden emin olmak gerekmektedir[1]. Sanal ortamdaki güvenlik önlemlerinin ne derece artması gerektiği hususu internet teknolojilerindeki saldırıların yeniliği ve çeşitliliği ile ölçülebilir hale gelmiştir. Günümüzde yapılan siber ataklar artık geçmişe göre daha nitelikli ve neredeyse iz bırakmadan yapılmaktadır. Bunun yanında, istihbarattan askeri alanlara kadar artık savaşlar siber ortama taşınmıştır. Dolayısıyla artık ülkeler daha az paranın harcanıp daha etkili sonuçlar alınan internet teknolojilerine yatırım yapmaktadır. Yapılan atakların önlenmesi kritik verilerin korunması ve sürekliliği açısından önemli hale gelmiştir.

Symantec'in yayınlamış olduğu 2019 İnternet Güvenliği Tehdit Raporu'nda küresel istihbarattaki veriler analiz edilerek dünya çapında 157'den fazla ülkede 123 milyon saldırı cihazından kayıt alınmış ve günlük 142 milyon tehditin engellendiği görülmüştür[2].Her ne kadar bu tehditler engellenmeye çalışılsa da aslında tamamen tehditlerden korunmanın mümkün olmadığı görüşü ön plana çıkmaktadır. Özellikle yeni atak türlerinin geliştirildiği günümüzde bu atakların etkilerinin tam olarak belirlenebilmesi ve gerekli tepkinin ortaya konması belirli bir zamanı gerektirmekte ve bu da sistemin zaafiyete uğramasına neden olmaktadır. İşte tam bu noktada saldırıların tespiti ve önceden öngörülebilmesi ağ sistemleri ile web trafikleri açısından olukça önemlidir.

Anomali tespiti günümüzün internetinde herhangi bir ağın hayati bir parçası haline gelmiştir. Anormal olarak belirtilen ağ trafikleri, kötü amaçlı beklenmeyen saldırılardan, hizmet reddi ve ağ taramaları gibi ağ saldırılarına, ağın performans ve bütünlüğüne ciddi zarar verebilir. Sürekli yeni anormalliklerin ve saldırıların ortaya çıkması, ağ bütünlüğünü riske sokan olaylarla başa çıkmak için sürekli bir zorluk yaratmaktadır. Ayrıca, trafiğin yapısındaki karmaşıklık, protokol sayısının artmasına neden olmaktadır. Anomali tespiti, bilgisayar güvenliği alanında çekişmeli bir problem sınıfıdır. Sistemler daha yakından izlendiğinde ve saldırılara tepkide gittikçe daha fazla özen gösterildiğinde, uyarıları yükseltmek için geleneksel kural tabanlı sistemler yetersiz kalmaktadır. Bu nedenle, izleme sistemlerini daha dinamik ve uyumlu hale getirmek için makine öğrenmesine dayalı anomali tespit teknikleri göz önünde bulundurulmalıdır.

Ağ üzerinde yaygın olarak Denial of Service(hizmeti engelleme), Probe(yoklama), User to Root(Root Kullanıcısı ele geçirme) ve Remote to User(Kullanıcıya uzaktan bağlanma) saldırları mevcuttur.Web trafiği sağlanan hizmet türüne, kullanıcı bağlantı şekillerine ve veri dağılımındaki düzensizliğe bağlı olarak farklı özelliklere sahiptir. Veri sınıflandırmasında anormal trafiğin tespiti noktasal veya toplu gibi farklı yollarla yapılabilmektedir[1]. Günümüzde bu saldırıların tespiti ve alınacak önlemlerin neler olabileceği hususu hala tartışılmaya devam etmektedir.Bu bakımdan veri sınıflandırması ve verilerin gerçekten zararlı olup olmadığının belirlenmesi gerçekten önem arz etmektedir. Dolayısıyla öncelikle tespit etme ve sonrasında ise aksiyon alma prosedürleri, şirket ve/veya kişilerin verilerinde herhangi bir kayıp olmaması bakımından önemlidir. Bu kaybın yaşanmaması için çeşitli çözümler mevcuttur. Bu çözümler sayesinde verinin güvenliğini sağlamak ve kurum politikalarını uygulamak çok daha kolay hale gelmektedir.

Son yıllarda düzenlenen hem hedefli hem de otomatik saldırılarda bir artış görülmüş ve bu tehditleri azaltmak için karmaşık ve katmanlı savunma mekanizmalarına ihtiyaç duyulmaktadır. Web sunucularda anormal trafik verilerinin tespiti genellikle bir zaman serisi problemi olarak değerlendirilmektedir. Bu problem sinyal pencereleri üzerinde özellik çıkarılarak sınıflandırma yapılarak değerlendirilmektedir. Ancak web trafik verilerinin genel bir karakteristik örüntü yapısı olmadığından, çözüm yöntemleri de farklılıklar içermektedir. Literatürde ağ üzerinde anormal verilerin sınıflandırılması için yapılmış birçok çalışma bulunmaktadır. Bunlardan birisinde, Münz vd. [3] çalışmalarında K-Ortalamalar(K-Means) kümeleme algoritmasını kullanarak ağ üzerinde anomali tespiti yapmışlardır. Gerçek verinin istatistiksel özelliklerini analiz ederek bir kümenin merkezini hesaplamışlardır. Bir merkez ile trafik değeri arasındaki mesafeyi hesaplayarak ağ verilerinden anomali tespitini gerçekleştirmişlerdir. Bir diğer çalışmada Thill vd. [4] Yahoo Webscope S5 verisetinde anomali tespiti için çeşitli çevrimiçi algılama algoritmaları karşılaştırılması yapılmıştır. Burada, Regresyon Analizi ile elde edilen sonuçların diğer anomali dedektörlerine kıyasla oldukça başarılı olduğu gözlemlenmiştir.

Kim ve Cho [1] tarafından yapılmış olan bir çalışmada, trafik verilerinde yer alan ve bir boyutlu zaman serisi sinyali olan mekansal ve zamansal bilgilerin etkin bir şekilde modellenmesi için bir C-LSTM sinir ağı kullanılmıştır. C-LSTM yönteminin, evrişimli bir sinir ağını (CNN), geniş kısa süreli belleği (LSTM) ve derin sinir ağını (DNN) birleştirerek daha karmaşık özellikler çıkarabileceği gösterilmiştir. Akbal ve Ergen [5] tarafından yapılan bir diğer çalışmada, kablosuz ağlarda saldırı tespitine farklı bir yönden yaklaşılarak tespit işlemi tüm kullanıcılar üzerinde yapmak yerine erişim noktası üzerinden kontrol işlemi gerçekleştirilmiştir. Ağ üzerinde uzun süreler yapılan gözlemler neticesinde yapay bağışıklık sisteminin başarılı bir şekilde aksiyon verdiği ve çalışma süresi ne kadar uzarsa o kadar başarılı çalıştığı sonucuna varılmıştır.

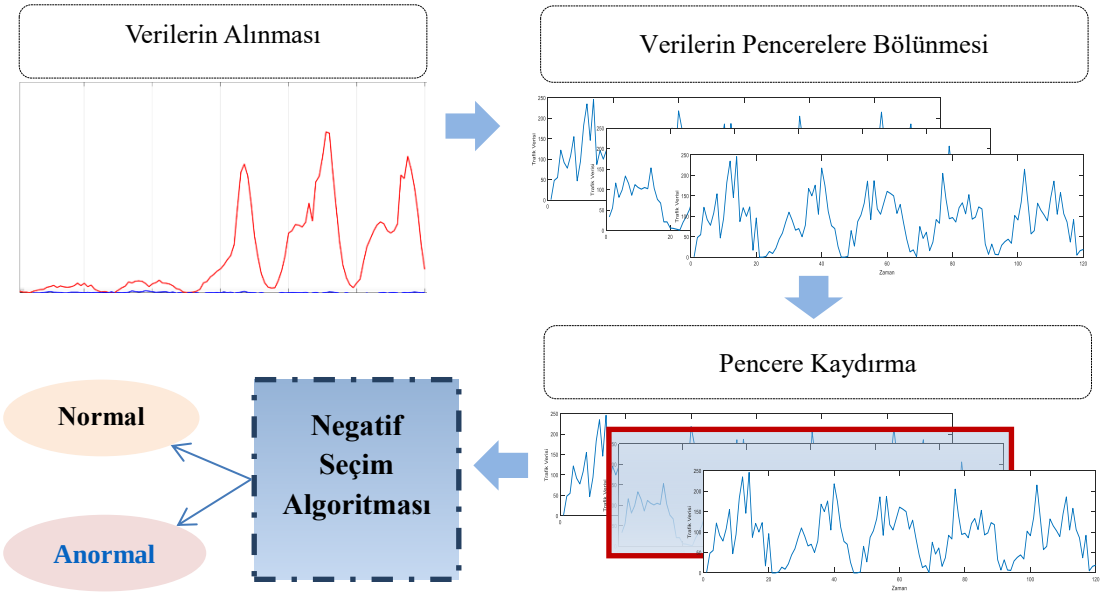
Alkasassbeh vd. [6] tarafından yapılmış olan çalışmada, modern saldırı türlerini içeren yeni bir veriseti toplanmış, toplanan veriler, uygulama ve ağ katmanlarını hedef alan farklı saldırı türleri için kaydedilmiştir. Toplanan veri setine Distributed Denial of Service(Hizmet engelleme) saldırı türlerini sınıflandırmak için üç makine öğrenme algoritması (Multi Layer Perceptron, Random Forest ve Naive Bayes) uygulanmıştır. Multi Layer Perceptron(Çok katmanlı perseptron) sınıflandırıcısının en yüksek doğruluk oranını elde ettiği görülmüştür. Dutt vd.[7] yapmış olduğu çalışmada yapay bağışıklık sistemi kullanılarak network sistemine yapılan kötü niyetli atakların etkili bir şekilde tespit edildiği görülmüştür.Virüs,solucan vb. zararlılarının belirli bir sunucuya belli sayıda bilgisayar tarafından bulaştırılmaya çalışılmış ve testlerin sonucunda gelen anormal trafiğin adedine bağlı olarak sistemin zararlıyı ne

oranda yakalıyıp yakalayamadığı tespit edilmiştir. Aziz vd.[8] tarafından yapılan çalışmada genetik algoritma tarafından oluşturulan dedektörler kullanılarak ağdaki anormal aktiviteyi tespit etmek için bir yaklaşım uygulanmamıştır. Minkowski mesafe fonksiyonu, algılama işlemi için Öklid mesafesine karşı test edilmiştir. Minkowski mesafesinin Öklid mesafesinden daha iyi sonuçlar verdiği ve daha az tim kullanarak çok iyi sonuçlar verdiği gösterilmiştir.

Web sunucularda anormal trafik verilerinin tespiti genellikle bir zaman serisi problemi olarak değerlendirilmektedir. Bu problem sinyal pencereleri üzerinde özellik çıkarılarak sınıflandırma yapılarak değerlendirilmektedir[9]. Ancak web trafik verilerinin genel bir karakterstik örüntü yapısı olmadığından, çözüm yöntemleri de farklılılar içermektedir. Özellikle tüm verilerin internet ortamında kolaylıkla ulaşılabildiği günümüzde belirli saldırılar için kullanılarak zamanında yapılan tespitle bilgi kaybının önüne geçilebilir. Literatürde ağ üzerindeki anormal trafiğin tespiti adına çalışmalar bulunmaktadır. Ancak yapay bağışıklık algoritmalarında negatif seçilim algoritmasını kullanarak durumu tespit eden çok az sayıda çalışma mevcuttur. Bu çalışmada, web trafiklerini gösteren Yahoo Webscope S5[10] veriseti kullanılarak, anormal durum tespiti için pencere kaydırma ile yapay bağışıklık sistemlerinin negatif seçim algoritmasına dayalı bir yöntem önerilmiştir. Ayrıca kullanıcı dostu arayüze sahip bir yazılım da geliştirilmiştir. Bu yazılım ile ağ verilerinin zaman düzlemlerindeki trafik değerleri kullanılarak, ağda oluşan anormal durumların tespiti, trafik değerlerinde hangi zaman adımlarında anormal trafiğin oluştuğunun tespiti başarılı bir şekilde gerçekleştirilmiştir. Çalışmanın sonraki aşamaları şu şekilde organize edilmiştir. İkinci bölümde deneyler ve veriseti ile kullanılan algoritmalar çalışmanın materyal ve metot kısmı olarak sunulmuştur. Üçüncü bölümde deneysel çalışmalar ve elde edilen bulgular detaylı olarak analiz edilmiş ve son bölümde ise sonuçlar tartışılmıştır.

2. Materyal ve Yöntem

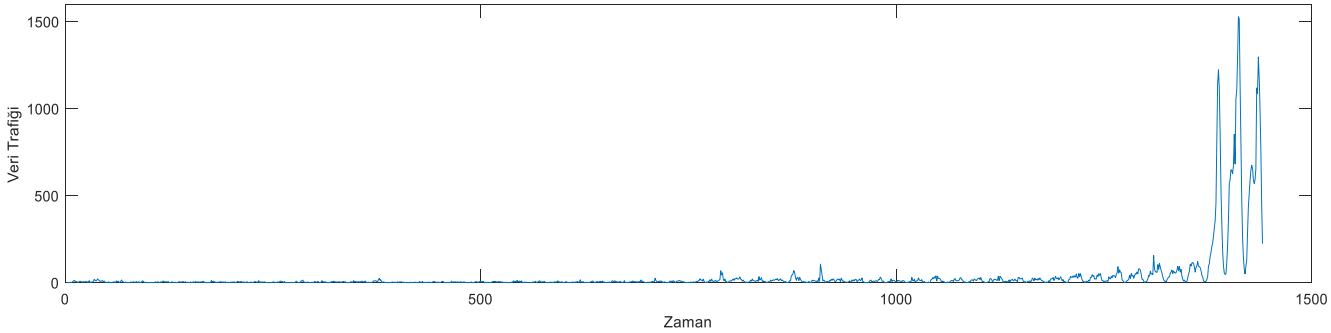
Bu çalışmada önerilen yöntem ile bir boyutlu zaman serileri şeklinde oluşturulan Yahoo Webscope S5[10] veriseti kullanılarak Yapay Bağışıklık Sistemlerinin Negatif Seçim Algoritması kullanılarak anormal web trafik verilerinin tespit edilmesi sağlanmıştır. Önerilen yöntemin tespit yapısını ve akışlarını gösteren açık diyagram Şekil 1'de sunulmuştur. Şekil 1'de görüldüğü gibi ilk önce verisetinden veriler alınır. İkinci adımdabu veriler belli sayıdapencerelere bölünür ve pencere kaydırma işlemi ile her bir pencereye tespit prosedürü uygulanır. NSA algoritması ile eğitim ve test pencereleri belirlendikten sonra, son aşamada aktifleşen dedektörler sayesinde her bir penceredeki anormal web trafik verilerinin tespiti gerçekleştirilir.



Şekil 1. Önerilen YBS-NSA destekli yöntemin tespit yapısı.

2.1. Web Trafik Veriseti

Web trafik verilerinde anormal durumların tespiti genellikle bir zaman serisi sinyali olarak ele alınmaktadır. Bu çalışmada, zaman serisi şeklinde web trafiklerini gösteren Yahoo Webscope S5[10] veriseti kullanılmıştır. Bu veriseti A1, A2, A3 ve A4 olmak üzere toplam dört farklı sınıf ve 367 adet sinyal örüntüsünden oluşmaktadır. Her bir sinyal örüntüsü ortalama 1500 veri içermekte olup toplamda dört farklı sınıfta 5050000 veri bulunmaktadır. A1 sınıfında gerçek veriler bulunmakta iken, diğer sınıflardaki web trafik verileri sentetik olarak oluşturulmuştur. Yahoo Webscope F5 datasetindeki A1 sınıfında 67 adet gerçek web trafiği dosyası bulunmaktadır. Şekil 2'de A1 sınıfında bulunan bir sinyal örüntüsü gösterilmiştir. Bu çalışmada A1 sınıfındaki anormal web trafiğinin tespiti sağlanmıştır. Tablo 1'de Yahoo Webscope F5 datasetindeki verilere ait detaylı bilgiler sunulmuştur.



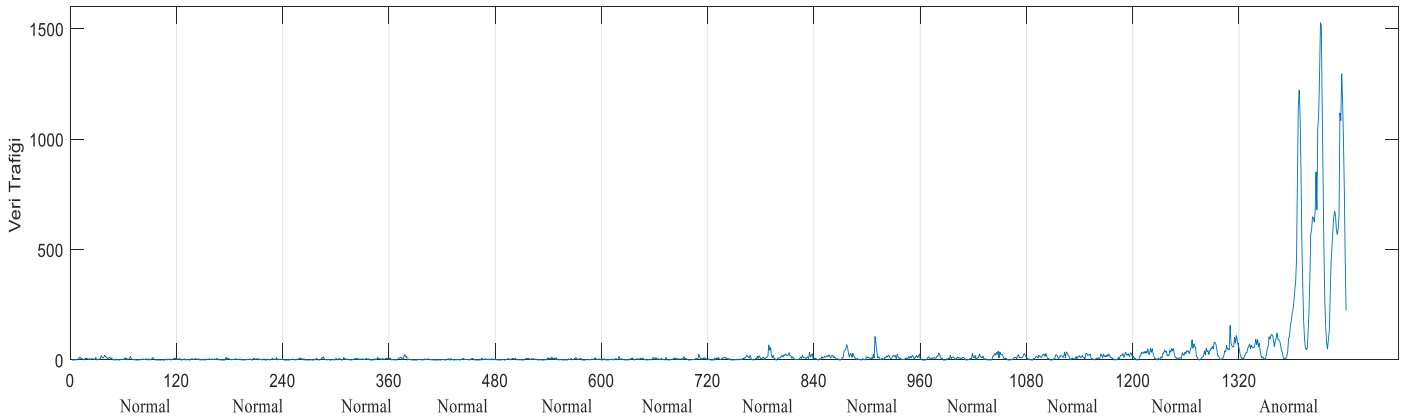
Şekil 2. Yahoo Webscope S5 verisetinden örnek bir sinyal örüntüsü.

Tablo 1. Yahoo Webscope S5 verisetinin detayları.

Sınıf	Gerçek / Sentetik Trafik (G / S)	Toplam veri	Toplam Anomali
A1	G	94866	1669
A2	S	142100	466
A3	S	168000	943
A4	S	168000	837

2.2. Verilerin Pencerelere Bölünmesi

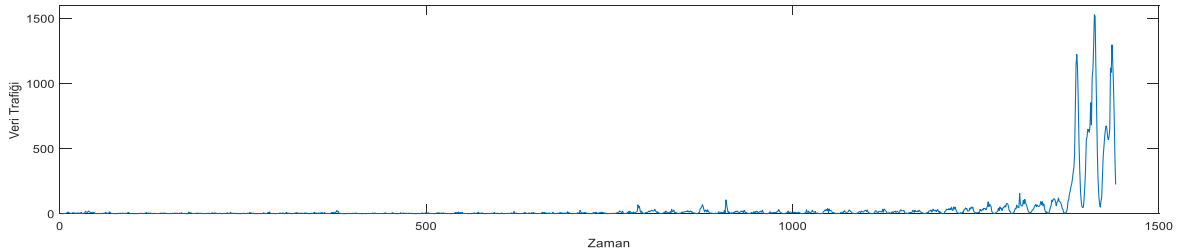
Yahoo S5 veri seti içinde A1 sınıfına ait setler anormal verinin bulunması amacıyla 12 pencereye bölünmüştür. Her pencere zaman dillimi 120 olarak ölçeklendirilmiştir. Buradaki pencerelemenin amacı anormal trafiği bulmak için normal trafik değerlerine ihtiyaç duyulmasıdır. Buradaki pencerelemenin amacı anormal trafiği bulmak için normal trafik değerlerine ihtiyaç duyulmasıdır. Bu pencerelenen normal trafik değerlerini eğitim verisi olarak kullanarak deneylerimizdeki anormal trafiğin bulunması önerilen yöntem ile sağlanmıştır. Şekil 3'te örnek bir sinyal örüntüsünün pencerelere bölünmesi, normal ve anormal trafik verilerinin olduğu alanlar gösterilmiştir.



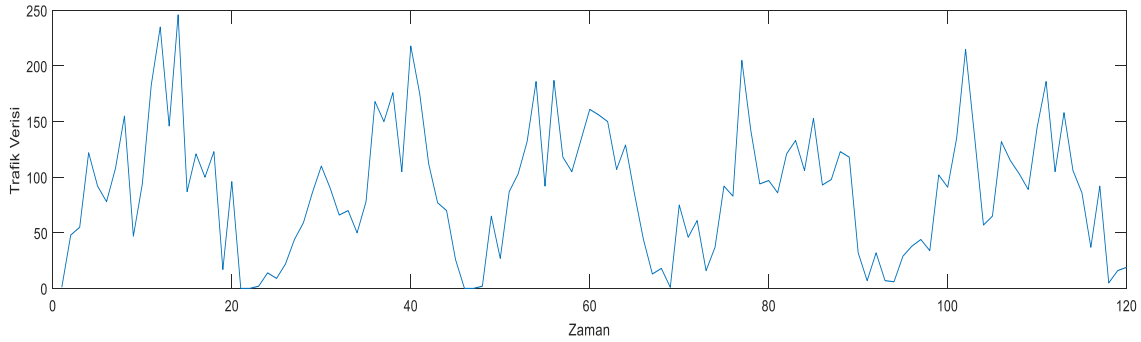
Şekil.3.Verilerin pencerelere bölünmesi

2.3. Pencere Kaydırma

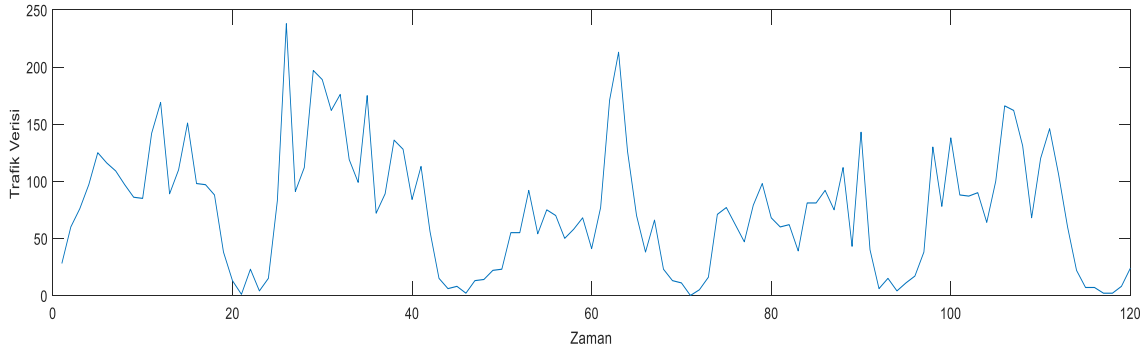
Zaman serisi şeklinde elde edilen sinyal örüntülerinin pencerelere bölünmesi ile zaman serisi verilerinin boyutları ve önemli özellikleri korunarak yeni değerler elde edilir. Burada pencerelere bölmenin temel amacı orijinal zaman serisindeki hata payını azaltmak ve en iyi veriyi elde etmektir. Bu ana yaklaşımda parçalara ayrılan sinyal pencereleri arasında pencere kaydırma yaparak tüm sinyalin işlemden geçirilmesi uygulanır. Çeşitli zaman serisi uygulamaları için hava durumu, finans ve sağlık vb. pencere kaydırma işlemi sıklıkla kullanılmaktadır. İlk değerden itibaren parçalar tanımlanır. İlk oluşturulan parçadan sonra verilen kritere göre sonraki parça işlenir. Bu işlem son parçada işlem yapılıncaya kadar devam eder. Bu metot sezgisel ve basittir. Temel amaç belirli miktarda verilen verinin tüm tahmini hatalarını azaltmaktır [11]. Bu çalışmada, her birisinde 120 tane veri olan pencereler oluşturulmuş ve çoklu segmentlere ayrılmıştır. İlk olarak birinci segment işlenmekte ardından sonraki segmente geçilerek tahmini değerler üretilmiştir. Her segmentte algoritma çalıştırılarak ağ öğrenmesi gerçekleştirilerek sonuç olarak anomali tespiti yapılarak saldırı yapılan alanlar bulunmuştur. Şekil 4(a)'da kullanılmış olan veri setindeki tüm veriler çizdirilmiş, Şekil 4(b), Şekil 4(c), Şekil 4(d), Şekil 4(e)' de ise sliding window(pencere kaydırma) yapılarak veriseti 12 segmente ayrılmış ve bu segmentlerin gösterimi verilmiştir.



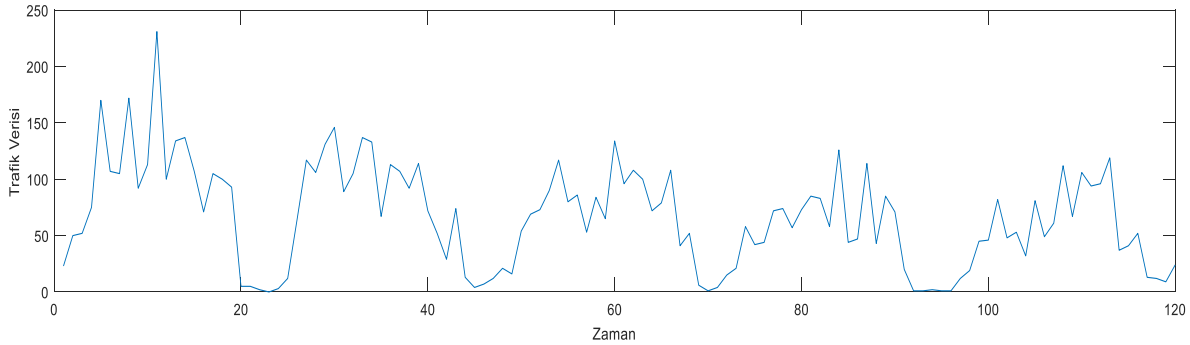
Şekil 4(a). İçerisinde anormal web trafiği olan pencerelere ayrılacak tam bir sinyal örüntüsü.



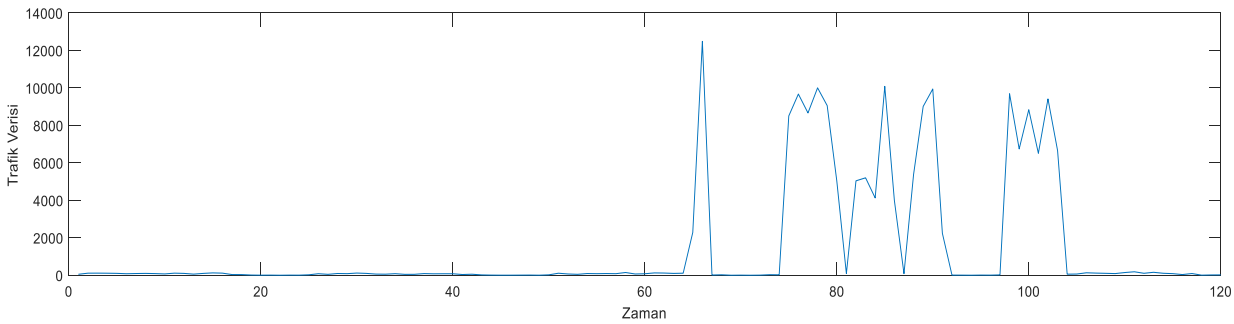
Şekil 4(b). Normal veri bulunan birinci pencereye ait sinyal örüntüsü.



Şekil 4(c). Normal veri bulunan ikinci pencereye ait sinyal örüntüsü.



Şekil 4(d). Normal veri bulunan altıncı pencereye ait sinyal örüntüsü.



Şekil 4(e). Anormal veri bulunan onikinci pencereye ait sinyal örüntüsü.

2.4. Yapay Bağışıklık Sistemleri

Canlılarda bağışıklık sisteminin ana görevi vücuttaki hatalı hücreleri ve yabancı hücre organizmaları araştırmaktır. Omurgalıların bağışıklık sistemi çok çeşitli moleküllerden, hücrelerden ve vücudun her yerine yayılmış organlardan oluşmaktadır. Bağışıklık sisteminin fonksiyonlarını izleyen herhangi bir organ bulunmamaktadır. Vücuda girdiğinde antikor oluşmasına yol açan virüs, bakteri, parazit gibi protein yapısında maddelere antijen adı verilmektedir. Bağışıklık sistemi tanınan antijen ve vücut dışı antijen ayrımını yapabilmelidir. Reseptör moleküller bu kısımda devreye girer. Bu reseptörler B ve T hücreleri olarak iki gruba ayrılır[12]. Bu iki tür hücre aslında oldukça benzer yapıdadır, ancak antijenleri nasıl tanıdıkları ve rollerini nasıl belirledikleri kısmında ayrılırlar. İnsan bağışıklık sisteminde, göğüs kemiğinin arkasında bulunan timus T hücrelerinin olgunlaşmasında önemli bir rol oynar ve bu olgunlaşma sırasında, tanınan antijenlerin tanımlanmasında tüm T hücreleri, T hücre popülasyonundan çıkarılır; bu olaya negatif seçilim adı verilir. Eğer bir B hücresi, kendisiyle özdeş olmayan bir antijenle karşılaşırsa, hafıza ve efektör hücrelere çoğalır ve farklılaşır. Bu olaya ise klonal seçilim adı verilir[12].

Yapay bağışıklık sistemi, gerçek hayattaki bağışıklık sisteminden yola çıkılarak geliştirilmiş genel amaçlı bir sezgisel yöntemdir. Canlılardaki bağışıklık sistemi genel olarak yorumlanıp belirli bir sistematiğe indirgenerek çalışmalarda kullanılmıştır. Günümüzde bu yöntem geliştirilerek çalışmalarda kullanılmaktadır. YBS, insanın bağışıklık sistemi baz alınarak geliştirilmiş yapay zeka alanıdır. Bağışıklık sistemimize herhangi bir zararlı nüfus ettiği zaman sistemin vermiş olduğu tepki modellenerek çalışmalarda anormal durumlar tespit edilmeye çalışılmaktadır. Modellemeler bilgisayar sistemine aktarılmış ve algoritmalar geliştirilmiştir. YBS' nin, Negatif / Pozitif Seçim Algoritması, Klonal Seçim Algoritması, Bağışık Ağ Modelleri, Antikor Ağ Modeli olmak üzere dört ana algoritması bulunmaktadır[13]. Bu çalışmada yapay bağışıklığın Negatif Seçim Algoritması kullanılmıştır.

2.4.1. Negatif Seçim Algoritması

Canlıların bağışıklık sistemlerinde vücuda giren zararlı organizmanın tanınması, kemik iliğinde üretilen iki lenfosit olan B ve T hücreleri ile yapılmaktadır. Kemik iliğinde üretilen bu hücrelerden T hücreleri timüste negatif seçim diye adlandırılan sürece tabi tutulur. Bu hücreler gözlemlenerek elde edilen sonuçlar neticesinde bu algoritma geliştirilmiştir. Bu işlemden esinlenerek geliştirilen NSA algoritmasının işlem adımları aşağıdaki gibi listelenebilir.

i. Öncelikle veriseti içinden bir eğitim kümesi(self-set) belirlenir.

ii. Sonraki adımda rastgele belli sayıda aday detektör üretilir

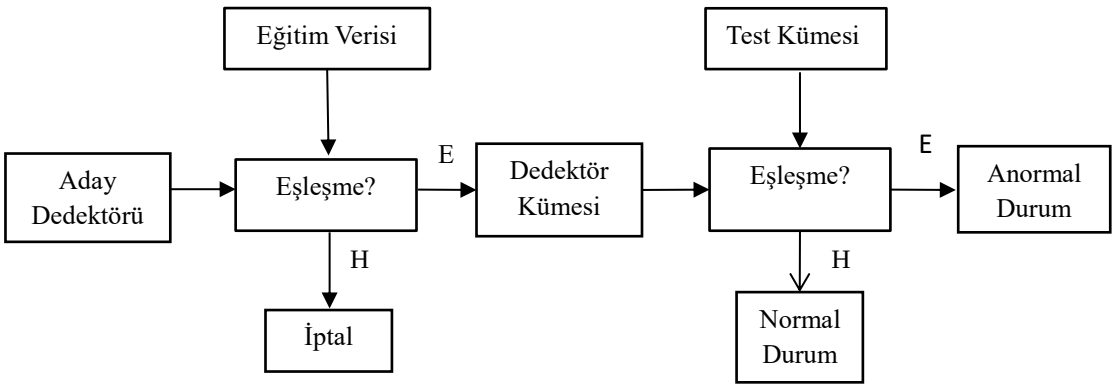
iii. Aday detektörlerden eğitim kümesi ile belirlenen eşik değerine göre eşleşenler aday detektör kümesinden çıkarılır. Eşleşmeyenler detektör kümesine atılarak eğitilir. Bir aday detektör ile eğitim veya test kümesi arasındaki eşleşmenin hesaplanması için Denklem 1' de verilen Öklid (Euclidian) mesafe ölçümü kullanılmıştır. Bu denklemde A bulunan mesafeyi, l data sayısını, Ab test veya eğitim kümesini, Ag ise detektör kümesini belirtmektedir.

$$\boldsymbol{A = \sqrt{\sum_{i=1}^{l} (Ab_i - Ag_i)^2}}$$

iv. Test aşamasında, yine veriseti içerisindentest kümesi(test-set) oluşturulur. Oluşturulan detektör kümesindeki elemanla, test kümesindeki eleman arasında belirlenen eşik değerine göre eşleşme olduğunda anormallik, eşleşme olmadığında normal durum olduğu tespit edilir.

v. Test aşamasının sonuçları gösterilerek işlem sonlandırılır.

NSA'nıngenel olarak işlem akışlarını gösteren şeması Şekil 5'de gösterilmiştir.

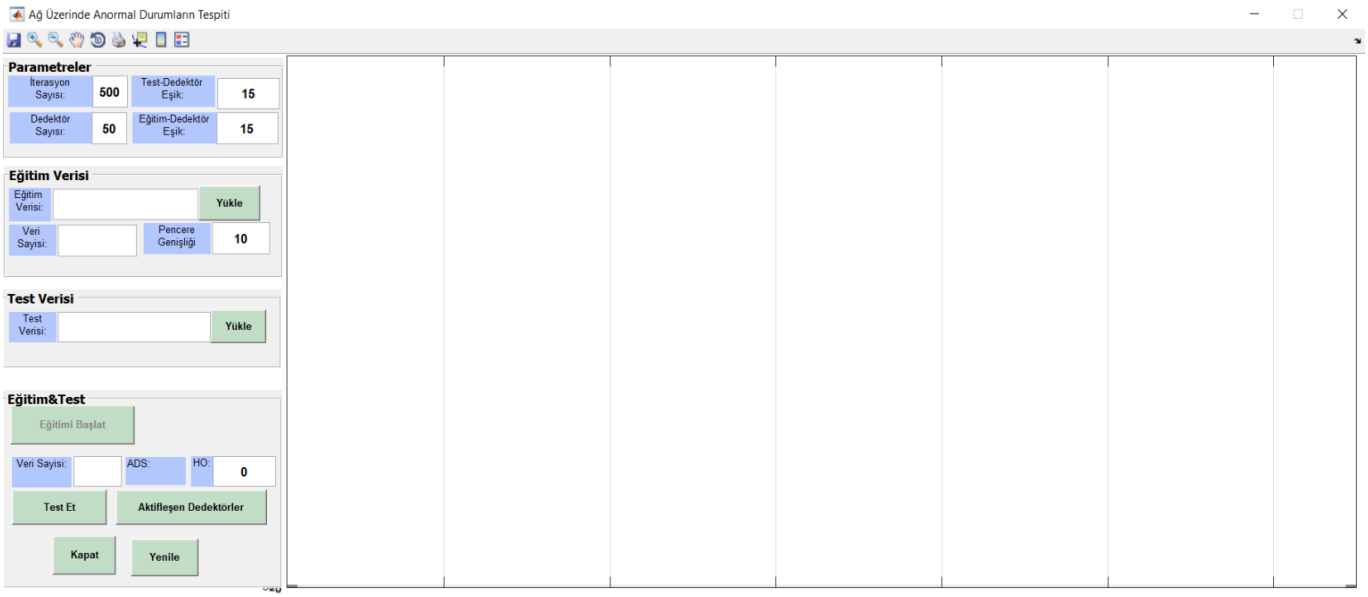


Şekil 5.Negatif Seçim Algoritmasının akış şeması.

3. Geliştirilen Uygulama ve Deneysel Sonuçlar

3.1. Yazılım Altyapısı

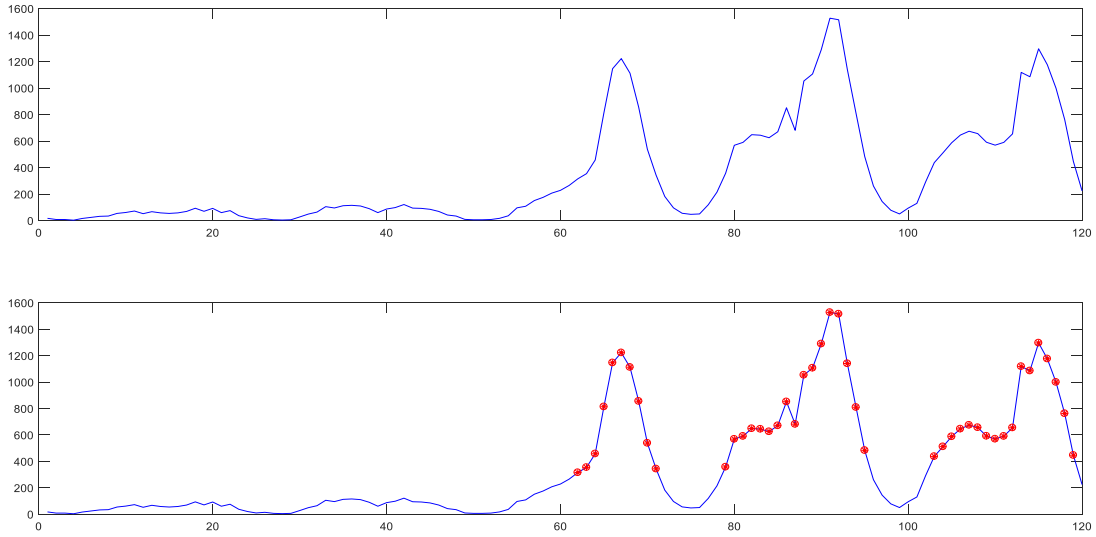
Bu çalışmada ağ üzerinde anormal trafiğin tespit edilmesi için Matlab ile tasarlanan kullanıcı arayüzü Şekil 6' da gösterilen bir yazılım geliştirilmiştir. Eğitim ve test verileri tanıtıldıktan sonra test sonuçları yine aynı arayüz üzerinden görüntülenip grafiksel değerlendirmeler yapılabilmektedir. Yazılımda, ilk önce eğitim verisi kısmına pencereleme yapılan normal değer olarak kabul edilen veriler yüklenir. Test kısmına ise yine pencereleme yapılan ve anormal olarak kabul edilen veriler yüklenir. Eğitimi başlat butonuna basıldıktan sonra normal olan veriler ile eğitim aşaması gerçekleşir. Daha sonra test et butonuna ve aktifleşen dedektör butonuna tıklandıktan sonra uygulamanın anormal trafik için ürettiği veriler ekranda gösterilmektedir.



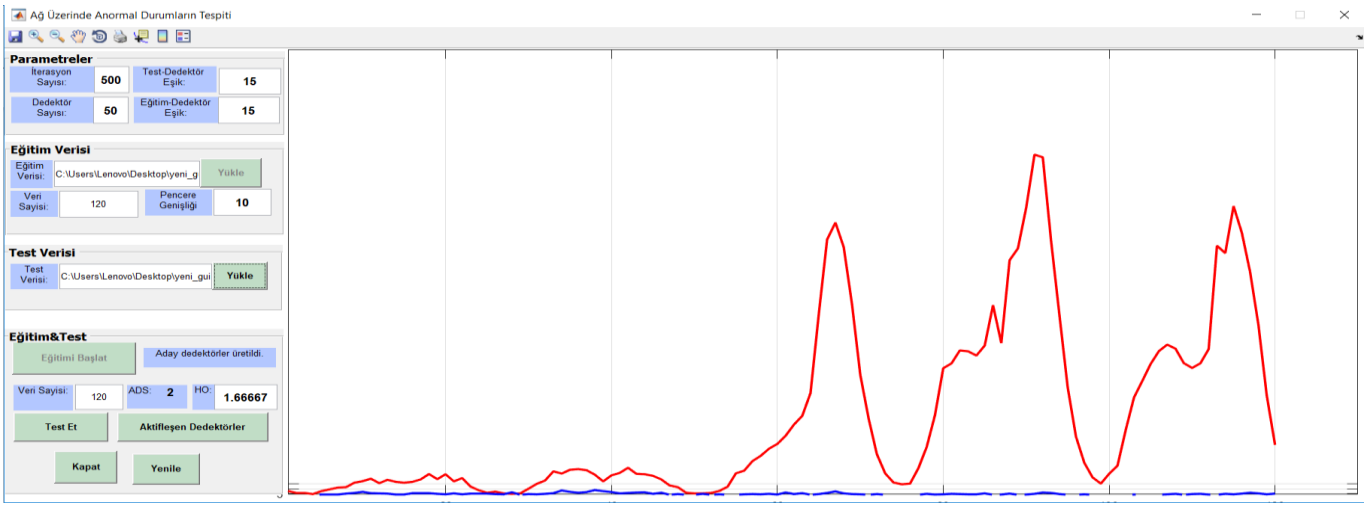
Şekil 6. Web trafik verilerinde anomali tespiti için geliştirilen yazılımın arayüzü

Çalışması kapsamında yapılan deneylerde YBS parametreleri belirlendikten sonra daha önceden ayarlanmış olan eğitim ve test verilerine göre ağda oluşmuş olan anomali yazılım tarafından tespit edilmektedir. Ağda anormal trafik verilerinde oluşan hata yüzdesi ve bu aşamada gerçekleşen aktifleşen detektör sayısına göre belirlenir. Uygulamanın doğru sonuç verdiğini göstermek adına Yahoo Wenscope S5 verisetinde bulunan gerçek web verileri üzerinde deneysel çalışmalar yapılmıştır. Bu deneylerde tablolarda birinci 120'lik veri P1 ikinci 120'lik veri P2,üçüncü 120'lik veri P3, dördüncü 120'lik veri P4, beşinci 120'lik veri P5, altıncı 120'lik veri P6, yedinci 120'lik veri P7, sekizinci 120'lik veri P8, dokuzuncu 120'lik veri P9, onuncu 120'lik veri P10, on birinci 120'lik veri P11, on ikinci 120'lik veri P12 olarak adlandırılmıştır. Burada toplam veri parametresi bölünmüş pencerelerin boyutunu temsil etmektedir. Bu deneylerde anormal trafiğin olduğu pencere ile yapılan deneyler dışındaki tüm deneyler normal trafik verisiyle yapıldığı için Anomali sayısı, aktifleşen dedektör sayısı ve doğru bulunan anomali sayısı 0 olur ve doğruluk oranı %100 olarak hesaplanır. Bu deneyler için eğitim eşik ve test eşik değerleri ile en uygun sonuçlar hesaplanmıştır.

Yahoo Webscope veriseti üzerinde ilk deneysel çalışma, verisetinde bulunan A1 gerçek veri sınıfına ait olan 42. sinyal örüntüsü ile gerçekleştirilmiştir. Şekil 7'de 42. veriye ait anormal trafik veri penceresi gösterilmiştir. Burada segmentlere ayrılmış 42. veri setinin son segmenti yani anormal trafiğin olduğu ve test verisi olarak kullanılan verilerden kırmızı yuvarlak ile işaretlenmiş olanlar anormal verileri göstermektedir. Birinci test işlemi 42. veri ile yapılan deneyi içermektedir.Anormalliklerin belirlenmesi için ilk aşama, Şekil 8'deki gibi gerçekleştirilen sisteme eğitim ve test verilerin yüklenmesidir. İkinci aşamada NSA algoritmasının parametrelerinin değerleri tanımlanır. Şekil 8'de her biri 120 parçaya bölünmüş pencereler kullanılmış olup normal veriler eğitim verileri olarak (P1 penceresi), anormal veriler ise test verisi (P12 penceresi) olarak yazılıma yüklenmiştir

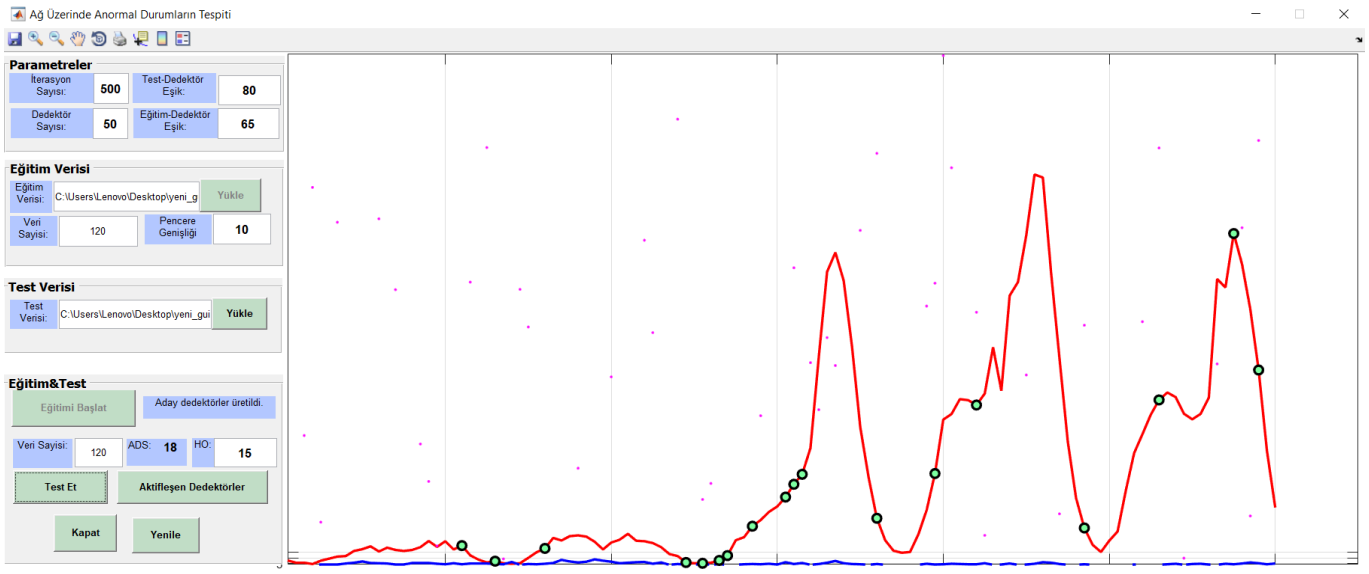


Şekil 7. Anormal web trafik verisinin bulunduğu sinyal örüntü penceresi



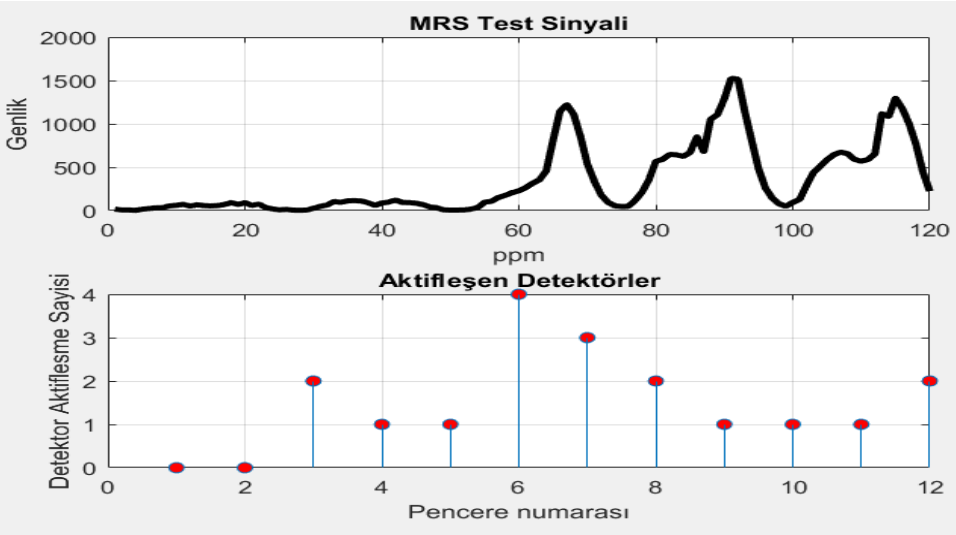
Şekil 8. Uygulama yazılımına eğitim ve test verilerinin yüklenmesi

Şekil 9'daki test işleminde eğitim verisi olarak 42. verinin ilk 120 verisi kullanılmış, test verisi olarak 42. verinin son 120 verisi yani anormal veriler kullanılmıştır. İterasyon sayısı 500, dedektör sayısı 50, eğitim dedektör eşik sayısı 65, test dedektör eşik sayısı ise 80 olarak belirlenmiştir.



Şekil 9. Test verisi üzerinde anormal trafik verilerinin NSA ile tespiti

Şekil 10'da anormal trafiğin olduğu veriler yani test verileri grafiğe dökülmüş, ayrıca aktifleşen dedektör sayısı yine grafiksel olarak verilmiştir. Buradan da görülebileceği gibi sinyal üzerinde anomalinin fazla olduğu bölgelerde aktifleşen dedektör sayısı da artmaktadır.



Şekil 10. Anormal web verilerinin bulunduğu sinyalde aktifleşen detektörler

3.2. Deneysel Sonuçlar

Gerçekleştirilen çalışmada anormal trafiği yüksek oranda tespit ettiğini göstermek adına 42. ve 58. veriler için yapılan deneysel çalışmalarla bulunan anormal trafik tespit edilmeye çalışılmış ve gösterilmiştir. Bu deneylerde tablolarda birinci 120'lik veri P1 ikinci 120'lik veri P2,üçüncü 120'lik veri P3, dördüncü 120'lik veri P4, beşinci 120'lik veri P5, altıncı 120'lik veri P6, yedinci 120'lik veri P7, sekizinci 120'lik veri P8, dokuzuncu 120'lik veri P9, onuncu 120'lik veri P10, on birinci 120'lik veri P11, on ikinci 120'lik veri P12 olarak adlandırılmıştır. Burada toplam veri parametresi bölünmüş pencerelerin boyutunu temsil etmektedir. Bu deneylerde anormal trafiğin olduğu pencere ile yapılan deneyler dışındaki tüm deneyler normal trafik verisiyle yapıldığı için Anomali sayısı, aktifleşen dedektör sayısı ve doğru bulunan anomali sayısı 0 olur ve doğruluk oranı %100 olarak hesaplanır. Bu deneyler için eğitim eşik ve test eşik değerleri ile en uygun sonuçlar hesaplanmıştır. Tablo 2'de 42. veri için normal ve anormal trafiğin tespiti için yapılan testler neticesinde eğitim verisiyle normal ve anormal trafik değerleri teste tabi tutulmuştur. Anormal trafik ile yapılan deneyde toplam anomali sayısı 44 iken aktifleşen dedektör sayısı 18, doğru bulunan anomali sayısı 41 ve doğruluk oranı %93.18 olarak bulunmuştur.

Tablo 2. 42. veri için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçların ve kullanılan parametrelerin detaylı analizi

Doğruluk Matrisi	Eğitim Verisi	Test Verisi	Toplam Veri	Anomali Sayısı	Aktf. Det. Say	Doğru Bul. Ano. Say.	Doğruluk Oranı	Eğitim Eşik Değeri	Test Eşik Değeri
Veri No:42	P1	P2	120	0	0	0	100	65	80
	P1	P3	120	0	0	0	100	65	80
	P1	P4	120	0	0	0	100	65	80
	P1	P5	120	0	0	0	100	65	80
	P1	P6	120	0	0	0	100	65	80
	P1	P7	120	0	0	0	100	65	80
	P1	P8	120	0	0	0	100	65	80
	P1	P9	120	0	0	0	100	65	80
	P1	P10	120	0	0	0	100	65	80
	P1	P11	120	0	0	0	100	65	80
	P1	**P12**	**120**	**44**	**18**	**41**	**93.18**	**65**	**80**

Tablo 3'te 58. Veri üzerinde normal ve anormal trafiğin tespiti için yapılan testler neticesinde eğitim verisiyle normal ve anormal trafik değerleri teste tabi tutulmuştur. Anormal trafik ile yapılan deneyde toplam anomali sayısı 43 iken aktifleşen dedektör sayısı 22, doğru bulunan anomali sayısı 41 ve doğruluk oranı %95.34 olarak bulunmuştur.

Tablo 3. 58. veri için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçların ve kullanılan parametrelerin detaylı analizi

Doğruluk Matrisi	Eğitim Verisi	Test Verisi	Toplam Veri	Anomali Sayısı	Aktf. Det. Say	Doğru Bul. Ano. Say.	Doğruluk Oranı	Eğitim Eşik Değeri	Test Eşik Değeri
Veri No:58	P1	P2	120	0	0	0	100	40	60
	P1	P3	120	0	0	0	100	40	60
	P1	P4	120	0	0	0	100	40	60
	P1	P5	120	0	0	0	100	40	60
	P1	P6	120	0	0	0	100	40	60
	P1	P7	120	0	0	0	100	40	60
	P1	P8	120	0	0	0	100	40	60
	P1	P9	120	0	0	0	100	40	60
	P1	P10	120	0	0	0	100	40	60
	P1	P11	120	0	0	0	100	40	60
	P1	**P12**	**120**	**43**	**22**	**41**	**95.34**	**40**	**60**

4. Tartışma ve Sonuçlar

Siber saldırıların tespiti ve alınacak önlemlerin neler olabileceği hususu hala tartışılmaya devam etmektedir. Günümüzde veri sınıflandırması ve verilerin gerçekten zararlı olup olmadığının belirlenmesi gerçekten önem arz etmektedir. Dolayısıyla öncelikle tespit etme ve sonrasında ise aksiyon alma şirket ve/veya kişilerin verilerinde herhangi bir kayıp olmaması için önemlidir. Bu çalışmada, ağ üzerindeki anormal web trafiklerinin tespiti için Yapay Bağışıklık Sistemlerinin Negatif Seçim Algoritmasına (NSA) dayalı bir yaklaşım önerilmiş ve tespit işleminin gerçekleştirilebilmesi için bir uygulama yazılımı geliştirilmiştir.

Web trafiği için Yahoo Webscope S5 verisetinde bulunan gerçek veriler kullanılmış ve pencere kaydırma yöntemi kullanılarak veriler pencerelere ayrılmıştır. Yapılan deneysel çalışmalarda, web trafik verilerinde oluşan anormal trafik verilerinin tespiti, NSA'nın yapısında bulunan aktifleşen detektör sayılarındaki değişimin izlenmesi ile sağlanmıştır. Negatif Seçilim Algoritması kullanılarak yapılan deneyler sonucunda yüksek doğruluk oranına ulaşılmıştır.Ayrıca kayan pencere yöntemini kullanarak setler 12 segmente ayrılmış, normal ve anormal trafik üzerinde deneyler yapılmıştır. Deneysel çalışmalar kapsamında verisetinde bulunan ve gerçek veriler olan 42. ve 58. olmak üzere zaman serileri şeklinde sunulan sinyal verileri kullanılmıştır Önerilen NSA destekli yöntem ile bir web trafik verisi içerisindeki anomalileri doğru bulma konusunda 42. veri için %93.18, 58. veri için ise %95.38 başarım elde edildiği görülmüştür.

Bu çalışmada kullanılan veri setinde verilerde oluşan anomalinin belirlenmesi için kullanılan veriler farklı zamanlarda alınmıştır. Gerçekleştirilen yazılım ile veriler incelendiğinde saldırı algılandığında zaman anomali trafiği oluşmaktadır. Bu durumda YBS yazılımında aktifleşen detektör ve bulunan hata yüzdesinde bir artışa neden olmaktadır. Verilerinin bu davranışı incelenerek YBS ile anormal trafik kolay bir şekilde belirlenebilmektedir.

Daha sonraki çalışmalarda çalışmanın perspektifi büyültülerek ağ katmanlarının özelliklerine göre veriler elde edilerek yapay bağışıklık sisteminin Negatif Seçilim Algoritması uygulanarak anormal trafiğin tespiti yapılabilir. Ayrıca veri sınıflandırılmasında ve uygulamada daha seçici özellikler eklenerek uygulama geliştirilebilir. Genel olarak atakların analizi yapılarak daha fazla atak tipinde anormal trafiğin tespiti yapılabilir.

5. Teşekkür

Bu çalışmanın yazarları, çalışmada deneysel testler için kullanılan Yahoo Webscope S5 Veriseti kullanımına izin verdiği için Yahoo şirketine teşekkürlerini sunmaktadır.

References

[1] Kim, T.-Y. and Cho S.-B. (2018). Web traffic anomaly detection using C-LSTM neural networks, Expert Systems with Applications, 106, pp. 66-76.

[2] Symantec. Symantec Internet Security Threat Report, https://www.symantec.com/content/dam/symantec/docs/reports/istr-24-executive-summary-en-apj.pdf, Erişim Tarihi:09.06.2019.

[3] Münz, G., Li, S., and Carle, G. (2007). Traffic anomaly detection using k-means clustering, GI/ITG Workshop MMBnet, pp. 13-14.

[4] Thill, M., Konen, W., and Bäck, T. (2017). Online anomaly detection on the webscope S5 dataset: A comparative study. Evolving and Adaptive Intelligent Systems (EAIS), pp. 1-8.

[5] Akba, E. and Ergen, B. (2019). Kablosuz Yerel Alan Ağlarında Yapay Bağışıklık Sistemi ile Saldırı Tespiti ve Performans Analizi, http://www.emo.org.tr/ekler/8947fab05bee9c5_ek.pdf , Erişim Tarihi: 09.06.2019.

[6] Alkasassbeh, M., Al-Naymat, G., Hassanat, A., and Almseidin, M. (2016). Detecting distributed denial of service attacks using data mining techniques. International Journal of Advanced Computer Science and Applications, 7, pp. 436-445.

[7] Dutt, I., Borah, S., and Maitra, I. (2016). Intrusion Detection System using Artificial Immune System. International Journal of Computer Applications,144.

[8] Aziz, A. S. A., Salama, M. A., ella Hassanien, A., and Hanafi, S. E.-O. (2012). Artificial immune system inspired intrusion detection system using genetic algorithm Informatica. 36.

[9] Zheng, Y., Liu, Q., Chen, E., Ge, Y., and Zhao, J. L. (2014). Time series classification using multi-channels deep convolutional neural networks. International Conference on Web-Age Information Management, pp. 298-310.

[10] Computing Systems Data. S5-A Labeled Anomaly Detection Dataset. https://webscope.sandbox.yahoo.com, Erişim Tarihi: 09.06.2019.

[11] Yahmed, Y. B., Bakar, A. A., Hamdan, A. R., Ahmed, A., and Abdullah, S. M. S.(2015). Adaptive sliding window algorithm for weather data segmentation. Journal of Theoretical and Applied Information Technology, 80, p. 322.

[12] De Castro, L. N., and Timmis, J. (2002). Artificial immune systems: a new computational intelligence approach. Springer Science & Business Media.

[13] Dandıl, E. and Güngör, O. (2012). Yapay Bağışıklık Algoritmaları ile CNC Kesici Takım Aşınmalarındaki Değişimin Belirlenmesi. Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu(ASYU 2012), Trabzon.