BİLECİK
ŞEYH EDEBALİ ÜNİVERSİTESİ

Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

WEB TRAFİK VERİLERİNDE YAPAY BAĞIŞIKLIK ALGORİTMALARI İLE ANOMALİ TESPİTİ

Kadir İLHAN
Yüksek Lisans

Tez Danışmanı
Dr. Öğr. Üyesi Emre DANDIL

BİLECİK, 2019
Ref. No:10278133

BİLECİK
ŞEYH EDEBALİ ÜNİVERSİTESİ

Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

WEB TRAFİK VERİLERİNDE YAPAY BAĞIŞIKLIK ALGORİTMALARI İLE ANOMALİ TESPİTİ

Kadir İLHAN
Yüksek Lisans

Tez Danışmanı
Dr. Öğr. Üyesi Emre DANDIL

BİLECİK, 2019

BILECIK
SEYH EDEBALI UNIVERSITY

Graduate School of Sciences
Department of Computer Engineering

ANOMALY DETECTION IN WEB TRAFFIC USING ARTIFICIAL IMMUNE ALGORITHMS

Kadir İLHAN
Master's Thesis

Thesis Advisor
Asst. Prof.Dr. Emre DANDIL

BİLECİK, 2019

BİLECİK ŞEYH EDEBALİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS
JÜRİ ONAY FORMU

Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulunun 10/07/2019 tarih ve 36-04 sayılı kararıyla oluşturulan jüri tarafından 29/07/2019 tarihinde tez savunma sınavı yapılan Kadir İLHAN'ın "Web Trafik Verilerinde Yapay Bağışıklık Algoritmaları ile Anomali Tespiti" başlıklı tez çalışması Bilgisayar Mühendisliği Anabilim Dalında YÜKSEK LİSANS tezi olarak oy birliği ile kabul edilmiştir.

JÜRİ

ÜYE
(TEZ DANIŞMANI) : Dr. Öğr. Üyesi Emre DANDIL

ÜYE : Prof. Dr. Ecir Uğur KÜÇÜKSİLLE (JÜRİ BAŞKANI)

ÜYE : Dr. Öğr. Üyesi Süleyman UZUN

ONAY

Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulunun …./…./…..... tarih ve ……/…… sayılı kararı.

İMZA/ MÜHÜR

TEŞEKKÜR

Çalışmalarımı yönlendiren, araştırmalarımın her aşamasında bilgi, öneri ve yardımlarını esirgemeyen danışman hocam sayın Dr. Öğr. Üyesi Emre DANDIL'a, veri setini bizimle paylaşan Yahoo Webscope ve çalışmalarımda her zaman yanımda olan ve manevi desteklerini hiçbir zaman esirgemeyen aileme sonsuz teşekkür ederim.

Kadir İLHAN

BEYANNAME

Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kılavuzu'na uygun olarak hazırladığım bu tez çalışmasında, tez içindeki tüm verileri akademik kurallar çerçevesinde elde ettiğimi, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun olarak sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunulduğunu, tezde yer alan verilerin bu Üniversite veya başka bir üniversitede herhangi bir tez çalışmasında kullanılmadığını beyan ederim.

…../…./ 2019

Kadir İLHAN

ÖZET

Günümüzde internet dünyasında farklı türdeki tehditler ve saldırılar artarak devam etmekte ve buna paralel olarak alınan güvenlik önlemlerinde de önemli gelişmeler olmaktadır. Ağ ve web trafiğinde artan kullanıcı sayısı ile paylaşılan veri miktarında meydana gelen ciddi güvenlik zafiyetleri, doğrudan veri sızıntısına sebep olabilmektedir. Bu veri sızıntılarının büyük oranda önlenmesi, çözülmesi gereken önemli bir sorun haline gelmiştir. Özellikle bu alanda yapılan çalışmalar göz önünde bulundurularak hata tespitinin insan hata toleransından çıkarılarak bir sistematiğe bağlanması ve önlem alınması önem taşımaktadır. Bu nedenle, 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 tez çalışmasında, 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. Tez çalışması kapsamında önerilen NSA destekli yöntem ile bir web trafik verisi içerisindeki anomalileri doğru bulma konusunda ortalama %94.30, genel sınıflandırma oranında ise ortalama %97.69 başarım elde edildiği görülmüştür.

Anahtar Kelimeler: Ağ Güvenliği, Web Trafik Verileri, Anomali Tespiti, Yapay Bağışıklık Sistemleri, Negatif Seçim Algoritması.

ABSTRACT

In recent years, different types of threats and attacks continue to increase in the internet world. There are also important developments in the security measures as a result of this situation. Increased number of users in network and web traffic and serious security vulnerabilities in the amount of shared data can directly lead to data leak. Preventing these data leaks to a large extent has become an important problem to solve. In particular, considering the studies conducted in this field, it is important to take error detection out of human error tolerance and connect it to a systematic and take precautions. Therefore, the rapid and accurate detection and prevention of abnormal changes in the rate of online visitors and web traffic data shown as time series is of great importance. 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 removing the feature. In this thesis, 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 contained in the Yahoo Webscope S5 dataset is used and the data is split into windows using the window sliding method. In the experimental studies, the detection of abnormal traffic data in the web traffic data is realized by monitoring the changes in the number of activated detectors in the structure of the NSA. It is observed that the average performance of finding anomalies in a web traffic data is 94.30% and the overall classification rate is 97.69%.

Key Words: Network Security, Web Traffic Data, Anomaly Detection, Artificial Immune Systems, Negative Selection Algorithm

İÇİNDEKİLER

Sayfa No

BEYANNAME ..... **i**
ÖZET ..... **I**
ABSTRACT ..... **II**
İÇİNDEKİLER ..... **III**
SİMGELER ve KISALTMALAR ..... **V**
ÇİZELGELER DİZİNİ ..... **VII**
ŞEKİLLER DİZİNİ ..... **VIII**
1. GİRİŞ **1**
1.1 Literatür Araştırmaları 4
1.2 Tezin Amacı 7
1.3 Hipotez 7
2. WEB GÜVENLİĞİ ve SALDIRI TİPLERİ **8**
2.1 SQL Saldırısı (SQL Injection) 10
2.2 XSS Saldırısı 10
2.3 Yanlış Güvenlik Yapılandırması 11
2.4 Yetersiz Saldırı Kontrolü 12
2.5 XML Harici Girişler 13
2.6 Hassas Veri Sızıntısı 14
2.7 Güvensiz Serileştirme 15
2.8 Bozuk Kimlik Doğrulama 15
2.9 Yetersiz Kayıt ve İzleme 16
2.10 DDOS Saldırıları 17
3. MATERYAL ve YÖNTEM **18**
3.1. Web Trafik Veriseti 18
3.2. Verilerin Pencerelere Bölünmesi 19
3.3. Pencere Kaydırma 20
3.4. Yapay Bağışıklık Sistemleri(YBS) 24
3.4.1. Negatif Seçim Algoritması 27
4. GELİŞTİRİLEN UYGULAMA ve DENEYSEL ÇALIŞMALAR **31**
4.1. Deneysel Çalışma 1 32
4.2. Deneysel Çalışma 2 35

4.3. Deneysel Çalışma 3 39
4.4. Deneysel Çalışma 4 42
4.5. Deneysel Çalışma 5 45
4.6. Deneysel Çalışma 6 47
4.7. Deneysel Çalışma 7 51
4.8. Deneysel Çalışmalar Üzerinde Genel Değerlendirmeler 53
5. TARTIŞMA ve SONUÇLAR 55
KAYNAKLAR 57
ÖZGEÇMİŞ 61

SİMGELER ve KISALTMALAR

Simgeler

A :Öklid mesafe ölçüm denkleminde mesafe

l :Öklid mesafe ölçüm denkleminde veri sayısı

Ab :Öklid mesafe ölçüm denkleminde eğitim veya test kümesi

Ag :Öklid mesafe ölçüm denkleminde dedektör kümesi

Kısaltmalar

ABIDS :Anomaly Based Intrusion Detection System/Anomali Tabanlı Saldırı Tespit Sistemi

ACK : Acknowledgement/Alındı

AIS :Artificial Immune System/Yapay Bağışıklık Sistemi

API :Application Programming Interface/Uygulama Programlama Arayüzü

ARPANET :Advanced research projects agency network/Gelişmiş Araştırma Projeleri Dairesi Ağı

BBN :Bolt Beranek and Newman

CNN :Convolutional Neural Network/Evrişimli Sinir Ağı

C-LSTM :CNN, DNN ve LSTM'in birleşimi

CSI :Computer Science Institute/Bilgisayar Güvenliği Enstitüsü

DOS :Denial of Server/Hizmet Engelleme Saldırısı

DDOS :Distributed Denial of Service/Dağıtık Hizmet Engelleme Saldırısı

DISA :Defense Information Systems Agency/Savunma Bilgi Sistemleri Ajansı

DNN :Deep Neural Network/Derin Sinir Ağı

FBI :Federal Bureau of Investigation/Federal Soruşturma Bürosu

FDCC :Federal Desktop Core Configuration /Federal Masaüstü Çekirdek Konfigürasyonları

HTTP :Hyper Text Transfer Protocol/Hiper Metin Transfer Protokolü

HTTPS :Hyper Text Transfer Protocol Secure/Güvenli Hiper Metin Transfer Protokolü

HTML :Hyper Text Markup Language/ Hiper Metin İşaret Dili

IDS :Intrusion Detection System/Saldırı Tespit Sistemi

IIS :Internet Information Service/İnternet Bilgi Servisi

IP	:Internet Protocol/Internet Protokolü
IPSec	:Internet Protocol Security/İnternet Protokol Güvenliği
LSTM	:Long Short Term Memory / Uzun Kısa Vadeli Hafıza Ağları
MAC	:Media Access Control/Ortam Erişim Kontrolü
MIT	:Massachusetts Institute of Technology
MLP	:Multi Layer Sensor/Çok Katmanlı Algılayıcı
NCP	:Network Control Program/Ağ kontrol Programı
NIST	:National Institute of Standards and Technology/Uluslararası Teknoloji ve Standart Enstitüsü
NSA	:Negative Selection Algoritm/Negatif Seçim Algoritması
OMB	:Yönetim ve Bütçe Ofisi
OSI	:Open Systems Interconnection/Açık Sistem Ara Bağlantısı
OWASP	:The Open Web Application Security Project/Açık Web Uygulama Güvenliği Projesi
PHP	:Hypertext Preprocessor/Hiper Metin Önişlemcisi
SOAP	:Simple Object Access Protocol/Basit Nesne Erişim Protokolü
SSL	:Secure Socket Layer/Güvenli Soket Katmanı
SQL	:Structured Query Language/Yapılandırılmış Sorgu Dili
STIG	:Güvenlik Teknik Uygulama Kılavuzları
SYN	:Synchronize/Senkronize
Tbps	:Terabips
TCP/IP	:Transmission Control Protocol/Internet Protocol-İletişim kontrol protokolü / internet protokolü
TLS	:Transport Layer Security/Taşıma Katmanı Güvenliği
UC	:Universty of California/Kaliforniya Üniversitesi
URL	:Uniform Resource Locator/Standart Kaynak Bulucu
VPN	:Virtual private network/Sanal Özel Ağ
YBS	: Yapay Bağışıklık Sistemleri
WAF	:Web Application Firewall/Web Uygulama Güvenlik Duvarı
XML	:Extensible Markup Language/Genişletilebilir İşaretleme Dili
XSS	:Cross Side Scripting/Siteler Arası Komut Dosyası Çalıştırma

ÇİZELGELER DİZİNİ

Sayfa No

Çizelge 3.1. Yahoo Webscope S5 verisetinin detayları. 19

Çizelge 4.1. Deneysel Çalışma 1 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçları ve kullanılan parametrelerin değerleri 34

Çizelge 4.2. Deneysel Çalışma 1 için gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi 35

Çizelge 4.3. Deneysel Çalışma 2 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçları ve kullanılan parametrelerin değerleri 38

Çizelge 4.4. Deneysel Çalışma 2 için 58. sinyal üzerinde gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi 38

Çizelge 4.5. Deneysel Çalışma 3 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçları ve kullanılan parametrelerin değerleri 41

Çizelge 4.6. Deneysel Çalışma 3 için 17. sinyal üzerinde gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi 42

Çizelge 4.7. Deneysel Çalışma 4 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçların ve kullanılan parametrelerin değerleri 44

Çizelge 4.8. Deneysel Çalışma 4 için 22. sinyal üzerinde gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi 44

Çizelge 4.9. Deneysel çalışma 5 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçları ve kullanılan parametrelerin değerleri 47

Çizelge 4.10. Deneysel çalışma 5 için 25. sinyal üzerinde gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi 47

Çizelge 4.11. Deneysel Çalışma 6 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçları ve kullanılan parametrelerin değerleri 50

Çizelge 4.12. Deneysel çalışma 6 için 40. sinyal üzerinde gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi 50

Çizelge 4.13. Deneysel Çalışma 7 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçların ve kullanılan parametrelerin değerleri 53

Çizelge 4.14. Deneysel Çalışma 7 için 67. sinyal üzerinde gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi 53

Çizelge 4.15. Deneysel çalışmalar ile elde edilen sonuçların özet tablosu 54

ŞEKİLLER DİZİNİ

Sayfa No

Şekil 2.1. Güvenlik yapılandırma yönetimi(Lumension,2009) 12

Şekil 2.2. Hassas veri araştırılması (F5 Networks,2017) 15

Şekil 3.1. Önerilen YBS-NSA destekli yöntemin tespit yapısı. 18

Şekil 3.2. Yahoo Webscope S5 verisetinden örnek bir sinyal örüntüsü..... 19

Şekil 3.3. Web trafik verilerin pencerelere bölünmesi 20

Şekil 3.4. İçerisinde anormal web trafiği olan pencerelere ayrılacak tam bir sinyal örüntüsü..... 21

Şekil 3.5. Normal veri bulunan birinci pencereye ait sinyal örüntüsü 21

Şekil 3.6. Normal veri bulunan ikinci pencereye ait sinyal örüntüsü 22

Şekil 3.7. Normal veri bulunan üçüncü pencereye ait sinyal örüntüsü 22

Şekil 3.8. Normal veri bulunan dördüncü pencereye ait sinyal örüntüsü..... 22

Şekil 3.9. Normal veri bulunan beşinci pencereye ait sinyal örüntüsü 22

Şekil 3.10. Normal veri bulunan altıncı pencereye ait sinyal örüntüsü..... 23

Şekil 3.11. Normal veri bulunan yedinci pencereye ait sinyal örüntüsü..... 23

Şekil 3.12. Normal veri bulunan sekizinci pencereye ait sinyal örüntüsü..... 23

Şekil 3.13. Normal veri bulunan dokuzuncu pencereye ait sinyal örüntüsü 23

Şekil 3.14. Normal veri bulunan onuncu pencereye ait sinyal örüntüsü..... 24

Şekil 3.15. Normal veri bulunan onbirinci pencereye ait sinyal örüntüsü 24

Şekil 3.16. Anormal veri bulunan onikinci pencereye ait sinyal örüntüsü..... 24

Şekil 3.17. Negatif seçim algoritmasının akış şeması..... 28

Şekil 3.18. Negatif seçim algoritması akış diyagramı..... 30

Şekil 4.1. Web trafik verilerinde anomali tespiti için geliştirilen yazılımın arayüzü .. 31

Şekil 4.2. Anormal web trafik verisinin bulunduğu sinyal örüntü penceresi 32

Şekil 4.3. Uygulama yazılımına eğitim ve test verilerinin yüklenmesi..... 33

Şekil 4.4. Test verisi üzerinde anormal trafik verilerinin NSA ile tespiti 33

Şekil 4.5. Anormal web verilerinin bulunduğu sinyalde aktifleşen detektörler 34

Şekil 4.6. Anormal web trafik verisinin bulunduğu sinyal penceresi (P12)..... 36

Şekil 4.7. Uygulama yazılımına Deneysel Çalışma 2 için eğitim ve test verilerinin yüklenmesi 36

Şekil 4.8. Sinyal penceresi üzerinde anormal web verilerinin NSA ile tespit edilmesi37

Şekil 4.9. Anormal web verilerinin bulunduğu sinyal penceresinde aktifleşen detektörler 37

Şekil 4.10. Deneysel Çalışma 3 için anormal web trafik verisinin bulunduğu sinyal örüntüsü (P12) 39

Şekil 4.11. Uygulama yazılımına Deneysel Çalışma 3 için eğitim (P1) ve test verilerinin(P12) yazılıma yüklenmesi..... 40

Şekil 4.12. Deneysel çalışma 3 için sinyal penceresi üzerinde anormal web verilerinin NSA ile tespit edilmesi 40

Şekil 4.13. Deneysel çalışma 3 için anormal web verilerinin bulunduğu sinyalde aktifleşen detektörler (17. Sinyal örüntüsü) 41

Şekil 4.14. Deneysel çalışma 4 için anormal web trafik verisinin bulunduğu sinyal örüntüsü (P12) 42

Şekil 4.15. Deneysel çalışma 4 için anormal web verilerinin bulunduğu sinyalde aktifleşen detektörler (22. Sinyal örüntüsü) 43

Şekil 4.16 Deneysel Çalışma 5 için anormal web trafik verisinin bulunduğu sinyal örüntüsü (P12) 45

Şekil 4.17. Deneysel Çalışma 5 için anormal web verilerinin bulunduğu sinyalde aktifleşen detektörler (25. Sinyal örüntüsü) 46

Şekil 4.18. Deneysel Çalışma 6 için 40. sinyal üzerinde anormal web trafik verisinin bulunduğu sinyal örüntüsü (P10)..... 48

Şekil 4.19. Uygulama yazılımına Deneysel Çalışma 6 için eğitim (P1) ve test verilerinin(P10) yazılıma yüklenmesi..... 48

Şekil 4.20. Deneysel Çalışma 6 için sinyal penceresi üzerinde anormal web verilerinin NSA ile tespit edilmesi 49

Şekil 4.21. Deneysel Çalışma 6 için anormal web verilerinin bulunduğu sinyalde aktifleşen detektörler (40. Sinyal örüntüsü) 49

Şekil 4.22. Deneysel Çalışma 7 için 67. sinyal üzerinde anormal web trafik verisinin bulunduğu sinyal örüntüsü (P12)..... 51

Şekil 4.23. Deneysel Çalışma 7 için anormal web verilerinin bulunduğu sinyalde aktifleşen detektörler (22. Sinyal örüntüsü) 52

1. GİRİŞ

İnternet dünya çapında yayın yapma yeteneği, bilginin yayılması için bir mekanizma ve coğrafi konuma bakılmaksızın bireyler ve bilgisayarların arasında işbirliği ve etkileşim ortamıdır. İnternet, sürdürülebilir yatırımın yararına ve bilgi altyapısının araştırılması ve geliştirilmesine olan bağlılığının en başarılı örneklerinden birini temsil etmektedir. 1990 yılından beri hayatımızda aktif olarak kullandığımız internet teknolojileri, büyük bir gelişim sürecine girmiş olup özellikle hayatımızın her alanında yer edinmiştir. Özel hayattan pek çok alana kadar hayatımıza girmiş ve sürekli gelişim sürecinde olmuştur(Leiner vd.,1997).

İletişim kontrol protokolü / internet protokolü(TCP/IP)' nin gelişmesi ve paralel olarak internet teknolojilerindeki gelişmelerle birlikte, 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(Kim vd.,2018). 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 Raporunda küresel istihbarattaki veriler analiz edilerek dünya çapında 157'den fazla ülkede 123 milyon saldırı cihazından kayıt altına alınmış ve günlük 142 milyon tehditin engellendiği görülmüştür (Symantec,2019). 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 zafiyete uğramasına neden olmaktadır. İşte tam

bu noktada saldırıların tespiti ve veriler üzerindeki anormal durumların önceden öngörülebilmesi ağ sistemleri ile web trafikleri açısından olukça önemlidir.

Günümüzde bu kadar saldırganın olduğu düşünülürse ve sistemlerin zafiyetinden yararlanmak isteyenlerin neler yapabilecekleri göz önüne alınırsa saldırı trafiğinin yani anormal trafiği tespit etmenin ne kadar önemli olduğu fark edilebilir. Anormal olan trafik çeşitli saldırılar sonucunda ağa zarar verebilir.

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. Bu karmaşıklıkta anomali tespit sisteminin görevini zorlaştırmaktadır. Şimdiye kadar önerilen çoğu ağ anomalisi algılama sistemi, yanlış öğrenme imza temelli algılama yöntemlerine dayanan anomali tespitini kullanmıştır (Mazel,2011). Buna benzer yaklaşım türleri, bilinmeyen anomalileri tespit ve karakterize edememektedir.

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 (Berger,2017). Amaç, potansiyel saldırı kritik noktaya ulaşmadan önce ağdaki anomaliyi tespit etmektir.

Ağ üzerinde yaygın olarak Hizmeti Engelleme Saldırısı(DOS), Yoklama(Probe), Root Kullanıcısı Ele Geçirme(User to Root) ve Kullanıcıya Uzaktan Bağlanma (Remote to User) saldırıları 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 olarak farklı yollarla yapılabilmektedir(Kim ve Cho, 2018). 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 duyulmuştur. İki hatlı savunma mekanizması yaygın bir örnektir. İlk savunma hattı genellikle ağ çevresine yayılır. Yönlendiriciler (router) ve güvenlik duvarları (firewall) dahil olmak üzere çeşitli internete bakan aygıtlardan oluşan yapı bu hattı oluşturur. Bunlar kolayca tanımlanabilen meşru olmayan trafiği engellemek için yapılandırılmıştır. İkinci savunma hattı, bir Saldırı Tespit Sistemi(IDS)'dir. IDS devam eden kötü niyetli faaliyetlerin tespiti için ağ içindeki olayları analiz etmek için kurulmuş bir sistemdir. Buralardan gelen iz kayıtlarına(log) göre saldırıya ait kanıtlar ağ yöneticilerine bildirilir, böylece buna göre tepki verilebilir(Potoček ve Rehák,2017).

İzinsiz giriş tespiti yaklaşımlarından olan anomaliye dayalı izinsiz giriş tespiti yaklaşımı ise, kötü niyetli faaliyetlerin beklenen davranışlardan önemli ölçüde farklı olduğunu ve bu farklılığa tepki göstermeyi temel alır. Gelen trafikte, normal verilerin diğer verilerden ayrılıp ayrılmadıklarını kontrol etmek için analiz eder. Anomali tabanlı sistemler bilinmeyen ve yeni saldırıların tespitini destekler ve ayrıca güvenlik açıklarının neden olduğu sorunları gidermek için eğitilebilirler. Anomali temelli bileşenler, farklı makine öğrenme tekniklerini kullanarak modellenebilir (Agarval ve Hussain,2018).

Web sunucularda anormal trafik verilerinin tespiti genellikle bir zaman serisi problemi olarak değerlendirilmektedir. Bu problem sinyal pencereleri üzerinde özellik çıkarılıp sınıflandırma yapılarak değerlendirilmektedir (Zheng vd.,2014). Ancak web trafik verilerinin genel bir karakteristik örüntü yapısı olmadığından, çözüm yöntemleri de farklılıklar içermektedir. Literatürde ağ üzerindeki anormal trafiğin tespiti adına çalışmalar bulunmaktadır. Ancak yapay bağışıklık algoritmalarında negatif seçilim algoritmasını kullanarak durumu tespit eden yok denecek kadar az sayıda çalışma mevcuttur.

Bu tez çalışmasında, web trafiklerini gösteren Yahoo Webscope S5 (Yahoo, 2019) 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.

1.1 Literatür Araştırmaları

Literatürde ağ üzerinde anormal verilerin sınıflandırılması için yapılmış birçok çalışma bulunmaktadır. Bunlardan birisinde, Münz vd. (2007) 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. Ayrıca, K-ortalama kümelenme algoritmasına dayanan yeni bir anomali tespit şeması ortaya atmışlardır. Etiketlenmemiş akış kayıtlarını içeren eğitim verileri, normal ve anormal trafik akış kümelerine ayrılmıştır. Çalışmada, veri madenciliği ve anomali tespit işlemlerinin ayrıntılı bir tanımı yapılmıştır. Kümeleme algoritmasını uygulamanın farklı servisler için ayrı ayrı algılama kalitesini arttırdığı gözlemlenmiştir.

Bir diğer çalışmada Thill vd. (2017) Yahoo Webscope S5 verisetinde anomali tespiti için çeşitli çevrimiçi algılama algoritmalarının karşılaştırılmasını gerçekleştirmişlerdir. Mevcut algoritmalara dayanarak, yenilikçi bir çevrimiçi mesafeye dayalı anomali saptama algoritması önermişlerdir. Burada, Regresyon Analizi ile elde edilen sonuçların diğer anomali dedektörlerine kıyasla oldukça başarılı olduğu gözlemlenmiştir.

Kim ve Cho (2018) tarafından yapılmış olan bir diğer 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. CNN katmanı, uzamsal bilgilerdeki frekans değişimini azaltmak için kullanılır; LSTM katmanı zaman bilgisini modellemek için uygundur ve DNN katmanı

verileri daha ayrılabilir bir alana eşlemek için kullanılır. C-LSTM yöntemi, daha önce sınıflandırması çok zor olduğu düşünülen benzer sinyaller için bile, web trafiği verileri için neredeyse mükemmel anomali tespit performansı sağlamıştır.

Akbal ve Ergen (2006) tarafından yapılan çalışmada, kablosuz ağlarda saldırı tespitine farklı bir yönden yaklaşılarak tespit işlemi tüm kullanıcılar üzerinde yapmak yerine erişim noktası üzerinden kontrol işlemi gerçekleştirilmiştir. Bu çalışmanın amacı saldırı tespitindeki yetersizliği ortadan kaldıracak yeteneğe sahip olan Yapay Bağışıklık Sistemi ile kablosuz ağları ve ağ cihazlarını uzun süreli olarak herhangi bir müdahaleye gerek kalmadan ve performans kaybı yaşamadan oluşan istenmeyen durumları tespit etmek ve ağı korumaktır. Ağ üzerinde uzun süreler yapılan gözlemler neticesinde YBS'nin başarılı bir şekilde aksiyon verdiği ve çalışma süresi ne kadar uzarsa o kadar başarılı çalıştığı sonucuna varılmıştır.

Alkasassbeh vd. (2016) 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 DDoS (Hizmet Engelleme Saldırısı) saldırı türlerini sınıflandırmak için üç makine öğrenme algoritması Çok Katmanlı Algılayıcı(MLP), Rastgele Orman(Random Forest) ve Bayes uygulanmıştır. Çok Katmanlı Algılayıcı (MLP) sınıflandırıcısının en yüksek doğruluk oranını elde ettiği görülmüştür. Deneysel sonuçlar MLP'nin % 98.63 doğruluk oranına ulaştığını göstermiştir.

Dutt vd. (2016) yapmış olduğu çalışmada yapay bağışıklık sistemlerini kullanılarak network sistemine yapılan kötü niyetli atakların etkili bir şekilde tespit edildiği görülmüştür. Virüs, solucan vb. gibi zararlı yazılımları, belirli bir sunucuya belli sayıda bilgisayar tarafından bulaştırılmaya çalışılmış ve testlerin sonucunda gelen anormal trafiğin adedine bağlı olarak sistemin zararlı yazılımı ne oranda yakalayıp yakalayamadığı tespit edilmiştir. 1 ile 1000 arasında gönderilen dosya sonucunda; 1 dosya gönderimi sonucu %100, 10 dosya gönderiminde %90, 100 dosya gönderiminde %98, 1000 dosya gönderimi sonucunda %99 oranında tespit işleminin gerçekleştiği görülmüştür.

Aziz vd. (2012) tarafından yapılan çalışmada genetik algoritma tarafından oluşturulan dedektörler kullanılarak ağdaki anormal aktiviteyi tespit etmek için bir yaklaşım uygulanmıştır. Minkowski mesafe fonksiyonu, algılama işlemi için Öklid

mesafe fonksiyonuna karşı test edilmiştir. Sistemin yapısında Minkowski mesafe fonksiyonunun, Öklid mesafe fonksiyonuna göre daha iyi sonuçlar verdiği ve daha az veri kullanarak daha iyi sonuçlar verdiği gösterilmiştir.

Kim ve Bentley (2002) yapmış olduğu çalışmada, sızma tespiti için YBS'nin negatif seçimin rolü incelenmiştir. Çalışma, bir şebeke trafiği anomali detektörü olarak negatif seçimin kullanımına odaklanmaktadır. Negatif seçilim algoritması deneylerinin sonuçları, gerçek ağ trafiği verilerinin işlenmesi için ciddi bir ölçeklendirme sorunu olduğunu göstermiştir. Çalışmada, YBS'de en uygun negatif seçim kullanımının, yetkili dedektörlerin(uyarıcı) üretilmesi yerine geçersiz dedektörler için bir filtre olduğunu öne sürerek sona ermektedir.

Zhang vd. (2003) yaptığı çalışmada, kablosuz ağların güvenlik açıklarını incelemiş ve mobil bilgi işlem ortamı için güvenlik mimarisinde saldırı tespitinin dahil edilmesi gerektiği savunulmuştur. Böyle bir mimari geliştirilmiş, değerlendirilmiş ve bu mimarideki anahtar bir mekanizma olan mobil geçici ağ için, simülasyon deneyleri aracılığıyla anomali tespiti yapılmıştır. Sınıflandırıcıları anomali detektörleri olarak hesaplamak için RIPPER ve SVM Light kuralları uygulanmıştır. Bu model hayata geçirilmiş ve performansını değerlendirmek için simülasyonlar yapılmıştır. Son olarak, bu dedektörlerin genel olarak iyi tespit performansına sahip olduğunu gösterilmiştir.

Silva vd. (2017) yapmış olduğu bir diğer çalışma, hata algılama ve izolasyon problemlerinde yeni nesil YBS'nin uygulanmasını amaçlamıştır. Yapılan çalışma ile üç YBS yaklaşımı üzerine bir inceleme sunulmaktadır. Gözden geçirilip bağlamsallaştırıldıktan sonra, değerlendirilen teknikler, ana parametreleri ve veri işleme şekilleri dikkate alınarak uygun bir şekilde ayarlanmış ve bu tekniklerin performans analizlerini, problemlerde uygulanabilirliklerine göre arıza tespit ve izolasyonu vaka çalışmasında uygulanmıştır.

Das vd. (2018) yapmış olduğu çalışmada çok sayıda alanda yaygın olarak kullanılan bir kavram olan anomali tespitini, Anomali Tabanlı Saldırı Tespit Sistemi (ABIDS)'ne dayandırarak, yeni güvenlik açıkları ve saldırıların sürekli göründüğü trafiklerde önemli olan trafiğin önceden bilinmeyen saldırılar olduğunu düşünerek bu anormal trafiğin tespit edilmesine yoğunlaşmışlardır. Ayrıca, ağ sistemlerine yeni saldırıları gözlemlemek için pratikte ve araştırma alanında yıllar içinde geliştirilen

birkaç anomali tespit tekniğini analiz etmişlerdir. İki farklı veri setini analiz etmişler ve anomali tespiti için en iyi YBS sınıflandırıcısını bulmaya çalışmışlardır.

Dokas vd. (2002) yapmış olduğu çalışmada, bilinenleri belirlemek için nadir sınıf tahmin modelleri oluşturma konusundaki araştırmasına genel bir bakış açısı sunmaktadır. İzinsiz girişler ve çeşitleri araştırılmış, anomali ve dışlayıcı tespit şemaları oluşturularak, bilinmeyen yeni saldırıları tespit etme amacı güdülmüştür. KDDCup's 99'da elde edilen deneysel sonuçlar, standart sınıflandırma tekniklerine göre nadir sınıf kestirim modellerinin izinsiz girişlerin tespitinde çok daha etkili olduğunu göstermiştir. DARPA 1998 veri setinde ve ayrıca Minnesota Üniversitesi'ndeki canlı ağ trafiğinde izinsiz girişlerin tespitinde yeni tekniklerin büyük oranda sonuç verdiği gösterilmiştir. Özellikle, yapılan çalışmanın son birkaç ayı boyunca SNORT gibi en gelişmiş araçlar kullanılarak tespit edilemeyen birkaç yeni izinsiz girişin belirlenmesi geliştirilen teknikler ile başarılı olmuştur.

1.2 Tezin Amacı

Bu tezin amacı, web trafiğine yapılan anomali trafiğinin tespit edilmesi ve tespit edilen bu trafiğin önlemesi için ağ trafiğindeki kötü amaçlı yazılımlar üzerindeki algılama etkinliğini arttırmaktır. Bunu gerçekleştirmek için, çeşitli web atak türlerini barındıran Yahoo S5 verisetindeki normal ve anormal veriler içeren trafik verileri kullanılacaktır. Web trafiklerinde anormal verilerin tespitinde YBS'e ait Negatif Seçim Algoritmasından yararlanılacaktır. Geliştirilecek yazılım üzerinde yapılacak deneysel çalışmalar ile web trafik verilerinden normal ve anormal trafiğin başarı oranı tespit edilecektir.

1.3 Hipotez

Yapay Bağışıklık Sistemlerinin Negatif Seçim Algoritması ile web trafiklerindeki anormal veriler yüksek başarım oranı ile tespit edilebileceği öngörülmektedir.

2. WEB GÜVENLİĞİ ve SALDIRI TİPLERİ

Web trafiği, bir bağlantıya tıklamak gibi basit görünen ama arka planda üst düzey parametre ve protokollerin çalışması ile başlar, ağ anahtarları ve kablolar arasında gezinmek gibi düşük düzey elemanlarla devam eder. Başka bir deyişle, Web trafiği genellikle kullanıcılar tarafından web tarayıcıları kullanılarak başlatılır. Trafik akışı, tarayıcı bilgilerini önceden belirlenmiş kuralları ve kullanıcı tarayıcı isteklerini elde etmek için yöntemler kullanan bir sunucuya gönderen bir fare tıklamasıyla başlar. Bu kurallara dayanarak, sunucu daha sonra hangi işlemin yapılması gerektiğine karar verir. Günümüzde, web trafiğindeki artış her yıl internet kullanıcılarındaki sayıyla paralellik göstermektedir. Web trafiği ile aynı oranda bu trafiğe yapılan saldırı da artmaktadır (Pande,2014).

Her gün milyarlarca kullanıcı herhangi bir ürünü satın almakta, para transfer etmekte, bilgi almakta ve web üzerinden birbirleriyle iletişim kurmaktadır. Web uygulamaları insan tarafından yapıldığı için, insan hatası olan çok fazla boşluk içermektedir. Ağdaki bu boşlukları ve bilgisayar sistemlerindeki bu boşlukları bulma, bunlardan yararlanarak veri elde etmeye bilgisayar korsanlığı denir. Uygulamaların popülaritesi, bilgisayar korsanlarını kendine çekmektedir. Web uygulaması saldırılarına karşı tespit, önleme ve çözüm bulma çalışmaları internet dünyasında önemli bir yer tutmaktadır (Adhyaru, 2016).

Web güvenliği tarafında bilgisayar korsanlarını engellemek adına bazı güvenlik önlemleri vardır. Web güvenlik protokollerinden Sanal Özel Ağ(VPN), verileri şifrelemek ve iletmek için tünelleme yapar. Bir paket iletilmeden önce, yeni bir başlık ile yeni bir paket içinde kapsüllenir. Kapsüllenmiş paketlerin içinden geçtiği bu mantıksal yola tünel adı verilir. Oluşan yeni başlık, paylaşılan veya halka açık bir ağı tünel bitiş noktasına ulaşmadan önce geçebilecek şekilde yönlendirme bilgisi sağlar. Her paket tünelin bitiş noktasına ulaştığında, “kapsüllenir” ve son hedefine iletilir. Her iki tünel uç noktasının da aynı tünel protokolünü desteklemesi gerekir. Tünel protokolleri, Açık Sistem Bağlantısı(OSI) ikinci katmanında (veri bağlantı katmanı) veya üçüncü katmanında (ağ katmanı) çalıştırılır.(The Government of the Hong Kong Special Administrative Region,2008).

Göreceli olarak genel amaçlı başka bir çözüm, güvenliği TCP'nin hemen üzerinde uygulamaktır. Bu yaklaşımın en önde gelen örneği, Güvenli Soket

Katmanı(SSL) ve Taşıma Katmanı Güvenliği(TLS) olarak bilinen internet standartıdır. SSL/TLS protokolleri, taşıma katmanı protokolü üzerinden güvenilir hizmetler sağlamak için kullanılır. SSL, günümüze kadar pek çok güncellemeden geçmiştir. SSL bağlantısı, istemcilerle sunucular arasında bağlantı kurmak için taşıma katmanında çalışır ve verileri şifreler. Her SSL oturumu bir SSL bağlantısıyla ilişkilidir. SSL bağlantısı sertifika ile çalışır ve bu sertifika dünyada bazı otoriteler tarafından verilir. Alınan sertifika ile yayın yapan web sitesi adresinin başına Güvenli Hiper Metin Transfer Protokolü(https) eklenir ve bu sayede artık gelen ve giden trafik şifreli ve güvenli bir şekilde aktarılır. SSL 'in iyi bir koruma sağlamasındaki mantık veri transferi yapılan nokta ile veri transferini başlatan nokta arasında ikili şifreli anahtar uyumudur. Yani genel ve özel anahtar mantığı. Öyle ki bir noktadaki şifrelenmiş veri çözülse bile diğer taraftaki şifrelenmiş veri çözülmediğinden yani anahtar kilit uyumu gerçekleşmediğinden dolayı veriler elde edilemeyecektir. TLS ise, SSL'nin güncellenmiş hali olduğundan, aynı mimari ve protokoller dışında bazı değişiklikler barındırır. Bunlar, güvenlik parametreleri, Ortam Erişim Kontrolü (MAC) hesabı, dijital imza, anahtar bloğu ve daha gelişmiş şifreleme algoritması gibi.

Güvenlik önlemleri trafiği güvenli bir şekilde taşıyıp veri transferinde kayıp olmasına engel olabilir. Ancak kötü niyetli olarak adlandırılan bilgisayar korsanları SSL ve VPN gibi güvenlik önlemlerinde olan açıklıklar nedeniyle bir şekilde sisteme erişmenin yolunu bulmuşlardır. Buradan şu sonuca varılabilir ki sadece SSL ve VPN gibi protokoller web güvenliğini sağlamada yetersiz kalmaktadır. SSL veya TLS trafiğini kullanamayan siteler için zaten tüm trafik açık olarak geçeceği için bilgi güvenliğini ve sistem sürekliliğini sağlamak oldukça zorlaşacaktır. Alınan güvenlik önlemlerinin yetersizliği arka planda farklı güvenlik cihazlarının ve uygulamalarının zorunluluğunu ortaya çıkarmıştır. Örneğin, ön planda tüm trafiği karşılayan IDS sistemleri, sunucuların önüne entegre edilen Web Uygulama Güvenlik Duvarı(WAF) ve güvenlik duvarı başta gelenler olur.

Bilgisayar Güvenliği Enstitüsü(CSI) ve Federal Soruşturma Bürosu(FBI) tarafından yürütülen bilgisayar suçlarıyla ilgili ortak bir araştırma, işletmelerdeki en ciddi kayıpların içerdekilerin yetkisiz erişimle gerçekleştiğini ve katılımcıların %71'inin yetkisiz erişim tespit ettiğini göstermektedir. Bu nedenle, yeni erişim kontrol modelleri geliştirmek veya güvenlik tehditlerini etkisiz hale getirmek ve web tabanlı

uygulamaların farklı güvenlik gereksinimlerini karşılamak için mevcut olanları genişletmek gerekmektedir (Joshi,2001). Aşağıda sıklıkla karşılaşılan bazı web atakları açıklanmıştır.

2.1 SQL Saldırısı (SQL Injection)

SQL Saldırısı, saldırganın bir uygulamanın veritabanında yaptığı sorguları çalıştırmasına izin veren bir web güvenlik açığıdır. Genellikle saldırganın normalde alamadığı verileri görüntülemesini sağlar. Saldırgan bu verileri değiştirebilir, silebilir veya veritabanında istediği sorguyu çalıştırabilir. Bu da uygulamanın içeriğinde veya davranışında kalıcı değişikliklere neden olabilir. Ayrıca kritik verilerin dışarı sızmasına neden olabilir. Bazı durumlarda, saldırgan, ana sunucuyu tehlikeye atmak veya DOS gerçekleştirmek için bir SQL saldırısını artırabilir (PortSwinger Ltd.,2019).

2011 yılında, Sony, Sony PlayStation Network, Sony Müzik Japonya ve Sony Pictures defalarca saldırıya uğramıştır. PlayStation Network'e yapılan saldırı 100 milyon kullanıcının kişisel bilgilerini tehlikeye atmıştır ve 171 milyon $ 'a mal olduğu tahmin edilmektedir. Sony Music'e yapılan saldırı hassas bilgilerin ihlal edilmesine yol açmazken, Sony Pictures saldırısı bir milyon kullanıcının kişisel bilgilerinin açığa çıkmasına neden olmuştur (Horner,2017).

2.2 XSS Saldırısı

Siteler Arası Komut Dosyası Çalıştırma (XSS) güvenlik açıkları çok yanlış anlaşılmakta ve tedarikçiler tarafından hak ettikleri endişe ve dikkat gösterilmemektedir. Basitçe söylemek gerekirse, XSS'e karşı hassas bir web uygulaması, kullanıcının yanlışlıkla bu uygulama aracılığıyla kendisine kötü niyetli veri göndermesini sağlar. Bu bağlantılar, saldırganın kurbanının tarayıcısında çalıştırmayı seçtiği istemci tarafındaki komut dosyası dilleriyle (VBScript, JavaScript vb.) olur. XSS güvenlik açıkları, web uygulamasındaki kullanıcı girişini doğru şekilde doğrulamaması nedeniyle ortaya çıkar (Endler,2002).

XSS tüneli kullanarak, saldırgan kurbanın bilgisayarını kontrol edebilir. Bilgisayar korsanları, kötü amaçlı kodu XSS güvenlik açığı olan bazı ünlü web sitelerine enjekte ederse, sayfayı ziyaret edenler saldırganlar tarafından kontrol edilebilir. Pek çok kullanıcı cihazlara virüs bulaşabilir. Bu, XSS'in ne kadar yıkıcı

olabileceğini göstermektedir. Bu tür saldırıları önlemek için XSS güvenlik açıklarına daha fazla dikkat edilmesi gerekmektedir.

2.3 Yanlış Güvenlik Yapılandırması

İyi güvenlik anlayışı, uygulama için tanımlanmış ve şekillendirilmiş güvenli bir yapılandırma gerektirir. Özellikle uygulama sunucusu, veritabanı sunucusu gibi önemli ve hassas verilerin bulunduğu ve her daim saldırıya açık platformların güvenlik ayarları tanımlanmalı, uygulanmalı ve korunmalıdır. Sisteme yeni dahil olan cihazların ayarları, genel olarak varsayılan olduğu ve güvenlik riski oluşturduğu için konfigürasyonlarının ve güncellemelerinin yapılması gerekmektedir.

Şekil 2.1'de gösterilen Lumension'un yayınladığı Güvenlik Yapılandırma Yönetimi aşağıda detaylıca listelenmiştir:

1.Keşif: Heterojen ağ ortamı için tam görünürlük kazandırılmalı. Ayrıntılı taramalar ve esnek gruplandırma ve sınıflandırma seçenekleriyle hem yönetilen hem de yönetilmeyen tüm BT(Bilgisayar Teknoloji) varlıkları en ince ayrıntısına kadar keşfedilmelidir.

2. Değerlendirme: En iyi uygulamalara ve standartlara dayalı politikalara karşı güvenlik yapılandırma sorunları tanımlanmalıdır. Windows Güvenlik Kılavuzları, Uluslararası Teknoloji ve Standart Enstitüsü (NIST) Özel Yayın, Savunma Bilgi Sistemleri Ajansı (DISA), Güvenlik Teknik Uygulama Kılavuzları (STIG), Ulusal Güvenlik Ajansı (NSA), Yönetim ve Bütçe Ofisi (OMB), Federal Masaüstü Çekirdek Konfigürasyonları (FDCC) gibi güvenlik yayınları takip edilebilir.

3. Önceliklendirme: İlk önce en kritik güvenlik cihazları ve risklerine odaklanılmalıdır.

4. Düzeltme: Tüm ana platformları ve uygulama ortamlarını sürekli izleyip, tespit ederek ve düzelterek güvenli bir ortam sağlama sürecini basitleştiren otomatik politikalar oluşturulmalıdır.

5. Rapor: Güvenlik yapılandırması politika ihlallerine ilişkin bütünsel bir görünüm kazandırılmalıdır. Tek bir yönetim konsolunda keşif, değerlendirme ve düzeltme bilgilerini birleştiren kapsamlı ve operasyonel yönetim raporları hazırlanmalıdır(Lumension,2009).

Şekil 2.1.Güvenlik yapılandırma yönetimi(Lumension,2009)

Güvenlik kontrolleri çok farklı sebepler yüzünden yanlış yapılandırılmış olabilir. Kullanıcının karşılaştığı hata veya eksikliklerden yola çıkarak sistem yöneticilerinin yaptığı hatalar örneklendirilebilir. Doğamızda olan ve insan olmanın verdiği yanlış yapma olasılığımız nedeniyle aslında bunu normal olarak karşılanabilir. Özellikle çoğu web uygulamaları diğer yazılımlara, veritabanlarına, kütüphanelere bağlıyken (Apache, İnternet Bilgi Servisi (IIS), Hiper Metin Önişlemcisi (PHP), Oracle gibi) bu karmaşık yapıda hata yapma oranını da arttığı söylenebilir. Ancak tüm bu yapı kontrolü belirli bir sistematiğe oturtulduğunda en azından verilen zafiyet de o kadar düşecektir.

2.4 Yetersiz Saldırı Kontrolü

Kötü niyetli aktörler yeni güvenlik açıklarından yararlanmak ve büyük çapta savunmasız sistemleri tespit etmek ve saldırı başlatmak için değişik yollara başvururlar. Bu kategori güçlü yeni saldırı vektörlerine ve anormal durumlara zamanında tepki verebilme yeteneğine odaklanır. Uygulamanın güvenliği şu sorularla ve alınan yanıtlarla tespit edilebilir:

- Ne zaman saldırıya uğradığını tespit edip tanıyabiliyor mu?
- Sıra dışı istek modelleri veya yüksek hacimli trafiklere karşı sistem otomatik olarak algılayan bir mekanizmaya sahip mi?
- İstenmeyen trafiğe tepki verme ve engelleme yeteneği nedir? (Amazon,2017).

Burada önemli olan her zaman saldırıya maruz kalınacağı psikolojisiyle hareket etmek ve bu doğrultuda gerekli önlemlerin alınmasıdır. Uygulamanın çalıştığı ağ topolojisine göre yapılandırılmış olan güvenlik cihazları üzerindeki loglar gerektiğinde detaylıca incelenmeli kritik olarak görülen ve sistemi ağır zafiyete uğratabilecek saldırıların her zaman gelebileceği unutulmamalıdır. Diğer bir husus ise kurumlarda çalışan ağ güvenlik ekibinin bu konuda eğitilmeli, saldırı öncesinde, anında, sonrasında nasıl önlem alınması gerektiği anlatılmalı ve gerekirse eğitimlerle desteklenmelidir. Unutulmamalıdır ki zincirin en zayıf halkasını insan unsuru oluşturmakta ve bu unsur ne kadar gelişirse sistemler o kadar güvenli olacaktır.

2.5 XML Harici Girişler

Genişletilebilir İşaretleme Dili (XML), yapılandırılmış veri nesnelerini insan tarafından okunabilir metin olarak göstermek için kullanılan, Hiper Metin İşaret Dili (HTML)'in de temel aldığı eXtensible Markup Language anlamına gelen işaretleme dilidir. XML, verilerin depolanması ve iletilmesi için bir format olarak tasarlanmış olup herhangi bir uygulamaya uyarlanabilmesi için genişletilebilir. Verilerin nasıl düzenlendiğini ve temsil edildiğini tanımlar (Kohfelder vd.,2018).

Açık Web Uygulama Güvenliği Projesi (OWASP), bu saldırıyı XML saldırısı, XML girişini ayrıştıran uygulamaya karşı yapılan bir saldırı olarak tanımlar. Bu saldırı, gizli verilerin ortaya çıkarılmasına, hizmet reddine, sunucu tarafı talebinde sahtecilik yapılmasına, port taramasına ve diğer sistem ataklarına neden olabilir (Haboob,2018).

Bir saldırgan, harici bir varlık referansı içeren XML girişi sağlayarak, XML ayrıştırıcısının başvurulan verileri okumasına ve sonuçta XML verilerine işlemesine neden olabilir. XML Dış Varlık, harici site adreslerinden alınacak değiştirme değerlerinin bir aracıdır. Böylece dosyalara ve ağ kaynaklarına potansiyel olarak erişebilir. Elde edilen verileri açığa çıkarmak için bir yol varsa, saldırgan XML ayrıştırıcı işleminin erişim ayrıcalıklarından yararlanarak verileri boşaltmayı başarabilir. Alternatif olarak, çok büyük bir veri kaynağına başvurarak, Hizmet Reddi Saldırısı (DOS)'na da yol açabilir (Kohfelder vd.,2018).

Yetersiz bir şekilde konfigürasyon yapılmış ve yanlış yapılandırılmış XML ayrıştırıcıları ve parametreleri, XML belgeleri ve XML girişini kabul eden Basit Nesne Erişim Protokolü (SOAP) servisleri gibi uygulamalar harici varlık referanslarının istemeden sisteme girişine izin verebilir. XML parametrelerinin verileri dahili olarak

ortaya çıkarmasına neden olan dosya paylaşımları, kod yürütme, tarama başlatma ve DOS saldırıları beklenmeyen dış referans ve komutlar olarak nitelendirilebilir. Hangi uygulamanın ilk sırada gösterileceği ve XML açıklarının giderilmiş olması burada büyük önem arz etmektedir. Bir web uygulamasından önce topolojide yapılandırılmış WAF bu tür atakların önlenmesinde büyük rol oynar (F5 Networks,2017).

2.6 Hassas Veri Sızıntısı

Kredi kartı numaraları, sağlık verileri ve şifreler gibi hassas veriler ekstra korumaya ihtiyaç duyulan verilerdir. Veri hassasiyetini korumak için tasarlanmış düzenlemeler ve standartlar mevcuttur. Ancak bu hassas veriler sağlıksız bir şekilde saklanır, aktarılır veya korunursa saldırganların bu verileri ele geçirmesi günümüz koşullarında hiç de zor olmaz. Veriler düz metin olarak saklanır veya aktarılırsa, daha eski ve zayıf şifreleme yöntemleri kullanılmışsa veya veriler dikkatsiz bir şekilde çözülmüşse bu durumda saldırganlar erişim kazanabilir ve verilerden yararlanabilir. Bu yüzden verileri saklarken ve aktarımını yaparken oluşacak tüm risklerin göz önüne alınması gerekmektedir (Hackerone,2017).

Birçok web uygulaması, kredi kartları, vergi kimlikleri ve kimlik doğrulama bilgileri gibi hassas verileri düzgün şekilde koruyamaz. Hassas veriler, aktarılırken şifreleme gibi ek korumayı ve değiştirilirken özel önlemleri hak eder. Bu riskin neden olduğu sorunlarla saldırganlar, kredi kartı sahtekarlığı, kimlik hırsızlığı veya diğer suçları işlemek için zayıf korunan verileri çalabilir veya değiştirebilir. Bu açıklığı gidermek için, şifreleme algoritmaları kullanarak gerekli görülen verileri şifrelemek doğru bir yöntemdir. Veritabanında depolanmadan önce kredi kartları, şifreler ve diğer hassas veriler Şekil 2.2'de gösterildiği gibi bilgilerin şifrelenmesi gerekmektedir (F5 Networks,2017).

Şekil 2.2. Hassas veri araştırılması (F5 Networks,2017)

2.7 Güvensiz Serileştirme

Seri hale getirme, bir nesneyi daha sonra geri yüklenebilecek veri biçimine veya bayt akışına dönüştürme işlemidir. Seri biçimde yapılır. Bu serileştirilmiş veri alınır ve tekrar bir veri nesnesine dönüştürür. Bir nesnenin durumunu koruyabildiğinden faydalıdır (Messina,2018) .

Uygulamalar, bir saldırganın sağladığı düşmanca veya tahrif edilmiş nesneleri seri hale getirirlerse savunmasız olacaktır. Bu, iki ana saldırı türüne neden olabilir. Saldırganın uygulama mantığını değiştirdiği seri kaldırma sırasında veya sonrasında davranışını değiştirebilecek sınıflar varsa, rasgele uzaktan kod yürütme gerçekleştirdiği nesne ve veri yapısıyla ilgili saldırılar ve mevcut veri yapılarının kullanıldığı ancak içeriğin değiştirildiği erişim kontrolü ile ilgili saldırılar gibi tipik veri kurcalama saldırıları. Seri hale getirme, uygulamalarda aşağıdakiler için kullanılabilir.

- *Uzaktan ve süreçler arası iletişim*
- *Tel protokolleri, web servisleri, mesaj kırıcıları*
- *Önbellekleme / Kalıcılık*
- *Veritabanları, önbellek sunucuları, dosya sistemleri*
- *HTTP çerezleri, HTML form parametreleri, API kimlik doğrulama belirteçleri*

(Cheatography,2018).

2.8 Bozuk Kimlik Doğrulama

Bozuk Kimlik Doğrulama, oturum yönetiminin yanlış yapılandırılması nedeniyle oluşan bir tür web güvenlik açığıdır. Bir kimlik doğrulama işlemi tamamlandıktan

sonra, sunucu ile doğrulama işlemi gerçekleştirilen kullanıcı arasında veri iletişimi için etkinleştirilecek olan bir oturum oluşturulur. Herhangi bir davetsiz misafir, kimlik doğrulama işlemini atlayarak, belirli bir kullanıcının aktif oturumuna erişebiliyorsa, yanlış kimlik doğrulama olarak değerlendirilir. Neredeyse tüm web uygulamaları, kullanıcılarına kaliteli hizmet ve iletişim sağlamak için ayrı ayrı kullanıcı profili oluşturmaktadır. Yanlış kimlik doğrulama ve oturum yönetimi sorunu, web uygulamasının gizliliğinin en büyük engellerinden biridir (Hassan vd., 2006).

Kimlik doğrulama ve oturum yönetimindeki kusurlar kimlik bilgilerinin ortaya çıkarılması demektir ki bu da kişisel verilerin çalınması anlamına gelir. Bu yüzden kullanıcı ve idari hesapların yetkilendirilmesi ve sorumluluğunu doğru şekilde yapmak gerekir, diğer türlü gizlilik ihlallerine neden olacağından dolayı veri kaçırılmasına yol açabilir. Ana kimlik doğrulama mekanizmasındaki kusurlar sistemin zayıf yönü olarak düşünülebilir. Bu mekanizmanın güvenli bir şekilde çalışması için yardımcı kimlik doğrulama fonksiyonları çalıştırılmalı, özellikle şifre yönetimi, zaman aşımı, beni hatırla, hesap güncellemesi ve gizli soru gibi tedbirler alınmalıdır. Buradaki amaç uygulamanın kullanıcıları doğrulaması ve kimlik bilgilerini güvenli bir şekilde saklamasıdır (Veracode,2019).

2.9 Yetersiz Kayıt ve İzleme

Yetersiz kayıt, tespit edememe, izleyememe ve aktif cevap verememe sistemin veya uygulamanın güvenlik açığı olarak değerlendirilebilir. Bu güvenlik açığına aşağıdaki maddeler neden olmaktadır.

•Başarısız giriş ve yüksek değerli işlemler gibi denetlenebilir olaylar günlüğe kaydedilmemesi,

•Uyarılar ve hatalar, yetersiz veya net olmayan loğlar olarak günlük mesajlarına kaydedilmesi,

• Şüpheli faaliyetler için uygulamaların kayıtlarının izlenmemesi,

• Günlüklerin sadece yerel olarak depolanması,

•Uygun uyarı eşikleri ve cevap süreçleri mevcut veya etkili olmaması,

•Penetrasyon (Sızma) testleri ve taramalar uyarıları tetiklemez ve gerekli önlem alınmazsa,

•Uygulamanın, gerçek zamanlı olarak aktif saldırıları tespit edememesi ve uyaramaması(Cheatography, 2018).

2.10 DDOS Saldırıları

DDoS ataklarının çoğunluğu hacimseldir, ağı tıkamak, servisleri çökertmek ve genellikle web sitesi sunucularını etki altına alıp cevap vermemesini sağlamak amacıyla yapılır. Bu tür bir saldırı çok az karmaşıklık ve beceri gerektirir, bu da kötü niyetli insanlar için ideal bir durumdur. DDOS saldırıları, genellikle büyük bir botnet (saldırı amaçlı yazılımlar) ve yansıtma saldırısı kombinasyonu kullanarak gerçekleştirilir ve bu saldırı belirli siteler aracılığıyla çok az miktar para karşılığında zombi (ele geçirilmiş) bilgisayarlar kiralanarak gerçekleştirilebilir. Üç tür DDOS saldırı atak tipi vardır:

i. Volümetrik(Hacimsel) ataklar; genel olarak, botnetler, zombi(ele geçirilmiş) istemciler v.s. gibi kaynaklardan çok yoğun trafik üretilerek, sunucunun bant genişliğini tamamen tüketmek üzere kurulu atak türüdür.

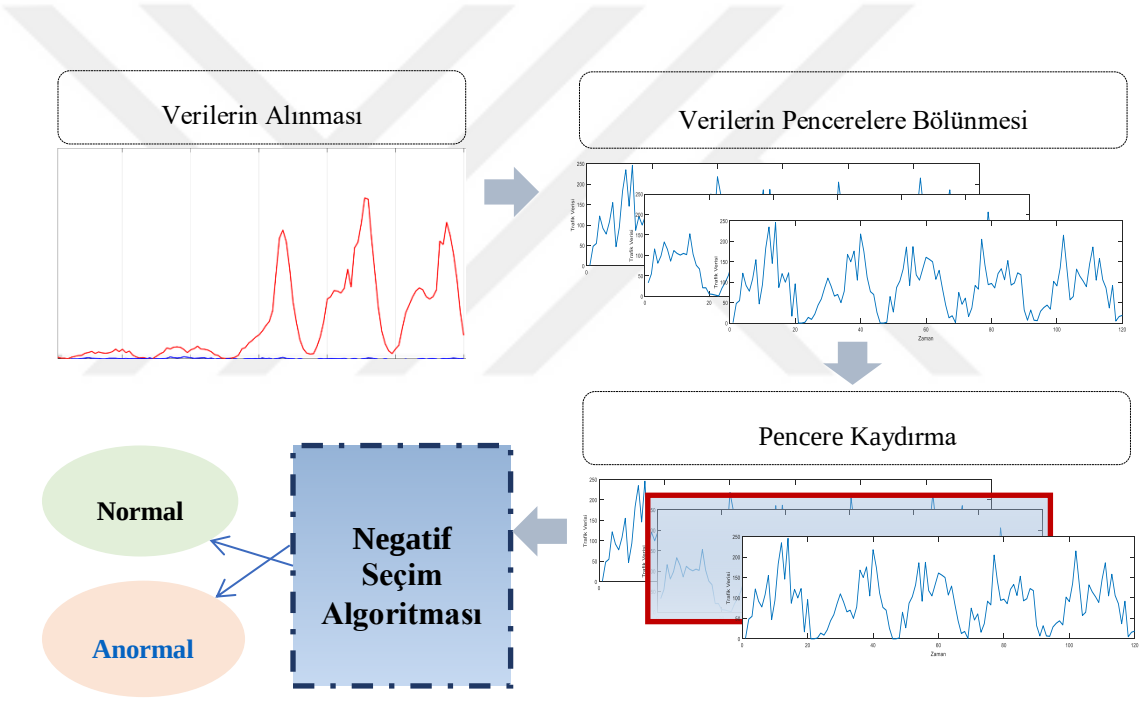
ii. TCP Tüketme Atakları; üçlü el sıkışmanın zafiyetini kullanarak oluşturulan ataklardır. Örnek olarak üçlü el sıkışma SYN paketi ile başlar hedef SYN-ACK cevabı döner, kaynak gelen pakete cevap vermeden yeniden SYN paketi yollar dolayısıyla hedef sürekli cevap beklemek zorunda kalır bu olaya SYN istilası denir.

iii. Uygulama Katmanı Atakları; uygulama ve/veya sunucudaki zafiyetlerden yararlanarak yapılan daha karmaşık saldırılardır. Bu saldırı metodunu tespit etmek daha zordur çünkü birçok istemcinin saldırıya geçmesi gerekmediğinden, yasal olduğu düşünülen düşük bir trafik oranı oluşturur (Sakar,2018).

Tarihte yapılan en büyük DDOS atağı 28 Şubat 2018 tarihinde GitHub firmasına yapılan 1.35 Tbps 'lik DDoS saldırısıdır. Bu boyutta bir trafik için botnet ağı kullanılmamış, saldırının daha güçlü hale gelmesi adına yanlış yapılandırılmış memcahce (genel amaçlı bir dağıtılmış bellek önbellekleme sistemi) uygulanmış sunucular hedef alınmıştır.

3. MATERYAL ve YÖNTEM

Bu tez çalışmasında önerilen yöntem ile bir boyutlu zaman serileri şeklinde oluşturulan Yahoo Webscope S5 (Yahoo, 2019) veriseti ve YBS'nin NSA'sı kullanılarak anormal web trafik verilerinin tespit edilmesi sağlanmıştır. Önerilen yöntemin tespit yapısını ve akışlarını gösteren açık diyagram Şekil 3.1'de sunulmuştur. Şekil 3.1'de görüldüğü gibi ilk önce verisetinden veriler alınır. İkinci adımda bu veriler belli sayıda pencerelere 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.

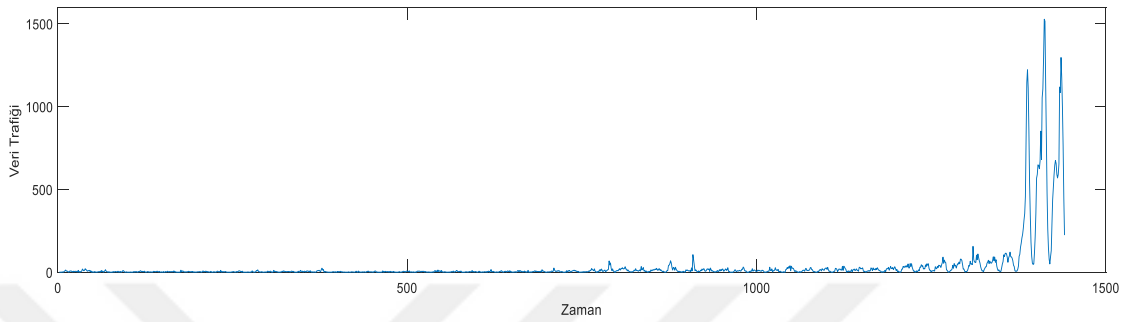


Şekil 3.1.Önerilen YBS-NSA destekli yöntemin tespit yapısı.

3.1. 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 (Yahoo, 2019) veriseti kullanılmıştır. Bu veriseti A1, A2, A3 ve A4 olmak üzere toplam dört farklı sınıf ve 367 adet zaman serisi sinyal örüntüsünden oluşmaktadır. Her bir sinyal örüntüsü ortalama 1500 veri noktası içermekte olup toplamda dört farklı sınıfta 5050000 veri noktası bulunmaktadır. A1

sınıfında gerçek veriler bulunmakta iken, diğer sınıflardaki web trafik verileri sentetik olarak oluşturulmuştur. Yahoo Webscope S5 veri setindeki A1 sınıfında 67 adet gerçek web trafiği dosyası bulunmaktadır. Şekil 3.2'de A1 sınıfında bulunan bir sinyal örüntüsü gösterilmiştir.



Şekil 3.2. Yahoo Webscope S5 verisetinden örnek bir sinyal örüntüsü.

Bu çalışmada A1 sınıfındaki anormal web trafiğinin tespiti sağlanmıştır. Çizelge 3.1'de Yahoo Webscope S5 veri setindeki verilere ait detaylı bilgiler sunulmuştur.

Çizelge 3.1. Yahoo Webscope S5 verisetinin detayları**.**

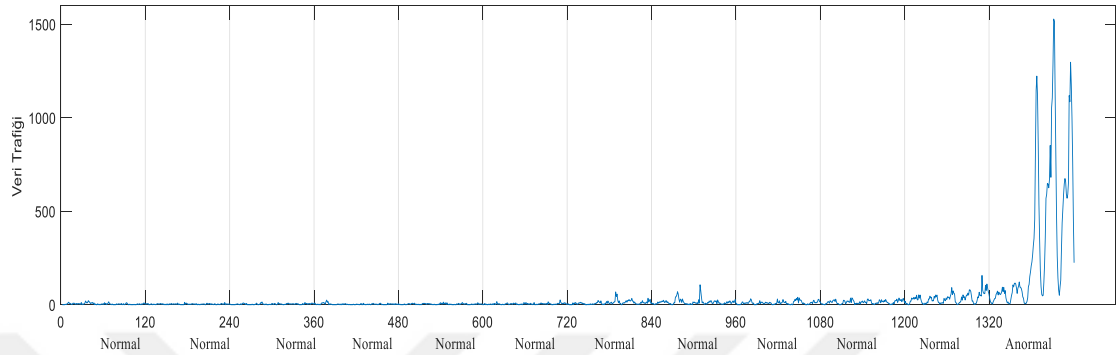
Sınıf	Gerçek / Sentetik Trafik (G / S)	Toplam Uzunluk	Toplam Anomali
A1	G	94866	1669
A2	S	142100	466
A3	S	168000	943
A4	S	168000	837

Tüm veri kümeleri (zaman serisi), üç sütun içeren CSV dosyaları biçiminde verilmiştir ve buradaki değerler; zaman damgası, değer, anomalidir. Zaman damgası trafiğin oluştuğu zaman değerlerini karşılarken, değer ağda oluşan anlık veriler ve anomali diye bahsedilen kısım ise anomalinin olup olmadığını bize belirten kısımdır.

3.2. 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. Bu pencerelenen normal trafik değerlerini

eğitim verisi olarak kullanarak deneysel çalışmalardaki anormal trafiğin bulunması önerilen yöntem ile sağlanmıştır. Şekil 3.3'te örnek bir sinyal örüntüsünün pencerelere bölünmesi, normal ve anormal trafik verilerinin olduğu alanlar gösterilmiştir.



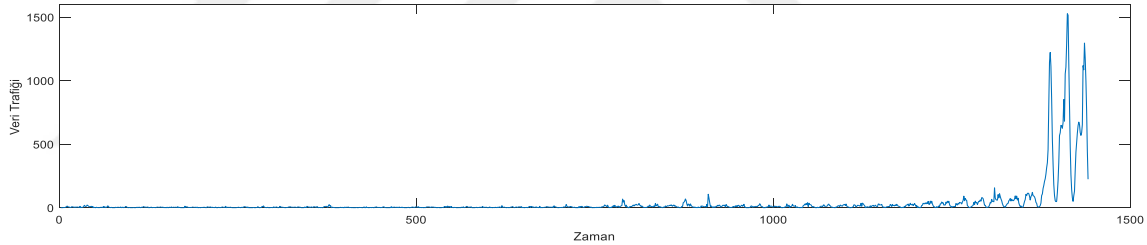
Şekil 3.3. Web trafik verilerin pencerelere bölünmesi

3.3. Pencere Kaydırma

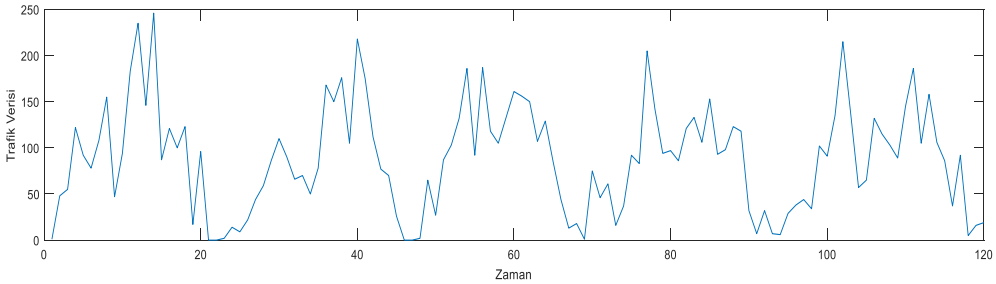
Kayan pencere tekniği, sürekli bir zaman serisi akışı verildiğinde, en son veri noktalarını inceler ve pencere adım boyutunu yeni ölçümler geldiğinde adımları zaman ekseni boyunca hareket ettirir. Bu teknik, hiç bitmeyen bir veri akışını saklamak zorunda kalmaması avantajına sahiptir, ancak aynı zamanda ölçümlerin yalnızca mevcut pencerede bulundukları sürece veri analizi için göz önüne alınabileceği anlamına da gelir. Genel olarak, sürgülü pencere filtresi bir veri akışındaki en son ölçümleri dikkate alır. Düşünülen ölçümler genellikle bir sınıflandırıcıya geçirilir. Mevcut sıranın bilinen bir kategoriye veya sınıfa atanması durumunda, ilgili bir eylem tetiklenir ve bir sonraki örtüşmeyen pencere, zaman ekseni boyunca adımlar ile incelenir. Bu geleneksel sürgülü pencere tekniğinin ana sınırlaması, büyümekte olan pencere boyutu, artan adım boyutu, daha yüksek örnekleme hızı ve daha büyük eğitim seti ile artan hesaplama karmaşıklığıdır.

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 sağlanır. Çeşitli zaman serisi uygulamaları için hava durumu, finans ve sağlık vb.

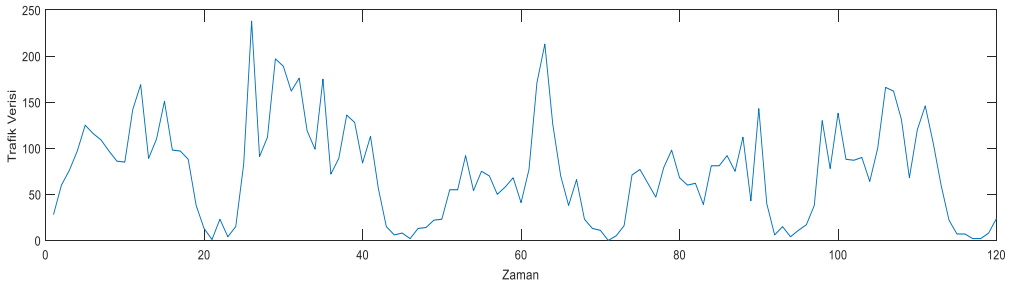
pencere kaydırma işlemi sıklıkla kullanılmaktadır. İlk değerden itibaren parçalar tanımlanır. İlk oluşturulan parçadan sonra verilen kritere göre sonraki parça işlenir. Bu işlem son parçada işlem yapılıncaya kadar devam eder. Bu metot sezgisel ve basittir. Temel amaç belirli miktarda verilen verinin tüm tahmini hatalarını azaltmaktır (Yahmed,2015). Bu tez çalışmasında, 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 gerçekleştirilmiş, sonuç olarak anomali tespiti yapılmış ve saldırı yapılan alanlar bulunmuştur. Şekil 3.4 de kullanılmış olan veri setindeki tüm veriler çizdirilmiş Şekil 3.5, Şekil 3.6, Şekil 3.7, Şekil 3.8, Şekil 3.9, Şekil 3.10, Şekil 3.11, Şekil 3.12, Şekil 3.13, Şekil 3.14, Şekil 3.15, Şekil 3.16'da ise sliding window(pencere kaydırma) yapılarak veriseti 12 segmente ayrılmış ve bu segmentlerin gösterimi verilmiştir.



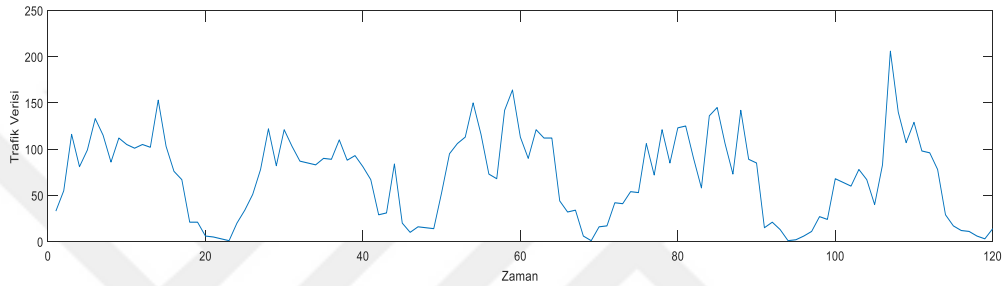
Şekil 3.4. İçerisinde anormal web trafiği olan pencerelere ayrılacak tam bir sinyal örüntüsü



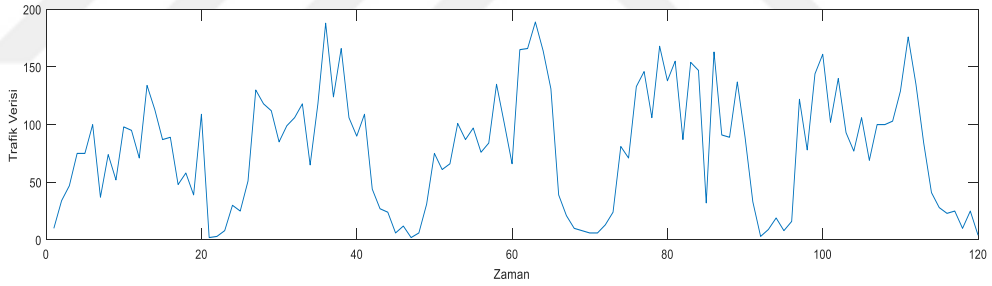
Şekil 3.5. Normal veri bulunan birinci pencereye ait sinyal örüntüsü



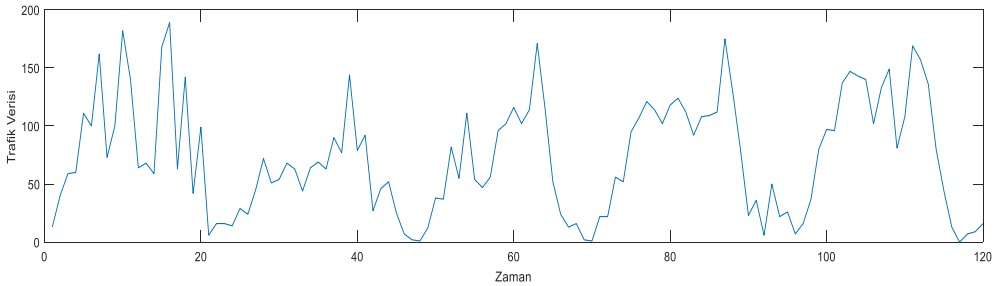
Şekil 3.6. Normal veri bulunan ikinci pencereye ait sinyal örüntüsü



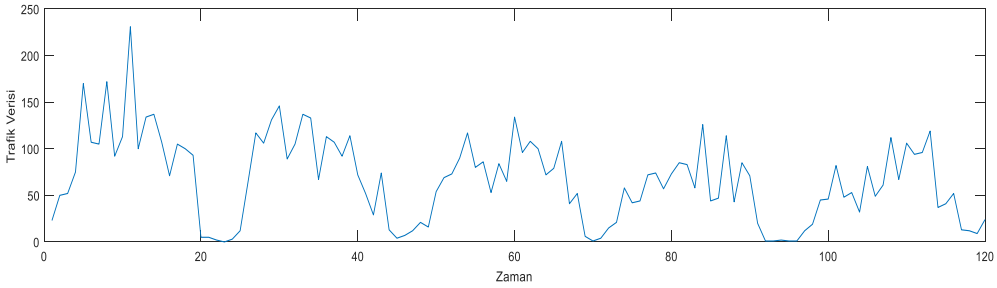
Şekil 3.7. Normal veri bulunan üçüncü pencereye ait sinyal örüntüsü



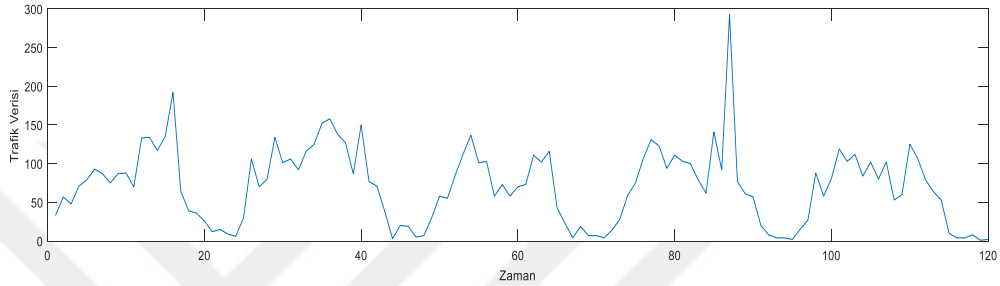
Şekil 3.8. Normal veri bulunan dördüncü pencereye ait sinyal örüntüsü



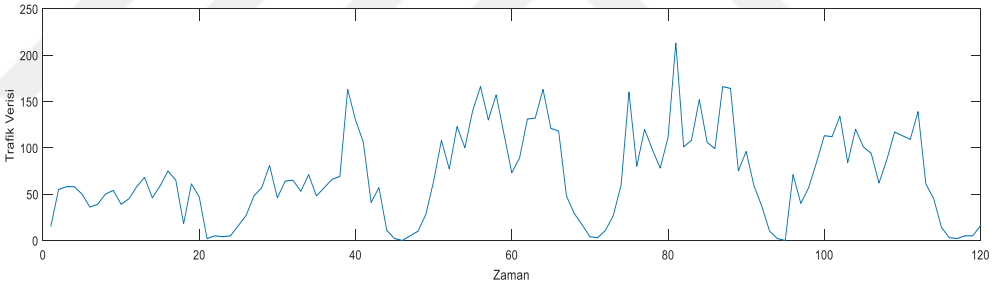
Şekil 3.9. Normal veri bulunan beşinci pencereye ait sinyal örüntüsü



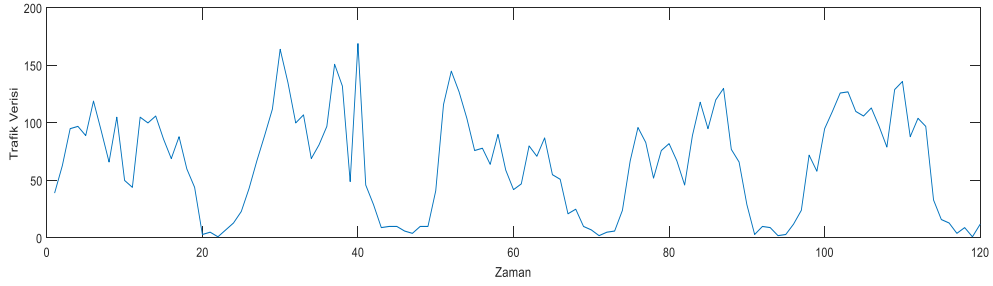
Şekil 3.10. Normal veri bulunan altıncı pencereye ait sinyal örüntüsü



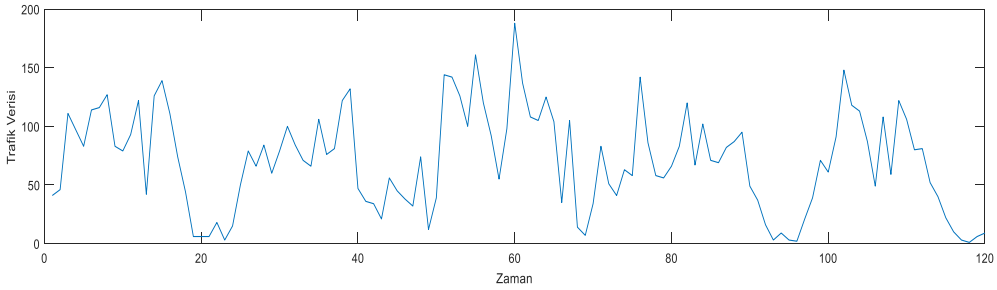
Şekil 3.11. Normal veri bulunan yedinci pencereye ait sinyal örüntüsü



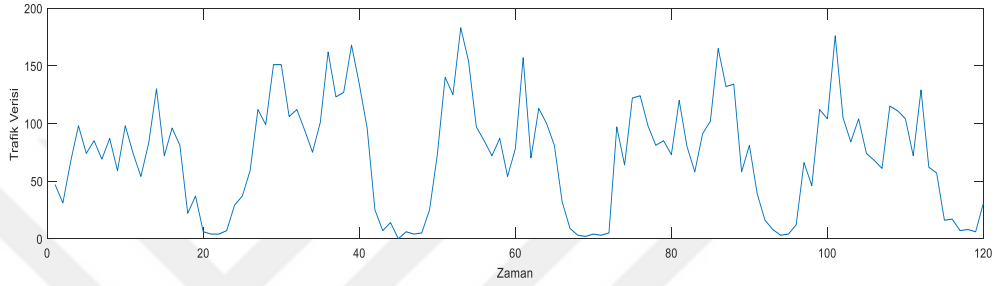
Şekil 3.12. Normal veri bulunan sekizinci pencereye ait sinyal örüntüsü



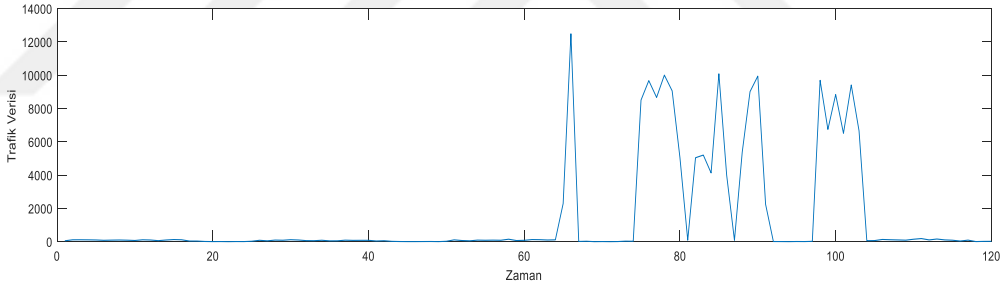
Şekil 3.13. Normal veri bulunan dokuzuncu pencereye ait sinyal örüntüsü



Şekil 3.14. Normal veri bulunan onuncu pencereye ait sinyal örüntüsü



Şekil 3.15. Normal veri bulunan onbirinci pencereye ait sinyal örüntüsü



Şekil 3.16. Anormal veri bulunan onikinci pencereye ait sinyal örüntüsü

3.4. Yapay Bağışıklık Sistemleri(YBS)

YBS alanında genel olarak, bağışıklık sistemi ve makine öğrenmesi üzerine Farmer vd. (1986) önemli miktarda araştırma yapmış, bu araştırmalar, 1990'ların başında biyolojik bağışıklık sistemini diğer alanlardaki sorunlara çözüm olarak ilham kaynağı olarak kullanan bir dizi bağımsız grup tarafından yürütülmüştür. YBS, doküman sınıflandırması, sahtekarlık tespiti, ağ ve ana bilgisayar tabanlı izinsiz giriş tespiti dahil çok çeşitli uygulama alanları için kullanılmaktadır. YBS'ler, uyguladıkları mekanizmaya bağlı olarak geniş bir şekilde iki kategoriye ayrılabilir: ağ temelli modeller ve popülasyon temelli modeller. Bu ayrımın yanında, birçok hibrit model de bulunmaktadır. Bu kategorilerden ilki, Jerne ve Towards (1974)'ün aptalpik ağ teorisine

dayanan sistemleri ifade eder. Bu YBS'ler, antikorlar ve antikorlar arasında olduğu kadar, antikorlar ve antijenler arasında da etkileşimlerin olduğunu kabul eden algoritmalar üzerine kuruludur. Nüfus temelli modeller, bir detektör popülasyonu üretme ve sürdürme yöntemi olarak negatif veya klonal seçim kullanır. Bu yaklaşım birkaç izinsiz giriş tespit sistemi oluşturmak için kullanılmıştır(Twycross, 2007).

İnsan bağışıklık sistemi, insan vücudunu lenfosit hücreleri kullanarak zararlı ve daha önce görülmeyen yabancı hücrelere karşı korur. Yabancı hücrelere bakteri ve virüs gibi antijenler denir. YBS, hesaplama sistemi için tasarlanmıştır ve insan bağışıklık sisteminden ilham almıştır. Bilgi güvenliği, özellikle izinsiz giriş tespiti alanındaki çeşitli problemlerin çözümünde uygulanır. Çeşitlilik, hata toleransı, dinamik öğrenme, adaptasyon ve öz-izleme dahil olmak üzere, insan bağışıklık sisteminin birçok özelliğini içerir. "öz" (sisteme ait hücreler) ve "öz olmayan" (sisteme yabancı varlıklar) arasında izinsiz girişler olarak ayrım yapar. Aynı şekilde, lenfositlere benzer dedektörler bilgisayar sisteminde konuşlandırılmıştır. Kötü amaçlı faaliyetleri engellemek ve bildirmek görevindedir (Hosseinpour vd., 2014).

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 (Castro vd., 2002). 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(Castrol vd., 2002).

Bağışıklık sistemi, çok sayıda farklı modeli tanıma, tanımlama ve cevaplama yeteneğine sahiptir. Ek olarak, bağışıklık sistemi kendi kendine işlev görmeyen ve kendi kendine zarar vermeyen hücreler arasında farklılıklar gösterebilir(Timmis, 2004).

Bağışıklık sisteminde çeşitliliğin üretilmesi ve korunmasında rol oynayan iki ana süreç vardır. Birincisi, gen kütüphanelerinden gen bölümlerinin birleşmesi yoluyla reseptör moleküllerinin üretilmesidir. Bir sonlu-kümeden genleri yeniden birleştirerek, bağışıklık sistemi neredeyse sonsuz sayıda değişen tipte reseptör üretme yeteneğine sahiptir, böylece bağışıklık sistemine antijenlerle ilgili geniş bir kapsama alanı kazandırır. Bağışıklık sistemindeki çeşitliliğe yardımcı olan ikinci süreç, somatik hipermutasyon olarak bilinir. Bağışıklık hücreleri kendilerini istilacı antijenlere cevap vermeden çoğaltırlar. Üreme sırasında, yeni reseptör molekülleri kalıplarının oluşturulmasına izin veren ve böylece bağışıklık reseptörlerinin çeşitliliğini artıran yüksek oranda somatik bir mutasyon işlemine tabi tutulurlar (Kepler ve Perelson, 1993).

Belirli bir antijene bağışıklık sistemi yanıt verdikten sonra, bazı hücre ve molekül kümeleri, aynı veya benzer antijenlerin gelecekteki enfeksiyonlarına daha hızlı ve daha güçlü yanıt sağlamak için yaşam alanlarını arttırırlar. Bağışıklık tepkisinin olgunlaşması olarak bilinen bu işlem, bu hücrelerin ve moleküllerin antijenleri tanımada başarılı bir şekilde sonuç almasını sağlar. Bu, tıpta aşılama prosedürlerinin ana prensibidir. Zayıflamış veya ölü bir antijen örneği, antijeni hafızaya almak için hafıza hücreleri ve moleküller üretmek üzere bir bağışıklık yanıtını (hastalık belirtisi olmadan) teşvik etmek üzere bir bireye aşılanır. Bazı hücre ve molekül kümeleri, aynı veya benzer antijenlerin gelecekteki enfeksiyonlarına daha hızlı ve daha güçlü bağışıklık tepkileri sağlamak için yaşam ömrünün uzamasını sağlar. Bağışıklık tepkisinin olgunlaşması olarak bilinen bu işlem, bu hücrelerin ve moleküllerin antijenleri tanımada başarılı bir şekilde rol oynamasını sağlar (Timmis, 2004).

Bağışıklık sistemi yüksek oranda dağınık, uyarlanabilir, doğada kendi kendini düzenleyen, geçmişte karşılaşılanların hatıralarını koruyan ve sürekli olarak yeni karşılaşmalar hakkında bilgi edinme yeteneğine sahiptir. Bağışıklık sistemi, bilgisayar bilimcileri ve mühendisleri için ilham kaynağı olacak birçok bilgi sunmaktadır. Hesaplamalı problemler karmaşıklaştıkça, insanlar bu sorunlara yeni yaklaşımlar aramakta ve çoğu zaman ilham almak için doğaya başvurulmaktadır. Omurgalı bağışıklık sistemi, potansiyel bir ilham kaynağı olarak büyük bir ilgi kaynağı olmakta,

burada farklı anlayışların ve alternatif çözümlerin biyolojik olarak esinlenilen diğer yöntemlerin üzerinde ve üstünde toplanabileceği düşünülmektedir. Bağışıklık sisteminde dikkat çeken bu yükselişi göz önüne alındığında, bu alanı biraz ayrıntılı bir şekilde araştırmak uygun görünmektedir(Timmis, 2004).

Canlılardaki bağışıklık sistemi genel olarak yorumlanıp belirli bir sistematiğe indirgenerek çalışmalarda kullanılmıştır. Günümüzde bu yöntem geliştirilerek çalışmalarda kullanılmaktadır. YBS, insanın bağışıklık sistemi temel alınarak geliştirilmiştir. 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 (Dandıl ve Güngör, 2012). Bu çalışmada yapay bağışıklığın Negatif Seçim Algoritması kullanılmıştır.

3.4.1. Negatif Seçim Algoritması

Forrest ve Perelson (1994) ilk NSA (Negatif Seçim Algoritması)'yı önerdiğinde normal ve anormal alanı temsil etmek için ikili kodlamayı kullanmıştır. Daha sonra gerçekten değerli bir yaklaşım sunulmuş ve anormal boşluğu kapatmak ve iyi dedektörler üretmek için genetik teknik önerilmiştir. NSA'lar hakkındaki son çalışmalar arasında, Zhou ve Dipankar (2004) gerçek değerler ile V-detektörü, Balachandran vd. (2007) çok şekilli bir dedektör negatif seçim algoritması önermiştir. NSA, YBS'deki en başarılı yöntemlerden biridir ve tipik uygulamaları arasında değişiklik tespiti, hata tespiti ve ağ girişi tespiti bulunmaktadır. NSA'nın alternatif yöntemlerden farklı bir işleme sahip olduğuna ve mevcut en etkili algoritma olduğuna inanılmaktadır. NSA'nın birçok başarılı uygulaması olmasına rağmen, YBS ve NSA'nın yoğun bir şekilde uygulanmasını önlemek için bazı problemler devam etmektedir (Jinquan vd., 2009).

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

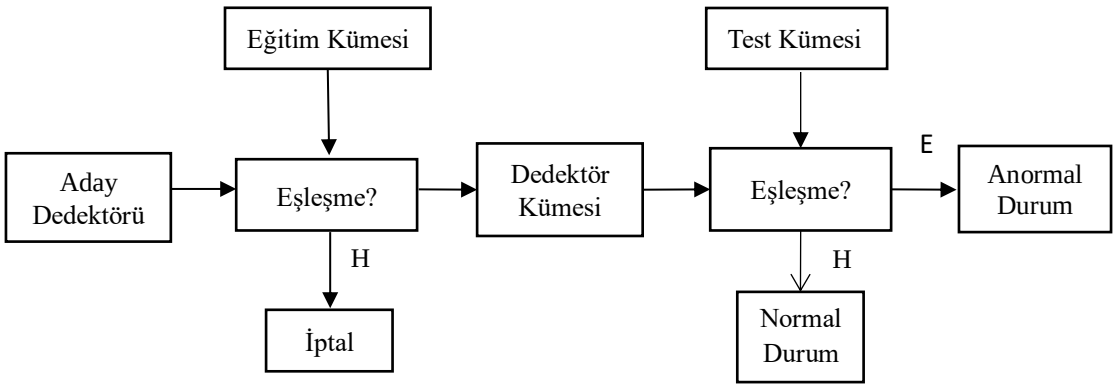
iii. Aday detektörlerden eğitim kümesi ile belirlenen eşik değerine göre eşleşenler aday detektör kümesinden çıkarılır. Eşleşmeyenler detektör kümesine atılarak eğitilir. Bir aday detektör ile eğitim veya test kümesi arasındaki eşleşmenin hesaplanması için Denklem 3.1' de verilen Öklid (Euclidian) mesafe ölçümü kullanılmıştır. Bu denklemde *A* bulunan mesafeyi, *l* data sayısını, Ab test veya eğitim kümesini, Ag ise detektör kümesini belirtmektedir.

$$A = \sqrt{\sum_{i=1}^{l}(Ab_i - Ag_i)^2} \tag{3.1}$$

iv. Test aşamasında, yine veriseti içerisinden test 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.

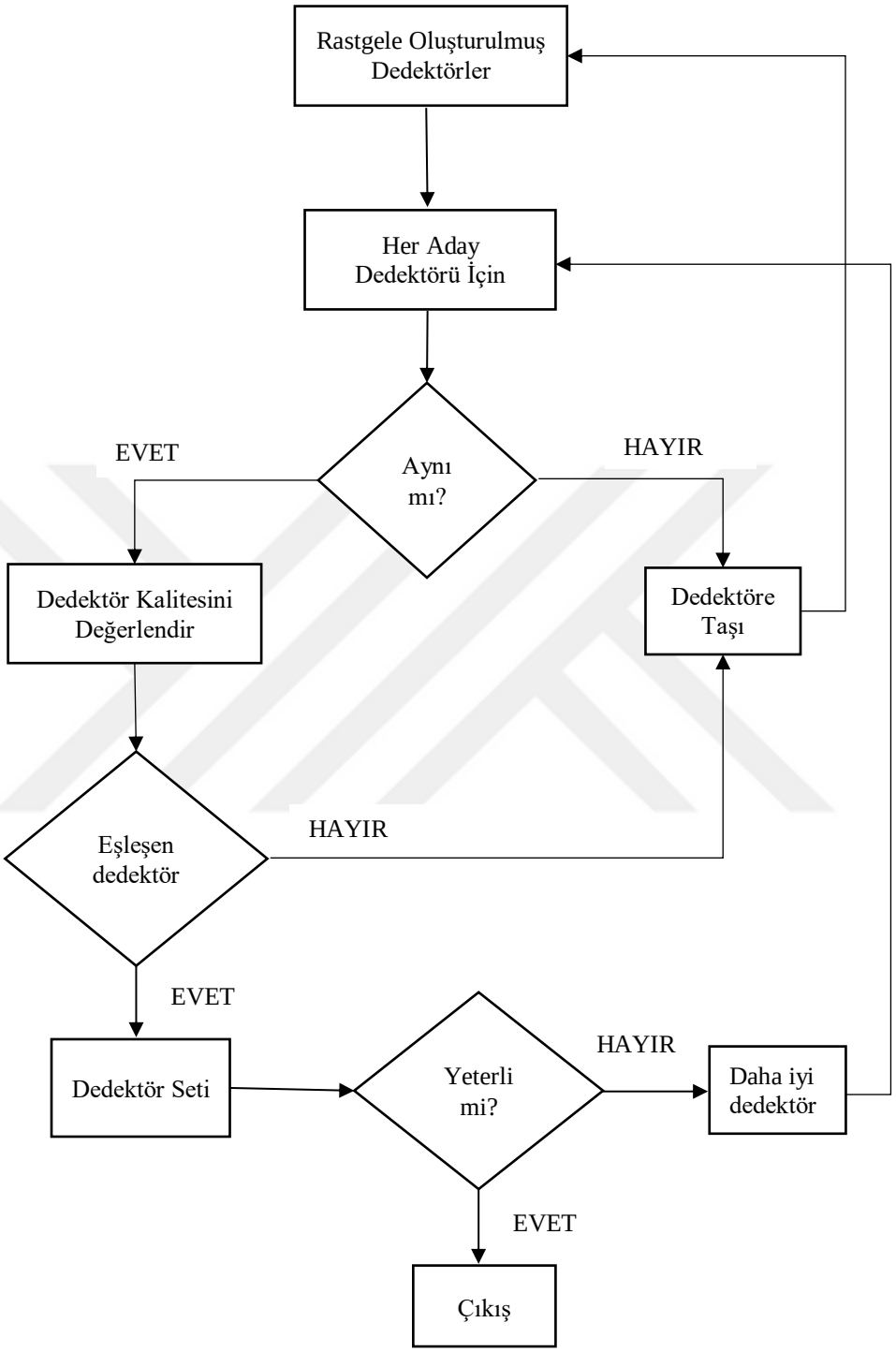
NSA'nın genel olarak işlem akışlarını gösteren şeması Şekil 3.17'de gösterilmiştir.



Şekil 3.17. Negatif seçim algoritmasının akış şeması.

NSA dedektörü üretimi, yinelemeli bir işlemle olgunlaşan bir aday dedektör popülasyonuyla başlar. Özellikle, her bir dedektörün merkezi rastgele seçilir ve yarıçap, dedektörün boyutunu (m-boyutlu alanda) belirleyen değişken bir parametredir. NSA

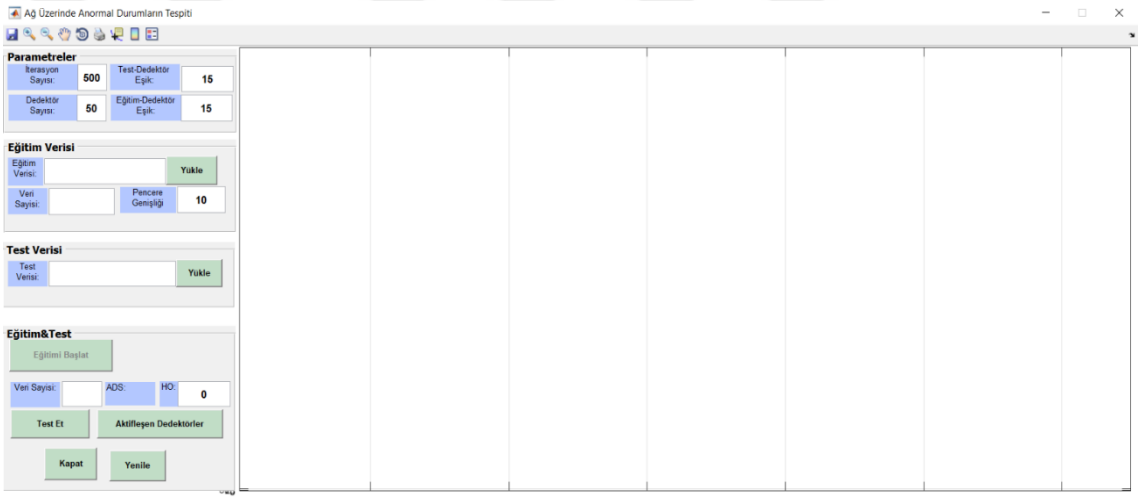
detektörü üretme algoritmasının temel algoritmik adımları, Şekil 3.18'de verilmiştir. Her bir yinelemede, ilk önce, her bir aday detektörün mesafesi hesaplanır ve kendi kendine düşen bölgelere girenler taşınır. Aynı olmayan dedektör seti daha sonra depolanır ve mesafelere göre sıralanır. Daha büyük mesafeye sahip (ve diğer dedektörlerle daha küçük çakışan) dedektörler daha uygun olarak kabul edilir ve bir sonraki nesle geçmek için seçilir. Ancak çok küçük mesafeye sahip dedektörler daha uygun dedektörlerin klonları tarafından değiştirilir. Seçilen bir dedektörün klonları yakın çevresinde yeni dedektörler üretmek için sabit bir mesafede hareket ettirilir. Dahası, bazı rastgele dedektörlerin tanıtılmasıyla, yeni alanlar keşfedilmiştir. Dedektör oluşturma işleminin tamamı, bir olgunluk kümesi (en az üst üste binmeyen) detektörler ile gelişir ve bu da kendisine ait alanın dışındaki alanları önemli ölçüde tespitine yaramaktadır.



Şekil 3.18. Negatif seçim algoritması akış diyagramı

4. GELİŞTİRİLEN UYGULAMA ve DENEYSEL ÇALIŞMALAR

Bu tez çalışmasında ağ üzerinde anormal trafiğin tespit edilmesi için MATLAB GUI ile tasarlanan arayüz Şekil 4.1' de gösterilmektedir. 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 4.1. Web trafik verilerinde anomali tespiti için geliştirilen yazılımın arayüzü

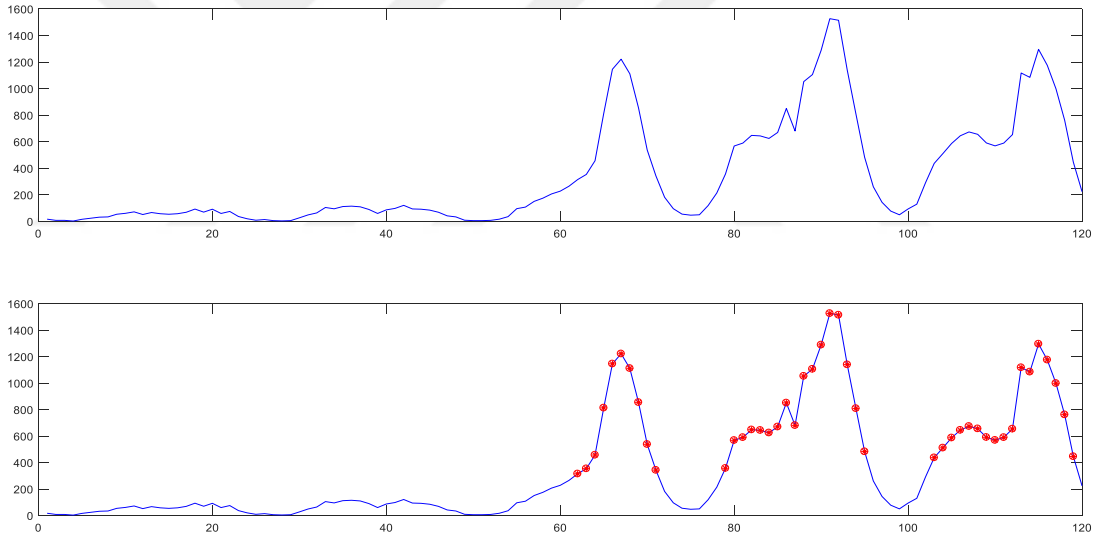
Tez çalışması kapsamında yapılan deneylerde YBS parametreleri(Eğitim Dedektör Eşik Sayısı, Test Dedektör Eşik Sayısı vs.) belirlendikten sonra daha önceden ayarlanmış olan eğitim ve test verilerine göre ağda oluşmuş olan anomali yazılım tarafından tespit edilmektedir. Ağda anormal trafik verilerinde oluşan hata yüzdesi ve bu aşamada gerçekleşen aktifleşen detektör sayısına göre belirlenir.

Uygulamanın doğru sonuç verdiğini göstermek adına Yahoo Webscope 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.

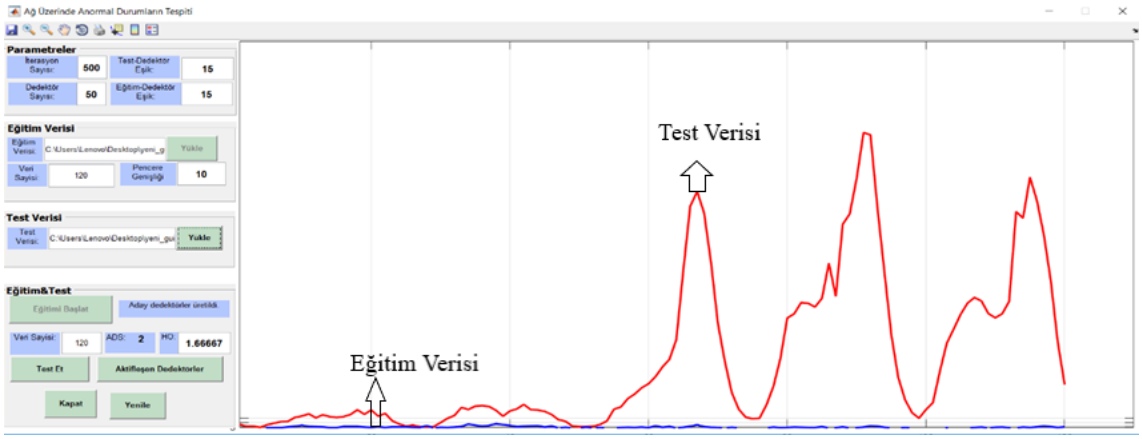
4.1. Deneysel Çalışma 1

Yahoo Webscope veriseti üzerinde ilk deneysel çalışma, verisetinde bulunan A1 gerçek veri sınıfına ait olan 42. sinyal örüntüsü(veri) ile gerçekleştirilmiştir. Şekil 4.2'de 42. veriye ait anormal trafik veri penceresi gösterilmiştir. Burada segmentlere ayrılmış 42. verinin son segmenti yani anormal trafiğin olduğu ve test verisi olarak kullanılan verilerden kırmızı yuvarlak ile işaretlenmiş olanlar anormal verileri göstermektedir.



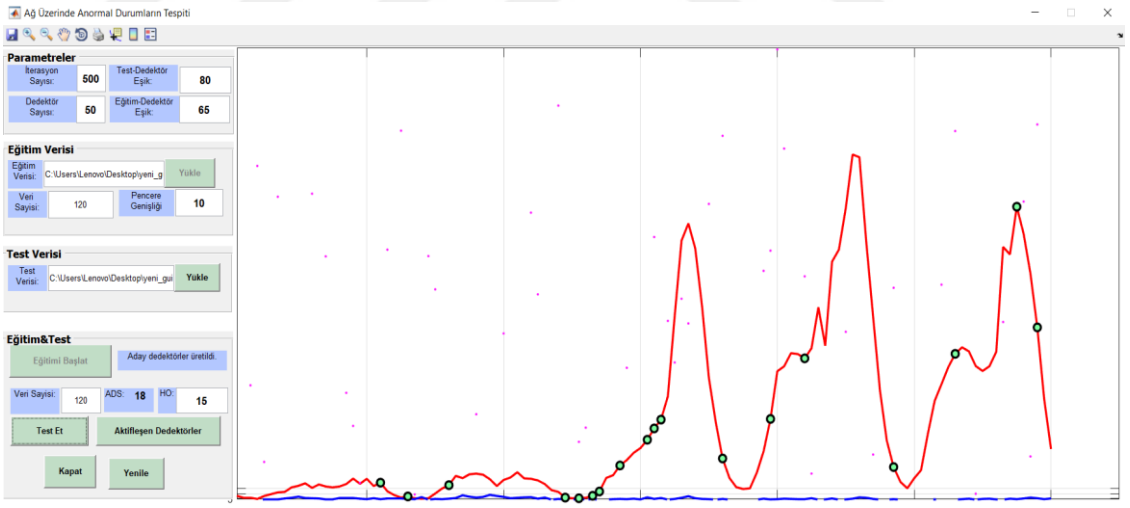
Şekil 4.2. Anormal web trafik verisinin bulunduğu sinyal örüntü penceresi

Birinci test işlemi 42. veri ile yapılan deneyi içermektedir.Anormalliklerin belirlenmesi için ilk aşama, Şekil 4.3'teki gibi gerçekleştirilen sisteme eğitim ve test verilerin yüklenmesidir. İkinci aşamada NSA algoritmasının parametrelerinin değerleri tanımlanır. Şekil 4.3'te 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



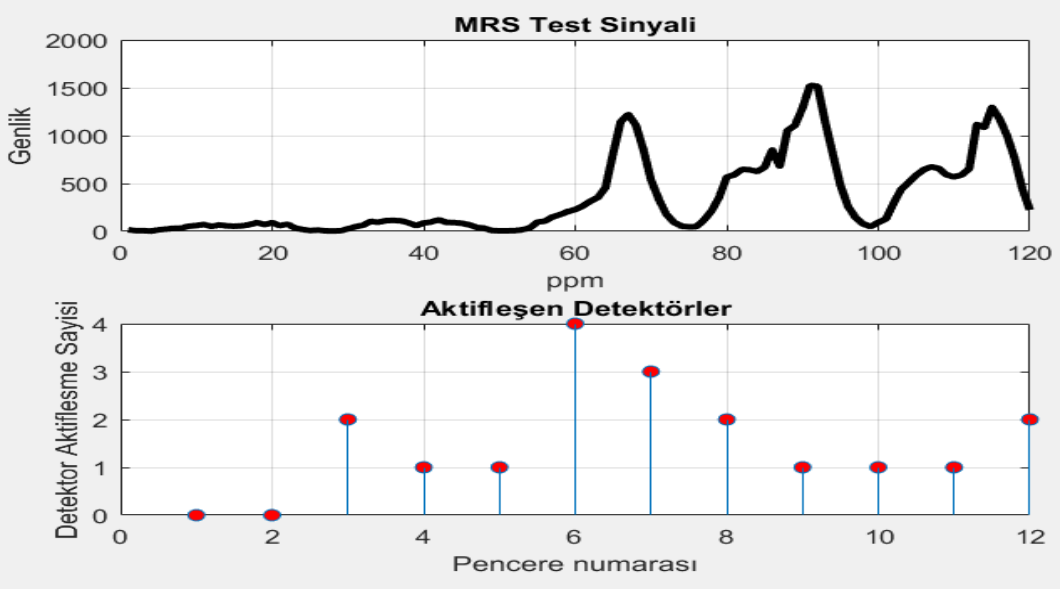
Şekil 4.3. Uygulama yazılımına eğitim ve test verilerinin yüklenmesi

Şekil 4.4'teki test işleminde eğitim verisi olarak 42. verinin ilk 120 verisi kullanılmış, test verisi olarak 42. verinin son 120 verisi yani anormal veriler kullanılmıştır. Tüm deneylerde iterasyon sayısı 500, dedektör sayısı 50 olarak belirlenmiştir.Eğitim dedektör eşik sayısı 65, test dedektör eşik sayısı ise 80 olarak belirlenmiştir.



Şekil 4.4. Test verisi üzerinde anormal trafik verilerinin NSA ile tespiti

Şekil 4.5'te 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 4.5. Anormal web verilerinin bulunduğu sinyalde aktifleşen detektörler

Çizelge 4.1'de görüldüğü gibi, 42. veri için normal ve anormal trafiğin tespiti için 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.

Çizelge 4.1. Deneysel Çalışma 1 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçları ve kullanılan parametrelerin değerleri

Test 1	**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**

Uygulama çalıştırıldığı zaman aktifleşen dedektörlerin anomali trafiği tespitinde göstermiş olduğu performans Çizelge 4.2'deki doğruluk çizelgesinde gösterilmiştir. Buradan da görülebileceği gibi, P12 penceresindeki 120 web trafik verilerindeki 76 normal verilerin tamamı normal (Doğru Pozitif, DP) olarak bulunurken, 44 anormal verinin ise 41 tanesi anormal (Doğru Negatif, DN) olarak tespit edilmiştir. Bunlara ek olarak, normal kategorideki tüm veriler doğru sınıflandırıldığı için anormal (Yanlış Negatif, YN) olarak bulunan veri olmamış, ancak anormal verilerden 3 tanesi Normal (Yanlış Pozitif, YP) olarak sınıflandırılmıştır. Sınıflandırma işleminin başarım değerlendirmesi Eşitlik 4.1'de gösterilen doğruluk metriği ile hesaplanmaktadır. dolayısıyla Çizelge 4.2' deki doğruluk çizelgesine göre Deneysel Çalışma 1 için sınıflandırma doğruluğu %97.50 olarak bulunur.

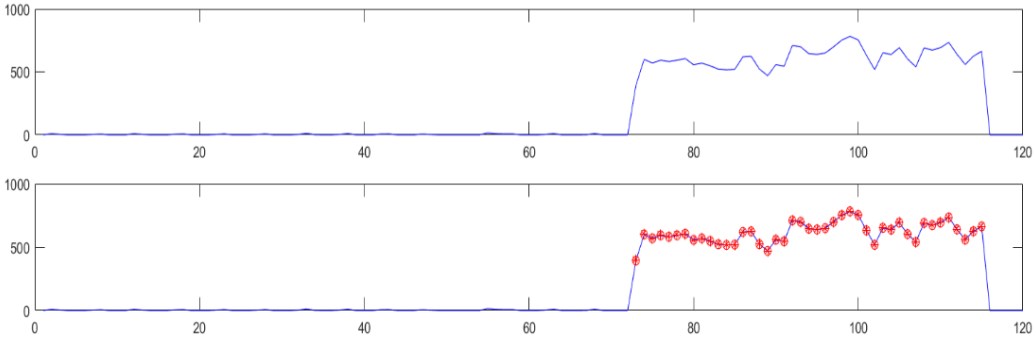
$$\text{Doğruluk (Accuracy)} = \frac{DP+DN}{DP+DN+YP+YN} \tag{4.1}$$

Çizelge 4.2. Deneysel Çalışma 1 için gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi

Doğruluk Çizelgesi	Tahmin Edilen **Normal**	Tahmin Edilen **Anormal**	**Toplam**
Gerçek **Normal**	**76 (DP)**	0 (YN)	76
Gerçek **Anormal**	3 (YP)	**41 (DN)**	44
Toplam	79	41	**120**

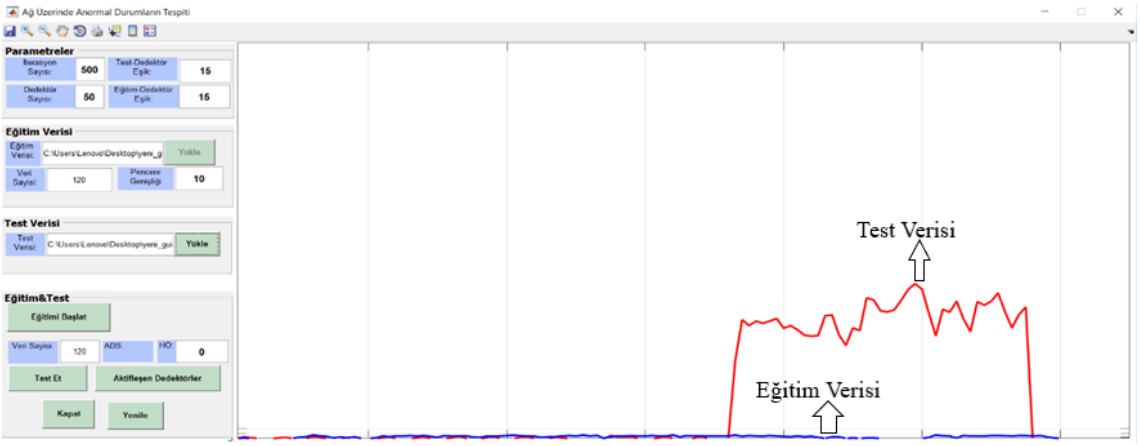
4.2. Deneysel Çalışma 2

Tez çalışması kapsamında web trafik verileri üzerinde ikinci test işlemi 58. sinyal örüntüsü ile gerçekleştirilmiştir. Şekil 4.6'da 58. veriye ait 12. pencereye için anormal trafik verileri gösterilmiştir. Burada pencerelere ayrılmış 58. verinin son penceresi yani anormal trafiğin olduğu test verisi olarak kullanılan noktalar kırmızı renkli işaretçi ile gösterilmiştir.



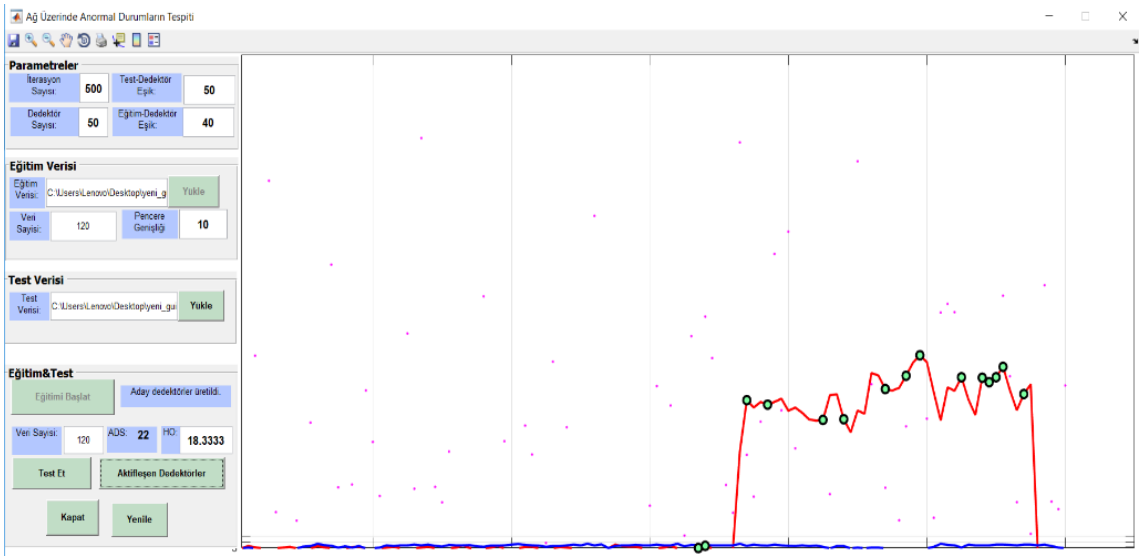
Şekil 4.6. Anormal web trafik verisinin bulunduğu sinyal penceresi (P12)

Deneysel Çalışma 2 için eğitim verisi olarak 58. sinyalin ilk 120 verisi (P1), test verisi olarak son 120 verisi (P12) kullanılarak geliştirilen uygulamaya yüklenmiş hali Şekil 4.7'de gösterilmiştir. Bu test işleminde eğitim dedektör eşik sayısı 40, test dedektör eşik sayısı ise 60 olarak belirlenmiştir.



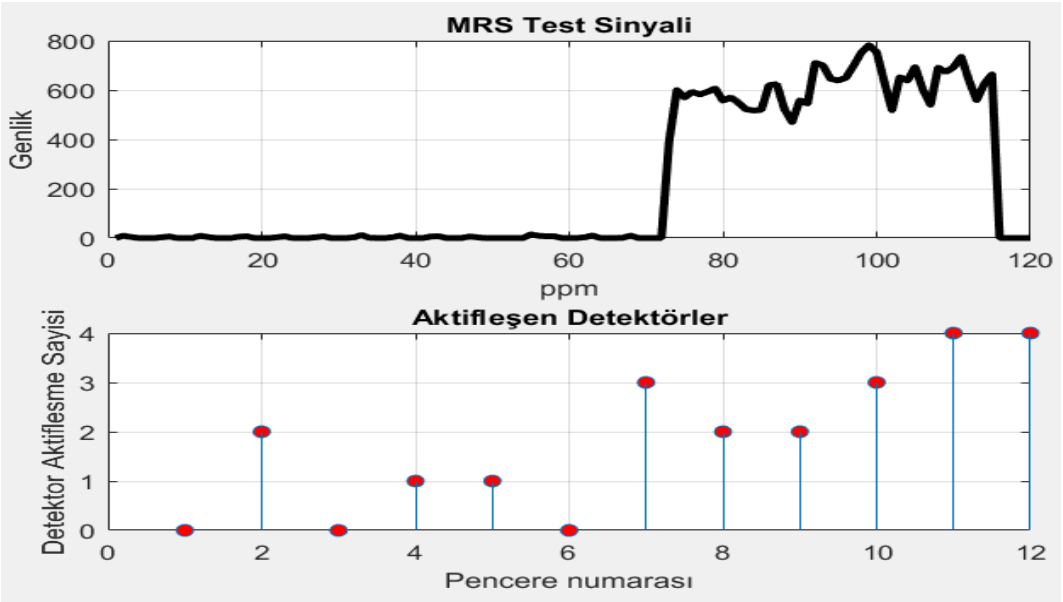
Şekil 4.7. Uygulama yazılımına Deneysel Çalışma 2 için eğitim ve test verilerinin yüklenmesi

Deneysel Çalışma 2 için 58. sinyaldeki anormal verilerin NSA ile tespit edilmiş hali Şekil 4.8'de sunulmuştur. Ağda anormal trafik verilerinde oluşan hata ve bu aşamada gerçekleşen aktifleşen detektör sayısı ile orantılı olarak değiştiği görülmektedir.



Şekil 4.8. Sinyal penceresi üzerinde anormal web verilerinin NSA ile tespit edilmesi

Şekil 4.9'da Deneysel Çalışma 2 için anormal trafiğin olduğu test verileri için aktifleşen dedektörler ve sinyal penreresindeki konumları gösterilmiştir. Yazılımda ağdaki verilerin incelenmesi sırasında oluşan aktifleşen detektörler anomaliyi belirleyen ana unsurdur. Aktifleşen detektörler ayrıca anomali verilerinin hangi zaman adımlarında hata olduğunu da göstermektedir. Buradan da görüleceği üzere, sinyal üzerinde anomalinin fazla olduğu bölgelerde aktifleşen dedektör sayısı da artmaktadır.



Şekil 4.9. Anormal web verilerinin bulunduğu sinyal penceresinde aktifleşen detektörler

Çizelge 4.3'te görüldüğü gibi, Deneysel Çalışma 2 için 58. Veri üzerinde normal ve anormal trafiğin tespiti için 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.

Çizelge 4.3. Deneysel Çalışma 2 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçları ve kullanılan parametrelerin değerleri

Test 2	**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**

Çizelge 4.4. Deneysel Çalışma 2 için 58. sinyal üzerinde gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi

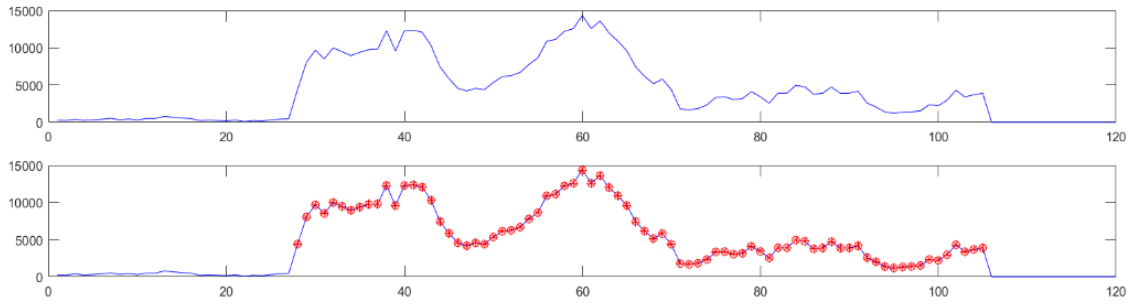
Doğruluk Çizelgesi	Tahmin Edilen **Normal**	Tahmin Edilen **Anormal**	**Toplam**
Gerçek **Normal**	**77 (DP)**	0 (YN)	77
Gerçek **Anormal**	2 (YP)	**41 (YP)**	43
Toplam	79	41	**120**

Deneysel Çalışma 2 için zaman aktifleşen dedektörlerin anomali trafiği tespitinde göstermiş olduğu performans Çizelge 4.4' te gösterilmiştir. Buradan da görülebileceği gibi, P12 penceresindeki 120 web trafik verilerindeki 77 normal verilerin tamamı Normal (Doğru Pozitif, DP) olarak bulunurken, 43 anormal verinin ise 41 tanesi

anormal (Doğru Negatif, DN) olarak tespit edilmiştir. Bunlara ek olarak, normal kategorideki tüm veriler doğru sınıflandırıldığı için anormal (Yanlış Negatif, YN) olarak bulunan veri olmamış, ancak anormal verilerden 2 tanesi normal (Yanlış Pozitif, YP)olarak sınıflandırılmıştır. Sınıflandırma işleminin başarımı ise, Deneysel Çalışma 2 için sınıflandırma doğruluğu %98.33 olarak bulunmuştur.

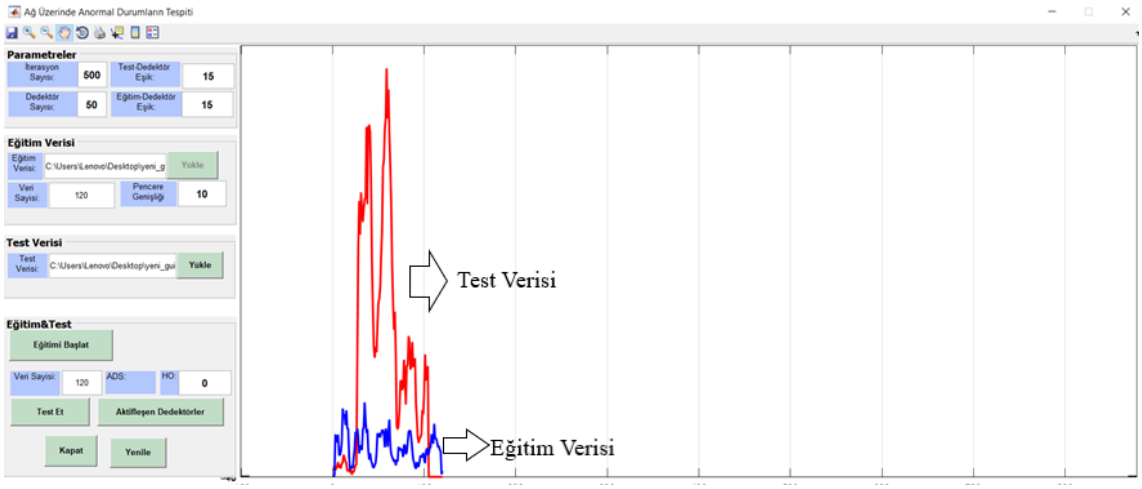
4.3. Deneysel Çalışma 3

Web trafik verileri üzerinde üçüncü test işlemi 17. sinyal örüntüsü ile gerçekleştirilmiştir. Şekil 4.10'da 17. veriye ait 12. pencereye için anormal trafik verileri gösterilmiştir. Burada pencerelere ayrılmış 17. verinin son penceresi yani anormal trafiğin olduğu test verisi olarak kullanılan noktalar kırmızı renkli işaretçi ile gösterilmiştir.



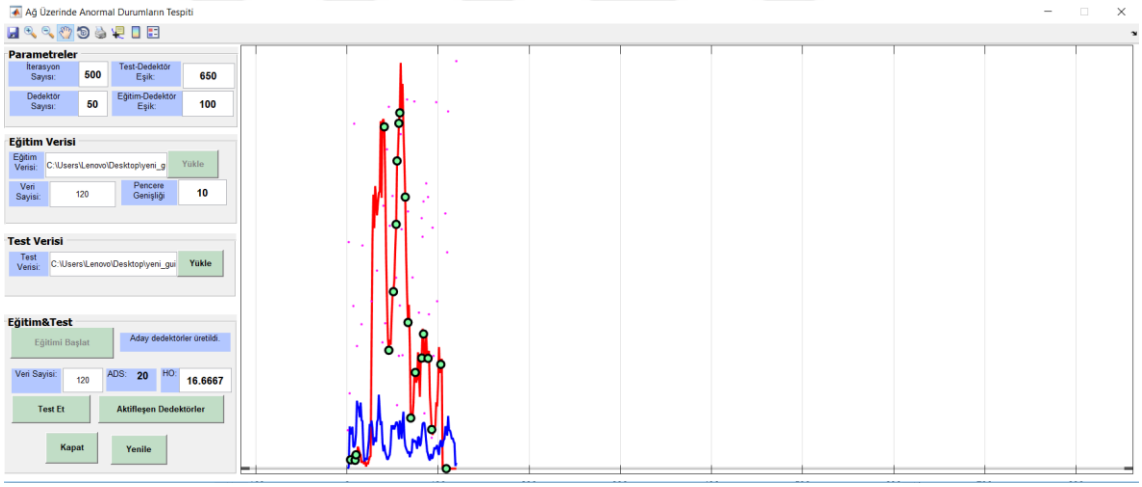
Şekil 4.10. Deneysel Çalışma 3 için anormal web trafik verisinin bulunduğu sinyal örüntüsü (P12)

Deneysel çalışma 3 için eğitim verisi olarak 17. sinyalin ilk 120 verisi (P1), test verisi olarak son 120 verisi (P12) kullanılarak geliştirilen uygulamaya yüklenmiş hali Şekil 4.11'de gösterilmiştir. Bu test işleminde eğitim dedektör eşik sayısı 10, test dedektör eşik sayısı ise 65 olarak belirlenmiştir.



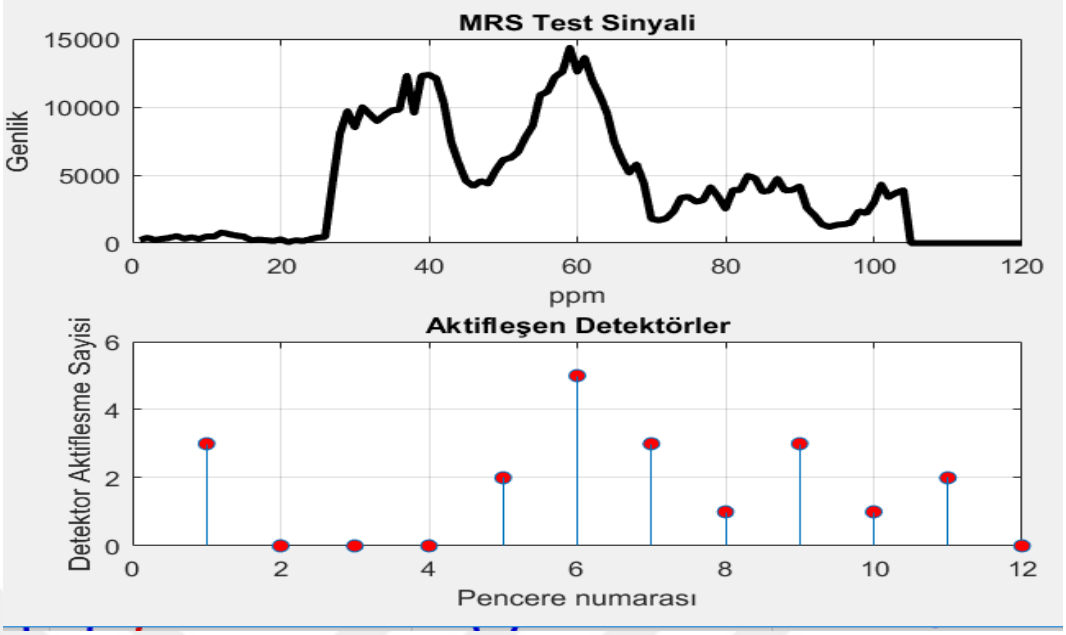
Şekil 4.11. Uygulama yazılımına Deneysel Çalışma 3 için eğitim (P1) ve test verilerinin(P12) yazılıma yüklenmesi

Deneysel çalışma 3 için 17. sinyaldeki anormal verilerin NSA ile tespit edilmiş hali Şekil 4.12'de sunulmuştur.



Şekil 4.12. Deneysel çalışma 3 için sinyal penceresi üzerinde anormal web verilerinin NSA ile tespit edilmesi

Şekil 4.13'te Deneysel Çalışma 3 için anormal trafiğin olduğu test verileri için aktifleşen dedektörler ve sinyal penreresindeki konumları gösterilmiştir. Aktifleşen detektörler ayrıca anomali verilerinin hangi, zaman adımlarında hata olduğunu da göstermektedir. Buradan da görüleceği üzere, sinyal üzerinde anomalinin fazla olduğu bölgelerde aktifleşen dedektör sayısı da artmaktadır.



Şekil 4.13. Deneysel çalışma 3 için anormal web verilerinin bulunduğu sinyalde aktifleşen detektörler (17. Sinyal örüntüsü)

Çizelge 4.5'te görüldüğü gibi, Deneysel Çalışma 3 için 17. veri üzerinde normal ve anormal trafiğin tespiti için eğitim verisiyle normal ve anormal trafik değerleri teste tabi tutulmuştur. Anormal trafik ile yapılan deneyde toplam anomali sayısı 79 iken aktifleşen dedektör sayısı 20, doğru bulunan anomali sayısı 74 ve doğruluk oranı %93.67 olarak bulunmuştur.

Çizelge 4.5. Deneysel Çalışma 3 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçları ve kullanılan parametrelerin değerleri

Test 3	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:17	P1	P2	120	0	0	0	100	10	65
	P1	P3	120	0	0	0	100	10	65
	P1	P4	120	0	0	0	100	10	65
	P1	P5	120	0	0	0	100	10	65
	P1	P6	120	0	0	0	100	10	65
	P1	P7	120	0	0	0	100	10	65
	P1	P8	120	0	0	0	100	10	65
	P1	P9	120	0	0	0	100	10	65
	P1	P10	120	0	0	0	100	10	65
	P1	P11	120	0	0	0	100	10	65
	P1	**P12**	**120**	**79**	**20**	**74**	**93.67**	**10**	**65**

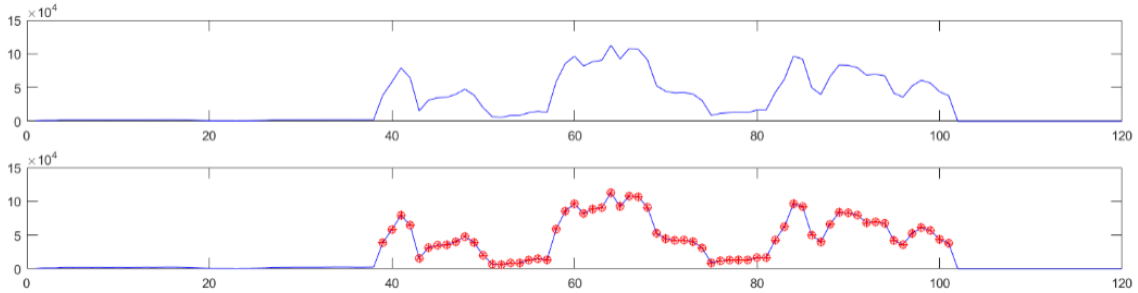
Çizelge 4.6. Deneysel Çalışma 3 için 17. sinyal üzerinde gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi

Doğruluk Çizelgesi	Tahmin Edilen **Normal**	Tahmin Edilen **Anormal**	**Toplam**
Gerçek **Normal**	**41 (DP)**	0 (YN)	41
Gerçek **Anormal**	5 (YP)	**74 (DN)**	79
Toplam	46	74	**120**

Deneysel Çalışma 3 için zaman aktifleşen dedektörlerin anomali trafiği tespitinde göstermiş olduğu performans Çizelge 4.6'da gösterilmiştir. Buradan da görülebileceği gibi, P12 penceresindeki 120 web trafik verilerindeki 41 normal verilerin tamamı normal (Doğru Pozitif, DP) olarak bulunurken, 79 anormal verinin ise 74 tanesi anormal (Doğru Negatif, DN) olarak tespit edilmiştir. Bunlara ek olarak, normal kategorideki tüm veriler doğru sınıflandırıldığı için anormal (Yanlış Negatif, YN) olarak bulunan veri olmamış, ancak anormal verilerden 5 tanesi normal (Yanlış Pozitif, YP) olarak sınıflandırılmıştır. Sınıflandırma işleminin başarımı ise Deneysel Çalışma 3 için sınıflandırma doğruluğu %95.83 olarak bulunmuştur.

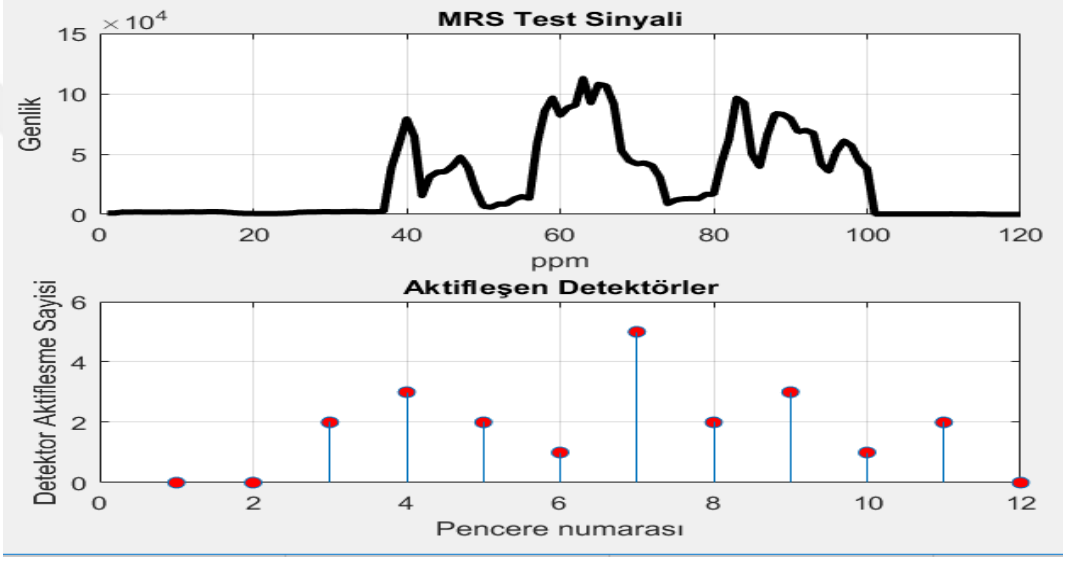
4.4. Deneysel Çalışma 4

Web trafik verileri üzerinde dördüncü test işlemi 22. sinyal örüntüsü ile gerçekleştirilmiştir. Şekil 4.14'te 22. veriye ait 12. pencereye için anormal trafik verileri gösterilmiştir. Burada pencerelere ayrılmış 22. verinin son penceresi yani anormal trafiğin olduğu test verisi olarak kullanılan noktalar kırmızı renkli işaretçi ile gösterilmiştir.



Şekil 4.14. Deneysel çalışma 4 için anormal web trafik verisinin bulunduğu sinyal örüntüsü (P12)

Deneysel çalışma 4 için eğitim verisi olarak 22. sinyalin ilk 120 verisi (P1), test verisi olarak son 120 verisi (P12) kullanılmıştır. Bu test işleminde eğitim dedektör eşik sayısı 35, test dedektör eşik sayısı ise 55 olarak belirlenmiştir. Şekil 4.15'te deneysel çalışma 4 için anormal trafiğin olduğu test verileri için aktifleşen dedektörler ve sinyal penreresindeki konumları gösterilmiştir. Aktifleşen detektörler ayrıca anomali verilerinin hangi, zaman adımlarında hata olduğunu da göstermektedir. Buradan da görüleceği üzere, sinyal üzerinde anomalinin fazla olduğu bölgelerde aktifleşen dedektör sayısı da artmaktadır.



Şekil 4.15. Deneysel çalışma 4 için anormal web verilerinin bulunduğu sinyalde aktifleşen detektörler (22. Sinyal örüntüsü)

Çizelge 4.7'de görülebileceği gibi, Deneysel Çalışma 4 için 22. veri üzerinde normal ve anormal trafiğin tespiti için eğitim verisiyle normal ve anormal trafik değerleri teste tabi tutulmuştur. Anormal trafik ile yapılan deneyde toplam anomali sayısı 63 iken aktifleşen dedektör sayısı 21, doğru bulunan anomali sayısı 59 ve doğruluk oranı %93.65 olarak bulunmuştur.

Çizelge 4.7. Deneysel Çalışma 4 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçların ve kullanılan parametrelerin değerleri

Test 4	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:22	P1	P2	120	0	0	0	100	35	55
	P1	P3	120	0	0	0	100	35	55
	P1	P4	120	0	0	0	100	35	55
	P1	P5	120	0	0	0	100	35	55
	P1	P6	120	0	0	0	100	35	55
	P1	P7	120	0	0	0	100	35	55
	P1	P8	120	0	0	0	100	35	55
	P1	P9	120	0	0	0	100	35	55
	P1	P10	120	0	0	0	100	35	55
	P1	P11	120	0	0	0	100	35	55
	P1	**P12**	**120**	**63**	**21**	**59**	**93.65**	**35**	**55**

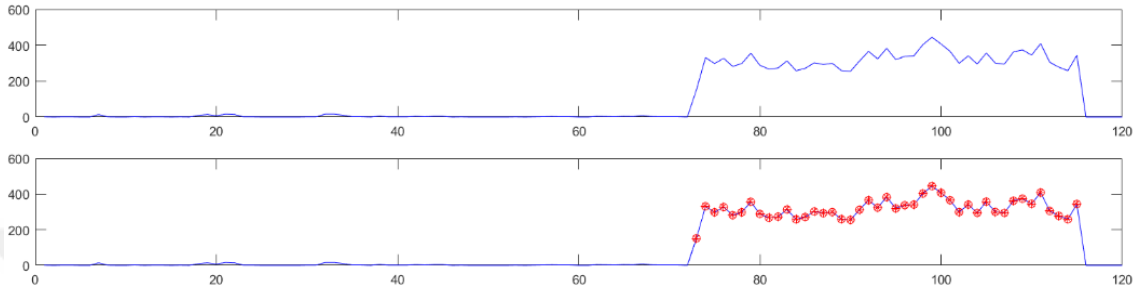
Deneysel Çalışma 4 için zaman aktifleşen dedektörlerin anomali trafiği tespitinde göstermiş olduğu performans Çizelge 4.8'de gösterilmiştir. Buradan da görülebileceği gibi, P12 penceresindeki 120 web trafik verilerindeki 57 normal verilerin tamamı normal (Doğru Pozitif, DP) olarak bulunurken, 63 anormal verinin ise 59 tanesi anormal (Doğru Negatif, DN) olarak tespit edilmiştir. Bunlara ek olarak, normal kategorideki tüm veriler doğru sınıflandırıldığı için anormal (Yanlış Negatif, YN) olarak bulunan veri olmamış, ancak anormal verilerden 4 tanesi normal (Yanlış Pozitif, YP) olarak sınıflandırılmıştır. Sınıflandırma işleminin başarımı ise Deneysel Çalışma 4 için sınıflandırma doğruluğu %96.67 olarak bulunmuştur.

Çizelge 4.8. Deneysel Çalışma 4 için 22. sinyal üzerinde gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi

Doğruluk Çizelgesi	Tahmin Edilen **Normal**	Tahmin Edilen **Anormal**	**Toplam**
Gerçek **Normal**	**57 (DP)**	0 (YN)	57
Gerçek **Anormal**	4 (YP)	**59 (DN)**	63
Toplam	61	59	**120**

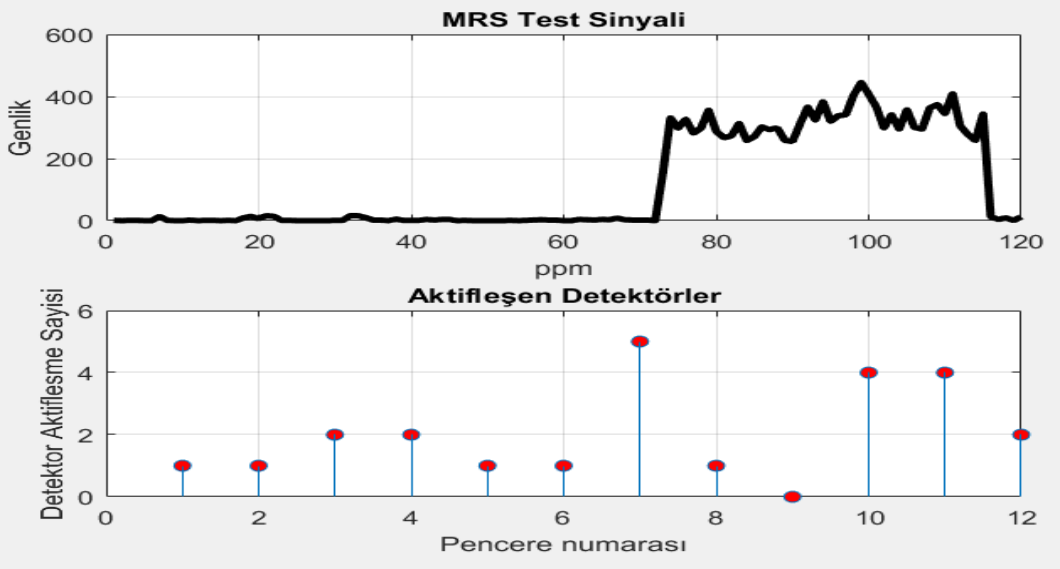
4.5. Deneysel Çalışma 5

Web trafik verileri üzerinde beşinci test işlemi 25. sinyal örüntüsü ile gerçekleştirilmiştir. Şekil 4.16' da 25. veriye ait 12. pencere için anormal trafik verileri gösterilmiştir. Burada pencerelere ayrılmış 25. verinin son penceresi yani anormal trafiğin olduğu test verisi olarak kullanılan noktalar kırmızı renkli işaretçi ile gösterilmiştir.



Şekil 4.16 Deneysel Çalışma 5 için anormal web trafik verisinin bulunduğu sinyal örüntüsü (P12)

Deneysel Çalışma 5 için eğitim verisi olarak 25. sinyalin ilk 120 verisi (P1), test verisi olarak son 120 verisi (P12) kullanılmıştır. Bu test işleminde eğitim dedektör eşik sayısı 15, test dedektör eşik sayısı ise 35 olarak belirlenmiştir. Şekil 4.17'de Deneysel Çalışma 5 için anormal trafiğin olduğu test verileri için aktifleşen dedektörler ve sinyal penreresindeki konumları gösterilmiştir. Aktifleşen detektörler ayrıca anomali verilerinin hangi, zaman adımlarında hata olduğunu da göstermektedir. Buradan da görüleceği üzere, sinyal üzerinde anomalinin fazla olduğu bölgelerde aktifleşen dedektör sayısı da artmaktadır.



Şekil 4.17. Deneysel Çalışma 5 için anormal web verilerinin bulunduğu sinyalde aktifleşen detektörler (25. Sinyal örüntüsü)

Çizelge 4.9'da görüldüğü gibi, Deneysel Çalışma 5 için 25. veri üzerinde normal ve anormal trafiğin tespiti için eğitim verisiyle normal ve anormal trafik değerleri teste tabi tutulmuştur. Anormal trafik ile yapılan deneyde toplam anomali sayısı 42 iken aktifleşen dedektör sayısı 24, doğru bulunan anomali sayısı 40 ve doğruluk oranı %95.20 olarak bulunmuştur.

Deneysel Çalışma 5 için zaman aktifleşen dedektörlerin anomali trafiği tespitinde göstermiş olduğu performans Çizelge 4.10' da gösterilmiştir. Buradan da görülebileceği gibi, P12 penceresindeki 120 web trafik verilerindeki 78 normal verinin tamamı normal (Doğru Pozitif , DP) olarak bulunurken, 42 anormal verinin ise 40 tanesi anormal (Doğru Negatif, DN) olarak tespit edilmiştir. Bunlara ek olarak, normal kategorideki tüm veriler doğru sınıflandırıldığı için anormal (Yanlış Negatif, YN) olarak bulunan veri olmamış, ancak anormal verilerden 2 tanesi Normal (Yanlış Pozitif, YP) olarak sınıflandırılmıştır. Sınıflandırma işleminin başarımı ise Deneysel Çalışma 5 için sınıflandırma doğruluğu %98.33 olarak bulunmuştur.

Çizelge 4.9. Deneysel çalışma 5 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçları ve kullanılan parametrelerin değerleri

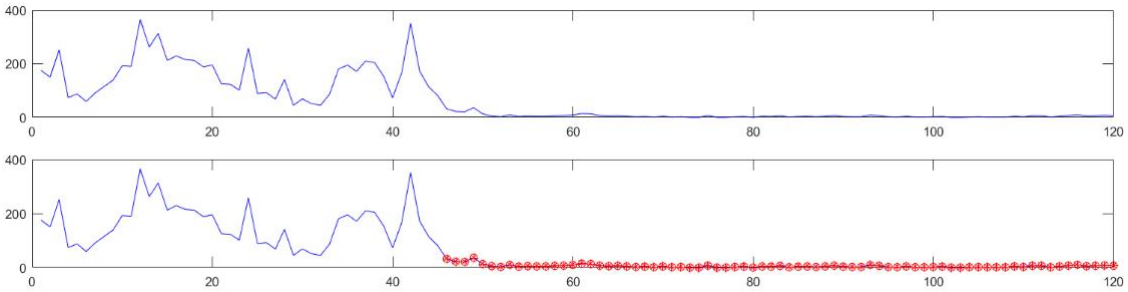
Test 5	Eğitim Verisi	Test Verisi	Toplam Veri	Anomali Sayısı	Aktf. Det. Say	Doğru Bul. Ano. Say.	Doğruluk Oranı	Eğitim Eşik Değeri	Test Eşik Değeri
Veri No:25	P1	P2	120	0	0	0	100	15	35
	P1	P3	120	0	0	0	100	15	35
	P1	P4	120	0	0	0	100	15	35
	P1	P5	120	0	0	0	100	15	35
	P1	P6	120	0	0	0	100	15	35
	P1	P7	120	0	0	0	100	15	35
	P1	P8	120	0	0	0	100	15	35
	P1	P9	120	0	0	0	100	15	35
	P1	P10	120	0	0	0	100	15	35
	P1	P11	120	0	0	0	100	15	35
	P1	**P12**	**120**	**42**	**24**	**40**	**95.20**	**15**	**35**

Çizelge 4.10. Deneysel çalışma 5 için 25. sinyal üzerinde gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi

Doğruluk Çizelgesi	Tahmin Edilen **Normal**	Tahmin Edilen **Anormal**	**Toplam**
Gerçek **Normal**	**78 (DP)**	0 (YN)	78
Gerçek **Anormal**	2 (YP)	**40 (DN)**	42
Toplam	80	40	**120**

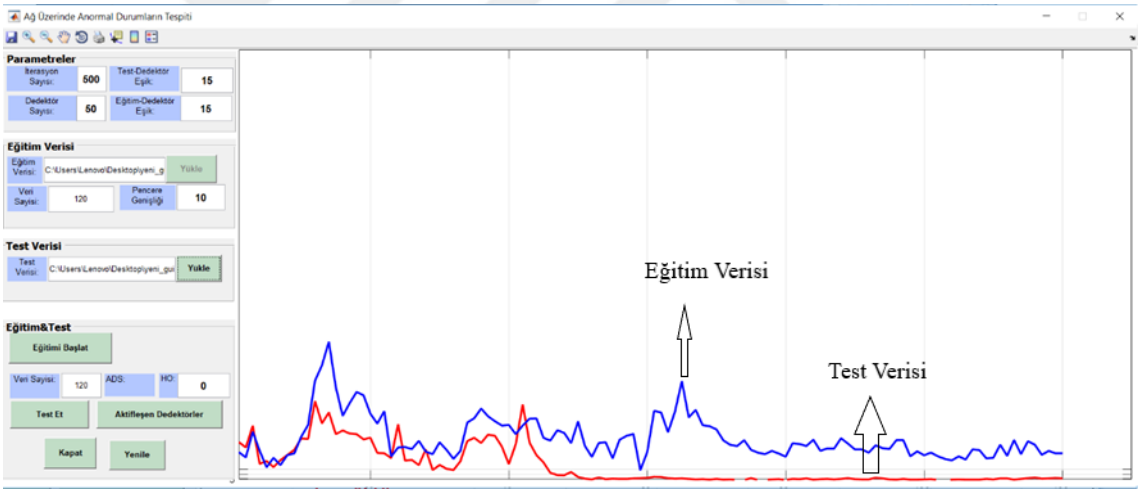
4.6. Deneysel Çalışma 6

Web trafik verileri üzerinde altıncı test işlemi 40. sinyal örüntüsü ile gerçekleştirilmiştir. Şekil 4.18'de 58. veriye ait 10. pencereye için anormal trafik verileri gösterilmiştir. Burada pencerelere ayrılmış 40. verinin anormal trafiğin olduğu test verisi olarak kullanılan noktalar kırmızı renkli işaretçi ile gösterilmiştir.

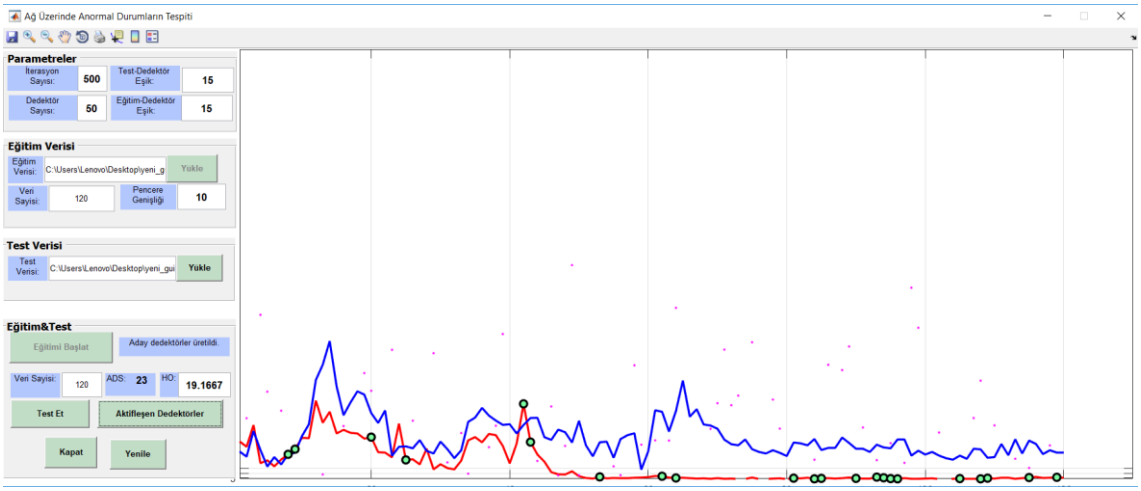


Şekil 4.18. Deneysel Çalışma 6 için 40. sinyal üzerinde anormal web trafik verisinin bulunduğu sinyal örüntüsü (P10)

Deneysel Çalışma 6 için eğitim verisi olarak 40. sinyalin ilk 120 verisi (P1), test verisi olarak 10. penceredeki 120 veri (P10) kullanılarak geliştirilen uygulamaya yüklenmiş hali Şekil 4.19'da gösterilmiştir. Bu test işleminde eğitim dedektör eşik sayısı 15, test dedektör eşik sayısı ise 15 olarak belirlenmiştir. Deneysel Çalışma 6 için 40. sinyaldeki anormal verilerin NSA ile tespit edilmiş hali ise Şekil 4.20' de gösterilmiştir.

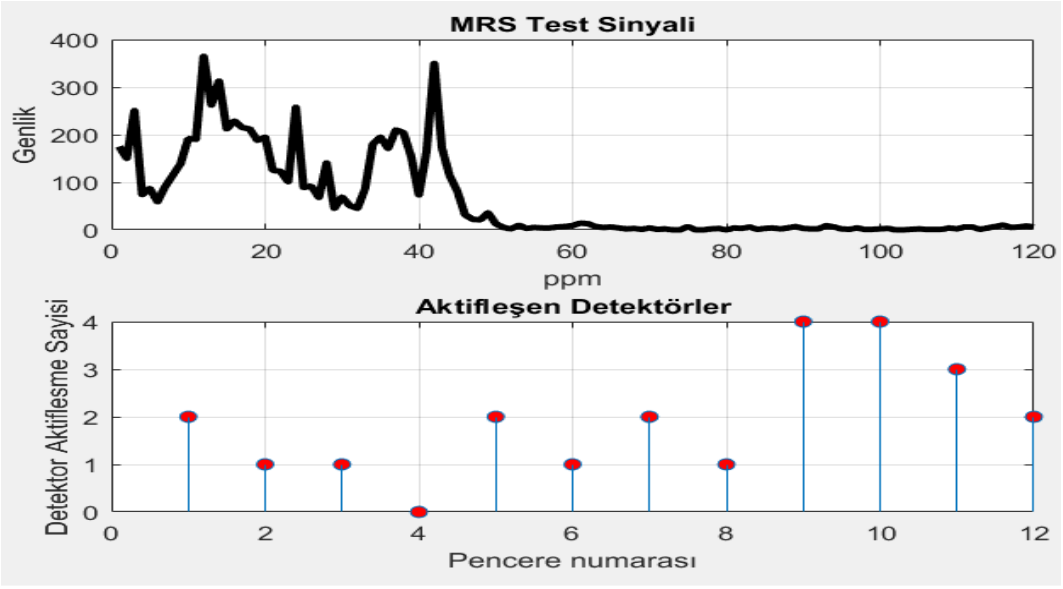


Şekil 4.19. Uygulama yazılımına Deneysel Çalışma 6 için eğitim (P1) ve test verilerinin(P10) yazılıma yüklenmesi



Şekil 4.20. Deneysel Çalışma 6 için sinyal penceresi üzerinde anormal web verilerinin NSA ile tespit edilmesi

Şekil 4.21'de deneysel çalışma 6 için anormal trafiğin olduğu test verileri için aktifleşen dedektörler ve sinyal penreresindeki konumları gösterilmiştir. Aktifleşen detektörler ayrıca anomali verilerinin hangi, zaman adımlarında hata olduğunu da göstermektedir. Buradan da görüleceği üzere, sinyal üzerinde anomalinin fazla olduğu bölgelerde aktifleşen dedektör sayısı da artmaktadır.



Şekil 4.21. Deneysel Çalışma 6 için anormal web verilerinin bulunduğu sinyalde aktifleşen detektörler (40. Sinyal örüntüsü)

Çizelge 4.11'de Deneysel Çalışma 6 için 40. 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 76 iken aktifleşen dedektör sayısı 23, doğru bulunan anomali sayısı 71 ve doğruluk oranı %93.42 olarak elde edilmiştir.

Çizelge 4.11. Deneysel Çalışma 6 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçları ve kullanılan parametrelerin değerleri

Test 6	**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:40	P1	P2	120	0	0	0	100	15	15
	P1	P3	120	0	0	0	100	15	15
	P1	P4	120	0	0	0	100	15	15
	P1	P5	120	0	0	0	100	15	15
	P1	P6	120	0	0	0	100	15	15
	P1	P7	120	0	0	0	100	15	15
	P1	P8	120	0	0	0	100	15	15
	P1	P9	120	0	0	0	100	15	15
	P1	**P10**	**120**	**76**	**23**	**71**	**93.42**	**15**	**15**
	P1	P11	120	0	0	0	100	15	15
	P1	P12	120	0	0	0	100	15	15

Çizelge 4.12. Deneysel çalışma 6 için 40. sinyal üzerinde gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi

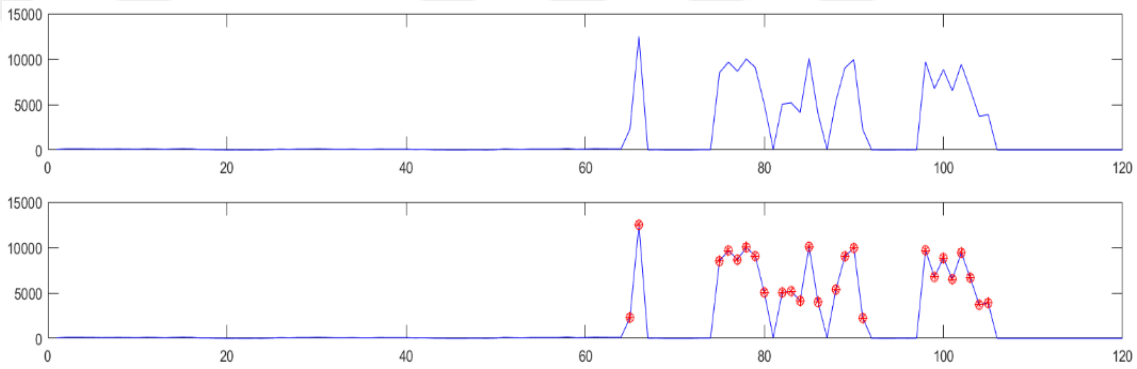
Doğruluk Çizelgesi	Tahmin Edilen **Normal**	Tahmin Edilen **Anormal**	**Toplam**
Gerçek **Normal**	**44 (DP)**	0 (YN)	44
Gerçek **Anormal**	5 (YP)	**71 (DN)**	76
Toplam	49	71	**120**

Deneysel Çalışma 6 için zaman aktifleşen dedektörlerin anomali trafiği tespitinde göstermiş olduğu performans Çizelge 4.12' de gösterilmiştir. Çizelde, P10 penceresindeki 120 web trafik verilerindeki 44 Normal verilerin tamamı normal (Doğru Pozitif, DP) olarak bulunurken, 76 anormal verinin ise 71 tanesi anormal (Doğru

Negatif, DN) olarak tespit edilmiştir. Bunlara ek olarak, normal kategorideki tüm veriler doğru sınıflandırıldığı için anormal (Yanlış Negatif, YN) olarak bulunan veri olmamış, ancak anormal verilerden 5 tanesi normal (Yanlış Pozitif , YP) olarak sınıflandırılmıştır. Sınıflandırma işleminin başarımı ise Deneysel Çalışma 6 için sınıflandırma doğruluğu %95.83 olarak bulunmuştur.

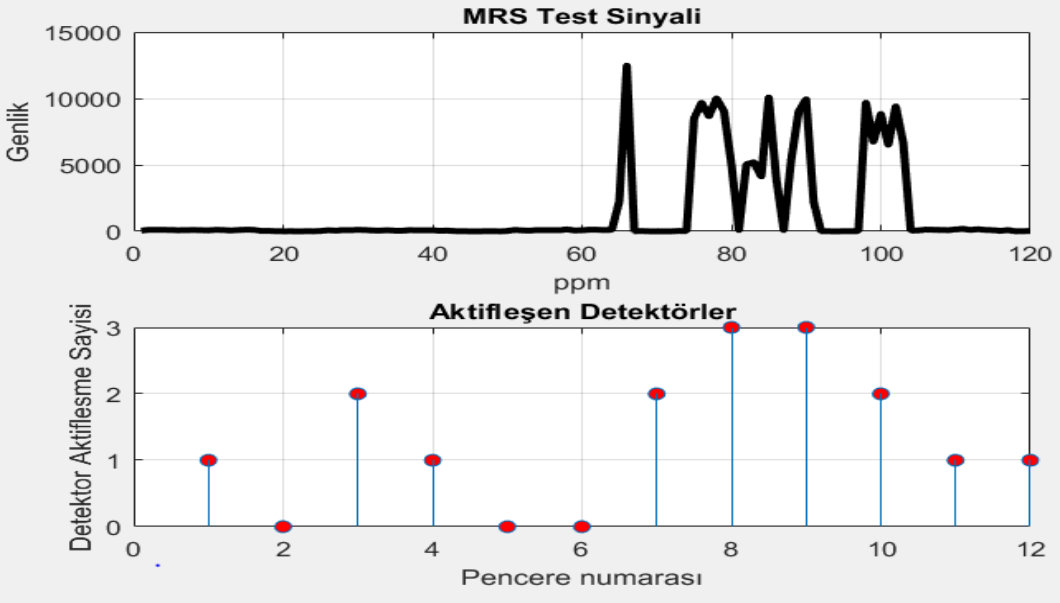
4.7. Deneysel Çalışma 7

Web trafik verileri üzerinde yedinci test işlemi 67. sinyal örüntüsü ile gerçekleştirilmiştir. Şekil 4.22'de 67. veriye ait 12. pencere için anormal trafik verileri gösterilmiştir. Burada pencerelere ayrılmış 67. verinin anormal trafiğin olduğu test verisi olarak kullanılan noktalar kırmızı renkli işaretçi ile gösterilmiştir.



Şekil 4.22. Deneysel Çalışma 7 için 67. sinyal üzerinde anormal web trafik verisinin bulunduğu sinyal örüntüsü (P12)

Deneysel Çalışma 7 için eğitim verisi olarak 67. sinyalin onbirinci 120 verisi (P11), test verisi olarak son 120 verisi (P12) kullanılmıştır. Bu test işleminde eğitim dedektör eşik sayısı 20, test dedektör eşik sayısı ise 60 olarak belirlenmiştir. Şekil 4.23'te Deneysel Çalışma 7 için anormal trafiğin olduğu test verileri için aktifleşen dedektörler ve sinyal penreresindeki konumları gösterilmiştir. Aktifleşen detektörler ayrıca anomali verilerinin hangi, zaman adımlarında hata olduğunu da göstermektedir. Buradan da görüleceği üzere, sinyal üzerinde anomalinin fazla olduğu bölgelerde aktifleşen dedektör sayısı da artmaktadır.



Şekil 4.23. Deneysel Çalışma 7 için anormal web verilerinin bulunduğu sinyalde aktifleşen detektörler (22. Sinyal örüntüsü)

Çizelge 4.13'te görüldüğü gibi, Deneysel Çalışma 7 için 67. veri üzerinde normal ve anormal trafiğin tespiti için eğitim verisiyle normal ve anormal trafik değerleri teste tabi tutulmuştur. Anormal trafik ile yapılan deneyde toplam anomali sayısı 23 iken aktifleşen dedektör sayısı 16, doğru bulunan anomali sayısı 22 ve doğruluk oranı %95.65 olarak bulunmuştur.

Deneysel Çalışma 7 için zaman aktifleşen dedektörlerin anomali trafiği tespitinde göstermiş olduğu performans Çizelge 4.14' te gösterilmiştir. Çizelgeden görülebileceği gibi, P12 penceresindeki 120 web trafik verilerindeki 97 normal verilerin 96 tanesi normal (Doğru Pozitif, DP) olarak bulunurken, 23 Anormal verinin ise 22 tanesi anormal (Doğru Negatif, DN) olarak tespit edilmiştir. Bunlara ek olarak, normal kategorideki 1 veri doğru sınıflandırılmadığı için anormal (Yanlış Negatif, YN) değeri 1 ve anormal verilerden 1 tanesi normal (Yanlış Pozitif, YP) olarak sınıflandırılmıştır. Sınıflandırma işleminin başarımı ise deneysel çalışma 7 için sınıflandırma doğruluğu %98.33 olarak bulunmuştur.

Çizelge 4.13. Deneysel Çalışma 7 için anormal web verilerinin bulunduğu pencere üzerinde elde edilen sonuçların ve kullanılan parametrelerin değerleri

Test 7	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:67	P11	P1	120	0	0	0	100	20	60
	P11	P2	120	0	0	0	100	20	60
	P11	P3	120	0	0	0	100	20	60
	P11	P4	120	0	0	0	100	20	60
	P11	P5	120	0	0	0	100	20	60
	P11	P6	120	0	0	0	100	20	60
	P11	P7	120	0	0	0	100	20	60
	P11	P8	120	0	0	0	100	20	60
	P11	P9	120	0	0	0	100	20	60
	P11	P10	120	0	0	0	100	20	60
	P11	P11	120	0	0	0	100	20	60
	P11	**P12**	**120**	**23**	**16**	**22**	**95.65**	**20**	**60**

Çizelge 4.14. Deneysel Çalışma 7 için 67. sinyal üzerinde gerçek ve tahmin edilen normal ve anormal web verilerinin doğruluk çizelgesi

Doğruluk Çizelgesi	Tahmin Edilen **Normal**	Tahmin Edilen **Anormal**	**Toplam**
Gerçek **Normal**	**96 (DP)**	1 (YN)	97
Gerçek **Anormal**	1 (YP)	**22 (DN)**	23
Toplam	97	23	**120**

4.8. Deneysel Çalışmalar Üzerinde Genel Değerlendirmeler

Bu tez çalışması kapsamında geliştirilen bir uygulama üzerinde, Yahoo Webscope S5 veriseti ile web trafik verileride anomali tespiti için NSA destekli bir yaklaşım önerilmiştir. Deneysel çalışmalar kapsamında verisetinde bulunan ve gerçek veriler olan 17., 22., 25., 40., 42.,58. ve 67. olmak üzere zaman serileri şeklinde sunulan sinyal verileri kullanılmıştır. Herbir veri üzerinde elde edilen sonuçları gösteren özet tablo Çizelge 4.15'te sunulmuştur. Bu özet tablodan da görülebileceği gibi, tez çalışması kapsamında önerilen yöntem ile bir web trafik verisi içerisindeki anomalileri doğru

bulma konusunda ortalama %94.30, genel sınıflandırma oranında ise ortalama %97.69 başarım elde edildiği görülmektedir. Böylece tez çalışması kapsamında önerilen yöntemin yüksek başarım ile web trafik verilerindeki anomalileri bulmada başarılı olduğu söylenebilir.

Çizelge 4.15. Deneysel çalışmalar ile elde edilen sonuçların özet tablosu

Test Numarası	**Veri Numarası**	**Anomali Doğru Bulma Oranı (%)**	**Genel Sınıflandırma Oranı (%)**
1	42	93.18	97.50
2	58	95.34	98.33
3	17	93.67	95.83
4	22	93.65	96.67
5	25	95.20	98.33
6	40	93.42	98.83
7	67	95.65	98.33
	Ortalama	**94.30**	**97.69**

5. TARTIŞMA 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 tez çalışmasında, ağ üzerindeki anormal web trafiklerinin tespiti için YBS'nin NSA (NSA)'ya 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. NSA 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 17., 22., 25., 40., 42.,58. ve 67. olmak üzere zaman serileri şeklinde sunulan sinyal verileri kullanılmıştır Önerilen NSA destekli yöntem ile bir web trafik verisi içerisindeki anomalileri doğru bulma konusunda ortalama %94.30, genel sınıflandırma oranında ise ortalama %97.69 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.

Tez çalışması kapsamında uygulanan yöntem sayesinde anormal trafik için geliştirilen uygulamanın IDS ve hatta sunucuların önüne konumlandırılan wafın görevini yapabileceği düşünülmektedir. Daha kapsamlı bir çalışmayla pek çok atağın yol açtığı anormal trafiğin tespitinde, önerilen bu yöntem önemli rol oynayabilir.

Günümüzde pek çok saldırı çeşidi bulunmaktadır. Bu tez çalışmasında her saldırı türüyle ilgili veri bulunamadığı için sadece Webscope S5 verisetindeki veriler üzerinden

çalışma yürütülmüştür. Bu kısım da çalışmamızın eksik olarak görülen yönü olarak söylenebilir. Web tarafında yapılan atakların anomali tespiti yapılmış olup diğer network katmanlarına ait verilerin olmaması sebebiyle deneyler sadece web trafik verileriyle sınırlı kalmıştır.

Daha sonraki çalışmalarda çalışmanın perspektifi büyültülerek ağ katmanlarının özelliklerine göre veriler elde edilerek YBS'nin NSA 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.

KAYNAKLAR

Adhyaru, P.R.(2016). Techniques For Attacking Web Application Security,*International Journal of Information Sciences and Techniques (IJIST)* ,6.

Agarwal, N.,& Hussain, Z. S.(2018). A Closer Look at Intrusion Detection System for Web Applications, *Hindawi Security and Communication Networks Volume*, 2018, 27.

Akbal E. ,& Ergen B.(2006).Kablosuz Yerel Alan Ağlarında Yapay Bağışıklık Sistemi ile Saldırı Tespiti ve Performans Analizi

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

Amazon(2017). *Use AWS WAF to Mitigate OWASP's Top 10 Web Application. Vulnerabilities*,https://d1.awsstatic.com/whitepapers/Security/aws-waf-owasp.pdf(24.06.2019).

Aziz, A., Salama, M., Hassanien, A., Hanafi, O.(2012),Artificial immune system inspired intrusion detection system using genetic algorithm, *Informatica*, 36

Balachandran, S., Dasgupta, D., Nino, F.(2007).A framework for evolving multi-shaped detectors in negative selection. *Proceedings of the 2007 IEEE symposium on foundations of computational intelligence, Honolulu*,401–8

Berger, V.(2017). *Anomaly detection in user behavior of websites using Hierarchical Temporal memories*.Yüksek Lisans Tezi, Degree Protect in Computer Science and engineering, Sweden.

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

Cheatography (2018),*OWASP Top 10 Vulnerabilities Cheat Sheet*, https://www.cheatography.com//clucinvt/cheat-sheets/owasp-top-10-vulnerabilities/pdf/(24.06.2019)

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

Das, R. K.,Panda, M.,Dash, S.,Dash, S. S.(2018).Application of Artificial Immune System Algorithms in Anomaly Detection, *Progress in Computing, Analytics and Networking*, 687-694

KAYNAKLAR (Devam Ediyor)

Dasgupta D., KrishnaKumar K.,Wong D., Berry M.(2004).Negative Selection Algorithm for Aircraft Fault Detection ,*3rd International Conference on Artificial Immune Systems*

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

F5 Networks (2017). *Preparing For The New OWASP Top 10 And Beyond*, https://interact.f5.com/rs/653-SMC-783/images/EBOOK_Owasp-Top-10-and-Beyond.pdf(25.06.2019)

Farmer, J., D., Packard N., H.,Perelson, A., S.(1986). The immune system, adaptation, and machine learning. *Physica D*,187–204

Forrest, S., Perelson, AS.(1994). Self-nonself discrimination in a computer,*Proceedings of IEEE symposium on security and privacy*, Oakland,202–13

Haboob(2018).*XXE Explanation and Exploitation*,https://www.exploit-db.com/docs/english/45374-xml-external-entity-injection---explanation-and-exploitation.pdf(25.06.2019).

Hackerone (2017).*OWASP TOP 10 2017 A Flash Card Reference Guide to the 10 Most Critical Web Security Risks of 2017*, https://www.hackerone.com/sites/default/files/2017-12/OWASP%20Top%2010%20Flash%20Cards.pdf(25.06.2019)

Hassan, M., Nipa, S. S., Akter, M., Haque, R.,Deepa, N. F., Rahman, M., Siddiqui1, A.,Sharif, H.(2006). Broken Authentication and Session Management Vulnerability: A Case Study Of Web Application, *ISSN*, 1473-8031

Horner, M.,Hyslip, T.(2017). SQL Injection: The Longest Running Sequel in Programming History, Journal of Digital Forensics, *Security and Law*,12.

Hosseinpour, F.,Amoli, V. P., Farahnakian, F., Plosila J., Hämäläinen T.(2014). Artificial Immune System Based Intrusion Detection: Innate Immunity using an Unsupervised Learning Approach, *International Journal of Digital Content Technology and its Applications(JDCTA)*,8, 5

Jerne, N., Towards, K.(1974).A Network Theory of the Immune System. *Annals of Immunology (Institut Pasteur)*, 373–389

Jinquan,Zeng,Xiaojie,Liu,Tao,Li,Caiming,Liu,Ling,xiPeng,Feixian,Sun(2009). A self-adaptive negative selection algorithm used for anomaly detection, *Progress in Natural Science* ,19,261-266

Joshi, J., Aref, W., Ghafoor, A.,Eugene, H., Spafford(2001).*Communications Of The Acm Security Models For Web-Based Applications*, *Communications Of The Acm*,44.

KAYNAKLAR (Devam Ediyor)

Kepler, T., Perelson, A.(1993). Cyclic re-entry of germinal center B cells and the efficiency of affinity maturation, *Immunology Today*,14, 412-415

Kim J.,& Bentley P.(2002).Towards an artificial immune system for network intrusion detection: An investigation of dynamic clonal selection, *Evolutionary Computation, CEC '02. Proceedings of the 2002 Congress*,2.

Kim, T.,& Cho, S.(2018),Web traffic anomaly detection using C-LSTM neural networks; *Expert Systems with Applications*, 106, 66-76.

Knight, T.,Timmis, J., De Castro, L.,Hart E.,(2004).An Overview of Artificial Immune Systems, *DOI: 10.1007/978-3-662-06369-9_4*

Kohnfelder, L., Heymann E., Miller B.(2019). Introduction to Software Security Chapter 3.8.4: *XML Injection Attacks*, https://research.cs.wisc.edu/mist/SoftwareSecurityCourse/Chapters/3_8_4-XML-Injections.pdf(25.06.2019)

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, E. R., Kleinrock, L., Lynch, D. C.,Postel, J.,Roberts, L. G., Wolff, S.(1997).*Brief History of the İnternet,*https://www.internetsociety.org/internet/history-internet/brief-history-internet/(20.06.2019).

Lumension(2009*). Security Configuration Management*,http://www.aspirantinfotech.com/sg/download/lumension/brochure/LEMSS---Security-Configuration-Management---11-23-09.pdf(25.06.2019).

Mazel, J.(2011). Unsupervised network anomaly detection.*Networking and Internet Architecture.*

Messina G.(2018). *10 Steps to Avoid Insecure Deserialization*, https://resources.infosecinstitute.com/10-steps-avoid-insecure-deserialization/#gref(25.06.2019)

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

Pande, P.V.(2014). Negative Selection and Niching by an Artificial Immune System for Network Intrusion Detection, *International Journal of Computer Science and Mobile Computing,* 3, 900-907.

PortSwinger Ltd.(2019). *SQL injection*,,https://portswigger.net/web-security/sql-injection(25.06.2019).

Potoček, P.,& Rehák, M.(2017). *Detection of Malicious Network Behaviour in Encrypted Network Traffic* ,Yüksek Lisans Tezi, Faculty Of Electrical Engineering Departman Of Computer Science, CZECH.

KAYNAKLAR (Devam Ediyor)

Sakar, G.(2018).*Gerçekleşen En büyük DDoS Saldırısı ve Memcahce zafiyeti*,https://bilgiguvenligi.saglik.gov.tr/Haberler/Detay/54/gerceklesen-en-buyuk-ddos-saldirisi-ve-memcahce-zafiyeti(24.06.2019).

Silva, C. G., Caminhas, M. W., Palhares, M. R.(2017).Artificial immune systems applied to fault detection and isolation: A brief review of immune response-based approaches and a case study, Applied Soft Computing,57.

Symantec. (09.06.2019). *Symantec Internet Security Threat Report*., https://www.symantec.com/content/dam/symantec/docs/reports/istr-24-executive-summary-en-apj.pdf(21.06.2019).

The Government of the Hong Kong Special Administrative Region(2008).*VPN Security*, https://www.infosec.gov.hk/english/technical/files/vpn.pdf

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

Twycross, J. P.(2007).*Integrated Innate and Adaptive Artifcial Immune System Applied To Process Anomaly Detection*, Doctoral dissertation, the University of Nottingham

Veracode(2019).What is OWASP and the OWASP Top 10,https://www.veracode.com/directory/owasp-top-10(24.06.2019)

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

Yahoo, WebScope S5 Computing Systems Data (2019). S5 - *A Labeled Anomaly Detection Dataset,Available*,https://webscope.sandbox.yahoo.com(09.06.2019)

Zhang, Y., Lee, W., ve Huang, Y., Intrusion Detection Techniques for Mobile Wireless Networks, *Wireless Networks* ,9 , 545-556.

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

Zhou, J., Dipankar, D.(2004). Real-valued negative selection algorithm with variable-sized detectors. *Proceedings of the genetic and evolutionary computation conference*,3102,287–98.

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Kadir İLHAN

Doğum Yeri ve Tarihi : Erzincan / 29.04.1990

Eğitim Durumu

Lisans Öğrenimi : Süleyman Demirel Üniversitesi Bilgisayar Mühendisliği

Bildiği Yabancı Diller : İngilizce

Bilimsel Faaliyetleri :

İş Deneyimi

Stajlar :

Projeler :

Çalıştığı Kurumlar : Milli Savunma Bakanlığı

İletişim

Adres : Milli Savunma Bakanlığı, ANKARA

E-Posta Adresi : kadir.ilhan929@gmail.com

Akademik Çalışmaları

- Dandıl, E., **İlhan, K.**, "Yapay Bağışıklık Algoritmaları ile Web Trafik Verilerinde Anomali Tespiti", *International Congress on Human-Computer Interaction,Optimization and Robotic Applications (HORA 2019)*, 5-7 Temmuz 2019, Ürgüp, Türkiye.

Tarih: 29/07/2019