Bilgisayar Bilimleri ve Mühendisliği Dergisi

Cilt-16
Sayı-2
2023

nDPI Derin Paket İnceleme Aracı Üzerinde Bir Çalışma

A Study on the nDPI Deep Packet Inspection Tool

Zehra Nur ÖZBAY
Ege Üniversitesi
Uluslararası Bilgisayar Enstitüsü
İzmir, Türkiye
zehra.nur.ozbay@ege.edu.tr
ORCID: 0000-0001-5680-1227

Mehmet Emin DALKILIÇ
Ege Üniversitesi
Uluslararası Bilgisayar Enstitüsü
İzmir, Türkiye
mehmet.emin.dalkilic@ege.edu.tr
ORCID: 0000-0003-3932-5155

Öz

Derin paket inceleme (DPI), ağ paketlerini beş katmanlı ağ modelinde bulunan uygulama katmanına kadar analiz ederek ağda bulunan uygulama protokollerini tespit etmek için kullanılan ileri seviye bir paket tanımlama yöntemidir. Bu çalışmada DPI yöntemi kullanarak paket tanımlaması yapan açık kaynak kodlu nDPI kütüphanesi ele alınmıştır. Bu kütüphane üzerinden yeni uygulama protokolü tespiti ve eksikliği bulunan bazı protokollere de eklemeler sağlanarak alana katkıda bulunulması hedeflenmiştir. Ayrıca, nDPI tarafından yanlış kategorize edildiği tespit edilen ağ paketleri için yeni kural tanımlarının yapılarak bu gibi durumların düzeltilmesi sağlanmıştır. Tüm bunlar için, ağ trafiği Wireshark paket yakalama aracıyla kaydedilip paket içerikleri analiz edilerek yeni kurallar oluşturulmuş ve yeni bulunan uygulama protokolleri nDPI kütüphanesine eklenmiştir. Son olarak, nDPI'da uygulama protokolü tespiti için yapılan çalışmaların kısmi bir otomasyonu yazılmıştır.

Anahtar sözcükler: Derin paket inceleme, DPI, nDPI

Abstract

Deep packet inspection (DPI) is a an advanced packet identification method used to detect application protocols found in the network by analyzing network packets up to the application layer in the five-layer network model. In this study, the open source nDPI library, which uses DPI method to identify packets, is discussed. It is aimed to contribute to the field by detecting new application protocols and adding some missing protocols through this library. In addition, new rule definitions were made for network packets that were found to be miscategorized by nDPI, and such cases were corrected. For all these, network traffic was recorded with Wireshark packet capture tool, packet contents were analyzed, new rules were created and newly found application protocols were added to the nDPI library. Finally, a partial automation of the work done for the application protocol detection in nDPI was written.

Keywords: Deep packet inspection, DPI, nDPI

Gönderme, düzeltme ve kabul tarihi: 20.02.2023 - 29.03.2023 – 20.05.2023
Makale türü: Araştırma

1. Giriş

Her geçen gün yeni bir uygulama ağda yerini alırken tanımlanması ve sınıflandırılması gereken protokol sayısı da artmaktadır. Bunun yanı sıra internete bağlı cihazların sayısı da nesnelerin interneti (IoT) ile birlikte muazzam seviyelere ulaşmaktadır. 2030 yılına gelindiğinde yaklaşık 30 milyar IoT cihazının kullanımda olacağı tahmin edilmektedir [1]. Cihaz ve uygulama sayılarındaki bu kontrolsüz artış ağ güvenliği problemlerini de beraberinde getirmektedir. Güvenlik duvarları, Sızma Tespit Sistemleri (IDS) ve Sızma Önleme Sistemleri (IPS) ağa yönelik saldırılara karşı koymak için sıklıkla tercih edilen ve kullanılan ağ güvenliği cihazlarıdır. Genellikle iç ağ ile dış ağ arasındaki sınır bölgelere konumlandırılan güvenlik duvarları, iç ağdan dış ağa ya da tersi yönde hareket eden ağ paketlerinin başlık bilgilerini kontrol ederek kullanıcının önceden tanımladığı kurallar ile uyumlu paketlere geçiş hakkı verirken, uyum sağlamayanları düşürür. Güvenlik duvarları ağ giriş ve çıkışlarında klasik paket filtreleme için kullanılır. Fakat "ince taneli" saldırı paketlerini tespit edemeyeceği için bu filtreden içeriye sızma mümkündür. Diğer bir ağ güvenliği cihazı olan IDS ise dinleme yapan duyargalar aracılığı ile iç ağda görev yapar. IDS imza veri tabanını kullanarak iç ağdaki şüpheli ve düzensiz trafiği tespit ederek ağ yöneticisine rapor eder. Genellikle güvenlik duvarlarının tam arkasına konumlandırılan IPS, IDS'nin paketler üzerindeki pasif işlevinin aksine, ağda aktif rol alarak

paketleri engelleme ve düşürme gibi yetkilere de sahip olup IDS cihazlarının bir uzantısıdır. Tüm bu cihazlar kuşkusuz ağ güvenliğinin ana unsurlarını oluşturmaktadır. Fakat bu cihazlar tek başlarına ele alındığında yeni tip ağ saldırılarına karşı etkileri düşük olacaktır. Örneğin, güvenlik duvarlarının işlem gücü sınırlı olup büyük hacimli paketleri işlemek için yeterli değildir. Ayrıca, bir güvenlik duvarı durum bilgisi tutsa dahi beş katmanlı ağ modelinin son basamağı olan ve asıl veriyi içeren uygulama katmanındaki bilgileri göremez. Bu tıpkı elimize aldığımız bir kitabın içeriğinden haberdar olmadan sadece kitabın başlığını okuyarak kitabı değerlendirmemize benzer [2]. DPI ise bu noktada ağ trafiğini çözerek uygulama katmanı dâhil tüm katmanlarda inceleme yapmayı ve böylece paket içeriğinde yer alan veriyi de, şifrelenmemiş ise, okumayı sağlar. Diğer bir yandan, saldırganlar hedef ağa gönderecekleri paketleri daha küçük parçalara ayırıp saldırı niteliğindeki imzanın tek bir pakette yer almasını engelleyerek IDS tarafından tespitini imkânsız hale getirebilirler. DPI analizinde ise saldırı imzası birçok pakete ayrılarak izi silinmeye çalışılmış olsa bile bu paketler bir araya getirilerek imza tespit edilir. Dolayısıyla, DPI teknolojisi ağ güvenliği cihazlarının temel bir bileşeni haline gelmiştir. Bir DPI aracı ya da kütüphanesinin kullanımı ağ içerisinde dinleme yapan paket yakalama araçlarıyla birlikte çalışması şeklinde olabileceği gibi daha yaygın kullanımlar olan DPI tanımlı yeni nesil güvenlik duvarı (NGFW) ya da IDS/IPS cihazları şeklinde de olabilir.

Yayınlardaki çalışmalar incelendiğinde hem uygulama protokolü tespiti hem de uygulamayı engelleme amaçlı derin paket inceleme kullanan çalışmalar bulunmaktadır. Renals ve Jacoby tarafından yapılan çalışmada [3], DPI kullanılarak P2P bir mesajlaşma uygulaması olan Skype'ın nasıl engellendiği gösterilmiştir. Çalışmada, Skype paketlerini tanımlamak için Skype'ı diğer uygulamalardan ayırt eden özellikleri kullanılarak bir kural kümesi oluşturulmuştur. Bu kural kümesini inşa edebilmek için önce -olası tüm durumları hesaba katmak üzere- farklı bilgisayarlar üzerinde, farklı zamanlarda ve farklı Skype hesapları ve versiyonlarında Wireshark ağ dinleme aracı kullanılarak veri toplanmıştır. Elde edilen Skype trafiğine ait veriler, aynı anda en fazla üç farklı oturum trafiğini karşılaştırabilen, Araxis Merge isminde bir analiz programı kullanılarak karşılaştırılmıştır. Bu analiz Skype kurallarının, anahtar kelimeler, port numaraları ve içerik olmak üzere, üç kategoride şekillenmesiyle sonuçlanmıştır. Anahtar kelimeler içeriğinde "Skype" gibi belirli dizgiler içeren paketleri, port numarası bağlantı noktası 33033 olan TCP trafiğini ve içerik ise "16 03 01 00 00" veya "17 03 01 00 00" ASCII dizgilerini içeren paketleri tanımlar. Bu kurallar, içerisinde DPI tanımlı bir IPS cihazı olan açık kaynak kodlu Snort'ta tanımlanır. Bu kuralları uyguladıktan sonra Wireshark ile trafik yeniden kaydedilerek Skype'ın davranışı gözlemlenir. Skype'ın bağlantı kurmaya çalışıp yeni dizgiler kullandığı tespit edilir. Son olarak bu yeni dizgiler de kural tablosuna eklenerek Snort çalıştırıldığında bir önceki davranışından farklı bir yol denemeyen Skype'ın bloke olduğu gözlemlenmiştir. Diğer bir çalışmada ise [4], internette anonim kimlikle iletişimi sağlayan Soğan Yönlendirme (TOR) yazılımının DPI kullanılarak ağda kullanımının tespit edilmesi ve elde edilen analiz sonuçlarıyla bu uygulamanın nasıl engellendiği ele alınmıştır. Bunun için, yeni bir TCP bağlantısı kurulduğu zaman bazı komutları çalıştırabilecek olan (olaya dayalı) ve DPI özellikli bir Bro-IDS sunucusu, Wireshark ve bir Squid vekil sunucusu kullanılmıştır. TOR trafiğini, TLS protokolü ile şifrelenmiş durumda olduğundan, karakterize edebilmek amacıyla TOR bağlantı kurulumu için gerekli olan ve veri alım ve gönderimi öncesi gerçekleştirilen iki süreç analiz edilmiştir. Birincisi TOR kullanıcısı ile TOR ağı arasındaki -SYN, SYN-ACK ve ACK paketlerinin iletimi olarak da bilinen- üç yönlü el sıkışma süreci, ikincisi ise TLS oturumu kurma sürecidir. Üç yönlü el sıkışma, TOR dışındaki diğer bağlantılar ile karşılaştırıldığında, standart bir trafik olarak seyretmiştir. Fakat TLS el sıkışması olarak da adlandırılan TLS bağlantı kurulumunda ise bazı farklılıklar tespit edilmiştir. Bunlardan istemcinin sunucuya gönderdiği "ClientHello" mesajlarında TOR'u karakterize eden iki özellik gözlemlenmiştir: Birincisi, şifreleme paketi kombinasyonlarının -yani istemci (TOR tarayıcısı) tarafından teklif edilen şifreleme algoritmalarının- hep aynı olması, diğeri ise sunucu adı formatının her zaman www.<rastgele_dizgi>.com veya "www.<rastgele_dizgi>.net" şeklinde olmasıdır. Diğer bir farklılık, "ClientHello" mesajlarından sonra sunucunun gönderdiği sertifikadaki bilgilerinde de "ClientHello" mesajındaki aynı dizgi formatının yer almasıdır. TOR trafiğinin, çıkarılan bu özellikleri Bro-IDS cihazında kural olarak yazıldıktan sonra, bloke edilme süreci ise bir internet kullanıcısı ile vekil sunucu arasındaki trafiğin tam bir kopyasının Bro-IDS tarafından analiz edilmesiyle başlar. TOR karakterine sahip paketlerin bazı özellikleri çıkartılarak Bro-IDS'in kayıt dosyasına yazılır. Bro-IDS, bu yazım olayı gerçekleşince, ilgili paketlerin varış IP adreslerini bir erişim listesi dosyasına yazma ve vekil sunucuyu -bloke edilecek bu adresler ile- yeniden yükleme olmak üzere iki fonksiyonunu aktive eder. Böylece TOR trafiği engellenmiş olur. Son olarak, ağ trafiğindeki Adobe Creative Suite, Microsoft Sharepoint, Salesforce, Yammer ve Zendesk uygulamalarının nDPI kullanılarak tespit edilmesini ele alan bir çalışmada [5] otomatik bir betik yazılarak uygulamaların ilgili web sitelerinde gezinti yapılmış ve en çok kullanılan 20 IP adresi ve alan adı kaydedilmiştir. Daha sonra bu IP adreslerinin ve alan adlarının gerçekte kime ait olduğunu doğrulamak için bir whois aracı kullanılmıştır. Bazı IP adreslerinin iki uygulama tarafından da kullanımda olduğu bir durumda bu adresler, tek bir uygulamayı net bir şekilde ifade edemeyeceği için, kural kümesine dâhil edilmemiştir. Diğer yandan, bir IP adresinin sunuculuğu başka bir firmaya ait olup kullanım hakkı bakımından sahipliği uygulamaya aitse bu adresler dâhil edilmiştir.

Bizim çalışmamız ise açık kaynak kodlu nDPI kütüphanesi [6] ve Wireshark [7] ağ dinleme cihazının beraber kullanıldığı bir senaryo üzerinden gerçekleştirilmiştir. Kullanılan yöntem yukarıdaki çalışmalara benzer olup, Radityatama ve arkadaşlarının araştırmasına [5] daha yakındır. Fakat bu çalışmadan farklı olarak IP adresleri kullanılmadan yeni protokoller tespit edilip eklenmek suretiyle açık kaynak kodlu nDPI kütüphanesine katkı sağlanması hedeflenmiştir. IP adreslerinin neden eklenmediği kural çıkarma alt başlığında açıklanmıştır.

2. Temel Tanım ve Kavramlar

Akış kavramı, bir TCP/IP bağlantısını oluşturan aynı beşli değer kümesine sahip bir veya birden fazla paketi ifade eder. Bu beşli değer kümesinin elemanlarını; kaynak IP adresi/port numarası, hedef IP adresi/port numarası ve IP başlık bilgisinde yer alan protokol bilgisi (TCP, UDP, ICMP vb.) oluşturur. Bir TCP/IP bağlantısını eşsiz olarak bu beşli belirler. Bir diğer ifadeyle, aynı beşli değere sahip TCP/IP paketleri aynı akışa aittir.

Ağ trafiği sınıflandırması bir ağda dolaşımda olan uygulamaları tanımlamak ve analizini yapmak için kullanılan temel bir metottur. Bu teknik ağın performans ölçümü, güvenliği, daha iyi yönetimi ve trafik mühendisliği için özellikle internet servis sağlayıcıları (ISP) ve ağ operatörleri tarafından sıklıkla kullanılır. Böylece, ISP'ler uygulama veya abone bazında trafiği tanımlayabilir, izleyebilir ve kontrol edebilirler. Ağ trafiğinin, genel olarak, web, bulut, multimedya, mesajlaşma, e-posta, oyun, müzik, veritabanı, dosya paylaşımı ve IP üzerinden ses (VoIP) gibi daha birçok sınıf tanımı yapılmaktadır. Örneğin, bu çalışmadaki uygulama protokolleri alışveriş ve eğlence kategorileri kapsamındadır.

2.1 DPI

Derin paket inceleme (diğer bir adıyla Tam Paket Denetimi veya Bilgi Çıkarma [8]); ağ yönetimi, güvenlik, filtreleme, yönlendirme, sansür, istatistik, hizmet kalitesi (QoS) ve deneyim kalitesi (QoE) gibi birçok farklı amaç doğrultusunda kullanılan; genellikle güvenlik duvarı, yönlendirici veya dağıtıcı gibi ağ düğümleri üzerinde önemli bir dişli olarak çalışan; paketlerin sadece başlık bilgilerine değil uygulama katmanında yer alan yük içeriklerine de bakarak "derinlemesine" analiz eden bir paket tanımlama tekniğidir. Kısaca, derin paket incelemesi içeriğe duyarlı işleme yaparak ağ trafiğinin doğru bir şekilde sınıflandırılmasını sağlar.

DPI metodu, kendi içerisinde dar ve geniş olmak üzere iki farklı kavram olarak ele alınır [9]. Dar anlam desen eşleştirme yöntemini ifade etmektedir. Bu yöntem de kendi içinde dizgi ve düzenli ifade eşleştirmesi olarak ikiye ayrılmaktadır. Bu tanıma işlemi, paket yükünün önceden tanımlanmış imzalarla eşleştirilmesi ile uygulanmaktadır. Geniş anlamda ise; istatiksel analiz, port tabanlı eşleştirme ve protokol çözümleme yöntemleri ele alınmaktadır. Bu çalışmada uygulanan da, dar kapsamında sayılan ve nDPI'da ndpi_content_match.c.inc isimli dosya üzerinden gerçekleştirilen dizgi eşleştirme yönteminin kullanılmasıdır.

DPI metodu, paket yükünü de öğrenmeye dayalı olduğundan yüksek doğruluk sağlasa da kullanıcı gizliliği ve güvenliği açısından sakıncalı bulunabilmektedir. Ağ trafiğinin şifreli yapıya geçişi de bu yöntemin üzerinde baskılayıcı bir rol oynamaktadır. Çünkü SSL/TLS trafiğinde şifre bilinmediği sürece sadece el sıkışma fazındaki anahtar değişiminden faydalanılabilir. Bu çalışmadaki uygulama protokollerinin hepsi şifreli bağlantı kullanmaktadır. Dolayısıyla sadece SSL/TLS el sıkışma fazındaki anahtar değişiminden elde edilen veriler kullanılarak ağ trafiği tanımlaması gerçekleştirilmiştir.

Hem şifreli hem de gerçek zamanlı ağ trafiğinin işlenmesi ise üzerinde çalışılması gereken ayrıca bir konu olup literatürdeki boşluğunu korumaktadır [10]. Ağ katmanlarının en üstünde yer alan bu problemden dolayı belki de sınıflandırma işlemlerini ağ katmanının en alt basamağı olan fiziksel katmandaki bitlere kadar indirgeyerek trafiği direkt izleme imkânı sunacak bir ağ trafiği sınıflandırması gerçekleştirme bu alanın geleceğindeki en büyük zorluklarından biri olarak görülmektedir [11].

DPI teknolojisinin iki temel fonksiyonu vardır. Birincisi, tanımlama yapmak, diğeri ise bu tanımlama sonucuna göre bir eylemde bulunmaktır [9]. Tanımlama, ağ paketinin karakteristik özelliğini çıkarmaktır. Bu karakteristik; uygulama protokolü, multimedya uygulaması, bulut yazılımı gibi normal trafik akışını ya da kötü amaçlı yazılım, virüs, siber saldırı gibi zararlı içeriği işaret edebilir. Uygulanan eylem ise güvenlik araçlarıyla kurulan ilişkidir. Paket tanımlama tamamlandıktan sonra, IDS, IPS veya güvenlik duvarı aktif hale getirilerek; uyarı sinyali verme, bloke etme, yeniden yönlendirme veya günlüğe kaydetme aksiyonları gerçekleştirilir. Örneğin, bir paket DPI tarafından; güvenlik duvarındayken "zararlı yazılım" tanımlaması yapılmış ise düşürülebilir, IDS üzerinde "potansiyel tehlike" olarak bildirilmişse ağ yöneticisine rapor gönderilebilir veya IPS üzerindeyken "şüpheli" olarak tanımlanmış ise bağlantı bloke edilebilir. Bu çalışma tanımlama kısmını ele almaktadır.

2.1.1 DPI Uygulamaları

DPI teknolojisinin birçok farklı amaca yönelik kullanım alanları bulunur. Bunlardan bazıları; ağ güvenliği, bant genişliği yönetimi, deneyim kalitesi, gözetim ve sansürdür. Xu ve arkadaşlarının çalışması [9] ise DPI kullanılarak yapılan uygulamaları devlet, ISP ve şirket olmak üzere üç grubun kullanım amaçları doğrultusunda daha detaylı olarak ele alır. Bu kullanım amaçları arasında, kullanıcı profili üzerinden reklam enjeksiyonu gerçekleştirmek de bulunur. İlk olarak Bendrath'ın çalışmasında [12] geçen reklam enjeksiyonu, bir kullanıcının internette gezindiği veya alışveriş yaptığı web sitelerinin takip ve analizi yapılarak bıraktığı ayak izine ait bir reklam profilinin oluşturulması ve ona özel çevrimiçi reklam sunulmasıdır. Bu araştırma çerçevesinde bahsedilen yedi şirketin web sitesinden toplanılan veriler incelendiğinde ağ trafiğini büyük ölçüde reklam ve takip amaçlı yazılımlara ait paketlerin doldurduğu gözlemlenmiştir.

2.1.2 nDPI

nDPI ilk olarak 2013 yılında, ntop şirketi tarafından ağ trafiğini karakterize etmek için, C dilinde geliştirilen ve LGPL lisansına sahip bir DPI kütüphanesidir. nDPI, GPL lisansına sahip olan fakat artık güncelleme almayan OpenDPI yazılımından ilham alınarak geliştirilmiştir. nDPI, OpenDPI ile benzer olarak, hem Linux çekirdeğinde hem de kullanıcı programlarının bulunduğu kullanıcı alanında kullanılabilir. Linux dışında, Windows, MacOS X ve BSD ailesi gibi işletim sistemlerinde de kurulumu yapılabilir. En son sürümü olan nDPI 4.6 [13], Şubat 2023'te çıkmıştır. Bu araştırmada yer alan çalışmalar Linux tabanlı bir işletim sistemi olan Ubuntu üzerinde ve o esnada mevcut olan nDPI 4.2 sürümü kullanılarak yapılmıştır. nDPI, en son sürümüyle beraber, 300'ü aşkın uygulama protokolü

tanımlamanın yanı sıra TLS sertifikası, tarayıcı adı ve şifreleme paketi gibi bir akışla ilişkili meta verileri de raporlar.

nDPI'ın belli başlı özellikleri aşağıdaki gibidir:

- nDPI'da bir protokol, genelde, nDPI kütüphanesinde tanımlı ve C dilinde yazılan trafik ayrıştırıcısı ile tespit edilir. Fakat protokoller sadece ayrıştırıcı kullanılarak bulunmaz. Port numarası, IP adresi ve protokol özelliklerine göre de bulunabilir. Örneğin, Dropbox trafiği hem yerel alan ağı tabanlı bağlantılar için kullanılan ayrıştırıcı hem de sunucu adı bilgisinde ".dropbox.com" dizgisi geçen HTTP trafiğinin Dropbox olarak etiketlenmesiyle tanılanır [14]. Bu yüzden, yeni bir akış trafiğinin yaşam döngüsü öncelikle paketlerin üçüncü ve dördüncü katmanlarının çözümlenmesi ve bu arada ayrıştırıcıların denenmesiyle başlar. Modülerlik ve genişletilebilirlik özelliklerini sağlaması için her biri ayrı bir c uzantılı dosyada kodlanan ayrıştırıcıların uygulanış sırası, trafik türüne göre akışla eşleşme olasılığı en yüksek olandan başlayarak gerçekleşir. Örneğin, TCP/80 için önce HTTP ayrıştırıcısı denenirken, 53 port numaralı TCP/UDP protokolleri için ise önce DNS ayrıştırıcısı denenir. Her akış, eşleşmeyen ayrıştırıcıların durum bilgisini gelecek olan iterasyonlarda atlamak için saklar. Analiz, bir eşleşme bulunana kadar veya belirli sayıda denemeden sonra sona erer. Bu çalışmada, ele alınan web sitelerinin hepsi TLS/HTTPS kullandığı ve nDPI'da da HTTP ayrıştırıcısı bulunduğu için ayrıca bir protokol ayrıştırıcısı yazılmamıştır.

- Deri ve arkadaşlarının çalışmasında [14], nDPI'da bir uygulama protokolünü tespit etmek veya bilinmeyen olarak etiketlemek için gerekli paket sayısının en fazla sekiz olduğu belirtilmiştir. Bunun anlamı, bir akış protokolü imzasının tanımlanabilmesi için bir araya getirilmesi gereken paket sayısının en fazla sekiz olmasıdır. Ayrıca, çalışma sürecinin ne zaman başlatıldığı da önemlidir. Örneğin, nDPI akış başladıktan sonra çalıştırıldığı bir durumda akışa ait ilk paketler analiz edilmediğinden o akış sınıflandırılmamış olarak işaretlenebilir.

- nDPI kütüphanesinde her uygulama protokolü benzersiz bir protokol numarası ve sembolik bir protokol adı ile tanımlanır. Yeni bir protokol kütüphaneye ekleneceği zaman o protokol için daha önce hiç kullanılmamış bir numara ve isim bulunmalıdır. nDPI dilinde sadece TLS, DNS gibi temel internet ağı protokolleri değil, Skype, Facebook veya Youtube gibi uygulamalar da protokol olarak adlandırılır. Bu yüzden bir protokol aslında ağ ve uygulama olmak üzere iki türlü tanımlanır. Bu protokol türleri sırasıyla majör ve minör olarak da ifade edilebilmektedir. Örneğin, ağ protokolü TLS ve uygulama protokolü Facebook olan bir TCP paketi proto: 91.119/TLS.Facebook olarak raporlanır. Buradaki 91 TLS'e, 119 ise Facebook'a özel tanımlanmış protokol numaralarıdır. Aynı şekilde, en çok karşılaşılan, 5 numara DNS'e aittir.

- nDPI, iki uç sistem arasındaki bir SSL bağlantısında ortak bir anahtar yardımıyla gerçekleştirilen şifreli iletişim başlamadan önceki ilk anahtar değişimi kısmı için kod çözümleyicisi bulundurur. Bu sayede bağlantı kurulan uç sistemin sunucu adını çıkarabilir ve şifreli paketlerin hangileri olduğunu, içeriklerini okuyamasa da, belirleyebilir. Sunucu adı bilgisi, tıpkı HTTP bağlantılarında ana bilgisayar adının HTTP başlık bilgisinden alınması gibi, nDPI'ın akışa ait meta verilerinde tutulur. Bu çözümleyici sayesinde hem sunucu adlarına göre protokoller belirlenebilir hem de kendinden imzalı SSL sertifikaları bulunabilir. Kendinden imzalı sertifikalar, kullanıcıların kendi adlarına verdiği, herhangi bir sertifika yetkilisi tarafından onaylanmamış ortak anahtar sertifikalarıdır. Bu tür imza içeren bir bağlantı güvenli sayılmadığından bu bilgi değerlidir.

- nDPI'da protokollerin güvenli, kabul edilebilir, potansiyel tehlikeli, tehlikeli gibi türlere ayrıldığı ndpi_typedefs.h dosyası bulunur. Aynı dosyada, ayrıca, protokolleri anlamlı bir şekilde gruplamak için; alışveriş, oyun, dosya paylaşımı, reklamlar, çevrimiçi bulut hizmetleri, sosyal ağlar, anlık mesajlaşma uygulamaları, kötü amaçlı yazılım ve yasaklı web sitesi gibi soyut kategoriler tanımlıdır.

- nDPI/example dosyasında bulunan ndpiReader demo uygulaması, komut satırından çalıştırılan ve nDPI kütüphanesinin bazı özelliklerini gösteren bir test aracıdır. Bu uygulamayı kullanmak için veriler ndpiReader tarafından belirli bir süre canlı trafiğin yakalanmasıyla veya Wireshark gibi ağ dinleme cihazlarıyla daha önceden kaydedilen hazır pcap dosyalarının girdi olarak verilmesiyle sağlanabilir. Bu yönden, ndpiReader aynı zamanda bir ağ dinleyicisi olarak da çalışır. ndpiReader aracının ihtiyaca göre kullanılabilecek çeşitli komut satırı seçenekleri bulunur. Örneğin, -i bayrağı ndpiReader aracına girdi olarak verilecek önceden kaydedilmiş bir pcap dosyasının veya canlı trafik yakalaması yapılacak ise de kullanılacak cihazın veya arayüzün isminin belirlenmesini sağlarken; -v <1|2|3|4> bayrağı 1 numara için normal ayrıntılı, 2 için daha çok ayrıntılı, 3 için port istatistikleri ve 4 için özüt (hash) istatistiklerine göre görüntüleme yapmayı sağlar. Tüm diğer seçenekler ve bazı nDPI özellikleri yardım anlamına gelen -h bayrağı kullanılarak elde edilebilir.

- Akış analizi sonuçları -C bayrağı kullanılarak csv formatında bir dosyaya kaydedilebilir. Bu csv dosyasında; akış numarası, kaynak/varış IP numarası, kaynak/varış port numarası, eşsiz protokol numarası, sunucu ismi, TLS versiyonu, her iki yöndeki paket ve byte sayısı gibi daha birçok alan bulunur. Bu şekilde csv formatındaki veriler düzenlenip sıralanarak trafik analizi süreçlerine faydalı olabilir. Bu formatta bir çıktı elde etmek için komut satırından aşağıdaki komut yazılır:

  ndpiReader -i <dosya_adı.pcapng> -C <dosya_adı.csv>

  Bu şekilde elde edilen csv dosyası, linux komut satırına

SQL sorgulama yetisi kazandıran q-text-as-data uygulamasına [15] girdi olarak kullanılırsa, trafik hakkında bazı çıkarımlarda bulunmamızı sağlar. Örneğin, belirli bir web sitesinde en çok hangi kaynak IP numarasının veya istemcinin, zaman harcadığını bilmek istediğimizi varsayalım. Aşağıdaki kod yardımıyla giden ve gelen trafikteki toplam byte sayısına göre siteyle iletişim kuran tüm IP numaralarını q uygulamasıyla sıralayabiliriz:

```
q -H -d ‘,’ "select src_ip,
SUM(s_to_c_bytes+c_to_s_bytes)
from <csv_dosya_yolu> where
server_name_sni like ‘<sunucu_ismi>’
group by src_ip"
```

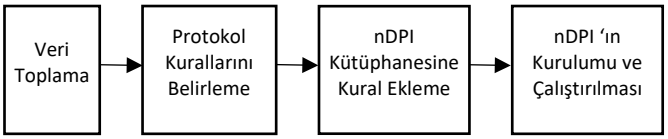
Burada -H bayrağı, dosyanın sütunları adlandırmak için kullanılan bir başlık satırı içerdiğini belirtir. -d bayrağı ise girdi sınırlayıcı olarak çalışır. Sunucu isminde "trendyol" geçen paketler için uygulanan bir örnek sorgunun çıktısı Şekil-2’de verilmiştir. Veriler kişisel masaüstü bilgisayar üzerinde toplandığı için karşılık gelen tek bir IP adresi, 155.223.40.20, görünmektedir. Bu şekilde, csv dosyasında bulunan diğer verilerden farklı farklı sorgular ile hedeflenen amaca yönelik ağ trafiği analizi gerçekleştirilebilir.

```
zehra@zehra:/usr/src/nDPI-4.2$ q -H -d ',' "select src_ip,SUM(s_to_c_bytes
+c_to_s_bytes) from /usr/src/nDPI-4.2/trendyol_for_q_test.csv where server
_name_sni like '%trendyol%' group by src_ip"
155.223.40.20,1805491
```

Şekil-2: Bir q-text-as-data sorgusuna ait örnek çıktı

3. Gereç ve Yöntem

Çalışmanın genel çerçevesi Şekil 1‘de verilmiştir.



Şekil-1: Çalışmanın genel çerçevesi

Bu kapsamda, tespit edilmesini istediğimiz uygulamanın gerçek zamanlı kullanımı sonucu ağda oluşan paketlerin Wireshark aracı yardımıyla yakalanıp pcap veya pcapng uzantısıyla kaydedilerek detaylı paket içerik analizlerinin yapılması için kurulumu yapılan nDPI aracına girdi olarak verilmiştir. Hem nDPI aracından elde edilen sonuçlar hem de kural çıkarma alt başlığında bahsedilecek olan diğer metotlar kullanılarak alan adlarından elde edilen bir kural listesi oluşturulmuştur. Son adımda nDPI cihazında eksik olan bu uygulama protokolleri kütüphaneye kural olarak eklenerek cihazın tespit edebildiği protokol sayısı artırılmıştır. Bu protokoller, Fast Company dergisinin 2020 Şubat ayında yayınlanan ve gelirini sadece internet üzerinden sağlayan Türkiye’deki en büyük 100 internet şirketi listesindeki [16] ilk yedi şirketin kullandığı web sitelerini kapsamaktadır. Bu şirketler büyüklük sırasına göre sırasıyla; Trendyol, Hepsiburada, Nesine.com, N11, Bilyoner, GittiGidiyor ve Yemeksepeti’dir. Protokol seçiminin bu listeden yapılmak istenmesinin sebebi, bu şirketlerin kullanıcıyla etkileşiminin sadece internet üzerinden olması ve ilgili web sitelerinde yoğun veri akışı gerçekleşmesi sebebiyle ağ trafiğindeki tespitlerinin önemli olduğunun değerlendirilmesidir. Son olarak, yapılan çalışma Selenium web sürücüsü ve Python programlama dili kullanılarak kısmi olarak otomatize edilmiştir.

3.1 Veri Toplama

Veriler iki farklı ana bilgisayar kullanılarak toplanmıştır: Bunlardan biri Windows 10 ve Ubuntu 20.04.5 LTS olmak üzere çift işletim sistemine sahip olan, Ethernet kablosuyla internete bağlanan ve dört çekirdekli 3.60 GHz Intel Core i7-4790 işlemciye sahip 64 bitlik bir masaüstü bilgisayardır. Diğeri ise, Windows Education serisi işletim sistemine sahip olan, internet ağına kablosuz bağlanan ve çift çekirdekli 2.50 GHz Intel Core i5-3210M işlemciye sahip 64-bitlik bir dizüstü bilgisayardır. Wireshark kullanılarak veri toplama amaçlı gerçekleştirilen web sayfası gezintilerinde, her iki bilgisayarda da hem Chrome ve hem de Firefox tarayıcısından faydalanılmıştır. Burada iki farklı bilgisayarın, işletim sisteminin, bağlantının ve tarayıcının kullanım amacı farklı sistem kombinasyonlarının yeni veri getirme olasılığını hesaba katmaktır. Fakat bu tercihlerden birinin diğerine göre fazladan veri sağladığı bir durum gözlemlenmemiştir. Bununla birlikte veri çeşitliliğine en çok katkı sağlayan fonksiyon, kuşkusuz, web sayfasında bulunan farklı web elemanlarına erişimin sağlanması olmuştur. nDPI kütüphanesine eklenecek kuralların belirlenebilmesi için öncelikle Trendyol, Hepsiburada, Nesine.com, N11, Bilyoner, GittiGidiyor ve Yemeksepeti’ne ait web sitelerinde Wireshark açılarak paket yakalaması başlatılmış ve ilgili web sitelerin her birinde bir süre gezinti yapılmıştır. Bu gezintilerde, bir web sayfasına ait tüm farklı alan adlarının elde edilebilmesi için, olabildiğince fazla sayı ve türdeki sayfa elemanlarına erişilmesi gözetilmiştir. Gezinti sonucu oluşan dosyalar protokol, tarayıcı ve bağlantı türüne göre çeşitli isimler verilerek .pcapng uzantısıyla işletim sistemi ayrımına göre bir klasöre kaydedilmiştir. Örneğin, n11_chrome_ethernet.pcapng veya trendyol_firefox_wifi.pcapng gibi. Bu şirketlerden eBay’ın sahibi olduğu GittiGidiyor 20 Haziran 2022 tarihinde Türkiye pazarından çekildiğini duyurarak 18 Temmuz 2022 itibarı ile de internet alışveriş sitesini kullanıma kapatmıştır. Bu nedenle, çalışmanın başlangıcında bu sitenin protokolü üzerinde çalışılmış ve kuralları çıkartılmış olsa da, daha sonra yazılan otomasyon sürecinde dâhil edilmemiştir.

3.2 Kural Çıkarma

Çalışmanın başlangıcında trafik verilerini analiz etmeye ilk olarak yedi şirketin tarayıcıda açıkça görünen ve kullanıcılar tarafından da iyi bilinen alan adlarının IP adreslerine ulaşılması hedeflenmiştir. Bunun için, öncelikle bu alan adlarının kaydını gerçekleştiren kayıtçı kuruluşlar ve ilgili web adresleri whois veri tabanı sorgusuyla bulunmuştur. Daha sonra kayıtçı kuruluşların web adreslerinde tekrar whois sorgusu yapılarak şirketlerin kullanıcılara sunduğu web sayfalarının yetkili ad sunucusu bilgileri ve karşılık gelen IP adresleri listesine ulaşılmak istenmiştir. Yapılan sorgularda kayıtçı kuruluşların; yetkili ad sunucuları, kayıt olan

organizasyonun adı, sistem yetkilisinin adı ve iletişim bilgileriyle kuruluşun adresi gibi çeşitli bilgileri döndürmesine rağmen IP adreslerini vermediği gözlemlenmiştir. Bunun üzerine, bu adresleri IP adres aralıkları şeklinde elde etmeyi sağlayan başka bir yöntem denenmiştir. Bunun için, ARIN, APNIC ve RIPE NCC kuruluşlarının veri tabanında whois sorgusu yapılmıştır. Fakat burada da sağlıklı sonuçlar elde edilmemiştir. Örneğin, Trendyol şirketinin RIPE veri tabanındaki sorgusunda 159.20.112.0/24, 159.20.113.0/24, 159.20.114.0/24, 159.20.115.0/24 ve 159.20.116.0/24 IP adres aralıkları elde edilmiştir. Fakat Wireshark ile ağ dinlenip, ip.addr == 159.20.116.0/16 gibi, oldukça geniş IP aralığının varlığını kontrol eden birçok filtre uygulanmasına rağmen whois sorgusundan elde edilen yukarıdaki IP adreslerinden hiçbiri ağ trafiğinde tespit edilmemiştir. Fakat https://dnschecker.org/ip-location.php adresinden IP adresi-konum sorgusu yapıldığında bu IP adreslerinin İstanbul'un çeşitli ilçelerine ait olduğu ve organizasyon adı olarak da Trendyol'un sahibi olan DSM Grup Danışmanlık İletişim ve Satış Ticaret Anonim Şirketi'nin [17] isminin geçtiği de gözlemlenmiştir. Dolayısıyla üçüncü bir seçenek olarak, uzun ama daha sağlıklı ve kesin çözüm üretebilecek olan, Wireshark'ta toplanan trafik verilerindeki tüm DNS sorgu ve cevaplarının tek tek incelenmesi yolu izlenmiştir. DNS sorgu ve cevaplarına kolay ve hızlı bir şekilde ulaşabilmek için Wireshark filtreleme yöntemleri kullanılmıştır. Örneğin, bir akıştaki tüm DNS sorgu ve cevaplarını çekmek için dns.qry.class filtresi kullanılırken, paketler arasından "n11" dizgisini içeren DNS sorgu ve cevaplarını elde etmek için dns contains "n11" filtresi uygulanmıştır. Bu gibi sorgular sonucunda Wireshark'ta çıktı olarak sıralanan paketlerin içeriğindeki veriler incelenerek hem alan adları hem de karşılık gelen IP adresleri kaydedilmiştir. Ayrıca, Wireshark sonuçlarında çıkmamasına rağmen nslookup yöntemiyle elde edilen farklı alan adları ve IP adresleri olmuştur. Fakat bu yöntem tüm IP adreslerini vermediği için yüksek ihtimalle var olduğu düşünülen fakat Wireshark ile bulunamamış alan adı olduğu zaman başvurulmuştur. Örneğin, Trendyol için Wireshark üzerinden collect.trendyol.com, collect2.trendyol.com, collect3.trendyol.com ve collect5.trendyol.com gibi alan adları tespit edilmiştir. Bu nedenle, collect1.trendyol.com ve collect4.trendyol.com gibi alan adlarının da var olma olasılığı doğduğundan, komut satırından nslookup collect1.trendyol.com şeklinde sorgulama yapılmıştır ve gerçekten var oldukları tespit edilmiştir. Trendyol ve diğer şirketlerin tüm alan adları yukarıda anlatılan DNS sorgu ve cevaplarının çekilmesi yöntemiyle, nDPI-4.2 kütüphanesine eklemek üzere, bulunup kaydedilmiştir. Fakat diğer yandan, sadece Trendyol için -RIPE veri tabanından elde edilenler de eklenirse- toplam 2095 IP adresinden oluşan bir liste oluşmuştur. Bunun üzerine, IP adreslerinin hepsi ikinci bir kontrolden geçirilmek üzere IP adreslerinin gerçek sahibinin kim olduğunun sorgulanabildiği https://dnschecker.org/ip-whois-lookup.php adresinden araştırılmıştır. Bu adresteki sorgu aracı aynı zamanda bir whois sorgusu yaptığından IP adresinin hangi kayıtçı kuruluşa -ARIN, APNIC, RIPE NCC gibi- ait olduğu bilgisini de vermektedir. Dolayısıyla, adreslerin kime ait olduğu, son olarak kayıtçı kuruluşlarda tekrar whois sorgusu yapılarak netleştirilmiştir. Tüm bu sorgulama sürecinin sonunda, Wireshark'tan elde ettiğimiz, Trendyol'un kullandığı IP adresleri listesindeki -RIPE'dan elde edilenler hariç- tüm adresler içerik dağıtım ağı (CDN) hizmeti sunmasıyla bilinen Alibaba Cloud ve CloudFlare gibi şirketlere ait çıkmıştır. Aynı şekilde, Hepsiburada, N11 ve Bilyoner şirketlerinin de CDN servisiyle ünlü Akamai şirketine ait IP adreslerini kullandıkları bulunmuştur. CDN hizmeti artık birçok firma tarafından tercih edildiği için birden fazla uygulamanın aynı adres aralığındaki farklı IP adreslerini kullandığı durumlarla karşılaşılabilir. Bu durumu en açık şekilde gösteren bir örnek Çizelge-1'de verilmiştir. Buna göre, nDPI-4.2 kütüphanesinde tanımlı TOR protokolünün belirleyicisi olarak kullanılan IP adresleri ile Trendyol'un kullandığı collect.trendyol.com alan adının Wireshark'tan elde edilen IP adres bilgilerinin hepsinin, APNIC kayıtçısındaki whois sorgularında, Alibaba şirketinin sağladığı Alibaba Cloud bulut hizmetinin IP adres aralığına denk geldiği anlaşılmaktadır. Dolayısıyla, bu bize ele aldığımız e-ticaret sitelerinin birden çok IP adresi kullanabileceğini ve bu IP adreslerinin de değişmeye müsait yapıda olduğunu göstermiştir. Bu durum, tespit etmek istediğimiz uygulama protokollerin belirleyicisi olarak IP adreslerinin seçilmek istenmemesinin temel nedenini teşkil etmiştir.

nDPI'ın 4.2'den hemen sonra çıkan 4.4 sürümünde yapılan daha sonraki incelemelerde TOR'un kullanmakta olduğu fakat Alibaba Cloud, Amazon gibi bulut servisi veren şirketlere ait IP adresinin kaldırıldığı fark edilmiştir.

Çizelge-1: Aynı bulut servisinden faydalanan iki farklı uygulama: TOR ve Trendyol

Trendyol Adresleri	TOR Adresleri
8.209.80.30 8.209.81.21 8.209.88.204 8.209.89.108	8.209.79.125 8.209.93.160 8.209.94.85 8.210.144.170
APNIC Sonuçları IP adres alanı aralığı: 8.209.64.0-8.209.127.255 Ağ Adı: ALICLOUD-DE	APNIC Sonuçları IP adres alanı aralığı: 8.209.64.0-8.209.127.255 Ağ Adı: ALICLOUD-DE

Paketlerin analizi yapılırken karşılaşılan ve IP adreslerinin yol açtığı somut bir problem ise ndpiReader aracının çıktılarında birçok alan adının yanlış kategori başlıkları altında etiketlendiğinin gözlemlenmesi olmuştur. Bunun sebebi, Cloudflare, Amazon AWS, Google ve Microsoft Azure gibi bulut hizmeti veren şirketlerin IP adreslerinden faydalanan organizasyonların alan adları nDPI-4.2 kütüphanesinde bulunmazken bulut hizmetini veren bu şirketlerin IP adreslerinin ise bulunmasıdır. Bu duruma birkaç örnek Çizelge-2'de verilmiştir.

Çizelge-2: nDPI-4.2'de yanlış kategorize edilen alan adları ve buna sebep olan bazı bulut hizmeti uygulamaları

Alan Adı ve IP Adresi	ndpiReader Çıktısı	Kayıtlı nDPI Protokolünün Adı ve IP Bilgileri
sync.srv.stackadapt.com 54.87.192.123	Proto: 91.265/TLS.Amazon AWS Cat: Cloud/13	NDPI_PROTOCOL_AMAZON_AWS 54.87.0.0/16

js.appboycdn.com 104.18.22.230	Proto: 91.220/TLS.Cloudflare Cat: Web/5	NDPI_PROTOCOL_CLOUDFLARE 104.16.0.0/12
api-js.mixpanel.com 107.178.240.159	Proto: 91.126/TLS.Google Cat: Web/5	NDPI_PROTOCOL_GOOGLE 107.178.192.0/18

Bu tabloya göre, .stackadapt.com ile başlayan alan adlarının reklam amaçlı veri toplayan StackAdapt yazılım şirketine, .appboycdn.com ile başlayan alan adlarının daha önce AppBoy olarak bilinen fakat şimdiki adı Braze olan bulut tabanlı yazılım şirketine ve .mixpanel.com ile başlayan alan adlarının ise kullanıcı etkileşimlerini izleyerek veri toplayıp raporlayan Mixpanel şirketine ait olduğu bulunmuştur. nDPI 4.2'de, bu örnekte olduğu gibi, yanlış kategorize edilen diğer tüm alan adları kütüphaneye doğru bir şekilde eklendikten sonra ağ paketlerinin cat: Advertisement/101 veya cat: Media/1 olarak tespit edilmesi sağlanmıştır.

Alan adlarının bulunabileceği ikinci bir yöntem ndpiReader uygulamasının kendisini kullanmaktır. nDPI'ın ayrıntılı modu seçildiği zaman, -v2, internet sunucu adı ile hizmet veren tüm sunucu adlarına ulaşılabilir. Otomasyon sürecinde bu yöntem kullanılmıştır.

Üçüncü ve son olarak, Türkiye'deki şirketlerin ticaret sicil numarası, ünvanı, adres bilgileri ve internet adresi gibi bilgilerine ulaşmayı sağlayan E-şirket şirket bilgi portalından [18] ve ilgili şirketin web adresinde hakkında veya biz kimiz gibi bölümlerinden alınan bilgiler incelenip karşılaştırıldığında, ağ trafiğinde o kuruma ait birkaç alan adının daha bulunması söz konusu olabilir. Örneğin, diğer şirketler için yapılan aramalarda bulunamasa da Trendyol için yapılan incelemede şirketin ticaret ünvanı DSM Grup Danışmanlık İletişim ve Satış Ticaret Anonim Şirketi olarak karşımıza çıkmaktadır. DSM grubu sadece Trendyol e-ticaret sitesinin sahibi olduğu için, Wireshark'ta dns contains "dsm" şeklinde filtrelediğimizde elde edilen aşağıdaki alan adları da Trendyol özelinde kaydedilmiştir:

- cdn.dsmcdn.com
- img-dsmncdn.mncdn.com
- dsmgrup.com
- static.dsmcdn.com

3.3 Kural Yazımı ve nDPI Kurulumu

nDPI kurulumundan önce yedi büyük şirket için kütüphaneye eklenmesi gereken iki kural kümesi vardır:

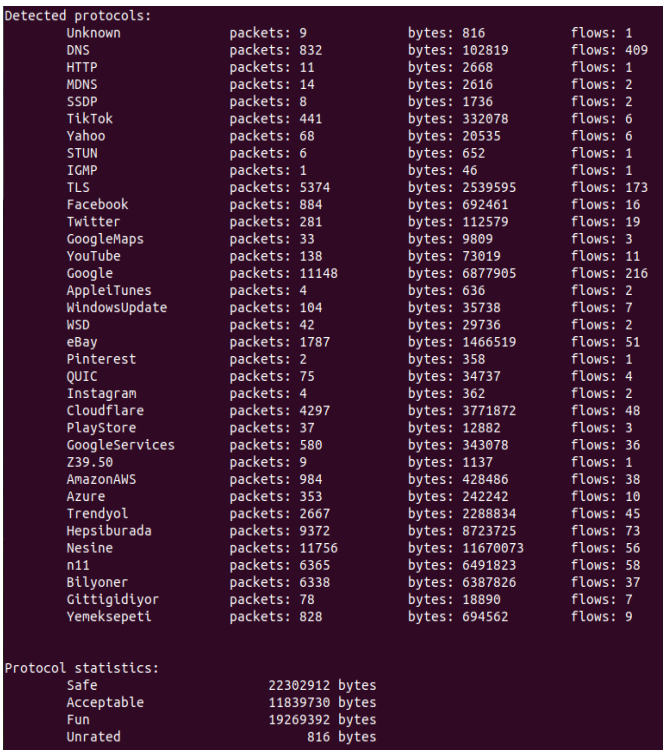
- /src/include klasöründeki ndpi_protocol_ids.h dosyasına her bir protokol için daha önce kullanımda olmayan kimlik numarası niteliğinde bir sayı yazılır. nDPI 4.2 versiyonunda en son 281 numarası kullanımda olduğu için, yedi sitenin protokolleri için sırasıyla karşılık gelecek şekilde, 282'den başlayarak 288'e kadar protokol adı ve kimlik numarası yazılır.
- İkinci olarak, her bir hedef protokolün ilgili alan adları /src/lib klasöründe yer alan ndpi_content_match.c.inc dosyasına sırasıyla alan adı, tespit edilmesi halinde ndpiReader'da nasıl isimlendirileceği, protokol adı, protokol kategorisi ve protokol türü bilgilerini içerecek şekilde kural satırları olarak yazılır. Bu kuralların birkaç örneği aşağıda verilmiştir:

- {"trendyol.com", "Trendyol", NDPI_PROTOCOL_TRENDYOL, NDPI_PROTOCOL_CATEGORY_SHOPPING, NDPI_PROTOCOL_SAFE, NDPI_PROTOCOL_DEFAULT_LEVEL}
- {"nesine.com", "Nesine", NDPI_PROTOCOL_NESINE, NDPI_PROTOCOL_CATEGORY_GAME, NDPI_PROTOCOL_FUN, NDPI_PROTOCOL_DEFAULT_LEVEL}
- {"yemeksepeti.com", "Yemeksepeti", NDPI_PROTOCOL_YEMEKSEPETI, NDPI_PROTOCOL_CATEGORY_SHOPPING, NDPI_PROTOCOL_SAFE, NDPI_PROTOCOL_DEFAULT_LEVEL}

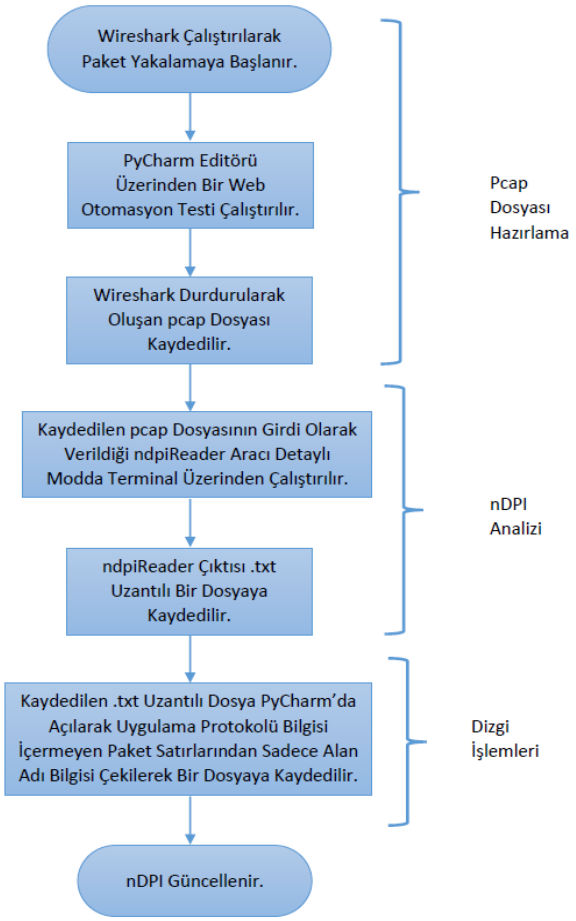
Bütün kural girdileri tamamlanınca artık nDPI kurulumu aşağıda yer alan komutlar sırasıyla çalıştırılarak gerçekleştirilebilir:

```
cd <nDPI_klasör_yolu>
sudo ./autogen.sh
sudo ./configure
sudo make
sudo make install
```

Kuralların eklenmesi ve nDPI kurulumunun yapılması sonrası ndpiReader aracının tespit ettiği uygulama protokollerine ait bir çıktının ekran görüntüsü Şekil-3'de verilmiştir.



Şekil-3: Eklenen protokollerin nDPI tarafından tespit edilmesi



Şekil-4: Otomasyon akışı

4. Otomasyon

Otomasyon sürecinin akış şeması Şekil 4'de verilmiştir. Buna göre, otomasyon üç bölümden oluşur; pcap dosyası hazırlama, nDPI analizi ve dizgi işlemleri. Bunlardan ilki, daha önce her bir web sayfasında manuel olarak yapılan gezintilerden elde edilen pcap veya pcapng uzantılı dosyaların web otomasyon testleri kullanılarak gerçekleştirilmesidir. Bunun için açık kaynak kodlu Selenium web sürücüsü ve Python programlama dili kullanılarak her bir web sayfasına ait bir adet test dosyası ve html raporunun yer aldığı bir Selenium web otomasyon çerçevesi oluşturulmuştur. Web otomasyon testleri çalıştırıldıktan sonra Wireshark durdurulur ve oluşan pcap dosyası kaydedilir. Böylece, otomasyon akışının ilk aşaması olan pcap dosyası hazırlama bölümü biter.

İkinci aşama nDPI analizidir. Pcap dosyası, -i anahtarı kullanılarak, komut satırından çalıştırılan ndpiReader uygulamasına girdi olarak verilir. nDPI tarafından ayrıntılı paket sonuçlarına ulaşmak için ayrıntılı mod, -v2 anahtarı kullanılır. nDPI analizi bitince terminal çıktısı .txt uzantılı bir dosyaya kaydedilir. Tüm ikinci aşama işlemleri için aşağıdaki komut satırının yazılması yeterlidir:

ndpiReader -i <dosya_adı>.pcapng -v2 >> <dosya_adı>.txt

Üçüncü aşama dizgi işlemleridir. İkinci aşamada kaydedilen metin belgesi herhangi bir Python IDE'si kullanılarak açılır. Tüm paket satırları çekilerek bir metin dosyasına kaydedilir. Daha sonra, sadece paket bilgilerini içeren bu metin belgesi açılıp nDPI tarafından uygulama protokolü bulunamamış olan paketler tespit edilir. Bunun için paket satırları taranarak TLS.<uygulama_protokolü> ya da DNS.<uygulama_protokolü> gibi dizgiler içermeyip bunun yerine [proto: 91/TLS] ya da [proto: 5/DNS] dizgilerini içeren

```
41  Hostname/SNI: guvendamgasi.org.tr
42  Hostname/SNI: bam.nr-data.net
43  Hostname/SNI: assets.humanz.com
44  Hostname/SNI: sync.crwdcntrl.net
45  Hostname/SNI: bilyoner.webinstats.com
46  Hostname/SNI: s.thebrighttag.com
47  Hostname/SNI: secure.adnxs.com
48  Hostname/SNI: vars.hotjar.com
49  Hostname/SNI: vars.hotjar.com
50  Hostname/SNI: se.semasio.net
51  Hostname/SNI: s.ad.smaato.net
52  Hostname/SNI: ad.360yield.com
53  Hostname/SNI: pixel.onaudience.com
54  Hostname/SNI: sslwidget.criteo.com
55  Hostname/SNI: pm.w55c.net
56  Hostname/SNI: sync.crwdcntrl.net
57  Hostname/SNI: an.yandex.ru
58  Hostname/SNI: criteo-sync.teads.tv
59  Hostname/SNI: bam.nr-data.net
60  Hostname/SNI: ads.betweendigital.com
61  Hostname/SNI: creativecdn.com
```

satırlar varsa bu satırlardaki "Hostname/SNI:<alan_adı>" dizgileri çekilir. Böylece, nDPI tarafından uygulama protokolü tespit edilmemiş olan paketler bulunarak yeni alan adları ayrı bir metin dosyasına kaydedilmiş olur. Dizgi işlemleri için yazılan betiğin örnek bir çıktısı Şekil-5'de verilmiştir. Bu şekle göre, 45. satırdaki bilyoner.webinstats.com alan adı yedi protokolden Bilyoner'e ait yeni bir alan adı olarak karşımıza çıkmaktadır.

Şekil-5: Dizgi işlemleri sonucu elde edilen çıktı bir örneği

Son olarak, yeni elde edilen alan adlarının kural olarak eklenip eklenmeyeceği belirlenerek nDPI'da güncelleme işlemleri yapılır. Burada önemli bir nokta, elde edilen yeni alan adları içerisinde, ndpi_protocol_ids.h dosyasında tanımlı eşsiz bir numaraya sahip olmayan -reklam veya CDN ağı gibi- uygulama protokollerine ait olan alan adlarının da bulunmasıdır. Bunun sebebi, bu tür protokollerin kategori bazlı tanımlanmasından dolayı paket içeriklerinde [proto: 91/TLS] ya da [proto: 5/DNS] dizgilerinin yer almaya devam etmesidir. Bu alan adları, nDPI'da yoksa, nDPI soyut kategorilerinden CUSTOM_CATEGORY_ADVERTISEMENT ve NDPI_PROTOCOL_CATEGORY_MEDIA kullanılarak eklenebilir. Bu şekilde 6 adet CDN ağına ait, 74 adet reklam ve takip amaçlı yazılımlara ait alan adları bulunmuştur. Bu alan adları nDPI'ın Şubat 2023'te çıkan son sürümü nDPI-4.6'da da bulunmamaktadır. Dolayısıyla, yanlış kategorize edilen alan adları probleminin devam etmemesi için bu çalışma kapsamında bulunan reklam ve takip amaçlı yazılım içeren alan adlarının nDPI kütüphanesine eklenmesine olan ihtiyaç devam etmektedir. nDPI-4.6'da –önceki 4.2 ve 4.4 sürümlerinin aksine- reklam, takip veya veri analitiği yapan şirketlere ait alan adları soyut kategori olarak ele alınmamıştır. Bu türdeki alan adlarının özelinde kullanılabilecek bir protokol adı ve buna karşılık bir kimlik numarası belirlenmiştir. Bu sebeple, yeni sürümde

NDPI_PROTOCOL_ADS_ANALYTICS_TRACK protokol adı ve NDPI_PROTOCOL_TRACKER_ADS protokol türü kullanılarak aşağıdaki kural şablonuna göre ekleme yapılmalıdır:

{ "<alan_adı>", "ADS_Analytic_Track",
NDPI_PROTOCOL_ADS_ANALYTICS_TRACK,
CUSTOM_CATEGORY_ADVERTISEMENT,
NDPI_PROTOCOL_TRACKER_ADS,
NDPI_PROTOCOL_DEFAULT_LEVEL }

5. Tartışma

Bu çalışmada, Türkiye'de yüksek kullanıcı sayısına sahip ve oldukça popüler e-ticaret siteleri ele alınmıştır. Bu tür uygulamaların ağdaki varlığının tespitinin üç yönden önemli olduğu söylenebilir.

Birincisi, daha etkili bir bant genişliği yönetimi için bir yerel ağda bulunan uç sistemlerin bu tür sitelere erişiminde, belirli dönemlerde veya saatlerde, sistem yetkilileri tarafından ölçeklendirilmesi ihtiyacı doğabilir. Kayıt yenileme başvuru veya sonuç öğrenme dönemlerinde yüksek yoğunluk yaşayan bir kampüs ağı bu duruma örnek olarak verilebilir.

İkincisi, e-ticaret sitelerinin kişisel veya bölgesel kullanım oranlarına ait istatiksel bilgiler çeşitli yönlerden faydalı olabilir. Örneğin, bir ISP'ye ait CDN ağında kullanımda olan uygulama protokollerinin hacmi göz önüne alınarak önceliklendirme tarifeleri uygulanabilir ve böylece hizmet kalitesi artırılabilir. Nitekim dört pcap dosyası üzerinden yapılan incelemelerde, Trendyol bağlantılarının akış sırası takip edildiğinde TCP bağlantı kurulumunun başında SYN ile başlayan üç yönlü el sıkışma ve hemen ardından onu takip eden TLS bağlantı kurulumu gerçekleştikten sonra bir CDN ağına bağlanıldığı tespit edilmiştir.

Üçüncü ve son olarak, gizlilik ve güvenlik gerekçeleriyle bu sitelerin yüksek güvenlik gereksinimi duyulan bazı ağlarda kısıtlanması ve hatta engellenmesi gerekebilir. Bu duruma en somut örneklerden birisi Yemeksepeti'ne ait bir web uygulama sunucusuna erişilerek gerçekleştirilen 18 Mart 2021 tarihli siber saldırıdır. Ele geçirilen veriler arasında ad, soyad, doğum tarihi, telefon numarası, eposta adresleri, adres bilgisi ve SHA-256 algoritması ile özeti alınmış şifreler bulunmaktadır. Şifrenin kendisi bilinmese bile bu tür veriler kötü niyetli kişiler için aslında oldukça değerli bir kaynaktır. Örneğin, Keeper Security firmasının Amerika'da tam zamanlı çalışan statüsünde bulunan ve işle ilgili çevrimiçi hesabına bir şifre yardımıyla giren 1000 kişi üzerinde yaptığı araştırmaya göre, bu çalışanların %44'ünün hem kişisel ve hem de iş ile alakalı hesaplarının şifresinin aynı olduğu sonucuna ulaşılmıştır [19]. Ayrıca, uzmanlara göre tüm internet kullanıcılarının yaklaşık %50'sinin bütün çevrimiçi hesaplarında hala aynı şifreyi kullandıkları tahmin edilmektedir [20]. Yemeksepeti'nde yaşanan veri hırsızlığı olayına dönecek olursak, özeti alınmış şifreleri elinde tutan bir saldırgan Yemeksepeti uygulamasına –kimlik doğrulama protokolünde de özeti alınmış şifrelerin oturumdan oturuma sabit tutulması gibi bir zayıflık varsa- Özet Geçirme (Pass the Hash - PtH) saldırısı gerçekleştirebilir. PtH saldırısı, bir saldırganın bir sistemden ele geçirdiği özeti alınmış şifreyi, sanki yeni bir oturum açıyormuş gibi, o sistemde yer alan kimlik doğrulama protokolüne kendi kimliğini doğrulatmak amacıyla kullanıp sisteme sızmasıdır. Şifre özetlerinin farklı oturumlarda sabit kaldığı zayıf bir sistemde şifre değişmedikçe özetler de değişmeyeceği için, şifre değişimine kadar, saldırganın sistemde fark edilmeden dolaşıp zaman kazanması mümkündür [21]. Böylece, özeti alınmış şifreleri deneyerek hesaplara, buradan da şifrenin açık haline ve belki de buradan da kritik yerlerde çalışan kişilerin iş yerlerinde kullandıkları hesaplara veya şirkete ait hassas verilerin bulunduğu sunuculara ulaşabilecektir. Diğer bir yandan, sadece doğum tarihi ve isim bilgilerinden bile şifreleri bulmaya yönelik bir saldırı yapılması mümkündür. Yapılan bir araştırmaya göre [22], Amerika'daki internet kullanıcılarının %60'a yakınının çevrimiçi hesaplarının parolasını bir isim veya doğum tarihi içeren bir dizgi oluşturmaktadır. Tüm bu saldırı metotları ile birlikte Yemeksepeti'nin KVKK'ya resmi olarak bildirdiği - 21 milyon 504 bin 83 rakamı ile Türkiye nüfusunun çeyreği eden- saldırıdan etkilenen kişi sayısı ve bu veri hırsızlığının saldırıdan ancak bir hafta sonra -25 Mart 2021'de- tespit edildiği göz önüne alındığında, bunun oldukça ciddi bir güvenlik açığı oluşturduğunu söylemek mümkündür. Diğer yandan, ele alınan Trendyol, N11 ve Hepsiburada şirketlerine ait web sitelerinde -sadece Yemeksepeti örneğindeki gibi ad, soyad, doğum tarihi, telefon numarası, eposta adresleri ve adres bilgisi gibi veriler değil- ürün arama, konum bilgisi, satın alma bilgileri gibi bir araya getirildiklerinde kişiler hakkında daha birçok çıkarımda (tercihleri, eğilimleri, sağlık durumu gibi) bulunmayı sağlayacak veriler de yer almaktadır. Elektronik ortamda gözetim ve takip amaçlı da toplanabilen bu veriler üçüncü parti şirketler ile de paylaşılabildiğinden önemli konumlarda çalışan kişiler ve çalıştıkları kurumlarda kısıtlayıcı önlemler alınması gerekebilir.

6. Sonuç

Bu araştırmada paketlerin sadece başlık bilgilerine değil beş katmanlı ağ modelindeki en üst katman olan uygulama katmanındaki yük içeriklerinin analiz edilerek uygulama protokollerinin tespit edilebildiği bir yapı olan derin paket inceleme yöntemi ele alınmıştır. Bu çalışmada, bu modeli kullanan açık kaynak kodlu nDPI aracına yeni protokoller eklenmiştir. Ele alınan protokoller şifreli bağlantı kullandıkları için sadece -SSL el sıkışma fazında bağlantı şifreli hale gelmeden önceki- sunucu adı ve ana bilgisayar adı bilgilerinin olduğu veriler kullanılarak Trendyol, Hepsiburada, Nesine.com, N11, Bilyoner, GittiGidiyor ve Yemeksepeti şirketlerinin kullandıkları web sitelerinin ağ trafiğinde tanımlanması sağlanmıştır. İleride oluşabilecek yeni ve güncel alan adlarının çekilebilmesi için kısmi bir otomasyon yazılmıştır. Bu web sitelerinde veriler toplanırken 6 adet CDN ağına ait, 74 adet reklam ve takip amaçlı yazılımlara ait alan adları bulunmuştur. Bu yazılımlar ilgili web sitesine bağlanıldıktan hemen sonra çalışmakta ve bağlantı sonuna kadar etkileşim devam etmektedir. Bu nedenle, internet üzerinden kullanıcı verilerinin işlenmesini ve belki de üçüncü parti şirketlere iletimini gerçekleştirme potansiyeline sahip bu tür sitelerin kullanımında, gizlilik ihtiyacı duyulan hassas verilerin yer aldığı kurumlar veya bu kurumlarda üst düzey yönetici pozisyonundaki yetkili kişiler için, kısıtlamaya gidilmesi ihtiyacı doğabilir. Bu kısıtlama neticesinde Yemeksepeti gibi olası saldırı durumları veya bu bilgilerin

doğrudan e-ticaret sitesinin kendisi tarafından kullanılması veya üçüncü parti şirketler ile paylaşması kaynaklı oluşabilecek güvenlik zafiyetlerinin önüne geçilmesi sağlanabilir. Bu çalışmada, ayrıca, nDPI tarafından yanlış kategorize edildiği tespit edilen ağ paketleri için yeni kural tanımlarının yapılarak bu gibi durumların düzeltilmesi sağlanmıştır.

Gelecek çalışmalarda ise sahte site ve link tuzaklarının DPI yöntemleri kullanılarak tespitinin sağlanması ile alana katkıda bulunulması hedeflenmektedir.

Kaynakça

[1] Vailshery, L. S., *"IoT connected devices worldwide 2030"*, https://www.statista.com/statistics/1183457/iot-connected-devicesworldwide/ , 2021.

[2] Brook, C., "*What is Deep Packet Inspection? How it Works, Use Cases for DPI, and More"*, https://digitalguardian.com/blog/what-deep-packet-inspection-how-it-works-use-cases-dpi-and-more, 2018.

[3] Renals, P. and Jacoby, G. A., *Blocking skype through deep packet inspection*, 42nd Hawaii International Conference on System Sciences, 2009, pp. 1-5, doi: 10.1109/HICSS.2009.90.

[4] Saputra, F. A., Nadhori, I. U. and Barry, B. F., *Detecting and blocking onion router traffic using deep packet inspection*, 2016 International Electronics Symposium (IES), 2016, pp. 283-288, doi: 10.1109/ELECSYM.2016.7861018.

[5] Radityatama, G. A., Lim, C. and Ipung, H. P., *Toward full enterprise software support on nDPI*, 6th International Conference on Information and Communication Technology (ICoICT), 2018, pp. 1-6, doi: 10.1109/ICoICT.2018.8528792.

[6] ntop, "*nDPI*", https://github.com/ntop/nDPI.git, 2023.

[7] Wireshark, "*About Wireshark"*, https://www.wireshark.org/about.html, 2023.

[8] Wikibooks, "*Intellectual Property and the Internet/Deep packet inspection*", https://en.wikibooks.org/wiki/Intellectual_Property_and_the_Internet/Deep_packet_inspection, 2023.

[9] Xu, C., Chen, S., Su, J., Yiu, S. M. and Hui, L. C. K., *A survey on regular expression matching for deep packet inspection: Applications, algorithms, and hardware platforms*, IEEE Communications Surveys & Tutorials, 18(4), 2016, pp. 2991-3029, doi: 10.1109/COMST.2016.2566669.

[10] Papadogiannaki, E. and Ioannidis, S., *A survey on encrypted network traffic analysis applications, techniques, and countermeasures*, ACM Computing Surveys, 2021, 54(6), pp. 1-35, https://doi.org/10.1145/3457904

[11] Mellia, M., Pescapè, A. and Salgarelli, L., *Traffic classification and its applications to modern networks*, Computer Networks (Vol. 53), 2009, pp. 759-760, 10.1016/j.comnet.2008.12.007.

[12] Bendrath, R., Global technology trends and national regulation: Explaining variation in the governance of deep packet inspection, International Studies Annual Convention, New York, 2009, pp. 15–18.

[13] ntop, "nDPI 4.6 (Feb 2023)", https://github.com/ntop/nDPI/releases/tag/4.6, 2023.

[14] Deri, L., Martinelli, M., Bujlow, T. and Cardigliano, A., *nDPI: Open-source high-speed deep packet inspection*, 2014 International Wireless Communications and Mobile Computing Conference, 2014, pp. 617-622, doi: 10.1109/IWCMC.2014.6906427.

[15] Attia H. B., "*q - Run SQL directly on CSV or TSV files*", http://harelba.github.io/q/.

[16] Fast Company Türkiye, "*En büyük 100 internet şirketi"*, https://fastcompany.com.tr/calisma-hayati/en-buyuk-100-internet-sirketi/ , 2020.

[17] DSM Grup Danışmanlık İletişim ve Satış Ticaret A.Ş, https://www.dsmgrup.com/.

[18] E-Şirket, https://e-sirket.mkk.com.tr/esir/.

[19] Whitney, L., "*How poor password habits put your organization at risk*", https://www.techrepublic.com/article/how-poor-password-habits-put-your-organization-at-risk/, 2021.

[20] Crafford, L., "*7 Bad Password Habits to Break Now*", https://blog.lastpass.com/2021/01/7-bad-password-habits-to-break-now-2/, 2021.

[21] Delinea, "*What is a Pass-the-Hash attack?*", https://delinea.com/what-is/pass-the-hash-attack-pth, 2022.

[22] Google and The Harris Poll, "*The United States of P@ssw0rd$*", https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/PasswordCheckup-HarrisPoll-InfographicFINAL.pdf, 2019.