D^2PI: Identifying Malware through Deep Packet Inspection with Deep Learning

Ronald Cheng, rscheng@cs.umd.edu
Gavin Watson, gkwatson@cs.umd.edu
University of Maryland, College Park

2017

İçerik

- Makale Tanımı ve Kapsamı
- Hangi Probleme Yönelik Geliştirilmiştir?
- Tasarım Detayları
- Sonuçları
- Neler Yapılabilir?

Derin Öğrenme Yöntemi ile Malware Tespiti

Günümüzde internet kullanımının büyümesi ile saldırganlar ağ üzerinden;

- bilgisayarlara arka kapı açan zararlı yazılım bulaştırma (trojan backdoors),
- kredi kartı bilgisi gibi önemli bilgileri çalma,
- Kullanıcıların kişisel verisini şifreleme ve fidye isteme (ransomware),

gibi saldırılarda avantaj sağlamışlardır.

Hangi Probleme Yönelik Geliştirilmiştir?

Zararlı yazılım analizleri kurumlarda yaygın olarak vaka durumunda tersine mühendislik, statik kod analizi ile yapılmaktadır. Bu yöntem çok zaman aldığı için ne yazık ki zararlı yazılımların amacına ulaşmasını engelleyememektedir.

Bu problem için geliştirilmiş olan Network Intrusion Detection Sistemleri (NIDS) ağı real time olarak tarar ve saldırıları tespit etmeye yardımcı olur. (Snort)
Bir ağda zafiyetli servise sahip bilgisayarları manipüle etmek, host tabanlı yetkisiz erişim ve yetki yükseltme saldırıları yapmak isteyen zararlı kişileri tespit eden bu NIDS yapıları yazılım ve/ve ya donanım tasarımlarından oluşur.

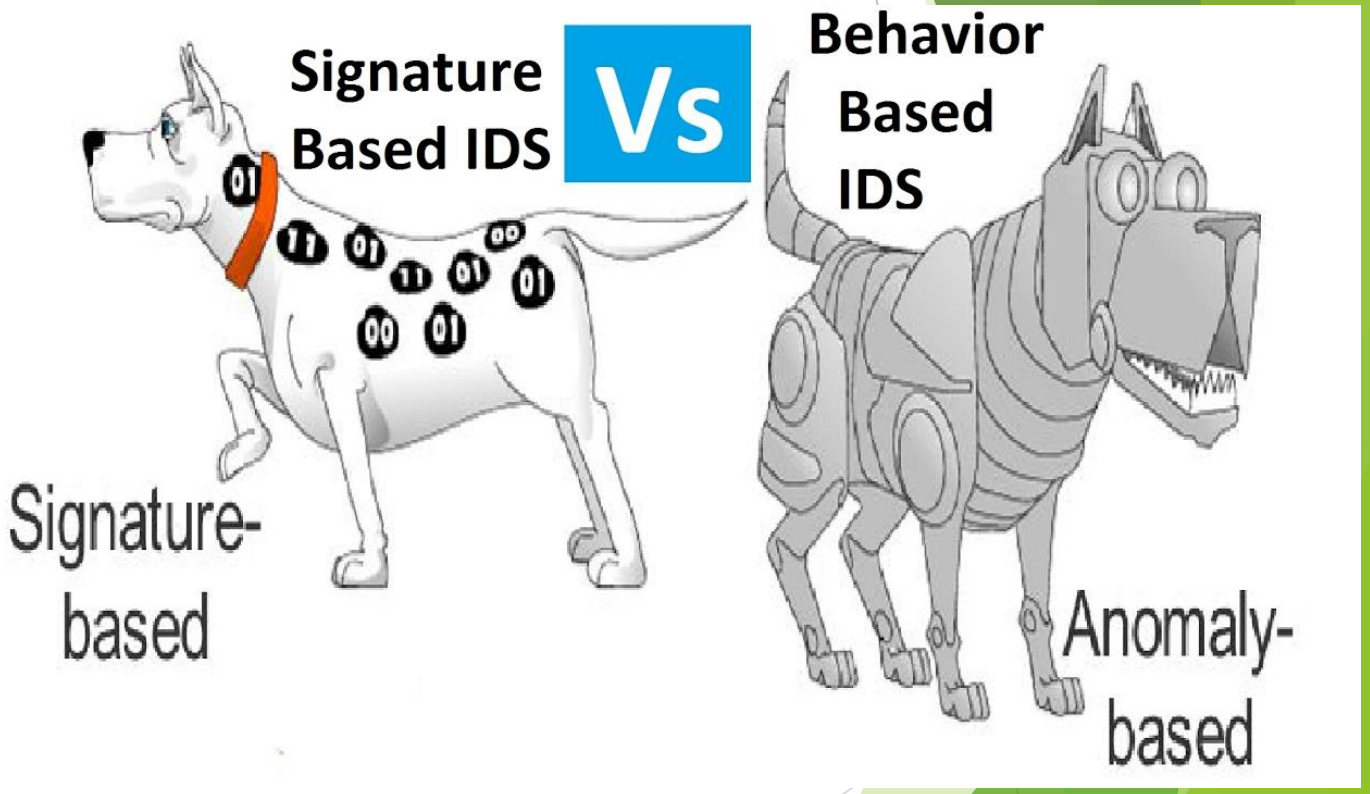
Genel Bilgiler - Çalışmalar

Bu tespit yöntemleri iki kategoride sınıflandırılır.

- Paket içeriğindeki bilgilerden ve imzalardan denetim yapmak
- Trafikte paket içeriklerini direk olarak tespit yöntemleri ile gözlemlemek (Dynamic (Çalıştırılır ve memory'deki durumuna bakılır (sandbox)) or Static (deep packet inspection vb.))

- Video: IDS - SIGNATURE based IDS Vs BEHAVIOR (Anomaly) based IDS

https://www.youtube.com/watch?v=dvDcwHPoD9w



Tasarım Detayları

Tüm bu yöntemler bilinen zararlıları tanımakta iyidirler fakat 0-day(henüz bilinmeyen) saldırıları tanımakta başarısızdırlar.

«Bu makalede ne malware operasyonlarının hakkında ön bilgiye sahip olmasına gerek olan ne de ağdan featureları extract eden Convolutional Neural Network (CNN) kullanılan yeni bir yaklaşım öne sürülmektedir. Çünkü featureları otomatik öğrenmektedir.»

Bu yöntem hem daha az maliyetlidir hem de yeni zararlı yazılımları tespit etmekte daha beceriklidir. Her yeni zararlı yazılım bulduğunda tekrar birkaç saat içinde kendini train eder. Fakat yöntemimiz çok gelişmiş değildir, ileriki çalışmalarda olacağını düşünüyoruz.

Bilgi: Bu makaledeki yönteme en yakın bir başka çalışma da VPN trafiği üzerinden zararlı Url leri CNN ile tanıyan Lotfollahi ve ark. [2] [3] Bu makaledeki sonuç güven vermiştir.

Payload Classifier

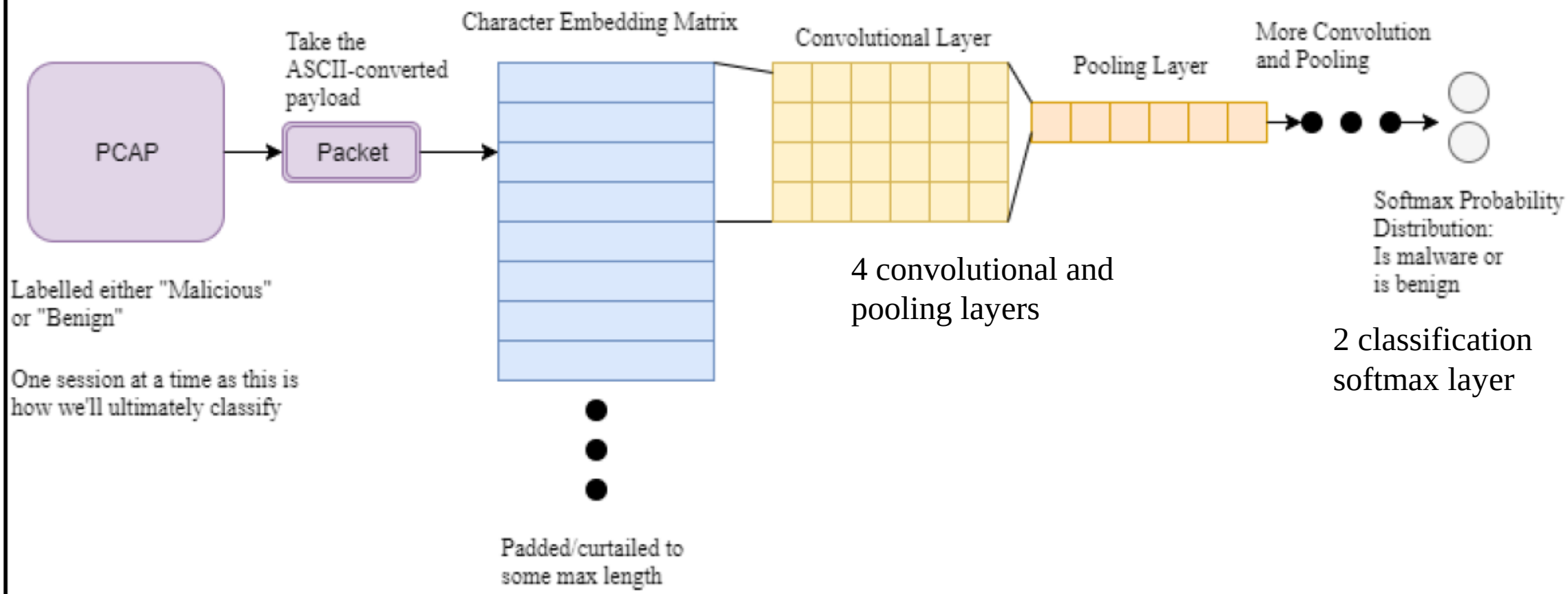


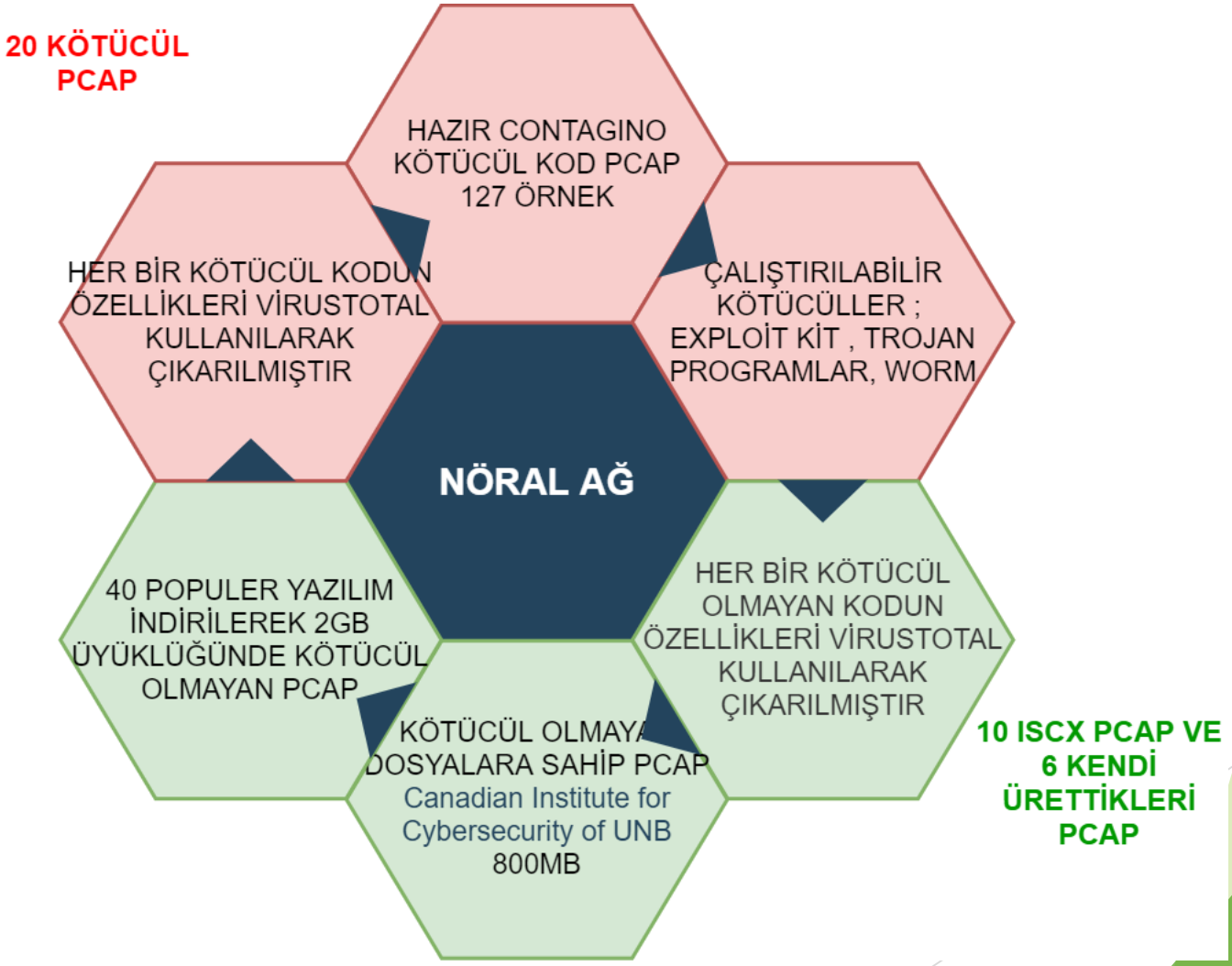
Figure 1

Tasarım Detayları

Bir önceki slaytta görülen tasarımın hepsi Tensorflow'un üstündeki Keras kütüphanesi aracılığıyla Python'da programlanmıştır.
Katman sayısı, window size, çekirdek sayısı, aktivasyon fonksiyonları ve pool katmanı boyutları gibi parametrelerin çoğu, önceki etkili çalışmaların tasarımlarına veya küçük bir veri kümesiyle yapılan test sonucunda bulunan değerlerdir.
Gömme katmanı, olası 128 ASCII karakterinin her biri için bağlamın bir vektörünü kodlayan 128x128 değerli önceden eğitilmiş bir karakter-karakter matrisidir. Tüm eğitim yükleri üç karakterden oluşan kayan bir bağlam penceresi ile yürütülür ve orta karakteri temsil eden vektöre, pencerede görülen diğer karakterlerin indeksinde ağırlık olarak bir ağırlık verilir. Gömme matrisindeki her vektör daha sonra bir büyüklüğe normalleştirilir. Bu, kelimeler yerine karakterler üzerinde çalışan bir word2vec modeline geçilerek iyileştirilebilecek nispeten basit bir gömme modelidir.

Tasarım Detayları

Tasarımlarında sadece TCP paketleri ile çalışmışlardır. Veri seti detayları yanda verilmiştir.



Sonuçlar

	ISCX Benign	Created Benign	Malicious
Predicted - Benign	131 (93.57%)	34 (100%)	36 (33.64%)
Predicted - Malicious	9 (6.43%)	0 (0%)	71 (66.36%)

Table 1: CNN results

	ISCX Benign	Created Benign	Malicious
Predicted - Benign	140 (100%)	34 (100%)	55 (51.40%)
Predicted - Malicious	0 (0%)	0 (0%)	52 (48.60%)

Table 2: Snort Results

$$F1 = \frac{2 * TruePos}{2 * TruePos + FalseNeg + FalsePos}$$

Equation 1

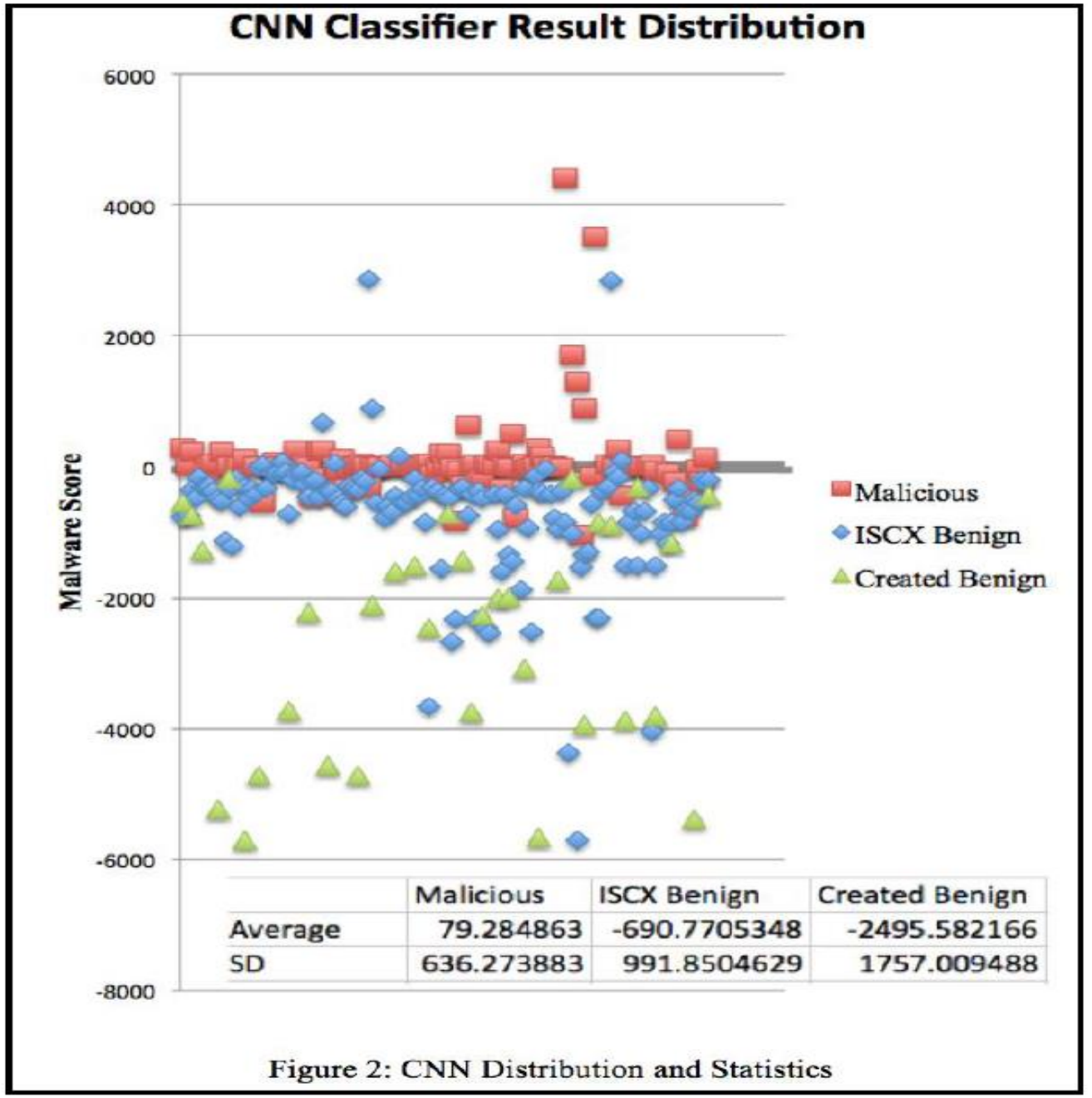
Sonuçlar

«Her ne kadar makaledeki tasarım sınıflandırıcısı bu çalışma sırasında canlı bir sistemde kullanılmasa da, analizimizin böyle bir ortamda iyi performans gösterebileceğine inanıyoruz» demişlerdir.

$$F1 = \frac{2 * TruePos}{2 * TruePos + FalseNeg + FalsePos}$$

Equation 1

Sınıflandırıcının F1 puanı 0,7724 iken, Snort'un değeri yalnızca 0,6003'tür. Tablo 1'de gösterildiği gibi, Snort'un iyi huylu trafiği tahmin ederken mükemmel olduğunu ve sınıflandırıcımızın olmadığını belirttiğimiz halde, bu sınıflandırıcı için umut verici bir sonuçtur.



	Malicious	ISCX Benign	Created Benign
Average	79.284863	-690.7705348	-2495.582166
SD	636.273883	991.8504629	1757.009488

Figure 2: CNN Distribution and Statistics

Appendix: Malware Predictions by File Name

Malicious Pcap Name	D^2PI prediction	Snort prediction
8202_tbd_ 6D2C12085F0018DAEB9C1A53E53FD4D1	Malicious	Benign
AlienSpyRAT_79E9DD35AEF6558461C4B93CD0C55B76	Malicious	Malicious
AlienspyRAT_DB46ADCFAE462E7C475C171FBE66DF82-WinXP	Malicious	Malicious
AlinaSpark_BE6371B8C90D8EECB749311373CEC0ED	Malicious	Benign
BIN_8202_6d2c12085f0018daeb9c1a53e53fd4d1	Malicious	Benign
BIN_9002_D4ED654BCDA42576FDDFE03361608CAA_2013-01-30	Benign	Benign
BIN_Alurewo_2502edca284bd8bf782a65123a22f9a6	Benign	Malicious
BIN_Andromeda_85F908A5BD0ADA2D72D138E038AECC7D_2013-04	Malicious	Benign
BIN_Bitcoinminer_12E717293715939C5196E604591A97DF-2013-05-12	Malicious	Benign
BIN_ChePro_2A5E5D3C536DA346849750A4B8C8613A-1	Benign	Benign
BIN_Cidox_Nuclear-EK_malware-traff-analysis-blog_2014-08-06	Benign	Malicious
BIN_CitadelPacked_2012-05	Malicious	Benign
BIN_CitadelUnpacked_2012-05	Malicious	Benign
BIN_Cutwail-Pushdo(1)_582DE032477E099EB1024D84C73E98C1	Malicious	Malicious
BIN_Cutwail-Pushdo(2)_582DE032477E099EB1024D84C73E98C1	Malicious	Malicious
BIN_DNSChanger_2011-12	Malicious	Malicious
BIN_DNSWatch_protux_4F8A44EF66384CCFAB737C8D7ADB4BB8_2012-11	Malicious	Benign
BIN_DarknessDDoS_v8g_F03Bc8Dcc090607F38Ffb3A36Ccacf48_2011-01	Malicious	Benign
BIN_Enfal_Lurid_0fb1b0833f723682346041d72ed112f9_2013-01	Malicious	Benign

Referanslar

[1] D2PI: Identifying Malware through Deep Packet Inspection with Deep Learning Ronald Cheng, rscheng@cs.umd.edu, Gavin Watson, gkwatson@cs.umd.edu, University of Maryland, College Park,2017

[2] M. Lotfollahi, R. Hossein Zade, M. Jafari Siavoshani and M. Saberian, "Deep Packet: A Novel Approach For Encrypted Traffic Classification Using Deep Learning", ARXIV, vol. 1709, no. 02656, 2017.

[13] S. Bagui, X. Fang, E. Kalaimannan, S. C. Bagui, J. Sheehan, Comparison of machine-learning algorithms for classification of vpn network traffic flow using time-related features, Journal of Cyber Security Technology (2017) 1–19.