

Olasılıksal SIP Ağı Benzetim Sistemi

A Probabilistic SIP Network Simulation System

Bariş Kurt¹, Çağatay Yıldız¹, Taha Yusuf Ceritli², Mehmet Yamaç³, Murat Semerci^{1,4}, Bülent Sankur³, Ali Taylan Cemgil¹

¹Bilgisayar Mühendisliği Bölümü, Boğaziçi Üniversitesi, İstanbul, Türkiye
{baris.kurt, cagatay.yildiz1, murat.semerci, taylan.cemgil}@boun.edu.tr

²Hesaplamalı Bilim ve Mühendislik Yüksek Lisans Anabilim Dalı, Boğaziçi Üniversitesi, İstanbul, Türkiye
{yusuf.ceritli}@boun.edu.tr

³Elektrik-Elektronik Mühendisliği Bölümü, Boğaziçi Üniversitesi, İstanbul, Türkiye
{mehmet.yamac, bulent.sankur}@boun.edu.tr

⁴NOVA Siber-Güvenlik, NETAŞ, İstanbul, Türkiye
{msemerci}@netas.com.tr

Özetçe —Bilgisayar ağları üzerindeki iletişim protokollerini geliştirmek ve başarımlarını irdelemek üzere büyük miktarda gerçek ağ verileriyle deneyler yapmak gereklidir. Ancak, gerçek ağlardan bu tür verinin elde edilmesi ve özellikle de “yer gerçekliği” ile işaretlenmesi, bazı durumlarda da olanaksız olacak denli zahmetlidir. Bu durumda benzetim modellerinden yola çıkarak çeşitli senaryolar için veri üretmek iyi bir alternatif olmaktadır. Bu çalışmada olasılıksal ağ modelleri kullanan bir SIP işletme benzetim sistemi geliştirdik ve bu sistemin SIP üzerindeki (D)DOS saldırılarının saptanmasında uygulamasını anlattık.

Anahtar Kelimeler—Ağ Benzetimi, SIP, Olasılıksal Ağ Modelleri

Abstract—Experimenting with large-scale real world data is crucial for the development of network protocol and investigate their performance. However, collecting such data from real networks, and especially to annotate them with ground truth proves to, if not impossible, too tedious. In such cases use of simulated data, generated for various network scenarios, becomes a plausible alternative. To this purpose, we have developed a SIP network simulation system that employs probabilistic network model and we have demonstrated its use for the performance analysis of statistical (D)DOS attack detectors.

Keywords—Network Simulation, SIP, Probabilistic Network Models

I. GİRİŞ

Büyük miktarda ağ verisi üzerinde deneyler yapmak, günümüz ağ uygulamalarının geliştirilmesi sürecinde oldukça önemlidir. Gerçek bir ağın gözlemlenmesi ve büyük miktarda verinin kaydedilmesi çeşitli zorlukları beraberinde getirir. Gözlemlenecek ağın altyapısına gerekli gözlemlene ve veri depolama sistemini kurmak genellikle zordur. Ayrıca ağ verileri kaçınılmaz olarak mahrem bilgi içerir ve kişiye özel bilgilerin gizliliğinin ihlalinin önlenmesi için isimleştirme gerekmektedir.

Bir ağ uygulamasını nihayetinde gerçek veriyle test etmek şart olsa da, geliştirme aşamasında benzetim süreçlerinden yararlanılabilir. Biz de bu ihtiyaçtan yola çıkarak, ağ üzerinde istatistik temelli örüntü tanıma uygulamaları geliştirilmesine

yardımcı olacak bir olasılıksal SIP (Session Initialization Protocol) ağı benzetim sistemi geliştirdik. SIP, her geçen gün daha da yaygınlaşan VoIP (Voice over IP) hizmetlerinin verilmesinde kullanılan iletişim protokollerinden biridir. Sistemimiz belirli sayıda kullanıcıları SIP iletişim kuralı kullanarak farklı senaryolar dahilinde birbirleriyle konuşturmayı amaçlamaktadır. Böylece geliştirilmekte olan uygulamalar, gerçek dünya verisiyle test edilmeden önce farklı senaryolarla irdelenebilmektedir.

Benzetim sistemimizin temelinde bir olasılıksal üretici ağ modeli yatar. Bu model kullanıcıların belirlediği bir takım üst düzey parametreler vasıtasıyla farklı davranışlar göstermeye yönlendirilir. Benzetim sistemimiz üretilen ağ trafiğinin geliştirilmekte olan ağ uygulamasına çevrimiçi ya da çevrimdışı olarak beslenmesi amacıyla bir yazılım arayüzü de sunmaktadır.

Makalemizin ikinci bölümünde benzetim sistemimizin ayrıntılarını anlattık. Üçüncü bölümde bir ağ saldırısı tespiti uygulamasının geliştirilmesi sürecinde benzetim sistemimizden nasıl yararlanılabileceğinden bahsettik. Son bölümde ise sistemin mevcut durumunu değerlendirip gelecekte eklenebilecek özellikleri tartıştık.

II. AĞ BENZETİM VE GÖZLEM SİSTEMİ

Sistem, üç ana birimden oluşmaktadır. Bunlardan ilki *benzetim birimi*dir. İkincisi oluşan SIP ağını gözlemlenmeye yarayan *gözlem birimi*dir. Üçüncü birim ise, abonelerin kayıtlı olduğu bir SIP sunucusudur. Biz bu çalışmamızda SIP sunucusu olarak ücretsiz bir sürümü mevcut olan, Asterisk tabanlı Trixbox [6] sistemini kullandık.

A. Benzetim Birimi

Benzetim birimi, olasılıksal üretici modelin belirlediği kurallar doğrultusunda gerçek zamanlı SIP konuşmaları üretir. Bu çalışmada SIP paketlerinin üretilmesi için PJSIP [7] adlı kütüphaneden faydalandık.

Olasılıksal üretici modeli, bir abonenin her durumda vereceği tepkilerin olasılık dağılımlarını içeren bir kütüphane olarak düşünebiliriz. Sistem kullanıcısı, üretici modelin

Parametre	Açıklaması	Ön Tanımlı Değeri
N	Abone Sayısı	100
K	Sosyal grup adedi	4
α	Grup dağılım parametresi	$[1, 1, 1, 1]$
a, b	Telefon defteri şekil parametreleri	2, 0.4
ρ, ϕ	Abone kaydı şekil parametreleri	2, 10
k, θ	Arama sıklığı şekil parametreleri	10, 10
f_{min}, f_{max}	Aramayı alt/üst limiti	0.9, 1
a_{min}, a_{max}	Cevap verme alt/üst limiti	0.8, 1
h_{min}, h_{max}	Beklemeye alma alt/üst limiti	0, 0.1
$\delta_{min}, \delta_{max}$	Ortalama konuşma süresi alt/üst limiti	20, 200

Tablo I: Sistem Parametreleri

parametrelerini (Tablo I) değiştirerek farklı ağ senaryolarının benzetimini yapabilir. Bu parametreler, aboneler arasındaki sosyal ağın yapısını ve abonelerin niteliklerini (her bir abone için sisteme kayıt olma zamanı, arama yapma sıklığı, gelen aramalara dönüş biçimi ve aramaların konuşma süresi) belirler.

Aşağıda bunlar ayrıntılarıyla açıklanmaktadır.

1) *Abone Sosyal Ağı*: Abonelerin birbirlerini arama sıklıklarını modellemek için rastlantısal öbek çizge modeli (stochastic block graph model) [2] kullanarak bir sosyal ağ oluşturduk. Bu modelde N adet abone K farklı gruba ayrılır. Bir abonenin gerek bulunduğu gruptaki kişilerle ve gerekse diğer gruplardaki kişilerle etkileşim sıklığı birer parametreyle belirlenir. Bir gerçek hayat örneği verecek olursak, bunu bir şirketin içinde aynı bölümde çalışanların birbirlerini belli bir oranda, diğer bölümlerdeki farklı oranlarda aramaları olarak düşünebiliriz.

Üretici model öncelikle abonelerin K adet gruba dağılımı belirler. Bir abonenin k grubuna bağlı olması olasılığına π_k diyecek olursak, K boyutlu π vektörünü bir α parametrelili bir Dirichlet dağılımdan örnekleyebiliriz.

$$\pi \sim \text{Dir}(\pi; \alpha) \quad (1)$$

$$p(\pi|\alpha) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k \pi_k^{\alpha_k - 1} \quad (2)$$

Bu sosyal ağdaki gruplaşmaları $N \times N$ boyutlu bir $G = (g_1^T, g_2^T, \dots, g_N^T)^T$ ikili matrisi ile ifade edelim, öyle ki her bir satır g_n bir abonenin bağlı olduğu grubu belirtsin. Abone ancak ve ancak k grubuna aitse $g_{n,k} = 1$ olacaktır. Bunun için bu matrisin her bir satırını π parametrelili bir kategorik dağılımdan örneklemek gerekir:

$$g_n \sim \mathcal{K}(g_n; \pi) \quad (3)$$

$$p(g_{n,k}) = \prod_k \pi_k^{g_{n,k}} \quad (4)$$

2) *Telefon Defteri*: Her abone, sosyal davranışı doğrultusunda sistemdeki diğer aboneleri hangi sıklıkta aradığını belirten bir telefon defterine sahiptir. Bu defteri oluştururken üretici model ilk olarak grupların birbirlerini arama eğilimlerini örnekler. $B_{i,j}$, i grubundan bir abonenin arama yapmak istediğinde j grubundan bir aboneyi arama olasılığı gösterebilir. Bu durumda $K \times K$ boyutlu B matrisini aşağıdaki şekilde örnekleyebiliriz:

$$B_{i,i} \sim \text{Beta}(B_{i,i}; a, b) \quad \forall i \in [1, K] \quad (5)$$

$$B_{i,j} \sim \text{Beta}(B_{i,j}; b, a) \quad \forall i \neq j \quad (6)$$

Burada Beta dağılımının a ve b parametreleri $a > b$ şeklinde verildiğinde grup içi konuşma ihtimalleri, gruplar arası konuşma ihtimallerinden daha fazla olacaktır.

Üretici model artık G abone grupları matrisi ve B gruplar arası iletişim matrisini kullanarak abonelerin telefon defterlerini oluşturabilir. Bu telefon defterini $N \times N$ boyutlu bir P matrisi ile gösterebiliriz, öyle ki m abonesinin herhangi bir arama yapmak istediğinde n abonesini arama olasılığı $P_{m,n}$ olsun. Bu durumda $\sum_n P_{m,n} = 1$ olacak şekilde aşağıdaki orantıyı yazabiliriz:

$$P_{m,n} \propto \prod_{i,j} B_{i,j}^{G_{m,i} G_{n,j}} \quad (7)$$

veya matris çarpımı ile ifade edecek olursak

$$P \propto G B G^T \quad (8)$$

Böylece oluşan abone sosyal ağı, tam bağlantılı ve ağırlıklı bir çizge oluşturur. Birbirlerini sık arayan aboneler arasındaki bağ kuvvetli iken, pek sık konuşmayan aboneler arasındaki bağ zayıftır.

3) *Abone Kaydı*: Benzetim başladığında tüm aboneler doğal olarak çevrimdışı durumdadır. Benzetim sistemi, ilk olarak aboneleri rastgele bir zaman beklettikten sonra SIP sunucusuna kaydederek çevrimiçi hale getirir. Bir n abonesinin kayıt olmadan önce beklediği süreye r_n diyelim. Üretici model bu süreyi bir Gamma dağılımından örnekler:

$$r_n \sim \mathcal{G}(\beta_i; \rho, \phi) \quad (9)$$

$$p(r_n|\rho, \phi) = \beta_i^{\rho-1} \frac{e^{-x/\phi}}{\phi^\rho \Gamma(\rho)} \quad (10)$$

4) *Arama Sıklıkları*: Herhangi bir n abonesi, sunucuya kayıt olduktan sonra veya mevcut konuşması bittikten sonra bir üstel dağılımdan örneklediği bir t_n bekleme süresi kadar bekledikten sonra yeni bir arama yapar. Abone, bu bekleme süresince kendisine bir arama gelirse aramayı cevaplayabilir.

$$t_n|\beta_n \sim \text{Exp}(t_n; \beta_n) \quad (11)$$

$$p(t_n|\beta_n) = 1/\beta_n \exp(-t_n/\beta_n) \quad (12)$$

Burada β_n abonenin arama yapma sıklığını belirten bir parametredir ve her bir abone için Gamma dağılımdan ayrı ayrı örneklenir:

$$\beta_i \sim \mathcal{G}(\beta_i; k, \theta) \quad (13)$$

$$p(\beta_i|k, \theta) = \beta_i^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)} \quad (14)$$

Böylece her abonenin arama yapma sıklığı farklı olmuş olur.

5) *Arama Yapma*: Aboneler, bekleme zamanları dolduğunda telefon defterlerinden rastgele bir abone seçip arama yapar. Bir n abonesinin, başka bir m abonesini arama ihtimali $P_{n,m}$ olduğundan, n abonesi arama yapacağı zaman arayacağı kişiyi P_n vektörünü kullanarak bir kategorik dağılımdan örnekleyecektir. Abonenin yaptığı seçime c_n diyecek olursak,

$$c_n|P_n \propto \mathcal{K}(c_n; P_n) \quad (15)$$

$$p(c_n|P_n) = \prod_{m=1}^N P_{n,m}^{[c_n=m]} \quad (16)$$

Burada, $[c_n = m]$ fonksiyonunu, c_n 'in m 'e eşit olduğu durumlarda 1'e, diğer durumlarda 0'a eşittir. Abone, aramak istediği diğer aboneye bir *INVITE* mesajı gönderir ve iletişimleri başlar.

6) *Aramalara Cevap Verme*: Bir abone, kendisine gelen aramayı kabul edebilir ya da reddedebilir, veya aramayı hiç farketmeyebilir. Sistemdeki bir n abonesinin gelen aramaları fark etme olasılığı, f_n , aramalara cevap verme olasılığı, a_n ve süren bir aramayı beklemeye alma ihtimali h_n , üretici model tarafından benzetim öncesinde aşağıdaki şekilde örneklenir:

$$f_n \sim \mathcal{U}(f_n; f_{min}, f_{max}) \quad (17)$$

$$a_n \sim \mathcal{U}(a_n; a_{min}, a_{max}) \quad (18)$$

$$h_n \sim \mathcal{U}(h_n; h_{min}, h_{max}) \quad (19)$$

Buna göre n abonesi benzetim sırasında kendisine arama geldiğinde eğer o anda başka bir konuşma yapmıyorsa, f_n olasılığıyla aramayı fark eder ve sonrasında a_n olasılığıyla aramaya cevap verir. Eğer arama geldiğinde kendisi de başka bir konuşma yapıyorsa, h_n olasılığıyla mevcut konuşmayı beklemeye alıp, gelen aramayı cevaplar. Bu durumda yeni konuşma bitince eski konuşmasına geri döner.

7) *Arama Süreleri*: Bir arama başarılı bir şekilde başladığı takdirde, konuşmanın süresi t_c abonelerce ortak bir şekilde belirlenir. Bir a abonesi başka bir b abonesini aradığında her iki abone de rastgele konuşma zamanları (t_a ve t_b) seçer ve bu zaman sonunda telefonu kapatır. Böylece konuşma kısa zamanı seçen tarafından sonlandırılır.

$$t_c = \min(t_a, t_b) \quad (20)$$

$$t_a | \delta_a \sim \mathcal{Exp}(t_a; \delta_a) \quad (21)$$

$$t_b | \delta_b \sim \mathcal{Exp}(t_b; \delta_b) \quad (22)$$

Abonelerin ortalama konuşma süreleri benzetim öncesinde üretici model tarafından birörnek dağılımdan örneklenir:

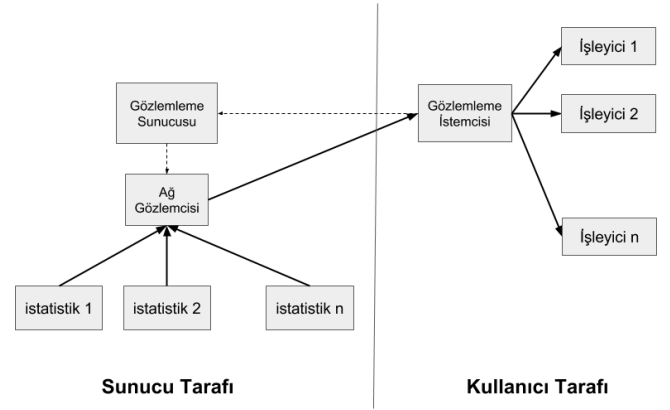
$$\delta_n \sim \mathcal{U}(\delta_n; \delta_{min}, \delta_{max}) \quad (23)$$

B. Gözlem Birimi

Yapılan benzetimin gerçek zamanlı izlenmesi, oluşan trafik verisinin kaydedilmesi ve bu verinin daha sonra tekrardan gerçek zamanlı veya hızlandırılmış şekilde işlenebilmesi amacıyla sisteme bir gözlem birimi ekledik. Bu birim SIP sunucusunun bulunduğu makine üzerinde çalışır ve SIP trafiğinin aktığı ilgili ethernet kanalını dinler. Kanaldan geçen her paket incelenir. Sunucuya gelen IP paketlerinin ve içlerindeki SIP mesajlarının istatistikleri toplanır. Toplanan SIP mesajlarından, o andaki konuşma trafiğinin haritası canlı olarak görselleştirilir.

Kimi uygulamalar ağ trafik akışıyla beraber gözlemlenen makinenin çalışma yüküyle de ilgilenmektedir. Biz de bu nedenle yukarıda sıralanan ağ tabanlı verilerin yanı sıra sunucunun kaynak kullanımıyla ilgili verileri (CPU, RAM kullanım yüzdeleri gibi) ve işletim sistemi istatistiklerini (iş parçacığı sayısı, TCP/UDP bağlantı sayısı, diske yazılan/okunan byte miktarı gibi) de topladık.

Şekil 1, gözlem biriminin yazılım mimarisini göstermektedir. Benzetim sisteminin kullanıcısı, amaçları doğrultusunda bir ya da birden çok işleyici örneği yaratır ve bunları gözlem istemcisine kaydeder. İşleyicilerinin ihtiyaç duyduğu veri



Şekil 1: Gözlem birimi mimarisini

tipleri, istemci aracılığıyla sunucudaki bir ağ kapısına İnternet üzerinden iletilir. Sunucuda çalışan gözlem sunucusu bu ağ kapısını 7/24 dinler. Buraya gelen mesaj ağ gözlemcisi örneğine iletilir ve böylece kullanıcının gözlemciye kaydı gerçekleşir. Ağ gözlemcisi, istatistik toplama modülleri yardımıyla kendisine kayıtlı kullanıcıların talep ettiği verileri toplar ve iletir.

Gözlem biriminin benzetim biriminden bağımsız tasarlanmış olması ve tüm trafik verilerini sadece IP paketlerinden takip etmesi, ona gerçek bir SIP trafiği altında da çalışabilme olanağı sağlar. Böylece sistemi kullanan geliştiriciler hem benzetilen hem de gerçek trafik altında deneylerini yapabilirler.

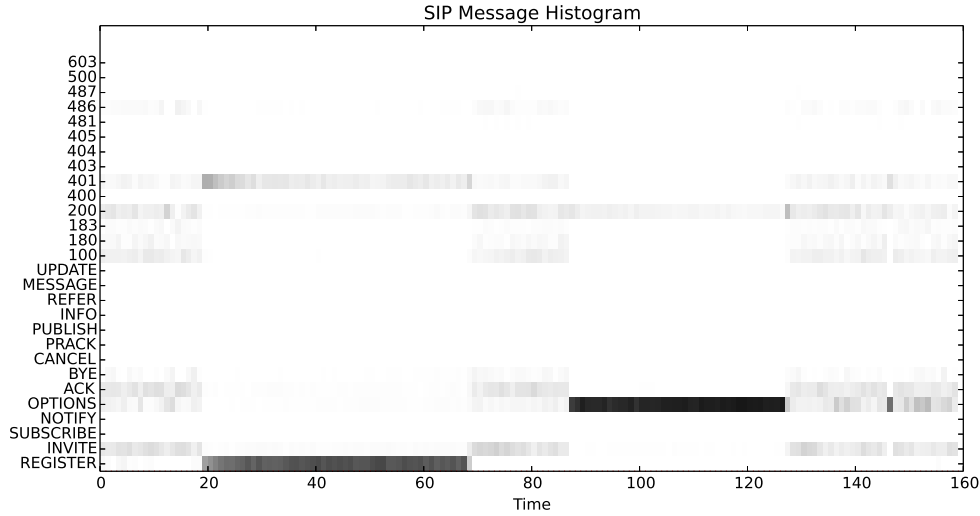
III. ÖRNEK UYGULAMA

Ağ benzetim ve gözlem sistemimizden, ağ trafiğinin izlenmesini gerektiren pek çok gerçek yaşam uygulamasında faydalanabiliriz. Bu bölümde, telefon ağını SIP üzerinden işleten bir kurumun Güvenlik İşletme Merkezi (GİM) inşasında benzetim sisteminin nasıl kullanılabileceğini ele alacağız. GİM'ler İnternet trafiğinin gözlemlendiği, kayıtların tutulduğu ve siber güvenlik tehditlerine karşı önlemlerin alındığı merkezlerdir.

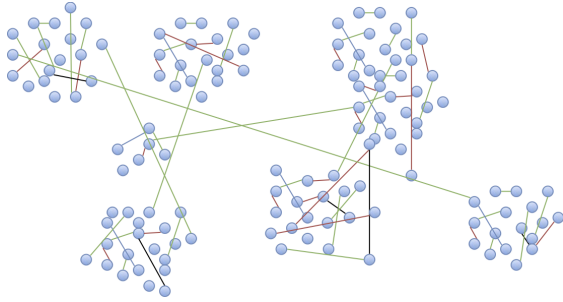
Geliştirmek istediğimiz uygulama bir SIP saldırı tespit uygulaması olduğundan, sistemi test etmek amacıyla hem günlük hayattaki normal SIP trafiğine, hem de SIP saldırıları trafiğine ihtiyaç duymaktayız. Bu testleri yapabilmek amacıyla benzetim sistemimizle normal SIP trafiği ve NETAŞ tarafından geliştirilen NOVA V-Spy adlı araçla [8] ise DDoS (Distributed Denial of Service) [3] saldırısı ürettik.

Benzetim sistemimiz ve NOVA V-Spy'nin ortaklaşa ürettiği trafiğe ait paket histogramı, Şekil 2'de verilmiştir. Bu histograma bakacak olursak, 20. ve 70. saniyeler arası yapılan *REGISTER* saldırısı, 90. ve 125. saniyeler arası yapılan *OPTIONS* saldırısıyla birlikte, benzetim sistemimizin ürettiği gündelik normal trafiği görebiliriz.

Normal trafik verisini üretmekte kullanılan rastgele üretilmiş sosyal ağ, Şekil 3'de verilmiştir. Burada her düğüm bir kullanıcıyı ve her ayrıt varolan bir konuşmayı temsil etmektedir. Benzetim parametrelerimizde ayarladığımız üzere gruplar arası konuşma sayısı, grup içi konuşma sayısına kıyasla daha seyrek olmaktadır.



Şekil 2: SIP Paket Histogramı



Şekil 3: Sosyal ağ ve konuşma grafiği

Benzetim sistemimiz veri üretmenin yanında tüm ağı dinleyip, SIP sunucu kaynak kullanımı bilgileri dahil çeşitli istatistikleri bize aktararak saldırı tespit aracı geliştirmemizde yardımcı oldu. Saldırı tespit yordamlarımızı ise diğer makalelerimizde anlattık [4], [5].

IV. DEĞERLENDİRME

Daha önce de bahsedildiği gibi, sunduğumuz benzetim sistemiyle birçok ağ uygulamasını sınamada kullanılabilecek sentetik veri üretmek mümkündür. Böylece gerçek bir ağ verisi gözlemlene ve bu veriyi mahremiyet korunarak kaydetme gibi uğraşlı bir yola pratik bir alternatif sunulmuş olur.

Olasılıksal modelin parametrelerinin ayarlanmasıyla birçok farklı SIP ağının benzetimi yapılabilir. Mesela, sosyal ağında iki grubun bulunduğu, grup içi konuşmaların nadiren gerçekleştiği ve aramaların neredeyse tamamının grupların birinden diğerine yapıldığı bir benzetimle çağrı merkezlerindeki akışı izlemek mümkün olabilir. Başka bir örnek olarak içinde irili-ufaklı çok sayıda grubun bulunduğu ve konuşma trafiğinin belirli bir periyotla değiştiği bir şirket içindeki telefon trafiğini düşünebiliriz. Son olarak, planlanan bir SIP altyapısının dayanıklılık testi için gerekli yoğun miktarda SIP trafiği de yine sistemimiz tarafından üretilebilir.

Örnek uygulamada gösterildiği gibi benzetim birimimiz bir saldırı tespit aracı geliştirilmesi sırasında günlük hayat trafiği üretmek amacıyla kullanıldı ve bu aracın gerçek veriyle test etme aşamasına hızlıca gelmesine yardımcı oldu. Benzetim sisteminin ikinci kısmı olan gözlem birimi ise, hem benzetim biriminin ürettiği hem de gerçek zamanlı üretilen veriyle çalışabildiğinden gerçek veri testlerine geçilebilmesi için fazladan kod yazılmasına gerek kalmadı.

Benzetim sistemimizin ileride açık kaynak kodlu olarak yayınlanması üzerine çalışmalarımız devam etmektedir.

V. BİLGİLENDİRME

Bu proje Boğaziçi Üniversitesi-NETAŞ ortak çalışması olup TEYDEB 3140701 nolu "Anomali Tespit ve Önlemede Öğrenen Sistem Mimarisi ile Birlikte Kalite Artırıcı ve Yüksek Hizmet Devamlılığı Sağlayan Zengin Servislerin VOIP Güvenlik Duvarı Ürünü Gerçekleştirilmesi" projesi kapsamında geliştirilmiştir. NOVA V-Gate ve NOVA V-Spy, NOVA Siber Güvenlik'in tescilli ticari ürünleridir.

KAYNAKÇA

- [1] H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, J. Rosenberg, "SIP: Session Initiation Protocol", RFC 3261, June 2002.
- [2] A. Goldenberg, A. X. Zheng, S. E. Fienberg, E. M. Airolidi, "A survey of statistical network models", Foundations and Trends in Machine Learning, 1-117, 2009.
- [3] S. Ehlert, D. Geneiatakis, T. Magedanz, "Survey of network security systems to counter SIP-based denial-of-service attacks", Journal of Computers & Security, 225-243, 2010.
- [4] Ç. Yıldız, T. Y. Ceritli, B. Kurt, B. Sankur, A. T. Cemgil, "Bayesçi Çoklu Değişim Noktası Modeliyle VoIP Ağlarda Saldırı Tespiti", 24. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı (Sunulacak), 2016.
- [5] M. Semerci, M. Yamaç, A. T. Cemgil, B. Sankur, D. S. Coşar, "Çevrimiçi Değişim Noktası Tespiti ile (D)DOS Saldırıların Seziimi", 24. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı (Sunulacak), 2016.
- [6] <http://www.fonality.com/trixbox>
- [7] <http://www.pjsip.org/>
- [8] <http://novacybersecurity.com/nova-vspy.html>