KÜTAHYA SAĞLIK BİLİMLERİ ÜNİVERSİTESİ
MÜHENDİSLİK VE DOĞA BİLİMLERİ FAKÜLTESİ

Thursday 14th March, 2024

YAPAY ZEKA DERSİ

Sistem LOG dosyalarını inceleyerek, sistemde aktif ya da öncesinde var olmuş yetkisiz girişlerin tespit edilmes

Barış AZAR **2118121004**

1 giriş

Güvenlik, günümüzde dijital sistemler için kritik bir öneme sahiptir. Bu proje, sistem güvenli§ini artırmak amacıyla yapay zeka kullanarak sistemde aktif olarak işlem yapan ya da geçmişte işlem yapmış izinsiz girişleri (hacker) tespit etmek amacıyla geliştirilecektir. Sistem loglarını inceleyerek, hackerları tespit etmeyi hedeflemektedir.

Linux sistem üzerinde bulunan ve tanımlı olarak(default) kaydedilen log dosyalarını analiz edilmesi planlanmaktadır. Programın aktif olarak çalışması değil, yalnızca istenilen zamanda çalışması planlanmaktadır.

2 Literatür Araştırması

SIEM ürünlerinin basit versiyonu olarak nitelendirebiliriz. Literatür daha çok ağ analizi üzerindedir.

3 Metodoloji

3.1 Veri Toplama ve Veri Seti

Proje, sistem loglarını toplamak için belirli bir metodoloji kullanacaktır. Veri setinde bir hacker tarafından izinsiz olarak ele geçirilen bilgisayarda olması muhtemel loglar ve izin verilen kullanıcı tarafından kullanılırken oluş- turulan loglar yer alacaktır. Hazır veri setleri ya da kendi oluşturduğum loglar veri setini oluşturacaktır.

Veri seti oluşturmak için bir hacker saldırı simülasyonu planım şu şekildedir: Tamamen yasal ve izinli olarak kendi bilgisayarım üzerindeki Linux işletim sistemi üzerinde önce hacker gibi davranarak bilgisayar ve sistem zaaflarını test edip bilgisayara sızarak logları toplayacağım. Sisteme giriş yaptıktan sonra içeride hala üzerine çalıştığım komut ve değişiklikleri yaparak ortaya çıkan logları alacağım.

Normal yani sistemi kullanılmasına izin verilen kullanıcı için log toplama planım ise şöyledir: Sisteme şifre ve parola ile önce yanlış sonra doğru olacak şekilde sisteme giriş yapacağım. Sistemde önce kullanıcı yetkisi ile sonra da süper kullanıcı yetkisi ile işlemler gerçekleştirip log toplayacağım.

2 farklı açıdan log toplama sebebim yapay zekanın yalnızca sisteme izinsiz giriş var demesi değil sistemde sorun yok çıktısı üretmesini sağlamak. Yani hastalıklı ve sağlıklı verinin bir arada bulunmasını örnek gösterebilirim.

3.2 Yapay Zeka Modeli

Literatür taraması ve yapay zeka modellerinin incelenmesi sonucunda; LOG analizi için RNN, özellikle de Long Short-Term Memory (LSTM) ile CNN modellerini paralel olarak kullanma kararı aldım. İki modelin özel avantajlarını birleştirerek sistemdeki potansiyel güvenlik ihlallerini daha etkili bir şekilde değerlendirmeyi amaçlıyorum. LSTM, metin verilerindeki uzun vadeli bağlantıları daha iyi anlama yeteneğine sahip olduğundan, zaman serilerini içeren LOG dosyalarındaki olayları modelleme konusunda güçlüdür. Ayrıca, CNN, metin verilerindeki özellikleri çıkarma konusunda etkili olduğu için, LOG dosyalarındaki belirli desenleri tespit etme kabiliyetiyle anomali tespiti ve güvenlik ihlallerini belirleme konusunda önemli bir avantaj sağlar.

Bu paralel kullanım, LSTM'in zamanla ilişkili olayları, CNN'in ise özellik çıkarma yeteneklerini birleştirerek, LOG dosyalarındaki hem zamanla ilişkili olayları hem de belirli desenleri daha etkili bir şekilde analiz etmeyi hedeflemektedir. Bu strateji, veri setindeki karmaşıklığı daha iyi ele alarak, güvenlik açısından daha sağlam bir sistem oluşturmayı amaçlamaktadır.

Ancak, her iki modelin de eğitim süreçleri, modelin karmaşıklığı ve kullanım senaryoları dikkate alınmalıdır. LSTM'nin uzun vadeli bağlantıları işleme kabiliyeti, zaman içindeki olayları modelleme açısından avantajlıdır, ancak eğitim süreci zaman alabilir. Aynı zamanda, CNN'in filtre boyutları ve mimari seçimi gibi parametrelerin dikkatlice ayarlanması gerekmektedir.

Sonuç olarak, LSTM ve CNN'nin paralel kullanımı, LOG analizi sürecini daha kapsamlı ve etkili hale getirerek, sistemdeki güvenlik açıklarını daha etkili bir şekilde tespit etmeye yönelik bir strateji olarak benimsenmiştir. Bu yaklaşım, her iki modelin avantajlarını birleştirerek güçlü bir analitik yetenek sunmayı amaçlamaktadır

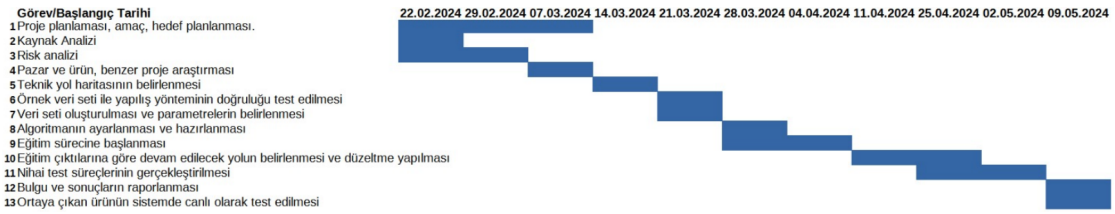
4 Sonuçlar ve Değerlendirme

Geliştirilen yapay zeka modeli ile elde edilen sonuçlar, projenin başarısını değerlendirecek. Modelin hassasiyeti, doğruluğu ve tespit yetenekleri, gerçek dünya senaryolarında nasıl performans gösterdiği üzerine odaklanacaktır. Ortaya çıkan ürünü canlı bir sistemde denemek için adımlar atılması düşünülmektedir.

4.1 GANTT Chart ve İş Akış Planı

İş akış şeması, zaman aralıkları ve yapılacak görevler henüz kesinleştirilmemiştir. Gelecek 2 hafta içerisinde iş akış planı, zaman aralıkları ve yapılacak görevlerin tanımları netleştirilecektir. Bu süre zarfında yapılan değişiklikler ilgili haftada bildirilecek ve nihai sonuçlar gelecek 2 hafta içinde sunulacaktır. Şekil 1'de görüldüğü gibi, GANTT Chart iş paketlerinin tanımını ve zaman çizelgesini göstermektedir.

Şekil 1: GANTT CHART



Şekil 1'de görebileceği üzere iş akış planı gösterilmektedir.

1. Kaynakça

- ADeep Learning Türkiye. (22 Kasım 2020). RNN Nedir? Nasıl Çalışır? Erişim tarihi: [5.03.2024]. https://medium.com/deep-learning-turkiye/rnn-nedir-nasıl-çalışır-9e5d572689e1
- TensorFlow. Zaman Serileri Yapılandırılmış Veri Eğitimi. Erişim tarihi: [10.03.2024]. https://www.tensorflow.org/tutorials/structured$_{d}ata/time_{s}eries?hl = tr$
- Application of Artificial Intelligence and Machine Learning in Security Operations Center Middle Georgia State University (2023). Erişim Tarihi: [10.03.2024] https://comp.mga.edu/static/media/doctoralpapers/2023Islam0516152253.pdf
- konvolüsyonel Sinir Ağları Tabanlı Türkçe Metin Sınıflandırma (2023). Erişim Tarihi: [13.03.2024] https://dergipark.org.tr/en/download/article-file/2609278
- Convolutional Neural Network (ConvNet yada CNN) nedir, nasıl çalışır? (02.10.2018). Erişim Tarihi: [13.03.2024] https://medium.com/@tuncerergin/convolutional-neural-network-convnet-yadacnn-nedir-nasil-calisir-97a0f5d34cad
- CCNN (Convolutional Neural Networks) Nedir? (30.09.2021). Erişim Tarihi: [13.03.2024] https://bartubozkurt35.medium.com/cnn-convolutional-neural-networks-nedira5bafc4a82a1