NetFlowScope: A novel Intrusion Detection System

Mohammed Y. Aldarwbi, Arash H. Lashkari, Ali A. Ghorbani

Canadian Institute for Cybersecurity (CIC), University of New Brunswick (UNB)

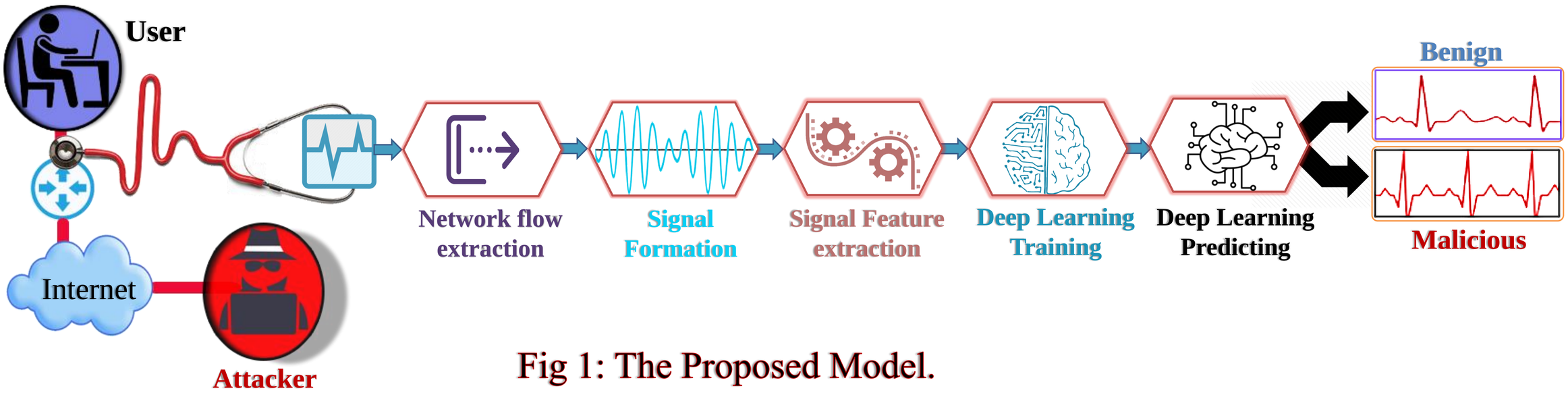


Fig 1: The Proposed Model.

Abstract

Currently there are many tools that allow to protect computer networks of possible attacks, these applications allow detect intruders. Among all the available security solutions Intrusion Detection Systems (IDS) are widely used for many purposes whether we are looking to monitor clients (host) or computer networks to identify the latest threats. With IDS, we can collect and use information of the various types of known attacks for the defence of all infrastructure, in addition to identifying points or attack attempts, allowing not only the registration but the continuous improvement of the security environment. The development of anomaly network intrusion detection systems (ANIDS) is a challenge for researchers, due to the growth of computer networks, and the constantly appear of new attacks. In this work, a novel ANIDS technique is proposed.

Idea

As a matter of fact, digital signal recognition using Deep Learning techniques is one of the most mature techniques in the artificial intelligence field. Therefore, the idea as presented in Fig.1 is to capture the network traffic and represent it as signals. Then, the power of deep learning techniques such as Convolutional Neural Network will be utilized in classifying the extracted signals as if they are sounds. The proposed idea have been validated using two well-known and recent data-sets, namly, ”NSL-KDD” and ”CICIDS2017”.

Notes

Network traffic flow is a sequence of network packets (sent and received) that share the same five tuple: Source IP, Source Port, Destination IP, Destination Port, and Protocol.

Data-sets: The most recent flow-based datasets , namely, NSL-KDD, and CICIDS2017 have been used. They contain benign traffic and the up-to-date common attacks, which resembles the true real-world data.

Nominal Features: The symbolic features such as (protocol, service, and flag) were expanded using 1-N encoding. These features have been converted into numeric by multiplying the index of the value by a constant number (e). The encoded data contains (3 from protocol, 64 from service and 11 from flag).

Service feature	
Key	Value
ftp_data	i*e
Smtp	i*e
telnet	i*e
http	i*e
:	:
:	:

Ref

[1] Yin, C., Zhu, Y., Fei, J., & He, X. (**2017**). A deep learning approach for intrusion detection using recurrent neural networks. Ieee Access, 5, 21954-21961.

[2] Gao, J., Chai, S., Zhang, B., & Xia, Y. (**2019**). Research on Network Intrusion Detection Based on Incremental Extreme Learning Machine and Adaptive Principal Component Analysis. Energies, 12(7), 1223.

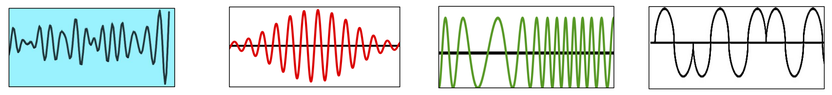
[3] Abdulhammed, R., Musafer, H., Alessa, A., Faezipour, M., & Abuzneid, A. (**2019**). Features Dimensionality Reduction Approaches for Machine Learning Based Network Intrusion Detection. Electronics, 8(3), 322.

[4] Ahmim, A., Maglaras, L., Ferrag, M. A., Derdour, M., & Janicke, H. (**2018**). A Novel Hierarchical Intrusion Detection System based on Decision Tree and Rules-based Models. arXiv preprint arXiv:1812.09059.

Signal Formation

Network Traffic flow **ranked** features (rank is a score of 8 different feature selection techniques)				
f_1	f_2	f_3	.....	f_n

$$Y = \sum_{j=1}^{n} \left(f_1 * \sin\left(2\pi * f_j * t + f_2\right)\right)$$



Where :

- (**n**) is the number of features of a flow
- (f_1) is the value of the top ranked feature representing the signal amplitude.
- (f_2) is the value of the second top ranked feature representing the signal phase shift.
- (f_j) is the value of the other features representing the signal frequency.
- (t) is the sampling rate
- (Y) is the superposition of the network flow features.

Signal Features

These are the extracted features out of the superposition signal of the flow:

- **Mel Frequency Cepstral Coefficients** (**MFCCs**) : MFCCs of a signal are a set of features which concisely describe the overall shape of a spectral envelope.
- **Mel-scaled spectrogram** (**Mel**): Mel represents an acoustic time-frequency representation of a sound, and the name mel comes from the word melody to indicate that the scale is based on pitch comparisons.
- **Chroma** is a typically a 12-element feature vector indicating how much energy of each pitch class, {C, C#, D, D#, E, ..., B}, is present in the signal. Chroma features are a well-established tool for processing and analysing music data.
- **Spectral Contrast** which could roughly reflect the relative distribution of the harmonic and non-harmonic components in the spectrum and it had good discrimination in music type classification.
- **Tonnetz** or **Harmonic Network** is a well known planar representation of pitch relations.

Results

These tables presents the preliminary obtained results of the proposed idea on two datasets.

NSL-KDD	
Method	Accuracy
Proposed	**85.54 %**
[1]	**83.28%**
[2]	**81.22%**

CICIDS2017	
Method	Accuracy
Proposed	**99.00%**
[3]	99.60%
[4]	96.66%

Conclusion

The proposed technique shows a promising results which encouraged us to do more research to come up with an enhanced signal formation model, evaluate the signal extracted features, and test several deep learning classifiers in order to find the best anomaly detection system ever.