On the Usage of Generative Models for Network Anomaly Detection in Multivariate Time-Series

Gastón García González
IE–FING, UDELAR, Uruguay
AIT Austrian Institute of Technology
gastong@fing.edu.uy

Pedro Casas
AIT Austrian Institute of Technology
pedro.casas@ait.ac.at

Alicia Fernández
IE–FING, UDELAR, Uruguay
alicia@fing.edu.uy

Gabriel Gómez
IE–FING, UDELAR, Uruguay
ggomez@fing.edu.uy

ABSTRACT

Despite the many attempts and approaches for anomaly detection explored over the years, the automatic detection of rare events in data communication networks remains a complex problem. In this paper we introduce *Net-GAN*, a novel approach to network anomaly detection in time-series, using recurrent neural networks (RNNs) and generative adversarial networks (GAN). Different from the state of the art, which traditionally focuses on univariate measurements, Net-GAN detects anomalies in multivariate time-series, exploiting temporal dependencies through RNNs. Net-GAN discovers the underlying distribution of the baseline, multivariate data, without making any assumptions on its nature, offering a powerful approach to detect anomalies in complex, difficult to model network monitoring data. We further exploit the concepts behind generative models to conceive Net-VAE, a complementary approach to Net-GAN for network anomaly detection, based on variational auto-encoders (VAE). We evaluate Net-GAN and Net-VAE in different monitoring scenarios, including anomaly detection in IoT sensor data, and intrusion detection in network measurements. Generative models represent a promising approach for network anomaly detection, especially when considering the complexity and ever-growing number of time-series to monitor in operational networks.

KEYWORDS

Deep Learning, Anomaly Detection, Multivariate Time-Series, Generative Models

1 INTRODUCTION

Network monitoring data generally consists of hundreds or thousands of counters periodically collected in the form of time-series, resulting in a complex-to-analyze multivariate time-series process (MTS). In particular, detecting anomalies in such multivariate, temporal data is challenging. Without loss of generality, we refer to the MTS as a set of n, non-iid time series sampled at the same rate, referred to as $x_t = \{x_t(1), x_t(2), \ldots, x_t(n)\} \in \mathbb{R}^n$. Current approaches to anomaly detection tackle this challenge by either focusing on univariate time-series analysis – running an independent detector for each time-series $x_t(i)$, or by considering multi-dimensional input data $x \in \mathbb{R}^n$ at each time t, neglecting the temporal aspects of the MTS. To improve the state of affairs we propose Net-GAN, a novel unsupervised approach to anomaly detection in MTS data, based on Recurrent Neural Networks (RNNs), trained through a Generative Adversarial Networks framework (GAN) [1].

The usage of generative models for semi-supervised anomaly detection helps to solve two major problems faced in this specific field: the high imbalance between normal operation and anomaly instances, as well as the lack of labeled instances for learning and validation purposes. Generative models such as Variational Auto-Encoders (VAEs) or Generative Adversarial Networks (GANs) are powerful approaches to learn the underlying distributions of data samples, in a purely data-driven, model-agnostic manner. Such models can be used in the practice to construct better baselines (i.e., profiles for normal operation) for the anomaly detection task, improving the identification of instances which deviate from this baseline. Most of previous work in this direction treats data as temporally independent samples, neglecting the information provided by causality and temporal correlation.

To capture the temporal correlations characterizing an MTS, we adapt the original GAN model proposed in [1], replacing the multilayer perceptrons by recursive, LSTM networks for both generator and discriminator models. The input data is therefore sequences of multi-dimensional measurements, of length T: $\{x_{t-T}, ..., x_t\}$. In a similar direction, we also explore the performance of other powerful generative models for anomaly detection, using in particular VAEs. Variational auto-encoders are a generative version of classical auto-encoders, but different from GANs, they make

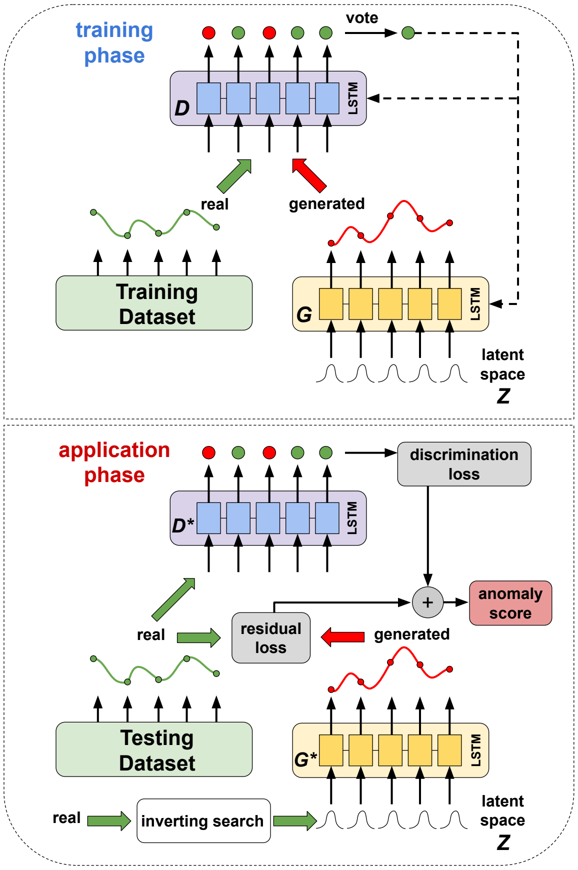


Figure 1: Net-GAN architecture and its application.

strong assumptions on the generative distribution they try to estimate. We refer to this flavor of Net-GAN as Net-VAE.

The reminder of the paper is organized as follows: Section 2 briefly overviews the related work; in Section 3, we describe the Net-GAN and Net-VAE approaches; Section 4 reports the preliminary evaluation results obtained with Net-GAN/Net-VAE models in the detection of anomalies in different datasets. Finally, Section 5 concludes the paper.

2 RELATED WORK

Generally speaking, operational network monitoring systems rely on rules and fingerprints to detect anomalous behaviors. There are nevertheless multiple extensive surveys on general domain anomaly detection techniques [5] as well as on network anomaly detection [6, 7], including machine learning-based approaches. There is a particularly extensive literature in the application of learning-based approaches for automatic traffic monitoring and analysis [4], including detection. Their main limitation as compared to our work is their (generally) supervised nature, which requires ground-truth data for learning. There is also a vast literature on clustering–based approaches for unsupervised network anomaly detection and analysis, mostly targeting the security domain [9–11].

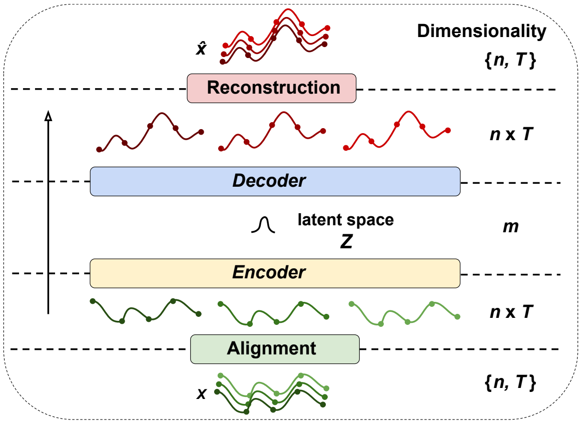


Figure 2: Net-VAE architecture.

When it comes to the application of generative models for anomaly detection, there are recent papers on GANs for time-series synthesizing and anomaly detection [8, 12, 13]. Examples of GANs for anomaly detection, as well as VAEs for anomaly detection, are presented in [2] and [3], respectively. A particularly interesting model for anomaly detection using GANs is BiGAN [14], which extends the original GAN architecture by adding the learning of the inverse mapping which maps the data back to the latent representation.

3 THE NET-GAN/VAE APPROACH

GANs are a framework for the estimation of generative models via an adversarial process in which two models, a discriminator D and a generator G, are trained simultaneously, in an adversarial manner. The generator G aim is to capture the – unknown and potentially complex, data distribution, while the discriminator D estimates the probability that a sample came from the training data rather than G. To learn a generative distribution p_g over the learning data x, the generator builds a mapping from a prior noise distribution p_z to a data space as $G(z)$. The discriminator outputs a single scalar $D(x)$ representing the probability that input x came from real data rather than from p_g.

Fig. 1 depicts the Net-GAN architecture and both the model training and anomaly detection procedures. In the ***training phase*** (top), the generator G draws synthetic sample sequences $G(z)$ from Gaussian noise – the latent space Z, with the objective of deceiving the discriminator D, which in turn learns to determine whether training samples are real or derived from the generative distribution. The classification result proposed by D is additionally fed back to G,

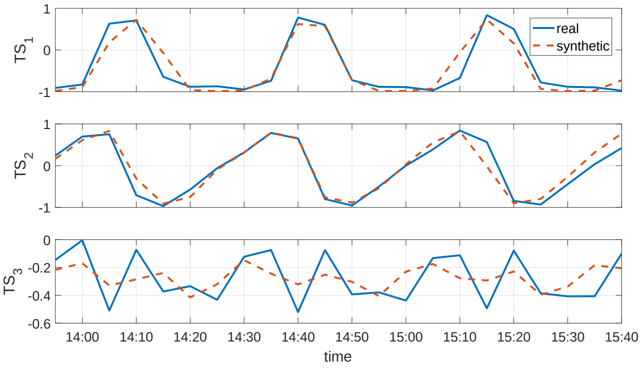


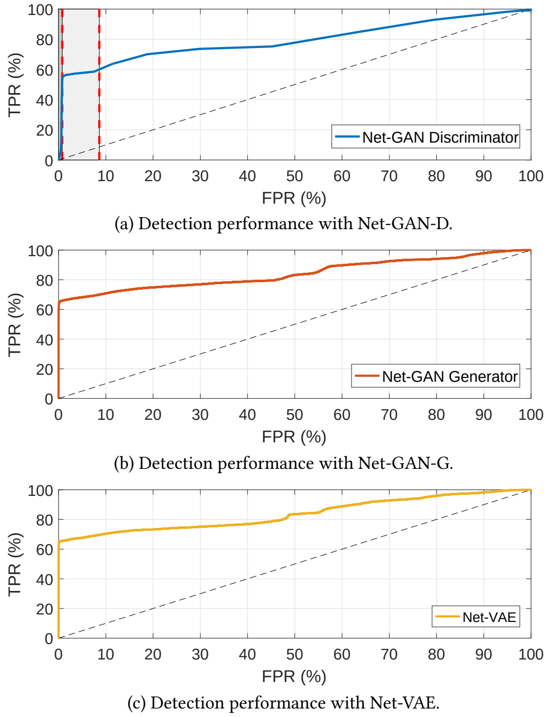
Figure 3: Net-GAN: synthetically generated time-series.

serving as a reinforcement loop to guide the generation process. As both G and D compete to achieve their adversarial tasks, synthetic samples become more and more "realistic", and the discriminator becomes robust to noise, improving the detection of non-conforming (i.e., out of the baseline) samples.

In the ***application phase*** (down), the trained discriminator D^* acts naturally as an anomaly detector, detecting deviations from the baseline, through a *discrimination loss* function. The trained generator G^* is also used to improve detection performance, serving as baseline generation; by doing an inverse search in the latent space – for example, constructing an inverse model for the generator [3, 14], we find the sample $z \in Z$ which generates the closest sample $\hat{x}$ to the tested one x, producing a *residual loss*. This step also be approximated by randomly sampling the latent space, and keeping the sample $\hat{x}$ which better approximates x. Both the discrimination and the residual loss functions can be combined into an *anomaly score*, which is compared to a calibrated threshold to take the final decision.

As we mentioned before, one of the salient features of Net-GAN is that we use LSTM networks for both G and D, instead of the traditionally used multilayer perceptrons or convolutional neural networks. Being recursive by conception, LSTMs can capture temporal dependencies in the data, which feed-forward networks fail to do. This is paramount when it comes to time-series analysis. In Section 4 we test separately the detection performance provided by the trained discriminator D^* and the trained generator G^*, using random sampling as reverse-search technique.

In the case of VAE, the architecture is composed of the standard encoder and decoder functions which form the auto-encoder. An auto-encoder is a type of neural network used to learn both efficient representations of the input data, typically for dimensionality reduction – the *Encoder*, along with (re)generation models, which generate representations from the reduced encoding latent space as close as possible to its original input – the *Decoder*. Fig. 2 depicts the Net-VAE architecture, which is composed of two data alignment and reconstruction layers – to pre-process the time-series and post-process the auto-encoded samples, and two three-layer, feed-forward neural networks, representing the encoder and the decoder, respectively. The detection with Net-VAE is simply done through the *residual loss* obtained by applying the trained VAE to the input testing sample x; if the difference between the input x and its reconstruction $\hat{x}$ is greater than a detection threshold, an anomaly is declared.



(a) Detection performance with Net-GAN-D.

(b) Detection performance with Net-GAN-G.

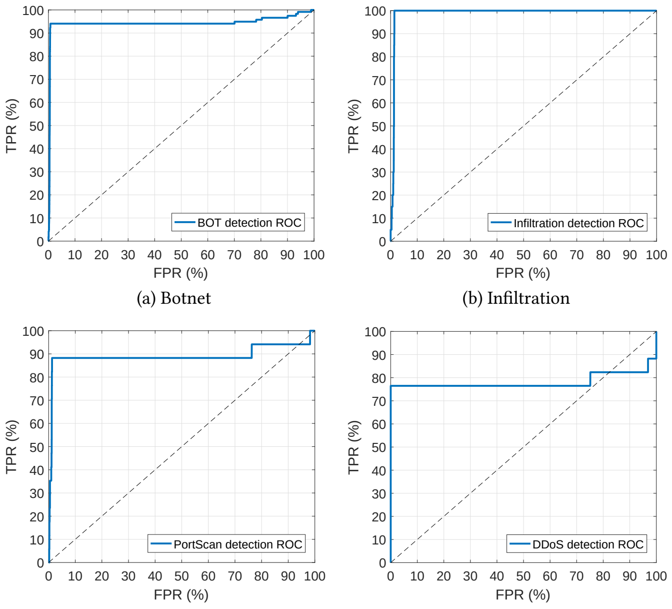
(c) Detection performance with Net-VAE.

Figure 4: Detection of anomalies in CPS data.

In terms of time-series preparation and processing, both Net-GAN and Net-VAE operate through a sliding window of T samples – using a unitary step. At each new step of the analysis, a matrix consisting of n chunks of T consecutive samples each is fed to the models, see Fig. 2. Finally, distance among time-series chunks is computed at a per-sample basis, and an anomaly is declared as soon as one or more of the samples deviate from the baseline by more than a detection threshold. To avoid false alarm due to spurious variations in the time-series, each sample T generally represents a temporal aggregation of measurements, e.g., a moving average.

4 PRELIMINARY EVALUATION RESULTS

We evaluate Net-GAN's detection performance on two different publicly available datasets, here referred to as CPS



(a) Botnet

(b) Infiltration

(c) Port Scan

(d) DDoS

Figure 5: Detection of attacks in SYN-NET data, using Net-GAN-D as underlying detection model.

[13] and SYN-NET [15]. The CPS dataset consists of synthetically generated attacks targeting industrial control systems, in particular a safe water treatment plant. It includes IoT sensor measurements for 51 different physical properties related to the plant and the water treatment process. In total, 946.722 samples are collected with a 1-second resolution, over 11 days. The SYN-NET dataset is a synthetically generated dataset for network intrusion detection, including normal operation traffic generated by a group of 25 users (e.g., HTTP/HTTPS browsing, FTP file transfer, SSH and mail, etc.), with controlled attacks over-imposed, of very different nature. In particular, we test Net-GAN for the detection of botnet traffic (0.2% of total flows), DDoS attacks (4.3%), port scan activity (16.6%), and infiltration activity (only 36 flows). SYN-NET consists of more than a million flows – 83%/17% benign/malign traffic. For the sake of completeness and performance-benchmarking, we also evaluate Net-VAE on the CPS dataset.

To show the generation capabilities of Net-GAN, Fig. 3 depicts some of the (min-max normalized) time series generated by the trained generator G^*, along with the corresponding real time-series. To reduce noise, samples are aggregated in time-windows of 10'. Time-series of higher magnitude are better reconstructed (TS_1 and TS_2), whereas noisy ones – such as TS_3, are more difficult to track.

Fig. 4 reports the detection performance achieved by Net-GAN and Net-VAE in the CPS dataset, in the form of ROC curves. Fig. 4(a) reports the obtained results when using Net-GAN's discriminator D^* as detector (Net-GAN-D), Fig. 4(b) uses Net-GAN's generator G^* as detector, and Fig. 4(c) uses Net-VAE.

While results are preliminary and depend on the size and quality of the analyzed datasets – we are currently working with bigger network measurement datasets, Net-GAN-D detects 56% of the attacks with a FPR below 1%, whereas both Net-GAN-G and Net-VAE detect close to 70% of the attacks without false alarms. This shows that the generative capabilities of both approaches are extremely useful when it comes to detecting deviations. Still, the overall performance is rather poor for this scenario, which we believe is linked to the quality of the data. As reference, detection performance obtained in previous work [13] for the same dataset is aligned with our results.

To conclude the paper, and to showcase the performance of Net-GAN in a different dataset, Fig. 5 reports the detection performance of Net-GAN-D in the SYN-NET dataset, for the four considered network attacks. About 93%, 100%, 89%, and 78% of the attacks are detected with a FPR below 1%, for botnet, infiltration, port scan, and DDoS, respectively, showing promising results in this specific scenario.

5 CONCLUDING REMARKS

In this paper, we have explored the application of modern generative models to the detection of anomalies in multivariate time-series. We have presented and evaluated *Net-GAN*, a novel approach to network anomaly detection in time-series, using RNNs and GANs. Different from the state of the art, which traditionally focuses on univariate measurements, Net-GAN detects anomalies in multivariate time-series, exploiting temporal dependencies through RNNs. Net-GAN discovers the underlying distribution of the baseline, multivariate data, without making any assumptions on its nature, offering a powerful approach to detect anomalies in complex, difficult to model network monitoring data. We complemented Net-GAN with an alternative approach based on variational auto-encoders, which also represent a powerful and appealing model for the specific task.

The evaluation of both detection approaches in two different monitoring scenarios, including anomaly detection in IoT sensor data, and intrusion detection in network measurements, shows promising results. Besides the specific performance attained in terms of detection properties and generation of false alarms, generative-based models for anomaly detection might prove extremely useful when confronted with the monitoring of complex and highly-dimensional systems – such as modern networks, where traditional modeling approaches would fall short to capture the underlying (joint) distributions of the data. A deeper and more comprehensive evaluation of Net-GAN and Net-VAE in real networking data at large scale is part of our ongoing work.

ACKNOWLEDGMENTS

This work has been partially supported by the ANII-FMV project with reference FMV_1_2019_1_155850 "Anomaly Detection with Continual and Streaming Machine Learning on Big Data Telecommunications Networks", and by Telefónica. Gastón García was supported by the ANII scholarship No. POS_FMV_2020_1_1009239.

REFERENCES

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems*, NIPS'14, 2014.

[2] S. Zavrak and M. Iskefiyeli, "Anomaly-based Intrusion Detection from Network Flow Features using Variational Autoencoder," *IEEE Access*, vol. 8, pp. 108346–108358, 2020.

[3] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, "Efficient GAN-based Anomaly Detection," *CoRR*, vol. abs/1802.06222, 2018.

[4] R. Boutaba, et al., "A Comprehensive Survey on Machine Learning for Networking: Evolution, Applications and Research Opportunities," *Journal of Internet Services and Applications*, 9(1):16, 2018.

[5] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: a Survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[6] M. Ahmed, A. N. Mahmood, and J. Hu, "A Survey of Network Anomaly Detection Techniques," *Journal of Network and Computer Applications*, vol. 60, pp. 19–31, 2016.

[7] W. Zhang, Q. Yang, and Y. Geng, "A survey of Anomaly Detection Methods in Networks," in *2009 International Symposium on Computer Network and Multimedia Technology*, 2009.

[8] F. D. Mattia, P. Galeone, M. D. Simoni, and E. Ghelfi, "A Survey on GANs for Anomaly Detection," *CoRR*, vol. abs/1906.11632, 2019.

[9] P. Casas, J. Mazel, and P. Owezarski, "Unsupervised Network Intrusion Detection Systems: Detecting the Unknown without Knowledge," *Computer Communications*, vol. 35, no. 7, pp. 772–783, 2012.

[10] J. Dromard, G. Roudière, and P. Owezarski, "Online and Scalable Unsupervised Network Anomaly Detection Method," *IEEE Transactions on Network and Service Management*, vol. 14, no. 1, pp. 34–47, 2016.

[11] M. Goldstein and S. Uchida, "A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data," *PloS one*, vol. 11, no. 4, p. e0152173, 2016.

[12] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs," *CoRR*, vol. abs/1706.02633, 2017.

[13] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S. Ng, "MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks," in *Artificial Neural Networks and Machine Learning - ICANN 2019*.

[14] J. Donahue, P. Krähenbühl, T. Darrell, "Adversarial Feature Learning," *CoRR*, vol. abs/1605.09782, 2016.

[15] I. Sharafaldin., A. H. Lashkari., and A. A. Ghorbani., "Toward Generating a new Intrusion Detection Dataset and Intrusion Traffic Characterization," in *Proceedings of the 4th International Conference on Information Systems Security and Privacy - Volume 1: ICISSP,*, pp. 108–116, 2018.