Use of Machine Learning Techniques for Network Assurance

Sujay Kanungo

Independent Researcher
Boston, USA
sujay.kanungo@gmail.com

Abstract:
For the past two decades Network Assurance and Network Automation have become the center stage to address the problems of the hyper scale of the enterprise and service provider networks. As the size of the networks grew that brought to the center the problem of managing and debugging these networks which requires a lot of human intervention. There has been effort by the networking industry to develop solutions which helps thousands of networking devices and nodes managed by controllers and monitor the networks performance but some of these tools and software rely mostly on static rule based analysis. For the past decade some use of Artificial Intelligence and Machine Learning has been done to create software and tools which are used to manage and troubleshoot the network infrastructure. In this paper we will analyze the current AI & ML methods used for network Assurance, the current limitations and the future scope of improvements.

Keywords: Network Assurance, Machine Learning, Networking, ML based Network Automation.

I. INTRODUCTION

The internet has grown to huge scale and complexity after the internet was first discovered by ARPANET and routing protocols were discovered in a dorm room of a college. If the projections are to be believed the connected devices in internet by 2030 will be around 120 Billion devices. Such a huge scale and complexity of these huge networks requires software and tools which are capable of managing the nodes , automated debugging of the nodes and predicting failures and degradation of the networks and nodes.

Thus such a huge scale requires modern tools and software which use the capabilities of the Machine Learning and Artificial Intelligence for network assurance.

Network Assurance deploys measurement software to monitor service quality and detect abnormal changes in network behavior, producing statistics and alerts when issues arise.

Machine Learning is well suited for data analysis , enabling efficient anomaly detection, traffic and network behavior classification. Machine Learning techniques can be used within network assurance to handle the huge scale of data produced by the networks; which enables the automatic analysis of this data set identifying network issues such as denial of service attacks, route leaks , resource failures and anomalies.

II. ML TECHNIQUES IN USE FOR NETWORK ASSURANCE

A. Supervised Learning

This method employs observation paired with associated outcomes to learn a function capable of inferring outputs from new inputs. This type of learning requires historical data which comprises numerical samples of this nature or form. In networking, supervised learning or algorithms can be used to extract information from labeled data sets to classify flows and estimate traffic volumes. Supervised learning enables the ranking of network points of presence by predicting the congestion patterns.

B. Unsupervised Learning

Unsupervised Learning offers the possibility for the automatic classification based on training dataset. This can be used for Association Rule Mining and clustering. Clustering aims to assemble groups of data points that exhibit high intra-group similarity while maintaining clear differences from other groups. Various

methods of clustering are centroid, density and hierarchical. Association rule mining extracts association rules from a dataset by identifying itemsets that satisfy specified support and confidence thresholds. Application of clustering in network assurance is in places where issues , logs can be grouped together to provide comprehension about the network problems. Grouping or clustering helps identifying problems or allows for further statistical analysis. K-means clustering has been employed to classify compromised smart meters , detect anomalies and malicious users.

III. DATA REQUIREMENTS OF BUILDING A NETWORK ASSURANCE SYSTEM USING ML

ML methods usually operate at the flow level; data preprocessing, feature extraction and feature engineering are best carried out by domain experts and constitute the main prerequisite for implementation of ML for network assurance. Similarly , ML techniques need large amounts of data, whose collection can raise significant privacy concerns. Best way to handle such problems is by encrypting any personally identifiable information. Planes such as the management and control can provide deep insights into network state without accessing actual data traffic; however if management data are handled by a third party, access policies need to comply with regulations or even contractual obligations.

A. Data Collection

The current evolution of cellular networks towards higher transmission rates, greater device density and a range of supporting services signifies a major shift in paradigm. As traffic volumes increase and network architectures change, effective and efficient quality of service (QoS) prediction becomes crucial. Machine Learning (ML) emerges as a natural methodology for predictive systems, enabling networks to behave proactively and sustain stringent requirements. ML research and development have long encompassed various network management and security related data. This prior work serves as a foundation for network operators seeking to adopt ML powered assurance platforms. Fault management represents a key aspect where ML techniques provide a promising avenue. The ability to adapt to network dynamicity is fundamental for both end-to-end network automation and more specifically, fault management. In the broader realm of network administration, ML techniques contribute to an assurance framework capable of detecting network faults, anomalies , and deviations, thereby promoting security and resilience. Network assurance aims to automate the design, execution, and assessment of tests across distributed points. The overarching goal is to utilize ML techniques to configure and deploy assurance tests that continuously evaluate network security and resilience at scale.

B. Data Preprocessing

Effective data preprocessing is essential before employing machine learning techniques for network assurance. Data is the foundation of any machine learning approach whether supervised, unsupervised, or reinforcement learning and appropriate transformation is necessary to enhance the performance of these algorithms.

Data collection involves gathering information about network conditions from various sources within the infrastructure. Depending on the application , this data may require filtering to remove irrelevant aspects and cleaning or transform the data into appropriate vectors, often through sampling or aggregation of time series. Domain knowledge related to network operation can inform the selection of these features.

Providing machine learning algorithms with high – quality , representative data is crucial, yet achieving this can be challenging in practice. Nonetheless, once preprocessing is complete, a machine learning approach can advance to model training and related procedures.

C. Feature Engineering

A dataset typically exhibits features of different types, which need to be treated in specific ways. Extracting precise features may be challenging in certain contexts, such as when sources are noisy or when relevant information is obscured within noisy data or complex patterns. Most ML techniques anticipate that valuable information will be encapsulated in a few features, which thus need to be prioritized and emphasized during feature engineering. High variance in certain features might indicate mechanisms distinct from those

governing relevant modes of variation; consequently , these features are often undesirable and may require manipulation or removal.

A feature can have a particular distribution, and the technique employed may rely on certain assumptions about this distribution. If a feature deviates significantly from the expected form, it may be beneficial to transform it into a space where the models assumptions are more valid. The representation of a feature within the learning algorithm is also crucial; for example, when employing classifiers that measure distance (e.g., nearest-neighbor or support vector machines), normalization over a comparable range can enhance effectiveness and prevent the distortion of results. Because features may be assembled from multiple physical sources, unit selection, base level, and scale can vary and need to be adjusted appropriately.

IV. Challenges of Implementing ML for Network Assurance

Despite significant breakthroughs, the adoption of ML approaches for network assurance is currently hindered by a number of interconnected challenges. One frequently-encountered obstacle is associated with data. The availability of large, up-to-date datasets greatly improves the performance of ML models; at the same time, such information is often unfeasible or imprudent to disclose, as it may contain confidential user details or security vulnerabilities. Thus, practitioners must resort to non-disclosive methodologies that may produce incomplete or noisy logs. Moreover, creating, curating and labeling datasets is a laborious process that is difficult to automate, which severely compounds the complexity of many ML-assisted solutions.

Besides data availability, operational maintenance also poses important issues during the implementation of ML models. Major difficulties are related to how the decisions and recommendations provided by the algorithms are presented: in particular, when the models behave as black boxes, the lack of justification hinders the ability to understand and act on their outputs without resorting to human trial and error. Finally, more general challenges center around scalability and processing capabilities, since the operational requirements of certain models may easily become unfit for on-board enforcement, while monitoring the behavior of multi-tier infrastructure and systemic patterns at scale may exceed present-day capabilities.

A. Data Privacy Concerns

Machine-Learning (ML) techniques such as deep learning have gained wide popularity and practical use because of (i) their convenient training and testing procedures, (ii) their ability to learn complex representations directly from data, and (iii) their ability to learn features directly from raw data with minimal manual feature engineering. The growing real-time data collected by networks and information technologies has made ML readily applicable to network assurance. Model-based network assurance methods become too slow to cope with unpredictable scenarios such as cyber attacks and sometimes rely on data unavailable for such scenarios. By contrast, data analytic approaches leverage real-time data availability to differentiate between normal and abnormal traffic patterns and between benign and potentially harmful events. ML methodologies such as anomaly detection , cluster analysis , traffic classification and prediction analysis can be used across various network-assurance functions such as security, monitoring, planning and configuration maintenance. Specifically anomaly detection can identify irregularities in network traffic, log files, workloads, and users. Cluster analysis can group similar instances to detect anomalies or segment traffic and hosts for in-depth analysis. Traffic classification tags unknown traffic using characteristics, whereas prediction analysis assists in forecasting traffic matrix and link availability. The deployment of ML methods for network assurance significantly turns attention to data, requiring extensive data collection , preprocessing, and feature extraction to characterize network behaviors efficiently. However, data privacy poses a serious challenge. ML models leak private information about their training sets through predictions and parameters. Meeting regulatory frameworks such as the European Union's General Data Protection Regulation (GDPR) necessitates fair, transparent ML models and restricted access to training data. Achieving adequate accuracy without full access to the training data or models is a complex inverse problem. Considering privacy requirements from the start, adding privacy-preserving measures during and after training, and quantifying the models privacy risks are essential for regulatory compliance and widespread adoption.

B. Model Interpretability

Network assurance uses distinct techniques to address network security, maintain dependability , and ensure optimal performance to support vital resources such as workflows and applications. Machine Learning automates network assurance tasks such as anomaly detection, classification, and prediction. ML is valuable for automation because it identifies indicators of issues through pattern extraction without explicit scripting, thus providing more robust solutions than simply engineering rules from historical data. Model interpretability enables network operators to understand the rationale behind model outputs and thus facilitates greater adoption and enhancement of ML-based mechanisms. Many real-world settings demand that ML models not only possess high precision but also enable inspection of the reasoning behind their decision. ML interpretability techniques strive to explain model decisions to alleviate issues related to biased datasets, data drifts, and the opacity of complex interactions. Generally, interpretable models are categorized as either ante-hoc or post-hoc. Ante-hoc are inherently transparent while post-hoc analyses apply interpretation methods to opaque models. Network operators addressing assurance challenges often favour transparent models, while those solving sophisticated problems resort to opaque models for superior performance. Post-hoc interpretation techniques for unfamiliar input samples. Model-agnostic approaches comprise a category of post-hoc techniques that generate explanations independently of the underlying model. Such explanations elucidate the behavior of a model, including its biases, through visualization, which is particularly useful for complex systems in which the operational intricacies remain unknown. Model agnostic strategies outperform widespread post-hoc methods such as partial dependence plots by virtue of their applicability to all model types, a sophistication that better reflects fragile local model behaviors.

C. Scalability Issues

Scalability issues represent a significant challenge for network assurance due to the limited resources in network equipment that constrain the size of prediction models. Many automated recovery solutions have been proposed, but none consider the scalability of the prediction methods. Current knowledge of scalability comes mainly from big data techniques or platforms employed to overcome large data pools; however, these do not account for network equipment constraints. For example, large data volumes and high-speed links in optical transport networks already pose scalability problems to current assurance solutions. Fault management based on machine learning (ML) methods offers a promising approach for end-to-end network automation, yet scalability issues remain largely unaddressed.

V. CURRENT USE-CASES OF ML IN NETWORK ASSURANCE

Real-world implementations of machine learning demonstrate promising results, highlighting its value as a primary solution to network assurance challenges. Two specific case studies illustrate the effectiveness of ML in handling anomaly detection and predictive maintenance. Real-time traffic monitoring alongside active probing data further enhances the capacity to detect and localize issues.

Preliminary analysis underscores the importance of timely fault identification. When ML algorithms flag potential problems, operators can intervene before faults materialize—a capability that significantly curtails operational expenses and reduces customer disruptions. Leveraging an ML-aided network assurance framework within overarching control entity designs thus delivers both efficiency gains and elevated reliability. Let's discuss some of the current use cases which are implemented by some of the big network equipment manufacturers.

A. Baselining of Wireless and Wired entities for enabling Anomaly Detection.

Baselining applications in wireless networks allows for baselining of key performance indicators

Based on Cisco Systems' solution, some key performance indicators for wireless networks are deviations in radio-frequency, upper-layer metrics flagging potential malfunctioning devices impacting the user experience, onboarding latency, signal-to-noise ratio and application response time. The solution utilizes unsupervised learning to cluster normal network behaviors from telemetry data and deviations are computed by comparing new incoming values against these percentile based statistical models. Time-series forecasting uses exponential smoothing and regression ensembles to predict threshold crossings, with corrective suggestions integrated.

Another organization, Juniper Networks, uses unsupervised learning for anomaly detection by clustering network telemetry data to identify deviations from normal patterns, enabling proactive identification of issues like connectivity drops or unusual traffic spikes. Additionally, unsupervised methods are used in virtual Bluetooth Low Energy location services; Juniper's vBLE uses unsupervised Gaussian mixture models and density-based clustering to resolve locations within 1-3 meters without requiring fixed beaconing hardware.

Cisco Systems' assurance solution for wired networks uses ML and AI for detailed endpoint visibility. It employs supervised learning for multi-factor classification (MFC), using telemetry like DHCP fingerprints and DPI results to accurately label devices. For unknown devices, unsupervised clustering called "smart grouping" , which creates profiling rules for policy enforcement via Cisco ISE. The platform also uses ML-based anomaly detection to identify spoofing by flagging behavioral deviations from established baselines. Detection of spoofing attack may trigger the device to be put under quarantine.

B. Radio Resource management using ML Techniques

Both Cisco and Juniper have products that support a feature called radio resource management using a form of Reinforcement Learning. Both solutions have a daily learning cycle and reward functions related to packet loss, retransmission rate, and bandwidth consumption. Adjustments to the channels and other configurations are applied in real-time to adjust the radio resources . There are fallback mechanisms which use rule-based learning.

C. ML in Wide Area Networks

Machine learning techniques have the capability to develop use cases to forecast network failures and SLA violations before they occur. For failure prediction , models like gradient boosted trees incorporate features to identify patterns indicative of failures such as connectivity loss or degraded performance. These algorithms are trained on large datasets from millions of paths across MPLS, Internet, DSL , fiber , satellite, and 4G links , optimizing for high precision to minimize "false positives".

Cisco's Catalyst SD-WAN WAN Insights feature which applies similar ML/AI for analytics in Software defined wide area networks. Moreover, for bandwidth forecasting, statistical time-series forecasting models process historical usage data, including ingress and egress metrics aggregated daily, to predict future needs. These models incorporate seasonality and trend analysis, requiring weeks of historical data for generation and output lower, upper, and mean bandwidth levels for comparison with actual usage.

D. Using ML for Network Anomaly Detection and Outage Detection

Though still being developed as a solution where reinforcement learning is being used to predict network degradation and further predict outage detection.

This technique applies anomaly detection algorithms in tandem with RL learning methods and makes a prediction by correlating the anomalies and the previous internet outages to predict future degradation and outages.

E. Using ML/AI for Network Security

Network Security has been traditionally a solution based on static rule based enforcement starting from Firewalls, Intrusion Detection System, Data Encryption Management Systems and Identity management systems.

With innovations in ML organizations such as Cisco , Palo Alto Networks etc innovating products for solving some of the following problems

1. Data Breach Detection – by using both supervised and unsupervised learning methods and using classification methods we can distinguish between legitimate and unauthorized outbound traffic.
2. Ransomware identification often employs techniques like Convolutional Neural Networks to analyze behavioural patterns from file activity and network traffic. CNNs can detect malicious activities such as unauthorized encryption or command- and-control communication. This technique is used by Palo Alto Networks.
3. Phishing detection requires checking of URLs, email headers and the body of the email for deceptive patterns. Deployed systems utilize a broad spectrum of techniques, including traditional models (e.g., Naive Bayes, Decision Trees, SVM, KNN), ensemble methods (Random Forests, Gradient Boosting), and deep learning architectures (CNNs, RNNs, Transformers). Hybrid models and Generative Adversarial Networks (GANs) are also employed to improve accuracy and resilience.

VI. FUTURE USE CASES OF ML IN NETWORK ASSURANCE

Though we have seen a few use cases up to now in the current implementations or use of ML in network assurance, let's look at a few future applications of ML in network assurance.

A. Using ML for Root Cause Analysis – Even though we are currently able to derive anomalies from time series network telemetry data and can correlate these events with the outages, the real value will be the capability to do root cause analysis of the network infrastructure nodes. This requires identification of ML models which can do this. This also requires identification of new data sources which can be labeled for getting more context. We will need complex neural networks in combination with other mathematical models to develop a solution for this problem.

B. Using ML for Quality of Service and Traffic Engineering –

Using ML models and feeding the traffic patterns to these models we can predict the networks quality of service and traffic engineering requirements and apply the policies dynamically to control the traffic flow.

C. Using ML to adapt dynamic routing protocols-

Today routing protocols rely on static rules and management policy rules and also among different wide area and local area protocols today cannot interact with each other; using ML-based models will allow collecting stats and predict packet routing behaviors and change the routes in the networking nodes.

VII. CONCLUSION

Machine Learning techniques play a critical role in network assurance through automated analyses and intelligent decision-making. ML provides tools that enhance network status prediction , service monitoring, and anomaly detection. Before applying ML across tasks such as classification, regression and clustering , key considerations include understanding available algorithms, selecting appropriate training data, and choosing suitable network features. Different ML approaches can then be applied effectively to various assurance problems ranging from anomaly detection to predictive maintenance. As networks grow in size and complexity, ML solutions offer scalable assurance support that becomes increasingly important when manual processes are insufficient or unavailable.

Network assurance relies heavily on continuous monitoring, fault diagnosis, and fault prediction since the success of any system depends on its guaranteed service. Large telecommunication networks such as cellular networks and cloud infrastructure are particularly vulnerable to undesired events like node failures, misconfigurations, capacity overload, bugs, or cyberattacks. These events may lead to network outages if unaddressed, resulting in significant revenue loss for service providers and dissatisfied customers. Additional operational challenges include resource efficiency, security, customer quality of experience, and energy consumption. Proper network assurance can alleviate many of these concerns. Effective assurance enables supervisors to address issues before they escalate into failures and to locate affected components once failures occur.

REFERENCES:

[1] K. M. Sivalingam, "Applications of Artificial Intelligence, Machine Learning and related techniques for Computer Networking Systems," 2021. [PDF]

[2] L. Domingo Velasco Esteban and D. Rafique, "Fault management based on machine learning," 2019. [PDF]

[3] O. Ibitoye, R. Abou-Khamis, M. el Shehaby, A. Matrawy et al., "The Threat of Adversarial Attacks on Machine Learning in Network Security - A Survey," 2019. [PDF]

[4] M. Azmi Umer, K. Nazir Junejo, M. Taha Jilani, and A. P. Mathur, "Machine Learning for Intrusion Detection in Industrial Control Systems: Applications, Challenges, and Recommendations," 2022. [PDF]

[5] J. Jurandir Alves Esteves, A. Boubendir, F. Guillemin, and P. Sens, "On the Robustness of Controlled Deep Reinforcement Learning for Slice Placement," 2021. [PDF]

[6] N. H. Rotman, M. Schapira, and A. Tamar, "Online Safety Assurance for Deep Reinforcement Learning," 2020. [PDF]

[7] M. F., R. C., N. A., M. I. et al., "An Overview on Application of Machine Learning Techniques in Optical Networks," 2019. [PDF]
[8] J. Wang, D. Rossell, C. G. Cassandras, and I. Ch. Paschalidis, "Network anomaly detection: a survey and comparative analysis of stochastic and deterministic methods," 2013. [PDF]
[9] J. Zhang, F. Li, F. Ye, and H. Wu, "Autonomous Unknown-Application Filtering and Labeling for DL-based Traffic Classifier Update," 2020. [PDF]
[10] M. Shafiq, X. Yu, A. K. Bashir, H. N. Chaudhry et al., "A machine learning approach for feature selection traffic classification using security analysis," 2018. [PDF]
[11] K. N. Nguyen, A. Sehgal, Y. Zhu, J. Choi et al., "Towards Intelligent Network Management: Leveraging AI for Network Service Detection," 2023. [PDF]
[12] S. Maheshwari, S. Tiwari, S. Rai, and S. Vinayak Daman Pratap Singh, "Comprehensive Study Of Predictive Maintenance In Industries Using Classification Models And LSTM Model," 2024. [PDF]
[13] G. Heiler, T. Gadermaier, T. Haider, A. Hanbury et al., "Identifying the root cause of cable network problems with machine learning," 2022. [PDF]
[14] D. Rafique and L. Domingo Velasco Esteban, "Machine learning for network automation: Overview, architecture, and applications [invited tutorial]," 2018. [PDF]
[15] A. Palaios, C. L. Vielhaus, D. F. Külzer, C. Watermann et al., "Machine Learning for QoS Prediction in Vehicular Communication: Challenges and Solution Approaches," 2023. [PDF]
[16] F. Wilhelmi, M. Carrascosa, C. Cano, A. Jonsson et al., "Usage of Network Simulators in Machine-Learning-Assisted 5G/6G Networks," 2020. [PDF]
[17] S. Kumar Murakonda and R. Shokri, "ML Privacy Meter: Aiding Regulatory Compliance by Quantifying the Privacy Risks of Machine Learning," 2020. [PDF]
[18] S. Miriyala Reddy and S. Miriyala, "Security and Privacy Preserving Deep Learning," 2020. [PDF]
[19] W. Briguglio and S. Saad, "Interpreting Machine Learning Malware Detectors Which Leverage N-gram Analysis," 2020. [PDF]
[20] J. Mitros and B. Mac Namee, "A Categorisation of Post-hoc Explanations for Predictive Models," 2019. [PDF]
[21] M. Tulio Ribeiro, S. Singh, and C. Guestrin, "Model-Agnostic Interpretability of Machine Learning," 2016. [PDF]
[22] J. Couchet, E. Ferreira, D. Manrique, and A. Carrascal, "Anomaly detection using prior knowledge: application to TCP/IP traffic," 2012. [PDF]
[23] P. Casas, "Two Decades of AI4NETS-AI/ML for Data Networks: Challenges & Research Directions," 2020. [PDF]
[24] K. Sultan, H. Ali, and Z. Zhang, "Call Detail Records Driven Anomaly Detection and Traffic Prediction in Mobile Cellular Networks," 2018. [PDF]
[25] A. Padmanabha Iyer, I. Stoica, M. Chowdhury, and L. Erran Li, "Fast and Accurate Performance Analysis of LTE Radio Access Networks," 2016. [PDF]
[26] I. Y. Chen, E. Pierson, S. Rose, S. Joshi et al., "Ethical Machine Learning in Health Care," 2020. [PDF]
[27] M. Sicart, I. Shklovski, and M. Jones, "Can Machine Learning be Moral?," 2021. [PDF]
[28] J. Pombal, A. F. Cruz, J. Bravo, P. Saleiro et al., "Understanding Unfairness in Fraud Detection through Model and Data Bias Interactions," 2022. [PDF]
[29] M. Pawlicki, M. Choraś, R. Kozik, and W. Hołubowicz, "On the Impact of Network Data Balancing in Cybersecurity Applications," 2020. ncbi.nlm.nih.gov
[30] L. Risser, A. Picard, L. Hervier, and J. M. Loubes, "A survey of Identification and mitigation of Machine Learning algorithmic biases in Image Analysis," 2022. [PDF]
[31] JP Vasseur, "Beyond Protocols: Why AI is Networking's Overdue Paradigm Shift"