International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(23s) 2026 pp. 1771-1784
https://doi.org/10.70917/ijcisim-2026-5728
Article

Baselines and Anomalies in Network Test Data for Degradation Detection

Sujay Kanungo

Independent researcher, india
Email: sujay2002@gmail.com

Abstract: Network performance degradation may occur without anomalous events in individual network measurements. This paper presents a baseline-and-anomaly framework for detecting gradual degradation in network test data. The approach establishes probability-density and time-series baselines from historical measurements and quantifies observed degradation relative to expected behavior. It considers statistical, machine-learning, hybrid, and probabilistic methods, while accounting for temporal, spatial, seasonal, and diurnal variability. A state-machine representation is also used to characterize transitions from healthy operation through minor, moderate, and severe degradation to failure. Evaluation considerations include synthetic and real-world datasets, robustness to controlled perturbations, detection metrics, operational thresholding, interpretability, and deployment constraints. The discussion highlights data quality, privacy, transferability, and adaptation to evolving traffic patterns as continuing challenges.
Keywords: network degradation detection, anomaly detection, baseline modeling, network measurements, time-series analysis, machine learning, traffic variability

1. INTRODUCTION

Network performance degradation—prompted by an increase in traffic volume, changes in the number of flow sources, or the introduction of new applications—may not be accompanied by anomalous events in network measurements. To recognize such gradual degradations, operators require the ability to establish and maintain probability density function (PDF) baselines from test data automatically. Once a baseline PDF has been defined, the task of anomaly detection changes to quantifying a degradation signal observed in test data with respect to the baseline. When the model is employed, the concept of degradation also extends beyond performance and can be detected in other dimensions such as security. The proposed statistical methodologies have been assessed on laboratory test beds as well as through field deployments dealing with traffic from customer data.

2. FOUNDATIONS OF BASELINE MODELING IN NETWORK MEASUREMENTS

The measurement of quality and performance in modern telecommunication networks, such as Internet Protocol (IP) networks and cellular networks, is vital for Service Providers to continuously monitor Quality of Service (QoS) / Quality of Experience (QoE) metrics. Many Service Providers employ active measurement tools that actively generate and monitor dedicated traffic in order to continuously check, record, and apply analysis techniques to recorded metrics. Some commonly monitored metrics include One-way Delay (OWD), Delay Variation (DV), and Packet Loss Rate (PLR). These metrics have been shown to be good indicators of VoIP audio quality and are therefore extensively collected together with data such as Packet Inter-arrival Time (PIT) [1].

Active measurement tools such as OWAMP, TIAMO, and the Test Traffic Generator of D-ITG are capable of generating active test measurements on various network links. Multitudes of active test traffic and performance metrics have been widely published by several researchers, and extensive measurement studies provide valuable, publicly available insights into the quality of the Internet. These insights cover numerous parameters, including autonomous

Copyright: © 2026 by the authors

system, country, and Internet Service Providers (ISPs) [2]. Due to different traffic patterns, the Quality of Service levels between these metrics and other parameters can also exhibit variations and legitimate differences [3].

A. Definitions and scope

Network monitoring, diagnostics, and restoration are increasingly important in today's interconnected digital economy. The role of a robust network monitoring system that continuously evaluates key performance metrics such as bandwidth, throughput, and latency is crucial to maintaining service continuity and avoiding expensive outages. Performance degradation detection encompasses measuring actual network performance and comparing it with the acceptable performance baseline. Accurate knowledge of the baseline enables monitoring systems to effectively identify emerging degradation. The definitions of baseline and anomaly adopted in the degradation detection context clarify the scope of the approach developed.

Network measurements taken at individual probes and across network sites vary in nature and complexity. Despite this, there exist a number of broad categories that capture most of the commonly monitored performance attributes. Delay and loss measurements are perhaps the most traditional focus of such evaluations, particularly for Connection-Oriented Transport Layer services. Bandwidth is now a significant consideration for many communications applications, including file downloads and video streaming. The monitoring focus for Transmission Control Protocol is also being broadened to include Throughput and Round Trip Time, with the latter often being obtained via Transport Control Protocol timestamp options. Measurements of Active Queue Length provide insight into network congestion levels. Many of these measurements are influenced by the source and destination, traffic patterns, and temporal constraints. In addition to network performance degradation, there are also application performance indications such as web page fetch time and Domain Name Service lookup times.

Anomaly detection examines whether observed measurements significantly deviate from expected behaviour, or normal samples. Although degradation detection may instead be viewed as the detection of anomalies in the appropriate data time series, the terms "baseline" and "anomaly" specifically capture the intended monitoring goal. "Deviation" is avoided to prevent potential confusion with common signal processing use. Baseline and anomaly are therefore the only terms used throughout to describe the part of the data on which the monitoring operates. [3]

B.Data collection and preprocessing

Network measurements collected for testing purposes maintain varietly in volume and frequency, from high-resolution per-second captures at a single location over a limited period to low-resolution records collected across an entire production network. Regardless of the discrepancy, every sample can have its normal behaviour uncovered and reproduced, enabling detection of degradation metrics influencing user performance.

When the aim is to detect degradation rather simply characterise traffic, few processing actions are necessary. Outliers for defined maximum or minimum threshold limits should be removed. The remaining data are divided into training and test subsets, reserving a portion of the data for the application of the chosen anomaly detection algorithm. Baselines and express test-set variances measure the data's ability to separate anomalies easily. However, attention must be given to ensure that no unusual detected behaviour in the test set occurred during the test period. Repeating the process several times avoids bias in anomaly detection methods that react to different baseline parametres.

C. Baseline establishment methods

Statistical modeling has been widely applied to distinguish abnormal network behaviors from typical patterns of network traffic. [3] proposed multiple methods of baseline estimation which can be employed to predict network conditions. The approach relied on a general statistical model of time-series analysis proximal to ARMA, whose parameters vary in real time and are generated dynamically from training periods. Data-driven models aware of temporal dynamics were applied by to predict network traffic variables indicating the absence of anomalies or equipment malfunction.

Statistical techniques based on autoregressive modeling can provide temporal baseline estimates for time-correlated traffic variables. Specific methods include linear regression, least-squares fit, and Kalman filtering. Basic and enhanced Kalman models are popular methods to conduct temporal filtering of time-dependent measurements and extract baseline estimations. They assume both the system state and the measurement noise have Gaussian distributions, which can be hard to validate. Simple integrators provide lowcomplexity estimation of temporal trends with a single adjustable time constant. Even simpler techniques consist of averaging measurements for a fixed-size window, either centered or at all past time slots. These approaches do not properly consider the causality in the time series, filtering in one direction only. Various ARMA models have also been considered to perform time-series extrapolation of one or multiple variables simultaneously, assuming that the bases are known. Proximal extensions of ARMA models can continuously learn the parameters on realistic timescales and perform decay on historical observations.

3. CHARACTERIZING ANOMALIES IN TEST DATA

A proper characterization of anomalies provides the essential foundation for detecting performance degradations in telecommunication networks. Degradation indicators can be classified into the following three types: (1) measurable performance values deviating from the established baseline (normal range) and remaining outside it during pre-established time periods; (2) performance values returning to the normal range after exceeding it, indicating potential problems with yet unknown causes that may require solving; (3) sustained alterations to the normal pattern of one or multiple performance values, which could be constant changes, cyclical variations, or other different fluctuations [2].

Various anomaly detection paradigms have been applied to network monitoring. (1) Myopic detection considers only the most recent observations up to the current time observation, based on the premise that performance values within the past time interval can provide information that correctly estimates the normal behavior. (2) Non-myopic detection utilizes the entire available historical trace with the expectation that past behavior influences future performance across long time lags. Both myopic and non-myopic detection paradigms aim to classify the current time observation as normal or anomalous [1].

Technical and operational constraints often limit the extent to which degradation indicators can be observed from the specific test data available. The focus is therefore on distinguishing normal fluctuations of monitored performance values from credible degradation signals. The anomaly detection problem at hand can be reframed as separating the normal variance from the additional, less-frequent variance associated with degradation signals. Conventional baseline modeling is designed chiefly for estimating expected behavior and identifying short-lived transient phenomena. It typically employs single-variable statistical models, rendering many models unsuitable due to the cross-dependencies among multiple variables. Moreover, an out-of-support observation must not be marked as anomalous, yet many statistical models such as Gaussian or mixture-based models fail to fulfil this requirement. Existing works also typically ignore diurnal or seasonal patterns, yet the variation of many telecommunications performance metrics exhibits both diurnal and longer seasonal behavior.

A. Types of degradation indicators

Network management requires effective monitoring and analysis of performance metrics, such as throughput, packet loss, latency, and jitter. This monitoring can detect degradation indicators, such as increased latency or reduced throughput. Performance indicators are unpredictable and exhibit inherent variability. Undetected, these indicators can cause premature replacement of devices or inefficient maintenance operations. Anomalies in the evolution of these indicators indicate possible degradation and an increased risk of failure. Data-driven anomaly detection models are essential for automatic monitoring of test data issued by transmission devices [2]. These models help entities optimize operation and prevent failures.

Many solutions to the degradation-detection problem have been proposed, including statistical, supervised machine learning, and hybrid approaches. Most of these solutions focus on temporal data, neglecting the topology of

interconnected networks [4]. In the field, widely adopted approaches struggle to filter out normal variance without prior knowledge of the expected baseline. Such approaches may erroneously trigger alerts.

B. Anomaly detection paradigms

An anomaly detection approach focuses on identifying significant deviations from a norm; in the context of network test data used for degradation detection it is the task of spotting significant modifications in the behavior of normal communications [5] as represented by a statistical baseline. Commonly encountered forms of anomalies either indicate a failure in the monitored equipment (e.g. service interruptions, working configuration errors, material change in the monitored traffic) or signal the emergence of additional entities (e.g. new subscribers, addtional IP address subnets, introduction of overhead traffic) that change the global traffic pattern [1]. Baseline modeling coupled with anomaly detection permits flagging improperly behaving streams on a network element as the communications volume pattern of such streams is expected to deviate significantly from normal when a degradation occurs.

C. Separation of normal variance from degradation signals

Consider a scalar-valued measurement as a time series S(t), where t denotes time. The time series can be split into three additive components: (1) the baseline signal B(t) which reflects the desired traffic or service characteristics, (2) the degradation signal D(t) which indicates the physical or functional deterioration, and (3) the disturbance signal E(t) which refers to the inherent variability in the network measurements. The decomposition is then given by:

Upon careful examination of the degradation process, both the types of degradation indicators and the corresponding forms of observations can be classified. Throughout the degradation process, Baseline signals B(t) keep evolving, whereas disturbance signals E(t) remain either stationary or exhibit limited variations.

The disturbances also manifest themselves as white noise but are typically of a smaller amplitude compared to the degradation indicators. In recent years, various anomaly detection methods for time series data have been presented [6]. Such approaches aim to detect the presence of a departure from an expected fluctuation range, which can ultimately assist in the identification of degradation problems within the network. Although these methods are effective in detecting anomalies in univariate status data and can accordingly match the mentioned demands, they do not apply well to the degradation indicators in the context of common network measurements. A relaxed safety condition allowing for normal fluctuation within a prescribed ε-neighborhood of the moving baseline must be incorporated instead to facilitate an effective monitoring system.

4. EVALUATION FRAMEWORKS FOR DEGRADATION DETECTION

Anomaly indicators in network measurements can signal an impending fault, but also may simply denote short-term fluctuations in an otherwise stable system. A sound evaluation framework is essential for assessing how well a detection technique can separate an anomalous signal from background noise. This framework defines the types of degradation indicators to consider, the anomaly detection paradigms to employ, a suitable method for separating normal variance from a potential degradation signal, and the relevant metrics for quantifying detection performance.

The metrics of interest are typically binary: does the detection indicate an anomaly when one is present? Or is the detection clear of an anomaly when no issues have occurred? Decision thresholds are therefore required for any system that outputs a continuous detection score, presenting a further decision point that can be tuned to minimize false positives or missed detections, depending on operational priorities. Similar issues that complicate detection design also arise during evaluation. Ideally, the testing set should contain a representative mix of normal conditions and anomalies. Experiments are generally conducted on synthetic datasets to prove the core signal separation capability, before using real-world data from unattended deployed sensors for validation. Such datasets inevitably have a limited number of test anomalies, but they are often complemented by an extensive "no-anomaly" baseline.

A. Metrics and benchmarks

A comprehensive set of metrics provides a robust benchmark for validating different methods. Candidates for modeling normal network behavior are evaluated by assessing both the temporal stability of the estimated model parameters and the accuracy of the estimated baseline value. For modeling anomalies, an explicit event set is defined to allow testing by a supervised paradigm. To determine how well a model can isolate normal time-variant behavior from degradation events, a third, unsupervised, criterion quantifies the extent to which the remainders of tested models contain distinguishing signatures of detected degradation. Different classes of baseline models and anomaly indicators are investigated and recommendations for practical deployment are offered. The results show that baselines and anomalies in test data for degradation detection are best addressed separately but can be integrated into a single framework that accounts for these relationships when characterizing the performance of candidate approaches.

Degradation detection is essential for the effective management of networks and systems. Timely identification of problems relies on the accurate detection of anomalies, whether in the form of exceedances of a normal operating threshold or in the form of significant departures from normal levels. In many operational settings, including the monitoring of Internet quality, network test data is often used as it is easier to collect than operational measurements directly related to performance.

B. Experimental design for robustness

The specifications of an evaluation framework for the automatic detection of performance degradations in modeled data are described here. Robustness to alterations in test datasets is paramount in gauging the effectiveness of any detected deviations. The framework accounts for two dimensions of variation: the percent of values that fall outside the present baseline and the scale of conditioned perturbation. Test datasets are deliberately engineered to adhere to the targeted percent-exceedance to facilitate more intuitive evaluations and comparisons against the framework's perspective.

The assessment utilizes a framework developed to ascertain the performance of anomaly-detection methods applied to network traffic measurements [7]. At the nexus of modeling the normal behavior and detecting the abnormal behavior, the percent-exceedance of the prescribed number of specimens remains invariant to the anticipated degree of alteration in normality that accompanies the injection of anomalies into recorded datasets.

C. Validation with synthetic and real-world data

Robustness to different datasets is crucial for evaluating the capacity of monitoring approaches to identify performance regressions and configuration changes. A synthetic-testing simulation was developed for assessing test-data models and pre-detection baselines. This simulation generates sequences of synthetic time-series traffic, incorporating realistic patterns based on attributes extracted from actual time-series datasets. A library of synthetic models was compiled to capture diverse forms of anomalies, including both gradual and abrupt shifts. Seven real-world datasets evidencing different forms of baseline-variance anomalies were also selected for testing.

Two exhaustive baseline modelling and three automatic-detection methodologies were examined. When assessing real-world datasets, the pre-detection methods were executed at the monitoring location and the synthetic abnormalities were omitted. The network involved diverse operations and the datasets presented distinctive statistical characteristics. A combination of the statistical modelling and the machine-learning detection approaches yielded particularly notable results on these real-world cases [2].

5. STATE MACHINE MODEL FOR NETWORK HEALTH

In this section, we present a state machine model that characterizes the health status of a network as it transitions from a healthy state to a degraded state. This model includes several intermediate states to illustrate the gradual decline in network performance.

A. States

Healthy (S_0): The network is operating optimally without any issues.

Degraded – Minor (S_1): The network shows slight performance issues but is still functional.

Degraded – Moderate (S_2): The network experiences noticeable performance degradation, affecting some services.

Degraded – Severe (S_3): The network is significantly impaired, causing major disruptions to services.

Failure (S_4): The network is completely non-functional, requiring immediate attention.

B. Inputs

Performance Drop: Identified issues that initiate the transition from a healthier state to a degraded state.

Recovery Action: Actions taken to rectify minor or moderate degradation.

Complete Failure: A severe incident leading to total network failure.

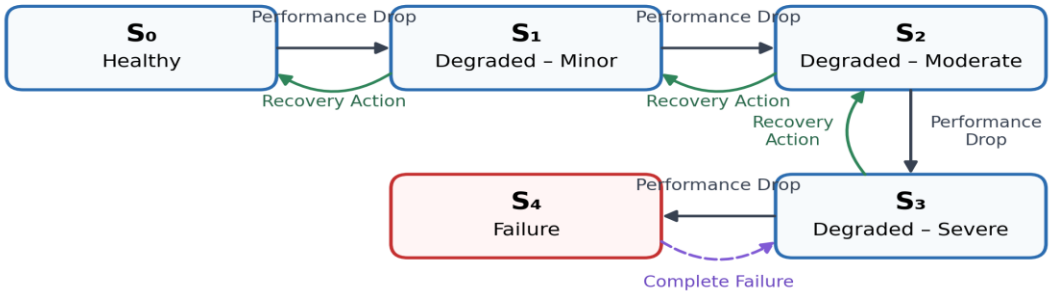

Fig. 1. Network health state-transition model.

C. Transitions

The transitions between these states are formally defined by the transition function T. The function uses the current state S and input I to determine the next state.

The transition function T(S, I) is defined by the following rules:

Table 1. Transition function T(S, I).

Here, S represents the current state and I represents the input.

D. State Update

To update the state of the network, we use the following equation:

$S_{new} = T(S_{old}, I)$ (1)

E. Explanation

This state machine model provides a structured approach to understanding the health status of a network. Through the defined states and transitions, we can analyze how performance drops affect the network and how recovery actions can potentially restore its health.

6. MODELING APPROACHES FOR BASELINES AND ANOMALIES

Degradation in telecommunication networks is a major concern for network operators. Automated detection of degradation is expected to minimize cumbersome troubleshooting tasks and to avoid an impact on subscribers. Test data from dedicated probes, such as packet loss measurements, is often available before degradation becomes critical,

allowing for timely corrective actions. Models for establishing baselines and for detecting anomalies form the core of degradation detection systems.

Network measurements exhibit a variety of temporal and spatial characteristics. Substantial changes in traffic patterns, due to events such as the release of new applications, pandemic lockdowns, and others, occur throughout the life of the network; these must be carefully considered by models for both baselines and anomalies. Seasonal effects, such as monthly or weekly cycles, are also present in many measurements. Robust characterization of non-stationarities is critical for reliable degradation detection.

Statistical models rely on classical definitions of independence across space and time. These models produce separate baselines for each equipment type—such as routers, switches, and links—each service type—such as Internet, VPN, and VoIP—and each measurement type—such as round-trip times, jitter, and packet loss. Measurements exhibit different baselines across days of the week, hours of the day, and network load levels [2].

Machine Learning (ML) techniques have recently emerged as new solutions to a broad spectrum of monitoring and surveillance problems. These techniques learn to detect anomalies and events from historical data without the need for establishing an analytical mathematical model of the system behaviour. They thus accommodate complex nonlinear and time-variant trends, taking into account disturbances from any source and allowing for multivariate analysis on multichannel data. However, traffic time series exhibit a disturbing combination of both gradual and sudden anomalies [5].

A. Statistical models for baseline estimation

Baseline measurements are essential for understanding traffic behaviour in a monitoring period. A commonly used statistical approach is to estimate baselines from historical monitoring data using distribution models such as Gaussian distributions for performance measurements like throughput and round-trip-time and time series forecasting methods such as Seasonal-Trend decomposition using LOESS to handle periodic measurements [6] [3].

B. Machine learning methods for anomaly detection

Anomaly detection aims to identify rare items, events, or observations that raise suspicions by differing significantly from the majority of the data, perceiving them as patterns that do not conform to expected behaviour (Iglewicz & Hoaglin, 1993). Estimating the underlying distribution of samples allows significant deviations to be detected, leading to the definition of anomalies. Characterisation of anomalies is uncertain and imprecise, identifying the need for the establishment of influential guidelines (Chandola et al., 2009). Anomalies fall into either point anomalies or contextual anomalies, which allow data points to be considered outliers only in specific contexts (Hodge & Austin, 2004).

Anomalies attract significant focus in network measurements. Malicious activities such as DDoS attacks create overloads in computer performance, causing noticeable degradation [8]. Natural events such as floods and storms can indiscriminately damage certain infrastructure and significantly disrupt services [9]. System failure, adding new nodes, breaking links, component malfunction, data loss, topology change, or unexpected behaviour all influence system variation and create potential sources of operational anomalies (Dhanya & Jayashree, 2024). The introduction of machine learning applications enables the separation of unwanted data from other information, preventing computational interference. Various machine learning models undertake the detection of anomalies to assist and automate detection.

C. Hybrid and probabilistic approaches

Specialized hybrid and probabilistic approaches, combining ideas from the previous sections, have also been proposed for the joint modeling of baselines and anomalies in time and traffic. [8] leverage a time-dependent hidden Markov model (TDHMM) to model normal traffic cycles and at the same time model anomalies that cause type 1 degradation events, such as packet delays or loss. At each time step, the state of the TDHMM specifies an expected

packet size and inter-arrival time, while noisy observations of the packet size and inter-arrival time are assumed to follow mixture-of-Gaussian distributions. Non-packet-level synthetic data from a multi-stage queuing system under constant-rate traffic is used to validate the TDHMM by demonstrating that the state transition pattern and degradation modeling capabilities match those of the underlying queuing model. Anomaly detection is based on likelihood ratios, leading to non-cyclic and off-cycle anomalies being detected simultaneously.

[3] observe that in communication networks, system-level event detection can be viewed as a low-rank change-point detection problem in the dynamic network time series (DNTS) framework. They formulate a probabilistic graphical model for characterizing DNTS measurements. In this model, the traffic matrix to communication links can be represented by a low-rank tensor, which can be also viewed as a highly sparsified form of the traffic matrix and hence fits well with description of network traffic. The change-point detection problem is introduced by modelling the dynamic traffic matrix as a piecewise-constant matrix, and the work aims to estimate both the change points and the corresponding background traffic matrices. In this framework, the detection of systemic anomalies corresponds to the detection of change points and is processed with an efficient algorithm in a streaming fashion.

7. HANDLING TEMPORAL AND SPATIAL VARIABILITY

Incorporating temporal and spatial variability is essential for reliable degradation detection. At a general level, anomaly detection relies on separating normal variance from degraded behavior. Although the duration of network tests is chosen to reduce the chance of encountering degradation, time-series methods adapt the normal-variance separation process to help isolate degraded behavior. The structure of network topology and traffic patterns, however, is constant in the test data.

Despite removal of diurnal and seasonal patterns, connections with extreme diagonal orientation in the topology potentially exhibit seasonal and diurnal deviation, particularly in WAN traffic. Sharp, significant, and confined seasonal-diurnal changes amplify the effect of the test measurement duration. Although gradually varying deviations may not introduce detectable degradation, they act against the desired normality for detection of more abrupt test degrades. Therefore, dead-band thresholds for times of no diurnal and seasonal diurnal pattern are typically recommended in alerting systems. Separated normal variance also reduces the danger of suppressed significant degradation, for example, a high-priority huge file transfer during normal-variance peaks.

The incorporation of temporal-correlation structure into normal-variance also brings potential seasonal-diurnal simplicity. Anomaly detection can be confined to seasonal-diurnal maxima-minima-only tests, with covariance indicated not only by the angles configured for normal-variance separation but also by the temporal-grouping structure of monitored boot-and-shoot connections, including that normal templates are constant in the season-diurnal period.

A. Time-series considerations

Network measurements possess inherent temporal and spatial variability. Time-series analyses enable the explicit incorporation of time as an independent variable, facilitating adaptation to this variability. However, physical constraints impose additional restrictions applicable to any set of traffic measurements. Each network node has a limited ingress capacity, and the total arrival count of packets or paths can only diminish. Consequently, temporal and spatial dependencies can be discarded once a measurement session concludes. These physical and operational constraints allow the use of a large spectrum of models and techniques found in regressions, time series, and supervised learning.

Degradation detection in contemporary communication systems involves measuring the performance of network components and analyzing traffic types. Markovian models enable estimation of the expected values of measurements and capture the degree of variability around the mean. Packet-switched networks generate, move, and drop packets as they traverse paths, resulting in observations on the edge in-between nodes. The arrival time of these packets is also recorded. Performance indicators focus on packet counts, delays, delay variances, and jitter determined by packet arrivals. Degradation consists of congestion impeding the nodes' ability to move packets or the duplication of packets along paths [6].

B. Network topology and traffic pattern effects

Network topology and traffic patterns significantly influence the factors determining baseline values and the characteristics of anomalies [10]. For instance, in a LAN environment, the baseline values and anomaly characteristics are markedly different from those observed in a WAN environment. In the former, traffic from workstation-to-workstation exchanges constitutes a large proportion of total network traffic; baseline values and anomalies reflect periods of active usage and inactivity at individual workstations, typically occurring during office hours, weekends, and vacation periods. In the latter case, a considerable proportion of traffic is directed toward external networks, and real-time or highly transient accesses to servers such as web, FTP, and Telnet generate bursty traffic. Under these conditions, the baseline values exhibit greater variation, while the anomalous characteristics taken into account for detecting access denial, unauthorized real-time on-demand video pattern change, and upload of large files, for example, diverge fundamentally from those in a LAN.

C. Seasonal and diurnal influences

When baseline or anomaly-estimation procedures are applied to longer time spans, anomalous behavior can occur that is not indicative of performance degradation. For example, experiences of a network may include seasonal, diurnal, or weekly-type patterns. With user verification, these issues may be ignored or an ensemble of detectors designed to act as a traffic model can be employed.

Such temporal patterns can have other implications in regard to the identified variances as well. For instance, traffic patterns can exhibit differing variances during the periods of day and week (e.g., the week-end versus the work days). The traffic volume itself can affect the performance and thus the anomalies detected. If readily identifiable, these differences can be captured by multiple baselines or deviations, monitoring both the degree of anomaly and the differential impact of distance and direction and allowing for more tailored alerts for specific users.

8. DEGRADATION DETECTION IN PRACTICE

Detection of traffic degradation on operational networks presents unique challenges compared to experimentation on test beds. In practical use, network operation involves multiple performance metrics, widely varying thresholds, very strict requirements, and little flexibility. A normalized version of performance metrics may therefore be required for monitoring. Fleets of diverse equipment types participate in network operations, leading to a wide variety of usage scenarios and performance guarantees across different Service Level Agreements. Consequently, broad anomalies such as traffic surges, detected on many links simultaneously, do not show clear degradation for individual performance metrics. Monitoring is thus routinely performed on a per-customer, per-site, or per-equipment-type basis. Measurements are typically compiled hourly, accumulated over both calendar and business hours, and back-validated against established base cases. Many companies adopt proprietary monitoring and remediation solutions. As a common practice, alarms appear upon exceeding minimum thresholds, often supplemented by neighbouring metrics, where neighbours can be defined topologically or in terms of aggregated activity [5].

A.Operational deployment considerations

Unlike traditional network performance countermeasures triggered by static threshold violations, systems for degradation detection must adjust to evolving baselines. For instance, a deterioration of test data commonly occurs paradoxically following a bandwidth boost. Detections may thus be aberrantly raised or depressed upon arbitrary contrast between values and static thresholds. These challenges arise within operational monitoring systems reliant on measuring the volume of active test packets traversing the network. Degradation indicators can result from suddenly blocking packets, restricted traffic courses, or a fallback from higher-order Web Real-Time Communication packets. The number of packets transmitted quite exceptionally follows the operational trajectory of the network, constrained by technological infrastructure or planned cycling.

Unrestricted test packet deployment can adhere closely to an unbinding envelope. The filling ratio from zero is seldom entirely drained, as spare capacity often persists after service adjustments. When higher-layer traffic alters,

network test packets on electrical paths exhibit temporal coherence, more so than those on multiplexed light paths. Preserving ordinary packet destinations further enhances cohesion. In practice, dynamic threshold schemes accompany detections of downgrading and up-gradation. For illustration, variation bands on a decaying trajectory conserves stability for periods exceeding substantive modifications.

A further practical hurdle resides in an interpretative gap from detection to requisite remedial measures and parameter tuning. Emerging awareness of the condition motivates operators to seek pertinent information about recently-seized deterioration detections within conventional monitoring frameworks devoted to query tracking rather than ticket opening [11].

B. Threshold selection and alerting strategies

Degradation detection usually requires monitoring of network measurements over time to identify anomalies or deviations from expected performance levels. Specifying the measurements to observe and the conditions that should trigger alerts is a prerequisite for any operational implementation of the detection methods described. The following subsections discuss practical suggestions for determining alerting thresholds and strategies that have been used to indicate the detection of degradations on monitored links.

Selecting appropriate alerting thresholds can be challenging, since the nature of the expected performance anomalies may not always be clear. Different traffic loads may be experienced when the measurements of interest are acquired, meaning that traffic deviations may not yet be evident, even if the measured situation has deteriorated. On the other hand, the initial load may have the same performance level as other degraded configurations, making it impossible to say if the situation has improved or worsened.

The presence of multiple, non-exclusive performance factors can complicate the assessment further. For example, a performance drop may first be observed under light loads before the classification of the impairment is known, while a warning of an incipient problem could then also be included for other values. Multiple devices acquiring information on different destinations and with independent, partially overlapped delays frequently exhibit a more complex collection of symptoms. Performance dropping on a link shared with other external points may simultaneously affect the evaluation of any or all of the network's devices and locations, and false signs of improvement could confuse the apparent links between the monitored performance indicators. For such situations, determining the relevant thresholds that entail a "degradation" indicator and hence trigger a degradation alert involves combining assessment from separate analytics to identify at least one degrading performance indicator, with other types of alerts indicated as future performance may drop or recover. [1]

C.Interpretability and explainability of detections

Machine learning approaches for anomaly detection often encourage human involvement by incorporating explanations into the model. Such explanatory frameworks have proven useful in other domains and can facilitate the investigation of degraded network performance. The goal of degradation detection is to identify changes with minimal human scrutiny. However, when formal investigations are necessary, relevant explanatory information should accompany all replies and answers.

Two complementary models can describe the underlying degradation detection mechanism. A deterministic outline indicates the precise thresholds defining normality and abnormality intervals in numerical measurements. For example, successful transmission attempts, previously characterized as a statistically constant test interval, are determined to be normal if they are above a given threshold, and abnormal otherwise. A probabilistic model quantifies the degree of normality of the measurements. Measurements with high normality likelihood receive normal scores, while those with a low likelihood return abnormal scores [12]. When applying the probabilistic scoring to post-processed residuals, the residual score serves as approximate measurements for degradation. Consequently, determination of the normative time series prior to residual analysis or degradation detection is required.

9. CASE STUDIES AND EMPIRICAL FINDINGS

Degradation detection on an optical transport network operating in a production environment is demonstrated using a passive network probe and a test-set-interval scheme based on the open-source framework [3]. Degradation detection is implemented on external jitter monitored on test data from diverse interfaces. These jitter traces have periodic baseline characteristics and exhibit degradations arising from failure of components or lines. For each test interval, the baseline and degradation anomaly are identified using the evaluation framework described alongside statistical modeling of the observed data [13].

Successes and ongoing challenges in degradation detection are discussed, along with practical guidance on how to address the particularities of an operator's network and data. Baseline modeling of external jitter collected over a decade is conducted on an operator's backbone network. The packet-trace-interval data set encompasses a wide-ranging set of routing paths including intra-, inter-, and pan-European edges, with approximately 10,000 samples per path falling within each cloud interval. Both periodic and non-periodic components are present and a two-part solution for degradation detection is proposed.

A. Laboratory benchmarks

Assessment of detection methods typically occurs with separately mined anomalies based on recorded traffic, aided by tagging or logging of modifications, and compared to selected video sequences; tools like Darshak or ViaGraph, MhD, and Smokey have facilitated independent anomaly analysis [14]. Traffic modeling within shared arenas remains an active research domain, as demonstrated by resolution of lingering debate juxtaposing transmission-centric and generation-point modeling frameworks and purposefully crafted design and measurement datasets—MGEN and ns include stochastic generation based on repetitions of pre-recorded files [15]. Insufficient frame rate and inadequate transport-level protocol sensitivity persist as motivating and revealing concerns, notwithstanding intrinsic measurement complexities or synthesized sequence freedom from modeling ordinariness [16].

B. Field deployments

Two customer deployments of the detection system to support degradation detection in enterprise access networks were conducted. The first deployment involved a proof-of-concept installation at one customer site. The traffic matrix included three different access points used by multiple devices and was collected over more than 180 hours. The second deployment incorporated lessons learned from the proof-of-concept installation. A more complex traffic matrix was used, containing eleven access points monitored at one-second granularity, capturing a wide variety of traffic patterns over several days [17]. A dedicated testbed was established, allowing unsupervised detection of anomalies in a fully online mode.

Another system deployed on a university campus was used to evaluate different anomaly-detection frameworks under varying conditions. The testbed consists of a single access point serving up to eleven clients of various types—personal computers, tablets, and smartphones—connected via wired and wireless links to the network core. By applying different configurations, multiple degradation scenarios were simulated, such as connectivity losses due to planned maintenance or the installation of improperly configured devices, so that the models could be assessed under both normal and degraded situations [18].

C. Lessons learned and best practices

Achieving successful operationalization of degradation detection in engineering networks demands careful consideration of various factors. The test subject, methodology, instrumentation, results, and means of dissemination must align with the expectations of both technical and non-technical audiences. Having participated in more than 20 such case studies covering both laboratory and production environments over a two-year period, several lessons have emerged regarding aspects of the detection technique most important for successful deployment [19].

Attention to engineering details enhances the likelihood of a successful deployment. Knowledge of operational needs is essential for guiding the selection of a relevant test subject. To improve applicability across multiple domains, the originally proposed approach operates on a defined set of KPI values. Nonetheless, further encapsulation of the

analysis area would reduce generality and therefore, applicability. When seeking feedback following test execution, summarizing results with visual elements and delving deeper into technical minutiae on a secondary layer promotes productive engagement with recipients possessing differing expertise levels. Recipients often retain higher-level concerns longer than they do lower-level interest in technical details. Subsequent to detection within a specific window, emphasizing subsequent behaviours and summarizing forward-looking experimental plans complements the initial exposition. Finally, a more consistent and wider lens on end-to-end traffic journeys would enhance the evaluation of deployed cases [5].

10. CHALLENGES AND FUTURE DIRECTIONS

Degradation detection presents challenges that remain relevant for future research. The availability and quality of network test data directly impacts the feasibility of degradation-acquisition methods [2]. In scenarios where the dataset is either limited or does not comply with specific quality standards (e.g., excessive noise), it becomes unfeasible to obtain representative baseline models or extract informative anomalies. Even when high-quality measurement series appear to be accessible, organizations may be restricted from exploiting the collected data due to privacy, confidentiality, or intellectual-property considerations. A further question is the degree of degradation-acquisition transferability. Even within a specific test case, network properties and the associated set of parameters can evolve. Consequently, acquired knowledge might only exhibit partial generalizability to new records. In a more general sense, the potential transferability of the established approaches between different test cases raises a fundamental concern. The evolution of traffic matrices is an additional challenge for future studies. Deterioration models usually assume that the underlying traffic matrix remains relatively stationary, enabling the characterization of performance changes over stable conditions. Yet, in many instances of practical relevance, the statistical nature of the traffic matrix can undergo significant changes. Consequently, performance alterations are evidenced by variations in the statistical properties of the measurements—affecting, for example, the variance, the baseline fit, or the frequency of degradation events. As a result, the definition of a prior baseline, alongside the detection of alterations, can become problematic—essentially reverting the challenges to the initial stages of the characterization framework.

A. Data quality and privacy implications

An effective degradation monitoring system requires the collection of measurements of a particular network performance metric over time to establish a baseline for normal behavior. Once an initial period of stable operation has passed, historical data can be analyzed to derive the baseline model and automatically segment the time series into testing periods. Since the network performance is impacted by many factors, establishing a baseline and detecting anomalies is a challenging task. Baseline modeling and anomaly detection can be performed in the time-domain, in the frequency-domain, and on a combination of the two. Statistical models estimate baselines and their variations and indicate whether a new measurement is anomalous or not based on the probability of occurrence through likelihood ratios. Machine learning models provide sophisticated frameworks for semi-automatic and automatic anomaly detection but require the availability of a representative labeled dataset containing both normal observations and anomalies, in many instances with binary or continuous labels indicating the degree of abnormality. When dealing with unlabeled datasets, one-class classification or novelty detection strategies can be explored. Characteristics such as non-stationarity, seasonality, trend, and global and local periodic patterns are taken into account in both approaches. Hybrid specifications combine complementary models to capture a wide range of complex degradation behaviors with different modeling assumptions. The models can be probabilistic in nature, allowing the incorporation of prior knowledge to improve the performance in small data scenarios.

Degradation detection and monitoring systems can be deployed in networks operated by multiple service providers. Data collection for degradation detection and monitoring may trigger concerns regarding data quality and privacy. Data quality concerns include whether the collected data, processing stages, and models influence the output results and how to handle such situations to avoid severe constraints on degradation detection and monitoring. Network operators typically use private and commercial probes for monitoring. Inputs from participating service providers should be aggregated at the traffic matrix level across different layers, and at the same time, privacy preservation

technologies such as differential privacy can be applied at the input, output, or both stages [17]. Beyond the introduction of new traffic patterns, observed data drift can occur due to changes in the traffic routing, user behavior, or external parameters. In such cases, updating the degradation detection and monitoring process either continuously or periodically is essential. Exploiting user signals, historical data, and prior knowledge about the network can further mitigate data scarcity and drift problems. Traffic distributions and temporal patterns may change when network expansions or reconfigurations take place. Continuous adaptation of the degradation detection and monitoring process is then required to provide consistent support from the beginning. Modifications in the traffic across different timescales may also originate from new services or the switching of widely-used applications [3].

B. Transferability across networks

Degradation detection approaches heavily depend on the equipment and network service under consideration. The spread of pervasive networking technologies, such as 5G, has resulted in the creation of a wide variety of dedicated communications services—e.g., vehicular ad hoc networks, smart grids, intelligent traffic systems, and the Internet of Things. Each service usually involves specialized measurement types based on the associated Quality of Service (QoS). Given therefore, that the different techniques address distinct topics, modelling techniques are often service-specific.

C. Adaptation to evolving traffic patterns

The methods and models presented for establishing baselines and detecting anomalies enable the definition of a network's traffic patterns and the identification of deviations from them. Due to their temporal nature, models of both baselines and anomalies should adapt to evolving traffic patterns. However, advances are still needed in realizing these adjustments automatically and immediately, without requiring retraining.

Many existing baseline and anomaly detection methods such as median, mean filters, and Holt-Winters, operate and adapt easily in real time. In contrast, other methods, such as those based on time series decomposition, which leverage the strength of different decomposition strategies and the capability to detect simultaneous anomalies in several components, generally need to be periodically trained off-line to allow adaptation. Nevertheless, training is crucial for understanding anomalies. It allows filtering and avoids confusing benign with malicious deviations, but even methods infeasible for real-time adaptation can be applied on previously trained models to yield insightful and relevant analyses.

Congestion at links, leading to packet web-stream degradation in VoIP or video applications constitutes the major capacity-requirement factor [1]. Traffic patterns depend on the carry capacity and utilization at links. Variation without trend in the proportion of remarkably more or less performed on VoIP homogeneous-call rates concerning intra-simulated-week or inter-school-weektraffic call rates relates, respectively, to the three basic data-traffic models identified [13]. Traffic patterns also evolve over time as new services emerge, indicating a need to improve adaptation schemes.

11. CONCLUSION

Network measurement data is characterized by significant and inherent variability, which complicates degradation detection based on historical baselines. Not all changes in measured variables signify an anomaly; yet subtle deviations may indicate a potential problem. Anomaly detection in this context aims to capture such degradation indicators while deliberately ignoring variations attributable to normal conditions. A two-stage approach, earlier in the literature [13] and empirically in practice, separates the modeling of baselines from the detection of deviations.

Despite an extensive academic foundation of anomaly detection that encompasses numerous methods, practical deployments frequently repurpose previously developed models towards a new domain (degradation detection) while continuing implementations based largely on classical, long-established statistical techniques. Addressing this latency between academia and operation is crucial for the effective transfer of research into everyday engineering.

TABLE I

TRANSITION FUNCTION T(S, I)

Current state (S)	**Input (I)**	**Next state T(S, I)**
S_0	Performance Drop	S_1
S_1	Performance Drop	S_2
S_2	Performance Drop	S_3
S_3	Performance Drop	S_4
S_1	Recovery Action	S_0
S_2	Recovery Action	S_1
S_3	Recovery Action	S_2
S_4	Complete Failure	S_3
S	Otherwise	S (unchanged)

REFERENCES

1. J. Jiang, "Detection of network anomalies and novel attacks in the internet via statistical network traffic separation and normality prediction," 2005.
2. K. Bordeau-Aubert, J. Whatley, S. Nadeau, T. Glatard et al., "Classification of Anomalies in Telecommunication Network KPI Time Series," 2023.
3. T. La Fond, J. Neville, and B. Gallagher, "Size-Consistent Statistics for Anomaly Detection in Dynamic Networks," 2016.
4. R. R. Zhou, N. Serban, and N. Gebraeel, "Degradation modeling applied to residual lifetime prediction using functional data analysis," 2011.
5. E. Mengoli, Z. Yao, and W. Wei, "Develop End-to-End Anomaly Detection System," 2024.
6. C. Ge, X. Chen, M. Wang, and J. Wang, "Deep Baseline Network for Time Series Modeling and Anomaly Detection," 2022.
7. B. Aliashrafi Jodat, S. Nejati, M. Sabetzadeh, and P. Saavedra, "Learning Non-robustness using Simulation-based Testing: a Network Traffic-shaping Case Study," 2022.
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.
9. S. Salem and S. Asoudeh, "A hybrid IndRNNLSTM approach for real-time anomaly detection in software-defined networks," 2024.
10. J. Wang, S. Jia, H. Zhao, J. Xu et al., "Internet Anomaly Detection based on Complex Network Path," 2017.
11. J. Antonio Trujillo, I. de-la-Bandera, J. Burgueño, D. Palacios et al., "Active Learning Methodology for Expert-Assisted Anomaly Detection in Mobile Communications," 2022. ncbi.nlm.nih.gov
12. S. Szymanowicz, J. Charles, and R. Cipolla, "X-MAN: Explaining multiple sources of anomalies in video," 2021.
13. F. Mata, P. Zuraniewski, M. Mandjes, and M. MELLIA, "Anomaly Detection in VoIP Traffic with Trends," 2012.
14. A. Wilmet, T. Viard, M. Latapy, and R. Lamarche-Perrin, "Degree-based Outliers Detection within IP Traffic Modelled as a Link Stream," 2018.
15. F. Chong Tat Chua, E. P. Lim, and B. A. Huberman, "Detecting Flow Anomalies in Distributed Systems," 2014.
16. T. Lukaseder, J. Fiedler, and F. Kargl, "Performance Evaluation in High-Speed Networks by the Example of Intrusion Detection," 2018.
17. R. Bourgerie and T. Zanouda, "Fault Detection in Telecom Networks using Bi-level Federated Graph Neural Networks," 2023.
18. A. Allahdadi and R. Morla, "Anomaly Detection and Modeling in 802.11 Wireless Networks," 2017.
19. H. Bodor, T. V. Hoang, and Z. Zhang, "Little Help Makes a Big Difference: Leveraging Active Learning to Improve Unsupervised Time Series Anomaly Detection," 2022.