Reliability Engineering and System Safety 193 (2020) 106606

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Quantitative approaches for optimization of user experience based on network resilience for wireless service provider networks

Deepak Kakadia[a,b], Dr. Jose Emmanuel Ramirez-Marquez[a,*]

[a] *School of Systems Engineering, Stevens Institute of Technology, Hoboken, NJ, United States*
[b] *Google, 1600 Amphitheatre Way, Mountain View, CA, United States*

ABSTRACT

Since the 1980′s and in particular 1996, telecom operators and recently mobile operators have been facing increasingly fierce competition, combined with flat subscriber growth and increased data usage resulting in tremendous downward pressures on profitability, forcing operators to differentiate themselves by trying to offer network services with better customer experience at lower operational costs. Wireless operators are challenged with measuring user experience which in itself is subjective, in a manner that accurately reflects the functional and emotional aspects of perceived quality and linking to Network Resiliency which characterizes the network behavior as it responds to disruptions. Current network faults and alarms only consider device failures and do not consider actual impact to user experience. For instance a failed router may not impact the users experience due to built in redundancies in the network. Studies to date, have proposed methods and models that focus on specific aspects of user experience in wired and cellular networks. However, to the best of our knowledge, there is currently very little research that connects linking poor user network experience to root cause. Previous recent work in this area focus on identifying what and where measurements to gage subscriber OoE, modeling and high level concepts, but do not address realistic challenges and approaches that can be automated to materially impact improved customer experiences at lower operational expenses. *There is a gap on how operators can automatically associate poor user experience, relevant network metrics and root causes with a suitable model that can be analyzed and optimized.* We propose a general framework for a solution that links these entities together, with a quantified approach to optimize user network experience by optimizing network resilience using a model that can be analyzed and optimized using machine learning methods to improve resilience and hence user experience. Results of directly applying existing machine learning algorithms for identifying root causes to network telemetry data have proven to be ineffective in practice due to the fact that existing machine learning algorithms are designed for prediction, classification and ranking not for identifying causal relationships and further complicated by the fact that these algorithms have assumptions on the data and in reality the network data distributions vary wildly during network disturbances. The proposed general framework combines existing methods for anomaly detection and machine learning algorithms, however the novel contribution centers on improving the accuracy of finding associated root causes by dynamically selecting the optimal machine learning algorithm based on the network telemetry data features that are recomputed before, during and after network disturbances. The proposed approach then allows us to automate the time consuming manual tasks of network engineers that proactively monitor key performance metrics for anomalies, correlate with other data sources to ultimately determine actionable insights to maintain a certain acceptable level of user experience by dynamically selecting the appropriate machine learning algorithm for the given data characteristics or features. We describe an example case study specific to wireless provider environment, illustrating the potential viability with results from actual wireless(approx 8 million monthly subscribers) operations data showing promising results by applying the proposed approach. The prototype implementation was able to programmatically detect anomalies, identify potential root causes using different algorithms suitable for the given data and time frame, which dramatically increased the accuracy and efficiency of the small network engineering team, and hence improved the user experience by improving network resiliency.

1. Introduction

Telecom deregulation starting from the 1980s to 1996 in the United States and then in other nations has benefited consumers but devastated incumbent, slow to adapt, operators due to increased competition and lower Average Revenue Per User(ARPU). Wireless Network Service Providers which have emerged from telecom operators, scramble to differentiate themselves by providing better user experience, while being under tremendous pressure to deliver more data volume, at faster rates, and above all at lower costs than their competitors. Network Operations Costs consume a significant portion of the overall network operating budget, making it even more challenging to provide high quality user experience. Looking closer at the underlying reasons of high operational costs, are experienced network engineers, that follow manual, time consuming processes to resolve network issues often, without knowing if the particular issue directly impacts user experience or not. Current technologies in Network Operations Tools are silo solutions that do not integrate well, focus on a single or few aspects of the network and require high skilled, expensive network engineers to correlate and resolve issues after there has been a reported user complaint.

* Corresponding author.
E-mail addresses: dkakadia@gmail.com, dkakadia@google.com (D. Kakadia), jmarquez@stevens.edu (D.J.E. Ramirez-Marquez).

https://doi.org/10.1016/j.ress.2019.106606
Received 15 January 2019; Received in revised form 22 July 2019; Accepted 9 August 2019
Available online 30 August 2019
0951-8320/

Prior research has described user experience from a single aspect rather than a network wide perspective. For instance, many studies concentrate on the optimization of only one application such as video and one network segment such as the Radio Access Network(RAN) and adjusting parameters without considering adverse impacts to other parts of the network or user experience. By not knowing how this benefits the user experience of all users is of limited value to operators.

Second, prior research has discussed extensively on how to optimize the availability of the network at various layers from the optical fiber, DWDM, IP, MPLS and other availability mechanisms. The operators are now finding that some parts of the network are over engineered at high cost with little incremental benefit to users experience and other parts of the network that are single points of failure and under engineered which impact user experience significantly. By not aligning the network availability with user experience with a validated model is something operators are no longer able to afford. Third, recent research have proposed several approaches to measure subscriber QoE at an application aggregate level or 3GPP interface level, which will not capture specific subscriber flows, and conceptual frameworks to detect subscriber QoE issues and a QoE manager but do not address how the operator will realistically resolve the offending network issues without expensive manual, time consuming efforts.

For these reasons network operators are struggling to find innovative ways to maximize user experience and simultaneously minimize operational costs in order to differentiate themselves and win in a competitive marketplace. There is a gap that links User Experience to Network Metrics to Root cause that can be automated to reduce manual efforts and costs which improve the resilience of the network and ultimately user experience.

In this paper we propose a general framework that detects anomalies and associated root causes by dynamically selecting the most suitable machine learning approach indexed by the network data characteristics or features computed in time windows before, during and after network disturbances. We found network telemetry data characteristics or features such as linearity, stationarity, normality and latent variables vary widely between steady state and episodes of network disturbances. We further propose a method on how to connect the user experience(i.e. deviations from expected levels of service by detecting anomalies of the service level) to direct actionable network root causes(determined by dynamically selecting the most suitable machine learning algorithm) by defining quantified models for user experience and network resiliency and a model that connects them, which can then be optimized and automated using machine learning methods. The key contribution of this paper is that we propose a method that can be automated to materially reduce the time and effort it takes the operator to take corrective action in the network in direct response to detecting drops in user experience using dynamically optimal selected Machine Learning algorithms based on the network data characteristics. Network telemetry data is wildy dynamic and machine learning algorithms vary in assumption of the data. By characteristing the network data dynamically we can select the most appropriate machine learning algorithm, hence yielding most accurate results. *We introduce connecting user experience with network resiliency by defining Key Performance Indicators(KPI)s that abstract metrics of the network that directly impact the user experience*. Once we have these metrics, we can construct an optimized model and dynamic selection of the most suitable machine learning algorithm for the given data in the processing window at that time, which can be automated that has the most impact amongst all alarms that are causing a drop in user experience, hence we automatically correlate and prioritize alarms as well as determine root cause. We build on the definition of "*System Resilience*" [1] which discussed the quantification of the performance of a system before and after a disruption, as "*Network Resiliency*" being defined as the capacity of the network to maintain adequate levels of service to its users by withstanding or recovering promptly to network disturbances that materially impact user experience as quantified by key network performance indicators with target service level objectives(SLO). The KPIs are defined as metrics that abstract key aspects of the network as service, which is composed of subservices. The deeper we can define these subservices, the more detailed we can identify root causes. For instance, in order for a user to access and use a particular network, they must first be authenticated by the network onboarding process and then access the internet by traversing the transport network of the provider which is, itself composed of several segments spanning user to internet. The availability aspect of the network will be composed of metrics pertaining to the network being up and working at all layers required to transport user packets. The performance of the network will be composed of metrics pertaining to the degree of congestion at various locations in the network. The research then describes models that link user experience with network resilience and the optimization function.

The remainder of this paper is organized as follows: In Section 2 we review prior art in the area of user experience in wireless operator networks, network resiliency and machine learning which we will use to automate the identification of network disturbances and root causes. In Section 3 we formally define the Network User Experience Model, including the important KPIs that link and quantify the user experience to network measurements. In Section 4 we build on Section 3 by formally defining the Network Causal Model, that organizes and links the most relevant network measurements to user experience but whose structure can be exploited to identify root cause in an automated manner. In Section 5 we formally define the Network Resilience Model which allows us to connect user experience, to relevant network measurements in such a way we can now see what needs to be optimized. In Section 6 we tie everything together to formally define the Network User Experience and Resilience Optimization function that considers all aspects of the network and user experience. Section 7 describes results from an example prototype implementation of the proposed approach that show improved user experience and improved network resilience by automating the detection and dynamically selecting the optimal machine learning algorithm based on the network data characteristics. Finally we conclude on the findings of our proposed approach including limitations which require further research.

2. Related work

At time of writing, there is no integrated published work on user experience, reliability and remediation. Prior art in wireless network user experience [2–5,15,17,39] centered on single aspects, significant predictors, metrics but did not discuss corrective action on remediation. Sun et al. [31–36] discuss network resilience in terms of network protocols, or domain specific aspects such as SDN, IP, Optical, DWDM, multi layers, Telecom, BGP. Operators are now concerned by moving away from silo approaches to network resiliency, that some parts of the network may be over engineered and other parts under engineered as these approaches are not tied directly to user experience. Marnerides et al. [37] however discusses resilience in terms of users, applications, network and systems but in terms of defense against malware. Bai et al. [38] state that the resiliency of a network is correlated to recovery time, however they do not discuss how to optimize the resilience function. Typically the operator requires manual expertise to find root cause of degraded user experience. Prior art in Network Reliability [6] only focussed on devices, no discussion on how reliability is correlated to user experience. Anomaly Detection [13] is concentrated identifying outliers which don't distinguish features such as busy hour times or congested cities vs rural areas. Resiliency Frameworks [7–10] lays the foundation of this paper where we extend to include user experience and corrective action to optimize user experience. Recent work [46–49] propose methods to connect subscriber user experience to network corrective action but, do not address network impairment root causes that result in operator actions to materially improve customer experience. Ahmad et al. [46] describes the need for a standard interface between ISPs and OTT, Liotou et al. [47] describes a theoretical QoE monitor and QoE Manager that upon detection of QoE, corrective action is taken without any guidance on how the operator will actually determine root cause in order to take corrective action. Baraković and Skorin-Kapov [48] is limited to describing where to instrument data collection in a service provider network, Aggarwal et al. [49] applies machine learning techniques to train a model to correlate bad user experience to associated network measurements but correlation does not imply causation, so the operator will still need to manually determine root cause. Prior art in Causality [11–14] describes algorithms and approaches to find Root Cause and anomaly detection, based on observational data and often requires interventions which is not practical in a live production network. There is no integrated approach known to date, that ties all these concepts together end to end, from user experience to root cause and optimizing network resiliency. Bao et al. [16] describes an approach that uses logistic regression to identify connect network parameters to network bandwidth, based on the presumption that bandwidth alone causes Mean Opinion Score or notion of user experience and claim by tuning these parameters user experience can be improved. The notion of correlation or association of features to response variable is different from causation, which may be in fact due to confounders not considered and application neutral analysis is overlooked. Bao et al. [18] discusses the causal relationship is claimed, not based on the core logistic regression algorithm but indirectly by verifying user experience has improved or not after treatment effect. The limitation of this approach is that by the time parameters have been changed and measurements on the effect commence, network conditions may have changed that contaminate the study. The causal relationships should be determined from the outset, only then are we sure that modification of the causal variables do in fact directly impact the effect. The notion of user experience should consider not only network bandwidth but other key quantifiable measures such as onboarding and network availability which far outweigh network performance in many cases.

3. Network user experience model

Intuitively, it is clear that the network user experience is directly tied to the network as shown in Fig. 1. Traffic from the User Device on the far left must traverse the Radio Access Network(RAN) then the backhaul, metro, core and then to the Internet via Peering Router onwards. It is not obvious, however how to quantify or optimize the network user experience. We propose a quantified model that captures the end to end network user experience in terms of network resilience and an approach to optimize this model.

3.1. Definitions

- *Service Level Objective(SLO)* - Target Level of promised service that the service provider must meet or exceed. The Units depend on the service under consideration. Availability is expressed as a percentage such as 99.99%, latency is expressed in units of milliseconds or seconds. The definition of the SLOs quantify user experience. If the delivered service meets or exceeds SLO targets, then the customer experience is labelled as good.
- *Service Level Indicator(SLI)* — The actual measured value representing the level of service actually delivered. This measurement is usually compared against promised target SLOs to determine if the service is compliant or not.
- *Resilience* - defined as the point in time when the system is not providing a level of service as promised(SLO) to the point in time the system recovered and is meeting or exceeded the promised level of service. We extend the definition of resilience to include metrics: Onboarding, Network Availability, Network Performance(*Latency, Bandwidth, Packet Loss Rate etc.*). In particular, resilience is defined as the time taken to recover from a noncompliant state to compliant state

The network user experience can be modelled and estimated by the following 3 wireless network service Key Performance Indicators:

1 *Availability* - How often is the network service available and working.
2 *Network performance*- How well does the network transport perform. The network transport can be characterized by the composite metrics:
   a Latency - Average end to end network delay experienced by users.
   b Bandwidth - Average network bandwidth provided to user traffic.
   c Packet Loss Rates - Average end to end packet drops experience by users.
3 *Onboarding* - How well does the network service perform the onboarding process. The network onboarding process can be characterized by the following composite metrics:
   a AP Association - How long does it take for the User Device to associate with the WiFi Access Point(AP)
   b DHCP latency - How long does it take for user device to obtain IP address and other Dynamic Host Control Protocol configuration information from the network
   c DNS - How long does it take for the network to resolve a canonical host name to a network IP address.

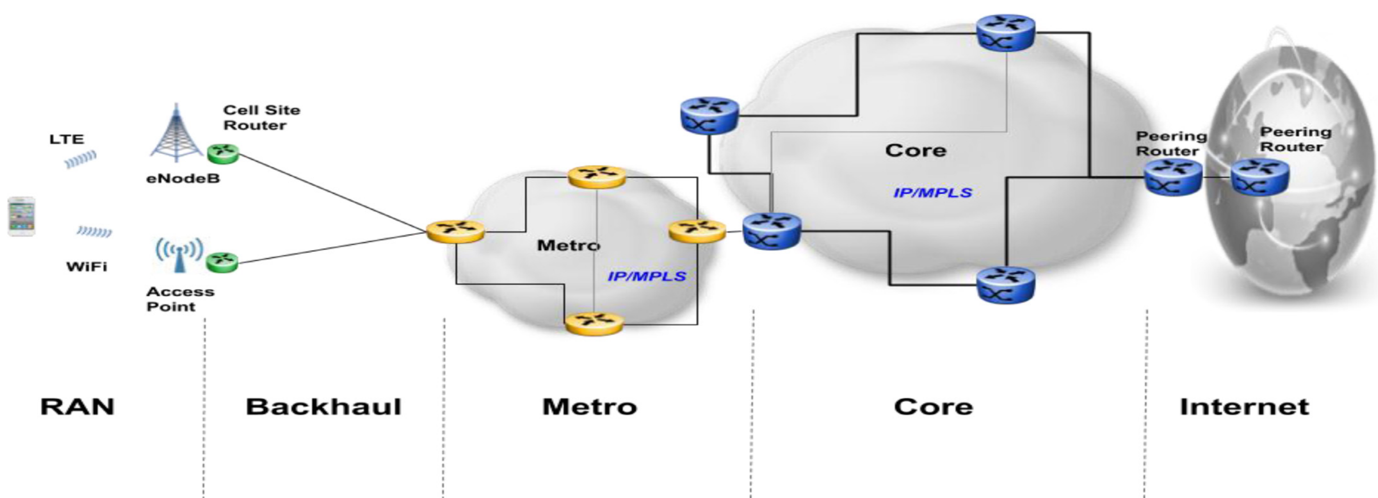

Fig. 1. Network user experience is directly correlated to network resilience of network and segments: Radio Access Network(RAN), backhaul, metro, core and internet. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

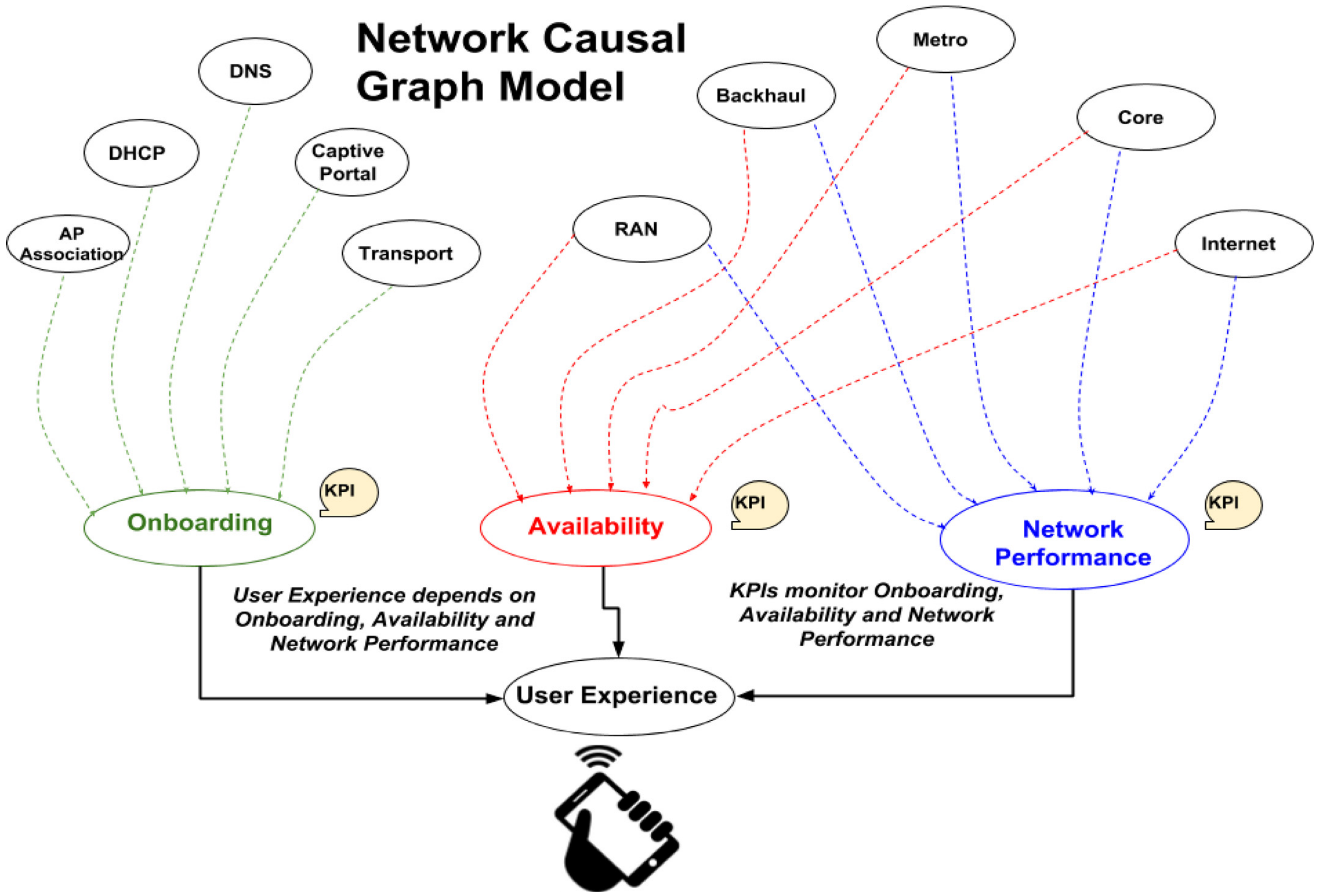

Fig. 2. Network causal model for each metric: availability, network performance and onboarding and causal structural relationships between user experience and root causes of non compliant metric performance. Fig. 3 below describes example metric measurements of Availability and the importance of time scales and measurement points. Station B and Station A are being measured individually for availability, such as synthetic pings for reachability, and we see that with granular measurements, we can identify exactly when a particular location is not performing as promised by meeting the target SLO, in this case 99% availability. Station A is chronically below SLO whereas Station B momentarily falls below the 99% threshold.

d Captive Portal - How long does it take for the network to authenticate a user in order to gain internet access to the network.
e Transport - Latencies in the network during the onboarding process

Fig. 2 below shows the network user experience model and how user experience is composed of a composite score that capture the main components of the network service that impact user experience based on Network Key Performance Indicators: Onboarding, Availability and Network Performance. Availability is based on the network transport being up and functioning, from the physical layer to IP layer. Network Performance is based on the capacity and load of the network. In cases where the load exceeds the capacity of the network, there is congestion, the network will be available, but traffic will be delayed due to buffering up to a point, beyond that packet will be dropped if queue buffers are filled beyond capacity. The user experience depends on onboarding, availability and network performance, where each component will be used to compute a composite score that enables the quantification of the overall user experience.

4. Network causal model

The diagram in Fig. 1 described a typical wireless service provider network architecture which consists of separate segments that can be owned and managed by the same or other partner service provider companies. User Data sends and receives data from a server, the network transports this data end to end, for an example application - youtube where data from youtube servers in the Internet are streamed towards a user traversing Core, Metro, Backhaul then finally to the Radio Access Network(RAN) segment to User. The service provider sees thousands of network disturbances in the network each day, but does not know which ones materially impact user experience, further which segment in the network is causing this disturbance. Identifying disturbances that materially impact user experience and root cause is time and resource consuming for the operator. Most operators typically focus on device outage as highest priority without any consideration to factors such as the network having redundant routers that can easily absorb all rerouted traffic, hence negligible impact on user experience. On the other hand, user impacting events may get ignored, for instance there may be an AP that is intermittently congested, impacting hundreds of users' experiences throughout the day each day. Our approach will address this gap.

Fig. 2 below describes an example Network Causal Graph(NCG) model that captures the structural relationships between user experience and network characteristics: Onboarding, Availability and Network Performance. These characteristics directly impact the user experience from a network operators perspective. Each one of these characteristics can be broken down further into subcomponents which are shown to have a causal relationship. For instance, for wifi users initially trying to access the network, there is an onboarding process, that first starts with the user device connecting with the wifi access point, which is then followed by a DHCP process to be allocated a

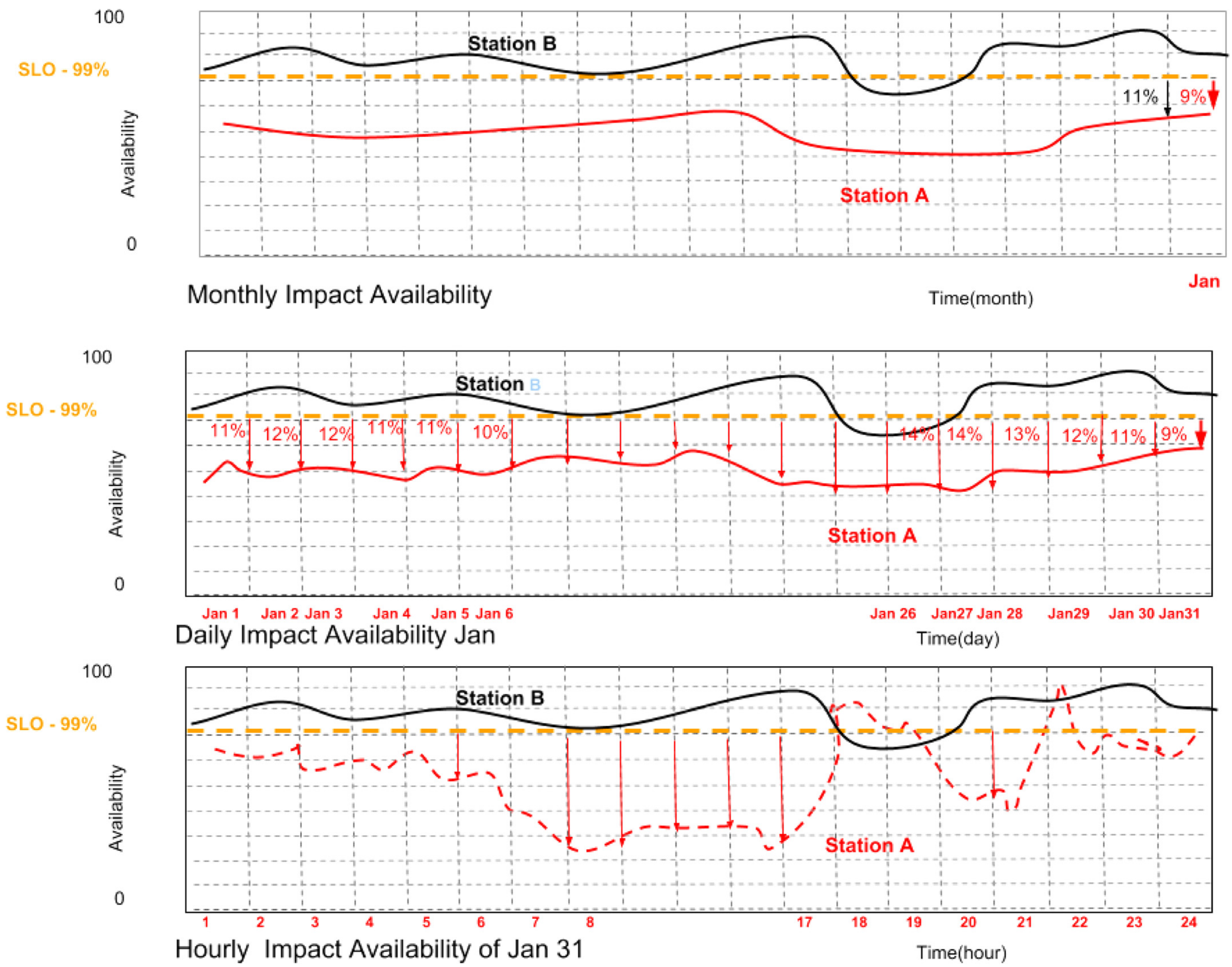

Fig. 3. Availability measurements: actual vs SLO target on a per location basis allows for root cause identification of components that are impacting user experience.

temporary IP address and then DNS resolution to resolve the IP address of the captive portal where the user must log in and get authenticated. This 5 onboarding sub processes can be abstracted by an onboarding Key Performance Indicator that simply measures the success rate and time to users to start the AP association to the final Captive Portal grant after successful authentication. The KPI can detect poor onboarding experience when the onboarding KPI sees a high failure rate or delayed onboarding time to drill down which subcomponent is the root cause. The transport abstracts all onboarding interconnection from IP layer and below. Similarly Availability and Network Performance have associated KPIs that measure how well the network is performance in terms of availability or speed and associated causal segments of the network are modeled to be able to identify root causes for non-compliance. The structural component relationships for each metric which will be used to identify root cause when a particular Infrastructure sub Service(*Onboarding, Availability, Network Performance*) is out of compliance. The availability NCG can be cross referenced with Fig. 1 where each segment directly impacts the transportation of user packets. In this paper, we will show, if a service is unavailable, we will use machine learning to automate the workflow and processes to identify the root cause, which the operator will then know what repairs are required in order to stabilize the system, saving costly resource intensive manual detection and diagnostic phases. The small nodes in blue prefixed by the letter "U" denote unknown, or noise.

Wireless Operators are struggling with finding methods to predict, identify, and diagnose failures in the network, in an effort to restore services to subscribers. Rafique et al. [40–44] describe various approaches to solving this problem, including using granger causality, optical hierarchical methods, Bayesian, and others [45] describes an approach to estimate root cause of faults by manually constructing a model using managed objects to represent the physical network elements, with relationships between layers of the network, organized in hierarchies. The root cause is determined by invoking methods in the managed objects using the pre programmed relationships. This approach cannot scale on large networks where new objects are being added and removed, and more importantly it is not practical to programmatically capture all rules for all possible scenarios. Each approach has different characteristics, complexities and time to recovery. For the purposes of this paper, the details of the optimal algorithm is beyond the scope of this paper however, we will use the fact that there are a finite set of different approaches to formulate our model for optimization.

5. Model: user experience and network resilience

We now extend Resilience Frameworks [7–10] to [19–30] model user experience and network resilience which will be correlated with the Network Causal Model.

5.1. Model definitions

R(t) - Resilience of system at time t- ratio of recovery at time t to loss suffered by system at some previous point in time t_d
R(t) = Recovery(t) / Loss(t_d)

Value of $\mathbf{R}_F(t_r|e_j)$ corresponds to a specific Figure of Merit $\mathbf{F}(t_r|e_j)$ evaluated at time $t_r$where $t_r \in (t_d, t_f)$ under disruptive event e_j is computed as:

$$\mathbf{R}_F(t_r|e_j) = \frac{\mathbf{F}(t_r|e_j) - \mathbf{F}(t_d|e_j)}{\mathbf{F}(t_{0j}) - \mathbf{F}(t_d|e_j)},\ e_j \in D \quad (1)$$

From Dessavre and Ramirez-Marquez [10] we have the following definitions:

System S is subjected to a disruptive event e_j D is the set of all Disruptive events that materially impact the network such that user experience is impacted. The total elapsed time from start to recovery = τ with the following breakdown:

1 **Reliable time frame** τ_1 (S, e_j) = [t0, te).
2 **Vulnerable time frame** τ_2 (S, e_j) = [te, td).
3 **Disrupted time frame** τ_3 (S, e_j) = [td, ts).
4 **Recovery time frame** τ_4 (S, e_j) = [ts, tf).

From Fig. 4, we see that the net effect of the Optimized System as compared to the Original System is the difference between the downtimes between systems $\tau_{original} - \tau_{optimized}$. We formalize the System Downtime time difference on each subcomponent timeframe as enumerated above as follows:

$$\tau_{difference} \sum_i (\tau_i(S, e_j) - \tau_i(S_c, e_j))\ \text{for}\ i \in \{2, 3, 4\},\ e_j \in D, S_c \in \{S_{c1}, S_{c2}, ..S_{cn}\} \quad (2)$$

, where $\tau_{difference}$ denotes difference in overall system downtime using a particular Causal System S_c, which is an element from the set of all available causal systems {S_{c1}, S_{c2},.. S_{cn}} compared to the original system S overall system downtime. Again, S_c denotes a system equipped with proactive monitoring and root cause analysis functions for all events e_j . Note we only compute time frames τ_2, τ_3, τ_4, that fall in the time frames when the system is down S_f and S_r. If the value of $\tau_{difference}$ negative then we are better off with the original system. This value must be positive in order to improve the system resiliency as a result of proactive monitoring and causal inference that reduces system downtime.

Fig. 4 The effect of Causal system with reduced overall system downtime as compared with the original system by reducing individual component times:

t_0 - *system stable time*
t_e - *time of disruptive event*
t_d - *time system in disruptive state*

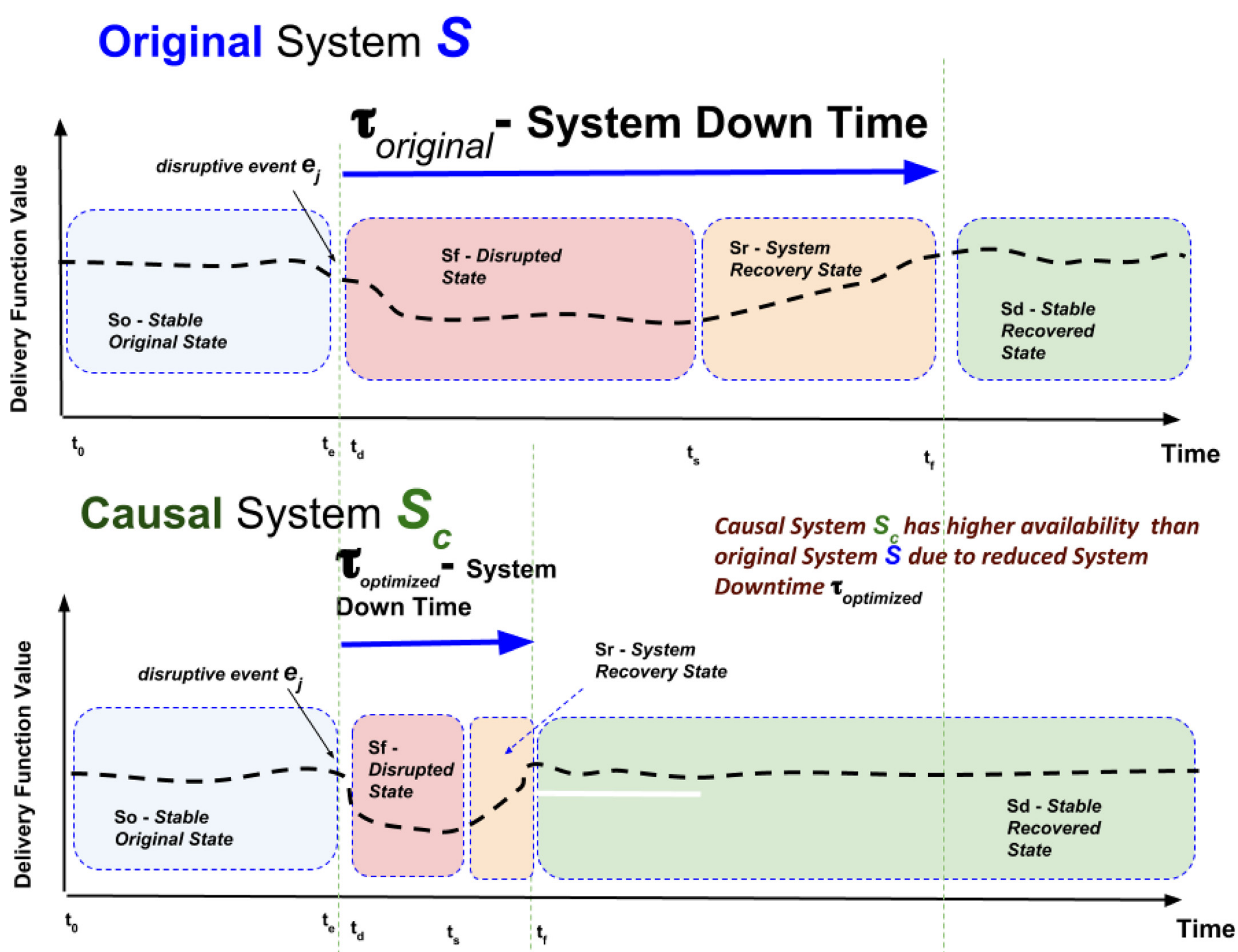

Fig. 4. Below shows the effect of optimizing the original system as described in [7–10].

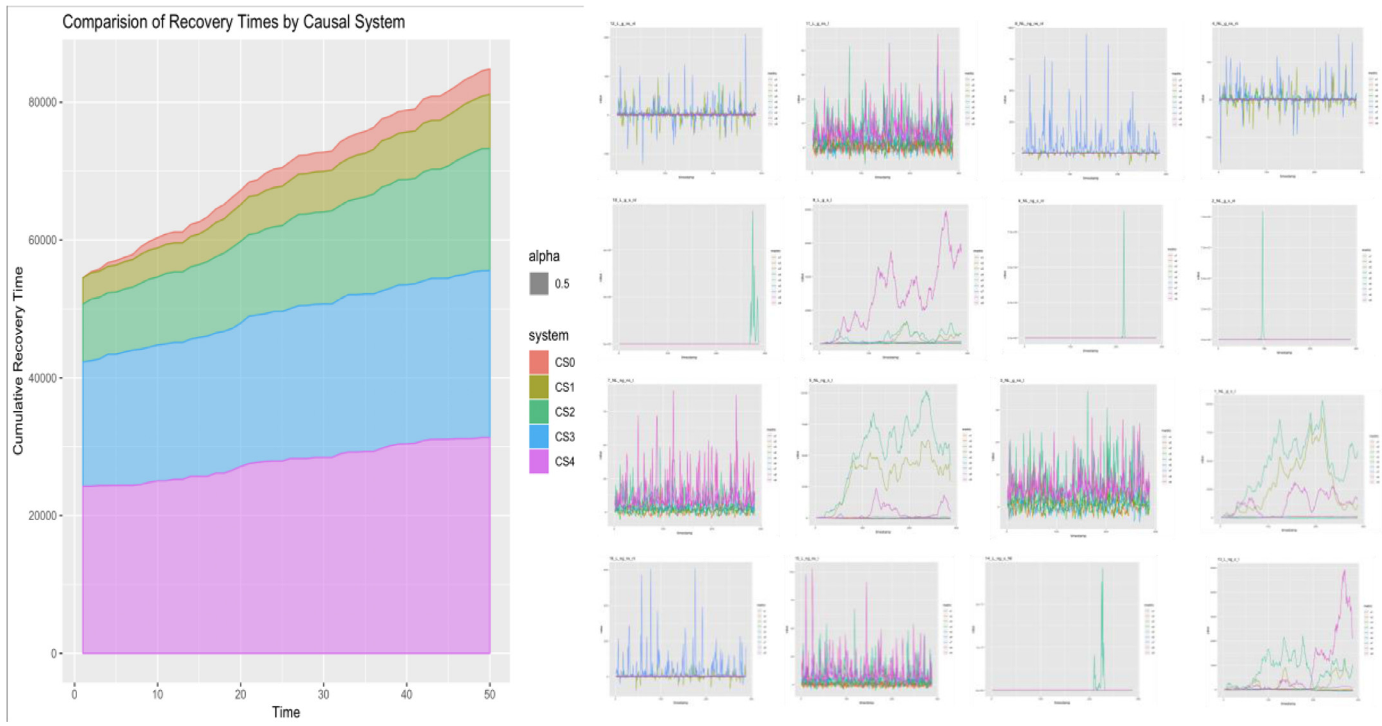

Fig. 5. *Right*: shows different network data distributions: latent variables, Gaussian distribution, linearity, stationarity. *Left*: cumulative recovery time (*vertical axis*) vs network time(*horizontal axis*), different causal systems S_{cs0}, S_{cs1}, S_{cs2}, S_{cs3}, S_{cs4}, S_{cs5} each with a different approach to anomaly detection and root cause analysis with different cumulative recovery times and sub components:.

t_s - *time system goes into recovery*
t_f - *time at which system has recovered*

From Dessavre and Ramirez-Marquez [10] we see that one of the most important factors of a resilient system is the reduction in overall cumulative system downtime and from (2) we want to maximize $\tau_{\text{difference}}$. However there are many different optimized systems S_c or Causal Inference approaches to choose from, examples are described in [40–44] and illustrated in Fig. 5 below.

- T1: *Reliable time frame* τ_1.
- T2: *Vulnerable time frame* τ_2
- T3: *Disrupted time frame* τ_3
- T4: *Recovery time frame* τ_4

From Fig. 5 above illustrates that the network has dynamic network characters(Right) and the differences in performance between algorithms in terms of accuracy based on the data being analyzed from the network. As an illustration, we see that Causal System $S_{cs0,}$ denoted as CS0, has the overall fastest cumulative recovery time, hence this system will be selected in the optimal configuration to maximize user experience. Each potential Causal System has different characteristics with different times to resolve, (*as each causal system was constructed based on different characteristics of the network and trained with different data and algorithms, explained briefly later in this paper*) with cumulative sum of all time components being the most important. Further each system will have different characteristics potentially by disruptive event e_j, characterizing that particular approach for a particular disruptive event e_j and hence corresponding cumulative recovery time. To find the best possible Causal System, a distribution must be considered not only across the 4 component times mentioned above, but *also across likely disruptive events* $e_{j,}$ *potential to occur* against the system under consideration. This is the reason why we do not write: $S_c \in \mathbf{min}\{S_{c1}, S_{c2}, .. S_{cn}\}$. We don't simply blindly choose the minimum *Sc* from the set because that does not consider how the various causal system implementations behave under different events. The next section will build on these models to formulate an optimization objective function.

6. Optimize user experience & network resilience

In this section we combine the models for User Experience and Network Resilience to formulate the Network User Experience Resilience Model Optimization objective. Fig. 6 shows the updated model of the subsystems of the Network which Network User Experience depends on.

The Network has been decomposed into separate subsystems, each with its own resiliency function. We show generic users, as opposed to a single user. If any on the subsystems are not functioning, then users will be directly impacted. The user experience is directly related to the amount of time of any of the sub systems being down or in the S_d state. We first describe the optimization function from a single network KPI perspective, which in 6.1 describes in general, how to choose the best causal system for a particular objective KPI and disruptive event e_j. Section 6.2 then extends the general optimization function obtained from 6.1 to all user experience related network KPIs and across all event types ej.

6.1. Network resilience optimization function

$$\text{Argmax}\{\sum \tau_{\text{difference}}(S_c, e_j)\}$$
$$\text{for } i \in \{\tau_2, \tau_3, \tau_4\},\ e_j \in D,$$
$$S_c \in \{S_{c1}, S_{c2,}. .,S_{cn}\}$$
– *set of all potential causal systems*

Since the original system ***S*** Reliability, Vulnerability, Disruption and Recovery times are relatively constant, the optimization function simplifies to minimizing the Vulnerability, Disruption and Recovery times of $S_c \in \{S_{c1}, S_{c2}, .. S_{cn}\}$, we then get the following:

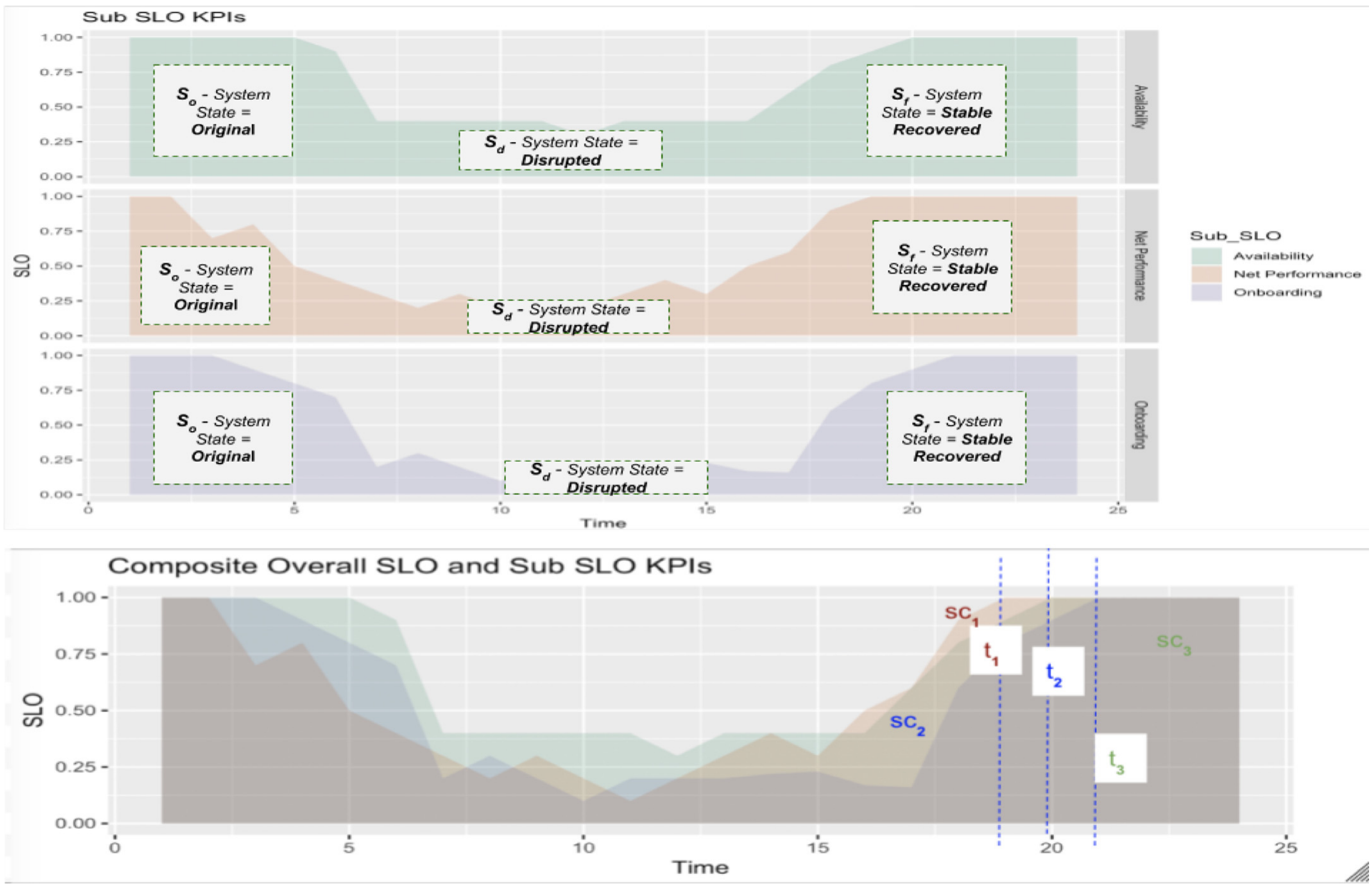

Fig. 6. Extended network user experience resilience model where events represents points in time where SLOs are violated and we have decomposed the system into subsystems, each with its own resilience model. Top portion shows individual Sub SLOs having its own resiliency state model for each KPI and bottom portion showing 3 different causal systems, S_{c1} with t_1, S_{c2} with t_2, S_{c3} with t_3, clearly S_{c1} is the fastest to recover.

$$\text{Argmin}\{\textstyle\sum \tau_i(S_c, e_j)\}$$
$$\text{for } i \in \{\tau_2, \tau_3, \tau_4\},\ e_j \in D,\ \ S_c \in \{S_{c1}, S_{c2,}. .,S_{cn}\} \tag{3}$$

From Dessavre and Ramirez-Marquez [10] Theorem 3, the solution to the optimization of the partial problem is equivalent to the solution of the general optimization problem

In order to use the above models in a wireless network, the network has vastly different characteristics, from different network traffic parts in different regions such as rural vs congested cities, different network functions such as simple layer 2 wirespeed switches to proxies that terminate socket connections. Different data distributions and relationships between dependent and independent variables, examples include linear, non-linear, Gaussian, non-Gaussian, stationary and non-stationary. These all require different algorithms and training data and models for both anomaly detection and root causal discovery. This is the motivation for different Causal Systems S_{ci}. In this paper we define a Causal System as the combination of anomaly detection algorithm, machine learning algorithm and related components to detect network disturbances and identify probable root cause. The construction of the different causal systems are beyond the scope of this paper, but will be described with sufficient details, as needed later in this paper. The main point is that no single generic system is optimal for different data and network characteristics. Different causal systems that incorporate anomaly detection and root cause discovery based on a particular set of data will yield optimal results trained and tuned for that classification of data characteristics.

6.2. User experience network resilience optimization function

In order to optimize Network User Experience, we minimize the time in S_d state for each subsystem(Onboarding, Availability, Network Performance). This is equivalent to maximizing the positive time difference between original and optimized times in State S_d of Eq. (3) for each subsystem of Onboarding, Availability and Network Performance.

From Eq. (3) we expand for each parameter Onboarding, Availability and Network Performance:

$$\begin{aligned}
&\textbf{Optimized Ux} \\
&= \textbf{Argmin}\{\textstyle\sum \tau_i(S_c, e_j)\}_{\text{Onboarding}}{}^{*} \\
&\qquad \textit{for } i \in \{\tau_2, \tau_3, \tau_4\},\ e_j \in D,\ S_c \in \{S_{c1}, S_{c2}, \ldots, S_{cn}\} \\
&\quad \textbf{Argmin}\{\textstyle\sum \tau_i(S_c, e_j)\}_{\text{Availability}}{}^{*} \\
&\qquad \textit{for } i \in \{\tau_2, \tau_3, \tau_4\},\ e_j \in D,\ S_c \in \{S_{c1}, S_{c2}, \ldots, S_{cn}\} \\
&\quad \textbf{Argmin}\{\textstyle\sum \tau_i(S_c, e_j)\}_{\text{NetworkPerformance}}{}^{*} \\
&\qquad \textit{for } i \in \{\tau_2, \tau_3, \tau_4\},\ e_j \in D,\ S_c \in \{S_{c1}, S_{c2}, \ldots, S_{cn}\} \\
&= \Pi_{\text{Argmin}}\{\textstyle\sum \tau_i(S_c, e_j)\} \\
&\textit{for } i \in \{\tau_2, \tau_3, \tau_4\},\ e_j \in D,\ S_c \in \{S_{c1}, S_{c2}S_{cn}\}, \\
&j \in \{Onboarding,\ Availability,\ NetworkPerformance\}
\end{aligned} \tag{4}$$

Each component Network Onboarding, Availability and Performance is contributing to the user experience and is considered equally important(*or weighted*) in a serial manner the same way probability of failures are multiplied if there is no alternate path. Hence we multiply since the user experience is composed of each one of these components. From Eq. (4), we see that the Network User Experience Optimization Function maximizes the spread of time between the

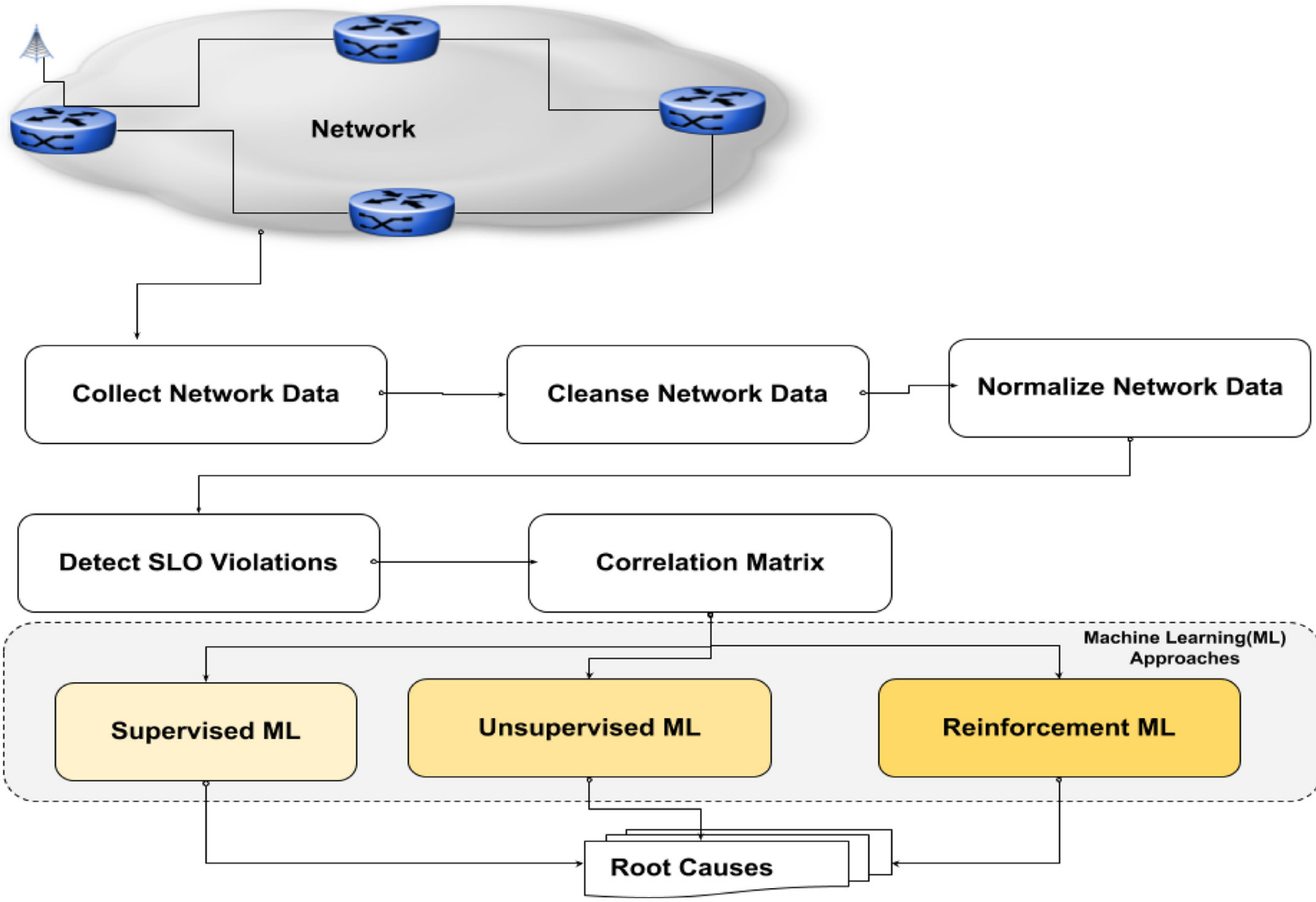

Fig. 7. Network user experience resilience optimization implementation to proactively predict network events and diagnose root cause using ensemble of statistical machine learning approaches.

original down time for each subsystem S and the new optimized subsystems S_c. The means S_c must be minimized as much as possible approaching zero. Fig. 6 below describes our extended model, the top figure shows individual sub SLO which in composite result in the overall composite SLO shown in the lower half, which consists of 3 different causal systems S_{c1}, S_{c2}, and S_{c3} each with corresponding time to recover t_1, t_2, t_3 and the best one is S_{c1}, with the fastest recovery time $t_1 < t_2 < t_3$.

In this example we can see as availability drops to 50%, the performance and onboarding KPIS are severely reduced and not meeting min SLO thresholds, which is modeled as Disrupted State using the Resilience Model Framework [7–9].

We propose to minimize S_d time, based on the lowest S_c found in comparison to all other causal systems and to drive S_d to as close to zero as possible by reducing the time to detect and diagnose events$_j$ for each of the sub states or reduce ts(*time the system starts recovery, see* Fig. 4), as close to t_e(*time disruptive event occurred, see* Fig. 4) as possible: $t_s - t_e < \varepsilon$, where ε is max threshold tolerable time between the event e_j and time to start restoration efforts, at which point operators know root cause and know what is required to repair outage. Current industry practice is that network operations wait for catastrophic events or customer calling in to complain, and lack the tools to proactively detect network events$_j$. We propose leveraging streaming data analysis and machine learning to achieve this at a relatively low cost with reasonable performance.

7. Case study

In this section, we present a specific case study of the general proposed framework, with an example prototype system that we implemented for an actual wireless deployment, for illustrative purposes only. For comparison, we show a typical network operations dashboards, illustrating the limitations in that it is practically impossible for a human to proactively scan all metrics to detect anomalies and identify potential causes due to velocity and volume of data. We then show how the proposed prototype, using the proposed approach materially improved user experience by detecting anomalies or deviations from expected service levels, then focussing on network issues that are directly causal to user experience disturbances. Causal Inference for time series data is beyond the scope of this paper. However, in our results we show how we used existing machine learning algorithms that worked best and noted the network data distributions to validate our claims that best algorithms work when then assumptions are valid. Our prototype first detect anomalies, using thresholds or existing algorithms that statistically compute outliers. Characteristics or Features: Stationarity, Linearity, Normality and of the predictors are computed in chunks of time frames before, during and after the network disturbance that caused the anomaly. We then dynamically select the optimal machine learning approach based on the characteristics of the predictors to determine root causes. Machine learning algorithms are designed for prediction, classification and not specifically for identifying causal relationships. We attempted to estimate root cause during periods of disturbances using different machine learning approaches and techniques particular to that algorithm. We first reduce the set possible candidate root causes using correlation then we intersect with results from the machine learning approaches, that try to find the covariates that explain the dependent variable the most by permuting one covariate at a time. We found that Linear Regression and Random Forest had built in variable importance measurements which was used to find most probable root causes by permuting covariates, computing the drop in F-test and p-values. In Random Forest importance variables are used to identify which covariate contributes most to the reduction in residual squared errors. Support Vector Machines and Simple neural network have little or no interpretability and an indirect method must be used to find potential root causes. Generally we found that linear regression did not perform well for non-linear data whereas support vector machines did better, but interpretability required another library that measured change in explanation of variance, by modifying one covariate, keeping others constant and ranking the covariates. These are estimates only for illustrative purposes. The example demonstrates the novelty of the paper in the sense we propose an approach where we can show the feasibility of a data pipeline that can continuously monitor user experience and once there is a drop, and dynamically select the optimal machine algorithm based on data characteristics to determine the most probable root cause. The different causal systems based on the different machine learning algorithms perform differently based on the data distributions and characteristics such as stationarity, Gaussian distribution, latent variables, and linearity.

7.1. Practical realization

We implemented a prototype system that collected actual WiFi operators network data, processed the data using Google Compute Platform, performed analysis using machine learning techniques to detect anomalies to User Experience Network SLOs and identify the root cause, the basic flow is shown in Fig. 7 below:

7.2. Current tools to address network resiliency - Manual, time consuming, scalability limitations

In this section we illustrate how network operators actually address Mean Time to Recovery(MTTR). From Fig. 4, in the context of this paper, MTTR is the time from when the operator knows a system is down, assigns a network engineer to diagnose the fault, determine root cause and take corrective action to the point in time the system has recovered. Mean Time to Repair is often confused with Mean Time to Recovery which is slightly different as the former spans the time to perform repairs, after diagnosis of the fault. The diagram below, Fig. 8 shows a typical network telemetry dashboard, where operators are in a reactive mode requiring constant human monitoring and expert diagnosis. Network operators must manually look at alarms that are prioritized based on device criticality, not user impact and then the operator must use experience and sift through datasets to deduce root cause. The problem with this approach is that the alarm being addressed may not matter to user experience, the alarms that do matter may be hidden the massive list of alarms and the operator is spending manual efforts in diagnosing the issue, resulting in a long Mean Time to Recovery (MTTR), reducing system resilience.

7.3. Example causal systems

We now describe some hypothetical practical realizations of example causal systems. Fig. 9 below describes a typical system that ingest Network Telemetry data for Network Performance, Availability and Onboarding, each signal has potentially different distributions. The diagram shows 3 example Causal Systems, with different algorithms to detect anomalies, diagnose root causes and optimized for different traffic distributions. We see some systems are able to detect and diagnose at different rates. For traffic distribution A, Causal System 1 was best, For another traffic distribution another causal system using different algorithms will be best, some other ones may not even work correctly. The actual construction of the different causal systems is beyond the scope of this paper, however a brief explanation will be provided in the next section.

7.4. Causal systems - Brief Explanation

The actual construction of the optimal causal system is beyond the scope of this paper, hower we will provide a brief explanation. The construction of causal system involves 2 phases:

(1) Classifier - This phase uses the network telemetry training data as predictors and accuracy of the machine learning approach as the labels. This classifier finds the best machine learning approach for the given network telemetry data characteristics.
(2) Forwarder - This phase takes in the real time network telemetry data in windows: before, during and after the anomaly is detected. Using the network characteristics computed in the window(during the network disturbance, immediately after the anomaly is detected), we can use the classifier in real time to identify the most accurate probable root causes.

For different parts of the network, there may for example different levels of congestion. For example in rural networks, the utilization is typically always low, and for a congested city, during rush hours, the network traffic along main streets is usually high. Due to differences in network characteristics: Stationarity, Linearity, Normality, or Latent Variables different models with different training data will be constructed and the classification will be based on the performance of different models with different predictors i.e. network telemetry data features or characteristics. These different data characteristics are the key reason different algorithms and different causal systems produce different results due to the foundational assumptions of the underlying algorithms. In Fig. 9, we basically see different causal systems producing different results because of the way the different causal systems were constructed on different training data, with different hyper parameters and assumptions.

Our proof of concept implementation systems tried a variety of approaches, using linear regression, random forests, support vector machines, simple neural networks against different network traffic distributions of varying degrees of Normality, Stationarity, Latent Variables and Linearity, all producing different results, i.e. that support vector machines and simple neural networks was more accurate(using ROC) for highly non-linear data distributions, whereas linear regression and random forests produced better results on linear, Gaussian, stationary data with no latent variables.

8. Conclusions and further work

In this paper, we addressed a gap in network operations for Network Service Providers where user experience is disconnected from network operations, by proposing a model to link the two domains and formalized an optimization model which is achievable by selecting the most appropriate machine learning approach based on the network data characteristics. The model showed the optimization function is achieved by minimizing the time between when a network event occurs and when recovery completes. This optimization was achieved by a proposed generalized automated framework that detects anomalies or deviations from acceptable service levels and finding associated causes. We developed a model to incorporate reducing the Time to Recover to include proactive prediction and detection of network faults and identifying root cause which enables operations to immediately jump straight to the repair process. We presented a case study including an example prototype implementation validating the proposed approach in a real wireless service provider which was composed of 3 services being monitored and analyzed for anomalies and root causes: on-

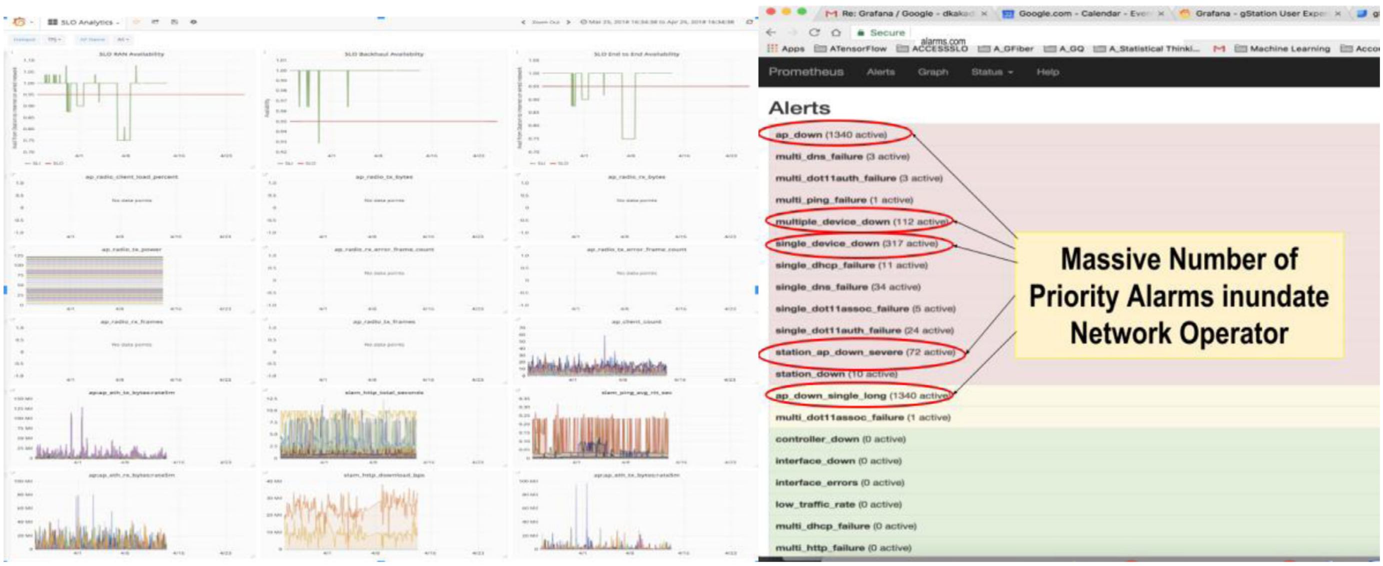

Fig. 8. Typical wireless network service providers network operations dashboard left: metrics time series graphs, right: raw uncorrelated alarms based on fixed thresholds. Current network operations require manual reactive efforts and experience to understand raw metrics, correlate alarms to determine root cause, resulting in longer MTTR and reduced resilience. By the time the customer calls in to support and complains, an assigned network engineer figures out the root cause elapsed time is typically from Θ(hours) to Θ(days). Proposed approach shortens this time frame to Θ(minutes).

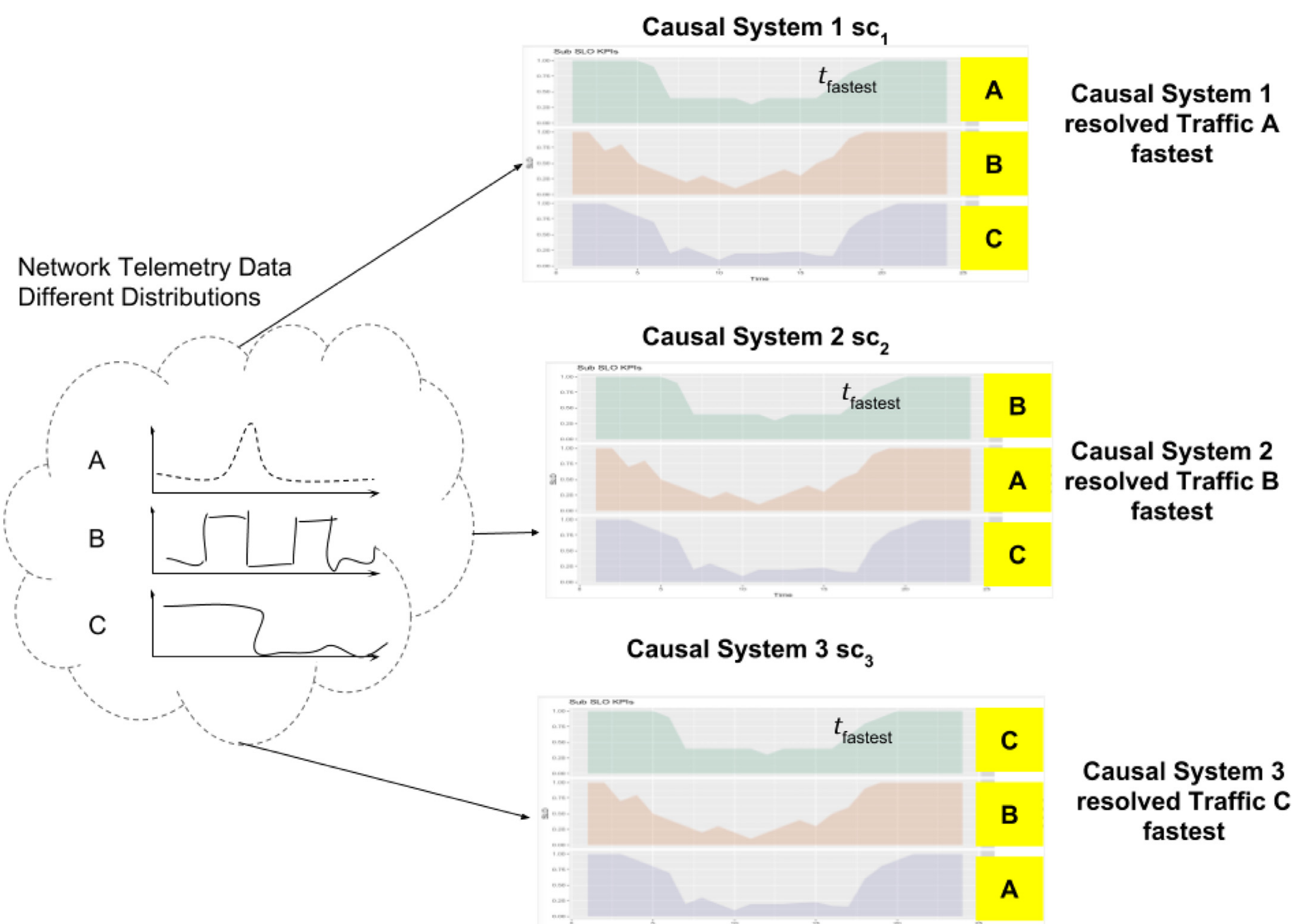

Fig. 9. Example practical realization of causal system selection. Optimal machine learning algorithm is dynamically selected based on past training that, based on the network telemetry data features such as stationarity, linearity, normality the optimal algorithm is chosen that produces the most accurate results of variable importance for that feature set or network telemetry data characteristics.

boarding, availability and network performance. The proposed approach allows for automation which dramatically reduces the time to recover, reduces operational costs and results in improved user experience. The proposed approach is being operationalized and changing previous workflows which involved customers complaining to call centers who file a ticket then notify operations to proactive monitoring and detection of faults that impact user experience and prompt repair, where in many cases users are prevented from impact. Further we used a combination of statistical machine learning approaches, correlation matrix, Linear Regression and Random Forest, Support Vector Machines, Simple Neural Networks to infer root cause showing by dynamically selecting the most suitable approach based on the network data characteristics yields promising results, but still require more research.

In Conclusion we proposed and illustrated a promising approach to optimize network customer user experience by increasing the resiliency of network and subsystems extending the framework of [7–10]. Causal Inference is beyond the scope of this paper however further research is needed to further increase the accuracy of determining root causes.

References

[1] Hosseini S, Barker K, Ramirez-Marquez JE. A review of definitions and measures of system resilience. Reliab Eng Syst Saf 2016;145:47–61.
[2] Xin L, Sridharan A, Machiraju S, Seshadri M, Zang H. Experiences in a 3G network: interplay between the wireless channel and applications. Proceedings of the fourteenth ACM international conference on mobile computing and networking 2008. https://doi.org/10.1145/1409944.1409969. September 14-19.
[3] Balachandran A, Sekar V, Akella A, Seshan S, Stoica I, Zhang H. Developing a predictive model of quality of experience for internet video. Proceedings of the SIGCOMM. 2013.
[4] Bouch A, Kuchinsky A, Bhatti N. Quality is in the eye of the beholder: meeting users A- Z requirements for internet quality of service. Proceedings of the CHI. 2000.
[5] Tullis T, Albert W. Measuring the user experience: collecting, analyzing, and presenting usability metrics. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc; 2008.
[6] Baltrunas D. Measuring the reliability of mobile broadband networks. Proceedings of the IMC'14. 2014. November 5–7.
[7] Ramirez-Marquez JE, Rocco C. Towards a unified framework for network resilience. Proceedings of the third international engineering systems symposium, CESUN 2012. Delft University of Technology; 2012. 18-20 June 2012.
[8] Barker K, Ramirez-Marquez JE, et al. Resilience-based network component importance measures. Reliab Eng Syst Saf 2013;117:89–97.
[9] Henry Devanandham, Ramirez-Marquez Jose Emmanuel. Generic metrics and quantitative approaches for system resilience as a function of time. Reliab Eng Syst Saf 2012;99:114–22.
[10] Gama Dessavre D, Ramirez-Marquez JE. Multi-Event Resilience Optimization Formulations. 27th European Conference on Operational Research. University of Strathclyde; 12-15 July 2015.
[11] Pearl J. Causality. Oxford University Press; 2000.
[12] Gregorutti, B., Michel, B., Saint-Pierre, P., Correlation and variable importance in random forests, arXiv:1310.5726, 10/2013.
[13] Diaz-Uriarte, R., Alvarez de Andres, S., Variable selection from random forests: application to gene expression data, arXiv:q-bio/0503025, 03/20.
[14] J. G., Indrayudh, Hooker, G., Boosting random forests to reduce bias, one-step boosted forest and its variance estimate, arXiv:1803.08000 , 03/2018.
[15] Vainio-Mattila K, Vainio T, Maehr W, Joshi D, Shrestha S. User experience of mobile internet: analysis and recommendations. International Journal of Mobile Human Computer Interaction 2009;1(4):4–23.
[16] Bao Y. Data-Guided approach for learning and improving user experience in computer networks. Proceedings of the ACML JMLR: Workshop and Conference Proceedings. 30. 2015. p. 1–16.
[17] Schatz R, Egger S. AN annotated dataset for web browsing QoE. Proceedings of the 2014 sixth international workshop on quality of multimedia experience (QoMEX). 2014.
[18] Bao Y, Wu H, Liu X. From prediction to action: improving user experience with data-driven resource allocation. IEEE J Sel Areas Commun MAY 2017;35(5).
[19] Kumar D, Nowicki D, Ramirez-Marquez JE, Verma D. A Goal Programming Model For Optimizing Reliability, Maintainability And Supportability Under Performance Based Logistics. Int J Reliab Qual Saf Eng 2007;14(03):251–61https://doi.org/10.1142/S0218539307002623.
[20] Ramirez-Marquez JE, Coit D. Multistate component criticality analysis for reliability improvement in multistate systems. Reliab Eng Syst Saf 2007;92(12):1608–19 Special Issue on ESREL 2005.
[21] Kumar D, Ramirez-Marquez JE, Nowicki D, Verma D. Reliability and maintainability allocation to minimize total cost of ownership in a series-parallel system. Int J Risk Reliab 2007;221(2):133–40.
[22] Ramirez-Marquez JE, Coit D. Optimization of system reliability in the presence of common cause failures. Reliab Eng Syst Saf 2007;92(10):1421–34.
[23] Sols A, Ramirez-Marquez JE, Verma D, Vitoriano B. Evaluation of full and degraded mission reliability and mission dependability for intermittently operated multi-functional systems. Reliab Eng Syst Saf 2007;92(9):1274–80.
[24] Assefa B, Ramirez-Marquez JE. Element substitution algorithm for general two-terminal network reliability analyses. IIE Trans 2007;39(3):265–75.
[25] Ramirez-Marquez JE, Jiang W. On improved confidence bounds for system reliability. IEEE Trans Reliab 2006;55(1):26–36.
[26] Ramirez-Marquez JE, Jiang W. Confidence bounds for the reliability of binary capacitated two-terminal networks. Reliab Eng Syst Saf 2006;91(8):905–14.
[27] Ramirez-Marquez JE, Rocco C, Assefa B, Coit D, Tortorella M. New insights on multi-state component criticality and importance. Reliab Eng Syst Saf 2006;91(8):894–904.
[28] Ramirez-Marquez JE, Coit D. A monte-carlo simulation approach for approximating multistate two-terminal reliability. Reliab Eng Syst Saf 2005;87(2):253–64.
[29] Ramirez-Marquez JE, Coit D. A heuristic for solving the redundancy allocation problem for multistate series-parallel systems. Reliab Eng Syst Saf 2004;83(3):341–9.
[30] Ramirez-Marquez JE, Coit D, Konak A. Reliability optimization of series-parallel systems using a max-min approach. IIE Trans 2004;36(9):891–8.
[31] Kim TH, Tipper D, Krishnamurthy P, Swindlehurst AL. Improving the topological resilience of mobile ad hoc networks. Proceedings of the seventh international workshop on design of reliable communication networks. 2009. p. 191–7.
[32] Sun S, Li R, Wang L, Xia C. Optimization of network resilience under attacks based on Simulated Annealing. Proceedings of the 33rd Chinese Control Conference, CCC 2014 2014. p. 6431–6. https://doi.org/10.1109/ChiCC.2014.6896049.
[33] Gunkel M, Autenrieth A, Neugirg M, Elbers JP. Advanced multilayer resilience scheme with optical restoration for IP-over-DWDM core networks. Proceedings of the fourth international congress on ultra modern telecommunications and control systems. 2012. p. 657–62.
[34] Alves RJM, Rodrigues MM, Paiva MHM, Segatto MEV. A note on resilience of telecommunication networks. Proceedings of the CHILEAN conference on electrical, electronics engineering, information and communication technologies (CHILECON). 2017. p. 1–6.
[35] Machado CC, Granville LZ, Schaeffer-Filho A. ANSwer: combining NFV and SDN features for network resilience strategies. Proceedings of the IEEE Symposium on Computers and Communication (ISCC). 2016. p. 391–6.
[36] Bachmann I, Reyes P, Bustos J, Silva A. Multidimensional network resilience analysis. IEEE Lat Am Trans 2016;14(6):2912–4.
[37] Marnerides AK, Bhandari A, Murthy H, Mauthe AU. A multi-level resilience framework for unified networked environments. Proceedings of the IFIP/IEEE international symposium on integrated network management (IM). 2015. p. 1369–72.
[38] Bai Y, Huang N, Xu K, Zhang X. Resilience analytics of networks with dependency groups. Proceedings of the second international conference on system reliability and safety (ICSRS). 2017. p. 486–90.
[39] Balachandran A, Aggarwal V, Halepovic E, Pang J, Seshan S, Venkataraman S, Yan H. Modeling Web Quality-of-Experience on Cellular Networks. ACM MobiCom. Hawaii, USA: Maui; Sep. 2014.
[40] Rafique D, Szyrkowiec T, Autenrieth A, Elbers JP. Analytics-driven fault discovery and diagnosis for cognitive root cause analysis. Proceedings of the optical fiber communications conference and exposition (OFC). 2018. p. 1–3.
[41] Miyazawa M, Nishimura K. Scalable root cause analysis assisted by classified alarm information model based algorithm. Proceedings of the seventh international conference on network and service management. 2011. p. 1–4.
[42] Mfula H, Nurminen JK. Adaptive root cause analysis for self-healing in 5G networks. Proceedings of the international conference on high performance computing & simulation (HPCS). 2017. p. 136–43.
[43] Zasadzinski M, Muntés-Mulero V, Simo MS. Actor based root cause analysis in a distributed environment. Proceedings of the IEEE/ACM 3rd International Workshop on Software Engineering for Smart Cyber-Physical Systems (SEsCPS). 2017. p. 14–7.
[44] Hu J, Zhang L, Wang A, Li S. Accident prevention by fault propagation analysis and causal fault diagnosis based on granger causality test. Proceedings of the thirteenth international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD). 2017. p. 1554–8.
[45] Miyazawa M, Otani T. Real-time root cause analysis in OSS for a multilayer and multi-domain network using a hierarchical circuit model and scanning algorithm. Proceedings of the IFIP/IEEE international symposium on integrated network management. 2009. p. 141–4.
[46] Ahmad A, Floris A, Atzori L. QoE-centric service delivery: a collaborative approach among OTTs and ISPs. Comput Netw 2016;110:168–79.
[47] Liotou E, Tsolkas D, Passas N, Merakos L. Quality of experience and management in mobile cellular networks: key issues and design challenges. Commun Mag IEEE 2015;53:145–53.
[48] Baraković S, Skorin-Kapov L. Survey and challenges of QoE management issues in wireless networks. J Comput Netw Commun 2013;2013:165146. Article ID.
[49] Aggarwal V, Halepovic E, Pang J, Venkataraman S, Yan H. Prometheus: toward quality-of-experience estimation for mobile apps from passive network measurements. Proceedings of the HotMobile. 2014.