Reactive versus Predictive Networks

JP Vasseur (jpv@cisco.com), PhD - Cisco Fellow Engineering Lead – Gregory Mermoud (gmermoud@cisco.com), PhD – Principal Engineer – Vinay Kolar (vinkolar@cisco.com), PhD Principal Engineer – Eduard Schornig (eschorni@cisco.com), Sr Technical Leader

Abstract: **Over the past 35 years, several protocols and technologies have been developed to make the Internet more reliable. What they all have in common is to be reactive: failures (usually lack of connectivity) are first detected as quickly as possible, and traffic is then rerouted along alternate paths that are either computed on-the-fly (restoration) or pre-computed (protection). Such a reactive approach has been dominant across all layers: physical/link layer (e.g., Optical, Ethernet, WiFi) and IP/MPLS (e.g., IP FRR, MPLS FRR, IGP). The concept of Predictive Networks was first introduced in 2021 as a new paradigm: the ability for networks to learn, predict and plan. Using a variety of data sources, models are computed that can be used to predict issues/failures at multiple (before they happen) thus allowing for proactively rerouting traffic and thus avoiding the issue. Still both approaches are complementary since all issues cannot be predicted.**

1. Predictive vs Reactive Networks

The concept of Predictive Networks was first introduced in 2021 as a new paradigm for networking. In summary, such networks rely on a predictive engine in charge of computing (statistical, machine learning) models of the network using several telemetry sources. Such models are used to predict various types of issues in the network (e.g., Service Level Agreement (SLA) violation, lack of connectivity) along a set of paths. Predictions can then be assessed by another engine in charge of path selection to potentially trigger a pro-active rerouting of the traffic, should an alternate path be available that significantly reduce the risk of issues/failures.

In contrast, existing reactive mechanisms rely on the detection of an issue (e.g., lack of connectivity, degraded SLA using probes) before rerouting the traffic on some alternate paths whose characteristics are usually unknown under new conditions (after rerouting the traffic along the impacted path) [2].

What does that mean in terms of user experience?

For all issues/failures that are predicted and for which an alternate path exists, issues are simply **avoided**. In contrast, reactive strategies first need to detect the issue, and then only address it by re-routing the traffic onto some alternate path. Thus, by definition, this means that the network has experienced an issue, and the user experience has already been degraded.

How long does it take for a reactive system to reroute? It depends. In the case of a complete loss of connectivity, Keep-alive (KA) mechanisms usually react within a few seconds (depending on the configuration). If the issue relates to SLA violations, requiring the use of probes (see Grey

Failure [3]), then the system must first assess the Quality of Experience (QoE) for some time before triggering a reroute (which takes usually several minutes). One may use aggressive timers to reduce the convergence time; this comes at the risk of introducing oscillations. An over-reactive approach may very likely trigger excessive rerouting (thus impacting the user experience). Along paths where transient issues are frequent (a very common situation in the Internet), such an approach would lead to constant rerouting and potential oscillations that are highly undesirable.

Moreover, there is no knowledge of the QoE on the alternate path once traffic has been rerouted; predictive engines, in contrast, can use models that can be tested against new conditions before rerouting.

Studies on large number of networks have shown a high number of grey failures (degraded user experience) compared to dark failure (loss of connectivity). For each of those grey failures, a reactive approach would have to assess SLA for a given time interval during which SLA violation is experienced before rerouting.
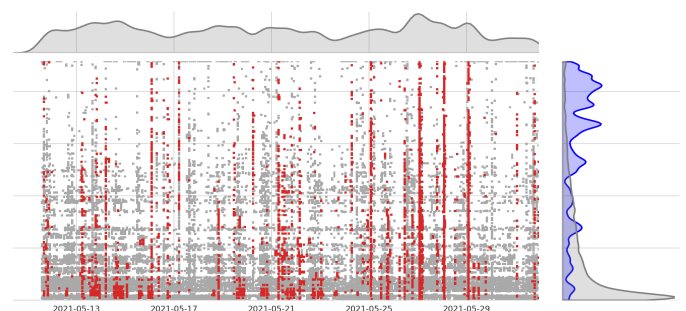

Figure 1: Dark versus Grey Failures

Consider a path across the internet suffering from daily congestion at a given time; a reactive approach would lead to the repetitive impact for the user every day at the same time, whereas a predictive approach would be able to predict and avoid. Of course, seasonality is simply an example signal to that can be used for predictions. There are usually other early signs (e.g., fluctuations in jitter) that can also be used to predict gray failures before they occur.

Figure 2 illustrates how a user does not completely avoid a bad experience in a reactive approach, while a predictive approach attempts to avoid the poor user experience altogether. The Y-axis shows the SLA violation, and the green line shows the ground-truth SLA violation over a path. The blue dashed line shows the predicted SLA violation by the predictive approach, and the red dotted line shows the reactive approaches to estimate SLA violation based on the past. A reactive

approach may of course catch up after the onset of a grey failure, but it will invariably fail to eliminate completely the disruption.
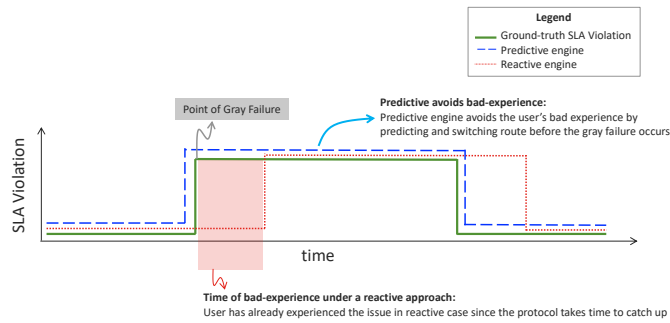

Figure 2: Illustration of predictive vs reactive protocol for avoiding bad - experience.

2. Can Predictive Networks predict all issues and how about accuracy?

No. None of the existing systems can predict all potential issues with 100% accuracy and Predictive Networks are no exception.

Let us first focus on the nature of issues that such a system is designed to predict. In theory, most issues should be predictable or at least preceded with some early signs that could be used by a model to forecast the issue. Why? Because strictly random events are simply extremely rare in nature. Even a fiber cut can be predicted, but only a few milliseconds before breaking (using advanced optics telemetry)! Router crashes are oftentimes predictable, but this would require unreasonable amount of telemetry data to be obtained from many parts of the system. In such situations (signal available a few milliseconds before the issue, or lack of granular telemetry), early signs of an issue, even if present, are simply not available to the predictive engine. Said differently, even if early signs exist, this may require telemetry that is not available to predict such issues.

Moreover, any system that predict events using statistical or ML models must be tuned with specific objectives in mind. Let us consider the classic example of predicting events using a classification ML algorithm. ML engineers always face the challenge of optimizing their system for the highest precision or recall (see Figure 3).
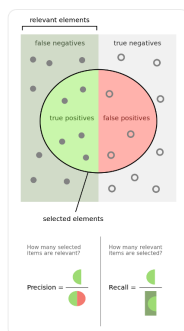

Figure 3: Notion of Recall & Precision

Although the objective is to optimize both precision and recall, there is a tradeoff between precision and recall; increasing recall usually leads to decreasing precision and vice-versa.

Cisco's predictive engine has been optimized for high precision; i.e., it attempts to avoid as many disruptions as possible (precision should always be close to 1). Said differently, the objective is not to predict all issues, but rather to ensure that predicted issues almost surely occur (that is, the system rarely generate false positives).

Why? Simply because precision is an absolute must to enable trusted automation in self-healing networks.

Furthermore, existing reactive networks (and the Internet as a whole, as a matter of fact) achieve a recall of zero since no issue is ever predicted and avoided before it occurs. We further discuss the actual performance of the system in our study of Predictive Networks [1].

3. Can we quantify the benefits of Predictive Versus Reactive?

Reactive approaches, by definition, cannot predict the onset of application disruptions. To quantify the benefits of a predictive engine, we introduce the notion of Application Failure (AF). As pointed out, reactive strategies cannot avoid AFs: traffic along the impacted path will experience degraded conditions for as long as the detection is not confirmed (between a few minutes and an hour depending on the conditions) and then only the traffic will be re-directed onto a **(potentially)** non-violating path. Figure 4 illustrates (a) the fraction of AFs successfully predicted by current version Cisco's predictive engine and (b) the fraction of how many sessions, session-minutes or users could have been forecasted to have an AF.
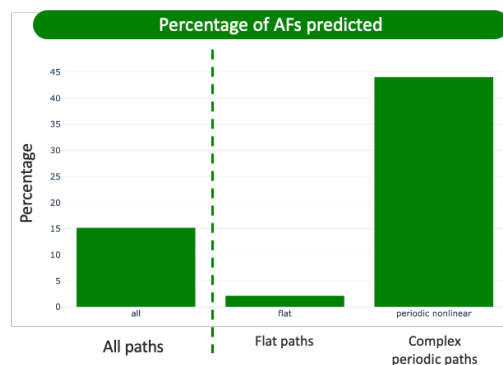

Figure 4: Performance of our predictive engine on 11 enterprise networks, for a total of more than 108 000 paths.

Across all paths, 15% of AFs are predicted correctly. As dicussed in details in prediction accuracy varies with the nature of the path. For so-called *flat* paths that have only very few issues, the engine predicts only 2% of the AFs, most likely because those are due to very unpredictable mechanisms. However, on a subset of the paths that exhibit a periodic pattern, the model predicts more than 44% of the AFs correctly. This demonstrates the need for predictive methods in situations where patterns of violation exist. Note again that reactive approaches have, by definition, a recall of 0%.

4. Will Predictive Networks replace reactive strategies?

No. Predictive systems will never be able to predict and forecast all issues, thus a reactive approach is required. If the issue is predicted, then traffic will be proactively rerouted onto an alternate path. If the issue is not predicted the reactive mechanism will be triggered.

This explains why both mechanisms work in concert and are totally complementary. Since the system is optimized for high precision (no False Positive) all actions proactively triggered by the predictive mechanism will have a positive impact while other non-predictive issues will be handled as in today's networks.

5. Why the future of the Internet will be predictive

Exclusively relying on reactive approaches no longer suffics. The fundamental reason for developing and deploying predictive engines in the Internet is not merely their advantages over reactive methods. They are also a stepping-stone towards truly application-driven networks that rely on other inputs than active probes for path selection, self-configuration, and self-healing. For instance, feedback from the application and its users is necessarily delayed by several minutes or hours, which makes it unusable in reactive settings. Therefore, networks that are driven by such feedback from the application must rely on predictive methods to make decisions based on historical data using various types of models.

Prediction is not merely about the future: it may also consist in assessing alternate scenarios, which are key in answering questions such as "would it help if I made that change?". A basic example of this situation occurs upon re-routing traffic onto a path with limited capacity: reactive approaches are caught between a rock and hard place wherein they oscillate between a path with a poor SLA and a path that has not enough capacity to support the re-routed traffic. Learning capabilities are key in removing these limitations. The same logic applies to any situation where automating a remediation is envisioned.

6. Conclusion

Predictive Engines are designed to avoid poor user experiences (such as grey failures) that reactive protocols could never completely avoid. A reactive protocol would, by definition, react to poor experience after the user experiences a degradation. As shown in previous section, the number of "poor experience" impacting users are extremely high; thus avoid such issues drastically improves the user experience.

In contrast, Predictive technologies are designed to predict such bad experience in advance, where possible, and then switch to a better path to save the user from a poor experience. However, the system uses both predictive and reactive approaches together. Reactive approach is a fail-safe mechanism; the reactive component reduces the bad user experience in case the predictive algorithm fails to predict failures.

7. Bibliography

[1] Cisco, "Cisco SD-WAN: Application-Aware Routing Deployment Guide," Cisco, 21 July 2020. [Online]. Available: https://www.cisco.com/c/en/us/td/docs/solutions/CVD/SDWAN/cisco-sdwan-application-aware-routing-deploy-guide.html. [Accessed 28 Aug 2021].

[2] J. P. Vasseur, "From Dark to Grey Failures in the Internet," Cisco Systems, 2021.

[3] J. P. Vasseur, "Towards a Predictive Internet: A new world, with new challenges," Cisco Systems, 2001.

[4] J. Vasseur, "Predictive Networks: Networks that Learn, Predict and Plan," 2022.