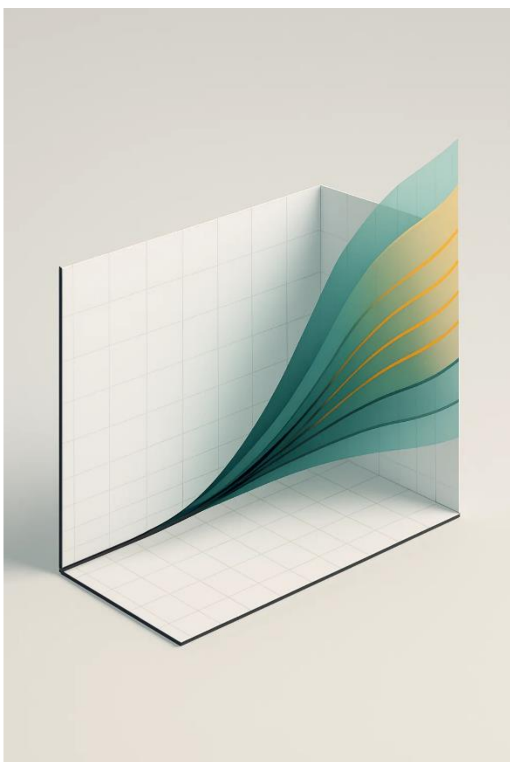
Probabilistic Traffic Modeling: Moving Beyond Static Forecasts

Introduction – Accurate traffic forecasting is the bedrock of network planning. Traditionally, planners have relied on **static forecasts** – a single predicted number (or a fixed growth rate) for future traffic demand – to guide capacity upgrades. For example, a planner might assume "traffic will grow 30% next year" and design the network accordingly. While straightforward, this deterministic approach treats the future as a single scenario. In reality, traffic growth is uncertain and can fluctuate due to myriad factors: user behavior, new applications, economic shifts, even global events. Static forecasts often prove wrong, either **overshooting or undershooting** actual demand. The consequence is either over-provisioned networks (wasted capital on capacity that isn't used) or under-provisioned networks (congestion and scrambling to add capacity). To address this, the industry is moving toward *probabilistic traffic modeling* – embracing uncertainty by forecasting a **range of possible futures** rather than one definitive outcome. By moving beyond static forecasts, network planners can make decisions that are robust under many scenarios, balancing risk and efficiency in a smarter way.



The Problem with Deterministic Forecasts

Relying on a single "most likely" traffic projection is inherently risky. Even sophisticated forecasts can't perfectly predict the impact of unpredictable events (for instance, the sudden shift to remote work during 2020 which caused traffic surges in residential networks). Static forecasts also tend to bake in assumptions that all parts of the network will grow uniformly and steadily, which is rarely true. Different services or regions can grow at different rates, and usage can be bursty rather than smooth.

Historically, because of these uncertainties, network planners compensated by **over-designing for worst-case scenarios**. Essentially, they would take a static forecast and then add a big safety margin (see Blog 1 on margin stacking). One large operator noted that forecasting traffic more than six months out is so challenging that their "original approach was to handle traffic uncertainties by dimensioning the network for worst-case assumptions and sizing for a higher percentile, say P95" engineering.fb.com. In practice, that meant building as if the top

5% highest traffic load might be the norm. This approach avoids surprises – but it also means that 95% of the time, the network is underutilized because the worst-case rarely happens continuously. Facebook (Meta) engineers recently observed that tracking every service's traffic spike individually led to lots of false alarms, since "most of these surges are harmless because not all services surge simultaneously" engineering.fb.com. In other words, **not everything peaks at once**, and treating a combination of worst-cases as a single scenario leads to substantial over-provisioning.

The limitations of static forecasts become evident in two failure modes:

- **Overestimation (Overbuilding):** If the forecast overshoots actual growth, the operator ends up deploying excess capacity that may stay idle for quite some time. This was common in the past where, for instance, long-term data traffic was over-forecasted and fiber was deployed far ahead of demand. The **cost impact** of overestimation is huge – capital tied up in unused infrastructure and higher operating costs for maintenance and power with little revenue to show for it. One analysis of legacy optical networks noted that engineering to worst-case "creates significant inefficiencies and capital waste… underutilized equipment and rarely used capacity" blueplanet.com. Static forecasts leaning high effectively multiply the margin stacking problem, as planners add capacity "just in case" that demand appears.
- **Underestimation (Capacity Crunch):** On the flip side, if a static forecast is too conservative, demand can exceed capacity sooner than planned. This results in scrambling – emergency installs, expedited orders for new circuits, or temporary fixes that might be costlier. It can also degrade customer experience if congestion occurs. Underestimation tends to happen when an unforeseen driver emerges (for example, a viral new app or higher-than-expected uptake of a broadband service). With static forecasting, there's little insight into *risk* – planners might know the single forecast value, but not the probability of exceeding it. So they either get it wrong or play very safe. Both outcomes are suboptimal.

Embracing Uncertainty: Probabilistic Models

Probabilistic traffic modeling aims to solve these issues by treating traffic forecasts not as a single number, but as a **distribution of possibilities**. Instead of saying "we expect 100 Gb/s next year," a probabilistic model might say "there's a 50% chance traffic will be between 95–105 Gb/s, a 30% chance it's higher, up to 130 Gb/s, and a 20% chance growth is slower, perhaps 80–95 Gb/s." This could be derived from statistical models, trends with confidence intervals, or simulation of different scenarios. The planner then has a richer picture: not just a point estimate, but a range with associated probabilities.

There are a few approaches to implement probabilistic forecasts in practice:

- **Scenario Planning:** Planners define a set of distinct scenarios (e.g. *Low*, *Medium*, *High* growth cases), each representing a narrative of how the future could unfold. For example, *High* might assume a new video streaming service launches, doubling data usage, whereas *Low* assumes economic slowdown and modest growth. Some network operators already use this method – making "a series of forecasts, each being a scenario which describes a possible future situation" link.springer.com. For each scenario, the network design is evaluated. The goal is to ensure the plan is **robust**, meaning it works reasonably well across all plausible scenarios, not just one. Scenario-based planning acknowledges uncertainty explicitly and forces consideration of extremes without assuming any single view is 100% right.

- **Statistical Distributions and Confidence Intervals:** Using historical data and statistical models, planners can project not just an expected value but a confidence band. For instance, using methods like quantile regression or time-series models, one can derive a forecast that says: demand in a region next year will most likely be ~200 Tb/month, but it could realistically range ±20% with 90% confidence. This method gives a probability to any given demand level. Planners can then decide a **risk threshold** – e.g., plan capacity such that there's only a 5% chance that demand will exceed it. This way, a small risk of congestion is accepted to avoid major overbuilding. It's a calculated risk approach, akin to how insurance or finance might accept certain risk levels. The benefit is a much more **optimized investment**, because you're not building out to the absolute worst-case (which might be a 1-in-100 chance event), only to a more moderate risk level.
- **Monte Carlo Simulation:** This is a computational approach where many random trials are run to simulate traffic growth under various random influences. By running thousands of simulations with different parameters (growth rates, adoption of services, etc., drawn from probability distributions), one obtains a distribution of outcomes. Monte Carlo methods can capture complex, multifactor uncertainties and produce an empirical probability distribution of traffic. Planners can then extract insights like, "In 80% of simulations, traffic stayed below X amount by 2025," and use that to guide capacity decisions.

By embracing these methods, planners shift from asking *"What is **the** forecast?"* to *"What are the **range of possible** forecasts and their likelihoods?"*. This mindset change is profound – it transforms planning into a risk management exercise. Instead of blindly overbuilding, planners can **quantify the risk** of congestion or wasted capacity and make informed trade-offs. For example, they might decide that a 5% risk of needing emergency upgrades is acceptable if it saves, say, $10 million in upfront deployment that might not be used. Or conversely, they might identify that a certain corridor has extremely unpredictable demand (wide uncertainty range) and thus provision a bit of extra capacity there while holding off elsewhere.

Benefits to Network Planning and Operations

Adopting probabilistic traffic modeling leads to several concrete benefits:

- **Higher Efficiency with Confidence:** Networks can be run "hotter" (i.e., with less idle headroom) without significantly increasing the risk of running out of capacity. This is because the planner knows the probability distribution – maybe they plan for the 90th percentile of demand instead of the 100th. The **hidden savings** can be substantial: rather than 30% over-provisioning "just in case," maybe only 10% is added, freeing capital. Over a large network, that difference can be millions of dollars. And if an unexpected surge comes, there's a plan for that (e.g., expedited augment only if the 95th percentile is breached).

- **Better Timing of Investments:** Probabilistic models often highlight not just how much demand *might* grow, but also the range of *when* thresholds might be crossed. Instead of upgrading too early "because our single forecast said so," operators can wait until certain confidence triggers are met. For instance, an upgrade might be initiated only when demand actually hits the 80th percentile forecast, thereby **delaying spend** until necessary. Conversely, if there's a chance demand could spike earlier than median case, contingency plans (like quick-turn augment contracts) can be in place. The net result is a more **flexible, just-in-time approach** to capacity adds.
- **Risk Management and SLAs:** Many network operators have service level agreements (SLAs) or internal reliability targets (e.g., never exceed 70% link utilization during peak to ensure low latency). Probabilistic planning allows them to express these as risk probabilities. For example, "We want less than 1% chance that any core link exceeds 70% utilization in the next year." They can then plan capacity to meet that probabilistic SLA. This is a much more nuanced method than guessing a safe headroom. It also aligns network planning with corporate risk tolerance – akin to how financial portfolios are managed.
- **Adaptability and Learning:** Once a probabilistic framework is in place, it can continuously learn and update. Each time new data comes in (monthly traffic figures, new product launch info, etc.), the model can adjust the forecast distribution. Planners can then re-run scenarios regularly. Over time, the organization builds a **feedback loop**: comparing actual outcomes with forecast distributions to refine their models. This leads to better accuracy or at least better understanding of uncertainty drivers. It also encourages a culture that is not "surprised" by variance, but expects it and plans for it.

Toward a Dynamic Planning Process

Moving beyond static forecasts is closely tied to the broader trend of making network planning a **dynamic, continuous process**. Probabilistic models lend themselves to automation – a software system can automatically generate scenarios or distributions and even recommend actions. For instance, an AI-driven planning tool might regularly output: "Based on current trends, there's a 20% probability that traffic between Data Center A and B will exceed capacity in 3 months. Consider augmenting by 100 Gbps if utilization continues on the 90th percentile trajectory." This marries forecasting with proactive planning in near real-time.

Moreover, probabilistic traffic modeling complements other modern planning techniques like multi-layer optimization (from Blog 2) and real-time network telemetry. If the network is instrumented to provide live utilization data, the planning system can constantly recalibrate forecast vs. actual. This is essentially how **network digital twins** incorporate uncertainty – by simulating not one future, but many, and guiding decisions that perform well across that spread. We can already see this approach in large content provider networks: for example, Meta (Facebook) shifted to a *"network hose"* model, forecasting total data center ingress/egress rather than precise flows, which inherently builds in statistical multiplexing and uncertainty tolerance engineering.fb.com engineering.fb.com. By abstracting demands in this probabilistic way, they simplified planning and let natural variations even out without micromanaging every flow.

Adopting probabilistic modeling does require a shift in mindset and tools. Planners need some statistical training, and organizations need to accept that plans will be expressed in probabilistic terms ("there is a 90% confidence we won't exceed X") rather than absolute certainties. However, the improved outcomes – **leaner networks that still meet user needs** – are compelling. In competitive markets, the ability to avoid both over-investment and service shortfalls is a major differentiator.

Conclusion – Static forecasts served as a compass in an earlier era of networking, but in the face of cloud-era volatility and exponential growth, a single compass point is no longer enough. Probabilistic traffic modeling provides a **map of many possible futures**, enabling network planners to navigate with agility and informed risk management. By planning for a range of outcomes, operators can strike a savvy balance: enough capacity to satisfy demand *almost* always, but not so much that assets sit idle perpetually. This approach dovetails with the broader trends of automation and intelligent planning – it's about using data smarter, not just throwing more hardware at the problem. As networks continue to evolve, embracing uncertainty may be one of the best ways to ensure infrastructure is used efficiently and customers are kept happy. In sum, moving beyond static forecasts turns planning into a proactive, **future-proofing exercise** rather than a gamble on a single predicted trajectory. In the fast-changing world of telecom, that could make all the difference between a network that's overbuilt or overwhelmed, and one that's **adaptive and optimized** for whatever comes next.