# Network Computing - Notes - v0.5.0

260236

September 2025

## Preface

Every theory section in these notes has been taken from the sources:

- Course slides. [2]

About:

GitHub repository

These notes are an unofficial resource and shouldn't replace the course material or any other book on network computing. It is not made for commercial purposes. I've made the following notes to help me improve my knowledge and maybe it can be helpful for everyone.

As I have highlighted, a student should choose the teacher's material or a book on the topic. These notes can only be a helpful material.

# Contents

# 1 Datacenters

## 1.1 What is a Datacenter?

A **Datacenter** is a specialized **facility that houses multiple computing resources**, including servers, networking equipment, and storage systems. These **resources are co-located** (placed together in the same physical location) to ensure efficient operations, leverage shared environmental controls (such as cooling and power), and maintain physical security.

So the main characteristics are:

- **Centralized Infrastructure**: Unlike traditional computing models where resources are scattered, datacenters consolidate thousands to millions of machines in a single administrative domain.
- **Full Control over Network and Endpoints**: Datacenters operate under a single administrative entity, allowing customized configurations beyond conventional network standards.
- **Traffic Management**: Unlike the open Internet, datacenter traffic is highly structured, and the organization can define routing, congestion control, and network security policies.

| Feature | Datacenter Networks | Traditional Networks |
|---|---|---|
| **Ownership** | Fully controlled by a single organization | Usually spans multiple independent ISPs |
| **Traffic** | High-speed internal communication (east-west traffic) | Lower-speed, external client-based traffic (north-south) |
| **Routing** | Customizable (non standard protocols) | Uses standard internet protocols (BGP, OSPF, etc.) |
| **Latency** | Optimized for ultra-low latency | Variable latency, dependent on ISPs |
| **Redundancy** | High redundancy to ensure failover and fault tolerance | Often limited by ISP policies |

Table 1: Difference between Datacenters and other networks (e.g., LANs).

### Why are datacenters important?

Datacenters are the backbone of modern cloud computing, large-scale data processing, and AI/ML workloads. They provide **high computational power and storage** for various applications, such as:

1. **Web Search & Content Delivery**. For example, when a user searches for “Albert Einstein” on Google, the request is processed in a datacenter where:

   (a) The query is parsed and sent to multiple servers.
   (b) Indexed data is retrieved.
   (c) A ranked list of results is generated and sent back to the user.

2. **Cloud Computing**. Services like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud offer computation, storage, and networking resources on-demand.
    - Infrastructure as a Service (IaaS): Virtual machines, storage, and networking.
    - Platform as a Service (PaaS): Databases, development tools, AI models.
    - Software as a Service (SaaS): Google Drive, Microsoft Office 365.
3. **AI and Big Data Processing**. Large-scale computations like MapReduce and deep learning training rely on distributed datacenter resources.
4. **Enterprise Applications**. Datacenters host internal IT infrastructure for businesses, including databases, ERP systems, and virtual desktops.

### Evolution of Datacenters

While the concept of centralized computing dates back to the 1960s, the modern datacenter model emerged with cloud computing in the 2000s. Notable developments include:

- 1970s: IBM mainframes operated in controlled environments similar to early datacenters.
- 1990s: Rise of client-server computing required dedicated server rooms.
- 2000s-Present: Hyperscale datacenters by Google, Microsoft, and Amazon revolutionized networking, storage, and scalability.

### What's new in Datacenters?

Datacenters have been around for decades, but modern datacenters have undergone significant changes in scale, architecture, and service models. The primary **factors driving these changes** include:

- ✓ The exponential **growth of internet services** (Google, Facebook, Amazon, etc.).
- ✓ The **shift to cloud computing** and on-demand services.
- ✓ The need for **better network scalability, fault tolerance, and efficiency**.

One of the most striking changes in modern data centers is their massive scale:

- Companies like Google, Microsoft, Amazon, and Facebook operate **datacenters with over a million servers at a single site**.
- **Microsoft alone has more than 100,000 switches and routers** in some of its datacenters.
- **Google processes billions of queries per day**, requiring vast computational resources.
- **Facebook and Instagram serve billions of active users**, with every interaction generating requests to datacenters.

Another major change is the **shift from owning dedicated computing infrastructure** to **renting scalable cloud resources**. Datacenters no longer just host enterprise applications, **they now offer computing, storage, and network infrastructure as a service**. The most common cloud computing models are:

- **Infrastructure as a Service (IaaS)**. User rent virtual machines (VMs), storage, and networking instead of maintaining their own physical servers (e.g., Amazon EC2).
- **Platform as a Service (PaaS)**. Provides a platform with pre-configured environments for software development (databases, frameworks, etc.).
- **Software as a Service (SaaS)**. Full software applications hosted in datacenters and delivered via the internet (e.g., Google Drive).

The move to cloud computing has fundamentally changed datacenters, shifting the focus to resource allocation, security, and performance guarantees. They are also moving from multi-tenancy to single-tenancy:

- **Single-Tenancy**. A client gets **dedicated infrastructure** for their services.
- **Multi-Tenancy**. **Resources are shared among multiple clients while ensuring isolation**.

✗ **Implications**. But this massive scale brings new challenges:

- **Scalability**: The need for **efficient network designs** to handle rapid growth.

  Traditional datacenter topologies, such as three-based architectures, are inefficient at scale. New designs, like **Clos-based networks (Fat Tree)** and **Jellyfish (random graphs)**, are being developed to:

  ✓ Ensure **high bisection bandwidth** (allow any-to-any communication efficiently).

  ✓ Provide **scalable and fault-tolerant networking**.

- Cost management: More machines mean **higher power, cooling, and hardware costs**.

  Datacenters are **expensive to build and maintain**, requiring:

  - **Efficient resource utilization** (prevent idle servers from wasting power).
  - **Energy-efficient cooling solutions** (cooling accounts for a *huge* portion of operational costs).
  - **Automation to reduce human intervention** (e.g., AI-based network optimization).

- Reliability: Hardware failures become **common at scale**, requiring **automated fault-tolerant solutions**.

  At the scale of modern datacenters, **hardware and software failures are common**. A key principle is: "*In large-scale systems, failures are the norm rather than the exception.*" (Microsoft, ACM SIGCOMM 2015).

  Thus, new **automated failover mechanisms** are required to:

  - Detect failures **quickly**.
  - Redirect traffic **seamlessly**.
  - Ensure **minimal service disruption**.

- Performance & Isolation Guarantees: In modern datacenters, **customers expect strict performance guarantees** for applications like: low-latency financial transactions, high-bandwidth video streaming, machine learning model training.

  To meet these demands, datacenters implement:

  ✓ Performance Guarantees: Allocating bandwidth and compute power dynamically.

  ✓ Isolation Guarantees: Ensuring one user's workload does not interfere with another's.

  But this requires **advanced networking techniques**, such as:

  - **Traffic engineering** to avoid congestion.
  - **Load balancing** to distribute workloads efficiently.
  - **Software-defined networking (SDN)** for centralized control over traffic flows.

**Key Takeaways: What is a Datacenter?**

- **Datacenters centralize** computing resources for performance, security, and scalability.
- **They differ from traditional networks** by offering more control, lower latency, and higher redundancy.
- **Applications include cloud services, AI, and enterprise computing**.
- **Scalability is a key challenge**, with hyperscale datacenters hosting millions of machines.
- **Efficiency and cost containment are major concerns**, requiring innovative architectures.

## 1.2 Datacenter Applications

Modern datacenters host a variety of applications that range from web services to large-scale data processing. These **applications can be classified based on their traffic patterns and computational needs**.

### ? Customer-Facing Applications (North-South Traffic)

Customer-facing applications involve direct interaction with users. This type of traffic follows a **North-South communication model**, meaning that **data flows between external users and the datacenter**.

**Example 1: North-South Traffic**

Examples include:

- **Web Search** (e.g., Google, Bing)
  - A user submits a query (e.g., "Albert Einstein").
  - The request is routed through the datacenter's frontend servers.
  - Backend database and indexing servers fetch relevant results.
  - The response is assembled and sent back to the user.
- **Social Media Platforms** (e.g., Facebook, Instagram, X (ex Twitter))
  - Users interact with content hosted in the datacenter (e.g., loading a feed, liking posts).
  - Each interaction requires queries to databases and caching systems.
  - Content delivery is optimized using load balancers.
- **Cloud Services** (e.g., Google Drive, Dropbox, OneDrive)
  - Users upload, store, and retrieve files.
  - Requests must be efficiently distributed across storage nodes.

### ? Large-Scale Computation (East-West Traffic)

Unlike customer-facing applications, backend computations do not involve direct interaction with external users. Instead, they focus on **processing massive datasets within the datacenter**. This type of traffic is known as **East-West traffic** because it occurs **between servers inside the datacenter** *rather than between the datacenter and the external world.*

**Example 2: East-West Traffic**

Examples include:

- **Big Data Processing** (e.g., MapReduce, Hadoop, Spark)
  - Large datasets are distributed across multiple servers.
  - Each server processes a portion of the data in parallel.
  - Results are combined to generate insights (e.g., web indexing, analytics).
- **Machine Learning & AI Training** (e.g., Deep Learning Models)
  - AI models are trained on massive datasets using clusters of GPUs/TPUs.
  - The process requires high-bandwidth, low-latency communication.
  - Synchronization between nodes is critical (e.g., gradient updates in distributed training).
- **Distributed Storage & Backup Systems** (e.g., Google File System, Amazon S3)
  - Data is replicated across multiple locations for reliability.
  - Servers frequently exchange data to ensure consistency and fault tolerance.

### ⚖ Key differences between North-South and East-West traffic

| **Feature** | **N-S traffic** | **E-W traffic** |
|---|---|---|
| **Direction** | External users $\leftrightarrow$ Datacenter | Within datacenter |
| **Examples** | File downloads | AI training |
| **Bandwidth Needs** | Moderate | Very High |
| **Latency Sensitivity** | High | Critical |
| **Traffic Type** | Query-response | Bulk data transfer |

Table 2: Differences between North-South and East-West traffic.

In terms of latency sensitivity, North-South traffic is high because user interactions must be fast. On the other hand, East-West traffic is critical because synchronization delays affect computation.

## Traffic Patterns and Their Impact on Networking

The way data moves within a datacenter **heavily influences network design**. The main **goal** is to **ensure high bandwidth, low latency, and efficient resource utilization**.

- **Any-to-Any Communication Model**
  - In **large-scale distributed applications**, **any server should be able to communicate with any other server at full bandwidth**.
  - **Network congestion can severely degrade performance**, especially for AI/ML workloads and big data processing.
- **High-Bandwidth Requirements**
  - Applications like **MapReduce** and **deep learning** require **high data transfer rates**.
  - If bandwidth is insufficient, **bottlenecks occur**, leading to delays.
- **Latency is a Critical Factor**
  - **Low-latency networking is essential** for interactive applications and distributed computing.
  - AI training, for example, requires nodes to synchronize frequently; a delay in one node slows down the entire process.
- **Worst-Case (Tail) Latency Matters**
  - It's not enough for **most requests** to be fast; **the slowest request** can delay the entire computation.
  - **Minimizing tail latency** is crucial for efficient AI model training and database queries.

## Challenges in Datacenter Traffic Management

The massive scale and complexity of modern datacenters introduce **several networking challenges**, including:

- **Network Congestion and Bottlenecks**. When multiple servers communicate simultaneously, **some network links become overloaded**, leading to congestion.

  For example, if many AI training jobs share the same network path, it can become a bottleneck, slowing down training.

  This can be a **critical issue for applications requiring real-time performance** (e.g., financial transactions, cloud gaming).

- **Load Balancing and Traffic Engineering**. How do we distribute traffic *efficiently* across network links? The solutions are: **Equal-Cost Multipath Routing** (ECMP, spreads traffic across multiple paths); **Dynamic Traffic Engineering** (adjusts paths in real time based on congestion levels).

- **Avoiding Link Over-Subscription**. If too many servers send data over a single link, the available **bandwidth is divided**, leading to **slow performance**. Modern datacenters aim for **full-bisection bandwidth**, meaning **any server can talk to any other server at full capacity**.
- **Scaling Challenges**. Traditional datacenter network architectures do not scale well beyond a certain point. **New network topologies** (e.g., Fat Tree, Jellyfish) are being adopted to address these limitations.

**Key Takeaways: Datacenter Applications**

- Datacenters handle **two major types of applications**:
  1. **Customer-facing applications (North-South traffic)** involve external users.
  2. **Large-scale computations (East-West traffic)** occur within the datacenter.
- **Traffic patterns** affect **bandwidth, latency, and congestion control**.
- **Managing congestion and ensuring high bandwidth** is critical for performance.
- **New network topologies and routing techniques** help address scaling challenges.

## 1.3 Network Architecture

The **primary goal** of a datacenter network is to **interconnect thousands to millions of servers** efficiently. Unlike traditional networks, which focus on wide-area communication, datacenter networks emphasize:

- **High throughput**: Supporting massive data transfers.
- **Low latency**: Ensuring real-time performance for applications.
- **Scalability**: Accommodating rapid growth without performance degradation.
- **Fault tolerance**: Handling hardware failures with minimal disruption.

Datacenter **networks physically and logically connect servers through a multi-tiered architecture**. This hierarchical structure ensures that servers in different racks, pods, or clusters can communicate efficiently.

### Traditional Three-Tier Datacenter Network

Most datacenter networks follow a **Three-Tier design**, which is optimized for scalability and efficiency. The three tiers are:

- **Edge Layer (Access Layer)**
  - Located at the **bottom of the hierarchy**, closest to the servers.
  - Consists of **Top-of-Rack (ToR) switches** that connect servers within a rack.
  - ✓ Purpose: Aggregates traffic from multiple servers and forwards it to the higher layers.
  - Typically uses **high-speed links (10-100 Gbps per port)** to connect servers.
- **Aggregation Layer (Distribution Layer)**
  - Intermediate layer **between the edge and core layers**.
  - **Connects multiple ToR switches** within a datacenter pod.
  - ✓ Purpose: Helps distribute traffic efficiently **without overwhelming core routers**.
  - Implements **load balancing, redundancy, and failover mechanisms**.
- **Core Layer (Backbone Layer)**
  - The **top layer** of the hierarchy.
  - Composed of **high-capacity, high-speed switches and routers**.
  - ✓ Purpose: Responsible for:
    - **Routing large volumes of traffic** between different aggregation switches.

    * **Connecting the datacenter to external networks** (e.g., the Internet or private backbones).
  - Core switches often run **at 100 Gbps or higher per port** to support high aggregate bandwidth.

Key **characteristics of the Three-Tier model**:

- **Position**:
  - **Edge Layer**: Closest to servers.
  - **Aggregation Layer**: Intermediate between edge and core.
  - **Core Layer**: Backbone layer.
- **Primary Function**:
  - **Edge Layer**: Connects servers within racks.
  - **Aggregation Layer**: Aggregates ToR traffic.
  - **Core Layer**: Routes traffic between datacenters or externally.
- **Switch Type**:
  - **Edge Layer**: Top-of-Rack (ToR).
  - **Aggregation Layer**: Aggregation switches.
  - **Core Layer**: Core routers.
- **Speed (per port)**:
  - **Edge Layer**: 10-100 Gbps.
  - **Aggregation Layer**: 40-100 Gbps.
  - **Core Layer**: 100 and more Gbps.
- **Fault Tolerance**:
  - **Edge Layer**: Redundant paths to aggregation layer.
  - **Aggregation Layer**: Load balancing across core switches.
  - **Core Layer**: High redundancy & backup links.

## ⚠ Limitations of the Traditional Three-Based Model

Although widely used, the traditional three-tier model faces **scalability and performance challenges** as datacenters grow.

- **Scalability Issues**. **Traditional networks are hierarchical**, meaning most communication **must pass through the core layer**. As datacenters scale, **core switches become bottlenecks** due to increased traffic.
- **Bandwidth Bottlenecks**. The model assumes that the **most traffic is North-South** (client to server). However, modern workloads involve **high East-West traffic** (server-to-server communication).

  **Over-subscription occurs** when the network cannot handle full-bisection bandwidth.

- **Over-Subscription Problem**. **Over-Subscription** refers to **the ratio of worst-case achievable bandwidth to total bisection bandwidth**. For example:
  - If 40 servers per rack each have a 10 Gbps link, total demand is 400 Gbps.
  - If the uplink capacity to the aggregation layer is only 80 Gbps, we have a 5:1 over-subscription.
  - This means only 20% of the potential bandwidth is available, causing congestion.

  Over-subscription ratios in large-scale networks can reach 50:1 or even 500:1, **severely limiting performance**.

- **Performance Issues in High-Density Environments**. High latency when traffic must **traverse multiple hops** to reach other racks. **Failures in core routers** can impact a large number of servers. **Inconsistent network performance** due to congestion in aggregation switches.

### ✓ Modern Datacenter Network Designs

To overcome the **scalability and congestion challenges** of traditional three-based networks, modern datacenters use alternative architectures.

✓ **Fat Tree (Clos Network)**. Fat Tree is a **multi-stage switching architecture** designed to:
  - **Ensure full-bisection bandwidth**: Every server can communicate at full capacity.
  - **Provide multiple paths** between any two servers (high redundancy).
  - **Balance traffic dynamically** to avoid congestion.

  It uses K-ary fat tree topology where each pod consists of aggregation and edge switches, and core switches connect multiple pods. The advantages are:
  - **Scalability**: Expands easily by adding more pods.
  - **Fault Tolerance**: Multiple paths prevent failures from disrupting traffic.
  - **Better Load Balancing**: Traffic is evenly distributed.

✓ **Jellyfish: Random Graph-Based Topology**. Instead of a strict hierarchical structure, Jellyfish uses a **randomized topology**. The advantages are:
  - **Higher network capacity** with **lower cost**.
  - **More flexible scaling** than Fat Tree.
  - **Better fault tolerance** since the network adapts dynamically.

✓ **BCube: Datacenter Network for Cloud Computing**. Designed for high-performance cloud computing environments. It is optimized for: multi-path communication, resilience against failures and lowe latency compared to hierarchical models.

**Key Takeaways: Network Architecture**

- Traditional **three-tier datacenter networks** include **Edge, Aggregation, and Core layers**.
- **Core switches bottlenecks** as datacenters scale.
- **Over-subscription limits bandwidth**, causing congestion.
- **Modern topologies like Fat Tree and Jellyfish improve scalability, fault tolerance, and load balancing**.

## 1.4 High and Full-Bisection Bandwidth

**Why is High-Bandwidth important in Datacenters?**

Modern datacenters **handle massive amounts of data** due to applications like AI training, cloud services, and big data processing. These workloads *require*:

- **High-bandwidth connections** to support fast data transfers.
- **Low latency** to ensure real-time performance.
- **Scalability** to accommodate increasing workloads.

Unlike traditional networks, where traffic primarily flows between users and servers (North-South), **datacenters experience heavy East-West traffic** (server-to-server communication). This shift **demands high-bandwidth and scalable network designs**.

***One step at a time*: What a Bisection Bandwidth is and why Full-Bisection Bandwidth is important**

**Bisection Bandwidth** is a key metric that measures the **total bandwidth available between two halves of a network**.

**Definition 1: Bisection Bandwidth**

If a network is split into two equal halves, the **Bisection Bandwidth** is the **total data transfer rate available between them**.

**Definition 2: Full-Bisection Bandwidth**

The **Full-Bisection Bandwidth** is when every server can communicate with every other server at **full network speed**.

In other words, bisection bandwidth can be thought of as cutting a data center network in half and measuring the total capacity of the links connecting the two halves. This tells us how much data can flow between the two sections simultaneously.

**Example 3: Understand what bisection bandwidth is**

Imagine a 1000-server datacenter, where 500 servers are processing data while 500 servers store the results. If the bisection bandwidth is **low**, the **data transfer between processing and storage nodes will be delayed**. This results in slow machine learning model training or delayed database queries.

As we can imagine, the full-bisection bandwidth is a real and critical aspect:

- **Prevents bottlenecks**: Ensures high-throughput communication across racks and clusters.
- **Essential for AI/ML training**: AI models require massive parallel computations with continuous data exchanges.

- **Optimized for cloud computing**: Services like AWS, Google Cloud, and Azure depend on fast, reliable inter-server communication.

**⚠ Then try to get high-bandwidth all the time! Yes, but there are some challenges...**

Ideally, high-bandwidth should be the ultimate goal, but unfortunately, there are some problems with traditional three-based networks:

✗ **The Problem with Traditional Three-Based Networks**. The standard **three-tier (core-aggregation-edge) topology** struggles to scale due to:

1. **Over-subscription** (definition on page 16): The ratio of available bandwidth to required bandwidth is too high.
2. **Core congestion**: Core routers become bottlenecks as traffic grows.
3. **Single points of failure**: A failure in a core switch can affect a large portion of the datacenter.

✗ **Over-Subscription and Its Impact on Network Performance**. A naive solution would be to use over-subscription to solve these problems, but this limits performance. **Over-Subscription** happens when the **network is provisioned with less bandwidth than needed** to cut costs.

$$\text{Over-subscription} = \frac{\text{Total server bandwidth demand}}{\text{Available bandwidth at aggregation/core layer}}$$

Common over-subscription ratios are:

- 5:1, only 20% of host bandwidth is available.
- 50:1, only 2% of host bandwidth is available.
- 500:1, only 0.5% of host bandwidth is available.

At 500:1 over subscription, congestion becomes severe, **limiting network efficiency**.

✗ **The cost problem: scaling is expensive!**

- Increasing bisection bandwidth requires **more high-performance network hardware**.
- **Scaling traditional networks** (adding more core switches) is extremely costly.
- **Energy consumption rises** with additional hardware.

Thus, **alternative solutions** are needed to achieve high-bandwidth networking **without excessive costs**.

### ✓ Solutions to Achieve High and Full-Bisection Bandwidth

To overcome these challenges, researchers and engineers have designed **new network architectures**.

- ✓ **Fat Tree (Clos Network) - The Scalable Solution**. Unlike traditional three-based designs, Fat Tree provides **multiple paths** for traffic.

  ✓ **Advantages**

  - ✓ **Ensure full-bisection bandwidth** by allowing traffic to take alternative routes.
  - ✓ **Eliminates single points of failure** using redundant paths.
  - ✓ **Load balancing** optimizes network utilization.

- ✓ **Jellyfish - A More Flexible Approach**. Uses a **randomized, non-hierarchical** topology instead of a fixed three structure.

  ✓ **Advantages**

  - ✓ **Better bandwidth scaling** as new servers are added.
  - ✓ **More resilient to failures** (no single critical point of failure).

- ✓ **BCube - Optimized for Cloud Services**. Designed for high-performance cloud environments with **massive inter-server communication**.

  ✓ **Advantages**

  - ✓ **Fast re-routing** in case of failures.
  - ✓ **Low-latency communication for cloud applications**.

**Key Takeaways: High and Full-Bisection Bandwidth**

- **High-bandwidth networking** is essential for modern datacenters.
- **Full-bisection bandwidth** ensures servers communicate at **full speed**.
- **Over-subscription** creates **bottlenecks**, limiting performance.
- **New network architectures** (Fat Tree, Jellyfish, BCube) solve scalability issues.

## 1.5 Fat-Tree Network Architecture

A **Fat-Tree** is a **multi-layer, hierarchical network topology** that provides *high scalability*, *full-bisection bandwidth*, and *fault tolerance*. It is a **special type of Clos Network**[1], designed to **overcome bandwidth bottlenecks** in traditional three-based networks.

The key idea is: Instead of a traditional tree where higher levels become bottlenecks, Fat-Tree ensures equal bandwidth at every layer by **increasing the number of links as we move higher in the hierarchy**.

### Structure of a K-Ary Fat-Tree

A **K-ary Fat-Tree** consists of **three layers**:

1. **Edge Layer (Top-of-Rack, ToR switches)**:
   - Connects directly to the servers.
   - Each edge switch connects $\frac{k}{2}$ servers and $\frac{k}{2}$ aggregation switches.
2. **Aggregation Layer**
   - Connects multiple edge switches.
   - Ensures **local traffic routing** between racks before sending to the core.
   - Each aggregation switch connects $\frac{k}{2}$ edge switches and $\frac{k}{2}$ core switches.
3. **Core Layer**
   - The backbone of the Fat-Tree, interconnecting multiple aggregation layers.
   - Consists of $\left(\frac{k}{2}\right)^2$ core switches, where each connects to $k$ pods.

**Example 4: Fat-Tree with $k = 4$**

- Each pod contains:
  - $\left(\frac{4}{2}\right)^2 = 4$ servers.
  - 2 layers of 2 2-port switches (Edge and Aggregation).
- Each Edge Switch connects 2 servers and 2 aggregation switches.
- Each Aggregation Switch connects 2 Edge switches and 2 Core switches.
- The Core Layer consists of $\left(\frac{k}{2}\right)^2 = 4$ core switches.

[1] A **Clos Network** is a type of multistage **switching topology that enables high-bandwidth and fault-tolerant communication by interconnecting multiple small switches instead of relying on a few large ones**. It is commonly used in datacenter networks (e.g., Google Jupiter Fabric) to maximize scalability and minimize congestion.

As a result, multiple paths between servers ensure no single point of failure and full-bisection bandwidth.

## Why Use Fat-Tree in Datacenters?

### ✓ Cost-Effective Scaling

- Can be built using **cheap, commodity switches** instead of expensive core routers.
- All switches operate at **uniform capacity**, simplifying hardware requirements.

### ✓ Full-Bisection Bandwidth

- Each switch and server has **equal access to bandwidth**, preventing bottlenecks.
- Every packet has **multiple available paths**, ensuring **load balancing**.

### ✓ High Fault Tolerance

- If one **switch or link fails**, traffic is rerouted through **alternative paths**.
- **No single point of failure**, unlike traditional three-based architectures.

### ✓ Efficient Load Balancing

- **Multipath Routing** ensures traffic is evenly distributed.
- **No congestion at higher layers**, as each pod has equal bandwidth allocation.

## Problems in Fat-Tree Networks

Fat-Tree is a highly scalable and efficient network topology, but **practical challenges exist** when handling real-world workloads.

- **Many flows running simultaneously**. In large datacenters, multiple applications generate concurrent flows. Some flows are **small but latency-sensitive** (mice flows), while others are **large data transfers** (elephant flows). The Fat-Tree must **efficiently balance all these flows** across available paths.
- **Traffic locality is unpredictable**. Some services (e.g., Facebook/Meta workloads) have localized communication within a rack, while others require data exchange across the entire network. Fat-Tree must **dynamically adapt to different workload patterns**.

- **Traffic is bursty**. Some applications **generate sudden traffic spikes**, leading to temporary congestion. This is problematic for routing since **congestion-aware path selection is difficult**.
- **Too Many Paths Between a Source and Destination**. Unlike traditional network that have a single best route, Fat-Tree networks offer multiple equal-cost paths. *Which path should be used?* Random selection might lead to congestion.
- **Random Path Selection Leads to Collisions**. If routing randomly assigns traffic flows, two large elephant flows may end up on the same link. This creates a congestion hotspot, even though other links remain underutilized.
  - **Ideal case**: Traffic should be spread evenly across all available links.
  - **Reality**: Without congestion awareness, routing **cannot react to traffic conditions dynamically**.
- **Short-Lived vs. Long-Lived Flows Create Conflicts**. An **ideal routing scenario** would be to evenly distribute all flows. However, if a short, latency-sensitive flow suddenly appears on a congested link, its performance suffers. The key problem is that **Fat-Tree does not inherently prioritize latency-sensitive flows**.

## ⚠ TCP Incast: A Major Issue in Fat-Tree Datacenters

Large-scale parallel requests cause network congestion. In fact, some workloads (e.g., distributed storage systems, AI training) involve a **single client requesting data from multiple servers simultaneously**. This means that all servers respond at once, **overwhelming the switch's buffer capacity**. This results in **packet loss and retransmissions**, significantly increasing latency.

**Definition 3: TCP Incast**

**TCP Incast** is a **network congestion issue** that occurs in datacenters when multiple servers send data to a single receiver simultaneously, overwhelming the switch's buffer capacity and causing severe packet loss and performance degradation.
In other words, TCP Incast happens when many-to-one communication causes network congestion, leading to packet loss, TCP retransmissions, and increased latency.

But in this scenario, how does **TCP Incast** happen?

1. A client application requests data from multiple storage servers.
2. All storage servers respond **simultaneously**.
3. The switch **cannot handle all packets at once**, causing **buffer overflow**.

4. **Packet loss triggers TCP retransmissions**, further slowing down performance.

This involves several issues:

- Causes **severe latency spikes**, affecting (AI training and large-scale cloud) workloads.
- Traditional TCP **was not designed for this kind of bursty traffic**.
- **Fat-Tree cannot solve this issue alone**, it requires transport-layer optimizations.

## ✓ Google's Approach to Solving Fat-Tree Challenges

Google faced severe scalability, congestion, and failure recovery challenges in its datacenters. Instead of using a traditional Fat-Tree model, they **developed a Clos-based architecture** known as **Google Jupiter Fabric**. The key challenges that Google is addressing are:

- **Scalability**. Traditional networks could not handle Google's exponential growth. Needed a network that scales gracefully by adding more capacity in stages.
- **Failure Tolerance**. A single failure should not impact traffic significantly. Needed path redundancy to ensure seamless operations.
- **Performance and Cost**. High-performance custom-built switching to support full-bisection bandwidth. Used commodity merchant silicon (off-to-shelf networking chips) instead of proprietary network devices, reducing costs.

The solutions adopted by Google are:

✓ **Clos Topology for Scalability & Fault Tolerance**. Google moved from traditional Fat-Tree to Clos networks to improve scalability.

- **Multiple layers of switches**, with multiple paths between every two endpoints.
- **Graceful fault recovery**: if one switch fails, traffic is rerouted dynamically.
- **Incremental scalability**: new switching stages can be added without network downtime.

A Clos network was chosen because, unlike Fat-Tree, which suffers from static oversubscription, **Clos networks offer more flexible bandwidth allocation**.

Note that Fat-Tree inherits the scalability and fault tolerance of Clos, but its hierarchical and structured nature leads to congestion, routing complexity, and TCP Incast problems. Google recognized that Fat-Tree had structural limitations, so they modified Clos into the Jupiter Fabric.

✓ **Custom Hardware: Merchant Silicon Instead of Proprietary Switches**. Google avoided vendor lock-in by using commodity hardware (merchant silicon). The reasons are:

- **Lower cost** than custom ASIC-based routers.
- **Faster** hardware **upgrade cycles**.
- **More control** over network design and software stack.

✓ **Centralized Control for Routing and Network Management**. In traditional datacenters, routing is distributed, meaning each switch makes independent routing decisions. This approach does not scale well in Clos networks with thousands of switches.

The solution is **precomputed routing decisions**. Instead of switches making their own decisions, Google precomputes traffic flows centrally and pushes them to switches.

✓ **Advantages**

✓ **Improves traffic engineering**: Load balancing decisions are optimized globally rather than per switch.

✓ **More predictable performance**.

✓ **Less congestion**: Can react dynamically to network failures.

**Key Takeaways: Fat-Tree Network Architecture**

- **Fat-Tree is a special type of Clos Network** that overcomes bottlenecks in traditional tree networks.
- **K-ary** Fat-Tree has three layers (Edge, Aggregation, Core), **ensuring equal bandwidth for all nodes**.
- **Fat-Tree** provides multiple paths, but **routing is difficult due to unpredictable traffic patterns**.
- **Collisions between large flows create network hotspots**.
- **TCP Incast is a major issue**, where too many responses at once cause packet loss.
- Google's Datacenter Network Strategy:
  - **Moved from Fat-Tree to Clos topology** for better scalability and failure recovery.
  - **Used merchant silicon instead of proprietary hardware** to cut costs and improve flexibility.
  - **Implemented centralized control for routing** to optimize traffic flows.
  - **Designed the Jupiter Fabric** to handle **Google-scale workloads** with incremental scalability.

# 2 Software Defined Networking (SDN)

## 2.1 Introduction

**Software-Defined Networking (SDN)** is an **architectural shift** in networking that **separates** the **control plane** from the **data plane**, allowing for centralized control and programmability. Unlike traditional networks, where control logic is embedded in individual devices, **SDN introduces a centralized software controller that dynamically manages the entire network**.

The importance of SDN lies in its ability to:

- Improve **network flexibility** by enabling real-time changes.
- Simplify **network management** through automation.
- Reduce **hardware dependency** by allowing software-driven policies.
- Support **rapid innovation**, making networks more adaptable.

### ⚖ Traditional Networking vs. SDN

In ***traditional networking***, routers and switches **contain both**:

- **Control Plane**, decides how traffic should be forwarded (e.g., routing decisions, firewall rules).
- **Data Plane**, physically forwards packets based on the control plane's decisions.

Challenges of traditional networking:

a. ✗ **Rigid Configuration**: Any changes require manual updates on multiple devices.

b. ✗ **Vendor Lock-in**: Hardware manufacturers impose proprietary limitations.

c. ✗ **Slow Innovation**: Implementing new networking features takes years due to hardware constraints.

d. ✗ **Complex Management**: Network engineers must configure each device individually.

How ***SDN differs***:

a. ✓ **Control Plane is centralized** in an SDN controller.

b. ✓ **Data Plane remains distributed** across switches and routers.

c. ✓ **Network logic is programmable**, making updates and changes easier.

This shift makes networks more dynamic, scalable, and easier to manage.

## 2.2 Legacy Router & Switch Architecture

Legacy network devices, such as routers and switches, are **built with integrated control and data planes**, meaning each device independently makes forwarding decisions. These devices consist of:

1. **Hardware Components**
   - **Application-Specific Integrated Circuits (ASICs)**, specialized chips for packet forwarding.
   - **Memory (buffers & TCAMs)**, stores forwarding tables and processing queues.
   - **Network Interfaces (NICs, Ports)**, physical ports for connecting network cables.
2. **Software Components**
   - **Router OS (Operating System)**, runs network protocols and management interfaces.
   - **Routing Protocols (OSPF, BGP, RIP)**, determines paths for packet forwarding.
   - **Forwarding Table**, maps destination addresses to outgoing ports.
3. **Management and Control Interfaces**
   - **Command-Line Interface (CLI)**, used for configuring routers manually.
   - **SNMP (Simple Network Management Protocol)**, enables monitoring and automation.

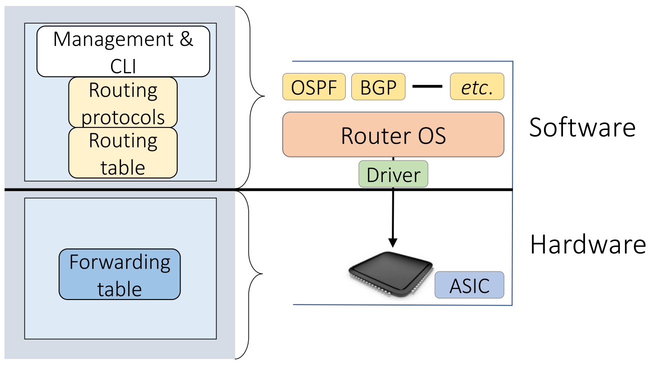


Figure 1: Legacy router and switch architecture.

Traditional network devices have two primary operational planes:

- **Control Plane**: makes forwarding decisions based on routing protocols.
- **Data Plane**: physically forwards packets based on control plane decisions. For example MAC lookup and IP forwarding.

Each router operates autonomously, using routing tables built through protocols like OSPF and BGP. These protocols dynamically learn network paths and update the forwarding tables, ensuring efficient packet delivery.

### Packet Processing in a Legacy Router

1. **Lookup Destination IP** → Find matching entry in the forwarding table.
2. **Update Header** → Modify packet headers if needed (e.g., TTL decrement).
3. **Queue Packet** → Send packet to the appropriate output interface.

Since **every device handles its own control and forwarding**, **large-scale changes require individual device updates**, making traditional networking complex and inflexible.

### Challenges in Traditional Network Management

- ✗ **Complex Configuration & Management**. **Each network device** has to be **configured individually**. Protocols like BGP and OSPF require manual tuning for optimal performance. Network engineers must interact with vendor-specific CLIs, which vary by manufacturer.
- ✗ **Limited Innovation & Vendor Lock-In**. New network **features** require **firmware or software updates from vendors**. Custom networking solutions are **difficult to implement due to proprietary hardware and software**.
- ✗ **Slow Response to Failures & Traffic Changes**. Routing adjustments depend on distributed algorithms that can take seconds to minutes to converge. Manual troubleshooting is often needed when failures occur.
- ✗ **Scalability Issues**. Growing networks require more hardware and manual configurations. **Updating** policies across multiple routers is **time-consuming and error-prone**.

**Key Takeaways: Legacy Router and switch architecture**

- Legacy networking relies on **autonomous devices** with tightly integrated **control and data planes**.
- Routing is handled by protocols like OSPF and BGP, which operate **independently on each device**.
- Challenges include manual configuration, vendor lock-in, slow failure response, and scalability issues.
- These limitations paved the way for SDN, which offers centralized, programmable networking.

## 2.3 SDN Architecture

The **core** concept of **Software-Defined Networking (SDN)** is the **separation** of the **control plane** from the **data plane**:

- In SDN, **network devices** (switches/routers) become **simple forwarding elements**, executing decisions made by a centralized **controller**.
- The **SDN Controller** is a **software-based system** that manages, programs, and monitors the entire network.

Key architecture:

- **Data Plane (Forwarding Engine)** → Located on switches; handles packet forwarding.
- **Control Plane (SDN Controller)** → Runs on external servers; computes forwarding rules.
- **Communication Channel** → Allows the controller to instruct the data plane; typically uses OpenFlow.

But *why decouple?* Enables centralized decision-making, consistent policy enforcement, and simplified management. Facilitates dynamic updates to the network without hardware changes.

### The Role o the SDN Controller

The SDN Controller is the **central brain** of the network. It performs:

- ✓ **Network State Monitoring**: Gathers real-time information from all forwarding devices.
- ✓ **Decision-Making**: Calculates the best routes, applies policies, and enforces security.
- ✓ **Rule Installation**: Pushes flow rules to switches, determining how packets should be handled.

Controllers provide a **global, up-to-date view of the entire network**, enabling smarter control than traditional distributed routing.

### Communication Interfaces

SDN uses two types of APIs to manage communication between layers:

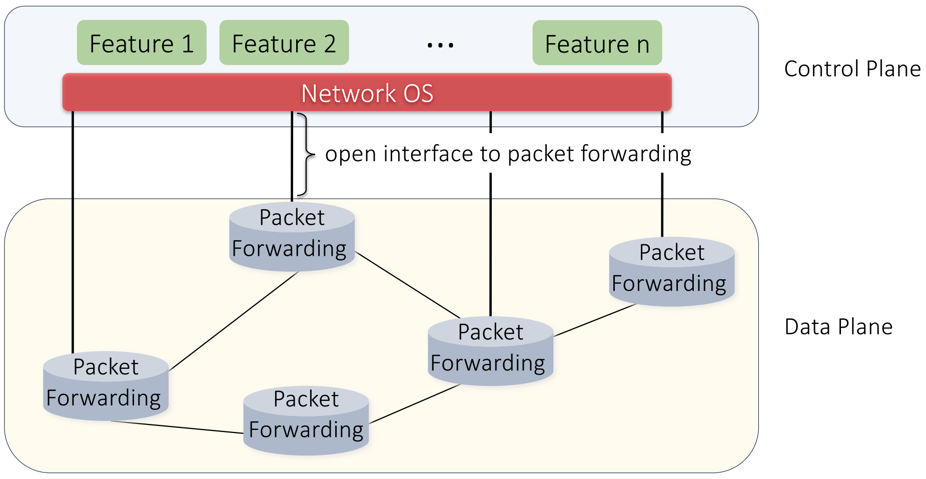
- **Southbound Interface**: Connects the controller to the data plane devices (e.g., **OpenFlow protocol**). It instructs switches via OpenFlow or similar protocols, installing/removing flow rules and collecting stats.
- **Northbound Interface**: Allows **applications to interact with the controller** via APIs (e.g., REST APIs). Applications (e.g., security monitoring, load balancing) query and command the controller to implement network policies.

### Network Operating System (Network OS)

The controller runs a Network OS, providing:

- **Abstractions** over the physical network (e.g., topology view, link status).
- **Programmatic Interfaces** for developing control programs.
- **Consistency & Global View**: All decisions are made based on coherent, synchronized data.

The Network OS simplifies the task of writing network control logic by exposing standardized APIs.



**Key Takeaways: SDN Architecture**

- **Traditional networking** embeds the control plane within each device; SDN **centralizes control** in software.
- The **SDN Controller** dynamically manages the **data plane devices** using a **communication protocol**.
- **OpenFlow** is the primary protocol used to communicate between the controller and switches.
- **Network OS** provides an **abstraction layer** and **programming environment** for writing control logic.

## 2.4 OpenFlow

**OpenFlow** is the **first and most widely adopted protocol** used in Software-Defined Networking (SDN) to **enable communication** between the **SDN Controller** (control plane) and the **data plane devices** (e.g., switches, routers, they are the forwarding engine). It allows the controller to **program flow tables** in the switches and **control how packets are forwarded**, enabling centralized management of traffic.

In other words, OpenFlow is the practical implementation of SDN, standardizing how controllers manage packet forwarding.

### How OpenFlow works

Each **OpenFlow switch** contains:

1. **Flow Table**: Contains rules in the form of $Match \rightarrow Action$ pairs.
2. **Communication Interface**: Connects to the SDN controller via the OpenFlow protocol.
3. **Stats Module**: Collects statistics about packet flows.

**Example 1: Flow Rules**

1. If $Header = p \Rightarrow$ send to port 5.
2. If $Header = q \Rightarrow$ modify header to $r$, then send to ports 6 and 7.
3. If $Header = p \Rightarrow$ send packet to the controller.

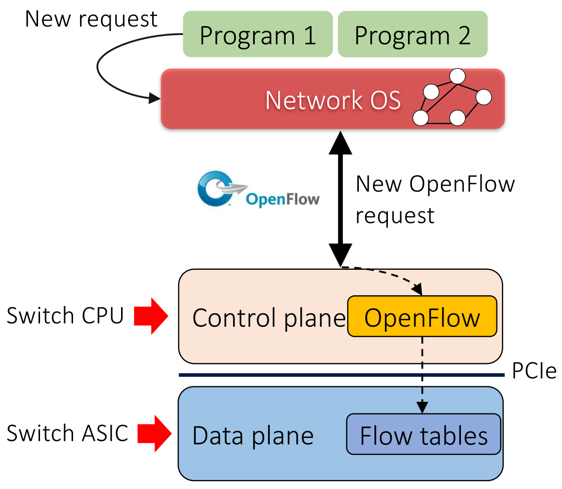
**Flow table operation**:

1. Packet arrives at the switch.
2. Switch checks for a **matching rule** in its flow table.
3. If matched $\rightarrow$ **apply action** (e.g., forward, modify, drop).
4. If no match $\rightarrow$ **send packet header to controller** for instructions.

**Example 2: OpenFlow**

1. New packet arrives at switch.
2. Match?
   - Yes $\rightarrow$ forward according to rule.
   - No $\rightarrow$ forward header to controller.
3. Controller analyzes packet and installs a new rule in switch.

4. Next packets of same type → directly processed by switch using newly installed rule.



### Actions in OpenFlow

OpenFlow supports many types of actions, such as:

- **Forwarding** to one or multiple ports.
- **Dropping packets**.
- **Modifying headers** (e.g., VLAN tags, IP addresses).
- **Sending packets to controller**.
- **Statistics collection** for flow monitoring.

This **flexibility** allows SDN to implement advanced functions like load balancing, traffic shaping, and security filtering without specialized hardware.

### Reactive vs. Proactive Flow Rules

In OpenFlow, the **controller installs flow rules** into switches to determine how packets are processed. There are **two main modes** of operation for rule installation: Reactive and Proactive. These **modes define when and how flow entries are populated in the flow tables** of switches.

- Reactive Mode

  **How it works?**

  1. When a **new flow** (a new type of packet) arrives at a switch, and **no rule matches** it, the **packet header is sent to the controller**.

2. The **controller analyzes the packet** and **decides what rule** should be installed in the switch to handle it.
3. After the controller sends back a rule, the switch installs it and **forwards the packet** accordingly.

Key Characteristics

* **Efficient use of flow tables** - Only rules for active flows are installed.
* **Every new flow** incurs **small setup time** (controller interaction delay).
* **Switch depends on the controller** for flow rule installation.
* If the **controller connection is lost**, the switch has **limited utility** for new flows.

Advantages

✓ **Dynamic** and **adaptive** to real-time network traffic.

✓ **Minimizes unused flow entries**.

Disadvantages

✗ Adds **latency** for the **first packet** of each flow.

✗ **High control plane load** in environments with many short flows.

- Proactive Mode

How it works?

1. The **controller pre-installs flow rules** in the switches before any packet arrives.
2. The switch **immediately processes packets** using pre-defined rules without contacting the controller.

Key Characteristics

* **Zero setup delay** for packet processing - packets are forwarded **immediately**.
* Requires **aggregated or wildcard rules** to efficiently use flow table space.
* **Independent of controller connectivity** - continues to operate even if controller is unreachable.

Advantages

✓ **Fast packet forwarding** with **no initial delay**.

✓ **No dependency** on controller for flow rule installation during packet arrival.

✓ Ideal for **predictable traffic patterns** or **mission-critical environments**.

Disadvantages

✗ Can **waste flow table space** if many pre-installed rules are unused.

✗ Requires **good planning** of rules; less flexible to dynamic traffic changes.

In summary, reactive mode is adaptive, but introduces latency and higher controller load; proactive mode is fast and resilient, but requires advance planning of rules.

**Key Takeaways**

- **OpenFlow** enables the SDN controller to manage **flow tables** in switches.
- **Flow rules** define how packets are handled, allowing **centralized, programmable networking**.
- OpenFlow supports **fine-grained traffic control** via a wide range of **match/action rules**.
- Two operation modes:
  - Reactive (dynamic but with latency).
  - Proactive (fast but needs good planning).

## 2.5 OpenFlow limitations

OpenFlow, conceived around 2007, introduced centralized control by standardizing how switches expose forwarding behavior to an SDN controller (as we discussed in the previous section, page 31). The **insight** at that time was that **most switches perform similar tasks** (Ethernet switching, IPv4 routing, VLAN tagging, ACL enforcement) **all via fixed**, **predictable behaviors**.

OpenFlow capitalized on this fixed-function approach. Controllers could install flow rules into switches, dictating how they process known packet headers. However, a **critical limitation** emerged: **we couldn't add new protocols or processing capabilities easily**. Because **OpenFlow assumes a static data plane**, **hardcoded to process only a predefined set of protocols and headers**.

### ✓ Expanding OpenFlow: pushing its limits

As networking needs evolved, particularly in **virtualized environments** and **cloud datacenters**, operators needed more **specialized packet processing**. For example, `VXLAN`, used to identify tenants in multi-tenant environments, wasn't supported in early OpenFlow.

✓ To address this, vendors and the OpenFlow community developed **new versions** (`1.1`, `1.2`, `1.3`, . . . ). Each iteration **added support for more header types**, up to 50 different header types, but **the process was slow and cumbersome**. Each new feature needed:

- New OpenFlow specification extensions.
- New ASICs in hardware to support the processing logic.

### ⚠ Hardware Bottlenecks: The ASIC Development Bottleneck

Here lies the **core problem**: even with updated protocols, **switches couldn't adapt until vendors redesigned and shipped new ASICs** (Application-Specific Integrated Circuits).

This hardware dependency meant:

- ✗ New features took **years to reach production**.
- ✗ Network owners **couldn't simply get a software upgrade**.
- ✗ The result: **slow innovation** in data plane capabilities.

**Example 3: `VXLAN`**

Virtual Extensible LAN (`VXLAN`) was urgently needed by cloud providers and datacenters to enable multi-tenant network virtualization.
Despite this high demand, hardware vendors took $\approx 4$ years to support `VXLAN` in switches due to ASIC development cycles and the fixed-function nature of OpenFlow switches.
Even though vendors delayed its release, once `VXLAN` was available, it became a standard requirement in data centers.
But attention! In the meantime, network operators used complex software overlays or kludges to simulate `VXLAN` functionality, increasing network complexity and cost.

### The Cost of Delay: Workarounds and Complexity

When vendors take years to deliver a new feature, network engineers often **develop complex workarounds**, **increasing network complexity** and **technical debt**. Even when the vendor releases the official feature:

- The workaround may already be deeply integrated.
- The official solution may no longer solve the problem.
- Worse, it may require a forklift upgrade, replacing hardware at high cost.

This inertia locks networks into **suboptimal solutions** and **impedes the agility promised by SDN**.

### The Missing Ingredient: Programmability at the Data Plane

The shift from fixed-function to programmable data planes mirrors other computing domains:

| **Domain** | **Hardware** | **Compiler/SW Stack** |
|---|---|---|
| General Computing | CPU | Java, C, OS Kernels |
| Graphics | GPU | OpenCL, CUDA |
| Signal Processing | DSP | Matlab Compiler |
| Machine Learning | TPU | TensorFlow Compiler |
| **Networking** | **PISA Switch** | **P4 Language, P4 Compiler** |

Just as CPUs became programmable via compilers, **networking needs** flexible **data planes programmable via languages** like P4, running on PISA (Protocol Independent Switch Architecture).

**Key Takeaways: OpenFlow limitations**

- OpenFlow was a **revolution in control plane innovation**, but its **rigid data plane** became a bottleneck. The industry's response, iterative protocol updates and ASIC redesigns, proved **slow and reactive**.
- A true solution lies in programmable data planes, where **software defines packet processing**, and the network evolves **as fast as the application demands**.
- This transition is **not trivial**, it requires new hardware, new abstractions, and operator retraining, but it's essential to **fulfill SDN's promise** of **rapid, flexible, and scalable networking**.

# 3 Programmable Switches

## 3.1 Introduction

In the past, **network switches were designed with fixed-function pipelines**. These switches could process packets extremely fast, but their internal logic was essentially "hardcoded" by hardware vendors. This meant that the functionality they provided, things like Ethernet switching, IP routing, and basic ACLs, was rigid and **difficult to extend or modify**.

However, as networks evolved and application demands grew more complex, the limitations of these fixed-function switches became apparent. There was a **growing need for flexibility at the data plane**, the part of the switch responsible for real-time packet processing. Network operators started to ask: *what if we could* program *the switch behavior instead of relying on vendors to update the hardware every time we needed new features?* This is where the concept of **programmable switches comes into play**.

### ❓ Why Programmability?

The **motivation** behind programmable switches stems from the **increasing complexity and dynamism of modern networks**. Today's infrastructures must support custom protocols for emerging technologies like IoT, 5G, and machine learning. They must also be able to adapt quickly to changing requirements, detect and mitigate threats in real-time, and perform network telemetry and monitoring with high granularity.

With **traditional switches**, making such changes **often meant waiting months** (or even years) for new hardware to be designed and released. In contrast, **programmable switches allow network behavior to be redefined using software**, even after deployment. This ability to program the forwarding logic gives networks a software-like agility that was previously unthinkable at the data plane level.

### ⚖ Control Plane vs Data Plane

To understand the significance of programmable switches, it's useful to recall the basic architecture of a network device. Typically, a **switch is divided into two major components**:

- The **Control Plane**, which is **responsible for**:
  - Computing routing tables;
  - Handling management tasks;
  - Making decisions about where traffic should go.
- The **Data Plane**, which is **responsible for**:
  - **Forwarding packets** at line rate, based on the decisions made by the control plane.

Traditionally, most of the innovation in networking happened in the control plane, for example, with Software-Defined Networking (SDN), which centralized and virtualized control logic (section 2, page 26). But the data plane remained fixed and closed.

**Programmable switches** shift this dynamic. They open up the data plane to innovation, **allowing developers to express forwarding behavior in a high-level language such as P4**. This means we can now rethink how packets are processed inside the switch itself.

### The Rise of PISA

A key enabler of this shift is the **Protocol-Independent Switch Architecture (PISA)**. Proposed by Barefoot Networks (later acquired by Intel), **PISA is a flexible hardware architecture that allows the structure of the switch pipeline to be configured by software**. Using PISA, one can **define new packet formats, parsing rules, match-action logic, and even custom metadata fields**, all using a high-level language like P4.

With PISA-based switches, it is no longer necessary to hardcode support for every protocol in silicon. Instead, **developers can define how packets are handled at runtime**. This brings about a level of protocol independence and reconfigurability that was previously reserved for general-purpose processors, but with the performance and parallelism needed to operate at terabit speeds.

## 3.2 Why didn't programmable switches exist before?

In short, **programmable switches didn't make sense before** because we lacked the technical feasibility and practical justification. But now, due to advances in chip design and network complexity, it's finally possible, and necessary, to build them.

### $ In the past: Programmability was too expensive

In the past, the trade-off between programmability and cost was too high:

1. **Performance was too low**. Programmable hardware, like FPGAs or general-purpose CPUs, was much slower than fixed-function ASICs.

   ✓ A **fixed switch chip** could forward billions of packets per second.

   ✗ A **programmable one**? Too slow for line-rate performance.

   So **if we wanted programmability, we had to sacrifice speed**. That was a deal-breaker for core network equipment.

2. **Chip area and power cost were too high**. Fixed-function logic is compact and power-efficient. Programmable logic, by contrast, used to **take up more silicon and required more power**. Result: vendors and data center operators couldn't justify using programmable switches, they were too big, too hot, and too slow.

### ✓ What changed?

**Three technological trends** made programmable switches finally viable:

- **Chip speed caught up**. We now have programmable switch chips (like Barefoot Tofino) that can **run at line rate**, just like fixed-function ones. In other words, programmability no longer costs us speed.

- ⚠ **Network complexity exploded**. There are now **too many protocols and features** to hard-code everything into silicon:

  - New protocols, encapsulations (VXLAN, GTP, QUIC, etc.)
  - Monitoring, load balancing, AI, security; all need custom, real-time logic.

  Hard-coding all of this would take years, and would never be flexible enough.

- ✓ **Moore's Law made logic "free"**. Thanks to Moore's Law:

  - We can double the amount of logic in the same area every 2 years.
  - The **cost** of programmability in terms **of chip area** and **power** has become **negligible**.

  Now, the logic that makes a switch programmable barely takes up more space than a fixed-function design.

| **Factor** | **Before** | **Now** |
|---|---|---|
| Chip speed | Too slow for line-rate | Equal to fixed-function |
| Logic cost | Too expensive (area + power) | Basically free |
| Protocols | Few, stable | Too many to hard-code |
| Urgency | Low | High (cloud, IoT, 5G, ML) |

Table 3: Why didn't programmable switches exist before?

## 3.3 Data Plane Programming and P4

**Traditionally**, configuring a switch meant **writing static forwarding rules**, usually via vendor-specific commands or protocols like OpenFlow. But this was **not true programmability**. We could **configure behavior**, but we **couldn't change how the switch processes packets internally**. With P4 (Programming Protocol-independent Packet Processors), that changes.

### What is P4?

**P4 (Programming Protocol-independent Packet Processors)** is a **high-level**, **domain-specific programming language** designed to describe how packets should be processed by the data plane of a network device.

Unlike general-purpose languages like C or Python, P4 is not Turing-complete. Instead, it is built to:

- Define **how to parse packet headers**
- Specify **how to match on those headers**
- Decide **what actions to take**

The **goal of P4 is to describe the behavior of the switch pipeline**, not to implement general algorithms. Specifically, P4 was designed with four main goals in mind:

1. **Reconfigurability**: We should be able to change switch behavior *after* deployment.
2. **Protocol Independence**: The switch should not be tied to Ethernet/IP/TCP. We define the packet format.
3. **Target Independence**: The same P4 program should run on different hardware (ASICs, FPGAs, software switches).
4. **Flexibility** and **Abstraction**: Developers write in P4, and the compiler maps it to the switch's low-level pipeline architecture.

### P4 is so cool, but OpenFlow is not the same?

We already discussed what OpenFlow is in Section 2.4, page 31. The short answer is no, P4 is different.

- **OpenFlow** is a **control protocol** for configuring predefined forwarding behavior.
- **P4** is a **programming language** for defining the forwarding behavior itself.

Let's make an analogy to understand the difference.

- **OpenFlow is like the driver of a regular car**. The driver can:
  - ✓ Steer left or right
  - ✓ Press the gas or brake
  - ✓ Use turn signals, radio, windshield wipers

  But the driver can't:
  - ✗ Change how the engine works
  - ✗ Reprogram how turning the wheel affects the tires
  - ✗ Add a new driving mode (e.g., "turbo boost")

  That's OpenFlow. We're in control of what happens (where to drive, how fast), but how the car works internally is fixed. We're controlling pre-built behavior, we're not changing the system.

- **P4 is like the car engineer or mechanic**. The car engineer can:
  - ✓ Redefine how the steering works (e.g., make left turn rotate only one wheel)
  - ✓ Change how the engine responds to the pedal
  - ✓ Add entirely new modules (e.g., self-driving mode, rocket engine, etc.)

  That's P4. We're not just driving the car, we're deciding what the car is capable of doing in the first place. We write the "rules" for how the system should behave.

## Workflow

1. Before starting to write a P4 program, is **necessary to know the P4 Architecture Model**. The **P4 Architecture Model** is a **logical interface** between:
   - The **P4 program** written by the developer.
   - The underlying **hardware target** (e.g., ASIC, FPGA, software switch)

   This model tells the compiler: "here's what the hardware looks like, these are the building blocks our P4 program can use.". This abstracts away hardware details and makes P4 programs portable across multiple targets.

   It's pretty obvious that the P4 architecture model is defined by the hardware switch we have. Because if our switch doesn't support some feature (e.g. packet cloning, a second pipeline), we can't use it.

2. **Write the P4 Program**. The network operator or developer writes a P4 program to describe:
   - Which **packet headers** to parse (e.g., Ethernet, IP, or custom)
   - What **tables** to build (match fields, actions)

- How the **control flow** works (pipeline logic)
- What actions to perform (forward, drop, modify, etc.)

This is written in a `.p4` file.

3. **Compile the P4 Program**. The P4 program is passed to a **P4 Compiler**, which does two main things:

   (a) **Generates a device-specific binary**. This is tailored to the target hardware (e.g., Tofino, FPGA, software switch like MBv2).

   (b) **Produces a runtime API**. This allows a controller (or CLI) to: install rules (e.g., match on `dstIP=10.0.0.1` forward to port 3), modify tables dynamically.

   The result is something the switch can understand and execute.

4. **Deploy to the Switch (Target)**. The compiled **output is loaded onto a P4-capable target**, such as: an ASIC (e.g., Barefoot Tofino), an FPGA-based switch, a software simulator (e.g., BMv2). At this point, the switch now knows how to: parse packets, match them in tables, take programmed actions.

5. **Runtime Table Configuration**. Once the program is installed, we still need to:

   - **Populate the tables** with actual forwarding rules.
   - This is usually done via a **controller**, using a runtime API (e.g., gRPC, Thrift, P4Runtime)

   It's like programming the switch with policy, after the logic has been defined.

Finally, the user is only concerned with the P4 program and the controller (to populate the tables). Instead, the P4 compiler, the P4 architecture model, and the switch (e.g., ASIC) are provided by the vendor.

## 3.4 PISA and Compiler Pipeline Mapping

**Protocol-Independent Switch Architecture (PISA)** is the **hardware abstraction** used by modern programmable switches (e.g., Barefoot Tofino). The idea behind PISA is simple but powerful: instead of building fixed-function blocks into hardware (e.g., IP routers, firewalls), **expose a generic pipeline of programmable stages**, and **let software define what each stage does**.

### PISA Architecture

A PISA switch consists of the following main components:

- **Parser**. **Extracts** packet **headers** and **creates** a structured **representation** (called a **Packet Header Vector**, or PHV). The PHV contains the keys for the match-Action units.
- **Multiple Match-Action Stages**. A pipeline of identical stages. Each stage:
  - Matches on some fields (using SRAM or TCAM)
  - Executes simple actions (via Arithmetic Logic Units - ALUs)
  - Modifies the PHV (e.g., changing a header field, setting a drop flag)
- **Deparser**. **Reassembles the packet** by combining the (possibly modified) headers and payload. Every packet flows through this pipeline, so the logic must be fully deterministic and parallelizable.

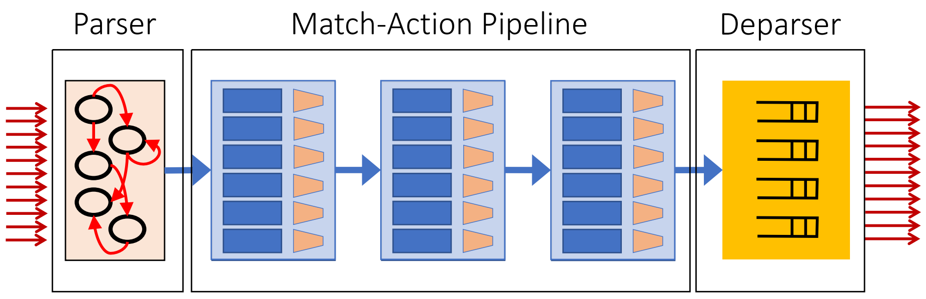


Figure 2: PISA architecture.

### Why use a pipelined architecture instead of a single processor?

A naive design would use **one CPU** to handle every packet: perform all lookups (routing, ACLs, NAT, etc.), apply all rules. But this would require an **unrealistically high frequency** to process billions of packets per second.

Just like in CPUs, we divide the processing into **stages**, each with: local memory (tables), local ALU, fixed resources. Each packet moves one stage forward per clock cycle, so we can **process many packets in parallel**.

### Protocol Independence

One of **PISA**'s most powerful features is that the chip **knows nothing in advance**.

- It doesn't recognize IP, Ethernet, TCP, or any protocol at all.
- The **programmer defines everything**: what headers to parse, what fields to match, what actions to perform.

This is what makes it **protocol-independent**, and feature-proof.

### What does the compile do?

Here's the key part of PISA and P4: we don't directly **program the pipeline**, **the compiler does**. We write a logical program in P4, and the P4 compiler:

- Analyzes dependencies between operations:
  - **Match dependency**: A table needs data generated by a previous match.
  - **Action dependency**: An action needs a value produced by a previous action.
- Packs logic into stages without violating resource limits
- Ensure parallelism and no data hazards

# 4 Data Structures

## 4.1 Introduction

Modern network devices, particularly programmable switches (PISA, page 45), implement a **packet processing pipeline** composed of three main blocks:

1. **Parser**: Extracts relevant headers from incoming packets.
2. **Match-Action Pipeline**:
    - **Match**: Uses lookup tables to compare extracted headers against known values.
    - **Action**: Applies logic (e.g., modify headers, make routing decisions).
3. **Deparser**: Reassembles the final packet for transmission.

This flow is deterministic and must maintain constant processing latency per stage, as switches are often implemented as hardware pipelines (one packet per stage per clock cycle).

### ❓ Layer 3 (L3) Router

The L3 router is a classic example used to explain the packet matching process:

- **Input**: IP destination address from packet.
- **Match Logic**: Find the Longest Prefix Match (LPM) in a routing table.
- **Action Logic**: Forward the packet to the correct output port, and adjust MAC address accordingly.

**Longest Prefix match (LPM)** is a fundamental concept in IP routing, where the **goal is to find the most specific route** (i.e., the one with the longest matching prefix) **for a given IP destination address**.

When a router receives a packet, it checks the **destination IP address** and compares it to entries in its **routing table**, which typically contain IP prefixes like:

- `192.168.0.0/16`
- `192.168.1.0/24`
- `192.168.1.128/25`

The router selects the entry whose prefix **matches the destination address** and has the **longest subnet mask** (i.e., most specific match).

In high-speed routers or programmable switches, **LPM must be done very quickly**, ideally in constant time. The naive solution for LPM is linear search over all routing entries. However, with thousands of entries, this is computationally infeasible at line rate. So the key questions in this section are:

- How do we **efficiently implement LPM**?
- Which data structures allow **fast lookups in a predictable and limited time**?

## 4.2 Ternary Content Addressable Memory (TCAM)

**Ternary Content Addressable Memory (TCAM)** is a specialized kind of (hardware) memory that works differently from standard RAM. Instead of accessing data by address, TCAM lets we input data and instantly tells we **if and where it's stored**. This is called **associative memory** or **content-based lookup**. It is built specifically for fast parallel search.

Unlike binary memories (which store 0s and 1s), **TCAMs can store 0, 1, or a third state** (ternary) called "*don't care*". This third value allows flexible and partial matching, making TCAMs very effective for operations like Longest Prefix Match (LPM) in IP routing.

### ⚖ RAM vs TCAM

- RAM (Random Access Memory):
  - Ask: "**What** is stored at address `X`?".
  - Classic address-value access.
- TCAM:
  - Ask: "**Where** is the value `X` stored?".
  - The memory searches all entries in parallel and returns the matching address in constant time. In other words, it **returns the address where the value is stored**.

This associative search is **very fast**, which is why TCAM is often used in **packet classification** and **routing tables** in high-speed switches. Usually these two hardware are **put together** because the TCAM gives the index, we use it to index the RAM, and we get the information.

### ✓ Pros

- ✓ **Speed**: Lookup happens in constant time, regardless of the number of entries.
- ✓ **Wildcard Matching**: TCAMs handle "don't care" bits, allowing prefix and pattern-based lookups.
- ✓ **Ideal for Match-Action Pipelines**: TCAM is a good fit for hardware pipelines like those found in P4-programmable switches.

### ✗ Cons

- ✗ **High power consumption**: Every lookup checks all entries in parallel.
- ✗ **Expensive**: Due to the hardware complexity and power demands.

**Example 1: TCAM in packet routing**

Imagine a TCAM storing IP prefixes:

- `0: 192.168.3.0/24`
- `1: 192.168.1.0/24`
- `2: 192.168.2.0/24`

If an incoming packet has destination IP `192.168.2.1`, the TCAM instantly finds that it matches entry 2.
However, this match index alone isn't enough to decide what to do. So, we usually pair TCAM with a RAM block that stores the actual action:

1. TCAM gives the index
2. Use it to index RAM
3. Get forwarding info, output port, etc.

## ⚠ Dealing with Multiple Matches

Sometimes a destination IP can match multiple entries. For example:

- `Entry 2: 192.168.2.0/24`
- `Entry 3: 192.168.2.0/28`

Both may match the same address, but **only one result is returned**. Depending on the hardware, this could be:

- The **lowest matching index** (first match)
- The **highest matching index** (last match)

To ensure correct behavior (e.g., always choosing the most specific prefix), **entries need to be** carefully **ordered**. This introduces extra logic during configuration or compile time.

## Extra Hardware: SRAM

Alongside TCAM, **each pipeline stage** may also have **SRAM**. It's used for:

- Storing values linked to TCAM matches.
- Keeping state (e.g., counters, flags).
- Performing fast value retrieval during match-action processing.

SRAM is faster and cheaper than TCAM, but does not supper associative lookup, so it **complements TCAM** rather than replacing it.

**Key Takeaways**

- TCAM is **fast**, parallel, and supports wildcards, great for networking.
- It's **costly and power-hungry**, so it's used sparingly and carefully.
- Works in tandem with **SRAM** for decision and action pipelines.
- Entry **ordering matters** to get correct behavior (e.g., longest prefix match).

## 4.3 Deterministic Lookup with Probabilistic Performance

### Problem Setup

We want to store a collection of elements (a set) in memory, and be able to:

- **Insert** new elements.
- **Check** if an element exists.
- **Do it fast**, ideally in constant time.

There are two broad strategies:

- **Deterministic** (this section):

  ✓ **Always** gives the **correct answer** (*deterministic lookup*, answer is always correct).

  ✗ **Slower** or **require more memory** (*probabilistic performance*, time depends on insertion history).

  With this type of data structures, we **always get the correct answer** (`true` if present, `false` if not), but the **number of steps** (e.g. in Separate Chaining, explained below, the steps are determined by traversing a chain) is **not fixed** because it depends on: collisions, load factory, quality of the hash function.

- **Probabilistic** (section 4.4, page 54):

  ✓ **Uses less memory** or **is faster** (*deterministic performance*, always the same number of operations).

  ✗ Might give **false positives/negatives** (*probabilistic lookup*, result might be wrong with some small probability).

  With this type of data structures, the **time is constant**, we have a fixed number of bit checks (usually one or a few), but the **answer can be wrong**. We can get a false positive (return `true` if the element isn't in the set), or we never get a false negative (if it says `false`, the element definitely wasn't inserted).

In this section we analyze the deterministic approach, so an output that we know what is, but the number of operations required is unknown (probabilistic). This is not suitable for network computing because it is detached from the PISA idea, but we present it for academic purposes.

### Hash Table

A **Hash Function maps data** of arbitrary size (e.g., strings like "hello") **into a fixed-size integer space**. This **integer** is then **used as an index in an array** called a **Hash Table**.

Pseudo-code:

```
index = hash("hello")
hash_table[index] = "hello"
```

But **collisions can occur**, multiple inputs may hash to the same index. To handle this, we **use separate chaining**.

## Separate Chaining: The Basic Idea

The basic idea of **Separate Chaining** is as follows. If two values hash to the same index, we **chain** them **together in a list**:

$$\texttt{Index 10} \rightarrow \texttt{``hello''} \rightarrow \texttt{``port''} \rightarrow \texttt{``fire''}$$

So instead of storing just one value per index, we allow **each index to store a linked list** (or vector, or queue).

## Performance Analysis

Let's say we're inserting $N$ elements into a table with $M$ buckets.

- **Average list size:** $\frac{N}{M}$
- Best case (uniform distribution): All chains are of similar length $\rightarrow$ fast lookups.
- Worst case: All $N$ elements hash to the same bucket $\rightarrow$ one long chain $\rightarrow O(N)$ lookup time.

So the **load factor** $\frac{N}{M}$ is key to understanding performance.

## Collision Probability

How likely is it to avoid collisions at all?

- 1st insertion: no collision.
- 2nd: no collision with probability $1 - \frac{1}{M}$
- 3rd: no collision with probability $1 - \frac{2}{M}$
- ...
- $N$-th: no collision with probability $1 - \frac{N-1}{M}$

Multiply all together to get the probability of **zero collisions**:

$$P(N, M) = \prod_{i=0}^{N-1} \left(1 - \frac{i}{M}\right) \tag{1}$$

This means that even if $M = 10000$ and $N = 100$, the chance of having at least one collision is about 40%! In other words, **collisions are almost inevitable unless** $M \gg N$.

✓ **Pros** and ✗ **Cons** of Separate Chaining

- ✓ **Output deterministic and accurate**, no false positives/negatives.
- ✓ Simple and well-understood.
- ✓ **Performs well if load factor is low**.
- ✗ **Memory usage can grow** if many chains form.
- ✗ **Slower when many elements are inserted and collisions increase**.
- ✗ Not ideal for extremely large-scale or memory-constrained environments.

## 4.4 Probabilistic Data Structures

### 4.4.1 1-Hash Bloom Filters

We've just seen **Separate Chaining**, which gives **accurate answers** but has **unpredictable performance**, not ideal for hardware pipelines. Now we flip the perspective.

This section introduces **probabilistic data structures**, where:

✓ Insertions and lookup have a fixed, **deterministic number of operations**, typically 1.

✗ However, the **lookup result is probabilistic**, so it can produce false positives with a small probability.

Why this trade-off? Because in networking hardware (e.g., PISA architecture), we care more about **fixed latency** tan occasional inaccuracies.

#### ✓ A simple bit-based Data Structure

Let a set implemented as a simple bit array:

- An array of $M$ 1-bit cells, all initially set to 0.
- To insert an element:
  1. Compute a hash function `hash(x)`
  2. Set the bit of the result of the hash function to 1: `bit[hash(x)] = 1`
- To check if `x` is in the set, we simply: `bit[hash(x)] == 1`

This data structure is often called a **1-hash Bloom Filter** because it has only **one hash function** and only **one bit per element**.

**Example 2: Single-Hash Bloom Filter**

Let an array of $M$ 1-bit cells, all initially set to 0, we insert:

1. "`Rust`" $\rightarrow$ sets 1 bit of the array to 1
2. "`Hello`" $\rightarrow$ sets another bit to 1
3. "`Fine`" $\rightarrow$ sets another bit to 1. Now 3 bits are set

Now we will try some lookups.

- "`Hello`" $\rightarrow$ `bit[hash(Hello)] == 1`? $\xrightarrow{\text{YES}}$ ✓ return `true`
- "`Bye`" $\rightarrow$ `bit[hash(Bye)] == 1`? $\xrightarrow{\text{NO}}$ ✗ return `false`
- "`P4`" $\rightarrow$ `bit[hash(P4)] == 1`? $\xrightarrow{\text{YES}}$ ✓ return `true`, but we never inserted it. It is a **false positive**

### % Probabilistic Analysis

Let:

- $N$: number of **inserted elements**
- $M$: number of **bit cells**

**Probability** that an **element maps to a** particular **bit** is:

$$\frac{1}{M}$$

So:

- **Probability** that an **element doesn't map to a bit**:

$$1 - \frac{1}{M} \tag{2}$$

- **Probability** that a **bit stays <u>0</u> after** $N$ **insertions**:

$$\left(1 - \frac{1}{M}\right)^N \tag{3}$$

- **Probability** that a **bit becomes <u>1</u>**, called **False Positive Rate (FPR)**:

$$\text{FPR} = 1 - \left(1 - \frac{1}{M}\right)^N \tag{4}$$

- Finally, the **False Negative Rate** is 0. Bloom filters (1-hash or multiple hashes) **guarantee that they cannot return a false negative**. Suppose our hash function returns a value of 3 when we put in the string "`Rust`" (`hash(Rust) = 3`); if we put the word "`Rust`" into the bit array, we have `bit[hash(Rust)] = 1` $\Rightarrow$ `bit[3] = 1`. Later, when we query "`Rust`", the data structure will always return `true`, because `bit[hash(Rust)] = bit[3] == 1 ? true`.

### Pros

✓ Simple

✓ **Fast**, constant-time insertion and query

✓ **Deterministic performance**, perfect for hardware pipelines

### Cons

✗ **Not always accurate**, there can be false positives.

✗ To keep FPR low (e.g. 1%), **we need** 100× **more memory than elements**.

### 4.4.2 Bloom Filters

A **Bloom Filter** is a space-efficient probabilistic data structure used for **membership queries**:

- **Fast** insertions lookups.
- **No false negatives**, but may return **false positives**.
- This trade-off is ideal for **fixed-latency, high-speed systems** (like programmable switches).

#### Generalization of the 1-hash Bloom filter to $k$-hash

To **reduce false positives**, we **extend the 1-hash Bloom Filter**:

- Instead of just 1 hash function, we use $K$ different hash functions.
- Each function maps the input element to a different positions in the bit array.

#### How Insertion Works

Let's say we want to insert "`Rust`":

1. Compute $K$ hash functions in parallel:

$$h_1(x) \qquad h_2(x) \qquad \dots \qquad h_K(x)$$

2. For each $h_i(x)$, set the bit at position $h_1(x)$ to `1`.
   - $h_1(x) = 1$
   - $h_2(x) = 1$
   - ...
   - $h_K(x) = 1$

#### How Lookup Works

To check whether an element is in the set:

- Compute all $K$ hashes
- If **all corresponding bits are set to 1**, we return `true` (element may be present)
- If **at least one bit is 0**, we return `false` (element is definitely not present)

## ⚠ False Positives

Let's say "`Fire`" is not inserted but happens to have all its hash bits already set by "`Rust`", "`Hello`", or "`Fine`". The filter will wrongly return `true`, a false positive. Still, **no false negatives** can occur: if an element was inserted, all bits are set, and it will always return `true`.

## % Probability Analysis

Let:

- $N$: number of **inserted elements**
- $M$: number of **bits in the filter**
- $K$: number of **hash functions**

Then:

- **Probability** a particular **cell is still 0 after inserting $N$ elements**:

$$\left(1 - \frac{1}{M}\right)^{(K \cdot N)} \tag{5}$$

- **Probability** of a false positive (all $K$ bits set for a non-inserted element), the **False Positive Rate (FPR)**:

$$\text{FPR} = \left(1 - \left(1 - \frac{1}{M}\right)^{(K \cdot N)}\right)^{K} \tag{6}$$

Just as an idea, with $1'000$ elements inserted, $10'000$ bits in the filter (cells), and 7 hash computations, we get a probability of FPR of only 0.82%. And if we increase the bits in the filter ($M$) to $100'000$, the FPR is about 0%! So, with a **moderate increase in memory and hash computations**, we can get extremely low FPRs.

## ✓ Pros

- ✓ Very **memory-efficient**, uses up to 10× less memory **than separate chaining**.
- ✓ Lookup and insertions are **predictable and fast**, constant time with $K$ steps.
- ✓ Still **no false negatives**.

## ✗ Cons

- ✗ Requires **more computation** than the single-hash version (e.g., 7 hash functions).
- ✗ Slightly **more complex to implement in hardware**.

### 4.4.3 Dimensioning a Bloom Filter

We want to design a Bloom Filter that:

- Stores $N$ elements
- Uses $M$ bits (memory size)
- Applies $K$ hash functions

But we also to **control the False Positive Rate (FPR)** and avoid unnecessary computation.

There are three parameters in play:

1. **Memory** $M$: more bits $\Rightarrow$ lower FPR
2. **Number of Hashes** $K$: more hashes $\Rightarrow$ lower FPR, but higher computational cost
3. **False Positive Rate** (FPR): we want this to be as low as possible.

Improving one usually worsens another. This is the classic **space/time/error trade-off**.

#### ✓ Asymptotic Approximation for FPR

In our case, the Asymptotic Approximation is a simplified mathematical expression that **estimates the False Positive Rate (FPR)** of a Bloom Filter when the number of **cells** $M$ **is large**. It's derived from the exact expression but uses limits and approximations that hold when $M \gg N$. It's much easier to work with and very accurate in practice.

If we insert $N$ elements into a Bloom filter with $M$ bits and use $K$ hash functions, the **exact False Positive Rate (FPR)**:

$$\text{Exact FPR } = \left(1 - \left(1 - \frac{1}{M}\right)^{(K \cdot N)}\right)^{K} \tag{7}$$

This expression can be tedious to compute, especially for large values of $M$, $N$, and $K$. By using the approximation:

$$\left(1 - \frac{1}{M}\right)^{K \cdot N} \approx e^{\left(-K \cdot \frac{N}{M}\right)} \qquad \text{when } M \gg 1$$

The **Asymptotic Approximation of False Positive Rate (FPR)** is:

$$\text{FPR } \approx \left(1 - e^{\left(-K \cdot \frac{N}{M}\right)}\right)^{K} \tag{8}$$

This approximation is easier to analyze and is widely used in practice.

#### ? Finding the Optimal Number of Hash Functions

The optimal number of hash functions $K$ **minimizes the FPR** for given $M$ and $N$. We can find it by minimizing the FPR formula:

$$K_{\text{opt}} = \frac{M}{N} \cdot \ln(2) \tag{9}$$

### 4.4.4 Counting Bloom Filters

In the standard Bloom Filter:

- Inserting an element means setting multiple bits to `1`.
- But we **never know which element caused a bit to be** `1`, because multiple elements may share the same hash outputs.

#### ⚠ What happens if we try to delete?

Let's say we inserted: "`Rust`" and "`Hello`". And now we want to delete "`Rust`". If "`Rust`" and "`Hello`" both caused a bit (say, index 9) to be set to `1`, and we reset it to `0` to delete "`Rust`", now:

- When we query "`Hello`", it might show a `0` in one of its position.
- This creates a **false negative**, which violates one of the core guarantees of Bloom filters!

So, **manually unsetting bits can remove evidence of other elements**.

#### ✓ Solution: Counting Bloom Filters

To enable deletion, we **upgrade each bit into a counter**, this structure is called a **Counting Bloom Filter**. It works like this:

- Instead of a bit array, we use an **array of small integers**.
- When **inserting**, for each hash $h_i(x)$, increment `counter[`$h_i(x)$`]`.
- When **deleting**, for each hash $h_i(x)$, decrement `counter[`$h_i(x)$`]`.

We can safely decrement counters, knowing that **only when the last element that hashed to that index is deleted will the counter reach zero**. All previous analyses about false positives, FPR formula and $K$ optimal are still valid, but now we **use more memory** and **add increment/decrement logic**.

#### ⚠ Risk: Counter Overflow

Counters must be large enough:

- If they **overflow** (e.g., go above 255 for 8-bit counters), the **filter can become corrupted**.
- Worse, if a counter **underflows** (e.g., we delete too many times), we might **accidentally remove bits** for elements still in the set $\Rightarrow$ **false negatives**.

### 4.4.5 Invertible Bloom Lookup Tables (IBLTs)

With Count Bloom Filters, we can:

- Insert elements
- Delete them
- But we **can't list what's inside**, or **retrieve keys/values**, the information is "smeared" across the structure.

Now we want something more powerful that can also list all entries or recover a specific key-value pair.

#### What is an IBLT?

An **Invertible Bloom Lookup Table** is a data structure that:

- Stores **key-value pairs**
- Supports **deletion** and **enumeration (listing)**
- Is inspired by Bloom Filters, but has a richer cell structure.

Each cell contains three values:

1. **Count**: how many key-value pairs map to this cell.
2. **KeySum**: XOR (or sum) of all keys that mapped here.
3. **ValueSum**: XOR (or sum) of all values that mapped here.

We hash the key using multiple hash functions, just like a Bloom filter, and update each corresponding cell.

#### Insertion

To **insert a key-value** pair:

1. Use $K$ hash functions to map the key to $K$ cells.
2. For each cell:
   (a) Increment the **count**
   (b) Add the key to **KeySum**
   (c) Add the value to **ValueSum**

### Deletion

To **delete a key-value** pair:

1. Use the same $K$ hash functions.
2. For each cell:
   - **Decrement** the **count**
   - **Subtract** the key from **KeySum**
   - **Subtract** the value from **ValueSum**

If the key was inserted, this will perfectly remove it.

### Lookup and Recovery

To **find a value for a key**:

1. Try to find a cell where `count == 1` and the `KeySum == input key`
2. If found, then `ValueSum` gives the value associated with that key

But:

- If the key is mixed with other keys in all $K$ cells, recovery is hard.
- That's why **some keys may not be recoverable immediately**.

### Enumerate everything stored in it

Once the structure is filled with multiple key-value pairs, we may want to enumerate everything stored in it, not just individual lookups. This process is known as **decoding** or **peeling** the IBLT. This restore operation is often used in real-world scenarios, for example, when we want to compare two sets of two different devices.

The **decoding algorithm** is:

1. Scan the table for a cell where:
   - `count == 1`
   - `KeySum` and `ValueSum` correspond to an actual key-value pair
2. When found:
   - Add the pair to output
   - Simulate deletion: subtract this key and value from all corresponding cells
   - Update the IBLT

For example:

1. Initial IBLT contains:

| Count | KeySum | ValueSum |
|---|---|---|
| 1 | 7 | 98 |
| 2 | 202 | 48 |
| 3 | 209 | 146 |
| 2 | 159 | 101 |
| 1 | 50 | 45 |

2. First, a cell with `count = 1` reveals:
   - `(7, 98)` ⇒ added to output
   - Remove it from the IBLT (as if deleting it)

| Count | KeySum | ValueSum |
|---|---|---|
| **0** | **0** | **0** |
| 2 | 202 | 48 |
| **2** | **202** | **48** |
| **1** | **152** | **3** |
| 1 | 50 | 45 |

3. After update:
   (a) Next, find `(152, 3)` ⇒ decode and remove

| Count | KeySum | ValueSum |
|---|---|---|
| 0 | 0 | 0 |
| **1** | **50** | **45** |
| **1** | **50** | **45** |
| **0** | **0** | **0** |
| 1 | 50 | 45 |

   (b) Finally, we can easily retrieve the last one which is `50, 45`.

4. The final result is:

$$\{(7,\ 98),\ (152,\ 3),\ (50,\ 45)\}$$

This process works **only if at least one of the key-value pairs is initially recoverable**, and then the remaining pairs become recoverable as the IBLT gets simplified.

## ⚠ Decoding problems

Sometimes, the decoding process **gets stuck**:

- All cells have `count > 1`, or are tangled with other keys
- We cannot isolate any key-value pair

When this happens:

- Listing **fails**
- The **IBLT** is said to be in a **non-decodable state**
- This usually happens when the **load factor is too high** (i.e., too many elements for the number of cells)

So IBLTs are powerful because allowing insertion, deletion, lookup and enumeration; but we need to allocate enough space, because if we overloaded, we risk failure to decode.

| **Feature** | **Standard Bloom** | **Counting Bloom** | **IBLT** |
|---|---|---|---|
| Insert | ✓ | ✓ | ✓ (key-value) |
| Delete | ✗ | ✓ | ✓ |
| Membership Test | ✓ (yes/no) | ✓ | ✓ (via decoding) |
| False Negatives | ✗ | ✗ | ✗ (unless corrupted) |
| False Positives | ✓ | ✓ | ✗ (when decoding works) |
| Listing Elements | ✗ | ✗ | ✓ (if decodable) |
| Memory Efficiency | Very high | Moderate | Lower (more fields) |

Table 4: IBLT vs Bloom Filters.

### 4.4.6 Count-Min Sketch

The **Count-Min Sketch** is a **probabilistic data structure** used to estimate the **frequency of elements** in a stream.

- We don't store each element individually.
- Instead, we use a **compat structure to maintain approximate counts**.
- It's designed for efficiency, especially when tracking millions of elements would be too memory-intensive.

#### How does it work?

We create a 2D array of counters with:

- $d$ rows (one per hash function)
- $w$ columns (size of each hash domain)

This gives a table of size $w \times d$, much smaller than a full hash table for all possible items. **Each row** has a **different hash function**.

#### Insertion

To insert an element (e.g. `ip.dest1`):

1. **Hash the element** with each of the $d$ hash functions.
2. **Each hash** gives we a **column index in its row**.
3. **Increment** the corresponding **counters**.

So we increment **1 counter per row**, total $d$ counters updated.

#### Querying the Frequency

To estimate the count of an element:

1. Hash it again with the same $d$ hash functions.
2. Get the counter values from the same positions.
3. **Return the minimum** of those $d$ counters.

The **minimum** because:

- Collisions with other elements can cause overestimation (counters get inflated).
- But the **minimum is never less than the true count**, so it's a **safe lower bound**.

That's where the name comes from: count, because it estimates the frequency, and min, because it takes the minimum over multiple counters.

### ✓ Advantages

- Sublinear space: uses **much less memory** than a full table.
- **Fast**: insertions and queries are both $O(d)$ time (constant if $d$ is fixed).
- Suitable for **high-speed data streams** (e.g., network flows, telemetry, monitoring).

| **Feature** | **Value** |
|---|---|
| Use case | Approximate frequency counts |
| Memory | Sublinear ($w \times d$) |
| Insertion time | $O(d)$ |
| Query time | $O(d)$, returns minimum |
| Overestimates | Possible |
| Underestimates | Never |
| Similar to | Counting Bloom Filter |

Table 5: Count-Min Sketch summary.

# 5 Datacenter Monitoring

## 5.1 Why Datacenter Monitoring Matters

Image we're running a distributed application in a datacenter, and performance suddenly degrades. The possible root causes can be multiple:

- A software bug in the application logic.
- Network congestion between the servers.
- A broken fiber cable disrupting communication.
- A hardware failure, e.g. broken switch.
- A network misconfiguration that reroutes traffic inefficiently.
- A bug in the routing protocol.
- And many more...

There is a huge space of possible issues, and **pinpointing the root cause without visibility is extremely difficult**.

Many papers describe the importance of monitoring in datacenters.

- In the *Pingmesh* [4] article, they point to research that has begun to investigate **how to distinguish network problems from application-level bugs**. They highlights the diagnostic ambiguity in complex systems. Without monitoring, it's extremely hard to tell whether a slowdown is due to:
  - Software bugs;
  - Application overload;
  - Or actual network failures.

  **Monitoring systems must disambiguate the root cause across layers**, application vs network.
- In the "*Understanding and Mitigating Packet Corruption in Data Center Networks*" [8] article, they showing how **minor misconfiguration or failures** (e.g., wrong routing entry) can ripple through a system, **creating major outages**. It stresses that even low-level, seemingly unimportant events mu be visible to prevent or debug large-scale issues. For example, a single corrupted forwarding rule in a switch might cause traffic loss affecting thousands of users.

  **Monitoring must include fine-grained data** (like per-packet or per-flow telemetry) **to detect these small but critical problems**.
- In the "*Flow Event Telemetry on Programmable Data Plane*" [7] article, they show that **performance degradation often happens silently**, with no clear immediate failures. These "gray failures" don't crash systems but hurt performance. They're invisible without high-resolution monitoring (latency histograms, queue lengths, retransmits, etc.).

  **Monitoring should detect subtle deviations**, not just crashes or timeouts.

- In the *CloudCluster* [6] article, they push toward deep programmability and visibility withing the network. This points to the **evolution of monitoring tools**:
  - From passive logs and SNMP stats;
  - To programmable packet tracing and real-time telemetry;
  - That help pinpoint network issues quickly and accurately.

  **Visibility must be deep, dynamic, and distributed across the system**.

## 5.2 Network Monitoring

Network Monitoring is the **continuous observation of a computer network to detect slowdowns or failures in components**. Its purpose is to detect, localize and respond to faults before they impact users.

### Monitoring Scope

Monitoring spans the **entire network path**. Each router (or switch/server) is a point where failures or slowdowns can occur:

- A switch could drop the packet silently.
- A routing issue could cause the packet to loop.
- A delay could occur due to congestion in queues.

**To effectively detect and diagnose problems**, the monitoring system must observe not just endpoints, but the entire path, or at least enough of it to: detect *where* things go wrong, or understand *why* a packet failed to reach its destination. This is why datacenter monitoring often tries to trace or mirror packets at different points in the network, to reconstruct the packet's journey and find anomalies.

### Monitoring Techniques

There are many ways to monitor a network. It can be done:

1. **From Switches**:
   (a) **Built-in Features** (section 5.3, page 70)
      - *NetFlow*: collects IP traffic statistics.
      - *Mirroring*: duplicates selected packets for analysis.
      - *SNMP (Simple Network Management Protocol)*: polls device stats.

   (b) **Programmable Switches**
      - Use *data plane programmability* (e.g., P4 language) to define custom monitoring behaviors.
      - Enables *custom counters*, tagging, filtering, or tracing at wire speed.
2. **From Servers**:
   (a) **Standard Tools**
      - `netstat`: network connections and stats.
      - `tcpdump`: packet capture and inspection.
      - `traceroute`: path tracing and latency.

   (b) **Ad-hoc Monitoring Services**
      - Lightweight daemons or agents tailored for the datacenter.
      - Export performance metrics or send alerts.

There is no single way to monitor, a **mix of passive and active, centralized and distributed methods is used**. Monitoring systems must collect data from multiple vantage points to build a full picture of the network's health.

### ⚠ Why traditional monitoring isn't enough

In large-scale datacenters many failures are subtle and effect only specific flows of packets:

- **Silent packet drops**. Packets are dropped but not reported by switches. The causes are software bugs or faulty hardware.
- **Silent blackholes**. Traffic is blackholed without showing in forwarding tables. The causes are corrupted TCAM entries.
- **Inflated end-to-end latency**. Packet flow experiences unexpected delays. The causes are congestion or queuing.
- **Loops**. Packets circulate endlessly. The causes are middleboxes modify headers or breaking routing logic.

These failures are:

- **Not visible** in flow-level stats.
- **Not logged** by switches.
- **Hard to localize** with only endpoint observations.

✓ We need **per-packet visibility** to detect and understand them.

### ? So can we monitor every packet on the network?

Tracing all packets in large datacenters is not scalable:

- Aggregate traffic can exceed 100 terabit per seconds.
- Microsoft estimated 3200 servers needed just to collect and analyze th data (in 2015).

To make packet-level telemetry practical, some strategies are required:

1. Monitoring must be **selective and smart** (e.g., sample important flows).
2. Diagnosing problems often requires **correlating behaviors across multiple hops**.
3. **Passive tracing alone is insufficient**:
    - It may miss transient problems.
    - It lacks the context to localize root causes effectively.

## 5.3 Everflow

### 5.3.1 What is Everflow?

Everflow [3] is Microsoft's system for **packet-level telemetry in production datacenters**, and it is built around three key concepts:

1. **Match and Mirror on the Switch**. Everflow leverages the match-action capability of commodity switches. It **defines rules to match specific packets and then mirror (copy) them to a monitoring collector**. Three matching rules:

   - `TCP SYN / FIN / RST`: to trace connection setup/teardown.
   - **Special debug bit**: used to flag packets for tracing.
   - **Protocol traffic**: such as BGP or other control plane packets.

   This allows the system to **monitor important or suspicious traffic patterns without touching every packet**.

2. **Switch-Based Reshuffler**. Mirroring packets from all switches generates huge data volumes. A single analysis server can't handle this load. The solution is to **use one or more intermediate switches** (reshuffler) that:

   (a) Receive mirrored packets.

   (b) Distribute them intelligently across multiple collectors.

   This **balances load and scales the telemetry infrastructure**.

3. **Guided Probing**. The system can **inject specific test packets into the network**. These packets are crafted to **explore or verify behaviors** (e.g., path correctness, loss, latency). They are useful because:

   - Helps when match and mirror alone misses packets (e.g., for complete TCP flow analysis).
   - Can reproduce or test suspected failures.
   - Distinguishes between persistent and transient issues.

   Uses DSCP bits (in IP headers) and parts of the IPID field to mark and sample packets.

| Idea | Purpose | Key Technique |
|---|---|---|
| **Match and Mirror** | Capture relevant packets | Matching on SYN / FIN / debug bits; mirroring to collectors. |
| **Reshuffler** | Scale analysis | Distribute mirrored packets across servers. |
| **Guided Probing** | Actively test network behavior | Inject custom packets using special bit fields. |

Table 6: Summary of Everflow concepts.

### 5.3.2 How it works

Everflow isn't just a single-purpose tool, it's an extensible framework that supports different debugging applications, all coordinated through a central controller.

1. **Everflow is Application-Driven**. Operators use Everflow to **run specific troubleshooting tasks**, such as:

   - Latency profiling
   - Packet drop debugging
   - Loop detection

   Each task is handled by an Everflow application tailored to that goal.

2. **The Controller as the Central Brain**. The **controller** coordinates the full debugging process:

   - It receives:
     (a) The operator's request (e.g., trace all flows to a web server).
     (b) The expected network routing.
   - It then:
     (a) `[init]` **Installs match-and-mirror rules** in selected switches.
     (b) `[config]` **Configures the analyzers** to process mirrored traffic.
     (c) `[debug]` **Sets the debug bits** (e.g., using DSCP or IPID) in custom probes if needed.

     This modular design allows Everflow to adapt to the operator's intent dynamically.

3. **Data Collection via Reshuffler and Analyzers**. Once rules are deployed:

   - **Mirrored packets** from the switches **go to a Reshuffler**.
   - The **Reshuffler distributes the traffic to** multiple **analyzers** (to balance the load).
   - The **analyzers inspect** the packet streams for signs of abnormal behavior.

4. **Smart Storage: Only Save What's Important**. Even with match-and-mirror, the system can generate **a massive amount of trace data**. For optimization, the **analyzers write to memory only packets with the debug bit set**, or **packets that show anomalies** (e.g., unusual delays, missing responses). This filtering prevents overload and ensures only useful diagnostic data is saved.

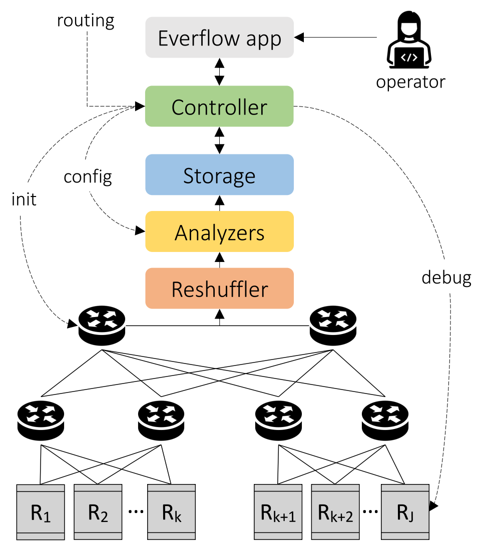


Figure 3: Summary of Everflow's End-to-End operation:

- **Operator Request**. Chooses a debugging goal (e.g., latency analysis).
- **Controller**. Interprets request, maps it to rules and config.
- **Switch Configuration**. Match and mirror rules installed.
- **Probing (Optional)**. Probes injected with debug bits.
- **Data Reshuffling**. Mirrored packets routed to analyzers.
- **Analysis**. Analyzers check for problems.
- **Selective Storage**. Only suspicious packets are saved.

**Example 1: Real episode**

Internal users reported that **some connections to a web service were timed out**. This violated the service level agreement (SLA). The root cause was suspicious: packet drops were occurring, but *where exactly*?

The service architecture involved multiple components: clients, load balancers, web servers, databases. All interconnected over the datacenter network. But the **datacenter is huge**, with many possible failure points.

The investigation begins.

1. Load Balancers showed no errors in their counters.
2. Some switches were checked manually, no issue found.
3. But the problem persisted, random connection timeouts were still

happening.

This is where Everflow comes in.

1. Everflow was used to **mirror TCP SYN packets** (which initiate connections) across the network.
2. Through its **trace analysis**, it was observed that:
   - Many SYN packets **never reached** the destination web server.
   - This only happened for **one specific web server**.
3. Further analysis reveled:
   - All SYN packets to that web server were **dropped at one switch**.
   - The switch showed **no error counters**, completely silent.

The root cause has been identified. The TCAM (Ternary Content Addressable Memory) on that switch was corrupted (TCAM stores forwarding table entries, used to decide where packets go). Because the corruption was silent:

1. The **switch dropped packets silently**.
2. **No logs**, **no alarms**, **no metrics**, traditional monitoring failed.

After a reboot of the switch, the issue disappeared.

## 5.4 FlowRadar

### 5.4.1 Architecture

FlowRadar [5] is a scalable, low-overhead solution for **per-flow monitoring** in datacenter networks. Its goal is to track **how much traffic each flow generates**, using **programmable switches** with fixed, minimal resources.

#### Why FlowRadar?

Monitoring networks at flow-level granularity is **valuable but expensive**:

- ✗ **Packet mirroring**. Every interesting packet is copied and sent to an external analyzer.

  ⚠ **Problem**: way **too much traffic**. In datacenters with 100 terabit per seconds, this would flood our monitoring system.

- ✗ **Per-flow counters at switches**. Maintain one counter per flow inside the switch.

  ⚠ **Problem**: switches have **very limited memory**. With millions of flows, we run our of space fast.

So FlowRadar sits in the sweet spot:

- ✓ It avoids mirroring massive amounts of data.
- ✓ It doesn't require full flow counters in the switches.
- ✓ It **works with fixed, limited operations per packet**, suitable for programmable hardware.

The main idea is to encode information compactly in the switch, then **decode it later at the collector**.

#### How it works

The FlowRadar works in three different ways:

1. **In the Switch**. Each switch maintains a **compressed data structure** to track flows and their counters. The **structure is similar to an Invertible Bloom Lookup Table** (IBLT, page 60), it records **flow IDs and counters** in a space-efficient way. Operations per packet are fixed and fast (ideal for hardware).
2. **Periodic Reports**. Switches **periodically export** their **encoded flow data to** central **collectors**.
3. **At the Collector**. Collectors receive the compressed data. Using multiple switch reports, they **correlate and decode** per-flow information. This allows the network operator to recover flow ID, and packet or byte count per flow.

### 5.4.2 Data Structure used in FlowRadar

**Switches have very limited memory**, but we want to count **how many packets are part of each individual flow**. Instead of using a separate counter per flow, which would consume too much space, FlowRadar stores **compressed aggregate information** in a fixed-size structure.

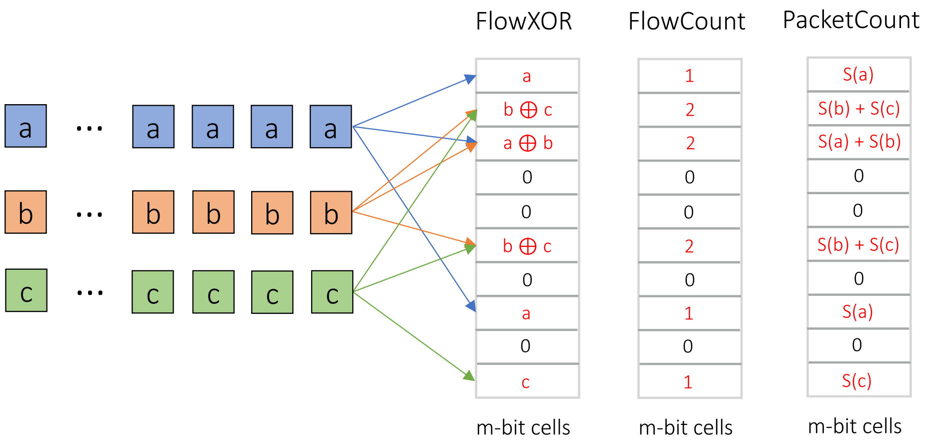
Each switch maintains three tables of $m$ cells:

- **FlowXOR**. XOR[2] of all flow IDs hashed into each cell.
- **FlowCount**. Number of flows that map to each cell.
- **PacketCount**. Total number of packets from those flows.

For example, assume flows A, B, and C are all seen by the switch. Each flow is hashed into multiple cells. The tables get updated like this:

- FlowXOR: $a \otimes b$, $b \otimes c$, etc.
- FlowCount: how many flows are hashed into each cell (1, 2, etc.)
- PacketCount: total packets seen in each cell (e.g., $S(a) + S(b)$)

So each cell contains a **mix of data from different flows**.



### ❓ Why use XOR?

Because the XOR operation is **reversible**. If we know the XOR of two values and one of them, we can recover the other. This **allows the collector to decode the original flows** by:

1. Getting reports from multiple switches.
2. Iteratively solving the system of XOR equations.

[2]XOR (Exclusive OR) is a binary operation denoted by $\otimes$, where the result is 1 if the two input bits are different, and 0 if they are the same.

## ⚠ What is Flow Filter and why do we need it?

The **Flow Filter** is a small data structure (like a Bloom Filter) used inside the switch to **remember which flows the switch has already seen**.

Let's say flow *a* sends 10 packets. All those packets will pass through the switch. But we don't want to treat each packet like a new flow, **we only want to register flow *a* once in the compressed counters**. If we update the XOR and FlowCount on every packet:

- The **FlowXOR** would get corrupted.
- The **FlowCount** would become too high.
- We'd **lose the ability to decode** the flows correctly later.

So the **Flow Filter helps us avoid this**.

## ❓ What does Flow Filter actually do?

For each packet:

1. The switch looks at the Flow ID (e.g., `source IP + dest IP + ports`).
2. It checks the Flow Filter:
    - If the flow is **new** (not in the filter yet):
        (a) It updates:
            – FlowXOR: add this flow's ID via XOR.
            – FlowCount: increment the count.
            – PacketCount: add 1.
        (b) It marks this flow as *seen* in the filter.
    - If the flow is **already known**:
        (a) It updates **only PacketCount**.

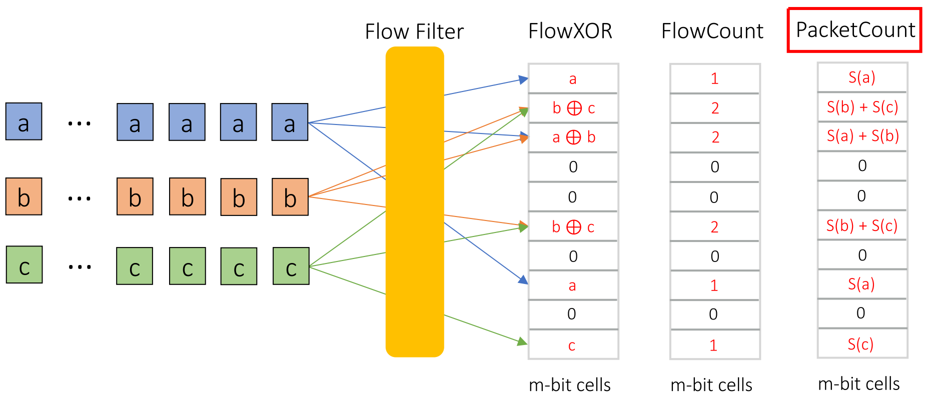


Figure 4: Correct representation of the data structure used in FlowRadar.

### 5.4.3 Collector Decode

Once the switch sends its tables (FlowXOR, FlowCount, PacketCount) to the collector, there are **two stages of decoding**:

1. **Single Decode** (local), performed at the collector level, recovers flows where the data structure is clean enough.
2. **Network-Wide Decode** (distributed), performed across multiple collectors/switches, combines views from many switches to fully decode the remaining flows.

#### Step 1 - Single Decode

**Single Decode** is the **local decoding stage**, it happens at one switch (or collector handling one switch's data). The goal is to **recover as many flows as possible** using **only the local FlowXOR, FlowCount, and PacketCount tables** collected from that switch. The key idea is to find "pure" cells (cells containing information from only one stream) and start the decoding process:

1. **Find a Pure Cell**. A cell is **pure** if `FlowCount = 1`. It means only one flow was hashed to that cell. For example, let the following FlowRadar table, this step identifies the first row:

| `FlowXOR` | `FlowCount` | `PacketCount` |
|---|---|---|
| $a$ | 1 | 5 |
| $a \otimes b$ | 2 | 12 |
| $b \otimes c \otimes d$ | 3 | 13 |
| 0 | 0 | 0 |
| 0 | 0 | 0 |
| $b \otimes c \otimes d$ | 3 | 13 |
| 0 | 0 | 0 |
| $a$ | 1 | 5 |
| 0 | 0 | 0 |
| $c \otimes d$ | 2 | 6 |

2. **Remove the Flow's Contribution from Other Cells**. Flow $a$ was hashed to multiple cells. So now we remove $a$'s effect from all its associated cells:

| `FlowXOR` | `FlowCount` | `PacketCount` |
|---|---|---|
| 0 | 0 | 0 |
| $b$ | 1 | 7 |
| $b \otimes c \otimes d$ | 3 | 13 |
| 0 | 0 | 0 |
| 0 | 0 | 0 |
| $b \otimes c \otimes d$ | 3 | 13 |
| 0 | 0 | 0 |
| 0 | 0 | 0 |
| 0 | 0 | 0 |
| $c \otimes d$ | 2 | 6 |

3. **Removal may produce purer cells**. By removing flow $a$, other cells might now have `FlowCount = 1` (pure cell). Repeat the previous steps until everything is decoded.

   ⚠ **Possible Stall**. Some flows are mixed together in such a way that no cell has `FlowCount = 1`. The solution here is to **apply** the second decoding stage, called **network-wide decode**.

## ✓ Step 2 - Network-Wide Decode

In the previous stage, each switch tries to decode as many flows as it can **locally**, by identifying *pure* cells. But sometimes decoding gets stuck because:

- Flow cells contain multiple flow mixed.
- No pure cells remain.

To solve this, we use **network-wide redundancy**: packets of the **same flow traverse multiple switches**, and those switches may store **different parts** of the encoded data. By combining these views, we can **solve flows that are undecodable at any single switch**.

For example, image the following situation:

| Switch 1 | | | | Switch 2 | | |
|---|---|---|---|---|---|---|
| `FlowXOR` | `FlowCount` | `PacketCount` | | `FlowXOR` | `FlowCount` | `PacketCount` |
| $a$ | 1 | – | | $a \otimes d$ | 2 | – |
| $a \otimes c \otimes d$ | 3 | – | | $a \otimes c$ | 2 | – |
| $b \otimes c \otimes d$ | 3 | – | | $b \otimes c \otimes d$ | 3 | – |
| $a \otimes b \otimes c$ | 3 | – | | $a \otimes b \otimes c$ | 3 | – |
| $b \otimes d$ | 2 | – | | $b \otimes d$ | 2 | – |

In both switches, single decode fails. But when **merged**, we have enough constraints to decode all flows. This is similar to solving a system of equations with more equations that unknowns. In other words, we don't need massive memory in each switch, we can use **small, compressed flow encodings per switch** and decode everything later by **combining views across the network**.

### ❓ But how do we know which switch saw which flow?

To decode accurately, we must know which switch processed which flow, because otherwise we might combine FlowXORs from switches that didn't see a particular flow (wrong result). The **solution is the Flow Filter**. Each switch has a Flow Filter, so we can:

1. Query the flow filter: "Did you see the flow $a$?"
2. If yes, we use that switch's data for decoding flow $a$.

This **guarantees correctness in multi-switch decoding**.

| **Step** | **Description** |
|---|---|
| 1. | Each switch reports its compressed counters (FlowXOR, FlowCount, PacketCount) |
| 2. | Some flows decoded via Single Decode |
| 3. | Remaining flows are solved by combining equations from multiple switches |
| 4. | Use Flow Filters to determine which switch saw which flows |
| 5. | Network-wide correlation fully decodes the remaining flows |

Table 7: Network-Wide Decode summary.

### ⚠ What if Switches Disagree Due to Packet Loss?

Image that:

- A packet of flow $f$ passes through Switch A and Switch B.
- Due to **transient issue**, one of the switches **misses that packet** (e.g., due to mirroring loss or memory overwrite).

Now, when both switches report their Flow Radar data structure:

- The values for flow $f$ might not match across switches.
- This creates **inconsistencies** when trying to decode.

✓ **Solution: Redundancy**. Even if packet counts differ, the set of flows seen by each switch can still be decoded. And more importantly, once we know which flows each switch saw, we can treat each switch's data as a system of linear equations and solve for the actual packet counts.

1. We **use the Flow Filter** to determine which flows each switch saw.

2. We **decode flow IDs** using the FlowXOR and FlowCount tables.
3. We set up a **linear system of equations** per switch:
   - Each **cell** gives us an equation:
     $$\texttt{XOR}\,(f_1, f_2, \dots) \Rightarrow \text{total packet count} = P$$
   - We solve for the **unknown packet counts** of individual flows.
4. If the same flow has **different counts** across switches:
   - It signals a possible **packet loss**;
   - Or a measurement **inconsistency**

## 5.5 In-Band Network Telemetry (INT)

### 5.5.1 What is INT?

In-Band Network Telemetry is a framework where the **data plane itself collects telemetry information** as packets traverse the network. Instead of relying on mirrored copies or external probes (like Everflow or FlowRadar), INT **embeds telemetry instructions directly into packets**. This means that the **packet *asks* switches along its path to record certain metadata** (e.g., delay, queue size, switch ID).

INT demonstrated the power and usefulness of programmable switches. It was one of the first real-world use cases where P4 offered something no traditional switch could do.

#### How it works?

1. Packets carry **INT headers**.
2. INT-capable devices **read the instructions** in those headers.
3. They ***collect*** specific **network state** and ***append* it to the packet** as it moves.
4. The telemetry data is delivered **in-band**, alongside the normal traffic.

The data collected cloud include: switch ID, input and output ports, queue occupancy, timestamp (arrival and departure), packet latency per hop.

#### Why INT?

INT **solves key limitations of older monitoring tools**:

- ✓ **No** need for **mirrored** traffic (like Everflow).
- ✓ **Real-time per-packet** information.
- ✓ **High visibility** into what happens at every hop.

This is especially useful for: fine-grained performance monitoring; diagnostic congestion, jitter, and path problems; dynamic traffic engineering.

### 5.5.2 Modes

Telemetry **data can be collected and exported in different ways**, depending on:

- Whether packets are modified or untouched
- Whether data is inserted in packets or sent to collectors separately
- How much pressure is put on switches vs collectors.

Each mode offers trade-offs between **performance**, **visibility**, and **system overhead**. There are three different modes:

1. **INT-XD (eXport Data)**. Switches do **not modify the packet**. Instead, they **send telemetry** data **directly to the collector** based on local configuration.

   ✓ Pros

   - ✓ **No changes** to the packet ⇒ avoids MTU issues.
   - ✓ Easier for legacy packet flows.

   ✗ Cons

   - ✗ Heavy **pressure on collectors** (they must gather data from all switches).
   - ✗ Telemetry query is based on **switch config**, not what the packet asks.

2. **INT-MX (eMbed instruct(X)ions)**. The **packet is marked** to indicate it wants telemetry. Switches **send telemetry data out-of-band** (to a collector), but **based on the packet's mark**.

   ✓ Pros

   - ✓ Telemetry is **packet-driven**, more dynamic than INT-XD.

   ✗ Cons

   - ✗ Still puts **pressure on collectors**.
   - ✗ Requires **modifying packets**, could affect headers, MTU.

3. **INT-MD (eMbed Data)**. The **packet itself is modified** to carry Telemetry metadata in-band. **Each switch inserts data into the packet** as it passes through.

   ✓ Pros

   - ✓ **No extra pressure** on collectors.
   - ✓ Telemetry query is **packet-dependent**, enabling full visibility along the path.

   ✗ Cons

   - ✗ **Packets are modified**, which may:
     (a) Break some applications;
     (b) Exceed MTU (Maximum Transmission Unit);
     (c) Require special handling at end hosts.

- Use `INT-XD` if we want no packet modifications and can tolerate heavy collector load.
- Use `INT-MK` for moderate flexibility but still out-of-band.
- Use `INT-MD` if we want **maximum in-path visibility** and can handle packet growth.

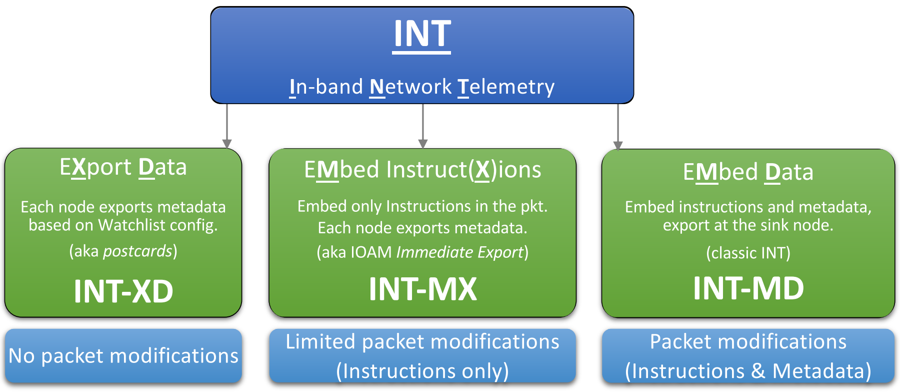

INT
In-band Network Telemetry
EXport Data
Each node exports metadata based on Watchlist config.
(aka *postcards*)
INT-XD
EMbed Instruct(X)ions
Embed only Instructions in the pkt. Each node exports metadata.
(aka IOAM *Immediate Export*)
INT-MX
EMbed Data
Embed instructions and metadata, export at the sink node.
(classic INT)
INT-MD
No packet modifications
Limited packet modifications (Instructions only)
Packet modifications (Instructions & Metadata)

Figure 5: INT modes.

| Mode | Packet Modified | Collector Load | Query Driven By | Pros | Cons |
|---|---|---|---|---|---|
| `INT-XD` | ✗ No | ✗ High | Switch Config | MTU-safe | Inflexible, collector-heavy |
| `INT-MK` | ✓ Yes | ✗ High | Packet Mark | Flexible | MTU risk, collector-heavy |
| `INT-MD` | ✓ Yes | ✓ Low | Packet | No collector pressure | MTU impact, complex parsing |

Table 8: Summary of INT modes.

# 6 Datacenter Layer 3 Load Balancing

## 6.1 Recap: Datacenters

### Characteristics of Workloads

Datacenters support very **different types of applications**, which shape the way networks are designed.

- **HPC (High-Performance Computing)**. Focus on scientific simulations, weather modeling, genome analysis. Workloads: large parallel computations. Require **low latency** and **high throughput** for inter-node communication.
- **Web services**. Think Google Search, Facebook, Instagram. Many short-lived requests/responses. Traffic is bursty, dominated by **small "mice flows"**. Latency-sensitive: users notice if it's slow.
- **Machine Learning (ML)**. Training large models across GPUs/TPUs. Requires frequent **synchronization** (gradient updates). Workloads dominated by **large "elephant flows"** (bulk transfers). Need predictable, low tail latency (slowest worker delays the whole training).
- **Big Data (MapReduce, Spark, etc.)**. Shuffle phase = massive all-to-all data exchange. Demands **high aggregate bandwidth**. Workloads are throughput-oriented, but also sensitive to stragglers.

### Traffic Variability

Datacenter traffic is not uniform. It's a **mixture of short and long flows**, and this mix complicates network design.

- **Short flows (mice)**. Few KBs to MBs, bursty, latency-sensitive. For example: a web request, RPC, or cache lookup. Critical: the user's experience depends on their fast completion.
- **Long flows (elephants)**. Hundreds of MBs to GBs, throughput-oriented. For example: dataset shuffling, ML gradient exchange, backups. Can easily congest network links if not managed carefully.
- **Traffic patterns**
  - **Shuffle traffic**: many-to-many, typical in MapReduce or Spark.
  - **Gradient aggregation**: many-to-one or all-to-all, in ML training.
  - **Incast**: one client requests data from many servers at once → bursts that overwhelm buffers.

Networks must serve both mice and elephants efficiently. Prioritizing one can hurt the other.

### ◎ Goal: High Bisection Bandwidth

Split the network into two equal halves; the **bisection bandwidth** is the total capacity of the links connecting them. Instead, **full-bisection bandwidth** means that every server can communicate with every other at full NIC speed, without bottlenecks.

### ❓ Why it matters?

- HPC: ensures all nodes in a parallel job can exchange data efficiently.
- Web services: prevents hotspots[3] when thousands of requests are routed.
- ML/Big Data: allows large-scale shuffles without stragglers.

The main **challenge** is achieving full bisection bandwidth at low cost, but it requires special topologies (Fat-Tree, Clos, Jellyfish).

In conclusion, datacenters run a **mix of workloads** (HPC, web, ML big data), which produce **variable traffic patterns** (mice vs. elephant flows, shuffle, gradient aggregation). The **main networking goal** is to deliver **high bisection bandwidth** so any-to-any communication can happen efficiently.

[3] A **Hotspot** in a datacenter network is a **point of congestion**, usually a specific link or switch that gets overloaded with too much traffic, while other parts of the network remain underutilized.

## 6.2 Introduction

**Remark: OSI model**

The **OSI (Open Systems Interconnection) Model**, developed by the International Organization for Standardization (ISO), is a conceptual framework that standardizes how different systems communicate over a network. It divides network communication into **seven layers**, each with specific responsibilities, enabling interoperability between diverse systems and technologies.

The Seven Layers of the OSI Model:

1. **Physical Layer**: Handles the physical connection between devices, transmitting raw bits over a medium. It defines hardware elements like cables, hubs, and transmission modes (e.g., simplex, half-duplex). Examples include USB and Ethernet.
2. **Data Link Layer**: Ensures error-free data transfer between nodes on the same network. It manages framing, physical addressing (MAC), and error detection. Devices like switches and bridges operate here.
3. **Network Layer**: Responsible for routing and forwarding data across different networks. It uses logical addressing (IP) to determine the best path for data packets. Routers work at this layer.
4. **Transport Layer**: Ensures reliable end-to-end communication. It segments data, manages flow control, and handles error recovery. Protocols like TCP and UDP operate here.
5. **Session Layer**: Manages sessions between devices, including establishing, maintaining, and terminating connections. It also handles synchronization and recovery.
6. **Presentation Layer**: Translates data into a format suitable for the application layer. It handles encryption, compression, and data formatting (e.g., JPEG, MPEG).
7. **Application Layer**: Closest to the user, it provides network services like file transfer, email, and directory services. Protocols include HTTP, FTP, and SMTP.

**What is Layer 3 Load Balancing?**

**Layer 3 (L3)** means the **network layer** in the OSI model, where **IP routing** happens. In a datacenter with many redundant paths (like a Fat-Tree), there are usually several routes of the **same cost** between a source and destination. **Layer 3 Load Balancing** is the process of:

- Choosing **which path** a packet (or a whole flow) takes among those equal-cost routes.
- With the goal of **distributing traffic** evenly across the network.

Unlike **Layer 4-7 load balancing** (used for applications, web servers, etc.), L3 load balancing doesn't care about *which service* is being accessed. It only cares about **routing packets** across the network fabric efficiently. It's a **network-wide optimization**:

- Not about picking *which server* handles a request,
- But about picking *which path* data takes to reach its destination.

## ? Why L3 Load Balancing is Needed

Modern datacenter topologies (like **Fat-Tree/Clos**) provide **many equal-cost paths** between any two racks. For example: rack A wants to send data to rack B → there may be 8, 16, or more paths that cost the same.

✗ If traffic always follows a single path, some links get congested (hotspots) and other links remain idle (wasted bandwidth).

✓ With **load balancing** the traffic is spread across multiple available paths. It ensures higher **bisection bandwidth utilization** and improves both **throughput** and **latency**.

The **goal** is to maximize the use of the parallel paths by distributing traffic wisely.

## ⚠ Challenges

Load balancing at Layer 3 (IP routing) is not trivial because of **traffic dynamics**:

- **Flows come and go quickly**. Millions of short-lived flows (RPCs, queries) coexist with long-lived flows (backups, ML training). If the algorithm reacts too slowly → short flows finish before rebalancing even happens.
- **Mix of short and long flows**
  - **Mice flows**: small, latency-sensitive, numerous.
  - **Elephant flows**: large, bandwidth-hungry, can dominate a link.

  Balancing both types is tricky:

  - If we spread elephants badly → collisions → hotspots.
  - If we treat mice like elephants → too much scheduling overhead.
- **TCP sensitivity**. TCP assumes packets of a flow arrive in order. If packets of the same flow are split across multiple paths → reordering happens → TCP slows down (false congestion signals). This rules out naïve strategies like packet spraying.

So the challenge is balance traffic in real-time while respecting the nature of **mice vs. elephant flows** and avoiding TCP issues.

## 6.3 Packet Spraying

**Packet Spraying** is a load balancing technique where **each packet** of a flow is sent over a **different path** in the network, instead of keeping the whole flow on a single path. This technique **balances the load immediately** across all available links and ensures that no path is left idle, thus optimizing network capacity utilization. However, there is one **significant issue**: packets of the same flow may arrive out of order due to the varying delays of the different paths taken. TCP is confused by out-of-order delivery and thinks packets are lost, resulting in unnecessary retransmissions and reduced throughput.

### Idea of Packet Spraying

Instead of assigning an entire flow to **one path**, packet spraying sends **each packet** of the flow independently across different available paths. For example, we have 4 equal-cost paths.

- Packet 1 → Path A
- Packet 2 → Path B
- Packet 3 → Path C
- Packet 4 → Path D

The intuition of Packet Spraying: By spreading packets at the *finest granularity*, all network links are used more evenly, and congestion is less likely to form.

### Pros and Cons

#### Pros

- ✓ **Great load distribution**. No path stays idle while another is overloaded. Utilizes all network capacity.
- ✓ **Simple logic**. Doesn't need complex flow classification or scheduling. Just a round-robin or randomized assignment of packets.
- ✓ **Fast reaction**. Even short flows (mice) benefit, because their few packets can be split across paths immediately.

#### Cons

- ✗ **TCP reordering problem**. TCP expects packets of a flow to arrive **in order**. With packet spraying, packets take different paths → different latencies → arrive out of order. TCP interprets out-of-order packets as **loss** → triggers retransmissions, reduces congestion window, lowers throughput.
- ✗ **Hardware complexity**. Switches need per-packet decisions at line rate, which can be expensive.
- ✗ **Not suitable for elephants**. Large flows generate so many packets that reordering overhead becomes very high.

### ⚠ The Reordering Issue (and why Packet Spraying is absolutely avoided in production)

Let's say *flow F* has 3 packets:

- P1 goes on Path 1 (latency 5 ms).
- P2 goes on Path 2 (latency 8 ms).
- P3 goes on Path 3 (latency 6 ms).

They arrive at the receiver as: P1 → P3 → P2. This out-of-order process produces the following errors:

- TCP sees missing sequence numbers (it expected P2 after P1).
- Receiver sends **duplicate ACKs** (saying "I didn't get P2").
- Sender wrongly assumes **congestion/loss** → retransmits and slows down.

As a result, throughput drops and latency increases. CPU is wasted on useless retransmissions. That's why it is **not used in production** as a general solution.

## 6.4 Equal Cost Multi Path (ECMP)

In datacenter networks (e.g., Fat-Tree/Clos topologies), there are **multiple equal-cost paths** between a source and a destination. Traditional IP routing normally picks **one path**, which wastes capacity. **Equal Cost Multi-Path (ECMP)** is the standard mechanism that allows a router/switch to use **all equal-cost paths**.

Main characteristics:

- **Per-flow load balancing**: ECMP does not spray packets individually (like packet spraying). Instead, it ensures that **all packets of the same flow follow the same path** and this avoids TCP reordering.
- **Hashing**: Switches compute a hash of packet header fields (usually 5-tuple: source IP, destination IP, source port, destination port, protocol). The hash value is mapped to one of the available next-hop paths. All packets of the same flow produce the same hash → go on the same path.

In summary, ECMP is a Layer 3 load balancing technique where each flow is assigned to one of the available equal-cost paths using a hash function on packet headers, ensuring packets stay in order and TCP remains happy.

**Example 1**

Say there are 4 equal-cost paths. A hash function outputs values 0-3.

- Flow A (src, dst, IP and port) hashes to 0 → path 1.
- Flow B hashes to 2 → path 3.
- Flow C hashes to 2 → also path 3.
- Flow D hashes to 1 → path 2.

Each flow is consistently mapped to one path. Packets stay in order.

### Hash Collisions and Inefficiency

In ECMP, the hash function maps each flow to one of the available paths. If two or more **large elephant flows** hash to the same path, that path becomes congested and other paths may stay underutilized. This is called a **Hash Collision**. Collisions are harmless for tiny mice flows, but disastrous when multiple elephants collide.

**Why Collisions Matter.** **Elephant flows dominate traffic volume**. Even if 90% of flows are small, the few elephants carry most of the bytes. If elephants collide on the same link, throughput is reduced and latency spikes for other flows sharing that links. It creates **hotspots** while parallel links sit idle.

**Inefficiency.** Hashing spreads flows *randomly*, not *evenly*. With $k$ paths, the load per path can vary widely, especially when:

- The number of elephant flows is small.
- A few unlucky hashes cluster them together.

So the network's **theoretical capacity** is high, but **effective throughput** is lower due to imbalance.

**Example 2: Hash Collisions and Inefficiency**

Imagine 4 equal-cost paths and 3 elephant flows.

- Flow A $\rightarrow$ hashes to path 1.
- Flow B $\rightarrow$ hashes to path 1.
- Flow C $\rightarrow$ hashes to path 3.

Path usage:

- Path 1: 2 elephants (overloaded).
- Path 2: empty.
- Path 3: 1 elephant.
- Path 4: empty.

Outcome:

- Path 1 congests $\rightarrow$ throughput limited.
- 50% of available network capacity wasted (paths 2 and 4 unused).

ECMP's reliance on static hashing leads to **hash collisions**, where multiple large flows land on the same path. This causes **inefficient bandwidth utilization** and **network hotspots**, even though other paths are free.

### There are two problems that ECMP still cannot resolve

Incast and rack skew are two problems that ECMP alone cannot solve. This is one of the reasons that pushes researchers to find a better solution.

- **Bursty Traffic (Incast)**. Incast happens when **one receiver asks for data from many servers at the same time**. For example, a storage node requests blocks from 50 servers; all 50 servers respond **at once** and their packets all converge on the **same final link** to the receiver.

  **Why ECMP doesn't help.** ECMP can spread traffic across multiple *upstream* paths. But the **last hop into the receiver** is always the same physical link. That link suddenly gets a burst of packets from 50 sources.

  **The consequence.** The buffer at that last-hop switch **overflows**. Packets are dropped, so TCP retransmits. Latency increases dramatically.

  So even if ECMP spreads flows earlier in the path, it **cannot prevent congestion at the final bottleneck** in incast scenarios.

- **Flow Skew Across Racks**. Skew means imbalance. Some racks **generate much more traffic** than others, depending on what services run there.

  For example:

  - Rack A: runs a database cluster → produces many **elephant flows**.
  - Rack B: runs lightweight web servers → produces mostly **mice flows**.

  ECMP hashes flows randomly, but Rack A already has more elephants, so its outgoing paths are **more likely to get congested**; Rack B's paths stay underutilized.

  **Why this is a problem.** ECMP doesn't adapt to traffic intensity differences between racks. It treats all flows equally, ignoring that some racks are "heavy hitters". So persistent **hotspots near busy racks** and wasted capacity elsewhere.

### ECMP in Production and Its Limitations

**Why ECMP Was Adopted.** There are three main reasons:

- **Industry standard**: ECMP is built into traditional routing protocols (OSPF, IS-IS, BGP).
- **Easy to deploy**: No special hardware or centralized controller needed.
- **Good enough for mice flows**: in web workloads (many small flows), ECMP spreads traffic fairly evenly. Avoids TCP reordering (a major plus over packet spraying).

⚠ **Observed Problems in Production.** But at hyperscale (tens or hundreds of thousands of servers), **ECMP inefficiencies become visible**:

- **Static & oblivious to congestion**. ECMP only looks at header hashes, not at link utilization. A congested link may still attract new elephant flows while other links remain idle.
- **Flow collisions**. In large topologies, even with thousands of equal-cost paths, collisions between elephants are common. A few unlucky hashes waste a lot of bisection bandwidth.
- **Wasted capacity**. Studies (e.g., on fat-tree topologies with $\approx$ 27k hosts) showed ECMP could waste **over 60% of available bisection bandwidth** on average due to imbalance.
- **Long-lived collisions**. Once a flow is hashed to a path, it stays there. If that assignment is bad, the flow suffers for its entire lifetime.

ECMP became the **default production solution** because it's simple, distributed, and TCP-friendly. But at datacenter scale, its **static, hash-based nature** makes it inefficient, prompting research into **smarter, traffic-aware load balancers** like Hedera (SDN-based, page 96) and HULA (P4-based, page ??).

## ✓ Pros and ✗ Cons

### ✓ Advantages

- ✓ **Simplicity**. Uses a straightforward hashing mechanism. No central controller or complex scheduling needed. Easy to implement in commodity switches.
- ✓ **Avoids packet reordering**. All packets of the same flow follow the same path. TCP sees packets in order, so it doesn't mistakenly trigger retransmissions.
- ✓ **Good for many short flows (mice)**. With millions of short, random flows, the hashing tends to spread them fairly well. This makes ECMP very effective in web-service workloads where flows are small and numerous.
- ✓ **Scalability**. ECMP is distributed: each switch does hashing locally. No centralized bottleneck, works across large-scale datacenters.
- ✓ **Widely supported**. ECMP is built into IP routing standards (OSPF, IS-IS, BGP). Already deployed in real datacenters today.

### ✗ Disadvantages

- ✗ **Hash collisions**. Two or more elephant flows (large flows) may hash to the same path. Result: some links get congested while others are idle $\rightarrow$ creates hotspots.
- ✗ **No congestion awareness**. ECMP assigns paths purely based on hash, not on current load. If one link is already overloaded, ECMP doesn't know $\rightarrow$ it may keep adding new flows there.

- ✗ **Unfairness**. Mice flows are fine, but a single elephant can dominate a link if unlucky with its hash. Other elephants hashed to that path suffer, while bandwidth on other links is wasted.
- ✗ **Static behavior**. Once a flow is mapped, it stays on that path until it finishes. ECMP doesn't migrate flows if conditions change.

ECMP is simple, scalable, and TCP-friendly, which is why it's the default in datacenters. But it's also **static and oblivious to congestion**, so it can lead to **hotspots** when elephant flows collide.

**Deepening: How ECMP Uses Hashing to Pick a Path**

Three steps:

1. **Multiple Equal-Cost Paths Exist**. Imagine a datacenter topology (like a Fat-Tree). From server **S1** to server **S2**, the routing protocol (e.g., OSPF, IS-IS) discovers that there are $k$ **different next-hop paths** that all have the **same cost**. For example, 4 paths, so the next-hops are `{N1, N2, N3, N4}`. The routing table on the switch stores:

   Destination S2 → Next-hops: N1, N2, N3, N4 (all cost 10)

   So, the switch knows it *can choose* any of them.

2. **Switch Computes a Hash of the Flow**. When a new packet arrives, the switch looks at the **flow identifier** (usually 5-tuple: source IP, destination IP, source port, destination port, protocol). It computes a **hash function**, for example:

   ```
   hash = H(srcIP, dstIP, srcPort, dstPort, proto)
   ```

   This gives an integer value.

3. **Map the Hash to a Next-Hop**. Now comes the key: the switch takes the hash value **modulo the number of available next-hops ($k$)**:

   ```
   path_index = hash % k
   ```

   - If `path_index = 0` → send packet to N1.
   - If `path_index = 1` → send packet to N2.
   - If `path_index = 2` → send packet to N3.
   - If `path_index = 3` → send packet to N4.

   All packets of the same flow have same 5-tuple, then same hash and same path. Instead, different flows have different hashes and likely different paths.

### Why This Works

The **paths themselves aren't "hashable"**. Instead, the switch maintains a *list of possible next-hops* for the destination. The hash selects an **index** into that list. This is why the hash needs to be "uniform" → to spread flows across all next-hops evenly.
In other words, ECMP doesn't hash the paths themselves. It hashes the **flow's header fields** and then uses the hash result to pick an **index** from the list of equal-cost paths in the routing table.

### Why ECMP Uses Modulo on the Hash Value

We want to assign each flow to **one of the available next-hops**. Suppose there are $k$ equal-cost paths (say 4). The switch needs a simple way to map the **huge range of hash outputs** (e.g., 32-bit integer) down to just 4 choices.
**The Problem.** Hash functions produce large numbers (e.g., $0 \dots 2^{32-1}$). But the switch only has a small number of next-hops ($k$ paths). We need a consistent, deterministic way to map "large space $\xrightarrow{to}$ small space".
**The Solution: Modulo.** Compute:

```
path_index = hash(flow_id) % k
```

The result is guaranteed to be in the range $0 \dots k-1$. That matches exactly the **indices of the next-hop list**. For example, with 4 paths:

```
hash(flow A) = 57 → 57 % 4 = 1 → path 2
hash(flow B) = 134 → 134 % 4 = 2 → path 3
hash(flow C) = 29 → 29 % 4 = 1 → path 2

```

**Why Modulo Works Well.** There are three main reasons:

1. **Uniformity**: If the hash function is good, the outputs are "random-looking", so modulo spreads flows fairly evenly across paths.
2. **Deterministic**: Same flow always hashes to the same path.
3. **Simple in hardware**: Modulo is fast and easy for switches to implement.

Modulo is used because it **compresses the large hash space into exactly the number of available paths**. That way, every flow gets assigned to one valid next-hop index.

## 6.5 Hedera: Dynamic Flow Scheduling

We just saw the weaknesses of **ECMP** (page 90):

- It **hashes flows blindly**, without knowing current congestion.
- Multiple **elephant flows** can collide on the same path, creating hotspots.
- Meanwhile, other links stay idle, wasted capacity.

Hyperscale datacenters[4] needed a **smarter, traffic-aware solution** to balance flows dynamically.

### What is Hedera?

**Hedera** is a **dynamic flow scheduling system for datacenter networks**, proposed in a research paper at NSDI 2010. [1]

**Key Insight of Hedera.** Most datacenter traffic volume is carried by a **small fraction of flows** (the elephants). Mice flows (small, latency-sensitive) are numerous but consume little bandwidth. If we can **identify and schedule only elephant flows** intelligently, we can fix most congestion while keeping the system lightweight.

**The Problem Statement.** Hedera addresses this question: "*how can we schedule large flows in a datacenter network so that they are spread across available paths, avoiding hotspots and using full network capacity?*". Specifically:

- Input: a set of **elephant flows** in a multi-path datacenter topology (e.g., Fat-Tree).
- Goal: assign each elephant to a path such that: network utilization is balanced; no link is overloaded while others are idle; small flows are not disrupted.

Hedera's motivation is that ECMP wastes bandwidth by ignoring flow sizes, so it proposes a system that **detects elephant flows and dynamically schedules them across paths** to avoid congestion.

### Hedera Architecture (SDN + OpenFlow, Centralized Controller)

Hedera introduces a **centralized SDN controller** (page 29) that has a **global view of the datacenter network**.

- Switches **report flow statistics** (e.g., which flows they see, how much bandwidth each uses).
- The controller runs a **scheduling algorithm** to compute optimal paths for **elephant flows**.

[4] **Hyperscale datacenters** are very large cloud facilities designed to support tens of thousands of servers and millions of virtual machines, built with uniform, modular infrastructure (compute, storage, networking) that can scale out efficiently to meet massive and dynamic workload demands.

- The controller then **installs forwarding rules** in the switches using **OpenFlow** (page 31).

So instead of random per-flow hashing (ECMP), Hedera makes **explicit scheduling decisions**.

The **key components** of the Hedera architecture are:

- **Commodity switches (OpenFlow-enabled)**: forward packets based on flow rules (see below) and export flow-level statistics (e.g., byte counts, duration) to the controller.
- **Centralized Controller (Hedera brain)**: collects network-wide flow information, identifies elephants and assigns paths to elephants to balance load.
- **Flow Rules**: installed dynamically by the controller into switches. They specify that packets of flow $F$ should go through next-hop $N$.

The detailed workflow is as follows:

1. **Flows arrive**: initially handled by ECMP.

   **Wait, so ECMP isn't being replaced?**

   No! Because Hedera needs a **lightweight default mechanism** for small flows, and a **smarter mechanism** only for the big ones. So, all flows start with ECMP:

   - New flows are hashed to a path immediately $\rightarrow$ very low latency to start forwarding.
   - No controller intervention required.
2. **Switch counters** reveal that some flows are big (exceed a threshold, e.g., $> 10$ MB). Switch counters report flow statistics to the Hedera controller.
3. **Controller detects elephants**. If a flow grows beyond a threshold, it's classified as an elephant. So the **controller computes a better path** for that flow.
4. **Controller installs new rules** in switches via OpenFlow.
5. **Elephants are moved** to less congested paths, while mice stay with ECMP.

This was one of the first real examples of **SDN applied to datacenter load balancing**.

### Elephant vs. Mice Flow Scheduling

First of all, *why do we need different treatment?*

- **Mice flows (tiny, short-lived)**
  - They are **the majority by count** but carry very little total traffic.
  - They finish so fast that trying to schedule them centrally would take longer than the flow itself.
  - They are **latency-sensitive** (user-facing requests).
- **Elephant flows (large, long-lived)**
  - They are **few in number** but carry most of the bytes.
  - If badly placed, they can congest links and hurt many other flows.
  - They are **throughput-sensitive** (bulk transfers, ML gradient sync, big shuffles).

So Hedera's philosophy: **let mice run free (ECMP), but carefully shepherd elephants**.

- **Scheduling Mice Flows**. Mice flows use ECMP (hash-based, per-flow).
  - ✓ Immediate forwarding, no controller involvement.
  - ✓ Keeps latency low.
  - ✓ Scales to millions of flows without overloading the controller.
- **Scheduling Elephant Flows**. The controller monitors flow statistics from switches. A flow is promoted to *elephant* if it exceeds a threshold (e.g., $>$ 10 MB transferred). The controller computes a less congested path for the elephant and installs OpenFlow rules to move it there.

  The main **goal** is spread elephants across available paths, avoid collisions (two elephants on the same path) and increase overall throughput and fairness.

### Flow Demand Computation Algorithm

Once Hedera identifies the **set of elephant flows**, it needs to estimate how much **bandwidth each elephant "wants"** (its demand) and assign flows to paths so that no single link is overloaded, and network utilization is balanced.

- **Estimating Flow Demands**. The **controller** collects statistics from switches: byte counters per flow and flow duration. From this, it computers the **demand**: expected bandwidth requirement of the flow. Demand is not just "how much data so far", it is an estimate of how much the flow *will need* in the near future.
- **Scheduling Algorithm**. Hedera then runs a **demand-aware placement algorithm**:

1. **Construct a demand matrix**, where rows are sources and columns are destinations. Each entry is the total demand (sum of flows) between that source and destination.
2. **Solve a multi-commodity flow problem (approximation)**. A multi-commodity flow occurs when many flows compete for shared resources, such as links. The algorithm tries to maximize utilization while respecting link capacities.
3. **Greedy assignment of flows to paths**. Place the largest-demand flows first; assign them to paths with available capacity; continue with smaller flows, updating remaining link capacity.

## ✓ Strengths and ✗ Weaknesses

### ✓ Strengths

- ✓ **Traffic-aware scheduling**. Unlike ECMP, Hedera looks at actual flow sizes. Elephants are spread across paths, it avoids collisions and hotspots.
- ✓ **Better utilization of network capacity**. Reduces wasted bandwidth and improves aggregate throughput in Fat-Tree/Clos topologies.
- ✓ **Hybrid design (ECMP + centralized scheduling)**
  - ∗ Mice flows: stay on ECMP → simple, fast, scalable.
  - ∗ Elephant flows: centrally scheduled → efficient use of resources.
- ✓ **Proof of concept for SDN in datacenters**. Hedera was one of the **first real SDN applications**. Showed that centralized control could improve load balancing.

### ✗ Weaknesses

- ✗ **Controller scalability**. Collecting flow statistics and computing assignments for many elephants is computationally heavy. Doesn't scale easily to Hyperscale datacenters with millions of flows.
- ✗ **Reaction time**. Detection of elephants takes time (flows must exceed a threshold). By the time scheduling decisions are made, network conditions may already have changed.
- ✗ **Centralization overhead**. All decisions come from one controller. In large networks, this becomes a bottleneck and a single point of failure.
- ✗ **Limited granularity**. Only elephants are scheduled; mice remain random. If mice collectively create congestion, Hedera doesn't help.

Hedera is an **improvement over ECMP** because it solves elephant collisions with centralized, demand-aware scheduling, which improves throughput and fairness. However, **controller overhead and slow reaction times limit its scalability** in real-world, production-scale data centers. This is why subsequent work (e.g., **HULA**) shifted toward **in-switch, decentralized, and faster load balancing** that leverages programmable data planes instead of heavy, centralized control.

# References

[1] Mohammad Al-Fares, Sivasankar Radhakrishnan, Barath Raghavan, Nelson Huang, Amin Vahdat, et al. Hedera: dynamic flow scheduling for data center networks. In *Nsdi*, volume 10, pages 89–92. San Jose, USA, 2010.

[2] Antichi Gianni. Network Computing. Slides from the HPC-E master's degree course on Politecnico di Milano, 2024.

[3] Albert Greenberg, Dave Maltz, Guohan Lu, Jiaxin Cao, Ratul Mahajan, and Yibo Zhu. Packet-level telemetry in large datacenter networks. In *SIGCOMM'15*, August 2015.

[4] Chuanxiong Guo, Lihua Yuan, Dong Xiang, Yingnong Dang, Ray Huang, Dave Maltz, Zhaoyi Liu, Vin Wang, Bin Pang, Hua Chen, Zhi-Wei Lin, and Varugis Kurien. Pingmesh: A large-scale system for data center network latency measurement and analysis. *SIGCOMM Comput. Commun. Rev.*, 45(4):139–152, August 2015.

[5] Yuliang Li, Rui Miao, Changhoon Kim, and Minlan Yu. Flowradar: a better netflow for data centers. In *Proceedings of the 13th Usenix Conference on Networked Systems Design and Implementation*, NSDI'16, page 311–324, USA, 2016. USENIX Association.

[6] Weiwu Pang, Sourav Panda, Jehangir Amjad, Christophe Diot, and Ramesh Govindan. CloudCluster: Unearthing the functional structure of a cloud service. In *19th USENIX Symposium on Networked Systems Design and Implementation (NSDI 22)*, pages 1213–1230, Renton, WA, April 2022. USENIX Association.

[7] Yu Zhou, Chen Sun, Hongqiang Harry Liu, Rui Miao, Shi Bai, Bo Li, Zhilong Zheng, Lingjun Zhu, Zhen Shen, Yongqing Xi, Pengcheng Zhang, Dennis Cai, Ming Zhang, and Mingwei Xu. Flow event telemetry on programmable data plane. In *Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication*, SIGCOMM '20, page 76–89, New York, NY, USA, 2020. Association for Computing Machinery.

[8] Danyang Zhuo, Monia Ghobadi, Ratul Mahajan, Klaus-Tycho Förster, Arvind Krishnamurthy, and Thomas Anderson. Understanding and mitigating packet corruption in data center networks. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, SIGCOMM '17, page 362–375, New York, NY, USA, 2017. Association for Computing Machinery.

# Index