

# CHAPTER 1

## INTRODUCTION

The Internet is adequate for providing connectivity between any two end-hosts across the globe, but is inadequate to provide the high-quality experience end-users expect when interacting with web services today. The reasons for the Internet's shortcomings are numerous. Long propagation delays that are fundamentally limited by the speed of light, transient and persistent congestion on network links, and sub-optimal path selection by the Internet routing protocols to name a few [15, 41, 97, 108].

The Internet's shortcomings are masked and high-quality user experiences are enabled by global networks of datacenters that form "overlays" on top of the Internet. Such an overlay network, commonly called a *content delivery network* (CDN), uses algorithms for optimizing data transfer from the origin of the content to the end-user. These algorithms make several decisions to that end. First, they place content at locations close to end users [41]. Second, to serve each request the "best" server is selected whereby overloaded servers are avoided, and so are servers that are far away from the end-user [56]. Third, if default paths in the Internet are performing poorly, then better alternate paths that are chosen that route traffic via nodes in the overlay network, an approach termed as *overlay routing* [15].

The genesis of this thesis lies in the following question: *what influence do the decisions at the overlay have on the routing strategies of the underlying network, and vice versa?*

This intriguing question is one of long-standing interest in computer science research [106, 95, 69, 42, 54, 121]. Several related questions have been put forth. Does the interaction between network and content delivery yield sub-optimal results? How can we design cooperative mechanisms to leverage these interactions? Do cooperative mechanisms yield benefit in an Internet-like environment? We find that nearly all previous research has focused either on the interaction of overlay routing with the network routing [106, 95], or that of server selection strategies with the underlying routing [69, 42, 54, 121]. Somewhat surprisingly, the role of placement strategies on this interaction has received little attention. Nonetheless, content placement is an important factor in determining the performance and cost objectives of networks and content delivery systems as discussed below.

## 1.1 Placement affects networks and content delivery systems

The role of placement strategies in improving performance and costs of content delivery is well known [107]. Placement strategies commonly replicate content and in doing so, they increase availability of content, and reduce the latency between an end-user and the content. Replication improves user-perceived performance but increases costs, e.g., replicating a large video file increases the storage requirements [19], and replicating dynamic objects increases the cost of propagating updates to all replicas [14]. Placement strategies used today achieve a good cost-performance tradeoff because they exploit the *locality* inherent in real-world workloads [109, 31, 66, 46]. Placement strategies exploit geographic and temporal locality patterns by replicating content at those times and at those locations where the content is more popular, and thereby improve performance while minimizing replication costs.

We observe that placement strategies also play a role in influencing cost and performance of the underlying network. A major cost component of the network is the cost of *provisioning capacity* on links in the network [6]. If a placement strategy replicates content at multiple locations, it enables users to fetch content from nearby locations, thereby reducing the aggregate traffic in the network. This traffic reduction helps network operators cut capacity provisioning costs. To ensure good performance, network operators use *traffic engineering* techniques [50], in which routes are computed based on network topology and traffic demand patterns to avoid congestion hotspots.<sup>1</sup> Placement strategies determine where content is available and therefore shape traffic demand patterns. Thus, placement strategies play an indirect role in traffic engineering decisions, and affect network performance as well.

The observation that placement strategies influence objectives of both networks and content delivery raises several questions that form the motivation of this thesis. How does a given placement strategy, especially widely-used placement strategies, such as least recently used (LRU) caching, affect the objectives of networks and content delivery systems? Can we design new placement strategies that better serve objectives of both content delivery and the underlying network? Finally, what is the relative importance of content placement vs. other decisions, such as server selection and traffic engineering, in determining objectives of networks and content delivery systems.

## 1.2 Thesis statement

We state the thesis statement as follows:

*Content placement is a powerful factor in shaping the network traffic and simple placement strategies are effective in improving cost-, performance- and energy-related metrics for networks and content delivery systems.*

---

<sup>1</sup>We use the terms routing, network routing, and traffic engineering interchangeably.

## 1.3 Research problems

In defense of the above thesis, we investigate four research problems. We study the interaction between network and content delivery in following three scenarios focusing on the role of placement strategies: an Internet service provider (ISP) network in which content is placed at a small number of randomly chosen locations (Section 1.3.1 & Chapter 3), an ISP that controls both the underlying network and content delivery on its network and therefore has full flexibility in designing placement strategies (Section 1.3.2 & Chapter 4), and a datacenter of a content delivery network (CDN) whose goal is energy-minimization, and makes both placement and routing decisions to that end (Section 1.3.3 & Chapter 5). As most network traffic is generated by static content, such as video, audio and images [33], we study the interaction between network and content delivery assuming that content is static. The placement of dynamic content, the fourth problem in this thesis, is considered solely as a content delivery problem, unrelated to its interaction with the network. We design and implement a placement strategy for dynamic content stored across geo-distributed datacenters (Section 1.3.4 & Chapter 6).

### 1.3.1 Performance and cost optimization in Internet service providers

The first problem studies the interaction of ISP traffic engineering with a placement strategy that replicates content at a small number randomly chosen locations. We assume that end-users leverage replicated content by downloading in parallel from all locations. We seek to understand how this interaction affects user-perceived metrics, such as file download time and Voice-over-IP (VoIP) call quality, as well as a traffic engineering metric. We consider several classes of TE strategies, and networks with varying degrees of replication of content. We conduct very large scale experiments simulating ISP traffic as a collection of TCP flows seeking to answer following questions:

- How do TE strategies compare in terms of user-perceived performance metrics?
- What is the effective capacity of different TE strategies to tolerate surges in traffic demand?
- How much benefit does TE give over a simple static routing?

### 1.3.2 Network CDN

A network CDN (NCDN) is an ISP that manages content delivery to users on its network. Unlike a traditional ISP (Section 1.3.1) that control only the routing, an NCDN has flexibility of choosing the placement strategy as well. We classify the possible NCDN strategies into *planned strategies* whose decisions are based on historic content demand patterns, and *unplanned strategies* that are oblivious to historic content demand patterns. We consider several strategies for an NCDN including a planned joint-optimization of placement and routing, an ideal planned

strategy with perfect knowledge of future demand patterns, and a strategy that makes both placement and routing decisions in an unplanned manner. We conduct trace-driven evaluations based on real ISP topologies and content access traces from a leading commercial CDN, Akamai, to answer following questions:

- How do simple, unplanned strategies compare to the ideal planned strategy?
- How much benefit do joint optimization strategies yield over simpler strategies as practiced today, and does the benefit warrant the added complexity?
- What is the relative importance of optimizing placement vs optimizing routing?

### 1.3.3 Energy efficiency of CDN datacenters

The third problem considers the design of an energy-efficient CDN datacenter while ensuring a minimal impact on user-perceived performance either due to cache misses or highly-loaded servers. Our key insight is that jointly optimizing underlying network routing and application-level policies of load balancing and server shutdown can increase datacenter energy savings. We are also exploring novel load balancing and server shutdown policies so that the resulting placement of content on servers ensures high cache hit rates and minimizes impact on user-perceived performance. We plan to conduct trace-driven experiments with traces from a commercial CDN datacenter as well as experiments with a prototype to answer following questions:

- What is the impact of energy-saving strategies on user-perceived performance? How do we design server shutdown and load balancing strategies to minimize this impact?
- How much additional energy savings can be obtain by a joint optimization of network and server energy?
- How do different types of content workloads, e.g, video, downloads, mixture of all traffic types, affect these results?

### 1.3.4 Geo-distributed placement of dynamic content

Several applications today generate dynamic content such as weather, stocks, status updates posted on social media websites [14, 10]. Placement of dynamic content is a challenging problem due to a fundamental cost-performance trade-off: dynamic content replication at multiple locations improves latency of content accesses but increases update propagation costs. State-of-art-systems either require manual placement configuration or use naive placement strategies, such as replicate-at-all-locations or replicate-at- $k$ -random-locations, incur excessively high update cost or sub-optimal latency [36, 99].

Our research problem is to design and implement a system that meets the following requirements. (1) The placement of replicas should minimize request latencies under pre-defined resource constraints across geo-distributed datacenters. (2) The design should not pose a fundamental limitation either on the number of datacenter locations or the number of objects stored in the system. (3) The system should provide flexible consistency semantics to meet requirements of wide-range of applications that need either strong or weak consistency semantics.

## 1.4 Research approach

We highlight the key elements of the approach we take to accomplish our research.

**Experimental methodology:** We take an empirical approach to evaluate our hypotheses, including experimentation with prototypes, and packet- and flow-level simulations. The choice of the experimental methodology is dictated by the research problem in most cases. For example, experimenting with an actual NCDN infrastructure is beyond our resources, therefore our evaluation for that problem uses flow-level simulations (Chapter 4). In evaluating TE strategies based on user-perceived metrics (Chapter 3), we perform packet-level simulations as they help us measure user-perceived metrics more accurately compared to flow-level simulations.

**Data collection:** We have performed extensive data collection from multiple ISPs including a Tier-1 US ISP, and from a leading commercial CDN, Akamai. From each ISP, we obtained point-of-presence-level (PoP-level) topology maps as well as PoP-to-PoP traffic demands represented in the form of traffic matrices. We have collected two sets of anonymized content access traces from Akamai CDN. The first set includes traffic for bit-intensive content such as video, and file downloads from users across the globe; these traces are used in evaluation of strategies for an NCDN. The second set of traces are collected at a single datacenter and include traffic of all content types; these traces will be used in evaluating the design of energy-efficient CDN datacenters. Our CDN traces contain logs of content accesses by millions of users, and reveal demand patterns at the level of individual objects, and therefore help us in evaluating placement strategies accurately.

**Evaluation metrics:** Our evaluation focuses on user-perceived performance metrics over system-level metrics, wherever possible. We find that a user-centric evaluation has potential to yield new insights. For example, Chapter 3 shows that common traffic engineering strategies, yield near-identical user-perceived performance, while the same set of strategies differed by up to  $2\times$  on a system-level metric of maximum link utilization. Along the same lines, we plan to conduct evaluation of end-to-end download performance in our ongoing work on designing energy-efficient CDN datacenters.

## 1.5 Contributions

This section presents a summary of contributions of this thesis.

1. We consider a placement strategy that replicates content at a small number of randomly chosen locations in an Internet service provider (ISP) network; end-users leverage the replicated content by downloading in parallel from all locations. Our experiments evaluate common traffic engineering strategies based on real topologies and traffic matrices. Our key contribution is to show that even a simple placement strategy that allows users to download content from just 2-4 locations enables all traffic engineering schemes to achieve near-optimal capacity to tolerate surges in traffic demand, and even simple static routing to be within 30% of optimal.
2. We model a single entity that controls both network and content delivery as in the case of a network CDN (NCDN). We classify the possible NCDN strategies into *planned strategies* whose decisions are based on historic content demand patterns, and *unplanned strategies* that are oblivious to historic content demand patterns. Our experiments on real network topologies and traces from a large CDN surprisingly show that unplanned placement strategies, e.g., LRU caching, combined with a simple static routing outperform (sometimes significantly) realistic (i.e., history-based) joint-optimization strategies and can achieve network cost and user-perceived latencies close to those of a joint-optimal strategy with future knowledge.
3. We design and implement Auspice, a system for dynamic content replication across geo-distributed datacenters. Our system infers pockets of high demand for a name and uses a heuristic placement strategy to provide low request latency, low update cost, and high availability. An application suited for Auspice is a global name service to store name-to-address mapping of mobile devices. Our extensive evaluation for an expected workload of such a global name service on multiple testbeds shows that Auspice significantly outperforms both commercial managed DNS services as well as DHT-based replication alternatives to DNS.

## 1.6 Organization

Chapter 2 presents the main concepts used in this thesis – traffic engineering techniques used by ISPs, algorithms used for content delivery including those for content placement and for request direction. This chapter also discusses prior work on the interaction between traffic engineering and content delivery.

Chapter 3, Chapter 4 and Chapter 6 present solutions and experimental results for the problems in Section 1.3.1, Section 1.3.2 and Section 1.3.4 respectively. These chapters describe the research we have already completed towards this thesis.

Chapter 5 surveys prior research on the design of energy efficient datacenters, and identifies key research questions that need to be answered in designing an energy-efficient CDN datacenter.

Chapter 7 presents a plan for future research to be accomplished towards the completion of this thesis.

## 1.7 Previous publications and collaboration

**Chapter 3** revises a previous publication: A. Sharma, A. Mishra, V. Kumar, A. Venkataramani. Beyond MLU: An Application-Centric Comparison of Traffic Engineering Schemes. *Proc. IEEE INFOCOM, April 2011*. Aditya Mishra and Vikas Kumar provided invaluable support in performing experiments for this work.

**Chapter 4** revises a previous publication: A. Sharma, A. Venkataramani, R. Sitaraman. Distributing Content Simplifies ISP Traffic Engineering. *Proc. ACM SIGMETRICS, June 2013*. Ramesh Sitaraman provided access to Akamai datasets for this work. A realistic experimental evaluation would not have been possible without these datasets.

**Chapter 6** revises a previous publication: A. Sharma, X. Tie, H. Uppal, D. Westbrook, A. Venkataramani, A. Yadav. A Global Name Service for a Highly Mobile Internetwork. *Proc. ACM SIGCOMM, August 2014*. This work also appears in Xiaozheng Tie's thesis, which describes the same placement algorithm and a simulation-based evaluation of the algorithm. The new material in this chapter includes (1) mechanisms to provide consistency of data and (2) experiments with an implementation of the placement algorithm in an emulation testbed and a geo-distributed testbed. Both Xiaozheng Tie and Hardeep Uppal have contributed to the placement algorithm. Hardeep Uppal, David Westbrook and Arun Venkataramani have contributed in implementing the Auspice system.