Inter-Datacenter Bulk Transfers with NetStitcher

Nikolaos Laoutaris
Telefonica Research

joint work with:
Michael Sirivianos, Xiaoyuan Yang and Pablo Rodriguez

Trend 1 – Fault tolerance to catastrophic failures

Perspectives

James Hamilton's Blog

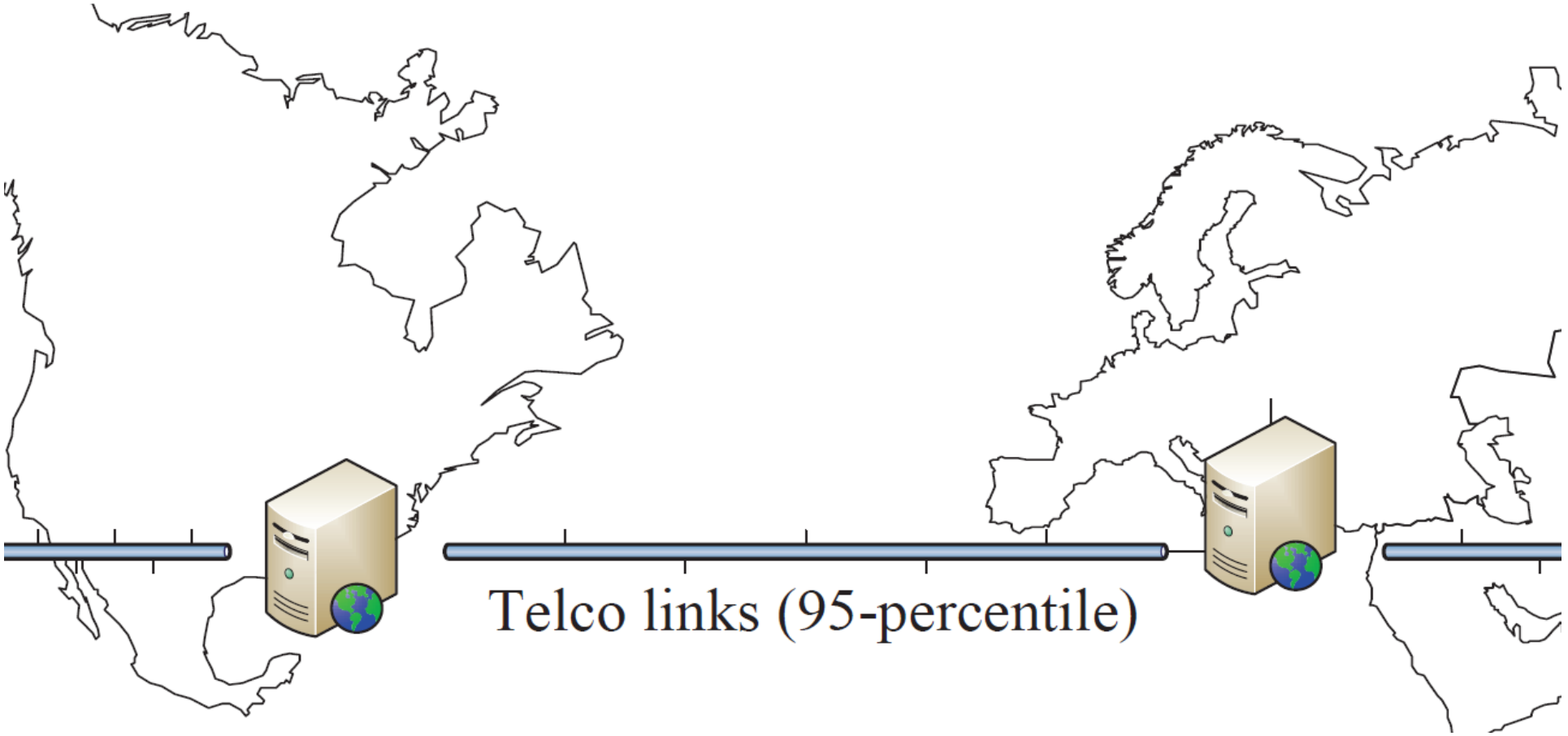
Monday, May 10, 2010

Inter-Datacenter Replication & Geo-Redundancy

With incredible redundancy, comes incredible cost. Even with incredible costs, failure modes remain that can eliminate the facility entirely. The only cost effective solution is to run redundantly across multiple data centers. Redundancy without physical separation is not sufficient and making a single facility bullet proof has expenses asymptotically heading towards infinity with only tiny increases in availability as the expense goes up. The only way to get the next nine is have redundancy between two data centers. This approach is both more available and considerably more cost effective.

Telefónica

Trend 2 – PoP replication for improved user QoS



Telefónica

The Issue: Traffic Growth Is Outpacing Business Growth

A Forrester Consulting Thought Leadership Paper Commissioned By Infineta Systems

The Future Of Data Center Wide-Area Networking

Why Skyrocketing WAN Bandwidth Requires A New Approach To Controlling Traffic

May 2010

FORRESTER

Headquarters | Forrester Research, Inc.
400 Technology Square, Cambridge, MA 02139 USA
Tel: +1 617.613.6000 | Fax: +1 617.613.5000 | www.forrester.com

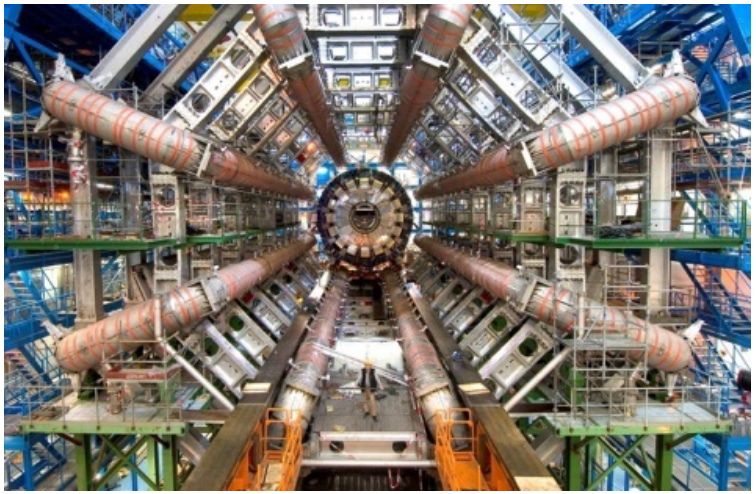
Forrester Consulting
Making Leaders Successful Every Day

Storage, and the replication of it, places enormous demands on the WAN. More than half of our respondents have over a petabyte of storage in their primary data center. Where does the traffic come from? The replication of all that stored data between sites. The vast majority (77%) of the firms we surveyed have three or more data centers running in an active/active or active/passive configuration, and most had replication processes for BC/DR, backups, archives, and data warehouses.

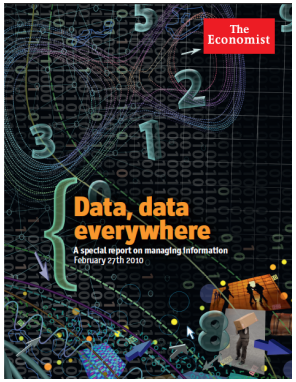
Telefónica

Additional applications

- Scientific computing



- Distributed production/delivery of movies

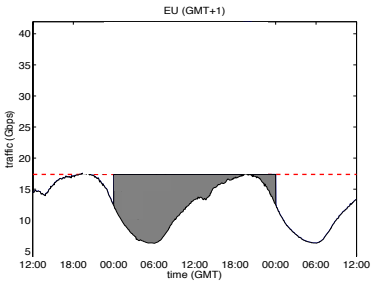


Telefónica

NetStitcher in a nutshell

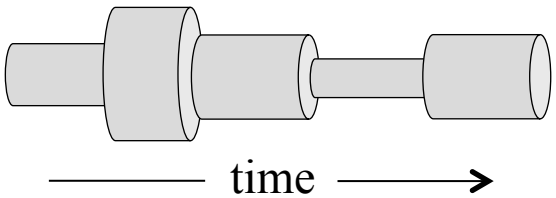
- A system for carrying bulk data for large customers
  - Volume ~ Tbytes / Pbytes
  - Delivery time ~ hours / days

- Main idea
  - Peak load dimensioning & backup paths ➔ lots of leftover bandwidth



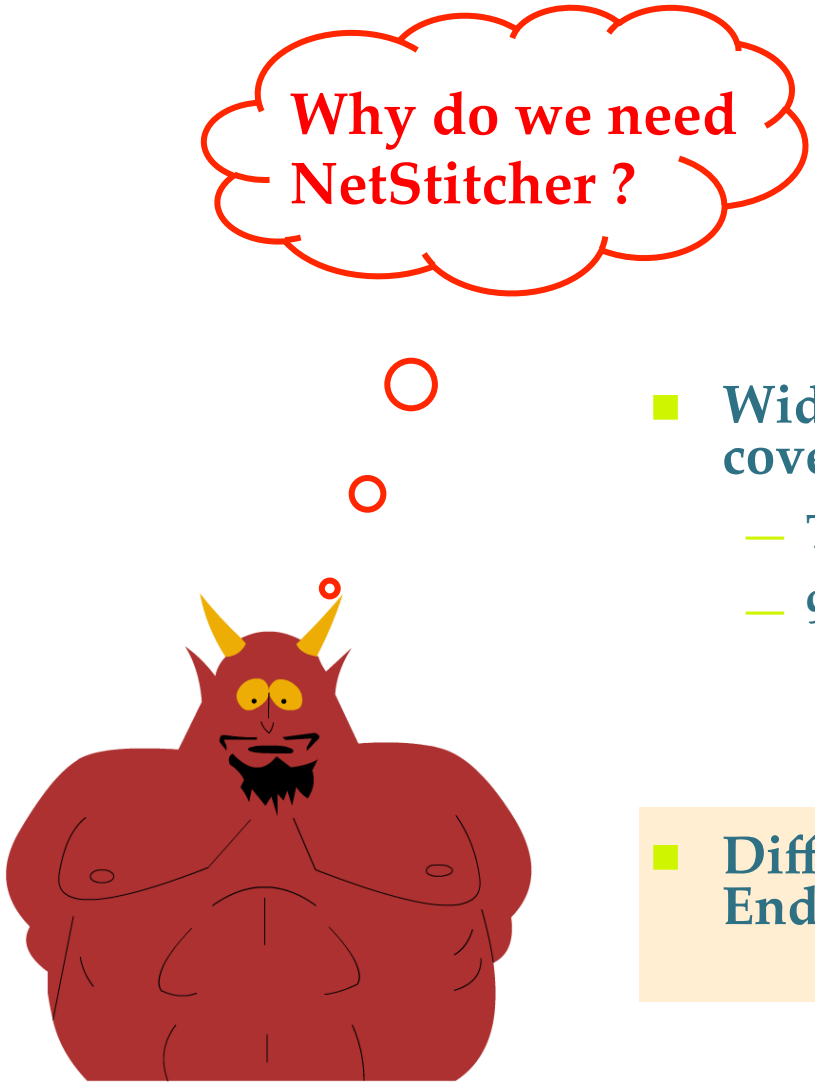
- Create a **volume service** for interconnecting datacenters

X TBs from A to B within the next Y hours



Telefónica

Let's play devil's advocate for a while

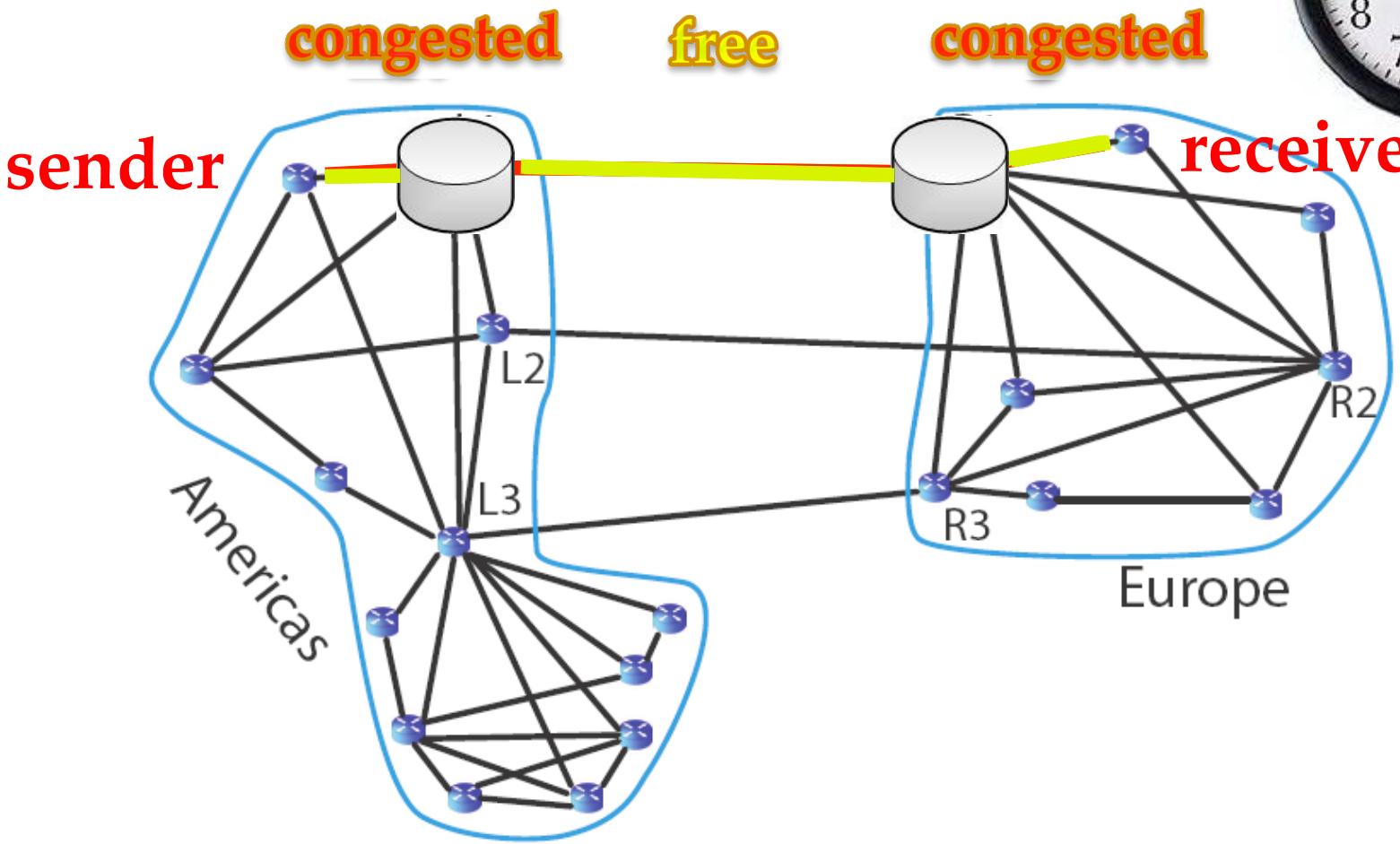


- Wide area bandwidth is not cheap enough to cover traffic growth
  - This applies independently of pricing scheme
  - 95-percentile based, flat, or owned lines

- Difficult to collect leftover bandwidth using End-to-End solutions

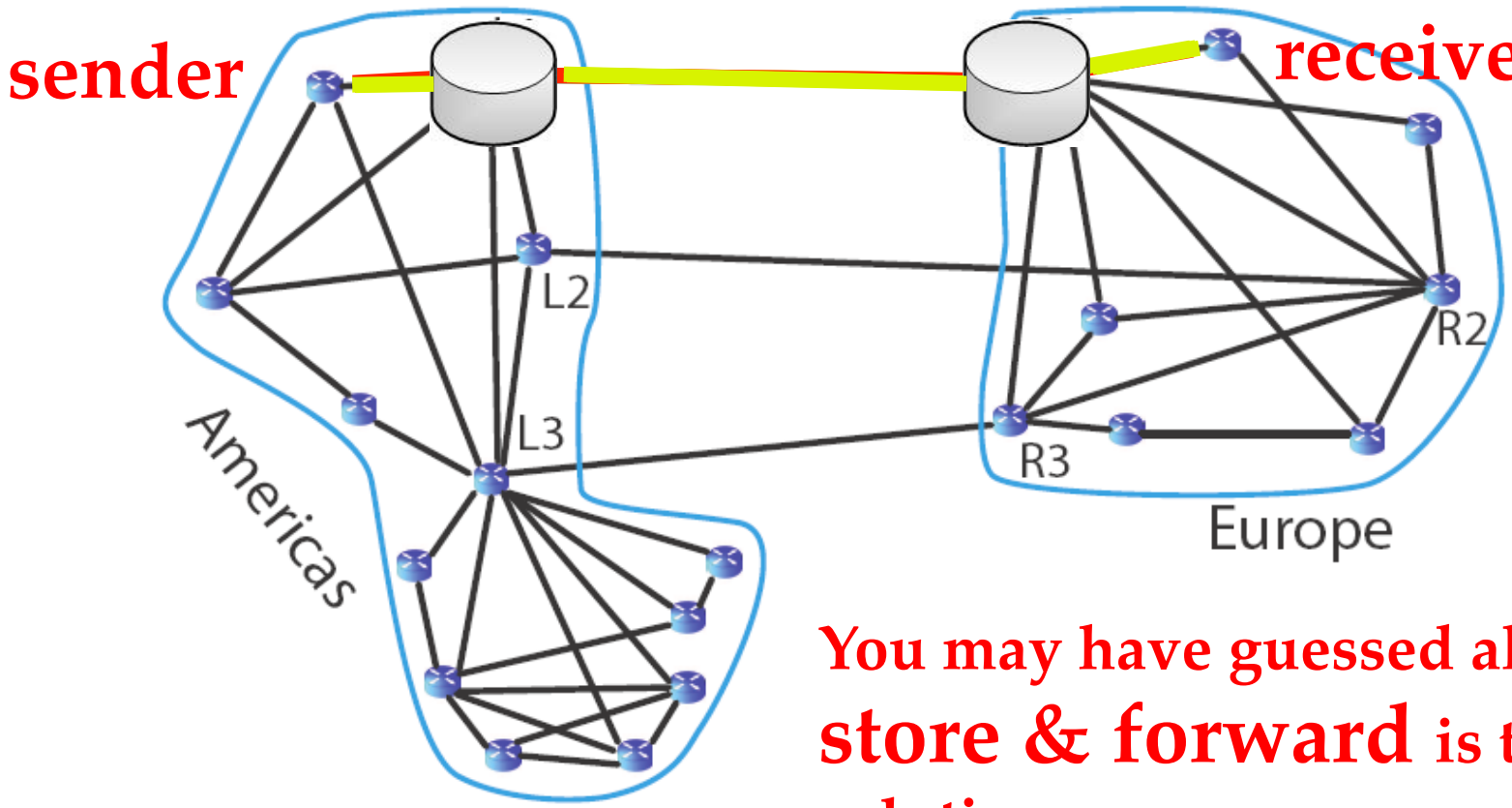
Telefónica

Leftover b/w appears whenever and wherever



Telefónica

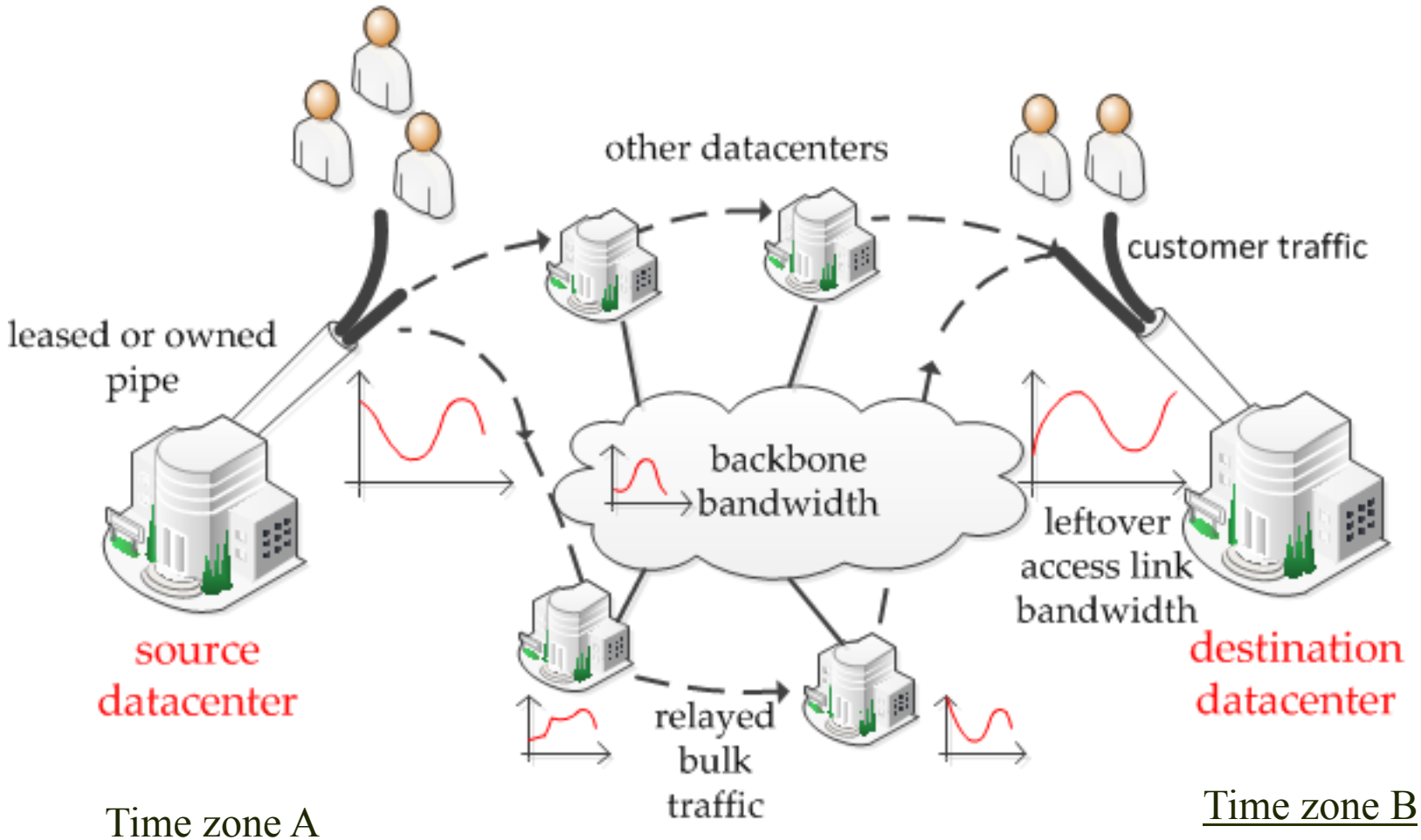
Leftover b/w appears whenever and wherever



You may have guessed already:
store & forward is the solution

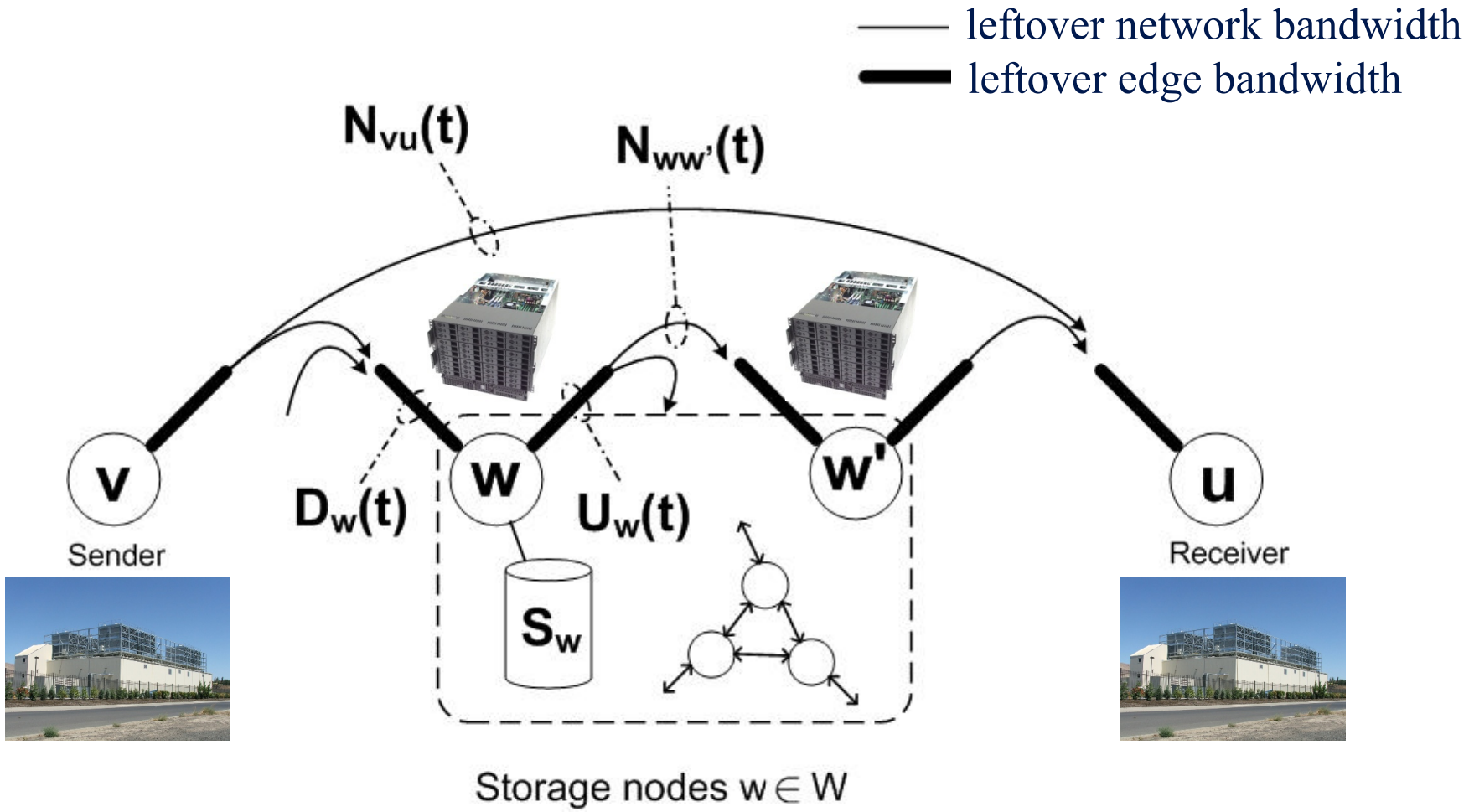
Telefónica

Stitching together leftover bandwidth is tricky



Telefónica

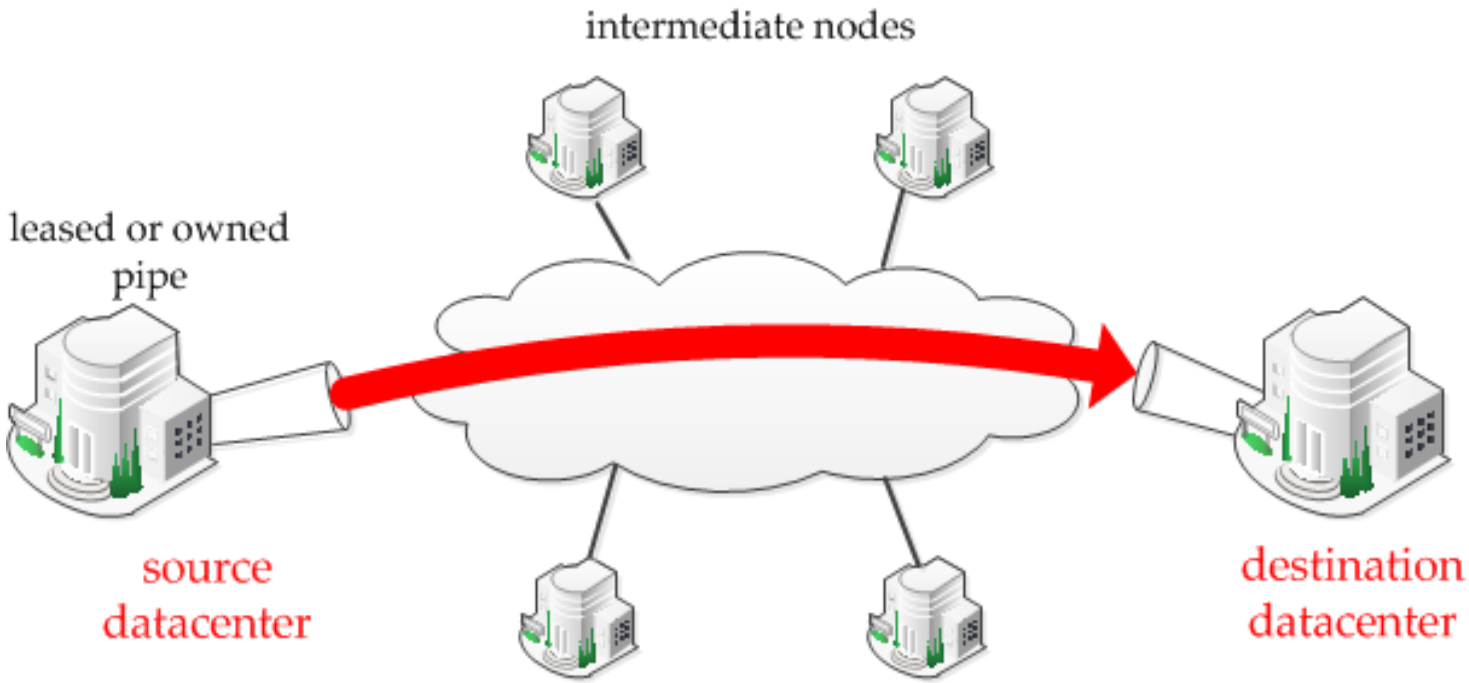
A storage overlay that is aware of leftover bandwidth



Telefónica

NetStitcher's bag of tricks

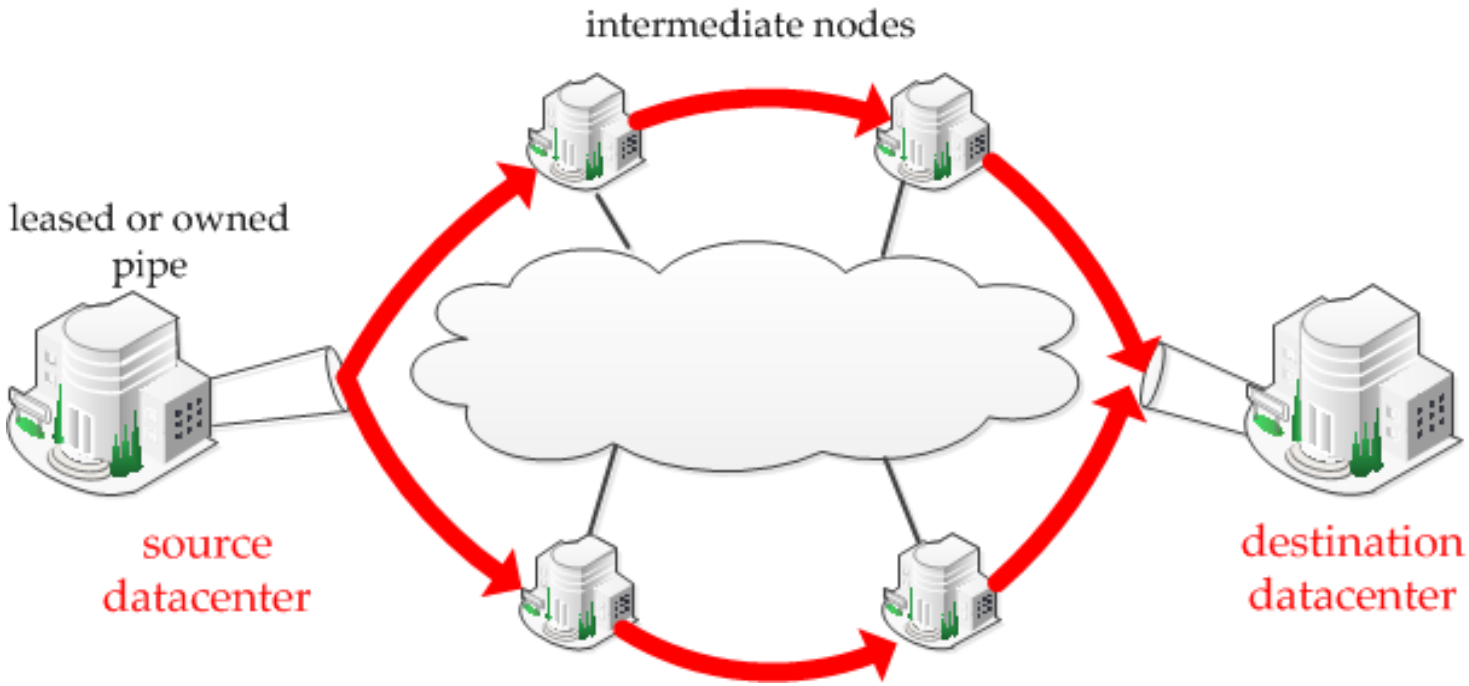
- No in-network constraints and time-aligned sender and receiver bandwidth availability
  - NetStitcher can perform end-to-end transfer



Telefónica

NetStitcher's bag of tricks

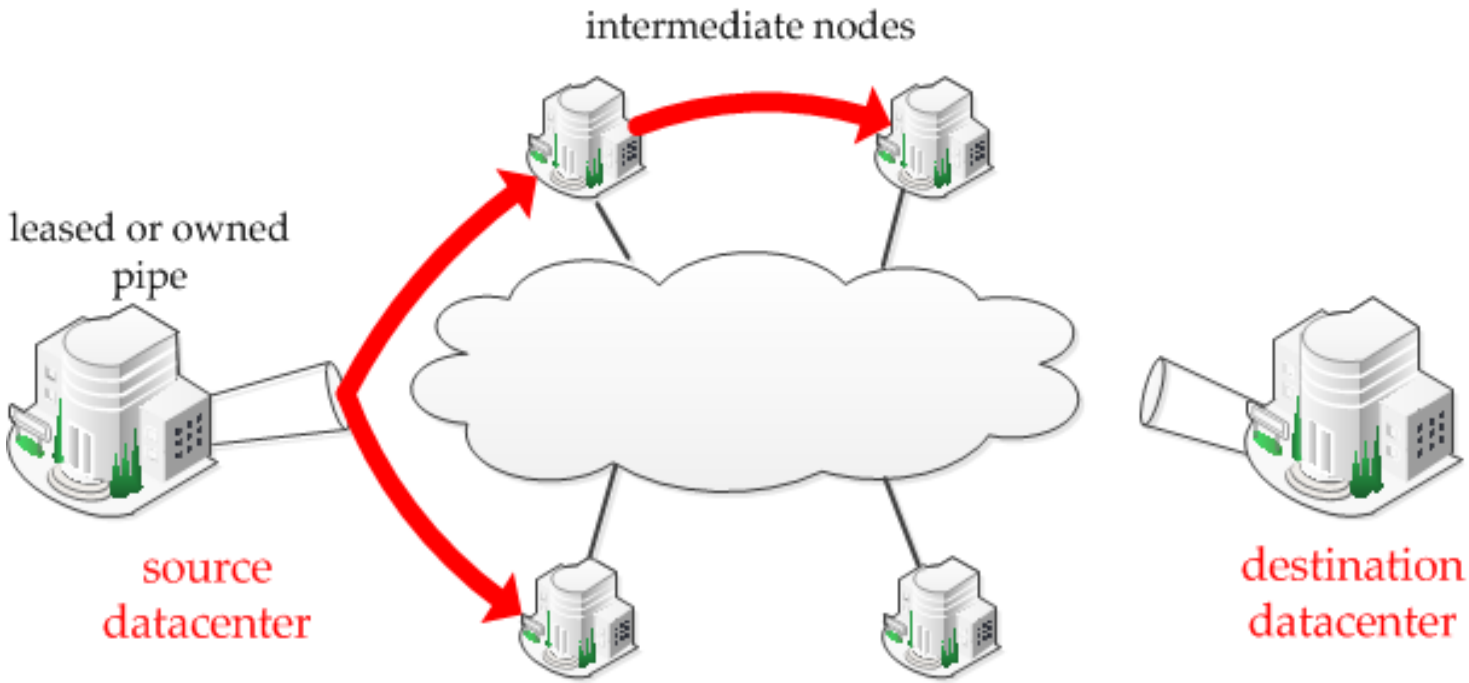
- In-network constraints and time-aligned sender and receiver bandwidth availability
  - NetStitcher can perform multi-path overlay routing



Telefónica

NetStitcher's bag of tricks

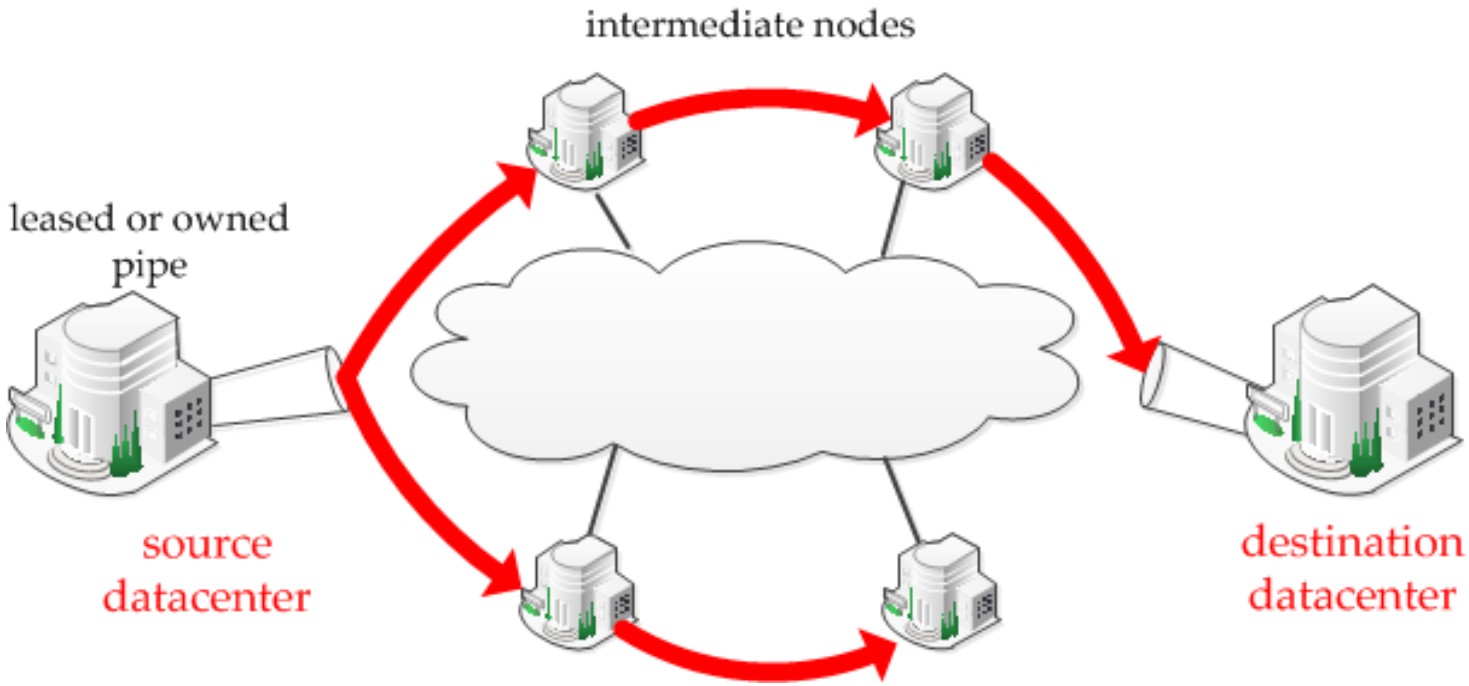
- In-network constraints and misaligned sender , receiver and intermediate node bandwidth availability
  - NetStitcher can perform multi-path and multi-hop store and forward



Telefónica

NetStitcher's bag of tricks

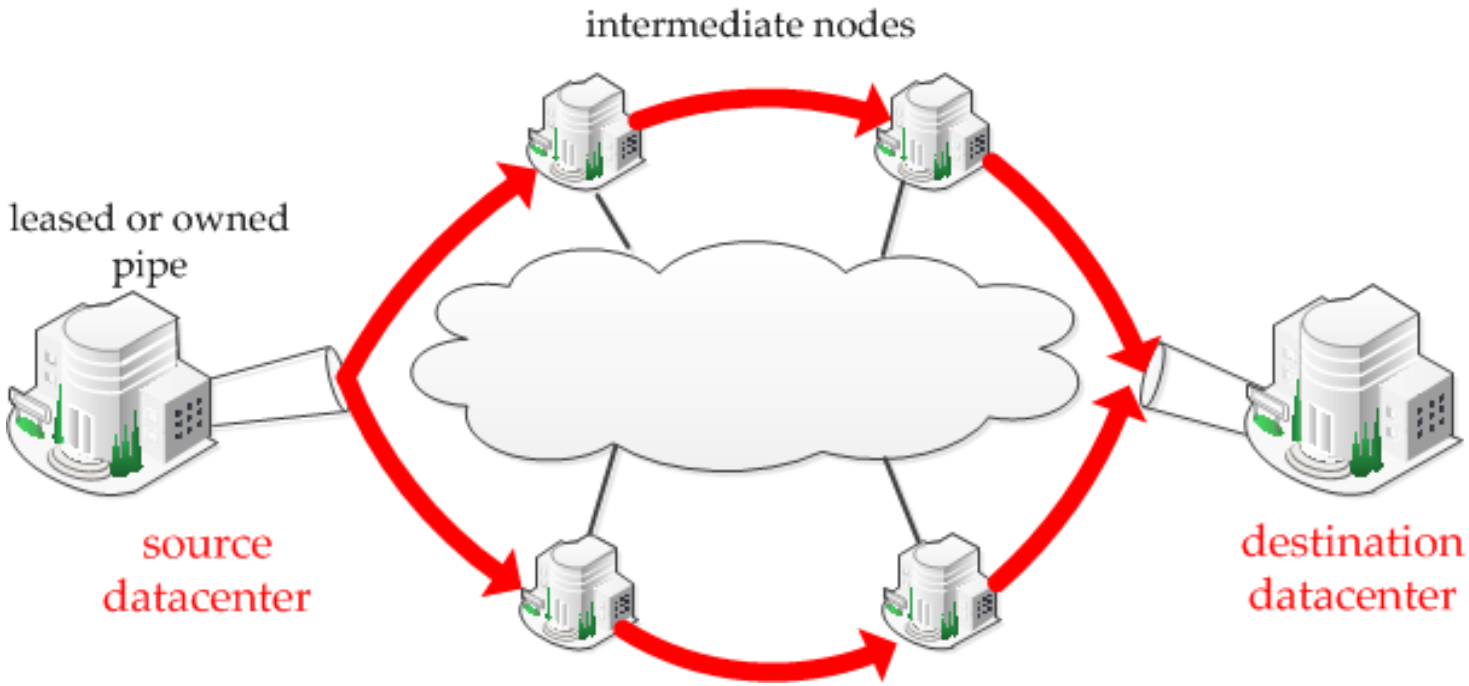
- In-network constraints and misaligned sender , receiver and intermediate node bandwidth availability
  - NetStitcher can perform multi-path and multi-hop store and forward



Telefónica

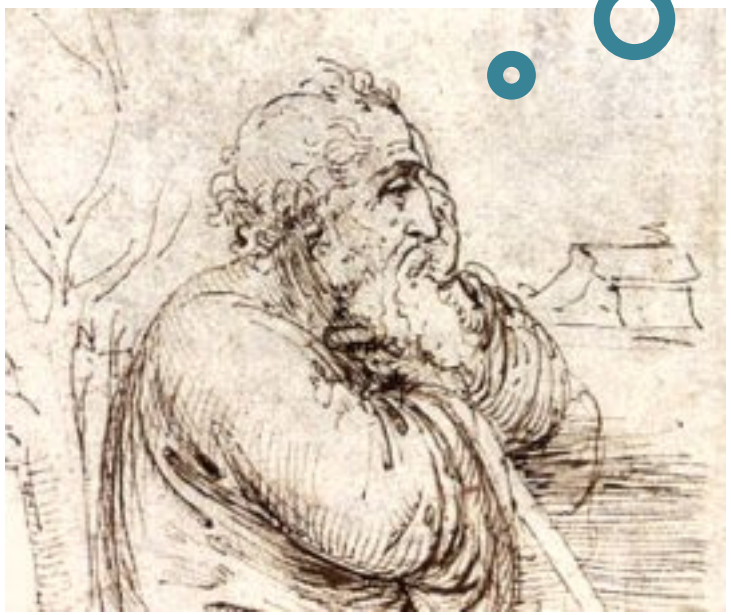
NetStitcher's bag of tricks

- In-network constraints and misaligned sender , receiver and intermediate node bandwidth availability
  - NetStitcher can perform multi-path and multi-hop store and forward



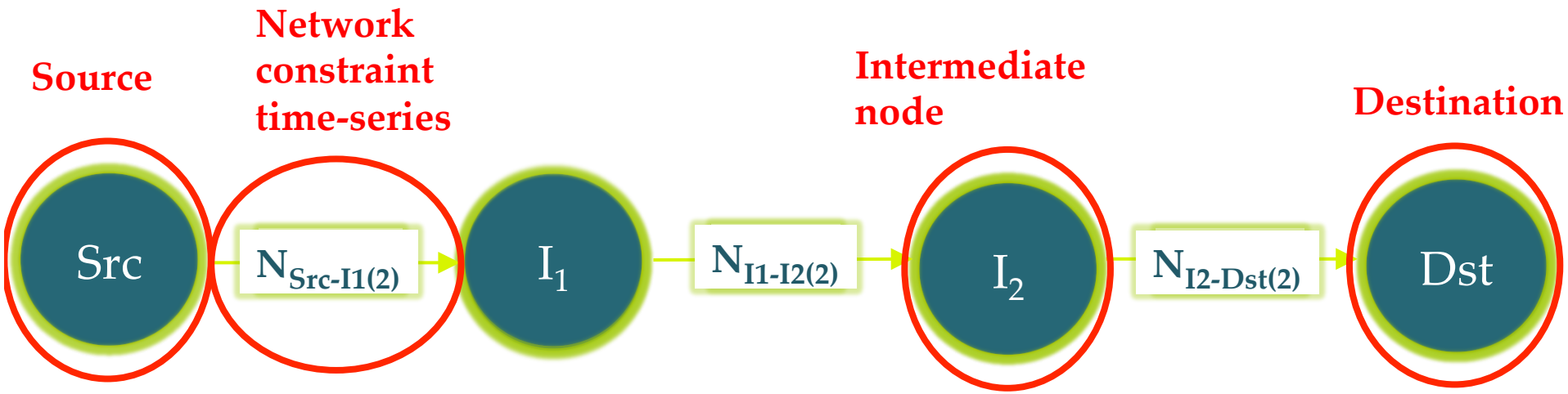
Telefónica

How do we schedule around all these constraints?



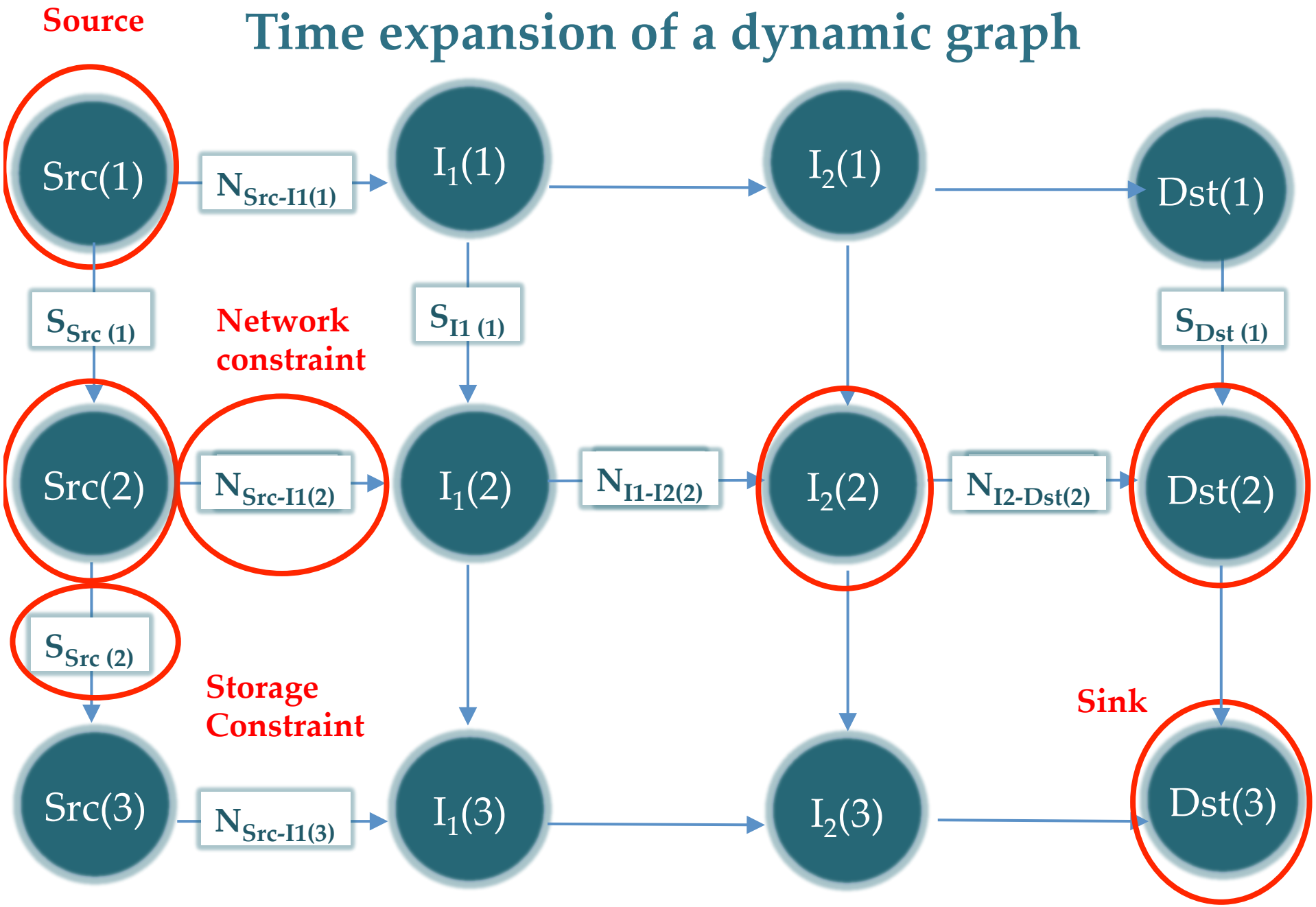
Telefónica

Time expansion of a dynamic graph

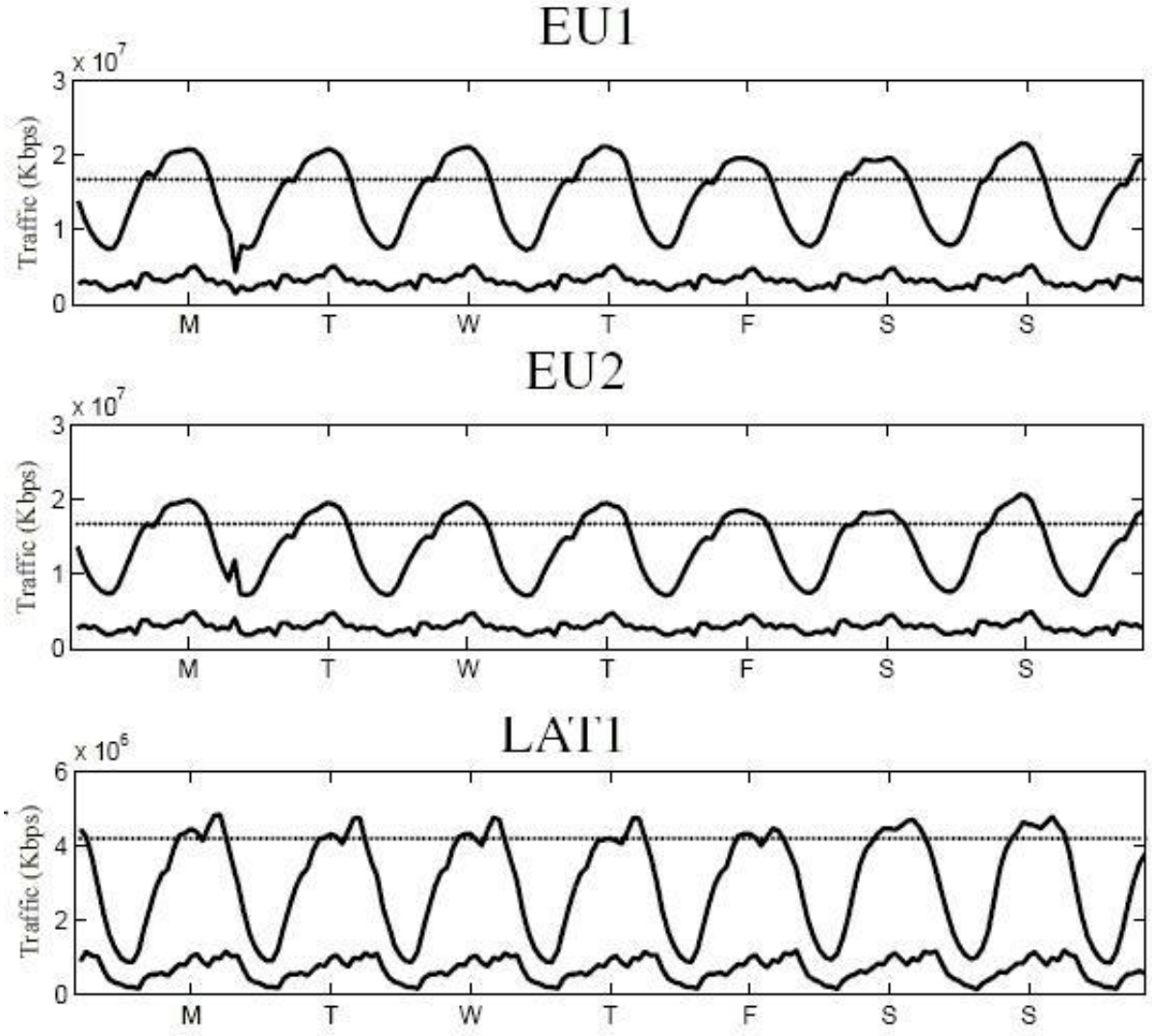


Telefónica

Time expansion of a dynamic graph



But we need to predict the future (of bandwidth)



International backbone traffic

Telefónica

Prediction is easy when data are bulk

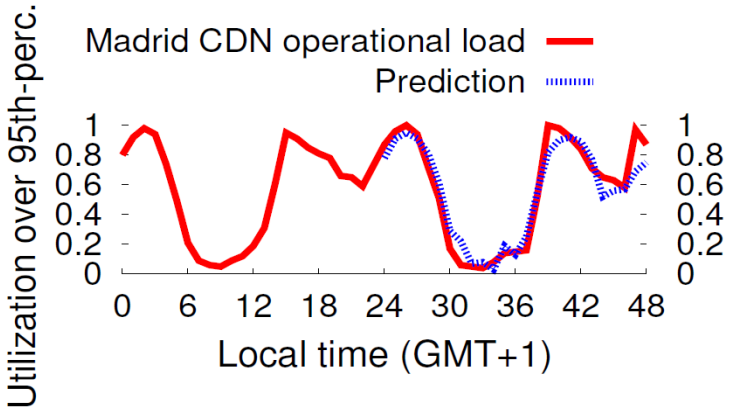
1. **Periodic patterns**

2. **We care about VOLUMES not RATES**

   - VOLUME = $\int_{t_0}^{t_0+T} \text{RATE}(t)\, d(t)$

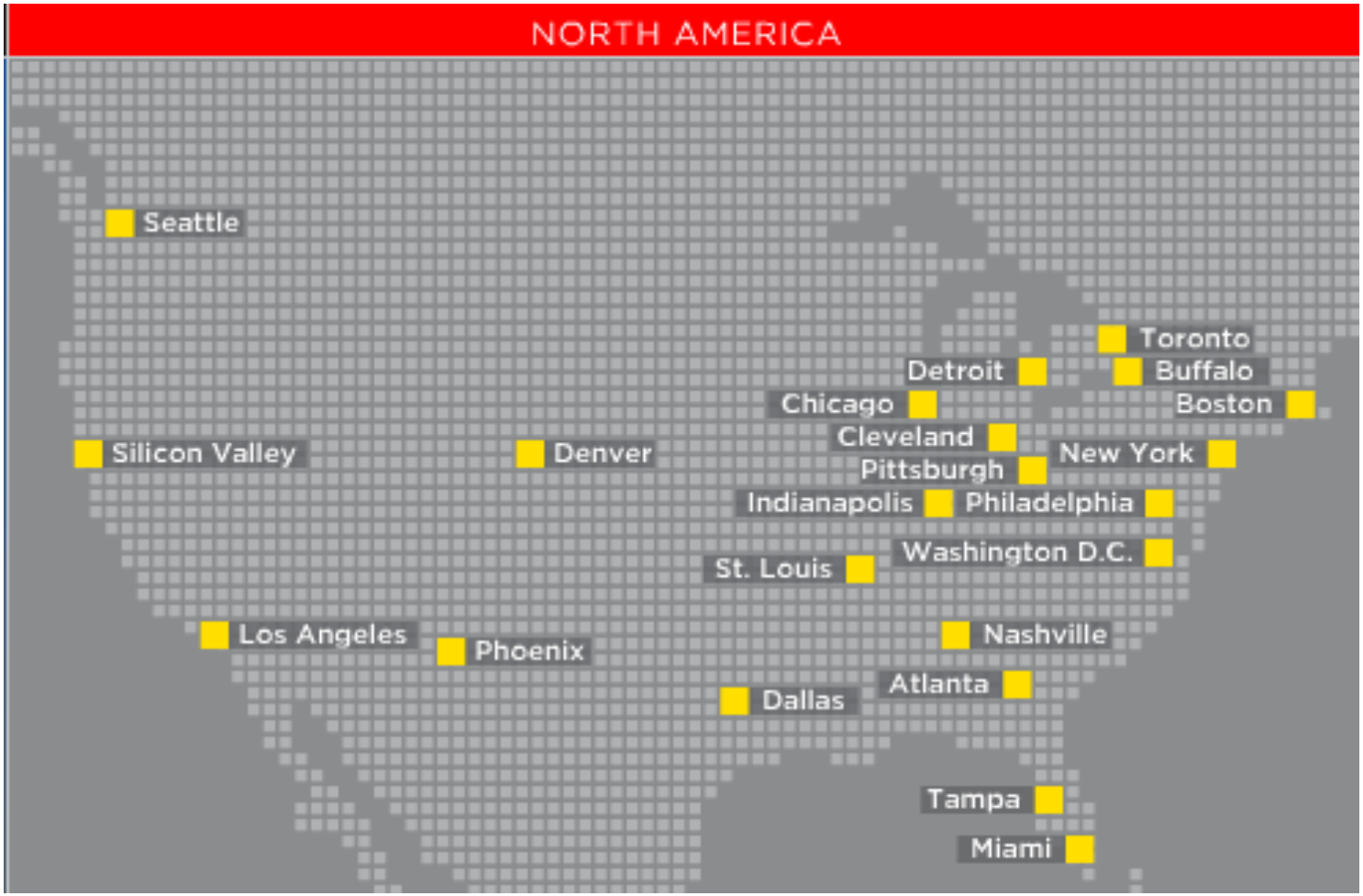
In our NetStitcher implementation we use:

- A simple Sparse Periodic Auto-regression Predictor (Chen et al., NSDI'08)
- Recomputation of transmission schedule
- End-game "pull mode" to handle occasional churn/prediction failure



Telefónica

Case study 1: Equinix datacenters

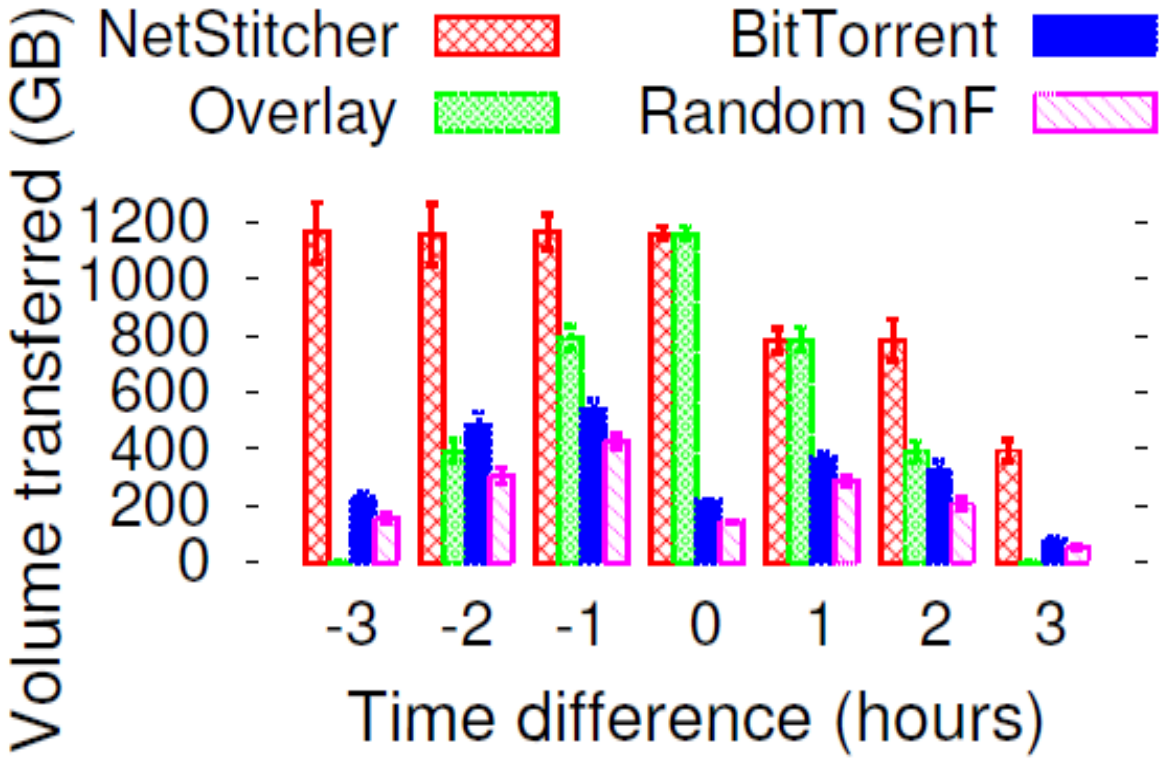


- 62 datacenter at 22 locations all over North America

Telefónica

How much data can we backup?

- 3 hours used for backup (3-6 am local time at the datacenter)
- 1 Gbps network access capacity



NetStitcher can move ×5 more bytes

Telefónica

Case study 2: Telefonica CDN

- 49 servers in Europe, Latin America and USA. GMT-1 to GMT-8
- Need to send a 4.2 TB file over 24h.
- Beyond leftover ➔ 95th-percentile pricing with $7/Mbps/month
- Storage cost: $0.055 GB/month

NetStitcher 80-90% cheaper between Europe & US

Telefónica

Conclusion: A practical application of DTNs

- The utilization of a network can be improved but for this we need:
  1. Delay elastic traffic to go into off-peak hours
  2. In-network storage
  3. High-level knowledge of traffic behaviour around the day

Telefónica

More info at: http://people.tid.es/Nikolaos.Laoutaris