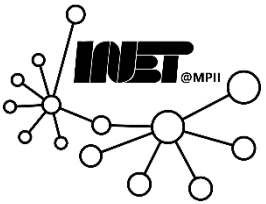
Content Delivery Networks

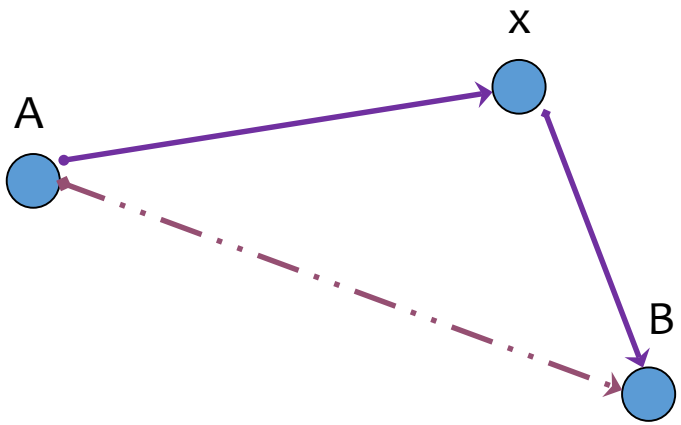
Prof. Anja Feldmann, Ph.D.

(Based on slide deck of Computer Networking, 7th ed., Jim Kurose and Keith Ross.)



Indirection

Indirection: rather than reference an *entity* directly, reference it ("indirectly") via another entity, which in turn can or will access the original entity



> *"Every problem in computer science can be solved by adding another level of indirection"*
> *— Butler Lampson*

Internet Content

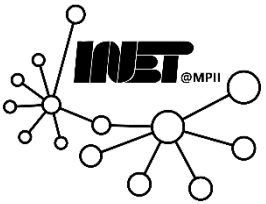
- ***Content***
  - Static web pages and documents
  - Images and videos, streaming, …

Content is King!

- 500 exabytes (10^{18}) created in 2008 alone *[Jacobson]*
- Estimated inter-domain traffic rate: 39.8 TB/s *[Labovitz]*
- Annual growth rate of Internet traffic: ~40%-60% *[Labovitz]*

- Much of web growth due to video

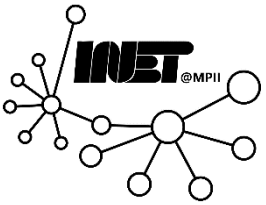
  (e.g, Flash, RTSP, RTP, and YouTube)



Internet Content

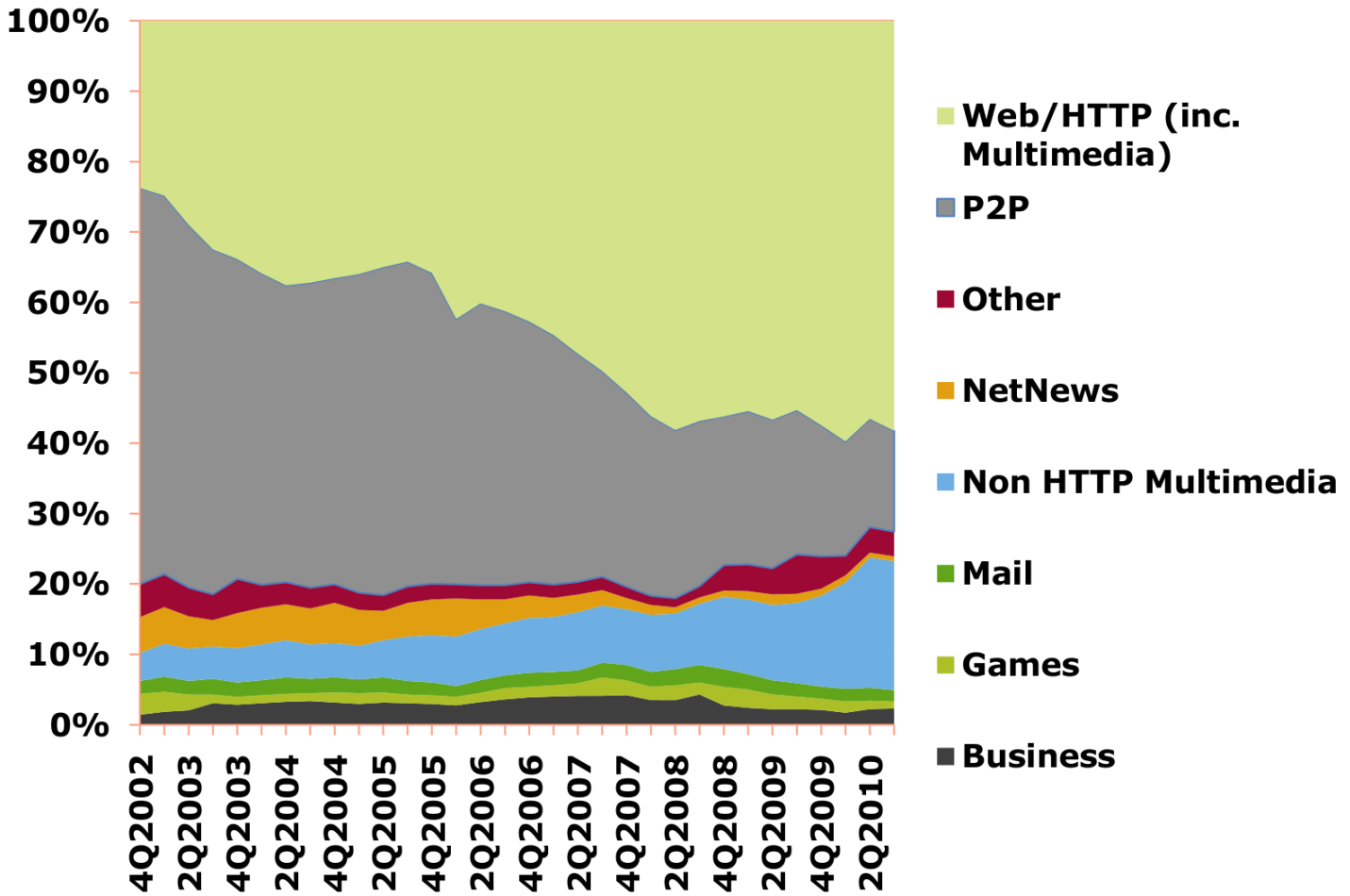
Content is King!

- *How to deliver content?*
- *How to cope with growth of content?*



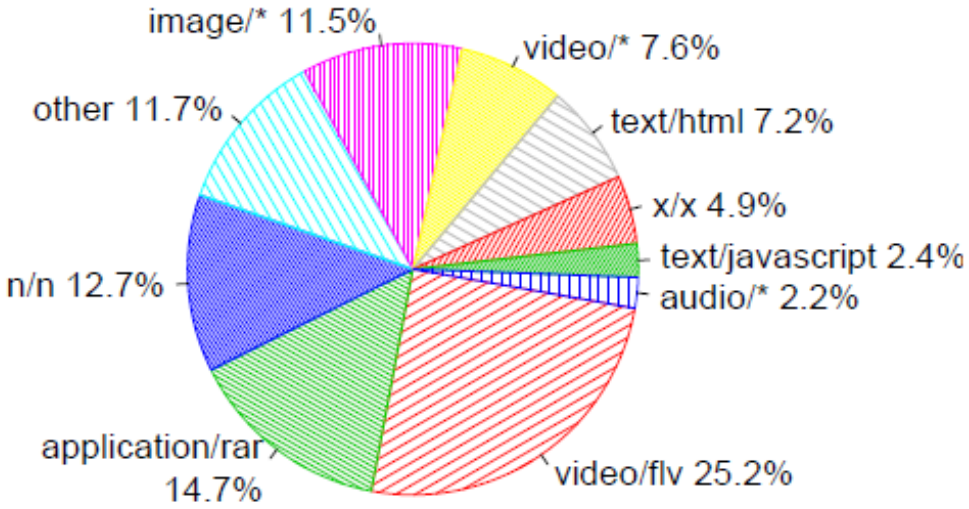
Application mix

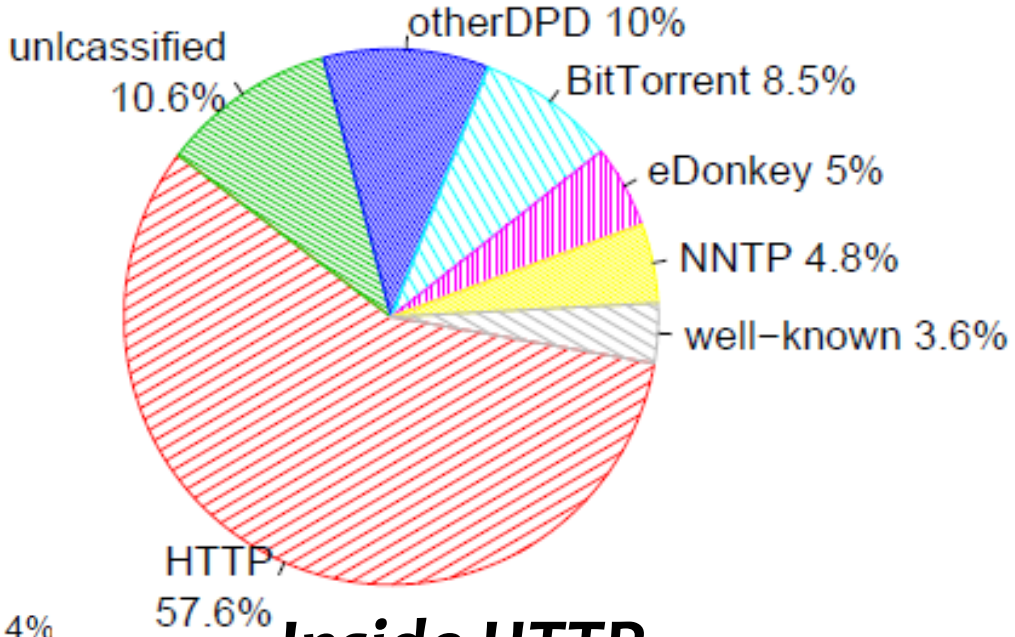
"Traffic Types and Growth in Backbone Networks", Alexandre Gerber and Robert Doverspike, AT&T Labs – Research 2011.



Application mix in 2009

HTTP dominates





Inside HTTP

- *Flash-video dominates*
- *Images and RAR files next*

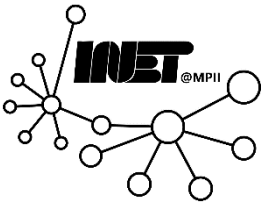
"On Dominant Characteristics of Residential Broadband Internet Traffic", Gregor Maier et al., IMC 2009.

Prevalence of CDNs

Content Delivery Networks (CDNs) will carry 71 percent of Internet traffic by 2021.

- *Seventy-one percent of all Internet traffic will cross CDNs by 2021 globally, up from 52 percent in 2016*

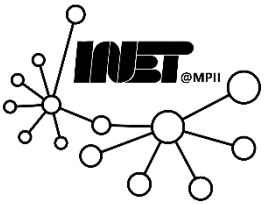
("The Zettabyte Era: Trends and Analysis", Cisco White Paper, 2017)



Serving content without CDN

Enormous demand for popular content

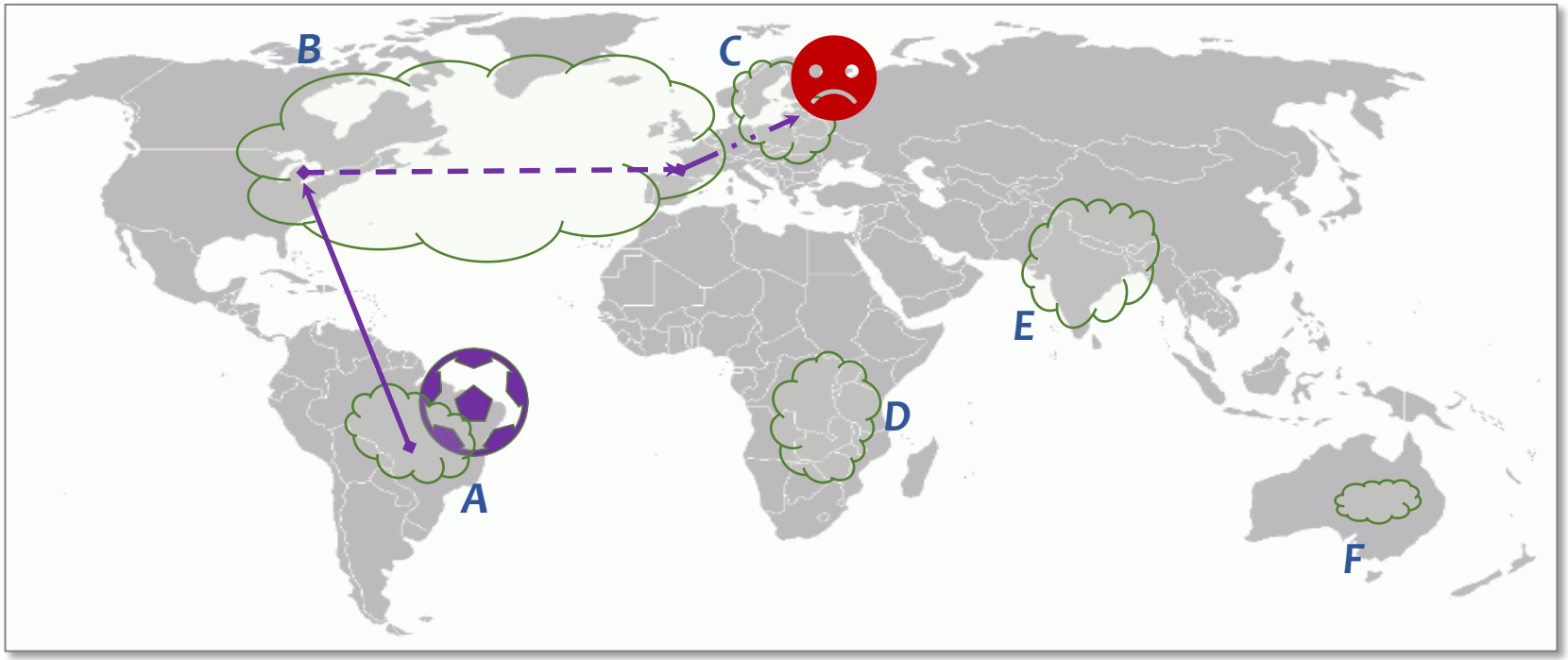
- ***Cannot*** be served from single server!



Serving content without CDN

Bad performance due to large distance

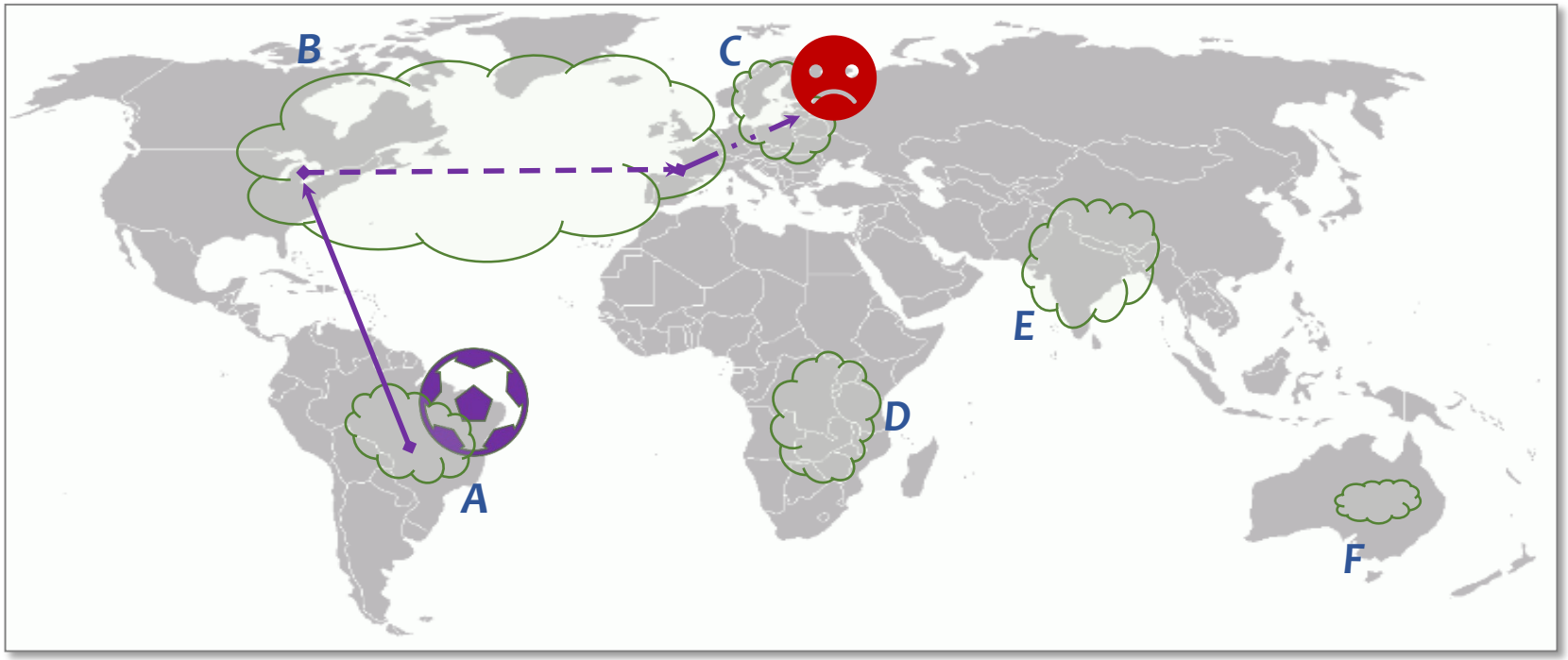
- *TCP-throughput depends on round-trip time!*
- Bad connectivity?



Serving content without CDN

Single point of "failure"

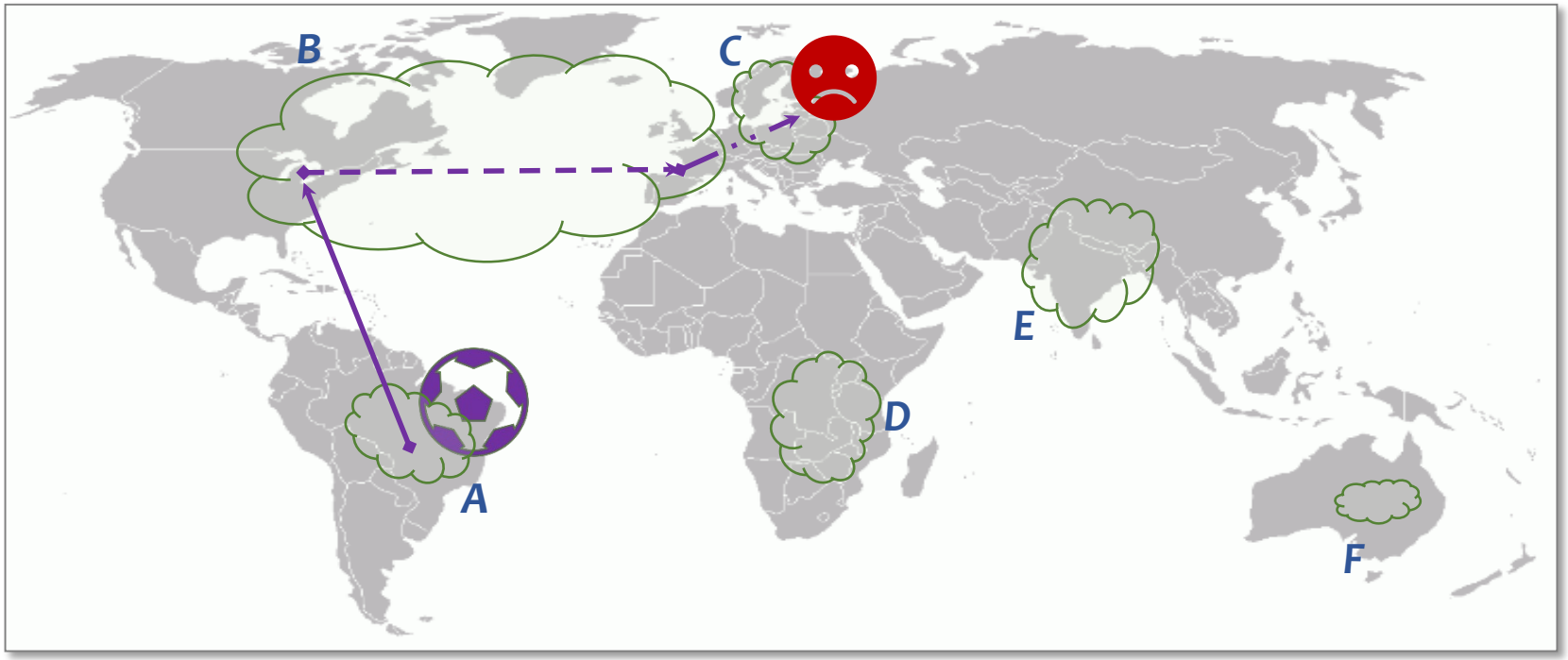
- High demand leads to crashes or high response times (*e.g., flash crowds*)



Serving content without CDN

High costs

- *Bandwidth and disk space to serve large volumes (e.g., videos)*

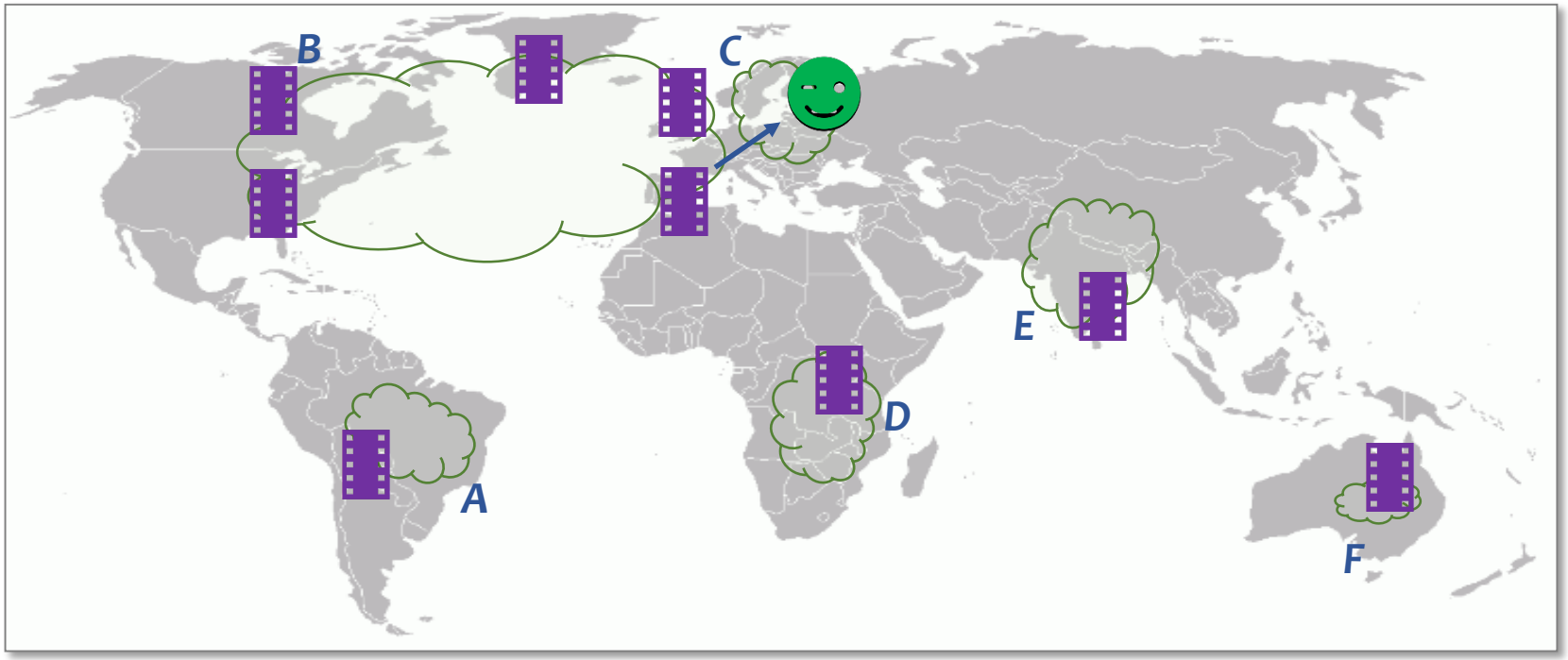


Approaches to content delivery

Replicate content

But where would you serve from?

Locally, or from a "nearby" location

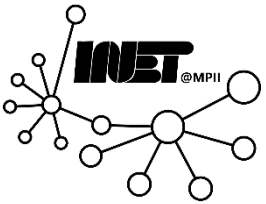


Approaches to content delivery

- ***Centralized hosting***

- ***Content delivery networks (CDN)***
  - ***Offload*** content delivery to large number of content servers
  - Put content servers ***near*** end-users

- ***Peer-to-peer networks***
  - In theory: infinite scalability
  - Yet, ***download capacity throttled by uplink capacity*** of end users

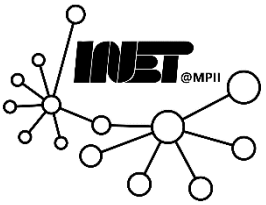


Akamai—one of the largest CDNs

- *Akamai (Hawaiian: "intelligent")*
  - Evolved out of MIT research effort: handle flash crowds
  - ***100,000+ servers*** located in ***72 countries***, ***1000+ networks***
  - Customers: *Yahoo!, Airbus, Audi, BMW, Apple, Microsoft, etc.*
- Why Akamai?
  - Content consumer: Fast download
  - Content provider: Reduce infrastructure cost, quick and easy deployment of network services
- Task of CDNs: Serve content
  - Static web content: HTML pages, embedded images, binaries …
  - Dynamic content: break page into fragments; assemble on Akamai server, fetch only non-cacheable content from origin website

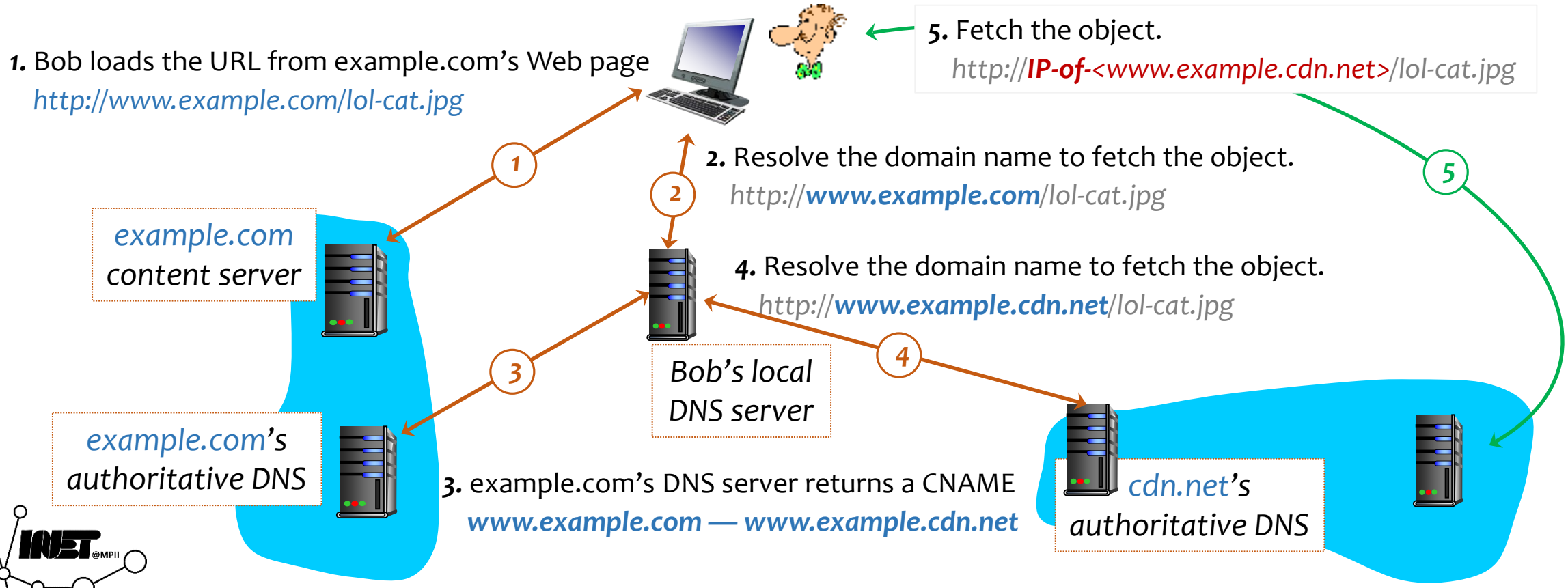
Akamai: Novel idea?

- Local server cluster
  - Bad if data center or upstream ISP fails
- Mirroring
  - Deploying clusters in a few locations
  - Each mirror must be able to carry all the load
- Multihoming
  - Using multiple ISPs to connect to the Internet
  - Each connection must be able to carry all the load
- Akamai vastly increases footprint
  - monitors and controls their worldwide distributed servers
  - directs user requests to appropriate servers
  - handles failures



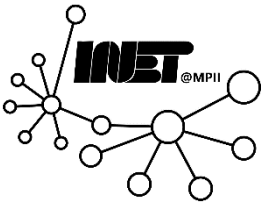
Serving content via CDN

- User requests image at URL *http://**www.example.com**/lol-cat.jpg*
  - Image stored in CDN at *http://**www.example.cdn.net**/12lol34cat56*



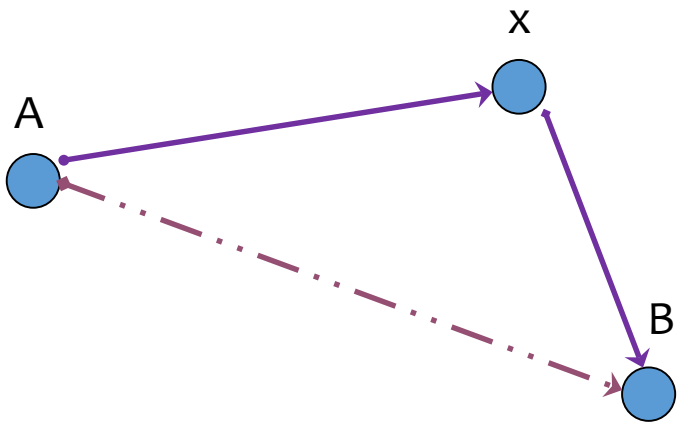
“Best” location?

- ***Service requested***
  - Server must be able to satisfy the request *(e.g., QuickTime stream)*
- ***Server health***
  - Up and running without errors
- ***Server load***
  - Server’s CPU, disk, and network utilization
- ***Network condition***
  - Minimal packet loss, sufficient bandwidth
- ***Client location***
  - Server should be close to client, e.g., in terms of RTT



Indirection

Indirection: rather than reference an *entity* directly, reference it ("indirectly") via another entity, which in turn can or will access the original entity



> *"Every problem in computer science can be solved by adding another level of indirection"*
> *— Butler Lampson*