Overlay Networks

Future Internet Communications Technologies

Prof. Dr. Panagiotis Papadimitriou

Institut für Kommunikations-Technik

Leibniz Universität Hannover

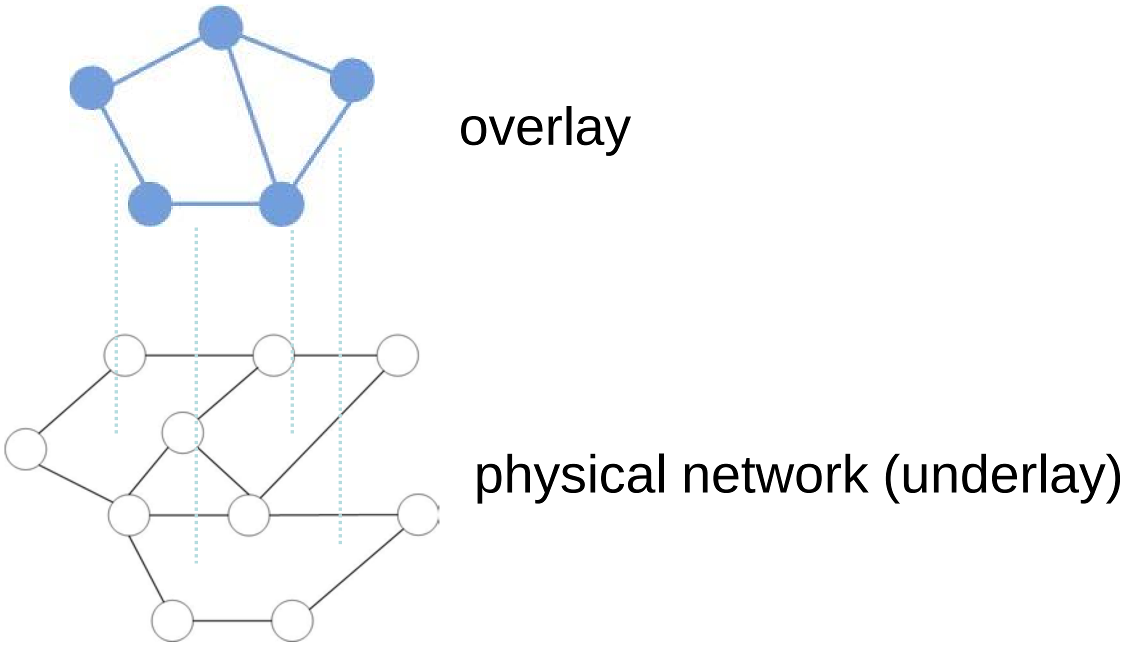
Outline

- Introduction to Overlay Networks
- Resilient Overlay Networks (RON)
- Content Distribution Networks

Introduction to Overlay Networks

Overlays

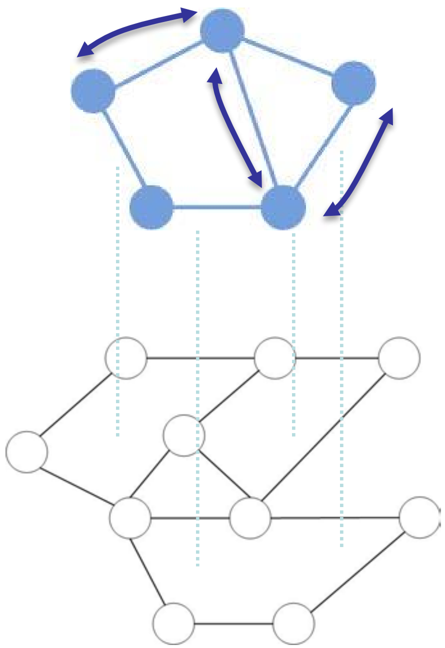
- Overlay is a logical network which is built on top of a physical network



- Overlays nodes:
  - are usually assigned at the network edge
  - are connected by virtual links (a.k.a. tunnels)

Overlay Routing

- Overlays:
  - use mechanisms above the transport layer to control the routing across the overlay nodes
  - but have no control on the routing between two overlay nodes



Deployments of Overlays

Institut für Kommunikations-Technik

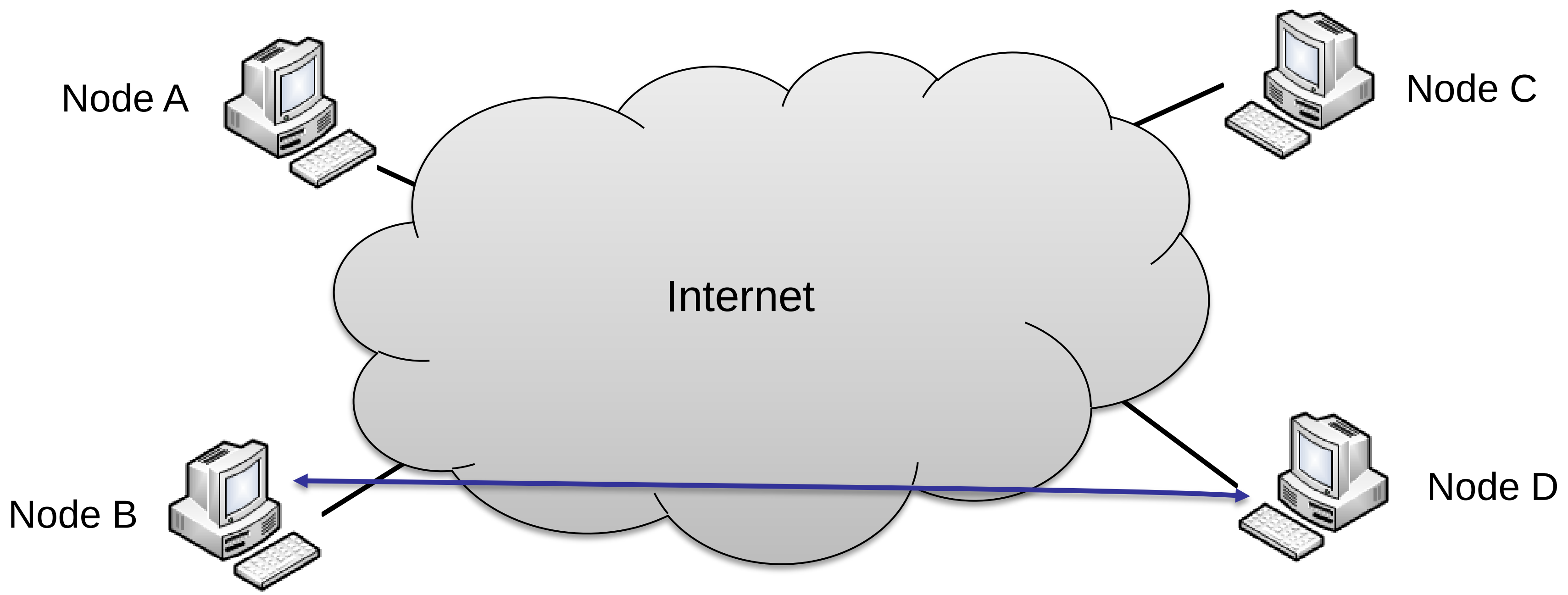
- Overlays have been used to enable:
  - Resilient routing
    - e.g., RON (Resilient Overlay Network)
  - Quality of Service (QoS)
    - e.g., OverQoS
  - Multicast
  - Content distribution
    - Akamai
    - YouTube
    - Limelight

Resilient Overlay Networks (RON)

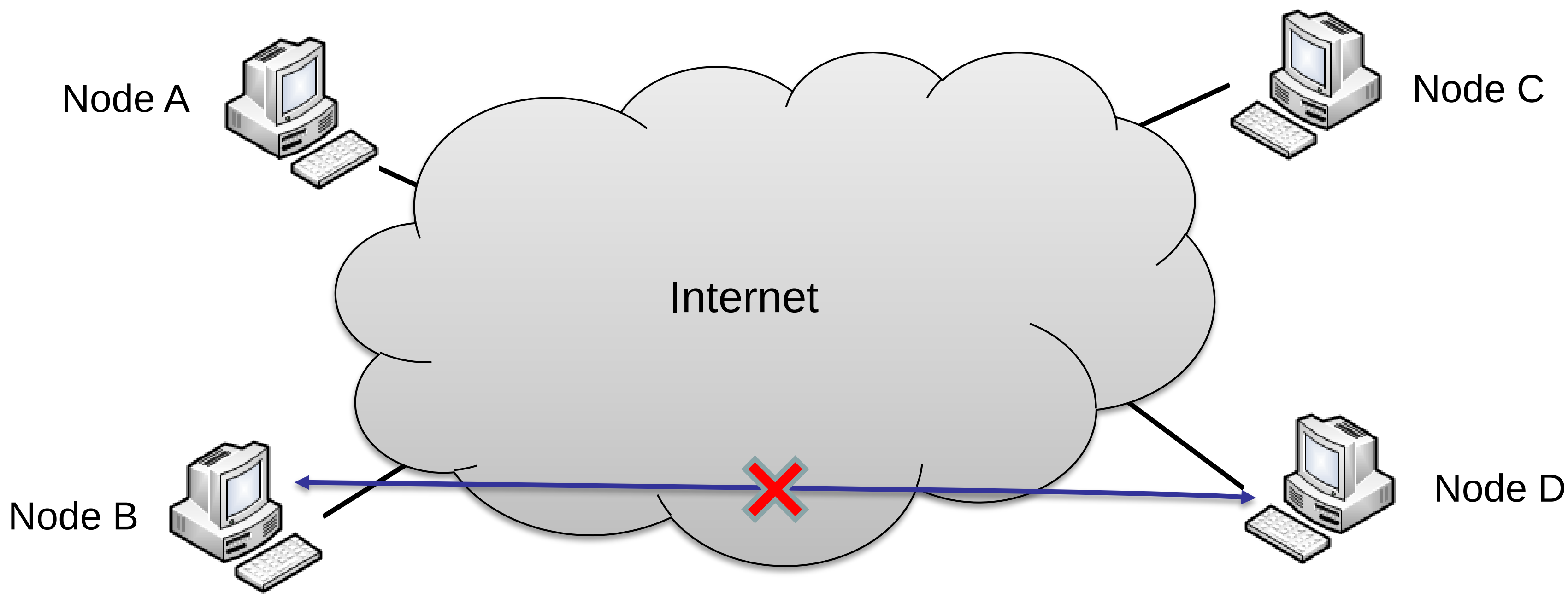
Resilient Overlay Networks (RON)

- RON deploys overlays to provide resilience:
  - Motivation:
    - End-hosts may have better information about performance and reachability than routers
    - End-hosts may be able to react faster to path failures (BGP might need several minutes to converge)
  - Main goal:
    - Improve reliability for a small number of hosts (< 50)
  - Main idea:
    - End-hosts discover path failures and cooperate to change flow routing

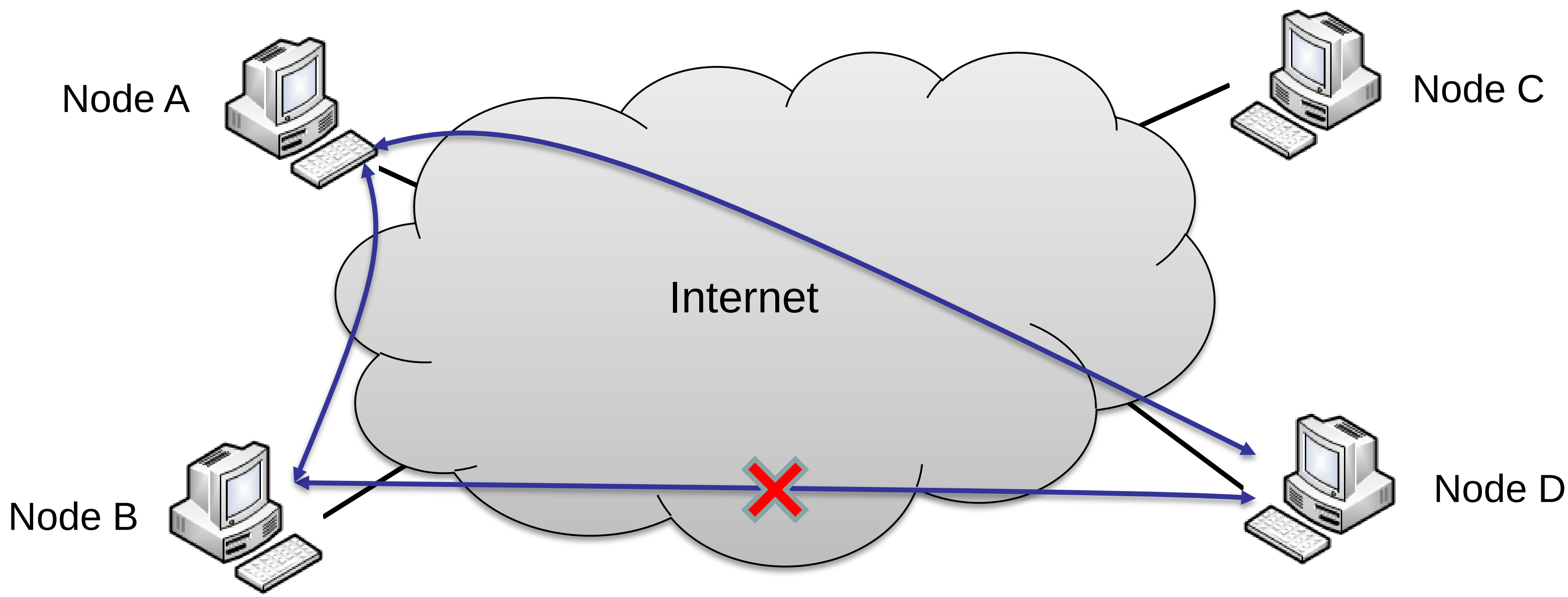
RON Example



RON Example

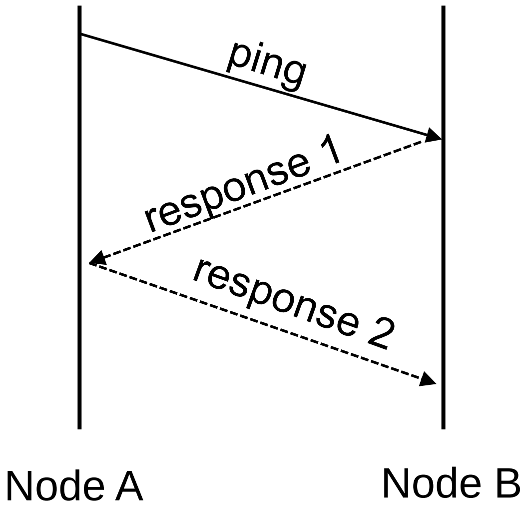


RON Example



RON Probing and Outage Detection

- Outage detection via active UDP-based probing and passive observation:
  - Sends probes every 14 sec (small overhead)
  - Transmits 3 packets
    - Both sides get RTT and reachability information
  - Lost probes are determined via timeout
  - If a probe is lost, RON transmits N consecutive probes
    - if all N probes timeout, the path is considered “dead”

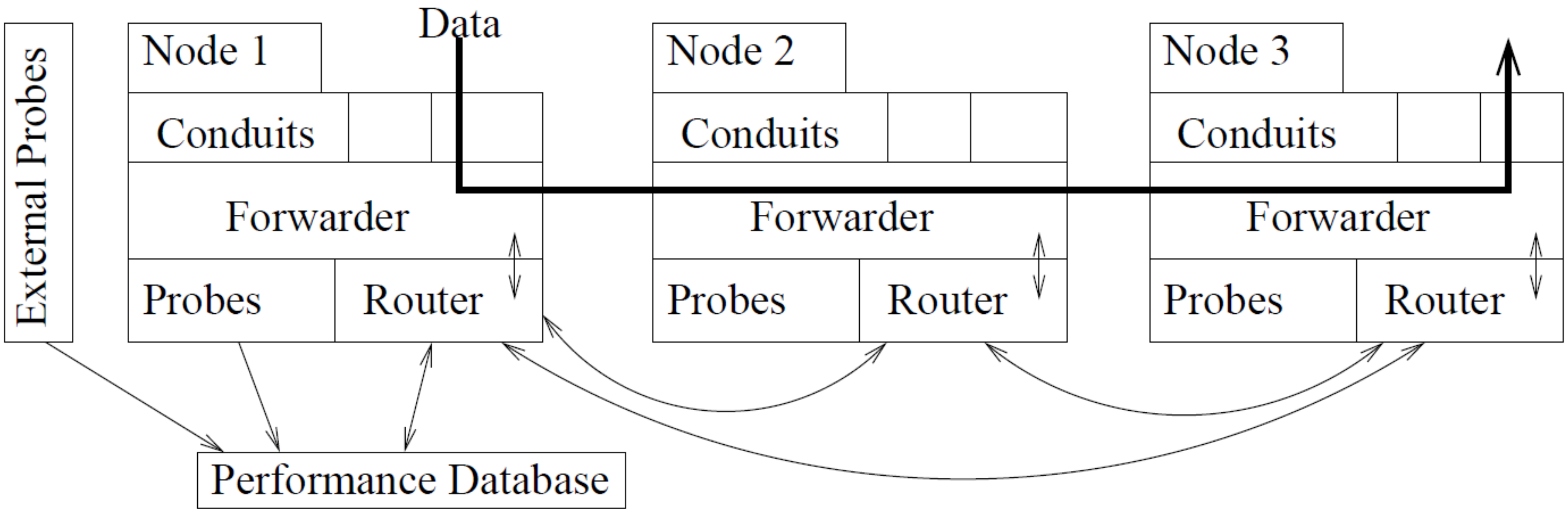


RON Routing

Institut für Kommunikations-Technik

- Link-state protocol to disseminate topology information across RON routers
- Policy-based routing
- Routing can be optimized according to specific application requirements:
  - 3 routing metrics are supported (separate routing tables for each metric):
    - Minimize latency
    - Minimize packet loss
    - Optimize throughput

RON Software Architecture



- API (*conduit*) is used by RON clients to send and receive packets
- Database maintains latency, throughput and packet loss information

RON Achievements

- RON can respond to path failures within a few seconds (~18)
  - BGP’s path/link failure recovery is in the order of minutes

- RON can achieve performance gains in terms of:
  - latency
  - packet loss (reduced by 5% for some connections)
  - throughput (increased up to 200%)

Content Distribution Networks

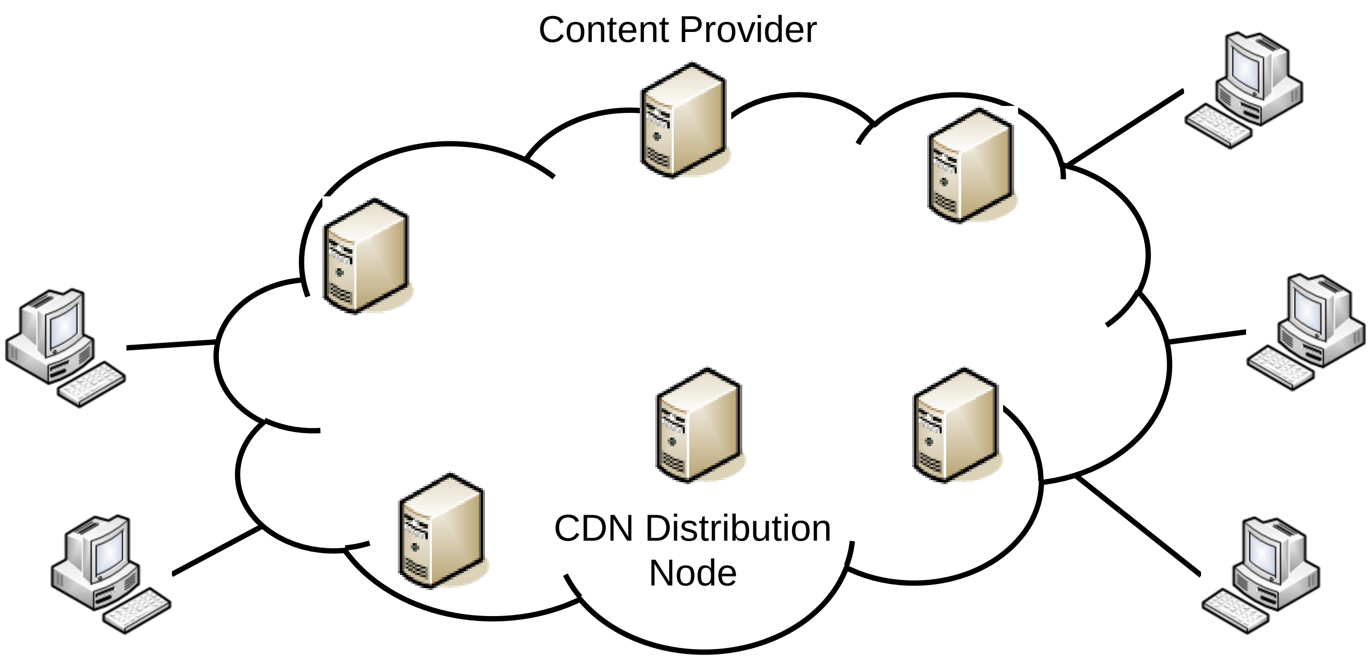
Content Distribution Networks (CDNs)

- Traditional client/server model relies on dedicated servers for storing content (e.g., video, images):
  - Increasing demand for content can result in bottlenecks and increase the server load
  - Users are dispersed across many geographical locations and many of them might experience delays while accessing content which is hosted in a particular server (i.e., far from their location)

- Content distribution networks (CDNs) replicate content at a large number of geographically dispersed servers:
  - Response time is reduced by redirecting requests to the servers that are proximate to the users
  - Server load is reduced and potential network bottlenecks are avoided by distributing requests across the available servers

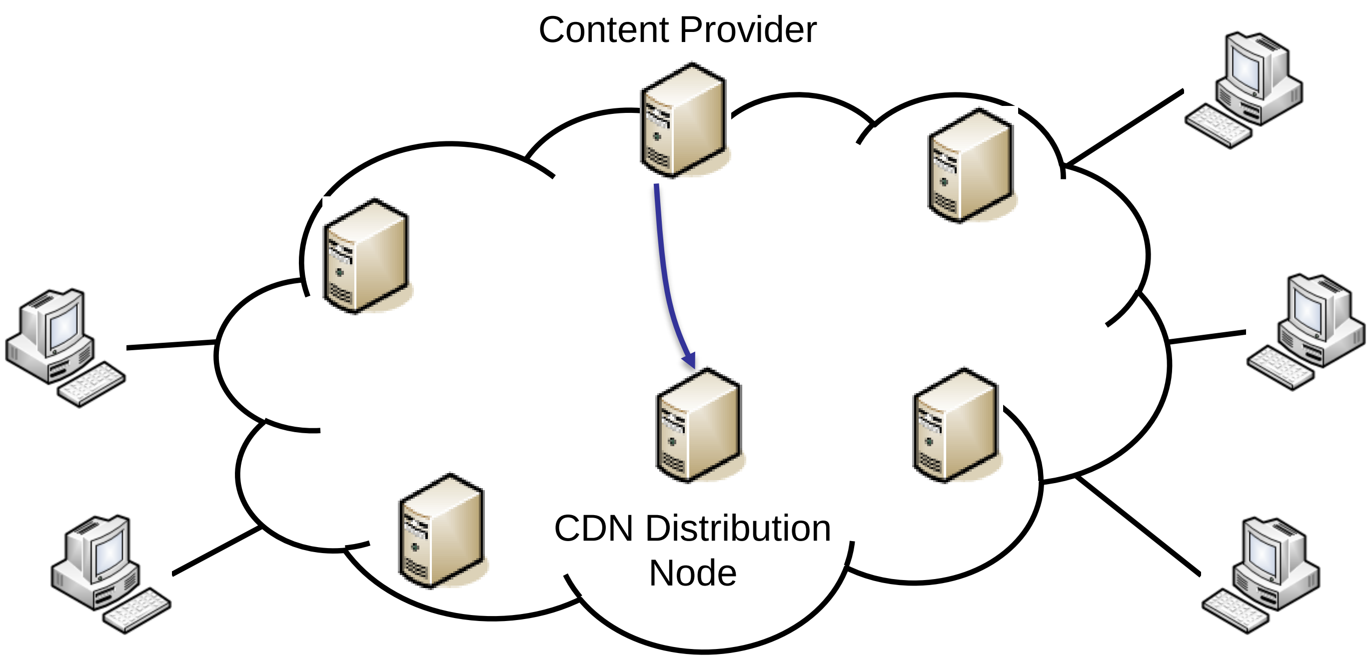
Content Distribution

- CDNs purchase servers at the edge of the network (usually near Points-of-Presence):
  - Content is cached to CDN servers beforehand, so that is readily available to the users
  - Content is redistributed when an update has been available



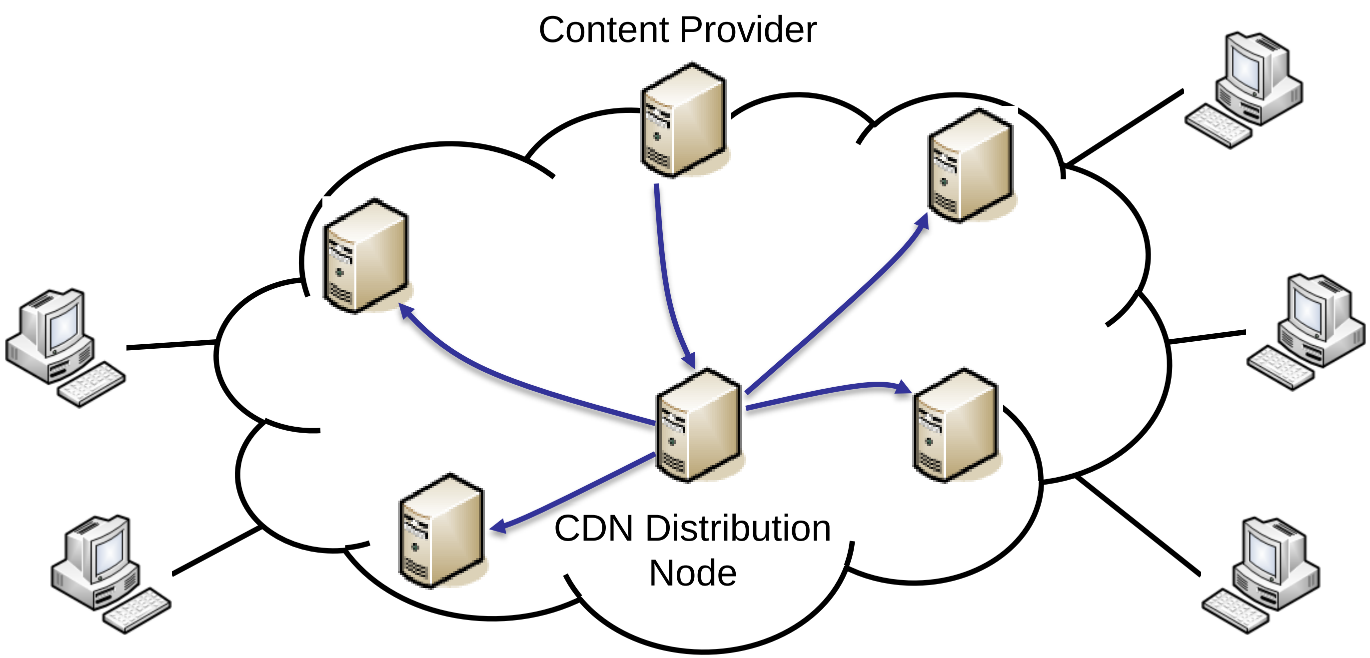
Content Distribution

- CDNs purchase servers at the edge of the network:
  - Content is cached to CDN servers beforehand, so that is readily available to the users
  - Content is redistributed when an update has been available



Content Distribution

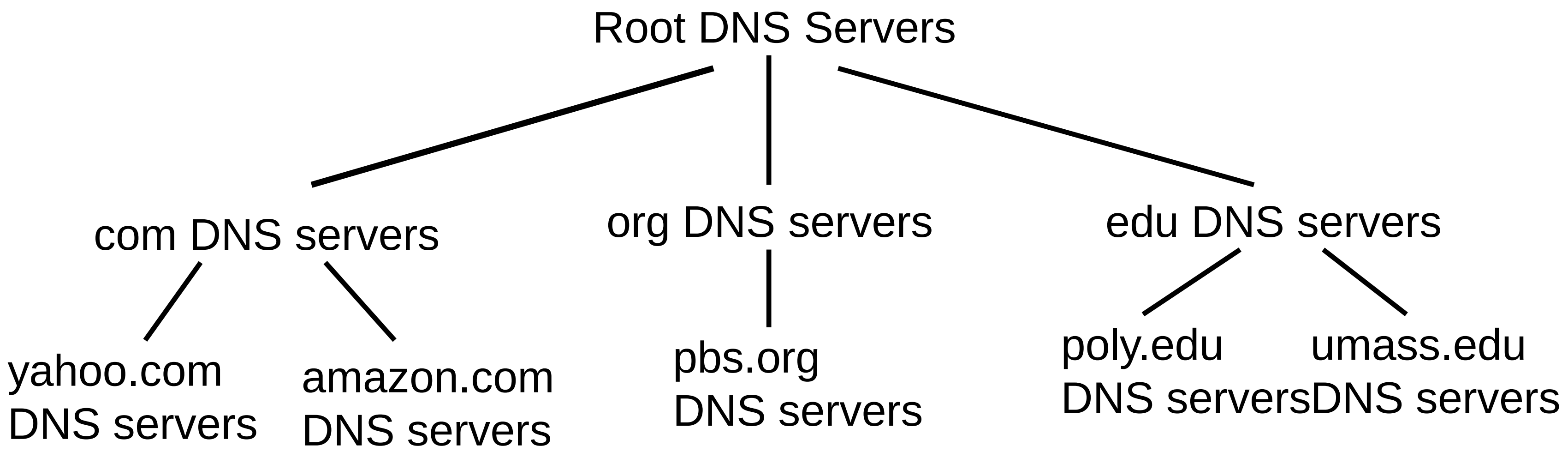
- CDNs purchase servers at the edge of the network:
  - Content is cached to CDN servers beforehand, so that is readily available to the users
  - Content is redistributed when an update has been available



Domain Name System (DNS)

Institut für Kommunikations-Technik

- DNS provides hostname to IP address translation
- DNS uses an hierarchical, distributed database



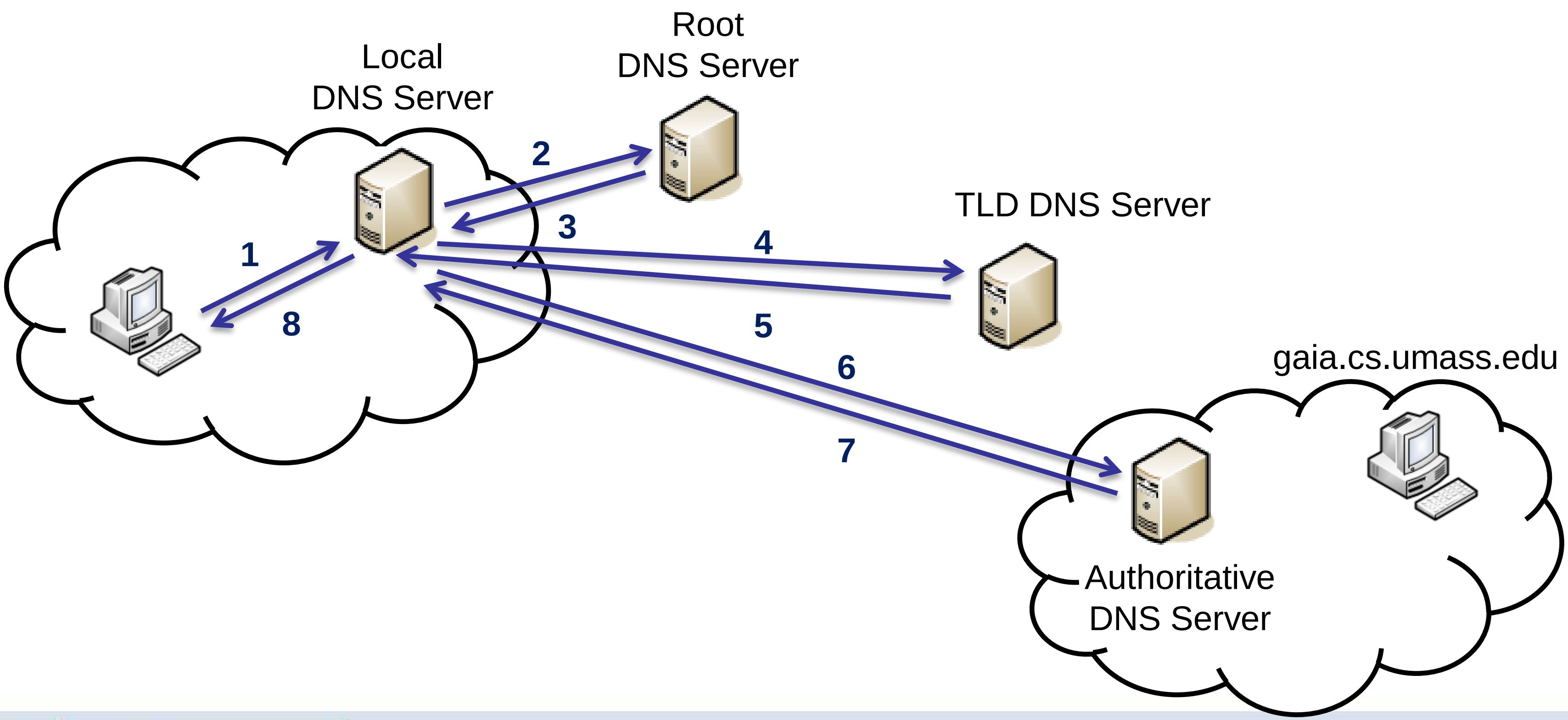
DNS Servers

Institut für Kommunikations-Technik

- Local DNS servers:
  - the first contact point for hostname resolution
  - forward queries to root DNS servers
  - cache hostname mappings locally
    - cache entries expire after a certain period
- Root DNS servers:
  - contacted by local DNS servers that cannot resolve hostname
- Top-level Domain (TLD) servers:
  - responsible for com, org, edu, etc. and all top-level country domains (e.g., de, uk, fr, jp)
- Authoritative DNS servers:
  - belong to organizations (e.g., enterprises, universities)
  - provide hostname resolution for organization’s servers (e.g., web, e-mail)

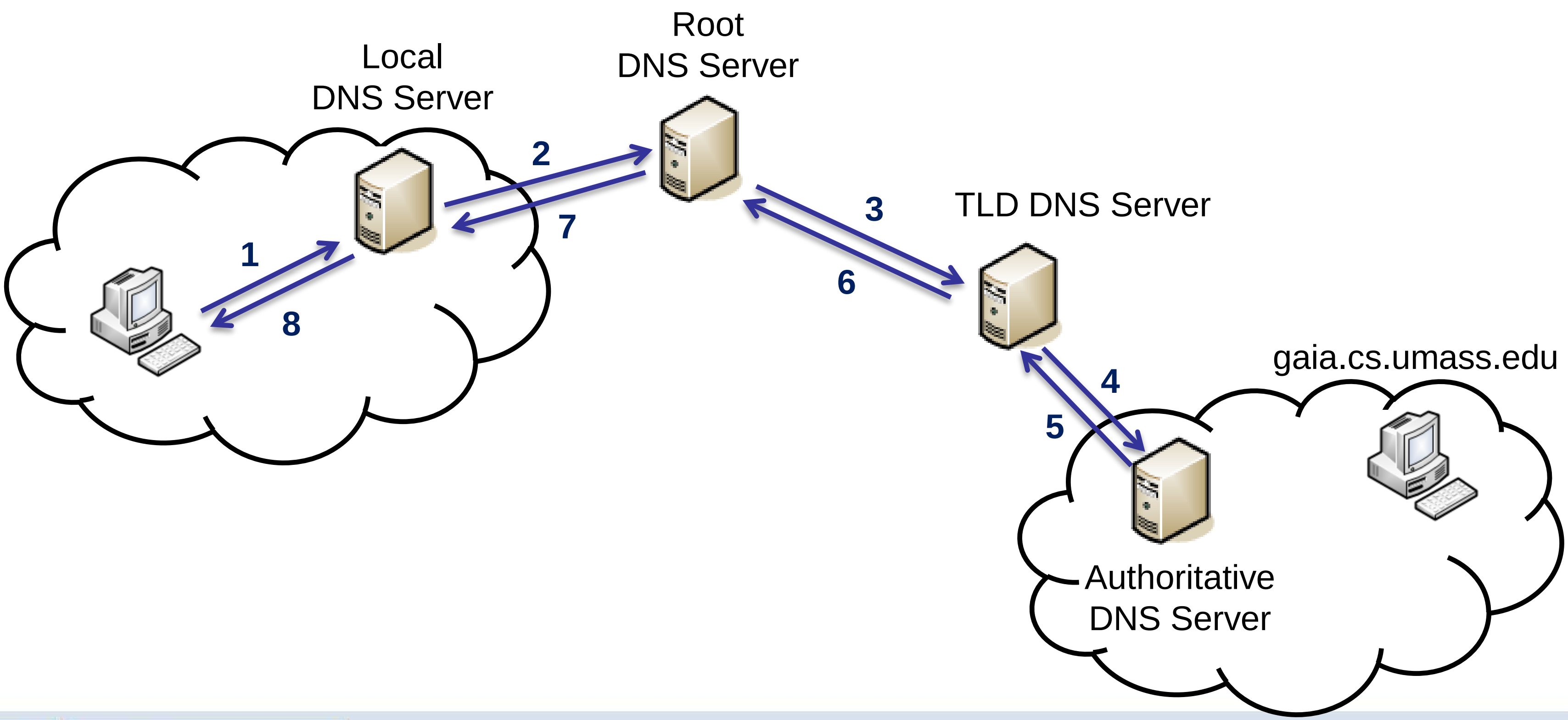
DNS Name Resolution

- The host **pc1.ikt.uni-hannover.de** queries the IP address for **gaia.cs.umass.edu**:
  - Iterated query



DNS Name Resolution

- The host **pc1.ikt.uni-hannover.de** queries the IP address for **gaia.cs.umass.edu**:
  - Recursive query

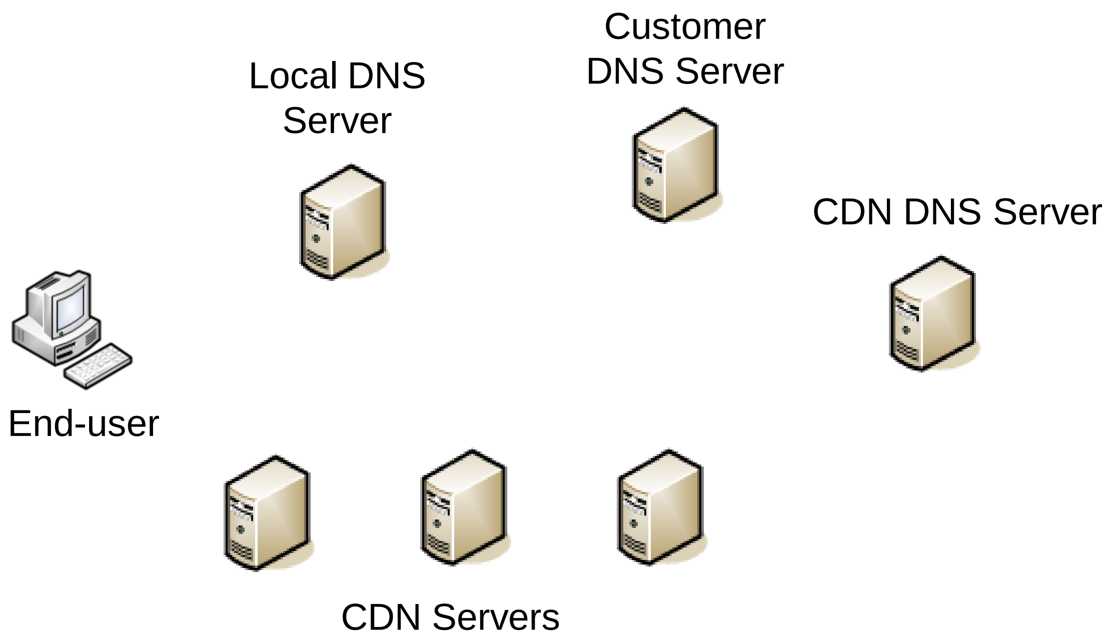


DNS Records

- DNS record format: `[name, value, type, TTL]`
- DNS record types:
  - Address (A) record for hosts
    - e.g., `[relay1.bar.example.com, 123.45.120.12, A, TTL]`
  - Mail Exchanger (MX) record for email servers
    - e.g., `[example.com, mail.example.com, MX, TTL]`
  - Name Server (NS) record for authoritative DNS servers of a domain
    - e.g., `[example.com, dns.example.com, NS, TTL]`
  - Canonical Name (CNAME) record:
    - `name` is an alias for a "canonical" name
    - e.g., `[example.com, relay1.bar.example.com, CNAME, TTL]`
    - allows the use of mnemonic hostnames
    - widely used for the redirection of client requests to a proximate server in content distribution networks (CDNs)

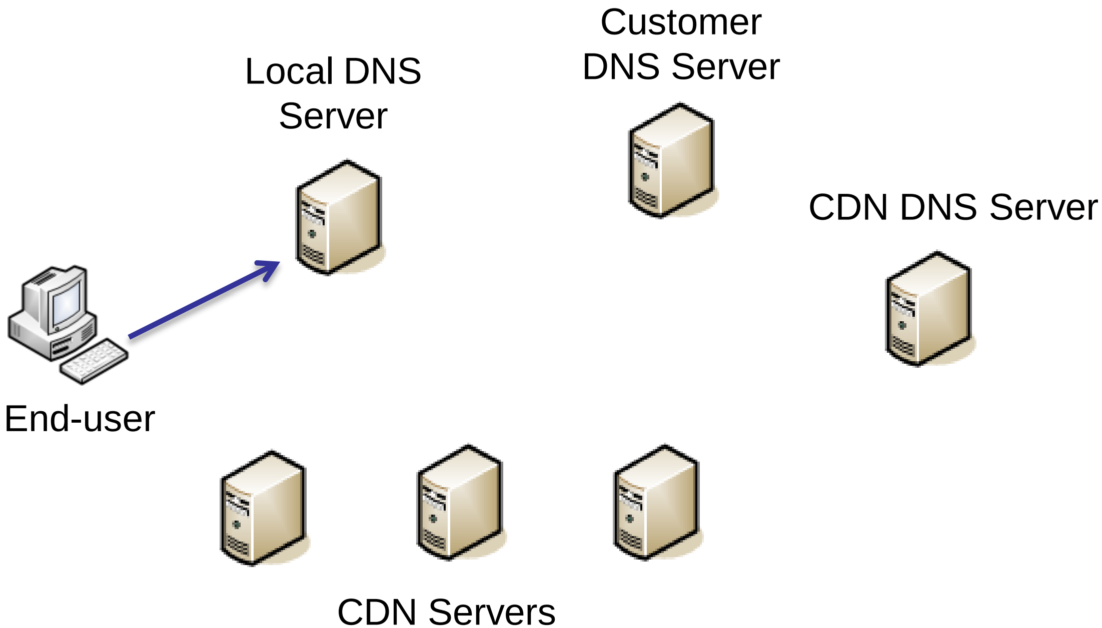
DNS Redirection

- Most CDNs use Domain Name System (DNS) to redirect end-user requests to the CDN:
  - Customer's DNS server redirects requests to the CDN DNS server
  - The CDN DNS server selects the most proximate CDN server



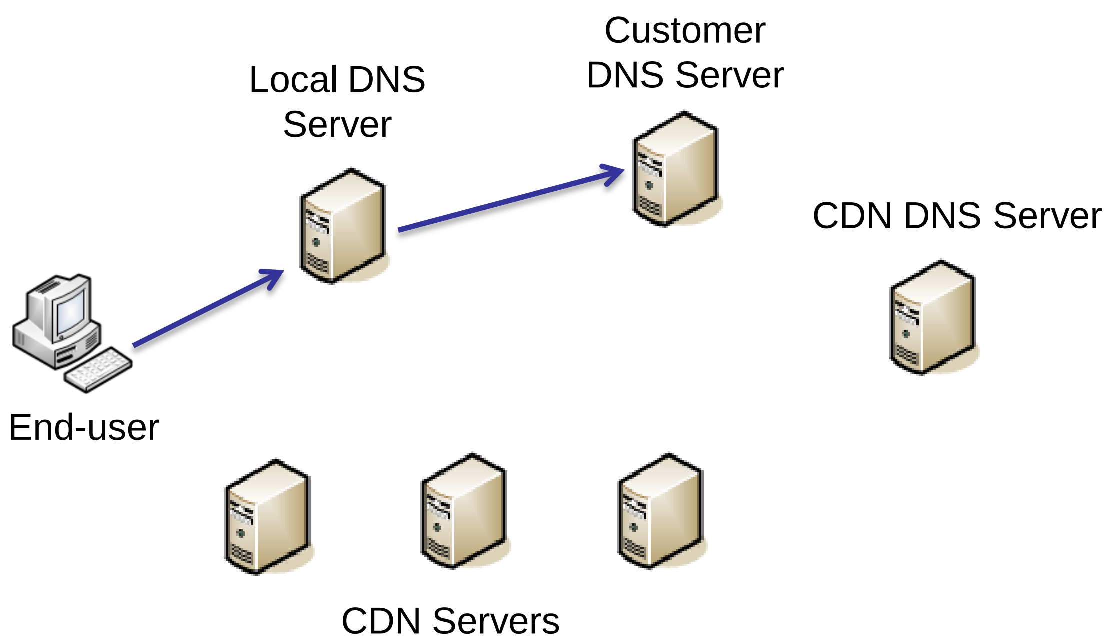
DNS Redirection

- Most CDNs use Domain Name System (DNS) to redirect end-user requests to the CDN:
  - Customer's DNS server redirects requests to the CDN DNS server
  - The CDN DNS server selects the most proximate CDN server



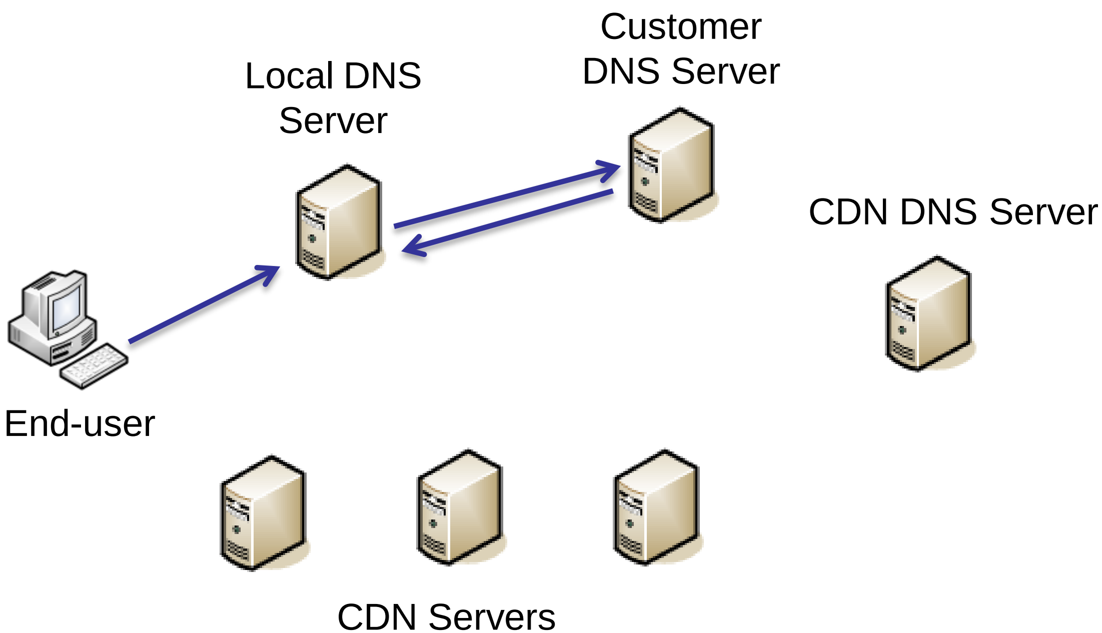
DNS Redirection

- Most CDNs use Domain Name System (DNS) to redirect end-user requests to the CDN:
  - Customer's DNS server redirects requests to the CDN DNS server
  - The CDN DNS server selects the most proximate CDN server



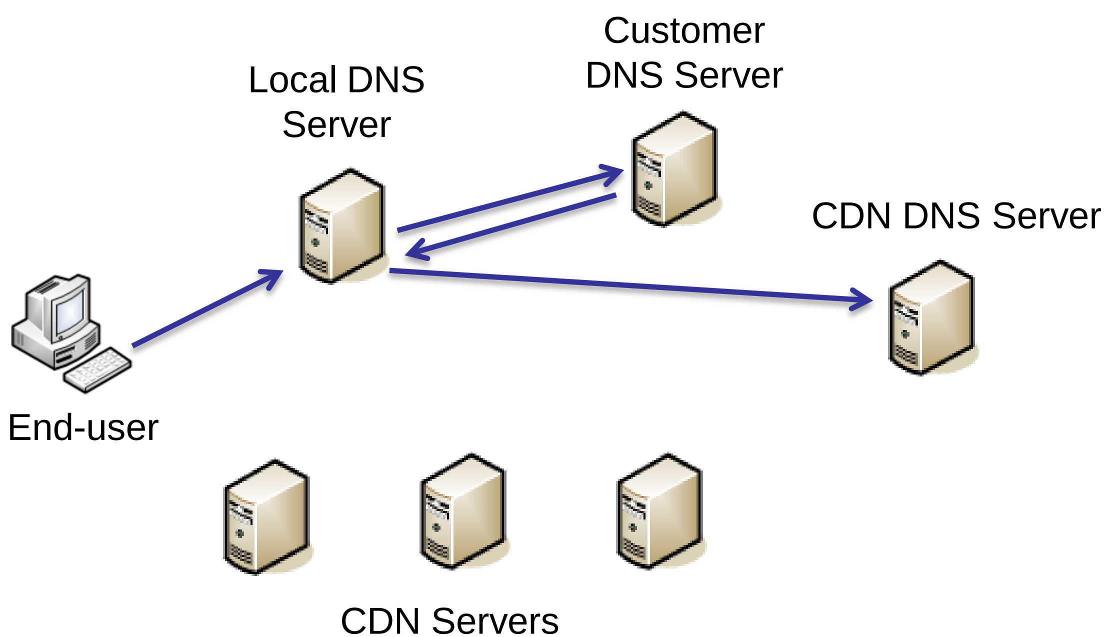
DNS Redirection

- Most CDNs use Domain Name System (DNS) to redirect end-user requests to the CDN:
  - Customer's DNS server redirects requests to the CDN DNS server
  - The CDN DNS server selects the most proximate CDN server



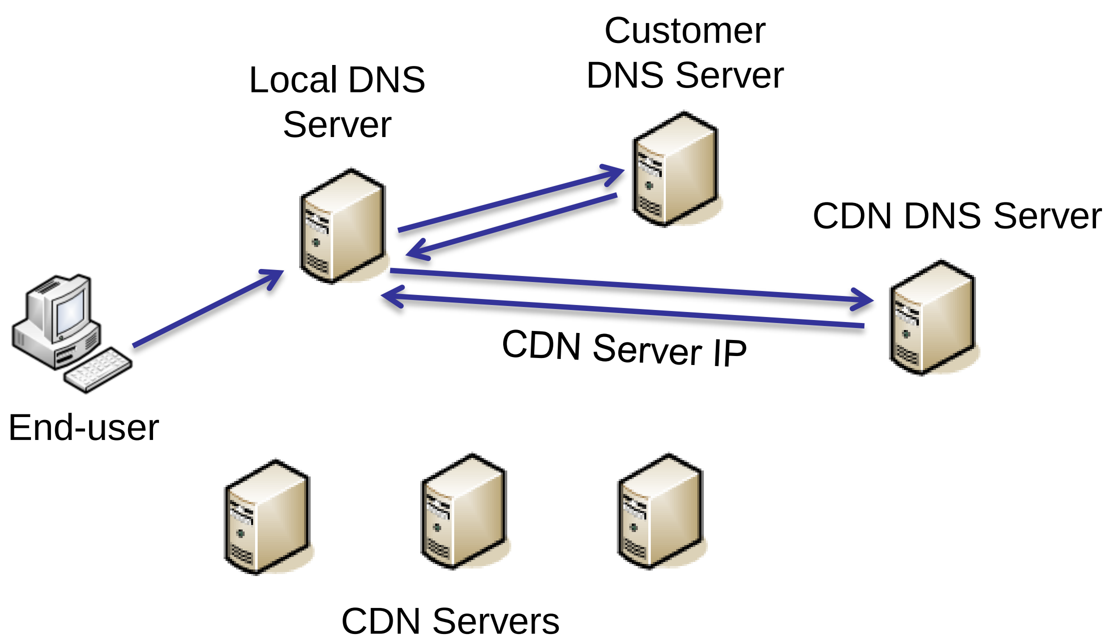
DNS Redirection

- Most CDNs use Domain Name System (DNS) to redirect end-user requests to the CDN:
  - Customer's DNS server redirects requests to the CDN DNS server
  - The CDN DNS server selects the most proximate CDN server



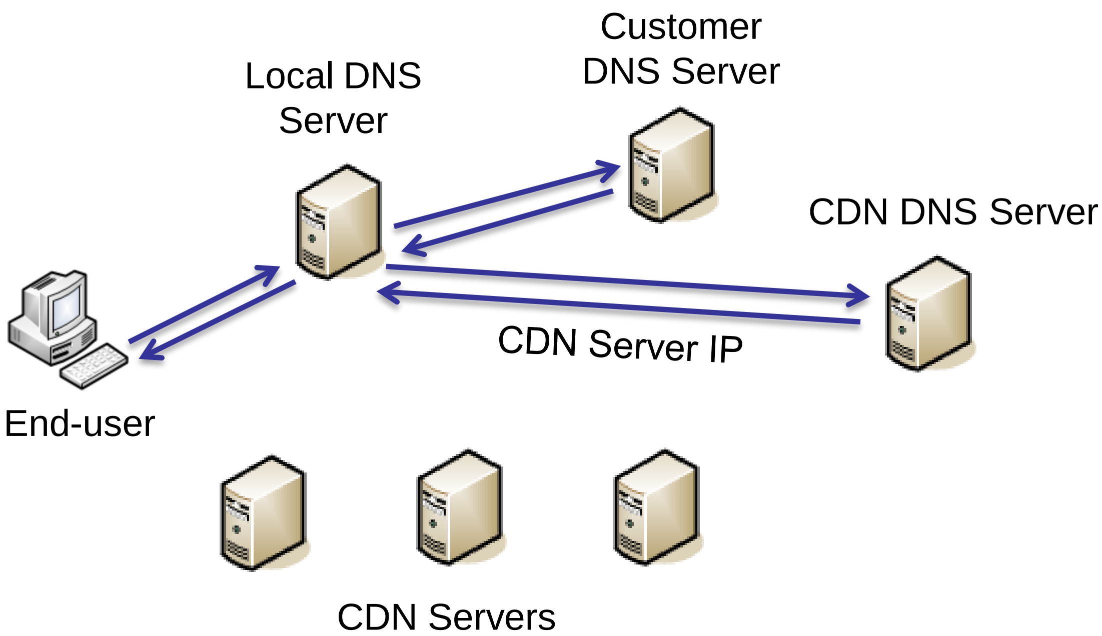
DNS Redirection

- Most CDNs use Domain Name System (DNS) to redirect end-user requests to the CDN:
  - Customer’s DNS server redirects requests to the CDN DNS server
  - The CDN DNS server selects the most proximate CDN server



DNS Redirection

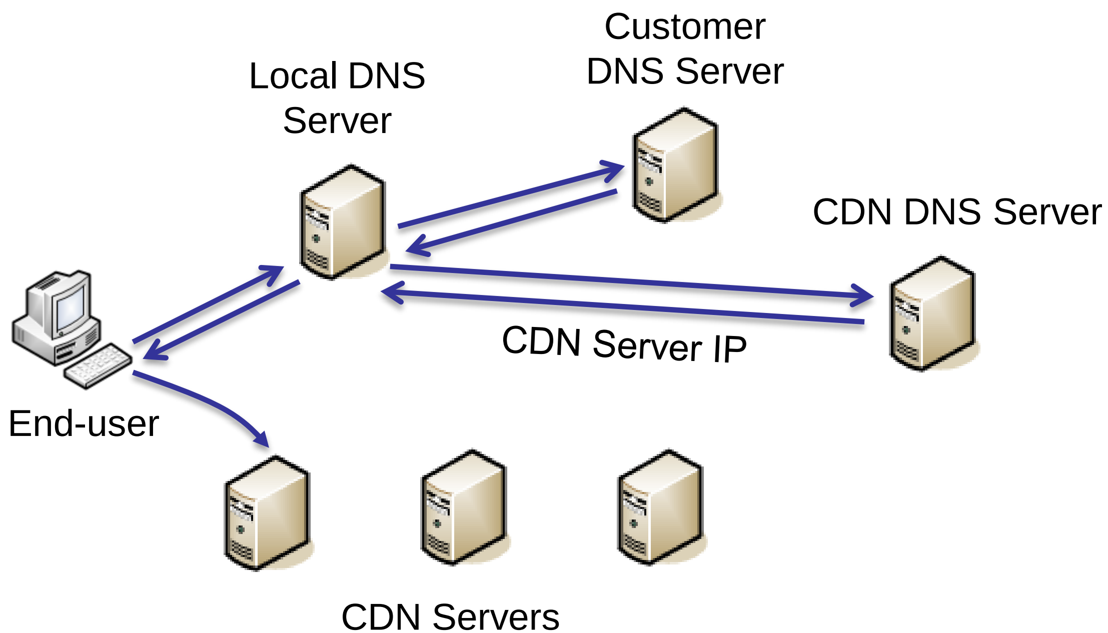
- Most CDNs use Domain Name System (DNS) to redirect end-user requests to the CDN:
  - Customer's DNS server redirects requests to the CDN DNS server
  - The CDN DNS server selects the most proximate CDN server



DNS Redirection

Institut für Kommunikations-Technik

- Most CDNs use Domain Name System (DNS) to redirect end-user requests to the CDN:
  - Customer's DNS server redirects requests to the CDN DNS server
  - The CDN DNS server selects the most proximate CDN server

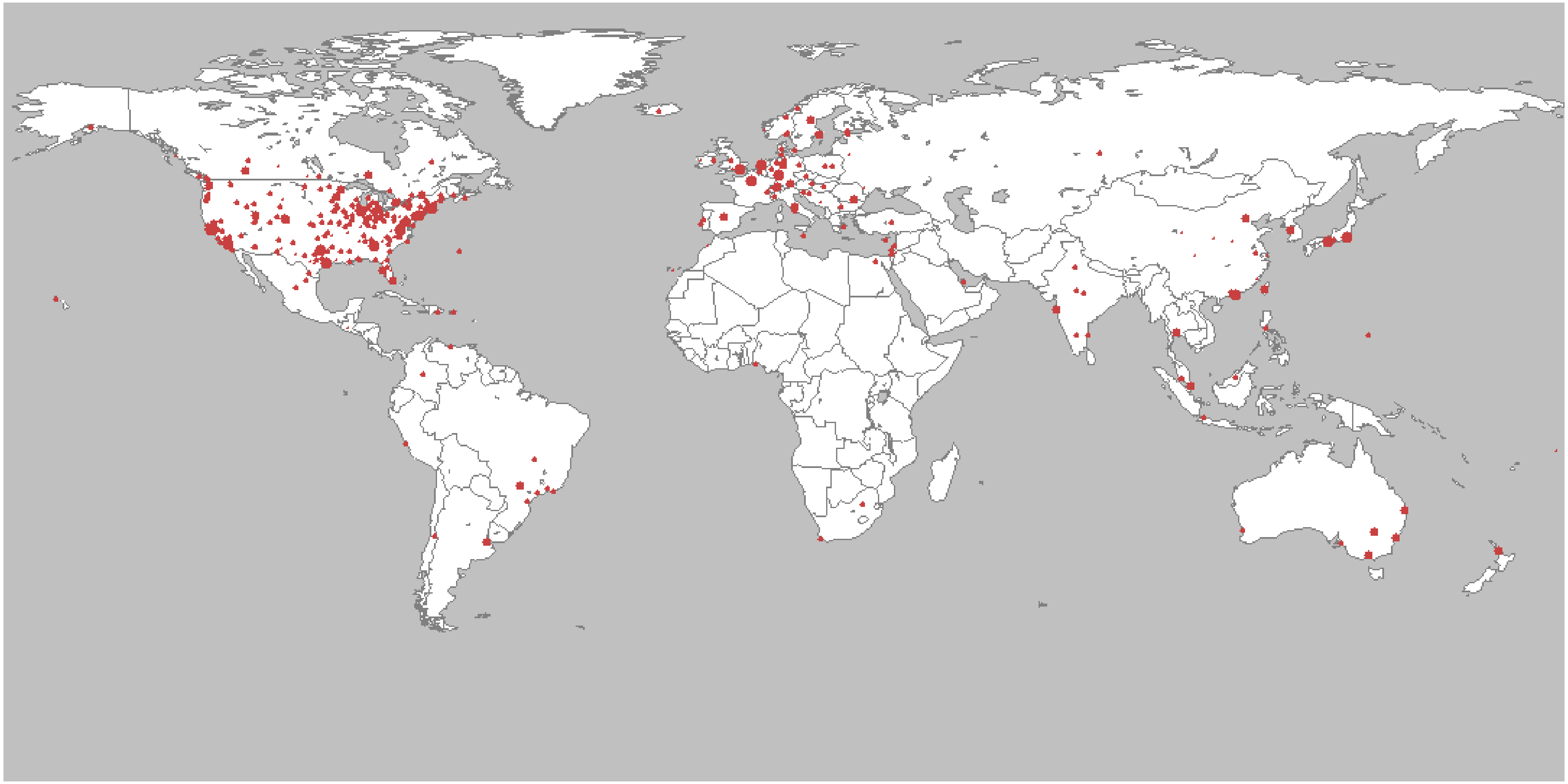


CDN Policies

- CDNs can use different policies for redirecting user requests:
  - Round-robin redirection
    - Load balancing
  - Redirection to the CDN server that is closer to the end-user
    - Minimize response time

- A CDN would want to achieve both goals (i.e., load balancing and reduced response time):
  - Akamai takes into account server proximity and load balancing:
    - Akamai DNS server returns the IP addresses of more than one (usually two) servers that are "close" to the end-user
    - The web client selects one of these servers to retrieve the requested content

Akamai Network



- Akamai network:
  - World-wide edge server locations

Akamai: Server Distribution

- Geographical distribution of Akamai CDN servers in 2008

Country	IPs	%
United States	6.661	57.7
Japan	865	7.5
United Kingdom	704	6.1
Germany	545	4.7
Netherlands	384	3.3
France	364	3.2
Canada	284	2.5
Australia	164	1.4
Hong Kong	158	1.4
South Korea	124	1.1
Others	1285	11.1

Akamai: DNS Redirection Example

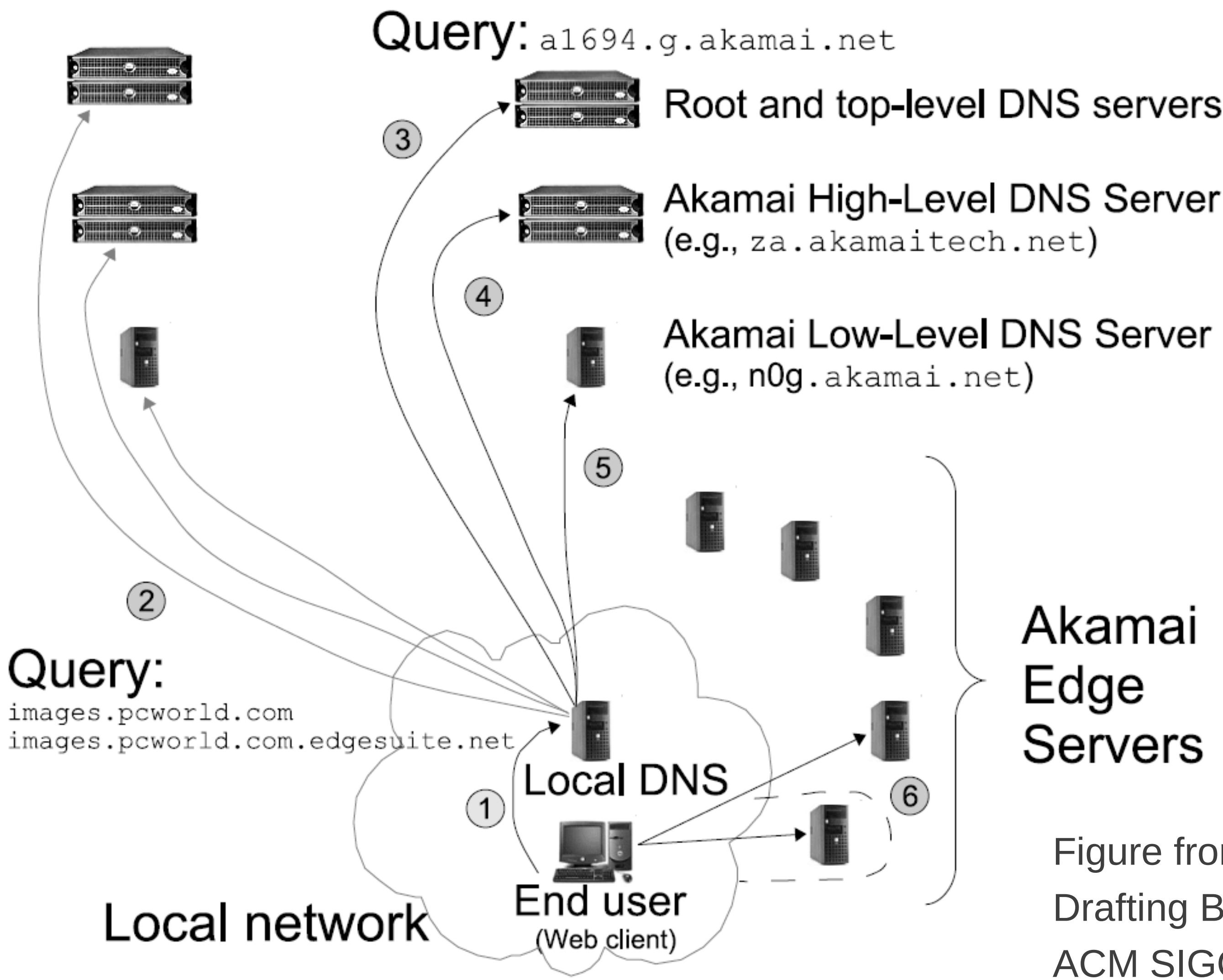
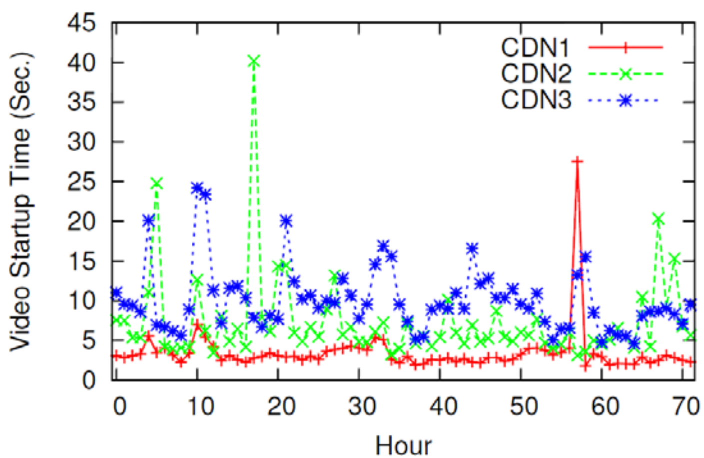


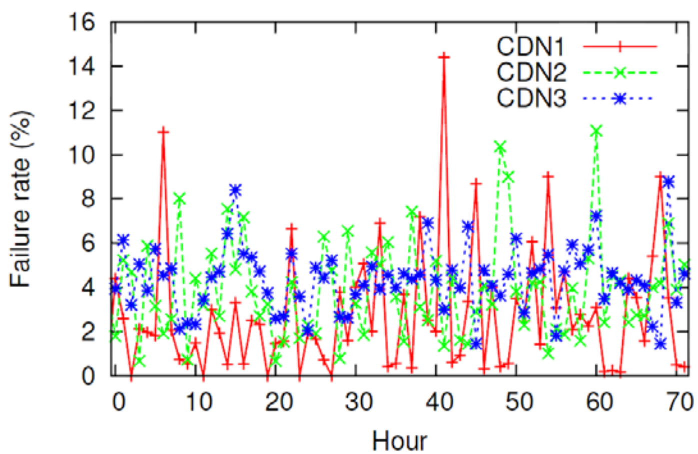
Figure from Su, et al.,
Drafting Behind Akamai,
ACM SIGCOMM 2006

CDN Issues

- CDN’s authoritative DNS server receives requests originating from the client’s DNS resolver (rather than from the client itself)
  - The DNS resolver may not be in proximity to the client.
  - Network bottlenecks and latency effects on the path between the client and the CDN server can degrade application performance

- CDN performance may vary across space and time, due to fluctuating CDN server load and network congestion

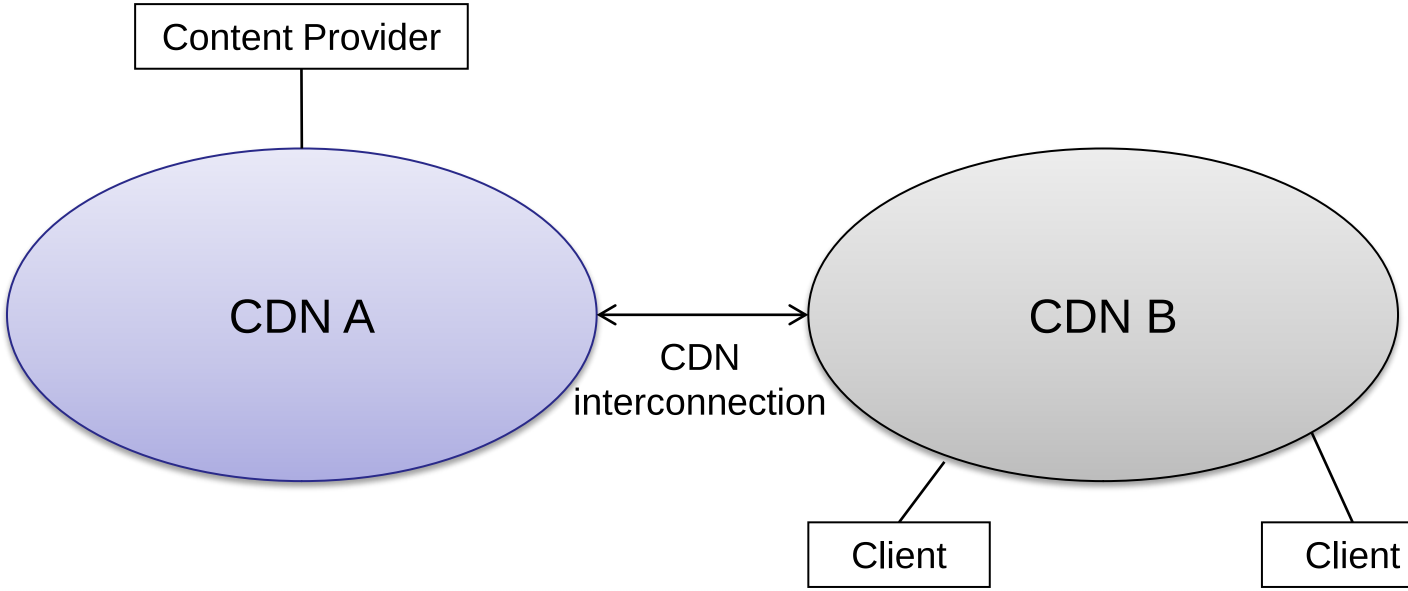




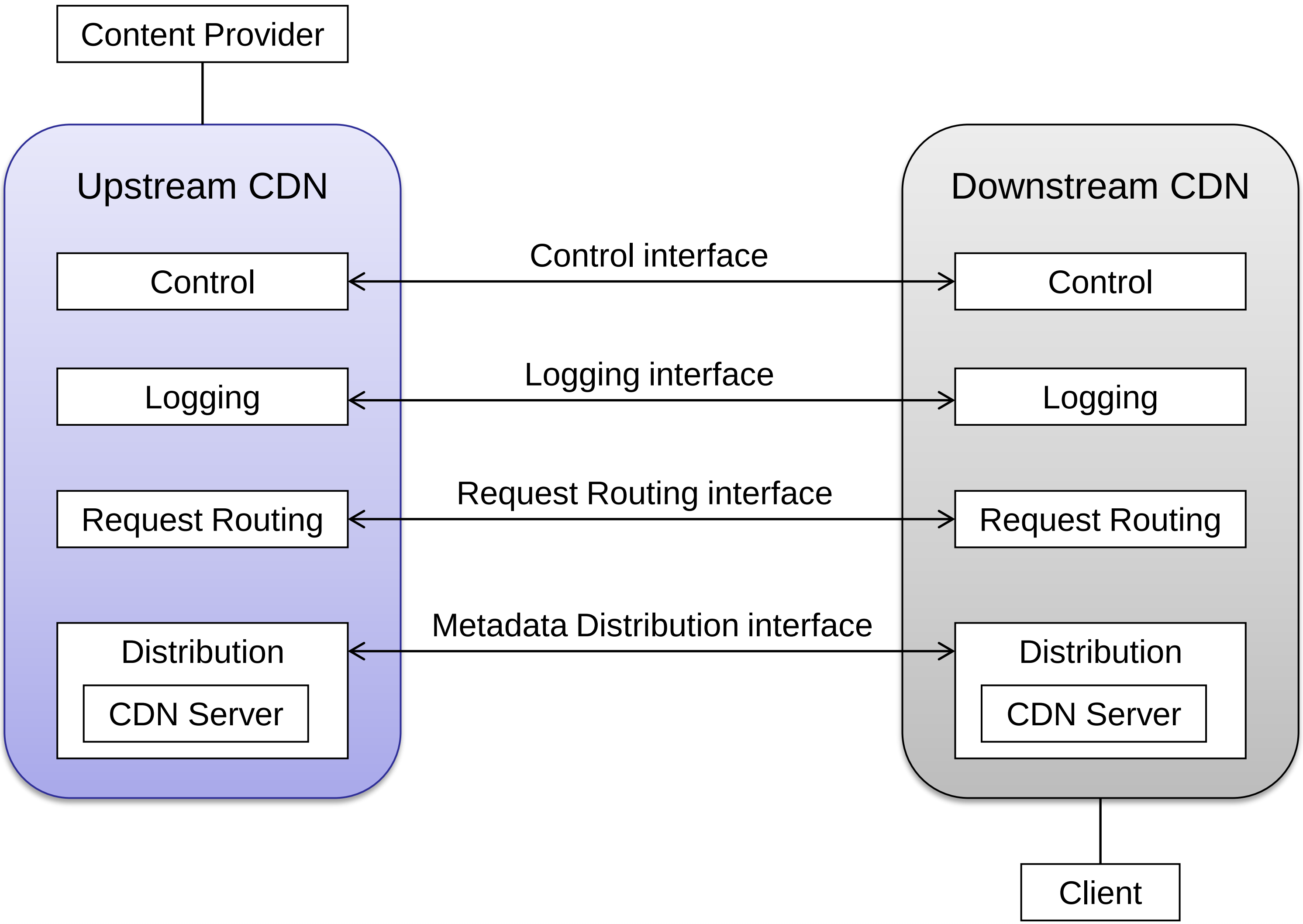
Liu, et al., A Case for a Coordinated Internet Video Control Plane, ACM SIGCOMM 2012

CDN Interconnection

- CDI interconnection (CDNi) enables content delivery delegation across collaborating CDNs
  - A CDN can extend its reach without purchasing or deploying new servers
  - Opportunity for small/medium CDNs that want to deliver content to clients out of their reach
  - Potential threat to wide-scale CDNs, e.g., Akamai



CDNi Interfaces



CDNi Interfaces

- Control Interface:
  - Allows the basic CDN control systems in interconnected CDNs to communicate
- Logging Interface:
  - Allows the logging systems in interconnected CDNs to communicate to exchange relevant activity logs in order to meet operational requirements such as billing
- Request Routing Interface:
  - Allows the request routing systems in interconnected CDNs to communicate to ensure that an end user's request can be redirected from an upstream CDN to a server in the downstream CDN
- Metadata Distribution Interface:
  - Allows the distribution system in interconnected CDNs to communicate to exchange information about content management

References

- D. Clark, et. al., **Overlay Networks and the Future of the Internet,** Communications and Strategies, 2006
- L. Subramanian, et. al., **OverQoS: An Overlay Based Architecture for Enhancing Internet QoS,** USENIX NSDI 2004
- D. Andersen, et. al., **Resilient Overlay Networks**, ACM SOSP 2001
- J. Dilley, et. al., **Globally Distributed Content Delivery**, IEEE Internet Computing, 2002
- A. Su, et. al., **Drafting Behind Akamai**, ACM SIGCOMM 2006
- IETF Content Delivery Network Interconnection (CDNi) Working Group http://datatracker.ietf.org/wg/cdni/
- IETF Application Level Traffic Optimization (ALTO) Working Group, http://datatracker.ietf.org/wg/alto/