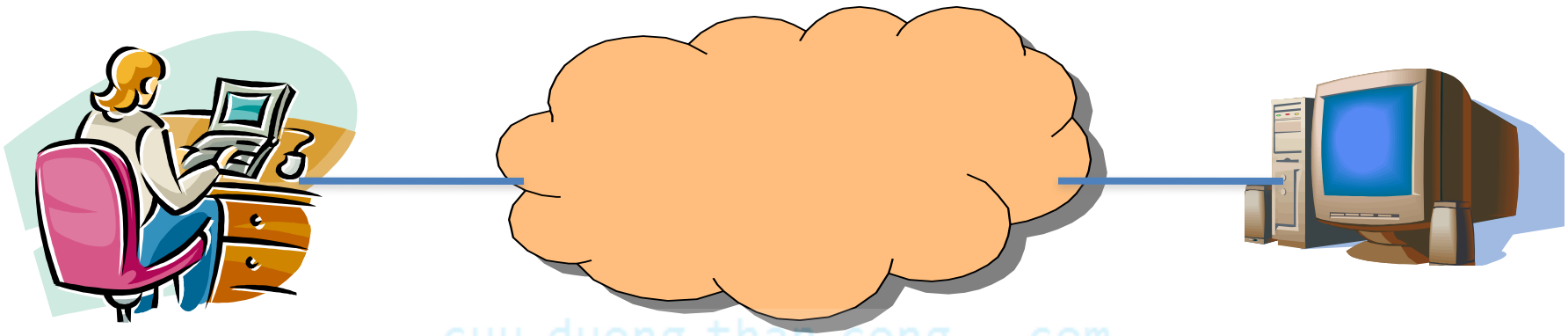
Content Distribution Networks (CDNs)

MẠNG MÁY TÍNH NÂNG CAO

Single Server, Poor Performance

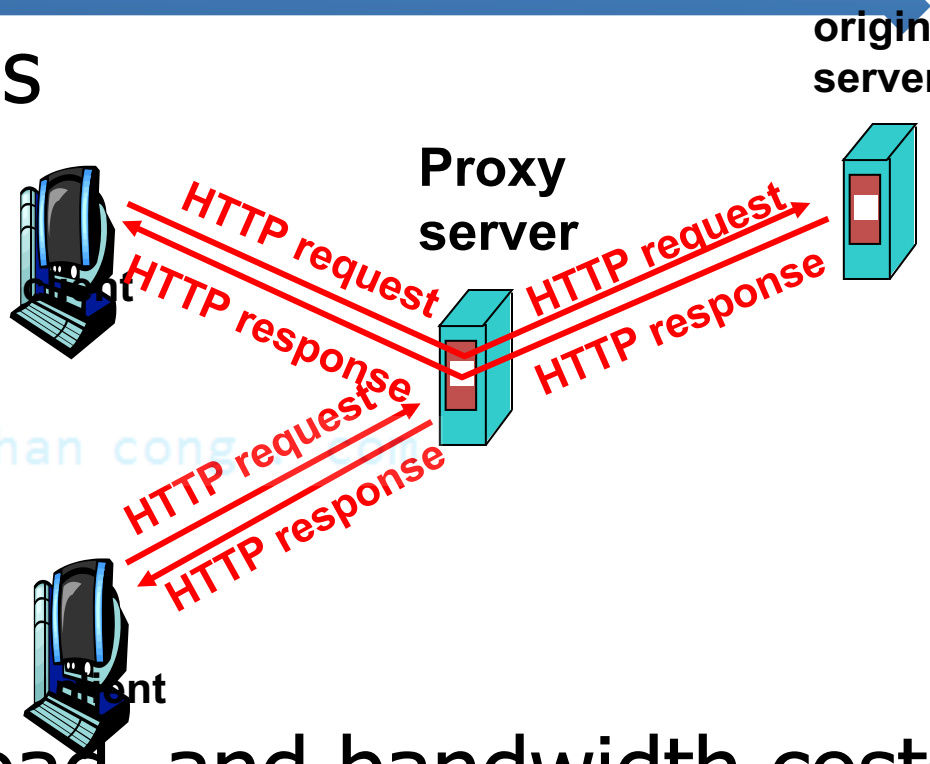


- Single server
  - Single point of failure
  - Easily overloaded
  - Far from most clients
- Popular content
  - Popular site
  - "Flash crowd" (aka "Slashdot effect")
  - Denial of Service attack

Web Caching

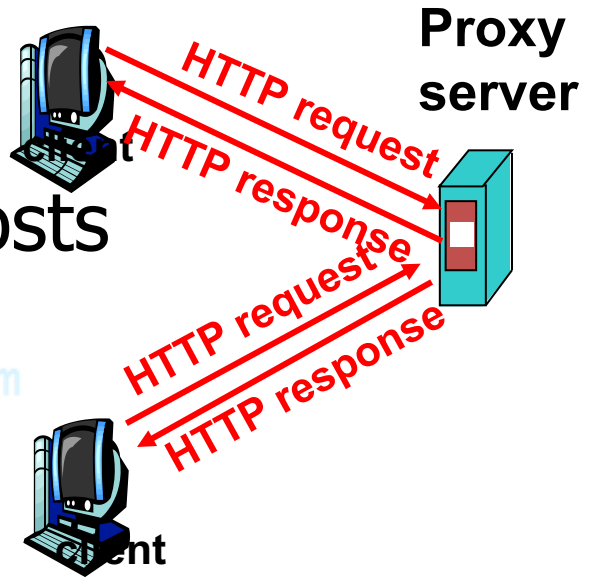
Proxy Caches

- Reactively replicates popular content
- Smaller round-trip times to clients
- Reduces load on origin servers
- Reduces network load, and bandwidth costs
- Maintain persistent TCP connections



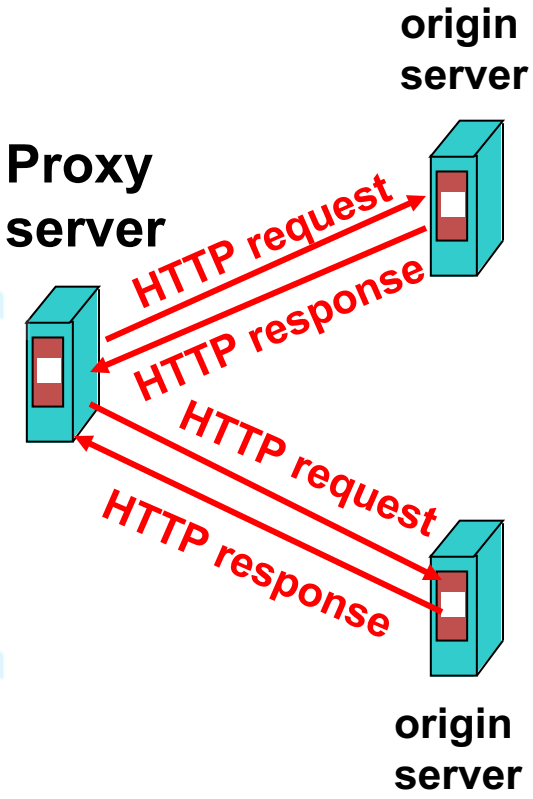
Forward Proxy

- Cache close to the client
  - Improves client performance
  - Reduces network provider's costs
- Explicit proxy
  - Requires configuring browser
- Implicit proxy
  - Service provider deploys an "on path" proxy
  - … that intercepts and handles Web requests

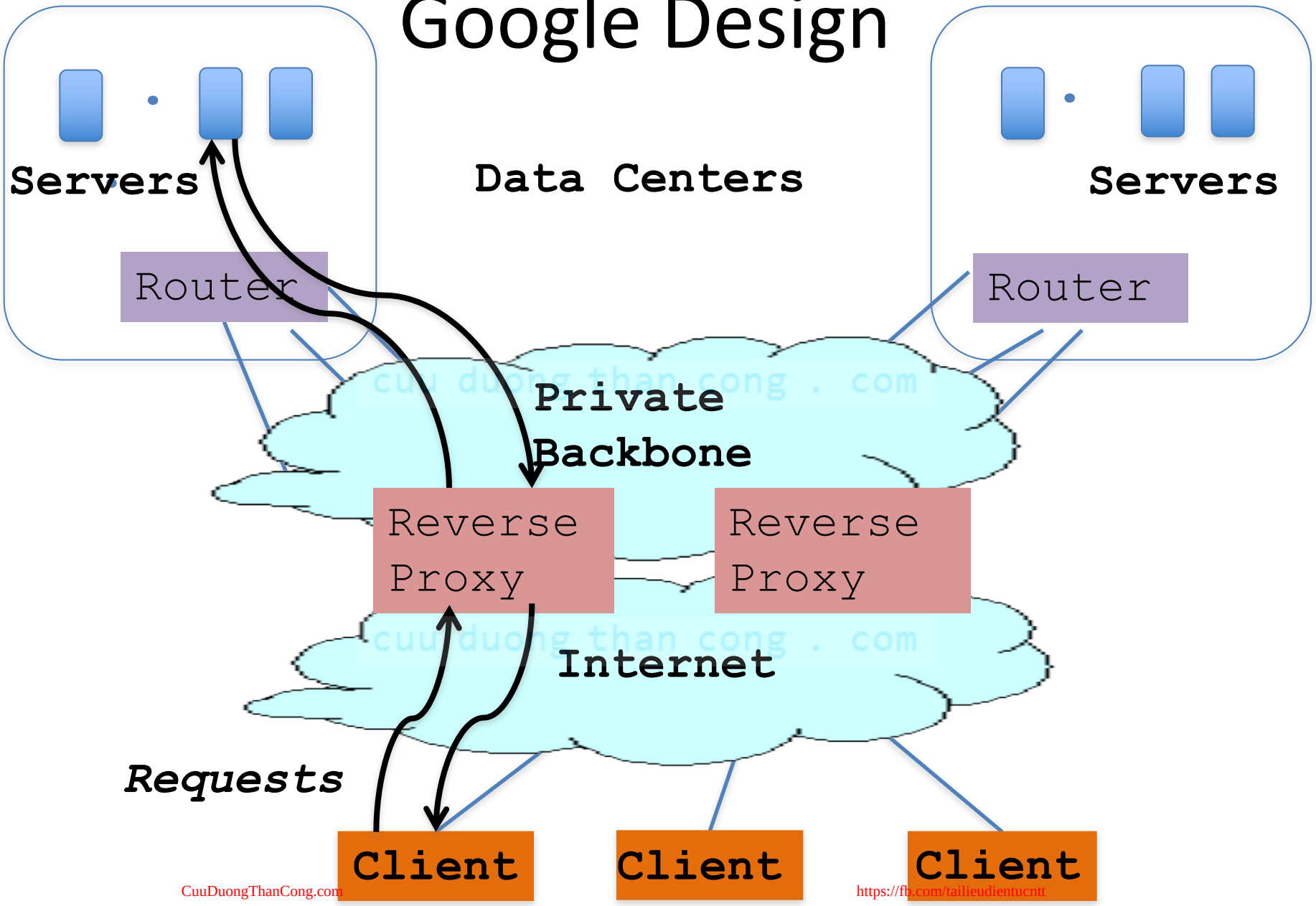


Reverse Proxy

- Cache close to server
  - Improve client performance
  - Reduce content provider cost
  - Load balancing, content assembly, transcoding, etc.
- Directing clients to the proxy
  - Map the site name to the IP address of the proxy



Google Design



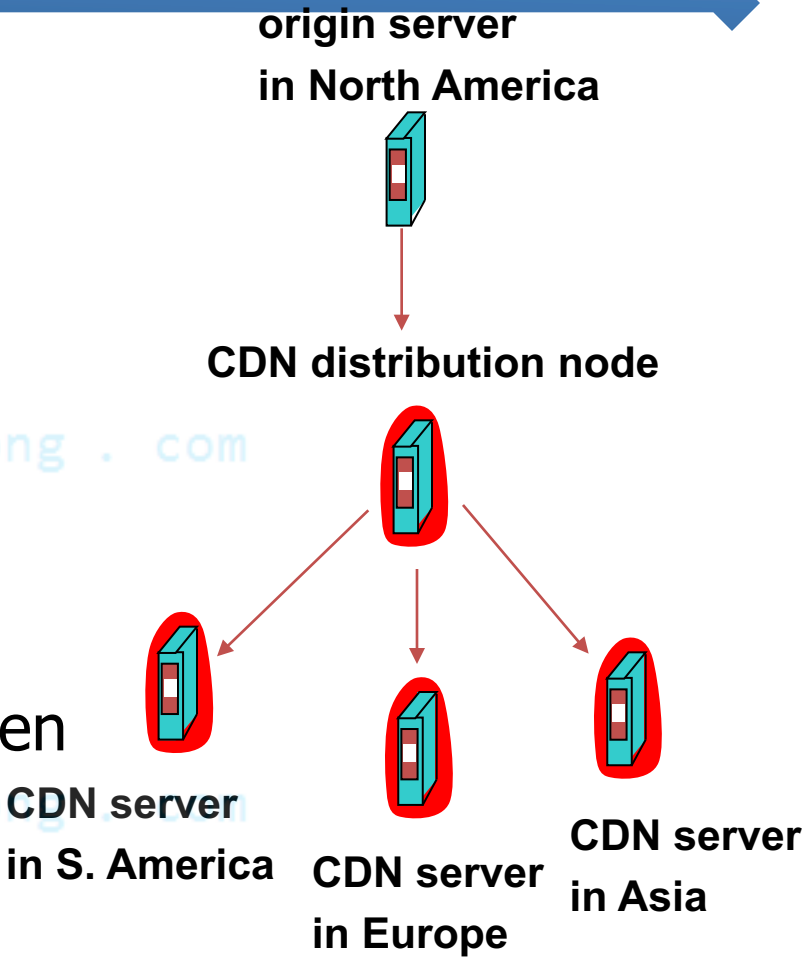
Limitations of Web Caching

- Much content is not cacheable
  - Dynamic data: stock prices, scores, web cams
  - CGI scripts: results depend on parameters
  - Cookies: results may depend on passed data
  - SSL: encrypted data is not cacheable
  - Analytics: owner wants to measure hits
- Stale data
  - Or, overhead of refreshing the cached data

Content Distribution Networks

Content Distribution Network

- Proactive content replication
  - Content provider (e.g., CNN) contracts with a CDN
- CDN replicates the content
  - On many servers spread throughout the Internet
- Updating the replicas
  - Updates pushed to replicas when the content changes



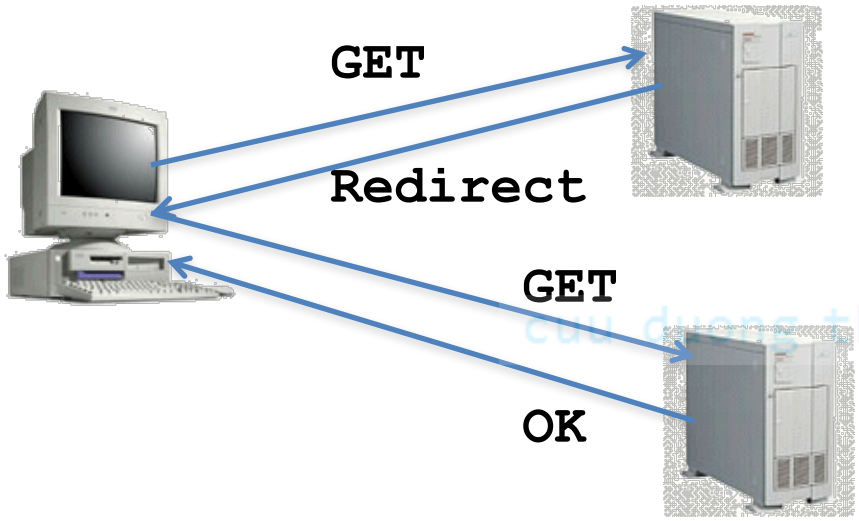
Server Selection Policy

- Live server
  - For availability
- Lowest load
  - To balance load across the servers
- Closest
  - Nearest geographically, or in round-trip time
- Best performance
  - Throughput, latency, …
- Cheapest bandwidth, electricity, …

Requires continuous monitoring of liveness, load, and performance

Server Selection Mechanism

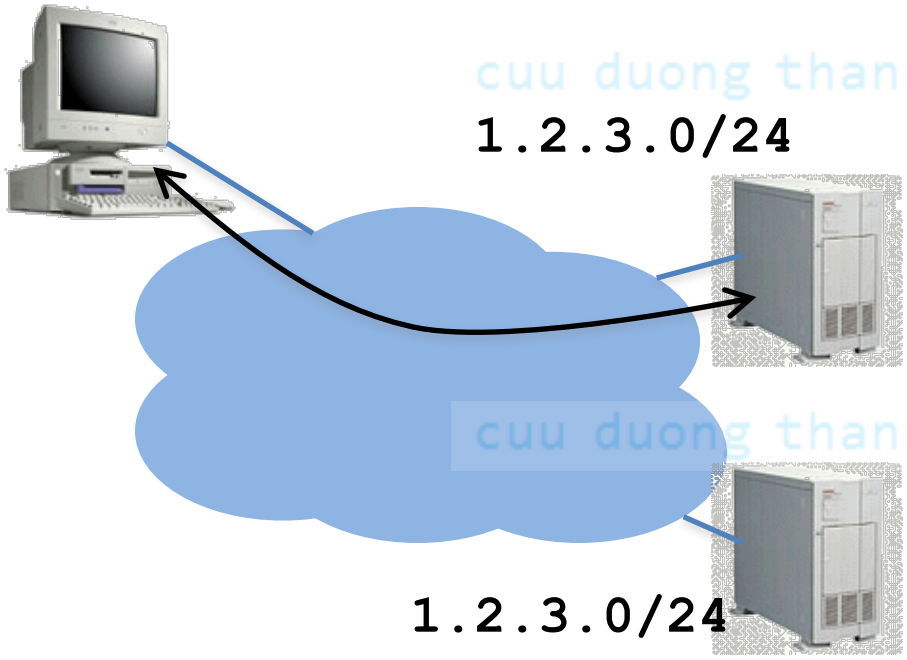
- Application
  - HTTP redirection



- Advantages
  - Fine-grain control
  - Selection based on client IP address
- Disadvantages
  - Extra round-trips for TCP connection to server
  - Overhead on the server

Server Selection Mechanism

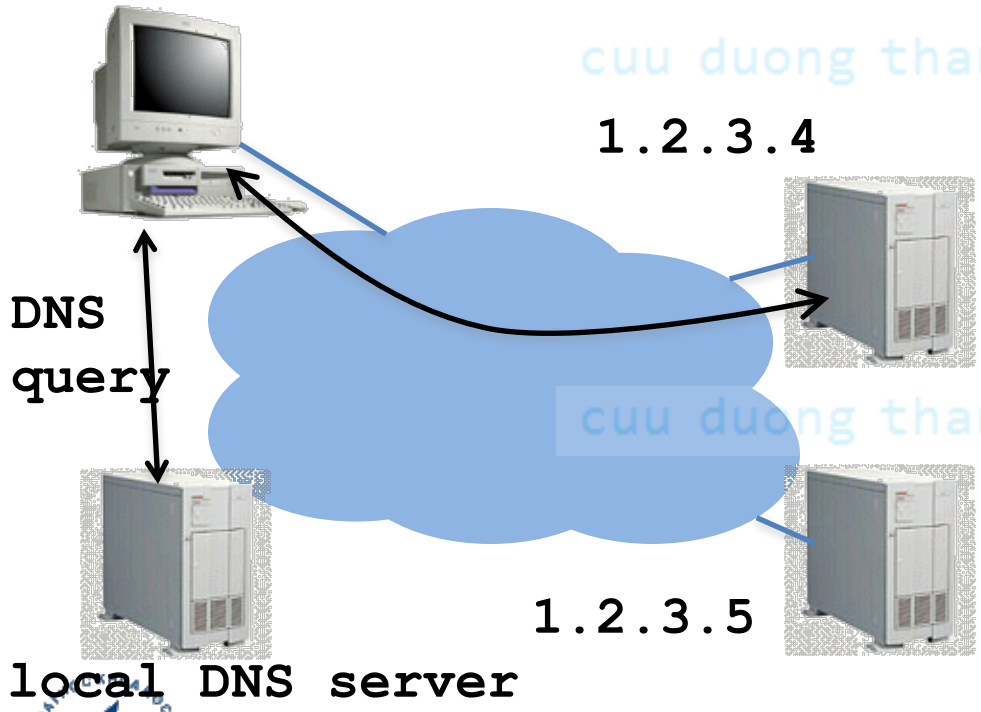
- Routing
  - Anycast routing



- Advantages
  - No extra round trips
  - Route to nearby server
- Disadvantages
  - Does not consider network or server load
  - Different packets may go to different servers
  - Used only for simple request-response apps

Server Selection Mechanism

- Naming
  - DNS-based server selection



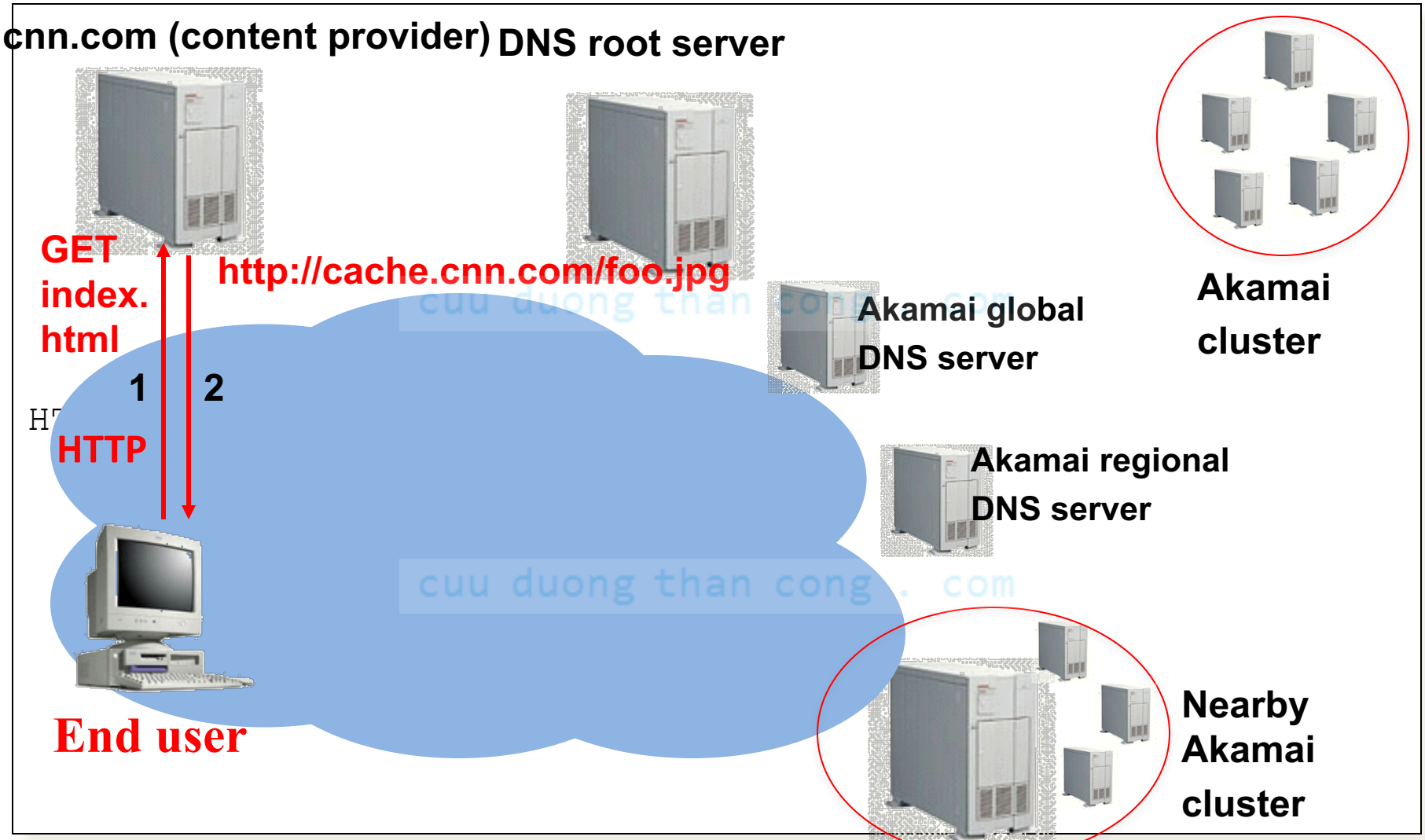
- Advantages
  - Avoid TCP set-up delay
  - DNS caching reduces overhead
  - Relatively fine control
- Disadvantage
  - Based on IP address of local DNS server
  - "Hidden load" effect
  - DNS TTL limits adaptation

How Akamai Works

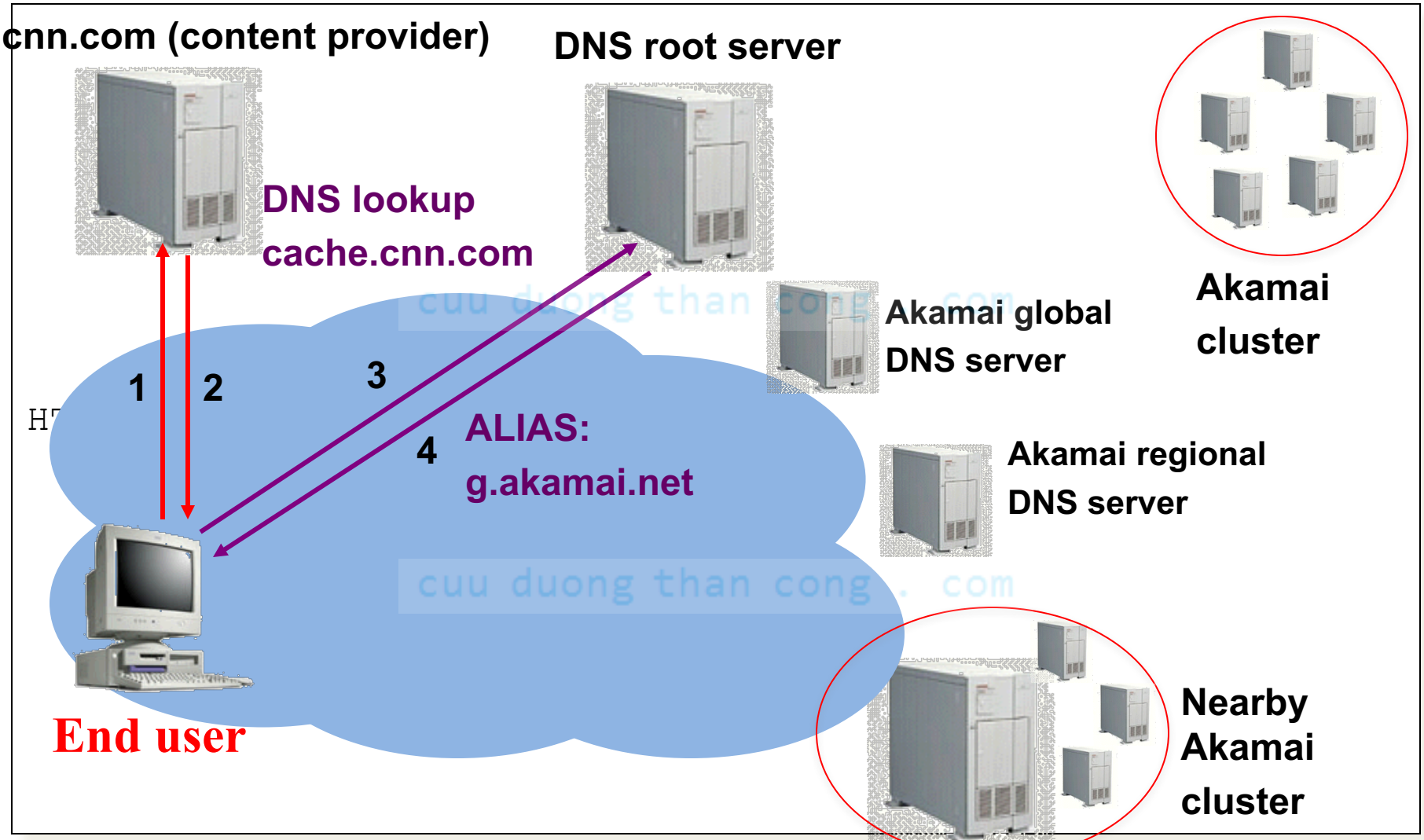
Akamai Statistics

- Distributed servers
  - Servers: ~61,000
  - Networks: ~1,000
  - Countries: ~70
- Many customers
  - Apple, BBC, FOX, GM IBM, MTV, NASA, NBC, NFL, NPR, Puma, Red Bull, Rutgers, SAP, …
- Client requests
  - Hundreds of billions per day
  - Half in the top 45 networks
  - 15-20% of all Web traffic worldwide

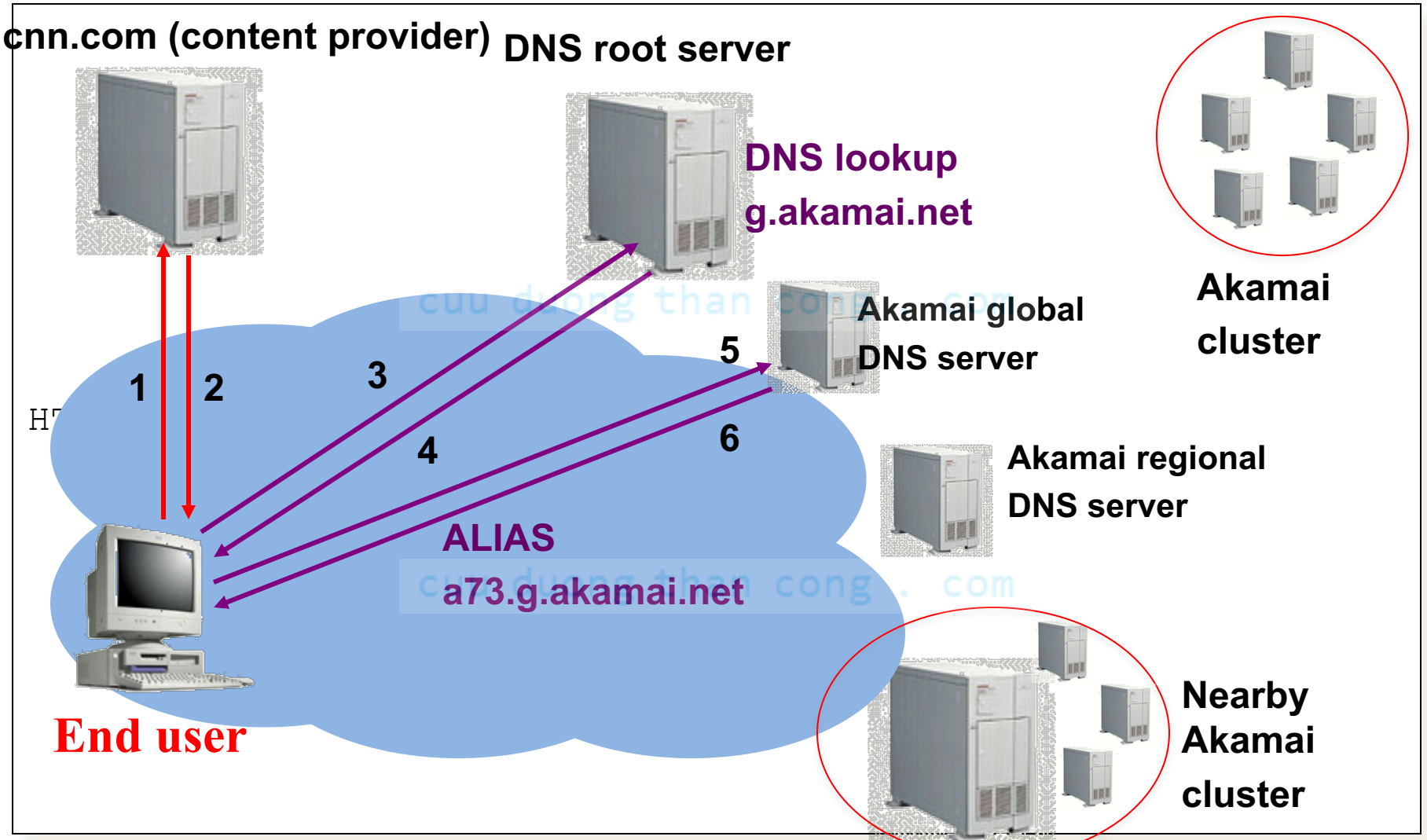
How Akamai Uses DNS



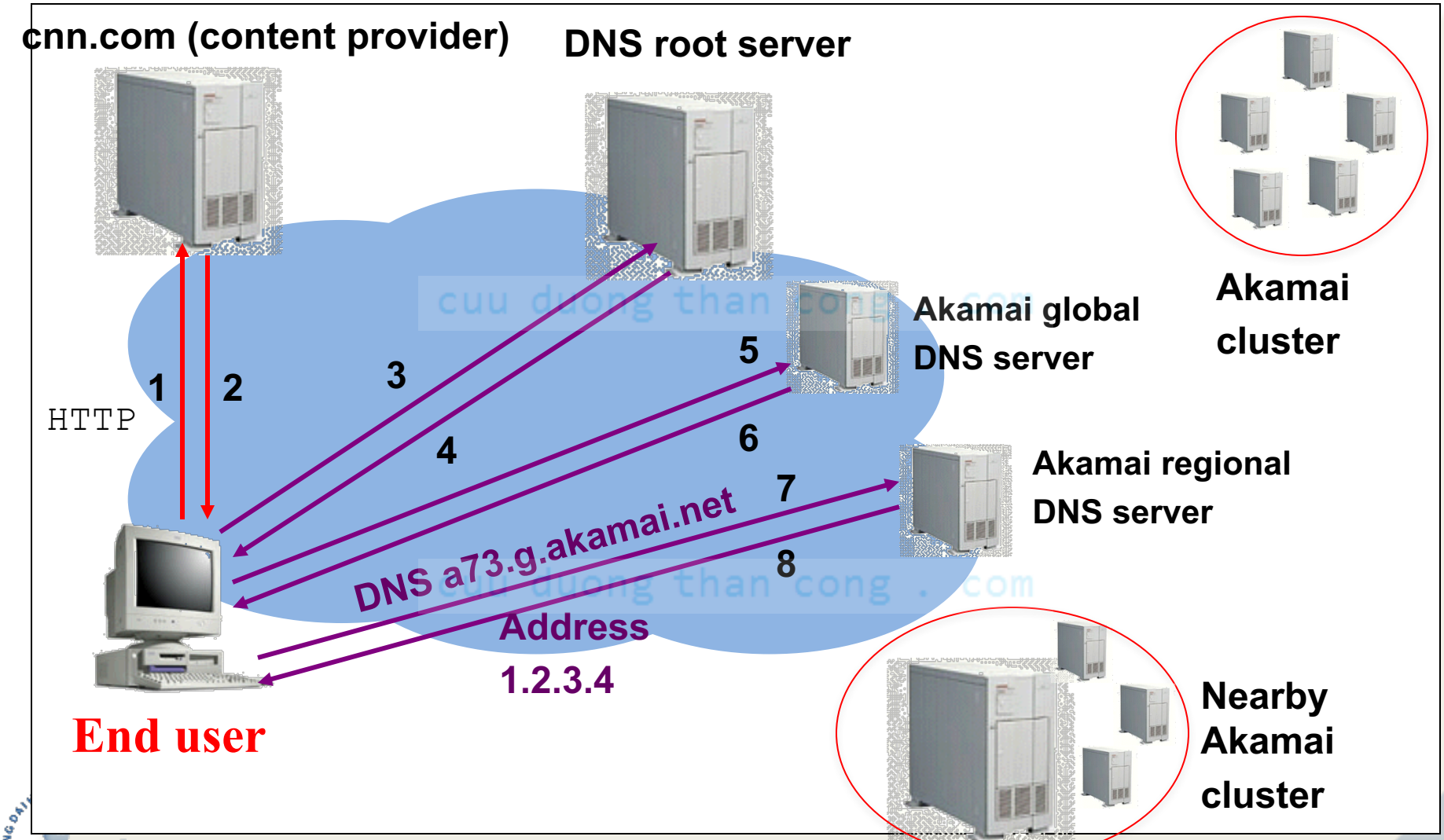
How Akamai Uses DNS



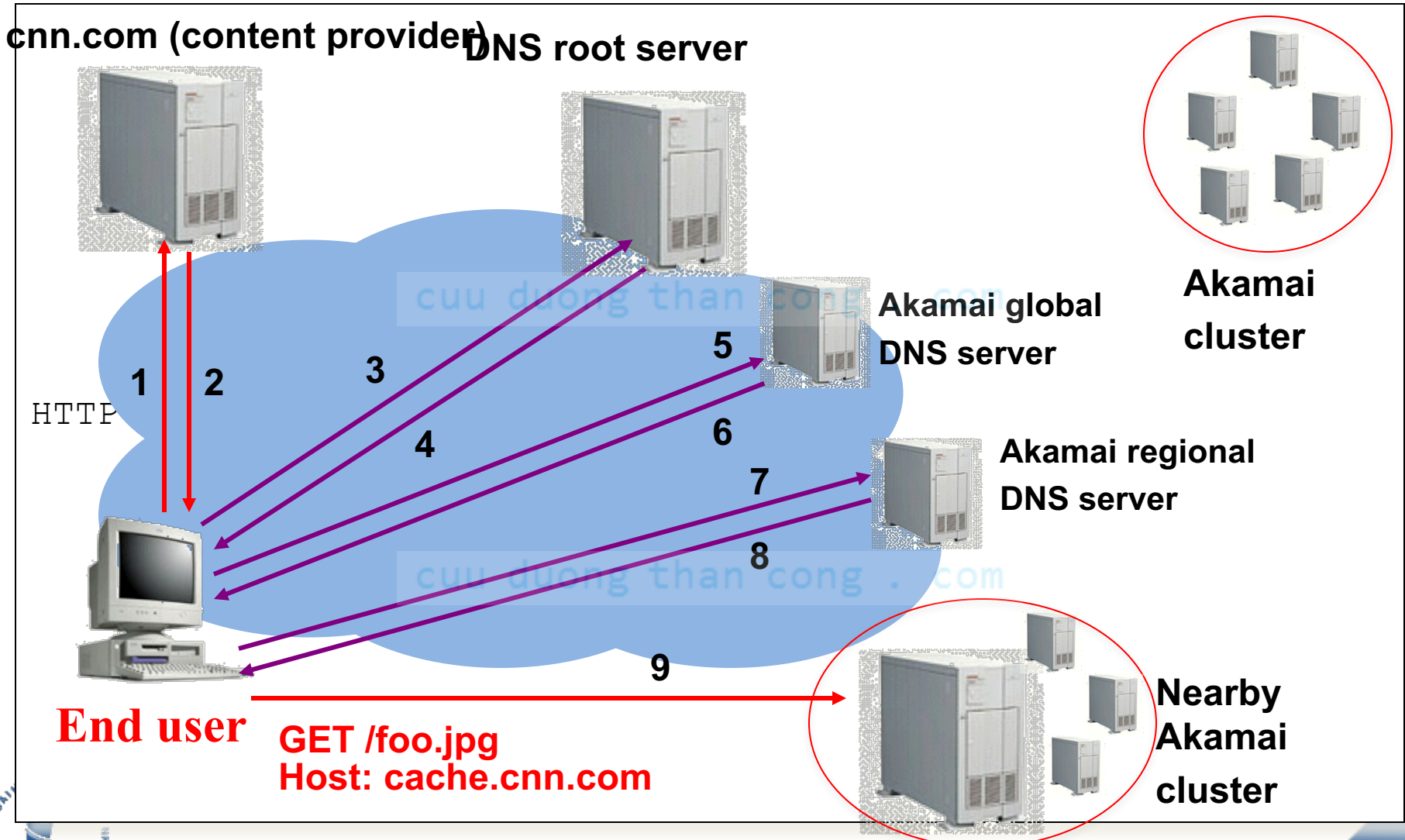
How Akamai Uses DNS



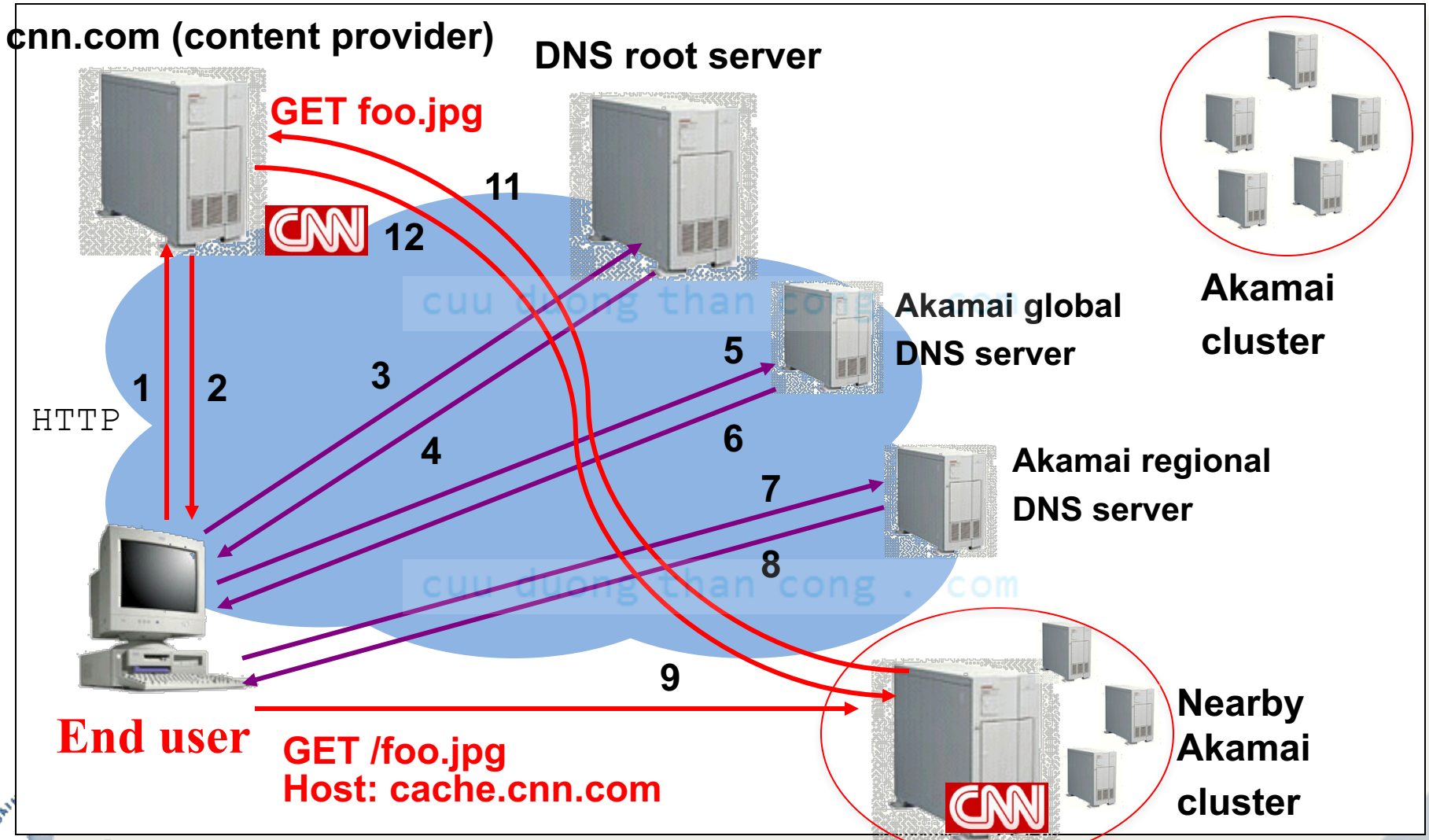
How Akamai Uses DNS



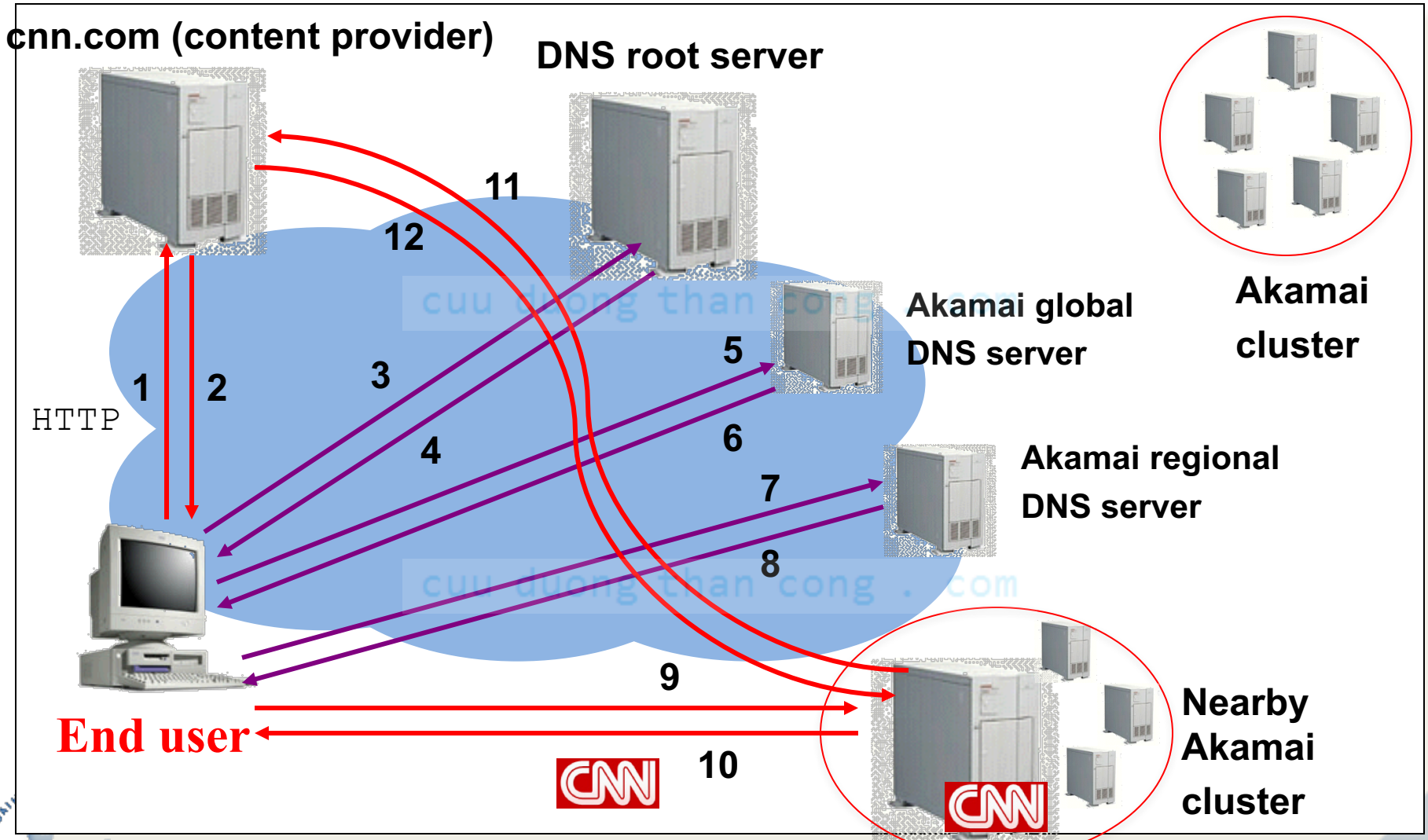
How Akamai Uses DNS



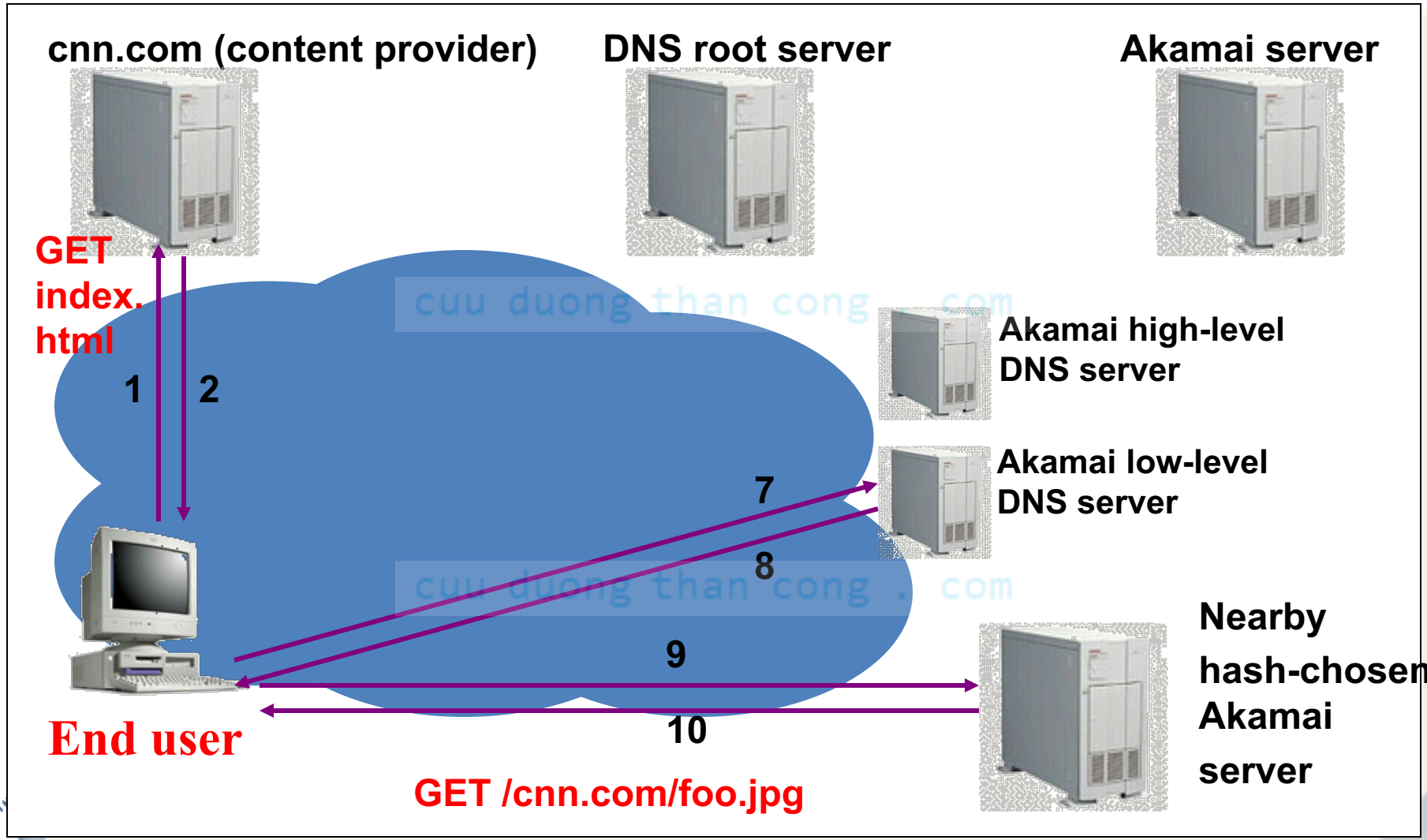
How Akamai Uses DNS



How Akamai Uses DNS



How Akamai Works: Cache Hit



Hỏi & Đáp