

# Skynet: A Cloud-Based Data Transfer Architecture



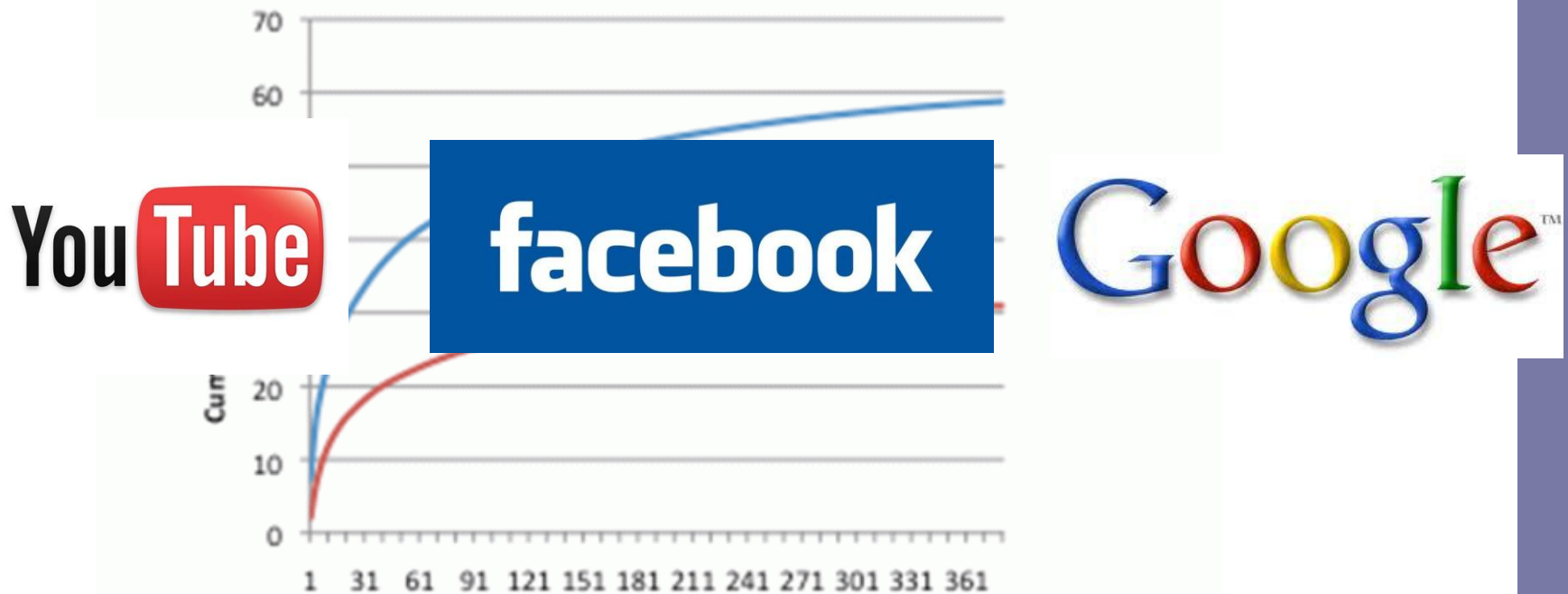
---

Aleksandar Kuzmanovic

<http://networks.cs.northwestern.edu>

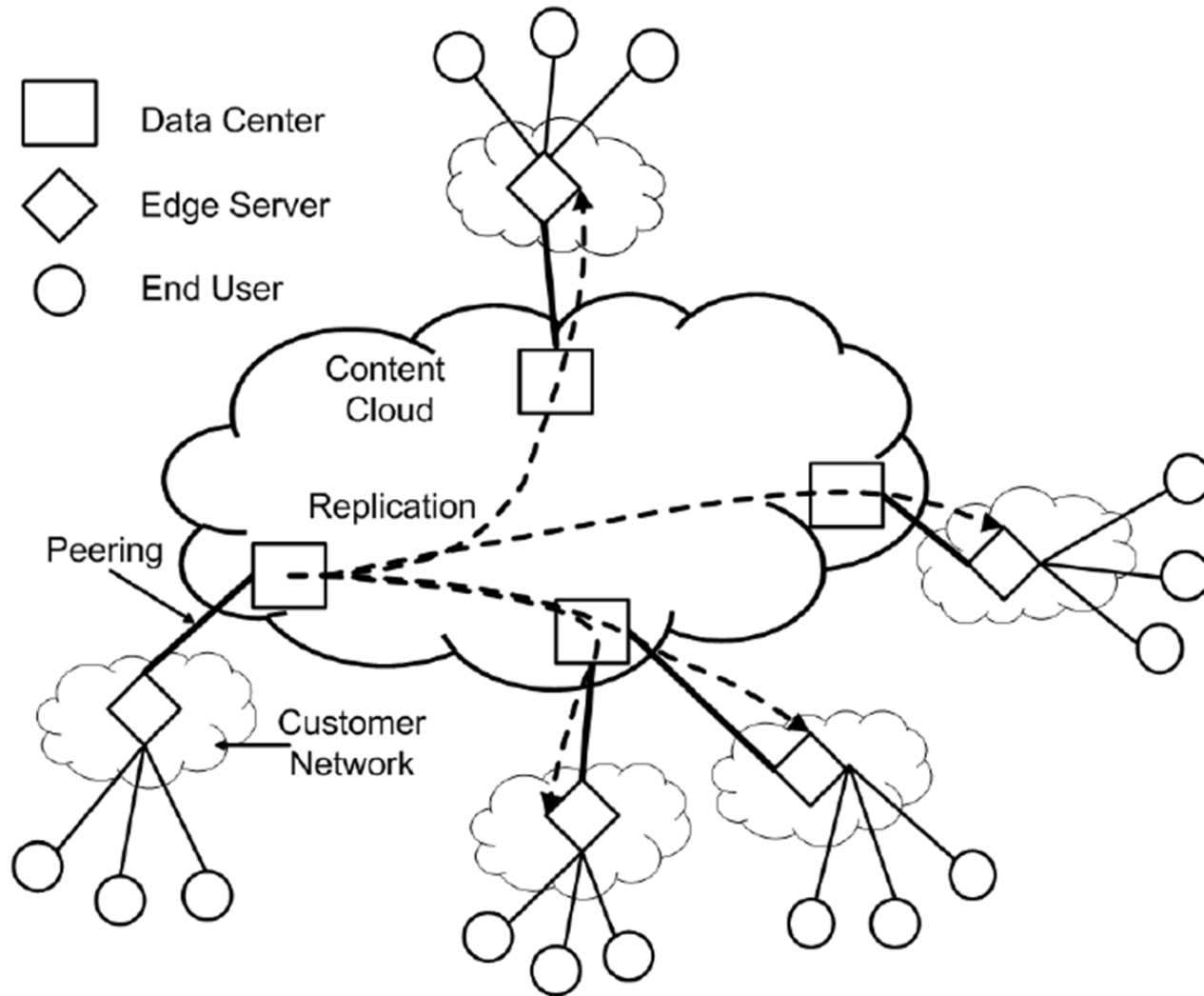
# The Rise of "Hyper Giants"

## Consolidation of Content



- Top 150 content providers account for 50% of traffic
- Top 30 content providers account for 35% of traffic

# Hyper Giants Morphing into CDNs



# Hyper Giants' CDN Examples

- More than **60% of traffic** flows directly between Google and consumer networks
  - Good for Google and consumer networks
- Facebook applies the same approach
  - More than 25% of Facebook runs through direct peerings with last-mile providers
- Microsoft started building its own CDN



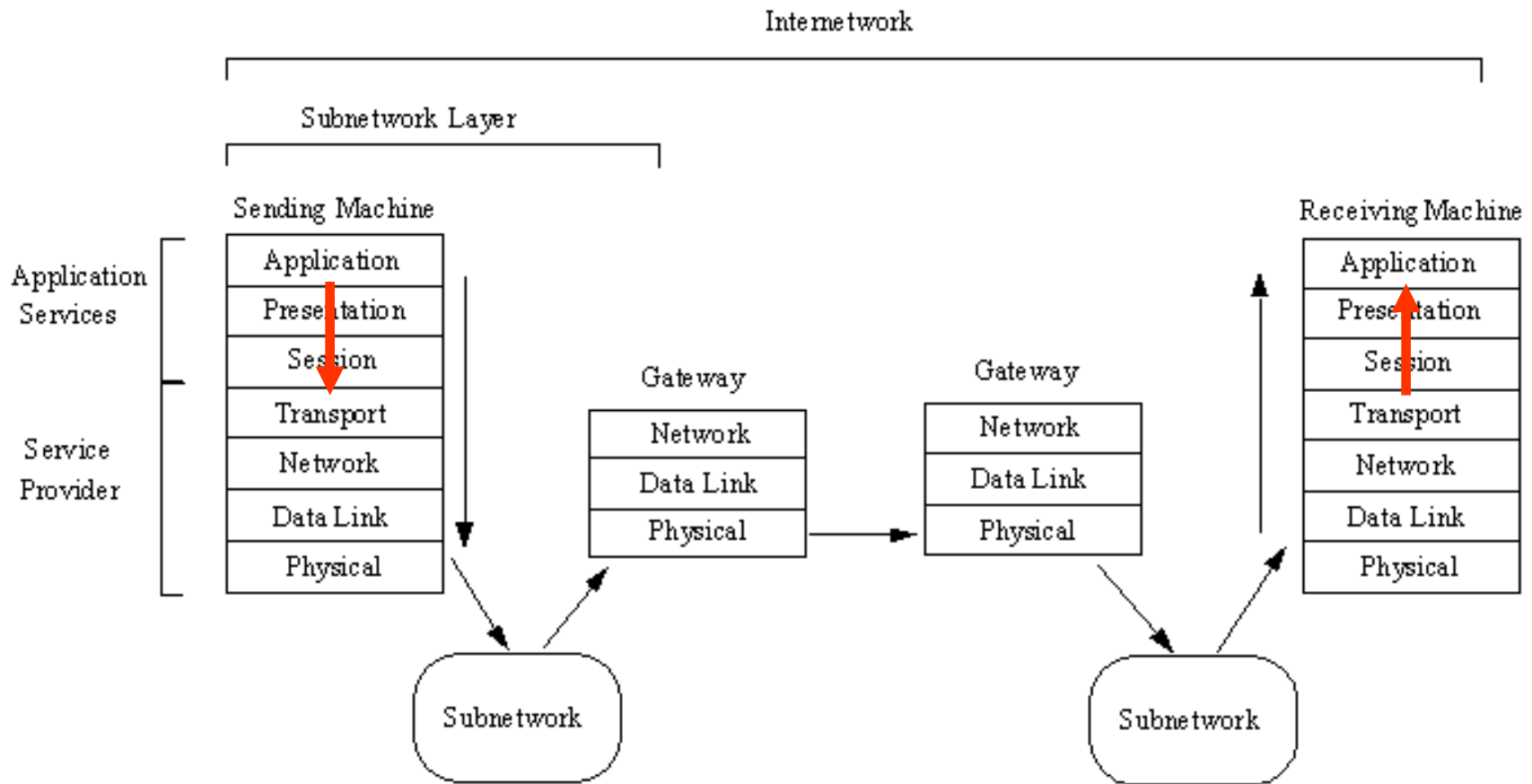
Google Global Cache Server

# Hyper Giants CDNs' Properties

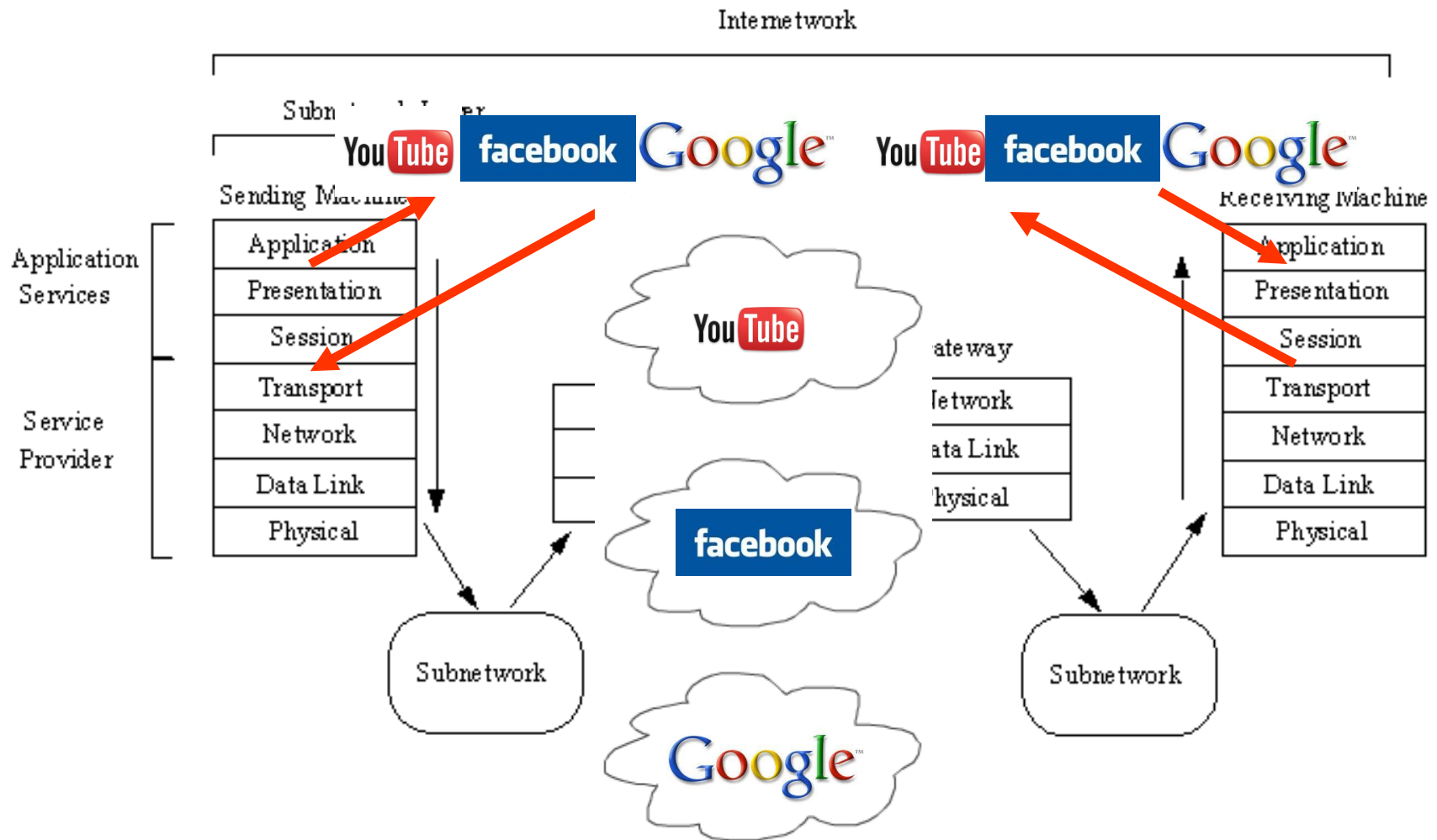
---

- Hyper giants' CDNs are different from traditional CDNs (e.g., Akamai's):
  - Support bi-directional data transfers, i.e., both uploads and downloads
    - Needed for various Web 2.0 apps
  - Agile replication of content towards anticipated receivers
- Research question:
  - Can we build a generic data transfer architecture on top of clouds?

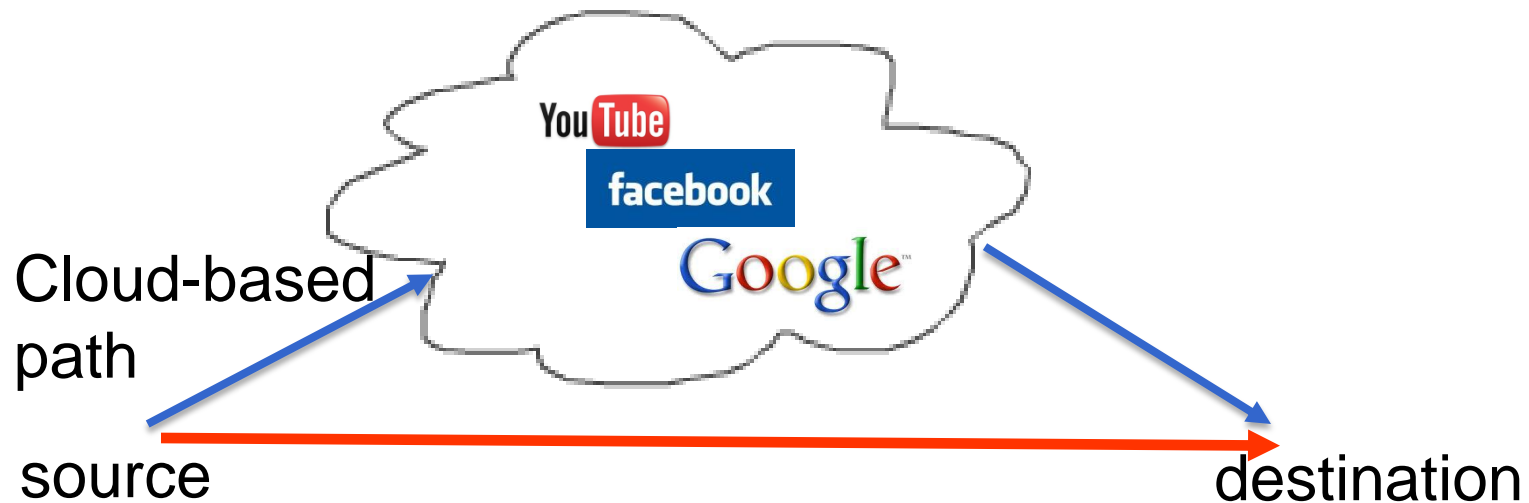
# Legacy Data Transfer



# Cloud-Based Data Transfer



# Why Should This Work?



- Cut the **e2e Internet path** into several shorter-RTT path chunks
- Avoid Internet bottlenecks
- Benefit from agile cloud replication

# Deployment Scenarios

---

- The bad

- “Free-riding” not likely to be liked by clouds:
  - Private resources used by 3<sup>rd</sup> parties
    - The problem is a transparent data transfer service provided by clouds
  - Hard to detect because:
    - Large amounts of data
      - » we were able to push 100 Gbytes with no problems
    - Opening multiple accounts is straightforward

- The good

- Potential for selling data acceleration services to 3<sup>rd</sup> parties
  - such that the performance of host applications is not degraded







# Remaining Outline

---

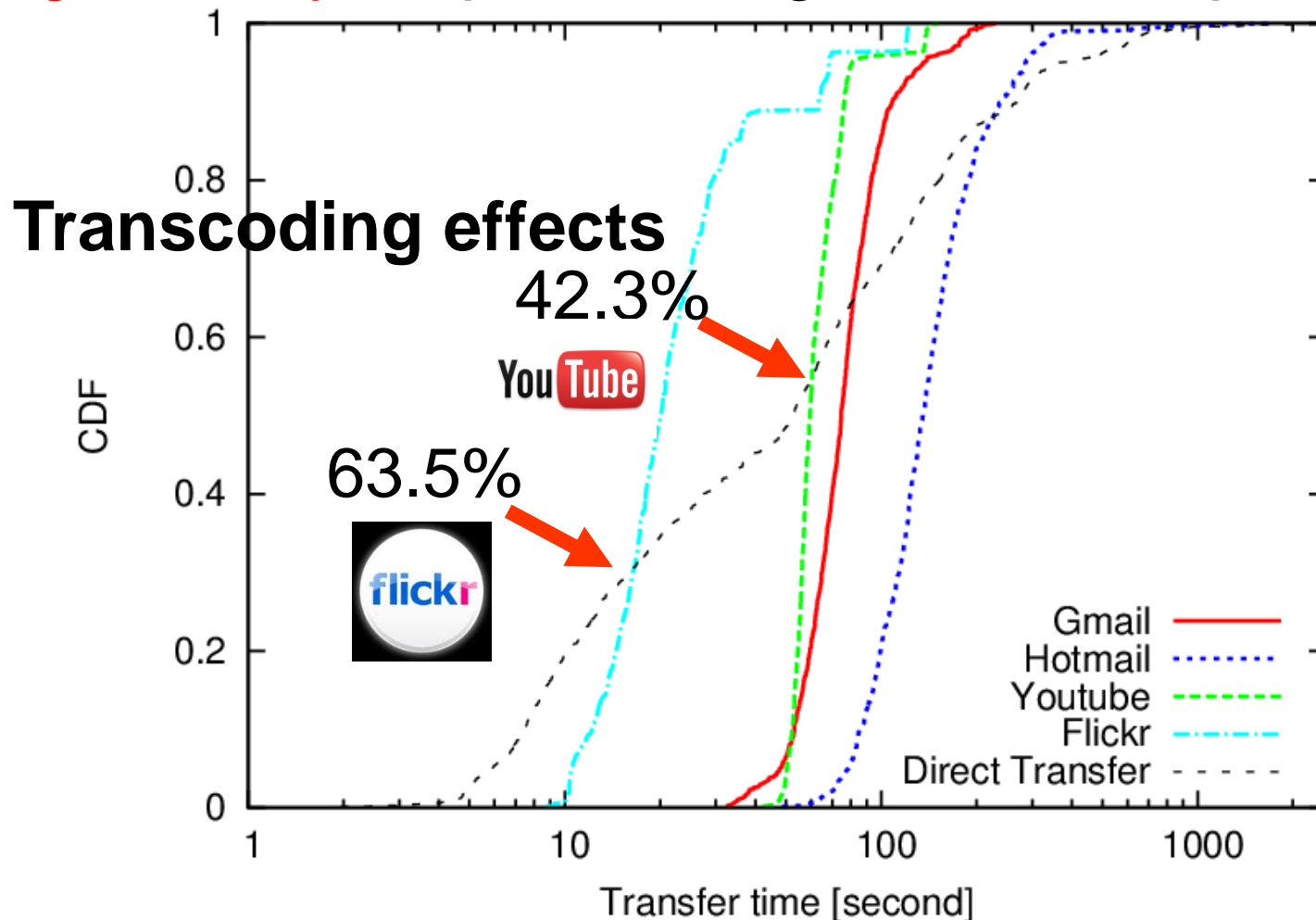
- Cloud selection and properties
- Per-cloud data transfer performance
- Decomposing cloud-based data transfers
- Multi-cloud homing
- Cloud-based multicast
- Cloud path consistency
- Cloud-hopping overlays
- Cloud-supported overlays

# Investigated "Carrier Applications"

Region	# of IPs							
								
	D	U	D	U	D	U	D	U
N. America	17	17	4	1	1917	15	12	1
Europe	6	6	0	0	2591	8	0	1
Asia	0	0	0	0	1186	4	0	0
S. America	0	0	0	0	154	1	0	0
Oceania	0	0	0	0	285	0	0	0
Unknown	0	0	2	1	99	0	12	2
Total	23	23	6	2	6232	28	1	2

# Per-Cloud Performance

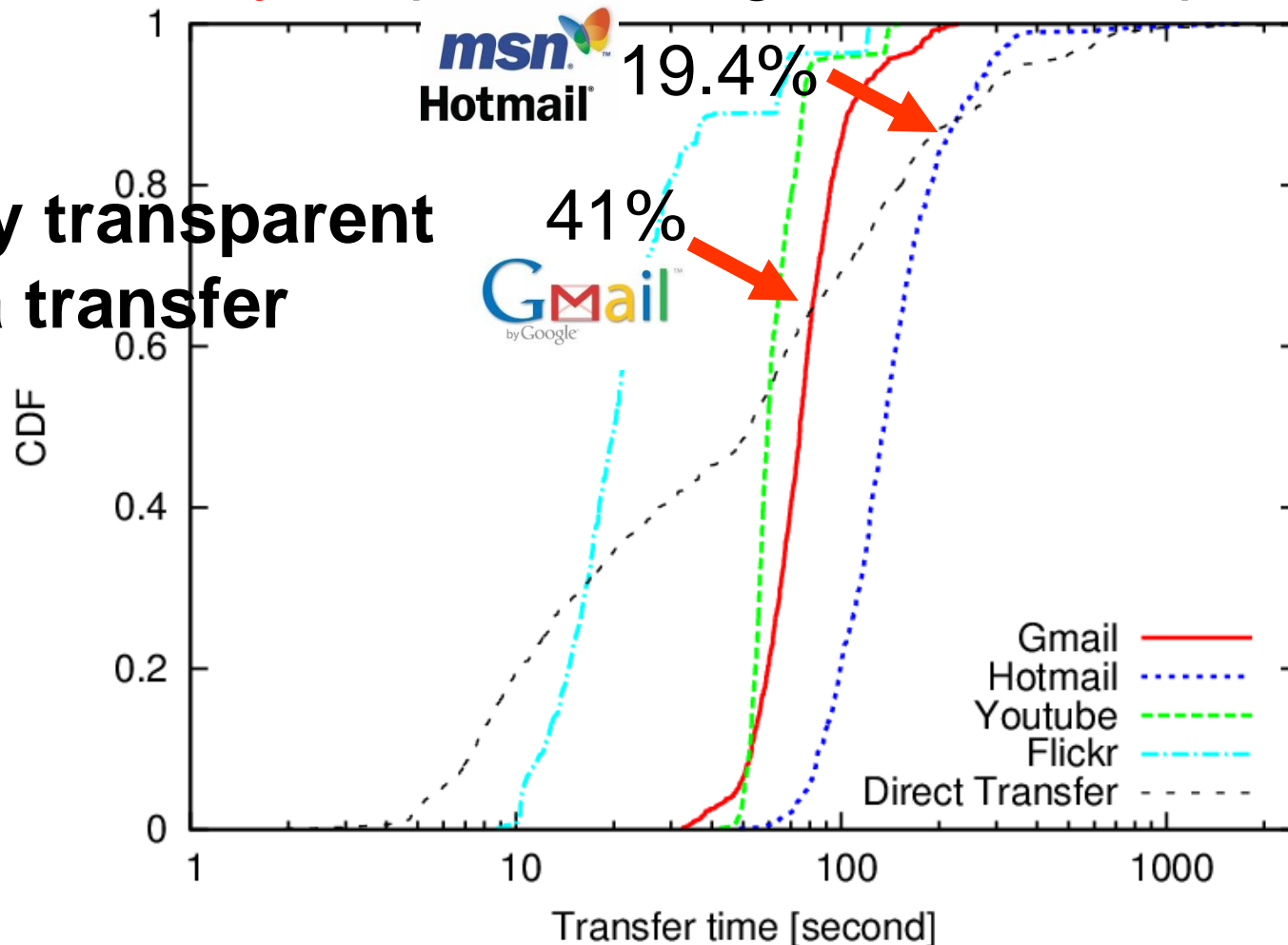
Cloud-based paths can *often*, *consistently* and *significantly* outperform regular Internet paths



# Per-Cloud Performance

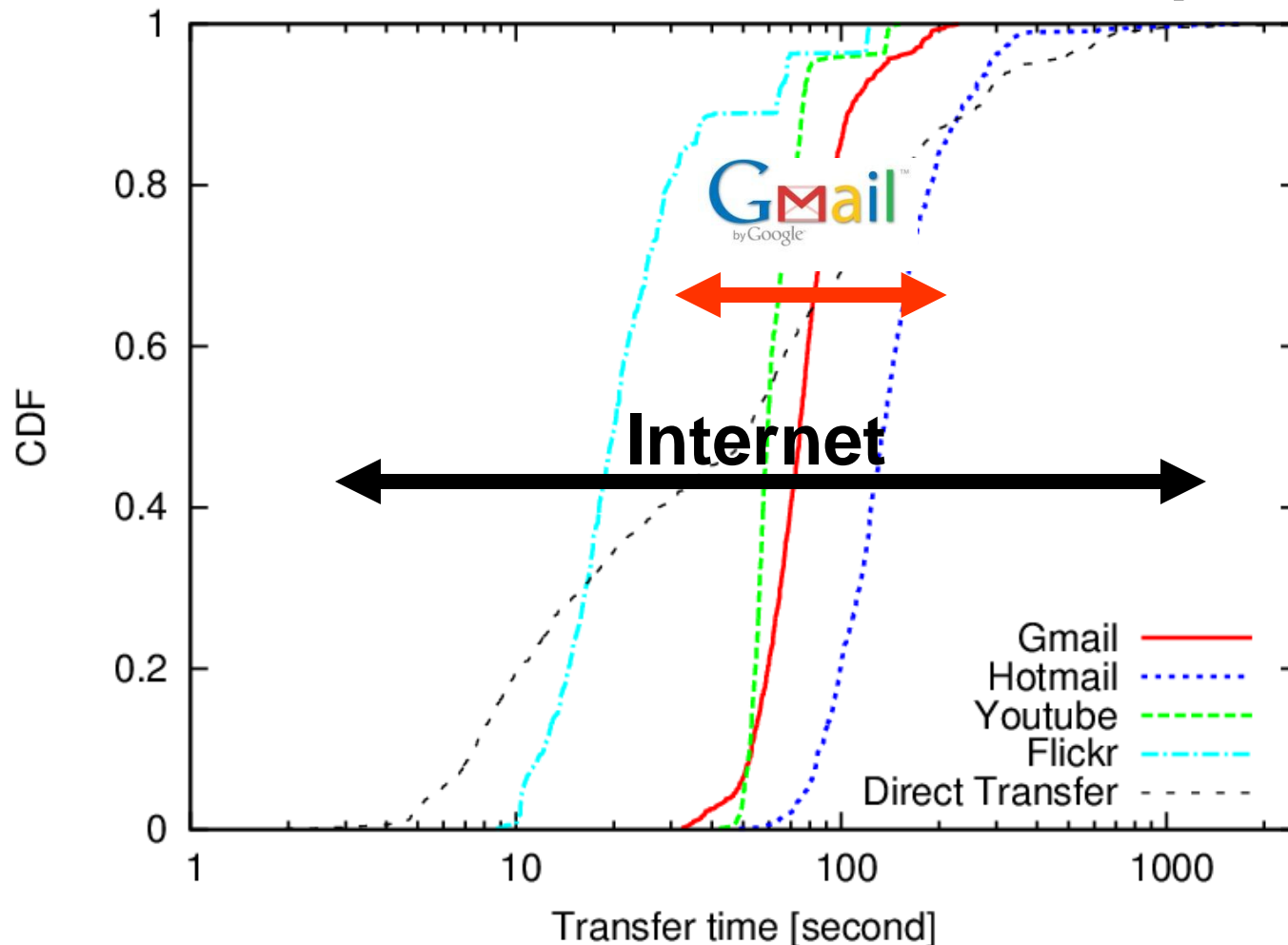
Cloud-based paths can *often*, *consistently* and *significantly* outperform regular Internet paths

Fully transparent  
data transfer



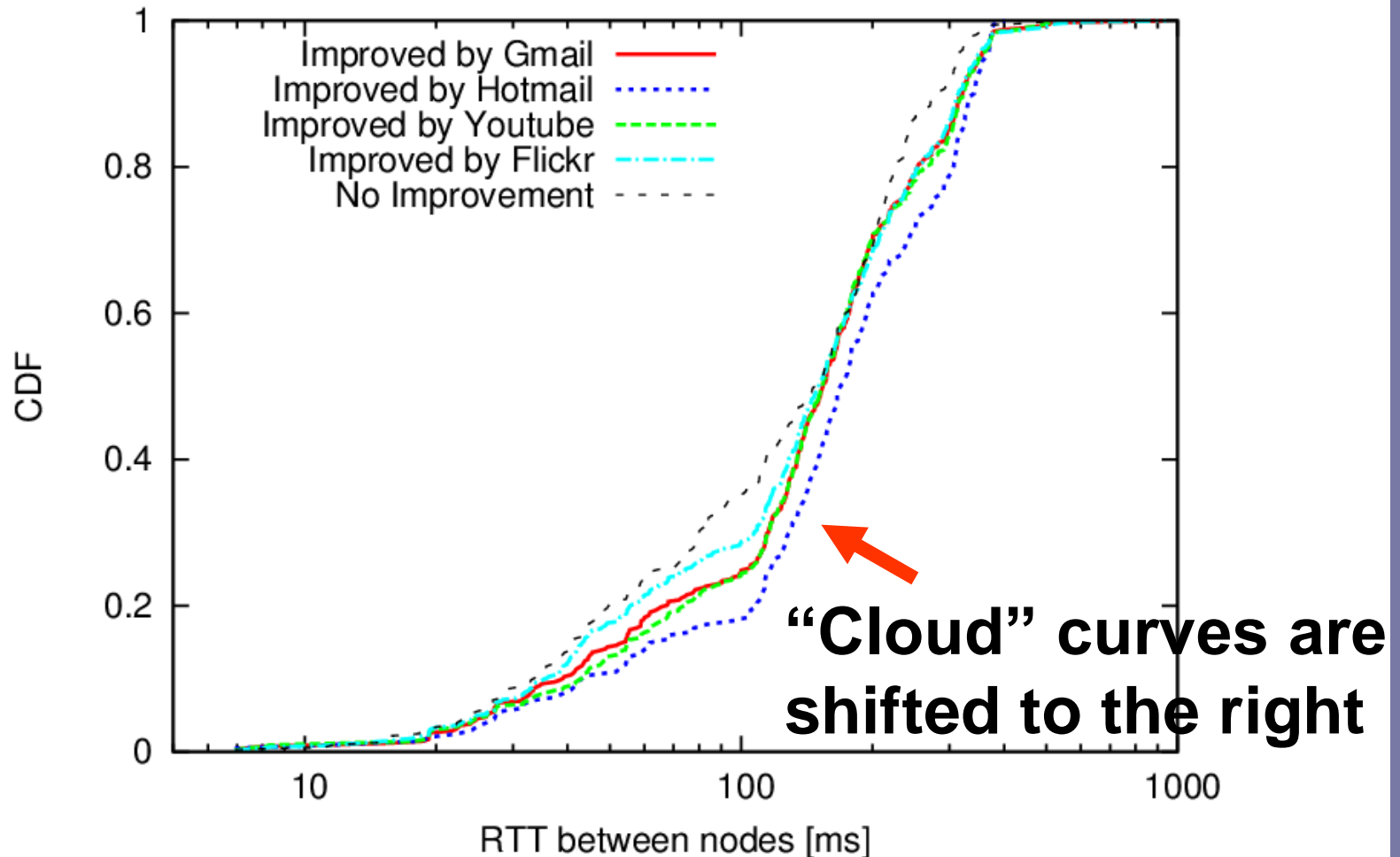
# Transfer-Time Variance

Variance of transfer times is much smaller for **cloud-based transfers** than for **Internet paths**



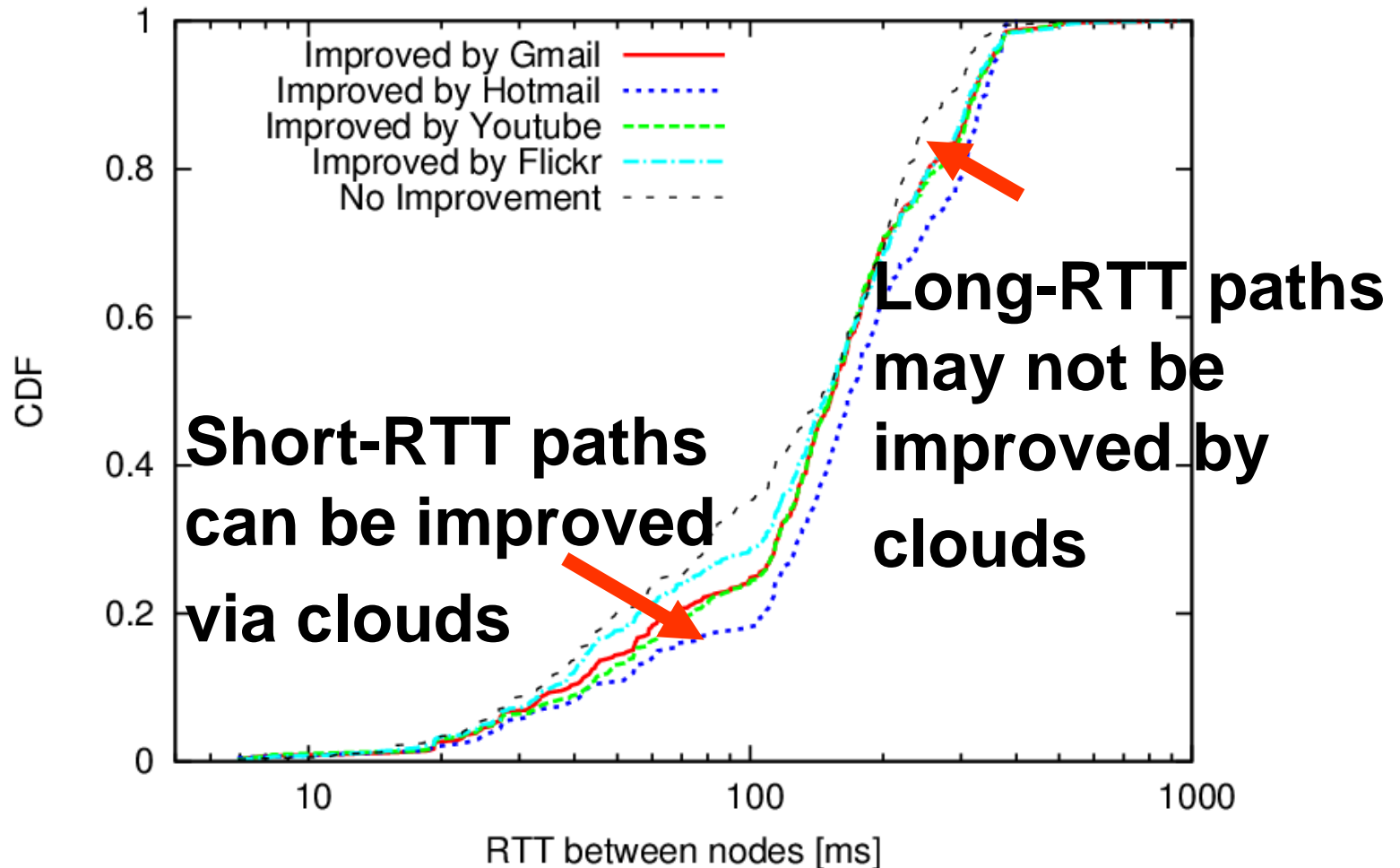
# The Role of RTTs

Longer-RTT paths are more likely to be improved via clouds, but other factors play the role as well



# The Role of RTTs

Longer-RTT paths are more likely to be improved via clouds, but other factors play the role as well



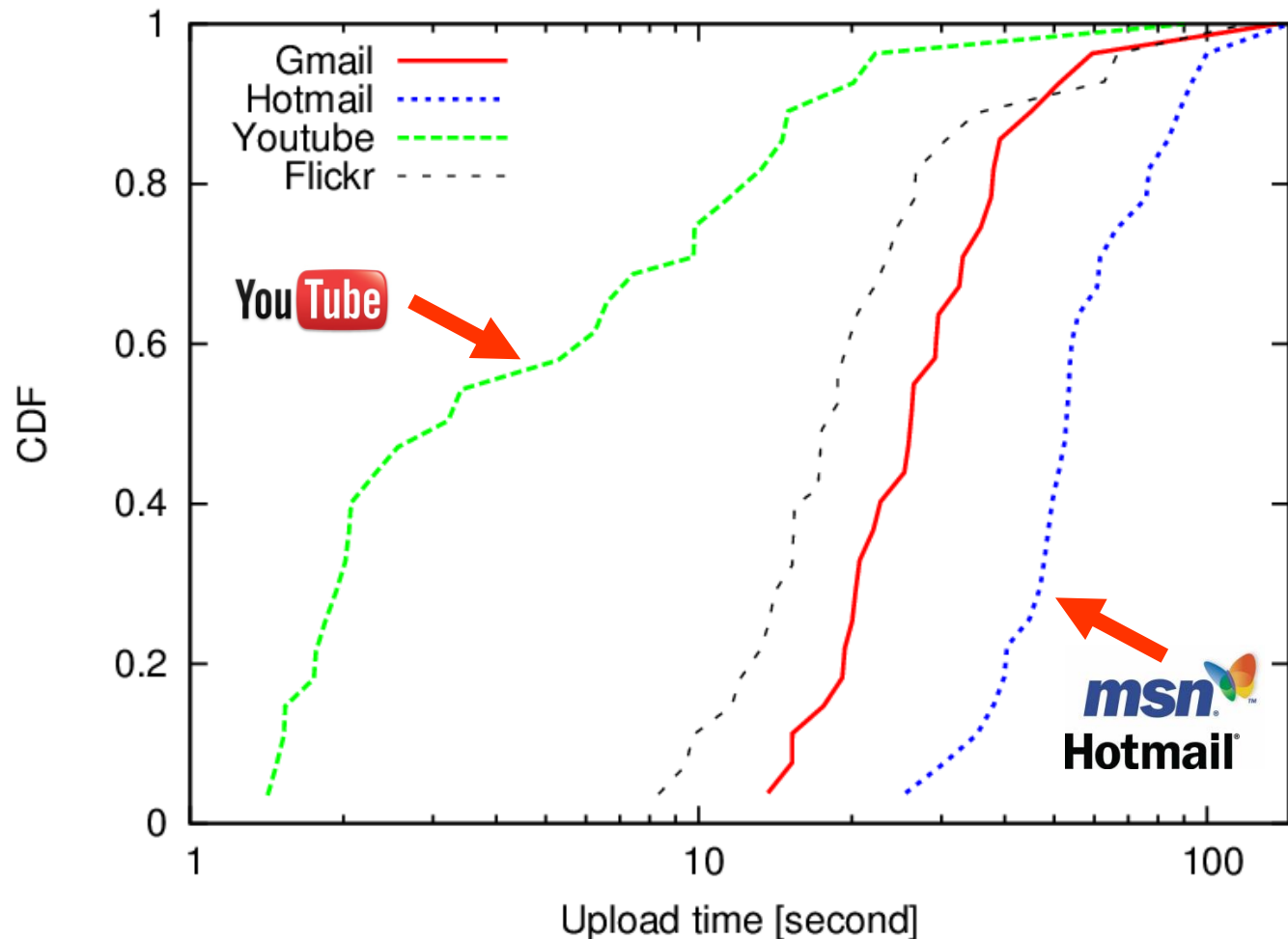
# Remaining Outline

---

- Cloud selection and properties
- Per-cloud data transfer performance
- **Decomposing cloud-based data transfers**
- Multi-cloud homing
- Cloud-based multicast
- Cloud path consistency
- Cloud-hopping overlays
- Cloud-supported overlays

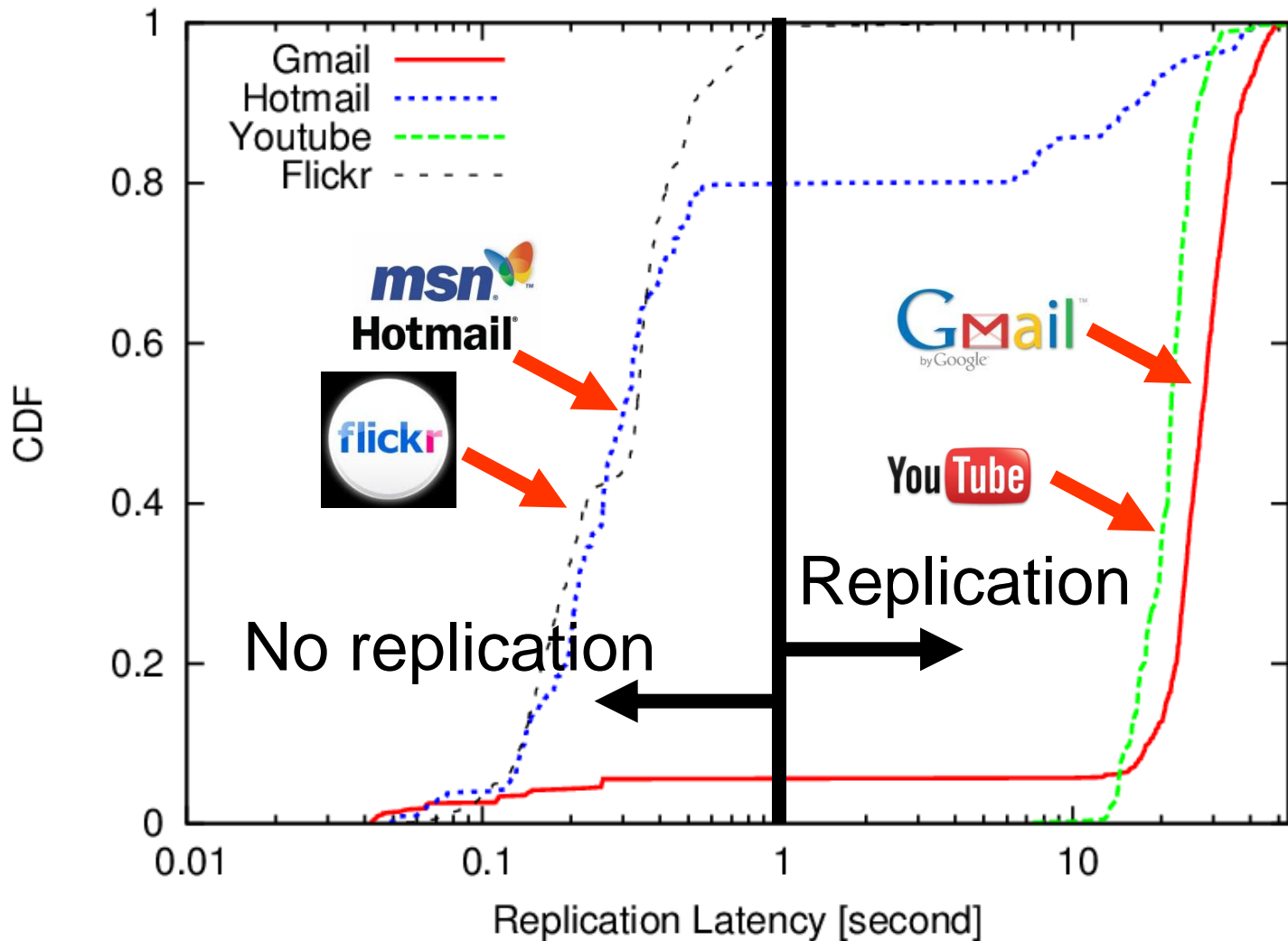
# Upload Latency

Upload latency depends upon the scale and distribution of upload servers



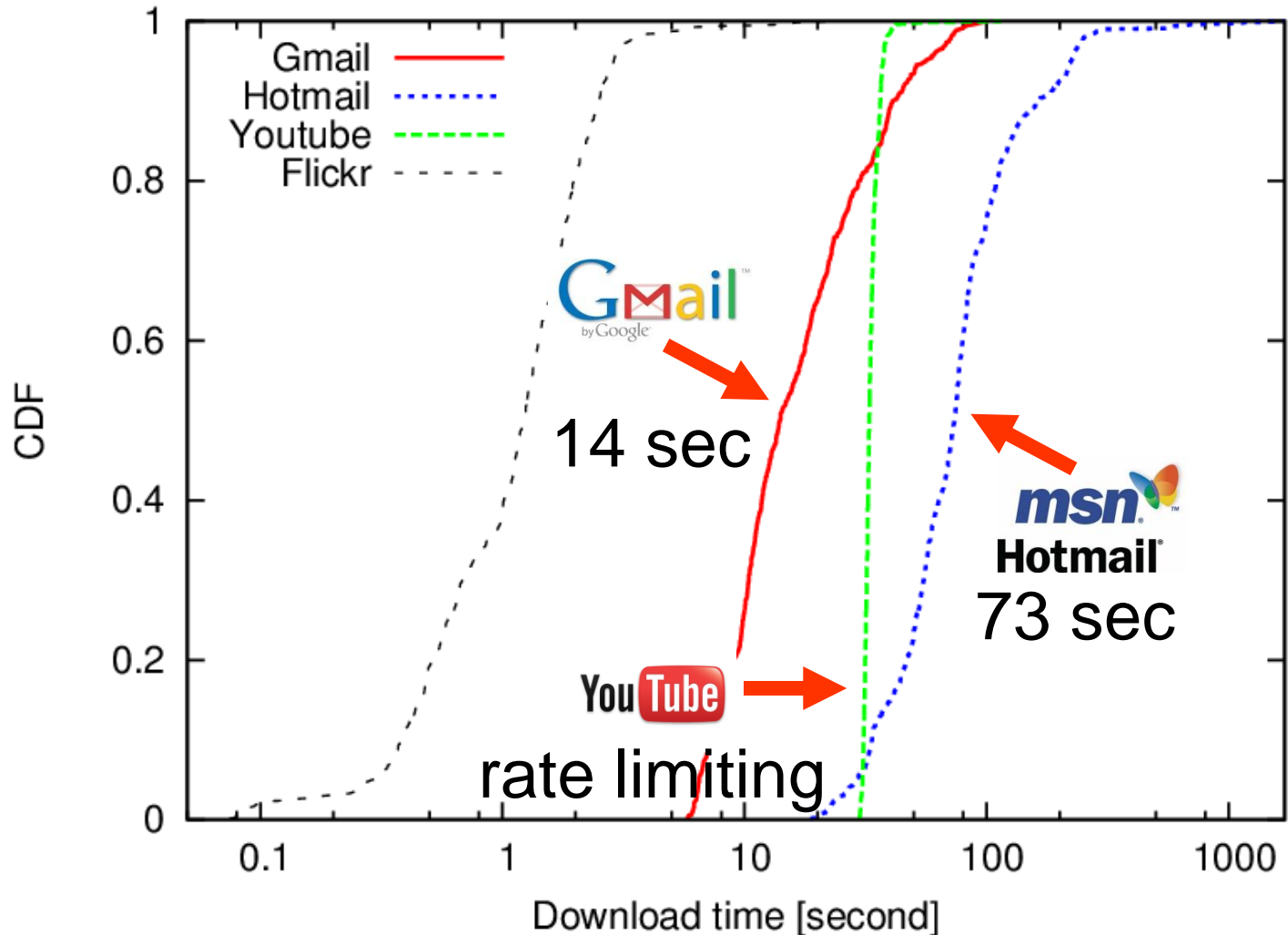
# Replication Latency

## Different replication policies



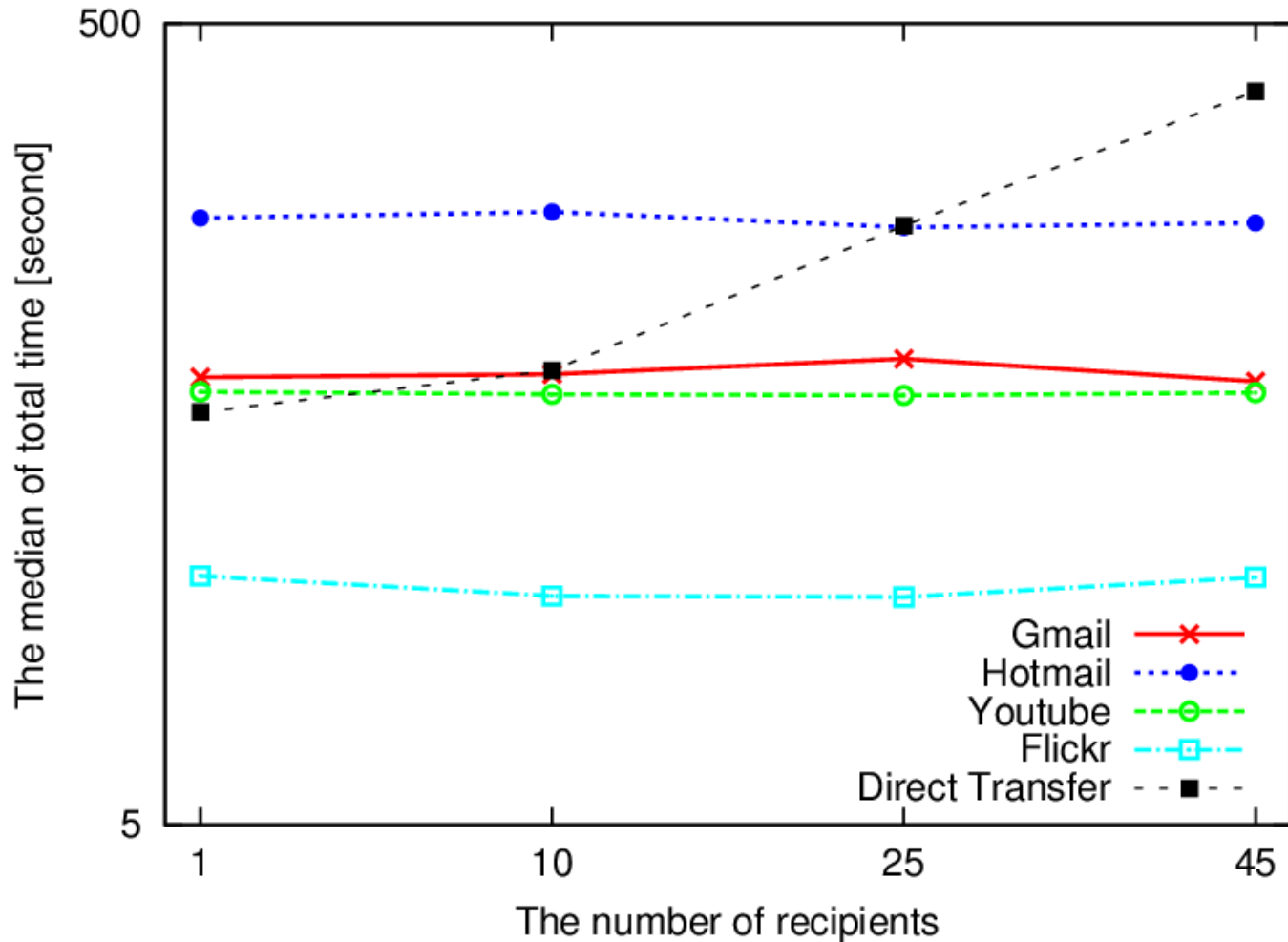
# Download Latency

Pushing data closer to end users pays off



# Cloud-Based Multicast

For multicast group sizes  $> 25$ , cloud-based multicast outperforms direct multicast for all clouds



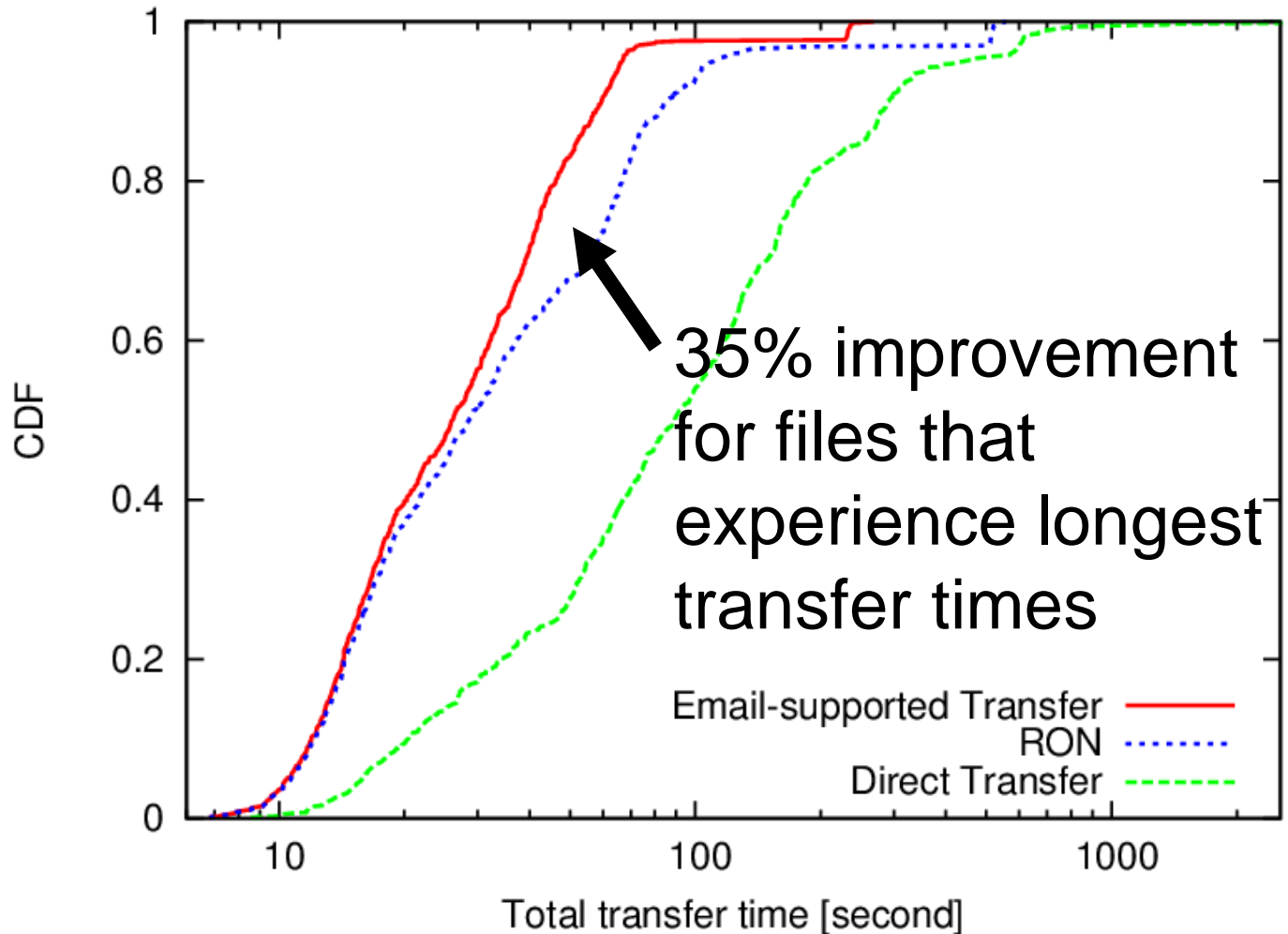
# Remaining Outline

---

- Cloud selection and properties
- Per-cloud data transfer performance
- Decomposing cloud-based data transfers
- Multi-cloud homing
- Cloud-based multicast
- Cloud path consistency
- Cloud-hopping overlays
- **Cloud-supported overlays**

# Cloud-Supported Overlay

Cloud-based paths bring additional benefits in overlay scenarios



# Remaining Outline

---

- Cloud selection and properties
- Per-cloud data transfer performance
- Decomposing cloud-based data transfers
- **Multi-cloud homing**
- Cloud-based multicast
- **Cloud path consistency**
- **Cloud-hopping overlays**
- Cloud-supported overlays

# Conclusions

---

- The rise of hyper giants and their transparency enable cloud-based data transfers
- We demonstrated the feasibility of building data transfer services on top of clouds and their superior performance
  - Opens novel security challenges and may lead to an “arms race” in this area
  - Shows significant potentials for building explicit ties between legacy- and the new Internet

# Other Projects

---

- Monitoring net neutrality (NSF and Google)
  - Design auditing tools to enable ISPs' transparency
  - Measurement Lab
- Auditing Internet content (NSF)
  - How do we know that the information on the Web is not biased?
- Analyzing human mobility with applications in networked systems (Narus Inc.)
- Sound fusion project

# Sound Fusion Project

- A huge amount of live concert recordings is available online
  - But the quality can be very poor
- Can we fuse these recordings to generate a high-quality signal?
- Complications:
  - No reference that defines which part is music and which noise
  - SNR unknown
  - No pilot signal



# Thank You!

---

# Questions?

<http://networks.cs.northwestern.edu>