Latency Evaluation of Anycast CDN with Limited Cache Sets

Futo Masai
Noriaki Kamiyama

Ritsumeikan University

Background

- CDN (Content Delivery Network)
  - Using geographically distributing cache servers (CSs)
  - Reducing latency and distributing load
- Anycast CDN
  - Assign the same IP address to multiple CSs
  - Route user requests via BGP to the closest CS
  - Enables simple operation
- Microsoft Anycast CDN
  - Maintains multiple CS sets called ring, each of which is assigned a unique IP address
  - Always choose the ring with the most CS

Problems of Anycast CDN

1. Suboptimal CS Selection
   - 10 ～ 15% of users route to distant CSs *1
   - Deploy hundreds of CSs and decrease the impact of suboptimal selection
2. Increasing the Probability of Suboptimal Selections
   - Deploying more distribution sites decreased the probability of routing to optimal CS
3. Ignore Regional Bias in Content Popularity

- Previous Study *2
  - Propose a method to create multiple rings composed of a small number of CSs, which considers regional content popularity by using a Genetic Algorithm (GA)

*1 Matt Calder, et al., Analyzing the Performance of an Anycast CDN, ACM IMC 2015
*2 Chihiro Kato, at el., Designing Server Sets for Anycast CDN Using Genetic Algorithm, IEEE ICC 2024

Overview

- Limitations of Previous Work
  - Actual latency reduction effect was not evaluated.
- Purpose of this Study
  - Evaluates the latency reduction of multi-ring anycast CDNs using measured Round Trip Times (RTTs)
- Approach
  - Measure RTTs for optimal and suboptimal CS selections
  - Calculate delivery latency per content and average latency
  - Compare this method with a conventional single-ring anycast CDN

Cache Sets (ring) Configuration Method

- Creating multiple rings with a small number of CSs, considering regional content popularity using a GA

- Ring Construction Method Using GA
  - Gene: The collection of rings composed of multiple AS
  - Fitness: The average coverage rate by the AS placement of genes among the top 5 demanding countries for each content
  - Using tournament selection to choose genes

- Key parameters of GA
  - Initial gene: 300
  - Mutation rate: 0.1
  - Generation: 100,000

RTT Measurement and Latency Calculation

- RTT Measurement Method
  - Using the RIPE Atlas internet measurement platform
  - Measure RTT from Probes in content demand countries to the ASes

- Calculation of Content Latency L_m

$$L_m = \frac{\sum_{i=1}^{5} \left((1-p) O_{m,i} + p S_{m,i} \right) T_{m, r_{m,i}}}{\sum_{j=1}^{5} T_{m, r_{m,j}}}$$

  - Variables

$r_{m,i}$	The i -th country with the highest demand for content m	$T_{m,r_{m,j}}$	Demand ratio of country $r_{m,i}$ for content m
$O_{m,i}$	RTT from $r_{m,i}$. to the **optimal** AS	p	Probability of a suboptimal AS being selected
$S_{m,i}$	RTT from $r_{m,i}$. to the **suboptimal** AS		

Determination of Delivery AS

1. Check if an AS exists within the country of content demand

2. If an AS is present:
   - Measure the RTT to each AS within that country
   - Set the AS with the best RTT as the optimal AS
   - Set the AS with the worst RTT as the suboptimal AS

3. If no AS is present:
   - Set the geographically closest AS as both the optimal and suboptimal AS

Evaluation Conditions

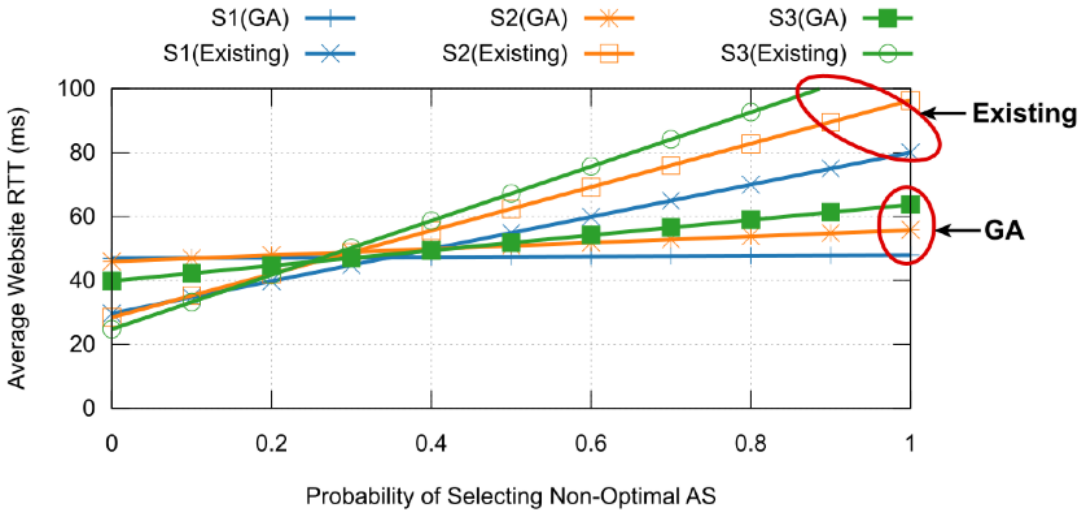
- GA Method
  - Multiple rings constructed from a small number of CSs
  - The number of rings is 20
  - The number of ASes constituting the ring is 5, 10, or 15

- Existing Method
  - Single ring composed of numerous CSs
  - Integrate Anycast CDN configured with GA into one ring

	$s1$		$s2$		$s3$	
	N	k	N	k	N	k
GA Method	20	5	20	10	20	15
Existing Method	1	84	1	154	1	198

Performance Evaluation

- The average RTT against the probability p of suboptimal AS selection
- The method using GA is stable
- The existing method increases RTT with increasing p
- Lines cross at
  - s1: $p = 0.35$
  - s2: $p = 0.31$
  - s3: $p = 0.25$



- 55% of users are routed to optimal CSs *
  - Using GA method is effective

* Matt Calder, et al., Analyzing the Performance of an Anycast CDN, ACM IMC 2015

Conclusion

- Evaluated the latency reduction effect of anycast CDNs that have multiple rings consisting of a small number of CSs using measured RTTs
- Compared with existing methods that have a single ring composed of multiple CSs
- When the probability p of suboptimal selection exceeds 0.25–0.35, the method using GA is effective.

Acknowledgment

- This work was supported by a grant from the Takahashi Industrial and Economic Research Foundation.

- This presentation is supported by the “Society for the Advancement of Science and Technology at Ritsumeikan”