

遺伝的アルゴリズムを用いた Anycast CDN の配信サーバ選択法

Designing Server Sets for Anycast CDN Using Genetic Algorithm

加藤 千尋

上山 憲昭

Chihiro Kato

Noriaki Kamiyama

立命館大学情報理工学部

College of Information Science and Engineering, Ritsumeikan University

1. はじめに

近年、Anycast CDN の効果的な運用に対する関心が高まっている。既存の Anycast CDN は、コンテンツの人気地域性を考慮せずに、すべてのコンテンツ要求に対して同一のキャッシュサーバ (CS) セットを回答する。また Anycast 配信では同一の IP アドレスに対して割り当てられるサーバの数の増加に伴い、配信サーバの選択の適正度が低下することが報告されている。そこで筆者らはこれらの問題を解決するために、コンテンツの人気の空間的局所性を取り入れた新しい Anycast 配信手法を提案した [5]。具体的には、空間的な人気の偏りを考慮して少数の CS から構成される複数の CS セットを作成し、各コンテンツに最適な CS セットを割り当てる貪欲法によるアルゴリズムを提案した。CS セットの構成 CS 数を最小限に抑えることで、広範囲の国々をカバーしながらも、CS 選択の適正度を向上させ、配信遅延時間の分散を低減することを目標としている。本稿では、CS セットを遺伝的アルゴリズムを用いて最適化することを提案する。そして高人気 Webpage の国別の要求数に関するデータと AS トポロジデータを収集し、各コンテンツにおける適切なリングを割り当てた際のカバー率を算出し、提案手法の実現可能性を評価する。

2. 提案手法

本稿では、遺伝的アルゴリズムを用いることで、地理的に分散した多様な自律システム (AS) を包括したリングを複数作成し、各コンテンツに適切なリングを割り当てるアルゴリズムを提案する。リングとは、地理的に分散した CS 群が一体となって構築される Anycast グループを指す。空間的に偏りがある AS を含めたリング集合を遺伝子とみなし、ランダムに生成された初期集団から開始し、各遺伝子に対して適応度を計算する。ここでの適応度は、収集されたコンテンツにおいてそれぞれ適切なリングを割り当てた際の平均需要カバー率として定義する。ただし需要カバー率は、コンテンツに割り当てられたリング内の AS の国とコンテンツのトラフィック上位国との一致度に対して各コンテンツの上位 5ヶ国の需要割合の和が 1 になるよう正規化したものである。一致する国ごとに、その上位国のトラフィックシェア率を合計することで、需要カバー率を算出する。すなわちリング作成数を N 、リング内に含まれる国数を k 、コンテンツ数を M 、コンテンツ m の国 i における需要割合を $T_{m,i}$ 、コンテンツ m の需要の上位 k 個の国の集合を $B_{m,k}$ 、コンテンツ m の需要が i 番目の国を $r_{m,i}$ とし、遺伝子 g における、コンテンツ m に割り当てられたリングの国集合を $R_m(g)$ とすると、遺伝子 g におけるコンテンツ m の需要カバー率 $C_m(g)$ と適応度 $A(g)$ は次式で定義される。

$$C_m(g) = \frac{\sum_{i \in (R_m(g) \cap B_{m,k})} T_{m,i}}{\sum_{j=1}^5 T_{m,r_{m,j}}} \quad (1)$$

$$A(g) = \frac{1}{M} \sum_{m=1}^M C_m(g) \quad (2)$$

以下に GA によるリング構成アルゴリズムを示す。

- 与えられた N と k に対して、ランダムに l 個の初期遺伝子を作成
- 適応度 i 位の遺伝子を選択する確率 p_i を各遺伝子の適応度に基づき $p_i = a \times b^{(i-1)}$ で計算 (a と b は $0 < a < 1$, $0 < b < 1$ を満たす定数で指数関数の確率分布のパラメータ)
- 選択確率を考慮し、確率的トーナメント法で次世代に残す遺伝子を x 個選択
- 選択された遺伝子に基づいて一様交叉処理を実行し、 y 個の子孫遺伝子を生成
- 一定の確率で突然変異を発生させ、遺伝子をランダムに再構成
- 上記の手順を G 世代繰り返し、最後の世代で最も適応度が高い遺伝子を最適リング案として選択

3. 性能評価

提案方式の有効性を検証するために計算機シミュレーションを用いた性能評価を行う。ただし、面積の大きな国に対して 1 つのエリアで CS 設置場所を考えると、需要カバー率の誤差が大きくなる。そのため面積が 5,000 万 km^2 以上の 6 か国に関しては、各国を 8 つのエリアに分割し、各エリアの人口比を基にして、コンテンツのトラフィック上位国とリング内の AS の国との一致時に、リング内の AS の緯度経度を元に該当エリアを特定する。そして、このエリアの人口比とトラフィックシェアを掛け合わせることで、エリア内の需要カバー率を計算する。世代数 $G = 10$ 、初期遺伝子数 $l = 300$ に設定し、リングの需要カバー率 $C_{N,k}$ を測定する。また CAIDA [3] の公開データセットから、AS rank に基づき降順に並べた AS の国名や座標に関する情報を取得し、上位 5 か国で需要比率が 9 割以上を占める Webpage の国ごとの需要比率のデータ [4] を用いた。上位 5ヶ国でトラフィックの 90% 以上を占めることから、これらの国々に対し CS を配置したリングを構成することで、大部分のユーザに対する効率的なコンテンツ配信が可能となる。

図 1 に各リング内の構成 CS 数 k に対して作成リング数 N を変化した場合の需要カバー率と、 N を 20 に固定した状態で、異なる k の値で生成されるリング集合に基づく各コンテンツの需要カバー率の累積分布関数 (CDF) を示す。ただしコンテンツ数を $n = 111$ とする。 $k=5$ の場合が最も低い需要カバー率を示しており、 k の値が増加するにつれて需要カバー率が向上していることが確認できる。特に $k=5$ から $k=20$ までの間で顕著な向上が見られた。しかしながら $k=25$ 以降は需要カバー率の向上があまり見られず、一定の CS をリングに含めることで需要カバー率は収束すると考えられる。また N の増加に伴い、需要カバー率は徐々に向上することが確認できた。しかし N が 30 を超えると需要カバー率の向上率は減少し、ほぼ一定となった。また N が増加するごとに、各 k の値における需要カバー率の CDF が上昇する傾向にある。

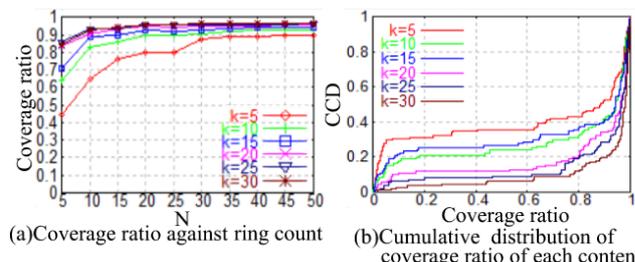


図 1: Cover ratio against number of rings

これらの結果から、高需要カバー率のコンテンツの累積分布は k の値が大きくなるにつれて増加していることが確認できる。これは k の増加がコンテンツの需要カバー率の向上に寄与していることを示している。また、リング作成数 N が増加することで全体的な需要カバー率の向上が見られたため、適切なパラメータの選定を行うことで提案手法が効果的な需要カバー率を実現することを確認できる。

謝辞 本研究成果は JSPS 科研費 21H03437 と 21H03437 の助成を受けたものである。ここに記して謝意を表す。

参考文献

- [1] Homas Koch, Ethan Katz-Bassett, John Heidemann, Matt Calder, Calvin Ardi, and Ke Li., Anycast in Context: A Tale of Two Systems, ACM SIGCOMM 2021
- [2] Zhihao Li, Dave Levin, Neil Spring, and Bobby Bhattacharjee, Internet anycast: performance, problems, & potential, ACM SIGCOMM 2018
- [3] CAIDA web page, <http://www.caida.org/data>
- [4] Similar web web page, <https://www.similarweb.com>
- [5] 加藤, 上山, Anycast CDN の配信サーバ群構成法の基礎検討, 信学会ソ大会 2023