

令和5年度 春学期 卒業研究3 (BI)
学士論文

題目 遺伝的アルゴリズムを用いた
AnycastCDNの配信サーバ選択法

指導教員 上山憲昭 教授

立命館大学 情報理工学部 セキュリティ・ネットワークコース

学籍番号 26002001054

加藤千尋

令和6年1月31日

概要

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、コンテンツの人気の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。そこで本稿で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 配信手法を提案する。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アルゴリズムを提案する。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指す。遺伝的アルゴリズムを利用して、空間的に偏った AS を取り入れた CS セットを構築し、計算機シミュレーションによりその効果を検証する。数値評価の結果、AS 数が 15 程度で高人気コンテンツの約 80% のトラフィックを効果的にカバーできることを示す。提案手法が少数の CS セットで大部分のコンテンツに対する要求を効果的にカバーでき、ユーザの体感品質の向上が期待される。

目次

概要	1
第1章 序論	3
1.1 研究の背景	3
1.2 研究の目的	4
第2章 関連研究	5
2.1 Anycast を用いた配信	5
2.2 Anycast グループのサイト数	5
第3章 提案方式	7
3.1 提案方式の概要	7
3.2 基本リング構成法	8
3.3 GA を用いたリング構成法	8
3.4 各コンテンツのリング割り当て法	10
第4章 性能評価	12
4.1 評価条件	12
4.2 需要カバー率	13
4.3 需要カバー率の累積分布	15
第5章 まとめ	17
謝辞	18

第1章 序論

1.1 研究の背景

Content Delivery Network (CDN) はインターネット上でコンテンツを効率的に配信する技術として広く普及している。CDN は世界中の地理的に分散した場所にキャッシュサーバ (CS) を設置し、要求に対してユーザに近い CS からコンテンツを配信することで、配信遅延時間の短縮やネットワーク内のトラフィック量の削減を実現する。CDN において優れたパフォーマンスを実現するには、配信要求に対して適切な CS を選択することが重要であり、最も一般的なメカニズムは DNS と Anycast である。DNS-based CDN では通常のホストのように、各 CS に各々異なる IP アドレスを付与し、各配信要求に対して CDN 事業者の DNS サーバが適切と思われる CS を選択して、その IP アドレスを応答することで CS 選択を実現する。CDN 事業者が CS を選択できるメリットがあるが、運用に大きな投資を必要とする [3]。そのため、新興 CDN の多くは DNS-based CDN に比べて安価で展開できる Anycast-based CDN を利用している。

Anycast-based CDN では同じ IP アドレスを複数の CS に対して付与し、インターネットの Anycast ルーティングプロトコルに基づいてルータが最も近いと推定される CS にパケットを転送する。Anycast を用いることで、ユーザと CS 間の転送時間を短縮することを目指している。しかし配信先が Border Gateway Protocol (BGP) のルーティングポリシーに大きく依存しており [14]、最適な CS に常にパケットが転送されるわけではない。事実、一部の配信要求は最も適切とされる CS にルーティングされない可能性が指摘されている [3]。さらに CS セットを構成する CS 数をただ増加するだけでは、ルーティングの適切さが必ずしも向上しないことも明らかになっている [4]。CS セット構成は地理的に分散した CS 群で構成される Anycast グループであり、Microsoft CDN など採用されている [1]。Microsoft CDN の場合、各国の法規上、各ユーザがアクセスできるコンテンツを限定する必要があるため、ユーザごとにアクセス可能なリングと呼ぶ CS セットを限定し¹、配信要求ユーザの拠点に配信可能なリングの中から、構成 CS 数が最大のリングを選択する方式が用いられている。また一方で、既存の Anycast CDN では、コンテンツの人気に地域性があるにも関わらず、すべてのコンテンツ要求に対して同一の CS セットを回答するという課題もある。つまり、ある地域で特定のコンテンツが非常に人気である場合、そのコンテンツはその地域の CS にキャッシュされるべきだが、Anycast ではこれを適切に行うのが難しい。

¹本稿でも以後、CS セットのことをリングと呼ぶ。

1.2 研究の目的

Anycast-based CDN はユーザと CS 間の転送時間を短縮することを目指している。しかし、既存の Anycast CDN は、コンテンツの人気の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 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。

そこで本稿では、コンテンツの需要の空間的局所性に基づき、Anycast 配信をより効果的に実現する新しい方法を提案する。具体的には、コンテンツの地域ごとの人気の偏りを考慮して、特定のリングを作成するアルゴリズムを構築する。作成したリングが各コンテンツに適切に割り当てられることで、ユーザのコンテンツ配信遅延の低減を目指す。さらに提案手法は CS 数を最小限に抑えることを特徴としている。これにより、広範囲の国に渡るユーザに対しても、遅延時間の分散が小さく安定した高品質なサービスを提供することが可能となる。さらにリングの構成に要する時間を低減するため、遺伝的アルゴリズムを用いたリング集合の生成アルゴリズムを提案する。本アルゴリズムでは、ランダムに生成したリング集合を各遺伝子、構成リング集合による平均需要カバー率を適応度として定義し、空間的に偏りがある AS を含めたリング集合を作成する。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。

第2章 関連研究

2.1 Anycast を用いた配信

Anycast 配信は、インターネット上でのコンテンツ配信品質を向上させる技術として注目されている。Anycast 配信は同じ IP アドレスを異なる地点に存在する複数のサーバに割り当て、ユーザの要求を「最も近い」と推定されるサーバへルーティングすることで、オンラインアプリケーションの性能向上や、ユーザデバイスとデータセンタ間の転送時間の短縮を実現する [5]。また、Anycast は高い接続の安定性を示しており、信頼性が高く、拡張性と管理の容易さを兼ね備えている [6]。

しかし Anycast 配信にはいくつかの課題が存在する。最も顕著なのは BGP ルーティングの制約により、Anycast の振る舞いや性能を細かくコントロールすることが難しい点である [7]。特にインターネットのルーティングは動的であり、その変動が Anycast の性能に与える影響は大きい。この問題を検証した Calder らの研究によると、多くの場面で Anycast は十分に良好な性能を示す一方、約 20% の要求に対し、最適とは言えないサーバ選択されたことが確認されている [3]。この原因の一つとして、BGP ポリシーでは ISP がピアルートよりも顧客ルートを優先することが指摘されている [9]。

2.2 Anycast グループのサイト数

Anycast 配信では一部の要求が適切なキャッシュにルーティングされないことが懸念されている。この問題は Anycast グループの CS 数を増やすだけでは、CS 選択時の適正度が必ずしも向上するわけではない [4]。Li らは Anycast において global sites が追加されるにつれて Anycast がどのようにパフォーマンスを向上させるかを測定しており、測定期間中に 77 サイトを追加した root-server のパフォーマンスが顕著に向上しないことを明らかにしている [4]。さらに Anycast は負荷の不均衡や不必要に長い距離を移動するクエリが存在することが明らかとなっており、Anycast において各グループの CS 数の増加によって、Anycast でのルート選択は利用可能な複数のサイトへのパスに影響され、これがパフォーマンスへのペナルティとなっていると結論づけられている。Schmidt らの研究でも CS 数の増加に伴いミスヒット率が上昇することが示されており、同様の原因が考えられる [11]。また Schmidt らは Vantage Points (VP) を活用して各 root-DNS server の遅延時間に関する調査を行っている。VP はネットワーク内の特定の観測地点を指す用語であり、その地点からのアクセスを通じてネットワークの性能や動作を評価することが可能になる。測定の結果としては、RTT の中央値は CS 数の変動による大きな影響を受けないことを示している。これは多数の CS を有する root server が特定の VP の改善を目的としてサーバ数を増やしているためと考えられる。しかし改善効果は一様ではなく、特定の国々で遅延の

悪化が確認される事例も存在し，少数の CS を持つ root server が多様な国で均一かつ安定したパフォーマンスを維持することを示している．この調査から，ルーティングポリシーが CS 数の増加による利益を制限している可能性が示唆され，少数の CS を持つリングが現時点での最適解であると結論づけている．

第3章 提案方式

3.1 提案方式の概要

空間的な人気の偏りを考慮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。これは、地域ごとのコンテンツの需要を反映したキャッシュ利用によってコンテンツの人気の空間的局所性を考慮した Anycast 配信の実現を目的としている。

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。Microsoft CDN では、各ユーザの国の法規上でアクセスできるコンテンツを限定する必要があるため、各ユーザに対して、アクセス可能なリングの候補を限定し、構成 CS の数が最も多いリングが選択される。

本稿では、現実的な時間で複数のリングを生成するための手法として遺伝的アルゴリズムを用いることで、図 3.1 に示すような地理的に分散した多様な自律システム (AS) を包括したリングの生成を目指している。コンテンツの高人気地域をカバーできるような少数のリングを作成することで、ルーティングテーブルのサイズを最小限に抑制し、Anycast 配信の効率化を図る。また、AS は地理的な位置だけでなく、ネットワークポリシーも反映しているため、実際のネットワークを想定した適応性の高いコンテンツ配信の実現に寄与する。

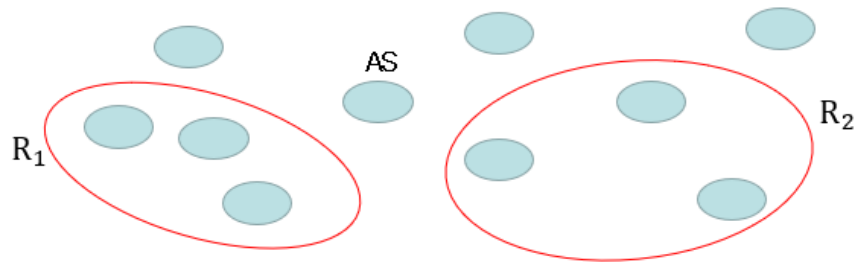


図 3.1: Examples of ring configurations

遺伝的アルゴリズムの利点を活かすことで、コンテンツの人気度に基づく空間的局所性を考慮したリング構成を可能にし、効率的な配信経路の最適化を図る。また、リングを構成する CS 数を少数に抑えることで広範囲の国にわたって遅延時間に分散が小さく安定したパフォーマンスの提供を図り、さらに、リング数を少数に抑えることでルーティングテーブルのサイズの抑制を図る。

3.2 基本リング構成法

各コンテンツ m の需要比率上位 k 国が含まれる集合を $B_{m,k}$ と定義する。本節では、まず、直感的なリング構成法として、生成されたリングがカバーする需要量が最大となるよう貪欲算法的にリングを構成する方法を提案する。提案アルゴリズムを以下に示す。

1. コンテンツ m を要求数の降順にソート
2. 最上位のコンテンツ m から順に、 B_m からトラフィックシェア k 位までの国を抽出
3. 抽出した国集合が、既に作成したリングと重複しない場合は新規リングとしてリングを作成
4. Step2 と Step3 を N 個のリングが作成されるまで反復。

本アルゴリズムは実装が容易であるが、リングの作成結果は収集するコンテンツのデータに依存する。そのため人気の高いコンテンツに対して特定の地域や国が含まれる傾向が強まり、その結果、リング構成の多様性が損なわれる可能性が高い。また、基本方式では国単位でリングを構成しているため、コンテンツに割り当てられたリングと一致する国はその国全てのトラフィックをカバーできるという前提をもとに評価を行っている。この前提は、各国内のトラフィック分布の不均一性を考慮していないため、実際のトラフィックパターンの多様性を適切に反映しきれないという課題がある。

3.3 GA を用いたリング構成法

基本リング構成法の問題を解決するため、本節では遺伝的アルゴリズムを用いたリング構成法を提案する。遺伝的アルゴリズム (GA) は自然の進化過程を模倣した計算手法として知られる。この手法は、生物の遺伝や選択、交叉、突然変異といった過程を計算上で模倣することで、多様な解の探索を可能とする。特に、多くの最適化問題や近似解探索において遺伝的アルゴリズムは広く用いられている。本研究では、コンテンツの人气が一定期間内で状況が変化しないという前提を置いている。この前提に基づき、逐次的な状況変化に対応する必要性が低いため、最適な解を見つけるための過程が比較的安定している。このような安定した問題設定において、近似解法として遺伝的アルゴリズムを選択することが適切であると判断した。

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.

$$C_m(g) = \frac{\sum_{i \in (R_m(g) \cap B_{m,k})} T_{m,i}}{\sum_{j=1}^5 T_{m,r_{m,j}}} A(g) = \frac{1}{M} \sum_{m=1}^M C_m(g) \quad (3.1)$$

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 カ国に関しては、図 3.2 に示すような 8 つのエリアに分割する操作を行う.

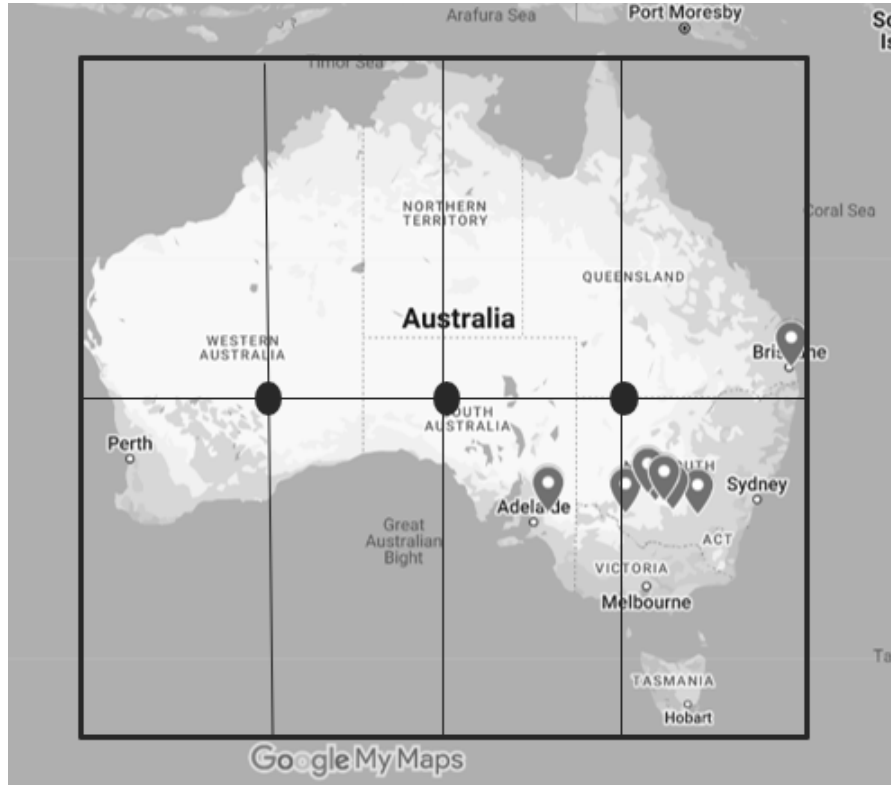


図 3.2: Examples of ring configurations

各国の最東端及び最西端の経度、および最南端及び最北端の緯度を用いて長方形の領域を定義し、長方形を上下に二分割する水平線を引く. さらに、二つの各長方形を四分分割することで、各国を 8 つのエリアに分割する. 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. エリア j の人口比を P_j とすると、分割された国の需要カバー率 C'_m を次式で定義する.

$$C'_m = \sum_{i \in (R_m \cap B_{m,k}) \text{ and } i \in j} P_j \times T_{m,i} \quad (3.2)$$

前提条件として、各コンテンツの国別需要比率上位 5 カ国の国名と、各国の需要比率が既知とする。要求発生数の上位 5 ヶ国以内で要求の 90%以上を占めるコンテンツ n 個に対し、提案アルゴリズムを用いてリングを作成する。GA によるリング構成アルゴリズムを、以下に示す。

1. 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
2. 各遺伝子の適応度を計算
3. 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タ)
4. 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
5. 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
6. 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
7. 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

本アルゴリズムで対象とするコンテンツは、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。

3.4 各コンテンツのリング割り当て法

遺伝的アルゴリズムによって構築されたリングの中から、各コンテンツに適切にリングを割り当てる際に、各コンテンツの需要比率の上位国の需要比率のデータを用いる。コンテンツ m に対して、リングの中に存在する AS の国が、コンテンツ m の上位国集合 $B_{m,k}$ に存在する場合、当該リングはその国のトラフィックをカバーするとみなす。また面積が 5,000 万 km^2 以上の 6 カ国については、分割した 8 つのエリアにおいて該当するエリアの需要比率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。以上の手順で、各コンテンツに対して最大限のトラフィックをカバーするリングを選択し、コンテンツごとに最適なリングを割り当てる。

しかし、 $B_{m,k}$ 内のすべての国がどのリングにも含まれていない状況においても、適切なリングの割り当ては必要となる。このような場合、まず、各コンテンツのトラフィック上位国と、既に構築されたリングに存在するすべての国との間で、地理的な距離 D を次式で計算する。ただし ϕ_1 と ϕ_2 は各々、2 つの地点間の緯度のラジアン値、 $\Delta\phi$ は 2 点間の緯度の差、 $\Delta\lambda$ は緯度の差のラジアン値、 R は地球の半径を示す。

$$d = 2R \times \arcsin \left(\sqrt{\sin^2 \left(\frac{\Delta\phi}{2} \right) + \cos(\phi_1) \times \cos(\phi_2) \times \sin^2 \left(\frac{\Delta\lambda}{2} \right)} \right) \quad (3.3)$$

上記で計算した地理的距離の中で最も短いものを持つ国を特定し、その国が属するリングを選択する。また、この選択プロセスでは需要比率がより上位国を重点的に考慮し、地理的に最も近接する国が含まれるリングを選択する。この方法を採用することにより、上位国集合 $B_{m,k}$ がすでに作成されたリングの国々と完全に一致しない状況でも最も適切なリングが割り当てられる。その結果、配信距離の不必要な増加を防ぐことができ、通信の遅延の増大を効果的に抑制することが可能となる。

第4章 性能評価

4.1 評価条件

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。この評価は、特に遺伝的アルゴリズムを使用してリングを作成する際の効果に焦点を当てて実施する。世代数 $G = 10$ 、初期遺伝子数 $l = 300$ に設定する。少数のリングを作成し、各コンテンツに最適なリングの割り当てを行うため、各コンテンツのトラフィック上位国が適切なリングによって需要をカバーできるかの結果が懸念される。そのため、3.2 節で述べた、上位コンテンツのトラフィック上位国 k 位までを各リングの構成要素とする方式を「基本方式」と呼ぶ。この基本方式と遺伝的アルゴリズムの方式の各々において、リングの需要カバー率 $C_{N,k}$ を比較する。

CAIDA[18] 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て用いる。また Similar web のサイト [19] から得られる、上位 5 개국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を用いた。

CAIDA の公開データセットから収集するデータは、Corn サイズにおける上位の AS トポロジ 300 個に加えて、コンテンツの需要比率上位にランクインしている国の AS が 1 つ以上存在するように追加で 32 個の AS を収集した。これはコンテンツのランキングに含まれる国の中に、AS rank 降順に収集した AS トポロジには含まれていない国が存在することによって、需要カバー率の減少に影響を及ぼすためである。具体的には、AS rank 300 個を収集し、シミュレーションを行ったところ収集したコンテンツが 5ヶ国でトラフィックシェアを 9 割超えるものの、リングを作成した場合の需要カバー率が 8 割近くで収束する傾向にあった。AS の追加を行うことによって、あるコンテンツの需要比率上位にランクインしている国が収集した AS の中に存在しないため、どのリングを割り当ててもその国のトラフィックを完全にカバーすることができないという状況を防ぐことを図る。また、本稿の目的は配信要求を最も効率的にカバーするリング構成を明らかにすることにあるため、今回の評価では特に大規模な AS に焦点を当て、重点的に扱うこととした。これにより、需要の多くを効率的にカバーするリングを構成可能となる。図 4.1 は収集した AS の分布を示しており、選択した AS は地球上のさまざまな地域に分散している。

また、要求発生数の上位 5ヶ国において、要求の 90%以上を占めるコンテンツを 111 個収集した。これらのコンテンツは、特定の地域で非常に高い人気を持っていることが示唆される。これらの特定地域での高い人気を持つコンテンツを中心に収集・分析することで、各地域のユーザのニーズや要求パターンの違いを深く理解し、それを基にしたリングの最適化を試みることが可能となる。

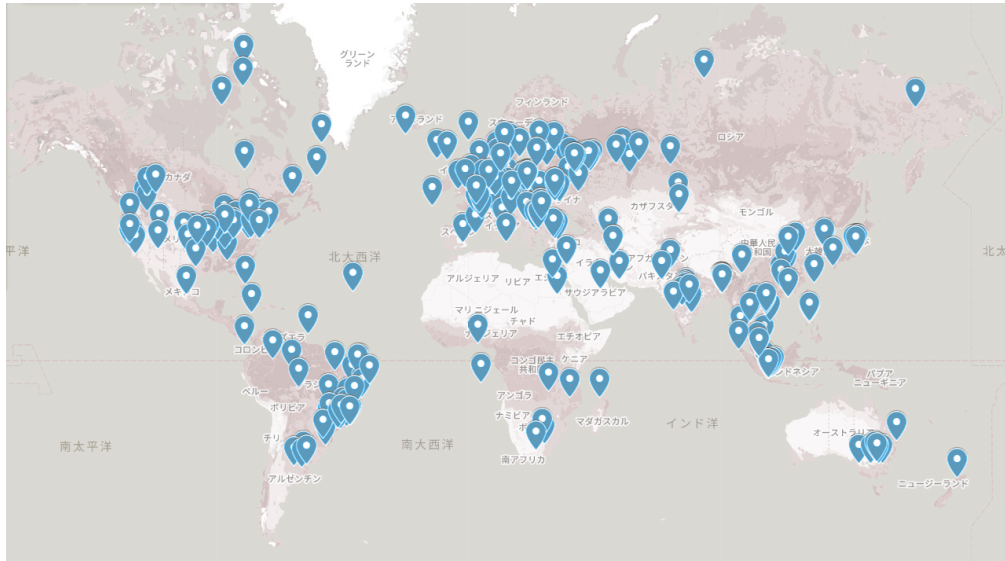


図 4.1: Graphical distributions of ASes used in numerical evaluation

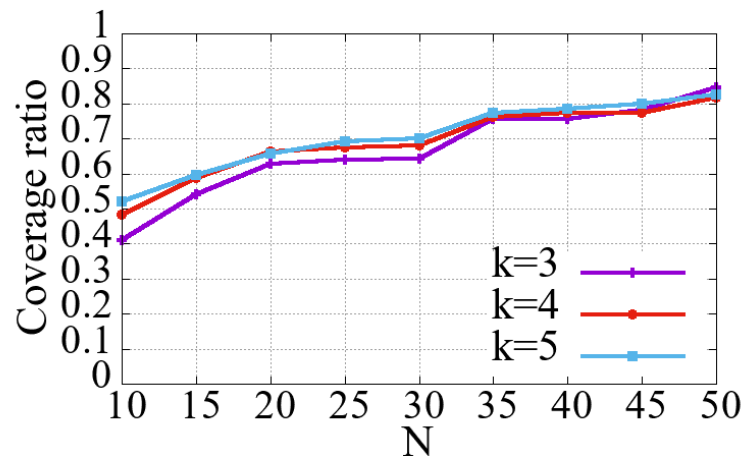
4.2 需要カバー率

図 4.2 に基本方式と提案方式を各々用いたときの，各リング内の構成 CS 数 k に対して作成リング数 N を変化させた場合の需要カバー率をプロットする．需要カバー率 $C_m(g)$ の定義を以下に示す．

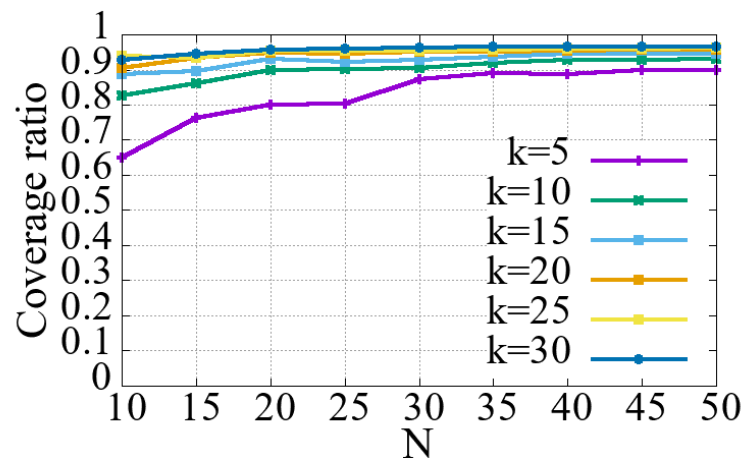
$$C_m(g) = \frac{\sum_{i \in (R_m(g) \cap B_{m,k})} T_{m,i}}{\sum_{j=1}^5 T_{m,r_{m,j}}} \quad (4.1)$$

両手法ともコンテンツ数を $n = 112$ とし，基本方式ではコンテンツデータに含まれる国を基にリングを構成するため k を最大 5 と制約を設けている．一方で遺伝的アルゴリズムを用いた手法では，リングはリング内構成 AS 数 k を最小 5，最大 40 とし， N を最小 5，最大 50 までの範囲で設定した． $k=5$ 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，基本方式と比較すると，より少ない N で高い需要カバー率を達成可能であることがわかる．この結果から，遺伝的アルゴリズムがリング内の要素の多様性を高め，トラフィックの地理的分布をより広くカバーすることが可能であることが確認できる．これは k の値が小さい場合でも効果的なリング構成を生成する能力を遺伝的アルゴリズムが有していることを意味する．



(a) Ring creation algorithm
using traffic country



(b) Ring creation using
genetic algorithm

図 4.2: Cover ratio against number of rings

4.3 需要カバー率の累積分布

さらにリングの数 N と需要カバー率の関係に着目し、パラメータ k の違いが各コンテンツの需要カバー率に与える影響を評価する．図 4.3 に、 N の値を 10, 15, 20 に固定した状態で、異なるパラメータ k の値によって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． 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．これはリング数の増加が需要カバー率の向上に直接的に寄与しており、より多くのコンテンツがカバーされる可能性が高まることを示唆している．

また、すべての N の値において、パラメータ k の増加が需要カバー率の向上に効果的であることが明らかになった．具体的には、 $k = 5$ から $k = 20$ までの増加による需要カバー率の向上は特に顕著で、それ以上の k の値では需要カバー率の向上が鈍化する傾向が見られた．これは、図 4.2 で観測された結果とも一致しており、 $k = 25$ 以降ではリングの追加による需要カバー率の向上効果が減少する．さらに N が 10 および 15 の場合において、 $k = 20$ までの需要カバー率の向上が特に著しく、それを超えると需要カバー率の向上が鈍化することが明らかとなった．一方で、 $N = 20$ の場合には、 $k = 20$ において需要カバー率が最大となり、それ以上の k の増加は需要カバー率の向上は限定的である．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した．

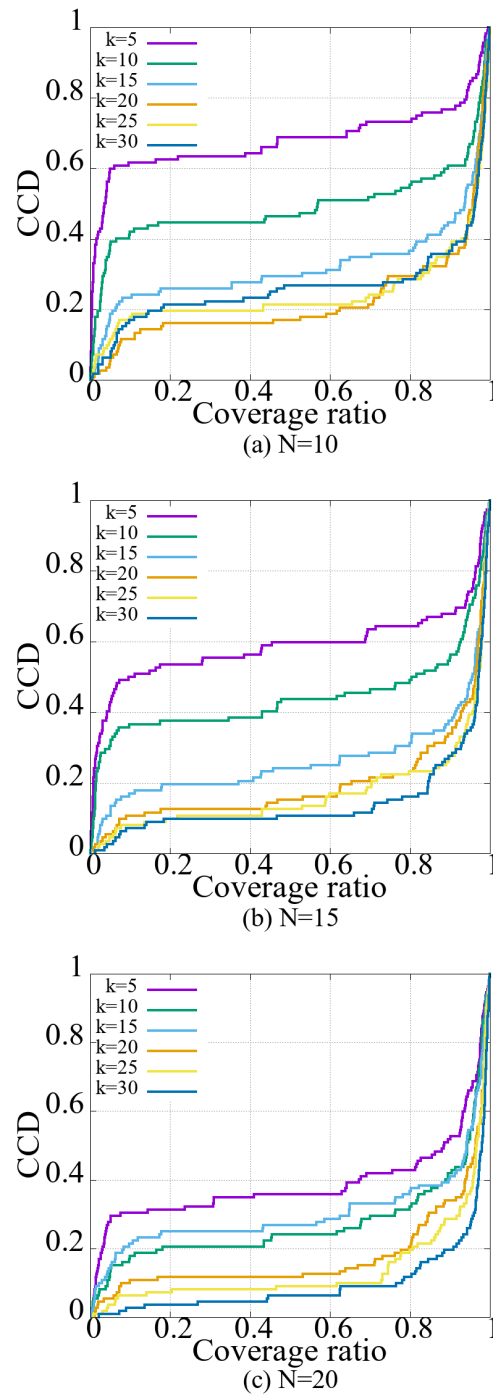


図 4.3: Cumulative distribution of coverage for each content

第5章 まとめ

トラフィック量の削減やユーザ品質の向上のため、地理的に分散した場所にキャッシュサーバを設置し、ユーザに近いキャッシュサーバからコンテンツを取得する CDN が広く普及している。CDN のパフォーマンスを最大限に引き出すためには、配信要求に対して適切なキャッシュサーバを選択することが重要であり、その手法の一つである Anycast-based CDN では、ルーティングプロトコルが BGP に依存しているため、一部の要求が適切なキャッシュにルーティングされず、また、Anycast 配信では CS 数増加に伴い CS 選択の精度が劣化することが指摘されている。つまり、コンテンツの人気の地域性に基づく少数 CS によるリング構成が有効であるとされる。そこで本論では、遺伝的アルゴリズムを用いたコンテンツの人気の空間的局所性を考慮したリングの作成手法を提案した。CAIDA のデータセットを基にした AS の所在地情報や、SimilarWeb を利用した地域性の強いサイトの国別需要比率のデータを用いた数値評価の結果、AS 数 15 程度で高人気コンテンツの約 70%～80% の需要をカバーできることが示された。提案手法を用いることで少数のリング数と CS 数でコンテンツの大部分をカバーすることができ、ユーザ品質の向上が期待できる。今後の研究においては、さらに高い需要カバー率を目指すためのリングの最適化手法の改善、さらなるデータ収集の検討を行う予定である。

謝辞

本研究を行うに当たり，ご指導を頂いた上山教授に感謝します．また日常，有益な議論をして頂いた研究室の皆様に感謝します．

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] Hussein A. Alzoubi, Seungjoon Lee, Michael Rabinovich, Oliver Spatscheck, and Jacobus Van der Merwe., Anycast CDNs Revisited. WWW 2008
- [3] Matt Calder, Ashley Flavel, Ethan Katz-Bassett, Ratul Mahajan, and Jitendra Padhye, Analyzing the Performance of an Anycast CDN, ACM IMC 2015
- [4] Zhihao Li, Dave Levin, Neil Spring, and Bobby Bhattacharjee, Internet anycast: performance, problems, & potential, ACM SIGCOMM 2018
- [5] Ashley Flavel, Pradeepkumar Mani, David A. Maltz, Nick Holt, Jie Liu, Yingying Chen, and Oleg Surmachev, FastRoute: A Scalable Load-Aware Anycast Routing Architecture for Modern CDNs, NSDI 2015
- [6] Ballani, H., and Francis, P, Towards a global IP Anycast service, ACM SIGCOMM 2005.
- [7] Q. Fu et al., Taming the Wild: A Scalable Anycast-Based CDN Architecture (T-SAC), IEEE Journal on Selected Areas in Communications, vol. 36, no. 12, pp. 2757-2774, Dec. 2018
- [8] Matt Calder, Manuel Schröder, Ryan Gao, Ryan Stewart, Jitendra Padhye, Ratul Mahajan, Ganesh Ananthanarayanan, and Ethan Katz-Bassett, Odin: microsoft's scalable fault-tolerant CDN measurement system, NSDI 2018
- [9] Minyuan Zhou, Xiao Zhang, Shuai Hao, Xiaowei Yang, Jiaqi Zheng, Guihai Chen, and Wanchun Dou, Regional IP Anycast: Deployments, Performance, and Potentials, ACM SIGCOMM 2023
- [10] Zhang, X., T. Sen, Z. Zhang, T. April, B. Chandrasekaran, D. Choffnes, B. M. Maggs, H. Shen, R. K. Sitaraman, and X. Yang., AnyOpt: Predicting and Optimizing IP Anycast Performance, ACM SIGCOMM 2021.
- [11] R. d. O. Schmidt, J. Heidemann, and J. H. Kuipers., Anycast latency: How many sites are enough?, PAM 2017
- [12] Hitesh Ballani, Paul Francis, and Sylvia Ratnasamy, A measurement-based deployment proposal for IP anycast, IMC 2006
- [13] Dina Katabi and John Wroclawski., A framework for scalable global IP-anycast (GIA). ACM SIGCOMM 2000
- [14] Rekhter, Yakov and Tony Li., A Border Gateway Protocol 4 (BGP-4). RFC 4271 (1994): 1-104
- [15] Anders Brodersen, Salvatore Scellato, and Mirjam Wattenhofer., YouTube Around the World: Geographic Popularity of Videos, WWW 2012
- [16] Zakaria Al-Qudah, Seungjoon Lee, Michael Rabinovich, Oliver Spatscheck, and Jacobus Van der Merwe. 2009. Anycast-aware transport for content delivery networks. WWW 2009
- [17] K. S. Tang, K. F. Man, S. Kwong and Q. He., Genetic algorithms and their applications, IEEE Signal Processing Magazine, vol. 13, no. 6, pp. 22-37, Nov. 1996
- [18] CAIDA web page, <http://www.caida.org/data>
- [19] Similar web web page, <https://www.similarweb.com>
- [20] Oak Ridge Natunal Laboratory, <https://landscan.ornl.gov/>

学会発表リスト

- 加藤千尋, 上山憲昭, “Anycast CDN の配信サーバ群構成法の基礎検討“, 電子情報通信学会 2023 年ソサイエティ大会, BS-4-7, 名古屋, 2023 年 9 月
- 加藤千尋, 上山憲昭, “遺伝的アルゴリズムを用いた Anycast CDN の配信サーバ選択法“ 電子情報通信学会インターネットアーキテクチャ (IA) 研究会, 沖縄, 2024 年 3 月
- 加藤千尋, 上山憲昭, “ 遺伝的アルゴリズムを用いた Anycast CDN の配信サーバ選択法 “ 電子情報通信学会 2024 年総合大会, 広島, 2024 年 3 月
- Chihiro Kato and Noriaki Kamiyama, “ Designing Server Sets for Anycast CDN Using Genetic Algorithm, ” accepted at IEEE ICC 2024, NGNI Symposium, Denver, June 2024