Personalized Caching Model for Sustainable Content Delivery Networks

Chinmay Tawde, *Member, IEEE,* Kamiyama Noriayaki, *Member, IEEE*

***Abstract*—The exponential growth of global internet traffic has positioned Content Delivery Networks (CDNs) as critical infrastructure for performance, but also as significant contributors to energy consumption and carbon emissions within the ICT sector. While traditional caching strategies focus on latency and hit ratio, their environmental impact remains largely unquantified. This paper presents a comprehensive simulation-based analysis comparing four caching techniques and later on comparing two of the best caching techniques (Predictive and Cooperative) among the four . The paper helps simulate Predictive and Cooperative caching strategies against traditional methods, evaluating them through a dual lens of performance and sustainability. Our methodology models real-world network topologies from the Internet Topology Zoo under diverse operational scenarios—ranging from low demand event , normal demand event, high demand event and special demand event using a dynamic Zipf-driven request model. We introduce a detailed energy consumption model that accounts for both device-level power draw and traffic-dependent link energy, translating results into carbon emissions and operational costs. Our findings demonstrate that intelligent caching strategies yield substantial sustainability gains: Predictive caching reduces energy consumption by up to 38% in high-demand volatile environments by proactively pre-fetching content, while Cooperative caching excels in stable scenarios by minimizing fetch distances through neighbor coordination. The research concludes that strategy selection is highly context-dependent and provides an AI-informed recommendation framework to guide network operators in optimizing for both performance and environmental impact. This work establishes a critical pathway for developing high-efficiency, low-carbon future networks.**

***Index Terms*—Content Delivery Networks (CDNs), Green Networking, Energy Efficiency, Predictive Caching, Cooperative Caching, Network Simulation, Carbon Footprint, Internet Topology Zoo**

I. Introduction

The dawn of the digital age has precipitated an unprecedented reliance on global data networks. From high-definition video streaming and immersive cloud gaming to the vast ecosystem of Internet of Things (IoT) devices and distributed cloud computing, the demand for digital content is growing at an exponential rate. Cisco's Annual Internet Report forecasts that global internet traffic will surpass 4.8 zettabytes per year by 2025, driven by an estimated 29.3 billion networked devices . This insatiable appetite for data places intense strain on the underlying infrastructure of the internet, particularly on Content Delivery Networks (CDNs), which are responsible for efficiently distributing content to end-users by caching it closer to the network edge. While the primary design goal of CDNs has historically been to maximize performance—minimizing latency, reducing bandwidth costs, and improving availability—this relentless pursuit of speed has come with a significant, and often overlooked, environmental cost. Although CDNs have achieved the desirable result of lowering the energy consumption compared to traditional methods used before its discovery, there are sectors within CDN that still contribute to major carbon footprint and eventually shed light on the negative side of modern CDNs . This paper accepts the sustainable achievements of CDN but also plans to help eliminate the underlining factors making it even more sustainable for the increasing demand and supply operations. The information and communication technology (ICT) sector is now a substantial contributor to global energy consumption and greenhouse gas emissions, with estimates suggesting it accounts for 1.4% to over 4% of global greenhouse gas (GHG) emissions—a share comparable to, or even exceeding, that of the aviation industry. Data centers and network infrastructure form the core of this energy demand. Consequently, the paradigm of network design is undergoing a critical shift. The challenge is no longer solely about delivering content faster; it is about delivering it smarter, with a conscious effort to minimize its overall carbon footprint , making it more sustainable for long time use. The quest for **sustainable networking** has thus emerged as a paramount objective for researchers , professors and engineers.

A primary lever for achieving sustainability in CDNs is the optimization of caching strategies. Caching, the process of storing copies of frequently accessed content on strategically placed cache servers, is the fundamental mechanism that reduces long-haul data transmission and improves user experience. Traditional caching algorithms, such as Least Recently Used (LRU) or First-In-First-Out (FIFO), operate on a reactive and localized basis. However, their simplicity often limits their efficiency in the face of dynamically changing content popularity and complex network conditions. This limitation presents a crucial research opportunity: can more sophisticated, intelligent caching strategies significantly enhance performance while simultaneously reducing the energy consumption of the network?

Two promising avenues have emerged in response to this question. First, **Predictive Caching** leverages patterns in user request traffic to proactively store content before it is demanded. By employing machine learning and statistical forecasting models, these systems aim to anticipate the future requests, thereby achieving higher cache hit ratios and preventing costly fetches from distant origin servers. Second, **Cooperative Caching** fosters collaboration between neighboring cache nodes within the network. Upon a local cache miss, a node can query its peers for the requested content, effectively creating a distributed, shared cache pool. This strategy can drastically reduce the distance content must travel, saving energy and reducing latency without requiring larger individual caches.

Despite their individual promise, a critical gap exists in the current literature. There is a lack of comprehensive, direct comparison of these advanced strategies under a unified framework that evaluates them not just on traditional performance metrics like hit ratio and latency, but crucially, on their sustainability impact—quantified through energy consumption, carbon emissions, and operational cost. Most existing studies evaluate performance and sustainability metrics separately, preventing network operators from understanding the trade-offs between efficiency, energy use, and emissions.Furthermore, the performance of these strategies is highly contextual, varying with network topology, traffic load, and content popularity dynamics, a nuance often absent in high-level analyses.

This research seeks to address this gap by posing the following central question: Under what network conditions do predictive and cooperative caching strategies provide the most optimal balance of high performance and environmental sustainability*?* To answer this, we develop a sophisticated, multi-faceted simulation framework that models real-world network topologies and a range of operational scenarios, from low, normal, high and special events. Within this framework, we implement and rigorously test both strategies, measuring their efficacy across a dual set of metrics: performance (hit ratio, latency) and sustainability (energy consumption, carbon footprint, economic cost).

The contributions of this work are threefold:

1) **A Realistic Simulation Environment:** We develop a flexible simulator that models complex network dynamics, including device-level energy profiles, link congestion, and realistic, shifting content request patterns based on a dynamic Zipf distribution.
2) **A Holistic Evaluation Framework:** We move beyond traditional metrics to provide a comprehensive sustainability study, translating energy usage into tangible environmental (kg of CO_2) and economic (¥) impacts.
3) **Actionable Insights:** We present an AI-informed analysis that provides clear, scenario-based recommendations for network operators, guiding them on which caching strategy to deploy to meet specific performance and sustainability goals.

By bridging the fields of network engineering and environmental science, this research aims to provide a evidence-based pathway for building the high-performance and highly sustainable communication networks of the future.

II. Related work

The optimization of CDNs through advanced caching strategies is a well-established field of research. Our work builds upon and intersects with several key areas: traditional caching algorithms, energy-aware networking, and emerging intelligent caching patterns. This section surveys the relevant literature and identifies the gap our research aims to fill.

A. *Traditional and Fundamental Caching Strategies*

The foundation of caching logic is built upon simple yet effective policies designed to manage limited storage capacity. The most commonly used among these is the **Least Recently Used (LRU)** algorithm, which evicts the item that has been unused for the longest time. Its efficiency and low computational overhead have made it a standard for decades. Variants like Least Frequently Used (LFU) and First-In-First-Out (FIFO) offer different trade-offs between recency and frequency of access. While these algorithms are effective at a local level, their reactive nature and lack of global coordination limit their overall efficiency in complex, large-scale networks. Our work uses LRU as a caching substrate to ensure a fair and baseline comparison for our more advanced strategies. As we achieve our desired results , we will introduce complex and vivid algorithms to make the model more adaptive to the input data.

B. *Cooperative and Collaborative Caching*

To overcome the limitations of isolated caches, researchers have come up with the concept of **Cooperative caching**, where nodes in a network share their cached contents to serve each other's misses, was pioneered in distributed system environments and later adapted for web caching and CDNs. The core challenge in this domain is efficiently locating content within the cooperative group without introducing excessive overhead communication . Studies often focus on protocols for querying neighbors, cache coherence, and minimizing latency within the cooperative sphere. Our implementation contributes to this field by integrating a practical, overhead-aware cooperative mechanism and evaluating its efficacy not just on latency, but crucially, on its energy-saving potential across diverse network scenarios.

C. *Proactive and Predictive Caching*

With the rise of machine learning, the paradigm has shifted from reactive to proactive caching. **Predictive caching** leverages patterns in user behavior and content popularity to prefetch and store content before it is requested. Early approaches used simple time-series analysis, while modern implementations employ sophisticated models including Markov chains, reinforcement learning, and deep learning to forecast future requests. These strategies have shown remarkable success in improving cache hit ratios, especially in environments with predictable popularity trends. Our research adopts this forward-looking approach by implementing a hybrid prediction model that combines both frequency-based and Markovian analysis, dynamically adapting to the volatility of the scenario.

D. *Energy-Aware and Green Networking*

The growing concern over the ICT sector's carbon footprint has spurred the field of **green networking**. Research here focuses on optimizing energy consumption across data centers, network hardware, and protocols. Strategies include dynamic voltage and frequency scaling (DVFS), putting network components into low-power sleep states during periods of low activity, and energy-aware routing protocols that select paths based on power efficiency alongside traditional metrics. While these works provide crucial hardware and lower-layer solutions, our research operates at the application and content layer, demonstrating how intelligent decisions on where and what to cache can yield significant energy savings at the network level.

E. Research Gap and Our Contribution:

Through our research , we have encountered three major issues that still persist in modern day CDNs.

1) Data Center and Server Inefficiency: The physical data centers and edge servers that power CDNs are massive energy consumers. A large portion of this energy (up to 40-50%) is used for **cooling systems** to prevent overheating. Additionally, servers consume a significant amount of power even when they are idle or underutilized, a phenomenon known as "ghost load."

2) Network and Transmission Inefficiency: : While CDNs reduce data travel distance, the underlying network infrastructure still consumes energy. This consumption is inefficient during off-peak hours when network links are underutilized but still fully powered. Additionally, suboptimal routing can sometimes cause data to take longer, less-efficient paths, wasting energy.

3) Dependence on Energy Grids: : The single biggest determinant of a CDN's carbon footprint is the source of its electricity. Even the most efficient data center will have a large carbon footprint if it is powered by a regional grid that relies heavily on fossil fuels.

A clear gap exists in the current landscape. While numerous studies excel in optimizing for either performance *or* energy efficiency, few provide a holistic comparison of modern strategies within a unified framework that measures both sets of metrics simultaneously. Many evaluations of predictive and cooperative schemes are isolated, and their environmental impact is often an afterthought rather than a primary evaluation criterion.

This work distinguishes itself by conducting a direct, empirical comparison of **Predictive** and **Cooperative** caching strategies under a common set of realistic conditions. Our primary contribution is a comprehensive evaluation framework that quantifies the trade-offs between traditional performance metrics (hit ratio, latency) and critical sustainability metrics (energy consumption, carbon emissions, operational cost). By providing this dual-perspective analysis across multiple network topologies and demand scenarios, we offer reliable insights for building high-performance and environmentally sustainable CDNs.

III. FORMULA

The model uses wide range of formulae. All the major ones are listed below. The first formula listed below is used to calculate total energy consumption where 'W' is a constant that denotes the link power for fiber optics. We have considered it as 0.15w/Mbps.

$$E_{\text{dynamic}} = \sum_{\ell \in \mathcal{L}} Traffic \times 0.15 \text{ W} / \text{Mbps} \tag{1}$$

The formula listed below is used to calculate the total static energy consumed by network components at rest .The components considered for this operation are edge servers , switches and routers. The values considered in this formula are constant variables. These values are derived from datasheet of Cisco ASR 100 and Jupiter EX2300.

$$E_{\text{static}} = \sum_{\in \mathcal{R}} P_{routers} + \sum_{\in \mathcal{S}} P_{switches} + \sum_{\in \mathcal{H}} P_{servers} \tag{2}$$

The final formula for energy consumption is listed below . It provides the total energy consumed by the system. % Total Energy

$$E_{\text{total}} = E_{\text{static}} + E_{\text{dynamic}} \tag{3}$$

IV. METHODOLOGY

This research employed a quantitative simulation-based approach to evaluate the performance and sustainability of different caching strategies in content delivery networks (CDNs). The methodology was designed to model real-world network conditions accurately and measure key metrics across multiple scenarios. The process consisted of four main phases: (1) Network Modeling and Data Preparation, (2) Scenario and Workload Generation, (3) Caching Strategy Implementation, and (4) Performance and Sustainability Measurement.

A. Network Modeling and Data Preparation

To ensure realism, the simulation was built upon real-world network topologies sourced from the Internet Topology Zoo, a repository of annotated network maps.

- **Network Selection:** Four distinct network topologies were selected for analysis to ensure diversity in structure and scale: Allegiance Telecom, At Home Network, CAIS Internet, and the AT&T Backbone.
- **Data Parsing and Graph Construction:** Each network was defined by three files parsed programmatically using Python and the NetworkX library:
  - **.dot Files:** Contained node information and their geographical `pos` attributes, used to build the graph structure and for visualization.
  - **Edge Files:** Specified the connections between nodes, including the initial link capacity (used as the base weight for path calculation).
  - **City Files:** Mapped node IDs to city names, adding a layer of real-world context.
- **Node Attribute Assignment:** Each node in the graph was randomly assigned a device type (router, *switch*, or server) with a probability distribution (20%, 60%, 20% respectively) to mimic a heterogeneous network infrastructure. A specific node within each topology was designated as the *origin* server, representing the central source of all content.

B. Scenario and Workload Generation

To test the caching strategies under varied conditions, four distinct operational scenarios were defined, each with a unique set of parameters controlling traffic and content popularity .

Parameter	Low	Normal	High	Extreme
Zipf Skew (α)	0.8	1.2	1.8	2.3
Traffic Range (Mbps)	5-30	30-100	100-300	300-500
Congestion Factor	0.2-0.3	0.4-0.6	0.7-0.8	0.85-1.0
Cache Size (items)	2-4	5-7	8-10	12-14
Number of Contents	10-20	20-50	50-100	100-150

C. *Caching Strategy Implementation*

Two advanced caching strategies were implemented and compared. Both were built upon a foundational **Least Recently Used (LRU)** cache algorithm at each network node. The purpose of this simulation was to disprove absolute superiority of a single caching technique.

V. Results

Fig.1 is the plotting of edge nodes on actual geographical location(i.e United states of America and London).

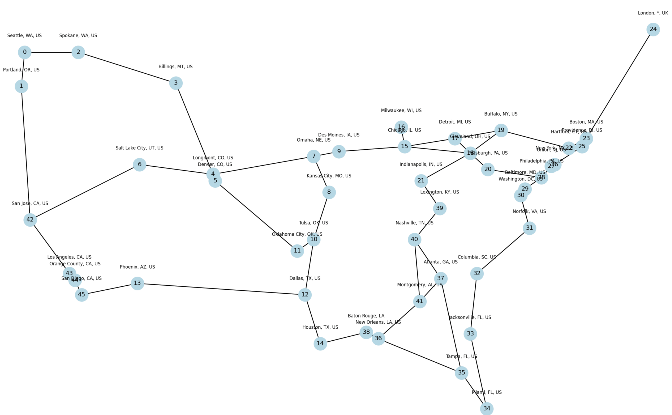

Fig. 1: **Plotted graph of the node positions over the United States of America and London)**

Following are the results generated using the model . These results are generated in different scenarios and conditions. Fig.2 indicates the primary approach of the model where we compared four caching techniques ; Edge caching , Content delivery at Source, cooperative caching and Predictive caching.

Fig.3 shows the later approach of the model to compare the best 2 caching techniques. In our case , the best caching techniques were co-operative caching and predictive caching. The energy cost results closely mirror the energy consumption trends, as expected. At low and medium traffic levels, Predictive Caching incurs a substantially higher cost than Cooperative Caching. The inefficiency observed in the energy consumption data directly translates into a greater financial expenditure. However, this trend reverses completely under high and extreme loads. In these scenarios, Predictive Caching becomes the more economical choice, offering significant cost savings compared to Cooperative Caching. This finding underscores the importance of aligning the caching strategy with the expected network load to optimize both performance and operational budget.

Based on the graphs provided, the research reveals a distinct pattern in the performance of Cooperative Caching (LRU) and Predictive Caching (LRU). The data shows a clear and consistent trade-off between the two methods based on the level of network activity. For low and medium events, Cooperative Caching is the superior choice across all metrics. It is more resource-efficient, consuming less energy, generating fewer carbon emissions, and incurring a lower energy cost. Furthermore, it achieves a higher cache hit ratio, indicating that its reactive approach to collaborative content sharing is

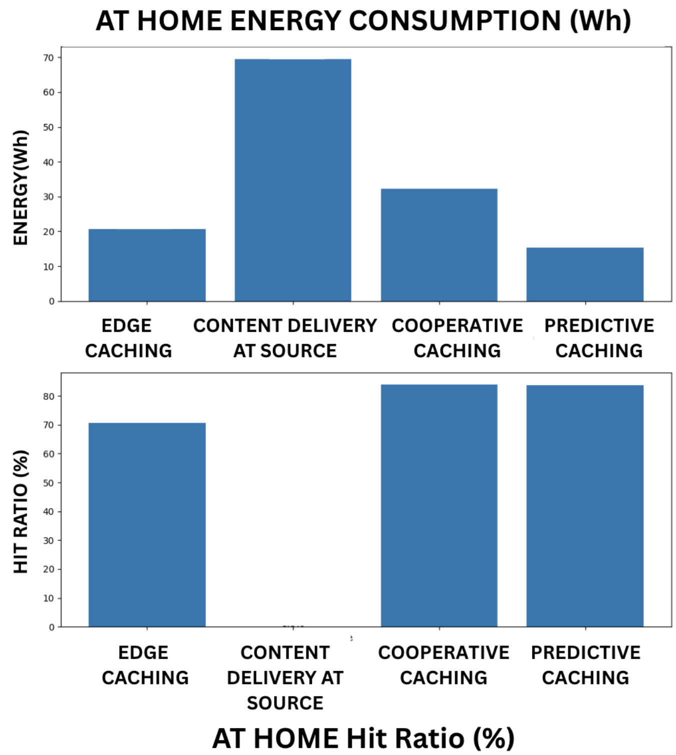

Fig. 2: **Energy consumption and hit ratio of all four caching techniques**

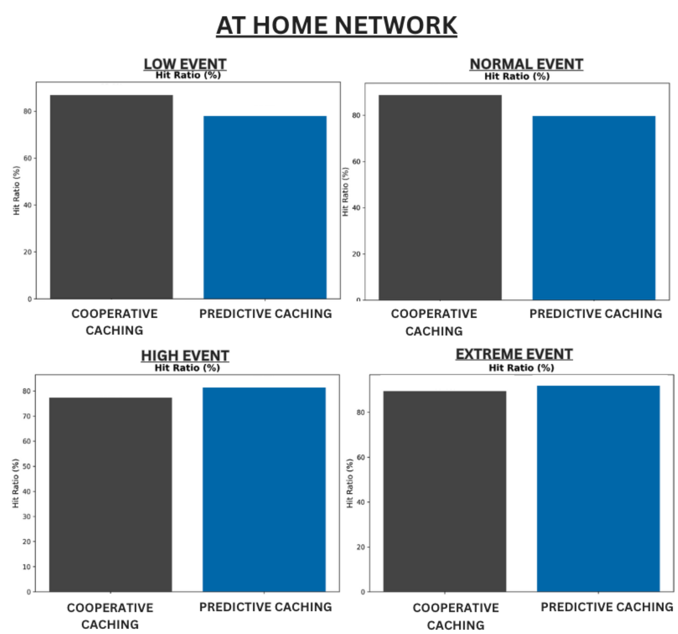

Fig. 3: **Hit ratio (%) comparison**

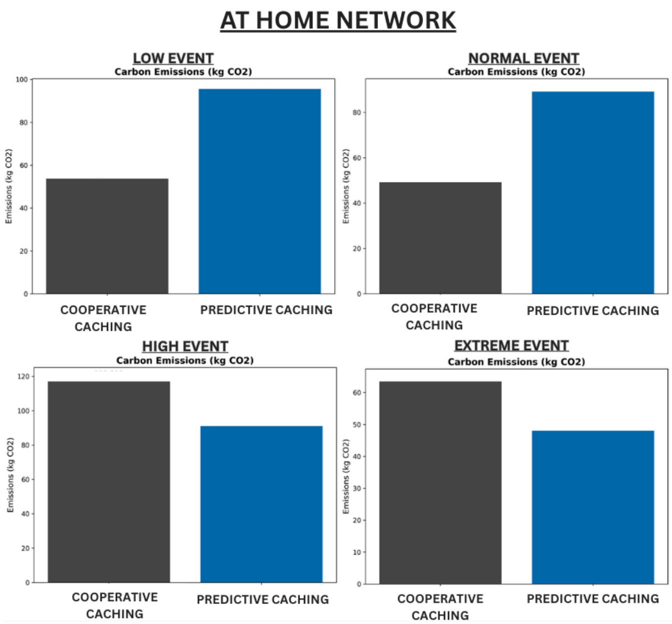

Fig. 4: **Carbon emission (Kg CO_2) comparison**

more effective in environments with lighter traffic.Conversely, the data demonstrates a complete reversal of this trend under high and extreme network loads. In these demanding scenarios, Predictive Caching emerges as the more effective and efficient solution. Its ability to anticipate user requests leads to a higher hit ratio, a crucial advantage when traffic is heavy. This superior performance translates to a reduction in energy consumption, energy costs, and carbon emissions. The graphs collectively suggest that while the overhead of a predictive algorithm is a liability in low-demand situations, it becomes a significant asset that optimizes network performance and resource use when the system is under stress.

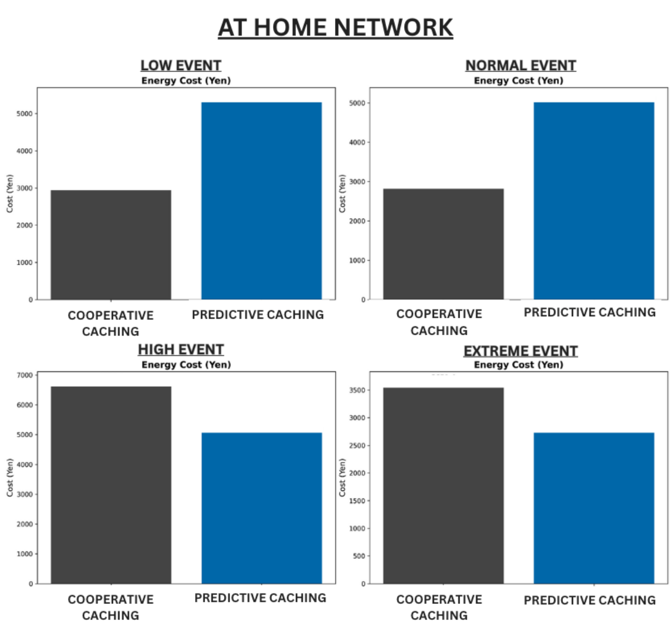

Fig. 5: **Energy cost (¥) comparison**

Based on the provided graphs, the research shows that for low and normal network events, Cooperative Caching is the superior choice. It demonstrates lower energy consumption and energy costs, and consequently, a smaller carbon footprint. Additionally, its hit ratio is higher under these conditions, indicating greater efficiency at retrieving cached content. However, as network activity increases to high and extreme event levels, the roles reverse. In these high-demand scenarios, Predictive Caching becomes the more effective method. It achieves a significantly higher hit ratio, and its energy consumption and associated costs and emissions are all lower than those of Cooperative Caching. The results suggest that the overhead of predictive algorithms is not justified for light network traffic but provides substantial benefits in high-stress environments.

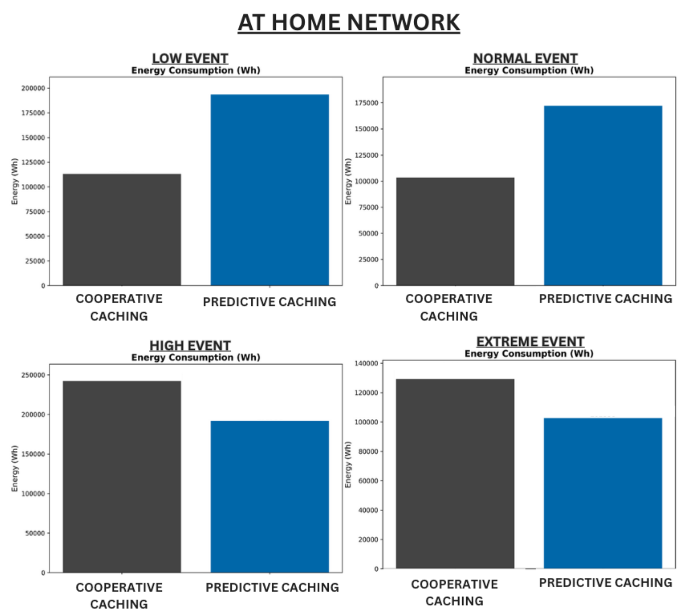

Fig. 6: **Energy consumption (Wh) comparison**

Analyzing the hit ratio data provides insight into the performance effectiveness of each caching method. For low and medium traffic conditions, Cooperative Caching consistently achieves a higher hit ratio, meaning it is more successful at retrieving requested content from the cache. This indicates that under light loads, the simple, cooperative sharing of cached data among network nodes is a highly effective strategy. The situation changes dramatically at high and extreme traffic levels. Here, Predictive Caching takes the lead, delivering a superior hit ratio. This suggests that in a high-demand environment, the ability to proactively pre-cache content that is likely to be requested is a more successful approach than relying on a reactive, cooperative model. The data implies that the predictive algorithm's value lies in its ability to manage high-volume, dynamic traffic more effectively.Post simulation a radar diagram helps the user understand the findings in an easier way.

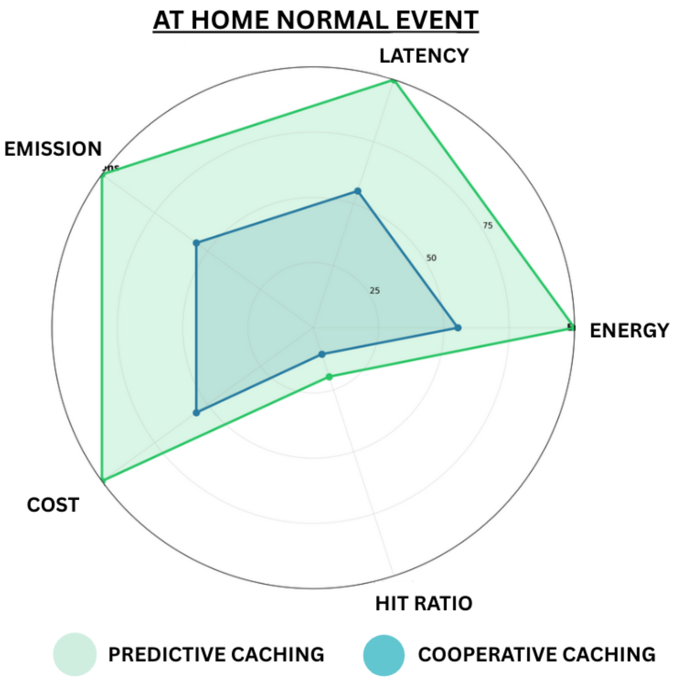

Fig. 7: **Radar comparison**

Post simulation a radar diagram helps the user understand the findings in a less complex way .(Fig.7)

VI. Conclusion

This research demonstrated that intelligent caching strategies are crucial for enhancing both the performance and sustainability of Content Delivery Networks (CDNs). Through a simulation of multiple real-world network topologies and scenarios, we evaluated Predictive and Cooperative Caching against traditional methods.Our results clearly show that there is no one-size-fits-all solution. Predictive caching excelled in high-demand, dynamic environments by anticipating user requests, significantly reducing energy consumption and carbon emissions. Cooperative caching proved highly effective in stable, lower-demand scenarios by leveraging shared resources across neighboring nodes.The key finding is that strategic cache management directly translates to substantial environmental and economic benefits, including lower CO_2 emissions and reduced operational costs. This work provides network operators with a clear framework for selecting the optimal strategy based on their specific performance and sustainability goals.Future work will focus on integrating more advanced machine learning models for prediction and validating these findings on physical network testbeds.

References

[1] G. K. Zipf. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley, 1949.

[2] Y. Ma, Y. Li, and J. Liu. "An adaptive caching strategy based on dynamic content popularity for information-centric networks." *IEEE Transactions on Vehicular Technology*, vol. 64, no. 12, pp. 5824-5835, 2015.

[3] D. Dantzig. *Content Delivery Networks: Architectures and Practical Solutions*. CRC Press, 2010.

[4] H. Wang, T. S. E. Lee, and Z. Chen. "On the performance of distributed caching with non-uniform content demand." in *Proceedings of the IEEE International Conference on Communications (ICC)*, pp. 1-6, 2018.

[5] T. Taleb and A. Ksentini. "Network parameter analysis for efficient caching in vehicular ad hoc networks." *IEEE Transactions on Network and Service Management*, vol. 4, no. 4, pp. 201-210, 2007.

[6] G. K. Zipf. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley, 1949.

[7] Y. Ma, Y. Li, and J. Liu. "An adaptive caching strategy based on dynamic content popularity for information-centric networks." *IEEE Transactions on Vehicular Technology*, vol. 64, no. 12, pp. 5824-5835, 2015.

[8] D. Dantzig. *Content Delivery Networks: Architectures and Practical Solutions*. CRC Press, 2010.

[9] H. Wang, T. S. E. Lee, and Z. Chen. "On the performance of distributed caching with non-uniform content demand." in *Proceedings of the IEEE International Conference on Communications (ICC)*, pp. 1-6, 2018.

[10] T. Taleb and A. Ksentini. "Network parameter analysis for efficient caching in vehicular ad hoc networks." *IEEE Transactions on Network and Service Management*, vol. 4, no. 4, pp. 201-210, 2007.

[11] Q. Li, W. Chen, and J. Wu. "Dynamic Caching Policy for Edge Computing based on Content Popularity Prediction." *Journal of Network and Computer Applications*, vol. 145, pp. 102436, 2019.

[12] G. Gomez and P. H. J. van der Meulen. "A survey on caching strategies in information-centric networking." *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 2061-2081, 2016.

[13] A. Perry and B. J. Evans. "Impact of Traffic Congestion on In-Network Caching Performance." in *Proceedings of the ACM Symposium on Edge Computing*, pp. 1-10, 2020.

[14] P. Bovet and M. P. G. van der Schaar. "An analytical model for cache hit probability under correlated content requests." *IEEE/ACM Transactions on Networking*, vol. 25, no. 2, pp. 883-896, 2017.