International Journal of Applied Engineering & Technology

A NOVEL APPROACH TO CAPTURE THE POPULARITY AND LOCATION OF THE CONTENT IN CACHE DISTRIBUTION NETWORKS

[1]Dr. N. Shanmuga Priya, [2]Mr. V. Nandeesh and [3]Ms. P. Nandhini.

[1]Associate Professor and Head, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

[2,3]PG Student, II MCA, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

ABSTRACT

With evolution of the Internet from an academic network to global infrastructure, the requirements of end-users have been widely changed. The emergence of Industry 4.0 and smart infrastructure require an effective communication backbone for all such applications. Thus, faster communication is required among the connected devices because end-users are more interested in accessing the content with minimal delay. However, traditional IP-based Internet architecture provides host-to-host based communication, which is inefficient to support rapid content distribution and its retrieval in the network. Additionally, current Internet architecture suffers from weak security, congestion control, inter-domain routing, and incompetent mobility support. Several caching schemes have also been proposed for IP-based networks to reduce network congestion. In information-centric networking, accurately predicting content popularity can improve the performance of caching. One of the foremost widely researched features extended by Information-Centric Networking (ICN) architectures are that the power to support packet-level caching at each node within the network. By individually naming every packet, ICN permits routers to invoke their lining-up buffers into packet caches and so tapping the network's existing storage resources. The proposed work is an analytical framework for distribution of popular content in an information centric network (ICN) that is comprised of access ICNs, a transit ICN, and a content provider. The performance of any caching system depends on content request came back by user. During this study we tend to propose a model to capture the recognition and placement of the content exploitation Short Noise Model (SNM). Also, the opposite attributes of the traffic that influence cache performances are studied.

Keywords: Information Centric Network (ICN), Packet, Short Noise Model (SNM), Content Distribution Network (CDN), IP Address.

1. INTRODUCTION

Recently, for common content, reducing the layoff in internet traffic through process caches to fulfill recurrent requests has long been a lively analysis topic. Outsized ranges of analysis examines the character of recent web traffic and have urged that caching has the potential to greatly slow down network load for a given traffic demand. Internet caches are important network components, landing common content nearer to the users, contributory to quicker information delivery, and reducing network and server load among ISPs and at giant stub networks [1]. Some analysis queries the efficiency of internet caches, reasoning that redundancy ought to be detected at a finer graininess, like packets, rather than objects. This style of caching additionally referred to as packet-level caches, which might be significantly simpler in wiping out continual content transfers. Also, they show important quantifiability and suppleness problems, like managing giant operation indexes, playing per packet lookups at wire-speed, operational in additional than one link and synchronizing operation indexes.

Most such weaknesses will doubtless be self-addressed by information-centric networking (ICN). ICN proposes a opportunity specification wherever all network operations concern data itself, in distinction to ip-based networking, wherever communication is endpoint-oriented. Most ICN initiatives adopt a model of receiver-driven content delivery of self-identified packets that may be quickly cached by routers, permitting routers to satisfy future requests for constant content. Still, ICN caching has not nonetheless met these expectations, receiving criticism for its potency, supported the debatable performance superiority of distributed in-network caching over freelance caches at the network edge, likewise as on the questionable support for packet-level caching by today's hardware [2].

The design and analysis of large-scale, interconnected systems of caches still desires clarity and understanding. Within the initial place, it is in ought to describe the traffic in terms of the sequence of contents' request generated by the users, and later processed by cache networks. Utilizing trace-driven simulations to assess the performance of cache design has many limitations and it is so extremely fascinating to own, initial of all, a correct model for the arrival method of contents' requests at the caches [3]. Because of the trend for present computing, creation of recent content publishers and customers, the web is changing into an progressively heterogeneous setting wherever completely different content varieties have different quality of service needs, looking on the content publisher/consumer. Such an increasing diversity in commission expectations advocates the necessity for content delivery infrastructures with service completely differentiation among different applications and content categories. Moreover, ancient cache management policies like LRU treat completely different contents during a powerfully coupled manner that creates it troublesome for (cache) service suppliers to implement differentiated services, and for content publishers to account for the valuation of their content delivered through content distribution networks which is shown in fig.1.1.

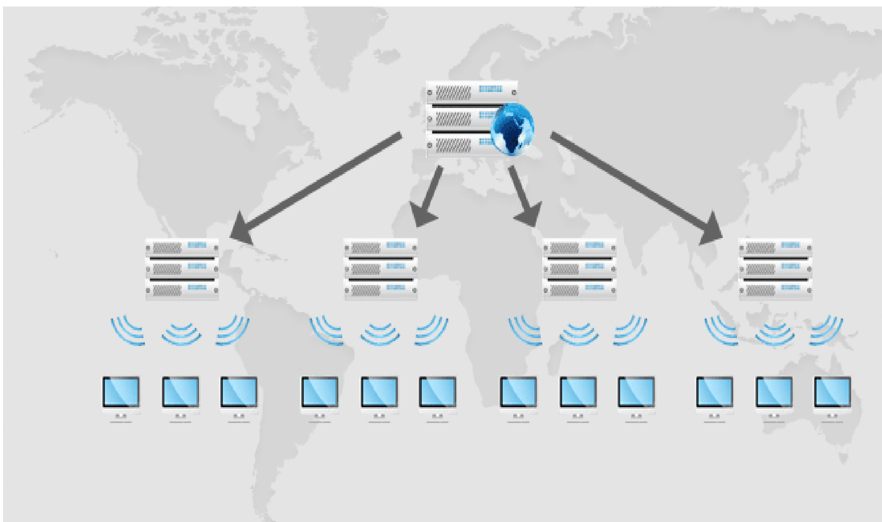

Fig 1.1: - Content Delivery System

Factors Impacting CDN Performance: CDNs became increasingly more important: the quantity of users with broadband web connections is continually increasing and, together with quicker property, return bigger expectations for higher content delivery. As associate example, requests of video content over the net are quite one billion per day. Although exploiting further resources provided by cdns, this demand puts a substantial pressure over the complete web [4]. This issue becomes even a lot of necessary if we tend to contemplate future trends: because the size of distributed content keeps growing, the space between server and shopper becomes a lot of important to the performance, since longer distances increase the probability of network congestion and packet loss that lead to longer transfer times.

In addition, the geographics of the requests influences the performance of cdns: it might be extraordinarily helpful to know whether or not an item becomes common on a planetary scale or simply during a explicit geographical region. A globally common content item ought to be replicated at each location, since it experiences several requests from all round the world. On the opposite hand, once content is barely domestically common, it ought to be cached solely within the locations that may expertise most requests. The key to such a technique is having the ability to predict quickly whether or not a bit of content is changing into domestically common so as to optimize its placement over the cdn before it undergoes the recognition soar.

2. RELATED WORKS

In IP networks, packet-level caching needs the detection of redundancy in absolute packets at wire-speeds. The machine price for avoiding replication via, say, suppressing replicated knowledge, deep packet examination and/or delta cryptography, has prevented internet caches from occupancy this direction. Interest in packet-level caching was rejuvenated by a computationally economical technique for locating redundancy in internet traffic, wherever Rabin fingerprints are accustomed observe similar, however not essentially identical, data transfers in real time. As this methodology is protocol freelance, it's going to even eliminate redundancy among totally different services, therefore greatly widening the scope of application caches. Sadly, this theme encompasses a restricted scope of applicability: it needs putting pairs of caching points at opposite ends of a physical link, exchange redundant knowledge with a special symbol as packets enter and leave that link. The two caching points should conjointly keep their search indexes synchronous [5]. A number of years later, the appliance of this method were explored in an inter-domain situation. Although the theme performed much better than a normal object cache, it had been another time finished that this answer will solely be applied to limited-scale deployments across specific network links. The authors argued that the quality of this method may well be increased by new network protocols that will leverage link-level redundancy elimination [6].

The identifying feature of ICN is that the placement of knowledge within the center of network operations, in distinction to endpoint-oriented science networks [7]. In ICN the functions of requesting, locating and delivering data area unit directly supported the data itself, instead of on the hosts providing the content. In most ICN proposals, data travels through the network as a group of self-verified knowledge chunks that carry a statistically distinctive symbol. This symbol, that is typically a concatenation of the content's name and also the packet's rank/order within the content, is placed within the packet header, relieving ICN nodes from the machine prices of sleuthing identical packets; if 2 packets have identical symbol, then they need to (statistically) carry identical content. Within the overwhelming majority of ICN studies, a chunk refers to the Most Transfer Unit (MTU) of the network, that is, the most packet allowed, hence, we are going to use below the terms packet and chunk as synonyms. ICN transport protocols area unit largely receiver-driven [8, 9], finishing a transmission via varied freelance transfers of self-verified chunks. Every transfer is triggered by a selected request packet and is consummated by the transmission of the corresponding knowledge packet.

The pull model permits exploiting on-path caches: ICN routers that use their queuing buffers as temporal repositories for packets will directly reply to later requests for these packets.

ICN has nice potential for exploiting packet-level caches, so several researchers have investigated the gains of present caching [10, 11, 12, and 13]. The authors of those papers attempt to mixture the caching potential of all on-path routers into a distributed caching system, specializing in achieving the foremost profitable distribution of content across these routers. However, expertise with distributed caching systems suggests that dedicated caching super-nodes at the sides of the network will have identical impact as caching at each in network node [12]. Additionally, some authors advocate caching content solely at a set of network nodes that satisfy bound spatial relation necessities [19], whereas others argue that an "edge" caching preparation provides roughly identical gains with a universal caching design [18].

To the simplest of our information, there's just one study within the literature handling the inner details of ICN packet caches [14]. This study proposes a two-layer cache model with the goal of up latent period. Specifically, it suggests that teams of chunks ought to be pre-fetched from the slow memory (ssd) to the fast one (dram) so as to reply quicker to subsequent chunk requests. However, the authors propose this style just for edge routers, because of its storage necessities and static content catalogue. For in-network routers they argue that each sram and dram ought to be utilized for wire-speed operation. Most different analysis merely assumes a Least Recently Used (LRU) replacement policy [9, 10, and 13] or novel policies for the correct distribution of the cached content on the trail [8, 11], while not evaluating whether or not router-cache performance is proscribed by the dimensions of its quick memory.

3. PROPOSED WORK

To better perceive the impact on cache performance on the geographical neck of the woods, the request method within the cdn ought to be treated as stratified of the many freelance processes, every touching on a private content [20]. As such, the arrival method of a given content is specific by three physical parameters, the time instant at that the content enters the system, the typical range of requests generated by the content and also the “popularity profile”, describing however the request rate for content evolves over time. For now, to induce an initial understanding of the projected work, take into account a content with a regular quality profile and this quality is predicated on life-span that we have a tendency to fix manually. on the far side the life-span, the recognition is reset and another time the populariry is computed. conjointly the life span value of content helps in understanding the realtime quality of that content.

The new contents become out there within the system in step with a unvaried poisson method rate of your time instants forms a customary poisson method. It has a tendency to visit this model as shot noise model (snm), since the method of requests arrival is understood as a poisson shot-noise method. we have a tendency to underline that the on top of poisson assumptions, on the (instant) generation method of requests for every content, and on the arrival method of latest contents, are basically introduced for the sake of analytical tractableness [21]. However, it's not very vital to capture complicated arrival dynamics at short time-scales. For a given content, the snm needs us to specify its entire quality profile within the kind of the perform that, given the issue in estimating quality profiles from a trace, may be thought of as a limitation. However, it's not essential to precisely discover the form of the content with relevance time. In fact, a straightforward first-order estimate, in step with that we have a tendency to specify the typical content life-span, is enough to get correct predictions of cache performance [16]. Finally, the content nonuniformity is taken into consideration by relating every individual content to its life-span beside the typical range of requests. this suggests that, on arrival of every new contents, we will every which way choose the combine of parameters from a given assigned joint distribution.

For each given content within the trace initial we have a tendency to calculate the full range of discovered requests then we tend to estimate the content life-span. The interested user will found the small print concerning the way to estimate from the traces. Contents are then partitioned off into categories, on the premise of previous values life-span and requests. category contains all the contents with tiny range of requests that we have a tendency to cannot derive a reliable estimate of their life-span. To handle the case of multiple interconnected caches, within which edge caches receive the requests generated by subsets of users presumably having quite completely different interests from one point to a different geographical neck of the woods. The extension of the fundamental model will be, in essence, allotted in its most general kind by associating to each content and entry purpose, a tuple, so that, at entry purpose, the requests for content arrived is given by the content, life-span and also the profile for a given time.

The popularity of a given content is assumed to follow a similar profile at each entry purpose. the actual form of the recognition profile represent every content by its start line of availableness, its life-span and a group of parameters denoting the volumes of requests attracted by content at completely different entry points. To specify the volumes of requests generated by contents at the assorted entry points is given by, for every content we tend to assign a world volume denoting the full range of requests generated within the whole system and it's diagrammatical as volume of the fraction of requests for content at any entry purpose. This approach permits an outsized degree of flexibility in describing the geographical neck of the woods of contents.

Class	Classification rule	Trace	%Reqs	%Videos	$E[\hat{L}_m]$	$E[\hat{V}_m]$
Class 0	$\hat{V}_m < 10$	Trace 1	74.60	98.588	-	1.41
		Trace 2	72.64	98.401	-	1.42
		Trace 3	72.53	98.210	-	1.44
		Trace 4	67.30	97.778	-	1.49
Class 1	$\hat{V}_m \geq 10$, $\hat{L}_m \leq 2$	Trace 1	2.34	0.044	1.14	86.4
		Trace 2	2.71	0.083	1.09	76.2
		Trace 3	2.60	0.067	1.04	76.0
		Trace 4	2.81	0.077	1.06	74.0
Class 2	$\hat{V}_m \geq 10$, $2 < \hat{L}_m \leq 5$	Trace 1	1.72	0.069	3.36	41.9
		Trace 2	3.43	0.125	3.34	50.7
		Trace 3	1.77	0.082	3.32	43.3
		Trace 4	2.01	0.093	3.41	48.0
Class 3	$\hat{V}_m \geq 10$, $5 < \hat{L}_m \leq 8$	Trace 1	1.49	0.041	6.40	59.5
		Trace 2	1.84	0.070	6.31	44.9
		Trace 3	1.66	0.052	6.42	63.3
		Trace 4	1.64	0.062	6.45	60.3
Class 4	$\hat{V}_m \geq 10$, $8 < \hat{L}_m \leq 13$	Trace 1	1.39	0.062	10.53	36.9
		Trace 2	2.96	0.128	10.86	39.6
		Trace 3	1.33	0.066	10.62	39.5
		Trace 4	1.75	0.103	10.65	37.8
Class 5	$\hat{V}_m \geq 10$, $\hat{L}_m > 13$	Trace 1	18.46	1.196	24.61	25.7
		Trace 2	16.41	1.193	19.29	25.3
		Trace 3	20.11	1.523	28.19	25.8
		Trace 4	24.49	1.887	24.59	28.1

Fig. 3.1: - Cross Classification of Contents

To show however the life-span of content measured in numerous traces will vary in size, for a similar two pairs of traces thought of before, the fraction of contents that, given associate initial classification in one trace, were categoryified into a given class within the second trace. By construction a table with rows and columns and assignment these two traces sums to 1. Moreover, the fraction of contents that the category index differs by quite one is negligible. Note that contents whose category index differs specifically by one ought to be smitten care, as a result of their life-spans could be near the brink price separating two neighboring categories, which may simply cause a misclassification. to examine however the non-perfect synchronization existing in real traces will have an effect on the ensuing cache performance. particularly, if we have a tendency to thought of a straightforward cache network composed by one root and two leaves. we have a tendency to assume that the foundation takes hour of the cache size and every leaf takes 20% of it. Then we have a tendency to fed the left leaf by trace one and also the right leaf by trace three. These traces wherever chosen as a result of they contain similar request volumes and have an outsized overlap. within the case of a miss, requests are forwarded to the foundation. Contents are replicated on all caches lined by letter of invitation.

Final class		Class 1	Class 2	Class 3	Class 4	Class 5
Initial class		Trace 7				
Trace 6	Class 1	0.69	0.23	0.04	0.02	0.02
	Class 2	0.11	0.64	0.16	0.05	0.04
	Class 3	0.03	0.21	0.44	0.16	0.16
	Class 4	0.01	0.07	0.18	0.35	0.39
	Class 5	0.00	0.00	0.01	0.02	0.97
		Trace 8				
Trace 3	Class 1	0.86	0.06	0.04	0.00	0.01
	Class 2	0.30	0.56	0.08	0.03	0.03
	Class 3	0.10	0.27	0.39	0.13	0.11
	Class 4	0.02	0.12	0.26	0.44	0.16
	Class 5	0.00	0.01	0.02	0.03	0.94

Fig 3.2: - Cross Classification of Contents from Two Traces

Based on the above observations, we realize that caching indiscriminately does not necessarily guarantee the highest cache hit rate. On the other hand, this result cannot be used as conclusive evidence that caching less is better since the string topology constrains to a large extent the diversity of the content delivery paths (i.e., all delivery paths are fully or partially overlapping), a fact that indirectly increases the probability of a cache hit. Following this argument, we propose a novel caching scheme based on the concept of betweenness centrality [17] which measures the number of times a specific node lies on the content delivery path between all pairs of nodes in a network topology. The basic idea is that if a node lies along a high number of content delivery paths, then it is more likely to get a cache hit.

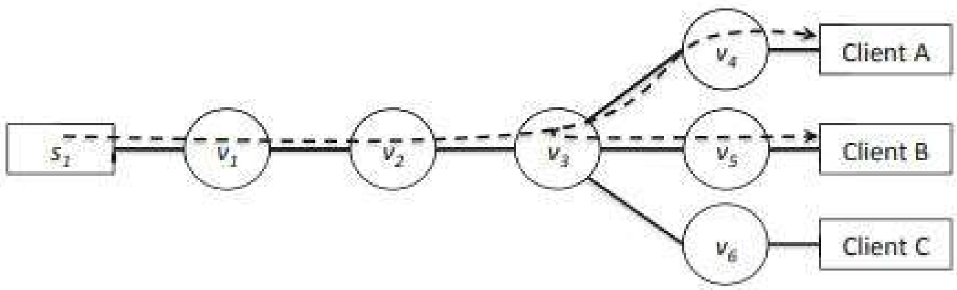

Fig. 3.3: - an example topology with optimal caching location at V_3

However, with a bird's eye view, it is clear that caching the content only at V_3 is sufficient to achieve the best gain without caching redundancy at other nodes. This can be verified by using the betweenness centrality, whereby V_3 has the highest centrality value with most content delivery paths passes through it (i.e., 9 paths).

4. CONCLUSION

In general, when believe content distribution systems spanning massive areas, together with terribly different time zones, profiles having little life-span mustn't be thought of synchronized, however properly changed to require collectively into consideration geographical neck of the woods and periodic patterns. The shot noise model provides a straightforward, versatile and correct approach to drawing the geographical neck of the woods of cache performance. The projected work would facilitate realize cache performance will considerably get pleasure from the presence of even a comparatively little portion of extremely widespread and native contents, and whether or not geographical neck of the woods plays conjointly a very important role within the marking of distributed caching systems and therefore the overall impact on cache performance of the distribution of the quantity of requests attracted by the contents is considerably intense with regard to traditional stationary models like IRM.

5. REFERENCES

[1] Liu, W. X., Zhang, J., Liang, Z. W., Peng, L. X., & Cai, J. (2017). Content popularity prediction and caching for ICN: A deep learning approach with SDN. IEEE access, 6, 5075-5089.

[2] Hajimirsadeghi, M., Mandayam, N. B., & Reznik, A. (2017). Joint caching and pricing strategies for popular content in information centric networks. IEEE Journal on Selected Areas in Communications, 35(3), 654-667.

[3] Nguyen, M. T., Le, D. H., Nakajima, T., Yoshimi, M., & Thoai, N. (2019, August). Attention-based neural network: A novel approach for predicting the popularity of online content. In 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS) (pp. 329-336). IEEE.

[4] Li, L., Xu, Y., Yin, J., Liang, W., Li, X., Chen, W., & Han, Z. (2019). Deep reinforcement learning approaches for content caching in cache-enabled D2D networks. IEEE Internet of Things Journal, 7(1), 544-557.

[5] Kumar, S., & Tiwari, R. (2020). Optimized content centric networking for future internet: dynamic popularity window based caching scheme. Computer Networks, 179, 107434.

[6] Sadeghi, A., Wang, G., & Giannakis, G. B. (2019). Deep reinforcement learning for adaptive caching in hierarchical content delivery networks. IEEE Transactions on Cognitive Communications and Networking, 5(4), 1024-1033.

[7] Tan, Z., Hu, W., Zhang, Y., & Ding, H. (2019, December). Online popularity prediction of video segments: Towards more efficient content delivery networks. In 2019 IEEE Global Communications Conference (GLOBECOM) (pp. 1-6). IEEE.

[8] Salahuddin, M. A., Sahoo, J., Glitho, R., Elbiaze, H., & Ajib, W. (2017). A survey on content placement algorithms for cloud-based content delivery networks. IEEE Access, 6, 91-114.

[9] Qazi, F., Khalid, O., Rais, R. N. B., & Khan, I. A. (2019). Optimal content caching in content-centric networks. Wireless communications and mobile computing, 2019.

[10] Peng, G. (2004). CDN: Content distribution network. arXiv preprint cs/0411069.

[11] Anjum, N., Karamshuk, D., Shikh-Bahaei, M., & Sastry, N. (2017). Survey on peer-assisted content delivery networks. Computer Networks, 116, 79-95.

[12] Liu, H., Chen, Z., Tian, X., Wang, X., & Tao, M. (2014). On content-centric wireless delivery networks. IEEE Wireless Communications, 21(6), 118-125.

[13] Abdelkrim, E. B., Salahuddin, M. A., Elbiaze, H., & Glitho, R. (2016, December). A hybrid regression model for video popularity-based cache replacement in content delivery networks. In 2016 IEEE Global Communications Conference (GLOBECOM) (pp. 1-7). IEEE.

[14] Kim, Y., & Yeom, I. (2013). Performance analysis of in-network caching for content-centric networking. Computer networks, 57(13), 2465-2482.

[15] Sahoo, J., Salahuddin, M. A., Glitho, R., Elbiaze, H., & Ajib, W. (2016). A survey on replica server placement algorithms for content delivery networks. IEEE Communications Surveys & Tutorials, 19(2), 1002-1026.

[16] Kirilin, V., Sundarrajan, A., Gorinsky, S., & Sitaraman, R. K. (2019, August). Rl-cache: Learning-based cache admission for content delivery. In Proceedings of the 2019 Workshop on Network Meets AI & ML (pp. 57-63).

[17] Tuncer, D., Sourlas, V., Charalambides, M., Claeys, M., Famaey, J., Pavlou, G., & De Turck, F. (2016). Scalable cache management for ISP-operated content delivery services. IEEE Journal on Selected Areas in Communications, 34(8), 2063-2076.

[18] Berger, D. S., Sitaraman, R. K., & Harchol-Balter, M. (2017). {AdaptSize}: Orchestrating the Hot Object Memory Cache in a Content Delivery Network. In 14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17) (pp. 483-498).

[19] Domingues, G., e Silva, E. D. S., Leão, R. M., Menasché, D. S., & Towsley, D. (2017). Enabling opportunistic search and placement in cache networks. Computer Networks, 119, 17-34.

[20] Li, J., Phan, T. K., Chai, W. K., Tuncer, D., Pavlou, G., Griffin, D., & Rio, M. (2018, April). Dr-cache: Distributed resilient caching with latency guarantees. In IEEE INFOCOM 2018-IEEE Conference on Computer Communications (pp. 441-449). IEEE.

[21] Jošilo, S., Pacifici, V., & Dán, G. (2017). Distributed algorithms for content placement in hierarchical cache networks. Computer Networks, 125, 160-171.