INVITED PAPER *Special Section on Internet Architectures and Management Methods that Enable Flexible and Secure Deployment of Network Services*

A Survey of Caching Networks in Content Oriented Networks

Miki YAMAMOTO[†a)], ***Fellow***

SUMMARY Content oriented network is expected to be one of the most promising approaches for resolving design concept difference between content oriented network services and location oriented architecture of current network infrastructure. There have been proposed several content oriented network architectures, but research efforts for content oriented networks have just started and technical issues to be resolved are still remained. Because of content oriented feature, content data transmitted in a network can be reused by content requests from other users. Pervasive cache is one of the most important benefits brought by the content oriented network architecture, which forms interconnected caching networks. Caching network is the hottest research area and lots of research activities have been published. This paper surveys recent research activities for caching networks in content oriented networks, with focusing on important factors which affect caching network performance, i.e. content request routing, caching decision, and replacement policy of cache. And this paper also discusses future direction of caching network researches.

key words: *cache networks, content oriented networks, in-network cache*

1. Introduction

According to Cisco whitepaper [1], total Internet traffic is predicted to be over 2 zettabytes by 2019 and CDN (Content Delivery Network) traffic will be nearly two thirds of total traffic. User generated traffic, such as YouTube traffic, is also growing rapidly, which will accelerate increase speed of content distribution traffic. As content distribution traffic is becoming major traffic in the current Internet, effective content distribution is one of the most important technical problems in networking research fields.

For content distribution services, several services have offered a content-oriented services in the current Internet. These include CDN [2] and P2P [3]. In CDN, a user's request is redirected to the adequately selected replication server. In P2P, content file is divided into small size of chunks and each chunk can be obtained from different peers. In these content distribution services, a user is just interested in the desired content itself and not in where it is obtained, which is the content-oriented feature. In this sense, the current Internet has already serves content-oriented services to network users. However, its fundamental network architecture is still location-oriented, because each IP packet is identified by IP address which is location-based ID.

Recently, content oriented network has been one of the hottest research fields in network engineering. Content oriented network is a clean slate approach for new generation networks and changes network architecture itself towards content-oriented one. It bridges enormous conceptual gap between network services and network architecture in the current Internet. In content oriented networks, a data packet is generally identified by content name and can be re-used for another request for the same content. So, content cache is universally equipped in a router. In-network caching is widely recognized as one of the most important elements in content oriented networks.

In-network caching is not a new concept. It is well-studied for web caching, such as hierarchical cache [4], [5]. However, in-network caching in content oriented networks has unique aspect, pervasive cache. All (or lots of) routers in networks are equipped with cache storage and these caches form complicated total cache systems. Performance of complicated in-network caching system depends on many factors, such as popularity of contents, routing of content request packets, cache decision and cache replacement policy.

In this paper, after preliminary survey of content oriented network researches in Sect. 2, we would like to survey interesting pro and con discussion for caching network performance in Sect. 3. In this discussion, some papers insist that in-network caching brings little performance improvement. And there have been published some papers standing contrary position, i.e. performance improvement of in-network caching is unreasonably evaluated to be little in these con papers. Following this, we would like to survey research activities for caching networks in content oriented networks from the following aspects. Performance of cache system deeply depends on popularity distribution of contents. First, we would like to survey interesting measured reports for popularity of contents in Sect. 4. Caching network performance also depends on content request routing, caching decision and cache replacement policy. Content request routing plays an important role for a content request to encounter cache storage holding the requested content. In Sect. 5, we would like to survey published papers dealing with content request routing taking account of cache location. In Sect. 6, we survey many interesting approaches for caching decision with which each router decides which content to be stored in its local cache. In Sect. 7, replacement policy for local cache is surveyed. Replacement of cached content is activated when caching decision decides content to be cached and eviction of stored content is necessary for making a room for a newly cached content. Finally, we would like to conclude this survey paper with some discussion about future direction.

Manuscript received December 7, 2015.
Manuscript revised January 17, 2016.

[†]The author is with the Faculty of Engineering Science, Kansai University, Suita-shi, 564-8680 Japan.
a) E-mail: yama-m@kansai-u.ac.jp

DOI: 10.1587/transcom.2015AMI0001

2. Content Oriented Networks

Since large portion of network traffic transferred in the current Internet is web-related content distribution, such as video distribution, content oriented network which is expected to be a promising architecture to enable effective content distribution, is one of the hottest research fields in network engineering. In this section, we would like to introduce short historical background of content oriented networks and its technical relationship to caching networks.

2.1 Content Oriented Approaches on the Internet

As content distribution becomes major traffic in the Internet, content distribution service changes its style in order to resolve content request concentration to popular web servers. CDN (Content Delivery Network) [2] is the first approach for improving user-perceived performance for content distribution. In CDN, a content request sent from a user is redirected to the adequately selected replicated server by DNS name resolution. CDN provides implicit content-oriented *service* to end-users because users have an impression that their content request is sent to the original server. In the sense that end-users do not care about "where the content is obtained", CDN provides *content-oriented service*.

P2P [3] provides more sophisticated style of content-oriented service than CDN. In P2P, content file is divided into fixed size of chunks and each chunk can be obtained from any peer providing it. This means chunks constructing the same file can be downloaded from different peers. In this sense, P2P has more content-oriented feature, i.e. user do not care about "where nor from whom the content is obtained."

As shown in CDN and P2P, content distribution services provided on the Internet have already been "content-oriented" service. However, architecture of underling network, i.e. the Internet, is still location-oriented one, which means there is a significant difference between design concepts of network services and the underling network infrastructure.

Research for content-oriented networks started around the beginning of 2000's. TRIAD [6], *i3* [7], and DONA [8] are the most famous approaches positioned as the launch of content-oriented networking research. In these approaches, content discovery in content-oriented manner is newly proposed. However, content transfer phase of all of these proposals is assumed to be still IP-based one. This is because in early 2000's, the Internet, i.e. IP networks, was securing a firm position for network infrastructure and timing of their proposals was too early to propose a new content transfer mechanism in content-oriented manner.

2.2 Clean Slate Approaches

After early stage of content-oriented networking research, e.g. *i3* and DONA, purely content-oriented networking architectures have been proposed. These proposals include PSIRP [9], PURSUIT [10], [11], SAIL [12], 4WARD [13], NetInf [14], [15], Mobility First [16], [17] and CCN/NDN [18]–[20]. CCN (Content Centric Networking: or called NDN (Named Data Networking)) is one of the most promising architectures for content oriented networks. This paper does not just focus on CCN/NDN, but we would like to pick up CCN/NDN and explain its architecture in detail.

CCN/NDN has a consumer-driven, i.e. pull-based, architecture. A user requiring content sends content request, called Interest packet. Each Interest packet requests a fixed-size part of content, called Data chunk, and brings an identifier of each required chunk. Content sources (called repositories) advertise their holding contents by a routing protocol and this advertisement of contents makes up routing table, FIB (Forwarding Information Base), at each router. When an Interest packet arrives at a router, it is forwarded to the interface selected from FIB entries (in CCN/NDN, FIB might have multiple entries for a content name). On forwarding an Interest packet, a router stores the interface from which the corresponding Interest packet arrives, to its pending table called PIT (Pending Interest Table). This table, PIT, is used for Data packet forwarding, which means in CCN/NDN, Data packets will track back on the exactly reverse path of the corresponding Interest packets.

In the IP networks, when a packet is forwarded, it is removed from a router buffer. This is because a packet will not be re-used in the IP networks. However, in content oriented networks, a Data packet might be re-used for another Interest packet, when an Interest packet for the same chunk arrives. This is content oriented feature, i.e. content can be obtained anywhere it is found. So, in CCN/NDN, a router is equipped with a cache for Data packets, called CS (Content Store).

A CCN router operates as follows (Fig. 1). When an Interest packet arrives, CS is checked first whether it has the corresponding Data chunk (Fig. 1.(a)). When CS caches it, the cached Data chunk is transmitted back to the consumer. Since the paper focuses only on cache, we explain only cache behavior. Other operation in CCN routers, such as PIT and FIB for Interest packets, please refer to CCN/NDN papers [19], [20]. As shown in Fig. 1(b), for a Data packet, a CCN

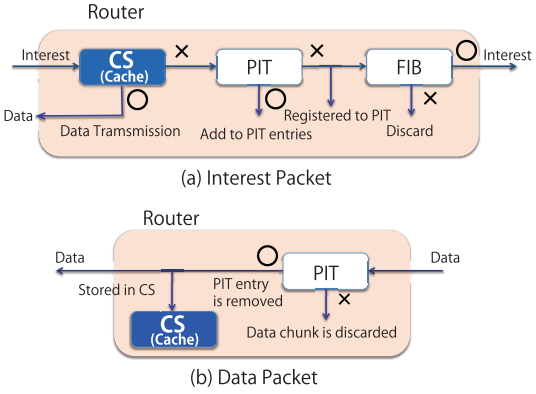

Fig. 1 Behavior of CCN router (CS is a local cache).

router forwards it by following PIT entry (when PIT is aggregated, entries for multicast). And this Data packet is cached at a local chache, CS, according to the caching decision policy.

In content oriented networks, a router generally has its local cache. So, ubiquitous in-network cache is one of the most important technical feature of the content oriented networks.

3. Caching Networks

In content-oriented networks, a router is equipped with cache storage and cache storages in routers construct caching networks.

3.1 Web Caching

Web caching was a hot research topic in 1990s. In web caching, cooperation of proxy caches and hierarchical cache have been proposed in many papers, such as [24]–[27]. These approaches for constructing total cache systems with structured caches are the first well-known example of caching networks. Several interesting insights have been reported for these organized caches. For example, N. Laoutaris et al. [28] reveal caching decision of LCD (Leave Copy Down) in web caching shows good performance (we explain about LCD in detail in Sect. 6).

For caching network performance, there have been published pro and con papers. P.Danzig et al. [29] evaluate cache performance using FTP trace. Even though this evaluation uses FTP traces, it reveals that caches in core-network play an important role for improving cache performance. This paper shows positive results for caching networks. However, A.Wolman et al. [30] present negative evaluation results of cooperative caching or hierarchical caching by trace-driven simulations. This negative conclusion [30] leads to also negative viewpoint papers in caching network of content oriented networks as shown in Sect. 3.3.

Although Web caching gives us several interesting insights for caching networks, it has significant difference from caching networks in content oriented networks. Web caches are well-organized while in content oriented networks, caches are not strictly organized (are organized in a distributed manner). So, more research efforts are necessary to understand caching network behavior in content oriented networks.

3.2 Caching Network in Content Oriented Networks

Performance of caching network depends on many factors. First, popularity of contents is very important factor. Y. Wang et al. [31] show that cache location at network edge is a good selection for skewed demand popularity and cache at network core is good for flat popularity distribution. For content popularity, we would like to discuss in detail in Sect. 4.

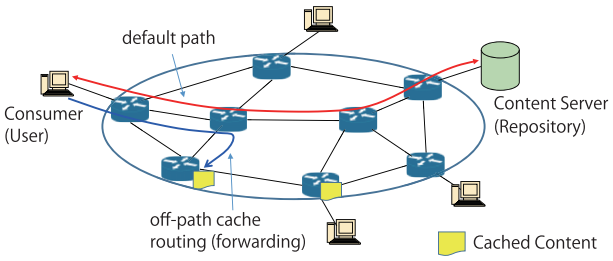

Fig. 2 Content request routing for caching networks.

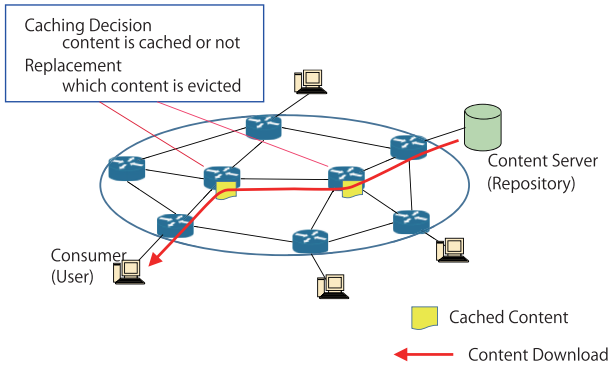

Fig. 3 Caching decision and replacement in caching networks.

G. Rossini and D. Rossi [32] show combination of content request routing (they call forwarding F in their paper), caching decision (meta-caching algorithm D), and replacement (replacement algorithm R) plays an important role for caching network performance.

Content request routing decides scope of content search. Almost papers assume default-path routing, where content requests are forwarded on the shortest path to the repository (the closest repository when there are multiple repositories). With the default-path routing, cached content not along the default path cannot be encountered by a content request, even when there are cached contents close to the consumer (Fig. 2). We discuss content request routing in detail in Sect. 5.

According to caching decision strategy, each router makes a decision whether passing content is to be cached at its local cache or not (Fig. 3). Caching decision determines the distribution of cached contents in a whole network. Widely used caching decision of TERC (Transparent En-Route Cache) has a tendency that only highly popular contents are widely spread inside a network, which means cache capacity in a whole network is not effectively utilized. To resolve this inefficiency of TERC, many interesting caching decision policies have been proposed as shown in Sect 6.

Replacement is also important factor which affects caching network performance. Replacement policy determines which stored content is to be evicted to make a room for newly stored content. This newly stored content is decided by caching decision, and these two policies, caching decision and replacement, totally affect cached content dis-

tribution in a whole network (Fig. 3).

3.3 Performance of Caching Networks

Pervasive cache in a network is one of the most promising aspects of content oriented networks. There have been published several papers discussing pros and cons of pervasive cache. Papers by A. Ghodsi et al. [22] and S.K. Fayazbakysh et al. [21] are the most famous papers for negative opinions for pervasive cache. Performance evaluation results in the latter paper show that small or little improvement can be obtained for pervasive cache when compared with cache location only at edge of the network. This means almost performance gain for caching popular contents is brought from edge caches. Their evaluation model is tree structure for topology and en-route cache for cache decision.

In contrast, several positive position papers for pervasive cache have already published. W.K. Chai et al. [77] show en-route cache where all routers along the default path store passing contents is not a good decision. They conclude that less cache inside a network brings more improvement. Their caching decision takes account betweeness (as explained in Sect. 6), which means caches at network-core are more important than edge cache. Their performance evaluation results show that pervasive cache is still a controversial issue.

G. Tyson et al. [23] evaluate in-network cache performance by trace-driven simulation. They show that 11% content requests are treated in the requesters' network in the case of cache located only at the network edge. They also show that 32% content requests are treated in it in pervasive cache case, which means network-edge cache can bring performance improvement but pervasive cache in content oriented network will bring more improvement.

G. Rossini and D. Rossi [32] focus on combination of content request routing and caching decision. Their performance evaluation results show that adequate combination of content request routing and caching decision will potentially bring more performance improvement. Negative position paper by S.K. Fayazbakysh et al. [21] assumes tree topology where there is small number of neighbor nodes, so G. Rossini and D. Rossi argued that caching performance will be improved more in the case where there is more neighbor nodes, such as for mesh network model.

According to this debate for pervasive cache in content oriented networks, more sophisticated performance evaluation are expected to be continuously investigated because caching performance depends on many factors, such as popularity of contents, network topology, caching decision, and content request routing.

4. Popularity Model

As described in the previous section, popularity of contents plays an important role for caching performance. Many reports of observed traffic have been published from the viewpoint of popularity of contents. These published papers are categorized in four classes from the viewpoint of content types (Fig. 4): Web content, P2P, VoD, and UGC (User Generated Content).

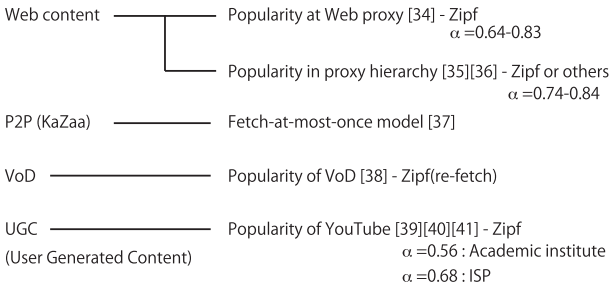

Fig. 4 Popularity model.

Web content popularity paper by L. Breslau et al. [34] is a very famous one and cited in lots of research papers. It analyzes 6 data sets obtained in academic, corporate, and ISP environments. These data sets are collection of data in several periods, between one day and 3months. Popularity of web content in data sets of this paper follows Zipf distribution with $\alpha = 0.64 - 0.83$. Zipf distribution [33]† is very famous popularity distribution, where the probability of a request to the $i-th$ most popular content, P_i has the following feature,

$$P_i \propto \frac{1}{i^{\alpha}}. \tag{1}$$

Value of α is reported to be slightly different for each data set. The papers by A. Mahanti et al. [35] and R. Doyle [36] analyze content popularity in different situation from the paper by L. Breslau [34], hierarchical web proxy. A. Mahanti et al. [35] show that web content popularity follows Zipf distribution with $\alpha = 0.74 - 0.84$. The further web proxy from users has the lower value of α. This is because content requests in higher hierarchy proxy (further proxy from users) are filtered by lower proxy, which makes Zipf distribution of higher web proxy flatter. R. Doyle et al. [36] show that content requests coming through web proxy have slight different distribution from Zipf distribution, i.e. popular contents distribution is flat. Reason for this flat distribution for popular content is also filtering of popular requests. R. Doyle et al. call this effect "trickle-down effect".

P2P has also similar flat distribution for popular contents [37]. K.P. Gummadi et al. [37] reveal that reason for this flat distribution of popular content in P2P is different from hierarchical web proxy. Web content is mutable and users have tendency to access many times to popular web site. However, in P2P system, content is immutable and users do not access to a same file after they have downloaded the file, which is the reason for flat distribution of popular content and is called "fetch-at-most-once".

K.P. Gummadi et al. [37] also claim that VoD traffic has similar distribution as P2P. However, H. Yu et al. [38] reveal

†In [34], the authors call Zipf distribution exactly when α is 1 and Zipf-like distribution when α is not 1. However, "Zipf distribution" is widely used even when α is not 1, so we use "Zipf distribution" in this sense, i.e. defined with widespread α.

that popularity of VoD traffic also follows Zipf distribution by analyzing their measured data set in Power Info VoD provided by China Telecom. In many VoD systems, video contents cannot be downloaded and stored at user side, so VoD system has no fetch-at-most-once effect, which leads to multiple accesses to popular contents from the same user and Zipf distribution.

For UGC (User Generated Content), such as YouTube, measured date sets show Zipf distribution of content popularity [39]–[41]. These three papers have different data sets; crawling meta information from YouTube website [39], measured data in academic institute (University of Calgary) [40], and measured in ISP (Orange IP backbone network in France) [41]. Parameter α has different value for each data set; 0.56 for academic institute and 0.682 for ISP. YouTube disallows downloads of Video content, so popular content is viewed multiple times by the same user, which causes Zipf distribution, not showing fetch-at-most-once feature.

According to these published papers on analysis of popularity model, Zipf distribution is quite reasonable model for content popularity for Web content and UGC, both of which are dominant traffic in the current Internet. M. Busari et al. [42] evaluate Web proxy cache performance with several parameters of Zipf popularity distribution. They claim that the steeper Zipf distribution, i.e. the larger value of paramenter α, gives the better hit ratio performance. This means Zipf distribution of content popularity brings good performance for caching network, which gives positive standpoint for caching network in content oriented networks.

Temporal locality of content requests generated by users has been reported in many papers [44]–[46]. S. Travelso et al. [43], [47] analyzed influence of temporal locality to cache performance by using trace-driven simulation. They compared cache size giving the same hit ratio for raw trace and shuffled trace (partially randomized trace). Their results show that shuffled trace requires more cache size, which means temporal locality has great influence to cache performance and gives positive impact towards caching performance. A. Mahanti et al. [48] also showed similar results that empirical temporal locality model gives better cache hit ratio than synthetic randomized model. Almost cache performance evaluated in research papers assumes no temporal locality. Evaluation results by S. Travalso et al. and A. Mahanti et al. show that caching performance evaluated thus far gives lower bound, which means cache performance of content oriented networks should be better in realistic situation.

5. Content Request Routing

For content oriented networks, several new routing protocols have been proposed. In this section, among them, we would like to pick up published papers which treat routing protocol or forwarding strategy taking account of cached content.

FIB is generally assumed to include the shortest path routing towards the repositories. For caching strategy, en-route cache is also widely applied. In en-route cache, downloaded content is cached on the path between the consumer and the repository, called *default path*. Both of content search by content request forwarding and content caching by data packet forwarding are managed only on the default path.

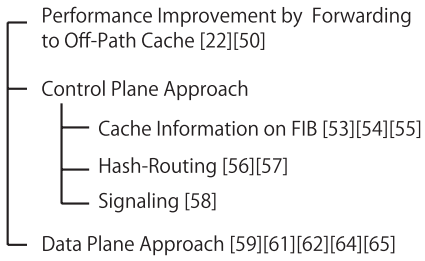

Fig. 5 Content request routing for caching networks.

Several papers discuss about *off-path caching* [22], [49], [50], [66]. These discussions show possibility of significant performance improvement by off-path caching where content request is forwarded to the closest cache which might not be on the default path. According to this discussion of off-path caching possibility, several interesting approaches for content request routing (or forwarding) have been published (Fig. 5).

5.1 Control Plane Approach

For control plane approach, published papers can be categorized into three groups, cache information on FIB, hash-routing and signaling. In the first one, FIB forms entries (or tentative entries) for caching contents. Second one utilizes hash-routing technique [51], [52] for distributing cached contents on total cache storages in ISP. Signaling approach makes use of control packets to form tentative forwarding table for adjacent routers.

[FIB approach]

In this FIB approach, FIB includes entries for cached contents. These entries can be formed by advertisement from cached contents.

Y. Wang et al. [53] proposed a new approach for content request forwarding which takes account of cached contents. Usually, FIB includes only static forwarding information, i.e. content source information for repositories (servers) and not for cached contents. This is because cached contents are transitory and FIB entries for cached content will not work well when stored content is replaced. To resolve this transitoriness of cached content, they proposed hierarchical cache structure where the higher level cache restricts the more content replacements and makes cached content the more stable. Each router advertises its storing content information. This information includes all cached contents and is compressed by Bloom Filter. Advertising area of this Bloom Filter routing information is restricted according to stability of cached content, i.e. level of cache hierarchy. For content request forwarding, when a router receives a content request and hashed value of its content name is matched to a Bloom Filter of advertised contents, it is forwarded to the

corresponding interface. When there is no matched Bloom Filter, this content request is just forwarded to the repository.

M. Lee et al. [54] proposed SCAN. SCAN assumes hybrid situation where a router can speak not only content-oriented forwarding protocol but also current IP routing. Each router advertises its content forwarding information. This information is a merged Bloom Filter of advertised information received from each interface and its cached contents. A content request is forwarded not only to interface (s) having matched Bloom Filter but also to the IP shortest path. Cached contents or the repository reply the content response which is a control packet. When a consumer receives multiple responses, it selects the best content provider to be retrieved from. IP forwarding is just only for fall-back mechanism for cache mis-hit due to replacement or false positive.

S. Lee et al. [55] proposed temporal FIB which contains cached content information. Their proposal is combination of cache decision and forwarding. Their cache decision limits its caching location only at a router on default-path which has the most highest CCV (Cache Capacity Value). CCV indicates how many contents are cached at a corresponding cache in a unit time, which can be interpreted as a kind of cache utilization. In order to utilize selectively cached content, i.e. limited number of cached contents, effectively, they introduced off-path cache routing. An Interest packet collects CCV values of routers along its transmission, and content data transmission can form temporal FIB at each router on the reverse path of Interest packet. This data transmission conveys advertisement information for temporal FIB. When future content request from other users occasionally encounters temporal FIB, they are forwarded to the corresponding interface. This proposal is based on CCN/NDN architecture and makes use of reverse path feature of an Interest packet and the corresponding Data packet[†].

[Hash-Routing approach]

Hash-Routing is a well-known content distribution technique for web-caching [51], [52]. L. Saino et al. [56] and S. Saha et al. [57] proposed hash-routing approach for content oriented networks (not specified to CCN).

L. Saino et al. [56] proposed Hash-Routing for content oriented networks. In their proposal, each content has its pre-designated router to be cached. This designated router is identified by hash function of its content name. This hash function generates N hash values, each of which is assigned to N routers in ISP. So, their approach assumes that information for all routers in ISP is given and their ID is predefined in a centralized manner; their approach can be categorized into control plane approach due to this centralized management. They proposed several data forwarding techniques, symmetric, asymmetric and multicast hash-routing. Their differences are in data download path; symmetric and asymmetric uses data path of reverse path of content request and the shortest path from the repository, respectively. Multicast uses both paths for data download. In symmetric and multicast, downloaded data can go through the designated router, which enables storing locally in its cache.

S. Saha et al. [57] extend the idea of Hash-routing also to Inter-ISP level. Each ISP advertises its designated cache contents and a content request is forwarded to the corresponding designated ISP. By designing external set of designated contents, effective distribution of cached contents among ISPs can be realized.

[Signaling approach]

J.M. Wang et al. [58] proposed content request forwarding by using "summary" information. Each router exchanges its storing cached contents. This exchanged information is called "summary" and is compressed by Bloom Filter. In their approach, only access router, i.e. the first router, can make use of content forwarding to its adjacent router by using "summary". This is because "summary" uses Bloom Filter and might cause cache miss-hit by false positive forwarding. Forwarding of content requests only to an adjacent router of the access router enables easy fall-back mechanism. The authors of this paper [58] describe that their proposed request forwarding is based on CCN/NDN architecture, but this idea can be applied to general content oriented networks.

5.2 Data Plane Approach

In data plane approach, content request forwarding is controlled in data plane without any control plane action. Almost proposals make use of data plane probing where request packet and/or data packet transmission is used for probing the transmission path.

C. Yi et al. [59], [60] proposed Adaptive Forwarding for CCN/NDN. In Adaptive Forwarding, when there are multiple entries for the same content, the best interface is selected from these FIB entries and an incoming Interest is forwarded to this best interface. Only with this operation, when the performance of the selected interface is dynamically changed and is no longer the best one anymore, a router has no way to know this change of situation. To resolve this difficulty of acquiring dynamic network situation, an Interest packet is forwarded to a randomly selected interface from the alternate corresponding FIB entries, and returned data chunk is used for measuring path performance. This "data probing" technique enables adaptive selection of the best interface.

R. Chiocchatti et al. [61] evaluated two forwarding approaches, exploitation and exploration. Exploitation is forwarding by FIB, i.e. static forwarding state for repositories. Exploration is discovery of dynamic content, e.g. cached contents, by using flooding of request packets. For extended exploration approach, they also treated temporal FIB approach (they call "soft-state in FIB") where the best exploited path discovered by exploration is registered in FIB and used for subsequent request packets.

[†]We explicitly describe its specific platform when proposed content routing, caching decision, and replacement policy assume some specific architecture, e.g. CCN/NDN. Unless otherwise noted, all works surveyed in this paper assume general content oriented networks.

Same authors [62] also proposed an extended exploration approach, called INFORM. INFORM is based on NDN/CCN. In INFORM, exploitation and exploration phase are alternately repeated. In exploration phase, reinforcement learning of Q-leaning is applied to obtain the best path (interface) [63]. In this phase, an Interest packet is forwarded towards not only the best selected interface obtained in the last exploration phase but also a randomly selected alternative interface which is used for refreshing Q-value. The reason for forwarding also towards the current best path is that there is a possibility of loss of the randomly forwarded packet due to cache miss-hit.

S. Wang et al. proposed a similar idea of limited flooding of content request, Forwarding with Shallow Flooding (FSF) [64]. In FSF, a content request is flooded towards all other interfaces than the shortest path. Flooding area of these flooded content requests is limited to a specific depth.

Congestion-aware caching [65] is an interesting approach for caching decision and is also explained in detail in the next section of caching decision. In congestion-aware caching, congestion-aware search is proposed for content request forwarding. A content request is broadcast in a limited region by flooding and a control packet is returned when the requested content is found. When multiple content sources including cached content are found, the consumer selects the best content source with taking account of content retrieval throughput which is mainly regulated by congestion on the bottleneck link. This content routing finds the cached content providing the best throughput.

6. Caching Decision for Caching Networks

In caching networks, caching decision, whether a passing content is to be cached or not, at each router is a very important issue which governs caching performance. There have been proposed many interesting approaches for caching decision (Fig. 6).

6.1 Technical Problem for TERC

The most simple caching decision is TERC (Transparent En-Route Caches). In TERC, contents are cached at every router along the download path from the server to a requesting user.

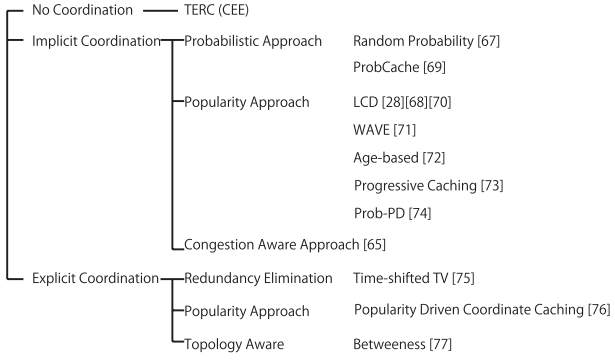

Fig. 6 Caching decision in caching networks.

This strategy is also called "Cache Everything Everywhere" in many literatures. A content request is generally transmitted towards the content server along the shortest path. When content data is downloaded on the shortest path, every cache on this path, called the default path, stores this content data. TERC is very simple and requires no explicit or implicit coordination among content routers. So, many published papers, e.g. [21], [66], employ this strategy for caching decision.

In TERC, downloaded content is stored at local cache on all routers along the download path. From this feature, it has the following technical problems.

1. Same content is redundantly cached along the default path.
   Same content is widely distributed along a default path. This might cause adjacent routers have similar contents, which leads to inefficient resource usage among content routers. When stored contents at close routers are managed to be different, redundancy of stored contents is eliminated and cache storage capacity is efficiently used, which might improve total performance of caching networks.
2. Unpopular content is not discriminated against popular one.
   Even when unpopular content passes through a content router, it stores this content, which might cause eviction of stored popular content. In TERC, content replacement rate is generally high and popular content cannot be stored in a stable fashion, especially at a router close to the content server.

W.K. Chai et al. [77] comparatively evaluate TERC and random caching where content chunk is locally stored only at randomly selected single cache on the content download path. This random selection of caching place means somewhat no policy, but performance evaluation in [77] shows this simple caching decision brings performance improvement from the viewpoint of server load and the number of hops for content download. This simple evaluation result intuitively triggers more sophisticated caching decision which is expected to bring more performance improvement.

Along these research trends, to improve cache performance from two technical viewpoints listed above, many interesting approaches have been proposed (Fig. 6). These can be classified into two categories, implicit coordination and explicit coordination.

6.2 Implicit Coordination

In caching decision categorized in implicit coordination approach, each cache is managed in a distributed fashion, i.e. it is independently controlled. For the first technical problem of redundancy of cached content, several probabilistic approaches have been proposed as explained in the first part of this subsection of probabilistic approach. To resolve the second technical problem related to content popularity, lots of papers have been published and interesting approaches are

explained in detail in the latter part of popularity approach. In the last part of this subsection, we explain about another interesting approach of congestion aware caching decision.

[Probabilistic Approach]

S. Arianfar et al. [67] describe random content placement with fixed probability at each router along the download path. They call this caching decision "random autonomous caching". Similar approach has also been picked up as performance comparison method in web caching analysis (e.g. $Prob(p)$ in [68]). Several papers call this caching decision strategy $Fix(p)$ decision where p is the probability of storing the content in local cache.

In $Fix(p)$ decision, probability of caching is identical for all routers along the download path. I. Psaras et al. [69] propose ProbCache where caching probability at a router is calculated dependently on its location in the download path. In ProbCache, probability that a content router caches an incoming chunk, is calculated as follows.

$$ProbCache = TimesIn \times CacheWeight, \quad (2)$$

where $TimesIn$ is estimated caching capacity of the path and $CacheWeight$ is a weight which increases with getting closer to the user. $TimesIn$ estimates the number of times that the remained default path can cache the chunk. $CacheWeight$ is the ratio of the number of total hops of the path from the content source to the user to the number of hops of the remained path. In ProbCache, probability of caching an incoming chunk increases with getting close to the user. When content is cached close to the user, content retrieval delay is expected to be improved.

These two simple approaches are frequently chosen as basic performance of caching decision [32], [71], in performance comparison in many papers.

[Popularity Approach]

Probabilistic approach described above can distribute stored content even with simple cache management, but it does not care about popularity of content. This means arrival of unpopular content might cause eviction of stored popular content. Several interesting caching decisions of distributed (implicit coordination) popularity approach have been proposed.

LCD (Leave Copy Down) [28], [68], [70] can effectively distribute contents according to their popularity even with simple mechanism. In LCD, cache hit of a content chunk enables storing this content chunk at local cache of a one-hop downward content router (Fig. 7). Popular content will have many content requests, which brings cached content closer to the user with LCD caching decision. And rather unpopular content can be stayed at core location of a network because content requests for popular contents can be filtered by cached contents closer to the users. In the literatures describing LCD, there is no explicit way to implement LCD, but LCD can be implemented with simple mechanism, for example, just one bit notification of cache hit at a higher level cache, i.e. one hop upward router on a default path.

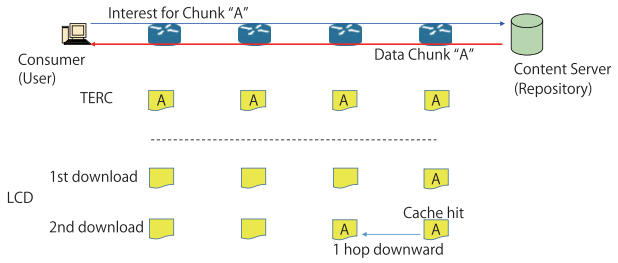

Fig. 7 LCD (Leave Copy Down).

LCD requires coordination between only two adjacent content routers and obtains popularity-based content location even with this simple mechanism.

In LCD, content file download causes all chunks of this content file will be stored at the one hop downward router. WAVE [71] realizes more moderate (gentle) distribution of popular content. It makes use of marking mechanism at each content router. Each router and the original server have a capability of marking data chunks. Only marked chunk can be cached at each router. When content file composed of several chunks is initially downloaded from the original server, the server marks only the first chunk. When this chunk arrives at the first content router, it will be cached there and its mark is removed. So, this chunk is never cached afterward on the download path. When this content file is downloaded for the second time from the original server, two new chunks are marked. At the first router, when the first chunk of a content file arrives again, it is marked here and is forwarded to the next content router. When the second and third chunks (marked at the original router) arrives, these are cached locally and their marks are removed. The number of marked chunks increases exponentially with the number of downloads. By these distributed marking operation, popular content is distributed around the network very fast. However its content distribution intensity is rather lower than LCD because WAVE only allows marked chunks to be cached at the downward router.

Age-based Cooperative Caching [72] makes use of Age-based mechanism for widely distributing popular content. In Age-based caching, when a content chunk arrives, a stored content chunk is replaced with an arrived chunk at a content router if and only if age of a stored chunk is expired and cache is full. Age of a content chunk is initially computed at the first content router as an initial value and assigned initial age is increased with multiplied with weight at each downward content router. This means cached content at the closer content router to the user has the larger age initially. So, in a steady state, the more popular content is expected to be located the closer to users.

These approaches can expand popular content widely in the network. When popular content is located close to the user, content request for popular content is filtered by these cached content and rather unpopular content can be cached in core location of the networks. However, in these approaches, cached content is managed without taking account of cache location, so there might be a possibility of

redundantly wide distribution of popular content. To resolve this technical problem, caching decision taking not only popularity of content but also cache location into account has been proposed.

Progressive Caching [73] is caching policy extended from LCD. So, a content chunk is cached at a local cache of a router located one-hop downwards from a cache-hit router. Progressive Caching is specifically proposed for CCN/NDN. In addition to this LCD caching, a content chunk is moreover cached at a content router according to the following conditions. For this additional caching policy, progressive caching has different policies for edge routers and intermediate routers. At intermediate routers, a content chunk passing through a router is cached locally when the number of PIT entries for a corresponding content chunk is larger than or equal to threshold θ_1. This is because the content is expected to be popular enough to be cached with sufficient number of PIT entries (PIT aggregation). At edge routers, a content chunk is cached when the number of generated content requests is larger than or equal to threshold θ_2. This is also because a corresponding data chunk is estimated to be sufficiently popular. With this edge caching decision, popular contents are to be cached earlier than LCD because LCD requires cache hits almost equal to the hop-number of default path for edge caching. Progressive caching also proposes an interesting cache replacement policy for intermediate and edge routers, which is described in Sect. 7.

A. Ioannou et al. [74] proposed Prob-PD which is basically a probabilistic approach and takes both content popularity (P) and node location (D:distance from the source) into account. Probability that node i caches a content chunk of content j is calculated by the following equation.

$$Prob - PD_{i,j} = (\text{measured popularity of content} j) \times \frac{d_{i,src}}{d_{dst,src}}, \quad (3)$$

where $d_{dst,src}$ and $d_{i,src}$ denote distance from the content source to the content destination (the source of interest) and to a node calculating this probability, respectively. With Prob-PD, the more popular content has the higher probability to be cached and the closer node to the content request node has the higher probability of caching.

[Congestion Aware]

Almost proposed cache decisions have been designed to obtain good performance from network operator's perspective, e.g. cache hit ratio. This network-centric performance has, of course, implicit relationship to user-perceived performance, e.g. content retrieval throughput and delay. Recently, an interesting approach having explicit and strong relationship with user-perceived performance, congestion-aware caching [65] has been proposed. In congestion-aware caching, user-perceived performance of content retrieval delay is a key factor for caching decision. Each router measures the number of flows passing through it and calculates its locally measured popularity of contents. This measured number of active flows is also used for local calculation

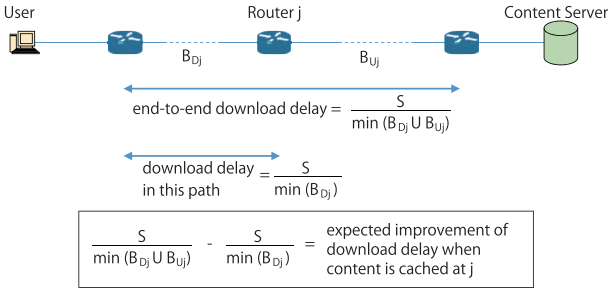

Fig. 8 Congestion aware caching decision.

of fair-shared bandwidth of each link. Each request and data packet brings the running minimum value of available fair-shared link bandwidth for all links that it is transferred over. This bandwidth information is brought in a piggyback manner. Utility function for caching decision at each node (router) is calculated as follows.

$$U_{i,j} = \left\{ \frac{S}{min(B_{Dj} \cup B_{Uj})} - \frac{S}{min(B_{Dj})} \right\} * P_{ij}, \quad (4)$$

where S is content file size, B_{Dj} and B_{Uj} are a set of available bandwidths of respective downstream and upstream link set for node j (Fig. 8), and P_{ij} is popularity of content i locally measured at node j. Equation 4 calculates expected improvement of content download delay when content is cached at node j (Fig. 8). When a data chunk of content i is going through router j, this chunk is cached only when calculated utility function of content i is larger than the minimum value of utility function of current cached contents. So, congestion-aware caching decision has a tendency to cache a content more popular and bringing more improvement for content download delay. This congestion aware approach is actually an integrated approach of not only caching decision, but also replacement policy and interest forwarding, as well.

6.3 Explicit Coordination

In implicit coordination described in the last section, each node is controlled in a distributed manner. On the contrary, in explicit coordination, caching decision in each node is managed with explicit coordination with other nodes.

[Redundancy Elimination]

Z. Li et al. [75] proposed explicitly coordinated caching which enables coordinated content chunk localization. Their proposed caching decision is based on CCN/NDN architecture. Each content router cooperates with its $k-1$ nearest routers. At initial stage, each router is labeled with integer in $[0, k-1]$ so that no its $k-1$ nearest routers have the same label. And each router is assumed to know these $k-1$ routers and their labels. CCN router has two additional tables, Collaborative Router Table (CRT) and Collaborative Content Store (CCS). When a data chunk arrives at a router, it checks whether modulo k of a chunk name is equal to its label. When this check is true, a data chunk is forwarded to the PIT entry as normal CCN router operation. Otherwise, a

router additionally forwards a data chunk to its nearest router in CRT having a matched label and adds this content chunk information in CCS. This forwarded router stores data chunk in its local cache. By these caching managements, k routers cooperatively store exclusive $\frac{1}{k}$ part of content chunks.

When an interest packet arrives, a router checks CS and PIT entry as normal CCN operation. When these entries have no exact match to an interest packet, in normal CCN, FIB entry is then checked. In this coordinated caching, CCS entry is checked prior to FIB. And an interest packet is forwarded to the nearest router having matched CCS entry, i.e. a router having the matched label (molule k). This forwarding strategy is categorized into off-path caching (Sect. 5), which fundamentally requires fall-back mechanism when cache miss-hit occurs. This cooperative caching proposes piggy-back transmission of control information for supporting CCS consistency which prevents cache miss-hit.

[Popularity Approach]

J. Li et al. proposed popularity-driven coordinated caching [76] which aims at minimizing inter-ISP traffic and user's average access latency. Their proposed caching decision specifically assumes CCN/NDN. They propose two caching algorithms, TopDown Caching and AsympOpt Caching. In TopDown Caching, the more popular content is located the closer to the gateway, i.e. the further from the user. In AsympOpt Caching, on the contrary, the more popular content is located the closer to users. According to their performance evaluation, cache hit ratio has similar good performance, i.e. both of them realizes reduction of inter-ISP traffic, and AsympOpt has better access latency performance. In these algorithms, popularity of contents is measured with coordination among routers and caching decision is based on network topology, which means explicit coordination is fundamentally required.

[Topology Aware]

W. Chai et al. proposed caching decision, based on topology information, especially on network centrality of betweeness [77]. Betweeness [78], [79] of a network node is one of the network centralities and is defined as the number of total passing shortest paths between all pairs of nodes. Intuitively, when betweeness of a network node is high, larger number of content requests will go through this node and higher cache hit rate is to be expected. In their proposed caching decision, called "*Betw*" in their paper, an interest packet records the highest value of betweeness among intermediate nodes along its transmission path. Content data packet stores this highest betweeness value and only a node having the same betweeness can store this content packet at its local cache. *Betw* requires all network nodes (routers) know their betweeness, which requires explicit coordination or centralized approach to obtain this whole network topology information.

7. Cache Replacement

As G.Rossini and D.Rossi described in their interesting paper of performance evaluation of coupling caching and forwarding [32], cache replacement plays limited role for performance improvement. However, there have been published lots of papers concerning cache replacement. This technical problem has been also treated widely in web caching. For web caching, lots of proposals for cache replacement have been published [80]. In survey paper by S. Podlipnig et al. [80], 11 proposals were explained even only for recency-based approach. They categorized recency-based, frequency-based, recency/frequency-based, function-based, and randomized approaches, each of which includes many proposals.

In many papers of content oriented networks, LRU (Least Recently Used) is widely used for cache replacement policy. This is because LRU is rather simple to be implemented. Several proposals for cache replacement have been published, which include Progressive Caching [73] and Congestion-Aware Caching [65]. These two papers also propose a new caching decision approaches, which are explained in detail in Sect. 6.

In Progressive Caching, at an edge node, hop-count to caching content is taken into account for cache replacement policy. Cache is divided into classes. When chunk is stored in cache, its class is set to the number of hops towards the corresponding cached content. And a content in the lowest class is evicted (replaced). In this eviction of cached content, class value for all the other contents are decremented by the class value of this evicted content. When a cached content is an old one, it is likely to be removed because its class value might be decremented many times by evictions of other contents. And a cached content whose original content is located closer, might has small original class value and is also likely to be removed.

In Congestion-Aware Caching, for replacement policy, the content chunk having the minimum value of utility function is replaced with the arrival chunk with larger utility function value. Utility function takes account popularity and improvement of content download delay. When a content arriving at a router brings more improvement for content download delay, which means this content has already passed the congested link, this content is likely to be cached with this Congestion-Aware Caching. A content is likely to be replaced when it brings less improvement, which means it has no significant congested link between this router and the original repository.

These two replacement policy will bring performance improvement for cache network because contents with less improvement (of content download hop in Progressive Caching and of content download delay in Congestion-Aware Caching) is likely to be evicted by cache replacement.

8. Conclusions

In this paper, we survey interesting research activities for caching networks in content oriented networks. In content oriented networks, cache storage is pervasively distributed inside a network. These tangled caches in a network make it very difficult to manage caching network effectively. In order to manage this complicated caching networks, not only technical issues surveyed in this paper, but also other important technical issues, such as, traffic engineering and congestion control are carefully designed integrally with taking account cached contents. And security is also significantly important technical problem for cached content. To understand good combination of these complicated factors for designing caching network, performance evaluation of caching networks is also one of the most important technical issues. Due to space limitations, we cannot discuss about analytical method for caching networks, but almost analytical papers make simple assumption of independent reference model (IRM) [81]. Performance analysis for more complicated caching network model is also left for further research directions.

As surveyed in this paper, researches from many aspects including performance evaluation, content request routing, caching decision, and so on for caching networks are now being progressed around the world. Several works for implementation issues of caching storage have also published. Standardization activities for content oriented networks also started in both of ITU-T and IRTF. Caching networks will bring benefits not only for end-users but also network providers. For end-users, it improves user-perceived performance, e.g. content retrieval delay. For network providers, it reduces total network traffic. We believe that caching networks bringing significant benefits for both of network consumers and providers will continue to play an important role for future network research.

Acknowledgements

This work is partly supported by JSPS KAKENHI Grant Number 26280034.

References

[1] Cisco Virtual Networking Index, The Zettabyte Era — Trends and Analysis http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html
[2] F. Douglis and M.F. Kaashoek, "Scalable internet services," IEEE Internet Comput., vol.5, no.4, pp.36–37, July 2001.
[3] E.K. Lua, J. Crowcroft, M. Pias, R. Sharma, and S. Lim, "A survey and comparison of peer-to-peer overlay network schemes," IEEE Commun. Surv. Tutorials, vol.7, no.2, pp.72–93, 2005.
[4] A. Chankhunthod, P.B. Danzig, C. Neerdaels, M.F. Schwartz, and K.J. Worrell, "A hierarchical Internet object cache," 1996 USENIX Technical Conference, pp.153–163, San Diego, USA, Jan. 1996.
[5] P. Rodriguez, C. Spanner, and E.W. Biersack, "Analysis of Web caching architectures: Hierarchical and distributed caching," IEEE/ACM Trans. Netw., vol.9, no.4, pp.404–418, Aug. 2001.
[6] M. Gritter and D. Chariton, "TRIAD: A new next-generation Internet architecture," http://www-dsg.stanford.edu/triad, July 2000.
[7] I. Stoica, D. Adkins, S. Zhuang, S. Shenker, and S. Surana, "Internet indirection infrastructure," SIGCOMM Comput. Commun. Rev., vol.32, no.4, pp.73–86, Oct. 2002.
[8] T. Koponen, M. Chawla, B.-G. Chun, A. Ermolinskiy, K.H. Kim, S. Shenker, and I. Stoica, "A data-oriented (and beyond) network architecture," SIGCOMM Comput. Commun. Rev., vol.37, no.4, pp.181–192, Aug. 2007.
[9] N. Fotiou, D. Trossen, and G.C. Polyzos, "Illustrating a publish-subscribe Internet architecture," Telecommun. Syst., vol.51, no.4, pp.233–245, Dec. 2012.
[10] PURSUIT project website, http://www.fp7-pursuit.eu/PursuitWeb/
[11] N. Fotiou, P. Nikander, D. Trossen, and G.C. Polyzos, "Developing information networking further: From PSIRP to PURSUIT," Broadband Communications, Networks, and Systems, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol.66, pp.1–13, 2012.
[12] Scalable and Adaptive Internet Solutions (SAIL) webpage, http://www.sail-project.eu/
[13] The FP7 4WARD Project webpage, http://www.4ward-project.eu/
[14] Network of Information webpage, http://www.netinf.org/
[15] C. Dannewitz, D. Kutscher, B. Ohlman, S. Farrell, B. Ahlgren, and H. Karl, "Network of information (NetInf) — An information-centric networking architecture," Comput. Commun., vol.36, no.7, pp.721–735, April 2013.
[16] MobilityFirst Future Internet Architecture Project, http://mobilityfirst.winlab.rutgers.edu/
[17] A. Venkataramani, J.F. Kurose, D. Raychaudhuri, K. Nagaraja, M. Mao, and S. Banerjee, "MobilityFirst: A mobility-centric and trustworthy internet architecture," SIGCOMM Comput. Commun. Rev., vol.44, no.3, pp.74–80, July 2014.
[18] L. Zhang, D. Estrin, J. Burke, V. Jacobson, J. Thornton, D. Smetters, B. Zhang, G. Tsudik, K. Claffy, D. Krioukov, D. Massey, C. Papadopoulos, T. Abdelzaher, L. Wang, P. Crowley, and E. Yeh: "Named data networking (NDN) project," PARC Technical Report 2010-003, Oct. 2010. http://named-date.net/ndn-proj.pdf
[19] V. Jacobson, D.K. Smetters, J.D. Thornton, M.F. Plass, N.H. Briggs, and R.L. Braynard, "Networking named content," Proc. 5th International Conference on Emerging Networking Experiments and Technologies, CoNEXT'09, pp.1–12, 2009.
[20] Named Data Networking website, http://named-data.net/
[21] S.K. Fayazbakhsh, Y. Lin, A. Tootoonchian, A. Ghodsi, T. Koponen, B. Maggs, K.C. Ng, V. Sekar, and S. Shenker, "Less pain, most of the gain: Incrementally deployable ICN," Proc. ACM SIGCOMM 2013 Conference on SIGCOMM, SIGCOMM'13, pp.147–158, 2013.
[22] A. Ghodsi, S. Shenker, T. Koponen, A. Singla, B. Raghavan, and J. Wilcox, "Information-centric networking: Seeing the forest for the trees," Proc. 10th ACM Workshop on Hot Topics in Networks, HotNets'11, pp.1–6, 2011.
[23] G. Tyson, S. Kaune, S. Miles, Y. El-khatib, A. Mauthe, and A. Taweel, "A trace-driven analysis of caching in content-centric networks," 2012 21st International Conference on Computer Communications and Networks (ICCCN'12), pp.1–7, 2012.
[24] L. Fan, P. Cao, J. Almeida, and A.Z. Broder, "Summary cache: A scalable wide-area Web cache sharing protocol," SIGCOMM Comput. Commun. Rev., vol.28, no.4, pp.254–265, Oct. 1998.
[25] S. Michel, K. Nguyen, A. Rosenstein, L. Zhang, S. Floyd, and V. Jacobson, "Adaptive web caching: Towards a new global caching architecture," Comput. Netw. ISDN Syst., vol.30, no.22-23, pp.2169–2177, Nov. 1998.
[26] B.D. Davison, "A Web caching primer," IEEE Internet Comput., vol.5, no.4, pp.38–45, July 2001.
[27] J. Wang, "A survey of Web caching schemes for the Internet," SIGCOMM Comput. Commun. Rev., vol.29, no.5, pp.36–46, Oct. 1999.
[28] N. Laoutaris, S. Syntila, and I. Stavrakakis, "Meta algorithms for

hierarchical Web caches," IEEE International Conference on Performance, Computing, and Communications, 2004, pp.445–452, 2004.
[29] P.B. Danzig, R.S. Hall, and M.F. Schwartz, "A case for caching file objects inside internetworks," SIGCOMM Comput. Commun. Rev., vol.23, no.4, pp.239–248, Sept. 1993.
[30] A. Wolman, M. Voelker, N. Sharma, N. Cardwell, A. Karlin, and H.M. Levy, "On the scale and performance of cooperative Web proxy caching," Proc. 17th ACM Symposium on Operating Systems Principles, SOSP'99, pp.16–31, 1999.
[31] Y. Wang, Z. Li, G. Tyson, S. Uhlig, and G. Xie, "Optimal cache allocation for content-centric networking," 2013 21st IEEE International Conference on Network Protocols (ICNP), pp.1–10, 2013.
[32] G. Rossini and D. Rossi, "Coupling caching and forwarding: Benefits, analysis, and implementation," Proc. ACM 1st International Conference on Information-Centric Networking, ICN'14, pp.127–136, 2014.
[33] K. Zipf, Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology, Addison-Wesley Press, Cambridge, Mass., 1949.
[34] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: Evidence and implications," Proc. IEEE INFOCOM'99. Conference on Computer Communications, pp.126–134, 1999.
[35] A. Mahanti, C. Williamson, and D. Eager, "Traffic analysis of a Web proxy caching hierarchy," IEEE Network, vol.14, no.3, pp.16–23, May 2000.
[36] R.P. Doyle, J.S. Chase, S. Gadde, and A.M. Vahdat, "The trickle-down effect: Web caching and server request distribution," Comput. Commun., vol.25, no.4, pp.345–356, March 2002.
[37] K.P. Gummadi, R.J. Dunn, S. Saroiu, S.D. Gribble, H.M. Levy, and J. Zahorjan, "Measurement, modeling, and analysis of a peer-to-peer file-sharing workload," Proc. 19th ACM Symposium on Operating Systems Principles, SOSP'03, pp.314–329, 2003.
[38] H. Yu, D. Zheng, B.Y. Zhao, and W. Zheng, "Understanding user behavior in large-scale video-on-demand systems," SIGOPS Oper. Syst. Rev., vol.40, no.4, pp.333–344, 2006.
[39] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes," Proc. 7th ACM SIGCOMM Conference on Internet Measurement, IMC'07, pp.1–113, 2007.
[40] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "YouTube traffic characterization," Proc. 7th ACM SIGCOMM Conference on Internet Measurement, IMC'07, pp.15–28, 2007.
[41] F. Guillemin, B. Kauffmann, S. Moteau, and A. Simonian, "Experimental analysis of caching efficiency for YouTube traffic in an ISP network," Proc. 2013 25th International Teletraffic Congress (ITC), pp.1–9, 2013.
[42] M. Busari and C. Williamson, "On the sensitivity of Web proxy cache performance to workload characteristics," Proc. IEEE INFOCOM 2001, Conference on Computer Communications, pp.1225–1234, 2001.
[43] S. Traverso, M. Ahmed, M. Garetto, P. Giaccone, E. Leonardi, and S. Niccolini, "Temporal locality in today's content caching: Why it matters and how to model it," SIGCOMM Comput. Commun. Rev., vol.43, no.5, pp.5–12, 2013.
[44] S. Jin and A. Bestavros, "Sources and characteristics of Web temporal locality," Proc. 8th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, pp.28–35, 2000.
[45] V. Almeida, A. Bestavros, M. Crovella, and A. de Oliveira, "Characterizing reference locality in the WWW," 4th International Conference on Parallel and Distributed Information Systems, pp.92–103, 1996.
[46] R. Fonseca, V. Almeida, M. Crovella, and B. Abrahao, "On the intrinsic locality properties of Web reference streams," IEEE INFOCOM 2003, pp.448–458, 2003.
[47] S. Traverso, M. Ahmed, M. Garetto, P. Giaccone, E. Leonardi, and S. Niccolini, "Unravelling the impact of temporal and geographical locality in content caching systems," IEEE Trans. Multimedia, vol.17, no.10, pp.1839–1854, Oct. 2015.
[48] A. Mahanti, D. Eager, and C. Williamson, "Temporal locality and its impact on Web proxy cache performance," Performance Evaluation, vol.42, no.2-3, pp.187–203, Sept. 2000.
[49] M. Draxler and H. Karl, "Efficiency of on-path and off-path caching strategies in information centric networks," 2012 IEEE International Conference on Green Computing and Communications, pp.581–587, 2012.
[50] G. Rossini and D. Rossi, "Evaluating CCN multi-path interest forwarding strategies," Comput. Commun., vol.36, no.7, pp.771–778, April 2013.
[51] K.W. Ross, "Hash routing for collections of shared Web caches," IEEE Netw., vol.11, no.6, pp.37–44, Nov. 1997.
[52] D.G. Thaler and C.V. Ravishankar, "Using name-based mappings to increase hit rates," IEEE/ACM Trans. Netw., vol.6, no.1, pp.1–14, Feb. 1998.
[53] Y. Wang, K. Lee, B. Venkataraman, R.L. Shamanna, I. Rhee, and S. Yang, "Advertising cached contents in the control plane: Necessity and feasibility," 2012 Proc. IEEE INFOCOM Workshops, pp.286–291, 2012.
[54] M. Lee, K. Cho, K. Park, T. Kwon, and Y. Choi, "SCAN: Scalable content routing for content-aware networking," 2011 IEEE International Conference on Communications (ICC), pp.1–5, 2011.
[55] S.-W. Lee, D. Kim, Y.-B. Ko, J.-H. Kim, and M.-W. Jang, "Cache capacity-aware CCN: Selective caching and cache-aware routing," 2013 IEEE Global Communications Conference (GLOBECOM), pp.2114–2119, 2013.
[56] L. Saino, I. Psaras, and G. Pavlou, "Hash-routing schemes for information centric networking," Proc. 3rd ACM Workshop on Information-Centric Networking, ICN'13, pp.27–32, 2013.
[57] S. Saha, A. Lukyanenko, and A. Yla-Jaaski, "Cooperative caching through routing control in information-centric networks," 2013 Proc. IEEE INFOCOM, pp.100–104, 2013.
[58] J.M. Wang, J. Zhang, and B. Bensaou, "Intra-AS cooperative caching for content-centric networks," Proc. 3rd ACM Workshop on Information-Centric Networking, ICN'13, pp.61–66, 2013.
[59] C. Yi, A. Afanasyev, L. Wang, B. Zhang, and L. Zhang, "Adaptive forwarding in named data networking," SIGCOMM Comput. Commun. Rev., vol.42, no.3, pp.62–67, July 2012.
[60] C. Yi, A. Afanasyev, I. Moiseenko, L. Wang, B. Zhang, and L. Zhang, "A case for stateful forwarding plane," Comput. Commun., vol.36, no.7, pp.779–791, April 2013.
[61] R. Chiocchetti, D. Rossi, G. Rossini, G. Carofiglio, and D. Perino, "Exploit the known or explore the unknown?: Hamlet-like doubts in ICN," Proc. ACM 2nd Workshop on Information-Centric Networking, ICN'12, pp.7–12, 2012.
[62] R. Chiocchetti, D. Perino, G. Carofiglio, D. Rossi, and G. Rossini, "INFORM: A dynamic interest forwarding mechanism for information centric networking," Proc. 3rd ACM Workshop on Information-Centric Networking, ICN'13, pp.9–14, 2013.
[63] J.A. Boyan and M.L. Littman, "Packet routing in dynamically changing networks: A reinforcement learning approach," in Advances in Neural Information Processing Systems 6, pp.671–678, Morgan Kaufmann, 1994.
[64] S. Wang, J. Bi, J. Wu, Z. Li, W. Zhang, and X. Yang, "Could in-network caching benefit information-centric networking?," Proc. 7th Asian Internet Engineering Conference on AINTEC'11, pp.112–115, 2011.
[65] M. Badov, A. Seetharam, J. Kurose, V. Firoiu, and S. Nanda, "Congestion-aware caching and search in information-centric networks," Proc. ACM 1st International Conference on Information-Centric Networking, ICN'14, pp.37–46, 2014.
[66] E.J. Rosensweig and J. Kurose, "Breadcrumbs: Efficient, best-effort content location in cache networks," IEEE INFOCOM 2009, The 28th Conference on Computer Communications, pp.2631–2635, 2009.

[67] S. Arianfar, P. Nikander, and J. Ott, "On content-centric router design and implications," Proc. Re-Architecting the Internet Workshop on, ReARCH'10, pp.1–6, 2010.
[68] N. Laoutaris, H. Che, and I. Stavrakakis, "The LCD interconnection of LRU caches and its analysis," Performance Evaluation, vol.63, no.7, pp.609–634, July 2006.
[69] I. Psaras, W.K. Chai, and G. Pavlou, "Probabilistic in-network caching for information-centric networks," Proc. ACM Second Workshop on Information-Centric Networking, ICN'12, pp.55–60, 2012.
[70] V. Martina, M. Garetto, and E. Leonardi, "A unified approach to the performance analysis of caching systems," IEEE INFOCOM 2014, IEEE Conference on Computer Communications, pp.2040–2048, 2014.
[71] K. Cho, M. Lee, K. Park, T.T. Kwon, Y. Choi, and S. Pack, "WAVE: Popularity-based and collaborative in-network caching for content-oriented networks," 2012 Proc. IEEE INFOCOM Workshops, pp.316–321, 2012.
[72] Z. Ming, M. Xu, and D. Wang, "Age-based cooperative caching in information-centric networks," 2012 Proc. IEEE INFOCOM Workshops, pp.268–273, 2012.
[73] J.M. Wang and B. Bensaou, "Progressive caching in CCN," 2012 IEEE Global Communications Conference (GLOBECOM), pp.2727–2732, 2012.
[74] A. Ioannou and S. Weber, "Towards on-path caching alternatives in information-centric networks," 39th Annual IEEE Conference on Local Computer Networks, pp.362–365, 2014.
[75] Z. Li and G. Simon, "Time-shifted in content centric networks: The case for cooperative in-network caching," 2011 IEEE International Conference on Communications (ICC), pp.1–6, 2011.
[76] J. Li, H. Wu, B. Liu, J. Lu, Y. Wang, X. Wang, Y. Zhang, and L. Dong, "Popularity-driven coordinated caching in named data networking," Proc. 8th ACM/IEEE Symposium on Architectures for Networking and Communications systems, ANCS'12, pp.15–26, 2012.
[77] W.K. Chai, D. He, I. Psaras, and G. Pavlou, "Cache "less for more" in information-centric networks," Comput. Commun., vol.36, no.7, pp.758–770, April 2013.
[78] L.C. Freeman, "Centrality in social networks conceptual clarification," Soc. Netw., vol.1, no.3, pp.215–239, 1978–1979.
[79] V. Latora and M. Marchiori, "A measure of centrality based on network efficiency," New J. Phys., vol.9, no.6, pp.188–188, June 2007.
[80] S. Podlipnig and L. Böszörmenyi, "A survey of Web cache replacement strategies," ACM Comput. Surv., vol.35, no.4, pp.374–398, Dec. 2003.
[81] J. Kurose, "Information-centric networking: The evolution from circuits to packets to content," Comput. Netw., vol.66, pp.112–120, June 2014.

Miki Yamamoto received his B.E., M.E., and Ph.D. in communications engineering from Osaka University in 1983, 1985, and 1988. He joined the Department of Communications Engineering at Osaka University in 1988. He moved to the Department of Electrical Engineering and Computer Science of Kansai University in 2005, where he is a professor. He visited the University of Massachusetts at Amherst in 1995 and 1996 as a visiting professor. His research interests include content oriented networks, high-speed networks, wireless networks, and the evaluation of performance of these systems. He is a member of IEEE, ACM, and IPSJ.