ARTICLE IN PRESS

Computer Networks xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

A content-based publish/subscribe framework for large-scale content delivery

Mohamed Diallo [a], Vasilis Sourlas [b,*], Paris Flegkas [b], Serge Fdida [a], Leandros Tassiulas [b]

[a] UPMC Sorbonne Universités, Paris, France
[b] University of Thessaly & CERTH-ITI, Greece

ARTICLE INFO

Article history:
Received 11 April 2012
Received in revised form 28 September 2012
Accepted 20 November 2012
Available online xxxx

Keywords:
Content-based publish/subscribe networking
Large-scale content delivery
Caching
Information overload

ABSTRACT

The publish/subscribe communication paradigm has become an important architectural style for designing distributed systems and has recently been considered one of the most promising future network architectures that solves many challenges of content delivery in the current Internet. This work is concerned with scaling decentralized content-based publish/subscribe (CBPS) networks for large-scale content distribution. A fundamental step for CBPS networks to reach the large-scale is to move from the current exhaustive filtering service model, where a subscription selects every relevant publication, to a service model capturing the quantitative and qualitative heterogeneity of information consumers requirements. Moreover, the proposed work aims at leveraging caching for increasing the communication efficiency of CBPS operating at large-scale characterized by widely spread information consumers with heterogeneous requirements, large number of publications and scarcity of end-to-end bandwidth. We propose and design a service model for addressing the consumers' requirements for content-based information retrieval and describe the relevant protocols necessary to implement such a service. We evaluate the proposed approach, by using realistic workload scenarios and comparing different content and interest forwarding strategies as well as caching policies in terms of resource efficiency and user perceived QoS metrics.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Content-centric publish/subscribe (pub/sub) networking is a flexible communication model that meets the requirements of the content distribution trends in the Internet. Pub/sub networking shifts the communication paradigm of the internetworking layer from endpoints to information, where information is addressed by semantic attributes rather than origin and destination identities. The pub/sub architecture enables information consumers to register their information interests to a mediation entity that will prospectively retrieve content relevant to those interests.

There exist different pub/sub models which are classified according to the semantic of the subscription language. *Channel-based* pub/sub allows information consumers to subscribe to publications originating from specific channels or feeds similarly to *ATOM* and *RSS* standards. *Topic-based* pub/sub enables information consumers to register to a set of predefined topics, while *content-based* pub/sub (CBPS) supports subscriptions following an attribute/value schema or defined as vectors of keywords. In the CBPS communication model, the mediation between information providers and consumers is realized at the network level in a flexible, decentralized and loose coupled manner. CBPS provides efficient dissemination channels for a wide range of applications such as event notification, news dissemination or file sharing services.

* Corresponding author.
E-mail addresses: mohamed.diallo@lip6.fr (M. Diallo), vsourlas@uth.gr (V. Sourlas), pflegkas@uth.gr (P. Flegkas), serge.fdida@lip6.fr (S. Fdida), leandros@uth.gr (L. Tassiulas).

1389-1286/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.comnet.2012.11.009

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

Current CBPS proposals support an *exhaustive filtering semantic*, i.e. a consumer registering its interests will receive all the corresponding matching publications. Such semantic is appropriate for a wide range of applications including distributed games, stock quotes or monitoring applications. However, for applications such as news distribution or content sharing, the amount of relevant available publications may be overwhelming as information consumers have a limited attention span. Implementing the same exhaustive filtering semantic for these applications would result in a huge information overload and communication overhead. Another important requirement so as to meet application delay requirements is high-throughput forwarding.

This paper is concerned with scaling decentralized CBPS for large-scale content distribution, which requires dealing with several issues. First, information consumers have different requirements. In fact, most consumers are likely to be overwhelmed by the amount of relevant information available in the network and would like the system to deliver matching publications at a maximum rate, while some would be interested in every publication matching their interests. Moreover, CBPS should enable information consumers to select previously published publications. Thus, large-scale CBPS services should be able to capture the heterogeneity of consumers' needs. Second, bandwidth is a critical resource. In fact, the design of bandwidth-thrifty solutions has been and remains one of the major challenges of CBPS research.

We envision a mediation network providing the mediation between information providers and information consumers and involving different service providers and administrative entities. The mediation network is a distributed infrastructure supported by a set of interoperating mediation routers (MRs), or brokers or simply routers (throughout the paper the terms mediation routers, brokers and routers are used interchangeably). We assume that MRs have important storage resources available and support a store-and-forward model [1]. This work aims at leveraging caching in order to meet the above stated requirements of decentralized and efficient large-scale content distribution. Particularly, we focus on increasing the communication-efficiency of CBPS operating on large-scale with an emphasis on large number of widely spread consumers with heterogeneous requirements. The same emphasis is given on the large number of publications and the scarcity of end-to-end bandwidth.

Our contributions are threefold. First, we design a service model that captures the heterogeneity of information consumers' requirements. This service model, which generalizes our previous work [2–4], seamlessly supports content-based information retrieval and dissemination and enables information consumers to express their actual information needs by the maximum number of publications desired per service period. Second, we describe the protocols required to efficiently implement the service. Finally, we investigate and evaluate different caching policies in order to regulate the trade-off between caching efficiency and the QoS offered.

The rest of the paper is organized as follows. In Section 2 a brief related work on content-centric pub/sub architectures and caching is given, while in Section 3 we present the functionality of the proposed service model. Section 4 is devoted to performance evaluation via simulations, while finally in Section 5 we conclude the paper.

2. Related work

Content-centric pub/sub networking is becoming increasingly popular for content access and dissemination, especially nowadays that Internet's usage has considerably changed from a resource sharing mechanism between a pair of hosts to a content distribution and retrieval mechanism. There are several research efforts that develops an overlay event notification service like IBM's Gryphon [5], Siena [6], Elvin [7], and REDS [8] which implement the pub/sub architecture. Moreover there are also several research efforts, among others PURSUIT [9], NDN/CCN [10] and SAIL [11], where information is explicitly named so that anybody who has relevant information can potentially participate in the fulfillment of requests for said information in a way to achieve scalability, security and performance. Finally the pub/sub architecture paradigm is already used in many commercial applications such as Google Alerts through the usage of a simple, open, server-to-server web-hook-based pub/sub protocol called PubSubHubbub protocol [12].

In the context of Web caching, NLANR designed the Internet Cache Protocol (ICP) [13] and the HTCP [14] protocol, to support discovery, retrieval and management of documents from neighboring caches as well as parent caches. In the area of Delay Tolerant Networks (DTNs) in [15] authors study the performance of caching by the nodes. Generally DTN can provide ad hoc communication services within (sparse) mobile user communities when end-to-end IP communication is not available. This implies that the nodes cache the data for some time, as DTN operates as a store-carry-and-forward network. The data that is being "carried" can also be used to serve requests from other nodes before its lifetime has expired.

In the area of packet caches in routers, [16] explores the benefits of deploying packet-level redundant content elimination as a universal primitive on all Internet routers. In order to have the most benefits from the caching scheme, redundancy-aware, intra- and inter-domain routing algorithms are recommended, lowering intra and inter-domain link loads by 10–50%. The proposed redundancy-aware routing algorithms are somewhat similar to multicast routing algorithms [17]. Moreover, CacheCast [18] is a mechanism to cache data in normal data streams by allowing senders to identify the packet content by a payload ID.

Additionally, in [19] authors present CONIC, a network architecture designed for efficient data dissemination using storage and bandwidth resources in end systems. CONIC exploits available storage located in end hosts and uses it as caches. Cached content is then advertised to the network and routers forward interests to the caches if the topological distance to the cache is smaller than the distance to the object's originator. Finally, the Cache and Forward architecture (CNF) [20] is a content centric network architecture operating as an overlay on top of

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

the current Internet Infrastructure. CNF consists of two components: a hop-by-hop transport service and in-network caches. Content objects are split in chunks and are transported between CNF routers in a hop-by-hop manner, instead of end-to-end TCP byte streams. In CNF, every router is equipped with enough memory to serve as a cache. Each router may operate autonomously and cache objects as they pass by.

Caching in content-centric pub/sub networking exhibits fundamental differences from the traditional caching schemes and poses new challenges [21,22]. In general, caching in pub/sub enables caching of information items [23,24] in every cache-equipped node [25] and replacement of cached items at line-speed [26]. The cache-everything-everywhere scheme presented in CCN [24] has already raised doubts and some authors have already questioned this aggressive strategy [27]. In that direction in [28] authors instead of caching the same item at every node along the delivery path investigate if caching only in a subset of node(s) along the delivery path can achieve better performance in terms of cache and server hit rates by exploiting the concept of betweenness centrality. Moreover, in [29] authors formulate the caching problem into a Linear Programming problem and propose a novel caching policy named as *Least Benefit*, which takes into account the benefit of a cache hit instead of simply counting the hit number as the Least Frequently Used (LFU) policy does.

In [30] authors propose a new cooperative caching strategy for the CCN architecture that has been designed for the treatment of large video streams with-on demand access. Their aim is to minimize the amount of queries for time-shifted TV that are treated by servers outside the ISP network and the proposed caching strategy manages to halve the cross-domain traffic. In a similar scenario authors in [31] evaluated the performance of a two-layer cache hierarchy under a demand model that reflects a realistic traffic mix. Their results demonstrate that caching Video on Demand in access routers offers a highly favorable bandwidth memory tradeoff but the other types of content that they considered (web, file sharing and user generated content) would likely be more efficiently handled in very large capacity storage devices in the core. Finally, in [32] authors proposed static and dynamic storage management policies with the common objective of providing service differentiation to multiple applications sharing the same CCN infrastructure. The proposed dynamic storage allocation mechanisms enhances service differentiation by adding flexibility in storage resources sharing, yielding high system efficiency while avoiding memory underutilization.

Every research attempt regarding caching in content-centric pub/sub architectures was based in the CCN architecture, where each interest in the network matches only one item and not a spectrum of items that matches the attribute/value subscription of the CBPS scheme that is assumed here. This implies that the aforementioned research attempts does not capture the quantitative and qualitative heterogeneity of the information consumers that is the main target of this work. Actually the proposed work is not an add-on for efficient caching in a well know architecture, but a complete framework which aims to increase the efficiency of CBPS on large-scale scenarios where consumers have heterogeneous requirements and bandwidth is scarce. Finally, all the research attempts, where the CCN architecture is used, assume the presence of a hosting server for each information item making the retrieval of previously published information a straightforward procedure. In contrast, in CBPS publishers upload their publications to the network only once and might disappear from the network requiring additional functionality for the retrieval of such information which is also part of the proposed framework.

Information filtering literature provides directions to filter a stream of information and deliver the most relevant information items to users given their predefined profiles in order to deal the issue of Information overload. Information filtering leverages several information retrieval ranking techniques [33]. Particularly `SIFT` [34] focuses on efficient indexing and matching in a centralized manner, while two distributed version of `SIFT` are sketched in [35].

Leveraging content-based publish/subscribe for content distribution has been considered in [36] where authors study the problem of how to pace the dissemination of information between a publisher and a proxy server when usage-patterns and subscription information are available. Although their effort is original, their solution is centralized and not distributed as the one presented in this paper. Other work that have considered caching to increase the availability of publications in decentralized `CBPS` systems implement an exhaustive filtering semantic [37].

Pacing the dissemination process to information needs has also been previously addressed. `Corona` [38] is a pub/sub system which provides high performance and scalability through optimal resource allocation. `Corona` aims at improving the performance of web syndication through cooperative polling. In spite of its effectiveness, `Corona` applies only to channel-based pub/sub systems.

3. The content-based mediation network

This section describes a framework for large-scale content-based networking characterized by the heterogeneity of information consumer requirements, the information overload and the scarcity of end-to-end bandwidth. The framework introduces a service model which captures the quantitative and qualitative heterogeneity of information consumer needs, and addresses the key design issues in implementing efficiently the service model at large-scale.

3.1. Content-based networking (`CBN`): Overview

Content-based networking involves three types of entities: *subscribers*, *receivers* or *information consumers*, *publishers* or *information providers* and *mediation routers (MRs)*, or *brokers*. Each receiver submits its interests by sending a subscription to the network where MRs acting as proxies are responsible for returning the corresponding matching pieces of data. The first MR to handle the subscription advertised by a receiver, *i.e.* one of the proxies mentioned above, is called a *home mediation router* (or *home broker*).

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

Note that receivers select their home MRs on the basis of trust or proximity. Publishers upload their publications so that they can be disseminated to interested receivers. A publication consists of a data item and a metadata description, while an interest is described by a predicate over the metadata space. Predicates and metadata typically follow an attribute/value schema. A publication *P* matches a subscription *S*, whenever the metadata describing *P* matches the predicate defined by *S*. MRs cooperate to efficiently disseminate data items corresponding to uploaded publications towards interested receivers (Fig. 1 depicts a typical mediation network).

Content-based forwarding (CBF) is the algorithm that based on the information established by the routing algorithm, processes incoming messages to decide on which interface an incoming message should be forwarded. That information is compiled in the forwarding table which associates each interface to a filter combining the predicates of the descendants in the dissemination tree via that interface. We define a filter as a compact representation of a set of predicates. Efficient data structures for forwarding tables are mentioned in [39,40]. Carzaniga et al. [41], describes two content-based forwarding schemes requiring a spanning tree rooted at each sender that can be configured through shortest-path trees or reverse-path forwarding. However, these CBF schemes are correct only if spanning trees verify the *all-pairs path symmetry* property, i.e. only in the case where shortest-paths are unique or routes between routers are symmetric. In practice, it is difficult to enforce such properties. Consequently, the deployment of such protocols is realistic only atop a global spanning tree. In order to reduce the complexity of content-based forwarding protocols, which requires evaluating publications against the index of advertised subscriptions at every hop of the dissemination tree, other work has proposed to implement matching only at the publishing brokers and switching at subsequent brokers of the dissemination tree on the basis of explicit forwarding directives. For instance, the DV/DRP protocol [42] is proposed with an optimization which consists in doing matching only at the source of publications and appending a particular structure which identifies the recipients of the message. When a message has to be pushed on two or more interfaces, the destination set structure is affected before being attached to the message and forwarded to the destinations downstream each interface. However, DV/DRP uses a compact bit vector data structure to address the message, whose size equals the number of sink nodes in the system. This assumes a small number of potential receivers and is thus not scalable. LIPSIN [43] is a forwarding protocol that encodes dissemination trees in message headers as Bloom filters and achieves line-speed forwarding. However, LIPSIN requires that each link of the topology be assigned an identifier and requires either a separation of the control plane from the forwarding plane or/and that each router has a global overview of the topology.

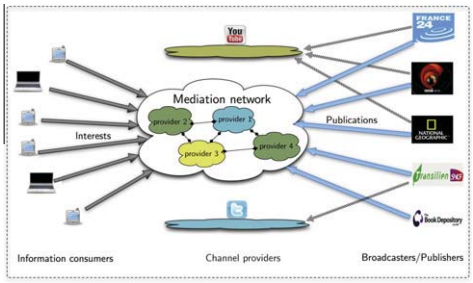


Fig. 1. Mediation network.

Content-based routing (CBR) is the distributed algorithm that collects, propagates and assembles receivers' interests as well as topological information to the router forwarding functions. Existing content-based routing (CBR) schemes are designed to support an exhaustive semantic where receivers register for all relevant publications matching their interests. Typically, routing consists in broadcasting subscriptions within the network in order to configure the dissemination tree required to efficiently forward publications from senders to receivers. Content-based routing requires a broadcast layer for operation on top of arbitrary topologies, which can be implemented through spanning trees.

3.2. e-CBN: An enhanced service model for large-scale content-based networking

We consider a mediation network constituted of MRs, involving independent stakeholders, interconnected according to an arbitrary topology that captures the specific features of content networks, and that provides efficient dissemination channels for *information providers* to reach *information consumers* at large-scale.

Information providers (or publishers) upload their publications to the mediation network so that they can be disseminated towards interested receivers. We assume that authoritative publishers upload publications once to the mediation network and each publication is assigned a unique identifier. The purpose of restricting publications to authoritative publishers is to guarantee that two publications with different identifiers embed different content.

A receiver *r* advertises its information needs to its home MR *R*, as a subscription *S* (*predicate*, *max*, *lifetime*, *freshness*) where *predicate* defines its information interest, *max* specifies the maximum amount of publications admissible by *r* over a period of time *lifetime*, and *freshness* the maximum age for a relevant publication matching the information interest. The *freshness* of a publication is the elapsed time since its initial upload in the mediation network. *S* is constituted of more attributes which are introduced throughout the section.

A publication *P matches* a subscription *S*, whenever the metadata describing *P* matches the predicate defined by *S* and the age of *P* is less than *freshness* at the end of the service period. Mediation routers cooperate to efficiently disseminate data items corresponding to uploaded publications towards receivers which have advertised relevant interests.

We will refer to *max* as the *selectivity* of the subscription and *lifetime* can be interpreted as the delay allocated by receivers to the content-based network to satisfy an interest. The content-based network delivers to receiver *r* (via

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

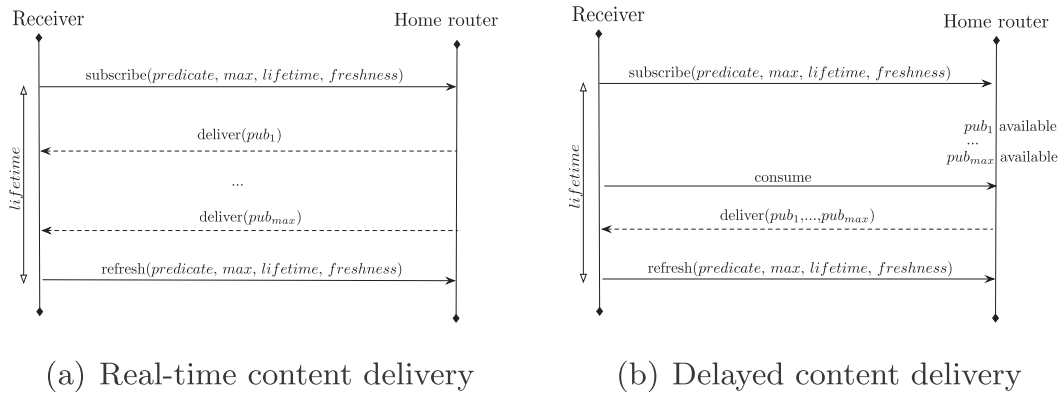


Fig. 2. Interactions between receivers and home mediation routers.

Table 1
Examples of interests for different applications.

Application	Predicate	Selectivity	Lifetime	Freshness
Notification services	(terms = ‘‘RER + C+Infos")	All	7 days	0
News alerting services	(type = article, terms = ‘‘election + 2012")	10	24 h	24 h
Content retrieval	(type = article, terms = ‘‘sophia + perennis")	All	24 h	Any

home MR *R*), at most *max* relevant publications before *lifetime* expires. Depending on *r* preferences, content delivery is either performed in real-time or delayed until *lifetime* expires at the latest or until *r* requests the publications retrieved on its behalf by *R*. This latter possibility requires to provision a minimum amount of caching resources at home MRs, but provides further temporal decoupling between receivers and their home MRs allowing temporary disconnections of the receivers.

These two variants are depicted by Fig. 2. Content delivery from home MRs to receivers require that the former maintain some states regarding their registered customers. For the sake of clarity, we assume that retrieved publications are delivered to receivers in real-time.

The service model described above captures the attention span of information consumers by allowing them to specify the maximum amount of relevant information they would like to receive over a period of time. Also, it captures the qualitative heterogeneity of information consumers.

When *freshness* equals zero, the interest has the conventional semantic of a subscription, *i.e.* the interest selects only future publications. When *freshness* and *lifetime* are both positive, the interest is a *loose subscription* which differs from a conventional subscription by the fact that requesters are only interested in publications that they did not consume previously. Loose subscriptions can be refreshed after *lifetime* expires. Then, the system guarantees to the requester that the refreshed subscription is not satisfied with previously consumed publications. Finally, in the case where *lifetime* equals zero, then the interest is non-persistent, *i.e.* a request to the mediation network to retrieve immediately up to *max* available publications.

Table 1 provides typical parameter settings for various applications, while Fig. 3 illustrates the different semantics of the proposed service model.

This paper focuses on the efficient processing of loose subscriptions which have not been previously studied in the literature. Efficient processing of non-persistent interests (resp. subscriptions) in an unstructured overlay have been largely studied in [44] (resp. [45,41]) and previous work can be leveraged. Note that within the framework, non-persistent interests can be processed similarly to loose subscriptions advertised with a very small lifetime.

Receivers are allowed to refresh their interests (*loose subscriptions*) when they expire. As such, an important requirement for the usability of the service is defined as follows:

Req 1. The service should guarantee that a refreshed interest will be satisfied at most once by any publication over its successive lifetimes. This condition should be enforced without having to track an exhaustive history of all interests satisfied with a publication or of all publications already consumed by a subscription.

To allow MRs to differentiate refreshed interests from new ones, we assume the existence of an agreement between MRs and receivers for such purpose. Control messages exchanged by MRs to advertise interests include a *refresh flag* indicating whether interests are refreshed or not (see Fig. 4).

Definitions 1. Let *S* (*predicate*,*max*,*lifetime*,*freshness*) be a subscription registered at t_0 issued by *r*, N_S be the number of publications notified to *r* by $t_0 + lifetime$ and M_S be the total number of publications uploaded between t_0 and $t_0 + lifetime$ and matching *S*.

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

- S is satisfied when:

$$N_S = max. \quad (1)$$

- Starvation occurs when:

$$N_S < max \leqslant M_S \quad (2)$$

Starvation occurs due to congestion or due to the service failing to timely satisfy interests. The starvation probability, *i.e.* the frequency of occurrence of starvation, is the metric used to characterize the quality of service offered by an implementation of the service model compared to an implementation of the exhaustive filtering semantic. Starvation does not account for interests which are not satisfied due to content unavailability.

Problem statement: We aim at minimizing with a low state complexity, the starvation probability and the communication costs required to satisfy a workload of loose subscriptions assuming bandwidth, storage and processing resource constraints.

3.3. e-CBN: Architecture and mechanisms

In order to implement the service model, the e-CBN framework addresses the following design issues:

Content forwarding schemes supporting both content retrieval and dissemination traffic. Content dissemination traffic should be handled by leveraging existing content-based forwarding schemes. Integrating content retrieval and dissemination in the same middleware is not straightforward. An important requirement is the ability to take into account bandwidth and resource constraints by assigning priorities to transiting content. Such forwarding algorithms should minimize starvation in presence of congestion.

Dissemination strategies defining the conditions making a publication eligible for dissemination towards interested receivers. Our service model requires strategies pacing the dissemination process to information needs.

Interest forwarding strategies necessary to support content forwarding schemes and dissemination strategies that should take advantage of locality in content availability patterns as well as attention span quantification in order to minimize communication costs, unlike existing content-based routing schemes that broadcast subscription messages to satisfy an exhaustive filtering semantic.

Caching policies increasing content availability as well as communication-efficiency.

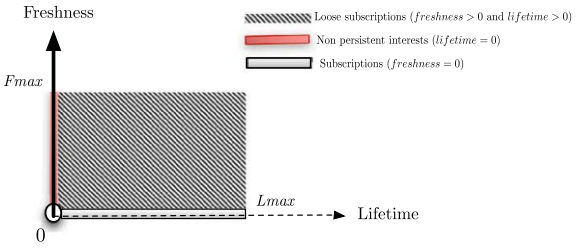


Fig. 3. Illustration of the different semantics of the service model.

3.3.1. Mediation router model

The model of a MR is depicted by Fig. 5. Each MR needs the following data structures to operate:

Pending interests table (PIT), which is constituted of an *index* of interests advertised to this MR and a *forwarding table* which provides information about the origin of the interest. The *index* supports a matching method which provides the identifiers of the interests matching a given publication and the *forwarding table* associates interest identifiers to the originating MR. We assume that once an interest is satisfied, the corresponding home MR removes the interest states from the PIT.

Pending publications table (PPT) references pending publications, which have just been uploaded at some MR from publishers and that are waiting for opportunities to be further disseminated *i.e.* which have not been

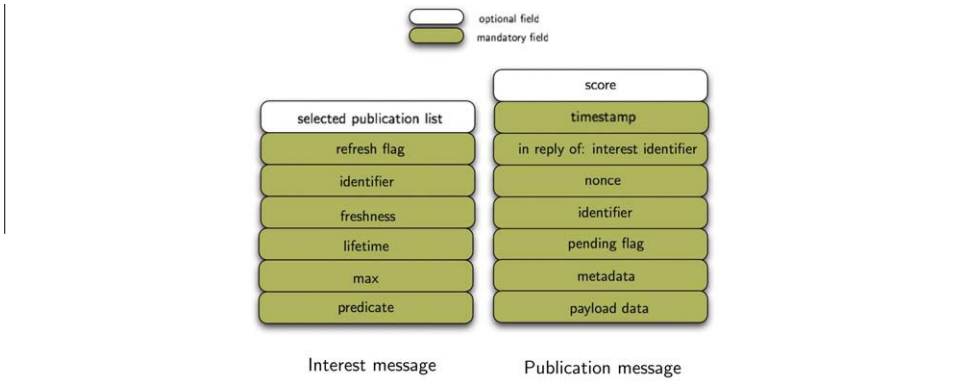


Fig. 4. Interest/subscription and publication messages structure.

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

used to satisfy remote interests. A pending publication may have been used to satisfy interests, which are local to a MR. In order to avoid that refreshed interests consume the same publication at their home MR, we add a *dispatched flag* (DF) in the PPT indicating whether a pending publication has been used to satisfy local interests. Note that when a publication is disseminated for the first time, it is forwarded with the *pending flag* (PF) enabled.

Overflowing publications table (OPT) references overflowing publications, which have been disseminated to remote subscribers, but that never served locally. In fact, they may be useful for refreshed interests in future lifetimes. New entries are added to the OPT for pending publications selected by remote interests with dispatched flag DF disabled and for publications incoming with pending flag PF enabled without matching local interests. The latter situation occurs whenever the content-based network returns more publications than requested or the states corresponding to an interest are still in the forwarding table or the *en-route caching* optimization is enabled (see Section 3.3.2).

Disseminated publications table (DPT) references publications which have been disseminated towards remote MRs and used to satisfy local interests.

Boxes table (BT), which references abstractions called boxes and are used to keep the preferences of the consumers as well as to monitor the service offered to them. For instance, MRs will use boxes to monitor the set of publications selected to satisfy local interests during their lifetime and use that information to detect duplicates. Also, they will use the boxes to detect that interests are satisfied and stop advertising them in their PIT and optionally advertise overload for that interest in the mediation network.

Buffers upstream and downstream the forwarding decision modeling the two bottlenecks of the MRs. The first bottleneck is related to the matching method of the PIT and the second bottleneck is related to the transmission of publications.

The cache indexes introduced above have been designed to effectively support Req. 1, while maximizing opportunities to satisfy refreshed interests. Fig. 6 depicts the lifecycle of a publication inside the cache of a MR *i.e.* the transitions between the indexes. Fig. 7 depicts the processing of an incoming (uploaded or transiting) content by a MR. Besides the *new*/*refreshed* classification, from the perspective of a MR, a *remote interest* denotes an interest advertised by downstream MRs, while a *local interest* denotes an interest registered by a local receiver.

3.3.2. Caching strategies

We assume that mediation routers are provisioned with a finite amount of caching resources. One might argue that it is always possible to extend MRs caching resources by distributing MR behaviors over a cluster of COTS servers as performed today by major online service providers. In practice, the size of a cluster is limited by practical reasons such as the lack of space or some energy considerations. Also, the size of the cache impacts the size of the cache indexes which should ideally fit into main memory. Consequently, beyond some amount of caching resources, it may be necessary to partition the cache indexes over several nodes. But, the response time of every MR to satisfy an interest will increase with the number of nodes involved in the processing and may impact timely delivery of content.

A publication incoming at a MR *R* and originating from a publisher or another MR is *cacheable* whenever, *R* is the MR that originally advertised the interest(s) that triggered the retrieval/dissemination of the content in the mediation network, or *R* is the MR where the publication is originally uploaded or, *R* is allowed to opportunistically cache transiting publications *i.e.* when the en-route/flushing option(s), introduced later in this section, are enabled.

Selection and replacement policies. Each MR executes a *selection policy* to determine which publication to select first for an incoming interest, and a *replacement policy* to determine which publication to replace first in case of cache overflow. Ideally, selection and replacement policies should achieve an optimal trade-off between the following tussles: receivers privileging fresh information, publishers wanting to reach the widest possible audience with their publications and network operators willing to minimize communication costs and maximize the quality of service offered to consumers.

Selection policies should guarantee that new interests are satisfied with any available publication, while refreshed interests are served only with publications that

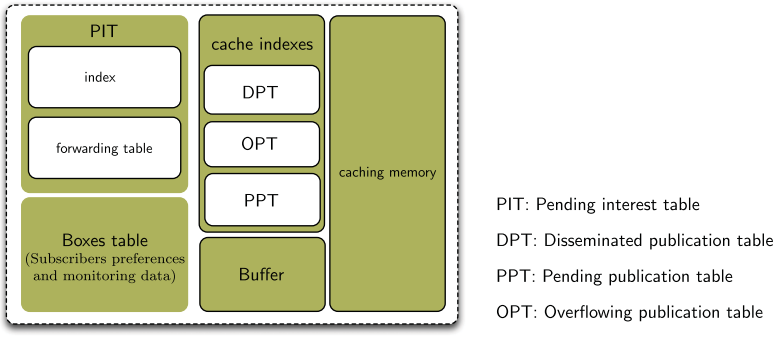


Fig. 5. Mediation router (MR) model.

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

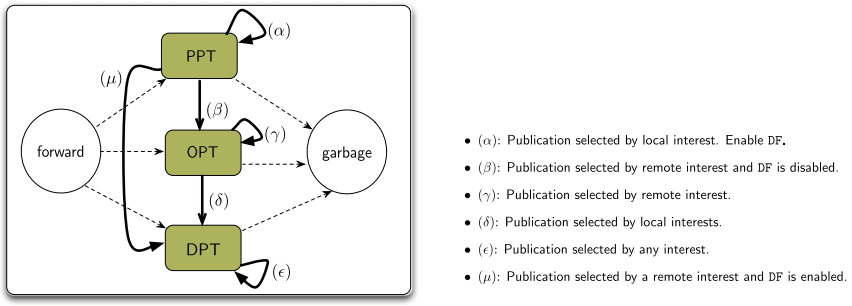
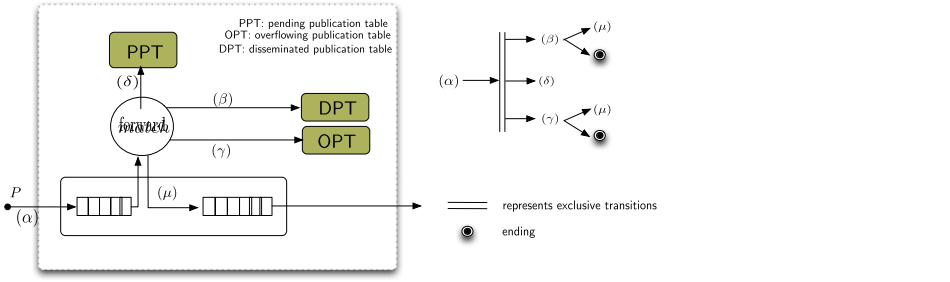


Fig. 6. Publication lifecycle inside the cache of a MR.



- $\alpha \rightarrow \beta$: P is a publication requested by R and used to satisfy local interests.
- $\alpha \rightarrow \beta \rightarrow \mu$: P is a transiting publication that matches interests advertised by downstream MRs/receivers while the NRT policy is enabled or while R is a recipient of the publication and PF disabled or P is an uploaded publication matching local and remote interests.
- $\alpha \rightarrow \delta$: P is an uploaded publication matching local interests or no interests.
- $\alpha \rightarrow \gamma \rightarrow \mu$: P is an uploaded publication selected by remote interests only or P is a transiting publication matching remote interests only, while the NRT policy and PF are enabled.
- $\alpha \rightarrow \gamma$: P is a transiting publication with PF enabled matching no interest in the table or matching already satisfied local interests.

Fig. 7. Caching and forwarding decisions inside a MR *R*.

have not been delivered to them during previous lifetimes. In order to be consistent with (Req. 1), refreshed interests should not be satisfied with publications indexed in DPTs or in OPTs of remote MRs. More precisely, refreshed interests are satisfied first with overflowing and pending publications available at the originating MR with DF disabled, then with pending publications available at remote MRs of the network. Consequently, new interests have more opportunities to be satisfied than refreshed ones. We can increase content availability for refreshed interests by making pending publications more persistent. At high-level, selection and replacement policies may discriminate or not publications according to their type (disseminated, pending, overflowing).

We consider two high-level discriminating policies simply called *discriminators*:

- DPF (*disseminated publications first*), which returns first disseminated publications, then overflowing and finally pending ones.
- PPF (*pending publications first*), which returns first pending publications, then overflowing and finally disseminated ones.

We note discriminator-selection policy/discriminator-replacement policy the combination of policies executed by a MR. The first discriminator in the notation applies to the selection policy and the second one applies to the replacement policy. In the case where no discriminator applies, the discriminator field of the notation and the following dash are left blank.

Selection and replacement policies should be designed/chosen in order to balance the trade-off between new and refreshed interests as well as to meet the expectations of information consumers, information providers and network operators:

- In order to balance the tradeoff between new and refreshed interests, the following combinations of policies can be considered: (DPF-∗/DPF-∗), and (PPF-∗/DPF-∗) where ∗ may refer to one of the following policies: *most recently used* (*MRU*), *least recently used* (*LRU*), *most frequently used* (*MFU*), *most fresh* (*MF*) or *least fresh* (*LF*). These policies make pending publications more persistent than other publications, which is risky as pending publications may correspond to unpopular publications.

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

Table 2
Selection and replacement policy configurations to investigate.

Configuration	Expected properties	Name
LFU/MFU	Fairness	LFU
MRU/LRU	Communication-efficiency and availability	MRU
MF/LF	Availability of fresh content	MF
LF/LF	Availability of fresh content	LF
DPF-MF/DPF-{LF,LRU}	Increases availability of content for refreshed interests	DPF,DPFu
PPF-MF/DPF-{LF,LRU}	Increases availability of content for refreshed interests	PPF,PPFu

- Information consumers request most recent publications (*freshness*) and want their interests to be satisfied (*availability*). Consequently, selecting most fresh information first and replacing least fresh information first *i.e.* an MF/LF policy, will make most fresh publications more persistent in the mediation network. An LF/LF policy is also worth investigating w.r.t. the availability of fresh information.
- Information providers want to reach the widest possible audience. *Fairness* among information providers in terms of content availability for content of similar popularity is important in order that some popular content providers do not get discriminated. An interesting policy to investigate to achieve such fairness is LFU/MFU.
- From the perspective of network operators, communication-efficiency and content availability for consumers can be achieved by enforcing an MRU/LRU policy.

In the next section, we evaluate the configurations presented in Table 2.

Caching policies. Besides the selection and replacement policies, MRs can enforce one of the following caching policies:

Default: With this policy enabled, publications are cached only at the publishing or requesting MRs.
En-route caching (NRT): With this policy enabled, MRs are allowed to cache transiting publications according to the enforced replacement policy and if the content items are not already present into the cache. The evaluation in the next section clarifies situations where MRs have incentives to cache transiting content. Selecting publications replicate those publications at several routers and consequently increase their availability. Replication and persistence of publications are expected to increase when the NRT policy is enabled.

3.3.3. Interest/Subscription forwarding

Interests have a persistent and/or temporary lifetime. In their temporary lifetime, *i.e.* during their propagation in the network, interests are satisfied with cached publications. Instead, in their persistent lifetime, interests are satisfied with pending publications which have just been uploaded or selected for dissemination. Advertising an interest consists in flooding the interest in the mediation network until it is satisfied or all MRs are notified with the interest.

Note that in order to avoid loops and duplicates with arbitrary topologies, we assume that interest messages embed a globally unique identifier to detect cycles. Such a scheme can be achieved by generating hierarchical identifiers where the prefix is the identifier of the mediation router. This is more convenient than operating over global spanning trees. In our work, we assumed that during their lifetime, interests are advertised once in the mediation network, otherwise a nonce would have been necessary. Moreover, we assume that interests have different identifiers throughout their successive lifetimes.

Mediation routers exchange interests using the following procedure: Upon reception of $S(max,*)$ MR/Broker C (Fig. 8) checks if the interest identifier has already been advertised in the PIT and if so drops the interest. Otherwise, if the number of relevant publications available into the cache exceeds or equals *max*, then *max* publications are selected for delivery and the propagation is stopped. Otherwise, S is further advertised with the *max* parameter decremented by the number of matching publications offered by C.

Fig. 8 illustrates a scenario where starvation may occur: Three publications relevant to I are available but only two of them are forwarded to Broker_C. This is due to the bad selection decision of Broker_P which returns c_4 instead of c_5 that would have contributed to satisfy the interest. Moreover, forwarding c_4 in the mediation network generates an overhead that should not be neglected. Note that the requesting MR/Broker can always detect duplicate publications retrieved from the network for the same interest using the states available in the BT.

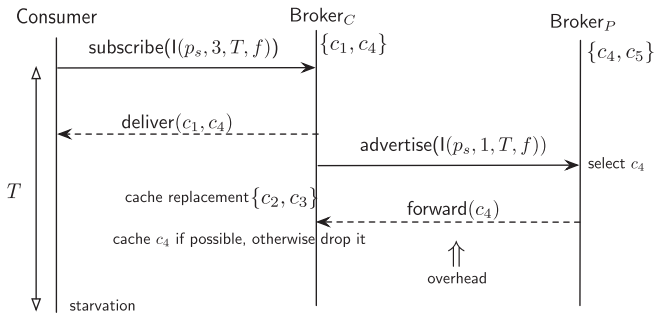
In order to attenuate that overhead, we introduce in Section 3.3.6 the *in-network duplicate dropping* (IDD) mechanism, where a MR drops a publication already present in the cache because it may have already responded with that content. In order to reduce starvation due to bad selection decisions, we also propose in Section 3.3.6, the *proactive duplicate dropping* (PDD) mechanism or *duplicate avoidance*, where interests are forwarded with the list of publication identifiers already used to satisfy them along the forwarding path. These schemes are not perfect and may infer themselves starvation and an overhead of publication messages. This situation underlines the sensitivity of selection policies on starvation.

3.3.4. Dissemination strategies

We discuss below, two simple yet effective strategies in the trade-off between satisfaction and communication-efficiency.

Push/pull and explicit overload notification (EON). *Overload* corresponds to the situation where the number of publications retrieved from the network w.r.t. an interest exceeds the selectivity of the subscription. A publication P is disseminated by a MR R, whenever at upload time, P matches an interest in the forwarding table or if a matching interest transits through the node while the publication is pending (*push*/*pull*). When an interest is satisfied i.e. *max* or more publications have been retrieved, the corresponding home MR advertises an overload notification message in order that remote MRs remove the corresponding states from their tables (*overload notification*). Fig. 9

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009



Note: We assume that c_1, c_4, c_5 are relevant to interest I. Broker$_C$ is the home router of Consumer and the cache replacement event may correspond to the upload of new content forcing the cache replacement policy to apply.

Fig. 8. Starvation illustration.

describes the interest forwarding strategy operation with the EON mechanism enabled.

Pull/delayed push (PDP). Unlike EON, PDP does not use explicit notification messages to notify remote MRs of *overload*. PDP uses the propagation of new and refreshed interests by the content-based routing protocol to pace the dissemination process (*pull*) instead of pushing publications additionally. In order to avoid that starvation occurs in some cases, we compute a *most lately publication time* for each uploaded publication. Let P be a publication uploaded at a MR R at time t_0 and S_P be the set of matching interests advertised in R's forwarding table (excluding local interests), the most lately publication time t_P associated to P is given by the relation:

$$t_P = t_0 + min_{S \in S_P}(deadline(S)) - \beta. \quad (3)$$

βbeing a system parameter and $deadline(S)$ being the remaining time before S expires. At last, any publication that may contribute to satisfy an interest not yet satisfied, is finally disseminated. Note that β should be engineered such that it is greater or equal to the time required to forward a publication between two endpoints of the network. In the case where a scheduled publication is eligible for replacement by another one, the scheduled publication is immediately disseminated and the incoming publication is cached.

A publication scheduled with PDP is disseminated at scheduling time even though a new interest selects the publication before its dissemination. But, the publication stops being pending, but overflowing or dispatched, depending on whether there exists local/remote recipients. Thus, while the publication is scheduled it may be selected by other interests and forwarded into the mediation network. Consequently, publications disseminated with PDP are forwarded with the pending flag disabled.

3.3.5. Publication forwarding

Publications are forwarded in the content-based network depending on whether they are selected by pending or transiting interests. In Fig. 4, the *nonce field* is used to identify publication messages and the *identifier field* is used to identify content embedded in publication messages. Nonces are used to detect duplicates on cyclic topologies.

Unicast delivery: A publication selected by a transiting interest is forwarded on the reverse path only towards the requesting broker. The information of the broker that advertised a particular interest identifier is provided by the PIT.

Multicast delivery (**content dissemination**): When a pending publication is selected by pending interests, it is disseminated towards all the receivers that have advertised the interest. Each broker which belongs to the dissemination tree DT of the pending publication rooted at the publishing broker, can determine the next hop in DT by evaluating the publication against the set of interests advertised in the PIT. Content dissemination can be realized according to any of the baseline content-based forwarding schemes described in Section 3.1, except that we prefer the usage of nonces embedded in publication messages in order to avoid forwarding loops instead of global spanning trees.

3.3.6. Forwarding optimizations

Congestion-aware forwarding. Congestion may appear at some MRs. It is important in these conditions to maximize the satisfaction of the service. This can be done by assigning priorities to transiting publications. The priorities are used to schedule publications in buffers upstream and downstream the forwarding decision.

For this purpose, we compute a score for each publication embedded in publication messages. The score is recomputed by every hop on the delivery path. Let R be a MR and Π be the set of publications buffered at R that are waiting to be further disseminated. For each publication $P \in \Pi$, we define S_P the set of interests matching P downstream R, and compute a score Cbf(P) giving higher priority to *popular* and *urgent* publications. We estimate popularity by the number of matching interests advertised downstream R and urgency by the minimum deadline among matching interests. Cbf(P) is defined by the following equation:

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

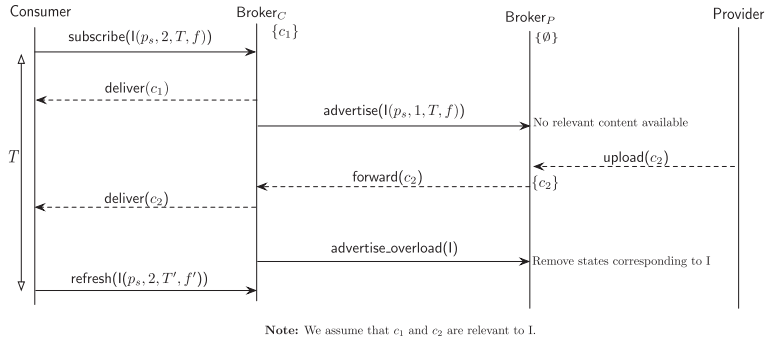


Fig. 9. Interest forwarding and content dissemination.

$$Cbf(P) = \frac{|S_p|}{\min_{I \in S_p}(\text{deadline}(I))}. \quad (4)$$

Note that in the unicast case, the score associated to a publication selected by an interest I is simply $1/deadline(I)$. Obviously, in any case, to avoid a division by zero in Eq. (4), interests whose deadlines have been reached are removed from the PIT and not considered in the computation of the score.

Handling duplicate responses. When an interest is advertised in the content-based network, different MRs may reply with the same (overflowing or dispatched) publication. In the worst-case, the duplicate responses will be detected by the requesting MR. This may infer starvation in scenarios such as the one depicted by Fig. 8, as well as a communication overhead. In order, to mitigate that overhead, we propose the mechanisms described below.

In-network duplicate dropping (*IDD*). Every MR, upon the arrival of a publication, checks whether the publication already appears in its cache and if this is true, it discards it. Otherwise, it forwards the publication according to the technique described in Section 3.3.5. The reason for searching the cache of a MR upon the arrival of a publication, is because publications follow the reverse path that interests follow. This means that the interest that selected the corresponding publication has also been processed by the MR under question which may have responded to that interest with the same cached publication (s). Note that this mechanism is not always effective as depicted by Fig. 10: The intermediate router Broker$_P$ cannot detect that content c_2 has already been forwarded to Broker$_C$ (that itself forwarded). As a consequence, Broker$_P$ redundantly forwards c_2 to Broker$_C$ which finally drops it thanks to the states installed in the BT.

Note that the duplicate dropping heuristic does not thoroughly solve the problem addressed but are expected to alleviate the communication overhead. Fig. 10 exhibits an extreme case where in-network duplicate generates both starvation and communication overhead.

Duplicate avoidance or Proactive Duplicate Dropping (*PDD*). In order to improve the effectiveness of the in-network duplicate dropping scheme, we propose a proactive counterpart such that a MR processing an interest and selecting cached publications to serve an interest, append the list of identifiers of the selected publications to the corresponding field in the interest message before further forwarding the interest. With this mechanism, duplicate responses will be eliminated from every branch of the tree within which the interest is forwarded.

3.4. Discussion

This section described the e-CBN framework for large-scale content-based networking. The framework introduces a service model capturing the quantitative and qualitative heterogeneity of information consumers and addresses the key design issues in implementing efficiently the service model. In the next section, we quantify the communication gains of e-CBN over baseline content-based networking under realistic workload assumptions

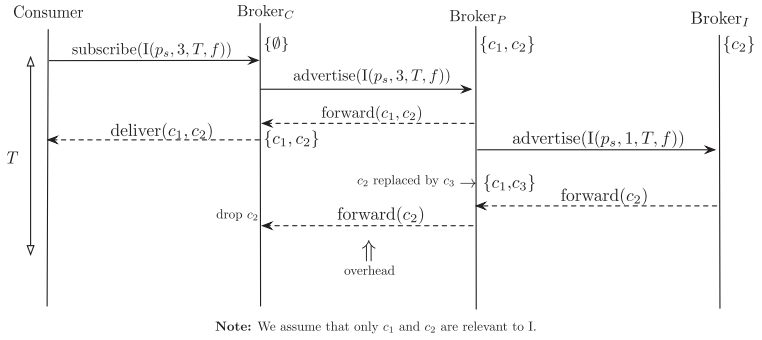


Fig. 10. Duplicate dropping mechanism limitations.

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

and evaluate the effectiveness of each mechanism implemented by the framework. Table 3 depicts all the used acronyms of the e-CBN framework.

4. Performance evaluation

This section provides a characterization of the framework with respect to the quality of service offered to consumers and communication-efficiency. Firstly, we quantify the gains of the framework (e-CBN) over a baseline scheme implementing the exhaustive filtering semantic (EF). We assume in the EF case that receivers caching resources are unlimited which is an extremely unfavorable assumption for e-CBN. Secondly, we evaluate the efficiency of eight policies introduced by Table 2 and discuss the benefits of en-route caching (NRT). Finally, we measure the performance of the dissemination methods and congestion-aware forwarding scheme.

Performance is characterized using the following metrics:

- **Satisfaction**, which is the percentage of interests satisfied as defined by Eq. (1). We use this metric mainly to compare an instance of the framework to EF.
- **Starvation probability** (SP), which is the frequency of occurrence of starvation. SP is selected when comparing two instances of the framework. Unlike the semantics of *satisfaction* as defined above, SP does not account for interests which are not satisfied due to the unavailability of relevant publications. SP accounts only for interests which are not satisfied, while EF would have satisfied the interest assuming unlimited caching resources at the receivers.
- **Message traffic**, which is the total number of publication messages forwarded into the mediation network. We also characterize the message traffic by the **bandwidth saved**, which is how much bandwidth has been saved for a given scenario using e-CBN instead of EF. This metric is also indicative of the efficiency of the two duplicate dropping mechanisms (IDD and PDD) presented in Section 3.3.6.
- **Control traffic**, which is the number of interest messages forwarded into the mediation network. We characterize the control traffic also by the **control overhead**, which is the ratio between the control traffic generated by an instance of e-CBN and the control traffic generated by an instance of EF.

4.1. Workload characterization

When evaluating content-based publish/subscribe approaches, workload assumptions in terms of *popularity* and *locality* significantly impact measured performances. Due to the lack of publicly available datasets of large-scale CBPS systems, synthetic workload generation has been widely accepted. The resulting challenge is how to choose the parameters of the workload model in order to generate a workload consistent with a given application profile.

For instance, content-based routing (CBR) [45] is preferable over pure broadcast only in scenarios with a sparse density of receivers. For this reason, previous CBR research has been evaluated under specific *popularity* and *locality* assumptions.

The seminal work on content-based routing [45] is evaluated assuming that the density of receivers equals 75%, and the popularity distribution is characterized by the matching message distribution, which represents the number of messages matching a percentage of predicates. Most messages match 5–15% of the predicates, a significant number of messages do not match any predicate, and no message matches more than 25% of the predicates. Another significant work, [46] was characterized for comprehensiveness with four different *popularity* and *volume* distributions reported in the publish/subscribe literature.

Majumder et al. study in [47] the impact of locality patterns on the communication-efficiency of several clustering algorithms for content-based routing. The workload is tuned from a localized subscription model, *i.e.* similar subscriptions originating from the same region, to a uniform model. In fact, the locality of similar subscriptions has a significant impact on the efficiency of multicast-based schemes and the efficiency of optimizations such as subscription covering [40].

The sensitivity of topologies on the communication-efficiency of content-based routing schemes is discussed in [48], where a run-time algorithm to adapt the topology to the application demand is proposed.

A very interesting fact about the importance of workload assumptions on the conclusions one can draw from the evaluation of its solution is the positioning of Riabov et al. in [49] regarding the conclusions on the benefits of multicast in one of the Gryphon papers [50]. Riabov et al., demonstrated that the conclusions from the Gryphon papers were not always true and that they depend

Table 3
Acronyms used in the description of the e-CBN framework. In brackets is the section where each acronym was introduced.

Acr.	Name	Acr.	Name
PIT	Pending interests table (Section 3.3.1)	LRU	Least recently used (Section 3.3.2)
PPT	Pending publications table (Section 3.3.1)	MFU	Most frequently used (Section 3.3.2)
OPT	Overflowing publications table (Section 3.3.1)	MF	Most fresh (Section 3.3.2)
DPT	Disseminated publications table (Section 3.3.1)	LF	Least fresh (Section 3.3.2)
BT	Boxes table (Section 3.3.1)	NRT	En-route caching (Section 3.3.2)
DF	Dispatched flag (Section 3.3.1)	EON	Explicit overload notification (Section 3.3.4)
PF	Pending flag (Section 3.3.1)	PDP	Pull/delayed push (Section 3.3.4)
DPF	Disseminated publications first (Section 3.3.2)	IDD	In-network duplicate dropping (Section 3.3.6)
PPF	Pending publications first (Section 3.3.2)	PDD	Proactive duplicate dropping (Section 3.3.6)
MRU	Most recently used (Section 3.3.2)		

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

on the assumptions made on locality and similarity properties.

4.2. Evaluation methodology

For the evaluation of the framework, we assume a news dissemination application such as Google news implemented with e-CBN distributed over a network of MRs. The workload is generated in order to meet this application profile. In order to validate our algorithms and design choices, we implemented the framework in the PEERSIM [51] simulator which is a common choice for the evaluation of large-scale publish/subscribe solutions.

We consider that a set of events generate both publications and interests, and that three parameters characterize each event: *popularity*, *locality* and *volume*. The popularity of an event refers to the number of interests related to it, its volume to the number of related publications and its locality to the regions of the topology likely to originate related interests and publications. A similar methodology is used in [46]. We use this methodology because it fits the target application model, and because it allows an easier control of popularity and locality assumptions.

In the generation of our workload, where the parameter is not the size of the network, we assume an arbitrary MR topology of 100 nodes characterized by average degree 4, diameter 6 and following a power-law distribution with few nodes of high degree and many nodes of small degree (see Fig. 11). We assume that each MR is provisioned with some amount of caching resources and can cache up to CS objects.

We assume that most popular events are characterized by larger and broader audiences in terms of number of interests, and are also likely to trigger larger volume of publications. In other terms, event popularity and volume are generated according to the same distribution. Note that we evaluate only loose subscriptions.

We investigate two different locality patterns for the workload generation:

Random: Publications and interests originate from random locations.
Realistic: Most popular events are more widely spread in the topology.

Let us consider e_i ($1 \leqslant i \leqslant E$), an event of popularity p_i, volume v_i and locality l_i. p_i is sampled from a power-law distribution and the volume v_i is such that $v_i = P * p_i$. When the random locality pattern is enforced, publications and interests associated to e_i can be issued by any of the N MRs. When the realistic locality pattern is enforced, publications and interests associated to e_i can be issued only from a set of nodes computed using l_i. In that case, we define l_i such that $l_i = p_i$ and such that $\lceil l_i * N \rceil$ MRs are potential issuers (hosting interested receivers) of interests related to e_i. This set of MRs is computed by choosing a random root node and $\lceil l_i * N \rceil - 1$ additional nodes among the closest MRs to the root. We assume a constant arrival rate for interests (resp. publications) equals to $r_s = S/T$ (resp. $r_p = P/T$).

For each publication (resp. interest), we randomly select an event and a location among the MRs eligible to generate traffic related to that event. When the volume (resp. popularity) associated to an event is reached, it is removed from the set of events that can be used to generate new publications (resp. interests). We generate selectivity and lifetime values associated to interests according to Poisson distributions of average reported by Table 4. Randomizing selectivity and lifetime values is necessary to model heterogeneous consumer requirements.

We assume that each interest is refreshed at the end of its lifetime with a probability P_r called *refreshed probability*, otherwise a new interest related to the same event is generated. For the simulation of PDP, we set β to 1, which is an upper bound of the end-to-end transfer delay between any pair of MRs in absence of bottleneck.

Simulation time is set to 100. Consequently, every time unit, 500 new publications are uploaded followed by the generation of 1000 new interests. Note that the PEERSIM simulator does not provide any guarantee about the order in which the scheduler processes the scheduled events.

Evaluation settings. In order to model the event popularity distribution used to generate the traffic, we measured the volume distribution of the top stories reported by the French front page of Google news during three consecutive days. It is interesting to note that on the 14th April, the top story reached a popularity of 45% (see Table 5) illustrating the fact a single event may represent a major fraction of traffic.

Fig. 12 shows that the measured volume distribution can be approximated by a power-law distribution of exponent 1.0 (12th April) or 2.4 (14th April). For workload generation, we chose exponent 1.0 in order to avoid a too much favorable scenario, where a single event would generate most of the traffic. The peak popularity equals 30% followed by a peak of 10% *i.e.* the two most popular events generate 40% of publications and subscriptions traffic. We use the same series of event popularity in all our experiments. In fact, Fig. 13c shows that there is much volatility in the generated series and demonstrates the importance to use the same series for the comprehensiveness of the results.

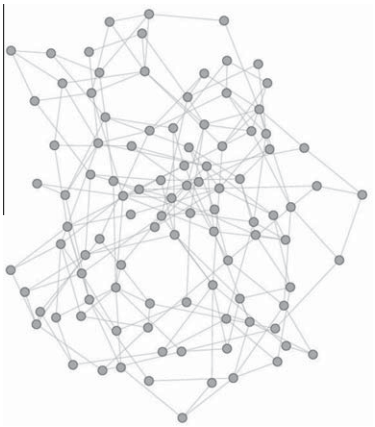


Fig. 11. Evaluation topology.

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

Regarding the refresh probability, we choose $P_r = 1.0$ *i.e.* each interest is automatically refreshed after expiration. This is inline with the fact that RSS traffic has been reported to be sticky contrarily to web traffic [52]. The number of events and publications simulated has been sized on the basis of measures reported in Table 4. Without explicit mention, the reader should refer to Table 4 for the default parameters values.

Workload characterization. We characterize the workload using three distributions: *matching distribution*, *density of recipients* and *gap distribution*. The matching distribution represents the CCDF of the popularity distribution of publications, which associates each publication to the percentage of matching interests. The density of recipients represents the CCDF of the topological popularity distribution of publications, which associates each publication to the percentage of MRs which have requested the publication. The gap distribution characterizes the gap existing between consumers' attention span and the volume of information available.

Fig. 13b characterizes the density of recipients and shows that no publication is requested by more than 24% of MRs with realistic locality enabled. Conversely, with random locality enabled, most publications have a topological popularity superior or equal to 90%. Realistic locality provides a less extreme scenario than random locality. Moreover, this is inline with what has been reported from Le Fessant et al. [53], that for 60% of files in an e-Donkey network, more than 80% of the replicas were in the same country. Fig. 13a shows that in the workload there is no publication with popularity larger than 2.5%. The gap distribution depicted by Fig. 13d is positive and describes a situation of information overload.

4.2.1. Results

Comparison of an instance of the framework to a variant of EF. First, we compare an instance of e-CBN to EF for different load levels obtained by varying the selectivity and with realistic locality enabled.

In Fig. 14a, e-CBN satisfies 13% more interests than EF for various load levels. e-CBN satisfies more interests than EF because e-CBN uses cached publications to satisfy the interests while EF does not take advantage of the cache. Consequently, more publications are available with e-CBN than with EF. Fig. 14b shows that for small selectivity values, e-CBN saves almost 100% of bandwidth and up to 40% of bandwidth for larger selectivity values. These gains are obtained while still generating less control traffic than EF.

Fig. 15 shows that with the MF policy and larger cache sizes, more interests are satisfied (Fig. 15a), and communication-efficiency is improved. Note that Fig. 15a does not display the percentage of interests satisfied by EF because it is obvious that EF does not take advantage of the cache to satisfy more interests (see Fig. 14a). Also, there is no incentive to add caching resources beyond 1% of the total volume of publications. This is not a surprising result, since with more caching resources, MRs have more opportunities to satisfy local interests with cached publications.

Finally, Fig. 16a shows that the when the size of the network increases also increases the satisfaction ratio of the two methods. Of course this increase is not very significant (double satisfaction ratio when the number of nodes is six times larger). More nodes/brokers might mean more different cached messages, but on the other hand more brokers also increase the replication degree of the cached messages which does not allow the retrieval of unique cached items. At any case we observe that e-CBN performs on average 10–34% better than the EF. Also, is obvious (see Fig. 16b) that for the default selectivity value and the usage of the two duplicate dropping mechanisms the e-CBN saves almost 88% of bandwidth for small networks and up to 98% of bandwidth for larger network, while generating less control traffic (see Fig. 16c).

Dissemination methods evaluation. Fig. 17 compares the performances of PDP to EON for different load levels. We observe that PDP and EON display close performances with respect to the quality of service (less than 2% even for extremely large selectivity values) and the amount of traffic generated by the content-based network. It is noticeable that PDP generates slightly less control traffic than EON.

Enhanced forwarding scheme evaluation. We compare the performances of the enhanced forwarding scheme (CBF) to a simple FIFO policy for different buffer sizes. Fig. 18a shows that FIFO achieves a bad quality of service and does not take advantage of larger buffer sizes. CBF instead takes better advantage of the available bandwidth and processing capacity to improve the QoS offered to consumers.

4.3. Caching policies evaluation

We observe that all selection/replacement policies achieve comparable quality of services for a wide range of cache sizes and within acceptable ranges (Fig. 19a). MF

Table 4
Default parameter values for the evaluation.

Parameter	Default value
Dissemination policy	EON
Number of interests	100,000
Number of publications	50,000
Cache size (CS)	500
Refresh probability	1.0
Number of events	50
Zipf exponent	1.0
Buffer size	50000
Average selectivity	10
Simulation time	100
Average lifetime	20
Freshness	50
Network Size	100
Locality pattern	Realistic
Selection/replacement policy	MF
Caching policy	Default
β	1

Table 5
Statistics from the front page of Google news.

Date	Monitored events	Volume	Peak popularity (%)
12-04-2011	33	9507	30
13-04-2011	31	7941	40
14-04-2011	29	8875	45

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

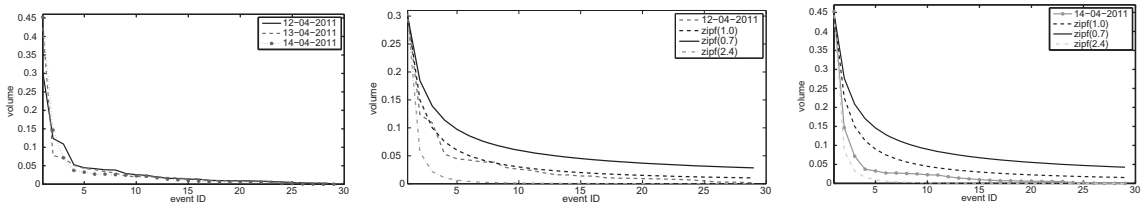
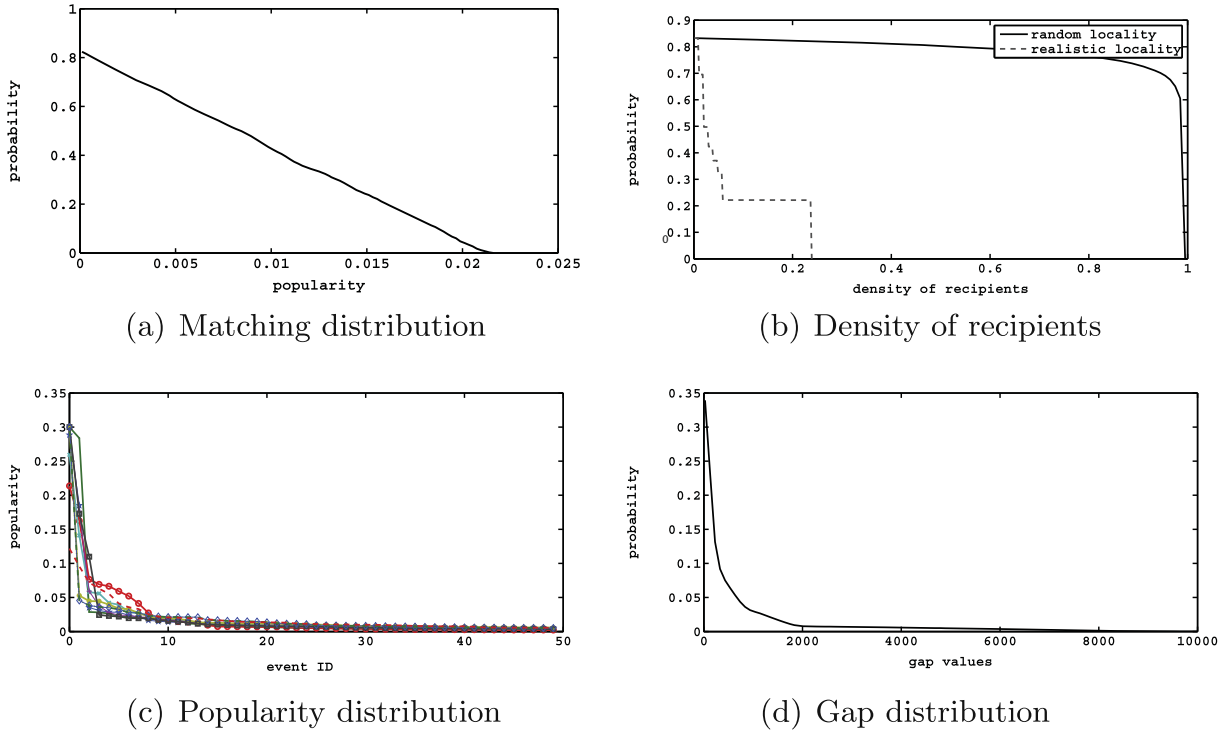


Fig. 12. Measures from Google news. Volume distribution (left), volume distribution fitting I (middle), volume distribution fitting II (right).



(a) Matching distribution (b) Density of recipients

(c) Popularity distribution (d) Gap distribution

Fig. 13. Workload characterization.

and LF perform slightly better than other policies with respect to the quality of service offered. It is worth mentioning the exceptional performances of PPF which generates very low publication message traffic, while providing a satisfying quality of service. MRU and DPF are the less communication-efficient policies. The better performances of PPF over DPF can be explained by the fact that PPF is more successful than DPF in guaranteeing the availability of pending publications for refreshed interests.

MF and LF achieve a good trade-off between quality of service and com-munication-efficiency, while still generating a lower control overhead than other policies (see Fig. 19). We did not observe a significant impact of the caching policies on the communication gains excepted for MF, LF and MRU. Also, LFU generates significantly less control traffic as cache size increases.

Also in the experiments, we did not observe any improvement with en-route caching enabled. We will recommend using the default caching policy for experiments in realistic configurations.

Impact of freshness on caching policies. We evaluate the impact of freshness on MRU, MF and LF policies that all provide a bad quality of service, with starvation probability larger than 10%, for small freshness values (Fig. 20). For the same freshness values, less message traffic is generated with a higher control overhead. This is due to the fact that with lower freshness values, there are less publications available in the mediation network. As a consequence, more interests are forwarded into the mediation network. Finally, the MRU policy is less communication-efficient than the MF and LF policies for a wide-range of freshness values.

Impact of background traffic. We inject voluntarily a volume of publications that will never be requested *i.e.* will remain pending, and investigate the resilience of the caching policies w.r.t. to this kind of traffic. In Fig. 21a, we observe a slight increase of the starvation probability when increasing the percentage of background traffic. But, all policies achieve less than 1% of starvation even for 70% of background traffic. In Fig. 21b, we observe that the communication gains of PPF, LFU, MRU are insensitive to the

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

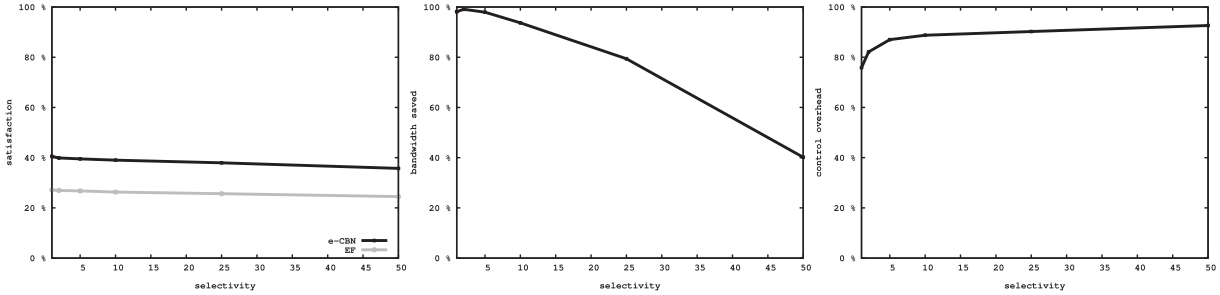


Fig. 14. Performance comparison of an instance of e-CBN to EF for different load levels.

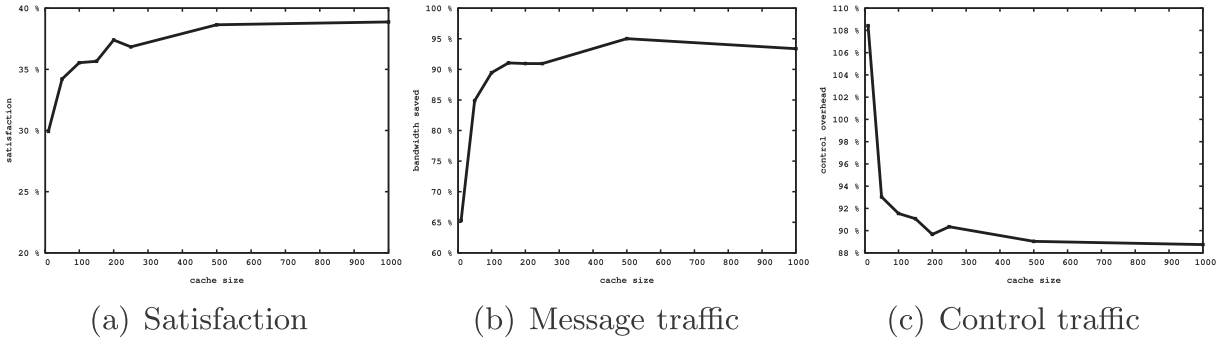


Fig. 15. Performance comparison of an instance of e-CBN to EF for different cache sizes.

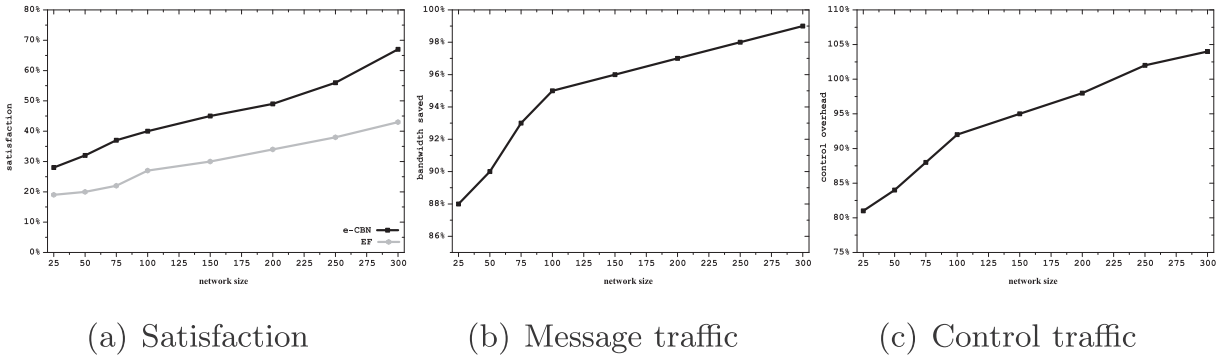


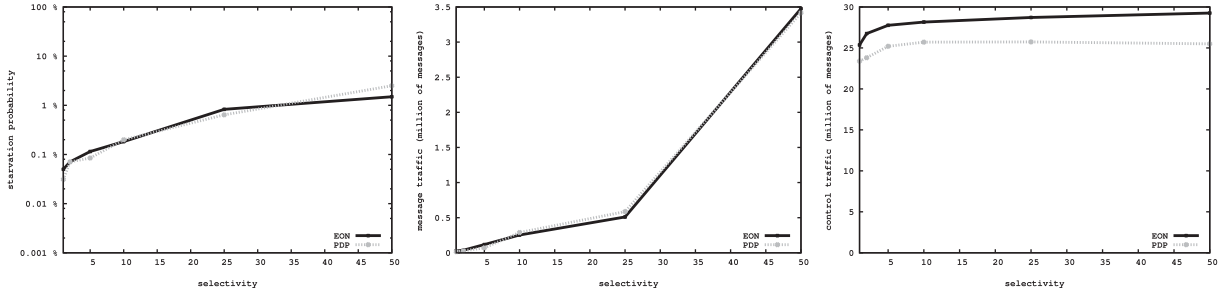
Fig. 16. Performance comparison of an instance of e-CBN to EF for different sizes of the network.

fraction of background traffic injected. Instead, the communication gains of MF, LF and DPF decreases significantly with the ratio of background traffic injected. The bad performances of DPF results from the fact that it gives higher priority in the cache to background traffic over dispatched publications which have a higher probability to be useful to future interests.

Impact of refresh probability. The proportion of refreshed interests over new interests increase with the refreshed probability. Fig. 22 shows that DPF communication gains drastically collapse when there are fewer refreshed interests. This may be due to the fact that when new interests dominate the traffic, they consume pending publications faster, reducing the opportunities for refreshed ones additionally to the fact that DPF is less efficient than PPF in guaranteeing the availability of pending publications. PPF generates very low communication traffic even for small values of the refresh probability. We also observed that for all the evaluated policies in Fig. 22, the control traffic increases with the number of refreshed interests. This is a consequence of the fact, that refreshed interests have less opportunities to be satisfied than new ones.

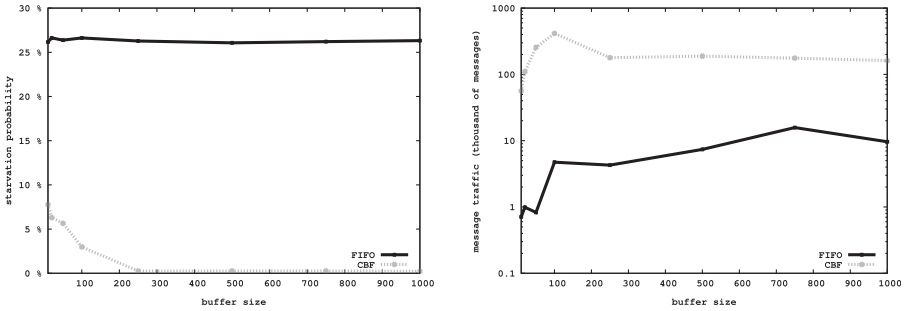
The evaluation of the framework under realistic workload assumptions reveals that e-CBN improves signifi-

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009



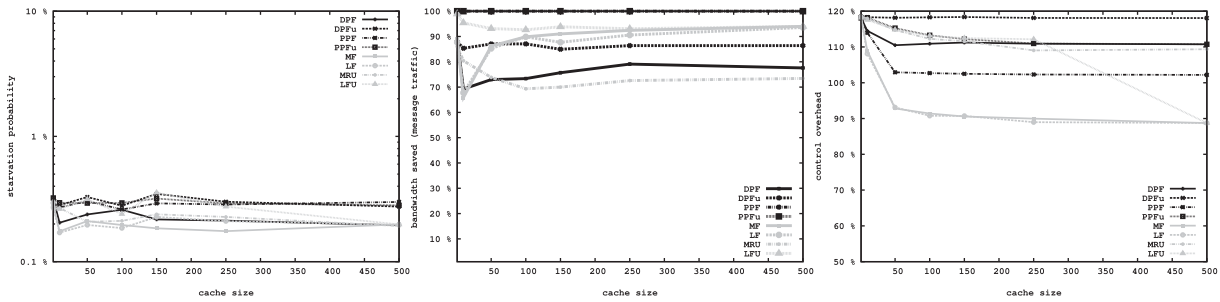
(a) Starvation (b) Message traffic (c) Control traffic

Fig. 17. Performance comparison of EON and PDP.



(a) Starvation (b) Message traffic

Fig. 18. CBF performances.



(a) Starvation (b) Message traffic (c) Control traffic

Fig. 19. Performance evaluation of selection/replacement policies.

cantly the communication-efficiency of content-based networking, while providing a high-quality of service. PPF, MF and LF achieve an excellent trade-off between quality of service and communication-efficiency for a wide-range of configurations and the CBF forwarding algorithm takes better advantage of the available processing and forwarding resources than a simple FIFO policy. Finally, PDP generates slightly less control traffic than EON.

5. Conclusions and future work

In this paper we proposed a service model for large-scale content based publish/subscribe networking. CPBS is characterized at large-scale by the heterogeneity of information consumer requirements, the information overload and the scarcity of end-to-end bandwidth. The proposed service model, called e-CBN, aims at leveraging caching in order to meet the above characteristics. The evaluation of the model under realistic workload assumptions reveals that e-CBN improves significantly the communication efficiency of CBPS networking, while providing a high quality of service. As future work, we plan to combine the proposed content and interest forwarding strategies and caching policies with CDN-like replication techniques designed for CBPS networks and enable mobility for both the information consumers and the mediation routers.

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

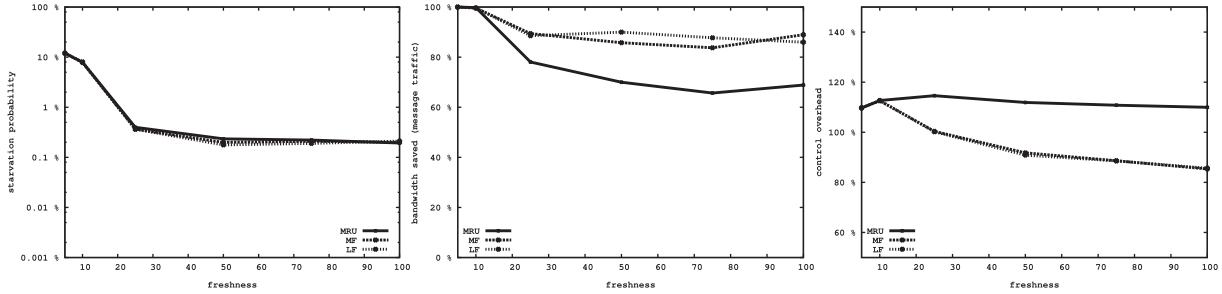


Fig. 20. Impact of freshness on caching policies (`CS = 100`).

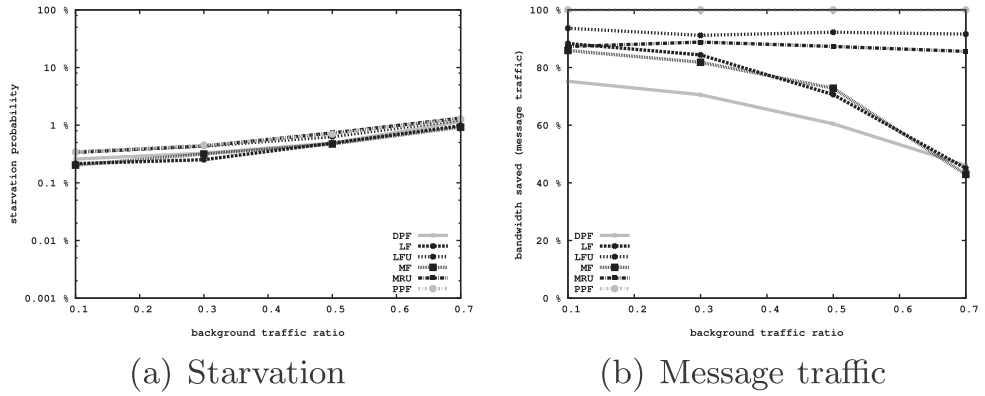


Fig. 21. Impact of background traffic on caching policies (`CS = 100`).

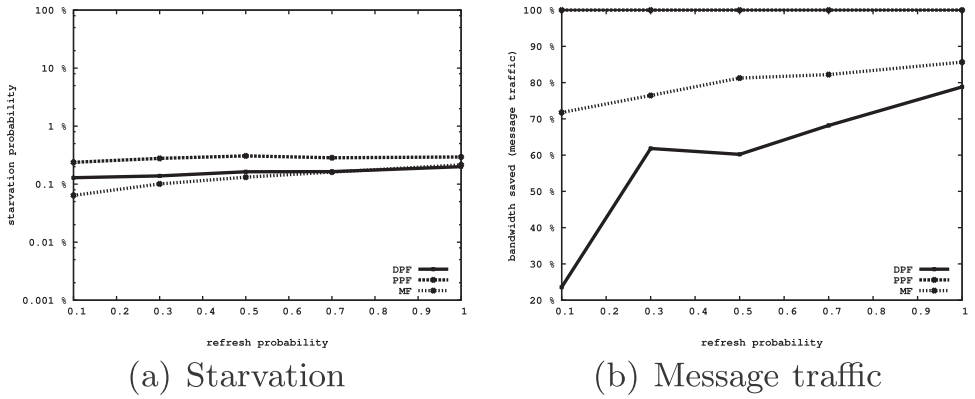


Fig. 22. Impact of refresh probability on caching policies (`CS = 100`).

Acknowledgments

V. Sourlas' research has been co-financed by the European Union (European Social Fund, ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) – Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund. This work has also been supported by the European Commission through the FP7 PURSUIT Program, under Contract ICT-2010-257217.

References

[1] S. Fdida, M. Diallo, The network is a database, in: Proc. 4th Asian Conference on Internet Engineering, Bangkok, Thailand, November 2008, pp. 18–20.

[2] M. Diallo, S. Fdida, IOA-CBR: information overload-aware content-based routing, in: Proc. 4th ACM Inter. Conference on Distributed Event-Based Systems, Cambridge, UK, 2010.
[3] M. Diallo, S. Fdida, Avalanche: towards a scalable content-based pub/sub network service, in: Proc. of the 3rd ACM Inter. Conference on Distributed Event-Based Systems, Nashville, TN, 2009.
[4] M. Diallo, S. Fdida, V. Sourlas, P. Flegkas, L. Tassiulas, Leveraging caching for Internet-scale content-based publish/subscribe networks, in: IEEE ICC 2011, Kyoto, Japan, June 2011.
[5] M.K. Aguilera, R.E. Strom, D.C. Sturman, M. Astley, T.D. Chandra, Matching events in a content-based subscription system, in: 18th ACM Symposium on Principles of Distributed Computing (PODC '99) Atlanta, GA, May 4–6, 1999, pp. 53–61.
[6] A. Carzaniga, D. Rosenblum, A. Wolf, Design and evaluation of a wide-area event notification service, ACM Transaction on Computer Systems 19 (2001) 332–383.
[7] B. Segall, D. Arnold, Elvin has left the building: a publish/subscribe notification service with quenching, in: Proceedings of AUUG97, Brisbane, Australia, September 3–5, 1997, pp. 243–255.
[8] G. Cugola and G. Picco, REDS, a reconfigurable dispatching system, in: 6th International workshop on Software Engineering and Middleware, Oregon, 2006, pp. 9–16.
[9] PURSUIT Project. <http://www.fp7-pursuit.eu>.
[10] Named Data Networking (NDN) Project. <http://named-data.net>.
[11] SAIL Project. <http://www.sail-project.eu/>.
[12] http://pubsubhubbub.googlecode.com/svn/trunk/pubsubhubbub-core-0.3.html.
[13] D. Wessels, K. Claffy, Applications of Internet Cache Protocol (ICP), v.2, ITF, May 1997. <http://tools.ietf.org/html/draft-wessels-icp-v2-02>.
[14] P. Vixie, D. Wessels, RFC 2756: Hyper Text Caching Protocol, January 2000.
[15] M. Pitkanen, J. Ott, Enabling opportunistic storage for mobile DTNs, Elsevier, Pervasive and Mobile Computing 4 (2008) 579–594.
[16] A. Anand, A. Gupta, A. Akella, S. Seshan, S. Shenker, Packet caches on routers: the implications of universal redundant traffic elimination, SIGCOMM Computer Communication Review 38 (4) (2008) 219–230.
[17] T. Ballardie, P. Francis, J. Crowcroft, Core based trees (CBT), SIGCOMM Computer Communication Review 23 (4) (1993) 8595.
[18] P. Srebrny, T. Plagemann, V. Goebel, A. Mauthe, CacheCast: eliminating redundant link traffic for single source multiple destination transfers, in: 2010 International Conference on Distributed Computing Systems, pp. 209–220.
[19] Y. Zhu, M. Chen, A. Nakao, CONIC: content-oriented network with indexed caching, in: IEEE INFOCOM, 15–19 March 2010; pp.1-6.
[20] L. Dong, H. Liu, Y. Zhang, S. Paul, D. Raychaudhuri, On the cache-and-forward network architecture, in: ICC '09, 14–18 June 2009.
[21] L. Muscariello, G. Carofiglio, M. Gallo, Bandwidth and storage sharing performance in information centric networking, in: ACM SIGCOMM ICN Workshop, 2011, pp. 2631.
[22] U. Lee, I. Rimac, V. Hilt, Greening the internet with content-centric networking, eEnergy (2010) 179.
[23] E.J. Rosensweig, J. Kurose, Breadcrumbs: efficient, best-effort content location in cache networks, in INFOCOM, 2009.
[24] V. Jacobson, D.K. Smetters, J.D. Thornton, M.F. Plass, N. Briggs, R. Braynard, Networking named content, in: CoNEXT 2009, Rome, Italy, December 1–4, 2009.
[25] D. Perino, M. Varvello, A reality check for content centric networking, in: ACM SIGCOMM ICN Workshop, 2011, pp. 4449.
[26] S. Arianfar, P. Nikander, J. Ott, On content-centric router design and implications, in: ReArch Workshop, vol. 9, ACM, 2010, p. 5.
[27] Al. Ghodsi, T. Koponen, B. Raghavan, S. Shenker, An. Singla, J. Wilcox, Information-centric networking: seeing the forest for the trees, in: ACM Workshop on Hot Topics in Networks (HotNets-X), Cambridge, MA, November 2011.
[28] W.K. Chai, I. Psaras, G. Pavlou, Cache less for more in information-centric networks, in: IFIP NETWORKING 2012, Prague, Czech Republic, May 2012.
[29] S. Wang, J. Bi, J. Wu, Z. Li, W. Zhang, X. Yang, Could in-network caching benefit information-centric networking? in: Proc. of AINTEC '11.
[30] Z. Li, G. Simon, Time-shifted TV in content centric networks: the case for cooperative in-network caching, in: Proc. of IEEE ICC 2011.
[31] C. Fricker, P. Robert, J. Roberts, N. Sbihi, Impact of traffic mix on caching performance in a content-centric network, in: Proc. of IEEE NOMEN 2012.
[32] G. Carofiglio, V. Gehlen, D. Perino, Experimental evaluation of memory management in content-centric networking, in: Proc. of IEEE ICC2011.
[33] N.J. Belkin, W.B. Croft, Information filtering and information retrieval: two sides of the same coin?, Communications of the ACM (1992) 29–38
[34] T.W. Yan, H. Garcia-Molina, SIFT: a tool for wide-area information dissemination, in: Proc. of the USENIX 1995 Technical Conference Proceedings (TCON'95), 1995, pp. 15–15.
[35] T.W. Yan, H. Garcia-Molina, The SIFT information dissemination system, ACM Transactions on Database Systems (TODS) (1999) 529–565.
[36] M. Chen, A. LaPaugh, J. Pal Singh, Content distribution for publish/subscribe services, in: Proc. of the ACM/IFIP/USENIX 2003 International Conference on Middleware (Middleware '03), 2003, pp. 83–102.
[37] V. Sourlas, G.S. Paschos, P. Flegkas, L. Tassiulas, Caching in content-based publish/subscribe systems, in: IEEE GLOBECOM, 2009.
[38] V. Ramasubramanian, R. Peterson, E. Gün Sirer, Corona: a high performance publish-subscribe system for the world wide web, in: Proc. of the 3rd Conference on Networked Systems Design & Implementation (NSDI'06), vol. 3, 2006.
[39] A. Carzaniga, A.L. Wolf, Forwarding in a content-based network, in: Proc. of the 2003 Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications, ACM, 2003, pp. 163–174.
[40] S. Tarkoma, J. Kangasharju, Optimizing content-based routers: posets and forests, Distributed Computing, vol. 19, Springer, 2006, pp. 62–77.
[41] A. Carzaniga, A.J. Rembert, A.L. Wolf, Understanding content-based routing schemes, 2006.
[42] C.P. Hall, A. Carzaniga, A.L. Wolf, DV/DRP: a content-based networking protocol for sensor networks, 2006.
[43] P. Jokela, A. Zahemsky, C. Esteve Rothenberg, S. Arianfar, P. Nikkander, LIPSIN: line speed publish/subscribe inter-networking, in: Proc. of the ACM SIGCOMM 2009 Conf. on Data communication, ACM, 2009, pp. 195–206.
[44] B. Yang, H. Garcia Molina, Improving search in peer-to-peer networks, in: 22nd International Conference on Distributed Computing Systems, 2002, pp. 5–14.
[45] A. Carzaniga, M.J. Rutherford, A.L. Wolf, A routing scheme for content-based networking, in: INFOCOM, 2004, pp. 918–928.
[46] F. Cao, J.P. Singh, Efficient event routing in content-based publish-subscribe service networks, in: Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM, 2004, pp. 929–940.
[47] A. Majumder, N. Shrivastava, R. Rastogi, A. Srinivasan, Scalable content-based routing in pub/sub systems, in: INFOCOM 2009, pp. 567–575.
[48] M. Migliavacca, G. Cugola, Adapting publish-subscribe routing to traffic demands, in: Proc. of the 2007 Inaugural International Conf. on Distributed Event-Based Systems (DEBS 2007), ACM, 2007, pp. 91–96.
[49] A. Riabov, Z. Liu, J.L. Wolf, P.S. Yu, L. Zhang, Clustering Algorithms for Content-Based Publication-Subscription Systems, IEEE Computer Society, 2002.
[50] L. Opyrchal, M. Astley, J. Auerbach, G. Banavar, R. Strom, D. Sturman, Exploiting IP multicast in content-based publish-subscribe systems, in: IFIP/ACM International Conference on Distributed Systems Platforms, Springer-Verlag, New York, 2000, pp. 185–207.
[51] http://peersim.sourceforge.net/.
[52] H. Liu, E.G. Sirer, A measurement study of a publish Subscribe System, 2005.
[53] F. Le Fessant, S. Handurukande, A. Kermarrec, L. Massoulie, Clustering in peer-to-peer file sharing workloads, in: Lecture Notes in Computer Science, Peer-to-Peer Systems III, vol. 3279, Springer, 2005, pp. 217–226.

Mohamed Diallo received his Ph.D. from the Network and Performance Analysis (NPA) group of the LIP6 laboratory, advised by Professor Serge Fdida in 2011. He also obtained an M.S. degree in Computer Networks with honors from UPMC in 2007, and an engineering degree in Telecommunications from INPT Rabat in 2006.

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009

Vasilis Sourlas was born in Athens, Greece, in 1980. He received his Diploma degree from the Computer Engineering and Informatics Department, University of Patras, Greece, in 2004 and the M.Sc. degree in Computer Science and Engineering of the Computer Engineering and Informatics Department, University of Patras, Greece in 2006. Since 2007 he is a Ph.D. student in the Department of Computer and Communication Engineering, University of Thessaly (Volos), Greece.

Paris Flegkas is currently an adjunct lecturer and a post-doctoral researcher at the Department of Computer Engineering and Telecommunications, University of Thessaly, Greece. He received a Diploma in Electrical and Computer Engineering from the Aristotle University, Thessaloniki, Greece, an M.Sc. with distinction in Telematics (Communications & Software) and a Ph.D. from the University of Surrey, UK, in 1998, 1999 and 2005 respectively. He has been a Research Fellow working in EU and UK national projects for more that 4 years in the Centre for Communications Systems Research (CCSR), UK and 4 years working on EU and Greek projects in CERTH. His research interests are in the areas of policy-based networking, management technologies, traffic engineering, service management, IP QoS and publish/subscribe networks.

Serge Fdida is a professor at the University Pierre et Marie Curie (Paris) since 1995. He received his Ph.D. in 1984, and the Habilitation à Diriger des Recherches (HDR) in Modelling of Computer Networks in 1989 from the University Pierre et Marie Curie. From 1989 to 1995, he was a Full Professor at the University Rene Descartes (Paris). His research interests are in the area of high speed and mobile networking, pervasive communication, resource management and performance analysis. He is heading the Network and Performance Analysis group of the LIP6 Laboratory (CNRS-University Pierre and Marie Curie). Professor Fdida was a Visiting Scientist at IBM Research (Raleigh, USA) during the 1990/1991 academic year. Serge Fdida has been an advisor for the Communication and Information Science and Technology (STIC) Department of CNRS since it was established in 2000 up to 2005. He was also a member of the CNRS National Committee (Section 7), the Evaluation Committee at INRIA. He has been a reviewer for the NSF and the European Commission among others. He also was the Vice-President of RNRT (Rèseau National de la Recherche en Tèlècommunications). Serge Fdida is a senior member of IEEE, a member of ACM and also involved in two IFIP working groups on networking. Serge Fdida was the Co-Director of EURONETLAB, a joint laboratory established in 2001 up to 2007, between University Paris 6, CNRS, THALES and BLUWAN, developing research and development work on "QoS Routers", "Radio Routers" and Security. Finally Serge Fdida was a co-founder of the QOSMOS company.

Leandros Tassiulas obtained the Diploma in Electrical Engineering from the Aristotelian University of Thessaloniki, Greece in 1987, and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Maryland, College Park in 1989 and 1991 respectively. He is Professor in the Dept. of Computer and Telecommunications Engineering, University of Thessaly, since 2002. He has held positions as Assistant Professor at Polytechnic University New York (1991–1995), Assistant and Associate Professor University of Maryland College Park (1995–2001) and Professor University of Ioannina Greece (1999–2001). His research interests are in the field of computer and communication networks with emphasis on fundamental mathematical models, architectures and protocols of wireless systems, sensor networks, high-speed Internet and satellite communications. He is a Fellow of IEEE. He received a National Science Foundation (NSF) Research Initiation Award in 1992, an NSF CAREER Award in 1995, an Office of Naval Research, Young Investigator Award in 1997 and a Bodosaki Foundation award in 1999. He also received the INFOCOM 1994 best paper award and the INFOCOM 2007 achievement award.

Please cite this article in press as: M. Diallo et al., A content-based publish/subscribe framework for large-scale content delivery, Comput. Netw. (2012), http://dx.doi.org/10.1016/j.comnet.2012.11.009