INVITED PAPER *Special Section on New Paradigm on Content Distribution and Sharing*

# Towards Extreme Scale Content-Based Networking for the Next Generation Internet

**Mohamed DIALLO**[†a)], **Serge FDIDA**[†], *and* **Prométhée SPATHIS**[†], *Nonmembers*

**SUMMARY** In this paper, we are concerned about content-based networking (CBN) at extreme scales, characterized by a large number of widely spread consumers, heterogeneous consumer requirements, huge volume of publications, and the scarcity of end-to-end bandwidth. We extend CBN with a generic service model that allows consumers to express their interests in future publications including cached content, but also to quantify the maximum amount of information they are willing to consume. We take advantage of this knowledge to pace the dissemination process and therefore, enhance the service efficiency. Early evaluation results show gains of up to 80% compared to a baseline CBN model.
***key words:*** *publish-subscribe, content-based networking, corres-ponding*

## 1. Introduction

Many recent studies on the current Internet have focused on the analysis of its major flaws as well as its main challenges in the years to come. They all come to the conclusion that opposite forces are trying to reshape the Internet in regard to its new usages, resulting in requirements hard to satisfy with the current architecture (trust, security, mobility, data-centric, ...). The design principles enforced in the Internet architecture 40 years ago, such as host-centricism, packet switching or stateless packet switches, were motivated by well-defined priorities [3]. Nowadays, content retrieval is the principal source of traffic in the Internet. This activity involves several layers of indirections including DNS resolution, which may become a bottleneck if improperly engineered. Moreover, traffic crossing the middle mile of the Internet usually suffers from significant delays and losses [4]. As a consequence, caching proxies, DNS caching and content distribution networks (CDN) have emerged as solutions of choice for meeting the requirements of content-related applications at global scale. Following this trend, the receiver-driven approach has been introduced to fill data-intensive applications needs. This approach is in contrast with source-driven communications that rely on host reachability. Instead, the receiver-driven approach focuses on content access resulting in the choice of stateful switches and in-networking caching. In this context, the flow of data is implicitly steered according to receivers' registered interests, rather than explicitly relying on the binding of the corresponding data objects to their host location. Data objects are directly addressed through names, which may offer different levels of expressiveness. Several receiver-driven architectures and research projects have emerged under the banner of information-centric networking as a consensus label in the corner of the research community working on receiver-driven architectures for global communications*. We discuss below the most relevant ones:

**PSIRP,** the publish/subscribe Internet routing paradigm aims at providing global scale mediation between information providers and consumers [6]. In PSIRP, information can be anything including data chunks. One original aspect of PSIRP is to associate explicitly every publication to a scope, which can be geographical (e.g. campus) or logical (e.g. social network). Scopes are associated with policies that can be used to implement access control mechanisms enforced by trusted rendezvous nodes [29]. PSIRP separates the control plane from the data plane, and already claims topic-based publish/subscribe forwarding at line-speed [12].

**DONA,** the data-oriented network architecture [30] allows a client to request a piece of data by its name (a flat self- certifying label), rather than the owner's address. The architecture supports two basic primitives FIND and REGISTER. To support these primitives, DONA introduces Data Handlers (DHs), which are the network entities responsible for the name resolution and data caching functions. Collectively, DHs assume the responsibility for routing clients' requests to nearby copies of data.

**NetInf,** [26] is an architecture proposed as part of the 4WARD project [5]. A piece of information in NetInf is represented by an Information Object (IO) and stored as one or more Bit-level Objects (BO). IOs may be structured into hierarchies and metadata can be provided as sets of attribute-value pairs. NetInf provides name resolution, search and event services. For instance, NetInf users can search IOs using SPARQL [28] and register to updates about a collection of objects using a SIENA-based [27] event service.

**Content-centric networking (CCN)** is a receiver-driven architecture [7], where receivers request data chunks using self-certifying names. Chunks may be retrieved either from the original server or in-network caches. For adoption and scalability reasons, CCN borrows many principles from IP networks including routing protocols and prefix-based forwarding.

Manuscript received March 24, 2011.
Manuscript revised May 23, 2011.
[†]The authors are with UPMC Sorbonne Universités, France.
a) E-mail: mohamed.diallo@lip6.fr
DOI: 10.1587/transcom.E94.B.2706

*From a conversation with Jim Thornton.

**Content-based networking (CBN)** [2] enables receivers to register their information interests according to expressive subscription languages. As a consequence, CBN implements a decentralized content-based publish/subscribe (CBPS) communication model and provides efficient dissemination channels for a wide range of applications such as event notification, news dissemination or file sharing services. In fact, CBN has basically been an attempt to push SIENA-like middlewares down the stack.

We believe that the Next Generation Internet (NGI) will provide a substrate supporting several communication architectures, including TCP/IP. It will enforce isolation, as well as interoperability depending on stakeholders policies. The NGI will involve network operators participating in the basic substrate, and virtual operators deploying and managing global communication networks. Experimentation platforms such as GENI[9], ONELAB[10] and openflow [8] are already implementing part of this vision. Competing receiver-driven architectures should demonstrate strong incentives for adoption by virtual network operators and users as well as achieve satisfying performances.

This paper focuses on increasing the communication-efficiency of content-based networking (CBN) operating on extreme-scale characterized by a large number of widely spread consumers with heterogeneous requirements, a large number of publications and the scarcity of end-to-end bandwidth. Existing CBN proposals [11] 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 quote or monitoring applications. However, for applications such as news distribution or content sharing, the amount of relevant publications available at global scale may be overwhelming as information consumers have limited attention span. Implementing the same exhaustive filtering semantic for these applications would result in a huge information overload and communication overhead.

Our contributions are threefold. *First*, we extend CBN by defining a generic service model that allows consumers to express their interests in already cached or coming up publications, but also to quantify their attention capabilities. *Secondly*, we develop several strategies that take advantage of this model to minimize communication costs. Early evaluation results show gains of up to 80% compared to a baseline CBN model for a wide range of consumers requirements. *Thirdly*, we conclude the paper with open research problems.

## 2. Content-Based Networking Overview

Content-based communications involve three types of entities: *receivers* or *information consumers*, *publishers* or *information providers* and *routers*. Routers are interconnected according to an arbitrary topology that captures the specific features of *content-based networks*. A receiver submits its interests by sending a subscription to the network where routers acting as proxies are in charge for returning the corresponding matching pieces of data. The first router to handle the subscription advertised by a receiver is called a *home router*. Publishers upload their publications so as 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 $S$. Routers cooperate to efficiently disseminate data items corresponding to uploaded publications towards receivers that have subscribed to this content.

*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 packet 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 [15], [16].

*Content-based routing* (CBR) is the distributed algorithm that collects, propagates, assembles and transforms 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 interest. 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.

Carzaniga et al. [17], 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. As a consequence, it is realistic to assume that such protocols operate only on top of a global spanning tree.

## 3. Proposal for Enhanced Efficiency

Baseline content-based routing schemes [11] flood predicates in the network to implement the associated exhaustive semantic. As a consequence, each subscription is replicated at all routers and all publications matching a subscription are delivered to the receiver. For many applications, information consumers have a limited attention span and are in most cases interested in few responses, while a huge volume of information is available at global scale. With the exhaustive semantic, it would be necessary for the usability of the service that receivers process a ranking function in order to display only the most relevant publications to consumers. However, ranking for decentralized publish/subscribe services is yet not well-understood and questionable. Actually, using an attribute/value schema reduces the need for ranking, and online filtering service such as `Google alerts` [1] poorly performs. Therefore, we can anticipate that implementing the exhaustive semantic at global scale would result in a huge communication overhead.

Shifting from the exhaustive semantic towards a semantic quantifying consumers attention capability, open new opportunities to reduce the amount of traffic carried by content-based networks as well as to limit routers forwarding tables complexity and communication delays. We also believe that supporting non-persistent requests is an interesting feature which can only increase the attractivity of content-based networks, which requires that routers be provisioned with a caching memory. This section describes a new service model allowing receivers to quantify their attention span and to request cached publications, as well as future ones.

### 3.1 Service Model

A receiver *r* advertises its information interest to a router *R*, as a subscription *S* defined by a *predicate*, a specification of the maximum amount of publications *max* admissible over a period of time $lifetime$ and $freshness$ the maximum age for a matching publication. The *freshness* of a publication is the elapsed time since its initial upload in the system. The *max* parameter is called the selectivity of the subscription. The content-based network delivers to receiver *r* (via home router *R*), at most *max* publications before $lifetime$ expires. We assume that a publication is uploaded once by an authoritative publisher and that receivers are allowed to refresh subscriptions when they expire. As such, an important requirement for the usability of the service is defined as follows: The service should guarantee that a refreshed subscription 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 subscriptions satisfied with a publication or with all publications already consumed by a subscription. For the rest of this paper, we refer to this requirement as (`RQ`). To allow routers to differentiate refreshed subscriptions from new ones, we assume the existence of an agreement between routers and receivers for this purpose.

**Proposition 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.

- S *is satisfied when the following relation is verified:*

$$N_S = min(M_S, max) \quad (1)$$

- *Starvation happens when:*

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

The $lifetime$ parameter can be interpreted as the delay allocated by receivers to the content-based network to satisfy an interest. Starvation occurs due to congestion or due to the service failing to timely satisfy subscriptions. The frequency of occurrence of *starvation* is the metric to characterize the quality of service offered by an implementation of the service model compared to an implementation of the exhaustive filtering semantic.

### 3.2 Problem Statement and Key Design Issues

Our objective consists in exploring the trade-off between communication costs incurred and the quality of service offered to consumers. The key design issues impacting this trade-off are discussed below and addressed in the next section (Sect. 4).

**Caching policies** implemented by routers play a key role in minimizing communication costs by maximizing sharing between similar subscriptions.

**Dissemination strategies** defining the conditions making a publication eligible for dissemination towards interested receivers is a key element to pace the dissemination process to information needs.

**Content-based forwarding schemes** extending baseline forwarding [15] with the ability to take into account bandwidth and resource constraints by assigning priorities to transiting subscriptions. Such forwarding algorithms should minimize starvation under congestion.

**Content-based routing protocols** that should take advantage of the *max* parameter and the availability of matching publications in the neighbourhood of requesting routers to minimize routing overhead.

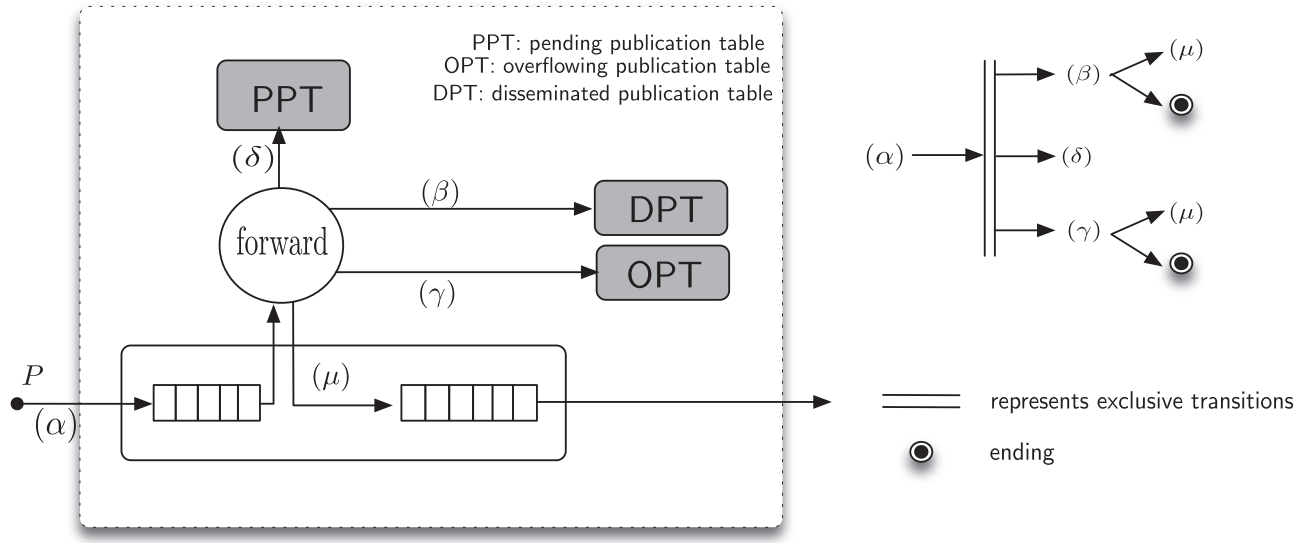
## 4. Framework and Algorithms

In order to achieve our objectives, we have designed a framework supporting the service model described above.

### 4.1 Data Structures and Caching Policies

For each registered subscription $S(max, lifetime, ...)$, the corresponding home router *R* instantiates a *box* $B_S$ such that $B_S$ monitors the number of matching publications cached in order to satisfy $S$. $S$ belongs to the set of subscriptions local to *R*. $B_S$ is *full*, when it references *max* publications. Whenever $B_S$ is *full*, *R* considers that $S$ is satisfied and removes the corresponding states from the forwarding table. We describe below the different types of publications stored by routers and the indexes required to manage them.

**Transiting publications,** which are buffered in the *buffer*




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

**Fig. 1** Router model and publication message flow inside a router.

before and prospectively after the forwarding decision.

**Pending publications,** which have been uploaded at some router from publishers and that are waiting for opportunities to be further disseminated are referenced in the *pending publication table* (PPT). A pending publication may have been used to satisfy subscriptions which are local to a router. In order to avoid that refreshed subscriptions consume the same publication, we define a *dispatched flag* (DF) in the PPT indicating whether a pending publication has been used to satisfy local subscriptions. Entries from the PPT are removed once the corresponding publications have been selected by a remote subscription.

**Disseminated publications,** which have been disseminated towards remote routers and used to satisfy local subscriptions are referenced in the *disseminated publication table* (DPT). When a publication is disseminated for the first time, it is forwarded with a *pending flag* (PF) set to true. The utility of this flag is explained in the next paragraph.

**Overflowing publications,** which have been disseminated to remote subscribers, but that never served locally are referenced in the *overflowing publication table* (OPT) as they may be useful for refreshed subscriptions in future lifetimes. New entries are added to the OPT for pending publications selected by remote subscriptions with dispatched flag DF equals to false and for publications incoming with pending flag PF equals to true without matching local subscriptions. The latter situation occurs whenever the content-based network returns more publications than requested or the states corresponding to a satisfied subscription are still in the forwarding table or the *en-route caching* optimization is enabled (See Sect. 4.1.2).

The indexes introduced above, namely DPT, PPT and OPT do not overlap and have been designed to effectively support (RQ), while maximizing the opportunities to satisfy refreshed subscriptions. Figure 1 depicts the processing of an incoming publication inside a router.

#### 4.1.1 Selection and Replacement Policies

Each router executes a *selection policy* to determine which publication to select first, 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.

Selection policies guarantee that new subscriptions are satisfied with any available publication, while refreshed subscriptions are served only with publications that have not been delivered to them during previous lifetimes. In order to be consistent with (RQ), refreshed subscriptions should not be satisfied with publications indexed in DPTs or in OPTs of remote routers.

Consequently, new subscriptions have more opportunities to be satisfied than refreshed ones. A reasonable high-level selection policy to regulate opportunities between new and refreshed subscriptions is to satisfy new subscriptions in the following order with dispatched, overflowing and pending publications available at the originating router, and pending publications available at remote routers if necessary, and to satisfy refreshed subscriptions first with overflowing and pending publications available at the originating router with DF sets to false, and finally, pending publications available at remote routers of the network.

Different high-level replacement policies can be envisaged. For instance, replacing pending publications first, then overflowing publications and finally disseminated ones. The reverse order is also applicable, as well as not differentiating publications on their types but rather on usages.

Selection and replacement policies can also discriminate publications which belong to the same type of publications according to different policies such as *most recently used* (MRU), *least recently used* (LRU), *most frequently used* (MFU), *most fresh* (MF) or *less fresh* (LF). Finding the right policies strongly depends on workload characteristics and requires an in-depth evaluation. We provide some insight on the impact of the ratio of new to refreshed subscriptions on performances in a companion paper [14] and will study this issue in-depth in future work.

#### 4.1.2 Caching Optimizations

In order to improve the availability of publications, we consider the following optimizations:

**En-route caching:** With this policy enabled, routers are allowed to cache transiting publications according to the enforced replacement policy.

**Flushing:** Whenever the cache of a router $R$ overflows, *flushing* allows routers to copy replaced publications towards remote routers that may be interested in caching them. The caching decision depends on the replacement policy. This can be achieved for instance by flooding the replaced publication with the appropriate signalling in all the network or within $R$'s neighbourhood.

### 4.2 Content-Based Routing

We describe a content-based routing scheme that has the ability to scope the propagation of subscriptions by enabling newly registered subscriptions to benefit from publications available into routers cache. Routers exchange subscriptions which are processed as described below. Upon reception of $S(max, *)$ by router $R$, if the number of 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 propagated with the $max$ parameter decremented by the number of matching publications offered by $R$. Note that in order to avoid loops and duplicates with arbitrary topologies, we assume that subscription messages embed a unique identifier, as well as a nonce. This is more convenient than operating over a global spanning tree.

### 4.3 Dissemination Strategies

We did not thoroughly explore dissemination strategies' design space. Those discussed below, are simple but yet effective in the trade-off between satisfaction and communication-efficiency.

#### 4.3.1 Push/Pull and Explicit Overload Notification (EON)

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

#### 4.3.2 Pull/Delayed Push (PDP)

Unlike EON, PDP does not use explicit notification messages to notify remote routers of *overload*. PDP uses the propagation of new and refreshed subscriptions 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 router $R$ at time $t_0$ and $S_P$ the set of matching subscriptions advertised in $R$'s forwarding table (excluding local subscriptions), the most lately publication time $t_P$ associated to $P$ is given by the relation:

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

$\beta$ being a system parameter and $deadline(S)$ being the remaining time before S expires. At last, any publication that may contribute to satisfy a subscription not yet satisfied, is finally disseminated. Note that $\beta$ should be greater or equal to the time required to forward a publication between two endpoints of the network.

### 4.4 Content-Based Forwarding

The algorithm described here extends the basic CBF algorithm described in 3. First, given the fact that different routers may respond with the same publication to a given request, it is fundamental that a router drops a publication it receives when that publication is already present in its cache. For this purpose, we assume that publication messages embed a unique identifier that can be used for such purpose. Second, congestion may appear at some routers. 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 that can be embedded in a header. The score is recomputed by every hop on the delivery path. Let $R$ be a router and $\Pi$ 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 subscriptions matching $P$ downstream $R$, and compute a score $\mathtt{Cbf}(P)$ giving higher priority to *popular* and *urgent* publications. We estimate popularity by the number of matching subscriptions advertised downstream $R$ and urgency by the minimum deadline among matching subscriptions. $\mathtt{Cbf}(P)$ is defined by the following equation:

$$Cbf(P) = \frac{|S_P|}{\min\limits_{S \in S_P}(\text{deadline}(S))}. \tag{4}$$

## 5. Evaluation

We are still in the initial evaluation phase of the framework. In this paper, we present the first coarse-grained results. We use our PEERSIM [13] implementation of the framework to measure its performances comparatively to an ideal content-based routing scheme [17] implementing an *exhaustive filtering* semantic (CBR) under realistic workload assumptions. We did not yet evaluate the content-based forwarding algorithm, nor the whole space for all the caching policies. The results were obtained with the following metrics:

**Saved bandwidth,** which is the fraction of bandwidth saved by using the framework instead of CBR in terms of publication traffic.

**Control traffic overhead,** the ratio between the subscription traffic generated by the framework and the subscription traffic generated by CBR.

**Starvation probability,** the percentage of unsatisfied subscriptions as defined in Sect. 3: A subscription is not satisfied when the number of matching publications available by the time the subscription expires at the home router is less than the subscription's selectivity (*max*), while there were sufficient opportunities to retrieve enough publications.

### 5.1 Methodology

It is widely acknowledged that content-based pub/sub research lacks public data sets for meaningful evaluation. Thus, synthetic workload generation is largely accepted in the field, assuming that the generated workload meets some realistic assumptions. We consider that a set of events generate both publications and subscriptions, and that three parameters characterize each event: *popularity*, *locality* and *volume*. The popularity of an event refers to the number of subscriptions related to it, its volume to the number of related publications and its locality to the regions of the topology likely to originate related subscriptions and publications.

**Table 1** Parameters used for performance evaluation.

| Parameters | Definition | Value(s) |
|---|---|---|
| $T$ | Simulation length | 1000 |
| $N$ | Network Size | 100 |
| $E$ | Number of events | 500 |
| $S$ | Total number of basic subscriptions | $10k$ |
| $Max$ | Maximum selectivity (default) | 10 |
| $P$ | Total number of publications | $500k$ |
| $Freshness$ | Maximum freshness (age) | 1000 |
| $Lifetime$ | Maximum lifetime | 50 |
| $P_r$ | Refresh prob. | 0.5 |
| $B_s$ | Buffer size | 2500 |
| $\beta$ | Parameter of Eq. 3 | 1 |

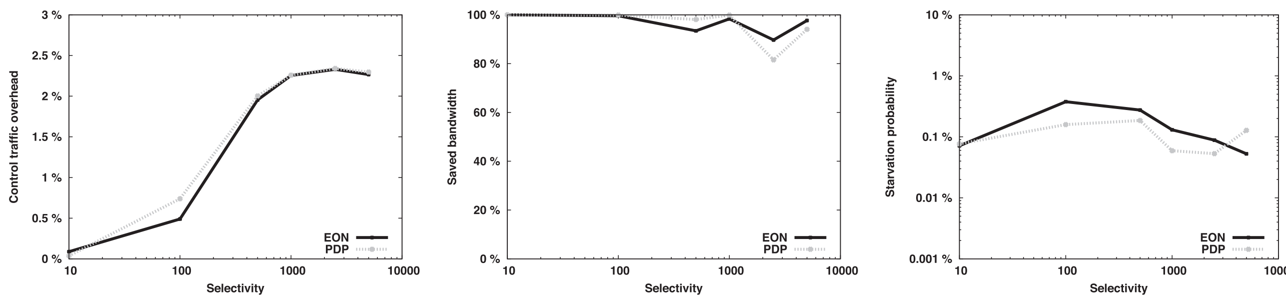
In the generation of our workload, we assume an arbitrary router topology of average degree 2 following a power-law distribution with few nodes of high degree and many nodes of small degree. Popular events are characterized by larger and broader audiences in terms of number of subscriptions, and are also likely to trigger larger volume of publications.

Please, refer to Table 1 for the definitions of the variables introduced in this paragraph. Let us consider $e_i$ ($1 \leq i \leq E$), an event of popularity $p_i$, volume $v_i$ and locality $l_i$. $p_i$ is sampled from a power law distribution of exponent 0.7 and the volume $v_i$ is such that $v_i = P * p_i$. Publications associated to $e_i$ can be issued by any of the $N$ routers, while subscriptions only from a set of nodes computed using $l_i$. We define $l_i$ such that $l_i = p_i$ and such that $\lceil l_i * N \rceil$ routers are potential issuers (hosting interested receivers) of subscriptions related to $e_i$. This set of routers is computed by choosing a random root node and $\lceil l_i * N \rceil - 1$ additional nodes among the closest routers to the root. We assume a constant arrival rate for subscriptions (resp. publications) equals to $r_s = S/T$ (resp. $r_p = P/T$). For each publication, we randomly select an event and a location among all routers. When the volume associated to an event is reached, it is removed from the set of events that can be used to generate new publications. Locations for subscriptions are selected among the set of routers related to the event. Subscriptions are generated using the $Max$, $Freshness$ and $Lifetime$ parameters defined by Table 1. We assume that each subscription is renewed at the end of its lifetime with a probability $P_r$. For the simulation of PDP, we set $\beta$ to 1, as our simulation model neglects simulation transfer delays.

Finally, we assume that selection and replacement policies do not differentiate publications according to their type (pending, overflowing or disseminated), but that both apply a *less fresh* (LF) policy.

### 5.2 Results

The results described in this section are obtained with the settings of Table 1. We measured the performance of the framework using EON and PDP relatively to CBR. We varied




(a) Control traffic overhead (b) Bandwith saved (c) Starvation probability

**Fig. 2** Framework performances.

the maximum selectivity *Max* between 10 and 5000 in order to obtain different load levels. Note that scenarios with larger selectivity values are the most favourable to CBR.

Figure 2(a) and Fig. 2(b) demonstrate the communication-efficiency of our framework, and Fig. 2(c) shows that these gains are not obtained at the expense of the quality of service offered to subscribers. In fact, EON (resp. PDP) can save up to 85% (resp. 80%) of bandwidth compared to CBR and for a wide range of selectivity values, PDP and EON provide very close quality of services and satisfy more than 99% of subscriptions. Also, EON and PDP produce less than 3% of the control traffic generated by CBR that shows the effectiveness of our content-based routing algorithm. Note that PDP generates less control traffic than EON for selectivity values less or equal to 10. In fact, for these values most subscriptions are satisfied before their lifetime expires and EON will generate a lot of *overload notification* messages. But, as selectivity increases, the number of occurrences of overload decreases and thus the number of control messages generated by EON.

## 6. Discussion

### 6.1 Related Work

To the best of our knowledge, we are the first to propose a service model that captures consumers'needs in order to improve the communication-efficiency of content-based networking (CBN) or decentralized content-based publish/subscribe.

Online services such as `Google alerts` [1] implement a similar service model to deal with information overload. `Google alerts` allows users to request content of some *type* by a set of *keywords* and to specify the notification *frequency* and *volume*. However, `Google alerts` is a centralized system.

Previous research has also considered the issue of Information overload. Information filtering literature provides directions to filter a stream of information and deliver the most relevant information items to users given their predefined profiles. Information filtering leverages several information retrieval ranking techniques [21]. The baseline system for Information filtering is SIFT [18]. It focuses on efficient indexing and matching. SIFT is basically a centralized system. But, two distributed versions of SIFT are sketched in [19]. Our work differs from SIFT in two points. The first difference is that SIFT searches the terms of a profile into a document. In our work, we assume that every information item is described by some metadata. This reduces considerably the complexity of the matching process without losing accuracy. The second difference is that in order to address Information overload, SIFT requires that for each query Q, users specify a relevance threshold R defined such that a document is relevant if the score returned by the matching process is above R. End-users would probably be reluctant to tune themselves this parameter.

Some of the mechanisms implemented by our framework resembles to mechanisms that have been applied to information dissemination problems in other environments. For instance, directed diffusion [22] defines many mechanisms for data dissemination in a sensor network. The authors propose one-phase pull diffusion, a protocol that provides mechanisms for routing queries and sensor data. Directed diffusion shares with us the fact that data dissemination is triggered by the propagation of matching interests. However, the protocol described in [22] is specific to sensor applications generating data streams, thus not generalizable to our problem.

Leveraging content-based publish/subscribe for content distribution has been considered by Chen et al. [20]. They studied 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 thus out of the scope of this paper. Other work that have considered caching to increase the availability of publications in decentralized CBPS systems implement an exhaustive filtering semantic [24].

Pacing the dissemination process to information needs has also been previously addressed. Corona [23] is a publish/subscribe 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 publish/subscribe systems.

### 6.2 Open Challenges

Several issues remain to be addressed in order that content-based networking operates at global scale.

#### 6.2.1 Heterogeneous Applications Support

Heterogeneous applications should be supported for global-scale CBN mediation between information providers and consumers. Given that different applications require different naming schemas and semantics, one important issue is the ability to multiplex heterogeneous applications over the same content-based network. This raises several exciting issues such as designing generic yet efficient data structures, as well as engineering content-based routers to provide different quality of services to applications with heterogeneous delay requirements.

#### 6.2.2 High-Throughput Forwarding

Although our framework tremendously reduces the amount of control traffic forwarded into the content-based network, and thus the amount of states covered by forwarding tables, it is unclear whether it will be sufficient to reduce tables' size and thus achieve high-throughput forwarding. In fact, several established optimizations such as covering-based and merging-based routing [16] are not applicable here because it is difficult to establish covering relationships between subscriptions having different temporal ($lifetime$) and quantitative ($max$) requirements. Thus, designing high-throughput content-based forwarding for our service model remains an open and interesting issue.

## 7. Conclusion and Ongoing Work

We introduced a novel solution to increase the efficiency of content-based networking. We propose a framework that takes advantage of consumers' attention span and in-networking caching to pace the dissemination process and reduce the communication costs of content-based networking. Simulation results have shown that our framework could save up to 85% of bandwidth compared to an optimal content-based routing scheme supporting the usual exhaustive service semantic. The gain is obtained while still generating very low control traffic and maintaining a high service level. This promising approach deserves further analysis in order to isolate the gains obtained thanks to caching policies from those of the dissemination methods and the forwarding algorithm.

### References

[1] Google Alerts service — http://www.google.fr/alerts
[2] Content-based networking project, http://www.inf.usi.ch/carzaniga/cbn
[3] D. Clark, "The design philosophy of the DARPA internet protocols," Proc. Communications architectures and protocols, (SIGCOMM'88), 1988.
[4] T. Leighton, "Improving performance on the Internet," Commun. ACM, vol.52, no.2, pp.44–51, Feb. 2009.
[5] N. Niebert, S. Baucke, I. El-Khayat, M. Johnsson, B. Ohlman, H. Abramowicz, K. Wuenstel, H. Woesner, J. Quittek, and L.M. Correia, "The way 4WARD to the creation of a future Internet," IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, 2008.
[6] S. Tarkoma, M. Ain, and K. Visala "The publish/subscribe Internet routing paradigm (PSIRP): Designing the future Internet architecture," Towards the Future Internet, 2009.
[7] V. Jacobson, D.K. Smetters, J.D. Thornton, M.F. Plass, N.H. Briggs, and R.L. Braynard, "Networking named content," Proc. 5th international conference on Emerging networking experiments and technologies (CoNEXT'09), pp.1–12, 2009.
[8] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: Enabling innovation in campus networks," ACM SIGCOMM Computer Communication Review, pp.69–74, 2008.
[9] A. Bavier, N. Feamster, M. Huang, L. Peterson, and J. Rexford, "In VINI veritas: Realistic and controlled network experimentation," Proc. 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, pp.3–14, 2006.
[10] S. Fdida, T. Friedman, and S. MacKeith, "OneLab: Developing future Internet testbeds," Lect. Notes Comput. Sci., vol.6481, pp.199–200, 2010.
[11] A. Carzaniga, M.J. Rutherford, and A.L. Wolf, "A routing scheme for content-based networking," Proc. 23rd Conf. IEEE Computer and Communications Societies, pp.918–928, 2004
[12] P. Jokela, A. Zahemszky, C. Esteve Rothenberg, S. Arianfar, and P. Nikander, "LIPSIN: Line speed publish/subscribe internetworking," Proc. ACM SIGCOMM conference on Data communication, pp.195–206, 2009.
[13] A. Montresor and M. Jelasity, "Peersim: A scalable p2p simulator," IEEE 9th International Conference on Peer-to-Peer Computing (P2P'09), pp.99–100, 2009.
[14] M. Diallo, S. Fdida, V. Sourlas, P. Flegkas, and L. Tassiulas, "Leveraging caching for Internet-scale content-based publish/subscribe networks," Proc. Int. Conf. Commun. 2011 (IEEE ICC2011), pp.1–5, 2011.
[15] A. Carzaniga and A.L. Wolf, "Forwarding in a content-based network," Proc. 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, pp.163–174, 2003.
[16] S. Tarkoma and J. Kangasharju, "Optimizing content-based routers: Posets and forests," Distributed Computing, pp.62–77, 2006.
[17] A. Carzaniga, A.J. Rembert, and A.L. Wolf, "Understanding content-based routing schemes," University of Lugano Faculty of Informatics Technical Report 2006/05, 2006.
[18] T.W. Yan and H. Garcia-Molina, "SIFT: A tool for wide-area information dissemination," Proc. USENIX 1995 Technical Conference Proceedings (TCON'95), pp.175–186, 1995.
[19] T.W. Yan and H. Garcia-Molina, "The SIFT information dissemination system," ACM Trans. Database Syst. (TODS), pp.529–565, 1999.
[20] M. Chen, A. LaPaugh, and J.P. Singh, "Content distribution for publish/subscribe services," Proc. ACM/IFIP/USENIX 2003 International Conference on Middleware (Middleware'03), pp.83–102, 2003.
[21] N.J. Belkin and W.B. Croft, "Information filtering and information retrieval: Two sides of the same coin?," Commun. ACM, pp.29–38, 1992.
[22] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: A scalable and robust communication paradigm for sensor networks," Proc. 6th Annual International Conference on Mobile Computing and Networking (MobiCom'00), pp.56–67, 2000.
[23] V. Ramasubramanian, R. Peterson, and E.G. Sirer, "Corona: A high performance publish-subscribe system for the world wide web,"

Proc. 3rd conference on Networked Systems Design & Implementation, (NSDI'06), vol.3, 2006.

[24] V. Sourlas, G.S. Paschos, P. Flegkas, and L. Tassiulas, "Caching in content-based publish/subscribe systems," Proc. Global Telecommunications Conference (GLOBECOM 2009), pp.1–6, 2009.

[25] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, "A survey of information-centric networking," Proc. Dagstuhl Seminar, 2011.

[26] K. Palmskog, A.G. Prieto, C. Meirosu, R. Stadler, and M. Dam, "Scalable metadata-directed search in a network of information," Proc. Future Network and Mobile Summit, pp.1–8, 2010.

[27] A. Carzaniga, D.S. Rosenblum, and A.L. Wolf, "Achieving scalability and expressiveness in an internet-scale event notification service," Proc. 19th Annual ACM Symposium on Principles of Distributed Computing, 2000.

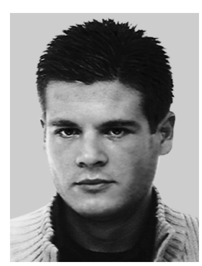
[28] B. Quilitz and U. Leser, "Querying distributed RDF data sources with SPARQL," The Semantic Web: Research and Applications, pp.524–538, 2008.

[29] N. Fotiou, D. Trossen, and G.C. Polyzos, "Illustrating a publish-subscribe Internet architecture," Telecommunication Systems, pp.1–13, 2009.

[30] T. Koponen, M. Chawla, B.G. Chun, A. Ermolinskiy, K.H. Kim, S. Shenker, and I. Stoica, "A data-oriented (and beyond) network architecture," ACM SIGCOMM Computer Communication Review, vol.37, pp.181–192, 2007.

**Mohamed Diallo** is preparing a Ph.D. degree at UPMC Sorbonne universités since 2007. He received an engineering degree in telecommunication from INPT (Rabat, Morocco) in 2006. His research is related to network architecture, information-centric networking, publish/subscribe systems design and evaluation, and content distribution at large. http://www-rp.lip6.fr/~diallo/

**Serge Fdida** received his Ph.D. in 1984 and Habilitation a Diriger des Recherches in 1989 both from the University Pierre et Marie Curie (Paris 6). His research interests are related to innovative wireless networks, large-scale content distribution systems as well as the design of federated test-beds to support experimentally driven research. From 1989 to 1995, he was a Full Professor at the University Rene Descartes (Paris). He was a Visiting Scientist at IBM Research (North Carolina) during the 1990/91 academic year and then joined the University Pierre et Marie Curie in 1991. He is/was leading or involved in many research projects in High Performance Networking in France and Europe, such as WIP (IST FP6 - An All Wireless Mobile Network Architecture), ONELAB and ONELAB2 (IST FP6&7 - An Open Laboratory Supporting Network Research Across Heterogeneous Environments). He was chairing the french national research group on high-speed networking in France for 8 years and the COST264 Action on Internet Architecture in Europe (98-02), and was on the steering committee of the European FP6 network of Excellence ENEXT. Serge Fdida is a senior member of IEEE and a Distinguisehd ACM Member. Serge Fdida was the Director of EURONETLAB, a joint laboratory established in 2001, between University Paris 6, CNRS, THALES, ENST and BLUWAN, developing collaborative research on the wireless internet. He is also a co-founder of the Qosmos company. From 2000 to December 2005, he was working part-time as a scientific adviser with CNRS-STIC (French National Scientific Research Center/Information Science & Technology). He was also the Vice-President of the RNRT (French national Research Network in Telecommunications), involving the industrial, academic and government actors in France. http://www-rp.lip6.fr/~sf/



**Prométhée Spathis** is an associate professor at the Université Pierre et Marie Curie (Paris 6) in the Network and Performance Analysis group. His current main research interests are content- and user-centric networking applied to various environments such as vehicular networks or urban environments. He received his Ph.D. in Computer Science from the Université Pierre et Marie Curie (Paris 6) in 2003. He has been involved in EU-funded and national research projects. He also participates to scientific-technological cooperation programs with South America. He is currently co-advising two Ph.D. students in the research areas of content-based architectures and the analysis of the community-created data of user-generated content systems.