Towards A Scalable Content-Based Pub/Sub Network Service

[Extended abstract for DEBS PhD Workshop] *

Mohamed Diallo
Laboratoire d’Informatique de Paris 6
UPMC Paris Universitas
mohamed.diallo@lip6.fr

ABSTRACT
Contributing to the current discussion of the design of a Network of Information supporting a Data-Centric approach, we explore the ability to extend the "Content-Based Publish and Subscribe" service model to the Internet scale (Widthness and Information volume). While the baseline Content-based routing algorithm appears as an ideal fit for this service, it incurs a huge complexity, only justified for applications requiring an exhaustive (v.s. a partial) matching semantics, where a subscription selects every matching publication. However, in the case of content distribution applications potentially deployable at this scale, subscribers are likely to have limited expectations in face of an increasing volume of information. Thus, implementing an exhaustive matching semantics would incur a huge overhead. In this paper, we attempt to answer to the question of how to take advantage of the knowledge of subscribers’ expectations to mitigate this overhead. For this purpose, we propose Avalanche, an original dissemination scheme, combining push-pull mechanisms and partial data matching. Avalanche reduces this overhead at the price of a lower complexity and a negligible service degradation.

Categories and Subject Descriptors
C.2.4 [Computer Communication Networks]: Distributed systems

General Terms
performance design

Keywords
Data-centric Internet, publish-and-subscribe

1. INTRODUCTION

Contributing to the current discussion of the design of a Network of Information supporting a Data-Centric approach, we explore the ability to extend the "Publish and Subscribe" (Pub/Sub) service model to the Internet scale. The future Internet will be polymorphic and will enable the deployment of concurrent network services on top of the same physical infrastructure. Realizing the mediation between information providers and information consumers "en route" at the network level will enhance the value of the network with respect to many information dissemination applications. In this mediation system, clients submit subscriptions to an information dissemination service that is directly managed as an underlay. The search activity will provide a means for competition and will support an architecture for participation where various types of content can be disseminated at different scales. Enabling routers to access and process high-level information such as data generated by applications opens a new avenue with promising functionalities and a set of difficult challenges.

The main difficulties are to cope with Internet-scale workloads and the heterogeneity of application requirements. The communication efficiency and complexity of "content based publish and subscribe" (CBPS) systems are tightly bound to the semantics of the target applications. Attempts to build a one-design-fits-all scalable CBPS system are doomed to failure. While baseline Content-based routing algorithms [5][4][6] appears as ideal fits for a wide range of applications within our vision, it incurs a huge complexity, only justified for applications requiring an exhaustive matching semantics, where a subscription selects every matching publication and every matching publication should be delivered timely to every matching subscription. However, in the case of content distribution applications potentially deployable at the Internet scale, subscribers are likely to have limited expectations in face of an increasing volume of information. Thus, implementing an exhaustive matching semantics would incur a huge overhead.

In this paper, we describe our research consisting in developing mechanisms that enforce a partial matching semantics at a lowest cost and complexity than any alternative implementing an exhaustive matching semantics. For this purpose, we state the problem and main design issues in the context of a network-supported service. Then, we explore and discuss promising approaches within the solutions space.

2. PROBLEM STATEMENT
Our content-based publish-and-subscribe service enables content providers to deliver content, over a network, to interested entities. The service operates at Internet scale and under heavy workload. Clients express their interests (subscriptions) as persistent keyword-based queries. Each subscriber is assigned a home rendezvous point, which is its access point to the service and its delegate within the system. Subscribers can be assigned to rendezvous points according to various criteria such as proximity and trust. Publications refer to content associated with descriptions including a set of keywords. A subscription matches a publication when all its keywords are included in the publication description. The service is responsible for delivering published content towards subscribers’ home rendezvous points. Subscribers’ notification of relevant content availability at their home rendezvous points and actual delivery of content from home rendezvous points to subscribers is beyond the scope of this work, and can be performed by any possible means.

We consider that devices named Mediation Routers (MR) support the CBPS service. They operate as an overlay, called the Mediation Network (MN), responsible for all the operations providing this service. Each rendezvous point knows about his local subscribers expectations: interests, min, max, period, where min and max are bounds specifying the amount of information expected by subscribers over a period (hourly, daily, weekly...).

We introduce an abstraction that models the number $N_i$ of content items delivered in regard to a subscription $S_Q$, where $t$ is the elapsed time since the beginning of the period. We instantiate this abstraction for each subscription and call it a Box. Boxes are mainly useful for monitoring the contract between subscribers and the system. The contract of a subscription $S_Q$ is fulfilled when at the end of each period the following relation is verified: $\min \leq N_{\text{period}}, \text{while no more than min items have been published in the system over the considered period}$, otherwise we say that starvation occurred. Note that the box is full, when $N_i \geq \max$. At the end of the period, the retrieved content items are consumed by the subscribers. It is worth-mentioning that services such as Google Alerts[2] implement a similar interface.

Given such service semantics, in the presence of Internet-scale workloads likely to meet the information overflow property, informally more information available than needed to operate the service, it is worth investigating how to reduce the overhead that would incur any content-based routing algorithm implementing an exhaustive matching semantics. We acknowledge the fact that workloads are difficult to predict for publish-and-subscribe applications and share the idea that reconfiguration of publish-and-subscribe systems matter[7][6][8]. But knowing how the CBPS service can be implemented effectively for a plausible workload pattern matters as much. Dealing with the information overflow issue requires that content-based publish-and-subscribe systems take advantage of the knowledge of subscribers’ expectations in order to meet communication-efficiency and scalability.

The main design issues are:

**Dissemination trigger.**

Baseline approaches disseminate publications as soon as available. This does not give any control over the information overflow issue. How to exploit the knowledge of subscriber’s expectations to pace the dissemination process to information needs is an interesting and non-trivial question. **Filtering strategy.**

The main purpose of any filtering strategy is to meet communication-efficiency. Filtering matters especially when there is either a small density of recipients within the Mediation Network or interests are localized within the topology. These are common workload assumptions behind most content-based publish-and-subscribe systems research [5]. A Filtering strategy is mainly characterized by the matching semantics, which defines the set of recipients of a publication, and the matching execution, which specifies how nodes collaborate to compute this set of recipients. Designing a filtering strategy for a partial matching semantics, to the limit of our knowledge, is an original problem. **Filtering state management.**

Filters are required to implement filtering strategies. This is about what information is embedded in filters, and how filters are represented and managed. The trade-off is that maintaining a tight consistency of filters with subscriptions might result in a costly control plane (important control traffic, huge state complexity, dynamic and real-time index management, etc.), but would give a more precise control over the effectiveness of the service, while tolerating inconsistent or approximate filters would result in a more scalable control plane while losing accuracy. Premature optimization being the root of all evil, we firstly consider options enforcing some consistency between the states in the mediation network and the existing subscriptions. **Content forwarding scheme.**

A content forwarding scheme defines how contents are disseminated towards their recipients. The main options worth-considering are Content-based forwarding (CBF) and Label-based reverse-path forwarding (LRPF). In CBF, dissemination is achieved by matching publications against routers’ filters on a hop-by-hop basis [5]. CBF is a very effective and quite well understood paradigm, in order to realize an exhaustive matching semantics, but how to extend it to implement a partial matching strategy that would be effective in tackling the Information Overflow issue is an open research question. With LRPF, content to disseminate is addressed by a set of labels computed by the source of the publication and identifying the paths towards interested parties. **Control plane.**

In our research, we focus essentially on subscription state forwarding, which enables the implementation of filtering strategies and dissemination triggers. Thus, its design is mainly driven by these building blocks. The fundamental design question is about performing scalable, robust, and manageable broadcast and enforcing reliable soft-state mechanisms. An important additional design consideration is that the control plane should be simple, distributed and decentralized. Thus, we do not consider offline approaches, which are not scalable from an internet-scale deployment perspective. Also, we target solutions that would perform well on generic P2P topologies.

### 3. AVALANCHE SCHEME

Having defined the problem, our work focus on exploring...
and characterizing the most promising options within the solutions space. In this section, we present Avalanche an original dissemination scheme, combining a source-based matching strategy with a pull-based trigger. Avalanche takes advantage of subscriptions state propagation, to trigger the dissemination of information and deliver it opportunistically to interested parties. This approach has shown to be quite effective in mitigating the information overflow issue, without much degradation of the service quality.

3.1 Design

Subscription states forwarding.
When a subscriber is looking for a given content, it will subscribe to it following certain rules not described here. As a consequence, a query $Q_n$ (keywords, max, lifetime) will then be issued. This query is broadcasted within the network until pending publications are hit at some location (up to max items are retrieved). During this phase, label-based reverse path states are installed in the network for a period of time. This has the nice property to localize subscription states when there is some popularity locality in the network, unlike the baseline subscription-forwarding algorithm. This contributes in reducing state and matching complexities.

Dissemination trigger.
When a publication is issued at a mediation router, it is dispatched to local matching subscriptions and not further disseminated. It remains pending until it is selected by a propagating query. Some content might be never disseminated, this is a marginal issue and each mediation router can implement its own caching policy. We compared Avalanche performance to an in-network filtering strategy implementing an exhaustive matching semantics combined with an immediate push-based trigger. In our scenario, Avalanche clearly mitigates the overhead by more than a factor of 4 compared to this baseline strategy (Figure 3).

Content dissemination.
The set of matching subscriptions is computed at the source of a publication. When a content $D_c$ has been located by a request $Q_n$, not only will it follow the reverse path $R_n$ of that request, but it will also irrigate (be disseminated through) every forwarding path that has been set up by a similar query and that intersects $R_n$ at the source of the publication. This dissemination method is called Avalanche, as from the injection of $D_c$ in $R_n$ in answer to $Q_n$, other forwarding paths are also irrigated with $D_c$.

3.2 Analysis and Evaluation

We implemented the Avalanche scheme in PEERSIM[1], a discrete event simulator written in Java and a common choice for simulating large-scale peer-to-peer scenarios. We investigated several scenarios with different combinations of parameters. The results presented in this section are obtained from the workload characterized in Figure 2 and Figure 1. Figure 1 illustrates the overflow phenomenon, captured by the gap distribution, representing the difference between subscribers’ expectations and the amount of available information in their box at lookup time. This distribution is clearly positive. Figure 2 shows that the publications have an heterogeneous popularity distribution. The scenario in-
cludes a random topology of 50 nodes, 1000 subscriptions, and a large volume of publications captured in Figure 2. The simulation models the information consumption process by the subscribers, which consists in performing lookups in their box at the end of the service period.

Figure 3 shows clearly the gain of Avalanche compared to an optimal exhaustive matching strategy. We also observed that the starvation probability was about 5% which is tolerable, and the cost of the subscription forwarding scheme reduced drastically. The side-effects of limiting the broadcast of subscription states combined with source-based matching of Avalanche are to reduce the matching and state complexities. In this era of energy-thriftiness, this not a trivial result.

4. CONCLUSION

In this paper, we gave an overview of our research investigating the issue of designing a scalable Content-based Publish-and-Subscribe network service for Content Distribution. For this purpose, we introduced a loose semantics that better suits many content distribution applications, thus creating opportunities to increase the scale of the content-based publish-and-subscribe framework. The design space being large, and the problem having multiple dimensions we defined the problem and the key design issues. Finally, we described Avalanche, an original solution that implements effectively a partial matching semantics, at a reasonable complexity and without much service degradation. Besides investigating other alternatives within the solution space, there are many other tough and interesting issues worth considering, to which we hope to bring more focus in the future, such as extending our model to network constraints, reliable soft-state management and relevance-related issues.

5. REFERENCES
