HAL
open science

Poster : Modèle générique pour le routage orienté contenu

Yosra Barouni, Prométhée Spathis, Serge Fdida

To cite this version:

Yosra Barouni, Prométhée Spathis, Serge Fdida. Poster : Modèle générique pour le routage orienté contenu. Colloque Francophone sur l'Ingénierie des Protocoles (CFIP 2008), Mar 2008, Les Arcs, France. hal-00260342

HAL Id: hal-00260342

https://hal.science/hal-00260342v1

Submitted on 4 Mar 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modèle générique pour le routage orienté contenu

Yosra Barouni* — **Prométhée Spathis*** — **Serge Fdida***

**Laboratoire d'informatique de Paris 6*
CNRS/Université Pierre et Marie Curie
104 avenue du président kennedy, 75016 PARIS
{yosra.barouni, promethee.spathis, serge.fdida}@lip6.fr

RÉSUMÉ. *Nous proposons un modèle générique pour capturer les propriétés distinctives des diverses approches existantes en matière de distribution de contenu. A partir de ce modèle, nous introduisons une terminologie permettant d'analyser les fonctionnalités propres aux approches existantes. Ensuite, nous proposons une architecture de routage orientée contenu effectuant un filtrage atomique des requêtes basé sur un vecteur optimal calculé pour un ensemble de métriques de filtrage.*

ABSTRACT. *In this paper, we propose a generic model that captures the inherent properties of the existing approaches that resolve the content routing and distribution field. From this model, we propose a terminology that allows us to analyse the functionalities implemented in these solutions. Then, we propose a content based routing architecture that filters client request using an optimized filtering vector.*

MOTS-CLÉS : *Réseaux de contenu, routage, filtrage, optimisation.*

KEYWORDS: *Content based networks, routing, filtering, optimization*

1. Introduction

Le succès que connaît l'Internet est largement dû la quantité de contenu mis è la disposition de tous. Pourtant, l'infrastructure aussi bien matérielles que logicielles de l'Internet ainsi que son modèle orienté hôte n'ont pas été architecturés pour permettre la livraison de contenu à son échelle. Pour être rendus, les services en rigueur requièrent qu'une connexion entre adresses IP soit préalablement établie. L'intérêt grandissant dans les architectures orientées contenu comme les systèmes pair-à-pair, les réseaux de distribution de contenu ou les systèmes basés sur les publications / abonnements se situe au niveau de la proposition d'un modèle de routage centré sur les contenus plutôt que les hôtes. Chacune des ces approches ayant été proposées pour répondre aux besoins spécifiques de certaines applications, il n'existe pas de solution générique qui s'accommode des réseaux et d'éléments ainsi que la complexité d'interaction entre ces approches différentes. Il est donc nécessaire de développer des solutions architecturales globales permettant le routage et la distribution du contenu et pouvant interconnecter les différentes infrastructures de gestion de contenu et distribuer le contenu à travers elles.

Dans cet article, nous présentons un modèle générique qui capture les propriétés distinctives des diverses approches proposées en matière de distribution de contenu. Notre modèle introduit une terminologie à partir de laquelle nous analysons les fonctionnalités propres aux approches existantes. A partir de cette analyse, nous proposons une architecture de routage orientée contenu basée sur le filtrage atomique des requêtes. L'acheminement des requêtes est réalisé de proche en proche en se basant sur un vecteur optimal calculé pour un ensemble de métriques de filtrage.

2. Travaux similaires

Dans la littérature, plusieurs solutions typiques aux routage orienté contenus ont été proposées comme les réseaux de distribution de contenus (CDN) ou les systèmes basés sur les événements dont ceux basés sur les publications/abonnements.

Réseaux de distribution de contenu Au niveau des réseaux de distribution de contenu (CDNs) [Hul 02], le routage consiste à distribuer le contenu à partir du meilleur emplacement vers le client qui le demande. Le processus de filtrage et de décision revient donc à choisir le meilleur emplacement de distribution du contenu plutôt que de filtrer le contenu adéquat aux intérêts des clients [DCT 03]. Plusieurs métriques peuvent affecter cette décision comme la proximité en terme de sauts entre les clients et les serveurs de contenu ou la proximité géographique. Les CDNs constituent une solution qui optimise la distribution des contenus en terme de rapidité et de disponibilité mais ne sont toujours pas centrés sur les propriétés des contenus dans le processus de routage et ne peuvent être déployés tels qu'ils sont conçus actuellement à l'échelle d'Internet.

Systèmes basés sur les événements Au sein des systèmes basés sur les publications / souscriptions [EFG 03], les clients notifient leur abonnement à un type de contenu à l'infrastructure de gestion d'événements. Cette infrastructure se charge de faire correspondre les abonnements reçus par rapport aux publications qu'elle stocke. Ensuite, les

clients sont notifiés chaque fois qu'un contenu correspondant à leurs critères d'abonnement est trouvé. Les avantages d'une telle architecture sont la persistance des abonnements ainsi que le découplage entre les requêtes clientes et les réponses reçues. Mais elle n'offre pas pas de filtrage optimisé et pas de tri des réponses les plus optimales envoyées au client.

Pour résoudre la problématique du routage centré exclusivement sur le contenu et non sur les hôtes, nous proposons de router les contenus uniquement à la base de leurs descriptions tout en optimisant un ensemble de métriques de filtrage. Notre solution est différente car elle imbrique les mots clés décrivant le contenu dans le processus du routage lui-même. Le routage réel et le filtrage du contenu sont combinés en un même processus permettant ainsi un gain conséquent en terme de messages échangés sur le réseau. Elle diffère par sa récursivité et a structure hiérarchique qui permettent de prendre en considération les changements de topologies et d'informations sur les contenus.

3. Modèle générique pour la routage orienté contenu

Les réseaux orientés contenu font référence à la capacité donnée aux applications et à leurs utilisateurs d'accéder aux données indépendamment de leur localisation. Les contenus sont accédés sur la seule base de leur type ou de leur description . Ces réseaux s'affranchissent de la nécessité d'identifer dans un système de nommage ou d'adressage les hôtes hébergeant les contenus ou souhaitant les accéder.

Entités impliquées Les réseaux orientés contenu sont constitués de noeuds qui diffèrent selon le rôle qui leur est attribué. On distingue les fournisseurs d'origine, les clients, les routeurs de médiation et les serveurs de contenu.

Processus d'accès au contenu Les fournisseurs d'origine disséminent les informations décrivant suivant une structure donnée les contenus qu'ils souhaitent rendre accessibles. Pour être en mesure de recevoir des contenus pertinents, un client définit et structure ses centres d'intérêt avant de les annoncer. Entre ces deux acteurs, les serveurs de contenu et les routeurs de médiation jouent des rôles primordiaux. Les serveurs de contenu gèrent la réplication et la mise en cache des données les plus accédées dans le but de minimiser leur temps d'accès et de diminuer la charge de traitement nécessaire aux fournisseurs d'origine pour satisfaire les requêtes de client. Les routeurs de médiation quant à eux gèrent la dissémination vers leurs pairs des requêtes de client et des annonces de fournisseurs. Ils stockent aussi des tables de correspondance entre les états concernant une partie ou la totalité des contenus disponibles et la localisation des fournisseurs d'origine hébergeant ou répliquant ces contenus. Ces tables permettent aux routeurs de médiation d'exécuter des algorithmes de filtrage de contenu et de redirection des requêtes de client de manière centralisée ou distribuée dans le but de sélectionner le(s) contenu(s) correspondant aux intérêts annoncés des clients. Notons que les quatre rôles que nous renons de définir peuvent être assumés par un même et unique nœud comme dans les réseaux de capteurs. Ce processus d'accs au contenu peut être déployé selon les contextes d'application à travers modèles différents : le modèle applicatif travers un réseau virtuel comme les systèmes pair à pair, le modèle réseau travers un filtrage XML intégré dans les routeurs niveau 3 et le modèle hybride combinant les deux comme dans le cas de certains réseaux CDNs.

4. Routage orienté contenu : algorithme et optimisation

L'idée sous-jacente à notre proposition est de substituer le routage classique basé sur les adresses des destinations par un routage saut-par-saut du meilleur contenu décrit avec un ensemble de termes $\{K_1...K_n\}$. La solution que nous proposons passe le facteur d'échelle en organisant les routeurs de médiation dans une structure hiérarchique en domaines. La dissémination des informations décrivant les contenus se limite au voisinage intra-domaine du routeur de médiation.

D'après le modèle général proposé dans la section 3, le problème de routage centré sur le contenu peut donc être réduit aux points suivants :

– Quelles sont les métriques à partir desquelles un routeur de médiation filtre les requêtes ?

– Dans quel type de structure ces métriques sont-elles stockées et comment sont-elles calculés ?

– Comment se fait le filtrage des contenus au niveau du routeur médiateur lui-même ?

– Quelles sont les politiques de routage et de filtrage inter et intra-domaines ?

Les types de requêtes pouvant varier d'une application cliente à une autre, nous commençons par transformer les requêtes reçues par l'infrastructure de routage orienté contenu en un format générique composé d'un ensemble de mots-clés $\{K_1...K_n\}$. Chaque mot-clé K_i décrit les propriétés susceptibles d'être intéressantes dans un contenu. Une relation de correspondance entre cette description générique et le format initial est stockée grâce dans une structure dédiée au sein de routeurs de médiation spécifiques. Ainsi, nous garantissons la compatibilité des requêtes de client indépendamment des applications à l'origine de la recherche initiée.

Une des contributions de notre solution est de baser le filtrage sur le calcul d'un vecteur de filtrage noté $[D, P, M]$ dont chaque coordonnée correspond à une métrique de filtrage de contenu. D représente la distance entre le client et les potentiels contenus correspondant à sa requête en nombre de sauts. P représente la popularité du contenu en nombre d'accès total pour des requêtes similaires. Quant à M, elle représente le taux de requêtes satisfaites par le système. Le calcul de la popularité d'un contenu utilise les statistiques d'accès au contenu réalisées à partir de l'historique. Nous nous basons aussi sur un système de gestion de la *réputation* des contenus pour calculer les valeurs de la troisième métrique. Les trois métriques que nous venons de présenter sont stockées au niveau de la table de filtrage. Les métriques sont reliées à l'identifiant local d'une interface de sortie ce qui permet de diriger la requête générique préalablement créée dans la direction indiquée par le vecteur $[D, P, M]$ optimal associé tel que représenté la figure 1.

Dans cette figure, une requête est reçue sur une des interfaces entrantes du routeur de médiation. Un module de filtrage retire les mots-clés contenus dans cette requête pour les insérer dans deux sous-requêtes atomiques. Chacune d'elles est traitée séparément et le mot-clé qu'elle contient est comparé aux valeurs contenues dans la colonne des mots-clés $Dest_Key$. Pour chaque sous-requête atomique, le module retourne l'in-

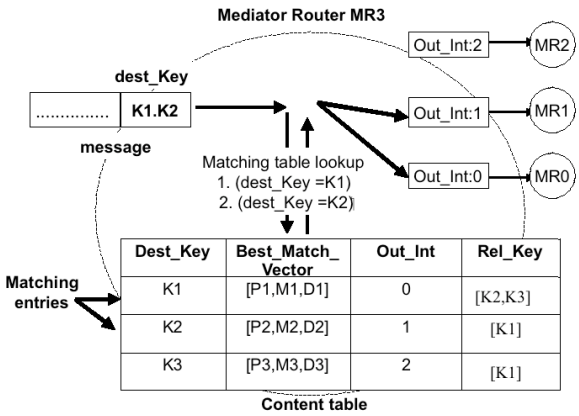

Figure 1. *Routage et filtrage du contenu à l'intérieur d'un routeur de médiation*

terface sortante correspondant au champ Ou_Int. Les résultats des sous-requêtes sont agrégés afin d'éliminer les interfaces redondantes et les doublons. Ensuite, la requête est redirigée alors dans sa totalité vers les différentes interfaces sortantes résultantes. Notons que le filtrage tel que présenté dans la figure 1 est réalisé de proche-en-proche. Cette propriété permet de prendre en considération les changements de topologie dûs à la mobilité des noeuds ou aux défaillances du réseau ainsi que la latence de propagation des informations de routage dans le réseau.

5. Conclusion et perspectives

Nous avons présenté dans ce document un modèle de réseau qui capture les propriétés des réseaux orientés contenu qui nous as permis de les étudier dans des contextes différents. En nous basant sur les fonctionnalités et la terminologie qu'introduit notre modèle, nous avons décrit une proposition de filtrage de contenu. Notre solution présente la particularité d'être résistante au facteur d'échelle et de s'accommoder de l'hétérogénéité des protocoles existants. Nous prévoyons de définir les comportements exacts nécessaires à l'implémentation de notre proposition.

Bibliographie

[DCT 03] DAY M., CAIN B., TOMLINSON G., RZEWSKI P., *RFC3466, A Model for Content Internetworking (CDI)*, RFC Editor, 2003.

[EFG 03] EUGSTER P., FELBER P.A., GUERRAOUI R., KERMARREC A.M., *The many faces of publish/subscribe*, *ACM Comput. Surv.*, vol. 35, n° 2, 2003, p. 114-131.

[Hul 02] HULL S., *Content Delivery Networks : Web Switching for Security, Availability, and Speed*, McGraw-Hill Inc, 2002 .

[PGM 06] PLAGEMANN T., GOEBEL V., MAUTHE A., MATHY L., TURLETTI T., URVOY-KELLER G., *From content distribution networks to content networks - issues and challenges*, *Computer Communications (COMCOM)*, vol. 29, n° 5, 2006, p. 551-562.