HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Techniques for Content Subscription Anonymity with Distributed Brokers

Sasu Tarkoma, University of Helsinki

Christian Prehofer, Fraunhofer Munich

22.9.2010

Contents

Introduction

Content-based routing and publish/subscribe

Partial-order-based anonymity definitions

Anonymity detection with distributed brokers

Conclusions

Introduction

Information targeting and delivery is crucial for Internet and mobile services

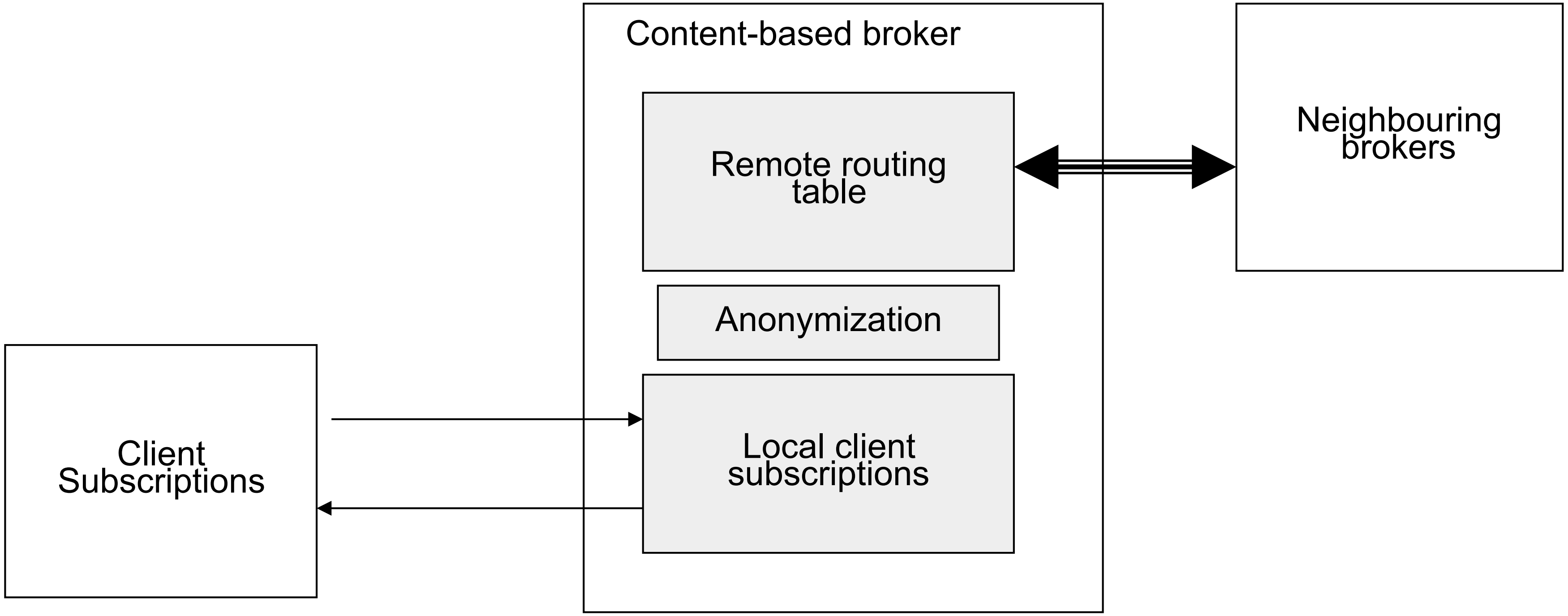
Publish/subscribe is a frequently used paradigm, in which subscriber register their interest for content supplied by producers

Content-based pub/sub allows expressive interest specification (queries, filters)

k-location anonymity is a well-known technique for ensuring location privacy

- User is not distinguishable from k-1 other users in some region

Basic System Model



k-filter anonymity

We define the **k-filter anonymity** that generalizes k-location anonymity by generalizing filters by using a partial order derived from filter containment / covering

Example: x > 10 covers x > 15 covers x > 20

Key idea: the partial order provides a natural way to generalize subscriptions and it allows to determine k for various subspaces of the content space

The partial order can be managed using several different data structures

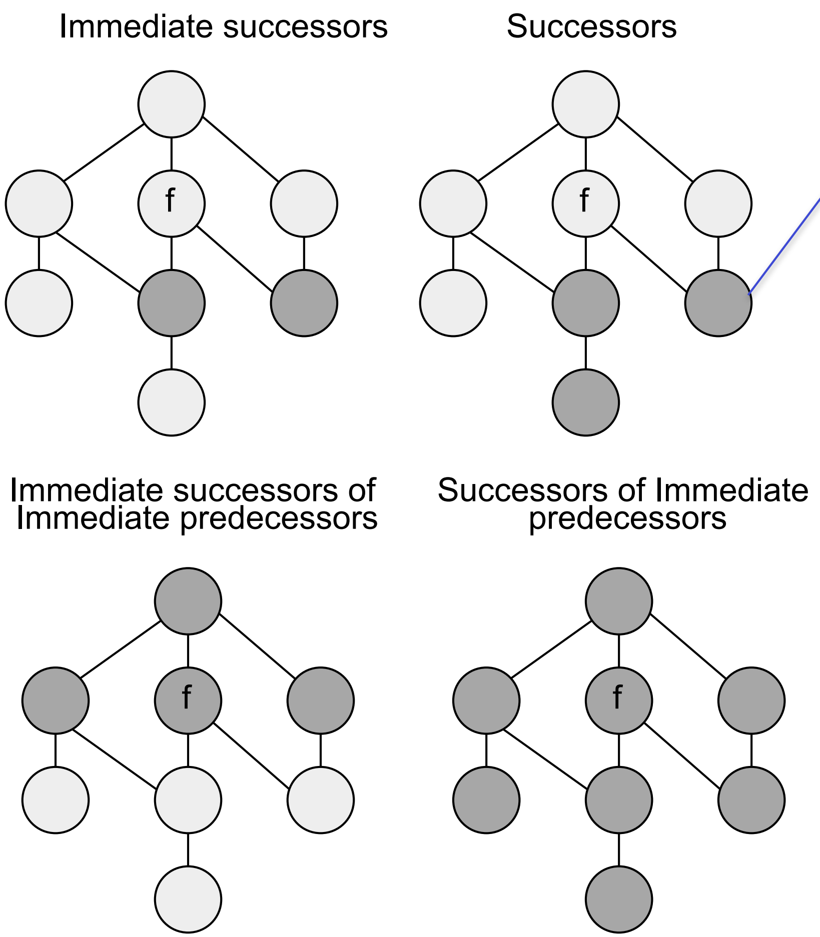
poset, poset-derived forest

Research Questions

- How to ensure that a given content subscription is k-anonymous (giving certain privacy protection)?

- How to ensure that in the network the subscription is not given to any broker that might violate the k-anonymity condition?

Definitions for k-filter anonymity



Immediate successors

Successors

Immediate successors of Immediate predecessors

Successors of Immediate predecessors

The union of the source interfaces of the grey nodes determine the value of k

k-filter anonymity in a distributed environment

Subscriber privacy is enhanced by guaranteeing that a subscriber cannot be distinguished from a set of subscribers when the interests and matching content is delivered by the network

This delivery can happen in the form of broadcast within a certain area, or delivered using unicast or multicast across multiple brokers

Physical broadcast can be implemented in such a way that specific recipient identifiers are omitted; however, given the knowledge that only a single entity is interested in the data is sufficient to pinpoint the subscriber

Therefore we are motivated in enhancing the privacy of the interest registration service

A Solution

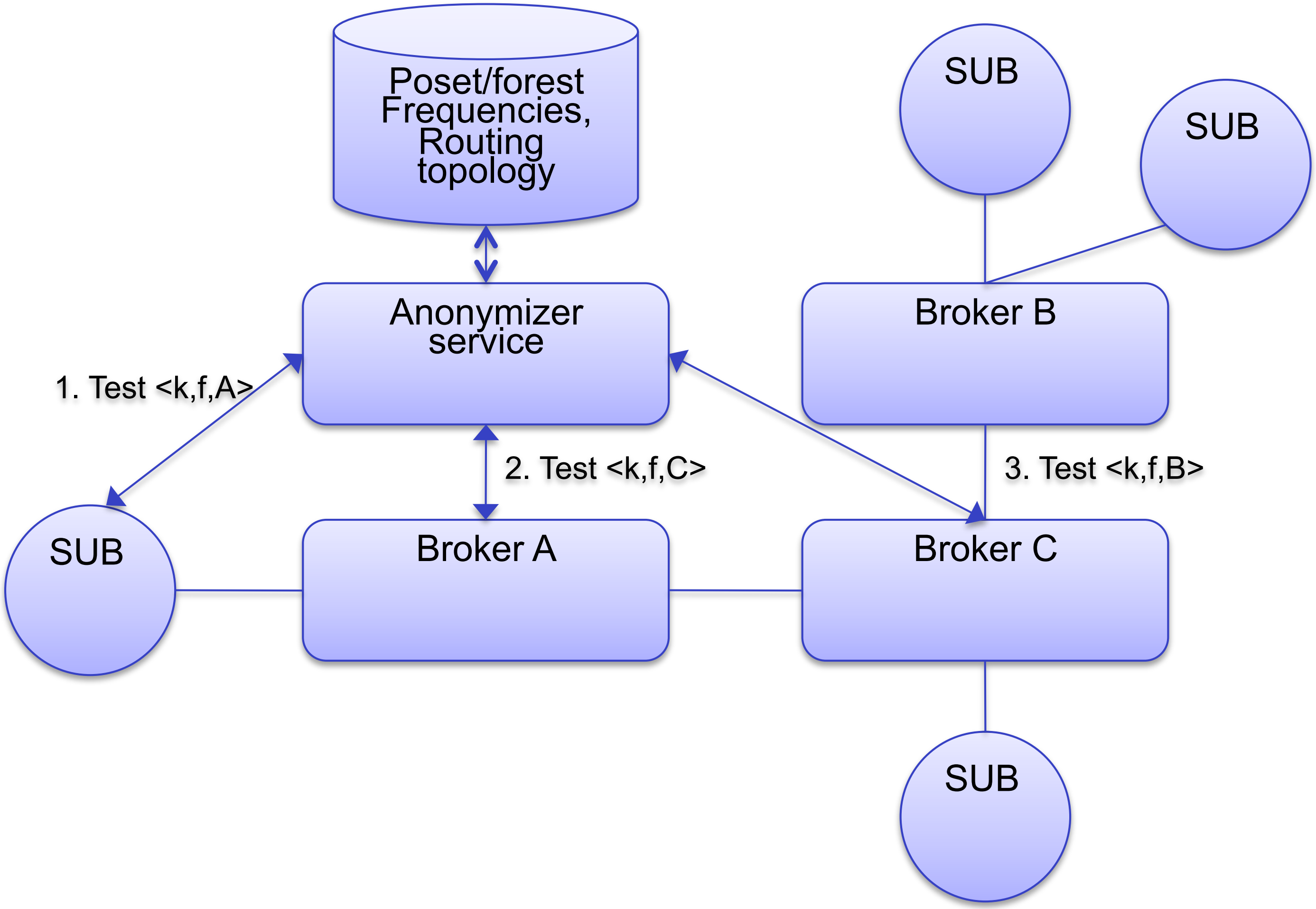
A logically centralized anonymizer broker accepts queries pertaining to filter anonymity

- Anonymizer is trusted, other brokers are not
- Clients can assess the level of k-anonymity for their subscriptions across the distributed system
  - A way to perform **content anonymity tomography**
- Anonymizer is required to perform some book-keeping regarding subscriptions and the value of k for a specific source, destination pair

Bootstrapping the system

- Important parameters:
  - Structure of partial order, value of k, network configuration

- How to bootstrap the system and allow new subscriptions that do not yet have subscribers?
  - Creation of **bogus subscriptions**
    - Create sufficient number of bogus subscriptions and place them either to the local broker or other brokers
    - Optimization problem
  - **Adaptive probing**
    - Probe the network in order to find a suitable tradeoff between generality and level of k



Poset/forest Frequencies, Routing topology

SUB

SUB

Anonymizer service

Broker B

1. Test <k,f,A>

2. Test <k,f,C>

3. Test <k,f,B>

SUB

Broker A

Broker C

SUB

Conclusions

The definitions for k-filter anonymity generalize k-location anonymity by allowing the generalization of filters using the containment relation

A proxy service is needed if the anonymity property needs to be verified with untrusted brokers

We briefly outlined two techniques for verifying the property, namely creation of bogus subscriptions and adaptive probing

The notion of k-filter anonymity appears to be useful in determining and maintaining certain levels of anonymity in distributed content-based systems