Adaptive Web Caching

Sally Floyd
Lawrence Berkeley National Laboratory
http://www-nrg.ee.lbl.gov/floyd/

DARPA Active Networking PI Meeting
June 19-20, 1997

This project is a collaboration between Lixia Zhang (UCLA) and Sally Floyd and Van Jacobson (LBNL).

URL http://irl.cs.ucla.edu/awc.html.

What is the project about?

There are three separate levels to this work that inform each other, and that each have to be kept in mind:

1. Our vision of a global data dissemination infrastructure of the future, with a self-organizing data transport substrate and self-organizing information flow.

2. The incremental deployment of this vision in the current web caching infrastructure in the Internet.

3. The general question of self-organization in networks.

Our vision of a global data dissemination infrastructure:

1. Data has unique names, and data integrity is a property of the data.

2. Servers make data available.

3. Clients ask for data.

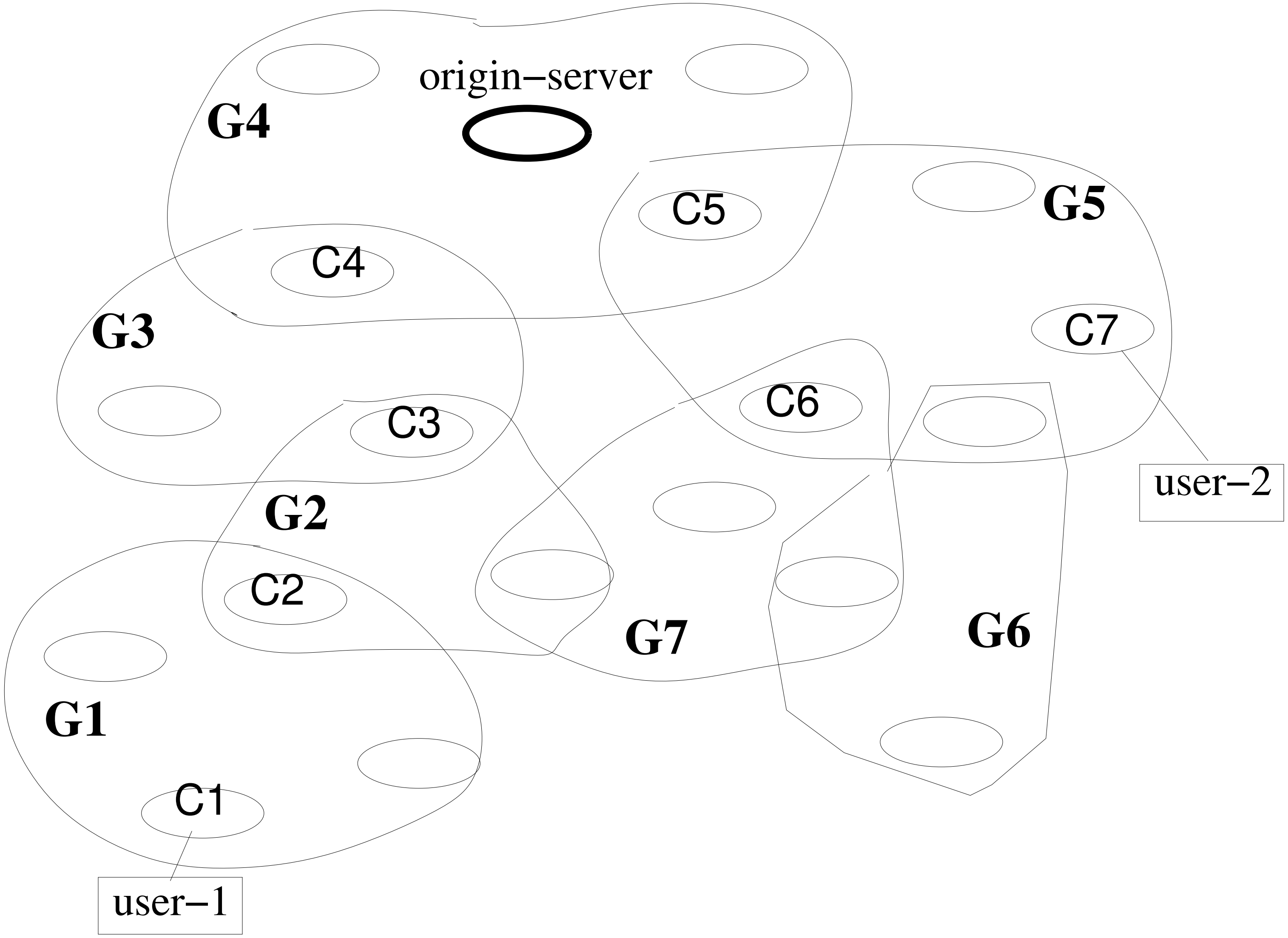
4. The web caching/data dissemination infrastructure conspires to deliver the data to clients.

Why use the word “self-organizing”?

1. Instead of manual configuration of communications between nearby web caches, web caches find neighboring web caches, and organize themselves into overlapping multicast groups.

2. Instead of the manual configuration of paths for requests for data, web caches exchange information among themselves and, for each request, determine the appropriate forwarding path.

3. This allows scalability, robustness, and adaptation to changes in topology and load.


origin−server
G4
C5
G5
C4
G3
C7
C6
C3
user−2
G2
C2
G7
G6
G1
C1
user−1

Question: What does this have to do with active networking?

Answer: The packets are not active.
However, the *network* is very active.

The burden is not on applications, but on the network itself. Applications just supply and request data. The *network* has to figure out how to organize itself to supply the data to users.

A comparison between active packets and active networks:

1. **Active packets**: Clients send requests to the origin server. Caches tell nearby routers to “capture” requests and send them instead to the cache.

2. **Active networks**: Clients send requests to a nearby cache or group of caches. If the cache does not have the data, it actively decides where to forward the request:

- to a cache closer to the origin server,
- to a cache that specializes in that type of request,
- to a cache that generally has a high hit rate,
- or to the origin server itself...

We are working on approach (2), not approach (1).

General research on self-organization in networks:

1. In web caching, caches organizing themselves into overlapping multicast groups, and organizing the flow of requests and of data.

2. In multicast, session members organizing themselves into groups for sending global session messages.

3. In reliable multicast, session members organizing themselves into groups for local recovery of packet losses.

4. In network management, self-organization to detect and recover from failures.

5. ...