

feedtree

Sharing Web micronews with
peer-to-peer event notification

Dan Sandler

Alan Mislove, Ansley Post, Peter Druschel

Rice University

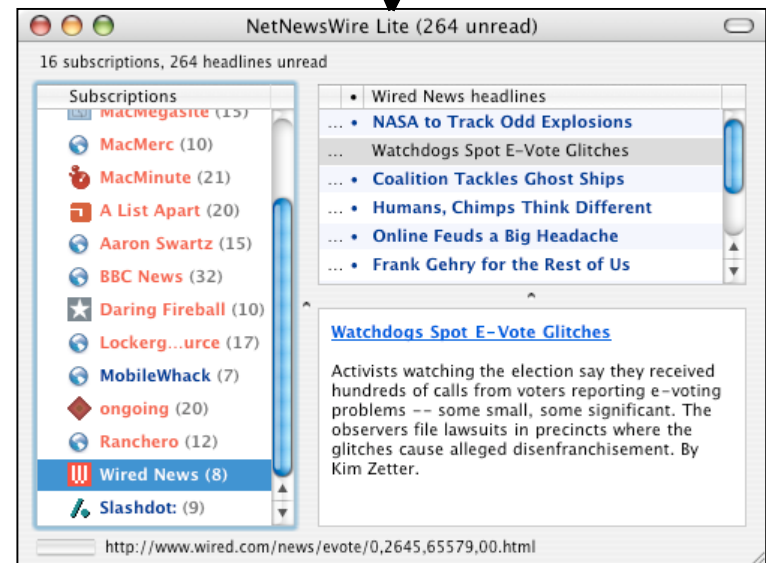
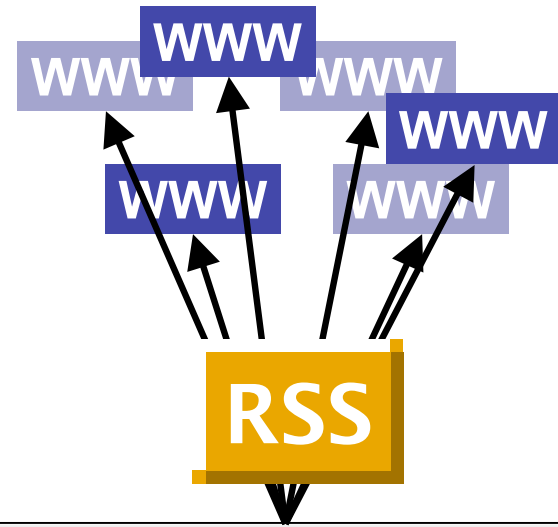
February 24, 2005

Surfing the Web

- Used to be fun, but now it's hard work!
 - Not enough just to reload a few bookmarks
 - Thousands of niche websites
 - Personal weblogs
- Trend: “Micronews”
 - Frequent, irregular updates

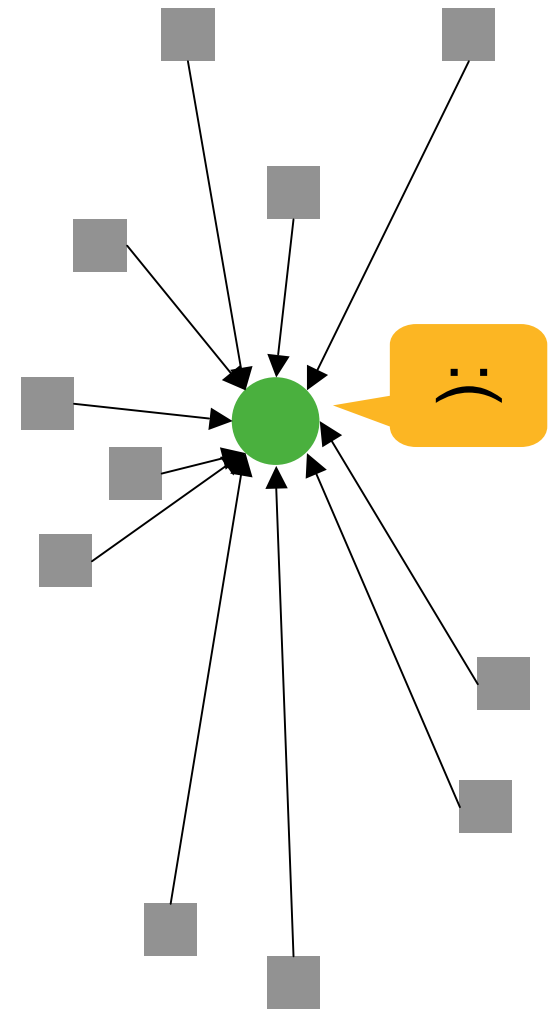
RSS News Feeds

- XML summary of headlines, links, stories
- 2004: Explosion in popularity of RSS feeds
 - 800,000+ feeds (Feedster)
 - Major news outlets
 - (NYT, BBC, MSNBC)
 - Minor news outlets
 - (Slashdot, Wired)
 - Weblogs



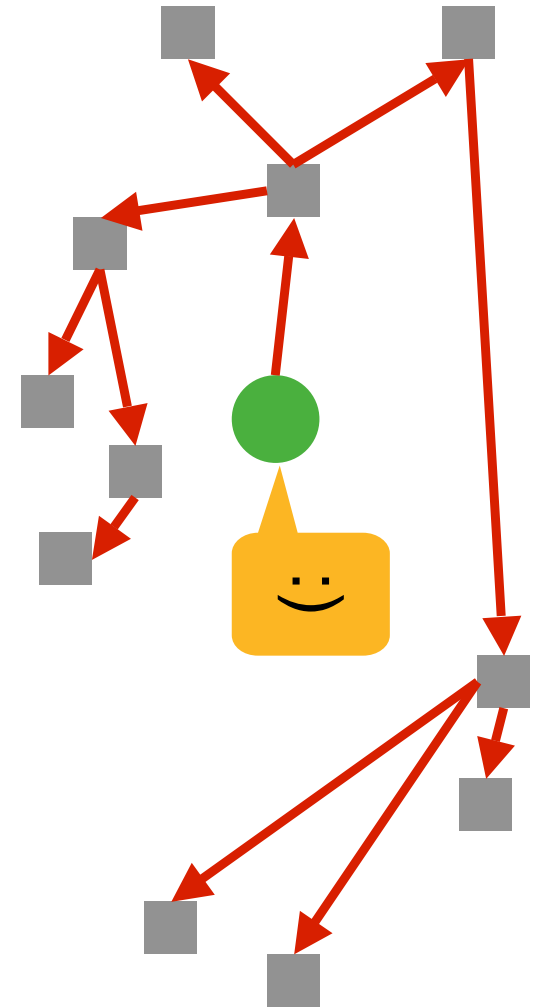
“Breaking News”

- Applications must *poll* RSS feeds
 - Frequently (for freshness)
 - Polling one resource does not scale
 - Popularity curse (“sticky” traffic)
 - ➔ **Bandwidth problem** for publishers
- Some attempts to control the problem
 - Scaling back RSS service (*e.g.* MSDN)
 - Tracking, punishing greedy clients (*e.g.* Slashdot)
 - The user base continues to increase



“Fixing News”

- Perfect fit for **p2p multicast**
 - Many participants
 - Continuously running client applications
 - Don't need extremely low latency
 - Still much better than polling:
10 sec. \ll 60 min.
- Goal: cooperative p2p dissemination of micronews *as it happens*



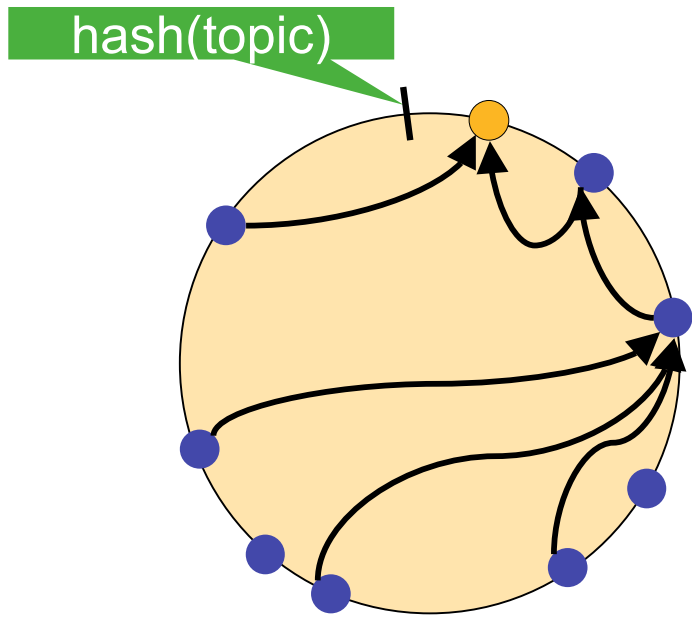
FeedTree

- Propose **FeedTree**, a p2p micronews distribution system built on Scribe
 - Use Scribe's multicast service to disseminate RSS news, ASAP, without polling
- Scribe: Batteries included
 - Decentralized membership
 - Cheap maintenance of multicast trees
 - Efficient multicast messages

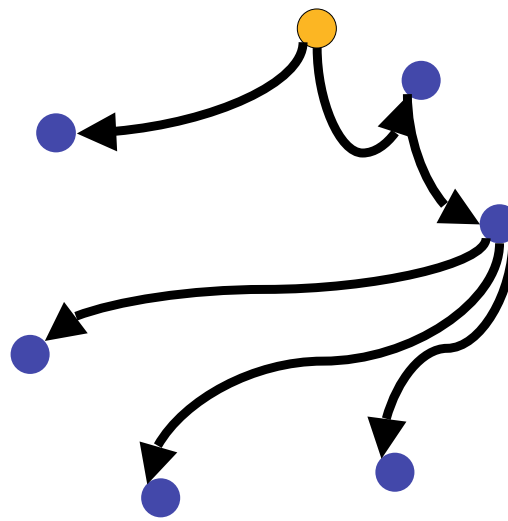


Scribe

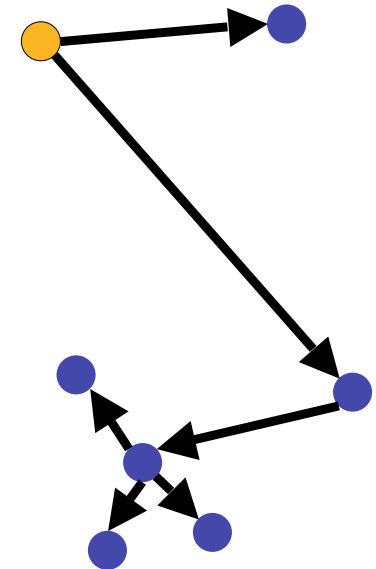
- Uses Pastry routing to create trees
- Root = nearest node to $\text{hash}(\text{topic})$
- Tree = $(\text{union}(\text{all routes to group root}))^R$



Pastry routes



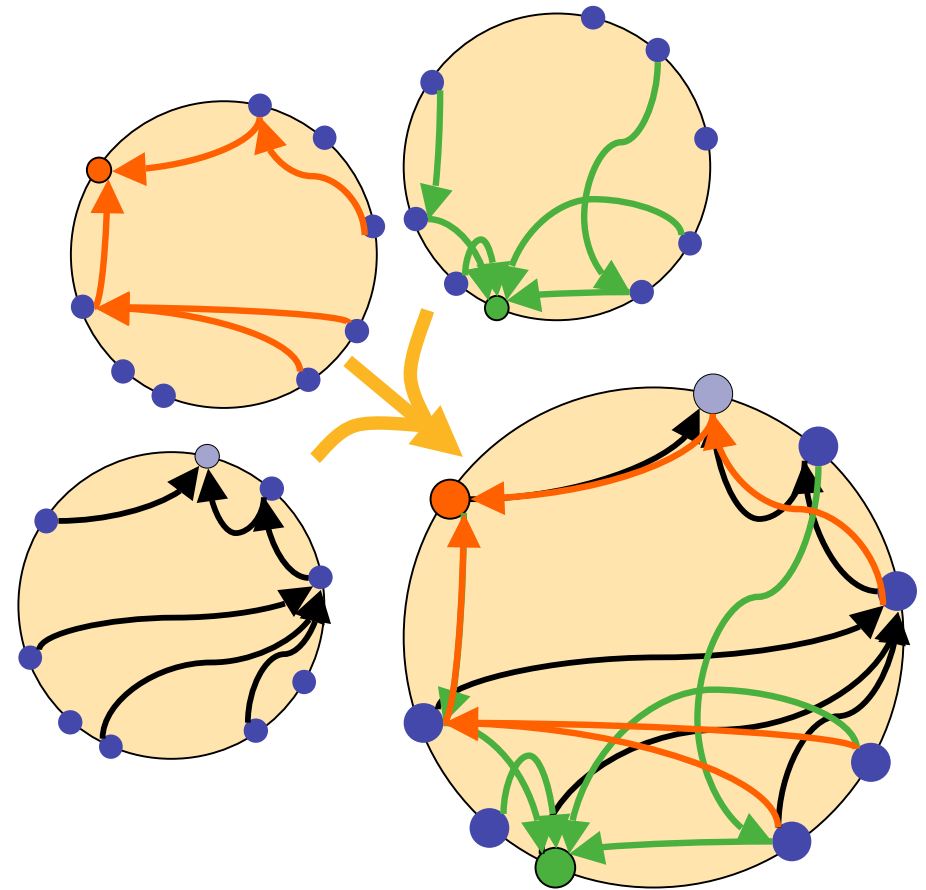
Scribe tree



Network proximity

Many Trees Make Light Work

- One ring, many trees
- Low maintenance cost
 - Group trees embedded in existing overlay graph
 - Cheap joins
 - Keepalives subsumed by traffic
 - Each node is interior in 1/16 of its trees



FeedTree Architecture

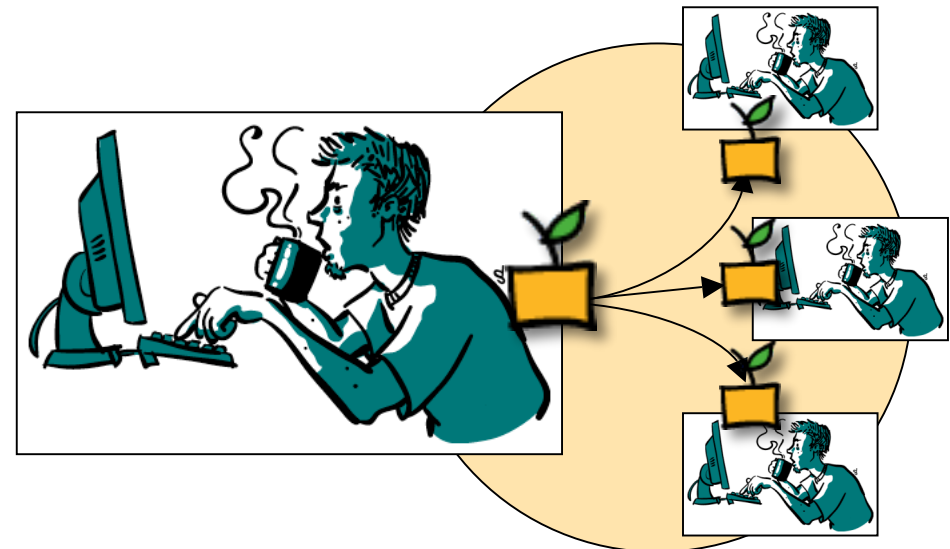
- Each RSS feed gets a Scribe group
- Publishers
 - Multicast micronews immediately
- FeedTree aware clients
 - Subscribe to multicast groups for each feed
 - Rely on multicast for all news content delivery
- Conventional RSS feed
 - URL provides Scribe topic
 - Recovery of recent lost items

Incremental Adoption Scenario

- Early adopters use existing RSS clients with a local FeedTree **proxy application**
- Proxies look for relevant Scribe groups to join
 - If no group exists, poll the RSS as usual, and start a group, multicasting each new item



**POLL
(HTTP)**

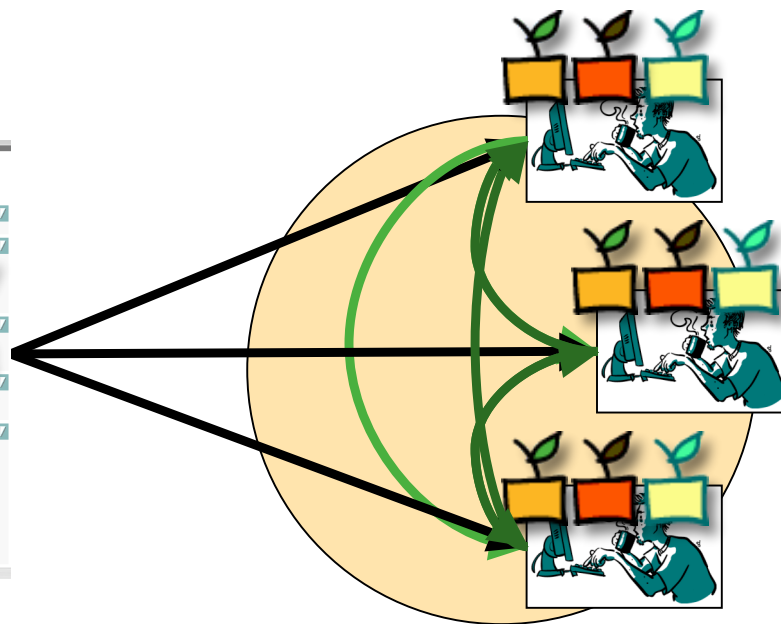


Development progress: ft-proxy

- A symmetric client
 - Listens for news via FeedTree notification
 - Polls legacy feeds; shares news with others
- Full HTTP proxy
 - Browse the Web and read news as normal
 - RSS, Atom requests detected and satisfied by FeedTree
- ft-proxy acts as a local feed cache
 - Clients may poll it arbitrarily often for news
 - Effect for users: Same software, better delivery

Feed Overload

- Multiple ft-proxy instances cooperatively poll the same legacy feed for better event resolution



N=3
T ≈ 30min

Everyone Loves Screenshots



Proxy server: 128.42.129.12:8500
Pastry node: 128.42.129.12:5200
Uptime: 17h17m18s

STATISTICS


ft-proxy
web interface

Subscribed feeds: 19
Selected for polling: 19

FEEDS

Feed	Window	Entries	Polling?	Events	Overlay Messages			
					Previous	Current	Next Slice	
MetaFilter (source, homepage)	30m	26	yes		30m	2	0	Mon Feb 21 10:10:13 CST 2005
del.icio.us/popular (source, homepage)	30m	83	yes		30m	0	1	Mon Feb 21 10:10:13 CST 2005
Cat and Girl (source, homepage)	30m	1	yes		30m	2	1	Mon Feb 21 10:10:13 CST 2005
dsandler.org (source, homepage)	30m	22	yes		30m	2	2	Mon Feb 21 09:53:08 CST 2005
ePOST News (source, homepage)	30m	11	yes		30m	2	1	Mon Feb 21 10:10:43 CST 2005
jwz (source, homepage)	30m	26	yes		30m	2	1	Mon Feb 21 10:16:23 CST 2005
Gibson Blog (source, homepage)	30m	2	yes		30m	2	1	Mon Feb 21 10:07:14 CST 2005
Boy On A Stick And Slither (source, homepage)	30m	2	yes		30m	2	2	Mon Feb 21 09:51:10 CST 2005
The RSS Blog (source, homepage)	30m	2	yes		30m	2	1	Mon Feb 21 10:18:12 CST 2005
Slashdot (source, homepage)	30m	2	yes		30m	2	2	Mon Feb 21 09:57:52 CST 2005
kottke.org remain	30m	2	yes		30m	2	1	Mon Feb 21 10:07:13 CST 2005
Scripting News	30m	2	yes		30m	2	2	Mon Feb 21 09:56:54 CST 2005
All Things Distrib	30m	2	yes		30m	2	1	Mon Feb 21 10:07:17 CST 2005

Event timeline
FeedTree messages seen during the last period



Volunteer mode
Are we polling this feed?

Polling period

Feed info
(name, URL, homepage)

Publisher Adoption Stages

- Desirable for content authentication
- Publishers must join FeedTree
 - **Basic:** Sign existing RSS content
 - **Incremental:** Add a server-side proxy which polls existing RSS (very frequently)
 - **Eventual:** Add FeedTree to RSS-aware publishing tools



Security Considerations

- Need help from publisher
 - The only trustable entity
- Detectable attacks
 - Authenticity of content
 - *Needs*: Signed content (by publisher)
 - Suppression/censorship
 - *Needs*: Signed content, sequence numbers, heartbeats

Security Considerations (cont.)

- Recovery
 - Route around faulty nodes
 - Re-join tree elsewhere
 - Recover missed data
 - Recent events cached along tree
 - Ask your new parent

Incentives to forward traffic

- Forwarding bandwidth expected to be small
- Expect most nodes to be a member of many trees (forwarder in few)
- “Tit-for-tat” bandwidth-trading incentive mechanisms [Ngan *et al.* 03]

Motivations for Adoption

- Cheaper, richer offerings for publishers
 - Dramatically lower bandwidth demands
 - Offer richer or differentiated content
 - As compared with conventional RSS
- Better RSS service to end users
 - **Immediate delivery**
 - Richer content if offered by publishers

Summary

- RSS
 - Users love it
 - Death of a thousand cuts for publishers
- p2p content distribution
 - A component of the “right” solution
 - Benefits for publishers and users
- FeedTree—a new application of structured overlays!
 - Clear steps toward a polling-free system
- Prototype: `ft-proxy`
 - Instant benefits for end users
 - Wider deployment soon (in contact with publishers)

One More Thing...

- Try it out!



- Set your Web or RSS reader proxy to one of:

demo1.feedtree.net:8501

demo2.feedtree.net:8501

demo3.feedtree.net:8501

...and use your RSS reader application as normal.

- Feedback/help/more info: <http://feedtree.net>

This presentation brought to you by epostmail.org.



Bonus Round!

Future Investigations

- Use DHT to create a persistent RSS archive
 - Offline clients will be able to “catch up”
 - Recovery of (very old) missed items
- Anonymous news feeds
 - Use anonymizing p2p routing schemes
- Distribute more sophisticated content
 - Media, software

Some Entirely Real Data

- “Slashdot has about **180K unique IPs** pulling RSS feeds, totalling about **10 GB**, every day. A little more on weekdays, less on weekends.”
 - Source: Jamie McCarthy of Slashdot, by email, 12/16/04
- **Boing Boing: 3 GB/day**
 - HTML files consumed less bandwidth than RSS over this period!
 - Source: BB web stats page (period 11/01–11/14/04)
 - Since then, BB has started using feedburner.com, a third party, to suffer the bandwidth

Some Not Entirely Fictional Data

- *Subscriber numbers from Bloglines.com*
- Slashdot: 17,700 subscribers
 - 2 KB headline-only feed * 30-min polling period
 - **1.7 GB daily**
- New York Times: 24,000 subscribers (all feeds)
 - 3 KB feeds -> **3.5GB/day**
- Boing Boing: 11,500 subscribers
 - Rich feeds: 40 KB
 - **22 GB/day**
- BBC News: 18,000 subscribers
 - “Updated every minute of every day”
 - Let’s not even speculate about THAT polling schedule