BubbleStorm: Powerful Peer-to-Peer Search

TECHNISCHE UNIVERSITÄT DARMSTADT

DVS

Christof Leng, Wesley Terpstra

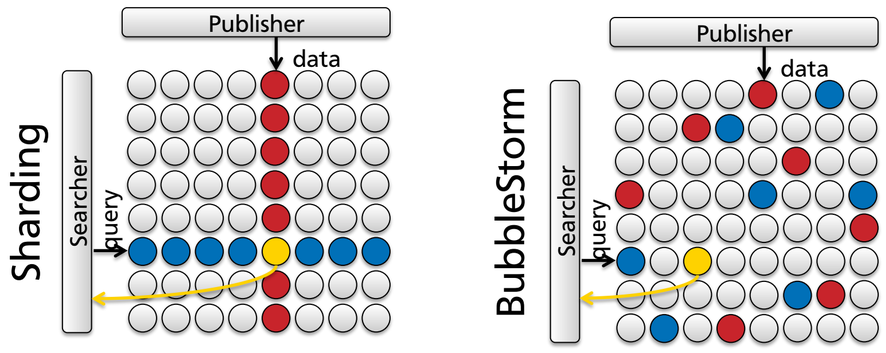
BubbleStorm

Motivation

- Decentralized P2P systems need expressive search
  - keyword search, XPath, SQL
- Data should be available beyond individual peer sessions
  - Data has to be widely replicated
- Open membership P2P overlays are an extremely unstable environment
  - Deterministic greedy routing (e.g. DHTs) leads to dead ends
  - Unstructured approaches are more robust

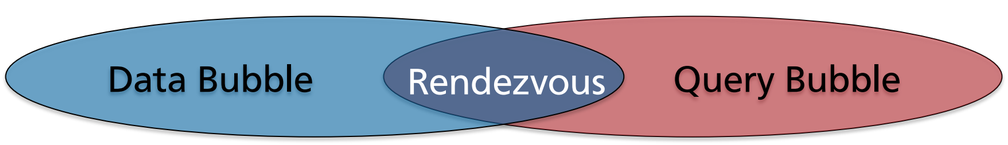
The Rendezvous Approach

- Every query meets every datum at least once somewhere in the network (at the **rendezvous** node)
- Similar to sharding in cluster search engines
- Arbitrary match functions that evaluate individual data are possible
- Easy to program API



The BubbleStorm Rendezvous

- Probabilistic rendezvous with $\mathbf{O(\sqrt{\lambda}\, n)}$ replicas for query and data (**bubbles**)
- Randomly placed on nodes in the network
- Tunable success guarantees
- Uses random graph topology to ensure randomness
- Supports heterogeneous capacities with individual node degrees

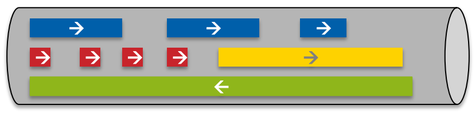


$$1 - e^{-\frac{\lambda}{D_2}} \le \left(1 - e^{-\frac{|data|}{D_2}}\right)\left(1 - e^{-\frac{|query|}{D_2}}\right) \qquad D_2 = \sum d_i^2$$

λ^2	1	2	3	4
P(ok)	63.2%	98.2%	99.99%	99.99999%

The CUSP Transport Protocol

- Novel transport protocol for routing overlays, event-driven and concurrent networking
- Based on ideas of SCTP and SST
  - Many streams multiplexed in one channel
  - No head-of-line blocking for concurrent streams
- Built-in encryption and authentication
- Sophisticated QoS features
- Native NAT traversal techniques

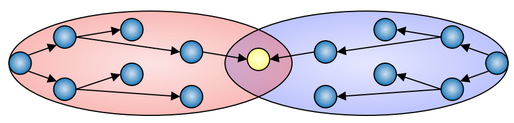


Bubble Classes

	persistent	mutable	indexed	use cases
instant	✖	✖	✖	queries
fading	(✔)	✔	✖	short-lived data
managed	(✔)	✔	✖	buddy list, shared files
durable	✔	✔	✔	wiki articles, long-term storage

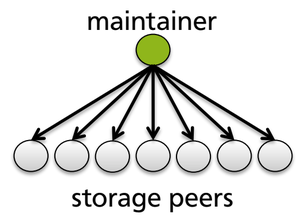
Instant & Fading Bubbles

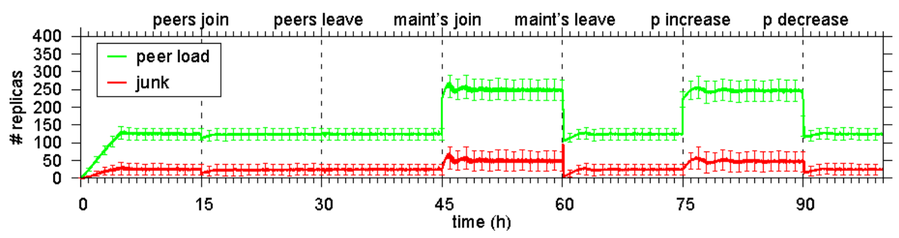
- Instant bubbles for data that is not stored at all
  - e.g. queries in query/data
  - e.g. publications in pub/sub
- Fading bubbles for short-lived data
  - e.g. position updates in gaming
  - e.g. short-lived subscriptions
- Dissemination with **bubblecast**
  - node-constraint, fixed-degree random “flooding”
  - superior to random walks and classic flooding
- No replica maintenance, fading bubbles will disappear over time



Managed Bubbles

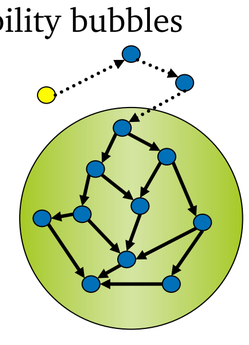
- For data that is inherently associated to a peer, e.g.
  - list of shared files
  - presence information for instant messaging
  - pub/sub subscriptions, continuous queries
- Can be updates and deleted
- will disappear when the owner leaves
- Every peer maintains a set of storage peers for his managed bubbles
- Replica maintenance against churn, bubble size and network size changes





Durable Bubbles

- For data that should be available indefinitely, e.g.
  - wiki articles
  - “cloud” storage
- Every peer is responsible for a set of bubble IDs (**key-based routing**)
- Responsibility is published as a managed bubble
- Peers build a routing tables based on the received responsibility bubbles
- To find a responsible node, ask a few neighbors
  - essentially an unstructured key-value lookup (**index**)
- To update or delete all replicas of a bubble
  - receivers forward it to all responsible nodes they know
  - the graph of responsible nodes is flooded



References:

1. Wesley W. Terpstra, Jussi Kangasharju, Christof Leng, Alejandro P. Buchmann. **BubbleStorm: Resilient, Probabilistic, and Exhaustive Peer-to-Peer Search**. Proceedings of the ACM SIGCOMM Conference, Kyoto, Japan, August 2007
2. Wesley W. Terpstra, Christof Leng, Alejandro P. Buchmann. **Brief Announcement: Practical Summation via Gossip**. Proceedings of the ACM PODC Conference, Portland, Oregon, USA, August 2007
3. Christof Leng, Wesley W. Terpstra, Bettina Kemme, Wilhelm Stannat, Alejandro P. Buchmann. **Maintaining Replicas in Unstructured P2P Systems**. Proceedings of the ACM CoNEXT Conference, Madrid, Spain, December 2008
4. Wesley W. Terpstra, Christof Leng, Max Lehn, Alejandro P. Buchmann. **Channel-based Unidirectional Stream Protocol (CUSP)**. Proceedings of the IEEE INFOCOM Mini Conference, San Diego, California, USA, March 2010
5. Christof Leng, Max Lehn, Robert Rehner, Alejandro Buchmann. **Designing a Testbed for Large-scale Distributed Systems** (Poster). Proceedings of ACM SIGCOMM, Toronto, Canada, ACM, August 2011

QuaP2P

Contact: Christof Leng, cleng@dvs.tu-darmstadt.de | http://www.bubblestorm.net/