Agenda

- Gnutella scaling *(Jordan Ritter:Why Gnutella Can't Scale. No, Really)*
- Freenet
- Freenet scaling compared to Gnutella

Notations

- P –users connected to Gnutella
- N – Connections to other servents
- T – TTL, Time To Live

Reachable users

- N – Connections to other servents
- T – TTL, Time To Live
- Table shows reachable users when everybody has specified N and T

	T=1	***T=2***	***T=3***	***T=4***	***T=5***	***T=6***	***T=7***	***T=8***
N=2	2	4	6	8	10	12	14	16
N=3	3	9	21	45	93	189	381	765
N=4	4	16	52	160	484	1,456	4,372	13,120
N=5	5	25	105	425	1,705	6,825	27,305	109,225
N=6	6	36	186	936	4,686	23,436	**117,186**	585,936
N=7	7	49	301	1,813	10,885	65,317	**391,909**	2,351,461
N=8	8	64	456	3,200	22,408	156,864	**1,098,056**	7,686,400

Generated Bandwidth

- Example: Search for “grateful dead live”
- Search Query Package is 83 bytes
- Table shows generated bandwidth for this package: (N = Connections / T = TTL)

	T=1	***T=2***	***T=3***	***T=4***	***T=5***	***T=6***	***T=7***	***T=8***
N=2	166	332	498	664	830	996	1,162	1,328
N=3	249	747	1,743	3,735	7,719	15,687	31,623	63,495
N=4	332	1,328	4,316	13,280	40,172	120,848	362,876	1,088,960
N=5	415	2,075	8,715	35,275	141,515	566,475	2,266,315	9,065,675
N=6	498	2,988	15,438	77,688	388,938	1,945,188	**9,726,438**	48,632,688
N=7	581	4,067	24,983	150,479	903,455	5,421,311	**32,528,447**	195,171,263
N=8	664	5,312	37,848	265,600	1,859,864	13,019,712	**91,138,648**	637,971,200

Statistics / mean numbers

- a – mean % of users who share content
- b – mean % of users who have responses to queries
- r – mean number of search responses pr. respondent
- l – mean length of search responses pr. respondent
- Example: Table shows generated bandwidth for a = 30%, b=40%, r = 10, l = 60

	T=1	***T=2***	***T=3***	***T=4***	***T=5***	***T=6***	***T=7***	***T=8***
N=2								
N=3	283.68	1,418.4	4,822.56	13,900.3	36,594.7	91,061.3	218,150	508.638
N=4	378.24	2,647.68	12,860.2	53,710.1	206,897	758,371	2,688,530	9,306,220
N=5	472.8	4,255.2	26,949.6	147,986	753,170	3,658,050	17,214,200	79,185,000
N=6	567.36	6,240.96	48,793	332,473	2,105,470	12,743,500	**74,798,500**	429,398,000
N=7	661.92	8,604.96	80,092.3	651,991	4,941,123	35,823,800	**252,002,000**	1,734,360,000
N=8	756.48	11,347.2	122,550	1,160,440	10,242,000	86,526,900	**709,521,000**	5,693,470,000

Totals: Query pr. sec (qps)

- Statistics shows 8000 users generating 5 queries pr. Second
- Example: Table shows bandwidth rates for 5 qps (S = 83 bytes, a,b,r,l same as before)

	T=1	***T=2***	***T=3***	***T=4***	***T=5***	***T=6***	***T=7***	***T=8***
N=2								
N=3	2.7KBps	10.8KBps	32.8KBps	88.1KBps	221.6KBps	533.7KBps	1.2MBps	2.9MBps
N=4	3.6KBps	19.9KBps	85.9KBps	335KBps	1.2MBps	4.4MBps	15.3MBps	52MBps
N=5	4.4KBps	31.7KBps	178.3KBps	916.3KBps	4.5MBps	21.1MBps	97.4MBps	441.3MBps
N=6	5.3KBps	46.1KBps	321.2KBps	2.1MBps	12.5MBps	73.4MBps	**422.6MBps**	2.4GBps
N=7	6.2KBps	63.4KBps	525.4KBps	4MBps	29.2MBps	206.2MBps	**1.4GBps**	9.6GBps
N=8	7.1KBps	83.3KBps	802KBps	7.1MBps	60.5MBps	497.7MBps	**4GBps**	31.7GBps

Conclusion

- When scaling from 8000 users to 1 million:
- 1.000.000 users would generate 625 qps.
- 5 qps generate 4GBps
- 1.000.000 users generate 500 GBps
- Result: 500 Kps pr. user.

Freenet

The Purpose of Freenet:

- To allow people to distribute material anonymously.
- To allow people to retrieve material anonymously.
- To make the removal of material almost insuperably difficult.
- To operate without central control.

Similarities between Freenet and Gnutella

- Peer2peer sharing of files
- Searches goes back and forth along the p2p path
- Queries has a TTL stamp that decrements along the search

Differences in Freenet compared to Gnutella

- Requests are replicated and cached through the network
- Searches executes depth-first (in Gnutella breadth-first)
- Free-text search not possible.
- Replication: Data is replicated and cached through the peers (Least-requested data “dies” first)
- Availability: Data is available on more than one node.
- Persistence: Trying to delete data will replicate data.
  Not-used data will be erased in time.
- Anonymity: Data is not visible to anyone else than the requester.

Hash-key routing

- Searching p2p using hash-key of data
- Caching requests and most requested files on nodes along the p2p path
- Routing of queries using most-matching hash-key
- Cannot modify documents because of hash-key
- (Name service is explained later)

Key	Data	Address
Fg3466ue56uaehjr	45hljktherltj	Tcp/5.34.27.4:6473
Dfgh655h5uyu	Sdfg546456	Tcp/89.34.36.3:6473
3ty4556ftyert	435efg54y457	Tcp/aa.bb.cc.dd:eeee
Fsfwer3w434		Tcp/aa.bb.cc.dd:eeee
2sdfsdf36sdfsdf		Tcp/aa.bb.cc.dd:eeee
354sdfwethe45		Tcp/aa.bb.cc.dd:eeee

Freenet key types

- Content Hash Keys (CHKs)
  Hash key of data (contents) of file
- Keyword Signed Keys (KSKs)
  Key representing the name of the document eg. “text/books/1984.html”.
  Original name cannot be generated from key!
  Points to a CHK.
- Signature Verification Keys (SVKs)
  Public/private key
  Making “subspace” of Freenet with documents guaranteed by auther of key.
  (prefixing KSK name with key)

Use of keys

When inserting a document into Freenet:

- CHK of the document is generated
- SVK that points to the CHK is generated from the name of the document
- The document is encrypted (typical with the name of the document)
- The document is stored on a node

Freenet Scaling

- Depth-first search mechanism
- Intelligent routing: To searches on same name will “tend” to follow the same path
- Routing acts organized in hash-key order
- Result: Fewer hops for each query (entire network will not be searched)

Conclusion: Freenet and Gnutella

- Freenet scales, Gnutella does not.
- Only Gnutella can be free-text searched.
- Freenet data is anonymous/encrypted and cannot be altered.
- Freenet replicates lots of data.

Questions

- What is the main design difference between Napster and Gnutella?
- How does the Gnutella network avoid duplicated messages?
- How does Gnutella route queries?
- Can you store a file into the Gnutella network, turn your machine of and find the file from another machine?
- How could Gnutella scaling be enhanced ? (by using some of Freenet principles)
- How does Freenet route queries ?
- What is the purpose of KSKs ?
- Can you suggest a change in the Freenet protocol in order to support free-text queries ?
- Is it possible to remove a file from Freenet?