Ian Clarke Oskar Sandberg

Routing in the Dark: Scalable Searches in Dark P2P Networks

It has become apparent that the greatest threat toward the survival of peer to peer, and especially file sharing, networks is the openness of the peers themselves towards strangers. So called "darknets"—encrypted networks where peers connect directly only to trusted friends—have been suggested as a solution to this. Some, small-scale darknet implementations such a Nullsofts WASTE have already been deployed, but these share the problem that peers can only communicate within a small neighborhood.

Utilizing the small world theory of Watts and Strogatz, Jon Kleinbergs algorithmic observations, and our own experience from working with the anonymous distributed data network Freenet, we explore methods of using the dynamics of social networks to find scalable ways of searching and routing in a darknet. We discuss how the results indicating the human relationships really form a "small world", allow for ways of restoring to the darknet the characteristics necessary for efficient routing. We illustrate our methods with simulation results.

This is, to our knowledge, the first time a model for building peer to peer networks that allow for both peer privacy and global communication has been suggested. The deployment of such networks would offer great opportunities for truly viable peer to peer networks, and a very difficult challenge to their enemies.

Ian Clarke is the architect and coordinator of The Freenet Project, and the Chief Executive Officer of Cematics Ltd, a company he founded to realise commercial applications for the Freenet technology. Ian is the co-founder and formerly the Chief Technology Officer of Uprizer Inc., which was successful in raising \$4 million in A-round venture capital from investors including Intel Capital. In October 2003, Ian was selected as one of the top 100 innovators under the age of 35 by the Massachusetts Institute of Technology's Technology Review magazine. Ian holds a degree in Artificial Intelligence and Computer Science from Edinburgh University, Scotland. He has also worked as a consultant for a number of companies including 3Com, and Logica UK's Space Division. He is originally from County Meath, Ireland, and currently resides in Edinburgh, Scotland.

Oskar Sandberg is a post graduate student at the Chalmers Technical University in Gothenburg, Sweden. He is working on a PhD about the mathematics of complex networks, especially with regard to the small world phenomenon. Besides this he has an active interest in distributed computer networks and network security, and has been an active contributor to The Freenet Project since 1999.

Routing in The Dark Scalable Searches in Dark P2P Networks

Ian Clarke and Oskar Sandberg

The Freenet Project

Ian Clarke & Oskar Sandberg - 2005 - p.1/33

Introduction

- We have long been interested in decentralised "Peer to Peer" networks. Especially Freenet.
- But when individual users come under attack, decentralisation is not enough.
- Future networks may need to limit connections to trusted friends.
- The big question is: Can such networks be useful?

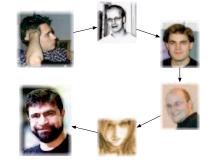
Ian Clarke & Oskar Sandberg - 2005 - p.2/33

Overview of "Peer to Peer" networks

- Information is spread across many interconnected computers
- Users want to find information
- Some are centralised (eg. Napster), some are semi- centralised (eg. Kazaa), others are distributed (eg. Freenet)

Ian Clarke & Oskar Sandberg - 2005 - p.3/33

The Small World Phenomenon



- In "Small world" networks short paths exist between any two peers
- People tend to form this type of network (as shown by Milgram experiment)
- Short paths may exist but they may not be easy to find Ian Clarke & Oskar Sandberg 2005 p.4/33

Navigable Small World Networks

- Concept of similarity or "closeness" between peers
- Similar peers are more likely to be connected than dissimilar peers
- You can get from any one peer to any other simply by routing to the closest peer at each step
- This is called "Greedy Routing"
- Freenet and "Distributed Hash Tables" rely on this principal to find data in a scalable decentralised manner

Ian Clarke & Oskar Sandberg - 2005 - p.5/33

Light P2P Networks

- Examples: Gnutella, Freenet, Distributed Hash Tables
- Advantage: Globally scalable with the right routing algorithm
- Disadvantage: Vulnerable to "harvesting", ie. people you don't know can easily discover whether you are part of the network

Ian Clarke & Oskar Sandberg - 2005 - p.6/33

Dark or "Friend to Friend" P2P Networks

- Peers only communicate directly with "trusted" peers
- Examples: Waste
- Advantage: Only your trusted friends know you are part of the network
- Disadvantage: Networks are disconnected and small, they typically don't scale well

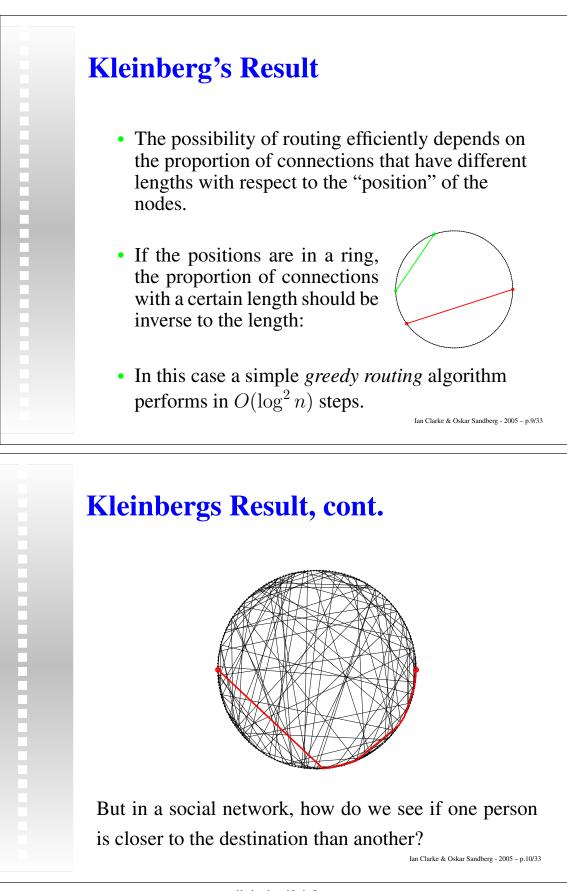
Ian Clarke & Oskar Sandberg - 2005 - p.7/33

Application

How can we apply small world theory to routing in a Dark peer to peer network?

- A Darknet is, essentially, a social network of peoples trusted relationships.
- If people can route in a social network, then it should be possible for computers.
- Jon Kleinberg explained in 2000 how small world networks can be navigable.

Ian Clarke & Oskar Sandberg - 2005 - p.8/33



Application, cont.

Is Alice closer to Harry than Bob?

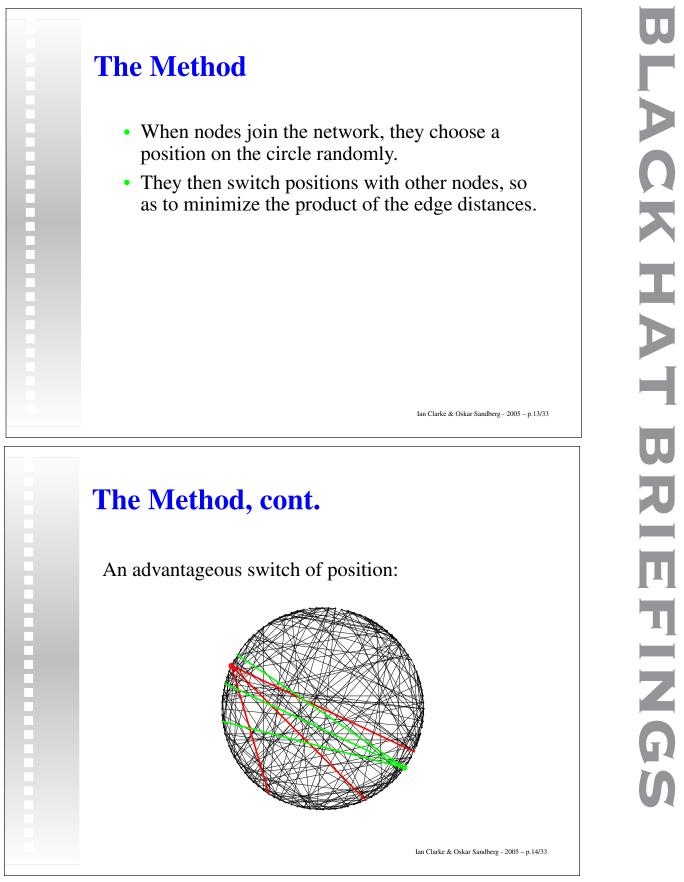
- In real life, people presumably use a large number of factors to decide this. Where do they live? What are their jobs? What are their interests?
- One cannot, in practice, expect a computer to route based on such things.
- Instead, we let the network tell us!

Ian Clarke & Oskar Sandberg - 2005 - p.11/33

Application, cont.

- Kleinberg's model suggests: there should be few long connections, and many short ones.
- We can assign numerical identities placing nodes in a circle, and do it in such a way that this is fulfilled.
- Then greedy route with respect to these numerical identities.

Ian Clarke & Oskar Sandberg - 2005 - p.12/33



The Method, cont.

Some notes:

- Switching is essential!
- Because this is an ongoing process as the network grows (and shrinks) it will be difficult to keep permanent positions.

Ian Clarke & Oskar Sandberg - 2005 - p.15/33

The Algorithm

- Two nodes are chosen in some random fashion, and attempt to switch.
- They calculate ℓ_b as the product of all the lengths of their current connections. Then they calculate ℓ_a as the product of what all their respective connection lengths would be after they switched.
- If \(\ell_b > \ell_a\) they switch. Otherwise they switch with probability \(\ell_b / \ell_a\).

Ian Clarke & Oskar Sandberg - 2005 - p.16/33

The Algorithm, cont.

Let d(z) give the degree (number of connections) of a node z, and let $e_i(z)$ and $e'_i(z)$ be distance of z's i- th connection before and after a switch occurs. Let nodes x and y be the ones attempting to switch. Calculate:

$$p = \frac{\ell(a)}{\ell(b)} = \frac{\prod_{i=1}^{d(x)} e_i(x) \prod_{i=1}^{d(y)} e_i(y)}{\prod_{i=1}^{d(x)} e'_i(x) \prod_{i=1}^{d(y)} e'_i(y)}$$

x and y will complete the switch with probability min(1, p). Otherwise we leave the network as it is.

Ian Clarke & Oskar Sandberg - 2005 - p.17/33

The Algorithm, cont.

- This is an application of the Metropolis- Hastings algorithm.
- Because there is a greater chance of moving to positions with shorter connection distances, it will tend to minimize the product of the distances.
- Because the probability of making a switch is never zero, it cannot get stuck in a bad configuration (a local minima).

Ian Clarke & Oskar Sandberg - 2005 - p.18/33

1gs ces. IAUX HAT

U

2

FING

The Algorithm, cont.

- How do nodes choose each other to attempt to switch?
- Any method will work in theory, but some will work better than others. Only switching with neighbors does not seem to work in practice.
- Our current method is to do a short random walk starting at one of the nodes and terminating at the other.

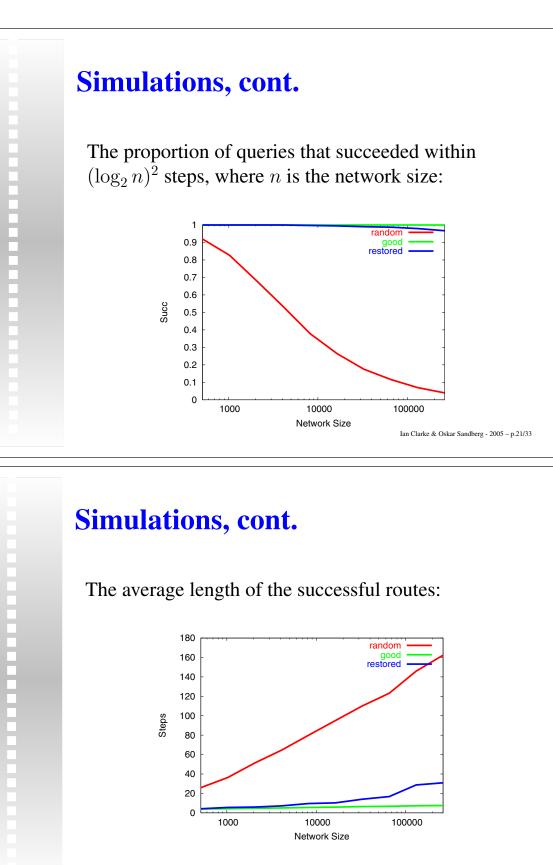
Ian Clarke & Oskar Sandberg - 2005 - p.19/33

Simulations

We have simulated networks in three different modes:

- Random walk search: "random".
- Greedy routing in Kleinberg's model with identities as when it was constructed: "good".
- Greedy routing in Kleinberg's model with identities assigned according to our algorithm (2000 iterations per node): "restored".

Ian Clarke & Oskar Sandberg - 2005 - p.20/33



Results

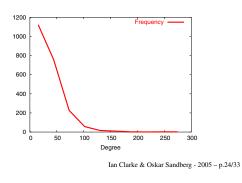
- Simulated networks are only so interesting, what about the real world?
- We borrowed some data from orkut.com. 2196 people were spidered, starting with Ian.



Ian Clarke & Oskar Sandberg - 2005 - p.23/33

Results, cont.

- The set was spidered so as to be comparatively dense (average 36.7 connections per person).
- It contains mostly American techies and programmers. Some are probably in this room. (No Brazilians...)
- The degree distribution is approximately Power-Law:



Results, cont.

Searching the Orkut dataset, for a maximum of $\log_2(n)^2$ steps.

	Success Rate	Mean Steps	
Random Search	0.72	43.85	
Our Algorithm	0.97	7.714	

Ian Clarke & Oskar Sandberg - 2005 - p.25/33

Results

Clipping degree at 40 connections. (24.2 connections per person.)

	Success Rate	Mean Steps	
Random Search	0.51	50.93	
Our Algorithm	0.98	10.90	

Our algorithm takes advantage of there being people who have many connections, but it does not depend on them.

Practical Concerns

- So the theory works, but how does one implement such a network in practice?
- Key concerns:
 - Preventing malicious behaviour
 - Ensuring ease of use
 - Storing data

Ian Clarke & Oskar Sandberg - 2005 - p.27/33

Preventing Malicious Behaviour

Threats:

- Selection of identity to attract certain data
- Manipulation of other node's identities

Ian Clarke & Oskar Sandberg - 2005 - p.28/33

Ensuring ease of use

- Peers will need to be "always on"
- Peer introduction
 - Email
 - Phone

- Trusted third party
- What about NATs and firewalls
 - Could use UDP hole- punching (as used by Dijjer, Skype)
 - Would require third- party for negotiation

Ian Clarke & Oskar Sandberg - 2005 - p.29/33

Storing Data

- We can store data as in a caching Distributed Hash Table (similar to Freenet)
- We can also route directly between two peers if we know their identities
 - Problem: Identities change
- We could employ a "crossing paths" approach
 - Both peers route towards the same random identity
 - When paths cross a connection is established

Conclusion

We believe very strongly that building a navigable, scalable Darknet is possible. *And we intend to do it!*

- There is still much work to do on the theory.
 - Can other models work better?
 - Can we find better selection functions for switching?
 - It needs to be tested on more data.

Ian Clarke & Oskar Sandberg - 2005 - p.31/33

Conclusion, cont.

- We have learned the hard way that practice is more difficult than theory.
 - Security issues are very important.
 - How the network is deployed will affect how well it works.

People who are interested can join the discussion at *http://freenetproject.org/*.

Ian Clarke & Oskar Sandberg - 2005 - p.32/33

