



# Pass it on: Social Networks Stymie Censors

(Unblocking the Internet: Social networks foil censors)

Yair Sovran, Alana Libonati, Jinyang Li

The International workshop on Peer-To-Peer Systems (IPTPS) 2008



## Jinyang Li 李金扬

He got a B.S. in computer science from National University of Singapore (1998). Undergraduate adviser is Y.C. Tay.

He got a Ph.D. from MIT (2005), working under the guidance of Robert Morris and Frans Kaashoek as a lucky PDOS member.

From 2005-2006, He was a postdoc at Berkeley with Scott Shenker.



### Safari 打不开页面 (?)

Safari 无法打开页面“[www.google.com.hk/url?sa=p&hl=zh-CN&pref=hkredirect&pval=yes&q=http://www.google.com.hk/&ust=1386124230425069&usg=AFQjCNE88yEawS1gVHcOnkTIh4ltX1DOVg](http://www.google.com.hk/url?sa=p&hl=zh-CN&pref=hkredirect&pval=yes&q=http://www.google.com.hk/&ust=1386124230425069&usg=AFQjCNE88yEawS1gVHcOnkTIh4ltX1DOVg)”，因为服务器意外中断了连接。这种情况在服务器忙时有时会出现。请等待几分钟，然后再试一次。

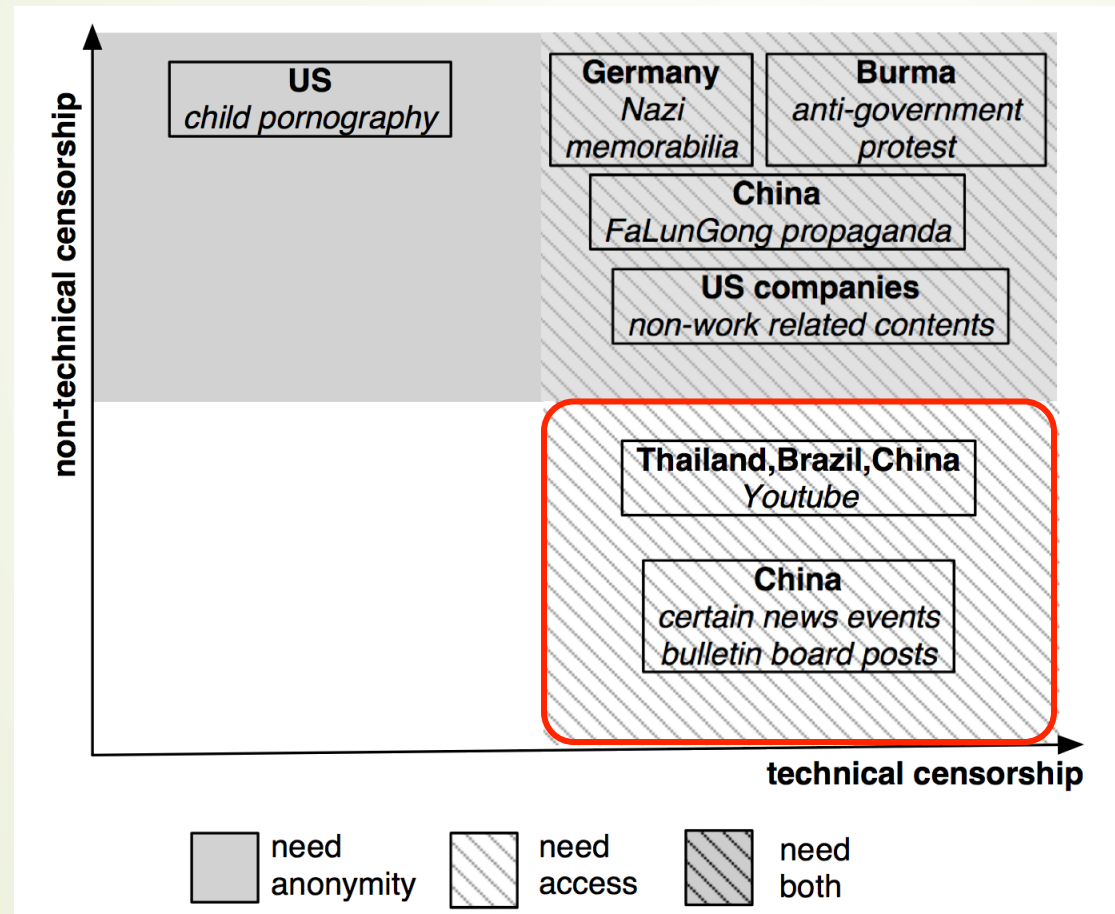
# Internet censorship

**Internet censorship** is the control or suppression of what can be accessed, published, or viewed on the Internet.

--- From Wikipedia

## ➤ Non-Technical

- Threat of jail terms
- Violence



## ➤ Technical

- IP address blocking
- Deep packet inspection
- DNS poisoning
- TCP resetting





# Problem : Collateral Damage

Because technical censorship is never **precise**, many blocked contents are simply the result of collateral damage as in the case of the blocking of Google, YouTube, and Flickr.

Such as : 胡萝卜 摄政王 美国军队



Censorship

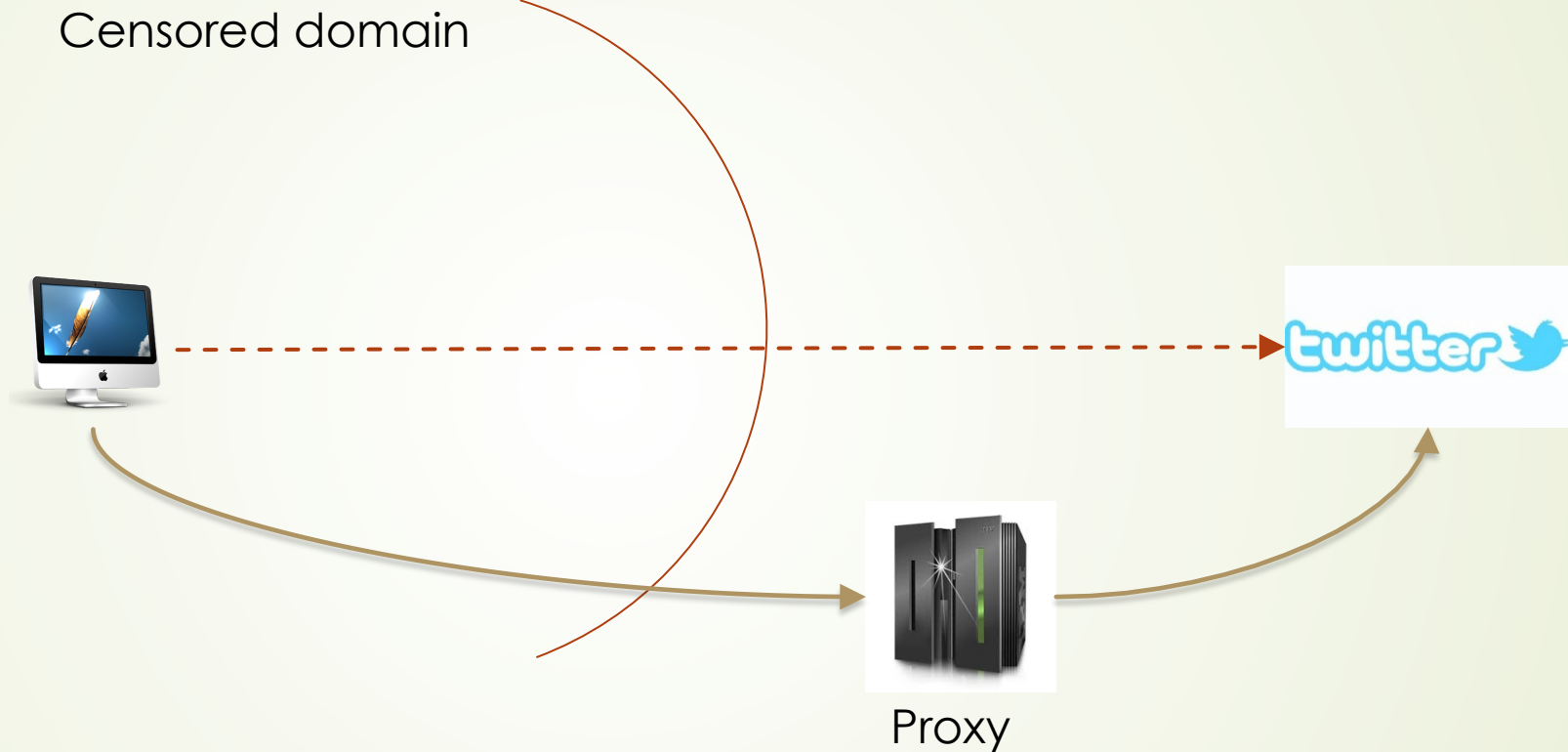


**VS**

Censorship-resistant



# Censorship Circumvention System



Although direct communication is blocked, users can access banned sites via unblocked proxies located outside of the censor's reach.



# Requirement:

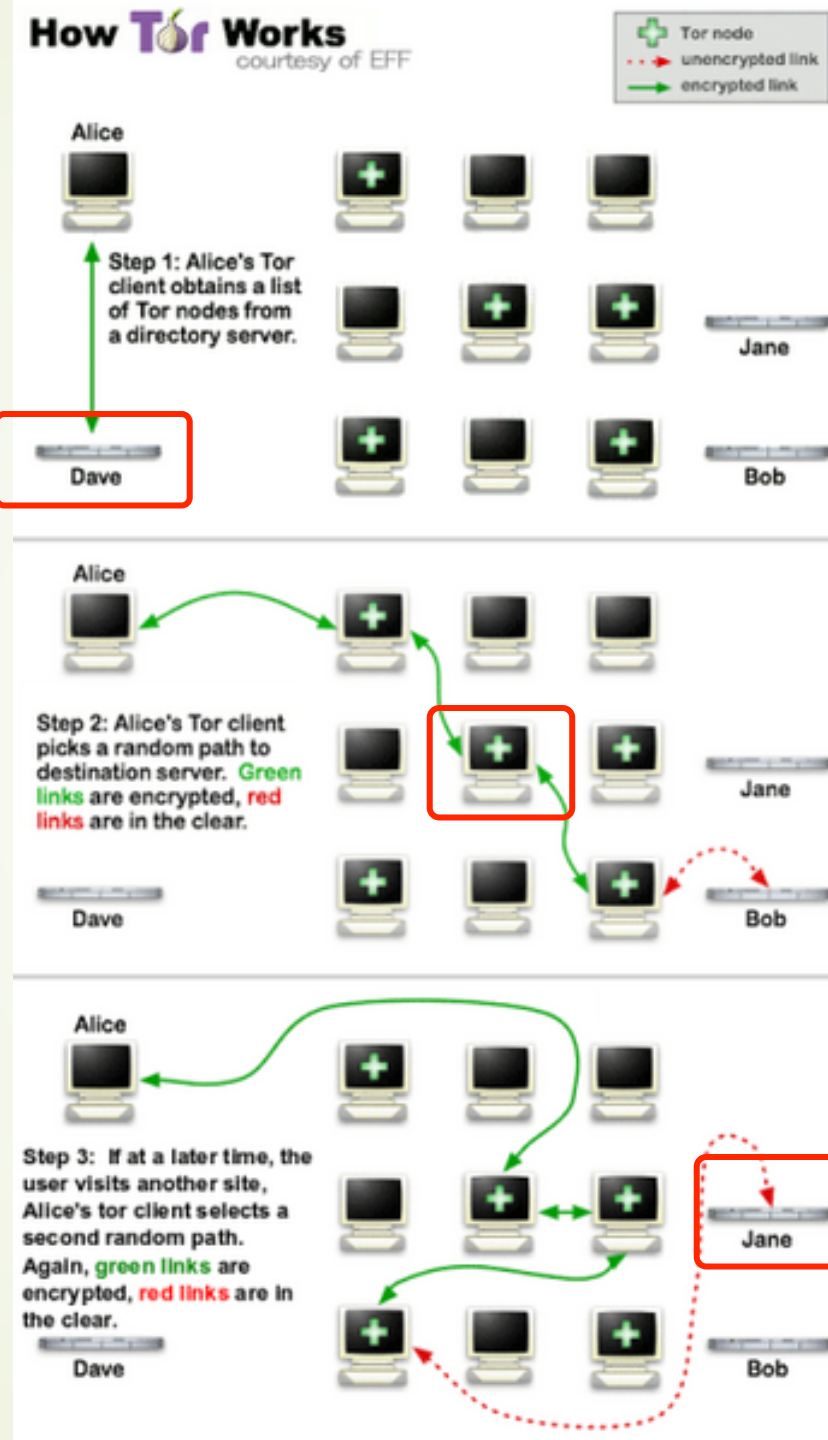
1. To avoid content-based blocking, the connection to the proxy must be encrypted.
2. Using a DNS server outside of the censored domain avoids the problem of DNS poisoning.
3. Preventing the censor from simply adding the proxies' addresses to the list of blocked sites.

Methods ?





# Sybil attacks



1. The censor can easily disable any centralized component in a circumvention system by blocking either the IP address or DNS name of the centralized component.

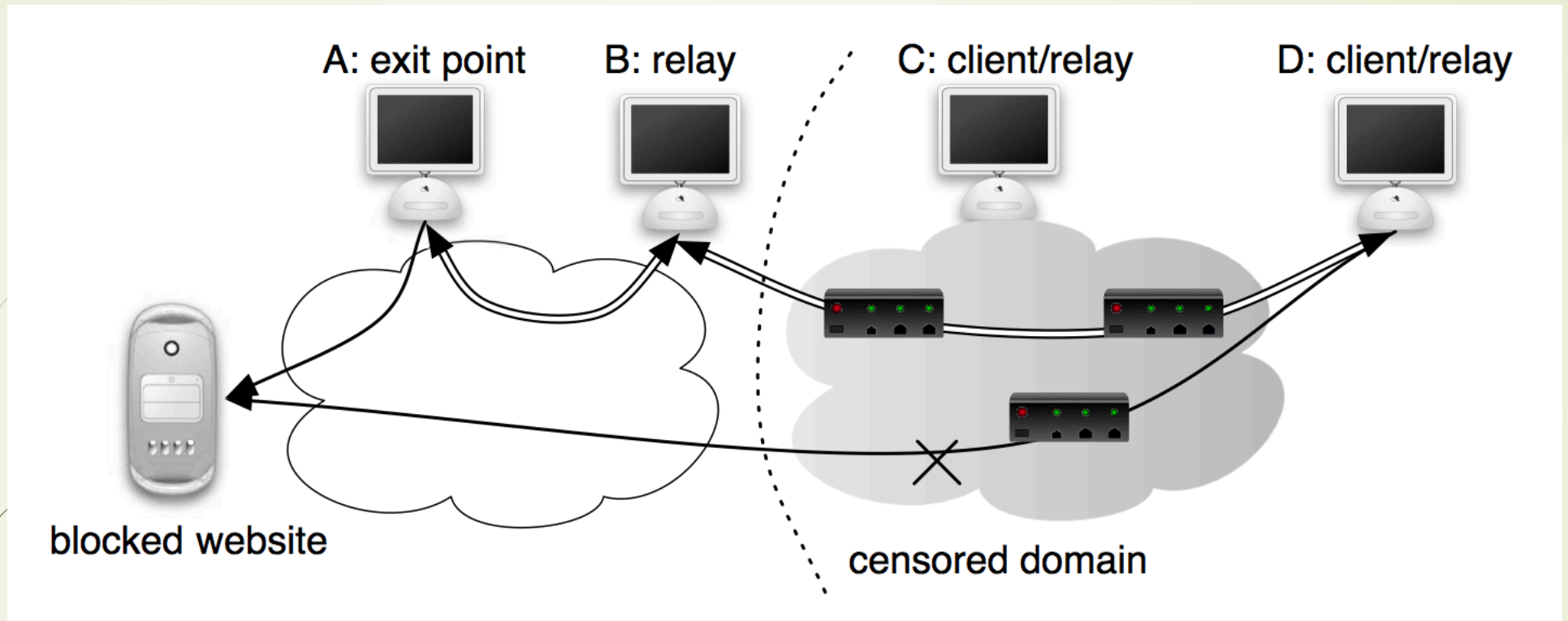
2. The censor can pose as a legitimate user in order to discover as many relays located outside the censored domain as possible and to block them.

3. The censor could also pose as a legitimate exit point or web site to attract traffic from many unsuspecting relays or users in order to track them.



# Kaleidoscope

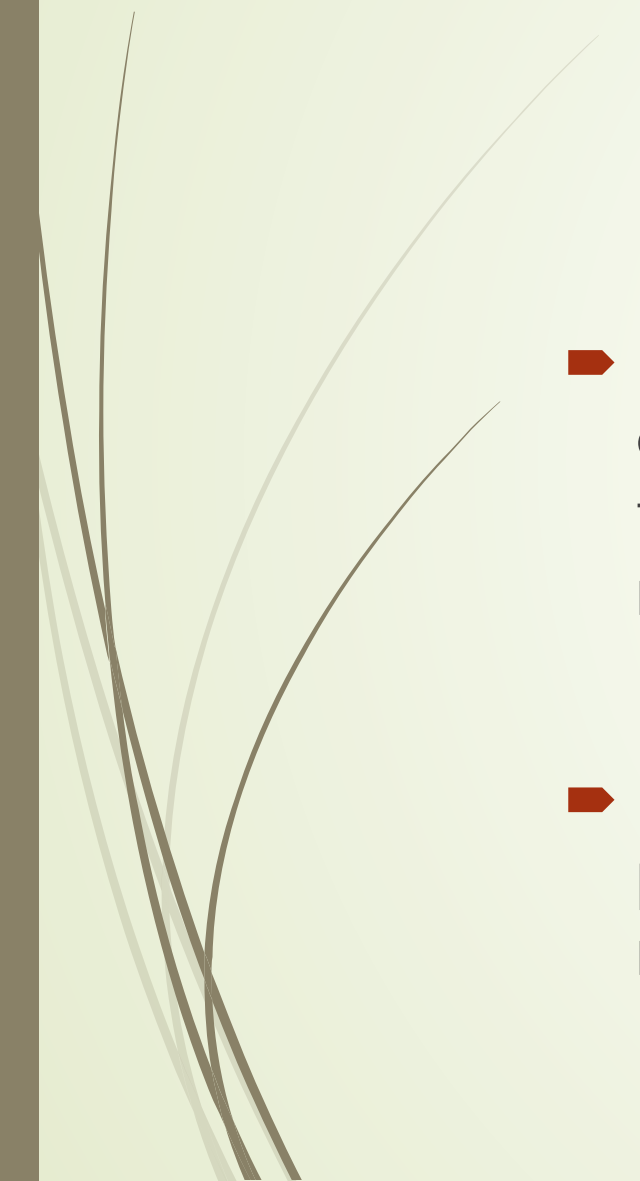
- **Decentralization**. A fully decentralized system foils any attempt to bring down the system by blocking its centralized component.
- **Limited relay exposure**. The system must partition knowledge of all relays (including exit points) among users.
- **Limited user exposure**. The system must restrict each relay or exit point to serving (hence, learning) a small number of relays or users.



- **End-Users**
- **Relay Nodes** : All nodes participating in the Kaleidoscope trust network by default act as relay nodes.
- **Exit Points** : A subset of relay nodes outside the censored domain volunteers to act as exit points (like proxies).



# Kaleidoscope relies on a trust network


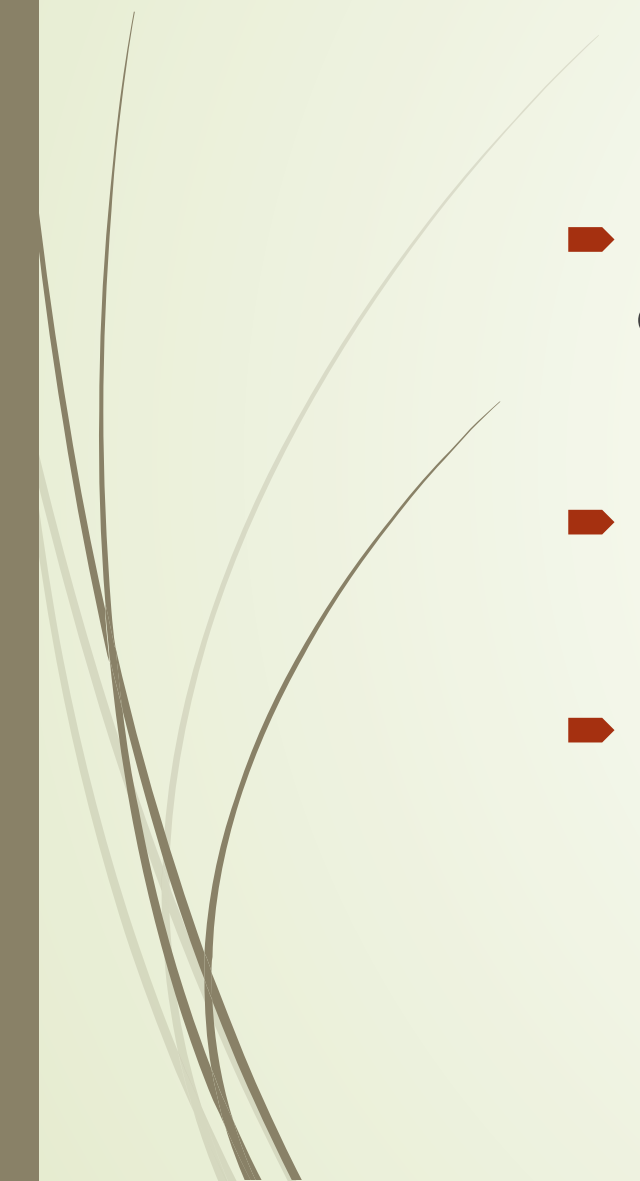
- It have a *sizeable* user population spread both within and across the censored domain; if this number is small, the censor can easily block all communication to these nodes upon detecting them.
  - Relay nodes forward traffic between end-users and exit points and these routing paths in the peer-to-peer relay network *need not involve the original trust links.*
- 



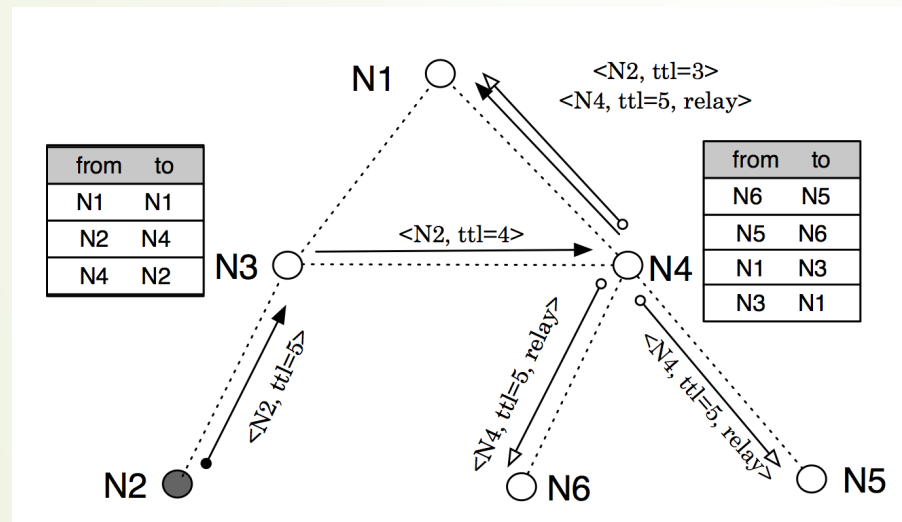
# Restricted service discovery

- Kaleidoscope disseminates the identities of relays and exit points along links *in the trust network*.
  - This requires users to establish trust links with care: subverted trust links cause a node to lose its known relays and exit points.
  - Social links also provide the important benefit of allowing users to utilize offline communications to bootstrap the system in a completely decentralized fashion.
- Each relay or exit point in Kaleidoscope advertises its address to a small number of other nodes along a series of trust links by performing a few short random walks beyond its immediate neighborhood.
- Multi-hop traffic forwarding allows each exit point to serve more users via a few intermediary relays without directly revealing its address to them.



- 
- 
- How it leverages the trust network to disseminate relays' addresses ?
  - How a node finds a multi-hop path to an exit point ?
  - How to set various parameters in Kaleidoscope ?

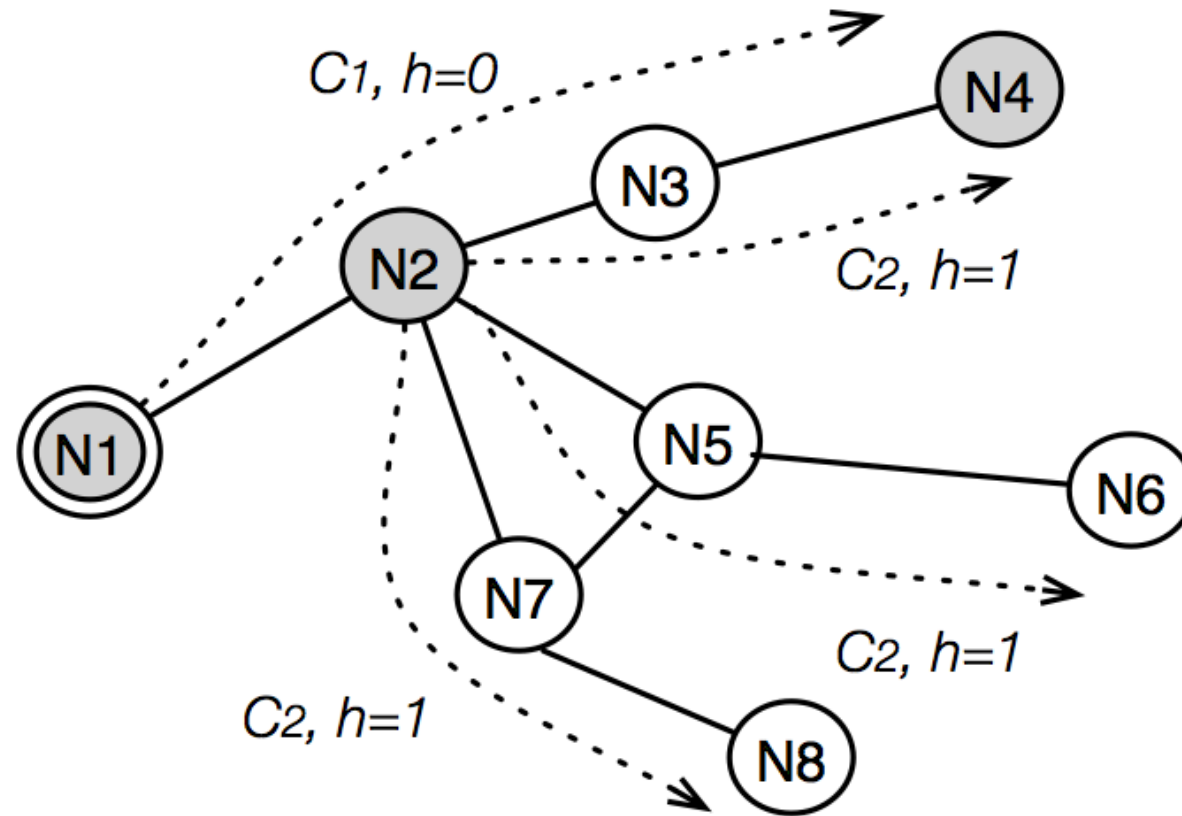
# Limited relay advertisement



IP address, port number  
and **an access cookie**

- Each relay and exit point should advertise its address to  **$r$**  nodes (*the target reach*) over the trust network, where  **$r$**  is larger than a node's immediate trust neighborhood.
- A relay with degree  **$d$**  can perform  **$d$**  random walks of length  **$w = r/d$**  via each neighbor.
- Random walks should be made repeatable.

# Traffic forwarding




# Parameter Setting

## Assumes

- Adversarial nodes collect relay addresses by receiving their advertisements.
- Done for a regular graph where each node has exactly  $d$  neighbors.
- A relay's advertisements reach  $r$  random nodes

Parameter	Meaning	Default Value
$r$	Targeted number of nodes a relay's advertisements should reach.	100
$h_{max}$	Maximum traffic forwarding hops.	3-4
$w_{max}$	Maximum allowed random route length.	20
$w_{min}$	Minimum route length used by a relay.	7

Network	Direct connection to an exit point	One hop to an exit point	Two hops to an exit point
YouTube	0.14	0.42	0.44
Flickr	0.15	0.41	0.44
LiveJournal	0.24	0.39	0.37
Synthetic	0.15	0.56	0.29



1. The target reach of a relay ( $r$ ) is dependent on  $f$ , the level of infiltration to the trust network denoting the fraction of all attack edges between the censor and honest users.

$$f = 0.2\%$$

2. In a network where a fraction  $f$  of links are attack edges,  $f/2$  fraction of all nodes are colluding with the censor.

3. The probability that none of  $r$  recipients is adversarial

$$(1 - f/2)^r$$

4. The expected number of advertisements from an unexposed relay

$$EX = r(1 - f/2)^r$$

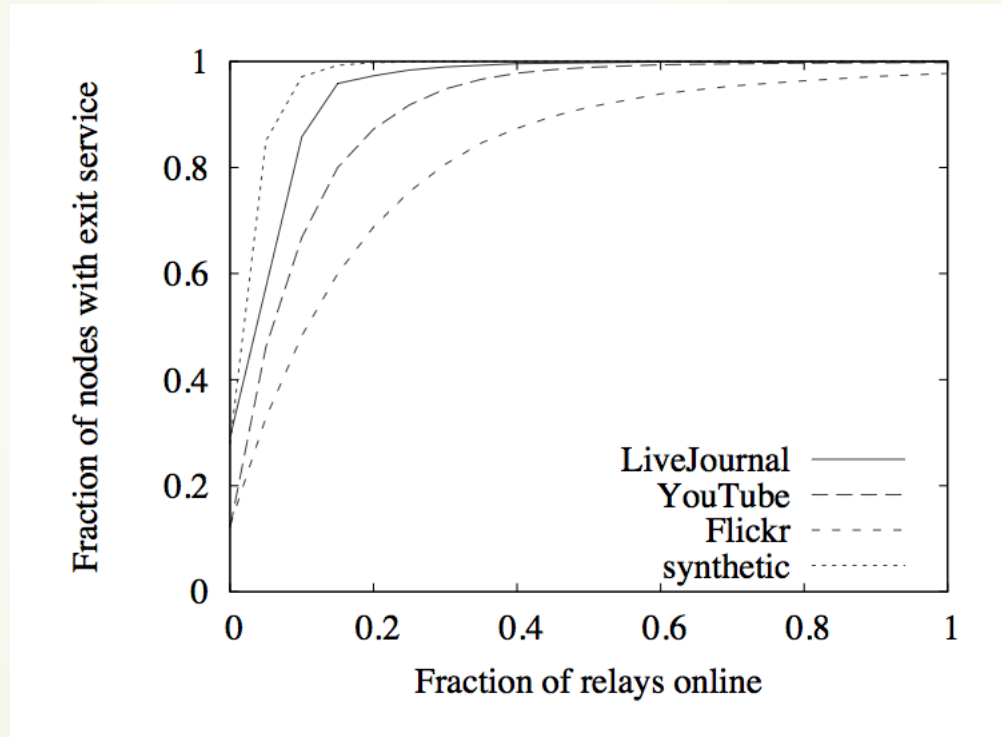
5. The optimal  $r$

$$r_{\downarrow opt} = -1 / \ln(1 - f/2) \approx 2/f = 100$$

$$W_{\downarrow max} = r/d = 100/5 = 20$$



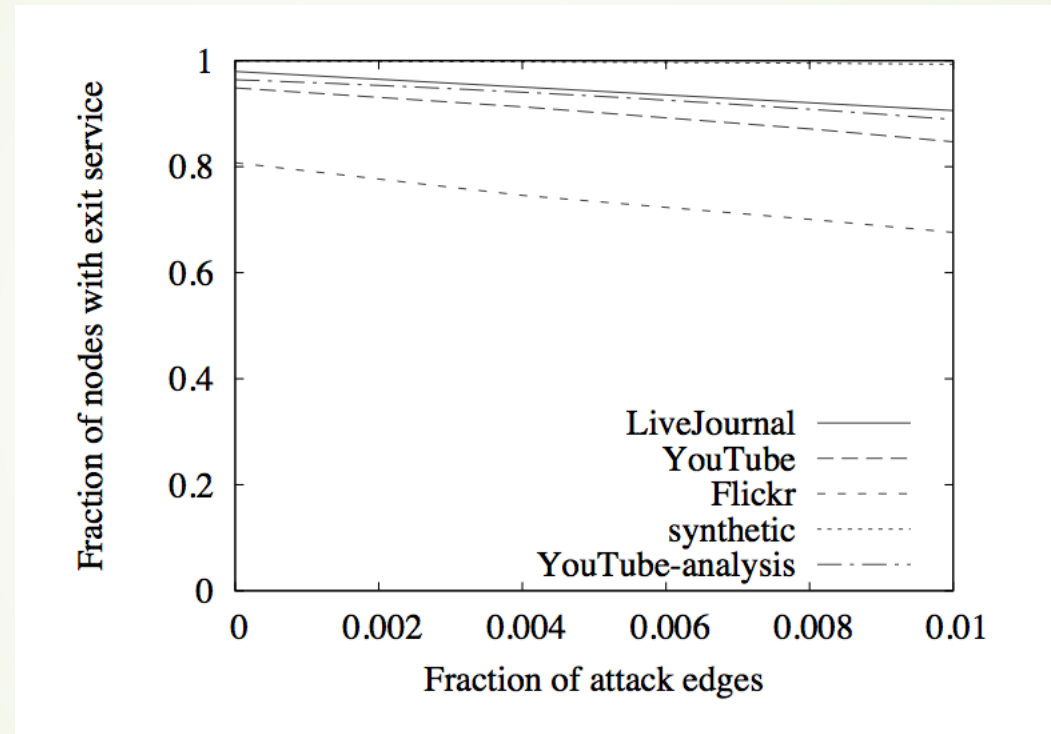
# Evaluation



When more than 40% relays are online, the service coverage is very high (~98%) for all networks except Flickr.

# Evaluation

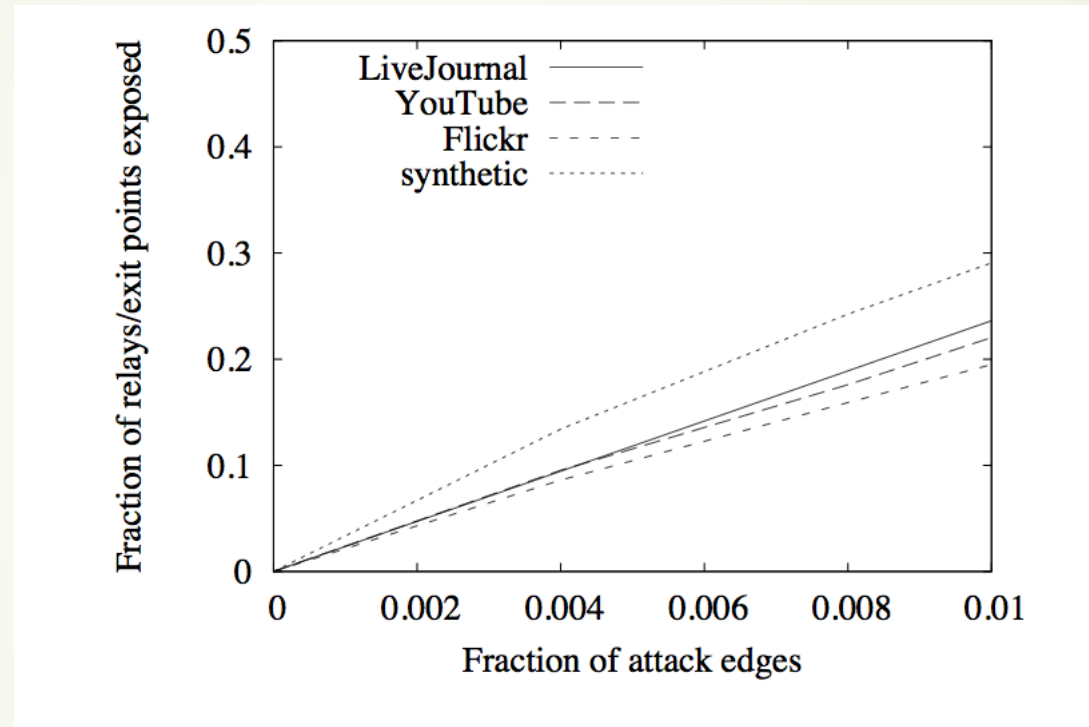
passive adversarial nodes



when the censor manages to attach as many as 0.5% attack edges, more than 90% of users can still find paths of unexposed relays and exit points.

# Evaluation

passive adversarial nodes



20 – 30% relays and exit points are exposed when the fraction of attack edges is 1% for all networks.

# Evaluation

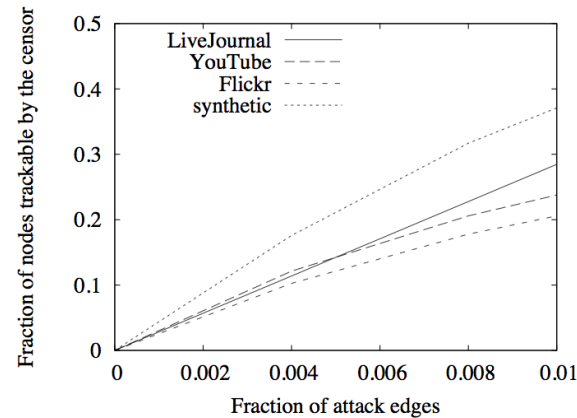


Figure 9: (Left) The number of clients that could potentially request service from a decoy exit point is bounded by the number of attack edges and  $w_{max}$ .

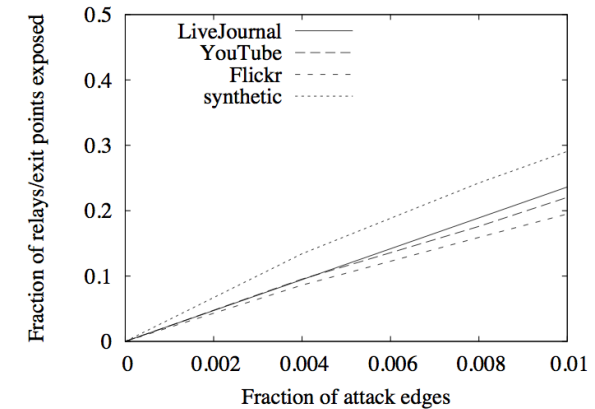


Figure 10: (Right) When new nodes join the network and activate trust links, the censor can learn of more relays via each of its attack edge as a result of the routing table changes at honest nodes. However, the rate of increase in the censor's knowledge is slow. (0.5% attack edges).

Actively participate as decoy relays or exit points



# Thanks

Any Question?