Chipping Away at Censorship Firewalls with User-Generated Content

S. Burnett, N. Feamster, S. Vempala
Web Censorship

Various governments around the world restrict internet access:
- Blocking activist websites
- Filtering content sites
- Monitoring web traffic

Image borrowed from Georgia Tech Slides
Collage

- Design Considerations:
  - Robustness: resistant to censorship efforts
  - Deniable: difficult to identify users
  - No dedicated infrastructure: utilizes existing infrastructure

- Aims to be an effective alternative to less effective methods of avoiding censorship
Collage versus Alternatives

• Proxies:
  – Proxies like Tor can be blocked if the proxy list is public
  – Not deniable if encryption is incriminating
  – Requires a dedicated infrastructure in the form of a proxy network
Collage versus Alternatives

- Covert Channels:
  - Not completely robust against blocking
  - Most stegonography algorithms have been broken; can never be fully secure
  - Requires dedicated infrastructure in the form of web servers
Layers of Collage

Sec. 7
Application Layer

Message Data

Application

send

receive

Block Block Block Block

encode

Vector

decode

Sec. 4.2
Message Layer

Sec. 4.1
Vector Layer
Layers of Collage

- Vector Layer:
  - Storage for data chunks
  - The medium for embedding a message

- Message Layer:
  - Specifies protocol for vector layer to send and receive messages

- Application Layer:
  - Applications built on the collage channel
The Collage Process: Send

- Obtain a message
- Pick a message identifier
  - Application specific
  - Similar to encryption key
- Obtain vector media
- Embed message into vectors
- Upload user generated content (UGC) to host
  - flickr, twitter, youtube, domestic equivalents
The Collage Process: Receive

- Find and fetch vectors from UGC host
- Decode message from UGC if it contains data encoded for given identifier
Tasks

- Collage uses tasks to create rendezvous between senders and receivers
- Corresponding HTTP requests
  - Sender task: upload photos tagged with flowers to flicker
  - Receiver task: search for photos tagged with flowers on flickr
Meeting Design Goals

• Robust
  – Erasure Coding
  – Uses too many content hosts to be completely blocked

• Deniable
  – Traffic only to/from content hosts
  – Depends on tasks

• No dedicated infrastructure
  – Messages stored on content hosts
Performance Evaluation

- Significant overhead in the Collage process
- Production stegonography tools achieve encoding rates from .01 - .05
  - 20-100 factor increases in storage
- Tasks can impose additional traffic and time
  - Finding UGC with encoded data can possibly represent an overhead of several hundred percent
Performance Evaluation

- Graph based on tests of 10 tasks
- Logarithmic Y-axis
  - Higher redundancy comes at the cost of storage required increasing extremely quickly past a certain point
  - 10.0x send rate means data can be retrieved at 90% vector block rate
- At 1.0 vector encoding efficiency
  - Storage required for 1.1x, 1.5x, 2.0x ranges from 50-110 KB
- High slope at low efficiency
  - Possibility of significant improvement with incremental advancement of encoding
Performance Evaluation

- Based on 1 MB vector storage
- Slope increases with higher block rate
  - Larger increase from 60%-80% than other increments
- Linear increase in traffic
  - More vectors blocked means more tasks needed to get data
Performance Evaluation

- Overall transfer time given various overheads
- Total Transfer time – Overhead is a constant value dependent on network speed
  - 768/384 Kbps: 100 seconds
  - 6000/1000 Kbps: 50 seconds
  - 768/10000 Kbps: 25 seconds

Paper does not give values for simulated message size
- Effects of size on time overhead unknown