Climbing China's Great Firewall

Adam Casey

Department of Computer Science
University of Minnesota Morris
UMM CSci Senior Seminar Conference

April 15th 2018

Overview

- People inside of China cannot access popular sites such as Facebook
- In most cases Chinese alternatives to popular websites exist
- Tools are being developed and updated by citizens to navigate around censorship
- At the same time the Chinese government is developing more advanced censorship tools

Outline

Outline

Introduction

- Citizens in countries such as Syria, Iraq, Iran and China experience government internet censorship
- 1.3 Billion people live in China
- China's internet censorship mechanism referred to as the Great Firewall of China (GFW)

Outline

Outline

Background

The TCP Protocol

- What is a DNS request?
- What is a TCP packet?
- Three-Way Handshake
- Connection Termination
- TCP Control Block (TCB)

What is a DNS Request

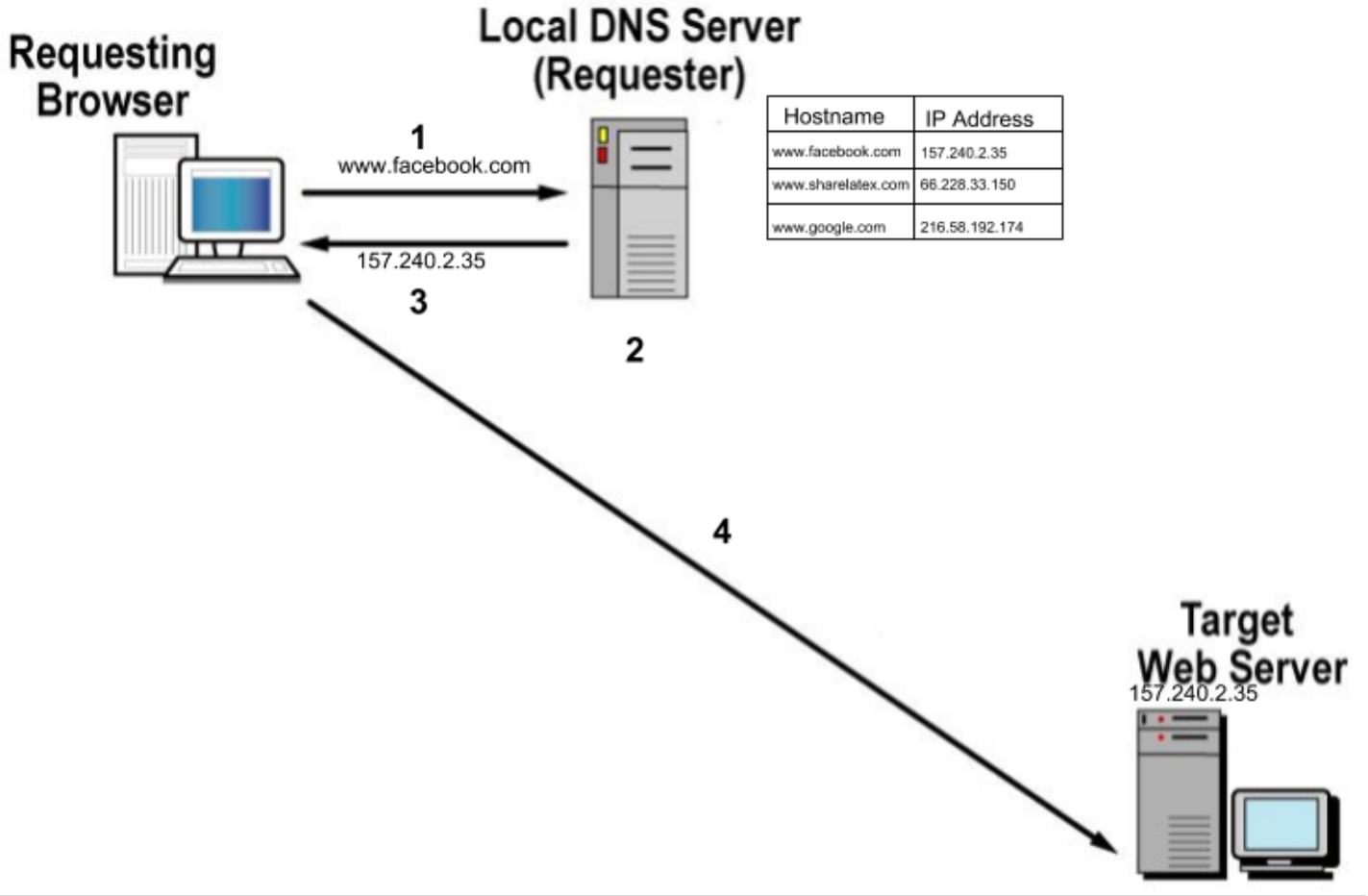


Figure: Simplified diagram of a DNS request taken from [Ric]

What is a TCP Packet?

- Data broken up into discrete parts called packets
- Each packet has a header, the data payload, and sometimes a trailer for error correction
- Header indicates type of packet, what port it's heading to and other data
- Each packet has a time to live or TTL

What is a TCP Header?

TCP Header

Offsets	Octet	0								1								2								3							
Octet	Bit	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	0	Source port																Destination port															
4	32	Sequence number																															
8	64	Acknowledgment number (if ACK set)																															
12	96	Data offset				Reserved 0 0 0			N S	C W R	E C E	U R G	A C K	P S H	R S T	S Y N	F I N	Window Size															
16	128	Checksum																Urgent pointer (if URG set)															
20 ...	160 ...	Options (if *data offset* > 5. Padded at the end with "0" bytes if necessary.) ...																															

Figure: Diagram of a TCP Header taken from [unk18]

What is a TCP Header?

TCP Header

Offsets	Octet	0								1								2								3							
Octet	Bit	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	0	Source port																Destination port															
4	32	Sequence number																															
8	64	Acknowledgment number (if ACK set)																															
12	96	Data offset				Reserved 0 0 0				N S	C W R	E C E	U R G	A C K	P S H	R S T	S Y N	F I N	Window Size														
16	128	Checksum																Urgent pointer (if URG set)															
20 ...	160 ...	Options (if *data offset* > 5. Padded at the end with "0" bytes if necessary.) ...																															

Figure: Diagram of a TCP Header taken from [unk18]

What is a TCP Header?

Bit	107	109	110	111
Flag	ACK	RST	SYN	FIN

Figure: Close-up of the relevant flags

Three Way Handshake

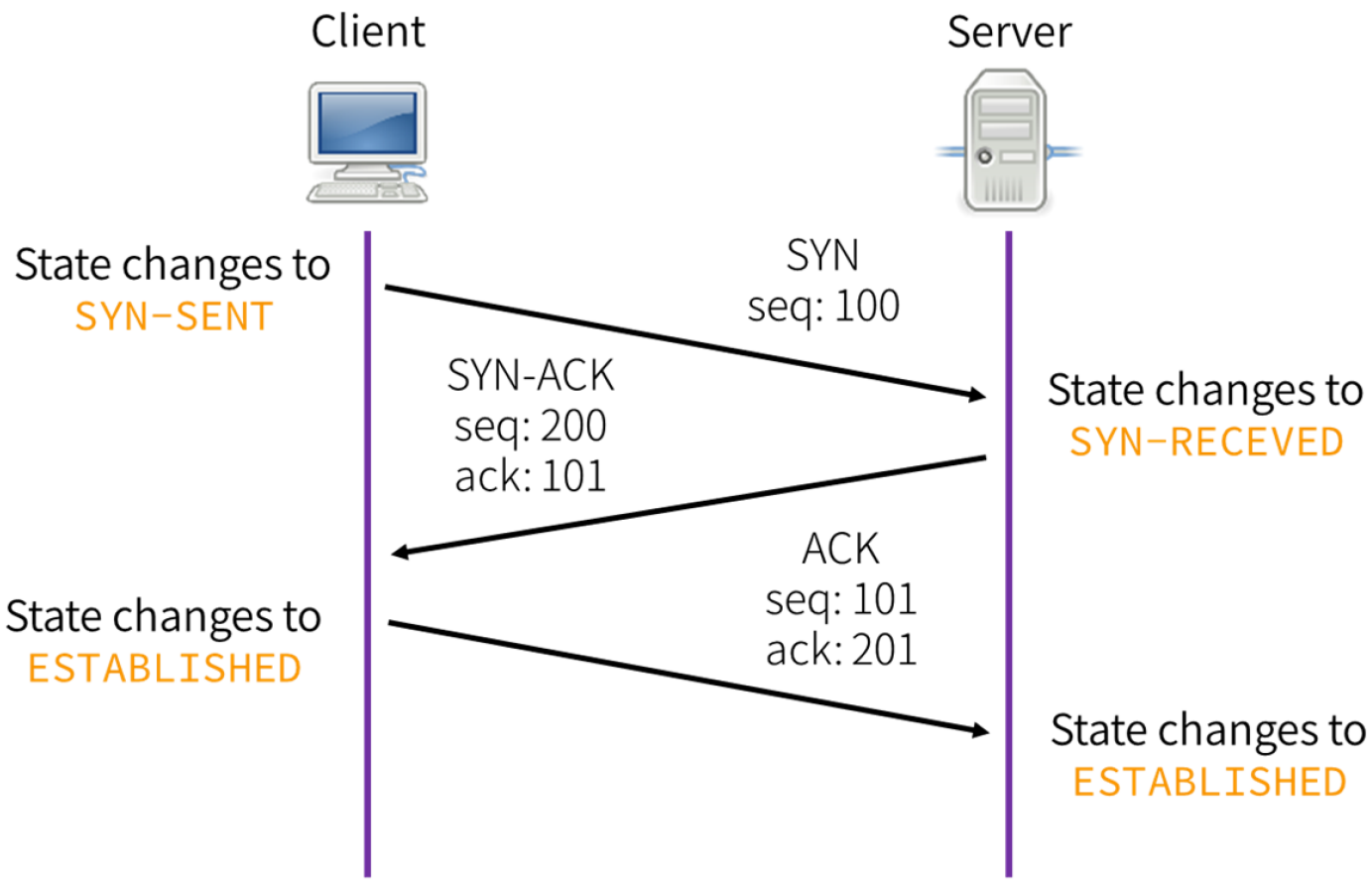


Figure: TCP Three Way Handshake taken from [FHHC16]

Connection Termination

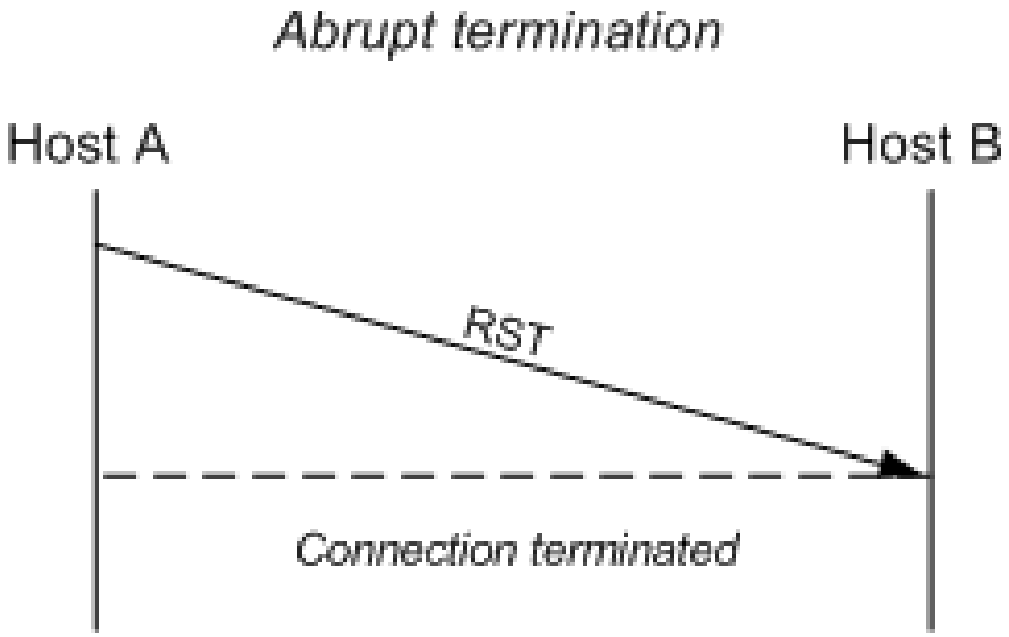


Figure: Diagram of TCP connection termination taken from [Unkb]

TCP Control Block

- Data structure created by the TCP protocol
- Keeps track of multiple connections outgoing and incoming
- TCB control block on GFW used in combination with packet inspection to terminate connections with sensitive keywords

Outline

Background

CDNs

- Content Delivery Network
- Run by third party companies

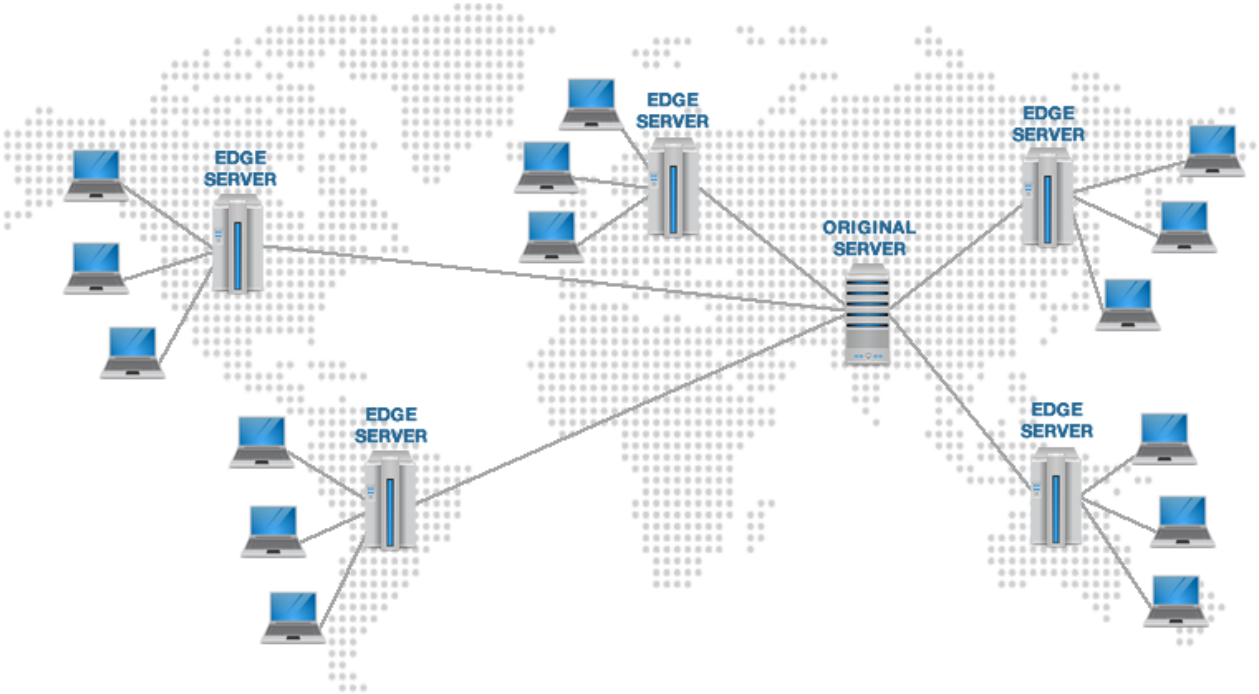


Figure: CDN layout taken from [unka]

Outline

Background

Tor

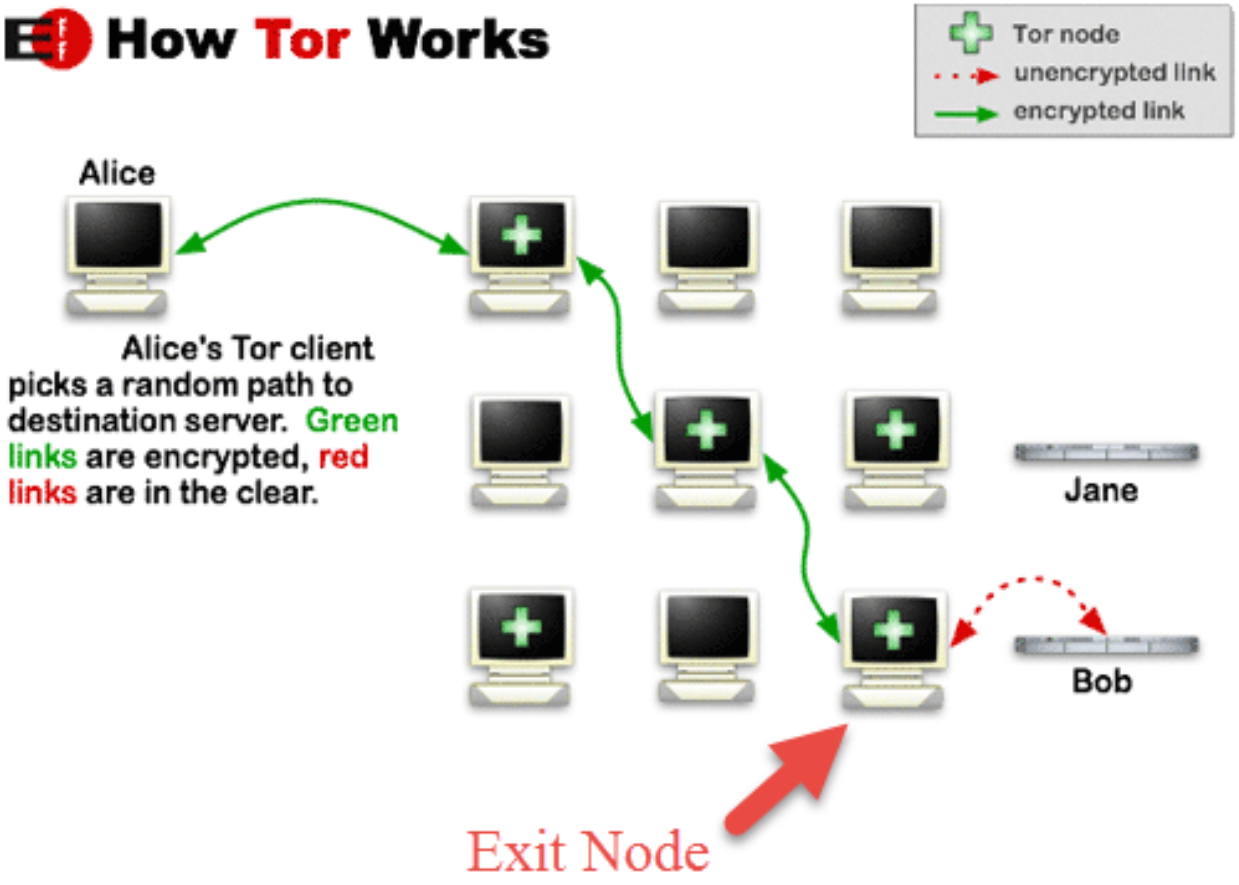


Figure: Diagram of Tor nodes take from [Des16]

Outline

Outline

INTANG

Strategies

- Tool developed by Wang et al.
- Packet manipulation
- False TCB creation
- TCB teardown
- Data reassembly

INTANG

Strategies

False TCB creation

- Send SYN insertion packet with modified sequence number
- Packet has low TTL and/or wrong checksum and will not be accepted by server
- Initiate connection with correct sequence number
- Traffic will be ignored by GFW due to unexpected sequence number
- Each packet is given a default Time to live (TTL)

TCB Teardown

- Uses the same idea as false TCB creation to create packets that are rejected by server
- Packet has low TTL and/or wrong checksum and will not be accepted by server
- TCB on GFW will be torn down when it recieves RST, RST/ACK or FIN packet
- Connection to server kept alive

Outline

INTANG

Results

- Do INTANG's strategies actually work?
- 77 websites
- 50 trials each

Vantage Points	Strategy	Success		
		Min	Max	Avg.
Inside China	Improved TCB Teardown	89.2%	98.2%	**95.8%**
	Improved In-order Data Overlapping	86.7%	97.1%	**94.5%**
	TCB Creation + Resync/Desync	88.5%	98.1%	**95.6%**
	TCB Teardown + TCB Reversal	90.2%	98.2%	**96.2%**
	INTANG Performance	93.7%	100.0%	**98.3%**

Figure: Packet manipulation strategy success rates taken from [WCQ+17]

Outline

Outline

Cachebrowser

Overview

- Tool developed by John Holowczak and Amir Houmansadr
- Browses through CDNs for cached content
- Gets around IP address filtering

Cachebrowser

CDNs

- Multiple websites at one IP
- IPs change very frequently (sometimes as frequently as once a minute)
- One website's content is on multiple different edge servers to ensure quick access

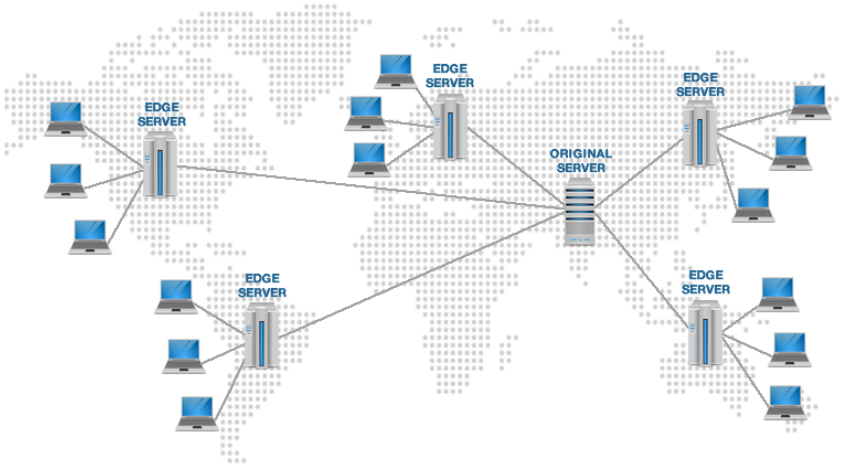


Figure: CDN layout taken from [unka]

Cachebrowser

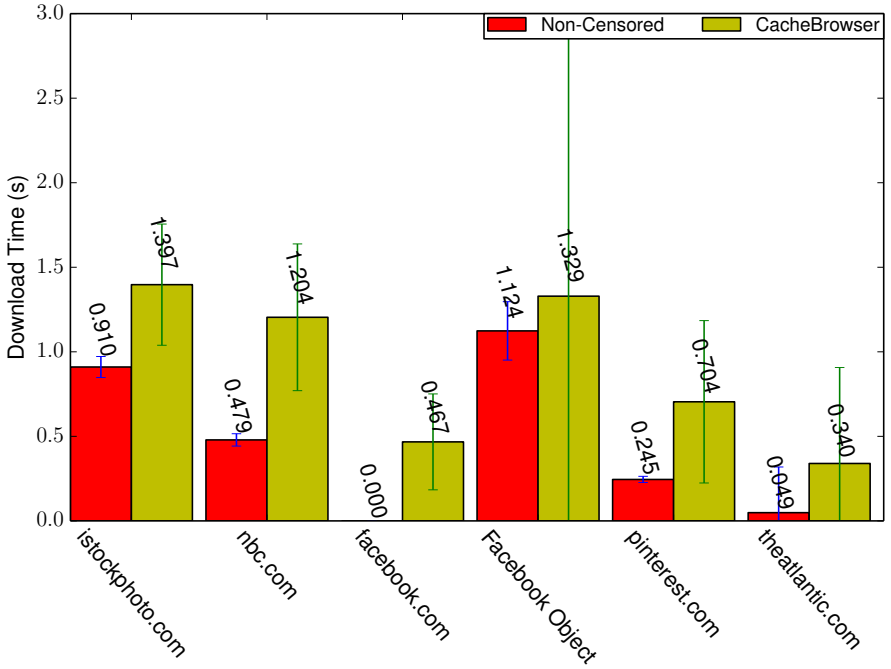
Strategies

- Keeps and internal database of CDN hosted alternatives to websites
- Makes requests to free unblocked DNS resolver website
- If request to DNS resolver fails makes request to remote server using SWEET

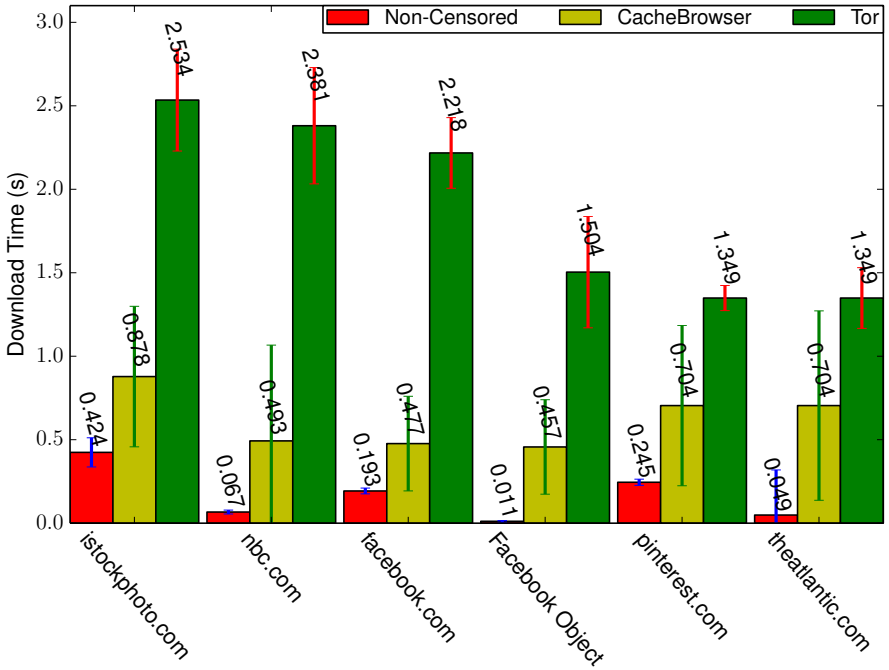
Outline

Cachebrowser

Results



(a) Client in China



(b) Client in Amherst, MA, U.S.

Figure: Graph of Cachebrowser latency versus alternative methods taken from [HH15]

Outline

Conclusions

- All discussed methods work so what is best?
- Tor works in a way that makes viable in the long term, but it is comparatively slow
- INTANG works well for now but the GFW could be modified
- Cachebrowser works only for content hosted on a CDN

Conclusions

- All discussed methods work so what is best?
- Tor works in a way that makes viable in the long term, but it is comparatively slow
- INTANG works well for now but the GFW could be modified. Does not avoid IP address filtering.
- Cachebrowser works only for content hosted on a CDN

Questions?

References I

- Nandan Desai, *Unclosing the dark web - post 2*, jun 2016.
- Cheng-Yu Tsai Wei-Tai Cai Chia-Hao Lee Fu-Hau Hsu, Yan-Ling Hwang and KaiWei Chang, *Trap: A three-way handshake server for tcp connection establishment*, Nov 2016.
- John Holowczak and Amir Houmansadr, *Cachebrowser: Bypassing chinese censorship without proxies using cached content*, Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security (New York, NY, USA), CCS '15, ACM, 2015, pp. 70–83.
- Robert Rich, *Anatomy of a web request*.
- unknown, *How content delivery networks work*.
- Unknown, *Network programming in linux*.

References II

unknown, *Transmission control protocol*, Mar 2018.

Zhongjie Wang, Yue Cao, Zhiyun Qian, Chengyu Song, and Srikanth V. Krishnamurthy, *Your state is not mine: A closer look at evading stateful internet censorship*, Proceedings of the 2017 Internet Measurement Conference (New York, NY, USA), IMC '17, ACM, 2017, pp. 114–127.