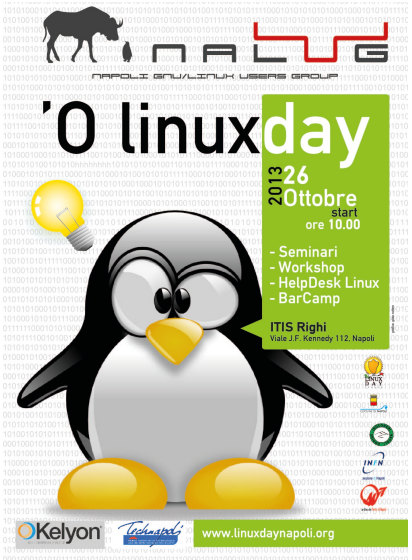


Monitoring Internet Censorship

part 2

Linux Day 2013

Napoli, October 26 2013

Giuseppe Aceto, giuseppe.aceto@unina.it

Premise – quick check

- We are going to see some “mechanics” of the Internet to explain **how censorship works**
- The description will avoid technicisms when possible
- No deep knowledge of computer networks will be implied but possibly a little introduction to “the Internet” basics

SO

  - How many of you knows what the DNS is?
  - … and IP?
  - … and HTTP?
  - … IDS?

Premise – quick check

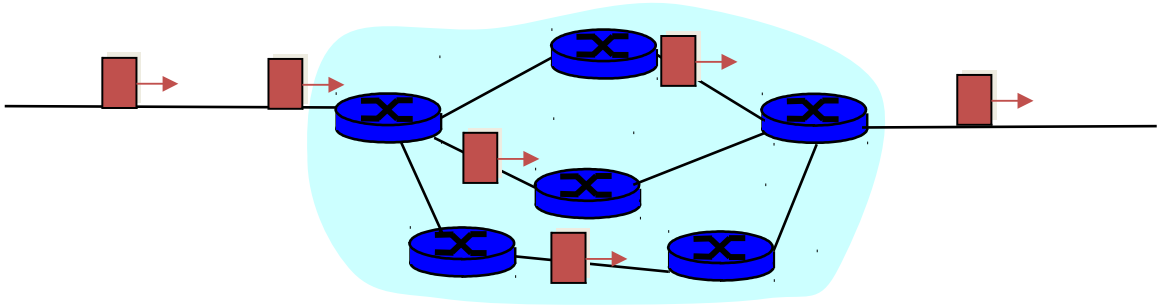
- We are going to see some “mechanics” of the Internet to explain **how censorship works**
- The description will avoid technicisms when possible
- No deep knowledge of computer networks will be implied but possibly a little introduction to “the Internet” basics

SO

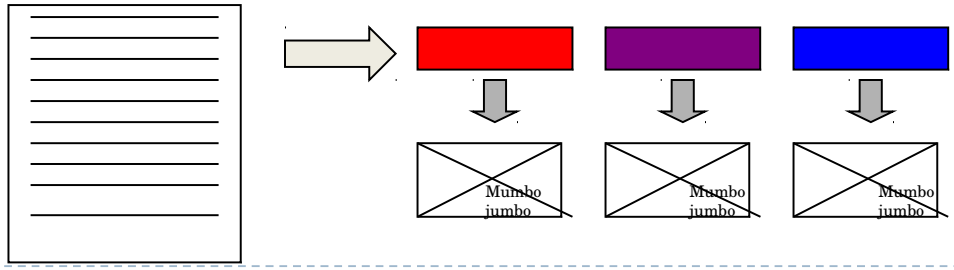
  - How many of you knows what the DNS is?
  - … and IP?
  - … and HTTP?
  - … IDS?

Basics about Internet: packets

- Information is divided in packets
- Packets are forwarded along a path *SENDER-to-RECEIVER* by several (usually<30) intermediate devices “routers”

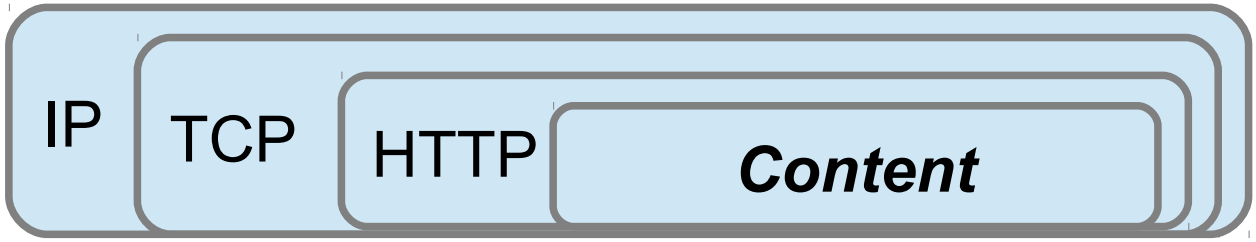


- Packets are “labeled” with SENDER and RECEIVER addresses (sort of *phone numbers*, called “IP addresses”)
  - plus other details useful for the service



Basics about Internet: protocols

- The rules to write, split, label, forward, reassemble, read the messages are called **protocols**
  - Each one has its own business, some work together: often are wrapped in layers

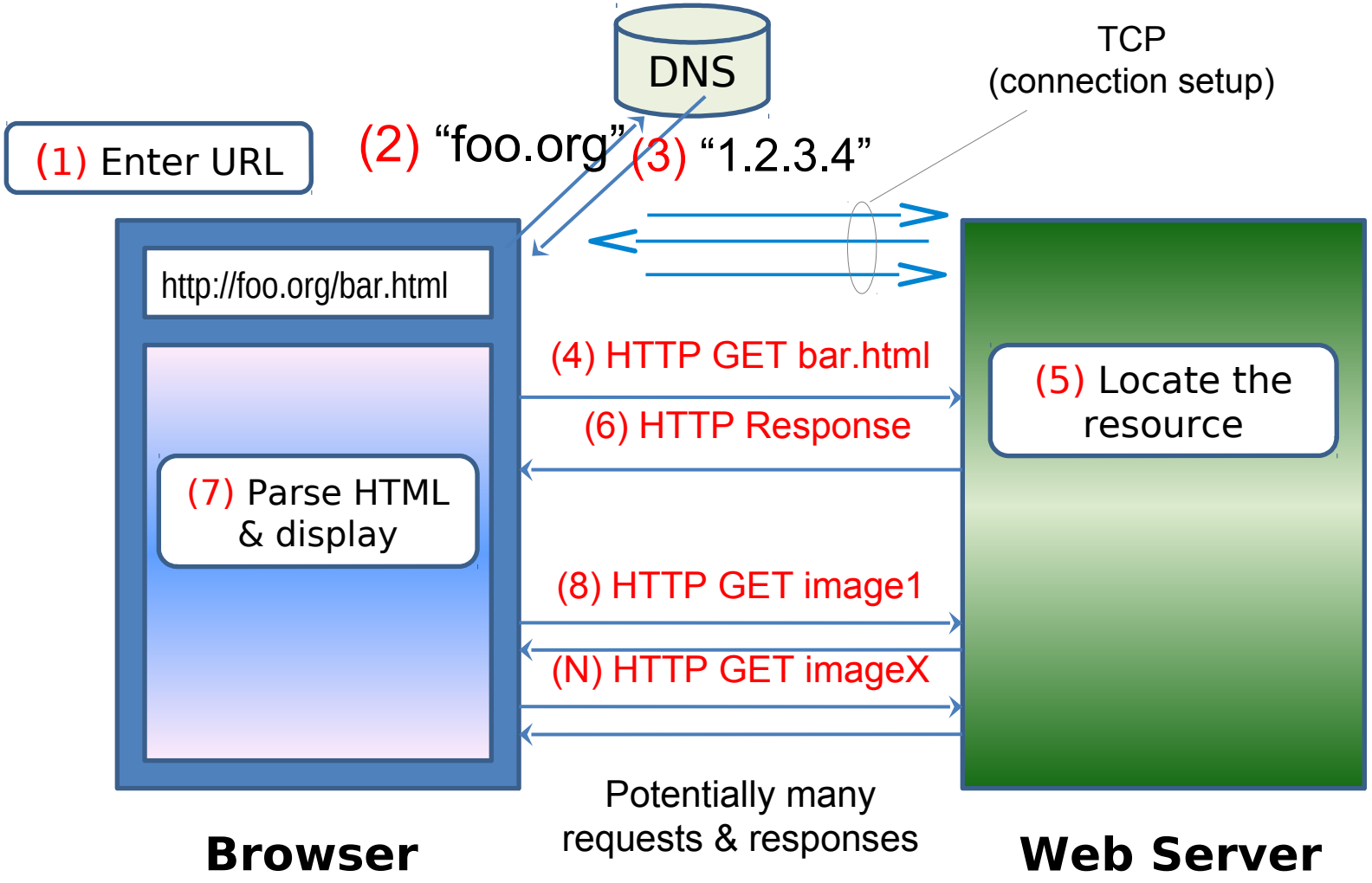


- **H**yper**T**ext **T**rasfer **P**rotocol is concerned with providing web pages, identified by an *Uniform Resource Locator* (**URL**):

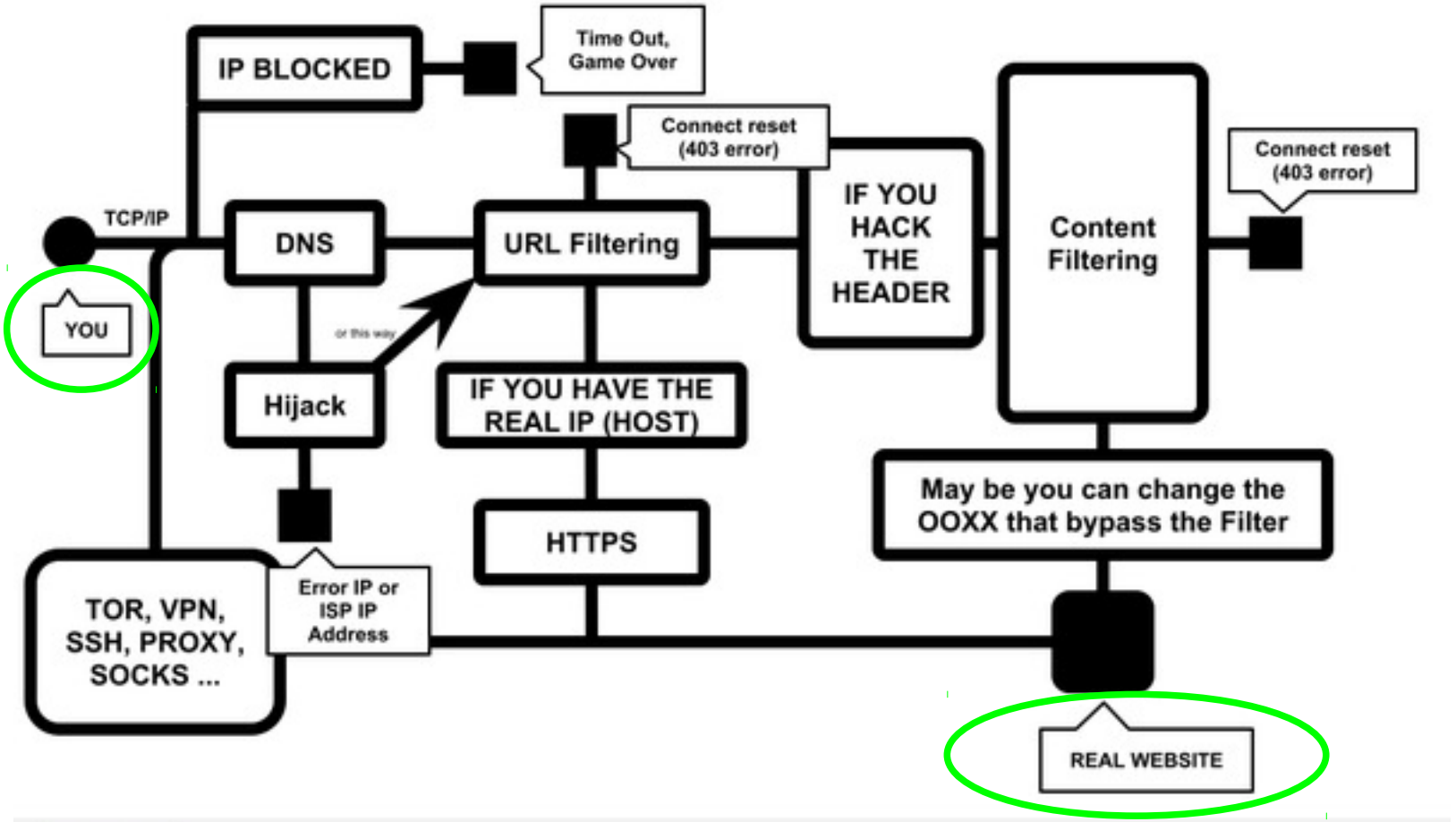
http://www.foo.org/abc/bar.html

protocol hostname path

Basics about Internet: a lot happens

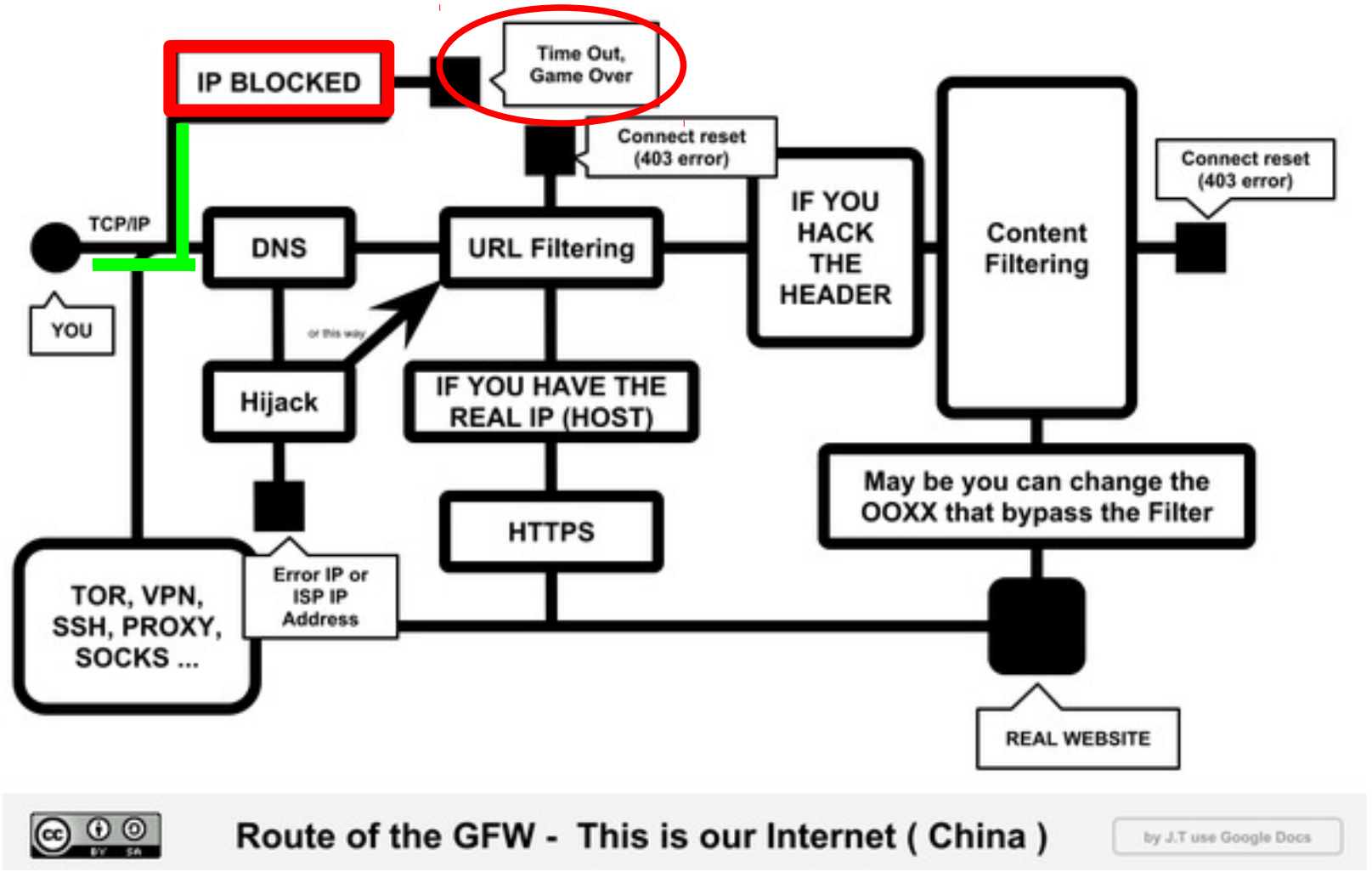


Access to content: an obstacle course

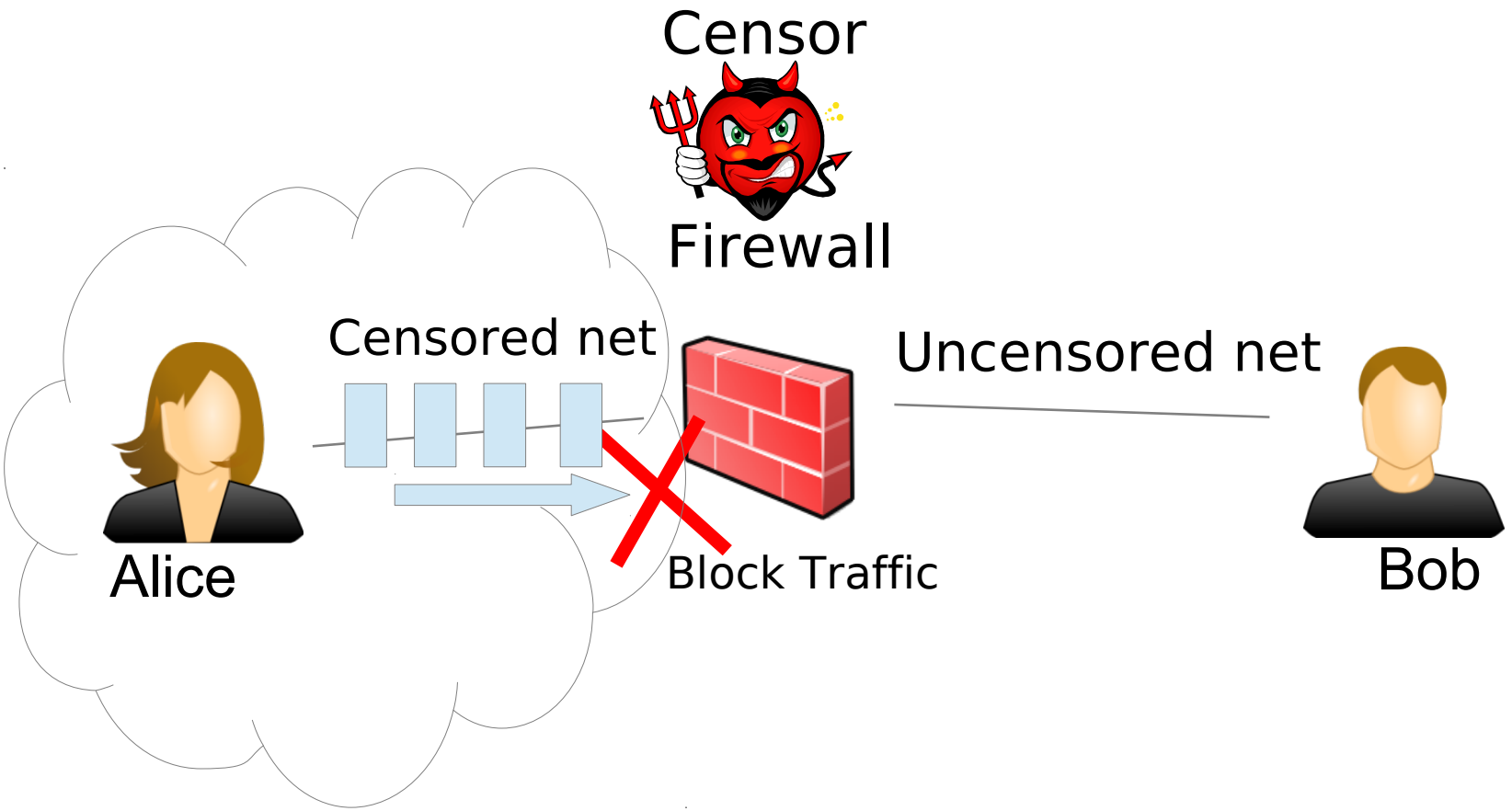


http://thenextweb.com/asia/2010/01/05/route-gfw-flow-chart/

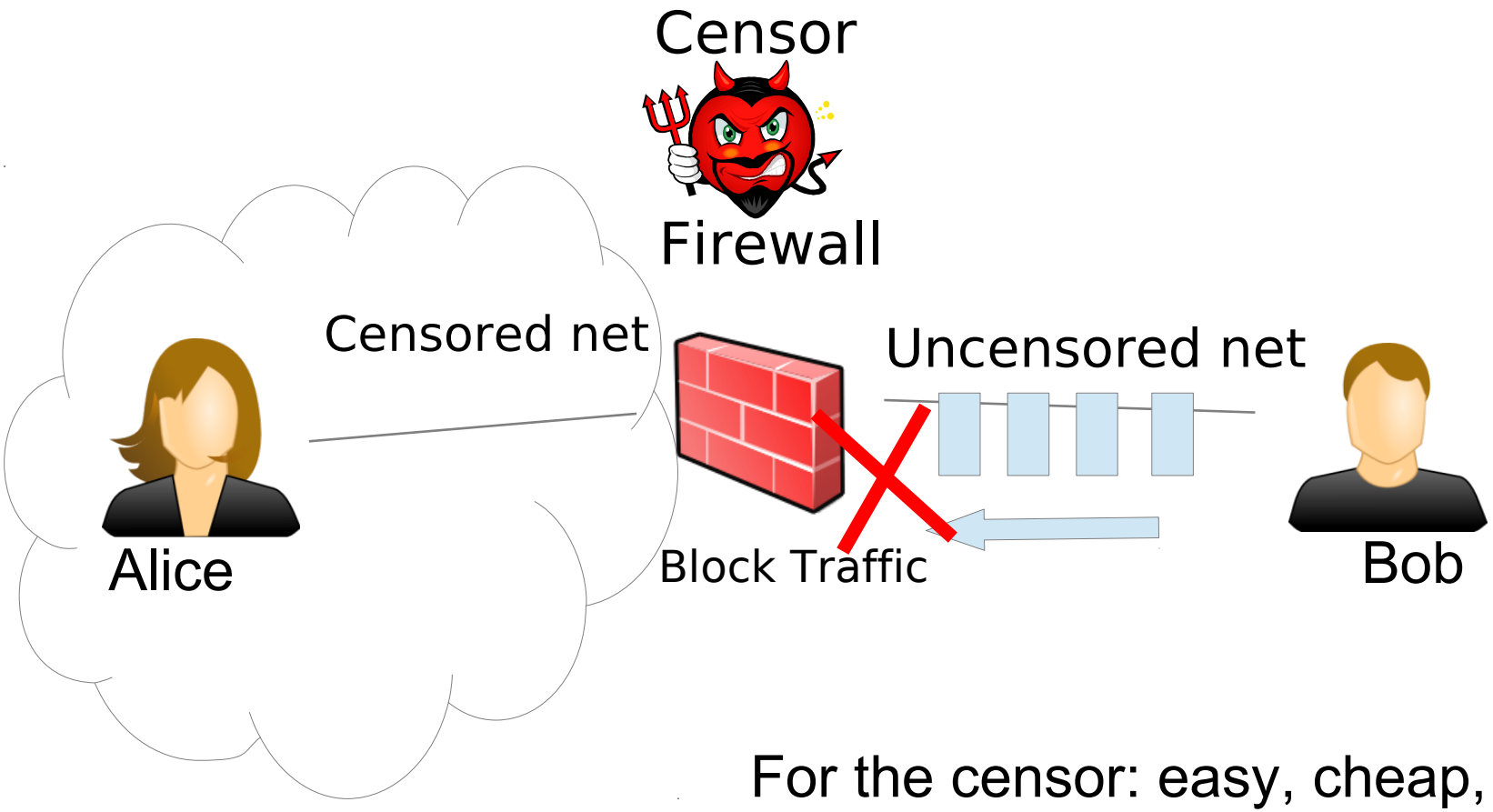
IP blocking


IP BLOCKED
Time Out, Game Over
Connect reset (403 error)
TCP/IP
DNS
URL Filtering
IF YOU HACK THE HEADER
Content Filtering
Connect reset (403 error)
YOU
or this way
Hijack
IF YOU HAVE THE REAL IP (HOST)
May be you can change the OOXX that bypass the Filter
HTTPS
TOR, VPN, SSH, PROXY, SOCKS ...
Error IP or ISP IP Address
REAL WEBSITE
Route of the GFW - This is our Internet (China)
by J.T use Google Docs

Conventional Internet Censorship (1/2)

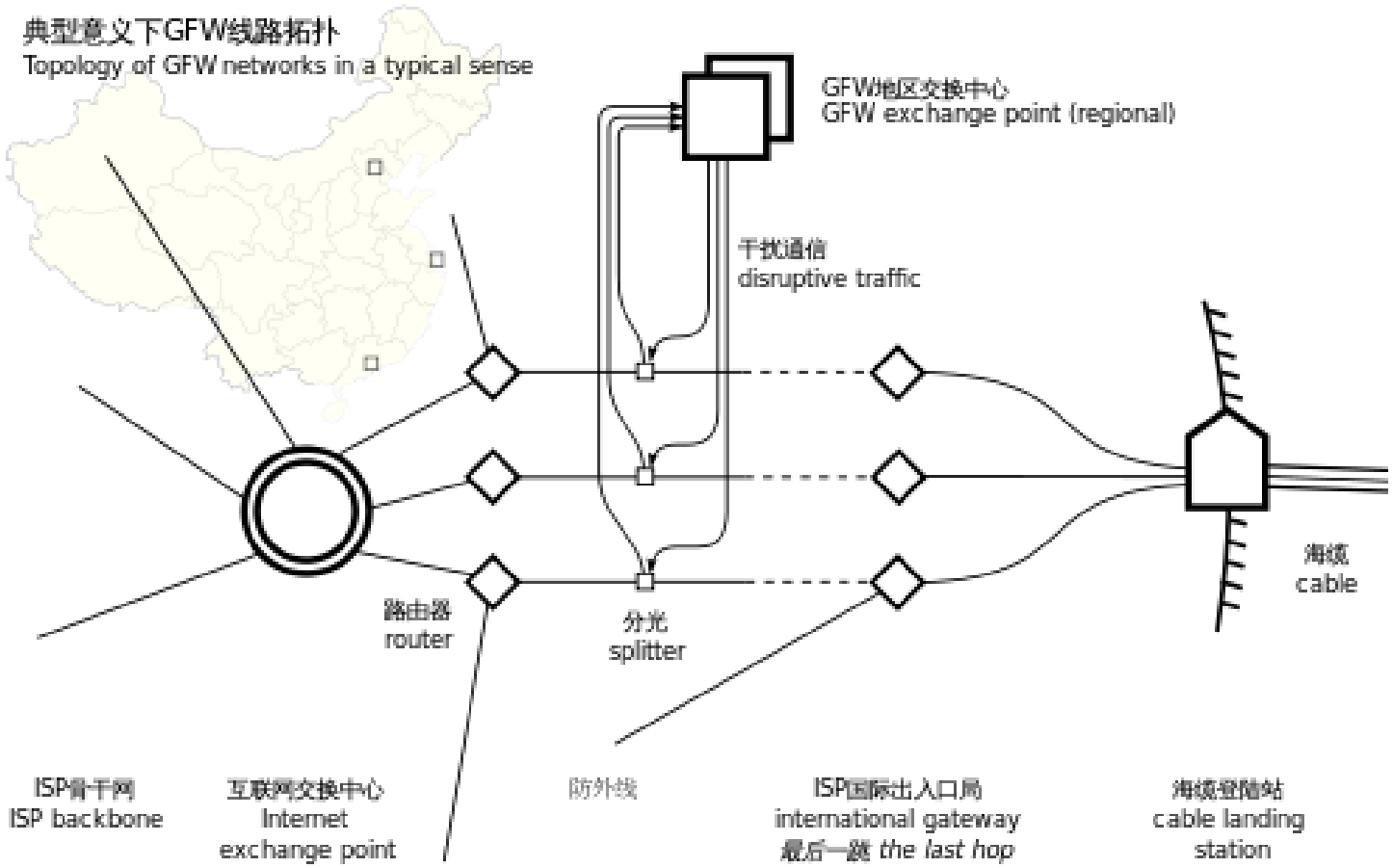


Conventional Internet Censorship (2/2)

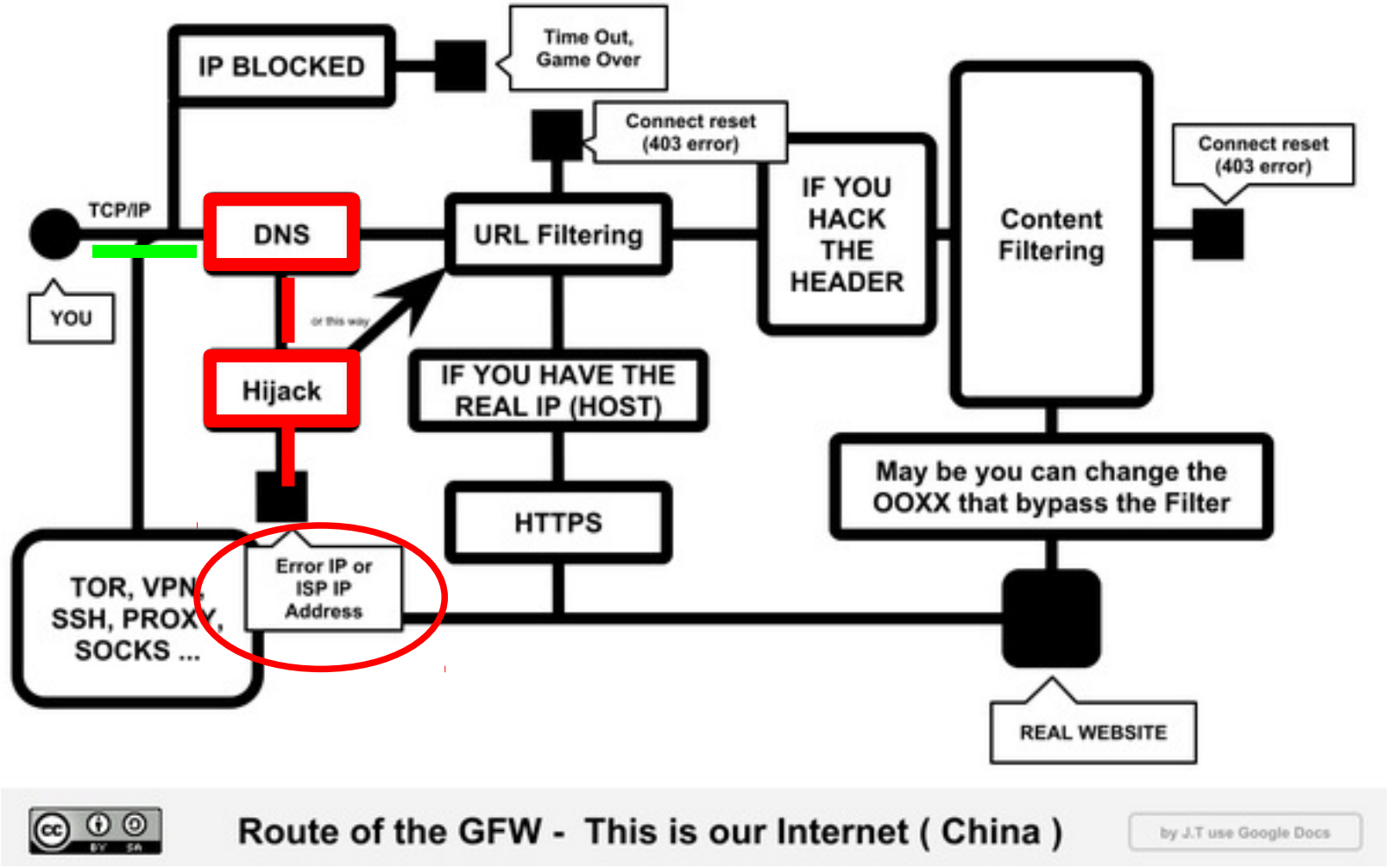


For the censor: easy, cheap,
but OVERBLOCKING issues

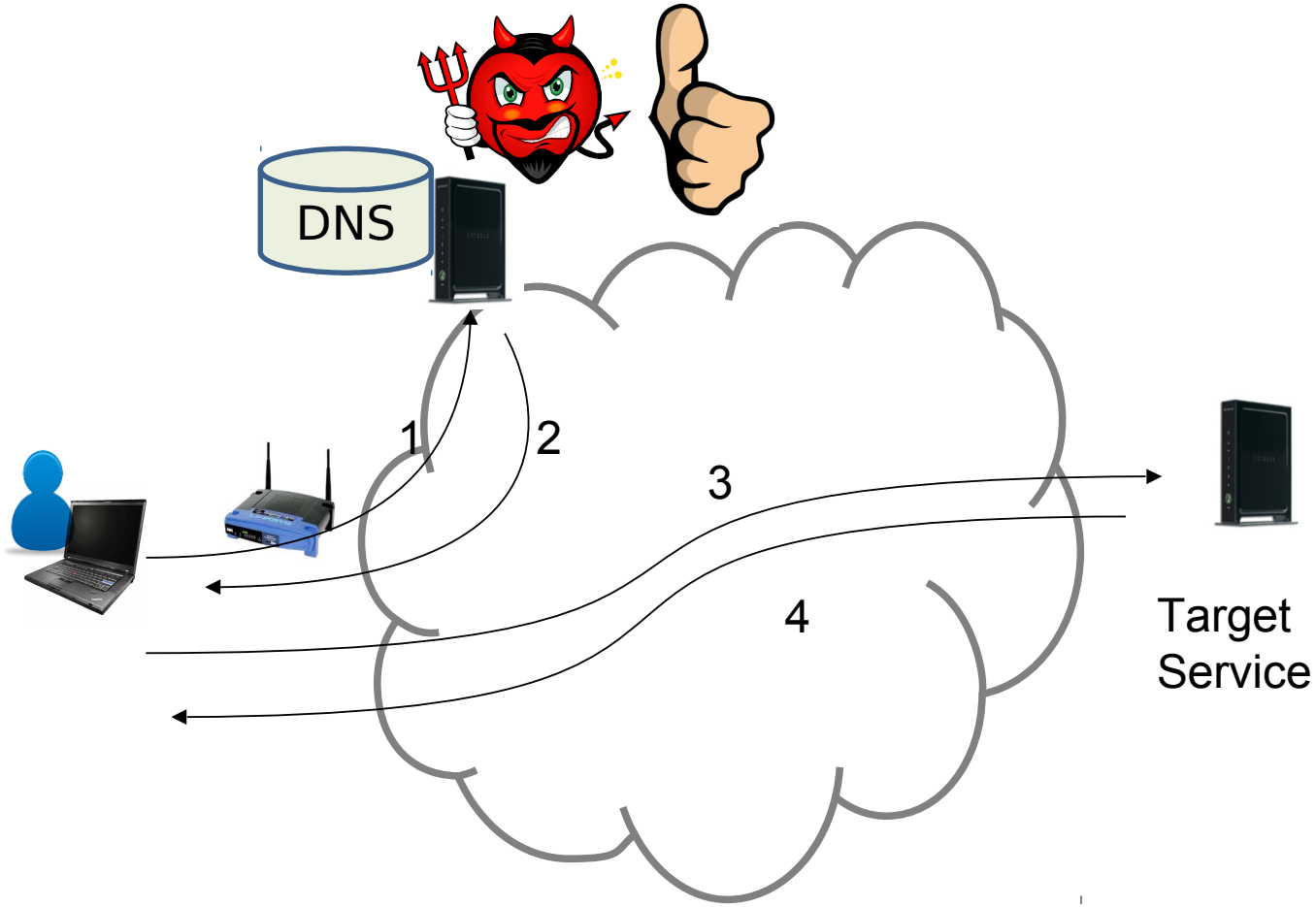
Real deployment can be more complex



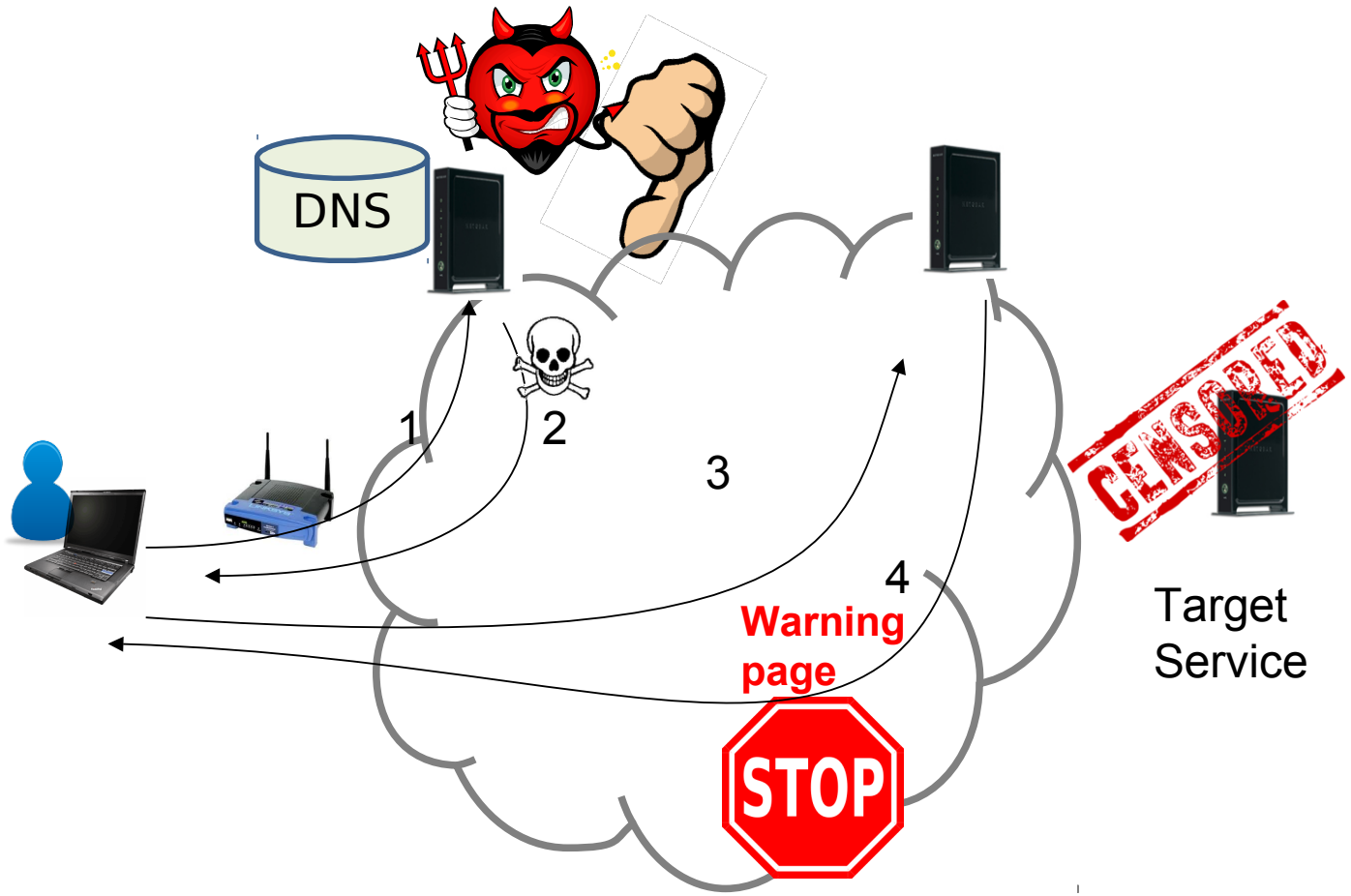
DNS blocking (hijacking)



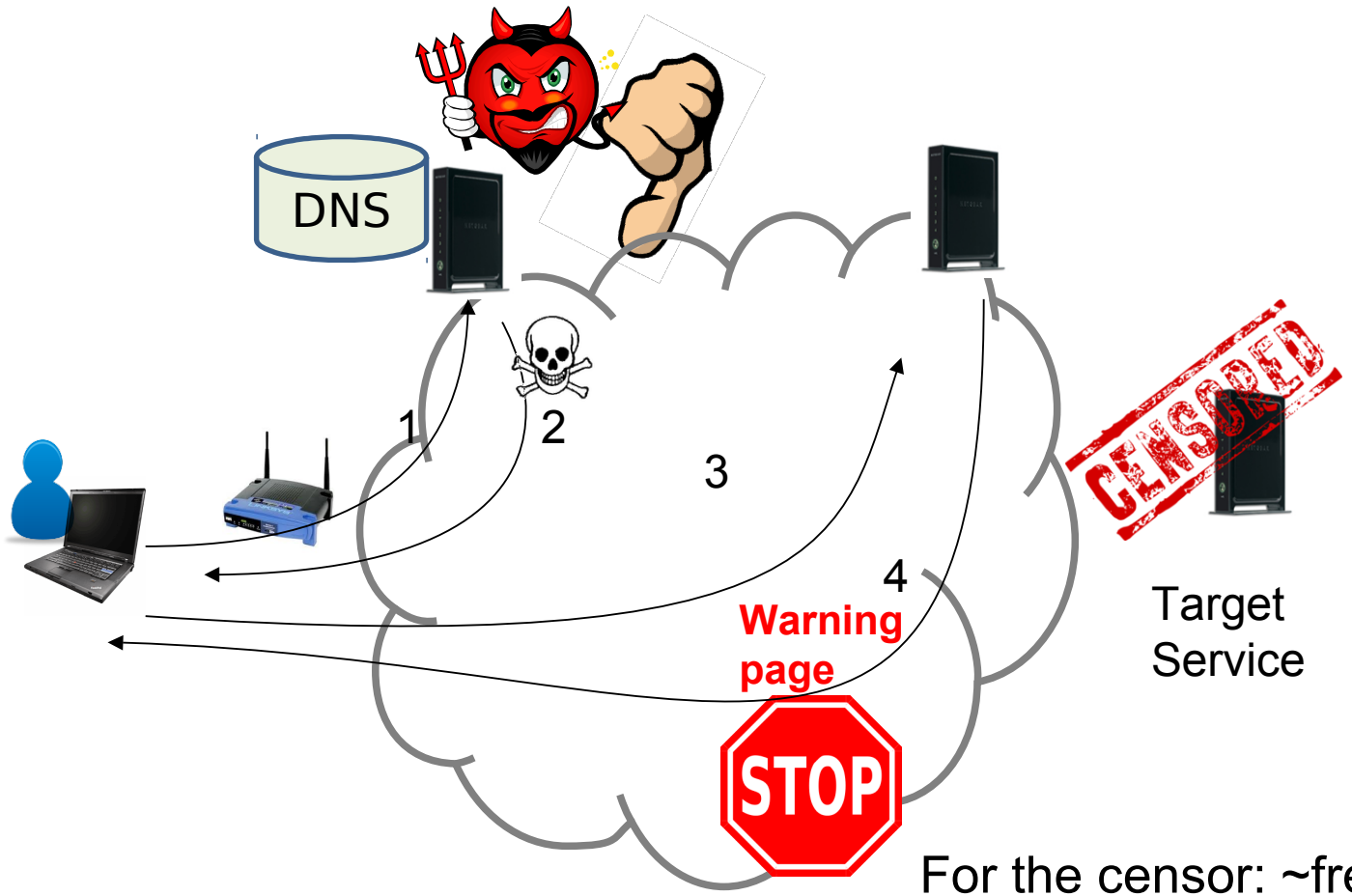
Domain Name System : normal behavior (and allowed hostnames)



Domain Name System : hijacking for blocked hostnames

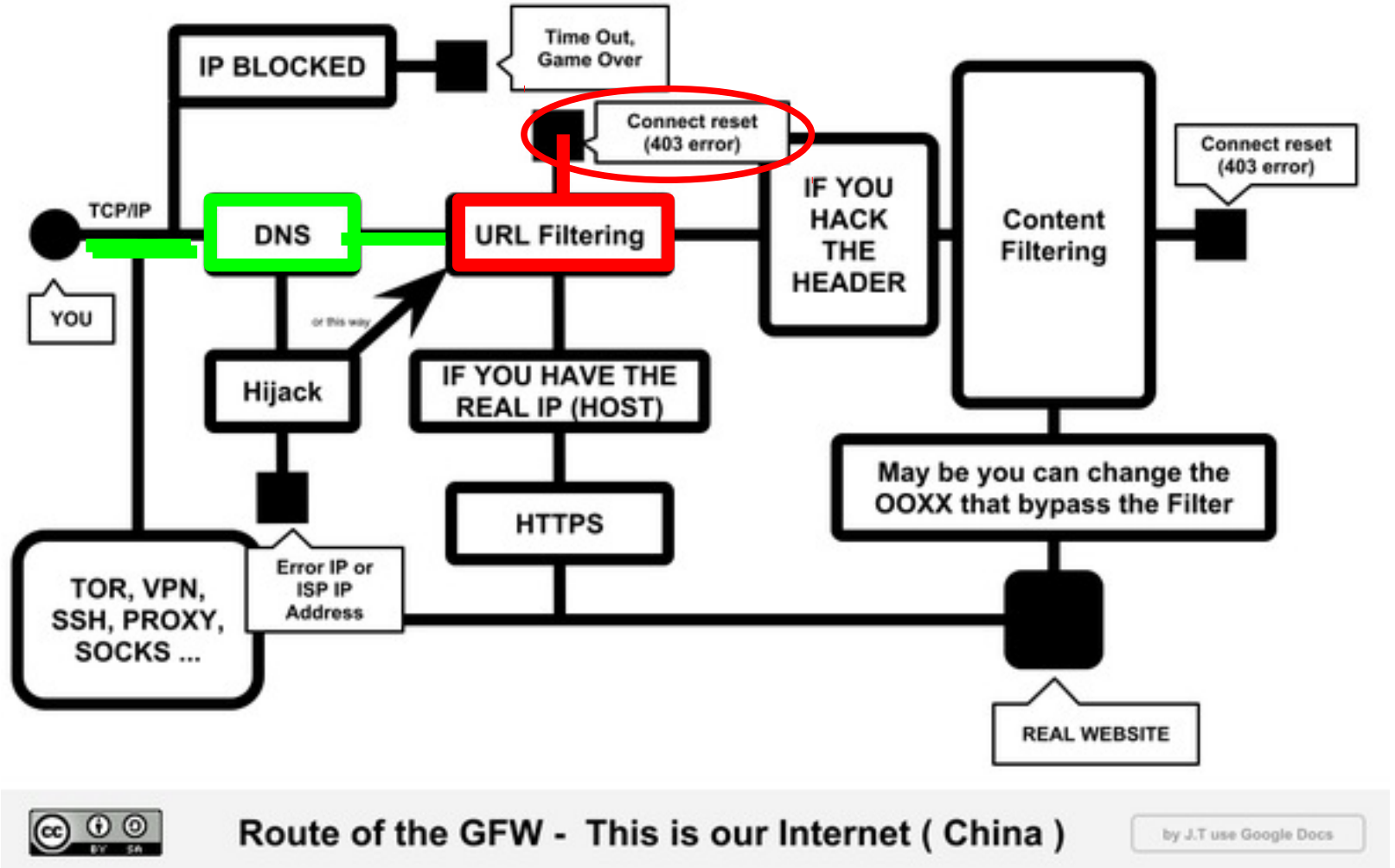


Domain Name System : hijacking for blocked hostnames

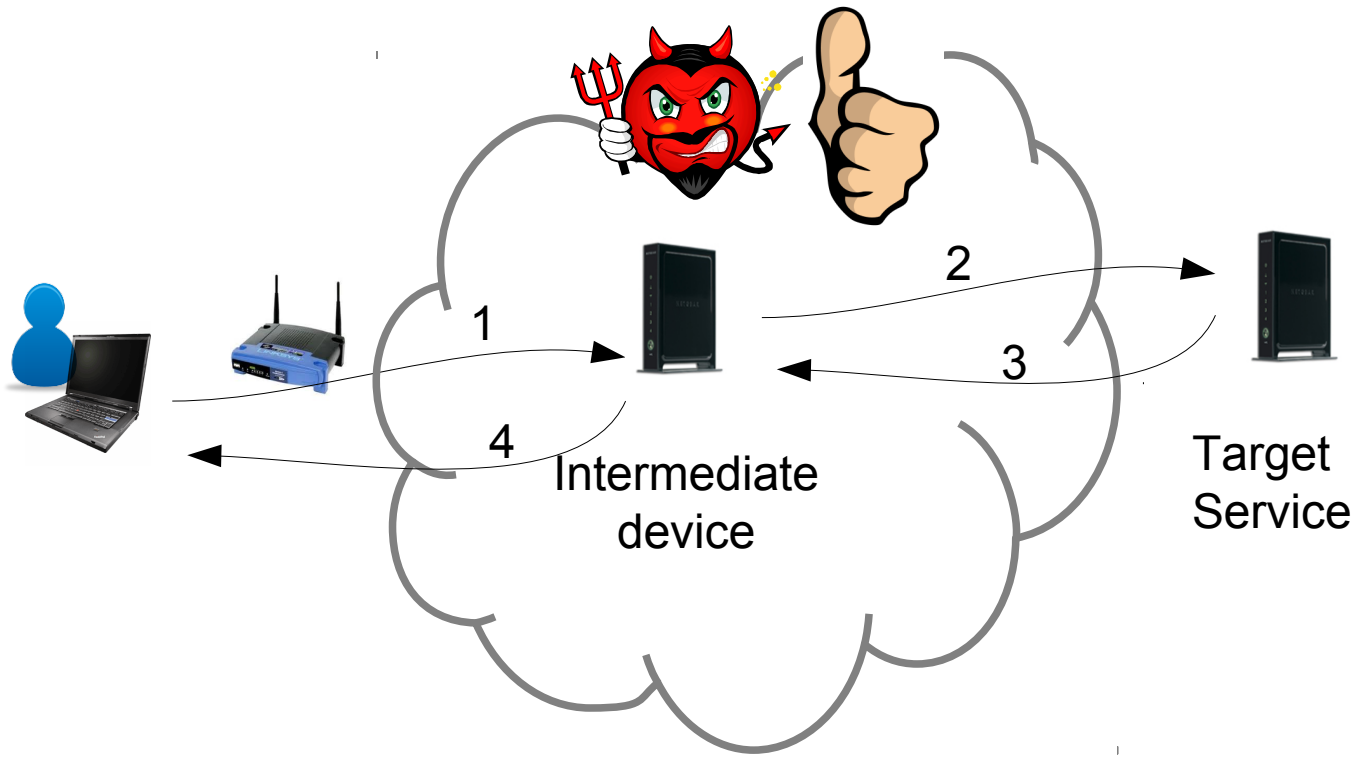


For the censor: ~free, easy but still OVERBLOCKING

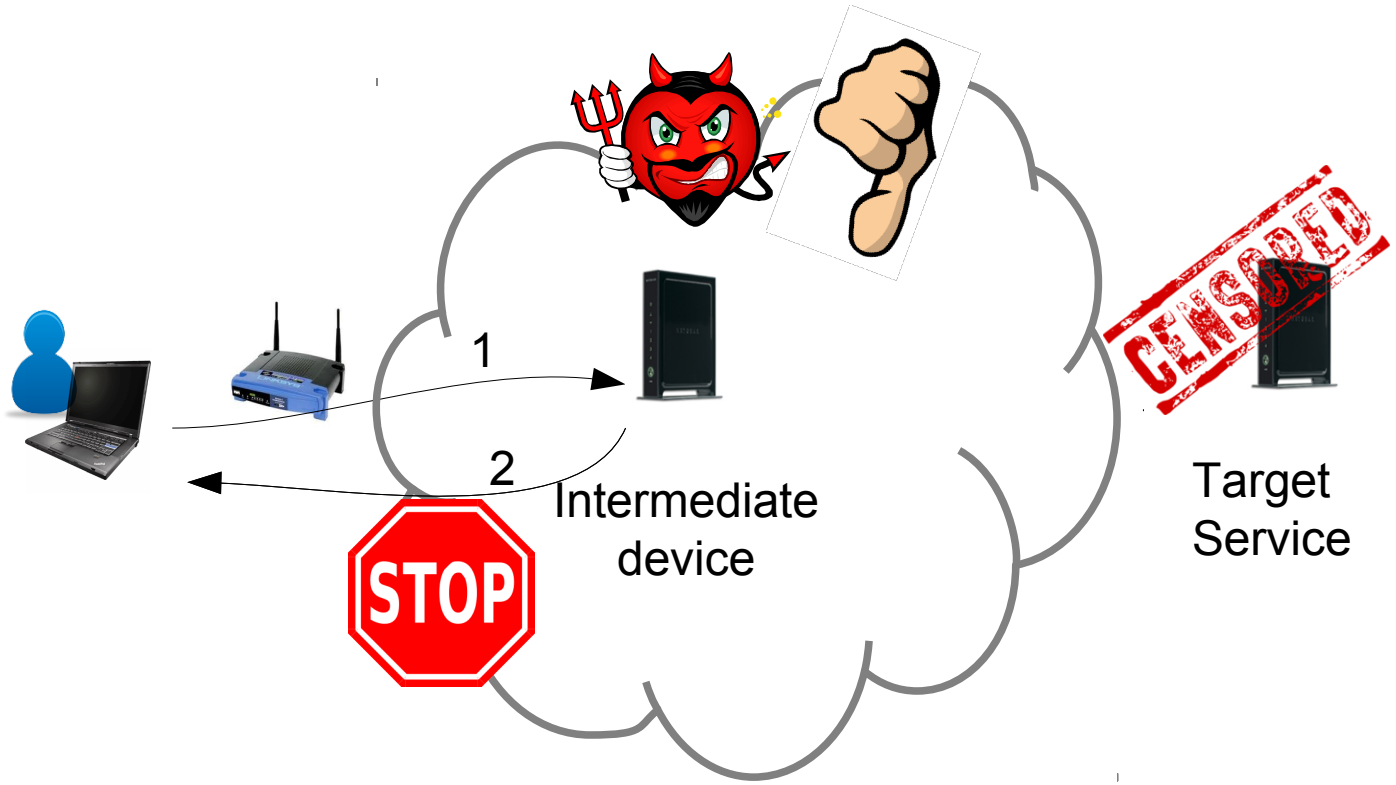
URL filtering



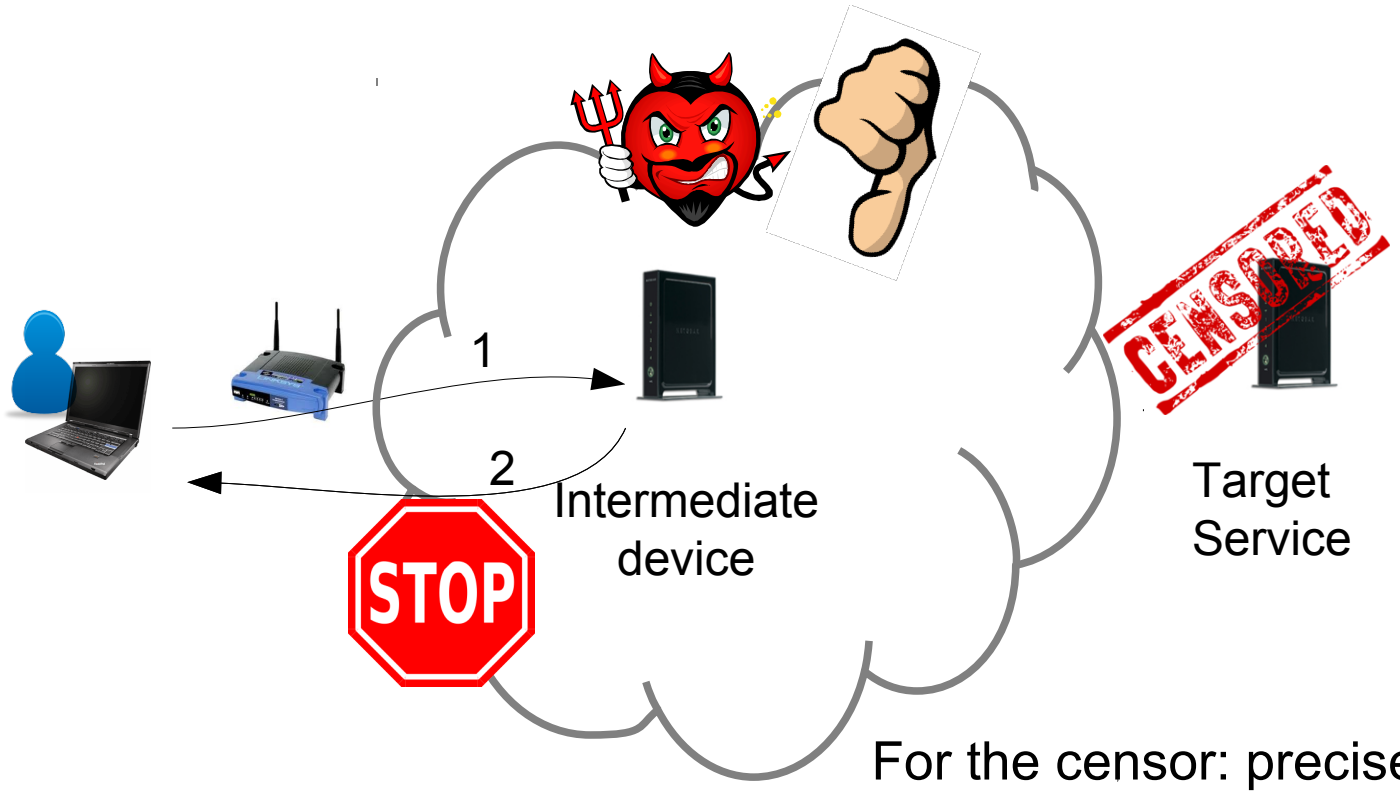
URL filtering: URL is ok



URL filtering: URL is blocked

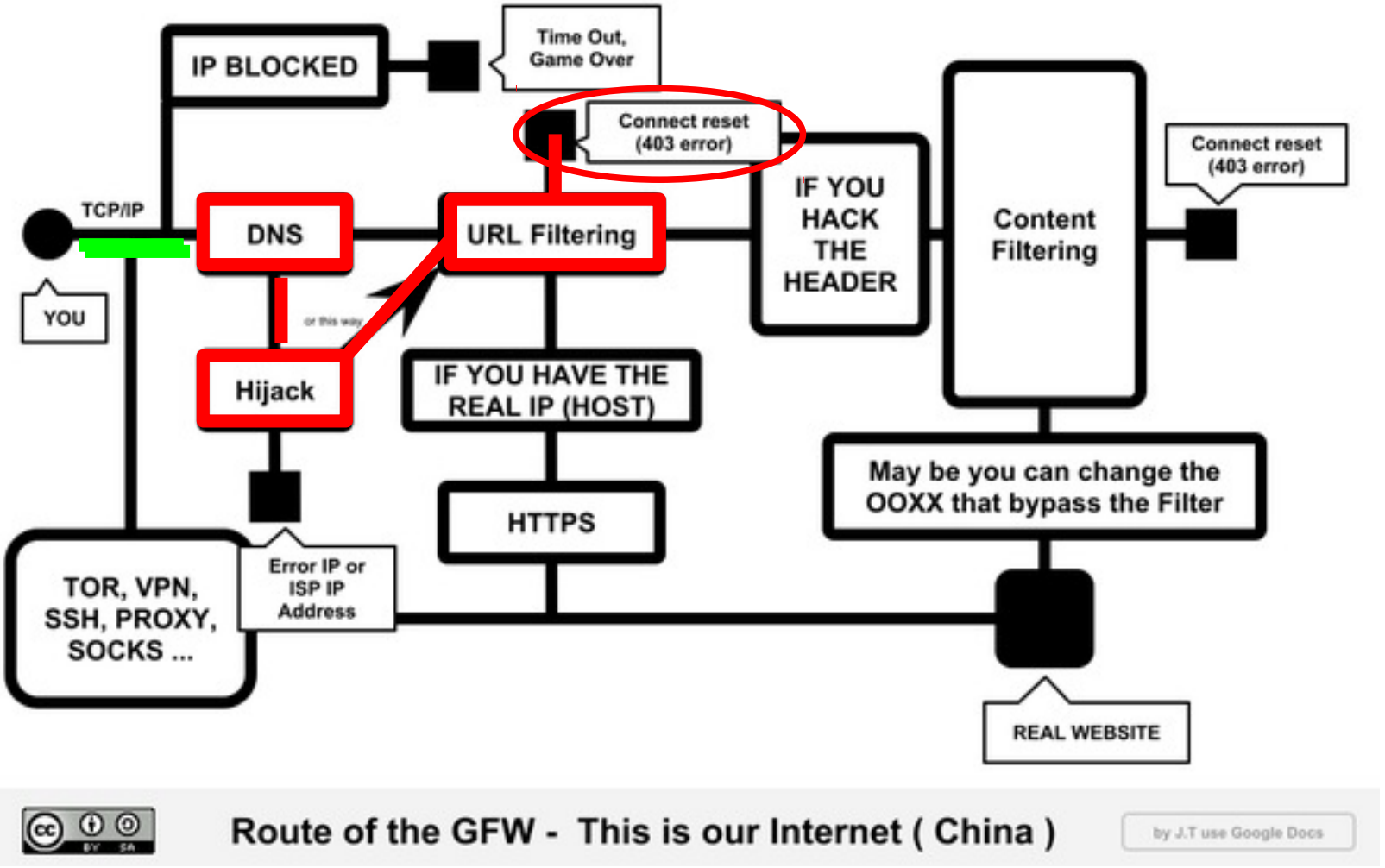


URL filtering: URL is blocked

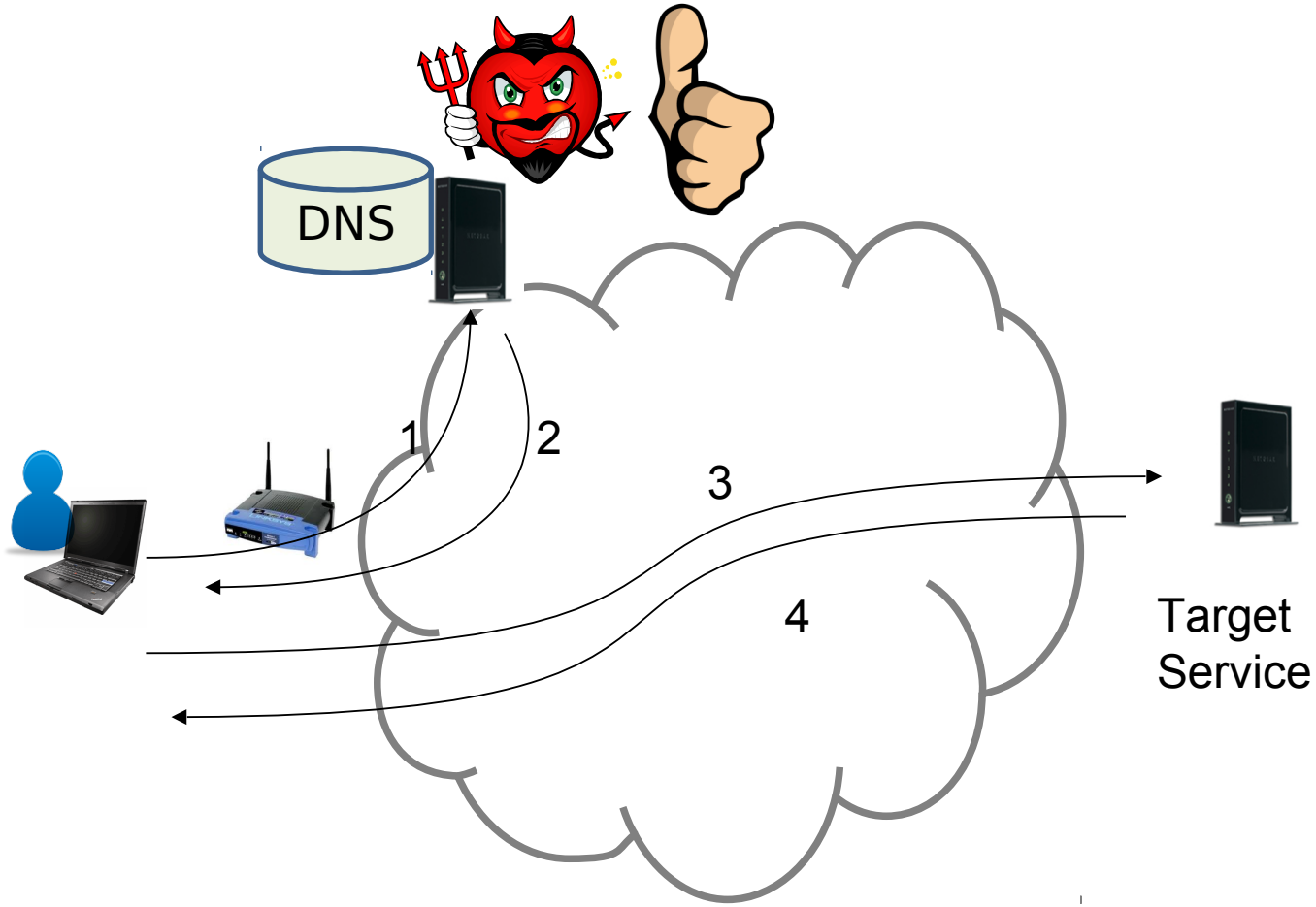


For the censor: precise, but expensive, performance issues

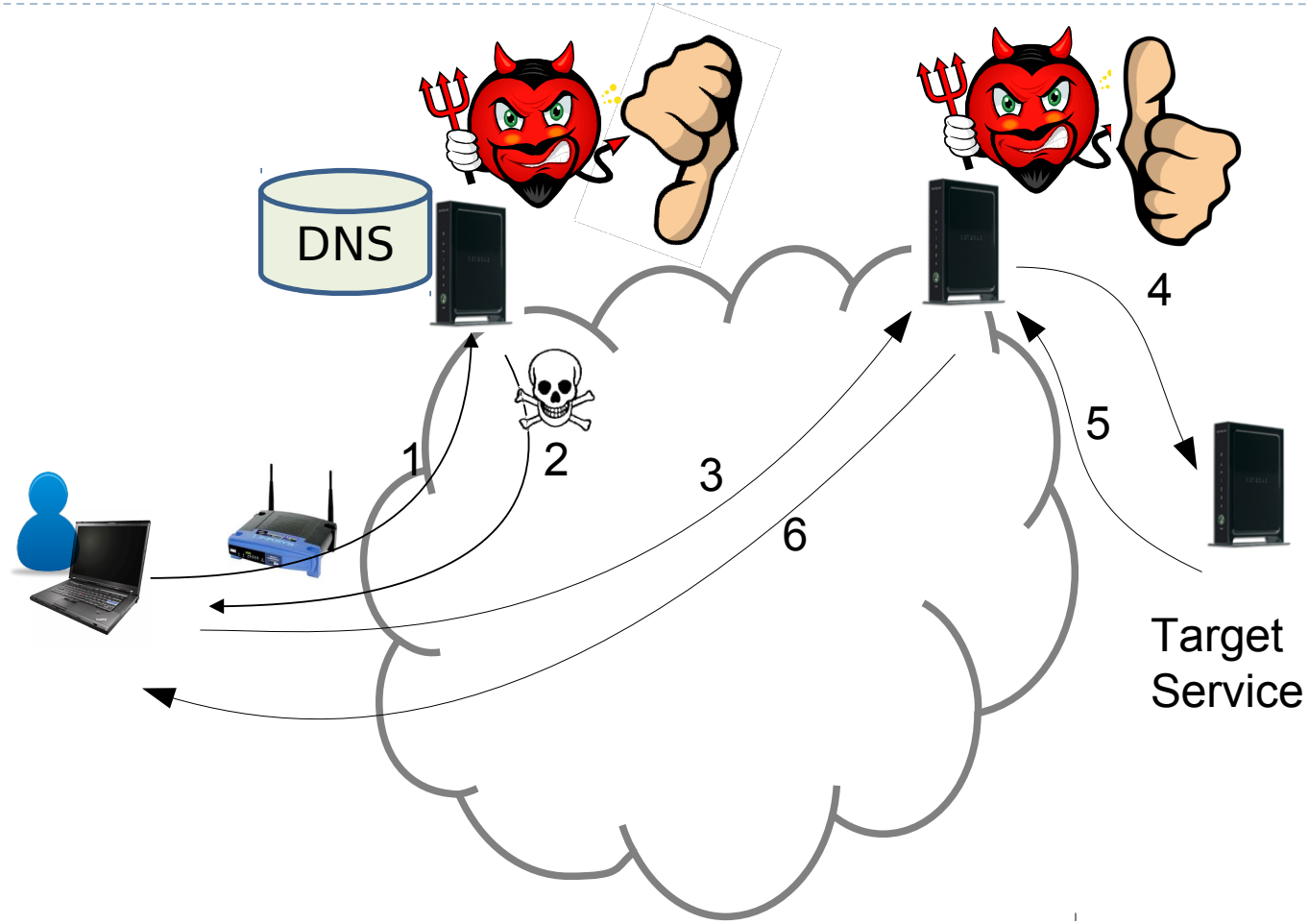
URL filtering (with DNS hijacking)



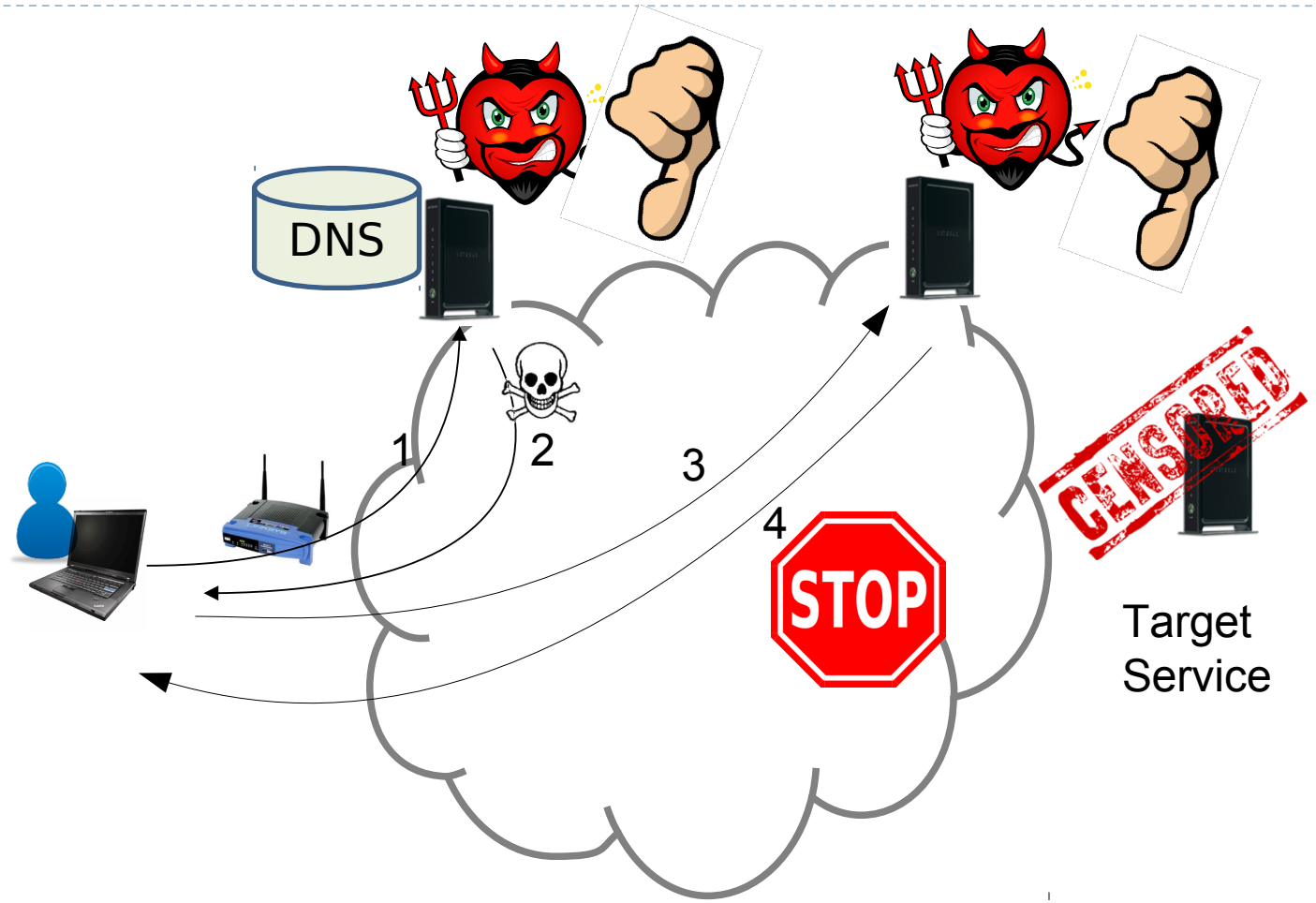
DNS hijacking + URL filtering: hostname is ok



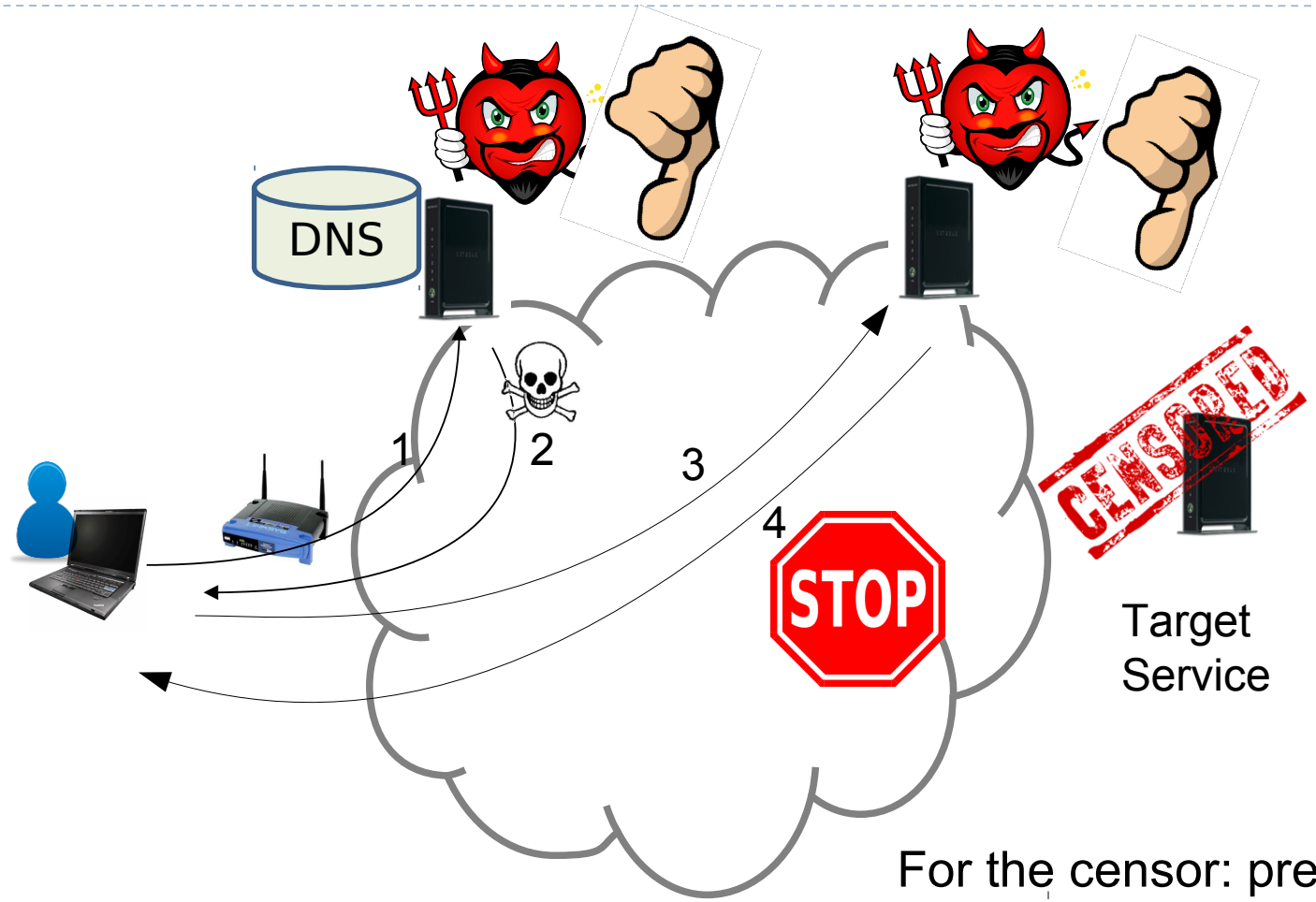
DNS hijacking + URL filtering: hostname with blocked pages – but ok



DNS hijacking + URL filtering: URL was censored

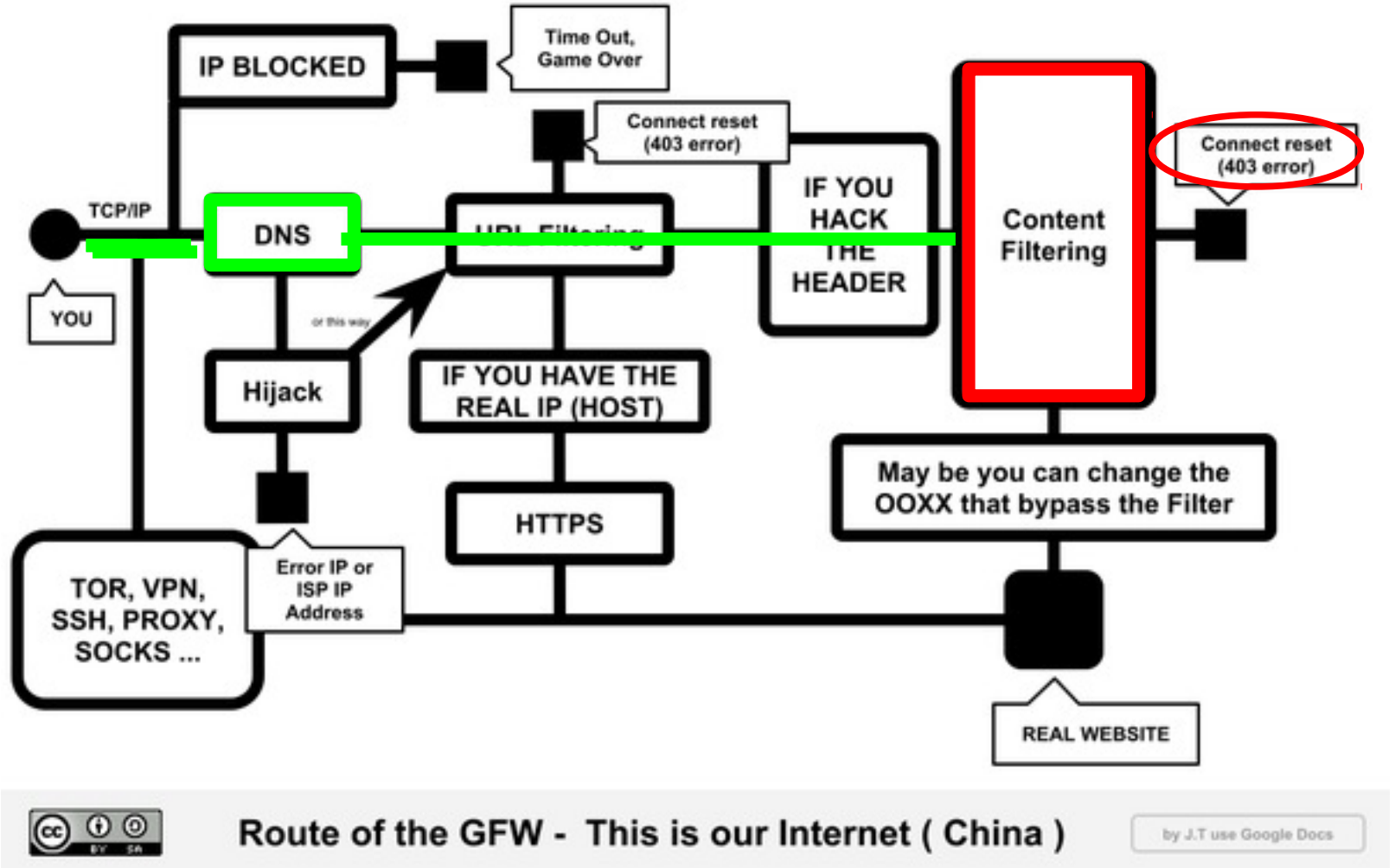


DNS hijacking + URL filtering: URL was censored

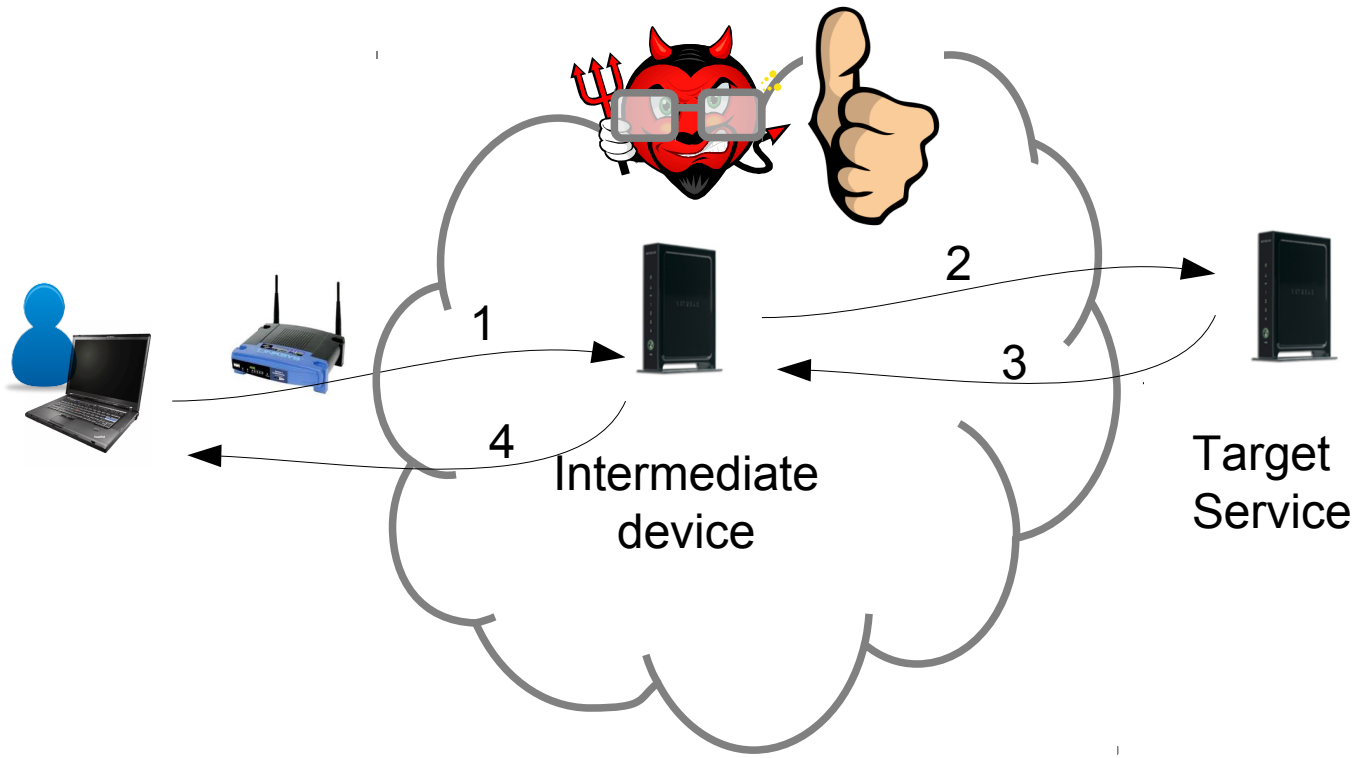


For the censor: precise, less performance issues, still expensive

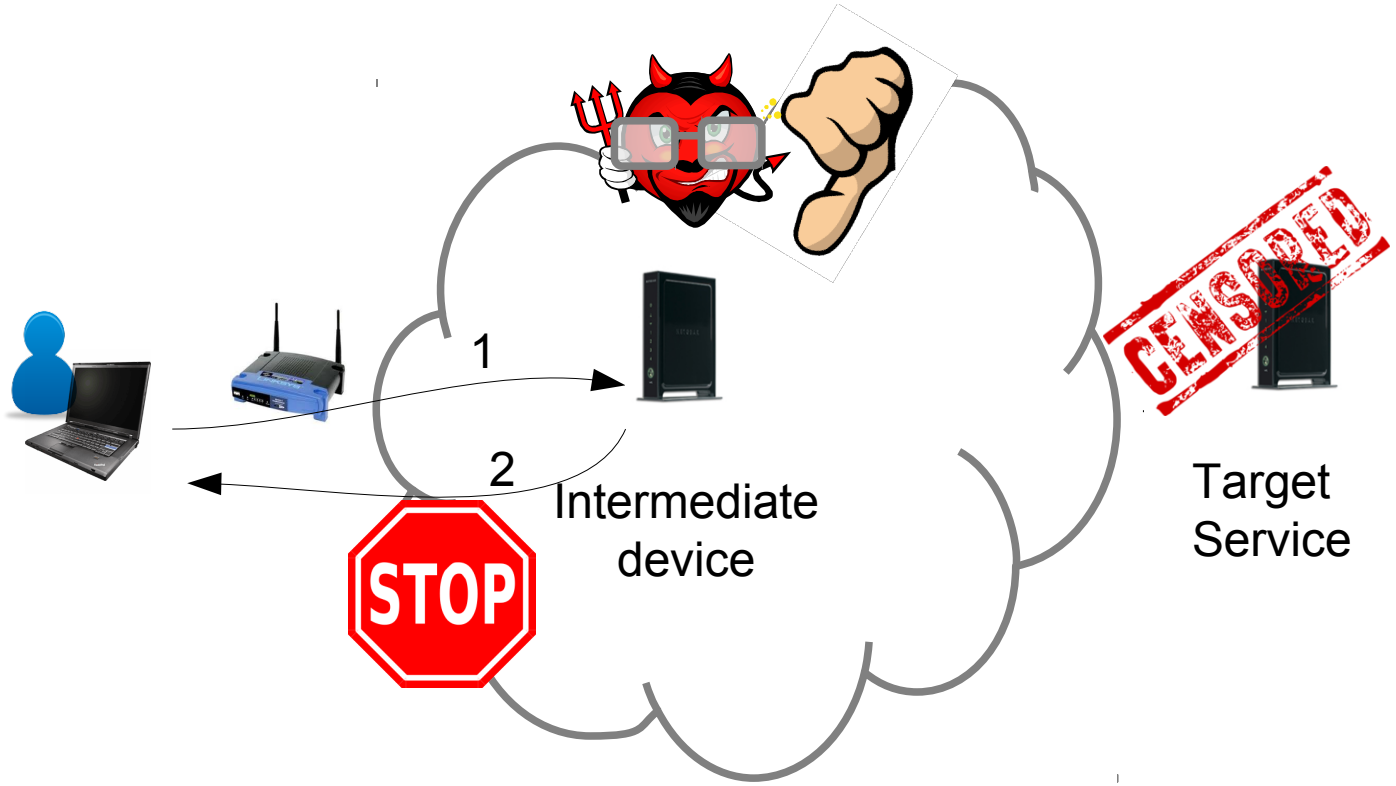
Content filtering



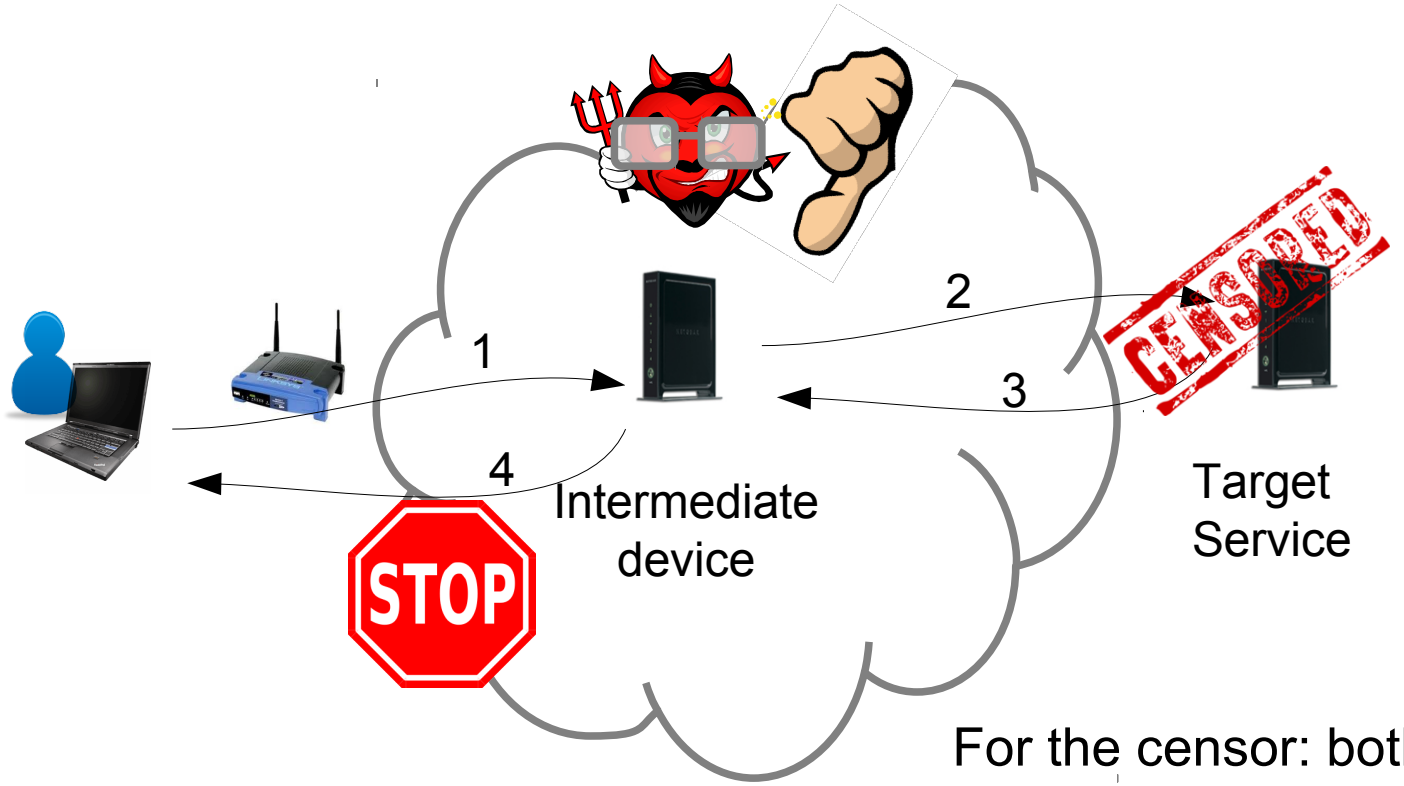
Content filtering: content is ok



Content filtering: *sender* content is blocked



Content filtering: *receiver* content is blocked

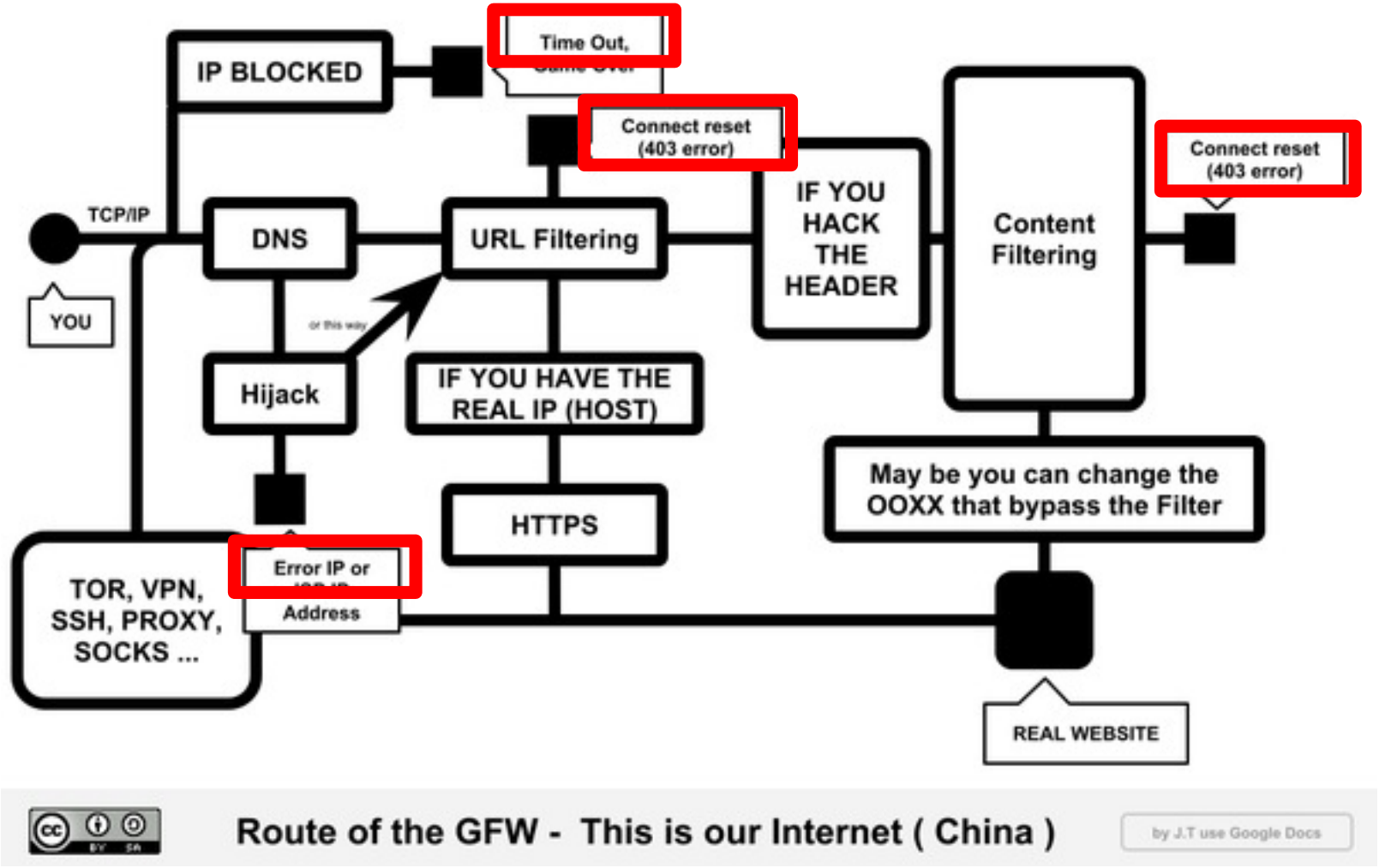


For the censor: both very precise or dramatic failure, most expensive

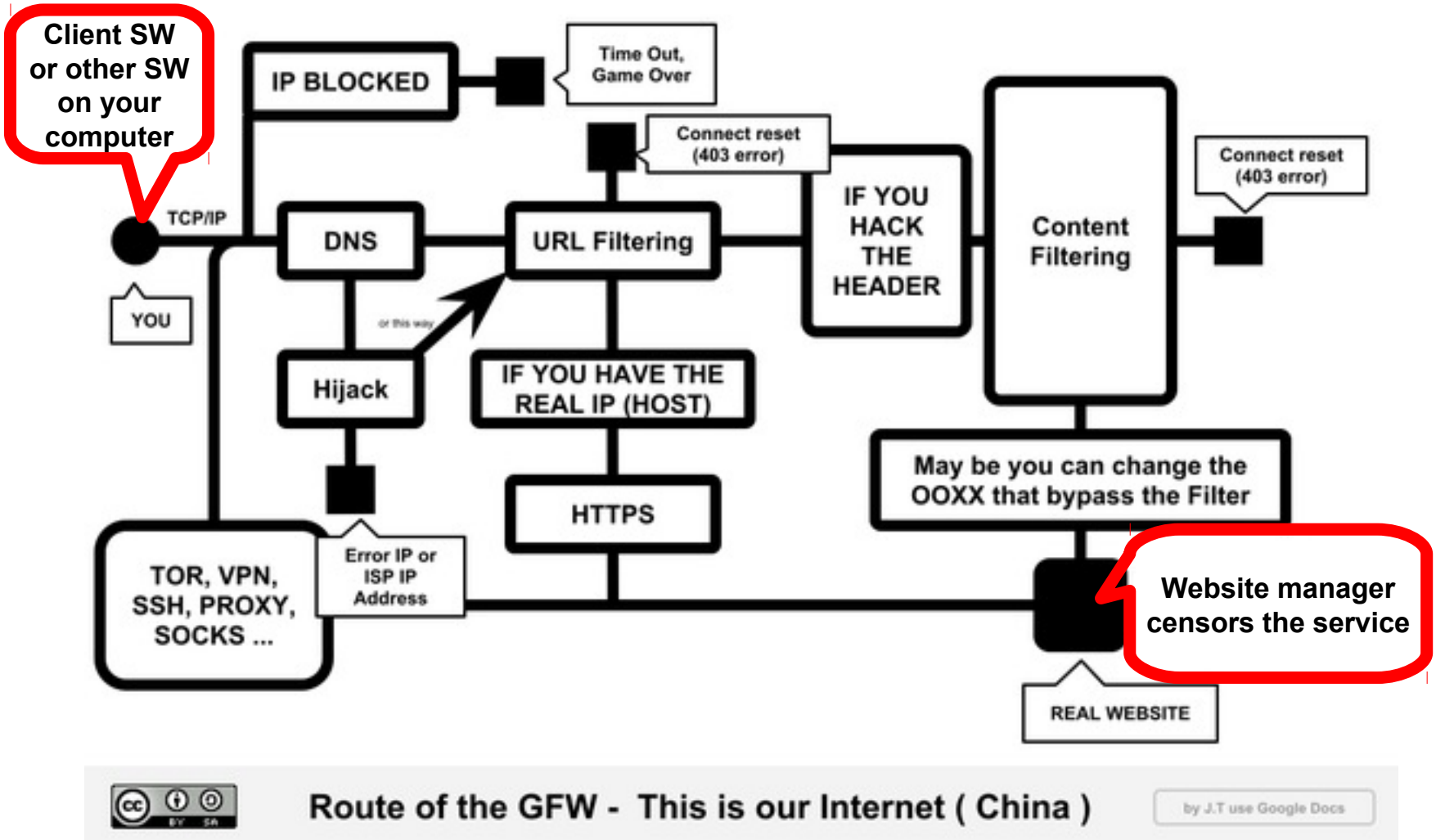
Detection challenges: observing censorship is not easy

- **Targets** of censorship change from country to country
  - … and sometimes in regions inside the country
  - … and in time
- Censorship **techniques** change in time, too
- To observe censorship we need (usually) a vantage point **inside the censored network**
- … but also other vantage points **from outside** the censored network, to tell censorship from outages
- Some kinds of censorship are **inherently** hard to detect
  - server-side self-censorship
  - User client software censorship
  - User self-censorship

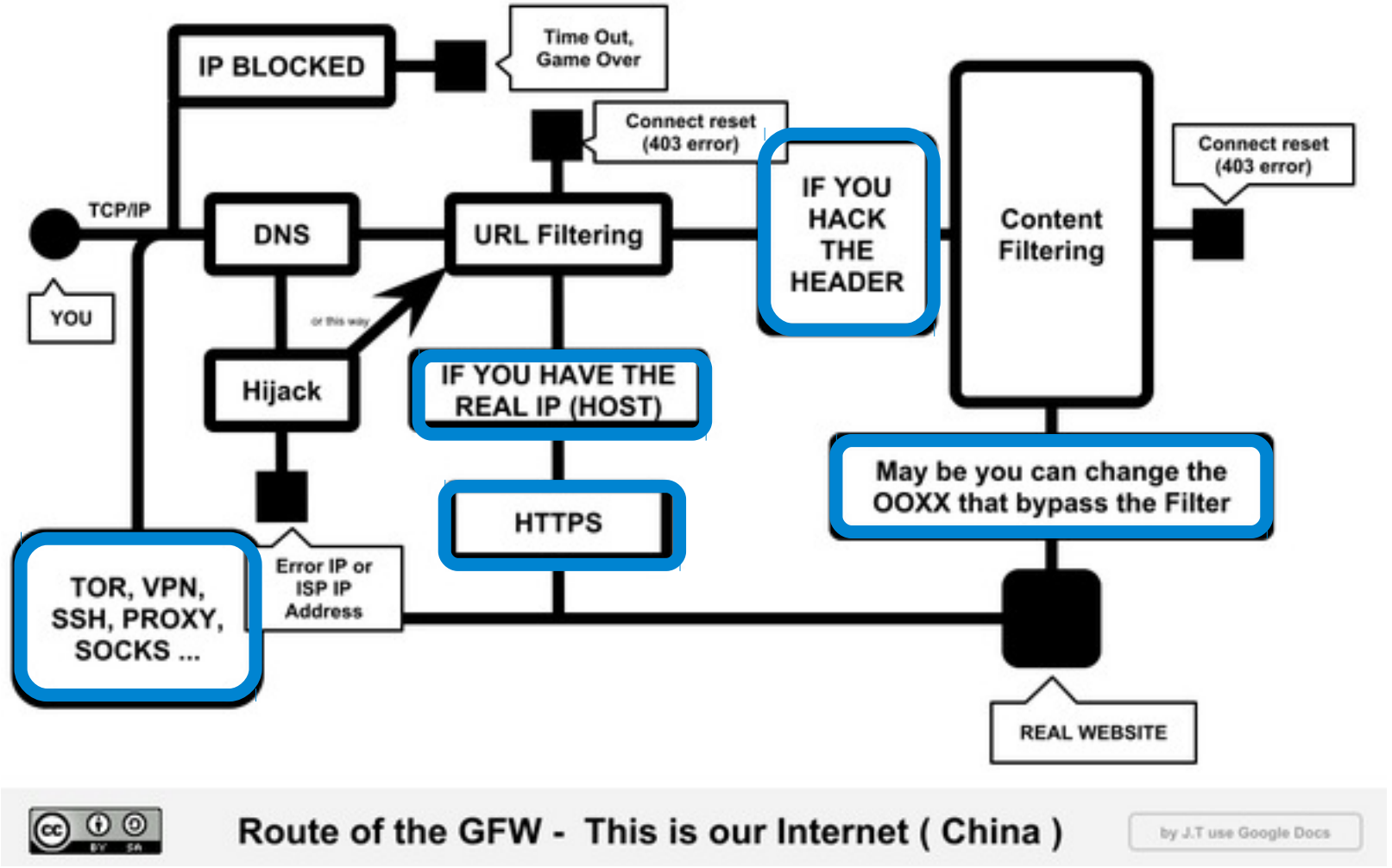
Detection challenges: similar symptoms, same as outages



“outside-the-Internet” blocking



Detection importance: awareness, circumvention



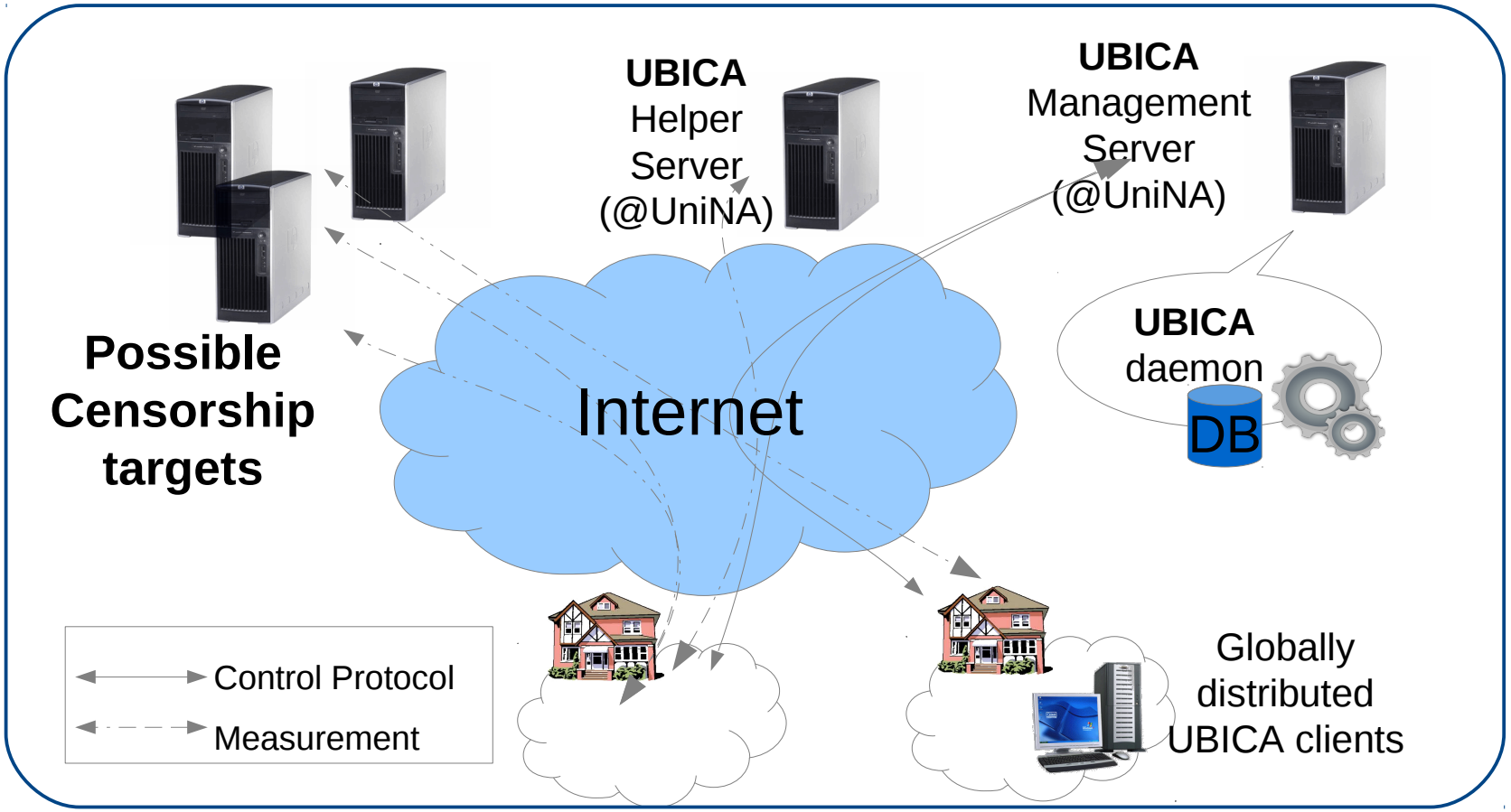
Our client-based approach: UBICA

- Multi-platform application
  Linux, Mac OSX, Windows

- Extensible measurement framework
  any underlying measurement tool supported
  flexible experiments

- Real-time reporting at different aggregation levels

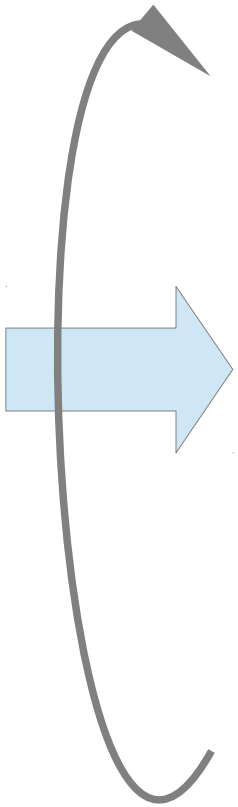
UBICA architecture



UBICA monitoring cycle

1. **Collection of Targets**
2. ***Scheduling* of evidence collection**
3. ***Evidence collection* by probes**
4. ***Evidence reporting* and data export**
5. **Censorship *Tests***
6. ***Update* Targets and Scheduling**

UBICA monitoring cycle



1. **Collection of Targets**
2. ***Scheduling* of evidence collection**
3. ***Evidence collection* by probes**
4. ***Evidence reporting* and data export**
5. **Censorship *Tests***
6. ***Update* Targets and Scheduling**

Evidence collection

- **TCP connectivity**
- **DNS lookup**
- **HTTPscan (URL retrieval)**
- **Keyword-based HTTP retrieval**
- **Topology**
- **Performance**

UBICA monitoring cycle

1. **Collection of Targets**
2. ***Scheduling* of evidence collection**
3. ***Evidence collection* by probes**
4. ***Evidence reporting* and data export**
5. **Censorship *Tests***
6. ***Update* Targets and Scheduling**

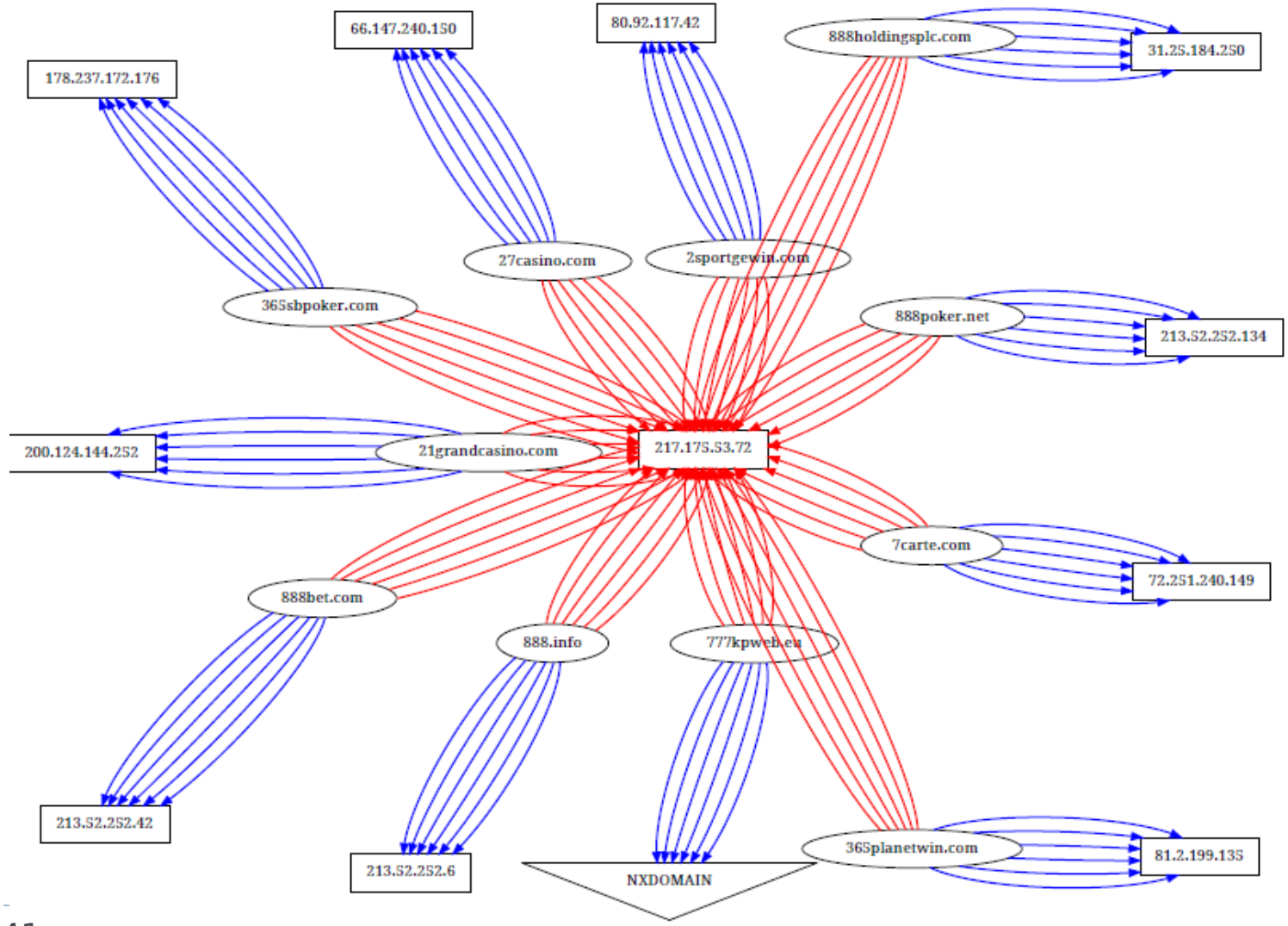
Censorship Checks

- **IP / port filtering**
- **DNS tampering**
- **URL filtering**
- **Localization**
- **Content mangling**

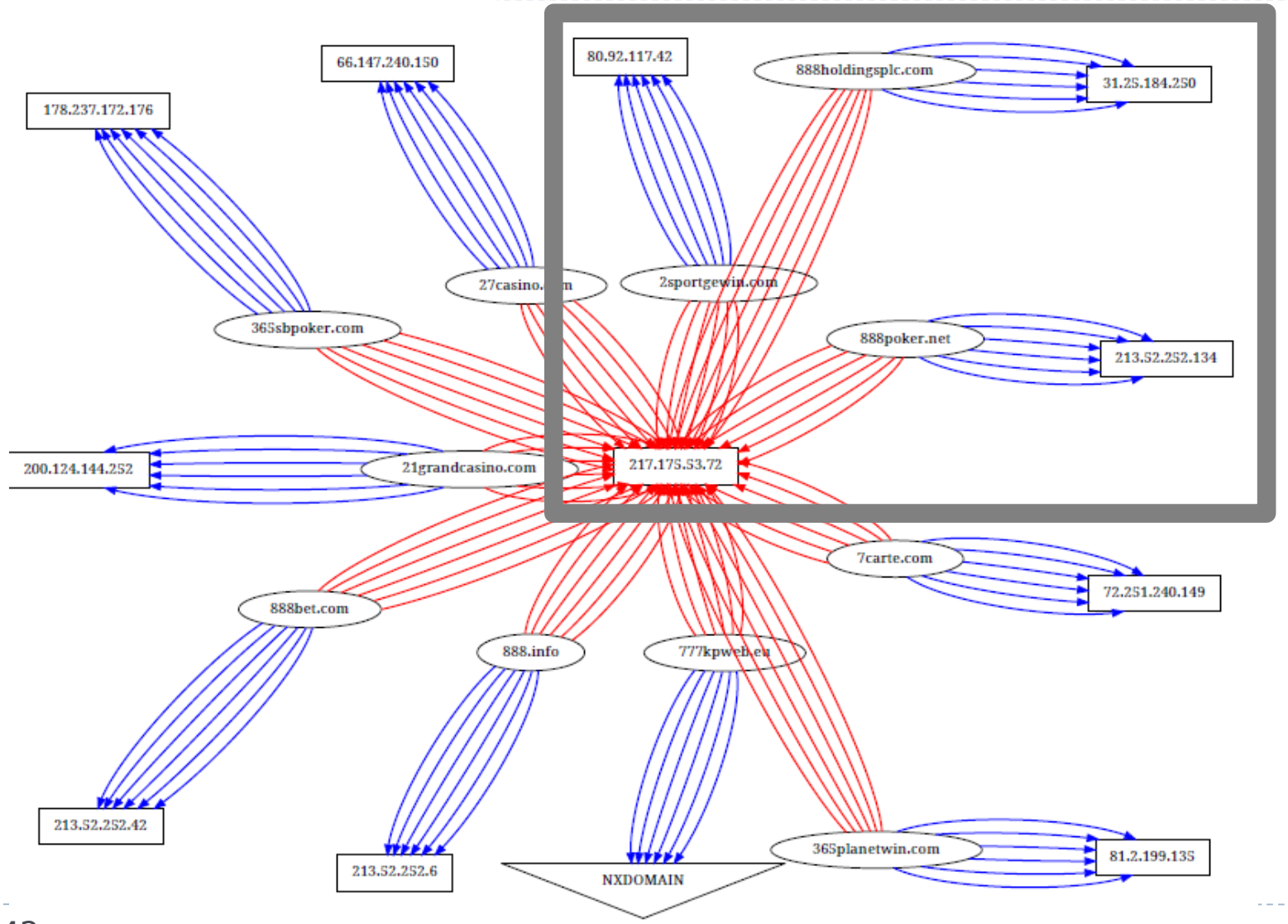
Test: DNS censorship

- DNS resolution
  - **Collection**: an A query for the target hostname is requested to the local default DNS server and to a small number of open resolvers
  - **Check local:** for each hostname compare the set of IP addresses from the default resolver against the ones from the open resolvers
  - **Check global:** for each hostname compare the set of resolved IP addresses inside a zone against other ones
  - **Check graph:** considering the bipartite graph of hostnames and their resolved IP addresses, with edges representing the name resolution relation, compute the in-degree and out-degree of nodes inside a zone and compare with other ones

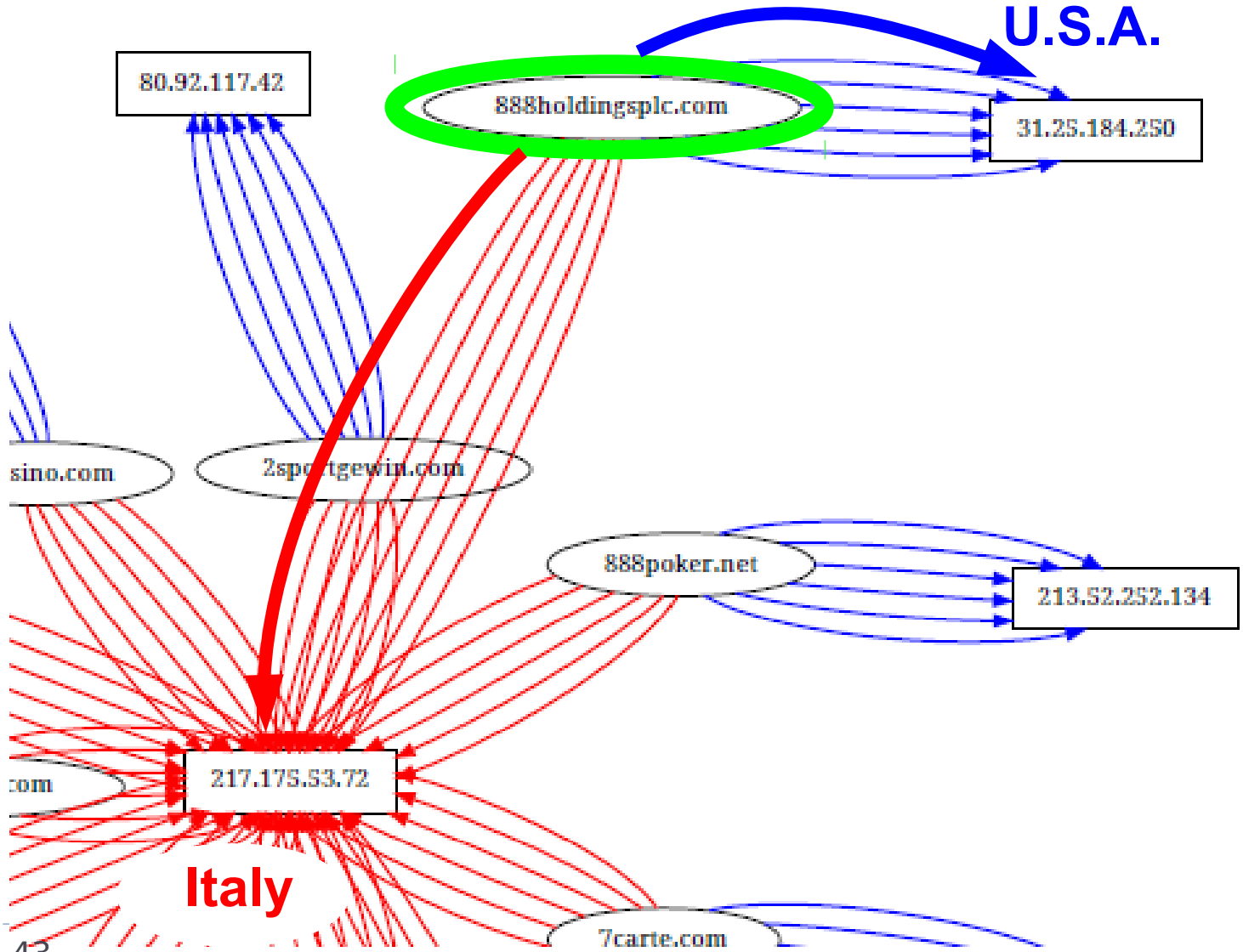
Evidences and visualization of DNS hijacking in Italy



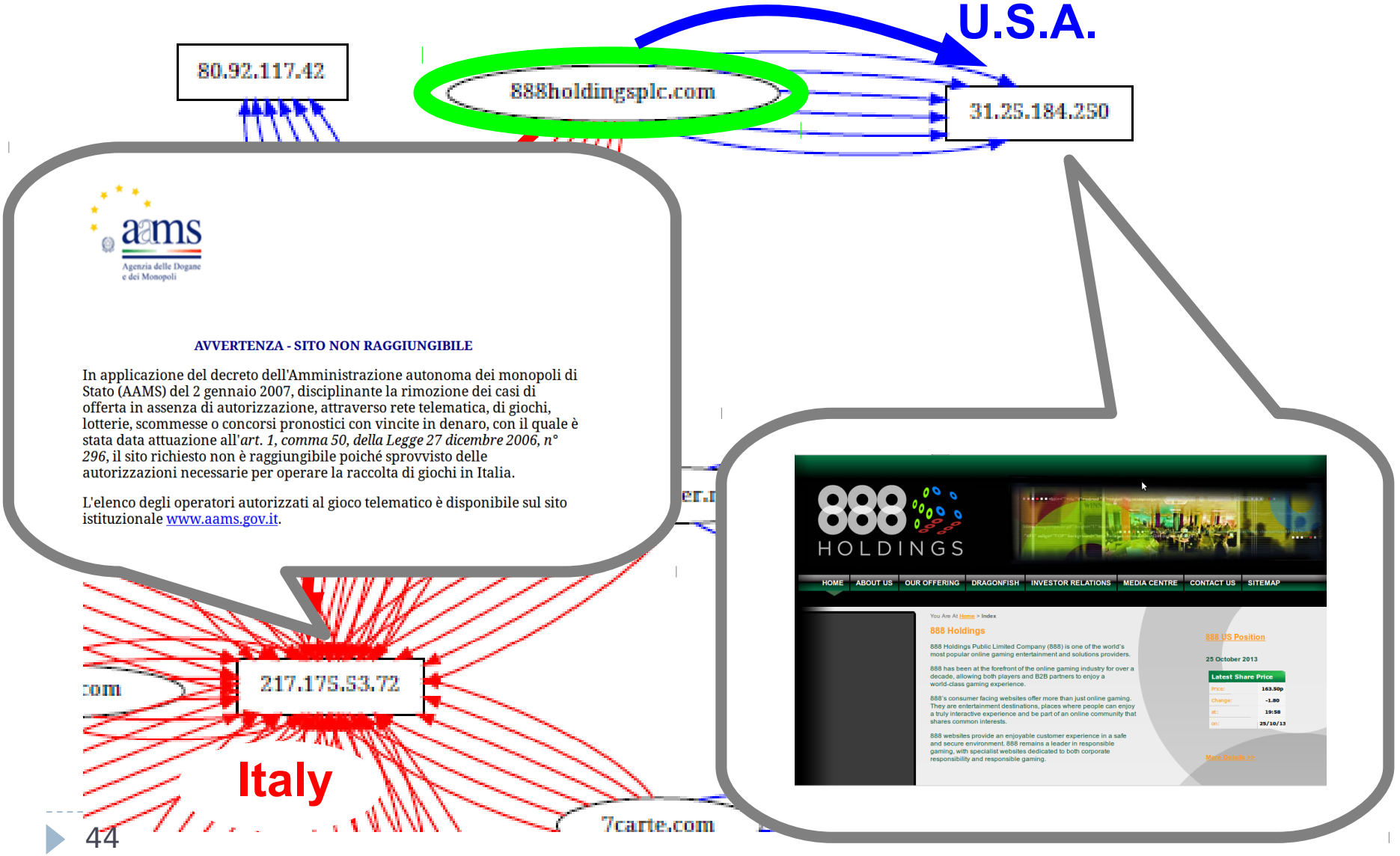
Evidences and visualization of DNS hijacking in Italy



Evidences and visualization of DNS hijacking in Italy



Evidences and visualization of DNS hijacking in Italy



Tests: TCP connection

- TCP reachability
  - **Collection**: a TCP handshake is initiated towards the target IP address, timeout, reset response and network errors are collected
  - **Check errors**: compare percentage of unreachability issues inside a country/ISP/AS against others
  - **Check RST**: compare percentage of RST received inside a country/ISP/AS against others

Tests: HTTP evidence

- HTTP content
  - **Collection**: HTTP GET of the target URL is requested to the target hostname (specifying the IP too);
    HTTP headers and downloaded content are saved
  - **Check errors**: compare percentage of unreachability issues (no content retrieved) inside a country/ISP/AS against others
  - **Check size**: compare average size inside a country/ISP/AS against a Ground Truth (selected reference country/ISP/AS)

Tests: HTTP evidence

cc	avgreachperc	avgunreachperc	span(m)
US	76.81	23.19	213.9
CN	35.23	64.77	212.8
JP	75.11	24.89	207.5
NZ	60.49	39.51	207.2
UY	71.16	28.84	209.3
BR	80.97	19.03	208
CA	83.98	16.02	207.1
RU	70.3	29.7	207.6
XX	77.96	22.04	207.3
JO	0	100	201.4
GB	71.85	28.15	206.7
HK	85.25	14.75	209.2
00	19.3	80.7	207.6
BD	35.69	64.31	206.9
IN	85.31	14.69	207.1
KR	79.21	20.79	205.3
AU	79.83	20.17	206.4
AR	76.04	23.96	205.9
CZ	85.09	14.91	209.1
TR	85.54	14.46	207.2
EC	85.2	14.8	208
SE	81.94	18.06	209.6

No response to HTTP request or response empty, grouped by country
"Expected" (CN, BD) and unexpected (NZ) countries
Currently under investigation!!!

Different html size retrieved for the same site from different countries
Block page is typically small
The size ratio may be a good censorship detector

URL	$< size_{PK} >$	$< size_{US} >$	Ratio
https://barenakedislam.wordpress.com	453.0	49095.63	0.01
http://www.internationalfreepresssociety.org	443.5	38085.32	0.01
http://ninjaproxy.com	342.45	14085.42	0.02
NinjaProxy.com	342.39	13154.06	0.03
http://www.similarsites.com	375.33	13701.44	0.03
http://www.youtube.com	4183.91	144177.2	0.03
http://www.freefacebookproxies.com/	9041.17	241485.33	0.04
http://friendlyatheist.com	7881.34	205294.23	0.04
http://www.loonwatch.com	2661.73	65075.19	0.04
http://www.sodahead.com	3575.67	73969.7	0.05
http://www.hotspotshield.com/	731.8	10789.91	0.07
http://face-of-muhammed.blogspot.com/	6208.7	85342.93	0.07
http://www.foxnews.com	4705.53	63425.26	0.07
http://www.buzzfeed.com	22097.93	287001.77	0.08

Tests: Tor censorship - collection

- Integrated test
  - **Collection**: DNS resolution of
    - Tor home page and mirrors
    - Tor-based circumvention techniques and tools pages
    - Tor overlay node list webpage
  - **Collection**: HTTP GET of
    - Tor home page and mirrors
    - Tor-based circumvention techniques and tools pages
    - Tor overlay node list webpage
  - **Collection**: TCP reachability of
    - Tor nodes (*relays*)
    - Tor indexing servers (Directory Authorities)

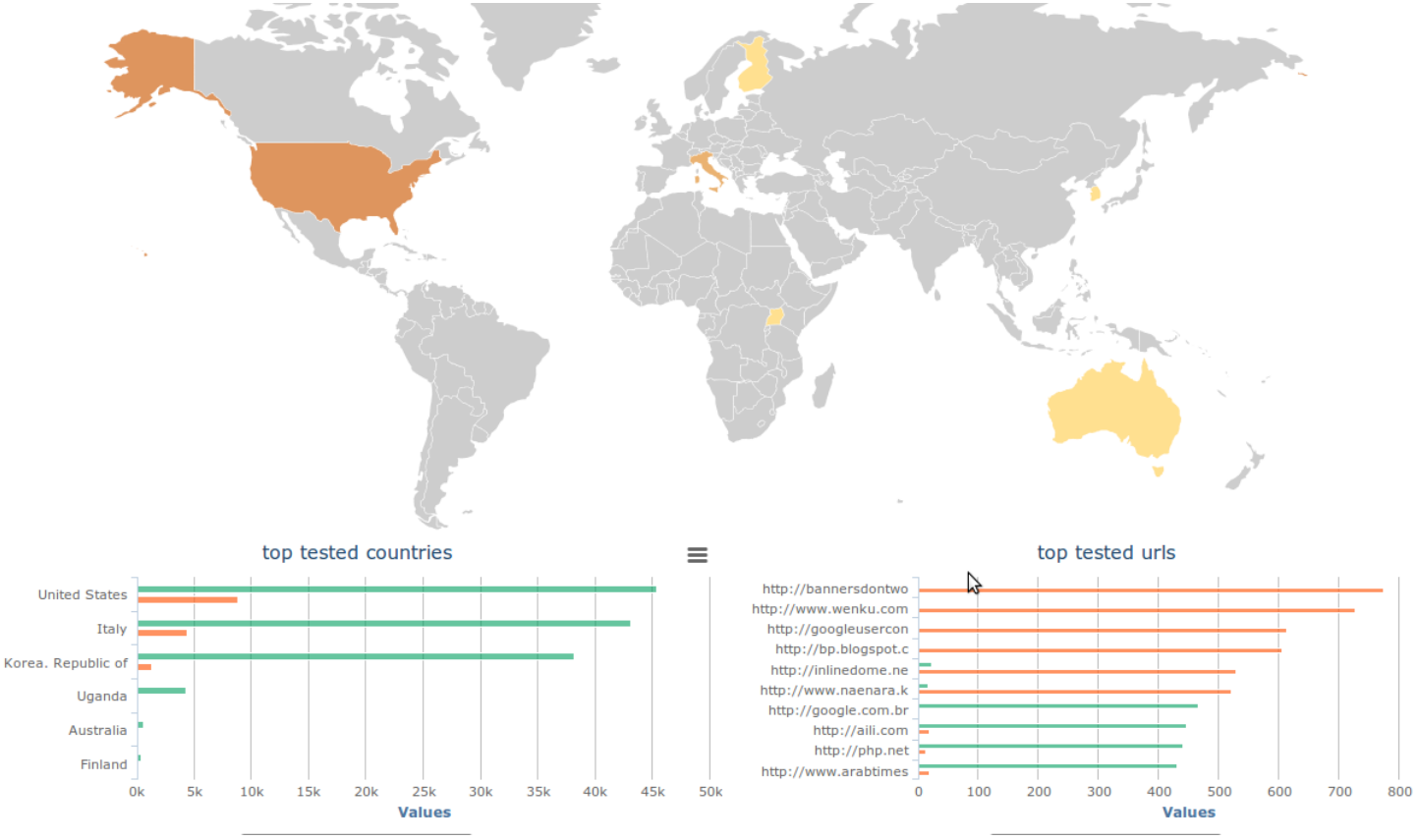
Tests: Tor censorship - checks*

- Integrated test Check:

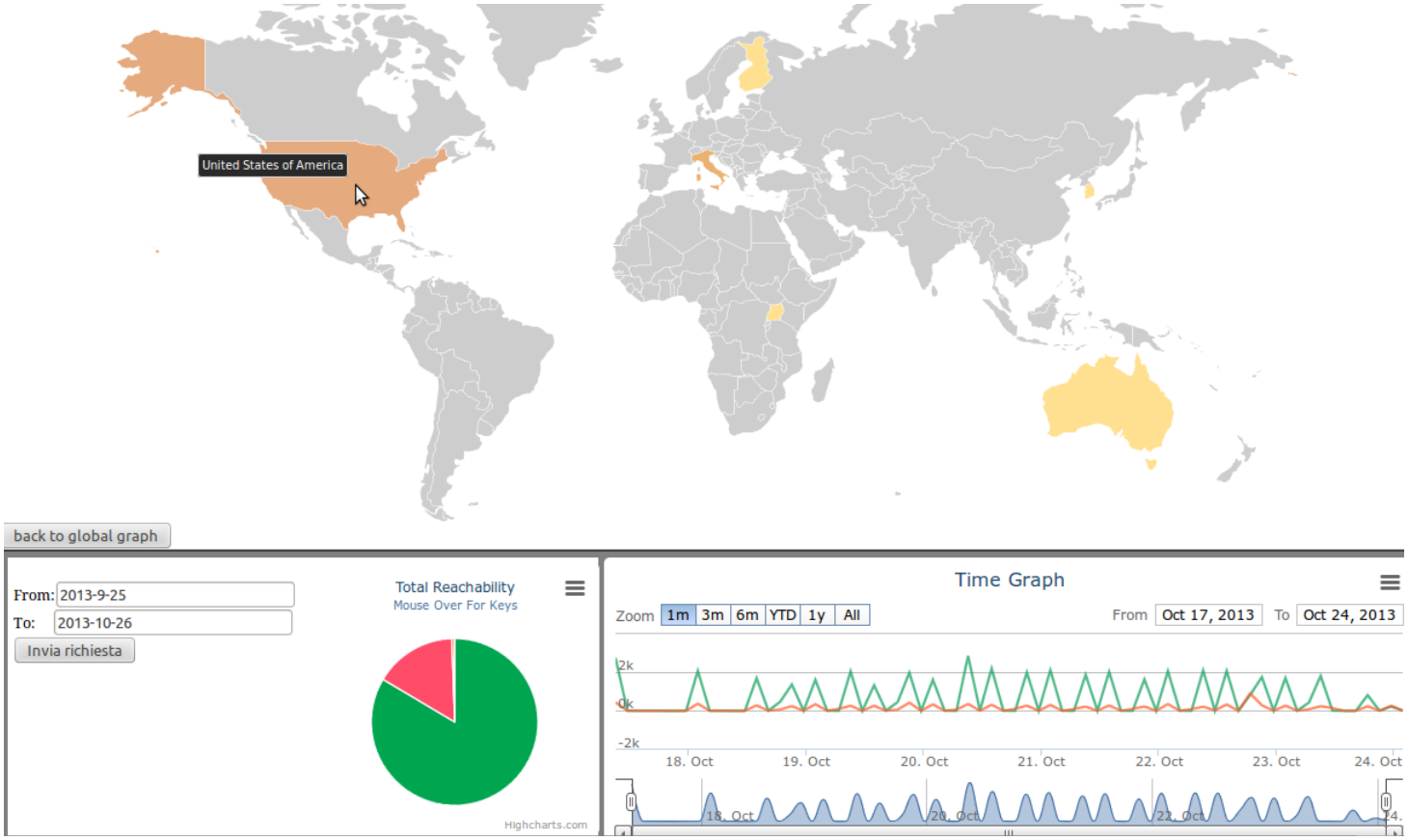
  DNS response for webpages is different from ground truth
  OR
  content of all webpages is unreachable
  OR
  X% of Directory Authorities is not **TCP-reachable**
  OR
  Y% of relays is not **TCP-reachable**
  THEN ***Tor is censored***

* variations of the method in: Philipp Winter, “Design Requirements for a Tor Censorship Analysis Tool”, Tor Tech Report 2013-02-001 February 6, 2013

UBICA monitoring interface (1/2)



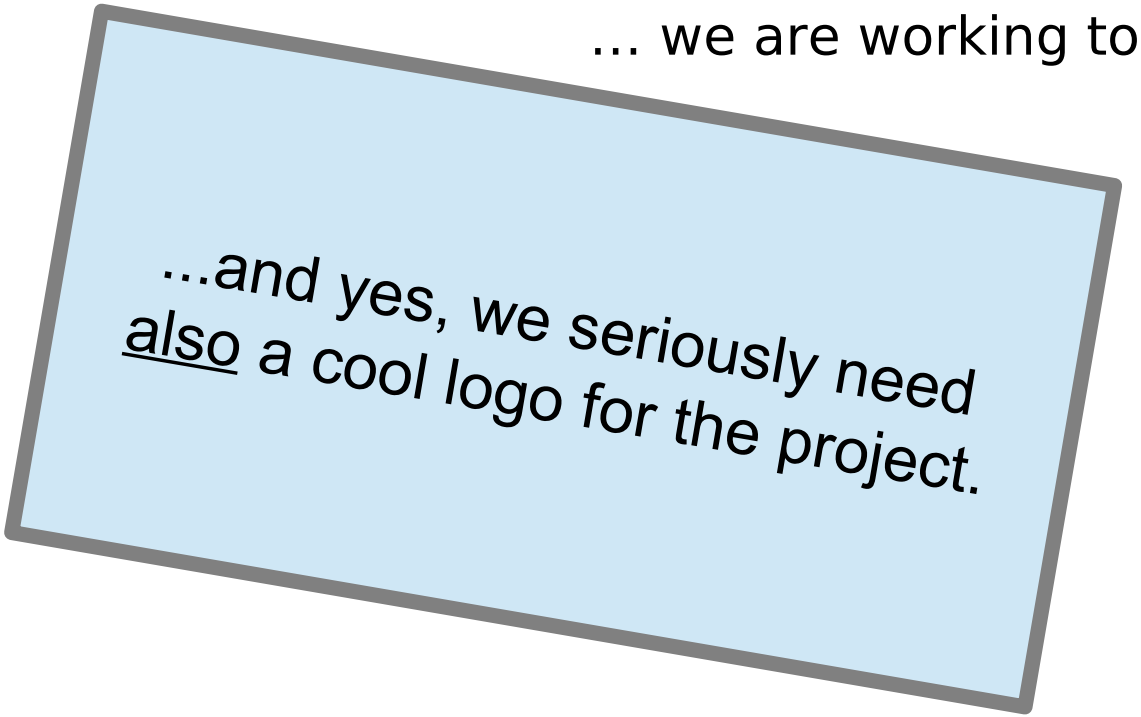
UBICA monitoring interface (2/2)



UBICA at the moment (yesterday night)

- Just a prototype: few tests, few vantage points, but has
  - Full-cycle automation
  - Actual data
  - Quasi-real-time reporting

... we are working to improve it!



References, credits

"Measuring & Circumventing Internet Censorship" – Nick Feamster, talk @ ETH ZISC Future Internet Security Workshop Oct 16th 2013 http://goo.gl/CHV5uJ

Philipp Winter, “Design Requirements for a Tor Censorship Analysis Tool “,Tor Tech Report 2013-02-001 February 6, 2013

Gill, Phillipa, et al. "Characterizing Censorship of Web Content Worldwide." 2013

Z. Nabi, "The anatomy of web censorship in pakistan," arXiv preprint arXiv:1307.1144, 2013.

Sfakianakis, Andreas, Elias Athanasopoulos, and Sotiris Ioannidis. "CensMon: A Web censorship monitor." FOCI’11: USENIX Workshop on Free and Open Communications on the Internet. 2011.

C. Anderson, "Dimming the internet: Detecting throttling as a mechanism of censorship in iran," Jun. 2013. [Online]. Available: http://arxiv.org/abs/1306.4361

Philipp Winter – selection of papers on Internet censorship
http://www.cs.kau.se/philwint/censorbib/

Thanks for your attention!

Any Questions?