TMA Workshop
Barcelona 24 April 2015

Monitoring Internet Censorship with UBICA

Giuseppe Aceto[1], Alessio Botta[1], Antonio Pescapè[1], Nick Feamster[2],
M. Faheem Awan[3], Tahir Ahmad[3], and Saad Qaisar[3]

[1] University of Napoli Federico II (Italy)
[2] Georgia Institute of Technology (GA, USA)
[3] NUST SEECS (Pakistan)

Outline

- Motivations
- UBICA platform
- Size Ratio test
- Case study
  - Pakistan
  - Korea
  - Italy
- Conclusions

Several ethical/political issues

- Little transparency (if any)
- Laws lag behind governments practice
- Risk of misuse / disproportionate use
- International relations / law implied
- Freedom of speech threatened

...

Several ethical/political issues

- Little transparency (if any)
- Laws lag behind governments practice
- Risk of misuse / disproportionate use
- International relations / law implied
- Freedom of speech threatened

...

NOT MY BUSINESS HERE

Internet Censorship tampers with protocols

Exploits *good-faith* early design choices, for

- DoS
- MITM

Requires *middle boxes* for surveillance and censorship enforcement.

Different stakeholders

Consequences for

- **users** as viewers or publishers

Different stakeholders

Consequences for

- **users** as viewers or publishers
- providers of censored **content/services**
- **access** and **transport** providers
  - deployment and management of middle boxes
  - circumvention traffic

Different stakeholders

Consequences for

- **users** as viewers or publishers
- providers of censored **content/services**
- **access** and **transport** providers
  - deployment and management of middle boxes
  - circumvention traffic

Moreover

- unexpected side effects*

* Anonymous "*The collateral damage of internet censorship by DNS injection.*" ACM SIGCOMM CCR 42.3 (2012).

M. L. Mueller. *China and global Internet governance*. In *Access contested: security, identity, and resistance in Asian cyberspace*, ed. R. J. Deibert, et al., pages 177–194, 2012.

Internet Censorship *Detection*

Analyzing network traffic

*to reveal **impairments***
in the access to content and services

*caused **by a third party***
(neither the client system nor the server hosting the resource or service) and

***not** justifiable as an **outage**.*

G. Aceto, A. Pescapè, "*Internet Censorship Detection: A Survey*", Computer Networks (2015)

Heterogeneous tools performing detection

- Herdict
  - **Web portal** to submit inaccessibility reports
- GreatFire
  - Monitoring the Great Firewall of China
  - No user involvement, **provides database** of results
- OONI
  - A platform for definition of censorship detection tests
  - Integrated with **Tor** anonymous overlay network
  - Integrated with **M-lab**

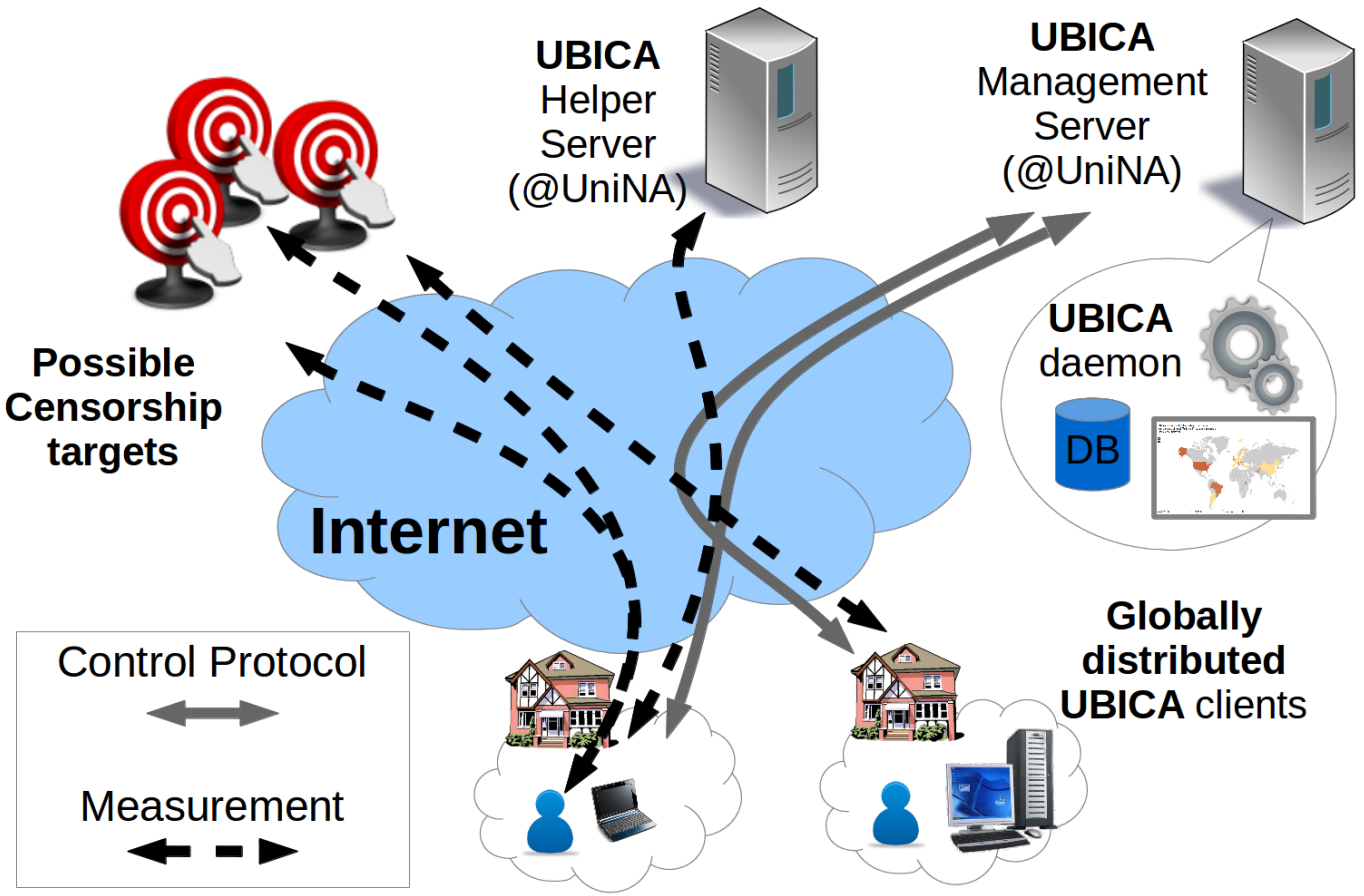
Heterogeneous tools performing detection

- Herdict
  - Web portal to submit inaccessibility reports
- GreatFire
  - Monitoring the Great Firewall of China
  - No user involvement, provides database of results
- OONI
  - A platform for definition of censorship detection tests
  - Integrated with Tor anonymous overlay network
  - Integrated with M-lab
- UBICA
  - More than just “layer 8” testing
  - Not only China, and crowdsourced
  - Geographically distributed agreement instead of GT

Outline

- Motivations
- **UBICA platform**
- Size Ratio test
- Case study
  - Pakistan
  - Korea
  - Italy
- Conclusions

User-Based Internet Censorship Analysis



Probe types

- Home gateways
  - BISMark platform (custom OpenWRT)
- GUI-client
  - Linux, Windows, OSX
- Headless client
  - Linux

DNS tests

- **Collection**: an A query for the target hostname is requested
  - to the default DNS server and
  - to a small number of open resolvers
- **Local Check**: returned IPs are checked against a list* of known *troublemakers*
  - non-publicly routable
  - failing to provide resource/service
  - explicit block pages

*highlighted by UBICA, manually validated

TCP tests

- **Collection**: a TCP handshake is initiated towards the target IP address (repeated up to 3 failures)

- **Local Check**: consider symptoms
  - Timeout (15s)
  - RST response
  - Network errors

HTTP tests

- **Collection**: GET of the target URL is requested
  - User Agent string is specified (taken from a list)
  - HTTP 30X redirects followed (max 50)
  - does *not* parse/execute JavaScript
  - HTTP headers and body are saved
- **Local check**: response HTTP code
- **Global check**: evaluate content size (*size ratio* below a threshold)

Related work: Jones, Ben, et al. "*Automated Detection and Fingerprinting of Censorship Block Pages.*" ACM Internet Measurement Conference, 2014.

Outline

- Motivations
- UBICA platform
- **Size Ratio test**
- Case study
  - Pakistan
  - Korea
  - Italy
- Conclusions

Size Ratio

For a given URL ***u***, and a set of probes ***P*** (single probe, ISP, Country), we define the *size ratio* ***r(u,P)*** as

$$r(u,P) = \frac{avg_P[s(u)]}{avg_{P\prime}[s(u)]}$$

where ***s(u)*** is the size of resource at *u*, and ***P'*** is the complement of *P* in the set of all probes from where *u* was tested

Threshold on Size Ratio

Goal:

Choose threshold that maximizes
a reliability* metric for detection algorithm

Ground Truth is needed

for training and validation

*defined later on

Threshold setting - methodology

Ground Truth: classification by manual inspection

Preprocess filters: (part of the detection algorithm)
- at least 3 different ISPs per target
- at least 3 measurements per target

Dataset:
- 19 different countries
- avg ~5 different ISPs per target
- 3.6K measurements, randomly chosen
- 2 weeks long window

Evaluation: Leave-One-Out cross-validation

Agorithm performance

Classification metrics adopted for evaluation

$$Prec = \frac{TP}{TP + FP} \qquad Rec = \frac{TP}{TP + FN}$$

Agorithm performance

Classification metrics adopted for evaluation

$$Prec = \frac{TP}{TP + FP} \qquad Rec = \frac{TP}{TP + FN}$$

$$NPV = \frac{TN}{TN + FN} \qquad TNR = \frac{TN}{TN + FP}$$

Agorithm performance

Classification metrics adopted for evaluation

$$Prec = \frac{TP}{TP + FP} \qquad Rec = \frac{TP}{TP + FN}$$

$$NPV = \frac{TN}{TN + FN} \qquad TNR = \frac{TN}{TN + FP}$$

Summary metrics:

F1

F2

$$F_\beta = (1 + \beta^2) \cdot \frac{Prec \cdot Rec}{(\beta^2 \cdot Prec) + Rec}$$

Agorithm performance

Classification metrics adopted for evaluation

$$Prec = \frac{TP}{TP + FP}$$

$$Rec = \frac{TP}{TP + FN}$$

$$NPV = \frac{TN}{TN + FN}$$

$$TNR = \frac{TN}{TN + FP}$$

Summary metrics:

F1

F2

$$F_\beta = (1 + \beta^2) \cdot \frac{Prec \cdot Rec}{(\beta^2 \cdot Prec) + Rec}$$

Informedness

$$Info = Rec + TNR - 1$$

Agorithm performance

Classification metrics adopted for evaluation

$$Prec = \frac{TP}{TP + FP} \qquad Rec = \frac{TP}{TP + FN}$$

$$NPV = \frac{TN}{TN + FN} \qquad TNR = \frac{TN}{TN + FP}$$

Summary metrics:

F1

F2

$$F_\beta = (1 + \beta^2) \cdot \frac{Prec \cdot Rec}{(\beta^2 \cdot Prec) + Rec}$$

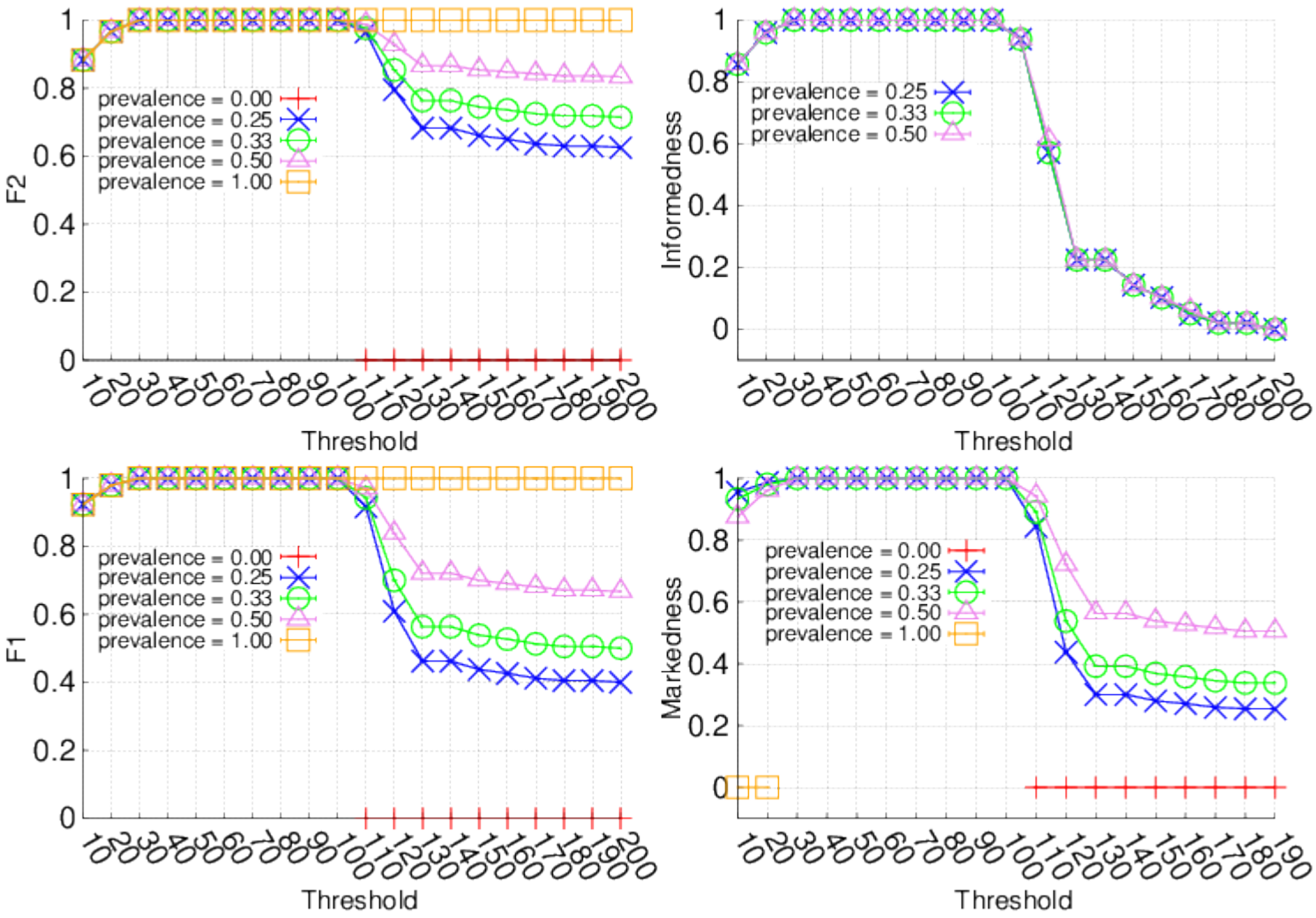
Informedness

$$Info = Rec + TNR - 1$$

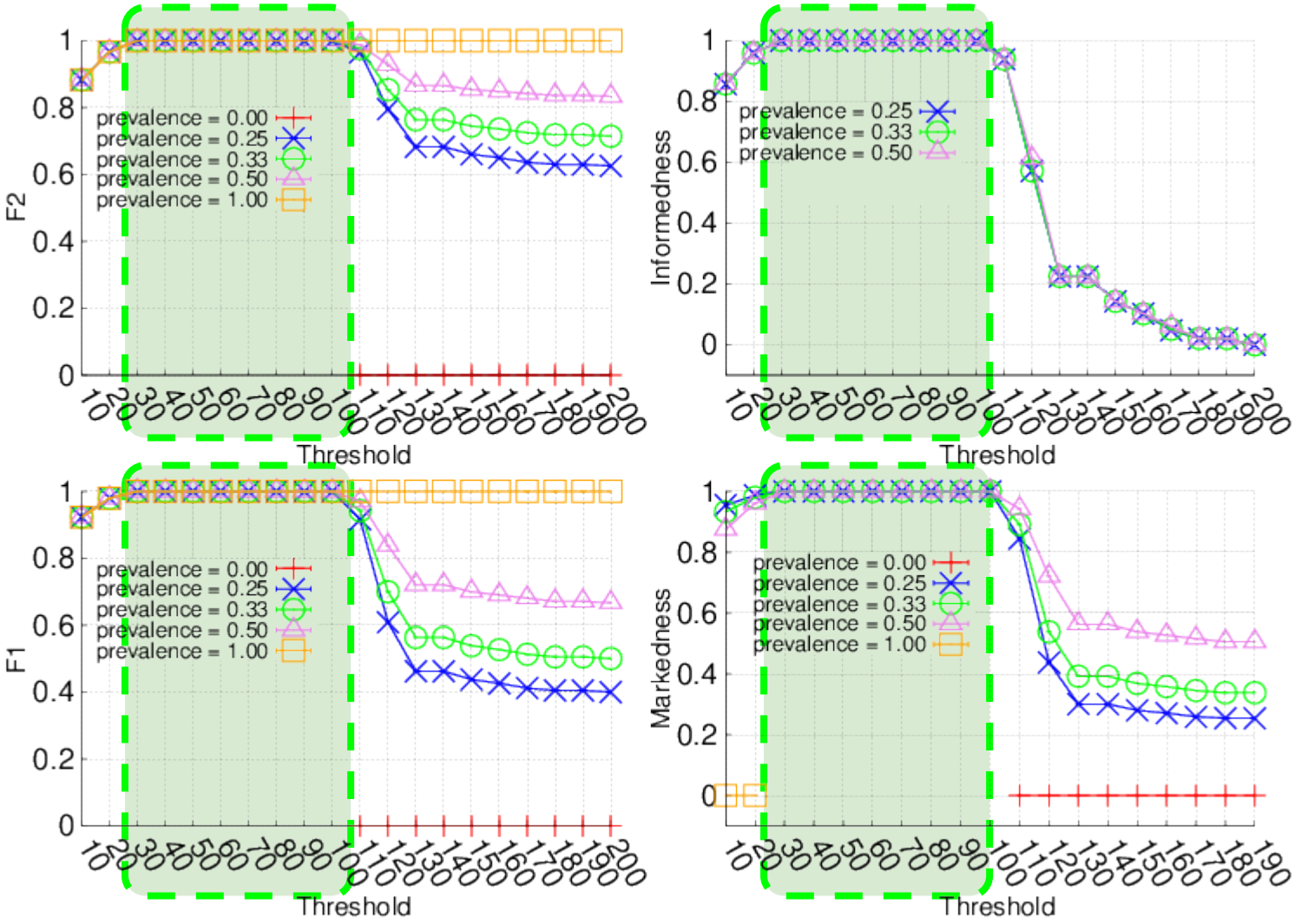
Markedness

$$Mark = Prec + NPV - 1$$

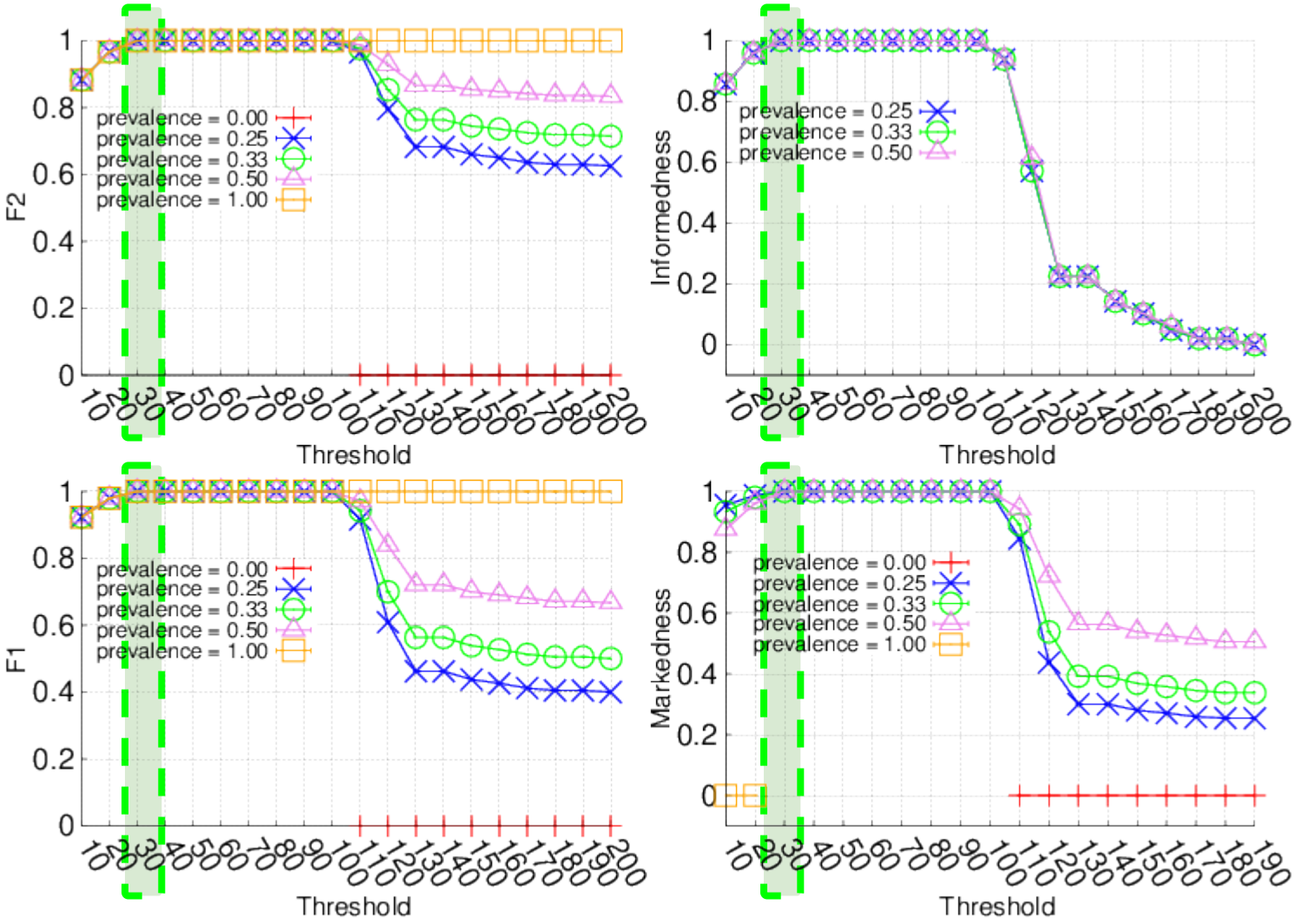
Threshold setting and prevalence



Robust to prevalence



Best threshold = 30%



Outline

- Motivations
- UBICA platform
- Size Ratio test
- **Case study**
  - Pakistan
  - Korea
  - Italy
- Conclusions

Experimental setup

More than 200 probes

- 47 clients with GUI
  (run by volunteers both in Italy and abroad);
- 188 headless clients
  (of which 19 run by volunteers worldwide
  and 169 in PlanetLab nodes);
- 16 BISmark home routers run by volunteers
  (mostly in Pakistan)

probes running from 31 different countries

Targets

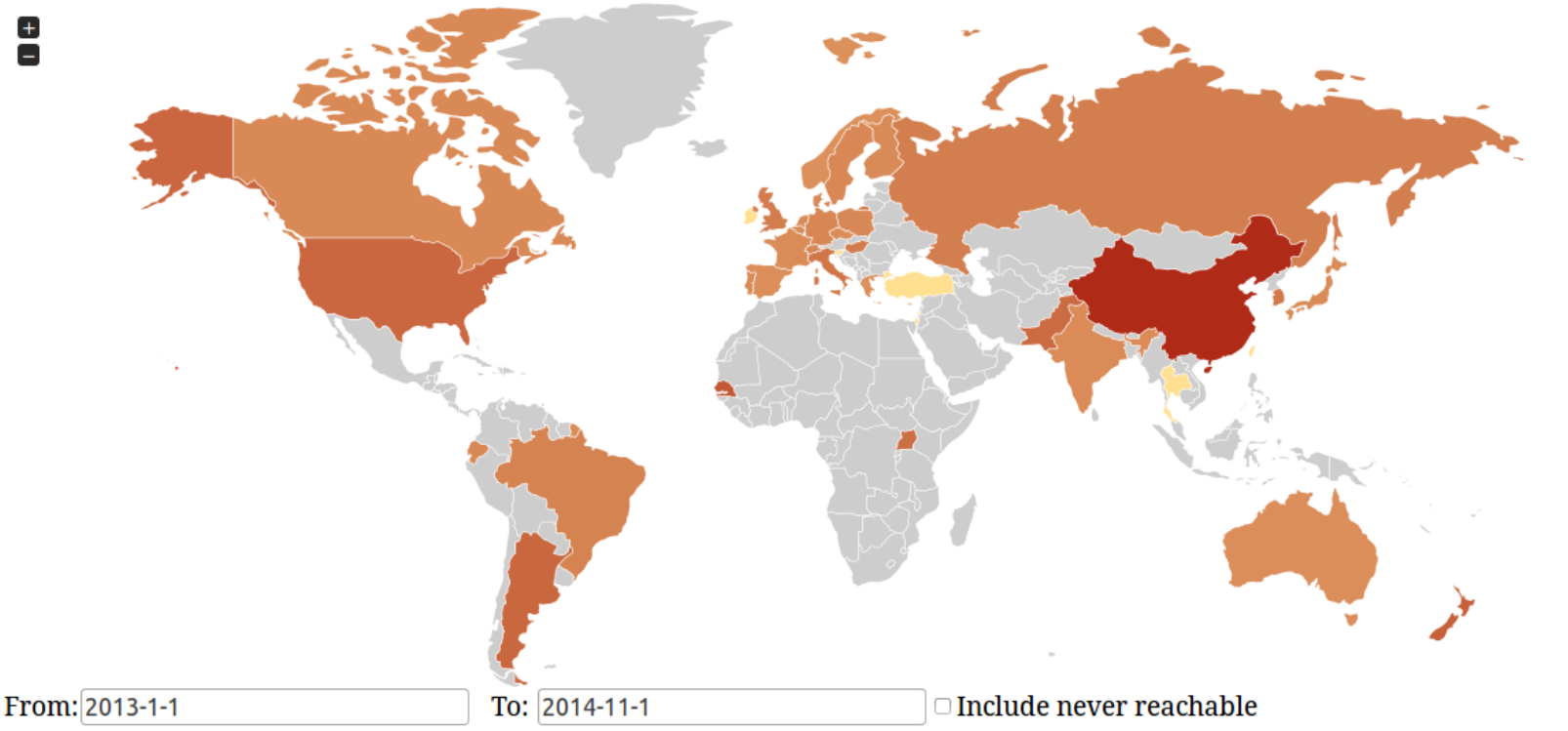
The target lists for each country included

- *Herdict* URL list for the country
- a list of worldwide top accessed websites
- and URLs suggested by local volunteers

testing more than 16K different targets
(~ 15K different domains)

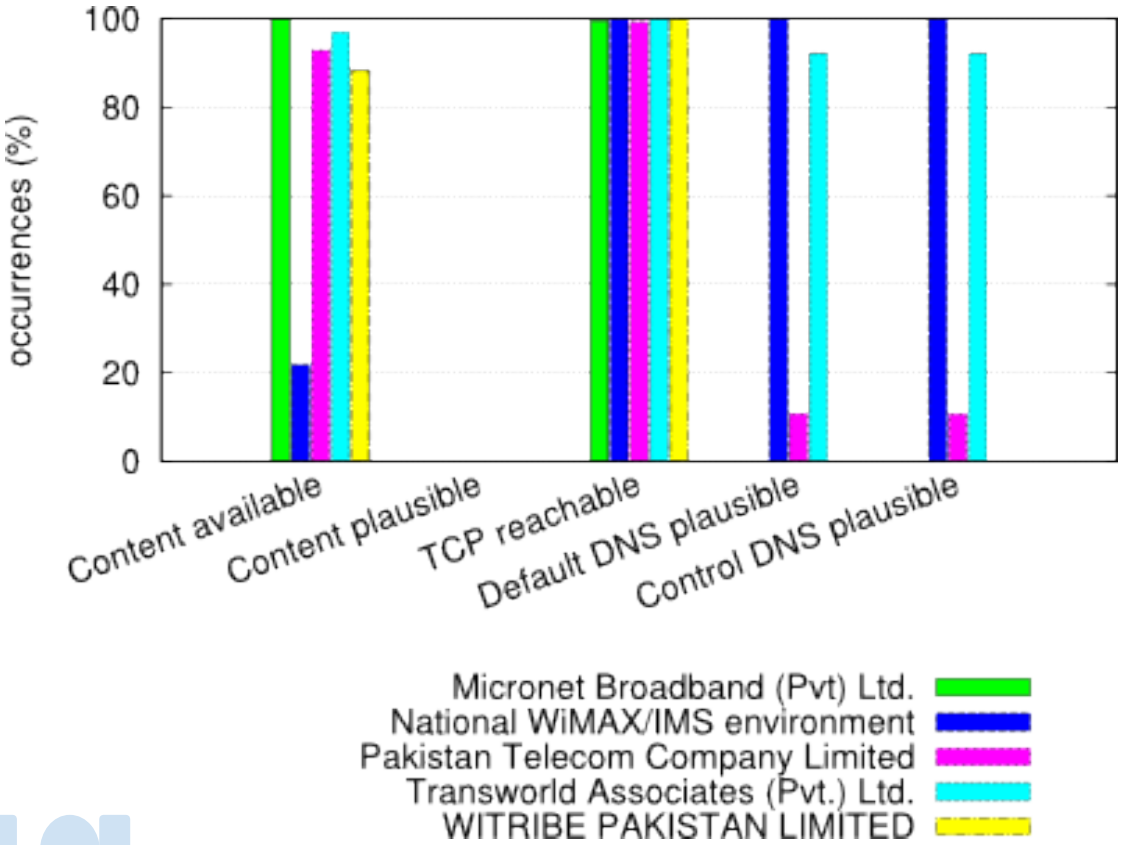
timespan of 4 months

Report interface (detail:global map)



Pakistan: YouTube

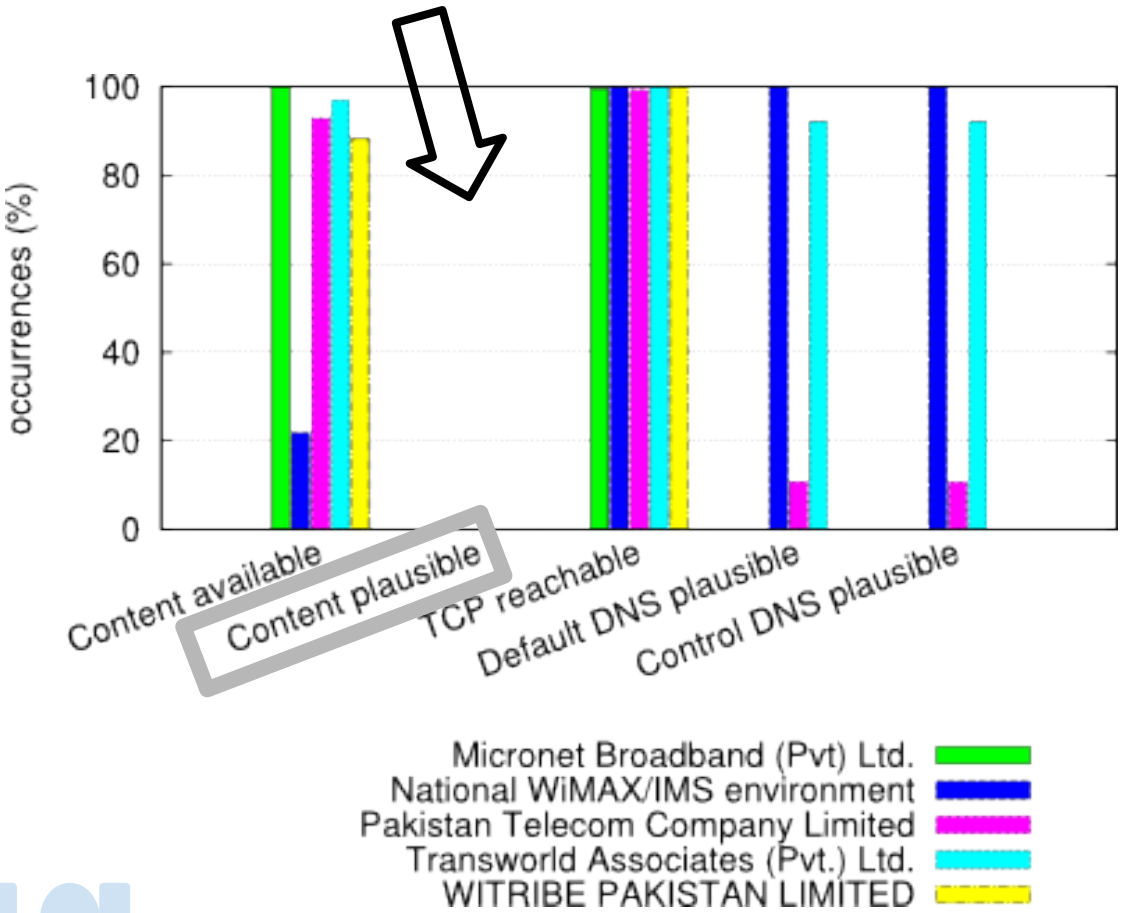
techniques overview



Pakistan: YouTube

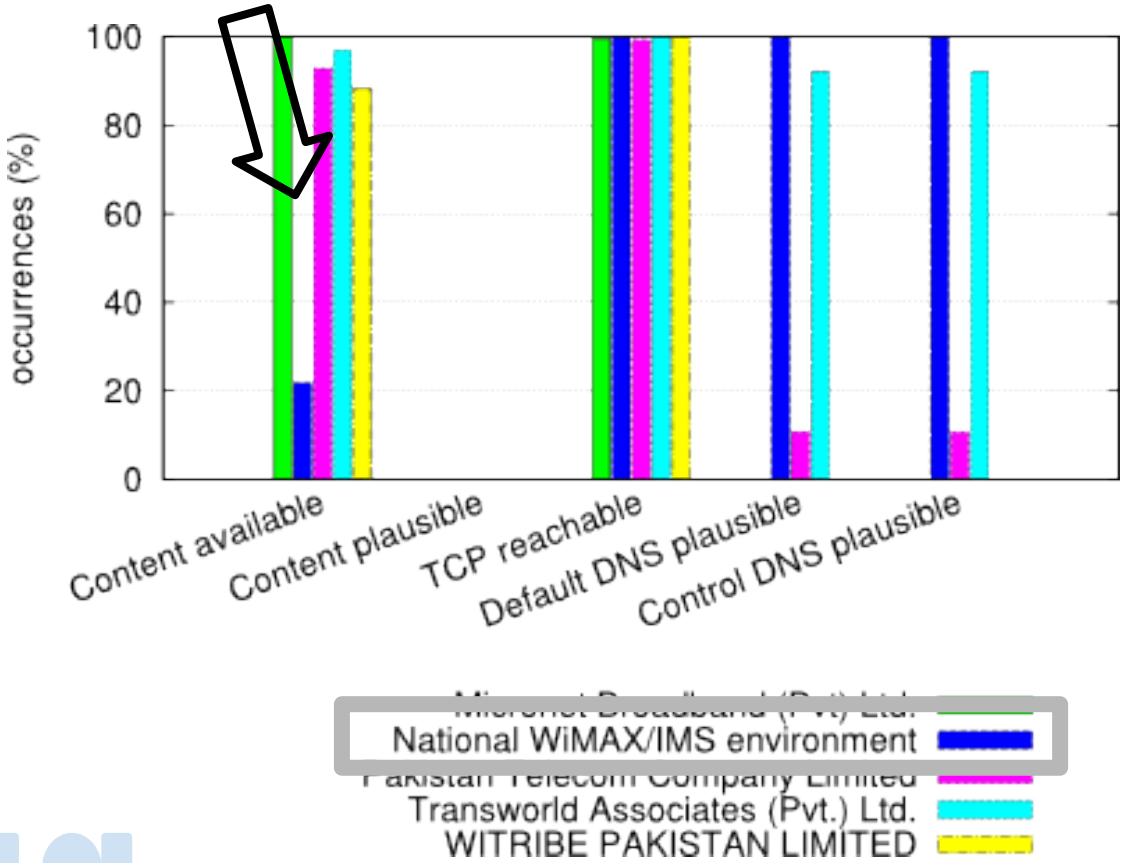
for all ISPs
Content never plausible

techniques overview



Pakistan: YouTube

techniques overview

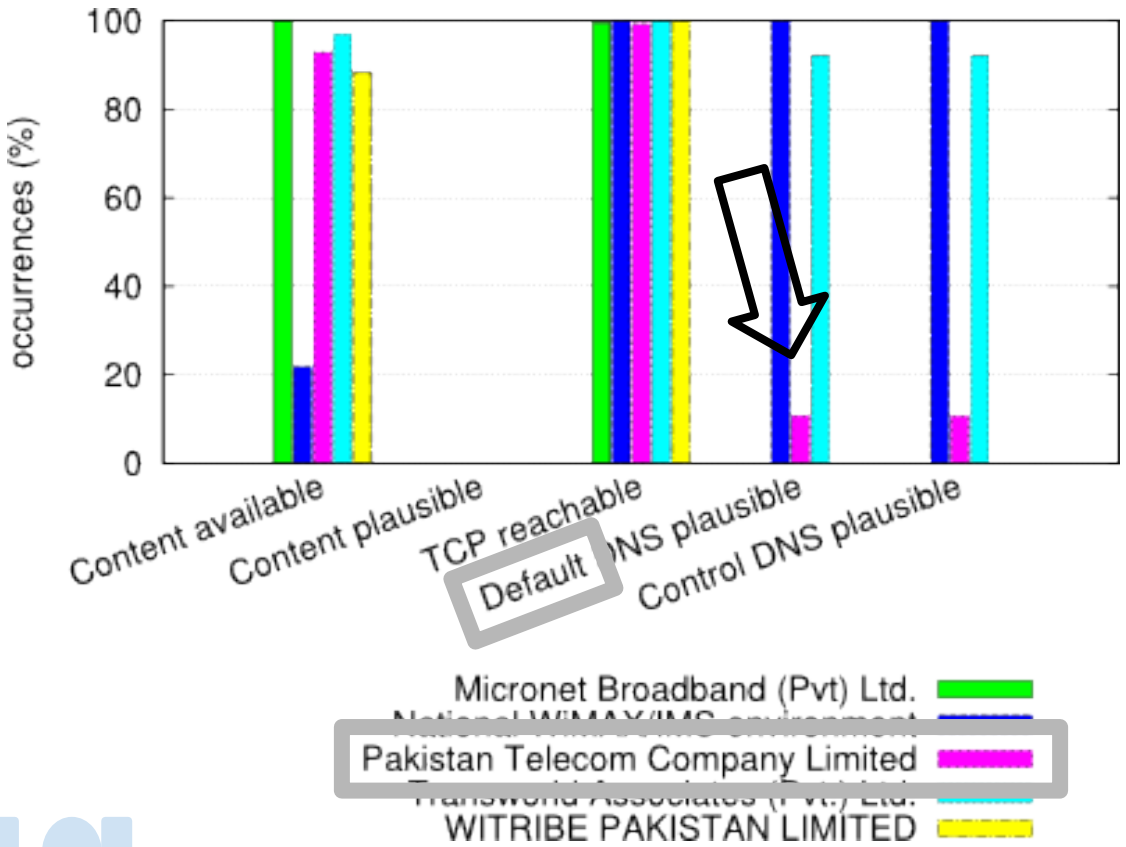


for all ISPs
Content never plausible

for WiMAX
Content mostly unavailable

Pakistan: YouTube

techniques overview



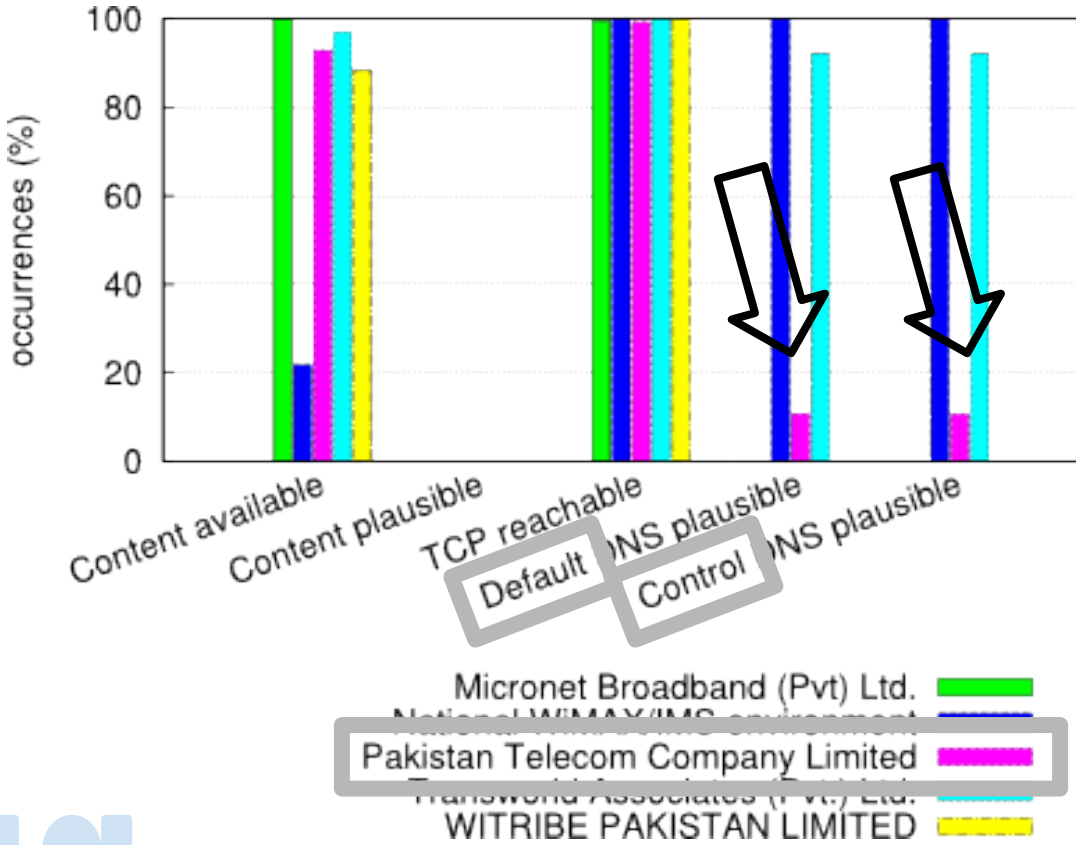
for all ISPs
Content never plausible

for WiMAX
Content mostly unavailable

for PTCL
DNS rarely plausible

Pakistan: YouTube

techniques overview



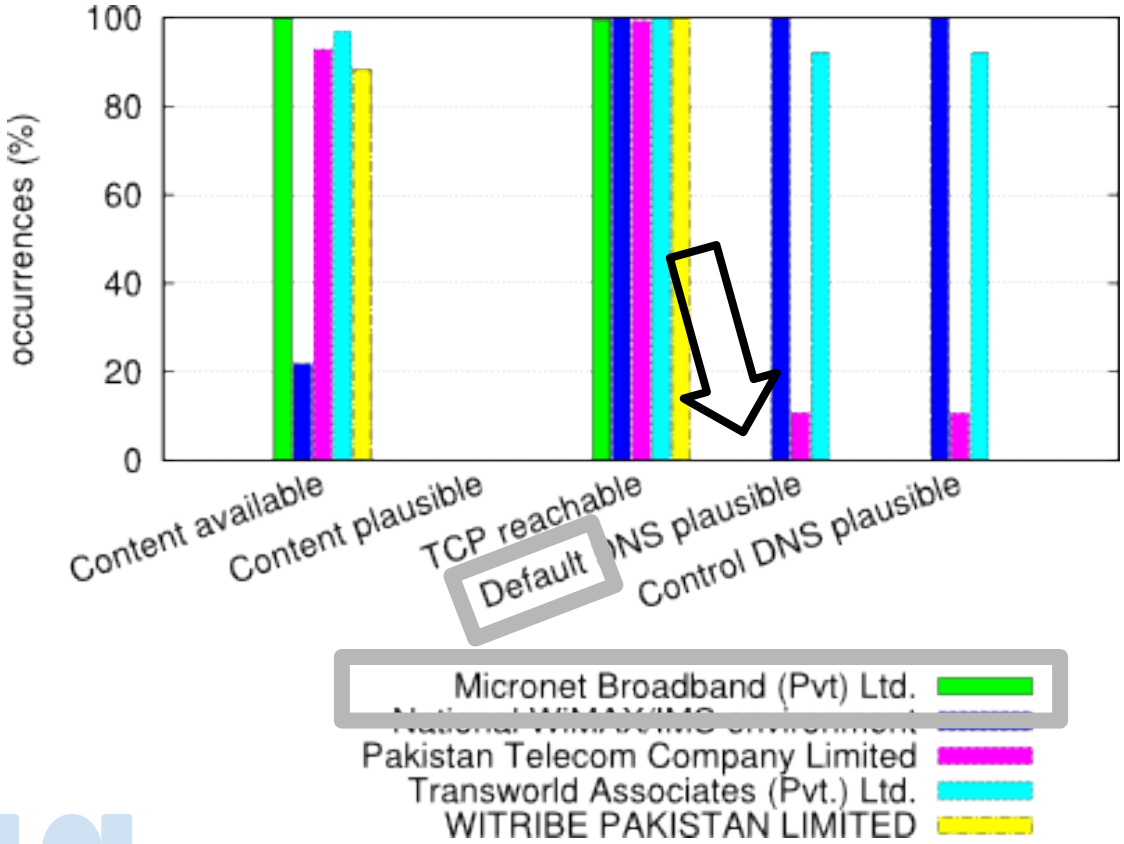
for all ISPs
Content never plausible

for WiMAX
Content mostly unavailable

for PTCL
DNS rarely plausible
technique: **injection**

Pakistan: YouTube

techniques overview



for all ISPs
Content never plausible

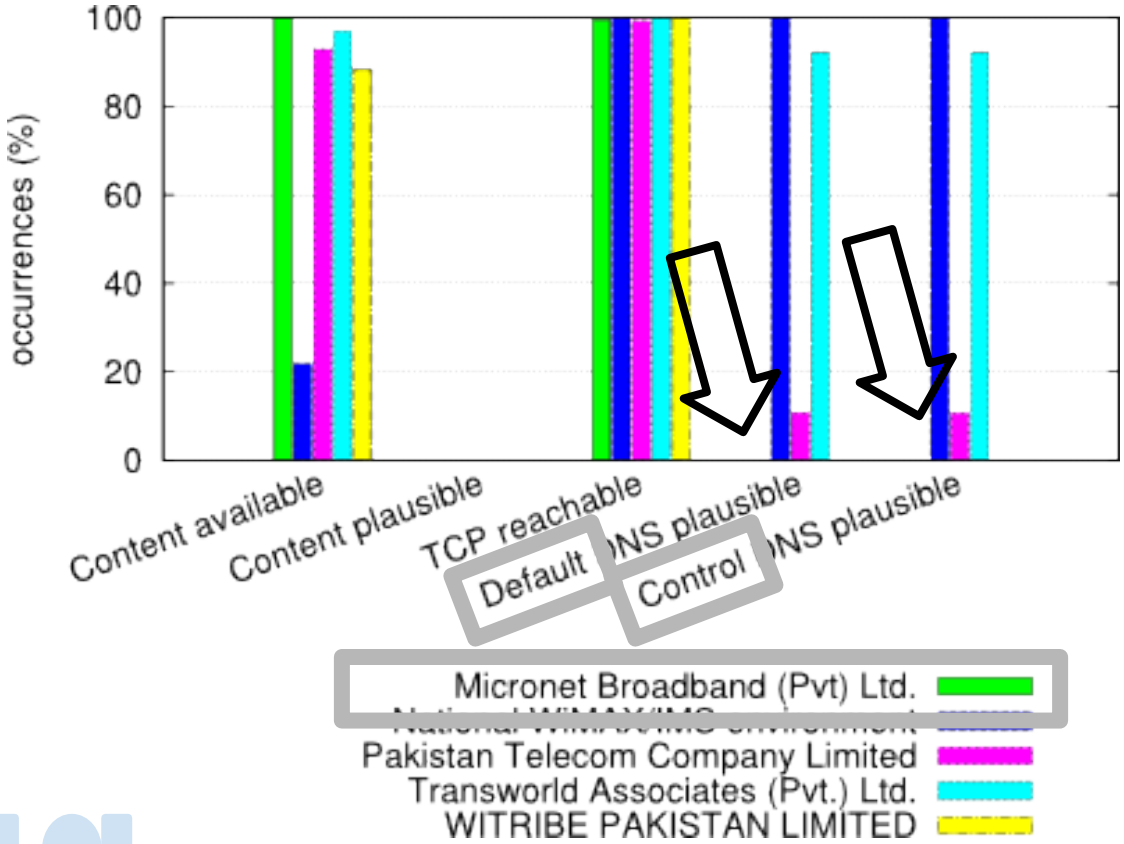
for WiMAX
Content mostly unavailable

for PTCL
DNS rarely plausible
technique: **injection**

for Micronet
DNS never plausible

Pakistan: YouTube

techniques overview



for all ISPs
Content never plausible

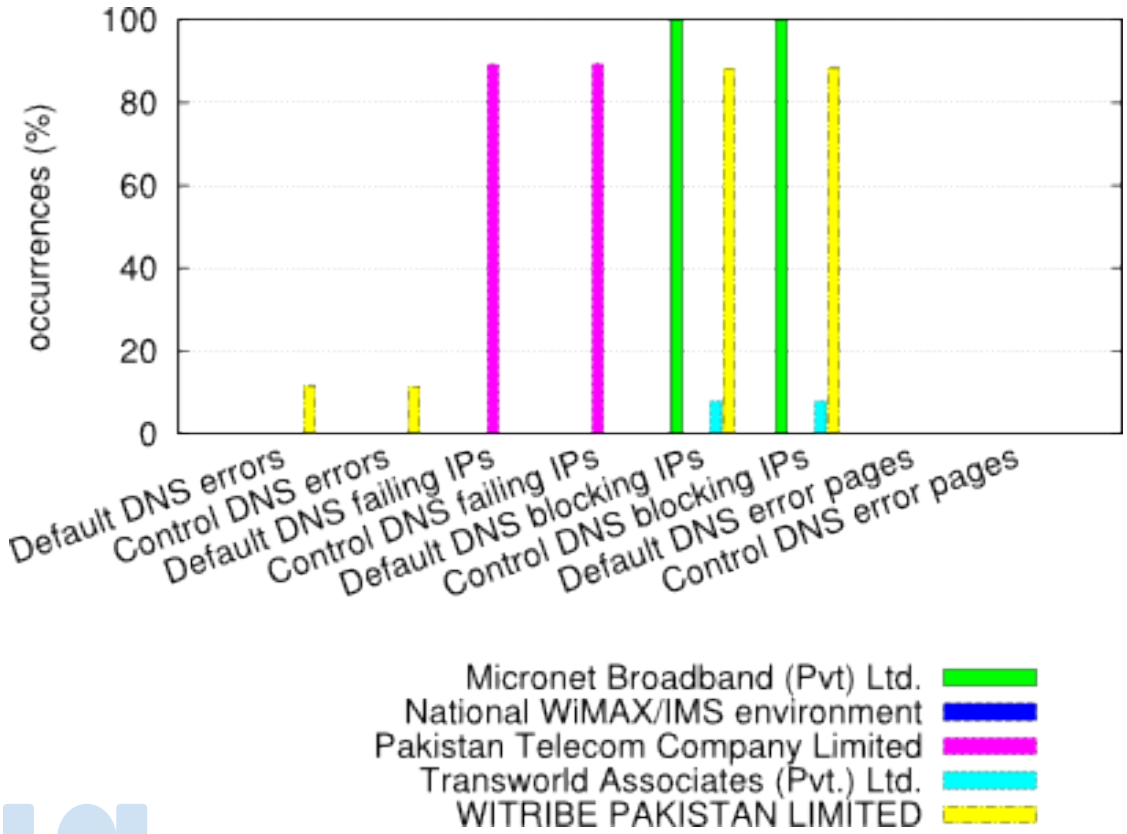
for WiMAX
Content mostly unavailable

for PTCL
DNS rarely plausible
technique: **injection**

for Micronet
DNS never plausible
technique: **injection**

Pakistan: YouTube

DNS analysis detail



for all ISPs
Content never plausible

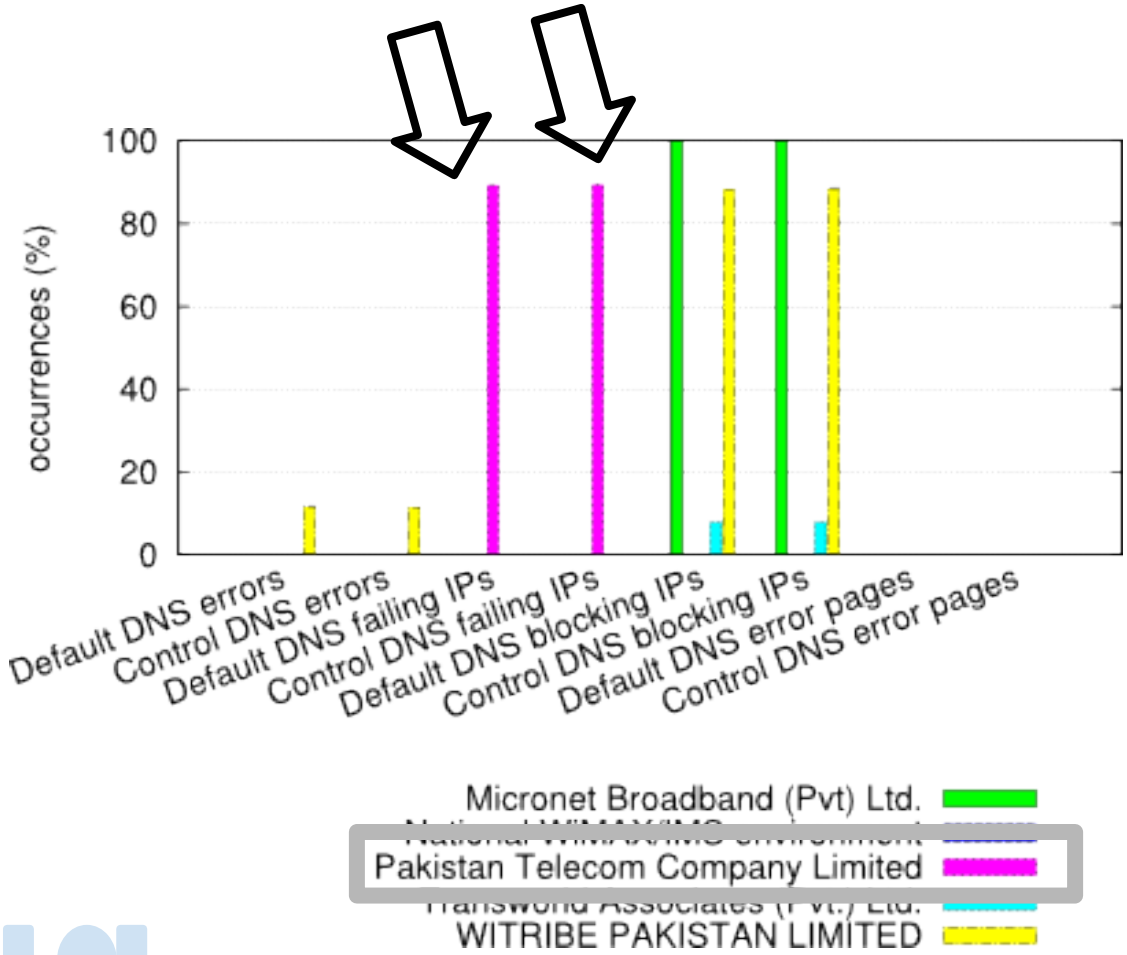
for WiMAX
Content mostly unavailable

for PTCL
DNS rarely plausible
technique: **injection**

for Micronet
DNS never plausible
technique: **injection**

Pakistan: YouTube

DNS analysis detail



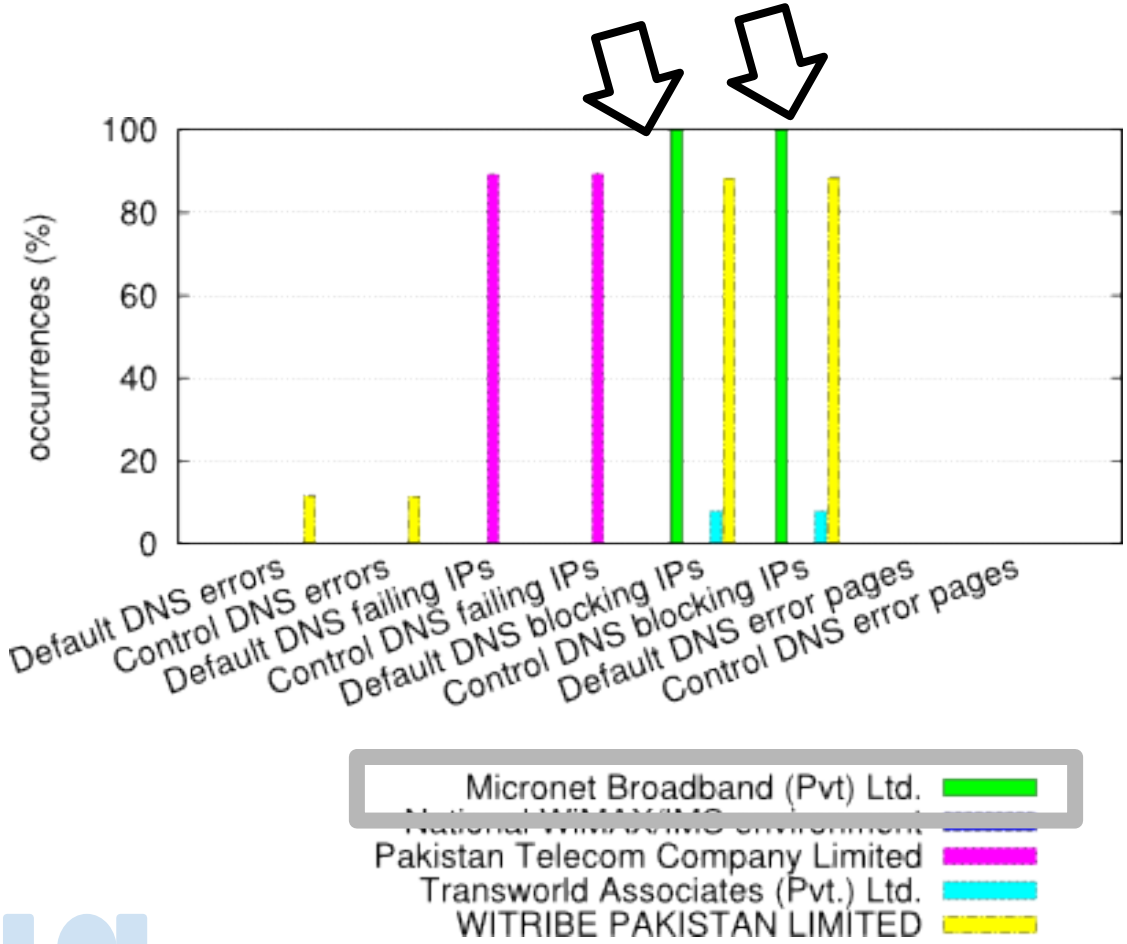
for all ISPs
Content never plausible

for WiMAX
Content mostly unavailable

for PTCL
DNS rarely plausible
technique: **injection**
symptom: **failing IP**

for Micronet
DNS never plausible
technique: **injection**

Pakistan: YouTube



for all ISPs
Content never plausible

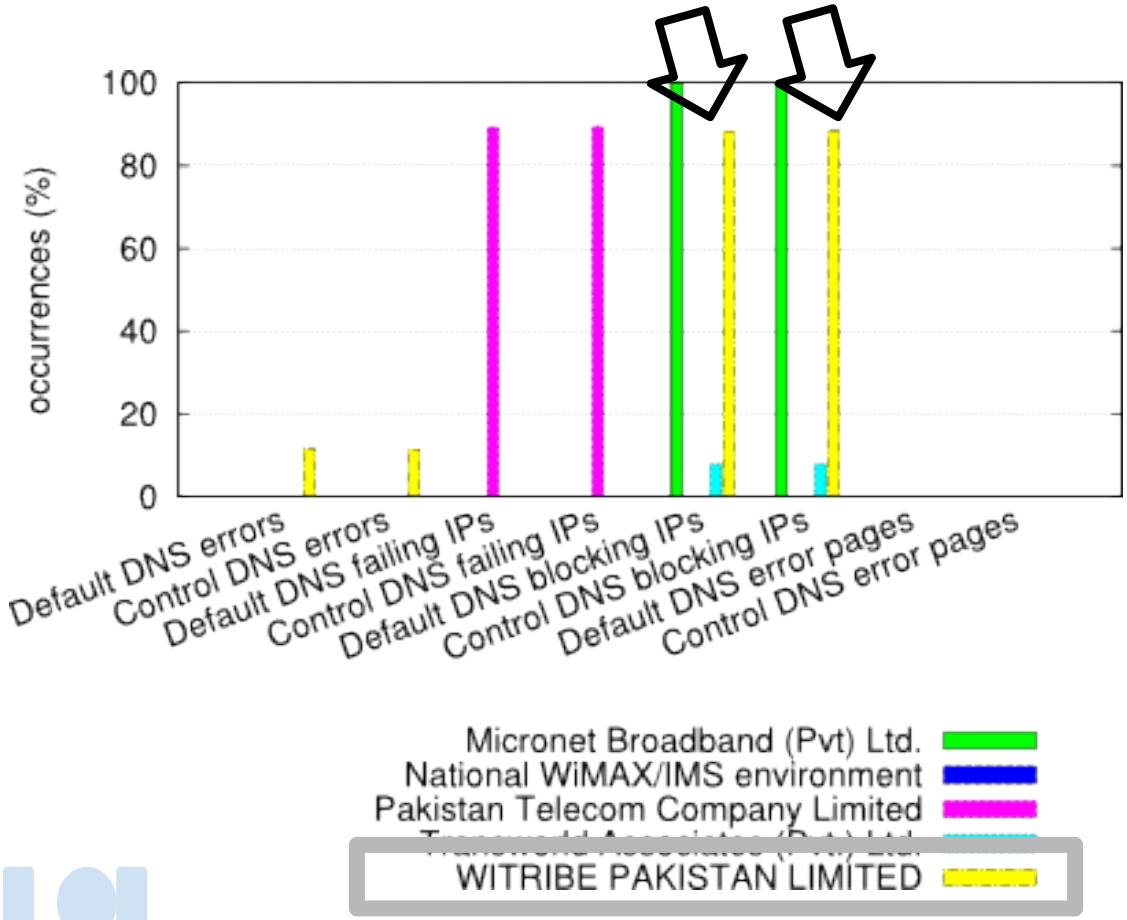
for WiMAX
Content mostly unavailable

for PTCL
DNS rarely plausible
technique: **injection**
symptom: **failing IP**

for Micronet
DNS never plausible
technique: **injection**
symptom: **block page**

Pakistan: YouTube

DNS analysis detail



for all ISPs
Content never plausible

for WiMAX
Content mostly unavailable

for PTCL
DNS rarely plausible
technique: **injection**
symptom: **failing IP**

for Micronet
DNS never plausible
technique: **injection**
symptom: **block page**

for WITRIBE
DNS never plausible
technique: **injection**
symptom: **block page**

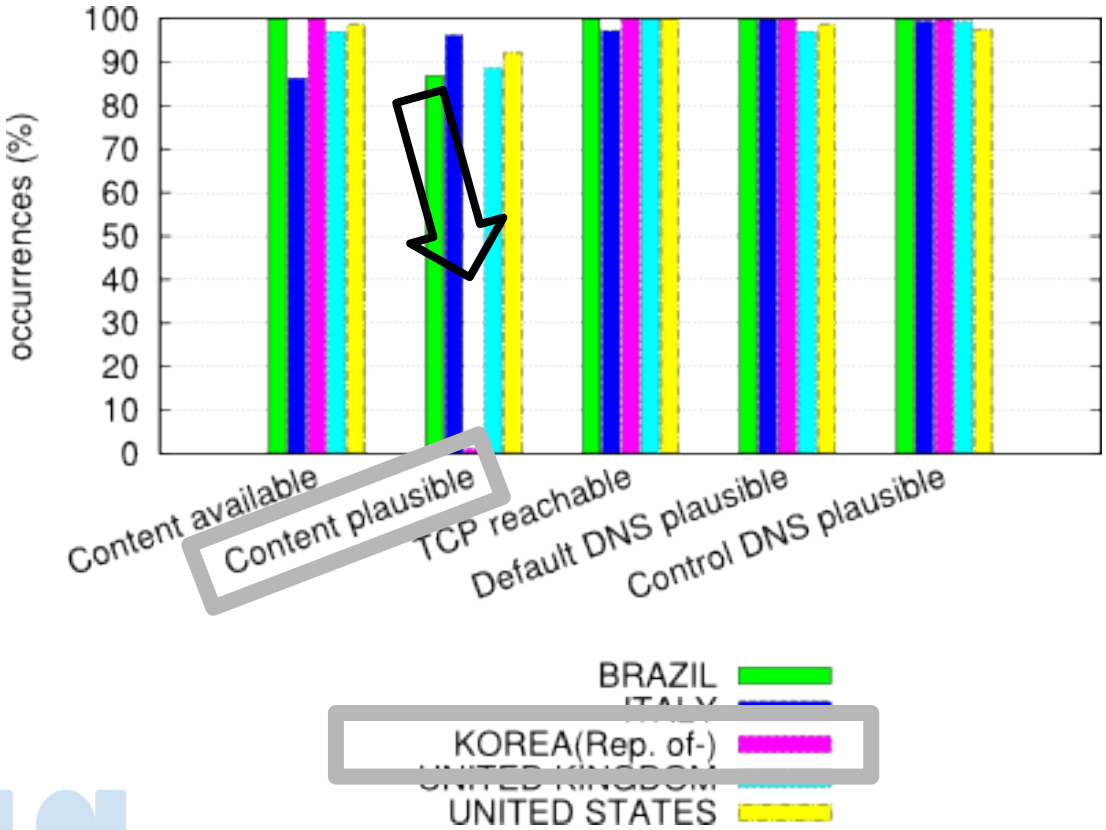
Pakistan: YouTube

UBICA responses

Micronet Broadband	DNS injection - send to *blocking page*
National WiMAX	HTTP tampering
PTCL	DNS injection - send to *failing IP*
WITRIBE	DNS injection - send to *blocking page*

Rep. of Korea: porn websites

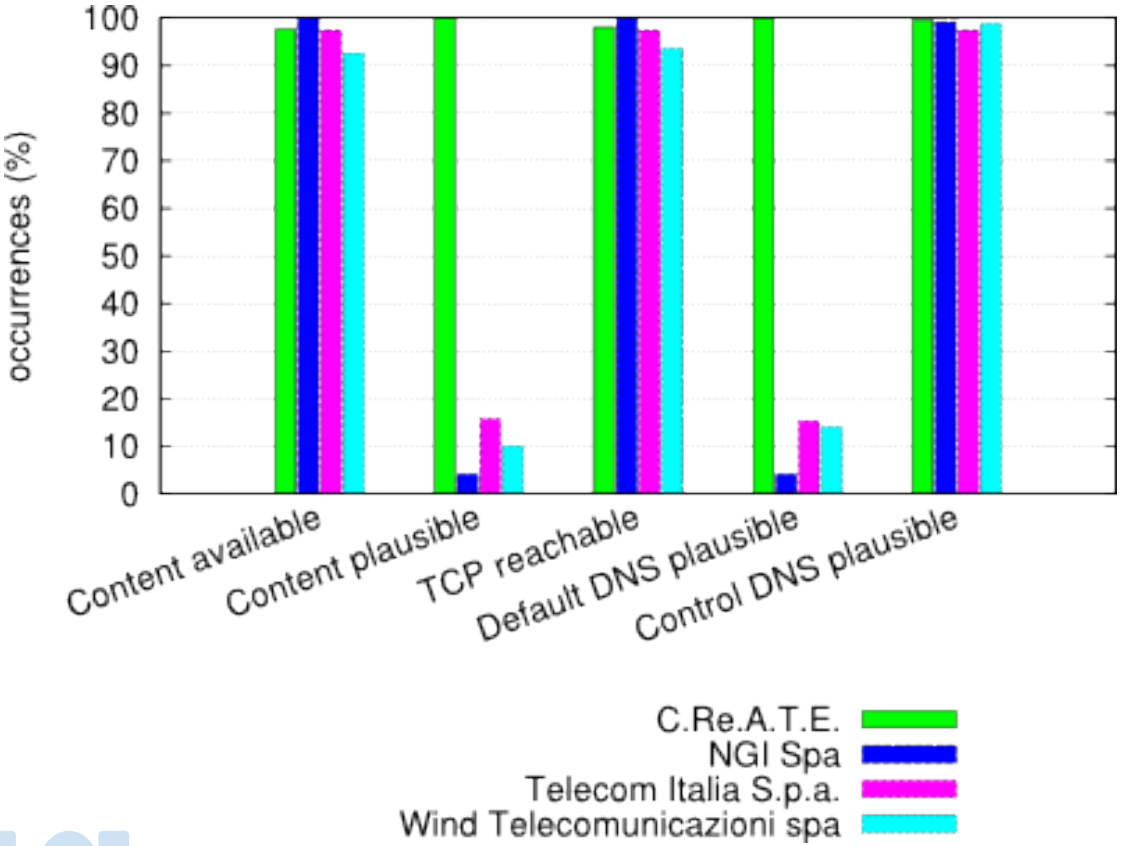
techniques overview inter-countries



No **Content** plausible:
HTTP tampering

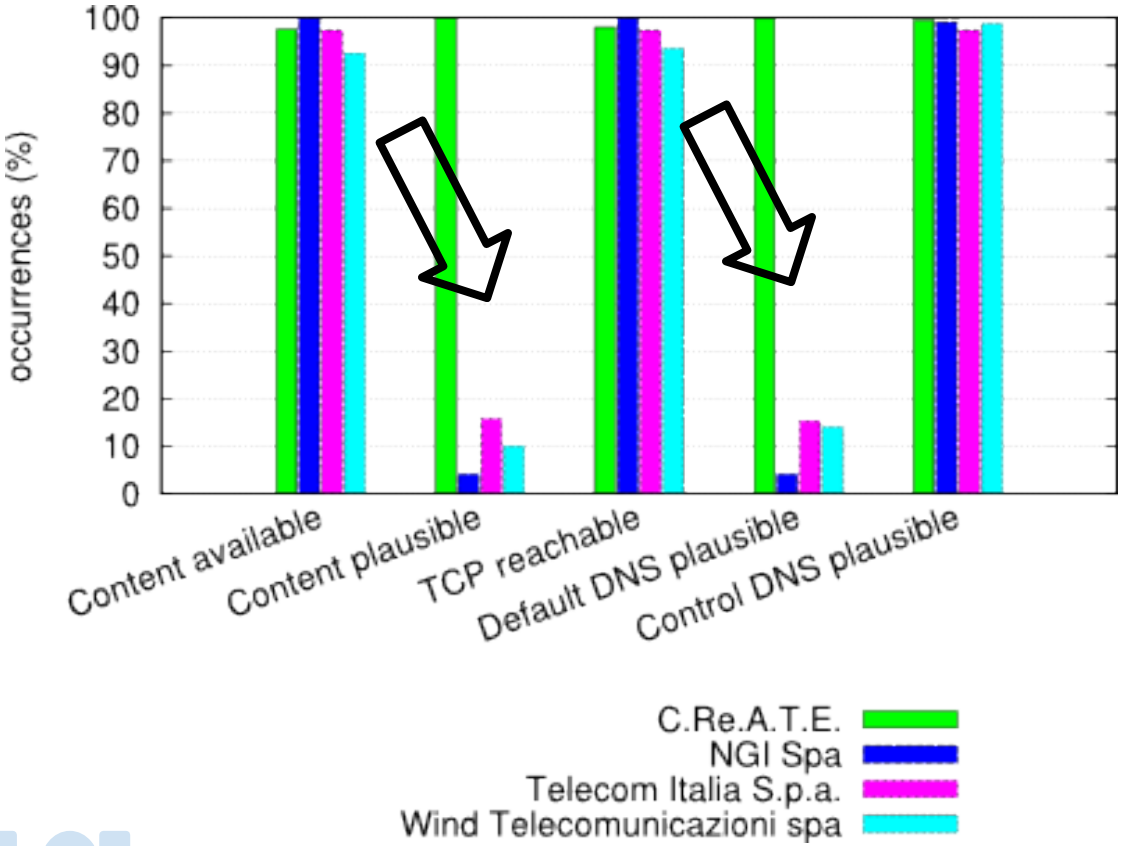
Italy: gaming and betting websites

bet365.com - techniques overview



Italy: gaming and betting websites

bet365.com - techniques overview



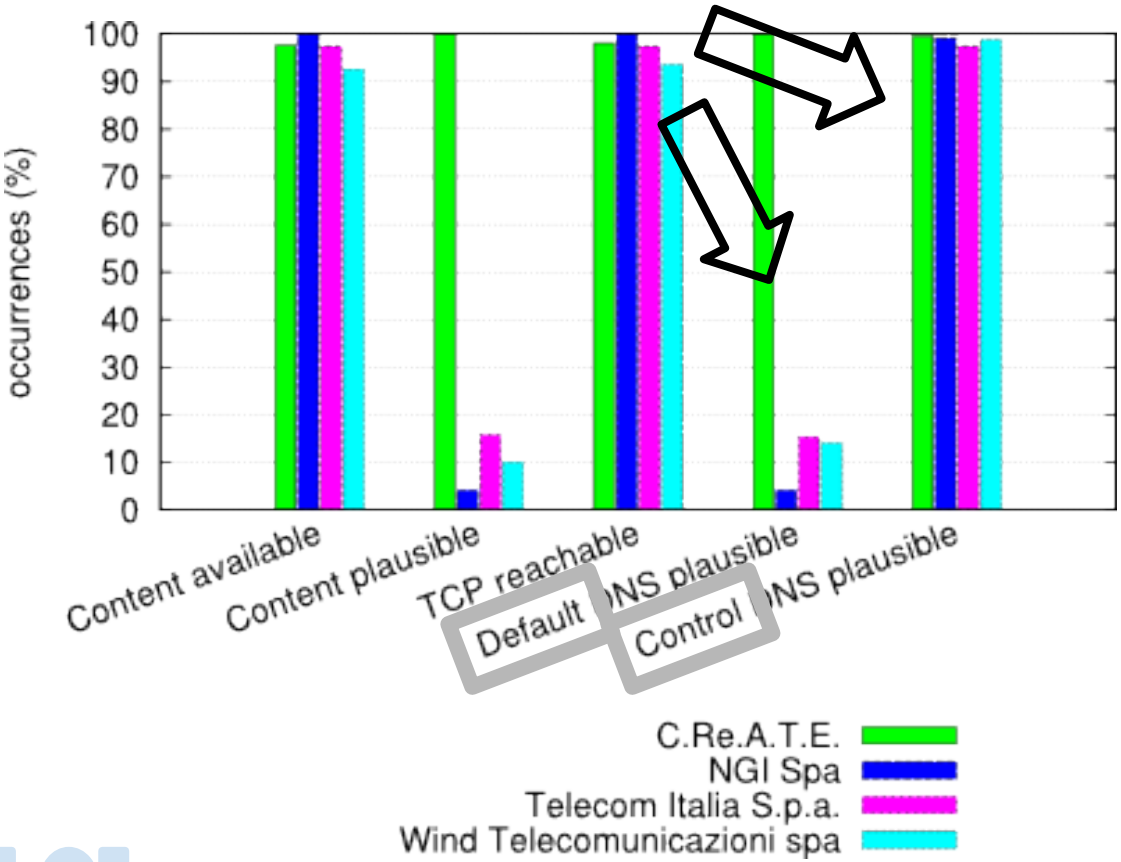
for all but one ISP

DNS and **Content** both low:

DNS tampering

Italy: gaming and betting websites

bet365.com - techniques overview



for all but one ISP

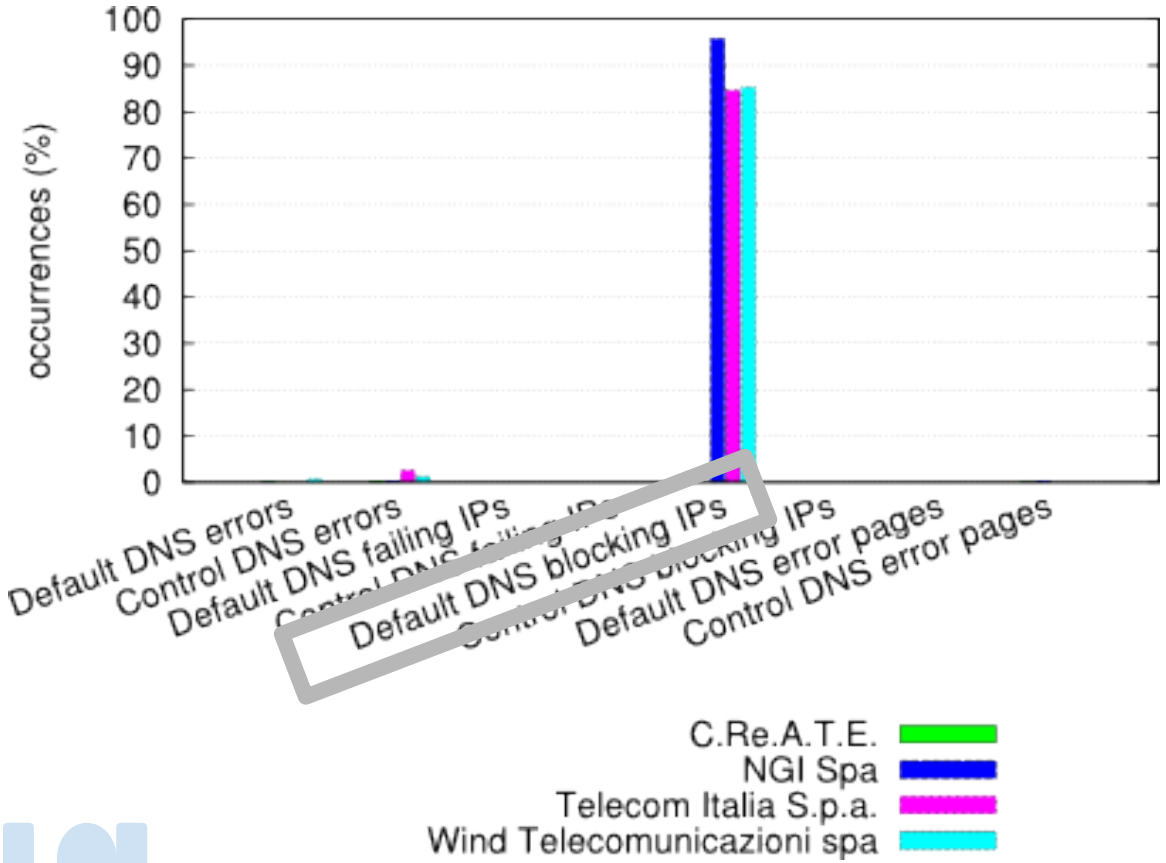
DNS and **Content** both low:

DNS tampering

of subtype: **hijacking**

Italy: gaming and betting websites

bet365.com - DNS detail



for all but one ISP

DNS and **Content** both low:

DNS tampering

of subtype: **hijacking**

symptom: **block page**

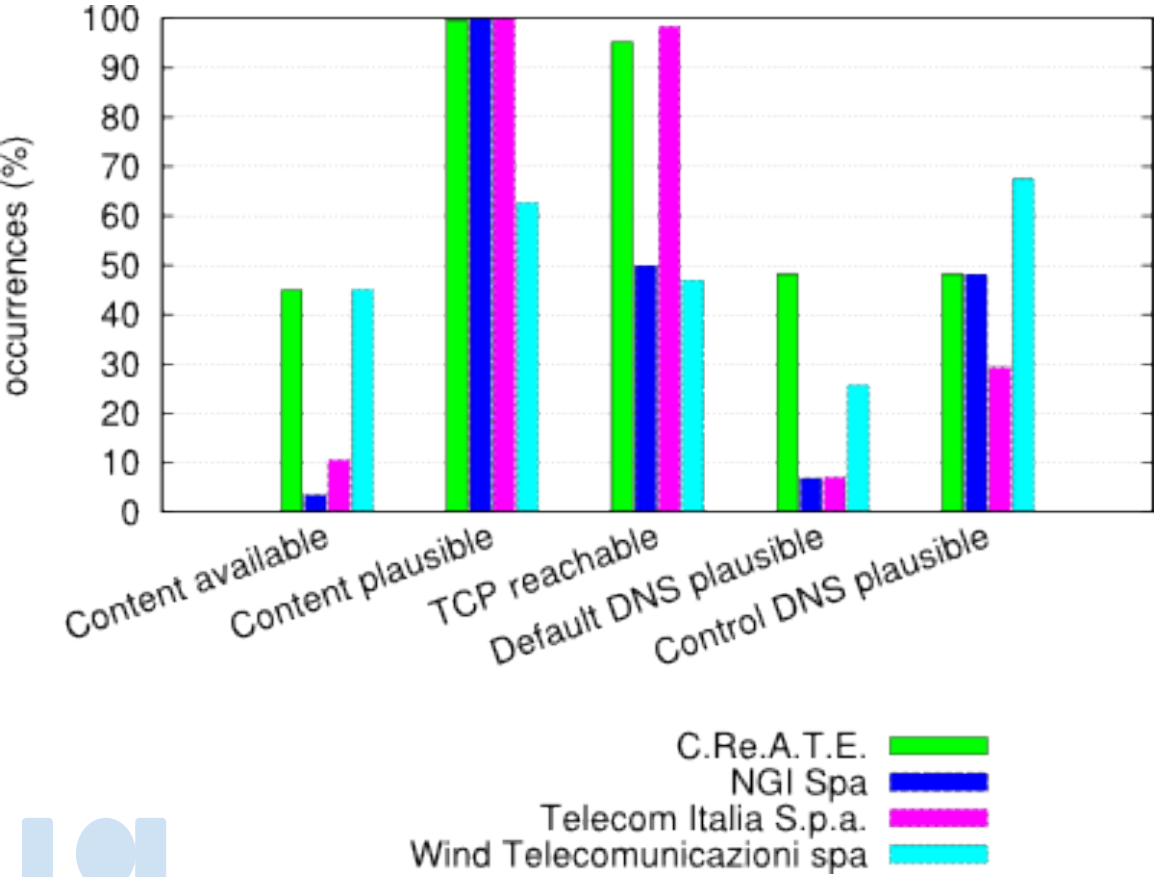
Italy: gaming and betting websites

UBICA responses

C.Re.A.T.E.	No censorship
NGI	DNS hijacking - send to *block page IP*
Telecom IT	DNS hijacking - send to *block page IP*
WIND	DNS hijacking - send to *block page IP*

Italy: file sharing websites

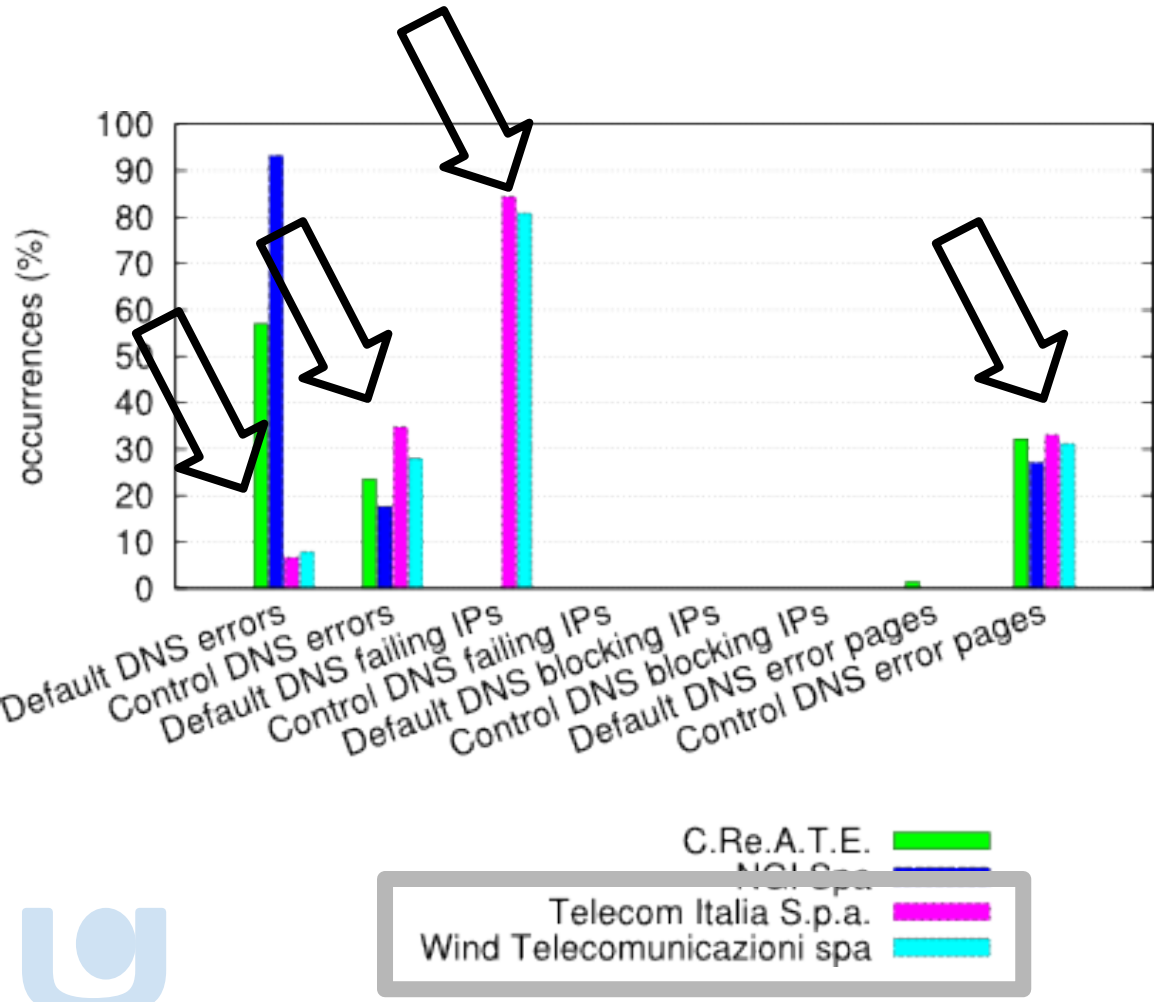
thepiratebay.sx - techniques overview



for all ISPs
Content partially available

Italy: file sharing websites

thepiratebay.sx - DNS analysis detail

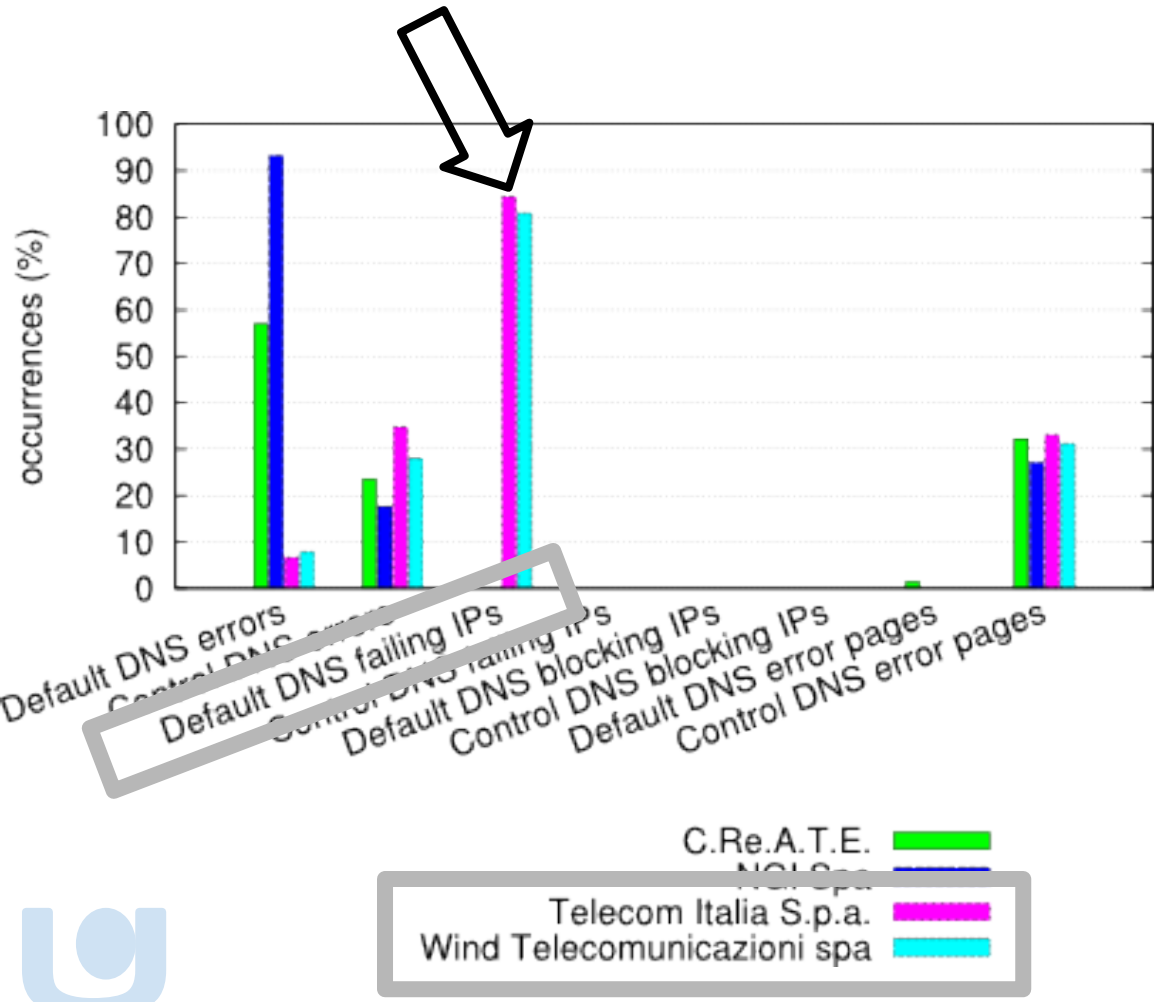


for all ISPs
Content partially available
No blocking pages

both Telecom IT
and WIND Tlc
Different DNS errors

Italy: file sharing websites

thepiratebay.sx - DNS analysis detail

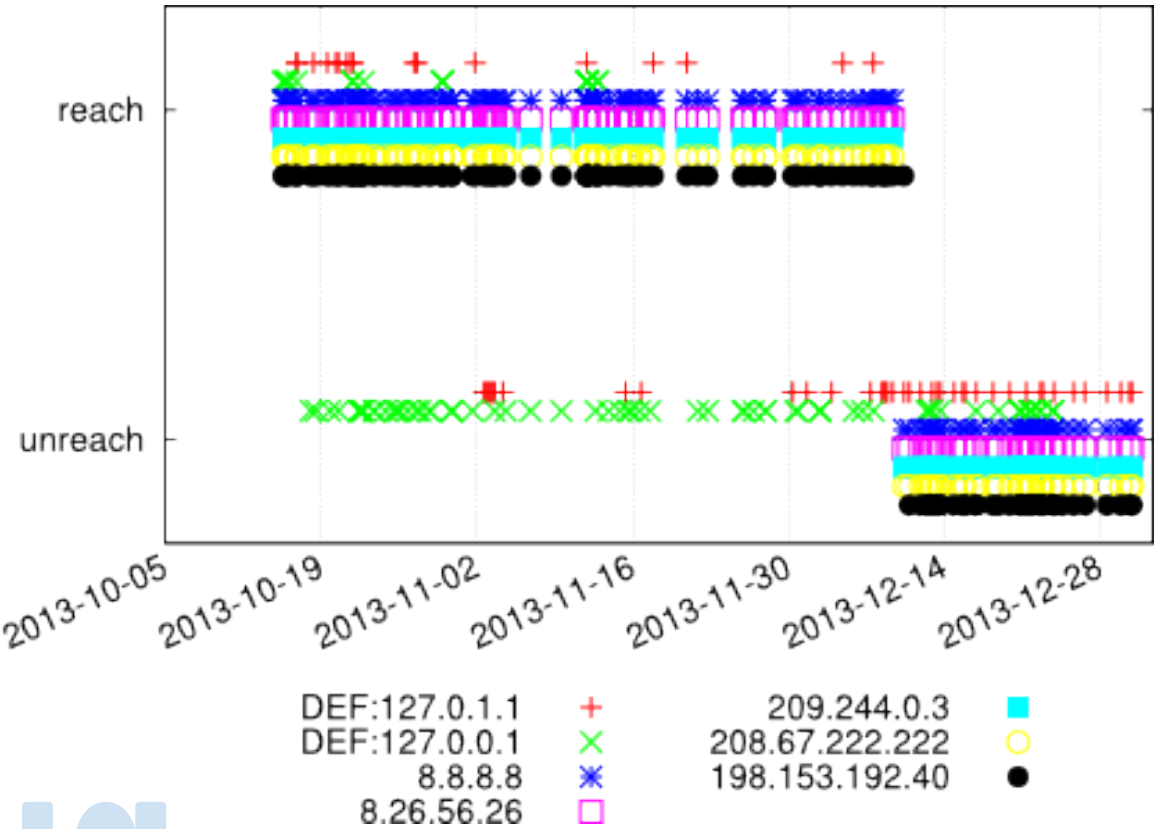


for all ISPs
Content partially available
No blocking pages

both Telecom IT
and WIND Tlc
Different DNS errors
censor with **Failing IP**

Italy: file sharing websites

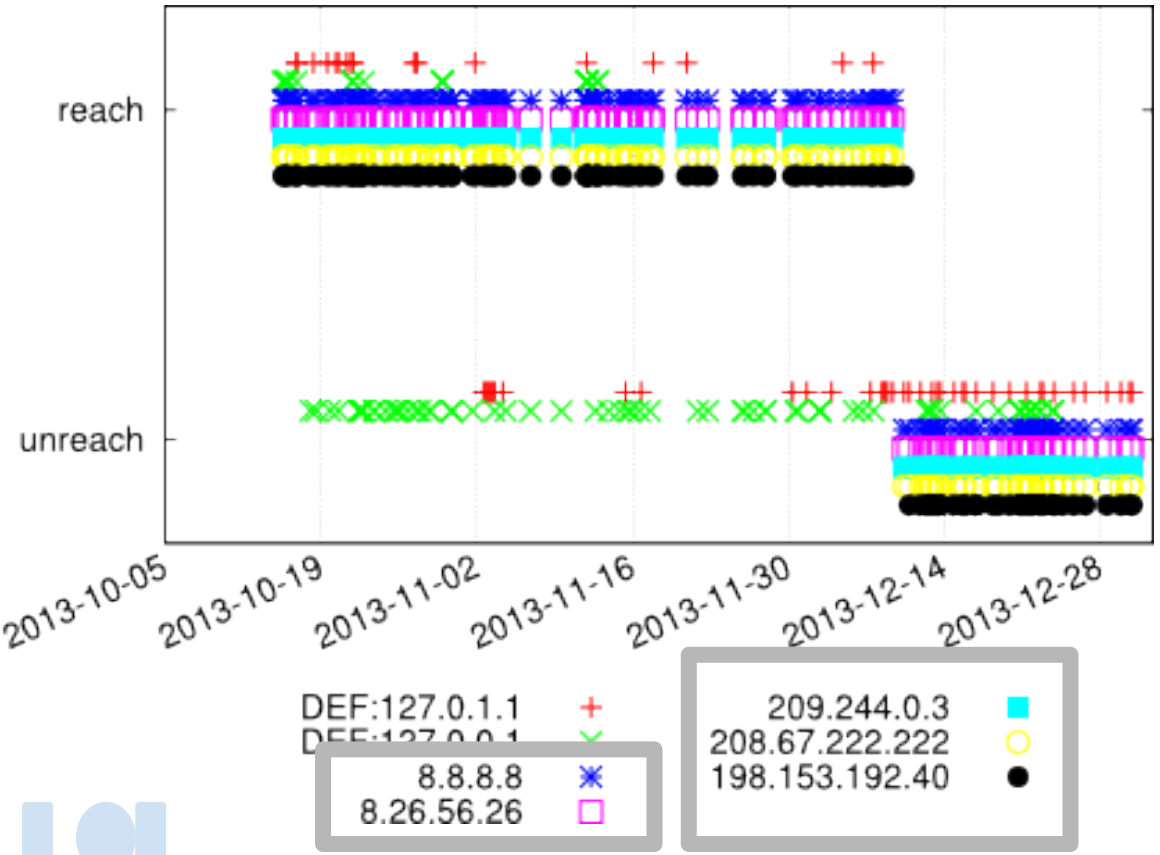
thepiratebay.sx - DNS time analysis



WIND ISP

Italy: file sharing websites

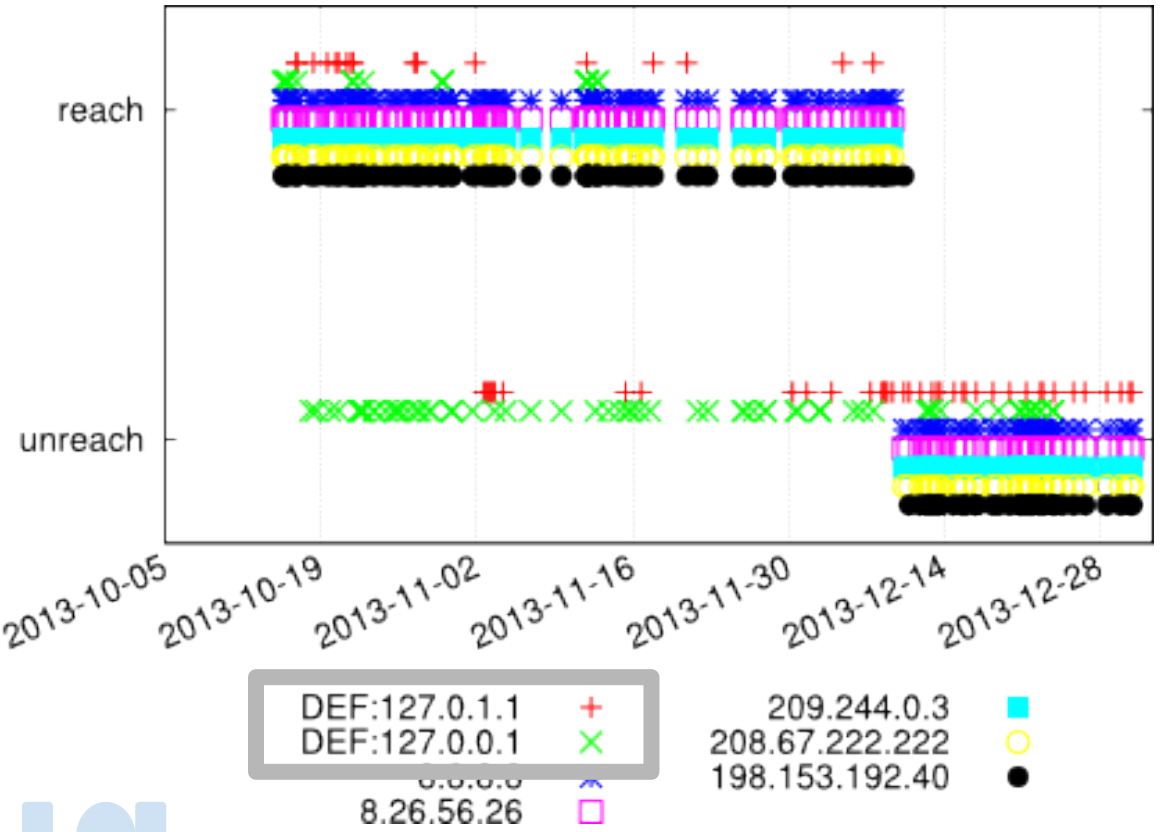
thepiratebay.sx - DNS time analysis



WIND ISP
Control resolvers OK

Italy: file sharing websites

thepiratebay.sx - DNS time analysis

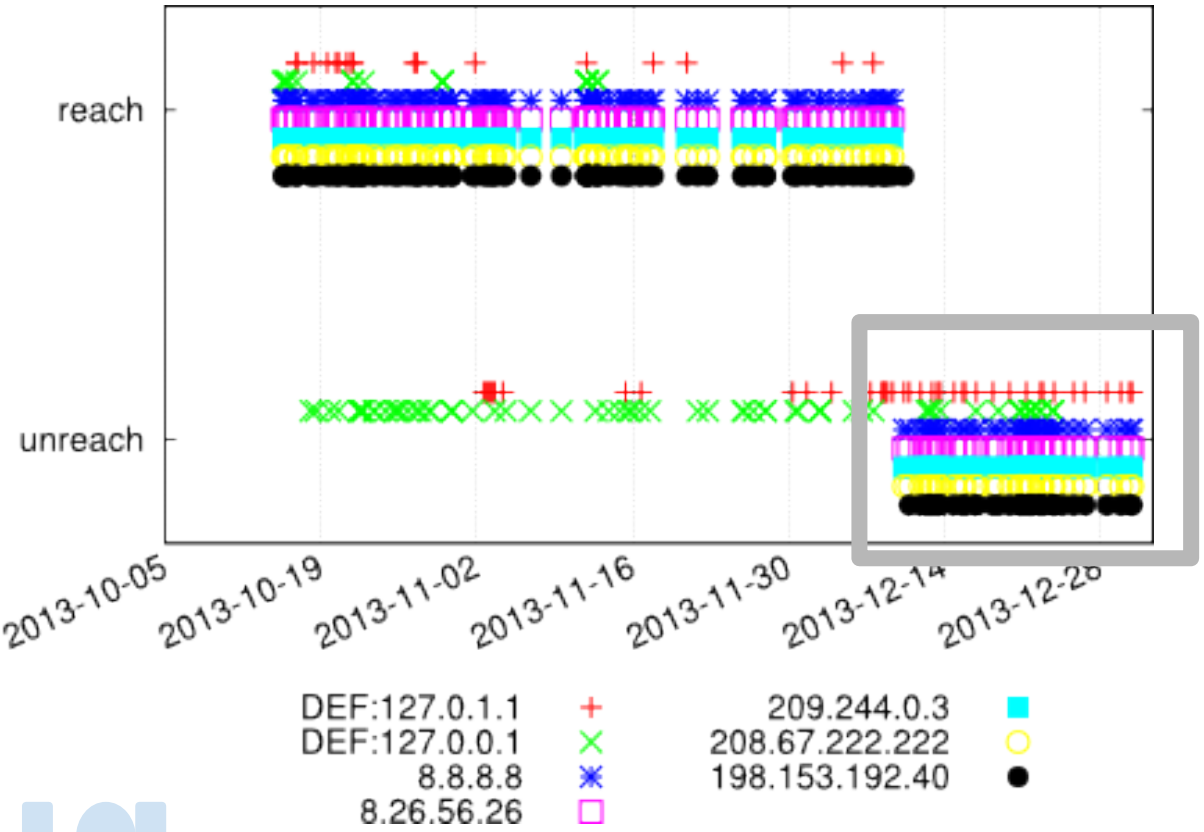


WIND ISP
Control resolvers OK

Default resolvers
one **hijacking**
one **obscillates**

Italy: file sharing websites

thepiratebay.sx - DNS time analysis



WIND ISP
Control resolvers OK

Default resolvers
one **hijacking**
one **obscillates**

All agree on
unreachability

Italy: file sharing websites

thepiratebay.sx - response

C.Re.A.T.E.	No censorship - domain expired
NGI	No censorship - domain expired
Telecom IT	DNS hijacking - send to *failing IP*
WIND	DNS hijacking - send to *failing IP*

Outline

- Motivations
- UBICA platform
- Size Ratio test
- Case study
  - Pakistan
  - Korea
  - Italy
- **Conclusions**

Concluding remarks

- Variability over time and space (ISP,country)
- Different censoring techniques with similar symptoms (e.g.RST from wrong host)
- Internet Censorship detection can be tricky

Concluding remarks

- Variability over time and space (ISP,country)
- Different censoring techniques with similar symptoms (e.g.RST from wrong host)
- Internet Censorship detection can be tricky

UBICA has proved effective in monitoring it

Undergoing/future work

- new collection / tests
  - throttling detection (download time)
  - topology information collection (traceroute to target)
  - TLS tampering (SSL certificate check)
  - Tuning *window size* for time analysis
- features
  - tomography setup (with helper server)
  - publish open data

    anonymize probe IP?
    remove/degrade timestamp?

“Side” open issues

- How to compare different countries?
- How much censorship is “high”?
- Involving users can be dangerous for them?

Collaborations more than welcome

- UBICA is crowdsourced (mostly academic): accuracy and coverage depend on participation
- New tests, analyses, features...

Preview of report interface (a snapshot run), and headless probe:

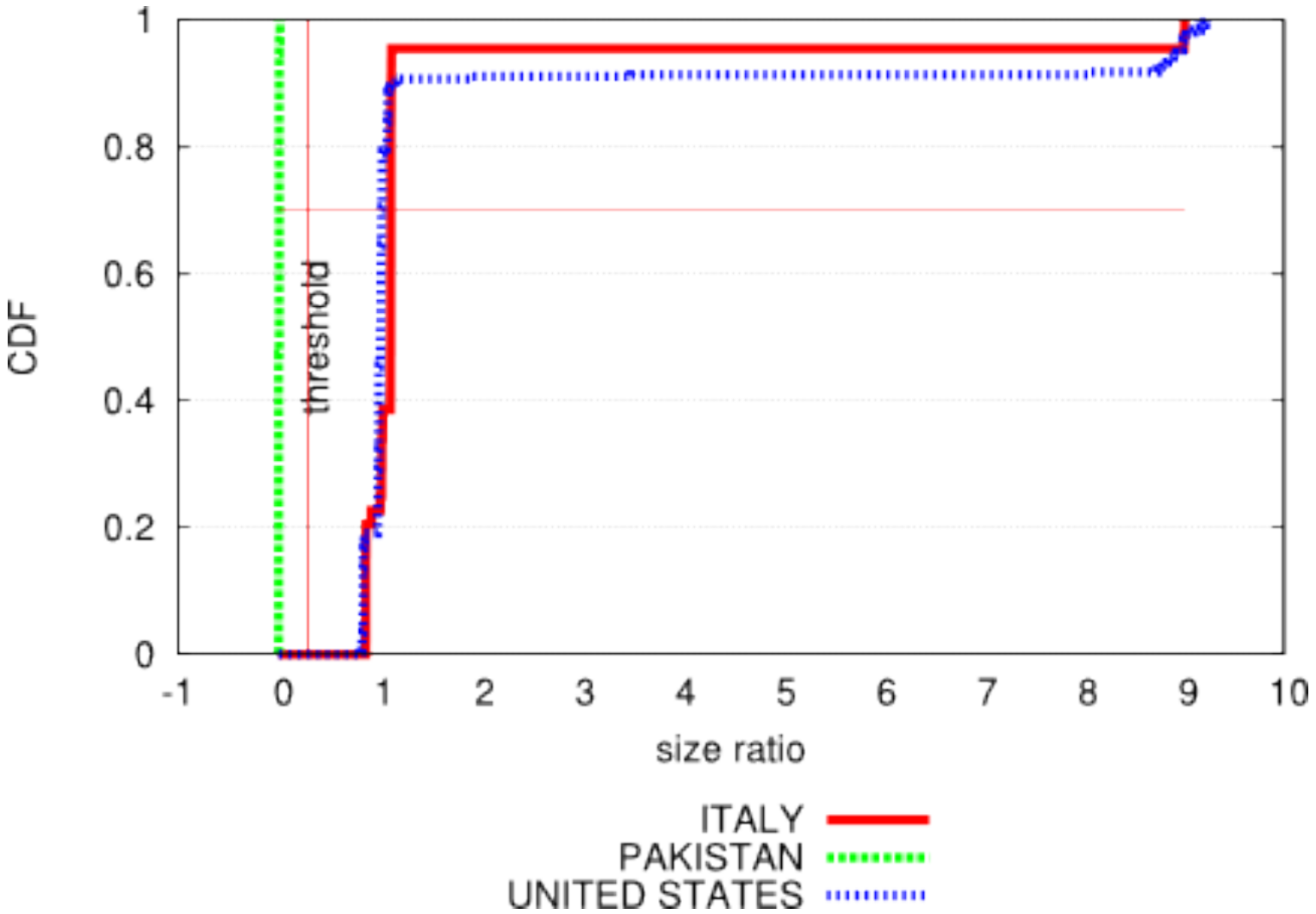
http://traffic.comics.unina.it/Traffic/ubica.php

drop an email! giuseppe.aceto@unina.it

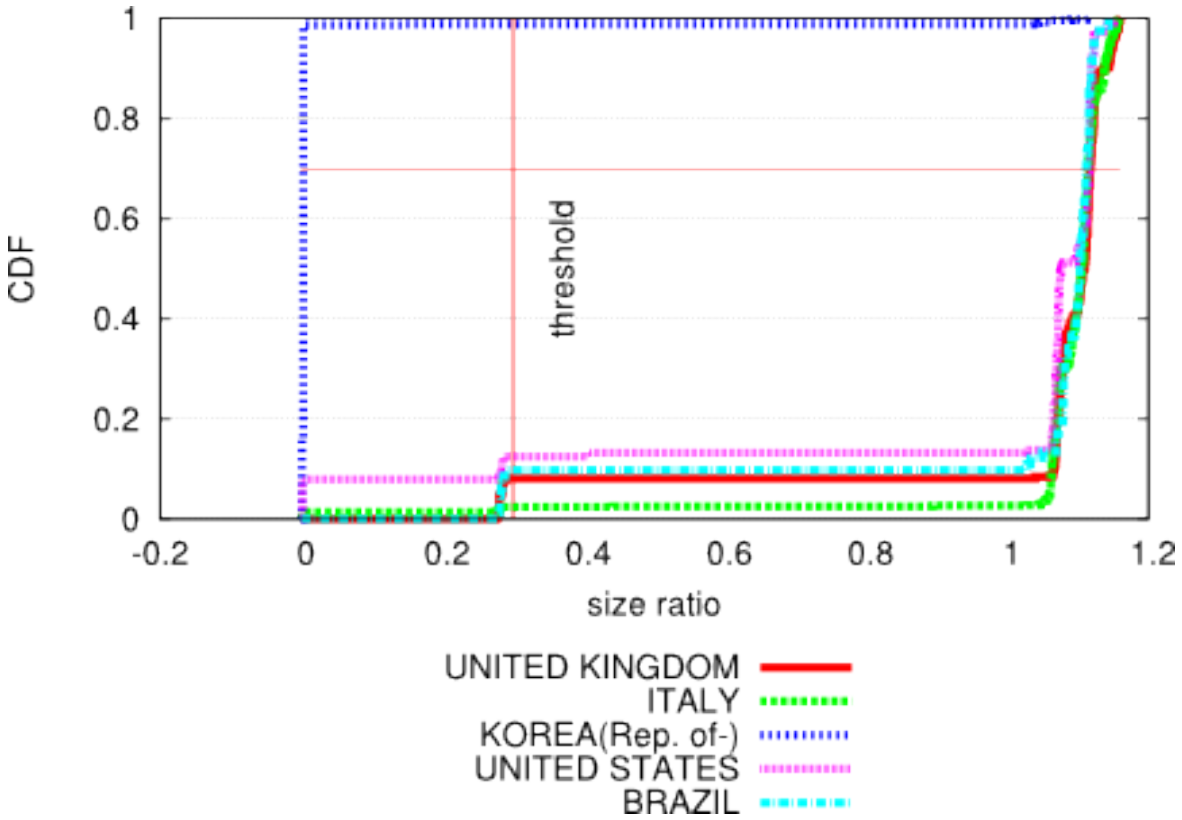
Thanks for your attention

backup stuff

Censorship in Pakistan: the case of YouTube



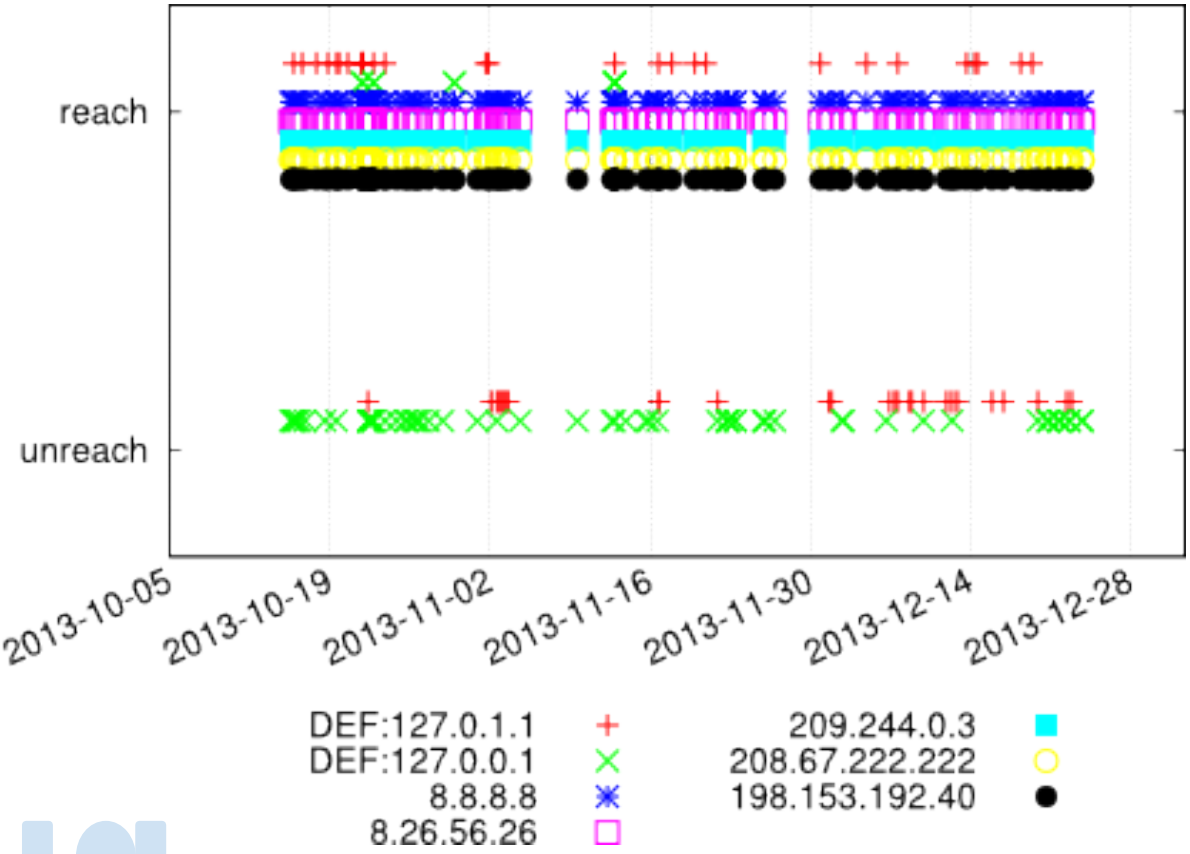
Censorship in Korea: porn websites



injected JavaScript section:
redirect to the address http://warning.or.kr

Italy: gaming and betting websites

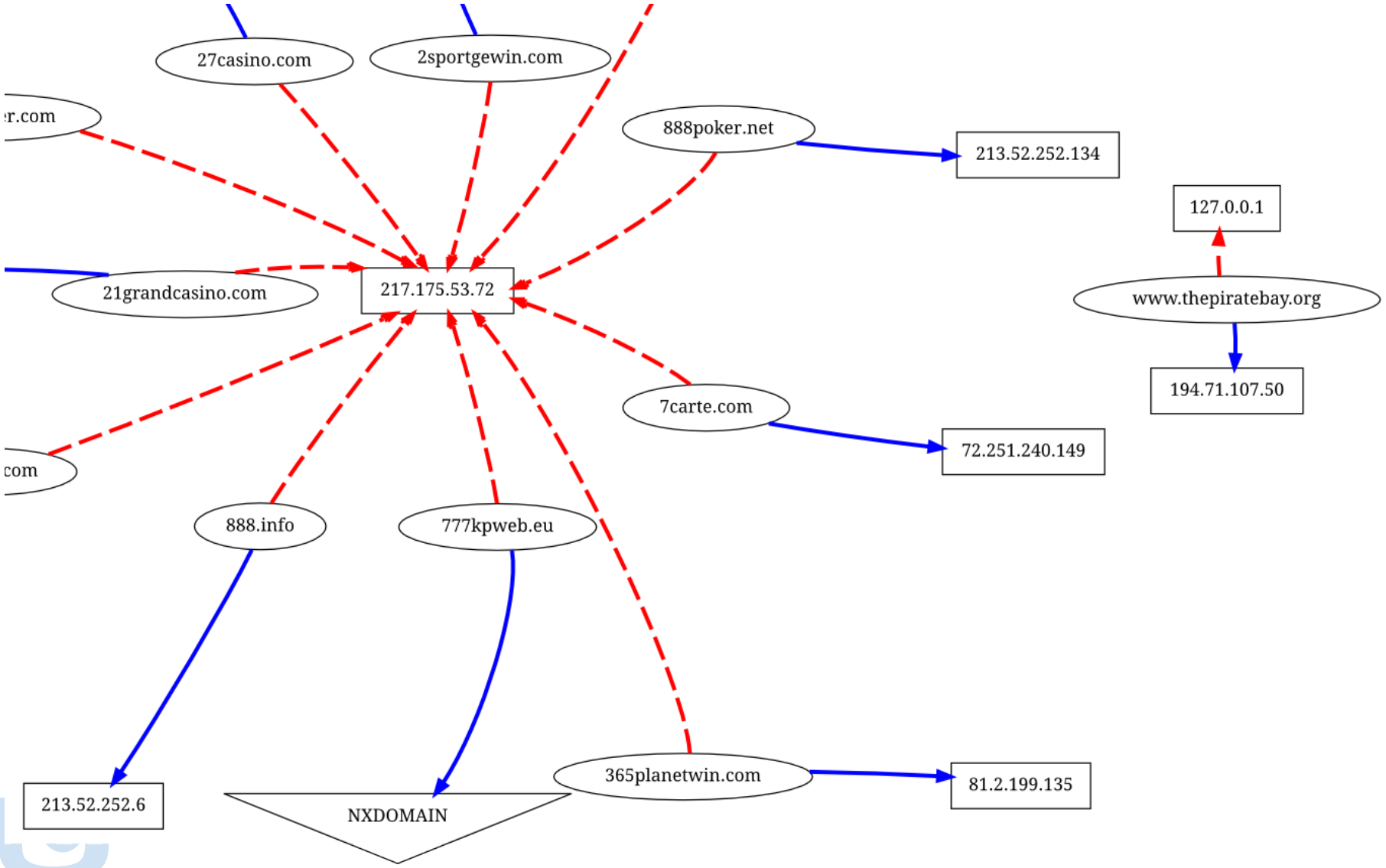
bet365.com - DNS time analysis



Default DNS obscillates:

possibly **open resolvers** used by some probe **as default**

DNS hijacking in Italy: betting websites



Internet Censorship is widespread

Countries may differ in

- motivations
- targets
- duration
- techniques

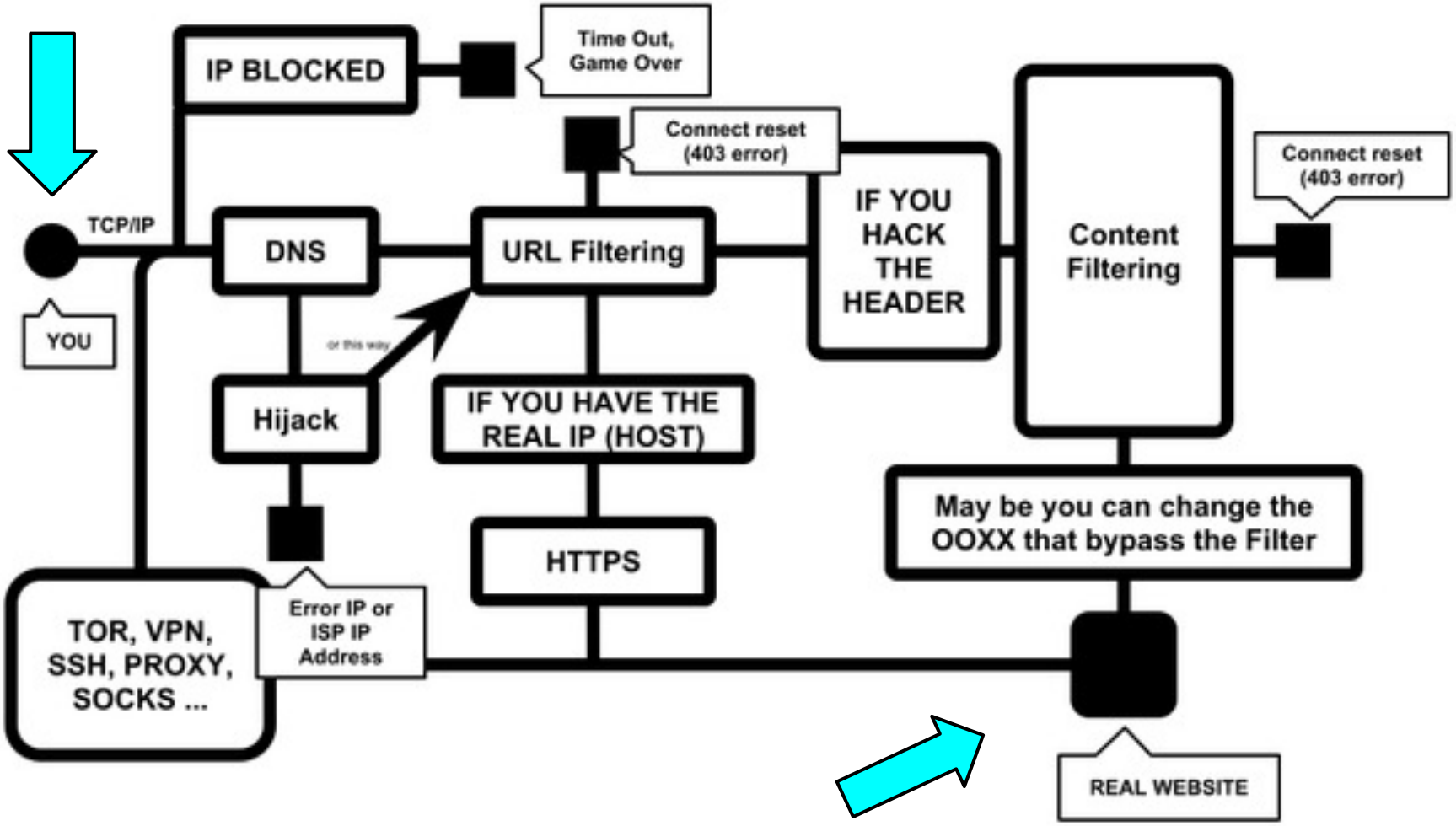
but great many* of them censor the Internet

*OpenNet Initiative http://map.opennet.net
Deibert, Ronald. *Access denied: The practice and policy of global Internet filtering*. Mit Press, 2008.

UBICA control 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

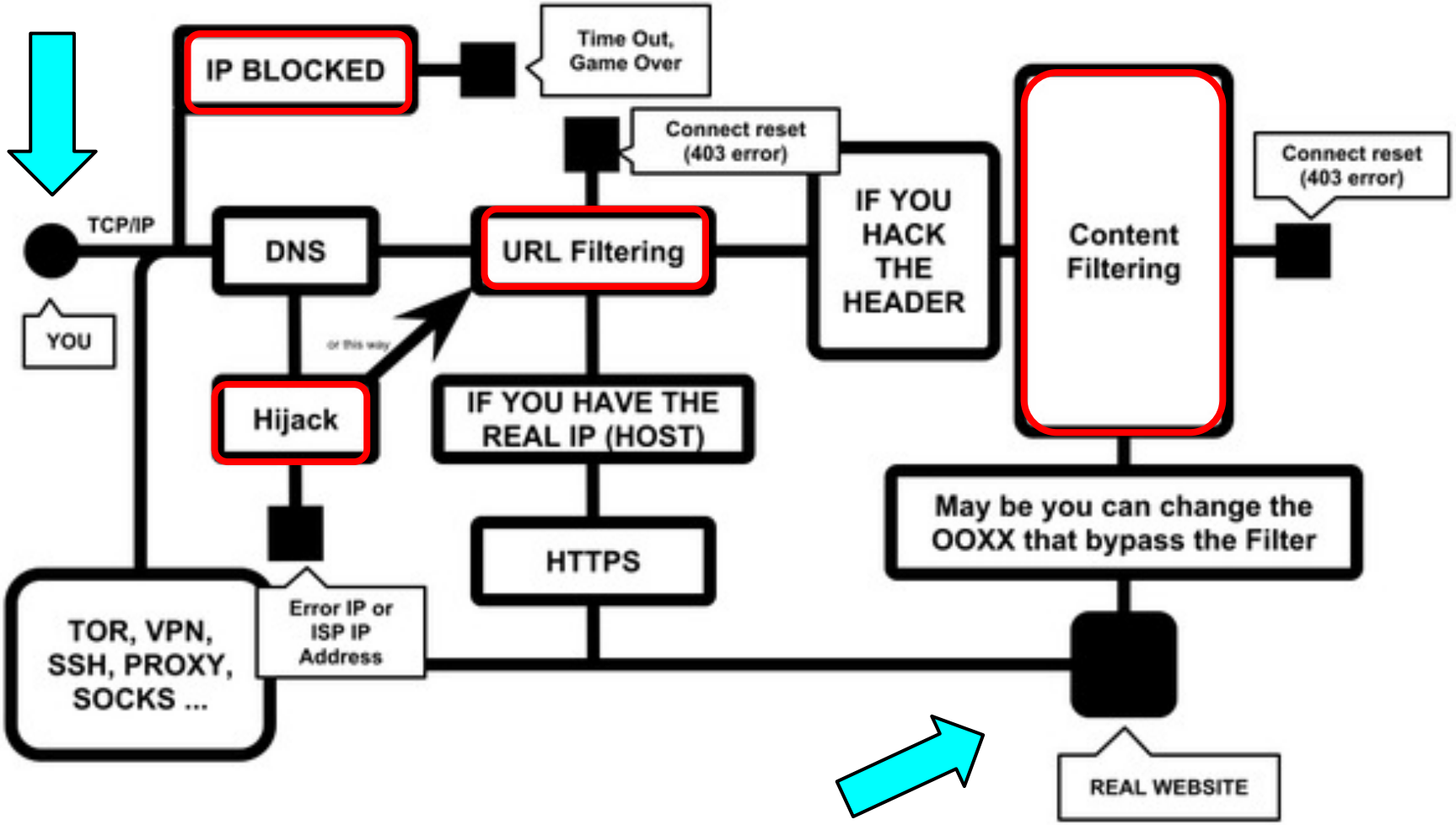
Several vulnerable phases of communication



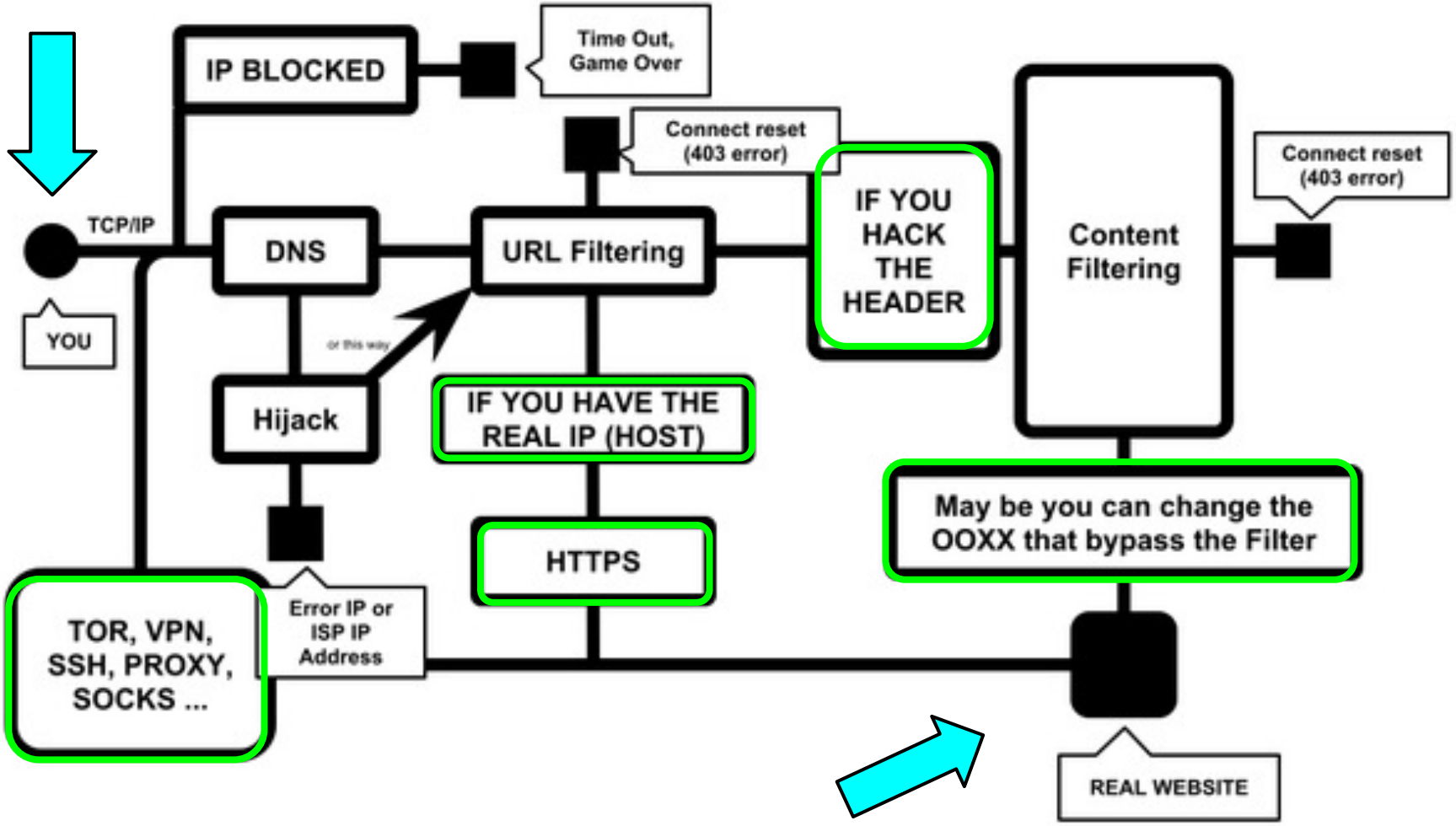
Route of the GFW - This is our Internet (China)

by J.T use Google Docs

Several techniques to *enforce* censorship



... and methods to *circumvent* censorship



Route of the GFW - This is our Internet (China)

by J.T use Google Docs