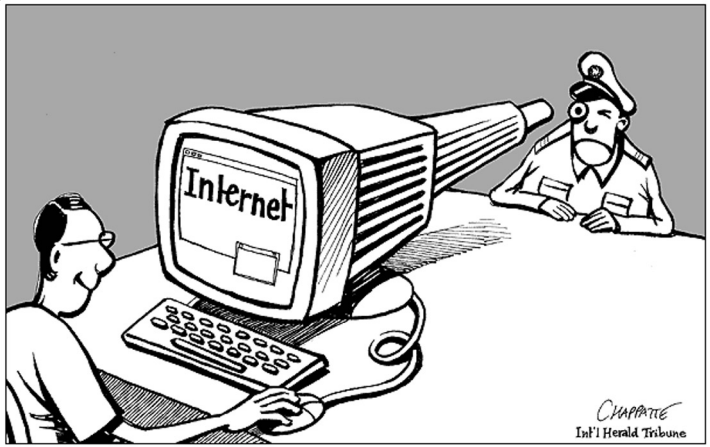
# Internet Censorship & Surveillance




UNIVERSITY OF WATERLOO

# Motivations for censorship and surveillance

# The Panopticon

## Jeremy Bentham's prison design

- Prisoners are aware of the presence of authority at all times, even though they never know when they are being observed
- Prisoners discipline themselves because someone might be watching
- A very cost-effective way to keep order

# The Panopticon Effect

**Michel Foucault**

- **Extends the Panopticon into a symbol of social control**
  - **Visibility** reaching deep into individuals' **everyday life**
- **A disciplinary society builds around rules and obedience**
  - Even without repercussions, **individuals self-impose a set of rules**
    - e.g., not speeding when no police car is visible
- **Just a thought exercise in the 1970's**
  - How was a figure of authority supposed to monitor everyone constantly?
  - **What about now?**

# The Internet Panopticon

- **The Internet enables authority figures to track (and act upon) multiple records of intellectual activities** ⇨ **Remember NSA's PRISM?**
- **Surveillance prevents "intellectual privacy"**
  - Interfere with the generation and maturing of ideas
  - Thoughts and beliefs get driven to:
    - the boring
    - the bland
    - the mainstream

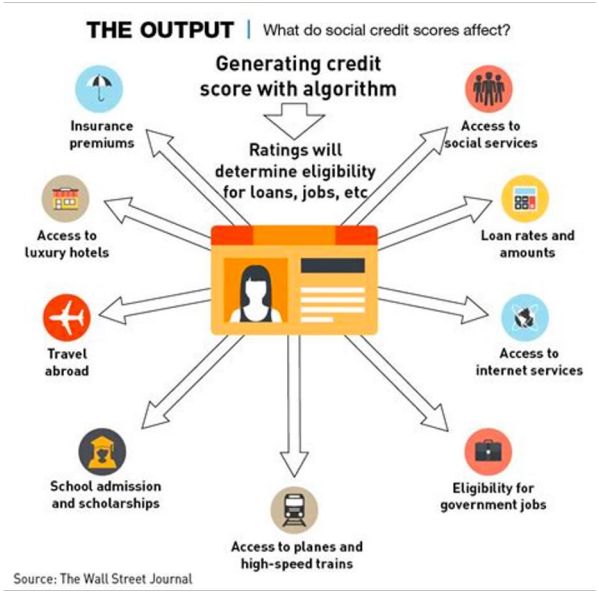
  } **Chilling effect**
  **Self-censorship**

**Anyways, probably too extreme to be put in practice, right?**

**Right?...**

# The Chinese Social Credit System

- **A push towards standardizing individuals' behavior**
  - Rewards for following the norm
  - Punishments for deviating from the norm

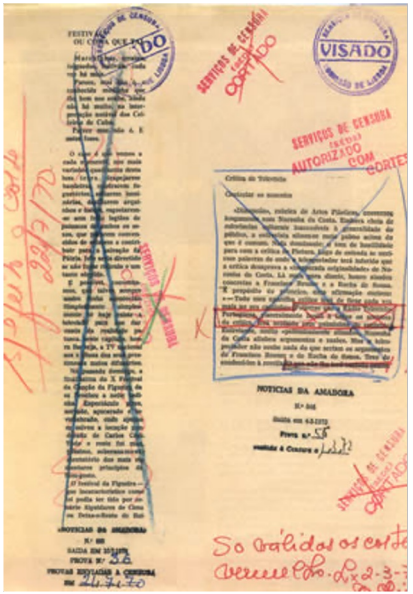






UNIVERSITY OF WATERLOO

# When the Panopticon is not enough to contain those who dare…

- **Prevent access to information via censorship**
  - Thwart the free discussion of ideas
  - Disempower local communities
  - Stifle contradictory opinions and dissent
  - Impose political and moral agendas
- **As real in today's Internet as in old-days vetoing**






UNIVERSITY OF WATERLOO

# Internet Censorship in Thailand: User Practices and Potential Threats

Genevieve Gebhart*†‖, Anonymous Author[2], Tadayoshi Kohno†
*Electronic Frontier Foundation †University of Washington
gennie@eff.org
yoshi@cs.washington.edu

bstract—The "cat-and-mouse" game of Internet censorship
d circumvention cannot be won by capable technology
ne. Instead, that technology must be available,
prehensible, and trustworthy to users. However, the field
ely focuses only on censors and the technical means to
mvent them. Thailand, with its superlatives in Internet
and government information controls, offers a rich case
for exploring users' assessments of and interactions with
rship. We survey 229 and interview 13 Internet users in
nd, and report on their current practices, experienced
erceived threats, and unresolved problems regarding
hip and digital security. Our findings indicate that
circumvention tools were adequate for respondents to
locked information; that respondents relied to some
n risky tool selection and inaccurate assessment of
content; and that attempts to take action with
content on social media led to the most concrete
ith the least available technical defenses. Based on
ines and in direct response to these problems, we
shifting objectives in anti-
and future

security community has proposed novel circumvention methods in response [10, 25, 38].

The goal of circumventing censorship and attaining freer access to information, however, relies on those circumvention methods being available, comprehensible, and trustworthy to users. Only by meeting users' needs can circumvention tools realize their full technical capabilities. With this goal in mind, the field lacks sufficient inquiry into the range of user perceptions of and interactions with censorship. How do users assess censored content? What is the range of their reactions when they encou
censorship? How does censorship affect the way th
only access but also produce information?

In addition to guiding more thorough anti-circumvent
strategies, these questions about users and censorship ca
act as a lens into broader security issues. Users
perspectives on censorship have wide-ranging implications
for security behaviors both on and offline [51, 55],
especially in the politically repressive, low-resource
ments in which common-sense censorship
re most needed. Looking

# Analyzing Internet Censorship in Pakistan

seppe Aceto, Alessio Botta, Antonio Pescapé
University of NapoliFederico II (Italy),
and NM2 S.r.l. (Italy),
giuseppe.aceto, a.botta, pescape}@unina.it

M. Faheem Awan, Tahir Ahmad, Saad Qaisar
National University of Science and Technology, NUST (Pakistan),
{10mscsemawan,11mscsstahmad,saad.qaisar}@seecs.edu.pk

—Internet Censorship is unceasingly increasing in
ies worldwide in order to restrict web contents within
remises. According to latest Open Net Initiative
almost 50 countries are involved in web censorship,
tan. This paper presents the methodology and the
alysis based on publicly available censored URLs
viding both qualitative and quantitative results
or ISPs are censoring web content in Pakistan.
udy in literature analyzing and comparing the
ISPs in Pakistan using automated detection
tive probing measurements. Our results show
CTL, and Nayatel block content by using DNS
Watten and Qubee apply filtering, using
comment on these results by considering
e of the forced censorship mechanisms.
University closed survey in order to find
iques adopted by users in Pakistan and
users try to evade censorship by

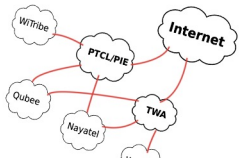



Fig. 1: Interne
sidera

Kevin Bock

Abstract

nsorship arms race advances, so
ing "censorship-in depth," compos
censorship mechanisms. This can ma
both measure and evade censorship. Ea
loyed their *protocol filter* that permits on
ocols (DNS, HTTP, and HTTPS) and cens
ising any other protocol. Iran composes the
with their standard censorship, threatening th

# Decentralized Control: A Case Study of Russia

Reethika Ramesh*, Ram Sundara Raman*, Matthew Bernhard*, Victor Ongkowijaya*,
Leonid Evdokimov†, Anne Edmundson†, Steven Sprecher*, Muhammad Ikram‡, Roya Ensafi*
*University of Michigan, {reethika, ramaks, matber, victorwj, swsprec, ensafi}@umich.edu
‡ Macquarie University, †Independent, leon@darkk.net.ru

*Abstract*—Until now, censorship research has largely f
on highly centralized

# Where The Light Gets In: Analyzing Web Censorship Mechanisms in India

Tarun Kumar Yadav*
IIIT Delhi, India
tarun14110@iiitd.ac.in

Akshat Sinha*
IIIT Delhi, India
akshat14132@iiitd.ac.in

Devashish Gosain*
IIIT Delhi, India
devashishg@iiitd.ac.in

Piyush Kumar Sharma
IIIT Delhi, India
piyushs@iiitd.ac.in

Sambuddho Chakravarty
IIIT Delhi, India
sambuddho@iiitd.ac.in

ABSTRACT

In this work we present a detailed study of the Internet censorship mechanism in India. We consolidated a list of potentially blocked websites from various public sources to assess censorship mechanisms used by nine major ISPs. To begin with, we demonstrate that existing censorship detection tools like OONI are grossly inaccurate. We thus developed various techniques and heuristics to correctly assess censorship and study the underlying mechanism used by these ISPs. At every step we corroborated our finding manually to test the efficacy of our approach, an exercise largely ignored by several others. We fortify our findings by adjudging the *coverage* and *consistency* of censorship infrastructure, broadly in terms of average number of network paths and requested domains the infrastructure censors. Our results indicate a clear disparity among ...ow they install censorship infrastructure. For instance, ...served the censorious middleboxes in over ...whereas for Vodafone, it is as ...evising our own novel ...third party tools ...locked

mechanism that are employed by such nations *i.e.* — describing the network location of the censorship infrastructure — what triggers them — and how are clients notified of such filtering.

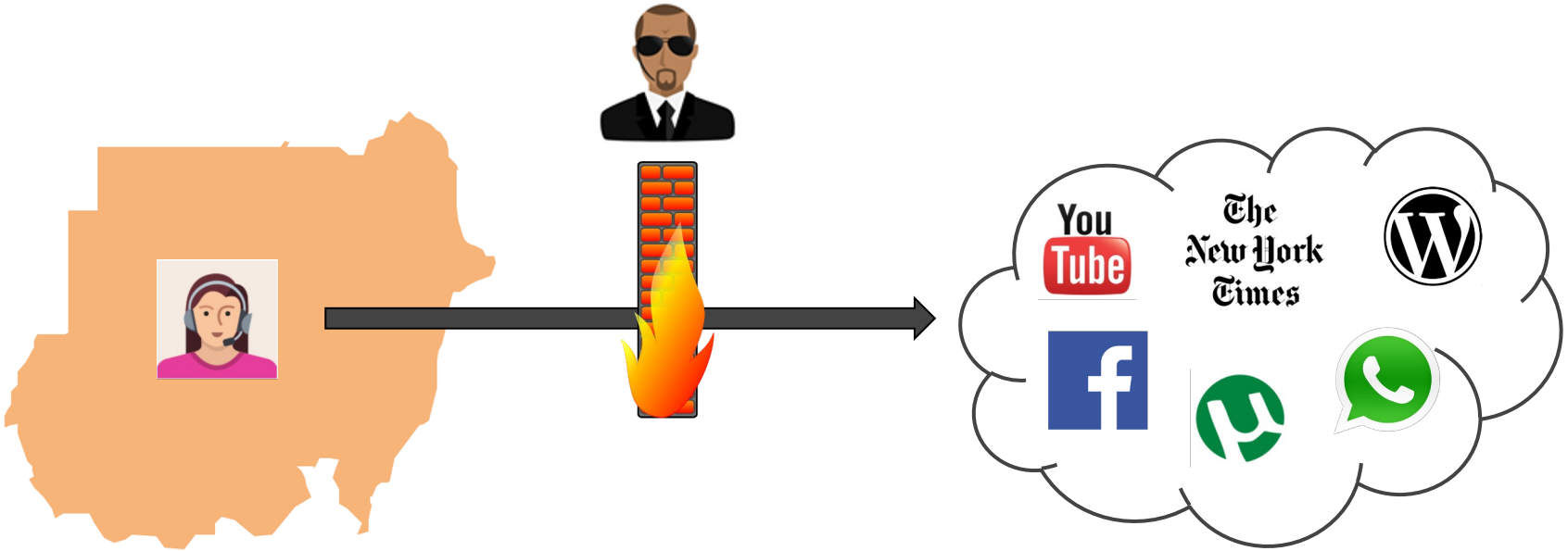
Through our studies over the past few years, we discovered that even democratic nations like India, have slowly, and rather covertly, evolved an infrastructure for large-scale Internet censorship, involving several privately and federally operated ISPs. India's Internet censorship policies have remained arbitrary (at best ambivalent)[1]. Over time several networks have upped their barriers against users accessing sites, which the administration "believes" to be "unfit for consumption", resulting in enough citizens facing web censorshi

Our previous work [21] emphasized on hypothetical scenari of potential (future) large scale censorship (or surveillance) by state. A mere preliminary report was also presented highligh the inconsistent web censorship policies amongst ASes.

We thus formally approached the authorities, filing a *Information* [15] request (RTI), inquiring about the polic mechanism the government uses to block content. In respo authorities shared that while the censorship policies are ...tial, the onus of implementing them lied with the indivi ...who could employ any mechanism they chose.

...ambiguous answer from authorities motivated us ...led analysis of the different censorship ...ctors of the country employ. ...s label is increasing ...gionally dive

# The typical state-level Internet censorship scenario

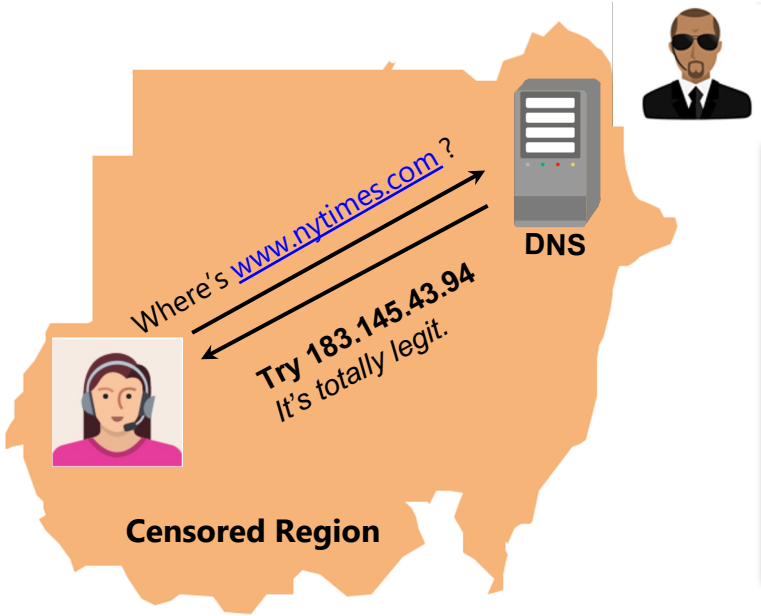


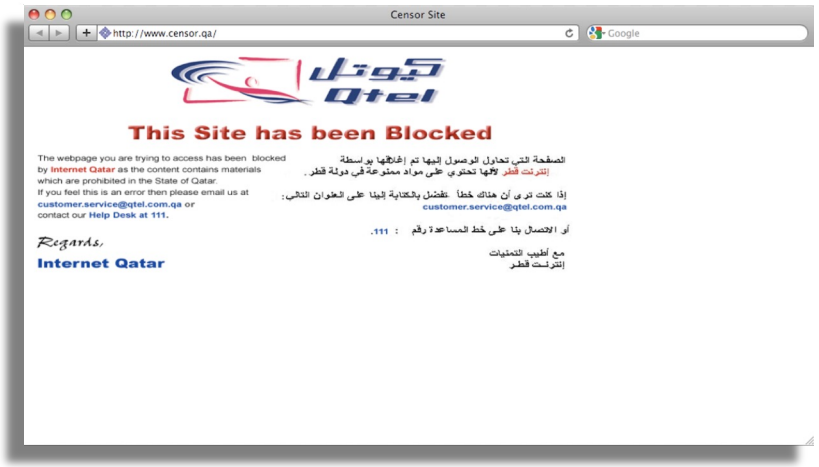



UNIVERSITY OF WATERLOO

# How do Censors Block Network Traffic?

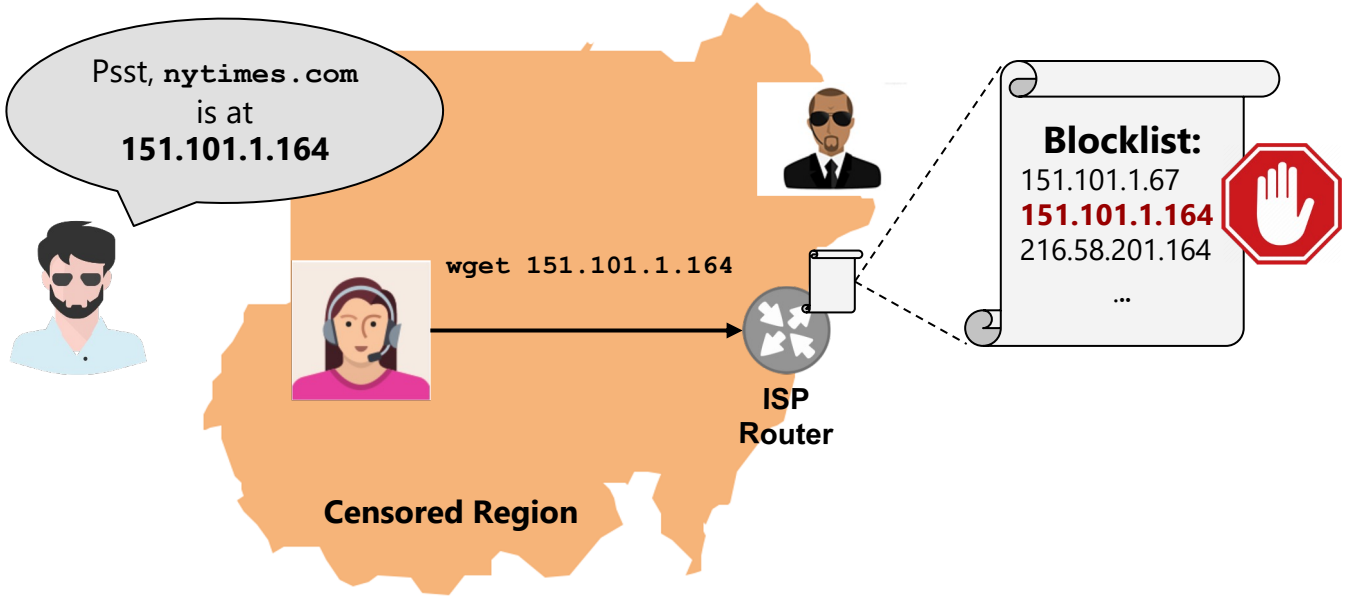
## Thwart IP address translation



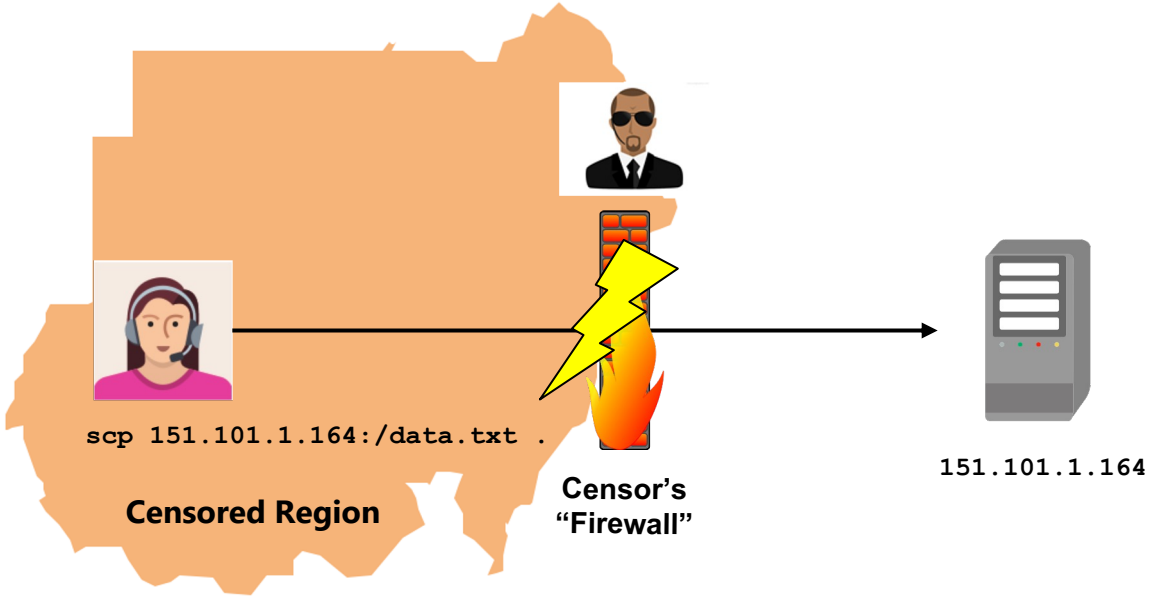
# How do Censors Block Network Traffic?

## Block IP addresses



# How do Censors Block Network Traffic?

## Slowdown network protocols

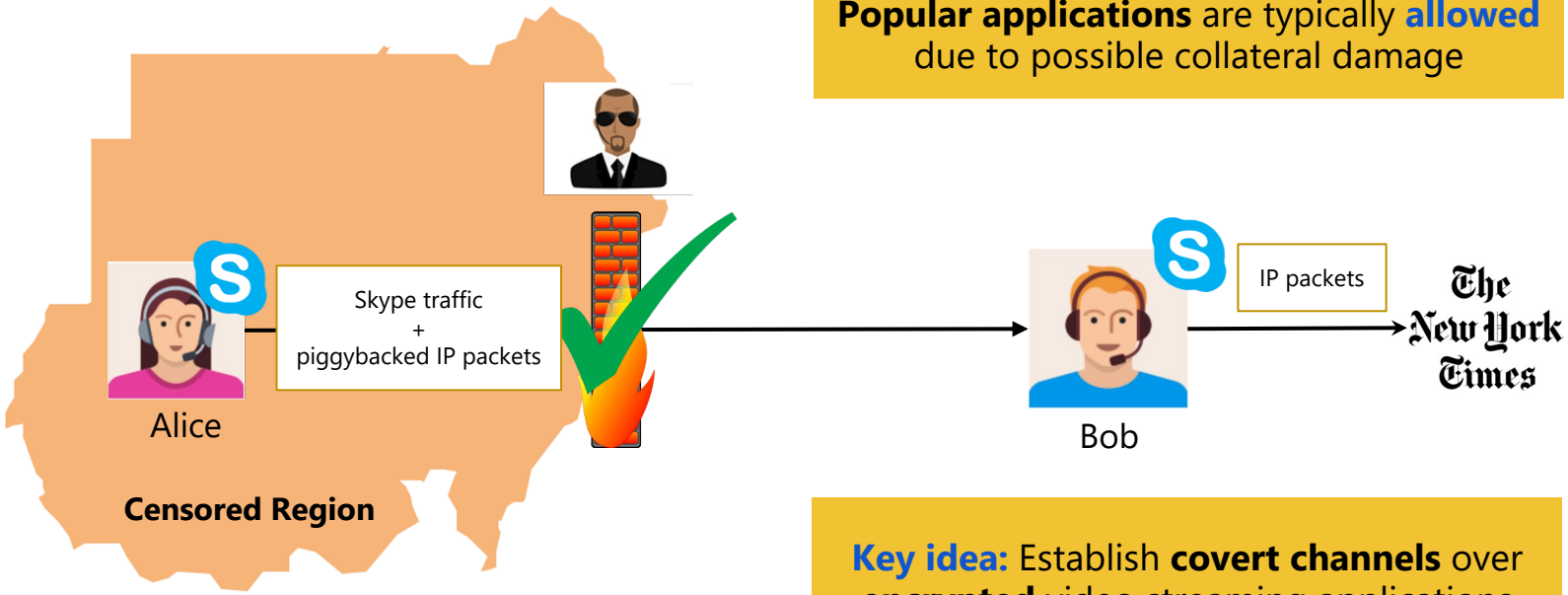


# How to get around Internet censorship?

- **Subliminal communication channels**
  - A broader notion of covert channels / steganography
    - Euphemisms on social media
    - “Abuse” popular Internet protocols


- **Thwart censors’ inspection abilities**
  - Break censors’ censorship mechanisms

# Not All Protocols and Destinations are Blocked

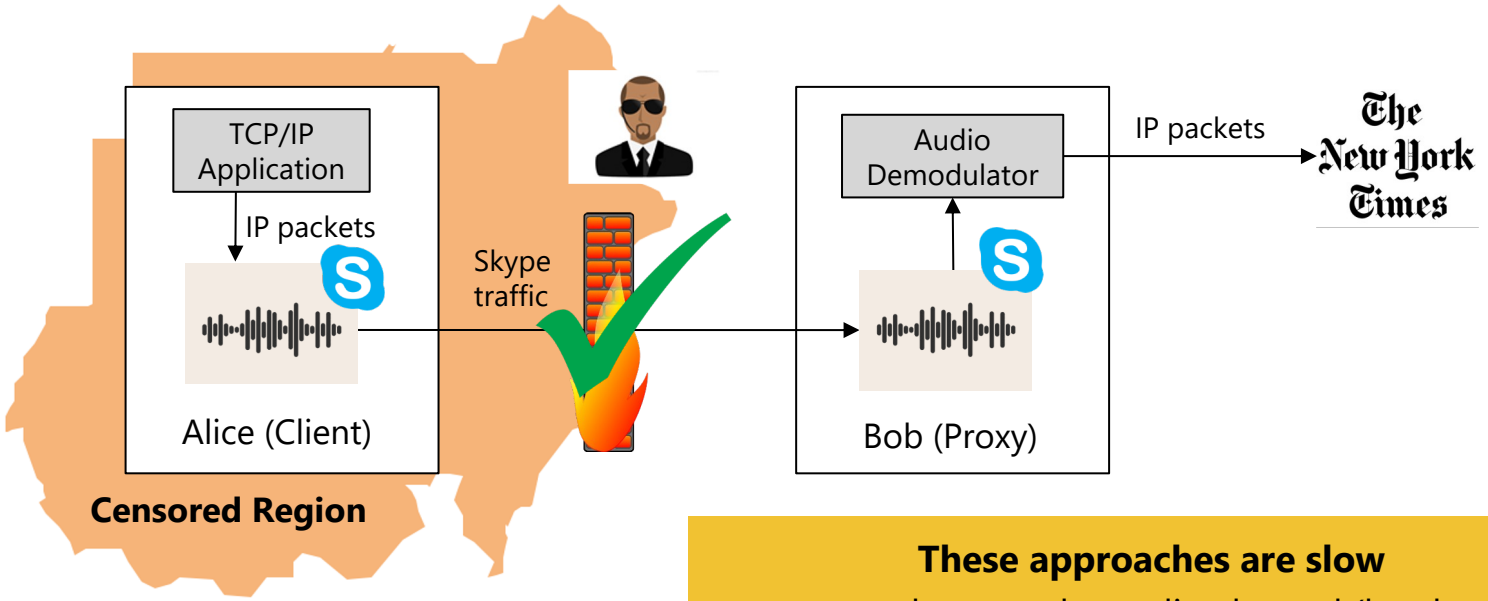



**Popular applications** are typically **allowed** due to possible collateral damage

**Key idea:** Establish **covert channels** over **encrypted** video streaming applications

UNIVERSITY OF WATERLOO

# We Can Tunnel Covert Data over Multimedia Protocols

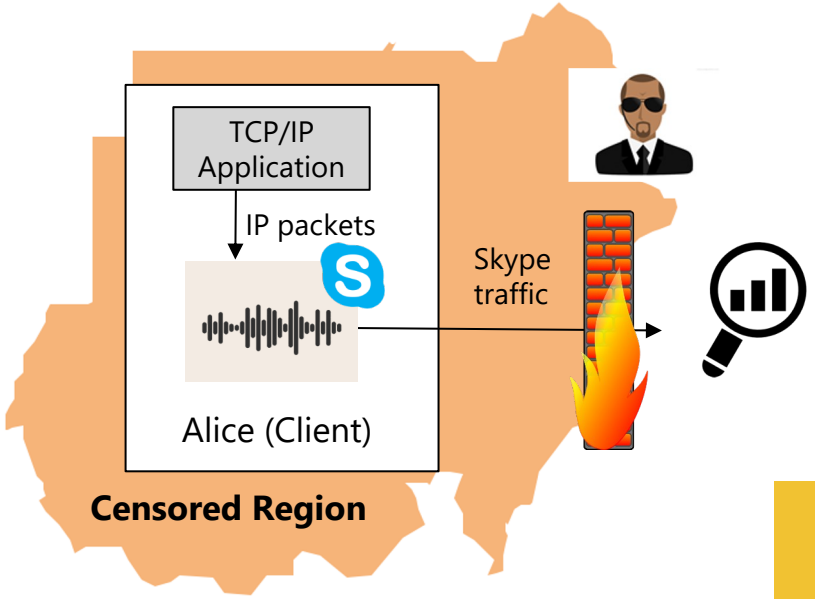



**These approaches are slow**

- they use the audio channel (low bw.)
- they require redundancy (further reducing bw.)

UNIVERSITY OF WATERLOO

# Censors Can Detect Covert Channels with Traffic Analysis




**Encrypted Traffic Analysis**
Statistical analysis of:

- Packet lengths
- Packets inter-arrival time

**Previous approaches are vulnerable**
e.g., covert data transmission detected by checking Skype **packet length std dev.**

UNIVERSITY OF WATERLOO

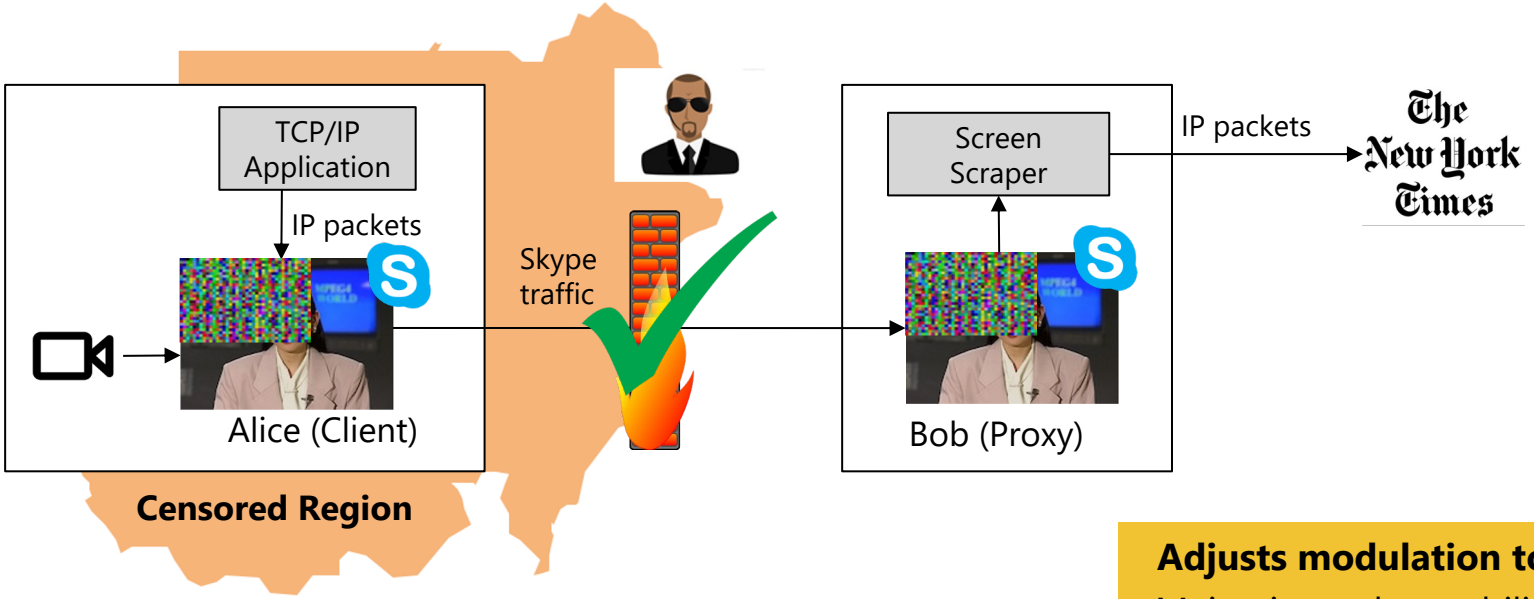
# Overarching Goal of Multimedia Protocol Tunneling (MPT)

High Throughput & Strong Resistance Against Traffic Analysis

also named **Unobservability**

# DeltaShaper: An Improved Tunneling Approach

MSc. Thesis & [PETS'17]




**Adjusts modulation to:**

Maintain unobservability

Increase throughput

UNIVERSITY OF WATERLOO

# Recurrent Issues of Multimedia Protocol Tunneling Tools

- **Network performance is very poor**
  - Low throughput: **~7 Kbps**
  - High latency: **~3s RTT**

- **Evaluation is performed with similarity-based classifiers**
  - Depend on small (and similar) sets of traffic features
  - Have not been compared in the literature

- **Poor evaluation may lead to optimistic unobservability claims**
  - Users of censorship-resistant tools may be endangered

UNIVERSITY OF WATERLOO

# Contributions

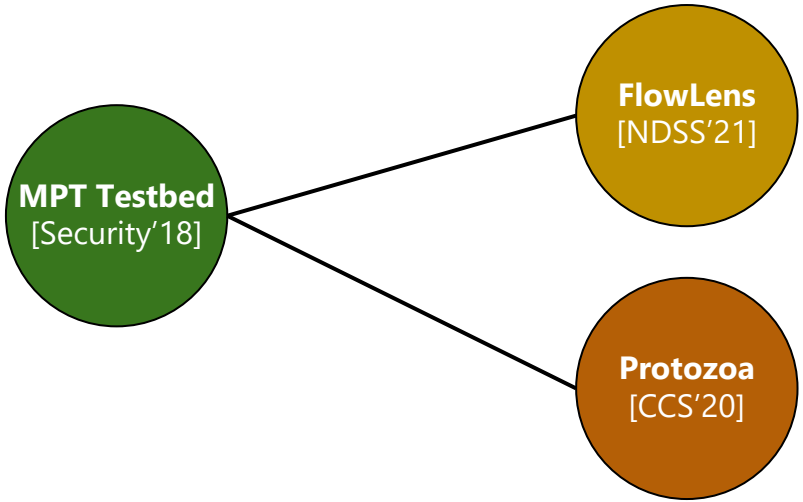
A. Found that the latest MPT tools were vulnerable to ML-based traffic analysis

A. Showed that ML-based traffic analysis can be widely deployed by ISPs

A. Developed a tool that offers unobservability / high-throughput (over WebRTC)

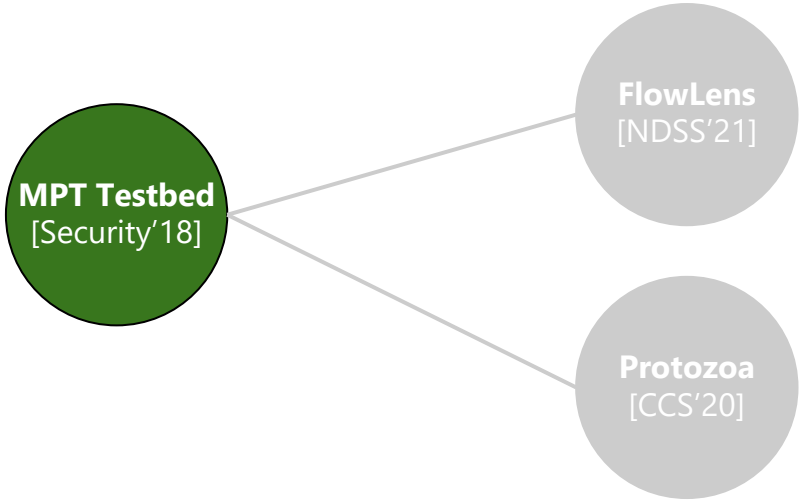
UNIVERSITY OF WATERLOO

# Roadmap




UNIVERSITY OF WATERLOO

# Roadmap



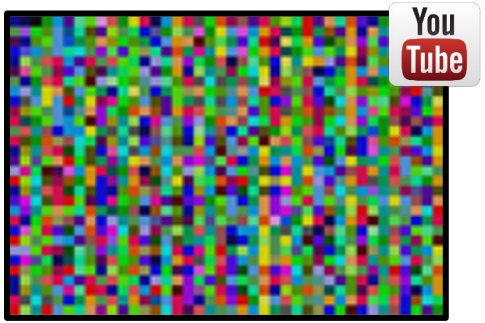

UNIVERSITY OF WATERLOO

# Can we Detect MPT Tools using ML-based Traffic Analysis?

- **The first extensive experimental study of the unobservability of covert channels produced by state-of-the-art MPT tools**

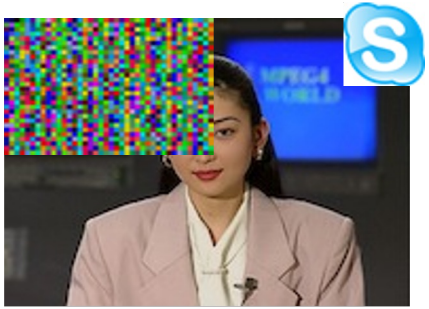
**System 1**
Facet [WPES'14]

Unidirectional (A/V)
Video Transmission




**System 2**
CovertCast [PETS'16]

Unidirectional (V)
Censored Websites Transmission



**System 3**
DeltaShaper [PETS'17]

Bidirectional (V)
Arbitrary Data Transmission

UNIVERSITY OF WATERLOO

# How was Unobservability Evaluation Performed?

- **Previous systems were evaluated with similarity-based classifiers**
  - **System 1** : Pearson's Chi-squared Test ($\chi 2$)
  - **System 2** : Kullback-Leibler Divergence (KL)
  - **System 3** : Earth Mover's Distance (EMD)

- **Feature sets are similar (quantized frequency distributions)**
  - **System 1** : Packet size bi-grams
  - **System 2** : Packet size, inter-arrival delay
  - **System 3** : Packet size, inter-arrival delay

# How Effective were Existing Detection Techniques?

| Protocol Tunneling System | [redacted] Classifier (acc%) | KL Classifier (acc%) | EMD Classifier (acc%) |
|---|---|---|---|
| **System 1** | **74.3** | 57.5 | 57.5 |

**$\chi^2$ is the most accurate classifier**

**Nearly random guess**
**Recent classifiers offer worse accuracy**

**None of the classifiers is able to detect covert channels with high accuracy**

UNIVERSITY OF WATERLOO

# ML-based Techniques To Detect Covert Channels

- **Assess the effectiveness of multiple decision tree-based classifiers**
  - Decision Trees
  - Random Forests
  - **eXtreme Gradient Boosting (XGBoost)**

Iterative generation of an ensemble of decision trees where new trees optimize previous predictions

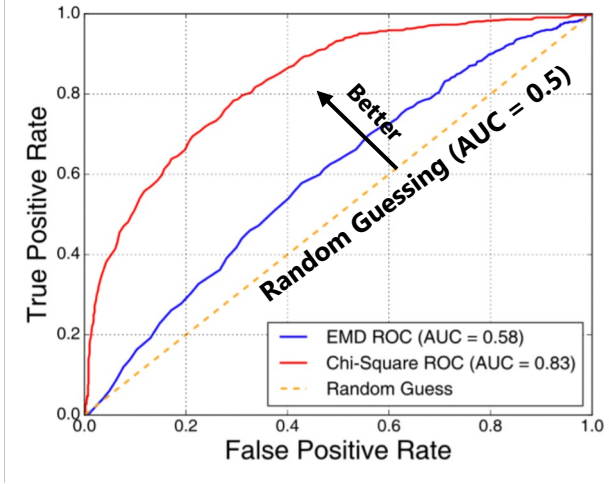

- **Models are easily interpretable**
- **Provide the ability to assess feature importance**

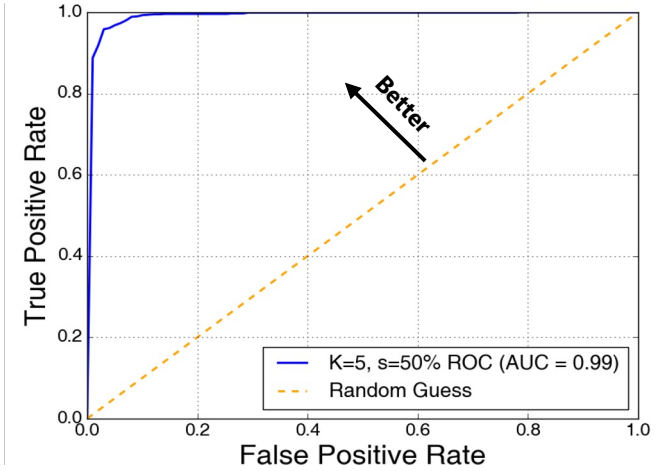
# Which Features Could an Adversary Use?

- **Feature set 1: summary statistics (ST)**
  - Total of 166 features, including simple statistics (e.g., max, min, percentiles), high order statistics (e.g., skew), and bursts

- **Feature set 2: quantized packet lengths (PL)**
  - Quantized PL frequency distribution for the flow carrying covert data
  - Each K size bin acts as an individual feature (K = 5 bytes)

# Detection of System 1




$\chi^2$: 90% TPR = **45% FPR**

**XGBoost-PL:** 90% TPR = **2% FPR**

**XGBoost-PL reduces the FPR when flagging the same amount of covert channels**

**The same trend can be observed for Systems 2 and 3**

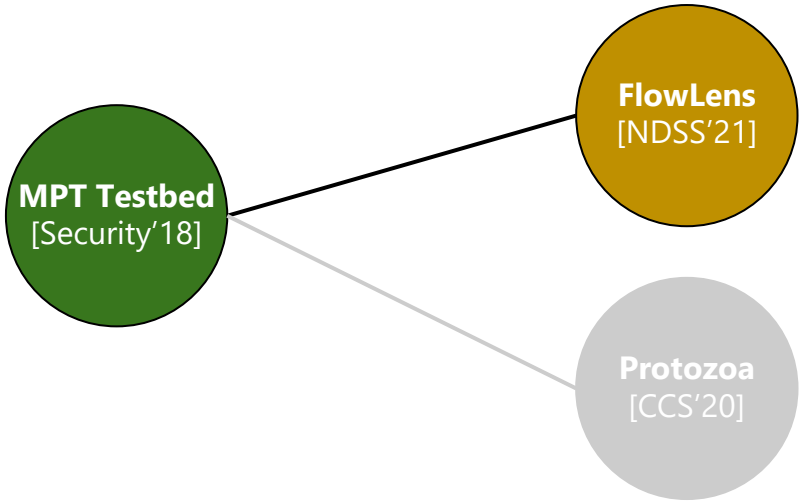
UNIVERSITY OF WATERLOO

# Censors can Accurately Detect MPT Tools

- **Previous unobservability claims were flawed**
  - ML-based techniques can detect MPT tools with high accuracy
  - Similarity-based provide optimistic unobservability guarantees

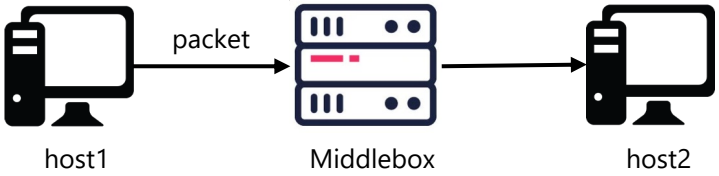
- **Can censors leverage these techniques in practice?**
  - In high-speed, large-scale networks

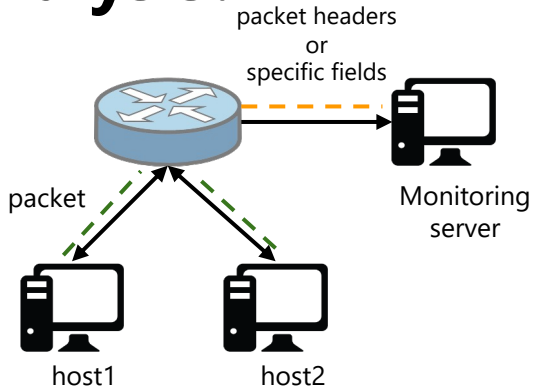
# Roadmap




UNIVERSITY OF WATERLOO

# Can Censors Efficiently Deploy ML-based Traffic Analysis?




Middleboxes

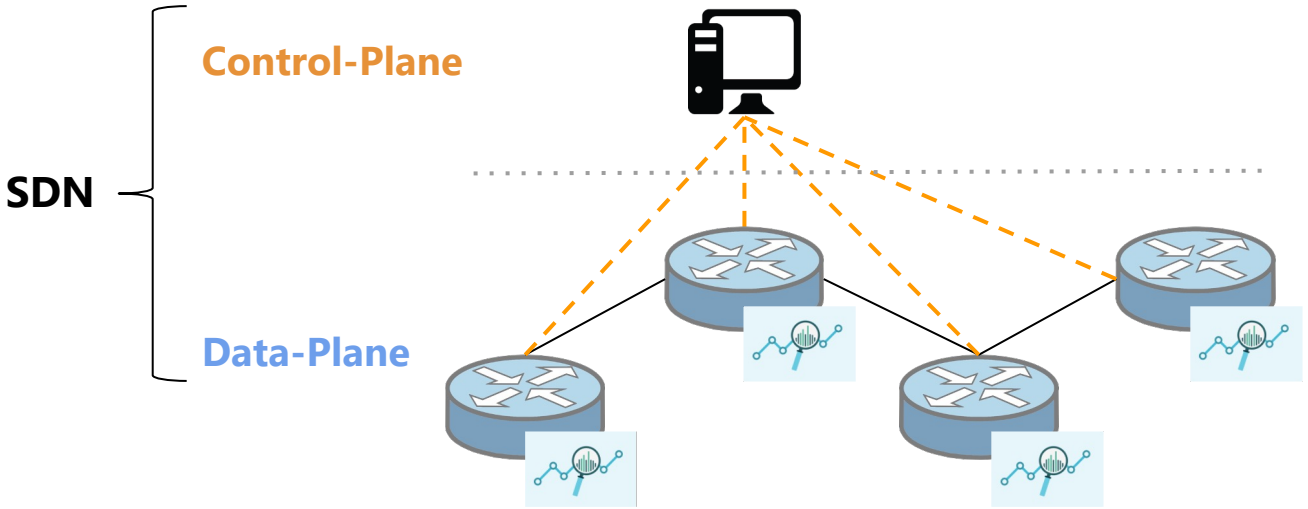
**Latency**
**Management Complexity**
**Infrastructure Costs**

Port Mirroring / Packet Aggregation

**Large Bandwidth Costs**

UNIVERSITY OF WATERLOO

# Programmable Switches Can Gather and Classify Packet Distributions Efficiently

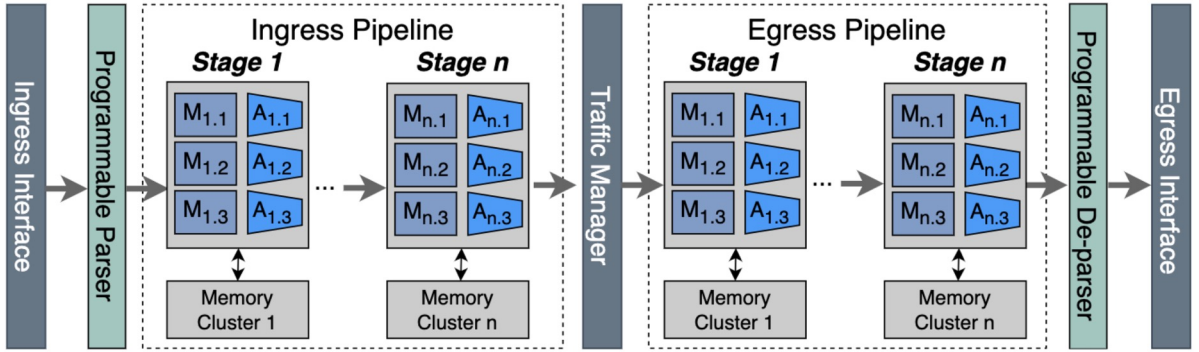



- Line speed
- No additional infrastructure
- Less management costs

# Programmable Switches

## Protocol Independent Switch Architecture (PISA)

- **Programmable packet parsing**
- **Process packets through match-action tables**
  - Arranged in stages
  - Perform an action upon matching some packet field
  - Actions may change packet headers or metadata



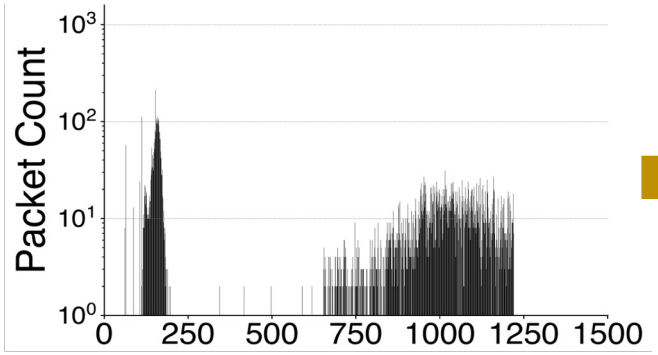
# Collecting Packet Distributions in the Switches is Hard

- **Stateful memory is severely limited**
  - ~100 MB SRAM
  - No memory for storing many flows
- **Packets must be processed at line speed (actions < 1ns)**
  - No multiplications or floating point operations
  - Existing packet distribution compression techniques **do not work**
- **We need a packet distribution representation that:**
  - Provides **high accuracy** and requires **small amount of memory**
  - Can be **implemented efficiently** in programmable switches

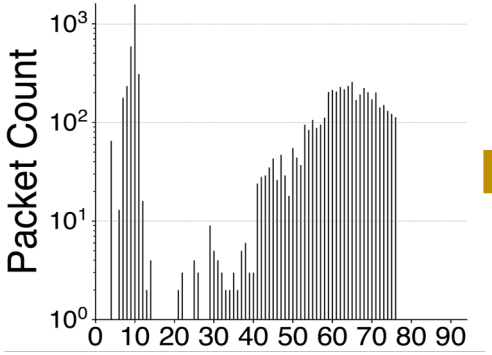
# Efficient Method to Compress Packet Distributions

- **Produce flow markers with two simple operators:**
  - **Quantization** - discretize the packet distribution into bins
  - **Truncation** - select the most relevant bins for classification

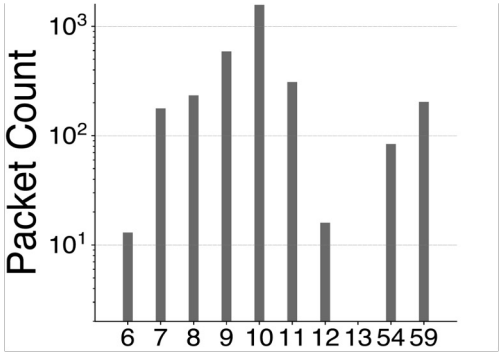
**Up to 150x size reduction**




**Raw packet size distribution**




**Quantized distribution**
**QL = 16**

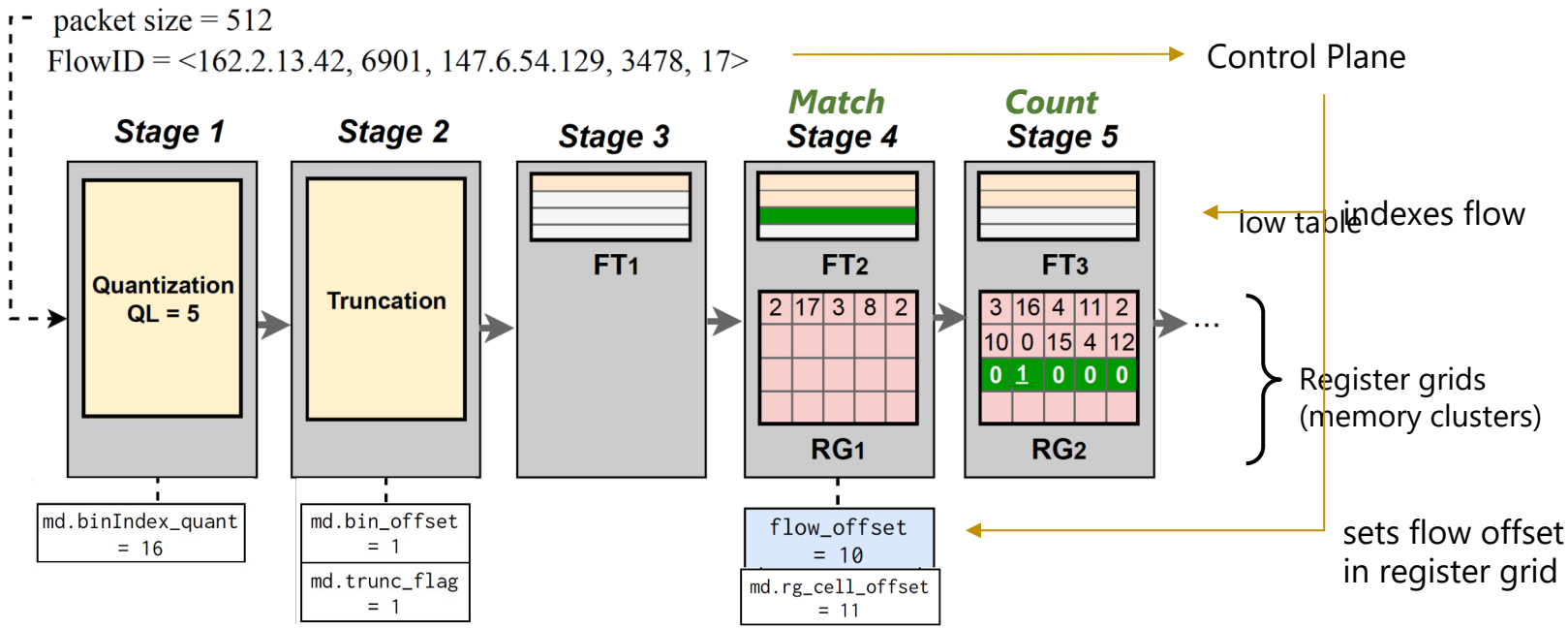



**Truncated distribution**
**Top-10 bins**

UNIVERSITY OF WATERLOO

# How are Flow Markers Collected in the Switch?

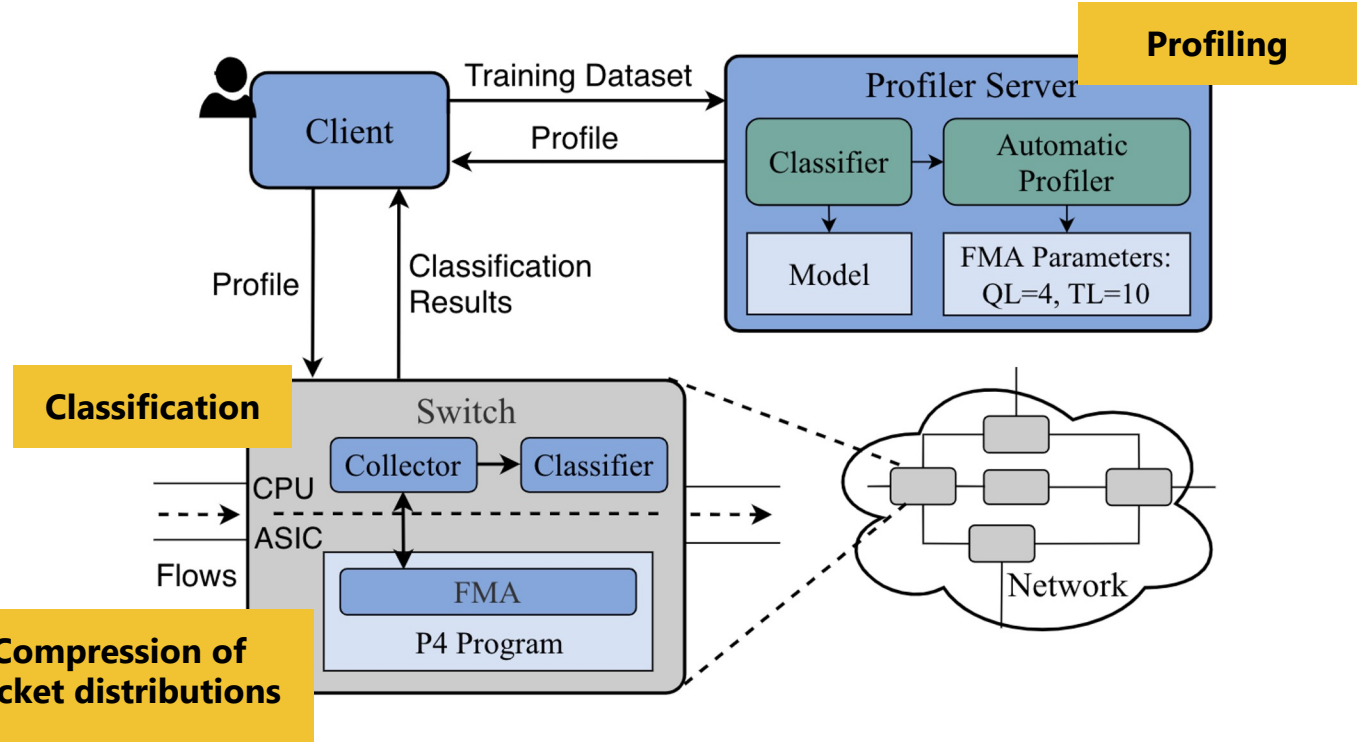
## Implementation in the Barefoot Tofino Switch



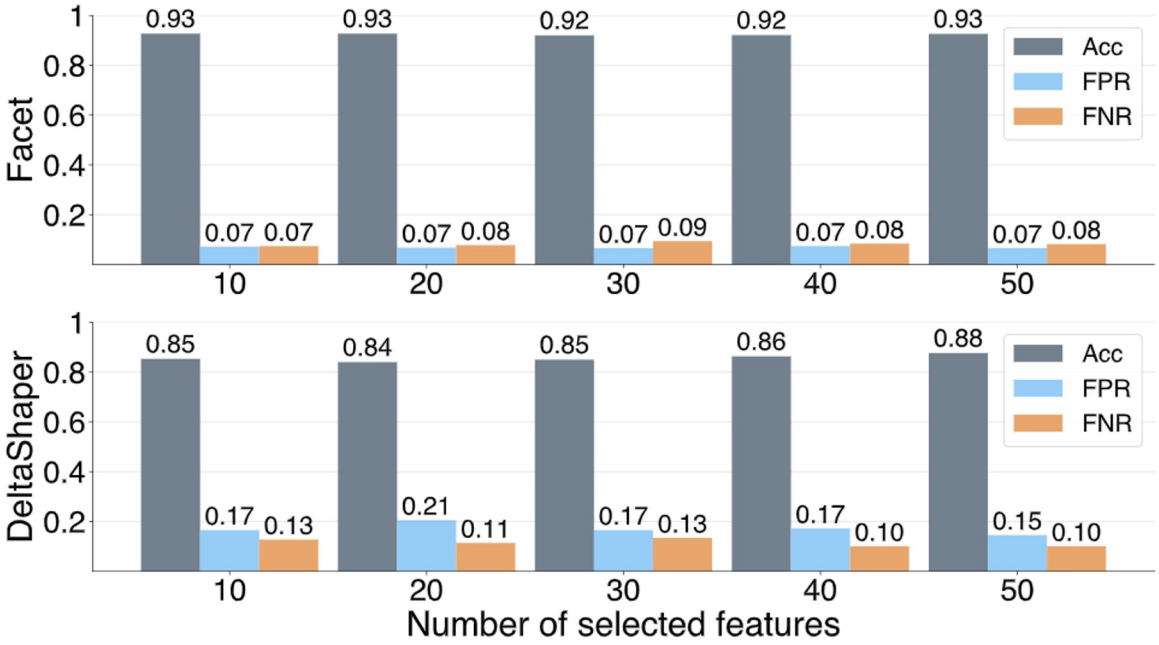
# Automatic Discovery of Quant. and Trunc. Parameters

- **Automate the configuration choice**
  - Large configuration space = **Quantization x Truncation**
- **Leverage Bayesian Optimization**
- **Three different criteria for selecting a configuration**
  - Smaller marker for target accuracy
  - Best accuracy given a size constraint
  - Compromise between marker size and accuracy

# FlowLens Architecture



# FlowLens can Accurately Detect MPT Tools




Full information = **3000B**
**Facet: 96% acc.**
**DeltaShaper: 87% acc**

Quant + Trunc = **20B**
**Facet: 93% acc.**
**DeltaShaper: 85% acc**

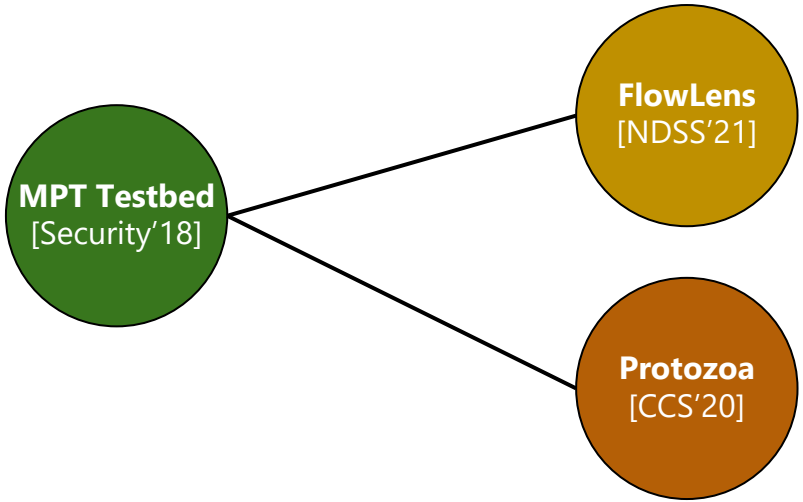
**Only up to - 3% accuracy**
**150x less memory**

UNIVERSITY OF WATERLOO

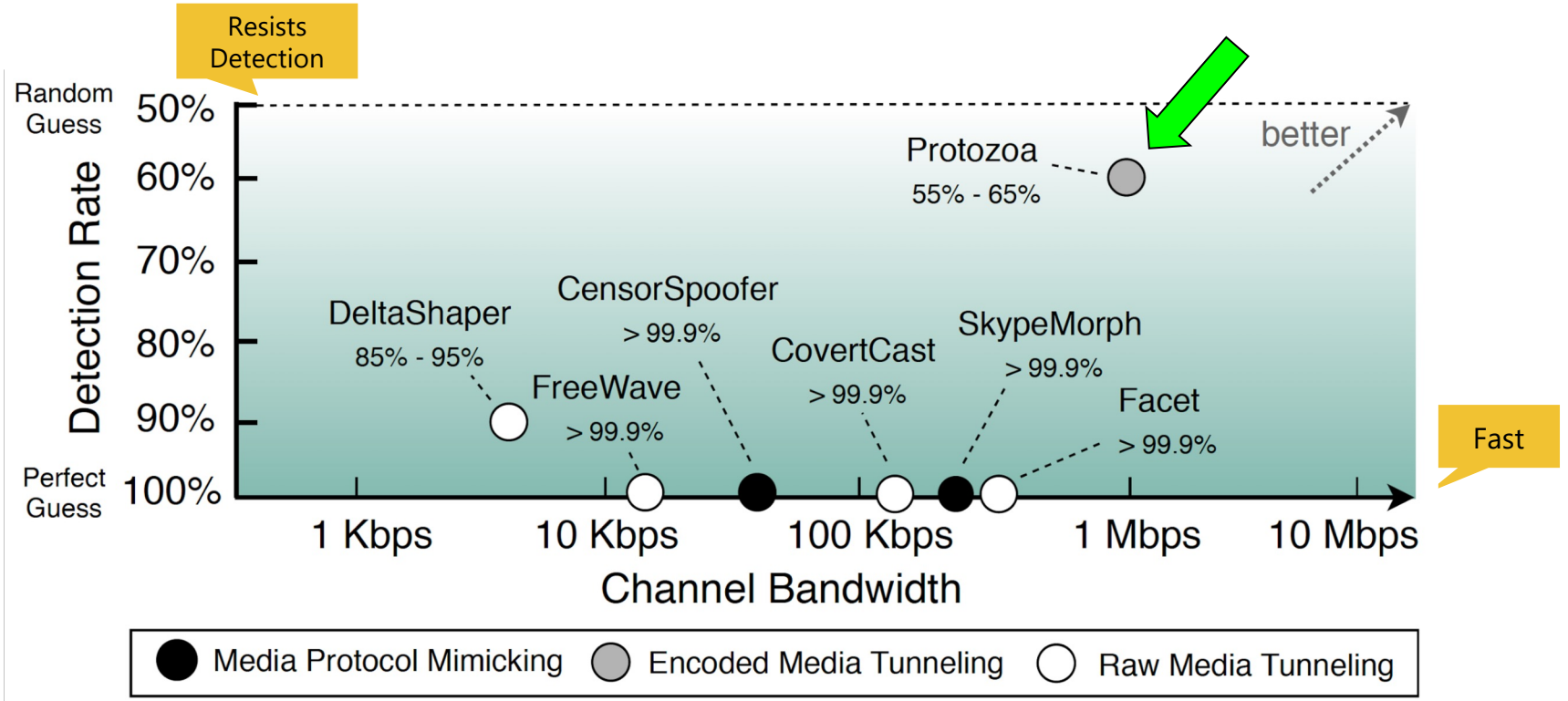
# Censors can Detect MPT Tools in Tbps Networks

- **FlowLens cuts down traffic analysis infrastructure costs**
  - Data collection and processing performed within programmable switches
- **FlowLens is able to collect flows at line speed in Tbps networks**
- **Censors can use FlowLens to detect MPT tools efficiently**
- **How can we devise an MPT tool that resists against traffic analysis?**

# Roadmap



# Can We Build a Fast and Unobservable MPT Tool?



# WebRTC

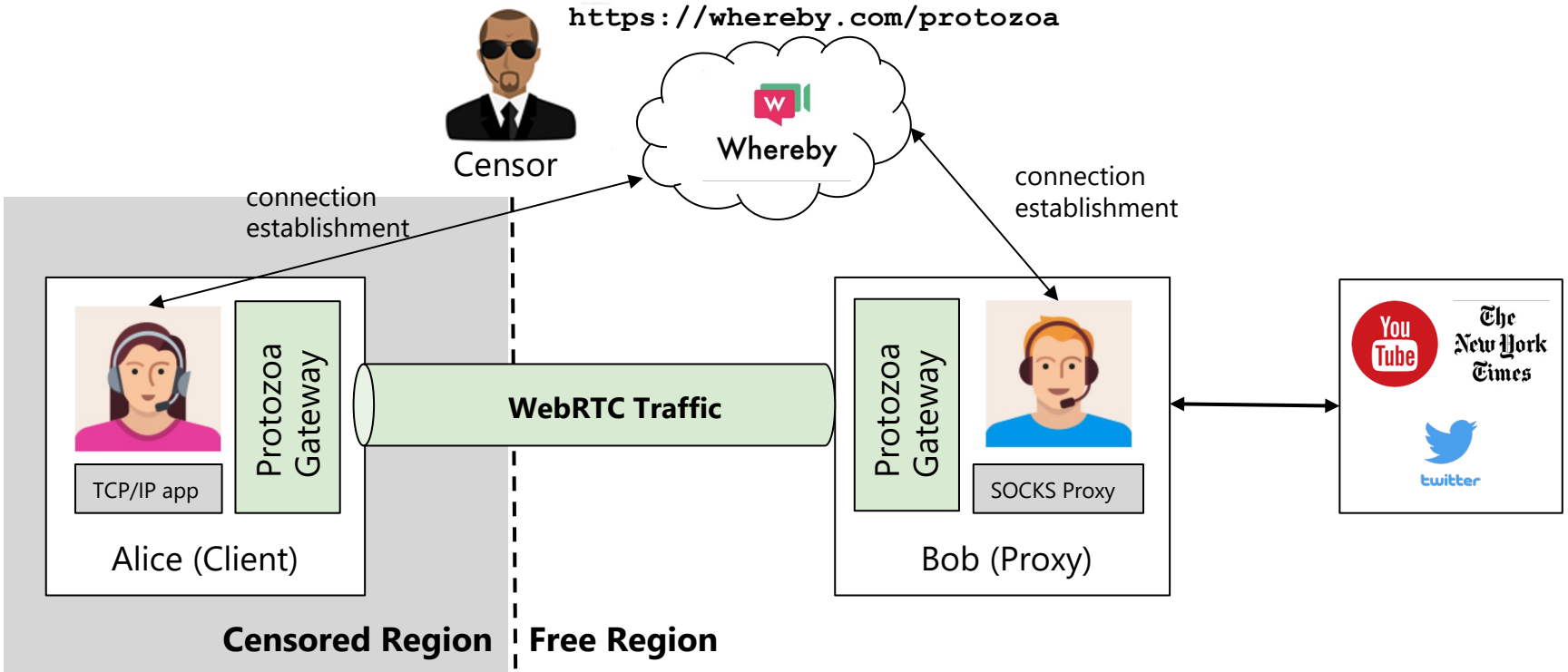
- **Framework that provides real-time communication capabilities**
  - Exposes a set of JavaScript APIs on **all major browsers**
  - Used by an **increasing number of trending applications**
  - **Open-source**



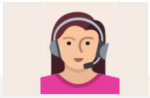




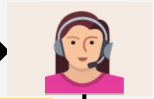

# Protozoa: a New Censorship Circumvention Tool



# How Does Protozoa Encode Covert Data?



Alice (Client)



Capture Video

Browser

WebRTC

VP8 Encoder

Encode

H P1 P2

VP8 Encoded Frame

Replace

Protozoa Hook

H Protozoa Data

Packetize & Encrypt

Transport

SRTP Packet

SRTP Packet SRTP Packet

IP Application e.g. curl

IP Packets

Protozoa Data Encoder


Browser

WebRTC

VP8 Decoder

Decode

H P1 P2

Jitter Buffer

Extract & Replace

H P1 P2

H Protozoa Data

Transport

Reassemble

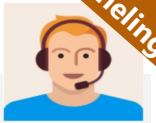
SRTP Packet

SRTP Packet SRTP Packet

SOCKS Proxy

IP Packets

Protozoa Data Decoder

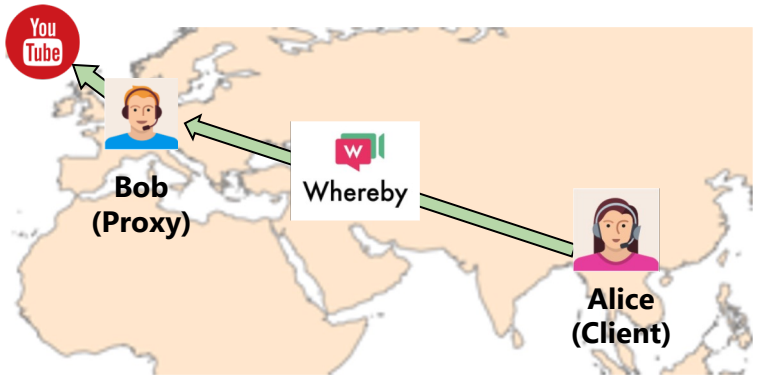


Bob (Proxy)

UNIVERSITY OF WATERLOO

# Protozoa in the Wild

| WebRTC Application | Reachability | | |
|---|---|---|---|
| | China | Russia | India |
| appr.tc | - | ✓ | ✓ |
| aws.amazon.com/chime | ✓ | ✓ | ✓ |
| codassium.com | ✓ | ✓ | ✓ |
| coderpad.io | ✓ | ✓ | ✓ |
| discordapp.com | - | ✓ | ✓ |
| gotomeeting.com | ✓ | ✓ | ✓ |
| hangouts.google.com | - | ✓ | ✓ |
| messenger.com | - | ✓ | ✓ |
| slack.com | ✓ | ✓ | ✓ |
| whereby.com | ✓ | ✓ | ✓ |

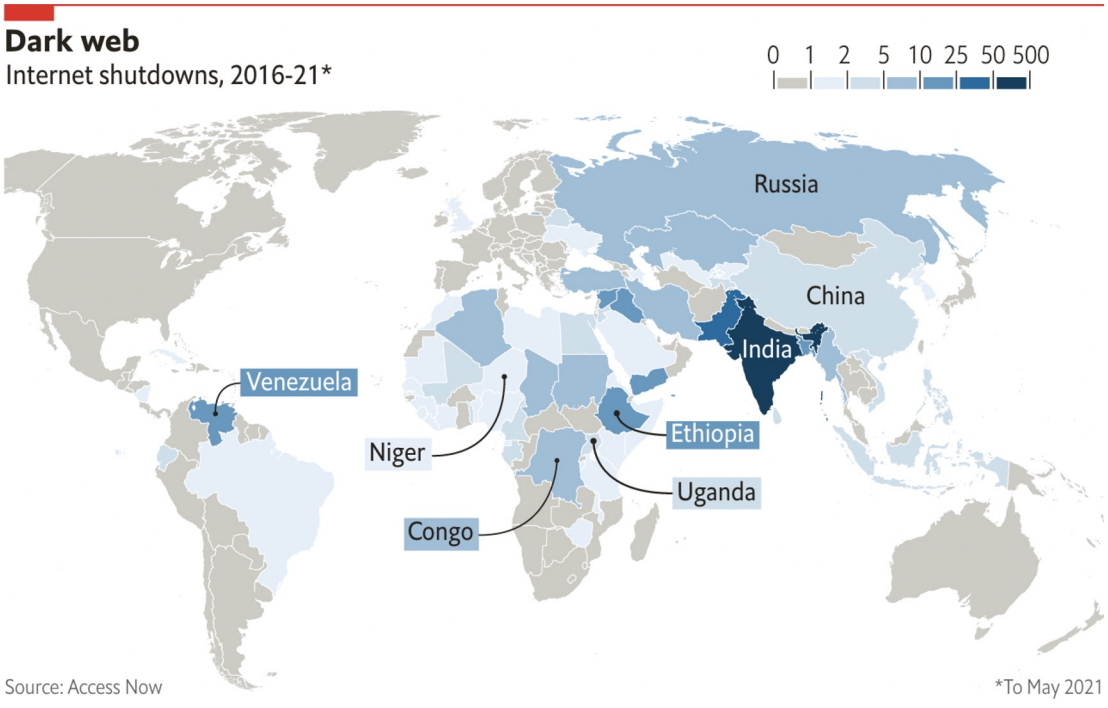



**Multiple WebRTC apps are available in countries known to experience Internet censorship**

**Protozoa makes it possible to access blocked content / services (e.g. YouTube)**

# Internet Blackouts

# How prevalent are Internet shutdowns?






UNIVERSITY OF WATERLOO

# How costly are these shutdowns?

**Data for 2022:**

| Rank | Country | Total Cost | Duration (Hrs) | Internet Users Affected |
|---|---|---|---|---|
| 1 | Russia | $21.59BN | 7,407 | 113,000,000 |
| 2 | Iran | $773M | 7,171 | 71,940,000 |
| 3 | Kazakhstan | $410.3M | 144 | 16,106,250 |
| 4 | Myanmar | $241.4M | 17,520 | 16,695,800 |
| 5 | Uzbekistan | $219.7M | 5,630 | 1,279,872 |
| 6 | India | $184.3M | 1,533 | 120,743,890 |
| 7 | Ethiopia | $145.8M | 8,760 | 1,022,983 |
| 8 | Nigeria | $82.7M | 287 | 104,400,000 |
| 9 | Cuba | $30.9M | 14 | 7,006,000 |
| 10 | Turkmenistan | $29M | 40 | 2,010,000 |

https://www.top10vpn.com/research/cost-of-internet-shutdowns/

# How to get around Internet blackouts?

- **Local P2P mesh networks**
  - FireChat, Bridgefy
  - Bluetooth, WiFi, etc.
  - **Anix [IEEE S&P'25]**
- **(Roaming) SIM cards**
  - Cellular infrastructure tends to remain available
- **Sneakernets**
  - Physical delivery of information (e.g., via USB drives or portable HDDs)

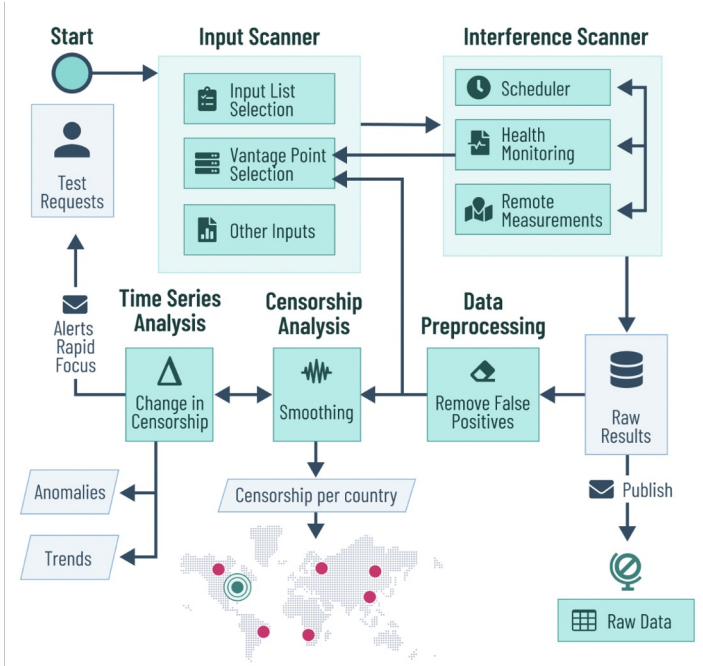
# Censorship measurements

# Why can't I access <website_name>.com?

- **Censorship measurements**
  - Can we get to example.com? If not, why?
- **Understand what is blocked**
  - Specific keywords/messages/topics
  - Websites or specific webpages **[WWW'21]**
  - Services & protocols
- **Understand how it is blocked**
  - Endpoint-based
  - Content-based
  - Behavior-based

# A deep dive on measurement techniques

- Measurement techniques
- Challenges on data validation
- Ethical considerations
- Global and longitudinal data collection
- Visualization of measurements






UNIVERSITY OF WATERLOO

# Censored topics & disinformation

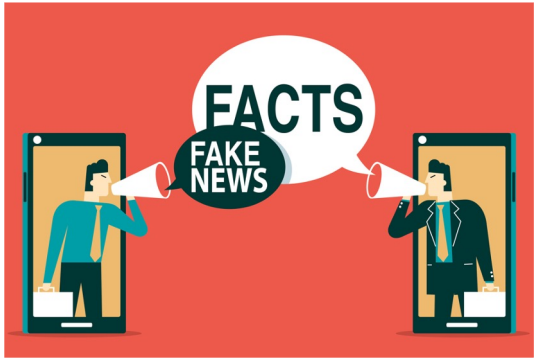
# Can we predict what will get blocked?

- Perennial topics deemed sensitive
- Curation of censorship test lists
- Correlation between blocked topics
- Evolution of a censored topic's definition [WPES'22]

Mentions of Tiananmen Square, June 4, and even the numbers "46" and "64" -- short for "4-6" and "6-4," or references to June 4 -- are studiously censored on the mainland.

# Can we filter the signal from the noise?

- **Disinformation leads to a confused audience**
  - Generates political and social discord
  - Benefits actors aiming to gain an advantage from such confusion
- **How are disinformation campaigns orchestrated?**
- **How to tackle disinformation campaigns?**






UNIVERSITY OF WATERLOO

# References


*Diogo Barradas, Nuno Santos, Luís Rodrigues*
**DeltaShaper: Enabling Unobservable Censorship-resistant TCP Tunneling over Videoconferencing Streams**
In *Proc. on Privacy Enhancing Technologies (PoPETS)*, **2017**

*Diogo Barradas, Nuno Santos, Luís Rodrigues*
**Effective Detection of Multimedia Protocol Tunneling using Machine Learning**
In *Proc. of USENIX Security Symposium (Security)*, **2018**

*Diogo Barradas, Nuno Santos, Luís Rodrigues, Salvatore Signorello, Fernando Ramos, André Madeira*
**FlowLens: Enabling Efficient Flow Classification for ML-based Network Security Applications**
In *Proc. of Network and Distributed Systems Symposium (NDSS)*, **2021**

*Diogo Barradas, Nuno Santos, Luís Rodrigues, Vítor Nunes*
**Poking a Hole in the Wall: Efficient Censorship-Resistant Internet Communications by Parasitizing on WebRTC**
In *Proc. of ACM Conference on Computer and Communications Security (CCS)*, **2020**

*Diogo Barradas, Nuno Santos*
**Towards a Scalable Censorship-Resistant Overlay Network based on WebRTC Covert Channels**
In *Proc. of 1st Intl. Workshop on Digital Infrastructure for Common Good (DICG)*, **2020**

Raymond Rambert, Zachary Weinberg, *Diogo Barradas, Nicolas Christin*
**Chinese Wall or Swiss Cheese? Keyword filtering in the Great Firewall of China**
In *Proc. of the 30th The Web Conference (TheWebConf)*, **2021**

Sina Kamali, Diogo Barradas.
**Anix: Anonymous Blackout-Resistant Microblogging with Message Endorsing.**
*In Proc. of the 46th IEEE Symposium on Security and Privacy*, **2025**

Asim Waheed*, Sara Qunaibi*, Diogo Barradas, Zachary Weinberg. (co-first authors)
**Darwin's Theory Of Censorship: Analysing the Evolution of Censored Topics with Dynamic Topic Models.**
In Proc. of the Workshop on Privacy in the Electronic Society, **2022**


UNIVERSITY OF WATERLOO