Discussed papers weeks 4+5

TDTS21, vt 2021

Niklas Carlsson, *Linköping University*

How can we measure the Internet?

- Active measurements
  - Probes: Traceroute, ping, packet trains
  - Application simulation
- Passive measurement
  - Logs (WWW)
  - Monitors, sniffers

- What is measured?
  - Everything …

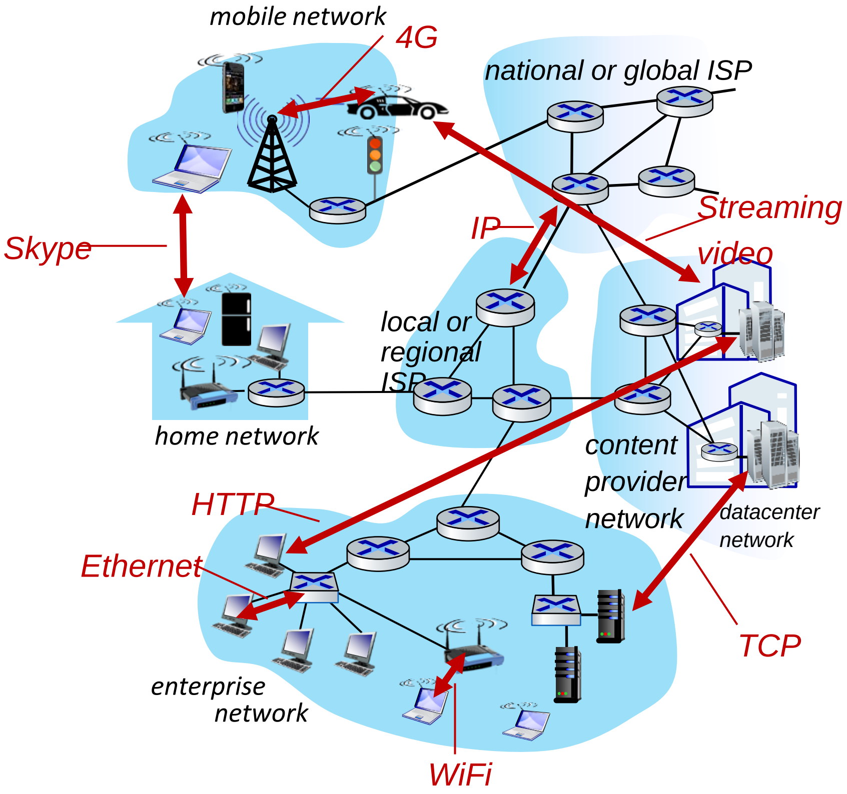
“you can't manage what you can't measure”

– quote claimed by many/several

- Where to measure? …

The Internet: a "nuts and bolts" view

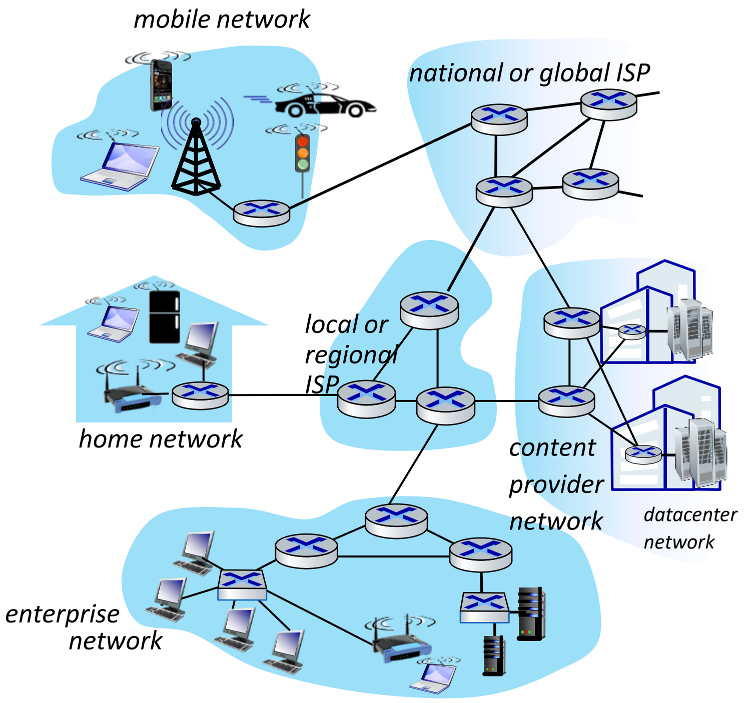
- *Internet:* "network of networks"
    - Interconnected ISPs
- ***protocols** are everywhere*
    - *control sending, receiving of messages*
    - *e.g., HTTP (Web), streaming video, Skype, TCP, IP, WiFi, 4G, Ethernet*
- ***Internet standards***
    - *RFC: Request for Comments*
    - *IETF: Internet Engineering Task Force*



Slide from intro in in TDDE35/TDTS11/TDTS04/TDTS06 …

Internet structure: a “network of networks”

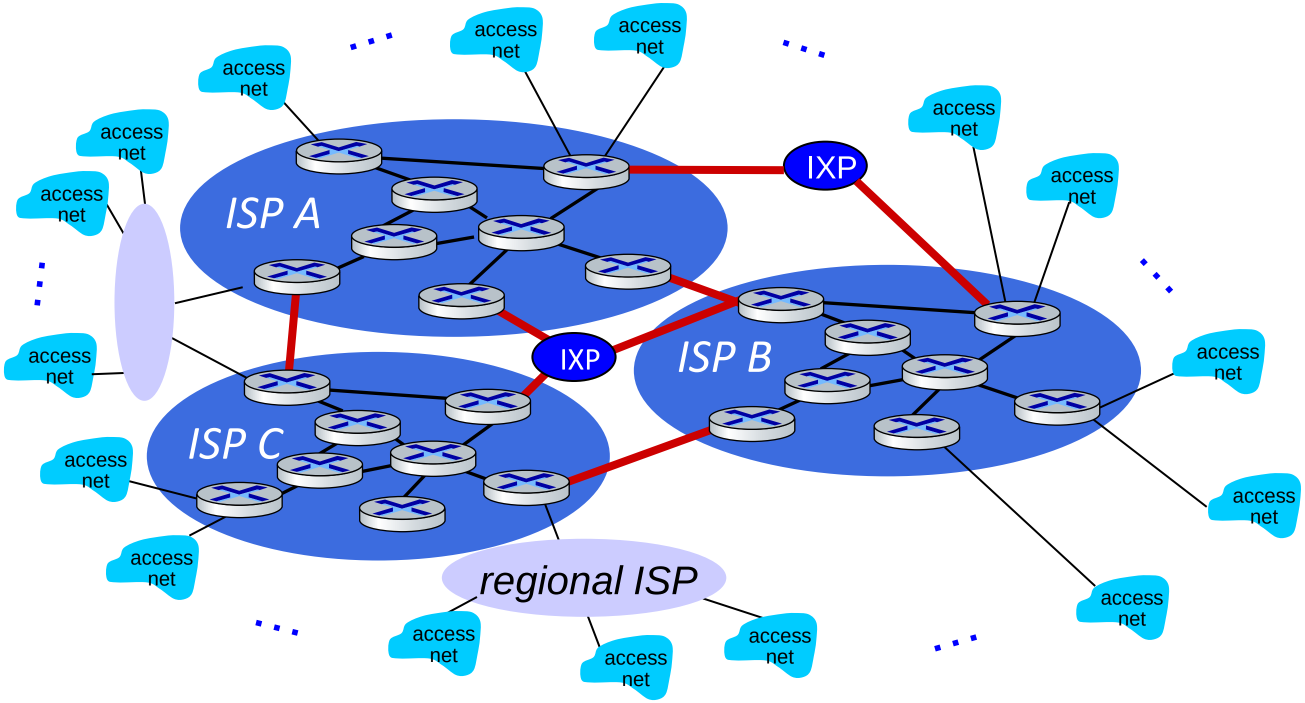
- hosts connect to Internet via access Internet Service Providers (ISPs)
- access ISPs in turn must be interconnected
  - so that *any* two hosts *(anywhere!)* can send packets to each other
- resulting network of networks is very complex
  - evolution driven by economics, national policies



Slide from intro in in TDDE35/TDTS11/TDTS04/TDTS06 …

Internet structure: a “network of networks”

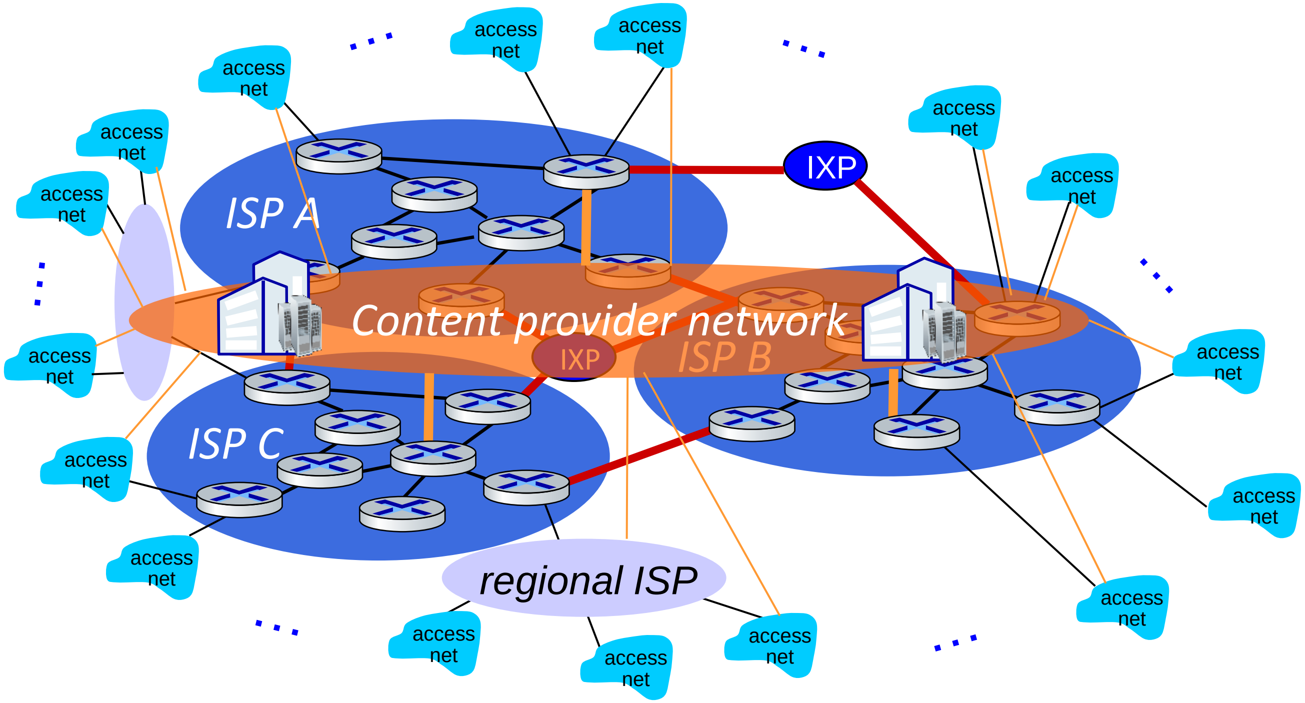
… and regional networks may arise to connect access nets to ISPs



Slide from intro in in TDDE35/TDTS11/TDTS04/TDTS06 …

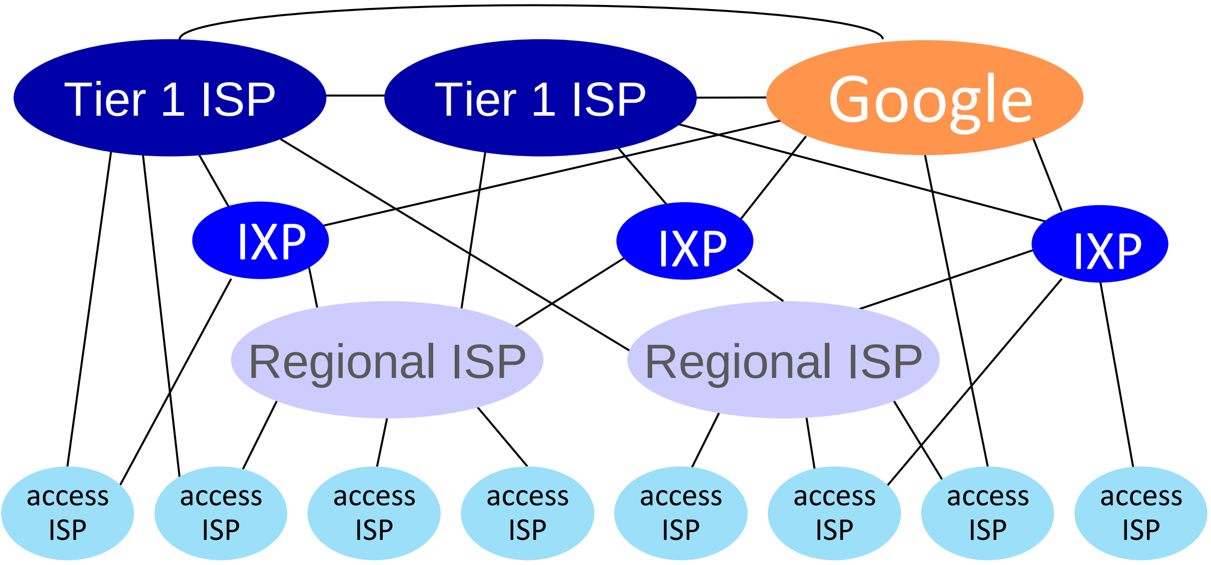
Internet structure: a “network of networks”

… and content provider networks (e.g., Google, Microsoft, Akamai) may run their own network, to bring services, content close to end users



Slide from intro in in TDDE35/TDTS11/TDTS04/TDTS06 …

Internet structure: a “network of networks”



At “center”: small # of well-connected large networks

- “tier-1” commercial ISPs (e.g., Level 3, Sprint, AT&T, NTT), national & international coverage
- content provider networks (e.g., Google, Facebook): private network that connects its data centers to Internet, often bypassing tier-1, regional ISPs

When should we measure the Internet?

- Diurnal and weekly traffic cycles
- Time scales depend on “what”, “how”, and “why”?
- Passive measurement are typically continuous
  - Can generate **huge** datasets
  - Log access problems
  - Privacy concerns
- Active measurements are typically discrete
  - Important characteristics can be missed
  - Probes can be filtered and/or detected

Publishing Internet Measurement Studies

- All major networking + security conferences & journals accept measurement papers
    - ACM SIGCOMM, IEEE INFOCOM, ACM SIGMETRICS
    - ACM CCS, NDSS, IEEE S&P, Usenix Security
    - IEEE/ACM ToN, IEEE TPDS
- Dedicated meetings
    - **ACM Internet Measurement Conf. (IMC)**
    - Passive & Active Measurements Conf. (PAM)
    - (TMA)

Carl Magnus

Why this paper?

What you liked most?

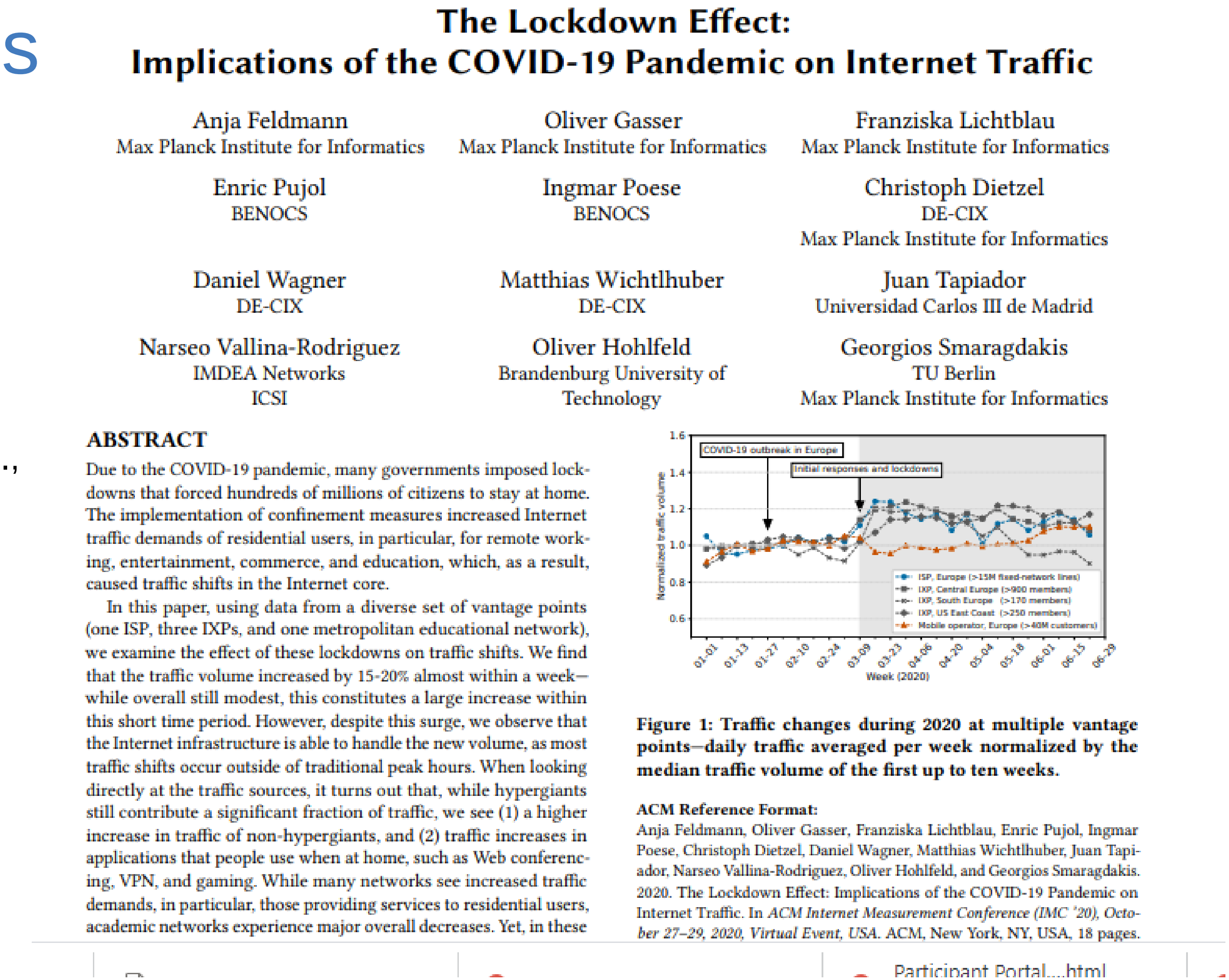
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



The Lockdown Effect: Implications of the COVID-19 Pandemic on Internet Traffic

Anja Feldmann
Max Planck Institute for Informatics

Oliver Gasser
Max Planck Institute for Informatics

Franziska Lichtblau
Max Planck Institute for Informatics

Enric Pujol
BENOCS

Ingmar Poese
BENOCS

Christoph Dietzel
DE-CIX
Max Planck Institute for Informatics

Daniel Wagner
DE-CIX

Matthias Wichtlhuber
DE-CIX

Juan Tapiador
Universidad Carlos III de Madrid

Narseo Vallina-Rodriguez
IMDEA Networks
ICSI

Oliver Hohlfeld
Brandenburg University of Technology

Georgios Smaragdakis
TU Berlin
Max Planck Institute for Informatics

ABSTRACT

Due to the COVID-19 pandemic, many governments imposed lockdowns that forced hundreds of millions of citizens to stay at home. The implementation of confinement measures increased Internet traffic demands of residential users, in particular, for remote working, entertainment, commerce, and education, which, as a result, caused traffic shifts in the Internet core.

In this paper, using data from a diverse set of vantage points (one ISP, three IXPs, and one metropolitan educational network), we examine the effect of these lockdowns on traffic shifts. We find that the traffic volume increased by 15-20% almost within a week—while overall still modest, this constitutes a large increase within this short time period. However, despite this surge, we observe that the Internet infrastructure is able to handle the new volume, as most traffic shifts occur outside of traditional peak hours. When looking directly at the traffic sources, it turns out that, while hypergiants still contribute a significant fraction of traffic, we see (1) a higher increase in traffic of non-hypergiants, and (2) traffic increases in applications that people use when at home, such as Web conferencing, VPN, and gaming. While many networks see increased traffic demands, in particular, those providing services to residential users, academic networks experience major overall decreases. Yet, in these

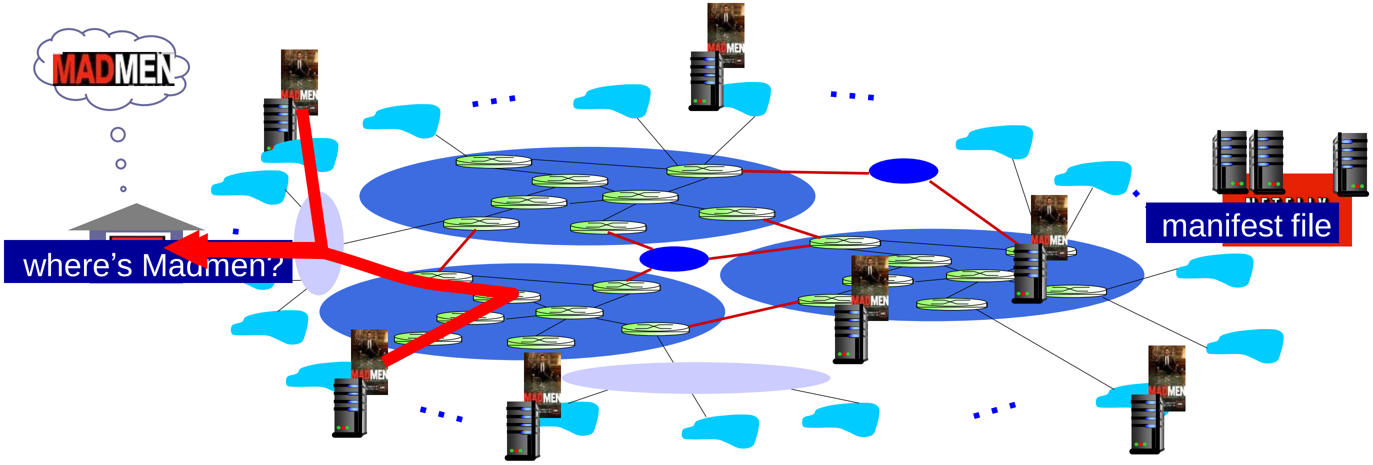
1.6
1.4
1.2
1.0
0.8
0.6
Normalized traffic volume
COVID-19 outbreak in Europe
Initial responses and lockdowns
ISP, Europe (>15M fixed-network lines)
IXP, Central Europe (>900 members)
IXP, South Europe (>170 members)
IXP, US East Coast (>250 members)
Mobile operator, Europe (>40M customers)
01-01
01-13
01-27
02-10
02-24
03-09
03-23
04-06
04-20
05-04
05-18
06-01
06-15
06-29
Week (2020)
Participant Portal....html

Figure 1: Traffic changes during 2020 at multiple vantage points—daily traffic averaged per week normalized by the median traffic volume of the first up to ten weeks.

ACM Reference Format:
Anja Feldmann, Oliver Gasser, Franziska Lichtblau, Enric Pujol, Ingmar Poese, Christoph Dietzel, Daniel Wagner, Matthias Wichtlhuber, Juan Tapiador, Narseo Vallina-Rodriguez, Oliver Hohlfeld, and Georgios Smaragdakis. 2020. The Lockdown Effect: Implications of the COVID-19 Pandemic on Internet Traffic. In *ACM Internet Measurement Conference (IMC '20), October 27–29, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 18 pages.

Content distribution networks (CDNs)

- CDN: stores copies of content (e.g. MADMEN) at CDN nodes
- subscriber requests content, service provider returns manifest
  - *using manifest, client* retrieves content at highest supportable rate
  - may choose different rate or copy if network path congested



Philip

Why this paper?

What you liked most?

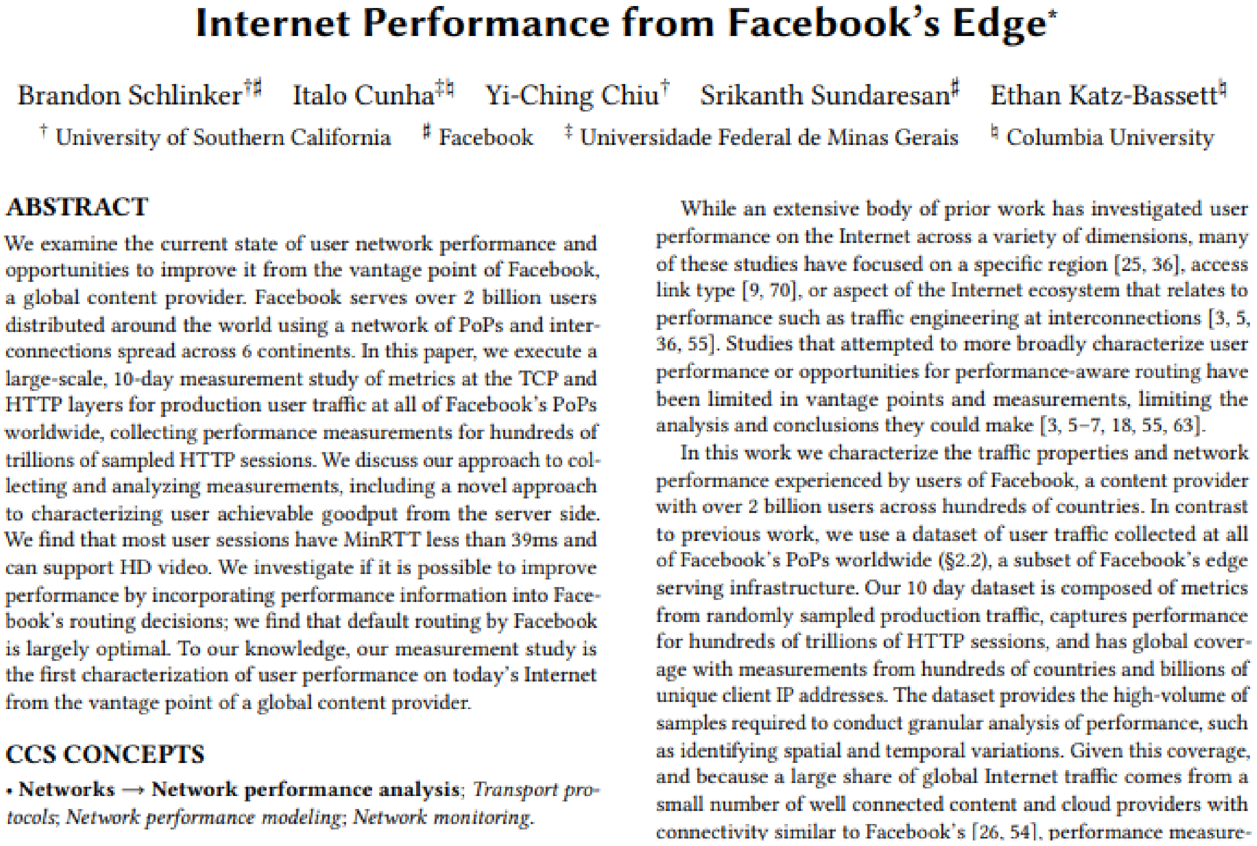
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



Internet Performance from Facebook's Edge*

Brandon Schlinker[†♯] Italo Cunha[‡♮] Yi-Ching Chiu[†] Srikanth Sundaresan[♯] Ethan Katz-Bassett[♮]

[†] University of Southern California [♯] Facebook [‡] Universidade Federal de Minas Gerais [♮] Columbia University

ABSTRACT

We examine the current state of user network performance and opportunities to improve it from the vantage point of Facebook, a global content provider. Facebook serves over 2 billion users distributed around the world using a network of PoPs and interconnections spread across 6 continents. In this paper, we execute a large-scale, 10-day measurement study of metrics at the TCP and HTTP layers for production user traffic at all of Facebook's PoPs worldwide, collecting performance measurements for hundreds of trillions of sampled HTTP sessions. We discuss our approach to collecting and analyzing measurements, including a novel approach to characterizing user achievable goodput from the server side. We find that most user sessions have MinRTT less than 39ms and can support HD video. We investigate if it is possible to improve performance by incorporating performance information into Facebook's routing decisions; we find that default routing by Facebook is largely optimal. To our knowledge, our measurement study is the first characterization of user performance on today's Internet from the vantage point of a global content provider.

CCS CONCEPTS

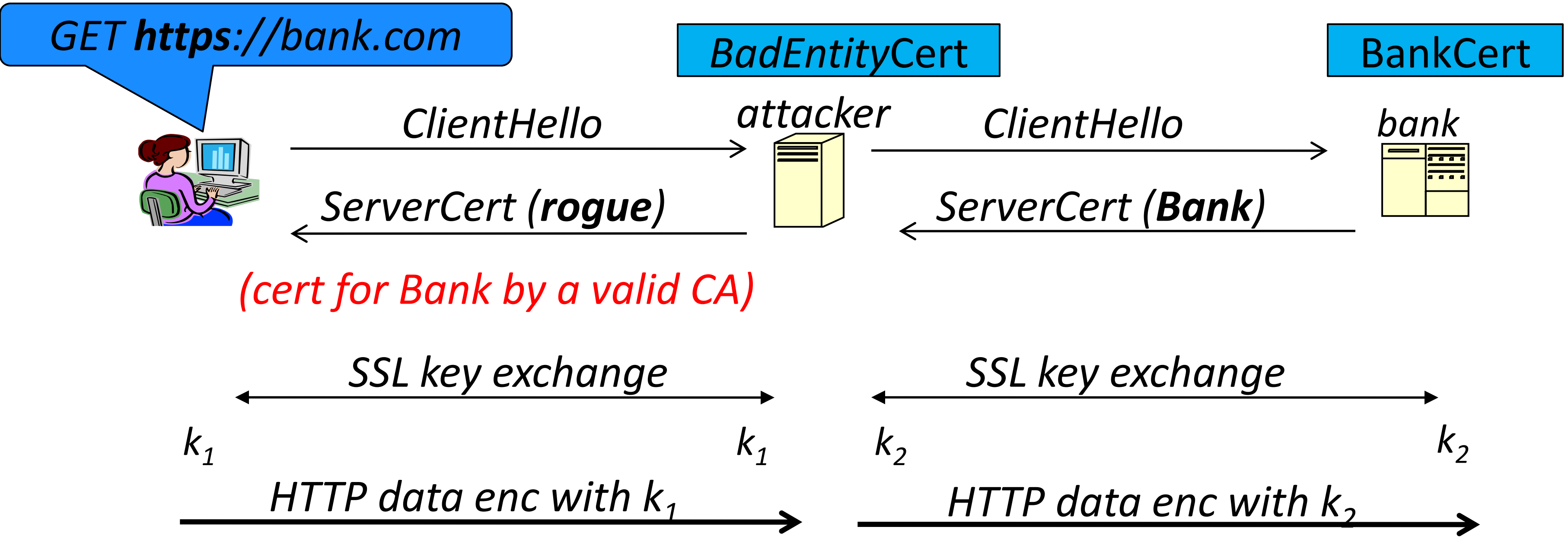
• **Networks** → **Network performance analysis**; *Transport protocols; Network performance modeling; Network monitoring.*

While an extensive body of prior work has investigated user performance on the Internet across a variety of dimensions, many of these studies have focused on a specific region [25, 36], access link type [9, 70], or aspect of the Internet ecosystem that relates to performance such as traffic engineering at interconnections [3, 5, 36, 55]. Studies that attempted to more broadly characterize user performance or opportunities for performance-aware routing have been limited in vantage points and measurements, limiting the analysis and conclusions they could make [3, 5–7, 18, 55, 63].

In this work we characterize the traffic properties and network performance experienced by users of Facebook, a content provider with over 2 billion users across hundreds of countries. In contrast to previous work, we use a dataset of user traffic collected at all of Facebook's PoPs worldwide (§2.2), a subset of Facebook's edge serving infrastructure. Our 10 day dataset is composed of metrics from randomly sampled production traffic, captures performance for hundreds of trillions of HTTP sessions, and has global coverage with measurements from hundreds of countries and billions of unique client IP addresses. The dataset provides the high-volume of samples required to conduct granular analysis of performance, such as identifying spatial and temporal variations. Given this coverage, and because a large share of global Internet traffic comes from a small number of well connected content and cloud providers with connectivity similar to Facebook's [26, 54], performance measure-

Slide from TDDE53 …

Man in the middle attack using rogue cert



Attacker proxies data between user and bank.
Sees all traffic and can modify data at will.

Jacob W.

Why this paper?

What you liked most?

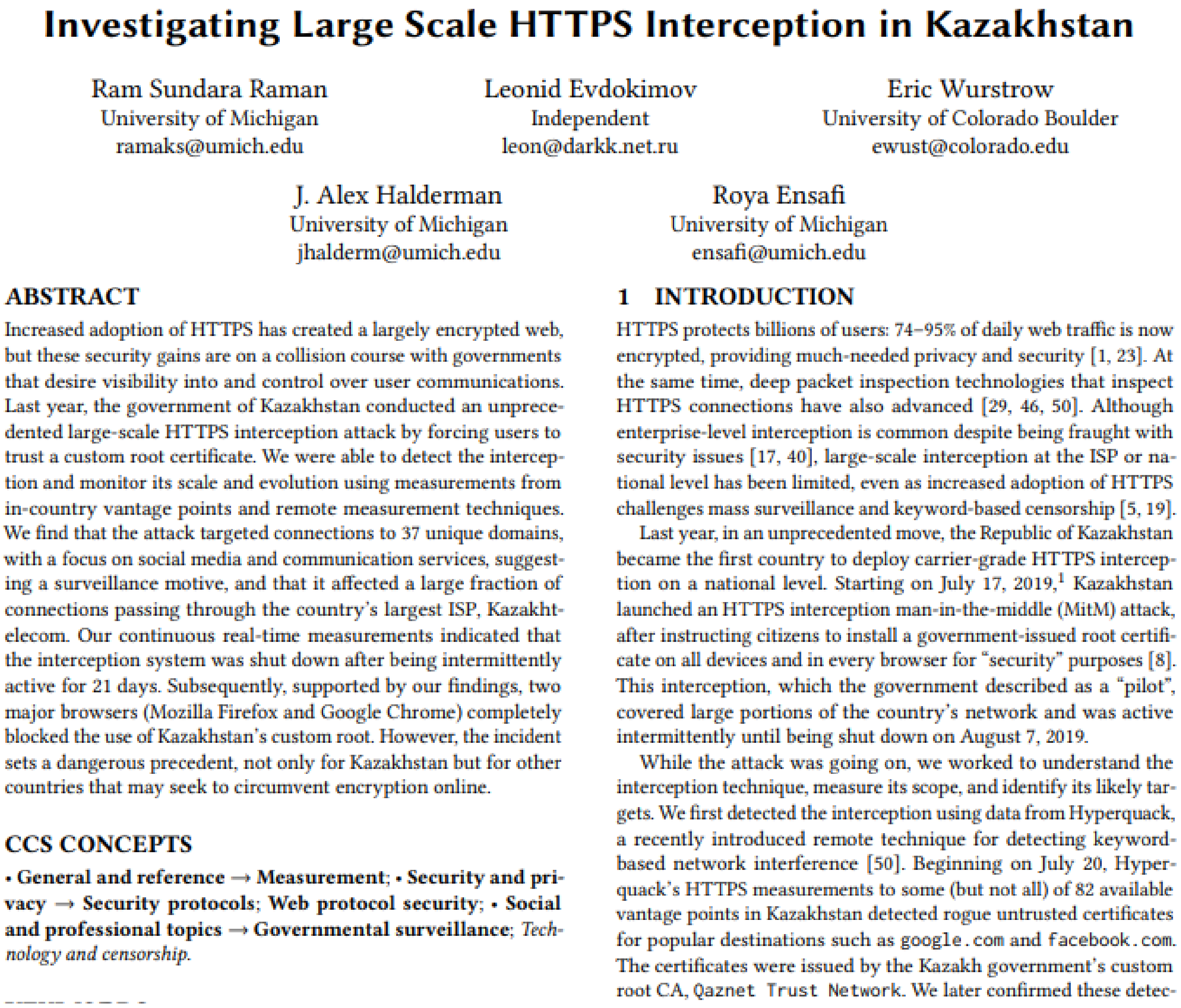
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



Investigating Large Scale HTTPS Interception in Kazakhstan

Ram Sundara Raman
University of Michigan
ramaks@umich.edu

Leonid Evdokimov
Independent
leon@darkk.net.ru

Eric Wurstrow
University of Colorado Boulder
ewust@colorado.edu

J. Alex Halderman
University of Michigan
jhalderm@umich.edu

Roya Ensafi
University of Michigan
ensafi@umich.edu

ABSTRACT

Increased adoption of HTTPS has created a largely encrypted web, but these security gains are on a collision course with governments that desire visibility into and control over user communications. Last year, the government of Kazakhstan conducted an unprecedented large-scale HTTPS interception attack by forcing users to trust a custom root certificate. We were able to detect the interception and monitor its scale and evolution using measurements from in-country vantage points and remote measurement techniques. We find that the attack targeted connections to 37 unique domains, with a focus on social media and communication services, suggesting a surveillance motive, and that it affected a large fraction of connections passing through the country's largest ISP, Kazakhtelecom. Our continuous real-time measurements indicated that the interception system was shut down after being intermittently active for 21 days. Subsequently, supported by our findings, two major browsers (Mozilla Firefox and Google Chrome) completely blocked the use of Kazakhstan's custom root. However, the incident sets a dangerous precedent, not only for Kazakhstan but for other countries that may seek to circumvent encryption online.

CCS CONCEPTS

• **General and reference** → **Measurement**; • **Security and privacy** → **Security protocols**; **Web protocol security**; • **Social and professional topics** → **Governmental surveillance**; *Technology and censorship*.

1 INTRODUCTION

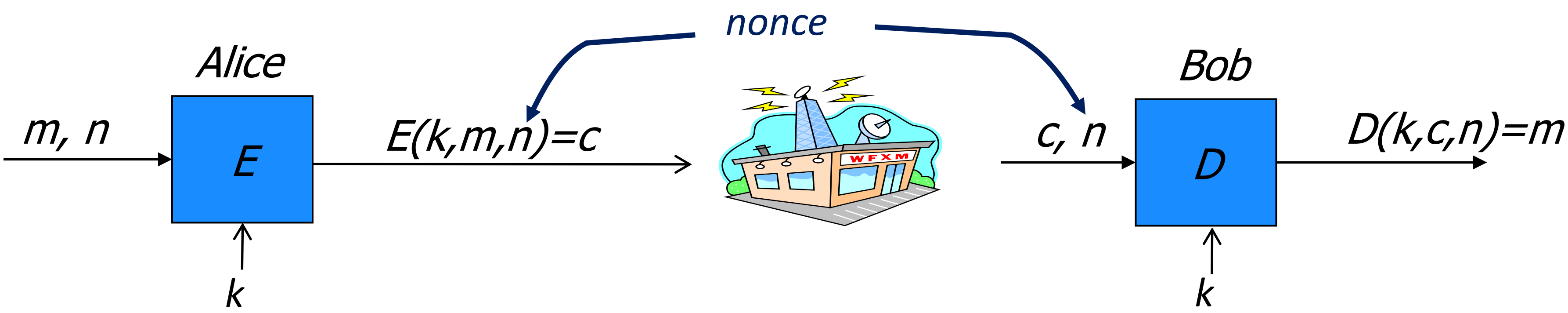
HTTPS protects billions of users: 74–95% of daily web traffic is now encrypted, providing much-needed privacy and security [1, 23]. At the same time, deep packet inspection technologies that inspect HTTPS connections have also advanced [29, 46, 50]. Although enterprise-level interception is common despite being fraught with security issues [17, 40], large-scale interception at the ISP or national level has been limited, even as increased adoption of HTTPS challenges mass surveillance and keyword-based censorship [5, 19].

Last year, in an unprecedented move, the Republic of Kazakhstan became the first country to deploy carrier-grade HTTPS interception on a national level. Starting on July 17, 2019,[1] Kazakhstan launched an HTTPS interception man-in-the-middle (MitM) attack, after instructing citizens to install a government-issued root certificate on all devices and in every browser for "security" purposes [8]. This interception, which the government described as a "pilot", covered large portions of the country's network and was active intermittently until being shut down on August 7, 2019.

While the attack was going on, we worked to understand the interception technique, measure its scope, and identify its likely targets. We first detected the interception using data from Hyperquack, a recently introduced remote technique for detecting keyword-based network interference [50]. Beginning on July 20, Hyperquack's HTTPS measurements to some (but not all) of 82 available vantage points in Kazakhstan detected rogue untrusted certificates for popular destinations such as `google.com` and `facebook.com`. The certificates were issued by the Kazakh government's custom root CA, `Qaznet Trust Network`. We later confirmed these detec-

Recent survey overview of TLS interceptions: https://arxiv.org/pdf/2010.16388.pdf

Building block: symmetric cipher



nonce

Alice

Bob

m, n

E

k

E(k,m,n)=c

c, n

D

k

D(k,c,n)=m

E, D: cipher   k: shared secret key (e.g., 128 bits)

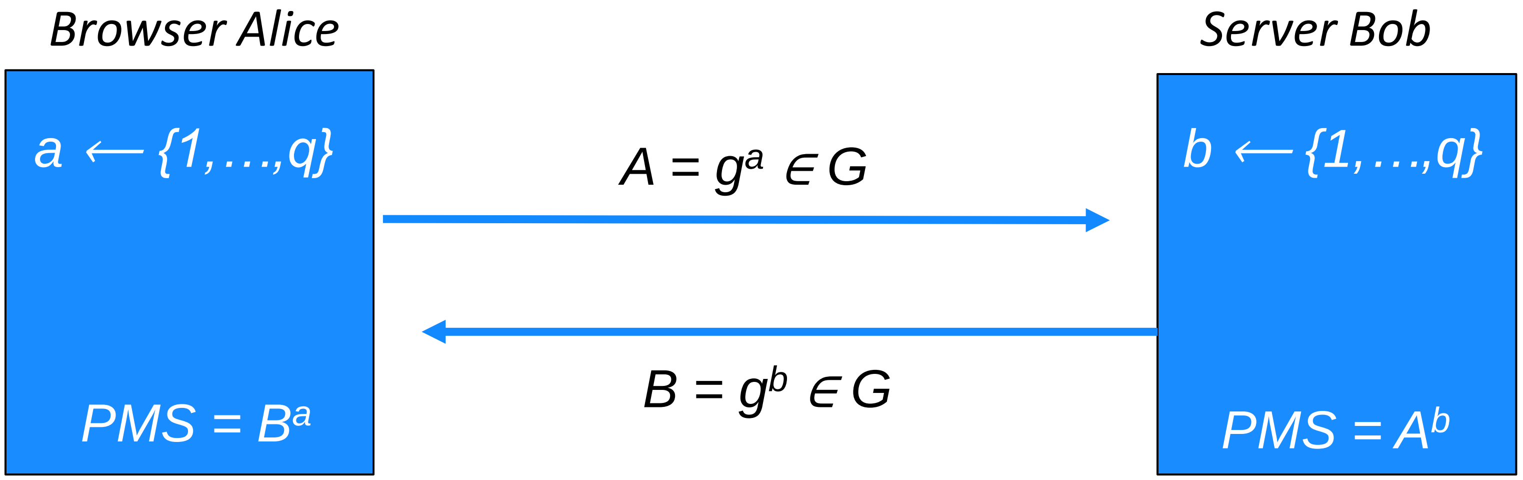
m, c: plaintext, ciphertext   n: nonce (non-repeating)

Slide from TDDE53 …

TLS overview: (1) DH key exchange

Anonymous key exchange secure against eavesdropping:

The Diffie-Hellman protocol in a group $G = \{1, g, g^2, g^3, \ldots, g^{q-1}\}$



Browser Alice

$a \leftarrow \{1,\ldots,q\}$

$PMS = B^a$

$A = g^a \in G$

$B = g^b \in G$

Server Bob

$b \leftarrow \{1,\ldots,q\}$

$PMS = A^b$

$$PreMasterSecret = g^{ab} = (g^b)^a = B^a = (g^a)^b = A^b$$

Used to establish a shared secret

Group G is publicly known

Tommy

Why this paper?

What you liked most?

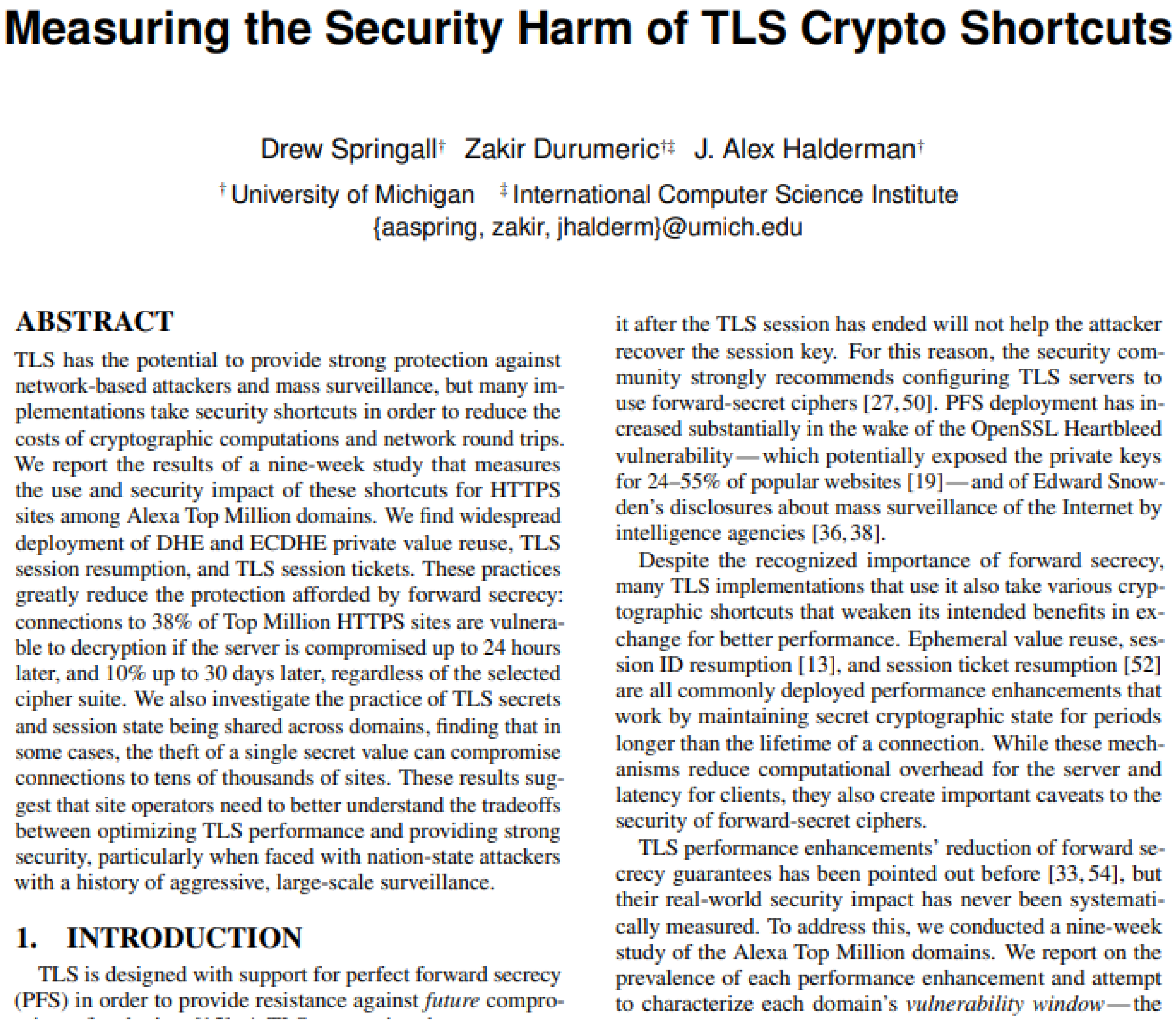
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



Measuring the Security Harm of TLS Crypto Shortcuts

Drew Springall[†] Zakir Durumeric[†‡] J. Alex Halderman[†]

[†] University of Michigan [‡] International Computer Science Institute
{aaspring, zakir, jhalderm}@umich.edu

ABSTRACT

TLS has the potential to provide strong protection against network-based attackers and mass surveillance, but many implementations take security shortcuts in order to reduce the costs of cryptographic computations and network round trips. We report the results of a nine-week study that measures the use and security impact of these shortcuts for HTTPS sites among Alexa Top Million domains. We find widespread deployment of DHE and ECDHE private value reuse, TLS session resumption, and TLS session tickets. These practices greatly reduce the protection afforded by forward secrecy: connections to 38% of Top Million HTTPS sites are vulnerable to decryption if the server is compromised up to 24 hours later, and 10% up to 30 days later, regardless of the selected cipher suite. We also investigate the practice of TLS secrets and session state being shared across domains, finding that in some cases, the theft of a single secret value can compromise connections to tens of thousands of sites. These results suggest that site operators need to better understand the tradeoffs between optimizing TLS performance and providing strong security, particularly when faced with nation-state attackers with a history of aggressive, large-scale surveillance.

1. INTRODUCTION

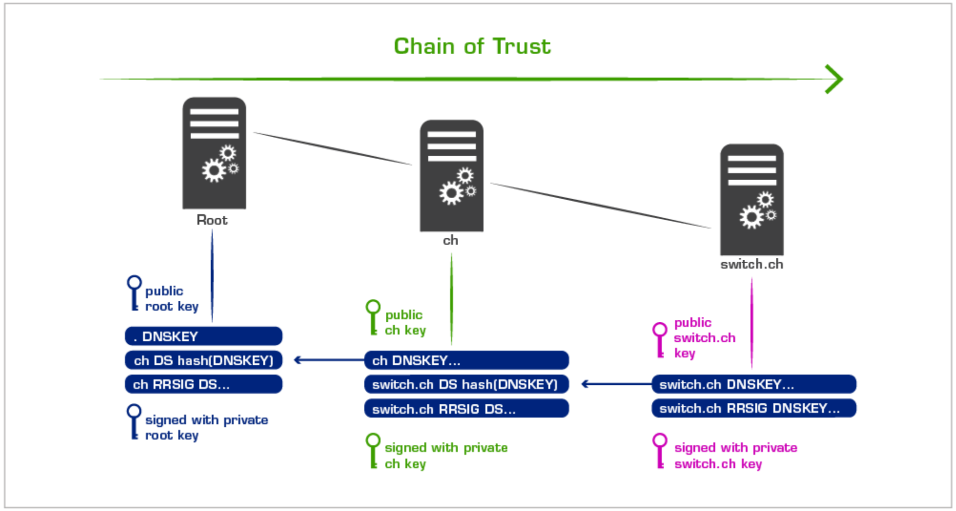
TLS is designed with support for perfect forward secrecy (PFS) in order to provide resistance against *future* compro-

it after the TLS session has ended will not help the attacker recover the session key. For this reason, the security community strongly recommends configuring TLS servers to use forward-secret ciphers [27, 50]. PFS deployment has increased substantially in the wake of the OpenSSL Heartbleed vulnerability — which potentially exposed the private keys for 24–55% of popular websites [19] — and of Edward Snowden's disclosures about mass surveillance of the Internet by intelligence agencies [36, 38].

Despite the recognized importance of forward secrecy, many TLS implementations that use it also take various cryptographic shortcuts that weaken its intended benefits in exchange for better performance. Ephemeral value reuse, session ID resumption [13], and session ticket resumption [52] are all commonly deployed performance enhancements that work by maintaining secret cryptographic state for periods longer than the lifetime of a connection. While these mechanisms reduce computational overhead for the server and latency for clients, they also create important caveats to the security of forward-secret ciphers.

TLS performance enhancements' reduction of forward secrecy guarantees has been pointed out before [33, 54], but their real-world security impact has never been systematically measured. To address this, we conducted a nine-week study of the Alexa Top Million domains. We report on the prevalence of each performance enhancement and attempt to characterize each domain's *vulnerability window* — the

DNSSEC vs DNS over HTTPS



Chain of Trust

Root

ch

switch.ch

public root key

. DNSKEY

ch DS hash(DNSKEY)

ch RRSIG DS...

signed with private root key

public ch key

ch DNSKEY...

switch.ch DS hash(DNSKEY)

switch.ch RRSIG DS...

signed with private ch key

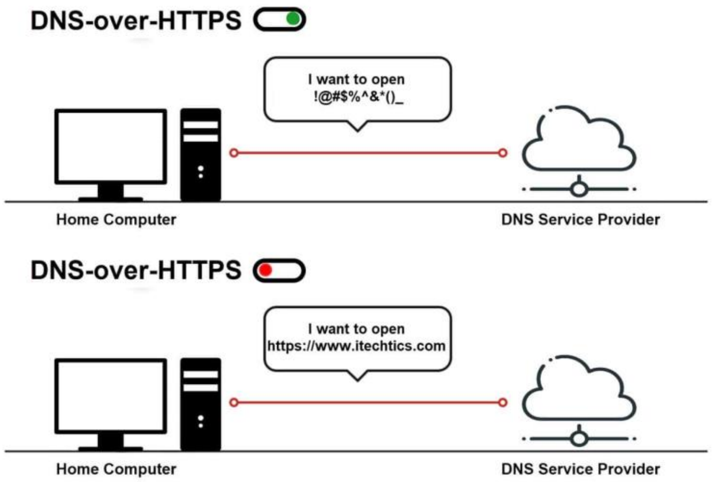
public switch.ch key

switch.ch DNSKEY...

switch.ch RRSIG DNSKEY...

signed with private switch.ch key

*Focus on **integrity** (via trust and authentication)*



DNS-over-HTTPS

I want to open !@#$%^&*()_

Home Computer

DNS Service Provider

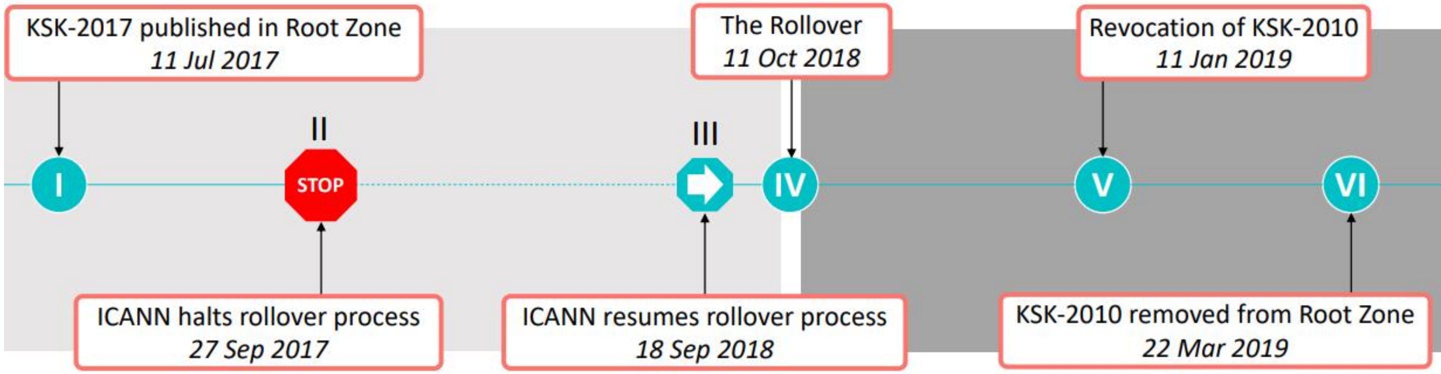
DNS-over-HTTPS

I want to open https://www.itechtics.com

Home Computer

DNS Service Provider

*Focus on **confidentiality***



KSK-2017 published in Root Zone
11 Jul 2017

I

II

STOP

ICANN halts rollover process
27 Sep 2017

III

ICANN resumes rollover process
18 Sep 2018

The Rollover
11 Oct 2018

IV

Revocation of KSK-2010
11 Jan 2019

V

VI

KSK-2010 removed from Root Zone
22 Mar 2019

Adam

Why this paper?

What you liked most?

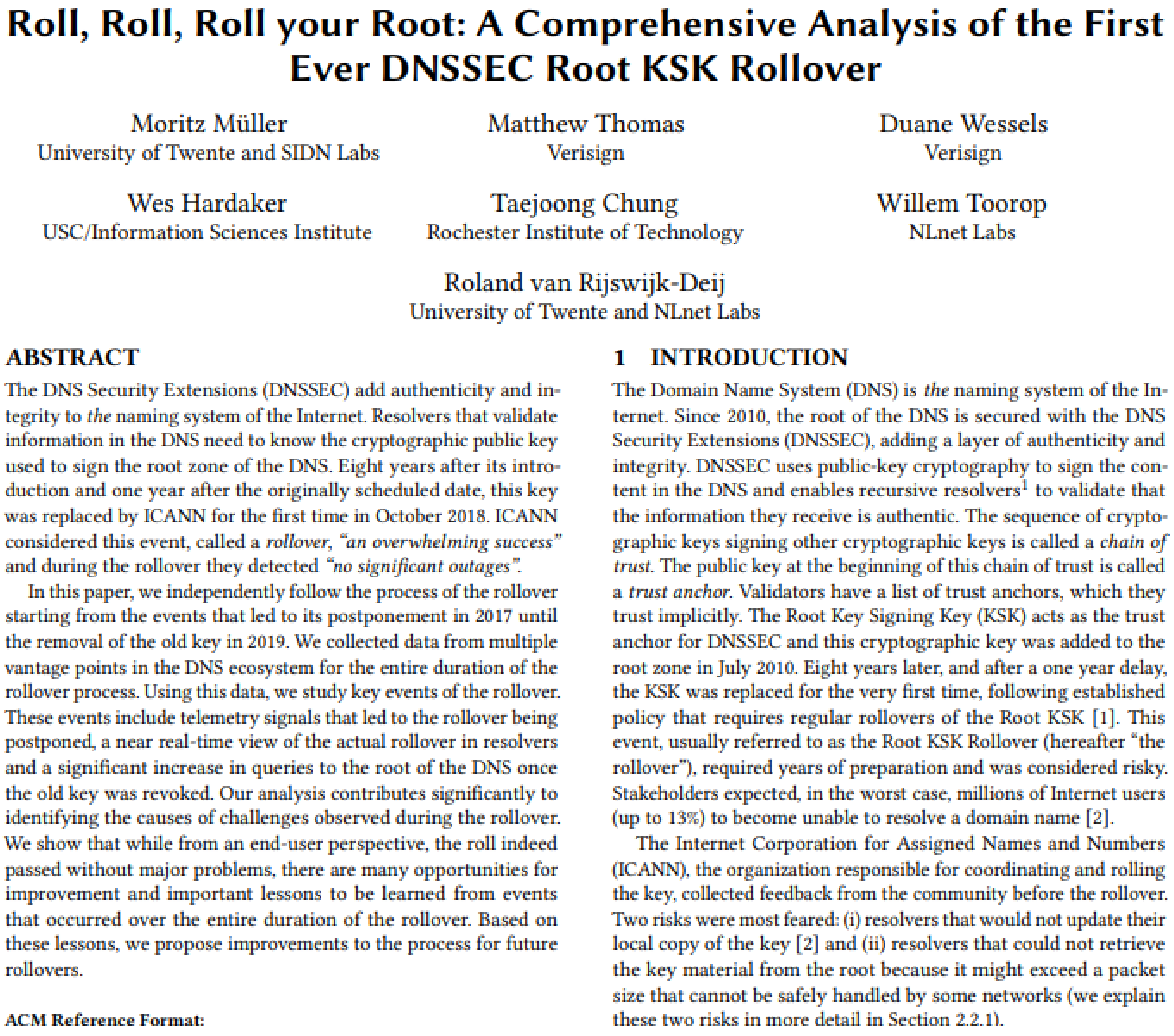
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



Roll, Roll, Roll your Root: A Comprehensive Analysis of the First Ever DNSSEC Root KSK Rollover

Moritz Müller
University of Twente and SIDN Labs

Matthew Thomas
Verisign

Duane Wessels
Verisign

Wes Hardaker
USC/Information Sciences Institute

Taejoong Chung
Rochester Institute of Technology

Willem Toorop
NLnet Labs

Roland van Rijswijk-Deij
University of Twente and NLnet Labs

ABSTRACT

The DNS Security Extensions (DNSSEC) add authenticity and integrity to *the* naming system of the Internet. Resolvers that validate information in the DNS need to know the cryptographic public key used to sign the root zone of the DNS. Eight years after its introduction and one year after the originally scheduled date, this key was replaced by ICANN for the first time in October 2018. ICANN considered this event, called a *rollover*, *"an overwhelming success"* and during the rollover they detected *"no significant outages"*.

In this paper, we independently follow the process of the rollover starting from the events that led to its postponement in 2017 until the removal of the old key in 2019. We collected data from multiple vantage points in the DNS ecosystem for the entire duration of the rollover process. Using this data, we study key events of the rollover. These events include telemetry signals that led to the rollover being postponed, a near real-time view of the actual rollover in resolvers and a significant increase in queries to the root of the DNS once the old key was revoked. Our analysis contributes significantly to identifying the causes of challenges observed during the rollover. We show that while from an end-user perspective, the roll indeed passed without major problems, there are many opportunities for improvement and important lessons to be learned from events that occurred over the entire duration of the rollover. Based on these lessons, we propose improvements to the process for future rollovers.

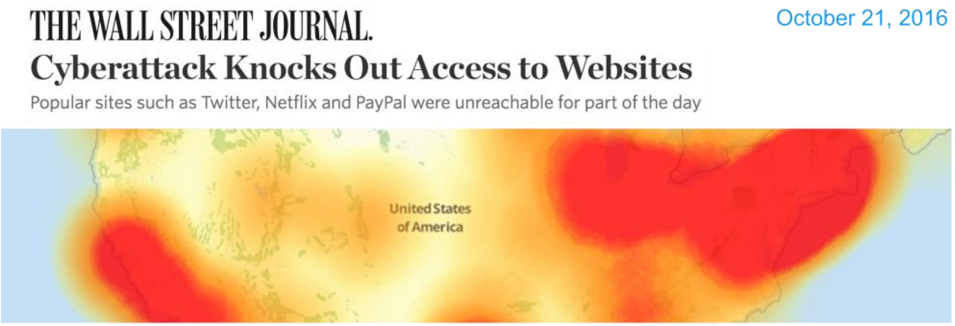
ACM Reference Format:

1 INTRODUCTION

The Domain Name System (DNS) is *the* naming system of the Internet. Since 2010, the root of the DNS is secured with the DNS Security Extensions (DNSSEC), adding a layer of authenticity and integrity. DNSSEC uses public-key cryptography to sign the content in the DNS and enables recursive resolvers[1] to validate that the information they receive is authentic. The sequence of cryptographic keys signing other cryptographic keys is called a *chain of trust*. The public key at the beginning of this chain of trust is called a *trust anchor*. Validators have a list of trust anchors, which they trust implicitly. The Root Key Signing Key (KSK) acts as the trust anchor for DNSSEC and this cryptographic key was added to the root zone in July 2010. Eight years later, and after a one year delay, the KSK was replaced for the very first time, following established policy that requires regular rollovers of the Root KSK [1]. This event, usually referred to as the Root KSK Rollover (hereafter "the rollover"), required years of preparation and was considered risky. Stakeholders expected, in the worst case, millions of Internet users (up to 13%) to become unable to resolve a domain name [2].

The Internet Corporation for Assigned Names and Numbers (ICANN), the organization responsible for coordinating and rolling the key, collected feedback from the community before the rollover. Two risks were most feared: (i) resolvers that would not update their local copy of the key [2] and (ii) resolvers that could not retrieve the key material from the root because it might exceed a packet size that cannot be safely handled by some networks (we explain these two risks in more detail in Section 2.2.1).

DoS attacks



THE WALL STREET JOURNAL.

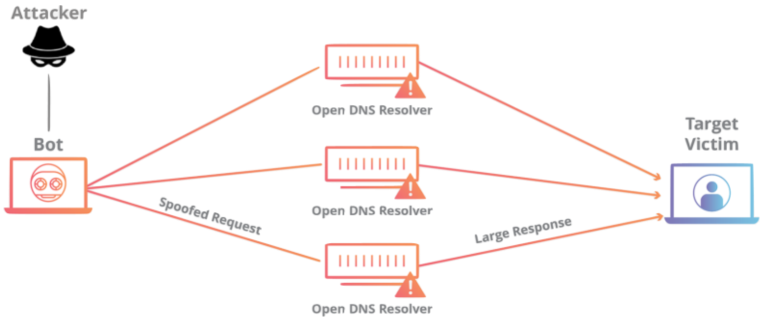
October 21, 2016

Cyberattack Knocks Out Access to Websites

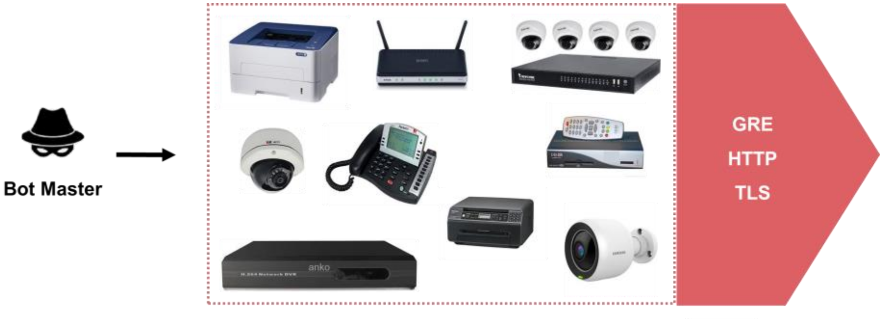
Popular sites such as Twitter, Netflix and PayPal were unreachable for part of the day

twitter amazon web services PayPal NETFLIX SOUNDCLOUD

Spotify GitHub reddit New York Times



Amplification



200K IoT devices

Control lots of machines

Joakim

Why this paper?

What you liked most?

Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?

Booting the Booters: Evaluating the Effects of Police Interventions in the Market for Denial-of-Service Attacks

Ben Collier
Ben.Collier@cl.cam.ac.uk
Department of Computer Science & Technology,
University of Cambridge, Cambridge, CB3 0FD, UK

Daniel R. Thomas
Daniel.Thomas@cl.cam.ac.uk
Computer & Information Sciences,
University of Strathclyde, Glasgow, G1 1XH, UK

Richard Clayton
Richard.Clayton@cl.cam.ac.uk
Department of Computer Science & Technology,
University of Cambridge, Cambridge, CB3 0FD, UK

Alice Hutchings
Alice.Hutchings@cl.cam.ac.uk
Department of Computer Science & Technology,
University of Cambridge, Cambridge, CB3 0FD, UK

ABSTRACT

Illegal *booter services* offer denial of service (DoS) attacks for a fee of a few tens of dollars a month. Internationally, police have implemented a range of different types of intervention aimed at those using and offering booter services, including arrests and website takedown. In order to measure the impact of these interventions we look at the usage reports that booters themselves provide and at measurements of reflected UDP DoS attacks, leveraging a five year measurement dataset that has been statistically demonstrated to have very high coverage. We analysed time series data (using a negative binomial regression model) to show that several interventions have had a statistically significant impact on the number of attacks. We show that, while there is no consistent effect of highly-publicised court cases, takedowns of individual booters precede significant, but short-lived, reductions in recorded attack numbers. However, more wide-ranging disruptions have much longer effects. The closure of HackForums' booter market reduced attacks for 13 weeks globally (and for longer in particular countries) and the FBI's coordinated operation in December 2018, which involved both takedowns and arrests, reduced attacks by a third for at least 10 weeks and resulted in lasting change to the structure of the booter market.

CCS CONCEPTS

KEYWORDS

denial of service attacks; DDoS; UDP-reflection; booter; stresser; cybercrime; police interventions

ACM Reference Format:
Ben Collier, Daniel R. Thomas, Richard Clayton, and Alice Hutchings. 2019. Booting the Booters: Evaluating the Effects of Police Interventions in the Market for Denial-of-Service Attacks. In *Internet Measurement Conference (IMC '19), October 21–23, 2019, Amsterdam, Netherlands.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3355369.3355592

1 INTRODUCTION

'Booter', or 'stresser', services provide Denial of Service (DoS) attacks as-a-service. DoS attacks generate large amounts of traffic which overwhelm end-users or web services, taking them offline or making legitimate access impossible [26]. Booter operators advertise customer-facing websites, where individuals can set up accounts and order attacks [54], with payments accepted using digital services such as PayPal or through transfers of cryptocurrency [22, 28]. A range of different packages and membership options are available, with $10 to $20 being typical for a month's worth of DoS attacks of sufficient size to disrupt an end-user connection or

Jakob

Why this paper?

What you liked most?

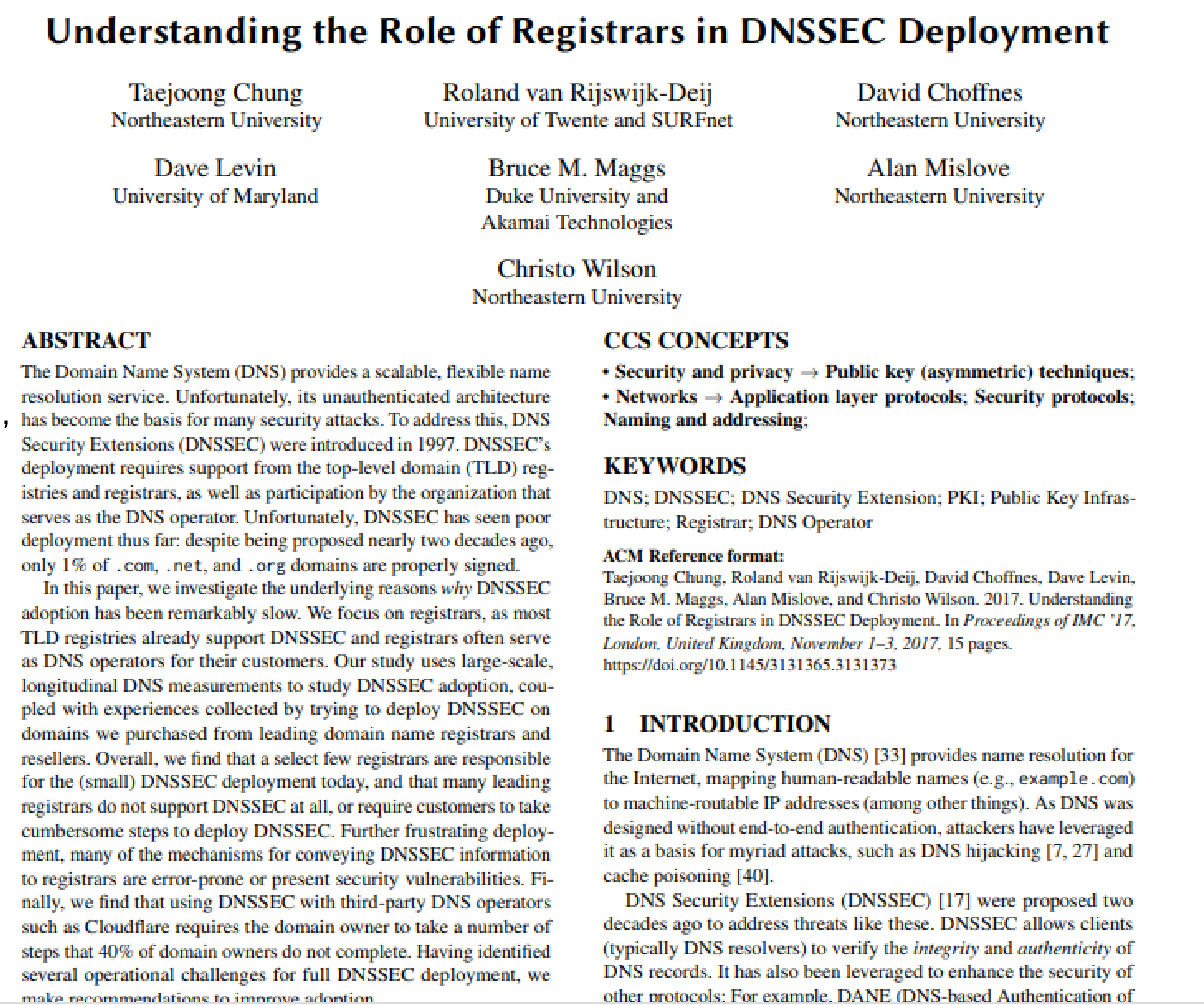
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



Understanding the Role of Registrars in DNSSEC Deployment

Taejoong Chung
Northeastern University

Roland van Rijswijk-Deij
University of Twente and SURFnet

David Choffnes
Northeastern University

Dave Levin
University of Maryland

Bruce M. Maggs
Duke University and
Akamai Technologies

Alan Mislove
Northeastern University

Christo Wilson
Northeastern University

ABSTRACT

The Domain Name System (DNS) provides a scalable, flexible name resolution service. Unfortunately, its unauthenticated architecture has become the basis for many security attacks. To address this, DNS Security Extensions (DNSSEC) were introduced in 1997. DNSSEC's deployment requires support from the top-level domain (TLD) registries and registrars, as well as participation by the organization that serves as the DNS operator. Unfortunately, DNSSEC has seen poor deployment thus far: despite being proposed nearly two decades ago, only 1% of .com, .net, and .org domains are properly signed.

In this paper, we investigate the underlying reasons *why* DNSSEC adoption has been remarkably slow. We focus on registrars, as most TLD registries already support DNSSEC and registrars often serve as DNS operators for their customers. Our study uses large-scale, longitudinal DNS measurements to study DNSSEC adoption, coupled with experiences collected by trying to deploy DNSSEC on domains we purchased from leading domain name registrars and resellers. Overall, we find that a select few registrars are responsible for the (small) DNSSEC deployment today, and that many leading registrars do not support DNSSEC at all, or require customers to take cumbersome steps to deploy DNSSEC. Further frustrating deployment, many of the mechanisms for conveying DNSSEC information to registrars are error-prone or present security vulnerabilities. Finally, we find that using DNSSEC with third-party DNS operators such as Cloudflare requires the domain owner to take a number of steps that 40% of domain owners do not complete. Having identified several operational challenges for full DNSSEC deployment, we make recommendations to improve adoption.

CCS CONCEPTS

• Security and privacy → Public key (asymmetric) techniques; • Networks → Application layer protocols; Security protocols; Naming and addressing;

KEYWORDS

DNS; DNSSEC; DNS Security Extension; PKI; Public Key Infrastructure; Registrar; DNS Operator

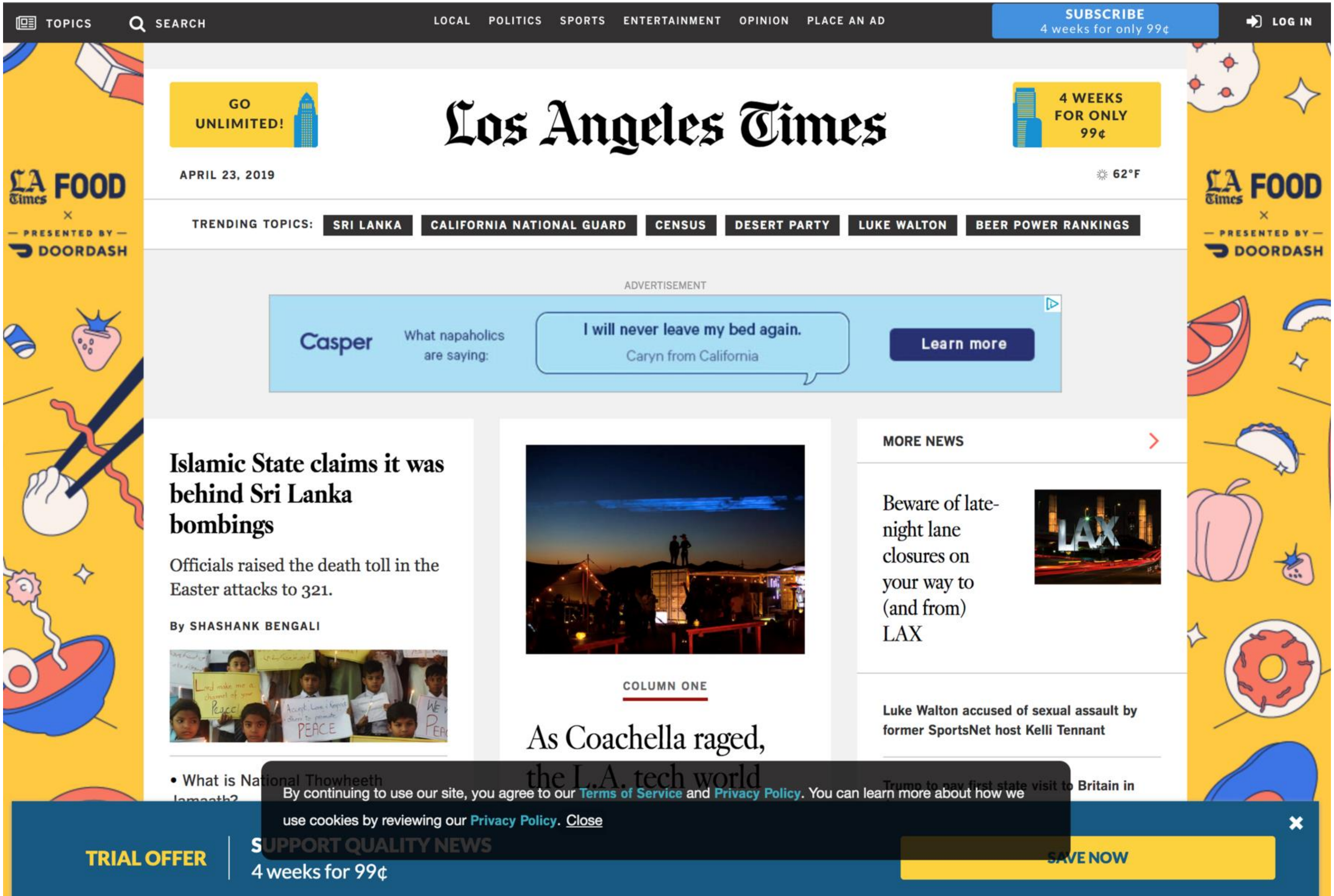
ACM Reference format:
Taejoong Chung, Roland van Rijswijk-Deij, David Choffnes, Dave Levin, Bruce M. Maggs, Alan Mislove, and Christo Wilson. 2017. Understanding the Role of Registrars in DNSSEC Deployment. In *Proceedings of IMC '17, London, United Kingdom, November 1–3, 2017,* 15 pages.
https://doi.org/10.1145/3131365.3131373

1 INTRODUCTION

The Domain Name System (DNS) [33] provides name resolution for the Internet, mapping human-readable names (e.g., example.com) to machine-routable IP addresses (among other things). As DNS was designed without end-to-end authentication, attackers have leveraged it as a basis for myriad attacks, such as DNS hijacking [7, 27] and cache poisoning [40].

DNS Security Extensions (DNSSEC) [17] were proposed two decades ago to address threats like these. DNSSEC allows clients (typically DNS resolvers) to verify the *integrity* and *authenticity* of DNS records. It has also been leveraged to enhance the security of other protocols: For example, DANE (DNS-based Authentication of

Modern Website



TOPICS SEARCH

LOCAL POLITICS SPORTS ENTERTAINMENT OPINION PLACE AN AD

SUBSCRIBE
4 weeks for only 99¢

LOG IN

GO UNLIMITED!

Los Angeles Times

4 WEEKS FOR ONLY 99¢

APRIL 23, 2019

62°F

TRENDING TOPICS: SRI LANKA | CALIFORNIA NATIONAL GUARD | CENSUS | DESERT PARTY | LUKE WALTON | BEER POWER RANKINGS

LA Times FOOD
×
PRESENTED BY
DOORDASH

Islamic State claims it was behind Sri Lanka bombings

Officials raised the death toll in the Easter attacks to 321.

By SHASHANK BENGALI

- What is National Thowheeth Jamaath?

COLUMN ONE

As Coachella raged, the L.A. tech world

MORE NEWS

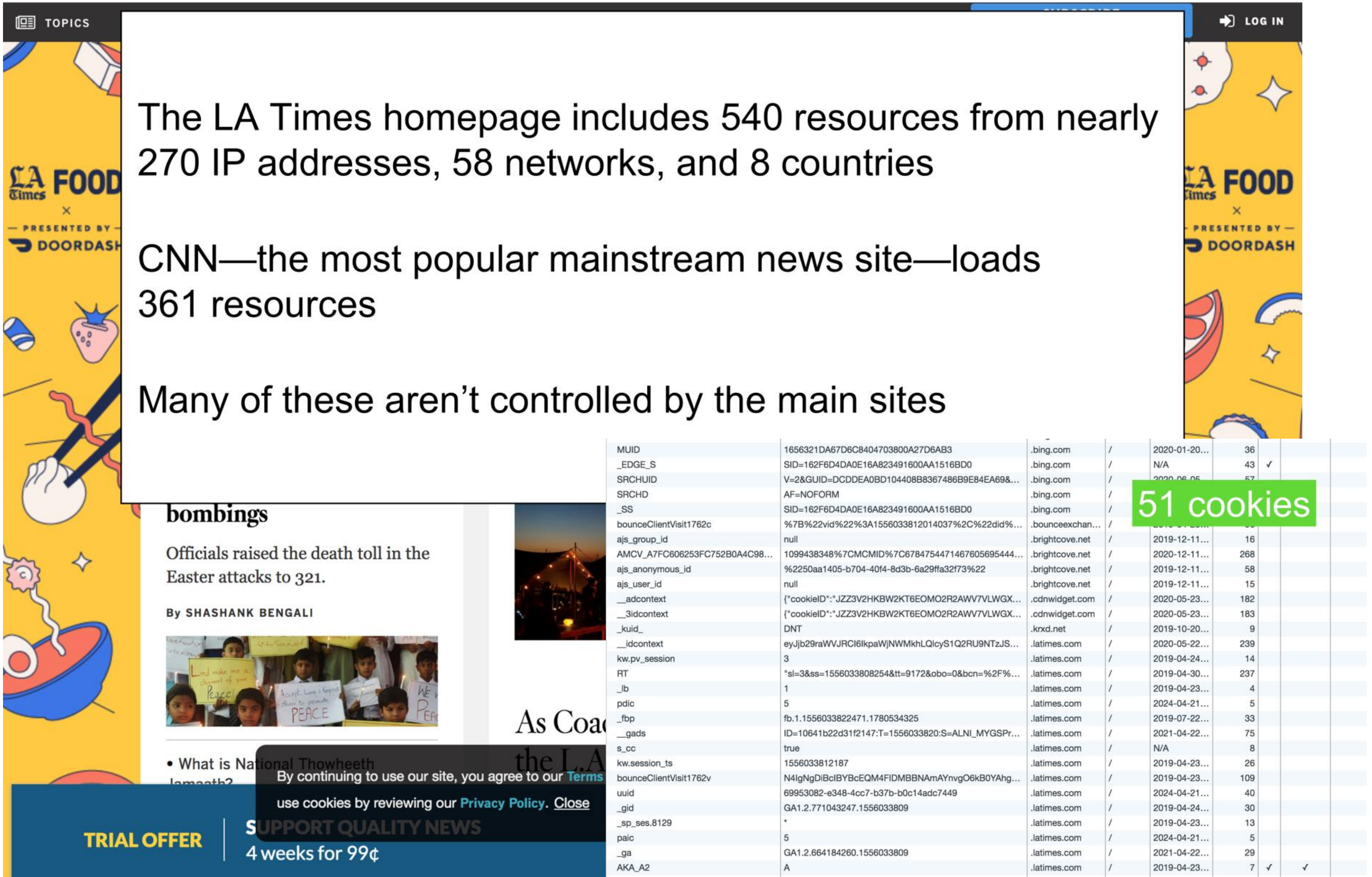
Beware of late-night lane closures on your way to (and from) LAX

Luke Walton accused of sexual assault by former SportsNet host Kelli Tennant

Trump to pay first state visit to Britain in

By continuing to use our site, you agree to our Terms of Service and Privacy Policy. You can learn more about how we use cookies by reviewing our Privacy Policy. Close

Modern Website



The LA Times homepage includes 540 resources from nearly 270 IP addresses, 58 networks, and 8 countries

CNN—the most popular mainstream news site—loads 361 resources

Many of these aren't controlled by the main sites

bombings

Officials raised the death toll in the Easter attacks to 321.

By SHASHANK BENGALI

• What is National Thowheeth Jamaath?

As Coa

the L.A

By continuing to use our site, you agree to our Terms

use cookies by reviewing our Privacy Policy. Close

TRIAL OFFER

SUPPORT QUALITY NEWS

4 weeks for 99¢

51 cookies

MUID	1656321DA67D6C8404703800A27D6AB3	.bing.com	/	2020-01-20...	36		
_EDGE_S	SID=162F6D4DA0E16A823491600AA1516BD0	.bing.com	/	N/A	43	✓	
SRCHUID	V=2&GUID=DCDDEA0BD104408B8367486B9E84EA69&...	.bing.com	/	2020-06-05	57		
SRCHD	AF=NOFORM	.bing.com	/	[illegible]	[illegible]		
_SS	SID=162F6D4DA0E16A823491600AA1516BD0	.bing.com	/	[illegible]	[illegible]		
bounceClientVisit1762c	%7B%22vid%22%3A1556033812014037%2C%22did%...	.bounceexchan...	/	[illegible]	[illegible]		
ajs_group_id	null	.brightcove.net	/	2019-12-11...	16		
AMCV_A7FC606253FC752B0A4C98...	1099438348%7CMCMID%7C6784754471467605695444...	.brightcove.net	/	2020-12-11...	268		
ajs_anonymous_id	%2250aa1405-b704-40f4-8d3b-6a29ffa32f73%22	.brightcove.net	/	2019-12-11...	58		
ajs_user_id	null	.brightcove.net	/	2019-12-11...	15		
__adcontext	{"cookieID":"JZZ3V2HKBW2KT6EOMO2R2AWV7VLWGX...	.cdnwidget.com	/	2020-05-23...	182		
__3idcontext	{"cookieID":"JZZ3V2HKBW2KT6EOMO2R2AWV7VLWGX...	.cdnwidget.com	/	2020-05-23...	183		
kuid	DNT	.krxd.net	/	2019-10-20...	9		
__idcontext	eyJjb29raWVJRCI6IkpaWjNWMkhLQlcyS1Q2RU9NTzJS...	.latimes.com	/	2020-05-22...	239		
kw.pv_session	3	.latimes.com	/	2019-04-24...	14		
RT	"sl=3&ss=1556033808254&tt=9172&obo=0&bcn=%2F%...	.latimes.com	/	2019-04-30...	237		
_lb	1	.latimes.com	/	2019-04-23...	4		
pdic	5	.latimes.com	/	2024-04-21...	5		
_fbp	fb.1.1556033822471.1780534325	.latimes.com	/	2019-07-22...	33		
__gads	ID=10641b22d31f2147:T=1556033820:S=ALNI_MYGSPr...	.latimes.com	/	2021-04-22...	75		
s_cc	true	.latimes.com	/	N/A	8		
kw.session_ts	1556033812187	.latimes.com	/	2019-04-23...	26		
bounceClientVisit1762v	N4IgNgDiBcIBYBcEQM4FIDMBBNAmAYnvgO6kB0YAhg...	.latimes.com	/	2019-04-23...	109		
uuid	69953082-e348-4cc7-b37b-b0c14adc7449	.latimes.com	/	2024-04-21...	40		
_gid	GA1.2.771043247.1556033809	.latimes.com	/	2019-04-24...	30		
_sp_ses.8129	*	.latimes.com	/	2019-04-23...	13		
paic	5	.latimes.com	/	2024-04-21...	5		
_ga	GA1.2.664184260.1556033809	.latimes.com	/	2021-04-22...	29		
AKA_A2	A	.latimes.com	/	2019-04-23...	7	✓	✓

Oceane

Why this paper?

What you liked most?

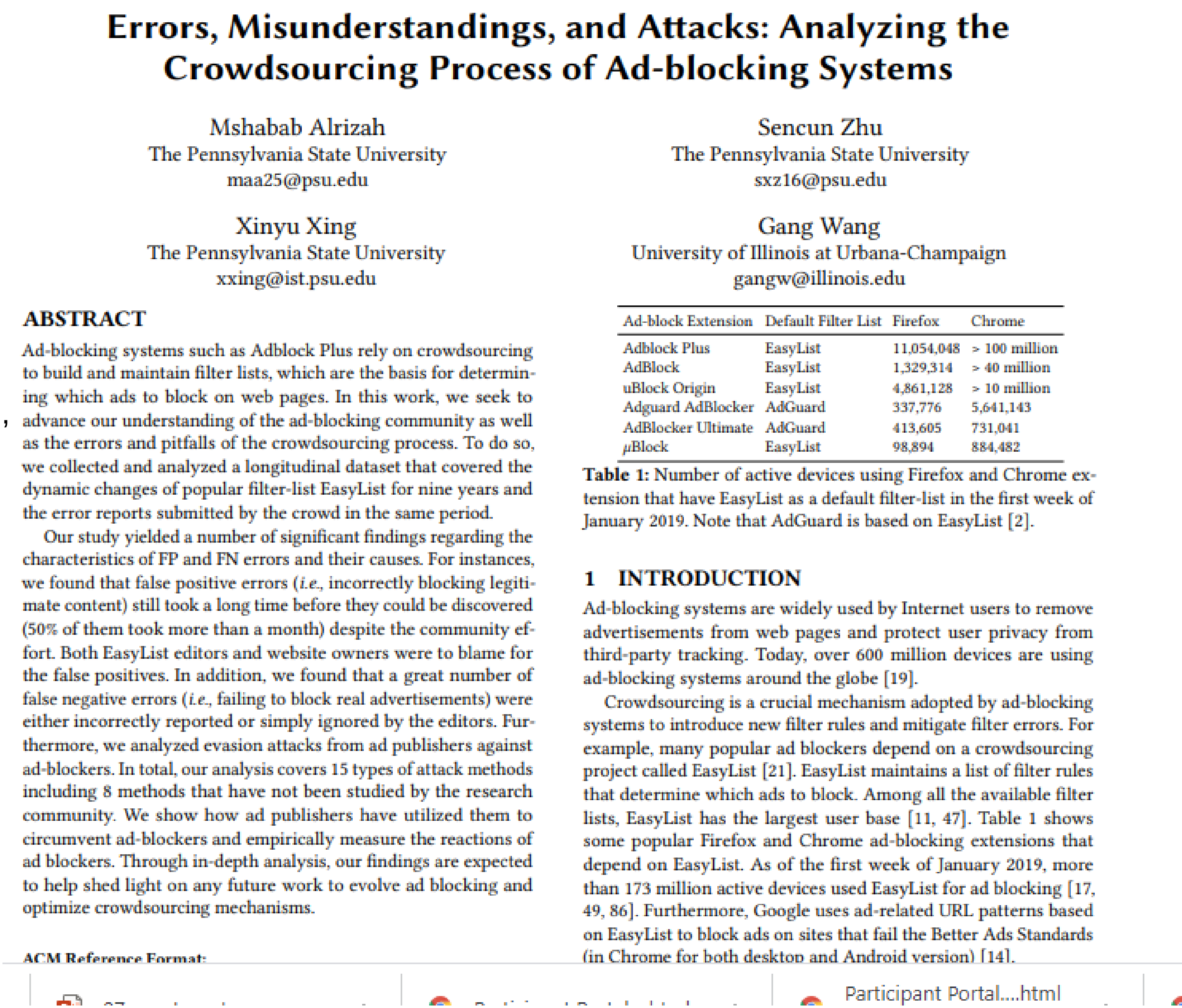
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



Errors, Misunderstandings, and Attacks: Analyzing the Crowdsourcing Process of Ad-blocking Systems

Mshabab Alrizah
The Pennsylvania State University
maa25@psu.edu

Sencun Zhu
The Pennsylvania State University
sxz16@psu.edu

Xinyu Xing
The Pennsylvania State University
xxing@ist.psu.edu

Gang Wang
University of Illinois at Urbana-Champaign
gangw@illinois.edu

ABSTRACT

Ad-blocking systems such as Adblock Plus rely on crowdsourcing to build and maintain filter lists, which are the basis for determining which ads to block on web pages. In this work, we seek to advance our understanding of the ad-blocking community as well as the errors and pitfalls of the crowdsourcing process. To do so, we collected and analyzed a longitudinal dataset that covered the dynamic changes of popular filter-list EasyList for nine years and the error reports submitted by the crowd in the same period.

Our study yielded a number of significant findings regarding the characteristics of FP and FN errors and their causes. For instances, we found that false positive errors (*i.e.*, incorrectly blocking legitimate content) still took a long time before they could be discovered (50% of them took more than a month) despite the community effort. Both EasyList editors and website owners were to blame for the false positives. In addition, we found that a great number of false negative errors (*i.e.*, failing to block real advertisements) were either incorrectly reported or simply ignored by the editors. Furthermore, we analyzed evasion attacks from ad publishers against ad-blockers. In total, our analysis covers 15 types of attack methods including 8 methods that have not been studied by the research community. We show how ad publishers have utilized them to circumvent ad-blockers and empirically measure the reactions of ad blockers. Through in-depth analysis, our findings are expected to help shed light on any future work to evolve ad blocking and optimize crowdsourcing mechanisms.

ACM Reference Format:

Ad-block Extension	Default Filter List	Firefox	Chrome
Adblock Plus	EasyList	11,054,048	> 100 million
AdBlock	EasyList	1,329,314	> 40 million
uBlock Origin	EasyList	4,861,128	> 10 million
Adguard AdBlocker	AdGuard	337,776	5,641,143
AdBlocker Ultimate	AdGuard	413,605	731,041
*μ*Block	EasyList	98,894	884,482

Table 1: Number of active devices using Firefox and Chrome extension that have EasyList as a default filter-list in the first week of January 2019. Note that AdGuard is based on EasyList [2].

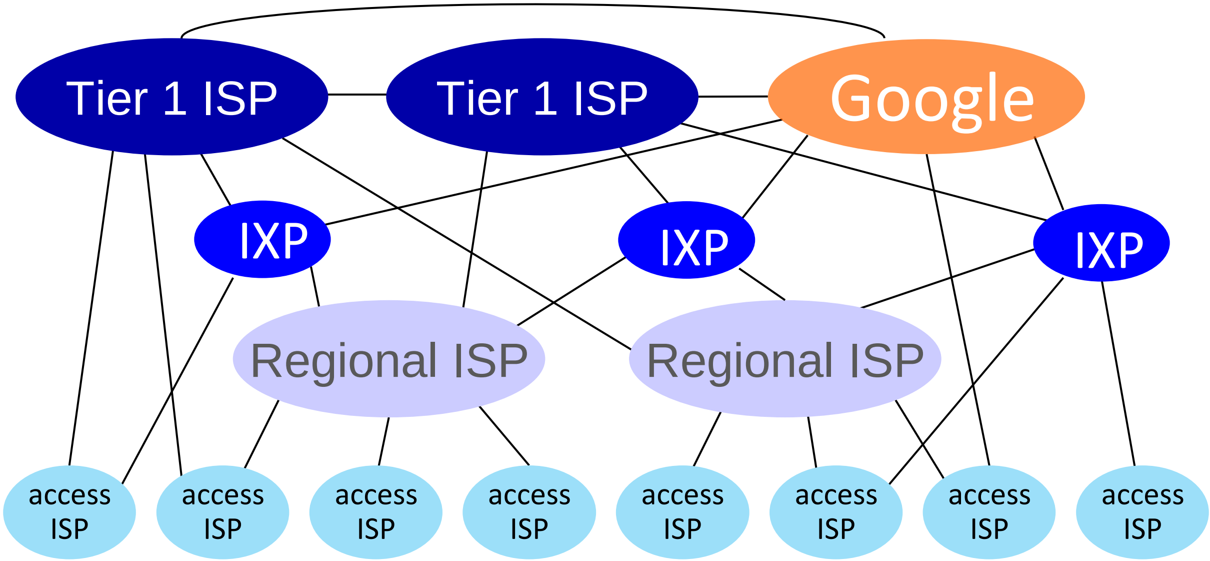
1 INTRODUCTION

Ad-blocking systems are widely used by Internet users to remove advertisements from web pages and protect user privacy from third-party tracking. Today, over 600 million devices are using ad-blocking systems around the globe [19].

Crowdsourcing is a crucial mechanism adopted by ad-blocking systems to introduce new filter rules and mitigate filter errors. For example, many popular ad blockers depend on a crowdsourcing project called EasyList [21]. EasyList maintains a list of filter rules that determine which ads to block. Among all the available filter lists, EasyList has the largest user base [11, 47]. Table 1 shows some popular Firefox and Chrome ad-blocking extensions that depend on EasyList. As of the first week of January 2019, more than 173 million active devices used EasyList for ad blocking [17, 49, 86]. Furthermore, Google uses ad-related URL patterns based on EasyList to block ads on sites that fail the Better Ads Standards (in Chrome for both desktop and Android version) [14].

Slide from intro in in TDDE35/TDTS11/TDTS04/TDTS06 …

Internet structure: a “network of networks”



At “center”: small # of well-connected large networks

- “tier-1” commercial ISPs (e.g., Level 3, Sprint, AT&T, NTT), national & international coverage
- content provider networks (e.g., Google, Facebook): private network that connects its data centers to Internet, often bypassing tier-1, regional ISPs

David J

Why this paper?

What you liked most?

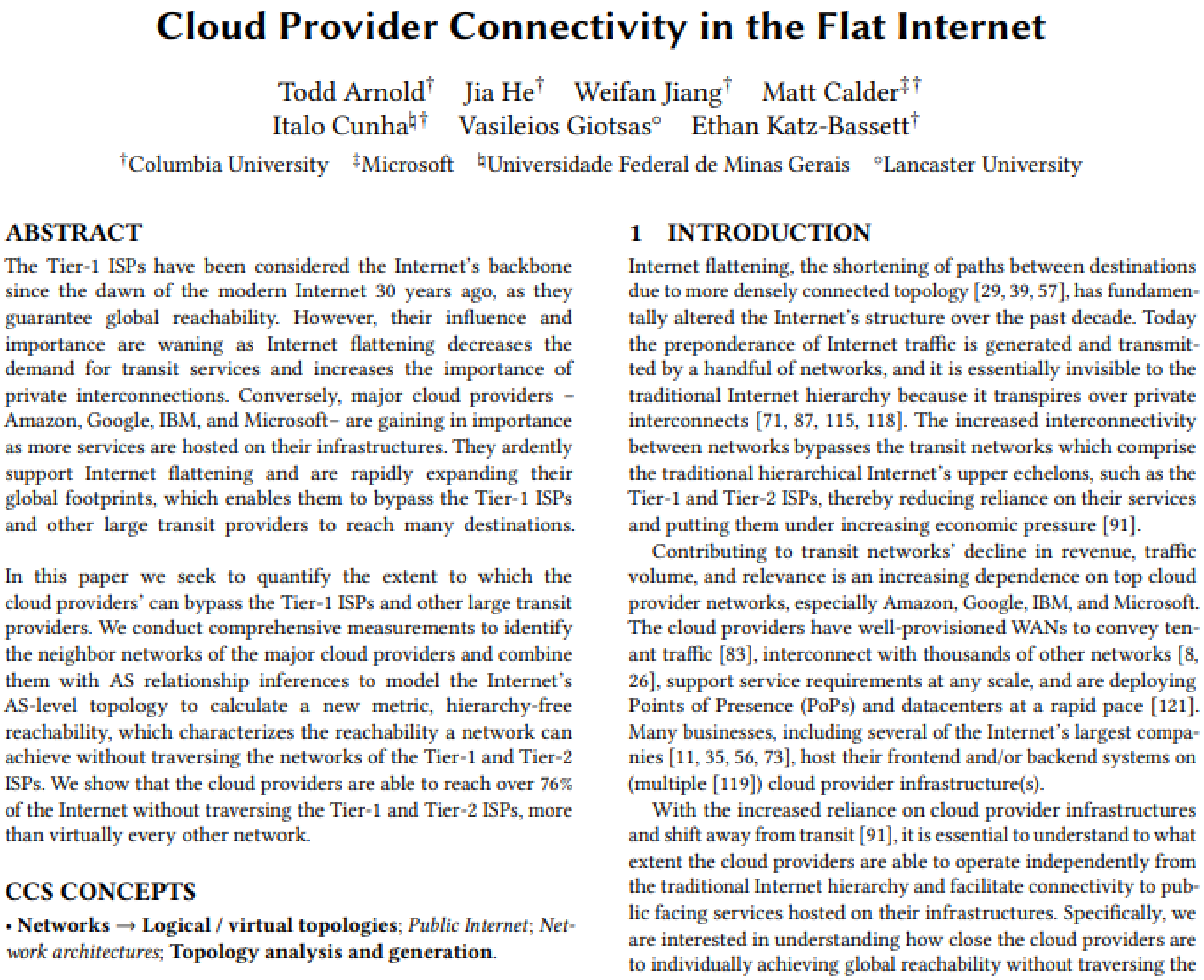
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



Cloud Provider Connectivity in the Flat Internet

Todd Arnold[†] Jia He[†] Weifan Jiang[†] Matt Calder[‡†]
Italo Cunha[♮†] Vasileios Giotsas[°] Ethan Katz-Bassett[†]
[†]Columbia University [‡]Microsoft [♮]Universidade Federal de Minas Gerais [°]Lancaster University

ABSTRACT

The Tier-1 ISPs have been considered the Internet's backbone since the dawn of the modern Internet 30 years ago, as they guarantee global reachability. However, their influence and importance are waning as Internet flattening decreases the demand for transit services and increases the importance of private interconnections. Conversely, major cloud providers – Amazon, Google, IBM, and Microsoft– are gaining in importance as more services are hosted on their infrastructures. They ardently support Internet flattening and are rapidly expanding their global footprints, which enables them to bypass the Tier-1 ISPs and other large transit providers to reach many destinations.

In this paper we seek to quantify the extent to which the cloud providers' can bypass the Tier-1 ISPs and other large transit providers. We conduct comprehensive measurements to identify the neighbor networks of the major cloud providers and combine them with AS relationship inferences to model the Internet's AS-level topology to calculate a new metric, hierarchy-free reachability, which characterizes the reachability a network can achieve without traversing the networks of the Tier-1 and Tier-2 ISPs. We show that the cloud providers are able to reach over 76% of the Internet without traversing the Tier-1 and Tier-2 ISPs, more than virtually every other network.

CCS CONCEPTS

• **Networks** → **Logical / virtual topologies**; *Public Internet*; *Network architectures*; **Topology analysis and generation**.

1 INTRODUCTION

Internet flattening, the shortening of paths between destinations due to more densely connected topology [29, 39, 57], has fundamentally altered the Internet's structure over the past decade. Today the preponderance of Internet traffic is generated and transmitted by a handful of networks, and it is essentially invisible to the traditional Internet hierarchy because it transpires over private interconnects [71, 87, 115, 118]. The increased interconnectivity between networks bypasses the transit networks which comprise the traditional hierarchical Internet's upper echelons, such as the Tier-1 and Tier-2 ISPs, thereby reducing reliance on their services and putting them under increasing economic pressure [91].

Contributing to transit networks' decline in revenue, traffic volume, and relevance is an increasing dependence on top cloud provider networks, especially Amazon, Google, IBM, and Microsoft. The cloud providers have well-provisioned WANs to convey tenant traffic [83], interconnect with thousands of other networks [8, 26], support service requirements at any scale, and are deploying Points of Presence (PoPs) and datacenters at a rapid pace [121]. Many businesses, including several of the Internet's largest companies [11, 35, 56, 73], host their frontend and/or backend systems on (multiple [119]) cloud provider infrastructure(s).

With the increased reliance on cloud provider infrastructures and shift away from transit [91], it is essential to understand to what extent the cloud providers are able to operate independently from the traditional Internet hierarchy and facilitate connectivity to public facing services hosted on their infrastructures. Specifically, we are interested in understanding how close the cloud providers are to individually achieving global reachability without traversing the

Elmedin

Why this paper?

What you liked most?

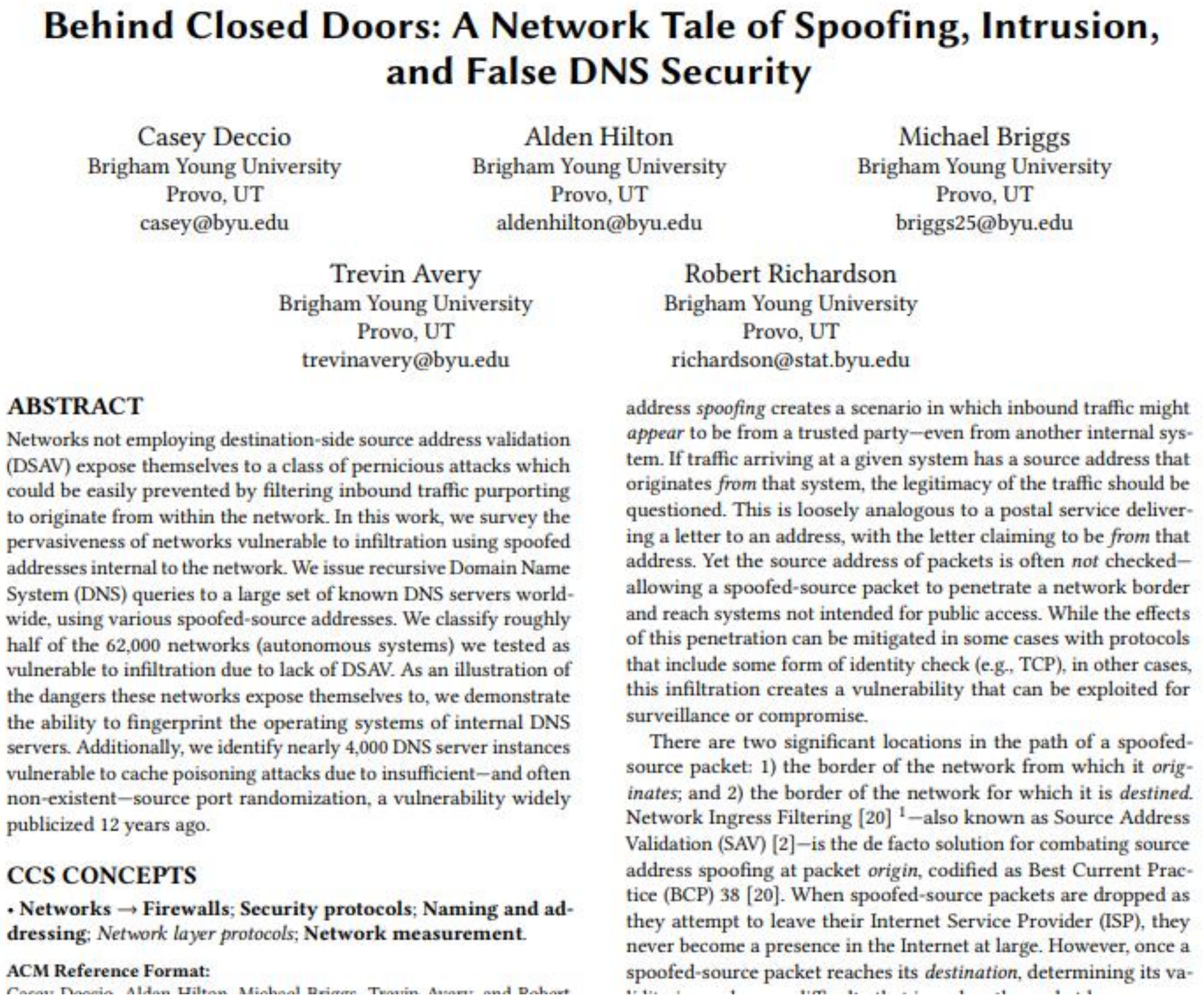
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



Behind Closed Doors: A Network Tale of Spoofing, Intrusion, and False DNS Security

Casey Deccio
Brigham Young University
Provo, UT
casey@byu.edu

Alden Hilton
Brigham Young University
Provo, UT
aldenhilton@byu.edu

Michael Briggs
Brigham Young University
Provo, UT
briggs25@byu.edu

Trevin Avery
Brigham Young University
Provo, UT
trevinavery@byu.edu

Robert Richardson
Brigham Young University
Provo, UT
richardson@stat.byu.edu

ABSTRACT

Networks not employing destination-side source address validation (DSAV) expose themselves to a class of pernicious attacks which could be easily prevented by filtering inbound traffic purporting to originate from within the network. In this work, we survey the pervasiveness of networks vulnerable to infiltration using spoofed addresses internal to the network. We issue recursive Domain Name System (DNS) queries to a large set of known DNS servers worldwide, using various spoofed-source addresses. We classify roughly half of the 62,000 networks (autonomous systems) we tested as vulnerable to infiltration due to lack of DSAV. As an illustration of the dangers these networks expose themselves to, we demonstrate the ability to fingerprint the operating systems of internal DNS servers. Additionally, we identify nearly 4,000 DNS server instances vulnerable to cache poisoning attacks due to insufficient—and often non-existent—source port randomization, a vulnerability widely publicized 12 years ago.

CCS CONCEPTS

• **Networks** → **Firewalls**; **Security protocols**; **Naming and addressing**; *Network layer protocols*; **Network measurement**.

ACM Reference Format:
Casey Deccio, Alden Hilton, Michael Briggs, Trevin Avery, and Robert

address *spoofing* creates a scenario in which inbound traffic might *appear* to be from a trusted party—even from another internal system. If traffic arriving at a given system has a source address that originates *from* that system, the legitimacy of the traffic should be questioned. This is loosely analogous to a postal service delivering a letter to an address, with the letter claiming to be *from* that address. Yet the source address of packets is often *not* checked—allowing a spoofed-source packet to penetrate a network border and reach systems not intended for public access. While the effects of this penetration can be mitigated in some cases with protocols that include some form of identity check (e.g., TCP), in other cases, this infiltration creates a vulnerability that can be exploited for surveillance or compromise.

There are two significant locations in the path of a spoofed-source packet: 1) the border of the network from which it *originates*; and 2) the border of the network for which it is *destined*. Network Ingress Filtering [20] [1]—also known as Source Address Validation (SAV) [2]—is the de facto solution for combating source address spoofing at packet *origin*, codified as Best Current Practice (BCP) 38 [20]. When spoofed-source packets are dropped as they attempt to leave their Internet Service Provider (ISP), they never become a presence in the Internet at large. However, once a spoofed-source packet reaches its *destination*, determining its va-

Rami

Why this paper?

What you liked most?

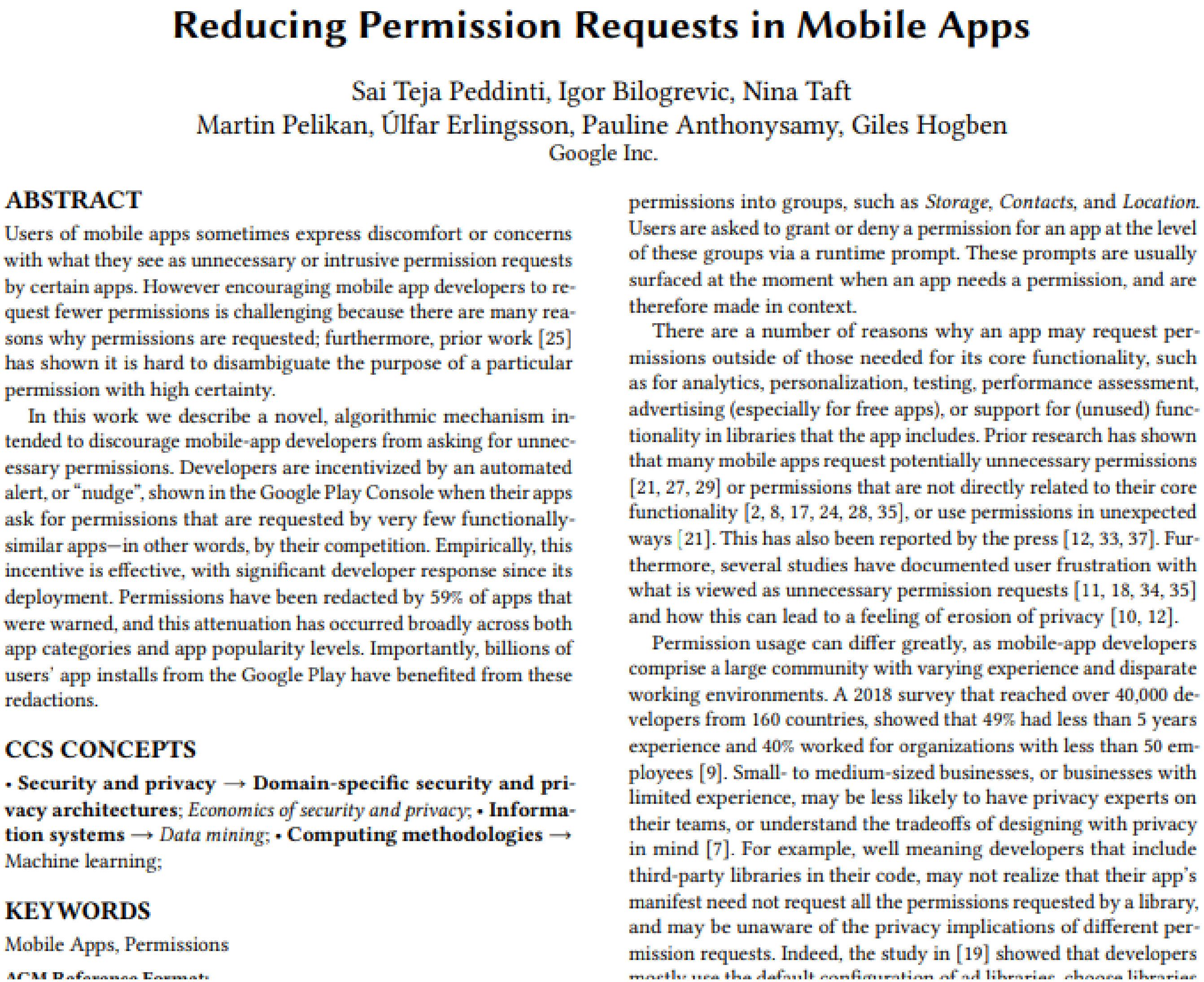
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



Reducing Permission Requests in Mobile Apps

Sai Teja Peddinti, Igor Bilogrevic, Nina Taft
Martin Pelikan, Úlfar Erlingsson, Pauline Anthonysamy, Giles Hogben
Google Inc.

ABSTRACT

Users of mobile apps sometimes express discomfort or concerns with what they see as unnecessary or intrusive permission requests by certain apps. However encouraging mobile app developers to request fewer permissions is challenging because there are many reasons why permissions are requested; furthermore, prior work [25] has shown it is hard to disambiguate the purpose of a particular permission with high certainty.

In this work we describe a novel, algorithmic mechanism intended to discourage mobile-app developers from asking for unnecessary permissions. Developers are incentivized by an automated alert, or "nudge", shown in the Google Play Console when their apps ask for permissions that are requested by very few functionally-similar apps—in other words, by their competition. Empirically, this incentive is effective, with significant developer response since its deployment. Permissions have been redacted by 59% of apps that were warned, and this attenuation has occurred broadly across both app categories and app popularity levels. Importantly, billions of users' app installs from the Google Play have benefited from these redactions.

CCS CONCEPTS

• **Security and privacy** → **Domain-specific security and privacy architectures**; *Economics of security and privacy*; • **Information systems** → *Data mining*; • **Computing methodologies** → Machine learning;

KEYWORDS

Mobile Apps, Permissions

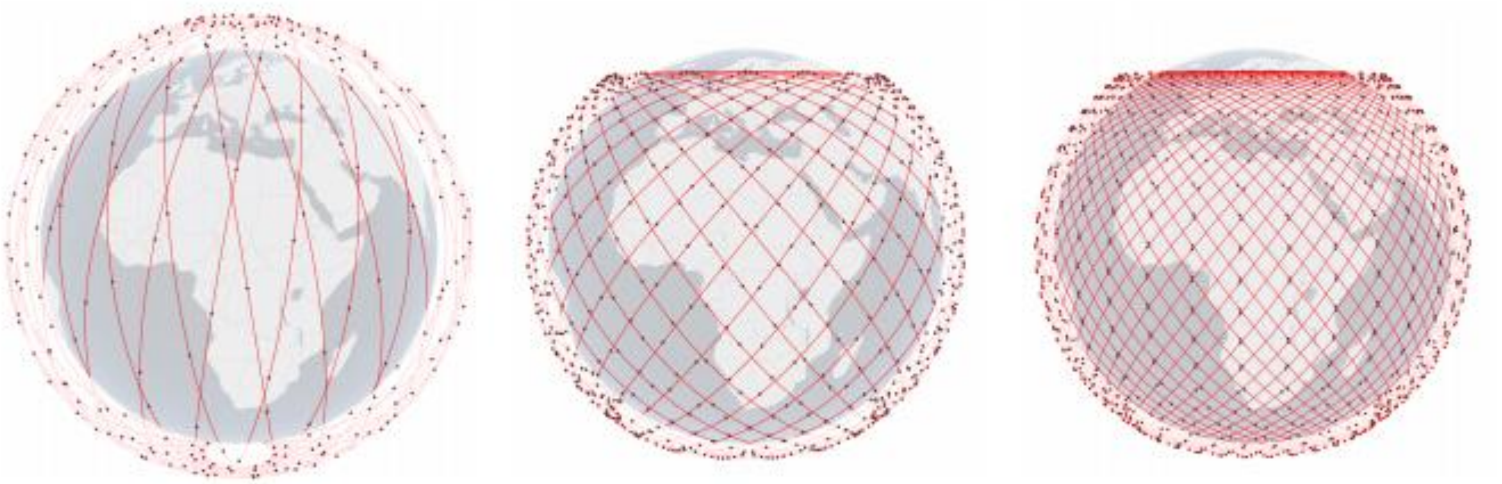
ACM Reference Format:

permissions into groups, such as *Storage*, *Contacts*, and *Location*. Users are asked to grant or deny a permission for an app at the level of these groups via a runtime prompt. These prompts are usually surfaced at the moment when an app needs a permission, and are therefore made in context.

There are a number of reasons why an app may request permissions outside of those needed for its core functionality, such as for analytics, personalization, testing, performance assessment, advertising (especially for free apps), or support for (unused) functionality in libraries that the app includes. Prior research has shown that many mobile apps request potentially unnecessary permissions [21, 27, 29] or permissions that are not directly related to their core functionality [2, 8, 17, 24, 28, 35], or use permissions in unexpected ways [21]. This has also been reported by the press [12, 33, 37]. Furthermore, several studies have documented user frustration with what is viewed as unnecessary permission requests [11, 18, 34, 35] and how this can lead to a feeling of erosion of privacy [10, 12].

Permission usage can differ greatly, as mobile-app developers comprise a large community with varying experience and disparate working environments. A 2018 survey that reached over 40,000 developers from 160 countries, showed that 49% had less than 5 years experience and 40% worked for organizations with less than 50 employees [9]. Small- to medium-sized businesses, or businesses with limited experience, may be less likely to have privacy experts on their teams, or understand the tradeoffs of designing with privacy in mind [7]. For example, well meaning developers that include third-party libraries in their code, may not realize that their app's manifest need not request all the permissions requested by a library, and may be unaware of the privacy implications of different permission requests. Indeed, the study in [19] showed that developers mostly use the default configuration of ad libraries, choose libraries

Competing satellite networks …



(a) Telesat T1

(b) Kuiper K1

(c) Starlink S1

Fig. 11: ***Constellation trajectories.*** *(a) Telesat T1* — *27 × 13, 1,015 km, 98.98° (b) Kuiper K1* — *34 × 34, 630 km, 51.9° (c) Starlink S1* — *72 × 22, 550 km, 53°. Satellites are black dots, while orbits are marked in red.*

Matteus

Why this paper?

What you liked most?

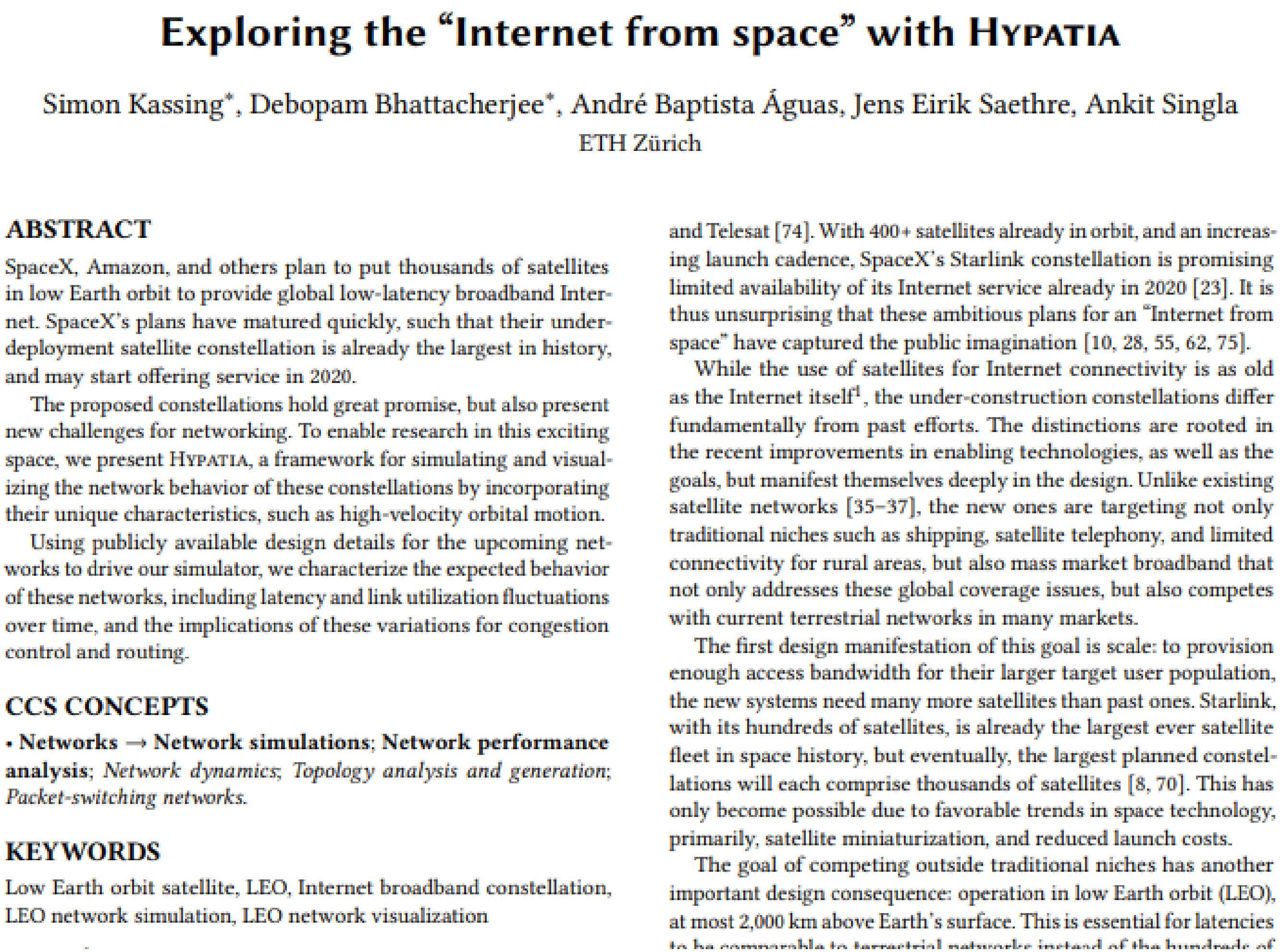
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



Exploring the "Internet from space" with HYPATIA

Simon Kassing*, Debopam Bhattacherjee*, André Baptista Águas, Jens Eirik Saethre, Ankit Singla
ETH Zürich

ABSTRACT

SpaceX, Amazon, and others plan to put thousands of satellites in low Earth orbit to provide global low-latency broadband Internet. SpaceX's plans have matured quickly, such that their under-deployment satellite constellation is already the largest in history, and may start offering service in 2020.

The proposed constellations hold great promise, but also present new challenges for networking. To enable research in this exciting space, we present HYPATIA, a framework for simulating and visualizing the network behavior of these constellations by incorporating their unique characteristics, such as high-velocity orbital motion.

Using publicly available design details for the upcoming networks to drive our simulator, we characterize the expected behavior of these networks, including latency and link utilization fluctuations over time, and the implications of these variations for congestion control and routing.

CCS CONCEPTS

• **Networks** → **Network simulations**; **Network performance analysis**; *Network dynamics*; *Topology analysis and generation*; *Packet-switching networks*.

KEYWORDS

Low Earth orbit satellite, LEO, Internet broadband constellation, LEO network simulation, LEO network visualization

and Telesat [74]. With 400+ satellites already in orbit, and an increasing launch cadence, SpaceX's Starlink constellation is promising limited availability of its Internet service already in 2020 [23]. It is thus unsurprising that these ambitious plans for an "Internet from space" have captured the public imagination [10, 28, 55, 62, 75].

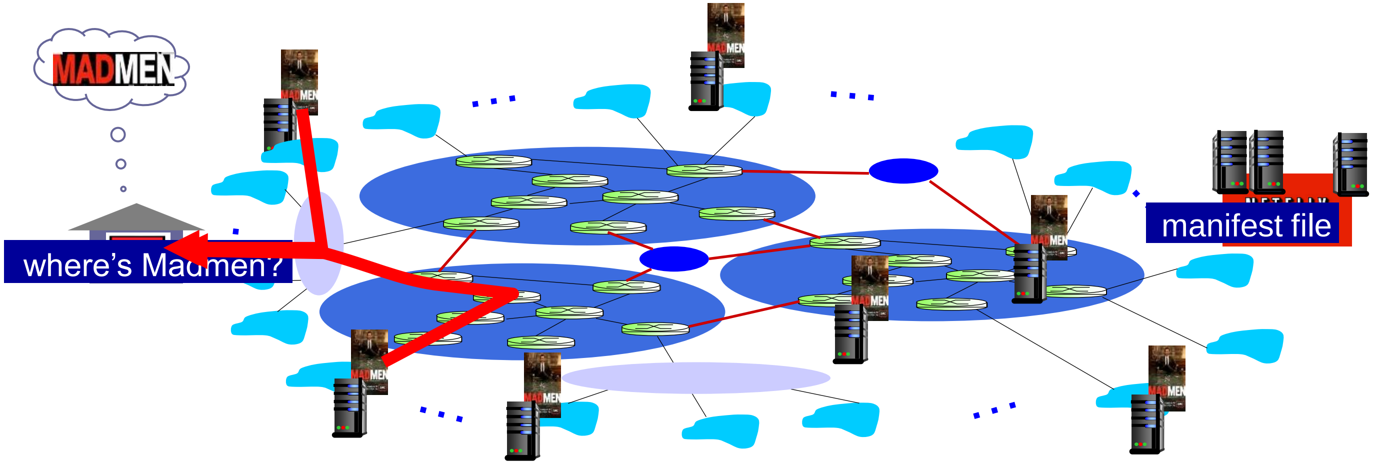
While the use of satellites for Internet connectivity is as old as the Internet itself[1], the under-construction constellations differ fundamentally from past efforts. The distinctions are rooted in the recent improvements in enabling technologies, as well as the goals, but manifest themselves deeply in the design. Unlike existing satellite networks [35–37], the new ones are targeting not only traditional niches such as shipping, satellite telephony, and limited connectivity for rural areas, but also mass market broadband that not only addresses these global coverage issues, but also competes with current terrestrial networks in many markets.

The first design manifestation of this goal is scale: to provision enough access bandwidth for their larger target user population, the new systems need many more satellites than past ones. Starlink, with its hundreds of satellites, is already the largest ever satellite fleet in space history, but eventually, the largest planned constellations will each comprise thousands of satellites [8, 70]. This has only become possible due to favorable trends in space technology, primarily, satellite miniaturization, and reduced launch costs.

The goal of competing outside traditional niches has another important design consequence: operation in low Earth orbit (LEO), at most 2,000 km above Earth's surface. This is essential for latencies to be comparable to terrestrial networks instead of the hundreds of

Content distribution networks (CDNs)

- CDN: stores copies of content (e.g. MADMEN) at CDN nodes
- subscriber requests content, service provider returns manifest
  - *using manifest, client* retrieves content at highest supportable rate
  - may choose different rate or copy if network path congested



Robert

Why this paper?

What you liked most?

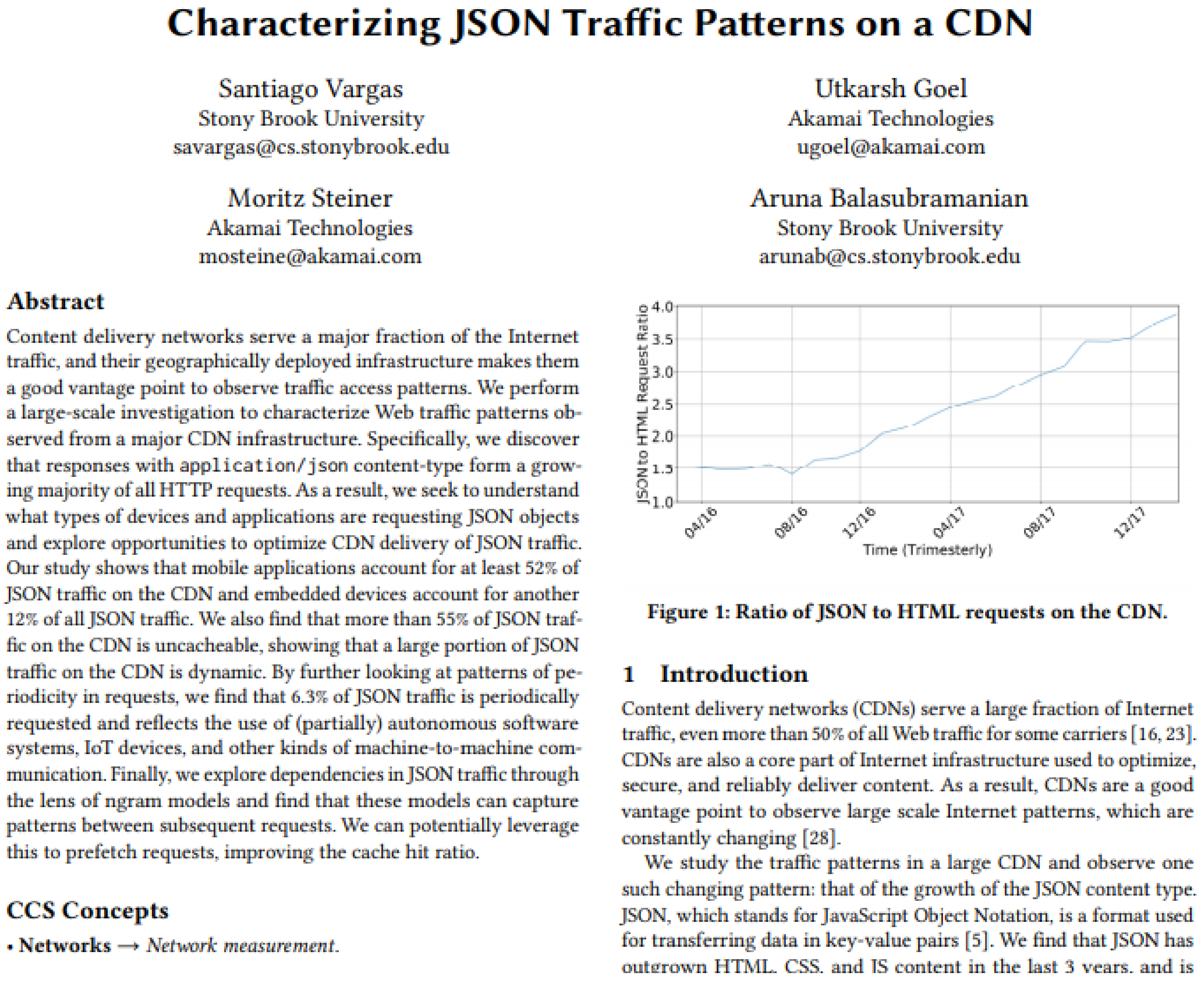
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



Characterizing JSON Traffic Patterns on a CDN

Santiago Vargas
Stony Brook University
savargas@cs.stonybrook.edu

Utkarsh Goel
Akamai Technologies
ugoel@akamai.com

Moritz Steiner
Akamai Technologies
mosteine@akamai.com

Aruna Balasubramanian
Stony Brook University
arunab@cs.stonybrook.edu

Abstract

Content delivery networks serve a major fraction of the Internet traffic, and their geographically deployed infrastructure makes them a good vantage point to observe traffic access patterns. We perform a large-scale investigation to characterize Web traffic patterns observed from a major CDN infrastructure. Specifically, we discover that responses with `application/json` content-type form a growing majority of all HTTP requests. As a result, we seek to understand what types of devices and applications are requesting JSON objects and explore opportunities to optimize CDN delivery of JSON traffic. Our study shows that mobile applications account for at least 52% of JSON traffic on the CDN and embedded devices account for another 12% of all JSON traffic. We also find that more than 55% of JSON traffic on the CDN is uncacheable, showing that a large portion of JSON traffic on the CDN is dynamic. By further looking at patterns of periodicity in requests, we find that 6.3% of JSON traffic is periodically requested and reflects the use of (partially) autonomous software systems, IoT devices, and other kinds of machine-to-machine communication. Finally, we explore dependencies in JSON traffic through the lens of ngram models and find that these models can capture patterns between subsequent requests. We can potentially leverage this to prefetch requests, improving the cache hit ratio.

CCS Concepts

• **Networks** → *Network measurement*.

Figure 1: Ratio of JSON to HTML requests on the CDN.

1 Introduction

Content delivery networks (CDNs) serve a large fraction of Internet traffic, even more than 50% of all Web traffic for some carriers [16, 23]. CDNs are also a core part of Internet infrastructure used to optimize, secure, and reliably deliver content. As a result, CDNs are a good vantage point to observe large scale Internet patterns, which are constantly changing [28].

We study the traffic patterns in a large CDN and observe one such changing pattern: that of the growth of the JSON content type. JSON, which stands for JavaScript Object Notation, is a format used for transferring data in key-value pairs [5]. We find that JSON has outgrown HTML, CSS, and JS content in the last 3 years, and is

Alexander

Why this paper?

What you liked most?

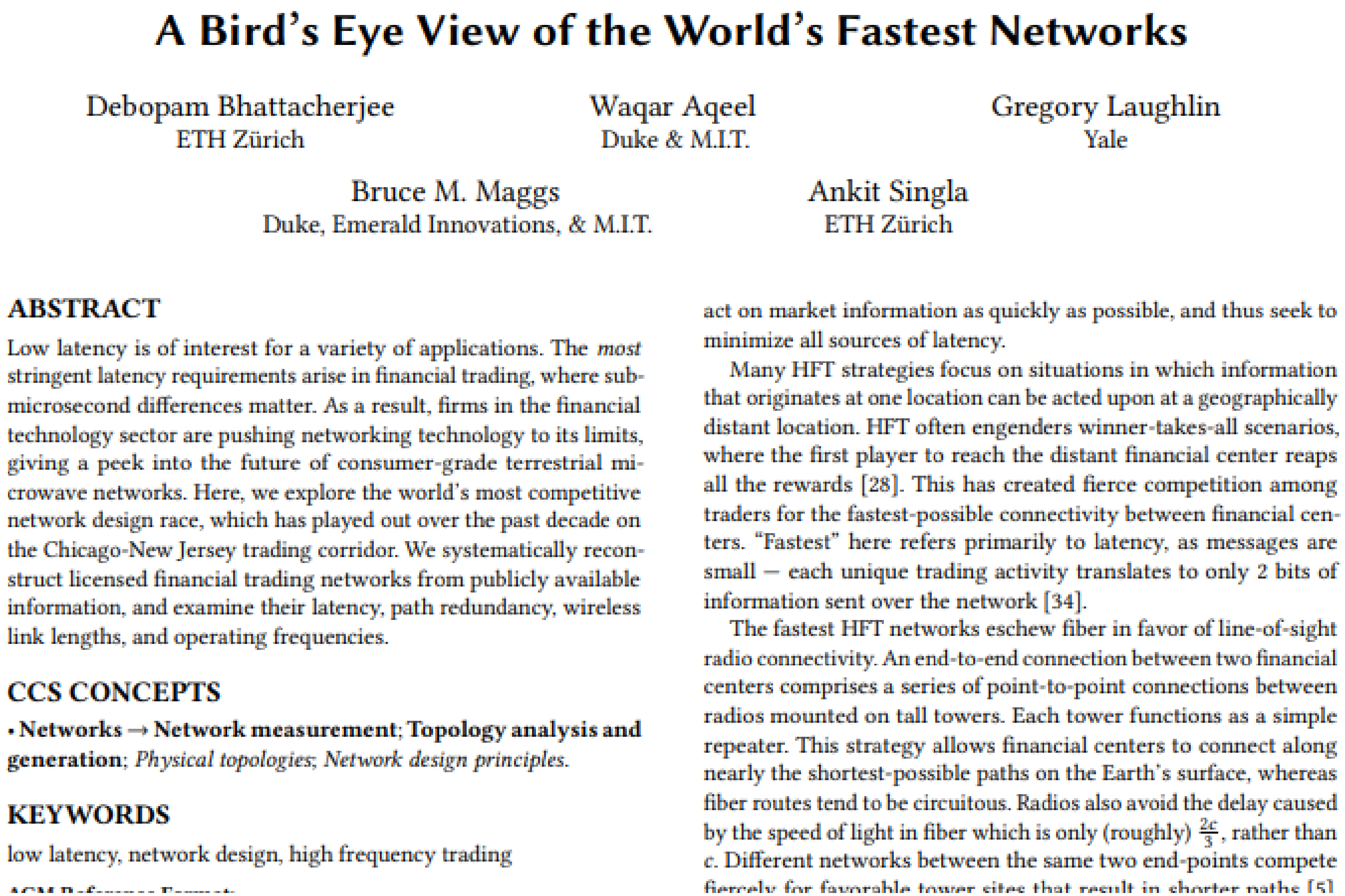
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



A Bird's Eye View of the World's Fastest Networks

Debopam Bhattacherjee
ETH Zürich

Waqar Aqeel
Duke & M.I.T.

Gregory Laughlin
Yale

Bruce M. Maggs
Duke, Emerald Innovations, & M.I.T.

Ankit Singla
ETH Zürich

ABSTRACT

Low latency is of interest for a variety of applications. The *most* stringent latency requirements arise in financial trading, where sub-microsecond differences matter. As a result, firms in the financial technology sector are pushing networking technology to its limits, giving a peek into the future of consumer-grade terrestrial microwave networks. Here, we explore the world's most competitive network design race, which has played out over the past decade on the Chicago-New Jersey trading corridor. We systematically reconstruct licensed financial trading networks from publicly available information, and examine their latency, path redundancy, wireless link lengths, and operating frequencies.

CCS CONCEPTS

• **Networks → Network measurement**; **Topology analysis and generation**; *Physical topologies*; *Network design principles*.

KEYWORDS

low latency, network design, high frequency trading

ACM Reference Format:

act on market information as quickly as possible, and thus seek to minimize all sources of latency.

Many HFT strategies focus on situations in which information that originates at one location can be acted upon at a geographically distant location. HFT often engenders winner-takes-all scenarios, where the first player to reach the distant financial center reaps all the rewards [28]. This has created fierce competition among traders for the fastest-possible connectivity between financial centers. "Fastest" here refers primarily to latency, as messages are small — each unique trading activity translates to only 2 bits of information sent over the network [34].

The fastest HFT networks eschew fiber in favor of line-of-sight radio connectivity. An end-to-end connection between two financial centers comprises a series of point-to-point connections between radios mounted on tall towers. Each tower functions as a simple repeater. This strategy allows financial centers to connect along nearly the shortest-possible paths on the Earth's surface, whereas fiber routes tend to be circuitous. Radios also avoid the delay caused by the speed of light in fiber which is only (roughly) $\frac{2c}{3}$, rather than c. Different networks between the same two end-points compete fiercely for favorable tower sites that result in shorter paths [5]

Janos

Why this paper?

What you liked most?

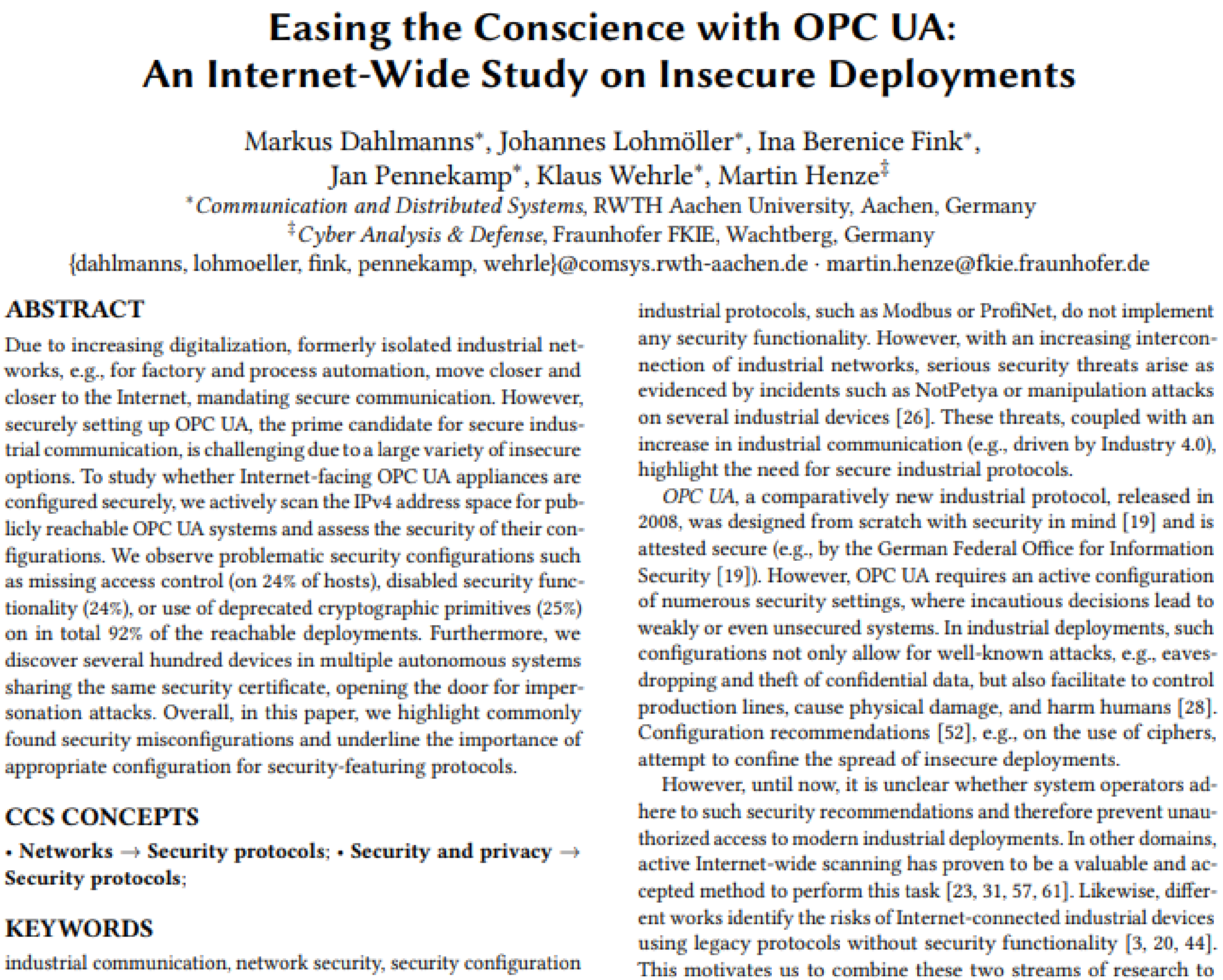
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



Easing the Conscience with OPC UA: An Internet-Wide Study on Insecure Deployments

Markus Dahlmanns[*], Johannes Lohmöller[*], Ina Berenice Fink[*],
Jan Pennekamp[*], Klaus Wehrle[*], Martin Henze[‡]
[*] *Communication and Distributed Systems*, RWTH Aachen University, Aachen, Germany
[‡] *Cyber Analysis & Defense*, Fraunhofer FKIE, Wachtberg, Germany
{dahlmanns, lohmoeller, fink, pennekamp, wehrle}@comsys.rwth-aachen.de · martin.henze@fkie.fraunhofer.de

ABSTRACT

Due to increasing digitalization, formerly isolated industrial networks, e.g., for factory and process automation, move closer and closer to the Internet, mandating secure communication. However, securely setting up OPC UA, the prime candidate for secure industrial communication, is challenging due to a large variety of insecure options. To study whether Internet-facing OPC UA appliances are configured securely, we actively scan the IPv4 address space for publicly reachable OPC UA systems and assess the security of their configurations. We observe problematic security configurations such as missing access control (on 24% of hosts), disabled security functionality (24%), or use of deprecated cryptographic primitives (25%) on in total 92% of the reachable deployments. Furthermore, we discover several hundred devices in multiple autonomous systems sharing the same security certificate, opening the door for impersonation attacks. Overall, in this paper, we highlight commonly found security misconfigurations and underline the importance of appropriate configuration for security-featuring protocols.

CCS CONCEPTS

• **Networks** → **Security protocols**; • **Security and privacy** → **Security protocols**;

KEYWORDS

industrial communication, network security, security configuration

industrial protocols, such as Modbus or ProfiNet, do not implement any security functionality. However, with an increasing interconnection of industrial networks, serious security threats arise as evidenced by incidents such as NotPetya or manipulation attacks on several industrial devices [26]. These threats, coupled with an increase in industrial communication (e.g., driven by Industry 4.0), highlight the need for secure industrial protocols.

OPC UA, a comparatively new industrial protocol, released in 2008, was designed from scratch with security in mind [19] and is attested secure (e.g., by the German Federal Office for Information Security [19]). However, OPC UA requires an active configuration of numerous security settings, where incautious decisions lead to weakly or even unsecured systems. In industrial deployments, such configurations not only allow for well-known attacks, e.g., eavesdropping and theft of confidential data, but also facilitate to control production lines, cause physical damage, and harm humans [28]. Configuration recommendations [52], e.g., on the use of ciphers, attempt to confine the spread of insecure deployments.

However, until now, it is unclear whether system operators adhere to such security recommendations and therefore prevent unauthorized access to modern industrial deployments. In other domains, active Internet-wide scanning has proven to be a valuable and accepted method to perform this task [23, 31, 57, 61]. Likewise, different works identify the risks of Internet-connected industrial devices using legacy protocols without security functionality [3, 20, 44]. This motivates us to combine these two streams of research to

Lukas

Why this paper?

What you liked most?

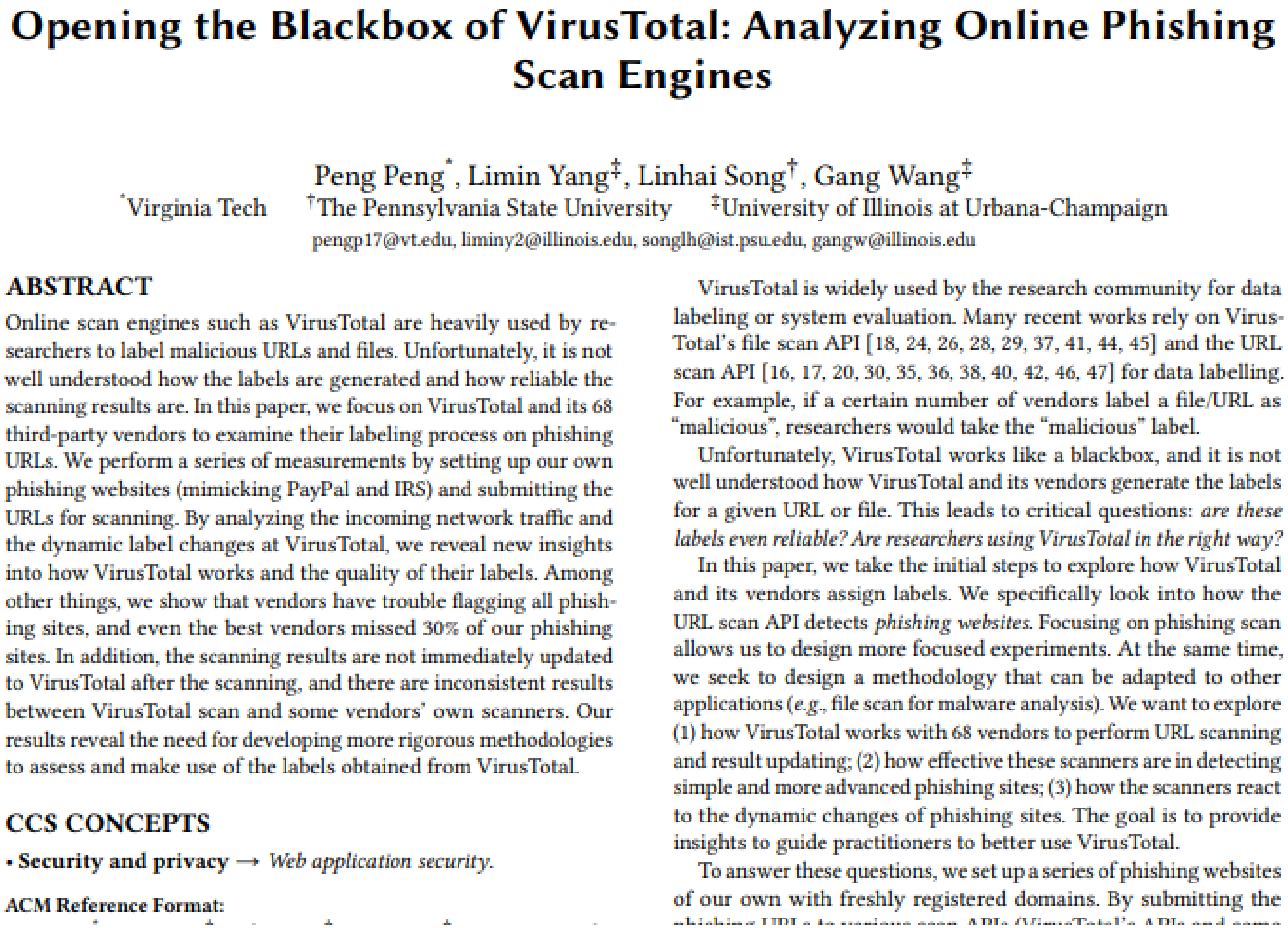
Main **contributions**?

Category and **Context**; e.g.,

- Type of measurements?
- Type of data?
- Type of analysis?
- Type of contribution?

Correctness and **Clarity**

- If not, why not?



Opening the Blackbox of VirusTotal: Analyzing Online Phishing Scan Engines

Peng Peng[*], Limin Yang[‡], Linhai Song[†], Gang Wang[‡]
[*]Virginia Tech [†]The Pennsylvania State University [‡]University of Illinois at Urbana-Champaign
pengp17@vt.edu, liminy2@illinois.edu, songlh@ist.psu.edu, gangw@illinois.edu

ABSTRACT

Online scan engines such as VirusTotal are heavily used by researchers to label malicious URLs and files. Unfortunately, it is not well understood how the labels are generated and how reliable the scanning results are. In this paper, we focus on VirusTotal and its 68 third-party vendors to examine their labeling process on phishing URLs. We perform a series of measurements by setting up our own phishing websites (mimicking PayPal and IRS) and submitting the URLs for scanning. By analyzing the incoming network traffic and the dynamic label changes at VirusTotal, we reveal new insights into how VirusTotal works and the quality of their labels. Among other things, we show that vendors have trouble flagging all phishing sites, and even the best vendors missed 30% of our phishing sites. In addition, the scanning results are not immediately updated to VirusTotal after the scanning, and there are inconsistent results between VirusTotal scan and some vendors' own scanners. Our results reveal the need for developing more rigorous methodologies to assess and make use of the labels obtained from VirusTotal.

CCS CONCEPTS

• **Security and privacy** → *Web application security.*

ACM Reference Format:

VirusTotal is widely used by the research community for data labeling or system evaluation. Many recent works rely on VirusTotal's file scan API [18, 24, 26, 28, 29, 37, 41, 44, 45] and the URL scan API [16, 17, 20, 30, 35, 36, 38, 40, 42, 46, 47] for data labelling. For example, if a certain number of vendors label a file/URL as "malicious", researchers would take the "malicious" label.

Unfortunately, VirusTotal works like a blackbox, and it is not well understood how VirusTotal and its vendors generate the labels for a given URL or file. This leads to critical questions: *are these labels even reliable? Are researchers using VirusTotal in the right way?*

In this paper, we take the initial steps to explore how VirusTotal and its vendors assign labels. We specifically look into how the URL scan API detects *phishing websites*. Focusing on phishing scan allows us to design more focused experiments. At the same time, we seek to design a methodology that can be adapted to other applications (*e.g.*, file scan for malware analysis). We want to explore (1) how VirusTotal works with 68 vendors to perform URL scanning and result updating; (2) how effective these scanners are in detecting simple and more advanced phishing sites; (3) how the scanners react to the dynamic changes of phishing sites. The goal is to provide insights to guide practitioners to better use VirusTotal.

To answer these questions, we set up a series of phishing websites of our own with freshly registered domains. By submitting the phishing URLs to various scan APIs (VirusTotal's APIs and some

Extra slide …