Internet Outage Detection & Analysis

http://www.caida.org/projects/ioda

Alberto Dainotti alberto@caida.org









FOCUS

Macroscopic Internet Outages

- Large-scale Internet connectivity disruption (keywords: Internet "outage", "black out", "shutdown", "kill-switch")
- E.g., a connectivity black-out significantly affecting a large network operator or a large geographical area
- Potential causes: natural disasters, cyber attacks, physical attacks (terrorism, war, ...), bugs and misconfigurations, government orders, ...



INTERNET OUTAGES why so relevant?

Public Safety

The Internet is a critical infrastructure

Virtually every element of modern life is now dependent on cyber infrastructure. As a result, our Nation's economic and national security relies on the security of the assets and operations of critical communications infrastructure. Past terrorist attacks and catastrophic natural disasters emphasized the need to focus our national attention on protecting the Nation's critical infrastructure and making it more resilient. Moving forward, it is essential that public and private sector partners adopt a coordinated approach to achieve joint goals for our communications infrastructure.

While the Communications Sector has few significant dependencies, other critical infrastructure sectors are dependent on the Communications Sector. As such, the Communications Sector is one of the few sectors that can affect all other sectors. At a minimum, each sector depends on services from the Communications Sector to support its operations and associated day-to-day communication needs for corporate and organizational networks and services (e.g., Internet connectivity, voice services, and video teleconferencing capabilities). Some sectors



Center for Applied Internet Data Analysis University of California San Diego



US Department of Homeland Security, National Infrastructure Protection Plan (NIPP) 2013

INTERNET OUTAGES

why so relevant?

Financial and reputational costs

Services are meant to be always on

CLOUD

5-minute outage costs Google \$545,000 in revenue

DYLAN TWENEY @DYLAN20 AUGUST 16, 2013 4:06 PM





Center for Applied Internet Data Analysis University of California San Diego

How Much Will Today's Internet Outage Cost?

Some companies lose tens of thousands of dollars for every *minute* of a DDoS attack.

ADRIENNE LAFRANCE | OCT 21, 2016 | TECHNOLOGY

INTERNET OUTAGES

why so relevant?

Human Rights

censorship and political violence



Libya Pulls the Plug







Center for Applied Internet Data Analysis University of California San Diego

HOME » NEWS » WORLD NEWS » AFRICA AND INDIAN OCEAN » EGYPT

How Egypt shut down the internet

Virtually all internet access in Egypt is cut off today as the gc to contain the street protests that threaten to topple Presider Mubarak.





Email



Police fire tear gas towards protesters in Suez, Egypt Photo: AFP/GETTY

INTERNET OUTAGES

why so relevant?

Human Rights

censorship and political violence

QUARTZ Africa

#KEEPITON

More African governments blocked the internet to silence dissent in 2016



OCTOBER 2016



Center for Applied Internet Data Analysis University of California San Diego Internet shutdowns cost countries \$2.4 billion last year

INTERNET OUTAGES So what's the problem?

There is lack of understanding of when, how often, why, how large Internet outages happen

There is lack of a general rigorous framework to obtain *empirical data* about - and to characterize - these events



IODA PROJECT

Started in Sep. 2012 with an NSF award from a program to *Transition to Practice* Cybersecurity research

Funding also provided by DHS S&T

• **Goal:** prototype an operational capability to monitor the Internet 24/7 to detect and analyze Internet blackouts affecting large networks / geographical areas

Project Website: <u>http://www.caida.org/projects/ioda</u>
Experimental service: <u>https://ioda.caida.org</u>







methodologies used for post-event manual analysis

 Country-level Internet Blackouts during the Arab Spring

> Dainotti et al. "Analysis of Country-wide Internet Outages Caused by Censorship" ACM Internet Measurement Conference 2011



EGYPT, JAN 2011 GOVERNMENT ORDERS TO SHUT DOWN THE INTERNET

• Natural disasters affecting the infrastructure

Dainotti et al. "Extracting Benefit from Harm: Using Malware Pollution to Analyze the Impact of Political and Geophysical Events on the Internet" ACM SIGCOMM CCR 2012



Center for Applied Internet Data Analysis University of California San Diego



<u>Japan, Mar 2011</u> Earthquake of Magnitude **9.0**

OUR METHODOLOGY combining various types of measurements

multiple types of sources for inference

- Routing Plane [BGP]
- Data Plane
 - Active probing
 - Passive traffic analysis [IBR]

• meta-data to extract liveness signals for various aggregations (e.g., countries, ASNs)

visualize and compare signals



BGP

IBR

IBR

"Extracting benefit from harm."

• Use Internet Background Radiation (IBR), mostly generated by malware-infected hosts as a "signal"

INFECTED HOST RANDOMLY SCANNING THE INTERNET

IBR

UCSD NETWORK TELESCOPE DARKNET XXX.0.0.0/8





$B(_{1}P$ Monitoring Global Internet Routing

- Operational BGP measurement projects establish peering sessions Router / Monitor Internet with ASes to receive their routing tables (no exchange of other traffic) Internet Exchange
 - RouteViews (Univ. Oregon): 371 peers • RIPE RIS (RIPE NCC): 508 peers



BGP



http://www.routeviews.org https://www.ripe.net/data-tools/stats/ris





hitting the news



TECH EUROPE

'Internet Background Radiation' Reveals Disasters and Censorship

By NICK CLAYTON

Mar 12, 2012 7:10 am GMT

There is something satisfying about finding something useful to do with garbage. Researchers at UC San Diego, California, have apparently found a way of using the data traffic generated by malware and and malicious scanning to detect Internet outages that may be caused by natural disasters or censorship.

Most Popular Vi

Super Bowl The Most B About Ads



hitting the news



Deploying Gasel Computing the to deale test commuting building taken and the barries test commuting test commuting test commuting ciety | DOI:10.1145/2330667.2330674

Garbage In, Info Out

Security researchers used malware to investigate large-scale Internet censorship in Egypt and Libya.



as governments scrambled to stem the flow of information among the people and with the outside world. Soon after the Arab Spring, an international team of computer scientists began analyzing precisely what happened in two of the affected nations, Egypt and Libya. Their fine-grained analysis of Internet censorship in these countries, which won this year's Applied Networking Research Prize from the Internet Research Task Force, emerged in part from a surprising source of data; malware.

"We've never before seen an entire country disappear from the Internet for several days," says Alberto Dainotti, lead researcher with the Cooperative Association for Internet Data Analysis (CAIDA) at the University of California,



Marrina Krakovsky

A crowd scene in Cairo's Tahrir Square during Egypt's five days of internet outage last year; the large sign on the KFC window says "Awiz Internet" ("We want Internet").



post-event manual analysis

4 months of work



Dainotti et al. "Analysis of Country-wide Internet Outages Caused by Censorship" ACM Internet Measurement Conference 2011



Center for Applied Internet Data Analysis University of California San Diego

Analysis of Country-wide Internet Outages Caused by Censorship

Alberto Daino ti Unversity of Napcil Federice I Claudie Squarcela. Roma Tre University Emile Aben FIPE NOC emile.aben@ripe.net alberts@unina.it squarce@dia.uniroma3.t Kimberly C. Clarty CADA/UCSD Marco Chiesa Roma Tro Universit kr@caida org chieza@cia uniroma3.it Mohele Russo Antonio Pescapé University of Napoli Federico II University of Napoli Federice I

ABSTRACT

In the first manhs of rapial in several Netti partexts and threats of a of these charaptions. In surplus on manipies of to academic resources data, association have p the material physics in an and MaxMind's protorate to determine the within each e-catry or est to BGF announced using publicity available rope. We then analyses performent discution control plane and data to names down which it phonented in a given p datasted what we belie based blocking before disconnection Our net direct outages errainilla



Categories and 5 C.2.) [Settork Operation] C.2.) [Local and Wide

General Terms Monument Socialty

Parameter to make digit-presendor disastronomic

presention Assessment nu-net made or distributes for hear function are the ten-renalitied, report to the ten-procession and ten to the DACTJ, Normalier 2-4, 2 Copyright 2011 ACM 978



Figure 2. UCD distance calls coming from labors Labols A, B, C in discuss the data compare. Aplica labolar D-1 and KL are due to index and former we derive the combine combine.

eisted to potests in the country. The web site of the Ministry of Communications (work, georg) was straight with a randomly spouled both areas, just before the source sourced, on humany 25 m. protocol of the state protocol of the state and MA REALIZED HOUSE.

Dra Bernary Erte well star of the Epyteian Ministry of Etector (treasonal cyptupe up) was impaired by two Bolf, stracker just of the fire and of the transmitter trans 1 - 42 to 12.39 (3417) and from 1238 YO LC H. 1891. THE DATE IF AMPLITUDE AND ADDRESS ADDRESS. time the fact after, from 00:00 to 00:42 GMT. In this case the estimated under transmoster smaller, around "to packets pre-second

\$2 Libya

2.2.1 CUSTARY Ldya's horner infrareation is over more protects manipul-tion famility of a pulping from in d year d one and formational connectivity is provided by only two subcarries cables, both en-ing in Trapol [20] subths because infrareation of the

(a) in Tep-D [10] works between infrastructure is durational by a single state recent, AE Wavehy found was relate AE as backups strate presenter in LDV, is successful to the contrast of the low of publicly devices and (Dyne 1). Figure 3 shows the infin-ed publicly devices are (Dyne 1). Figure 3 shows the infin-element by the USED network taken up to the low devices in the vol excerption in the low publicly and and a three will excerption in the low publicly and 20 show between by the USED network taken up to the low of the LD and Contrast true fullered taken in the low publicly and 20 show to the low duration in the low of the low of the low of the low of the state of the result of a collection interactivity and the low. the sight of the propiet is in difficult to interpret what a welly hap putting it Librar because of the cost with

3.2.2 Consers Wale inf

3.2.2 Unterpresentation of the second sec delenant in three you willdares. These tooles I'rd parties delignes in 24ye was withdraw. These today if the parking with according by cylindrace, the continuous approximation the semaning IP-4 paths are managed by heAS2. As of May 2011, here were so lively prefers in disPMC soldspace (The far-Colys, The Model all Dynelocities advects for the path D are-compared if anges in 24ye, all part of an ecompassing P-4.

e sectores stillets relation divide. (2.052 of becaused sites the Mullis Real, Asia and Africa. The covering DVI parts also covering 100 Pringeric second offer counties prelocations of the Nullis Real, we consider a second of the counties and the prelocation of the Second S UCM, carket generally downed a significant amount of unov Which will be coming to the first state of the first state of the first energy (draw Shapadan and lay). This level of hadapenad ref. is indexton a population control on any full rate yes often by Catilities as other advant, advanta inference of attents, con-Contract Tell Print the restorate along printers of a strategy of distant. Tell'rest the restorated stor particle anisotro of subar hipportal to folly arbitrar an exemption hand on addition system met mangest by the restoratements private. Comparing theorem 3.200 metals and thereit behavior that conflicts and operations resents (17): the second radies for end

cately creatily DC volumeds. The DCP darkers equiva-primary 15 around 2.131.35.077, easing manifesting for their of-creases: darked battle from Larns continue conducts with works an Islyon with over by their Keyners (2011) for trades approxinstely one loss lots, it 2005.22. In contrast, the latence con-age as shown by the seleccupe data and reported by the news [17] Initial unit approximately Person V & 4 6:12 UTC The Inflate suggests that a deliversit disruption inclusion on surfact delivering strongy apparently adopted indexeparity in the visial entropy and exceptional by the constant day would - way through being most day



Forms 11: The fructive Libran concess the maximum labels 0.750 , datase conting tools (Eqs. (b) shifting of Thipse Bod produce in RCF data have Research REF (PP) PD relations to the structure of the label of the data shifting the concentration of concentrating the second magnetized and the dataset transformed as a structure of the second structure of the second

IODA GOALS

applied research





IODA CHALLENGES

Why this is a tough problem

- refine/extend inference methodologies
- automate inference methodologies
- complex data
- noisy data
- big data
- heterogeneous data
- velocity
- lack of tools
- distributed system
- visualization for dashboards and data exploration
- lots of infrastructure to maintain/operate
-

Caloa

• all with relatively few money/people/time..

IODA FIRSTYEARS documenting events on our blog

Syria disappears from the Internet — Nov 2012



Time (UTC)



IODA FIRSTYEARS documenting events on our blog

Time Warner Cable outage 27th August 2014





IODA FIRSTYEARS documenting events on our blog

Time Warner Cable outage 27th August 2014





IODA AFTER 2YEARS

live Internet monitoring

In 2014 we made it possible for anybody to follow the North Korean disconnection almost live CAIDA @caidaorg · Dec 23

Follow outages in #NorthKoreaInternet in almost real-time (30min delay) at charthouse.caida.org/public/kp-outa...





Center for Applied Internet Data Analysis University of California San Diego

https://ioda.caida.org/public/kp-outage

IODA AFTER 4 YEARS (TODAY)

live detection and monitoring

loo Dashboard Exp	vlorer Alert Feed Projec	d Info Acknowledgements Help						Logged	in as alberto - I	loyart 🛄
a toryago - Now						Outage Severity Overview				
Country Outages										
Geographical Distribution 💠 🔤 Intend 💥 🖸						Show 10 \$ entries Secrets				
Overall Score BCP A	ctive Probing					Country	La Overall Score La	Active Probing	BGP J.	Darlmet 17
(+)		thong and	Paurin	Range 0.00	- 3313	Bolivia	589M	10.5k	56.3k	
L Drese	R	al Private Contract	1 5		587 33R	Philippines	30.3k		30.3k	
	North North	Constant (Amin) Konstante	Stangel a			New Zealand	25.6k		25.6k	
n Deck Sona	Ationtic	Spain Turks Unskiew	thin	an	Pecífic	Iszael	21.7k		21.7k	
Nata	N. I	- Mgeria LByn Fund Sault Balistan	man.		Decon	Argentina	20.5k		20.5k	
	5.00	end Hall Hax Subn	a dia			Guyana	18.4k	18.4k		
	Carriery to	Dirizente	inter variation of			Martinique	1B.Ok		18.0¥	
	Fer. Bet.	Tebre	- Storak	Texables Select		Bulgaria	16.9k		16.9k	
	50	ata Thispear Inc	9an			Reunion	16.2k		16.2k	
South	Atta	offe sunari de	an M	utrala	51	Syrian Arab Republic	5.65k	5.65k		
Ozean					G.	Showing 1 to 10 of 10 ontrice				Previous Meat
13			10	fe Majulii-Dişerkiret	Macconstatures					
a T STM	Nor 13 3AM	54W 22M	12°M	an	SIN					
ð Senes: 50 🐨 Points: 14420 Data w	asplution: minute		February 12, 20	177:07pm - February 19,	20177.07pm					



Center for Applied Internet Data Analysis University of California San Diego

https://ioda.caida.org

high-level system view







high-level system view





Center for Applied Internet Data Analysis University of California San Diego



high-level system view



high-level system view

Center for Applied Internet Data Analysis University of California San Diego

PART RUNS ON GORDON! NSF XSEDE allocation

high-level system view

high-level system view

high-level system view

Center for Applied Internet Data Analysis University of California San Diego

IODA DEMO

MANY SUB-SYSTEMS

each with its own challenges

AN EXAMPLE: BGPSTREAM

efficient scalable processing of Internet routing data

BGPSTREAM

A Border Gateway Protocol (BGP)

- BGP is the central nervous system of the Internet!
- There is almost 40 years of highly relevant research on BGP (and still going..)
- Operators collect, analyze and monitor BGP data to learn about and solve Internet routing problems
- There was no efficient way of processing large amounts of distributed and/or live BGP measurement data

BGPSTREAM

efficient scalable processing of Internet routing data

BGPSTREAM

efficient scalable processing of Internet routing data

Center for Applied Internet Data Analysis University of California San Diego

caida

BGPSTREAM IN IODA

the toolchain we needed to process routing data

BGPSTREAM IN IODA

32 BGPCorsaro instances processing data from ~500 routers

BGPSTREAM BGPSTREAM

A research + development project of its own

• We published a paper presenting BGPStream at the ACM Internet Measurement Conference 2016

• Includes analysis of massive amounts of historical BGP data using Apache Spark running on SDSC's Comet!

• Alistair has been awarded the **IRTF's Applied Networking Research Prize** for this paper and will present it at the next Internet Engineering Task Force (IETF) meeting.

• Users worldwide (including students), code contributions, and several collaborations:

• **Cisco Systems** awarded us ~\$100k to collaborate to extend BGPStream functionalities to support their open-source BGP Monitoring Protocol framework • • • • • • • • • • •

cisco

BGPStream: a software framework for live and historical BGP data analysis

Chiara Orsini ¹, Alistair King¹, Danilo Giordano¹, Vasileòs Giotsas¹, Alberto Dainotti ¹CAIDA, JC San Diego ²Politeorico di Torino

ABSTRACT

We present BGPB/ream, an open-source selfware frame work for the analysis of both historical and real-time Border Gateware Protocol (BGP) measurement data. Although BGP is a crucial operational component of the Internet infrastructure, and is the subject of research in the areas of Internet performance, security, topoltary, protocols, nonsumba, visc., there is not reflicing and of processing large amounts of distributed and/or live BGP measurement data. BGPS/ream fills this gap, enabling efficient lowestigation of events, rapid prototyping, and building complex tools and large-scale monitoring applications (e.g., detection of connectivity discuptions or BGP blacking attacks). We discuss the goah and auchinecture of BGPSream. We apply the components of the functivers to different scenarios, and we describe the development and depioyment of complex services for global Internet monitoring that we built or top of it.

1. INTRODUCTION

We present DEPStream, an open-score software framwork' for the analysis of historical and live Border Gato way Protocol (BGP) measurement class. Although BGP is a crucial operational component of the Internet in frastructure, and is the subject of findamental research (in the areas of performance, security, topology, protocols, economy, etc.), there is no efficient and easy way of processing large amounts of BGP measurement data. REPStream 6th this gap by walking auxiliable as st of

¹BGPStream is Satributed with the GPL v2 license and is available at bgstream.caida.org.

Perturbation to make digad of task copies of all or get, of this work for general or characters are is general without for provided the origins at too made a distributed for profile or consensual adventage and that origins to not made a distributed for profile or consensual adventage and that origins the advent of the field consistence for the page. Copyright is comparing the fields work remaining the profile or contemposite of the profile condition line, tragation prior operating profiles for page to be consense are consensing with condition, tragation prior operating profiles for page to be consensed on a particular base provided the particular start of the particular start of the particular start profiles are start or particular start of the particular start

IBNC 2016. November 14 – 16. 2016. Sama Monicy, CA, USA 0.2016 Copyright lask by the sensorbathenity. Meloatene rights hanned in ACM 1989 5761–490-4052-20141. Sci 100 DOI: http://dx.doi.org/10.1143/2987148.29871482 APIs and tokic for processing large amounts of live and historical dats, thus supporting investigation of specific events, napid prototyping, and unliding complex tools and efficient largo-scale monitoring applications (e.g., detection of sumscrivity disruptions of BGPdistribution). We discuss the goals and architecture of BGP-Stream and we show how the components of the framework can be used in different application accurates.

2. BACKGROUND BGP Data at Router Level

The Border Cateway Protocol (BGP) is the do-facto standard intro-domain routing protocol for the Internet: its primary function is to exchange reachability information among Actenorous Systems (ASPn) [1]. Each AS announces to the others, by means of BEP update messages, the routes to in local prefixes and the preferred routes learned from its acighbors. Such neusages provide information shout how a destination ran be reached through an ordered list of AS hops, called an AS peth.

an AS path. A BGP rovier maintains this rachability information in the Revealing Information Base (RHB) [21], which is structured in three sets:

- Adj-RIB-Ju: routes learned from inbound upcate messages from its neighbors.
- Loc-RD: routes selected from Adj-RIBs-In by applying local policies (e.g., shortest path, peering relationships with asighboot); the router will instell these contex in the routing table to catalidat where to forward packets.
- Adj-RHs-Out: routes selected from Loc-RIB, which the router will announce to its neighbors, for each neighbor the router creates a specific Ady-RJB-Out haved on local politics (e.g., poering relationship).

BGP Data Collection

Some operators make BGP routing information from their routers smallable for monitoring, troubleshoung and research purposes. BGP indeing glasses give users imited (e.g., read-only) access to a command line interface of a router, or allow them to download the ASCII

IODA SW SPIN OFFS

open-source frameworks of more general utility

ONGOING COLLABS

Academia, Industry, Government

Collaboration with Industry

We are collaborating with Comcast researchers, who are using IODA to support their own research on Internet reliability and performance. In addition, Comcast, through their Innovation Fund provided a research grant for the development of visual interfaces to monitor and characterize Internet outages.

cisco. <u>Cisco</u>

We established a collaboration with researchers at Cisco Systems, who are using BGPStream and are collaborating in extending it to support internal and open source projects carried out by Cisco, such as the OpenBMP implementation of the BGP Monitoring Protocol.

Public Safety

The Public Safety and Homeland Security Bureau (PSHSB) of the Federal Communications Commission (FCC) has the responsibility for ensuring that communications networks are reliable, resilient and secure. To accomplish this task, the PSHSB developed a data-driven process centered on collecting information on and performing analyses of communication outages. CAIDA had several meetings with the FCC to discuss results of the IODA project, providing the FCC with additional insight into the complexity of Internet outage monitoring and to discuss technology transfer of some of these research results and infrastructure capabilities.

Also, research collaborations with networking and poli-sci researchers

IODA FUTURE

next steps

- Collect feedback
- Provide to 3rd parties live+historical alert data feeds through the DHS "Information Marketplace for Policy and Analysis of Cyber-Risk & Trust" program
- Infrastructure Improvements/Maintenance/Documentation etc.
- Research on improving and cross-validating inferences
- Integrate other data sources

THANKS www.caida.org/projects/ioda

