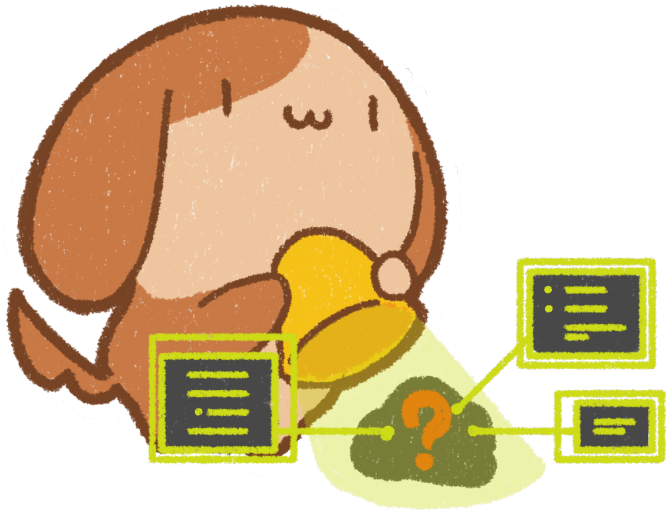
When BGP goes MAD: Multi-scale Anomaly Detection in Internet Routing

Justin Loye, Esteban Bautista, Romain Fontugne



When the Internet goes mad

- The Internet is a critical infrastructure
- Outages happen regularly and have economic and societal impacts:
  - Facebook 2021: all services down, massive loss of revenue
  - Italy 2023: 30% of its population disconnected
  - Cloudflare 2025: planes grounded worldwide

2021 Facebook outage

20 languages

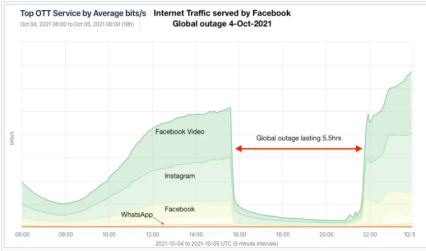
Article Talk Read Edit View history

From Wikipedia, the free encyclopedia

This article **may be excessively based on old or outdated news reports.** Please update this article with better sources. *(February 2025)* (*Learn how and when to remove this message*)

On October 4, 2021, at 15:39 UTC, the social network Facebook and its subsidiaries, Messenger, Instagram, WhatsApp, Mapillary, and Oculus, became globally unavailable for a period of six to seven hours.[1][2][3] The outage also prevented anyone trying to use "Log in with Facebook" from accessing third-party sites.[4]

During the outage, many users flocked to Twitter, Discord, Signal, and Telegram, resulting in disruptions on these sites' servers.[9] The outage was caused by the loss of IP routes to the Facebook Domain Name System (DNS) servers, which were all self-hosted at the time.[10][5] Border Gateway Protocol (BGP) routing was restored for the affected prefixes at about 21:50, and DNS services began to be available again at 22:05 UTC, with application-layer services gradually restored to Facebook, Instagram, and WhatsApp over the following hour, with service generally restored for users by 22:50.[11]



Traffic volume for Facebook services on October 4, 2021, with a drop during the global outage

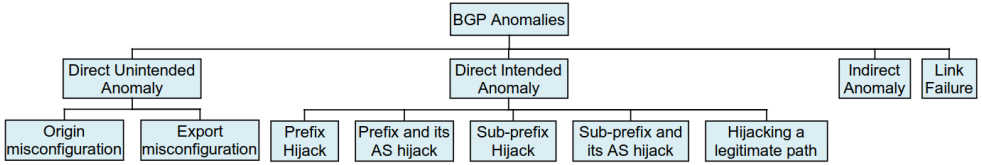
The New York Times

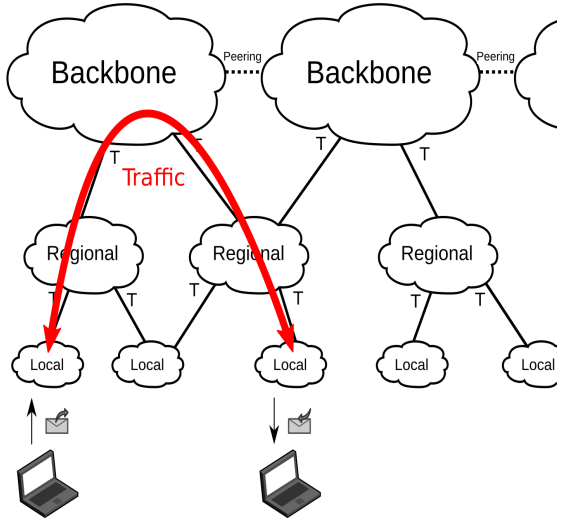
Cloudflare Says It Has Resolved Outage That Disrupted Parts of the Internet

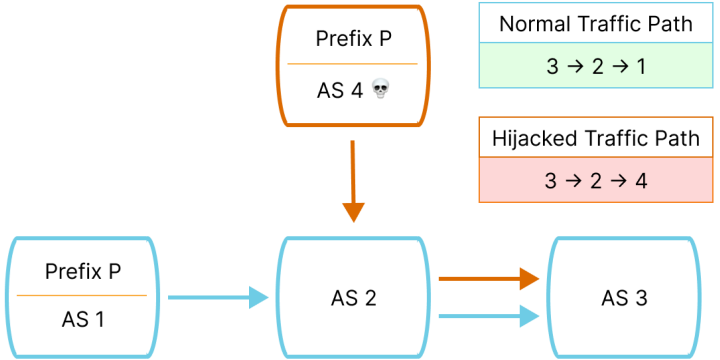
Services from Cloudflare, a software company, underpin thousands of websites, including X, Spotify and OpenAI. The company said a crash in a software system was to blame.

When BGP goes mad

- BGP, de facto protocol for routing Internet traffic. Coarsest Internet topology
  - Around 80k Autonomous Systems: ISPs, Universities, CDNs, ...
  - Internet connected with transit provider, customer and peer relationships
- Internet outages are linked to BGP:
  - Re-routing: BGP prevents connectivity loss
  - BGP as a culprit:
    - BGP misconfiguration: fat fingers, route leaks
    - BGP hijacks
- Can we use BGP routing information to detect outages?







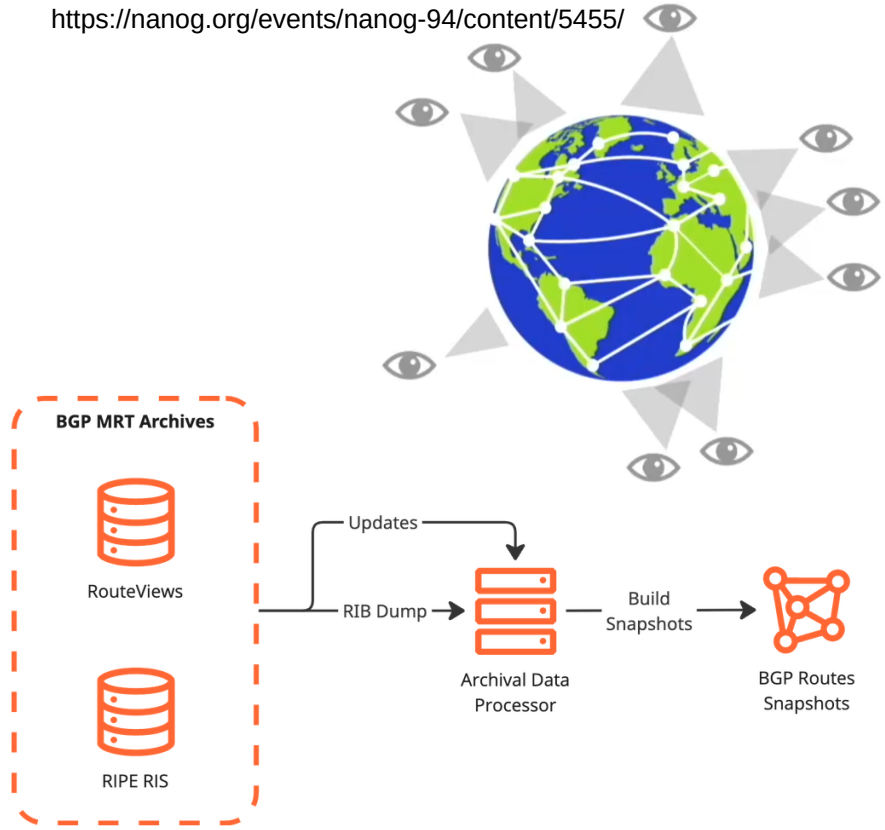
[1] Al-Musawi, Bahaa, Philip Branch, and Grenville Armitage. "BGP anomaly detection techniques: A survey." IEEE Communications Surveys & Tutorials 19.1 (2016): 377-396.

Source: https://blog.cloudflare.com/bgp-hijack-detection/

BGP for Internet monitoring

- 25 years of public BGP routing data is archived and available from RIPE RIS / RouteViews:
  - Peer ASes (e.g., IIJ) connect to a Route Collector (e.g., route-views.wide) and disclose their full routing table (RIBs)
- Data is published in near real-time
  - Every 2 hours: full RIB dump
  - Every 5 minutes: updates (ms resolution) of the lastest RIB

https://nanog.org/events/nanog-94/content/5455/



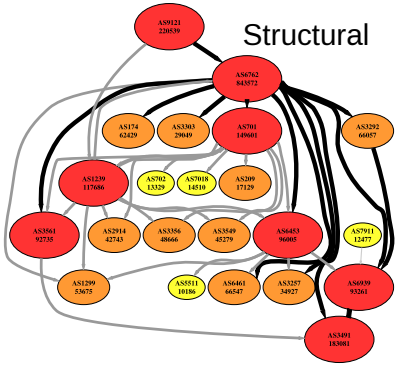
https://blog.cloudflare.com/bringing-connections-into-view-real-time-bgp-route-visibility-on-cloudflare/

BGP anomaly detection landscape

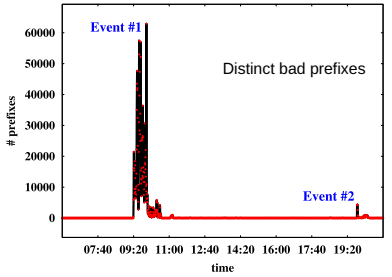
- Live monitoring
  - Commercial systems:
    - Cisco ThousandEyes
    - Cloudflare Radar
    - …
  - Prefix visibility with IODA
- Academia:
  - 20 significant methods surveyed in 2016 [1]
  - 178 surveyed in 2024 [2]
    ”large numbers of features, parameters, domain-specific tuning, and training, often contributing to an unacceptable computational cost”

- Challenges:
  - Near real-time constraint
  - Capture both temporal and structural patterns
  - Generality: avoid feature engineering
  - Unsupervised: Little ground-truth data, expensive training, concept drift

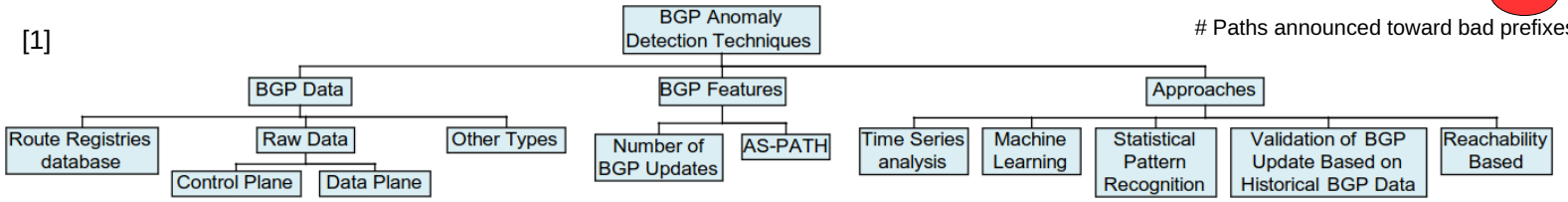
https://archive.nanog.org/meetings/nanog34/presentations/underwood.pdf



Paths announced toward bad prefixes



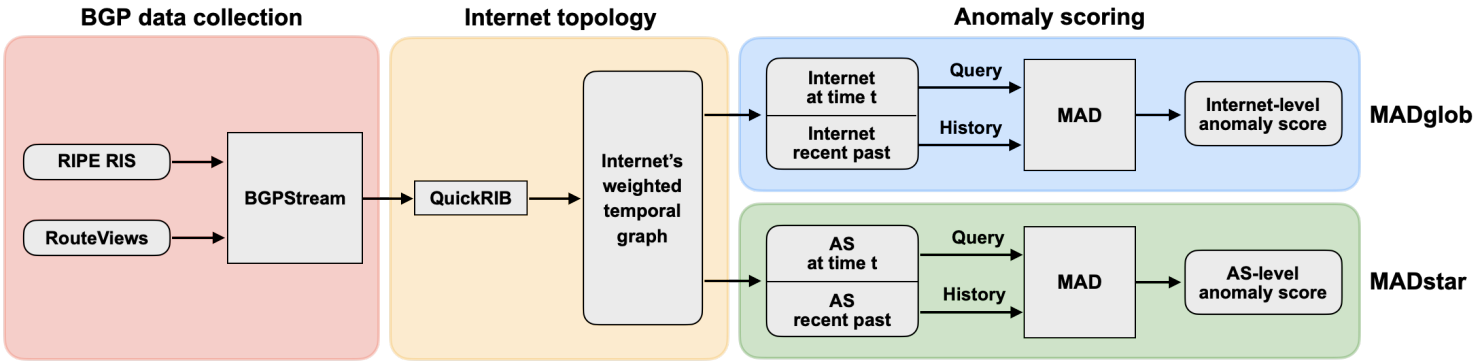
[1]



[2]Scott, Ben A., Michael N. Johnstone, and Patryk Szewczyk. "A survey of advanced border gateway protocol attack detection techniques." Sensors 24.19 (2024): 6414.

MADBGP: Multi-scale Anomaly Detection in BGP routing data

- We introduce MADBGP: end-to-end anomaly detection pipeline

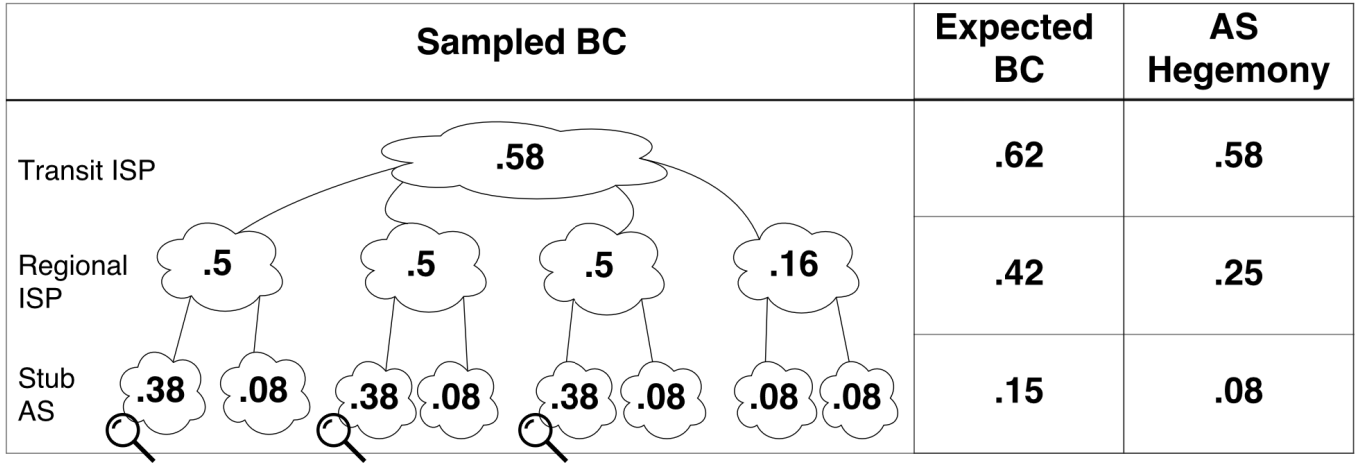


- Global AD
- AS-level AD

- Contributions
  - QuickRIB: an efficient, extensible tool for building high-frequency RIBs from public BGP data
  - Hegemony graphs: unbiased Internet topology representation
  - Extend the MAD graph anomaly detection algorithm to weighted graphs
  - Extend the benchmark dataset

Internet topology creation

- Topologies constructed from adjacency in AS-paths are biased and prone to noise
- AS-hegemony was introduced to reduce bias when computing a node (AS) centrality
- We adapt it to link centrality to build more realistic AS topologies

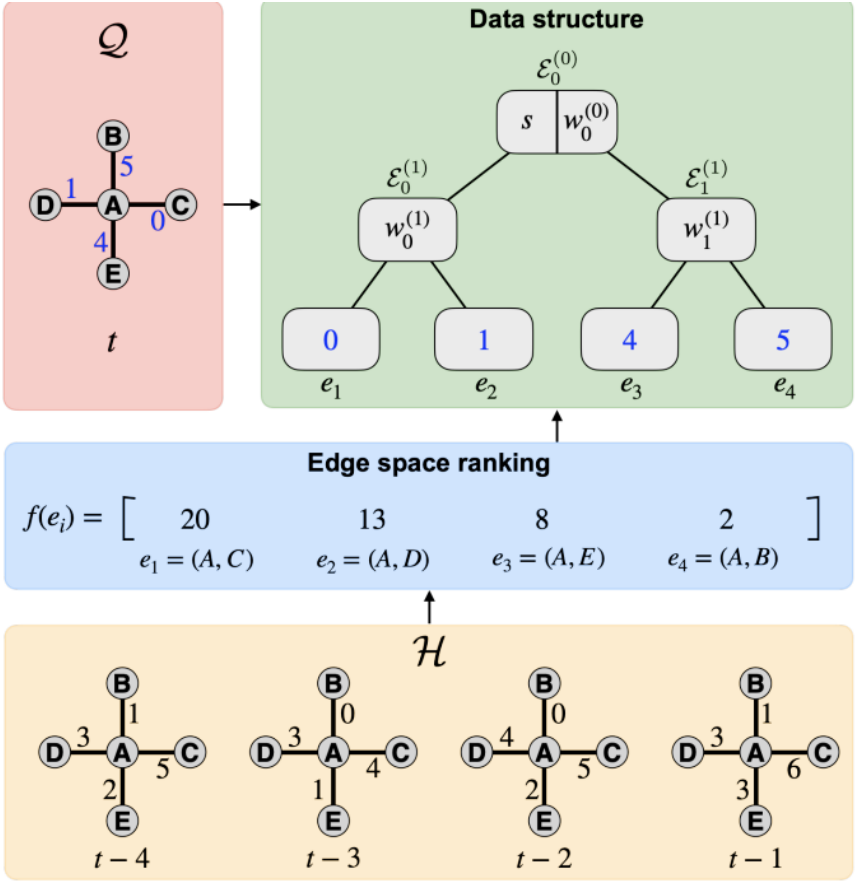


QuickRIB: efficient and friendly BGP data analysis

- Produce high frequency RIBs (few hours → min res):
  - Build a RIB from RIB files of multiple collectors
  - Use BGP updates to reconstruct intermediate RIB states
- Define analysis modules:
  - Applied to the whole RIB, updates, or other analysis modules
  - Built-in modules: IODA, Hegemony, BLT (update classifier)
- We use QuickRIB to build high frequency hegemony graphs
  - Lightweight computation: use BGP updates to update the topology

MAD: Multi-scale anomaly detection

- MAD: temporal graph anomaly detector
- Goal: quantify how active (resp. inactive) graph links become active (inactive)
- Achieved by applying a sliding-window temporal AD to a graph Haar wavelet decomposition



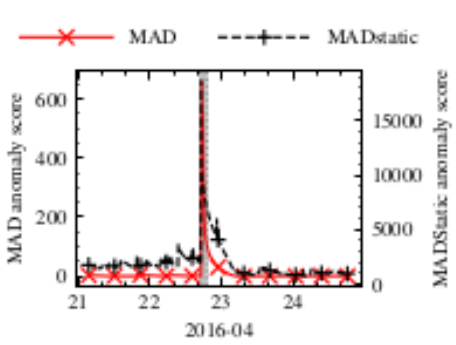
Evaluation setup

- We use 4 well studied global outages
- We introduce 6 new outages with more collectors
- For AS-level AD, we identify potentially impacted ASes
- Realistic time ranges:
  - BGP anomalies of 2 hours
  - Experiment of 5 days

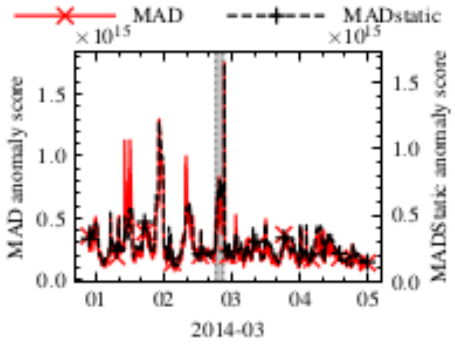
Scenario	Year	Start	End	Collectors	Outage start	Dur.
aws [29]	2016	04-20 16h	04-24 19h	rrc04	04-22 17:10	1:55
indosat [44]	2014	03-31 16h	04-05 00h	rrc04	04-02 18:25	2:30
tm [45]	2015	06-10 08h	06-14 16h	rrc04	06-12 08:43	3:02
ttnet [37]	2004	12-22 08h	12-27 00h	rrc05	12-24 09:20	10:27
verizon [41]	2019	06-22 08h	06-26 16h	rrc00,01	06-24 10:30	2:00
allegheny				rrc03,12		
iran [27]	2024	07-09 00h	07-13 16h	rrc00,01	07-10 11:00	3:20
				rrc03,12		
china [10]	2019	06-04 08h	06-08 16h	rrc00,01	06-06 09:43	2:00
europeanmobile				rrc03,12		
mainone [19]	2018	11-11 16h	11-16 00h	rrc00,01	11-13 21:13	2:00
				rrc03,12		
pakistan [39]	2008	02-22 16h	02-27 00h	rrc00,01	02-24 18:10	2:14
youtube				rrc03,12		
italy [42]	2023	02-03 00h	02-08 00h	rrc00,01	02-05 11:00	5:00
				rrc03,12		

Anomaly detection results

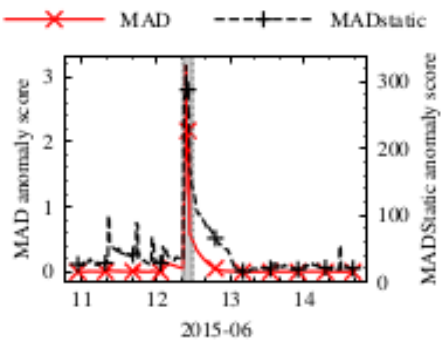
- Detect global anomalies with high ROC AUC / PR AUC



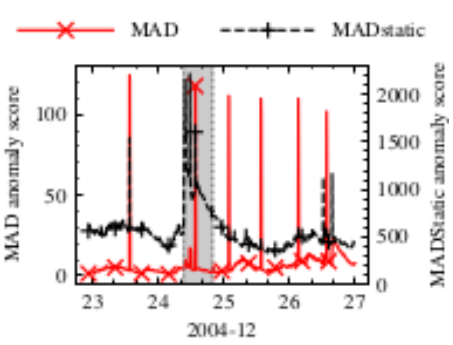
(a) aws



(b) indosat



(c) tm

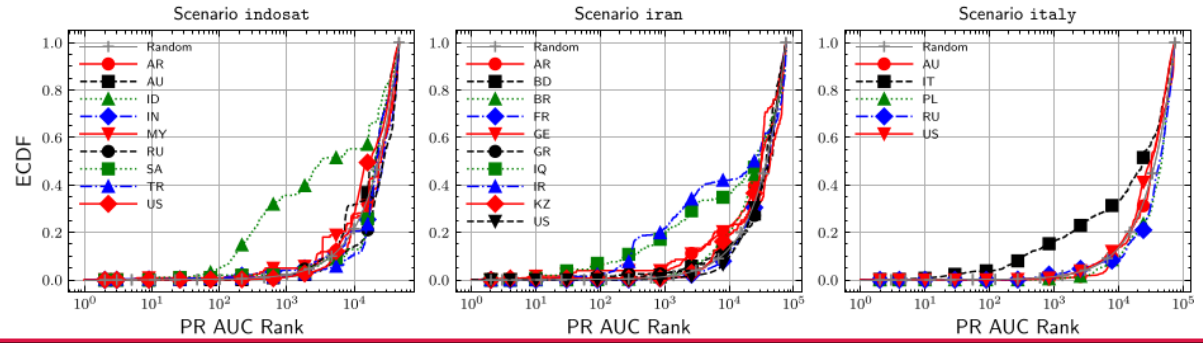


(d) ttnet

AS-level anomaly detection results

- We consistently achieve better than random and IODA baselines for potentially impacted ASes
- Grouping ASes per country, we observe geographical footprint of outages

Potentially impacted ASes			IODA		MADStar		Random	
ASN	Name	CC	PR AUC	ROC AUC	PR AUC	ROC AUC	PR AUC	ROC AUC
Scenario: `ttnet`								
9121	**TTNet**	**TR**	0.69	0.73	0.97	1.00	0.11	0.50
1239	Sprint	US	0.11	0.35	0.93	0.98	0.11	0.50
6762	SEABONE-NET	IT	0.11	0.51	0.78	0.80	0.11	0.50
3561	LVLT-3561	US	0.09	0.50	0.41	0.85	0.11	0.50
701	Verizon	US	0.09	0.26	0.33	0.85	0.11	0.50
174	COGENT	US	0.15	0.55	0.30	0.81	0.11	0.50
6453	TATA	US	0.19	0.78	0.29	0.65	0.11	0.50
6939	HURRICANE	US	0.09	0.50	0.25	0.82	0.11	0.50
3491	PCCW	US	0.09	0.50	0.13	0.61	0.11	0.50
Scenario: `aws`								
16509	**AMAZON-02**	**US**	0.02	0.13	0.96	0.99	0.02	0.50
200759	FLOW	CH	0.01	0.25	0.96	0.97	0.02	0.50
6939	HURRICANE	US	0.02	0.50	0.96	0.96	0.02	0.5
46841	FORKNETWORKING	US	0.02	0.50	0.81	0.98	0.02	0.
Scenario: `tm`								
4788	**TMNET-AS-AP**	**MY**	0.02	0.22	1.00	1.00	0.03	0
3356	LEVEL3	US	0.08	0.76	1.00	1.00	0.03	0
3549	LVLT-3549	US	0.03	0.52	0.98	1.00	0.03	0.
10753	LVLT-10753	US	0.03	0.46	0.77	0.98	0.03	0.50
Scenario: `indosat`								
4761	**INDOSAT-INP-AP**	**ID**	0.02	0.05	0.94	0.99	0.03	0.50
7018	ATT-INTERNET4	US	0.03	0.47	0.08	0.81	0.03	0.50
58682	LEVEL3-BD	BD	0.02	0.50	0.06	0.78	0.03	0.50
4637	ASN-TELSTRA-GLOBAL	HK	0.02	0.50	0.05	0.73	0.03	0.50
251	KAIAGLOBAL-AS	EU	0.02	0.49	0.02	0.32	0.03	0.50



Conclusion

- MADBGP, an end-to-end anomaly detection pipeline
- Hegemony graphs: unbiased Internet topologies
- QuickRIB: efficiently build high resolution datasets from public BGP data
- MAD: applied to real graphs for the first time and extended to weighted graphs
- Evaluated on weeks of data, billions of BGP elements
  - Identify global anomalies
  - Identify AS-level anomalies better than IODA BGP signal
  - Identify geographical footprint of outages
- Next steps
  - Investigate different graph decompositions (countries?)
  - Live deployment