33RD USENIX SECURITY SYMPOSIUM

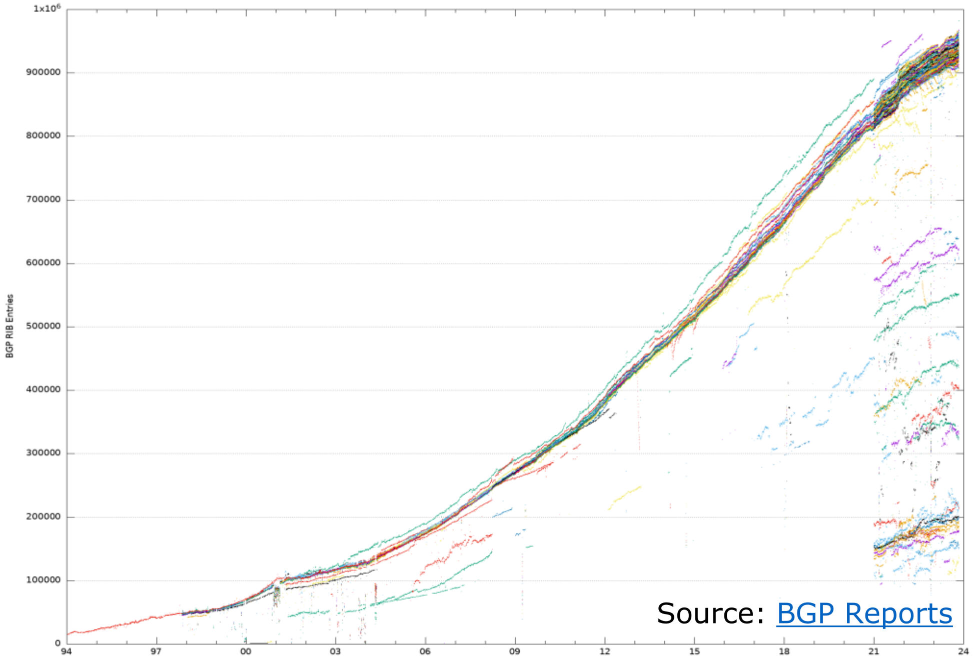
Learning with Semantics: Towards a Semantics-Aware Routing Anomaly Detection System

Yihao Chen, Qilei Yin, Qi Li, Zhuotao Liu, Ke Xu,

Yi Xu, Mingwei Xu, Ziqian Liu, Jianping Wu

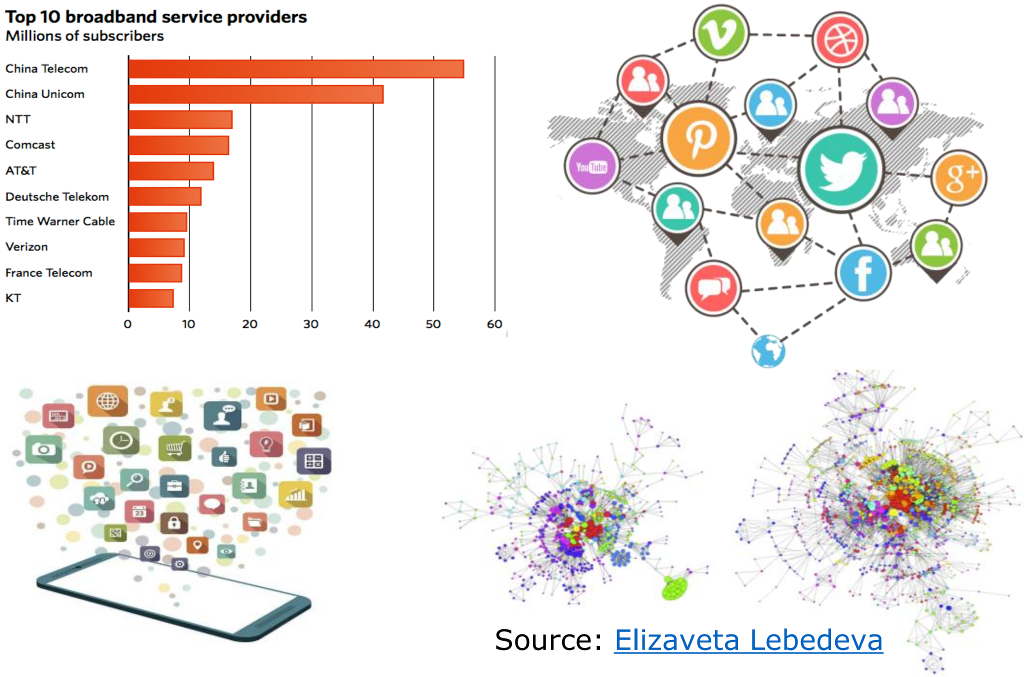
BGP in Today's Internet

- The foundational routing protocol maintaining **global Internet connectivity**.
- Over **960k interdomain routes** in operation, supporting more than **75k ASes**.



Source: BGP Reports

Rapid Growth of the BGP Table



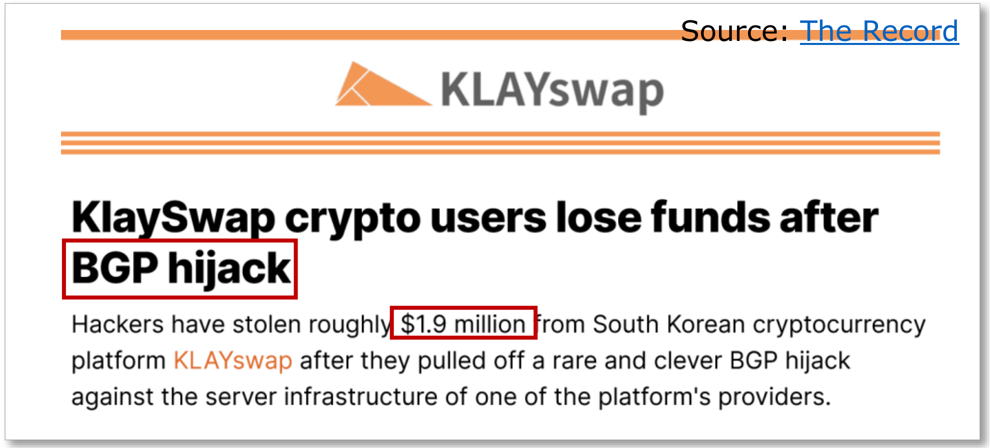
Source: Elizaveta Lebedeva

Vast Applications and Services over BGP

BGP (In)security

- BGP does not guarantee the **authenticity** and **integrity** of route announc
- RPKI and other security extensions, e.g., BGPsec, are **not widely deploye**

BGP Anomalies

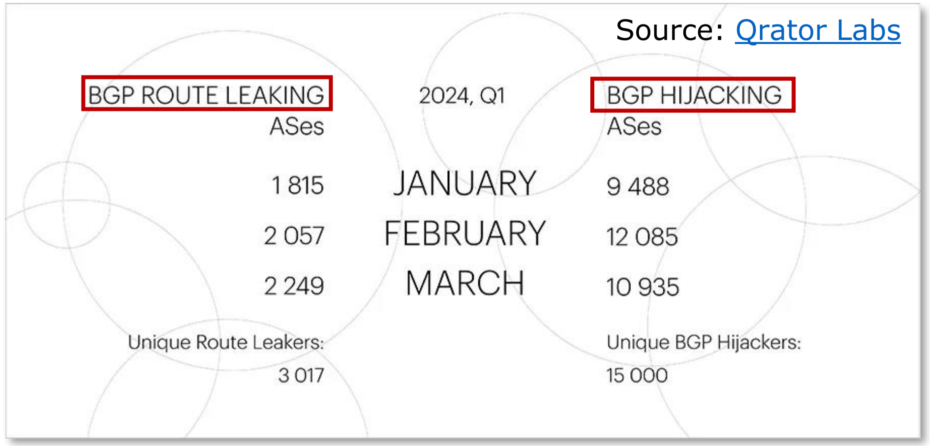

Source: The Record

KLAYswap

KlaySwap crypto users lose funds after BGP hijack

Hackers have stolen roughly $1.9 million from South Korean cryptocurrency platform KLAYswap after they pulled off a rare and clever BGP hijack against the server infrastructure of one of the platform's providers.

Loss of Millions in a Single BGP Anomaly

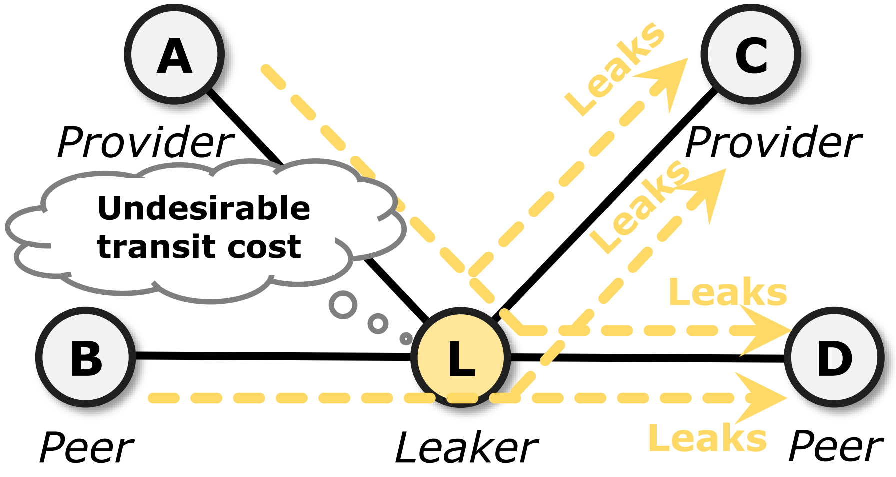


Thousands of BGP Incidents in Q1 2024

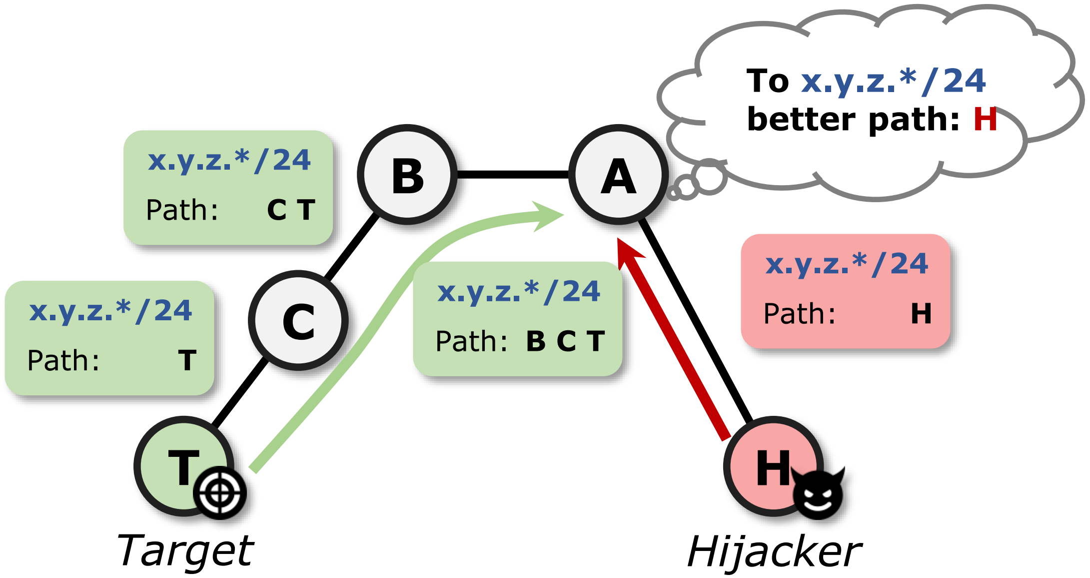
BGP routing anomalies are still a major threat to the current Internet!

BGP Routing Anomalies

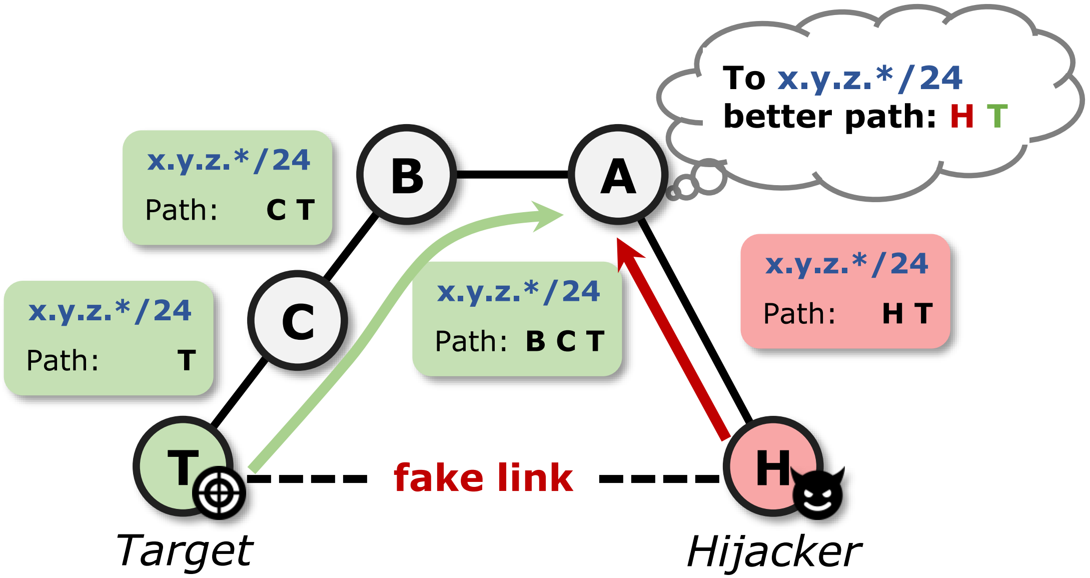
- Typical BGP anomalies
  - BGP hijacking (origin/path)
  - BGP route leak



BGP Route Leak Illustration



BGP Origin Hijacking Illustration



BGP Path Hijacking Illustration

BGP Anomaly Detection

- Traditional methods ARTEMIS [TON’18], HEAP [JSAC’16], Argus [IMC’12], BGPmon [CATCH’09], ……
  - ➢ Reliance on extensive **configuration** and/or **data probing**.
  - ➢ Require significant **manual investigation**.
  - ➢ Limited to **specific anomaly types**.
- ML-based methods AP2Vec [TNSM’22], Dong et al. [ICNP’21], Hoarau et al. [LCN’21], Testart et al. [IMC’19], ……
  - ➢ Significant training overhead: large-scale **data labeling** and **feature crafting**
  - ➢ No BGP semantics embedded -> **uninterpretable results**

Unsupervised and interpretable BGP anomaly detection, without requiring labeled training data and feature engineering?

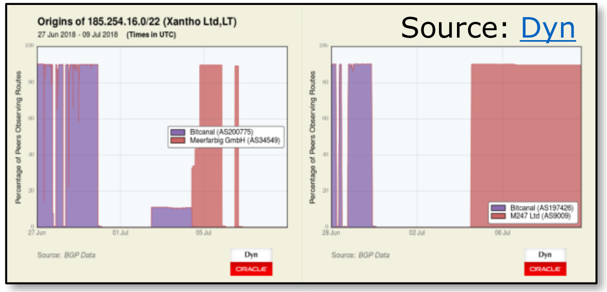
Our Intuition

- Indications of BGP anomalies?

⇒ **Drastic routing path changes.**

- Ways to profile the “drasticness” ……
  - Geolocation? **Inaccurate**
  - BGP statistics? **VP bias**

Geolocation



BGP statistics

⇒ **A quantifiable representation capable of characterizing an AS’s overall routing behaviors.**

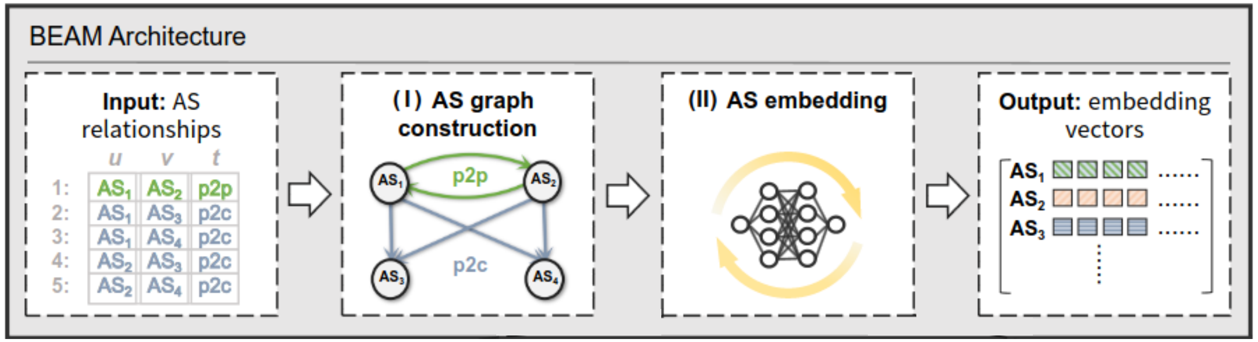
- What direct/indirect customers does it serve?
- What direct/indirect providers does it access?

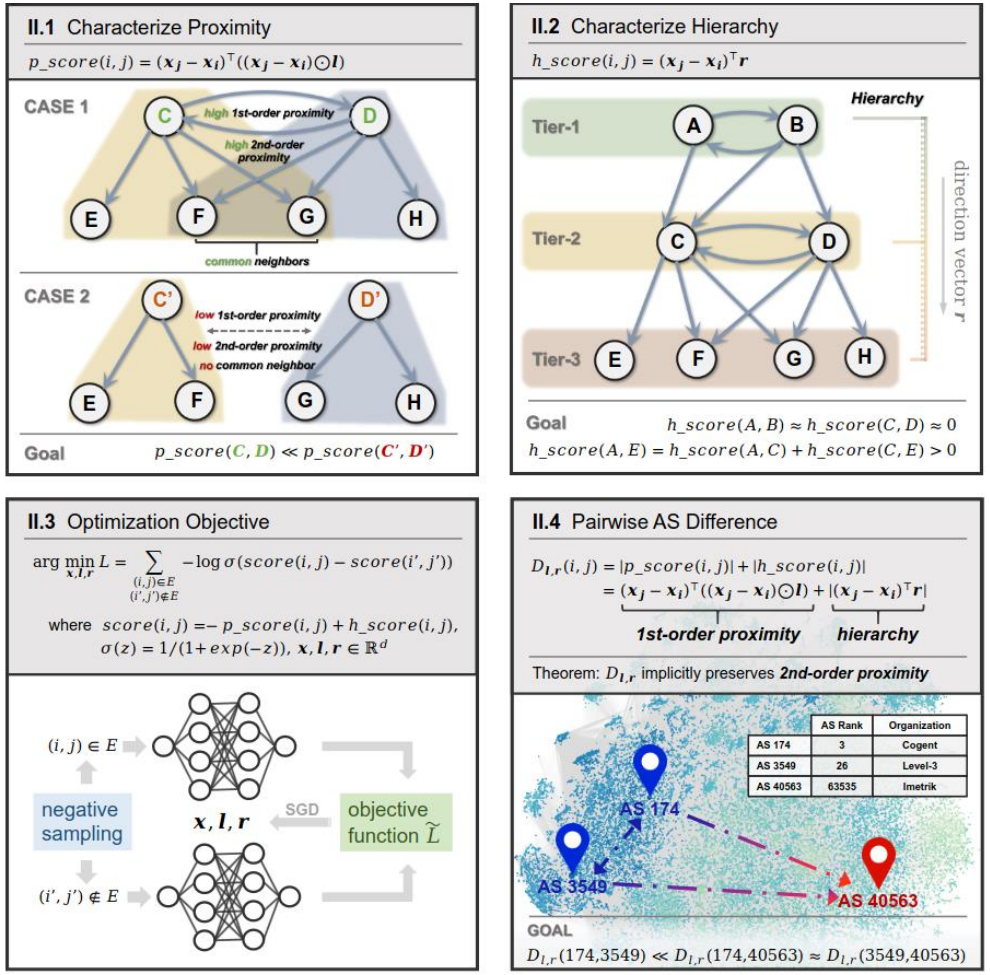
AS Routing Role: An AS’s overall routing behaviors, determined by its unique set of routing policies.

How to learn the routing role of each AS?

Introducing BEAM: An Embedding Model

- **B**GP s**E**m**A**ntics-aware network e**M**bedding
- Routing role: characterize overall relationships
- Two key properties
  - AS Proximity
  - AS Hierarchy
- A dedicated network representation learning model that fully integrates BGP semantics

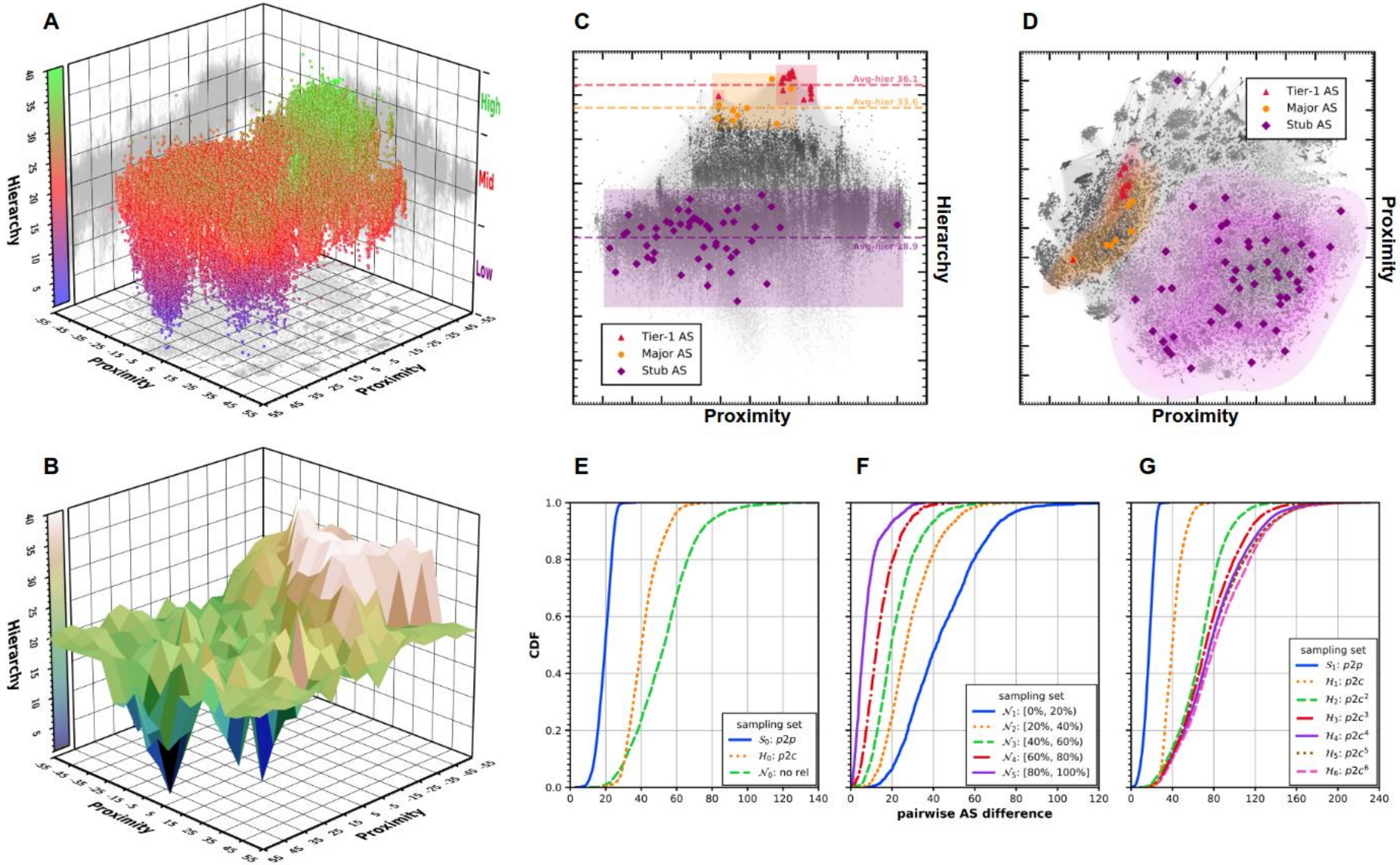




Routing Role Analysis

- 3D Visualization
  - Tier-1
  - Major but not Tier-1
  - Stub
- Pairwise AS Difference
  - 1st-order proximity
  - 2nd-order proximity
  - Hierarchy

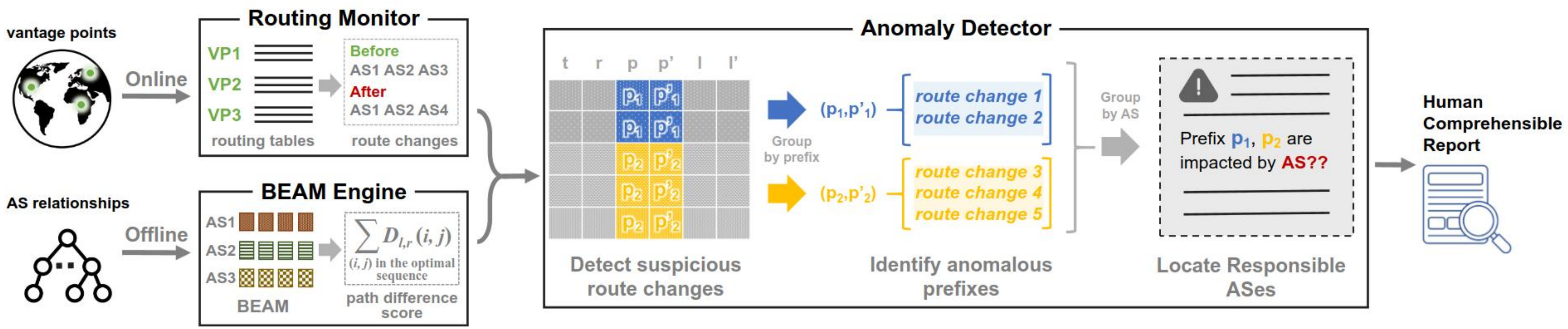
The intrinsic routing characteristics are reserved.



Detection System Based on BEAM

- **BEAM engine:** offline model training from monthly AS relationships
- **Routing monitor:** online route monitoring from multiple public vantage points
- **Anomaly Detector**
  - Detect suspicious route changes with high path difference
  - Identify anomalous prefixes by prefix-indexed grouping
  - Locate responsible ASes by frequent itemset mining

Interpretable Alarm Reports

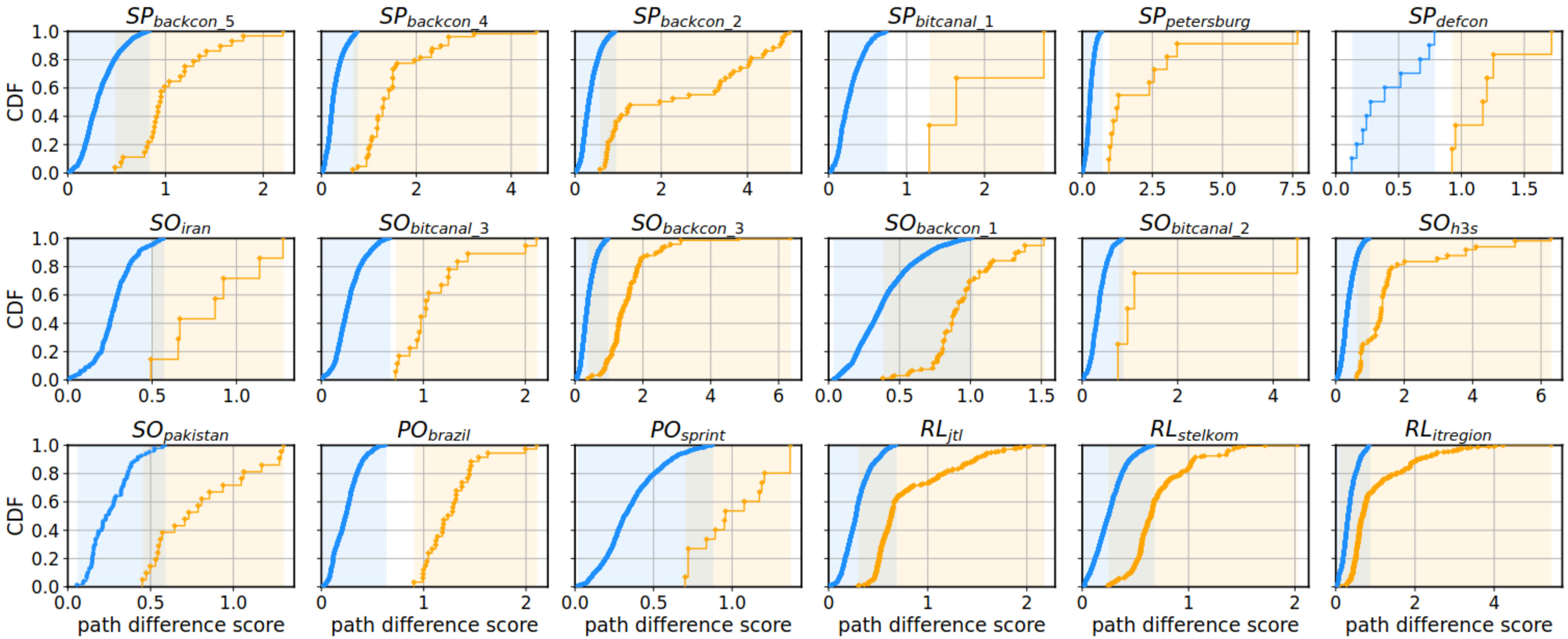


Experiment Setup

- **Ground truth:** 18 reports on historical incidents spanning from 2008 to 2021
  - 15 BGP hijacking (2 prefix and 13 subprefix hijacking)
  - 3 BGP route leak incidents
  - A total of 11,861,377,951 route announcements
- **Baseline**
  - 6 variants: Edit Distance, Jaccard Index, Line, Marine, node2vec, SDNE
  - 2 state-of-the-art ML-based system:
    - AV: using word embedding to represent ASN
    - LS: supervised LSTM based detector
- **Real-world deployment**
  - AS 4134 (rank top-100), China Telecom
  - One-month span of detection since Jan 1 2023

Evaluation: Path Difference Score

- 18 reports on historical routing anomalies spanning from 2008 to 2021



Anomalous route changes show significantly greater path difference.

Evaluation: Detection Performance

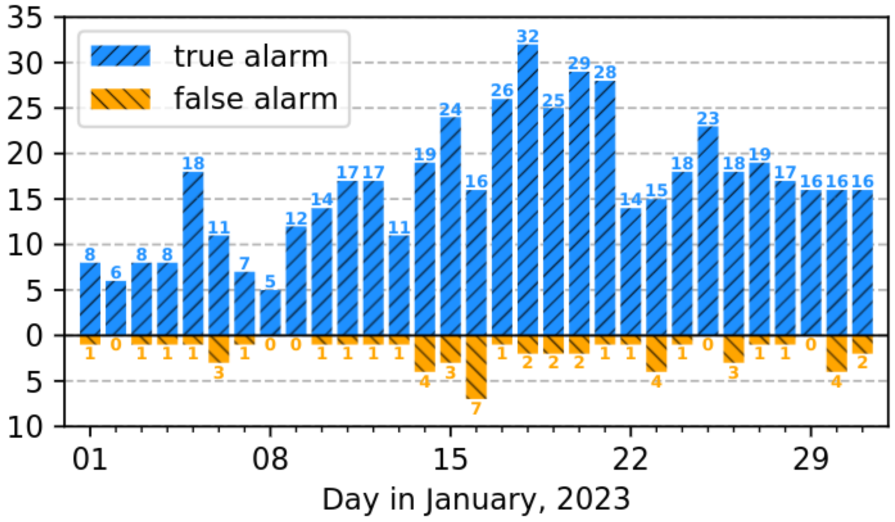
- All 18 anomalies are **correctly detected** by ours **within tens of alarms**.
- Ours reports **no false alarms for 6 datasets**, and only **5 false alarms in the worst case**.
- The baselines cannot detect all anomalies and raise more false alarms than ours.
- The baselines require many extra data to train the models.

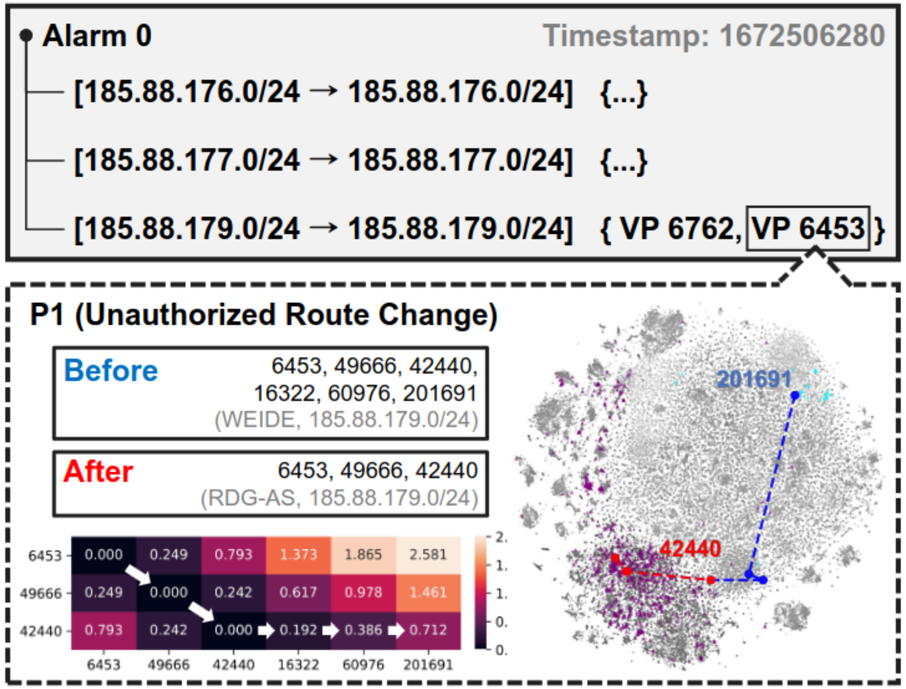
Dataset	Detected									#Alarms(#FalseAlarms)								
	ED	JI	Li	Ma	NV	SD	LS	AV	Ours	ED	JI	Li	Ma	NV	SD	LS	AV	Ours
$SP_{backcon\_5}$	✓	✓	✓	✓	✓	✗	✗	✗	✓	**18(5)**	**12(3)**	**21(4)**	**15(3)**	**17(2)**	9(1)	62(31)	5(1)	**34(2)**
$SP_{backcon\_4}$	✓	✓	✓	✓	✓	✗	✓	✓	✓	**14(1)**	**8(1)**	**15(2)**	**13(1)**	**12(1)**	7(1)	**42(17)**	**22(6)**	**21(0)**
$SP_{backcon\_2}$	✓	✓	✓	✓	✓	✗	✓	✗	✓	**29(7)**	**21(7)**	**24(5)**	**25(7)**	**21(4)**	8(3)	**38(13)**	23(18)	**37(1)**
$SP_{bitcanal\_1}$	✗	✗	✗	✓	✓	✗	✗	✗	✓	16(0)	16(0)	14(0)	**18(0)**	**17(0)**	7(0)	67(36)	30(10)	**16(0)**
$SP_{petersburg}$	✓	✓	✓	✓	✓	✗	✓	✓	✓	**22(3)**	**14(0)**	**21(3)**	**20(1)**	**16(0)**	12(2)	**66(28)**	**37(16)**	**24(0)**
SP_{defcon}	✓	✓	✓	✓	✓	✗	✓	✓	✓	**7(2)**	**7(2)**	**9(3)**	**9(3)**	**9(3)**	2(2)	**28(10)**	**17(9)**	**7(1)**
SO_{iran}	✓	✓	✓	✓	✓	✗	✗	✗	✓	**15(1)**	**8(1)**	**24(5)**	**12(2)**	**16(3)**	0(0)	21(11)	19(10)	**31(2)**
$SO_{bitcanal\_3}$	✗	✗	✗	✗	✓	✗	✗	✗	✓	26(4)	24(3)	29(6)	25(3)	**26(5)**	7(0)	44(19)	17(8)	**40(1)**
$SO_{backcon\_3}$	✓	✓	✓	✓	✓	✗	✗	✗	✓	**32(8)**	**23(4)**	**27(6)**	**34(8)**	**34(9)**	6(1)	49(27)	19(9)	**35(5)**
$SO_{backcon\_1}$	✓	✗	✓	✓	✓	✗	✗	✗	✓	**19(6)**	16(4)	**35(14)**	**18(4)**	**17(7)**	0(0)	63(35)	25(11)	**18(3)**
$SO_{bitcanal\_2}$	✗	✓	✓	✓	✓	✗	✗	✗	✓	16(1)	**15(1)**	**17(2)**	**15(1)**	**16(1)**	12(2)	39(14)	29(8)	**24(0)**
SO_{h3s}	✓	✗	✓	✓	✓	✗	✓	✗	✓	**11(1)**	3(0)	**15(3)**	**12(2)**	**9(0)**	5(1)	**38(22)**	27(8)	**14(0)**
$SO_{pakistan}$	✓	✓	✓	✓	✓	✗	✗	✗	✓	**12(4)**	**8(2)**	**9(2)**	**10(4)**	**8(1)**	1(0)	26(14)	2(0)	**10(1)**
PO_{brazil}	✓	✓	✓	✓	✓	✗	✗	✓	✓	**30(5)**	**32(4)**	**30(5)**	**37(5)**	**25(3)**	11(2)	52(25)	**28(11)**	**51(1)**
PO_{sprint}	✗	✗	✗	✗	✗	✗	✓	✗	✓	20(0)	18(0)	16(0)	22(2)	19(2)	10(2)	**84(24)**	33(8)	**29(0)**
RL_{jtl}	✗	✗	✗	✗	✗	✗	✓	✗	✓	17(1)	16(1)	21(2)	16(2)	17(1)	6(3)	**60(40)**	21(11)	**46(5)**
$RL_{stelkom}$	✗	✓	✗	✗	✗	✗	✓	✗	✓	25(4)	**21(2)**	34(6)	26(3)	23(3)	4(0)	**284(225)**	17(8)	**43(3)**
$RL_{itregion}$	✗	✗	✗	✗	✗	✗	✓	✗	✓	25(0)	21(1)	23(0)	20(0)	21(3)	6(2)	**74(44)**	23(9)	**44(4)**
Overall	11/18	11/18	12/18	13/18	14/18	0/18	9/18	4/18	18/18	354(53)	283(36)	384(68)	347(51)	323(48)	113(22)	1137(635)	394(161)	524(29)

Real-World Deployment

- Deployed in AS 4134 of China Telecom, detecting from January 1 to February 1 2023
- 152,493,303 live route announcements, 5,106,442 route changes, and **548 raised alarms**.
- On average, 17.68 alarms per day, including **1.65 false alarms**.

#Affected Routes	#Affected Prefixes	#Affected Origins
1,202	961	477



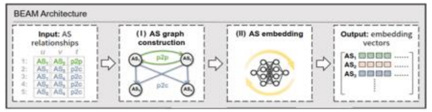


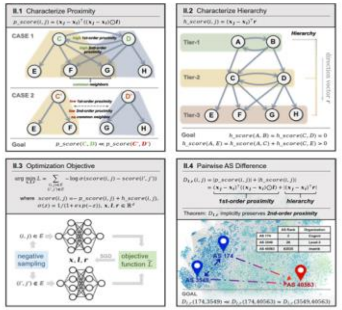
Summary

Intuition: represent AS routing roles

Introducing BEAM: A Novel Approach

- **B**GP s**E**m**A**ntics-aware network e**M**bedding
- Routing role: characterize overall relationships
- Two key properties
  - AS Proximity
  - AS Hierarchy
- The first dedicated network representation learning model that fully integrates BGP semantics

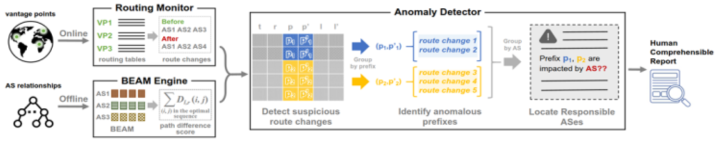




Workflow: learning, detection, interpretation

Anomaly Detection System

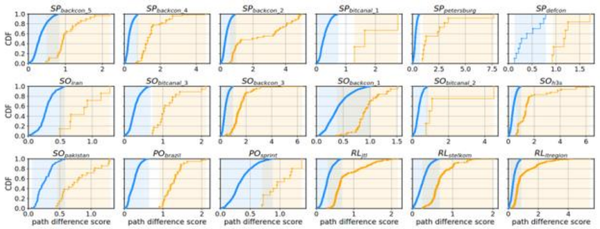
- BEAM engine: offline BEAM model training from montly AS relationships
- Routing monitor: online route announcement fetching from multiple public vantage points
- Anomaly Detector:
  - Detecting suspicious route changes with high path difference
  - Identifying anomalous prefixes by prefix-indexed grouping
  - Locating responsible ASes by frequent itemset mining



Evaluation: over 11 billion announcements

Evaluation: Path Difference Score

- 18 reports on historical routing anomalies spanning from 2008 to 2021
  - 15 BGP hijacking (2 prefix and 13 subprefix hijacking)
  - 3 BGP route leak incidents
  - A total of 11,861,377,951 route announcements



Deployment: large ISP real-world deployment

Real-World Deployment

- Deployed in a large ISP (top 100), detecting from January 1 to February 1 2023
- In total, the system processes 152,493,303 live route announcements, detects 5,106,442 route changes and raises 548 alarms.
- On average, 17.68 alarms per day, including 1.65 false alarms.

Table 2: The overall impact of the detected anomalies.

#Affected Routes	#Affected Prefixes	#Affected Origins
1,202	961	477

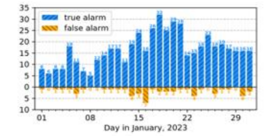


Figure 8: Daily alarm number in real-world deployment.

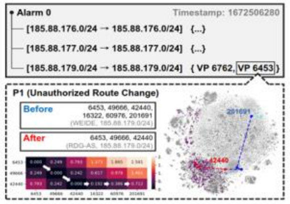


Figure 10: An anomaly report in real-world deployment.

Q&A

Thank you!

Key Takeaways

- AS routing roles capture overall routing behaviors of ASes.
- Routing anomaly detection can be achieved by detecting routing role churns.
- Our system can detect previous confirmed anomalies with minor false alarms.
- Routing-role change analysis enables visually interpretable anomaly alarms.

Contact: yh-chen21@mails.tsinghua.edu.cn

Code: https://github.com/yhchen-tsinghua/routing-anomaly-detection

This research is sponsored in part by Deng Feng Fund