IRRedicator: Pruning IRR with RPKI-Valid BGP Insights

Minhyeok Kang[†], Weitong Li[§], Roland van Rijswijk-Deij[§‡], Taekyoung "Ted" Kwon[†], Taejoong Chung[§]

[†]Seoul National University, [‡]University of Twente, [§]Virginia Tech

Border Gateway Protocol (BGP)

- BGP is one of the most crucial components for sustaining global network connectivity
  - However, BGP was not designed with security in mind (e.g., no route origin authentication)

THE POWER OF FALSE ADVERTISING —

How an Indonesian ISP took down the mighty Google for 30 minutes

Internet's web of trust let a company you never heard of block your Gmail.

SEAN GALLAGHER - 11/6/2012, 11:07 AM

Google's services went offline for many users for nearly a half-hour on the evening of November 5, thanks to an erroneous routing message broadcast by Moratel, an Indonesian telecommunications company. The outage might have lasted even longer if it hadn't been spotted by a network engineer at CloudFlare who had a friend in a position to fix the problem.

Border Gateway Protocol (BGP)

- BGP is one of the most crucial components for sustaining global network connectivity
  - However, BGP was not designed with security in mind (e.g., no route origin authentication)

THE POWER OF FALSE ADVERTISING —

How an Indonesian
mighty Google for

Internet's web of trust let a company you

SEAN GALLAGHER - 11/6/2012, 11:07 AM

Google's services went offline for
5, thanks to an erroneous routing
telecommunications company. The
spotted by a network engineer at

Catalin Cimpanu | February 14, 2022

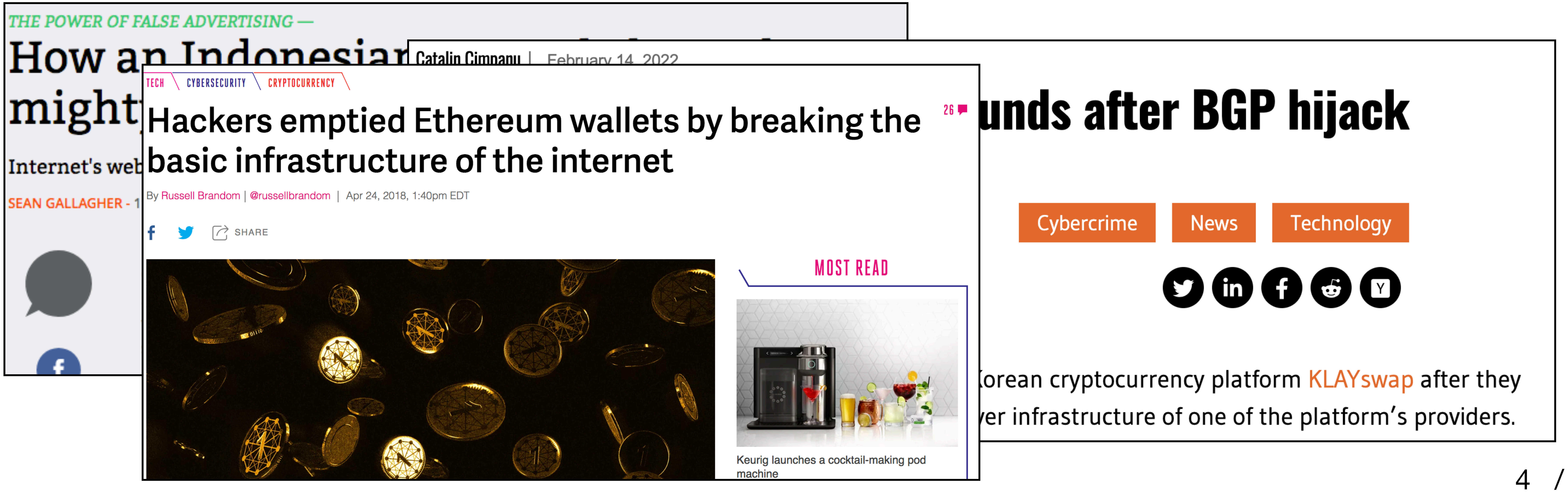
KlaySwap crypto users lose funds after BGP hijack

Cybercrime News Technology

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.

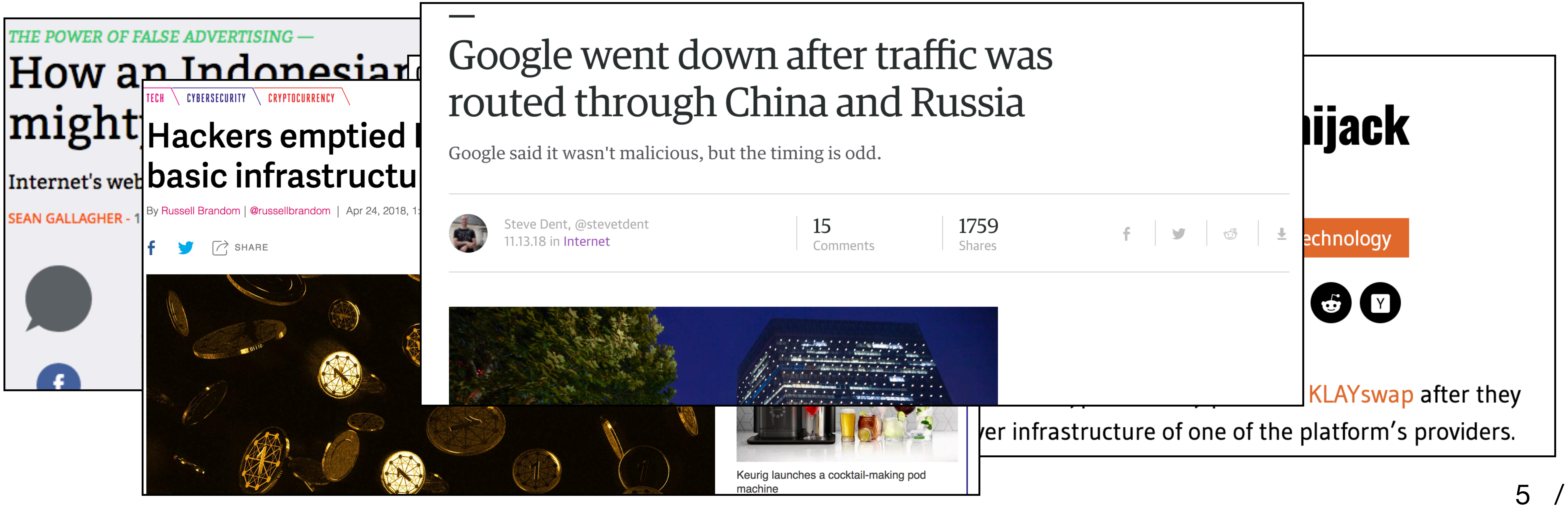
Border Gateway Protocol (BGP)

- BGP is one of the most crucial components for sustaining global network connectivity
  - However, BGP was not designed with security in mind (e.g., no route origin authentication)



Border Gateway Protocol (BGP)

- BGP is one of the most crucial components for sustaining global network connectivity
  - However, BGP was not designed with security in mind (e.g., no route origin authentication)



Efforts to improve BGP security

- **Internet Routing Registry (IRR) (1995)**
  - widely used for sharing global routing information (> 68% of ASes)
  - lacks an authentication mechanism & has many outdated entries

- **Resource Public Key Infrastructure (RPKI) (2008)**
  - provides a cryptographically verifiable method of binding IP prefixes to their respective origin ASes
  - narrower coverage than IRR
    - has certificate dependencies in the hierarchy of RPKI
    - configuration issues in Route Origin Authorization (ROA) objects

Efforts to improve BGP security

- **Internet Routing Registry (IRR) (1995)**
  - widely used for sharing global routing information (> 68% of ASes)
  - lacks an authentication mechanism & has many outdated entries

Take the strengths of both IRR and RPKI
in order to improve the BGP security

- r
  respective origin ASes
  - narrower coverage than IRR
    - has certificate dependencies in the hierarchy of RPKI
    - configuration issues in Route Origin Authorization (ROA) objects

Datasets

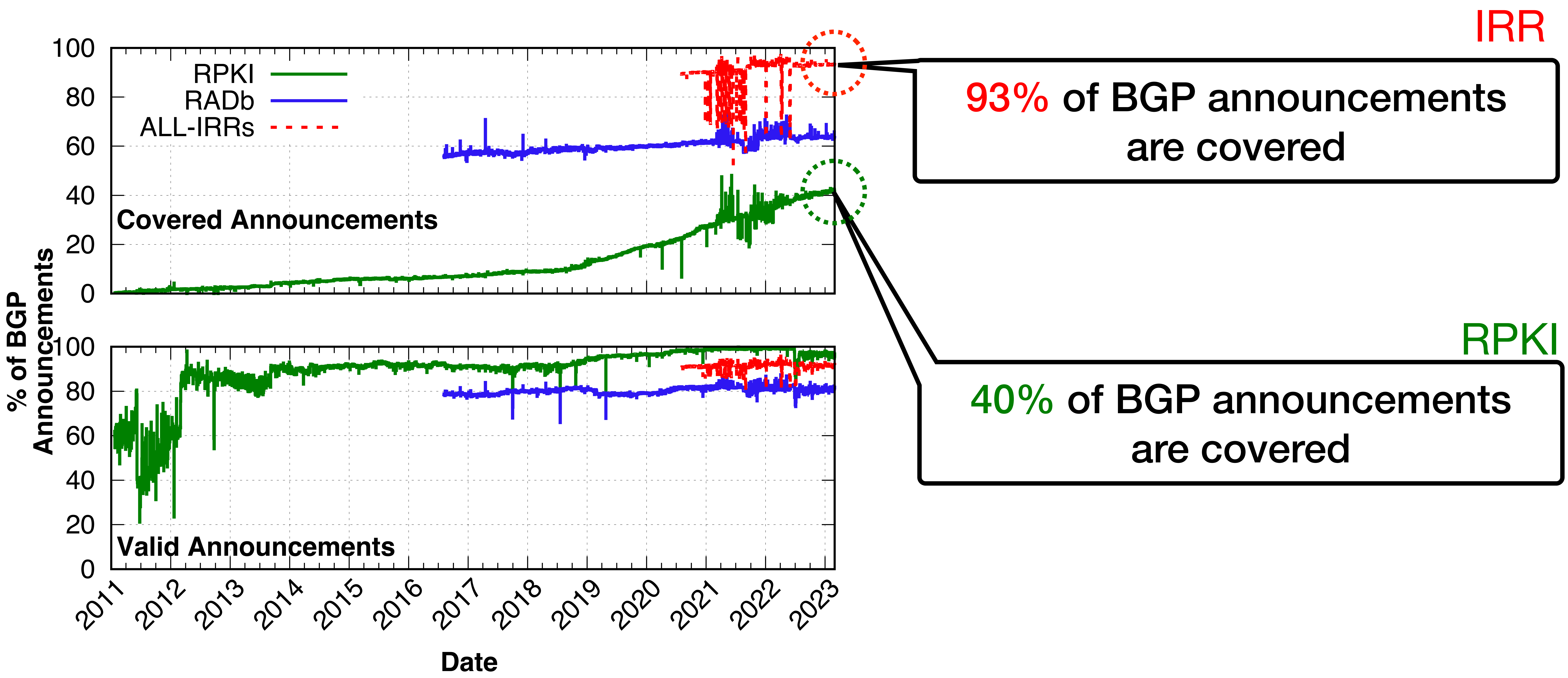
RPKI

IRR

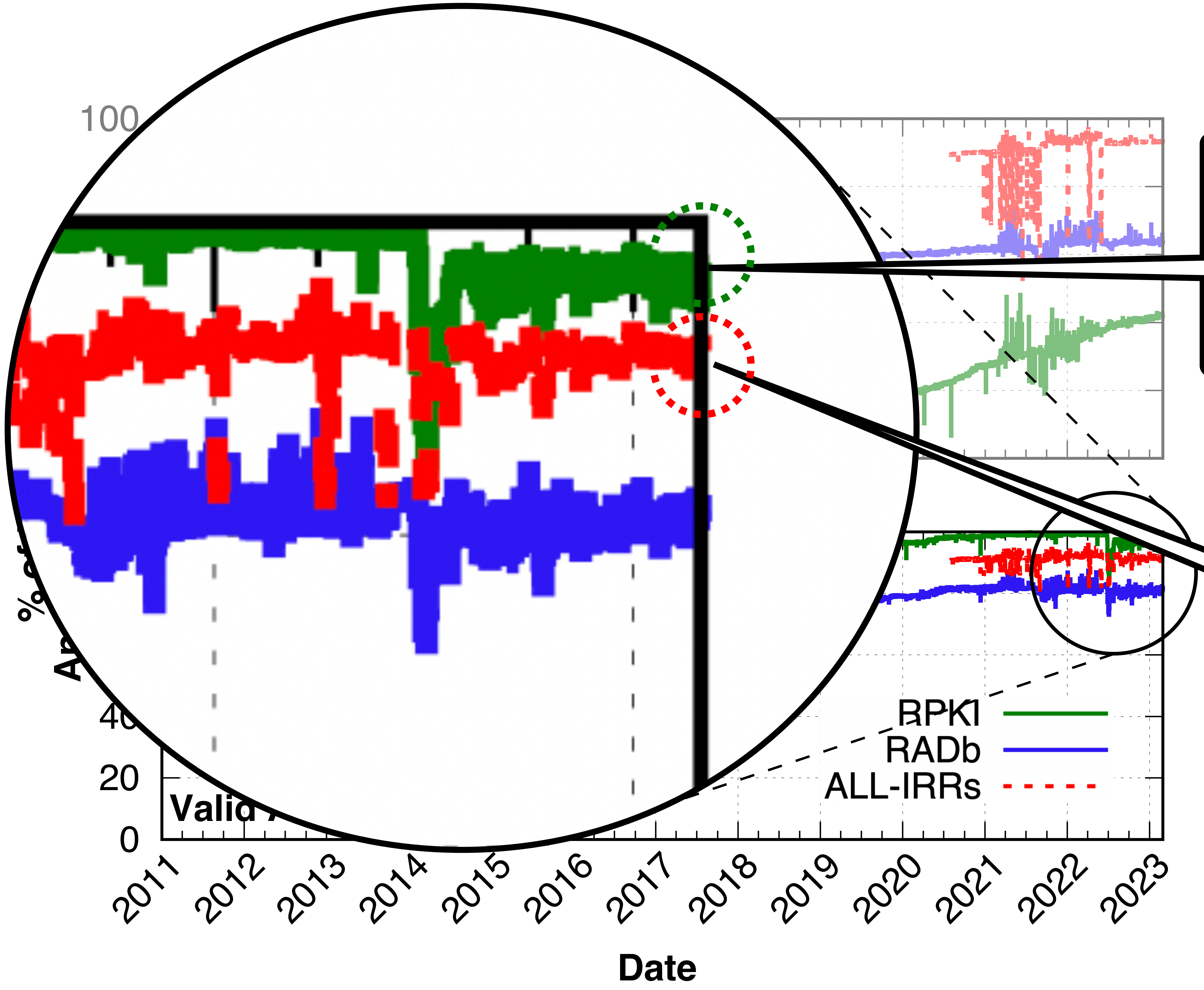
Auth. Objects	Measurement Period	# of Objects
ROA	2011/01 – 2023/03	333 K
RADb	2016/08 – 2023/03	1.43 M
ALL-IRRs	2019/12 – 2023/03	2.69 M

RADb + IRRs of Regional Internet Registries

The deployment status of IRR and RPKI



The deployment status of IRR and RPKI



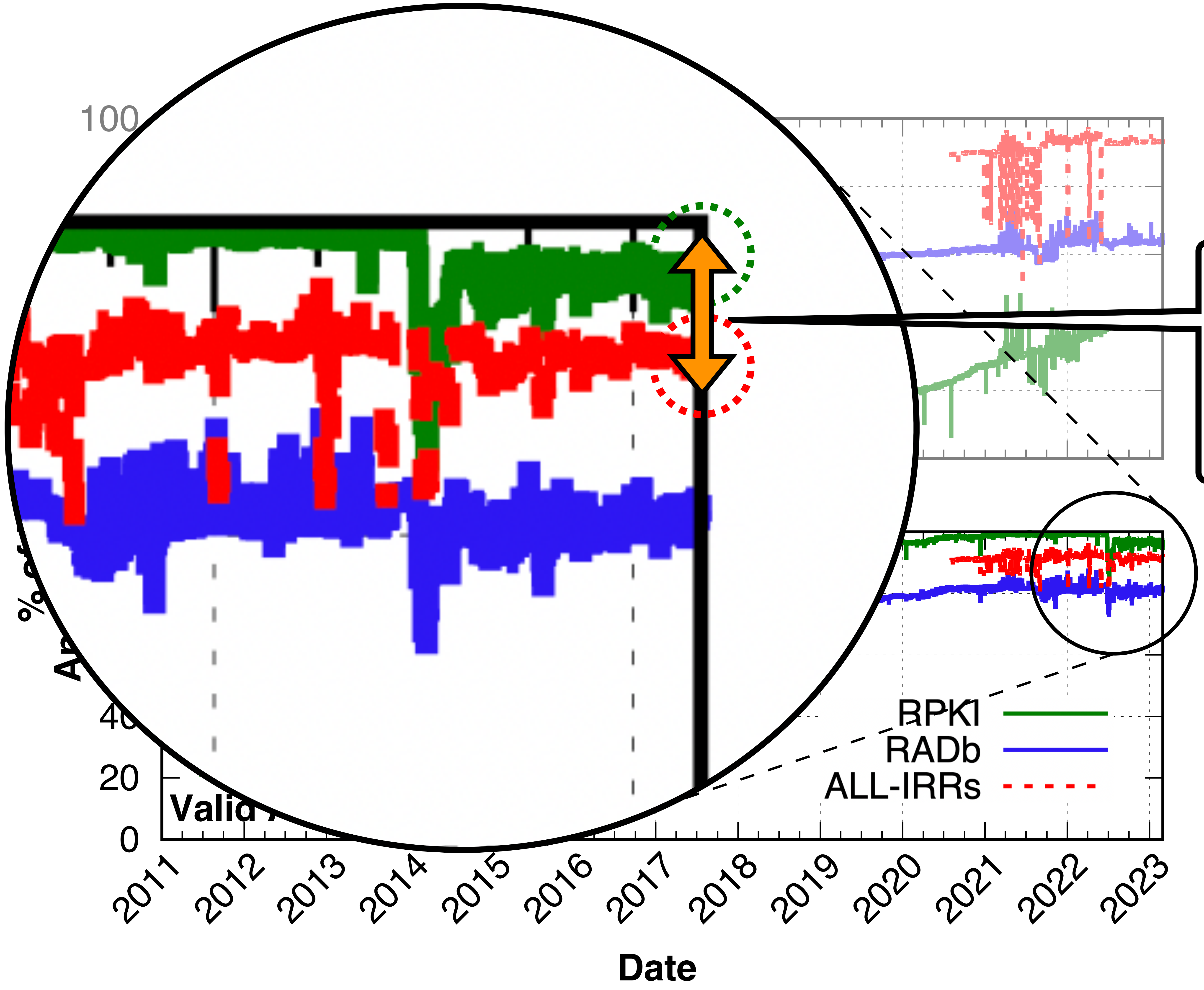
RPKI

98% of covered BGP announcements are valid

IRR

90% of covered BGP announcements are valid

The deployment status of IRR and RPKI

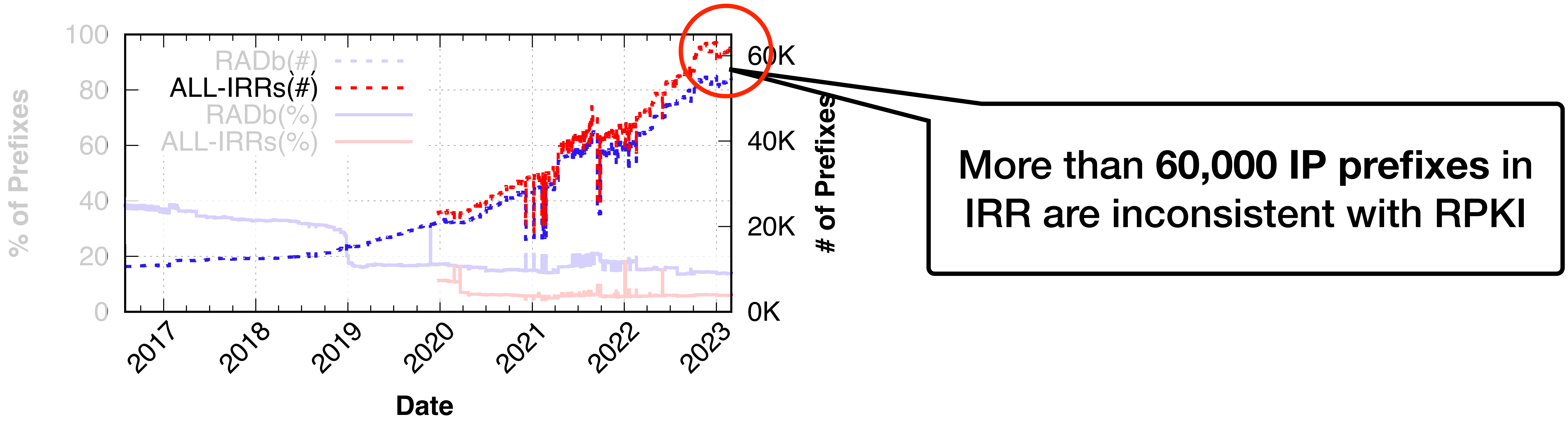


a gap between the percentages of valid BGP announcements

Inconsistency?

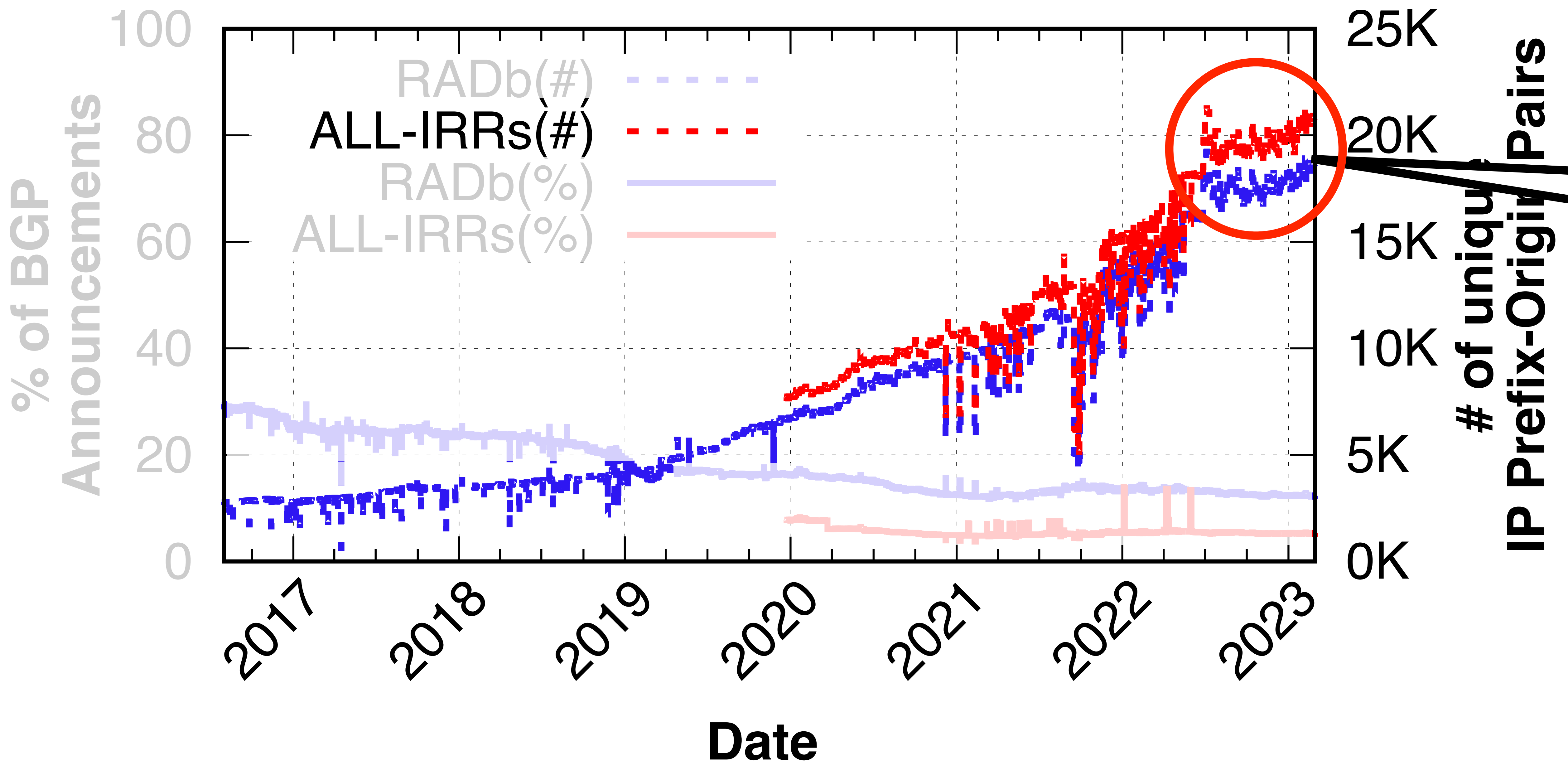
Are they consistent with “each other”?

- For IP prefixes registered in both IRR and RPKI, we examine whether they have the same origin AS as the one registered in RPKI



Do inconsistent IP prefixes appear in BGP announcements?

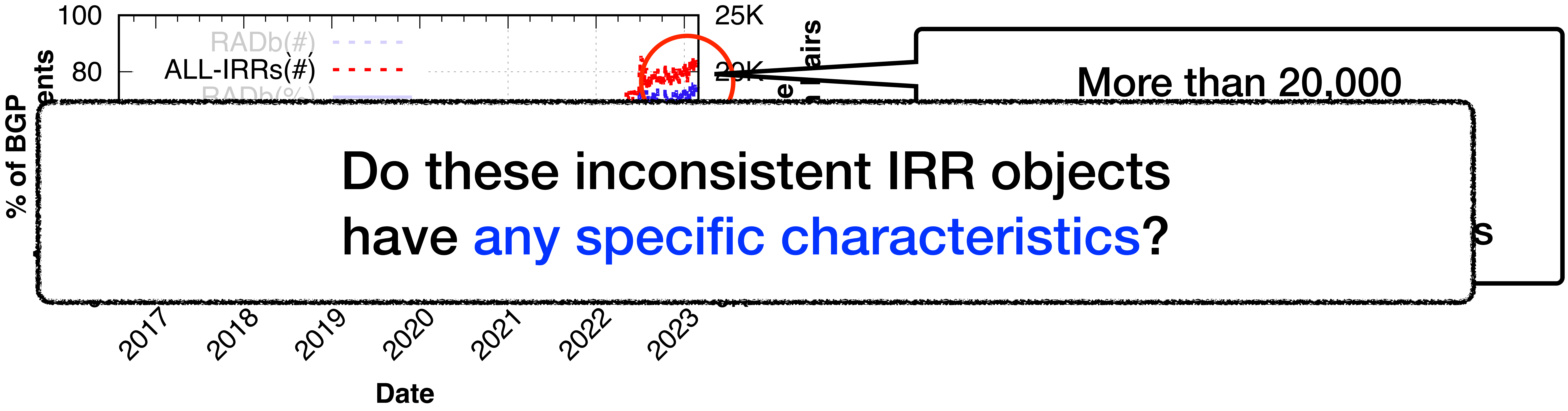
- For BGP announcements verifiable through both RPKI and IRR, we track their frequency over time



More than **20,000**
BGP announcements
are covered by
the inconsistent IP prefixes

Do inconsistent IP prefixes appear in BGP announcements?

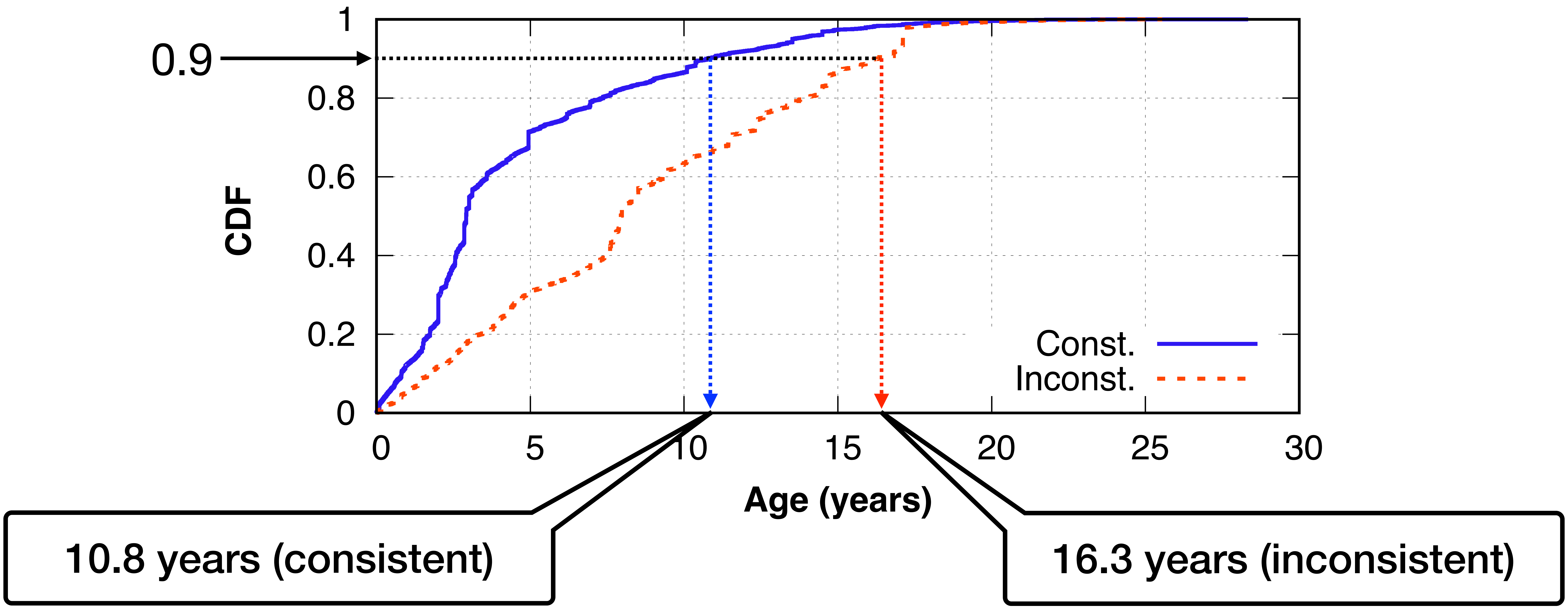
- For BGP announcements verifiable through both RPKI and IRR, we track their frequency over time



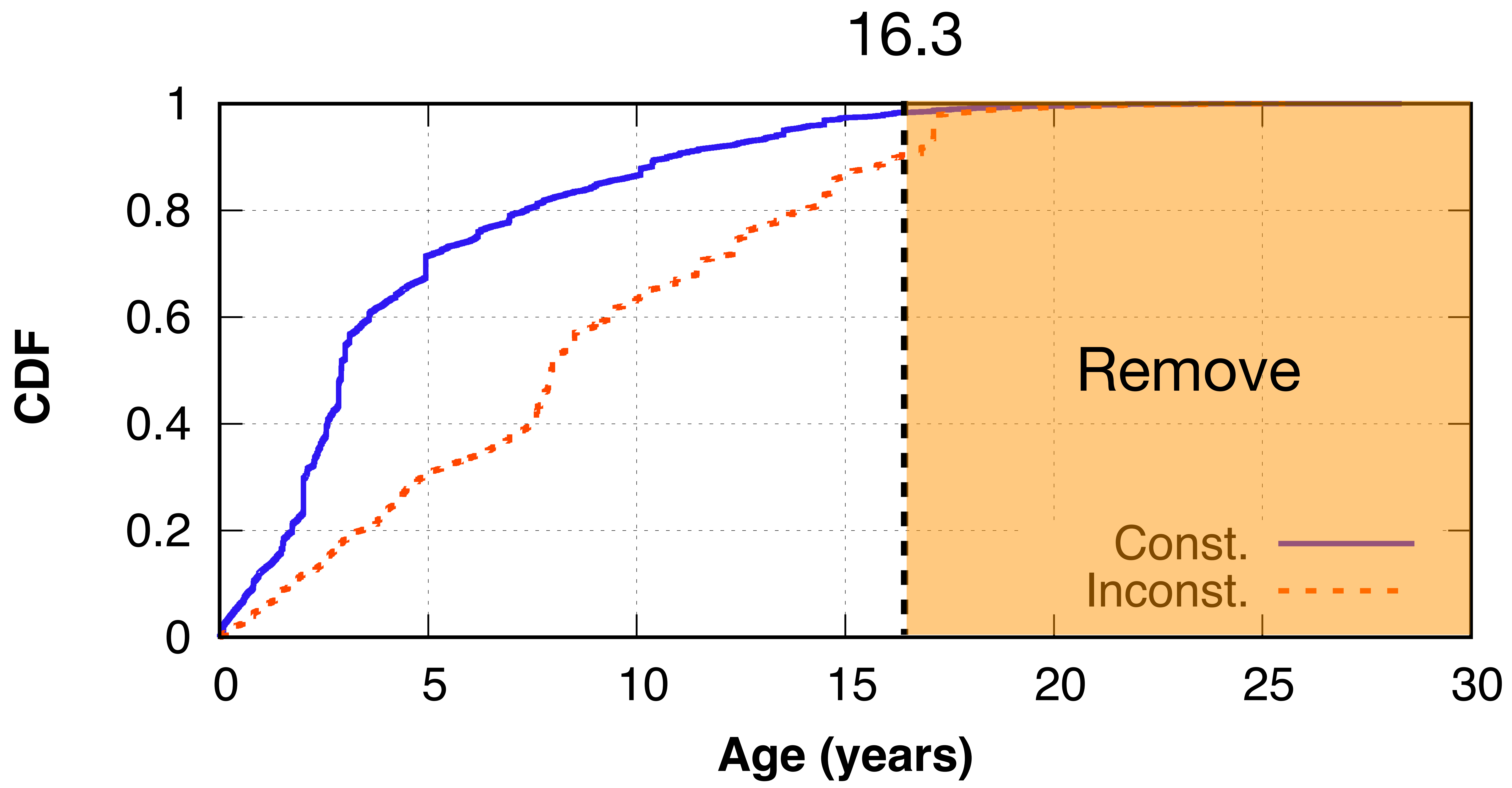
Do these inconsistent IRR objects have any specific characteristics?

Age of IRR objects

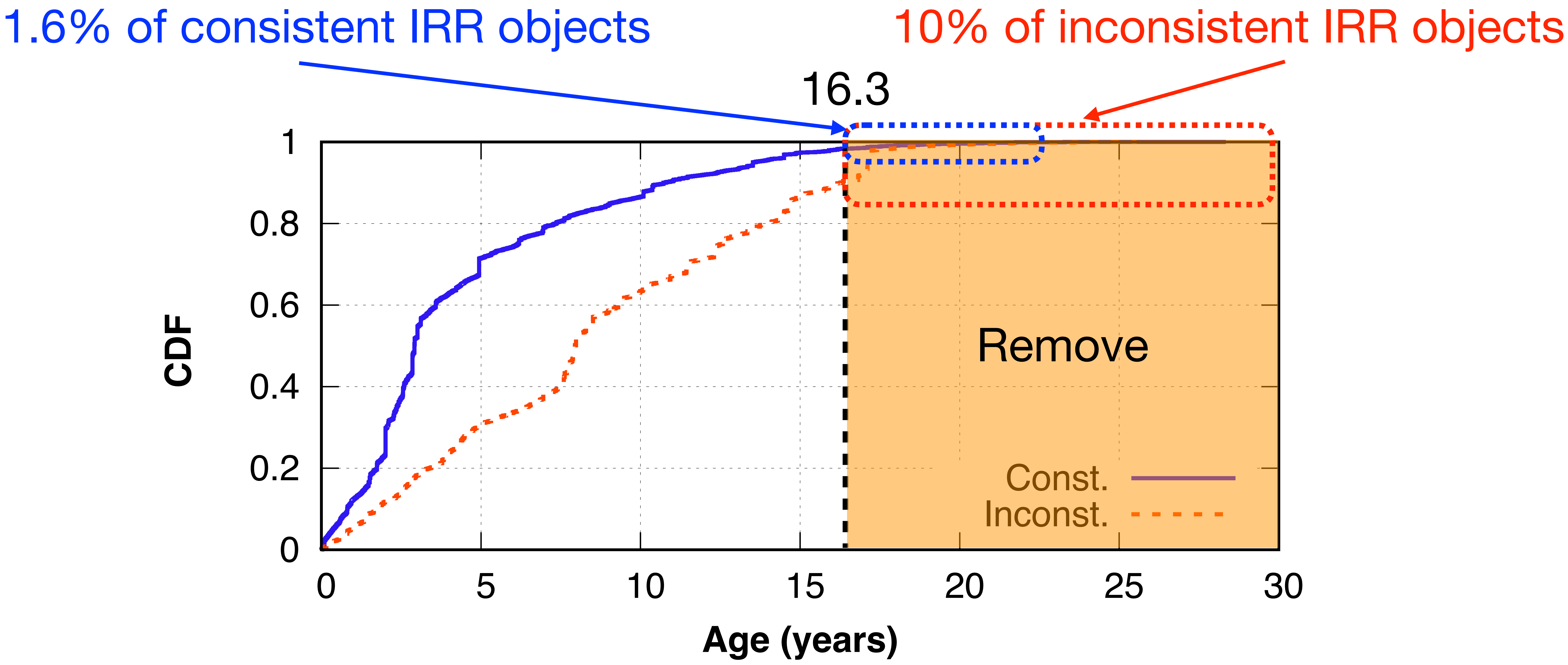
Age = latest date of our dataset - last modified date



Example: filtering with age



Example: filtering with age

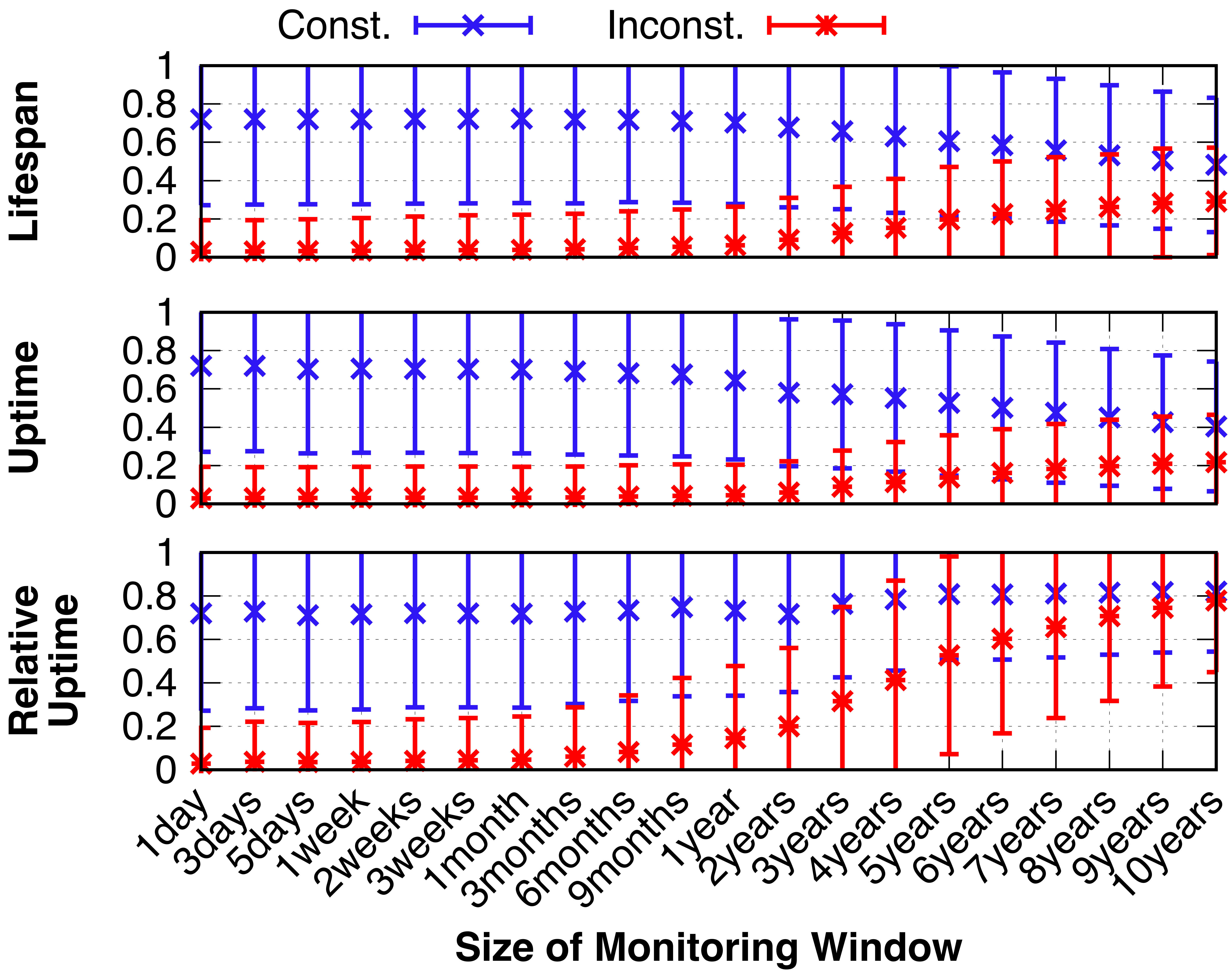


How to deal with inconsistent IRR objects?

- Filtering IRR objects with their ages
  - setting a “good” threshold is challenging
    - conservatively → low coverage, aggressively → high mis-classification
- Utilizing RPKI to filter out inconsistent IRR objects
  - RPKI only covers 44% of IRR objects
- Leveraging patterns of BGP announcements datasets to identify inconsistent IRR objects
  - can be applied to all IRR objects!

BGP announcement pattern

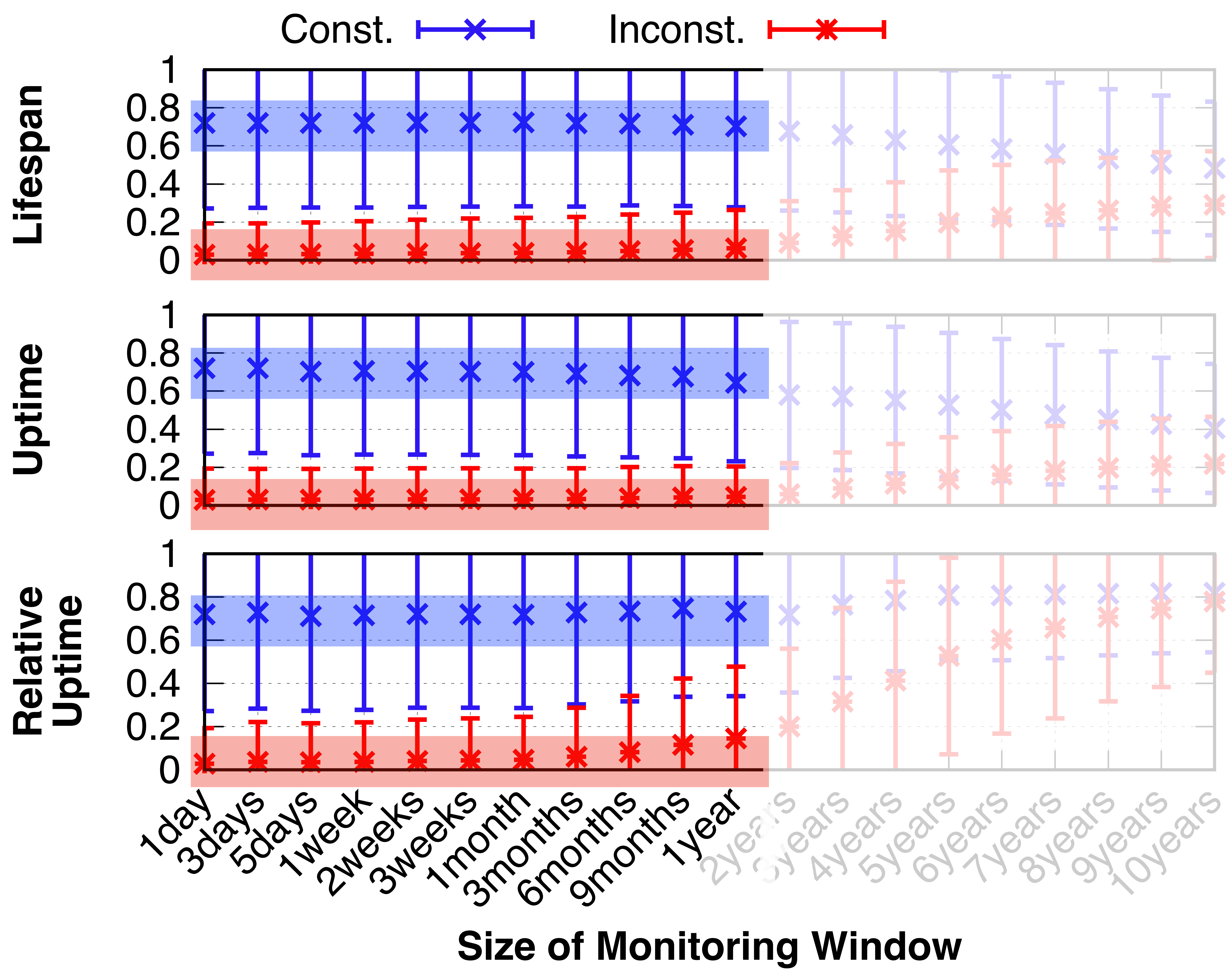
Monitoring window
a time period from the start time t to the latest date of our dataset



- **Lifespan**
  the difference in dates between the first and last observations, divided by a monitoring window size
- **Uptime**
  the number of days that BGP announcements have been observed, divided by a monitoring window size
- **Relative uptime**
  uptime/lifespan

BGP announcement pattern

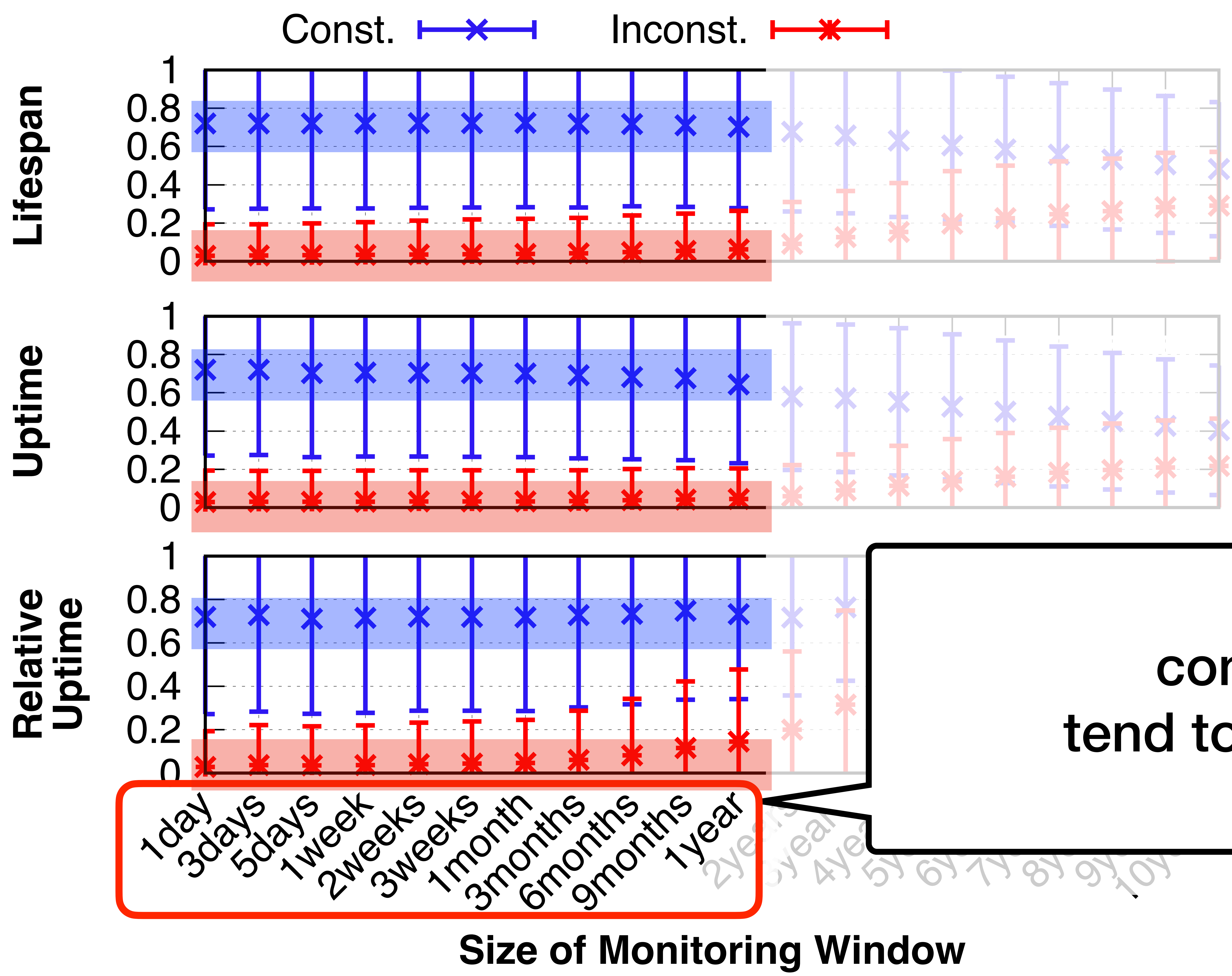
Monitoring window
a time period from the start time t to the latest date of our dataset



- **Lifespan**
  the difference in dates between the first and last observations, divided by a monitoring window size
- **Uptime**
  the number of days that BGP announcements have been observed, divided by a monitoring window size
- **Relative uptime**
  uptime/lifespan

BGP announcement pattern

Monitoring window
a time period from the start time t to the latest date of our dataset

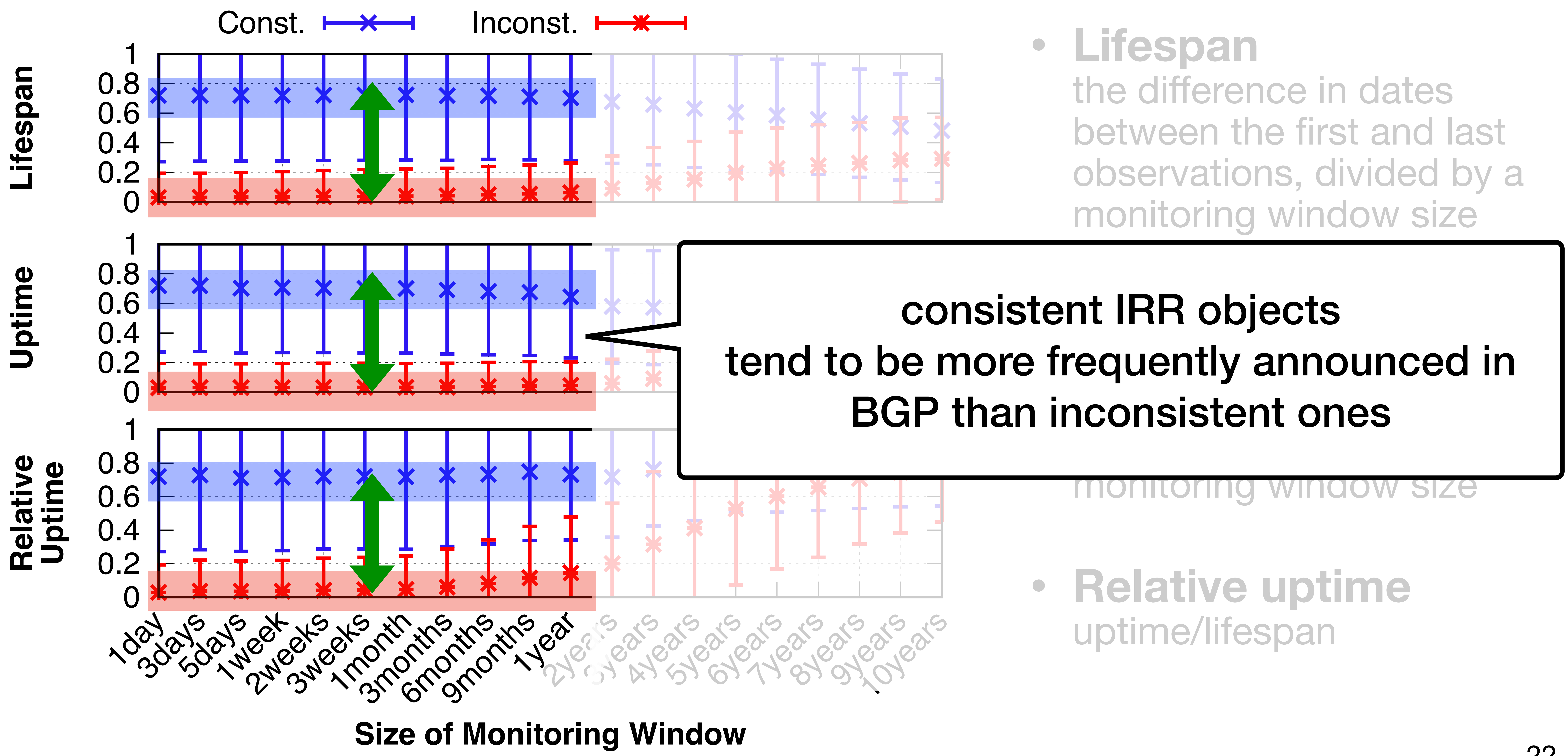


- **Lifespan**
  the difference in dates between the first and last observations, divided by a monitoring window size
- **Uptime**
  the number of days that BGP announcements have

consistent IRR objects tend to be more recently used

BGP announcement pattern

Monitoring window
a time period from the start time t to the latest date of our dataset

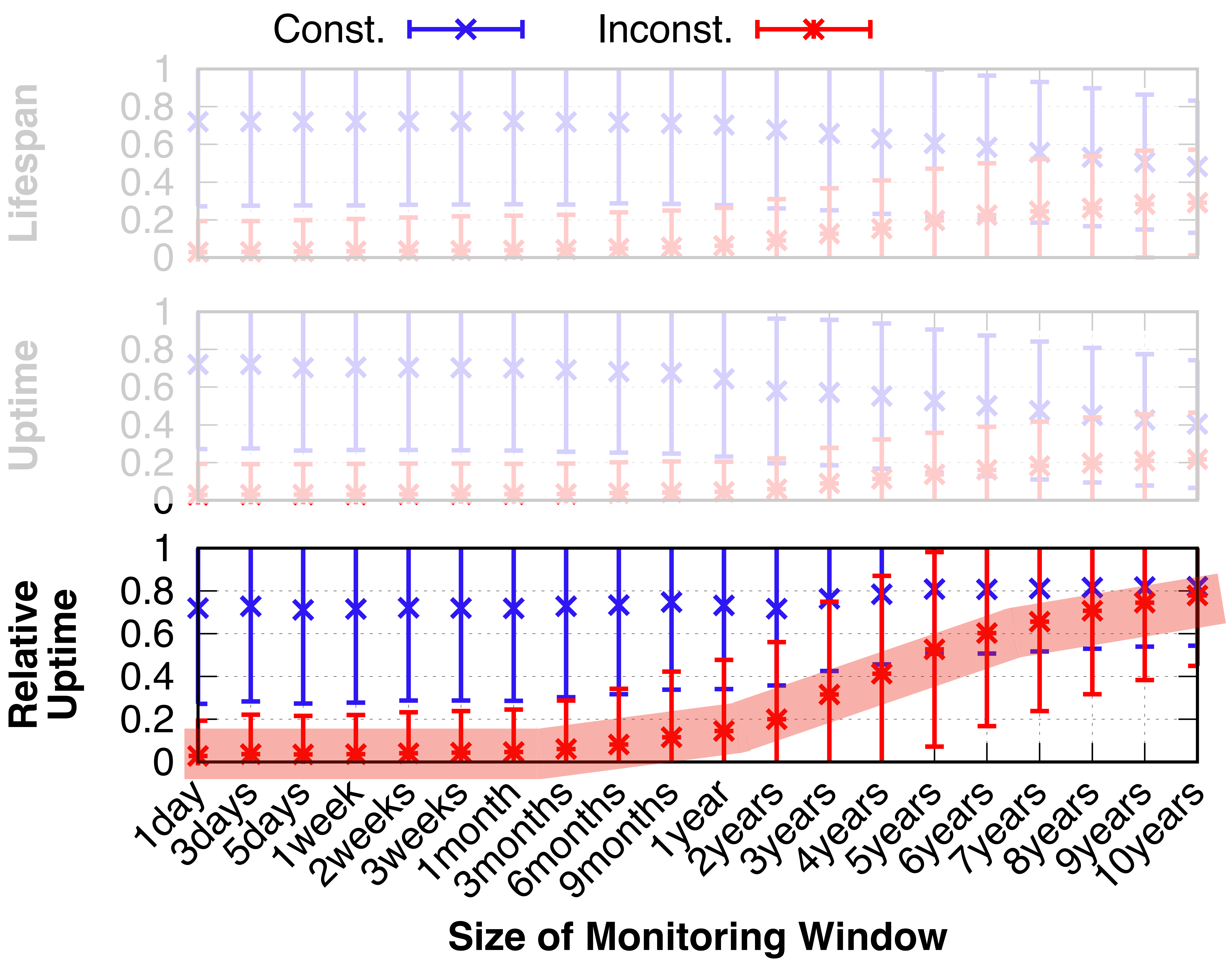


- **Lifespan**
  the difference in dates between the first and last observations, divided by a monitoring window size

monitoring window size

- **Relative uptime**
  uptime/lifespan

BGP announcement pattern

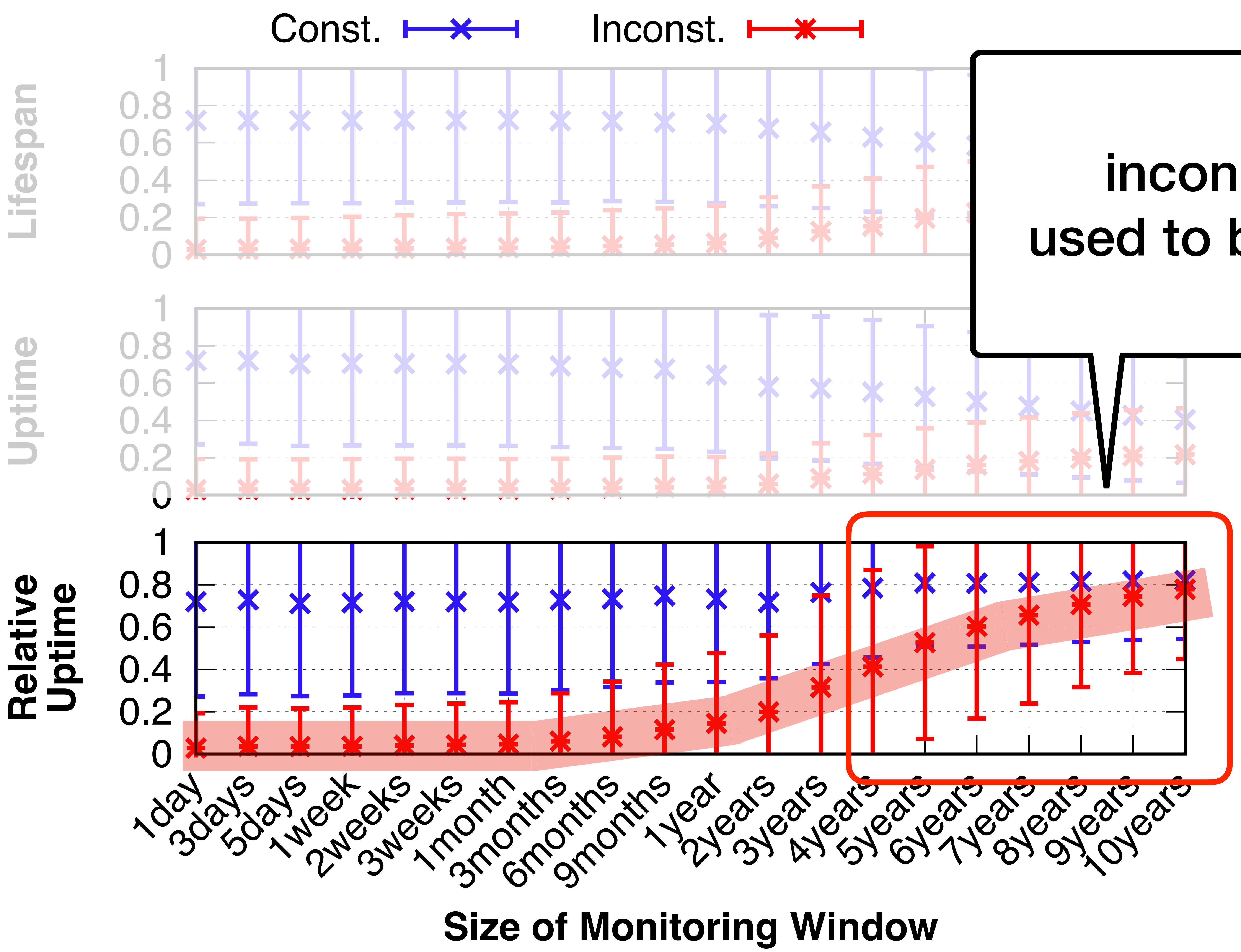


Monitoring window
a time period from the start time t to the latest date of our dataset

- **Lifespan**
  the difference in dates between the first and last observations, divided by a monitoring window size
- **Uptime**
  the number of days that BGP announcements have been observed, divided by a monitoring window size
- **Relative uptime**
  uptime/lifespan

BGP announcement pattern

Monitoring window
a time period from the start time t to the latest date of our dataset

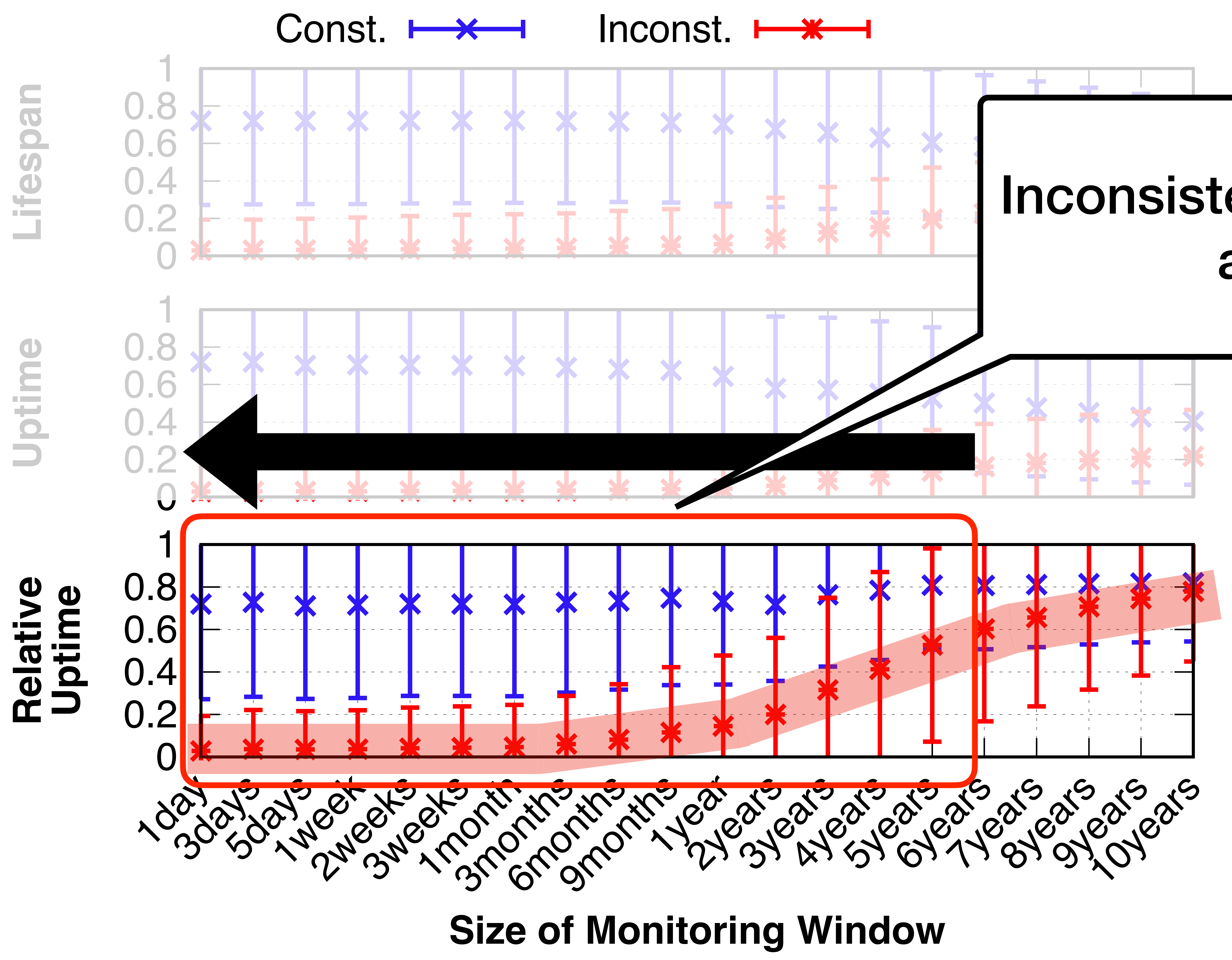


inconsistent IRR objects used to be actively announced

- **Uptime**
  the number of days that BGP announcements have been observed, divided by a monitoring window size
- **Relative uptime**
  uptime/lifespan

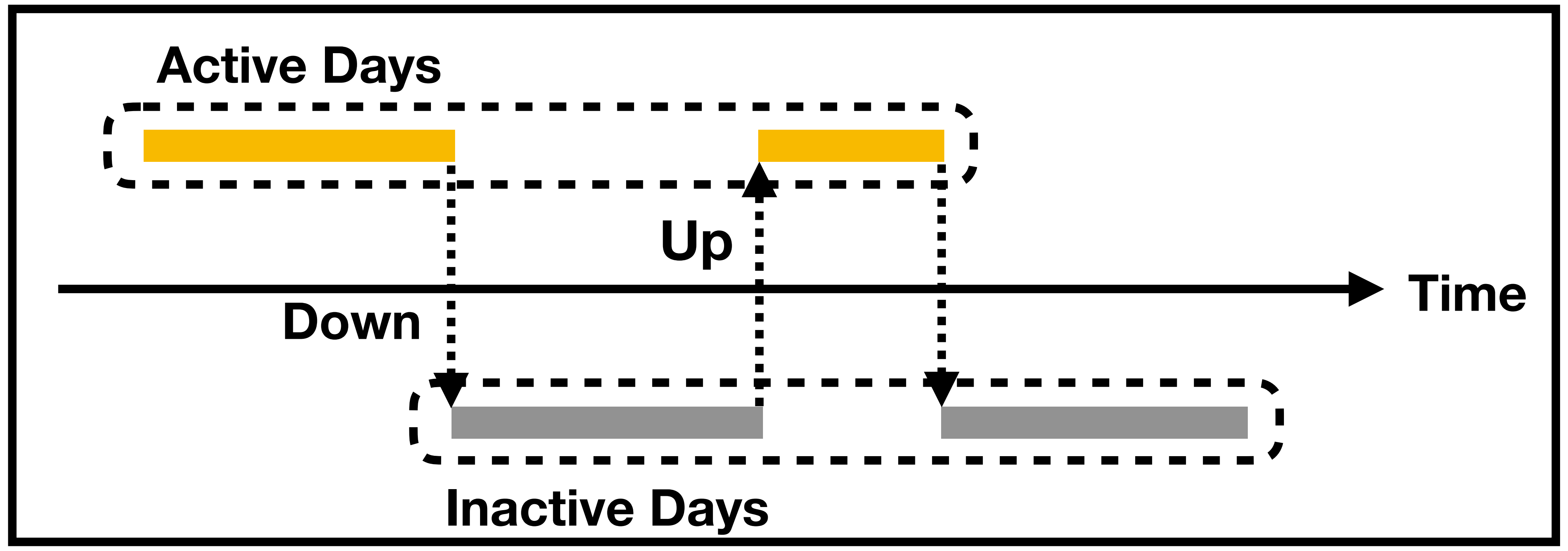
BGP announcement pattern

Monitoring window
a time period from the start time t to the latest date of our dataset



- Lifespan
- Uptime
  the number of days that BGP announcements have been observed, divided by a monitoring window size
- Relative uptime
  uptime/lifespan

Features



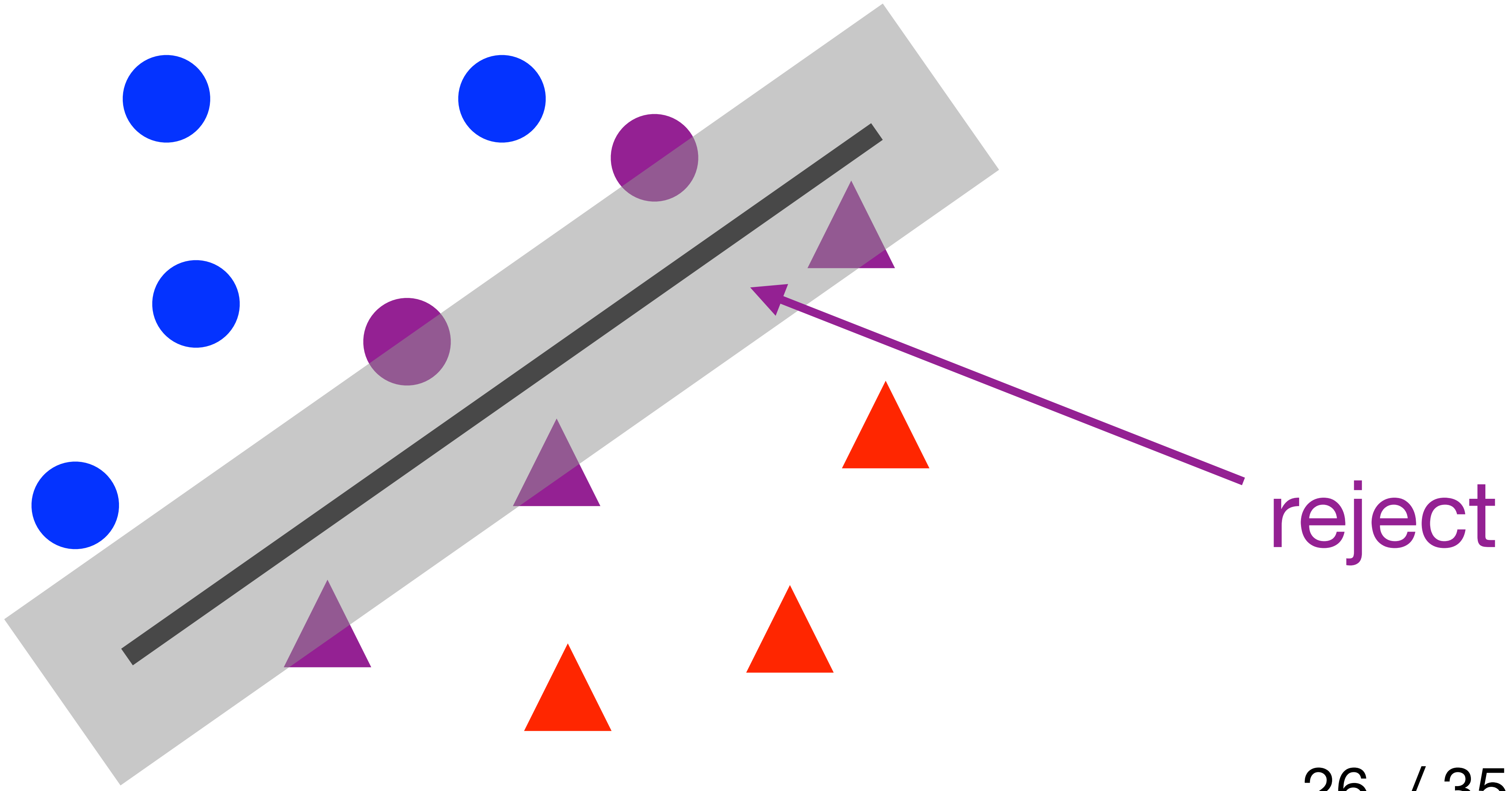
- Features
  - More than 300 features for each prefix-origin pair in IRR
- Metrics:
  - Lifespan, Uptime, Relative uptime
  - # of Ups/Downs, Active/Inactive Days

Model

LightGBM

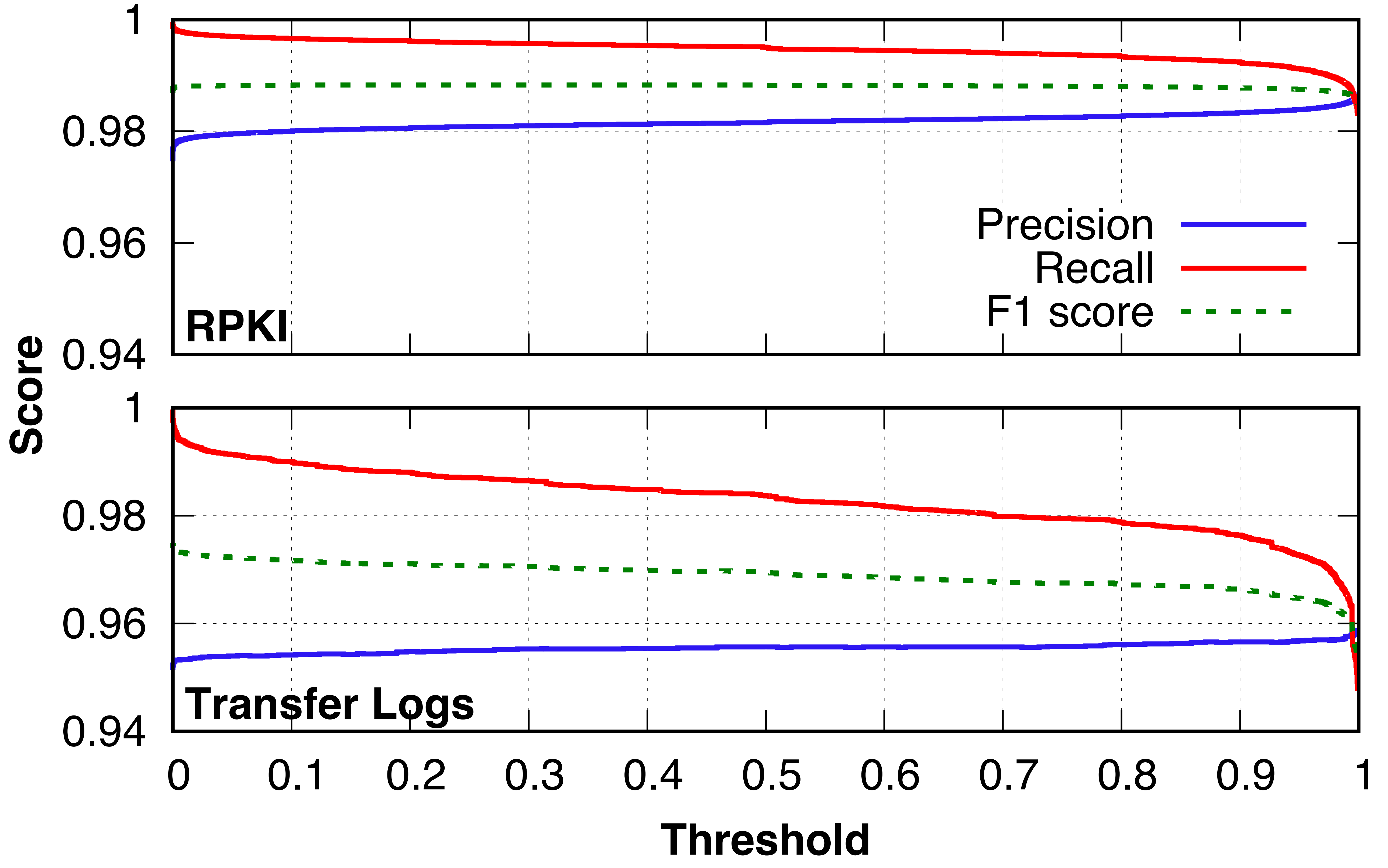
+

Classification with rejection

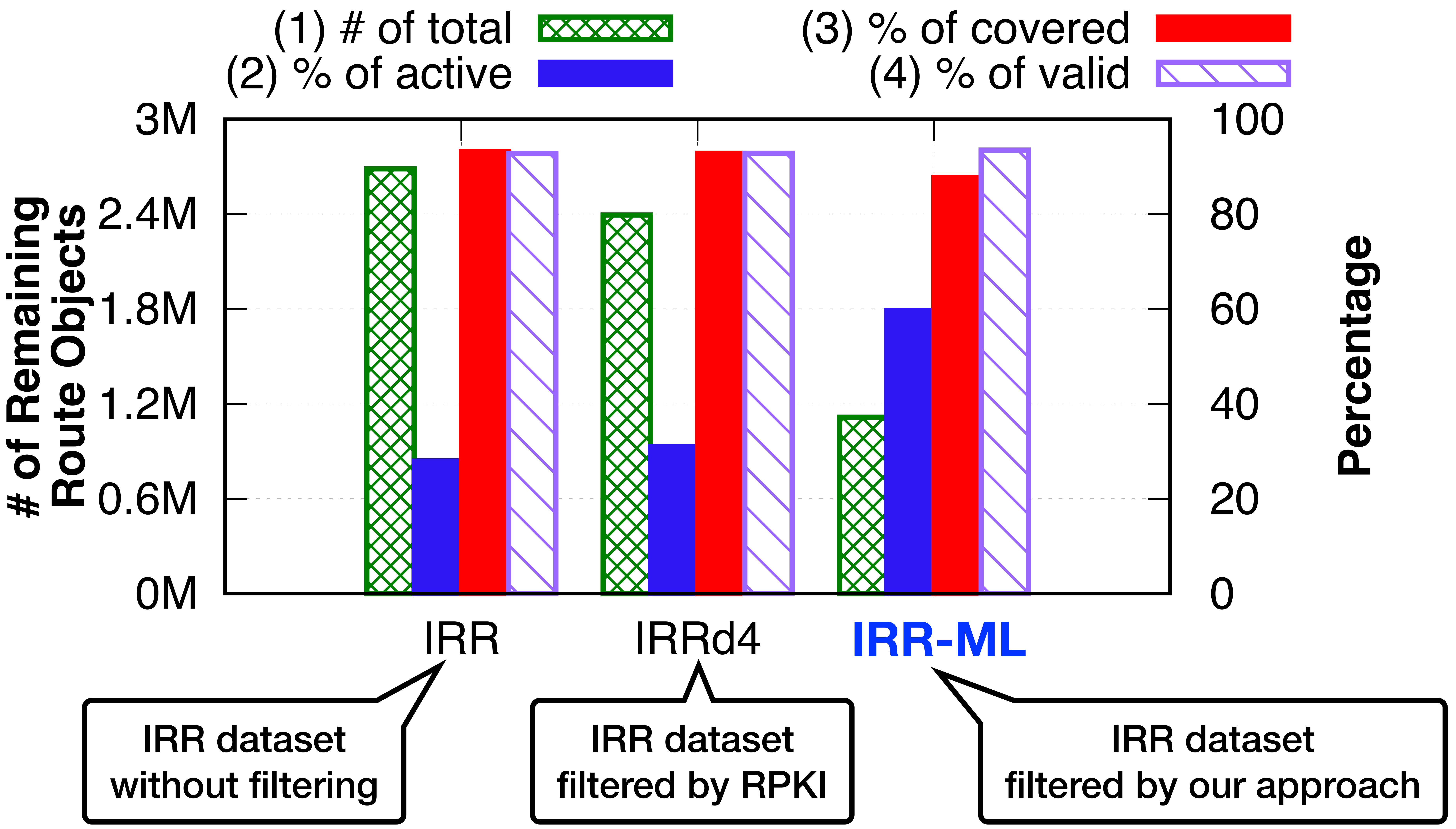


Evaluation with two ground truth datasets

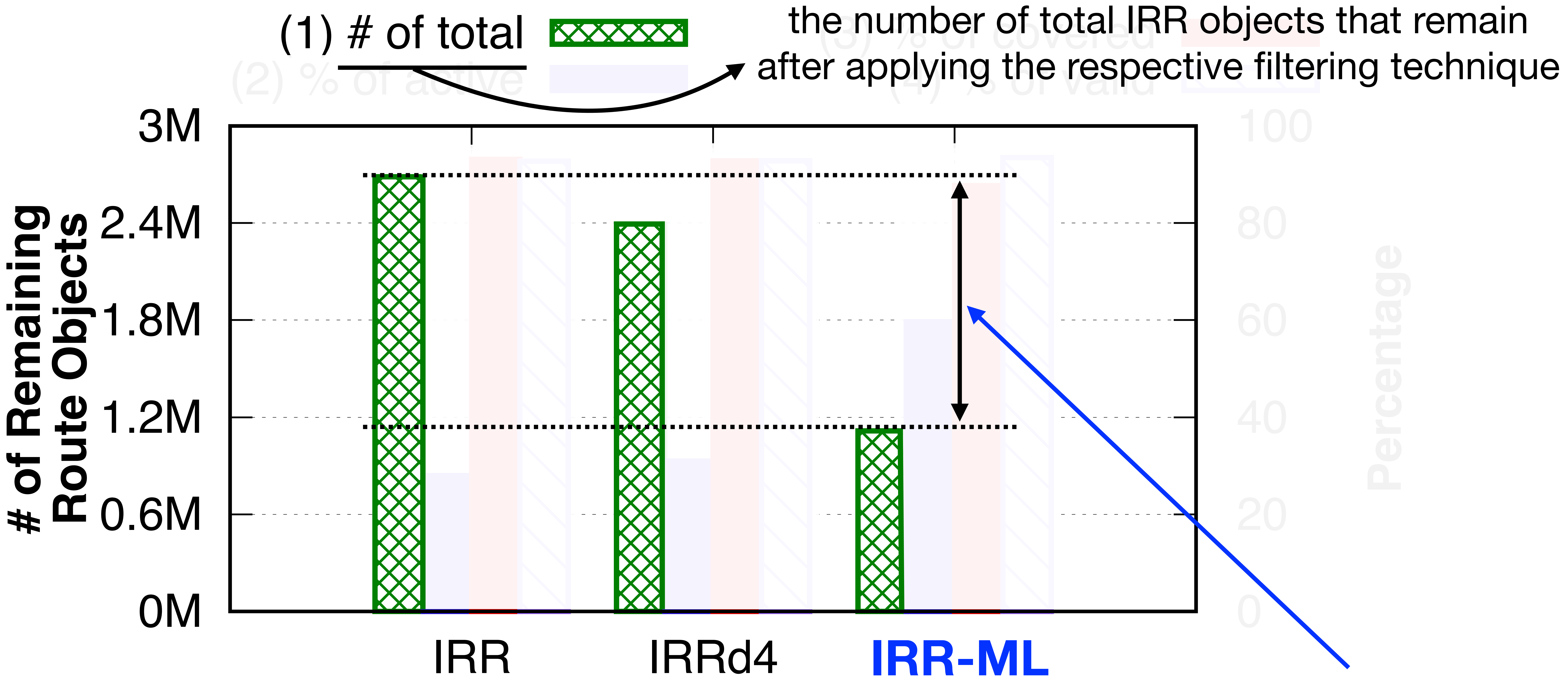
- RPKI
  - ROAs: IP prefix, origin AS
  - Origin AS is the owner of the IP prefix
- Transfer logs from RIRs
  - IP prefixes can be transferred between organizations
  - Transfer logs: IP prefix, source and recipient organizations
  - Recipient organization is the owner of the IP prefix



Comparison with original IRR and RPKI-filtered IRR (IRRd4)



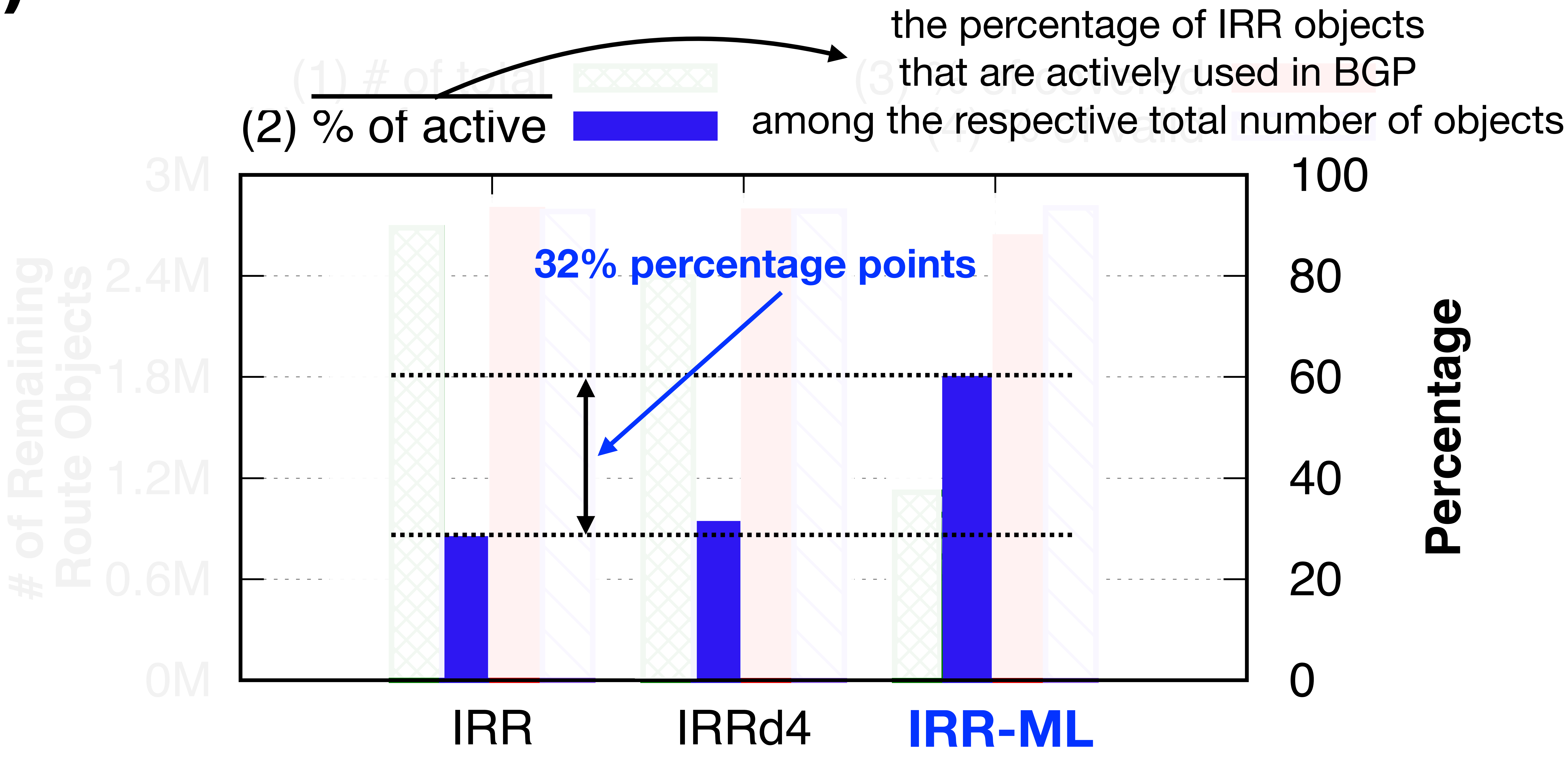
Comparison with original IRR and RPKI-filtered IRR (IRRd4)



59% of IRR objects are filtered

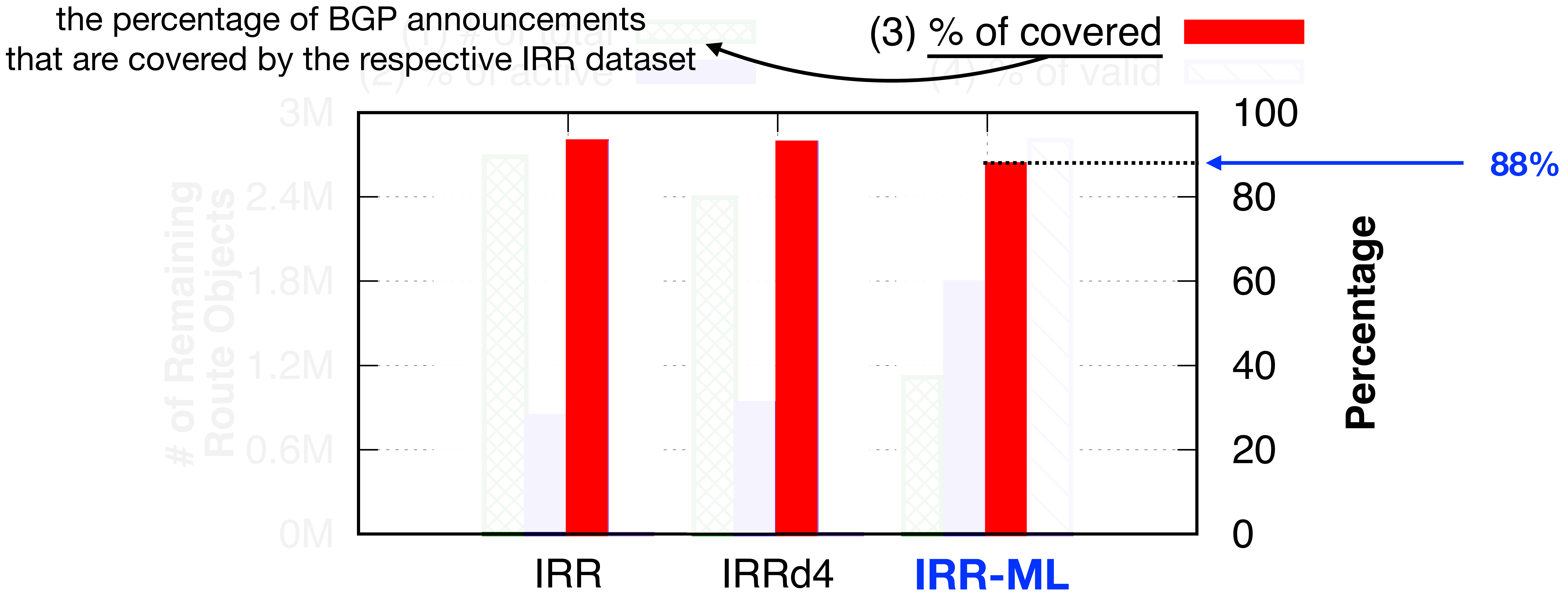
59% of IRR objects in original IRR dataset

Comparison with original IRR and RPKI-filtered IRR (IRRd4)



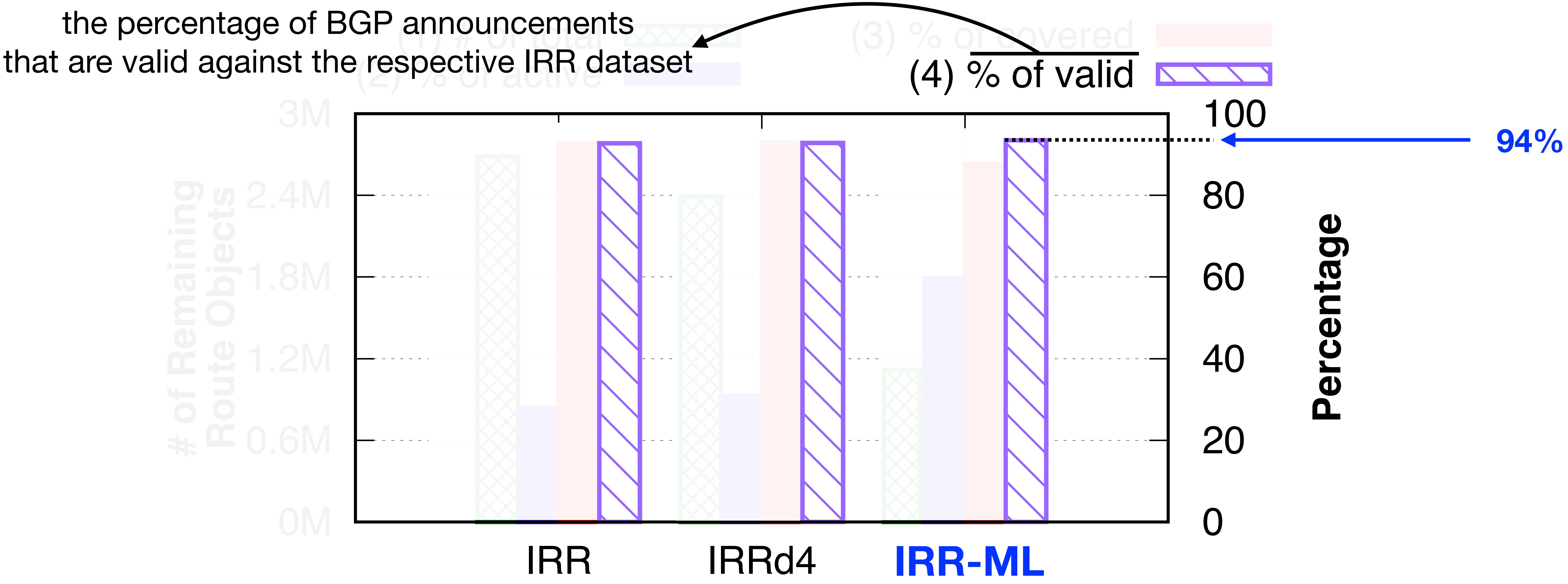
60% of IRR objects are used

Comparison with original IRR and RPKI-filtered IRR (IRRd4)



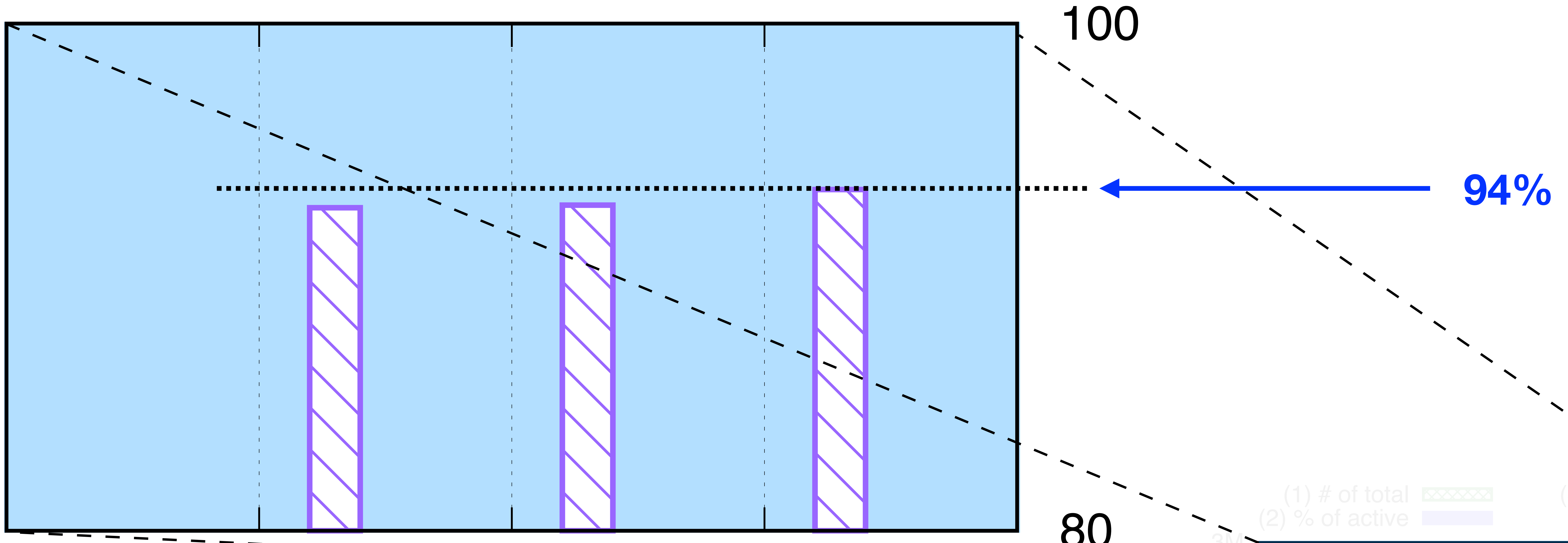
While filtering out 59% of IRR objects,
we still cover 88 % of BGP announcements

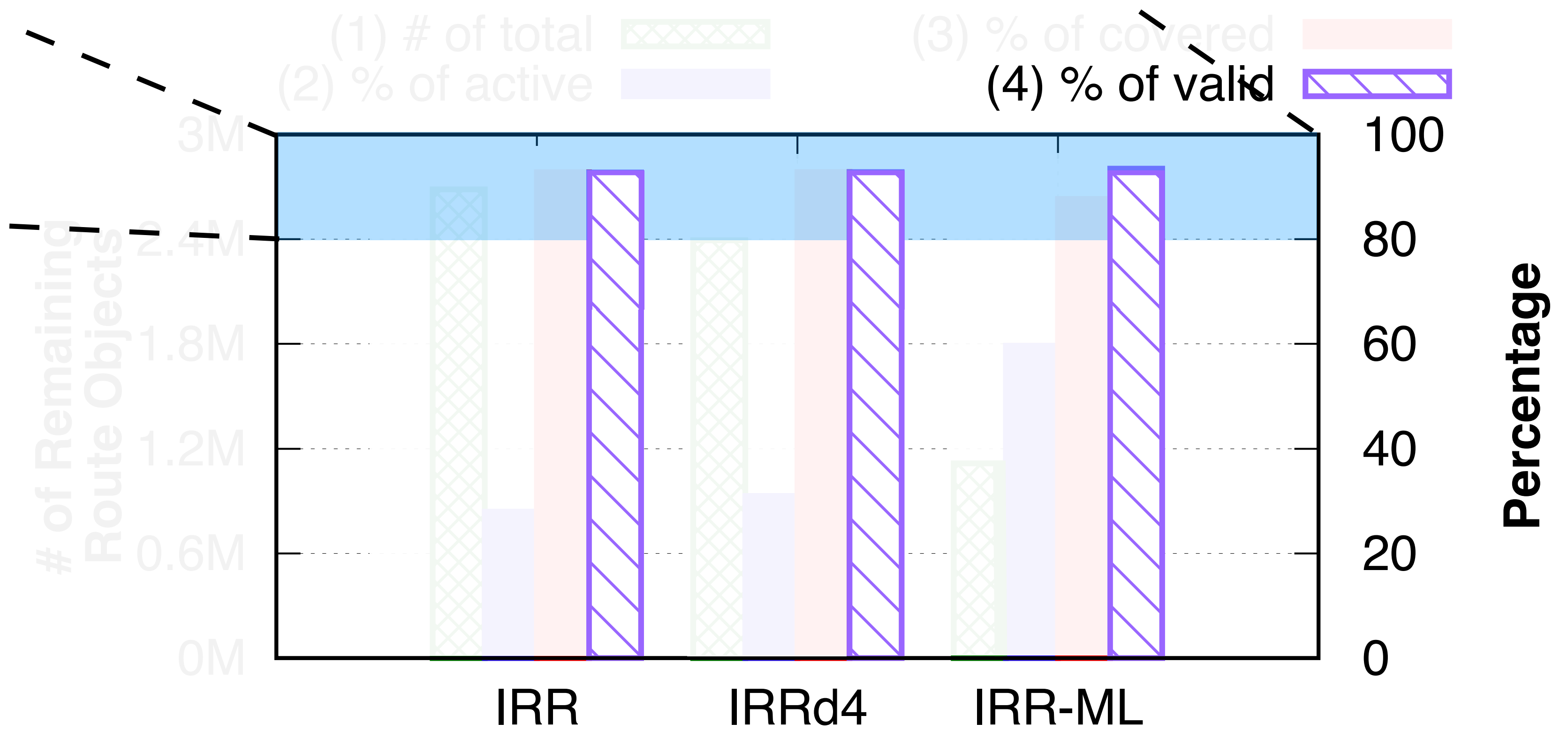
Comparison with original IRR and RPKI-filtered IRR (IRRd4)



94% of the covered BGP announcements are valid against IRR dataset filtered by our approach

Comparison with original IRR and RPKI-filtered IRR (IRRd4)





the percentage of valid BGP announcements slightly increases

Discussion and future work

- Who would be responsible for applying our technique?
  - IRR vs. network operators

- Reducing false negatives
  - Grouping IRR objects by the prefixes and select the most up-to-date IRR object for each IP prefix

- Source code and dataset are publicly available
  - irredicator.netsecurelab.org

Conclusion

- Conduct a **longitudinal study of the inconsistencies** between IRR and RPKI
  - found that the number of inconsistent IRR objects increases
- Analyze the **characteristics of the inconsistent IRR objects**
  - captured distinct patterns between consistent and inconsistent IRR objects
- Propose an **ML-based IRR pruning technique**
  - successfully filtered out stale IRR objects (58.5% of the entire IRR)

Thank you

Minhyeok Kang

mhkang@mmlab.snu.ac.kr

mmlab.snu.ac.kr/~mhkang/

irredicator.netsecurelab.org/

Backup

Features

- 13 metrics
  - Uptime, Lifespan, Relative Uptime (=3)
  - # of Ups / Downs (=2)
  - min, max, avg, and std of Active/Inactive days (=8)

- Total 312 features
  - Window based features
    - 13 metrics * 20 monitoring windows = 260 features
  - Statistics for each metric
    - 13 metrics * 4 statistics = 52 features