A Comparative Analysis of BGP Anomaly Detection and Robustness Algorithms

Kotikapaludi Sriram, Oliver Borchert, Okhee Kim,

Patrick Gleichmann, and Doug Montgomery

National Institute of Standards and Technology

(Contact: ksriram@nist.gov; dougm@nist.gov)

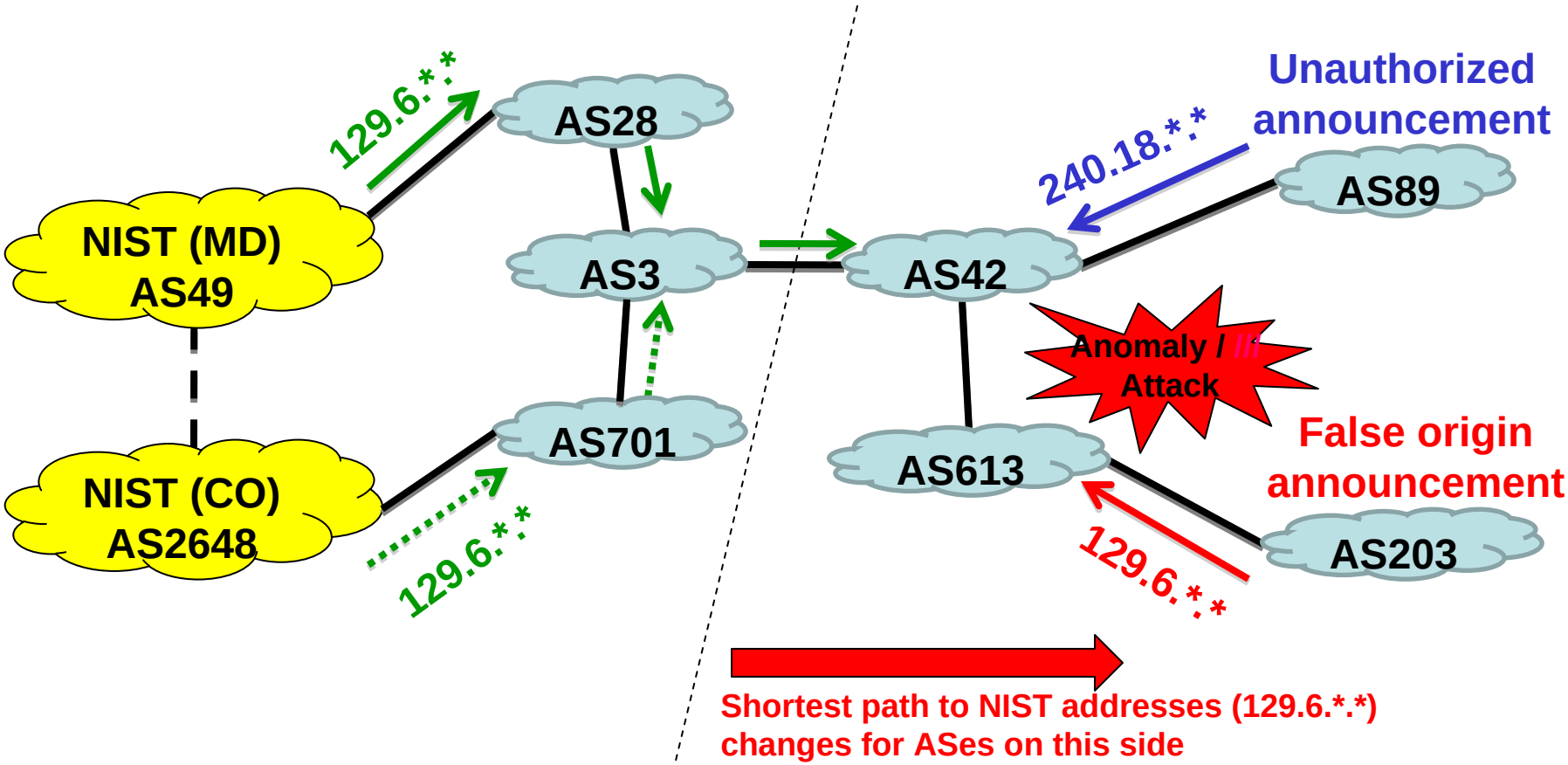
CATCH 2009, March 3-4, 2009

This research was supported by the Department of Homeland Security under the Secure Protocols for the Routing Infrastructure (SPRI) program and the NIST Information Technology Laboratory Cyber and Network Security Program.

Outline of the Talk

- **Problem statement**
- **Known / New BGP robustness schemes**
- **Evaluation of BGP robustness algorithms**
  - **Comparative analysis of utility**
  - **Quantitative results**
- **Conclusions / Future Work**

BGP Robustness Problem Space



Data Driven BGP Robustness

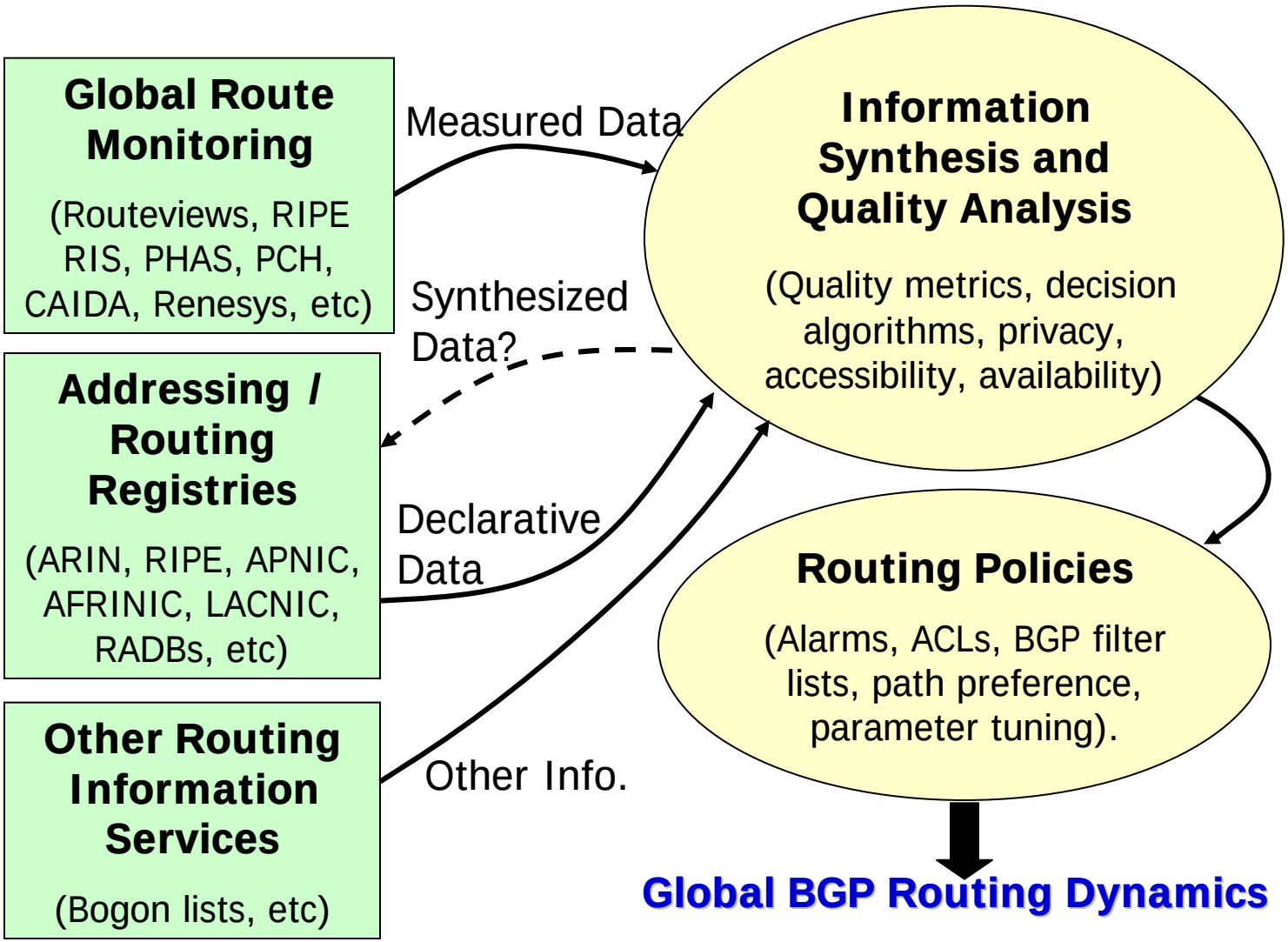
What are the Data Sources?

- Addressing Registries
  - global databases of address block and autonomous system number assignments.
- Routing Registries
  - loosely maintained global databases of contractual relationships for routing services.
- Monitoring Data
  - public BGP monitoring and measurement projects that collect BGP protocol exchanges at various spots around the Internet.

Why is this hard?

- Registries
  - known to be incomplete and inaccurate, and are maintained in differing formats, by differing processes in different regions of the world.
- Robustness Algorithms
  - to be effective, must make precise policy decisions from highly imperfect data.
- Needle in a Hay Stack
  - millions of BGP update messages per day, millions of registry entries, rare but potent threats.

Solution Components / Players



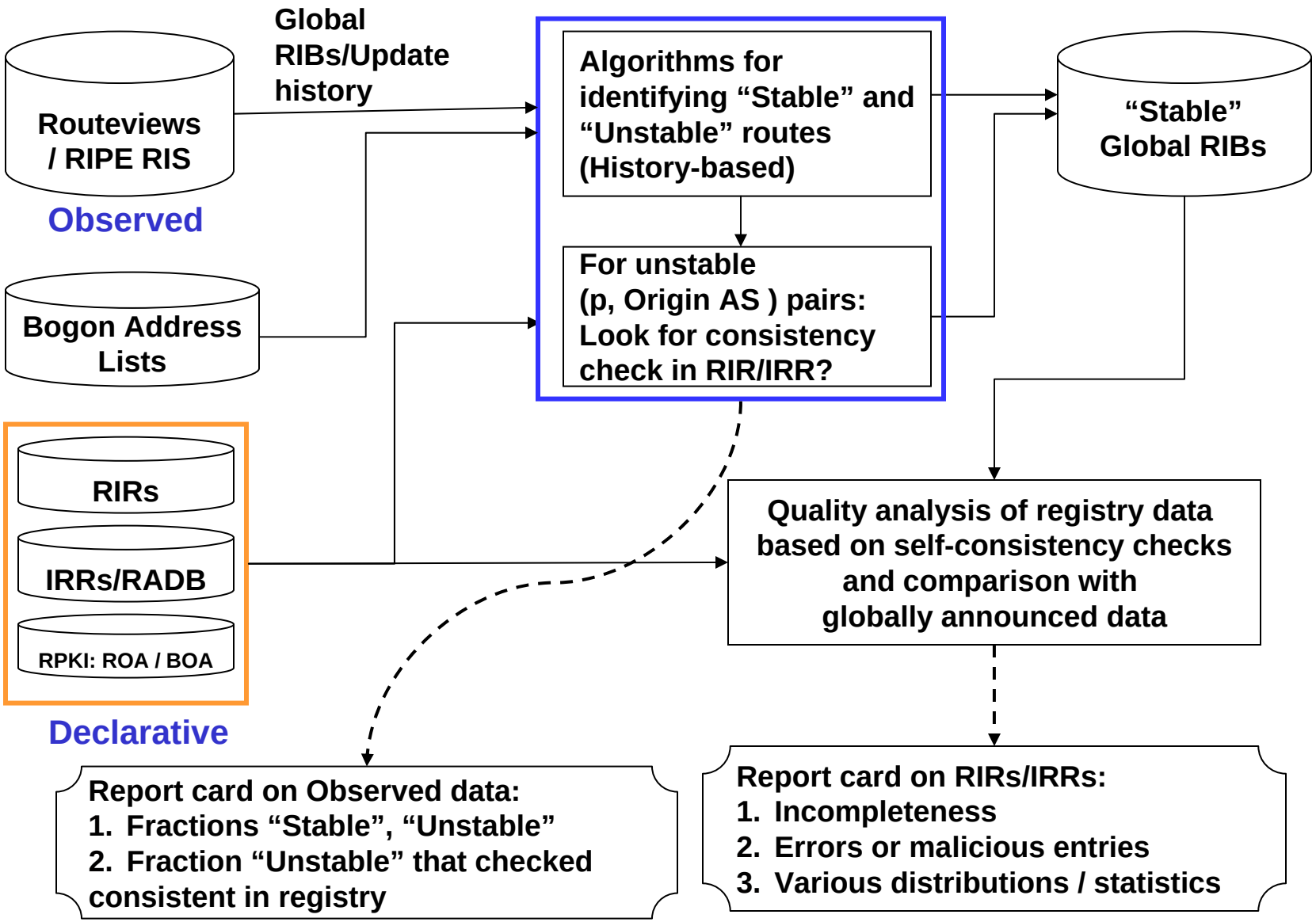
Outline of the Talk

- Problem statement
- Known / New BGP robustness schemes
- Evaluation of BGP robustness algorithms
  - Qualitative / comparative analysis of utility
  - Quantitative results
- Conclusions / Future Work

Known BGP Robustness Algorithms

- General goal: Validate an observed (p, Origin AS) pair
- Nemecis: Compare with registered objects (route, inetnum, autnum)
- PHAS: Compare with historically observed (p, Origin AS) pairs, AS-paths:
  - Identify origin changes, subprefix announcements; generate alerts
- Pretty Good BGP (PGBGP): Compare with historically observed (p, Origin AS) pairs
  - Influence forwarding or holding back of updates in real-time in BGP processing

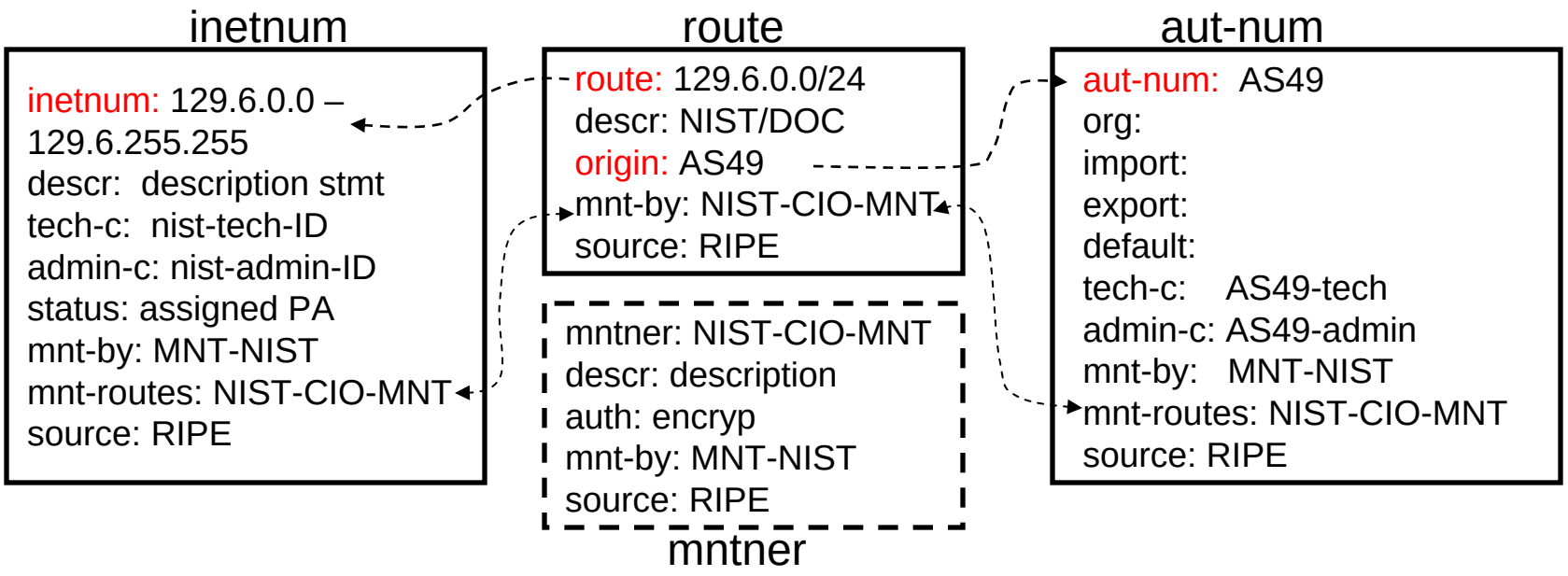
New Integrated Approach



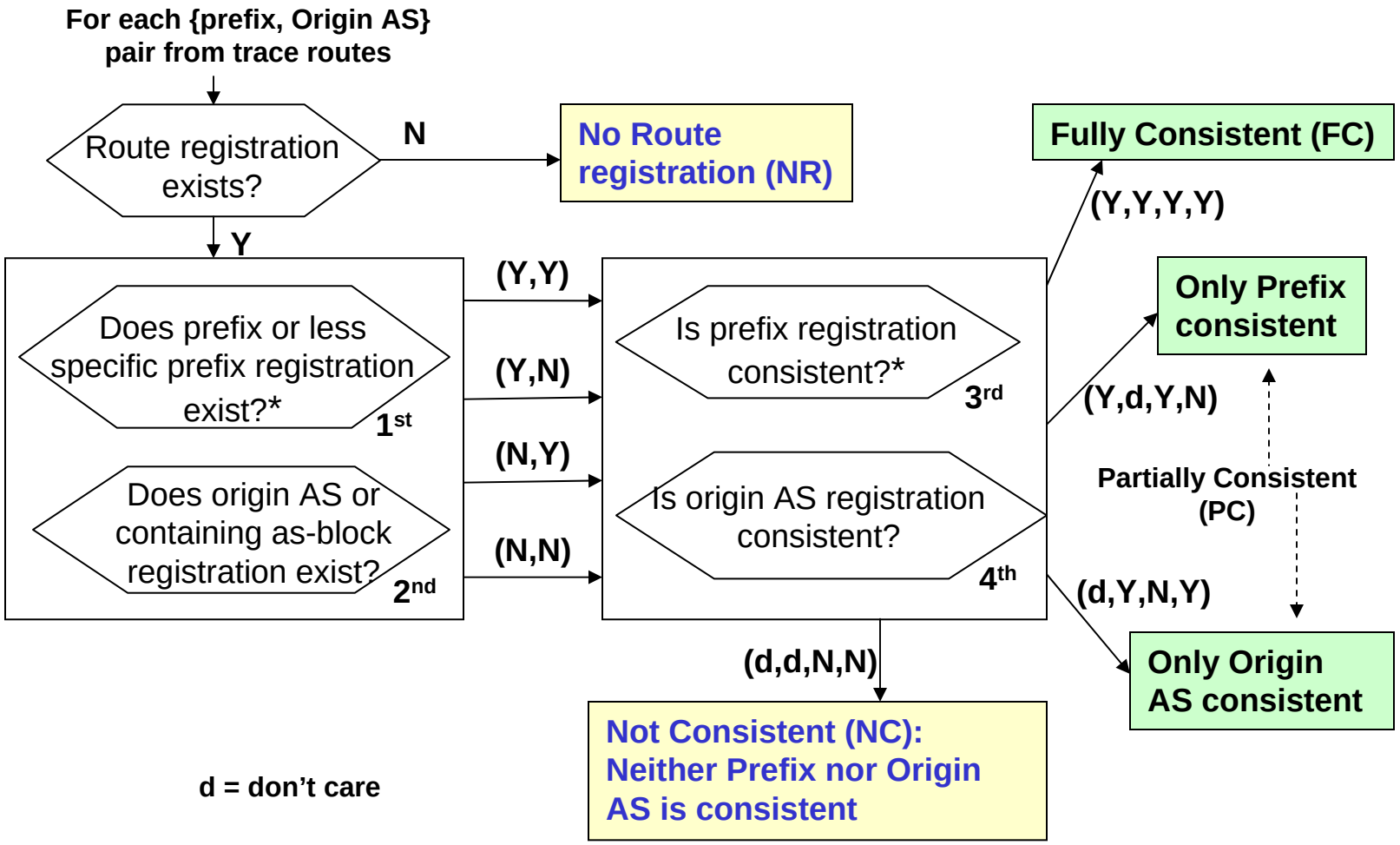
ROA: Route Origin Attestation

BOA: Bogon Origin Attestation

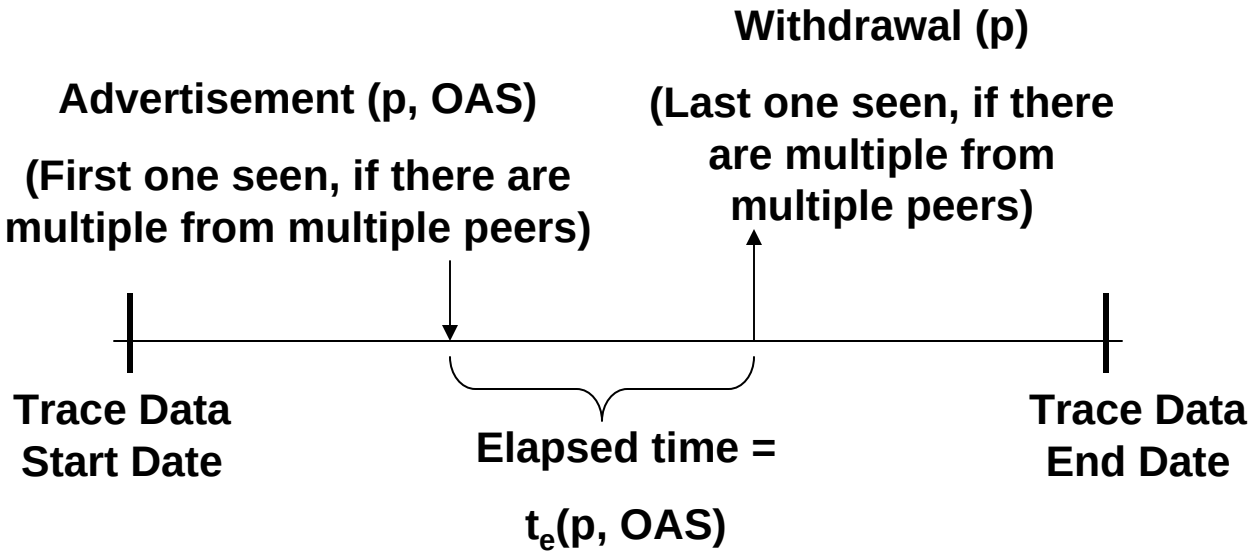
Checking Consistency of a Registered Route with Corresponding Inetnum and Aut-Num



Registry-Based Algorithm for Scoring Routes Observed in Trace Data

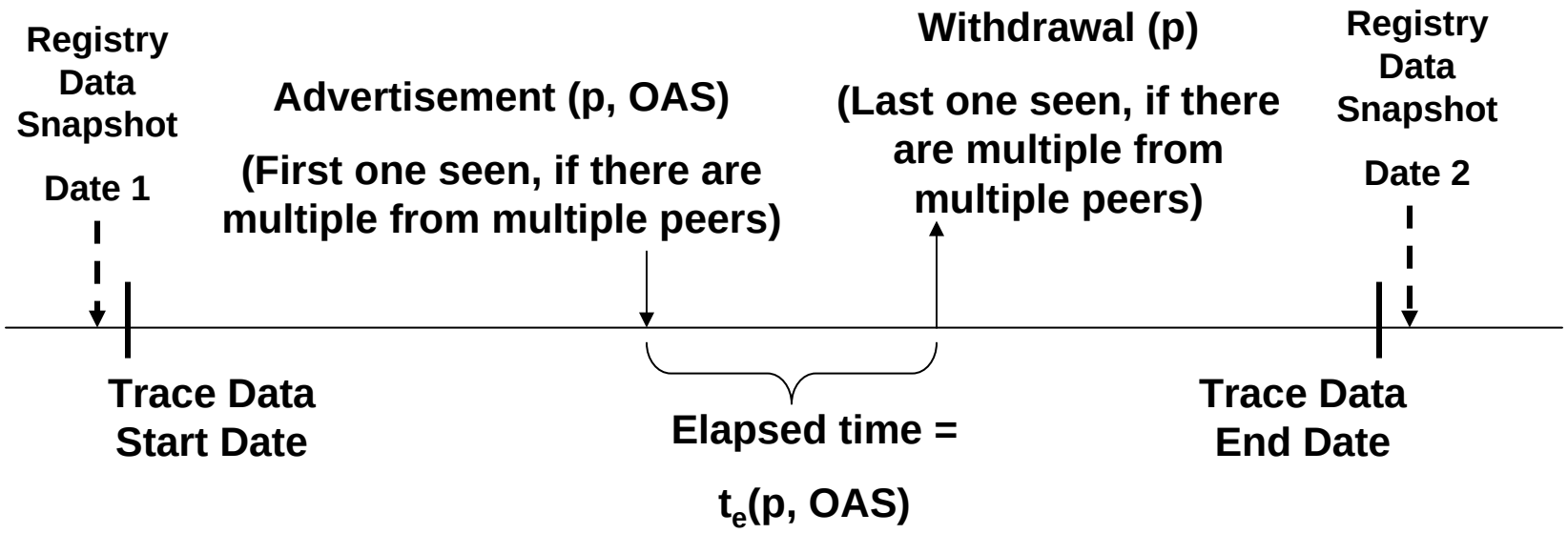


Enhanced History-Based Algorithm for Determining Stability of (p, OAS) in the Trace Data



- If t_e(p, OAS) $\geq$ 48 hours, then (p, OAS) is a stable (prefix, Origin AS) pair
- If t_e(p, OAS) < 48 hours, then (p, OAS) is an unstable (prefix, Origin AS) pair
- Update data is initialized with stable (i.e., persistent for $\geq$ 48 hours) RIB entries

Enhanced Hybrid Algorithm for Validating (p, OAS) in the Trace Data



- Use enhanced history-based (i.e., trace-data-based) algorithm as in previous slide
- Complement it with combined results of the registry-based algorithm with data from two dates (close to start and end dates of the history algorithm)
- Result: Better performance of anomaly detection algorithms

Outline of the Talk

- Problem statement
- Known / New BGP robustness schemes
- Evaluation of BGP robustness algorithms
  - Comparative analysis of utility
  - Quantitative results
- Conclusions / Future Work

Comparative Analysis of Existing and Enhanced Algorithms

- We have encoded Registry-based, Enhanced Trace-data-based and Enhanced Hybrid algorithms for evaluation
- Algorithms are run on top of the NIST TERRAIN* framework
  - Unified database of Registry / Trace data (RIRs, IRRs, RIPE-RIS, Routeviews)
- Tested and compared the algorithms

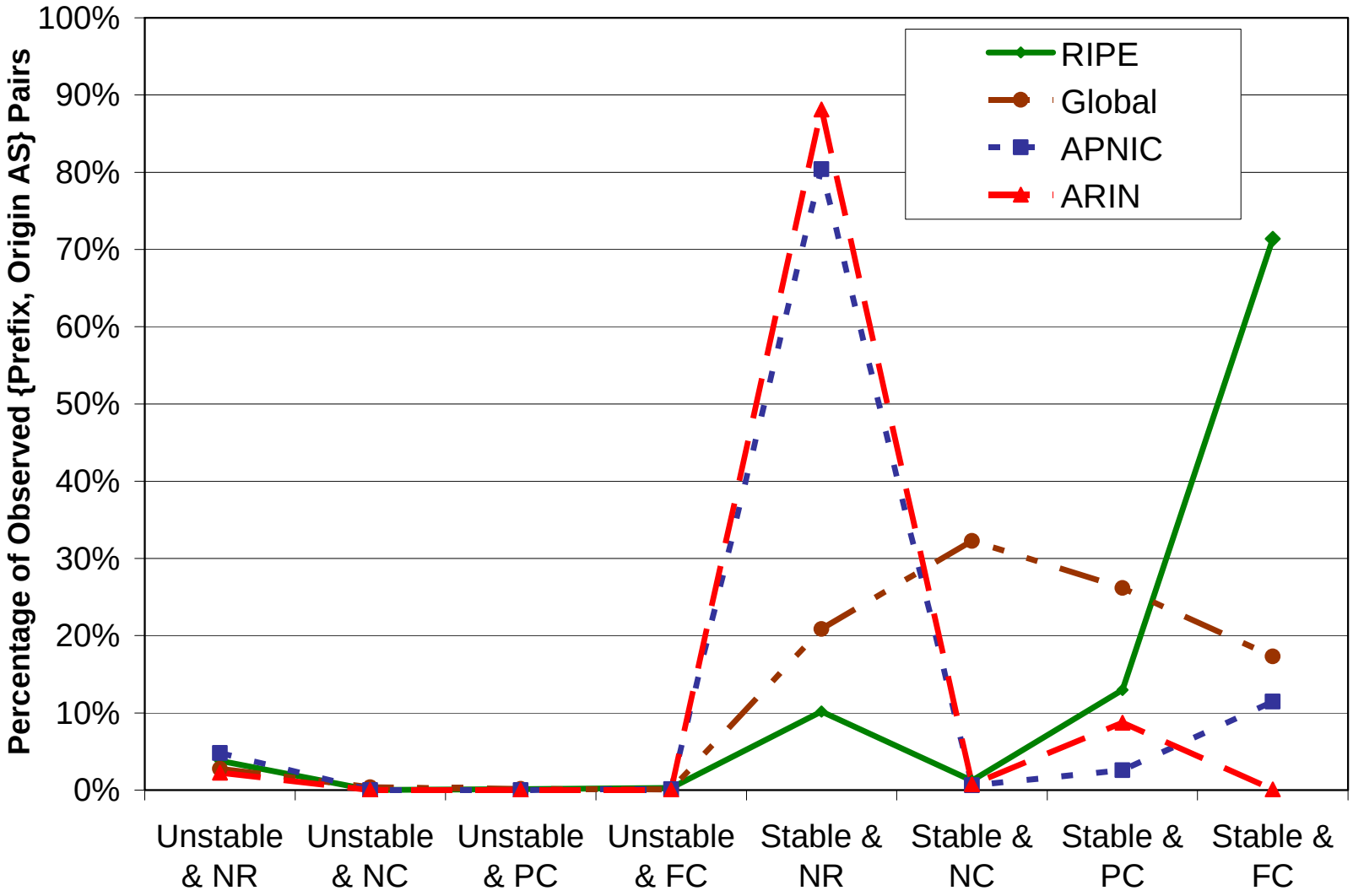
* TERRAIN: Testing and Evaluation of Routing Robustness in Assurable Inter-domain Networking

Comparative Analysis of Existing and Enhanced Algorithms (Contd.)

For the purpose of this presentation:

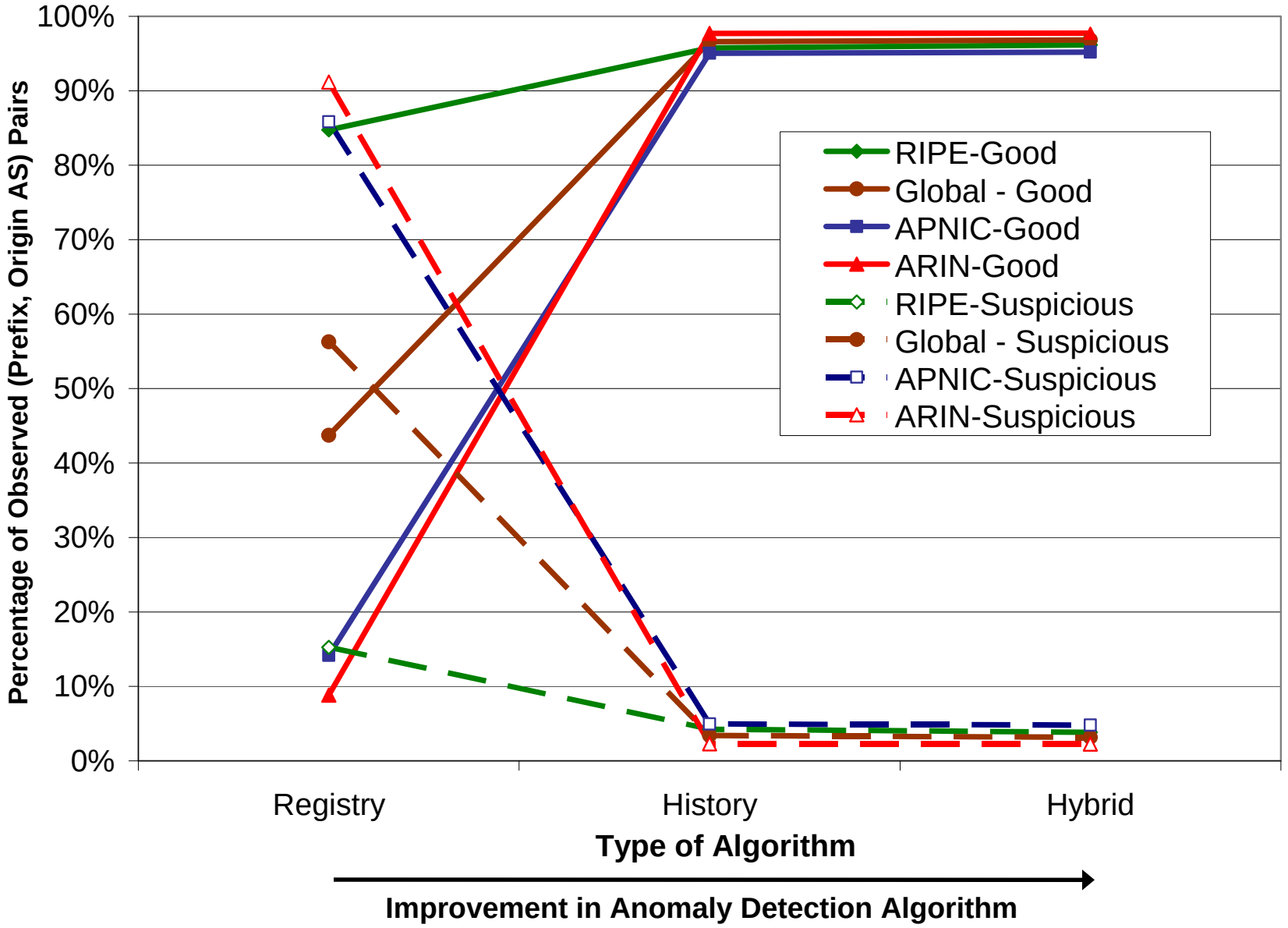
- Results focus on Origin AS validation
- Results are reported globally for all prefixes as well as selectively for regional (RIPE, ARIN, …) prefixes
- Six-month trace-data window (January through June 2007); initialized with stable RIB entries
- Registry data – two dates prior to and towards the end of the six-month window (December 12, 2006 and June 18, 2007)

Classification of Observed (p, OAS) Pairs According to Stability / Consistency Scores

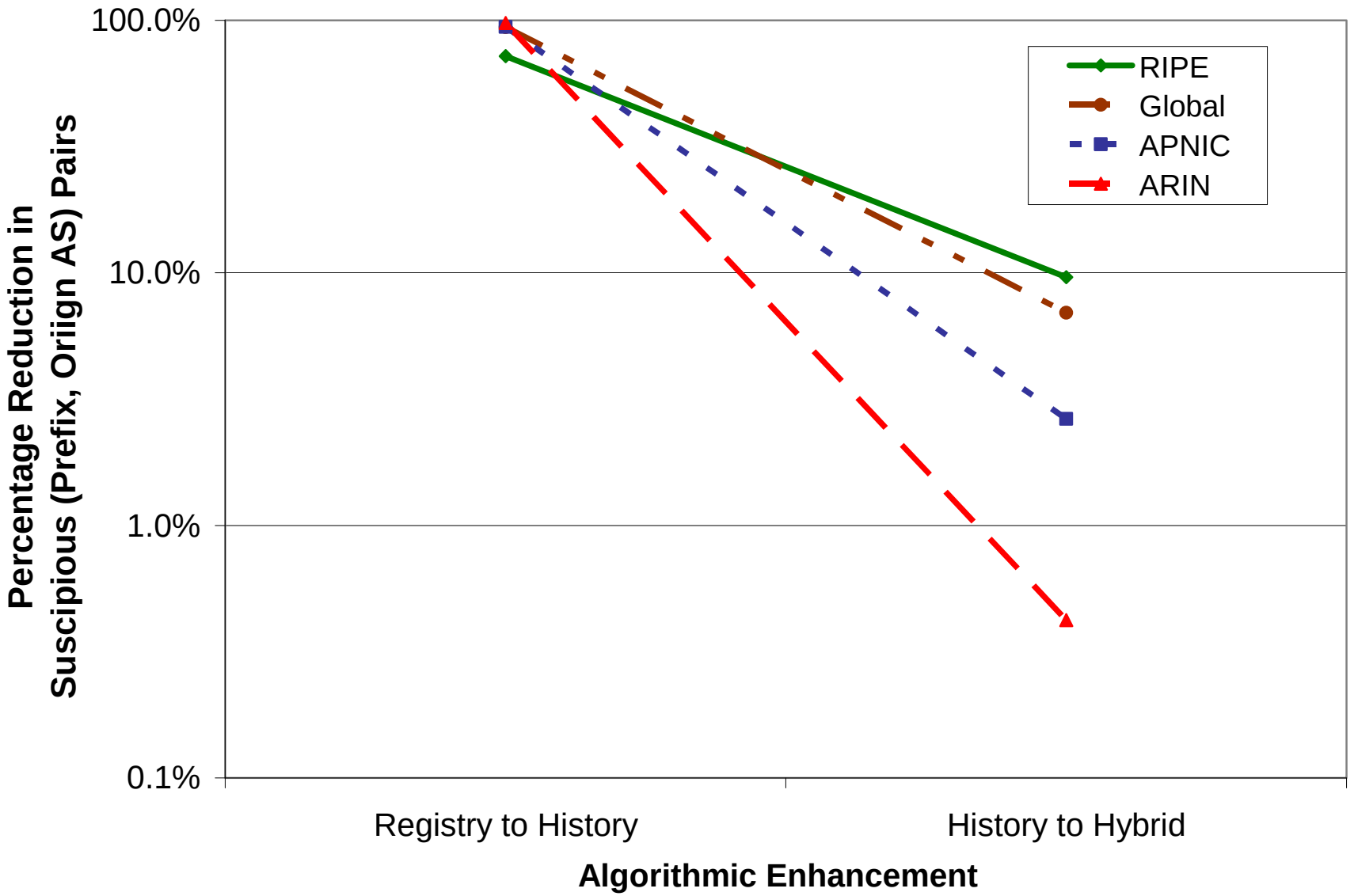


p = prefix; OAS = Origin AS; FC = Fully Consistent; PC = Partially Consistent; NC = Not Consistent; NR = Not Registered

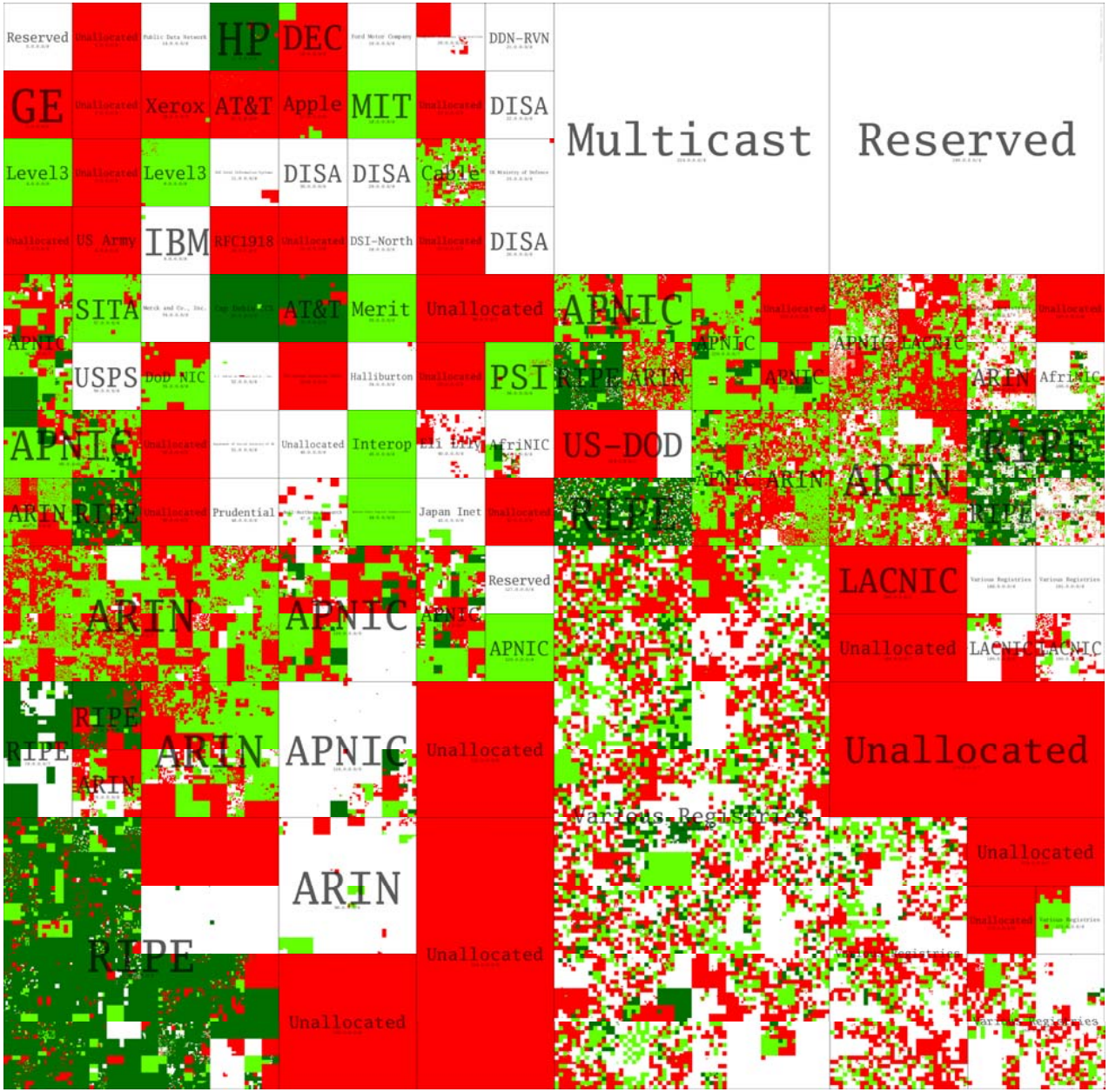
Comparative Performance of Algorithms



Comparative Performance of Algorithms



Checking Origin AS : Comparison of Algorithms



Registry-based Algorithm

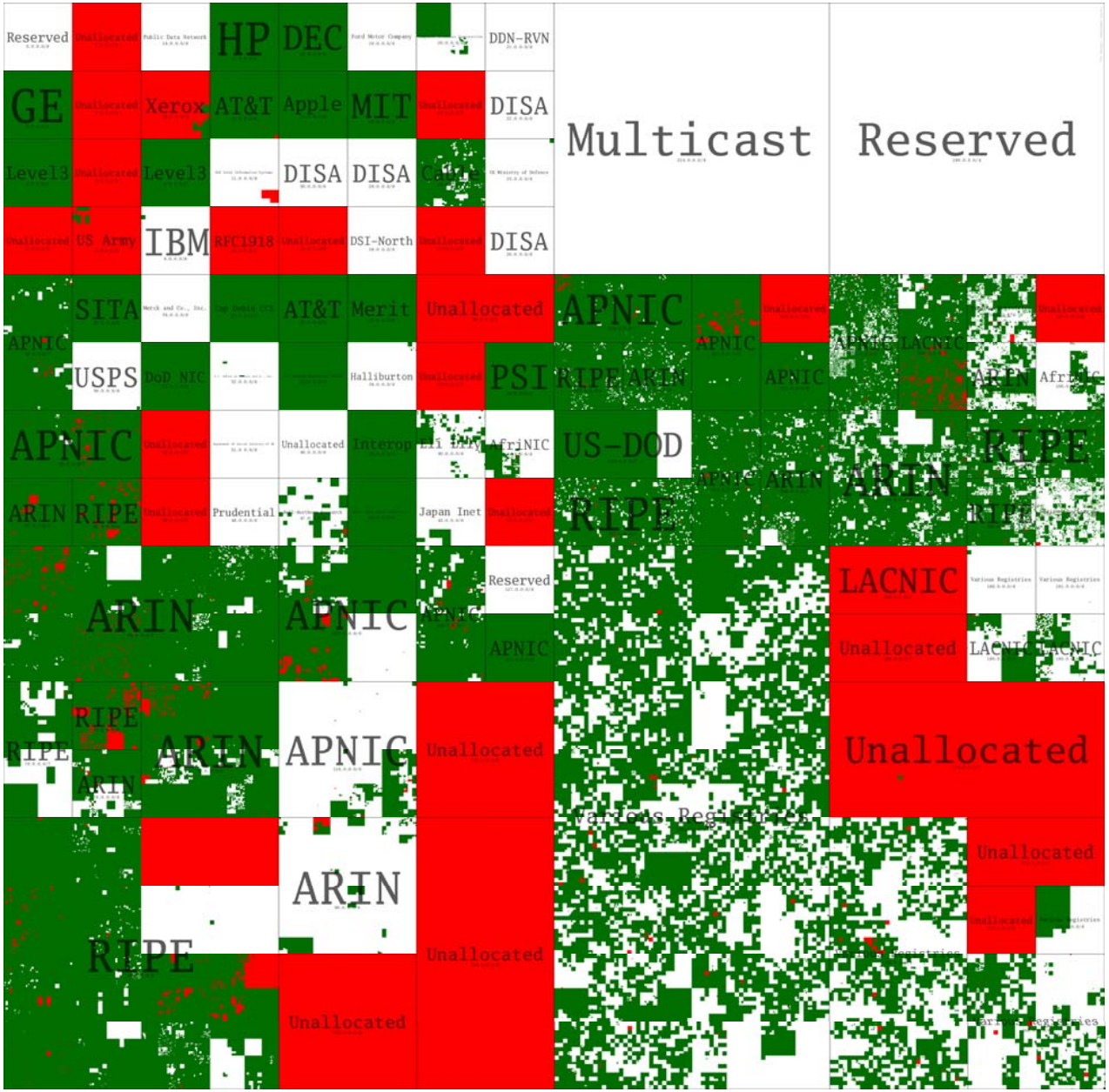
Green: Good / FC

Light Green: Good / PC

Red: Suspicious

White: Not found in trace data

Checking Origin AS : Comparison of Algorithms



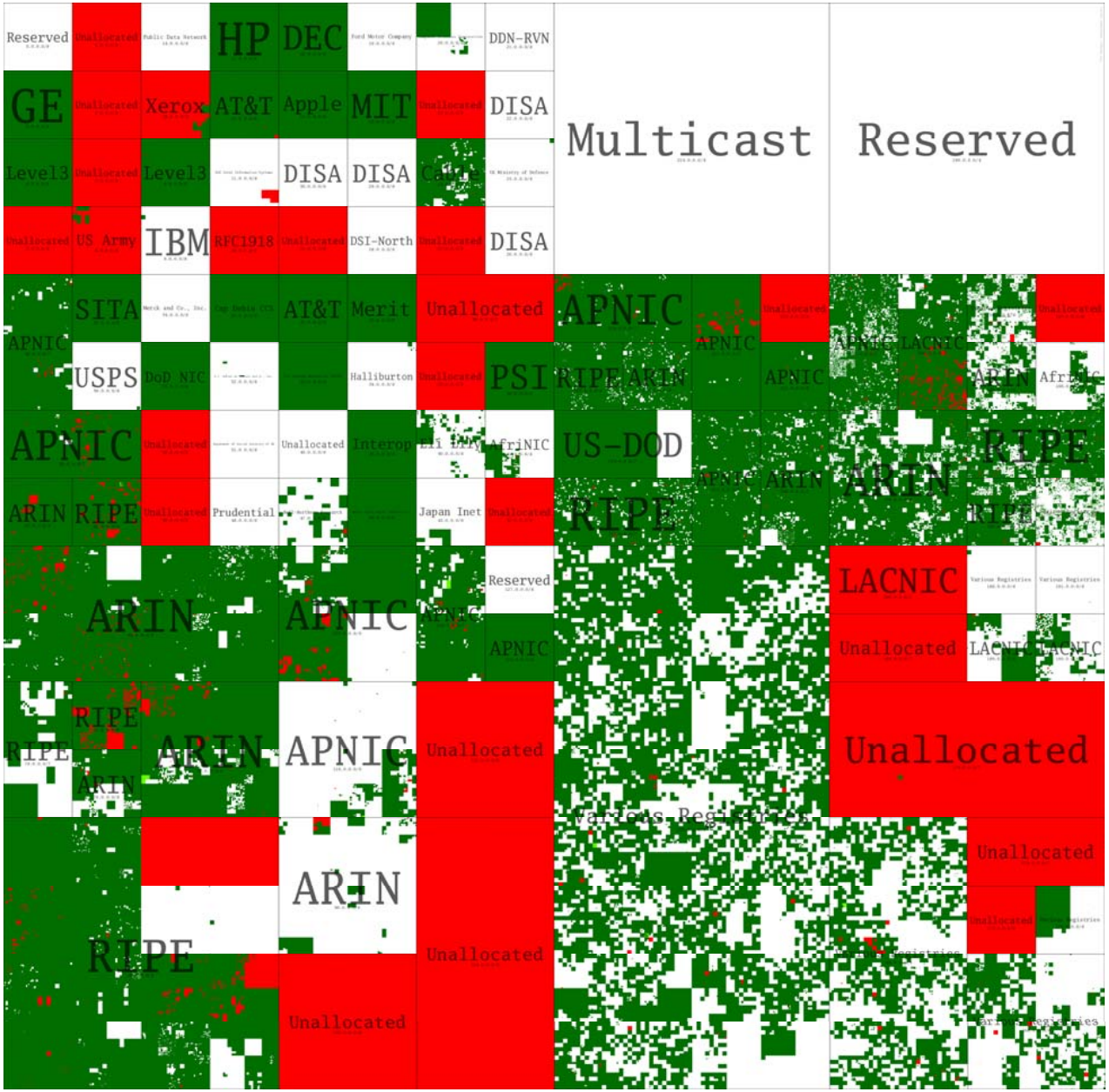
Enhanced trace-data-based Algorithm

Green: Good

Red: Suspicious

White: Not found in trace data

Checking Origin AS : Comparison of Algorithms



Enhanced Hybrid Algorithm

Green: Good / FC

Light Green: Good / PC

Red: Suspicious

White: Not found in trace data

Prefixes with Multiple Origin ASes

# Origin ASes	# Prefixes
1	476243
2	55673
3	10419
4	2683
5	965

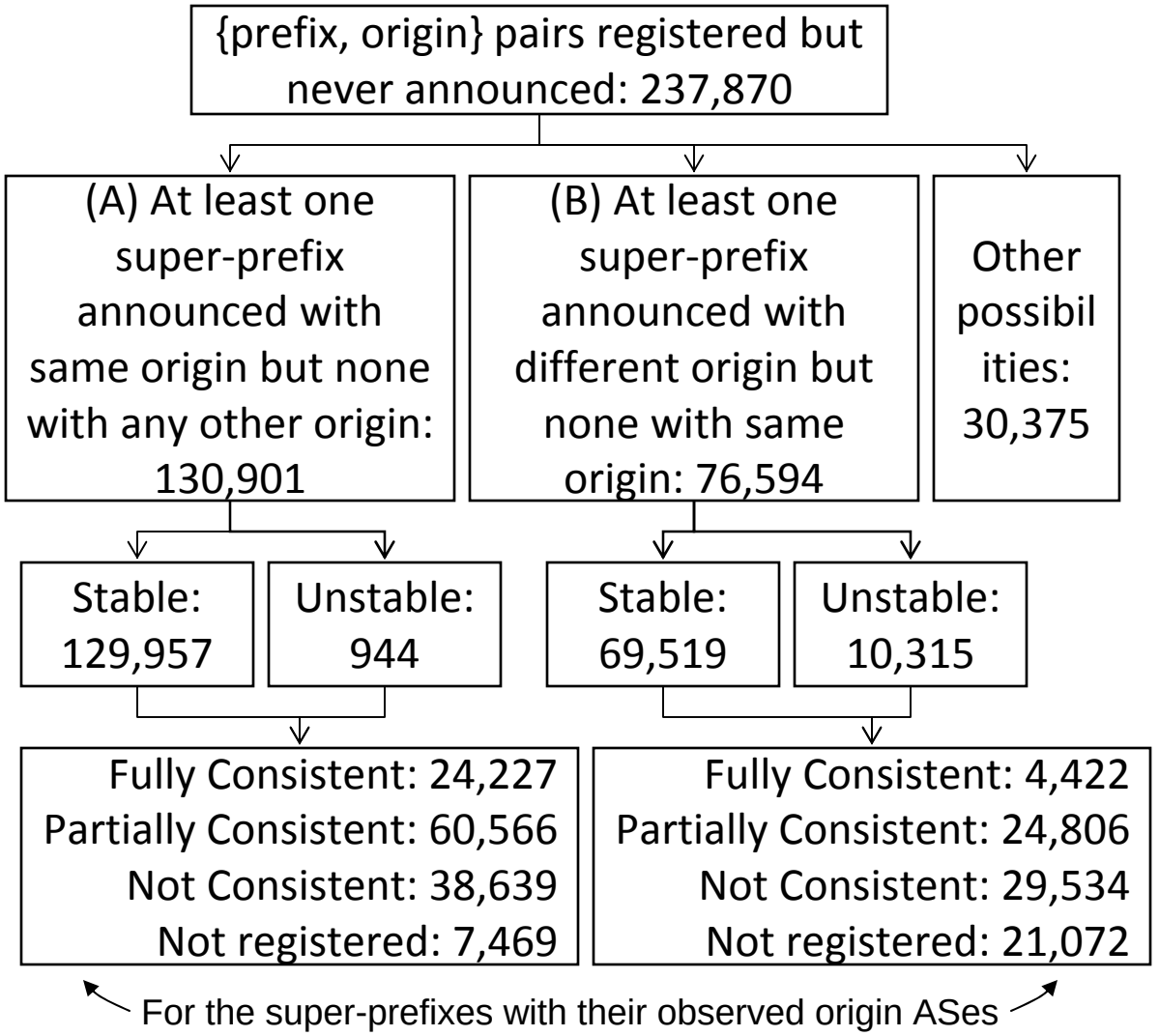
For prefixes with two Origin ASes:

OAS1	OAS2	# Prefixes
FC + Stable	FC/PC + Unstable	23
PC + Stable	FC/PC + Unstable	41
NC + Stable	FC/PC + Unstable	104
NR + Stable	FC/PC + Unstable	0
Total		168

- Statistics of prefixes with two Origin ASes where the primary path is stable (with or without consistency in the registry), while the secondary (failover) path is transient (unstable) but consistent in the registry

Analysis of Registered But Unobserved Routes

- Large number of {prefix, origin} pairs registered but never announced
- In most cases, super-prefixes are announced with the same origin AS (as in registered route) or a different origin AS
- Is it due to aggregation by a higher tier ISP?



Conclusions and Planned Future Work

- Enhanced hybrid algorithm – history and registry data have complementary influence on improvement in origin validation
- Some ***caveats*** apply in the reported results (To do list)
  - Consideration of new NetHandle format in ARIN which includes origin AS information
  - Consideration of multiple trace-data collectors
- Further testing for robustness of the algorithms will be performed with extensive real and synthetic trace data
- Help industry understand implications of proposals emerging from various ongoing R&D projects

Thank you!

Questions?