APPLE: Alias Pruning by Path Length Estimation

Alex Marder

UC San Diego / CAIDA

Overview

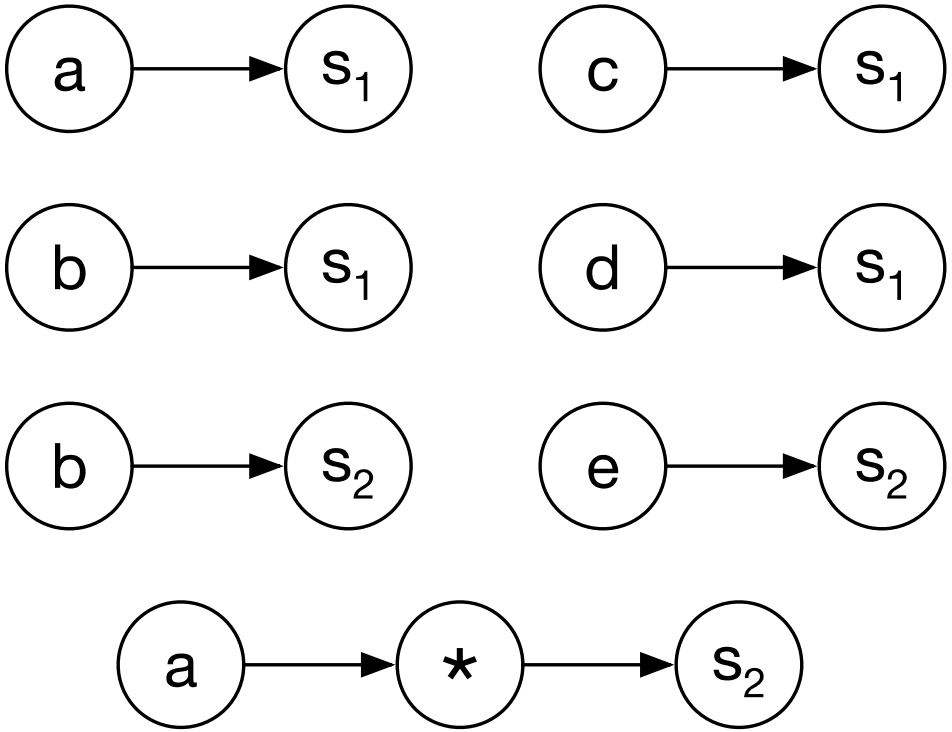
1. What is alias resolution and why do we need it?
2. Current state-of-the-art
3. APPLE methodology
   A. Common successor groups
   B. Prune with pings from many vantage points (VPs)
4. Results from two ground truth networks
5. Conclusion

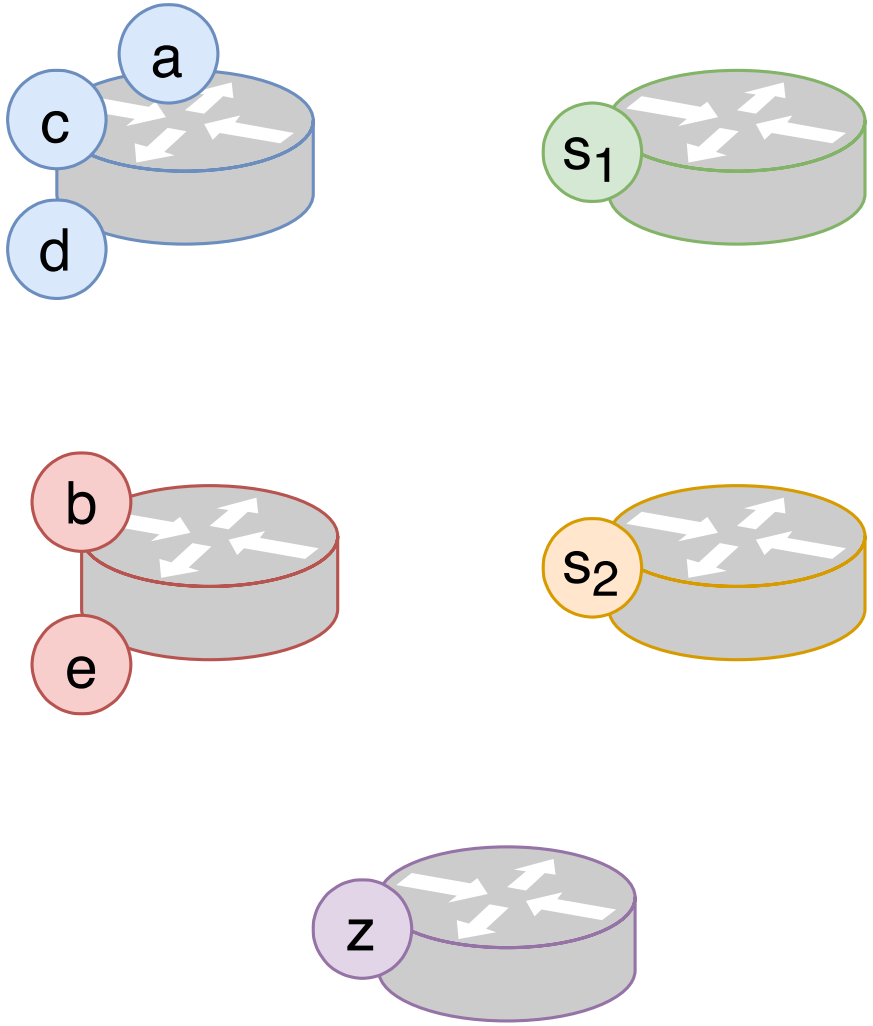
Overview

1. **What is alias resolution and why do we need it?**
2. Current state-of-the-art
3. APPLE methodology
   A. Common successor groups
   B. Prune with pings from many vantage points (VPs)
4. Results from two ground truth networks
5. Conclusion

What is Alias Resolution?

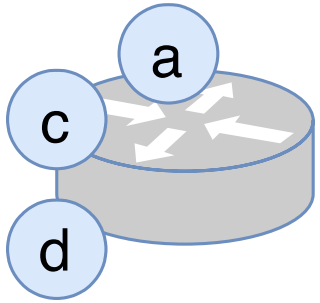
- Grouping addresses by their physical router

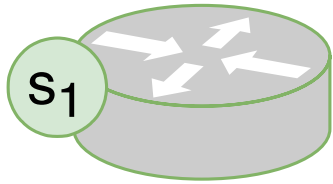




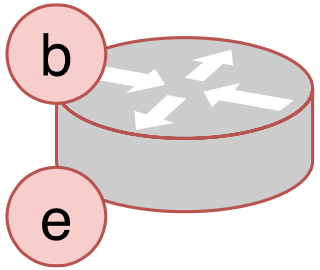
Why is Alias Resolution Important?

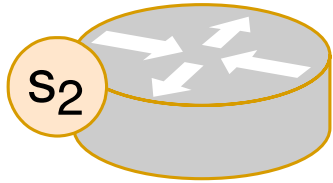
- **Traceroute provides interface addresses**
  - **No router identifiers**

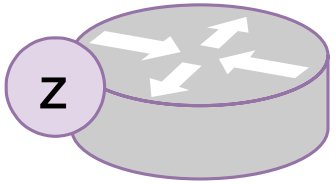




- Alias resolution helps provide router graph
  - More complete view of the network

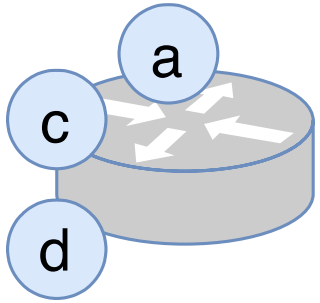


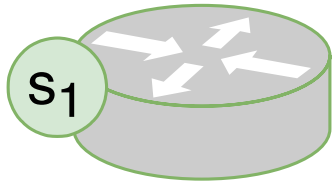




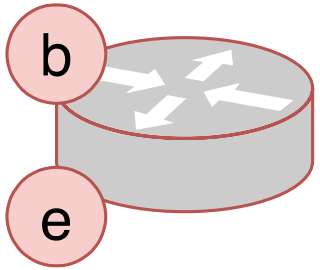
Why is Alias Resolution Important?

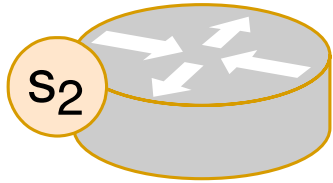
- Traceroute provides interface addresses
  - No router identifiers

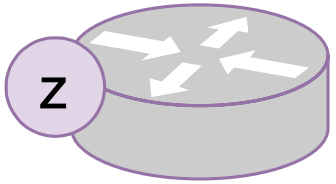




- **Alias resolution helps provide router graph**
  - **More complete view of the network**





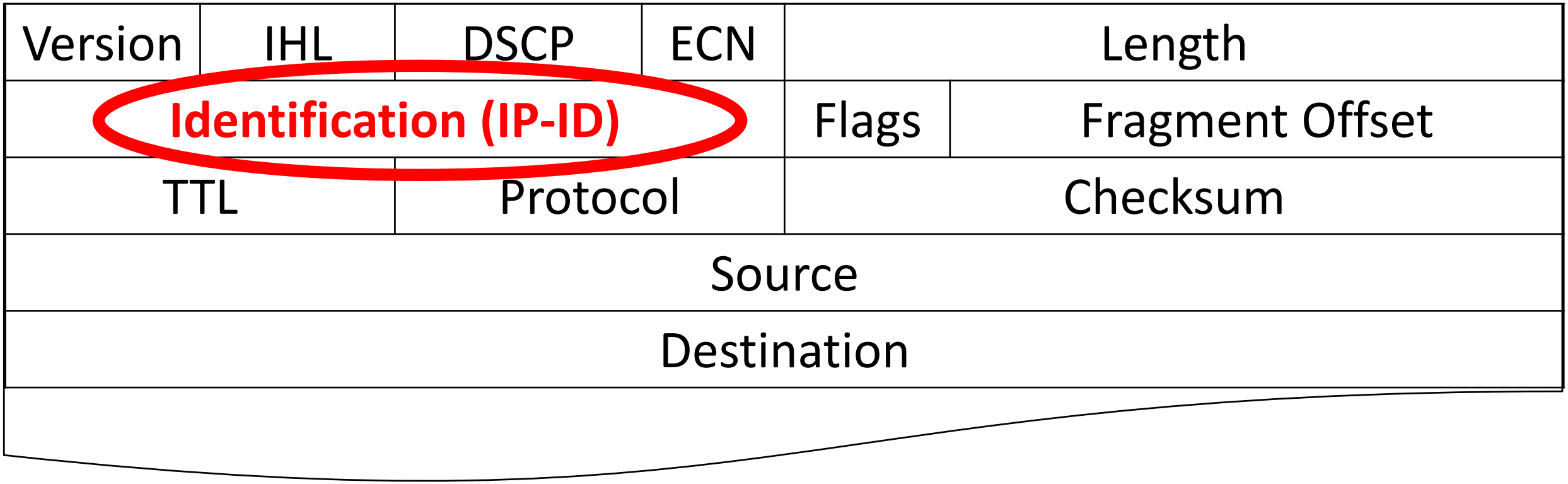


Overview

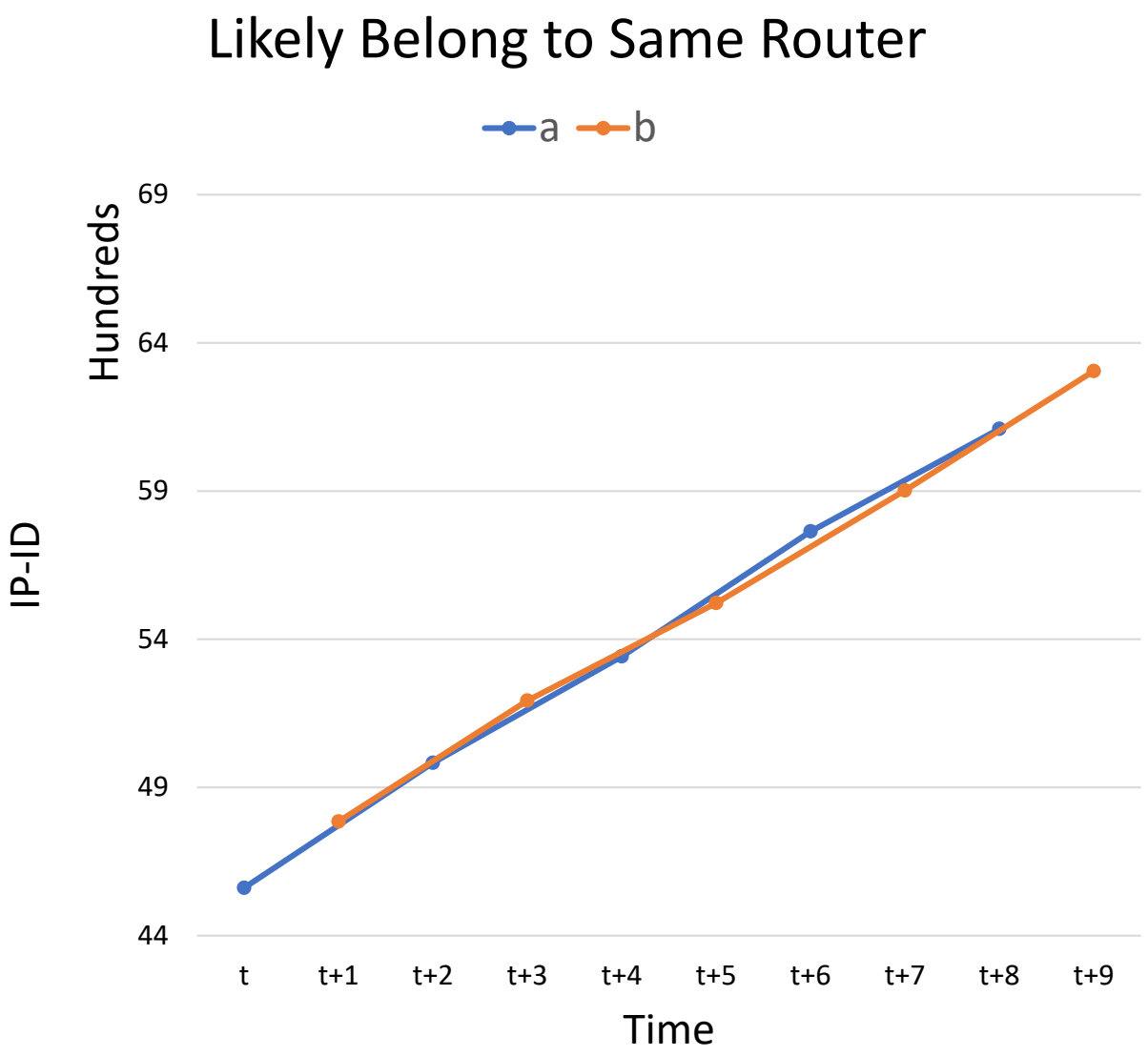
1. What is alias resolution and why do we need it?
2. **Current state-of-the-art**
3. APPLE methodology
   A. Common successor groups
   B. Prune with pings
4. Results
5. Conclusion

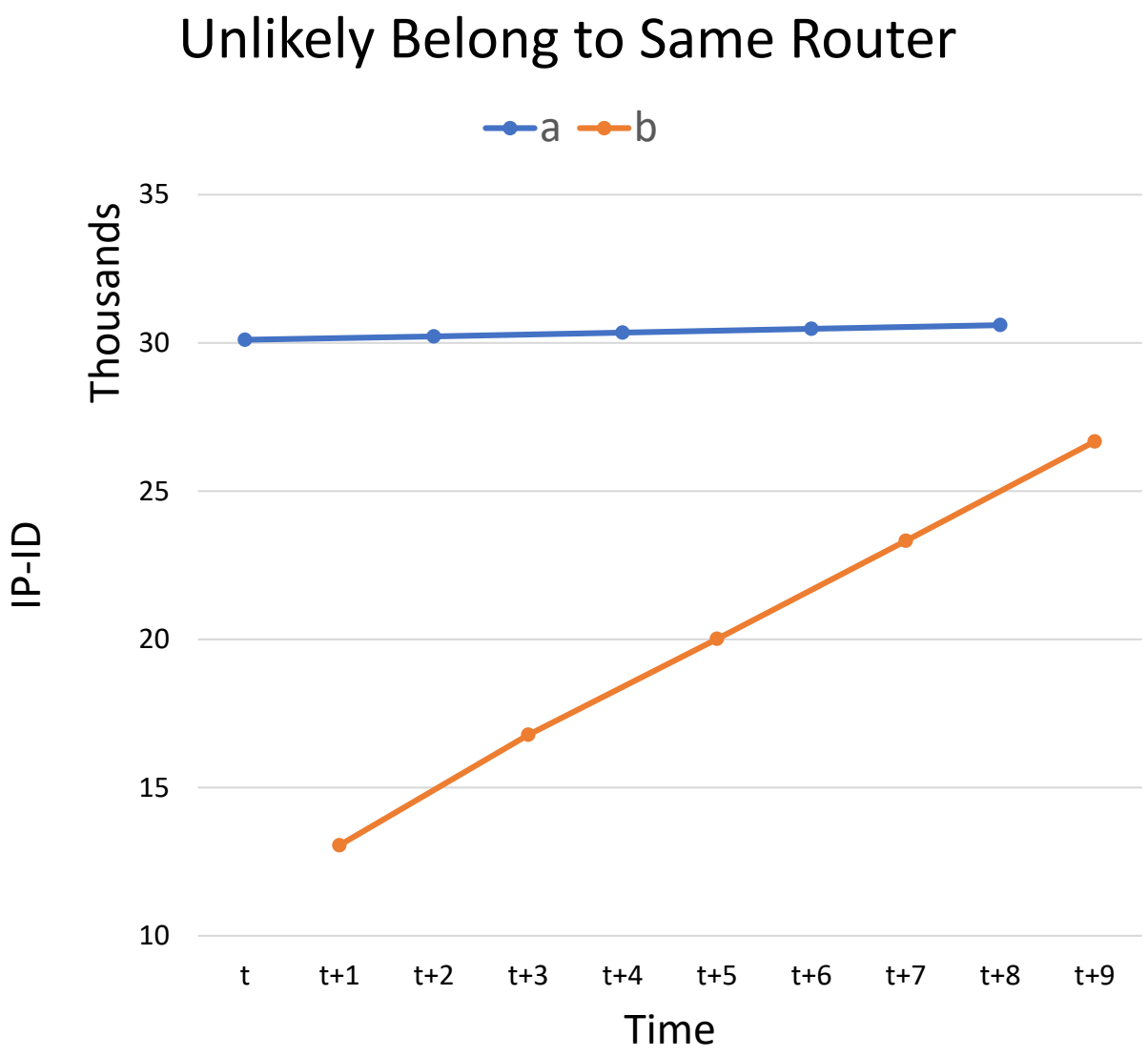
Current State-of-the-Art: IP-ID Velocity

IPv4 Packet Header



Current State-of-the-Art: IP-ID Velocity





Current State-of-the-Art: `hoiho` Regex Learning [1]

- Uses IP-ID router groupings as training

- Automatically learn regexes for router IDs in hostnames

- ^[a-z]+-[^\.]+\..+\.([a-z\d]+\.net)\.internet2\.edu$
  - et-9-1-0.4079.rtsw.**phil.net**.internet2.edu -> **phil.net**
  - et-11-0-0.103.rtsw.**phil.net**.internet2.edu -> **phil.net**
  - ae-3.4079.rtsw.**dall3.net**.internet2.edu -> **dall3.net**

[1] Luckie, M., Huffaker, B., et al.: Learning regexes to extract router names from hostnames. In: IMC (2019)

Current State-of-the-Art: `hoiho` Regex Learning [1]

- **Uses IP-ID router groupings as training**

- Automatically learn regexes for router IDs in hostnames

- ^[a-z]+-[^\.]+\..+\.([a-z\d]+\.net)\.internet2\.edu$
  - et-9-1-0.4079.rtsw.**phil.net**.internet2.edu -> **phil.net**
  - et-11-0-0.103.rtsw.**phil.net**.internet2.edu -> **phil.net**
  - ae-3.4079.rtsw.**dall3.net**.internet2.edu -> **dall3.net**

[1] Luckie, M., Huffaker, B., et al.: Learning regexes to extract router names from hostnames. In: IMC (2019)

Current State-of-the-Art: `hoiho` Regex Learning [1]

- Uses IP-ID router groupings as training

- **Automatically learn regexes for router IDs in hostnames**

- ^[a-z]+-[^\.]+\..+\.([a-z\d])+\.net\.internet2\.edu$
  - et-9-1-0.4079.rtsw.**phil**.net.internet2.edu -> **phil**
  - et-11-0-0.103.rtsw.**phil**.net.internet2.edu -> **phil**
  - ae-3.4079.rtsw.**dall3**.net.internet2.edu -> **dall3**

[1] Luckie, M., Huffaker, B., et al.: Learning regexes to extract router names from hostnames. In: IMC (2019)

Do We Need More?

- IP-ID techniques rely on specific IP implementations
  - IPv6 is even worse
- Cannot reverse lookup many addresses
  - Others lack router identifiers
- No one technique to rule them all

Do We Need More?

- **IP-ID techniques rely on specific IP implementations**
  - IPv6 is even worse
- Cannot reverse lookup many addresses
  - Others lack router identifiers
- No one technique to rule them all

Do We Need More?

- IP-ID techniques rely on specific IP implementations
  - **IPv6 is even worse**
- Cannot reverse lookup many addresses
  - Others lack router identifiers
- No one technique to rule them all

Do We Need More?

- IP-ID techniques rely on specific IP implementations
  - IPv6 is even worse

- **Cannot reverse lookup many addresses**
  - **Others lack router identifiers**

- No one technique to rule them all

Do We Need More?

- IP-ID techniques rely on specific IP implementations
  - IPv6 is even worse
- Cannot reverse lookup many addresses
  - Others lack router identifiers
- **No one technique to rule them all**

Overview

1. What is alias resolution and why do we need it?
2. Current state-of-the-art
3. **APPLE methodology**
   A. Common successor groups
   B. Prune with pings from many vantage points (VPs)
4. Results from two ground truth networks
5. Conclusion

APPLE Methodology

- Goal: Reduce reliance on specific router implementations of IP
- Step 1: collect candidate router alias groups from traceroute
- Step 2: prune the groups using pings

APPLE Methodology

- **Goal: Reduce reliance on specific implementations of IP**

- Step 1: collect candidate router alias groups from traceroute

- Step 2: prune the groups using pings from many VPs

APPLE Methodology

- Goal: Reduce reliance on specific router implementations of IP
- **Step 1: collect candidate router alias groups from traceroute**
- Step 2: prune the groups using pings from many VPs

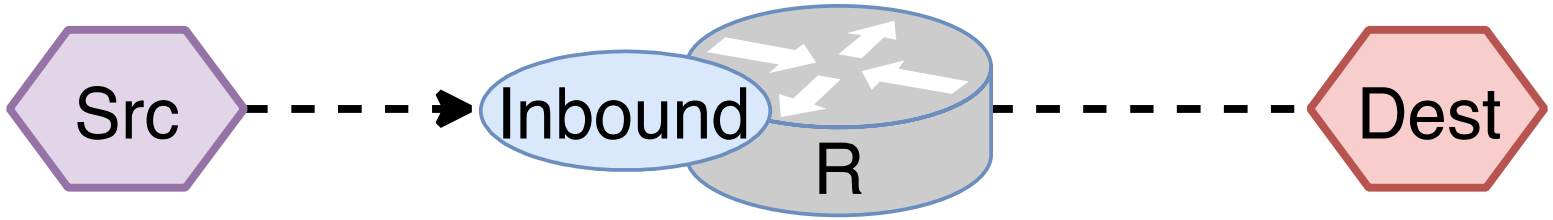
APPLE Methodology

- Goal: Reduce reliance on specific router implementations of IP

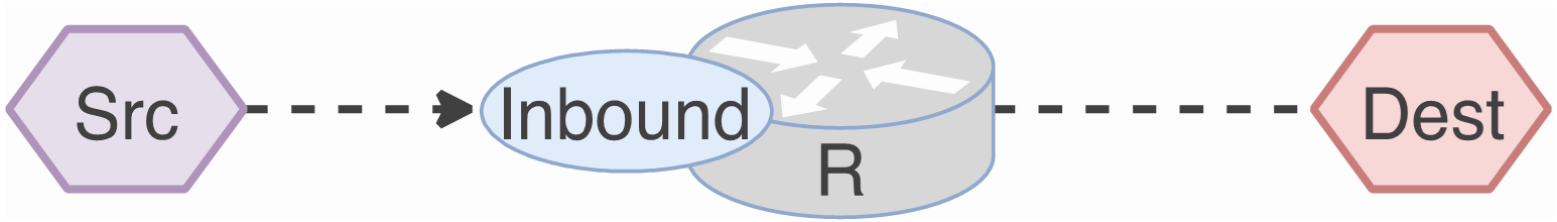
- Step 1: collect candidate router alias groups from traceroute

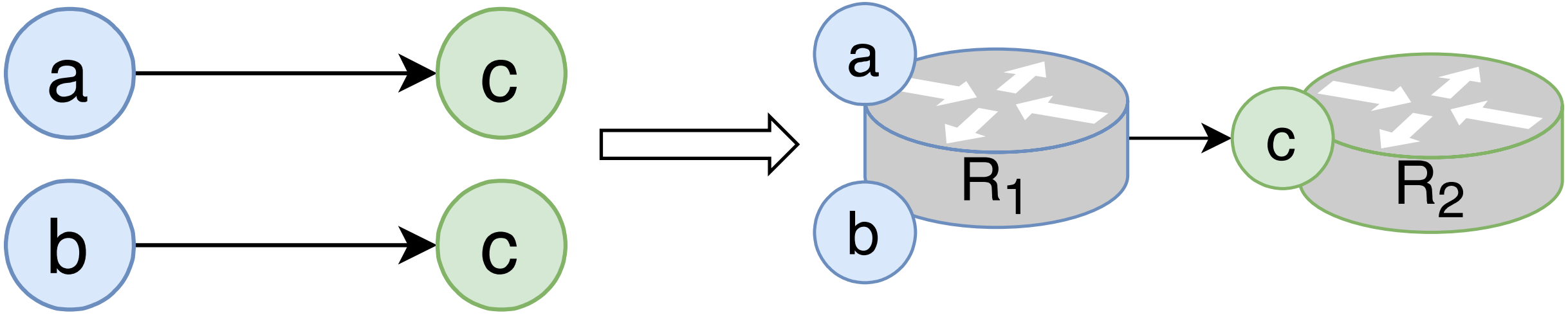
- **Step 2: prune the groups using pings from many VPs**

Traceroute Responses Provide Inbound Address

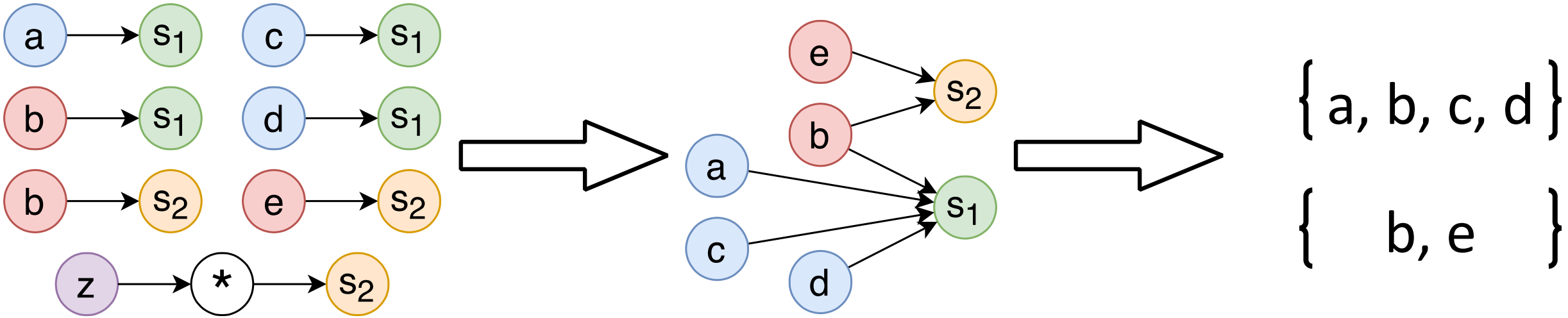


Common Successors Provide Alias Hints



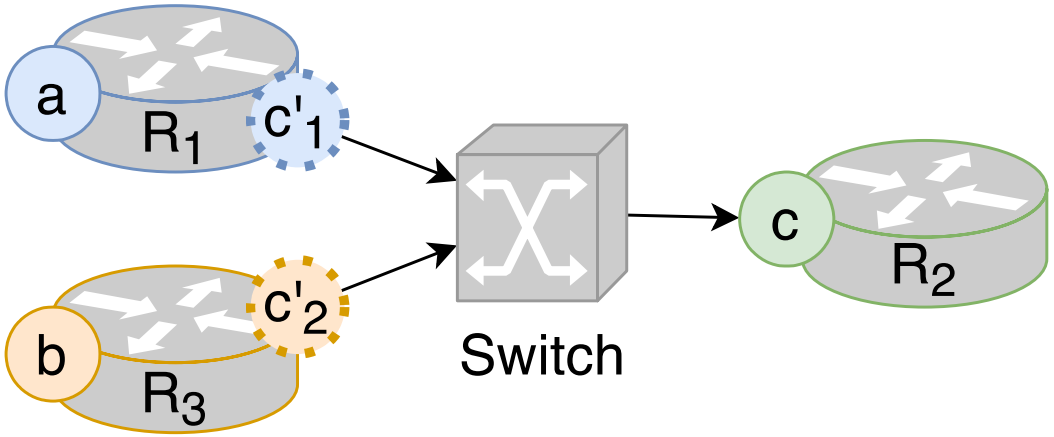


Step 1: Build Common Successor Sets

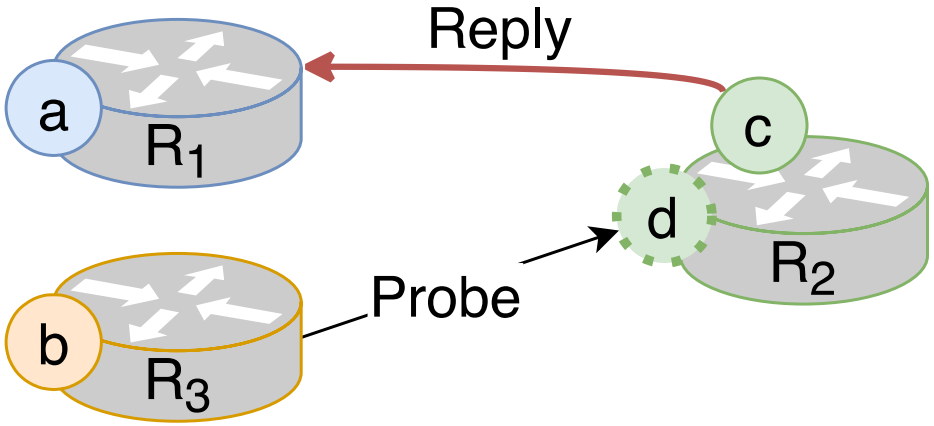


Well Known Problems With Common Successor

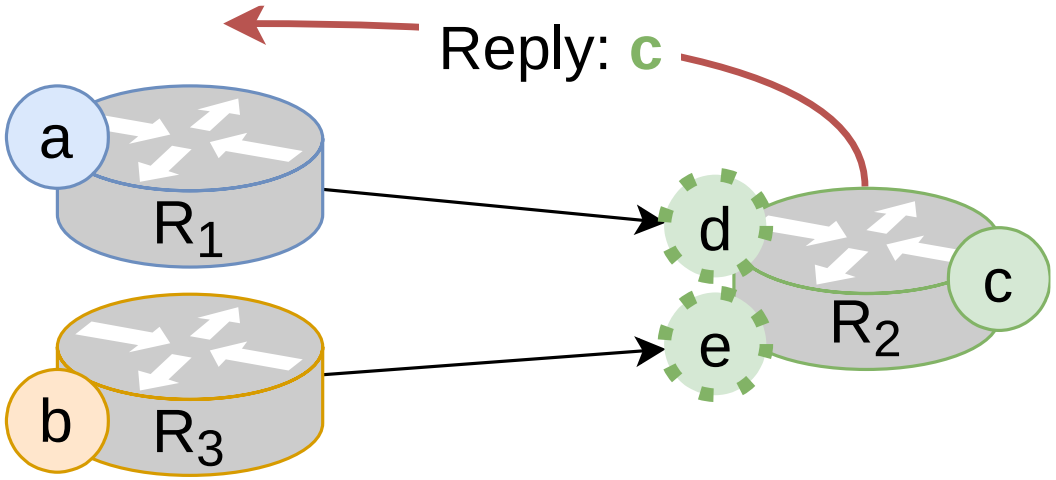
Multipoint-to-point Links



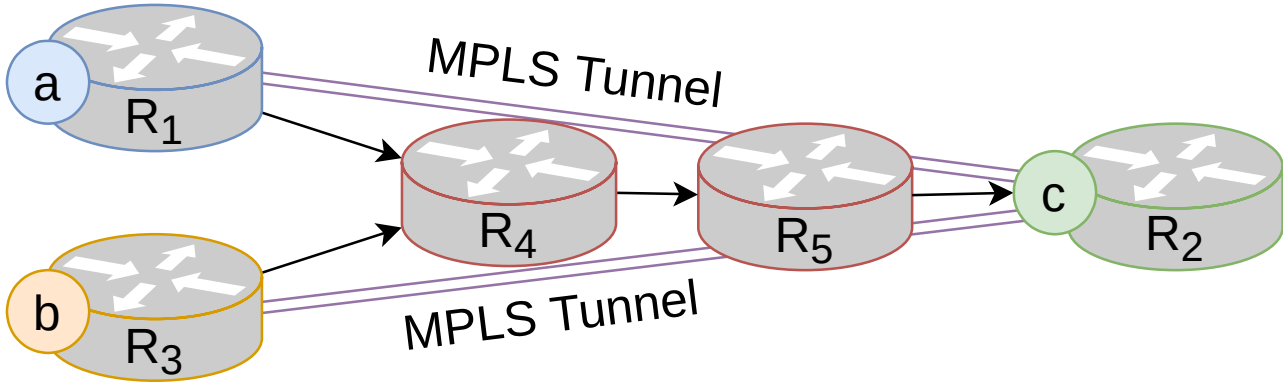
Off-Path Responses



L3VPN Outbound Addresses

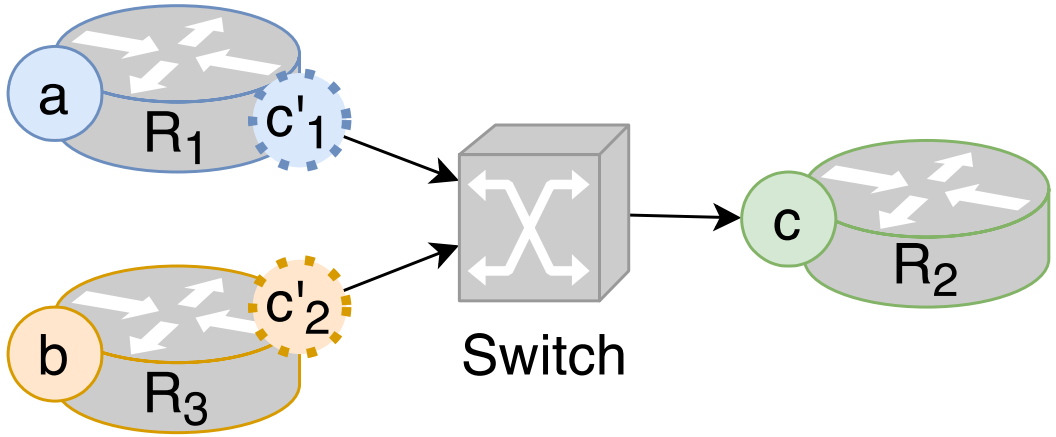


Invisible MPLS Tunnels

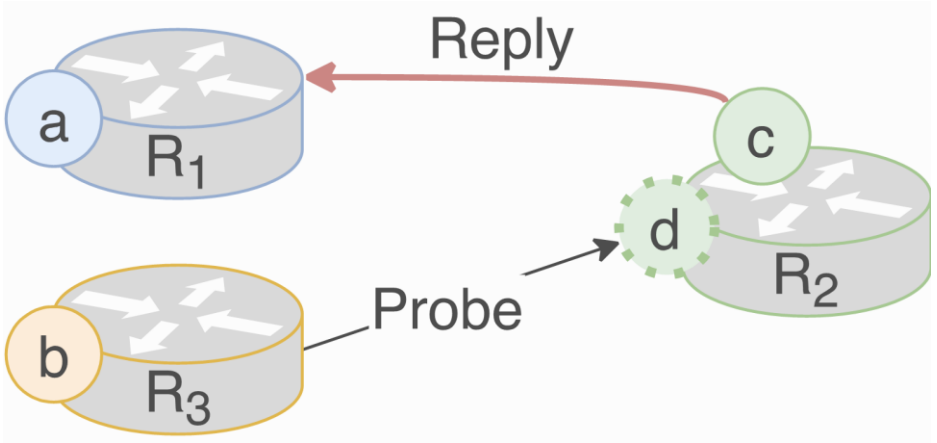


Well Known Problems With Common Successor

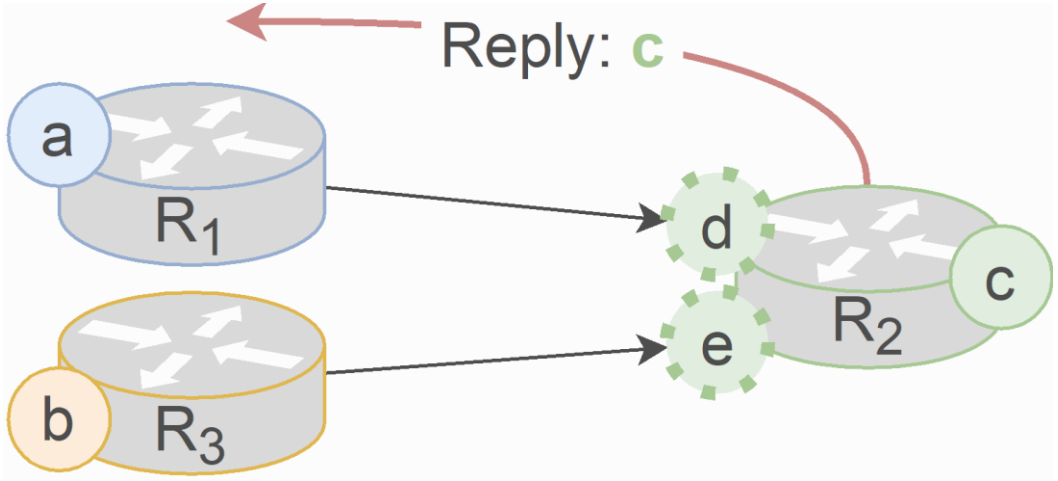
Multipoint-to-point Links



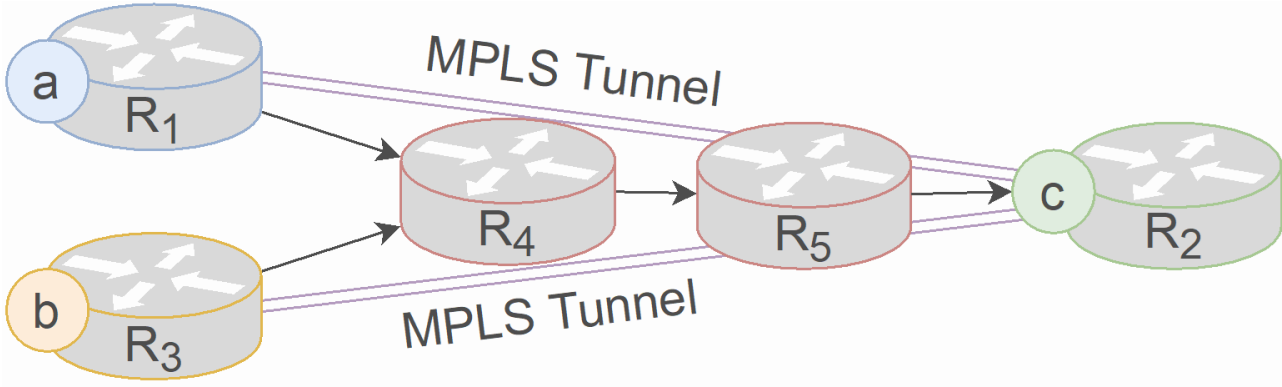
Off-Path Responses



L3VPN Outbound Addresses

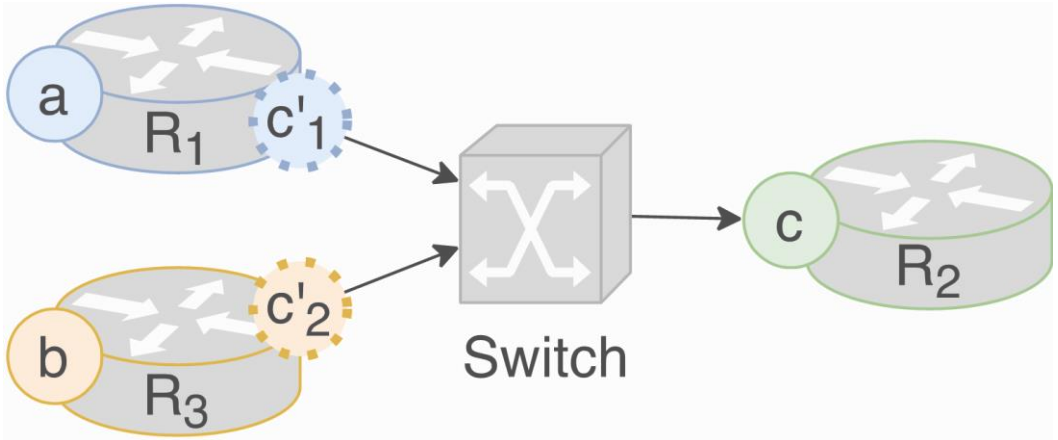


Invisible MPLS Tunnels

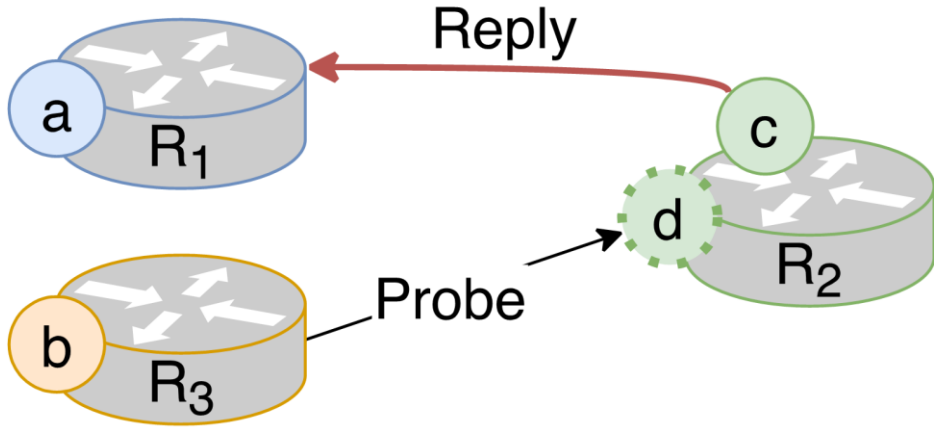


Well Known Problems With Common Successor

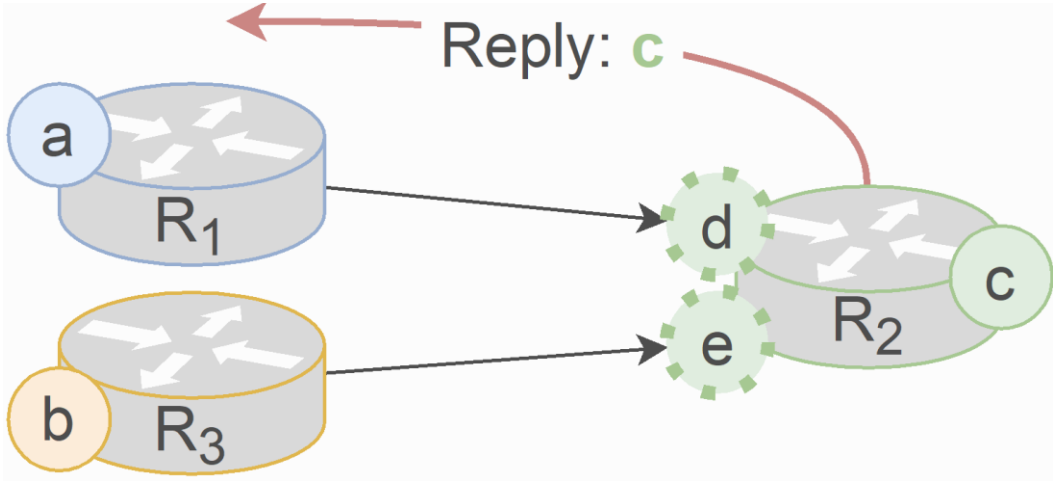
Multipoint-to-point Links



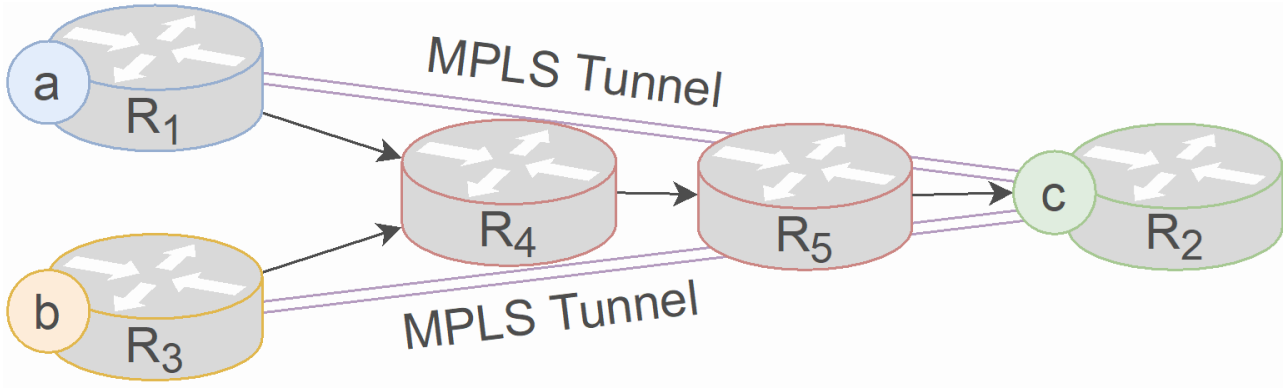
Off-Path Responses



L3VPN Outbound Addresses

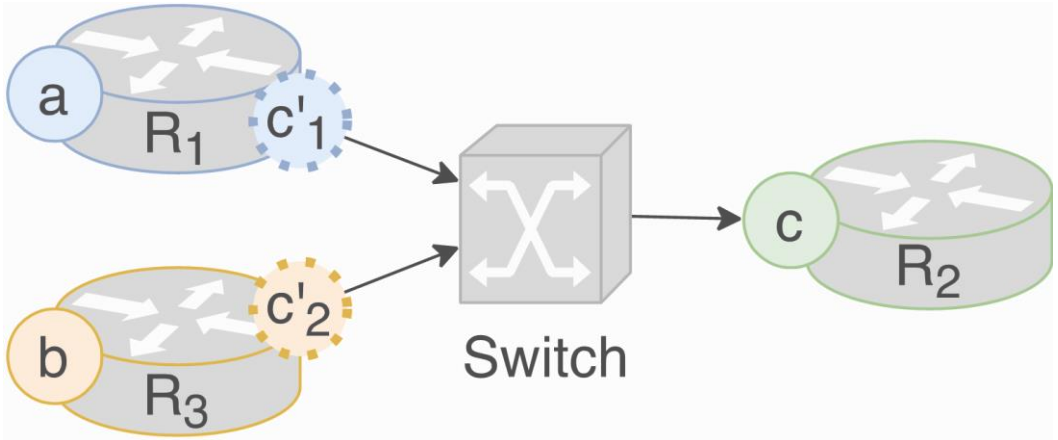


Invisible MPLS Tunnels

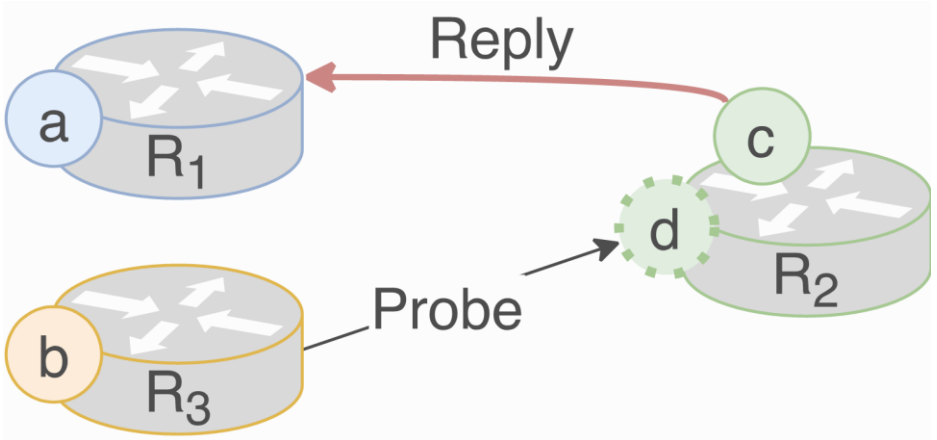


Well Known Problems With Common Successor

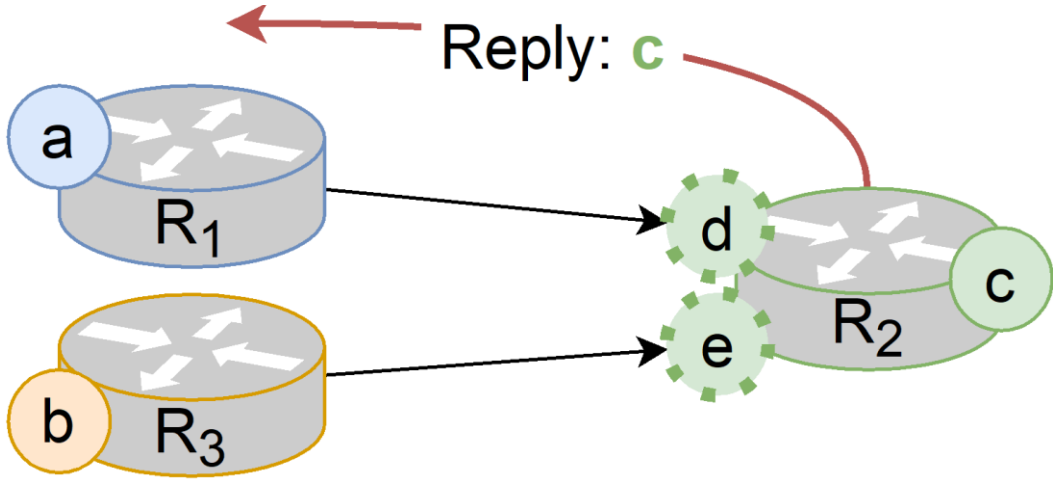
Multipoint-to-point Links



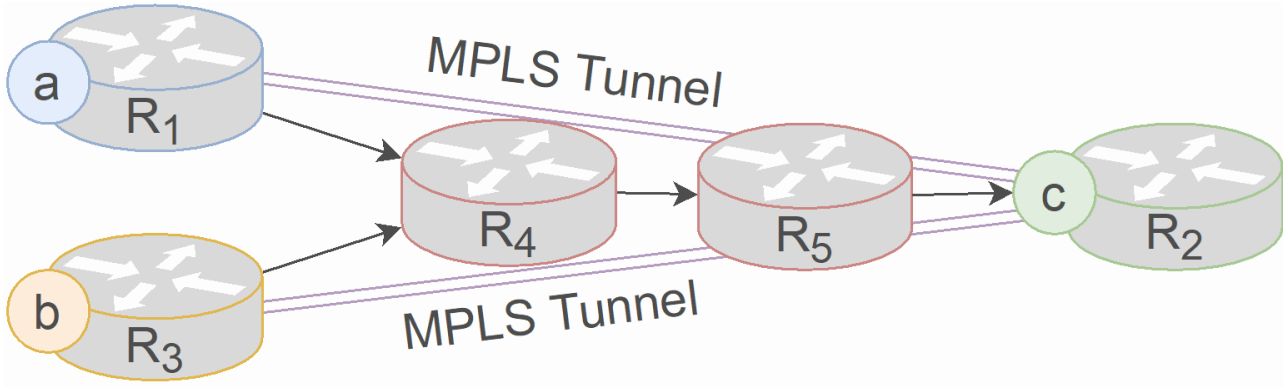
Off-Path Responses



L3VPN Outbound Addresses

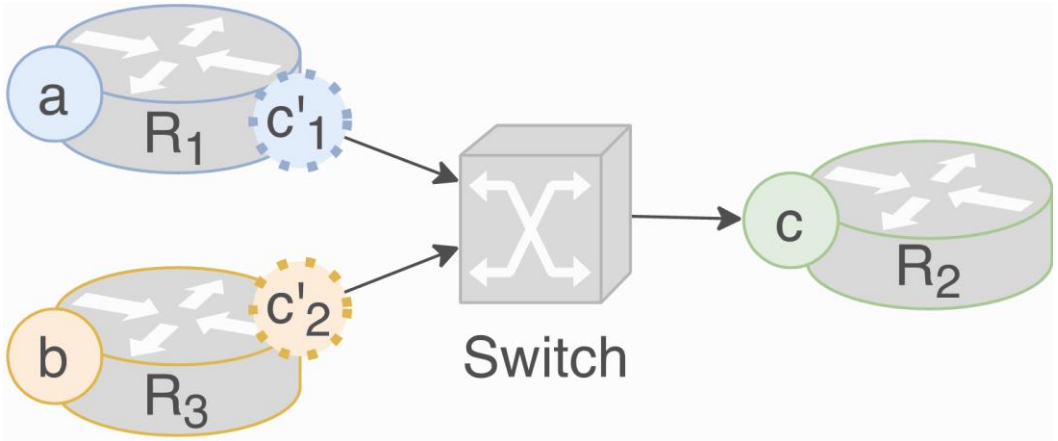


Invisible MPLS Tunnels

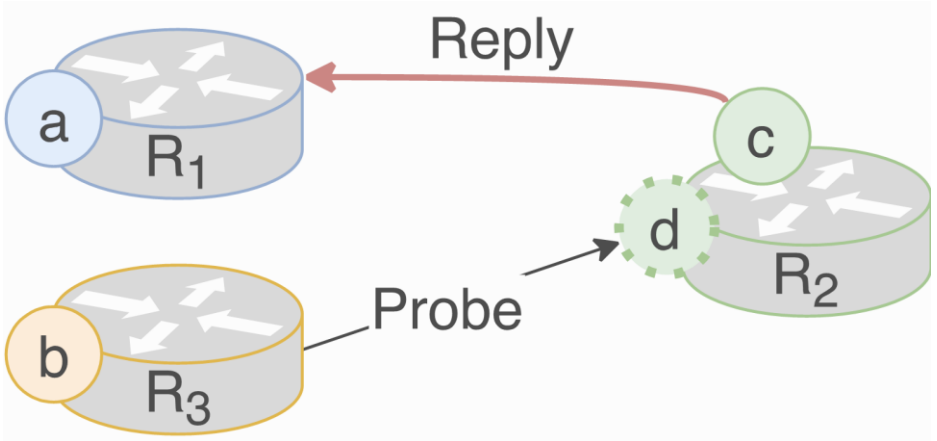


Well Known Problems With Common Successor

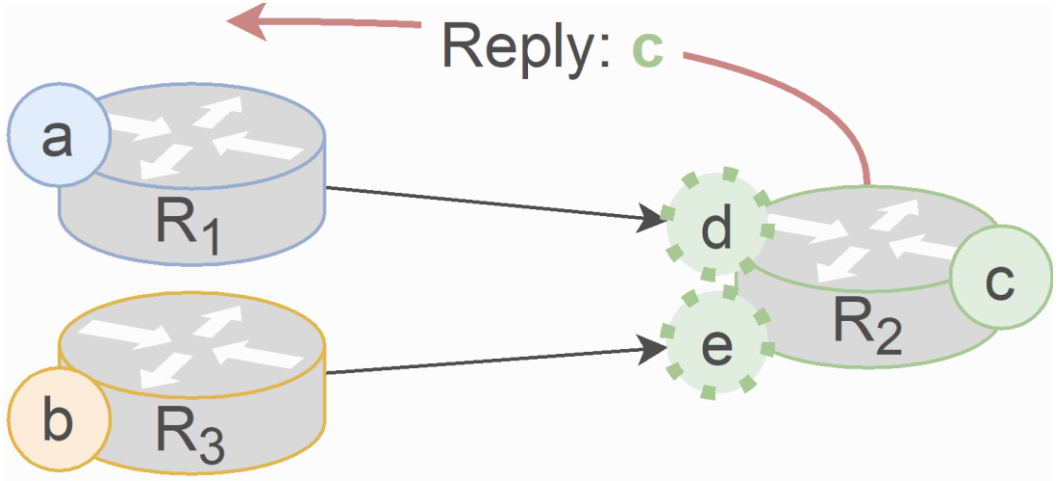
Multipoint-to-point Links



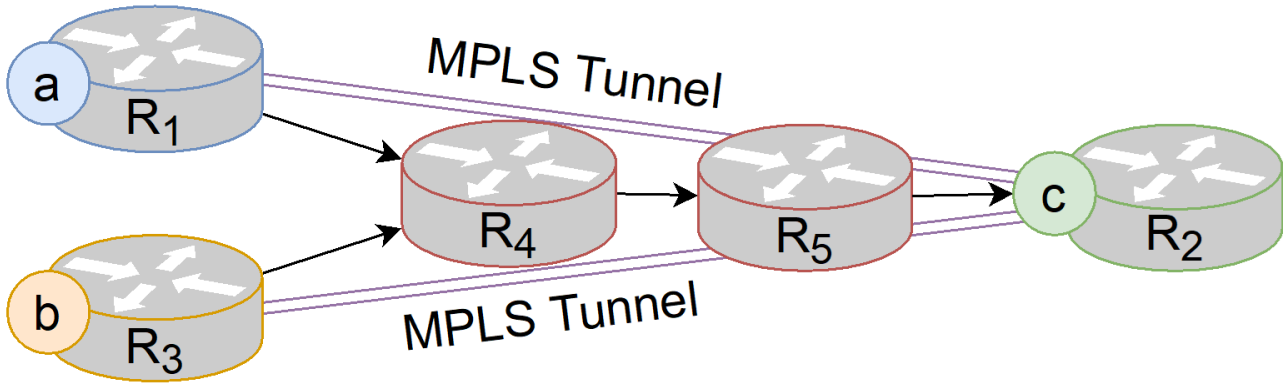
Off-Path Responses



L3VPN Outbound Addresses



Invisible MPLS Tunnels



APPLE Methodology

- Goal: Reduce reliance on specific router implementations of IP

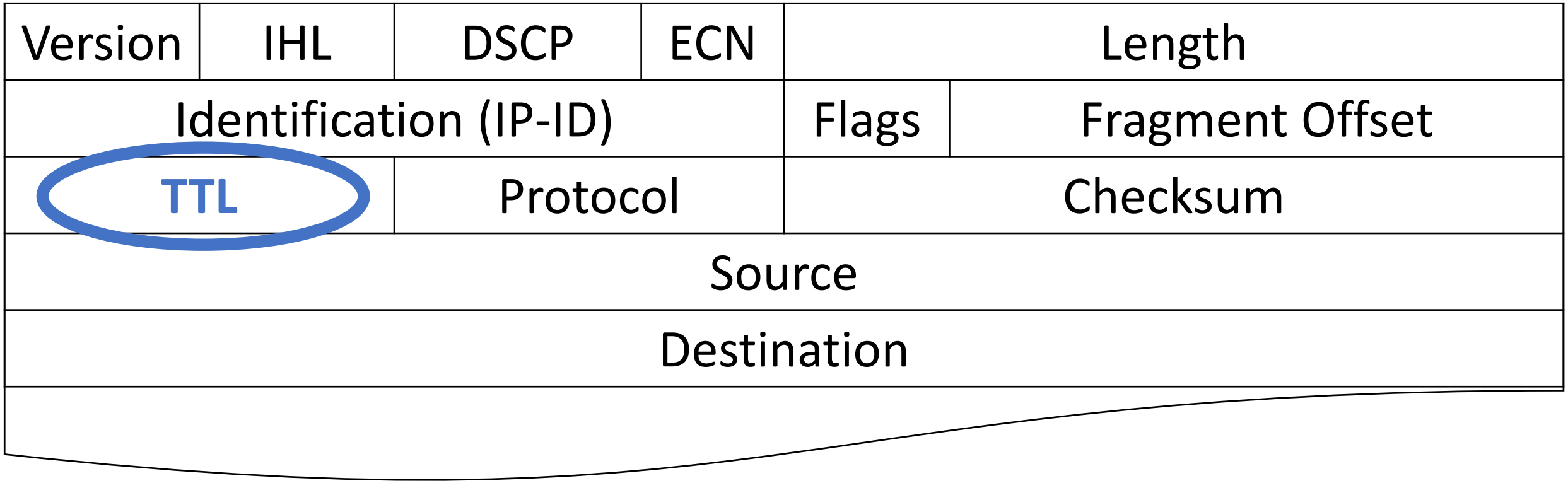
- Step 1: collect candidate router alias groups from traceroute

- **Step 2: prune the groups using pings from many VPs**

Step 2: Pruning

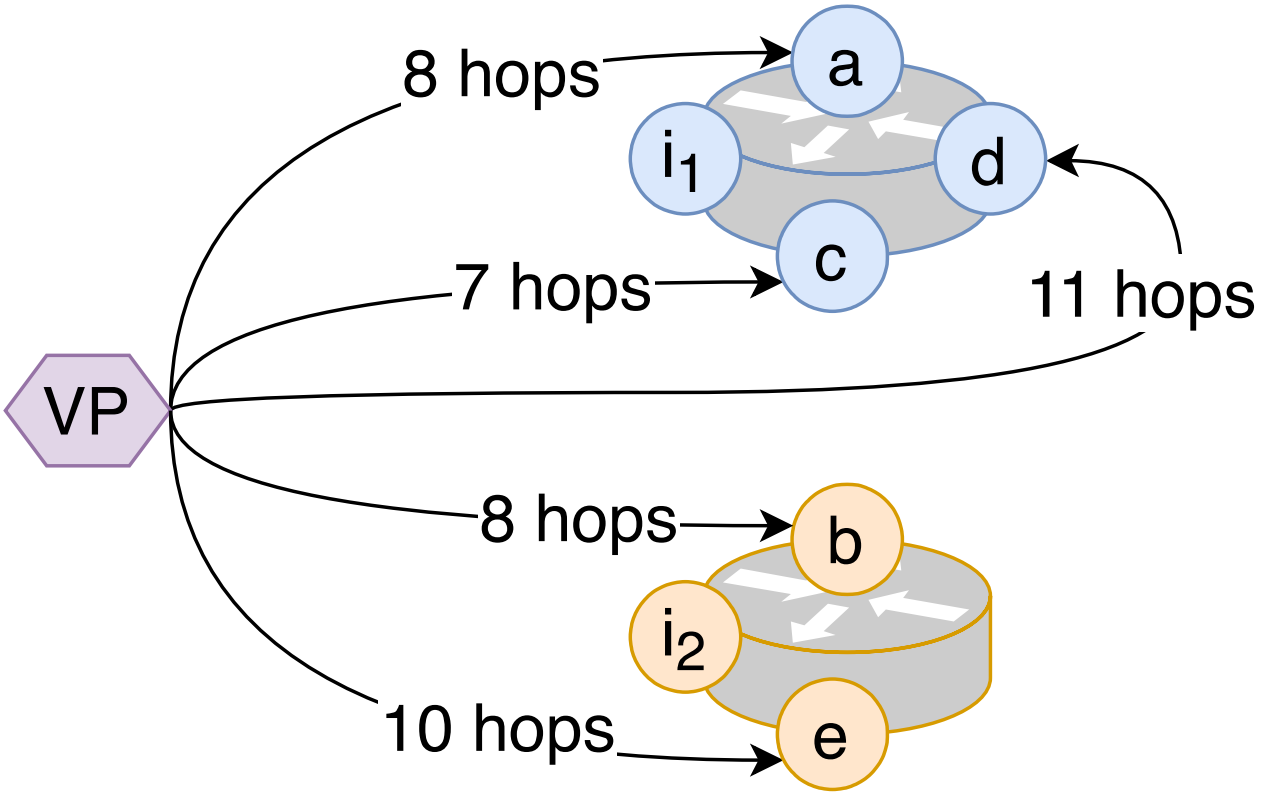
- Prune common successor groups using reply path length

IPv4 Packet Header

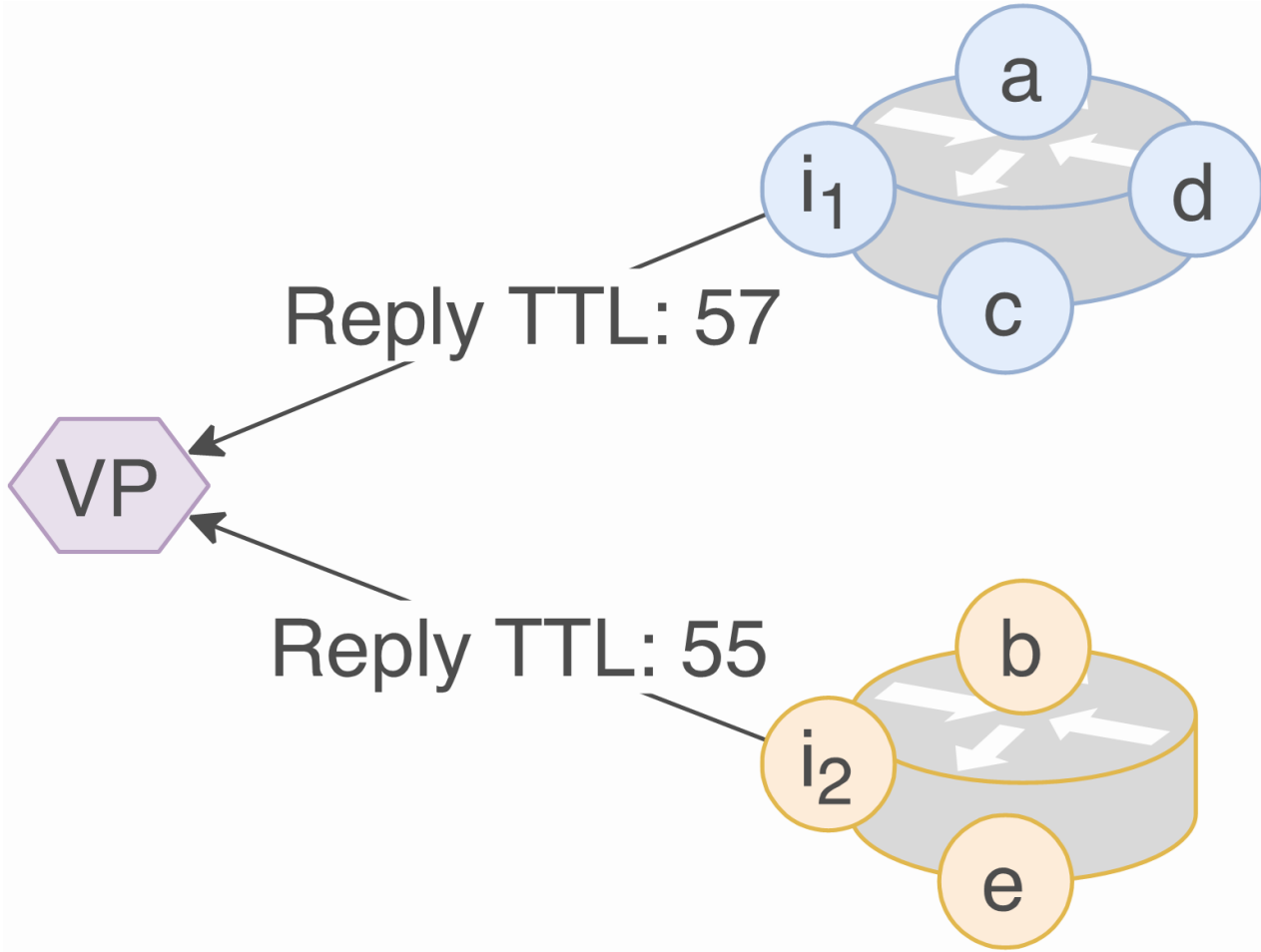


Intuition: Destination-Based Forwarding

Route to routers

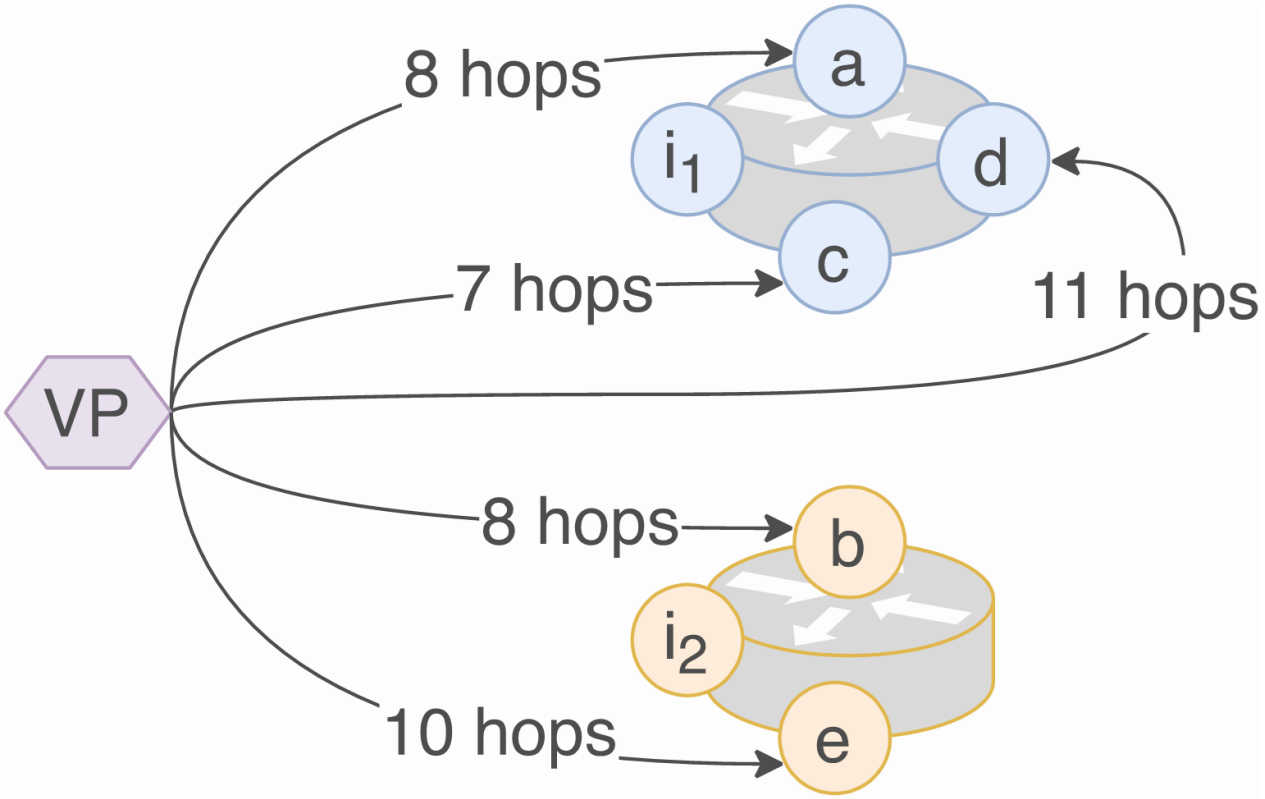


Route from routers

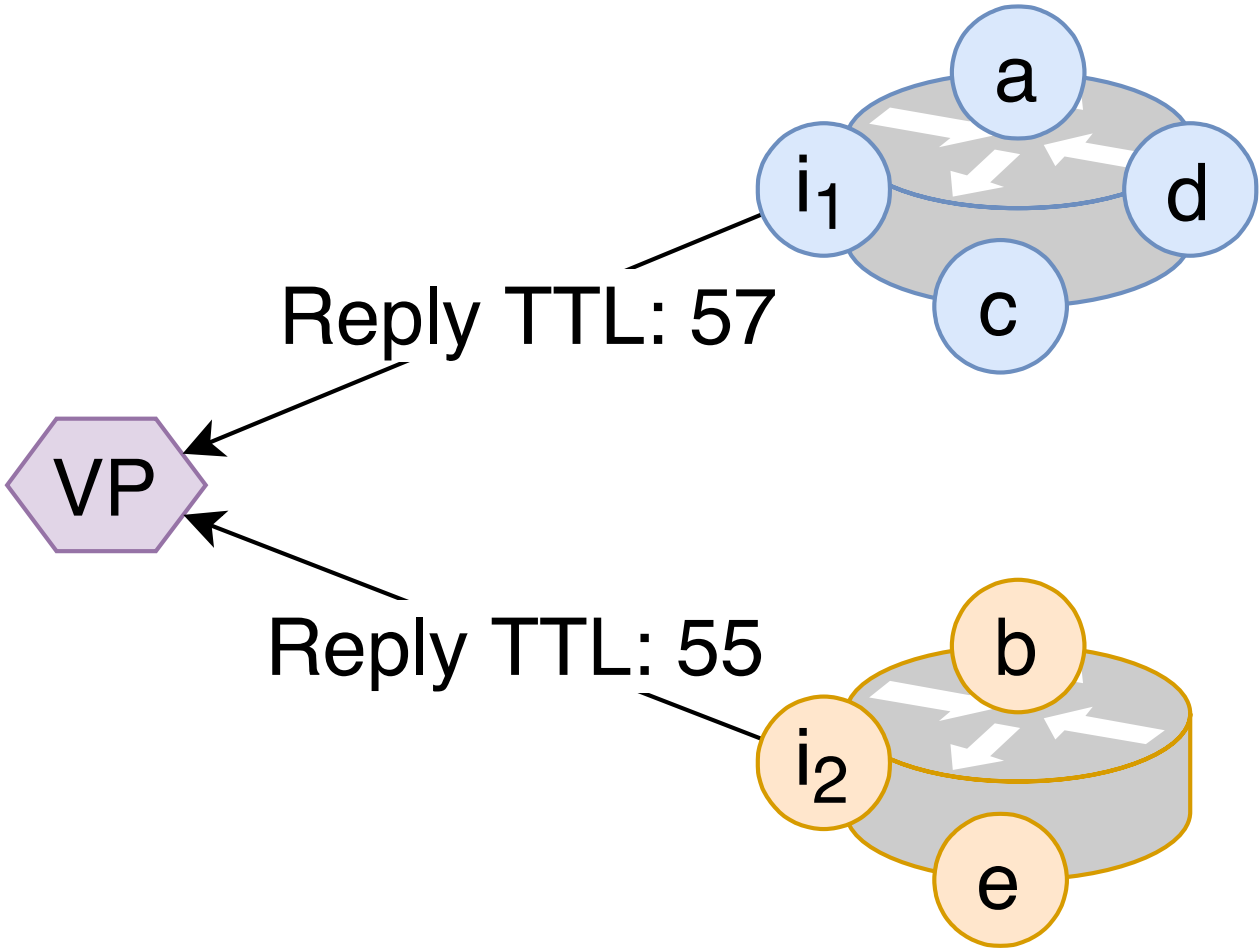


Intuition: Destination-Based Forwarding

Route to routers

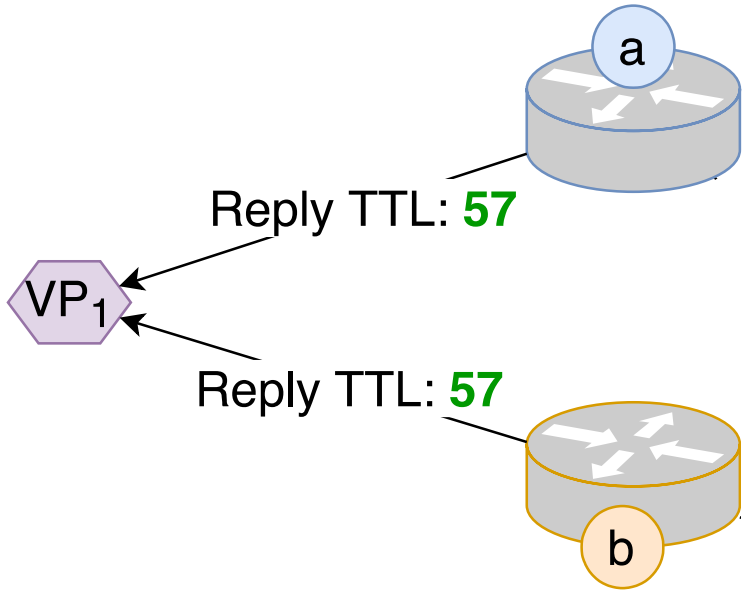


Route from routers



Random Reply TTL Collisions

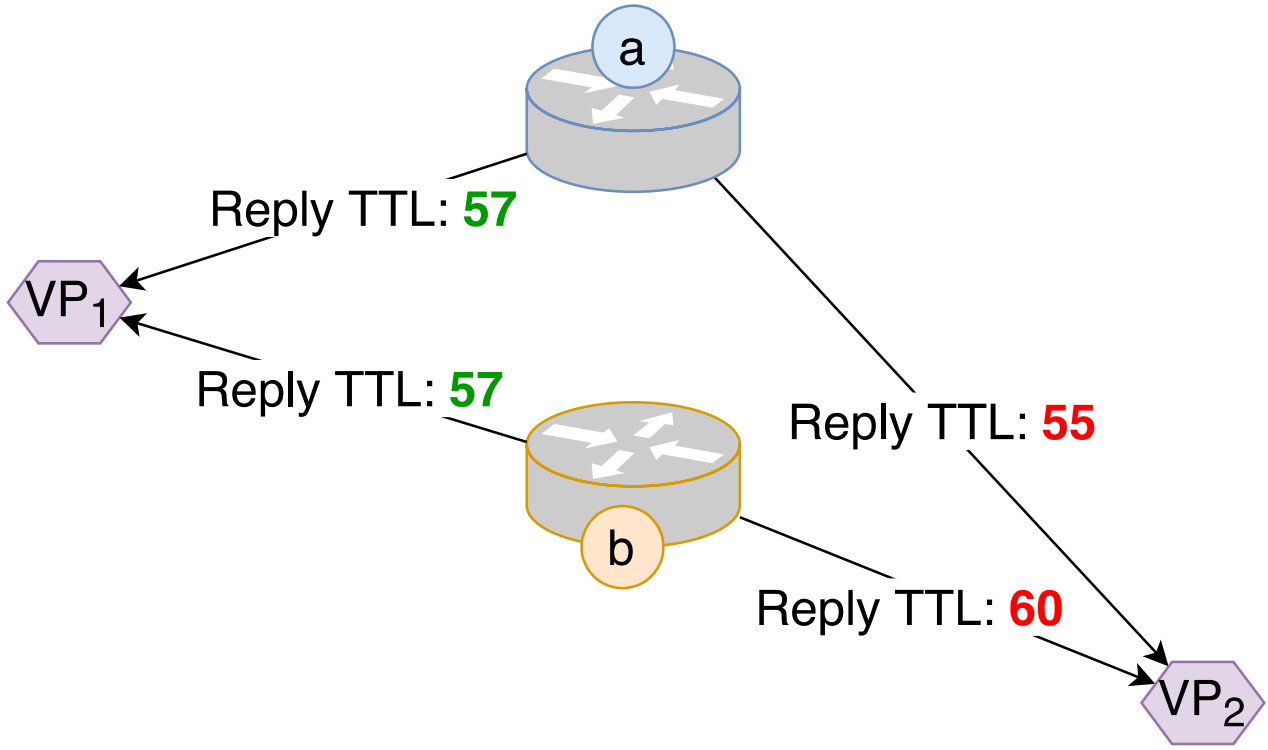
- **Finite range of TTL values**
  - **8-bit TTL field**
  - **256 possible values**



- More VPs reduces chances of collision

Random Reply TTL Collisions

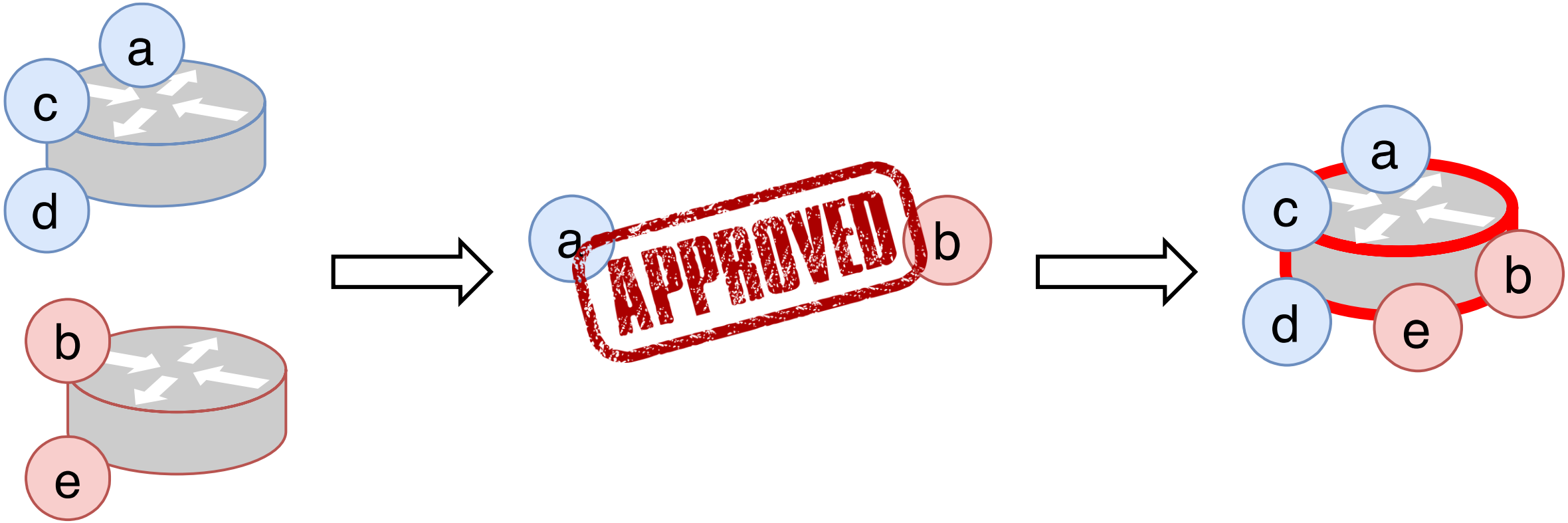
- Finite range of TTL values
  - 8-bit TTL field
  - 256 possible values



- **More VPs reduces chances of collision**

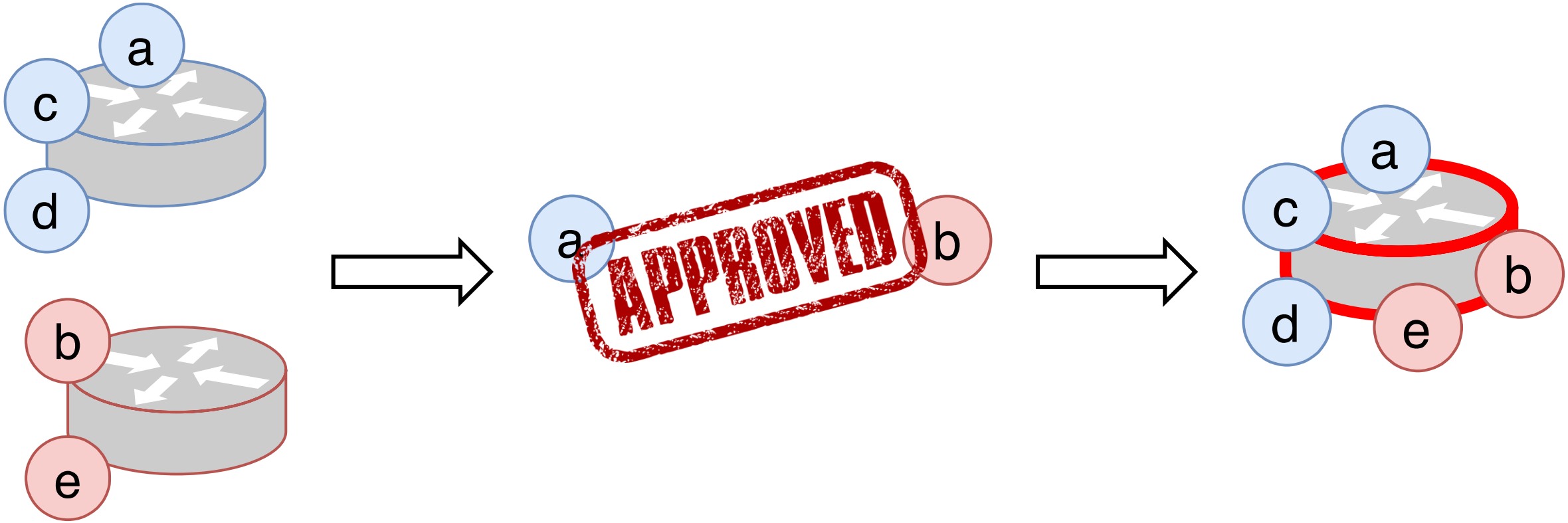
Coverage vs False Aliases

- **Tradeoff between including valid aliases and including false aliases**
- Better to exclude invalid aliases – alias resolution is transitive



Coverage vs False Aliases

- Tradeoff between including valid aliases and including false aliases
- **Better to exclude invalid aliases – alias resolution is transitive**



APPLE Thresholds Control the Tradeoff

1. *Minimum match threshold*: minimum number of matches required to keep alias pair

2. *Acceptance threshold*: minimum ratio of matches to comparisons

APPLE Thresholds Control the Tradeoff

1. ***Minimum match threshold*: minimum number of matches required to keep alias pair**

2. *Acceptance threshold*: minimum ratio of matches to comparisons

APPLE Thresholds Control the Tradeoff

1. *Minimum match threshold*: minimum number of matches required to keep alias pair

2. ***Acceptance threshold*: minimum ratio of matches to comparisons**

Minimum Match Threshold: Birthday Problem

- Probability that any 2 people in a group share a birthday

- $p(n, d) \approx 1 - \exp\left(\frac{-n(n-1)}{2d}\right)$
  - n – number of people
  - d – number of days

Minimum Match Threshold: Birthday Problem

- Probability that any 2 people in a group share a birthday

- $\boldsymbol{p(n,d) \approx 1 - \exp\left(\frac{-n(n-1)}{2d}\right)}$
  - n – number of people
  - d – number of days

Apply Birthday Solution to APPLE

$$p(n,d) \approx 1 - \exp\left(\frac{-\boldsymbol{n(n-1)}}{\boldsymbol{2}d}\right)$$

$\frac{n(n-1)}{2}$: people pairs to compare

Apply Birthday Solution to APPLE

$$p(a, d) \approx 1 - \exp\left(\frac{-a}{d}\right)$$

a: number of potential alias pairs

Apply Birthday Solution to APPLE

$$p(a, d) \approx 1 - \exp\left(\frac{-a}{d}\right)$$

a: number of potential alias pairs

d: number of days

Apply Birthday Solution to APPLE

$$p(a, r, v) \approx 1 - \exp\left(\frac{-a}{r^v}\right)$$

a: number of potential alias pairs

r: number of possible reply TTL values

v: number of vantage points

Apply Birthday Solution to APPLE

$$p(v) \approx 1 - \exp\left(\frac{-a}{r^v}\right) < \frac{1}{a}$$

v: number of vantage points

Set the minimum match to limit the probability of a random collision to < 1/a

Acceptance Threshold

- Order VPs by minimum RTT to potential alias pair addresses
- Compare reply TTLs until finding required number of matches
- Is $\frac{\text{matches}}{\text{comparisons}}$ >= acceptance threshold?

Acceptance Threshold

- **Order VPs by minimum RTT to potential alias pair addresses**
- Compare reply TTLs until finding required number of matches
- Is $\frac{\text{matches}}{\text{comparisons}}$ >= acceptance threshold?

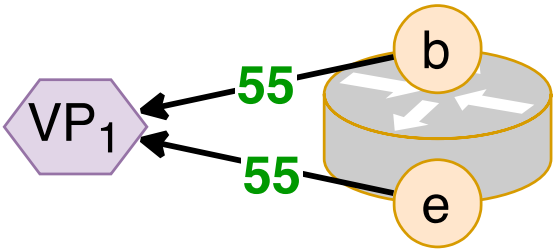
Acceptance Threshold

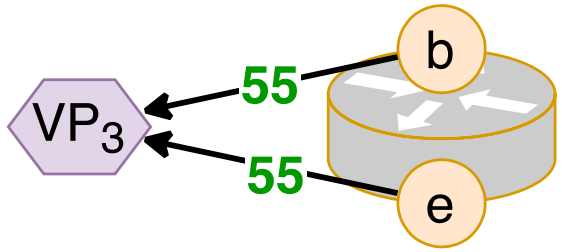
- Order VPs by minimum RTT to potential alias pair addresses
- **Compare reply TTLs until finding required number of matches**
- Is $\frac{\text{matches}}{\text{comparisons}}$ >= acceptance threshold?

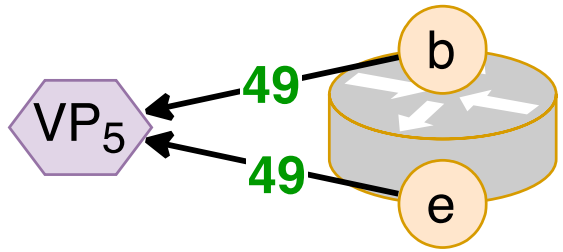
Acceptance Threshold

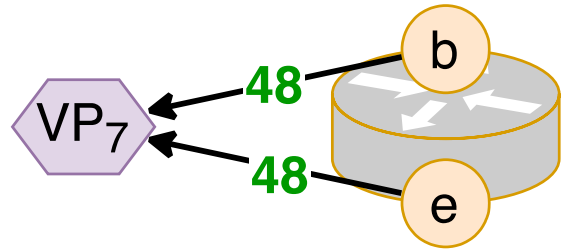
- Order VPs by minimum RTT to potential alias pair addresses
- Compare reply TTLs until finding required number of matches
- **Is $\frac{\text{matches}}{\text{comparisons}}$ >= acceptance threshold?**

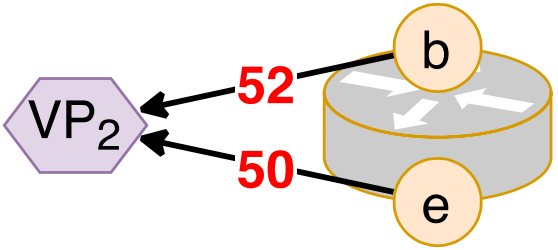
Ping Each Address From Every VP

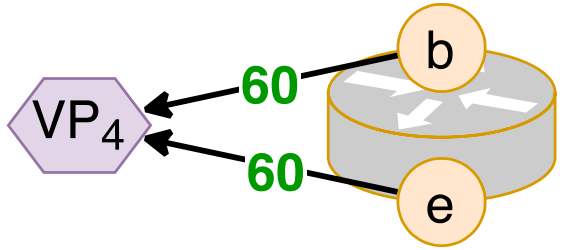


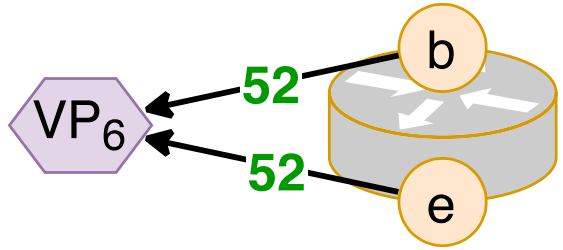


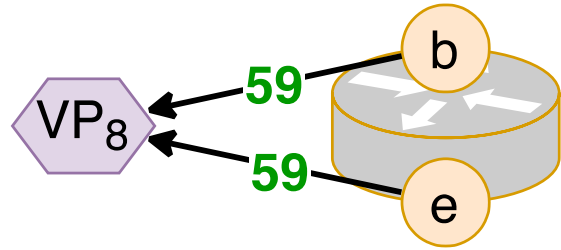






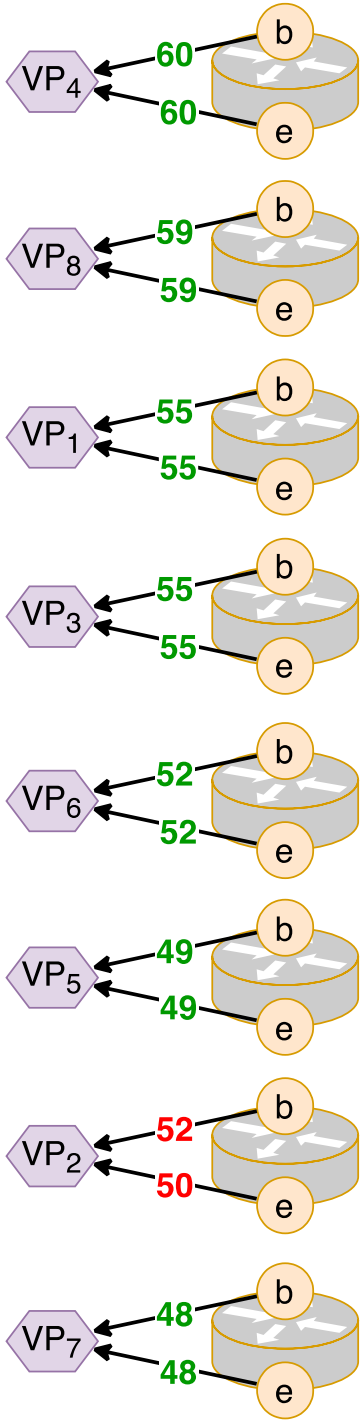






Example

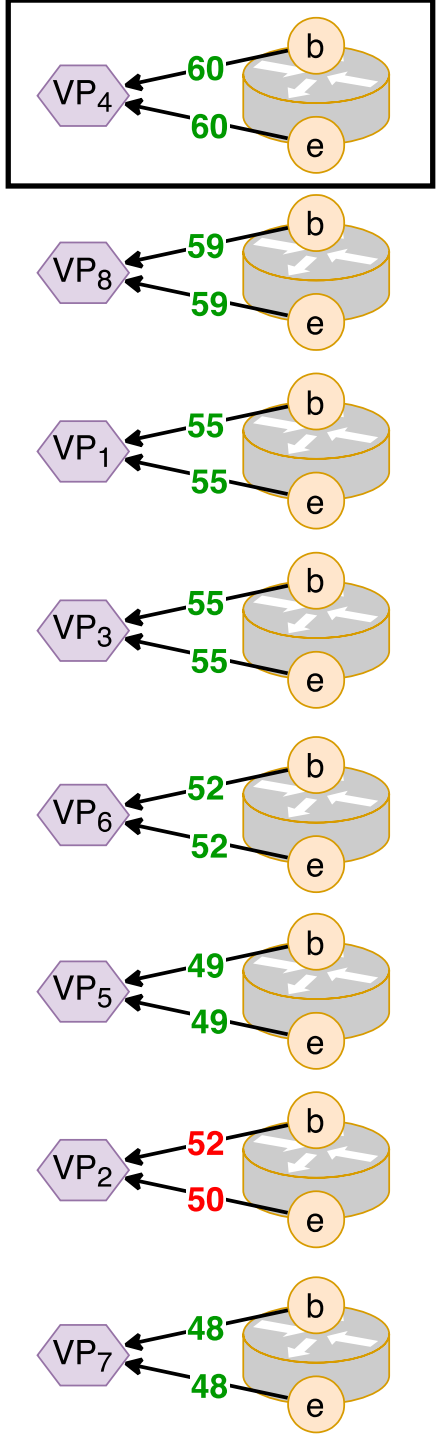
- Minimum match = 6
- Acceptance = 100%



Example

- Minimum match = 6

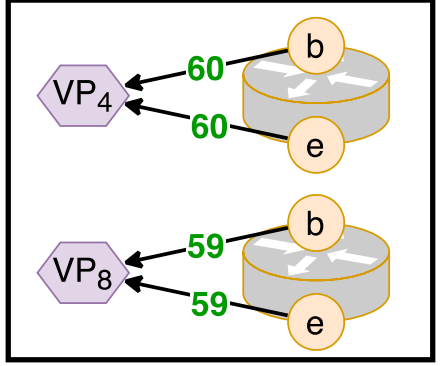
- Acceptance = 100%
  - 1/6 = **16.7%**

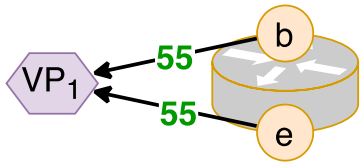


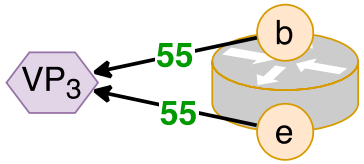
Example

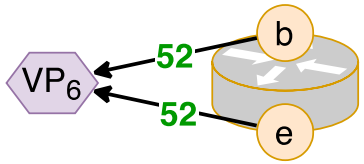
- Minimum match = 6

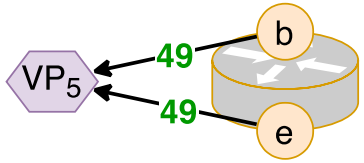
- Acceptance = 100%
  - 2/6 = **33.3%**

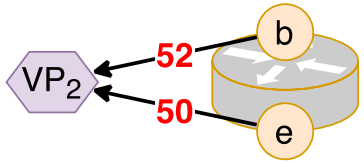


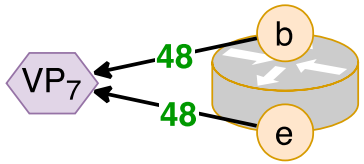








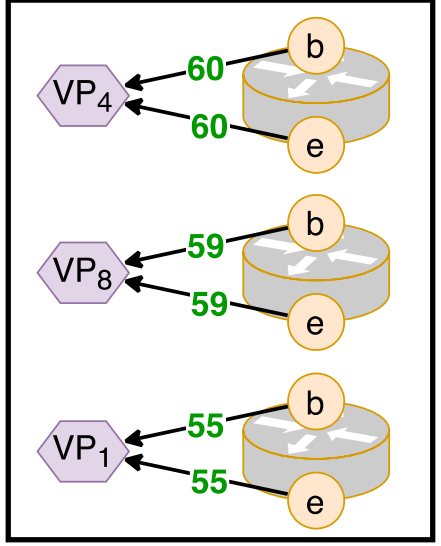


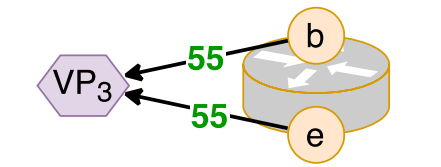


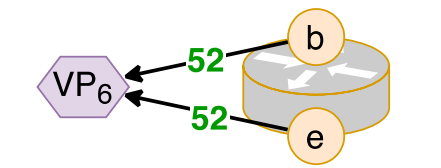
Example

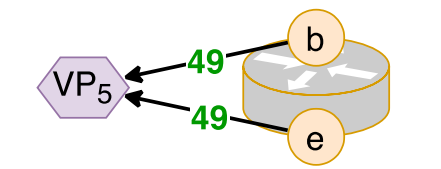
- Minimum match = 6

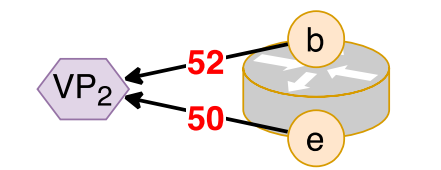
- Acceptance = 100%
  - 3/6 = **50%**

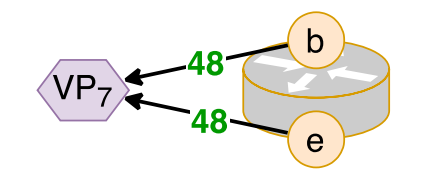








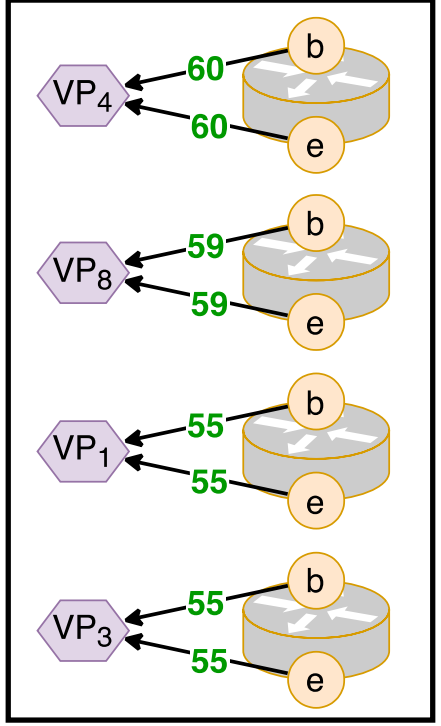


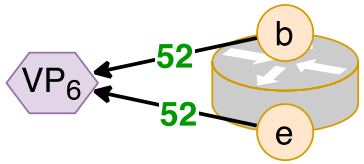


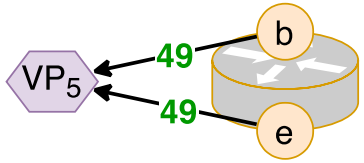
Example

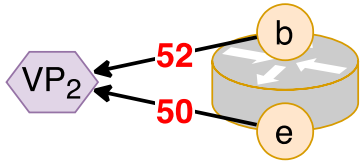
- Minimum match = 6

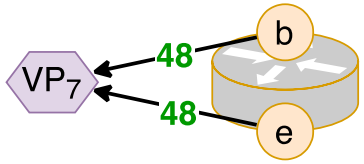
- Acceptance = 100%
  - 4/6 = **66.7%**







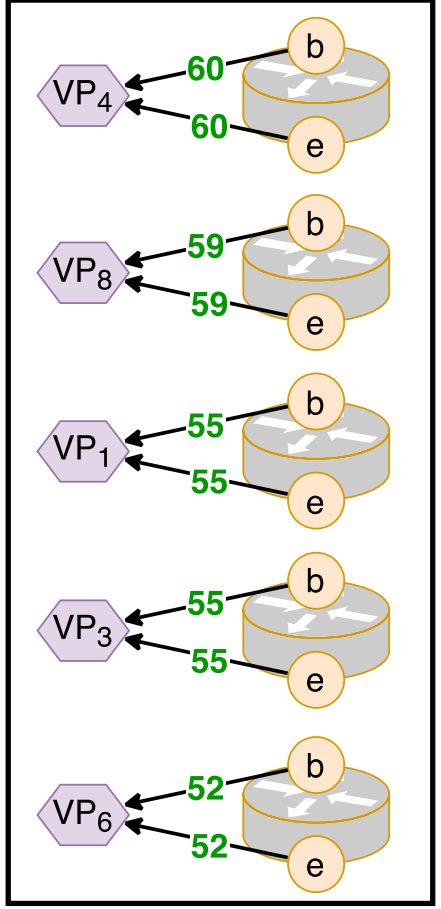


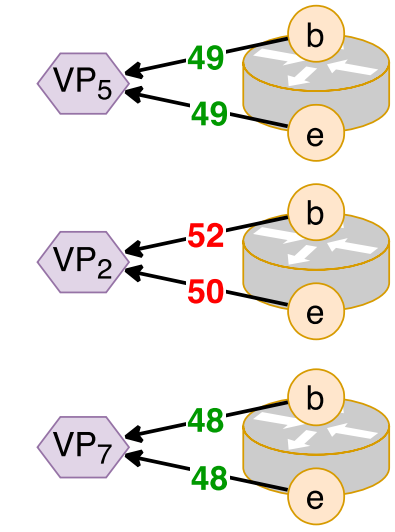


Example

- Minimum match = 6

- Acceptance = 100%
  - 5/6 = **83.3%**

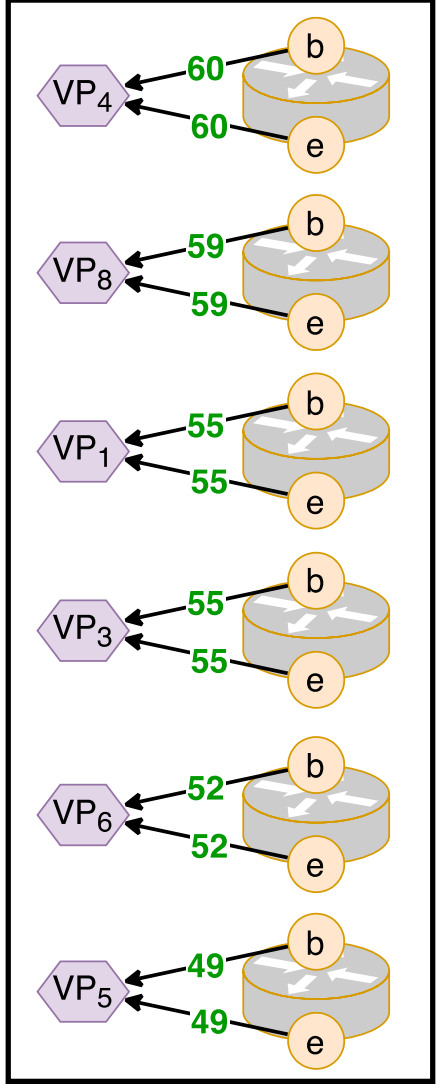


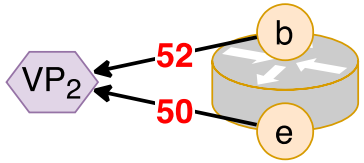


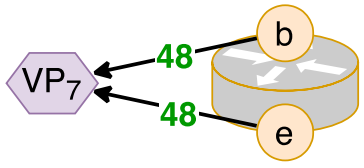
Example

- Minimum match = 6

- Acceptance = 100%
  - 6/6 = **100%**



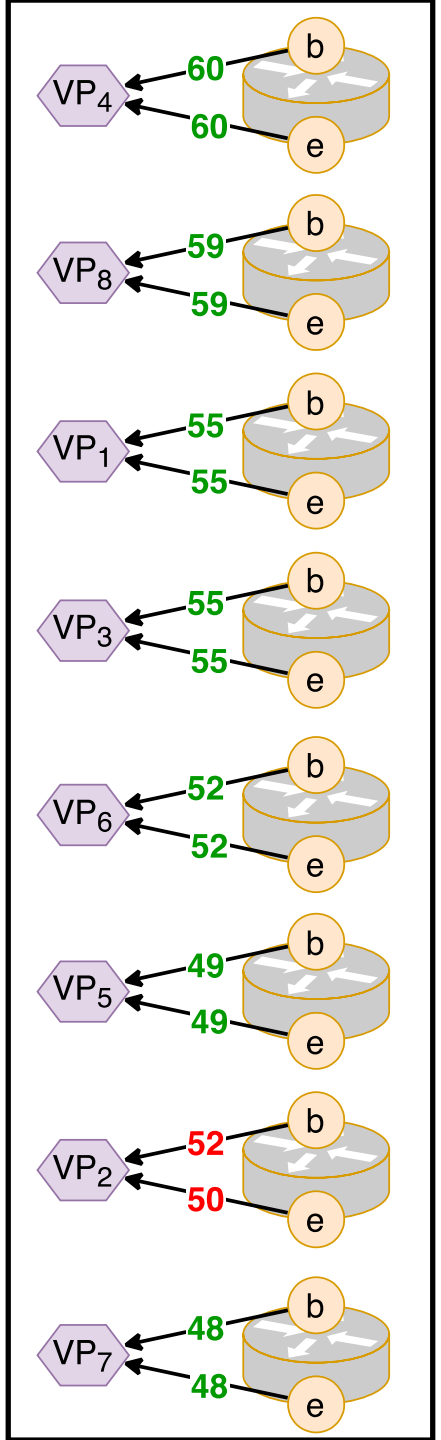




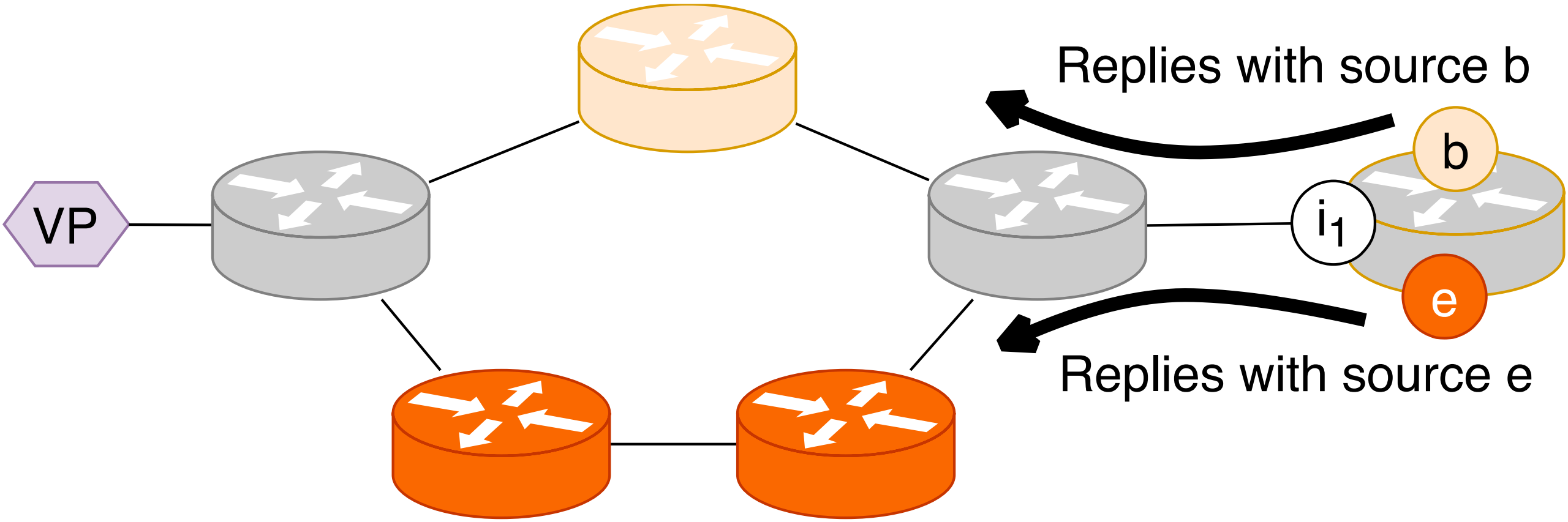
Example

- Minimum match = **7**

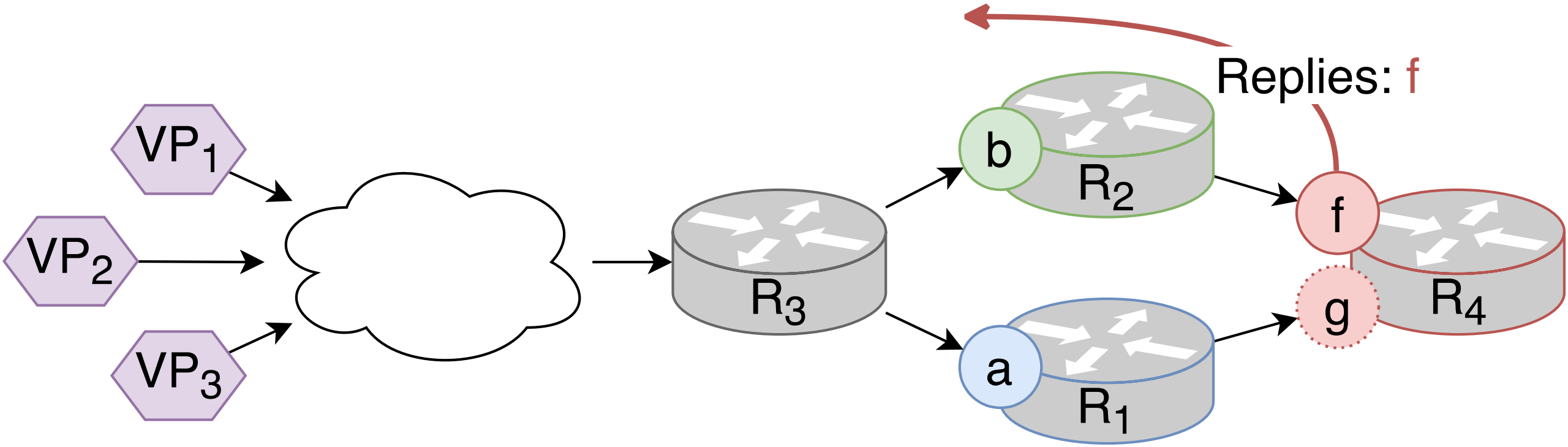
- Acceptance = 100%
  - 7/8 = **87.5%**



Load Balancing: Prune Valid Alias Pair



Load Balancing: Keep Invalid Alias Pair



Overview

1. What is alias resolution and why do we need it?
2. Current state-of-the-art
3. APPLE methodology
   A. Common successor groups
   B. Prune with pings from many vantage points (VPs)
4. **Results from two ground truth networks**
5. Conclusion

Experiments

	ITDK	Pings Sent	Probed	Responses	Resp. %	Pairs
IPv4	201904	201904	366,469	292,141	79.7%	5,022,839
IPv6	201901	201905	76,098	59,778	78.6%	563,489

https://www.caida.org/data/internet-topology-data-kit/

	VPs	ASNs	Countries	Cities
IPv4	99	71	37	83
IPv6	78	61	29	63

https://www.caida.org/projects/ark/locations/

Ground Truth

- 2 ground truth networks: Internet2 & large US R&E network

	Routers		**Alias Pairs**		**Probed**		**Responses**	
	IPv4	IPv6	IPv4	IPv6	IPv4	IPv6	IPv4	IPv6
Internet2	42	41	2176	1095	719	616	646	536
R&E	25	16	1651	137	352	137	352	137
Total	**67**	**57**	**3737**	**1232**	**1071**	**753**	**998**	**673**

How Popular is the Most Popular Reply TTL?

- For each VP:
  - Select most frequent reply TTL
  - Compute % of all replies at that VP

VP_1	VP_2
56	59
56	58
57	61
54	58
	57
50%	**40%**

How Popular is the Most Popular Reply TTL?

- For each VP:
  - Select most frequent reply TTL
  - Compute % of all replies at that VP

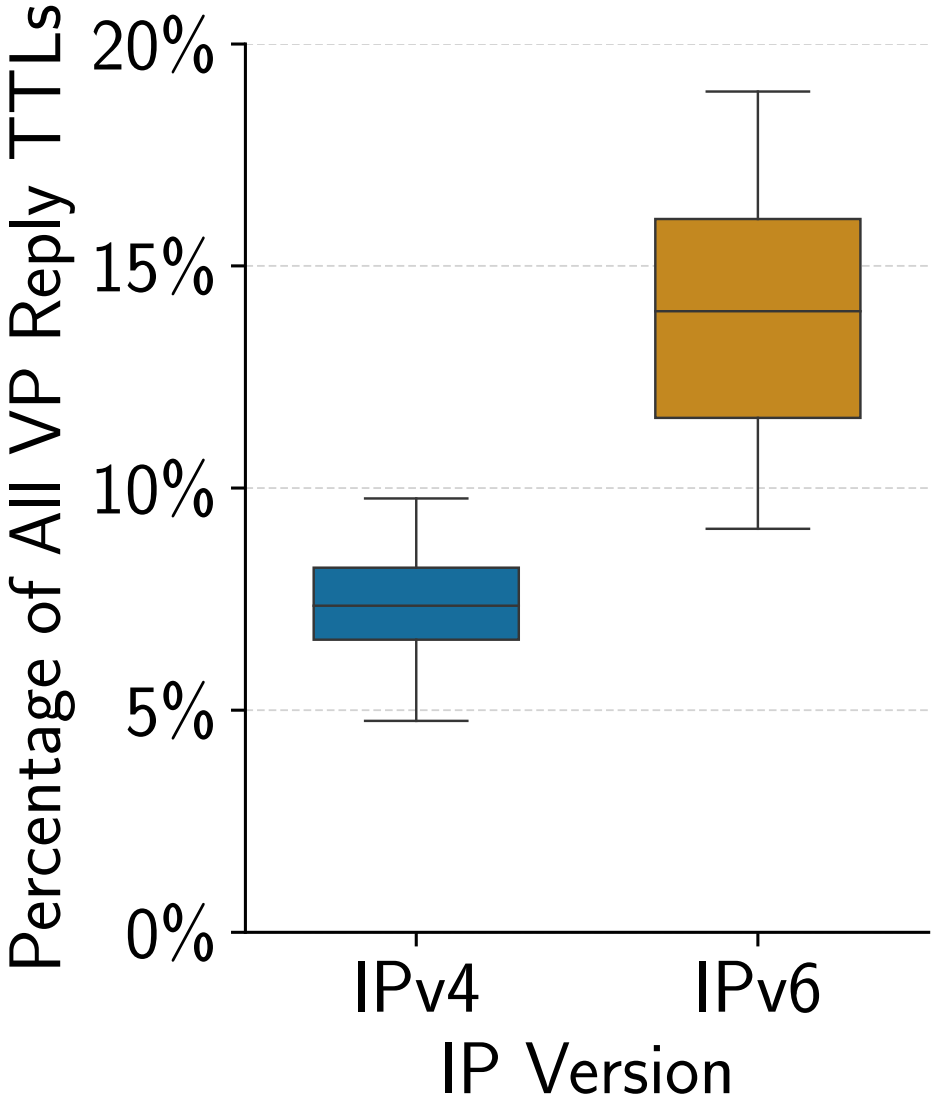
VP_1	VP_2
56	59
56	58
57	61
54	58
	57
50%	**40%**

How Popular is the Most Popular Reply TTL?

- For each VP:
  - Select most frequent reply TTL
  - Compute % of all replies at that VP

VP_1	VP_2
56	59
56	**58**
57	61
54	**58**
	57
50%	**40%**

Maximum % of Individual Reply TTLs



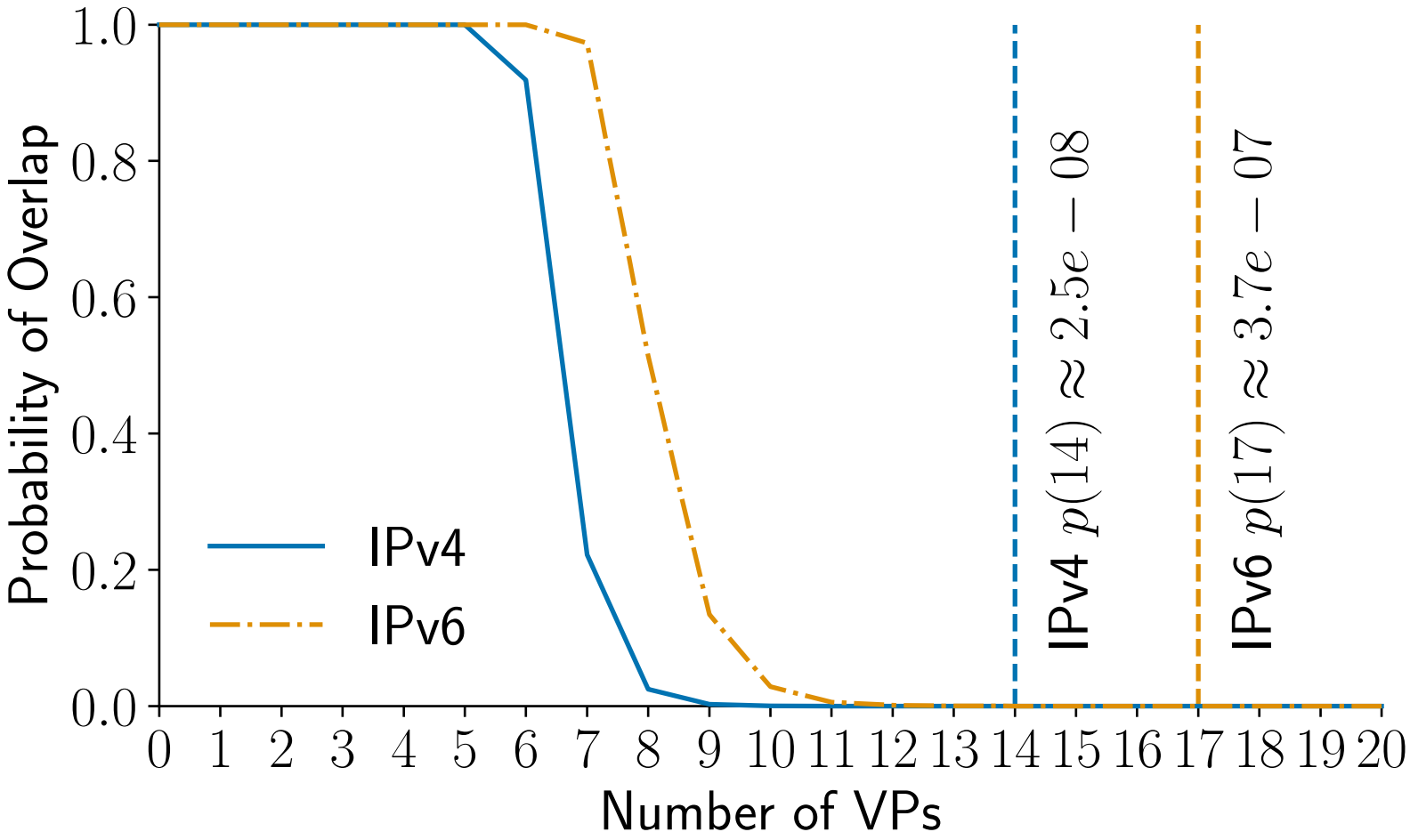
- IPv4: most frequent reply TTL <**10%** of replies at any VP
  - Estimate 10 possible reply TTLs
- IPv6: most frequent reply TTL <**20%** of replies at any VP
  - Estimate 5 possible reply TTLs

Setting the Minimum Match Threshold

Input Parameters

	IPv4	IPv6
Alias Pairs	5 M	563 K
Reply TTLs	10	5

Estimated Probability of Random Reply TTL Overlap

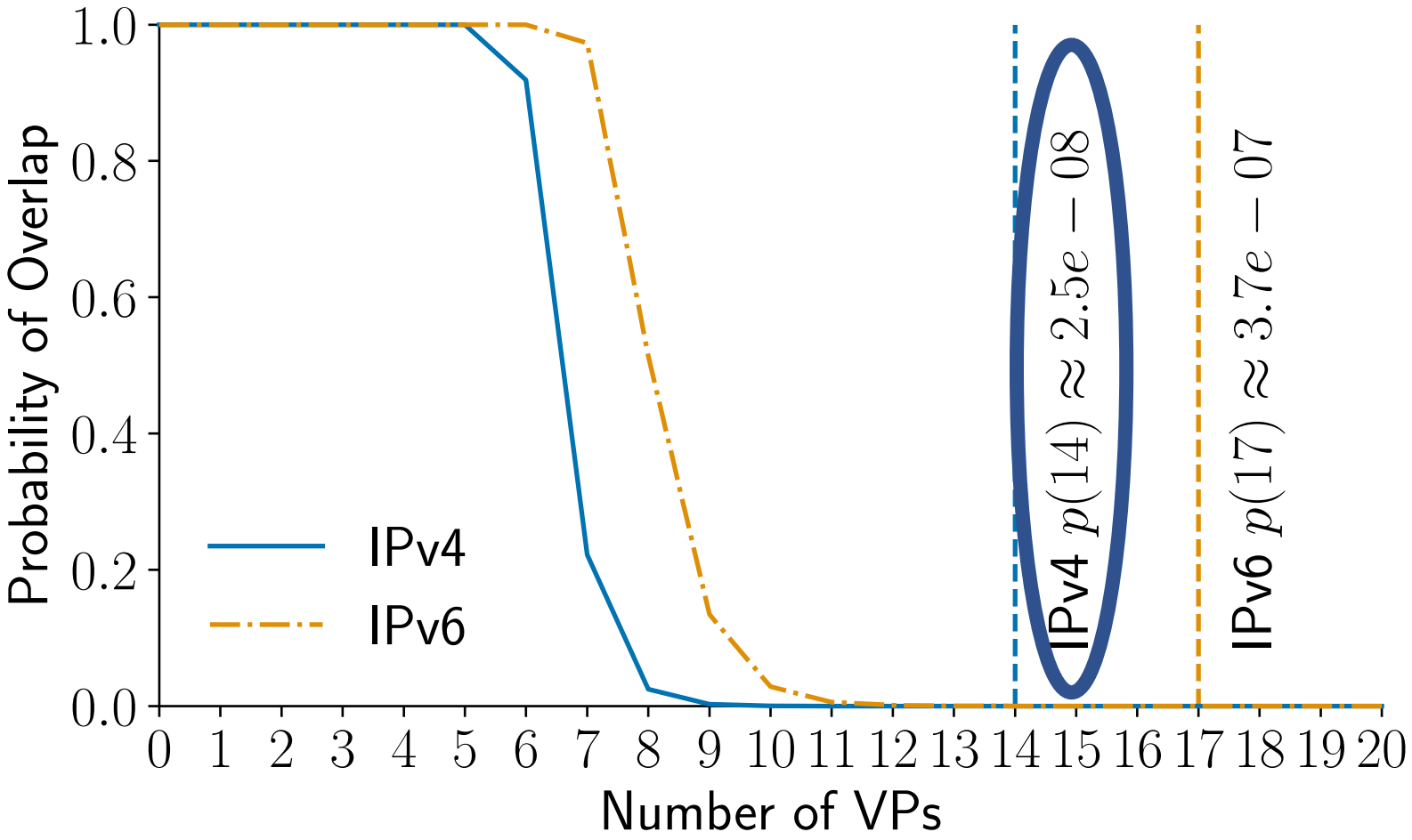


Setting the Minimum Match Threshold

Input Parameters

	IPv4	IPv6
Alias Pairs	5 M	563 K
Reply TTLs	10	5

Estimated Probability of Random Reply TTL Overlap

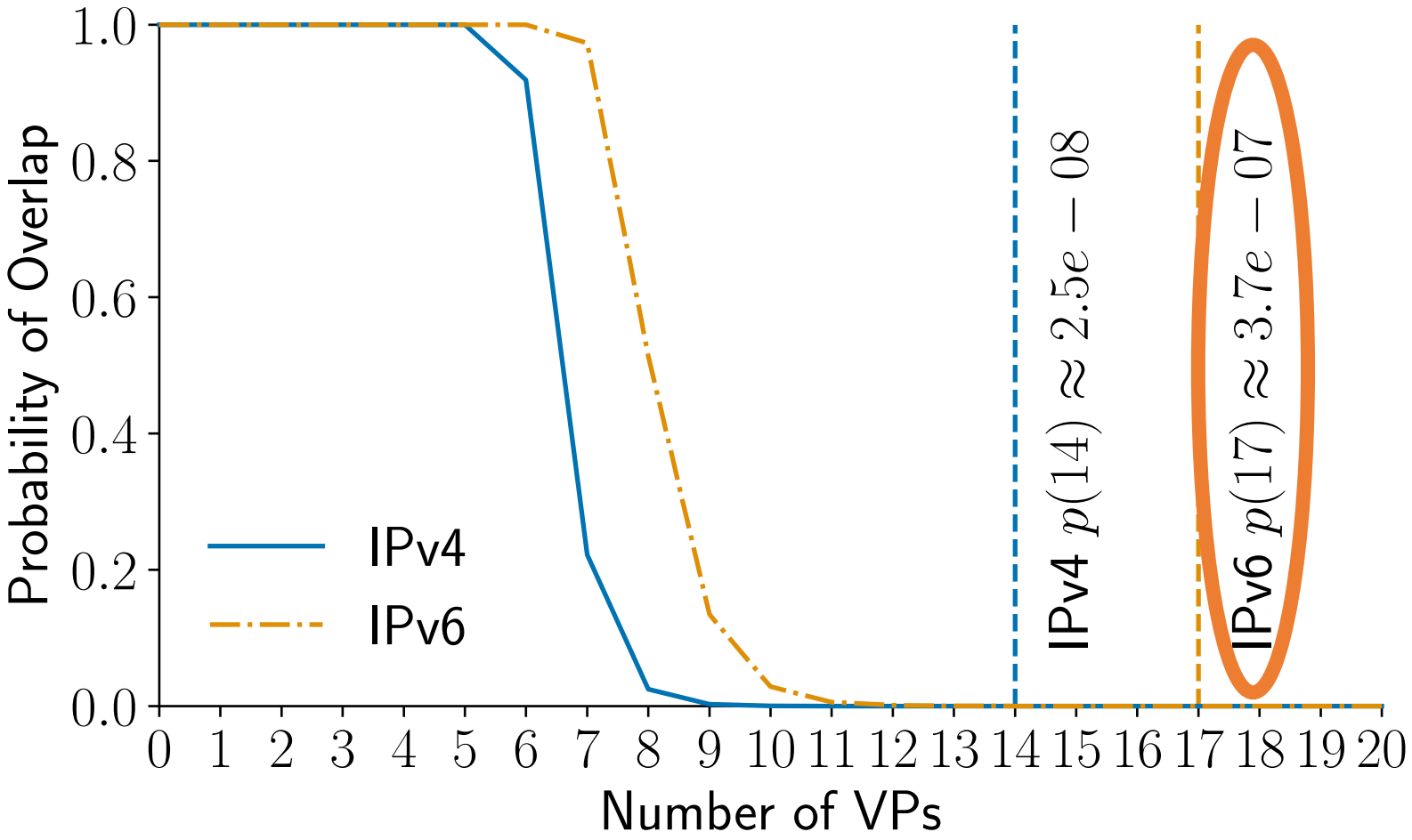


Setting the Minimum Match Threshold

Input Parameters

	IPv4	IPv6
Alias Pairs	5 M	563 K
Reply TTLs	10	5

Estimated Probability of Random Reply TTL Overlap

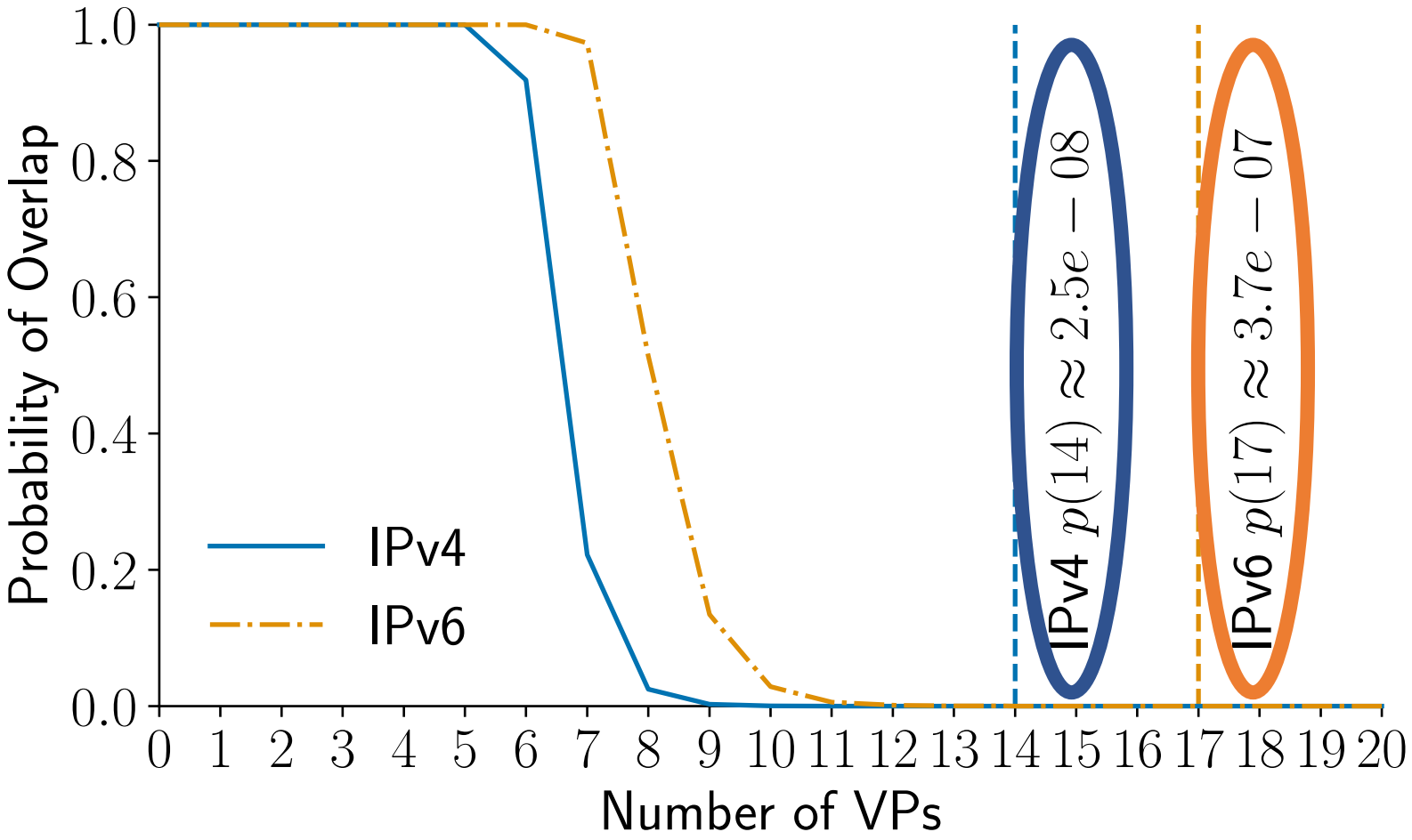


Setting the Minimum Match Threshold

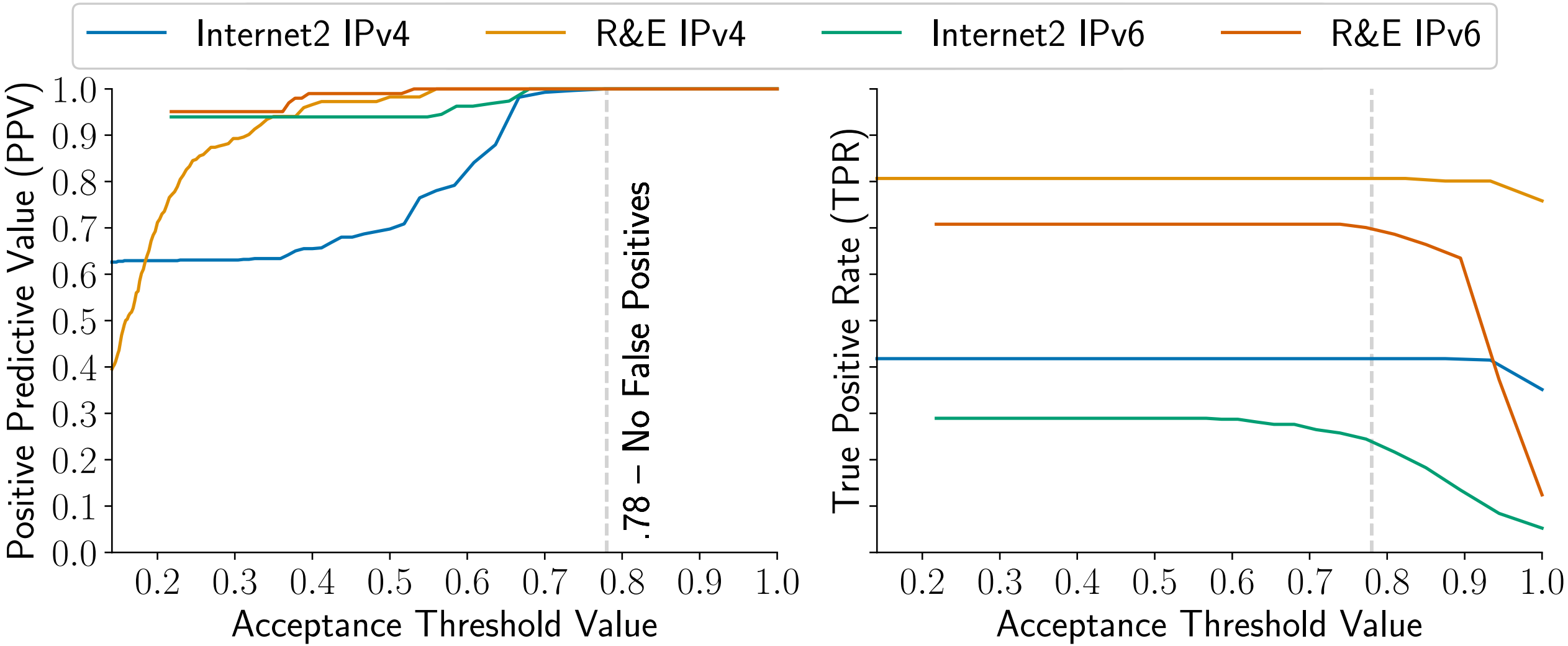
Input Parameters

	IPv4	IPv6
Alias Pairs	5 M	563 K
Reply TTLs	10	5

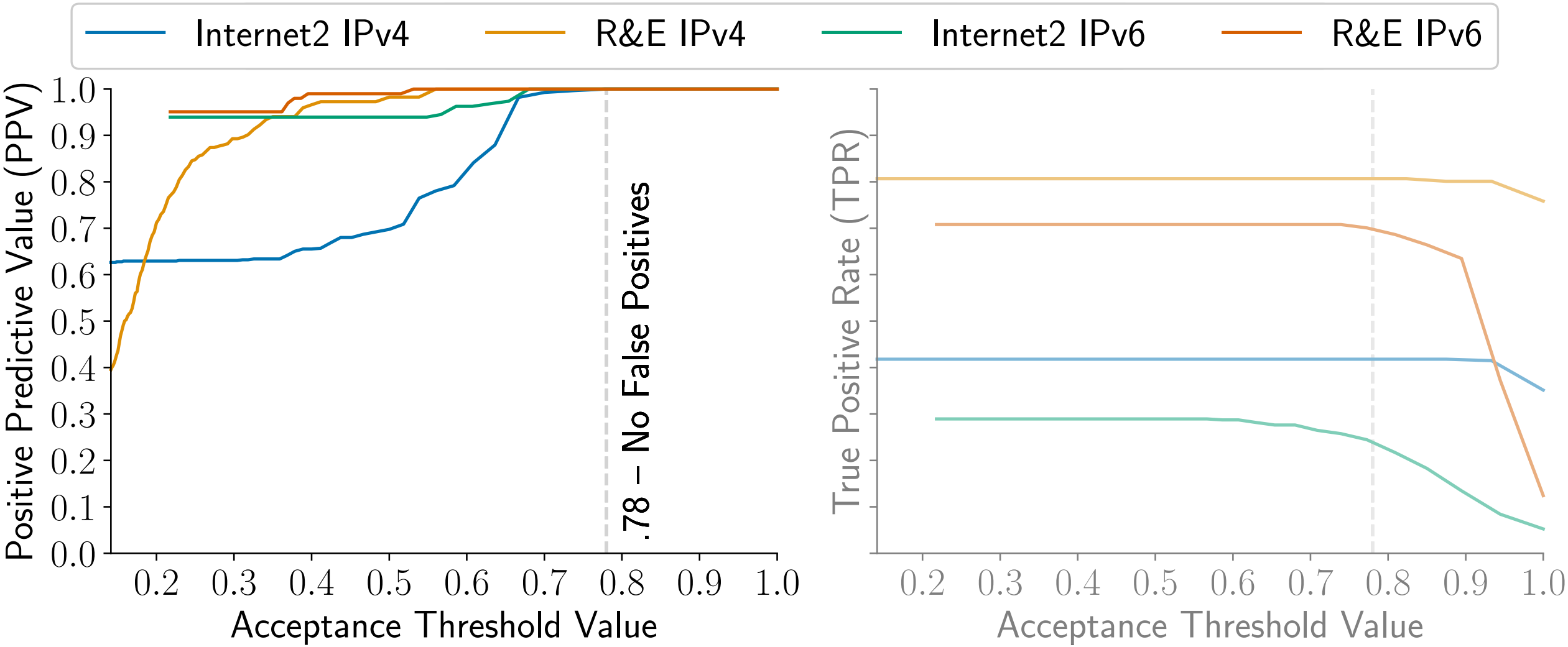
Estimated Probability of Random Reply TTL Overlap



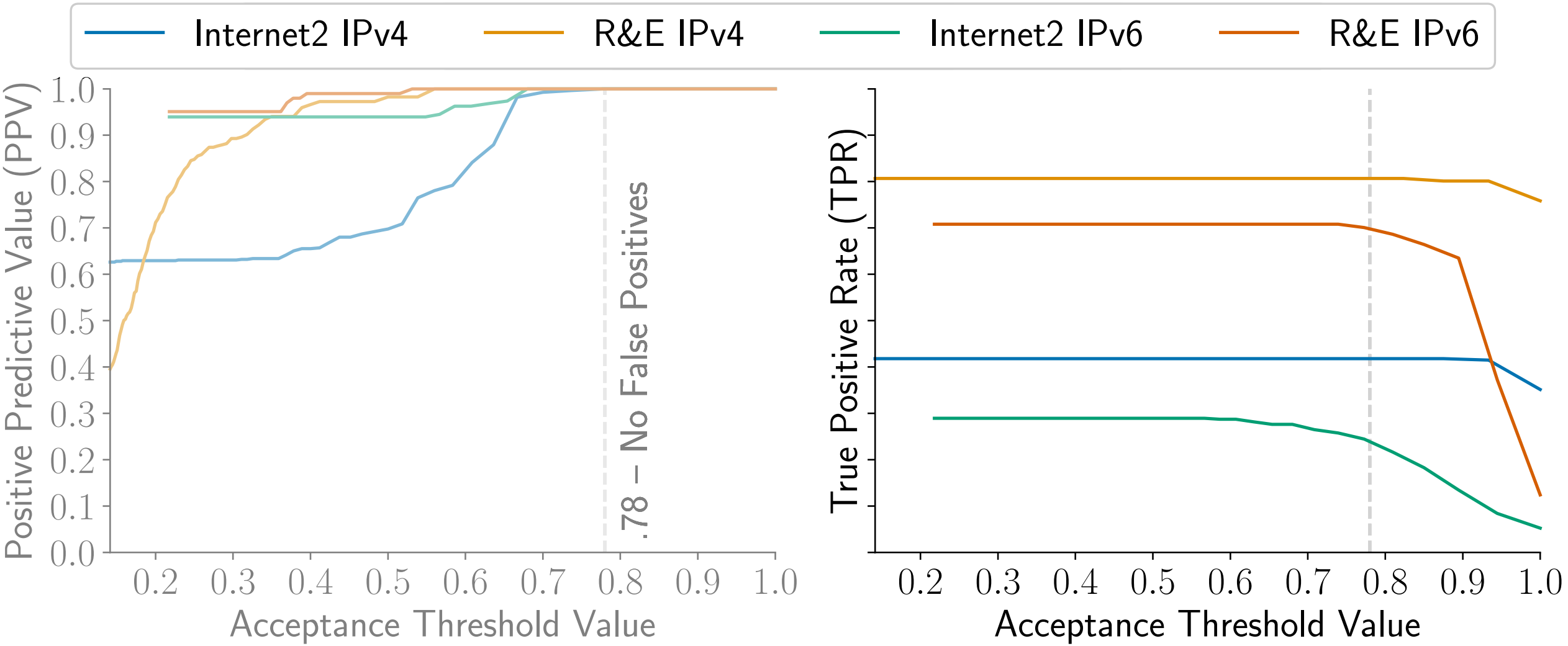
Setting the Acceptance Threshold



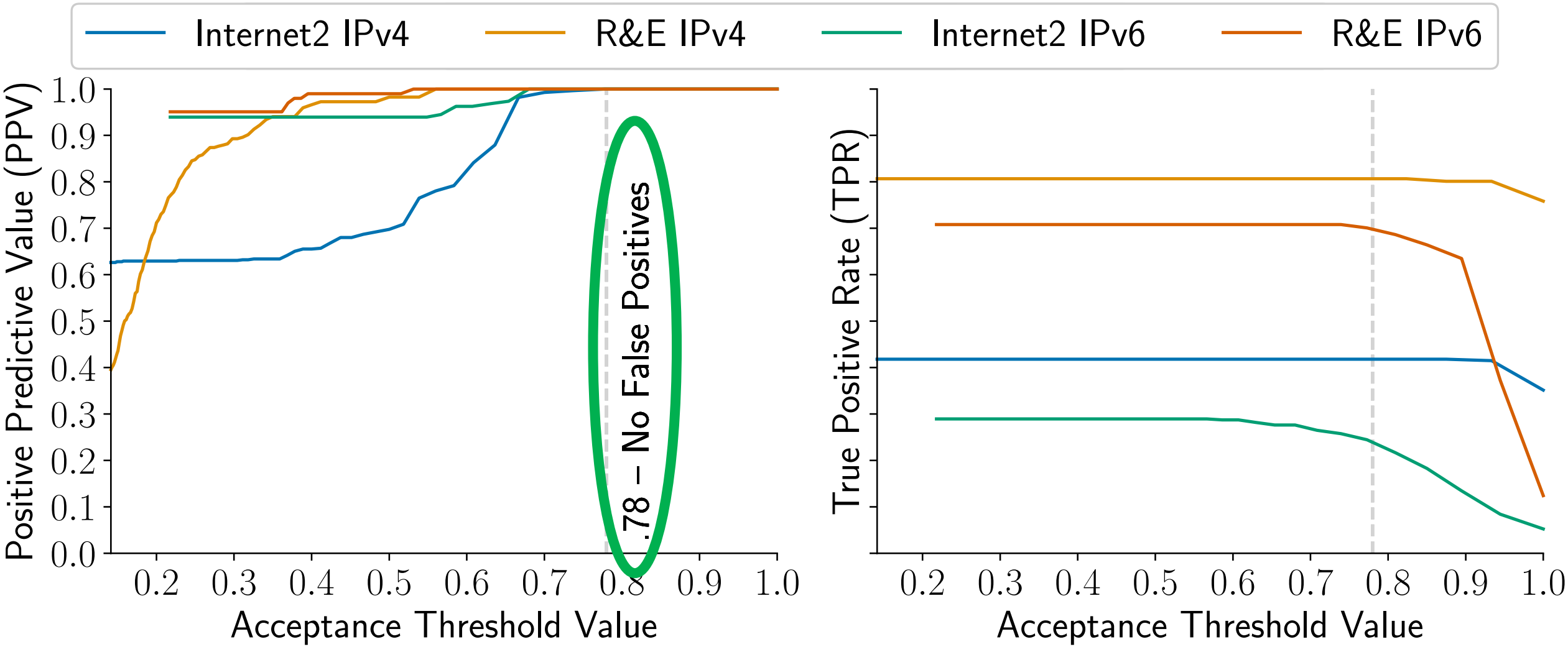
Setting the Acceptance Threshold



Setting the Acceptance Threshold

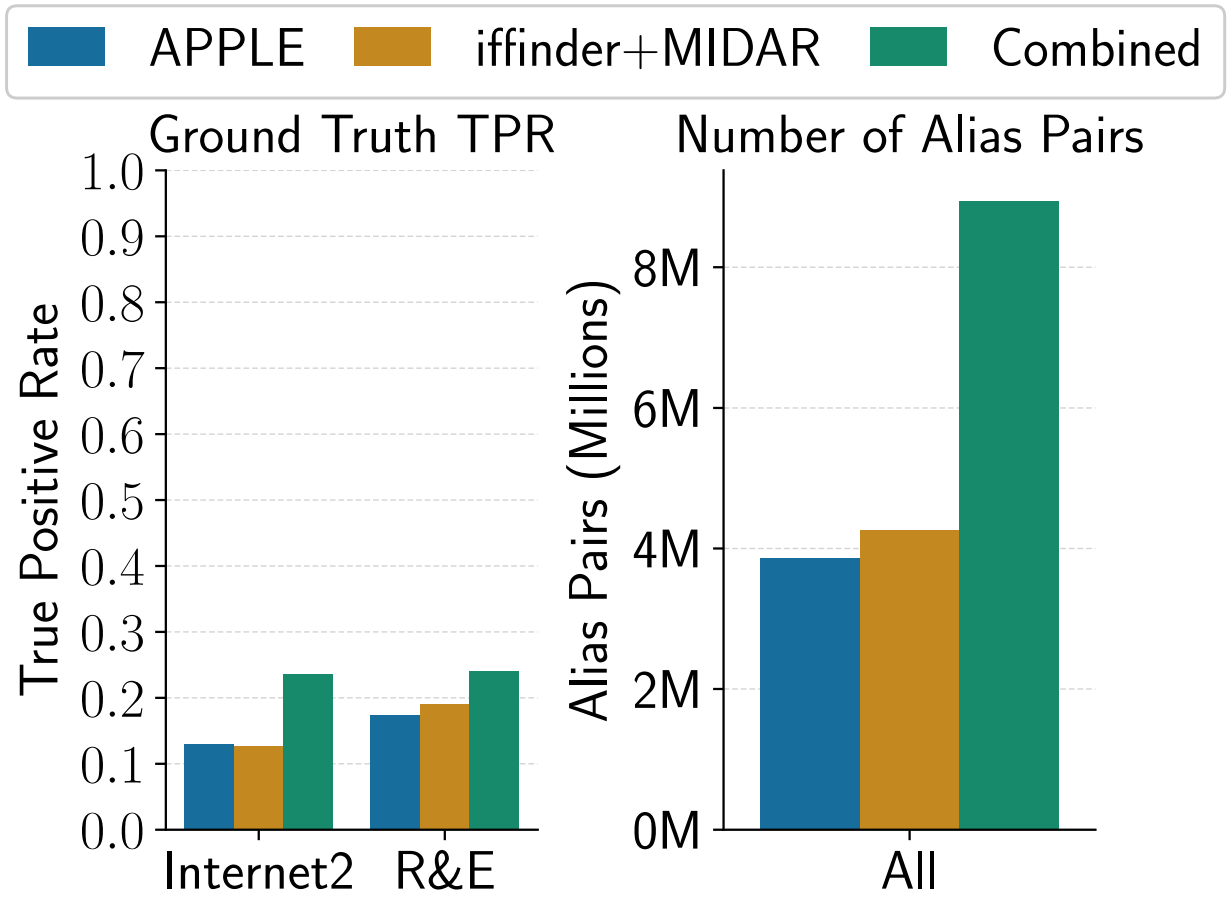


Setting the Acceptance Threshold

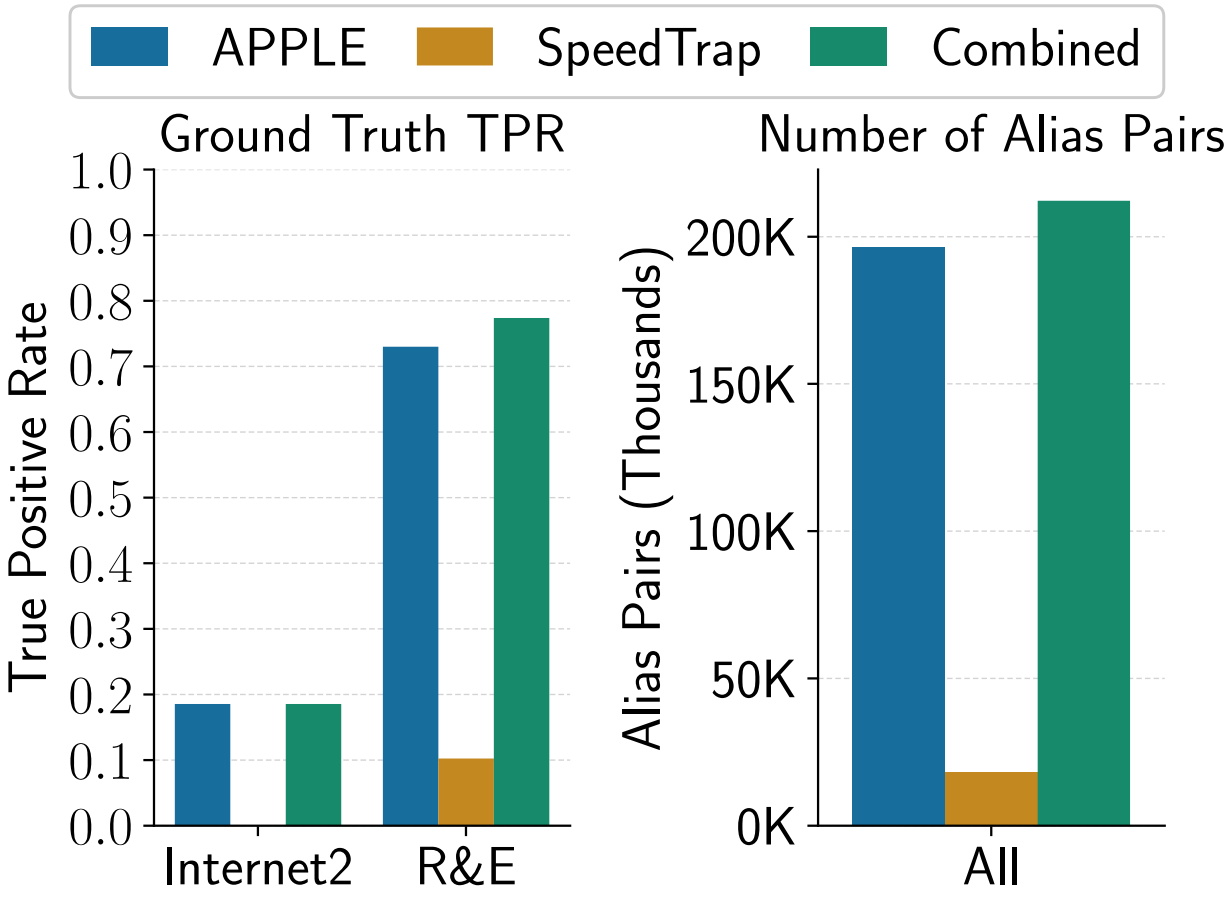


True Positive Rate (TPR) and Number of Alias Pairs

IPv4

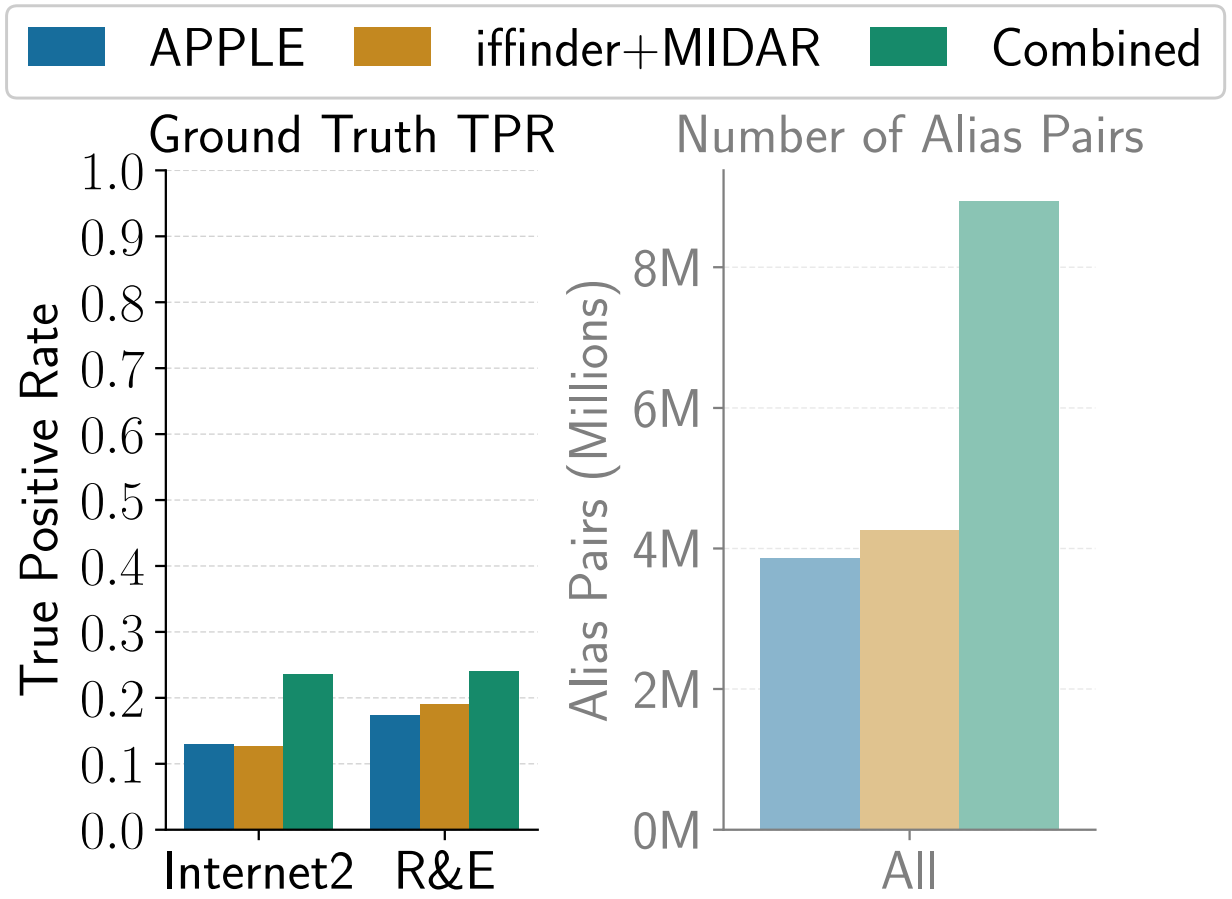


IPv6

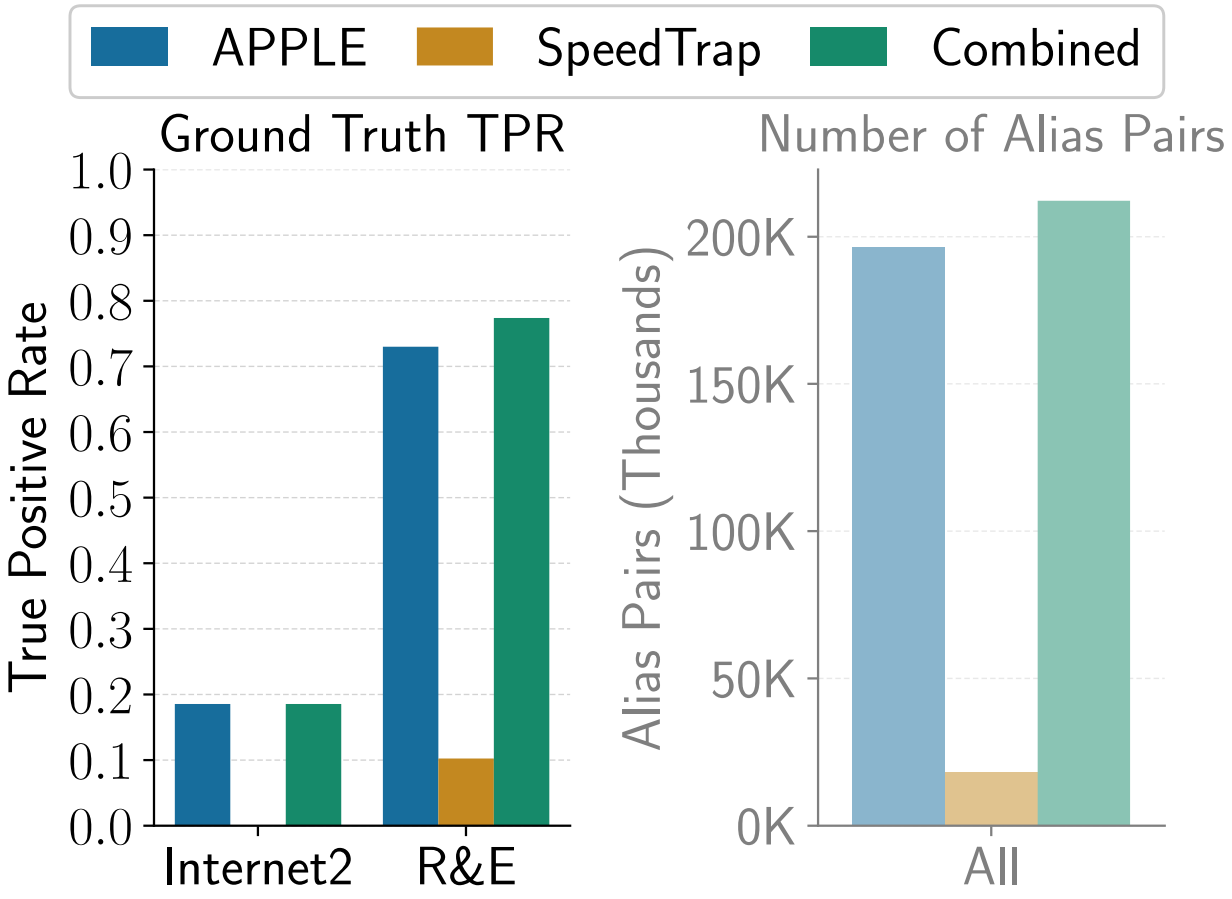


True Positive Rate (TPR) and Number of Alias Pairs

IPv4

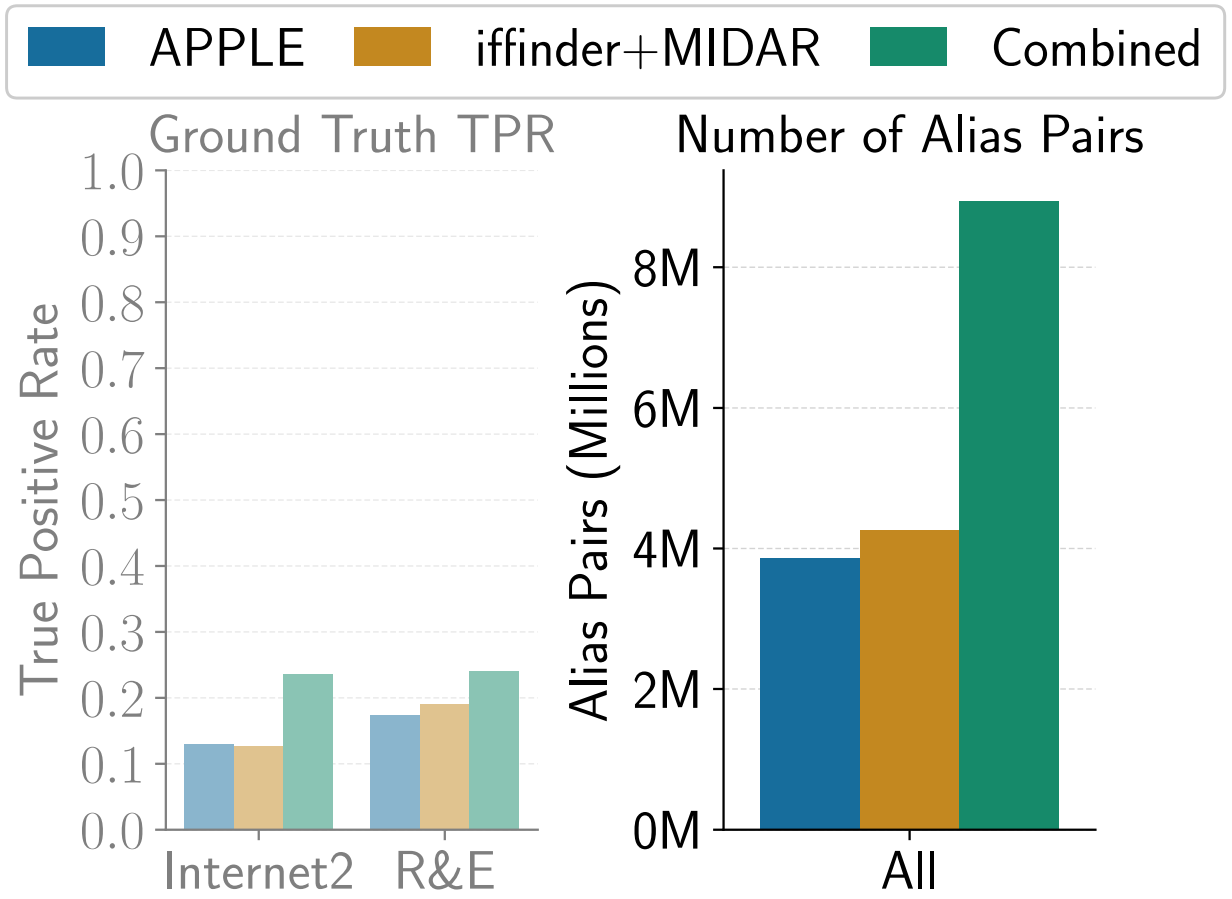


IPv6

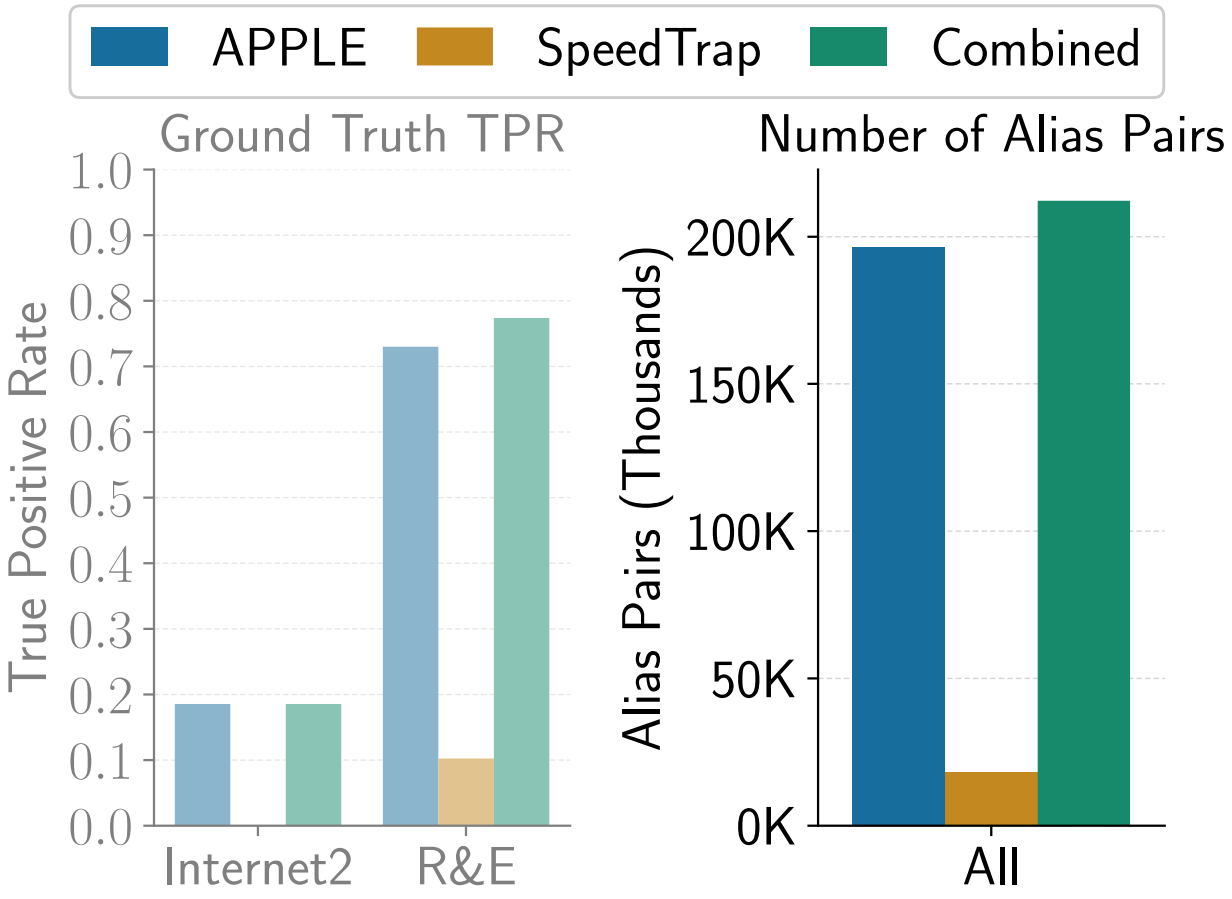


True Positive Rate (TPR) and Number of Alias Pairs

IPv4

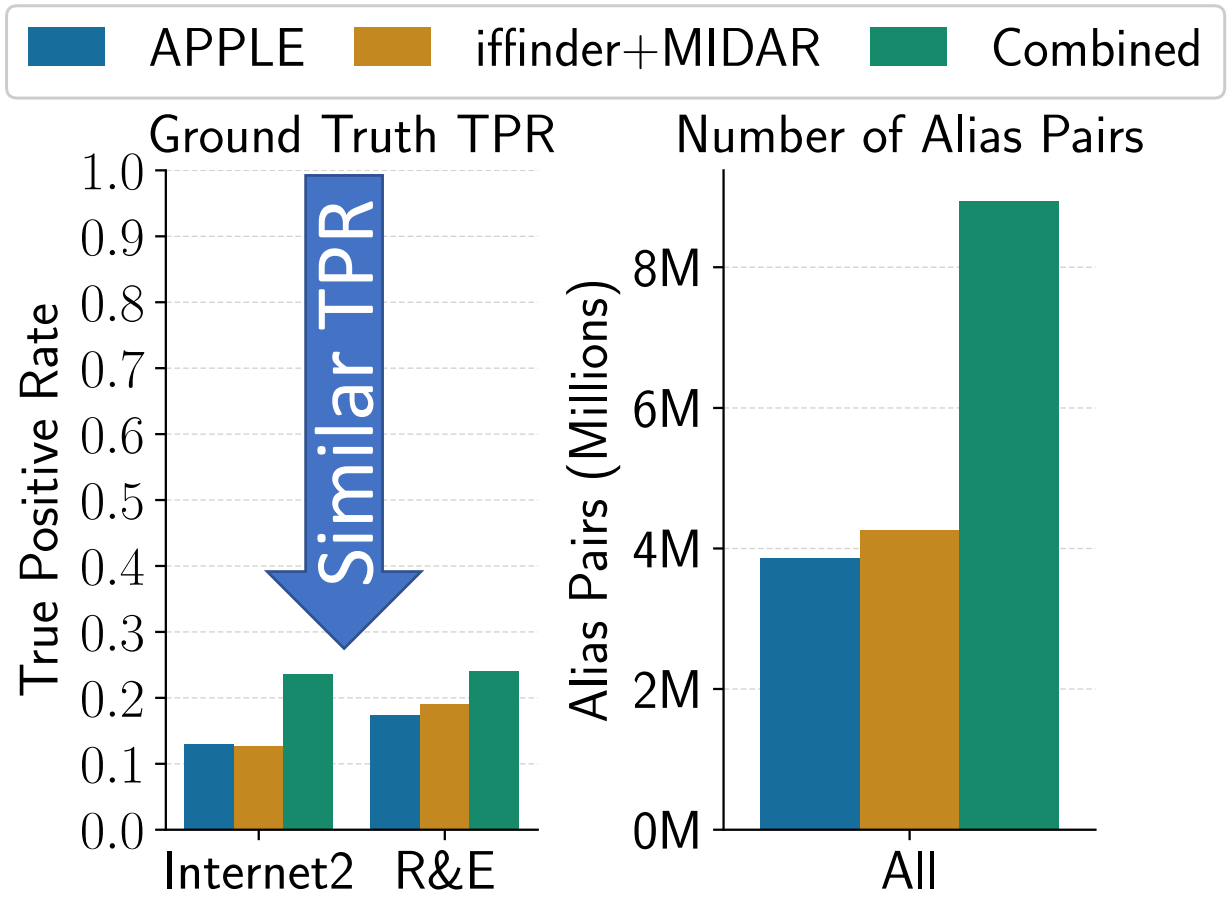


IPv6

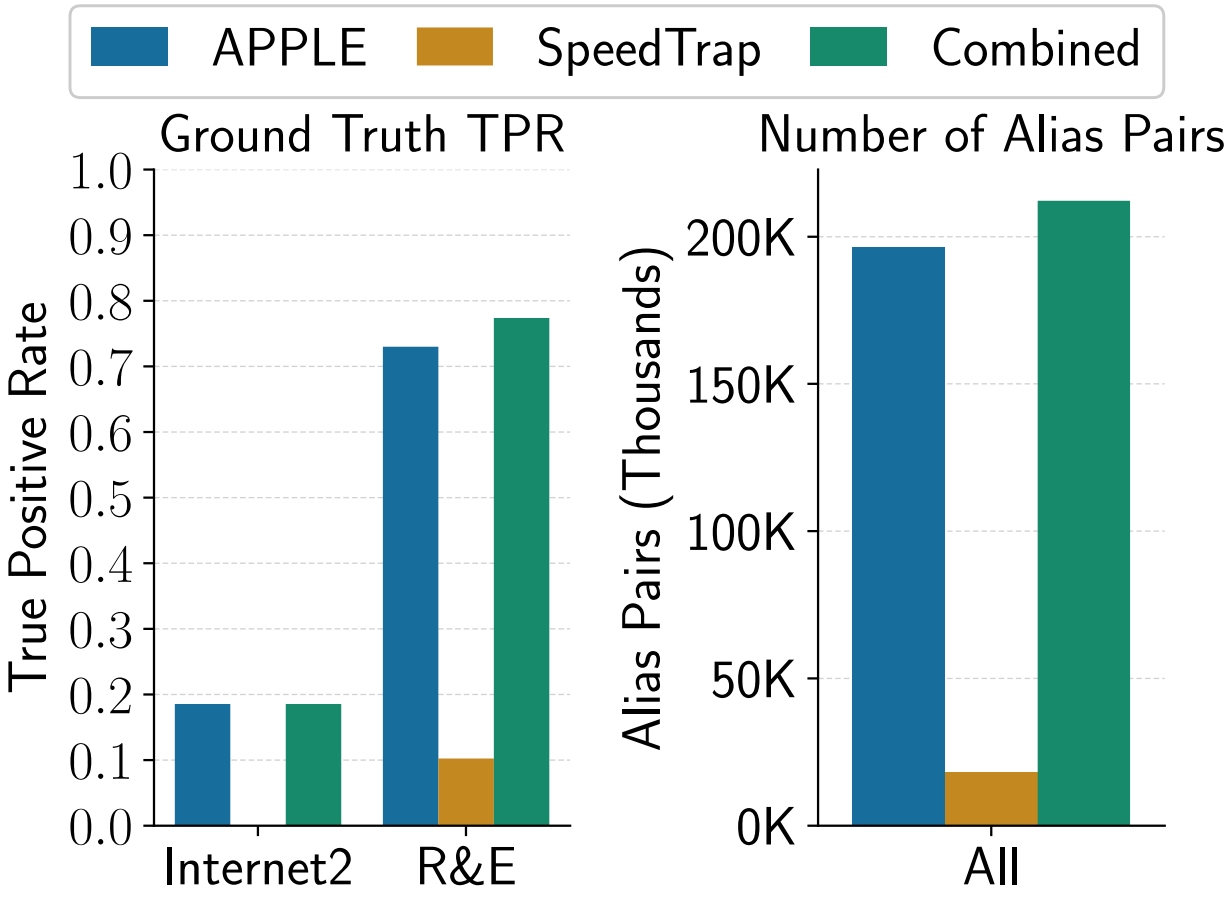


True Positive Rate (TPR) and Number of Alias Pairs

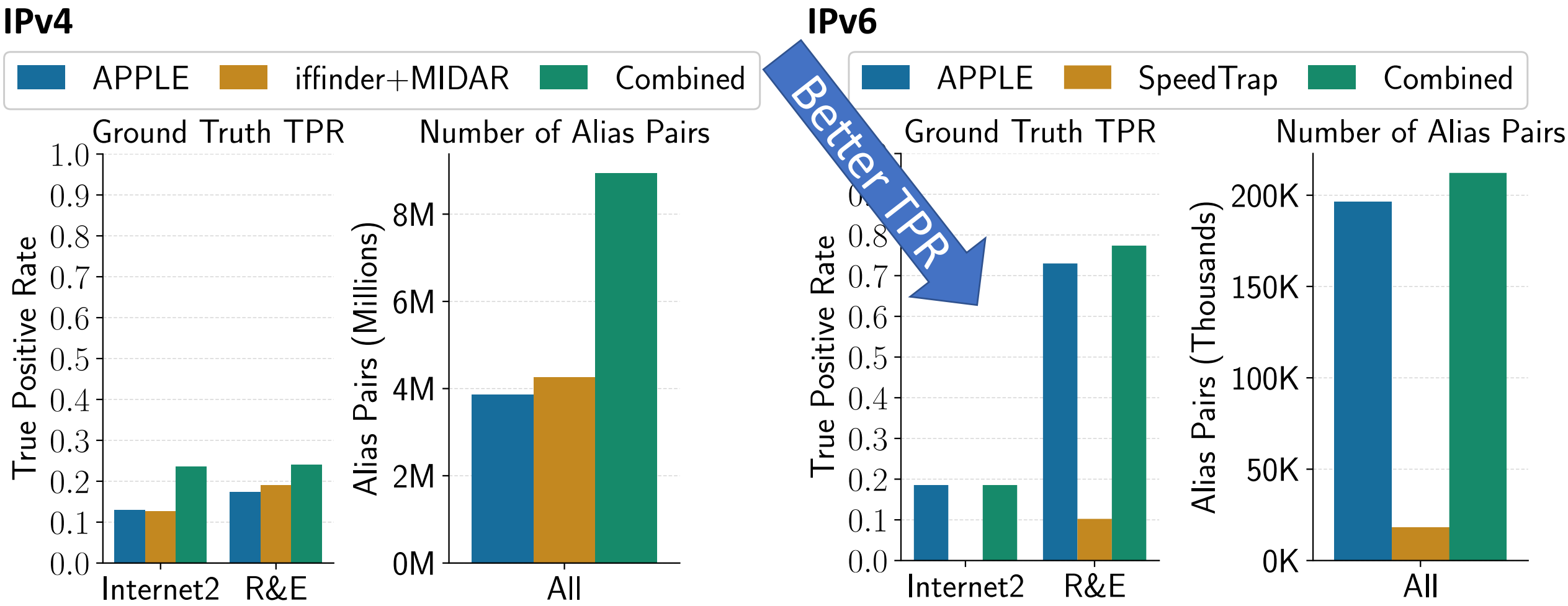
IPv4



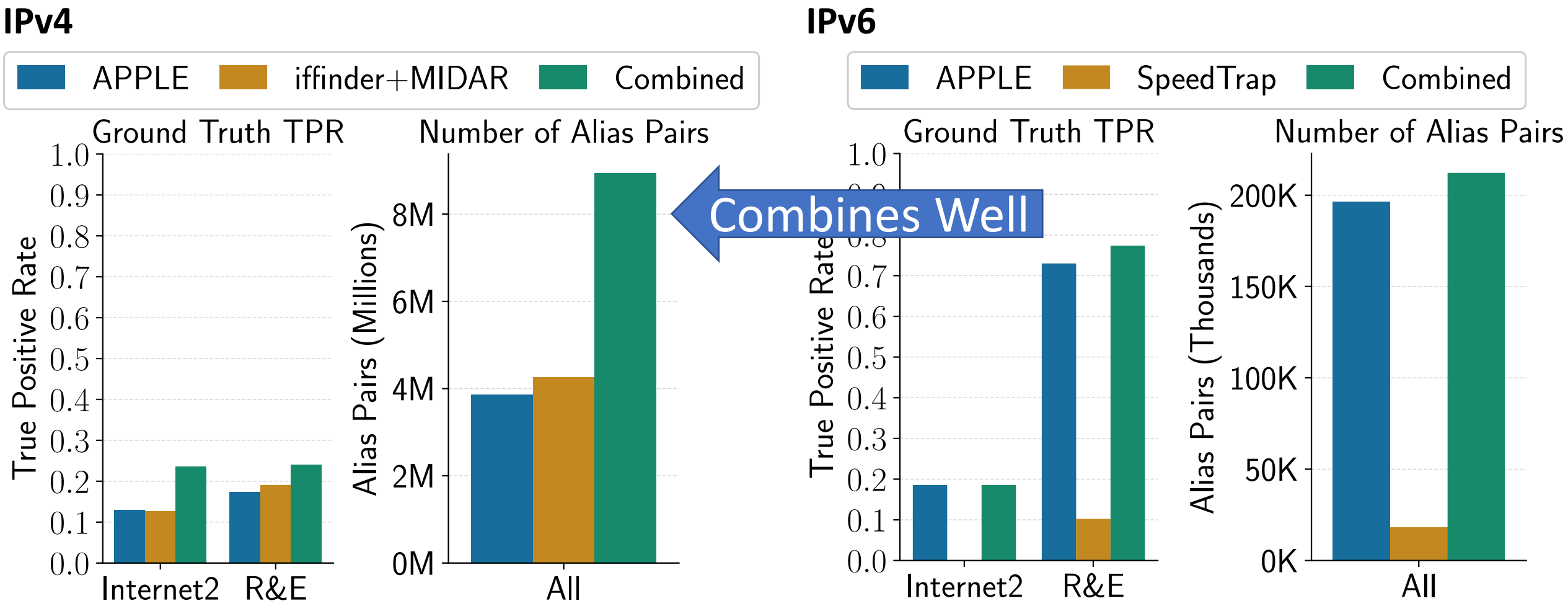
IPv6



True Positive Rate (TPR) and Number of Alias Pairs

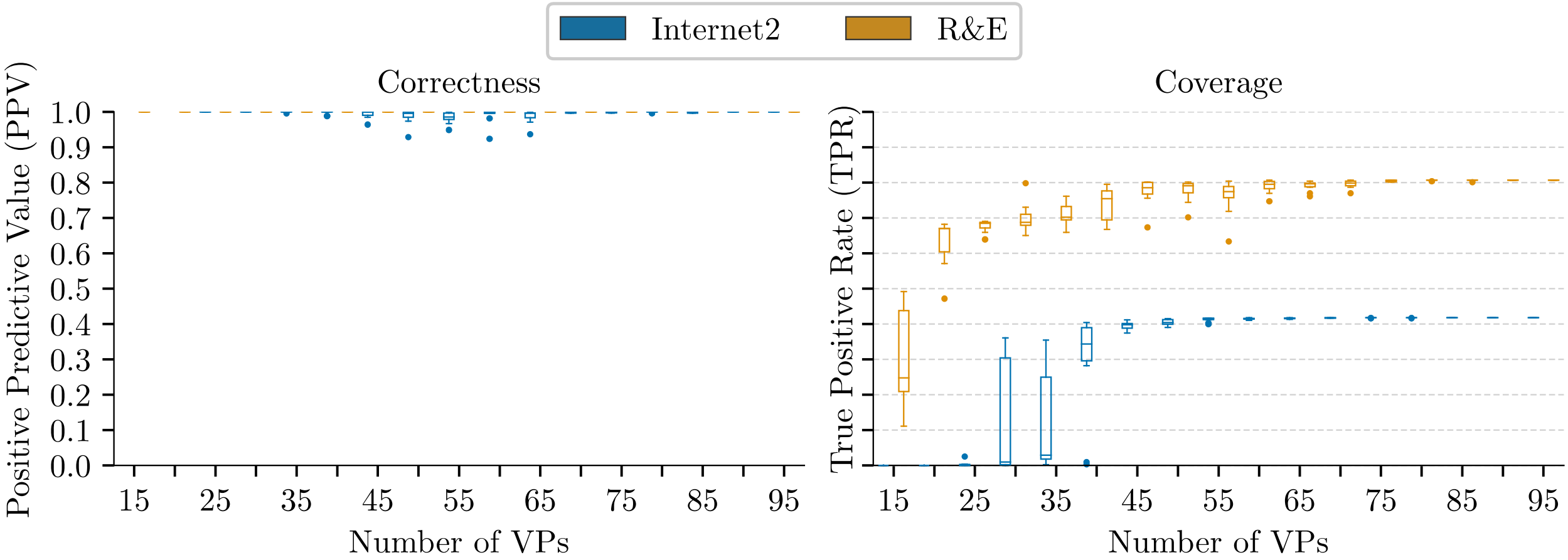


True Positive Rate (TPR) and Number of Alias Pairs



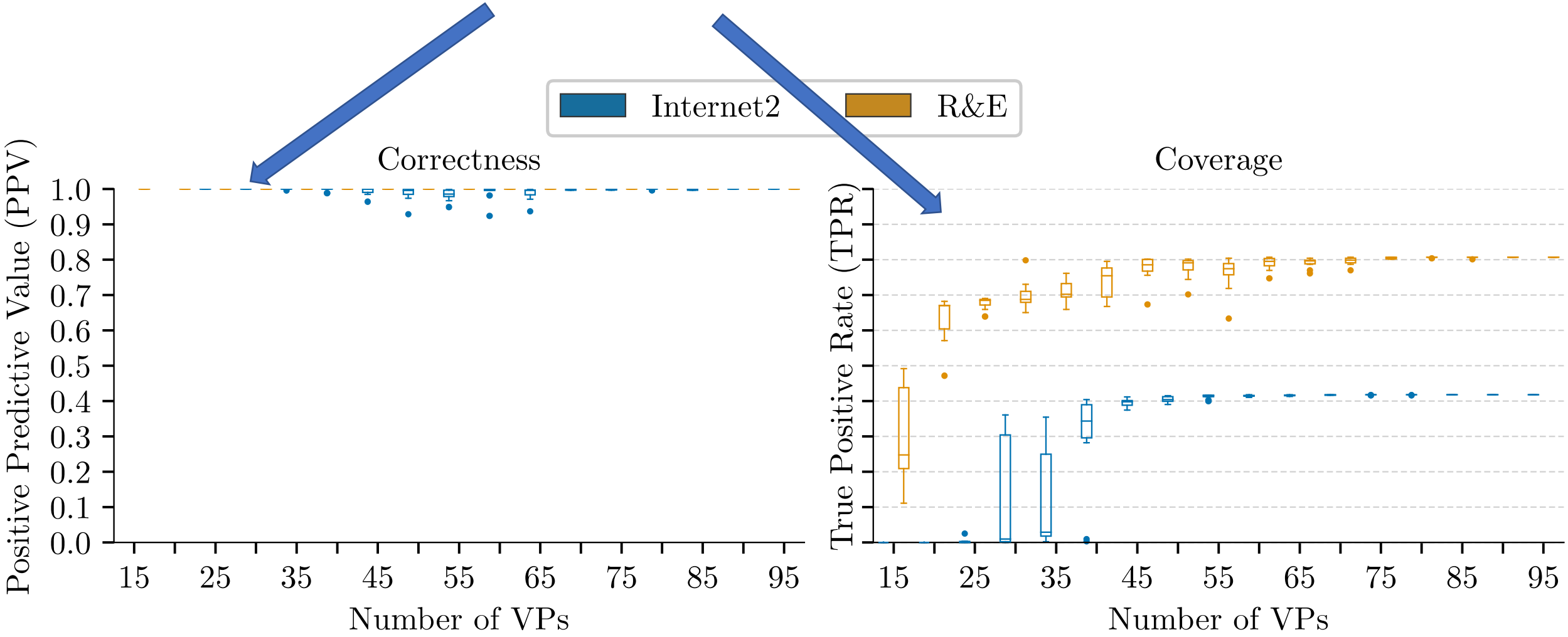
Impact of VPs on PPV and TPR

The number of VPs directly affects correctness and coverage



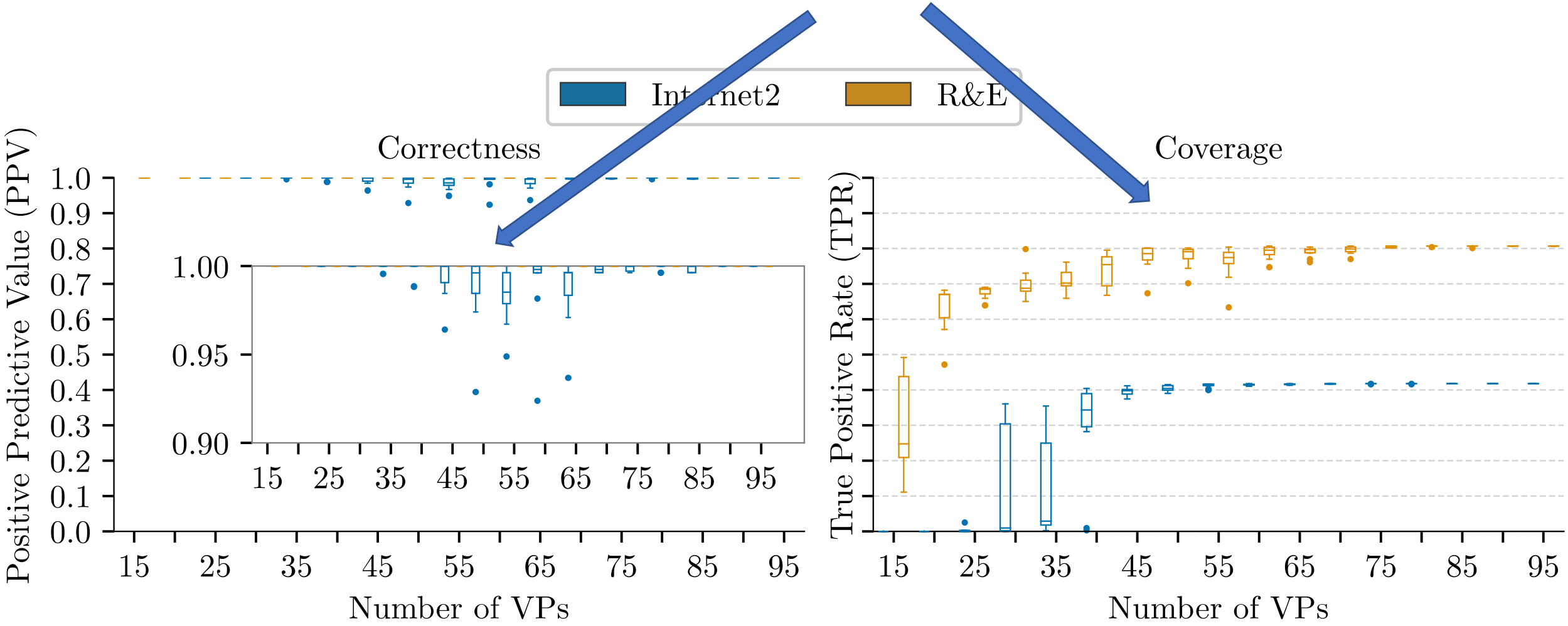
Impact of VPs on PPV and TPR

No false aliases but low coverage for few VPs



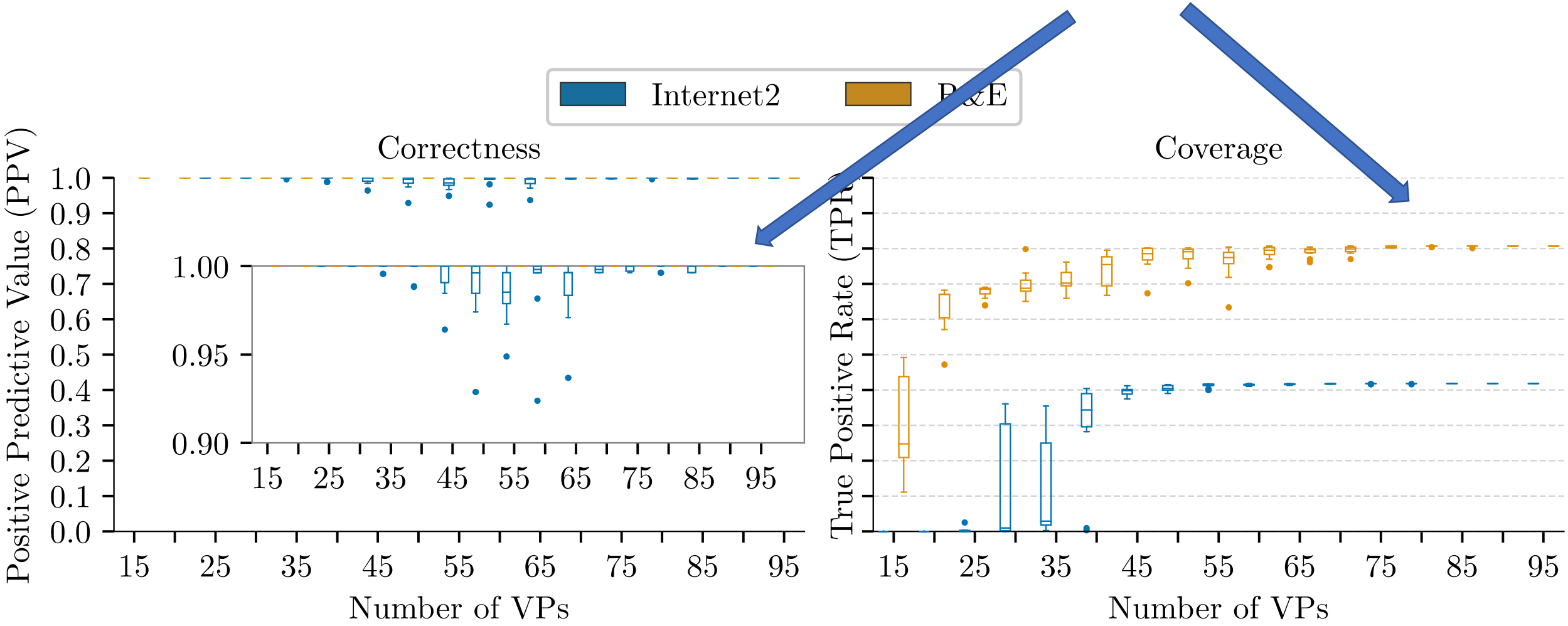
Impact of VPs on PPV and TPR

With more VPs we introduce errors but increase coverage



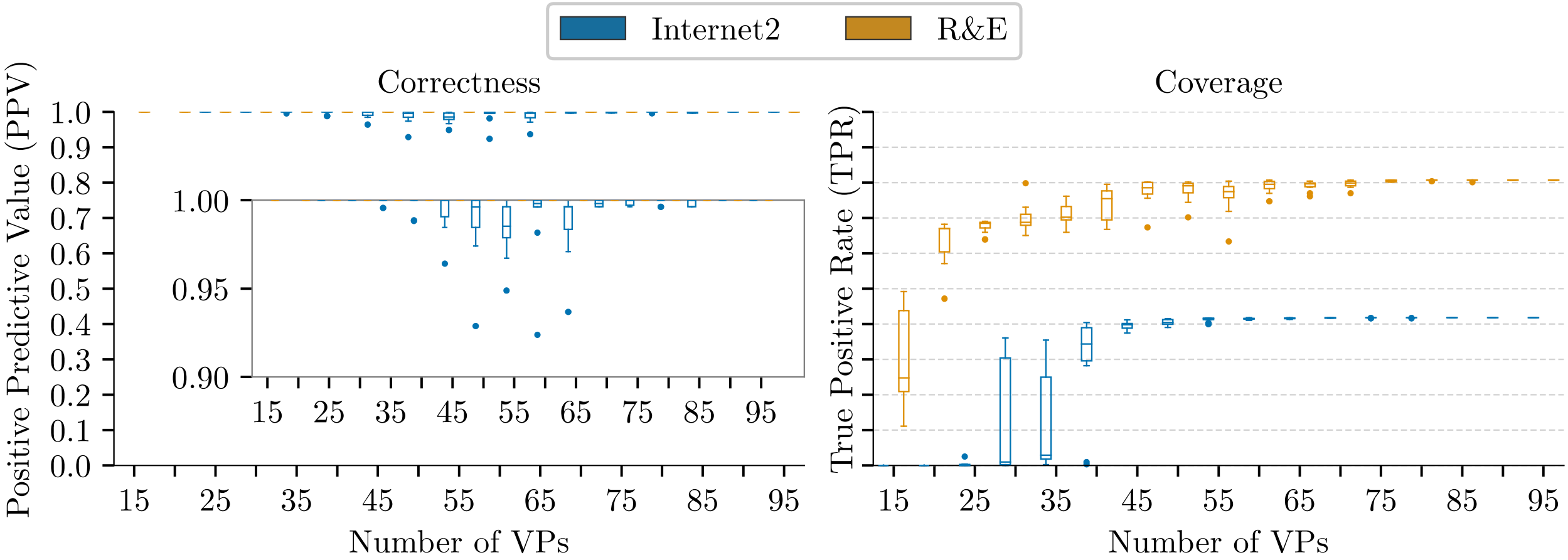
Impact of VPs on PPV and TPR

No more false aliases by 90 VPs



Impact of VPs on PPV and TPR

No more false aliases by 90 VPs



Conclusion and Future Work

- Similar TPR to Midar+iffinder
- Even better TPR for IPv6 than IPv4
- Combines well with existing alias resolution

- Future work
  - More ground truth – use hoiho regexes as pseudo-ground truth
  - Improve pruning to decrease false alias pairs

Conclusion and Future Work

- Similar TPR to Midar+iffinder
- Even better TPR for IPv6 than IPv4
- Combines well with existing alias resolution

- Future work
  - More ground truth – use hoiho regexes as pseudo-ground truth
  - Improve pruning to decrease false alias pairs