Association for Computing Machinery

ACM Internet Measurement Conference 2025

Understanding and Characterizing Intermediate Paths of Email Delivery: The Hidden Dependencies

Ruixuan Li[1], Chaoyi Lu[1], Baojun Liu[1], Yanzhong Lin[2], Haixin Duan[1],

Qingfeng Pan[2], Jun Shao[3]

清華大學 Tsinghua University

Coremail

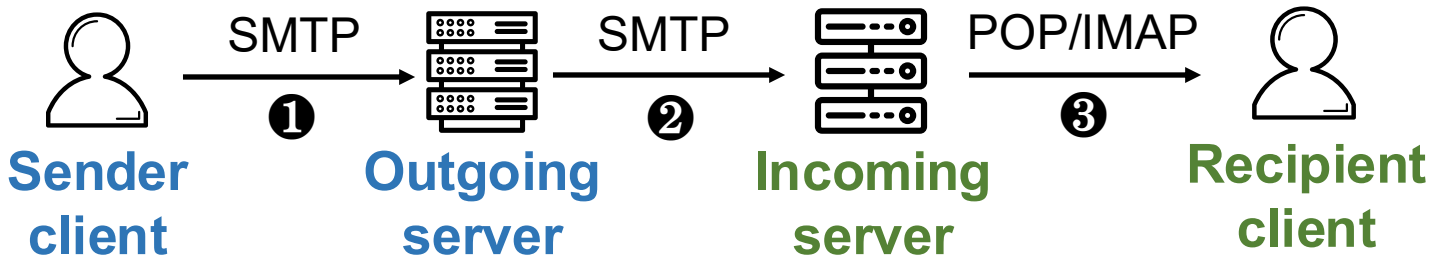
浙江工商大學 ZHEJIANG GONGSHANG UNIVERSITY

[1]*Tsinghua University,* [2]*Coremail Technology Co. Ltd,* [3]*Zhejiang Gongshang University*

Email delivery: “end-to-end” to “segment-to-segment”

Traditional email delivery mode (end-to-end)

Emails are sent directly from the sender’s server to the recipient’s server

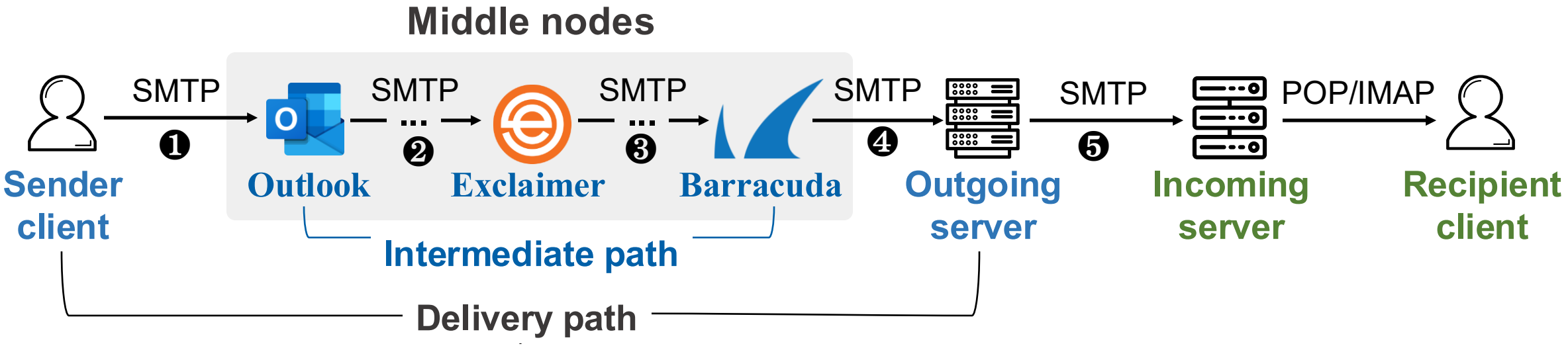


Email delivery: "end-to-end" to "segment-to-segment"

Emerging email delivery mode (segment-to-segment)

In the cloud era, hosting-based email services have become a common business

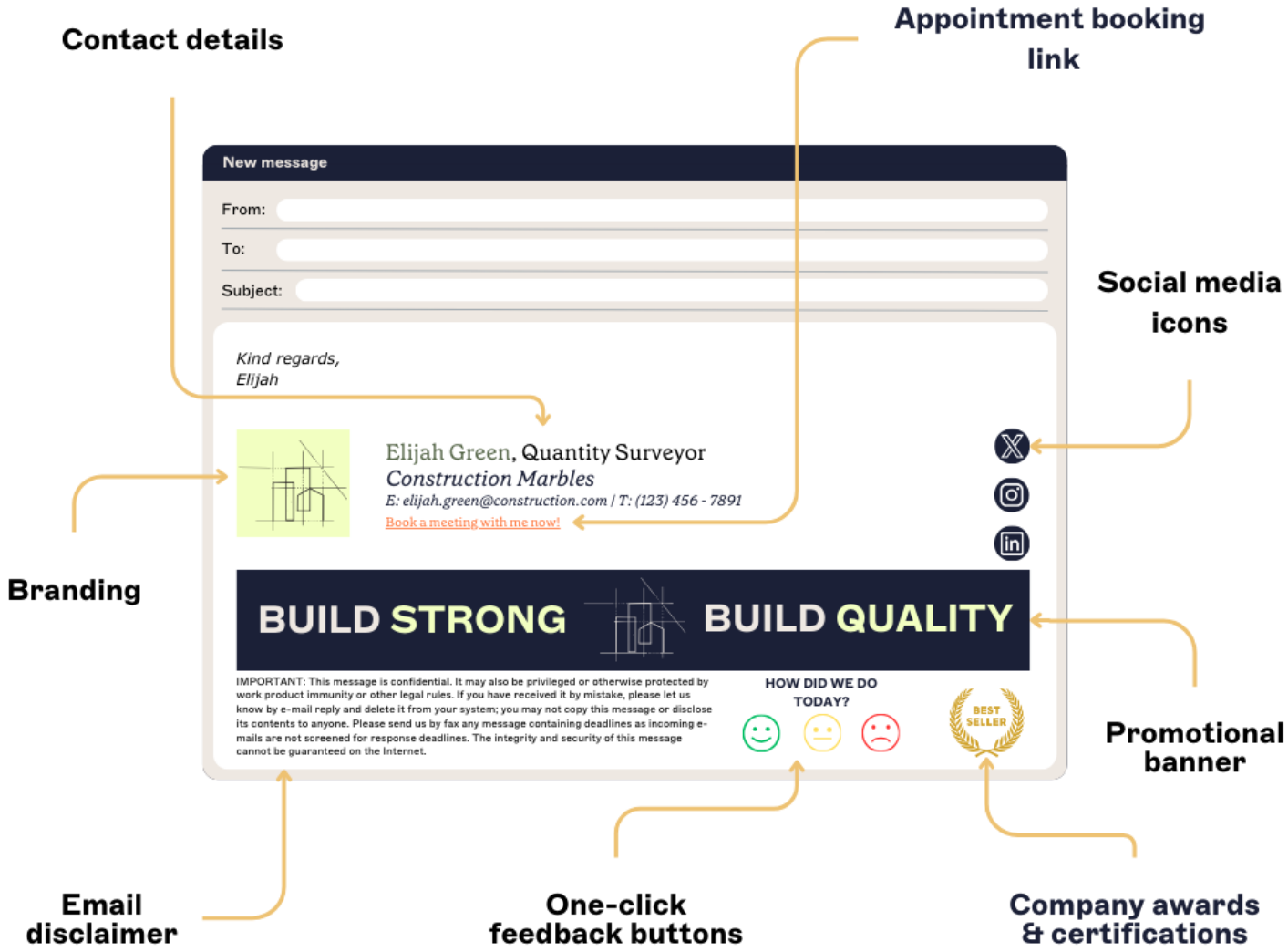
Emails traverse multiple middle nodes: hosting providers, forwarding servers, security vendors, and email signature provider, etc.



Note on email signature provider

- An email signature refers to the consistent content appended to the end of each email body, often including personalized text or graphics such as company logos, job titles, and contact information.

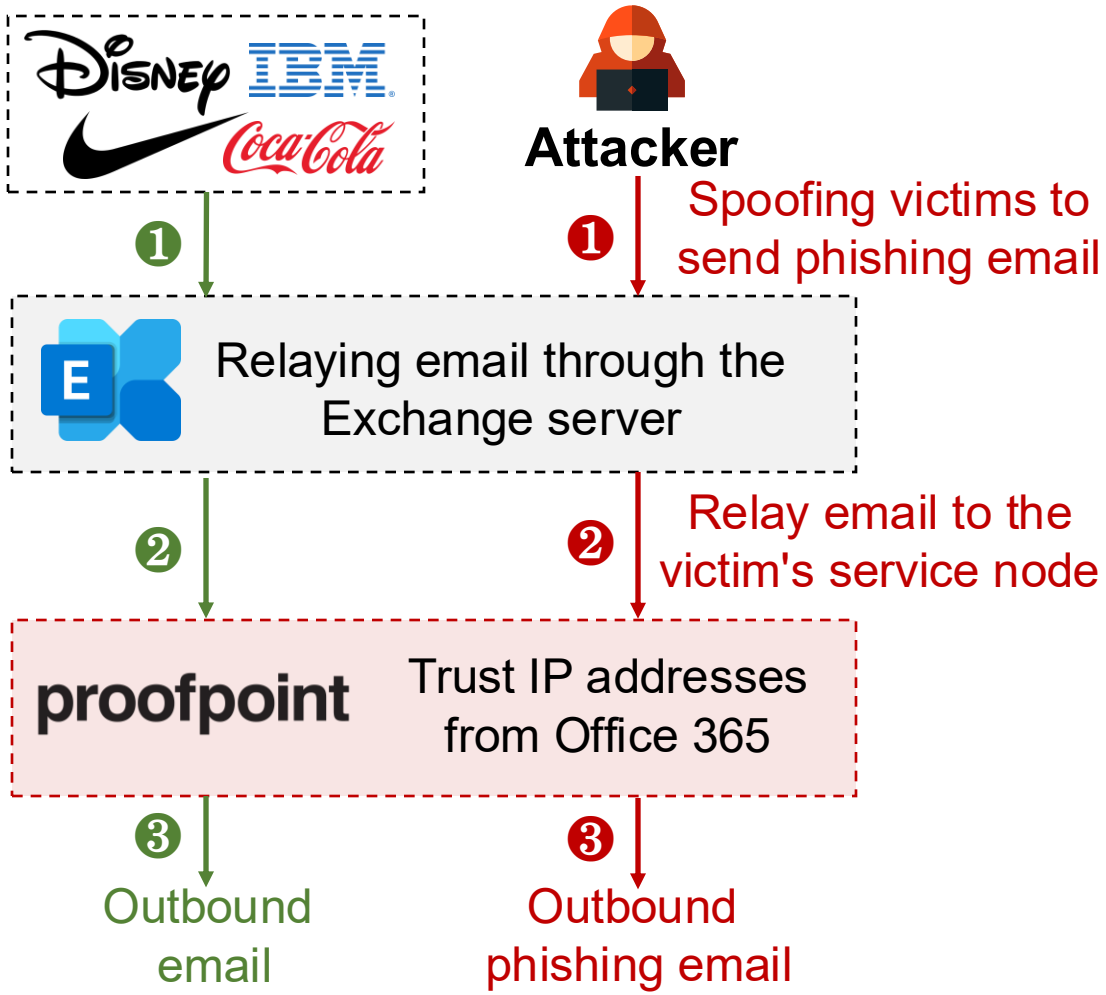
- Companies like Exclaimer and CodeTwo offer brand and signature management for outbound emails of many Fortune 500 companies



Source: https://exclaimer.com/blog/what-is-an-email-signature/#what-makes-up-an-email-signature

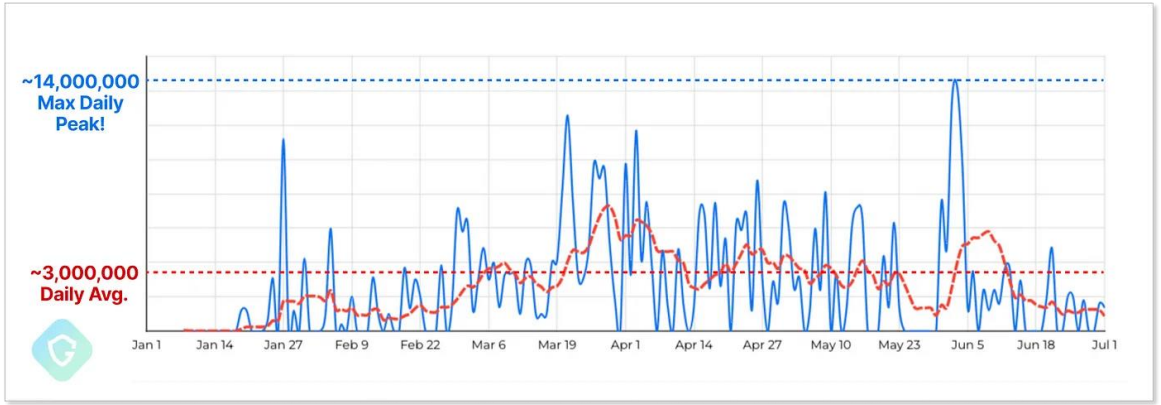
Vulnerable middle nodes affect the security of the entire email delivery path

EchoSpoofing attack: Abuse the lax source verification policies of middle nodes



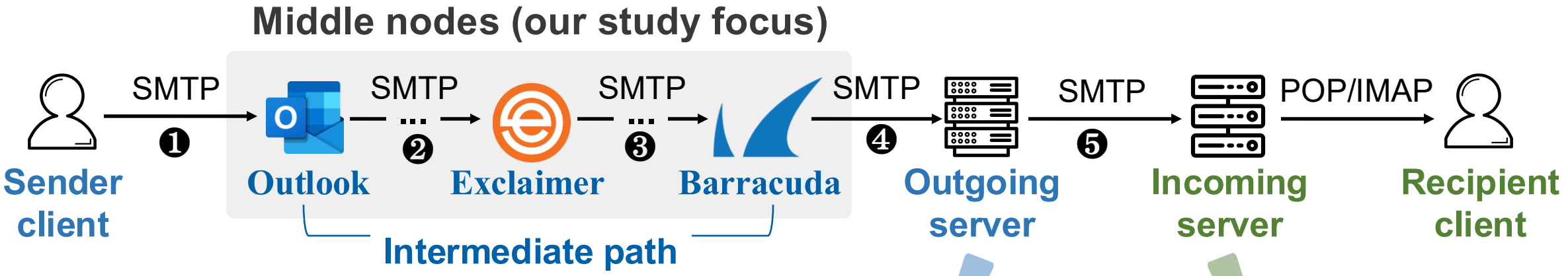
Guardio

The attack campaign began around January 2024 with an average of 3 million perfectly phishing emails per day



Source: https://labs.guard.io/echospoofing-a-massive-phishing-campaign-exploiting-proofpoints-email-protection-to-dispatch-3dd6b5417db6

Previous works focus on incoming and outgoing ends of email delivery paths



What is the landscape of email intermediate paths *(Hidden Dependencies)*

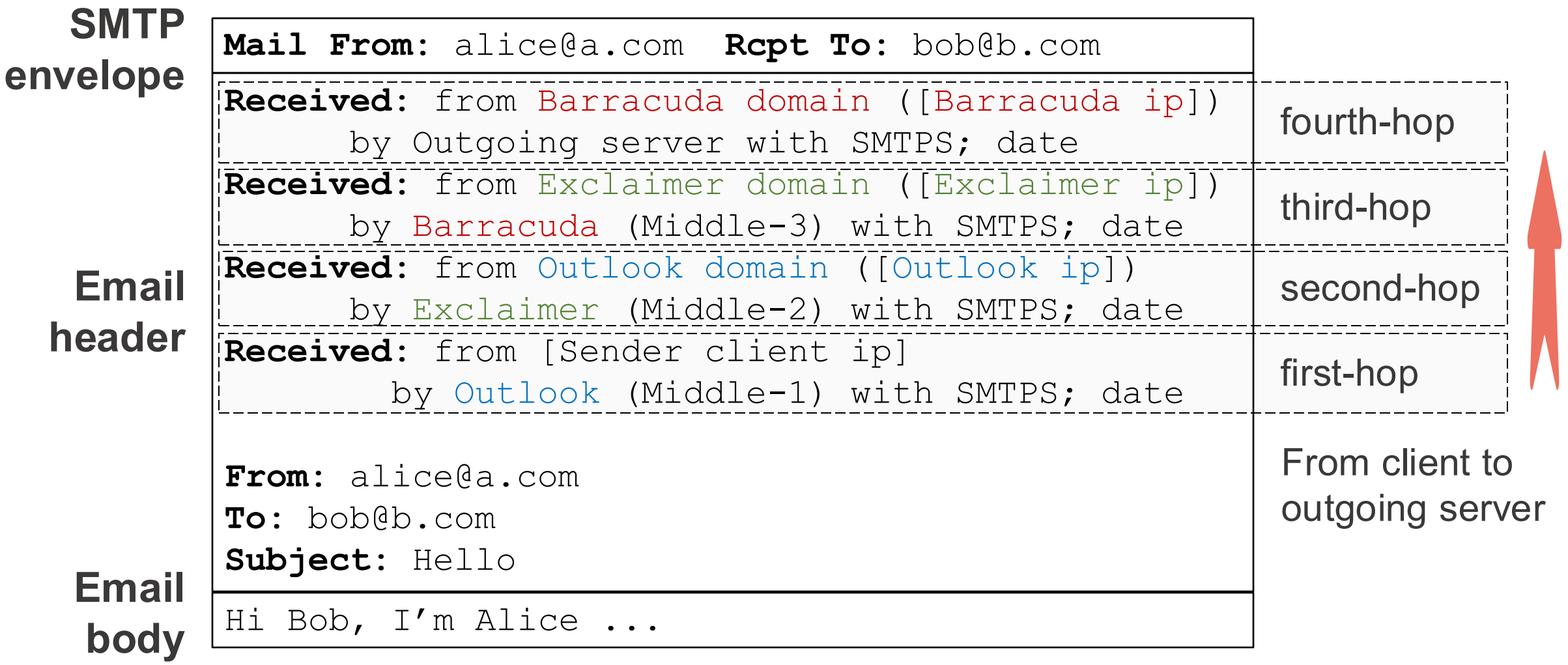
Scan SPF records of domains

Scan MX records of domains

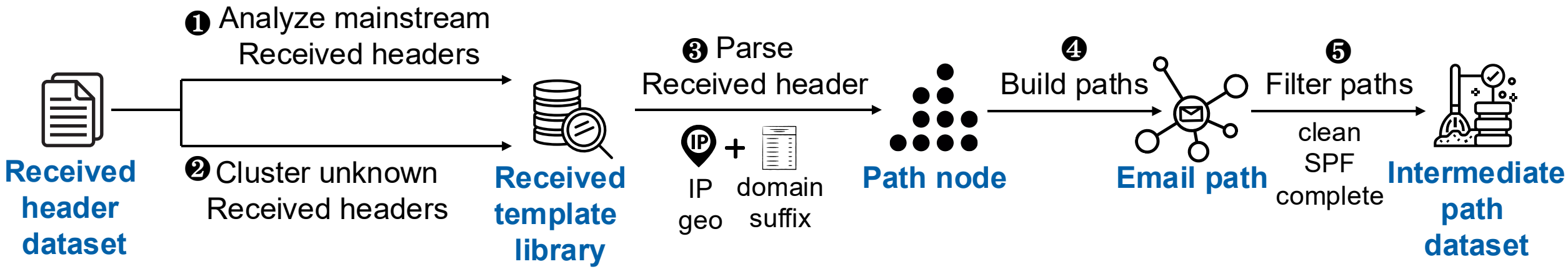
Collect incoming and outgoing nodes

Previous works

We obtain middle nodes through Received headers



Constructing email intermediate path dataset



- Obtain Received header dataset from a large email service provider
- Generate a template library to parse Received headers and extract path nodes
- Build and filter email intermediate path

Received header dataset from Coremail

Coremail

A large email service provider in China, offers email services for more than 20K organizations

Received header dataset example

```
{
  "Mail From": "a.com", "Rcpt To": "b.com", // Only domain
  "Receive_time": "2022-06-14 16:30:35",
  "outgoing_ip": "ip1",
  "Received_headers": {"from xxx by xxx"...}, // Only domain
  "spf": "PASS",
  "email_flag": "Spam"
}
```

- **Time span**：9 months, from May 1, 2024 to November 30, 2024
- **Number of emails**：2,446,933,441 (2.4 billion)

https://www.coremail.ai/, https://www.coremail.cn/

Parse Received headers and extract path nodes

We built a template library with 54 regular expressions, which can match 96.8% of Received headers in our dataset

E.g., `from\s+(?P<from_name>[\w\.\-]+)\s+by\s+(?P<by_name>[\w\.\-]+)\s+with\s+(?P<protocol>\w+)\;`

Received: from Outlook domain by Exclaimer domain with SMTPS;

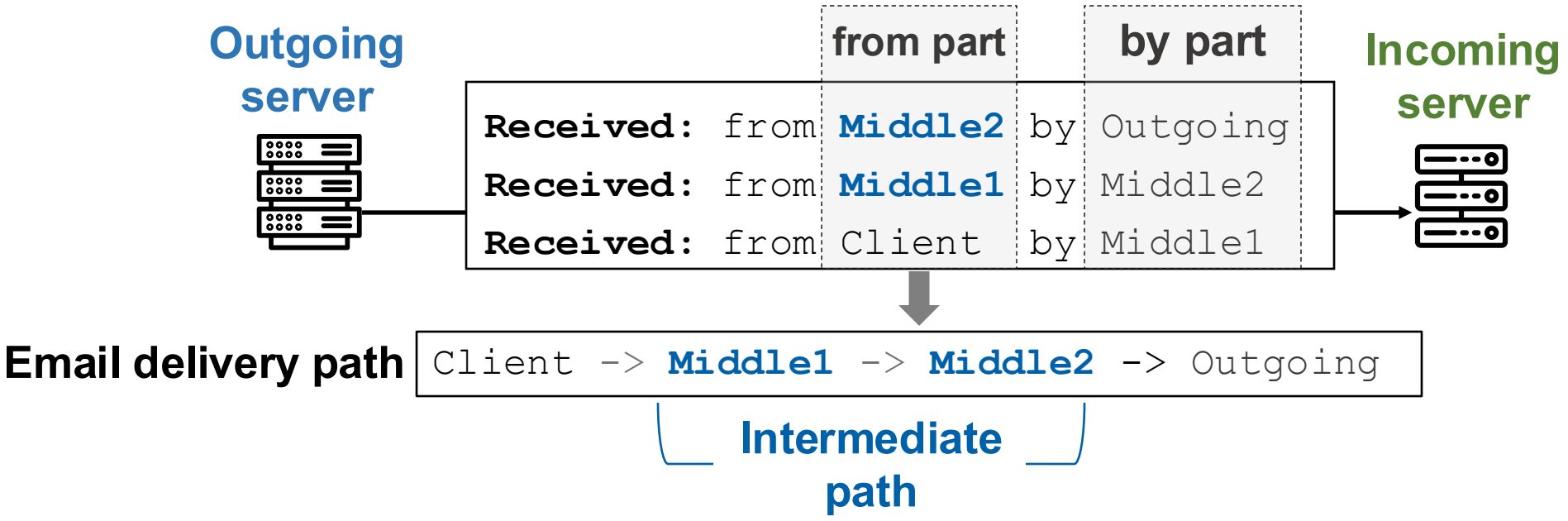
from_name: Outlook domain
by_name: Exclaimer domain
protocol: SMTPS

Path nodes are the IP address and domain name of the from and by parts in each Received header

		from part		by part
Received:	from	**Middle2**	by	Outgoing
Received:	from	**Middle1**	by	Middle2
Received:	from	Client	by	Middle1

Build and filter email intermediate path

Considering that email servers may hide or falsify their identities[1], we use the from part of each Received header to indicate the information of the previous node



Filter dataset: spam, SPF verification failed, without/incomplete email intermediate path

[1] E. Luo, L. Young, G. Ho, M. Afifi, M. Schweighauser, E. Katz-Bassett, and A. Cidon. Characterizing the Networks Sending Enterprise Phishing Emails. PAM 2025.

Overview of email intermediate path dataset and study

- Our intermediate path dataset involves 105 million emails, including 42,478 middle node SLDs and 881,669 middle node IP addresses
- 32.8% of the emails were transmitted exclusively within China ("domestic email"), while the rest were from outside China ("international email")

We aim to unveil the picture of intermediate paths of email delivery, find hidden dependencies, and evaluate the degree of centralization

- ❖ What are the **identities and distribution** of email middle nodes?
- ❖ What is the **dependency structure and regionality** of email intermediate paths?
- ❖ What are the **centralization degrees and cross-country differences** of email intermediate paths?

Distribution of email middle nodes

Most middle nodes belong to ESPs, with outlook.com accounting for more than half of the emails

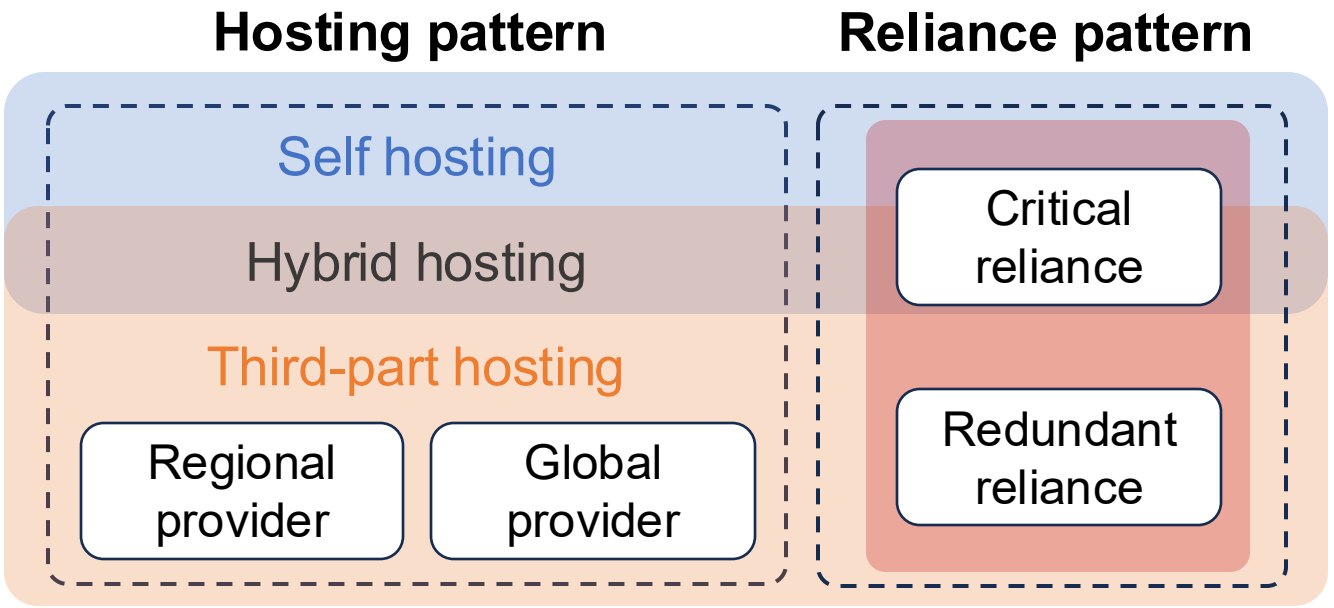
Top 10 providers of middle nodes with high sender SLD dependencies

Top 10 providers	Type	# SLD	# Email
outlook.com	ESP	51.5%	66.4%
exchangelabs.com	ESP	4.4%	4.6%
icoremail.net	ESP	2.3%	0.4%
yandex.net	ESP	1.7%	0.5%
exclaimer.net	**Signature**	1.6%	1.3%
google.com	ESP	1.4%	0.6%
codetwo.com	**Signature**	1.2%	0.8%
qq.com	ESP	0.5%	0.2%
aliyun.com	ESP	0.4%	0.2%
secureserver.net	**Security**	0.4%	0.1%

Suggestion

We suggest that future work conduct in-depth analyses of previously underexplored middle nodes in email delivery paths, focusing on their operational roles and potential implications for security and resilience in global email infrastructure

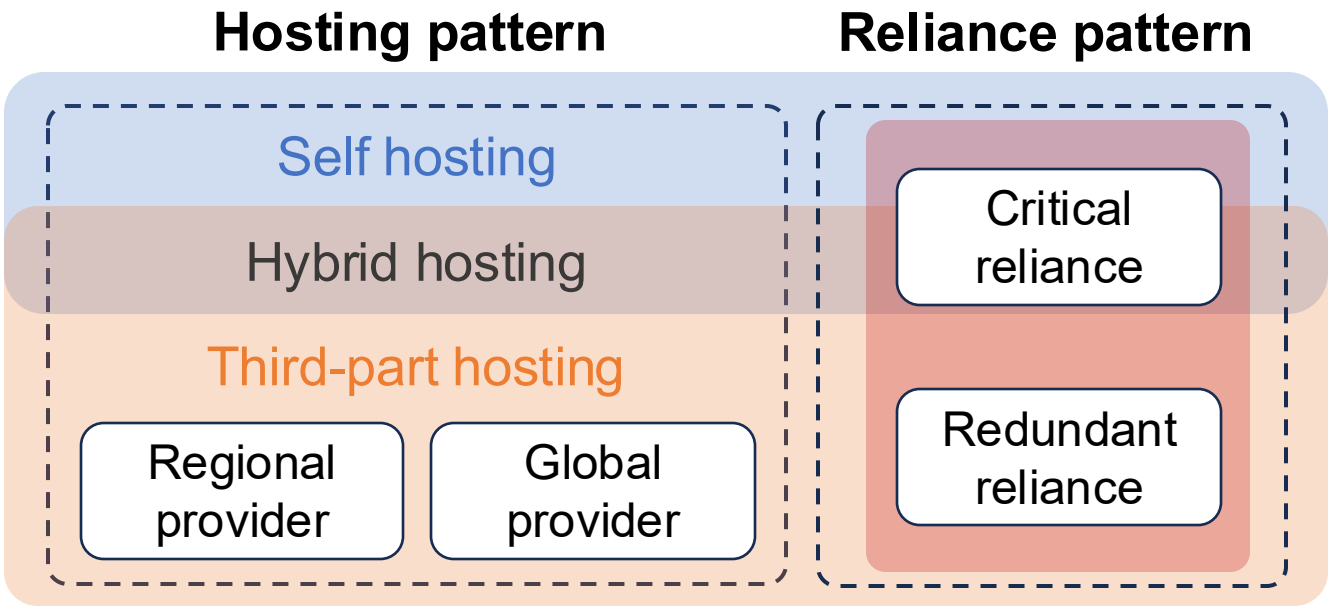
Dependency pattern of email intermediate paths



❖ **Hosting pattern** describes the relationship between middle nodes and the sender domain, reflecting the extent to which a domain relies on third-party providers

- **Self-hosting:** Domain uses its own infrastructure to handle the email intermediate path
- **Third-party hosting:** Email intermediate path is completely dependent on third-party providers
- **Hybrid hosting:** Email intermediate path involves both self-hosted and third-party infrastructure

Dependency pattern of email intermediate paths



❖ **Reliance pattern** refers to the number of distinct providers involved in an email intermediate path, reflecting the complexity of the intermediate path

- ➢ **Single reliance:** Email intermediate path involves only one provider
- ➢ **Multiple reliance:** Email intermediate path involves multiple providers

Overview of dependency patterns of intermediate paths

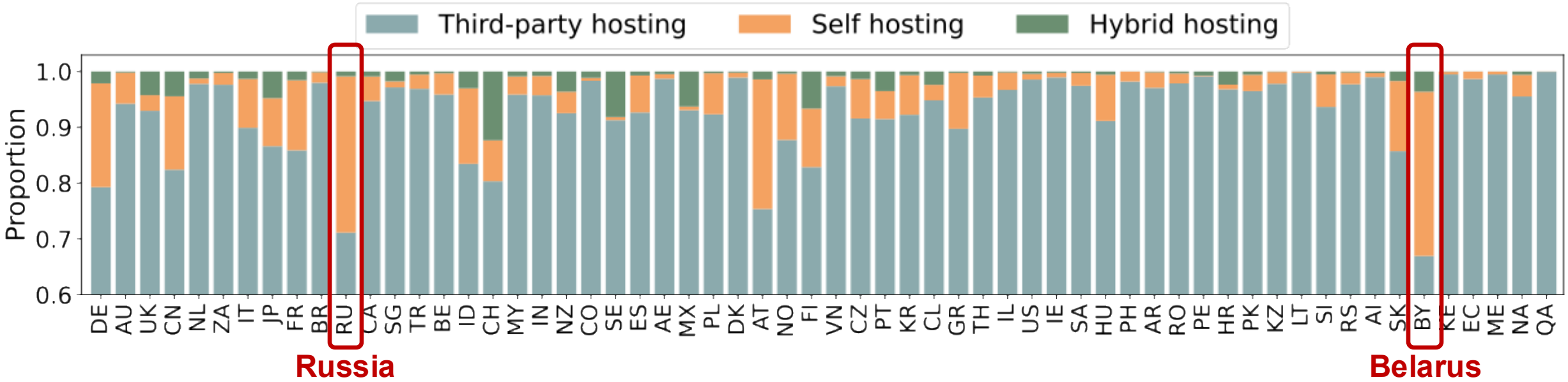
The email Intermediate path is largely dominated by third-party hosting providers

In most cases, a single vendor handles the majority of intermediate relays.

	# SLD	# Email
Hosting pattern		
Self hosting	17.7K (4.3%)	15.1M (14.3%)
Third-party hosting	**399.1K (96.8%)**	**86.9M (82.7%)**
Hybrid hosting	7.5K (1.8%)	3.2M (3.0%)
Reliance pattern		
Single reliance	**384.5K (93.3%)**	**96.0M (91.3%)**
Multiple reliance	52.8K (12.8%)	9.1M (8.7%)

Dependency patterns of country domains

The proportion of Third-party hosting in email intermediate paths for various countries exceeds 60%, highlighting the email dependency on hosting providers



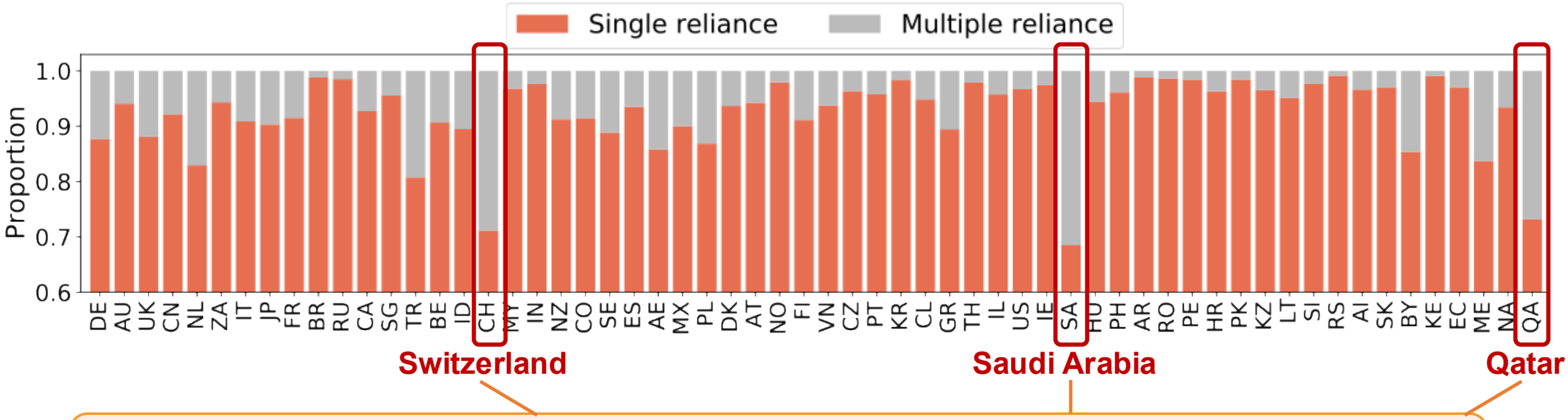
Intermediate paths from Russia and Belarus exhibit the Self hosting proportion of about 30%

"Following the Russia-Ukraine conflict, Russia reduced its Internet dependency (e.g., DNS and PKI) on foreign hosting services[1]."

[1] M. Jonker, G. Akiwate, A. Affinito, k. claffy, A. Botta, G. Voelker, R. Rijswijk-Deij, and S. Savage. Where .ru?: assessing the impact of conflict on russian domain infrastructure. ACM IMC 2022.

Dependency patterns of country domains

The majority of countries' email intermediate paths rely on a single provider, with the proportion of Single reliance typically exceeding 80%

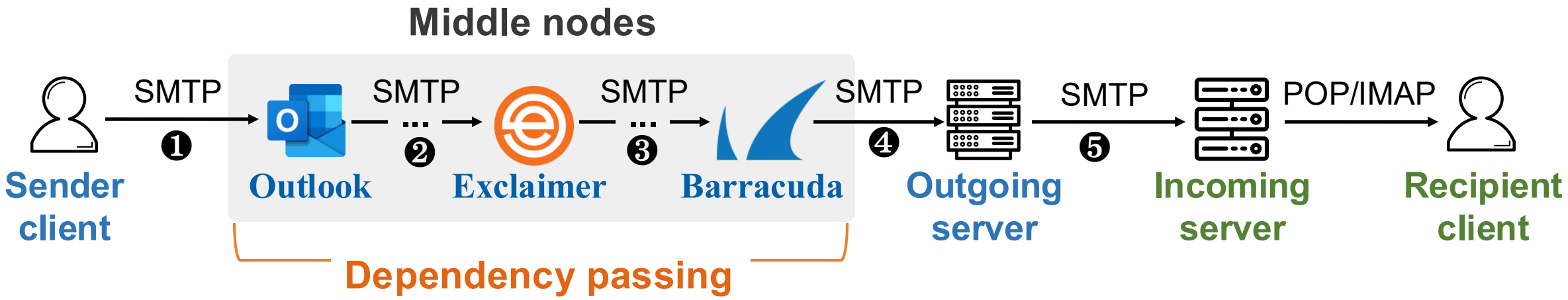


For Switzerland, Saudi Arabia and Qatar, the proportion of Multiple reliance exceeds 30%

This is primarily due to the inclusion of email signatures and security filtering providers in intermediate paths.

Dependency passing in email intermediate paths

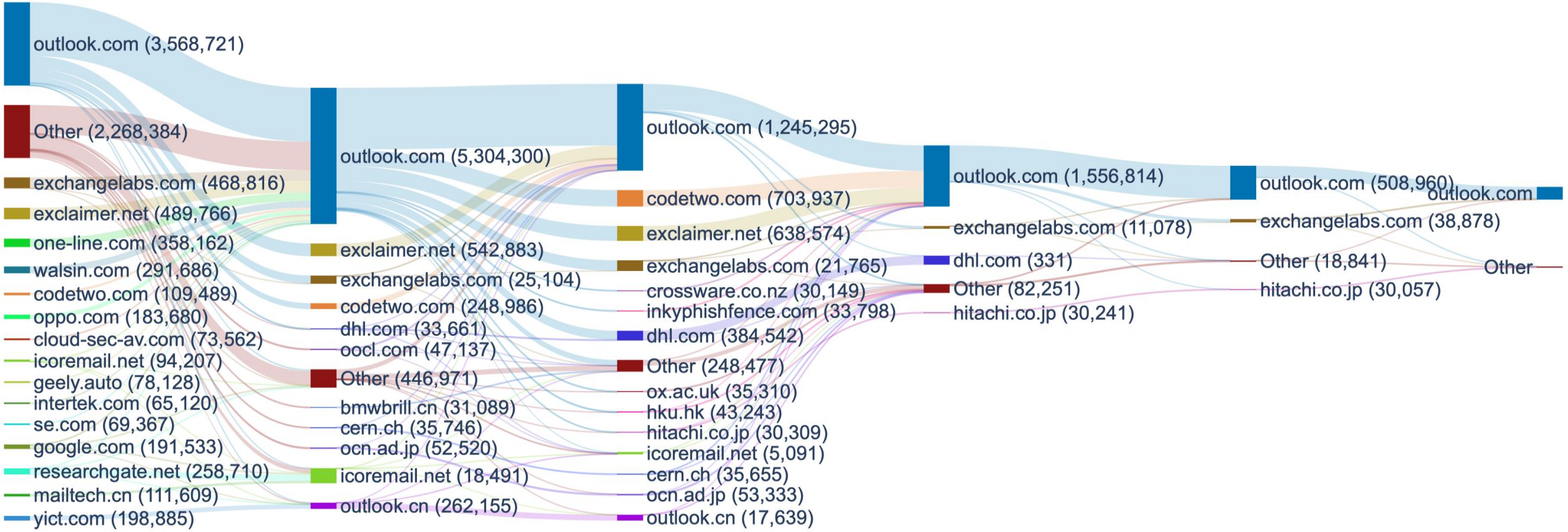
The email intermediate path involves **different SLDs**, meaning that **dependencies are passed between various suppliers**, such interactions may harbor potential security risks (e.g., EchoSpoofing)



- We analyze the dependency passing in the 9.1 million multiple dependency intermediate path
- If two email intermediate paths contain the same set of middle node SLDs (regardless of order), we consider them to belong to the same dependency passing relationship

Analyze dependency passing relationship

- In total, we identify 28,359 distinct dependency passing relationships, among which 55.8% involve two SLDs, 25.8% involve three SLDs, and 18.4% involve more than three SLDs
- In email intermediate paths of each hop, **a significant proportion of the emails rely on outlook.com for transmission**



Analyze dependency passing relationship

- In total, we identify 28,359 distinct dependency passing relationships, among which 55.8% involve two SLDs, 25.8% involve three SLDs, and 18.4% involve more than three SLDs
- In email intermediate paths of each hop, **a significant proportion of the emails rely on outlook.com for transmission**
- Most prevalent dependency passing occurs between email service providers (ESP) and email signature providers ("ESP-Signature")

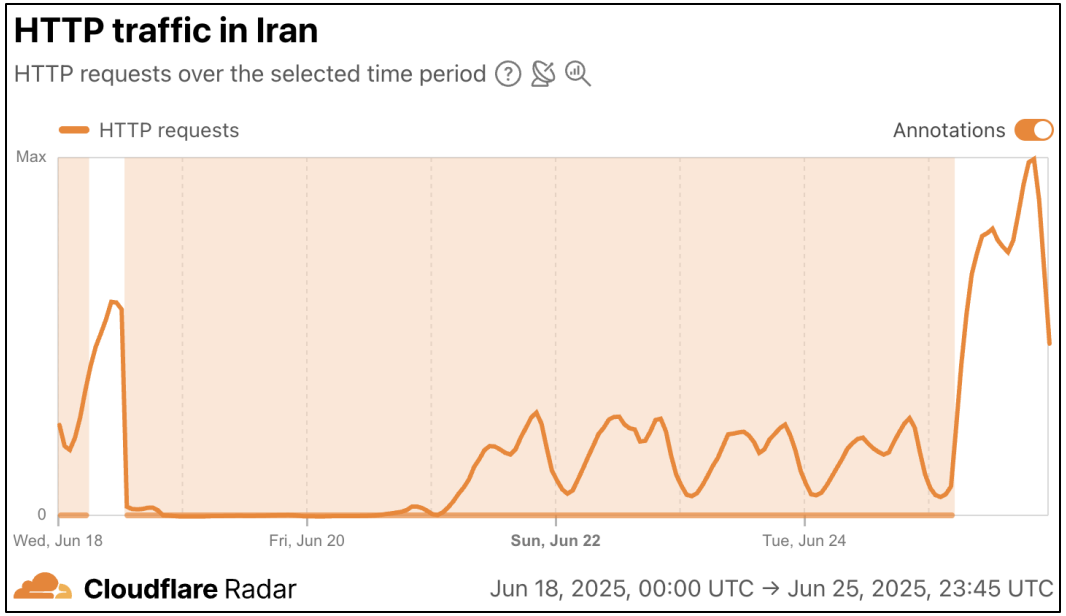
Type	# SLD	# Email
ESP-Signature	**16.4K (31.2%)**	**2.7M (29.7%)**
ESP-ESP	8.3K (15.8%)	1.2M (13.3%)
ESP-Security	2.8K (5.4%)	237.8K (2.6%)
ESP-Signature-ESP	1.5K (2.9%)	192.1K (2.1%)
ESP-Security-ESP	950 (1.8%)	146.3K (1.6%)
ESP-Signature-Security	580 (1.1%)	82.3K (0.9%)

Regional dependency of email intermediate paths

We focus on analyzing the dependence of email intermediate paths from domain of different countries or continents on external regions

Suggestion

We suggest that stakeholders pay closer attention to critical points of dependency along intermediate paths, as they may pose significant risks of service disruption under geopolitical tensions or cross-border regulatory shifts



In June 2025, Iranian blocked access to the Internet, affecting network services that depended on it

Source: https://blog.cloudflare.com/zh-cn/q2-2025-internet-disruption-summary/

Regional dependency across countries

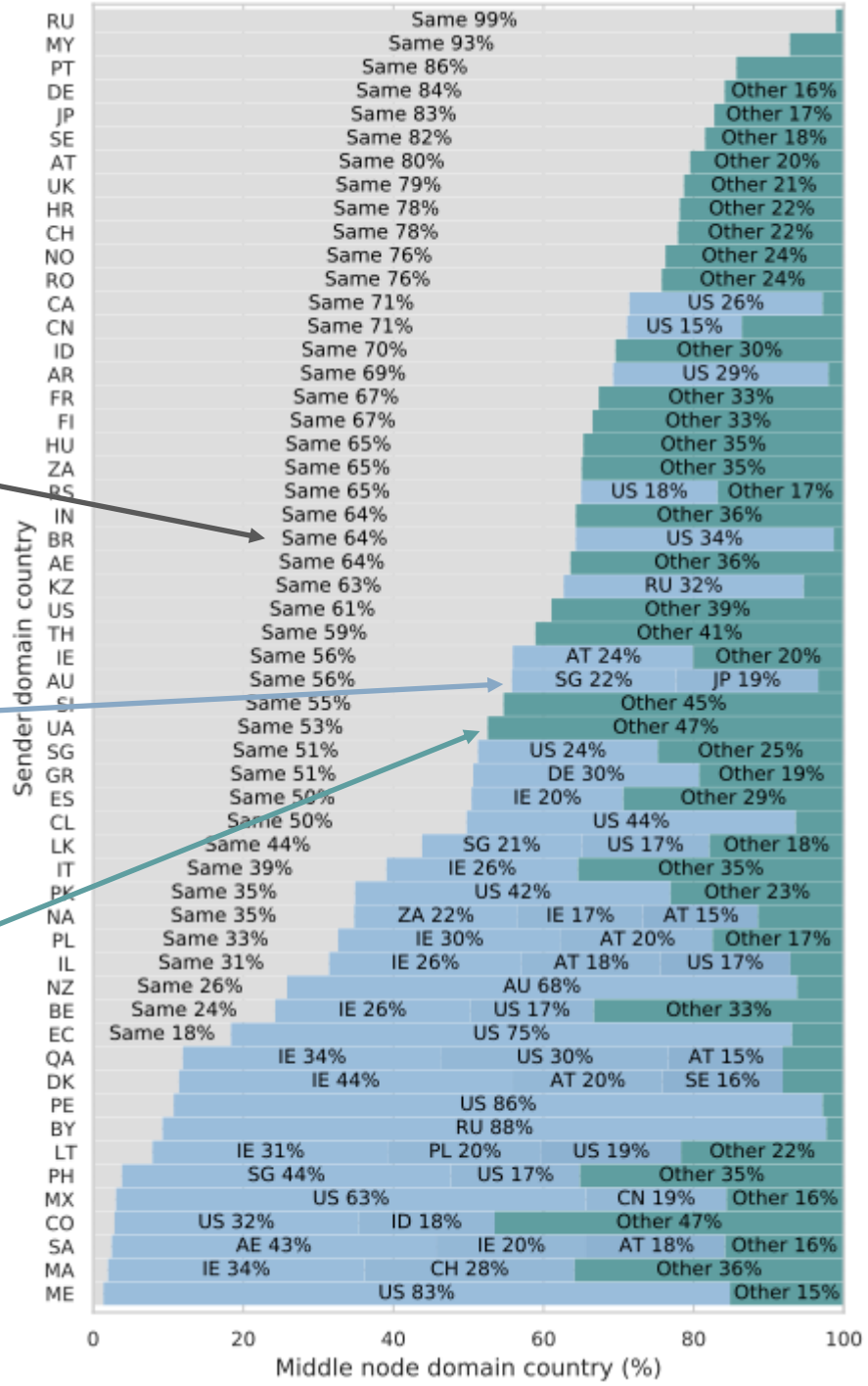
Same

Email middle nodes belong to the same country as the sender domain

xx% of email intermediate paths from sender domains in a country depend on another country

Other

Countries accounting for less than 15% are grouped under "Others"



Type of regional dependency

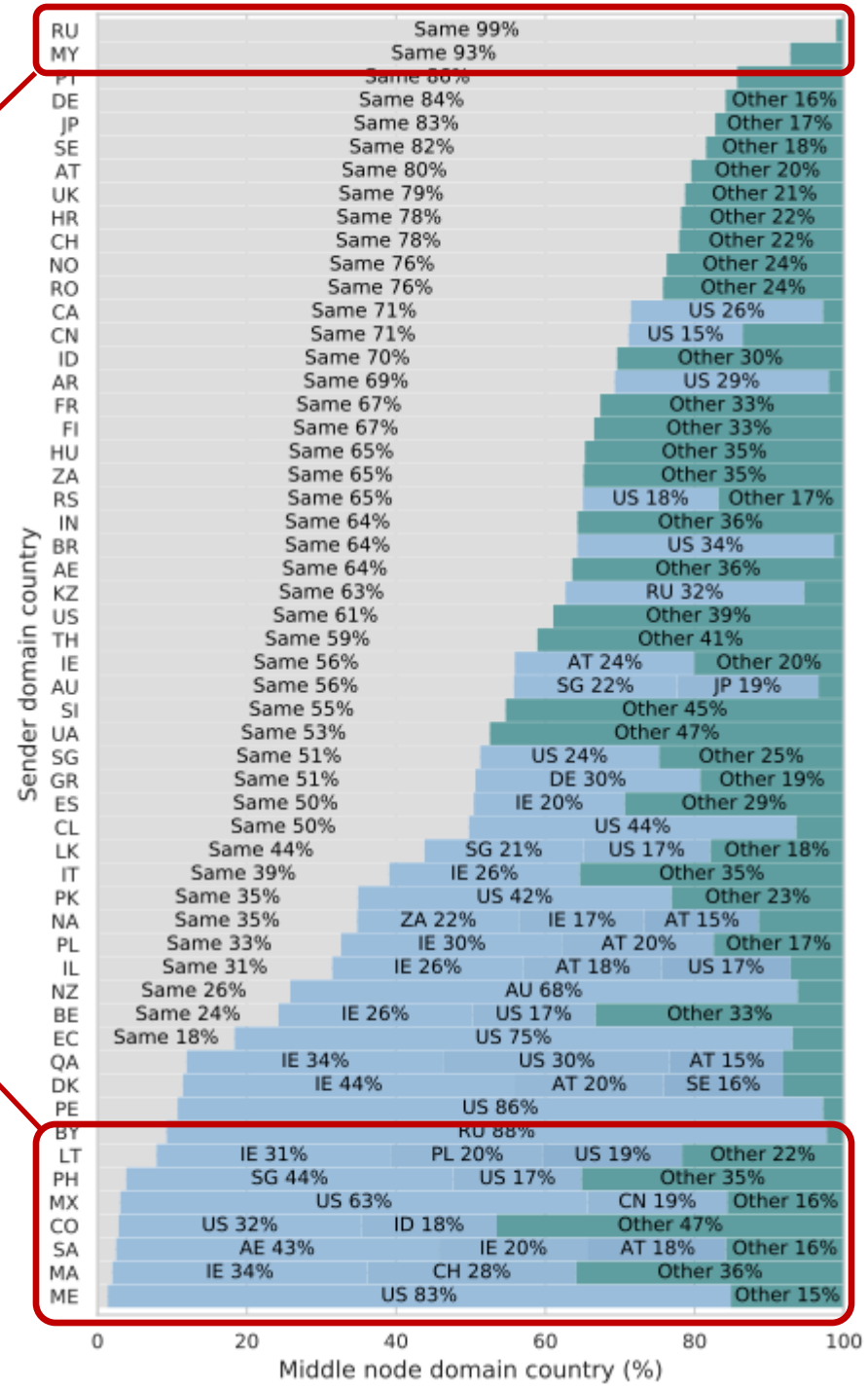
Regional dependency patterns vary across countries

- **High dependence on domestic infrastructure**

  Such as: Russia and Malaysia (>90%)

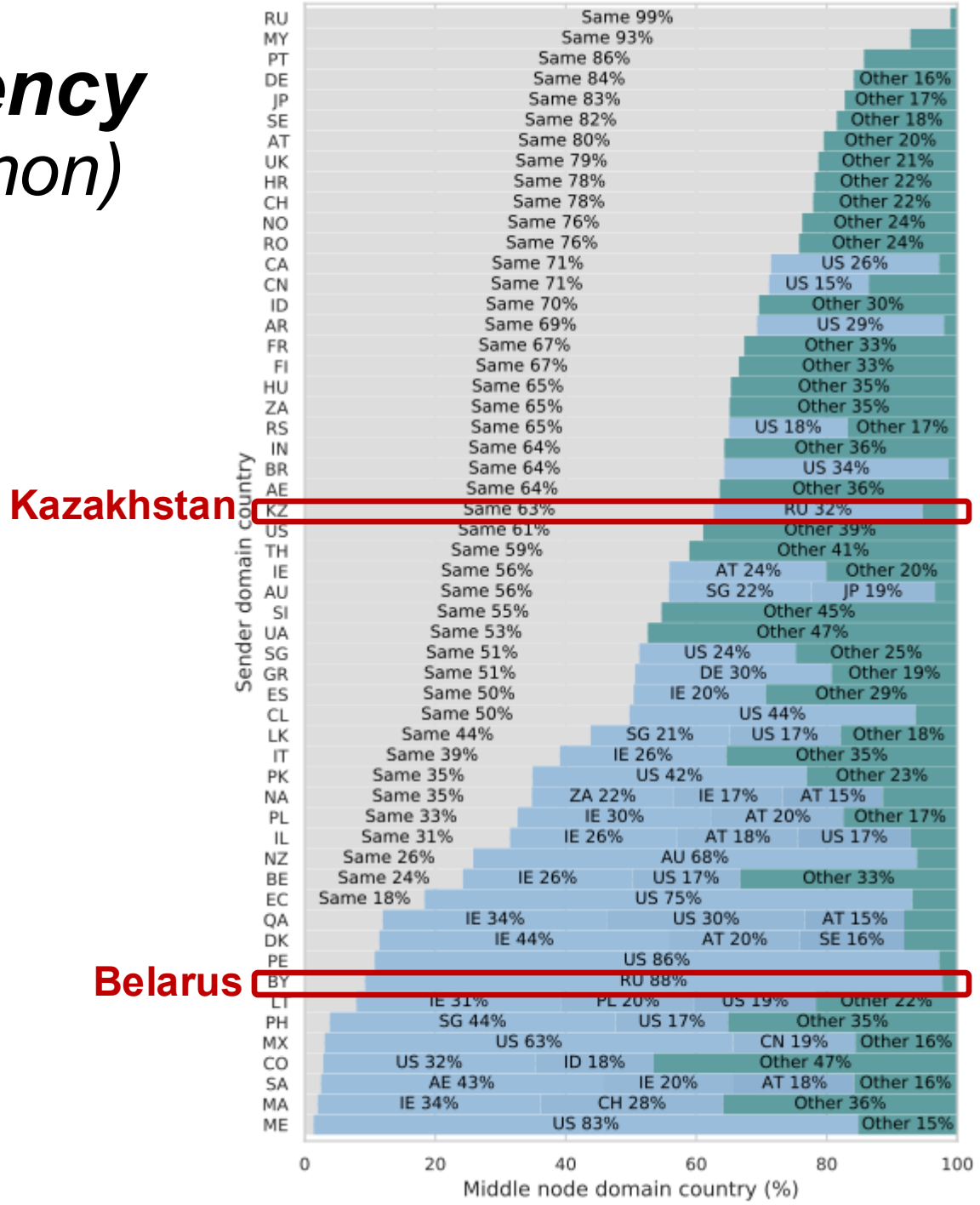
- **High dependence on foreign infrastructure**

  Such as: Montenegro and Morocco (>90%)



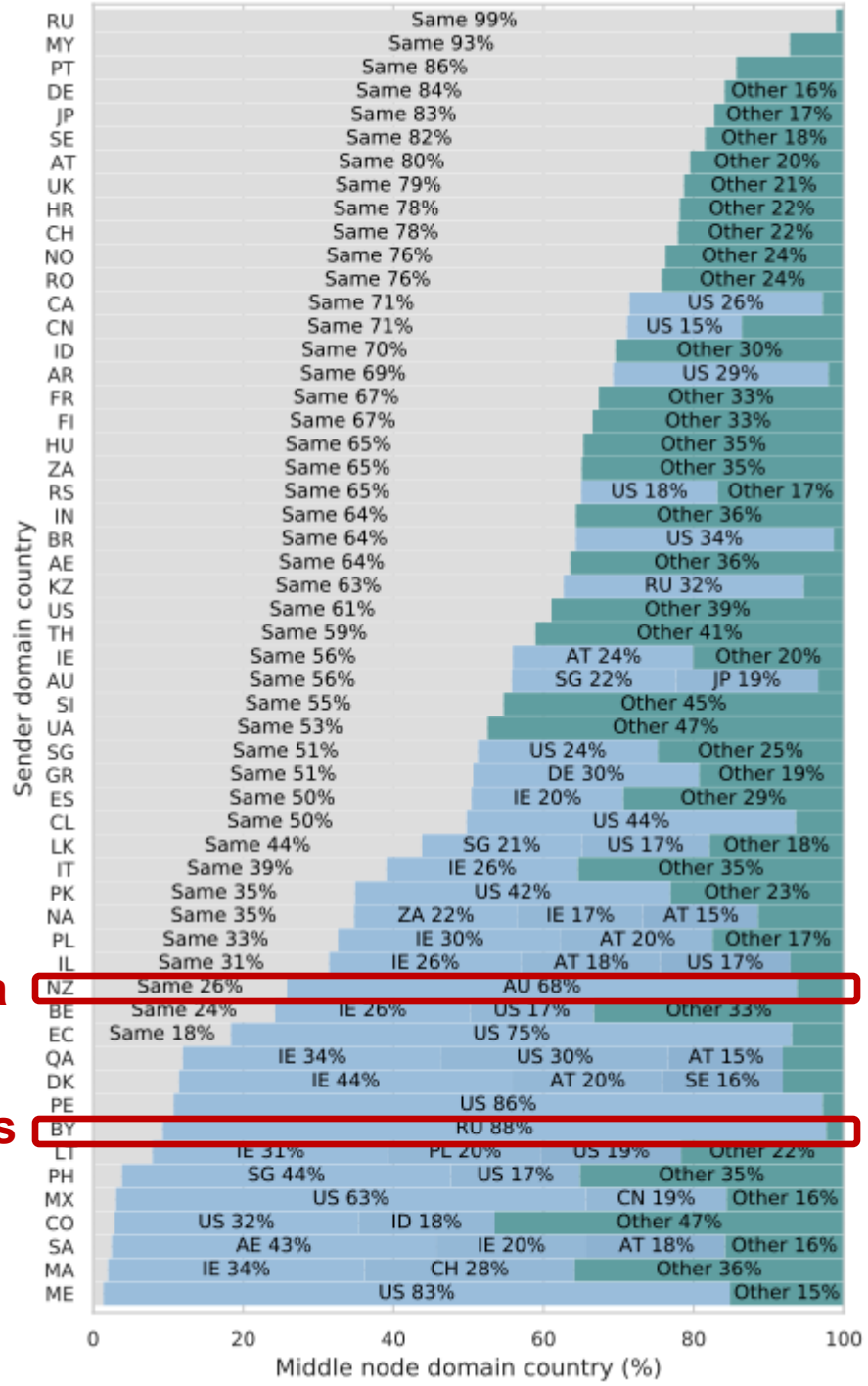
Reasons of regional dependency
(We try to infer from the phenomenon)

- Countries belonging to the Commonwealth of Independent States (CIS), formed after the collapse of the Soviet Union, significantly rely on Russia's email infrastructure



Reasons of regional dependency
(We try to infer from the phenomenon)

- Email intermediate paths often reflect regional dependencies between geographically proximate or linguistically similar countries

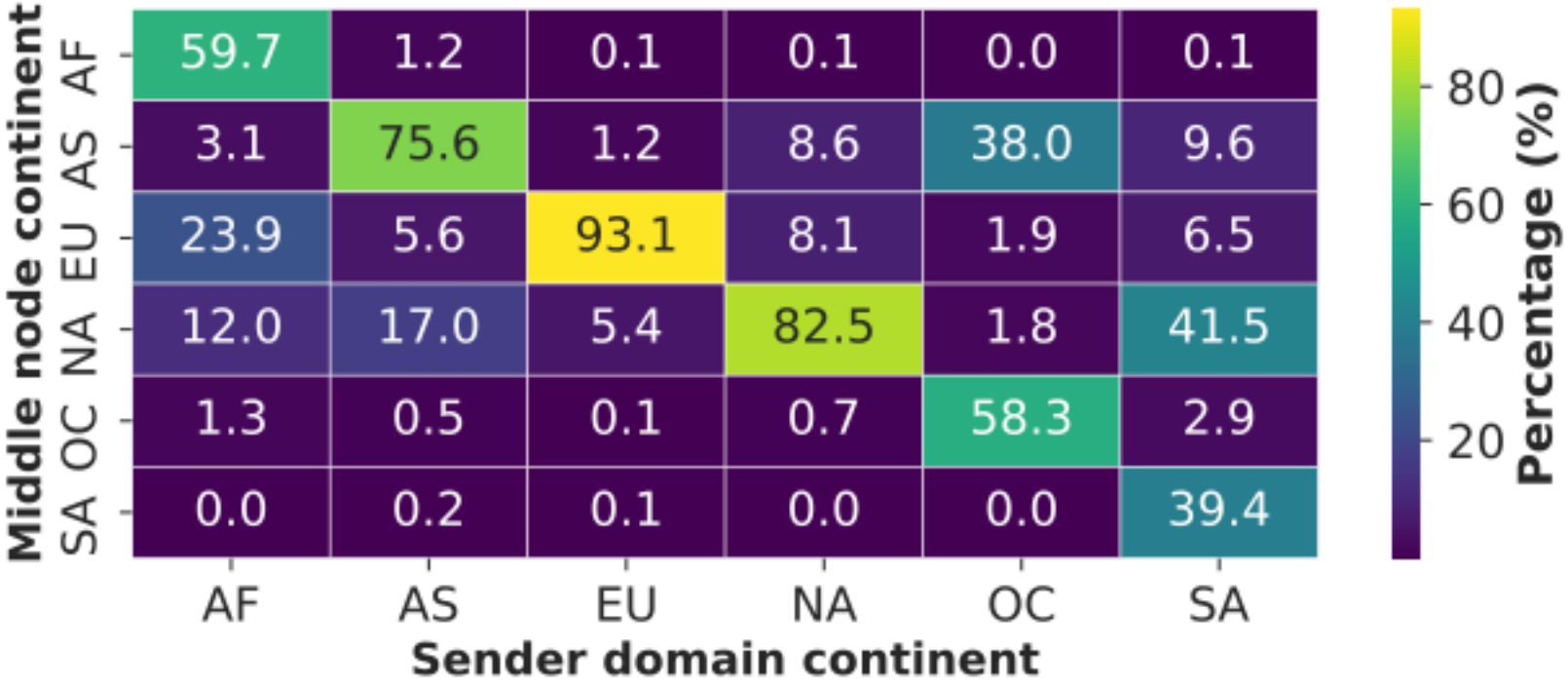


New Zealand depends on Australia

Saudi Arabia depends on United Arab Emirates

Regional dependency across continents

- The majority of emails originating from Asia, Europe, and North America have middle nodes located **within the same continent**, with Europe accounting for as much as 93.1%
- Email intermediate paths from **Africa heavily depend on Europe and North America**, while those from **South America are highly dependent on North America**

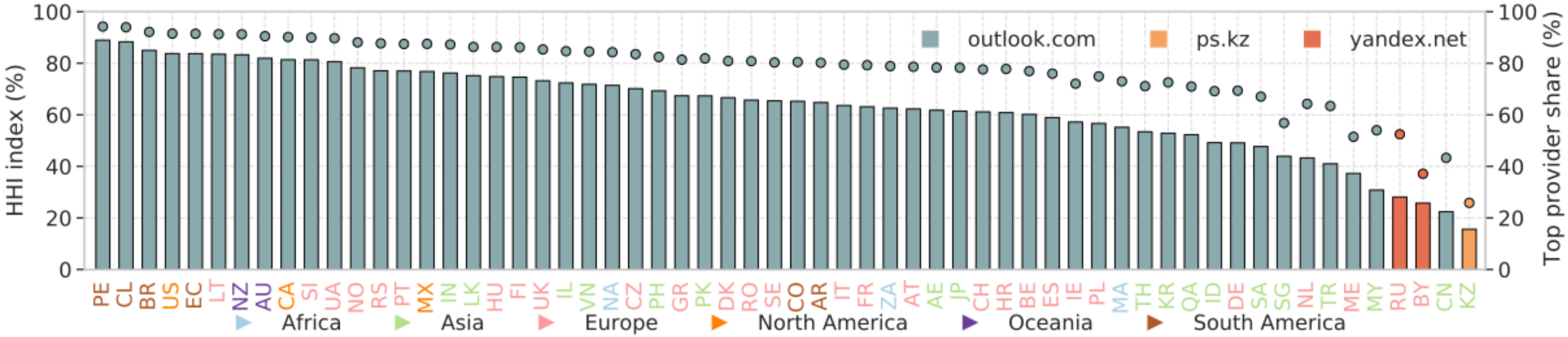


Overview of centralization of email intermediate paths

- We use the Herfindahl-Hirschman Index (HHI) to evaluate the market concentration of email middle nodes
- A higher HHI indicates a more monopolistic market structure: an HHI of 10% indicates moderate concentration, while a value above 25% indicates high concentration
- **Considering all email intermediate paths, we obtain an HHI of 40% for the middle node market, which indicates a highly concentrated market**
- **Microsoft dominates the overall email middle node market, participating in about 70% of email intermediate paths**

Centralization of intermediate paths across countries

- **HHI varies greatly between countries**: Peru is 88% and Kazakhstan is 16%
- **outlook.com dominates the market share in most countries, typically exceeding 60%**.
- yandex.net is the primary provider in Russia and Belarus. In the case of Kazakhstan, ps.kz, a local cloud service provider, holds 26% of the intermediate email path market.



Sort in descending order according to the value of HHI,
and mark the supplier with the highest market share in a country with a circle

Conclusion

- Using a unique and large-scale industrial email dataset, we **unveil middle nodes** and **intermediate paths of email delivery**, one missing piece from previous studies

- We systematically analyze **hidden dependencies** and evaluate the **centralization degree** of email intermediate paths

- We **publish our email path extractor and intermediate path dataset** (at https://github.com/RUI-XUAN-LI/Email_Path) for facilitating future research

Thanks for Listening!

Ruixuan Li

Email: *lirx25@mails.tsinghua.edu.cn*

Website: *https://ruixuanli.com/*

清華大學 Tsinghua University

Coremail

浙江工商大學 ZHEJIANG GONGSHANG UNIVERSITY