ACM Internet Measurement Conference 2024

Bounce in the Wild: A Deep Dive into **Email Delivery Failures** from a Large Email Service Provider

Ruixuan Li[1], Shaodong Xiao[2], Baojun Liu[1], Yanzhong Lin[3], Haixin Duan[1],

Qingfeng Pan[3], Jianjun Chen[1], Jia Zhang[1], Ximeng Liu[2], Xiuqi Lu[4], Jun Shao[5]

Tsinghua University

FUZHOU UNIVERSITY

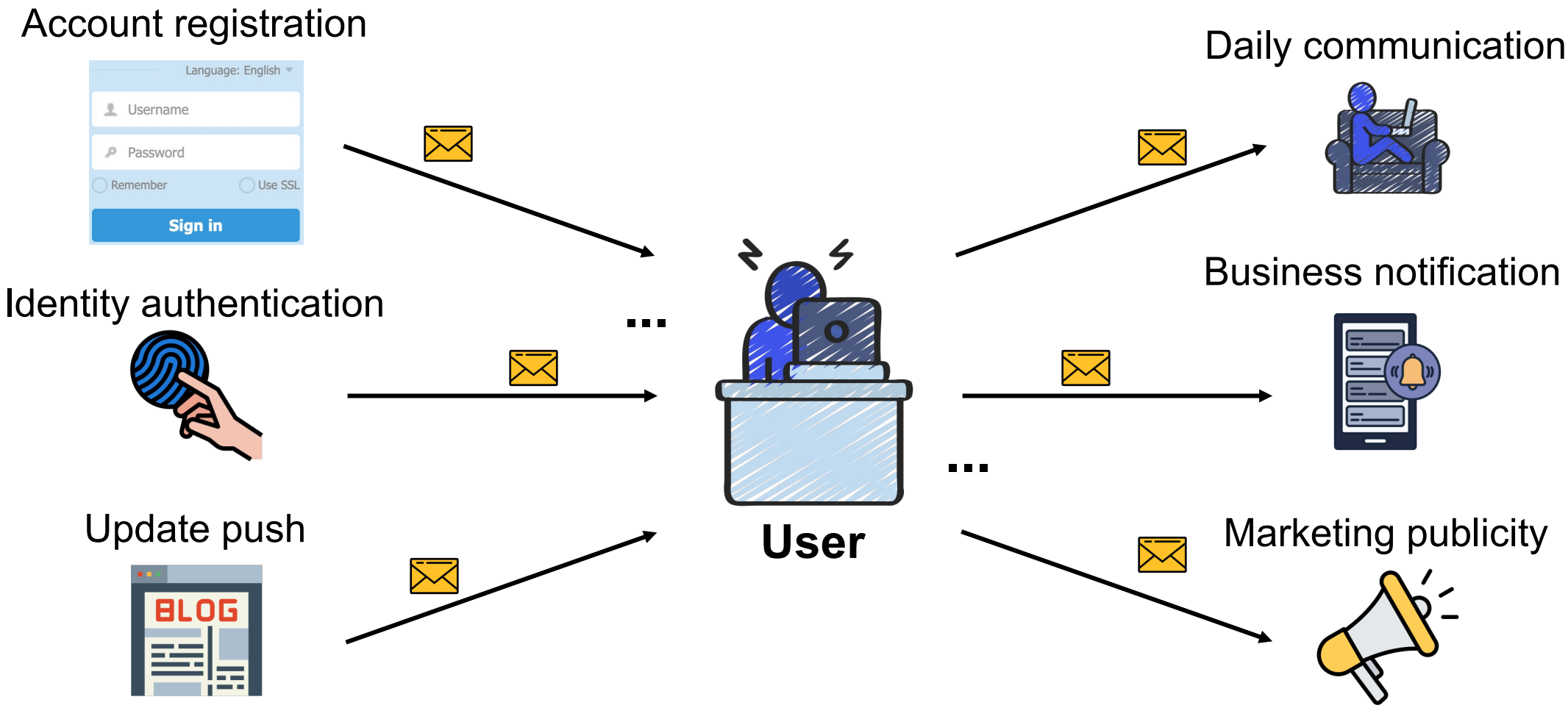
Coremail

華中科技大學 HUAZHONG UNIVERSITY SCIENCE AND TECHNOLOGY

浙江工商大學 ZHEJIANG GONGSHANG UNIVERSITY

[1]Tsinghua University, [2]Fuzhou University, [3]Coremail Technology Co. Ltd,
[4]Huazhong University of Science and Technology, [5]Zhejiang Gongshang University

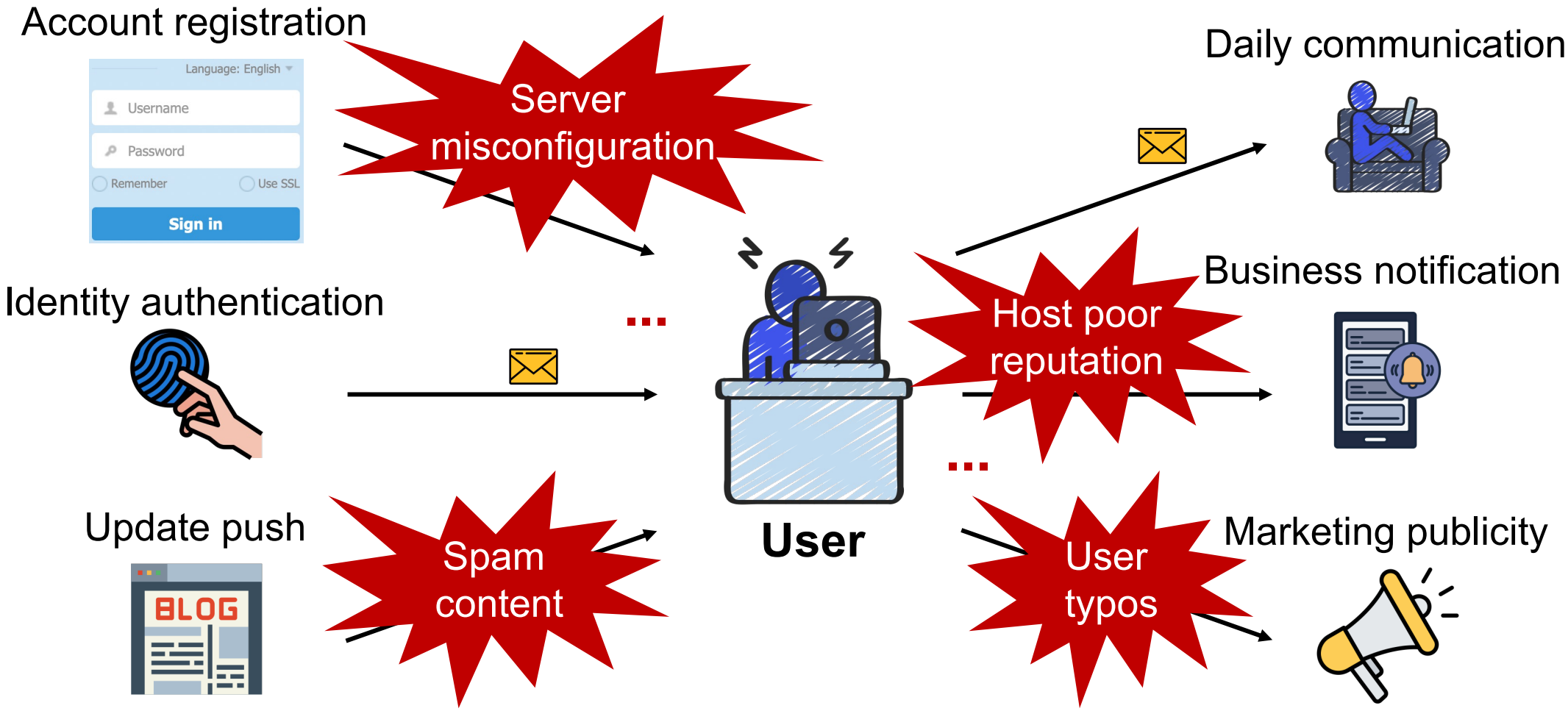
Email delivery is everywhere



Email users exceeded 4.3 billion and are expected to reach 4.84 billion by 2027[1]

[1]https://www.emailisnotdead.com/

Email delivery failure is everywhere



Email delivery failure is everywhere

Microsoft | Community Products Get Started Buy Microsoft 365

Search within Outlook

OG

Created on July 23, 2020

550 5.7.1 Service unavailable, Client host [46.140.66.xxx] blocked using Spamhaus

Microsoft | Community Products Get Started Buy Microsoft 365

Search within Microsoft 365 and Office

BR

Created on August 6, 2020

550 5.4.1 Recipient address rejected: Access denied. AS(201806281) [DB5EUR01FT018.eop-EUR01.prod.protection.outlook.com]

Google Account Help | Describe your issue

"Delivery Status Notification (Failure)" and 552 5.2.2 email over quota but have 9.09G of 15G left

Sep 22, 2022

550 5.1.1 The email account that you tried to reach does not exist. Please try double-checking the r

Gmail Help | Describe your issue

May 31, 2023

This mail is unauthenticated, which poses a security risk to the ender and Gmail users, and has been

Gmail Help | Describe your issue

May 26, 2021

550-5.7.28 Our system has detected an unusual rate of unsolicited mail

Quora

Why can't I receive an email reply from a friend in China? I use 163.com as my email account. My friend received my email but I cannot receive email from this friend.

Answer Follow · 3 Request

Microsoft

Search within Out

AR

Created on October 13, 2017

550 5.7.1 Unfortunately, messages from weren't sent. Please contact your Internet service provider since part of their network is on our block list (AS3150).

What happens to email during delivery

There are three types of email delivery status

First delivery is success (*non-bounced*)

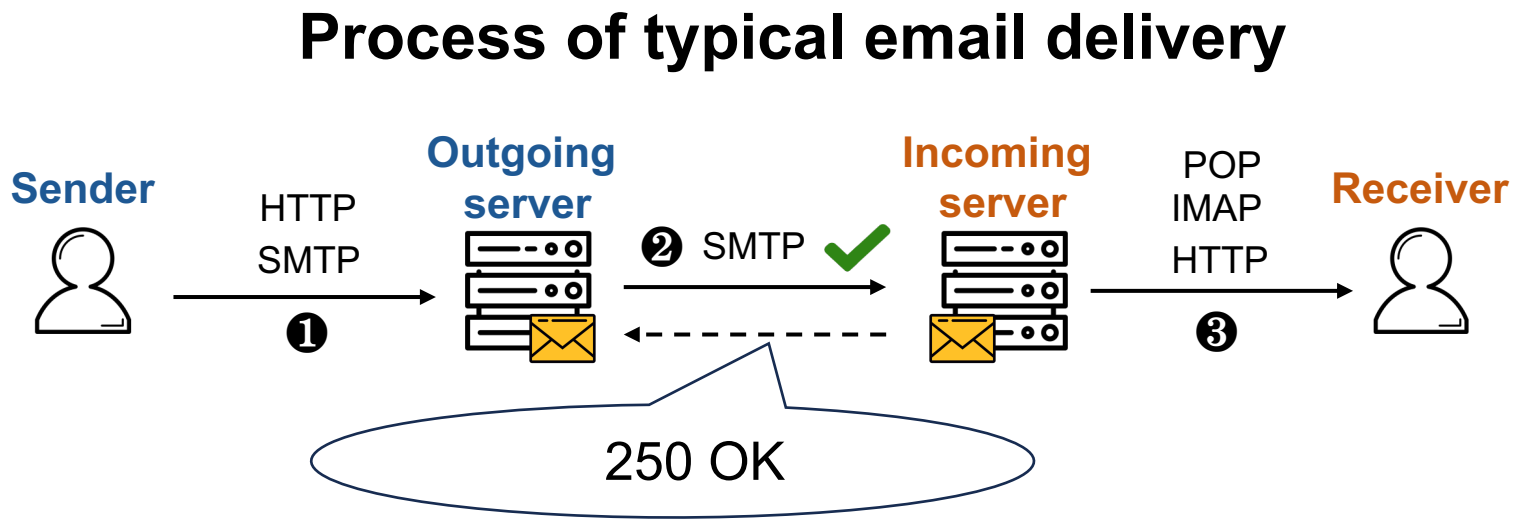
First delivery fail, but success after retries (*soft-bounced*)

Undeliverable, even after multiple attempts (*hard-bounced*)

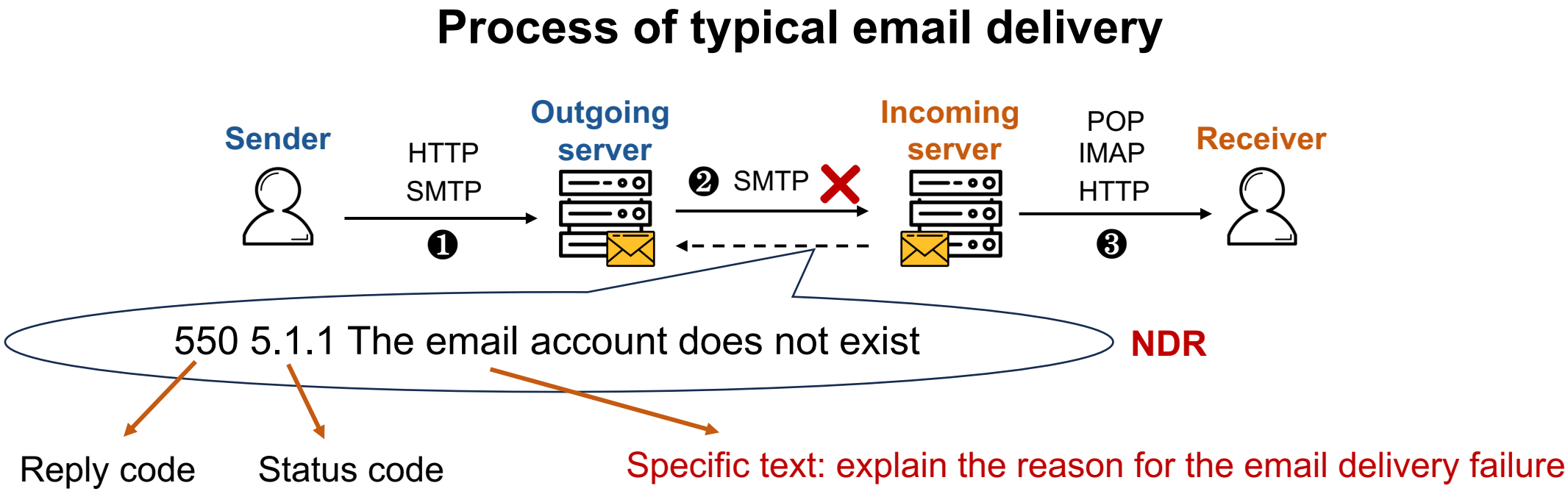
What about email delivery states in the real world?

What are the root causes for delivery failure?

Non-delivery report (NDR)



Non-delivery report (NDR)



Key idea: Understanding email deliverability and bounce reasons from NDR messages

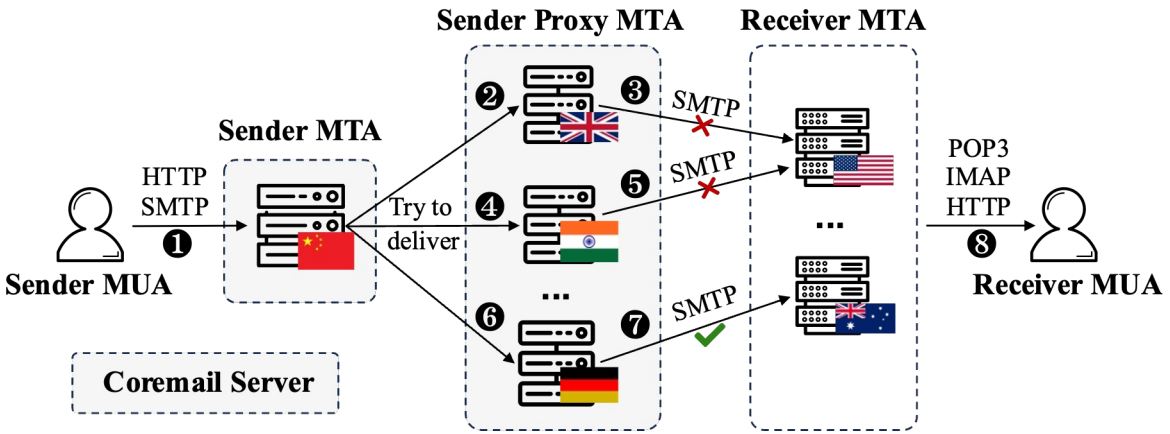
Challenges

How to obtain a representative dataset of NDR messages?

Active delivery is unrealistic: ethical risks; comprehensiveness

Global email delivery logs from a large ESP

Coremail's distributed mail proxy policy

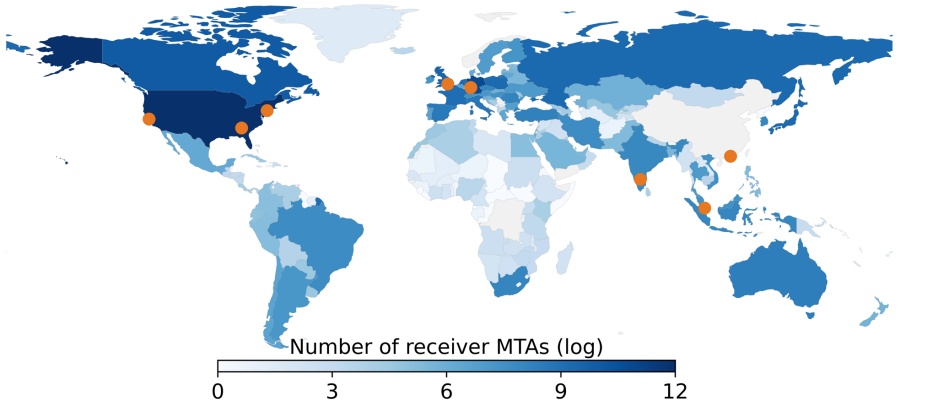


Email delivery dataset example

```
{
  "from": "alice@a.com", "to": "bob@b.com",
  "start_time": "2022-06-14 16:30:35",
  "end_time": "2022-06-14 16:45:19",
  "from_ip": ["proxy1_ip", ... , "proxy5_ip"],
  "to_ip": ["dest1_ip", ... , "dest5_ip"],
  "delivery_result": ["550 Mail rejected", ... , "250 OK"],
  "delivery_latency": [54854, ... , 28320],
  "email_flag": "Spam"
}
```

- **Time span**：15 months
- **Number of emails**：298M
- **Email recipient/sender**：3M / 68K
- **Outgoing server**：34 IP addresses, six countries
- **Incoming server**：574K IP addresses, 141 countries

Geographic distribution of outgoing/incoming servers



Challenges

How to obtain a representative dataset of NDR messages?

How to identify the types of many NDR messages?

NDR messages exhibit inconsistent formats

NDR classifier based on pre-trained language model

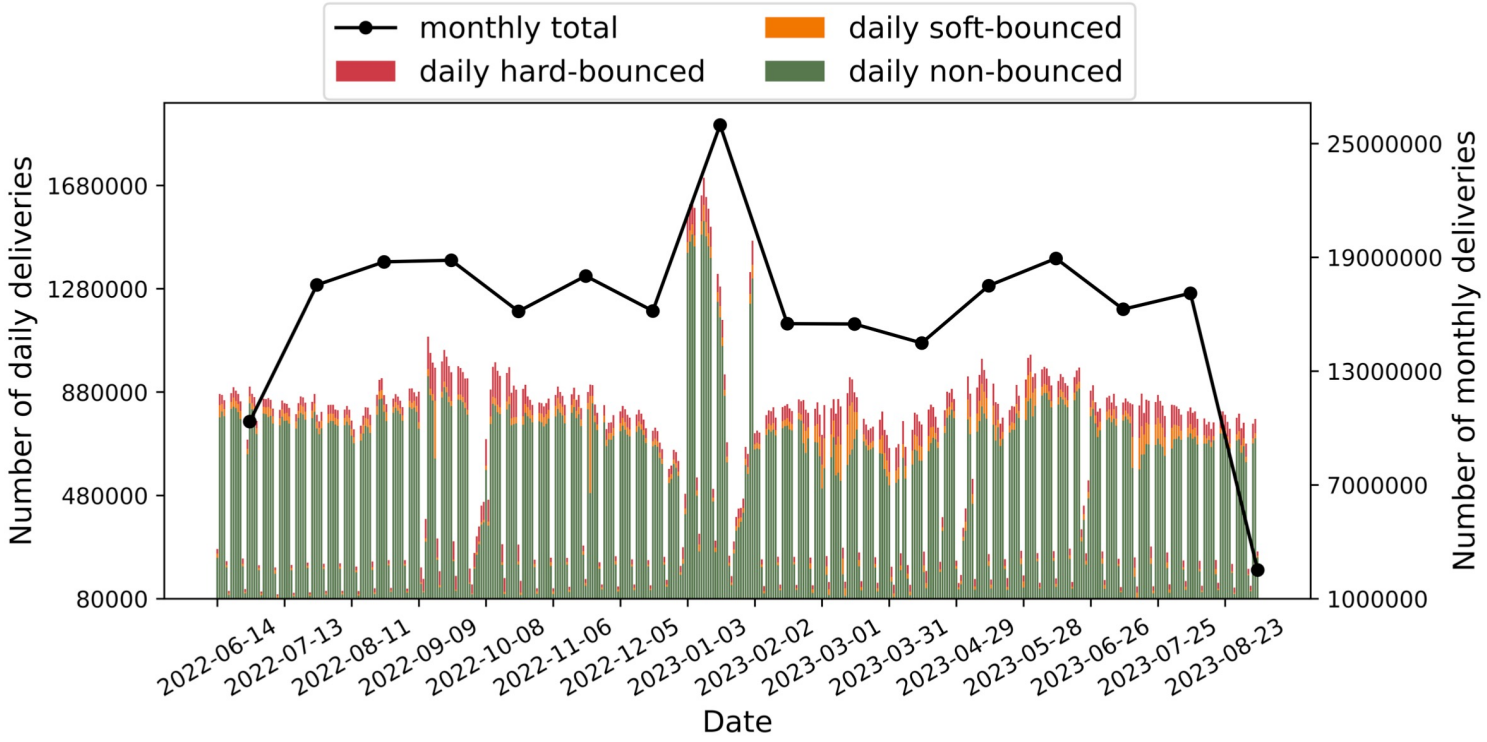
Three steps of classifier construction:

- **Cluster** 190M NDR messages into 10,089 templates using text clustering algorithm (Drain)
- **Manually analyze** the Top 200 templates to divide bounce reasons into 16 types
- **Train the classifier** using the Bert language model and predict all NDR messages (92% accuracy)

<table>
<tr><th>DNS query failure</th><th>Violate protocol standard</th><th>Restrict email source</th><th colspan="2">Refuse email reception</th><th>SMTP connection error</th></tr>
<tr><td rowspan="2">DNS record of sender domain failed to resolve</td><td rowspan="2">sender violates DKIM, SPF, or DMARC</td><td>sender MTA listed in blocklists</td><td>non-existence of the receiver email address</td><td rowspan="2">email is too large</td><td rowspan="2">SMTP session timeout</td></tr>
<tr><td rowspan="2">sender MTA blocked by greylisting</td><td rowspan="2">Receiver mailbox is full</td></tr>
<tr><td rowspan="2">DNS record of receiver domain failed to resolve</td><td rowspan="2">Sender MTA incorrectly implements STARTTLS</td><td rowspan="2">email content is considered as spam</td><td rowspan="2">SMTP session interruption</td></tr>
<tr><td>sender MTA delivering too fast</td><td>excessive recipient count of email</td></tr>
</table>

Overview of email deliverability

Among 298M emails, 24M (8%) are soft-bounced, and 14M (4%) are hard-bounced



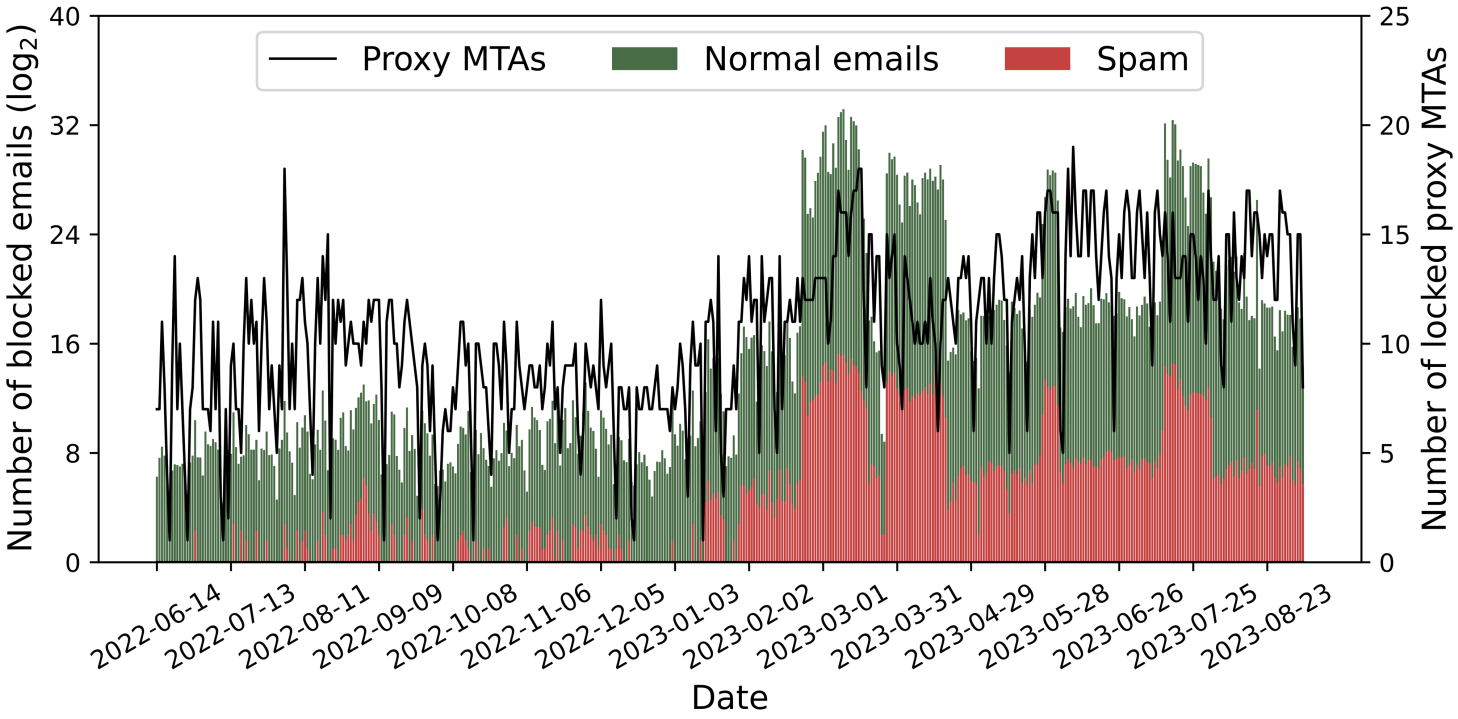
Five root causes:

- ❖ Malicious Email Behavior
- ❖ Spam Blocking Policy
- ❖ Server Manager Misconfiguration
- ❖ Improper User Operation
- ❖ Poor Email Infrastructure

Main reason for soft-bounced: hit blocklists

Because of the poor reputation of outgoing servers, 10M (31.10%) emails are soft-bounced

Example of NDR: *Service unavailable, Client host blocked using Spamhaus*



288K domains adopt Spamhaus blocklist

An average of 16K emails per day are undeliverable because the outgoing server hits the Spamhaus blocklist

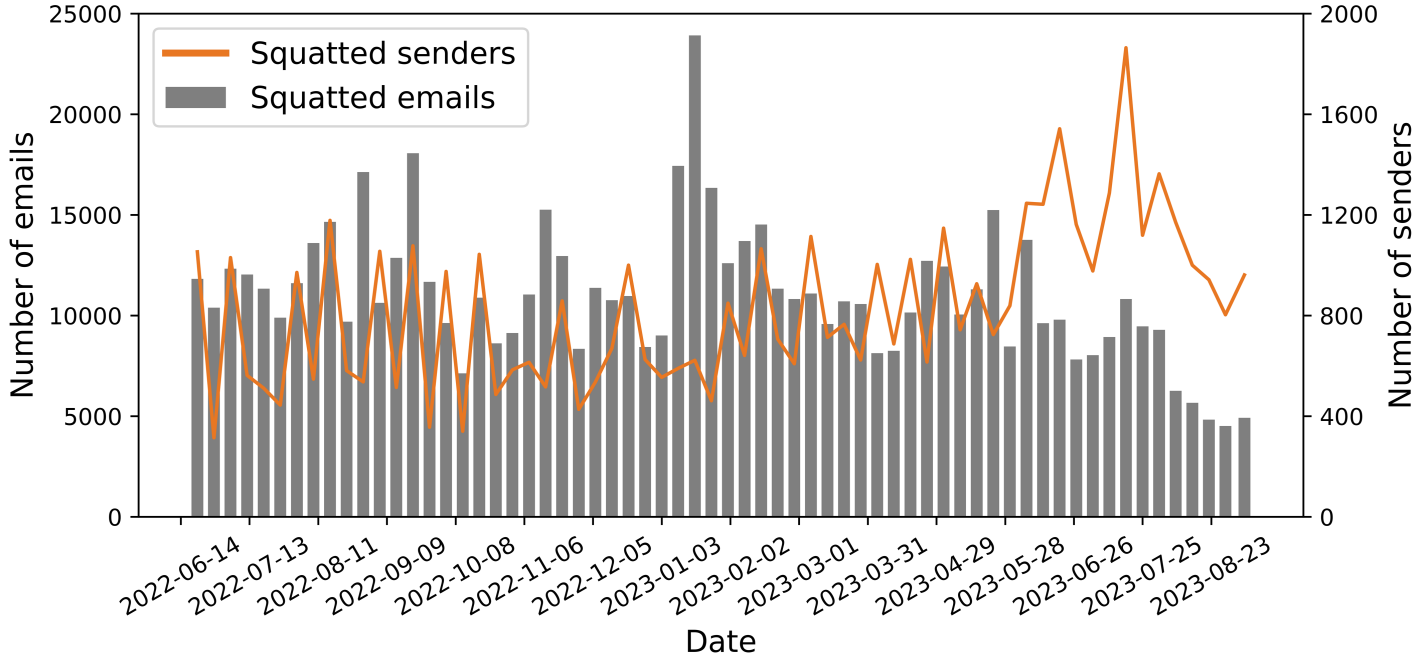
Coremail's outgoing servers and outbound emails blocked by Spamhaus blocklist

Main reason for hard-bounced: user typos

2M (6.85%) emails deliveries failed due to receiver domain and username typos

The most common error is "*omission*", e.g., "*yahoo.com.cn*" to "*yaho.com.cn*"

Many email addresses with significant residual value can be squatted



Number of senders and emails vulnerable to squatting attacks per week

3K registrable domains have received 150K emails in history

In a small-scale experiment, more than a third of the usernames could be registered

Malicious activities seen in email delivery

NDR messages associated with some senders or IPs are all "*email address does not exist*"

Mutated usernames

to****_p@***.tj	to****@***.tj
to****.p@***.tj	pr****@***.tj
p_to****@***.tj	p_to****@***.tj
p.to****@***.tj	p.to****@***.tj
pto****@***.tj	pr**-to**@***.tj
p-to****@***.tj	pr**to**@***.tj

Attackers combining social engineering guessed 39 email addresses, with a success rate of 0.91%

Protocol implementation seen in email delivery

The NDR message indicates the verify support and deployment of DKIM/SPF

701K (2.19%) emails are hard-bounced due to sender authentication failure

The average time for the sender domain to fix DKIM/SPF error is 12 days

DKIM and SPF checks failed:

421-4.7.0 This message does not pass authentication checks (SPF and DKIM both do not pass)

DKIM or SPF checks failed:

550-5.7.26 This message does not have authentication information or fails to pass authentication checks (SPF or DKIM)

DMARC check failed:

550-5.7.26 Unauthenticated email from (.) is not accepted due to domain's DMARC policy*

For other reasons ...

- Email content detected as spam (6.87%)
- User gets too much email (1.35%)
- Mailbox is full or inactive (2.06%)
- Network quality issues (10.20%)

...

View our paper for details

https://dl.acm.org/doi/10.1145/3646547.3688425

Conclusion

- We present the **landscape of global email delivery**, and evaluate the security mechanism deployment and real user behavior from the **email bounce perspective**

- We systematically **explore the root causes of email delivery failures** in the wild, providing new insights for communities to improve the email ecosystem

- We widely reveal the **squatting risks of email domain names and usernames** in the real world, and report security threats to relevant entities

Thanks for Listening!

Ruixuan Li

Email: *liruixuan@mail.tsinghua.edu.cn*

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