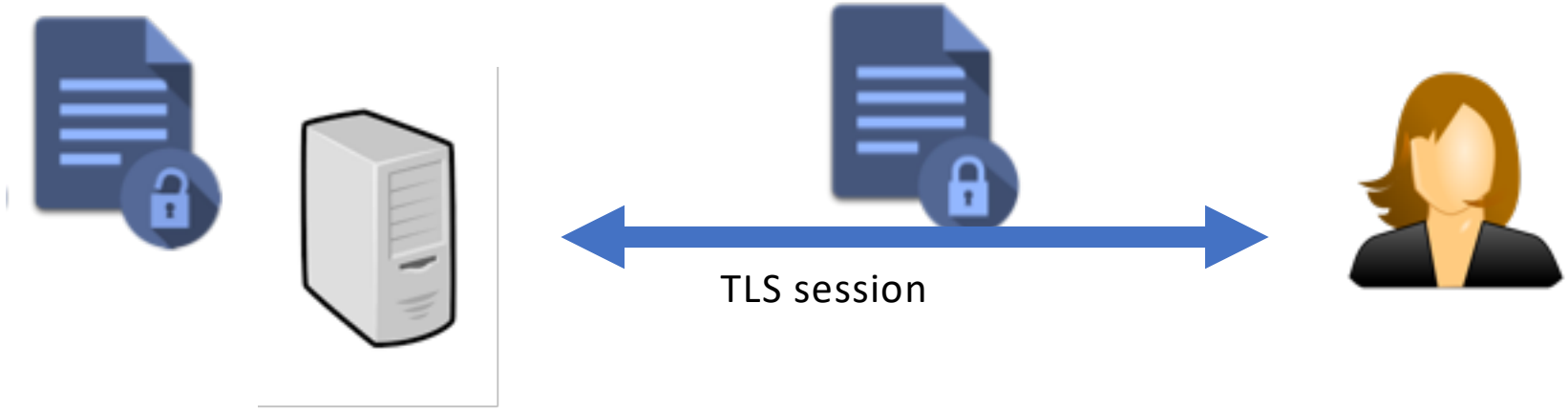
Pretzel: Email encryption and provider-supplied functions are compatible

Presenter: Jianan Lu

Email Services Today



The email server becomes a central point of attack:

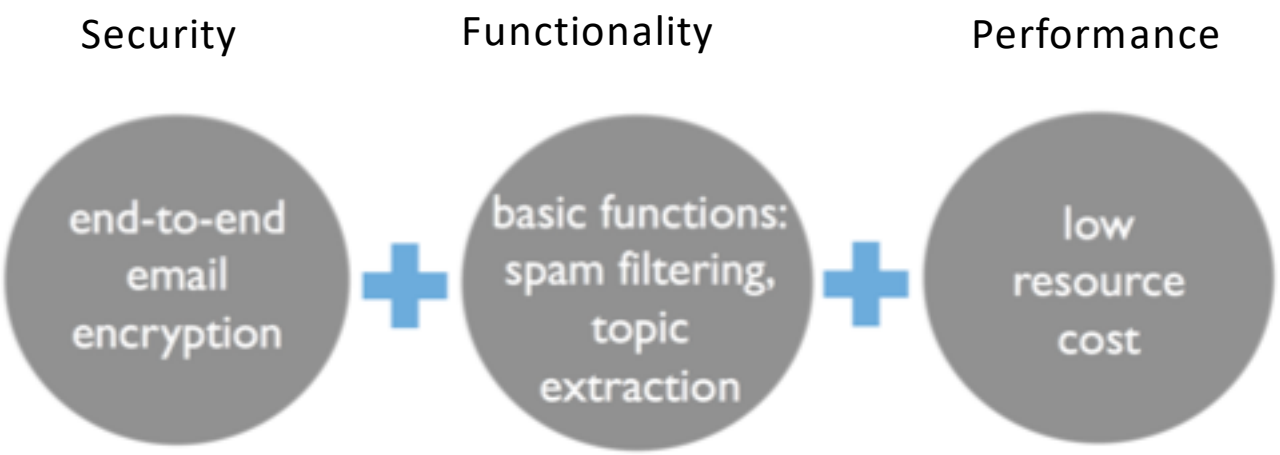
- Malicious insiders
- Hacks
- Government agencies

What About End-To-End Encryption?

- Some suitable contexts:
  - Secure Messaging: WhatsApp, Signal
  - Network communication: TLS protocol
- What encryption scheme for emails have you seen in this class?
  - PGP: trusted key server + web of trust
  - Challenge: how to perform valuable services (functions) on encrypted emails such as spam filter, search, and a lot more?

What About End-To-End Encryption?

- Possible Solutions:
  - Fully Homomorphic Encryption
  - Searchable Encryption
  - …
- Motivation: Can we do better?



[taken from Gupta et al. SIGCOMM 2017 slides]

System Overview

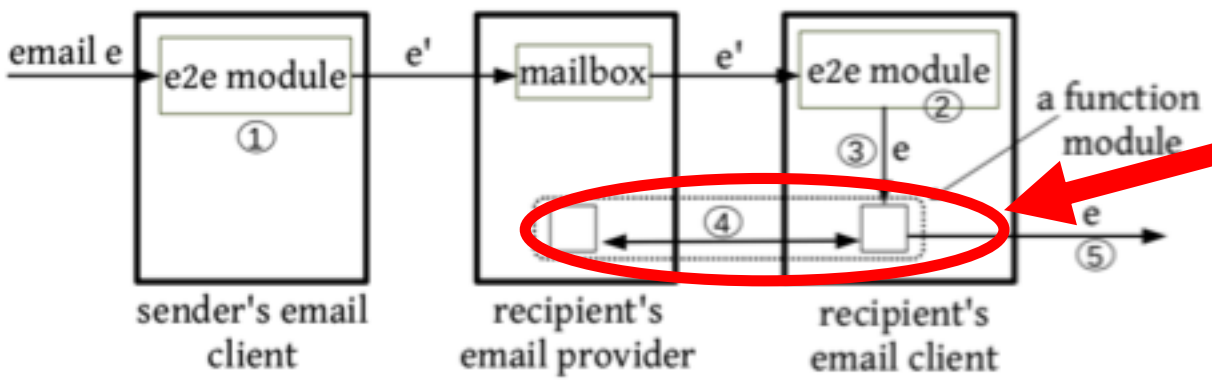
- Assumption:

key management (CONIKS etc.) and end-to-end encryption scheme (PGP etc.)

- Security Guarantee:

Both the email provider and the user cannot directly see the other's private data

- System Architecture:



How to implement secure email functions in an efficient way?

Figure 1: Pretzel's architecture. e denotes plaintext email; e' denotes encrypted email. The sender's provider is not depicted.

Building Blocks

- Secure Dot Product
  - Use GLLM in Pretzel
- Additive Homomorphic Encryption(AHE)
  - For any two messages m1, m2, Enc(m1) * Enc(m2) = Enc(m1 + m2).
- Yao’s Garbled Circuits

Multinomial Naïve Bayes Linear Classifier

- Linear classification can be used to build more feature-rich services: spam filtering, topic extraction and etc.
- Informal Definition
  - For j=1….B, denote C_j as the j-th category/label.
  - The input data x (feature vector) is N-dimensional. Denote x_i as its i-th entry and it is Boolean value (0/1) indicating the presence/absence of this i-th feature.
  - Given an input data x, output the j*-th category with maximum conditional probability. That is, $j^* = \text{argmax}_j \, p(C_j \mid \vec{x})$

Multinomial Naïve Bayes Linear Classifier

- With some derivation, it is equivalent to find the category C_{j^*} with $j^*=j$ that maximizes the equation below:
  - $p(t_i \mid C_j)$ is the probability that i-th feature is present given the data, x, belongs to category C_j.

$$\underbrace{\underbrace{\left(\sum_{i=1}^{i=N} x_i \cdot \log p(t_i \mid C_j)\right)}_{\text{Dot Product}} + 1 \cdot \log p(C_j).}_{\text{Comparison that outputs the category with highest probability}} \qquad (2)$$

Apply the building blocks

Overview of Yao+GLLM

Model Structure
Feature Extraction Methods

Model Parameters

Party X
(email provider)

Setup

Computation

Data (emails)

Party Y
(email user)

Secure Dot Product

Additive
Homomorphic
Encryption

Yao's Garbled Circuits

Construction of of Yao+GLLM

- On board
- Q: why does the user add noise?
- Revisit the security guarantee:

Both the email provider and the user cannot directly see the other’s private data

Conclusion

- Security: E2E Encryption
- Functionality and Performance: Yao + GLLM