RSAConference2022

San Francisco & Digital | June 6 – 9

TRANSFORM

SESSION ID: OST-M01

Building Safe End-to-End Encrypted Services for Business

a Google Workspace perspective

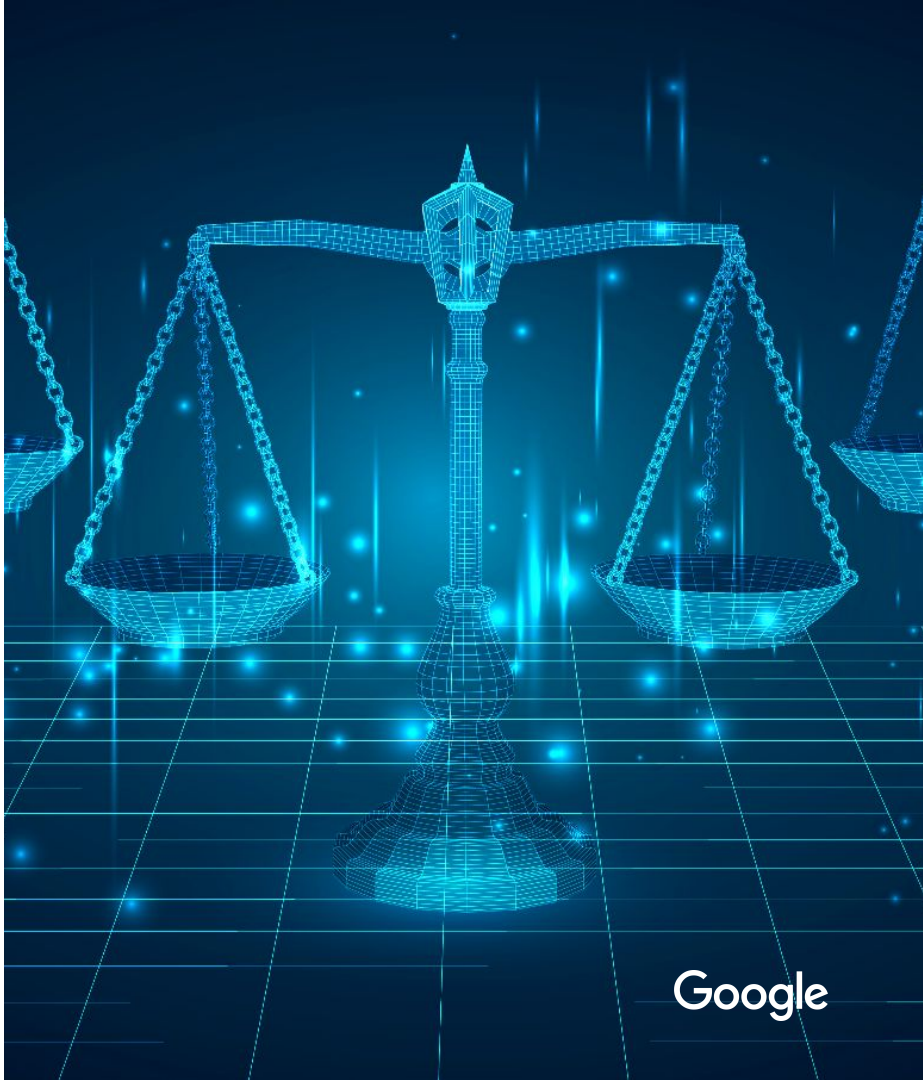
Elie Bursztein

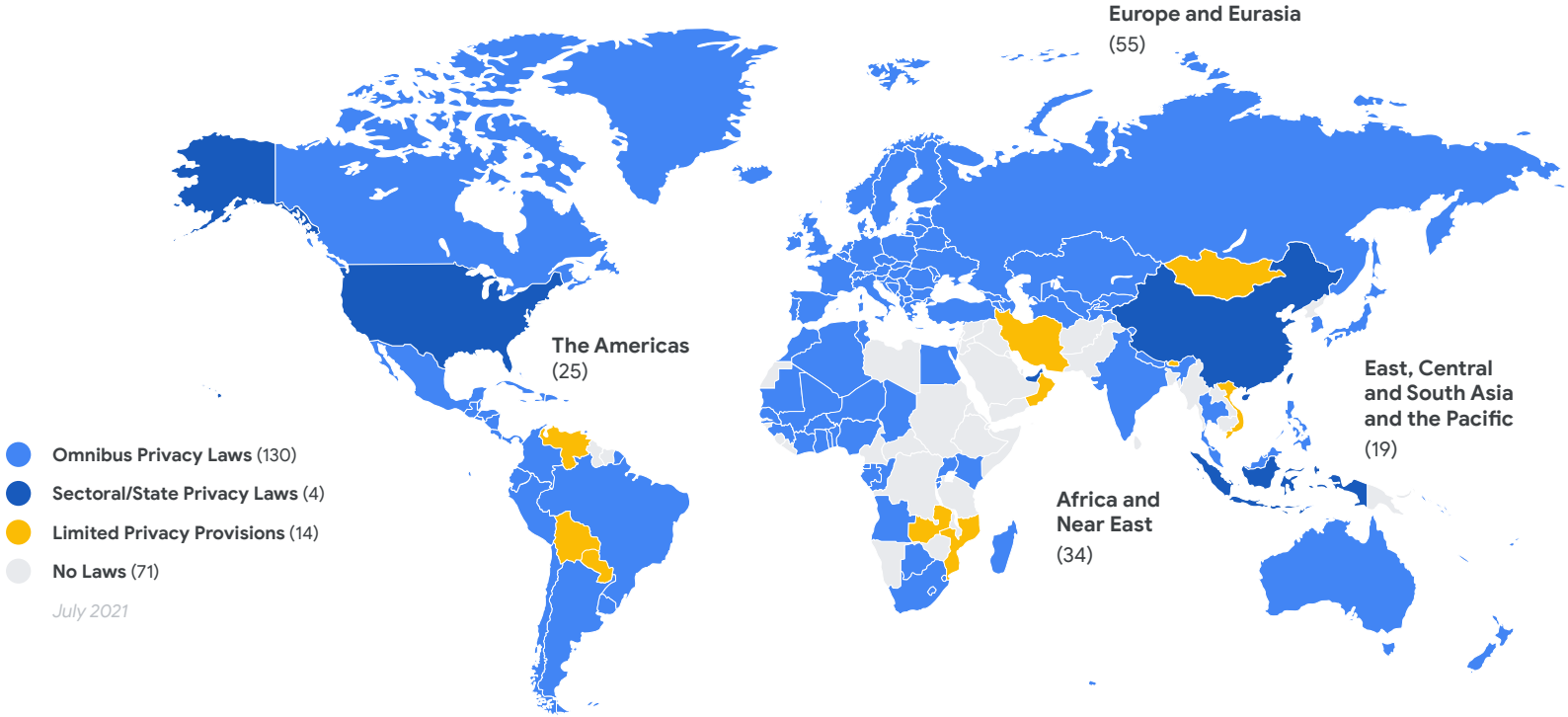
Cybersecurity Research Director
Google
@elie

Nicolas Lidzborski

Workspace CSE Engineering Lead
Google

Stronger data privacy needs and upcoming data protection regulations are reshaping the world





Data regulations and privacy needs are rising

https://www.mofo.com/resources/insights/210127-data-privacy-day.html

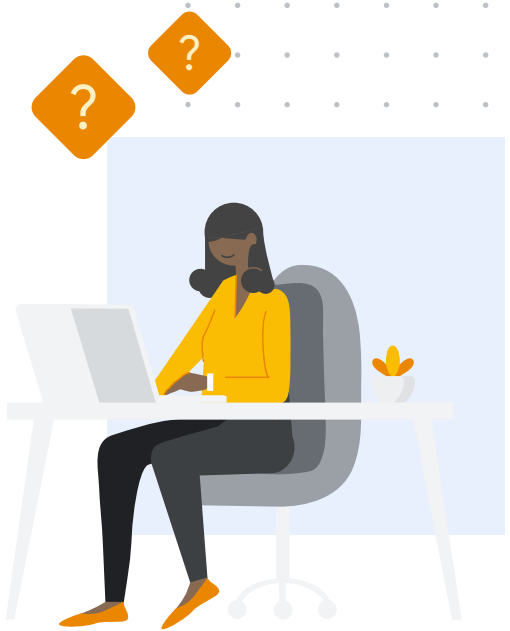
RSAConference

Today

Client-Side Encryption is one of the key technology that can help meet those new requirements.

Google

What is **Client-Side Encryption (CSE)?**



Client-side encryption (CSE) is the set of end-to-end encryption crypto-systems that enterprises can use to ensure that only authorized users can access, authenticate and decrypt specific pieces of data.

Customer feedback

> Certain email contents or recipients are **required** to be end-to-end encrypted (so **Google cannot access the data** under any circumstance). Drive is the same way - for certain information”

> ... **I just want Google to come up with a solution** so that we do not have to use any third parties ... Just one system that covers **data storage and communication security**.”

Key questions for today

How can CSE **help**?

How does CSE **work in practice**?

What are the **trade-offs**?

What are the **challenges**?

Bad news?

This presentation **does not** contain any blockchain, NFT or cryptocurrency related information.

Agenda

Client-Side Encryption for enterprise

Google Workspace CSE case-study

CSE protection challenges

Malicious URLs detection case study

Client Side Encryption for enterprise

How can CSE help protect enterprise data?

> Client-side encryption ensures that **data stored in the cloud can only be viewed by the company employees** since data is encrypted before being uploaded to cloud providers servers.

Client-side encryption use cases

Mitigate data breaches and insider risk

Separation of duty increases resilience to compromise

Offer data sovereignty control

Prevent data processing outside of a specific jurisdiction

Support regulatory compliance

E.g.: ITAR, CJIS, TISAX, IRS 1075, EAR,...

Key challenges

01.

Key deployment

Provisioning and management of keys is typically complex and requires additional software and services

02.

Interoperability

Most entreprise CSE solutions rely on proprietary infrastructure not allowing easy collaboration with others.

03.

Smart features

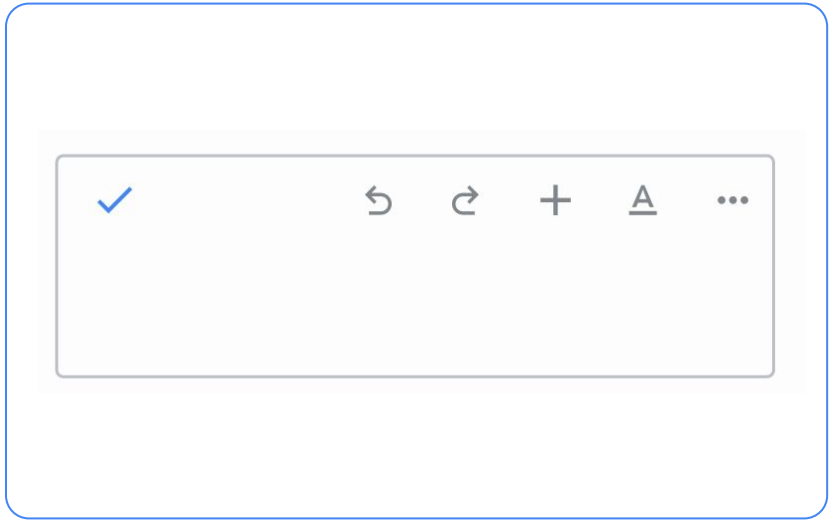
Advanced capabilities require inference with large ML models hosted server-side.

04.

Anti-abuse protection

User safety features are mostly based on complex and proprietary processing on servers.

Writing suggestions and anti-abuse are examples of server-side powered features that need to be reinvented

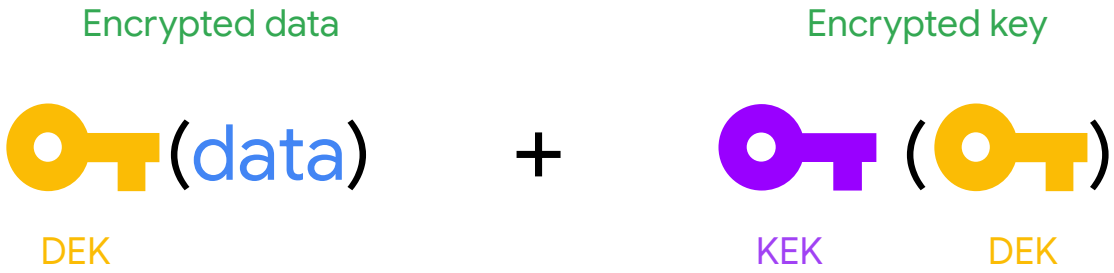


How do you encrypt the data while keeping cloud benefits?

CSE leverages envelope encryption

Data Encryption Key (DEK): key generated on end-user endpoint used to encrypt data (email, doc, file)

Key Encryption Key (KEK): common key used to encrypt and protect many DEKs



Encrypted data and DEK keys encrypted with the KEK keys can be stored safely in the cloud to enable collaboration and ensure data durability & reliability

Benefits of envelope encryption

Private and performant: data encryption is done on the endpoints

Flexible: Allows sharing with different group of users without data re-encryption

Trustworthy: Auditable, highly-protected and delegable to 3rd parties

What are the trade-offs made when implementing envelope encryption?

Access control options

Immutable access control

Or

Dynamic access control

Option 1

Data readers are set at send time

Asymmetric cryptography using recipient public keys (S/MIME, PGP, Signal,...)

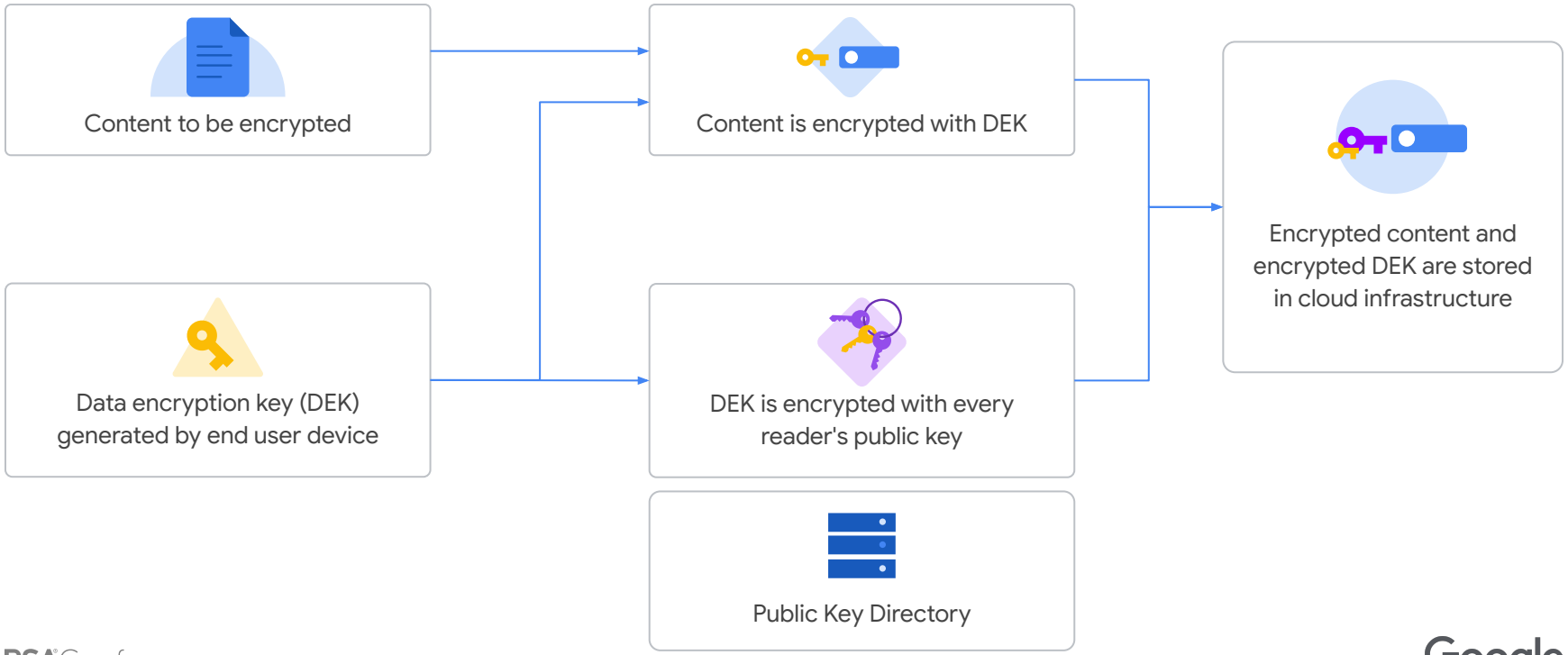
Pros:

- Portability and interop
- Sender controls readers
- Asynchronous and offline

Cons:

- Immutability
- Provisioning
- Discovery

Data readers are set at send time



Option 2

Access evaluated at decryption time

Key service solutions (Google Drive CSE, Microsoft DKE,...)

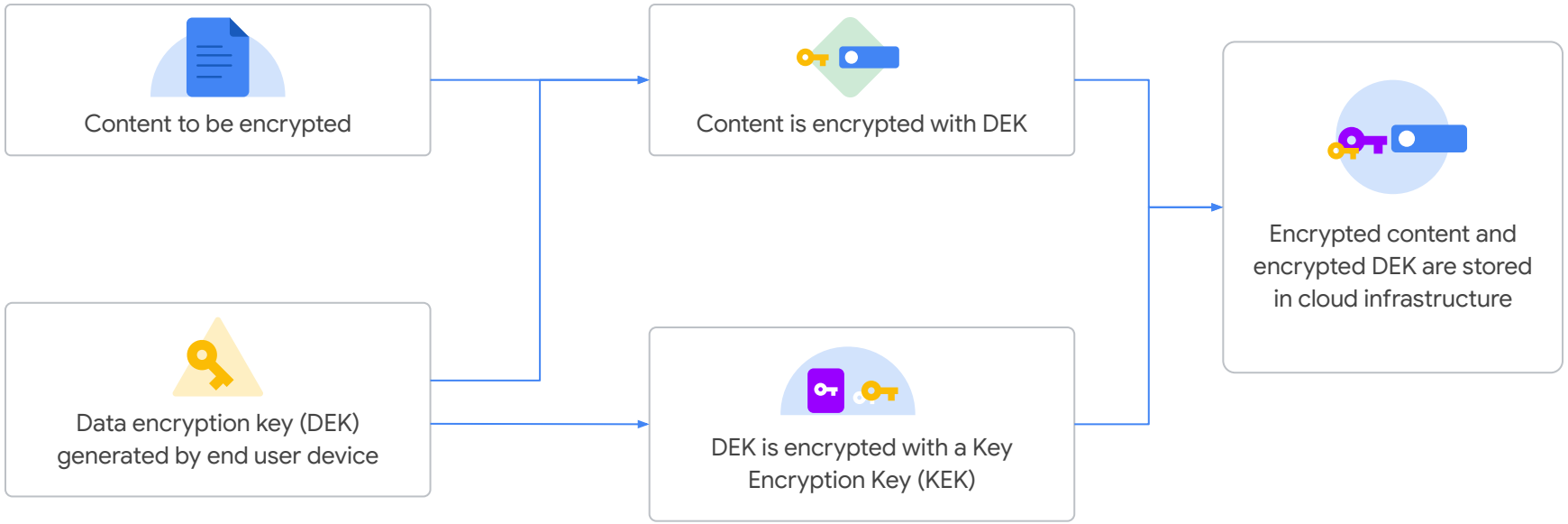
Pros:

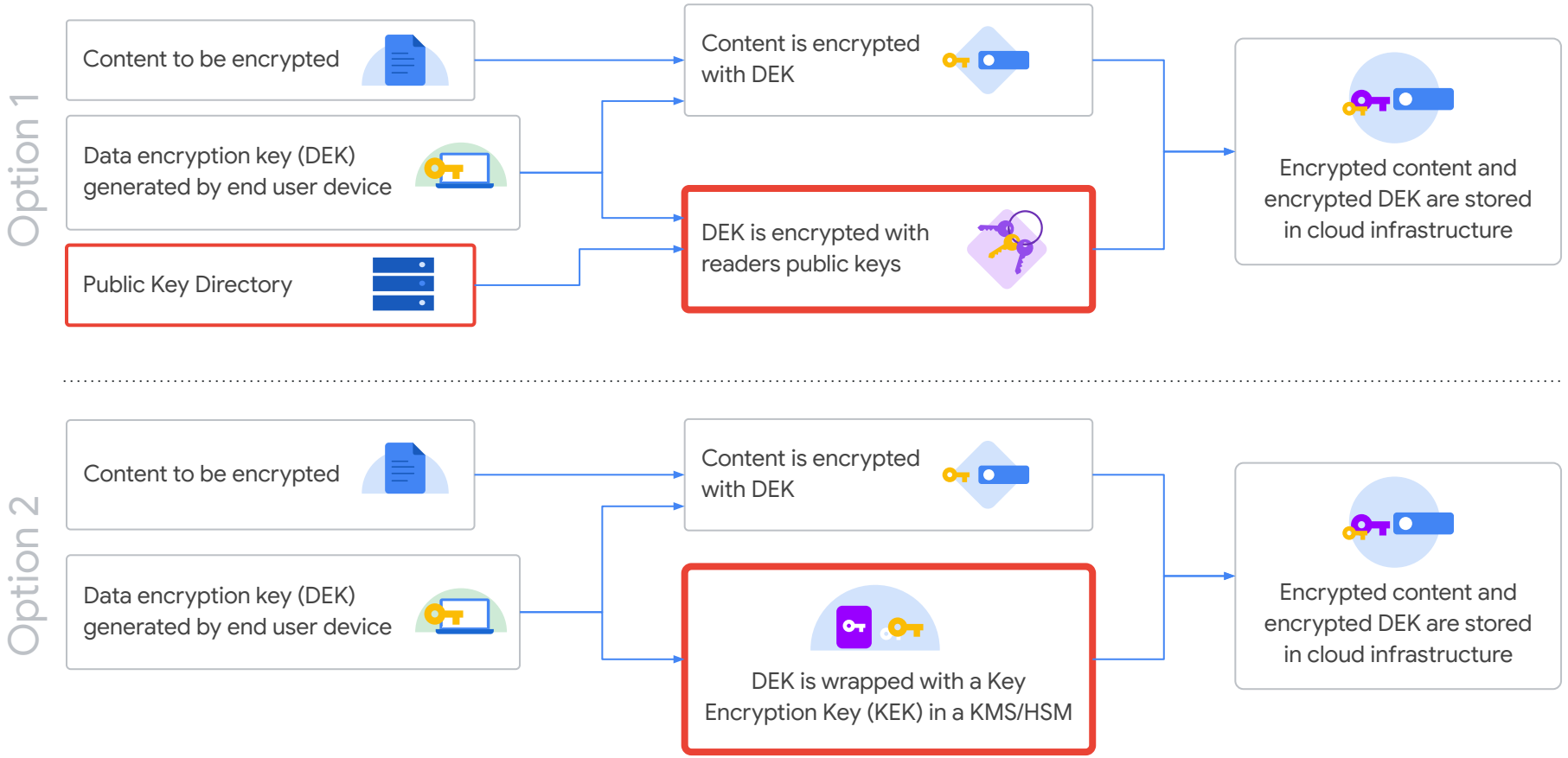
- Access with user identity
- Dynamic access control

Cons:

- Requires key service
- Harder interop/portability

Access evaluated at decryption time




Option 1
Content to be encrypted
Data encryption key (DEK) generated by end user device
Public Key Directory
Content is encrypted with DEK
DEK is encrypted with readers public keys
Encrypted content and encrypted DEK are stored in cloud infrastructure
Option 2
Content to be encrypted
Data encryption key (DEK) generated by end user device
Content is encrypted with DEK
DEK is wrapped with a Key Encryption Key (KEK) in a KMS/HSM
Encrypted content and encrypted DEK are stored in cloud infrastructure

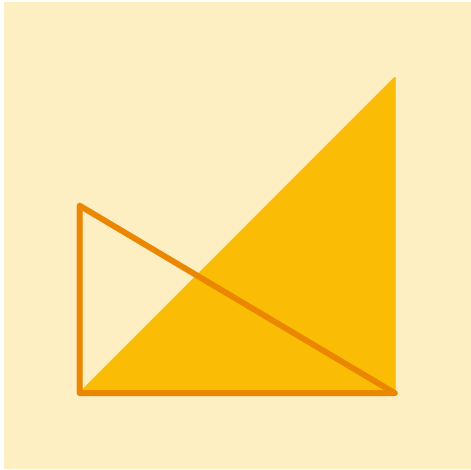
Takeaways

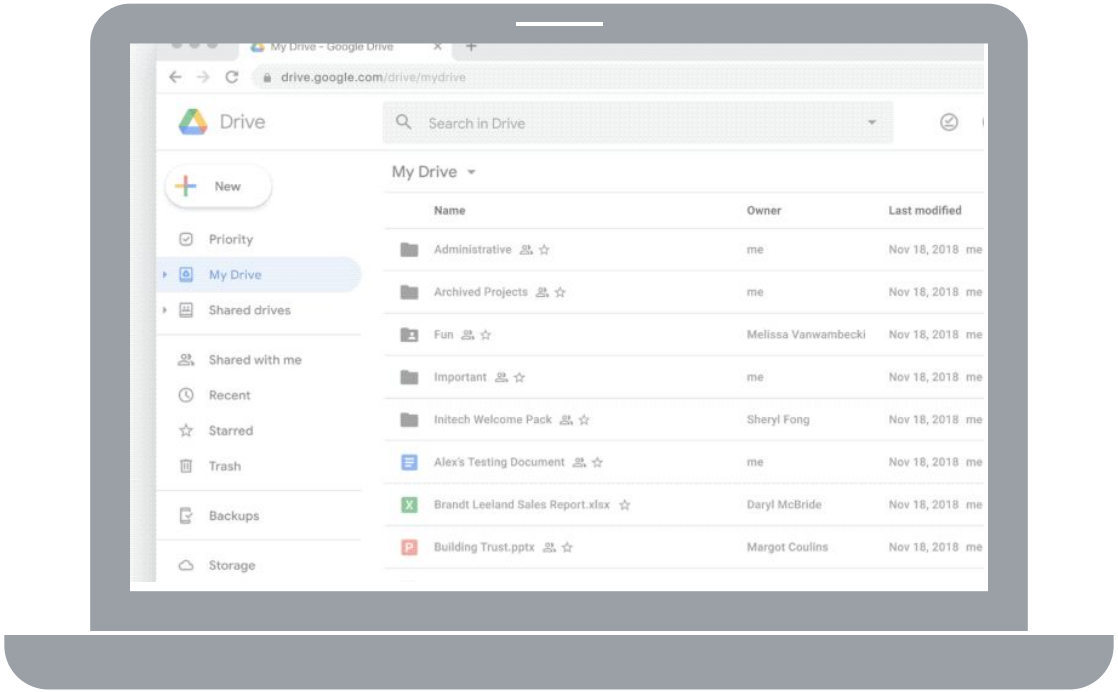
CSE solutions helps enterprises meet regulatory requirements

CSE offers technical options for data sovereignty

Building secure and smart CSE products is extremely challenging

Google Workspace CSE case study





[Google Workspace Client-side encryption launched on Drive, Docs, Sheets, and Slides in 2022]

Workspace’s approach to encryption

Whitepaper →

Sovereignty of data

Authoritative control over data through customer control of encryption keys

No server side access to content

Ensure that data is only accessible by the customer’s employees

Preserve user experience

Maintain the same high-quality experience without the need for legacy desktop clients

Workspace CSE Key ACL Service properties

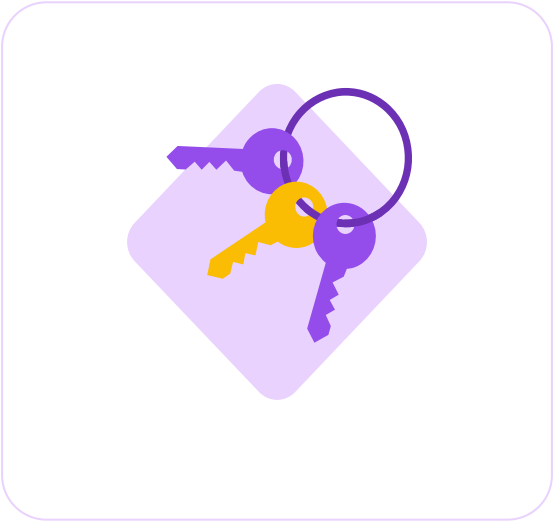
Secure service controlled by the customer

Encrypts and decrypts DEK using KEK

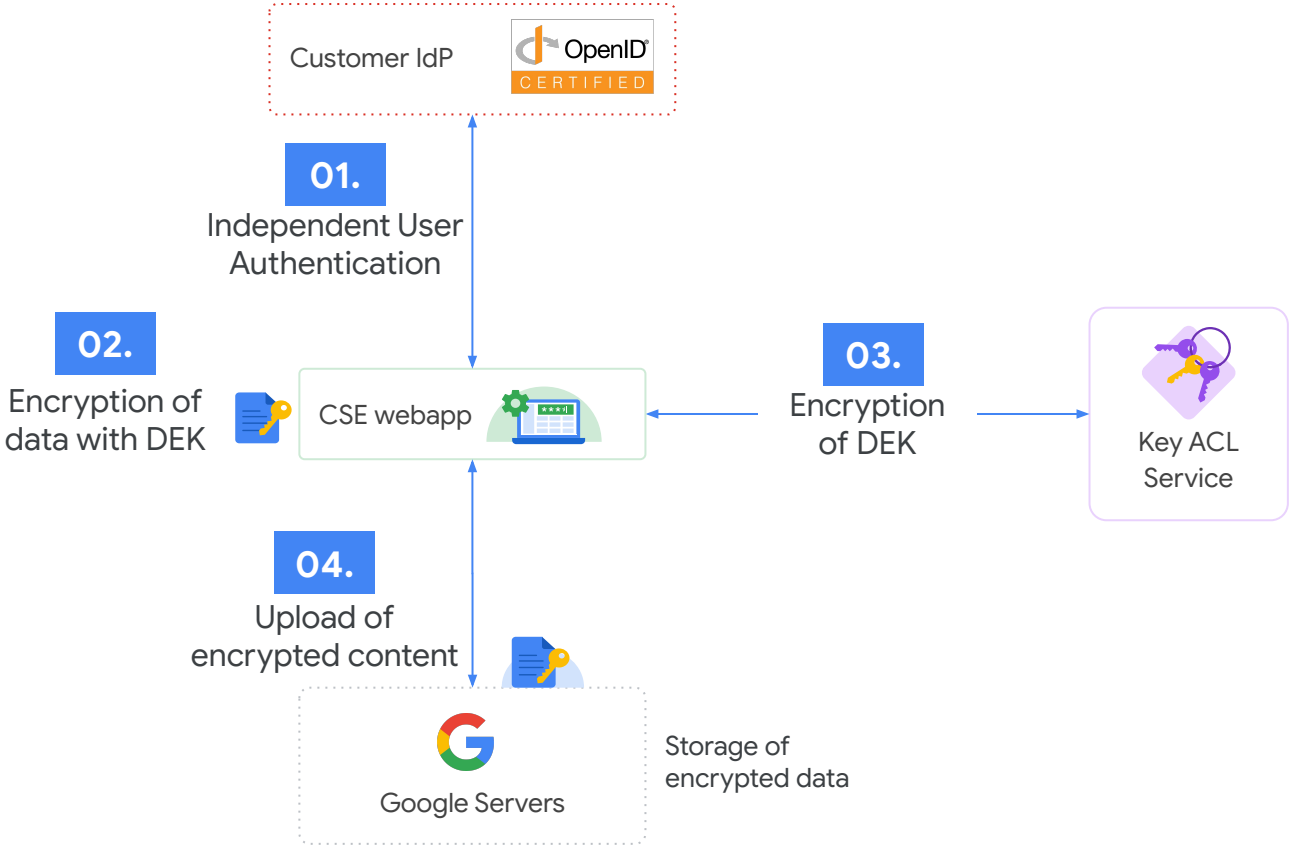
Requires strong user authentication

Provide dynamic access control

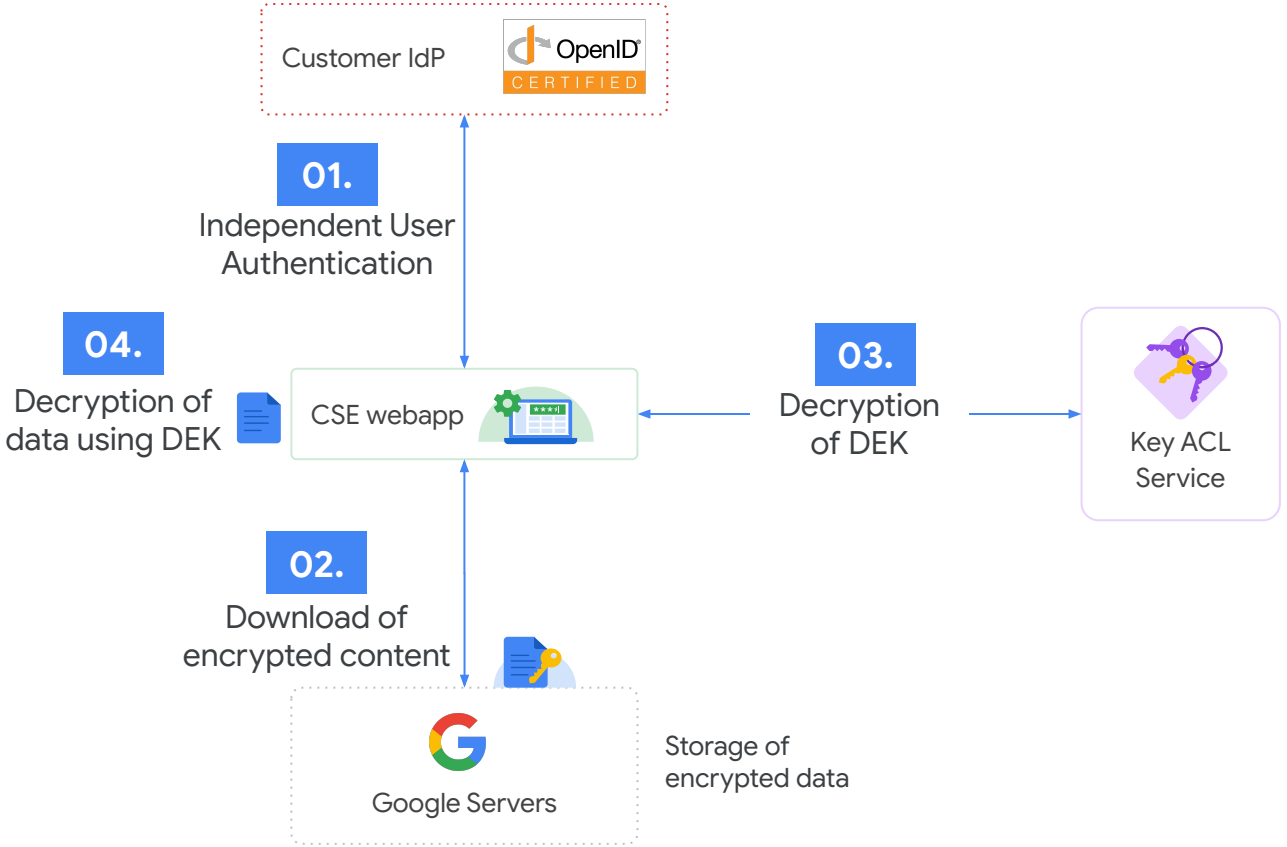
Public API allowing partners to build services



Encryption with a Key ACL Service



Decryption with a Key ACL Service



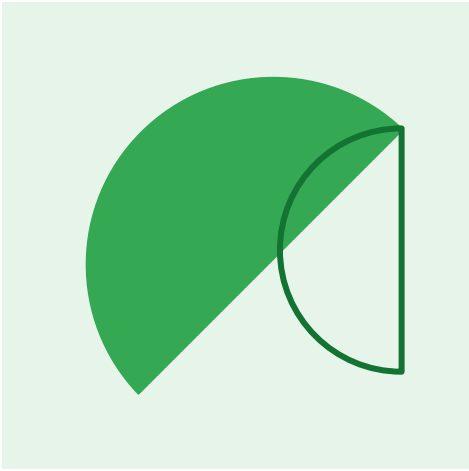
Takeaways

User experience first

Own core components to ensure reliability and security

Build on APIs and openness to maximize interoperability and transparency

CSE protection challenges



How do you protect users without server side detection?

Potential directions

01.

On-Device Business Logic

Need to rebuild product business logic to run on clients (parsing, processing and presentation layers).

02.

On-Device ML Processing

Create and deliver ML models suitable for clients (small, efficient and private).

03.

Confidential Computing

Clients can call remote services that are trusted to process content confidentially.

04.

Private Computing

Clients can use cryptographic tools (FHE, Private Intersection-Sum,..) to get remote service assistance without revealing confidential content.

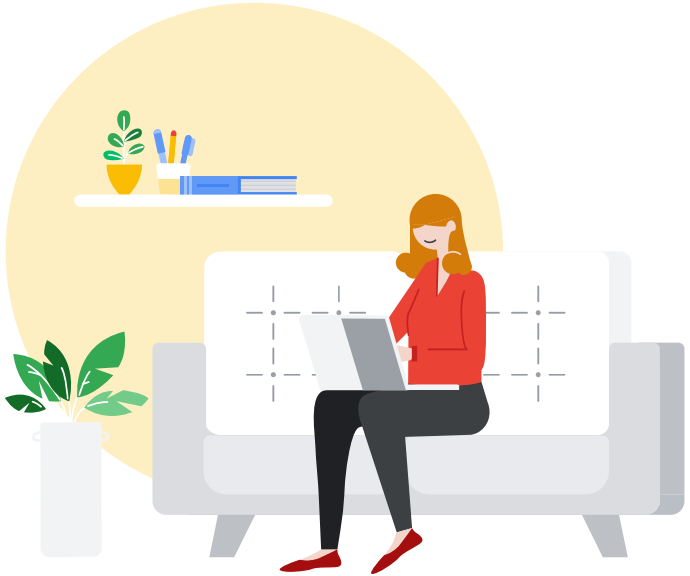
Today: two explored approaches

Privacy preserving blocklist
Combine PIR and Safe Browsing

On-device AI based detection
Model trained to recognize patterns of abuse

Enc(Query)

Enc(Resp).. Enc(shaft)



PIR tradeoffs

Pros:

Strong privacy guarantee

PIR security and privacy guarantees and limitations are well understood and researched.

Cons:

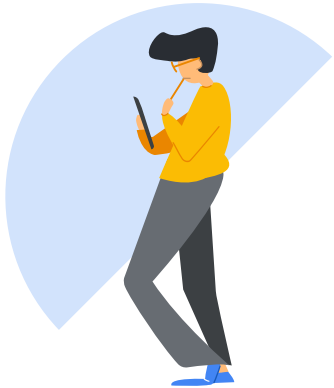
Limited operation

Can only do exact matching - No fuzzy search for example

Cons:

Scalability

Large scale databases require heavy computation and privacy trade-offs to scale

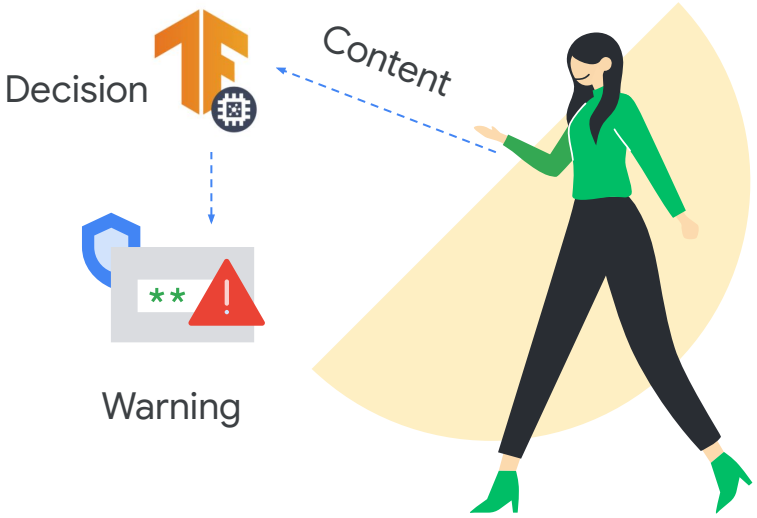


Netflix Popcorn PIR database

> Popcorn's overheads are high when compared to a non-private baseline: for each request, Popcorn **consumes 1080x more computational resources, about 14x more I/O bandwidth**, and **2x longer network transfers**.

https://www.usenix.org/system/files/conference/nsdi16/nsdi16-paper-gupta-trinabh.pdf

How an on-device model works



On-device model tradeoffs

Pros:

Strict privacy guarantees

Model operates on-device guaranteeing the strict privacy of the detection

Cons:

Resource intensive

Models require significant device compute resources and initial download, but there are techniques to help out.

Cons:

Adversarial attacks

Having an on-device model makes it easier for attackers to develop effective adversarial attacks

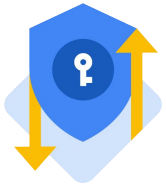
Cons:

Accuracy tradeoff

Fitting models on-device requires scaling down size, which can lead to an accuracy drop.

Takeaways

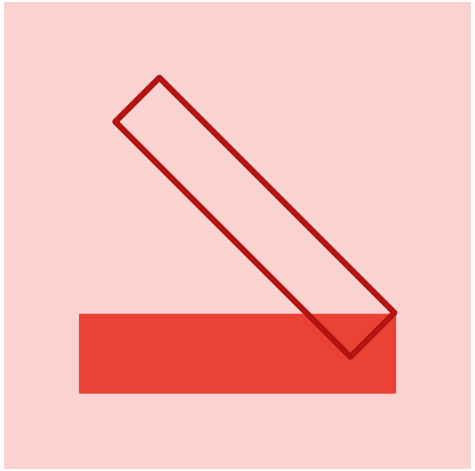
There is no silver-bullet to protect end-to-end encryption



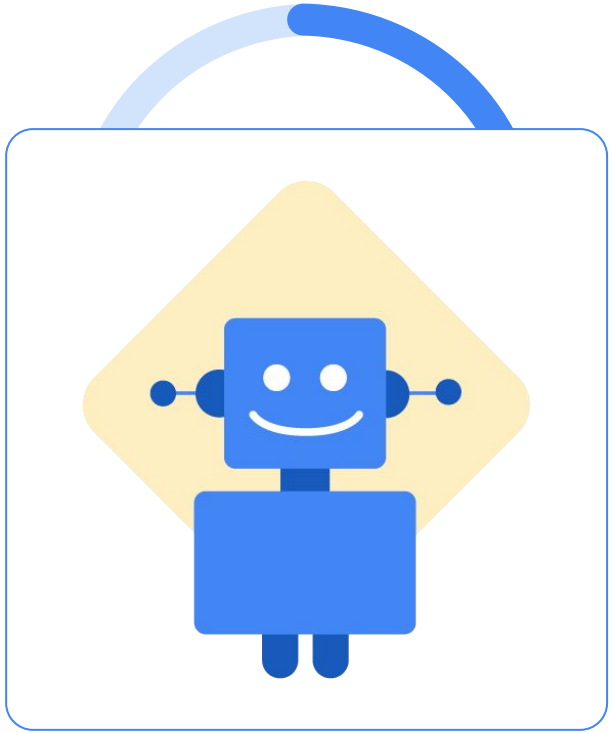
A few promising directions with strong theoretical foundations exist.

On-device protection requires significant additional research

Experimenting with malicious URLs detection case-study

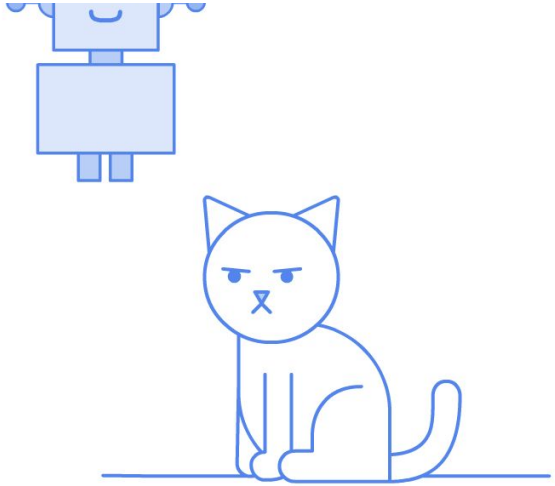


How do you protect users against malicious links without server side detection?



Current idea

Rely on an on-device model

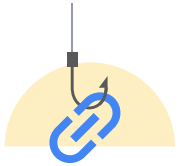


AI? Really?

Model capabilities

Malicious links detection

Predict if a url leads to a malicious website that will attack the user's machine.



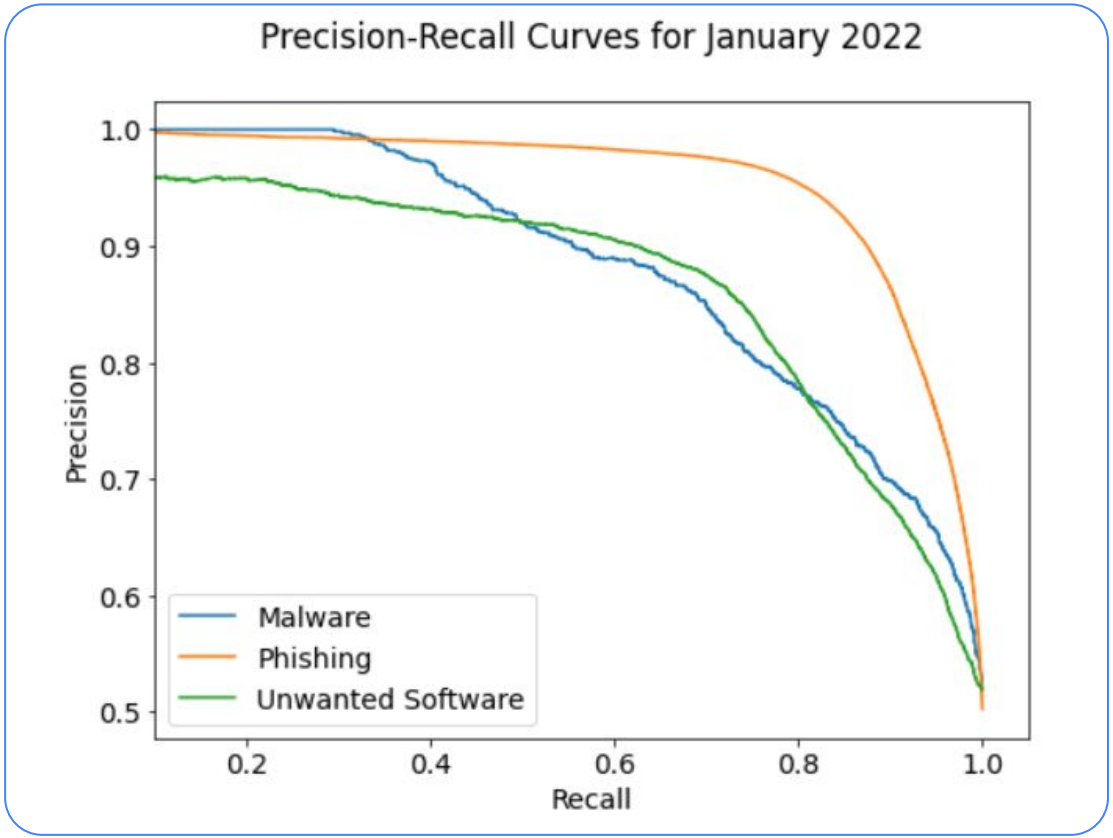
Social engineering link detection

Predict if a URL leads to a site that will phish the user.

Experimental on-device model seems accurate and fast enough to be a good base solution

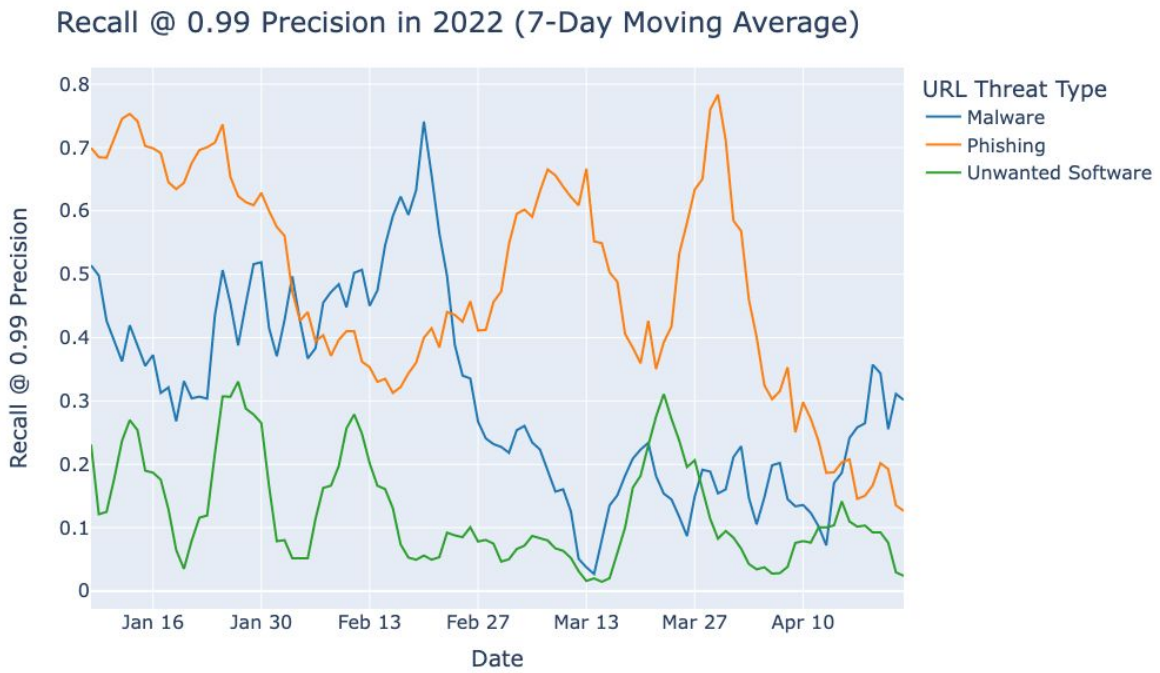
Experimental model performance

Parameters	500k
Size	2MB
Inference time	~20ms
Phishing link accuracy	90.3%
Malware link accuracy	86.45%
Unwanted software	79.41%



Model precision / recall curve

Rely on an on-device model



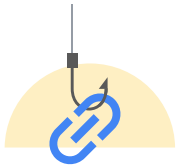
Re-imagining protections

On-device models give us the chance to reimagine Workspace protection and push the boundary of what is possible.

Model capabilities

Malicious links detection

Predict if a url lead to a malicious website that will attack the user machine.



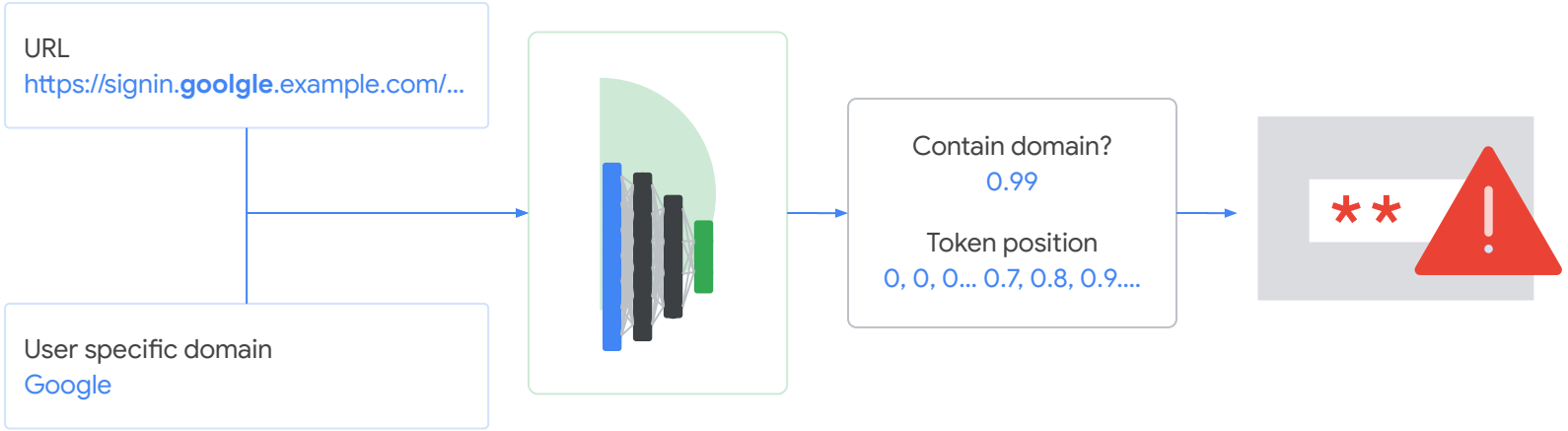
Social engineering link detection

Predict if a URL lead to a site that will phish the user.

Personalized impersonation detection

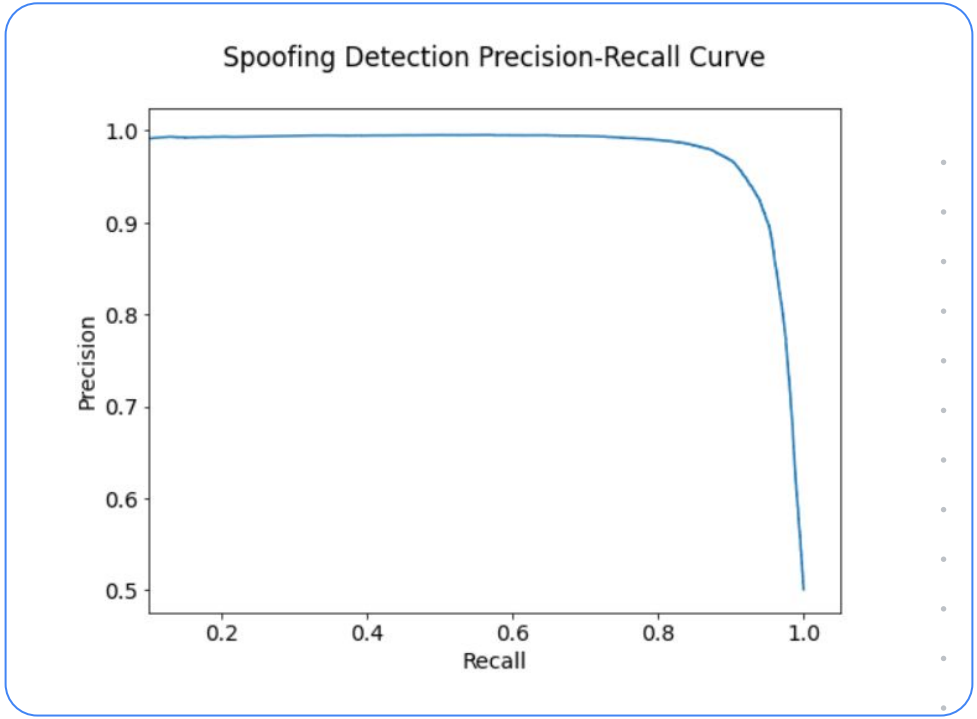
Detect if user's specific company is impersonated.

How personalized impersonation detection works

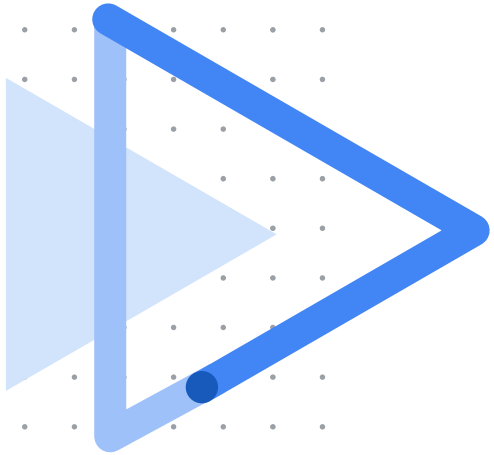


more details in our paper

Personalized impersonation precision vs recall



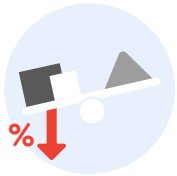
more details in our paper



Experimental model demo from this very laptop

Research paper and evaluation data to be released open-source

Takeaways



Pushing detection on device can reduce protection accuracy

On-device models offer a viable path to malicious link detection

Model generalization and resilience needs to be further researched

Takeaways

Protecting end-to-end encrypted services is challenging

There are many promising directions but no silver-bullet

Building advanced CSE protections is a very active research area

CSE services are becoming a critical part of business data protection strategy. They introduce new unique operational challenges that require innovative solutions to offer strong usability, safety, reliability and functionality.

Thank you