財團法人電信技術中心出國報告

出席瑞士
The 23th Privacy Enhancing Technologies Symposium
出國報告書

單位名稱：財團法人電信技術中心

姓名職稱：鄧宇翔 工程師

梁博硯 助理工程師

派赴國家：瑞士洛桑

出國期間：**2023** 年 **7** 月 **8** 日至 **2023** 年 **7** 月 **18** 日止

報告日期：**2023** 年 **8** 月 **17** 日

摘要

2023 年隱私強化技術期刊（Proceedings on Privacy Enhancing Technologies，PoPETs）舉辦第 23 屆隱私強化技術研討會，會議時間自 2023 年 7 月 10 號至 7 月 15 號，為期 6 天，地點位於瑞士洛桑大學。該研討會聚集如 Google、Microsoft 研究院、IBM 研究院、麻省理工學院、洛桑理工學院等知名產學界專家共同參與，並發表 124 篇與隱私強化技術相關之學術研究論文。本次會議分為 17 個不同主題，包括資料隱私、網頁隱私、匿名化，法律與政策等面向，透過參與本次研討會，本團隊將鎖定三種面向之議題，分別為：隱私強化技術之評估、落地與驗測，藉此從中學習國際前沿知識與經驗，以助本計畫執行隱私強化技術導入，建立國際專家之鏈結，並提升我國隱私強化技術之發展能量。

目錄

表目錄

圖目錄

壹、 目的

本中心為執行數位發展部多元創新司「建構隱私強化技術與數據公益合規機制計畫」補助計劃案，了解國際在隱私強化技術之發展趨勢、現行可用技術，以及欲導入隱私強化技術所需之知識與經驗。本次由鄧宇翔工程師與梁博硯助理工程師出席於瑞士舉辦，第 23 屆隱私強化技術研討會 (The 23th Privacy Enhancing Technologies Symposium)。本團隊於參與研討會前，先行盤點出該研討會所發表之主題，並整理歸納出「評估」、「落地」與「驗測」三個主題。

在評估方面，著重於導入隱私強化技術前之情境評估，並從研討會所發表之主題中獲取於評估階段所需之知識，預期會從中學到包含：評估前之需求訪談重點、適用技術評估應考慮面向，以及評估流程與規劃等，有望可從相關研究或案例中獲取經驗，以優化我方隱私強化技術導入之評估工作。

在落地方面，著重於各隱私強化技術之情境導入實作與案例研究，本團隊預期從相關研究中，獲取不同隱私強化技術之實作細節，例如：使用之開源工具、適用之技術變種與技術實作方法。對於落地案例之研究，本團隊預期從各案例導入之技術中，獲知不同情境中技術導入之效益、不同情境與資料樣態之處理方式，以及各技術落地之限制與挑戰。上述知識可作為本計畫隱私強化技術導入合作之參考依據。

在驗測方面，由於隱私強化技術仍屬前沿領域，各技術之隱私保護力與資料可用性之驗證，較少有相關文獻可供參考，故本團隊預期從該研討會中，獲取各隱私強化技術驗測相關知識，例如：各技術驗測之評估指標、相關開源工具、驗測方法論等，藉此提升本計畫隱私強化技術驗測之發展能量。

貳、 行程

表格 1：出國行程表

日期	行程
2023 年 7 月 8 日 23:35 至 2023 年 7 月 9 日 13:20	差旅時間：臺灣桃園國際機場至瑞士蘇黎世機場 (搭乘阿聯酋航空 EK0367，經杜拜轉機)
2023 年 7 月 10 至 14 日 9:00-18:00	參加 The 23th Privacy Enhancing Technologies Symposium 五天議程 主要參與議題： • How Much Privacy Does Federated Learning with Secure Aggregation Guarantee? • Lessons Learned: Surveying the Practicality of Differential Privacy in the Industry • Differentially Private Publication of Trajectory Data • Differentially Private Speaker Anonymization • A Unified Framework for Quantifying Privacy Risk in Synthetic Data • SoK: New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks
2023 年 7 月 13 日 18:00-21:00	參與研討會 Gala Dinner 晚宴
2023 年 7 月 16 日 15:25 2023 年 7 月 17 日 16:15	差旅時間：瑞士蘇黎世機場至臺灣桃園國際機場 (搭乘阿聯酋航空 EK0088，經杜拜轉機)

表格 2：7/10 會議參與進程

2023 年 7 月 10 日	
時間	**主題**
09:00 –09:15	FOCI Welcome 開幕演說
09:15 –10:15	FOCI paper session: Censorship Analysis ● A Worldwide View of Nation-state Internet Censorship ● Detecting Network Interference Without Endpoint Participation ● Towards a Comprehensive Understanding of Russian Transit Censorship
10:15 – 10:45	Break with snacks
10:45 –11:45	FOCI Keynote： ● Motivation and Challenges for Censorship Research (the Indian perspective)
11:45 –13:30	Lunch
13:30 - 14:30	FOCI paper session: New strategies for evasion： ● The Use of Push Notification in Censorship Circumvention ● Proteus: Programmable Protocols for Censorship Circumvention ● Voiceover: Censorship-Circumventing Protocol Tunnels with Generative Modeling
14:30 - 15:00	Break
15:00 – 16:00	FOCI Keynote
16:00 –16:15	Break with snacks
16:15 –17:15	FOCI paper session: Real-world considerations： ● Rethinking Realistic Adversaries for Anonymous Communication Systems ● Running a high-performance pluggable transports Tor bridge

	● Crowdsourcing the Discovery of Server-side Censorship Evasion Strategies
17:15 – 17:30	Closing

表格 3 : 7/11 會議參與進程

2023 年 7 月 11 日	
時間	**主題**
09:00 – 10:30	Session 1B: Privacy and AI/ML I ● Efficient decision tree training with new data structure for secure multi-party computation ● On the Privacy Risks of Deploying Recurrent Neural Networks in Machine Learning Models artifact ● How Much Privacy Does Federated Learning with Secure Aggregation Guarantee? ● ezDPS: An Efficient and Zero-Knowledge Machine Learning Inference Pipeline ● Differentially Private Simple Genetic Algorithms artifact ● Towards Sentence Level Inference Attack Against Pre-trained Language Models
10:30 – 11:00	Break
11:00 – 12:00	AMA with Serge Egelman, Susan Landau, and Carmela Troncoso
12:00 – 13:30	Lunch
13:30 – 15:00	Session 2B: Data Privacy ● Distributed GAN-Based Privacy-Preserving Publication of Vertically-Partitioned Data ● Lessons Learned: Surveying the Practicality of Differential Privacy in the Industry ● SoK: Managing risks of linkage attacks on data privacy ● Strengthening Privacy-Preserving Record Linkage using Diffusion artifact ● SoK: Membership Inference is Harder Than we Previously Thought artifact ● DP-SIPS: A simpler, more scalable mechanism for differentially private partition selection
15:00 – 16:00	Town Hall

16:00 – 17:00	Birds of a Feather sessions： ● Welcome to PETS (ANT-1031), contact: Susan McGregor ● New directions in censorship research (ANT-1129), contacts: Ram Sundara Raman and David Fifield ● LGBTQ+ in PETS (ANT-2024), contact: Carmela Troncoso ● Hallway track with refreshments
17:00 – 18:00	Birds of a Feather sessions： ● Ethics concerns in S&P research (ANT-1031), contact: Lujo Bauer ● Cryptographic Tools for Privacy (ANT-1129), contact: Nadim Kobeissi ● Women in PETS (ANT-2024), contacts: Anna Lorimer and Bailey Kacsmer ● Board Games (ANT-3021), contact: Rebekah Overdorf ● Hallway track with refreshments

表格 4 : 7/12 會議參與進程

2023 年 7 月 12 日	
時間	**主題**
09:00 – 10:30	Session 3B: Privacy and AI/ML II ● Exploring Model Inversion Attacks in the Black-box Setting ● Individualized PATE: Differentially Private Machine Learning with Individual Privacy Guarantees artifact ● Private Multi-Winner Voting for Machine Learning ● Find Thy Neighbourhood: Privacy-Preserving Local Clustering ● How to Combine Membership-Inference Attacks on Multiple Updated Machine Learning Models ● PubSub-ML: A Model Streaming Alternative to Federated Learning
10:30 – 11:00	Break
11:00 – 12:00	Speed Mentoring
12:00 – 13:30	Lunch
13:30 – 14:45	Session 4C: Personal Safety ● Blind My - Robust Stalking Prevention in Apple's Find My Network ● Creative beyond TikToks: Investigating Adolescents' Social Privacy Management on TikTok artifact ● SoK: Differentially Private Publication of Trajectory Data ● On the Role and Form of Personal Information Disclosure in Cyberbullying Incidents ● Track You: A Deep Dive into Safety Alerts for Apple AirTags
14:45 – 15:00	Break
15:00 – 16:00	Poster Session
16:00 – 16:30	Break
16:30 – 17:30	Rump Session

17:30 – 18:00	Break
18:00 -	Gala Dinner

表格 5 : 7/13 會議參與進程

2023 年 7 月 13 日	
時間	**主題**
09:00 – 10:30	Session 5C: Genomes and Biometrics ● Differentially Private Speaker Anonymization ● I-GWAS: Privacy-Preserving Interdependent Genome-Wide Association Studies ● StyleID: Identity Disentanglement for Anonymizing Faces artifact ● "My face, my rules": Enabling Personalized Protection against Unacceptable Face Editing ● Two-Cloud Private Read Alignment to a Public Reference Genome artifact ● Examining StyleGAN as a Utility-Preserving Face De-identification Method
10:30 – 11:00	Break
11:00 – 12:00	Birds of a Feather sessions： ● Innovation strategy and building PETs in industry ● Parents in PETS ● NLP for Privacy and Privacy for NLP ● Privacy Preference Signals ● (Un)solvable privacy gaps and challenges
12:00 – 13:30	Lunch
13:30 – 15:00	Session 6B: Privacy and AI/ML III ● HeLayers: A Tile Tensors Framework for Large Neural Networks on Encrypted Data artifact ● MIAShield: Defending Membership Inference Attacks via Preemptive Exclusion of Members ● SoK: Secure Aggregation based on cryptographic schemes for Federated Learning ● Private Graph Extraction via Feature Explanations artifact ● Losing Less: A Loss for Differentially Private Deep Learning ● Privacy-Preserving Federated Recurrent Neural Networks
15:00 – 16:00	Break
16:00 – 17:00	

	Session 7B: Measurement and Data Privacy ● Differential Privacy for Black-Box Statistical Analyses ● DPrio: Efficient Differential Privacy with High Utility for Prio artifact ● A Unified Framework for Quantifying Privacy Risk in Synthetic Data artifact ● SenRev: Measurement of Personal Information Disclosure in Online Health Communities ● SoK: New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks artifact ● CERTainty: Detecting DNS Manipulation at Scale using TLS Certificates
16:00 – 17:00	Closing Remarks

表格 6 : 7/14 會議參與進程

2023 年 7 月 14 日	
時間	**主題**
09:00 – 09:15	HotPETs Opening Remarks
09:15 – 10:35	● Data for good: US data breach visualization and classification ● Female mHealth Apps: Opportunities and Challenges ● Dataveillance: An Overlooked Side of Dark Patterns
10:35 – 11:00	Coffee Break
11:00 – 12:00	Keynote: The Power of Community Participation in Shaping Digital Rights Tech
12:00 – 13:30	Lunch
13:30 – 15:00	● Impacts of Growing Tor – A Neglected Aspect of Tor Evolution? ● Are PETS and algorithmic accountability at loggerheads? ● Who are users in privacy discussions?
15:00 – 16:00	Ice Cream!
16:00 – 17:00	● Where Do Threat Models Come From? Challenging Implicit Assumptions ● Bridging the Gap between Privacy Incidents and PETs
17:00	HotPETs Closing Remarks and Award

參、 驗測相關研究議題

1. A Unified Framework for Quantifying Privacy Risk in Synthetic Data

本研究介紹一個名為 Anonymeter 的統計框架，其用於共同量化合成表格資料集中的不同類型的隱私風險。該框架基於歐洲通用數據保護法規 (GDPR) 所提及之三種隱私攻擊，包括單獨識別 (singling out)、可鏈結性 (linkability)，與推斷 (inference)，並針對上述隱私攻擊發展出隱私風險評估。

單獨識別攻擊指旨在資料集中識別出特定個體，使攻擊者能夠確定該個體的特定敏感屬性，從而損害其隱私。舉例來說，某資料集中，存在一筆資料，其有多個屬性的唯一組合 (例如：男性，65 歲，郵遞區號 30305，心臟病發作次數 4 次)。下圖擷取該研究之演講簡報，說明單獨識別攻擊，與其攻擊演算法。

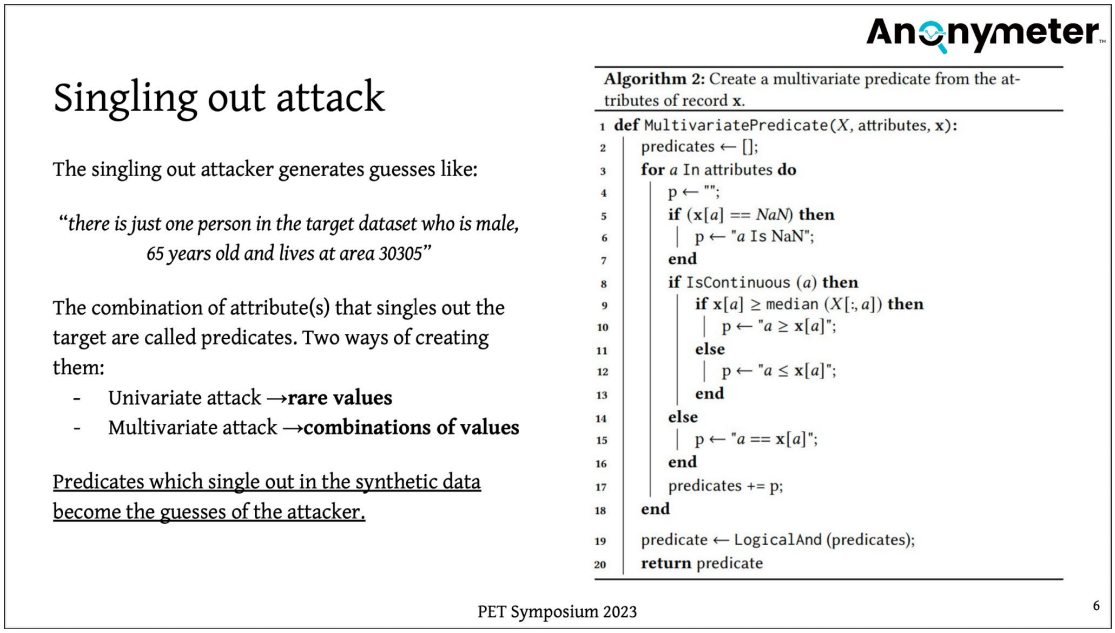


圖 1、單獨識別攻擊介紹

可鏈結性是指將兩個或多筆資料（同資料集或不同資料集中）與同一個體或一組個體聯繫在一起的可能性。換句話說，即在不同資料中找出關聯性，並將之交叉比對串連出特定隱私資訊。下圖擷取該研究之演講簡報，說明單獨識別攻擊，與其攻擊手法。

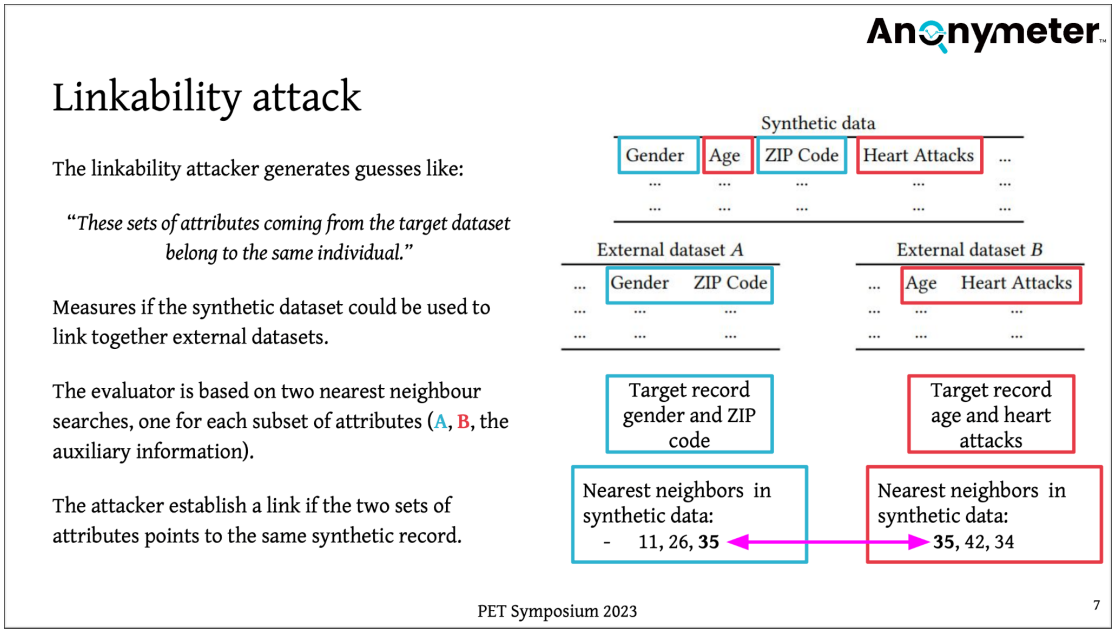


圖 2、可鏈結性攻擊介紹

推斷攻擊為當攻擊者取得部分欄位資料，即可依照背景知識或其他資訊推斷出隱蔽之敏感資訊。舉例來說，現有一病例資料，一名患者被記錄抽菸多年，並患有癌症，其資料集中，抽菸多年為公開資訊，患有何種癌症為隱私資訊，但依照邏輯可推斷出其罹患之癌症很大機率為肺癌。下圖擷取該研究之演講簡報，說明推斷攻擊。

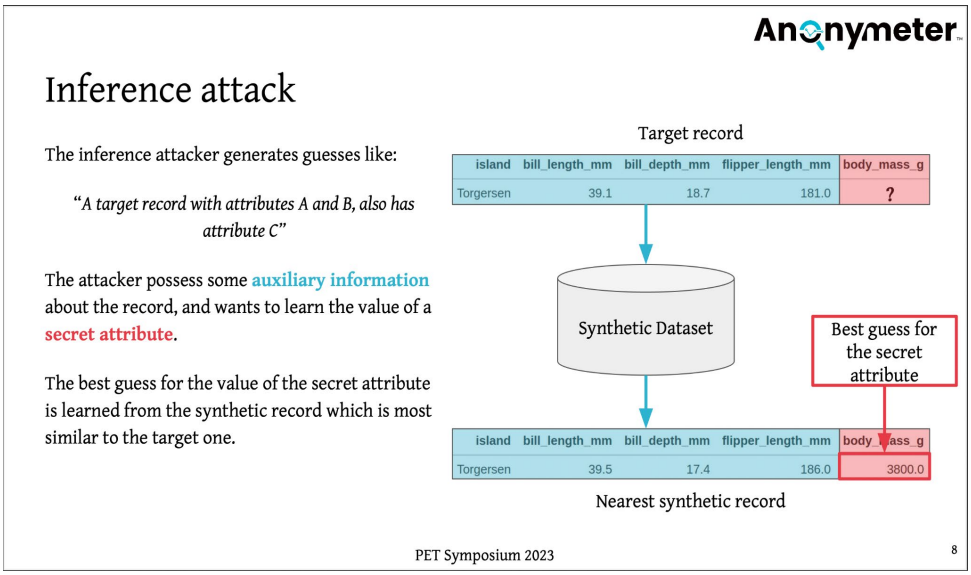


圖 3、推斷攻擊

儘管該研究所提出之驗測指標雖然著重於探討合成資料，但該指標適用於任何表格行之資料。換句話說，只要適用於表格資料之技術 (例如：K 匿名、差分隱私等)。同時，該研究亦發展出開源框架，對於我方驗測方法論之設計與實作皆具有很大參考價值。

2. Lessons Learned: Surveying the Practicality of Differential Privacy in the Industry

主要探討自 2006 年以來，差分隱私技術於學術界已成為重要量化資料隱私之統計工具。然而，儘管有大量的研究和開源工具，但差分隱私在企業領域的廣泛度仍然有限。該研究通過詳細訪談 9 家主要公司的 24 名隱私從業者，揭示產學界差距的根本原因。並於研究中嘗試回答一個最核心的問題：即「差分隱私在學術界受到高度重視，但為何在企業領域的採用仍然有限？」

針對此問題，該研究分析出企業導入差分隱私之困難有以下幾點：

1. 冗長的官僚審核程序
2. 影響既有之分析流程
3. 分析師擔憂資料精準度流失

雖在文章中，作者針對上述三個困難提出相應解決之道。但我們認為在實務上之複雜度可能高於作者之想像。雖如此，要能解決問題，得先認識與清楚問題之定義。因此，我們認為此文章最大貢獻在於協助企業認識與定義導入差分隱私技術之實務問題。此外，此文章亦提到，雖有許多差分隱私開源工具，但企業在使用差分隱私技術上仍非廣泛。作者認為，可能有一些要求與限制：

要求包括：

1. 高度的安全性：工具應該提供足夠的保護，以防止資料洩露或不正確的使用。
2. 易於使用：工具應該對非專家友好，並且有清晰的文件和指引。
3. 靈活性：工具應該能夠適應不同的資料集和使用情境。
4. 效能：工具應該能夠在合理的時間內處理大量資料。

限制包括：

1. 缺乏整合能力：工具可能難以與現有的資料基礎設施或其他工具整合。
2. 高成本：工具的採用和維護可能需要大量的資源和時間。

3. 缺乏培訓和支持：企業可能缺乏必要的知識和技能來正確使用和維護這些工具。

因此，若欲讓差分隱私技術在企業中獲得廣泛使用，我們或許可以在隱私強化技術指引上，加強上述章節，降低限制或困難產生的影響，進而提升企業的使用意願。

3. SoK: Differentially Private Publication of Trajectory Data

該研究主要發展差分隱私框架，該框架將差分隱私運用於特殊資料形式「軌跡資料」之隱私強化。相較於一般常見之資料形式，軌跡資料具有許多特性，例如：不同資料庫之軌跡資料皆不相同、結構與定義不統一。此外，研究發現，軌跡資料雖具有相當價值，可作為交通管理分析、路線建議或基礎設施發展。然而，其屬於使用者生活蹤跡之資訊，若該資訊洩漏，所造成之危害無法被忽視，因此該資料被視為重要隱私資料。為解決軌跡資料之隱私問題，該研究探討如何以差分隱私技術，對軌跡資料進行隱私強化。

該研究首先說明軌跡資料之分析結果，以解釋軌跡資料之價值與隱私問題，分析面相包含以下幾點：

- **軌跡資料之重要性：**該資料是否值得投入成本進行隱私保護。
- **軌跡資料之隱私風險：**該資料洩漏之風險與危害程度。
- **軌跡資料之複雜性：**該資料不同軌跡、位置間之關聯複雜。
- **軌跡資料之時效性：**該資料之更新頻率極高，然而每次的更新意味著洩漏部分資訊，使隱私保護更具挑戰。

資料分析後，接著探討差分隱私為何適用於軌跡資料之保護，發展其保護機制，並開發一差分隱私系統框架 (SoK)，最後利用真實個人軌跡資料進行實驗以確保其可用性。

該研究可作為我方進行差分隱私落地案例之重要參考依據，當案例所遇上之特殊資料類型，此研究提供了幾個分析方向，例如：特殊資料之分析面向、可用性評估等。此外，對於特殊資料類型，如何比較不同隱私強化技術之限制與對於該資料類型之適用性，該研究亦可作為重要參考依據。

4. A Worldwide View of Nation-state Internet Censorship

該研究主題為「從全球角度看待的國家網路審查」。探討了網路審查的全球現象，並深入研究了不同國家如何實施審查。

隨著網路成為人類歷史上最重要的通訊工具，其影響已超越了電視、報紙和廣播。然而，一些國家在其影響範圍內對網路通訊進行了審查。無論網路審查的動機是什麼——無論是基於意識形態、專制、法律、社會還是其他原因——網路審查研究已成為一個廣泛的跨學科努力（interdisciplinary effort）。

國家對其網路通訊施加了各種程度的審查。隨著全球人口對網路資源的訪問增加，一些政府更願意過濾內容、限制連接或拒絕訪問其影響範圍內的網路資源。為了更好地理解這一現象，作者從多個研究社群和學科中提取資料，提供了一個更全面的網路審查方法的視角。

講者還深入探討了網路審查的不同方法，如網路關閉、內容拒絕和深度封包檢查。這些方法可以是粗糙和直接的，例如網路關閉，或者更為精密和目標導向的，例如內容過濾。

為了更好地理解這些審查方法如何在全球範圍內實施，作者參考了多個資料來源，包括 OpenNet Initiative、OONI、Censored Planet Lab、ICLab、Citizen Lab 和 Freedom House。這些來源提供了關於如何和在哪裡發生網路審查的深入見解。

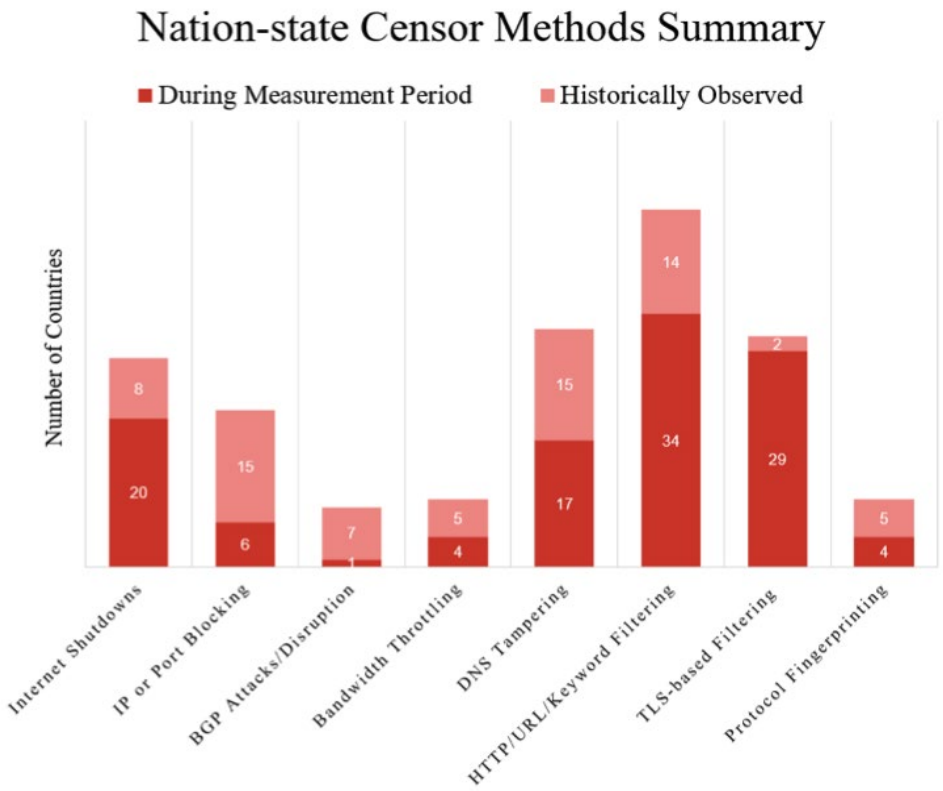


圖 4、進行審查國家數量長條圖
底部長條圖表示測量期間使用特定方法進行審查的國家數量；
頂部長條圖表示有審查歷史紀錄的國家數量（不包括測量期間）

5. Detecting Network Interference Without Endpoint Participation

該研究主要探討如何在不依賴終端參與的情況下檢測網路干擾。在全球，許多專制政權部署了全面的審查裝置來限制網路上的自由和開放通訊。過去的研究已經揭示了審查者如何運作，但這些方法主要集中在大國，如中國、伊朗和俄羅斯。對於小國，由於低網路滲透率、小人口和公共伺服器或遠程觀測點的稀缺，使用這些方法變得具有挑戰性。

為了解決這一問題，研究者提出了一種系統，該系統可以長時間地檢測網路審查，而不依賴於審查國家內的本地終端或志願者。該系統的核心思想是利用許多審查網路中間設備（middlebox）不完全遵循 TCP 協議的事實。作者使用 Geneva，一種基於遺傳演算法的工具，直接對審查者進行訓練，以發現可以達到特定目標的數據包序列。

研究中還介紹了一些初步結果，特別是在文萊和塔吉克斯坦的應用。在文萊，作者發現一個可以觸發審查的數據包序列。而在塔吉克斯坦，他們發現了一種策略，其中他們發送了兩次帶有被審查內容的 PSH+ACK 數據包，然後收到審查者注入的 RST+ACK 響應。

研究者提出了一種新的審查檢測方法，該方法不依賴於終端參與，並利用審查網路中間設備的設計選擇來實現。這種方法有望填補審查測量領域的一個空白，並為未來的研究提供了一個新的方向。

6. Towards a Comprehensive Understanding of Russian Transit Censorship

該研究核心是深入了解俄羅斯的網路過境審查。過去，有關俄羅斯過境審查的觀察大多是基於個案研究，而講者旨在提供一個更全面的研究。在過去的十年中，俄羅斯顯著增加了其審查強度，這引起了一個問題：俄羅斯的網路服務提供商是否不僅對在俄羅斯終止的流量實施審查政策，還對僅經過俄羅斯的流量實施審查政策。

論文中提到，俄羅斯在過去的十年中對其訊息控制進行了倉促的擴展。例如，2018 年，俄羅斯封鎖了 Telegram，導致封鎖了數百萬屬於 Amazon 和 Google 的雲托管平台的 IP 地址，進一步封鎖了許多無關的網站。

研究者在這項工作中首次對俄羅斯審查的「附帶損害」進行了全面研究。他們掃描了 18 個圍繞俄羅斯的國家的 IP 位址，試圖從俄羅斯審查者那裡獲得回應。研究發現，俄羅斯的附帶損害至少影響了這 18 個國家中的 9 個國家的部分流量，並且至少有 7 個 AS（自治系統）對過境流量進行了審查。

本研究提供了對俄羅斯過境審查「附帶損害」的初步但全面的了解，並強調了全球範圍內進一步研究附帶損害的必要性。

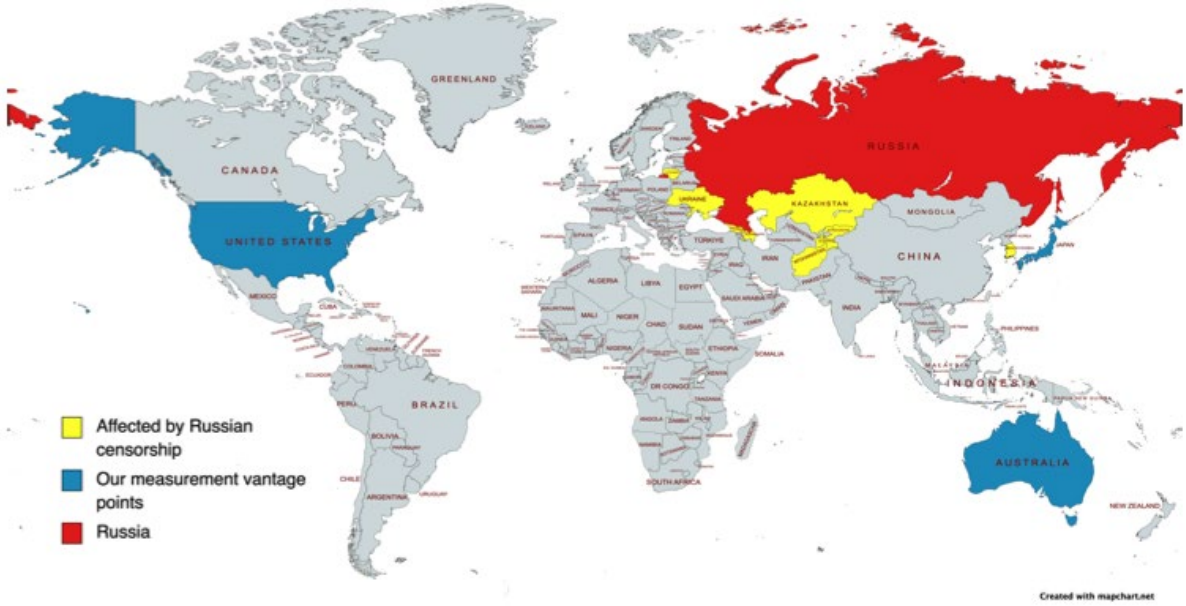


圖 5、受俄羅斯過境審查影響國家（黃色），測量觀察點（藍色）

7. The Use of Push Notification in Censorship Circumvention

該研究探討了推送通知在繞過網路審查中的應用。近年來，推送通知在行動和桌面平台上得到了廣泛的採用，因為它們為應用程式提供了一種直接向使用者提供時效性訊息的方式。作者認為，由於封鎖推送通知會造成重大的附帶損害，因此它們是理想的隧道繞過流量的候選者。

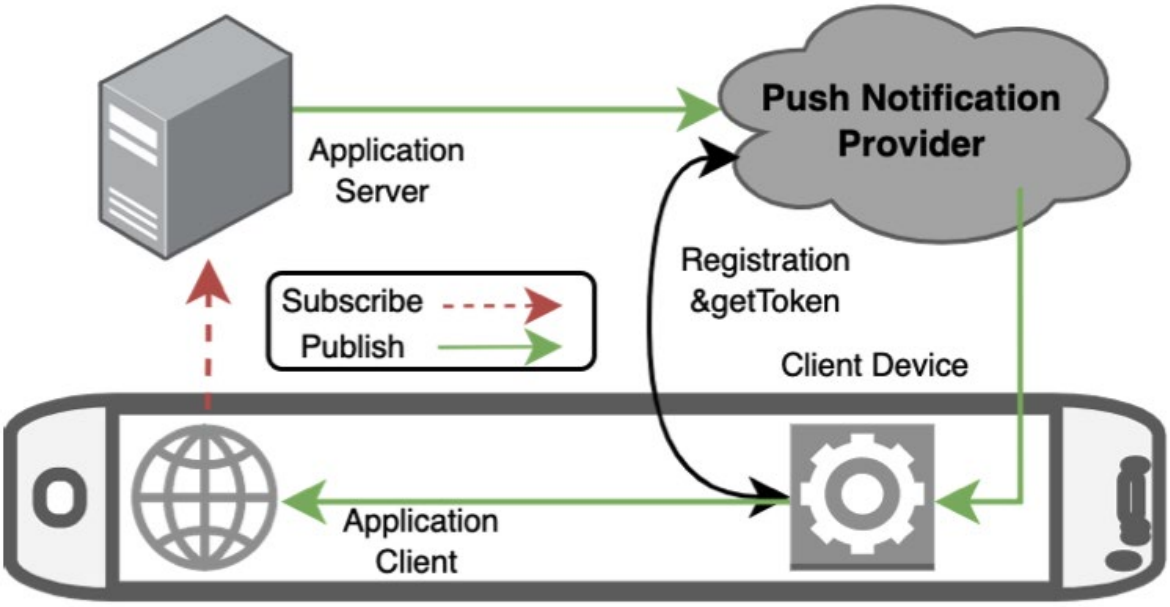


圖 6、通送通知服務之流程

論文中介紹了兩個利用推送通知作為傳輸的繞過審查系統。第一個是 PushRSS，它是一個阻擋抵抗的內容聚合器，可以通過推送通知網路隧道 RSS 更新。一旦啟動，即使伺服器 IP 被完全封鎖，該工具仍然可以執行。第二個是 PushProxy，它是一個通用的代理，可以通過推送通知服務路由使用者的下行流量，同時保持上行流量為獨立通道。通過解耦下行和上行，PushProxy 減少了網路對手（ network adversaries ）進行每流量分析（ per-flow analysis ）的能力，同時提供了與流行的對稱代理相當的性能（ offering performance comparable to popular symmetric proxies ）。

研究者通過大規模、長期的測量來評估推送通知作為繞過流量的傳輸的潛力。他們的發現表明，只有少數 AS （ Autonomous System ）積極干擾推送通知連

接，大多數封鎖事件都發生在中國的政治敏感時期。有趣的是，他們還發現了證據表明，中國的審查者對長期完全封鎖推送通知服務持謹慎態度，並在封鎖後僅幾天內做了白名單例外，這可能是由於會造成大量的附帶損害。

8. Proteus: Programmable Protocols for Censorship Circumvention

該研究主要內容是介紹 Proteus 系統，一個用於規避審查的可程式協議環境。隨著網路審查成為政治和社會控制的常見工具，反審查社群開發了多種工具來規避審查。Proteus 提供了一種環境，其中新的通訊協議可以被表達為簡潔且容易理解的規範文件。這種設計允許客戶端和代理快速回應新的審查策略，只需安裝新的規範文件。

Proteus 改進了先前的可程式設計，提高了主機安全性，防止惡意規範，提供了一種完整且對非技術人員容易理解的規範語言，並支持代理的多個同時協議，以實現版本控制和本地化。

研究者還提到了 Marionette 系統，一個通訊協議的可程式系統，它提供了一種語言和工具，使客戶端和代理容易寫入和安裝新的協議。這種設計允許通過重新配置代理協議快速規避新的審查方法。

Proteus 系統之目的是使使用者能夠快速應對不斷變化的審查環境。其主要設計目標是：使兩方能夠進行通訊、提供協議可程式性、提供安全的協議更新，以及允許平滑的更新。

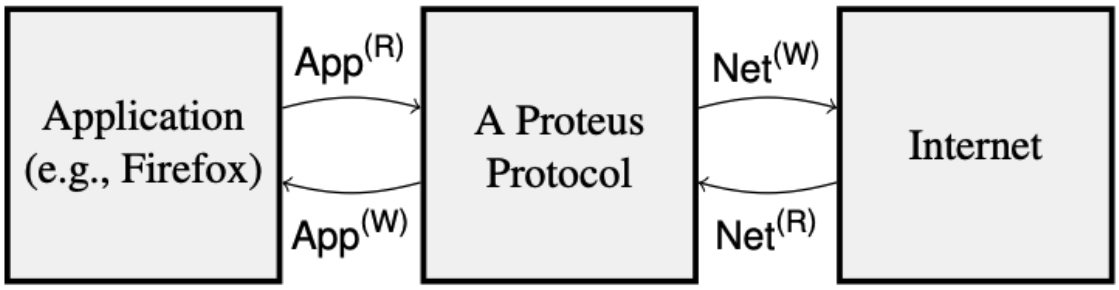


圖 7、讀取緩衝區和寫入緩衝區與 Proteus 協議的關係。
該協議通過讀取緩衝區接收資料，並通過寫入緩衝區輸出資料。

9. Voiceover: Censorship-Circumventing Protocol Tunnels with Generative Modeling

該研究探討了如何使用生成模型來規避審查的協議隧道。隨著審查制度不斷採用和部署先進技術來檢測和起訴網路上的公開通訊，多媒體協議隧道的技術旨在通過合法的音訊/影片通訊系統直接處理隱秘的數據通訊。然而，這些方法可以被審查者輕易地識別出來。

為了解決這個問題，作者們提出了一種新的流量整形技術（ traffic-shaping technique ），該技術模擬正常媒體內容，並將其屬性應用於隱秘內容。他們構建了一個生成的機器學習模型，以限制隱秘數據傳輸，使其時序屬性與真實的雙人對話學到的屬性相匹配。評估發現，模擬應用層內容中的時序屬性可以減少加密網路流量中的區別特徵，從而減少對粗粒度時序屬性的內容不匹配攻擊。

Voiceover 是他們對這個問題的方法，它是一個基於音訊的多媒體協議隧道。Voiceover 通過卷積編碼和正交幅度調製處理數據作為音訊。該音訊由 Skype 音訊通話在兩個端點之間傳輸，以越過審查者的邊界。他們的新方法是使用生成對抗網路學習和模擬對話時序屬性的分佈。這個模型塑造了 Voiceover 通過 Skype 傳輸的音訊，使得我們的隱秘協議隧道傳輸的時序屬性與真正的 Skype 通話的音訊相匹配。

研究者提出了一種新的設計，使用生成模型來減少在規避審查的協議隧道中的內容不匹配攻擊。此外，他們完全實現了這一設計作為一個開源原型進行評估，並促進未來的工作。

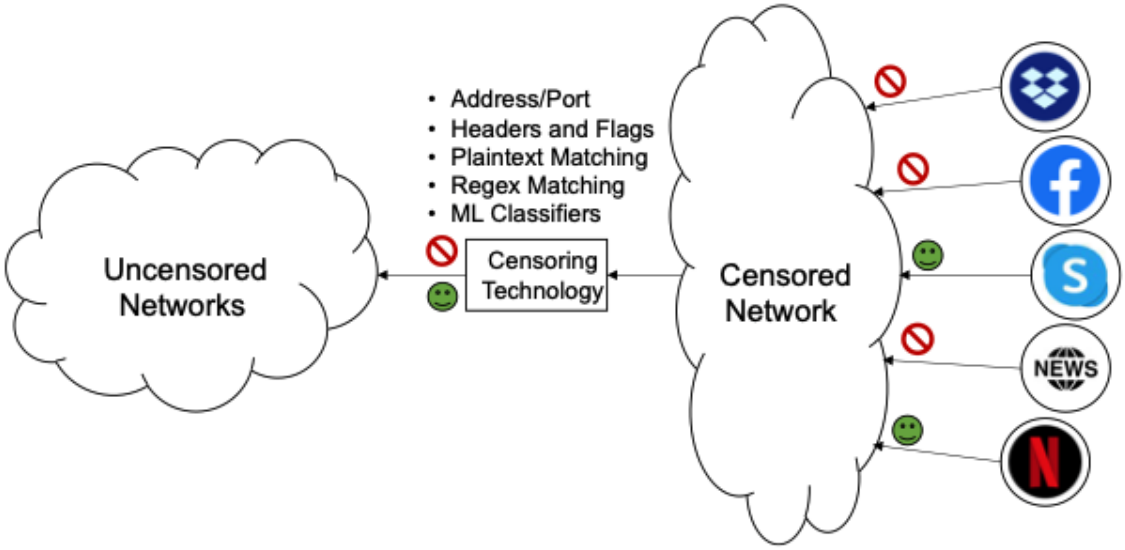

Address/Port
Headers and Flags
Plaintext Matching
Regex Matching
ML Classifiers
Uncensored Networks
Censoring Technology
Censored Network
NEWS

圖 8、威脅模型

10. Rethinking Realistic Adversaries for Anonymous Communication Systems

該研究主要內容是重新思考匿名通訊系統中的真實敵人。隨著網路的普及，人們越來越依賴於網路來進行通訊，但網路的設計並不支持隱私或匿名性。過去的研究已經開發了多種匿名系統，但這些系統通常都是基於某些特定的敵人模型來設計的。然而，現實中的敵人可能與這些模型有所不同，因此有必要重新考慮這些模型。

講者首先介紹了匿名通訊的背景，指出由於網路的設計，使用者必須依賴於覆蓋網路來獲得匿名性。過去的研究通常考慮了全局的被動或主動敵人，但這些敵人可能與現實中的敵人有所不同。講者接著討論了不完美的敵人，指出現有的文獻可能忽略了許多實際的限制，例如敵人的覆蓋範圍、觀察時間、互相不信任等。

接著提出了一些新的敵人模型，這些模型基於真實的監控計劃的限制。作者認為，為了擴展現有的敵人模型的現實性，我們必須引入影響範圍的概念。此外，敵人模型應該反映出擴展影響範圍是有成本的，這些成本可能包括時間、金錢、計算、存儲、政治資本等。

研究者重新考慮了匿名通訊系統中的真實敵人，並提出了一些新的敵人模型。這些模型更加符合現實中的敵人，因此可以幫助設計更加安全和有效的匿名系統。

11. Running a high-performance pluggable transports Tor bridge

該研究主要焦點是 Tor (The Onion Router) 的可插拔傳輸模型，該模型通過在本地的進程間通信中執行獨立的繞過代碼，將匿名性和繞過的問題分開。這種模型在尺度上存在問題，特別是對於像 meek 和 Snowflake 這樣的傳輸，它們的阻塞抵抗不依賴獨立管理的橋接器，而是將所有流量轉發到一個或幾個集中的橋接器。講者探討了當橋接器從 500 到 10,000 的同時用戶，然後從 10,000 到 50,000 的同時用戶進行擴展時會出現什麼瓶頸，並展示了基於他們執行 Snowflake 橋接器的經驗來克服這些瓶頸的方法。主要的想法是在橋接器主機上平行執行多個 Tor 進程，並使用外部同步的身份鍵。

研究中介紹了 Tor 的橋接器和可插拔傳輸如何為其核心功能的匿名性增加了阻塞抵抗（繞過審查）。橋接器是其網路地址不為全球所知的中繼，意味著它們應該難以被審查者發現和通過地址阻塞。可插拔傳輸是模組化的隧道協議，它封裝和偽裝一個內部協議，從而防止審查者識別 Tor 協議並基於此阻塞連接。

接下來討論了如何垂直擴展可插拔傳輸的 Tor 橋接器。關鍵技術是在同一主機上執行多個 Tor 進程，並使用相同的身份鍵。這解決了單個 Tor 進程受到 CPU 限制的最大瓶頸，但也引起了一些複雜性。此外，還需要考慮其他資源限制，如文件描述符和短暫的端口號的限制。這些建議來自於他們從 2021 年 12 月到 2023 年 2 月執行 Snowflake 橋接器的經驗，期間平均同時用戶數從 2,000 增加到大約 100,000。

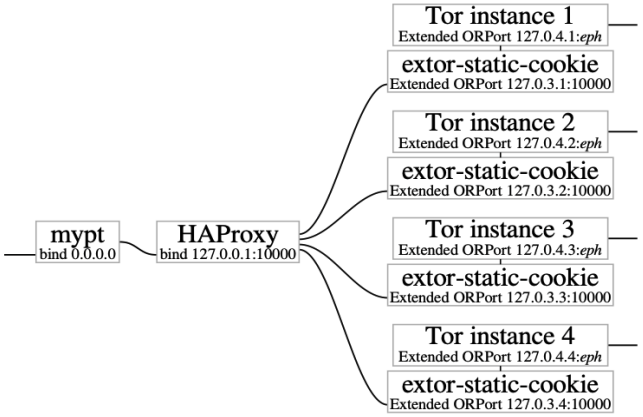


圖 9、多 Tor 設置

12. Crowdsourcing the Discovery of Server-side Censorship Evasion Strategies

該研究探討了伺服器端審查規避的最新進展。在這種方法中，伺服器在 TCP 三方握手過程中操作封包，代表客戶端繞過審查。這種策略的主要優點是客戶端不需要採取任何額外的行動或安裝額外的軟體。實際上，即使客戶端不知道他們正在被審查，他們也可以避免審查。這使得部署更加容易和安全，並已被多個審查規避工具採用。

儘管如此，伺服器端規避的一個主要限制是首先發現規避策略的困難。過去，研究人員研究了審查系統，並手工製作了可以繞過審查的封包操作策略。最近，研究人員提出了多種自動發現這些封包操作策略的解決方案，如 Geneva、Alembic 和 SYMTCP。但這些工具在訓練階段都需要客戶端的儀器化。

為了解決這個問題，作者提出了一種分佈式訓練模型，允許用戶訪問一個網站，使用未經修改的網頁瀏覽器，並在提供知情同意後，幫助發現審查規避策略。這意味著用戶的瀏覽器可以生成到我們控制的伺服器的禁止請求，我們可以利用與審查器的這些互動來發現審查規避策略。

此外，論文還討論了 Geneva 的訓練過程。Geneva 是一種基因演算法，可以發現封包操作策略來繞過審查。在伺服器端訓練中，客戶端只發送未修改的請求，觸發審查；所有的封包操作都在伺服器端進行。

最後，作者強調了保護用戶的重要性。他們要求所有用戶在參與訓練之前提供知情同意。此外，為了減少每個用戶觸發審查的次數，系統會限制用戶生成禁止請求的頻率。此外，為了保護用戶的隱私，只會收集最少量的用戶訊息。

作者提出了一種新的方法，允許用戶使用未經修改的瀏覽器參與伺服器端審查規避策略的發現，同時確保了用戶的安全和隱私。

13. Efficient decision tree training with new data structure for secure multi-party computation

本研究主要探討決策樹，這是一種經典的機器學習方法。儘管它在計算上相對簡單且容易解釋，但它仍然被廣泛使用。此外，決策樹也是其他機器學習方法，如梯度提升決策樹和隨機森林的重要組成部分。自從 Lindell 和 Pinkas 在隱私保護資料挖掘的早期工作以來，關於多方計算（MPC）協議的決策樹訓練已有大量的研究。

研究者提出了一種安全的多方計算（MPC）協議，該協議為給定的秘密共享資料集構建一個秘密共享的決策樹。先前基於 MPC 的決策樹訓練協議需要 O(2hmn log n) 的比較，其中 h 是樹的高度，n 和 m 分別是數據集的行數和屬性數。這篇論文通過安全的資料結構解決了這一問題，該結構使我們能夠計算每個組的聚合值，同時隱藏分組訊息。使用這種資料結構，我們可以在隱藏中間資料大小的情況下訓練決策樹。

此外，研究者還提出了一種安全的資料結構，可以在保持分組訊息私密的情況下計算每個組的聚合值。這種資料結構可以用於計算每個組內的聚合值，如總和和最大值，即使在決策樹訓練之外的情況也是如此。

為了展示其協議的實用性，研究者在基於三方秘密共享方案的 MPC 框架上實現了它。講者實施結果顯示，對於具有 220 行和 11 個屬性的數據集，其協議在 404 秒內訓練了一個高度為 4 的決策樹。

14. How Much Privacy Does Federated Learning with Secure Aggregation Guarantee?

研究者先提到隨著資料隱私和安全性的日益重要，聯合學習（Federated Learning，縮寫為 FL）逐漸受到了學術界和工業界的關注。FL 允許多個使用者在本地私有資料上協同訓練機器學習模型，而不需要與中央伺服器共享他們的私有本地資料。然而，即使使用者的資料保持在設備中且永遠不與伺服器共享，隱私仍然不能得到保證。這是因為使用者的訓練資料上的計算以訓練好的本地模型的形式被共享，這些本地模型可能會洩露有關本地資料集的訊息。

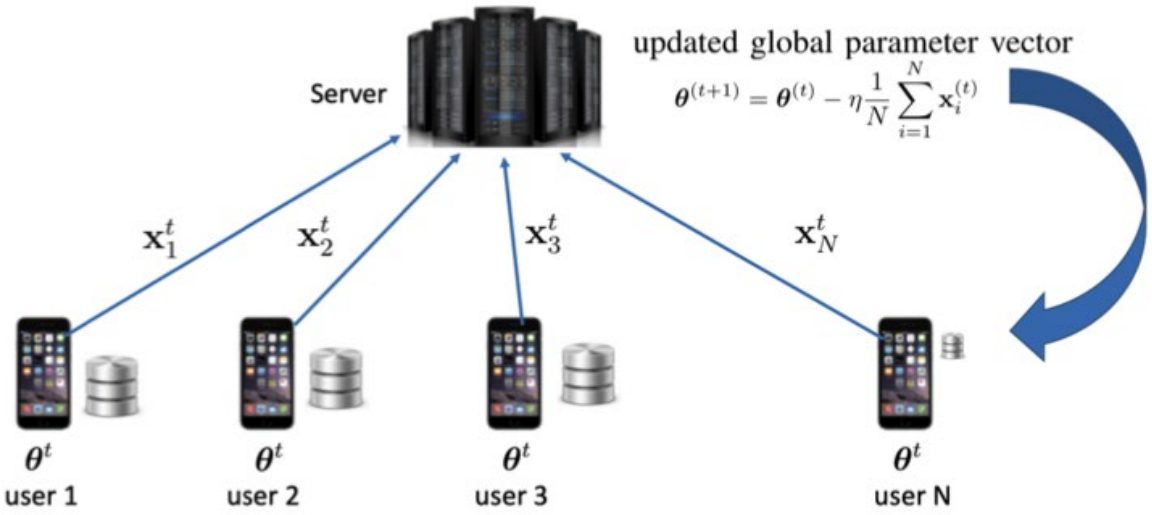


圖 10、聯合學習之訓練過程

為了解決這一問題，研究者提出了安全聚合（Secure Aggregation，縮寫為 SA）作為一種框架，以在 FL 中保護隱私。SA 確保伺服器只能學習全局聚合的模型更新，而不是單個模型的更新。但是，儘管如此，FL 與 SA 實際上可以提供多少隱私仍然沒有正式的保證。

研究者首次對 FL 與 SA 的正式隱私保證進行了分析。他們使用交互資訊（Mutual Information，縮寫為 MI）作為一種量化指標，來衡量在 FL 中，通過聚合模型更新可能洩露的每個使用者資料集的訊息量。具體來說，他們推導出了關於每個使用者資料集可以通過聚合模型更新洩露的訊息的上界。

當使用 FedSGD 聚合算法時，研究者的理論界限顯示隱私洩露的數量與參與 FL 與 SA 的使用者數量成線性減少。為了驗證他們的理論界限，研究者使用了一個交互資訊神經估計器來實證性地評估在不同 FL 設置下的隱私洩露，包括在 MNIST 和 CIFAR10 資料集上。

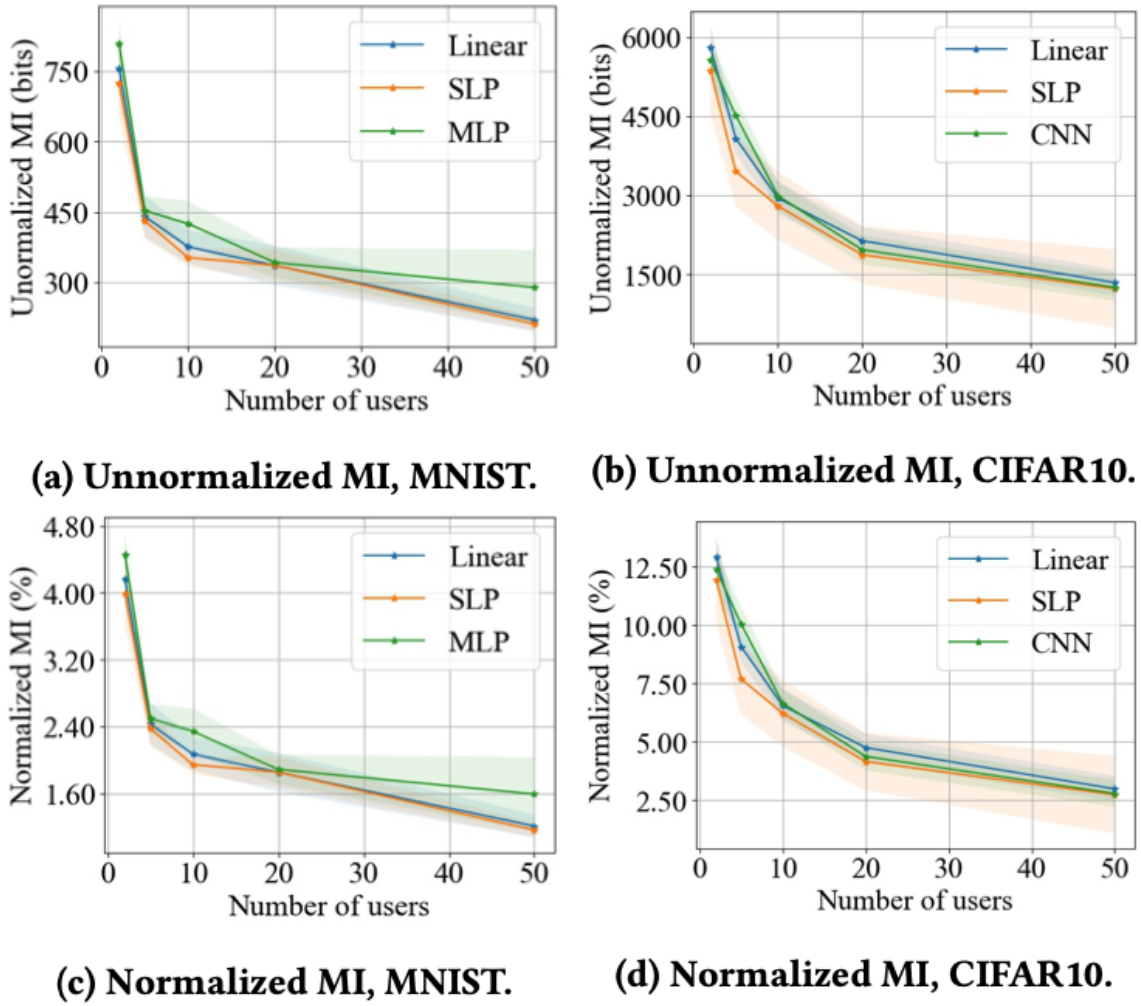
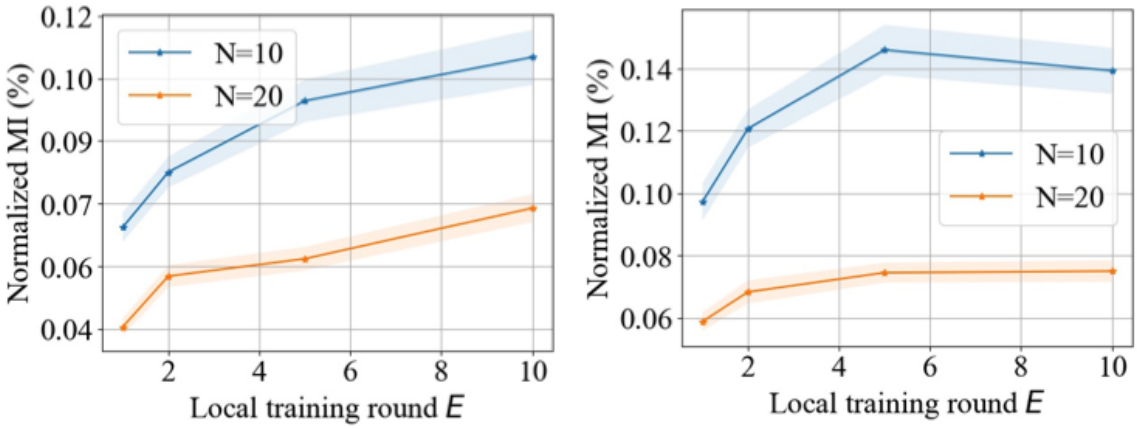


圖 11、使用 FedSGD 時使用者數量（N）的影響。

此外，該研究者還探討了如何使用交互資訊神經估計器來抽象地量化訊息洩露的量，而不是重建圖像的數量。他們還實證性地研究了系統參數對隱私洩露的影響，並發現隨著使用者數量和本地批量大小的增加，隱私洩露減少，而隨著訓練輪次的增加，隱私洩露增加。



(a) Normalized MI, MNIST. **(b) Normalized MI, CIFAR10.**

圖 12、使用 FedAvg 時，本地訓練次數（E）對其影響。

研究者提供了一個深入的分析，關於在聯合學習中可能洩露的隱私訊息量，並使用交互資訊作為一種量化指標來衡量這一洩露。這項研究為想要在不洩露使用者資料的情況下進行機器學習的研究者和工程師提供了有價值的見解。

肆、 心得與建議

- **本次研討會參與心得：**

  此次研討會我們觀察到較有趣的主題包含兩個面向：隱私強化技術、資料隱私。對於隱私強化技術面，許多研究探討機器學習與 AI 技術之訓練模型使用之資料可能具有隱私洩露之風險，包含：訓練資料若為敏感資訊，是否會造成隱私問題，以及各種針對機器學習與 AI 模型之資料竊取攻擊，相關研究有：「On the Privacy Risks of Deploying Recurrent Neural Networks in Machine Learning Models」與「Towards Sentence Level Inference Attack Against Pre-trained Language Models」等。因此，隨著機器學習與 AI 之快速發展，其隱私問題之討論日漸增加，故隱私強化技術導入機器學習與 AI 已為一種趨勢。

  於資料隱私面，許多研究開始探討現今之資料型態已不同以往，只包含數據與文字等資訊，如今影、音、圖之資料亦被廣泛使用 (例如：圖片生成 AI 之訓練資料)。此外，地理軌跡、生物辨識、基因資訊等隱私問題也被許多研究發起探討，相關研究包含：「"My face, my rules": Enabling Personalized Protection against Unacceptable Face Editing」、「I-GWAS: Privacy Preserving Interdependent Genome-Wide Association Studies」以及「SoK: Differentially Private Publication of Trajectory Data」等。我們認為，資料隱私問題與 AI 隱私問題具有一定程度之關聯，由於機器學習與 AI 仰賴大量資料進行模型訓練，除模型本身之隱私風險外，用於訓練之資料亦具有隱私疑慮，且如今新型資料型態使用需求廣泛，如何將隱私強化技術應用於此類資料亦成為一種趨勢。

- **對數位發展部未來發展隱私強化技術之心得與建議：**

  為利數位部推廣各公部門與 NPO/NGO 組織共享資料，以達到數據公益之目的，隱私強化技術之推廣亦相當重要。從此次研討會中，部分研究表明隱私強化技術在學術領域之發展良好，但欲在企業導入隱私強化技術則遇上許多困難，相關研究如「Lessons Learned: Surveying the Practicality of Differential Privacy in the Industry」，該研究探討企業在導入隱私強化技術之限制，包含：缺乏整合能力、高成本與缺乏培訓支持。欲解決這些限制，建議於隱私強化技術導入企業之前，先將各個技術之工具進統合，並設計隱私強化技術之框架。接著，以該框架為基礎，發展各技術之範例情境，以利企業欲導入時可作為參考，以解決「缺乏整合能力」之限制。在技術工具之完善後，建議撰寫完整指引，並於指引中詳述導入之流程，以降低企業導入隱私強化技術之成本，最後搭配指引安排企業員工之培訓，推廣隱私強化技術之應用。如此一來，在前期有完整技術發展，並且有詳細文獻參考，應使企業對於導入隱私強化技術之意願提高，以利促進數據公益。

- **對 TTC 未來發展隱私強化技術之心得與建議：**

  於本次研討會中，許多議題探討隱私強化技術之驗測方法，可作為我方驗測技術之參考。相關研究如「A Unified Framework for Quantifying Privacy Risk in Synthetic Data」提及合成資料之驗測方法，其研究中參考歐洲 GDPR 規範，並發展出合成資料之評估指標，而該研究雖然著重於合成資料驗測之探討，但其驗測指標亦適用於各類表格資料。因此，諸如 K 匿名、差分隱私等隱私強化技術，我方亦可作為參考，以補強、優化現有驗測方法論。根據該研討會的眾多研究主題，皆有提到參考 GDPR 之規範，因此建議 TTC 未來在發展隱私強化技術可優先針對 GDPR 進行研析，以利加強我方對於隱私議題之知識基礎。此外，從研討會之議題可觀察出，AI 導入隱私強化技術之趨勢，建議 TTC 對於相關議題之研究亦可先行部署研究，以加強 TTC 對於 AI 與隱私強化技術之研發能量，為後續計畫打下基礎。

伍、 附錄

The 23rd Privacy Enhancing Technologies Symposium (PETS) *July 10–15, 2023*

SoK: New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks

Charles Gouert*, **Dimitris Mouris***, Nektarios G. Tsoutsos

* The first two authors have equal contribution.

cgouert@udel.edu, ***jimouris@udel.edu***, *tsoutsos@udel.edu*

Trustworthy Computing Group

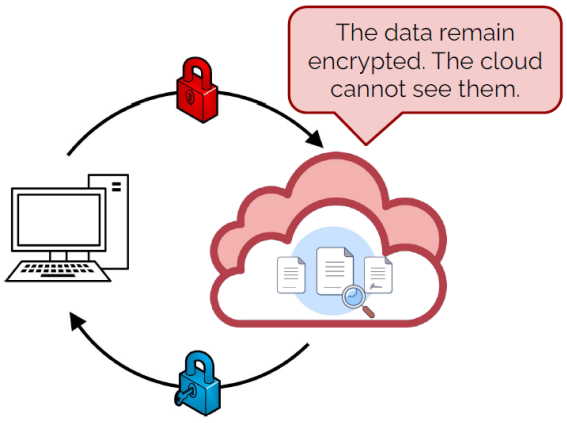
University of Delaware

圖 13、 New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks

Why care about Homomorphic Encryption (HE)?

Homomorphic Encryption (HE) enables computing directly on encrypted data!

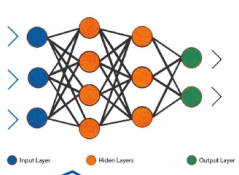
"Outsource your data processing, not your privacy".



2

圖 14、New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks

Motivation



Privacy-preserving ML

Encrypted Sorting

Encrypted Search

- Many HE libraries and schemes are actively being developed.
- Each scheme has its own benefits *(batching, bootstrapping, precision, primitive operations)*
- Each library exposes its own API.

How do you choose which HE library to use for a given application?

3

圖 15、New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks

HE Types & Schemes

Partial (PHE): Addition OR multiplication.

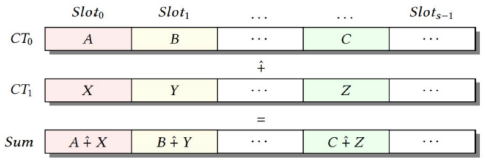
Leveled (LHE): Bounded addition AND multiplication.

Fully (FHE): Unbounded addition AND multiplication.

Uses **bootstrapping** to allow unlimited operations

Binary (CGGI): Encrypt each bit in one ciphertext.

- Compose algorithms as Boolean circuits.
- Fast bootstrapping.
- Low throughput.



Integer (BGV/BFV): Encrypt integers modulo *p*.

- HE add/mult → plaintext modular add/mult.
- Slow bootstrapping.
- High throughput.

Floating-Point (CKKS): Encrypts FP numbers (up to a precision).

- HE add/mult → plaintext real number add/mult.
- Slow bootstrapping.
- High throughput.

4

圖 16、New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks

HE Libraries

HElib: BGV, CKKS

Lattigo: BFV, BGV, CKKS

PALISADE (OpenFHE): leveled BFV, BGV, CKKS, CGGI, DM

SEAL: leveled BFV, BGV, CKKS

TFHE: CGGI

Other Libraries: tfhe-rs/Concrete, FHEW, HEAAN

Each library implements different schemes and exposes different API

Library	Language
Concrete [26]	Rust
FHEW [35]	C, C++
HEAAN [23]	C, C++
HElib [48]	C, C++
Lattigo [78]	Go
PALISADE [69]	C, C++
SEAL [72]	C, C++, .NET
TFHE [25]	C, C++

Different optimizations

Different programming languages

5

圖 17、New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks

Our Contribution

How do you choose which HE library to use for a given application?

T2 is a framework for comparing HE libraries/schemes.

- ↪ **One** high-level language (called T2DSL)
- ↪ T2DSL **compiler** to **many HE** libraries/schemes.
- ↪ **Benchmark suite** (in T2DSL).

6

圖 18、 New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks

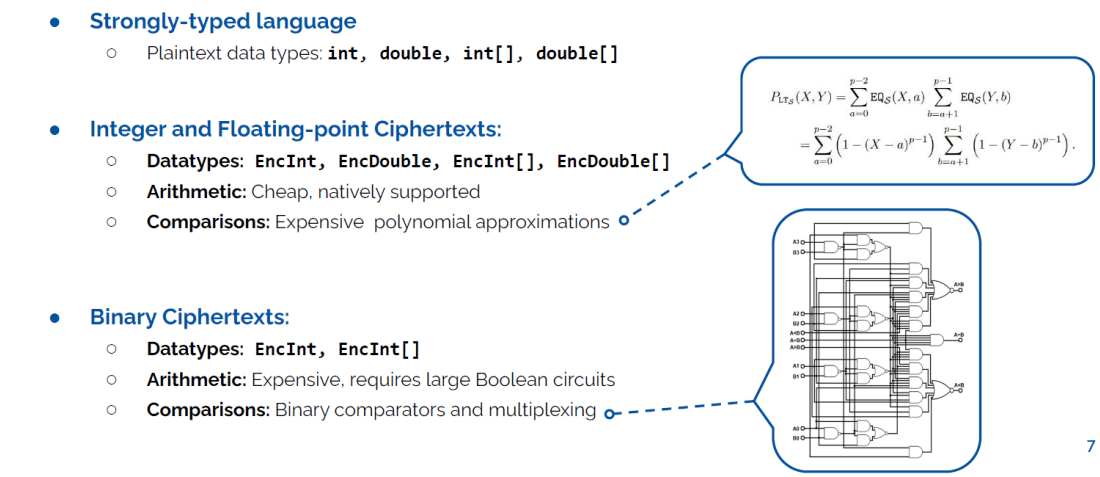


圖 19、New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks

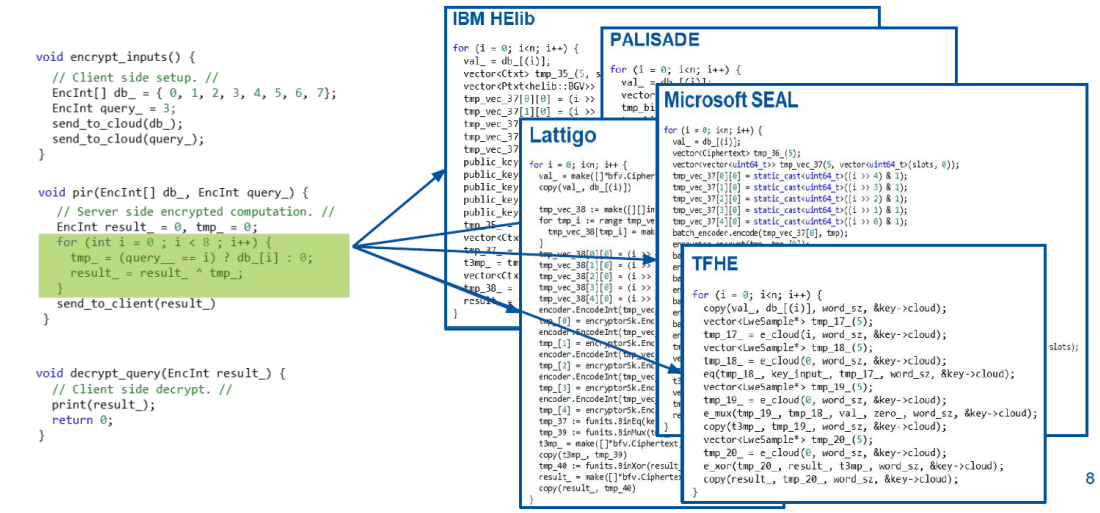


圖 20、New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks

T2 Benchmark Suite

Arithmetic: contain integer/floating-point additions and multiplications

Bitwise: composed primarily of logic gate operations as well as shifts.

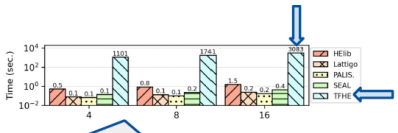
Relational: contain numerous comparisons over encrypted data

Arithmetic	Bitwise	Relational
Chi-Squared	Binary Manhattan Distance	Hamming Distance
Machine Learning Inference	Binary Private Information Retrieval (BinPIR)	Relational Manhattan Distance
Logistic Regression (LR)	Cyclic Redundancy Check (CRC)	
Matrix Multiplication	Batcher Sorting Network	
Batched Private Information Retrieval (BaPIR)	Insertion Sort	
Squared Euclidean Distance		

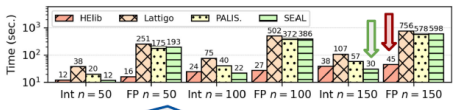
9

圖 21、New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks

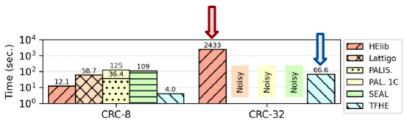
Experimental Evaluations



Logistic Regression: TFHE performs poorly due to the large amount of arithmetic operations (need Boolean circuits).



Neural Network: HElib exhibits great performance for encrypted floats, while TFHE can not evaluate in a reasonable amount of time.



CRC: TFHE is the fastest in this case. Due to the computational depth, some libraries cannot compute CRC-32.

More experiments in the paper!

10

圖 22、New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks

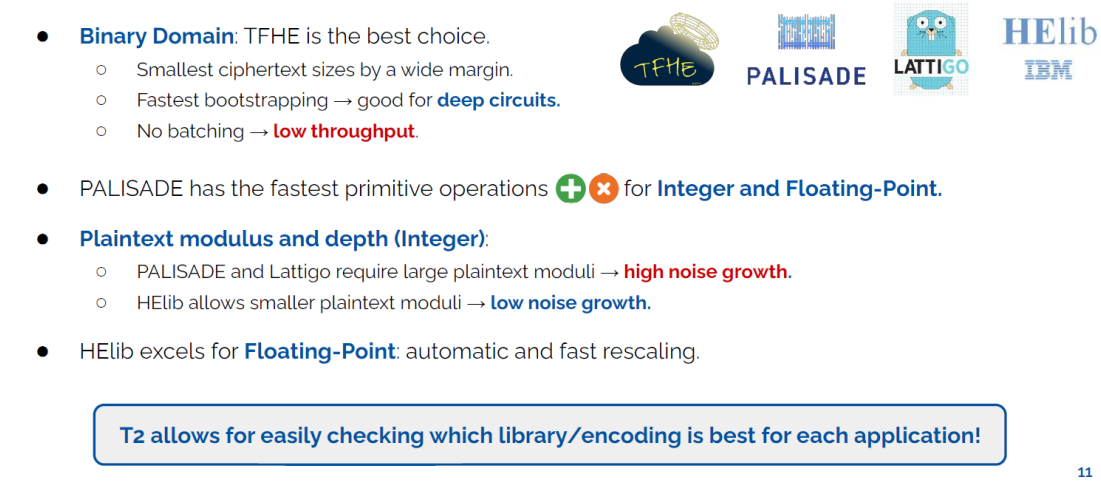


圖 23、New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks

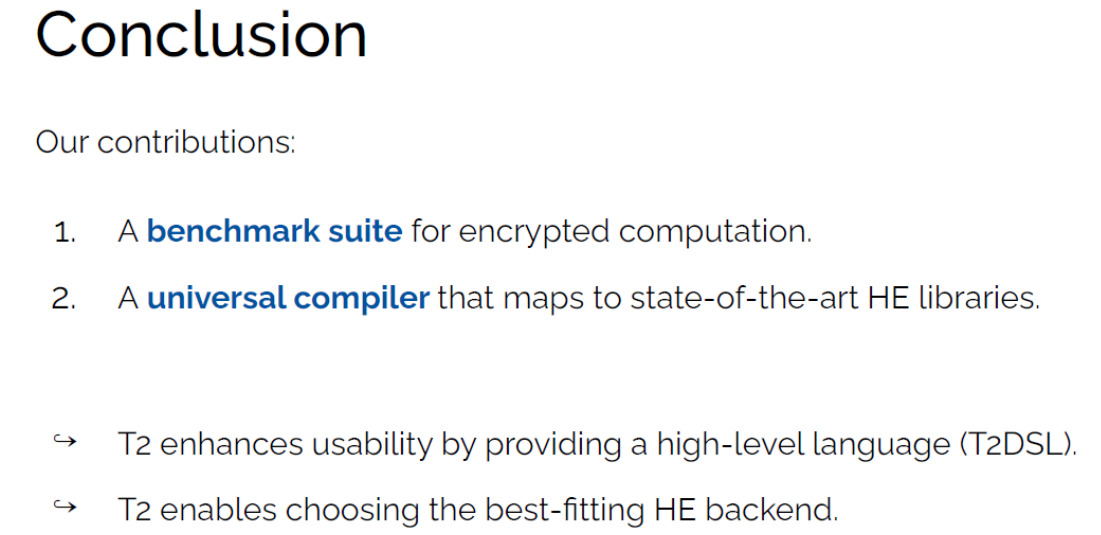


圖 24、New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks

Anonymeter

A Unified Framework for Quantifying Privacy Risk in Synthetic Data

Matteo Giomi, Franziska Boenisch
Christoph Wehmeyer, Borbála Tasnádi

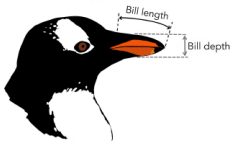
Statice
ANONOS

PET Symposium 2023

VECTOR INSTITUTE

圖 25、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

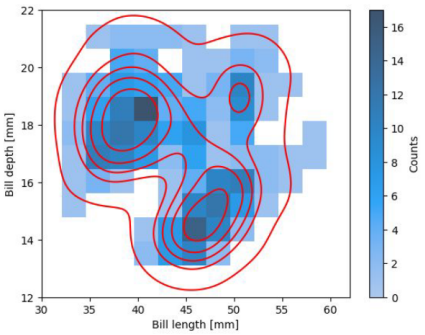
Synthetic data in a nutshell



A) learn the probability distribution of the dataset, **B)** draw samples from it.

	bill_length_mm	bill_depth_mm
0	41.5	18.5
1	35.3	18.9
2	43.6	13.9
3	51.3	14.2
4	36.5	18.0

ORIGINAL DATA



	bill_length_mm	bill_depth_mm
0	39.6	17.2
1	49.0	19.5
2	42.0	19.5
3	47.8	15.0
4	38.3	19.2

SYNTHETIC DATA

PET Symposium 2023

2

圖 26、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

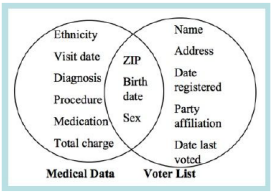
Anonymeter

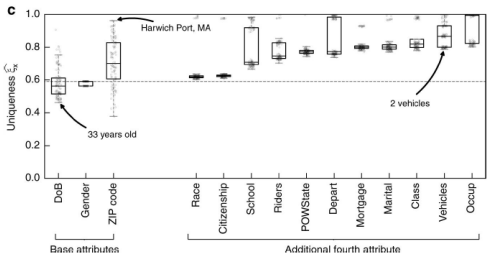
Synthetic data and privacy

Synthetic data as some advantages with respect to "traditional" anonymization techniques:

- It **breaks 1-1 mapping** between sensitive records and "protected" ones.
- It does not need to **distinguish quasi-identifiers** from other attributes.

However, one cannot assume that generating a synthetic data eliminates all privacy risks.





Rocher, L., et al. Estimating the success of re-identifications in incomplete datasets using generative models.
Sweeney, Latanya. Weaving Technology and Policy Together to Maintain Confidentiality.

PET Symposium 2023 3

圖 27、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

Anonymeter

Anonymeter

Anonymeter is a software that measures the privacy risks for **tabular synthetic data**. It measures the three privacy risks indicated by the Article 29 Data Protection Working Group Party as essential to anonymization:

- **Singling Out**
- **Linkability**
- **Inference**

It has been positively evaluated by the CNIL (French data protection authority):

> *"The results produced by the tool Anonymeter should be used by the data controller to decide whether the residual risks of re-identification are acceptable or not, and whether the dataset could be considered anonymous."*

PET Symposium 2023 4

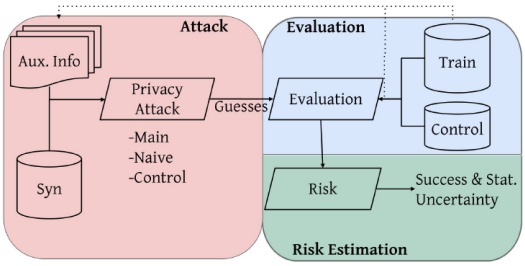
圖 28、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

Anonymeter

Methodology

Attack phase:

- The attacker has **access to the full synthetic data**, and some **auxiliary information** coming from the original dataset.
- the attacker **generates guesses** on a set of target records.



Evaluation phase: the guesses from the attacker are evaluated comparing them with the truth.

Risk estimation: comparing attacker success on different population of targets we separate utility from privacy and derive an estimate of the risk.

圖 29、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

Anonymeter

Singling out attack

The singling out attacker generates guesses like:

"there is just one person in the target dataset who is male, 65 years old and lives at area 30305"

The combination of attribute(s) that singles out the target are called predicates. Two ways of creating them:

- Univariate attack →**rare values**
- Multivariate attack →**combinations of values**

Predicates which single out in the synthetic data become the guesses of the attacker.

Algorithm 2: Create a multivariate predicate from the attributes of record x.

```
def MultivariatePredicate(X, attributes, x):
    predicates ← [];
    for a In attributes do
        p ← "";
        if (x[a] == NaN) then
            p ← "a Is NaN";
        end
        if IsContinuous (a) then
            if x[a] ≥ median (X[:, a]) then
                p ← "a ≥ x[a]";
            else
                p ← "a ≤ x[a]";
            end
        else
            p ← "a == x[a]";
        end
        predicates += p;
    end
    predicate ← LogicalAnd (predicates);
    return predicate
```


圖 30、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

Anonymeter

Linkability attack

The linkability attacker generates guesses like:

"These sets of attributes coming from the target dataset belong to the same individual."

Measures if the synthetic dataset could be used to link together external datasets.

The evaluator is based on two nearest neighbour searches, one for each subset of attributes (A, B, the auxiliary information).

The attacker establish a link if the two sets of attributes points to the same synthetic record.

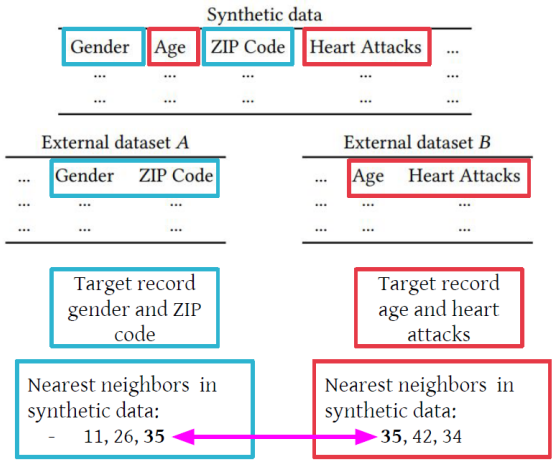


圖 31、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

Anonymeter

Inference attack

The inference attacker generates guesses like:

"A target record with attributes A and B, also has attribute C"

The attacker possess some **auxiliary information** about the record, and wants to learn the value of a **secret attribute**.

The best guess for the value of the secret attribute is learned from the synthetic record which is most similar to the target one.

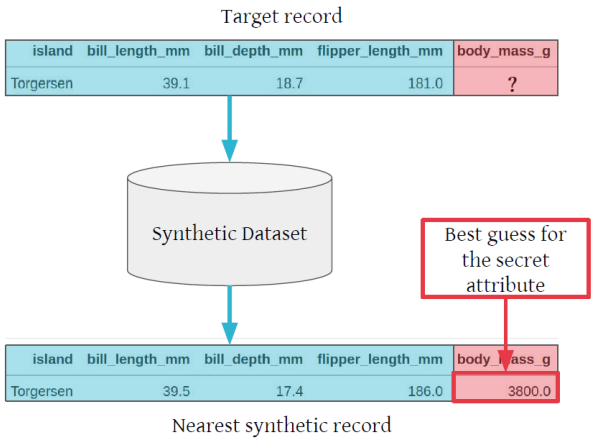


圖 32、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

Anonymeter

From attack success to risk estimate

Measuring the success of a privacy attack is **not enough** to derive a reliable measure of the privacy risk.

There are two more questions that one needs to answer:

- Can I trust my attack?
  - Implements **naive attacks** for each of the three risks.
  - They provide a **statistical baseline** to estimate the attacker strength.
  - Only attacks which are better than the naive one can be "trusted"

- How can I distinguish privacy from utility?

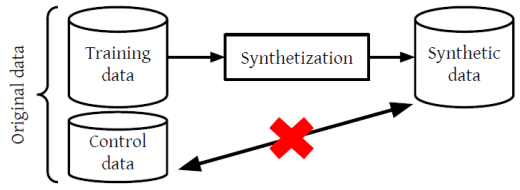
圖 33、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

Anonymeter

How to distinguish privacy from utility

Anonymeter makes use of a control dataset: a set of original records which was not used to generate the synthetic data.

There is **no information specific to the control records** in the synthetic data. Whatever the attacker learns about the control records is **only due to utility**.

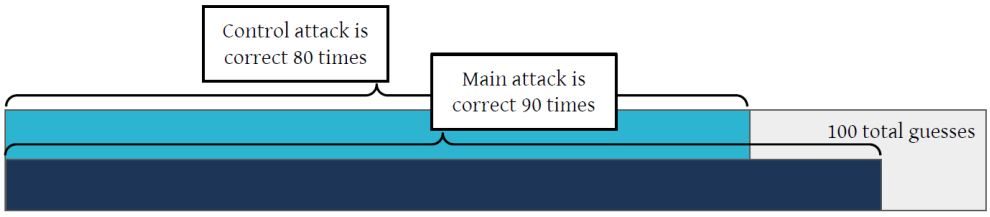


There is a privacy risk if the attacker learns more about the training dataset than about the control one.

圖 34、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

Anonymeter

From success rates to risk estimate



- 80% of the correct guesses is due to utility.
- **On top of that**, the attacker is able to succeed another 10% of the times.
- The most powerful attacker in the world can't be better than 100% correct. That is, 20% better than against control.
- The risk is 50%.

Excess of correct guesses

$$R = \frac{R_A - R_C}{1 - R_C}$$

Margin of improvement

圖 35、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

Anonymeter

Benchmarking Anonymeter

Leaky synthesization: starting from a zero-risk situation (the just a few scenario) we add records from the training set into the released one → ground truth for the risk!

The leaky synhtetizations are evaluated modeling attackers with various amount of auxiliary knowledge.

The metrics of anonymeter are **well calibrated**: the risk scales linearly with the fraction of leaks.

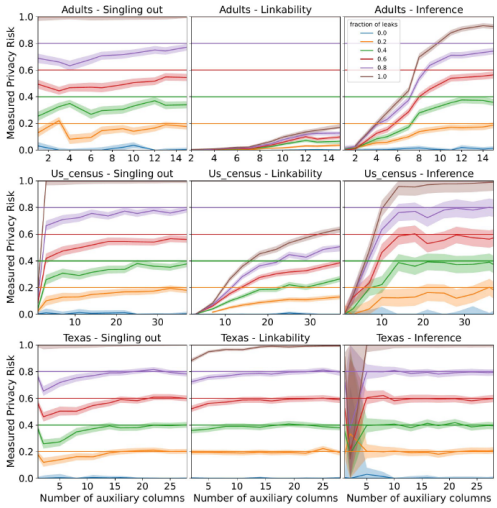
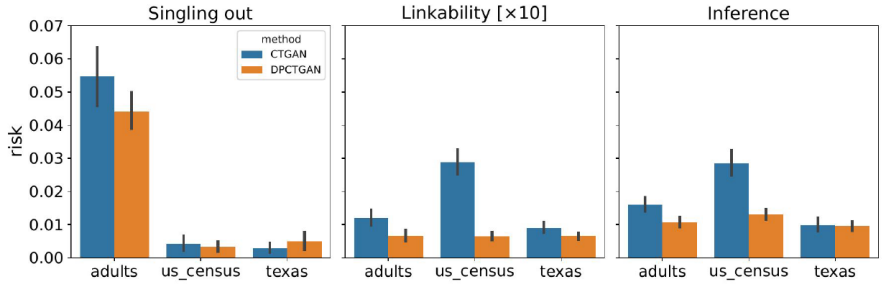


圖 36、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

Anonymeter

Evaluating the impact of DP on synthetic dataset

Evaluated synthetisations from CTGAN with and without differential privacy.



- DP effective in reducing all the risks.
- Linkability risk is much lower than the other two in all cases.

圖 37、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

Anonymeter

Using Anonymeter

The code is open source under BSD Clear.

The evaluators have a common interface and a few parameters to tweak:

- **Statistical uncertainties**
- **Attack model** (AUX info, target)

To test it, just install it:

```
pip install anonymeter[notebooks]
```

```python
from anonymeter.evaluators import InferenceEvaluator

evaluator = InferenceEvaluator(ori: pd.DataFrame,
                               syn: pd.DataFrame,
                               control: pd.DataFrame,
                               aux_cols: List[str],
                               secret: str,
                               n_attacks: int
                              )
evaluator.evaluate()
risk = evaluator.risk()
```

and play with the example notebook (data included).

圖 38、A Unified Framework for Quantifying Privacy Risk in Synthetic Data

圖 39、Differentially Private Speaker Anonymization

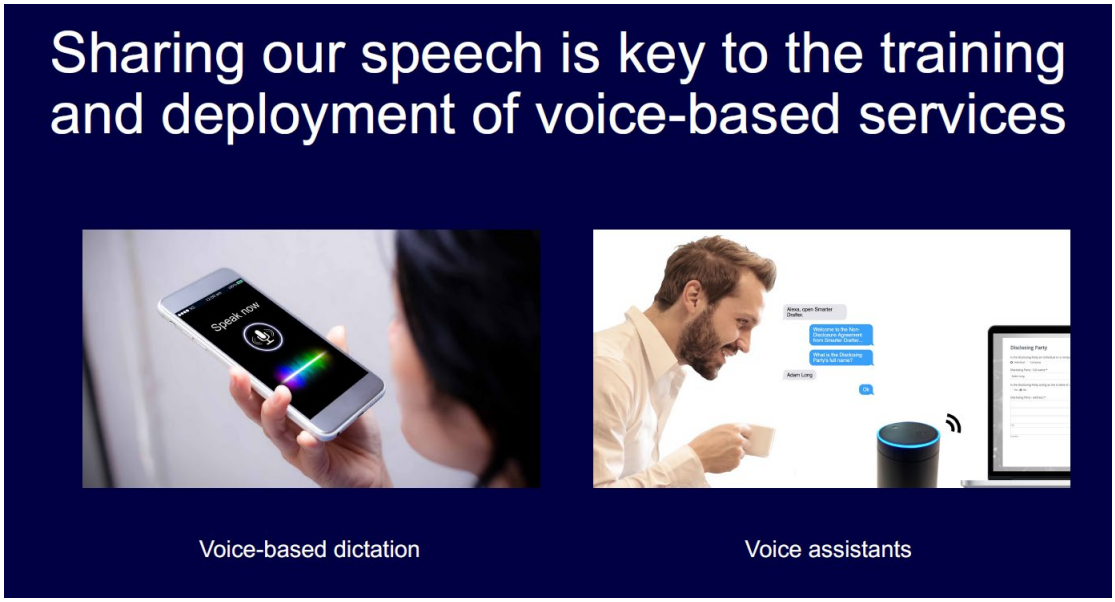


圖 40、Differentially Private Speaker Anonymization

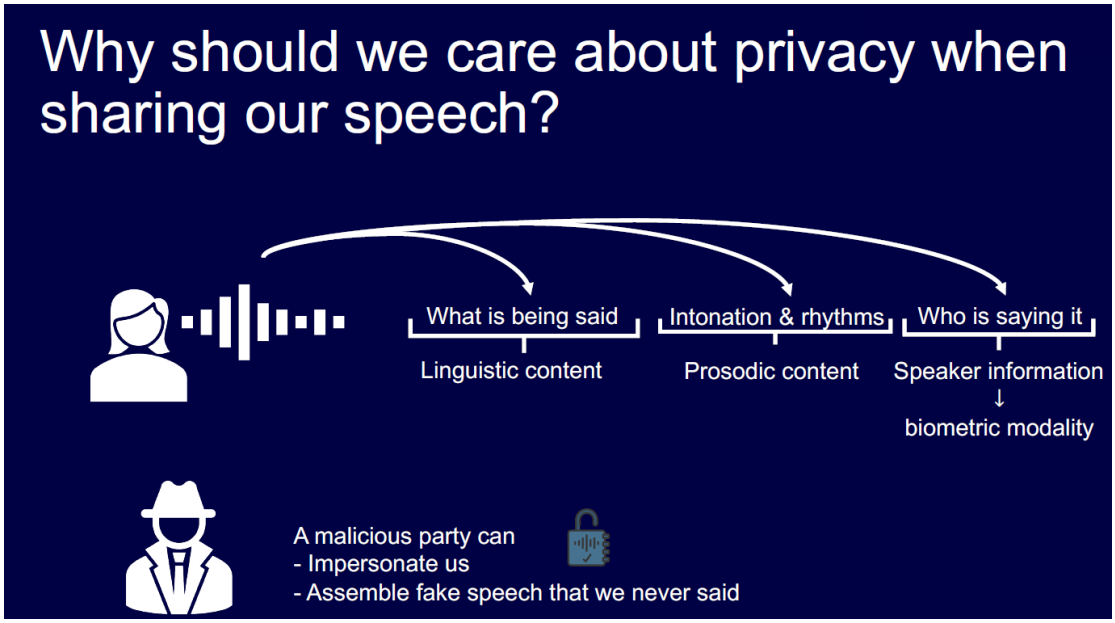


圖 41、Differentially Private Speaker Anonymization

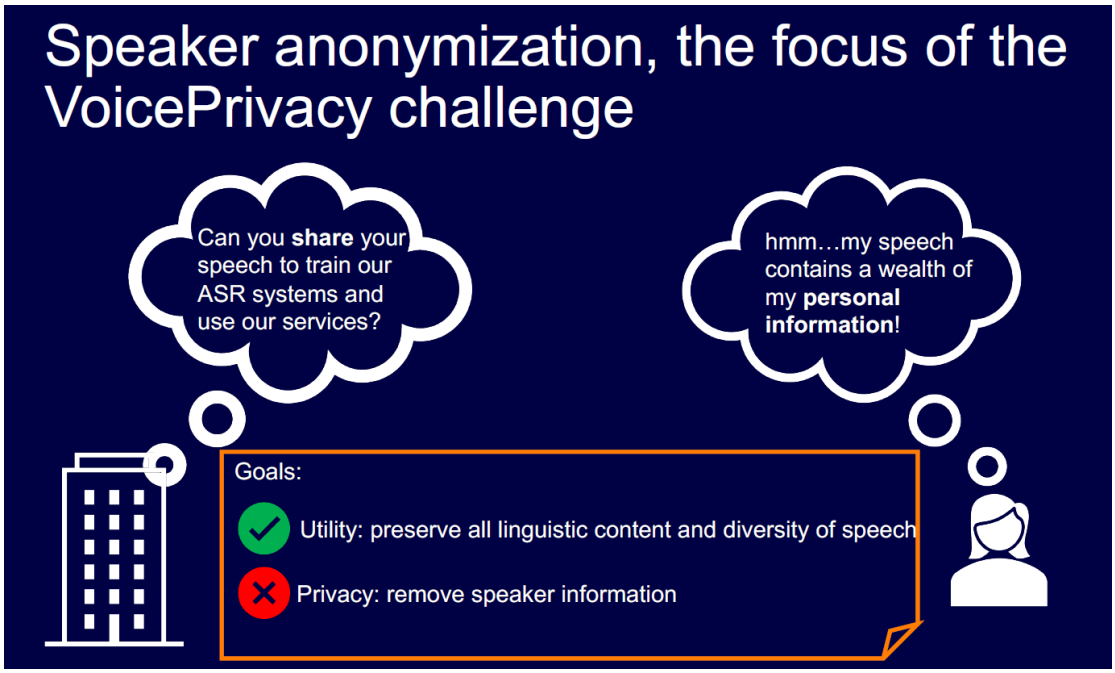


圖 42、Differentially Private Speaker Anonymization

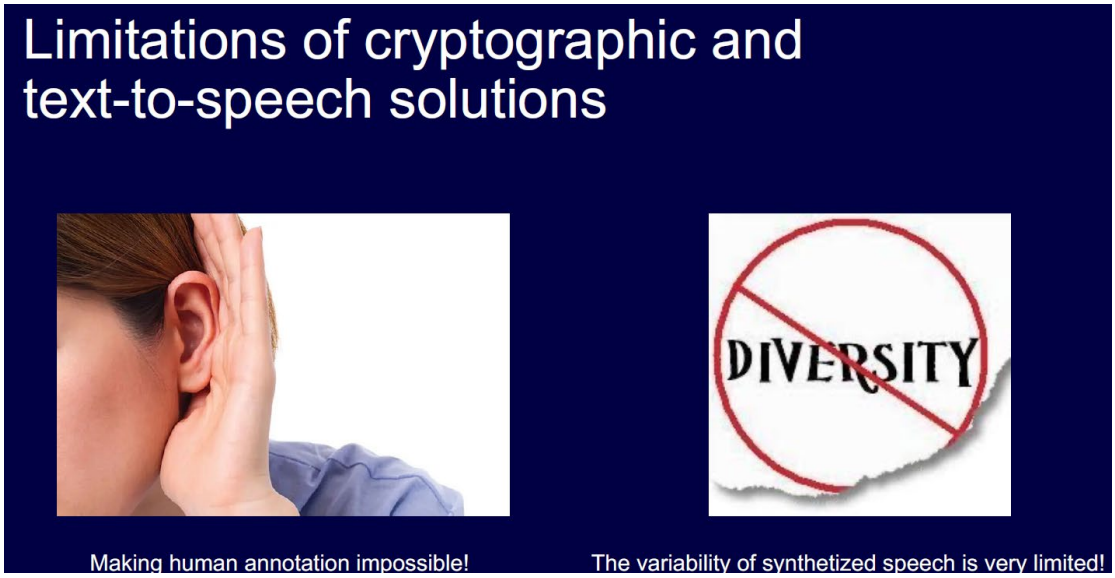


圖 43、Differentially Private Speaker Anonymization

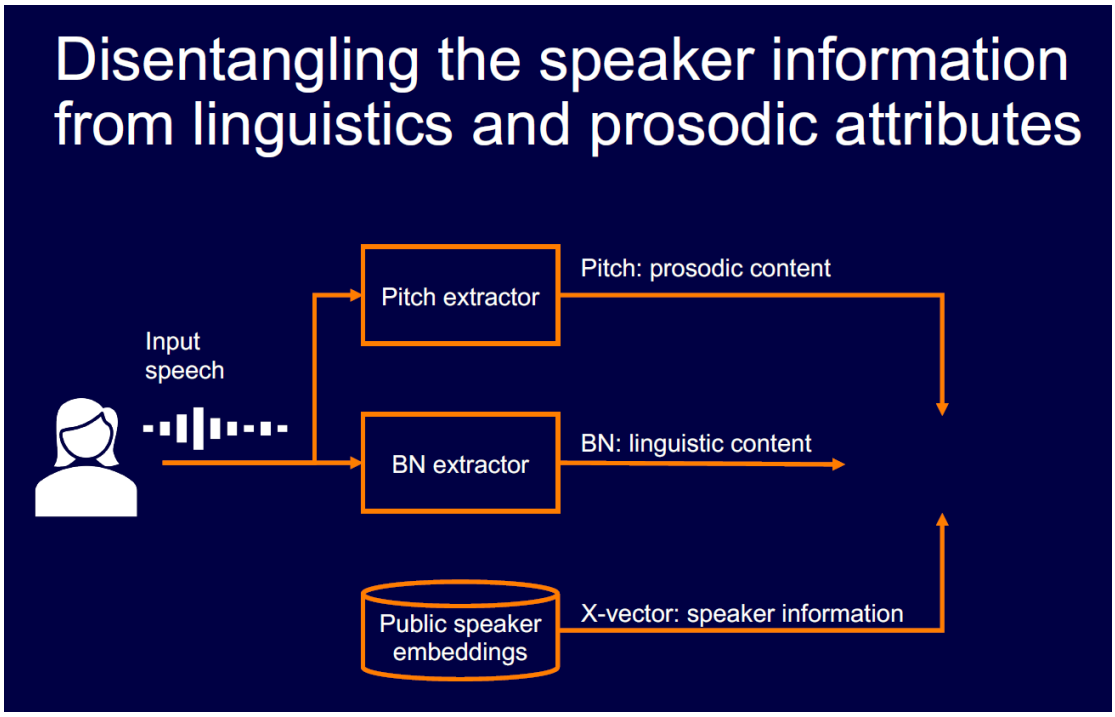


圖 44、Differentially Private Speaker Anonymization

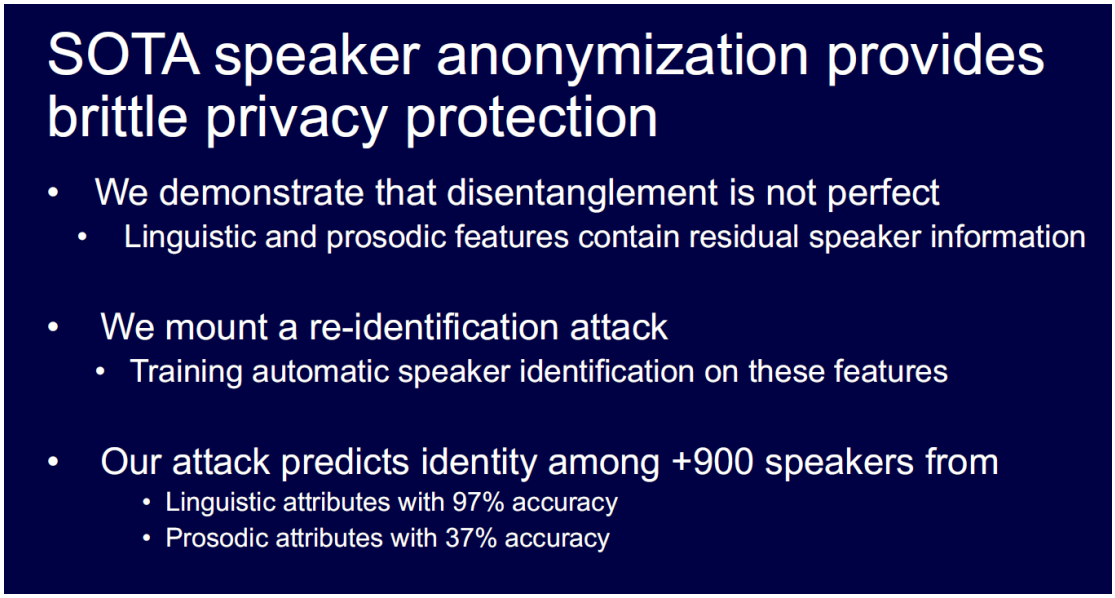


圖 45、Differentially Private Speaker Anonymization

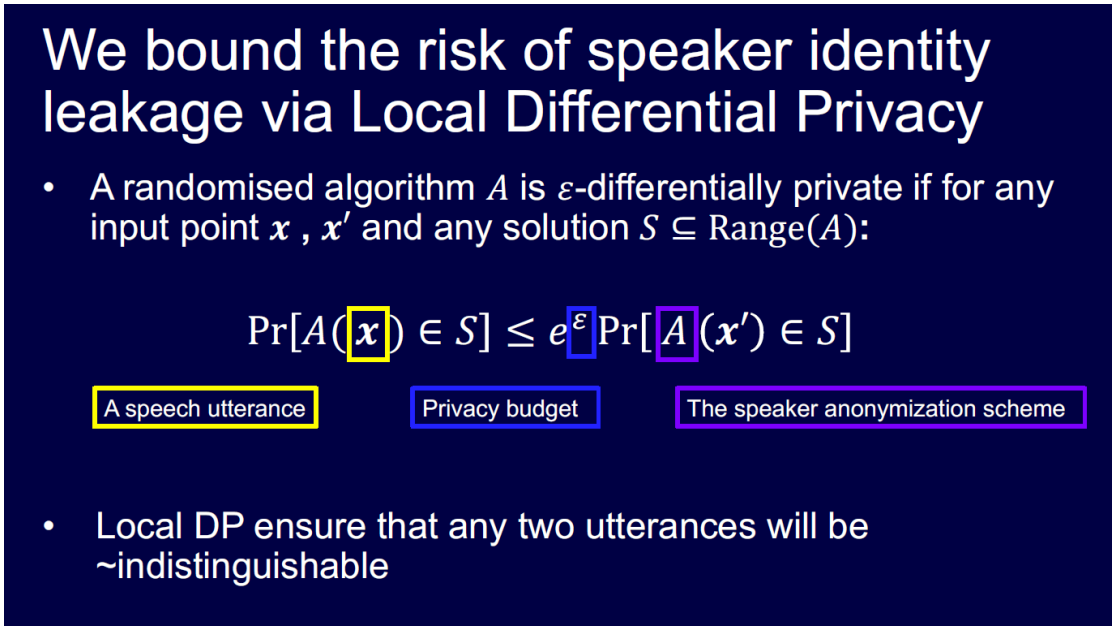


圖 46、Differentially Private Speaker Anonymization

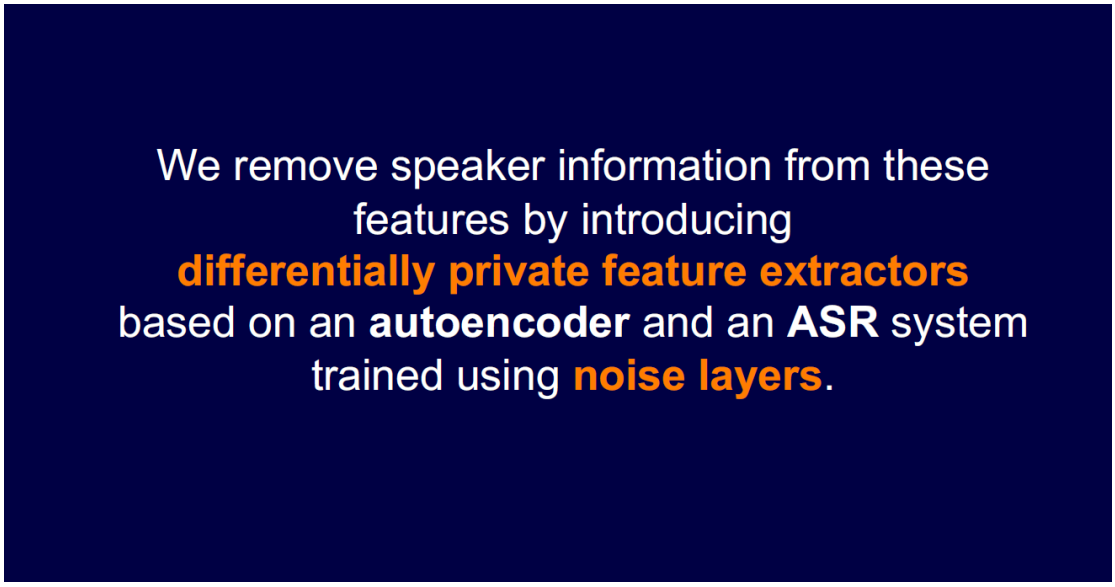


圖 47、Differentially Private Speaker Anonymization

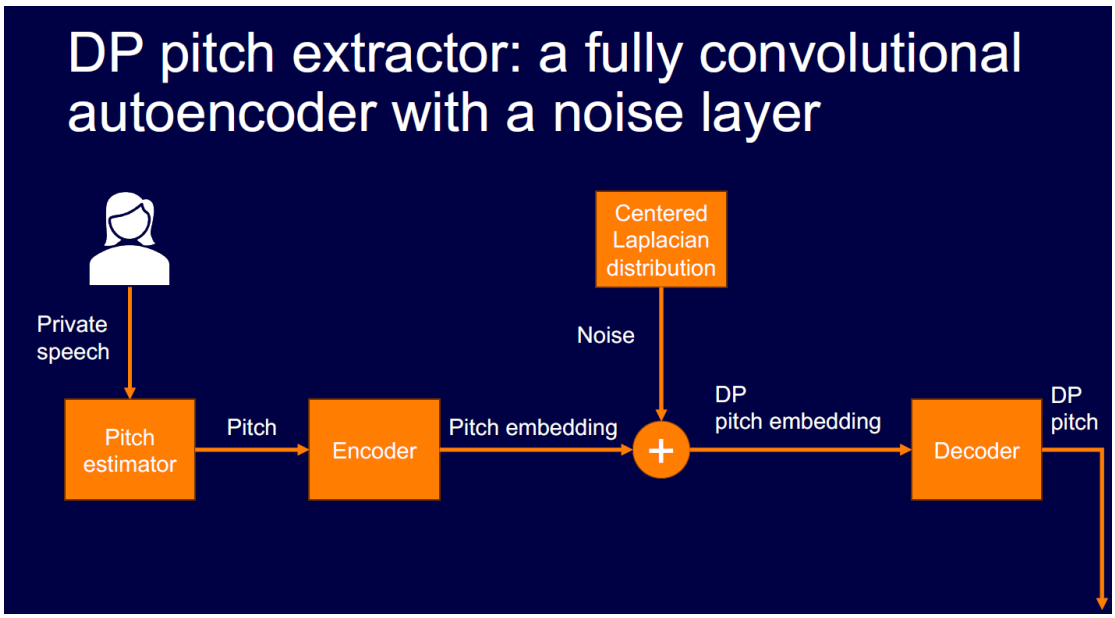


圖 48、Differentially Private Speaker Anonymization

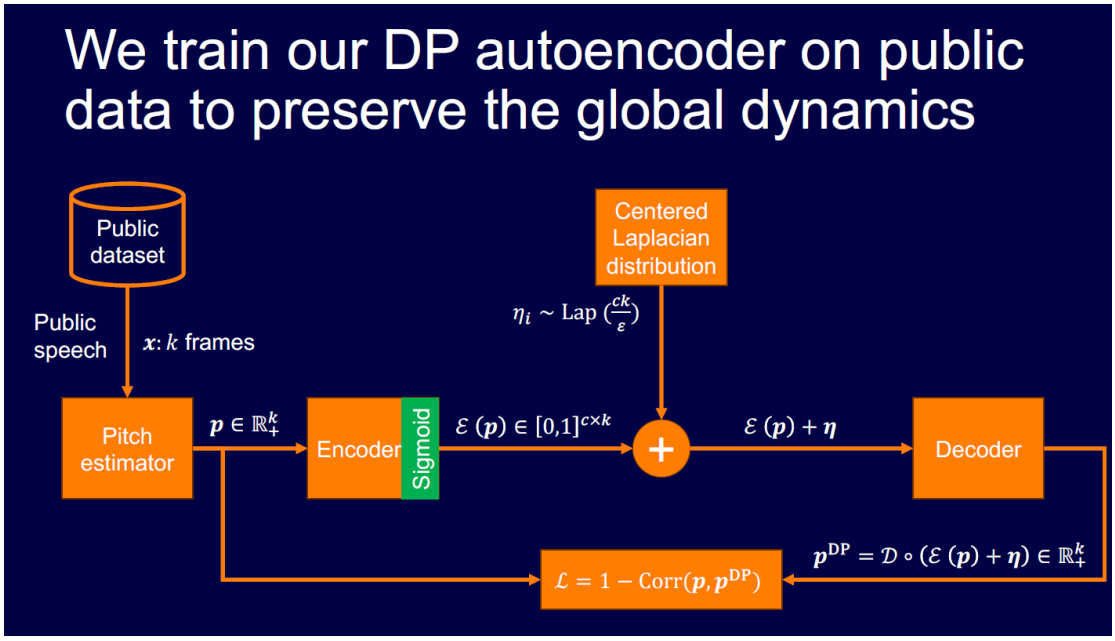


圖 49、Differentially Private Speaker Anonymization

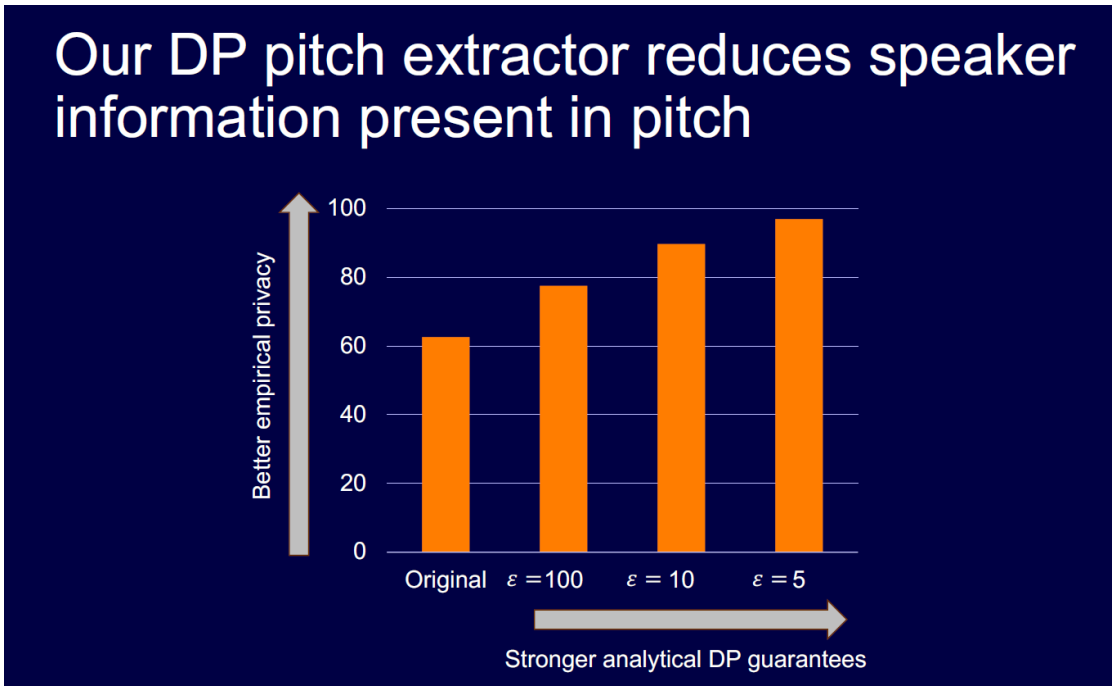


圖 50、Differentially Private Speaker Anonymization

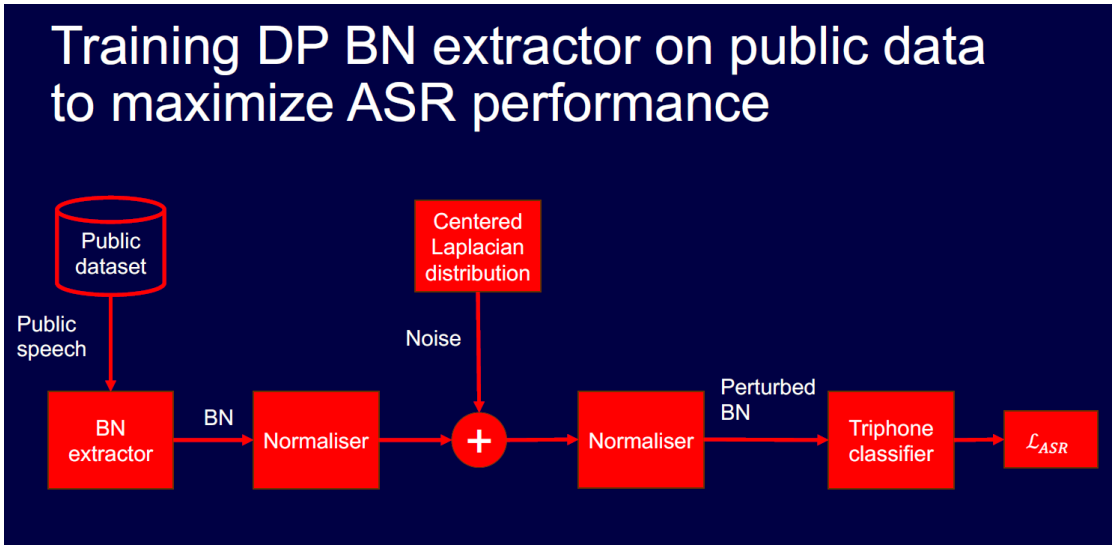


圖 51、Differentially Private Speaker Anonymization

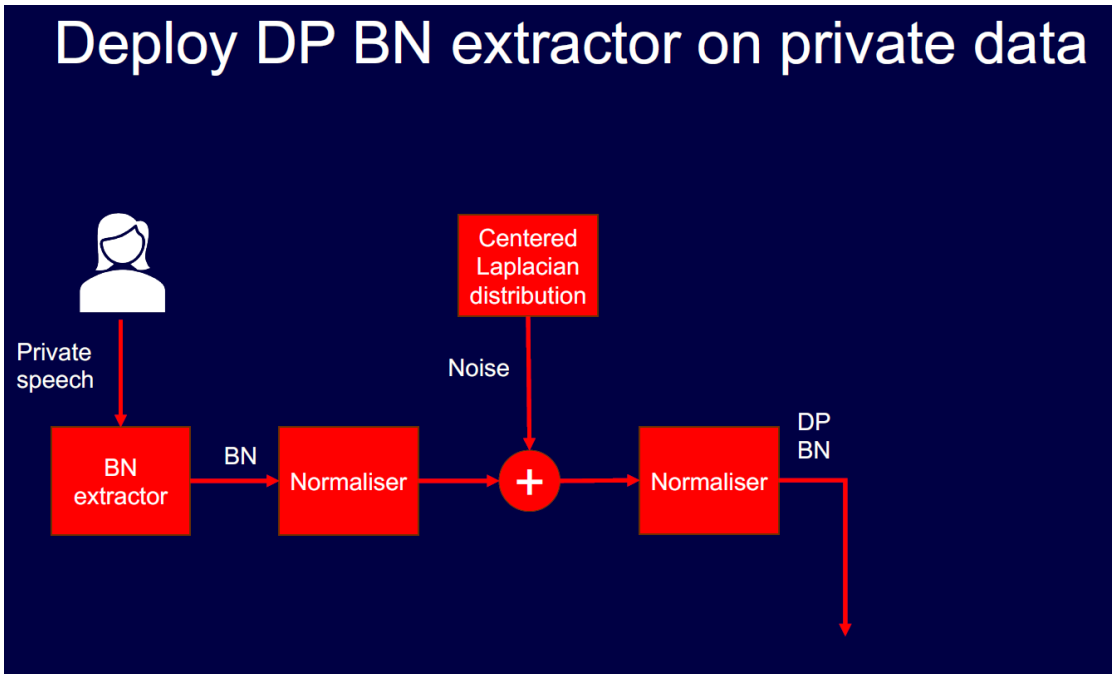


圖 52、Differentially Private Speaker Anonymization

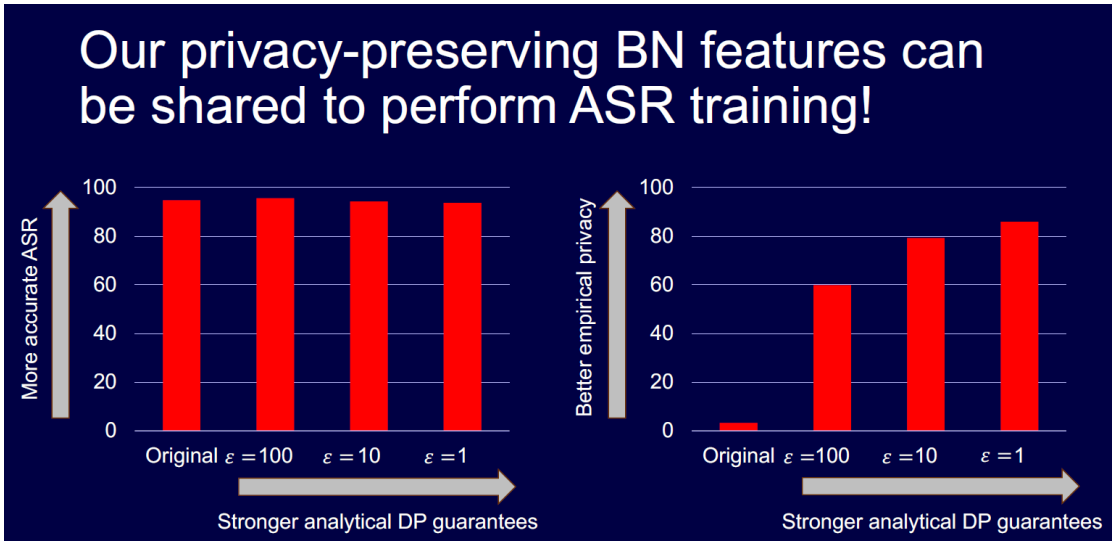


圖 53、Differentially Private Speaker Anonymization

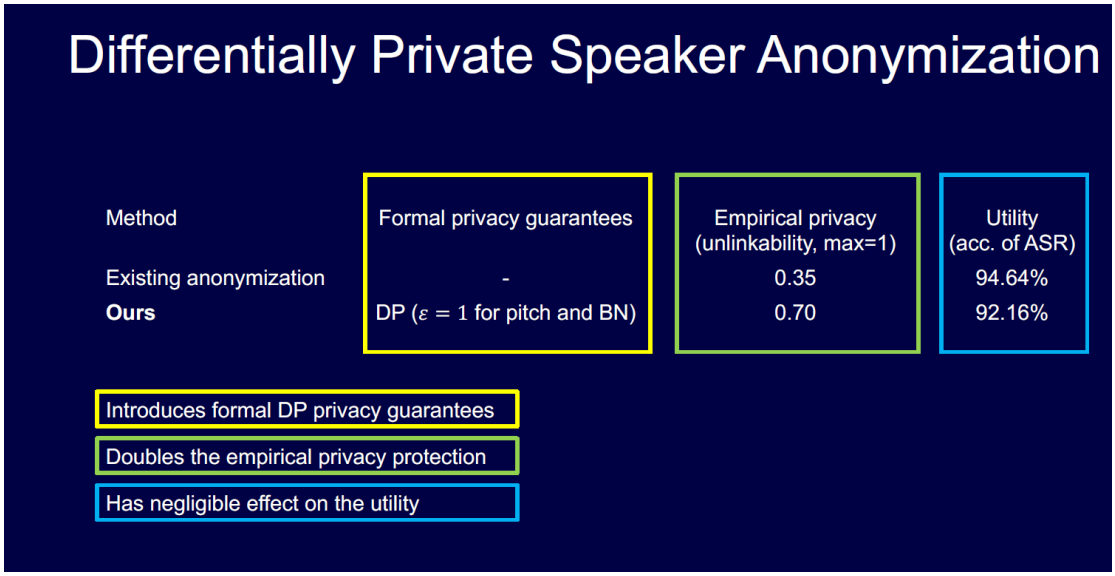


圖 54、Differentially Private Speaker Anonymization

圖 55、Creative beyond TikToks: Investigating Adolescents Social Privacy Management on TikTok

圖 56、On the Role and Form of Personal Information Disclosure in Cyberbullying Incidents

圖 57、Track You: A Deep Dive into Safety Alerts for Apple AirTags

圖 58、SoK: Differentially Private Publication of Trajectory Data