Privacy Mediators:

Helping IoT Cross the Chasm

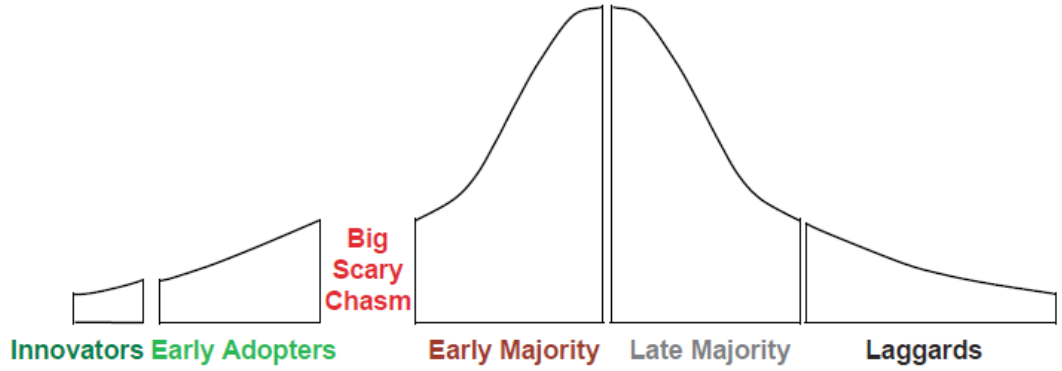
一、 動機

隨著物聯網（IoT）的快速發展，許多生活上常見的物品都有了感知技術與網路連線能力。從手錶到冰箱甚至汽車，萬物皆能連上網路，並透過即時數據的收集，能主動在適當的時機與地點提供所需要的資訊。物聯網的高速發展大大地改變人類的生活習慣，但也衍伸出了新的議題——隱私。在「Crossing the Chasm」中，Geoffrey Moore 曾提出警告，當一個新科技欲由發展早期的小衆市場（科技愛好者、擁有 IoT 需求的公司）往主流市場邁進時，和早期的愛好者不同，主流市場的使用者會因新科技所可能帶來的問題而卻步，讓新科技的發展面臨挑戰，物聯網正面臨這個挑戰。

物聯網裝置爲了能無時無刻收集使用者資料以便分析，生活周遭遍布了各式各樣的 sensor，而這些 sensor 會皆露出各種使用者可能知道或不知道的習慣、特徵等敏感資料，如：智慧手錶可以蒐集使用者心跳、脈搏等生理數據，許多隱私便在無法控制的情況下揭露出來，而形成了另一種形式的監控。

根據 2015 年 6 月的消費者對物聯網隱私的意見調查顯示，高達 78%的消費者對於 IoT 裝置商可能會利用裝置來蒐集資料，並轉給第三方進行運用的問題有所疑慮。因爲消費者的不安，將使得物聯網市場的擴展受阻，隱私問題也就成爲 IoT 發展中亟需面對的挑戰。

因此作者於本篇提出一個系統架構，用以保護物聯網設備的使用者隱私，讓使用者的隱私揭露程度能由自己來控制，從原先的「是否願意提供資訊」轉變爲「願意提供多少資訊」。



二、 議題

在隱私方面，對於 IoT 裝置的持續監測及儲存 sensor 歷程記錄於雲端中，都讓使用者感到憂慮。根據調查，使用者最在意的部分爲「由使用者決定服務或應用程式看到多少資料」以及「資料能夠與誰共享」。使用者希望能擁有控制這些 IoT 裝置在蒐集、上傳內容的權力，並決定分享內容及限定分享對象。因此，可將隱私控管分爲以下六項需求：

(1) Deletion and Denaturing：

使用者應能夠針對他們產生的數據片段進行 denaturing(變性)和刪除，因此系統必須提供相關選項，將敏感的數據部分進行預先處理（模糊化、修改、刪除）後再行發布。例如：有兒童出現的影片中，對兒童臉部加上馬賽克防止身分曝光。同時，使用者也得以控制裝置蒐集資訊的時段，而非讓裝置無時無刻皆在進行資訊的蒐集。

(2) Summarization(摘要)：

讓裝置僅提供數據簡單、摘要性的內容，而非鉅細靡遺地將使用者所有資訊上傳至雲端。摘要性資料又可區分爲時間性摘要資訊與空間性摘要資訊。前者爲裝置蒐集的資訊並非每分每秒的詳細資訊，而是在一長段時間區間中的摘要資訊，如：智慧型手錶量測脈搏的數據，僅發布當天最大與最脈搏數，而非每分鐘測量值。而後者通常代表位置的摘要，如：由裝置分享的 GPS 數據並非完整詳細資訊，可能僅是概略、具代表性的內容(如:郵遞區號)。

(3) Inference(可推論性)：

使用者需要能夠控制數據可能被使用的方式，例如：溫度與光的 sensor 能夠幫助了解房間中的居住情況，由數據可推論出如:房間居住者的步行時間等資訊。因此，使用者需要在資訊發布前進行一些控制，使其無法進行推論。

(4) Anonymization(匿名性)：

使用者可能更偏好在匿名的情況下，提供對社會有助益的數據，例如：匿名後才將數據提供給醫學研究，因此系統設計須滿足資訊提供者的匿名性。

(5) Mobility(移動性)：

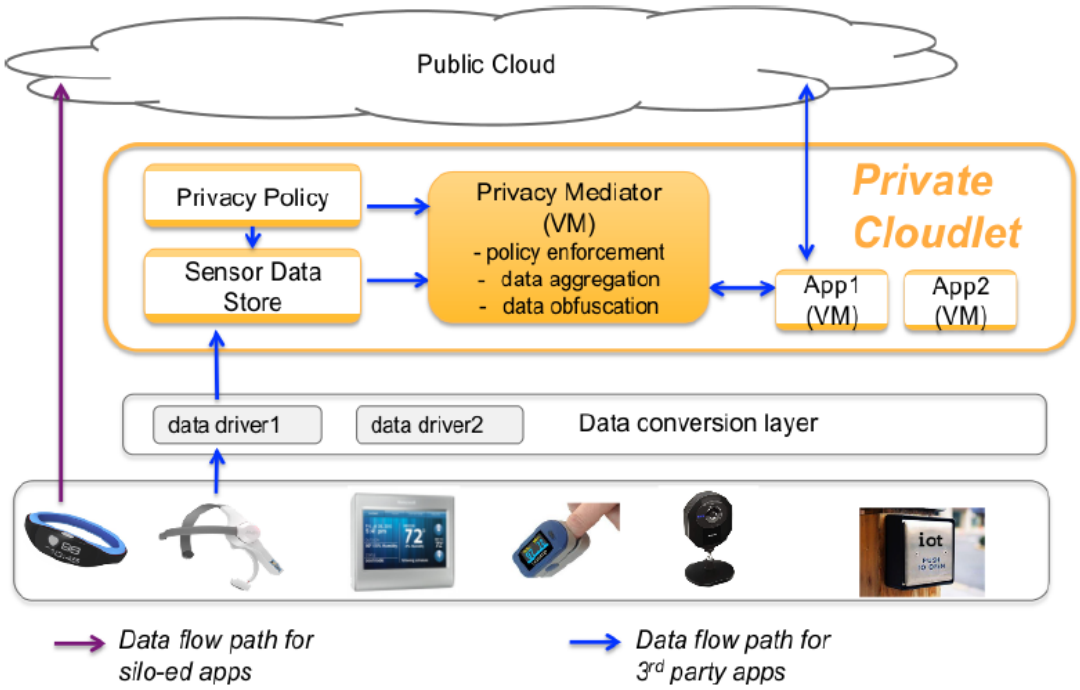
使用者可能會希望即便遇到了自家以外的公共場合中的裝置，也能滿足前述幾點需求，以得到上傳資訊的控制權。

(6) Ease of use(易用性)：

由於根據研究顯示，使用者對於隱私控制通常不太了解，因此需要使用在給定的顆粒度(granularity)下，對於隱私控制有意義的語意來表示，令隱私控管相關操作容易上手，讓一般使用者能快速理解並有效利用相關權力控管自身隱私。

三、 研發方向

本文作者設計的架構核心便是利用布建於各個 network edge 的 cloudlets，以其做爲 local 的小型 data center，負責 denaturing 及資料的儲存。爲了消弭使用者對於物聯網數據隱私上的疑慮，作者提出了一個將 **Privacy Mediators**（爲具有可信的軟體的 plug-in）插入 data distribution pipeline 的解決方案。所有被傳送至使用者信賴之 cloudlets 上的資料都會經過 Privacy Mediators，並依據使用者定義的隱私政策(policy)對數據進行編輯、操作。Privacy Mediators 會如同防火牆一般，將不符合政策的部分進行阻擋，而儲存於 cloudlets 上的資料則會透過一組隨機產生的 private key 進行加密，當需要運用時才將其解密。方案架構如下圖所示，以下將逐一說明方案中的各架構：



(1) Cloudlets：

Cloudlets 是一個位於 internet edge 的小型數據中心，靠近 sensor 和 mobile device，並且座落於使用者的信任域中。

(2) Privacy Mediator：

Privacy Mediator 負責進行各類型的隱私控制，可針對單一類別的 IoT sensor 或許多不同的 IoT sensor 所產生之數據進行操作。

使用者能夠控制 sensor data 到 privacy mediator 的 routing 以及個別 privacy mediator 的配置策略，並且過濾發布的數據、依據使用者的 granularity 配置儲存 sensor data、對 sensor data 進行發布前的預先處理（模糊、刪除、修改、denaturing）等。

(3) Mediator and Trust(信任機制)：

在此方案設計上，當中有個重要的考量要素爲：生產、部署 privacy mediator 與其相關設備的驅動程式。作者認爲 mediator 的生產應有很多家，且須爲獨立的第三方進行開發，而非由提供 sensor 和雲服務的組織來進行生產。否則，仍就會造成使用者在隱私保證上的疑慮，資料仍爲被某些廠商掌握著。作者認爲，只要皆遵循特定資料交換格式，整個系統仍可正常運行。同時，應用程式開發者也僅需要遵循資料格式與通訊協定，便不須太關注隱私系統架構運作(由 mediator 廠商負責即可)，而得以專注於開發應用程式上。此外，作者亦預期將有大量的開源 mediator 出現供大家運用與檢視。

(4) Data Storage：

每個 mediator 需維護自己的數據緩衝區，不提供 mediator 存取在 cloudlet 上的所有 IoT 數據。這樣的方式能夠減少 IoT 架構上的數據風險，當 rough mediator 存在時，只危及該 rough mediator 能夠存取的數據。

數據刪除部分，在一些僅發布 summary data 系統中，當數據一發布後便可立即刪除相對應的原始數據，或在有限時間內暫存在 cloudlet 中。

(5) User Policy：

使用者可以維護或建立，那些套用於應用程式或 IoT 設備的隱私配置文件，讓 mediator 來進行使用者的 IoT 隱私政策。

爲了讓使用者方便建立此隱私配置文件，作者也提出可參考一篇能夠尋找一份捕獲許多使用者偏好，並且讓使用者自定義隱私配置文件的研究，能夠幫助使用者產生隱私配置。

當中，Privacy mediators 是依據使用者所訂定的 privacy control policy 來進行不同的處理。因此必須設計一套標準——「**policy language**」。用於記錄使用者的需求，並同時能讓 privacy mediators 解讀其內容，如同在規定網站隱私的 P3P 及 APPEL。然而，若要制訂新的 policy language 也將會面臨新的挑戰。

第一，因 IoT 服務、裝置大量且分散於各地，若使用者突然遇到新的應用服務時，可能會因爲各處 cloudlets 的更新速度不同，而使新服務無法立刻生效。除此之外，sensor 不若智慧型手機，無法訂定如此繁多、複雜、客製化的政策。因此作者認爲，應將裝置分層級，依據不同層級以訂定不同的政策，來精簡政策訂定對象的範圍。

第二，使用者或許會希望依據不同的資料接收者，而提供不同的隱私政策。如：使用者可能較爲願意提供完整電力資料給電廠，然而僅願意提供摘要性資料給政府。

四、 未來預測

在本文中也提出另一種使用情境：透過將作者提出的架構應用於企業或組織中，由管理者來設定隱私政策，可以讓 IoT 的使用更無疑慮，從而移除讓 IoT 發展受阻的隱私疑慮因素。如：在擁有攝影機的 IoT 設備的教室中，透過避免發布兒童臉部的隱私政策，能夠讓影像數據在發布前先進行兒童臉部的模糊化。而在醫院的使用情境中，透過此架構能夠讓醫生在保護病患隱私的同時，發布 summary data 來幫助改進醫學研究。另一方面，若作者所提出的 mediator 概念能被實現，加上 edge computing 技術的普及，智慧程式、智慧家庭等概念便可更容易落實於生活之中，物聯網技術便可以更快速、隱私的方式另我們擁有更好的生活。

然而隨著 IoT 裝置快速普及，將有越來越多 sensor 遍布於我們生活當中。在裝置將資料送上 cloudlets 進行 denaturing 時，如何在隱私性與資訊價值間達到平衡，以及面對選擇性的隱私政策，都考驗著架構的設計與相關輔助技術的效能。

此外，或許我們也應設想當 IoT 相關資訊落於有心人士之中(如:中間人攻擊、擷取)，透過大數據分析與人工智慧預測的方式，我們會否也暴露於危險之中。因此不僅隱私，資訊安全中的保密性(Confidentiality)、完整性(Integrity)、可用性(Availability)，也皆是值得注意的地方。