What do Information Centric Networks, Trusted Execution Environments, and Digital Watermarking have to do with Privacy, the Data Economy, and their future?*

Nikolaos Laoutaris
IMDEA Networks Institute, Spain
nikolaos.laoutaris@imdea.org

Costas Iordanou
Cyprus University of Technology
costas.iordanou@eecei.cut.ac.cy

ABSTRACT

What if instead of having to implement controversial user tracking techniques, Internet advertising & marketing companies asked explicitly to be granted access to user data by name and category, such as Alice→Mobility→05-11-2020? The technology for implementing this, already exists, and is none other than the Information Centric Networks (ICN) developed for over a decade in the framework of Next Generation Internet (NGI) initiatives. Beyond named access to personal data, ICN's in-network storage capability can be used as a substrate for retrieving aggregated, anonymized data, or even for executing complex analytics within the network, with no personal data leaking outside. In this opinion article, we discuss how ICNs combined with trusted execution environments and digital watermarking can be combined to build a personal data overlay inter-network in which users will be able to control who gets access to their personal data, know where each copy of said data is, negotiate payments in exchange for data, and even claim ownership, and establish accountability for data leakages due to malfunctions or malice. Of course, coming up with concrete designs about how to achieve all the above will require a huge effort from a dedicated community willing to change how personal data are handled on the Internet. Our hope is that this opinion article can plant some initial seeds towards this direction.

CCS CONCEPTS

• **Networks** → **Network privacy and anonymity**; • **Information systems** → **Web services**; *Online advertising*; • **Computing methodologies** → **Distributed computing methodologies**; *Machine learning*; • **Computer systems organization** → **Distributed architectures**; • **Security and privacy** → *Cryptography*; **Access control**; **Privacy-preserving protocols**; *Database and storage security*; **Distributed systems security**; **Privacy protections**;

KEYWORDS

Information Centric Networking, Overlay Network, Personal Data Management System, Online Privacy

1 INTRODUCTION

The Internet and the Web were built decades ago to fulfill objectives and requirements far different from? those of today. Both have managed to exhibit tremendous evolvability and extendibility and have succeeded in supporting services and capabilities that could hardly be imagined in the 60s (Internet) or even the 90s (Web). This has mostly been achieved via *layering*, *standardization*, and *openness*. Every time that new applications (content distribution, video conferencing) or capabilities (broadband connectivity, mobility) become critically important, additional layers are introduced on the top or the bottom of the protocol stack to implement the necessary new functionality and requirements. Such layers are often standardized via the open and flexible standardization process of the Internet (IETF, RFCs, etc.), or are at least made unofficial de facto standards via the support of large corporations and groups. Many thousands of independent companies, public and private organizations, and government, collaborate in all the above activities.

A landmark moment on the history of the Internet and the Web was the appearance of online advertising and marketing. This sector grew at tremendous rate, starting from a simple advertising banner for AT&T data services in Hot Wired in '94 [14], to becoming an entire industry that has overtaken broadcast and print advertising [49], and is currently funding a large part of the so-called free services of the Internet [12]. Of course, with online advertising, came user tracking and all the data protection and privacy problems that have challenged the Internet and the Web during their online advertising era [43].

2 TECHNOLOGICAL CONSEQUENCES

Like many other things, the Internet and the Web were not designed to support online advertising, especially advanced versions of it, such as targeted (or behavioral) advertising [9, 37], in which marketing offers are optimized on a per user basis based on complex recommendation and auctioning algorithms driven by detailed personal data collected online from millions of individuals.

2.1 Advertising sector

The digital advertising industry has built an immensely complex business ecosystem [47] and set of tools, protocols, and layers for empowering user tracking and advertising on a massive scale. These include third party tracking cookies and complex digital fingerprinting techniques [29], the Real Time Bidding (RTB) [7] and Cookie Synching protocols [19], scores of proprietary user profiling, campaign planning, auctioning, and advertisement delivery and rendering mechanisms and algorithms. Part of all this complex data collection and advertising apparatus has been standardized (e.g., cookies and RTB), while at the same time, lots of it remains proprietary, evolving, non-transparent, and highly fragmented.

2.2 Privacy response and its three major shortcomings

Looking at the other side of the coin, data protection and privacy attempts are even less standardized, and even more fragmented

*Title inspired by Jaron Lanier's, "Who Owns the Future?" [42].

than the data collection and tracking approaches that they attempt to counter-balance. Consensus and consolidation exist only at the legislative layer, mainly due to GDPR [23] and similar attempts elsewhere outside Europe. At the technology layer, however, privacy and protection tools for citizens are hopelessly fragmented, ad hoc, and proprietary. There exist tracking and advertising blocking solutions [1, 28, 55], privacy-preserving browsers [5, 13] & virtual private networks [6, 25], personal information management systems (PIMS) [20], and scores of data marketplaces [15]. Each one of the above attempts, is trying to deliver some level of compromise between privacy and the personal data consumption needs of the advertising and marketing ecosystem.

What is common to all these attempts, is that they all fail to encompass one or more of the fundamental principles that drove the growth of the the Internet and the Web. Starting with *openness*, we see that it is nowhere to be found. Most often, an individual company, be it an ambitious startup, or a large established corporation, aspires and promises to solve on its own, all the complex privacy and data protection challenges of our times. This is in stark contrast to how the Internet and the Web came to be, and how they evolved in an open manner that invited the participation of different stakeholders such as ISPs, backbone networks, publishers, and web-domains. Even the advertising ecosystem [24], to which privacy solutions are supposed to be a counterweight, is more open and encompassing, by allowing various stakeholders such as Demand Side Platforms (DSPs), Supply Side Platforms (SSPs), Advertising Exchanges, and Data Management Platforms (DMPs) [57] to collaborate for the delivery of targeted advertising.

Lack of *standardization* follows closely in the list of ways in which data protection and privacy solutions are deviating from the path taken by other successful paradigms of the Web. Whereas the advertising ecosystem has standardized mechanisms such as RTB [7] and Cookie Sync [19], the privacy protection side has failed to standardize any meaningfully successful mechanisms. Do-Not-Track [21] was an early attempt from W3C, but it never reached maturity, nor was ever widely adopted. The advertising sector has introduced the YourAdChoices [59] mechanism for explaining to users why they see a particular display advertisement. Still, the mechanism, although standardized, provides very limited transparency in terms of why its explanations are what they are and whether they are sufficient and/or complete [45].

With openness and standardization being in a rather poor state, *layering*, i.e., implementing a solution as a separate layer of the Internet/Web protocol stack, is nowhere to be seen for the time being.

2.3 How to fix the Web (without having to scrap it first)

From the above discussion, one may be tempted to conclude that with privacy, the Internet and the Web have finally "met their master" and that the end of their evolvability has been reached. Thus, maybe we should be heading back to the draw-boards to design a new Web and Internet around a new set of primitives pertaining to fundamental aspects such as identity, directionality of links, and connection initiation, in order to be better prepared to address modern privacy challenges. Indeed, a magnitude of data protection, privacy, ownership of personal information, and other related challenges we face today have their roots in fundamental design choices of the past. For example, had the web been built around bidirectional links (e.g., like the original Xanadu hypertext proposal of Ted Nelson in the 60s [51]), it would be possible for an individual to release a piece of information, and know who links to it, who has seen it, and how many times. This would allow individuals to revoke released data, negotiate compensations based on the number of times that their data have been used, define exclusion lists, etc. Stronger/more flexible notions of identity for users and devices, as well as stronger trust models for devices and protocols would also help in addressing many privacy problems.

The Internet and the Web, however, are by now too big, too important, and too expensive to be decommissioned in favor of better alternatives. Technological revolutions are, unfortunately, more expensive than scientific ones [39]. What is needed, therefore, is a revolutionary approach to how privacy and personal data management are conducted, but in a way that remains incrementally deployable on top of the existing web. Such an approach would have to address the following challenges:

Challenge 1: Allow users to make their data available to third parties in a controlled manner. This challenge has two parts: a) make it easy for third parties to locate their data, b) allow users to set precise rules about who, and under which circumstances can have access to said data. Currently, a) is done via tracking, i.e., implicitly, inefficiently, and non-transparently. b) is almost impossible to do, or reduces into the crude all-or-nothing logic of ad- and tracking-blockers such as AdBlockPlus [1], uBlock [55], and Ghostery [28], among others.

Challenge 2: Allow users to negotiate and receive compensation for their data in a fair, usage-based, data economy. This is important for multiple reasons: (a) the direct value for users, i.e., the income stream that it can create, and its connection with wider societal issues like the future of labor in the times of automation and machine learning [42, 44], (b) payments and prices are additional, beyond-technology, means of avoiding negative and parasitic behaviors, such as information over-collection, unethical use of collected information, etc. To build sound economics of information, it is imperative to be able to count credibly how many times a piece of information has been used and by whom. This is currently impossible to do in today's web.

Challenge 3: All of the above needs to be built on top of the existing Web using robust and well-thought paradigms and technologies. The list of fundamental difficulties and challenges for designing an additional layer for the web to handle personal data are so many that starting everything from scratch does not seem like a good or efficient idea to pursue. Instead, we should strive to re-use tried and trusted building blocks as individual components of the solution. **Challenge 4: Minimize user involvement and cognitive load.** The most successful technologies are those that minimize the cognitive load and effort that they require from adopters. Labor-intensive tasks like data ingestion, configuration, and management should be as automated as possible, while always remaining under the control of users.

Challenge 5: Follow Internet and Web governance and expansion principles, including openness, standardization, and

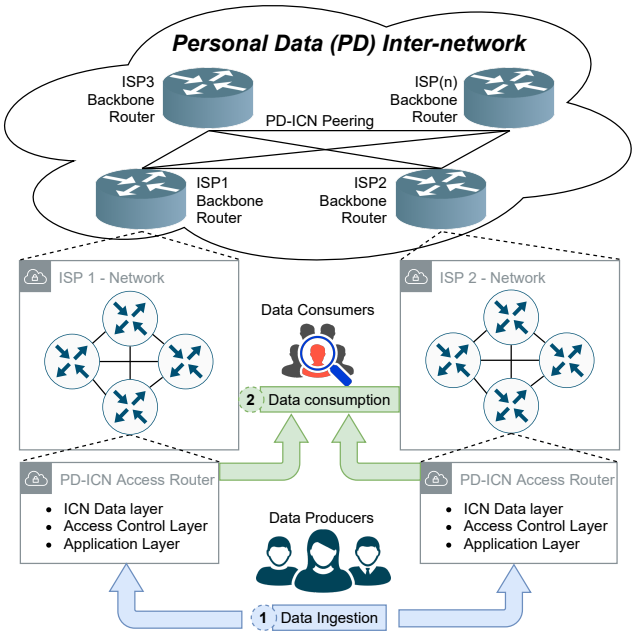

Figure 1: PD-ICN Overlay Network and Basic Interactions.

layering. The problem is so big that we should not expect that a single company, or a "wall-gardened" ecosystem, will solve it for everyone. This is what existing approaches are attempting, be it data marketplaces, personal information management systems (PIMS) [20], consent management solutions [33], etc. Even if one such company managed to become the one-stop-shop for privacy, we would, at best, be left with yet another monopoly on the Internet. Instead, we should aim to have a personal data inter-network that follows the plurality of the Internet (composed of independent autonomous systems (AS)) and the Web (composed of independent domains). In the same fashion, independent communities of users could interconnect to control, protect, and even trade their data, as opposed to all handing them over to a single, or few global entities.

3 HOW TO BUILD AN INTER-NETWORK FOR PERSONAL DATA

In this section we will discuss a number of concepts for addressing the challenges presented in the previous section. We will discuss these concepts using as basis the high level architecture of a hypothetical inter-network for personal data that we will call PD-ICN.

The main idea behind PD-ICN is a Private by Design (PD) overlay network on top of the existing Web that will allow users to share with services personal data in a controlled and fair manner. Given its overlay nature, and with the right type of APIs, PD-ICN can be incrementally deployed over the current Web, in which personal data are "shared" implicitly via tracking, rather than explicitly as is our intention with PD-ICN.

The high-level architecture of PD-ICN is depicted in Figure 1. PD-ICN *access routers* hosted and owned by different Internet Service Providers (ISPs) are used as the entry point of personal data & access rules from users. Such PD-ICN access routers could be collocated with access routers used for IP connectivity. Each ISP also has one or more PD-ICN *backbone routers* to interconnect with different ISPs. Backbone PD-ICN routers can be co-located with backbone IP routers used in peering between ISPs for basic connectivity. PD-ICN backbone routers will implement personal data reachability between ISPs, i.e., they will allow data consumers in one ISP to locate and retrieve data from data producers in a second ISP.

Each PD-ICN router is based on Information Centric Networking (ICN) principles [2] that allow named personal data to propagate to other PD-ICN access routers of the same ISP, or even to remote ISPs (PD-ICN backbone routers) via personal data peering (the PD equivalent of connectivity peering).

At the bottom of Figure 1 we depict the Data Producers (Owners) and the Personal Data Consumers (e.g., Ad Tech, Digital Marketing Companies) which are the end-users of the PD-ICN platform. The Data Producers can interact with the system by (1) inserting their personal data and the desired access rules to their ISP's PD-ICN access router. The Personal Data Consumers can then issue (2) an Access Request to any PD-ICN access router in the overlay network to get access to personal data. Depending on the type of the request, the access rules of individual users and the credibility of the requester, the PD-ICN access router will (3) reply back to the requester with the appropriate information. Note that the request can be received on a different PD-ICN access router and the final data included in the response may need to be replicated from multiple PD-ICN access routers including PD-ICN backbone routers across different ISPs.

3.1 Driving concepts

In this subsection, we present the various concepts and considerations that have motivated PD-ICN's architecture.

Concept 1: Proactive rather than reactive. Instead of letting personal data leak in an uncontrolled manner (e.g., through cookies, digital finger-printing, device identifiers) [22] and then trying to figure out who has collected them, where they are, and how have they been used, it is better to instead be proactive, and control data spread before it happens. When data is allowed to flow, it should happen only towards well known and vetted Personal Data Consumers that have accepted clear terms & conditions prior to being given access to the PD-ICN platform.

Concept 2: Named Personal Data. Using ICN principles solves multiple problems, including (1) finding personal data and routing requests using human understandable names, (2) storing personal data close to where they are needed, i.e., close to Personal Data Consumers that issue frequent queries against them, including delay sensitive ones. Via ICN naming, an individual consumer can request information about a particular named user, such as `Alice` → `Mobility` → `05-11-2020`, or an unnamed one, such as `WomanBelow30[1]` → `Mobility` → `05-11-2020`, in which case `WomanBelow30[1]`, is just an alias for the first member of an audience comprising users that are women with age below thirty years. Such audiences can be selected and cached by running appropriate analytics at the applications layer of PD-ICN, as will be explained later.

Concept 3: In-network personal data processing. The initial CCN proposal [26] and its various ICN derivatives [50] have advocated in-network caching of popular content, such as video, images, and other content. PD-ICN will take the concept one step further, and not only cache personal data in the network, but also allow its processing for the extraction of useful analytics, such as audience creation, without having to let the data leave the trusted environment of PD-ICN. The system, of course, will also allow copying and extracting raw data from the platform, but only if, and only if, the policy file of a user allows it.

Concept 4: Trusted access control, personal data hosting, and distributed counting. Instead of having arbitrary entities allowed to set third-party cookies, and other tracking code, at the devices of users, PD-ICN will allow only vetted Personal Data Consumers, that have registered with the platform and have passed a thorough vetting procedure, to connect to the platform and explicitly ask for specific types of end-user personal data. Then depending on the policy file of a user, the request will be granted or not. If it is granted, the access will be logged, for monitoring and (optionally) data access payment computation. Data access can be kept inside the confines of PD-ICN, with all the processing happening at the application layer, or may lead to an actual copy of the data being taken outside the platform, if the initial request indicates so, and the policy file of the user permits it. Personal data copies, including the original one, and its replicas, can be in cleartext or encrypted. In the latter case, prior to any access, key-exchange and/or renewal may need to take place. All PD-ICN routers will run on top of trusted execution environments to make sure that they do not deviate from the protocol and, e.g., leak data without consulting policy files, fail to update access counters, etc. By contrast, in the current web, nobody knows where, and how far, collected personal data have been transmitted [38]. Even if data gets leaked in PD-ICN through a compromised router, watermarks will be used to identify the point of leakage a posteriori.

Concept 5: Personal data policy automation and one-stop monitoring and control. Instead of putting a high cognitive load on users in order to set complex privacy policy settings, or interrupting them constantly with accept/reject decisions, notifications, and pop-ups, PD-ICN will be equipped with common predefined privacy policies appropriate for most users, and with machine learning algorithms for improving them with use. Additional policy import & export functionalities will allow users to re-use highly optimized privacy policies constructed by experts or crowdsourced based on common behaviors observed across large populations of users. Similarly, simple intuitive dashboards will keep users informed about who is using their data, in which geographic areas, what type of data are used the most, including access counts per data item, potential offers and payments received for releasing data, and other useful information for re-assuring users about the use of their data and allowing them to adjust their privacy policies accordingly (including revoking granted accesses to some data, or releasing new data that was not previously available).

Concept 6: Incremental deployment / Coexistence. PD-ICN is a revolutionary approach to protecting citizens' privacy without starving socially indispensable and valued online services (including but not limited to advertising, search, real time traffic monitoring, etc.) of the data that they run upon. As such, PD-ICN has been designed to be useful from day 1 of deployment and interoperable with the current advertising and tracking approaches, which it aspires to eventually substitute. The deployment and proliferation of PD-ICN can mimic more or less the path taken by protocols such as HTTPS [31] and Quic [41] and their symbiotic relationship with HTTP [30] and TCP [10]. Basically, whenever a tracker attempts to set or read a cookie, the tracking server can be automatically redirected to PD-ICN. If the tracker does not support PD-ICN, then the user can set in her policy file the default action, e.g., accept, reject, or forward the decision to the local ad/tracking blocker. In terms of user base, PD-ICN does not require that many or a majority of users use it before it becomes useful (which is the case for example for messaging, or social network services). Even with a single PD-ICN access router online, Personal Data Consumers will already be able to access the user base behind this first PD-ICN access router (e.g., the users of an ISP). If a second ISP adopts PD-ICN as well, then it will only take for a personal data peering to be established between the two PD-ICN backbone routers to allow Personal Data Consumers from ISP1 to request data from users behind ISP2. In effect, PD-ICN is mimicking the peering and interconnection paradigm used by the Internet for connectivity. This is a process that can grow gradually, based on easy to establish bilateral agreements, and without requiring any cumbersome global consensus for the system to get bootstrapped.

Concept 7: Enhanced security and fewer data breaches via homogenization. Having one type of personal data router makes it easier to build stronger security and thus protect from intrusion and data breaches compared with having one's personal data stored in a myriad of different systems, each one pursuing its own security policy.

Concept 8: Bootstrapping using already collected data. Having PD-ICN access routers be provided by data controllers that already manage data on behalf of large populations of data owners (users), has the additional advantage of easing the bootstrapping process, by pre-populating the platform with already collected data. Of course, this requires having the consent of users, but data controllers such as telecommunication companies and banks, are in much better position to elicit and obtain such a consent from large populations of users.

3.2 Architecture and open questions

In this section we discuss a series of open architectural and systems questions for implementing PD-ICN in practice. We do that using as basis the three different layers depicted in Figure 2.

3.2.1 Data Layer (DL). The Data Layer provides the core functionality of the PD-ICN router to handle all low-level functionalities such as data routing and replication, data encryption and storage and trusted code execution. Following Figure 2 (center box), next we provide details for each component of the Data Layer and how these components address the different challenges and concepts.

Trusted Computing Module: It is responsible to verify the software stack of a PD-ICN access router participating on the overlay network. This will allow other PD-ICN backbone routers from other ISPs to remotely attest the authenticity of the remote PD-ICN router and its software. In this way, PD-ICN routers will continuously check that remote routers do not deviate from the protocol

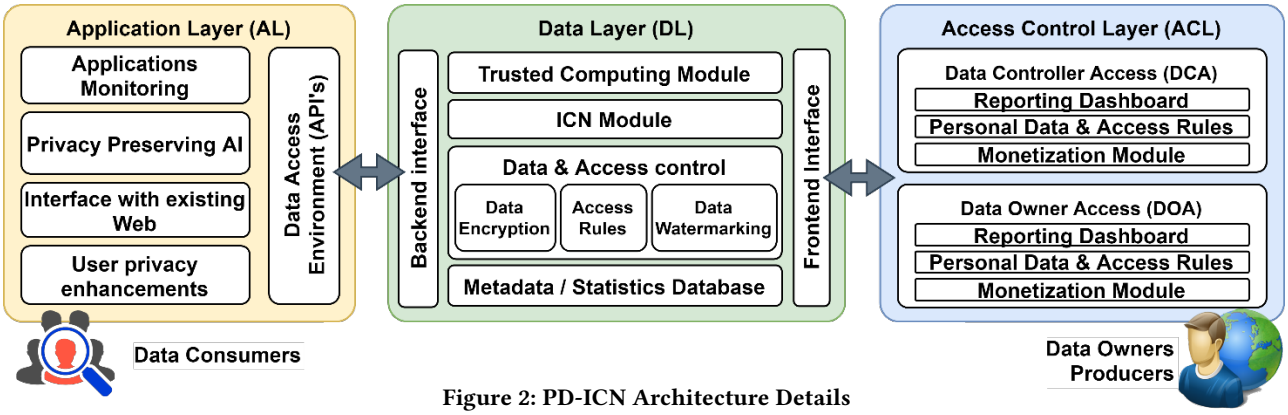

Figure 2: PD-ICN Architecture Details

by, e.g., leaking data outside the system, not respect the privacy policies of different users, or keep unfaithful counts of how many times each piece of personal data has been requested by different Personal Data Consumers.

Open Questions: In order to implement such functionality one needs to decide which parts of the PD-ICN software stack need to be executed inside a trusted enclave [11]. Also, the right technology for remote attestation needs to be picked (hardware, software, hybrid, or virtualization-based, such as docker [17]). Such decisions will depend a lot on whether PD-ICN nodes will be realized as standalone routers, or will be software driven hosted at a datacenter/IXP.

ICN Module: It encapsulates all low-level networking functionalities such as Naming, Routing, Replication and Storage. Naming & Routing defines a unique Uniform Resource Identifier (URI) for the personal data storage of each user participating in the PD-ICN, while Replication & Storage includes all the logic related to personal data storage and replication across all PD-ICN routers upon request from other PD-ICN routers.

Open Questions: Different privacy preserving naming schemes already exists based on different encryption schemes, such as, hash chain [58] and symmetric encryption [3, 48], nevertheless, due to the complex functionality of PD-ICN, two open question need to be answered. (1) How to construct a general purpose naming scheme for personal data that allows all the different functionalities that PD-ICN provides? (2) What adaptations (of any) of generic ICN routing and caching algorithms should be applied for handling personal data?

Data & Access Control: This module includes all required functionality related to the raw personal data. In total, this module comprises three submodules (Data Encryption, Access Rules and Data Watermarking):

- **Data Encryption:** This submodule handles all the details pertaining to data encryption, such as, private and public keys, access tokens, cryptographic algorithms, privacy preserving aggregation algorithms, etc. The raw personal data of each user are encrypted as soon as the user provides data to the system. For each personal data attribute (i.e. age, gender, interests, etc.) this module provides individual access tokens in order to allow partial decryption and sharing of such attributes, allowing the traceability of each attribute leading to data provenance and reporting. Access tokens are granted if and only if the corresponding access rules of each user allows it.

  Open Questions: Data encryption in ICN networks is heavily depended on the Naming and Routing scheme [3, 48, 53, 58]. In addition, in PD-ICN we also have to deal with additional constrains. (1) What type of encryption scheme to use in order to protect personal data? (2) Should it be one size-fits-all, or should the encryption level depend on the type of personal data? The answer to the above questions should take into consideration both aspect, personal data protection and the ICN naming and routing part of PD-ICN. Additional approaches from different domains should also be considered [32, 35, 40].
- **Access Rules:** The core component of this submodule is the Access Control Language (ACL), which is responsible to translate and enforce the user rules on the corresponding personal data. The ACL supports a hierarchical data structure scheme in order to provide flexibility and expandability of the system.

  Open Questions: Instead of building a new ACL from scratch, another approach is to use existing languages from different domains, such as, distributed systems [56]. All-in-all, the open questions for this submodule are: (1) What type of access control language to use for allowing users to express their privacy policies? (2) Do access rules need to be propagate across all PD-ICN nodes or should they stay at the user's entry node?
- **Data Watermarking:** This module allows PD-ICN to track and safeguard personal data belonging to its users and choose to share them with other entities in a raw format. In such cases the PD-ICN access router will first apply watermarking techniques to the data prior to the release and create a record in the metadata database (see below) with the corresponding attributes that the user allowed to be shared, the name of the entity that will receive the data, and finally the signature of the watermark(s). In the case of data leakage from the intended recipient, the PD-ICN can identify which external entity is responsible for the leakage and the affected users.

  Open Questions: Since PD-ICN needs to provide fine grain analytics for different type of personal data, which in some cases can be a single attribute of a user (i.e. age, gender, etc), an open question is: (1) How to watermark data that, unlike media files, are smaller in size and typically contain non binary data such as strings and numbers?

Metadata / Statistics Database: It holds all the logs and statistics related to the personal data, including access tokens statistics, watermarking logs, access rules definitions, etc. Each PD-ICN access router is responsible to store and maintain up-to-date logs and statistics for all the local users that have the specific PD-ICN access router as their entry point to the overlay network. All access requests towards such users are evaluated and granted access only by the entry point PD-ICN access router of each user in order to allow consistency and traceability.

3.2.2 Access Control Layer (ACL). This layer is the main interface between Data Owners and Data Controllers. It allows Personal Data Ingestion, either by individual Data Owners, or in bulk by Data Controllers. It provides the visual interfaces and dashboards to report information statistics and other information as needed.

The Data Controller (DCA) and Data Owner (DOA) Access Layers allows data controllers and data owners to interact with the system, respectively. This layers provide functionalities such as reporting (Reporting Dashboard), easy manipulation of access rules on personal data (Personal Data & Access Rules) and monetization functionalities (Monetization Module) pertaining the financial value of personal data and the financial compensation of the end users.

Open Questions: With respect to personal data monetization, we need to address the following open questions: (1) What type of data monetisation principles to implement? (2) Should it be seller initiated fixed price for data? (3) Auction-based? (4) Hybrid schemes depending on the fidelity of data? Should we also consider privacy preserving (targeted) advertising schemes [8, 18, 52, 54]?

3.2.3 Application Layer (AL). This layer implements the appropriate interfaces, APIs, algorithms and environment to allow Personal Data Consumers to interact with personal data and extract useful information required for their daily business activities.

The Application Monitoring module allows the development, deployment and monitoring of different applications running on top of personal data stored within PD-ICN Overlay Network. Similarly, the Privacy preserving AI includes a set of algorithms, Machine Learning (ML) frameworks and privacy preserving data structures to provide anonymous data statistics (i.e. Male -> Interest.Sports.('Tennis') -> Location('Italy') -> Percent() => 2%). It exposes easy to use ML and aggregation APIs to the application layer. For the smooth integration of PD-ICN with the existing Web and end users the "Interface with Existing Web" module and "User Privacy Enhancements" module are responsible to interact with the existing monetization model of the Web and enhance end users privacy, respectively. Finally, the "Data Access Environment (APIs)" acts as a direct link between the backend interface of the Data Layer and the Application Layer

Open Questions: Different approaches can be found in the literature pertaining anonymous data analysis though aggregation techniques [16, 27, 60] and Secure, Privacy Preserving Machine Learning (SPP-ML) [4, 34, 36, 46], focusing on different aspects of SPP-ML such as ML as a Service (MLaaS) [36], ML with multiple data providers [46] and verifiable outsourcing schemes [34]. The open questions under the suggested PD-ICN setting are: (1) How to develop a *"sandboxed"* environment (or combine existing solutions) to execute ML algorithms on top of PD-ICN data at the application layer? (2) Can we implement such a solution without copying any raw data outside the trusted environment of PD-ICN?

4 CONCLUSION

In this article we have sketched the architecture of an overlay for disseminating personal data in a controlled manner. Via this design, we have argued that ICN, TEE, and DW are fundamental technologies for implementing PD-ICN, and repeating the benefits discussed in Section. 3.1. Each one of these technologies opens up new possibilities for making the next step in personal data management and addressing the challenges listed at the beginning of the article. For Information Centric Networking in particular, we believe that PD-ICN may be a great opportunity for re-purpusing all the great work that has taken place in the area in the last 10 years. The magnitude of the challenge, partially reflected in the list of open questions that we have provided, will require work from an entire community dedicated to building PD-ICN. Our hope is that this article will serve as a call to arms for building this new important layer of the Internet.

REFERENCES

[1] adblockplus.org. Adblock Plus. Surf the web without annoying ads! https://adblockplus.org/, 2020. [Online: accessed 4-August-2020].

[2] Bengt Ahlgren, Christian Dannewitz, Claudio Imbrenda, Dirk Kutscher, and Börje Ohlman. A survey of information-centric networking. *IEEE Communications Magazine - IEEE Commun. Mag.*, 50:26–36, 07 2012.

[3] M. Bilal and S. Pack. Secure distribution of protected content in information-centric networking. *IEEE Systems Journal*, 14(2):1921–1932, 2020.

[4] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, page 1175–1191, New York, NY, USA, 2017. Association for Computing Machinery.

[5] Brave. Brave - Much more than a browser. https://brave.com/, 2017.

[6] Brave. A Privacy-Preserving Distributed Virtual Private Network. https://brave.com/vpn0-a-privacy-preserving-distributed-virtual-private-network/, 2019.

[7] Interactive Advertising Bureau. OpenRTB API Specifications. https://www.iab.com/wp-content/uploads/2016/03/OpenRTB-API-Specification-Version-2-5-FINAL.pdf, 2016. [Online: accessed 29-November-2020].

[8] Hui Cai, Fan Ye, Yuanyuan Yang, Yanmin Zhu, and Jie Li. Towards privacy-preserving data trading for web browsing history. In *Proceedings of the International Symposium on Quality of Service*, IWQoS '19, New York, NY, USA, 2019. Association for Computing Machinery.

[9] Juan Miguel Carrascosa, Jakub Mikians, Ruben Cuevas, Vijay Erramilli, and Nikolaos Laoutaris. I always feel like somebody's watching me: Measuring online behavioural advertising. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*, CoNEXT '15, New York, NY, USA, 2015. Association for Computing Machinery.

[10] Vinton G. Cerf and Robert E. Icahn. A protocol for packet network intercommunication. *SIGCOMM Comput. Commun. Rev.*, 35(2):71–82, April 2005.

[11] Guoxing Chen, Yinqian Zhang, and Ten-Hwang Lai. Opera: Open remote attestation for intel's secure enclaves. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, CCS '19, page 2317–2331, New York, NY, USA, 2019. Association for Computing Machinery.

[12] Eric Clemons. Business models for monetizing internet applications and web sites: Experience, theory, and predictions. *Journal of Management Information Systems*, 26:15–41, 09 2009.

[13] Cliqz GmbH. Cliqz - The no-compromise browser. https://cliqz.com/, 2018.

[14] Karla Cook. A Brief History of Online Advertising. https://blog.hubspot.com/marketing/history-of-online-advertising, 2020. [Online: accessed 18-October-2020].

[15] datarade.ai. Data marketplaces & exchanges. https://datarade.ai/platform-categories/data-marketplaces-exchanges, 2020.

[16] Amit Datta, Marc Joye, and Nadia Fawaz. *Private Data Aggregation over Selected Subsets of Users*, pages 375–391. Springer, Cham, 10 2019.

[17] Marco De Benedictis and Antonio Lioy. Integrity verification of docker containers for a lightweight cloud environment. *Future Generation Computer Systems*, 97, 02 2019.

[18] Erdong Deng, H. Zhang, Peilin Wu, Fei Guo, Z. Liu, Haojin Zhu, and Z. Cao. Pri-rtb: Privacy-preserving real-time bidding for securing mobile advertisement in ubiquitous computing. *Inf. Sci.*, 504:354–371, 2019.

[19] developers.google.com. Cookie Matching. https://developers.google.com/authorized-buyers/rtb/cookie-guide, 2020. [Online: accessed 29-November-2020].

[20] European Data Protection Supervisor (EDPS). Personal Information Management Systems (PIMS). https://edps.europa.eu/data-protection/our-work/subjects/personal-information-management-system_en, 2020. [Online: accessed 29-November-2020].

[21] Electronic Frontier Foundation EFF. Do Not Track. https://www.eff.org/issues/do-not-track, 2017. [Online: accessed 18-October-2017].

[22] Steven Englehardt and Arvind Narayanan. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, page 1388–1401, New York, NY, USA, 2016. Association for Computing Machinery.

[23] EU. The EU General Data Protection Regulation. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&qid=1490179745294&from=en, 2016. [Online: accessed 29-November-2020].

[24] Content European Commission, Directorate-General for Communications Networks and Technology. An overview of the Programmatic Advertising Ecosystem Opportunities and Challenges. https://platformobservatory.eu/app/uploads/2019/10/An-overview-of-the-Programmatic-Advertising-Ecosystem-Opportunities-and-Challenges.pdf, 2019. [Online: accessed 29-November-2020].

[25] ExpressVPN. The vpn that just works. https://www.expressvpn.com/, 2020.

[26] Internet Engineering Task Force. Ccnx 0.x first software release. https://wiki.fd.io/view/Cicn, 2009.

[27] Badih Ghazi, Pasin Manurangsi, Rasmus Pagh, and Ameya Velingker. Private aggregation from fewer anonymous messages. In Anne Canteaut and Yuval Ishai, editors, *Advances in Cryptology – EUROCRYPT 2020*, pages 798–827, Cham, 2020. Springer International Publishing.

[28] ghostery.com. Ghostery - Cleaner, faster, and safer browsing. https://www.ghostery.com/, 2020. [Online: accessed 4-August-2020].

[29] Alejandro Gómez-Boix, Pierre Laperdrix, and Benoit Baudry. Hiding in the crowd: An analysis of the effectiveness of browser fingerprinting at large scale. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, page 309–318, Republic and Canton of Geneva, CHE, 2018. International World Wide Web Conferences Steering Committee.

[30] Internet Engineering Task Force Network Working Group. Hypertext transfer protocol – http/1.1. https://www.w3.org/Protocols/rfc2616/rfc2616.html, 1999.

[31] Internet Engineering Task Force Network Working Group. Http over tls. https://tools.ietf.org/html/rfc2818, 2000.

[32] T. Hardjono and A. Pentland. Data cooperatives: Towards a foundation for decentralized personal data management. *ArXiv*, abs/1905.08819, 2019.

[33] Denuwanthi Hasanthika. Consent Management. https://medium.com/identity-beyond-borders/consent-management-5e5a6f4e28a9, 2020. [Online: accessed 29-November-2020].

[34] A. Hassan, R. Hamza, H. Yan, and P. Li. An efficient outsourced privacy preserving machine learning scheme with public verifiability. *IEEE Access*, 7:146322–146330, 2019.

[35] Nurmamat Helil and Kaysar Rahman. Cp-abe access control scheme for sensitive data set constraint with hidden access policy and constraint policy. *Security and Communication Networks*, 2017:1–13, 09 2017.

[36] Ehsan Hesamifard, Hassan Takabi, Mehdi Ghasemi, and Rebecca N. Wright. Privacy-preserving machine learning as a service. *Proceedings on Privacy Enhancing Technologies*, 2018(3):123 – 142, 01 Jun. 2018.

[37] Costas Iordanou, Nicolas Kourtellis, Juan Miguel Carrascosa, Claudio Soriente, Ruben Cuevas, and Nikolaos Laoutaris. Beyond content analysis: Detecting targeted ads via distributed counting. In *Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies*, CoNEXT '19, page 110–122. Association for Computing Machinery, 2019.

[38] Costas Iordanou, Georgios Smaragdakis, Ingmar Poese, and Nikolaos Laoutaris. Tracing Cross Border Web Tracking. In *Proceedings of ACM IMC 2018*, Boston, MA, October 2018.

[39] Thomas Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, 1962.

[40] J. Kurihara, E. Uzun, and C. A. Wood. An encryption-based access control framework for content-centric networking. *2015 IFIP Networking Conference (IFIP Networking)*, pages 1–9, 2015.

[41] Adam Langley, Alistair Riddoch, Alyssa Wilk, Antonio Vicente, Charles Krasic, Dan Zhang, Fan Yang, Fedor Kouranov, Ian Swett, Janardhan Iyengar, Jeff Bailey, Jeremy Dorfman, Jim Roskind, Joanna Kulik, Patrik Westin, Raman Tenneti, Robbie Shade, Ryan Hamilton, Victor Vasiliev, Wan-Teh Chang, and Zhongyi Shi. The quic transport protocol: Design and internet-scale deployment. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, SIGCOMM '17, page 183–196. Association for Computing Machinery, 2017.

[42] Jaron Lanier. *Who Owns the Future?* SIMON & SCHUSTER, 2013.

[43] N. Laoutaris. Data transparency: Concerns and prospects. *Proceedings of the IEEE*, 106(11), 2018.

[44] N. Laoutaris. Why online services should pay you for your data? the arguments for a human-centric data economy. *IEEE Internet Computing*, 23(5):29–35, 2019.

[45] Pedro Giovanni Leon, Justin Cranshaw, Lorrie Faith Cranor, Jim Graves, Manoj Hastak, Blase Ur, and Guzi Xu. What do online behavioral advertising privacy disclosures communicate to users? In *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, page 19–30, New York, NY, USA, 2012. Association for Computing Machinery.

[46] Ping Li, Tong Li, Heng Ye, Jin Li, Xiaofeng Chen, and Yang Xiang. Privacy-preserving machine learning with multiple data providers. *Future Generation Computer Systems*, 87:341 – 350, 2018.

[47] lumapartners.com. Luma provides resourceful content that delivers insightful and thoughtful observations on the complex and dynamic digital media and marketing ecosystem. https://lumapartners.com/luma-content/, 2020. [Online: accessed 29-November-2020].

[48] Michele Mangili, Fabio Martignon, and Stefano Paraboschi. A cache-aware mechanism to enforce confidentiality, trackability and access policy evolution in content-centric networks. *Comput. Netw.*, 76(C):126–145, January 2015.

[49] Price Waterhouse and Coopers (PwC). IAB Internet Advertising Revenue Report 2010 Full Year Results. https://www.iab.com/wp-content/uploads/2015/05/IAB_Full_year_2010_0413_Final.pdf, 2011. [Online: accessed 29-November-2020].

[50] Wikipedia the free encyclopedia. Information-centric networking. https://en.wikipedia.org/wiki/Information-centric_networking, 2020.

[51] Wikipedia the free encyclopedia. Project Xanadu. https://en.wikipedia.org/wiki/Project_Xanadu, 2020. [Online: accessed 18-October-2020].

[52] V. Toubiana, A. Narayanan, D. Boneh, H. Nissenbaum, and Solon Barocas. Adnostic: Privacy preserving targeted advertising. In *NDSS*, 2010.

[53] R. Tourani, S. Misra, T. Mick, and G. Panwar. Security, privacy, and access control in information-centric networking: A survey. *IEEE Communications Surveys Tutorials*, 20(1):566–600, 2018.

[54] Theja Tulabandhula, S. Vaya, and Aritra Dhar. Privacy-preserving targeted advertising. *ArXiv*, abs/1710.03275, 2017.

[55] ublock.org. uBlock - The Fastest, Most-Powerful Ad Blocker. https://ublock.org/, 2020. [Online: accessed 4-August-2020].

[56] Frank Wang, Ronny Ko, and James Mickens. Riverbed: Enforcing user-defined privacy constraints in distributed web services. In *Proceedings of the 16th USENIX Conference on Networked Systems Design and Implementation*, NSDI'19, page 615–629, USA, 2019. USENIX Association.

[57] Jun Wang, Weinan Zhang, and Shuai Yuan. Display advertising with real-time bidding (rtb) and behavioural targeting, 2017.

[58] Kaiping Xue, Peixuan He, Xiang Zhang, Qiudong Xia, David S. L. Wei, Hao Yue, and Feng Wu. A secure, efficient, and accountable edge-based access control framework for information centric networks. *IEEE/ACM Trans. Netw.*, 27(3):1220–1233, June 2019.

[59] youradchoices.com. YourAdChoices Gives You Control. http://youradchoices.com/, 2020. [Online: accessed 4-August-2017].

[60] Q. Zhou, X. Qin, G. Liu, H. Cheng, and H. Zhao. An efficient privacy and integrity preserving data aggregation scheme for multiple applications in wireless sensor networks. In *2019 IEEE International Conference on Smart Internet of Things (SmartIoT)*, pages 291–297, 2019.